

**METODOLOGÍA DE INVESTIGACIÓN EN CIENCIAS  
SOCIALES**

Gilda Gómez-Peresmitré y Lucy Reidl

U.N.A.M.

# **METODOLOGÍA DE INVESTIGACIÓN EN CIENCIAS SOCIALES**

Gilda Gómez-Peresmitré y Lucy Reidl

U.N.A.M.

## **Indice**

### **I. CIENCIA, MÉTODO Y TEORÍA**

#### **1.1. Conceptos Generales**

##### **1.1.1. Ciencia**

##### **1.1.2. Método**

##### **1.1.3. Teoría**

#### **REFERENCIAS**

### **II. MÉTODOS DE INVESTIGACIÓN**

#### **2.1. Introducción**

#### **2.2. Clasificación de los Métodos de Investigación**

##### **2.2.1. Método Experimental**

###### **2.2.1.1. Experimento de Laboratorio**

###### **2.2.1.2 Experimento de Campo**

###### **2.2.1.3 Experimento Natural**

##### **2.2.2. Expost-Facto y Correlacional**

##### **2.2.3. Método Cuasi-Experimental**

##### **2.2.4 Método No Experimental**

###### **2.2.4.1. Estudio de Campo y Encuesta**

#### 2.2.4.1.1. Tipos de Encuesta

Encuesta Descriptiva

Encuesta Explicativa o Analítica

Encuesta Diagnóstica

Encuesta Predictiva

Encuesta Exploratoria

#### 2.2.5 Diseños de Encuesta

Transversal No Ponderada

Transversal Ponderada

Muestras Contrastadas

Longitudinales

Antes-después

Encuestas de Tendencias

Superposición de Muestras

#### 2.2.5.1 Comentarios adicionales sobre el control en el proceso de investigación

#### REFERENCIAS

### III. EL PROCESO DE INVESTIGACIÓN

#### 3.1. Problema de Investigación

#### 3.2. Proyecto de Investigación

#### 3.3. Desarrollo de la Investigación

##### 3.3.1. Revisión Bibliográfica

##### 3.3.2. Problema

##### 3.3.3. Hipótesis

##### 3.3.3.1. Hipótesis Conceptual

##### 3.3.3.2. Hipótesis de trabajo

##### 3.3.3.3. Hipótesis Estadísticas

3.3.4.Variables

3.3.5.Control de Variables

3.3.5.1.Eliminación

3.3.5.2.Constancia de Condiciones

3.3.5.3.Balanceo

3.3.5.4.Contrabalanceo

3.3.5.5.Aleatorización

3.3.6.Diseño de Investigación

3.3.7.Muestreo

3.3.7.1.Muestras Probabilísticas

3.3.7.1.1.Muestreo al Azar Simple

3.3.7.1.2.Muestra Estratificada al Azar

3.3.7.1.3.Muestreo en Racimo

3.3.7.2.Muestras No Probabilísticas

3.3.7.2.1.Muestras Accidentales

3.3.7.2.2.Muestreo por Cuota

3.3.7.2.3.Muestras Intencionadas

3.3.7.3.Tamaño de la Muestra

3.3.7.3.1. Cálculo del tamaño de la muestra partiendo de porcentajes estimados

3.3.7.3.2.Cálculo del tamaño de la muestra tomando como base medias estimadas

3.3.7.3.3. Tamaño de la muestra tomando como base el valor de la comparación entre medias y su nivel de significancia

3.3.7.3.4.Tamaño de la muestra dependiendo del nivel del efecto esperado de los tratamientos experimentales

3.3.7.3.5. Tamaño exacto de la muestra conociendo la desviación estándar de la población

REFERENCIAS

## IV.DISEÑOS DE INVESTIGACION, PROCEDIMIENTOS ESTADISTICOS NO PARAMETRICOS

4.1.Introducción

4.2.Estadística Paramétrica

4.3.Estadística No Paramétrica

4.4.El Uso de la Estadística en la Investigación

4.4.1.La Hipótesis Nula

4.4.2.La Elección de la Prueba Estadística

4.4.3.El Nivel de la Significancia y el Tamaño de la muestra

4.4.4.La Distribución Muestral

4.4.5.La Región de Rechazo

4.4.6.La Decisión Estadística

4.4.7.Errores que se cometen en la Toma de Decisiones Estadísticas

4.5.Diseño de Investigación y Pruebas no Paramétricas

4.5.1.Diseño de una Sola Muestra

4.5.1.1.Prueba No Paramétrica Para el Diseño de una Sola Muestra :  $X^2$  (Chi Cuadrada)

4.5.2.Diseño de Dos Muestras Independientes

4.5.2.1.Prueba No Paramétrica Para Dos Muestras Independientes : Chi Cuadrada ( $X^2$ )

4.5.2.2.Prueba No Paramétrica Para Dos Muestras Independientes : Prueba de la Mediana

4.5.3.Diseño de Dos Muestras Relacionadas

4.5.3.1.Prueba No Paramétrica Para Dos Muestras Relacionadas : Prueba de Wilcoxon

4.5.4.Diseño de Más de Dos Muestras Independientes

4.5.4.1.Prueba No Paramétrica para k Muestras Independientes Chi Cuadrada  $X^2$

4.5.4.2.Extensión de Prueba de la Mediana

4.5.4.3. Análisis de la Varianza de Una Entrada de Kruskal-Wallis

4.5.5. Diseño de Más de Dos Muestras Relacionadas

4.5.5.1. Prueba No Paramétrica con k Muestras Relacionadas : Análisis de Varianza de Dos Entradas de Friedman

REFERENCIAS

## V. DISEÑOS DE INVESTIGACIÓN: PROCEDIMIENTOS ESTADÍSTICOS PARAMÉTRICOS

5.1. Conceptos y procedimientos Básicos de la Estadística Paramétrica

5.1.1. Medidas de Tendencia Central y de Dispersión

5.1.2 Normas del Uso Adecuado de la Media ( $\bar{x}$ ) y de la Desviación estándar (S)

5.1.3. Grados de Libertad

5.2. Curva normal y sus áreas

5.2.1 Puntajes Z

5.2.2 Error estándar de la media

5. 2.3 Prueba Z e Intervalos de Confianza

5.2.4 Prueba t para una sola muestra

Supuestos básicos de las Pruebas z y t de una sola muestra

5. 3 Revisando los datos: detección de errores y cumplimiento de los supuestos

Ausencia inesperada de la información

Variabilidad inusual

Patrones de respuesta

Normalidad de las distribuciones

5.4 Transformando las distribuciones

5.5 Prueba paramétrica para el diseño de dos muestras independientes: t

5.6 Pruebas t para dos muestras relacionadas:

5.7. Pruebas Paramétricas Para k Muestras Independientes:  
Rangos de Duncan y Análisis de Varianza

5.7.1. Prueba de Rangos de Duncan

5.7.2. Análisis de Varianza

Supuestos del Análisis de Varianza en el Modelo de los efectos fijos

Diferentes Métodos de Comparación de Medias

Comparaciones Ortogonales o Independientes

Decisiones Estadísticas

Comparaciones Múltiples Post- Hoc

Problemas relacionados con la Tasa de Error

Pruebas de Newman-Keuls y Duncan

5.7.3. Prueba Paramétrica Para k Muestras Relacionadas:  
Análisis de Varianza de Muestras con Puntajes Relacionados (F)

5.7.4. Prueba Paramétrica para el Diseño de k Variables Independientes: F. Diseño Factorial

5.7.4.1. Interacción Ordinal

Estimaciones de Varianza en los Diseños Factoriales

5.7.5. Solución Práctica al problema de No-Independencia de las razones F, con un diseño factorial de 2 X 3 X 4

5.7.5.1. Consideraciones Finales

REFERENCIAS

VI. ASPECTOS TEORICOS DE LA MEDICION Y MÉTODOS DE RECOLECCIÓN DE DATOS

6.1. ¿Qué es medir ?

6.2. Niveles y Clases de Medición

- 6.2.1.Escala Nominal
- 6.2.2.Escala Ordinal
- 6.2.3.Escalas de Intervalo o Intervalares
- 6.2.4.Escala de Razón o Proporción
- 6.3. Clases de Medición
  - 6.3.1.Medición Fundamental
  - 6.3.2.Medición Derivada
  - 6.3.3.Medición por Autoridad
- 6.4.Técnicas o Procedimientos de Recolección de Datos
  - 6.4.1.Encuestas y Cuestionarios
    - 6.4.1.1.Planeación de la Encuesta
    - 6.4.1.2.Aplicación de la Encuesta.
    - 6.4.1.3.Tipos de Reactivos de los Cuestionarios
      - 6.4.1.3.1.Información de identificación
      - 6.4.1.3.2.Antecedentes Sociales o Datos Actuales de Tipo Censal
      - 6.4.1.3.3.Pregunta sobre el tópico de la Investigación
        - 6.4.1.3.3.1.Tipos de Pregunta
          - Preguntas abiertas
          - Preguntas Dicotómicas
          - Listas de Chequeo
          - Ordenación de reactivos
          - Preguntas de opción múltiple
          - Escalas de clasificación de tres puntos
          - Escalas de clasificación de cinco puntos

## REFERENCIAS

## VII. CONSTRUCCION DE ESCALAS DE ACTITUD

- 7.1.Escalas de Actitud
- 7.2.Características Operantes de los reactivos



### 7.3.Diferentes Aproximaciones de Escalameinto

#### 7.3.1.Aproximación Centrada en el Sujeto

#### 7.3.2.Aproximación Centrada en el Estímulo o Aproximación de Juicio

#### 7.3.3.Aproximación Centrada en la Respuesta

Actitud

Objeto psicológico

Atributo

Continuo psicológico

Estimulos

Valor escalar

Escala monodimensional

### 7.4.Especificación del Atributo, Continuo o Dimensión de la actitud

#### 7.5.Afirmaciones

##### 7.5.1.Contenido

##### 7.5.3. 7.5.2.Redacción

### 7.6 Diferentes procedimientos de recolección de datos

#### 7.6.1. Entrevista Directiva

#### 7.6.2. Observación Directa

### 7.7.Escalamiento da la Actitud

#### 7.7.1. Afirmación de Hecho

#### 7.7.2. Afirmaciones Favorables y Desfavorables

#### 7.7.3. Afirmaciones Ambiguas

### 7.8. Leyes en el Escalamiento de la Actitud

#### 7.8.1. La Ley del juicio Comparativo

##### 7.8.1.1. Supuestos de la Ley del Juicio Comparativo

#### 7.8.2. Ley del Juicio Categórico

### REFERENCIAS

## VIII. ESCALAS DE ACTITUD: APROXIMACION CENTRADA EN EL ESTIMULO

### 8.1 Método de los pares comparados

#### 8.1.1 Ventajas y desventajas de este Método

### 8.2 Método de los intervalos aparentemente iguales

#### 8.2.1 Ventajas y Desventajas de este Método

### 8.3 Método de los intervalos sucesivos

#### 8.3.1 Análisis de la Dispersión Discriminativa

#### 8.3.2 Prueba de Consistencia Interna

#### 8.3.3 Selección de Afirmaciones, Puntajes de Actitud e Interpretación

#### 8.3.4 Valores Escalares Para Afirmaciones con 50% o Más de los Juicios en el Primer o Último intervalo

#### 8.3.5 Ventajas y Desventajas de este Método

## REFERENCIAS

## IX. ESCALAS DE ACTITUDES: APROXIMACION CENTRADA EN LA RESPUESTA

### 9.1 Método de los Rangos Sumarizados

#### 9.1.1 Formato de la Escala

#### 9.1.2 Proceso de Computación y Análisis de Datos

#### 9.1.3 Selección de las Afirmaciones

#### 9.1.4 Puntajes de Actitud e Interpretación

#### 9.1.5 Transformación de los Puntajes de Actitud a Puntajes T

#### 9.1.6 Ventajas y Desventajas del Método de Rangos Sumarizados

### 9.2 Análisis de Escalograma

9.2.1. Proposiciones Fundamentales que Subyacen al Análisis Escalar

9.2.2 Técnica de Cornell

9.3 Técnica de Goodenough

9.3.1 Afirmaciones con Categorías de Respuestas No Dicotómicas

9.3.1.1 Coeficiente Mínimo de Reproducibilidad Marginal

9.3.1.2 Patrones de respuesta: Escalares y No Escalares

9.3.1.3 Ventajas y Desventajas

REFERENCIAS

## X. ESCALAS DE ACTITUD : DIFERENCIAL SEMÁNTICO

10.1 El Diferencial semántico como Escala de Actitud

10.2 Ventajas y Desventajas

REFERENCIA

## XI. CARÁCTERÍSTICAS PSICOMÉTRICAS DE LOS INSTRUMENTOS: CONFIABILIDAD.

11.1. Confiabilidad

11.1.1. Estabilidad temporal

11.1.2. Muestreo de reactivos

11.1.3 Homogeneidad de los reactivos

11.1.4 Confiabilidad del examinador y el calificador

11.1.5. Fuentes Principales de Error

11.1.5.1 Variación dentro de una Prueba

11.1.5.2. Variaciones entre pruebas

11.2 Métodos Experimentales para obtener la confiabilidad de una prueba.

- 11.2.1. Uso de Formas Paralelas
  - 11.2.2 "Retest" (Replicación) con la misma forma
  - 11.2.3 Consideraciones Generales de los Métodos por Mitades
    - 11.2.3.1 Mitades o tercios sucesivos
    - 11.2.3.2 División por reactivos nones contra pares
    - 11.2.3.3 Subpruebas apareadas al azar
  - 11.2.4 Correlación Intraclase
  - 11.2.5 Consistencia Interna
    - 11.2.5.1 Alpha de Cronbach
    - 11.2.5.3 Análisis de Reactivos
- REFERENCIAS

## XII. CARACTERISTICAS PSICOMETRICAS DE LOS INSTRUMENTOS: VALIDEZ

- 12.1. Validez Predictiva
  - 12.1.2. Coeficiente de Validez
  - 12.1.3 Comparación de grupos
  - 12.1.4 Comparación de la Distribución de la variable Predictiva y un arreglo de la Distribución de la Variable Criterio
- 12.2 Comparación con Grupos Conocidos
- 12.3 Validez de Contenido
- 12.4 Validez de Construcción
  - 12.4.1 Dominio de Observables
  - 12.4.2 Relaciones entre observables
  - 12.4.3 Relaciones entre constructos
  - 12.4.4 Explicación de los constructos
  - 12.4.5 Análisis Discriminante
- 12.5 Validez de fase
- 12.6 Lugar que ocupa el Análisis Factorial

12.7 Relación entre Confiabilidad y Validez

REFERENCIAS

### XIII. EL REPORTE DE INVESTIGACIÓN

13.1 Raporte de Investigación para Publicación

13.1.1. Elaboración del Reporte de Investigación: Reglas Principales

13.2 Formato del Reporte de Investigación.

13.2.1 Resumen

13.2.2 Introducción

13.2.3 Método

13.2.4 Resultados

13.2.5 Discusión

13.2.6 Referencias Bibliográficas

13.3 Reporte de Investigación Aplicada Institucional

13.3.1 Introducción

13.3.2 Método

13.3.3 Resultados

13.3.4 Conclusiones

13.4 Reporte de Investigación de Tesis

13.4.1 Introducción

13.4.2 Antecedentes

13.4.3 Marco Teórico de Referencia

13.4.4. Metodología

13.4.5 Interpretación y Discusión de resultados

13.4.6 Referencias Bibliográficas

## **METODOLOGÍA DE INVESTIGACIÓN EN CIENCIAS SOCIALES**

Gilda Gómez-Peresmitré y Lucy Reidl

U.N.A.M.

### **I. CIENCIA, MÉTODO Y TEORÍA**

#### **1.1. Conceptos Generales.**

Una revisión de los términos que forman el título se hace necesaria, si se considera, por un lado, la multiplicidad de significados que a ellos se les refiere y por otro, si se toma en cuenta que constituye, en su conjunto, el antecedente fundamental del proceso de investigación; esto es, el objetivo que guía y da cuerpo al tema central del presente trabajo.

De esta manera, se tratará en primer término, de proporcionar los elementos que permiten distinguir el concepto de ciencia dentro del sistema total de la misma, haciendo referencia especial a las llamadas ciencias sociales, para integrarlos posteriormente junto con los elementos de los conceptos de método y teoría, al proceso de investigación.

**1.1.1. Ciencia.** Aún cuando el término de ciencia es uno de los más frecuentemente utilizados, también es uno de los más incorrectamente empleados, en torno al cual han surgido, no sólo definiciones variadas sino también contradictorias (Rosenbleuth, 1971; Rosenthal y Rosnow, 1991). Entre algunas de las razones y/o características del trabajo científico que se consideran han contribuido a lo anterior se encuentran las siguientes:

- la mayoría de las personas que se dedican a la investigación científica, no podrían formular con precisión el concepto de lo que es la ciencia, ya que lo consideran como algo dado;
- la utilización de criterios populares como son, que la ciencia busca el beneficio del hombre y de la humanidad, o bien que la ciencia trata de lograr el dominio, el poder y la fuerza, pasando por encima de las reglas éticas tradicionales;
- la retórica (lenguaje) de la ciencia no constituye una representación exacta de la intuición y presupuestos, esto es, de las ideas lógicas como también de las ilógicas que los científicos utilizan al crear o planear su investigación;
- a pesar del interés de los científicos por la lógica su perspectiva también está caracterizada por supuestos ocultos que

pueden no ser estrictamente lógicos y si presuposiciones intuitivas en torno a la naturaleza de la realidad que filósofos como Apel (1982) y Rosnow (1983) entre otros llaman **principios regulativos**;

- existen limitaciones en la empresa misma de la ciencia. Una de sus principales limitaciones tiene que ver con nuestra capacidad cognoscitiva para visualizar y expresar completamente nuestras experiencias. Asimismo, de las limitaciones lógicas y temporales de los métodos empíricos se derivan otras limitaciones de la ciencia ya que hay aspectos de la realidad que van más allá de los procedimientos empíricos particulares.

De acuerdo con Kerlinger (1982), el modo más fácil de caracterizar a la ciencia moderna, es contraponiendo el conocimiento científico al vulgar y al conocimiento especulativo. Así, de acuerdo con el autor, el conocimiento vulgar es la serie de conceptos y esquemas conceptuales satisfactorios para los usos prácticos de la humanidad. El conocimiento vulgar atribuye importancia a la autoridad y carece de un espíritu crítico, de duda. En cambio, el conocimiento científico trata de entender de una manera sistemática los fenómenos estudiados, y se preocupa, en gran medida, por examinar y probar teorías e hipótesis. Por otra parte, la diferencia entre las ciencias empíricas modernas y las construcciones teológico-especulativas, radica en la manera cómo se pretenden comprobar las teorías e hipótesis. La teología se apoya en la autoridad bíblica o eclesiástica; la ciencia empírica insiste en el recurso sistemático y controlado de la experiencia.

En esta misma línea de pensamiento, Rosenbleuth (1971) afirma que es necesario tener en cuenta que el sentido común hace generalizaciones que no comprueba, mientras que la ciencia exige que sus generalizaciones se basen en pruebas rigurosas para efectos de verificación.

Así, existen, de acuerdo con este autor, algunos criterios populares que frecuentemente se utilizan para calificar un procedimiento o un estudio como científico, por ejemplo:

-Descripción. Aunque la estrategia de la observación forma parte del método científico, la mera descripción de lo observado no es criterio suficiente para el calificativo de científico.

-Sistematización. La sistematización, al igual que la descripción, sólo constituye una etapa preliminar en el desarrollo del conocimiento. Cualquier clase de hechos pueden ser sistematizados, y no obstante, ser completamente ajenos a los propósitos de la ciencia.

-Medición. Aunque frecuentemente se considera la medición como una tarea de la ciencia, por sí sola no constituye una aportación científica. Es decir, el que los datos sean tratados matemáticamente, no asegura resultados científicos.

-Explicación de los hechos. Explicar algo a alguien es procurarle una satisfacción subjetiva, lo cual sólo es incidental en los propósitos de la ciencia; por lo que este criterio tampoco resulta útil si se considera de manera aislada.

-Predicción. Aunque la ciencia permite hacer predicciones también el sentido común las hace, de aquí que por sí sola no sea acreedora del estatuto de la ciencia.

-Conocimiento del universo. Constituye uno de los criterios más aceptados aunque se le señala como una noción subjetiva, además de las diversas acepciones que se le dan. De aquí que de acuerdo con el autor sea necesario agregar al término conocimiento el calificativo de científico ya que como tal tiene particularidades que lo singularizan y lo separan de otros modos de conocimiento.

Podría señalarse de paso, como una de las particularidades del conocimiento científico (desde luego, postulada por una posición frente a la ciencia diferente a los autores antes referidos), el hecho de que éste, no obstante de partir del concreto real, del sujeto de la realidad, logra a través de un proceso de elaboración que lleva al análisis de las abstracciones más simples, transformar ese punto de partida en un concepto que ha roto con el dato sensorial, con la apariencia.

Una ciencia es una disciplina que utiliza el método científico con la finalidad de hallar estructuras generales (leyes). Por tanto, la ciencia es un estilo de pensamiento y acción. Aspira a ser racional, objetiva, crítica y coherente; adaptable a hechos, en vez de permitirse especulaciones sin control. Sin embargo, de acuerdo con Piaget (1970), no es posible limitarse a una jerarquía o una clasificación lineal. Ninguna de las ciencias puede ser situada en un sólo plano, sino que cada una de ellas admite distintos niveles jerárquicos según se considere:

- el objeto o contenido material de su estudio;
- sus interpretaciones conceptuales o técnicas-teóricas;
- su epistemología interna o análisis de sus fundamentos;
- su epistemología derivada o análisis de las relaciones entre el sujeto y el objeto en conexión con las demás ciencias.

Se presenta entonces, el problema de la especificidad o reductibilidad de los fenómenos estudiados en las distintas ramas



del saber.

Conceptos tales como los de interacción e interdependencia han ido sustituyendo a las series lineales. De esta manera, la pregunta que surge es si hay una tendencia a la asimilación general o a una serie de modos relacionales o dialécticos de interpretación que tengan en cuenta tanto las oposiciones como las analogías. Así mismo, al encontrar diferencias de escala, el problema que surge es el de determinar: si los mecanismos de escala superior son reductibles a los de escala inferior; si los primeros son simplemente irreductibles, o incluso si hay alguna relación inteligible entre los dos. También está el problema de si las ciencias son abiertas o si por el contrario, hay fronteras definidas y estables que separan los problemas científicos de aquellos de carácter filosófico.

Piaget (1970) señala que existe una tendencia a considerar que las ciencias son abiertas, por lo que resulta vano tratar de trazar fronteras inmutables entre un grupo de nociones consideradas como las únicas científicas y otro grupo que representaría a la filosofía. De acuerdo con el mismo autor, los factores que han contribuido al desarrollo de las ciencias sociales, son los siguientes:

- Una ciencia no se reduce a ir acumulando conocimientos de un modo aditivo a partir de una centralización inicial, sino que supone que dicha adición va acompañada de sistematizaciones. La primera condición para una sistematización objetiva es un distanciamiento con respecto al punto de vista propio. Este distanciamiento asegura la actitud de comparación ampliando las exigencias normativas hasta subordinarlas a múltiples sistemas de referencia.
- La tendencia histórica o genética. Esta tendencia señala que los estados sociales o individuales que aparentemente dan lugar a un conocimiento intuitivo o inmediato, son producto de una historia o un desarrollo cuyo conocimiento es necesario para comprender los resultados. Se trata de un distanciamiento, que además de la posibilidad de comparación, proporciona un medio de explicación en tanto que los desarrollos en cuestión, están relacionados causalmente.
- Los modelos ofrecidos por las ciencias de la naturaleza. Si los primeros pasos de este acercamiento se caracterizan sobre todo por la tendencia reduccionista, la continuación del desarrollo histórico de las investigaciones ha puesto de manifiesto que la utilización de modelos tomados de las ciencias naturales no excluía tener en cuenta la especificidad de los fenómenos de orden superior, y que varias de las técnicas elaboradas en el terreno de

las ciencias del hombre han influido a su vez, en las disciplinas biológicas, e incluso en las fisicoquímicas.

-La tendencia a la delimitación de problemas. La ciencia comienza tan pronto como es posible delimitar un problema de tal forma que su solución queda subordinada a una serie de comprobaciones accesibles a cualquiera, separándolo de las cuestiones de evaluación o convicción.

-La elección de los métodos. Lo que ha dado unidad a las ciencias sociales ha sido la voluntad común de verificación cuya precisión aumenta precisamente en relación con los controles recíprocos e incluso con las críticas. La fase científica comienza cuando el investigador, separando lo verificable de lo que es reflexivo o intuitivo elabora métodos especiales, adaptados a su problemática, que son a la vez métodos de análisis y de verificación.

Ahora bien, en relación con los procedimientos específicos que desarrolla la ciencia, ésta puede ser, de acuerdo con Bunge (1969), formal o factual. La primera estudia ideas, es autosuficiente en cuanto a su contenido y método, se refiere a fórmulas analíticas, es la lógica matemática. Al respecto Torgerson (1958), afirma que la ciencia está compuesta de un espacio teórico (que corresponde a la definición que hace Bunge de ciencia formal), en el cual, se encuentran un conjunto de constructos conectados entre sí (relaciones lógicas y formales) y el espacio de la naturaleza, de los datos observables, con los que se conectan, a través de otro tipo de relaciones (definiciones operacionales, reglas de correspondencia o interpretación), los constructos del espacio teórico. El interjuego de estas dos esferas constituye el avance de la teoría científica. El espacio teórico, por sí solo, sin relación con los datos de la naturaleza, constituye de acuerdo con Torgerson, un modelo (que tiene las características de autosuficiencia señaladas por Bunge), éste se convierte en una teoría científica en el momento en que se contrasta (se relaciona) con los datos observables de la naturaleza.

La segunda, la ciencia factual, estudia hechos, depende de éstos en cuanto al contenido y método, se refiere a fórmulas sintéticas y puede ser natural o cultural. Por otra parte, en cuanto al objetivo y alcance de la ciencia, puede clasificarse en pura y aplicada: la ciencia pura tiene un fin meramente cognoscitivo y acrecienta la teoría; mientras que la ciencia aplicada, tiene fines prácticos, y explica hechos concretos.

**1.1.2. Método.** El método en su expresión genérica es un procedimiento que se sigue para tratar un problema o un

conjunto de ellos. Sin embargo, en el ámbito del conocimiento, se le dan diferentes connotaciones. El método científico permite la más adecuada expresión de los procesos del pensamiento a través de enunciados y argumentos que explican los fenómenos del universo, las relaciones internas entre sus elementos y sus conexiones con otros fenómenos mediante el raciocinio y la comprobación a través de la demostración y verificación. Debe distinguirse, sin embargo método de técnica. Esta se refiere a actividades u operaciones prácticas aplicables con diferentes propósitos determinados por el método y la teoría. El método, en cambio, puede verse como compuesto por varias fases o entenderse distinguiendo sus diferentes niveles.

Por ejemplo, y de acuerdo con Grawitz (1974), la causa de la multiplicidad de sentidos que se dan al método, reside en que se usa para designar diversos procedimientos situados a muy distintos niveles, a diferentes grados de abstracción, o al modo como influyen en las etapas más o menos concretas, del proceso de investigación empírica.

Grawitz distingue cuatro sentidos fundamentales:

a) Método en un sentido filosófico: es el nivel más alto de abstracción, designa los procedimientos lógicos, independientes de todo contenido concreto.

b) Método como actitud concreta frente al objeto (teoría): dicta los modos concretos de organizar una investigación.

c) Método ligado a una tentativa de explicación (metodología): se refiere más o menos a determinada posición filosófica y como tal influye en las etapas de la investigación.

d) Método ligado a un ámbito en particular (técnicas): se refiere a un dominio específico e implica una manera propia de actuar. La metodología ocupa un lugar menos abstracto y dependiente que la posición filosófica y la teoría. Al respecto Jamous (1968, citado por Grawitz, 1974) señala que uno se encuentra ante un problema de método, cuando el investigador, se interroga sobre la elección, la coordinación y la integración de objetivos específicos ligados a técnicas particulares para responder a dicho problema.

Por otra parte, se postula que la unidad en la ciencia consiste en la existencia de un método científico válido para todas las disciplinas, independientemente de cuál sea su materia u objeto de estudio. De esta manera, se habla de método general, para buscar la verdad con una sola estrategia y de métodos especiales que

dependen de la naturaleza del objeto de estudio (técnicas).

Así por ejemplo, Bunge (1969), considera que el método científico es una estrategia que afecta el ciclo completo de investigación, es independiente del tema de estudio, aunque como lo señala el autor, la ejecución concreta de cada una de sus operaciones dependerá del grado de conocimiento que se haya desarrollado en relación a éste.

**1.1.3. Teoría.** La formulación de la teoría constituye uno de los principales objetivos de la ciencia. En general, puede decirse que una teoría científica está constituida por un conjunto de leyes ordenadas sistemáticamente, que permiten explicar el comportamiento de los procesos estudiados por una rama dada de la ciencia. Esta definición de teoría, habla de un sistema deductivo axiomatizado.

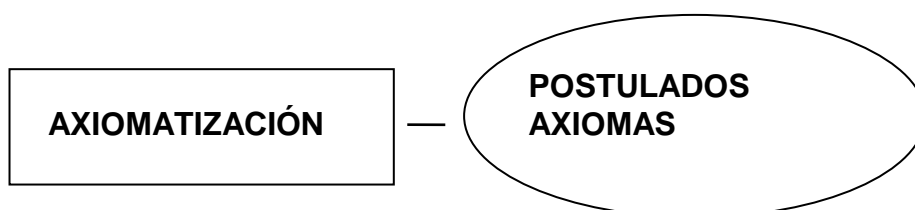
En las ciencias sociales y conductuales, las teorías científicas pueden definirse como un sistema de símbolos descriptivos y/o explicativos.

De acuerdo con Madsen (1967), una teoría axiomatizada es aquella en la que sus enunciados se han ordenado o sistematizado en primarios (axiomas, premisas o principios), y secundarios (consecuencias, conclusiones o teoremas). Los enunciados secundarios se derivan por deducción de los primarios o de otros secundarios siguiendo ciertas reglas lógicas.

Los postulados son conjuntos de expresiones que pueden representar experiencias y que se dan por sentado. Pueden estar formados por palabras definidas o indefinidas que se combinan en oraciones según reglas lingüísticas comunes o siguiendo normas sintácticas especiales. Frecuentemente las palabras indefinidas son términos abstractos o hipotéticos que se definen por otros menos abstractos; y por último, mediante términos empíricos. La sintaxis y la semántica son los componentes de la teoría.

Forman parte de la sintaxis, la axiomatización, simbolización y formalización (véase la Figura 1).

S  
I  
N  
T  
A  
X



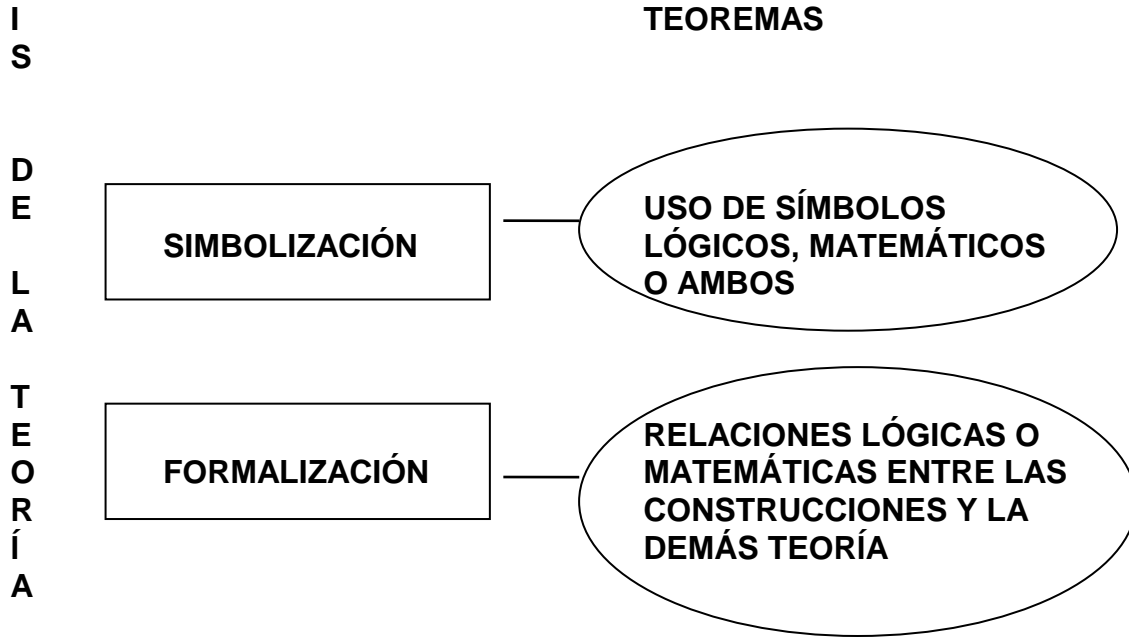


Figura 1. Sintaxis de la Teoría

Cabe señalar, que dejó de aceptarse que los axiomas fueran proposiciones sin demostración, esto es, evidentes **a priori**. Sin embargo, fue dentro de la misma matemática, el reino de los axiomas, en donde las verdades **a priori** desaparecen como tales. Así por ejemplo, uno de los postulados de Euclides, el más difícil de demostrar (las paralelas que se extienden indefinidamente sin juntarse), fue sometido a numerosas pruebas basadas en el método de reducción a lo absurdo (demostrar que la afirmación de la proposición contraria a la que se quiere demostrar, conduce a una contradicción con los otros axiomas o postulados, o con otra posición ya demostrada a partir de ellos). En ninguno de los intentos se encontró la inconsistencia que se esperaba.

La ciencia moderna no acepta axiomas ni verdades absolutas apriorísticas; acepta, sin embargo, un número importante de postulados definidos como principios sugeridos por la experiencia y aceptados sin prueba previa para ser desechados si la experiencia ulterior estuviera en desacuerdo con ellos.

Las reglas sintácticas indican la manera en que los términos de la teoría forman una red explicativa, describen el funcionamiento de los términos teóricos, para relacionarlos entre sí.

Dentro de la psicología se encuentran, de acuerdo con Madsen (1967), teorías con una estructura formal adecuada, por ejemplo la de Lewin, que desarrolla constructos como valencia, fuerza y tensión; que se relacionan unos con otros de acuerdo con reglas

formales de relación. Es coherente o consistente, al margen de que la realidad que trata de explicar se comporte como lo especifica la teoría. Dentro de estas mismas consideraciones se encuentra la teoría hipotético-deductiva del aprendizaje por habituación de Hull, pero nuevamente sus posibilidades empíricas, son mínimas.

Por otro lado, a la vinculación de la estructura formal (sintaxis), con los datos empíricos (semántica), se le suele llamar anclaje. Así, puede decirse que la semántica constituye sistemas de anclaje de la teoría con la realidad o hechos empíricos.

Las construcciones teóricas deben relacionarse semánticamente con las construcciones empíricas, que constituye lo que se conoce como lenguaje de los datos, compuestos por la tecnología necesaria para describir observaciones y operaciones. Así por ejemplo, en psicología, el lenguaje de los datos permite describir conductas junto con las situaciones en donde éstas ocurren.

El problema de la semántica de una teoría, no sólo está presente en la relación de las construcciones teóricas y las empíricas, sino también de éstas (lenguaje de los datos) a los hechos de la experiencia, el lenguaje de las cosas físicas. Idealmente el lenguaje de los datos y el de las cosas físicas deberían estar relacionados a través de definiciones operacionales o por reducciones, en donde un término se define por sus propiedades funcionales.

La Figura 1.1 expresa las relaciones sintácticas y semánticas de una teoría.

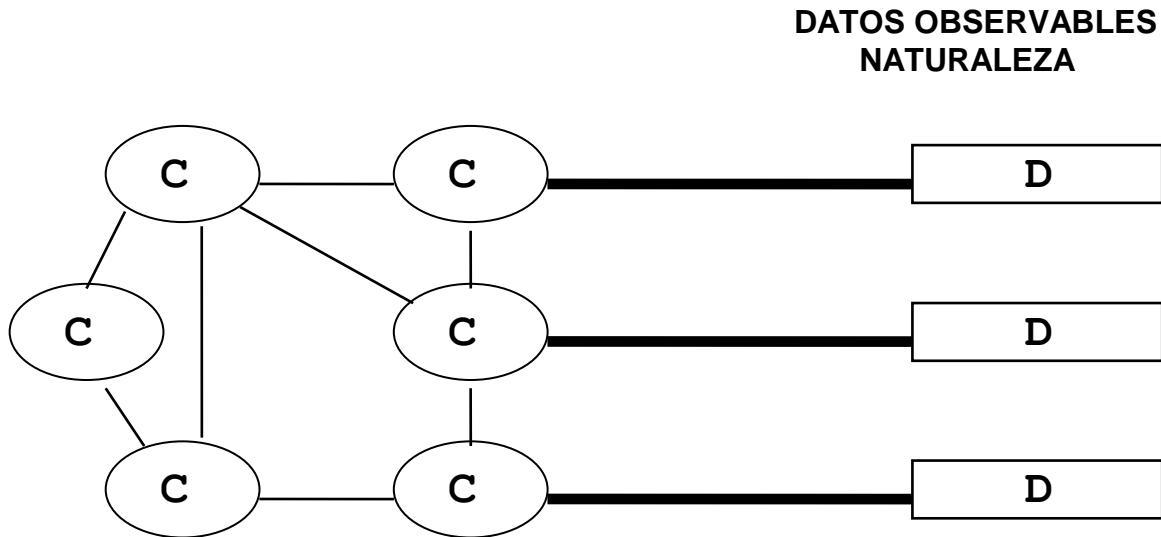
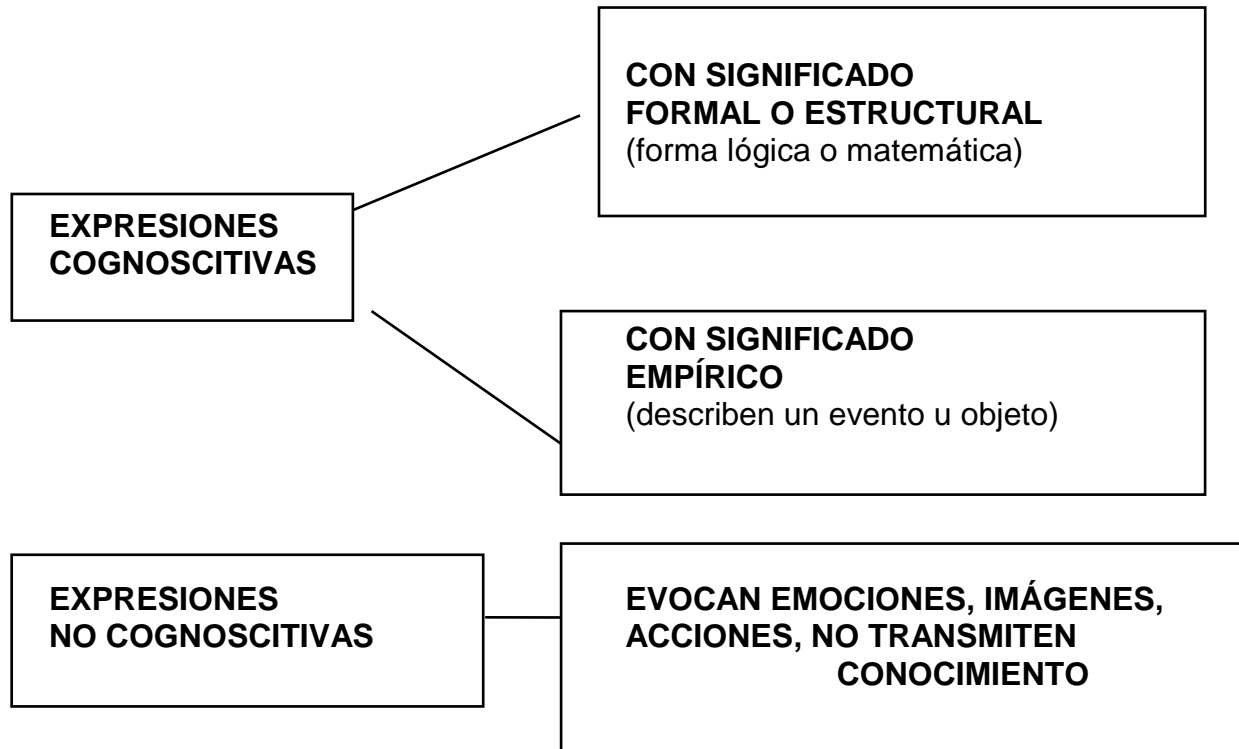


Figura 1.1. Representación esquemática (tomada de Torgerson, 1958, p. 3) que ilustra la estructura de una ciencia bien desarrollada. Los círculos corresponden a los Constructos; las líneas sencillas son las conexiones formales entre constructos; las líneas dobles son las reglas de correspondencia (definiciones operacionales) que unen a ciertos constructos con los datos de la naturaleza.

Aunque el problema de la semántica es antiguo, fue traído nuevamente al tapete de la discusión debido por un lado, al desarrollo de las matemáticas formales y de la lógica simbólica o matemática, que culminó en la apreciación de que éstas no sólo se relacionaban íntimamente, sino que formaban una sola disciplina; y por el otro, a la revolución en la manera de pensar de los físicos ocasionada por la aparición de la teoría de la relatividad de Einstein y posteriormente por la aparición de la teoría cuántica. Así, los hombres de ciencia tuvieron que revisar sus principios básicos y por supuesto, las relaciones entre los hechos y experimentos y las teorías elaboradas a partir de ellos.

Dicha revisión llevó a la conclusión de que en el lenguaje existen distintos tipos de proposiciones que se agrupan en distintos universos de discurso:



Así, debido a que la ciencia estudia hechos concretos, requiere de un gran número de afirmaciones sintéticas, pero como también ésta busca relaciones formales entre distintas variables, se hace necesario el uso de afirmaciones de carácter analítico.

De ésta manera, los positivistas lógicos proporcionaron un criterio para atribuir significado a una proposición cognoscitiva empírica. Pierce (1955), la formuló pragmáticamente como sigue: Una proposición tiene significado empírico cuando y sólo cuando la afirmación o negación de dicha proposición implica una diferencia que puede ser sometida a una prueba experimental. Por su parte, Carnap (1937) y Fiegl (1951), distinguen dentro del lenguaje cognoscitivo empírico, el lenguaje de los datos y el de los conceptos físicos. En términos generales, el lenguaje de los datos es el que expresa las sensaciones y percepciones directas e inmediatas de los observadores. Aunque este lenguaje es introspectivo (lo que el observador percibió), puede adquirir universalidad, si se da un acuerdo satisfactorio entre diversos observadores.

Desde otra perspectiva, Lazarsfeld (1970) propone hablar de reflexión analítica en lugar de teoría. Según este autor, cuando se habla de teoría en ciencias sociales se hace referencia a alguna de las siguientes operaciones:



- establecimiento de esquemas de clasificación precisos
- formulación de conceptos complejos que orientan al observador hacia hechos interesantes
- formulación de problemas de investigación de una gran importancia desde el punto de vista de la sociedad
- formulación de ideas generales acerca de la manera en que se producen o pueden ser provocados los cambios sociales
- hipótesis
- interpretación: puesta en relación de unos hechos empíricos con otros hipotéticos.

Las teorías existentes en las disciplinas sociales, también se clasifican como teorías de rango medio, término usado por Merton en 1957 para caracterizar a las teorías que emplean conjuntos de construcciones laxamente relacionados y un limitado conjunto de hipótesis que derivan más o menos rigurosamente de las suposiciones. Estas teorías son más informales que formales en su estructura interna y generalmente pertenecen a un aspecto más o menos limitado de la conducta social.

Los modelos formales que existen dentro de las Ciencias Sociales están restringidos a aspectos particulares de la conducta social tales como: teoría de la decisión, del juego y de la comunicación, las cuales han tenido un éxito limitado cuando se ha tratado de validarlas con observaciones del mundo real.

También es necesario hacer notar que la complejidad en la investigación social está dada en parte por la naturaleza de sus variables; dos características importantes de la conducta social residen en su naturaleza causal y en su historicidad. Es decir, la conducta social es multicausal, más de una variable la produce y la historia de las relaciones pasadas influye en las respuestas actuales, por lo que no es fácil categorizar a las variables sociales en dependientes e independientes, en función de simples modelos estímulo-respuesta. Por otro lado, está el hecho de que están operando procesos, lo que determina que se debe tener presente las relaciones temporales. Es decir, en un momento dado algo puede ser una consecuencia de procesos previos o una causa de efectos posteriores, de tal modo que la estrategia del investigador social está necesariamente determinada por la yuxtaposición de variables en su teoría o en las ideas que lo orientan. Se debe tener en cuenta además, que las relaciones entre las variables en el orden social no son estables en tiempo y espacio.

La complejidad de la investigación social es aún más evidente cuando vemos que el investigador mismo se convierte en una

variable que puede afectar los resultados, por lo que requiere de un control especial.

Ahora bien, las técnicas y los instrumentos que utiliza el investigador están proyectados y justificados con la ayuda de las teorías. Es decir, la información empírica no se estima nunca en un vacío teórico, sino que tal evidencia tiene que juzgarse a la luz de la teoría utilizada para proyectar y llevar a la práctica la técnica con la cual se ha obtenido la información. Así como ninguna teoría factual se sostiene por sí misma, así tampoco hay datos que constituyan por sí mismos evidencia a favor o en contra de una teoría.

Una de las hipótesis sobre la que la ciencia factual trabaja, es la cognoscibilidad limitada. Esto obliga exclusivamente a esperar que los efectos del azar (sus probabilidades), puedan finalmente calcularse y que puedan analizarse y reducirse en alguna medida las lasitudes objetivas o subjetivas. En otras palabras, la ciencia factual reconoce que existe objetivamente el azar en todos los niveles y que no es posible ninguna certeza definitiva. El problema epistemológico genuino, no es si podemos conocer, sino en qué medida conocemos la realidad que nos rodea, y si podemos ampliar las actuales fronteras de lo conocido.

Cuando una teoría no se limita a describir ciertos esquemas sino que intenta describir mecanismos, se dice que es una teoría profunda (representacional). Estas son más específicas y por lo tanto más informales, igualmente que son las mejor contrastables por lo que son más aptas para adquirir o perder una buena fundamentación empírica. La posesión de tres atributos que caracterizan mejor la profundidad de las teorías son: a) construcciones de nivel alto; b) presencia de un mecanismo; y c) gran capacidad de explicación.

Las teorías menos profundas están más próximas a los fenómenos por lo que se les conoce como teorías fenomenológicas. Sin embargo, esto no quiere decir que estas teorías sean totalmente descriptivas y que no tengan capacidad explicativa. Debido a que este nombre puede introducir malentendidos, se ha preferido llamarlas teorías del tipo **caja negra**. Estas teorías atienden al comportamiento del sistema y no toman en cuenta la estructura interna, tratándola como unidad simple. Este tipo de teorías da razón del comportamiento general basándose en relaciones entre variables globales como causas netas (**inputs**) y efectos netos (**outputs**), y se asume que entre causa y efecto existe mediación de variables intervinientes que carecen de referentes.

Metodológicamente las teorías del tipo caja negra son muy

estimadas como contrastadoras de las correspondientes teorías representacionales; además suelen ser manejables, mientras que las representacionales, al ser más complejas son más difíciles de manejar.

Las aplicaciones cognoscitivas de las teorías, es decir, las explicaciones y las predicciones preceden a su aplicación práctica. En otras palabras, antes de poder hacer algo relacionamente con un objetivo práctico, se tiene que entender qué es, y por qué es, y cómo puede comportarse. Esto es lo que se llama la acción racional; esta acción no es sino la consumación de una elección hecha a la luz de las teorías o a la ejecución de un plan construido con la ayuda de la teoría. En otros términos, puede decirse que toda explicación racional de un hecho es una secuencia hipotético-deductiva que contiene generalizaciones e información. Si éstas son lógicamente válidas, se hace referencia a una explicación científica y nomológica (Piaget, 1970).

El objetivo principal de la explicación científica es ampliar, profundizar y racionalizar el conocimiento ordinario. Todo lo que puede explicar la ciencia en un momento dado, se explica con la ayuda de las teorías. No toda explicación es causal: hay tantas clases de explicación como clases de teoría haya. Por último, nunca se consigue una explicación perfecta, completa y precisa. Sin embargo, toda explicación es perfectible, puede mejorarse y corregirse. Toda nueva explicación aumenta la sistematicidad del conocimiento porque transforma el dato aislado o la generalización suelta en una pieza de un todo conceptual.

La esencia del procedimiento de investigación reside en la relación fundamental entre teorías, métodos y técnicas utilizados para obtener pruebas adecuadas que permitan su verificación.

Las teorías sobre la conducta social, por lo general, acentúan las características de un individuo, como consecuencia de sus experiencias pasadas o, por otro lado, acentúan las características de sus situación social inmediata (tradición psicológica vs. tradición sociológica). Las primeras estudian las respuestas típicas de una persona en diversas situaciones y relaciones sociales; las segundas estudian a los individuos como seres influidos por presiones que se originan en las demandas de la actual estructura social, incluidas las expectativas de los otros. Estos dos enfoques tienden a estimular formas diferenciadas de investigación. Sin embargo, existen pruebas evidentes de la confluencia de ambos enfoques (Lazarsfeld, 1970).

Otras distinciones teóricas dentro de las Ciencias Sociales son las que se establecen entre las teorías cognoscitivas y las

teorías conductuales. Las teorías conductuales se ocupan de las formas de conducta que varían de acuerdo a los estímulos ambientales y pueden demostrarse en forma sistemática. Por otro lado, se tienen las teorías cognoscitivas, que se ocupan de las experiencias internas del individuo, sin querer decir que se descuide la conducta, sino más bien, se concibe a los factores cognoscitivos como mediadores, es decir, como variables que intervienen entre estímulo y respuesta.

Existen también teorías que ponen énfasis en la consciencia que el sujeto tiene de sus experiencias y sobre cuya base actúa. En el otro extremo están aquellas en las que se acentúa lo inconsciente, el no darse cuenta de las experiencias, y la conducta se explica por la determinación de los problemas no resueltos en los primeros años de vida.

Aunque estos señalamientos no agotan las clasificaciones hechas al respecto de las teorías, el propósito es hacer notar que cada una de las teorías dirige la investigación en cuanto al método a usar; determina cuáles son las variables relevantes a seleccionar y controlar y contribuye además a decidir si éstas, las variables, se someterán o no, a medición. En suma, el papel de la teoría es esencial para determinar la metodología; esto es, método y técnicas de investigación así como la interpretación que se hará de los resultados.

De esta manera puede decirse que existe una íntima interrelación entre teoría y método; que ciertas técnicas son más compatibles con algunas suposiciones teóricas que con otras; que el investigador, por el simple hecho de seleccionar una determinada metodología (métodos y técnicas) necesariamente asume una particular teoría; y por último, que es a través de la actividad científica (el proceso de investigación que subsume teoría y método), como se produce un conocimiento "**objetivo**" del mundo.

#### **REFERENCIAS**

Apel, K. (1982). C.S. Pierce and the post-Tarskian problem of an adequate explication of the meaning of truth: Towards a transcendental pragmatic theory of truth, part I. **Transactions of the Charles S. Pierce Society, 18, 3-17.**

Bunge, M. (1969): **La investigación Científica.** Barcelona: Ediciones Ariel.

Carnap, R. (1937): **Logical Syntax of Language.** Londres: Routledge and Kegan Paul.

- Feigl, H.(1951): Principles and Problems of Theory Construction in Psychology. En W. Dennis (De.): **Current Trends of Psychological Theory**. Pittsburg: University of Pittsburg Press.
- Grawitz, M.(1974): **Methodes de Sciences Sociales**. Paris: Dalloz.
- Kerlinger, F.N.(1982): **Fundamentos de la Investigación del Comportamiento**. México: Nueva Editorial Interamericana.
- Lazarsfeld, P.S.(1970): La Sociología. En J. Piaget, W.J.M. Makenzie, P.S. Lazarsfeld y otros (Eds.): **Tendencias en la Investigación en Ciencias Sociales**. Capítulo 3. Madrid: Alianza.
- Madsen, K.B. (1967): **Teorías de la Motivación**. Buenos Aires: Paidós.
- Merton, R.K. (1957): **Social Theory and Social Structure**. Nueva York: The Free Press.
- Piaget, J. (1970): La Situación de las Ciencias del Hombre dentro del Sistema de las Ciencias. En J. Piaget, W.J.M. Mackenzie, P.L. Lazarsfeld y otros (Eds.): **Tendencias de la Investigación en las Ciencias Sociales**. Madrid: Alianza Editorial. pp. 44-120.
- Pierce, C.S. (1955): General Theory of Probable Inference. En J. Buchler (Ed.): **Philosophical Writings of Pierce**. Nueva York: Dover.
- Rosenbleuth, A. (1971): **El Método Científico**. México: Consejo Nacional de Ciencia y Tecnología.
- Rosenthal,R., y Rosnow, R.L.(1991). **Essentials of behavioral research.Methods and data analysis**.Nueva York: McGraw-Hill
- Rosnow. R.L.(1983). Von Osteris horse, Hamlets question, and the mechanistic view of causality: Implications for a post-crisis social psychology. **Journal of Mind and Behavior, 4, 319-338**.
- Torgerson, W.S. (1958): **Theory and Methods of Scaling**. Nueva York: John Wiley.

## **II. MÉTODOS DE INVESTIGACIÓN**

### **2.1. Introducción.**

La investigación dentro del campo de las disciplinas sociales, se propone los siguientes objetivos:

- a) Conocer y/o producir conocimiento en relación a fenómenos, eventos o situaciones de carácter social. Con esto se quiere decir: explicar, o por lo menos describir las razones, causas y relaciones que intervienen en la producción de ese fenómeno, evento o situación, con el objetivo de integrarlo al campo del conocimiento.
- b) Responder a preguntas que plantea la realidad; esto es, tratar de descubrir el por qué, cómo y cuándo de las cosas que nos rodean para aplicar dicho conocimiento en la solución de los problemas que enfrenta el hombre.

### **2.2. Clasificación de los Métodos de Investigación.**

El proceso de investigación puede clasificarse desde diversos puntos de vista. Uno de ellos se divide en dos clases, a saber:

- a) investigación pura o básica, b) investigación aplicada.
- La primera se interesa en el descubrimiento de las leyes que rigen el comportamiento de ciertos fenómenos o eventos; intenta encontrar los principios generales que gobiernan los diversos fenómenos en los que el investigador se encuentra interesado. La segunda trata de responder a preguntas o problemas concretos que se presentan al investigador con el objeto de encontrar soluciones o respuestas que puedan aplicarse de manera inmediata en contextos o situaciones específicas.

Esta última, la investigación aplicada suele clasificarse como sigue:

- a) exploratoria, b) descriptiva y c) confirmatoria

En la investigación de carácter exploratorio el investigador intenta, en una primera aproximación, detectar variables, relaciones y condiciones en las que se da el fenómeno en el que está interesado. En otros términos, trata de encontrar indicadores que puedan servir para definir con mayor certeza un fenómeno o evento, desconocido o poco estudiado. Esta clase de investigación, que se lleva a cabo en relación con objetos de estudio para los cuales se cuenta con muy poca o nula información, no puede

aportar, desde luego, conclusiones definitivas ni generalizables, pero sí permite definir más concretamente el problema de investigación, derivar hipótesis, conocer las variables relevantes. En suma, proporcionará la información necesaria para aproximarse al fenómeno con mayor conocimiento en un estudio posterior, en la investigación propiamente dicha. Idealmente toda investigación debería incluir una fase exploratoria.

En la investigación descriptiva, por otra parte, se trata de describir las características más importantes de un determinado objeto de estudio con respecto a su aparición y comportamiento, o simplemente el investigador buscará describir las maneras o formas en que éste se parece o diferencia de él mismo en otra situación o contexto dado. Los estudios descriptivos también proporcionan información para el planteamiento de nuevas investigaciones y para desarrollar formas más adecuadas de enfrentarse a ellas. De esta aproximación, al igual que de la del estudio exploratorio, tampoco se pueden obtener conclusiones generales, ni explicaciones, sino más bien descripciones del comportamiento de un fenómeno dado.

En cambio la investigación confirmatoria, como su nombre lo indica, tiene como función principal confirmar o desconfirmar una teoría o aproximación teórica que intenta explicar el por qué del fenómeno que se estudia. En relación a éste, pueden existir diferentes teorías que tratan de explicarlo; en esta situación el propósito del investigador es probar la fuerza o capacidad de explicación de alguna o algunas de ellas. Este tipo de investigación proporciona principios generales de explicación.

Desde el punto de vista del método frente al objeto de estudio se puede hablar de investigaciones experimentales y no experimentales (asociativas o correlacionales). El investigador centra su interés en la rigurosa recolección de datos y en el análisis refinado mediante el uso ya sea del experimento o de estudios asociativos o correlacionales. El primero abarca las siguientes categorías: experimentos de laboratorio, de campo y naturales. Los segundos abarcan estudios de campo y encuestas (Kerlinger, 1982). En el siguiente esquema (Figura 2.1) se proporcionan las categorías correspondientes de investigación en función del método:

Figura 2.1. Clasificación de los métodos en función del control experimental ejercido sobre las variables a estudiar.

---

**EXPERIMENTALES**

**LABORATORIO:** SE EJERCE EL MÁXIMO CONTROL EN ESCENARIOS NO NATURALES, DIFICULTA LA VALIDEZ EXTERNA.

**CAMPO:** SE EFECTÚAN EN ESCENARIOS NATURALES. DISMINUYE LA ARTIFICIALIDAD, FACILITA LA VALIDEZ EXTERNA.

**NATURALES:** SE PRODUCE UN EVENTO QUE SE CONVIERTE EN LA VARIABLE INDEPENDIENTE (V.I.)

---

**EX-POST-FACTO:** (NO PROPORCIONA EXPLICACIONES FUNCIONALES ).

**CUASI-EXPERIMENTALES:**NO PERMITEN EL CONTROL PROPIAMENTE DICHO DE LA VARIANZA EXTERNA.

---

**ESTUDIOS DE CAMPO:** (INTENSIVOS NO INTERESA EL MUESTREO PROBABILISTICO).

**CORRELACIONALES:**(PROPORCIONA PREDICCIONES).

**NO EXPERIMENTALES**

**ENCUESTA:** (POR LO GENERAL ES UN ESTUDIO EXTENSIVO REPRESENTATIVO).

**OBSERVACIONES NATURALES:** (NO INTERVECIÓN DEL OBSERVADOR).

---



### **2.2.1. Método Experimental.**

#### **2.2.1.1. Experimento de Laboratorio.**

En el experimento de laboratorio el investigador crea una situación con las condiciones exactas que desea y en la cual controla algunas variables y manipula otras. Por lo general en un experimento de laboratorio no se intenta replicar situaciones de la vida real, más bien se quiere crear una situación que permita ver claramente cómo operan las variables en condiciones bien definidas. Esto es, se crean situaciones artificiales para tener la posibilidad de probar, elaborar y refinar el conocimiento con el propósito de incrementar la comprensión y el conocimiento del comportamiento de las variables relevantes que intervienen en los fenómenos sociales. Sin embargo, la aplicación de los resultados encontrados en condiciones experimentales a situaciones de la vida real requieren de experimentación y estudios adicionales. No se puede extrapolar mecánicamente del laboratorio a la vida real, pues en el experimento de laboratorio se gana precisión en el conocimiento del comportamiento de las variables y se pierde realismo, que es el contexto social y complejo en el que éstas se dan. Los componentes principales del experimento son: la observación, el control y la medición. Se provoca deliberadamente algún cambio y se controla, se mide y se interpretan sus resultados con la finalidad de obtener conocimiento.

La observación abarca tres momentos principales: el objeto de observación, el observador y un sistema de registro cuantitativo o cualitativo que conjunta la observación con lo observado.

En cuanto al control, el experimentador lo ejerce tanto en los estímulos que rodean al objeto de observación como sobre el objeto mismo. El tipo de medición utilizado permite, por otra parte, hablar de experimentos cuantitativos o cualitativos. Así, si el control de factores y productos no se lleva a cabo cuantitativamente, el experimento es de tipo cualitativo. Esto es, cuando la presencia o ausencia de las variables o de los factores se tienen en cuenta pero no se miden, el experimento es cualitativo. La mayoría de los experimentos exploratorios que se refieren al descubrimiento de hechos nuevos o a la contrastación preliminar de nuevas teorías son experimentos cualitativos o semi-cuantitativos. En el esquema que sigue (Fig. 2.2) aparecen las características distintivas de la observación natural en comparación con el experimento.



FIG. 2.1

<b>OBSERVACIÓN NATURAL</b>	<b>VS</b>	<b>EXPERIMENTO</b>
Ausencia de control		Máximo control
Grado mínimo de artificialidad		Grado máximo de artificialidad
Ocurrencia de las conductas con independencia del observador.		Produce presencia / ausencia de las condiciones críticas antecedentes de un fenómeno. Atenta contra validez externa por lo tanto contra la generalización .
Capacidad mínima de explicación.		Capacidad máxima de explicación causal.

Por otra parte, las técnicas experimentales por lo general proponen hipótesis, postulados, sobre los factores o variables que se asume pueden intervenir en las relaciones que se estudian. Dentro de los estudios experimentales ocupan un lugar especial dos clases de variables: las variables que son creadas o deliberadamente manipuladas por el experimentador se llaman variables independientes; son los estímulos o tratamientos que provocarán respuestas o valores esperados.

Aquellas variables cuyos valores cambian como consecuencia de las variaciones de las primeras, se llaman variables dependientes; cambian en función de las variables independientes. Para establecer las diferencias introducidas mediante cambios de los valores de éstas se requiere de grupos de control. El grupo control no se encuentra sometido al estímulo, variable independiente o tratamiento que se supone es causa de la variación que se produce en el grupo experimental. Estos grupos experimental y control, tienen que ser homogéneos en los factores o variables relevantes. Lograr esa homogeneidad es tanto más difícil cuanto más complejos son los sistemas que se estudian. Sin embargo, en la medida en que se logra dicha homogeneidad, en esa medida los resultados tienen significado. Es decir, se pueden atribuir a los tratamientos del experimento y no a causas aleatorias. Para alcanzar dicha homogeneidad se puede recurrir a dos clases de control: el individual y el colectivo. El primero se refiere al apareamiento simultáneo de individuos para formar ambos grupos. Este apareamiento simultáneo es difícil y costoso cuando el grupo

es de grandes dimensiones, en cuyo caso se puede apelar al control estadístico cuyas principales variantes son: control de distribuciones y el de aleatorización. El primero se utiliza cuando se cuenta con ciertos parámetros (desviación standard, medias de las poblaciones). En tales casos, los grupos se forman por dichas propiedades estadísticas (por ejemplo, muestras de niños de edad promedio semejante, o con igual promedio de escolaridad). La aleatorización y otras técnicas con mecanismos causales, minimizan la probabilidad de obtener relaciones tendenciadas, de la misma manera que facilitan la utilización de pruebas estadísticas más poderosas, y hacen más consistentes las contrastaciones.

También se pueden formar grupos homogéneos con un control combinado (individual en algunos aspectos y estadístico en todos los demás) utilizando el apareamiento simultáneo para las variables observadas, las que se toman en cuenta explícitamente, mientras que la aleatorización se aplica a las variables que explícitamente no son tomadas en cuenta, pero que se considera pueden dar lugar a alguna diferencia.

Por ejemplo, primero, se igualan los Sujetos en una variable extraña importante, *v.gr.*, en inteligencia, midiéndola a todos y ordenándoles del más al menos inteligente. Posteriormente, al azar, se decide cuando uno pertenece al grupo control y cuando al experimental. De esta manera, todas las demás posibles variables extrañas de interés, quedan repartidas aleatoriamente en ambos grupos.

En cuanto al estímulo en relación al grupo experimental, éste puede actuar por presencia (positivo) o por ausencia (negativo). En el primer caso se aplica o se "hace algo"; en el segundo se elimina algún elemento que generalmente está presente en el grupo control. En algunas ocasiones no se pueden aplicar los estímulos a sistemas reales, por razones de limitación técnica o de carácter ético. En estos casos se trabaja con modelos: el ingeniero construye modelos de edificaciones en pequeña escala y los somete a cargas, movimientos, o cualquier otra acción a los estímulos que interesan para luego extrapolar los resultados al sistema original. El farmacólogo usará monos o ratas como modelos. En algunos casos se recurre a modelos conceptuales que replican algunos de los rasgos de interés del sistema real: se hacen cambios y se interpretan las reacciones apoyándose en alguna teoría. Esta proyección de modelos conceptuales se conoce como simulación.

Las técnicas de simulación pueden substituir al experimento real cuando el propósito no es contrastar teorías, sino aplicarlas.

Estas técnicas no producen teoría ni la someten a prueba. La computadora es la aliada más efectiva en esta clase de experimentos.

Las técnicas experimentales deben contrastarse o convalidarse teórica y empírica o experimentalmente. Experimentalmente porque pueden no funcionar, y teóricamente, porque aunque funcionen, dicho funcionamiento puede deberse a alguna causa distinta de la asumida. La contrastación teórica de una técnica experimental consiste en explicar con ayuda de teorías independientemente corroboradas, cómo funciona el procedimiento; aún más, una técnica que resultó exitosa no puede considerársela como tal, mientras no se pueda explicar su éxito y limitación sobre la base de leyes o principios generales. Por otra parte, los procedimientos empíricos tienen que ser por principio perfectibles y analizables, y por lo tanto susceptibles de comprobación, comparación, y modificación; en suma, susceptibles de contrastaciones válidas y confiables. Debe recordarse, por último, que en el experimento se trabaja con grupos iguales o igualados, idealmente en todas las variables, a los que el investigador tratará diferente (dando valores distintos a la variable independiente). En el esquema que sigue (Fig. 2.3) se resumen algunas de las particularidades más importantes del diseño experimental.

---

### FIG. 2.3. PARTICULARIDADES DEL DISEÑO EXPERIMENTAL:

Si se maximiza el control: en la ejecución de 2 grupos estos deben ser tratados inicialmente igual en todo con una sola excepción (la variable independiente.)

Así cualquier diferencia posterior se atribuirá a aquello que se trató diferente.

### **AI INICIO LA "H<sub>0</sub>." ES VERDADERA.**

Las variables extrañas se aleatorizan por lo que se logran conclusiones más fuertes y se eliminan hipótesis e interpretaciones alternas.

La asignación de los sujetos a los diferentes tratamientos al azar permite el equiparamiento de los grupos.

*Si se tienen variables extrañas sistemáticas se afectan los resultados como sigue: \_*

- a) Los efectos se pueden atribuir a la **VARIABLE INDEPENDIENTE**;
  - b) Se pueden cancelar los efectos de la manipulación de la **VARIABLE INDEPENDIENTE**.
- 

#### **2.2.1.2. Experimento de Campo.**

En el experimento de campo se hace una manipulación real de las condiciones por parte del experimentador, para determinar relaciones funcionales. La manipulación de la variable independiente no se deja a la naturaleza; por el contrario, es ideada por el experimentador. Este debe planear el diseño por adelantado. El contexto de un experimento de campo es natural y se estudian fenómenos sociales concretos. Se puede decir que un experimento de campo es un proyecto de investigación con orientación teórica en el que el experimentador manipula una o más variables independientes en alguna situación social real, con la finalidad de probar hipótesis. En este caso, se trata de aplicar, hasta donde es posible, el control del laboratorio en situaciones reales.

#### **2.2.1.3. Experimento Natural.**

El experimento natural se lleva a cabo, cuando ocurre un fenómeno o cambio social sin que en éste haya intervenido la acción del investigador. Lo que el investigador hace, es aprovechar de manera oportuna el acontecimiento, elaborando para ello, un diseño experimental. En la Fig. 2.1, experimento natural quedó fuera del cuadro que engloba la noción de método experimental, debido a que no comparte lo que caracteriza a éste, la creación y manipulación de la variable independiente.

Por otra parte, no deben confundirse los estudios de campo con los

experimentos de campo, pues las técnicas y sus implicaciones son diferentes; lo único que tienen en común es el ambiente natural en el que se llevan a cabo.

Por ejemplo, en el caso del meteoro Paulina de 1997 en Acapulco. Si un investigador hubiera deseado estudiar la actitud de solidaridad de la población ante una situación de desastre, podría levantar su información en Acapulco, y en otro centro turístico semejante donde no hubiera llegado el meteoro, usando al segundo centro como grupo control.

### 2.2.2. **Expost-Facto y Correlacional.**

Con respecto al método ex-post-facto ha habido tradicionalmente desacuerdo ya que para algunos forma parte del diseño experimental cuando por su propia naturaleza retrospectiva impide ejercer control experimental por lo que, entonces, forma parte de los métodos no experimentales (Ver esquema de la figura 2.4). El método correlacional estudia la **co - variación, la co - relación** conjunta entre 2 o más variables y permite la predicción: **Si X entonces Y**. Las correlaciones o relaciones predictivas están abiertas a explicaciones alternativas, cuestión que no sucede con el diseño experimental.

---

### **Fig. 2.4. PARTICULARIDADES DEL MÉTODO EX-POST- FACTO.**

El experimento comienza con grupos iguales y los trata diferente; mientras que el ex post facto ( en el que la VI se escoge después del hecho sin manipulación directa del investigador) comienza con grupos desiguales y luego se los trata igual.

<b>INICIALMENTE DESIGUALES</b>	<b>TRATAMIENTO IGUAL</b>
Fumadores - No Fumadores	☹ Respuestas al estrés
NSE alto - NSE bajo	misma Prueba de personalidad
☠ Alcoholismo - No alcoholismo	misma Prueba de identidad sexual

En consecuencia **No** se puede inferir causalidad o relación funcional, ya que los

## sujetos son diferentes en muchas otras variables además de la VI.

En el "experimento" *expost-facto*, el investigador procede en dirección opuesta a la del experimento natural, pues trata de remontarse de los efectos hasta las causas que los produjeron. Es decir, que el investigador entra en acción (recogiendo información) una vez que ha ocurrido un acontecimiento. Para algunos estudiosos de la materia, este método es considerado como experimental; para otros, sin embargo, es **correlacional**, en tanto, al igual que en el natural, no se da la manipulación de la variable independiente, como tampoco el ejercicio de una serie de controles que caracteriza al método experimental.

En lugar de levantar la información en dos centros turísticos como en el caso antes descrito, se levanta información en sólo uno de ellos, el afectado por el meteoro, y se registran las respuestas que los Sujetos dan a preguntas referidas al pasado, comparándose con lo que sucedió en el presente.

En el caso de este método podría decirse que la investigación se inicia con dos grupos desiguales (por ejemplo, uno con nivel socioeconómico bajo y el otro con alto) y se les trata igual, se les aplica los mismos instrumentos.

### **2.2.3. Método Cuasi-Experimental.**

Este método se ajusta a un diseño aplicable a situaciones sociales naturales. Es, podría decirse, un experimento de campo en el que el control de las variables y de las condiciones experimentales (del **set** experimental) no se puede ejercer en la medida que exigen los estándares del método experimental.

En el caso del cuasi-experimento, el investigador debe conocer, de manera precisa, cuáles son las variables específicas que el diseño que está manejando no controla, para detectar en los aspectos pertinentes las interpretaciones distintas que puedan hacerse de sus datos (Campbell y Stanley, 1973). Este tipo de aproximación, requiere de mayor experimentación, precisamente en aquellos aspectos que hacen que los resultados puedan ser interpretados en más de una forma, o que contribuyen a resultados equívocos.

Por ejemplo, en un estudio donde los grupos fueran igualados estadísticamente, controlando así una o mas variables extrañas



importantes, pero en el cual se tuvieran que aplicar dos instrumentos que miden diferentes variables, se podría controlar el efecto del orden de presentación por medio de un balanceo, para asegurar que los efectos o las diferencias encontradas no se debieran a esta situación sino a la variable independiente.

## **2.2.4. Método No Experimental.**

### **2.2.4.1. Estudio de Campo y Encuesta.**

El estudio de campo requiere un contacto directo con los individuos en un medio natural mientras que los estudios de encuesta se ocupan de personas numerosas y dispersas. Entre las diferencias principales de ambos métodos de investigación deben señalarse las siguientes: la encuesta tiene mayor alcance pero menor profundidad. Asimismo, mientras la encuesta trata de representar algún universo conocido, de tal modo que la representatividad es fundamental, en el estudio de campo se intenta hacer una descripción completa de los procesos investigados y por lo tanto la representatividad, puede no ser importante.

En el estudio de campo se trata de estudiar una comunidad o grupo específico, tomando en cuenta las interrelaciones que se establecen entre aspectos de la estructura y la interacción social que se produce. Por otro lado, en la encuesta, los procesos que interesan, su comportamiento y desarrollo se infieren de los resultados estadísticos. En los estudios de campo el análisis de los datos puede ser cuantitativo y/o cualitativo. Así, se puede emplear estadística descriptiva o inferencial; análisis del contenido o del discurso. En general, son aplicables a esta clase de investigaciones, la amplia gama de dispositivos que existen al interior de dichos procedimientos analíticos. La encuesta sólo puede proporcionar asociaciones y difícilmente se podrían interpretar los resultados en términos de relaciones funcionales o causales. Podrían obtenerse, sin embargo, dichas relaciones si se realiza un estudio longitudinal con un diseño antes y después, entrevistando a la misma muestra por lo menos dos veces, antes y después de una dada condición y si además se recurre a dos o más grupos control. Pero entonces, como puede verse, se ha modificado un método asociativo o correlacional hasta transformarlo a uno experimental, o para ser más precisos, se han combinado ambos métodos.

Si el interés es establecer causa-efecto ya se ha señalado que el diseño más apropiado es uno que se ajuste al método experimental, en tanto permite establecer un mejor control sobre las variables. Estos dos tipos de aproximación, experimental y no-experimental (estudios de campo, encuestas, en suma, estudios asociativos o correlacionales) tienden a ser contrastados de tal forma que aquellos que favorecen la encuesta, se oponen al experimento poniendo énfasis en su no representatividad y en su dificultad de extrapolar los resultados obtenidos en una situación artificial a una de la vida real (problemas de validez experimental). Por su

parte, los experimentadores hacen resaltar la dificultad que la encuesta impone para poder controlar las variables importantes, así como la imposibilidad que el investigador tiene para provocar el evento a estudiar. La crítica principal, sin embargo, se hace en torno a su imposibilidad de establecer causalidad. Haciendo a un lado estos argumentos, cabe señalar que el criterio que debe seguirse en la selección del método a utilizar depende de cuál es el más apropiado y en el que sobresale por ejemplo, el carácter del problema a investigar. Una posición más flexible y más fructífera en el campo de la investigación, consideraría a ambos métodos en una relación de complementareidad, utilizables en diferentes etapas de la investigación, para producir conclusiones precisas (experimento) y generalizables y representativas.

#### **2.2.4.1.1. Tipos de Encuesta.**

Como ya lo señalamos la encuesta puede definirse como la recolección sistemática de datos a través de entrevistas o de la aplicación de otros instrumentos. Por lo general se aplica a grupos grandes y dispersos (*encuesta descriptiva*) a fracciones representativas (*encuesta por muestreo*) a toda la población (censo).

**Ventajas de la encuesta:** es útil para describir características de grandes poblaciones, son flexibles y económicas si se toma en cuenta el tamaño de las muestras.

##### **C encuesta descriptiva**

. Proporciona una representación exacta del fenómeno buscando regularidad del mismo y surgen las bases para formular hipótesis. El propósito de la encuesta descriptiva o encuesta de tipo censo, como también se le conoce, es obtener información en relación a los atributos de una o más variables, ya sea de toda una población, o de una muestra representativa a partir de la cual se hacen inferencias de la población de la que dicha muestra se extrajo. Cuando éste es el caso, se tiene que recurrir a técnicas de diseño de muestreo (diseños estadísticos) que producirán los resultados óptimos en función del tipo del problema, precisión que se desee, y del tiempo y del dinero con que cuenta el investigador. Cuando las encuestas son representativas, es posible establecer relaciones entre variables y predicciones, comparando los resultados de las encuestas obtenidas en diferentes tiempos, las encuestas descriptivas no responden a los "por qué", lo que hacen es describir. Se utilizan cuando la información requerida no puede obtenerse de otras fuentes. Los principales usuarios de este tipo de encuestas son las instituciones gubernamentales y las grandes empresas comerciales: en el primer caso, se utilizan por ejemplo, para la solución de problemas sociales relacionados con el crecimiento demográfico; en el segundo un ejemplo podría ser, para conocer el "rating" de un programa

(sólo por señalar, en ambos casos, un ejemplo de entre muchos). En el esquema que sigue (fig. 2.5) se exponen los tipos de encuesta más frecuentemente utilizadas.

---

Fig. 2.5. Principales tipos de encuesta

**C explicativa o analítica..** Sigue el modelo de los experimentos con la diferencia que busca representar ese diseño en un medio natural, prueba hipótesis, trabaja con muestras homogéneas lo que equivale a la igualación por pareamiento en el experimento.

**C diagnóstica.** Busca causas posibles en ámbitos relativamente desconocidos.

**C predictiva.** Sirve para estimar situaciones futuras.

**C exploratoria.** Se utiliza para obtener información básica en áreas de estudio poco conocidas .

---

### **Encuesta Analítica**

El interés en la encuesta analítica reside en los "por qué", es decir, en la explicación de relaciones entre variables particulares; en la búsqueda de asociaciones y comportamiento de las mismas. Con la encuesta analítica al igual que con los diseños experimentales, se comprueban hipótesis. Aún cuando estas encuestas tienen mucho en común con el experimento, no deben perderse de vista sus diferencias (que hacen que sean más o menos adecuadas para diferentes tipos de problemas) que intervienen en las conclusiones, predicciones y generalizaciones que pueden hacerse. La encuesta analítica cuenta con procedimientos que permiten estudiar las relaciones de las variables bajo diferentes condiciones (especificación); encontrar asociaciones contingentes, clarificar el verdadero valor de la relación, controlar factores contaminantes, especificar las condiciones que facilitan las relaciones, así como aquellas que las inhiben y determinar la naturaleza de la variable independiente (Rosenberg, 1968). Debe señalarse que toda la serie de controles que se pueden llevar a cabo se hacen a nivel estadístico. Es decir, una vez que se han recolectado los datos, (correlaciones, correlaciones parciales, análisis de regresión, tabulaciones cruzadas, clasificación de subgrupo, apareamiento de submuestras, etc.). Esta es una de las diferencias que tiene la encuesta analítica con el experimento, donde el control descrito, se realiza en la misma situación experimental. Sin embargo en la encuesta analítica se intenta ir

más allá de la sola prueba de hipótesis. Las razones que se ofrecen son las siguientes (Rosenberg, 1968):

a) a pesar de que la hipótesis proviene de una teoría y esté respaldada por los datos, éstos no prueban la teoría que dio lugar a la hipótesis, sino que solamente la respaldan; es decir, los hallazgos pueden también ser consistentes con otras teorías;

b) aún si la hipótesis ha sido confirmada en el análisis de la encuesta, todavía existe la posibilidad de que las relaciones entre las variables sea espúrea, y

c) se desperdicia información valiosa si el investigador se limita a la prueba de hipótesis previamente formuladas. Es más importante producir nuevas hipótesis. En relación a la prueba y producción de hipótesis, se recurre en la encuesta analítica al procedimiento de elaboración de las relaciones encontradas entre dos variables. Una de las estrategias de dicho procedimiento es introducir una tercera variable en la relación original para especificarla haciéndola significativa y más exacta. Este procedimiento permite contestar a preguntas tales como ¿por qué? y ¿bajo qué condiciones?. Conforme avanza el análisis van surgiendo nuevos datos, así como nuevas ideas. Con este tipo de **reanálisis** se desarrolla un interjuego entre teoría y datos; es decir, el procedimiento impide la separación entre ellos. El papel del investigador se vuelve más flexible y aún cuando se deja conducir por los datos, es él quien da dirección al curso del análisis.

Dada la importancia del proceso de elaboración en los diseños de encuesta analítica, se abordará más específicamente sobre el tema siguiendo los lineamientos planteados por Rosenberg (1968).

Cuando el investigador lleva a cabo encuestas analíticas, su primer propósito es encontrar y determinar el tipo de relaciones existentes entre las variables. Estas pueden ser simétricas, recíprocas y asimétricas.

Una relación simétrica es aquella en que las variables no se influyen entre sí. Se pueden encontrar diferentes clases de relaciones simétricas: a) ambas variables son indicadores alternativos del mismo concepto. Por ejemplo, la dilatación de la pupila y la aceleración del ritmo cardiaco, aparecen juntas en un estado de intoxicación farmacológica; ninguna influye en la otra y ambas son indicadores de dicho estado; b) las dos variables son efecto de una misma causa: por ejemplo, problemas de aprendizaje y deserción pueden tener como causa común, la pertenencia a la clase social económicamente desprotegida (pobre alimentación y la necesidad de incorporarse a la fuerza de trabajo); c) ambas

variables muestran la interdependencia funcional de los elementos de una unidad necesarios para la sobrevivencia de la misma. Por ejemplo, trabajo asalariado y capital, no puede existir uno sin el otro, y ambos funcionan distintamente para que sobreviva el sistema; d) las dos variables son partes no interdependientes de un sistema o complejo común; por ejemplo, vestir de acuerdo con el "último grito de la moda" e ir a las "discotecas" forman parte de un "estilo de vida" de la juventud de cierta clase social; e) son simplemente relaciones fortuitas: por ejemplo, la aparición del **rock and roll** y el inicio de la era espacial.

Por otra parte, una relación recíproca es aquella en la que ambas variables se influyen entre sí. Cada variable es causa y efecto: inflación ---> aumento del costo de la vida ---> incremento de salarios ---> inflación posterior, etc.

Las relaciones asimétricas, por otro lado, son las de mayor interés para los investigadores. En éstas, una variable (denominada independiente) es esencialmente responsable de otra (la variable dependiente). Aquí es pertinente recordar que el concepto de causalidad que maneja el científico social no lo es en un sentido estricto. En las ciencias exactas la causalidad sólo se da cuando se cumplen los principios de unidad, aislamiento, productividad, invariabilidad y unidireccionalidad. Mientras que el sentido de causalidad en las ciencias sociales es en realidad percibido desde un punto de vista extenso. Lo que significa que el principio de unidireccionalidad sólo, se plantea como suficiente, quedando ésta, la causalidad, determinada por un lado, por la temporalidad de las variables (la que ocurre antes en el tiempo es más probable que sea la causa, que la que se da después); y por el otro, por lo estático o alterable de las mismas (en este caso, las variables más estables o menos alterables tienden a determinar la dirección de la causalidad). Por ejemplo, el sexo y color de la piel son más estables (aunque la ciencia y la tecnología hacen tambalear este aserto si se toma en cuenta lo frecuente que se está volviendo el cambio quirúrgico del sexo y la blancura de la piel de Jackson...) que la clase social.

Las relaciones asimétricas pueden ser de diferentes clases. La primera se refiere a la relación entre un estímulo y una respuesta, y un ejemplo sería la contracción pupilar ante la estimulación lumínica. La segunda se refiere a la relación que se da entre una disposición y una respuesta. Entre las disposiciones, Rosenberg (1968) señala a las actitudes, las habilidades, los reflejos, los hábitos, las pulsiones y las características de personalidad. Un ejemplo sería la relación entre el liberalismo y la participación en las demostraciones a favor de los derechos humanos. Una tercera clase de relación asimétrica es la que se da entre una propiedad del individuo y una disposición. Entre las

propiedades del individuo, el mismo autor señala, entre otras, la raza, el sexo, la edad y la nacionalidad. Estas se distinguen de las disposiciones en que no dependen de las circunstancias, tienen un carácter de perdurabilidad, mientras que las disposiciones se consideran como tendencias o probabilidades de hacer o decir algo. Un ejemplo de este tipo de relación sería la que se da entre edad y conservadurismo. Un cuarto tipo de relación es aquella en donde la variable independiente es una precondition necesaria de la dependiente; un ejemplo sería la relación que existe entre alcanzar cierta edad y el derecho a votar (sin embargo, la primera no es la causa de la segunda, sólo la hace posible). La quinta se refiere a la relación inmanente entre dos variables. El autor propone como ejemplo de esta relación, el hecho constatado en diferentes ocasiones, en el que una organización que originalmente fue motivada por la democracia degenera eventualmente en una oligarquía. Esto es, las cualidades inherentes en la naturaleza de un organismo da lugar a ciertas consecuencias. Una sexta clase se refiere a la asociación entre los medios y los fines; un ejemplo en este caso sería la relación que existe entre el avance tecnológico y la aceleración del grado de industrialización en alguna sociedad o nación.

Cuando el investigador plantea una hipótesis, está asumiendo la existencia de una relación asimétrica. La relación asimétrica original que plantea la analiza o somete a prueba siguiendo un procedimiento conocido como proceso de elaboración. Este consiste en introducir una tercera variable o factor de prueba, con el propósito de determinar si la relación original entre **x** (variable independiente) **y** (variable dependiente) se debe o no, a otra variable, a **z**. En otros términos, el investigador intenta asegurarse de que la relación que estudia es "genuina", se mantiene; o si por el contrario, cambia o desaparece, por que ella está dada por su relación con otra variable (la tercera, o factor de prueba, **z**).

La forma en que se determina si la relación original se mantiene es controlando o manteniendo constante al factor de prueba. Esto se logra estratificando dicho factor y examinando las relaciones contingentes respecto a la original. Estratificar significa dividir al factor de prueba en sus categorías componentes.

Por ejemplo, si la relación original es entre la actitud hacia la planificación familiar y el género, estratificar en una tercera variable o factor de prueba (hijos: si se tienen o no), significa reanalizar los datos por separado para aquellos Sujetos que tienen hijos y los que no tienen.

Las variables que se toman como factores de prueba son aquellas

que el investigador asume que se asocian con las variables dependiente e independiente. De acuerdo con Rosenberg (1968) no todos los factores de prueba tienen el mismo significado. Se pueden distinguir seis de ellos según se considere a la tercera variable como: variable extraña, componente, interventora, antecedente, supresora y distorsionadora.

**Variables Extrañas.**- Cuando existe una relación original entre dos variables, por ejemplo: las mujeres casadas ven más telenovelas que las solteras, y se introduce una tercera variable (escolaridad: alta y baja) y la relación original desaparece, se dice que la tercera variable o factor de prueba es una variable extraña. Es decir, la variable estado civil no era la causa explicativa de la exposición a las telenovelas. Más bien, que las mujeres vieran más o menos esta clase de programas, dependía del mayor o menor grado de escolaridad de las mismas.

**Variables Componentes.**- Muchas de las variables que se estudian en las ciencias sociales son variables complejas. En estos casos el investigador se pregunta cuál es el elemento decisivo de esa variable compleja que produce la relación entre la variable dependiente y la independiente. Por ejemplo, si se observa una relación original que señala que en una situación de trabajo la clase media tiene mayor capacidad de decisión que la clase trabajadora, se puede introducir una tercera variable como libertad para solicitar o no, supervisión. Si la relación original se mantiene en las asociaciones contingentes al controlar dicho factor de prueba, se puede decir que éste, el acceso libre a la supervisión, es una dimensión o componente importante de la variable compleja. Si en las asociaciones contingentes desapareciera la relación original, podría significar que la dimensión o elemento importante de la variable compleja es otro y no el factor de prueba; o bien, que la variable compleja como totalidad es la importante.

**Variables Interventoras.**- Las variables interventoras son aquellas que se encuentran localizadas dentro de la cadena causal, entre la variable independiente y la dependiente. (variable independiente - ---->variable interventora ---->variable dependiente). Para poder decir que una variable es interventora, se deben cumplir las siguientes relaciones: a) una relación original entre la variable dependiente y la independiente; b) una relación entre la variable independiente y la variable interventora (que funciona como variable dependiente); c) una relación entre la variable interventora (funcionando como variable independiente) y la variable dependiente. Un ejemplo sería: existe una relación original entre la edad de un grupo de profesores (jóvenes-maduros) y ausentismo en el trabajo (alto-bajo). Se introduce una tercera



variable como factor de prueba: desempeño de otros trabajos remunerados (un sólo trabajo-dos o más trabajos). Se observa que al controlarla, la relación original desaparece. Pero también se observa que cuando se emplea desempeño de uno o más trabajos como variable dependiente, la relación original se mantiene; y que si se emplea a esta misma tercera variable como variable independiente de la dependiente original, la relación original también se mantiene.

**Variables Antecedentes.**- Las variables antecedentes se localizan en la cadena causal antes de la variable independiente (variable antecedente ----->variable independiente ----->variable dependiente). Se les considera como una influencia verdadera y real; sin embargo, no explica la relación entre la variable independiente y la dependiente, sino que clarifica las influencias que precedieron a dicha relación. Para probar que la variable seleccionada es de hecho la variable antecedente, se requiere, de acuerdo con el autor, del cumplimiento de tres requisitos: a) las tres variables (antecedente, independiente y dependiente) deben relacionarse; b) cuando la variable antecedente se controla, la relación entre las variables independiente y dependiente no debe desaparecer y c) cuando se controla la variable independiente, la relación variable antecedente-variable dependiente, debe desaparecer.

La segunda condición implica, como ya se señaló, que la variable antecedente "no explica" la relación variable independiente-variable dependiente sino que la precede en la secuencia causal: afecta a la variable independiente, y ésta a la variable dependiente.

Por ejemplo, la relación original es que los adultos tienen mas preferencia por escuchar programas radiofónicos religiosos que los jóvenes. Como tercera variable o factor de prueba se tiene el nivel escolar formal alcanzado por los Sujetos. Cuando se analizan los datos controlando el nivel escolar formal alcanzado, se sigue dando la relación entre edad y preferencia por escuchar programas radiofónicos religiosos. Pero cuando se controla la edad de los sujetos, la relación entre escolaridad formal alcanzada por los sujetos y preferencia por escuchar programas religiosos desaparece.

**Variables Supresoras.**- En ocasiones se obtienen resultados negativos; es decir, no se encuentra la relación postulada entre la variable independiente y la dependiente. En este caso, se pueden dar esos resultados, por la intervención de una variable supresora. En otras palabras, cuando no se produce la relación original esperada, y se introduce una tercera variable o factor de

prueba, y de su control surge la relación original planteada, se dice que esta tercera variable o factor de prueba es una variable supresora. Por ejemplo, un investigador asume (basado en otros estudios realizados, en las teorías de interacción y liderazgo, etc.) que se dará una relación entre estudiantes que destacan como deportistas y liderazgo en sus grupos de clase. Al no encontrar una relación significativa entre ellas, mete como factor de control una tercera variable: desempeño académico (alto-bajo). Encuentra que la relación deportista-liderazgo aparece cuando el estudiante se desempeña bien académicamente.

**Variables Distorsionadoras.**- Cuando se da una relación original positiva (o negativa) entre la variable dependiente y la independiente, y se controla con una tercera variable o factor de prueba, y las relaciones contingentes se vuelven negativas (o positivas); esto es, cuando cambia la dirección original, se dice que la tercera variable es una variable distorsionadora. Un ejemplo sería el siguiente: un investigador encuentra que los niños que son juzgados por sus compañeros como más inteligentes, son también considerados como los menos agradables o simpáticos. Es decir, a mayor inteligencia menor simpatía (relación original negativa). Introduce como factor de control la variable accesibilidad y cooperación con los compañeros de clase (alta-baja). La relación original negativa se vuelve positiva cuando la variable inteligencia va acompañada con alta accesibilidad y cooperación. De este modo, inteligencia y simpatía mantienen una relación positiva.

Resumiendo, se puede decir que en un diseño de encuesta analítica, el investigador es guiado por hipótesis; sin embargo, el análisis también genera nuevas hipótesis. Se considera que la prueba de hipótesis previamente formulada es un paso importante, pero al mismo tiempo se asume que su poder es limitado por lo que se recurre a una serie de estrategias, algunas de las cuales ya se han señalado, con el propósito de lograr resultados más significativos.

#### **2.2.5. Diseños de Encuesta.**

Entre los principales diseños de encuesta están los siguientes:

**Transversal no ponderada.** es la mas frecuentemente utilizada. Se efectúa una sola vez y sirve para determinar las características de una población en un momento particular.

**Transversal ponderada.** Se trabaja con muestras que sobre - estiman deliberadamente la variable de interés. . Por ejemplo: con sobre - estimación del nivel socioeconómico alto. Esto es, tener mas sujetos de este nivel para así lograr un grupo con un tamaño aceptable que permita hacer comparaciones con sujetos de otros niveles socioeconómicos.

**Muestras contrastadas.** se tienen subgrupos extremos en la variable de interés. Esto es, se consideran los valores extremos de una distribución (los más bajos y los más altos) si no se encuentran diferencias estadísticamente significativas, entonces se concluye que la variable independiente (vi) no influye. Sin embargo, si las diferencias resultan significativas no implica relación lineal a lo largo de todos los valores de la distribución; se tienen que probar valores intermedios.

**Longitudinales. (secciones transversales sucesivas):** este diseño se utiliza para el estudio de cambios o de evolución en los fenómenos de interés.. Se entrevista dos o más veces a las mismas muestras de sujetos.

**Diseño antes-después.** El diseño mas simple de los estudios longitudinales es el de antes y después en el que se ve a los mismos sujetos dos veces.

**Encuesta de tendencias (trend analysis).** Implica dos o más mediciones. Lo que interesa es una población y cada vez se estudia una muestra de la misma aunque no necesariamente compuesta por los mismos sujetos. Permite ver la tendencia, la dirección que sigue el cambio, las respuestas dadas a un objeto determinado a lo largo de un tiempo determinado.

**Superposición de muestras (overlapping design).**

El diseño de superposición de muestras es muy útil para cubrir un amplio periodo de investigación en menor tiempo . Un ejemplo de este diseño es el siguiente en el que con un estudio de 6 años se cubrió un periodo de 12 años (Holtzman, Diaz-Guerrero, Swartz y Lara Tapia, citados por Gómez-Pérez-Mitré (1973).

A Ñ O S E S C O L A R E S												
EDAD DE INICIO		6 AÑOS										
GRUPO I	1	2	3	4	5	6						
GRUPO II				4	5	6	7	8	9			
GRUPO III							7	8	9	10	11	12
EDAD	DE		TERMINO							17 AÑOS		

Con el diseño de superposición de muestras se obtiene una curva continua de 6 a 17 años de edad que corrige tendencias del desarrollo debidos a efectos de la practica , aprendizaje, maduración de los sujetos, etc.

Un diseño como el de sobreposición permite preveer y disminuir algunos de los efectos mas limitantes de los estudios longitudinales como es el de la **mortalidad experimental** ya que se espera una perdida del 25% o mas después de un año debidos a varias razones como pueden ser :

traslados, perdida de interés por la investigación, etc. Y el efecto de practica, que se refiere a la sensibilización que sufren los sujetos por las entrevistas repetidas, aburrimiento o perdida de motivación por la misma razón.. *Los problemas que acabamos de señalar también pueden disminuirse utilizando muestras o grupos control.*

**Cohorte.** Con este diseño se pueden observar cambios en muestras de subpoblaciones especificas. De hecho son grupos de personas que comparten algo en común.

un ejemplo de este diseño es el siguiente: encuestas aplicadas cada 5 años para estudiar la conducta política de una cohorte compuesta por personas nacidas durante los años de la crisis política de 1968:

☐ muestra de sujetos de 15 a 20 años en 1983

☐ muestra de sujetos de 25 a 30 años en 1988;

☐ muestra de sujetos de 35 a 40 años en 1993;

*Como puede verse cada muestra representa sobrevivientes de la cohorte nacida entre 1968 y 1973.*

**Paneles.** Con este diseño se realizan entrevistas o aplicaciones reiteradas a los mismos sujetos. Se obtiene una medida mas sensible del cambio de lo que pueden aportar 2 muestras independientes aunque sean de la misma población

### **2.2.5.1 Comentarios adicionales sobre el control en el proceso de investigación.**

El propósito primero y permanente del investigador independientemente del diseño que utilice es: limpiar y evitar relaciones falsas. Distinguir las *variables explicativas* de las *exógenas* falsas, en otros términos detectar la presencia de **otras fuentes de variación:**

**I.** La fuente única de variación debe ser la relación de las

variables explicativas o experimentales que constituyen el interés del experimentador, a saber:

**Variable independiente / variable dependiente**

**II. Variables exógenas** que se pueden controlar a través de procedimientos de selección y/o de estimación.

**III Variables exógenas** no controladas que se pueden confundir con variables de la clase i..

**IV. Variables exógenas y no controladas** consideradas como errores aleatorizados (forma de control experimental). En experimentos ideales realmente se encuentran aleatorizados.; Con otros métodos solo se asume que están aleatorizados.

**El objetivo de los diseños eficientes (experimentales o no experimentales) es ubicar todas las variables exógenas que se pueda en la clase II.**

En el experimento se aleatorizan las variables exógenas para poner las variables de la clase III en la IV. (en un experimento ideal no existen variables de la III clase).

En la encuesta analítica se utilizan diferentes clases de controles para separar las variables de la III clase de las de la primera clase (I), mientras que en el experimento el esfuerzo se centra en las variables de la clase III.

En la encuesta analítica se miden los efectos de las variables de la clase i contrastando o estimando las de la clase IV.

En experimentos y encuestas ideales se utilizan pruebas estadísticas para separar los efectos de los errores aleatorios (clase IV) de los efectos de las otras variables.

La aleatorización se utiliza en experimentos y encuesta (explicativa o analítica). En los demás métodos esta ausente la aleatorización o el muestreo probabilístico.

Las diferencias entre experimento, encuesta analítica y demás métodos no residen en las técnicas estadísticas **sino en los métodos de manejo de las variables y en la selección de las muestras.**

Así, los problemas fundamentales del proceso de investigación son : **medición, representación y control.** En el esquema que sigue (Ver Fig. 2.7) se comparan experimento y encuesta según fuerza o debilidad explicativa, esto es en función **de limpiar la relación: variable independiente / variable dependiente.**

FIG. 2.7 . FUERZA Y DEBILIDAD DEL EXPERIMENTO Y DE LA ENCUESTA EN SU CAPACIDAD EXPLICATIVA

---

FUERZA DEL EXPERIMENTO

**CONTROL RELATIVO**

DEBILIDAD

**REPRESENTACIÓN**

FUERZA DE LA ENCUESTA

**REPRESENTACIÓN**

DEBILIDAD

**CONTROL RELATIVO**

---

Se ha utilizado en la figura 2.7 el término **control relativo** debido a la imposibilidad de controlar el total de las variables exógenas que pueden confundir la relación variable independiente / variable dependiente por lo que la ventaja del mejor control del experimento por encima de la encuesta, también es solo relativa. En la misma tónica puede decirse con respecto a la cuestión de la representación que **el problema del muestreo no forma parte de los propósitos del experimento.**

## **REFERENCIAS**

- Campbell, D.T. y Stanley, J.C. (1973): **Diseños Experimentales y Cuasi-experimentales de Investigación.** Buenos Aires: Editorial Amorrortu.
- Gómez, Pérez-Mitré, G. (1973). Family life styles and parental attitudes in Mexico and the United States. Thesis, M.A. Educational Psychology. The University of Texas at Austin.
- Kerlinger, F.N. (1982): **Fundamentos de la Investigación del Comportamiento.** México: Nueva Editorial Interamericana.
- Rosenberg, M. (1968): **The Logic of Survey Analysis.** Nueva York: Basic Books.

### **III. EL PROCESO DE INVESTIGACIÓN**

#### **3.1. Problema de Investigación.**

El proceso de investigación se inicia en el momento en el que la realidad sea de naturaleza física o social, plantea problemas a los estudiosos de un campo determinado del conocimiento, y éstos deciden encontrar las respuestas. Dichos problemas pueden expresarse de muy diversas maneras: como resultados contradictorios, como lagunas o resultados parciales; generalizaciones débilmente fundamentadas y aún carencia total de datos, todo esto con respecto a un fenómeno u objeto de conocimiento determinado. De esta manera, habiendo surgido un problema general de investigación, éste deberá traducirse en preguntas o interrogantes específicos y delimitados, como primer paso necesario en el intento que tiene como propósito darle una solución.

El planteamiento del problema de investigación, se refiere precisamente a su especificación y delimitación. El planteamiento de los problemas de investigación que forman parte de las disciplinas sociales debe reunir ciertas características, por ejemplo referirse a cuestiones resolubles, solucionables. Con esto se quiere decir que éstas deben ser susceptibles de investigación concreta; debe estar correcta y adecuadamente especificado, es decir, clara y precisamente definido; debe referirse a una cuestión que puede ser probablemente cierta o falsa. Decimos que probablemente cierta o falsa porque la ciencia es un campo de verdades relativas a las que hay que aproximarse en forma sucesiva.

Cuando un problema de investigación está mal planteado, impide aproximarse a cuestiones resolubles. Un problema mal planteado puede deberse a que la pregunta que se hace el investigador es demasiado general (éste debería limitarla antes de intentar aproximarse a su estudio); o bien, puede ser que los conceptos o términos con los que está planteada la pregunta no estén correctamente definidos.

#### **3.2. Proyecto de Investigación.**

Una vez que el investigador ha planteado correctamente su problema de investigación, deberá proceder a planear su investigación. El plan o proyecto de la investigación abarca los siguientes pasos, que deberán considerarse antes de iniciar el estudio:

- a) Revisión de la literatura
- b) Planteamiento del problema
- c) Establecimiento de hipótesis
- d) Definición de variables
- e) Determinación del diseño de investigación
- f) Determinación de la muestra que se estudiará; definición de la población de la que se extraerá la muestra; el procedimiento de selección de la misma; características de los sujetos a ser investigados; su localización; etc.
- g) Selección de los instrumentos que se emplearán, si ya existen, para recoger los datos pertinentes en este caso deberán especificarse los valores psicométricos; si los instrumentos van a ser construidos, se determinará el procedimiento que se empleará para elaborarlo, y se establecerán las características psicométricas pertinentes.
- h) Determinación de la forma en que se recogerán los datos o de las situaciones experimentales que se llevarán a cabo.
- i) Determinación de los procedimientos estadísticos que se emplearán para responder al problema de investigación planteado originalmente.

Planear la investigación de antemano tomando en cuenta los aspectos señalados, tiene dos ventajas fundamentales para el investigador: primero, conocer las posibilidades que se tienen de llevar a cabo la investigación de manera válida según los estándares que exige la producción de conocimiento científico, y segundo, determinar cuáles son los posibles obstáculos que enfrentará, para poder preverlos y superarlos previamente hasta donde sea posible, evitando con esto soluciones improvisadas que pudieran invalidar los resultados que se obtuvieran. Estas razones justifican plenamente la planeación y ejecución del proyecto de investigación.

### **3.3. Desarrollo de la Investigación.**

En seguida se describirán los pasos arriba señalados, que corresponden a los que generalmente se siguen cuando se planea una investigación. Esto se hará dentro de los límites planteados en el presente trabajo, que intenta proporcionar una guía práctica del proceso de investigación dentro del campo de las disciplinas sociales.

#### **3.3.1. Revisión Bibliográfica.**

Como primer paso en la realización de una investigación se tiene que revisar la literatura. Esto es, habrá de buscarse a través de todos los medios disponibles lo que se haya escrito respecto al problema de investigación, o se relacione con éste.



Existen un gran número de revistas especializadas, tanto nacionales como extranjeras que resultan de gran utilidad para el estudioso de este campo. Se cuenta también con un gran número de obras publicadas, que generalmente se conocen como Lecturas o Manuales que también proporcionan información especializada. Por supuesto también existen diversos centros de investigación tanto en el país como en el extranjero, así como centros de información computarizada a los que se puede recurrir para obtener el mayor número de datos sobre el problema de interés. La revisión de la literatura reditúa múltiples beneficios al investigador; entre ellos además de los mencionados anteriormente, están los siguientes:

a) Evita duplicaciones inútiles, pues posiblemente las respuestas que se buscaban ya han sido dadas por otros investigadores.

b) Informa sobre los obstáculos que otros investigadores enfrentaron al realizar un estudio similar al de interés, proporcionando con esto ideas y estrategias que facilitan el trabajo de investigación a emprender.

c) Aclara dudas sobre la mejor manera de obtener datos sobre instrumentos que se pueden utilizar o sobre formas de aproximación y recolección de datos.

d) Facilita el surgimiento de ideas con respecto a variables, hipótesis, formas de control, etc., que antes de la revisión bibliográfica no había pensado el investigador.

e) Informa de las limitaciones que investigadores anteriores no pudieron superar, por diversas razones, así como de las sugerencias que estos hacen al respecto y que por lo general se refieren a nuevas investigaciones.

En suma, la revisión bibliográfica permite conocer, profundizar y ampliar, las perspectivas con respecto al problema que se abordará; además esa información constituye de hecho, los antecedentes teóricos y empíricos del estudio a realizar; es decir, constituye el marco de referencia teórico conceptual que contribuirá a explicar los resultados que se obtengan.

### **3.3.2. Problema.**

Habiendo revisado la literatura pertinente y consultado con especialistas en la materia, el siguiente paso, el planteamiento del problema, resulta más sencillo. La información recopilada

facilitará el cumplimiento de los requisitos antes señalados. Se tendrá así, un problema de investigación correctamente planteado, clara y precisamente definido. Esto es, se podrá formular una pregunta, exactamente la que se quiere, y se podrá responder de manera adecuada.

### **3.3.3. Hipótesis.**

El siguiente paso es el establecimiento de las hipótesis. En este apartado se hablará de las hipótesis en general y de algunos tipos de hipótesis en especial. Una hipótesis representa la posible solución al problema de investigación. Se puede definir como la relación potencial que existe entre dos o más variables, y que puede ser probablemente cierta o falsa. En realidad el problema de investigación es una pregunta que se refiere a la posible existencia de una relación entre dos o más fenómenos o variables. De esta manera, la hipótesis expresa el establecimiento adecuado dentro de los requisitos de la ciencia de esa misma pregunta, y la investigación en su conjunto es un intento de dar respuesta a dicha pregunta (McGuigan, 1993).

Las hipótesis tienen funciones específicas en la investigación: la principal es que la orienta y la dirige. Por otra parte, algunas hipótesis pueden ser mejores que otras en cuanto a su planteamiento. Esta "bondad" de las hipótesis se logra en la medida en que se cumplen los siguientes requisitos:

-deben ser relevantes al problema de investigación; deben ser simples o parsimoniosas; sus términos (las variables) deben de ser susceptibles de cuantificación, o por lo menos de definición operacional.

#### **3.3.3.1. Hipótesis Conceptual.**

Esta hipótesis se refiere a la formulación teórica que respalda el trabajo de investigación; es la hipótesis que servirá para relacionar los resultados con el marco teórico o conceptual. Esta hipótesis por lo general es la que se deriva de una teoría, ley o principio general, y la que finalmente será confirmada o desconfirmada según los resultados de la investigación.

#### **3.3.3.2. Hipótesis de trabajo.**

Esta se conoce también como hipótesis empírica. Difiere de la conceptual en el sentido de que se plantea en términos de variables concretas, definidas operacionalmente, e indica

específicamente qué es lo que se realizará en la investigación. De esta manera, en la medida en que la hipótesis de trabajo esté bien planteada, en esa medida estará indicando cómo se debe llevar a cabo la investigación. Es la hipótesis del investigador y representa la traducción operacional de una hipótesis conceptual. Por lo general la hipótesis de trabajo postula la existencia de diferencias entre dos o más grupos, partiendo del supuesto de que éstas se deben a la influencia o efecto de la variable independiente que está presente en un grupo y no en el otro: a la existencia en cantidades diferenciales de esa variable en el grupo o grupos experimentales, o a la ausencia de ella en el grupo control. También puede plantear la existencia de una relación, positiva o negativa entre dos o más variables. Para la confirmación de las hipótesis de trabajo se sigue un procedimiento indirecto, que depende de la prueba de las hipótesis estadísticas: la hipótesis nula ( $H_0$ ) y la hipótesis alterna ( $H_1$ ).

### 3.3.3.3. Hipótesis Estadísticas.

La hipótesis nula es parte del diseño estadístico, y se le conoce como ( $H_0$ ) la hipótesis de "las no diferencias". Es decir, con ella se plantea que no existen diferencias entre una muestra y una población, o entre dos muestras de una misma población, o entre dos muestras de dos diferentes poblaciones. Genéricamente se expresa como sigue: *No habrá diferencias estadísticamente significativas entre dos grupos de datos, y si las hay se deben al azar, y no a la variable que el investigador postula como la responsable.* Por otra parte, la hipótesis alterna ( $H_1$ ), es la hipótesis alternativa a la nula; es decir, mientras esta última plantea la existencia de no diferencias, la alterna postula que sí se encontrarán diferencias estadísticamente significativas. Esta hipótesis siempre se expresa en la misma dirección que la hipótesis de trabajo, de hecho, es una réplica de esta; pero la que se somete directamente a prueba es la alterna.

Una vez que se han establecido las hipótesis se especificará el nivel de confianza con el que se aceptarán o rechazarán. Por lo general en las ciencias sociales se trabaja con niveles de confianza de  $\leq 0.05$  y  $\leq 0.01$ . En situaciones comunes se usa más el primero que el segundo. Si la hipótesis nula se rechaza se acepta la alterna y como consecuencia puede concluirse que se confirma la hipótesis de trabajo correspondiente. Sucede lo contrario cuando se acepta la hipótesis nula.

Las hipótesis alternas pueden plantearse con o sin dirección. Son sin dirección cuando las hipótesis alternativas sólo indican que habrá diferencia entre los grupos, sin estipular a favor de cuál grupo, mientras que las de dirección sí lo hacen. En el primer caso se trabaja con una región de rechazo llamada de *dos colas*, y

en el segundo con una de *una cola* (Siegel, 1976). (En los ejemplos del capítulo IV quedará claro este aspecto de la prueba de hipótesis, niveles de confianza, región de rechazo, etc.).

Ahora bien, la prueba de las hipótesis estadísticas cumplen las siguientes funciones:

-Como es de esperarse el investigador se encuentra más involucrado con la hipótesis de trabajo, por ser esta la que él formula. Esto conlleva a la formación muy entendible, por otra parte, de expectativas en la dirección de que ésta se confirme. De esta manera, la aceptación o rechazo de la hipótesis nula, que es la que se prueba directamente, aparte de disminuir el costo afectivo relacionado con una decisión que puede ser contraria a sus expectativas, contribuye a que las decisiones sean más objetivas.

-Por otro lado, la prueba de la hipótesis nula asegura lo siguiente: la diferencia puede darse, pero tiene que ser lo suficientemente grande como para que si este es el caso no se deba al azar, sino a la variable independiente que está manejando el investigador. Esto asegura que de encontrarse la diferencia entre los grupos, su magnitud sea tan grande que la única explicación que se pueda dar para ella sea la existencia de la variable independiente en un grupo y no en el otro.

-Asimismo, en caso de ser rechazada, es decir, cuando se encuentran diferencias entre los dos grupos, es de esperarse que si se repitiera la misma investigación, en situaciones similares, se encontraría que 95 o 99 de cada 100 veces, se producirán los mismos resultados que se obtuvieron en esa única investigación.

Debe recordarse también que existen un sin número de trucos o posibles manipulaciones de los datos para que los resultados salgan como el investigador quiere, es decir, siguiendo los deseos o hipótesis de éste. En cambio, si se trabaja con la hipótesis nula será más difícil que se intente influir consciente o inconscientemente sobre los resultados para que éstos concuerden con la manera de pensar del investigador y no de acuerdo con la realidad, que es el objetivo último de la investigación dentro de cualquier ciencia. Por último debe señalarse que las hipótesis nula y alterna se plantean y se prueban cuando se procede al análisis de los datos, al análisis estadístico.

#### **3.3.4. Variables**

Reiteradamente se ha hablado de variables. Una variable es un evento o fenómeno que existe en diversas cantidades en la

realidad. Esto es, puede adquirir diferentes valores. Las variables también se pueden clasificar desde diferentes puntos de vista. Pueden ser discretas o continuas. Las primeras se refieren al hecho de que ciertos fenómenos, eventos o cosas, sólo existen en términos de cantidades que representan unidades enteras, mientras que las segundas expresan fenómenos o eventos que pueden ser fraccionados, es decir, que son lógicamente susceptibles de ser representados por números enteros y toda la gama de fracciones que se requiera. Un ejemplo de las primeras sería el sexo (masculino o femenino). Este atributo carecería de sentido si se expresara en fracciones (1.8 masculino). Lo mismo sucede con la variable número de hijos (se tienen 3 o 4 hijos, difícilmente se puede hablar de 3.6 hijos). Un ejemplo de las segundas puede ser la variable edad que se puede fraccionar en años, meses, días, horas, etc.

Asimismo, el ingreso familiar constituye otro ejemplo, se puede fraccionar en pesos y centavos. Así pues, las variables continuas permiten una gradación casi infinita dentro de una cantidad a la siguiente, cosa que no es posible con las variables discretas.

Las variables también pueden clasificarse como organísmicas, de estímulo y de respuesta (McGuigan, 1993). Las primeras se refieren como su nombre lo indica, a los atributos del organismo o Sujeto; están dadas intrínsecamente o bien se pueden producir en un organismo. La edad, el estado de ansiedad, el nivel de inteligencia, etc., son algunos ejemplos de estas variables. Las segundas se refieren a las que generalmente existen en el medio ambiente, y que algunos autores conceptualizan como estímulos que impactan o influyen sobre el organismo o Sujeto. Un ejemplo sería el nivel de ruido que puede afectar a una comunidad; el nivel de iluminación, o las normas de un grupo, son otros ejemplos. Las terceras son en realidad otras respuestas que pueden estar sirviendo como determinantes de una respuesta consecuente o posterior. Como ejemplo de éstas se tiene el número de errores cometidos por un Sujeto y el efecto que esto tuviera en la ejecución posterior en una prueba de inteligencia.

Desde otro punto de vista, las variables también se pueden clasificar como dependientes, independientes y extrañas. Las primeras se refieren principalmente a la conducta, y se llaman así por que dependen para su producción, aparición, modificación o desaparición, de otras variables. Las segundas son aquellas de las que dependen éstas. Es decir, las variables independientes son las causas, razones, motivos, circunstancias o condiciones, que producen, motivan, causan, modifican, influyen, etc., a las primeras. Las terceras reciben el nombre de extrañas porque se desconocen o simplemente se supone su existencia para tratar de controlar su efecto en la relación variable independiente-variable

dependiente. Precisamente reciben el nombre de extrañas porque son ajenas a dicha relación. Un ejemplo podría ser tiempo de lectura (variable independiente) y grado de memorización (variable dependiente) Una variable extraña a esa relación sería que algunos de los Sujetos tuvieran problemas de lectura. Si no se controla esta variable, los resultados de esa relación estarían contaminados por ella, pues estarían mostrando no sólo capacidad de retención sino también inhabilidad de lectura.

Ahora bien, debe tomarse en cuenta que cualquier conducta es multicausada. Es decir, que en su aparición o producción, intervienen varios factores o variables. Independientemente de que la conducta sea multicausada ésta será siempre tratada como la variable dependiente. En una investigación que se ajusta a un diseño univariado (una sola variable independiente) o bivariado (una variable dependiente y una independiente), por lo general lo que se intenta es determinar el efecto de una o dos variables independientes manteniendo constantes las otras, lo cual no implica que se deje de reconocer el efecto que esas otras variables pueden ejercer sobre la conducta que se estudia. La lógica que se sigue es que dado que en ese momento, en esa investigación, sólo interesa el efecto de una variable independiente sobre una variable dependiente o conducta, el resto de las otras variables que potencialmente pueden estar actuando sobre ésta, pueden ser identificadas y controladas en mayor o menor medida. Para ejercer ese control existen diversos procedimientos.

Lo anterior es importante por que en muchas ocasiones no se puede investigar, por diversas razones, el efecto simultáneo de varias variables. Existen desde luego, diseños apropiados (multivariados) que aunque más complejos proporcionan soluciones satisfactorias a dichas necesidades (estudiar la conducta en relación a sus causas múltiples). Debe aclararse que aunque dichos diseños no forman parte de los propósitos del presente libro, en el Capítulo V se revisa el diseño factorial que ya corresponde a los multivariados.

### **3.3.5. Control de Variables.**

El control de variables abarca básicamente dos aspectos principales: se dice que se controla una variable cuando ésta es definida y producida por el investigador. Cuando de alguna manera éste decide a qué Sujetos o grupos se les aplicará una u otra cantidad de la variable independiente, y a cuáles no. El control en este sentido se refiere a que la producción, registro y cuantificación dependen del investigador. Por otra parte, el control en su otro significado se refiere al que antes hemos

señalado, al control de las variables que pueden influir en la relación variable independiente-variable dependiente que se está investigando. A este se le denomina control de las variables extrañas, es decir todas aquellas otras independientes que no se están controlando en el primer sentido. El propósito de esta segunda clase de control es evitar una investigación denominada *confundida* o *contaminada*; en otros términos, se quiere evitar adjudicar a los resultados de la investigación los efectos de una o unas variables, las independientes (las controladas en el primer sentido) cuando en realidad los resultados no se deben a éstas sino a las extrañas. Se hará referencia pues, a los procedimientos de control de dichas variables.

Se había señalado anteriormente, que las diferencias que se encuentran entre dos o más grupos se atribuyen a la existencia o no de la o las variables independientes. Se dijo también que se parte de la base de que los grupos son semejantes en un gran número de variables relevantes o aspectos posibles, menos en la variable o variables independientes, de tal manera que éstas son lo único que los diferencia.

El aseguramiento de esta semejanza previa, de los grupos a comparar, constituye el problema del control de las variables:

#### **MÉTODOS DE CONTROL DE LAS VARIABLES EXTERNAS O EXTRAÑAS:**

**Eliminación.** Cuando las variables pueden tener valor de 0.

**Constancia de Condiciones**

**Balanceo y Contra balanceo**

**Aleatorización o Randomización.** Igualación de grupos con respecto de variables extrañas sistemáticas y no sistemáticas.

**Variables extrañas no sistemáticas :** son variaciones de los sujetos (genéticas , de aprendizaje, motivacionales, etc.). La planeación sistemática del procedimiento experimental reduce el efecto de las variables extrañas no sistemáticas. Este tipo de variables reduce la confiabilidad y si no hay confiabilidad no hay validez.

Reiterando una vez más, estos procedimientos de control se utilizan para igualar hasta donde sea posible, a los grupos en cuanto a la presencia de variables extrañas; éstas son aquellas que se relacionan de una manera u otra con la variable dependiente en estudio, pero que de momento no se puede controlar en términos del primer sentido.

### 3.3.5.1. Eliminación.

Cuando existe una variable extraña que se asume (en base a teorías, a la lógica, a la experiencia, a las fuentes de información especializadas, etc.) que está relacionada, y por lo tanto puede afectar a la variable dependiente que se investiga, influyendo de manera diferencial en los grupos que se comparan, entonces debe procederse a eliminarla de ambos grupos para igualarlos. Un ejemplo muy simple: supóngase que se está investigando cuál de tres productos nuevos (galletas) tiene mayor aceptación (gusta más) entre los niños, con el propósito de introducirlo al mercado. Se forman varios grupos de niños (con las características socioeconómicas que requieren los futuros consumidores de dicho producto) de escuelas diferentes. Una variable extraña que podría influir con *clase de galletas y mayor o menor gusto por ellas* podría ser la influencia de los comentarios entre los niños: si uno dice: "¡qué rica!" otros podrían opinar así, sin que en verdad les gustara o viceversa. El investigador decide eliminar la variable *presión social*, recogiendo los datos individualmente y no en grupos.

### 3.3.5.2. Constancia de Condiciones.

Este tipo de control opera en sentido opuesto al anterior. En muchas ocasiones no se pueden eliminar estímulos, variables, entonces lo que se hace es mantenerlas constantes en todos los Sujetos o grupos. Tomando el mismo ejemplo, supóngase que el investigador considera que la variable extraña que puede intervenir en la relación que se está estudiando (clase de galleta-mayor o menor gusto por ella) es la variable *tiempo de privación de comida*. El razonamiento es que no existe la misma necesidad de ingestión de comida cuando los niños llegan a la escuela (acaban de desayunar), después del recreo (comieron golosinas o su **lunch**) o un poco antes de irse a sus casas. El mayor o menor gusto para unas de las galletas puede estar influido por la mayor o menor hambre del niño; como el investigador no puede eliminar la variable *tiempo de privación de comida*, decide que ésta se mantenga constante para todos los grupos de las diferentes escuelas. Escoge una determinada hora de entre las que los niños permanecen en la escuela y a esa misma hora se lleva a cabo el experimento en todos los grupos, ya sea en el mismo día o a lo largo de varios días.

### 3.3.5.3. Balanceo.



Supóngase ahora, que las personas que realizarán una encuesta sobre el uso de anticonceptivos en una comunidad, son de diferente sexo, y se sabe por otros estudios realizados que las mujeres son las más adecuadas para recoger este tipo de información. Sin embargo, no se cuenta con un número suficiente de encuestadoras. Si se investigan las actitudes hacia el uso de los anticonceptivos en dos grupos diferentes de dicha comunidad y los encuestadores entrevistarán a uno de los grupos y las encuestadoras al otro, se obtendrían resultados diferentes que no necesariamente se deberían a las diferencias de actitud entre los grupos, o por lo menos no se podrían atribuir a éstas porque dichas diferencias podrían ser producto del efecto que provocó en las respuestas el sexo del encuestador.

Este problema podría resolverse mediante el balanceo, que consiste en este caso específico, en que las mujeres encuestaran al 50% de cada grupo, mientras que los hombres encuestarían al otro 50% de esos grupos. De esta manera, el efecto diferencial que puede producir el sexo del encuestador, en cierto modo se anula o compensa al hacer intervenir dicho efecto de manera balanceada.

#### **3.3.5.4. Contrabalanceo.**

En algunas ocasiones, las investigaciones requieren que los mismos Sujetos respondan a varias tareas experimentales diferentes. Esto tiende a producir dos problemas:

- a) Efecto de fatiga por la repetida presentación del estímulo
- b) Efecto de aprendizaje por la repetida ejecución de la tarea que exige el estímulo.

A estas condiciones se les conoce como *acarreo de efectos* y como es de esperarse, si éste no se controla, las diferencias que pueden encontrarse entre las respuestas de los Sujetos sometidos a múltiples tareas experimentales no pueden atribuirse a los efectos de los estímulos o variables independientes que se están investigando, porque en las respuestas pueden estar influyendo los efectos de fatiga y aprendizaje.

Para resolver este problema se debe contemplar en el diseño de investigación el método de control denominado contrabalanceo. Esto quiere decir, que si los grupos de Sujetos tendrán que exponerse a más de una tarea experimental, cada grupo se deberá subdividir en tantos subgrupos según el número de combinaciones posibles pueda hacerse con los estímulos o variables independientes. De esta manera, cada subgrupo se expone a cada combinación de estímulos de manera tal que todos y cada uno de ellos sea presentado el mismo

número de veces en los diferentes órdenes posibles de aparición dentro de las combinaciones posibles; en otras palabras, esto significa que los Sujetos se enfrentarán a diferentes órdenes de presentación de esos estímulos. Este procedimiento anula, compensa o por lo menos disminuye la influencia en las respuestas (la variable dependiente) del acarreo de efectos entre los que se encuentran la fatiga y el aprendizaje. El ejemplo de la investigación de productos comestibles (el de las galletas) puede ser útil para explicar de manera muy sencilla el procedimiento de contrabalanceo. Supóngase que el investigador se enfrenta a limitaciones tales como la imposibilidad de reclutar a un gran número de Sujetos; el tiempo para entregar resultados es mínimo; dispone de un pobre presupuesto. Estas limitaciones le impiden probar cada producto en diferentes grupos. Debido a esto decide investigar cuál de los tres tipos de galletas tiene mayor aceptación en un sólo grupo.

Como tiene tres clases de galletas (A, B y C) procede a obtener el número de posibles combinaciones que pueden hacerse con ellas: A B C; B C A y C A B. En seguida subdivide al grupo en tres subgrupos, y los Sujetos de cada uno de éstos probarán las galletas en el orden indicado por las combinaciones. El investigador asigna al azar cada subgrupo a una de las combinaciones de los estímulos (tipos de galletas). Así, a los Sujetos que les tocó por azar la combinación A B C, probarán primero la galleta A, luego la B y finalmente la C; los Sujetos de los otros subgrupos probarán las galletas según el orden de la combinación que les tocó.

#### **3.3.5.5. Aleatorización.**

Este método de control de variables es el más empleado en el campo de la experimentación en las ciencias sociales. Parte de la premisa de que si la selección de los Sujetos en los diferentes grupos a contrastar o comparar se realiza al azar, siempre y cuando el número de Sujetos por grupo sea lo suficientemente grande, las posibles diferencias en cuanto a las variables extrañas se repartirán al azar, de la misma manera en todos los grupos, quedando éstos igualados mediante dicho procedimiento. El proceso a seguir es muy simple: se saca al azar una muestra de la población que interesa estudiar. Los Sujetos de esta muestra también se asignan al azar para formar los grupos control y experimentales asumiéndose así, que los grupos han quedado igualados. Aunque este procedimiento de control es el más sencillo y el más aceptable, no siempre es posible, por diversas razones prácticas, recurrir a él. Sus principales ventajas son que permite el uso de las pruebas estadísticas con mayor poder y su capacidad de generalización de los resultados.

Por otra parte, la sencillez de este procedimiento reside principalmente en que el investigador no tiene que preocuparse por las variables extrañas que pueden interferir en la relación que estudia, pues la selección de los Sujetos al azar cancela o neutraliza, a la larga, los efectos de dichas variables. La mayor dificultad, por otro lado, se relaciona fundamentalmente con lo siguiente: se debe conocer la población de la que se extraerá la muestra. En un caso tan sencillo en el que la población fuera los estudiantes de la Facultad de Psicología de la UNAM, el investigador debe tener a la mano el total de los nombres de dichos estudiantes. Después de determinar cuál será el tamaño de la muestra, sacará al azar  $x$  número de Sujetos (sus nombres). Para obtener los datos de su investigación deberá localizar a cada Sujeto y sólo a esos Sujetos que por azar resultaron parte de la muestra (a menos que se establezca desde el principio el tipo de substitución que se llevará a cabo en caso de no localizarlos; para mayores detalles consúltese Kish, 1975). Esto obedece al principio probabilístico de este procedimiento que dice: todos los Sujetos (de una población determinada) deberán tener la misma probabilidad de ser parte de la muestra. Como esto se cumple con la selección al azar, los que resultaron elegidos no pueden ser cambiados por otros, pues se perdería la validez fundamentada por dicho principio. La mayoría de las veces uno de los principales problemas del control por aleatorización es la localización de los Sujetos que resultaron formar parte de la muestra. Extrapolando este ejemplo a otra población, por decir algo, "usuarios de los servicios públicos del Sector Salud", puede pensarse en las dificultades que enfrenta este procedimiento. Existen sin embargo, otras alternativas que se verán en el apartado dedicado a Muestreo. Además, debe señalarse que este procedimiento es más utilizado en investigaciones de carácter experimental, en donde por lo general, las poblaciones que se utilizan son de menor complejidad.

### **3.3.6. Diseño de Investigación.**

El diseño de investigación se refiere al hecho de tomar una decisión al respecto del número de grupos a utilizar, el número de Sujetos dentro de cada grupo, y la asignación o selección de los grupos en cuánto a cuál o cuáles serán los experimentales y cuáles el o los controles. La selección del diseño es muy importante, pues de éste depende en gran medida la validez de los resultados. Los resultados pueden ser válidos desde dos puntos de vista: validez interna y externa de la investigación.

La primera se refiere al grado con el que se logra obtener efectos diferenciales de la variable independiente (como se puede ver esto

se relaciona con la capacidad de ejercer el control en sus dos significados) en la dependiente, conducta o respuestas de los grupos investigados, y por supuesto que dichos resultados permitan aceptar o rechazar con la menor probabilidad de error la hipótesis nula. La segunda se refiere a la posibilidad de generalizar los resultados obtenidos con la muestra o muestras con las que se trabajó en la investigación, a la o las poblaciones de donde proceden aquellas. Es decir, qué tanto y hasta dónde es posible extrapolar los resultados obtenidos con las muestras.

Existen diversos tipos de diseños, cada uno es más o menos adecuado según el problema y objetivos de la investigación. En la selección del diseño, también deben tomarse en cuenta, aspectos de carácter práctico: el tiempo con el que se cuenta para hacer el estudio; el personal del que se dispone; el presupuesto que se tiene; el número de Sujetos de investigación disponibles. Otro aspecto determinante en la selección del diseño es el número de variables, independientes y dependientes que se van a investigar.

Los diseños más utilizados en la investigación de las ciencias sociales reciben los siguientes nombres genéricos: diseño de una sola muestra; diseño de dos muestras independientes; diseño de dos muestras relacionadas; diseño de más de dos muestras independientes; diseños de más de dos muestras relacionadas; diseños factoriales. (Cada uno de estos diseños se tratarán ampliamente en el siguiente capítulo).

### **3.3.7. Muestreo.**

Se ha hablado de poblaciones y muestras; se revisará brevemente cada uno de estos conceptos. Una población se define como la totalidad de eventos existentes que tienen en común una o más características. También se les conoce con el nombre de universos. Una muestra es una porción de una población o universo que lo representa en una o algunas de sus características de manera más o menos adecuada. Es decir, la muestra presenta características idénticas, o por lo menos no lo suficientemente distintas de la población, como para considerarla como representativa de la misma. El investigador por lo general no puede, en gran medida por razones prácticas, trabajar con las poblaciones o universos en su totalidad; de aquí la importancia de las muestras y del desarrollo de diversos procedimientos de muestreo.

De esta manera, utilizando muestras adecuadas, el investigador puede estudiar un reducido número de Sujetos y generalizar, considerando las limitaciones del estudio, a la población que representa esa muestra (Warwick y Lininger, 1975).

En términos generales existen dos tipos de procedimientos de selección de muestras: muestras probabilísticas y muestras no probabilísticas.

### **3.3.7.1. Muestras Probabilísticas.**

La característica esencial del tipo de muestreo probabilístico es que se puede especificar para cada elemento de la población la probabilidad que tiene de quedar incluido en la muestra, que es la misma para todos y cada uno de ellos. En cambio en el muestreo de no probabilidad no existe la posibilidad de hacer esa determinación como tampoco la seguridad de que cada elemento tenga alguna probabilidad de ser incluido. El muestreo de probabilidad es el único que hace posible los planes de muestreo representativo. Permite que el investigador investigue el grado con el que los resultados basados en su muestra tienden a diferir de los que habría obtenido de haber estudiado toda la población. Con este tipo de muestreo puede especificar el tamaño de la muestra (o bien los tamaños de varios componentes de muestras complejas) que precisará, el grado de certeza con el que desea que los resultados de su muestra no difieran más allá de un determinado límite que se habría logrado si se hubiera estudiado a la población total. Algunos de los procedimientos que forman parte del muestreo probabilístico, son los siguientes:

#### **3.3.7.1.1. Muestreo al Azar Simple.**

Este tipo de muestra se selecciona mediante un proceso que no solamente da a cada elemento una oportunidad igual de ser incluido en la muestra, sino que también hace la selección de cualquier combinación posible del número deseado de casos igualmente semejantes. El procedimiento general es el de asignarle un número a todos y cada uno de los casos de la población en cuestión. Después se toma una Tabla de Números Aleatorios, y al azar se toma un punto determinado de ésta, y los casos cuyos números están por encima o por abajo; o a la derecha o a la izquierda (según se quiera) del número señalado, se toman como muestra hasta llegar a la totalidad de casos deseados.

#### **3.3.7.1.2. Muestra Estratificada al Azar.**

En este caso, la población se divide primeramente en dos o más estratos. Estos estratos pueden estar basados en un criterio único, o en una combinación de dos o más criterios. En el muestreo estratificado lo que se toma de cada estrato es una muestra al azar simple, y entonces se unen las submuestras para formar la muestra total.

### **3.3.7.1.3. Muestreo en Racimo.**

En el muestreo en racimo se llega a la última serie de elementos susceptibles de ser incluidos en la muestra mediante la obtención, en primer lugar, de las agrupaciones mayores (*racimos*). Los racimos son seleccionados por métodos de muestreo simple o estratificado por azar; y si no van a ser incluidos en la muestra todos los elementos de estos racimos, la selección última a partir de los racimos también se lleva a cabo con cualquiera de los dos procedimientos, de muestreo simple o estratificado por azar. Por ejemplo, en un estudio de familias urbanas se puede tomar una muestra de distritos (*racimos*), y dentro de cada distrito seleccionado, una muestra de familias. Si el muestreo es llevado a cabo en una serie de etapas es posible, desde luego, la combinación de muestreo de probabilidad y de no probabilidad en un solo esquema. Es decir, una o más de las etapas pueden ser llevadas a cabo de acuerdo con los principios de probabilidad, y el resto mediante principios de no probabilidad.

### **3.3.7.2. Muestras No Probabilísticas.**

A continuación se señalan algunos de los procedimientos a los que se puede recurrir en el muestreo de no probabilidad:

#### **3.3.7.2.1. Muestras Accidentales.**

En una muestra accidental simplemente se toman los casos que vienen a la mano, continuando el proceso hasta que la muestra adquiera el tamaño precisado.

#### **3.3.7.2.2. Muestreo por Cuota.**

Con este tipo de muestreo se intenta incluir a los diversos estratos de la población en la muestra, tomando en cuenta las proporciones con las que éstos ocurren en la población. La condición básica en el muestreo por cuota no es que los distintos estratos de la población sean representados por muestras en sus proporciones correctas, sino más bien que haya el suficiente número de casos de cada estrato para hacer posible una estimación del valor del estrato de la población, y se pueda conocer la proporción con la que se constituye en la población cada uno de tales estratos.

#### **3.3.7.2.3. Muestras Intencionadas.**

La hipótesis básica detrás del muestreo intencionado es que con buen juicio y una estrategia adecuada se pueden seleccionar fácilmente los casos a ser incluidos en la muestra y por lo tanto

desarrollar muestras que sean satisfactorias en relación con las propias necesidades. Una de las aproximaciones más generales consiste en tomar casos que se juzgan como típicos de la población en la que se está interesado, suponiendo que los errores de juicio en la selección tenderán a compensarse entre sí.

### **3.3.7.3. Tamaño de la Muestra.**

Existen diversos y variados procedimientos para establecer el tamaño adecuado de las muestras, desde los más simples y genéricos, hasta los más sofisticados como los utilizados en las encuestas por muestreo, pasando por técnicas funcionales y de dificultad mínima, como son las que utilizan tablas estandarizadas para tales propósitos. Entre los primeros se tienen aquellos en los que por consenso se acepta que si se tienen datos derivados de una investigación experimental (con control de varianza externa) una muestra con un tamaño mínimo de  $N = 30$  sería suficiente; mientras que si ese no fuera el caso ( estudio no experimental) entonces se requeriría por lo menos una  $N = 200$ . En lo que sigue veremos ejemplos de procedimientos más sofisticados, sin embargo, queremos subrayar siguiendo a Kraemer y Thiemann.(1987) que en el proyecto de investigación y de acuerdo con el problema al que se intenta dar solución, se decide el tipo de diseño, la observación y medición de las respuestas, el número y tiempos de mediciones por sujeto, tipo de análisis y pruebas que se aplicarán y número de sujetos que conformarán la muestra; todas estas decisiones determinarán que tan fuerte será la evidencia o el poder de la decisión final en torno a la respuesta que se dará al problema de investigación.

Ahora bien, siguiendo las recomendaciones de Goode y Hatt (1970) y de Parten (1950), procederemos a establecer otros criterios y requisitos para establecer el tamaño adecuado de las muestras que se usan en la investigación.

Como ya se señaló en el desarrollo del plan de una investigación se debe tomar una decisión respecto al número apropiado de casos que incluir en la muestra. Todas las fases de la investigación se verán afectadas por el tamaño de la muestra. El tamaño se puede estimar con base en la distribución probable de los datos.

Existen algunas ideas erróneas respecto al tamaño de las muestras. Entre ellas se tiene aquella que se refiere a que un número muy grande de casos aseguran resultados correctos. Otra tiene que ver con la idea de que las tendencias prejuiciadas se eliminan o corrigen simplemente incrementando el tamaño de la muestra. Otra idea errónea es aquella que dice que una muestra del 10% del tamaño de la población es la que se debe tener.

Las ideas anteriores se basan erróneamente, en que es el tamaño del universo de donde se saca la muestra la que determina su tamaño. En realidad, el énfasis se debe poner, no en el tamaño de universo, sino en el diseño de la muestra.

La muestra óptima de una investigación es aquella que reúne los requisitos de eficiencia, representatividad, confiabilidad y flexibilidad. La muestra mas eficiente es aquella que proporciona la máxima información por peso invertido, mas que por caso. Debe ser lo suficientemente pequeña para evitar gastos innecesarios, pero lo suficientemente grande para evitar errores de muestreo intolerables. Debe ser lo suficientemente grande como para producir resultados representativos y estadísticamente significativos en todas las tabulaciones de importancia que se requieran: pero no tan grande como para gastar dinero, retrasar el proyecto y alcanzar innecesaria precisión.

Los factores que afectan el tamaño de la muestra son varios. En primer lugar, la homogeneidad o heterogeneidad de la población o universo; entre mas homogéneo el universo, menor el tamaño que se requiere de la muestra, a mayor heterogeneidad, mayor deberá ser el tamaño. En segundo lugar, se tiene que tomar en cuenta el número de categorías y clases en los que se van a analizar y agrupar los resultados; entre mayor sea el numero de subgrupos universos menores, se requiere una muestra total mas grande para garantizar mediciones estadísticas confiables. En tercer lugar se tienen que tomar en cuenta los problemas del levantamiento, de modo que se pueda levantar la información con los fondos y el tiempo con que se cuenta. Es decir, que el tamaño de la muestra también se ve determinado por el tamaño de la encuesta, el número de encuestadores, la especificidad de las personas que habrán de ser entrevistadas, y el grado en el que se concentran o dispersan los casos en una zona, la tasa de rechazo, la pérdida de casos, el tipo de muestreo empleado, y el método de recolección de datos (por correo, teléfono, entrevista personal, etc.).

Cuando se emplean muestras por cuota con especificaciones o criterios de inclusión muy específicos, se tienen muestras pequeñas, al ser mas difícil localizar a los sujetos.

En cuarto lugar, se tienen al tamaño y tipo de muestreo. Ya se había mencionado que se requiere de una muestra menor para que sea confiable si la población es homogénea. Por lo tanto, se espera que las muestras sean mas pequeñas y suficientes, si se utiliza el muestreo estratificado al azar en comparación con el aleatorio simple. Esto se debe a que el efecto de la estratificación resuelve el problema de la relativa heterogeneidad total en un número de submuestras individualmente homogéneas. Entre mas



heterogéneo sea el universo, se pueden ahorrar un número mayor de casos por medio de la estratificación.

### **3.3.7.3.1. Cálculo del tamaño de la muestra partiendo de porcentajes estimados.**

Los factores anteriormente señalados: el grado de homogeneidad, el número de tabulaciones, y el tipo de muestreo afectan todos el número de casos que se deben tener para lograr la precisión que se desea. La cuestión práctica y concreta entonces, es la de determinar el tamaño de la muestra que se requiere para garantizar la precisión necesaria para la investigación. Las estimaciones que se hacen a partir de la media y los porcentajes son los mas confiables que se pueden hacer, y por lo tanto son los valores que se consideran en la estimación de los tamaños óptimos de las muestras precisas.

Para calcular el tamaño de la muestra el investigador debe hacer dos cosas: en primer lugar, obtener información preliminar de la población que pretende muestrear; y en segundo lugar, debe decidir el nivel de precisión que se desea tener. Esta decisión depende de: a) que tanto error se permitirá, y b) que tanta seguridad se requiere para que la estimación caiga dentro del error permisible.

El error permisible se refiere a que tanta desviación se permite entre la estimación del parámetro (valor en la población) y el de la muestra, para efectos de la investigación.

Asegurar que el valor estimado caiga dentro del error permitido es la segunda cuestión a determinar, y como esto depende de las fluctuaciones del muestreo, la medida adecuada que nos señala estas fluctuaciones es el error estándar. Es la medida de la variación de las estimaciones que se esperan en un muestreo aleatorio. Las muestras elegidas al azar tienden a seguir las leyes del azar, y las estimaciones de la muestra tienden a aglutinarse alrededor de los valores reales del universo o población, en forma de una campana, como en el caso de la curva normal. El error estándar no es igual al rango completo de la posible variación de la muestra, sino aproximadamente de tan sólo una sexta parte de ella. Y como las variaciones extremas ocurren muy rara vez, para propósitos prácticos, si se toman en cuenta dos errores estándar, en ambas direcciones alrededor del punto medio del rango, quedan incluidas la mayoría de las fluctuaciones muestrales. Esto significa que se puede esperar que el 95% de las estimaciones muestrales caigan dentro de los límites de más o menos dos errores estándar.

Por lo tanto, cuando lo que se va a estimar son porcentajes, se parte de la siguiente fórmula:

$$\sigma_{p.c.} = \sqrt{\frac{p.c.(100 - p.c.)}{n}}$$

dónde: p.c. = porcentaje

n = tamaño de la muestra

$\sigma_{p.c.}$  = error estándar del porcentaje

Esta fórmula se empleó para construir Tablas como las que aparecen en Parten (1949, pag.309).

Habiendo considerado el error permitido o la tolerancia, y las probabilidades de permanecer dentro del límite de tolerancia, se puede proceder al cálculo del tamaño de la muestra que permita cumplir con los requisitos previamente señalados.

La fórmula que se emplea para calcular el tamaño de la muestra, (que es una variación transpuesta de la fórmula del error estándar) habiendo tomado la decisión del error de muestreo que se desea tener y el intervalo de confianza de que la estimación caiga dentro de dos unidades de error estándar, es la siguiente:

$$n_s = \frac{p.c. (100 - p.c.) z^2}{T^2} \quad 0$$

dónde: p.c. = estimación preliminar del porcentaje

z = número de unidades de error estándar(2)

T = precisión o tolerancia requerida

Cuando es suficiente con estar moderadamente seguro (95%) de que la estimación caiga dentro de los límites prescritos, se pueden usar Tablas como la que aparece en Parte (1949) en las páginas 314-315.

### **3.3.7.3.2. Cálculo del tamaño de la muestra tomando como base medias estimadas.**

Otro caso importante para calcular el tamaño de la muestra es el caso de la estimación de la media. El problema ahora es calcular el tamaño de la muestra necesario para predecir la media del universo o población dentro de un cierto rango de error con una probabilidad dada. El investigador requiere de tres datos para calcular el tamaño necesario de la muestra: a) la precisión o error permisible (T); b) la probabilidad requerida (P) de caer dentro del rango de error (95%); y c) una estimación preliminar de la desviación estándar del universo ( $\sigma$ ). El tercer requisito demanda tener información del universo o población, que puede

obtenerse de experiencia pasada con la población, de un estudio piloto o de alguna otra fuente de información.

La fórmula para calcular el tamaño de la muestra es la siguiente:

$$n_s = \left( \frac{\sigma z}{T} \right)^2$$

dónde:  $n_s$  = tamaño de la muestra  
 $\sigma$  = estimación preliminar de la desviación estándar del universo o población  
 $T$  = error permisible o tolerancia

De esta manera, si se supone una desviación estándar en la población de digamos \$500.00 M.N. en ingreso, y se desea tener una tolerancia posible de  $\pm$  \$100.00 M.N., con un intervalo de confianza del 95%, (1.96 unidades de desviación estándar), se tendría, substituyendo en la fórmula, que requeriría de una muestra del tamaño:

$$n_s = (500 \times 1.96 / 100)^2 = 86$$

Si se quiere una muestra con un valor estimado de la media que tuviera mayores probabilidades de caer dentro del intervalo de confianza (99%), se coloca el valor de 2.57 en lugar de 1.96, y se obtendría un total de 165 Sujetos.

### **3.3.7.3.3. Tamaño de la muestra tomando como base el valor de la comparación entre medias y su nivel de significancia.**

Otra manera de calcular los tamaños idóneos de las muestras de investigación parten del supuesto de comparar dos medias de la muestra, que esperemos sean lo suficientemente diferentes como para arrojar un valor de  $t$  de Student, lo suficientemente grande como para que sea estadísticamente significativo al 0.05, para asegurar que las medias de la muestra caen dentro de los intervalos de confianza del 95%.

En este caso, la fórmula de la que se parte es la siguiente:

$$n = \frac{t^2 \times s^2}{E^2 \div \frac{t^2 \times s^2}{N}}$$

dónde:  $t$  = valor de la  $t$  de Student (1.96)  
 $s$  = desviación estándar  
 $E$  = error de muestreo  
 $N$  = tamaño de la población  
 $n$  = tamaño de la muestra

Empleando la fórmula anterior se han elaborado tablas como la Tabla 3.1.

Tabla 3.1. Tabla de Muestreo para valores de  $t$  de 1.96 con un nivel de significancia al 0.05.

## ERRORES DE MUESTREO

POBLACIÓN	0.01	0.02	0.03	0.04	0.05
150	147	141	131	120	107
200	195	184	168	150	131
250	243	226	202	176	151
300	290	266	234	200	168
350	337	305	263	221	183
400	384	342	290	240	195
450	429	378	316	257	207
500	475	413	340	272	217
550	520	447	362	287	226
600	564	480	384	300	234
650	608	511	403	312	241
700	652	541	422	323	248
750	695	571	440	333	254
800	738	600	457	342	259
850	780	627	473	351	264
900	822	654	488	360	269
950	864	680	502	367	273
1000	905	705	516	375	277
1500	1297	923	623	428	305
2000	1655	1091	695	461	322
2500	1983	1224	747	484	332
3000	2285	1333	787	500	340
3500	2565	1424	817	512	346
4000	2823	1500	842	521	350
4500	3064	1565	862	529	353
5000	3288	1622	879	535	356
5500	3497	1671	893	541	359
6000	3692	1714	905	545	361
6500	3876	1753	916	549	362
7000	4048	1787	925	552	364
7500	4211	1818	934	555	365
8000	4364	1846	941	558	366
8500	4509	1872	948	560	367
9000	4646	1895	953	562	368
9500	4775	1916	959	564	369
10000	4899	1936	964	566	369

### 3.3.7.3.4. Tamaño de la muestra dependiendo del nivel del efecto esperado de los tratamientos experimentales.

En este caso se toma en consideración para determinar el tamaño de la muestra por un lado, al número de tratamientos o condiciones, por otro, el tamaño esperado del efecto de los tratamientos (moderado o fuerte), y por último, el grado de control de las variables involucradas (Feldt, 1973).

De ello surge la Tabla 3.2, que se presenta a continuación.

Tabla 3.2 Tamaño mínimo de la muestra (por grupo de tratamiento) para experimentos de métodos/materiales.

NÚMERO DE TRATAMIENTOS O CONDICIONES	TRATAMIENTOS MODERADAMENTE EFECTIVOS		
	SIN CONTROL DE VARIABLES	REGULAR CONTROL DE VARIABLES	BUEN CONTROL DE VARIABLES
2 TRATAMIENTOS	235	200	135
3 TRATAMIENTOS	305	265	175
4 TRATAMIENTOS	325	280	190
	TRATAMIENTOS MUY EFECTIVOS		
2 TRATAMIENTOS	60	50	35
3 TRATAMIENTOS	80	65	45
4 TRATAMIENTOS	85	70	50

### 3.3.7.3.5. Tamaño exacto de la muestra conociendo la desviación estándar de la población.

Otro procedimiento relacionado con los anteriores utiliza una fórmula para calcular el tamaño exacto de la muestra, conociendo la desviación estándar de la población, la discrepancia o tolerancia entre la estimación real de la media y el valor del parámetro en la población, así como el intervalo de confianza, que en este caso se plantea al 95%.

La fórmula empleada es la siguiente:

$$\sqrt{N} = \frac{1.96 (\sigma)}{T}$$

dónde:

$N$  = el tamaño de la muestra

1.96 = el 95% de intervalo de confianza de que la media de la muestra caiga dentro del error estándar de la media de la población.

$\sigma$  = desviación estándar estimada o conocida de un parámetro de la población.

$T$  = tolerancia de la discrepancia entre el valor de la media de la muestra y la de la población aceptado por el investigador.

Substituyendo los valores adecuados dentro de esta fórmula, se puede tener información sobre el número exacto de sujetos que habrán de formar parte de la muestra de investigación de manera que se tenga un error de muestreo del 5% y la seguridad de que la media de la muestra, caerá dentro de los límites del error estándar de la media de la población (Blalock, 1972).

## REFERENCIAS

Blalock, H.M. (1972): **Estadística Social**. México: Fondo de Cultura Económica.

Feldt, L.S. (1973): What size samples for methods/materials experiments. **Journal of Educational Measurement**. Vol. 10, No. 3, pp. 221-226.

Goode, W.J. y Hatt, P.K. (1970): **Métodos de Investigación Social**. México: Editorial Trillas.

Kraemer, H. CH., y Thiemann, S. (1987). **How many subjects? Statistical power analysis in research**. Newbury Park California: SAGE Publications.

Kish, L. (1975): **Muestreo de Encuestas**. México: Editorial Trillas.

McGuigan F.J. (1993): **Experimental Psychology: Methods of Research**. Sixth Edition. Englewood Cliffs, N.J.: Prentice Hall.

Parten, M. (1949): **Surveys, Polls and Samples**. Nueva York: Harper and Row.

Rosenthal, R., y Rosnow, R.L. (1991). **Essentials of behavioral research. Methods and data analysis**. Nueva York: McGraw-Hill

Siegel, S. (1976): **Estadística No Paramétrica**. México: Editorial Trillas.

Warwick, D.R. y Livingier, Ch. A. (1975): **The Sample Survey: Theory and Practice**. Nueva York: McGraw Hill Book Company.

## **IV. DISEÑOS DE INVESTIGACIÓN, PROCEDIMIENTOS**

### **ESTADÍSTICOS NO PARAMÉTRICOS**

#### **4.1. Introducción.**

La investigación es un conjunto de acciones en proceso, en el que interactúan y se interrelacionan todos y cada uno de los momentos que la componen. Debe entenderse que sólo por razones didácticas y de exposición se separan las acciones y razonamientos involucrados en la misma.

En la selección del diseño de la investigación se prevé al mismo tiempo, tipo de muestras, técnicas e instrumentos para la recolección de la información, procedimientos estadísticos para el tratamiento de los datos y la prueba de hipótesis. Todo esto está, por supuesto, en función del problema de investigación, de las variables que se manejarán, y aún de consideraciones de otro orden, como son los recursos y tiempo del que se dispone.

La decisión de recurrir o no, a la aplicación de procedimientos estadísticos, así como al tipo de estadística que se utilizará está condicionada por el carácter de las variables, por el nivel de medición de las mismas, por el procedimiento de selección de muestras, en suma, por el diseño de investigación. Debe señalarse sin embargo, que el uso de la estadística no sólo está condicionado por el diseño, sino que la estadística utilizada, a su vez condiciona, en tanto su correcta o incorrecta utilización, optimiza o invalida su producto final, sus resultados, y con ello, el total de la investigación. El mejor diseño se desmorona ante un uso inadecuado de la estadística; esto es, si los datos que producen no cubren los requisitos que ésta impone.

Una de las divisiones que se hace de los procedimientos estadísticos se basa en los supuestos que parten del conocimiento teórico y/o práctico de los parámetros (especialmente de las medias y desviaciones estándar) de las poblaciones de las que provienen las muestras de investigación. De esta manera, se habla de estadística paramétrica y no-paramétrica, contando cada una de ellas con diversos y diferentes procedimientos. El uso adecuado de estas clases de estadística depende de que los datos en los que se aplican se ajusten a los supuestos de los que aquellas parten.

#### **4.2. Estadística Paramétrica.**



Los supuestos de la estadística paramétrica imponen mayores exigencias a los datos, pero también se considera la de mayor poder para la prueba de hipótesis. Dichos supuestos son los siguientes:

*-Distribuciones Normales.* Los valores de la variable que se estudia deben distribuirse en la población en forma normal (distribución de campana o de Gauss). En otros términos, pocos Sujetos tienen mucho o muy poco de la variable, estos forman las colas de la campana, mientras que la mayoría posee cantidades intermedias. Estatura, edad e inteligencia se distribuyen normalmente; el ingreso económico, por ejemplo, no se distribuye en forma normal. La literatura especializada, los centros de investigación, los departamentos de censo, constituyen algunas de las fuentes de información al respecto. Sin embargo, para una gran mayoría de las variables que interesan al investigador de las disciplinas sociales no existe conocimiento de cómo se distribuyen en la población, ni posibilidades de obtenerlo, por las grandes dificultades de orden diverso, que ello implica. En estos casos al investigador no le queda otra cosa que asumir que si los valores de las variables de sus muestras se distribuyen normalmente, entonces éstas también se encuentran normalmente distribuidas en las poblaciones correspondientes.

*-Homocedasticidad u Homogeneidad de la Varianza.* Las poblaciones de donde se extrajeron las muestras que se van a comparar deben tener igual varianza (desviaciones estándar al cuadrado), desde luego en relación a las variables de interés.

*-Nivel de Medición.* Los valores de la variable dependiente se deben obtener (medir o registrar) por lo menos a nivel intervalar.

*-Observaciones Independientes.* Las observaciones independientes se refieren, como su nombre lo indica, a observaciones o mediciones no relacionadas, de tal manera que la obtención de una no afecta la de otra. En otros términos, puede decirse que una observación no es independiente cuando un sujeto es medido dos o más veces. Por ejemplo, un diseño con mediciones **antes y después** es uno con observaciones dependientes ya que la primera medición puede afectar de muy diversas maneras a la segunda. Una diferencia fundamental entre medidas independientes y dependientes es que en la primera se trabaja con la media muestral o de grupo, ya que cada sujeto tiene sólo una medición o puntaje, mientras que en la segunda, se trabaja con las medias de los sujetos, puesto que cada uno de ellos tiene por lo menos dos mediciones o puntajes. En el capítulo que siguen regresaremos a este y otros conceptos de la estadística paramétrica.

### **4.3. Estadística No Paramétrica.**

La utilización de esta clase de estadística representa una alternativa de análisis y tratamiento de los datos, cuando estos no cubren los requisitos que expresan los supuestos en los que se basa la estadística paramétrica. La estadística no paramétrica no hace suposición alguna en relación a:

- la forma de las distribuciones
- la homocedasticidad u homogeneidad de las varianzas
- la independencia de las observaciones

Por otra parte, el nivel de medición puede ser nominal u ordinal.

### **4.4. El Uso de la Estadística en la Investigación.**

Antes de iniciar la presentación de las pruebas estadísticas mas empleadas para someter a prueba las hipótesis planteadas por el investigador, es conveniente hacer una somera revisión de algunos conceptos estadísticos, que permitirán una mejor comprensión de lo que se presenta mas adelante.

La investigación que se lleva a cabo tiene por objetivo determinar la medida en que son acertadas las hipótesis que se plantean a partir de las diferentes teorías existentes referidas al tema de estudio en cuestión. Para ello, se recogen datos empíricos que señalarán la medida en que las hipótesis planteadas son probablemente ciertas, ya que el significado de los datos permiten confirmarlas, revisarlas o rechazarlas, apoyando o no la teoría de las que derivaron.

Para poder determinar si las hipótesis particulares se confirman o no, se requiere de un procedimiento objetivo que se base en la información obtenida en la investigación así como en el margen de riesgo que esté dispuesto a aceptar si el criterio de decisión respecto a la hipótesis resulta incorrecto. Por lo general este procedimiento consiste en diversos pasos, que se presentan a continuación.

#### **4.4.1. La Hipótesis Nula.**

La hipótesis nula es una hipótesis estadística, de no diferencias, que se pretende rechazar, para poder aceptar la alterna y con esto confirmar la hipótesis de trabajo.

#### **4.4.2. La Elección de la Prueba Estadística.**

La elección de las pruebas estadísticas depende del número de muestras con que se cuente (una, dos o  $k$  muestras), del procedimiento de elección de las mismas (relacionadas o independientes), del nivel de medición de la(s) variable(s) dependiente(s), y del cumplimiento de los supuestos que subyacen a las mismas. Se debe elegir la prueba mas adecuada e idónea para responder a la pregunta problema de investigación y poder así, determinar si la hipótesis de investigación se confirma o no.

#### **4.4.3. El Nivel de Significancia y el Tamaño de la Muestra.**

Una vez planteadas las hipótesis, se procede a especificar el nivel de significación y el tamaño de la muestra. Es decir, antes de recoger los datos, se especifica el conjunto de todas las posibles muestras que se encuentran si la hipótesis nula es verdadera; de este conjunto, se determina un subconjunto de muestras con características tan extremas de manera tal que la probabilidad que tienen de deberse al azar es sumamente pequeña. La(s) muestra(s) de investigación se comparan con ellas, y en la medida que su comportamiento también tenga pocas probabilidades de deberse al azar, se puede decir que el comportamiento que muestran se debe al efecto de los tratamientos experimentales o de las variables independientes consideradas en el estudio.

A esta pequeña probabilidad simbolizada por  $\alpha$  se le llama nivel de significancia. Los valores mas usados de  $\alpha$  son  $0=05$  y  $0=01$ . Esto significa que si los valores asociados a los valores particulares producidos por las pruebas estadísticas son iguales o menores a los señalados por  $\alpha$ , se deberá rechazar la hipótesis nula y aceptar la alterna y confirmar la de investigación (hipótesis de trabajo).

#### **4.4.4. La Distribución Muestral.**

La distribución muestral es una distribución teórica formada por todos los valores de las muestras posibles, del mismo tamaño, que se pueden sacar de una misma población. Es decir, es la distribución conforme a la hipótesis nula, de todos los valores posibles que una estadística (por ejemplo una  $t$ , o una media) puede tomar cuando es calculada con muestras del mismo tamaño que la(s) de investigación. Esta distribución señala las

probabilidades asociadas conforme a la hipótesis nula a los diferentes valores numéricos obtenidos por el estadístico. En todos los casos, nos interesa conocer la probabilidad asociada al valor del estadístico obtenido por nosotros en el proceso de investigación, para poder determinar que tan extrema es aquélla.

#### 4.4.5. La Región de Rechazo.

La región de rechazo es una región de la distribución muestral constituida por los valores del estadístico cuya probabilidad de deberse al azar es igual o menor a  $\alpha$ . Si el valor especificado fue de 0.05, la región de rechazo incluye todos los valores del estadístico que tengan asociada una probabilidad debida al azar  $\leq$  que 0.05.

#### 4.4.6. La Decisión Estadística.

Cuando el valor de la prueba o estadístico obtenido por el investigador tiene una probabilidad asociada  $\leq$  a  $\alpha$ , se toma la decisión de que la hipótesis nula es falsa y se acepta la hipótesis estadística alterna. Si ésta está planteada sin dirección (solo se dice que existen diferencias entre los grupos), la región de rechazo se localiza en las dos colas (los extremos de la curva de distribución muestral dado el tamaño de N); si la hipótesis plantea dirección (señala cual de los dos grupos obtendría puntajes más altos o mas bajos en la(s) variable(s) dependiente(s)), la región de rechazo se localiza en un solo lado de la curva de distribución muestral. En este caso la decisión depende de si los valores obtenidos corresponden o no con lo señalado por la hipótesis alterna.

#### 4.4.7 Errores que se Cometen en la Toma de Decisiones Estadísticas.

De acuerdo con Keppel (1973) cuando el investigador llega a la fase de la prueba estadística de sus resultados se encuentra frente a dos posibles situaciones: la  $H_0$  es falsa o verdadera y la decisión que tiene que hacer es rechazarla o aceptarla. la decisión ideal es que se rechace la  $H_0$  cuando ,sta, de hecho, sea falsa o la acepte cuando ,sta, de hecho, sea verdadera. Desafortunadamente dado que el investigador se encuentra en la situación de hacer decisiones frente a información incompleta (por ejemplo trabajar solo con muestras) se espera que en la toma de las mismas se cometa error. Puesto que tal error es inevitable, lo único que queda es tratar de minimizarlo. Los tipos de errores que se pueden cometer son: **el error tipo I** (la  $H_0$  es falsamente rechazada) y **el error tipo II** (la

$H_0$  es falsamente aceptada). El poder de una prueba está dado por la probabilidad de rechazar la  $H_0$ , cuando, de hecho deba de ser rechazada ( $P= 1-B$ ).

El error tipo I está representado por **alpha** y el error tipo II por **beta**. Dependiendo del problema de investigación, el experimentador selecciona los valores para alpha y beta. Existen situaciones en donde el investigador recurre a establecer el valor de alpha extremadamente pequeño sin importarle mayormente beta. Es decir, el experimentador ejerce un control completo sobre el error tipo I. Piénsese, por ejemplo, en aquellos estudios sobre fármacos, cirugía y en general problemas relacionados con la aplicación de la medicina. En estos casos el investigador disminuye al máximo la posibilidad de error tipo I. (Rechazar la  $H_0$  falsamente). En tales condiciones es más seguro dejar que aumente la probabilidad de cometer el error tipo II (aceptar  $H_0$  falsamente). De hecho, el investigador tiene un control indirecto sobre el error tipo II y lo ejerce a través, de la selección del tamaño adecuado de la muestra y controlando el tamaño de la varianza de error condición que se logra con diseños más precisos.

Puede verse entonces que el investigador es absolutamente libre de establecer el error tipo I, mientras que el error tipo II solo lo controla indirectamente, y es un control que siempre implica costos. Por otra parte, hay situaciones diferentes de experimentación en las que se acostumbra utilizar los valores convencionales de alpha ( $\leq .05$  y  $\leq .01$ ) que han demostrado equilibrar la relación inversa que existe entre ambas clases de error.

#### **4.5. Diseños de Investigación y Pruebas No Paramétricas**

##### **4.5.1. Diseño de una Sola Muestra**

Con este diseño, como su nombre lo indica, se trabaja con un grupo, con la muestra de Sujetos que el investigador selecciona de la población que le interesa. Sin embargo, para efectos de la prueba de sus hipótesis, ésta, la población, se trata como si fuera otro grupo para efectuar la contrastación. El objetivo de los estudios que se ajustan a este diseño es determinar qué tanto la muestra se parece a la población de donde se extrajo, con el fin de poder extrapolar los resultados obtenidos con la muestra a la población.

Sin embargo, debe tenerse en cuenta que la validez de estos estudios depende de la representatividad de la muestra. Este tipo de diseño se utiliza más frecuentemente en estudios de carácter

descriptivo y exploratorio. También se utiliza como un primer paso para determinar la representatividad de la muestra que se está investigando. En el caso en el que el interés principal es extrapolar los resultados a la población, determinación que se tomaría si la muestra resultara igual a la población, entonces lo que interesa al investigador es aceptar la hipótesis nula, que es la de las no diferencias. A esta aproximación se le conoce como la de "bondad de ajuste" (entre la muestra y la población).

Esto procede cuando los Sujetos de la muestra no se someten a algún estímulo, tratamiento o variable independiente; en otros términos, cuando no se trata de establecer el efecto de una variable independiente sobre la dependiente, pues de lo contrario, necesariamente se encontrarían diferencias entre muestra y población. Por el contrario, cuando a la muestra se le aplica un tratamiento (una variable independiente), en este caso se toma implícitamente a la población tal y como se da en su estado natural, como si fuera el grupo control y la muestra viene a ser el grupo experimental. Lo que al investigador le interesa es rechazar la hipótesis nula, si esto es posible; dicho rechazo significa que la muestra y la población, originalmente iguales con base en el procedimiento de selección de la primera, ya no lo siguen siendo en función del tratamiento especial dado a la muestra. Esto es, el rechazo de la hipótesis nula indica la efectividad del tratamiento.

Como se señaló anteriormente, las estadísticas paramétrica y no paramétrica cuentan respectivamente con diversos y diferentes procedimientos. Pasaremos a revisar algunos de estas últimas.

Para probar las hipótesis de estudios que se ajustan a diseños de una sola muestra, la prueba de  $\chi^2$  (chi cuadrada de una sola muestra) es una de las más frecuentemente utilizadas.

#### **4.5.1.1. Prueba No Paramétrica Para el Diseño de Una Sola Muestra: $\chi^2$ (Chi Cuadrada).**

Con gran frecuencia se requiere conocer la forma en que diversos objetos, Sujetos o respuestas se clasifican en diferentes categorías. Por ejemplo, saber si existen diferencias en las frecuencias de respuestas dadas "a favor" o "en contra", con respecto a algún objeto de actitud en un grupo determinado, en el que se supone que las hay, en virtud de alguna circunstancia particular. Para analizar este tipo de datos resulta de gran utilidad la prueba de la  $\chi^2$ . Por principio, los datos se arreglan en categorías de tal manera que estas sean exhaustivas y excluyentes; es decir, que cada respuesta, Sujeto u objeto, pueda ser ubicado en una categoría y que cada uno de ellos (respuestas,

objetos o Sujetos) sólo puedan ubicarse en una categoría. Esto es, que al pertenecer a una de ellas, automáticamente queda excluido de las  $k-1$  categorías restantes. La  $\chi^2$ , en este caso busca la "bondad de ajuste" para probar la existencia de una diferencia significativa entre un número *observado* de objetos, Sujetos o respuestas de cada categoría, y un número *esperado*, basado en la hipótesis de nulidad.

Para poder comparar un número de frecuencias observadas con un número de frecuencias esperadas, se requiere conocer cuál será el número de las frecuencias esperadas. La hipótesis nula establece cuál es el número que al azar se espera en cada categoría. Con dos categorías, por ejemplo, se espera que por azar caigan en cada categoría  $N/2$  frecuencias. Si las categorías fueran tres, se esperaría que por azar cayeran  $N/3$  frecuencias en cada una, y así sucesivamente.

Mientras más difieran las frecuencias observadas de las esperadas al azar, más se puede esperar que exista una diferencia real entre las frecuencias observadas. Es decir, si la forma en que se distribuyen las frecuencias observadas, no se parece a la forma en que se distribuirían al azar, esto significa que la distribución de las primeras no es aleatoria y que por lo tanto las diferencias se pueden explicar por alguna otra razón o causa. Esta otra razón o causa, que no se debe al azar, es por lo general, la variable independiente que el investigador está manejando, que al tener influencia sobre la dependiente, hace que esta última se distribuya en forma diversa a la que lo haría si sólo el azar estuviera operando sobre los datos.

La hipótesis nula, la que establece no diferencias estadísticamente significativas entre las frecuencias de cada una de las categorías, se puede poner a prueba mediante esta prueba.

**Ejemplo.** un investigador supone que en una muestra de estudiantes de una escuela religiosa habrá una frecuencia mayor de individuos en contra de la legalización del aborto en comparación con los que estarán a favor o indecisos, así como en comparación de lo que se esperaría en la población general.

**La Hipótesis Nula,** señalaría que no hay diferencia en el número esperado de los Sujetos en cada una de las categorías de respuesta hacia la deseabilidad de legalizar el aborto, y las diferencias observadas son meramente variaciones aleatorias esperadas en una muestra al azar de una población rectangular donde  $f_1 = f_2 = f_3 = 0$ .

**La Hipótesis Alternativa**, las  $f_1$ ,  $f_2$ ,  $f_3$  son diferentes.

**La Prueba Estadística.** Al comparar los datos de una muestra con alguna población, se puede utilizar una prueba de una muestra que compare frecuencias observadas con esperadas en categorías discretas, que implican un nivel de medición nominal (clasificadorio), como la chi cuadrada.

**El Nivel de Significancia.** Sea  $\alpha = 0.01$  y  $N = 120$ , el número de estudiantes de la muestra.

**Distribución Muestral.** La distribución muestral de  $\chi^2$  calculada con su fórmula sigue la distribución con  $gl = k - 1$  (grados de libertad).

**Región de Rechazo.** La hipótesis nula se rechazará si el valor observado de  $\chi^2$  tiene una probabilidad asociada con su ocurrencia conforme a  $H_0$  con  $gl = 2$ , igual o menor que  $\alpha = 0.01$ .

Si se obtuviera una  $\chi^2 = 60$  con  $gl = 2$  y con una probabilidad de ocurrencia al azar menor que  $p = 0.0001$ , entonces siendo esta probabilidad menor que el nivel de significancia establecido ( $\alpha = 0.01$ ), se rechaza la  $H_0$  y se acepta la alternativa, decisión que permitirá confirmar la hipótesis de trabajo.

**Restricciones estadísticas para uso de la Prueba  $\chi^2$ .** Cuando  $gl = 1$ , es decir, cuando  $k = 2$ , las frecuencias esperadas deberán adquirir por lo menos un valor de 5; cuando  $gl$  es mayor que 1, o sea  $k$  es mayor que 2, se puede emplear esta prueba para el caso de una muestra sólo si el 80% de las celdillas obtienen valores de frecuencias esperadas de 5 o mayores y ninguna celdilla obtiene frecuencias esperadas menores a 1. Si los datos no cumplen con las restricciones anteriores, se deberá emplear otra prueba estadística.

#### **4.5.2. Diseño de Dos Muestras Independientes.**

Este diseño se utiliza cuando se comparan dos muestras de la misma población, obtenidas independientemente, y una de ellas será tratada como grupo control y la otra como grupo experimental; o bien, cuando se comparan dos muestras provenientes de dos poblaciones que se supone son diferentes. El objetivo en ambos casos es, por lo general, encontrar diferencias entre los valores de la variable dependiente de los dos grupos. En el primer caso, para demostrar que la diferencia se debe a la presencia o ausencia de la variable independiente, en el segundo, para asentar que



realmente las dos poblaciones de las que provienen las muestras son diferentes entre sí y que esta diferencia ejerce sus efectos sobre la variable dependiente de manera significativa. Este tipo de diseño es muy utilizado en la investigación cuasiexperimental y de campo.

#### **4.5.2.1. Prueba No Paramétrica Para Dos Muestras Independientes: Chi Cuadrada ( $X^2$ ).**

Cuando un investigador cuenta únicamente con frecuencias de aparición de datos organizados en categorías discretas (nivel nominal de medición) de dos muestras independientes, y desea saber si existen diferencias entre los dos grupos (en cuanto a la forma en que se distribuyen las frecuencias) puede recurrir a la prueba de  $X^2$  para dos muestras independientes. La hipótesis nula se somete a prueba con la siguiente fórmula correspondiente.

En esta ocasión los **gl** (grados de libertad) se calculan por medio de la fórmula  $gl = (r - 1)(k - 1)$ ; dónde  $r =$  hileras y  $k =$  columnas.

El objetivo de la prueba de  $X^2$  es determinar la cercanía o lejanía que existe entre las frecuencias observadas y las frecuencias esperadas. A mayor sea la diferencia entre ellas, menor será la probabilidad de que las frecuencias observadas se deban al azar. Si esto sucede, el valor obtenido de  $X^2$  será tal que la probabilidad de ocurrencia asociada al azar será muy pequeña, y por lo tanto el investigador podrá decir que la forma en que se distribuyen las frecuencias por el observadas se deben a las características o atributos que distinguen a sus grupos.

**Ejemplo.** Un investigador desea saber si las respuestas afirmativas a una pregunta que dice: **La mayor parte del tiempo me siento bien conmigo mismo** (y que él considera es un indicador de la autoestima de una persona) se distribuirán de manera diferente según la autoestima de las personas, la cual se supone relacionada con las características de liderazgo de las mismas. De acuerdo con esto, clasifica a las personas como líderes, adeptos o "inclasificables".

**Hipótesis Nula.** No hay diferencia entre los dos grupos (autoestima alta y baja) en la proporción de miembros que son clasificados como líderes, adeptos o "inclasificables". La hipótesis alterna dice: hay una proporción mayor de individuos con autoestima alta que son clasificados como líderes que entre el grupo de autoestima baja.

**Prueba Estadística.** Se escoge la prueba de  $X^2$  para dos muestras

independientes porque los datos de los grupos se obtuvieron independientemente y los puntajes son frecuencias en categorías discretas (líderes, adeptos, "inclasificables").

**Nivel de Significación.** Sea  $\alpha = 0.01$ , y  $N = 95$ , número de Sujetos observados en la situación experimental.

**Distribución Muestral.** La  $\chi^2$  calculada con la fórmula anteriormente señalada, tiene una distribución muestral que se aproxima a la distribución muestral de Chi Cuadrada con  $gl = 2$ .

La significación de  $\chi^2 = 10.67$  cuando  $gl = 2$ , tiene una  $p$  asociada  $< 0.01$ , por lo tanto se rechaza la hipótesis nula y se acepta la alterna, que establecía que habría una mayor proporción de individuos con alta autoestima clasificados como líderes, que dentro de aquellos con autoestima baja.

**Restricciones estadísticas en el uso de la Prueba  $\chi^2$ .** Cuando  $k$  es mayor que 2, puede usarse esta prueba si menos del 20% de las celdillas tienen una frecuencia esperada menor que 5 y ninguna celdilla tiene una frecuencia esperada menor que 1. Si este no es el caso, se puede seguir el procedimiento de colapsar hileras o columnas, siempre y cuando dicho procedimiento sea lógico o congruente con los contenidos de las hileras o columnas. Si aún así no se cumple con las especificaciones señaladas, se tendrá que recurrir a otra prueba estadística.

#### **4.5.2.2. Prueba No Paramétrica Para Dos Muestras Independientes. Prueba de la Mediana.**

Una prueba estadística adecuada para el análisis de datos de dos muestras independientes con un nivel de medición ordinal es la Prueba de la Mediana. La prueba de la Mediana pone a prueba una hipótesis que hace referencia a la igualdad o no de una de las medidas de tendencia central: la Mediana informa al investigador sobre la probabilidad que dos grupos independientes tienen de provenir de poblaciones iguales o similares. Por lo tanto, la hipótesis nula supone que los grupos provienen de poblaciones que tienen la misma Mediana. Las hipótesis alternas pueden plantear que la Mediana de una población es diferente de la de otra (prueba de dos colas); o que una Mediana es mayor que otra (prueba de una cola). La variable independiente que se investiga deberá haber sido medida por lo menos a nivel ordinal.

La prueba de la Mediana deriva de la prueba de Chi Cuadrada, siendo su fórmula la misma, aunque en este caso la clasificación de los datos parte de la Mediana combinada de las observaciones (de los grupos que se están sometiendo a comparación), dando como

resultado una matriz de 2 x 2. En este caso, al estar la variable dependiente medida a nivel ordinal, se transforman los puntajes crudos a rangos, teniéndose en cuenta que no se puede dar el mismo rango a dos puntajes crudos iguales. En este caso se habla de la existencia de ligas y habrá que deshacerlas o desligarlas.

El procedimiento de desligar es muy sencillo. Suponga que ya asignó los rangos del 1 al 8 a los ocho primeros Sujetos, y que los siguientes tres tienen el mismo puntaje crudo. Como no se les puede asignar el mismo rango (9) a los tres, se hace lo siguiente: esos tres casos deberían ocupar los rangos 9, 10 y 11; pero tampoco se puede asignar diferentes rangos a puntajes crudos iguales. Entonces lo que se hace es promediar los rangos 9, 10 y 11, sumándolos y dividiéndolos entre tres (el número de Sujetos u objetos ligados, o cuyos puntajes crudos son iguales). De esta manera, los tres casos obtendrían cada uno un rango de 10:  $(9 + 10 + 11 = 30 / 3 = 10)$ . Al siguiente objeto o Sujeto le correspondería el rango de 12, puesto que los rangos 9, 10 y 11 ya fueron empleados en los casos ligados.

Habiendo ordenado en rangos a todos los objetos o Sujetos del estudio, se procede a determinar la Mediana combinada para esto, simplemente se divide el rango más alto entre 2, y éste será el valor de la Mediana combinada; la mitad de los casos caen por encima y la otra mitad por abajo de ella. En el caso en que muchos puntajes caigan en la Mediana, si N es suficientemente grande, se pueden excluir del análisis. Si al excluirlos se reduce N de manera drástica, y esto trajera como consecuencia la imposibilidad de superar las restricciones señaladas para la prueba de Chi Cuadrada (que son las mismas para esta prueba), entonces lo que se puede hacer es clasificar dichos puntajes por debajo de la Mediana. Como tienen un valor igual que ella, en estos casos se clasifican en la categoría "no la exceden", o sea por debajo de ella.

**Ejemplo.** Un investigador desea poner a prueba la hipótesis que asume que las madres de niños hiperkinéticos que obtuvieron información acerca de este problema tienen una actitud más favorable hacia sus hijos que aquellas que no la recibieron. La actitud hacia los hijos fue registrada con una escala que midió a nivel ordinal, en la que los puntajes más altos indican una actitud más favorable, y los puntajes más bajos indican una actitud menos favorable.

**Hipótesis Nula.** No hay diferencias entre la Mediana de actitud de las madres que recibieron información sobre la hiperkinesia y la de aquellas que no la recibieron. La hipótesis alterna señala que la Mediana de actitud de las madres que recibieron información

sobre hiperkinesis será mayor que la de las que no la recibieron.

**Prueba Estadística.** Como la medición de la variable dependiente (actitud) es de nivel ordinal y los grupos son independientes, se puede emplear la prueba de la Mediana para muestras independientes.

**Nivel de Significancia.** Sea  $\alpha = 0.05$  y  $N = 38$ , el número estudiado de madres:  $n_1 = 18$ , madres que recibieron información sobre hiperkinesis; y  $n_2 = 20$ , madres que no recibieron la información.

**Región de Rechazo.** Como la hipótesis alterna (que es la misma que la de trabajo) señala una dirección (que la información recibida sobre hiperkinesis producirá una Mediana de actitud más alta que la que se obtendría sin la información), la región de rechazo consiste en todos los resultados de una tabla con división en la Mediana que estén en la dirección predicha y sean tan extremos que la probabilidad asociada con su ocurrencia conforme a la hipótesis nula sea igual o menor que  $\alpha = 0.05$ .

Una  $X^2 = 1.738$  con  $gl = 1$  tiene una probabilidad de ocurrencia conforme a la hipótesis nula de  $p < 1/2 (0.20) = p < 0.20$  para una prueba de una cola. Como el nivel de significancia establecido por el investigador era  $\alpha = 0.05$ , se tiene que aceptar la hipótesis nula y rechazar la hipótesis alterna. Para poder aceptarla,  $X^2$  debería haber tenido un valor de por lo menos 5.41.

#### **4.5.3. Diseño de Dos Muestras Relacionadas.**

Se dice que dos muestras son o están relacionadas si cubren una de las dos siguientes características:

a) Si el mismo grupo va a ser sometido a investigación en dos ocasiones diferentes, sirviendo cada Sujeto como su propio control. Este es el diseño que más se emplea en los estudios de antes y después. En este caso, se hace un registro de la variable dependiente en la primera ocasión, se somete a todo el grupo a la situación experimental o a los efectos de la variable independiente, y posteriormente se vuelve a registrar la variable dependiente de interés. Por lo general el objetivo de estas investigaciones es determinar si hubo un cambio en la variable dependiente, y en qué medida, comparando los resultados antes y después del tratamiento experimental o de la introducción de la variable independiente.

b) El otro sentido en el que pueden estar relacionados dos grupos se refiere al hecho de si fueron apareados, apareados o

igualados al respecto de una o más variables que están relacionadas con la variable dependiente en estudio. Si por ejemplo se sabe que el uso de fármacos está relacionado con la edad de los usuarios, y no se quiere formar grupos control y experimental arriesgándose a que debido al azar estos grupos queden constituidos de manera que Sujetos mayores en edad pertenezcan a uno de los grupos y no al otro, se debe aparear, aparejar o igualar a los Sujetos con respecto a la edad antes de proceder a realizar la investigación. De no hacerlo así, las diferencias que se encuentren entre ellos se podrían deber a la diferencia inicial entre los grupos en cuanto a su edad, y no a la presencia o ausencia de la variable independiente que interesa. Para igualarlos, se procede a registrar las edades de todos los Sujetos, y asignarlos a uno u otro grupo en forma tal que las edades promedio de ambos grupos sean iguales desde un punto de vista estadístico. De esta manera se controla la variable extraña edad, y cualquier diferencia que se encuentre entre ellos puede asumirse que se debe a la variable independiente y no a la edad. Así, igualando a los grupos se tiene un caso de dos muestras relacionadas en el segundo sentido del término.

En el caso de los diseños de dos grupos, ya sean independientes o relacionados, por lo general uno de los grupos será considerado grupo control y el otro experimental. El grupo control puede no tener variable independiente o un valor pequeño de la misma, mientras que el grupo experimental puede tener presencia de la variable independiente o una cantidad mayor de ella que el otro. En estos diseños se tratará de establecer el efecto que una variable independiente, con dos valores (ninguno o poco y algo o mucho) tiene sobre una variable dependiente.

#### **4.5.3.1. Prueba No Paramétrica Para Dos Muestras Relacionadas. Prueba de Wilcoxon.**

La prueba de Wilcoxon se conoce también como prueba de rangos señalados y pares igualados. Es útil para analizar datos medidos a nivel ordinal y permite analizar no sólo la dirección de las diferencias entre pares de datos, sino también la magnitud relativa de las mismas: es decir, permite dar mayor peso a un par que muestra una diferencia grande entre sus miembros que a un par que exhibe diferencias pequeñas. En otras palabras, el investigador puede saber cuál de los miembros de un par es mayor, conocer el signo de las diferencias en cualquier par, y, clasificar éstas por orden de tamaño absoluto. Como su nombre lo indica, la prueba es adecuada para analizar datos que provienen de dos muestras relacionadas (igualadas o apareadas).

Si se tienen dos conjuntos de datos, con  $n_g$  iguales, derivados de

dos tratamientos diferentes se pueden analizar las diferencias entre cada par de la siguiente manera: a) en primer lugar se establecen las diferencias entre los pares, asignando el rango de 1 a la diferencia menos, el de 2 a la que sigue, y así sucesivamente, sin tener en cuenta los signos de las diferencias. De esta manera, una diferencia de +1 tiene un rango menos a una de -2 ó +2. b) En segundo lugar, se añade a cada rango el signo de la diferencia original, obteniéndose, finalmente, diferencias rangueadas positivas, y otras negativas ( $d_i$ ).

Si la hipótesis nula es cierta, es decir, si no hay diferencia entre los efectos de los dos tratamientos sobre la variable dependiente de interés, se esperaría encontrar un número similar de  $d_i$  mayores en ambos tratamientos. Es decir, algunos de los rangos mayores procederían de las  $d_i$  positivas, y otros de las negativas. Si se suman los rangos que tienen signos positivos por un lado, y los de signo negativo por otro, las dos sumas deberían ser iguales de acuerdo con la hipótesis nula. Sin embargo, si la suma de las  $d_i$  positivas fuera mayor que la de las  $d_i$  negativas, significaría que los tratamientos tuvieron efectos diferenciales sobre la variable dependiente y por lo tanto se tendría que rechazar la hipótesis nula.

En algunas ocasiones, por lo general cuando las muestras son grandes, los puntajes de algunos pares son iguales dando por resultado  $d_i = 0$ . Estos pares deben ser descartados del análisis. Otra situación posible es que dos o más  $d_i$  puedan ser del mismo tamaño. En este caso, para asignarles rango, se procede a desligar los pares tal y como se señaló en el apartado dedicado a la prueba de la Mediana para dos muestras independientes (promediando rangos).

La prueba de Wilcoxon presenta dos versiones, para muestras pequeñas ( $N = 6$  a  $N = 25$ ) y otra para muestras grandes, cuando  $N > 25$ .

**Ejemplo.** (Muestras Pequeñas). Un investigador desea saber si la asistencia al jardín de niños tiene algún efecto sobre el establecimiento de relaciones afectivas con los adultos. Califica el establecimiento de este tipo de relaciones mediante láminas que representan diversas situaciones afectivas con adultos. Con este instrumento obtiene puntajes de 0 a 100. Un puntaje más alto representa una capacidad mayor de establecer una relación afectiva que un puntaje más bajo. Sin embargo, no está seguro de si su instrumento mide a nivel intervalar; es decir, no puede establecer la exactitud numérica entre las diferencias, pero si puede clasificarlas en orden de tamaño absoluto.

Para someter a prueba el efecto de su planteamiento consigue 8 pares de gemelos idénticos como Sujetos. Al azar determina cuál de los miembros de un par de gemelos asistirá al jardín de niños y cuál permanecerá en casa. Cuando termina el período escolar, aplica su instrumento a los 16 niños para determinar el grado de establecimiento de relaciones afectivas con adultos, en el total de los niños de la muestra.

**Hipótesis Nula.** No hay diferencia en la capacidad de establecer relaciones afectivas con adultos entre los niños que asistieron al jardín de niños y los que permanecieron en sus hogares. En términos de esta prueba, lo anterior significa que la suma de los rangos positivos será igual a la suma de los rangos negativos. La hipótesis alterna señala que la capacidad de establecer relaciones afectivas con adultos será mayor entre los niños que asistieron a la escuela, en comparación con los que permanecieron en casa. Es decir, se espera que la suma de los rangos positivos sean mayor que la de los rangos negativos.

**Prueba Estadística.** La prueba de rangos señalados y pares igualados de Wilcoxon es adecuada para un estudio de dos grupos relacionados y con una variable dependiente registrada, de tal manera que se obtienen puntajes de diferencia que pueden clasificarse en orden de magnitud absoluta.

**Nivel de Significancia.** Sea  $\alpha = 0.025$ , y N el número de pares (8), menos los pares cuya  $d_i$  resulte cero.

**Distribución Muestral.** Se consideran los valores críticos a partir de la distribución muestral de T, para  $N_{25}$ .

**Región de Rechazo.** Puesto que se predice la dirección de la diferencia (los niños que asisten a la escuela establecerán más fácilmente relaciones afectivas con adultos que aquellos que permanecen en casa), se requiere una región de rechazo de una cola. La región de rechazo se compone, entonces, de todos los valores de T tan pequeños que la probabilidad asociada a su ocurrencia conforme a la hipótesis nula es igual o menor que  $\alpha = 0.025$  para una prueba de una cola.

Si la suma menor de los rangos señalados fuera  $1 + 3 = 4 = T$ . La distribución muestral de la Prueba de Wilcoxon señala que para  $N = 8$ , una  $T = 4$  permite rechazar la hipótesis nula con  $\alpha = 0.025$  y por lo tanto aceptar la hipótesis alterna, confirmándose así, la hipótesis de trabajo.

Cuando  $N > 25$ , la fórmula de la Prueba de Wilcoxon que permite analizar los datos se modifica.

El nivel de significancia alcanzado por la prueba se determina por medio de una Tabla de Probabilidades Asociadas con valores tan extremos como los valores observados de  $Z$  en la distribución normal.

**Ejemplo.** (Muestras Grandes). Un investigador desea saber si el tiempo de permanencia en prisión influye en la aceptación de las normas del grupo informal de la misma (efecto de **prisonalización**); es decir, quiere probar si a mayor permanencia en prisión se da un mayor efecto de **prisonalización**. Para someter a prueba su hipótesis, iguala en escolaridad, edad, sexo y tipo de delito a 30 pares de prisioneros, de manera tal que lo que diferencie a cada miembro de cada par es el tiempo que ha permanecido en prisión. La **prisonalización** es medida con un conjunto de reactivos que arrojan calificaciones de 0 a 100, donde a mayor calificación, mayor **prisonalización**. El investigador no puede asegurar que su nivel de medición sea intervalar, pero sí que las diferencias encontradas entre los pares se pueden clasificar en orden de valor absoluto.

**Hipótesis Nula.** No hay diferencias en el efecto de **prisonalización** entre los Sujetos con poco tiempo y los que tienen más tiempo de estar en prisión; es decir, la suma de los rangos positivos es igual a la de los rangos negativos. La hipótesis alterna señala que a mayor tiempo de estancia en prisión, mayor es el efecto de **prisonalización**.

**Prueba Estadística.** Se aplica la prueba de rangos señalados y pares igualados de Wilcoxon dado que los dos grupos están relacionados y la medición de la variable dependiente es a nivel ordinal.

**Nivel de Significancia.** Sea  $\alpha = 0.01$  y  $N = 30$ , el número de pares de prisioneros. El valor de  $N$  se reducirá si existen  $d_i$  iguales a cero.

**Distribución Muestral.** Conforme a la hipótesis nula, los valores de  $z$  alcanzados con la fórmula de Wilcoxon para muestras grandes está distribuida normalmente, con una media igual a cero y una varianza igual a uno. La Tabla correspondiente contiene la probabilidad asociada con la ocurrencia conforme a la hipótesis nula de valores tan extremos como cierto valor obtenido de  $z$ .

**Región de Rechazo.** Como la hipótesis alterna tiene dirección, la



región de rechazo es de una cola. La región de rechazo está formada por todos los valores de  $z$  (obtenidos de los datos con T) tan extremos, que la probabilidad asociada con su ocurrencia conforme a la hipótesis nula es igual o menor que  $\alpha = 0.01$ .

Si se obtuviera un valor  $z$  tan extremo como  $-3.11$ , ésta tiene una probabilidad (de una cola) asociada con su ocurrencia conforme a la hipótesis nula de  $p = 0.0009$ . Como  $p$  es menor que  $\alpha = 0.01$  y el valor de  $z$  está en la región de rechazo, se rechaza la hipótesis nula y se acepta la alterna, que señala que el efecto de **prisonalización** (aceptación de las normas del grupo informal dentro de la prisión) para los prisioneros con más tiempo de estancia en ella, es mayor, en comparación con los que tienen menos tiempo.

#### 4.5.4. Diseño de Más de Dos Muestras Independientes.

En algunas ocasiones el problema de investigación requiere que se manejen varios valores de la variable independiente, por ejemplo, más de dos. En estos casos, es necesario tener más de dos grupos, uno como el grupo control y dos o más grupos experimentales. Supóngase que se quiere saber que relación existe entre el grado de integración familiar o clase de la misma (falta del padre, falta de la madre, falta de ambos) y el uso o abuso de fármacos de algún tipo. En este caso, se requeriría de un grupo control, donde la integración familiar fuera óptima, y quizá sería interesante tener otros dos grupos con diferentes grados de desintegración familiar. El objeto de la investigación podría ser el tratar de averiguar si, por ejemplo, a mayor desintegración familiar mayor uso de estupefacientes. Dicho objetivo requiere, por lo tanto, de tres grupos. En este caso los tres grupos son independientes por que cada uno de ellos proviene de diferentes poblaciones: una población de familias integradas, otro de familias menos integradas, y un último de familias muy desintegradas. Se compararía el uso e incidencia de estupefacientes (la variable dependiente en este caso) y si se encuentran diferencias entre los grupos, suponiendo que los grupos; son iguales en los demás aspectos o variables, se podría establecer que existe una relación entre el grado de uso de estupefacientes y el grado de desintegración familiar.

Este diseño es especialmente adecuado cuando se tiene más de un valor de la variable independiente y se quiere determinar sus efectos sobre la dependiente. La ventaja de usar este tipo de diseños sobre el de dos grupos o muestras reside en que las relaciones que se descubren entre las variables serán más finas y precisas. Si se quisiera averiguar esto mismo usando sólo dos

grupos, se tendrían que hacer varias comparaciones entre pares de grupos: una investigación se avocaría a comparar al grupo de familias integradas con el de familias ligeramente desintegradas; otra comparación, y por lo tanto otra investigación, sería la que se realizaría comparando a un grupo de familias muy integradas con otro de familias muy desintegradas. Por último, se tendría que hacer una tercera investigación entre familias ligeramente desintegradas y las muy desintegradas. Sólo así, mediante tres investigaciones de dos grupos cada una, se podrían hacer todas las posibles comparaciones entre las condiciones familiares señaladas. Como es claro, esto implica un alto costo en tiempo, dinero y esfuerzo, además de que no tendría sentido hacerlo así, si se cuenta con otra posibilidad mejor: usar un diseño de más de dos grupos independientes.

Calcúlese, sólo por curiosidad, el número de investigaciones de dos grupos cada una, que se tendrían que llevar a cabo si se deseara comparar los valores de la variable dependiente, entre un grupo control, y cinco experimentales (cada uno con diferentes valores de una variable independiente). Serían en total 15 investigaciones de dos grupos cada una, las que se requerirían para poder establecer todas las posibles comparaciones entre los seis grupos.

Una de las consecuencias de establecer tantas comparaciones tiene que ver con el efecto de las mismas sobre el nivel de significancia con que se termina, en virtud de lo que se conoce como el efecto de Bonferroni.

Otra de las razones por las que a veces es conveniente usar más de dos grupos en una investigación es el hecho de que la relación entre las variables no necesariamente es lineal. Se dice que la relación es lineal si se establece que a más de  $x$  más de  $y$ , o bien que a más de  $x$  menos de  $y$ . Aunque la ciencia postula que la relación detectada entre más simple sea es mejor, a veces se dan casos en los que no se puede asegurar que ésta sea la situación. Puede suceder, por ejemplo, que sólo hasta cierto punto a más de  $x$  más de  $y$ , y que después de determinado valor de  $x$ , el valor de  $y$  ya no sigue aumentando sino que disminuye. Si sólo se trabaja con dos grupos, uno control, con nada de  $x$ , (la variable independiente) y otro con mucho de  $x$ , puede darse el caso de que no se detecte una diferencia entre los valores de  $y$  de los dos grupos, pero no porque no la haya, sino por que se podría tener en el grupo experimental tanta  $x$  que su efecto sobre  $y$  haya disminuido, y nunca se podría averiguar, si no se contara con un tercer grupo: un grupo experimental con un valor intermedio de  $x$ . Lo que sucede es que del primer al segundo grupo la relación es lineal, pero ya no lo sigue siendo del primer al tercer grupo; debido a esto será

necesario entonces tener más de dos grupos, y es en este caso cuando este tipo de diseño muestra su gran utilidad.

**4.5.4.1. Prueba No Paramétrica Para K Muestras Independientes: chi cuadrada  $X^2$ .**

Cuando los datos con los que cuenta el investigador consisten sólo de frecuencias en categorías discretas (sean nominales u ordinales), la prueba de  $X^2$  puede determinar la significación de las diferencias entre  $k$  grupos independientes. Esta prueba es una extensión directa de la prueba de  $X^2$  para dos muestras independientes presentada anteriormente. Por tanto, las restricciones que se aplican a la prueba de  $X^2$  de  $k$  muestras independientes son las mismas que para dos muestras independientes, y consecuentemente se deberán tomar en consideración.

Para aplicar la prueba de  $X^2$  se dispone de frecuencias en una matriz  $K \times R$ . La hipótesis nula supone que las  $K$  muestras de frecuencias o proporciones provienen de la misma población o de poblaciones idénticas. La hipótesis de que las muestras difieren entre sí se pone a prueba aplicando la fórmula de chi cuadrada.

La distribución muestral de  $X^2$  calculada con la fórmula antes presentada, se aproxima a la distribución de chi cuadrada con  $gl = (k-1)(r-1)$ , dónde:

- k = número de columnas
- r = número de hileras

**EJEMPLO.** Un investigador desea saber si la preferencia por diferentes tipos de espectáculos se ve influida por la clase social a la que se pertenece. Clasifica a 390 personas de una comunidad en cuanto a su clase social en cuatro niveles: Clase Alta, Clase Media Alta, Clase Media y Clase Baja, y les pregunta por medio de una pequeña encuesta, a qué tipo de espectáculos prefieren asistir en su tiempo libre, pidiéndoles que sólo marquen uno de los tres que se señalan en la pregunta (conciertos, cine, deportes).

**Hipótesis Nula.** La preferencia de los diferentes tipos de espectáculos es la misma en todas las clases sociales. La hipótesis alterna señala que preferencia por diversos tipos de espectáculos difiere en función de la clase social a la que pertenezcan los Sujetos investigados.

**Prueba Estadística.** puesto que los grupos que se estudian son independientes, y más de dos, y los datos se encuentran en

categorías discretas, la prueba apropiada es la  $X^2$  para  $k$  muestras independientes.

**Nivel de Significancia.** Sea  $\alpha = 0.01$  y  $N = 390$ , el número de Sujetos cuya clase social y preferencia se determinó en el estudio.

**Distribución Muestral.** Conforme a la hipótesis nula, la  $X^2$  calculada con la fórmula correspondiente, se distribuye en forma aproximada a la de chi cuadrada con  $gl = (k-1)(r-1)$ , y la probabilidad asociada con la ocurrencia de acuerdo a la hipótesis nula de valores tan grandes como el valor observado de  $X^2$  queda establecido por su distribución muestral.

**Región de Rechazo.** La región de rechazo consiste en todos los valores de  $X^2$  tan grandes que la probabilidad asociada conforme a la hipótesis nula sea igual o menor que  $\alpha = 0.01$ . La región de rechazo es de dos colas puesto que la hipótesis alterna no tiene dirección.

Consultando la Tabla correspondiente, se observa que una  $X^2 = 69.2$  con  $gl = 6$ , es significativa a un nivel menor que 0.001. Como  $p = <0.001$  es menor que  $\alpha = 0.01$ , se rechaza la hipótesis nula y se acepta la alterna en el sentido de que la diferencia en la preferencia por distintos tipos de espectáculos se ve influida por la clase social a la que pertenece el individuo.

#### **4.5.4.2. Extensión de la Prueba de la Mediana.**

La extensión de la prueba de la Mediana permite que el investigador determine si  $k$  grupos independientes (con igual o diferente número de Sujetos) proceden de la misma población o de poblaciones que tienen Medianas iguales. Se utiliza cuando la variable dependiente ha sido medida por lo menos a nivel ordinal.

El procedimiento es muy semejante al de la prueba de la Mediana para dos grupos independientes. Es decir, en primer lugar se tiene que determinar la Mediana combinada de todos los puntajes, independientemente de a cuáles grupos pertenecen. En segundo lugar, se señala cuáles puntajes caen por encima del valor de la Mediana combinada, y cuáles por debajo. De existir puntajes que caen exactamente en el valor de la Mediana y éstos son muchos, se pueden colocar uno por encima y otro por debajo del valor de la Mediana, hasta agotarlos. Si son pocos los casos en esta situación, es decir, si el tamaño de  $N$  no se reduce grandemente, se pueden eliminar del análisis, modificándose tanto el Gran Total, como los Totales Marginales (tanto de columnas como de

hileras) correspondientes.

Debe recordarse que esta prueba también tiene las restricciones de uso de la prueba de  $X^2$  para muestras independientes, a saber: ninguna celdilla de la matriz puede tener frecuencias esperadas de 1 o menores; y sólo el 20% de las celdillas pueden tener frecuencias esperadas menores a 5. Si no se superan las anteriores restricciones, se puede, si la naturaleza de las categorías lo permite, colapsar columnas, siempre y cuando sea lógico y consistente hacerlo desde un punto de vista teórico formal o que no sea incongruente desde el punto de vista del sentido común.

En tercer lugar, deberán colocarse los puntajes señalados en una matriz de  $K \times 2$ . En cuarto lugar, se habrán de calcular las frecuencias esperadas. Por último, habrá de desarrollarse la fórmula para determinar el valor de  $X^2$  observada.

Es pertinente aclarar que en el caso de esta prueba, los grados de libertad serán siempre igual a  $(k - 1)(2 - 1)$ . Para determinar la probabilidad asociada al valor de  $X^2$  observado de acuerdo a la hipótesis nula se compara con una distribución muestral que se aproxima a la chi cuadrada con  $gl = (k - 1)(2 - 1)$ .

**Ejemplo.** Un investigador desea averiguar si el nivel de escolaridad formal alcanzado por las madres tiene relación, o produce algún efecto diferencial sobre el interés que muestran en la instrucción de sus hijos.

Se entrevista a las madres solicitando informen el número de veces que han acudido a la escuela de *motu proprio* a la escuela de sus hijos para conocer su desempeño escolar, y se desea saber si su interés, medida de esta manera, difieren en función de la escolaridad formal alcanzada por ellas.

**Hipótesis Nula:** El grado de interés por el desempeño escolar de los hijos es igual entre las madres de diferente escolaridad formal alcanzada. La hipótesis alterna señala que existen diferencias en el grado de interés mostrado por las madres en función de su escolaridad.

**Prueba Estadística.** Como el número de grupos a comparar es mayor de dos, y el número de visitas escolares es una medida ordinal del grado de interés en la instrucción del hijo, la extensión de la prueba de la Mediana es la adecuada.

**Nivel de Significancia.** Sea  $\alpha = 0.05$  y  $N = 44$ , el número de madres estudiadas.

**Distribución Muestral.** De acuerdo a la hipótesis nula, el valor de  $X^2$  observado está distribuido aproximadamente como chi cuadrada con  $gl = k - 1$  cuando  $r = 2$ .

**Región de Rechazo.** La región de rechazo consiste en todos los valores de  $X^2$  tan grandes que la probabilidad asociada a su ocurrencia de acuerdo con la hipótesis nula sea igual o menos que  $\alpha = 0.05$ ; la región es de dos colas, porque la hipótesis de trabajo no plantea dirección.

De acuerdo con los valores de la distribución muestral, si se observa que un valor de  $X^2 = 1.295$ , con  $gl = 3$ , tiene una probabilidad asociada de ocurrencia conforme a la hipótesis nula entre 0.80 y 0.70. Como  $p < 0.80 > 0.70$  es mayor que  $\alpha = 0.05$ , se acepta la hipótesis nula y se rechaza la alterna. Esto significa, que de acuerdo a los resultados de este ejemplo, el nivel de escolaridad formal alcanzado por las madres no se relaciona o no influye sobre el grado de interés que muestran en las instrucción de sus hijos. Es decir, el interés por la instrucción de los niños no varía en virtud de la escolaridad alcanzada por sus madres.

#### 4.5.4.3. Análisis de Varianza de una Entrada de Kruskal-Wallis.

La prueba de Kruskal-Wallis es útil para probar los resultados de **k** muestras que vienen de poblaciones diferentes. Por lo general los valores de las muestras, aún provenientes de una misma población, difieren entre sí hasta cierto punto. De esta manera, la prueba de Kruskal-Wallis se utiliza para determinar si las diferencias que se encuentran entre las muestras son reales o simplemente están dadas por el azar. La hipótesis nula que se somete a prueba mediante este análisis estadístico supone que las **k** muestras vienen de la misma población o de poblaciones cuyos promedios de rangos son idénticos. La prueba requiere que la variable dependiente sea medida por lo menos a nivel ordinal y que tenga como base una distribución continua.

La prueba de Kruskal-Wallis se presenta en dos versiones: una para muestras pequeñas ( $n's < 5$ ) y otra para muestras grandes ( $n's > 5$ ). Si las **k** muestras realmente vienen de la misma población o de poblaciones iguales, es decir, si la hipótesis nula es cierta, **H** (el valor resultante de la prueba) se distribuirá como chi cuadrada, con  $gl = k - 1$ , siempre que los tamaños de las diferentes **k** muestras sean grandes (o sea, sus  $n's$  sean mayores de 5). Cuando  $k = 3$  y sus  $n's$  son iguales a 5 o menores, los valores de **H** no se aproximan a la distribución de chi cuadrada, y se tiene que emplear una distribución muestral diferente.

En este caso, se trabaja con rangos. De esta manera, el primer paso consiste en que cada una de las  $N$  observaciones tiene que reemplazarse por rangos. Es decir, todos los puntajes de las  $K$  muestras combinadas (independientemente de la muestra a la que pertenezcan) se ordenan en una sola serie. En segundo lugar, se reemplaza el puntaje más pequeño por el rango 1, el siguiente en tamaño por el rango 2, y así sucesivamente, hasta darle el rango  $N$  al puntaje más alto. En este caso, particularmente cuando se tienen muestras grandes ( $k > 3$  y  $n_j > 5$ ), pueden encontrarse puntajes iguales o ligados. Cuando así sucede, se deberán desligar los puntajes por medio del procedimiento de obtención de promedios de rangos, tal y como se explicó en el apartado dedicado a la Prueba de la Mediana para dos muestras independientes. En seguida se llevan a cabo las operaciones señaladas por la fórmula de  $H$ , y se procede a consultar la distribución muestral de Chi Cuadrada, con  $gl = k - 1$  para el caso de muestras pequeñas, con objeto de determinar la probabilidad asociada de ocurrencia conforme a la hipótesis nula de un valor observado de  $H$ .

**Ejemplo.** (Muestras Pequeñas). Un investigador desea poner a prueba la hipótesis que señala que los profesores de una institución educativa poseen diferentes grados de autoritarismo dependiendo del grado u orientación que tengan hacia la enseñanza o hacia las tareas administrativas, porque supone que aún cuando algunos profesores tan sólo se dedican a labores docentes, sus aspiraciones pueden estar orientadas a éstas. El investigador clasifica a los profesores en tres grupos: aquellos que sólo les interesa dar clases; aquellos que quieren ser parte de la administración; y aquellos que ocupan puestos administrativos. Para obtener sus datos aplica a los Sujetos de estos grupos, la Escala F de Adorno y cols. (1950) que mide autoritarismo. Cuenta con una muestra de tan sólo 14 Sujetos que quedan clasificados en los tres grupos señalados. Su hipótesis de trabajo plantea que los tres grupos se diferenciarán con respecto a los promedios resultantes de las respuestas dadas a la Escala F.

**Hipótesis Nula.** La hipótesis nula señala que no hay diferencia entre los promedios de los puntajes  $F$  de los profesores, cualquiera que sea su orientación (hacia la docencia, la administración, o los que están laborando en ésta). La hipótesis alterna señala que sí se encontrarán diferencias entre dichos promedios.

**Prueba Estadística.** Como se tienen tres grupos medidos independientemente; un instrumento que mide a nivel ordinal la variable autoritarismo, y ésta es continua (de poco a mucho), la prueba adecuada es la de Kruskal-Wallis.

**Nivel de Significancia.** Sea  $\alpha = 0.05$ ,  $N = 14$ , el número total de Sujetos;  $n_1 = 5$ , número de profesores orientados hacia la enseñanza;  $n_2 = 5$ , número de profesores orientados hacia la administración; y  $n_3 = 4$ , número de administradores.

**Distribución Muestral.** Para  $k = 3$  y valores pequeños de  $n_j$ , la distribución muestral para valores observados de  $H$ , presenta las probabilidades asociadas con la ocurrencia conforme a la hipótesis nula, de valores tan grandes como el de un  $H$  observada.

**Región de Rechazo.** La región de rechazo está formada por todos los valores de  $H$  tan grandes, que la probabilidad asociada con su ocurrencia de acuerdo con la hipótesis nula es igual o menor que  $\alpha = 0.05$ .

Cuando las  $n_j$ 's son 5, 5, y 4; entonces  $H \geq 6.4$  tiene una probabilidad de ocurrencia conforme a la hipótesis nula de  $p < 0.049$ . En virtud de que esta probabilidad es menor que el nivel de significancia previamente establecido ( $\alpha = 0.05$ ), se decide aceptar la hipótesis alterna y rechazar la nula. Es decir, se confirma la hipótesis de trabajo del investigador que predijo que los grupos diferirían en autoritarismo, según fuera su orientación.

La fórmula de  $H$  sufre una modificación cuando los estudios involucran muestras grandes y se encuentran puntajes ligados. Ya se señaló que estas ligas deben desbaratarse conforme al procedimiento anteriormente explicado. Pero debido a que los puntajes ligados afectan el valor de  $H$ , la fórmula se modifica incluyendo una corrección por ligas.

**Ejemplo.** (Muestras grandes con puntajes ligados). Un investigador desea saber si el contenido proteico de la alimentación de madres embarazadas afecta el peso del niño al nacer. Investigó el contenido proteico de la alimentación consumida por 56 madres, y las clasificó en ocho grupos, de mayor a menor contenido proteico. Al nacer los niños, se registraron sus pesos y se analizaron los datos. Su hipótesis de trabajo predice que el contenido proteico de la alimentación de las madres influirá en el peso de los niños al nacer.

**Hipótesis Nula.** La hipótesis nula señala que no hay diferencia en el peso promedio de los niños, independientemente del grado del contenido proteico en la alimentación de las madres. La hipótesis alterna indica que los pesos promedio de los niños al nacer



difieren entre los grupos de distinto contenido proteico en la alimentación de las madres.

**Prueba Estadística.** Como se tienen datos de observaciones independientes de más de dos grupos, se optó por una prueba de  $k$  muestras independientes. Aunque la variable peso de los niños pertenece a una escala de razón (peso en kilogramos) el investigador se decide por la prueba de Kruskal-Wallis para evitar hacer suposiciones incorrectas concernientes a la homogeneidad de la varianza y la normalidad de la distribución en los grupos estudiados. Si tuviera bases para fundamentar tales supuestos, él podría recurrir a una prueba paramétrica. Como carece de éstas, la decisión hecha es la más adecuada. Por otra parte, la variable peso, cumple con el requisito de ser una distribución continua.

**Nivel de Significancia.** Sea  $\alpha = 0.05$  y  $N = 56$ , el número total de los niños nacidos de la muestra.

**Distribución Muestral.** Con una  $N$  grande y  $K > 3$ , la distribución muestral de  $H$  obtenida con la fórmula corregida por ligas, se distribuye aproximadamente como chi cuadrada con  $gl = k - 1$ . La probabilidad asociada de ocurrencia de acuerdo con la hipótesis nula de valores tan grandes como el de  $H$  observada puede determinarse empleando la distribución muestral.

**Región de Rechazo.** La región de rechazo consiste en todos los valores de  $H$  tan grandes que su probabilidad asociada de ocurrencia de acuerdo con la hipótesis nula para  $gl = k - 1 = 7$  es igual o menor que  $\alpha = 0.05$ . La región de rechazo es de dos colas por que la hipótesis alterna no señala dirección.

La distribución muestral de Chi Cuadrada indica que la probabilidad asociada con la ocurrencia de acuerdo con la hipótesis nula de un valor tan grande como  $H = 18.566$  con  $gl = 7$  es  $p < 0.01$ . Como esta probabilidad es menor que el nivel de significancia  $\alpha = 0.05$  previamente establecido, se decide rechazar la hipótesis nula y aceptar la alterna, que establecía diferencias en el peso de los niños al momento de su nacimiento, según los diferentes niveles de contenido proteico en la alimentación de las madres durante el embarazo.

#### 4.5.5. Diseño de más de Dos Muestras Relacionadas.

Este diseño es una extensión del diseño de dos grupos relacionados. La lógica de su utilización es la misma que la del diseño de más de dos grupos independientes. Obviamente, en la

medida en que el número de grupos sea mayor, más difícil será igualarlos, aparearlos o aparejarlos en aquellas variables extrañas que se relacionan de manera importante con la variable dependiente.

#### **4.5.5.1. Prueba No Paramétrica con k Muestras Relacionadas: Análisis de Varianza de dos Entradas de Friedman.**

La prueba de análisis de varianza de dos clasificaciones por rangos de Friedman se emplea cuando se tienen **k** muestras relacionadas, y la variable dependiente ha sido medida por lo menos a nivel ordinal. Al ser las muestras igualadas, el número de casos es el mismo en las **k** condiciones. Esta prueba, al igual que las anteriores, puede emplearse para muestras pequeñas y muestras grandes.

Para analizar los datos con la prueba de Friedman se colocan en una matriz de N hileras y K columnas. Inmediatamente después se procede de la siguiente manera: en primer lugar, los puntajes de la variable dependiente deberán transformarse en rangos. La transformación en rangos se hace para cada hilera, dando el rango de 1 al puntaje más bajo y el rango K al puntaje más alto de la hilera. Los puntajes de cada hilera se ordenan por separado, por lo que con **k** condiciones, los rangos de cualquier hilera van de 1 a K. La prueba de Friedman determina la probabilidad con la que las diferentes columnas (tratamientos o muestras) proceden de la misma población.

Si la hipótesis nula (que todas las muestras--columnas proceden de la misma población) es cierta, la distribución de los rangos en cada columna será aleatoria y los K rangos asignados deberán aparecer en cada una de ellas con igual frecuencia. Es decir, el conjunto de rangos en cada columna representa una muestra aleatoria de una distribución rectangular discontinua, y los totales de rangos de las diferentes columnas serán iguales; y en virtud de que todas las columnas tienen el mismo número de Sujetos o casos, las medias de los rangos de las diferentes columnas serán iguales de acuerdo con la hipótesis nula.

Ahora bien, dependiendo del tamaño de las muestras, se determina la distribución muestral conforme a la hipótesis nula de  $X^2_r$ . Para muestras pequeñas ( $k = 3$  y N de 2 a 9; ó  $k = 4$  y N de 2 a 4) los valores de  $X^2_r$  observados tienen probabilidades exactas conforme a la hipótesis nula en la distribución muestral de probabilidades asociadas con valores de  $X^2_r$  en el análisis de varianza de dos clasificaciones por rangos de Friedman (ver Siegel, 1976). Para las muestras grandes ( $K = 3$  o más, y  $N = 4$  o más), los valores de

$X^2_r$  tienen probabilidades asociadas conforme a la hipótesis nula semejantes a las de la distribución de chi cuadrada con  $gl = k - 1$ .

**Ejemplo.** (Muestra Pequeña). Un investigador desea saber si la atención que los individuos prestan a ciertos estímulos difiere en función del color que éstos tengan. Forma tres grupos de cuatro Sujetos cada uno, igualados o apareados en las variables pertinentes (agudeza visual, respuesta a estimulación cromática y destreza manual). Somete a cada uno de los cuatro Sujetos, asignándolos al azar, a una prueba de tiempo de reacción ante cuatro estímulos idénticos, pero de diferente color (amarillo, azul, rojo y verde).

**Hipótesis Nula.** La hipótesis nula señala que los diferentes colores de los estímulos no tienen efecto diferencial sobre los tiempos de reacción. La hipótesis alterna indica que sí se darán diferentes efectos producidos por el color.

**Prueba Estadística.** Debido a que se tienen tres grupos apareados o igualados se requiere de una prueba para  $k$  muestras relacionadas. A pesar de que la medición del tiempo pertenece a una escala de razón, dado el pequeño tamaño de  $N$  en cada condición (tamaño que impide la distribución normal de los puntajes), el investigador decide tomar la variable tiempo de reacción como un indicador a nivel ordinal del grado de atención que los estímulos pueden despertar. La prueba más adecuada es el análisis de varianza de dos clasificaciones por rangos de Friedman.

**Nivel de Significancia.** Sea  $\alpha = 0.05$  y  $N = 3$ , el número de hileras o grupos.

**Distribución Muestral.** Las probabilidades exactas asociadas a los valores observados de  $X^2_r$  conforme a la hipótesis nula tienen su distribución muestral

**Región de Rechazo.** La región de rechazo consiste en todos aquellos valores tan grandes de  $X^2_r$  cuyas probabilidades asociadas de ocurrencia de acuerdo con la hipótesis nula son iguales o menores que  $\alpha = 0.05$ . La región de rechazo es de dos colas puesto que la hipótesis alterna no establece dirección.

Consultando la distribución muestral de probabilidades exactas asociadas a valores obtenidos de  $X^2_r$ , se observa que cuando  $K = 4$  y  $N = 3$ , una  $X^2_r = 7.4$ , tiene una probabilidad asociada de ocurrencia conforme a la hipótesis nula de 0.033. Como  $p = 0.033$

es menor que el nivel de significancia previamente establecido,  $\alpha = 0.05$ , se rechaza la hipótesis nula y se acepta la alterna, que señala que los tiempos de reacción de los Sujetos se ven afectados diferencialmente por el color de los estímulos a los que responden.

**Ejemplo.** (Muestra Grande). Un investigador desea saber si tres diferentes métodos didácticos afectan diferencialmente el proceso de enseñanza-aprendizaje de un material determinado. Para esto iguala a 18 grupos de 3 Sujetos cada uno en las variables pertinentes (cociente intelectual y promedio de calificaciones en años escolares anteriores). Como no está seguro de que su instrumento que mide aprendizaje lo haga a nivel de una escala intervalar pero sí a nivel ordinal, establece que los puntajes más altos señalan mayor aprendizaje.

**Hipótesis Nula.** La hipótesis nula señala que no habrá diferencias de aprendizaje entre los tres grupos. La hipótesis alterna plantea que sí se encontrarán diferencias dependiendo del método de enseñanza.

**Prueba Estadística.** Como se tienen tres grupos apareados, se requiere de una prueba para  $k$  muestras relacionadas. Como  $N$  y  $K$  son grandes y la variable dependiente está medida tan sólo a nivel ordinal, la prueba adecuada es la de Friedman.

**Nivel de Significancia.** Sea  $\alpha = 0.05$  y  $N = 18$ , el número de grupos apareados.

**Distribución Muestral.** Cuando  $N$  y  $K$  son grandes, los valores de  $X^2_r$  se distribuyen en forma aproximada a los de  $X^2$  con  $gl = k - 1$ . La probabilidad asociada de ocurrencia conforme a la hipótesis nula de un valor tan grande como el obtenido con  $X^2_r$  se determina por medio de la distribución de valores críticos de chi cuadrada.

**Región de Rechazo.** La región de rechazo consiste en todos los valores de  $X^2_r$  tan grandes que su probabilidad asociada de ocurrencia conforme a la hipótesis nula es igual o menor que  $\alpha = 0.05$ .

Consultando la tabla de chi cuadrada con  $gl = k - 1 = 3 - 1 = 2$ , se observa que un valor de  $X^2_r = 8.4$  es significativo entre los niveles de 0.02 y 0.01. Como  $p < 0.02$  es menor que  $\alpha = 0.05$ , se rechaza la hipótesis nula y se acepta la alterna, que señalaba que los distintos métodos didácticos afectarían diferencialmente el proceso de enseñanza-aprendizaje.

**REFERENCIAS**

- Adorno, T., Frenkel-Brunswik, E., Levinson, D.J. y Stanford, R.N.  
(1950): **The Authoritarian Personality**. Nueva York: Harper.
- Keppel, G. (1973). **Design and analysis. A researcher's handbook**.  
New Jersey: Prentice-Hall
- Siegel, S. (1976): **Estadística No Paramétrica**. México: Editorial  
Trillas.

## V. DISEÑOS DE INVESTIGACIÓN: PROCEDIMIENTOS ESTADÍSTICOS PARAMÉTRICOS

### 5.1. CONCEPTOS Y PROCEDIMIENTOS BÁSICOS DE LA ESTADÍSTICA PARAMÉTRICA.

#### 5.1.1. Medidas de Tendencia Central y de Dispersión.

La Media junto con la Mediana y el Modo, forman parte de las medidas de Tendencia Central,. Son medidas sumarias que identifican los puntajes o calificaciones característicos, típicos o representativos de una distribución, o de manera más general, los de una muestra o población. La media ( $\bar{X}$ ) es la más precisa pues toma en cuenta toda la información disponible, sin embargo, también es la más sensible a la falta de normalidad de la distribución, lo que altera fácilmente su valor. De manera más precisa, la  $\bar{X}$  es el promedio aritmético de los puntajes de una distribución.

La mediana (Med) en cambio, divide la distribución en dos partes iguales a partir del 50% de la misma y es, además, la que menos se ve afectada por la no normalidad o por puntajes extremos. El Modo (Mo.) la tercera y última de las medidas más usuales de tendencia central, representa el punto de concentración de frecuencias de una distribución y es la única que puede repetirse dos o más veces determinando que las distribuciones sean multimodales. La presencia de múltiples Modos es un signo claro de no normalidad de la distribución, y puede indicar, en algunas ocasiones, la presencia de diferentes poblaciones. El Mo. se distingue de las otras medidas de tendencia central en tanto se le puede considerar como un promedio "real". A la pregunta: ¿Qué talla de zapatos es la más vendible en México? el Mo. proporcionaría el tamaño real solicitado, por ejemplo, talla 26 porque se encontró, que dentro de un determinado lapso de tiempo, fue la más frecuentemente solicitada o vendido; la  $\bar{X}$ ., y la Med., podrían proporcionar tallas inexistentes (v.gr., 25,7 o 26.3).

Por otro lado la desviación estándar ( $\sigma$  ó S) es considerada dentro de las medidas de variabilidad, más frecuentemente utilizadas, junto con los percentiles y el rango. Son medidas de dispersión, de alejamiento, o de heterogeneidad de los puntajes de una distribución, a partir de un punto de ésta, así por ejemplo, la desviación estándar ( $\sigma$ ) es la dispersión de los puntajes a partir de la  $\bar{X}$ ; no tendría sentido hablar de variabilidad sin saber con respecto a que, de aquí, que arriba se halla señalado el significado conjunto de los parámetros de la curva. Al igual que las medidas de tendencia central, las de variabilidad, tienen sus propias características, así como sus respectivas normas de uso adecuado u óptimo. La Desviación Estándar ( $\sigma$ ) es la más precisa en la medida que para su desarrollo requiere de toda la información disponible, pero al igual que la  $\bar{X}$ ., se ve

afectada por la falta de normalidad de la distribución. Es la medida de variabilidad con respecto a la  $\bar{X}$ ., de su distribución y a más pequeña sea la  $\sigma$ ., menor será la dispersión esto es, más compactados los puntajes en torno a la  $\bar{X}$ ., o más homogéneas fueron las respuestas de los sujetos. Siempre serán mejor las  $\sigma$ s pequeñas que las grandes, como se podrá deducir de lo que sigue a lo largo de este capítulo.

Otra medida de variabilidad son los percentiles. En una distribución de valores ordenados, de menor a mayor, o viceversa, el valor que la divide en dos partes porcentualmente complementarias, son los Percentiles (Pp.): Pp. 25 divide la serie dejando una cuarta parte de los valores con menor o igual magnitud a él y a las tres cuartas partes restantes con magnitudes más grandes o iguales a él. La Med., es igual o equivalente al Pp. 50% que divide la distribución, en dos partes de igual tamaño, por arriba y por abajo de ese punto. Usualmente este valor se ubica en la posición  $(N + 1)/2$ . Se pueden hacer dos o más divisiones percentuales, pero cualquier percentil se ubica en la posición  $(N \times P)/100$  y cualquier percentil representa un lugar, una ubicación, de un valor o puntaje en la distribución y al igual que la Med., es la menos afectada por la no-normalidad. Con la mediana y los percentiles se pueden utilizar variables cuantitativas discretas o cuantitativas continuas con distribuciones no normales y variables cualitativas con mediciones ordinales.

El rango, la otra medida de variabilidad, es la más común y la menos precisa, ya que sólo utiliza dos valores de toda la distribución: el puntaje más alto menos el más bajo. Con el modo y rango se pueden utilizar variables cuantitativas discretas y cualitativas con medición nominal.

Otra medida de variabilidad, menos utilizada que las anteriores, es el Coeficiente de Variabilidad (CV) es la desviación estándar expresada en porcentajes de la media, a la que también se le conoce como medida de variabilidad relativa:

$$C.V. = \frac{\sigma \times 100}{\bar{X}}$$

### **5.1.2 Normas del Uso Adecuado de la Media ( $\bar{x}$ ) y de la Desviación Estándard. (S)**

Debido a que, como ya se señaló, tanto la  $\bar{X}$ ., como la S., requieren de todos los datos de la distribución, es que se ven afectados por la forma en que se distribuyen los puntajes, de aquí. que la primera norma de uso de estas dos estadísticas sea que la distribución de los puntajes debe ser normal o por lo menos sin desviaciones serias de la normalidad, las variables cuantitativas continuas y la medición de las

variables de donde provienen los puntajes debe ser por lo menos intervalar (con propiedades de orden y distancia a partir de un punto cero arbitrario), aunque veremos más adelante que algunas mediciones ampliamente aceptadas y que han mostrado su utilidad práctica, son sólo mediciones ordinales.

La normalidad de una distribución de puntajes se ajusta al modelo teórico de la Curva Normal (Fig. 5. 1) . La no-normalidad se aplica a distribuciones positiva (Fig.5. 2) o negativamente coleadas (Fig. 5.3).

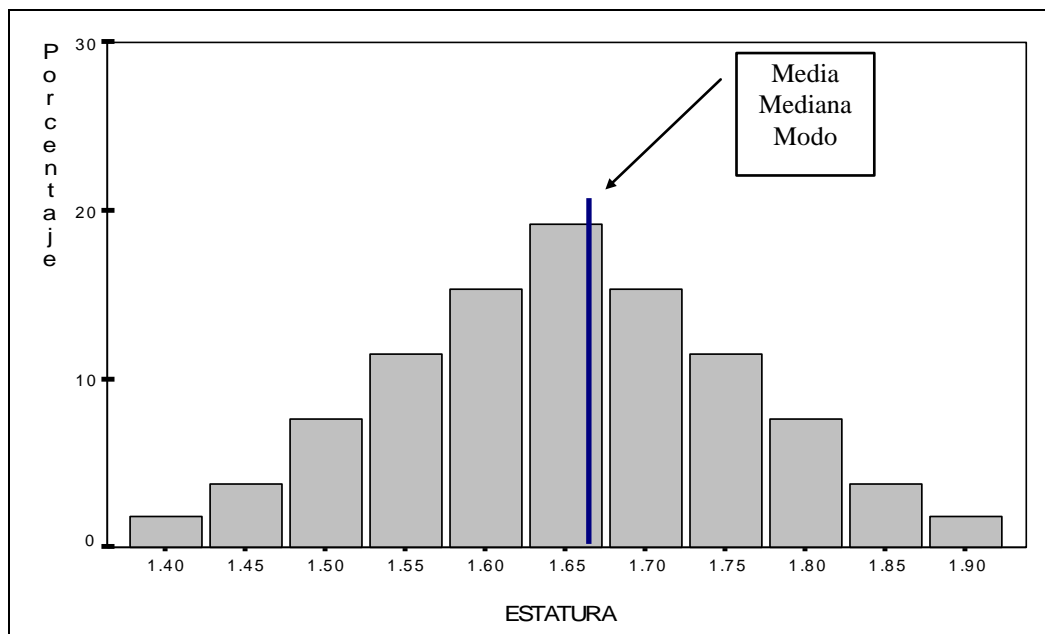


Fig 5.1 Curva con  $\bar{X}$ , Med y Mo. en el centro

En la Fig. 1 es evidente que la distribución normal es simétrica, la mitad de un lado replica la otra mitad y las tres medidas de tendencia central tienen el mismo valor . Mientras que en la Fig 2 puede verse cómo el grueso de los datos caen a la izquierda de la curva, esto es, puntajes con valores pequeños, y basta que uno o algunos puntajes tengan valores extremos, (puntajes altos,) para que se cargue la distribución. Si imaginamos que la curva normal, en lugar de ser un concepto abstracto, fuera un objeto concreto que pudiéramos sostener del centro, veríamos como se mantendría



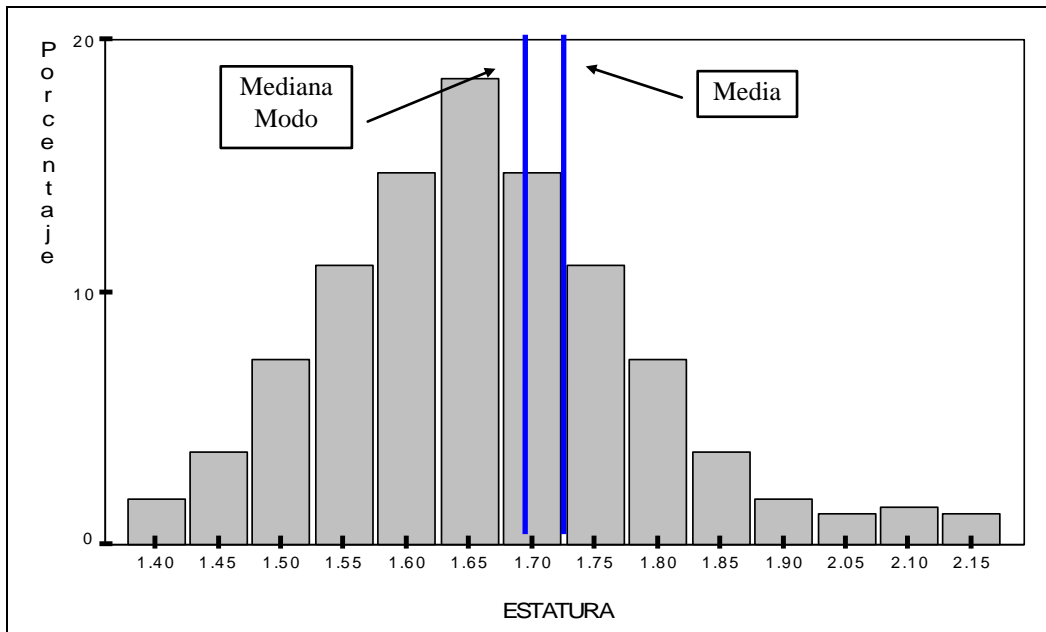


Fig 5.2 Distribución coleada a la derecha

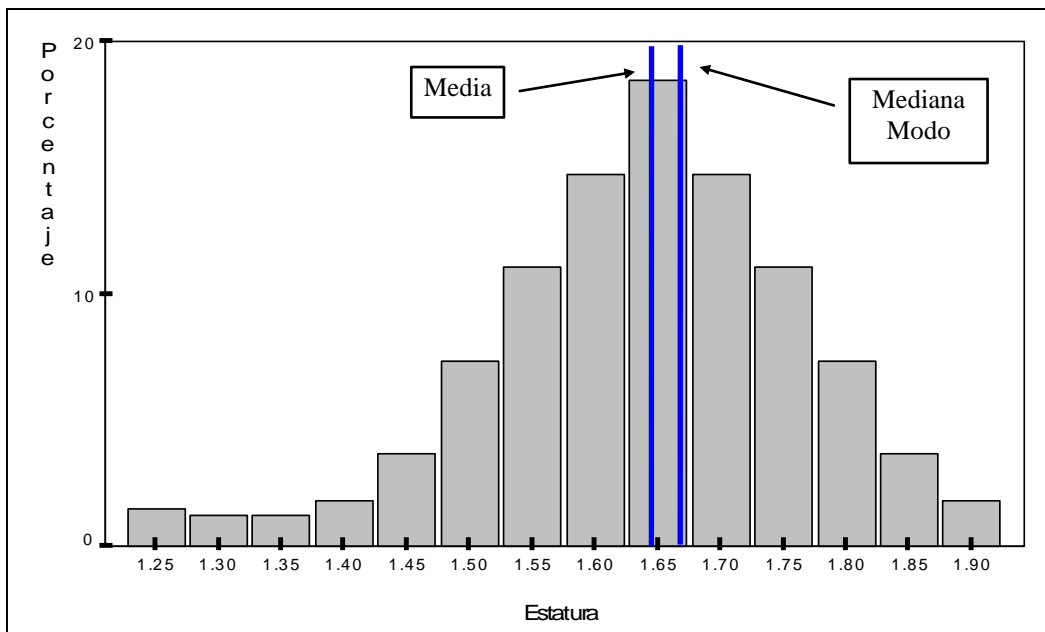


Fig 5.3 Distribución coleada a la izquierda

balanceada la curva; ahora, sí a la curva que estamos sosteniendo, le agregáramos datos extremos, altos veríamos como cesaría su estabilidad o balanceo; la curva se colearía o inclinaría y observaríamos deslizarse, hacia la derecha, a la Media; Lo mismo vale para una distribución negativamente cargada, la mayoría de los puntajes tendrían valores altos y unos cuantos estarían en el extremo izquierdo, esto es, puntajes extremos con valores pequeños, en este caso la Media, se deslizaría a la izquierda. Lo importante a rescatar es

que el uso de la media con puntajes que no se distribuyen normalmente o aproximadamente normal, producirá valores promedio espuriamente altos o bajos, en suma, valores medios alterados. Veamos un ejemplo sencillo y rápido en el que la variable es la edad de los sujetos de una pequeña muestra :

15		1		
17			<---	Media
		25		
17	< --- Med y Mo.	27	<---	Med y
		Modo		
25		27		
	< --- Media	30		
60		110		
134				

$$\bar{x} = 134 / 5 = 26.8 \quad 110 / 5 = 22$$

Puede verse, que en las dos condiciones los promedios no son representativos de sus distribuciones, en la primera, distribución positivamente coleada, la media es falsamente más alta, por el jaloneo del puntaje extremo (60 años) mientras que en la negativamente coleada, el promedio es espuriamente más bajo. por el puntaje extremo (1 año). Bajo estas condiciones los valores de la Mediana y del Modo serían promedios más representativos y por lo tanto los más adecuados.

### **DESVIACIÓN ESTÁNDAR (S) Y VARIANZA (S<sup>2</sup>)**

1. Como ya se señaló, S es la medida de variabilidad más precisa, representa un índice de variabilidad en unidades de la medición original. La varianza (S<sup>2</sup>) también es una medida de variabilidad con el mismo significado que el de la desviación estándar, con la única diferencia que es un índice obtenido por una transformación de S en la que cada uno de los valores de la desviación se eleva al cuadrado. Así en lugar de ser un índice en unidades originales de medición como la S, es un índice en unidades cuadradas, no aporta más información u otra información diferente a la de la desviación estándar. La varianza representa también un valor sumario que no es otra cosa que un promedio o media de variabilidad (Media Cuadrática o Cuadrada) que se obtiene al dividir: la suma de los valores de desviación elevados al cuadrado entre N (número de casos o de puntajes en la distribución) o entre N - 1 (grados de libertad o factor de corrección). Cuando se utilizan grados de libertad (gl) es porque la muestra se utilizará como punto de

estimación para hacer inferencias a la población, lo que corresponde a una estimación no-parcial (unbiased, ver fórmula **a**). No se utilizan  $gl$  cuando los datos de la muestra sólo tendrán un uso descriptivo, esto es, no inferencial (Ver fórmula **b**); La fórmula **c.**, se conoce como la fórmula de desviación (pero también podría ser de la varianza, si se elimina la radical) de puntajes crudos y es la más frecuentemente utilizada.

Tenemos entonces que la varianza se define como Media Cuadrada o como promedio de las desviaciones cuadradas en torno a la media. En cambio, la desviación estándar es igual a la raíz cuadrada de la varianza de una distribución o sea igual a la raíz cuadrada de la suma de desviaciones elevadas al cuadrado y dividida entre el número de casos, o entre el número de casos menos uno.

---


$$a) S^2 = \frac{\sum x^2}{N} - 1 \quad b) S = \frac{\sum x^2}{N} \quad c) S = \frac{\sum X^2 - (\sum X)^2}{N - 1} \quad (1)$$


---

La  $S$  puede obtenerse a partir de puntajes de desviación que indican qué tanto se aleja cada puntaje crudo ( $X$ ) de la  $\bar{X}$  de su distribución:  $\bar{X} - X = x$ . Debe recordarse aquí, la importancia del uso de las notaciones o símbolos estadísticos, ya que de ello depende la comunicación entre los miembros de las comunidades científicas y profesionales relacionadas con la investigación. Usualmente, la notación de los puntajes crudos es una "X" mayúscula, para distinguirla de los puntajes de desviación que se representan con una "x" minúscula, la sola "x" expresa la diferencia arriba señalada, esto es:  $x = \bar{X} - X$ . Igualmente,  $S$  y  $S^2$  se utilizan para datos muestrales mientras que como valores de población son:  $\sigma$  y  $\sigma^2$  respectivamente; de la misma manera  $\bar{X}$  es la notación de la media muestral, sin embargo la media poblacional se denota como  $\mu$ .

Una de las propiedades de los puntajes de desviación es que:  $\sum x = 0$ . Si  $X$  representa una distribución de puntajes de edad, la desviación estándar y la varianza correspondientes, se obtienen como sigue:

1 (X)	2 (x)	3)
15	-11.8	
17	- 9.8	
17	- 9.8	
25	- 1.8	
60	33.2	
<hr/>	<hr/>	<hr/>
$\Sigma X=134$	$\Sigma x= 0$	$\Sigma x^2=1436.80$

$$S^2=1436.80/ 5 = 287.36 \quad S^2= 1436.80/ 4 = 359.20$$

$$\bar{X}=134/ 5 = 26.8 \quad S = 16.9 \quad S =$$

Puede verse como la columna 2, la de las desviaciones, tiene valores positivos y negativos y cómo la suma de las mismas resulta en cero. Se requiere elevar al cuadrado cada una de esas desviaciones para obtener la varianza. Al elevar cada desviación al cuadrado no se afecta la distribución, sólo se transforma a la dimensión cuadrática para solucionar el obstáculo de la sumatoria cero, y también para evitar trabajar con números negativos y positivos. La simple aplicación de la raíz cuadrada la convierte de inmediato a los puntajes originales. La columna 3, la de las desviaciones cuadradas, muestra por qué se le conoce como media cuadrada, ya que se suman las desviaciones cuadradas y se dividen entre el número de casos, y también muestra que media cuadrada y varianza son lo mismo. Por otra parte, nótese que la obtención de la varianza y de la desviación estándar se hizo con la misma distribución coleccionada utilizada arriba. ¿Cuáles son las consecuencias? Por ejemplo si tomamos el valor de  $S = 18.9$  decimos que las edades de los sujetos se dispersan, en torno a la media con una desviación estándar de 18.9 años, lo que resulta falso, sí se observa cómo se distribuyen las edades, nuevamente el puntaje extremo de 60 incrementa falsamente la dispersión de los puntajes en torno a la media. Por otra parte, lo que se interpreta es la desviación estándar, tiene más sentido decir que los puntajes tienen una desviación de 18 años que de 359.20 años ya que este último valor, la  $S^2$ , representa años al cuadrado.

### 5.1.2. Grados de Libertad

Los grados de libertad,  $N-1$ , son un factor corrector que se utiliza porque la varianza y/ o la media de la muestra, (según sea o sean los parámetros que se están estimando) representa

el resultado de sólo un subconjunto de observaciones, que servirá como punto de estimación del parámetro de la población, esto es, cómo si se hubiera tomado el conjunto total de observaciones. La función correctora de los grados de libertad controla la intervención del azar imponiendo restricciones a la decisión de significancia o no significancia de los valores muestrales. A menos grados de libertad más altos deben ser los valores muestrales para ser significativos, consecuentemente, a más grados de libertad (que implica una muestra más grande) menores serán los valores de muestra requeridos para que sean significativos.

Se tiene entonces que en el proceso de inferencia de parámetros se utilizan factores correctores como los grados de libertad y como regla general siempre se pierde un grado de libertad por parámetro estimado. En términos técnicos los grados de libertad se traducen operativamente como sigue: todos los puntajes que componen una distribución tienen libertad de variar al azar, menos uno de ellos, cuyo valor queda determinado por el valor del parámetro que se estima, por ejemplo, por el valor de la media de la muestra:

DISTRIBUCIÓN	MEDIA
a) 3 + 4 + 2 + 1 + 5	= 15/ 5 = 3
b) 1 + 3 + 4 + 1 + (?)	
c) 4 + 1 + 2 + (?) + 3	
d) (?) + 5 + 1 + 3 + 4	

En el ejemplo se tiene una distribución (a) con 5 puntajes con una  $\bar{x} = 3$ ; en los incisos b, c, y d, se muestra que todos los puntajes pueden tomar diferentes valores (los que se dieron y muchos otros más que podrían darse) menos uno, los que aparecen con una interrogación: en b, necesariamente el valor tendría que ser 6, para que la media resulte igual a 3; en c y en d, por las mismas razones, los valores tendrían que ser, 5 y 2 respectivamente.

## **5.2. CURVA NORMAL Y SUS AREAS.**

La curva normal es un concepto teórico matemático que se define por 2 parámetros; la MEDIA ( $\mu$ ) y la DESVIACIÓN ESTÁNDAR ( $\sigma$  ó S), mismos que adquieren significado conjuntamente y, sin embargo, son independientes. Es un concepto central para la estadística paramétrica, en tanto, se utiliza como modelo de la realidad y sus propiedades son fundamentales para hacer inferencias estadísticas.

Como ya lo habíamos señalado los parámetros característicos y que definen la curva normal son, la media y la desviación estándar, misma que se define como:  $N(\mu)$  en donde  $N$  significa normal. El mero conocimiento de los valores de estos dos estadísticos permite construir la distribución normal. También se ha dicho que, aunque su significado depende de los dos, ya que poco sentido tiene una desviación sin la media o de ésta sin la desviación, son, no obstante, independientes: dos o más distribuciones pueden tener la misma media pero diferentes desviaciones o viceversa. Una distribución normal también es útil para propósitos meramente descriptivos. Por ejemplo, se puede especificar qué proporción del área se puede encontrar en cualquier región de la curva. Debe agregarse además que existen variables biológicas, como, concentraciones de diversas sustancias en sangre, el peso corporal, la talla, la presión sanguínea, o psicológicas, como el coeficiente intelectual, o demográficas, como la edad, que se distribuyen normal o aproximadamente normal; aunque las distribuciones, también pueden cambiar a través de las generaciones, o de las culturas. Por ejemplo, en la actualidad, en países como México, la edad como parámetro poblacional, no se distribuye normalmente, ya que la proporción mayor se encuentra entre los jóvenes, mientras que, estos, en otros países, por ejemplo en los europeos son los menos. Asimismo, las normas del coeficiente intelectual van cambiando, por lo general, incrementándose a lo largo de las generaciones. También existen variables coleadas o asimétricas en mayor o menor medida, por ejemplo, escolaridad o el ingreso económico: grandes proporciones de individuos de diversos y diferentes países tienen ingresos mínimos, mientras que proporciones muy pequeñas obtienen ingresos muy altos. Es importante hacer hincapié, también, en que, es con el uso del análisis paramétrico o inferencial en donde se requiere la normalidad de las distribuciones, misma que sólo puede ser aproximada, especialmente con algunos tipos de análisis "fuertes o robustos" como se verá más adelante, pero también, a través de diferentes procedimientos denominados transformaciones se pueden obtener distribuciones más o menos normales a partir de distribuciones originalmente no normales, como lo veremos en este mismo capítulo.

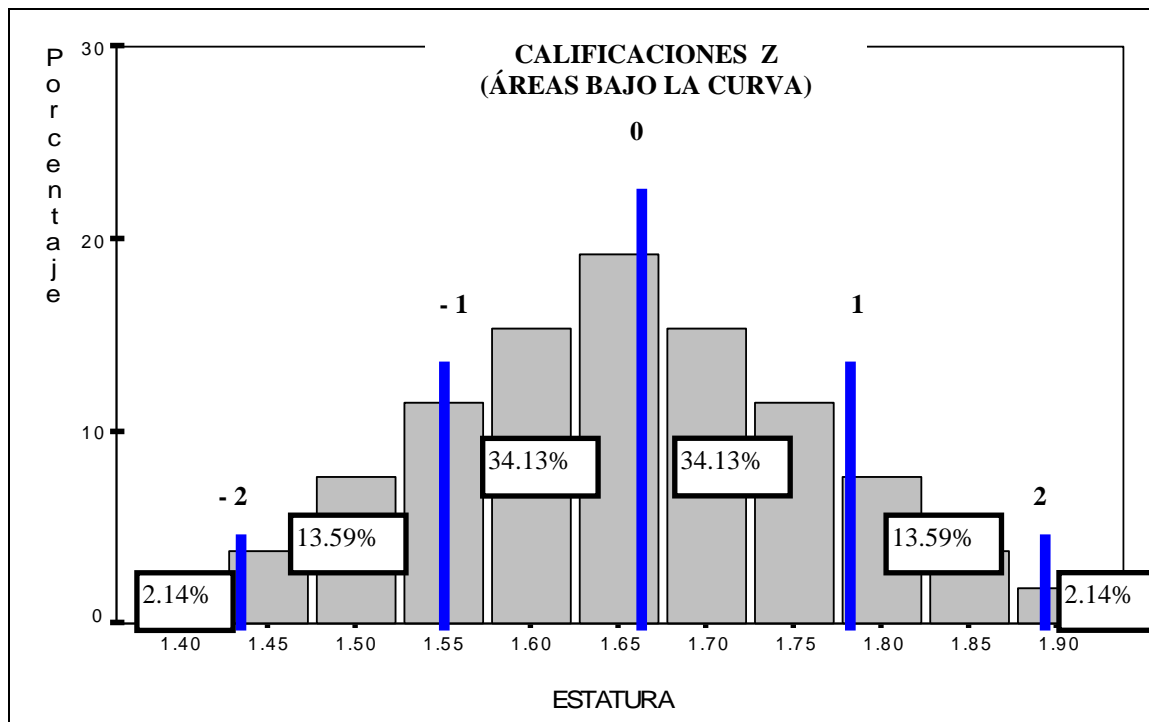


Figura 5. 4. Areas bajo la curva.

Ahora bien, la curva normal, que también se define como un polígono de frecuencias en forma de campana tiene un valor total igual a 1. Es simétrica con una media igual a 0 (es el promedio que se obtiene de los valores positivos y negativos de la abscisa) y con partes iguales con valor de .5 por abajo y por arriba de la media. Las áreas bajo la curva normal tienen las magnitudes que se muestran en la figura 4.

Puede verse que aproximadamente el 68% de los puntajes caen como sigue: Media  $\pm$  1 ; el 95%, Media  $\pm$  2 y el 99%, Media  $\pm$  3 ; el número de desviaciones estándar positivas y negativas podrían crecer al infinito, sin embargo, por cuestiones prácticas se acordó hasta 3 sigmas a cada lado de la media, ya que, las colas de la curva normal se extienden al infinito sin que nunca toquen la abscisa, de aquí, que se diga que son asintóticas.

Supongamos que la media de la distribución teórica del Coeficiente Intelectual (CI) de una determinada población es,  $\bar{X} = 100$  y la  $\sigma = 10$ . Conociendo estos valores puede afirmarse de inmediato que el 68% de todos los sujetos que respondieron la prueba tienen un CI que cae dentro de un rango que va de 90 a 110, esto es, que se distribuye a lo largo de una desviación por arriba y por abajo de la  $\bar{X} = 100$ .

### 5.2.1 Puntajes Z

También podríamos estar interesados en conocer qué porcentaje de los sujetos tendrían un CI entre 74.35 y 125.65.

En este caso se tiene que recurrir a los puntajes estandarizados Z mismos que forman parte de las Tablas de las áreas de la curva normal que se consultan para responder cuáles son los porcentajes de los sujetos que tienen ciertas características. En otros términos la curva normal puede considerarse, una curva normal estandarizada, que como tal se define:  $N(0, 1)$  que significa que tiene una media igual a cero y Desviación Estándar (DE) igual a uno, al igual que los puntajes Z. Así cualquier puntaje proveniente de una distribución normal puede transformarse en un puntaje Z ubicable en un punto de la abscisa de una curva normal estandarizada, de acuerdo con la fórmula que sigue:

$$Z = \frac{x - \bar{x}}{\sigma} = \frac{x}{\sigma} \quad (2)$$

Como puede verse el numerador de la fórmula Z es un puntaje de desviación, tal y como aparece a la derecha del signo igual (x) y el denominador es la desviación estándar cuyo valor sabemos, qué es, según el ejemplo que estamos revisando, igual a 10. Resolviendo la fórmula tenemos:

$$\begin{aligned} Z &= \frac{74.35 - 100}{10} = -25.65 \\ &= \frac{125.65 - 100}{10} = 25.65 \\ x/\sigma &= 25.65 / 10 = 2.56 \end{aligned}$$

El puntaje Z resultante fue de 2.56 y se busca en la tabla correspondiente qué porcentaje del área de la curva normal le corresponde. Encontramos que:  $Z = 2.56$  abarca un 49.48% de una mitad, pero como la curva es simétrica también abarca el 49.48% de la otra mitad, haciendo un total de 98.96%. La respuesta es, entonces, que un 98.96% de los sujetos de la muestra tendrán un CI que va de 74.35 a 125.65. De la misma manera sabemos que, puesto que el valor total de la curva es igual a 1, entonces un 0.52% de los sujetos tendrán un CI  $> 125.65$  y otro 0.52%  $< 74.35$ . Ahora bien ¿qué porcentaje de casos cae por abajo de 1, o de un CI de 110?. La respuesta es el 84%: 50 (del valor de la mitad de la curva) + 34 (el valor del área correspondiente a una desviación estándar por arriba de la media). ¿Qué porcentaje de casos caen por abajo de una  $Z = 1.96$  y de  $Z = 2.58$ ? En la tabla de las áreas bajo la curva normal encontramos que el porcentaje correspondiente a  $Z = 1.96$  es 47.5% y a  $Z = 2.58$  es 49.5%, a las que se les suma el porcentaje de la mitad de la curva por abajo de la media: 47.5% + 50% = 97.5%; y 50% + 49.5% = 99. Los valores clave en tanto son los más frecuentemente utilizados, como veremos más adelante, y que conviene memorizar son:



$$\bar{X} \pm \quad = 68\%$$

$$\bar{X} \pm 1.96 \quad = 95\%$$

$$\bar{X} \pm 2.58 \quad = 99\%$$

Los puntajes Z son, como ya se señaló, puntajes estandarizados con media = 0 y desviación estándar = 1. Además de traducir la desviación estándar en Z para efectos de trabajar con las áreas de la curva normal, una de sus aplicaciones de igual o mayor utilidad es la transformación de puntajes originales obtenidos con diferentes unidades de medición (por ejemplo, edad en años cronológicos, ingreso en miles de pesos, escolaridad en años de estudio etc.) en la misma unidad, es decir medirlos con el mismo "metro", con media 0 y desviación estándar de 1, para cumplir con un principio indispensable para finalmente hacer comparaciones relativas entre las diferentes distribuciones, variables o puntajes y que puedan interpretarse con sentido. Sólo con esta transformación podría alguien responder si es "más alto que pesado" (ejemplo citado por varios autores, entre otros, Young & Veldmann, 1979 ; Rosenthal y Rosnow, 1991; y retomado aquí por considerarlo altamente descriptivo , en el señalamiento del problema, esto es, hace evidente cómo se están mezclando dos diferentes unidades de medición: kilos con centímetros. Debemos recordar nuestras primeras lecciones en las que se nos enseñó que ¡no se podían sumar naranjas con manzanas! Es claro que existen diferentes "metros" de igualación o de estandarización de diferentes unidades de medición, por ejemplo, los puntajes T que tienen una media igual a 50 y desviación estándar igual a 10, entre muchos otros procedimientos de estandarización. También es importante hacer notar que para efectos de convertir diferentes unidades de medición en una común no es requisito necesario que las variables se distribuyan normalmente (Ver Rosenthal y Rosnow, 1991), cuestión indispensable sí lo que se busca son proporciones de puntajes que se encontrarán por arriba o por abajo de algún nivel previamente determinado.

### **5.2.2 Error Estándar de la Media**

El error estándar de la media es una medida de variabilidad de las medias muestrales. Se fundamenta en un concepto teórico que asume que si se tomaran todas las posibles muestras del mismo tamaño de una población y se obtuviera la  $\bar{X}$  de cada una, y se graficarán los valores de éstas, se obtendría una distribución normal en la que la  $\bar{X}$  de las  $\bar{X}$ s sería =  $\mu$ . Esto es, la media de las medias muestrales tendría el mismo valor que la media de la población. Cada  $\bar{X}$  podría tener diferentes valores unos más próximos y otros más alejados de la M, pero la  $\bar{X}$ s / N( $\bar{X}$ s) = M. A partir de este concepto se puede estimar

el Error Estándar (EEst.) que se cometerá en la estimación de la M de la población a partir de la media muestral, sin embargo es requisito que esa media muestral provenga de una muestra aleatoria, si esto se cumple, no importa que el tamaño de la muestra sea relativamente pequeño. En otros términos, independientemente del tamaño de la población el cálculo de la  $\bar{X}$  aún con una muestra pequeña será bastante exacto si la muestra es aleatoria.

Resumiendo: Si la muestra es aleatoria se asume:

a) que la  $\bar{X}$  es una  $\bar{X}$  seleccionada de una población hipotética de miles de medias. Por ejemplo, si el universo o población estuviera compuesto de sólo 30 casos, todas las posibles muestras de 10 observaciones que se podrían obtener serían nada menos que ; más de 30 millones

$$= {}^{30}C_{10} = \frac{30!}{10! (30 - 10)!} = 30,045,015$$

b) que la distribución de esa  $\bar{X}$ s de muestras aleatorias será normal;

c) que la M de la población se puede calcular con gran exactitud a partir de una muestra de esa población; y

d) que lo que se requiere es calcular la probabilidad del error de la  $\bar{X}$  con el EEst.

Como puede verse, la notación estadística del EEst. =  $\sigma_{\bar{x}}$ , nos indica que es una medida de variabilidad o de dispersión por la presencia de la "  $\bar{x}$  " pero esta es una desviación estándar de una distribución de medias de muestra ( $\bar{X}$ s) (como lo indica la  $\bar{X}$  que le sigue) mide la dispersión de  $\bar{X}$ s en torno a la  $\bar{X}$  de  $\bar{X}$ s, mientras que la "s" es una medida de dispersión de puntajes crudos (X) en torno a la media. La fórmula del EEst., es la siguiente:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}} \quad (3)$$

Los factores que influyen en el EEst son:

a) la  $\sigma$  de la población a  $\sigma > \sigma_{\bar{x}}$ ; y

b) la N de la población: a  $N < \sigma_{\bar{x}}$ .

Por otra parte, si las varianzas ( $S^2$ ) también se distribuyen normalmente, esto implica que la desviación estándar también tiene su propio error estándar, mismo que se representa como

$$\sigma_{\bar{s}} = \frac{S}{\sqrt{2N}}$$

El error estándar de la desviación tiene las mismas propiedades, que el error estándar de la media. De hecho la media y la desviación estándar deben ir acompañadas de sus respectivos errores estándar (  $\bar{X} \pm y$  y  $S \pm$  ).

### 5. 2.3 Prueba Z e Intervalos de Confianza

La prueba Z es una prueba para una sola muestra. Su fórmula es la siguiente:

$$Z = \frac{x - \bar{x}}{\sigma_{\bar{x}}} \quad (4)$$

Supongamos que consistentemente se ha encontrado que la media teórica del CI de una población determinada era  $M = 100$  y la Desv. Est. = 13. Al encontrar en una muestra aleatoria ( $N = 24$ ) de otra población una  $\bar{X} = 107$ , el investigador se propuso probar si habría, con un nivel 0.01, diferencias estadísticamente significativas entre esos valores de medias. Entonces, la hipótesis nula a probar fue:  $H_0. \bar{X} = M = 100$ .

De acuerdo con la fórmula (3) :

$$\frac{13}{\sqrt{24}} = 2.65$$

Siguiendo la fórmula (4) :  $107 - 100 / 2.65 = 2.64$ .

Ahora bien, una  $Z = 2.64$  es  $>$  que una  $Z = 2.58$  y sabemos que ésta abarca el 99% de la curva normal dejando fuera el .01%. Si lo que queda fuera se toma como el área crítica o de intervención del mero azar y el nivel de significancia se estableció al 0.01 , sería suficiente con que la Z resultante fuera de 2.58. Sin embargo fue mayor, por lo que sin buscar en tablas podemos afirmar con toda seguridad, que una  $Z = 2.64$  tiene una zona crítica,  $<0.01$ , esto es, una probabilidad (p) menor de una en 100 de que la diferencia entre  $\bar{X}$  y M ocurriría sólo por azar. En otros términos, la probabilidad de que la  $H_0$ . sea verdadera sólo por azar es  $p < .01$  razón por la cual la decisión estadística es rechazar la  $H_0$  y se concluye con una probabilidad  $> 99\%$  que  $M = 107$ .

Por otra parte, debe tenerse en cuenta que si se repitieran estas mediciones con diferentes muestras de la misma población, difícilmente encontraríamos otra vez el mismo valor de  $\bar{X} = 107$  por lo que la mejor opción es estimar un rango de

valores, para el cual se calcula la probabilidad con la que se esperaría encontrar, la M, dentro de ese rango. Las expresiones o fórmulas estadísticas más simples del intervalo de confianza y que producen los mismos resultados son:

$$a) \bar{X} \pm (\text{prob.}) (\sigma_{\bar{X}})$$

$$b) \bar{X} \pm (\text{prob.}) \frac{\sigma}{\sqrt{N}}$$

$$\text{De acuerdo con } \mathbf{a}: 107 - 1.96 (2.65) = 101.81$$

$$107 + 1.96 (2.65) = 112.19$$

$$\text{Intervalo de Confianza: } 101.81 < M < 112.19$$

Si la probabilidad. se define a 1.96 o a 2.58, se establece que la probabilidad de que el valor real de la M quede fuera de estos límites (101.81 a 112.19) es igual o menor que el .05 ( en el caso de Z = 1.96) o bien, puede decirse, que existe una probabilidad del 95% o más de que la verdadera media de la población caiga dentro de dichos límites; mientras que la probabilidad sería menor que el .01 de que la verdadera Media cayera fuera de esos límites o que el 99% cayera dentro, en el caso de Z= 2.58. Es importante hacer notar que el valor teórico de la Media igual a 100, no quedó incluido en el rango por lo que se rechaza con mayor seguridad que M = 100 sea el verdadero valor de la Media.

Si se tiene una M= 17;  $\bar{X} = 15$ ; N = 30 y  $\sigma = 5$ . De acuerdo con la fórmula **b**:

$$\frac{\sigma}{\sqrt{N}} = 5 / 5.47 = 0.91$$

$$15 \pm 1.96 (0.91) (25)$$

$$15 \pm 1.96 (0.91) =$$

$$15 \pm 1.78 =$$

$$\text{Intervalo al 95\% de confianza: } 13.22 < M < 16.78$$

La probabilidad de que la media real quede fuera de los límites del intervalo (13.22 a 16.78) es menor que 5 veces en 100, o bien, existe una probabilidad del 95 veces en 100 de que la media de la población se encuentre dentro de dicho intervalo. Nuevamente, puede verse que la M = 17 no quedó incluida dentro del intervalo por lo que existe mayor seguridad de que el verdadero valor de la media es igual a quince.

La prueba Z para una sola muestra se utiliza cuando los puntajes se distribuyen normalmente, cuando se conoce la media de la población o puede asumirse un valor teórico de la misma y además, también se conoce el valor de la DE.

### 5.2.3 Prueba t para una sola Muestra

Aunque la aplicación de la prueba t para una sola muestra requiere también de puntajes con distribución normal y el conocimiento del valor de la media de la población, no obstante, representa una alternativa con menores exigencias al uso de la prueba Z, en tanto, no se requiere el conocimiento del valor de la  $\sigma$  de la población y se puede utilizar en su

lugar, la S de la muestra (Fórmula ):  $S = \frac{S}{\sqrt{N}}$

$$t = \frac{\bar{x} - M}{S}$$

El valor de t requerido para que sea significativo al nivel de 0.05 se aproxima al 1.96 de la prueba Z, a más grande es la N. En general los mismos valores de la prueba Z de 1.96 y de 2.58 son significativos a los niveles de 0.05 y de 0.01 respectivamente a mayor sea N. Sin embargo, Lathrop (1969), entre otros investigadores, considera que basta que los grados de libertad sean  $>30$  para que se puedan utilizar dichos valores (1.96 y 2.58) como valores significativos a los niveles ya conocidos de 0.05 y de 0.01. La forma tradicional de la decisión estadística con respecto a la hipótesis nula es buscar en la tabla t con N-1 grados de libertad el valor crítico al nivel de significancia establecido.

#### Supuestos Básicos de la Pruebas z y t de una sola Muestra

1. Los sujetos de la muestra son seleccionados al azar;
2. Los puntajes provienen de una distribución normal; y
3. Las variables deben ser cuantitativas continuas y medidas por lo menos intervalarmente.

Consideraciones adicionales: la prueba Z y t para una sola muestra requieren del conocimiento del valor de la media de la población y de la DE en el caso de la primera mientras que en el caso de la segunda basta con la S de la muestra. Es aconsejable una N grande porque asegura un decremento del error estándar (aunque debe recordarse que estamos hablando de la raíz cuadrada de N por lo que estaríamos refiriéndonos a incrementos más grandes de la misma) además de propiciar la normalidad de la distribución.

### 5. 3 REVISANDO LOS DATOS: DETECCIÓN DE ERRORES Y CUMPLIMIENTO DE LOS SUPUESTOS.

Una vez que el investigador ha recogido sus datos y los ha codificado, lo primero que deberá hacer es revisar cuidadosamente sus datos antes de aplicar cualquier análisis

ya que, una parte importante del éxito de éste depende de esta primera tarea. que entre las varios propósitos, los principales son: a) detectar posibles errores de codificación o errores de "dedo", por ejemplo, códigos inválidos, como escribir un 3 en la variable sexo cuando 1 es para femenino y 2 para masculino. En este tipo de errores, algunos puede ser fácilmente detectados y corregidos, mientras que habrá otros que requerirán mayor esfuerzo y posiblemente algunos sean indetectables, como por ejemplo, capturar 25 años en lugar de 26; y b) observar el comportamiento de los datos de las distribuciones para tomar decisiones de eliminación de algunos de ellos, de adquisición de otros, de transformación de distribuciones, si es el caso, y de ejecución de los análisis más pertinentes. Esta tarea previa es insoslayable sin importar la sencillez o complejidad del problema que se investigue, ya que un análisis estadístico adecuado no sólo optimiza la información que puede derivarse, sino que su nivel de adecuación o de inadecuación contribuye a la validez o invalidez del mismo diseño de investigación.

Entre los comportamientos a observar están:

1). **Ausencia inesperada de información.** Por ejemplo, podría encontrarse que una o algunas de las categorías, rangos, o intervalos de la distribución aparecen vacíos; si no es por error mecánico y es porque no hubo respuestas, habría que averiguar la razón (por ejemplo, ver el tipo de preguntas, repilotear las mismas, reformularlas, agregar preguntas cedazo y redirigir la pregunta sólo a aquellos que tengan las precurrentes necesarias para contestarlas, entre otras posibles estrategias);

2). **Variabilidad inusual..** Desviaciones excesivamente grandes debidas a grandes dispersiones entre los puntajes y con respecto a la media lo que llevaría a revisar lo adecuado del tamaño de la muestra, pequeña y asimétrica o no normal podría ser una razón, pero también podría ser un rango muy amplio, esto es una gran diferencia entre el valor más grande y el más pequeño de la distribución, lo que aconsejaría eliminar pequeños porcentajes de valores a ambos extremos de la misma y utilizar medias "trimmed", o bien añadir categorías intermedias y aumentar el número de sujetos; también podría encontrarse mínima o cero variabilidad; si todos los sujetos reponden de la misma manera a una pregunta aparte de que no aportaría mayor información sería inútil también desde el punto de vista estadístico, habría que revisar si la pregunta podría ser una de conocimiento, si es obvia en la dirección de deseabilidad social, observar si podría ser una respuesta reactiva o defensiva. Por ejemplo este puede ser un problema frecuente en las categorías intermedias de respuesta, como sucede con la tercera categoría de cinco de los reactivos tipo Likert : " ni de acuerdo ni en desacuerdo" o bien "indecisa" " no estoy segura" y similares. Lo aconsejable es utilizar

números pares de categorías, por ejemplo, 6 en lugar de 5 , y evitar tales expresiones categoriales; también se ha logrado disminuir este problema solicitando a los sujetos que no utilicen para responder dicha categoría a menos que sea indispensable y realmente sea esa su respuesta. Seguramente esta advertencia tiene el efecto de que el encuestado ponga más cuidado y atención en sus respuestas.

3). **Patrones de respuesta.** Si un entrevistado tiene poco interés en participar o se ve comprometido a hacerlo, o no están claras las instrucciones y no se entiende la tarea, está cansado o le parece muy largo y aburrido el cuestionario, etc. todas estas condiciones pueden reflejarse en patrones de respuesta característicos: por ejemplo todas las respuestas en una misma ubicación, en la parte extrema derecha de los reactivos, independientemente del contenido de los reactivos; o siguiendo un patrón simétrico, unas cruces en la parte extrema derecha, otras en el centro, otras en la izquierda y vuelta al inicio; es aconsejable: eliminar todo aquel cuestionario que muestre un 20% o más de respuestas en las mismas categorías; incluir preguntas clave que pueden asumir muy diferentes formas, por ejemplo algunas se formulan como repetición de otras, parafraseadas de otra manera y distantemente colocadas, si se encuentran respuestas contradictorias es señal del poco cuidado o interés del encuestado lo que debe llevar a cuestionar la utilidad del mismo;

4). **Normalidad de las distribuciones.** Dentro de la observación del comportamiento de los datos, el de la normalidad es uno de los más importantes cuando se pretenden utilizar análisis estadísticos más potentes, para la prueba de hipótesis, para propósitos inferenciales y de generalización. Se cuenta con diferentes estrategias y procedimientos, numéricos y gráficos, para efectuar tal observación, entre ellos tenemos:

a). Una tabla de **Frecuencias** para los datos de cada variable. Como puede verse en la Tabla 5.1 se despliegan diferentes columnas: en las que aparecen los **valores** o códigos dados a cada categoría; las frecuencias o número de casos que cayeron en cada categoría, en la que debe checarsé que la suma de las frecuencias = N o tamaño de la muestra; el porcentaje de cada frecuencia (1) teniendo como base N o el total de los casos; el porcentaje que corresponde al porcentaje de cada categoría teniendo como base el total de respuestas dadas(2) que puede ser igual a N si no hay respuestas faltantes o en caso contrario, diferente a N ; y los porcentajes acumulados (3) que corresponden a la suma de los porcentajes de respuestas dadas.



Tabla 5.1. Distribución de Frecuencias de la variable estatura.

		P O R C E N T A J E S		
valores	frecue ncias	(1)	(2)	(3)
1.40	6	1.9	1.9	1.9
1.45	12	3.8	3.8	5.8
1.50	24	7.7	7.7	13.5
1.55	36	11.5	11.5	25.0
1.60	48	15.4	15.4	40.4
1.65	60	19.2	19.2	59.6
1.70	48	15.4	15.4	75.0
1.75	36	11.5	11.5	86.5
1.80	24	7.7	7.7	94.2
1.85	12	3.8	3.8	98.1
1.90	6	1.9	1.9	100.0
-----	-----	-----		
Total	312	100	100.0	

Media	1.650	Mediana	1.650
Modo	1.650	Desviación standard	.110
Kurto sis	-.383	Skewness	.000
Rango	.500	Mínimo	1.400
Máxim	1.900	Varianza	.012

o

- (1) Porcentajes de respuesta total (Base = N)
- (2) Porcentaje de respuestas (total de respuestas dadas)
- (3) Porcentaje Acumulado

En la parte inferior de la tabla aparecen algunas de las estadísticas más frecuentemente utilizadas. Para efectos de checar normalidad conviene revisar los valores de las 3 medidas de tendencia central, valores iguales o próximos nos hablan de distribución normal, lo mismo que desviaciones no muy grandes; sin embargo, los indicadores directamente relacionados con la normalidad son la **simetría** (skewness = **Sk.**) y la **kurtosis** = **Ks.** (grado de aplastamiento de la curva) aunque algunos autores señalan que este último estadístico no es de utilidad dentro del campo de las ciencias sociales y de la conducta (v.gr., Montemayor, 1973). Si estos valores, **Sk.**, y **Ks.** son igual a cero la distribución es perfectamente normal, sin embargo, la normalidad perfecta es ideal, en la realidad no existe. Debe señalarse, nuevamente, que para la mayoría de los análisis, basta que la distribución sea aproximadamente normal y debe recordarse que siempre es recomendable una N grande porque esta propicia la normalidad.

¿A qué tamaño de muestra se le puede considerar grande? Existen dos formas de respuesta: una general a la que se ha llegado por consenso basado en la empiria o experiencia y otras que se responden, en cada caso, a través de la solución de fórmulas o tablas específicas. De acuerdo con la primera, una muestra grande, o adecuada, en el caso de investigaciones experimentales, de laboratorio o de campo, en las que se mantiene control de la varianza extraña o de la varianza de error, es suficiente con  $N_s \geq 30$ ; en investigaciones no experimentales, por ejemplo, correlacionales o ex- postfacto, se considera que una muestra es grande con una  $N \geq 200$ . Por otra parte, ciertos análisis determinan un número de sujetos por cada reactivo para asegurar la variabilidad requerida según el tipo de análisis (v.gr., el Diseño Factorial y el Análisis Factorial, entre otros). Otras formas de respuesta al tamaño de la muestra se revisaron en el Capítulo 3.

Por otra parte un valor positivo de  $Sk$ . indica que la distribución está positivamente coleada mientras que valores negativos señalan distribuciones negativamente cargadas. Un valor positivo en kurtosis señala una distribución más puntiaguda que la normal. En términos generales se considera que una distribución es semejante a la curva normal si el valor de su asimetría calculada resulta dentro del rango:  **$Sk = -0.5$  a  $0.5$  y el de  $Ks = 2$  a  $4$ .**

**$Sk$ . Y  $Ks$ .** se obtienen como sigue:

$$Sk = \frac{\bar{x} - Mo}{S} \quad (8)$$

Si la media y el modo son iguales las diferencia será igual a cero y como ya se señaló no habrá asimetría, en otros términos, la distribución es normal.

Se ha encontrado que si la asimetría es moderada se establece una relación entre las 3 medidas de tendencia central como puede verse en la fórmula 8 (primer Coeficiente de Asimetría de Pearson). La fórmula 9 (segundo Coeficiente de Asimetría de Pearson) produce los mismos resultados, sin utilizar el valor del modo:

$$Sk = \bar{x} - Mo + 3(\bar{x} - Med.) \quad (9) \qquad Sk = \frac{3(\bar{x} - med.)}{S} \quad (10)$$

En términos generales se considera que una distribución es semejante a la curva normal si el valor de su asimetría calculada resulta dentro del rango

$$- 0.5 \quad \text{a} \quad + 0.5$$

Otras fórmulas para series agrupadas de valores mediante el cálculo de momentos son las siguientes:

$$m_2 = \frac{f(x - \bar{x})^2}{n} \quad (11)$$

$$m_3 = \frac{f(x - \bar{x})^3}{n} \quad (11a) \quad m_4 = \frac{f(x - \bar{x})^4}{n} \quad (11b)$$

$$\text{asimetria } a_3 = \frac{m_3}{(m_2)^3} \quad (12a) \quad \text{kurtosis } a_4 = \frac{m_4}{(m_2)^2} \quad (12b)$$

b) Una gráfica de barras (histograma) permite visualizar rápidamente, qué tanto se alejan las distribuciones de la curva normal, cómo se agrupan los valores, a la izquierda, a la derecha, si hay uno o más picos (modos) y si hay valores extremos. Si se superpone a la gráfica de barras la curva normal se facilita, todavía más, la captación inmediata de la información (Ver Fig. 5.2). La función y ventaja de las gráficas es que sólo a través de la observación se puede responder al problema de la normalidad, y sólo en contados casos hay que recurrir, además, a las fórmulas. De hecho, se ha dicho que las gráficas, los grabados, o desplegados de los números, son más instructivos que los mismos números (Tufte 1983, citado por Rosenthal & Rosnow, 1991).

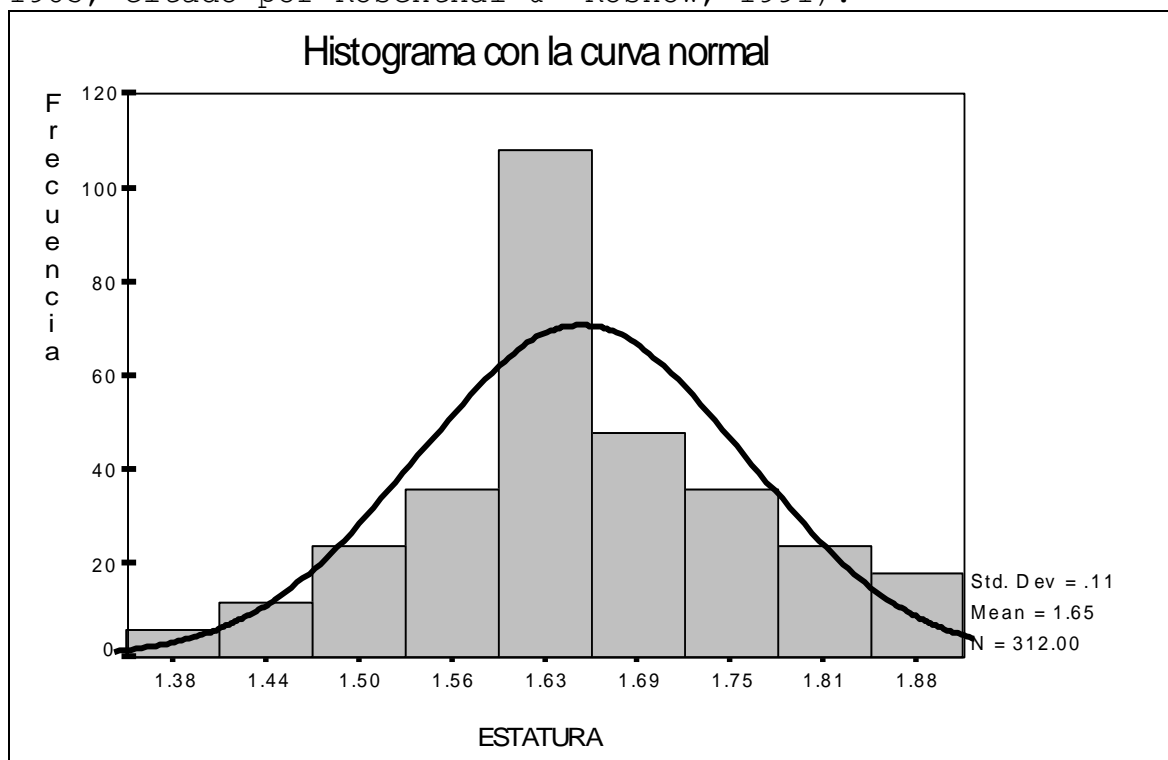


Figura 5.5. Histograma con la curva normal superpuesta.

c) Presentar los datos con **gráficas de caja** (Box Plots) desarrolladas por Tukey (1977), quien originalmente las

denominó "box-and-whisker". Es una de las representaciones visuales más ilustrativas o instructivas cuya utilidad sobresale con distribuciones grandes y cuando se requieren comparaciones.

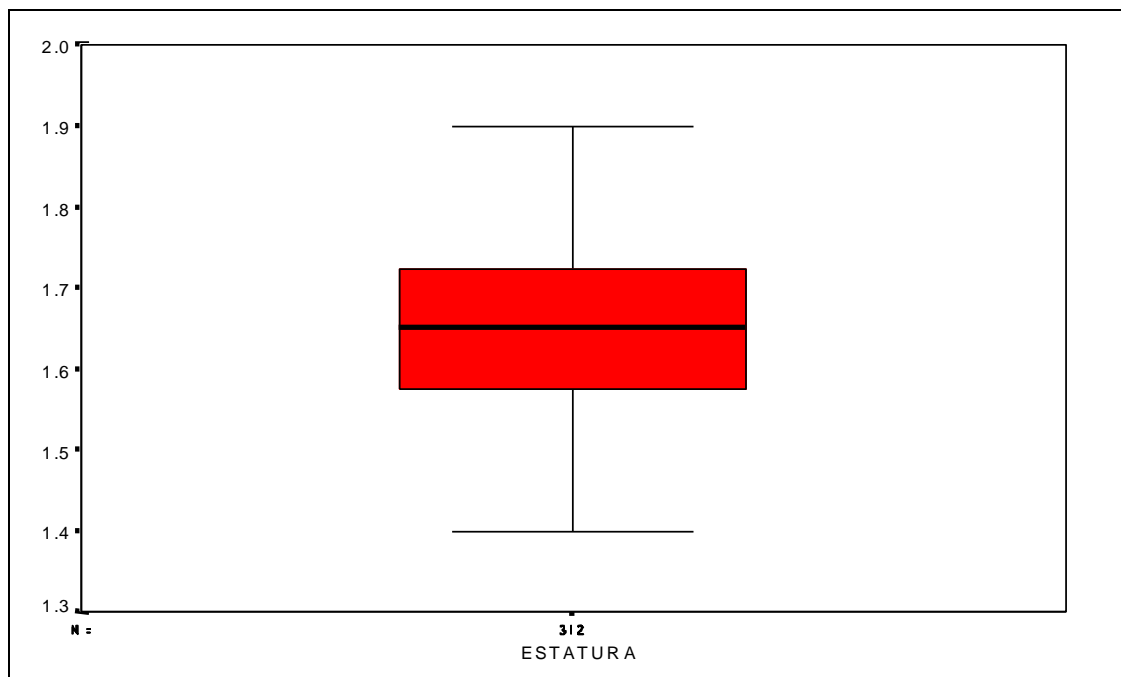


Figura 5.6. Gráfica de Caja representando una distribución normal

Con las gráfica de caja se despliegan estadísticas sumarias, no los valores crudos. Así, la división que aparece al interior de la caja representa la ubicación de la mediana (50%), la caja representa, en su límite superior el percentil 75% y en su límite inferior el percentil 25%; de esta manera el 50% de los casos tienen valores dentro de la caja. El cálculo de estos percentiles es diferente al tradicionalmente utilizado y se conocen como los "hinges" de Tukey.

El ancho de la caja nos indica la dispersión o variabilidad de los datos. También aparecen señalados en la gráfica de caja, valores que caen fuera de la misma, por arriba (valores altos) y por abajo (valores bajos): los más cercanos, los limitados por las líneas o colas fuera de la caja representan valores altos o bajos, pero no son extremos. Les siguen valores alejados, aunque no extremos, que se encuentran a 1.5 veces la longitud de la caja por arriba o por abajo de los límites de ésta y finalmente están los valores extremos que se alejan 3 veces la longitud de la caja, por arriba (valores altos) y por abajo (valores bajos) de los límites de ésta. Puede verse entonces que con esta información visual rápidamente puede decirse si hay normalidad que es cuando

el lugar de la división de la caja está en el centro. Si no es así, pueden aparecer en el fondo de la caja o en la parte superior de la misma; en el primer caso la distribución estará positivamente coleada, mientras que en el segundo, la distribución estará negativamente coleada; igualmente de inmediato se sabrá que tan grande es la variabilidad de los puntajes, según sea el tamaño de la caja.

Los **gráficas de caja** (Box Plot) son particularmente útiles para comparar la distribución de los valores en varios grupos o variables. Sólo como ejemplo, en la gráfica siguiente se presentan las cajas Box Plots de dos índices de masa corporal. El Índice de Masa Corporal (IMC) es un indicador obtenido a partir del peso y la estatura ( $\text{peso} / \text{estatura}^2$ ) que, entre otras cosas proporciona información sobre el peso corporal. Uno de estos índices fue calculado a partir de lo que los sujetos dijeron medir y pesar (llamado IMC estimado) y el otro a partir de lo que los sujetos medían y pesaban realmente (llamado IMC real). En la Fig. 5.3 podemos ver rápidamente que hay mayor variabilidad en los datos del IMC real (la caja es ligeramente más ancha, que la del IMC estimado) y que los sujetos tienden a reportarse más delgados de lo que son en realidad, la caja está ligeramente más abajo que la caja del IMC real), Sin embargo la distribución del IMC estimado se distribuye relativamente de manera "más" normal que la del IMC real. (ubicación de la rayas que dividen las cajas).

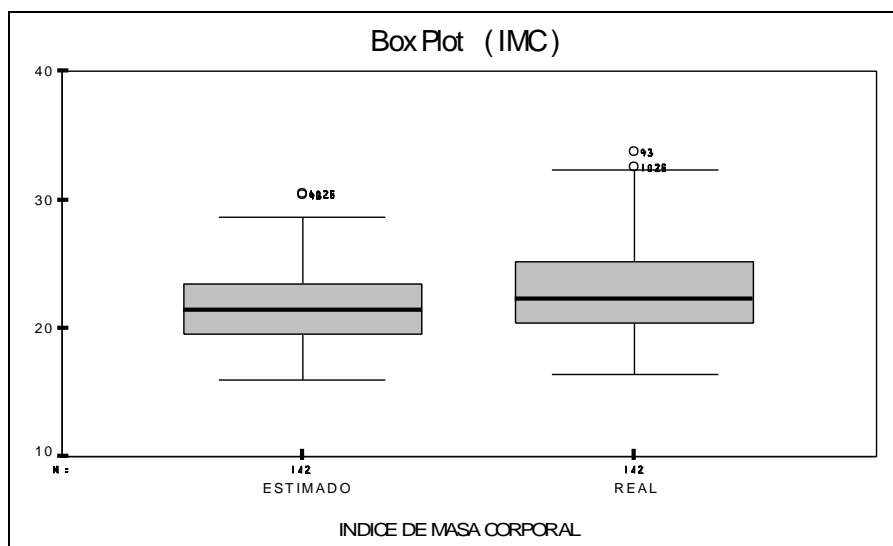


Figura 5.7. comparación de las distribuciones de dos variables con Gráficas de Caja.

d) Uso de estimadores fuertes o robustos de tendencia central. Sabemos, como lo señalamos al inicio de este Capítulo, que la media se ve significativamente afectada por los datos extremos. Por otra parte también habíamos señalado la ventaja de una medida como la mediana que no se ve afectada por los datos extremos, ya que depende de la parte principal o central de los datos por lo que se le considera una medida resistente en tanto su valor no se vería afectado por la no normalidad como tampoco por la adición o aumento no muy grande de datos, o bien por la sustracción o eliminación de los mismos.

Sin embargo, la mediana representa una medida simple e intuitiva (señala la ubicación de un punto) lo que puede mejorarse con una medida que utilice toda la información, por ejemplo la media, pero bajo supuestos no restrictivos sobre la distribución subyacente de los datos.

Uno de tales estimadores se puede obtener reduciendo (trimming) los datos para eliminar los valores que se encuentran alejados (extremos). Por ejemplo si se trabaja con una "media reducida al 15%" significa que se ha eliminado el 15% de los puntajes más bajos y el 15% de los puntajes más altos, lo que implica que la estimación se ha basado en el 70% de los datos centrales. Puede verse que una "media reducida utiliza más información de lo que haría la mediana, de hecho, puede decirse que la mediana es una media reducida al 50%.

Con la media reducida vimos que se eliminan los datos que se encuentran alejados del resto. Sin embargo, no siempre puede uno permitirse la pérdida de casos, o en otros términos, siempre es posible utilizar procedimientos menos drásticos. Por ejemplo recurriendo a otro tipo de estadísticos conocidos como Estimadores de Máxima Probabilidad (Nourisis,1992) que no es otra cosa que una media ponderada o pesada y que consiste en asignar pesos más pequeños a los valores alejados que a los casos que se encuentran más cercanos del centro de la distribución (en el procedimiento usual todos los valores de la distribución tendrían un peso igual a uno).

La regla general es asignar pesos en relación inversa a la distancia del centro de la distribución, esto es: menor peso a mayor distancia. Existen diferentes estimadores, entre ellos están los de Huber, Hampel, Tukey y Andrew (Ver Nourisis, 1992).

#### 5.4. TRANSFORMANDO LAS DISTRIBUCIONES.

En general no son serias las consecuencias cuando las suposiciones hechas por las pruebas paramétricas no se satisfacen estrictamente. Esto quiere decir que no debemos alarmarnos por variaciones moderadas a partir de las condiciones especificadas por tales suposiciones, ya que, por ejemplo, desviaciones menores de la normalidad y/o algunos grados de heterocedasticidad (falta de homogeneidad de las varianzas) sólo tendrán pequeños efectos en las pruebas comunes y en las inferencias resultantes siempre y cuando el tamaño de las muestras sea adecuado. De cualquier manera, es importante saber cómo tratar situaciones en las cuales los datos no se comportan de acuerdo con las suposiciones.

Cuando tratamos de hacer algo para lograr que nuestros datos se ajusten a las suposiciones paramétricas el procedimiento más usual es el de transformar los datos originales de modo que los datos transformados cumplan o cumplan mejor las condiciones especificadas.

Transformar los datos consiste en cambiar todos los valores de una variable de una forma similar y sistemática. Es decir vamos a realizar una misma operación matemática sobre todos y cada uno de los datos que tenemos en esa variable. Esa es la regla principal cuando se usan transformaciones: la transformación debe aplicarse a todos los datos que hay en la variable, no es válido que se transformen solo los datos de unos sujetos y de otros no.

Lo que hacemos cuando transformamos los datos es cambiar la escala que estamos utilizando. No inventamos ni creamos nuevos datos, simplemente estamos usando una "regla" diferente para medirlos. Los datos no cambian, lo que cambia es la forma de su distribución.

A continuación se presentan unos datos como ejemplo. Los números indican la distancia que un grupo de niños logro recorrer en un laberinto sin tocar las orillas. La primera columna presenta los datos originales (en pulgadas) y la segunda los mismos datos transformados (en centímetros)

Pulgadas	Centímetros
1	2.5
1.5	3.75
2.5	6.25
3	7.5
3.5	8.75
4.5	11.25
5	12.5
6	15
7	17

Podemos ver como al realizar la transformación los datos se distribuyen de una manera diferente, tal y como podemos ver en la siguiente gráfica.

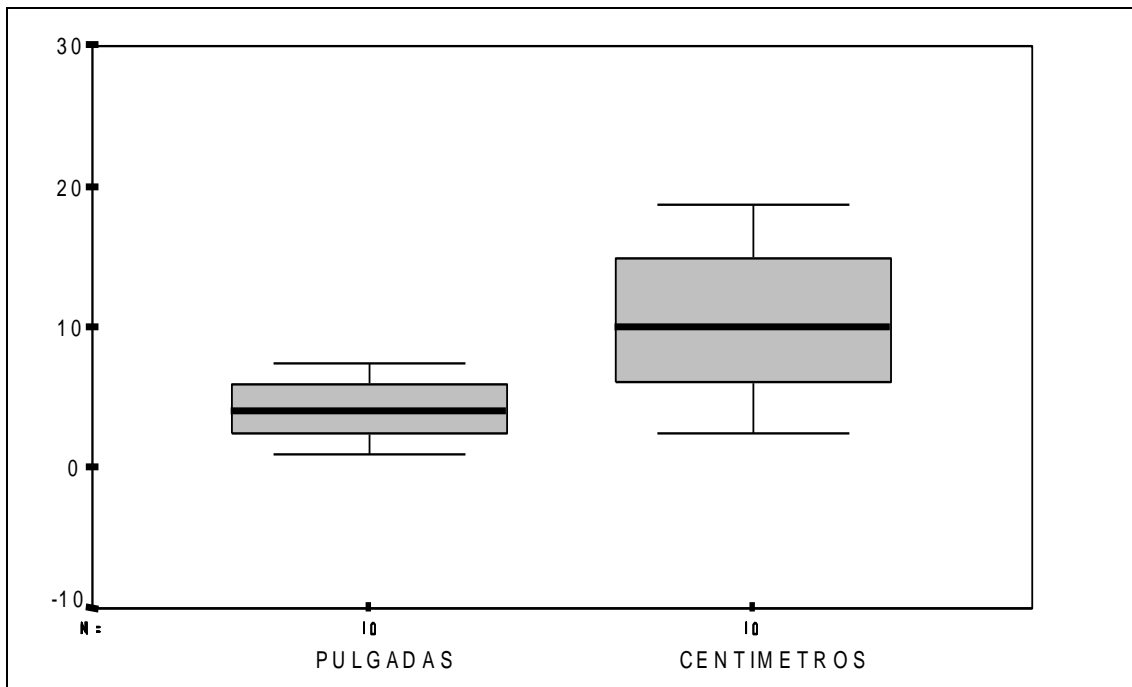


Figura 5.8. Gráficas de Caja para distribuciones de los mismos datos con diferentes escalas..

Es decir las transformaciones nos permiten trasladar nuestros datos a una escala diferente. Esto permite que la nueva distribución de los datos cumpla con los supuestos paramétricos. Afortunadamente, en la mayoría de los casos bastará una transformación para lograr que se cumpla con los supuestos. Es decir, regularmente no es necesario desarrollar una serie de transformaciones que corrijan diversas "deficiencias" de los datos. En general, el uso de una transformación para corregir una deficiencia particular (por ejemplo la falta de homogeneidad) también ayudará a corregir otra deficiencia (la no normalidad).

Aunque existen muchas posibles transformaciones, las más comunes son las expuestas en la tabla 5.2.

---



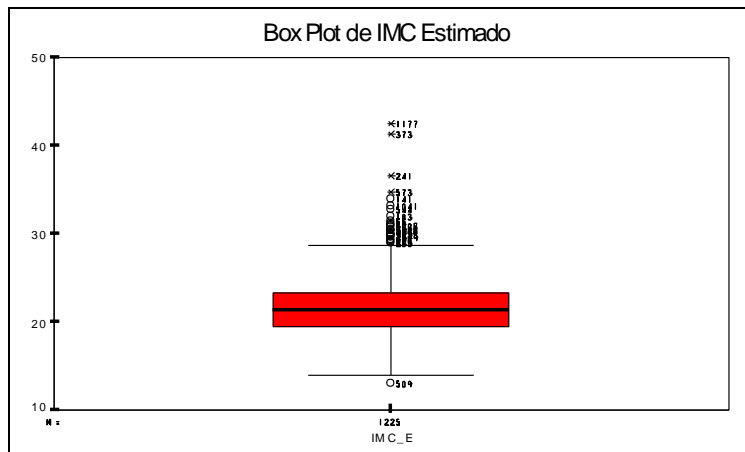
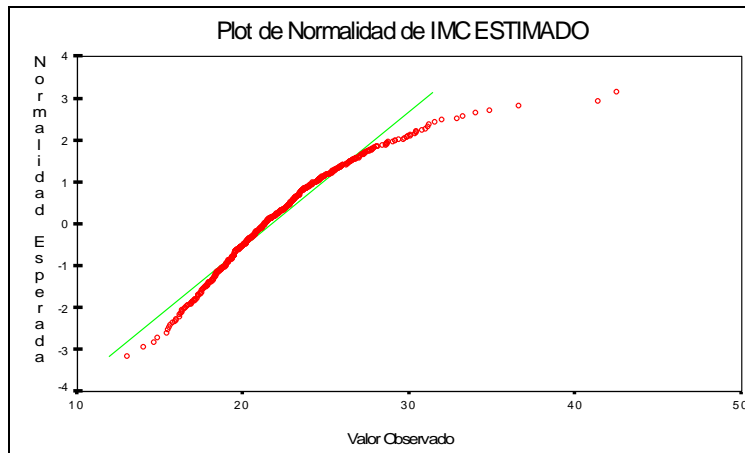
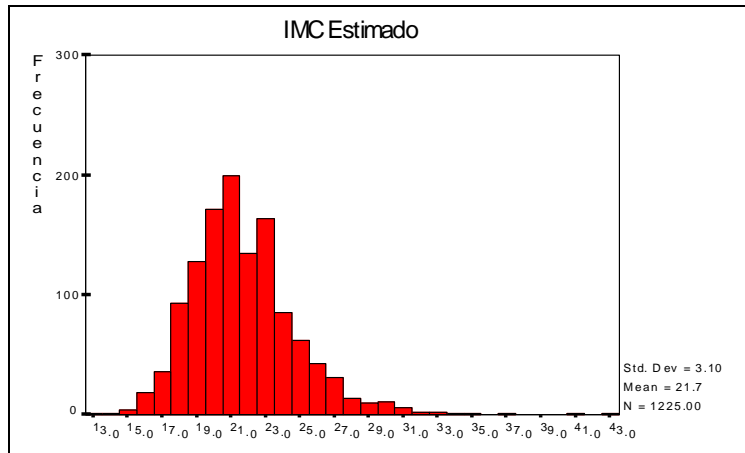
Tabla 5.2. Transformaciones más frecuentemente utilizadas.

Nombre	Ecuación	Potencia	Se usa cuando:	Efectos en la curva
Recíproco	$Y' = 1 / Y$	- 1	Exista cola en el lado derecho de la curva	Jala la joroba de la curva hacia la derecha solo que de forma aún más pronunciada
Recíproco de la raíz cuadrada	$Y' = 1 / \sqrt{Y}$	- 1/2	igual	Jala la joroba de la curva hacia la derecha solo que de forma más pronunciada
Logarítmica	$Y' = \log Y$	0	igual	Jala la joroba de la curva hacia la derecha.
Raíz cuadrada	$Y' = \sqrt{Y}$ ó $Y' = \sqrt{Y + 1}$	1/2	Exista cola en el lado izquierdo de la curva	Jala la joroba de la curva hacia la izquierda.
Normal	$Y' = Y$	1		Sin cambios
Cuadrado	$Y' = Y^2$	2	Exista cola en el lado izquierdo de la curva	Jala la joroba de la curva hacia la izquierda solo que de forma más pronunciada.
Cubo	$Y' = Y^3$	3	igual	Jala la joroba de la curva hacia la izquierda solo que de forma aún más pronunciada

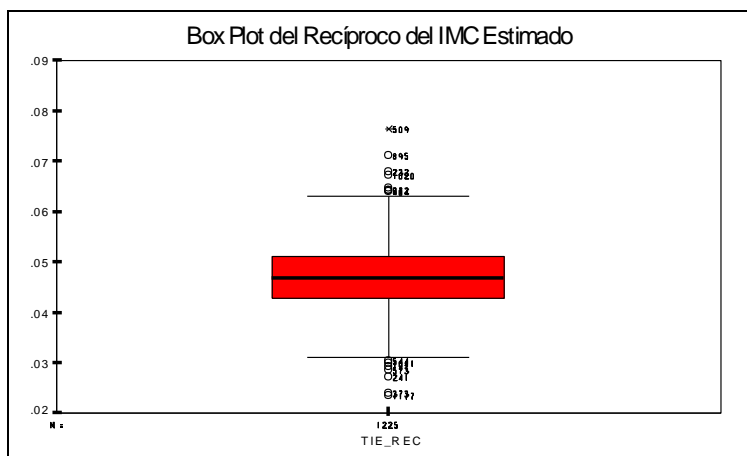
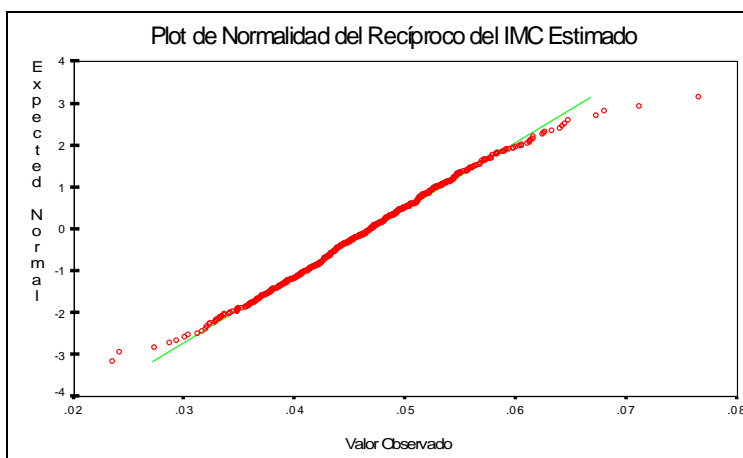
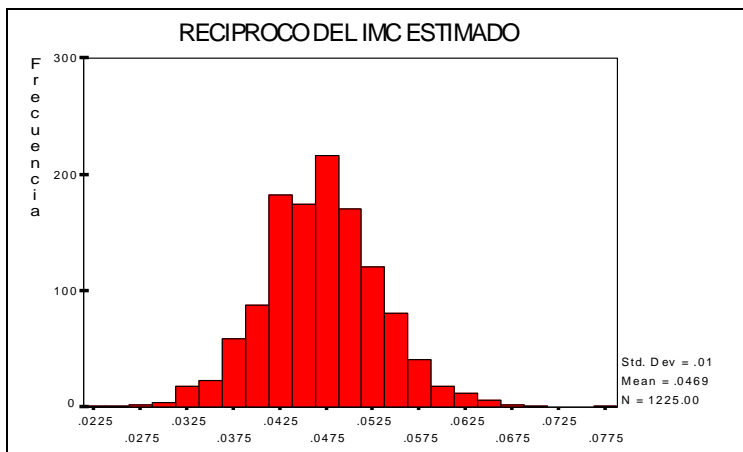
A continuación se presenta un ejemplo concreto acerca del uso de las transformaciones y su efecto en un conjunto de datos.

En este ejemplo se utilizó el recíproco del IMC (Índice de Masa Corporal) estimado de los estudiantes. El IMC es un índice que proporciona una medida del peso corporal. Como puede observarse, las gráficas antes y después de la transformación muestran que estas últimas mejoraron la distribución, aunque la distribución no es "perfectamente" normal.

GRÁFICAS DE LOS DATOS ORIGINALES, PREVIAS A LA TRANSFORMACIÓN:



## GRÁFICAS DESPUÉS DE LA TRANSFORMACIÓN:



## **5.5. PRUEBA PARAMÉTRICA PARA EL DISEÑO DE DOS MUESTRAS INDEPENDIENTES: t.**

Como se señaló anteriormente, en el diseño de dos muestras independientes pueden presentarse dos condiciones: a) las muestras provienen de la misma población, y b) las muestras provienen de diferentes poblaciones. Sin embargo, en cualquiera de estas dos situaciones el interés del investigador es el mismo: encontrar diferencias entre los valores de las variables dependientes.

El investigador maneja una variable independiente o tratamiento para el que establece por lo menos, dos valores o condiciones. Lo que intenta es probar si esos dos valores o condiciones del tratamiento o variable independiente producen efectos diferentes en la variable dependiente que investiga. Para esto procede como sigue: determina la población y extrae al azar  $N$  Sujetos quienes serán asignados, también al azar, a los grupos control y experimental. A los Sujetos del grupo experimental se les aplicará un valor de la variable independiente (el valor total o mayor) y a los Sujetos del grupo control el otro valor (cero o un valor menor).

El supuesto básico del que parte este diseño es que las varianzas de ambos grupos son iguales. Esto es, que antes de aplicar el tratamiento, las diferencias que puedan existir entre ellos no son estadísticamente significativas.

Dicho supuesto se apoya en lo siguiente:

- a) los Sujetos provienen de la misma población;
- b) los procedimientos aleatorios (muestras seleccionadas al azar, y sujetos asignados también al azar, a los grupos experimental y control) tienden a igualar a los Sujetos en las variables extrañas relevantes a la relación (variable independiente-dependiente) que se estudia.

El cumplimiento de dicho supuesto permite atribuir, con un determinado nivel de probabilidad, que la diferencia que se encuentre se debe a la aplicación del tratamiento. Esto es, si antes de éste las varianzas de los grupos eran iguales, y después de su aplicación se encuentran diferencias, éstas pueden adjudicarse al tratamiento.

*Ejemplo:* Supóngase que el investigador quiere probar los resultados que dos procedimientos de enseñanza tienen sobre el proceso de aprendizaje: un método programado comparado con el método tradicional. De acuerdo con su propósito aplica al grupo

experimental el método programado y al grupo control el método tradicional, obteniendo así, los valores de la variable dependiente de cada grupo. Es decir, los puntajes de aprendizaje, que para este ejemplo representan las respuestas correctas a una prueba de conocimiento. El investigador está seguro que ésta mide la variable dependiente a nivel intervalar.

La hipótesis del investigador (hipótesis de trabajo) predice que el aprendizaje del grupo experimental será superior al del grupo control (en este caso se establece dirección), o simplemente puede establecerla en términos de que se encontrarán diferencias (sin determinar dirección).

*Hipótesis Nula (H<sub>0</sub>).* No se encontrarán diferencias estadísticamente significativas entre las medias de los grupos experimental y control. Esta hipótesis es estadística y se le conoce como la hipótesis de las no diferencias. Esta junto con la Hipótesis Alternativa (H<sub>1</sub>), también de carácter estadístico, se prueban directamente en el proceso de análisis estadístico, y de sus resultados, esto es, de la decisión estadística a la que se llegue, se confirma o no la respectiva hipótesis de trabajo. Así, si se acepta la H<sub>0</sub>, se rechaza la H<sub>1</sub> y consecuentemente no se confirma la hipótesis de trabajo, si se rechaza la nula se acepta la alterna y se confirma la de trabajo.

En realidad lo que la hipótesis nula (H<sub>0</sub>) predice es que no habrá diferencias entre las medias de la población, entre las verdaderas medias ( $H_0 = M_1 = M_2 = 0$ ) ya que el propósito del experimento es hacer inferencias con respecto a la población, a partir de los resultados obtenidos con las muestras.

Ahora bien, con respecto al planteamiento de la hipótesis estadística alterna (H<sub>1</sub>) se toma en cuenta como se estableció la hipótesis de trabajo. Si ésta tiene dirección, la H<sub>1</sub> también la tendrá; por el contrario, si la hipótesis de trabajo sólo predice que se encontrarán diferencias (es decir, no establece dirección), así deberá formularse la H<sub>1</sub>. Hipótesis alterna: se encontrarán diferencias estadísticamente significativas, entre las medias siendo más alta la del grupo experimental.

Como puede verse, la H<sub>1</sub> establece dirección. La hipótesis alterna sin dirección se formularía como sigue: se encontrarán diferencias estadísticamente significativas entre la media del grupo experimental y la del grupo control.

**Prueba Estadística.** La elección de la prueba **t** se justifica porque los datos cumplen (el investigador tiene bases que le permite asumirlo así) con los supuestos de la estadística paramétrica señalados en el capítulo anterior. En este ejemplo,

sería: observaciones o mediciones independientes; igualdad de las varianzas; medición intervalar y distribución normal. Para cumplir con este último supuesto se requiere, para investigaciones de carácter experimental y la utilización de pruebas paramétricas, por lo menos una  $N = 30$ . Por otra parte, para favorecer la homocedasticidad (igualdad de varianzas) de los datos, es aconsejable que las muestras se seleccionen al azar, y las **n's** (tamaño de los grupos) sean iguales (ver por ejemplo, Gayen, 1950 y Keppel, 1973, entre otros).

Al interpretar el valor  $t$  lo que se busca es determinar si la diferencia entre las medias es significativa, esto es, si  $t = 6.42$  es significativa al nivel de significancia que previamente se haya establecido. Se obtienen los grados de libertad ( $gl$ ) que para la prueba **t** son  $gl = N - 2$ , donde  $N = n_1 + n_2$ . Así, para este ejemplo,  $gl = 5 + 5 = 10 - 2 = 8$ .

Si el nivel de significancia previamente establecido fue de  $\alpha = 0.05$ , el resultado arriba señalado indica que el valor encontrado de  $t = 6.42$  es significativo con una probabilidad, no solo al 0.05, sino que es significativo con una probabilidad de error mucho más baja ( $p < 0.001$ ).

Dicho nivel de probabilidad significa que si se repitiera el experimento 1000 veces, sólo una vez de 1000, ese valor de **t** estaría dado por el azar, o bien, que la diferencia entre las medias se podría atribuir en esa misma proporción (1:1000) a factores aleatorios. Por lo tanto, la intervención del azar quedaría descontada en 999 veces de 1000.

Como puede verse, el resultado de la prueba estadística es altamente significativo. Si se cumple el supuesto de la igualdad inicial de las medias, dicho resultado permite considerar que el tratamiento experimental (específicamente el método de enseñanza) es la causa de la diferenciación final entre las medias, siendo más efectivo el método programado ( $\bar{X}_1 = 9.2$ ,  $s = 1.8$ ) que el tradicional ( $\bar{X}_2 = 5.8$ ,  $s = 1.9$ ).

La decisión final tomada con base en el resultado de la prueba estadística remite al investigador, primero a la hipótesis estadística y después a la hipótesis de trabajo o empírica. Con respecto a la primera, se procede a rechazar la hipótesis nula de las no diferencias y aceptar la hipótesis alterna. En este punto se debe tener cuidado que los resultados obtenidos sean los postulados por dicha hipótesis. En el ejemplo se estableció una dirección: que la media del grupo experimental sería más alta que la media del grupo control. Si el valor de las medias hubiera sido inverso ( $\bar{X}_2 > \bar{X}_1$ ), aún cuando la diferencia seguiría siendo altamente significativa, y la hipótesis nula también tendría que ser rechazada, la hipótesis alterna no se

aceptaría, y con esto, la hipótesis de trabajo no se habría confirmado.

Por otro lado, cuando la hipótesis de trabajo (y consecuentemente la alterna) no establece dirección del cambio o diferencia, y ésta es significativa, se acepta la alterna y se confirma la de trabajo. Si se da la condición de una diferencia significativa, lo que resta es interpretarla. En el caso específico de este ejemplo, si la  $\hat{\alpha}_2 > \hat{\alpha}_1$ , entonces se concluiría que el método tradicional es más efectivo en la enseñanza que el método programado; o por el contrario, que el programado es mejor que el tradicional cuando  $\hat{\alpha}_1 > \hat{\alpha}_2$ , que es la decisión correspondiente al ejemplo presentado.

En el ejemplo que sigue se trabajó con una muestra formada por 1, 214 estudiantes de bachillerato (651 varones y 563 mujeres). A estos sujetos se les pidió que eligieran su figura ideal de entre 9 figuras que se les presentaban. Las nueve figuras forman una escala visual en un continuo de peso corporal en el que las siluetas 1 y 2 representan figuras muy delgadas, la 3 y 4 delgadas, la 5 figuras con peso normal, la 6 y 7 con sobrepeso y la 8 y 9 obesas. Se obtuvieron las **frecuencias** por separado, de las elecciones que hicieron los estudiantes hombres y las mujeres

Tabla 5.3: Distribución de frecuencias. Elección de figura ideal hecha por los sujetos del sexo masculino



valores	frecuencias	P O R C E N T A J E S		
		(1)	(2)	(3)
1.00	1	.1	.2	.2
2.00	2	.3	.3	.5
3.00	15	2.2	2.3	2.8
4.00	141	21.1	21.7	24.4
5.00	327	49.0	50.2	74.7
6.00	149	22.3	22.9	97.5
7.00	8	1.2	1.2	98.8
8.00	4	.6	.6	99.4
9.00	4	.6	.6	100.0
16	2.4	perdidos		
<u>Total</u>	<u>667</u>	<u>100.0</u>	<u>100.0</u>	

Media 5.018                      Mediana 5.000  
Modo 5.000      Desviación standard.891  
Kurtosis2.930                      Skewness .370  
Rango 8.000                      Mínimo 1.000  
Máximo9.000                      Varianza .794

- (1) Porcentajes de respuesta total (Base = N)
- (2) Porcentaje de respuestas (Base =total de respuestas dadas)
- (3) Porcentaje Acumulado

Tabla 5.4 Distribución de frecuencias de la variable selección de la figura ideal hecha por estudiante del sexo femenino.

Valores	Frecuencias	P O R C E N T A J E S		
		(1)	(2)	(3)
1.00	4	.7	.7	.7
2.00	20	3.5	3.6	4.3
3.00	108	18.7	19.2	23.4
4.00	316	54.6	56.1	79.6
5.00	103	17.8	18.3	97.9
6.00	7	1.2	1.2	99.1
7.00	4	.7	.7	99.8
8.00	1	.2	.2	100.0
.	16	2.8	Perdidos	
	-----	-----	-----	
Total	579	100.0	100.0	

Media 3.952                      Mediana 4.000  
 Modo 4.000      Desviación standard.850  
 Kurtosis2.312                  Skewness .039  
 Rango 7.000                      Mínimo 1.000  
 Máximo8.000                      Varianza .722

- (1) Porcentajes de respuesta total (Base = N)
- (2) Porcentaje de respuestas (Base =total de respuestas dadas)
- (3) Porcentaje Acumulado

Tanto en las distribuciones de frecuencias como en los descriptores estadísticos de las Tablas 5.3 y 5.4, puede observarse que existen diferencias en las respuestas dadas por hombres y mujeres.

La pregunta que nos planteamos es: ¿esas diferencias observadas en las respuestas dadas por los sujetos son estadísticamente significativas de tal manera que se pueda considerar que hombres y mujeres tienen figuras ideales diferentes o bien sólo son diferencias debidas al azar? Debe recordarse además que la pregunta básica es si las medias de las dos poblaciones (nótese que la pregunta no es sobre las muestras) son iguales. Para poder responder esta pregunta

debemos utilizar una prueba de hipótesis que nos permita comparar dos grupos con observaciones independientes y es cuando utilizamos, precisamente la prueba t para muestras independientes.

Sin embargo, antes de desarrollarla se requiere determinar si los datos cumplen o no con el supuesto de homocedasticidad (igualdad de varianza). Para cumplir con este propósito se cuenta con diferentes pruebas, por ejemplo la de Bartlett, de Cochran y la de Levene. En general se considera que la prueba de Bartlett es complicada, además de que, al igual que la de Cochran, se muestra muy sensible a la no-normalidad de las distribuciones. La de Levene, en cambio, es más sencilla en cuanto a su desarrollo y además poco sensible a los problemas relacionados con la normalidad. La prueba de Levene se desarrolla como un análisis de varianza con la única diferencia: parte de una matriz de puntajes de desviación ( $x = X - \bar{x}$ ) esto es desviaciones con respecto a la media dentro de cada grupo, mientras que el análisis de varianza arranca con una matriz de puntajes crudos.

Los resultados de la prueba de Levene ( $F = .977, p = .323$ ) permiten aceptar la  $H_0: VAR_1 = VAR_2 = 0$  lo que significa que los datos de nuestro ejemplo cumplen con el supuesto de homogeneidad y se procede, entonces a la aplicación de la prueba t correspondiente. para datos con homogeneidad de varianza:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_p^2}{N_1} + \frac{S_p^2}{N_2}}} \qquad S_p^2 = \frac{(N_1 - 1)S_1^2 + (N_2 - 1)S_2^2}{N_1 + N_2} \quad (13)$$

En donde:

$S_p^2 =$  varianza combinada = promedio pesado de las varianzas individuales que se obtiene con el desarrollo de  $S_p^2$

A continuación se presentan los resultados de la prueba t utilizada para comparar las respuestas de hombres y mujeres a la petición de elegir figura ideal.

Tabla 5.5 Resultados. Prueba t para muestras independientes. Comparación de la Figura Ideal de hombres y mujeres.

	<b>t</b>	<b>GL</b>	<b>Prueba 2 colas</b>	
	21.25	1212	..000 .	
<b>Variable</b>		<b>Numero</b>	<b>Media</b>	<b>Desviación</b>
<b>Error Estándar</b>				
<b>de casos</b>			<b>Estándar</b>	<b>de la Media</b>
FIG_IDE				
Masculino	651	5.0184	.891	.035
Femenino	563	3.9520	.850	.036

Lo que se calcula cuando se resuelve la ecuación de la prueba  $t$  es la probabilidad con la que ocurriría una diferencia del tamaño de la observada ( la que se obtuvo entre las medias) si las dos medias de la población ( $M_1$  y  $M_2$ ) fueran iguales .

Esta probabilidad se conoce como el nivel de significancia observado. De esta manera , si el nivel de significancia observado es pequeño ; Por lo general £0.05 o £ 0.01 se rechaza hipótesis de que las medias de la población son iguales ( $H_0: M_1 = M_2 = 0$ ).

Como se puede observar las diferencias entre los dos grupos son significativas ( **$t$  (1212) = 21.25,  $p=.000$** ). Puede decirse que la Figura Ideal de las mujeres mostró ser significativamente más delgada (Media = 3.95,  $S = .850$ ) que la de los hombres (Media = 5.02,  $S = .891$ ). Específicamente la figura ideal promedio de las mujeres cayó dentro de la categoría delgada, mientras que la de los hombres en la de peso normal. Estos resultados pueden apreciarse gráficamente en la Fig. 5.9.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{S_1^2/N_1 + S_2^2/N_2}} \quad (14)$$

En donde

$S^2$ =Varianza  
N= Tamaño de la Muestra

Subíndices 1 y 2 =Grupos 1 y 2

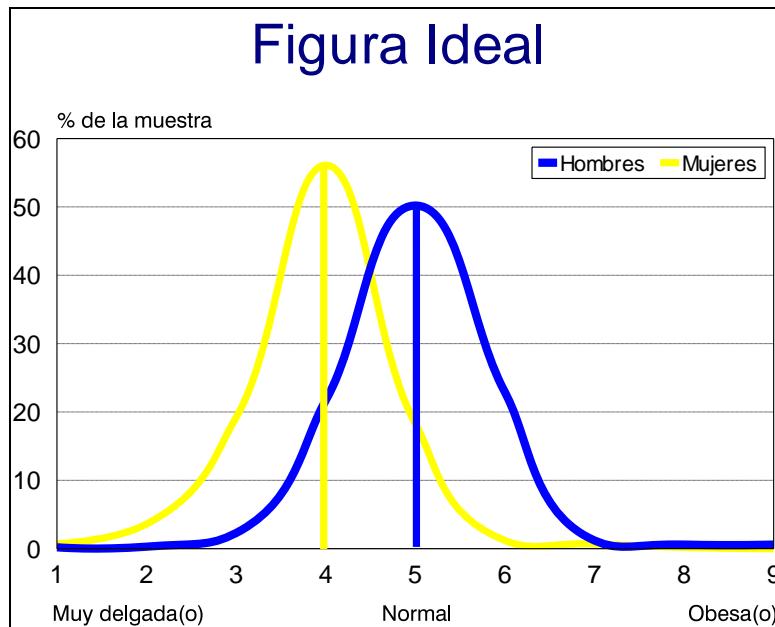


Fig. 5.9 Distribuciones de la Figura Ideal por Sexo.

de **t** para **n's** desiguales.

Sin embargo, se ha encontrado que cuando la diferencia no es muy grande, menor de 2 veces el tamaño de la más pequeña, (i.e.,  $n_1 = 20$  y  $n_2 = 32$ ), el uso de la **t** para **n's** iguales no produce errores importantes y por lo tanto es utilizable en tales casos. En cambio, cuando la diferencia es igual o mayor que 2 veces el tamaño de la **n** más pequeña, es aconsejable utilizar la ecuación para **n's** desiguales.

Debido a que lo tratado para el caso de **n's** iguales es válido para cuando  $n_1 = n_2$ , y lo único que cambia es el desarrollo de la ecuación, no tiene caso detenerse en otro ejemplo. Todo el procedimiento, sus supuestos, desarrollo e interpretación presentados, es el mismo para ambos casos.

### 5.6 PRUEBA **t** PARA DOS MUESTRAS RELACIONADAS:

En el capítulo anterior se trató lo referente al diseño de dos muestras relacionadas. Se señaló que este diseño puede responder a dos situaciones diferentes: las muestras se igualan o aparean en una variable importante, o bien, los Sujetos de

una misma muestra se someten a dos condiciones o mediciones, constituyendo dos muestras, en el sentido de que se tienen dos grupos de datos, fungiendo uno de ellos como control (los datos obtenidos antes de la aplicación de la variable independiente) que corresponde a lo que usualmente los investigadores denominan *pre-test*; y el otro como grupo experimental, que se refiere a los resultados obtenidos una vez aplicada la variable independiente (*post-test*).

Esta versión del diseño de dos muestras relacionadas, mejor conocido como diseño antes-después enfrenta al investigador con serios problemas a los que deberá dar solución para descartar interpretaciones alternativas derivadas de dichos problemas, y que oscurecen la relación que estudia. Estos problemas se derivan principalmente de los factores que pueden intervenir entre una y otra medición: los más obvios son aquellos que pueden darse en los Sujetos por el simple paso del tiempo y el efecto del aprendizaje.

Para intentar resolver dichos problemas, por lo general se decide dejar, entre una y otra medición, un lapso tan corto, que asegure el menor riesgo de que se den sus efectos -por ejemplo, maduración de los Sujetos en su sentido más amplio pero no tan corto, como para que intervenga el aprendizaje. Con respecto a éste, también se puede recurrir a la elaboración de condiciones o instrumentos paralelos. Sin embargo, no existe posibilidad de que dos condiciones se dupliquen exactamente iguales, por un lado, y por el otro, la construcción de instrumentos paralelos requieren de confiabilidad y validez probadas que implican un costo mayor --en tiempo, esfuerzo y dinero-- que en muchas ocasiones no es posible afrontar. Se han desarrollado otras alternativas de solución, en las que se utilizan grupos adicionales.

El diseño de cuatro grupos propuesto por Solomon (1949) intenta responder principalmente al problema encontrado con los diseños *pretest-posttest* en relación al cambio de actitudes. El *pretest* puede modificar la posición inicial de los Sujetos (hacerlos más sensibles, o también más resistentes) a la manipulación persuasiva. De esta manera el autor sugiere utilizar dos grupos control más: uno que recibe un tratamiento experimental y un *posttest* y otro que sólo recibe un *posttest*. El diseño completo (cuatro grupos) queda como sigue:

		PRETEST	M MANIPULACIÓN	POSTEST
DISEÑO	GRUPO EXPERIMENTAL	SI	SI	SI
ANTES-DESPUÉS	GRUPO CONTROL 1	SI	NO	SI
DISEÑO	GRUPO CONTROL 2	NO	SI	SI

SOLO DESPUÉS	GRUPO CONTROL 3	NO	NO	SI
--------------	-----------------	----	----	----

(Esquema tomado de Insko, 1967, p.4)

Los resultados obtenidos con este diseño de cuatro grupos se pueden someter a un análisis de varianza (diseño factorial) que permite (como se verá posteriormente en este mismo capítulo) probar los efectos principales, en este caso del *pretest* y del tratamiento experimental, así como los de interacción (tratamiento experimental X *pretest*). Solomon sugiere que los puntajes faltantes del *pretest* (para los grupos control II y III) se pueden obtener haciendo una estimación de ellos sobre la base de las medias de los grupos experimental y control I. Este procedimiento, sin embargo, disminuye los grados de libertad y viola el supuesto de independencia del que parten las pruebas de significancia, según la crítica de Campbell (1957). Este autor propone que el análisis de varianza se lleve a cabo sólo con los puntajes del *posttest* que proporciona el efecto de la interacción que se busca (*pretest* X tratamiento experimental) sin los problemas antes señalados. Sin embargo, Campbell hace mayor énfasis en el diseño de **sólo después** (los grupos control II y III de Solomon) ya que es más sencillo y no requiere de mucho esfuerzo o cooperación por parte de los Sujetos, de tal manera que puede utilizarse en investigaciones fuera del laboratorio (Insko, 1967, p.6).

Existen otros procedimientos en los que el interés se centra en eliminar las diferencias iniciales entre los Sujetos (reducir la varianza de error) más bien que en eliminar dicha interacción. Entre estos, el análisis de covarianza ha sido ampliamente utilizado, con el diseño tradicional de antes-después sin necesidad de recurrir a grupos adicionales. Sin embargo, a pesar de que existen múltiples recursos en el intento de solución al problema de los efectos "contaminadores" que pueden intervenir entre una y otra medición, no se ha encontrado todavía un procedimiento o un modelo satisfactorio de solución.

Ahora bien, en el diseño de dos muestras relacionadas se mantienen los mismos supuestos estadísticos y lógica de análisis señalados para la condición de dos muestras independientes, debido a que en ambos casos se está haciendo referencia a la aplicación de pruebas paramétricas. En relación específica al diseño en su versión de muestras igualadas o aparejadas, se utiliza una nueva variable que es precisamente la que permite igualarlas.

*Ejemplo.*

Retomando nuestro ejemplo se tiene: como variable independiente, método de enseñanza con dos valores o variaciones: programado y tradicional. Como variable dependiente, puntajes de aprendizaje (número de respuestas

correctas). Supóngase ahora que el investigador considera que la inteligencia de los Sujetos es una variable que puede intervenir en el aprendizaje y que por lo tanto, éste no sólo depende del método de enseñanza, sino también del grado de inteligencia de los Sujetos.

Decide así, igualar los grupos en la variable inteligencia. Antes de aplicar el tratamiento o variable independiente, aplica una prueba de inteligencia. En seguida ordena estos puntajes y forma pares de Sujetos con puntajes iguales; asigna al azar, al grupo control y al experimental, cada miembro de cada par de Sujetos con puntajes iguales. Al tener los dos grupos igualados en la variable inteligencia, se procede entonces a la aplicación del tratamiento experimental.

Con los valores obtenidos en la variable dependiente se ejecutan las operaciones indicadas en la fórmula de **t para dos muestras relacionadas**.

$$t = \frac{\bar{D}}{S_D/N} \quad (15)$$

En donde:

D = diferencias observadas entre las medias ;  
S<sub>D</sub> = desviación estándar de las diferencias observadas pareadas;  
N = Número de pares.

Se consulta la distribución muestral de valores **t**, con los grados de libertad correspondientes y al nivel de significancia establecido previamente. Para la prueba **t** de dos grupos relacionados, gl = n - 1. El valor tabulado de **t**, con  $\alpha = 0.05$ , y 4 grados de libertad, es igual a 2.77. Se tiene por lo tanto, el siguiente resultado:  $t = 6.67$  (valor observado)  $> t_{.05} = 2.77$  (valor tabulado) con 4 grados de libertad que permite hacer la decisión con respecto a la prueba de la hipótesis nula. Si el nivel de significancia se hubiera establecido al 0.05 ( $\alpha = 0.05$ ), bastaría con un valor obtenido de  $t = 2.77$ . Habiendo encontrado una  $t = 6.67$ , que es altamente significativa (con  $p < 0.05$ ) se rechaza la hipótesis nula y se acepta la alterna confirmándose así la hipótesis del investigador (la hipótesis de trabajo).

Cuando se elige una variable de igualación es importante que ésta se relacione con la variable dependiente y preferentemente que dicha relación sea positiva. Esto es, que los puntajes altos en la variable dependiente correspondan a puntajes altos en la variable de igualación y los puntajes bajos de la primera correspondan a puntajes bajos de la segunda (correlación



significativa y positiva). Es aconsejable así, probar la relación entre esas variables, de tal modo que si la relación es baja o no existe, es preferible recurrir a otro tipo de diseño.

También es importante que la variable de igualación no afecte o contamine los resultados del tratamiento, produciendo, por ejemplo: aprendizaje, cansancio o alertamiento de los Sujetos respecto al propósito del experimento.

Ahora bien, la ventaja del diseño de dos muestras relacionadas (v. gr., antes-después), en su versión de muestras igualadas o aparejadas es que prueba la existencia del supuesto de la igualdad inicial, por lo menos en la variable que el investigador considera más importante. Su desventaja, por otra parte, es la posibilidad de que se contaminen los resultados del experimento, por las razones que arriba se señalaron. Sin embargo, la prueba estadística de ese diseño, prevee hasta cierto punto, el error que puede producir la aplicación previa de la variable de igualación en este caso, y la dependencia de las medidas en la otra versión del diseño de dos muestras relacionadas, pues con éste se pierden más grados de libertad ( $n - 1$ ), que con el de dos grupos independientes ( $N - 2$ ).

En otros términos, esto significa que se requieren, para muestras relacionadas, valores  $t$  más altos para que sean significativos. De lo anterior se desprende que el diseño de dos muestras relacionadas requiere de  $N$ 's más grandes que el de dos muestras independientes.

Para probar la existencia de la relación, y de su magnitud, entre la variable dependiente y la de igualación, puede utilizarse el coeficiente de correlación Producto Momento de Pearson.

En el ejemplo que sigue, se trabajó con una muestra formada por 558 mujeres que se encontraban cursando estudios de bachillerato. A estas personas se les pidió que eligieran su figura ideal (como les gustaría estar) y su figura actual (como estaban en ese momento) de entre las mismas 9 figuras que hemos descrito en el ejemplo anterior.

A continuación se presentan las tablas de frecuencias elaboradas con las respuestas de estas mujeres en el momento de elegir su figura actual e ideal. Así mismo se presentan los principales estadísticos descriptivos.

Tabla 5.6 Distribución de Frecuencias de la Variable Selección de Figura Actual.

P O R C E N T A J E S				
Valores	Frecuen cias	(1)	(2)	(3)
1.00	36	6.2	6.3	6.3
3.00	202	34.9	35.2	41.5
5.00	218	37.7	38.0	79.4
7.00	115	19.9	20.0	99.5
9.00	3	.5	.5	100.0
.	5	.9	Perdido s	99.8
-----	-----	-----		100.0
Total	579	100.0	100.0	

Media	4.467	Mediana	5.000
Modo	5.000	Desviación standard	1.737
Kurtosis	-.661	Skewness	.015
Rango	8.000	Mínimo	1.000
Máximo	9.000	Varianza	3.017

(1) Porcentajes de respuesta total (Base = N)

(2) Porcentaje de respuestas (total de respuestas dadas)

(3) Porcentaje Acumulado

Tabla 5.7. Distribución de Frecuencias de la Variable Selección de la Figura Ideal.

Valores	Frecuencias	P O R C E N T A J E S		
		(1)	(2)	(3)
1.00	24	4.1	4.3	4.3
3.00	424	73.2	75.3	79.6
5.00	103	17.8	18.3	97.9
7.00	11	1.9	2.0	99.8
9.00	1	.2	.2	100.0
.	16	2.8	Perdidos	
-----	-----	-----		
Total	579	100.0	100.0	

Media	3.36	Mediana	3.000
	9		
Modo	3.00	Desviación	1.070
	0	standard	
Kurtosis	3.038	Skewness	1.183
Rango	8.00	Mínimo	1.000
	0		
Máximo	9.00	Varianza	1.144
	0		

(1) Porcentajes de respuesta total (Base = N)

(2) Porcentaje total (Base = total de respuestas dadas)

(3) Porcentaje Acumulado

En esta ocasión también existen diferencias evidentes en las respuestas dadas por las adolescentes en el momento de elegir su figura actual y su figura ideal, sin embargo nuevamente debemos preguntarnos si estas diferencias son debidas al azar o no. Para encontrar la respuesta utilizamos t para muestras relacionadas, porque cada estudiante tiene dos puntajes:

Tabla 5.8. Resultados Prueba t para muestras relacionadas. Comparación de las variables Figura Ideal y Figura Actual

Diferencias						
Media	Desviación Estandard	Error standard de la media_	t	Gl	Significancia	
.5663	1.398	.059	9.57	557	.000	
Variable	Numero de pares	Correlación	Sig. dos colas	Media	Desviación standard	Error standard de la media
FIG_ACT	558	.186	.000	4.5197	1.282	.054
FIG_IDE				3.9534	.846	.036

Los resultados muestran diferencias significativas  $t (557) = 9.57, p=.000$ . En la Estos mismos resultados se exponen en la gráfica siguiente::

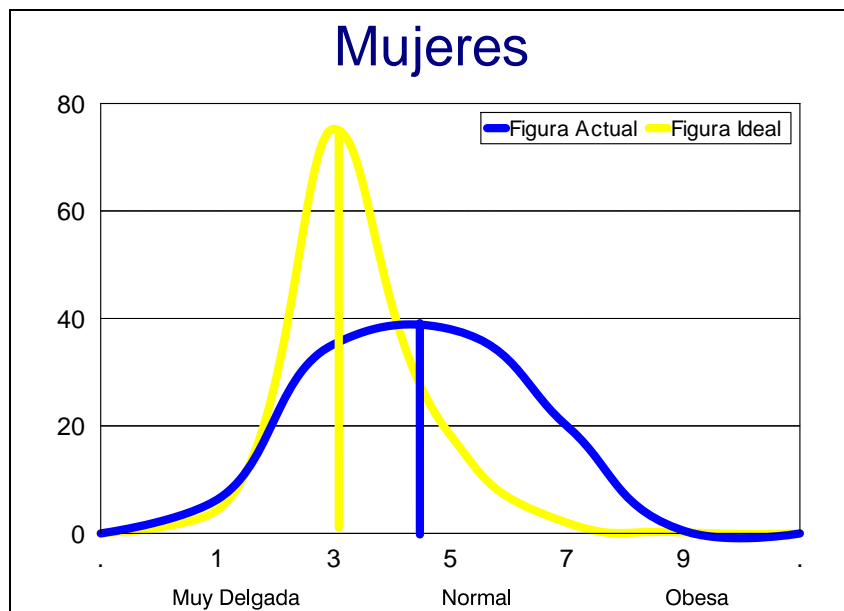


Figura 5.10. Distribución de la variables figura actual y figura ideal

#### Uso de la prueba t para n's desiguales.

Por otro lado, es importante aclarar que también se dispone de una fórmula  $t$  general utilizable tanto para el diseño de dos muestras independientes como para el de dos muestras relacionadas.

Esta fórmula se desarrolla en términos de desviaciones ( $S_1$  y  $S_2$ ) y de varianzas ( $S_1^2$  y  $S_2^2$ ). La varianza, no es más que el cuadrado del valor de la desviación estándar ( $S$ ).

La  $R^{12}$  de la fórmula general de  $t$ , se refiere al coeficiente de correlación Producto-Momento de Pearson correspondiente a la fórmula tradicional de correlación.

### **5.7. Pruebas Paramétricas Para k Muestras Independientes: Rangos de Duncan y Análisis de Varianza.**

Para los diseños de investigación que manejan más de dos grupos, se exponen dos procedimientos que pueden considerarse alternativos. Producen los mismos resultados y su aplicación tiene los mismos requerimientos; es decir, los que exige cualquier prueba estadística paramétrica y que ya se han señalado.

#### **5.7.1. Prueba de Rangos de Duncan.**

En seguida se presenta un ejemplo para tres grupos independientes; sin embargo, el procedimiento que se sigue es generalizable a cualquier número de éstos.

*Ejemplo.* Al ejemplo que se ha venido utilizando se le agrega un tercer valor o variación de la variable independiente, para lo cual se necesita de un tercer grupo. En este caso el investigador quiere probar qué método de enseñanza, el tradicional (A), el programado (B) o el audiovisual (C), es más efectivo en el aprendizaje de una tarea o tema particular.

Su propósito es probar si la diferencia entre las medias de las tres condiciones son diferentes y estadísticamente significativas, y cuál de las tres produce una media más alta, puesto que los valores más altos de la variable dependiente representan mayor aprendizaje.

Se procede a probar si las diferencias entre las medias son significativas. Para el caso de tres medias se prueban las siguientes diferencias:

- a) entre la media más alta y la más baja;
- b) entre la media más baja y la intermedia;
- c) entre la media más alta y la intermedia.

Como puede verse, y este es el procedimiento para cualquier número de medias, se prueban todas las posibles comparaciones. En este ejemplo, las posibles comparaciones son A-C; A-B y B-C.

Para llevar a cabo las pruebas se consideran los rangos estandarizados significativos mínimos o valores  $r_p$ .

En el símbolo  $r_p$ ,  $r$  se refiere al rango y  $p$  al número de medias que intervienen en cada comparación. Así, en la comparación de las medias extremas (A-C),  $p = 3$ ; por lo tanto  $r_p = 3$ . Se busca en una tabla de Duncan de valores  $r_p$  (ver McGuigan, 1993), el valor que se encuentra en la intersección de las columnas y las hileras correspondiente a los grados de libertad. Puede verse entonces que  $r_3$  adquiere un valor determinado para cada comparación que se efectúa.

En seguida se computan los valores  $R_p$ , que son los rangos significativos mínimos, obtenidos a partir de los propios datos. Puede decirse que estos son los rangos empíricos, mientras que los primeros son los teóricos.

Por último, se ordenan las medias de menor a mayor para sacar las diferencias entre ellas; compararlas y probar su significancia.

Para que las diferencias sean estadísticamente significativas con los grados de libertad (que en este caso son  $gl = N - r$ , número de Sujetos - número de grupos) y el nivel de significancia establecidos, las diferencias entre las medias que se estén comparando deben ser mayores a los rangos significativos mínimos empíricos ( $R_p$ ).

Los resultados permiten decidir si las diferencias C-B, B-A y A-C son estadísticamente significativas. Con base en los resultados el investigador puede rechazar la hipótesis nula, la de las no diferencias.

Como ya se había señalado, el procedimiento para el caso de tres grupos es generalizable a cualquier número de grupos independientes.

Por otra parte, la lógica del procedimiento señala que si la diferencia entre las medias extremas no resulta significativa (es decir, se encuentra una diferencia menor que el valor  $R_p$  correspondiente) no tiene caso probar las otras diferencias, pues éstas, al ser menores, necesariamente serán no-significativas.

Cuando las  $n$ 's no son iguales, aún cuando el procedimiento general es el mismo que para  $n$ 's iguales, las ecuaciones correspondientes sufren cambios. Por ejemplo, si se tienen tres grupos con  $n$ 's desiguales, la fórmula del error estándar cambia en el denominador, y en el caso de  $R_p$  también se dan ciertos cambios.

### 5.3.2. Análisis de Varianza.

Como ya se había indicado, el análisis de varianza es un procedimiento alternativo al de Rangos de Duncan cuando se manejan  $k$  grupos independientes. Este goza de gran aceptación entre los investigadores utilizándose, por lo tanto, con mucha mayor frecuencia que los Rangos de Duncan.

### **Supuestos del Análisis de Varianza en el Modelo de los Efectos Fijos.**

Los supuestos del análisis de varianza con el modelo de los efectos fijos constituyen la justificación teórica del análisis y de la prueba F. Ahora bien, puesto que en la realidad difícilmente se puede encontrar o probar la existencia de todos los supuestos, es importante conocer cuáles son las consecuencias tanto para el análisis como para la aplicación de la prueba F cuando los supuestos no se cumplen.

Un primer supuesto dice que: "El error de varianza debe distribuirse normalmente". Esto es, para cada tratamiento "j" se asume que el error "e<sub>ij</sub>" se distribuye normalmente. Para el modelo I (o de los efectos fijos), esto implica que los puntajes (Y<sub>ij</sub>) responden a una distribución normal. Arriba hemos descrito ampliamente algunos procedimientos para probar este supuesto

Por otra parte, para probar los efectos de la no-normalidad de las distribuciones, se han llevado a cabo, experimentos Monte Carlo. Keppel (1973) señala los resultados encontrados por Norton, (1952). Este investigador, trabajó con puntajes extraídos de distribuciones que tenían la misma forma y con puntajes de distribuciones con diferentes formas. Encontró que en las distribuciones homogéneas la probabilidad esperada para las distribuciones F, (teórica y empírica), mostraban una alta aproximación con la de las distribuciones heterogéneas: una ligera sobre estimación del 2% al 3% en el nivel  $\alpha = .05$ ; y del 1% al 2% con  $\alpha = .01$ .

Un segundo supuesto dice que el error de varianza debe tener el mismo valor para todos los tratamientos. El efecto del no cumplimiento de este supuesto disminuye, como ya se había señalado si el número de casos en cada grupo de la muestra es el mismo. Si esto no es así, el número de casos es diferente, y se presenta heterocedasticidad entonces pueden presentarse consecuencias serias con respecto a la validez de las inferencias que se hagan.

El desarrollo del análisis de varianza detecta diferencias entre las medias de los grupos. Se asume que los efectos del tratamiento experimental o de la variable independiente se refleja en las medias. De esta manera la (o las) hipótesis nula que se prueba establece que no se encontrarán diferencias estadísticamente significativas entre las medias.

La prueba que utiliza el análisis de varianza es la prueba **F**, que definida en términos técnicos no es otra cosa más que la razón de dos varianzas:

$$F = \frac{\text{Varianza Entre los Grupos}}{\text{Varianza Dentro de los Grupos}}$$

El numerador se refiere a las diferencias (varianza) que se encuentran entre los grupos como consecuencia de los efectos, si los hay, de los tratamientos a que fueron sometidos. Esta es la varianza del efecto real, pero además, está compuesta por la varianza de error.

Por otra parte, el denominador se compone, o sólo estima la varianza del error; ésta será mayor entre mayores sean las diferencias entre los Sujetos del mismo grupo. Así, si los Sujetos del grupo A muestran entre ellos grandes diferencias y lo mismo sucede entre los del grupo B y los del grupo C, entonces la varianza dentro de los grupos será grande y entre mayor sea ésta, mayor será la varianza de error del experimento. Como puede fácilmente inferirse, si la varianza dentro de los grupos es mayor que la varianza entre los grupos, la razón **F** será pequeña y por lo tanto, no será significativa.

El investigador espera reducir la varianza dentro de los grupos, para resaltar los efectos del tratamiento, si es que éstos se dan.

Ahora bien, lo que se señaló respecto a la prueba **t** vale también para la prueba **F**: ambas son paramétricas, y de hecho, una se puede definir en términos de la otra:  $t^2 = F$ , ó  $\sqrt{F} = t$ . Con dos grupos independientes se puede utilizar cualquiera de las dos pruebas. Sin embargo, cuando se tienen más de dos grupos, (**k** grupos independientes), la prueba adecuada es la **F**.

La **F** que, como ya se indicó, expresa la razón de dos varianzas, explica por sí misma lo que hace una análisis de varianza: se denomina así porque hace una partición de la varianza total y la analiza.

Por lo tanto, el propósito fundamental de este procedimiento estadístico, es determinar la varianza total y luego analizarla por partes. Para lograr este objetivo, considérense los siguientes razonamientos, que subyacen a las ecuaciones que se emplean para ello.

a) Se calcula una media basándose en todos los valores de la variable dependiente (ignorando el hecho de que algunos Sujetos están bajo una condición y otros bajo otra).



b) Las sumas totales de cuadrados miden la desviación de todas las calificaciones de esta media general.

c) La suma de cuadrados entre-grupos es una medida de la desviación de las medias de los diversos grupos partiendo de la media general.

d) La suma de cuadrados dentro de los grupos es una suma combinada de cuadrados basada en la desviación de las calificaciones de cada grupo a partir de la media de dicho grupo.

Por lo tanto, primero se deberá calcular la suma de cuadrados total y luego analizarla por partes: la suma de cuadrados entre los grupos ( $SC_{\text{entre}}$ ) y la suma de cuadrados dentro de los grupos o intra-grupos ( $SC_{\text{intra}}$  o dentro).

Para calcular las varianzas, se dividen las SC correspondientes entre sus respectivos grados de libertad (gl).

La obtención de los grados de libertad obedece a la siguiente regla: se pierde un grado de libertad por parámetro o estadístico que se estime. Como puede verse, se calcula la  $SC_{\text{total}}$  y se descompone en  $SC_{\text{entre}}$  y  $SC_{\text{dentro}}$ . Con los grados de libertad se procede de la misma manera:

$$SC_{\text{total}} \text{ con gl} = N - 1$$

$$SC_{\text{entre}} \text{ con gl} = r - 1$$

$$SC_{\text{dentro}} \text{ con gl} = N - r$$

dónde:

gl = grados de libertad

N = Número total de observaciones

r = número de grupos.

Los grados de libertad correspondientes a la  $SC_{\text{entre}}$  más los grados de libertad correspondientes a la  $SC_{\text{dentro}}$  son igual a los grados de libertad de la  $SC_{\text{total}}$ .

*Ejemplo.* Se presenta un ejemplo con tres grupos. Cada grupo es sometido a diferentes valores de una variable independiente (tratamientos experimentales) y se espera que haya diferencias entre ellos.

Se necesitan dos varianzas o medias cuadradas (MC): *entre* los grupos y aquella asociada a *dentro* de los grupos, como se señala claramente en la Tabla Sumaria de un ejemplo de Análisis de Varianza (5.9); con base a lo que es la prueba F.

-----  
 Tabla 5.9. Tabla Sumaria de un Análisis de Varianza  
 -----

Fuente de Variación	SC(1) Suma de Cuadrados	gl(2)	MC(3) Medias de Cuadrados	F	p
Entre gpos.	496.50	2	248.25	7.81	<0.05
Dentro gpos.	285.75	9	31.75		
TOTAL	782.25	11			

-----

En la Tabla Sumaria se tienen todos los datos y el resultado del análisis de varianza: en la columna (1) aparecen las sumas de cuadrados; en la (2) los grados de libertad; la división de cada suma de cuadrados entre sus respectivos grados de libertad dan como resultado los valores de las medias cuadradas (por ejemplo,  $496.5 / 2 = 248.25$ ) o varianzas. De acuerdo con la fórmula de  $F = a$  a la razón de dos varianzas; así, con la división de las medias de cuadrados (columna 3) se obtiene el valor de  $F$ :  $248.25/31.73 = 7.81$ . El valor de  $F = 7.81$  obtenido, por si mismo no dice nada; es decir, este valor puede o no ser significativo. Para hacer la decisión estadística se requiere consultar los valores de las distribuciones muestrales de  $F$  (ver por ejemplo McGuigan, 1993) contra los cuales se contrasta el valor obtenido al nivel de significancia previamente establecido. Se busca si la  $p$  (probabilidad) asociada con el valor obtenido de  $F$  es igual o menor que el nivel de significancia seleccionado (supóngase que éste sea  $\alpha = 0.05$  lo que significa que se aceptará un resultado en el que sólo haya como máximo, un 5% de intervención del azar). En la Tabla de valores  $F$  se busca aquel que se encuentra en el punto donde confluyen los grados de libertad correspondientes en el nivel de  $p = 0.05$ . En otras palabras, se busca en la columna de la Tabla el número 2 (en grados de libertad para el numerador de  $F$ ); en la hilera de la misma el número 9 (grados de libertad del denominador de  $F$ ) y en ese punto se ve el valor que corresponde a  $p = 0.05$ . Se encuentra que con  $gl(2,9)$ ,  $F = 4.26$  para un nivel de significancia del 5%. Puesto que  $F$  obtenida =  $7.81 > 4.26$ , se rechaza la hipótesis nula y se concluye que los grupos difieren significativamente (sus medias). Es decir, que la variación de la variable independiente influyó significativamente en la variable dependiente. Este es el significado de la última columna de la

Tabla Sumaria ( $p < 0.05$ ) que expresa que  $F = 7.81$  es significativa más allá del 5%. Habiendo establecido la significancia de  $F$  y dependiendo del problema de investigación que da significado a los valores de la variable dependiente se puede proceder a la toma de decisiones revisando los valores de las medias de los grupos.

Para terminar con el tema de este apartado es importante señalar lo siguiente: la prueba  $F$  para  $k$  grupos independientes proporciona un valor general, que no especifica en cuáles pares de grupos se encuentra la diferencia significativa. Esto es, al resultar una  $F$  significativa, lo único que sabe el investigador es que existe una diferencia significativa, por lo menos, entre un par de grupos, pero no cuáles son éstos. La solución sería aplicar tantas pruebas  $t$  como fuera necesario (según el número de grupos) una vez que se ha obtenido una  $F$  significativa. Sin embargo, este procedimiento presenta problemas de carácter estadístico, pues no proporciona el nivel de significancia proporcionado por el investigador. Con tan sólo tres grupos, ¿cuál es la probabilidad de obtener una  $t$  significativa cuando se consideran todas las pruebas? (entre los grupos 1 y 2; 1 y 3; 2 y 3; o entre los grupos 1 y 2 y también 1 y 3; 1 y 2 y 2 y 3 ; etc.). El problema reside en que tales pruebas  $t$  no son independientes, ya que el valor de  $t$  se encuentra relacionado al computar el valor de  $t$  entre los grupos 1 y 3, pues el grupo 1 está en ambas pruebas  $t$ ; lo mismo sucede con todas las pruebas por pares. Si se tiene un nivel de significancia de 0.05 en cada uno de dos experimentos o comparaciones, la probabilidad no es 0.05, sino de 0.0975 como probabilidad conjunta. La probabilidad de obtener por azar una  $t$  significativa de 0.05 en cada uno o en ambas comparaciones es de 975 veces en 10,000 y no de 5 en 100 como se está postulando. Esto lo explica la siguiente fórmula:

$$p_j = 1 - (1 - \alpha)^k$$

dónde:

$p_j$  = probabilidad conjunta (no individual)

$\alpha$  = nivel de significancia elegido

$k$  = número de experimentos independientes

Así, si  $\alpha = 0.05$  y  $K = 2$ , entonces:  $1 - (1 - 0.05)^2 = 0.0975$ .

Como puede verse, el problema se vuelve más complejo entre más grupos se manejen, pues se requiere obtener después de una  $F$  significativa, pruebas  $t$  para todas las posibles comparaciones entre pares de grupos (con 7 grupos se tendrían que computar 21 pruebas  $t$ ).

Desde luego esto no siempre es necesario, pues el investigador al obtener una **F** significativa sabe que por lo menos existe una diferencia significativa entre dos de sus grupos, y simplemente puede estar interesado en aquellos con medias más altas o más bajas, de acuerdo a su problema de investigación; o bien, puede escoger calcular sólo aquellas pruebas **t** que son independientes; o también identificar con precisión el nivel de significancia "real" de las pruebas **t** que ejecuta.

Es precisamente con respecto a este problema de la prueba **t** múltiple que usualmente se realiza con el análisis de varianza con **k** grupos independientes, en donde sobresale la ventaja de la Prueba de Rangos de Duncan, ya que permite realizar todas las posibles comparaciones entre pares de grupos, sin afectar el nivel de significancia.

### **Diferentes Métodos para la Comparación de Medias.**

La estructura del análisis de varianza simple no permite conocer al investigador cuál o cuáles son los tratamientos responsables del efecto encontrado (F significativa). Toda la información que se obtiene es que existen diferencias que no pueden atribuirse al azar. Si se tienen cuatro medias, no sabemos específicamente entre cuales de éstas se dan las diferencias reales y entre cuales no y cuál es la magnitud de dichas diferencias. Lo que se obtiene es un valor único que representa pruebas simultáneas de la hipótesis que dice que todas las posibles comparaciones entre pares de medias es igual a 0.

Para realizar comparaciones entre las medias se debe de tomar en cuenta la relación que existe entre el proceso de comparación y el error tipo I.

Sabemos que se pueden cometer dos principales clases de error y que el investigador, que se controla la posibilidad de cometer el error tipo I cuando selecciona un determinado valor de alpha. Así con  $\alpha = 0.05$  suponemos que rechazamos la  $H_0$ , cuando de hecho no existen diferencias reales entre las medias un 5% de las veces. Cuando se realiza un gran número de experimentos, por ejemplo: 20 experimentos y utilizamos el mismo nivel alpha del .05 por cada experimento, la probabilidad de cometer el error tipo I por cada uno de dichos experimentos es de 1:20, es decir, una vez en veinte rechazaremos falsamente la  $H_0$ . Podemos conocer el número promedio esperado de errores tipo I que se pueden cometer cuando la  $H_0$  es verdadera.

Este promedio esperado se obtiene multiplicando alpha por el número de experimentos que se conducen. Así; en veinte experimentos el número esperado es  $(.05) (20) = 1.00$ ; si se

conducen cien experimentos esperamos  $(.05)^{100} = 5.00$ , cometer cinco veces el error tipo I. Por lo tanto, si aumentáramos el número incrementaríamos también la cantidad de errores Tipo I.

Este mismo argumento es válido cuando en lugar de hablar de múltiples experimentos nos referimos a uno sólo en el que lo que se realiza son múltiples comparaciones entre pares de medias. A más comparaciones entre pares de medias más veces se cometerá el error tipo I.

Se simbolizamos con ECM la tasa de error tipo I por comparación entre medias entonces con base en lo anterior sabemos que la probabilidad de cometer el error tipo I es de .05 por cada comparación separada que se haga.

El investigador puede planear a priori las comparaciones que le interesan, de hecho, la mayoría de los experimentos se diseñan con hipótesis específicas. De hecho, la  $H_1$  indica las comparaciones a realizar. Por otra parte, las comparaciones planeadas pueden realizarse sea o no significativa la F general del análisis de varianza simple puesto que dichas comparaciones proporcionan la F por comparación que se haga por lo que se puede omitir la F general.

Otra forma de comparaciones son aquellas no planeadas conocidas con el nombre de **comparaciones Post-Hoc**. El investigador después de haber obtenido los datos se interesa por comparaciones específicas. En este caso se evalúan las diferencias entre pares de medias o cualquier combinación entre ellas una vez que la prueba F general ha mostrado significancia estadística.

Cualquiera de estas dos formas de comparación se pueden realizar con el modelo de los efectos fijos (modelo I) o con el mixto (modelo III) pero no con el modelo II, el de efectos al azar. En las comparaciones Post-Hoc se incrementa el error tipo I (ECM) y se trata en forma diferente a la de las comparaciones planeadas. En general, para cualquier comparación, una prueba es más poderosa cuando la comparación es planeada que cuando es Post-Hoc.

**Comparaciones Ortogonales o Independientes.** Teniendo K medias de muestras independientes no puede haber más comparaciones que K-1 si cada comparación debe considerarse independiente tanto de la media total, como entre comparaciones. El número de K-1 comparaciones independientes es también el número de grados de libertad de la varianza entre los grupos. En otras palabras, el número de grados de libertad para la media cuadrada entre los grupos es también el número de posibles comparaciones independientes. Así por cada grado de libertad que haya en la SCE existe una pregunta potencial. Se puede considerar que cualquier análisis de

varianza es equivalente a una partición de los datos en conjuntos de comparaciones independientes. Por otro lado, tenemos que la suma de cuadrados total (SCT) para cualquier K-1 comparaciones independientes es siempre igual a la suma de cuadrados entre K grupos.

Cualquier comparación ortogonal entre pares de medias se obtiene como sigue:

- 1) Dando peso a las medias. Estos pesos se obtienen multiplicando las medias por un conjunto de números o coeficientes  $y$ ,
- 2) Sumando algebraicamente las medias pesadas.

La determinación de la independencia de dos comparaciones entre medias es función de los pesos dados a éstas y no de las medias observadas. También es importante considerar que dichas comparaciones pueden hacerse con: 1) las medias o bien, 2) con las sumatorias (antes de dividir la suma entre el número de casos. Por otro lado, si la suma de los coeficientes es igual a 0 se puede afirmar que las comparaciones son independientes. Esta suma igual a 0 asegura la independencia de la media total o gran media del experimento: Las reglas que deben cubrirse para obtener coeficientes ( $c_i$ ) ortogonales son:

1.  $(\sum c_i = 0) = c_1 + c_2 + \dots + c_k = 0$ .
2.  $(\sum c_i)(\sum c_i^2) = 0$

La solución siguiendo las especificaciones del inciso 1 (la del inciso 2 se revisará adelante) para 3 comparaciones de 4 tratamientos o grupos es:

$$(\sum c_i = 0) = 3 \quad -1 \quad -1 \quad -1 = 0$$

Debe aclararse que también podrían ser:

$$\begin{array}{cccc} 1 & -1/3 & -1/3 & -1/3 = 0 \\ 780 & -260 & -260 & -260 = 0 \end{array}$$

Esto es, la solución puede ser cualquier relación de números, **siempre y cuando la suma sea igual a cero**, la conveniencia de la primera es meramente **práctica**: nos evita trabajar con fracciones o con números grandes.

Ejemplo. Tenemos 4 tratamientos o grupos (A,B,C,D) a los que se les aplicó un análisis de varianza con los siguientes resultados:

Fuente de Varianza	SC	GL.	MC	F	P
Entre Grupos	29.9	3	9.96	5.71	< 0.025
Dentro de los Grupos (error experimental)	14.1	8	1.76		
<b>Total</b>	<b>44.0</b>	<b>11</b>			

TRATAMIENTOS	A	B	C	D
MEDIAS	3	7	10	16

supongamos que el grupo A es el grupo control y que los otros tres grupos (B,C,D) representan diferentes condiciones del tratamiento experimental. Es de esperarse que nos interese comparar el grupo control A contra los niveles experimentales de los otros tres grupos.

COMP. 1 Media A - ( Media B + Media C + Media D / 3 )  
 La Comp. 1 nos mostraría sí hubo o no efecto experimental en general, grupo control vs. grupos experimentales. La siguiente comparación se haría entre grupo experimental B comparado con los grupos C-D:

**COMP. 2. B - (C+ D / 2).**

Por último, la Comp. 3 (puesto que para comparaciones ortogonales la regla establece K-1 comparaciones , en donde K = al número de grupos o de tratamientos) se haría entre los grupos experimentales C y l D:

**COMP. 3 : C-D**

Enseguida desarrollamos los coeficientes para cada comparación:

	A	B	C	D
Comp.1	3	-1	-1	-1=0
Comp.2	0	2	-1	-1=0
Comp.3	0	0	1	-1=0

Los ceros (0) que aparecen a la izquierda del signo igual (=) indican que esa o esas medias en particular no entran en la comparación.

Ahora bien aunque se cuenta con diferentes procedimientos y fórmulas para realizar comparaciones ortogonales, en nuestro ejemplo utilizaremos la fórmula F para computar al mismo tiempo la sumas de cuadrados entre los grupos (SCE) asociada y descompuesta o desagregada para cada comparación de pares de medias.

$$SC_{com} = \frac{[\sum (ci)(\bar{x}_j)]^2}{n[\sum (ci)^2]} \quad (16)$$

En donde :

ci= Coeficientes ortogonales

n= Tamaño de la muestra

Haciendo las sustituciones correspondientes se obtienen las sumas de cuadrados:

$$SC_{comp1} = \frac{[(3)(3) + (-1)(7) + (-1)(10) + (-1)(16)]^2}{3[3^2 + (-1)^2 + (-1)^2 + (-1)^2]} = \frac{576}{36} = 16$$

$$SC_{comp2} = \frac{[(0)(3) + (2)(7) + (-1)(10) + (-1)(16)]^2}{3[0^2 + (2)^2 + (-1)^2 + (-1)^2]} = \frac{144}{18} = 8$$

$$SC_{comp3} = \frac{[(0)(3) + (0)(7) + (1)(10) + (-1)(16)]^2}{3[0^2 + (0)^2 + (1)^2 + (-1)^2]} = \frac{36}{6} = 6$$

En la tabla que sigue ( 5.8.) se exponen las sumas de cuadrados de las comparaciones ortogonales recién hechas junto con los resultados del análisis de varianza de los mismos datos:

Tabla 5.8. Análisis de Varianza y Comparaciones Ortogonales

Fuente de la varianza	SC	gl.	MC	F	P
Entre grupos	(29.9)	(3)			
Comp 1	16.00	1	16.00	9.09	< .025
Comp 2	8.00	1	8.00	4.54	> .05
Comp 3	6.00	1	6.00	3.40	> .05
Dentro de los grupos (error)	14.1	8	1.76		
TOTAL	44.10	11			



Revisando el contenido de la tabla tenemos lo siguiente:

a) las tres comparaciones realizadas constituyen la suma de cuadrados entre grupos del análisis de varianza:

$$SC \text{ comp } 1 + SC \text{ comp } 2 + SC \text{ comp } 3 = SCE =$$

16.00 + 8.00 + 6.00 = 30 (en la tabla aparece 29.9 la diferencia se explica por el problema de redondeo).

b) Los grados de libertad asociados con cada comparación son igual a 1.

Puede verse que las comparaciones siempre se hicieron entre las medias de dos grupos.. La suma total de los gl. es igual al número de gl. de la SCE más los correspondientes a la SCD. Debe hacerse notar que el número de gl., de SCE (k-1) señala el número de comparaciones ortogonales a realizar.

c) La razón de los valores F para las comparaciones realizadas siempre tuvo como denominador la MCD (la media cuadrada o varianza dentro de los grupos) :  $F = MC_{\text{comp. } i} / MCD$  y debe quedar claro también, que esta condición señala una dependencia estadística ( uso del mismo denominador en las diferentes pruebas F) que contradice la supuesta independencia de las varianzas que se analizan a través de este modelo (Análisis de Varianza). Este problema puede reducirse incrementando los grados de libertad en la MCD.

**Decisiones Estadísticas.** Como de costumbre, para cada comparación se plantea la hipótesis nula de las no diferencias y sus correspondientes hipótesis alternas. En la Tabla 5.8 puede verse que sólo la comparación 1 resultó estadísticamente significativa (Comp. 1  $F(1,8) = 9.09, P < 0.025.$ ). Con base en estos resultados se rechaza la  $H_0$ . y se acepta la alterna. En otros términos puede decirse que, si comparamos al grupo control con los grupos experimentales encontramos una diferencia significativa que puede atribuirse al tratamiento. Sin embargo, las comparaciones específicas entre los grupos experimentales no mostraron diferencias significativas:

$$\text{Comp. 2 } F = 4.54 (1,8) \text{ gl. } P > 0.05$$

$$\text{Comp. 3 } F = 3.40 (1,8) \text{ gl. } P > 0.05$$

En ambas comparaciones se acepta la  $H_0$  y se rechaza la  $H_1$ .

Por último, nos falta mostrar a que se refiere la segunda regla que permite asegurar la ortogonalidad de las comparaciones: esto es,  $\sum (c_i) (c_i')$ . En la Tabla 5.9 registremos el conjunto de coeficientes utilizados en una tabla de dos entradas en donde las hileras representan las comparaciones y las columnas los grupos o tratamientos.

TABLA 5.9 Conjunto de Coeficientes Ortogonales.

TRATAMIENTOS					
Comparaciones	A	B	C	D	$\sum(C_i)$
1	3	-1	-1	-1	0
2	0	2	-1	-1	0
3	0	0	1	-1	0

• Se ha dicho que la ortogonalidad o independencia de las comparaciones está dada por los coeficientes que tienen como función ponderar las medias. Uno de los criterios aparece en la 5ª. columna de la Tabla 5.10 que indica que **la suma de los coeficientes i debe ser igual a 0** (se refiere a la independencia dentro de cada comparación) mientras que:  $\sum (C_i) (c_i') = 0$  se refiere a la independencia entre comparaciones, en donde: **ci** se refiere a los coeficientes de una comparación determinada y **ci'** a los de las otras comparaciones.

Si se toman las comparaciones de la Tabla 5.10: 1 y 2; 2 y 3; y 1 y 3 puede comprobarse la presencia positiva del segundo criterio:

$$C1 C2 = (3)(0) + (-1)(2) + (-1)(-1) + (-1)(-1) = 0$$

$$C2 C3 = (0)(0) + (2)(0) + (-1)(1) + (-1)(-1) = 0$$

$$C1 C3 = (3)(0) + (-1)(0) + (-1)(1) + (-1)(-1) = 0$$

**Comparaciones Múltiples Post-Hoc.**

A diferencia de lo que sucede con las comparaciones planeadas que pueden llevarse a cabo directamente, (sin conocer el valor de la F general), en el caso de las comparaciones Post-Hoc es necesario comprobar primero que el valor de la F es significativo, para entonces poder llevar a cabo todas las posibles comparaciones que se deseen. Por lo general, el investigador recurre a las comparaciones Post-Hoc cuando su interés se centra en la formulación de nuevas hipótesis. Llevar a cabo múltiples comparaciones, (todas las posibles combinaciones entre pares de medias), es utilizar uno de los mejores medios para extraer la máxima información de los datos. Las comparaciones Post-Hoc, a diferencia de las comparaciones planeadas no requieren de independencia en y entre comparaciones.

**Problemas Relacionados con la Tasa de Error.** En las Comparaciones Múltiples. Cada uno de los procedimientos para realizar comparaciones múltiples tiene sus propios medios para solucionar el problema del Error de las Comparaciones Múltiples (ECM). Este problema, se refiere a que entre más comparaciones se realicen más alta será la probabilidad de cometer errores tipo I. Una solución común a este problema es reducir el tamaño de la región de rechazo.

Para llevar a cabo todas las comparaciones posibles entre las medias se tienen entre otras pruebas, la de Tukey y la de Scheffé. La construcción de estas pruebas toma en cuenta la reducción del error por comparación y esto está previsto tanto para si se realizan o no todas las posibles comparaciones. La prueba de Duncan que ya revisamos y la de Newman-Keuls llevan a cabo el ajuste del ECM dependiendo del número de medias involucradas.

Se considera en general, que la prueba de Tukey y la de Scheffé, son las más estrictas y las que ofrecen mayor protección contra el error tipo I mientras que las pruebas de Duncan y Newman-Keuls ofrecen una protección intermedia.

El método de Scheffé ofrece múltiples ventajas al investigador ya que se le considera como relativamente insensible a problemas relacionados con distribuciones no-normales así como a los de la no-homogeneidad. De manera simple y resumida lo que se hace con esta prueba es utilizar un valor F específico para evaluar las comparaciones. Con este procedimiento, que enseguida se describe, se lleva a cabo, precisamente el ajuste para disminuir ECM. La F específica o especial se obtiene como sigue:

$$F' = (k - 1) F .05 (GLE, GLD)$$

En donde:

F' = F específica para las comparaciones hechas.

$k-1$  = número de tratamientos menos uno;

$F_{.05}$  = al valor tabulado de  $F$  con un nivel de significancia del .05;

GLE Y GLD = grados de libertad del análisis de varianza del experimento:

GLE para el numerador y GLD para el denominador.

Haciendo la sustitución se tiene entonces:  $F' = (4 - 1) F_{.05}(3,8) = 3(4.07) = 12.21$

El valor de  $F' = 12.21$  es el valor crítico que demanda la prueba de Scheffé. Es un valor crítico más estricto (por la reducción del tamaño de la región de rechazo de la  $H_0$ ) ya que una  $F$  al 0.05 sólo requiere un valor observado de  $F(3,8) = 4.07$  y la de Scefefeé es de 12.21.

**Pruebas de Newman-Keuls y Duncan.** La prueba de Duncan que revisaremos no requiere llevar a cabo todo el proceso de análisis antes descrito, sólo se aplica a las medias que se quieren comparar. La exposición que haremos se ajusta a los lineamientos seguidos por Keppel (1973) por encontrarlos claros y altamente didácticos. Utilizamos los mismos datos del ejemplo de las comparaciones ortogonales y los presentamos en la tabla (Tabla 5.10) en donde los totales de los tratamientos se registran en orden ascendente a lo largo de la primera hilera y columna. Los elementos dentro de la parte superior de la diagonal principal son las diferencias entre las sumas de los tratamientos (la suma de la columna menos la suma de la hilera). Aparecen otras tres columnas a la derecha de las diferencias entre los tratamientos que son los valores que se registran para evaluar la significancia de las diferencias (comparaciones). Así, tenemos:

$r$  = al número de grupos involucrados en cada comparación.  $r = 4$ . resulta del número de sumas en la hilera A más la que se eliminó en la diagonal.

$RC_{n-k}$  = rangos críticos ( $\alpha = 0.05$ ) para cada valor de  $r$  de acuerdo con la prueba de Newman-Keuls.

$RCD$  = Rangos críticos ( $\alpha = .05$ ) para cada valor de  $r$  de acuerdo con la prueba de Duncan.

Tabla 5.10. COMPARACIONES MULTIPLES CON LAS TECNICAS DE NEWMAN-KEULS Y DUNCAN.

	A	B	C	D		
SUMAS DE	3	7	10	16	$r$	$RC_{n-k}$ $RC_d$
TRATAMIENTOS						

A	3	-	4	7	13**4	10.37	7.94	
B	7		-	3	9**3	8.08	7.76	
C	10			-	6	2	7.46	7.46
D	16				-			

\*\*Valores Significativos con ambas pruebas ( Newman-Keuls y Duncan).

Obtención de los rangos críticos:

Fórmula:  $RC_{n-k} = q(r, glD) \sqrt{n(MCD)}$

En donde:

r = número de grupos involucrados en cada comparación.

glD = grados de libertad dentro de los grupos, valor que se obtiene del análisis de varianza correspondiente;

n = número de sujetos de cada grupo del experimento.

MCD = Media cuadrada (varianza) dentro de los grupos.

$$\sqrt{n(MCD)} = \sqrt{3(1.76)} = 2.298$$

El valor 2.298 es una constante por la que se multiplica el valor q.

La tabla de los valores q, está compuesta por columnas que indican diferentes números de grupos, e hileras en donde aparecen los glD y valores de alpha al .05 y al .01.

Así; para 8 gl y al .05:

$$r_4 = 4.53 \times 2.29 = 10.37 = RC_{n-k}$$

$$r_3 = 4.04 \times 2.29 = 8.08 = RC_{n-k}$$

$$r_2 = 3.26 \times 2.29 = 7.46 = RC_{n-k}$$

Los cálculos para los rangos críticos de la prueba de Duncan son los mismos, sólo que los valores de r se buscan en la tabla de los valores de la prueba de rangos de Duncan. Estos valores con 8 gl y al .05 son:

$$r_4 = 3.47 (2.29) = 7.94 = RCD$$

$$r_3 = 3.39 (2.29) = 7.76 = RCD$$

$$r_2 = 3.26 (2.29) = 7.46 = RCD$$

Evaluación de las comparaciones: La Tabla 5.11 contiene toda la información necesaria para proceder a la aplicación de las pruebas estadísticas. La evaluación se inicia con la diferencia (entre los tratamientos) más grande. En la Tabla 5.11 puede verse que ésta se encuentra en la parte superior de la esquina derecha con un valor igual a 13. Cuando la

diferencia más grande resulta significativa se continúa probando las diferencias que aparecen a la izquierda de ésta en la misma hilera. Si la diferencia más grande no resulta significativa no tiene caso ver las otras de esa misma hilera, se pasa inmediatamente a la siguiente: nuevamente se inicia por la diferencia más grande y después a la izquierda dentro de esta misma hilera y así sucesivamente.

Principiaremos por la prueba de Newman-Keuls. La diferencia más grande (13) excede a  $RC_{n-k} = 10.37$  por lo tanto la diferencia entre los tratamientos A y D es significativa. Avanzamos hacia la izquierda y encontramos una diferencia igual a  $7 < RC_{n-k}$ . Siendo esta diferencia no significativa nos movemos al siguiente renglón comenzando de derecha a izquierda. La diferencia más grande de esa hilera es (9)  $> RC_{n-k}$ , por lo tanto, la diferencia entre B y D es significativa y se procede en esta forma hasta terminar con los valores (diferencias) de la tabla. El mismo procedimiento se sigue para probar las diferencias para la prueba de Duncan. El criterio general es que los valores de las diferencias entre los tratamientos excedan a los correspondientes valores de los rangos críticos para poder rechazar la  $H_0$ . Nótese que la comparación **B Vs. D** (Comp. 2 de la Tabla 4) no fue significativa con la prueba ortogonal, mientras que con las pruebas de Newmann-Keuls y de Duncan es posible decidir que si existe diferencia significativa. La razón, como ya se señaló anteriormente, es que estas últimas son menos estrictas y por lo tanto más vulnerables al error Tipo I.

Existen además otras pruebas entre ellas están la de las diferencias mínimas significativas (LSD), con la que **no** hay protección contra el EMC. Otra es la prueba de Bonferroni que es una modificación de la anterior, basada en un procedimiento de comparaciones múltiples entre todos los pares de grupos por medio de una prueba t. Esta ajusta el nivel de significancia observado, el ECM, aunque de manera más laxa que la pruebas de Newman Keuls y la de Duncan. Por último debe señalarse que no sólo existen otras pruebas sino también variaciones de las aquí revisadas (Ver por ejemplo, Norusis, 1992) de tal manera que es posible encontrar aquella o aquellas que mejor se ajusten a las necesidades específicas de investigación

#### **5.7.2. Prueba Paramétrica Para k Muestras Relacionadas: Análisis de Varianza de Muestras con Puntajes Relacionados (F).**

El procedimiento que se revisará es una combinación del análisis de varianza y de la correlación, en donde la varianza total se descompone en:

1. varianza entre los grupos (diferencias debidas al efecto del tratamiento o variable dependiente);
2. diferencias sistemáticas entre los Sujetos (corresponde a la porción correlacionada); y
3. varianza de error.

El proceso de computación es igual al del análisis de varianza simple: se requiere de una  $Sc^{total}$  con la única diferencia de que se sustrae adicionalmente una SC que corresponde a la sumas de los puntajes (valores de la variable dependiente) que obtuvo cada Sujeto en cada tratamiento.

En este caso, el análisis de la varianza se parte o divide en un componente mas que corresponde a la varianza que muestran los sujetos a lo largo de los diferentes tratamientos a los que se ven sometidos. Por lo tanto, se puede ver que la varianza total queda constituida por la varianza experimental o explicada por los efectos de la variable independiente, la varianza intrasujetos, y la varianza de error, dentro o residual. La suma de cuadrados de error, que es aquella a partir de la cual se obtiene el valor del denominador de la razón F, es igual a la  $Sc^{total}$  menos la  $Sc_{entre}$  y menos la  $Sc_{sujetos}$ , de tal manera que se sustrae el efecto de los tratamientos sobre los sujetos, modificándose así el valor del denominador de la razón F.

*Ejemplo.* A un grupo de niños se les aplicó tres veces un tratamiento: en una ocasión se les pidió que se autoevaluaran ( $X_n$ ); en otra que evaluaran a su mamá ( $X_m$ ); y en la última que evaluara a su papá ( $X_p$ ). El instrumento estaba constituido por 19 adjetivos, y al lado de cada uno de éstos aparecía un dibujo con cinco escalones. El niño tenía que poner una marca en el primer escalón si consideraba que tenía poco de una adjetivo dado (puntaje de 1) o en el último si consideraba que tenía mucho de él (puntaje de 5); si la marca era puesta en el segundo, tercero o cuarto escalones, los puntajes asignados eran 2, 3 y 4 respectivamente. En suma, cada adjetivo aplicado a él, a la mamá o al papá, podía tener valores del 1 al 5; los puntajes más altos implicaban mejor evaluación (propia y de los padres).

Debido a que el mismo Sujeto fue sometido a tres condiciones o estímulos diferentes, el investigador utilizó en la aplicación de los instrumentos el contrabalanceo, de tal manera que los Sujetos respondieron a los mismos en órdenes diferentes (por ejemplo:  $X_n, X_m, X_p$ ;  $X_m, X_p, X_n$ ;  $X_p, X_m, X_n$ ; etc.). Uno de los problemas que le interesaba responder era ver si había o no diferencias entre las evaluaciones hechas por los Sujetos. De esta manera, decidió utilizar una análisis de varianza para puntajes relacionados o dependientes.

Supongamos que se obtuvieran los resultados que aparecen en la Tabla No. 5.11.:

Tabla 5.11. Tabla Sumaria del Análisis de Varianza con Puntajes Correlacionados

Fuente de Variación	SC	gl	MC	F	p
SC <sub>e</sub> (trat.)	364.660	2	182.330	7.814	<0.01
SC <sub>s</sub> (suje.)	478.750	11	43.523	1.865	>0.05
SC <sub>d</sub> (error)	513.340	22	23.334		
Total	1356.750	35			

Como puede verse en la Tabla Sumaria (5.11), los valores de **F** se obtuvieron dividiendo las MC de los tratamientos y de los Sujetos entre la MC del error. De esta manera, para probar las **F's** obtenidas se localiza en la Tabla de **F** los valores que caen en los puntos 2 y 22 gl y 11 y 22 gl. Con base en los resultados puede decirse que los puntajes o medias entre los tratamientos difieren significativamente, siendo la evaluación de la madre la más alta, aunque prácticamente igual que la del padre, y la autoevaluación la más baja. De hecho, la significancia de **F** está dada por la diferencia entre cualquiera de las evaluaciones del padre o de la madre, en comparación con la que los niños hacen de sí mismos.

Así mismo, se observa que la F asociada al efecto de las tres mediciones o registros efectuados por los sujetos investigados, no es estadísticamente significativa, ya que tiene un valor asociado de  $p > 0.05$ .

Es pertinente señalar que el uso de este diseño requiere de una estricta revisión metodológica. Es decir, recurrir a las estrategias de control que disminuyan el acarreo de efectos presente en los diseños en los que los mismos Sujetos se someten a dos o más estímulos o mediciones.

#### **5.7.4. Prueba Paramétrica para el Diseño de k Variables Independientes: F. Diseño Factorial.**

Hasta ahora sólo se han revisado diseños en los que se trata de determinar el efecto de una variable independiente, con dos o más valores, sobre otra dependiente. Pero al investigador le puede interesar averiguar el efecto que más de una variable independiente tiene sobre una dependiente. La necesidad de



estudiar y descifrar complejos de interrelaciones entre variables ha contribuido al desarrollo de los diseños factoriales. Estos diseños se manejan con análisis de varianza, por lo que su aplicación es efectiva cuando los datos se ajustan a los supuestos de la estadística paramétrica. El diseño factorial requiere de un número de casos iguales o proporcionales en cada celdilla, para separar los efectos de una variable de las otras. Estos diseños producen más información acerca de las relaciones entre las variables que ningún otro tipo de diseño. Los casos se usan una y otra vez para hacer comparaciones para cada factor en turno, y para las combinaciones e interacciones entre variables. En vez de estudiar variable por variable, mientras se controlan todas las otras el diseño factorial permite estudiar varias variables experimentales en combinación. No sólo provee más información, sino que aumenta las probabilidades de predicciones de los resultados bajo diferentes circunstancias. Este diseño requiere de un análisis cuidadoso de los casos. En algunas ocasiones el número de casos para algunas celdillas no llenan los requerimientos necesarios, de tal modo que se tiene que recurrir a algunos métodos de extrapolación estadística o replicación al azar.

El diseño factorial no pretende ser representativo, pues en su forma final es marcadamente no-representativo, debido a que se tienen que tener números iguales o proporcionales de Sujetos, en cada celdilla de tal modo, que aquellas condiciones raras (poco frecuentes) de determinantes están presentes en la misma proporción que aquellas combinaciones típicas. El diseño factorial también requiere, como otros diseños, del control de variables, lo cual se realiza mediante exclusión, constancia o por aleatorización; sin embargo, siempre quedan algunas variables que introducen confusión, por lo que es aconsejable que en las últimas etapas del análisis, las hipótesis de dichas variables se exploren por otros medios: tabulaciones cruzadas, apareamiento de submuestras, o por análisis de regresión.

El hecho de que el diseño factorial permita manejar simultáneamente  $k$  variables independientes resulta de gran utilidad, ya que en la realidad la conducta responde a múltiples causas, es decir, son diversas variables independientes las responsables de su aparición, producción o modificación. Además, debe tomarse en cuenta que estos conjuntos de variables responsables de una conducta o evento pueden tener cada una de ellas diferentes valores. Estas condiciones, varias variables independientes con diferentes valores cada una, actuando simultáneamente sobre la variable dependiente se resuelven con los diseños factoriales. Debe hacerse notar por otra parte, que no deben confundirse estos diseños con el análisis factorial o de factores, que responde a otro procedimiento.

Suponga que un investigador desea saber en que forma influyen diferentes variables independientes sobre el uso de cierto fármaco. Por ejemplo, el investigador piensa que para poder explicar la conducta *uso de fármacos*, tiene que tomar en cuenta variables tales como: nivel socioeconómico de los Sujetos; grado de integración familiar, educación formal alcanzada, edad y sexo. Considera además, que analizando dichas variables en sus diferentes niveles o valores, obtendrá mayor información sobre el problema que le interesa. Por lo tanto, decide estudiar la variable nivel socioeconómico con tres valores diferentes: bajo, medio y alto; la variable educación con cuatro: primaria, secundaria, preparatoria y profesional; la variable integración familiar con tres niveles: buena, regular y pésima; la variable edad con dos valores: jóvenes (de 25 a 35 años) y adultos (de 36 a mayor edad); y la variable sexo, en sus dos valores: masculino y femenino.

Como primer paso el investigador de este ejemplo tendría que definir las variables y las categorías de cada una de ellas. Esto lo haría usando diversos indicadores. Una vez que ya las tuviera definidas tendría que determinar el número de Sujetos necesarios para manejar todas las posibles combinaciones de variables en sus diferentes valores y formar así sus grupos. Este problema lo resuelve el tamaño del diseño factorial que utilizará y está dado por el número de variables y sus niveles. El tamaño del diseño factorial de este ejemplo es de  $3 \times 4 \times 3 \times 2 \times 2 = 144$ . Es decir, sus cinco variables independientes: la primera (nivel socioeconómico) con tres niveles; la segunda (educación) con cuatro; la tercera (integración familiar) con tres; y las dos últimas (edad y sexo), con dos niveles cada una. La multiplicación de estos números le darán el número de grupos que necesitará: es decir, un total de 144 grupos.

Cada grupo requiere como mínimo de 10 Sujetos, así, el investigador requeriría 1440 Sujetos. Cada uno de los 144 grupos (representados por cada celdilla de la Figura 5.12.) constituye una combinación de las diferentes variables con sus respectivos valores o niveles.

PRIMARIA			SECUNDARIA			PREPARATORIA			PROFESIONAL			ESCOLARIDAD		
B	R	P	B	R	P	B	R	P	B	R	P	INTEGR.	FAM.	
												NSE	EDAD	SEXO
1	2	3	4	5	6	7	8	9	10	11	12	B	J	F
											24	M	J	F
											36	A	J	F
											48	B	A	F
											60	M	A	F
											72	A	A	F

											84	B	J	M
											96	M	J	M
											108	A	J	M
								117			120	B	A	M
											132	M	A	M
											144	A	A	M

Figura 5.12. Esquema de un diseño factorial de 3 x 4 x 3 x 2 x 2.  
Acotaciones:

Sexo: M = masculino; F = femenino

NSE: Nivel socioeconómico: B =bajo; M = medio; A = alto

Edad: J = jóvenes; A = adultos

Integr. Fam.: Integración Familiar: B=buena; R=regular; P=pésima.

Como puede verse en el esquema, éste indica las características que deberán tener los Sujetos (la combinación de variables y sus niveles) que componen los diferentes grupos. Por ejemplo, los Sujetos del grupo 117 deberán tener las siguientes características: sexo masculino, adultos, de nivel socioeconómico bajo, con estudios de preparatoria y una pésima integración familiar.

Como puede inferirse del diseño utilizado en el ejemplo, a mayor número de variables y de niveles, mayor complejidad, así como mayor número de Sujetos. Por lo general se usan diseños factoriales más sencillos que el expuesto, pero en teoría, el número de variables independientes a investigar puede ser infinito.

Este tipo de diseños responde a tantas preguntas como variables independientes haya, más una pregunta más (para cada una de las posibles combinaciones de las mismas) que se refiere a si existe o no interacción o relación entre las diversas variables independientes, en su efecto sobre la dependiente. Respecto al diseño del ejemplo, se responderían las siguientes preguntas:

Con relación a la variable dependiente, esto es, con el consumo de drogas:

1. ¿hay diferencia o no entre los tres niveles socioeconómicos?

2. ¿hay diferencia o no entre los cuatro niveles educativos?

3. ¿hay diferencia o no entre los tres grados de integración familiar?

4. ¿hay diferencia o no entre las dos edades?

5. ¿hay diferencia o no entre los sexos?

Interacciones:

¿Hay interacciones o no entre las variables independientes? y si las hay, ¿de qué manera afectan el consumo de drogas?, como se tienen cinco variables, se harían muchas preguntas de interacción. Si a cada variable se le asigna una letra (A,B,C,D,E) las preguntas se harían sobre las siguientes interacciones tomando a las variables de dos en dos: AxB; AxC; AxD; AxE; BxC; BxD; BxE; CxD; CxE y DxE; también se podría ver si hay interacción tomando a las variables de tres en tres, de cuatro en cuatro, y por último, tomando a todas ellas juntas: AxBxCxDxE.

*Ejemplo.* Supóngase que un investigador asume que en la realización de una tarea (solución de problemas abstractos) intervienen las siguientes variables:

A = Autoestima (AE): baja ( $A_1$ ) y alta ( $A_2$ )

B = Tiempo: limitado ( $B_1$ ) y libre ( $B_2$ )

C = Sexo: femenino ( $C_1$ ) y masculino ( $C_2$ )

Como puede verse, el investigador está manejando un diseño factorial de  $2 \times 2 \times 2$ . Esto es, cada número señala los factores (AE, Tiempo, y Sexo), y la numerosidad de cada uno expresa los niveles de cada factor (AE: baja y alta; Tiempo: limitado y libre; Sexo: masculino y femenino). El producto de ellos ( $2 \times 2 \times 2 = 8$ ), indica el número de combinaciones o tratamientos experimentales que se investigarán.

Para proceder al análisis se necesita obtener las sumas de cuadrados (SC):

1) Suma de cuadrados total (SCT); Suma de cuadrados entre tratamientos (SCE); y Suma de cuadrados dentro de los tratamientos (SCD).

*Estimación de la Homocedasticidad.* Se puede hacer una estimación de la homogeneidad de la varianza (homocedasticidad) para saber si los datos de las diferentes condiciones provienen de la misma población.

Si las ocho varianzas difieren grandemente entre sí, es aconsejable aplicar una prueba de homogeneidad de varianza (v. gr. Prueba de Levene); si las diferencias son pequeñas, puede concluirse que ésta existe.

Es aconsejable realizar dicha prueba, ya que la  $F$  que se maneja en los diseños factoriales del análisis de varianza es sensible a la heterogeneidad de la varianza, especialmente cuando el número de Sujetos no es el mismo en cada condición experimental o tratamiento. En cambio, cuando se tiene el mismo número de

Sujetos, la homogeneidad inicial permite asumir que los efectos resultantes se deben a los tratamientos experimentales (o variables independientes) que se están manejando.

2) Prueba de Significancia de las Medias Cuadradas o Varianzas del Tratamiento.

Como paso siguiente se prueba si existen diferencias estadísticamente significativas en la varianza de los tratamientos. Se recurre a la prueba **F** que como se sabe, es una prueba paramétrica. Sin embargo, antes se necesitan determinar los grados de libertad (gl).

Hasta el momento se ha hecho una partición de la suma total de cuadrados, es decir:  $SCT = SCE + SCD$ . A esta descomposición de la SCT le corresponde también una partición de los grados de libertad: así, con la SCE se asocian  $k - 1$  grados de libertad.

Si en un ejemplo se tienen 8 tratamientos, si  $k =$  al número de tratamientos, entonces  $k - 1 = 8 - 1 = 7$  gl. Para la SCD, los grados de libertad se obtienen con  $k(n - 1)$ ; si  $k = 8$  y  $N = 10$  (número de observaciones o de Sujetos en cada tratamiento), entonces,  $k(n - 1) = 8(9) = 72$  gl; por último, a la SCT le corresponden  $N - 1$  grados de libertad; si  $N = 80$ , entonces  $N - 1 = 79$  gl.

En la Tabla 5.13 aparecen los resultados en la forma de un análisis sumario que proporciona toda la información de todas las operaciones hasta aquí realizadas.

-----

--

Tabla 5.13 Análisis sumario de la obtención de **F**

-----

--

Fuente de Variación	Suma de Cuadrados (SC)	Grados de Libertad (GL)	Media Cuadrada (MC)	F
Entre los Tratamientos	195.6	7	27.94	22.72
Dentro de los Trata- mientos	88.4	72	1.23	
Total	284.0	79		

-----

--

-----  
--  
En la Tabla 5.13 se han registrado en la primera columna los valores de las SC entre y dentro de los grupos. En la columna siguiente aparecen los grados de libertad correspondientes a cada SC. Las medias cuadradas o varianzas que aparecen en la tercera columna, se obtienen dividiendo cada suma de cuadrados entre sus respectivos grados de libertad. Finalmente, el valor de **F** es la razón resultante de la división de las medias cuadradas.

Para saber si el valor obtenido de **F** es significativo se consulta una tabla de valores **F** (véase por ejemplo a McGuigan, 1993), localizando el valor tabulado en el punto donde intersectan los grados de libertad del numerador (7) y del denominador (72) con el nivel de significancia que previamente se haya establecido ( $\alpha = 0.05$ ). En este punto de confluencia se encuentra que  $F_{0.05} = 2.32$ . Dado que el valor obtenido de **F** es  $F = 22.72$ , puede concluirse que  $F 22.72 > F_{0.05} = 2.32$  con  $p < 0.01$ . Es decir, el valor encontrado de **F** es altamente significativo, más allá de  $p = 0.05$ . En base a esta prueba puede decidirse que los tratamientos difieren significativamente.

3) Análisis de los Efectos Principales. Los efectos principales se refieren a los efectos de los factores que se están manejando. En el ejemplo son tres, cada uno con dos niveles, por lo tanto se comparan los niveles de cada uno: comparación de las sumas de cuadrados del factor A (SCA), es decir de la autoestima (alta y baja); la suma de cuadrados del factor B (SCB), tiempo limitado y tiempo libre; la suma de cuadrados del factor C (SCC), sexo de los Sujetos, masculino y femenino.

Debe entenderse que esta suma de los cuadrados de los factores y sus niveles forman parte de la suma de cuadrados entre los tratamientos (SC<sub>entre</sub>). Si como ya se señaló a la SCE se asocian 7 gl, entonces a cada comparación entre los niveles de cada factor corresponde 1 gl.

Las medias cuadradas o varianzas obtenidas para los niveles de los factores A, B y C, se conocen como los efectos principales de esos factores. Por otra parte, para obtener dichas MC se han ocupado 3 grados de libertad, de los 7 asociados a la Suma de Cuadrados Entre los grupos (SC<sub>entre</sub>).

4) Análisis de los Efectos de Interacción. Si se tienen tres factores, el análisis de la interacción entre ellos se realiza tomando en cuenta sus posibles comparaciones (con sus correspondientes niveles). De esta manera, las interacciones que

se analizarán para el caso del ejemplo que se está desarrollando son:  $AxB$ ;  $AxC$ ;  $BxC$ ; y  $AxBxC$ .

Teniendo las sumas de cuadrados se procede a la solución de las interacciones de los factores.

Para transformar las Sumas de Cuadrados de las interacciones a Medias Cuadradas o varianzas, se dividen entre sus correspondientes grados de libertad. La regla general para obtener éstos, en el caso de las interacciones, es multiplicar los grados de libertad asociados con los factores que intervienen en las mismas. Como ya se señaló se tiene un grado de libertad para cada factor; de esta manera, a cada interacción le corresponde 1 gl.

Debe recordarse que de los 7 gl de la SCE ya se ocuparon 3 para los efectos principales (SCA, SCB, y SCC). De los 4 restante, 3 más corresponden a la suma de cuadrados de las interacciones realizadas; el grado de libertad restante corresponde a la suma de cuadrados de la interacción faltante:  $AxBxC$ .

Se obtiene por medio de una simple substracción: debido a que la Suma de Cuadrados Entre los grupos de tratamientos (SCE) es igual a la suma de las sumas de cuadrados de los factores (SCA, SCB y SCC) y de las sumas de cuadrados de las interacciones ( $AxB$ ,  $AxC$ , y  $BxC$  y  $AxBxC$ ), entonces la suma de cuadrados de  $AxBxC$  se obtiene restando a la SCE las seis sumas de cuadrados ya obtenidas.

En la Tabla 5.14 se proporciona una Tabla Sumaria de un análisis factorial de varianza.

-----  
--

Tabla 5.14 Sumario del Análisis de Varianza de un diseño factorial de  $2 \times 2 \times 2$  ( $n = 10$  y  $N = 80$ )

-----  
--

Fuente de Variación	Suma de Cuadrados (SC)	gl	Medias Cuadradas (MC)	F
---------------------	------------------------	----	-----------------------	---

-----  
--

EFFECTOS PRINCIPALES:

A. Autoestima	88.20	1	88.20	71.71
B. Tiempo	28.80	1	28.80	23.41
C. Sexo	61.26	1	61.26	49.80

INTERACCIONES:

AxB	0.80	1	0.80	
AxC	11.25	1	11.25	9.15
BxC	4.05	1	4.05	3.29
AxBxC	1.24	1	1.24	1.01

ERROR:

(dentro de los tratamientos)

88.40                      72                      1.23

TOTAL                      284.00                      79

-----  
--

Los datos de la Tabla 5.14 se explican por sí mismos; los valores **F** son el resultado de la razón de dos medias cuadradas o varianzas: el resultado de la división de cada media cuadrada entre 1.23 que es el valor de la Media Cuadrada del Error. Es obvia la razón por la cual no se calculó el valor **F** para la interacción AxB (el más pequeño el valor de ésta, que el de la varianza del error).

5) Una vez que se tienen los datos de la Tabla 5.4 se procede a la decisión estadística; es decir, probar si los valores de **F** obtenidos son significativos al nivel de significancia previamente establecido. Supóngase que este nivel se escogió con  $\alpha = 0.05$ . Como ya se ha visto a lo largo de los ejemplos descritos, para este paso se debe recurrir a la tabla de los valores tabulados, que en este caso es la Tabla de Valores **F**. Como **F** es la razón entre dos varianzas, cada una de estas tiene sus correspondientes grados de libertad. Todas las varianzas del numerador tienen un gl, mientras que a la del error le corresponden 72 gl. Se localiza en la Tabla el valor que le corresponde a **F** con (1,72) gl al nivel de significancia de  $\alpha = 0.05$  (ver McGuigan, 1993). Se encuentra que se requiere una **F** = > 4.0. En la Tabla 5.11 puede verse que todas las **F** de los efectos principales (A, B, C) son altamente significativos, más allá de 0.05. Con respecto a las interacciones sólo AxC fue significativa ( $F = 9.15 > F_{0.05} = 4.0$ ), mientras que AxB, BxC y AxBxC resultaron no significativas.

6) Habiendo hecho la decisión estadística, esto es, decidido si las **F** son o no significativas y consecuentemente aceptado o



rechazado respectivamente las hipótesis nulas ( $H_0$ ), lo que queda es interpretar los resultados obtenidos.

*Efectos Principales.* Partiendo del hecho de que las F de éstos fueron altamente significativas, puede concluirse que los factores de AE (autoestima), tiempo y sexo influyeron significativamente en la tarea de solución de problemas abstractos (variable dependiente). Sin embargo, debido a que tales factores tienen cada uno dos niveles, esos valores F son globales y aportan de hecho, poca información, de aquí que se requiera analizar los valores de cada factor en sus dos niveles promediados con los niveles de los otros factores.

Existen dos posiciones al respecto: una que considera que cuando una interacción resulta significativa, no se deberán interpretar los efectos principales de los factores que intervinieron en dicha interacción. En otros términos, sólo se interpretarán los efectos principales de aquellos factores que cuando interactúan no son significativos (ver Tabla 5.4). La otra posición no impone dicha restricción, se interpretan los factores principales (siempre y cuando sean estadísticamente significativos) aún cuando los factores de éstos se encuentren involucrados en interacciones que también resultaron significativas. De acuerdo con la primera posición, los efectos principales de los factores A y C no se interpretarían, puesto que el valor de la variable dependiente se debió sólo a la influencia conjunta (interacción) de A y C (que resultó significativa).

Siguiendo adelante con el análisis de interpretación, debe considerarse que la Media Cuadrada del Factor A corresponde a la comparación de los dos niveles de Autoestima, baja y alta, promediados sobre los dos niveles de los factores B y C. Puede verse que  $X_{a1} = 4.20$  vs.  $X_{a2} = 6.30$ . Esto significa que los Sujetos de la condición de alta autoestima se desempeñaron mejor en la tarea que los de baja autoestima. Debe recordarse que en el análisis de varianza, los efectos del tratamiento experimental o de la variable independiente se refleja en las medias. Relacionando este resultado con las medias del factor B, se tiene que  $X_{b1} = 5.85$  y  $X_{b2} = 4.65$ ; es decir, cuando se limitó el tiempo de ejecución hubo mejores resultados en la variable dependiente. Entonces, puede concluirse que los Sujetos de alta autoestima con tiempo limitado tienen un mejor desempeño que los Sujetos con baja autoestima y tiempo libre.

La misma lógica se sigue con el factor C. Cuando  $X_{c1} = 6.13$  vs.  $X_{c2} = 4.38$ , que significa que las mujeres se desempeñaron mejor que los hombres. Así, las conclusiones obtenidas hasta aquí son: en la solución de problemas abstractos las mujeres con alta

autoestima y con límite de tiempo para la solución de tales problemas obtuvieron un desempeño superior comparado con el de los hombres, de baja autoestima y con el tiempo libre en la solución de dichos problemas. En otras palabras, lo que se sabe hasta esta parte del análisis es lo referente a los efectos principales de los factores en relación a sus niveles.

Factor A:  $A_2 > A_1$

Factor B:  $B_1 > B_2$

Factor C:  $C_1 > C_2$

Sin embargo, falta el análisis de las interacciones para concluir con la interpretación de los resultados. En la Tabla Sumaria (5.4) se encuentra que la F de AxB no es significativa ( $0.8/1.22 = 0.65$ ) Este resultado significa que las diferencias entre los niveles de A son independientes de las diferencias en los niveles del factor B. Es decir, que la autoestima de los Sujetos, ya sea alto o baja, no afectó o es independiente, de si se fija o no un tiempo límite en la solución de la tarea.

La F significativa de AxC implica que las diferencias entre  $A_1$  y  $A_2$  no son independientes de los niveles de C, y lo inverso también es cierto. En otros términos: la autoestima alta o baja sí se relaciona con el sexo de los Sujetos en la realización de la tarea:

Autoestima alta con mujeres y baja con hombres, siendo las primeras mejores en cuanto a la solución de los problemas, que los segundos. Una interacción significativa quiere decir que los valores de la variable dependiente se deben a la acción conjunta de los factores involucrados en ella, por lo que pierde sentido explicar la variable dependiente tomando cada uno de esos factores por separado.

Por último, con respecto a la interacción AxBxC que resultó no ser significativa, puede decirse que cuando se analiza la influencia de los tres factores en conjunto sobre la variable dependiente, éstos producen su efecto por separado, independientemente. Al analizar la acción conjunta de los tres factores se anula, por así decirlo, la acción de AxC. Esta es efectiva sólo así, como combinación de los efectos específicos de Autoestima y Sexo, pero no cuando se analizan los tres factores juntos. Lo anterior es congruente si se ve que AxB y BxC tampoco resultaron ser significativas. Entonces, un desempeño superior en la solución de problemas abstractos se relaciona con una alta autoestima y con el sexo de los Sujetos (mujeres). La interpretación del factor B se hace en términos de efecto principal dado que fue significativo. Siendo  $X_{b1} > X_{b2}$  puede decirse que la restricción del tiempo (imponer un tiempo

límite) en la solución de la tarea proporciona mejores resultados que cuando se deja un tiempo libre (desde luego, la interpretación irá respaldada por la teoría en la que se apoyó el investigador, ya sea que tales resultados vayan o no en la dirección de la misma). También es cierto que los niveles del factor B (ya sea tiempo límite o tiempo libre) son independientes de la autoestima y del sexo de los Sujetos.

### **Interacción Ordinal.**

Si los resultados de un diseño factorial muestran uno o más efectos significativos de interacción, entonces, la interpretación de los efectos principales pierde interés. En principio la interpretación de los efectos principales depende de la ausencia de efectos significativos de interacción. Un diseño factorial puede presentar, en sus resultados, las siguientes condiciones: sí A y B son dos variables independientes y se encuentra una interacción significativa (A x B) ésta puede acompañarse de cualquiera de las siguientes condiciones de los efectos principales de A y B:

- 1) A y B no significativas .
- 2) A significativa y B no significativa.
- 3) B significativa y A no significativa.
- 4) A y B significativas.

Como puede verse sólo en la condición 1 es inexistente el dilema de sí i se interpretan o no los efectos principales puesto que estos resultaron no significativos. En las otras condiciones (2,3 y 4) la pregunta es sí se interpretan los efectos principales significativos o basta con la interpretación de la interacción. Una respuesta más precisa (que la genérica que señala que no se interpretan los efectos principales que intervienen en una interacción significativa) parte del criterio de ordinalidad de la interacción.

**Interacción Ordinal.** En una interacción ordinal el rango de los niveles de uno de los factores no cambia en los diferentes niveles de otro. Esto implica que uno de ellos tiene mayor efecto que el otro: Por ejemplo:

### **INSERTAR FIG. 5.11**

En la Figura 5.11 puede verse que, las líneas no se cruzan dentro de los límites del Factor B y que el nivel **a1** muestra un valor más alto que el representado por el nivel **a2**; así mismo, se observa que la diferencia mayor entre **a1 y a2** se dió en **b2**. Se tiene entonces que. Así pues una interacción ordinal acredita por sí misma la interpretación de efectos principales (desde luego los que resultan significativos). Sin embargo antes habrá que probarse la **ordinalidad** de la interacción. Para

tal propósito se vuelven a graficar los resultados de una interacción ordinal, cambiando de eje los niveles de los factores: en el caso de la Figura 6, los niveles del Factor A en el eje de las **X** y los del factor B en el de la **Y**. Si desaparece la ordinalidad (se encuentra que cambian los niveles del factor que previamente no lo hacían, no cambiaban en los diferentes niveles del otro) se tendrá finalmente una interacción **no-ordinal** (Ver Fig. 5.12). En una interacción de este tipo (no ordinal) se fundamenta el criterio de la no interpretación de los efectos principales cuando se tienen interacciones significativas.

**INSERTAR FIG. 5.12**

### **Estimaciones de Varianza en los Diseños Factoriales.**

En los diseños factoriales de dos factores o variables independientes se obtienen dos varianzas (medias cuadradas) por separado para cada factor. Estas dos medias cuadradas proporcionan aspectos diferentes, no redundantes, del experimento.

Las medias cuadradas son independientes unas de otras: entre los factores por separado, entre éstas y la media cuadrada de la interacción y entre ésta y la del error o sea la media cuadrada dentro de los grupos. Bajo estas condiciones de independencia de todas las varianzas, el investigador tiene una visión clara de los factores que contribuyen en la variación de sus datos. Esto es precisamente lo que hace el análisis de varianza, permite obtener aspectos particulares significativos del total de los datos.

En un análisis de tres factores se tienen tres valores  $F$  para tres valores de varianza que se asumen independientes. Sin embargo, el problema es el que las tres  $F$ s tienen el mismo denominador: la media cuadrada dentro de los grupos (MCD) lo que significa que estadísticamente las  $F$ s no son independientes entre sí. Si fueran independientes, la probabilidad esperada de que sean significativas sólo por azar con una  $p..05$  sería de un 15% ( $3 \times .05$ ), o una probabilidad de aproximadamente el 14% ( $1-95$ ) de que por lo menos una de las pruebas mostrara una significancia falsa.

El problema resultante de la no independencia de los valores  $F$  es la imposibilidad de conocer el porcentaje esperado de la intervención del azar en los resultados de las pruebas ( en otros términos la probabilidad del error Tipo I) con el agravante de que a mayor complejidad de los diseños factoriales mayor será la probabilidad de obtener una o más pruebas falsamente significativas. Esto constituye una razón por la que no es aconsejable que el investigador centre su atención sólo en la significancia de sus resultados y menos aun, en

resultados aislados sino que debe dar igual o más importancia a la fuerza de asociación de sus variables.

**Solución práctica al problema de no-independencia de las razones F, con un diseño factorial de 2 X 3 X 4.**

Tomemos como ejemplo un experimento en el que se tienen tres factores A, B y C. Supongamos que el factor A tiene dos niveles (a1, a2), que B tiene tres (b1, b2, b3) y C tiene cuatro (c1, c2, c3, c4). Tendremos por lo tanto, un diseño factorial de  $2 \times 3 \times 4 = 24$  combinaciones de tratamientos.

Asumamos que es un diseño randomizado completo con tres observaciones para cada tratamiento o en cada celdilla (recuérdese que en la práctica se requieren por lo menos de 10 observaciones, lo que equivaldría a una  $N = 240$ ) por lo tanto la muestra total es  $N = 72$  sujetos con  $ns = 3$ . En la tabla 5.15 se muestran los resultados.

TABLA 5.15 RESULTADOS DE UN DISEÑO FACTORIAL DE  $2 \times 3 \times 4$ ,

FUENTE DE LA VARIANZA	SC	GL	MC	F
Entre los Grupos	(1056.00)	(23)		
Factor A	60.5	1	60.5	11.61 < .01
Factor B	366.75	2	183.38	35.20 < .001
Factor C	468.00	3	156.00	30.00 < .001
A x B	54.25	2	27.13	5.21 < .01
A x C	25.5	3	8.5	1.63 < .05
B x C	80.25	6	13.38	2.57 < .05
A x B x C	.75	6	.13	.02 < .05
<b><i>Dentro de los Grupos</i></b>				
(error)	250.00	(48)	5.21	
TOTAL	1306.00	71		
F.05 (1,48) = 4.04	F.05 (3,48) = 2.80			
F.01 (1,48) = 7.19	F.01 (3,48) = 4.22			
F.05 (2,48) = 3.19	F.05 (6,48) = 2.30			
F.01 (2,48) = 5.08	F.01 (6,48) = 3.20			

Puede verse que los tres efectos principales y dos de las interacciones, resultaron altamente significativas. Se obtuvieron 7 valores F con un denominador común: la Varianza dentro de los Grupos (MCD = 5.21) violándose, así, como ya se señaló, la independencia estadística de los valores F. Para disminuir la probabilidad de valores F falsamente significativos, se sigue un procedimiento especial (Ver Keppel, 1973). Se busca la interacción significativa del orden máximo y se utiliza como denominador para obtener nuevos valores F. Revisando las interacciones de las Tablas 5.\_\_\_\_ observamos que la interacción del orden máximo (AxBxC) resultó no-significativa. De las interacciones de primer orden, resultaron significativas dos de ellas. Entonces se toma como denominador la media cuadrada de A x B que es la que tiene el valor mayor de F. En la Tabla 5.16 se han reordenado los elementos correspondientes en la fuente de varianza de tal modo que la MC de la interacción A x B se utilizar como el denominador de las diferentes razones F.

TABLA 5.16 OBTENCION DE NUEVOS VALORES F TOMANDO COMO DENOMINADOR LA MC DE A X B

Fuente de la Varianza	SC	gl	MC	F
Factor A	60.5	1	60.5	2.23
Factor B	366.75	2	183.38	6.76
Factor C	468.00	3	156.00	5.75
A x C	25.5	3	8.5	0.31
B x C	80.25	6	13.38	0.49
A x B x C	0.75	6	0.13	0.00
A x B	54.25	2	27.13	
Residual	250.00	48		
TOTAL	1,306.00	71		

F.05 (1,2) = 18.51      F.05 (2,2) = 19.00  
 F.01 (1,2) = 98.49      F.01 (2,2) = 99.01

F.05 (3,2) = 19.16      F.05 (6,2) = 19.33  
 F.01 (3,2) = 99.17      F.01 (6,2) = 99.33

Puede verse que el hecho de tomar como denominador, la MC de la interacción A x B, en lugar del de la Media Cuadrada Dentro de los Grupos (MCD) lleva consigo una importante reducción de los grados de libertad: de 48 a 2. Como se recordará, a menor sea el número de G.L. mayores deberán ser los valores arrojados por la prueba para que resulten significativos. Como consecuencia de una reducción tan grande, ninguno de los valores F (de la tabla 5.\_\_\_\_) resultó significativo. Hasta aquí, el procedimiento, es extremo y por lo mismo, llevaría a cometer el otro error, el error Tipo II que conduce a aceptar falsamente las Ho. Se requiere, entonces, una solución que medie la probabilidad de cometer los errores Tipo I y Tipo II.

Con este propósito se procede a tomar los grados de libertad de la interacción no significativa del orden mayor (que en nuestro ejemplo corresponde a la de A x B x C) y se agregan a los G.L. de la interacción que se eligió como denominador. Procediendo a la correspondiente adición de G.L. y a la obtención de la MC que se utilizará como denominador se tiene:

Interacción	SC	gl	MC
A x B x C	0.75	6	
A x C	54.25	2	
TOTAL	55.00	8	6.87

Entonces, 6.87 será la media cuadrada que se utilizará como denominador, para obtener, nuevamente, todos los valores F tanto de los efectos principales como los de las interacciones.

TABLA 5.17 NUEVOS VALORES F UNA VEZ QUE SE HAN REDUCIDO LOS EFECTOS DEL PROBLEMA DE LA DEPENDENCIA ENTRE FS MULTIPLES.

Fuente de la Varianza	SC	gl	MC	F
Factor A	60.5	1	60.5	8.81* < .05
Factor B	366.75	2	183.38	26.69*** < .001
Factor C	468.00	3	156.00	22.71*** < .001

A x C	25.5	3	8.5	1.24	>.25
B x C	80.25	6	13.38	1.95	>.10
Error (A x B + (A x B x C)	55.00	8	6.87		
Error residual	250.00	48			
TOTAL	1,306.00	71			

F.05 (1,8) = 5.32	F.05 (3,8) = 4.07
F.01 (1,8) = 11.26	F.01 (3,8) = 7.59
F.05 (2,8) = 4.46	F.05 (6,8) = 3.58
F.01 (2,8) = 8.65	F.01 (6,8) = 6.37

Si se compara este último análisis con el primero que aparece en la Tabla 5.\_\_\_\_\_, se verá que los efectos principales (A, B y C) siguen siendo significativos, aunque se registra un aumento del nivel de probabilidad del .01 al .05 para el factor A.

Por otro lado, desaparece la significancia estadística de la interacción B x C. Es evidente que este procedimiento al ser más estricto aumenta el poder de la prueba F al reducir la probabilidad de cometer el error Tipo I. Los resultados finales muestran diferencias significativas en los efectos principales así como en la interacción A x B. El siguiente paso sería tomar las decisiones estadísticas correspondientes e interpretar los resultados comenzando siempre por la interacción significativa del orden más alto.

Consideraciones Finales. Anteriormente se señaló la interdependencia que existe entre el diseño y el análisis estadístico de los datos. Se dijo que un mal uso de la estadística puede invalidar el mejor diseño de investigación; sin embargo, también un mal diseño tiene efectos sobre un análisis estadístico adecuado. Un ejemplo claro al respecto es la influencia que puede ejercer, en la conclusión obtenida del análisis estadístico, el error sistemático derivado de variables extrañas no controladas, que al actuar sobre todos los Sujetos componentes de un grupo (varianza dentro de los grupos) impiden distinguir si las diferencias encontradas se deben al tratamiento (aplicación de la variable independiente), al error sistemático de las variables extrañas, o a ambas condiciones. Como el efecto del error sistemático se suma al efecto del



tratamiento es muy posible que la varianza del numerador aumente y se obtengan diferencias significativas que el investigador atribuirá erróneamente a su tratamiento. En otras palabras, el efecto del error sistemático puede conducir a cualquiera de los dos tipos de error (Tipo I o Tipo II) que pueden cometerse en las conclusiones que se hacen en torno a la prueba de las hipótesis estadísticas, y en general en relación a los resultados de una investigación.

Lo señalado se vuelve más claro si se revisan gráficamente los factores que intervienen en el análisis de los datos, considerando las pruebas estadísticas que se revisaron:

$$T \text{ ó } F = \frac{V.I. + EES + EM}{EV}$$

V.I.= efecto de la Variable Independiente (o del tratamiento)

EES = efecto de error sistemático debido a variables extrañas no controladas

EM = efecto del error de muestreo (si se sacaran todas las muestras del mismo tamaño [n] que componen una población determinada y se obtuvieran sus medias, se encontrarían pequeñas diferencias debidas al azar)

EV = error de varianza (o variabilidad dentro de los grupos para el caso de la prueba **F**). En este caso interviene el efecto del error aleatorio derivado de las variables extrañas no controladas.

Como puede verse, es claro que el factor EES, aparte de confundirse con el efecto que se busca (V.I.), puede incrementar el valor del numerador, lo cual puede dar como resultado una diferencia significativa que puede conducir a concluir que tal diferencia se debe al efecto de la V.I., cometiéndose así el error Tipo I (decidir que existen diferencias cuando en realidad no es cierto).

Por otra parte, puede darse el caso de que el efecto de la V.I. se cancele por un efecto opuesto del error sistemático (EES) disminuyéndose así, el valor de la razón estadística, y por lo tanto, concluir que no existe una diferencia significativa (error Tipo II), cuando de hecho si hay tal. En un caso se rechaza falsamente la hipótesis nula ( $H_0$ ) (Error Tipo I), y en el otro se le acepta, también falsamente (Error Tipo II). La falla que expresan ambos tipos de error es la incapacidad que tiene el experimentador en encontrar el verdadero efecto, el único en el que está interesado, el de la variable independiente que manejó. Puede verse entonces que un mal diseño de investigación también afecta al mejor (adecuado) análisis estadístico que se halla elegido. El reto al que se enfrenta

todo investigador es, de hecho, minimizar hasta donde sea posible los efectos del error.

#### **REFERENCIAS**

- Campbell, D. (1957): Factors relative to the validity of experiments in social settings. **Psychological Bulletin**, 54, pp. 297-312.
- Gayen, A.K. (1950): Significance of a difference between the means of two non-normal samples. **Biometrika**, 37, pp. 399-408.
- Insko, Ch. A. (1967): **Theories of Attitude Change**. Englewood Cliffs, N.J.: Prentice Hall.
- Keppel, G. (1973). **Design and Analysis. A researcher's handbook**. New Jersey : Prentice -Hall, Englewood.
- Lathrop, R.G. (1969). **Introduction to psychological research. Logic, design, analysis** Nueva York: Harper & Row
- McGuigan, F.J. (1993): **Experimental Psychology: Methods of Research**. Sixth Edition. Englewood Cliffs, N.J.: Prentice Hall.
- Montemayor, G. F. (1973). **Fórmulas de estadística para investigadores**. México: Colección Científica. Instituto Nacional de Antropología e Historia
- Norusis, M.J. (1992): **SPSS for Windows Base System User's Guide. Release 5.0**. Chicago, Illinois: SPSS Inc.
- Rosenthal, R., y Rosnow, R.L. (1991). **Essentials of behavioral research. Methods and data analysis**. Nueva York: McGraw-Hill
- Sandler, J. (1955): A test of the significance of the difference between the means of correlated measures, based on a simplification of Student's t. **British Journal of Psychology**, 46, pp. 225-226.
- Solomon, R. (1967): An extension of control group design. **Psychological Bulletin**, 46, pp. 137-150.
- Tukey, J. W. (1977). : **Exploratory data analysis**. Reading Mass: Addison-Wesley.
- Young, R.K., y Veldman, D.J. (1979). **Introducción a la estadística aplicada a las ciencias de la conducta**. México: Trillas.

## VI. ASPECTOS TEORICOS DE LA MEDICION Y METODOS DE RECOLECCION DE DATOS

Antes de abordar el tema referente a los diferentes métodos de recolección de datos que se presentan en este libro, es importante revisar, aunque sea brevemente, algunos aspectos de la medición y de los diferentes niveles de medición utilizados en las ciencias sociales.

### 6.1 QUE ES MEDIR

De una manera bastante general puede decirse que medir es un proceso en el que se aplica o utiliza un lenguaje común y universal--los números y operaciones asociadas-- para alcanzar una mejor y mas precisa descripción de la realidad--física o social-- que se intenta conocer. Badiou (1972,p.25) afirma al respecto, que medir es el proceso según el cual "el hecho se vuelve número".

Russell (1938, p.176) define a la medición de magnitudes como "cualquier método por medio del cual se establece una correspondencia única y recíproca entre todas o algunas de las magnitudes de un tipo y todos o algunos de los números, integrales, racionales o reales". Para Campbell (1938, p.126) medir es "el procedimiento de asignar números para representar propiedades de sistemas materiales, otros que el numérico, en virtud de las leyes que gobiernan estas propiedades". Por su parte Stevens (1951, p.22) define medir como "el hecho de asignar números a objetos y eventos de acuerdo a ciertas reglas". De la misma manera Nunnally y Bernstein(1995) establece que medir consiste en un conjunto de reglas para asignar números a objetos para representar cantidades de atributos.

De las definiciones se desprende que para que los números representen válidamente los hechos, es necesario que en el proceso de dicha transformación se cumplan ciertas reglas, que involucran problemas de carácter lógico y epistemológico,--que se refieren a la validez lógico-matemática y a la validez gnoseológica, respectivamente--, que exigen entender, entre otras cosas, la naturaleza misma de los números.

Los números son objetos formales abstractos y representan el producto de una práctica científica, y es esa práctica la que fija sus propiedades. Así, el número, ni es una abstracción ni un atributo de la cosa contada; es lo que permite contar, lo que abre la dimensión de la magnitud...el "concreto real". (Braunstein et al, 1978, p.162).

La medición en general, debe tener sentido. Al respecto Bachelard (1972) sugiere que hay que reflexionar para medir, y no medir para reflexionar. Un primer criterio que habrá de tomarse en cuenta es el principio del isomorfismo, con el que se asume, que los sistemas numéricos pueden representar modelos generales de las relaciones que se establecen entre los hechos --objetos, cosas, eventos, fenómenos, etc.-- que ocurren en la realidad física o social.

Es decir, si un objeto, evento o fenómeno del mundo empírico representa un sistema particular, y los aspectos observables del mismo son sus propiedades, entonces, las propiedades lógico-matemáticas de los números deberán tener un equivalente --reglas de correspondencia-- en las propiedades empíricas de los sistemas a medir. De esta manera y de acuerdo con el principio del isomorfismo, los números se asignan a los objetos, de tal manera que las relaciones entre los números reflejen las relaciones entre los objetos mismos, con respecto a una determinada propiedad.

Las características del mundo empírico, los aspectos observables, son las propiedades --altura, color, olor, intensidad, gravedad, masa, longitud, etc.-- y éstas pertenecen a un sistema --persona, casa, flores, luz, tonos, electrones, mesas, etc.--. Un sistema particular posee propiedades y son éstas las que se miden, y no los sistemas. Cuando se mide, se busca definir las propiedades (Torgerson, 1967,p.9). Sin embargo, debe tenerse en cuenta que no todas las propiedades de los sistemas son susceptibles de medición propiamente dicha; es decir, no todas las propiedades son cuantitativas. Esto implica que existen diferentes clases de medición, como se verá más adelante.

Ahora bien, ¿es posible encontrar una relación de uno a uno entre las propiedades de los números y el comportamiento de los eventos que se pretenden medir? Es claro que existen algunas propiedades del sistema numérico que difícilmente pueden encontrarse --someterse a prueba experimental-- no sólo en el campo de las ciencias sociales, sino aún dentro del terreno mismo de la física. La aditividad de los números constituye un ejemplo de dicha dificultad. Al plantear este problema, Plutchik (1958,p.227) afirma "...si se colocan en serie dos resistencias de 100 ohms, la resistencia combinada es de 200 ohms. Pero si éstas se suman en paralelo en vez de en serie, la resistencia combinada es de 50 ohms y no de 200. De la misma manera, si se tiene un recipiente de agua hirviendo a 100°C y se añade agua a la misma temperatura, 100°C, la temperatura del agua combinada sigue siendo de 100°C". Ya Guilford (1954), había señalado el problema que representaba probar experimentalmente la aditividad en relación a la temperatura y las limitaciones que enfrentan dichas pruebas, aún tratándose de longitudes o distancias. Después de todo, como Guilford lo dijera,

nadie ha tomado dos años luz y los ha colocado una después del otro, --tal como se hace por ejemplo, con dos reglas que midieran un metro cada una y se juntaran para probar que juntas suman dos metros-- como tampoco nadie ha demostrado experimentalmente la suma de distancias atómicas. De aquí y de acuerdo con Guilford, el supuesto de aplicabilidad de los números en todas las ciencias descansa sobre pruebas empíricas limitadas. Por lo tanto, el autor sugiere que a cada ciencia se le debe permitir proporcionarse su propia clase de evidencia empírica en relación al uso que hace de la medición.

Por otro lado, para que una medición pueda calificarse como útil o adecuada, no es necesario que se satisfagan todas las propiedades de los números, ya que los propósitos de investigación pueden variar y algunos de éstos se cumplen, por ejemplo, en ausencia del postulado de la aditividad. Sin embargo, no se deberá perder de vista, que cuando los requerimientos de la aditividad no se satisfacen, disminuye el significado de los números asignados, y en sentido estricto, se imposibilita la ejecución de algunas de las operaciones aritméticas. Como puede verse, el significado de la medición descansa, en gran medida, en el nivel o clase de medición realizada.

**6.2 NIVELES Y CLASES DE MEDICION.**

Se cuenta con diferentes clasificaciones o escalas de medición. Por ejemplo Torgerson (1967), parte de las propiedades de orden (los números se pueden ordenar), distancia (las diferencias entre los números se pueden ordenar), y origen (las series pueden tener un punto de origen, el número cero). Para distinguir cuatro diferentes escalas de medición dependiendo del número de propiedades que reflejan los números (ver Figura 6.1). Torgerson (1967, p.16) propone la siguiente clasificación tomando en cuenta sólo dos de las señaladas, ya que la propiedad de orden se encuentra invariablemente involucrada en la medición.

	Sin origen natural	Con origen natural
Sin distancia	Escala ordinal	Escala ordinal con origen natural
Con distancia	Escala intervalar	Escala de razón

Figura 6.1 Clasificación de acuerdo con las propiedades de orden, distancia y origen.

Puede verse en la Figura 6.1 que la escala ordinal cubre solamente la propiedad de orden; la escala ordinal con origen natural, aparte de cubrir la propiedad de orden, satisface otra propiedad, la de

origen; la escala intervalar satisface las propiedades de orden y distancia; por último, la escala de razón satisface las tres propiedades mencionadas: la de orden, origen y distancia.

En la siguiente Figura (6.2) se suman las propiedades de los números en las que se incluye la de la aditividad (postulados propuestos por Campbell, 1938).

PROPIEDADES	EXPRESION SIMBOLICA	EXPRESION VERBAL.
IDENTIDAD	1. $a=b$ ó $a=b$ 2. si $a=b$ , entonces $b=a$ 3. si $a=b$ y $b=c$ , entonces, $a=c$	Los números pueden ser idénticos o diferentes. La relación de igualdad es simétrica. Cosas iguales a la misma cosa son iguales entre sí.
ORDEN	4. si $a>b$ , entonces $b<a$ 5. si $a>b$ y $b>c$ , entonces $a>c$	La relación "mayor que" es asimétrica. Afirmación transitiva.
ADITIVIDAD	6. si $a=p$ y $b>0$ , entonces $a+b>p$ 7. $a+b=b+a$ 8. $a=p$ y $b=q$ , entonces $a+b=p+q$ 9. $(a+b)+c=a+(b+c)$	La adición de cero no modifica al número. El orden de las medidas no afecta el resultado. Objetos idénticos se pueden sustituir entre sí. El orden de las combinaciones no afecta los resultados.

Figura 6.2 Propiedades Numéricas

Regresando al tema de los niveles o escalas de medición, se ha visto que Torgerson elimina de su clasificación a la escala nominal y en su lugar, hace una diferenciación de la escala ordinal tomando en cuenta la propiedad de origen. Por su parte, Stevens (1946, 1951b) clasifica las escalas en nominal, ordinal, intervalar y de razón. Otra clasificación es la elaborada por Coombs (1952), quien considera una quinta categoría, la escala denominada parcialmente ordenada, que se ubica entre la nominal y la ordinal de la clasificación proporcionada por Stevens. Enseguida se proporciona

una descripción sumaria de las escalas de medición tomando en cuenta la clasificación de Stevens.

### **6.2.1 Escala Nominal.**

En este nivel de medición o con este tipo de escala, los números se usan para etiquetar una clase o categoría, para formar grupos de objetos. Los miembros de una clase son iguales o equivalentes en algún aspecto. Se puede hacer referencia a estos grupos simplemente como grupo 1, 2 y 3; es posible intercambiar los números que etiquetan a cada grupo y el propósito de la clasificación (medición) se sigue cumpliendo. Las reglas que se aplican a este nivel son: a) que todos los miembros pertenecientes a una clase tengan el mismo número; b) que no se asigne a dos clases el mismo número; c) que haya suficientes categorías que clasifiquen a todos los miembros. Los postulados que se aplican son los de identidad: un objeto se coloca en la clase "x" cuando no hay diferencia entre el objeto a clasificar y los ya clasificados en la clase "x". Se aceptan categorías de latitud relativamente amplias con el objeto de reducir el número de categorías. Esto significa que si por medio de la observación no se perciben diferencias, los objetos pueden considerarse iguales al respecto de la variable en cuestión. Esta es la forma lógica y más simple de medir.

Propiedades:

Relación de identidad. Todos los objetos que recibieron el mismo número poseen el mismo atributo. Atributo: sexo. Hombre=1; Mujer =2. Los números se pueden asignar a la inversa, o se pueden asignar letras u otros símbolos, siempre y cuando los poseedores del mismo atributo queden bajo el mismo símbolo.

Reglas Permisibles para Asignar Números a los Sujetos:

- A. Depende de las propiedades del atributo que se mide y no de las propiedades de los números.
- B. No se pueden determinar sin conocer las propiedades del atributo.

Operaciones Estadísticas:

- a) frecuencias; b) modo; c) Coeficiente de contingencia

Transformaciones:

Se pueden transformar los números de cualquier modo siempre y cuando se conserve la relación de identidad

### **6.2.2 Escala Ordinal.**

En este tipo de escala los números asignados tienen la propiedad de ordenar a los objetos por rangos. Los postulados numéricos que se aplican son los de orden. Es una especie de clasificación en cuanto a categorías cuantitativas. La distinción entre las categorías se basa en una cualidad o propiedad de los objetos clasificados. Si se desea tener un alto grado de discriminación en este tipo de escala, se coloca a un objeto en cada categoría; es decir, cada categoría tendrá una frecuencia de 1, y esto es lo que constituye una clasificación por rangos. También se puede tener en algunas o todas las categorías, frecuencias mayores de 1. En este caso se estaría manejando el método de categorías sucesivas. Debe tenerse en cuenta que las categorías ordenadas numéricamente están igualmente espaciadas en una escala, pero los intervalos entre ellas **no** son iguales. Por ejemplo, si se ordena a un grupo de sujetos en cuanto a su estatura, asignándole el número 1 al más alto, el 2 al que sigue y así sucesivamente, se tendría a los sujetos igualmente espaciados en la escala numérica, pero esto no garantiza que la diferencia de estatura entre el que tiene el número 1 y el que tiene el número 2 sea igual que la diferencia entre los sujetos que tienen los números 2 y 3. Las reglas que se aplican a este tipo de escalamiento son las siguientes: a) un conjunto de objetos o personas se ordenan de más a menos con respecto a la posesión de un atributo o variable determinada; b) sin indicación absoluta de qué tanto de la variable poseen los sujetos; c) y tampoco qué tan lejos están unos de otros.

Propiedades:

Asume la propiedad nominal mas una propiedad de orden. Se puede ordenar un conjunto de objetos con respecto a un atributo, asignando a cada objeto un sólo numeral que refleje su posición ordinal. También se puede usar una escala de orden y asignar a diferentes objetos el mismo número de orden.

Reglas Permisibles para asignar Números a los Sujetos:

Depende de que el atributo tenga la propiedad ordinal.

Operaciones estadísticas:

a) medianas; b) centiles; C) coeficientes de correlación de rangos.

Transformaciones:

Bajo cualquier regla que conserve el orden; que mantenga una función monótonica positiva de los números originales (multiplicar por una constante, sumarla, elevar al cuadrado, sacar raíz



cuadrada, logaritmos, etc.) Estas transformaciones se llaman monotónicas porque la relación funcional entre los números de los rangos originales y los valores transformados aumentan o disminuyen.

### 6.2.3 Escalas de Intervalo o Intervalares.

Este tipo de escala también se conoce con el nombre de escala de unidades iguales. En ella, las distancias numéricamente iguales representan distancias empíricamente iguales en algún aspecto o variable de los objetos. Aunque se puede hablar de la suma de intervalos, no se ha logrado el principio de aditividad en sentido absoluto porque el punto de origen o cero de la escala se coloca arbitrariamente. La cantidad de propiedad a la que se le asigna el cero probablemente no es el límite inferior de la variable que se está midiendo.

No se puede decir, por ejemplo, que una persona tiene una inteligencia de cero y que ese punto indica ausencia total de esa variable. Las reglas de este nivel son las siguientes: a) se conoce la ordenación por rangos de los objetos; b) se sabe qué tan lejos están unos objetos de otros; c) no se tiene información de la magnitud absoluta, porque no se sabe dónde está el cero de origen o cero real.

Propiedades:

Igualdad de intervalos. Si se agregan 6 objetos a 10 objetos, éstos serán 16; si se agregan 6 a 100 se tendrá 106. La diferencia entre 10 y 16 es la misma que entre 100 y 106. Los intervalos numéricamente iguales representan diferencias iguales en números. En los atributos psicológicos difícilmente se puede probar esta propiedad. Aunque se puede hablar de adición de intervalos, esto no implica o no tiene el significado de la propiedad aditiva porque el punto cero es arbitrario.

Reglas Permisibles para asignar Números a los Sujetos:

Aunque el atributo no tenga todas las propiedades de la escala de números, se debe poseer la propiedad de intervalos iguales.

Operaciones Estadísticas:

a) media; b) desviación estándar; c) correlación Producto Momento de Pearson; d) la única estadística común no aplicable es el coeficiente de variación.

Transformaciones:

Puede cambiarse el punto de origen. Se pueden aplicar los números, multiplicar por una constante. A través de cualquier función lineal positiva, se debe satisfacer la siguiente condición:  $Y = A + Bx$ . (La transformación debe ser no sólo monotónica sino también lineal.

#### **6.2.4 Escala de Razón o Proporción.**

Estas escalas tienen cero absoluto, donde el cero efectivamente representa nada de la variable que se mide. Se aplican los postulados de identidad, orden y aditividad. Las medidas de numerosidad por ejemplo, son medidas de proporción. Cuando se dice cero objetos se está diciendo realmente que no hay objeto. Este tipo de escala en realidad implica, en relación a una variable específica, que: a) se conoce el orden de clasificación de las personas u objetos; b) se conocen los intervalos entre las personas; y c) se conoce, por lo menos para una persona, su distancia del cero racional.

La investigación empírica implica clasificación o medición de las variables dependientes e independientes. Es importante que el investigador conozca a qué nivel de medición está trabajando para poder determinar adecuadamente el uso de sus procedimientos estadísticos. Esta información la puede obtener de la literatura especializada, y de los manuales de los instrumentos que utiliza. También la puede inferir a partir del modelo de medición que se haya utilizado en la construcción de los instrumentos que aplica para registrar y recoger sus datos.

Cuando un investigador incluye en su instrumento de recolección de datos preguntas referidas al sexo del respondiente, su religión, su estado civil, lugar de origen o su ocupación, está "midiendo" a nivel nominal. Si dentro de su instrumento contempla un conjunto de preguntas o indicadores dicotómicos (correcto-incorrecto) referidos al conocimiento de por ejemplo, los métodos anticonceptivos, y supone que el Sujeto que responda correctamente a un mayor número de preguntas de ese conjunto, tiene más conocimiento de métodos anticonceptivos, estará midiendo a nivel ordinal. Es decir, el investigador sólo supone que "a mayor número de respuestas correctas, más conocimiento de métodos anticonceptivos", y esta información sólo le permite ordenar en rangos a los Sujetos investigados. Si además dentro de su instrumento se contempla la existencia de, por ejemplo, una escala de actitudes tipo Thurstone, estará midiendo a nivel intervalar, con un origen cero arbitrario.

Lo mismo sucederá si su instrumento tiene alguna prueba o "test" que permita transformar puntajes crudos a puntajes estandarizados:

en este caso, el nivel de medición también es intervalar. Si por otro lado, algunas de las preguntas incluidas en el instrumento se refieren, por ejemplo, al número de hijos, habitaciones de la vivienda, ingreso económico, o cualquier otro atributo que se registre en términos de numerosidad, que implique la existencia de un "cero" real, el investigador estará midiendo a nivel de razón o proporción.

De esta manera, un instrumento construido para medir o registrar diferentes variables, lo puede estar haciendo a diferentes niveles, el investigador deberá tomar en cuenta en el momento de efectuar el tratamiento estadístico de sus datos.

Propiedades:

En este caso el numeral 20 es dos veces el 10; así como 24 es el doble de 12, estas dos razones son iguales. Esta escala contiene un cero absoluto.

Reglas Permisibles para asignar Números a los Sujetos:

Las operaciones aritméticas de multiplicación y división pueden ser aplicables al atributo que se mide. Se debe estar seguro de que se está midiendo un atributo que puede tener una carencia absoluta del mismo (por la existencia del cero absoluto).

Operaciones Estadísticas:

a) todas las operaciones numéricas y estadísticas, así como el coeficiente de variación.

Transformaciones:

Se pueden multiplicar los numerales por una constante de acuerdo con la siguiente ecuación :  $Y = by$ . En dónde  $b$  es la constante multiplicadora y puede ser mayor o menor que la unidad, pero no igual a cero.

### **6.3 CLASES DE MEDICION.**

Además de los niveles o escalas de medición, se cuenta con otra clasificación, que basándose en el significado que se le atribuyen a las propiedades numéricas de escalas particulares, permite distinguir entre diferentes clases de medición. Torgerson (1967) señala tres diferentes maneras en las que las propiedades numéricas de orden, distancia y origen pueden tener sentido. Estas maneras o procedimientos representan tres diferentes clases de medición:

1. Medición fundamental o medición de magnitudes A, (estas denominaciones fueron propuestas por Campbell en 1938)
2. Medición derivada o medición de magnitudes B,
3. Medición por autoridad (**measurement by fiat**). Guilford (1954) denomina esta clase de medición como índices.

### **6.3.1 Medición Fundamental.**

La primera, la medición fundamental, depende de las leyes que relacionan a las cantidades entre sí, que representan a una construcción o sistema determinado. Los números se asignan siguiendo las leyes naturales y no se presupone la medición de otras variables. La utilización de este procedimiento de medición permite atribuir a la construcción medida, dos clases de significado: un significado constitutivo que implica la posibilidad de definir construcciones en términos de otras. Por ejemplo, las ecuaciones que relacionan las propiedades de fuerza, masa y aceleración; y un significado operacional que permite definir una construcción directamente, en términos de datos observables. En este caso, los números reflejan las leyes naturales que relacionan diferentes cantidades de la propiedad que se mide.

Un ejemplo de la medición fundamental, se tiene en las propiedades extensivas de la física, longitud, número y volumen, que son medibles de acuerdo con dicho procedimiento. Para cada una de estas propiedades que se basan en leyes que relacionan las diferentes cantidades que se encuentran presentes en cada una de ellas es posible construir, de acuerdo con Torgerson, una escala de razón, y es en términos de dichas relaciones y no de las relaciones que se establezcan con otras variables, que las propiedades numéricas de orden, distancia y origen adquieren significado constitutivo. Sin embargo, una vez que se hayan establecido relaciones con otras variables, éstas se pueden utilizar en lugar del procedimiento fundamental que requiere la medición de la propiedad en sí misma.

### **6.3.2 . Medición Derivada.**

Con respecto a la medición derivada, Torgerson señala que las propiedades numéricas, que se han venido mencionando, adquieren significado a través de leyes que relacionan, no a las cantidades de una construcción, como en el caso de la medición fundamental, sino a las propiedades entre sí. La densidad constituye un ejemplo de esta clase de medición: la razón entre la masa y volumen para cualquier cantidad de una sustancia determinada es una constante. Por otro lado, el valor de dicha razón va a cambiar para diferentes sustancias. De aquí que se pueda tomar el valor de la razón como la densidad de una sustancia dada. La importancia de esta clase de medición es que hace posible, dentro de un determinado campo de

investigación, ampliar el número de construcciones constitutivas a partir de otras relaciones ya bien establecidas; además por lo general se encuentra que una igualdad de intervalos o de razones adquieren significado a través de leyes o simplemente de definiciones que relacionan una propiedad con otras variables.

### 6.3.3. Medición por Autoridad.

En la medición que Torgerson denomina por autoridad, las propiedades medidas adquieren significado a través de un procedimiento un tanto arbitrario que se basa, en gran medida, en la intuición del investigador. Se asume que existe una relación entre las observaciones realizadas y el concepto o construcción de interés. Un ejemplo de esta clase de medición se tiene en los índices que se utilizan en Psicología, Sociología y Economía. Se miden directamente otras propiedades o se obtienen promedios ponderados de propiedades que se asumen relacionadas con una construcción determinada.

Esta clase de medición, escalas de números-índice, como la denomina Guilford (1954), representan un procedimiento ampliamente utilizado, que en el campo de la Psicología se puede ejemplificar especialmente con el coeficiente intelectual. La razón entre edad mental y edad cronológica, en donde ambos terminos se expresan en unidades de tiempo, años o meses, proporciona un número que no se encuentra en la escala original de años, ya que ésta se cancela debido a que la unidad de tiempo aparece tanto en el numerador como en el denominador. Se tiene pues, que el resultado es un número de razón, lo cual no significa que la escala del coeficiente intelectual sea una escala de razón, ni justifica el supuesto de que corresponda a una de intervalos iguales. De hecho, si no existen otras pruebas al respecto y ciertas condiciones se mantienen (distribución normal, una muestra grande, etc.), se considera una escala ordinal.

Siguiendo a Guilford, el coeficiente de correlación también corresponde a un número-índice. Las unidades originales de la medición realizada y que aparecen en el numerador y denominador de la razón, se cancelan. El coeficiente  $R$  de Pearson representa una escala ordinal, ya que aún cuando tiene un punto cero de origen, no tiene unidades iguales. Sin embargo, como señala el autor, es posible hacer transformaciones de escalas de niveles inferiores a otras de niveles superiores. Así, la  $R$  elevada al cuadrado se convierte en un coeficiente de determinación,  $R^2$ , que viene a ser índice de las proporciones de la varianza total en una de las dos variables, controlada por la varianza de la otra variable. Se tiene así, que de una escala ordinal ( $R$ ), se obtiene una escala de razón ( $R^2$ ) (Guilford, 1954, pp. 17-18).

En las disciplinas sociales, cuando se utilizan correlaciones, medias, desviaciones estándar, regresiones, etc., se asume por lo general, que los atributos están siendo medidos por una escala intervalar o de razón, aún cuando inicialmente, el significado de las propiedades numéricas haya sido atribuido sólo por definición, es decir, por autoridad. Sin embargo, este procedimiento, de acuerdo con Torgerson, no puede considerarse erróneo o lógicamente incorrecto. Lo importante es establecer relaciones estables entre variables. Empero, no deberá perderse de vista que el mayor problema de la medición por autoridad, reside en que una construcción o concepto determinado puede ser definido de muy diversas maneras, por la carencia de sistemas teóricos previamente establecidos, evidenciando con esto, la necesidad de dedicar un mayor esfuerzo al logro de mediciones fundamentales para eliminar mucha de la arbitrariedad que actualmente se encuentra presente en el proceso de medición que siguen las ciencias sociales y las de la conducta.

#### **6.4 TECNICAS O PROCEDIMIENTOS DE RECOLECCION DE DATOS.**

Las técnicas de recoger datos o registrar variables dependen de varios aspectos: a) el objetivo de la investigación; b) las características de los Sujetos que se investigarán; c) el tiempo del que se dispone para realizar la investigación; d) el personal con el que se cuenta para recoger los datos; y e) el tamaño de la o de las muestras que se estudiarán.

Entre los diversos procedimientos con los que se cuentan para la recolección de datos, los más utilizados son: la entrevista, la observación, la encuesta, y el cuestionario, entre otros. Se revisarán brevemente los dos últimos, por ser los de mayor aplicación en las investigaciones de carácter social.

##### **6.4.1 Encuestas y Cuestionarios**

Estos dos procedimientos pueden considerarse como variaciones del mismo tema, ya que son parecidos, y la diferencia entre ellos estriba en el tipo de información que se va a recoger y en la amplitud de la muestra a la que se va a aplicar. Las encuestas recogen una mayor variedad de diferentes tópicos al mismo tiempo, y los cuestionarios se puede decir que tratan temas un poco más específicos. En cuanto a la amplitud de la muestra, los procesos de encuestamiento son por lo general a niveles macros, ciudades o naciones y los cuestionarios están dirigidos a grupos más reducidos. Como en realidad las delimitaciones entre ellos son difíciles de establecer, algunos autores piensan que son una y la misma cosa. Siguiendo esta posición, debe entenderse que las

referencias que aquí se hagan, son aplicables a ambos procedimientos.

La utilización de estos puede obedecer a diversos y variados objetivos: obtener información de los ingresos de la población; conocer los hábitos de consumo; la distribución de los empleos; los hábitos de salud; las condiciones habitacionales; la emigración; la composición familiar; el crecimiento demográfico; lo que la gente piensa, siente, planea hacer; las opiniones respecto de problemas sociales, económicos y políticos nacionales e internacionales; información sobre hábitos de exposición a los medios masivos de comunicación; estudios de mercado, etc., etc.

#### **6.4.1.1. Planeación de la Encuesta**

En la planeación del procedimiento de la encuesta se deberá tener en cuenta lo siguiente:

1. Qué cuestión o cuestiones deben ser contestadas por los datos de la encuesta; esto se refiere específicamente a cuál es el problema a investigar.
2. Qué tópicos se van a encuestar para qué poblaciones; para hacer esto se deben conocer las tendencias y eventos sociales, políticos y económicos de la población que será investigada. Los tópicos que se investiguen deben ser de interés público general para que haya información y opiniones que dar al respecto; y sobre todo, deben ser cosas entendidas y conocidas por esa población que se encuestará.
3. Cómo, por quién y para qué van a ser usados los resultados.
4. Se debe prever la posibilidad de que los datos pierdan interés o utilidad para cuando se entreguen los resultados.
5. Presupuestar el costo y determinar el tiempo que se llevará la encuesta.
6. Seguridad de que no se ha respondido ya a este problema.
7. Saber cómo se va a obtener la información (qué técnicas se emplearán).
8. Entrenar a los encuestadores.

#### **6.4.1.2. Aplicación de la Encuesta**

Cuando se van a utilizar instrumentos contruidos por el investigador, o instrumentos poco conocidos, o bien contruidos para poblaciones de otras culturas, es necesario pilotearlos. El pilotaje tiene como objetivo principal el detectar fallas en los instrumentos, en los aplicadores, en la muestra, en los términos utilizados y en la forma de aproximarse al problema. La información proporcionada por el pilotaje permite afinar el instrumento, aumentando así las posibilidades de éxito de la investigación. El pilotaje se lleva a cabo con Sujetos similares, es decir, con las

mismas características de las personas a las que se les aplicará la encuesta. El piloteo debe hacerse teniendo en mente que los datos que se proporcionen, además de ser útiles para corregir errores, contribuirán a estimar la confiabilidad y validez del instrumento.

Los principales puntos que se deben tomar en cuenta en la construcción de cédulas son los siguientes:

1. Quiénes van a contestarla; dependiendo de esto se tendrá que usar lenguaje simple y apropiado para estas personas
2. Qué formato tiene la cédula; es preferible que parezca "corta".
3. Cómo se deben parafrasear las preguntas; se deben hacer en la forma en que están escritas en la cédula.
4. En qué orden se hacen las preguntas: en el orden en el que están impresas.
5. Cuántas preguntas contiene la cédula. Si son pocas, no importará mucho cómo se ordenen; si son muchas, deberá hacerse de tal manera que contengan mucha información en pocas hojas.

#### **6.4.1.3 Tipos de Reactivos de los Cuestionarios**

Los reactivos de los cuestionarios se pueden clasificar **grosso modo** en tres grandes grupos: información de identificación, antecedentes sociales o datos factuales de tipo censal, y preguntas sobre el tópico de la investigación.

##### **6.4.1.3.1 Información de identificación.**

La información contenida en este tipo de preguntas es la siguiente:

- a) número de la cédula
- b) nombre de la encuesta
- c) nombre de la agencia o institución que la lleva a cabo
- d) nombre o identificación del individuo investigado
- e) sexo del individuo
- f) relación del informador con el jefe de la familia
- g) dirección, zona postal
- h) lugar dónde se llevó a cabo la encuesta o cuestionario
- i) teléfono, si se tiene o si es posible
- k) nombre del encuestador
- l) fecha y hora de la encuesta
- m) renta mensual o estimación del costo de la casa del informador

##### **6.4.1.3.2. Antecedentes sociales o datos actuales de tipo censal.**

La información contenida en este tipo de preguntas se puede referir a todos o algunos de los siguientes aspectos:

- a) edad del jefe y demás miembros de la familia
- b) lugar de nacimiento del jefe de la familia



- c) estado civil
- d) religión
- e) educación
- f) preferencia política
- g) pertenencia a algún sindicato o agrupación
- h) tipo de tenencia de la casa habitación
- i) renta mensual o valor de la casa
- j) localización rural, urbana o semiurbana
- k) tamaño de la familia y composición
- l) ocupación del jefe de la familia o del informador
- m) quiénes trabajan y quiénes no, en esa familia
- n) ingreso familiar
- ñ) propiedad mueble o inmueble
- o) nivel socioeconómico aparente de la familia.

#### 6.4.1.3.3. Preguntas sobre el tópico de la investigación.

Como primer paso se requiere que el investigador estime el grado de conocimiento que el informante tenga del tópico o tema que se está investigando, antes de recoger los datos. Es aconsejable hacer preguntas de sondeo para separar a los que saben de los que no saben respecto al tema que se investiga. Para hacer esto se pueden usar varias estrategias:

- a) preguntar: "¿Ha oído ud. hablar de ....."
- b) Hacer la pregunta en forma de "buscar consejo". Los informadores por lo general se sienten halagados y contribuyen a establecer un buen **rapport**.
- c) Otra aproximación es la de usar una "pregunta explicativa". En esta situación se le da al respondiente suficiente información al respecto del tema tratado, pero en ningún momento con contenido evaluativo, crítico o calificativo.

Para seleccionar las preguntas se recomienda tomar en cuenta los siguientes puntos ( Bradburn y Sudman 1979; Belson, 1981: Parten, 1949)):

1. Inclúyanse sólo preguntas que tienen relación directa con el problema o con la evaluación de la metodología usada en la investigación.
2. No se incluyan preguntas cuyas respuestas pueden ser obtenidas con mayor confiabilidad partiendo de otras fuentes a menos que se deseen usar como una prueba o cotejo de la muestra, o vayan a ser usadas en tabulaciones de nuevos datos recogidos en la investigación.
3. Manténganse en mente los planes de tabulación cuando se hagan las preguntas, de manera que las respuestas a ellas sean fácilmente tabulables posteriormente (aunque ahora con ayuda de las computadoras personales, este punto ya no es tan importante).

4. Cuando sea posible, asegúrese de tener datos comparables de los cuales poder obtener indicadores de qué preguntas usar, cuáles, para usar términos, definiciones y unidades cuantitativas de medición iguales a las usadas en otros estudios comparables anteriores.
5. Cuídese la forma de hacer preguntas personales o aquellas que puedan avergonzar al informante. Si las va a hacer, hágalo en forma tal que se llegue lógicamente a ellas; o se puede concertar una cita posterior con el informante para que éste, habiéndolo conocido y adquirido confianza, tenga menos resistencia a contestarlas.
6. Haga sólo preguntas que la mayoría de los informantes puedan responder. Lo que la mayoría de la gente recuerda y es capaz de reportar, depende de la impresión dejada por la experiencia misma y las condiciones bajo las cuales es recordada en el presente. Los factores principales que afectan el recuerdo son: primacía, recencia, frecuencia, duración, validez, interés, significatividad, escenario, predisposición.
7. Evite preguntas que sean factibles de ser respondidas en forma inexacta, a menos que se tomen las precauciones para no hacerlo así.
8. No haga preguntas que necesiten demasiado trabajo extra por parte del informante.
9. No use preguntas de opinión, a menos que sean opiniones las que está investigando.

#### **6.4.1.3.3.1. Tipos de Preguntas.**

Los tipos de preguntas que se pueden emplear son diversos. A continuación se señalan algunos y se ponen ejemplos demostrativos.

1. *Preguntas Abiertas.* Durante las fases exploratorias de una investigación es por lo general necesario emplear preguntas que permitan al respondiente una amplia latitud en sus respuestas. Sólo descubriendo cuáles son los aspectos de un tema que han causado impacto en los Sujetos, y cuáles son las palabras que éstos emplean para expresar sus reacciones, es que se puede planear en forma adecuada un conjunto de preguntas cerradas o categorizadas. En los estudios de gran escala, las preguntas abiertas producen una variedad tal de respuestas que las tareas de clasificación y análisis toman mucho tiempo y es difícil manejarlas estadísticamente. Por estas razones, las preguntas abiertas son más adecuadas para estudios pequeños o estudios piloto. Al conjunto de preguntas abiertas centradas alrededor de un tema de investigación se le conoce con frecuencia como "entrevista profunda". Esta puede tomar una hora o más de tiempo en su aplicación. Como ejemplo de este tipo de preguntas están las siguientes:

En general, ¿cuál cree usted que sea la causa principal de la violencia?

¿Qué piensa usted acerca del impuesto al valor agregado (IVA)?

2. *Historia Libre y Método de Casos.* En este caso, en lugar de utilizar preguntas específicas, los entrevistadores pueden estimar opiniones y actitudes a partir de conversaciones extensas con los informantes.

El método de caso más aplicable a los procedimientos de investigación es con certeza la entrevista oral controlada, llevada a cabo por entrevistadores bien entrenados. Después de una entrevista extensa con el informante, el investigador (entrevistador) escribe un resumen en el que señala con claridad las varias afirmaciones o conductas que parecen indicar opiniones o actitudes. Si se sigue un esquema, y las reacciones del informante se pueden clasificar en todas las preguntas de todos los temas del esquema, los datos pueden ser valiosos en los análisis estadísticos. Sin embargo, cuando no existe tal procedimiento sistemático, el material del caso obtenido de esta manera es poco probable que se pueda adaptar a un análisis cuantitativo.

3. *Preguntas Dicotómicas.* Por lo general las preguntas pueden presentar alternativas opuestas diseñadas para provocar una respuesta de sí o no, de acuerdo o en desacuerdo, cierto o falso, correcto e incorrecto, apruebo o desapruuebo, bueno o malo, justo o injusto, etc. La ventaja de las preguntas dicotómicas es su simplicidad, tanto desde el punto de vista de la entrevista, como desde el punto de vista de la manipulación estadística. Toma poco tiempo preguntar y proporciona una respuesta tajante que puede ser codificada y tabulada fácilmente. Las preguntas dicotómicas son muy útiles cuando la opinión se ha cristalizado en forma tal que el tema estudiado puede reducirse a una proposición específica. Una de las limitaciones de este tipo de preguntas surge del hecho de que una ligera mala interpretación del significado de la pregunta puede dar como resultado una inversión completa de una opinión favorable a una negativa y viceversa. Otra limitación sería la que se refiere a no poder incluir respuestas a preguntas dicotómicas en los análisis correlacionales multivariados, a menos de que se hayan codificado con 1 y 0. Como ejemplo de este tipo de preguntas se tienen:

¿Tiene ud, una marca preferida de pasta dental?  
Si \_\_\_\_\_ No \_\_\_\_\_

¿Está ud. de acuerdo o en desacuerdo con la plantación de arboles en la Plaza Mayor?

Acuerdo( ); Desacuerdo( )

¿Si continúa la crisis, votaría ud. por un partido de la oposición en las próximas elecciones presidenciales?

Si ( ); No ( ).

4. *Lista de Chequeo.* En cierto sentido, las listas de chequeo son preguntas de respuesta de opción múltiple. Por lo general, consisten en el establecimiento de un problema o pregunta seguida de una lista desde tres hasta quince posibles respuestas entre las que se les pide a los respondientes marquen sus respuestas. Como regla, estas listas se deben hacer después de alguna investigación preliminar o piloto, que haya permitido determinar qué tipo de respuestas pueden esperarse. Sólo por medio del piloteo puede el investigador asegurarse que la mayoría de las respuestas no caerán en la categoría "otro" que se proporciona generalmente dentro de una lista limitada de opciones. Por otro lado, las listas de chequeo pueden sugerir respuestas que el respondiente marca por alguna extraña razón, como puede ser el que sea la primera o la última de la lista, o por ser la "apropiada" (deseabilidad social). Si las listas se presentan en forma oral, éstas deberán ser cortas, de preferencia de menos de cinco opciones, para que el que responda pueda tenerlas en mente cuando de su respuesta. Si la lista es larga, o si cada opción es compleja o complicada, será mejor escribirla en una tarjeta separada que el entrevistador le puede proporcionar al respondiente para que él mismo las lea (cuando esto sea posible). Como ejemplo de este tipo de preguntas, se tienen las siguientes:

Si los precios vuelven a subir dentro de los próximos seis meses, ¿de quién cree ud. que sea la culpa?

- ( ) Cámara de Diputados
- ( ) de todos
- ( ) Iniciativa Privada
- ( ) de la SECOFI
- ( ) de los Sindicatos
- ( ) del Gobierno
- ( ) de la situación internacional
- ( ) de nadie
- ( ) no sé

¿Cuál de las siguientes revistas hojeó o leyó ud. el mes pasado?

- |                  |                 |
|------------------|-----------------|
| ( ) Claudia      | ( ) Kena        |
| ( ) Proceso      | ( ) Plural      |
| ( ) Cosmopolitan | ( ) Discoposter |

( ) Siempre

( ) Ninguna de ellas

5. *Ordenación de Reactivos.* Este tipo de preguntas es muy utilizado en la medición de actitudes. En este caso se le pide al respondiente que arregle u ordene afirmaciones, palabras, frases, dibujos u otro tipo de objetos de acuerdo con su preferencia. Entre las limitaciones o desventajas de este tipo de preguntas están las siguientes: en primer lugar, al tratar de obtener una calificación individual de actitud no se hacen suposiciones acerca del tamaño de los intervalos entre los diferentes pasos o etapas empleados por los Sujetos. En segundo lugar, se encuentra el número relativamente pequeño de reactivos que es posible incluir. Si la lista es demasiado grande, el respondiente tiende a volverse más descuidado en sus respuestas en la medida en que se acerca al final. Aunque se ha sugerido que tres elecciones dan los mejores resultados, probablemente el número óptimo de ellas varíe de acuerdo al tema en cuestión. La calificación de este tipo de reactivos puede hacerse de diversas maneras. Sin embargo el procedimiento común ( Bradburn y Sudman,1979; Lundberg, 1941) es calificar de acuerdo al grado en el que la respuesta concuerda con la asignación predeterminada del valor escalar de cada reactivo. Como ejemplo de este tipo de preguntas se tiene:

¿Qué tan importantes son para ud. las características de un buen Gobernante? Escriba un (1) junto a la más importante, un (2) junto a la siguiente más importante, y así sucesivamente.

- ( ) su programa político
- ( ) su honradez
- ( ) su ideología
- ( ) su capacidad
- ( ) su equipo de trabajo

6. *Preguntas de Opción Múltiple.* Este tipo de preguntas se formula de tal manera que el repondiente deba escoger de entre varias posibles respuestas aquella que mejor represente su opinión, o esté lo más cerca posible de ella. Son particularmente útiles cuando el tema de estudio no es muy claro y no puede representarse en forma exacta por medio de una pregunta dicotómica. Con la opción múltiple se debe dar oportunidad de expresar todos los grados de opinión y actitud; deberá por lo tanto tenerse mucho cuidado en plantear la pregunta de tal manera que se represente al rango total de opinión sobre el tema o tópico en cuestión. Es importante que la lista de alternativas o categorías sea lo suficientemente completa como para cubrir todas las respuestas ante el tema. Si la lista es incompleta existe el peligro de que el respondiente crea que una de las alternativas especificadas representa su punto de vista y lea en esa alternativa algo que pertenece a otra categoría completamente diferente. Es difícil elaborar opciones que sean mutuamente

excluyentes como para que el respondiente no se encuentre indeciso entre dos elecciones igualmente deseables. Aún cuando las alternativas se comprendan y se recuerden, las preguntas de opción múltiple pueden dar resultados inadecuados por la tendencia que muestran los Sujetos de escoger los términos medios, y así darle mucho peso a las categorías intermedias de la escala. Las preguntas de opción múltiple pueden adoptar diferentes formas, que se presentan a continuación. En general, a estas formas se les llama escalas de clasificación o de intensidad. En ellas se le pide al respondiente que escoja entre varios grados de opción sobre una cuestión particular. El número de grados que se presentan depende del juicio del investigador y de consideraciones tan generales como el propósito del estudio y la naturaleza del tema. Pueden variar de tres, al número máximo que el investigador suponga que los respondientes son capaces de diferenciar. En general, no se emplean más de cinco opciones.

a) Escalas de Clasificación de Tres Puntos: las palabras particulares que se emplean para expresar el grado de sentimiento depende del fraseo de la pregunta o del tema, así como del juicio de la persona que construye la escala. Como ya se señaló anteriormente, el punto medio o "muestral" deberá utilizarse con cuidado en virtud de la tendencia a responder sin comprometerse. Entre las palabras o frases que comúnmente se emplean en las escalas de tres opciones están:

Muy importante	Algo importante	Nada importante
Más	Igual	Menos
Difícil	Igual	Fácil
Mayor	Igual	Menor
Si	Depende	No
La mayoría	Muchos	Pocos
Bueno	Regular	Malo
De Acuerdo	Indeciso	En Desacuerdo

b) Escalas de Clasificación de Cuatro Puntos: entre las palabras o frases que se dan a las opciones están:

Muchos	Algunos	Pocos	Ninguno
Excelente	Bueno	Regular	Malo
Siempre	Muchas veces	Pocas veces	Nunca

c) Escalas de Clasificación de Cinco Puntos: en este caso el tipo de fraseo que se da a las diferentes opciones queda ejemplificado de la siguiente manera:

Totalmente de acuerdo	De Acuerdo	Indeciso	En Desacuerdo	Totalmente en desa-
-----------------------	------------	----------	---------------	---------------------

				cuerto
Ciertamente correcto	Probablemente correcto	Dudoso	Probablemente incorrecto	Ciertamente incorrecto
Casi todos	Muchos	La mitad	Pocos	Casi nadie
Gusta mucho	Gusta	Neutral	Disgusta	Disgusta mucho
A favor en todo	A favor en algo	Neutral	A favor en poco	A favor en nada

Por último, cabe señalar que deberá tenerse en cuenta mucho cuidado en el parafraseo de las preguntas. A continuación se enuncian algunas recomendaciones que habrán de tomarse en cuenta para evitar el uso de palabras que colorean la respuestas, las condiciones bajo las cuales las respuestas se vean menos afectadas por la connotación de las palabras y que se puedan obtener opiniones o respuestas reales o verdaderas.

1. Emplee palabras sencillas que sean familiares o conocidas por los respondientes potenciales
2. La pregunta debe ser lo más concisa posible
3. Formule la pregunta en forma tal que produzca exactamente la información que desea
4. Evite preguntas con doble significado
5. Evite preguntas ambiguas
6. Evite preguntas que dirijan la respuesta
7. Evite usar nombres de personas prestigiadas o conocidas en el cuerpo de la pregunta
8. Evite palabras "peligrosas", estereotipos o palabras con connotación emocional fuerte
9. Incluya si es preciso, preguntas indirectas
10. Sea cauteloso con el empleo de frases que puedan afectar el prestigio del respondiente
11. Decida si personalizar o nó, algunas de las preguntas
12. Permita que se den todas las posibles respuestas
13. Asegúrese que las alternativas de respuesta en las preguntas de opción múltiple sean realistas
14. Asegúrese que sea mínima la cantidad de escritura que se requiere en el cuestionario. Emplee instrucciones como: escriba Si o NO; ponga un número; escriba una "X"; subraye la respuesta; circule su respuesta
15. Considere incluir algunas preguntas de cotejo o prueba para poder determinar la exactitud, veracidad y consistencia de las respuestas obtenidas en su totalidad.
16. Evite preguntas aparentemente no razonables empleando un argumento breve que las justifique
17. Considere el comparar las respuestas de algunas preguntas aisladas con otras respuestas de preguntas que se refieran al mismo tema, pero en contextos diferentes.

18. Considere omitir la alternativa intermedia y utilizar un número par de alternativas de respuesta.
19. Evite preguntas de hecho o de conocimiento si lo que le interesa es evaluar actitudes.
20. Cuide el orden en el que aparecen las preguntas, teniendo en cuenta que los respondientes puedan ser sensibilizados en pro o en contra según los contenidos de estos (Belson,1981; Bradburn y Sudman,1979;Converse y Presser, 1991: Kahn y Cammell, 1957; Turner y Martin,1984).

Como última recomendación es útil tomar en cuenta la secuencia de las preguntas. Las preguntas deben ordenarse lógicamente para evitar confusión y malas interpretaciones; deben agruparse en tal forma que una lleve a la siguiente en forma lógica. La pregunta con la que se inicie el cuestionario debe ser interesante o causar interés en el respondiente, y debe ser respondida con facilidad. Las preguntas que pueden molestar al respondiente deben colocarse en el medio o al final del cuestionario. Cuando se cambie de un tema a otro, debe hacerse una introducción breve y justificatoria. En el cuestionario debe quedar claramente especificado cuál pregunta deberá ser la siguiente en los casos en que una de ellas no sea aplicable al respondiente, o cuando la respuesta a una de ellas vuelve impropio las que le siguen (por ejemplo, la persona respondió NO a una pregunta y las que le siguen están explorando alrededor del mismo tema para el caso en el que la respuesta fue afirmativa).

El cumplimiento con las consideraciones señaladas en este apartado asegura la construcción de un instrumento más adecuado, de fácil aplicación, codificación y tabulación

## REFERENCIAS

- Bachelard,G. (1972): **La Formación del Espíritu Científico**. Buenos Aires: Siglo XXI Editores.
- Badiou, A.:(1972): **El Concepto de Modelo**. Buenos Aires: Siglo XXI Editores.
- Belson, W. R. (1981). **The design an understanding of survey questions**.Aldershot England : Gover.
- Braunstein, N.A., Pasternac, M., Benedito, G., Saal, F. 1978):**Psicología, Ideología y Ciencia**. México: Siglo XXI Editores.
- Bradburn, N., Sudman,S.(1979) **Improving interview method and questionnaire desing**.San Francisco :Jossey Bass



- Campbell, N.R.(1938): Symposium: Measurement and its Importance for Philosophy. **Aristotelian Society Supplement**. Volumen 17.
- Coombs, C.H.(1952): **A Theory of Psychological Scaling**. Ann Arbor, Mich.: Engineering Research Institute, University of Michigan.
- Converse, J.M., y Presser, S. (1991). **Survey questions. Handerafting the standarized questionnaire**. Newbury Park California : Sage Publications.
- Guilford, J.P.(1954): **Psychometric Methods**. Nueva York: McGraw Hill Book Company.
- Lundberg, G.A. (1941): **Social Research**. Longmans, Green.
- Kahn, R., y Cannell, C.F. (1957). **The dynamics of interviewing**. Nueva York : John Wiley.
- Nunnally, J.C. y Bernstein, J.I. (1995): **Teoría Psicométrica**. México: McGraw Hill.
- Plutchick, R. (1968): **Foundations of Experimental Research**. Nueva York: Harper and Row.
- Russell, B.(1938): **The Principles of Mathematics**. Segunda Edición. Nueva York: Norton.
- Stevens, S.S.(Editor)(1951): **Handbook of Experimental Psychology**. Nueva York: Wiley.
- Stevens, S.S.(1951b): Mathematics, Measurement and Psychophysics. En S.S.Stevens (Ed.):(1951) **Handbook of Experimental Psychology**. Nueva York: Wiley.
- Stevens, S.S.(1946): On the Theory of Scales of Measurement. **Science**, 103, 670-680.
- Torgerson, W.S.(1967): **Theory and Methods of Scaling**. Nueva York: John Wiley and Sons, Inc.
- Turner, C .F ., y Martin, E.(Eds.)(1984).**Surveying subjective phenomena** (Vol II). Nueva York : Rusell Sage.

## VII CONSTRUCCION DE ESCALAS DE ACTITUD

Si bien las escalas de actitud forman parte de los procedimientos de recolección de datos y constituyen un conjunto de instrumentos para cuya construcción son válidos muchos de los aspectos señalados en el capítulo anterior (y de hecho su revisión pudo quedar incluida en éste) la decisión de tratarlas como un tema aparte se fundamentó en la siguiente consideración: la construcción y aplicación de las escalas de actitud constituyen un área de medición que surge fundamentándose en un campo teórico de especial importancia para las ciencias sociales (estudio de las actitudes y el cambio de las mismas) y desarrolla procedimientos específicos que deben tratarse como tales. Esto es, su revisión requiere de un espacio propio para tratar cada procedimiento en su especificidad (sin que esto signifique, que en su conjunto no compartan generalidades con otros procedimientos) lo que las convierte en un campo especial de medición al interior de algunas de las disciplinas sociales (psicología social y sociología principalmente).

Los procedimientos que originalmente se desarrollaron para medir actitudes, son generalizables de tal manera que también son utilizables para la elaboración de instrumentos de medición potencialmente aplicables a cualquier otra variable o construcción diferente de la actitud.

De esta manera, el conocimiento del proceso que lleva a la construcción de las escalas de actitud, habilita al mismo tiempo en la construcción de instrumentos de medición en general. Este hecho, cobra especial importancia, en tanto hace posible la obtención de instrumentos *ad-hoc* y con esto, la eliminación de los problemas que conlleva por un lado, la búsqueda--que la mayoría de las veces es infructuosa-- de un instrumento o prueba estandarizada que se adecue a una serie de necesidades particulares de investigación, y por el otro, la utilización --con sus consecuencias teórico-metodológicas-- de instrumentos que han sido desarrollados y validados en poblaciones distintas.

Las innegables ventajas que representa para el estudiante, el profesional, y para el investigador de las ciencias sociales la adquisición de esta competencia técnica que hace posible elaborar instrumentos propios de medición, es probablemente el aspecto más valioso de este capítulo.

### 7.1 ESCALAS DE ACTITUD

Existen muy diversas y variadas definiciones de actitud. Sin

embargo, se puede dar como una definición de actitud que aglutina a la mayoría de las existentes, la siguiente: es una disposición fundamental que interviene junto con otras influencias en la determinación de una diversidad de conductas hacia un objeto o clase de objetos, que incluye: declaraciones de creencias o sentimientos, así como acciones de aproximación y de evitación hacia dicho objeto u objetos.

Entre las variables que influyen en la determinación de la diversidad de conductas hacia un objeto o clase de objetos, además de la disposición actitudinal, se encuentran: a) otras características del individuo, disposiciones hacia otros objetos representados en la situación, los valores del Sujeto y de los que están representados en la situación, el estado motivacional, el estilo de expresión, etc.; b) las características de la situación, las prescripciones de la conducta adecuada, las expectativas de las consecuencias posibles, etc.

La actitud puede verse como una construcción hipotética o como una variable latente, ya que no puede medirse directamente, sino que sólo se puede inferir a partir de las acciones de los individuos, como serían: su conducta verbal (información de creencias y sentimientos); la ejecución de una tarea relacionada con el objeto psicológico; y las acciones hacia un representante de la clase del mismo.

El investigador trata de producir un índice o medida combinada que represente la mejor inferencia que puede hacerse de la multitud de operaciones empíricas acerca de las características subyacentes que se supone reflejan. Para esto, trabajará en el desarrollo de varias medidas diferentes para que en cada una se elimine de manera sistemática toda influencia identificable sobre respuestas de la actitud que interesa.

Existen diversas variables extrañas que pueden intervenir como fuentes de invalidez en la medición de actitudes. Entre ellas están: a) el efecto reactivo de la medición, es decir, el mismo hecho de la medición hace que el Sujeto piense sobre el tema y adopte una actitud, que antes podía no haber tenido, b) la tendencia a estar de acuerdo con las afirmaciones independientemente de su contenido, c) la intención de dar una autoimagen socialmente deseable d) la existencia de patrones de respuesta y el estilo de expresión e) los errores cometidos por el investigador (experimentales) y los errores del muestreo.

A continuación se señalan algunos procedimientos para controlar y disminuir la influencia de las variables extrañas, que se presentan en tres aspectos principales: en el metodológico, en los Sujetos y

en su contexto, y en el instrumento.

En cuanto al aspecto metodológico, se recomienda muestrear grupos opuestos en cuanto a la sensibilidad que estos tengan frente a las influencias extrañas; emplear diseños con grupos control; llevar a cabo un análisis interno de los datos para señalar cómo difieren las respuestas de los subgrupos cuando es eficaz un determinante y no otro; incluir técnicas de control de balanceo y contrabalanceo en los reactivos, en los Sujetos y en los aplicadores.

En lo concerniente a los Sujetos y su contexto, se recomienda evitar situaciones de presión social, ya que en ocasiones las presiones concretas coinciden con presiones internas; garantizar el anonimato; controlar el aspecto cognoscitivo (qué tanto se sabe o se conoce) en relación al objeto psicológico; y controlar aquellas variables antecedentes que se suponga se relacionan de manera importante con el objeto psicológico.

Por lo que respecta al instrumento, debe tomarse en cuenta la sensibilidad del mismo a otras influencias; en su construcción no debe olvidarse el supuesto que señala que la disposición subyacente hacia el objeto psicológico es el determinante de las respuestas dadas al instrumento. Puede existir una discrepancia entre la respuesta privada del individuo y la manifiesta; para evitarla, se recomienda disminuir el grado de evidencia del propósito de la medición por medio de una o algunas de las siguientes formas: incluir reactivos que no corresponden al objeto psicológico; incluir en los reactivos varios aspectos, además de aquel en el que se está interesado; informar que **no hay respuestas correctas o incorrectas**; informar que *Las personas difieren en sus opiniones*; motivar al Sujeto para que coopere en nombre del conocimiento científico solicitándole proporcione respuestas honestas; **transmitir** de alguna manera al Sujeto la aceptación incondicional y no evaluativa del investigador.

Otras maneras (no utilizables con las escalas de actitud) de reducir la discrepancia antes señalada, se relaciona con el empleo de instrumentos semiproyectivos o proyectivos para evitar así respuestas sujetas a control consciente; elaboración de reactivos de elección forzada donde se le pide al Sujeto que seleccione entre dos o más afirmaciones (igualadas en deseabilidad social pero diferenciales en implicación) aquella que mejor refleje su opinión. Por último, otras acciones tendientes a controlar las variables extrañas serían: variar la dirección de los reactivos adyacentes; emplear pares igualados de reactivos; incluir reactivos **clave**; incluir en una afirmación que provocará una actitud desfavorable, una justificación de la misma; incluir reactivos paralelos con determinantes opuestos.

En la elaboración de una escala de actitudes intervienen, en los casos más sencillos, por lo menos tres conjuntos de variables:

1. Estímulos (variable independiente): estos proporcionan una forma en la que el Sujeto expresa la cantidad de atributo que se halla en él mismo y no en el objeto psicológico.

2. Los Sujetos (a quienes se presentarán los estímulos): algunas técnicas ubican tanto a los estímulos como a los Sujetos en sus respectivas escalas a partir de los mismos datos.

3. Las respuestas (variable dependiente): los estímulos representan una cantidad predeterminada del atributo pudiendo entonces determinar la percepción que el Sujeto tiene de sí mismo por medio de sus respuestas al aceptar algunos estímulos como representativos de sí mismo, o deseables para él.

Se debe recordar que se construyen escalas para medir un atributo, una abstracción de la percepción, ya sea de estímulos, de personas o de respuestas. El interés primordial es el atributo abstraído, y por lo tanto se pueden elaborar escalas de cualesquiera de las tres variables básicas. Lo que se requiere en última instancia, es determinar las propiedades del atributo (por ejemplo, si tiene un cero absoluto o no) y posteriormente, asignarle numerales que establezcan una relación entre las propiedades del atributo y algunas de las de la escala de los números.

## **7.2 CARACTERÍSTICAS OPERANTES DE LOS REACTIVOS**

Las escalas de actitud se basan, en general, sobre la noción de continuos psicológicos cuantitativos, --valores escalares de la variable actitud-- y las actitudes se consideran, como ya se señaló, como variables latentes; variables que subyacen a las conductas o respuestas, asumiéndose que la conducta se encuentra en función de dichas variables.

En las escalas acumulativas así como en las escalas de localización consensual, se parte del supuesto que si a un individuo se le enfrenta a un conjunto de reactivos, --o afirmaciones como usualmente se les denomina en relación a las escalas de actitudes-- entonces, de la selección o aceptación que haga de los mismos se infiere la magnitud que del atributo en cuestión --al que se refieren las afirmaciones--, posee dicho individuo. De la misma manera se puede asumir, que las conductas o respuestas que un individuo hace en relación a un objeto actitudinal se pueden ordenar a lo largo de una dimensión o continuo que representa grados de favorabilidad o desfavorabilidad, de donde se puede inferir el lugar que el individuo ocupa en dicha dimensión, a partir de la magnitud de sus conductas o respuestas.

La localización del lugar que ocupan los individuos sobre el continuo, a partir de la selección de reactivos se ajusta a una función probabilística más bien que determinística, debido a la intervención de factores de variación --diferencias individuales y experimentales-- que imposibilitan una localización precisa. Los reactivos de un modelo de medición probabilístico, ocupan una posición que representa valores medios, valores promedio de todos los individuos que constituyen una población determinada.

La probabilidad de aceptar una determinada afirmación se puede considerar como una función de una variable de actitud latente hacia el objeto de interés. Si se grafica esta probabilidad de aceptación sobre los valores conocidos de esa variable latente, se tiene lo que se conoce como la característica operante de la afirmación. En términos generales, la característica operante de un reactivo, es la función que describe la probabilidad de aceptación del mismo.

Los reactivos se pueden clasificar de acuerdo con sus características operantes en dos clases principales: monotónicos y no-monotónicos. Si la probabilidad de aceptación para una determinada afirmación aumenta --o permanece constante dentro de intervalos limitados-- a medida que el valor de la variable latente también aumenta, entonces se tiene una función creciente monotónica. Si por otro lado, la probabilidad de aceptación decrece --o permanece constante dentro de intervalos limitados-- mientras que la variable latente aumenta, también en este caso se tiene una función monotónica, sólo que corresponde a una función decreciente. Dado lo anterior, es claro que los reactivos con características operantes monotónicas forman escalas acumulativas o sumativas. Esta característica operante se representa como una función sigmoide y la mayor o menor correspondencia entre la localización del reactivo y la localización del individuo que acepta dicho reactivo, se debe como ya se señaló, a factores de variación en los que interviene el azar.

Si por otro lado, la probabilidad de aceptación de una determinada afirmación, alcanza un valor máximo mostrando un incremento, --junto con un incremento en la variable latente-- y luego decrece, entonces se tiene una función normal no-monotónica. Las escalas con reactivos cuyas características operantes son no-monotónicas, corresponden a escalas no acumulativas.

### **7.3 DIFERENTES APROXIMACIONES DE ESCALAMIENTO**

Torgerson (1967, pp.45-48) clasifica el procedimiento de escalamiento en tres grandes aproximaciones, tomando como criterio

el lugar donde se localiza, --en los Sujetos, en los estímulos, o en ambos-- la variabilidad de las respuestas dadas a los estímulos.

### **7.3.1 Aproximación centrada en el Sujeto.**

En la aproximación centrada en el Sujeto, se atribuye a las diferencias individuales la variación sistemática que se encuentra en las reacciones de los Sujetos a los estímulos.

Los procedimientos que se siguen de acuerdo con esta aproximación, tienen como propósito principal a los Sujetos, serán a estos a los que se asignen valores escalares, mientras que los estímulos se consideran como replicaciones. De esta manera, si se añaden o quitan --aleatoriamente-- estímulos de la población original de estímulos, no se afectará el procedimiento ni los resultados. Dado que los estímulos son considerados como replicaciones, esta aproximación es análoga a la clasificación de una sola entrada del análisis de varianza, con varias observaciones por columna y en donde los sujetos corresponden a las columnas.

De acuerdo con Torgerson la aproximación centrada en el Sujeto no ha dado lugar, de manera importante, al desarrollo de modelos de escalamiento, ya que las aplicaciones que se han hecho representan en general, mediciones por definición, como es el caso de la gran mayoría de las pruebas de aptitud, de aprovechamiento e inteligencia, en donde el puntaje de los individuos viene a ser una suma simple o pesada del número de reactivos contestados correctamente. En los casos en los que el interés principal es el de escalar a los individuos, se requiere cambiar el proceso de medición por definición a uno de medición fundamental) y con esto se cambia también de una aproximación centrada en el Sujeto a una aproximación centrada en la respuesta.

En relación con las escalas de actitud, la técnica de Likert, que se revisará posteriormente, constituye un ejemplo de lo que se ha indicado. La tarea de los Sujetos está diseñada de tal manera que permite incrementar las diferencias individuales, es decir: como lo que se busca es dar valores escalares a los Sujetos, el investigador tiene que controlar los efectos de variación debidos a otras fuentes, seleccionando aquellos estímulos y respuestas que tienden a poner énfasis en las diferencias entre los Sujetos. Los Sujetos responden a cada afirmación sobre la base de su propia aceptación con el contenido de la misma. Sólo éstos reciben puntaje y los estímulos se seleccionan para incrementar las diferencias individuales. Es necesario hacer notar que aún cuando se pueden asignar pesos a los estímulos, esto se hace sólo con el propósito de poder diferenciar a los individuos y no con el de

asignarles una posición relativa en el continuo de actitud.

### **7.3.2 Aproximación centrada en el Estímulo o Aproximación de Juicio.**

En este caso, la variación sistemática --la reacción de los Sujetos a los estímulos--, se atribuye a las diferencias en los estímulos con respecto a un atributo determinado. El procedimiento que se sigue de acuerdo con esta aproximación, tiene como propósito principal a los estímulos, y serán éstos a los que se les asignen valores escalares. En este caso, los Sujetos se consideran como replicaciones, por lo tanto, aumentar o disminuir Sujetos seleccionados al azar de la misma población, no producirá efectos en el procedimiento ni en los resultados. También aquí, se puede ver una analogía con el análisis de varianza de una sola clasificación, aún cuando son los Sujetos los que se consideran como replicación.

En relación con la construcción de escalas de actitudes, esta aproximación se ve ejemplificada en el procedimiento que se sigue con la escala de intervalos aparentemente iguales. La tarea de los jueces, está diseñada para minimizar cualquier influencia que se deba a la propia posición del Sujeto con respecto al objeto de actitud, y por lo tanto, minimizar también las diferencias individuales. El investigador seleccionará grupos homogéneos o individuos a los que someterá a un gran número de ensayos. Cualquiera de estos tratamientos tenderán a minimizar la varianza debida a otros factores que no sean las diferencias interestímulos con respecto a un atributo de interés.

### **7.3.3 Aproximación Centrada en la Respuesta**

De acuerdo con esta aproximación, la variabilidad de las reacciones a los estímulos se atribuye tanto a la variación en los Sujetos como a la variación en los estímulos. De esta manera, se asignan valores escalares a ambos: a los Sujetos y a los estímulos, Por lo tanto, el propósito de este procedimiento puede ser escalar Sujetos y/o estímulos. En relación con la analogía que se ha venido haciendo con el análisis de varianza, esta aproximación puede equipararse con el de doble clasificación en el que se tiene por lo general, una sólo observación por celdilla. Ahora bien, en relación con las escalas de actitud, la aproximación de respuesta tiene su mejor aplicación con la técnica de Guttman --como se verá más adelante-- que se conoce como Análisis de Escalograma o Análisis Escalar. La tarea de los Sujetos consiste en responder a una afirmación sobre la base de la posición de ésta en relación a la posición de los Sujetos. Es decir, la actitud del Sujeto y la actitud que refleja el reactivo,



determinan la respuesta del individuo. Esta actividad es cualitativamente diferente a la tarea que realizan los Sujetos dentro de la aproximación centrada en el estímulo, en donde los Sujetos hacen juicios que consisten en determinar cuál o cuáles afirmaciones reflejan una actitud más favorable hacia el objeto de actitud. Así, un Sujeto puede seleccionar una afirmación determinada como la mas favorable independientemente de su propia actitud. En la parte dedicada específicamente a la construcción de escalas de actitud, podrá observarse claramente las diferencias que imponen a los Sujetos --juicios vs. respuestas-- estas dos aproximaciones de escalamiento: la centrada en el estímulo y la centrada en la respuesta.

Enseguida se proporcionan una serie de definiciones (acotaciones comunes en la construcción de escalas de actitud), con el propósito de unificar el significado de los términos que se utilizarán a lo largo de este capítulo.

*Actitud.* Para efectos de medición, el término de actitud se define como el grado de afecto positivo o negativo asociado a un objeto psicológico (Edwards, 1957; Thurstone, 1946).

*Objeto Psicológico.* (O.P.). Se entiende por objeto psicológico cualquier símbolo, frase, **slogan**, persona, institución, idea o ideal hacia los cuales la gente puede cambiar en cuanto al afecto positivo o negativo asociado a ellos (Edwards, 1957).

*Atributo.* Se refiere a una propiedad medible de la actitud. En este caso la actitud es el sistema. Este no es medible en sí mismo, sino que posee uno o más atributos potencialmente medibles que lo definen (Torgerson, 1967).

*Continuo Psicológico* (C.P.). Es el ordenamiento psicológico --en tanto se forma a partir de juicios, creencias, sentimientos, etc.--, de los estímulos según el grado en el que poseen el atributo. Se obtienen diferentes continuos psicológicos para diferentes atributos. Sin embargo, también es posible encontrar diferentes ordenamientos para un mismo atributo o no encontrar un C.P. para un determinado atributo.

*Estímulos.* Son las afirmaciones o declaraciones que se hacen con respecto a un objeto psicológico.

*Valor Escalar.* (V.E.). El valor escalar se refiere a la estimación promedio --media o mediana-- que se obtiene para cada estímulo --afirmación--. En el continuo psicológico las afirmaciones se ordenan de acuerdo a sus valores escalares.

*Escala Monodimensional.* Los estímulos o afirmaciones se ordenan a lo largo de una dimensión y la posición de estos se representa por un único número --el valor escalar--. Atributo, dimensión y continuo pueden utilizarse de manera intercambiable.

#### **7.4 ESPECIFICACION DEL ATRIBUTO, CONTINUO O DIMENSION DE LA ACTITUD**

La actitud es una variable compleja, compuesta, y que para efectos técnicos de su medición se representa como un sistema que puede descomponerse en varios atributos. Ahora bien, en la construcción de escalas de actitud se han desarrollado dos grandes modelos: el monodimensional y el multidimensional. El modelo de escalamiento monodimensional, como su nombre lo indica, ha sido elaborado para medir un sólo atributo o una única dimensión por vez. De aquí que habiendo seleccionado el objeto psicológico, se requiera como primer paso, delimitar, especificar, el atributo que se medirá.

Un ejemplo de delimitación dimensional puede ser el siguiente. Supóngase que el objeto psicológico es la actitud hacia la prohibición de los juegos de azar y que el investigador delimita la dimensión de esta variable al grado de restricción que debe imponerse a la libertad individual. Se decide que dicho grado va desde una restricción absoluta, pasando por una posición neutral, hasta una libertad absoluta. Supóngase ahora que este investigador pide a una muestra de Sujetos que den sus opiniones acerca de la prohibición de los juegos de azar y se encuentra con las siguientes respuestas: a) la prohibición de los juegos de azar ha disminuido el uso de drogas; y b) la prohibición de los juegos de azar ha incrementado el uso del tabaco.

Puede observarse que aunque ambas opiniones se relacionan con la prohibición, rebasan sin embargo, la dimensión particular que se pretende evaluar; no se restringe al continuo especificado por el investigador. Por supuesto que los atributos medibles de la actitud hacia dicha prohibición pueden ser varios, de tal manera que si se quiere medir más de uno, es posible elaborar una escala monodimensional para cada atributo, y después obtener correlaciones y comparaciones entre ellas; o también medir más de un atributo con una sola escala, pero en este caso se recurriría al escalamiento multidimensional.

#### **7.5 AFIRMACIONES**

En relación con la elaboración de las afirmaciones de las escalas

de actitud deberán tomarse en cuenta los siguientes aspectos:

### **7.5.1 Contenido.**

Es claro que la selección del objeto psicológico determina el tema del contenido, y que la especificación del atributo establece la delimitación del mismo. Por ejemplo, las siguientes afirmaciones se relacionan con un objeto psicológico y muestran dos extremos del atributo especificado:

Objeto Psicológico: actitud hacia la participación de la mujer en la política.

Atributo: amplitud del campo político abierto a la participación femenina.

Afirmaciones extremas sobre el atributo:

-Considero que lo que menos importa en la elección de un candidato presidencial es su sexo; es decir, el que sea un hombre o una mujer.

-La política es un mundo hecho por y para los hombres, un mundo en donde la mujer no tiene cabida.

### **7.5.2 Redacción .**

La importancia de las afirmaciones reside en su función de estímulos, que por un lado, deben representar diversos grados del atributo o continuo de la actitud que se estudia y por el otro, provocar o despertar la disposición subyacente --que se logrará en la medida en que las afirmaciones representen efectivamente el continuo de la actitud de dicha disposición-- de los individuos hacia el objeto psicológico en la forma de respuestas.

Uno de los aspectos tendientes a asegurar la función de las afirmaciones se relaciona con la redacción correcta de las mismas y con esto lograr la eliminación de posibles interpretaciones erróneas. Existen diversos trabajos (Oppenheim, 1966; Selltitz, Jahoda, Deutsch y Cook, 1965; Goode y Hatt, 1970) que se ocupan de los criterios que han sido aportados por algunos de los principales especialistas en el tema. Edwards (1957) resume las sugerencias dadas al respecto por Wang, 1932; Thurstone y Chave, 1929; Likert, 1932; Bird, 1940; Edwards y Kilpatrick, 1948:

-Seleccionar afirmaciones que cubran el rango total de la escala afectiva de interés

-Utilizar un lenguaje simple, claro y directo

-Las afirmaciones deberán ser preferentemente cortas, idealmente no más de veinte palabras

-Cada afirmación deberá contener sólo un pensamiento completo

-Siempre que sea posible deberá darse preferencia a oraciones simples

-Palabras tales como *solo, solamente, casi, meramente* y otras de naturaleza similar deberán utilizarse con cuidado y moderación.

-Evitar afirmaciones que sean respondidas de la misma forma por casi todos; o que no sean respondidas por ninguno de los sujetos a quienes se aplica la escala.

### **7.5.3 Recopilación.**

Se asume que el conjunto de afirmaciones que se elabora en relación con un determinado objeto psicológico, constituye una muestra representativa de su universo de contenido; es decir, de todas las posibles afirmaciones que podrían hacerse con respecto a dicho objeto psicológico.

Un recurso práctico tendiente a disminuir el problema que representa el cumplimiento de ese supuesto, se relaciona con la ampliación y diversidad de fuentes de información de las que se obtendrá el material que se utilizará para elaborar el contenido de las afirmaciones. Algunas de las fuentes que pueden utilizarse para la recopilación de dicho material son las siguientes:

- Los medios de comunicación social --radio, cine, televisión, revistas, periódicos, etc.

- Bibliografía especializada en el tema, principalmente de las investigaciones realizadas en esa área y campos afines.

- Opiniones de investigadores y especialistas, que de una u otra manera, se encuentran en relación con el tema del objeto psicológico que se está estudiando.

- Opiniones individuales e independientes, acerca del objeto psicológico de los miembros de una población con características similares a las de la población que se estudiará (aquella a la que se aplicará la escala de actitudes).

Es claro que una fuente importante de información la constituye la experiencia y las opiniones del propio investigador. Sin embargo, si la recopilación del material se limitara sólo a esta fuente, por diversas razones, disminuirían las probabilidades de alcanzar una muestra representativa del universo de contenido en cuestión.

Una fuente recomendada por algunos investigadores (Delgado y Gutiérrez, 1995) consiste en la realización de estudios piloto, donde el objetivo particular se refiera la obtención de opiniones relacionadas con el objeto psicológico, por medio de técnicas como entrevistas no estructuradas o libres, de preguntas abiertas, asociación libre, grupos focales, redes semánticas, etc.

La mayor diversidad de fuentes de información, no sólo proporciona mayores posibilidades, --cuantitativas y lo más importantes cualitativas-- de muestrear adecuadamente el universo de contenido del objeto psicológico, sino también permite cubrir otra exigencia necesaria en el escalamiento de las actitudes, a saber: contar con afirmaciones que cubran los diversos grados que se encuentran entre los extremos favorable-desfavorable del continuo de la actitud.

Por otra parte, el logro de las dos condiciones mencionadas, --relacionadas con el universo de contenido y con el continuo de la actitud--, aumenta también la probabilidad de encontrar un continuo psicológico para el atributo que se medirá. El mero hecho de aumentar dicha probabilidad representa en sí mismo una ventaja eminentemente práctica, ya que no encontrar finalmente un continuo psicológico para el atributo que se ha medido, significaría reiniciar todo el procedimiento.

## **7.6. DIFERENTES PROCEDIMIENTOS DE RECOLECCION DE DATOS**

Algunos de los procedimientos de recolección de datos que más se han utilizado en el estudio de las actitudes son: la entrevista, la observación y las escalas.

### **7.6.1. *Entrevista Directa.***

Aún cuando existen variantes del método de la entrevista y cada una de ellas puede ser utilizada para el propósito de recoger la información relacionada con un determinado objeto actitudinal, no obstante, el procedimiento más utilizado ha sido el de la entrevista directa. Parece ser que el razonamiento que subyace a esta mayor utilización es que si se desea conocer "algo", lo más lógico y natural, --además de ser sencillo y rápido--, es preguntarlo directamente. Sin embargo, es precisamente lo directo de la entrevista lo que produce los principales problemas metodológicos, que se traducen especialmente, en variables extrañas o ajenas a la relación que se está estudiando y que intervienen traslapándose con ésta y, en este sentido, contaminando u obscureciendo dicha relación.

Así, en la entrevista directa existe una mayor probabilidad de que el entrevistado tienda a dar una imagen socialmente aceptada; es decir, que se produzca con mayor frecuencia lo que se conoce como el efecto de deseabilidad social. De la misma manera, las respuestas pueden estar influidas por las características personales del entrevistador. Otros posibles efectos son las respuestas reactivas que se producen como consecuencia del

contenido y forma de ciertas preguntas que el entrevistado puede interpretar como intromisión --a su privacidad--, o como restricción a su libertad de acción o respuesta. Por otra parte, el efecto de la presión social se hace más patente en la entrevista directa que en otros procedimientos (respuestas falsas o ambiguas, así como un aumento del número de respuestas del tipo "no sé, "no entiendo" y otras similares).

### **7.6.2. Observación Directa.**

El método de la observación, al igual que el de la entrevista, puede llevarse a cabo a través de diferentes procedimientos. Por ejemplo, en algunas investigaciones sobre actitudes se ha utilizado la observación participante, pero dada la complejidad teórica y práctica, que entre otras cosas implica mayor tiempo, costo y esfuerzo, se ha otorgado mayor preferencia a la observación directa. Sin embargo, si se compara con la entrevista, la observación directa resulta un proceso todavía más largo y costoso, principalmente por el tiempo que pueda llevar a cada observador la producción del suceso (interacción del individuo observado con el objeto de la actitud) a registrar.

Con la observación directa se enfrenta un problema teórico fundamental: la conducta, el comportamiento que se observa no mantiene una relación unívoca con la actitud. Así, la conducta no es necesariamente la expresión del sentimiento, y frecuentemente, puede ser totalmente opuesta, debido a los factores situacionales en que se desarrolla.

Es necesario señalar, sin embargo, en relación con los problemas descritos, con respecto a ambos procedimientos, --la entrevista y la observación--, que existen diversos recursos técnicos que permiten disminuir, en mayor o menor medida, sus efectos contraproducentes. Empero, existe una limitación inherente, y por lo tanto característica de estos procedimientos de recolección de datos. La información que se obtiene a través de ellos, sólo permite clasificar a los Sujetos estudiados en tres únicas categorías, a saber: favorables, desfavorables e indecisos.

Una clasificación tan gruesa, conlleva necesariamente un desperdicio de información, ya que en cada categoría se pueden encontrar diferentes grados de favorabilidad-desfavorabilidad hacia el objeto de la actitud que implica una mayor información potencialmente utilizable para una mejor comprensión e interpretación del fenómeno que se estudia.

## **7.7 ESCALAMIENTO DE LA ACTITUD**

Las escalas de actitud presentan más ventajas que desventajas si se comparan con los procedimientos anteriores. Con éstas se supera específicamente el problema de la clasificación limitada a sólo tres categorías.

Las escalas de actitud permiten evaluar los diferentes grados de afecto que los individuos pueden asociar a un determinado objeto psicológico. Se aplican a grandes grupos y en su elaboración, al igual que en cualquier prueba psicológica estandarizada, se deben cubrir ciertos requisitos (que se verán más adelante), para asegurar su adecuada utilización.

Un objeto psicológico, como ya se señaló, posee un dominio o universo de contenido que está formado por todas las posibles afirmaciones que pueden hacerse en relación a él. Las afirmaciones que constituyen el universo de contenido de un determinado objeto psicológico, pueden agruparse de acuerdo con Edwards (1957) y en diferentes clases.

#### **7.7.1. Afirmaciones de Hecho.**

Parte del universo de contenido está formado por afirmaciones de hecho, cuyo contenido se refiere a aspectos de conocimiento que constituyen lugares comunes a grandes grupos. Así, afirmar que "X fue uno de los Presidentes de México", representa una afirmación de hecho, de conocimiento. La probabilidad de que dicha afirmación sea aceptada es la misma independientemente de si los Sujetos tienen actitudes favorables o desfavorables al objeto psicológico. Por esta razón, las afirmaciones de hecho se eliminan de las escalas de actitud. Se asume que las afirmaciones que formarán una escala de actitud, deberán elaborarse de tal manera que permitan mostrar las diferencias existentes entre los sistemas de creencias (formados de acuerdo con Edwards por todas aquellas afirmaciones que el individuo acepta) de Sujetos con actitudes favorables de los de aquellos con actitudes desfavorables, en relación a un mismo objeto de actitud.

#### **7.7.2 Afirmaciones Favorables y Afirmaciones Desfavorables.**

En el universo de contenido de un determinado objeto psicológico, se encuentran además de las afirmaciones de hecho, afirmaciones favorables y desfavorables. Se asume que los individuos responderán a esta clase de afirmación más sobre la base de su actitud, de lo que sienten hacia el objeto en cuestión, que sobre lo que ellos pueden considerar verdad-mentira de los contenidos expresados en las afirmaciones. Así, se espera que las afirmaciones favorables tengan mayor probabilidad de ser aceptadas por individuos con actitudes favorables que por aquellos con

actitudes desfavorables.

### **7.7.3 Afirmaciones Ambiguas.**

Una última clase de los componentes del universo de contenido son las afirmaciones ambiguas. Estas afirmaciones se pueden interpretar en más de una manera, razón por la cual, no deben formar parte de una escala de actitudes.

## **7.8 LEYES EN EL ESCALAMIENTO DE LA ACTITUD**

En el desarrollo de los procedimientos de escalamiento actitudinal se postularon dos leyes: la ley del juicio comparativo (Thurstone, 1927) y la ley del juicio categórico (Attneave, 1949), que hicieron posible la investigación cuantitativa de toda clase de valores y expresiones subjetivas. De hecho estas leyes fundamentan la medición que se lleva a cabo con la mayoría de las escalas de actitud.

### **7.8.1 La Ley del Juicio Comparativo.**

La ley del juicio comparativo proporciona los fundamentos para el ordenamiento de los estímulos (opiniones respecto a un objeto psicológico) a lo largo de un continuo psicológico. Para su mejor comprensión se empleará un ejemplo ilustrativo:

Dado un conjunto de **n** estímulos (el conjunto de afirmaciones hacia un objeto psicológico), se postula que cada uno de éstos posee en grado variable (desconocido para el investigador) algo del atributo en el que está interesado.

El problema u objetivo del escalamiento psicológico consiste en ordenar los **n** estímulos en un continuo psicológico de acuerdo al grado de atributo que poseen. De lo anterior se deriva que la pregunta es ¿Se pueden ordenar las opiniones en un continuo psicológico según el grado o nivel de atributo que represente cada una de ellas?

Se debe tener en cuenta que probablemente: a) no existe un continuo psicológico para todos y cada uno de los atributos de interés; b) que cualquier conjunto de **n** estímulos se puede ordenar de diferentes maneras; c) que estos estímulos puede diferir con respecto a más de un atributo y d) que su ordenamiento puede no ser el mismo (para los correspondientes atributos) sobre los varios continuos.

A un estímulo dado **i** se asocia más frecuentemente un proceso modal



discriminativo ( $S_i$ ). Este es un concepto teórico que representa la reacción media de un conjunto de Sujetos al enfrentarse al estímulo  $i$  cuando se les pide que hagan un juicio en relación a algún atributo.  $S_i$  es el proceso discriminativo, o sea la reacción de un Sujeto al estímulo  $i$ .

En la Figura 7.1 se presenta en forma gráfica lo expuesto anteriormente: la distribución de las respuestas individuales ( $S_i$ ) que forman la curva normal, y el valor promedio ( $S_i$ ) de esas respuestas.

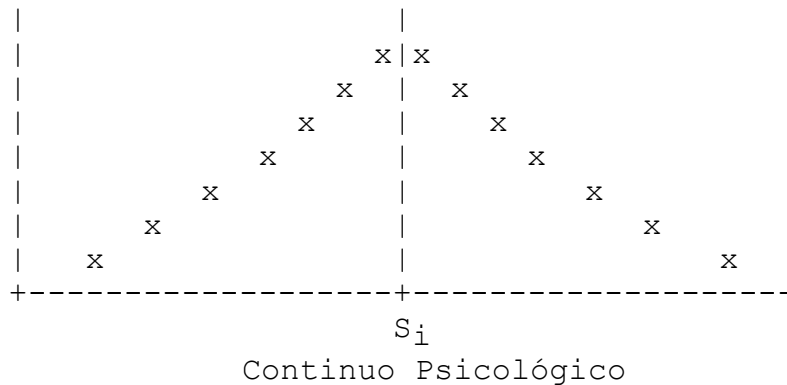


Fig.7.1. Distribución teórica normal de los Procesos Discriminativos  $S_i$  (provocada por el estímulo  $i$ ) distribuidos alrededor del Proceso Modal Discriminativo ( $S_i$ ).

Nota:  $S_i$  = media=mediana=modo (todos tendrán el mismo valor en una distribución normal).

**7.8.1.1. Supuestos de la Ley del Juicio Comparativo**

Los supuestos de los que parte esta ley son los siguientes:

- a) un estímulo dado  $i$  no siempre provocará el mismo proceso discriminativo, y
- b) los procesos discriminativos ( $S_i$ ) provocados por el estímulo  $i$ , se distribuyen normalmente alrededor del Proceso Modal Discriminativo ( $S_i$ ).

Ahora bien, el proceso modal discriminativo asociado al estímulo  $i$ , viene a ser su valor escalar. De esta manera,  $S_i$  = valor escalar; y valor escalar = Media = Mediana = Modo.

Se sabe que cualquier distribución normal se puede describir en términos de dos parámetros: la media y su desviación estándar. Se ha hablado del primero (valor escalar). El segundo es la Dispersión Discriminativa:  $\sigma_i$ . Esta se refiere al mayor o menor acercamiento que los  $S_i$  tienen respecto a  $S_i$ .

Si se mantiene constante el atributo del estímulo  $i$ , y se le introduce un segundo estímulo  $j$  para que sea juzgado con respecto al mismo atributo del estímulo  $i$ , y por otro lado, se asume para el estímulo  $j$  lo ya expresado para el estímulo  $i$ , o sea:  $S_j$ ,  $S_j$  y  $\sigma_j$ , los estímulos  $i$  y  $j$  pueden ser iguales o diferir con respecto a sus valores escalares ( $S_i = S_j$  ó  $S_i > S_j$ ), y lo mismo con respecto a su dispersión discriminativa ( $\sigma_i = \sigma_j$  ó  $\sigma_i > \sigma_j$ ).

Supóngase que a un grupo de Sujetos se les presentan las siguientes afirmaciones:

Estímulo  $i$ : "La mujer no tiene nada que hacer en el mundo de la política que fue hecho por y para los hombres"

Estímulo  $j$ : "En la elección de un candidato presidencial el que sea hombre o mujer es lo que menos importa", y se les pide que indiquen cuál de ellas es más favorable al objeto psicológico (amplitud de la participación política de la mujer). Con los juicios hechos por los Sujetos, los resultados podrían ser los siguientes:

- A. Que ambas son iguales:  $S_i = S_j$
- B. Que  $i$  es más favorable que  $j$ :  $S_i > S_j$
- C. Que  $j$  es más favorable que  $i$ :  $S_i < S_j$

En este caso las diferencias estarán dadas por la proporción de Sujetos que juzguen que  $i$  es igual, más o menos favorable que  $j$ . De esta manera, la separación escalar sobre el continuo psicológico de los procesos modales discriminativos ( $S_i$ ,  $S_j$ ) está dada en función de la proporción de juicios  $i$  y  $j$ .

Esquemáticamente, se puede establecer lo siguiente:

Frecuencias =  $f_{ij}$  = al número de veces que se escoge a los estímulos  $i$  como más favorables que  $j$

Proporciones =  $p_{ij} = f_{ij}/N$

Desviación normal =  $z_{ij}$

Si  $p_{ij} = 0.500$ , entonces  $z_{ij} = 0.000$

$p_{ij} < 0.5000$ , entonces  $z_{ij} = -z_{ij}$   
 $p_{ij} > 0.5000$ , entonces  $z_{ij} = +z_{ij}$

En forma gráfica, la situación anterior se vería como se muestra en la Figura 7.2.

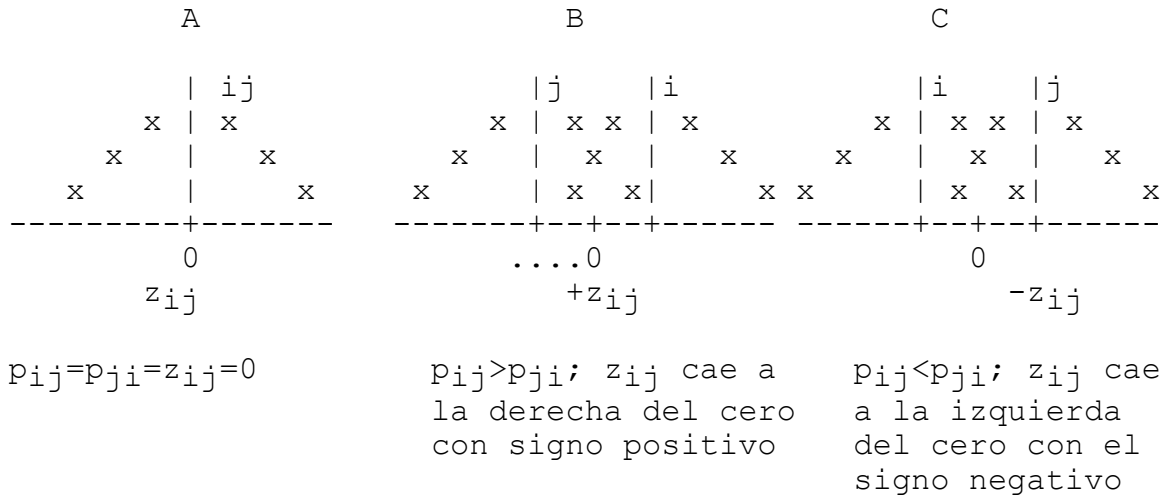


Figura 7.2 Representación gráfica de la distancia entre los estímulos *i* y *j* .

**7.8.2.Ley del Juicio Categórico**

Esta ley es paralela a la del juicio comparativo y se basa en los siguientes supuestos:

- a) El continuo psicológico se puede dividir en un número dado de categorías.
- b) Debido a diferentes factores, el límite de una categoría dada, no necesariamente se localiza siempre en un punto particular del continuo psicológico, sino que éste también se proyecta como una distribución normal de posiciones. Los diferentes límites de las categorías pueden tener diferentes medias de localización y diferentes dispersiones.
- c) El Sujeto juzga un estímulo por debajo de un límite dado de la categoría, siempre que el valor del estímulo sobre el continuo psicológico sea menor que el de ese límite. Es decir, en esta ley se parte del supuesto de que los límites de las categorías adyacentes se comportan como estímulos.

Estas dos leyes señalan, como se dijo anteriormente, los fundamentos sobre los cuales se desarrollaron los diferentes métodos o modelos de construcción de escalas de actitudes.

**REFERENCIAS**

- Attneave, F.(1949): A method of graded dichotomies for the scaling of judgements. **Psychological Review**, 56, 333-340.
- Bird, C.(1940): **Social Psychology**. Nueva York: Appleton-Century-Crofts.
- Delgado, J.M. y Gutiérrez, J.(Editores)(1995): **Métodos y Técnicas Cualitativas de Investigación en Ciencias Sociales**. Madrid: Editorial Síntesis, S.A.
- Edwards, A.L. (1957): **Techniques of Attitude Scale Construction**. Nueva York: Appleton-Century-Crofts.
- Edwards, A.L. y Kilpatrick, F.P.(1948): A Technique for the Construction of Attitude Scales. **Journal of Applied Psychology**, 32, 374-384.
- Goode, W.J. y Hatt, P.K.(1970): **Métodos de Investigación Social**. México: Editorial Trillas.
- Likert, R.(1932): A Technique for the Measurement of Attitudes. **Archives of Psychology**, 140.
- Oppenheim, A.N.(1966): **Questionnaire Design and Attitude Measurement**. Nueva York: Basic Books Inc.
- Selltiz, C., Jahoda, M., Deutsch, M. y Cook, S.W.(1965): **Métodos de Investigación en las Relaciones Sociales**. Madrid, Editorial RIALP.
- Summers ,G.F.(1976) **Medición de actitudes**, México: Trillas
- Thurstone, L.L. y Chave, E.J.(1929): **The Measurement of Attitude**. Chicago: University of Chicago Press.
- Thurstone, L.L.(1927): A Law of Comparative Judgement. **Psychological Review**, 34, 273-286.
- Thurstone, L.L.(1946): Comment. **American Journal of Sociology**, No. 258.
- Torgerson, W.S. (1958): **Theory and Methods of Scaling**. Nueva York: John Wiley.
- Wang, K.A.(1932): Suggested Criteria for Writing Attitude Statements. **Journal of Social Psychology**, 3, 367-373.

## VIII. ESCALAS DE ACTITUD: APROXIMACION CENTRADA EN EL ESTIMULO

En este capítulo se presentan los procedimientos operativos necesarios para la elaboración o construcción de las escalas de actitud de acuerdo con los métodos que corresponden a la aproximación centrada en el estímulo, y que son: el método de pares comparados, el de intervalos aparentemente iguales y el de intervalos sucesivos.

### 8.1 METODO DE PARES COMPARADOS

Este método deriva de la ley del juicio comparativo; es el primero que se desarrolla y el más fuerte desde el punto de vista teórico-estadístico. Sin embargo, como también es el más complejo y laborioso de desarrollar, resulta ser el menos utilizado, especialmente porque algunas de sus principales ventajas se obtienen con los métodos que se desarrollaron después y a partir de él, y porque estos superan su principal desventaja que se refiere a la gran cantidad de trabajo que exige a los Sujetos y al mayor tiempo que requiere su construcción. Por estas razones, sólo se proporcionará un panorama general de este método, para revisar detalladamente el resto de los métodos que conforman el campo de las escalas de actitud. Los lectores interesados en el Método de Pares Comparados encontrarán en la bibliografía de este libro, las fuentes a las que pueden acudir.

Con el método de pares comparados se requieren de 10 a 14 afirmaciones, la mitad de ellas favorables y la otra mitad desfavorables al objeto psicológico. Con estas afirmaciones se elaboran todas las posibles combinaciones de pares de afirmaciones de manera que cada afirmación quede apareada con todas las demás, sin que estos pares se repitan. Así, si se tienen por ejemplo, 10 afirmaciones, se habrán de formar 45 pares de ellas  $n(n-1)/2$ .

Estos pares de afirmaciones se aplican a un grupo de personas denominado sistema de jueces, que deberá ser una muestra representativa de la población que se va a estudiar. La tarea de estos jueces es la siguiente: de cada par de afirmaciones habrán de marcar cuál es la más favorable al objeto psicológico. Debe hacerse notar que lo que se pide a los jueces es un juicio objetivo, no su propia actitud o sentir al respecto de lo que se dice del objeto psicológico. Se recomienda que este sistema de jueces esté constituido por lo menos de 50 personas.

Con los resultados obtenidos (juicios comparativos), se procede a formar un cuadro de doble entrada donde se indica la frecuencia con que cada afirmación fue escogida como la más favorable en su

comparación con todas las demás. Este cuadro recibe el nombre de Matriz de frecuencias (Matriz, F).

Esta Matriz F se transforma en una Matriz de Proporciones (Matriz P), que consiste simplemente en transformar las frecuencias en proporciones. Para esta matriz tendrá que obtenerse la suma de proporciones de cada columna, y determinar si las sumas de las columnas se ordenan de la menor a la mayor, estando la menor del lado izquierdo de la matriz. Si esto no es así, se deberán rearrreglar las columnas y las hileras para que cumplan con ese requisito. Una vez que se tiene la Matriz P se transformará en una Matriz Z; esto se logra transformando cada proporción en una cantidad expresada en unidades de desviación estándar mediante el uso de la Tabla de desviaciones normales **z** correspondientes a proporciones **p** de una unidad de distribución normal dicotomizada (ver Edwards, 1957).

El siguiente paso será sumar las columnas de la Matriz Z y dividir cada suma entre el número de datos que tenga cada columna. Con esto se obtiene el Valor Escalar (V.E.) para cada afirmación, y se indica numéricamente qué lugar ocupa esa afirmación en un continuo que va de una calificación muy negativa (desfavorable) a una calificación muy positiva (favorable).

Como es difícil trabajar con números positivos y negativos, se hace una transformación de estos valores para que el número negativo mayor desaparezca, quedando como cero. Para lograr esto se le suman a todos los valores, el valor numérico del negativo mayor. Con esto se establece el origen de la escala.

Este método cuenta con pruebas de consistencia interna, pruebas de las suposiciones del modelo de medición utilizado, y el desarrollo de casos especiales para cuando no se cumplen los supuestos. Será conveniente llevar a cabo todas estas pruebas antes de continuar; estas pruebas aparecen en bibliografía especializada (p.e., Edwards, 1957).

Una vez que se conocen los V.E. de las afirmaciones y se cumplen los requisitos que establece el modelo, se puede decir que se tiene la escala de actitud definitiva o final. Estas afirmaciones se imprimen en un orden obtenido por selección aleatoria, escribiéndose las siguientes instrucciones a la cabeza de la hoja en que aparezcan:

"Escoja usted aquellas afirmaciones (pueden ser tres o cuatro) que mejor representen su punto de vista, opinión o sentir al respecto de X" (donde X es el objeto psicológico estudiado).

Debe hacerse notar que ésta es la escala final (la que se aplicará a los Sujetos que forman la muestra de investigación) y que se obtuvo con los valores encontrados con la muestra de jueces.

El procedimiento de calificación de la actitud que se sigue con este método es el mismo que se sigue con el Método de Intervalos Aparentemente Iguales, que se revisará detalladamente. El procedimiento con el que se obtiene el puntaje final de actitud corresponde al momento en que se le pidió a los Sujetos de investigación que escogieran aquellas afirmaciones con las que estaban más de acuerdo; con esto se obtiene la ubicación del Sujeto en el continuo de desfavorabilidad-favorabilidad hacia el objeto psicológico formado por los V.E. (de las afirmaciones) obtenidos con la muestra de jueces. De esta manera, el puntaje de actitud que un Sujeto obtiene, no es más que el promedio o la mediana de los V.E. de las tres o cuatro afirmaciones que escogió como las que mejor representaba su opinión o punto de vista.

### **8.1.1 Ventajas y desventajas de este Método**

Las ventajas de este método son las siguientes:

a) es el método más fuerte desde el punto de vista teórico y metodológico ;

b) posee el mayor número de pruebas que lo validan.

Las desventajas principales son:

a) sólo se puede usar un número pequeño de afirmaciones, ya que en la medida en que aumenta el número de éstas, se incrementa en forma exorbitante el número de comparaciones que los jueces tendrían que llevar a cabo para la obtención de los V.E. de cada afirmación;

b) el tratamiento estadístico de los datos para la obtención de los valores escalares es muy laborioso; además de que existen muchas situaciones que se consideran casos especiales para los que hay que seguir procedimientos especiales y diferentes.

## **8.2 METODO DE LOS INTERVALOS APARENTEMENTE IGUALES**

El método de los intervalos aparentemente iguales fue desarrollado por Thurstone y Chave en 1929, y aunque no en forma directa está sustentado en la ley del juicio categórico mencionada anteriormente. Para explicar el procedimiento que se sigue en la construcción de esta escala, al igual que las escalas que siguen, se recurre a ejemplos numéricos. Con el propósito de facilitar la exposición y comprensión de las mismas. se utilizan ejemplos con **n's** pequeñas y se describe paso por paso:

1. Se define el objeto psicológico. (O. P.) Supóngase que

este es: *actitud hacia el trabajo*; se debe entonces delimitar el atributo o dimensión de la variable actitud. El investigador señala como atributo, el siguiente: *la mujer que trabaja fuera del hogar*.

2. Se reúne la información necesaria para la elaboración de las afirmaciones que constituirán la escala. Esta información se puede obtener de diversas fuentes: pedir opiniones sobre el tema, a mujeres que trabajan fuera y dentro de sus hogares; obtener opiniones que al respecto hayan sido vertidas por los diversos medios de comunicación social (T.V., radio, revistas, periódicos); solicitar ideas y opiniones a especialistas; consultar libros, investigaciones o trabajos relacionados con el tema.

Con esta información se elaboran aproximadamente 100 afirmaciones que cumplan con los siguientes requisitos.

1. Usar palabras sencillas, que sean conocidas por todos los respondientes potenciales.
2. La afirmación debe ser lo más concisa posible, no teniendo más de 20 palabras.
3. Formular la afirmación de manera que produzca únicamente la información deseada; evitar cualquier pregunta o reactivo no relacionado con la variable.
4. Evitar afirmaciones ambigüas o que puedan ser interpretadas de diferentes maneras.
5. Evitar afirmaciones que sugieren la respuesta.
6. Cada reactivo o afirmación debe tener sólo una idea; oraciones simples.
7. Las sintaxis debe ser sencilla y clara: sujeto, verbo y complemento.
8. No se usen nombres de personas prestigiosas ( favorable o desfavorable ) en el cuerpo de la afirmación o reactivo.
9. Evitar palabras "peligrosas", pegajosas, esterotipos o palabras con connotación emocional.
10. Decidir si se va a emplear o incluir afirmaciones o reactivos indirectos.
11. Tener cuidado de no emplear

frases que puedan afectar el prestigio del respondiente.

12. Decidir si se va a personalizar algunos de los reactivos o preguntas, o si se redactarán en tercera persona o en forma impersonal.
13. Permitir que se den todas las posibles respuestas, principalmente en las afirmaciones o reactivos de opción múltiple.
14. Las alternativas de la preguntas de opción múltiple deben ser realistas.
15. Evitar hacer preguntas referidas al pasado, pues los





a) colocando en la parte superior de cada hoja una escala como la que aparecen en la Figura 8.1, y las afirmaciones listadas en el cuerpo de las hojas, precedidas por un paréntesis en el que habrá de escribirse la letra correspondiente de acuerdo al juicio del juez;

b) se pueden presentar las afirmaciones seguidas cada una de ellas de una escala como la que aparece en la Figura 8.1, donde el juez marca sobre ellas con una "cruz" o "paloma", el intervalo que juzgue le corresponde a cada afirmación, que se presentarán en orden aleatorio.

Las instrucciones que se dan a los jueces se imprimen en una hoja aparte, que se presenta antes de las afirmaciones, y debe contener los siguientes aspectos:

a) señalar que el Sujeto ha sido elegido para que funja como juez

b) explicar que el ser juez implica responder de una manera particular: el juez no va a señalar que tan de acuerdo o desacuerdo está en lo personal con cada afirmación que se le presenta, sino que debe indicar, según él, qué tan favorable o qué tan desfavorables o que tan bien o mal habla cada una de las afirmaciones acerca del O.P. Esto es, el Sujeto debe entender que lo que se le pide es un juicio objetivo, y no su propio acuerdo, sentir u opinión

c) presentar un ejemplo desarrollado, empleando una o varias afirmaciones hacia un objeto psicológico diferente al que se está investigando. Por último,

d) señalar que si se tiene alguna duda se pregunte al investigador. El propósito es asegurarse que el juez entendió bien las instrucciones, pues lo contrario invalida el instrumento.

Por lo tanto, el grupo de jueces se habrá de seleccionar con cuidado, pues de este grupo depende la validez y la confiabilidad del instrumento que se desarrolle. Para formar al grupo de jueces se sugiere que sean personas que cubran alguno o algunos de los siguientes criterios:

- a) que sean especialistas en el tema;
- b) que sean personas similares, por lo menos en algunas variables importantes, a los Sujetos de investigación;
- c) que sean personas que hayan tenido experiencia como jueces y/o que se suponga puedan desempeñar la clase de tarea que se les está pidiendo.

Una vez que el grupo de jueces (que habrán de ser 50 como mínimo) ha emitido sus juicios, esto es, que el investigador ya tiene los instrumentos respondidos, procede como sigue:

a) asigna números del 1 al 10 u 11 a las letras (A a J ó A a K) respectivamente. Así, el 1 quedará en el extremo desfavorable

del continuo (A) y el 10 u 11 en el extremo favorable (J ó K). En otras palabras, numera los intervalos de la escala gráfica;

b) revisará cada escala (cada cuestionario o instrumento que calificó cada juez) con el propósito de descartar a aquellos jueces que no entendieron lo que se esperaba de ellos.

De esta manera se eliminan aquellos cuestionarios en los que un juez colocó el 25% de sus respuestas en una sola categoría. Por ejemplo, haber enjuiciado 25 afirmaciones (de un total de 100), en la categoría 11; este es el criterio que se sigue para eliminar cuestionarios de jueces inadecuados.

Se eliminan también los cuestionarios de aquellos jueces que por ejemplo hayan empleado la misma categoría o intervalo de respuesta para enjuiciar un par de afirmaciones claramente opuestas. Es decir, supóngase que un juez calificó en el intervalo 11 (favorable), las siguientes afirmaciones: *Es muy recomendable que la mujer trabaje fuera del hogar* y *La mujer debe dedicarse únicamente a las labores del hogar*.

Esta situación está hablando de inconsistencia o descuido por parte del juez; esta inconsistencia puede deberse a que haya olvidado su función como juez y haya empezado a responder con su propio grado de acuerdo o desacuerdo.

4. En este paso se inicia el procesamiento del escalamiento de las afirmaciones propiamente dicho:

a) se procede a obtener las distribuciones de frecuencias (**f**), de proporciones (**p**) y de proporciones acumuladas (**p.a.**) para cada afirmación de la escala, y se concentran los datos en una tabla como la Tabla 8.1.

b) se calculan para cada afirmación, sus V.E. (la mediana) y su valor intercuartilar (Valor Q: dispersión cuartilar), susbtituyendo los valores correspondientes con la siguiente fórmula:

$$V.E. = L + ( 0.50 - \sum p_{ab} ) / (pd) i$$

dónde:

V.E. = valor escalar

L = límite inferior del intervalo que comprende el 0.50 o la mediana de la distribución p.a.

$\sum p_{ab}$  = suma de proporciones por debajo del intervalo que contiene a la mediana

pd = proporción del intervalo que contiene a la mediana (proporción dentro del intervalo en el que se encuentra la mediana)

i = 1.0, amplitud de los intervalos.

Se elabora una Tabla que contenga la información referida al número de veces que cada afirmación fue calificada o clasificada en cada uno de los intervalos, seguida de la representación del dato en términos de proporciones, y por último, de proporciones acumuladas. Se puede uno auxiliar de las siguientes aseveraciones:

f= número de veces con la que cada afirmación fue calificada en una dada categoría

Por lo tanto:

$$\sum f = N$$

p = cada f se divide entre N, así:

$$p = f/N \quad \text{ó} \quad p = (1/N)f$$

pa = la suma de las p

$$\sum p = 1.00$$

N = número de jueces

Para obtener los valores Q se necesitan los centiles 25 y 75, y sus fórmulas son las siguientes:

$$C_{25} = L + (0.25 - \sum p_{ab}) / (pd)_i$$

$$C_{75} = L + (0.75 - \sum p_{ab}) / (pd)_i$$

$$Q = C_{75} - C_{25}$$

Se deben calcular, de la misma manera, los centiles 25 y 75, para poder determinar el valor intercuartilar o de dispersión de las afirmaciones evaluadas. Los pasos a seguir son semejantes a los del cálculo de los Valores Escalares.

Posteriormente, se ordenan las afirmaciones de acuerdo con el Valor Escalar obtenido, junto con sus valores intercuartilares (Q), para escoger de entre ellas las mejores, Es decir, las que presenten dispersiones mas pequeñas.

El rango intercuartilar, Valor Q, es una medida de la dispersión del 50% medio de los juicios. Los valores grandes de Q señalan desacuerdo entre los jueces, y se toman como indicador de ambigüedad de las afirmaciones.

Además del procedimiento numérico para obtener valores escalares y valores Q antes expuesto, existe uno gráfico. Los pasos a seguir son los siguientes: se traza una gráfica para cada afirmación en la que:

- a) se traza el continuo psicológico en el eje de las **xs**;
- b) en el eje de las **ys** se distribuyen los valores **pa** (ver Figura 8.2); c) los puntos en las intersecciones (**pa** y **cp**) permiten trazar la curva;

d) se trazan líneas horizontales a partir de los valores 0.25, 0.50 y 0.75; y en el lugar que alcanzan la curva se traza líneas perpendiculares al eje de las **xs**. De esta manera se obtienen los valores escalares y los valores de los centiles 25 y 75;

e) el valor Q se obtiene midiendo la distancia entre el centil 25 y el 75. Las gráficas trazadas en papel milimétrico proporcionan valores escalares y valores Q tan exactos como los que se pueden obtener con el procedimiento numérico.

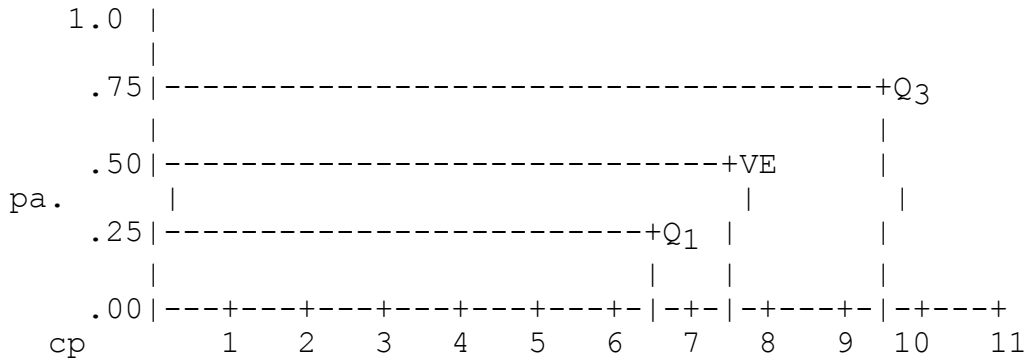


Figura 8.2. Gráfica de la proporción acumulada de una afirmación.

El último paso en la elaboración de una escala de actitudes de intervalos aparentemente iguales consiste en seleccionar las afirmaciones que constituirán la escala final o definitiva. Para esto se procede de la siguiente manera.

Primero: se ordenan las afirmaciones de acuerdo a sus valores escalares (del más pequeño al más grande) en una Tabla como la siguiente:

TABLA 8.1 Ordenamiento de los Valores Escalares.

Afirmaciones	Valores Escalares.	Valores Q
4	3.29	1.89
2	6.03	1.81
1	6.17	2.26
3	7.76	2.31
5	8.90	2.51

Segundo: se seleccionan de 20 a 25 afirmaciones de acuerdo con los siguientes criterios: a) afirmaciones cuyos valores escalares cubran el continuo psicológico de tal manera que queden igualmente espaciadas; b) afirmaciones con valores Q pequeños.

Si se observan los valores escalares de la Tabla 8.1 se puede ver que la afirmación 4 (V.E.=3.29) queda en la parte desfavorable del continuo psicológico; las afirmaciones 2 y 1 (V.E.= 6.03 y 6.17 respectivamente) cubren la parte neutral; mientras que las afirmaciones 3 y 5 (V.E.=7.76 y 8.90 respectivamente) cubren los primeros grados de favorabilidad del continuo psicológico. Es obvio que con sólo cinco afirmaciones es imposible cubrir a espacios iguales el continuo psicológico; con un número mayor de afirmaciones es perfectamente posible, si se tuvo cuidado en elaborar afirmaciones que así lo permitieran.

Por otro lado, se señaló que el valor Q debe ser pequeño, en relación a los valores Q obtenidos para ese conjunto de afirmaciones; las afirmaciones 2 y 1 tienen valores bastante similares (6.03 y 6.17) pero, sus valores Q son diferentes (1.81 y 2.26) respectivamente). Si se tuviera un número adecuado de afirmaciones se daría preferencia a la afirmación número 2, que tiene un valor Q menor que la número 1, para formar parte de la escala final.

Con esto se termina el procedimiento de elaboración de una escala de actitudes de intervalos aparentemente iguales. La segunda fase consistirá en administrar la escala final a los Sujetos de investigación. Para esto, las afirmaciones seleccionadas se imprimen en orden aleatorio en cuanto a sus valores escalares. Se pide a los Sujetos que indiquen con cuáles afirmaciones está de acuerdo y con cuáles está en desacuerdo. Una vez que los Sujetos han dado sus respuestas, se procede a calificarlos: se toman en cuenta sólo aquellas afirmaciones con las que estuvieron de acuerdo. Se establece el valor escalar de esas afirmaciones y entonces se puede seguir cualquiera de dos procedimientos de calificación: a) el procedimiento de la mediana, o b) el procedimiento de la media.

Supóngase que un Sujeto estuvo de acuerdo con las afirmaciones 2, 1, 3, y 5, cuyos valores escalares fueron: 6.03, 6.17, 7.76, y 8.9 respectivamente. Si se sigue el procedimiento de la mediana, se desechan las afirmaciones cuyos valores escalares fueron extremos (6.03 y 8.9) y se toman los valores escalares intermedios (6.17 y 7.76) y se procede de la siguiente manera, de acuerdo con la fórmula de la mediana:

$$6.17 + (7.76 - 6.17)/2 = 6.96.$$

Si el procedimiento por el que se opta es el de la media, simplemente se obtiene el promedio de los valores escalares de las afirmaciones seleccionadas (con los que estuvo de acuerdo) por el Sujeto. En este caso, y de acuerdo con la fórmula de la media:

$$\sum X/n = 6.03 + 6.17 + 7.76 + 8.9/4 = 7.21.$$

Lo anterior, para el caso en el que el Sujeto selecciona un número par de afirmaciones. Si sólo aceptara estar de acuerdo con por ejemplo tres afirmaciones, cuyos valores escalares hubieran sido: 6.17, 7.76, y 8.9, el procedimiento de la mediana indica escoger la afirmación cuyo valor escalar sea el intermedio (7.76), y ésta será su calificación individual de actitud. En el procedimiento de la media el proceso es igual que en ejemplo anterior: se obtiene el promedio de los valores escalares de las afirmaciones escogidas por el Sujeto, que en este ejemplo sería:  $22.83/3 = 7.61$ .

Los puntajes de actitud se interpretan como desfavorable-neutral-favorable de acuerdo con el lugar que éstas ocupen en el continuo psicológico. Los puntajes de 0.5 a 4.5 pueden considerarse como desfavorables; de 4.6 a 7.5, como neutrales; de 7.6 a 11.5 como favorables. Debe recordarse que el punto medio del continuo psicológico, o punto neutral corresponde al intervalo 6, mientras que el 1 es el extremo desfavorable y el 11 el favorable, en un continuo de 11 intervalos.

El que los puntajes individuales de actitud se puedan interpretar como se hace en este modelo, se justifica por la ley del juicio categórico que subyace a esta técnica: haber evaluado a un conjunto de afirmaciones iniciales por medio de un grupo de jueces, a lo largo de un continuo psicológico constituido por diferentes grados de desfavorabilidad-favorabilidad; haber seleccionado de acuerdo a los valores escalares y en los valores Q, afirmaciones que se distribuyeron a lo largo de ese continuo psicológico; y que los Sujetos hayan aceptado un cierto número de esas afirmaciones, permiten localizar a los Sujetos en el lugar que esas afirmaciones ocupan en el continuo psicológico, y por lo tanto interpretar sus puntajes individuales como desfavorables, neutrales o favorables al objeto psicológico estudiado.

### **8.2.1 Ventajas y Desventajas de este Método**

Las desventajas de este método son las siguientes:

a) la tarea que se les pide realizar a los jueces es extraña en el sentido de que probablemente nunca la hayan hecho antes, y les cueste un poco de trabajo hacerla bien.

b) la actitud de los jueces puede influir en los juicios que hagan de las afirmaciones modificando de alguna manera los valores escalares.

c) se requiere que los jueces sean buenos jueces, si no se corre el riesgo de que no entiendan lo que tienen que hacer.

Entre las ventajas tenemos las siguientes:

- a) a los jueces sólo se les pide un juicio para cada afirmación.
- b) se pueden usar muchas afirmaciones, de entre las cuales se escogerá a las mejores
- c) la escala final contiene un mayor número de afirmaciones que por ejemplo, el método anterior.
- d) se pueden construir formas paralelas o equivalentes de la escala, usando las mismas afirmaciones originales.

### **8.3 METODO DE LOS INTERVALOS SUCESIVOS**

El método de los intervalos sucesivos fue desarrollado por Edwards y Thurstone (1952) y su procedimiento estadístico de construcción deriva directamente de la ley del Juicio Categórico desarrollada por Thurstone.

Antes de continuar, cabe establecer la comparación entre este método y el de intervalos aparentemente iguales. La técnica de intervalos sucesivos comparte con la de intervalos aparentemente iguales una de sus principales ventajas: lo sencillo de la recolección de los datos de los jueces, en el sentido de que sólo se pide un juicio para cada afirmación. Sin embargo, esta es una técnica más precisa que viene a superar una de las principales limitaciones de la técnica de intervalos aparentemente iguales. Es decir, su procedimiento permite conocer la amplitud de cada intervalo, a diferencia de lo que pasa en intervalos aparentemente iguales, en donde se asume, sin poner a prueba, que cada intervalo tiene una amplitud igual a la unidad. Esta ventaja adicional que presenta el método de intervalos sucesivos explica la relación lineal que se encuentra entre sus valores escalares y los que se obtienen con el método de pares comparados para un mismo conjunto de afirmaciones. En cambio la relación entre los valores escalares obtenidos con intervalos aparentemente iguales y con pares comparados, se alejan de la linealidad, especialmente en los extremos. El método de pares comparados es uno de los procedimientos más precisos porque exige a cada uno de los jueces juicios más discriminativos, al pedirles a éstos que comparen todas las posibles combinaciones  $n(n-1)/2$  de pares de afirmaciones. Es por esta razón que el método de intervalos sucesivos adquiere su importancia, ya que elimina la principal desventaja de pares comparados (un número excesivo de juicios aún con pocas afirmaciones), pero, al mismo tiempo mantiene su precisión.

El procedimiento que se sigue en la recolección de datos es el



mismo que se lleva a cabo con el método de intervalos aparentemente iguales. Una vez que el grupo de jueces ha dado sus respuestas se asignan números a los intervalos. El número de intervalos que se puede emplear es variable, sin embargo se recomienda utilizar de 9 a 11. Los autores señalan que el número de intervalos debe ser *suficiente*, lo cual no es siempre posible anticipar de manera correcta. Es decir, deberán emplearse suficiente número de intervalos con objeto de disminuir la posibilidad de que alguna o algunas de las afirmaciones produzcan valores escalares que queden localizados en cualquiera de las categorías extremas. En otras palabras, si más del 50% de los juicios se localizan en cualquiera de los intervalos extremos, habrán de seguirse procedimientos adicionales para el cálculo de los valores escalares de estas afirmaciones. Debe señalarse, sin embargo, que existen otros procedimientos para calcular los valores escalares como los descritos por Attneave (1949) y Green (1954), quienes en vez de la mediana, utilizan la media. En este caso la obtención de los valores escalares no se ve afectada por el hecho de que más del 50% de los juicios pueda localizarse en cualquiera de los extremos.

Al igual que en intervalos aparentemente iguales, se sigue el mismo procedimiento para descartar las escalas de jueces que hayan sido inadecuados o descuidados.

El siguiente paso y los subsecuentes, se refieren al proceso de escalamiento propiamente dicho. En primer lugar, se procede a obtener para cada afirmación: frecuencias ( $f$ ), frecuencias acumuladas ( $f_a$ ) y proporciones acumuladas ( $p_a$ ).

En segundo lugar, se procede a calcular la amplitud de los intervalos. Habiendo obtenido las distribuciones de  $f$ ,  $f_a$ , y  $p_a$  para cada afirmación, se construye una matriz que está formada por todas y cada una de las distribuciones de  $p_a$ , para las cuales se supone una distribución normal, al igual que en el método de intervalos aparentemente iguales. Los valores escalares serán las medias o medianas de las mismas.

Los elementos de la Tabla 8.2 se denotan como proporciones  $p_{ij}$  en donde el subíndice  $i$  se refiere a las afirmaciones (en este ejemplo  $i = 1, 2, 3, 4, \text{ y } 5$ ) y el subíndice  $j$  señala el límite superior de los intervalos (para este ejemplo son 11 intervalos,  $j = 1.5, 2.5, 3.5, \dots, 11.5$ ). De esta manera  $p_{ij}$  indica la proporción de juicios dados a una afirmación  $i$  que se encuentra por abajo del límite superior de la categoría  $j$ . Por lo tanto,  $1 - p_{ij}$  corresponde a la proporción de juicios de la afirmación  $i$  que se encuentra por arriba del límite superior de la categoría  $j$ . Por

ejemplo, en la Tabla 8.2, la  $p_{ij} = p_{4,7} = 0.887$  indica que un 0.887 de los juicios dados a la afirmación 4 están por debajo del límite superior del intervalo 7 y que  $1 - 0.887 = 0.113$  de los juicios están por encima de éste (se omiten los decimales en la Tabla, por cuestiones de espacio).

TABLA 8.2 Distribuciones **pa** INTERVALOS SUCESIVOS

AFS	1	2	3	4	5	6	7	8	9	10	11
1	013	029	055	074	311	591	750	864	913	952	985
2	033	072	098	157	328	651	931	964	987	987	987
3	000	000	000	000	066	264	429	693	940	989	989
4	066	264	561	792	825	851	887	903	929	958	987
5	066	264	561	792	825	851	887	903	929	958	987

\*No se debe olvidar que la  $pa$  del último intervalo debe ser igual a la unidad. En el caso presente no se obtuvo debido a la constante utilizada  $1/N = 1 / 300 = .0033$  que no se aproximó a decimales para efectos de facilitar el desarrollo del ejemplo.

El siguiente paso consiste en transformar las  $p_{ij}$  a puntajes de desviación normal ( $z_{ij}$ ). Dicha transformación se realiza localizando en la Tabla de la Curva Normal (ver Edwards, 1957) el valor  $z_{ij}$  correspondiente a cada una de las  $p_{ij}$ ;  $z_{45}$  indica que el límite superior del intervalo 5 se ha transformado en una desviación normal en términos de la afirmación 4.

Cabe aclarar en este momento, que lo que se está haciendo es expresar el lugar en donde el grupo de jueces percibe al límite superior de los intervalos de las categorías asumiendo que las afirmaciones están fijas. Posteriormente, una vez conocidos los límites de los intervalos de las categorías (por medio de la estimación de las amplitudes de los mismos), se procede a determinar los valores escalares de las afirmaciones.

Continuando con el procedimiento, se debe señalar que al convertir  $p_{ij}$  a  $z_{ij}$ , se eliminarán todas aquellas  $p_{ij}$  que sean menores a 0.02 o mayores a 0.98, siempre y cuando el número de jueces sea menor a 200. Si el número de jueces es igual o mayor que 200, se pueden utilizar  $p_{ij}$  con valores hasta de 0.99 y 0.01.

Los valores de  $p_{ij}$  de 1.00 o 0.00 no se utilizan aún cuando el número de jueces sea mayor de 200, porque tales valores no permiten estimar la separación escalar de los valores escalares. Por otro lado, existe una razón estadística para eliminar **pa** mayores de 0.98 y menores de 0.02. Esta razón se refiere a las diferencias existentes entre dos valores  $z_{ij}$  que correspondan a dos valores  $p_{ij}$  localizados en las áreas centrales de una distribución normal puesto que no son iguales proporcionalmente hablando a la diferencia de dos  $p_{ij}$  que se encuentren en los

extremos de la distribución. Es decir, la diferencia entre  $p_{ij} = 0.98 = z_{ij} = +2.054$  y  $p_{ij} = 0.99 = z_{ij} = 2.326$ , entre ellas es de 0.272. Mientras que si  $p_{ij} = 0.50 = z_{ij} = 0.000$  y si  $p_{ij} = 0.51 = z_{ij} = 0.025$ , la diferencia entre estas dos es tan sólo de 0.025. Ambas difieren manejadas en unidades **z**; aunque en proporciones sean iguales, en **z** ya no lo son.

La transformación de las distribuciones **pa** de la Tabla 8.2 en distribuciones de desviaciones normales ( $z_{ij}$ ), se presentan en la siguiente tabla, la 8.3.

Tabla 8.3 Tabla de distribuciones normales  $Z_{ij}$  para los datos de la Tabla 8.2.

AFS	Intervalos Sucesivos										
	1	2	3	4	5	6	7	8	9	10	11
1	----	2.075	1.598	1.447	.493	.230	.674	1.098	.359	1.665	----
2	.838	1.454	1.293	1.007	.445	.388	.483	.799	----	----	----
3	----	----	----	----	.506	.631	.179	.504	.555	----	----
4	1.506	.631	.154	.813	.955	.041	.211	.299	.468	1.728	----
5	1.960	1.563	1.353	1.206	1.049	.860	.637	.269	.407	1.538	----

Dadas las razones anteriores se puede observar que en la Tabla 8.3 no aparecen valores para las  $p_{ij}$  menores a 0.02 y mayores a 0.98; por lo cual se puede eliminar al último intervalo.

La obtención de valores  $z_{ij}$  permite proceder a estimar la amplitud que tienen los intervalos. Para esto se construye una última matriz que estará formada por las diferencias algebraicas existentes entre los elemento de la Tabla 8.3 ( $z_{ij}$ ). La ecuación que representa la obtención de estas diferencias es la siguiente:

$$AMp_{ij} = Z_{ij} - Z_{i(j-1)}$$

dónde:

$AMp_{ij}$  = amplitud del intervalo **j** para la afirmación **i**

$Z_{ij} - Z_{i(j-1)}$  = diferencia entre un elemento dado  $Z_{ij}$  y otro que se encuentra en un intervalo anterior. Es decir, la diferencia entre los elementos del intervalo 2 y el intervalo 1; las del 3 y el 2; y así sucesivamente.

La obtención de las amplitudes de los intervalos se presentan en

la Tabla 8.4.

Tabla 8.4 Cálculo de las amplitudes de intervalo para los datos de la Tabla 8.3.

AFS	INTERVALOS SUCESIVOS								
	2-1	3-2	4-3	5-4	6-5	7-6	8-7	9-8	10-9
1	----	.477	.151	.954	.263	.417	.424	.261	0.306
2	.384	.161	.286	.562	.057	.095	.316	----	----
3	----	----	----	----	.875	.452	.325	.051	----
4	.875	.477	.659	.142	.086	.170	.088	.169	0.260
5	.397	.210	.147	.157	.189	.223	.368	.138	----
SUMA	1.656	.325	.243	.815	.464	.357	.521	.619	0.566
n	3	4	4	4	5	5	5	4	2
AM <sub>p<sub>ij</sub></sub>	.552	.331	.310	0.453	.292	.471	.304	.405	0.283
C.P.	.552	0.883	1.193	1.646	1.938	2.409	2.713	3.118	3.401

Una vez que se han obtenido las diferencias de los valores  $Z_{ij}$  (elementos de la Tabla 8.4), se ejecutan las operaciones señaladas en las últimas cuatro hileras de las mismas. A saber: se suman las columnas (sumas); se encuentra el número de celdillas ocupadas en cada columna (n); se obtienen las medias dividiendo las sumas de la hilera (1) entre los números de la hilera (2); estas medias (AM<sub>p<sub>ij</sub></sub>) proporcionan la amplitud de los intervalos.

El continuo psicológico (C.P.: cuarta hilera de la Tabla 8.4), se constituye sumando a la primera amplitud de intervalo (tercera hilera de la misma tabla) el siguiente valor (0.552 + 0.331 = 0.883); y después, a este valor (0.883), se le suma el siguiente valor de amplitud de intervalo (0.310), y así sucesivamente hasta terminar con el continuo. La tercera hilera, AM<sub>p<sub>ij</sub></sub>, es la amplitud del intervalo dentro del cual cae la mediana. Ahora sólo falta establecer para cada afirmación su valor escalar; pero se requiere también calcular los valores de las proporciones dentro del intervalo donde cae la mediana, y que se calculan de la siguiente manera:

- a la proporción acumulada del intervalo donde cae la mediana, se le resta la proporción acumulada anterior.

De esta manera, se tienen las siguientes proporciones dentro del intervalo donde cae la mediana:

Afirmación 1 = 0.591 - 0.311 = 0.280  
 Afirmación 2 = 0.651 - 0.328 = 0.323  
 Afirmación 3 = 0.693 - 0.429 = 0.264

$$\text{Afirmación 4} = 0.561 - 0.264 = 0.297$$

$$\text{Afirmación 5} = 0.658 - 0.394 = 0.264$$

Para sustituir el valor de **l** se busca en las distribuciones de pa (Tabla 8.2) el límite inferior del intervalo donde se encuentra la mediana; localizado dicho intervalo se busca su valor correspondiente en el continuo psicológico (cuarta hilera de la Tabla 8.4); el valor de la amplitud del intervalo se localiza en la tercera hilera de la Tabla 8.4 y corresponde al número del intervalo en el que cayó la mediana. A continuación se pone un ejemplo para la afirmación 1.

En la Tabla 8.2 se localiza la mediana y se observa que cae en el intervalo 6, por lo que el límite inferior de este intervalo se encuentra en el intervalo 5-4 de la Tabla 8.4, en la cuarta hilera de la misma (continuo psicológico) y le corresponde por lo tanto 1.667  $L_i = 1.646$

El valor  $AMP_{ij}$  se busca en la tercera hilera (amplitudes de intervalo) de la Tabla 8.4, el valor correspondiente al intervalo 6 que fue donde cayó la mediana. En esta Tabla 8.4, en el intervalo 6-5, el valor  $AMP_{ij} = 0.292$

El valor de **pab** se encuentra en la Tabla 8.2 y corresponde en este caso a 0.311. Por último, **pd** es igual al valor de la mediana (0.59) menos la suma de proporciones por debajo de donde cae la mediana (0.311), o sea: 0.280; ambos valores se localizan en la Tabla 8.2. Substituyendo en la fórmula, se tiene:

$$\begin{aligned} VE_1 &= 1.646 + (0.50 - 0.311/0.59 - 0.311)(0.693) = 0.292 \\ &= 1.619 \end{aligned}$$

En la Tabla 8.5 se presentan los valores escalares de todas las afirmaciones del ejemplo que se ha venido presentando.

-----  
 TABLA 8.5 Computación de los Valores Escalares  
 -----

AFIRMACIONES	$(0.50 - \sum p_{ab}) / (pd)$	$AMP_{ij}$	= V.E.
1	1.46+	$[(0.50 - 0.311)/0.280]$ (0.292)	=1.619
2	1.646+	$[(0.50 - 0.328)/0.323]$ (0.292)	=1.801
3	2.409+	$[(0.50 - 0.429)/0.264]$ (0.304)	=2.490
4	0.552+	$[(0.50 - 0.264)/0.297]$ (0.331)	=0.815
5	0.552+	$[(0.50 - 0.264)/0.297]$ (0.331)	=2.875

El siguiente paso, al igual que en el caso del método de los intervalos aparentemente iguales, consiste en calcular los valores Q. Para esto se sigue el procedimiento señalado en el método de los intervalos aparentemente iguales con los cambios correspondientes para el método de los intervalos sucesivos, señalados en los párrafos inmediatamente anteriores, al obtener los valores escalares. Existe otro procedimiento, el análisis de la dispersión discriminativa, que puede utilizarse al igual que los valores Q, como uno de los criterios para la selección final de las afirmaciones que constituirán la escala definitiva. Este procedimiento se presenta a continuación.

**8.3.1 Análisis de la Dispersión Discriminativa.**

En el análisis de la dispersión discriminativa se trabaja a partir de los datos contenidos en la Tabla 8.6, es decir, con las proporciones transformadas a puntajes de desviación normal.

A continuación se ejemplifica el cálculo de las dispersiones discriminativas con un ejemplo ficticio. En la Tabla 8.6, se esquematizan los pasos que se habrán de llevar a cabo.

Una vez obtenidos los valores de las dispersiones discriminativas se observa si sus magnitudes son semejantes o si alguna o algunas de ellas son demasiado diferente de las demás.

En el ejemplo que se presenta, se observa que los valores de las dispersiones discriminativas fueron:  $\sigma_1 = 0.694$ ;  $\sigma_2 = 0.743$ ;  $\sigma_3 = 1.153$ ;  $\sigma_4 = 1.181$ ; y  $\sigma_5 = 1.225$ . Como se puede ver, los valores de las afirmaciones 1 y 2 son diferentes de los valores de las dispersiones de las afirmaciones 3, 4 y 5. Uno de los criterios para eliminar afirmaciones es precisamente el alejamiento de

algunos de los valores de los demás valores. Sin embargo, antes de eliminarlas, es conveniente y necesario ver los valores de las amplitudes de intervalo que ellas obtuvieron. Se observa que en la Tabla 8.4, que las amplitudes de intervalo (hilera tres de la parte inferior de la Tabla) de las afirmaciones 1 y 2 no discrepan mucho de los valores de amplitud de intervalo del resto de las afirmaciones. Lo anterior permite concluir que el posible desplazamiento de estas afirmaciones (1 y 2) respecto a las demás (3, 4 y 5) se minimiza, y por lo tanto no es importante, con lo cual pueden permanecer en la escala.

-----  
 TABLA 8.6 Cálculo del Análisis de la Dispersión  
 Discriminativa con Datos Ficticios.  
 -----

Pasos de computación	Afirmaciones				
	1	2	3	4	5
1. Suma de los valores $z_{ij}$ elevados al cuadrado	14.81	13.98	5.37	13.59	14.34
2. Sumas de los valores $z_{ij}$	-0.408	-2.374	-.257	6.51	-6.93
3. Se dividen las sumas $z_{ij}$ de la hilera 2 entre $n$ , el número de afirmaciones	0.33	1.127	.013	8.48	9.6
4. Se restan los valores de la hilera 3 de los valores correspondientes de la hilera 1	14.77	12.85	5.35	5.11	4.74
5. Se dividen los valores de la hilera 4 entre el número de afirmaciones	2.954	2.570	1.07	1.022	0.948
6. Se saca raíz cuadrada de los valores de la hilera 5	1.718	1.603	1.034	1.010	0.975
7. Se obtiene el recíproco de los valores de la hilera 6	0.582	0.623	0.967	0.990	1.027
8. Se suman los valores de la hilera 7			4.189		

-----



9. Se obtiene el valor de la constante <b>k</b>			
n = número de afirmaciones			1.193
-----			
10. Se estima la dispersión discriminativa para cada afirmación	(.582)	(1.193)	(.99) (1.19)
		(.623) (1.193)	
		(.967) (1.193)	
		(1.027) (1.193)	
	$\sigma_1 = .694$		$\sigma_4 = 1.181$
	$\sigma_2 = .743$		
		$\sigma_3 = 1.153$	
			$\sigma_5 = 1.225$
-----			

Por último, se puede comprobar si se han cometido errores en el cálculo de las dispersiones, sumándolas. Esta suma debe ser igual al número de afirmaciones. En este caso, la suma de las dispersiones es igual a  $4.99 = 5$ .

### 8.3.2 Prueba de Consistencia Interna

La prueba de consistencia interna permite comprobar la suposición de que los valores escalares obtenidos por medio de este procedimiento son consistentes con los datos empíricos, y para esto se procede a establecer una comparación entre las proporciones teóricas derivadas del modelo de escalamiento y las proporciones empíricas provenientes de los juicios dados por la muestra que fungió como juez para la elaboración de la escala de actitud. Entre menor sea la discrepancia entre las proporciones teóricas ( $p'_{ij}$ ) y las proporciones observadas ( $p_{ij}$ ) mayor consistencia interna, en lo que se refiere a los supuestos del modelo.

Para llevar a cabo la prueba de consistencia interna se trabaja con los valores escalares de las afirmaciones y con las estimaciones de las amplitudes de sus intervalos. A partir de éstas se derivan un conjunto de proporciones que se denominan teóricas ( $p'_{ij}$ ).

En nuestro ejemplo se tienen cinco valores escalares, y 9 amplitudes de intervalo, en donde **k** es el número de intervalos sucesivos al que se le restan dos, ya que no se estima la amplitud del primero ni del último intervalo. De esta manera se tienen 14 parámetros que se emplearán para reproducir  $5(11-1) = 50$  proporciones (las  $p_{ij}$  de la Tabla 8.2).

Los pasos que se siguen para el cálculo de la consistencia interna

son los siguientes:

a) se construye una Tabla que tiene en la primera columna los valores escalares de las afirmaciones (Tabla 8.5); las columnas restantes consisten o tienen como encabezado, a las amplitudes de intervalo acumuladas, procedentes de la Tabla 8.4, aunadas a un origen arbitrario (0.000) que se coloca antes del primer valor de amplitud acumulado. Los elementos de esta nueva Tabla (8.7, de  $z'_{ij}$ ) se obtienen restando a cada amplitud de intervalo todos y cada uno de los valores escalares de las afirmaciones. De esta manera los elementos en el cuerpo de la Tabla son las ( $z'_{ij}$ ) distribuciones teóricas de las desviaciones.

-----  
 TABLA 8.7. Distribuciones Teóricas de Desviaciones, obtenidas con los valores escalares y las amplitudes de intervalo acumuladas.  
 -----

V.E.	Af.	Amplitudes Acumuladas									
		.000	.544	.909	1.219	1.667	2.360	2.836	3.211	3.750	4.315
2.13	1	-2.13	-1.585	-1.221	-.911	-.463	.230	.706	1.081	1.620	2.185
2.03	2	-2.03	-1.486	-1.121	-.811	-.363	.330	.806	1.081	1.720	2.285
2.93	3	-2.93	-2.386	-2.021	-1.711	-1.263	-.570	-.094	.281	.820	1.385
3.42	5	-3.42	-2.876	-2.511	-2.201	-1.753	-1.06	-.584	-.209	.330	.895

-----

b) El siguiente paso es construir otra tabla (Tabla 8.8). Para esto se busca en el cuerpo de la tabla de desviaciones normales  $z$  correspondientes a proporciones  $p$  de una distribución normal dicotomizada (ver Edwards, 1957) cada una de las  $z'_{ij}$  y se transforman a  $p'_{ij}$ , para dar como resultado la Tabla 8.8.

TABLA 8.8 Proporciones teóricas obtenidas a partir de las distribuciones teóricas de las desviaciones.

Afs.	Intervalos Sucesivos									
	1	2	3	4	5	6	7	8	9	10
1	.016	.056	.111	.463	.321	.591	.760	.861	.948	.986
2	.021	.068	.129	.467	.358	.630	.790	.882	.958	.989
3	.001	.008	.021	.043	.103	.284	.462	.611	.533	.917
4	.202	.386	.530	.650	.798	.936	.978	.992	.999	.999
5	.001	.002	.006	.013	.041	.144	.279	.417	.630	.815

c) El siguiente paso es obtener las diferencias absolutas entre  $p_{ij}-p'_{ij}$ ; es decir, a cada elemento de la Tabla 8.2 se le resta su elemento correspondiente de la Tabla 8.8, dando como resultado la Tabla 8.9.

TABLA 8.9 Diferencias absolutas entre proporciones observadas y teóricas (Entre los datos de las Tablas 8.2 y la 8.8).

Afs.	Intervalos Sucesivos									
	1	2	3	4	5	6	7	8	9	10
1	.003	.027	.056	.389	.010	.000	.010	.003	.036	.034
2	.012	.004	.031	.310	.030	.021	.141	.082	.029	.002
3	.001	.008	.021	.043	.037	.020	.033	.082	.407	.072
4	.136	.122	.031	.142	.027	.085	.091	.089	.070	.041
5	.025	.057	.082	.101	.106	.052	.017	.023	.028	.123

Debe hacerse notar que en las Tablas 8.8 y 8.9 sólo aparecen k-1 intervalos; lo anterior se debe a que en la Tabla original (8.2), sólo pueden variar libremente k-1 intervalos, ya que el último intervalo necesariamente es igual a la unidad.

d) Por último, el valor de la prueba de consistencia interna se obtiene sumando todos y cada uno de los elementos de la Tabla 8.9 y dividiendo esta suma entre el número total de celdillas o elementos (5 afirmaciones por 10 intervalos = 50). Para el ejemplo en cuestión se tiene entonces:

$$\sum |p_{ij} - p'_{ij}|/n = 3.402/50 = 0.068$$

En este caso se dice que se tiene un error absoluto promedio de 0.07; como esta prueba carece de pruebas de significancia, los resultados obtenidos se tendrán que comparar con los resultados reportados por otros investigadores. De esta manera, Edwards (1957), reporta que para 17 afirmaciones escaladas en 10 intervalos se obtuvo un valor de 0.021; y de 0.025 para 10 afirmaciones con 9 intervalos. En este caso el valor obtenido (0.07) es mayor que el reportado por otros investigadores, pero habrá de tomarse en consideración que el ejemplo en cuestión tiene un número muy pequeño de afirmaciones, ya que se desarrolló con propósitos didácticos buscando simplificar la exposición. Se esperaría que con un número mayor de afirmaciones se obtuviera valores que fueran de 0.021 a 0.025.

### **8.3.3 Selección de Afirmaciones, Puntajes de Actitud e Interpretación**

Las afirmaciones que se seleccionan para formar la escala final o definitiva son aquellas que tienen valores escalares que representan a todo el continuo psicológico, y cuyos valores  $Q$  o dispersiones discriminativas sean los más pequeños o los más semejantes en relación con el conjunto total de afirmaciones con las que se empezó el procedimiento escalar. Es decir, los criterios de selección de afirmaciones para la escala definitiva son los mismos que se siguen en el caso de los intervalos aparentemente iguales.

La obtención de los puntajes de actitud es semejante al caso de los intervalos aparentemente iguales, y su interpretación también. Se recomienda ver la sección correspondiente en ese capítulo.

### **8.3.4 Valores Escalares Para Afirmaciones con 50% o Más de los Juicios en el Primer o Último intervalo.**

Se puede dar el caso de que algunas afirmaciones tengan el 50% o más de los juicios en el primer o en el último intervalo. En este caso el cálculo de los valores escalares difiere del hasta ahora presentado. A continuación se presentan los procedimientos a seguir.

Se deberá examinar con cuidado las distribuciones de las proporciones acumuladas (Tabla 8.2) para determinar si la mediana de una o algunas afirmaciones se encuentra en el primer intervalo. Si el número de afirmaciones que se encuentra es pequeño se aconseja eliminarlas y proceder a la obtención de los valores escalares con el procedimiento ya señalado. Si el número de afirmaciones en estas condiciones es importante, o aun siendo pocas pero no se quieren eliminar por alguna razón, se procederá entonces a extender el continuo psicológico hacia la izquierda y a cambiar el punto de origen arbitrario del límite superior del primer intervalo al punto medio del mismo; y empleando éste como nuevo punto de origen se procede a encontrar el valor escalar de esa(s) afirmación(es). Los pasos a seguir son los siguientes:

a) Se ignorarán todas aquellas afirmaciones cuyas entradas o proporciones sean menores de 0.04 en el primer intervalo.

b) Se construye una Tabla en la que la primera columna tenga proporciones con valores igual a la mitad de los de las proporciones que originalmente constituyeron el primer intervalo; de hecho, por medio de esta operación se ha extendido el continuo psicológico hacia la izquierda.

c) La segunda columna estará constituida por las proporciones del primer intervalo con sus valores originales (los que se tenían antes de dividirlos entre dos para formar el punto medio o primera columna de esta nueva tabla).

d) Se obtienen valores **z** para estas dos columnas, produciéndose así una tercera y cuarta columnas. Los valores **z** se buscan en la misma Tabla que en la ocasión anterior (ver Edwards, 1957).

e) Se obtiene una quinta columna que contiene la diferencia entre los valores **z** de la columna (4) menos los valores **z** de la columna (3).

f) Se suman los valores de la quinta columna y se divide esta suma entre el número de celdillas ocupadas. Este resultado es la amplitud del primer intervalo que ya se ha extendido hacia la izquierda. Conociendo este valor se procede a obtener el valor escalar de las afirmaciones de acuerdo a la fórmula ya conocida.

$$V.E = \left( \frac{0.50 - \sum p_{ab}}{pd} \right) AM_{p_{ij}}$$

dónde:

V.E.= valor escalar para una afirmación dada

$\Sigma p_{ab}$  = valor de la proporción de la columna 1 correspondiente a la afirmación que tuvo el 50% o más de los juicios en el primer intervalo

pd= diferencia que existe entre el valor de la proporción de la columna 1 y el de la columna 2

AM<sub>pij</sub> = valor obtenido de la suma de los valores de la columna (4-3) dividido entre el número de celdillas ocupadas.

A continuación se presenta el esquema de una tabla con el contenido necesario para obtener los valores escalares de afirmaciones con el 50% o más de los juicios en el primer intervalo.

-----  
 Tabla 8.10 Esquema con el contenido que se requiere para la obtención de V.E. para afirmaciones con el 50% o más de juicios en el primer intervalo.  
 -----

AFS.	1	2	3	4	5
	Punto Medio	Prop. originales del 1er. intervalo	Valores z la columna 1	Valores Z la columna 2	Diferencias de las columnas 4-3
1	.200	.400	-.842	-.253	.589
2	.315	** .630	-.482	.332	.814
3	.190	.380	-.878	-.305	.573
4	.265	** .530	-.628	.075	.703
5	----	.030	----	----	----
					$\Sigma$ 2.679
					n = 4
					= 0.669
n	n	n	n	n	

\*Se ignora la proporción de esta afirmación por ser menor a 0.04

\*\*Las proporciones del primer intervalo de las afirmaciones 2 y 4 contienen más del 50% de los juicios.

$$V.E. = [(0.50 - 0.315) / 0.315] - 0.669 = 0.392$$

$$V.E. = [(0.50 - 0.265) / 0.265] - 0.669 = 0.595$$

Cuando el 50% de los juicios se encuentran en el último intervalo se tiene que extender el continuo psicológico hacia la derecha,

obteniendo el punto medio del penúltimo intervalo que corresponderá al límite superior de éste o al límite inferior del último. A continuación se presenta una tabla esquemática (Tabla 8.11) con la información necesaria para calcular los valores escalares de las afirmaciones que se encuentran en la situación señalada.

TABLA 8.11. Valores necesarios para obtener Valores Escalares de Afirmaciones con el 50% o más de juicios en el último intervalo.

AFS.	1	2	3	4	5	6
	Proporciones del límite inferior del último intervalo o (penúltimo)	Proporciones del límite inferior del último intervalo o	Punto medio de la Col.2 mas la Col. 1	Valores Z de la Col..842 11.282	Valores Z de la Col. 3	Diferencias
1	.8	.2	.9	.842	1.282	0.44
2	.9	.1	.95	1.282	1.645	0.363
3	.3	*.7	.65	-0.524	0.385	0.909
4	.4	*.6	.70	0.253	0.524	0.777
5	.98	.02				
6						
.						
.	n	n	n	n	n	Suma

La suma es =  $2.489/4 = 0.622$

\*Las proporciones del ultimo intervalo de las afirmaciones 3 y 4 contienen mas del 50% de los juicios.

Se procede a explicar la Tabla 8.11. Los valores que aparecen en la columna (1) son las proporciones originales del penúltimo intervalo.

En la columna (2) aparecen los valores que corresponden a restarle a la unidad los valores de la columna (1). Estas son las proporciones que cayeron en el último intervalo y las que tienen un 50% o más de juicios en este intervalo están señaladas con \*.

En la columna (3) aparecen los valores que resultan de dividir entre dos a los valores de la columna (2) más el valor de la

columna (1). Este es el nuevo punto medio; se observa que no aparece un nuevo punto medio correspondiente a la afirmación 5 por haber tenido ésta una proporción menor a 0.04. Las columnas (4) y (5) son los valores **z** de las columnas (1) y (3), respectivamente.

La columna (6) muestra las diferencias entre los valores **z** de la columna (5) menos los de la columna (4). Estas diferencias se suman y se dividen entre el número de celdillas ocupadas, obteniéndose así la amplitud del último intervalo al haber alargado, recorrido o desplazado el continuo psicológico hacia la derecha.

Por último, se procede a calcular los Valores Escalares de las afirmaciones en situación de tener 50% o más de los juicios en el último intervalo: afirmaciones 3 y 4, por medio de la fórmula:

$$V.E = \left( \frac{0.50 - \sum p_{ab}}{pd} \right) AM_{pij}$$

$$V.E_3 = 2.65 + \left[ \frac{(0.50 - 0.30)}{0.35} \right] 0.622 = 3.00$$

$$V.E_5 = 2.65 + \left[ \frac{(0.50 - 0.40)}{0.30} \right] 0.622 = 2.857$$

En este caso la distancia a partir del origen del continuo psicológico acumulado al límite inferior del último intervalo (el punto medio obtenido) fue de 2.650. Las proporciones del denominador corresponden a las diferencias que existen entre las proporciones de la primera y tercera columna de la Tabla 8.11.

### 8.3.5 Ventajas y Desventajas de este Método

Las desventajas son las mismas que las del modelo anterior, con excepción del inciso b); tiene además, otras desventajas:

- a) el continuo que queda puede ser muy reducido.
- b) el trabajo estadístico es laborioso, sobre todo en lo que se refiere al cálculo de los valores escalares de aquellas afirmaciones que caen en el primer o último intervalo, para las cuales hay que seguir el procedimiento especial ya señalado.

Entre sus ventajas, se tienen:

- a) es un modelo teóricamente fuerte ya que comparte aspectos del de pares comparados.
- b) tiene una prueba de consistencia interna.
- c) permite usar muchas afirmaciones, y comparte además, las ventajas del modelo anterior.

### REFERENCIAS



- Attneave, F. (1949): A method of graded dichotomies for the scaling of judgements. **Psychological Review**, 56, pp. 334-340.
- Edwards, A.L. (1957): **Techniques of Attitude Scale Construction**. Nueva York: Appleton Century Crofts.
- Edwards, A.L. (1952): The scaling of stimuli by the method of successive intervals. **Journal of Applied Psychology**, 36, pp. 118-122.
- Edwards, A.L. y Thurstone, L.L. (1952): An internal consistency check for scale values determined by the method of successive intervals. **Psychometrika**, 17, pp. 169-180.
- Green, B.F. (1954): Attitude Measurement. En: G. Lindzey (Ed.): **Handbook of Social Psychology**. Cambridge, Mass.: Addison-Wesley, pp. 335-469.
- Thurstone, L.L. y Chave, E. (1929): **The Measurement of Attitude**. Chicago: University of Chicago Press.

## **IX. ESCALAS DE ACTITUDES: APROXIMACION CENTRADA EN LA RESPUESTA**

En esta sección presentaremos dos métodos: el método de rangos sumariados y el del análisis de escalograma. El primero se puede clasificar estrictamente como de aquellos métodos centrados en la respuesta; el segundo, aunque es centrado en la respuesta, no lo es en sentido estricto puesto que también se puede considerar como centrado en el estímulo. Sin embargo, y con afán de balancear los capítulos dedicados a la medición de las actitudes, y en virtud de que tampoco se puede clasificar como perteneciente a los métodos multidimensionales, objeto del siguiente capítulo, lo presentamos en esta sección.

### **9.1 Método de los Rangos Sumarizados**

El método de los rangos sumariados fue desarrollado por Rensis Likert (1932) teniendo en mente la búsqueda de un modelo que no requiriera tanta laboriosidad en su construcción y que permitiera por medio del mismo, determinar alguna forma de validez de la escala que se construye, a diferencia de los métodos hasta ahora presentados.

A primera vista, este método es el más sencillo y su lógica se fundamenta en el procedimiento de obtención de validez concurrente por medio de grupos contrastados. Sin embargo, tiene un problema fundamental que los métodos vistos anteriormente no tienen. Este modelo carece de un continuo psicológico donde se escalen las afirmaciones que habrán de formar el instrumento final. Sin embargo y visto de otra manera este modelo no hace un supuesto difícil de comprobar como en el que se refiere al punto cero del continuo psicológico. En el apartado dedicado a la interpretación de calificaciones se hablará con mayor amplitud al respecto.

Este modelo consta también de dos fases: la primera corresponde a la elaboración de la escala propiamente dicha. La segunda, al empleo de la escala en la investigación o estudio de interés. Se iniciará en seguida la explicación de la primera fase.

El primer paso tiene que ver con la selección del objeto psicológico, para el cual se habrá de definir una dimensión de actitud.

El segundo paso consiste en elaborar 40 o más afirmaciones, la mitad de ellas favorables al objeto psicológico y la otra mitad desfavorables.

En el siguiente paso se selecciona una muestra de Sujetos, 100 o

más, que compartan características semejantes a aquellas de la muestra que participará en la investigación, semejantes en por lo menos aquellas variables que se consideren como las más importantes, de acuerdo con los objetivos de la investigación.

En seguida se presenta a esta muestra de sujetos el conjunto inicial de afirmaciones (40 o más) para que respondan a ellas en una de las siguientes cinco categorías u opciones de respuesta:

Muy de Acuerdo (MA)	De Acuerdo(A)	Indeciso(I)
En Desacuerdo (D)	Muy en Desacuerdo(MD)	

Cada Sujeto habrá de señalar para cada afirmación una de las categorías: aquella que mejor exprese su sentir u opinión. Para facilitar la tarea de los Sujetos, se les pone un ejemplo referido a otro objeto psicológico, haciendo hincapié en las diferencias existentes entre estar Muy de Acuerdo (MA) y De Acuerdo(A), así como En Desacuerdo(D) y Muy en Desacuerdo(MD).

**9.1.1 Formato de la Escala**

El formato de la escala puede ser diverso. Siempre se tendrán que establecer instrucciones referidas a lo que el Sujeto habrá de hacer, y un ejemplo aclaratorio o explicativo. A continuación se presentan cuatro formatos diferentes:

- a) las categorías de respuesta aparecen en la parte superior de la hoja y las afirmaciones precedidas de un paréntesis donde habrán de responder los Sujetos, se enlistan en la parte inferior de la hoja.
- b) las categorías de respuesta pueden tener iniciales : MA, A, I, D, MD; o números: MA=5, A=4, I=3, D=2, MD=1.
- c) las categorías de respuesta deben explicarse con claridad, ya sea que se empleen siglas o números.
- d) se escribe cada afirmación e inmediatamente abajo de cada una de ellas se colocan paréntesis con siglas o números; o líneas con siglas y números. Por ejemplo:

"Las mujeres deben participar activamente en la política"

( )MA;	( )A;	( )I;	( )D;	( )MD	ó
( )5	( )4	( )3	( )2	( )1	ó
_____	_____	_____	_____	_____	
MA	A	I	D	MD	ó

5
4
3
2
1.

Debe tenerse mucho cuidado en colocar las afirmaciones en forma aleatoria para evitar tener primero un conjunto de afirmaciones favorables y después un conjunto de afirmaciones desfavorables al objeto psicológico. Se recomienda poner una de cada una, alternadamente a lo largo de todo el instrumento, cuidando también que las opciones de respuesta unas veces se inicien con MA y otras con MD, con objeto de disminuir al máximo la posibilidad de que aparezcan predisposiciones a responder (*response sets*).

### 9.1.2 Proceso de Computación y Análisis de Datos

Los pasos a seguir para el cómputo y análisis de datos son los siguientes:

- a) identificar en cada escala aplicada la dirección (favorable-desfavorable) de cada afirmación.
- b) asignar pesos a las categorías señaladas en cada afirmación. El criterio de asignación de pesos es el siguiente: se dará el peso más alto a la categoría de respuesta que indique la actitud más favorable. Si se escogen pesos de 1 a 5, las afirmaciones con dirección favorable tendrán para sus opciones, los siguientes pesos: MA=5, A=4, I=3, D=2 y MD=1. Para las afirmaciones con dirección desfavorable, la asignación de pesos se invierte: MA=1, A=2, I=3, D=4 y MD=5.

Cuando alguien está muy de acuerdo (MA) con una afirmación favorable, esta opción indica la actitud más favorable al objeto psicológico; cuando alguien está muy en desacuerdo (MD) con una afirmación desfavorable, ésta es la opción que indica la mayor favorabilidad. De esta manera se cumple con el criterio de dar a la opción que indique mayor favorabilidad el peso más alto.

### 9.1.3 Selección de las Afirmaciones

Los pasos que se siguen para seleccionar las afirmaciones que compondrán la escala de actitud final se detallan a continuación.

En primer lugar se concentran las respuestas dadas por los Sujetos en una tabla como la que se presenta a continuación (Tabla 9.1).

-----  
 Tabla 9.1 Concentración de las respuestas dadas por los  
 Sujetos a cada una de las afirmaciones.  
 -----

**Afirmaciones**

	1	2	3	4	5	6	7	8	9	10	11	12	13	PTS	
<b>Ss</b>															
1	4	0	0	1	3	2	2	4	2	0	1	0	1	20	pb
2	3	3	4	3	2	3	4	4	3	3	2	4	3	44	pa
3	0	1	2	0	1	2	0	2	2	1	1	0	0	12	pb
4	4	3	3	4	4	3	2	2	3	4	4	3	3	42	pa
5	2	3	3	4	3	3	2	3	3	4	4	3	3	40	pa
6	3	2	3	3	3	4	4	2	2	3	3	4	4	40	pa
7	0	2	0	1	0	2	0	0	3	0	2	0	0	10	pb
8	2	3	4	1	3	2	0	2	1	3	2	3	2	28	
9	0	0	1	0	2	0	2	0	1	1	2	2	2	13	pb
10	4	3	4	3	4	4	3	3	4	3	3	3	4	45	pa
11	2	3	4	1	2	3	3	4	3	2	1	2	3	33	
12	1	0	2	0	3	0	3	0	1	2	1	2	2	17	pb
13	4	3	2	2	3	2	3	2	2	1	2	2	2	30	
14	0	1	2	3	3	2	2	3	2	3	1	2	0	24	
15	4	2	3	3	2	2	0	1	0	0	3	2	1	23	
16	3	3	4	4	3	3	4	2	1	0	0	2	2	31	
17	4	2	2	4	2	0	1	3	0	2	3	2	1	26	
18	2	3	1	0	4	2	3	4	2	1	0	3	2	27	
19	1	2	0	2	3	4	2	0	3	2	0	1	3	23	
20	0	0	3	4	2	3	2	1	3	2	4	1	2	27	
.															
.															
.															
n															

En segundo lugar, se seleccionan de la distribución de puntajes totales (PTS) el 25% de puntajes más altos y el 25% de puntajes más bajos. En la Tabla 9.1 se han señalado con p.a. y p.b. respectivamente.

Estos dos grupos contrastados de Sujetos (sus respuestas a las afirmaciones) constituirán los grupos criterio para seleccionar las afirmaciones que formarán la escala final.

El tercer paso consiste en evaluar individualmente todas las afirmaciones. Para esto se llevan a cabo las siguientes acciones:

a) Se concentran en una tabla (Tabla 9.2) los puntajes de los Sujetos de los grupos alto y bajo. La Tabla de concentración es la siguiente:

Tabla 9.2 Concentración de los Puntajes Altos y Bajos

	Afirmaciones													Puntajes Totales	
	Ss	1	2	3	4	5	6	7	8	9	10	11	12		13
Grupo	2	3	3	4	3	2	3	4	4	3	3	2	4	3	44
Puntajes Altos	4	4	3	3	4	4	3	2	2	3	4	4	3	3	42
	5	2	3	3	4	3	3	2	3	3	4	4	3	3	40
	6	3	2	3	3	3	4	4	2	2	3	3	4	4	40
	10	4	3	4	3	4	4	3	3	4	3	3	3	4	45
Grupo	1	4	0	0	1	3	2	2	4	2	0	1	0	1	20
Puntajes Bajos	3	0	1	2	0	1	2	0	2	2	1	1	0	0	12
	7	0	2	0	1	0	2	0	0	3	0	2	0	0	10
	9	0	0	1	0	2	0	2	0	1	1	2	2	2	13
	12	1	0	2	0	3	0	3	0	1	2	1	2	2	17

b) Se obtiene la frecuencia con que cada categoría de respuesta de cada afirmación fue seleccionada; se eleva al cuadrado el resultado de multiplicar la frecuencia por el peso de la categoría. En la Tabla 9.3 se presentan ejemplos para las afirmaciones 1, 2, 5 y 9.

Tabla 9.3 Evaluación individual de las afirmaciones tomando como criterio los puntajes de dos grupos extremos o contrastados (datos tomados de la Tabla 9.2)

Af.1		25% de puntajes bajos				25% de puntajes altos			
Cat.de Resp.	X	F	FX	FX <sup>2</sup>	Cat.de Resp.	X	F	FX	FX <sup>2</sup>
M.A.	4	1	4	16	M.A.	4	2	8	32
A.	3	0	0	0	A.	3	2	6	18
I.	2	0	0	0	I.	2	1	3	4
D.	1	1	1	1	D.	1	0	0	0
M.D.	0	3	0	0	M.D.	0	0	0	0
Sumas		5	5	17			5	16	54
Notación		NB	FX <sub>b</sub>	FX <sup>2</sup> <sub>b</sub>		NA	FX <sub>a</sub>	FX <sup>2</sup> <sub>a</sub>	
Af.2									
M.A.	4	0	0	0	M.A.	4	0	0	0
A.	3	0	0	0	A.	3	4	12	36
I.	2	1	2	4	I.	2	1	2	4

D.	1	1	1	1	D.	1	0	0	0
M.D	0	3	0	0	M.D.	0	0	0	0
Sumas		5	3	5			5	14	40
Notación		NB	FX <sub>b</sub>	FX <sup>2</sup> <sub>b</sub>		NA	FX <sub>a</sub>	FX <sup>2</sup> <sub>a</sub>	

c) Se procede a hacer un análisis de reactivos para determinar el poder discriminativo de cada afirmación. Se dice que una afirmación tiene poder discriminativo si se encuentran diferencias estadísticamente significativas entre las respuestas dadas por los dos grupos criterio (el grupo alto corresponde a Sujetos con una opinión favorable ---puntajes altos---hacia el objeto psicológico; el grupo bajo corresponde a Sujetos con una opinión desfavorable ---puntajes bajos--- hacia el objeto psicológico). Es en este sentido en el que se habla de grupos contrastados.

El poder discriminativo de las afirmaciones lo proporciona la prueba **t** cuya fórmula es la siguiente:

$$t = \frac{M_a - M_b}{\sqrt{\frac{\sum fx^2_a - (\sum fx_a)^2/NA + \sum fx^2_b - (\sum fx_b)^2/NB}{n(n-1)}}$$

dónde:

M<sub>a</sub> = media de respuestas a la afirmación x del grupo alto

M<sub>b</sub> = media de respuestas a la afirmación x del grupo bajo

$\sum fx^2_a - (\sum fx_a)^2/NA$  y  $\sum fx^2_b - (\sum fx_b)^2/NB$  = suma de cuadrados de la distribución del grupo alto (A) y bajo (B) respectivamente

n(n-1) = n = número de pares de Sujetos

A continuación se exponen en la tabla 9.4 los resultados de haber substituído en la fórmula anterior los datos correspondientes a las afirmaciones 1, y 2, del ejemplo.

Tabla 9.4 Substitución de valores en la fórmula **t**.

Datos tomados de la Tabla 9.3

Afirmaciones	t 0.05	sig.
1	2.55	no
2	4.91	si

d) Los resultados de la aplicación de la prueba **t** a los datos de los grupos criterio de este ejemplo muestran (ver Tabla 9.4) que la afirmación 2 tiene poder discriminativo y que por lo tanto sería seleccionada para formar parte de la escala final. Esta decisión se basa en la significancia estadística del resultado de la prueba **t**. Si el valor de **t** a un nivel de significancia de  $\alpha=0.05$  con grados de libertad =  $N-1$  es igual a 2.77, para las afirmaciones del ejemplo desarrollado, el valor obtenido de **t** debería de ser igual o más grande que el valor tabulado de 2.77 (consultar una Tabla de valores **t**), para que dichas afirmaciones se puedan considerar como suficientemente discriminativas y de esta manera formen parte de la escala final. Este procedimiento habrá de llevarse a cabo con todas y cada una de las afirmaciones iniciales (40 o más).

Likert (1932) estableció como criterio general para la aceptación de afirmaciones una  $t=1.75$  para grupos criterio con 25 Sujetos o más. Este valor de **t** corresponde a un nivel de significancia de aproximadamente 0.10.

e) Habiendo encontrado los valores **t** para todas las afirmaciones, se procede a seleccionar de entre ellas, a 20 o 25 con los valores **t** más altos, cuidando que aproximadamente la mitad sean favorables al objeto psicológico, y la otra mitad sean desfavorables. Es pertinente señalar en este momento que si se obtiene un número suficiente de afirmaciones con valores **t** estadísticamente significativos, se proceda a construir dos escalas paralelas, con **t's** similares y contenidos semejantes, con objeto de contar con elementos suficientes para poder determinar la confiabilidad de formas paralelas (estabilidad temporal y homogeneidad de varianza).

En este momento se termina la primera fase del proceso de construcción de una escala de actitudes con el método de rangos sumariados. La segunda fase consiste en aplicar estas 20 o 25 afirmaciones (la mitad favorables y la otra mitad desfavorables al objeto psicológico) con **t's** significativas desde el punto de vista estadístico, a la muestra de Sujetos de la investigación propiamente dicha.

Las instrucciones que se dan a los Sujetos, así como el formato de las escalas son los mismos que los utilizados en la primera fase. Obtenidas las respuestas a las afirmaciones, se procede a dar pesos a las categorías de respuesta siguiendo el mismo criterio señalado en la primera fase.



#### 9.1.4 Puntajes de Actitud e Interpretación

Los puntajes de actitud se obtienen sumando los pesos asignados a las categorías de respuesta seleccionadas por cada sujeto en todas y cada una de las afirmaciones que conformaron la escala final.

Es pertinente recordar que este modelo no tiene continuo psicológico, como se señaló anteriormente, y que por lo tanto no se puede suponer un punto de origen 0 en el continuo, y en consecuencia, es imposible con este método, interpretar puntajes individuales de actitud.

En otras palabras, si la escala final tiene por ejemplo, 20 afirmaciones, si se optó por dar un pesaje de 1 a 5 a las categorías, y si se dió el peso más alto (5) a la opción que refleja una actitud más favorable, y el más bajo (1) a la opción que refleja la actitud más desfavorable, sólo se pueden interpretar inequívocamente dos puntajes totales posibles: 20 y 100. En el primer caso el sujeto habría respondido en todas y cada una de las 20 afirmaciones (10 favorables y 10 desfavorables) en la opción cuyo peso hubiera sido (1) (el más bajo); en el segundo caso, el Sujeto habría respondido en todas las afirmaciones en la opción cuyo peso fuera de (5) (el mas alto). Sin embargo, como bien puede esperarse, este tipo de casos son raros; lo más frecuente es obtener puntajes que se encuentren entre estas dos calificaciones extremas. De esta manera, una calificación de 60 por ejemplo, puede corresponder a un Sujeto que conteste en la opción que vale 3 en todas las afirmaciones; o en 10 afirmaciones en la opción que vale 4 y en las otras 10 en la que vale 2; en fin, el número posible de combinaciones de respuesta en las diferentes opciones que pudieran sumar un total de 60 es muy alto, y la interpretación de esta calificación se complicaría mucho.

En virtud de lo anterior se puede decir que sería posible obtener, si así se requiriera, puntajes individuales de actitud, pero su interpretación sería relativa al grupo de pertenencia los Sujetos. Es decir, se puede obtener el puntaje promedio del grupo, empleando la siguiente fórmula:

$$X \text{ grupo} = \frac{\sum X}{n}$$

dónde: X = calificación de los individuos  
n = número de individuos en el grupo

Si se tiene como punto de origen la media del grupo, se puede comparar el puntaje de actitud de cada Sujeto con el de la media

del grupo. Los Sujetos con puntajes mayores a la media del grupo tendrán actitudes más favorables hacia el objeto psicológico que la actitud media del grupo; aquellos con puntajes más bajos, tendrán una actitud más desfavorable que la media del grupo.

### 9.1.5 Transformación de los Puntajes de Actitud a Puntajes **T**

Para efectos de comparación, cuando se desea comparar actitudes con otras pruebas u otros instrumentos que miden otras variables, es necesario estandarizar los puntajes. La distribución de puntajes **T** es una manera de hacerlo. Esta distribución tiene una media de 50 y una desviación de 10. Para transformar un puntaje de actitud obtenido con el método de rangos sumariados a puntajes estandarizados **T** se emplea la siguiente fórmula:

$$T = 50 + 10 (X - M/s)$$

dónde: X = puntaje de actitud de un Sujeto

M = media aritmética de la distribución de X

s = desviación estándar de la distribución de X.

Transformando los puntajes de actitud y los obtenidos por otras pruebas en otras actitudes y/o variables diferentes, se pueden comparar las posiciones relativas de los Sujetos respecto a las medias del grupo entre diferentes variables.

Por último, es conveniente hacer notar que los coeficientes de confiabilidad (equivalencia pares-nones) reportados para este tipo de escalas se encuentran alrededor de  $r_x = 0.85$ . Otro tipo de confiabilidad (estabilidad temporal) se obtiene correlacionando los puntajes obtenidos por los mismos Sujetos con las dos formas paralelas de la escala. La confiabilidad de consistencia interna se puede obtener para este tipo de instrumentos por medio del Coeficiente Alpha de Cronbach (1951) (véase la sección correspondiente de este libro).

Por lo que toca a la validez, se mencionó al principio que el procedimiento de selección de las afirmaciones que compondrán la escala final incluye lo que podría considerarse validez concurrente por medio de grupos contrastados. En realidad, las afirmaciones que constituyen el instrumento tienen validez concurrente por grupos contrastados, pero de manera individual, ya que el análisis de reactivos para poner a prueba su poder discriminativo (las pruebas **t** efectuadas), es el que determina dicha validez. Será conveniente, establecer el poder discriminativo del instrumento en su totalidad (utilizando las calificaciones totales obtenidas por los Sujetos), ya sea por

medio de grupos contrastados (puntajes altos y bajos), o grupos criterio, como se explicará mas adelante, en la sección correspondiente a la validez de los instrumentos. Sin embargo será importante contar con un criterio externo ajeno a la escala para determinar su validez, ya sea concurrente, predictiva o de construcción.

### **9.1.6 Ventajas y Desventajas del Método de Rangos Sumarizados**

Las principales desventajas de este método son:

- a) no proporciona puntajes individuales de actitud
- b) es el más pobre desde el punto de vista teórico
- c) carece de pruebas propias de consistencia de cualquier tipo

Entre sus ventajas están:

- a) los jueces realizan una tarea común, es decir: simplemente indican su grado de acuerdo o desacuerdo.
- b) es la escala de más fácil elaboración.
- c) los procedimientos estadísticos son muy sencillos.
- d) se puede usar un gran número de afirmaciones,
- e) se pueden construir escalas paralelas o equivalentes partiendo del mismo conjunto original de afirmaciones.

## **9.2 ANÁLISIS DE ESCALOGRAMA**

Como ya se mencionó, el Análisis de Escalograma pertenece a la aproximación centrada en la respuesta y en el estímulo, donde la variabilidad de las reacciones a los estímulos se atribuye tanto a la variación en los Sujetos como a la variación en los estímulos, asignándose valores escalares a los estímulos, y a los Sujetos. En esta técnica, la tarea de los Sujetos consiste en responder a una afirmación sobre la base de la posición de ésta en relación a la posición de los Sujetos. Es decir, la actitud del Sujeto, y la actitud que refleja el reactivo, determinan la respuesta del individuo.

Entre los problemas más importantes en todo proceso de medición unidimensional se encuentran los siguientes:

- a) para lograr una medición significativa es necesario que ésta se realice a lo largo de una sola dimensión por vez.
- b) una vez delimitada esa dimensión única, es importante encontrar algunos puntos que sirvan para anclar las medidas

realizadas a lo largo de un continuo

Guttman (1944,1945,1947a,1947b) proporciona una teoría y las técnicas derivadas de ésta para dar una respuesta a los problemas señalados. La técnica del análisis de escalograma que proporciona un procedimiento simple para probar la unidimensionalidad de una serie de reactivos cualitativos y, la técnica de la función de intensidad que se utiliza para encontrar un punto significativo de separación o de corte a lo largo de un sólo continuo. Este análisis de intensidad permite determinar objetivamente cuál de los reactivos no tiene una tendencia con respecto a la actitud como un todo y lo hace dividiendo a los sujetos en dos grupos con actitudes opuestas (favorables-desfavorables) al objeto psicológico

### **9.2.1. Propositiones Fundamentales que Subyacen al Análisis Escalar**

Un conjunto de afirmaciones puede considerarse una escala si una persona con un rango u orden más alto que otra, obtiene en cada afirmación puntajes tan o más altos que la otra. Otra definición equivalente de escala referida no al conjunto de afirmaciones, sino al interior de éstas dice que si una categoría de respuesta es más alta que otra, entonces todos los sujetos que acepten dicha categoría deberán tener rangos escalares más altos que aquellos sujetos que están en una categoría más baja. Una tercera definición de escala dice que cada afirmación debe ser una función simple de los rangos del sujetos. En otras palabras, esto significa que cada una de las respuestas de un Sujeto se podrá reproducir a partir del rango obtenido por éste. Esta definición es la que se considera más conveniente para los procedimientos prácticos de determinar la monodimensionalidad de un conjunto de reactivos o afirmaciones.

En la práctica no se espera encontrar escalas perfectas, por lo que se tiende a medir la desviación que las escalas guardan respecto a lo que sería una escala perfecta por medio del coeficiente de reproducibilidad. Este se define como la frecuencia relativa empírica con la que los valores de los atributos corresponden a los intervalos de una variable cuantitativa. En este caso el atributo es la variable cualitativa y los valores de un atributo son las categorías. Se pondrá un ejemplo para aclarar estos planteamientos. Supóngase que **Y** es un atributo: "gusto por las telenovelas", con tres valores:

- Y1= Me gustan las telenovelas
- Y2= no me gustan las telenovelas
- Y3= no sé si me gustan las telenovelas

y que **X** es una variable cuantitativa dividida en tres intervalos que mantienen una correspondencia de uno a uno con los valores de Y. Lo anterior se presenta en la siguiente tabla 9.5

		Intervalo 1	Intervalo 2	intervalo 3
Intervalos de una variable cuantitativa	X	0 1 2	3 4	5 6 7 8 9
Valores del atributo	Y	Y1 Y1 Y1	Y3 Y3	Y2 Y2 Y2 Y2
		Y1	Y3	Y2

TABLA 9.5. Correspondencia entre una variable cuantitativa y los valores de un atributo.

En la práctica se ha encontrado que las escalas con un coeficiente de reproducibilidad del 90% se han empleado como aproximaciones eficientes de escalas perfectas.

En el proceso de escalamiento no existe problema métrico, lo que significa que si considerara a la variable **X** del esquema anterior como una variable escalar se podría sumar o sustraer cualquier constante a cada puntaje **X**; también se podría multiplicar por una constante, sacarle raíz cuadrada o su logaritmo. Es decir, se puede utilizar cualquier transformación, continua o discontinua, en la medida en que la correlación del orden de rangos entre la **X** original y la transformación sea perfecta. En otras palabras, cualquier transformación produciría variables escalares en donde cada una de ellas sería igualmente buena para reproducir los atributos.

En la práctica, el ordenamiento de rangos ha sido utilizado como una variable escalar (de hecho es una métrica de cuadrados mínimos para una distribución rectangular de puntajes escalares), y se ha encontrado que tal ordenamiento es suficiente para los aspectos mecánicos con los que se prueba la escalabilidad de un conjunto de reactivos, así como para efectos de predicción externa.

Los reactivos utilizados en un análisis de escalograma deben tener la propiedad acumulativa. Un ejemplo muy simple de propiedad acumulativa de un conjunto de reactivos se resume en la siguiente escala hipotética de escolaridad.

1. ¿Estudió cursos superiores a los del nivel de la preparatoria?  
SI; NO

2.¿Sus estudios son superiores a los del nivel de la secundaria?

Si; NO

3.-¿Realizó estudios superiores a los del nivel de la primaria?

SI; NO

Si una persona contesta SI al primer reactivo deberá responder también SI, al siguiente, y también deberá decir SI al último. De esta manera si se da un puntaje de 2 al Sujeto que aceptó dos reactivos, se sabe exactamente cuales de éstos fueron respondidos por él. Para los reactivos señalados existen cuatro patrones posibles de respuesta y aparecen en la tabla 9.6.

TABLA 9.6. PATRONES DE RESPUESTA A 3 REACTIVOS DE CARACTER ACUMULATIVO

ORDEN-RANGO DE LOS SUJETOS	PUNTAJES	RESPUESTA SI REACTIVOS			RESPUESTA NP REACTIVOS		
		1	2	5	1	2	3
1	3	X	X	X			
2	2		X	X	X		
3	1			X	X	X	
4	0				X	X	X

NOTA -EL DIAGRAMA (CONJUNTO DE CRUCES), ES LO QUE SE CONOCE COMO ESCALOGRAMA Y ANÁLISIS DE ESCALOGRAMA SE REFIERE AL PROCEDIMIENTO.

Para determinar si una serie de reactivos forman una escala es necesario probar el grado en el que se pueden reproducir las respuestas a reactivos particulares a partir del valor de los puntajes escalares. Para llevar a cabo esto existen diferentes procedimientos; sin embargo, las siguientes técnicas han mostrado en la práctica resultados similares ya que todas ellas se basan, en lo fundamental, en la misma teoría. Los procedimientos son a) la técnica del tablero del escalograma, diseñado por Guttman (1946) que permite el manejo de los rangos, de los Sujetos y de las categorías de los reactivos, proporcionando la posibilidad de determinar visualmente la existencia de los patrones requeridos (si los hay) en forma de paralelograma; b) la técnica de Cornell, desarrollada por Guttman (1947b); c) la técnica de tabulación o de Goodenough (1944); y el método de los cuadrados mínimos (Guttman, 1947a). En este capítulo se desarrollarán las más frecuentemente utilizadas, las técnicas de Cornell y de Goodenough.

### 9.2.2 Técnica de Cornell

La técnica de Cornell se emplea para trabajar datos dicotómicos. Una vez definido el objeto psicológico y elaborado un conjunto (de 10 a 12) afirmaciones, se inicia el proceso de análisis. Cabe señalar que esta técnica no sólo se emplea para medir actitudes, de hecho se ha utilizado más para probar la escalabilidad o monodimensionalidad de escalas ya hechas, por lo que en estos casos no se elaboran afirmaciones, más bien se seleccionan de la escala que se va a someter a prueba.

Cada afirmación tiene dos categorías de respuesta: de acuerdo y desacuerdo. Estas afirmaciones se aplican a una muestra de 100 Sujetos y se les pide que respondan si están de acuerdo o en desacuerdo con cada una de ellas. Se inicia entonces el proceso de computación.

Una vez que los Sujetos han dado sus respuestas, se asigna a cada afirmación pesos de 1 y 0 a las categorías de respuesta. Se da el peso de 1 a aquellas categorías que representen más de aquello que se está midiendo. En el caso de que se estén midiendo actitudes, el peso de 1 se da a la opción que representa más favorabilidad hacia el objeto psicológico. El peso de 0 se asigna a la opción de respuesta que represente más desfavorabilidad hacia el objetos psicológico.

En seguida se obtiene un puntaje total para cada Sujeto, sumando los pesos asignados a las categorías de respuestas seleccionadas por él. Se procede a ordenar los cuestionarios de los Sujetos, de puntajes altos a puntajes bajos y se vacía esta información en una matriz de doble entrada en la que las hileras corresponden a los Sujetos y las columnas, dobles, cada una corresponde a una afirmación con sus dos categorías de respuesta.

A continuación se ilustra la forma que tendría una escala perfecta con cinco afirmaciones evaluadas o respondidas por 20 Sujetos. Las líneas horizontales señalan el punto de corte o el cambio de respuesta favorables (1) a desfavorables (0). (Tabla 9.20).

La escala perfecta equivale a una escala ideal inexistente; permitiría reproducir sin errores las respuestas de cada una de las afirmaciones conociendo solamente los puntajes totales.

TABLA 9.7 ILUSTRACIÓN DE UNA ESCALA PERFECTA CON 5 AFIRMACIONES ACUMULATIVAS EVALUADAS POR 20 SS.

SUJ.	AFIRMACIONES										PUNT.
	1		2		3		4		5		
	1	0	1	0	1	0	1	0	1	0	
1	X		X		X		X		X		5
2	X		X		X		X		X		5
3	X		X		X		X		X		5
4	X		X		X		X		X		5
5		X	X		X		X		X		4
6		X	X		X		X		X		4
7		X	X		X		X		X		4
8		X		X	X		X		X		3
9		X		X	X		X		X		3
10		X		X	X		X		X		3
11		X		X	X		X		X		3
12		X		X		X	X		X		2
13		X		X		X	X		X		2
14		X		X		X	X		X		2
15		X		X		X		X	X		1



16	X	X	X	X	X	1
17	X	X	X	X	X	0
18	X	X	X	X	X	0
19	X	X	X	X	X	0
20	X	X	X	X	X	0

Cuando se manejan datos empíricos es necesario medir el grado de reproducibilidad que presenta el conjunto de afirmaciones de interés. De esta manera, una vez que se han vaciado las respuestas de los Sujetos en una Tabla de N afirmaciones por N Sujetos (Tabla 9.7) se procede a encontrar los puntos de corte. Un punto de corte es el lugar en el que se produce un cambio brusco en la categoría de respuesta; es decir, de favorable (1) a desfavorable (0).

Para localizar los puntos de corte se debe cumplir con dos criterios. El primero se refiere a que el lugar donde se localiza el punto de corte debe minimizar el error. Un error es una respuesta en una categoría que queda en el lado opuesto de la mayoría de las categorías por arriba o debajo del punto de corte. Si se observa la tabla 9.7 se verá como en cada punto de corte (línea horizontales) realmente se cambia de una categoría de respuesta a otra, sin que queden respuestas (cruces) fuera del lugar indicado por el corte. Por ejemplo, para la afirmación 1 el punto de corte se localiza entre los Sujetos 4 y 5 ; todas las respuestas arriba del punto de corte están en la categoría 1 y todas las respuestas por debajo del punto de corte están en la categoría 0. En este caso no hay error. Sin embargo, si se encuentra que el Sujeto 2 da una respuestas de 0 en la afirmación 1, y los Sujetos 11 y 15 dieron respuestas de 1, entonces habrá tres errores, o respuestas que no caen donde el punto de corte señala debieran caer.

El segundo criterio se refiere a que se deberá localizar el punto de corte de manera tal que en ninguna de las dos categorías (0 y 1) de cada afirmación se vaya a tener más errores que no errores. En otras palabras, no debe haber un mayor número de repuestas fuera del lugar que teóricamente deberían ocupar que el número de respuestas que ocupan el lugar teóricamente adecuado. A continuación se explica este criterio, con la ayuda de un ejemplo.

TABLA 9.8 UN EJEMPLO DEL DESARROLLO DE LA TÉCNICA DE CORNELL APLICADA A 4 AFIRMACIONES QUE FUERON EVALUADAS POR 25 Sujetos.

SUJETOS	AFIRMACIONES								PUNTAJES
	1		2		3		4		
	1	0	1	0	1	0	1	0	
1	X		X		X		X		4
2	X		X		X		X		4
3	X			X	X		X		3
4	X		X			X	X		3
5	X			X	X		X		3
6	X		X			X	X		3
7	X		X			X	X		3
8		X	X		X		X		3
9	X		X		X		X		3
10	X			X		X	X		3
11		X	X	X	X		X		2
12		X	X	X	X		X		2
13		X	X	X	X		X		2
14		X	X	X	X		X		2
15		X	X	X	X		X		2
16	X		X	X				X	2
17		X	X		X	X	X		1
18		X	X		X	X	X		1
19		X	X		X	X	X		1
20		X	X		X	X	X		1
21		X	X		X	X	X		1
22		X	X		X	X	X		1
23		X	X		X	X	X		1
24		X	X		X	X	X		1
25		X	X		X			X	0
FREC.	15		18		13		2		CR= $1 \frac{-\sum e}{N_{respues}}$
p y q ERROR	.6 1		8 .72 2		8.52 4		2 .08 1		CR=.91 $e=\sum 9$

Como se observa en la tabla 9.8, el punto de corte para la afirmación 1 aparece entre las respuestas dadas por los sujetos 10 y 11. Al pie de la Tabla (donde dice Error) puede verse que se registró el número 1 para cada categoría. Estos errores se deben a que debajo del punto de corte cayó una respuesta (la del sujeto 16) que teóricamente debía estar arriba del punto de corte en la categoría correspondiente (1). El Sujeto 8 presenta una respuesta arriba del punto de corte, cuando debiera estar por debajo del

mismo en su lado correspondiente ( 0). El lugar que se seleccionó para el punto de corte minimiza el error. Si el corte se hubiera establecido entre los Sujetos 16 y 17, se hubieran tenido 0 errores para la categoría 1, pero 6 errores en la categoría 0. Si se hubiera localizado entre los Sujetos 7 y 8, se tendrían 0 errores para la categoría 0 pero 3 para la categoría 1.

Obsérvese ahora la afirmación 2. Si el punto de corte se encontrara entre las respuestas de los sujetos 2 y 3, se tendrían 0 errores en la categoría 0, pero 6 en la categoría 1. En este caso se estaría violando los dos criterios, ya que no se minimiza el error ( 0,2 vs. 6,0) y además habría más error que no error en la categoría 1 porque 6 respuestas quedan fueran de su lugar teóricamente correcto.

El siguiente paso en el empleo de la técnica de Cornell para la determinación del coeficiente de reproducibilidad consiste en obtener las frecuencias o número de respuestas en cada categoría para cada afirmación. En seguida se obtienen las proporciones de respuesta de la categoría 0 (q) empleando las siguientes fórmulas:

$$p = F/N = \text{número de Sujetos}$$

$$; q = 1 - p$$

En el ejemplo de la tabla 9.8, para la afirmación 1 se tendrá:  
 $\Sigma F = 10 / 25 = 0.4 = p; q = 1 - 0.4 = 0.6$

El siguiente paso consiste en sumar los errores de todas las afirmaciones:  $\Sigma e = 9$ . Por último, se estima el coeficiente de reproducibilidad con la siguiente fórmula:

$$C.R. = 1 - \Sigma e / (R = \text{número de respuestas}) (N \text{ Sujetos})$$

que en este ejemplo será:  $C.R. = 1 - 0.09 = 0.91$

Este coeficiente, 0.91 significa que para las afirmaciones evaluadas es posible reproducir las respuestas dadas a ellas con una exactitud del 91% o que si se intentaran reproducir las respuestas a las afirmaciones individuales a partir del conocimiento de los puntajes o calificaciones totales, se cometería un error menor al de 10%. Sin embargo, si se revisan los datos de la tabla 9.8, es evidente la imposibilidad de reproducir la respuestas individuales de cada afirmación con un error tan pequeño. Esta sobrevaloración del coeficiente de reproducibilidad se debe al hecho de haber localizado los puntos de corte de acuerdo con el criterio de minimización del error (Edwards, 1957). Para obtener un coeficiente de reproducibilidad más exacto, se sugiere la técnica de tabulación, más conocida como análisis de escalograma de Goodenough.

**9.3 Técnica de Goodenough**

Los pasos a seguir son los mismos que en la técnica de Cornell, hasta tener una matriz de N Sujetos X K afirmaciones como la que se presenta en la tabla 9.9 y se procede a la obtención de patrones teóricos o predichos de respuestas.

TABLA 9.9 MATRIZ de NXK PARA ESTIMAR EL COEFICIENTE DE REPRODUCIBILIDAD CON LA TECNICA DE GOODENOUGH.

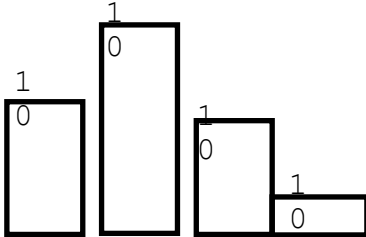
SUJETOS	AFIRMACIONES				PUNTAJES	ERRORES
	1	2	3	4		
1	1	1	1	1	4	0
2	1	1	1	1	3	0
3	1	0	1	1	3	2
4	1	1	0	1	3	0
5	1	0	1	1	3	2
6	1	1	0	1	3	2
7	1	1	0	1	3	2
8	0	1	1	1	3	0
9	1	0	1	1	3	2
10	1	1	0	1	3	0
11	0	0	1	1	2	0
12	0	0	1	1	2	0
13	0	0	1	1	2	0
14	0	0	1	1	2	0
15	0	0	1	1	2	2
16	1	0	1	0	2	0
17	0	0	0	1	1	0
18	0	0	0	1	1	0
19	0	0	0	1	1	0
20	0	0	0	1	1	0
21	0	0	0	1	1	0
22	0	0	0	1	1	0
23	0	0	0	1	1	0
24	0	0	0	1	1	0
25	0	0	0	0	0	
RECUENCIAS	10	7	12	23	52	12
PROPORCIONES	.4	.28	.48	.92		
q=1-p	.6	.72	.52	.08		C.R.=.88

A continuación se procede a obtener patrones teóricos o predichos de respuestas para cada puntaje de la siguiente manera.

En papel milimétrico se grafican barras, una para cada afirmación.

Las barras se dividen de acuerdo a las proporciones de respuesta ( $p$  y  $q$ ). En el punto  $p$  de cada barra se traza una línea continua horizontal que se continúa en forma puntuada a través de las otras barras. De esta manera se obtienen patrones teóricos o predichos para cada afirmación, como se puede observar en la tabla 9.10.

TABLA 9.10 CONSTRUCCIÓN DE BARRAS PARA OBTENER LOS PATRONES DE RESPUESTA TEÓRICOS CORRESPONDIENTES A LOS PUNTAJES DE LA TABLA 9.9.

AFIRMACIONES				PUNTAJES	PATRONES TEÓRICOS DE RESPUESTA			
1	2	3	4					
				4	1	1	1	1
				3	1	0	1	1
				2	0	0	1	1
				1	0	0	0	1
				0	0	0	0	0

Como puede verse en la Tabla 9.9, para un puntaje de 4, el patrón teórico de respuesta es : 1111. Para el puntaje 3, el patrón teórico se obtiene localizando ese puntaje siguiendo el espacio que marcan las líneas punteadas en el que éste se encuentra; se dirige la vista a la barra de la afirmación 1, en el espacio mencionado aparece el número 1 que corresponde al primero del patrón predicho. En la barra de la afirmación 2, dentro del espacio correspondiente al puntaje 3 aparece un 0, que será el segundo número de dicho patrón. En la barra 3, no obstante que está vacío el espacio del puntaje 3, el número que ocupará el tercer lugar será del patrón será el 1 ya que el estar arriba de la línea divisoria significa que toda esa parte de la barra corresponde a la proporción de respuestas con categoría de 1, así como abajo de dicha línea corresponde a la proporción de respuesta con categoría de 0. Por último, la barra de la afirmación 4 señala al número 1 que corresponderá al último número del patrón de respuestas predichas para un puntaje de 3. De esta manera, este patrón queda : 1011. Se sigue el mismo procedimiento para obtener los patrones predichos para todos y cada un de los puntajes.

Una vez que se tienen los patrones teóricos, se procede a compararlos con los patrones obtenidos (las respuestas de 1 y 0 registradas en la tabla 9.9): Cada desviación de los patrones obtenidos respecto al predicho teórico, constituye un error. Por ejemplo, para el sujeto 4 que obtuvo un puntaje de 3, se tiene que:

- (P.O.) Patrón Obtenido: 1101
- (P.T.) Patrón Teórico : 1011
- (D.P.T.) Desviación del patrón teórico: c.i.i.c, (c = correcto; i = incorrecto);
- por lo tanto: e = 2.

Esto se lleva a cabo para todos los Sujetos de acuerdo con el puntaje total que hayan obtenido. El número de errores para cada puntaje se registra en la última columna en la Tabla 9.9.

Se procede entonces a estimar el coeficiente de reproducibilidad con la fórmula ya señalada, y se observa que :  $C.R. = 1-12/100 = 0.88$ . Comparando los resultados obtenidos con ambas técnicas, es evidente que la técnica Cornell, al minimizar los errores ( $\sum e=9$ ) y la de Goodenough ( $\sum e=12$ ), que no lo hace, sobre estima el coeficiente de reproducibilidad (  $C.R$  Cornell = 0.91;  $C.R.$  Goodenough = 0.88). En el ejemplo desarrollado la diferencia entre los coeficientes de reproducibilidad es mínima, condición que puede no sostenerse en el caso de la práctica real.

### **9.3.1 Afirmaciones con Categorías de Respuestas No Dicotómicas**

Las afirmaciones pueden presentarse a los Sujetos con más de dos categorías de respuesta. Se ha observado que entre mayor sea el número de categorías de respuesta, mayor es la discrepancia que se encuentra entre los patrones teóricos o predichos y los empíricos u observados. Esto implica que la magnitud del coeficiente de reproducibilidad disminuye.

#### **9.3.1.1 Coeficiente Mínimo de Reproducibilidad Marginal**

Aunque un coeficiente de reproducibilidad de 0.90 constituye una condición necesaria para la prueba de escalabilidad (monodimensionalidad) de un conjunto de afirmaciones, sin embargo, no constituye una condición suficiente. Esto se debe a que el valor del coeficiente de reproducibilidad se ve influido por un alto número de frecuencias en uno o varias de las categorías. En otras palabras, la reproducibilidad de una afirmación dada nunca podrá ser menor que la de la frecuencia presente en una categoría modal ( la categoría con mayor número de frecuencias). En virtud de esto, se hace necesario estimar el coeficiente mínimo de reproducibilidad marginal (C.M.R.M.), ya que el valor de este depende directamente de las frecuencias de las categorías modales. De esta manera, entre más pequeño sea el C.M.R.M. y por lo tanto mayor sea la diferencia con el coeficiente de reproducibilidad, mayor seguridad se tiene en la prueba de escalabilidad. La importancia de C.M.R.M. se hará más evidente en los siguientes párrafos, donde se verán los problemas a resolver con el escalamiento de afirmaciones que tienen más de dos categorías de respuesta.

Para el ejemplo desarrollado, (datos de la tabla 9.9), el C.M.R.M. se obtiene simplemente localizando las proporciones de respuesta

en las categorías modales de cada afirmación: 0.60, 0.72, 0.52, 0.92. Se suman estos valores y se dividen entre el número de valores:  $2.76/4 = 0.69$ . Por lo tanto, el C.M.R.M. = 0.69.

Un procedimiento alternativo requeriría formar una patrón general de respuesta, y sin tomar en cuenta los puntajes totales, se compararían los patrones observados de cada Sujeto contra este patrón general, para obtener el número de errores. Si ahora se substituyeran los valores en la fórmula:  $C.M.R.M. = 1 - \sum e/N$ ; se obtendría el mismo valor de 0.69.

El patrón general de respuesta en el ejemplo en cuestión se obtiene de la siguiente manera:

		A F I R M A C I O N E S			
		1	2	3	4
CATEGORIAS	1=p				0.92
MODALES	0=q	0.60	0.72	0.52	
		Patrón General de Respuesta			
		0	0	0	1

Este patrón es el que se compara con cada patrón observado, ignorando el puntaje total. Se obtiene el número de errores, que en este ejemplo es de 30 y se tiene el  $C.M.R.M. = 1 - 30/100 = 1 - 0.3 = 0.7$ .

Para este ejemplo, el coeficiente de reproducibilidad fue de 0.88 y el C.M.R.M. fue de 0.69 ó 0.7. En este caso, se puede decir en principio, que un coeficiente de reproducibilidad de 0.88, aunque es ligeramente menor que el valor considerado como aceptable (0.90), es más seguro y aceptable dado que el C.M.R.M. fue de 0.69, que al no ser tan alto mantiene una diferencia relativamente grande respecto al coeficiente de reproducibilidad; la diferencia es de 0.18.

Como ya se mencionó anteriormente, entre mayor sea el número de categorías de respuesta, menor será el valor del coeficiente de reproducibilidad y por lo tanto las afirmaciones no formarán una escala. En estos casos, Guttman.(1945) sugiere que se proceda a hacer una combinación sucesiva de las categorías de respuesta, hasta alcanzar un coeficiente de reproducibilidad aceptable ( $\geq 0.90$ ), sin importar que para lograrlo se llegue a la dicotomización de las categorías en cada afirmación.

Por ejemplo, si con 6 categorías de respuesta se obtiene un coeficiente de reproducibilidad bajo, habrá que combinarlas:



Pesos iniciales de las categorías de respuesta	0	1	2	3	4	5
Nuevos pesos de la combinación de las categorías	0	1	2	3		

Se obtiene entonces una nueva matriz de puntajes con sus correspondientes patrones de respuesta. Se calcula el coeficiente de reproducibilidad. Si aún así, este coeficiente no fuera satisfactorio, se procedería a una nueva combinación:

Pesos de la primera combinación de categorías	0	1	2	3		
Pesos de la segunda combinación de categorías	0	1	2			

Otra alternativa es recombinar las categorías iniciales y proceder a darles nuevos pesos:

Pesos iniciales de las categorías de respuesta	0	1	2	3	4	5
Nueva recombinación y pesaje de las categorías de respuesta	0	1	2			

Si es necesario, se procede a combinar sucesivamente las categorías hasta llegar a respuestas con pesos de 1 y 0. Un criterio que puede seguirse para la combinación de categorías está dado por el grado con el que se superponen los pesos que se registran en las columnas de la matriz original.

Al trabajar con más de dos categorías de respuesta, el peso más alto se da a la respuesta más favorable. Así, para afirmaciones favorables al objeto psicológico, con tres categorías de respuesta: De Acuerdo, Indeciso, Desacuerdo, éstas categorías recibirán los pesos de 2,1 y 0, respectivamente. Los pesos de las categorías se invertirán para el caso de afirmaciones que sean desfavorables al objeto psicológico.

Las matrices originales tendrán estas columnas por afirmación como el número de categorías que tengan éstas. Por ejemplo, para tres afirmaciones con tres categorías de respuesta, aplicadas a 20 Sujetos, se tendrá una matriz de  $N(n \times c)=20(3 \times 3)=20(9)= 20$  hileras y 9 columnas. A continuación se presenta la tabla 9.11 de una matriz con estas características:

TABLA 9.11 ESQUEMA DE UNA MATRIZ PARA 3 AFIRMACIONES CON 3 CATEGORIAS DE RESPUESTA PARA N SUJETOS.

SUJETOS	A F I R M A C I O N E S									UNT.	e
	1			2			3				
	2	1	0	2	1	0	2	1	0		
1	X			X			X			6	
2	X				X		X			5	
3		X		X				X		4	
.			X		X		X			3	
.			X		X			X			
.			X			X		X			
.											
.											
N											
F											
P	.40	.30	.30	.50	.25	.25	.80	.20	.00		

Como se puede observar, a partir de este esquema (tabla 9.11), el proceso de cálculo es el mismo que se siguió en la Tabla 9.9 con sólo dos categorías. Para obtener el número de errores que aparecerían en la última columna de la Tabla se precederá a obtener los patrones teóricos de respuesta, tal y como aparece en la siguiente tabla (9.12).

TABLA 9.12 PATRONES TEORICOS DE RESPUESTA.

AFIRMACIONES			PUNTAJES	PATRONES TEÓRICOS DE RESPUESTA		
1	2	3				
			6	2	2	2
2						
	2					
1			5	1	2	2
0	1		4	1	1	2
	0		3	0	1	2
		2	2	0	0	2
		1	1	0	0	1

### 9.3.1.2 Patrones de respuesta: Escalares y No Escalares

Hasta ahora se ha visto que para probar la escalabilidad de un conjunto de afirmaciones se debe considerar el tamaño y relación de dos coeficientes diferentes: el coeficiente de reproducibilidad y el coeficiente mínimo de reproducibilidad marginal. Aunado a estos dos criterios, es necesario buscar una evidencia más. Esta se refiere a la presencia o ausencia de una segunda variable. Se supone que la escalabilidad o monodimensionalidad está dada por la presencia de una sola variable. Sin embargo, aún con un coeficiente de reproducibilidad de 0.90, existe un 10% de error y es necesario determinar si este porcentaje de error se debe a errores al azar o a la presencia de una segunda variable, distinta a la originalmente definida.

Debe hacerse notar que este 10% no es importante si se debe a errores aleatorios, ya que se ha aceptado desde un principio para el caso del análisis de escalograma que este porcentaje es bastante aceptable. Se sabe además que este tipo de errores tienden a cancelarse entre sí a la larga; y que la presencia de estos es inevitable en cualquier estudio de naturaleza empírica.

Por otro lado, si será importante si el error se debe a la presencia de una segunda variable no prevista, pues se viola el fundamento principal de lo que Guttman (1945) llama escalabilidad. Por esta razón se debe proceder a examinar los patrones de respuesta de los Sujetos, para encontrar la posible existencia de tipos no escalares.

Para determinar el número posible de patrones (escalares y no escalares) de respuesta se procede de la siguiente manera. El número total de posibles patrones de respuesta depende del número de afirmaciones y sus correspondientes categorías de respuesta. Con cuatro afirmaciones, cada una con dos categorías de respuesta, se tienen  $2^4$  posibles patrones de respuesta, es decir, 16, mismos que generan un rango de puntajes de 0 a 4. En otras palabras, el número posible de patrones de respuesta es igual al producto de las varias categorías de respuesta de las afirmaciones ( i.e.  $2 \times 2 \times 2 \times 2$  ).

De acuerdo a la definición de escala que hace Guttman (1945, 1947a), cada puntaje deberá de tener uno y solamente un patrón de respuesta. El número posible de patrones escalares se obtiene sumando el número de categorías de respuesta de cada afirmación, y sustrayendo el número de afirmaciones, a lo que se le suma la unidad. De esta manera, con cuatro afirmaciones cada una con dos

categorías de respuesta el número posible de patrones escalares es:  $2+2+2+2-4+1=5$ .

En la Tabla 9.10 pueden verse los cinco patrones escalares de respuesta. Para tres afirmaciones, con tres categorías de respuesta, el número posible de patrones escalares es  $3+3+3-3+1=7$ .

Debe tomarse en cuenta que no necesariamente aparecen todos los posibles patrones escalares y no escalares de respuesta en una muestra de Sujetos estudiados. Así, por ejemplo, en el segundo ejemplo mencionado arriba, el número de patrones escalares fue de 7, pero como se puede observar en la Tabla 9.12, sólo aparecen seis, ya que está ausente el patrón correspondiente al puntaje de 0, en la medida en que los Sujetos no dieron respuesta para esa categoría en la afirmación 3.

Los patrones no escalares son todos aquellos que difieren del patrón escalar correspondiente para un puntaje dado; se pueden ver patrones no escalares en la sección en donde se comparan los patrones observados y los predichos para determinar el número de errores. Después de examinar los patrones de respuesta para cada puntaje, si se encontrara un número substancial (no se sabe qué frecuencia constituye un número "substancial") de patrones no escalares, deberá pensarse en la existencia de otra variable que no corresponda a la originalmente definida. En tales circunstancias se pueden construir escalas para medir por separado las variables presentes y posteriormente determinar la correlaciones entre los puntajes de cada escala.

Por último, se debe señalar la existencia de cuasi-escalas. De acuerdo a Guttman (1945 y 1947a) éstas son aquellas cuyos coeficientes de reproducibilidad no son más altos, o lo son ligeramente, que su coeficiente mínimo de reproducibilidad marginal. Por lo general, el coeficiente de reproducibilidad de las cuasi-escalas es menor a 0.85, pero los patrones de respuesta no-escalares no muestran frecuencias substanciales. Es decir, aunque se tenga el 15% de error, el orden de los Sujetos está dado por la variable principal, no produciéndose de esta manera muchos patrones no escalares; si por el contrario, existen muchos patrones de respuesta no-escalares, la existencia de otra variable es la que está determinando el ordenamiento de los sujetos y por lo tanto, la escala original ya no constituye un cuasi-escala.

### **9.3.1.3 Ventajas y Desventajas**

Las desventajas principales son las siguientes: a) el método en sí no es para construir escalas de actitud; b) el rango posible de

calificaciones que pueden tener los Sujetos va de 0 al número total de afirmaciones, por lo que su poder discriminativo puede ser pobre, al requerirse de un conjunto acumulativo de reactivos, lo que posiblemente reduce el número posible de ellos; c) la calificación de actitud no indica favorabilidad o desfavorabilidad, ni graduaciones dentro de esas categorías, d) carece de un continuo psicológico e) como método de determinación de escalabilidad de un conjunto de reactivos ( por lo general pocos) requiere de por lo menos 100 jueces.

Entre las ventajas están: a) la tarea que se le pide a los jueces es realmente sencilla; b) asegura que la escala que se construye es unidimensional, c) permite cumplir con los requisitos de las escalas monodimensionales.

#### **REFERENCIAS**

- Cronbach, L.J. (1951). Coefficient Alpha and the internal structure of tests. **Psychometrika**, 16, 297-334.
- Goodenough, W.H. (1944). A technique for scale analysis. **Educational Psychological Measurement**, 4, 179-190.
- Guttman, L. (1944). A basis for scaling qualitative data. **American Sociological Review**, 9, 139-150.
- Guttman, L. (1945). **Questions and Answers about Scale Analysis**. Research Branch Information and Education Division. Army Service Forces. Report D-2.
- Guttman, L. (1946); An Approach for quantifying paired comparisons and rank order. **American Mathematical Statistics**, 17, 144 - 163.
- Guttman, L. (1947a). Suggestions for further research in scale and intensity analysis of attitudes and opinions. **International Journal of Opinion and Attitude Research**, 1, 30 - 35..
- Guttman, L. (1947b). The Cornell technique for scale and intensity analysis. **Educational Psychological Measurement**, 7, 247-280.

## X ESCALAS DE ACTITUD : DIFERENCIAL SEMÁNTICO

El diferencial semántico se desarrolló como una técnica de medición del significado connotativo de diversos estímulos (colores, objetos, dibujos, etc.), pero principalmente de estímulos verbales.

Los supuestos fundamentales de la teoría del diferencial semántico desarrollada por Osgood, Suci y Tannenbaum (1957) son los siguientes: el resultado de una evaluación o juicio puede concebirse como el lugar que el estímulo evaluado o enjuiciado ocupa en un continuo experiencial definido por dos términos (adjetivos) bipolares. Muchos de los continuos experienciales son esencialmente equivalentes y por lo tanto se pueden representar en forma unidimensional. Un número limitado de estos continuos puede utilizarse para definir el espacio semántico dentro del cual se puede especificar a cualquier estímulo.

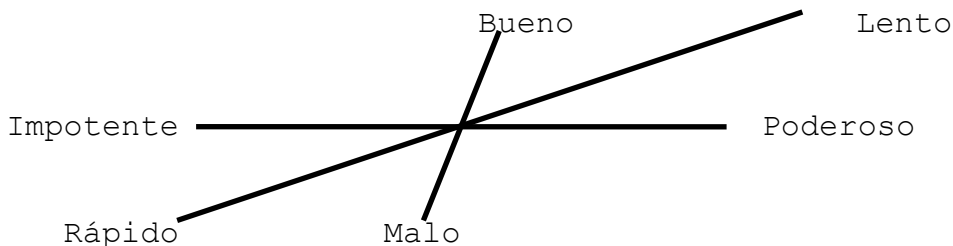
El espacio semántico ha sido estudiado por medio de diversos tipos de análisis factorial, y una y otra vez se ha encontrado la existencia de tres factores, que debido a su ortogonalidad pueden ser identificados como dimensiones independientes del significado. Los tres factores obtenidos son el evaluativo (E), el de potencia (P) y el de actividad (A). A éstos se les ha denominado estructura tridimensional EPA del significado connotativo. La importancia de cada factor en cuanto a la cantidad de varianza que explican corresponde al orden en que se presenta: EPA. Numerosos estudios han mostrado que la estructura EPA es válida para una gran diversidad de Sujetos, conceptos y escalas.

En 1967, May (citado por Díaz-Guerrero y Salas, 1975) propuso nueve escalas que correspondieron a las tres dimensiones (EPA) y se obtuvieron los resultados que se presentan en la siguiente Tabla 10.1

TABLA 10.1. "NUEVE ESCALAS PROPUESTAS POR MAY (1967) PESOS EN VALORES ABSOLUTOS.

DIMENSIONES	ADJETIVO-BIPOLARES	PESOS FACTORIALES EN VALORES ABSOLUTOS		
		E	P	A
EVALUACIÓN	SIMPATICO-ANTIPATICO	0.92	0.01	0.08
	ADMIRABLE-DESPRESIABLE	0.91	0.08	0.05
	BUENO-MALO	0.92	0.10	0.10
POTENCIA	FUERTE-DEBIL	0.06	0.54	0.07
	LARGO-CORTO	0.05	0.45	0.19
	MAYOR-MENOR	0.24	0.54	0.03
ACTIVIDAD	ACTIVO-PASIVO	0.01	0.21	0.57
	RAPIDO-LENTO	0.20	0.08	0.45
	JOVEN- VIEJO	0.41	0.22	0.46

Cuadro tomado de Díaz-Guerrero y Salas (1975).  
 Se puede pensar que las dimensiones EPA forman un espacio tridimensional que se puede representar como aparece en la Figura 10.1.



**Figura 10.1. Espacio Tridimensional EPA**

Cuando se han realizado estudios en lengua española con el diferencial semántico (ver por ejemplo, Díaz Guerrero y Salas, 1975,), las escalas que se han empleado típicamente son las que aparecen en la Tabla 10.2, donde también se presentan los pesos factoriales de los adjetivos bipolares que conforman las dimensiones EPA.

TABLA 10.2 ESCALAS TÍPICAS DEL DIFERENCIAL SEMÁNTICO EN ESPAÑOL (DÍAZ-GUERRERO Y SALAS, 1975).

DIMENSIONES	ADJETIVO-BIPOLARES	PESOS FACTORIALES
EVALUACION	BUENO-MALO	0.96
	ADMIRABLE-DESPRECIABLE	0.95
	SIMPATICO-ANTIPATICO	0.95
	AGRADABLE-DESAGRADABLE	0.96
POTENCIA	GRANDE-CHICO	0.78
	GIGANTE ENANO	0.80
	FUERTE-DEIL	0.66
	AMYOR-MENOR	0.78
ACTIVIDAD	ACTIVO-PASIVO	0.72
	BLANDO-DURO	0.45
	JOVEN-VIEJO	0.46
	RAPIDO-LENTO	0.55

El origen o centro del espacio del diferencial semántico o del sentido afectivo representa la neutralidad en las tres dimensiones. Al considerar las medidas EPA de un estímulo como coordenadas, permite que se sitúe al estímulo como un punto en el espacio. Dicho punto representa gráficamente la respuesta afectiva. Se puede medir la afectividad total de un estímulo, y comparar su semejanza afectiva con otros estímulos.

## 10.1 El Diferencial Semántico como escala de Actitud

Cuando se desea emplear el diferencial semántico como una escala de actitud, la pregunta principal que se debe hacer el investigador se refiere a cuáles escalas deben utilizarse. Las escalas habrán de seleccionarse tomando en cuenta dos consideraciones: a) lo adecuado o relevante que sean al estímulo; y b) su composición factorial, procurando emplear escalas con pesos factoriales elevados en la estructura EPA:

Existen dos procedimientos para identificar las escalas que se adecuan al concepto o estímulo, y a una muestra determinada de personas. El primero consiste en presentar a un conjunto de Sujetos un grupo de escalas pidiéndoles que las ordenen de acuerdo con la pertinencia que tengan al pensar en cierto estímulo; se emplean las escalas ordenadas en los primeros lugares. El segundo consiste en presentar a los Sujetos pares o triadas de estímulos, y preguntarles cómo difieren. Se forman escalas bipolares a partir de las distinciones hechas omitiendo aquellas puramente denotativas.

Algunos autores consideran que las actitudes se miden exclusivamente con escalas del Factor E; otros consideran más apropiado utilizar escalas de las tres dimensiones. Cualquiera que sea la alternativa que se seleccione, se requiere que las escalas tengan pesos factoriales elevados ( $> 0.40$ ) en sus respectivas dimensiones. Si se desea tener medidas relativamente puras de esas dimensiones, es decir, que cada escala tenga un valor alto solamente en una dimensión, la manera objetiva de hacerlo es con base en análisis factoriales, procedimiento que resulta un tanto costoso. Un procedimiento alternativo es seleccionarlas con base en resultados de análisis factoriales publicados que indiquen la composición factorial de las escalas del diferencial semántico. Si se opta por esta última alternativa, se deberá tener en cuenta el problema, siempre presente, de la estabilidad semántica. Este problema, y el de la adecuación son las razones principales que llevan a la realización de análisis especiales, sobre todo en regiones de nuevo contenido.

El número de escalas que habrá de utilizarse en cada dimensión será otra decisión que habrá de tomar el investigador. Se recomienda, en términos generales, emplear cuatro o cinco escalas en cada dimensión, procurando que la carga factorial de ellas sea alta en una dimensión ( a la que pertenece) y baja en las otras dos. De esta manera se probabiliza una mejor confiabilidad de las puntuaciones factoriales por un lado, y por el otro, una sensibilidad adecuada del instrumento, para la mayoría de los propósitos. Es conveniente, sin embargo, tomar en cuenta que las



escalas evaluativas son más confiables que las de potencia y las de actividad, por lo que se recomienda emplear un número ligeramente mayor de escalas para las dimensiones P y A.

A continuación se ejemplifican tres posibles formatos de presentación gráfica de escala y conceptos (estímulos). En el primer caso se presenta un concepto (estímulos). Se presenta un concepto, seguido cada uno de todas las escalas en las que habrá de evaluarse.

	MUJER							
Buena	-----	-----	-----	-----	-----	-----	-----	Mala
Lenta	-----	-----	-----	-----	-----	-----	-----	Rápida
Admirable	-----	-----	-----	-----	-----	-----	-----	Despreciable
Fuerte	-----	-----	-----	-----	-----	-----	-----	Débil
Agradable	-----	-----	-----	-----	-----	-----	-----	Desagradable

En el segundo caso, se presenta un concepto y una de las escalas; las diferentes combinaciones de conceptos y una de las escalas; se colocan al azar uno después de otra.

	MUJER							
Buena	___	;	___	;	___	;	___	Mala
	DINERO							
Fuerte	___	;	___	;	___	;	___	Débil
	HOGAR							
Agradable	___	;	___	;	___	;	___	Desagradable
	MUJER							
Fuerte	___	;	___	;	___	;	___	Débil
.								.
.								.
.								.
.								.
	DINERO							
Bueno	___	;	___	;	___	;	___	Malo

En el tercer caso, se presenta una sola escala junto con todos los conceptos.

	BUENO-MALO								
Mujer	___	/	___	/	___	/	___	/	___
Dinero	___	/	___	/	___	/	___	/	___
hogar	___	/	___	/	___	/	___	/	___
	FUERTE-DEBIL								
Mujer	___	/	___	/	___	/	___	/	___
Dinero	___	/	___	/	___	/	___	/	___
Hogar	___	/	___	/	___	/	___	/	___
	ADMIRABLE-DESPRECIABLE								
Mujer	___	/	___	/	___	/	___	/	___
Dinero	___	/	___	/	___	/	___	/	___
Hogar	___	/	___	/	___	/	___	/	___

Los estudios reportados señalan que las medidas difieren de un formato a otro. Se recomienda el primer formato presentado en el que no cuenta el orden de los conceptos y se facilita el proceso de computación. El menos conveniente es el tercero, ya que se presenta una ligera tendencia a que las evaluaciones de un concepto afecten las evaluaciones de otros.

En términos generales, se aconseja alternar las escalas de las dimensiones EPA, y alternar también los polos de la dirección de las escalas. Es decir, presentar una escala de E, otra de P y otra de A, etc.; la primera con el polo positivo a la derecha, la segunda con el polo positivo a la izquierda, la tercera a la derecha, y así sucesivamente, procurando que todas las dimensiones (EPA) tengan escalas con los adjetivos positivos, la mitad a la derecha y la mitad a la izquierda. Con esto se reduce al mínimo las predisposiciones a la respuesta, obligando además que el Sujeto, ponga atención en cada una de las escalas.

Las posiciones escalares del diferencial semántico pueden ir acompañadas de cuantificadores adverbiales. Los adverbios que se emplean con más frecuencia son "Extremadamente", "Bastante", y "Ligeramente". Otros que también se utilizan son: "Algunas Veces", "Frecuentemente", "Mucho"; "Algo", "Ligeramente". Se pueden emplear en todas las escalas o ser específicos de algunas de ellos. Se deberá tomar en cuenta, por supuesto, el calificativo al que acompaña.

La aplicación del diferencial semántico puede ser en grupo o individualmente. La primera forma es la manera más eficiente de obtener los datos; la segundo también produce resultados satisfactorios.

Las instrucciones que se dan a los Sujetos permite encubrir los verdaderos propósitos del estudio con mucha facilidad, ya que se les pide a los Sujetos que trabajen con rapidez y que no regresen a consultar respuestas anteriores. Es recomendable presentar un ejemplo desarrollado. Las instrucciones, además de señalar lo anteriormente expuesto, le indican al Sujeto que escriba una cruz (X) en el intervalo que para él más se asocie, o más le haga pensar en el estímulo evaluado, en cada una de las escalas bipolares. Si está evaluando el concepto o estímulo MUJER, por ejemplo, se le señala que si BUENA está muy asociado con MUJER, ponga una X en el intervalo más cercano a ese adjetivo; si la percibe como DÉBIL, pero no totalmente, su X deberá estar en el intervalo anterior al próximo a DÉBIL; si no la percibe ni LENTA ni RÁPIDA, deberá marcar el intervalo de en medio, etc.

A continuación se procede a calificar las respuestas dadas por los Sujetos para obtener puntuaciones de factor. Para esto, en primer lugar se le asignan valores numéricos a las posiciones escalares de manera que se otorgue el peso de +3 al extremo favorable del continuo bipolar; el de +2 al siguiente intervalo; el de +1 al que sigue; 0 al intervalo de en medio; -1 al siguiente; -2 al que sigue; y -3 al extremo desfavorable del continuo bipolar. EN segundo lugar, se califican todas las escalas que corresponden a un solo factor (obteniéndose tres puntuaciones, una para cada uno de los factores EPA). Estas puntuaciones son en realidad el promedio de las calificaciones de las escalas en cada factor. Este resultado, el promedio de cada factor, representa la reacción de un Sujeto a un concepto o estímulo en una de las dimensiones. Es pertinente aclarar que estas medias de estimaciones como puntuaciones de factor son válidas sólo si se cumple con los siguientes requisitos: a) las cargas factoriales de las escalas (en una dimensión dada) son todas grandes y similares en tamaño; b) si todas las escalas tienen cargas principalmente en una dimensión; c) si todas las escalas son igualmente adecuadas, lo que permite asegurar la homogeneidad de varianza de las estimaciones. De esta manera se obtiene la puntuación factorial de un concepto. En otras palabras, la descripción completa de una reacción afectiva en función de las dimensiones EPA.

Si tales supuestos son notoriamente incompatibles, convendrá seguir el procedimiento de valor diferencial. Este consiste en valorar cada escala con el valor factorial elevado al cuadrado o utilizar formas de regresión múltiple.

Si el objetivo es determinar las estimaciones de las puntuaciones factoriales verdaderas de un concepto, un grupo o de una cultura, se procede a encontrar las medias de grupo de las puntuaciones

factoriales de los diferentes conceptos. Es decir, se promedian las puntuaciones factoriales de los Sujetos investigados.

Para algunos propósitos no se necesita una información detallada (puntuación factorial) sino que basta una media general de la intensidad de la respuesta afectiva independiente de su carácter. En otras palabras, cuando se quiere conocer la emocionalidad de un concepto se tendrá que calcular la medida de polarización. La polarización es igual a la distancia entre el punto neutral del espacio del diferencial semántico y el concepto particular considerado. Para esto se emplea la siguiente fórmula:

$$P = e^2 + p^2 + a^2$$

dónde:

P=Polarización

e, p y a = medidas de puntuación de factor de un concepto en las tres dimensiones.

Cuando el objetivo es comparar las medidas de un concepto A en contraste con las de un concepto B (que pueden corresponder a grupos diferentes, estimaciones de antes y después, etc.) se analizarán los perfiles de respuesta en cada dimensión por separado.

En otros casos se puede requerir una medida de las diferencias combinadas en las tres dimensiones EPA, y para esto se utilizan las puntuaciones D. Las puntuaciones D representan la distancia entre dos conjuntos de medidas de los diferenciales semánticos, cuando ambos se representan como puntos en el espacio tridimensional. Se emplea la siguiente fórmula:

$$D = \sqrt{(e_1 - e_2)^2 + (p_1 - p_2)^2 + (a_1 - a_2)^2}$$

dónde:

D=distancia entre dos conjuntos de medidas

$e_1, p_1$ , y  $a_1$  = puntuación factorial del concepto 1, o del grupo, etc.

$e_2, p_2$ , y  $a_2$  medidas correspondientes al concepto 2, grupo2, etc.

Cabe señalar que las puntuaciones D deben ser utilizadas con cautela ya que ocultan el carácter de la diferencia. Es decir, una D grande puede ser resultado de una diferencia grande en una dimensión, o de pequeñas diferencias en las tres dimensiones.

Por otro lado, cuando existe un tipo particular de reacción afectiva que es de interés especial, se puede emplear una regresión múltiple:

$$y = a + b_1x_1 + b_2x_2 + b_3x_3$$

Respecto a la confiabilidad del diferencial semántico, se han encontrado coeficientes aceptables de test-retest en cuanto a: a) estimaciones de conceptos (desviación = 0.5 de unidad escalar); b) puntuación factorial ( $r = 0.87$  y  $0.97$ ) y c) medias de grupo ( $r = 0.98$ ), que son las más confiables y estabilizadas, aún con muestras tan pequeñas como  $N = 30$ .

En cuanto a su validez, diversos estudios demuestran que el diferencial semántico puede usarse para medir actitudes. La validez queda corroborada por el hecho de que produce los resultados predichos cuando se emplea con ese propósito, y por otros estudios donde se comparan medias del diferencial semántico con medidas de actitud que emplean escalas tradicionales. Por ejemplo, se han reportado  $r = 0.74$  y  $0.82$  al correlacionar escalas tipo Thurstone y diferencial semántico; y  $r = 0.78$  al correlacionar escalas tipo Guttman y diferencial semántico.

## 10.2 Ventajas y Desventajas

Entre las principales ventajas se tiene: a) su economía, ya que se emplean las mismas escalas para medir actitud hacia cualquier objeto psicológico; b) su rapidez, ya que se elaboran fácilmente y pueden estudiarse eventos sociales que se producen inesperadamente; c) su comparabilidad de conceptos ya que al medirse actitudes hacia diferentes objetos psicológicos con las mismas escalas, es posible comparar las diferentes actitudes; d) su métrica estándar: las actitudes se miden sobre la misma métrica usando las tres dimensiones EPA, por lo tanto todos los objetos pueden colocarse en un sólo espacio actitudinal y esto permite, análisis, comparaciones y obtención de conocimientos que no sería posible con las escalas tradicionales; e) pueden emplearse como un método indirecto o disfrazado de medición de actitudes.

Entre las limitaciones se tiene: a) las aplicaciones del diferencial semántico para medir actitudes se han basado principalmente en la dimensión E, Esta lamentable tradición ha limitado el conocimiento en el campo de las actitudes. Se ha encontrado que las escalas tradicionales contienen en sus medidas las dimensiones de potencia y actividad. Se ha observado también que las correlaciones múltiples de las estimaciones EPA con las escalas tradicionales son a menudo mucho mayores que las correlaciones de las estimaciones E con dichas escalas. Se

sugiere por lo tanto, obtener estimaciones en las tres dimensiones para medir actitudes, ya que se esperaba que la información completa EPA incrementara el poder de análisis.

#### **REFERENCIAS**

Díaz-Guerrero, R. y Salas, M:(1975): **El Diferencial Semántico del Idioma Español**. México; Editorial Trillas.

Osgood, Ch. E., Suci, G.J., Tannenbaum. P.H., **The Measurement of Meaning**. Urbana: University of Illinois Press.

## XI CARACTERÍSTICAS PSICOMETRICAS DE LOS INSTRUMENTOS : CONFIABILIDAD

El investigador en ciencias sociales se enfrenta por lo general a problemas *sui generis* que conllevan, a la necesidad de construir instrumentos (cuestionarios, escalas, cédulas, etc.) especiales, ad-hoc para el problema de investigación. Por otro lado, éste, el problema de interés puede involucrar una gran cantidad de variables, por lo que es difícil que la decisión de llevar a cabo todos los pasos necesario para la construcción de un instrumento que se aboque a la medición de una sola variable (véase, por ejemplo, Nunnally, y Bernstein,1993).la alternativa más frecuente por la que opta el investigador es elaborar un instrumento por medio del cual pueda indicar o medir de la mejor manera posible y con un mínimo de tiempo, la mayor cantidad de información relevante de esas variables. En esta tarea uno de los problemas que se debe resolver es el de las características psicométricas del instrumento y que se refieren a dos aspectos fundamentales: la confiabilidad y la validez de los mismos y que frecuentemente son dejados de lado. En este capítulo se plantean los conceptos teóricos y los mecanismos prácticos para determinar la confiabilidad y la validez de los instrumentos, teniendo en cuenta los objetivos del investigador.

### 11.1. Confiabilidad

La confiabilidad de una prueba o instrumento se refiere a la consistencia de las calificaciones obtenidas por los mismos individuos en diferentes ocasiones o con diferentes conjuntos de reactivos equivalentes.

Es un hecho que en cualquier conjunto de medidas se encuentra que éstas varían, es decir que se producen errores de medición. La confiabilidad y la validez se ocupan, aún cuando de manera relativamente distinta, del problema del error de medición.

Como es sabido, existen dos posibles clases de error, el error constante, sistemático y el error casual, variable, debido al azar. Estos errores forman parte de cualquier medida obtenida (X):

$$x=v+e$$

Donde: v es igual a la medida verdadera más cualquier error sistemático y "e" viene a ser el error residual, casual o variable. Como puede verse el enunciado X expresa el supuesto básico psicológico referido a una sola medida y dado que la aplicación de instrumento se traduce en un conjunto de medidas, resulta más conveniente traducir el enunciado X a su forma más general:

$$\alpha^2 = \sigma^2 + \epsilon^2$$

Donde:  $\alpha^2$  representa la varianza total;  $\sigma^2$  se conoce como la varianza verdadera o sistemática porque engloba las medidas verdaderas más los errores constantes y  $\epsilon^2$  se refiere a la varianza debida a los errores variables.

En el sentido más amplio, la confiabilidad indica el grado con el que las diferencias individuales en las calificaciones de las pruebas se atribuyen a errores aleatorios de la medición, y el grado con el que se atribuyen a diferencias reales de las características o dominio en consideración. En términos técnicos, la confiabilidad señala qué proporción de la varianza total de las calificaciones de una prueba es "varianza de error". Al respecto, es importante entender a que se refiere esa varianza de error, ya que los actores que pueden ser considerados como varianza de error para un propósito, pueden clasificarse bajo el rubro de "varianza verdadera" para otro. Por ejemplo, si se está interesado en medir fluctuaciones de estado de ánimo, entonces los cambios cotidianos en las calificaciones de una prueba que midiera alegría, depresión, serían relevantes para el propósito de la prueba y serían por lo tanto parte de la varianza verdadera de las calificaciones. Si, por otro lado, la prueba está diseñada para medir características de personalidad más duraderas o permanentes, las mismas fluctuaciones diarias caerían bajo el rubro de varianza de error.

Sin embargo, en esencia, puede decirse que cualquier condición que es irrelevante al propósito de la prueba representa varianza de error. Vista así, puede inferirse que cuando el examinador trata de mantener las condiciones de prueba uniformes, controlando el ambiente de la situación de prueba, ( las instrucciones, los límites de tiempo, el "rapport" y otros factores similares), su propósito es reducir a la varianza de error y con esto incrementar la confiabilidad de la prueba. Sin embargo, pese a las óptimas condiciones de prueba logradas, difícilmente puede obtenerse un instrumento perfectamente confiable. De aquí la importancia de que cada prueba establezca claramente su índice de confiabilidad. Tal medida de confiabilidad tiene sentido cuando la prueba es administrada bajo condiciones estándar y aplicada a Sujetos semejantes a aquellos que constituyeron la muestra normativa. De aquí que las características de tales muestras deberán estar perfectamente especificadas, junto con el tipo de confiabilidad que se utilizó.

Puede haber por supuesto, tantas variedades o tipos de confiabilidad como haya condiciones que afecten las calificaciones de las pruebas ya que cualquiera de tales condiciones pueden ser irrelevantes para ciertos propósitos y muy relevantes para otros. No obstante la gran diversidad de tipos de confiabilidad en la



práctica se utilizan unos cuantos. Los principales son los siguientes:

#### **11.1.1. Estabilidad Temporal**

Una fuente de varianza de error obvia para la mayoría de los propósitos de la aplicación de pruebas, es la de las fluctuaciones azarosas de la ejecución, que ocurren de una sesión de prueba a otra. Estas variaciones pueden deberse en parte a la falta de control de las condiciones de prueba; a los cambios en la condición del Sujeto mismo, (enfermedad, fatiga, tensión emocional, preocupación, experiencias recientes de naturaleza agradable o desagradable, etc.). La estabilidad temporal de una prueba, depende en parte, del lapso mayor o menor que interviene entre la primera y segunda medición.

Las fluctuaciones azarosas, a corto plazo, que ocurren durante intervalos que van de unas cuantas horas a algunos meses por lo general se incluyen dentro de la varianza de error de la calificación de la prueba. Con este tipo de confiabilidad, se aconseja que sea corto dicho intervalo, más aún si se trata de niños pequeños, ya que a edades tempranas se producen cambios de desarrollo más rápidos que en los adultos. Los estudiosos del tema consideran que el intervalo de tiempo entre la aplicación de las pruebas no deberá exceder a los seis meses.

Con períodos de tiempo mayores cualquier cambio adicional en la ejecución relativa de una prueba es probable que sea acumulativo y progresivo y no tanto debido al azar. En esas condiciones los cambios encontrados caracterizan un área de conducta más amplia que la que cubre la ejecución de la prueba. Así por ejemplo, el nivel general de un individuo en aptitud verbal, comprensión mecánica, o juicio artístico, puede verse apreciablemente alterado en largos períodos de tiempo debido al cúmulo de experiencias ocurridas, comunes o poco comunes durante ese tiempo. El status de un Sujeto puede haber cambiado en forma apreciable en relación a otros de su misma edad, pueden producirse cambios en el hogar, en el trabajo o en la escuela; cambios en el organismo (maduración fisiológica y psicológica, etc.)

El grado en el que tales factores pueden afectar el desarrollo psicológico del individuo plantea un importante problema para la investigación; esto sin embargo, no debe confundirse con la estabilidad de una prueba particular. Así, si se mide, por ejemplo la confiabilidad de una prueba de inteligencia, de personalidad, por lo general la prueba de la estabilidad se hace habiendo transcurrido tan sólo unas semanas. Se han hecho estudios en los que se han replicado las pruebas con intervalos de tiempo grandes, pero los resultados por lo general se discuten, o se habla de ellos, en términos de "constancia del nivel intelectual",

predictibilidad de la inteligencia del adulto a partir de la ejecución infantil, más que en términos de la confiabilidad de una prueba determinada. El concepto de confiabilidad por lo general se restringe a cambios azarosos a corto plazo, cambios que caracterizan el comportamiento de la prueba en sí misma.

### **11.1.2. Muestreo de reactivos**

Con toda seguridad, todos hemos pasado por la experiencia de tomar un examen en alguna materia y sentido que nos "iluminaba" la suerte porque muchos de los reactivos cubrían los temas que habíamos estudiado con más cuidado. En alguna otra ocasión, nos pudo haber sucedido lo contrario, es decir, nos encontramos con una gran cantidad de preguntas acerca de los temas que no habíamos revisado sino tan solo por "encima". Esta situación ilustra una segunda fuente de varianza de error en las calificaciones de las pruebas. ¿Hasta qué grado dependen las calificaciones de esta prueba, de los factores específicos de la selección particular de reactivos? Si un investigador diferente, trabajando en forma independiente prepara otra prueba de acuerdo con las mismas especificaciones, ¿qué tanto diferiría la calificación de un individuo en ambas pruebas?

Supóngase que se construye una prueba de vocabulario de 40 reactivos para obtener una medida general de comprensión verbal. Después se construye una segunda prueba con el mismo número de reactivos, solo que con palabras diferentes. Ambas tienen el mismo propósito, se ha tenido cuidado que en ambas, los reactivos cubran el mismo rango de dificultad.

Las diferencias en las calificaciones obtenidas por los mismos individuos en estas dos pruebas ilustran el tipo de confiabilidad que se está considerando. Debido a factores fortuitos en la experiencia pasada de diferentes individuos, la dificultad relativa de las dos pruebas variará algo de persona a persona. En esta forma, la primera prueba puede tener un mayor número de palabras desconocidas para el sujeto A que la segunda. Por otro lado, ésta puede contener demasiadas palabras desconocidas para el sujeto B. Si los dos individuos son aproximadamente iguales en sus conocimientos total de palabras (v.gr., en sus "calificaciones verdaderas"); de todas maneras B excederá a A en la primera prueba mientras que A excederá a B en la segunda. La localización relativa de estos dos sujetos en las dos pruebas será reversible, debido a la diferencia azarosa en la selección de los reactivos.

### **11.1.3 Homogeneidad de los reactivos**

La homogeneidad de una prueba se refiere esencialmente a **la consistencia de la ejecución de todos los reactivos dentro de una**

**prueba.** Por ejemplo, si una prueba tiene solo reactivos de multiplicación, mientras que otra comprende reactivos de sumas, restas, multiplicaciones y divisiones, la primera probablemente tendrá más consistencia entre sus reactivos que la segunda. En la segunda, la prueba más heterogénea, un sujeto puede responder mejor en la parte de las restas que en cualquiera de las otras operaciones aritméticas; otro sujeto puede responder relativamente bien en la parte de las multiplicaciones que en cualquiera de las otras operaciones aritméticas; otro, puede responder relativamente bien en la parte de las divisiones, pero en forma más pobre en sumas, restas y multiplicaciones; y así sucesivamente. Un ejemplo más extremo: una prueba con 40 reactivos de vocabulario, en contraste con otra que tiene 10 reactivos de vocabulario; 10 de relaciones espaciales, 10 de razonamiento aritmético y 10 de discriminación perceptiva. Es probable que en la segunda se encuentre poca o ninguna relación, en la ejecución del Sujeto, entre los diferentes tipos de reactivos.

Como es de esperarse las calificaciones de una prueba serán menos ambiguas cuando éstas se derivan de pruebas relativamente homogéneas. Supóngase que en la prueba de los 40 reactivos altamente heterogéneos, antes citada, los Sujetos A y B obtuvieron ambos calificación de 20. ¿Se puede concluir que las ejecuciones de estos dos Sujetos en dicha prueba fueron iguales?. Por supuesto que no. El Sujeto A pudo haber completado en forma correcta 10 reactivos de vocabulario, y los 10 reactivos de discriminación perceptiva y ninguno de los reactivos de razonamiento aritmético y de relaciones espaciales. En contraste, el sujeto B pudo haber obtenido una calificación de 20 respondiendo correctamente a cinco reactivos de cada uno de los cuatro tipos diferentes de reactivos. Se podrían producir muchas otras combinaciones que arrojarían una calificación total de 20. Pero la calificación tendría un significado muy diferente al haberse obtenido de combinaciones de reactivos tan disímiles. Por otro lado, en la prueba relativamente homogénea de vocabulario, una calificación de 20 probablemente significa que el Sujeto contestó correctamente aproximadamente las primeras 20 palabras, si los reactivos estaban ordenados por grado creciente de dificultad. Pudo haber fallado en dos o tres de las palabras más fáciles y respondido correctamente a dos o tres de las más difíciles, pero tales variaciones individuales, son ligeras en comparación con aquellas encontradas en la prueba más heterogénea.

Un aspecto relevante en relación a esto se refiere al grado de relativa homogeneidad o heterogeneidad de la misma variable que la prueba trata de medir. Aunque las pruebas homogéneas son preferidas debido a que sus calificaciones permiten una interpretación relativamente poca ambigua, una única prueba homogénea, obviamente, no es un predictor adecuado de un criterio altamente heterogéneo.

En la predicción de un criterio heterogéneo, la heterogeneidad de los reactivos de una prueba no necesariamente representa la de error. Las pruebas de inteligencia tradicionales proporcionan un buen ejemplo de pruebas heterogéneas; quizá sea más deseable construir varias pruebas relativamente homogéneas, cada una midiendo una fase diferente del criterio heterogéneo. En esta forma se podría combinar una interpretación de calificaciones sin ambigüedades y un cubrimiento adecuado del criterio.

¿En qué forma difiere la homogeneidad de la adecuación del muestreo de reactivos? Un ejemplo extremo servirá para resaltar la diferencia; supóngase que cada uno de los reactivos de cierta prueba, mide una función diferente y no relacionada. Sería totalmente posible construir otra prueba paralela a la primera, que contuviera el mismo tipo y distribución de reactivos. Teóricamente las calificaciones de estas dos formas podrían estar muy de acuerdo, indicándose en esta forma una alta confiabilidad de la prueba en términos de muestreo de reactivos. Sin embargo, la homogeneidad de esta prueba sería cercana a cero, ya que la consistencia de la ejecución de un reactivo a otro dentro de cualquiera de las formas no sería mejor que la dada por el azar.

#### **11.1.4 Confiabilidad del examinador y el calificador**

Deberá ser claro ahora que los diferentes conceptos de confiabilidad de una prueba dependen de los factores que se coloquen bajo el término "varianza de error". En un caso, la varianza del error abarca fluctuaciones temporales, en otro, se refiere a las diferencias entre conjuntos de reactivos paralelos; y en otro más incluye cualquier inconsistencia entre los reactivos. Por otro lado, los factores excluidos de las medidas de la varianza de error son de dos tipos: a) aquellos factores cuya varianza debiera permanecer en las calificaciones, ya que son parte de las diferencias reales o verdaderas bajo consideración; y b) aquellos factores irrelevantes que pueden ser controlados experimentalmente. Por ejemplo, no es común reportar los errores de medición que resultan cuando una prueba es administrada en condiciones distractoras o con tiempos límite menores o mayores que los especificados en el manual. Sin embargo, como los errores de tiempo y de distracción serios pueden eliminarse en forma empírica de la situación de prueba, no es necesario reportar coeficientes de confiabilidad especiales correspondientes a la "varianza de tiempo", o "varianza de distracción".

En forma semejante, la mayoría de las pruebas proporciona procedimientos altamente estandarizados para la administración y calificación por lo que se puede suponer que la "confiabilidad del examinador" y la "confiabilidad del calificador" son lo

suficientemente altos para propósitos prácticos. En esta forma, no existe ninguna necesidad especial de medir estos tipos de confiabilidad. Esto es particularmente cierto para pruebas de grupo, diseñadas para ser aplicadas a grandes grupos de sujetos y para ser calificadas por máquinas. En tales pruebas sólo se debe asegurar que se **sigan** en forma cuidadosa los procedimientos prescritos. El problema es por tanto de **control empírico de las condiciones**.

Sin embargo, en ciertas pruebas individuales, el papel del examinador es mucho más complejo. Como ejemplo se puede citar la prueba de Stanford Binet, y la mayoría de las pruebas preescolares. El procedimiento de prueba en tales casos no está tan rígidamente estandarizado. Mucho depende del éxito que el examinador tenga en el establecimiento de **rapport** y en la producción de la motivación adecuada. Con frecuencia la ejecución del Sujeto necesita ser evaluada por el examinador durante el mismo proceso de administración de la prueba, ya que tal ejecución determina en qué forma procederá el durante la prueba. Bajo tales condiciones es probable que aún examinadores muy calificados obtengan a veces resultados diferentes de los mismos sujetos. Estas variaciones en la calificación constituiría la varianza de error atribuible a idiosincrasia o diferencias individuales entre los examinadores.

En pruebas en las que la idiosincrasia del examinador puede jugar una parte apreciable, es deseable obtener alguna medida de la "confiabilidad del examinador", especialmente cuando han de combinarse los resultados obtenidos por varios examinadores. Para este tipo de pruebas se deberá dar igual importancia al índice de confiabilidad de examinadores, como se le da a otros tipos de índices de confiabilidad

#### **11.1.5. Fuentes principales de error**

Ya se ha señalado que los instrumentos de medición son confiables en la medida en que son repetibles, y que cualquier influencia azarosa que tienda hacer que las medidas sean diferentes de una ocasión a la siguiente es una fuente de error. En la práctica hay muchos factores que hacen que los instrumentos de medición no sean exactamente repetibles; el número y tipo de factores depende de la naturaleza de la prueba y de cómo se utilice ésta. Ahondando en algunas de las fuentes principales de error en la medición, se pasa a exponer algunos ejemplos.

##### **11.1.5.1 Variación dentro de una prueba**

Es importante hacer una distinción entre los errores de medición que producen variación en la ejecución de un reactivo al siguiente, dentro de una prueba y los errores que se manifiesten sólo en la

variación de la ejecución en diferentes formas de una prueba aplicada en diferentes tiempos u ocasiones.

La principal fuente de error de una prueba es debido al muestreo de reactivos. De acuerdo con el modelo domino-muestra, cada persona tiene una probabilidad particular de responder en forma correcta a cada reactivo, que depende de su calificación verdadera y de la dificultad del reactivo para la gente en general. En el caso más simple, si una persona tiene una calificación verdadera promedio y todos los reactivos tienen un índice de dificultad del 0.5 para la gente en general, esa persona tiene una probabilidad de 0.5 de responder correctamente cualquier reactivo seleccionado al azar dentro del dominio. Se esperaría que respondiera, en forma correcta la mitad de los reactivos en cualquier prueba que se sacara del dominio, pero esta expectativa estaría acompañada de algún error. Entre más reactivos tuviera la prueba, menor sería el error, la misma lógica se puede extender a los reactivos que no tienen una respuesta "correcta" (respuestas de sentimiento). Por ejemplo, en reactivos que se refieran a estar o no de acuerdo con ciertas afirmaciones. Se puede pensar que cada persona tiene una probabilidad establecida de estar de acuerdo con cada afirmación, lo que a su vez llevaría a un número esperado de respuestas de acuerdo dentro de una muestra de reactivos. Dependiendo del número de reactivos en cada muestra, habría alguna variabilidad en las calificaciones de una prueba a otra prueba.

El error debido al muestreo de reactivos es totalmente predecible a partir de la correlación promedio. En consecuencia, el **coeficiente alfa** sería la medida correcta de la confiabilidad para cualquier tipo de reactivo, y la versión especial de esa fórmula, la (KR-20, Kuder - Richardson 20), para pruebas de reactivos dicotómicos (Nunnally y Bernstein, 1993).

En las pruebas de elección múltiple, la adivinación es una fuente de error de la medición. Si por ejemplo, un individuo realmente no sabe la respuesta a dos preguntas, puede contestar correctamente una y no la otra debido a que adivinó. El adivinar produce alguna variación en la ejecución de un reactivo al siguiente, y esto tiende a disminuir la confiabilidad de la prueba. El adivinar es manejado con facilidad por el modelo **dominio-muestra**. Puede pensarse que el dominio está constituido por reactivos de opción múltiple. La correlación típica entre tales reactivos permitiría una estimación de la confiabilidad de cualquier muestra de reactivos. El adivinar podría servir para disminuir la correlación típica; pero una vez que ésta fuera estimada de las correlaciones dentro de una prueba, se podría usar para estimar la confiabilidad.

Además de la adivinación, muchos otros factores producen variación en las calificaciones de un reactivo al siguiente dentro de una

prueba. Por ejemplo, a un Sujeto, le puede empezar a doler la cabeza cuando está resolviendo una prueba; esto tenderá a disminuir sus calificaciones en los reactivos que contestó cuando apareció su malestar; otra persona puede tener la intención de marcar la alternativa **a** para un reactivo en particular, y debido a un error marcar en cambio la alternativa **b**: otra puede, inadvertidamente, alterar un reactivo que podría haber contestado en forma correcta; a la mitad de una prueba, una persona puede darse cuenta que mal interpretó las instrucciones de la forma en que se deberá responder y por no tener tiempo de regresar a los reactivos anteriores, resulta que su desempeño fue mejor en los reactivos sucesivos que en los primeros; también puede suceder que una persona que realmente sabe la respuesta a una pregunta, puede responder en forma incorrecta porque accidentalmente leyó "no es un ejemplo de" en lugar de " es un ejemplo de". El número de ejemplos para señalar factores que producen errores dentro de las pruebas, podría formar enormes listas, pero lo que interesa es dejar claro las causas de variación en las pruebas.

Hasta cierto punto, pueden estimarse los errores de calificación para una prueba. En las pruebas objetivas, los errores de calificación son puramente mecánicos, pero como tienden a disminuir las correlaciones entre los reactivos, caen dentro del campo del modelo dominio-muestra. En algunas pruebas la calificación es principalmente subjetiva, como por ejemplo, en los exámenes de tema o ensayo y en la mayoría de las pruebas proyectivas. El elemento de error de medición en este caso está causado por las fluctuaciones en la norma de calificación de un calificador individual, y por las diferencias en las normas de diferentes calificadores. Para el calificador individual, tales errores se manifiestan dentro de una prueba si cada reactivo es calificado Independientemente de los otros reactivos. Por ejemplo, en un examen de temas el instructor puede calificar todas las respuestas a la pregunta número uno; después calificar todas las respuestas a la pregunta dos, y así sucesivamente. Si tales calificaciones son independientes, la corrección promedio entre los reactivos puede usarse para obtener una estimación exacta de la confiabilidad.

Todos los errores que ocurren dentro de una prueba pueden ser fácilmente abarcados por el modelo dominio muestra. Las suposiciones del modelo pueden extenderse al caso donde las influencias situacionales son "asignadas" azarosamente a los reactivos. En esta forma, no solamente a cada persona se le administraría una muestra aleatoria de los reactivos del dominio si no que cada reactivo estaría acompañado por un conjunto azaroso de factores situacionales. Así, el que una persona pase o no cualquier reactivo obtenido al azar del dominio es función en parte, de la frecuencia con que un reactivo sea escogido y en parte, de la frecuencia de los factores situacionales que acompañan al reactivo.

Todas estas fuentes de error tenderán a disminuir la correlación promedio entre los reactivos de una prueba, pero la correlación promedio es todo lo que se necesita para estimar la confiabilidad.

#### **11.1.5.2. Variaciones entre pruebas**

Si se administran formas alternativas, equivalentes o paralelas de una prueba con un intervalo de dos semanas de tiempo transcurrido entre ellas, casi nunca correlacionarán en forma perfecta los dos conjuntos de calificaciones. El modelo dominio-muestra proporciona una predicción de la correlación, y como se dijo anteriormente, la predicción toma en cuenta no sólo el muestreo del contenido, sino también muchas fuentes de error dentro de cada sesión de prueba. Hay sin embargo, tres fuentes de error principales que intervienen entre la administración de diferentes pruebas que no son precisamente estimadas a partir de la correlación promedio de los reactivos dentro de cada prueba. El modelo dominio-muestra es un muestreo real de los reactivos de un dominio hipotético. Dos pruebas de ortografía construidas independientemente por dos personas pueden enfatizar diferentes tipos de palabras. Entonces, la correlación entre las dos pruebas puede ser menor que la predicha a partir de la correlación promedio entre los reactivos de cada prueba. En forma semejante, forma alternativas de un instrumento que mide actitudes hacia las Naciones Unidas, pueden ser sistemáticamente diferentes en contenido, y en consecuencia la correlación entre las dos formas sería menor que la predicha por el modelo dominio-muestra.

Un segundo factor que produce variación en las calificaciones en algunas pruebas de una ocasión a la siguiente, es debido a la subjetividad de la calificación. En un examen por temas o en una prueba proyectiva, el mismo examinador puede dar clasificaciones algo diferentes a las mismas personas, y aún diferencias mayores deberán esperarse entre las calificaciones dadas por diferentes examinadores. Previamente se dijo que parte del error debido a la subjetividad de la calificación de una persona podría estimarse a partir de la correlación entre los reactivos dentro de una prueba, si los reactivos se calificaran independientemente, pero esto tocará a tan solo una parte del error. El calificador puede cambiar sus normas en alguna forma de una ocasión a la siguiente. por ejemplo, entre las dos administraciones de la prueba, el examinador puede considerar un tipo particular de respuesta como más patológico de lo que previamente había considerado. Antes se dijo que si diferentes partes de una prueba son calificadas en forma independiente por diferentes examinadores, la correlación promedio entre los reactivos sería indicativa del error involucrado al usar diferentes examinadores; pero como rara vez colaboran dos examinadores en esta forma, difícilmente se conoce la cantidad de error que existe en un examinador.



Otra fuente de variación en la ejecución de una prueba de una ocasión a la siguiente es debida al hecho de que la gente realmente cambia al respecto del atributo que se está midiendo. Una persona puede sentirse mucho mejor de una ocasión a la siguiente, puede estudiar el contenido del domino, o puede cambiar su actitud hacia las Naciones unidas. Es razonable pensar que exista cierta fluctuación en las habilidades de un día a otro, dependiendo de factores fisiológicos y ambientales. Lo mismo sucede con los estados de ánimo, autoestima, y actitudes hacia la gente y cosas. Tales cambios en la gente tenderán a hacer que las correlaciones entre las formas alternativas de las pruebas sean menores que las predichas por la correlación promedio de los reactivos de cada prueba.

### **11.2 Métodos Experimentales para obtener la confiabilidad de una prueba**

La confiabilidad se puede definir como la "correlación entre pruebas paralelas". La definición de pruebas paralelas se expresa en términos de igualdad de medias, desviaciones estandar e intercorrelaciones.

El término confiabilidad fue introducido por Spearman en sus trabajos básicos de teoría de las pruebas (Spearman, 1904,1907,1910 y 1913). Desde entonces ha habido muchas discusiones de los varios factores que influyen sobre la confiabilidad en relación a los diferentes métodos de medida. Existen muchas formas diferentes de clasificar los factores que influyen sobre la confiabilidad y sobre los métodos para medirla. Entre ellas se cuentan los siguientes métodos:

- a) el uso de pruebas paralelas
- b) "retest" con la misma forma de la prueba
- c) varios métodos de mitades, tales como la primera contra la segunda mitad; reactivos pares contra nones, y el método de subpruebas al azar apareadas ( ya sea mitades o tercios).

Recientemente se han elaborado métodos para estimar la confiabilidad de una prueba de homogeneidad que no hacen uso de la correlación de calificaciones paralelas. En lugar de eso, estos métodos usan datos de análisis de reactivos para estimar la homogeneidad del grupo de reactivos de una prueba.

Aunque el error de medición es un concepto más básico en la teoría de las pruebas que el coeficientes de confiabilidad, se ha vuelto costumbre durante los últimos 50 años evaluar a las pruebas en términos del coeficiente de confiabilidad más que en términos del error de medición. Como existen ventajas y desventajas para cada una de estas medidas, se sugiere que se utilicen ambas para la

evaluación completa de cualquier prueba. Otis y Knollin (1921) señalaron que el error de medición es superior al coeficiente de confiabilidad ya que no varía con cambios en la heterogeneidad del grupo. Kelley (1921) indicó que, aunque el error de medición no varía con la heterogeneidad del grupo, sin embargo, la unidad en la que se expresa el error de medición sí varía de una prueba a la otra. Lincoln (1932) y (1933) señaló que la confiabilidad podría ser muy alta aun cuando las diferencias entre dos conjuntos de medidas fueran muy grandes.

Las pruebas o subpruebas que se correlacionen para determinar la confiabilidad de una prueba, deberán ser paralelas tanto en el sentido de que satisfagan los criterios estadísticos de pruebas paralelas (Gulliksen, 1950, capítulo 14) como en el sentido de que los reactivos requieren los mismos procesos psicológicos y el mismo tipo de aprendizaje por parte de los Sujetos. Este último criterio depende del juicio del técnico en pruebas y el experto en la materia, y será diferente para cada tipo de prueba. Aquí, se considera sólo los métodos generales del establecimiento de pruebas o subpruebas paralelas que son comunes a todos los tipos de material.

#### **11.2.1. Uso de las formas paralelas**

Para la mayoría de las situaciones, se ha encontrado que el mejor método para obtener la confiabilidad de una prueba es construir formas paralelas de la prueba y administrarlas en diferentes ocasiones al mismo grupo de sujetos. Así el método comúnmente usado será construir dos formas paralelas para este propósito; pero se sabe que con tres formas paralelas es posible hacer una evaluación más completa y además permite asegurar que las formas sean paralelas, no sólo con respecto a sus medias y varianzas, sino también en relación con sus correlaciones.

Existe sólo una situación en la que el uso de formas paralelas administradas en diferentes días no es aconsejable. Esto es, cuando la habilidad que está siendo probada cambia marcadamente en el intervalo de tiempo transcurrido entre las pruebas. Por ejemplo, si se quiere determinar la confiabilidad de una prueba de mecanografía administrando una forma a un grupo el lunes y otra forma el viernes, el método no funcionaría si el grupo estuviera practicando (y por lo tanto aumentando rápidamente su habilidad mecanográfica) durante el intervalo del tiempo transcurrido.

En la misma forma, este método no es adecuado si la primera prueba se da cuando los Sujetos están en excelente "condición" y la segunda se aplica cuando la habilidad de los Sujetos ha disminuido, por falta de práctica durante la semana transcurrida entre ambas aplicaciones.

El mismo tipo de consideración se aplica por ejemplo, a cualquier prueba de destreza física o habilidad o dominio muscular. Las dos administraciones de la prueba no pueden usarse para estimar la confiabilidad de la prueba si existe una buena razón para creer que los sujetos han mejorado o desmejorado en la variable que está siendo medida.

Para la mayoría de las pruebas de logro escolar y habilidad mental, es razonablemente fácil estar seguros de que los sujetos no han cambiado realmente en forma marcada durante el período que interviene entre las dos pruebas (Gulliksen 1950, pag. 195). Para otros tipos de ejecución, de los cuales las habilidades atléticas de varios tipos son un buen ejemplo, es muy difícil mantener a un grupo en un estado de excelencia uniforme. Es probable que la habilidad se deteriore con la falta de práctica, y/o pueda mejorar o pueda estancarse con la misma. En tales casos todo el "error de medición" no puede ser atribuido a la prueba. Mucho de lo que se manifiesta en la prueba estadística como error de medición es de hecho una variabilidad real de la habilidad. Sin embargo, desde otro punto de vista se debe reconocer que la medición de algunas habilidades es en extremo poco confiable (independiente de la causa de esta falta de confiabilidad); en ese caso, al utilizar cualquier de tales mediciones, se deben tratar, como se tratarían las mediciones muy poco confiables.

Sin embargo, si se está manejando un periodo de tiempo durante el cual la habilidad medida o la variable no cambiará en forma sistemática para los diferentes miembros del grupo, y se está trabajando con un grupo de Sujetos bajo condiciones tales que no es probable que la habilidad o variable cambie, la utilización de las diferentes formas de la prueba es el método más adecuado para indicar la confiabilidad.

Deberá de señalarse que las posibilidades de error anotadas arriba pueden detectarse con facilidad. Si el grupo se ha mejorado o se ha deteriorado, la media será más alta o baja en la segunda ocasión. Si alguna persona han mejorado y otras han desmejorado, la desviación estándar con toda probabilidad cambiará. Un conjunto complicado de influencias en el que algunas personas mejoran y otras se deterioran en tal forma que la media y la desviación estándar del grupo permanezcan iguales, es una posibilidad que puede existir, pero sin duda alguna sería muy extraña o rara.

En resumen, el método de aplicar pruebas paralelas con un intervalo de tiempo entre ellas, es un método que permite que las fuentes de error relevantes influyan sobre el coeficiente de confiabilidad. Si se utilizan las pruebas estadísticas de medias y desviaciones estandar, y si se satisfacen, el método es uno que

puede usarse en forma rutinaria con relativamente poco temor de que factores irrelevantes y no detectados estén produciendo que el coeficiente de confiabilidad obtenido sea uno espuriamente alto o espuriamente bajo.

Se debe notar, que el método de las formas paralelas es válido para las pruebas de velocidad. Una prueba de velocidad es una prueba compuesta de reactivos, muy fáciles. Reactivos tan fáciles que podrían ser contestados todos por todo el grupo si se les permitiera tiempo para hacerlo. Por ejemplo, un conjunto de suma de dos dígitos dados a alumnos de segundo de secundaria se aproximaría a ser una prueba de velocidad. Si se va a obtener un buen rango de calificaciones en tal prueba, es necesario que se tengan un gran número de reactivos, y establecer un tiempo límite tan corto que únicamente los mejores del grupo terminen, si es que lo hacen. En tal prueba, es importante el efecto de la práctica de una vez a la siguiente. A menos de que condiciones tales como cantidad de práctica y el uso de "ejercicio previo" estuvieran cuidadosamente estandarizados, no sería posible que se tuviera la misma media y varianza en las formas paralelas para el grupo. Sin embargo, si las medias y las varianzas son iguales, uno puede estar razonablemente en lo cierto al decir que la intercorrelación de las dos formas es una aproximación razonable del coeficiente de confiabilidad que debería tener la prueba.

La confiabilidad de formas paralelas también se puede obtener administrando ambas formas en la misma sesión. Una vez más, en algunas pruebas, puede haber una marcada diferencia en la ejecución debido al hecho de que la aplicación de la primera prueba influyó a la segunda, por ejemplo, si es una prueba de velocidad de sumas de dos dígitos, es probable que para muchas personas, particularmente las peores, la calificación en la segunda prueba sea mucho mejor debido a la práctica obtenida en la primera prueba. Por supuesto que esto puede detectarse fácilmente en los resultados ya que la media sería mayor para la segunda forma. Existen otras pruebas para las cuales la ejecución en la segunda forma es muy probable que sea peor que la ejecución en la primera. Cualquier prueba que produzca fatiga o cansancio en los sujetos claramente entra dentro de esta categoría, y una vez más, tal fatiga podría ser detectada con facilidad a partir de los resultados. El promedio sería menor para la segunda prueba que para la primera.

Si las anteriores, más bien obvias y fácilmente detectables, dificultades no estuvieran presentes, la dificultad principal con la confiabilidad obtenida mediante la administración sucesiva de formas paralelas, es que es muy alta. Esto se debe al hecho de que no hay posibilidad de que la variación debida a la variabilidad normal diaria disminuya la correlación entre las formas paralelas. Woodrow (1932) en su estudio de la variabilidad cotidiana demostró

que existen variaciones de un día a otro en la ejecución de las pruebas.

Algunos otros autores han señalado que algunas veces una baja correlación entre dos formas paralelas de una prueba indica que la prueba es una medida inestable de un rasgo estable; en otras ocasiones tal correlación baja, puede surgir de una medición estable de un rasgo inestable. La inestabilidad, ya sea en la prueba o en el rasgo, resultaría en una baja correlación entre las formas paralelas. Métodos para determinar la inestabilidad de un rasgo como algo diferente de la inestabilidad de un instrumento han sido sugeridas por Paulsen (1931), Thouless (1936) y (1939), Preston (1940). Se puede entonces concluir que si las formas paralelas de una prueba son aplicadas en el mismo día y que si se satisface el criterio estadístico de las pruebas paralelas, específicamente el de igualdad de medias y desviaciones estándar, la confiabilidad obtenida es probablemente más alta que la que hubiera sido obtenida si se hubiera permitido que la variabilidad de un día a otro hubiera afectado a la confiabilidad.

Hablando en términos generales, entonces, el uso de dos o tres formas paralelas administradas en diferentes ocasiones es el mejor método para determinar la confiabilidad de una prueba. Sin embargo, como con frecuencia no se dispone de varias formas paralelas, y como también es difícil algunas veces asegurar la cooperación de los Sujetos durante períodos extendidos de tiempo, se considerará la posibilidad de obtener una indicación de la confiabilidad cuando sólo se dispone de una forma de la prueba.

### **11.2.2 "Retest" (Replicación) con la misma forma**

Algunas veces, cuando las formas paralelas de una prueba no están disponibles, es posible obtener una estimación de la confiabilidad administrando la misma prueba dos veces. Por lo general, es preferible hacer esto con un intervalo de tiempo transcurrido "regularmente grande" entre ambas aplicaciones. Una vez más con este método se debe de estar pendiente al respecto del efecto de la práctica o fatiga que será detectado con facilidad en la mayoría de los casos, observando las distribuciones de las calificaciones de la prueba en la primera y segunda aplicación (administración). A parte de tal efecto, el peligro principal en esta técnica es que la confiabilidad será muy alta debido a que existe la tendencia por parte del Sujeto de repetir su ejecución previa. Esto es, si el Sujeto no conoce la respuesta a un reactivo, pero con suerte lo adivina y lo resuelve bien, es más probable que lo vuelva a adivinar, la siguiente vez y se asegure el crédito para un reactivo para el que en realidad no conoce la respuesta. En la misma forma, si comete alguna pequeña equivocación, y como resultado responde en forma incorrecta a un reactivo que en situaciones normales hubiera

contestado en forma correcta, es más probable que repita su ejecución cuando se le vuelva a administrar la prueba. Tal efecto no ocurrirá si la persona se estuviera sometiendo a una forma paralela que no contiene los mismos reactivos. En otras palabras, la ejecución en la repetición de una prueba tiende a parecerse más a la calificación original que la ejecución en una forma paralela misma prueba. Por tal razón, es aconsejable que este método de repetición de la misma prueba en diferentes ocasiones no se utilice, ya que producirá un coeficiente espuriamente alto, y no es fácil determinar el grado de error.

Puede haber excepciones con ciertas variables, por ejemplo con discriminación perceptual, para la cual no se pueden construir formas paralelas. Por ejemplo, una prueba de discriminación de tonos o una prueba de umbrales de audición para diferentes tonos puros, podría ser replicada sin que tal efecto ocurriera. La persona simplemente juzga cada vez si oye un tono o si no oye un tono. En una prueba como ésta parece no haber una forma o manera sencilla en la que la persona pudiera repetir en forma espuria sus errores y éxitos del conjunto previo de ensayos. Sin embargo, aún en temas tan simples, con frecuencia es deseable construir varias técnicas de medición diferentes y correlacionarlas, así como obtener la confiabilidad de una prueba repetida por el uso de cada método. En general, se puede decir que aún cuando parezca que la repetición de la misma forma es todo lo que se puede hacer, está bien que el constructor de la prueba use algo de su ingenio para presentar un factor dado en diferentes formas que sean comparables **grosso modo**, y que después vea que tanto acuerdo existe, entre las diferentes pruebas. Con frecuencia se descubren cosas nuevas de la variable estudiada al ser medida en esta forma.

### **11.2.3 Consideraciones generales de los Métodos pos mitades**

Por lo general cuando sólo se dispone de una sola forma de la prueba, la confiabilidad se determina por algún método de mitades. Esto quiere decir que los reactivos de la única forma se dividen para formar dos, cada una con la mitad de número de reactivos de la forma original. Típicamente, los sujetos no saben que la prueba va a ser calificado en dos parte, y no saben qué reactivos estarán en cuál de las mitades. El experimentador no necesita decidir y por lo general no lo hace, cómo van a dividirse los reactivos hasta que ve los resultados de la prueba. Sin embargo, desde el punto de vista de establecer procedimientos de calificación eficientes, es deseable, decidir sobre la división para formar las dos subpruebas antes de que se mande a imprimir la prueba.

Los métodos discutidos en la secciones previas (ya sea formas paralelas o reaplicación con la misma forma), proporcionan al experimentador dos componentes (conjuntos) de calificaciones. En

tal caso la confiabilidad está dada directamente por la correlación producto Momento de Pearson entre los dos conjuntos de calificaciones. Es necesario un método ligeramente modificado cuando se va a calcular la confiabilidad a partir de las calificaciones de dos subpruebas obtenidas de una sola prueba original. Un método, es correlacionar las calificaciones de las dos mitades y entonces substituir esta correlación en la fórmula de Spearman-Brown para el doble de la longitud. podemos escribir:

$$r'_{xx} = \frac{2r_{12}}{1+r_{12}}$$

donde:  $r'_{xx}$  Designa la confiabilidad de la prueba total estimada al corregir la correlación por mitades al doble de longitud.

$r_{12}$  designa la correlación entre las dos mitades de la prueba.

Otro método de obtener la confiabilidad de la prueba total partiendo de la información contenida en los dos conjuntos de calificaciones de las subpruebas es utilizando la fórmula presentada por Rulon (1939):

$$r''_{xx} = 1 - \frac{s_d^2}{x_x^2}$$

donde:  $s_d^2$  es la varianza de  $x_1 - x_2$ , la diferencia de las calificaciones en las dos mitades de la prueba.

$s_x^2$  es la varianza de las calificaciones en la prueba total, la suma de calificaciones de las dos mitades de la prueba:  $x = x_1 + x_2$

$r''_{xx}$  se utiliza para designar la confiabilidad de la prueba.

Es deseable no usar la correlación entre dos conjuntos de calificaciones de las subpruebas para la estimación de la confiabilidad, si no dividir la prueba total en tres o posiblemente cuatro partes, y probar la semejanza de estas partes, así como obtener la correlación entre ellas. Estas correlaciones pueden usarse en la fórmula general de Spearman-Brown:

$$R_{k\bar{k}} = \frac{kr_{11}}{1 + (k-1)r_{11}}$$

donde:  $r_{11}$  es la confiabilidad de la unidad de prueba.

$k$  es el número de reactivos de la prueba alargada dividida entre el número de reactivos de la unidad de prueba, y

$R_{k\bar{k}}$  es la confiabilidad de la prueba alargada.

En el caso particular que se está viendo,  $k$  sería igual a 3 o 4. Al utilizar este método, se sabe que se está usando una correlación entre tres subpruebas **paralelas** como base para la obtención de la confiabilidad. Esto significa que la confiabilidad encontrada no será muy baja porque no se han escogido subpruebas no paralelas como base para estimarla.

El principal problema al usar calificaciones de subpruebas para el propósito de estimar la confiabilidad, es dividir la prueba original en subpruebas equivalentes. A continuación se considerarán algunos de los métodos para la división de la prueba en subpruebas, así como las ventajas y desventajas de cada uno.

### 11.2.3 Mitades o tercios sucesivos.

Dividir una prueba en mitades o tercios equivalentes no es cosa sencilla. Por ejemplo, la manera más fácil de dividir la prueba es tomar la primera mitad de la prueba contra la segunda. Con frecuencia un método como éste, no dará como resultado pruebas paralelas. Por ejemplo, si la prueba es administrada en una sola sesión y es contra reloj, cualesquiera reactivos que no se contestaran por falta de tiempo se encontrarían en la segunda mitad de la prueba. La calificación de la segunda mitad, sería menor que la de la primera. Para una prueba de velocidad compuesta de reactivos fáciles, los resultados de graficar las calificaciones de la primera mitad contra las calificaciones de la segunda serían muy peculiares. Todos los sujetos que no llegaron a la segunda mitad, tendrían una calificación de cero en ésta; independientemente de la calificación que hayan obtenido en la primera mitad. Si la prueba es una prueba de velocidad pura, en el sentido de que la mayoría de los Sujetos pueden contestar correctamente a los reactivos si se enfrentan a ellos, en forma tal que los únicos errores serían "los reactivos no intentados", cualquier persona que termina la primera mitad obtiene una calificación perfecta o casi perfecta en ésta, independientemente de su calificación en la segunda mitad. Siempre que la calificación esté en **gran parte** determinada por el hecho de que el tiempo transcurrió antes de que muchos Sujetos hayan terminado, se aproxima uno a la situación antes descrita, y la primera contra la segunda mitad no serán "mitades comparables" adecuadas para obtener una estimación del coeficiente de confiabilidad.

Se puede pensar que, si todos los Sujetos terminaran las dos terceras partes de la prueba, se podría correlacionar el primer tercio de la prueba contra el segundo, y corregir este coeficiente para el triple de longitud. Sin embargo, un método tal es válido únicamente si el último tercio es paralelo a las dos mitades apareadas obtenidas de los primeros dos tercios. Si los reactivos difíciles están al final de la prueba, es imposible de hacer cualquier adivinación plausible al respecto de lo que sucedería si el límite de tiempo fuera aumentado en forma tal que todos pudieran terminar la prueba. Aún más, tal método no da la confiabilidad de la prueba con el límite de tiempo menor. Lo que hace es estimar la confiabilidad que se tendría si el límite de tiempo fuera tal que prácticamente todos terminarían la prueba. Si el límite de tiempo es



importante, se debe usar el método de formas paralelas para estimar la confiabilidad. Si el límite de tiempo es generoso en forma tal que la mayoría de los Sujetos terminen la prueba, es posible estimar la confiabilidad partiendo de las calificaciones de las subpruebas.

Además del problema de los límites de tiempo en una prueba, debe también considerarse el problema de la dificultad de los reactivos. Muchas pruebas están construidas con los reactivos fáciles al principio, los reactivos de dificultad promedio después, y los reactivos más difíciles al final de las mismas: es claro que si los reactivos de la prueba están ordenados de acuerdo a su grado de dificultad, la primera y segunda mitad no serán "mitades comparables". Se puede ver que si una prueba contiene un número de reactivos de dificultad promedio, y es alargada añadiendo reactivos más difíciles, la confiabilidad de la prueba disminuiría a pesar del aumento en la longitud de la prueba y en el tiempo límite. Los nuevos reactivos añadidos serán contestados en una base en el azar por la mayoría de las personas; siendo así que será accidental que contesten en forma correcta o incorrecta a los nuevos reactivos. A medida que se añade un gran número de reactivos difíciles, un componente mayor de la calificación se deberá al proceso de adivinación, y este componente disminuirá la confiabilidad de la calificación de la prueba alargada. Esto de ninguna manera contradice la formulación de Spearman-Brown, sobre la relación de la longitud de la prueba y la confiabilidad (esta formulación dice que entre más larga sea la prueba es más confiable), ya que esta formulación supone que el nuevo conjunto de reactivos es paralelo a los antiguos o anteriores. Esto significa que los reactivos tienen medias, desviaciones estándar y confiabilidad iguales. Los nuevos reactivos supuestamente añadidos aquí, serían reactivos difíciles con una media más baja, y como serían contestados al azar, la confiabilidad de esta nueva parte y su correlación con la parte más fácil de la prueba, estaría más cercana a cero (Gulliksen, 1960,p.203).

Partiendo de consideraciones como éstas, se ve que el efecto de aumentar el límite de tiempo de una prueba, es difícil de predecir. Aumentar el límite de tiempo permitirá que los Sujetos al no conocer la respuesta a los reactivos más difíciles al final de la prueba procederán a adivinar las respuestas a estos reactivos y añadir al azar su calificación. Este incremento no permanecerá estable de forma a forma; por lo tanto disminuirá la confiabilidad de la prueba.

Si se desea usar la primera y segunda mitades ( o los tercios sucesivos) de una prueba para calcular su confiabilidad, es posible planear una prueba en forma tal que se superen los problemas producidos por los tiempos límites y el grado de dificultad de los

reactivos: para el método de la primera contra la segunda mitad, por ejemplo, se arreglan los reactivos de la prueba de tal manera que el rango del grado de dificultad en la primera parte de la prueba sea respetado en la segunda parte. Entonces, si se da suficiente tiempo como para que todos o casi todos tengan la oportunidad de terminar la prueba, la primera y la segunda parte serán comparables. Si existe un efecto de **práctica o fatiga** a medida que el Sujeto avanza a lo largo de la prueba, pero si la prueba es administrada en dos sesiones, con tiempo entre ellas para descanso y relajación, si el grado de dificultad de los reactivos es igual en ambas sesiones, y se dan tiempos comparables para cada sesión, es probable que se pueda obtener una buena estimación de la confiabilidad correlacionando los resultados de la primera sesión contra los de la segunda.

### **11.2.3.2 División por reactivos nones contra pares.**

Con mucho, la forma más común de mitades comparables es la división de reactivos en pares y nones. Es probable que este método nunca de un valor muy bajo para el coeficiente de confiabilidad. Si hay error siempre será en la dirección de una confiabilidad que sea espuriamente alta. Algunas veces, como se verá, la confiabilidad de pares y nones sobrestima seriamente la confiabilidad de la prueba indicada por el método de formas paralelas.

Se puede ver fácilmente que, si los reactivos están ordenados de acuerdo con su grado de dificultad, los reactivos nones tendrán aproximadamente la misma dificultad promedio y dispersión de dificultad que los reactivos pares. Si existe alguna influencia, es posible que los reactivos nones sean en promedio ligeramente más fáciles que los reactivos pares.

Sin embargo, al usar este método, se debe asegurar que no existe dependencia de un reactivo con otro. En algunas pruebas se encuentran una serie de preguntas al respecto de un tópico dado, y es algunas veces difícil decidir si los reactivos son independientes, en el sentido de que el conocer la respuesta dependa principalmente de si se ha o no estudiando el tema o si existe una dependencia espuria entre ellos. En las pruebas de ejecución, cuando el Sujeto tiene que armar o desarmar un mecanismo, y se le califica en los diversos pasos, es muy probable que exista una relación espuria, en el sentido de que el Sujeto aprende o no un cierto conjunto de actos como una unidad, mientras que el examinador para poder calificar la ejecución en forma objetiva, establece una cantidad de divisiones más bien artificiales. En casos como estos, parece que la pregunta que se debe responder es: ¿establecería usted, como constructor de exámenes, tales mitades como pruebas separadas?. En un conjunto de afirmaciones que describen las características de los conos y los

bastones del ojo, por ejemplo, es posible que el constructo de la prueba pudiera usar sólo la mitad de las afirmaciones para tener una forma corta de la prueba. Sin embargo, es muy posible que los reactivos nones no constituyan una forma paralela satisfactoria de los reactivos pares. Los reactivos deberán de inspeccionarse para asegurar que el tipo de materia que cubren y la distribución de su dificultad de una de las mitades es **grosso modo** paralela al de la otra mitad.

La correlación pares-nones también es espuriamente alta en una prueba con un límite de tiempo muy pequeño porque un gran número de sujetos no terminan la prueba. Si un Sujeto no contesta los diez últimos reactivos de la prueba, obviamente no "pasa" ninguno de ellos. En esta forma obtiene cinco puntos más de error en su calificación non y también cinco puntos más de error en su calificación par. Es muy probable que una observación cuidadosa demuestre que muchas de las confiabilidades publicadas son espuriamente altas debido a este factor. Una vez más, este tipo de error queda ilustrado en forma muy clara en las pruebas de velocidad a las que se hizo referencia con anterioridad. Si cada Sujeto tiene todos los reactivos correctos hasta donde haya llegado, el que termine diez reactivos tendrá una calificación non de cinco y una calificación par de cinco, si termina once reactivos tendrá una calificación non de seis y una calificación par de cinco, y con doce reactivos, la calificación será de seis y seis. Esto es, la calificación nones y pares serán ya sea idénticas, o la calificación non será un punto más alta que la calificación par.

Deberá notarse que la confiabilidad pares - nones puede ser muy alta, aún cuando los reactivos estén ordenados de acuerdo a su grado de dificultad si les permite a los sujetos terminar la prueba, y los reactivos sean independientes uno de otro ( en el sentido de que cometiendo un error en uno de ellos no necesariamente aumenta la probabilidad de cometer un error en otro reactivo). La variabilidad debida a las variaciones de un día al otro en la habilidad, queda descontada y aún si la variación que pudiera ser causada por un efecto ligero de práctica o fatiga a medida que se progresa a lo largo de la prueba también queda descartada, si se usa el método de las formas paralelas como norma. La confiabilidad pares - nones, como se aplica por lo general a la mayoría de las pruebas, es muy probable que arroje un resultado bastante alto debido a que se pueden controlar diversas fuentes de variación y también debido a que por lo general, la mayoría de las pruebas tienen límites de tiempo logrando en esta forma, una buena proporción de la calificación ya que la mayoría de los Sujetos, no tienen oportunidad de intentar contestar los últimos reactivos.

En una prueba de velocidad en la que la calificación depende de qué tan rápido trabaja un Sujeto en el tiempo Límite dado, **no hay forma de estimar la confiabilidad** si no es aplicando una **prueba o forma paralela** una segunda vez. Ahora bien, diferentes métodos de medir la confiabilidad dan diferentes resultados: en general, la confiabilidad de formas paralelas es la más baja, y la pares - nones (corregida) es la más alta.

Se puede pensar que, si todos terminaron dos tercios de la prueba se podría usar una confiabilidad pares - nones en los dos primeros tercios, obtener la correlación entre estos dos, y corregirla al triple de longitud. Sin embargo, esto proporciona una estimación de la confiabilidad de la prueba total sobre la suposición de que **todos terminen** la prueba. No da una estimación del grado con el que un Sujeto alcanza la **misma tasa de velocidad** en diferentes administraciones de la prueba, y que por lo tanto, llegue al mismo punto en la prueba. No hay forma posible de estimar este factor con exactitud, excepto dando formas paralelas con **tiempos límites comparables** y bajo instrucciones estandarizadas, y cuidando el grado en el que las calificaciones sean las mismas.

#### 11.2.3.3 Subpruebas apareadas al azar:

Si las calificaciones son obtenidas en una sola aplicación de la prueba que va a usarse para estimar la confiabilidad de la misma, es necesario considerar a esta calificación única como dividida en dos, tres o cuatro calificaciones de subpruebas equivalentes. En las secciones anteriores se ha visto que bajo ciertas condiciones las mitades o tercios sucesivos de una prueba pueden ser razonablemente consideradas como formas paralelas, mientras que bajo otras condiciones los segmentos sucesivos de una prueba, no son paralelos en forma clara. De manera semejante, asignar cada segundo o tercer reactivos a una, dos o tres sub-pruebas puede ser un buen o mal método según diferentes condiciones para obtener subpruebas paralelas.

Si una prueba está compuesta de un gran número de reactivos independientes y es administrada con un tiempo límite "normal", se puede por lo general subdividir en subpruebas paralelas. Si una prueba tiene pequeños grupos de reactivos, como por ejemplo en la mayoría de la pruebas mecánicas o en pruebas que involucran escribir un párrafo, puede ser o no posible construir una prueba que esté compuesta por subpruebas paralelas. Si se usa un tiempo límite pequeño, no hay posibilidad de obtener alguna estimación válida de la confiabilidad usando calificaciones de sub-pruebas.

Si se tienen datos de análisis de reactivos de una prueba (que tiene un gran número de reactivos independientes y un tiempo límite liberal), los reactivos deberán de ser apareados en base en los

datos del análisis de reactivos y asignados a las subpruebas. Este es un método excelente de asegurar que las subpruebas sean paralelas. Por ejemplo, supóngase que se dispone del porcentaje de personas que contesta correctamente cada reactivo ( $p$ ) y se tiene también su correlación  $r_b$  biserial con el resto de los reactivos de la prueba. El mejor procedimiento para construir subpruebas paralelas es representar a cada reactivo por medio de un punto en un diagrama de dispersión, la abcisa para  $p$  y la ordenada para  $r_b$ . Para poder identificar los reactivos, cada punto, deberá estar señalado con el número del reactivo, como se demuestra en la fig. 11.1. Entonces los reactivos pueden ser apareados simultáneamente en  $p$  y  $r$ , trazando una línea alrededor de los pares, tríos o cuádruples apareados.

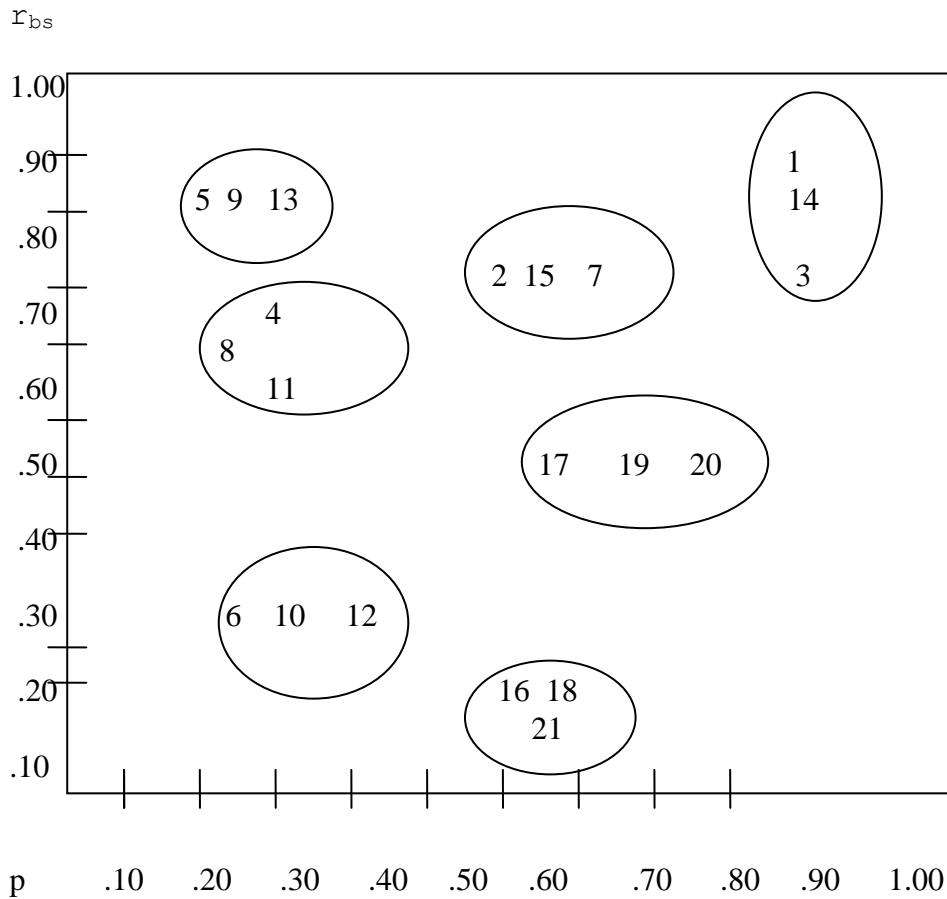


Figura 11.1 Cómo construir tres subpruebas o pruebas paralelas apareando simultáneamente reactivos sobre índices de dificultad y confiabilidad.

$p$  = proporción de Sujetos que pasa el reactivo o lo endosa afirmativamente

$r_{bs}$  = correlación reactivo-calificación total.

Es importante hacer notar que si la prueba es heterogénea con respecto al tipo de reactivos, o con respecto al tipo de material cubierto, es necesario aparear los reactivos en esos aspectos, así como en relación a p y r.

Se deberá entonces, asignar el azar, a un miembro de cada grupo a una subprueba dada. Por ejemplo, si solamente se forman dos subpruebas la asignación podría determinarse a través de un "volado", y asignar el reactivo con el número más pequeño del par a la forma A si "cae águila" y a la forma B si "cae sol". Al construir tres subpruebas paralelas, es necesario asignar a cada triada de reactivos a las diferentes tres subpruebas paralelas por medio de un procedimiento un poco más complicado. Por ejemplo los reactivos de cada tríada pueden identificarse a través de su número de reactivos como bajo, medianos y altos (B, M y A). Existen por lo tanto seis posibles manera de asignar estos tres reactivos, uno a cada una de las tres subpruebas. A cada uno de esos órdenes se les puede entonces asignar un número del 1 al 6 (1=BMA;2=BAM; etc.) y cada triada asignarse de acuerdo a la tirada de un dado.

Si la información del análisis de reactivos se encuentra disponible antes de que se constituya la prueba, se pueden simplificar las rutinas de calificación si los reactivos de una subprueba se ponen en primer lugar, y luego los de la segunda, etc.; o si los reactivos de las diferentes subpruebas se distribuyen con cierta secuencia en forma sucesiva a lo largo de la prueba.

Debe hacerse énfasis en que no importa en qué orden se emplean los reactivos, pero que sí es necesario permitir tiempo suficiente para que la mayoría de los Sujetos terminen casi todos los reactivos. No es posible estimar la confiabilidad de una prueba a partir de subpruebas paralelas si la calificación de la prueba se ve marcadamente influida por un tiempo límite.

#### 11.2.4 Correlación Intraclase

Uno de los procedimientos más utilizados en la estimación del coeficiente de confiabilidad es el que se sigue en la obtención del coeficiente de correlación intraclase:

$$p = 1 - \sqrt{\Sigma x_1 x_2 / 2 n \sigma^2 c^2}$$

donde:  $x_1$  es el puntaje obtenido, por un sujeto particular, en la primera medición y  $x_2$  es el puntaje obtenido por el mismo individuo en la segunda medición, n representa el número de sujetos que

respondieron al instrumento y  $\sigma^2$  corresponde a la varianza de los puntajes combinados de ambas mediciones.

Dos aspectos evidentes en la fórmula de la correlación intra clase son:

-que la confiabilidad depende de la precisión de las medidas, es decir, de la menor discrepancia entre los puntajes de la primera y segunda medición:

-que la confiabilidad depende de la población medida por el instrumento, en donde, el grado de precisión es relativo a la variabilidad que se da entre todos los puntajes. Una determinada discrepancia promedio producirá un alto coeficiente de confiabilidad si la variabilidad intersujeto es grande al compararla cuando ésta es pequeña.

Un alto coeficiente de confiabilidad indica que el instrumento de medición utilizado localiza de manera precisa a cada sujeto en relación a todos los otros sujetos que también fueron medidos. Sin embargo, en dicha localización sólo puede asumirse que ésta se ha hecho sobre la dimensión de un atributo. La determinación de si ese atributo corresponde efectivamente al propósito de la medición es problema de la validez.

### **11.2.5 Consistencia Interna.**

#### **11.2.5.1 Alpha de Cronbach.**

La ecuación Alpha de Cronbach para determinar la confiabilidad del tipo de consistencia interna inter-reactivos, es una de las deducciones más importantes de la teoría del error de medición (Nunally y Bernstein, 1993). La fórmula es la siguiente:

$$r_{kk} = k / k - 1 ( 1 - \Sigma\sigma_i^2 / \sigma_y^2 )$$

donde, K es igual al número de reactivos que componen el instrumento;

$\Sigma\sigma_i^2$  es igual a la suma de la varianza de los reactivos (en la matriz de covarianza se obtiene sumando los elementos de la diagonal principal).

$\sigma_y^2$  es igual a la varianza total, ( en la matriz de covarianza se obtiene sumando los elementos de la diagonal principal más dos veces la suma de los elementos que se encuentran fuera de dicha diagonal).

Se puede derivar la misma fórmula del modelo de las pruebas paralelas, y se pueden derivar fórmulas semejantes a partir de otros modelos matemáticos de la medición del error. Esta fórmula representa la correlación que se espera de una prueba con una forma

alternativa que contenga el mismo número de reactivos. La raíz cuadrada del coeficiente Alpha es la correlación estimada de una prueba con calificaciones verdaderas sin error. Esta fórmula debería aplicarse en forma rutinaria a todas las pruebas nuevas.

El coeficiente Alpha se podrá calcular para una prueba o escala con opciones de respuesta múltiple ( más de dos) siempre y cuando **todos** los reactivos de la prueba tengan el mismo número de opciones de respuesta. EL procedimiento es el siguiente:

1. Se califica cada reactivo con el peso de la opción marcada por el sujeto
2. Se califica toda la prueba, usando los pesos de los reactivos individuales: calificación total.
3. Se prepara una matriz de puntajes de acuerdo con la figura 11.2. En ella se escriben los puntajes que obtiene cada Sujetos en cada reactivo y la calificación total.

SUJETOS	REACTIVOS								C.T.	
	1	2	3	4	5	.	.	.		K
1										
2										
3										
.										
.										
.										
n										

FIGURA 11.2 MATRIZ DE CALIFICACIONES

4. Se obtiene la media para cada columna, incluyendo la de las calificaciones totales (C.T.). Se prepara entonces otra matriz de desviaciones, donde se anota en cada celdilla, la desviación de cada puntaje con respecto a su media, la de su columna.

5. Se procede entonces a elevar al cuadrado cada desviación y se suman éstas para cada columna dividiendo la suma entre el número de sujetos (n), obteniendo así la varianza de cada reactivo y la de calificación total ( $\sigma^2$ ).

6. por último se procede a substituir los valores en la ecuación y se obtiene  $r_{kk}$

7. Se busca el nivel de significancia de  $r_{kk}$  con N-K grados de libertad.

Cuando se efectúa la investigación de la confiabilidad de una prueba compuesta por reactivos dicotómicos (sí y no; falso-verdadero, correcto-incorrecto), el coeficiente Alpha adopta la siguiente fórmula especial:

$$r_{kk} = k/k-1 ( 1 - \Sigma pq / \sigma_{2y} )$$

donde:



$\sum pq$ : la suma del producto de la proporción de sujetos que contesten una opción (correcta, o aquella que tenga el peso de 1) por la proporción de sujetos que contestaron en la otra opción (o sea 1-p).

Los pasos a seguir para la determinación de  $r_{kk}$ .en la fórmula anterior son los siguientes:

1. Se encuentra el valor p de cada reactivo, que se multiplica entonces por 1-p.
2. Se suman estos productos
3. Se calcula la varianza de las calificaciones totales ( $\sigma^2$ ).
4. Se divide la  $\sum pq$  entre  $\sigma^2$ .
5. Se resta este número de 1
6. Se multiplica el resultado de (5) por la proporción del número de reactivos entre ese número menos 1
7. Se determina su nivel de significancia.

Esta versión del coeficiente Alpha se conoce también como la "fórmula 20 de Kuder Richarson" (KR-20).

Por otra parte se recordará que el coeficiente de confiabilidad de cualquier prueba es la correlación promedio estimada de esa prueba con todas las posibles pruebas de la misma longitud cuando se propone que las dos pruebas miden la misma cosa. El coeficiente Alpha también se puede derivar como la correlación que se espera entre una prueba real y una forma hipotética alternativa. Si llamamos X a la prueba real y "Y" a la prueba hipotética, entonces la matriz de la varianza total para todos los reactivos se puede esquematizar como se ve en la fig. 11.3.

	<b>X</b>	<b>Y</b>
	Cx	Cxy
<b>X</b>		
	Cxy	Cy
<b>Y</b>		

Figura 11.3: Matriz de la varianza total.

A partir del modelo dominio-muestra se espera que el término diagonal promedio en Cx sea el mismo que en Cy y que el promedio de los elementos fuera de la diagonal en las dos matrices sea el mismo También se espera que el elemento promedio a lo largo de Cxy sea igual al promedio del elemento fuera de la diagonal en Cx. Por lo

tanto, se puede derivar el coeficiente Alpha a partir de la correlación de suma, como sigue:

$$r_{xy} = \frac{\bar{C}_{xy}}{\sqrt{\bar{C}_x} \sqrt{\bar{C}_y}}$$

De acuerdo con el modelo,  $\bar{C}_x$  es aproximadamente igual a  $\bar{C}_y$ , de manera tal que la ecuación anterior se puede volver a escribir como sigue:

$$r_{xy} = \frac{\bar{C}_{xy}}{\bar{C}_x}$$

De acuerdo con el modelo, el coeficiente promedio en  $C_{xy}$  (y así la suma de los coeficientes) se puede derivar de  $C_x$ . Primero sería necesario restar de  $C_x$  las varianzas de los reactivos que se encuentran en la diagonal. Después sería necesario inflar el resultado por el factor desarrollado previamente, es decir,  $K/(K-1)$ , lo que nos lleva de nuevo al coeficiente alpha.

### 11.2.5.3 Análisis de Reactivos

El análisis de reactivos es otro procedimiento que se sigue en la búsqueda de consistencia interna. Es decir, cualquier operación que implique un análisis de la varianza de los elementos componentes de un instrumento, proporciona una estimación de la consistencia interna. Un procedimiento general y un tanto diferente al de al Alpha de Cronbach es el propuesto por Cureton (1966), que permite seleccionar reactivos que muestran correlaciones más altas con el resto de los mismos, y por lo tanto, los que se correlacionan más alto con la puntuación total. Una ventaja de este procedimiento es que es aplicable a reactivos dicotómicos o múltiples, además de que en el cálculo de las correlaciones también se prevé, la corrección necesaria que elimina el falso incremento del valor de los coeficientes que se produce cuando se incluye en el puntaje total el reactivo que se está analizando.

### REFERENCIAS

- Díaz Guerrero, R. y Salas, M. (1975). **El Diferencial Semántico del Idioma Español**. México: Editorial Trillas.
- Edwards, A.L. (1957). **Techniques of Attitude Scale Construction**. Nueva York: Appleton-Century-Crofts.
- Gulliksen, H. (1950). **Theory of Mental Tests**. Nueva York: John Wiley and Sons.

- Gulliksen, H.(1950b). The reliability of speeded tests. **Psychometrika**, 15, 259-269.
- Guttman, L.(1944). A basis for scaling qualitative data. **American Sociological Review**, 9, 139-150.
- Guttman, L. (1945). **Questions and answers about scale analysis**. Research Branch, Information and Education Division, Army Service Forces. Report D-2.
- Guttman, L.(1946). An approach for quantifying paired comparisons and rank order. **Annals of Mathematical Statistics**, 17, 144-163.
- Guttman, L. (1947a). Suggestions for further research in scale and intensity analysis of attitudes and opinions. **International Journal of Opinion and Attitude Research**, 1, 30-35.
- Guttman, L. (1947b). The Cornell Technique for scale and intensity analysis. **Educational Psychological Measurement**, 7, 247-280.
- Kelley, T.L.(1921). The reliability of test scores. **Journal of Educational Research**, 3, 370-379.
- Likert, R.(1932). A technique for the measurement of attitudes. **Archives of Psychology**. No. 140.
- Lincoln, E.A.(1932). The unreliability of reliability coefficients. **Archives of Psychology**, 140.
- Lincoln, E.A.(1933). Reliability coefficients are still unreliable. **Journal of Educational Psychology**, 24, 235-236.
- Nunnally, J.C. y Bernstein, I.J.T. (1993). **Teoría Psicométrica**. México: McGraw Hill.
- Osgood, Ch.E., Suci, G.J., Tannenbaum, P.H.(1957). **The Measurement of Meaning**. Urbana: University of Illinois Press.
- Ottis, A.S. y Knollin, H.E.(1921). The reliability of the Binet Scale and of pedagogical scales. **Journal of Educational Research**, 4, 121-142.
- Paulsen, G.(1931). A coefficient of trait variability. **Psychological Bulletin**, 28, 218-219.
- Preston, M.G.(1940). Psychophysical measurement methods. **Psychological Bulletin**, 35, 63-83.

- Rulon, P.J. (1939). A simplified procedure for determining the reliability of a test by split-halves. **Harvard Educational Review**.9, 99-103.
- Spearman, Ch.(1904a). The proof and measurement of association between two things. **American Journal of Psychology**, 15, 72-101.
- Spearman, Ch.(1904b). "General Intelligence" objectively determined and measured. **American Journal of Psychology**, 15, 201-295.
- Spearman, Ch.(1907). Demonstration of formulae for true measurement of correlation. **American Journal of Psychology**, 18, 161-169.
- Spearman, Ch. (1910). Correlations calculated with faulty data. **British Journal of Psychology**, 3, 271-295.
- Spearman, Ch. (1913). Correlations of sums and differences. **British Journal of Psychology**, 5, 417-426.
- Thouless,R.H.(1936). Test unreliability and function fluctuation. **British Journal of Psychology**, 26, 325-343.
- Thouless,R.H. (1939). The effects of errors of measurement on correlation coefficients. **British Journal of Psychology**, 29, 383-403.
- Woodrow, H.(1932). Quotidian variability. **Psychological Review**, 39, 245-256.

## **XII CARACTERISTICAS PSICOMETRICAS DE LOS INSTRUMENTOS: VALIDEZ.**

Después que se ha escogido un modelo para la construcción de un instrumento de medición y éste ya se construyó, el siguiente paso es determinar si el instrumento es útil o no. A este paso se le conoce como determinación de la validez del instrumento.

En un sentido muy general, un instrumento de medición es válido si mide lo que se intenta medir, o lo que se propone medir. La validación siempre requiere investigación empírica; la naturaleza de la evidencia que se requiere depende del tipo de validez. La validez es una cuestión de grado, más que una propiedad de todo o nada, y la validación es un proceso que no termina nunca; la mayoría de los instrumentos de medida deberían ser investigados constantemente, para determinar si se están comportando como debieran. La evidencia nueva puede sugerir modificaciones a un instrumento ya existente o el desarrollo de una aproximación nueva y mejor para medir el atributo en cuestión.

Hablando en forma estricta, uno no valida un instrumento o variable de medición, sino más bien el uso que se hizo de ese instrumento. Por ejemplo, una medición válida de la respuesta a unos tratamientos experimentales productores de angustia no necesariamente puede ser una medición válida de neuroticismo o de alguna otra cosa.

Aunque un instrumento de medida puede ser válido para diversos propósitos, la validez con la que sirve a cada propósito debe verse apoyada por evidencia empírica.

Los instrumentos de medición en psicología tienen por lo general tres propósitos fundamentales: a) establecer una relación funcional con una variable particular; b) representar a un universo de contenido específico; c) medir rasgos psicológicos. En forma correspondiente, existen tres tipos de validez: a) validez predictiva; b) validez de contenido; c) validez de construcción (Nunnally y Bernstein, 1993). Otro autores como Anastasi (1961) y Cronbach (1960), también hablan de una validez concurrente.

### **12.1 Validez Predictiva.**

La validez predictiva entra en juego cuando el propósito es usar un instrumento para estimar alguna forma importante de conducta, y nos referimos entonces, a ésta última como el **criterio**. Un ejemplo sería el de una prueba de selección de estudiantes que se empleara para predecir el éxito académico. En este caso, probablemente el criterio serían las calificaciones obtenidas en el ciclo escolar que se pretende predecir. Una vez que se obtiene el criterio, la

validez de una función predictiva es muy sencilla. Consiste principalmente en correlacionar las calificaciones de la prueba predictora con las calificaciones de la variable criterio. El tamaño o magnitud de la correlación es una indicación directa de la magnitud de la validez.

El término "predicción" se empleará en un sentido muy general (y no gramatical) para referirse a las relaciones funcionales entre las calificaciones de un instrumento, con eventos que ocurren antes, durante o después de que fue aplicado el instrumento. De esta manera una prueba aplicada a adultos podría emplearse para hacer "predicciones" acerca de eventos que ocurrieran durante su niñez. Una prueba que tiene la intención de "predecir" daño cerebral, por supuesto que no intenta predecir en el sentido estricto de la palabra, a quien va a surgir daño cerebral en el futuro, sino más bien, intenta predecir quien tiene daño cerebral en el momento en que se administra la prueba.

Cuando se usa una prueba para predecir el éxito académico, el término "predicción" sí significa hablar de eventos futuros. Otros autores, se han referido a la validez predictiva en estos tres puntos en el tiempo, como "post-dicción", "validez concurrente" y "predicción", respectivamente. El emplear diferentes términos sugiere que la lógica y los procedimientos de validación son diferentes, lo cual no es cierto. En cada caso se relaciona una media predictora con una media de criterio, y después de que se tienen disponibles los datos, no importa cuando fueron obtenidos.

La validez predictiva está determinada sólo por el grado de correspondencia entre las dos mediciones involucradas. Si la correlación es alta, no se requiere de otros estándares.

Mientras que es muy fácil hablar de correlacionar una prueba predictora con su criterio, en la realidad, obtener un buen criterio puede ser más difícil que obtener una prueba predictora. En muchas ocasiones, o no se tiene a la disposición al criterio, o los criterios que se tienen sufren de diversas fallas.

### **12.1.2 Coeficiente de Validez**

La validez de los instrumentos predictores individuales y de la combinación de instrumentos predictores queda determinada por el análisis correlacional y extensiones de éstos a análisis multivariados. El ejemplo más simple de un coeficiente de validez es la correlación de una prueba predictora individual con un criterio individual.

La forma adecuada de interpretar un coeficiente de validez es en

términos del grado en el que indica una posible mejora en la cualidad promedio de las personas que se podría obtener empleando el instrumento en cuestión. Esto quiere decir que en promedio, las personas que puntúan alto en la prueba tienen ejecuciones considerablemente mejores que aquellas que puntúan bajo.

Sin embargo, un indicador más adecuado que una simple correlación, es llevar a cabo una regresión ya sea simple (para instrumentos predictores individuales) o múltiple (para instrumentos predictores combinados), ya que uno de los usos específicos para los que se emplea la regresión es el de la predicción. Si el análisis de regresión es estadísticamente significativo y los coeficientes obtenidos por los predictores también lo son, el análisis de regresión permite predecir puntajes individuales en la variable criterio, con cierto intervalo de confianza establecido por el investigador.

El análisis de regresión simple se emplearía cuando se tuvieran un instrumento predictor (o prueba) que midiera una variable determinada, que supuestamente es un buen predictor del criterio, el que también se tendría que medir. El análisis de regresión múltiple se emplearía cuando se parte de la suposición de que el criterio queda mejor predicho por varias variables o predictores, para las cuales se han construido instrumentos o pruebas que pretenden medirlas. En ambos casos, la variable dependiente es el criterio y las independientes predictoras es o son los instrumentos o pruebas elaboradas con la finalidad de predecir el criterio.

### **12.1.3 Comparación de grupos**

Un método práctico y sencillo que frecuentemente se utiliza en la prueba de la validez predictiva es el que se conoce como el método de comparación de grupos. La comparación de grupos puede realizarse siguiendo diferentes procedimientos de grupos a comparar: partiendo de un control estadístico o a partir de un control experimental.

### **12.1.4 Comparación de la Distribución de la variable Predictiva y un arreglo de la Distribución de la Variable Criterio.**

La comparación de las distribuciones de las variables predictivas y criterio forma parte de los procedimientos que determinan los grupos siguiendo un control estadístico. Para comparar las distribuciones se procede como sigue:

-La distribución de puntajes de la variable predictiva se divide, teniendo como punto de referencia el lugar más próximo a la mediana, de tal manera, que se tengan dos grupos: uno con los

puntajes más altos y el otro con los puntajes más bajos;

-de la distribución de los puntajes de la variable criterio se forma un grupo con los puntajes superiores. En este caso, puntajes superiores no es necesariamente sinónimo de puntajes altos: un puntaje bajo en la variable criterio puede considerarse superior en tanto es lo que predice el instrumento;

-se calcula en cada uno de los grupos de la variable predictiva el porcentaje de sujetos que tuvieron puntuaciones superiores, es decir que cayeron dentro del grupo criterial.

-se prueba si existe diferencias estadísticamente significativa entre los porcentajes. Lawshe y Barker (1950) desarrollan una monograma que permite utilizar la "t" para la prueba de los porcentajes, pero se puede utilizar cualquier prueba de estadística no-paramétrica. La prueba "t" tradicional es correcta si no se usan porcentajes.

Si la prueba estadística arroja una diferencia significativa entre los porcentajes, este resultado se interpreta como confirmación de la validez del instrumento, en tanto, la relación que existe entre ambas variables, la predictiva y la criterio. Sin embargo, no proporciona información con respecto a la magnitud de esa relación y que se requiere como un índice de seguridad para la utilización práctica del instrumento.

Para efectos de evaluar la magnitud de la relación que se ha probado a través de la comparación de las distribuciones de la variable criterio y la predictiva se procede a estimar el porcentaje de superposición o traslape que existe entre los dos grupos de la variable predictiva tomando en cuenta el grupo criterial. En términos generales, se considera que la magnitud de la relación entre las variables predictiva y criterio será mayor, entre menor sea el porcentaje de superposición. En relación a este tema y en general con los procedimientos de comparación de esta clase, consultar a Guion, (1965).

## **12.2 Comparación con Grupos Conocidos**

Dentro del método de comparación de grupos, se puede utilizar para realizar dicha comparación, grupos que se denominan "grupos conocidos" debido a que el investigador los formará con base en el grado en que los sujetos poseen el atributo que se está midiendo, - de aquí que se diga que los grupos se forman a partir de un control experimental-. Por lo general, se forman dos grupos extremos: uno por sujetos que poseen el atributo en alto grado y otro en donde los sujetos muestren el atributo en cuestión en grado mínimo.

Aun cuando este procedimiento es operativamente diferente al



anterior, puede verse que en ambos los puntajes de la variable predictiva se comparan con los de la variable criterio; en el caso de los grupos conocidos los puntajes de la predictiva se comparan con dos condiciones diferentes de la variable criterio; en el anterior, dos condiciones de la variable predictiva se comparan con una condición de la criterio.

Los puntajes obtenidos de la aplicación del instrumentos, -de la variable predictiva-, a los grupos criterios extremos, se someten a una prueba de las diferencias significativas, -con la prueba t-, o a un índice de correlación, -el biserial-. La interpretación biserial está condicionada al cumplimiento de los requisitos que exige su adecuada utilización. (Ver al respecto Dawes 1962,1972).

En el tratamiento estadístico de la validez predictiva los procedimientos de la regresión,- especialmente la regresión múltiple- resultan de gran utilidad porque permiten estudiar relaciones múltiples, entre más de dos variables.

### 12.3 Validez de Contenido

Para algunos instrumentos, la validez depende principalmente de lo adecuado del muestreo de un dominio de contenido específico. Un ejemplo sería un examen final de un curso. El propósito de la prueba es medir en forma directa la ejecución o desempeño en una unidad de instrucción. La prueba debe sostenerse a sí misma como una medida adecuada de lo que se supone debe medir;. no se puede determinar la validez correlacionando la prueba con algún criterio, porque la prueba en sí misma es el criterio de ejecución.

Quizá sería más conveniente tratar de lograr la validez de este tipo de prueba por medio del plan y procedimiento de construcción que tratar de determinarlo después de que ha sido construida.

Las normas principales para asegurar la validez de contenido de un instrumento son a) tener un conjunto representativo de reactivos y b) métodos "sensatos" de construcción de pruebas. La primera se puede alcanzar **formulando** una colección de reactivos que representen en forma amplia el dominio del atributo ( o, en el caso de un examen para un curso), que representen a la unidad de instrucción. Para asegurar que los reactivos realmente representan el dominio, es necesario tener un bosquejo detallado de los tipos de reactivos (preguntas y/o problemas en el caso de un examen) que se incluirán. El simple muestreo al azar del contenido es poco realista porque la selección del contenido por lo general involucra cuestiones de valores que determinan el énfasis relativo

que se ponga en los diferentes tipos de contenidos.

La validez de contenido depende principalmente del razonamiento referido a qué tan adecuadamente se ha muestreado un contenido importante y qué tan adecuadamente se ha vertido este contenido en forma de reactivos.

Por otro lado, existe evidencia circunstancial que puede hablar de la validez de contenido de un instrumento dado. Se espera un nivel de consistencia interna por lo menos moderado entre los reactivos de una prueba; los reactivos deberían por lo menos medir algo en común. Al comparar la ejecución en una prueba antes y después de un periodo de entrenamiento, se debería manifestar un incremento en las puntuaciones en la segunda aplicación si es que la prueba o instrumento estuvo hecho para medir el progreso del entrenamiento; la mejoría en las calificaciones de los reactivos individuales se pueden considerar como evidencia de la validez de esos reactivos. Otro tipo de evidencia circunstancial puede ser la correlación entre diferentes pruebas que se proponen medir la misma cosa. Sin embargo, nada de lo anterior, garantiza la validez de contenido: ambas pruebas pueden medir las mismas cosas en forma equivocada.

#### **12.4 Validez de Construcción**

La validez predictiva es importante en los procesos de selección; la de contenido lo es para examinar la efectividad de diferentes métodos de entrenamiento. Al igual que las ciencias básicas, la ciencias sociales se preocupan por el establecimiento de relaciones funcionales entre variables importantes. Estas variables deben ser medidas antes de que puedan relacionarse. Antes de establecer relaciones que tengan significado, cada medida debe, en algún sentido, medir en forma válida lo que se propone medir. La validez predictiva y la de contenido, no son suficientes para el tipo de problemas a los que se enfrentan los científicos sociales.

El grado de dificultad que involucra la validación de variables está en proporción con el grado en el que la variable sea concreta o abstracta. Una variable muy concreta sería por ejemplo, el tiempo de reacción. La intención es medir una variable más o menos abstracta, y consecuentemente, la validez de cualquier medida particular está abierta a investigación. Esto es lo que sucede con la mayoría de las mediciones: representan esfuerzos por medir variables relativamente abstractas, variables que se supone se evidenciarán en una variedad de formas de conducta y de ninguna manera en forma perfecta en alguna de ellas.

Cuando una variable es abstracta, más que concreta, se conoce como

un **constructo**. Tal variable es literalmente una construcción del científico, ya que reúne algo a partir de su imaginación, algo que no existe como una dimensión del comportamiento, aislado y observable. Este constructo representa una hipótesis que señala que una variedad de conductas correlacionarán entre sí en los estudios de diferencias individuales, o se verán afectadas en forma semejante por los tratamientos experimentales.

Es importante darse cuenta de que las teorías en ciencia se refieren principalmente a afirmaciones acerca de constructos y no a variables específicas observables. Las teorías están pobladas de ellos y, los teóricos aún en conversaciones informales, les es difícil discutir su trabajo sin emplear palabras que se relacionan a constructos. Es importante recordar que no sólo las medidas propuestas para los constructos requieren ser validadas, sino que la ciencia se ocupa principalmente de desarrollar mediciones de los constructos y de encontrar relaciones funcionales entre las medidas de los diferentes constructos.

Los constructos varían ampliamente en el grado en el que el dominio de las variables observables relacionadas sea grande o pequeño y esté definido en forma rigurosa o no. Respecto a lo primero, en algunos casos el dominio de las variables relacionadas es tan pequeño que cualquiera de las pocas variables observables del dominio será suficiente para medir el constructo. Respecto a lo segundo, entre más grande sea el dominio de observables relacionadas con un constructo, más difícil tiende a ser definir qué variables pertenecen o no, a ese dominio. Por lo general, el científico sostiene una firme creencia acerca de las observables más predominantes relacionadas con el constructo, pero más allá de esto, sólo puede adivinar hasta dónde se extiende. Este es el caso de la mayoría de los constructos, los límites del dominio de observables relacionadas no es muy claro.

En virtud de que los constructos se refieren a dominios de observables, desde el punto de vista lógico, la mejor medición de cualquier constructo se podría obtener combinando los resultados de un número de mediciones de tales observables y comparándolos con cada uno de ellos en forma separada. Así, se puede pensar que cualquier medición en particular tiene un grado de validez de construcción que depende del grado en el que los resultados obtenidos al emplear la medición, serían semejantes si en alguna otra medición, o en forma hipotética, se hubieran empleado todas las mediciones del dominio en el experimento. En forma semejante, las calificaciones combinadas a partir de un número de mediciones de observables en el dominio, podría pensarse que tienen un grado de validez de construcción para el dominio como una unidad total.

Si la medición de los constructos es una parte vital de la actividad científica, ¿en qué forma se desarrollan y validan tales mediciones? Existen tres aspectos primordiales en este proceso: 1) especificar el dominio de observables; 2) determinar en qué grado todas, o algunas de estas observables correlacionan unas con otras o se ven afectadas en forma semejante por tratamientos experimentales; y 3) determinar si una o algunas de todas las mediciones de tales variables actúan como si midieran el constructo. El tercer aspecto consiste en determinar si la supuesta medición de un constructo correlaciona o no, de la manera esperada con otros constructos o si se ve afectada de la manera esperada por tratamientos experimentales particulares. El primer aspecto se refiere al esquema de observables del dominio relacionadas con el constructo, que tiene que especificar el investigador. Todos, o parte de tales esquemas del dominio habrán de someterse a investigación para determinar el grado en el que las variables del dominio tienden a medir la misma cosa, y esto es el segundo aspecto. El impacto de la teorización respecto al primer aspecto y los resultados de la investigación del segundo, tienden a influir sobre qué variables particulares se estudian en el tercer aspecto.

#### **12.4.1 Dominio de Observables**

El investigador debe esquematizar el dominio de observables antes de asumir que cualquiera de éstas se relaciona con el constructo. No existe ningún método preciso que pueda señalarse para esquematizar en forma apropiada el dominio de variables de un constructo.

La esquematización constituye esencialmente la forma en que las variables se relacionarán unas con otras. Esquematizar un constructo consiste esencialmente en establecer lo que quiera uno decir con palabras particulares. El esquema por lo general, consiste de una definición en la que la palabra que denota al constructo se relaciona con palabras de un nivel de abstracción menor. La importancia principal de este primer punto (esquematizar o delinear el dominio), es que le indica al investigador, qué hacer en el segundo punto (investigar la relación entre las diferentes medidas propuestas de un constructo).

#### **12.4.2 Relaciones entre observables.**

La forma en que se pone a prueba lo adecuado del esquema de un dominio referido a un constructo es determinar qué tan bien "van juntas" las mediciones de las observables a través de investigaciones empíricas. En los estudios de diferencias individuales, el primer paso es obtener calificaciones para una muestra de individuos en alguna de las medidas; después, cada

medida se correlaciona con las otras medidas. El análisis de las correlaciones resultantes proporciona evidencia acerca del grado con el que todas las mediciones tienden a medir lo mismo. Este es esencialmente un problema de análisis factorial.

En los experimentos controlados, lo que se investiga es el grado en el que las condiciones de tratamiento o experimentales tienen efectos semejantes sobre algunas de las medidas de las observables del dominio. La prueba de qué tan bien las diferentes medidas supuestas de las observables de un constructo "van juntas"; es el grado en el que tienen curvas de relación semejantes con una variedad de variables de tratamiento. No importa la forma de la relación con alguna variable o tratamiento particular siempre y cuando las relaciones supuestas del constructo se comporten de manera parecida. Si dos medidas se vieran afectadas exactamente en la misma forma por todos los posibles tratamientos experimentales, no importaría cuál de las dos se empleara en un experimento particular y consecuentemente uno podría hablar de ellas como midiendo la misma cosa. El grado en el que las dos medidas se vean afectadas en forma semejante por una variedad de tratamientos experimentales, en ese grado se puede decir que miden lo mismo. Cuando una variedad de mediciones se comportan en forma semejante a través de una variedad de tratamientos experimentales, tiene sentido hablar de medición de un constructo.

Los métodos para investigar la validez de construcción tanto en los estudios de diferencias individuales como en los experimentos controlados, involucran correlaciones. Las correlaciones se calculan entre las medidas de diferencias individuales. En el caso de los experimentos controlados, aunque en esencia no se pueden aplicar los métodos de correlación, la comparación entre las curvas obtenidas en las diferentes observables y el efecto de los tratamientos experimentales sobre ellas, es de hecho una correlación entre las curvas. Tales correlaciones proporcionan evidencia acerca de la estructura de un dominio de observables relacionadas con un constructo.

Los resultados de investigación como las anteriormente descritas llevaría a una de tres conclusiones. Si todas las medidas propuestas correlacionan altamente entre sí, se puede concluir que son medidas de la misma cosa. Si las medidas tienden a dividirse en agrupamientos en forma tal que los miembros de un agrupamiento correlacionan altamente entre sí y poco con los miembros de los otros agrupamientos, se puede concluir que se están midiendo diferentes cosas. Si las correlaciones entre las mediciones se acercan a cero, todas ellas miden diferentes cosas.

### **12.4.3 Relaciones entre constructos.**

En la sección anterior se discutieron formas para estudiar la validez de construcción en términos de la **consistencia interna**, es decir, diferentes medidas de un dominio tienden a proporcionar la misma información (tienden a correlacionar altamente entre sí y se ven afectadas en forma semejantes por diversos tratamientos experimentales). En el grado en el que los elementos de tal dominio presentan esta consistencia, se puede decir que se puede emplear **algún** constructo para dar cuenta de ella. Sin embargo, la consistencia es una condición **necesaria** pero no **suficiente** de validez de construcción.

Para poder demostrar la suposición de que es posible encontrar pruebas irrefutables de que un conjunto particular de variables miden a un constructo particular sería necesario recabar ciertas formas de evidencia. Antes que nada, es suficiente evidencia de la validez de construcción de un instrumento si las supuestas medidas del constructo (ya sea una sola o una combinación de observables) se comportan como se espera que lo hagan.

Primero vendría la prueba de consistencia interna de los elementos de dominio; después tendrían que hacerse varios estudios correlacionales y experimentos controlados. Existirá prueba de validez de construcción en el grado en el que las mediciones cubran las expectativas que se tengan de ellas.

Existe una falacia lógica obvia al proponer que tal evidencia es prueba de la validez de construcción. Para determinar la validez de construcción, la medición debe ajustarse a la teoría de la que parte el constructo; pero para usar esto como evidencia, es necesario **asumir** que la teoría es cierta.

Para poder asegurar que la teoría es cierta, ya que la evidencia que se encuentre es prueba de la validez de construcción, el investigador relaciona sus variables en situaciones donde está seguro de las suposiciones de las que se parte.

Desde un punto de vista estricto, la validez de construcción se refiere a una relación hipotética entre la supuesta medida de un constructo y una variable observable particular. Los estudios de validez de construcción deberían llevarse a cabo sólo cuando: 1) el dominio del "otro" constructo está bien definido, 2) cuando no se puede argumentar en contra de la suposición de la relación entre los dos constructos.

#### **12.4.4 Explicación de los constructos.**

Cuando se habla de validez de construcción, se hace referencia en

realidad a la **explicación de construcción**, con lo que se quiere decir: proceso de hacer explícita una palabra abstracta en términos de variables observables. Se debe cuidar de no caer en la concepción errónea de que los constructos tienen una **realidad objetiva** más allá de las observables particulares empleadas para medirlo.

Un conjunto de normas más rígidas para la "validez de construcción" empieza con la definición de un conjunto de medidas referidas a observables. Así, se podría decir que el conjunto A consiste de las medidas de observables particulares  $X_1, X_2, X_3$ , etc., y que el conjunto B de las observables particulares  $Y_1, Y_2, Y_3$ , etc. La validez de construcción consiste entonces de los siguientes pasos:

A través de una serie de estudios empíricos, se establecen una serie de afirmaciones probabilísticas entre las diferentes mediciones del conjunto A y lo mismo se hace para el conjunto B. Existen muchas formas de hacerlo, dependiendo de los tipos de estudio empíricos que se lleven a cabo y los tipos de afirmaciones probabilísticas que se creen son más significativas. El ejemplo más directo es aquel en el que se correlacionan una con otra, diferentes medidas de diferencias individuales.

Después de haber obtenido todas las posibles correlaciones entre las observables individuales, es posible entonces, deducir correlaciones entre diferentes combinaciones de variables del conjunto. También sería posible deducir la correlación entre cualquier medida particular en el conjunto, y la suma de todas las medidas que hubieran sido investigadas en éste.

Se puede hablar de toda la información obtenida de este tipo de estudios como aquellas que forman la estructura interna de los elementos de un conjunto. La estructura puede indicar que todas las variables miden más o menos lo mismo, lo que señalaría retener al conjunto tal y como se definió en un principio; o la estructura podría ser tal, que indicara que los miembros de un conjunto están midiendo dos o más cosas, y si éste fuera el caso, lo apropiado sería dividir el conjunto original en dos o más conjuntos (según fuera el caso) correspondientes a esas variables que realmente correlacionan bien unas con otras. Si todas las correlaciones entre los miembros de un conjunto son pequeñas, es ilógico continuar hablando de las variables como constituyendo un grupo, y el investigador debería dirigir su atención a otro conjunto de variables.

Cuando se ha logrado lo anterior, se conoce la consistencia

interna para los elementos del conjunto A. En forma semejante se determina la consistencia interna para los elementos del conjunto B. Supóngase que una variable particular en  $A_1, X_1$ , correlaciona con una variable particular en  $B_1, Y_1$ . Dependiendo del tamaño de la correlación sería posible hacer varios tipos de afirmaciones de probabilidad referidas a la correlaciones no conocidas entre cualquier otro de los miembros de A y cualquier otro de los miembros de B. De esta manera, existen estructuras internas para las variables del conjunto A y para el conjunto B por separado, y una trans-estructura entre las variable de los dos conjuntos. Si la estructura interna de cualquier conjunto es satisfactoria, le permite al investigador explorar trans-estructuras de ese conjunto con otros conjuntos. Si tales trans-estructuras son satisfactorias, puede decirse que se está progresando desde el punto de vista científico, ( se está poniendo a prueba teorías, o se están haciendo nuevos descubrimientos).

Se debe recordar que las palabras que emplean los científicos para denotar a sus constructos, no tienen contraparte real en el mundo observable. Tan sólo son instancias heurísticas para explorar observables. Aunque las palabras que se relacionan a los constructos son indiscutiblemente una buena ayuda para el científico, también lo pueden meter en problemas. Tales palabras son sólo símbolos que se refieren a colecciones de observables. Las denotaciones de una palabra no pueden ser más exactas que en el grado en que: 1) se especifiquen todas las posibles observables relacionadas, y 2) que todos los que empleen las palabras estén de acuerdo con la especificación.

Tomando en cuenta la inexactitud de las denotaciones de las palabras que se relacionan a los constructos, no es posible probar que cualquier colección de observables mide un constructo. Aunque en un sentido estricto, la "validez de construcción" es lógicamente imposible, existen formas de pruebas que demuestran algo muy parecido.

Hablando en un sentido estricto, los científicos nunca podrán estar seguros que se ha medido un constructo, o que se ha comprobado una teoría que se refiere a ese constructo, aunque sea útil hablar como si tal fuera el caso. Un constructo es sólo una palabra, y aunque la palabra puede sugerir exploraciones de la estructura interna de un conjunto o de variables interesantes, no existe forma de probar que cualquier combinación de esas variables realmente "miden" la palabra. Las teorías pueden sugerir investigaciones interesantes de trans-estructuras entre conjuntos de observables; la evidencia obtenida no es tanto prueba de la **verdad** de las teorías, como lo es de su **utilidad** como guías de la



realidad empírica. Llámese a la "medición" y a la "validación" de constructos como se quiera, éstos son: 1) conceptos que denotan constructos; 2) conjuntos de variables especificadas para tales constructos; 3) evidencia referida a la estructura interna de tales conjuntos; 4) palabras que se refieren a las relaciones entre constructos (teorías); 5) que sugieren **trans-estructuras** entre diferentes conjuntos de observables; 6) evidencia referida a tales **trans-estructuras**; y 7) más allá de esto, nada.

#### 12.4.5 Análisis Discriminante.

El análisis discriminante desarrollado por Fisher (1940), también resulta de gran utilidad para el proceso de la validación de constructos. A grandes rasgos, el análisis discriminante permite solucionar el problema del traslape o sobreposición que frecuentemente se da entre los puntajes de las variables que se están manejando. De esta manera, es posible ubicar con un grado mínimo de error, conjuntos de puntajes dentro de clases predeterminadas con propiedades ya definidas. En otros términos, es posible determinar a qué grupos pertenecen los sujetos medidos con base en los puntajes obtenidos. Es así como la validez discriminante se utiliza para conocer la magnitud de la relación.

Como primer paso, en el análisis discriminante, se obtienen las funciones discriminantes ( $D$ ) de cada grupo. El valor de estas funciones sirven de criterio para decidir a qué grupo pertenece cada uno de los sujetos. Para llevar a cabo esta ubicación con los puntajes de cada sujeto en cada variable en que fue medido, se calculan las funciones discriminantes individuales ( $d$ ). De esta manera, la pertenencia de cada individuo a un determinado grupo, depende de qué tan próximo es el valor de la  $d$  individual a la  $D$  del grupo.

Por otro lado, si la validez discriminante se desarrolla en términos de correlaciones, se esperarán correlaciones más bajas entre instrumentos y correlaciones más altas dentro de cada uno. Si los resultados se invierten, -a altas correlaciones inter-instrumentos y bajas correlaciones intra-instrumento-, entonces, se procede a un análisis de reactivos que permite una reubicación de reactivos en cada instrumento.

Campbell y Fiske (1959), han desarrollado un método de análisis denominado Multirasgo- Multimétodo que permite hacer comparaciones entre las confiabilidades de los instrumentos y las correlaciones inter-instrumentos. Proporciona una matriz con todas las intercorrelaciones resultantes cuando cada uno de los diferentes atributos es medido por distintos instrumentos. De esta manera, hace posible la detección de los efectos secundarios- debidos a la

cualidad de la medición utilizada; es decir los efectos del método; a las características individuales de respuesta: estilos de respuesta, ésto es, a la relación o contenido que se estudia y que se refleja en la varianza sistemática de la puntuaciones. (En relación a este tema, ver los trabajos de Muñiz, 1996; Scott, 1968; y Summer, 1976).

### **12.5 Validez de fase.**

La validez predictiva se ha llamado "validez empírica", "validez estadística"; a la validez de contenido se le ha llamado "validez intrínseca", "validez circular", "relevancia", y "representatividad"; y de la validez de construcción se ha hablado como "validez de rasgo" y "validez factorial".

En ocasiones el investigador se encuentra con el término de "validez de fase" que se refiere al grado en el que el instrumento "parece" que mide lo que intenta medir. La validez de fase se refiere a juicios acerca de un instrumento **después** de que fue construido. Como ya se dijo anteriormente, la forma apropiada de lograrla es a través del plan de contenido y del plan de construcción de los reactivos.

De esta manera, la validez de fase puede considerarse como un aspecto de la validez de contenido, que se refiere a una inspección del producto final para asegurarse que "nada salió mal" al transformar los planes en instrumentos terminados.

Con las funciones de predicción, la validez de fase es importante sólo en la formulación de hipótesis acerca de instrumentos que correlacionarán bien con sus criterios. Antes de que se lleve a cabo una investigación sobre un problema de predicción, debe existir alguna esperanza de que un instrumento particular funcionará. Tal esperanza se ve alentada si el instrumento parece que debería predecir el criterio.

También es cierto que las pruebas por lo general son mejores predictoras de un criterio, si su contenido de reactivos esta parafraseando en el lenguaje y términos de los objetos que realmente se encuentran presentes en el tipo particular de ejecución en cuestión. Por lo tanto, la validez de fase juega una parte importante en las decisiones acerca de los tipos de pruebas que se usarán como predictoras y en la construcción de los reactivos de esas pruebas.

### **12.6 Lugar que ocupa el Análisis Factorial.**

El análisis factorial está íntimamente relacionado con aspectos de

validez. Consiste esencialmente en un conjunto de métodos para descubrir agrupamientos de variables relacionadas. Cada uno de estos agrupamientos o factores, se denota por un grupo de variables cuyos miembros correlacionan más altamente entre ellos de lo que correlacionan con variables no incluidas en el agrupamiento. Se piensa en cada factor como un atributo unitario que va a ser medido en mayor o menor grado por instrumentos particulares, dependiendo del grado en el que estos correlacionen con el factor. Se habla de tales correlaciones como representando la **validez factorial** de las medidas. Quizá sería mejor hablar de tales correlaciones como de la **composición factorial** de las mediciones, debido a que la palabra "validez" puede llevar por un camino equivocado.

El análisis factorial es importante en la selección de los instrumentos que van a ser puestos a prueba como predictores. En lugar de construir un nuevo instrumento para cada problema aplicado que vaya surgiendo, el investigador selecciona un instrumento predictor de un "almacén" de instrumentos disponibles. El análisis factorial puede servir para construir tal almacén de mediciones con composiciones factoriales conocidas. En esta forma es mucho más fácil formular hipótesis acerca del poder predictivo de instrumentos desarrollados **ad hoc** para ese problema.

El análisis factorial proporciona evidencia circunstancial muy valiosa respecto a las mediciones que se intenta tengan validez de contenido. El análisis factorial se encuentra en el corazón de la medición de constructos psicológicos. Como se dijo anteriormente, la explicación de los constructos consiste principalmente en determinar: 1) la estructura estadística interna de un conjunto de variables que se dice que miden a un constructo y, 2) las **trans-estructuras** estadísticas entre las diferentes medidas de un constructo y aquellas de otros constructos. El análisis factorial se emplea directamente para determinar el punto (1), y algunos procedimientos relacionados con el análisis factorial son importantes para determinar el punto (2).

El análisis factorial juega un papel muy importante respecto a los tres tipos de validez, aunque en diferentes partes de cada uno de ellos. Respecto a la validez predictiva, es útil para sugerir predictores que funcionen bien en la práctica. Respecto a la validez de contenido, es importante para sugerir la forma en que habrán de mejorarse los instrumentos. Respecto a la validez de construcción proporciona las herramientas que son más útiles para determinar las estructuras internas y las trans-estructuras entre conjuntos de variables.

### **12.7 Relación entre Confiabilidad y Validez.**

Un coeficiente de confiabilidad alto implica, como se ha venido señalando a lo largo de este trabajo, un mayor control o eliminación de la varianza del error variable y en consecuencia una mayor proporción de varianza sistemática o verdadera, que por otra parte, es la única que se puede explicar y/o predecir. Así, dado que el trabajo de validación se efectúa sobre la varianza sistemática, es claro que el límite de ésta depende del grado de confiabilidad. De aquí la conocida afirmación: un instrumento nunca podrá predecir mejor un criterio que lo que "se predice" a sí mismo. En términos de Lord y Novick (1968, pag. 72), la validez, determinada a través de la relación con un criterio externo, "nunca puede exceder a la correlación de una puntuación observada con su puntuación verdadera).

El coeficiente de confiabilidad ( $r_{XX}$ ) representa una medida empírica; es la estimación obtenida por la correlación de dos distribuciones de puntajes. Para determinar la proporción de varianza verdadera que se encuentra en la varianza total se requiere de un coeficiente de determinación ( $r^2_{XX}$ ). Para entender el origen del coeficiente de determinación es necesario recurrir al índice de confiabilidad ( $r_{XX}$ ) que es un concepto teórico que simboliza una correlación de medidas hipotéticas con medidas observadas (x). Sobre estas bases, se interpretan los coeficientes de correlación: una correlación perfecta, (1.00) indicaría que la varianza total del instrumento únicamente reflejaría la varianza sistemática. Sin embargo, dado que una correlación perfecta sólo existe en la teoría, en la práctica es necesario determinar la proporción de la varianza total que se puede atribuir a la varianza sistemática o verdadera. Para este propósito, es posible estimar el índice de confiabilidad a partir del coeficiente de confiabilidad obtenido, y por lo tanto el coeficiente de determinación:

$$r_{XX} = r^2_{XX} \quad \text{ó} \quad \sqrt{r^2_{XX}} = r_{XX}$$

Si los coeficientes obtenidos de confiabilidad de dos instrumentos fueron:  $r_{XX} = 0.50$  y  $r_{XX} = 0.90$ , entonces, sus coeficientes de determinación respectivos son:  $r^2_{XX} = 0.25$  y  $r^2_{XX} = 0.81$ . Por lo tanto, en el primero el 25% y en el segundo, el 81% de la varianza total se debe a la varianza sistemática. Tenemos que la confiabilidad menor ( $r_{XX} = 0.50$ ), implica también, una menor proporción de varianza sistemática (25%) y con esto, también menor posibilidad de explicación o predictibilidad del origen de la varianza total, debido a la mayor intervención del error, (1.00-

0.25 = 0.75). Es decir, el 25% de la varianza total es atribuible a la varianza sistemática y el 75% a la varianza de error. En cambio, con una  $r_{xx} = .90$  el 81% se atribuye a la varianza verdadera y sólo el 19% se debe a la varianza de error.

Ahora bien, en relación a los coeficientes de validez, cuando se intercorrelacionan instrumentos o bien, instrumentos con criterios, se busca cuánto se relacionan sus varianzas verdaderas con las de los instrumentos entre sí o las de los instrumentos con las de los criterios y qué tanto se pueden predecir. Aquí se ve claramente expresada la relación confiabilidad- validez, ya que la varianza sistemática que interesa al proceso de validez ha quedado determinada por la confiabilidad  $r_{xy}$  no podrá exceder a  $r_{xx}$ , tal como se expresa en la fórmula de corrección por atenuación:

$$r_{x(v)y(v)} = \frac{r_{xy}}{\sqrt{r_{xx} r_{yy}}}$$

en donde;  $r_{x(v)y(v)}$  = a la correlación entre la varianza verdadera de X y Y.

$r_{xy}$  = al coeficiente de validez obtenido.

$r_{xx}$  y  $r_{yy}$  = a los coeficientes de confiabilidad de X y Y.

Por otro lado, la predicción que se basa en la correlación de los puntajes del instrumento, (variable predictiva) con los de la variable criterio, requiere por un lado, conocer la confiabilidad de esta última y por el otro, de una corrección, que se obtiene como sigue:

$$r_{xy(v)} = \frac{r_{xy}}{\sqrt{r_{yy}}}$$

esta fórmula proporciona el coeficiente de validez con la variable criterio ( $y_v$ ) corregida

Si el coeficiente de validez obtenido fuera  $r_{xy} = 0.45$  y el coeficiente de confiabilidad de la variable criterio :  $r_{yy} = 0.26$ , entonces, la correlación que se obtendría una vez corregida la variable criterio sería  $r_{xy(v)} = 0.88$ , casi el doble del valor  $r_{xy}$ . Comparando estos resultados en términos de coeficientes de

determinación tenemos:  $(0.88)^2 = 0.77$  y  $(0.45)^2 = 0.20$ . que muestran cómo la varianza en común (0.77) casi se cuadruplicó.

Es necesario señalar que la utilización de la fórmula de corrección por atenuación,- ya sea la total o únicamente en la variable criterio, presenta serias limitaciones cuando se desconocen los factores condicionantes de su interpretación ( En relación a este problema consultar Guilford, (1954) y Gulliksen (1950b). Así mismo, es aconsejable no perder de vista que la relación confiabilidad- validez es de naturaleza esencialmente matemática y en ella como lo hemos visto, existe una dependencia entre ambas. En cambio, en la práctica, de acuerdo con Campbell y Fiske, la validez es la correlación de medidas independientes, -lo contrario de lo que sucede con la confiabilidad-, y si bien esto no niega la relación entre confiabilidad-validez, sí previene de cometer un error, mal interpretando dicha relación, pensando que la validez de un instrumento- que va más allá de las operaciones de una fórmula, se puede obtener por una simple raíz cuadrada de su coeficiente de confiabilidad.

#### REFERENCIAS

- Anastasi, A. (1961): **Psychological Testing**. Nueva York: The MacMillan Company.
- Campbell, D.T. y Fiske, D.W.(1959): Convergent and discriminant validation by the multi-trait multi-method matrix. **Psychological Bulletin**, 56, 81-105.
- Cronbach, L.J.(1960): **Essentials of Psychological Testing**. Nueva York: Harper and Row Publishers Inc.
- Dawes, R.M.(1962): A note on base rates and psychometric efficiency. **Journal of Consulting Psychology**, 26, 422-424.
- Dawes, R.M. (1975): **Fundamentos y Técnicas de Medición de Actitudes**. México: Editorial Limusa.
- Fisher, R.A.(1936): The use of multiple measurements in taxonomic problems. **Annals of Eugenics**, 7, 179-188.
- Guilford, J.P.(1954): **Psychometric Methods**. Nueva York: McGraw Hill Book Company.
- Gulliksen, H.(1950): **Theory of Mental Tests**. Nueva York: John Wiley and Sons.

- Guion, R.M. (1965): **Personnel Testing**. Nueva York: McGraw Hill.
- Lawshe, C.H.(1942): A nomograph for estimating the validity of test items. **Journal of Applied Psychology**, 26, 846-849.
- Lord, F.M. y Novick, M.R. (1968): **Statistical Theories of Mental Test Scores**. Reading, Mass.: Addison-Wesley Publishing Company.
- Muñiz, J. (1996). **Psicometría** Madrid, España: Universitas
- Scott, W.A.(1968): Attitude Measurement. En G. Lindzey y E. Aronson(Eds.): **Handbook of Social Psychology**. Volumen II, Segunda Edición. Reading Mass.: Addison-Wesley Publishing Company.
- Summers, G.F.(1976): **Medición de Actitudes**. México: Editorial

### **XIII. EL REPORTE DE INVESTIGACIÓN**

El "último" paso en el proceso de investigación es el reporte de los resultados de la misma. Los formatos seguidos para hacerlo son diversos y dependen de las finalidades mismas de los estudios. En términos generales, se puede hablar de tres finalidades diferentes inherentes al proceso de investigación: la primera que incide sobre la intención de ampliar y difundir el conocimiento allegado por medio del proceso investigativo, entre los miembros de la comunidad científica particular de que se trate. La segunda, puede estar respondiendo a una necesidad planteada por una institución (pública o privada) que contrató al investigador para averiguar algo específico dentro del área de su especialidad, y los destinatarios de los hallazgos serán las instituciones y las personas que las dirigen. La tercera, se puede referir a la necesidad de culminar un proceso de educación formal, a cualquier nivel (licenciatura, maestría o doctorado), para lo cual se requiere como requisito parcial la presentación de una Tesis.

A continuación se presentan tres formatos generales para reportar investigaciones realizadas, ya sea para publicación en alguna revista científica de la especialidad correspondiente a la del investigador, para alguna institución que haya solicitado la realización del trabajo, o para alcanzar la culminación de un proceso formal de enseñanza en alguna institución educativa. Debemos aclarar que aunque existen ligeras modificaciones dependiendo de la institución de que se trate y del nivel educativo de las diferentes instituciones de educación superior, los formatos que presentamos, cubren los puntos esenciales de la posible diversidad de formatos existentes para la consecución de las finalidades planteadas anteriormente.

#### **13.1 REPORTE DE INVESTIGACIÓN PARA PUBLICACIÓN.**

El desarrollo de la ciencia depende no sólo del trabajo de investigación, sino también, entre otras cosas, de la publicación de los hallazgos y resultados de aquel.

Para que la publicación de lo encontrado en el proceso de investigación cumpla de manera óptima con sus funciones de difusión del conocimiento científico, es necesario que la elaboración del reporte cumpla con ciertos cánones y se ajuste a formas ( formatos) estandarizadas de presentación. Aun cuando estas últimas pueden variar relativamente según las normas establecidas por las asociaciones o grupos de investigación que se forman al interior de las diferentes disciplinas, las primeras,



las reglas, valen en general.

### **13.1.1. Elaboración del Reporte de Investigación: Reglas Principales**

El precepto modal alrededor del cual se agrupan las reglas de elaboración del reporte de investigación se compone de dos conceptos íntimamente relacionados aplicados a la comunicación o expresión verbal del estudio o experimentos que se reporta: *CONFIABILIDAD Y OBJETIVIDAD*.

La confiabilidad y objetividad son las características relevantes que deben cumplir la expresión verbal en un reporte científico. Si el propósito de confiabilidad lo describimos cómo lo que no es, entonces podemos decir que la confiabilidad es igual a la no transmisión de impresiones falsas, ambiguas, subjetivas o inferenciales. El propósito que busca una comunicación confiable es lograr que todos los lectores tengan **una misma interpretación básica**; de aquí que la confiabilidad en la comunicación del reporte requiera de la objetividad de la misma.

La objetividad en la comunicación de un reporte se refiere al isomorfismo o similitud que debe existir entre la expresión verbal y el trabajo de investigación en sí mismo. De esta manera una comunicación confiable es una comunicación objetiva o bien la objetividad en la comunicación da como resultado una comunicación confiable en tanto asegura una misma interpretación básica.

De los requisitos de confiabilidad y objetividad se derivan diferentes estrategias (reglas) relacionadas con ellas y encaminadas a su logro: v.gr., para disminuir el problema de diferentes interpretaciones de un mismo comunicado, es decir el problema semántico, se recomienda recurrir al uso del lenguaje operacional: describir en lugar de usar términos evaluativos; sustituir afirmaciones de juicio por hechos; y evitar expresiones con cargas emotivas.

Uno de los preceptos más útiles en el logro de reportes de investigación publicables se refiere a la regla de replicación que asegura que el autor proporcione la información adecuada suficiente para que otro investigador en cualquier parte del orbe pueda repetirla o replicarla.

Como puede verse, el requisito de replicación que debe cubrir todo reporte científico también se encuentra íntimamente relacionado con la confiabilidad y objetividad de la comunicación. Ya que se puede proporcionar la información necesaria para efectos de replicación, pero si no se da de manera adecuada (confiable y

objetiva) difícilmente se podría lograr.

### **13.2 Formato del Reporte de Investigación.**

Como ya se señaló, el investigador cuenta con diferentes formatos de publicación, sin embargo, la gran mayoría divide la información en seis rubros principales:

- 1.RESUMEN (ABSTRACT).
- 2.INTRODUCCIÓN
- 3.MÉTODO
- 4.RESULTADOS
- 5.DISCUSIÓN
- 6.REFERENCIAS BIBLIOGRÁFICAS.

El conjunto de los apartados o secciones arriba señalados tiene como propósito dar a conocer a la comunidad científica de un determinado campo del conocimiento, los hallazgos y resultados encontrados por los miembros de las mismas. Sin embargo, esos apartados también cumplen con propósitos específicos tendientes a facilitar la consulta de los lectores.

#### **13.2.1 Resumen**

El resumen o "abstract" debe proporcionar la información más importante, de tal manera, que la sola lectura del mismo permita decidir al lector si es de su interés y proceder o no a la lectura -total o parcial- del reporte.

En el resumen se redacta brevemente ( en aproximadamente 120 palabras) el problema de investigación, el diseño, y el o los resultados y la o las conclusiones más importantes; no se deben utilizar abreviaturas aún cuando en algunos formatos se hacen una excepción con respecto a la palabra sujeto o sujetos (S y Ss respectivamente). Esta restricción en el uso de abreviaciones se extiende a todos los apartados del reporte.

#### **13.2.2 Introducción**

La introducción puede ser de interés a otro investigador, por los antecedentes teóricos y prácticos y propósito del estudio que se reporta, aún cuando no de manera directa.

En la introducción se debe describir de manera clara y precisa el propósito de la investigación, la teoría o teorías en las que se apoya y los antecedentes más importantes -teorías y trabajos de investigación- con los que se relaciona.

En este apartado se justifica la importancia de la investigación, especificando su contribución -práctica, teórica o técnica- y se señala si el trabajo hace una aportación original, se aclara en qué consiste, o en caso contrario se explicita la razón que llevó a abordar un problema ampliamente estudiado.

Es importante cuidar no poner subtítulo a este apartado, ya que la posición o lugar que ocupa en el reporte indica por sí mismo, que es la introducción. Al relacionar el problema de investigación con teorías y trabajos antecedentes se deberá acreditar a los autores (autor o autores y año de publicación, esto último entre paréntesis).

### **13.2.3 Método**

El apartado dedicado al método puede ser consultado por otros investigadores total o tangencialmente interesados en el estudio que se reporta.

El problema que se debe resolver al escribir los contenidos de esta sección consiste en seleccionar la información que se va a reportar y en decidir cuál se va a omitir. Esta decisión debe ajustarse al criterio de replicación. Así, mientras que se pueden omitir los nombres de los sujetos que formaron parte del estudio, puede ser indispensable para efectos de que nuestro estudio se pueda repetir o replicar, reportar las edades, sexo, extracción social, nivel de educación, etc. De la misma manera puede ser que las instrucciones que se dieron a los sujetos no sean imprescindibles pero si lo sea el reporte del escenario y de los instrumentos que se utilizaron.

En la sección del método se reporta de manera adecuada la información necesaria para replicación sobre: el diseño de investigación (grupos independientes o dependientes; diseño correlacional, factorial, etc.); tipo de variables y su manejo; hipótesis, tipo de control de las variables extrañas; tipo de selección de las muestras (número, fuente y características específicas y relevantes de los sujetos); aspectos relevantes del procedimiento (aunque algunos formatos ubican lo referente al procedimiento en la sección de resultados), lo referente a los aparatos mecánicos o electrónicos (especificaciones de marca y modelo) o de otro tipo de instrumentos (estandarización, confiabilidad y validez).

### **13.2.4 Resultados**

Este apartado debe proporcionar, nuevamente tomando en cuenta el criterio de replicación, el tipo de datos y de análisis

estadísticos y los resultados, de tal manera, que otros investigadores puedan decidir la pertinencia de los mismos o aplicar otros análisis que consideren más apropiados (no es necesario proporcionar los datos crudos aunque sí se deberán poner a disposición de quien los solicite).

Esta sección debe contener una descripción detallada de los procedimientos de análisis utilizados, es decir, toda la manipulación estadística que se hizo de los datos y pruebas que se aplicaron, incluyendo representaciones gráficas - cuadros, tablas, etc. -indispensables. No se describen los modelos estadísticos y pruebas en sí mismos, a menos que se hayan hecho modificaciones especiales o sean muy recientes y por lo tanto poco conocidos. Se deben emplear la notación decimal y las abreviaturas correspondientes a las unidades físicas tipificadas (seg., db., etc.).

Si se utilizan gráficas (recuérdese que sólo deben ser las indispensables) debe tenerse cuidado de que sean claras, que estén debidamente rotuladas, y que por sí mismas expresen la información que se desea transmitir sin que se tenga la necesidad e recurrir al texto para comprenderlas.

### **13.2.5 Discusión**

Este es el único apartado en el que el autor expresa su propia interpretación de los resultados de su investigación. No sólo puede generalizar sus resultados- desde luego si su diseño se lo permite- más allá de los resultados específicos garantizados por los datos, sino también incluir afirmaciones de opinión sobre la base de las experiencias particulares y conocimiento ganado a través de la investigación.

Por otra parte, el lector puede o no estar de acuerdo con la interpretación del autor y llegar a sus propias conclusiones sobre los resultados del estudio, ya que forma parte del enriquecimiento de la ciencia ver los mismos datos de diferentes maneras y desde perspectivas distintas. Esta posibilidad de distintas conclusiones es la razón por la cual no se incluyen éstas en el mismo apartado de los resultados.

En esta sección se incluye información sobre la relación entre los resultados y el problema de investigación; la relación entre los resultados y los de otras investigaciones (las señaladas en la introducción); si se confirmaron o no las hipótesis y las implicaciones y generalización de los resultados.

También se especifican las limitaciones del estudio (que no es lo

mismo que dar una serie de excusas por una pobre planeación del mismo), y se hacen sugerencias para futuras investigaciones.

### **13.2.6 Referencias Bibliográficas**

Las referencias bibliográficas aparecen al final del reporte en orden alfabético. Aún cuando existen diferentes formas con respecto al orden en que se escriben los datos de las referencias, aquí expondremos la más usual: primero se escribe el apellido seguido por la primera letra del nombre del autor; en seguida se da la fecha (entre paréntesis); el título del libro, el lugar de la publicación (seguido de dos puntos) para finalizar con la editorial. Si se trata de un artículo se substituye el lugar de la publicación por el nombre de la revista o del **Journal** con sus respectivos datos de localización (volumen y/o número de la revista, y los números de inicio y término de las páginas del artículo). Si se trata del capítulo de un libro, después del nombre del capítulo, se señala que aparece **En** y se sigue con los nombres de los editores, pero ahora iniciando con la inicial del nombre y continuando con el apellido; se indica que los autores aquí mencionados son el editor (Ed.) ó editores (Eds.), del título del libro que aparece inmediatamente a continuación, seguido ahora sí por el lugar de la publicación y la casa editorial. Por último, se señalan los números de las páginas de inicio y terminación del capítulo en cuestión.

Los aspectos más importantes de la redacción de la bibliografía son: a) proporcionar los datos necesarios para que el lector localice fácilmente el material que se reporta y b) incluir únicamente las referencias citadas en el cuerpo del reporte (el error más frecuente es el omitir alguna de las referencias citadas o agregar otras que no se señalaron).

Por último, los datos que no deben olvidarse al elaborar un reporte es el título del estudio, el nombre del autor y la afiliación académica o dependencia a la que se encuentra adscrito. Se recomienda que el título sea corto pero que proporcione información clara respecto a lo que se investigó.

### **13.3 Reporte de Investigación aplicada Institucional**

Las características distintivas del reporte de investigación aplicada institucional son dos: su destinatario y su utilización. El primero es la firma o institución --pública o privada-- que contrata al investigador; el segundo se refiere a que la investigación que se realiza constituye una fuente de datos a partir de la cual se tomarán decisiones para la solución de problemas de carácter social.

Estos dos aspectos determinan cambios en el reporte de investigación tradicional, lo que no implica que por otra parte el proceso de investigación mismos sufra modificaciones, en tanto su desarrollo sigue ajustándose a los cánones que exige un trabajo científico.

Los cambios en el reporte son los siguientes: se elimina el resumen o **abstract**; los apartados de RESULTADOS Y DISCUSIÓN son los de mayor peso, y los dedicados a INTRODUCCIÓN Y MÉTODO son breves, de carácter sumario.

### **13.3.1 Introducción**

Esta debe contener una descripción breve del problema que se investigó y de los trabajos (resultados empíricos) importantes directamente relacionados con aquel. Por lo general el espacio dedicado a los fundamentos teóricos se eliminan o se reducen. Se especifican claramente los propósitos que tuvo la investigación.

### **13.3.2 Método**

En este apartado, también sumario, se da prioridad a todo lo relacionado con la muestra (diseño de muestreo, tamaño de características relevantes de los sujetos) y con los instrumentos (piloteo, coeficientes de confiabilidad y validez).

### **13.3.3 Resultados**

En esta sección se describen ampliamente, principalmente por medio de gráficas, cuadros y tablas los resultados. A diferencia del reporte de investigación para publicar, el uso de gráficas es de especial importancia, razón por la cual, éstas deben ser claras, de tal manera, que expresen los resultados a primera vista. La especificación de los procedimientos estadísticos, pruebas, etc., debe ser breve o bien, a veces se elimina.

### **13.3.4 Conclusiones**

En este apartado se deben completar de manera directa los resultados antes proporcionados. De hecho, se presenta la interpretación del investigador a manera de resúmenes de cada uno de los resultados. La importancia de este rubro reside en la especificación de los problemas encontrados (limitaciones), en la indicación de cómo pueden ser superados en la práctica (sugerencia) y en la anticipación que el investigador hace de los problemas que pueden presentarse (vg., en la entrevista; en la selección de las muestra; problemas de comunicación, etc.)

proporcionando al mismo tiempo, alternativas de solución.

Es importante señalar que los apartados dedicados a resultados y conclusiones se presentan, por lo general, con la ayuda de materiales didácticos de exposición colectiva o grupal (transparencias, cartulinas, videos, etc.) y que es el momento en el que el investigador debe aclarar dudas y dar solución a los problemas que surjan en las personas que harán uso de los resultados de la investigación.

### **13.4 Reporte de investigación de Tesis**

El objetivo de una tesis de investigación empírica es, en términos generales, reportar un estudio llevado a cabo por un alumno o un grupo de alumnos, bajo la asesoría de uno o varios profesores, con la idea de que será sometido a juicio ante un grupo de profesores o jurado calificador, que habrá de evaluar al o a los sustentantes en cuanto a diversos aspectos, entre los cuales destacan los siguientes: 1) El conocimiento teórico-metodológico adquirido durante una experiencia de enseñanza-aprendizaje formal. 2) La habilidad adquirida para extrapolar esos conocimientos a una situación práctica o de la realidad extracurricular. 3) La capacidad que el sustentante haya adquirido para establecer relaciones, semejanzas, diferencias e integraciones del conocimiento adquirido; 4) La habilidad para resolver los diferentes problemas surgidos durante el desarrollo del estudio, de acuerdo con el conocimiento que tiene y que se refleje en la bondad de las soluciones planteadas, y 5) La habilidad para crear, a partir del conocimiento existente, nuevas aproximaciones para el estudio y/o explicación de los fenómenos o problemas del área de conocimiento en cuestión. Por último, este tipo de reporte debe cumplir con una función didáctica, para otros alumnos o personas interesadas en el tema, que lo vayan a consultar.

Debido a las razones anteriormente mencionadas, la forma en que se reporta una tesis, es ligeramente diferente a las antes presentadas. Deberá recordarse, sin embargo, que las reglas generales presentadas al principio de esta sección, siguen siendo aplicables.

En términos generales, los apartados de un reporte de investigación tipo Tesis, son los siguientes:

- 1 INTRODUCCIÓN
- 2 ANTECEDENTES
- 3 MARCO TEÓRICO DE REFERENCIA
- 4 METODOLOGÍA

## 5 INTERPRETACIÓN Y DISCUSIÓN DE RESULTADOS

## 6 REFERENCIAS BIBLIOGRÁFICAS

Se debe aclarar que este formato es aplicable únicamente a aquellas tesis que involucran una investigación empírica de algún problema de carácter social. Este formato, a diferencia del que se usa para elaborar reportes para publicación, carece de resumen, y contiene dos apartados que anteriormente no aparecían: antecedentes y marco teórico de referencia.

**13.4.1 Introducción**

En esta parte se deberá establecer con toda claridad el o los objetivos de la investigación; los problemas que pretende resolver. La importancia teórico-metodológica y/o práctica que tiene el estudio, así como las implicaciones y posibilidades de aplicación de los resultados esperados.

También es pertinente señalar la forma en que está organizada la información que se presenta en el resto de los apartados del reporte.

**13.4.2 Antecedentes**

En esta parte habrán de reseñarse en forma crítica y analítica, todos aquellos estudios que tienen alguna relación con el que se está reportando. La relación de los estudios realizados por otros investigadores con el actual puede ser de varios tipos; a) la semejanza de los problemas planteados; b) la semejanza de las técnicas empleadas (aparatos, instrumentos, análisis de datos); c) la semejanza de la aproximación teórica; d) la semejanza en el tratamiento o inclusión de algunas variables; e) resultados de otros estudios que sirven de fundamentación para los resultados esperados en el presente estudio; etc.

Esta presentación deberá hacerse en forma integrada, secuencialmente organizada de manera tal que lleve al lector a los aspectos más importantes y relevantes de los estudios anteriores que estén directamente vinculados con el presente estudio.

**13.4.3 Marco Teórico de Referencia**

En este apartado se señalará la o las perspectivas teóricas que se emplearán como línea directriz en la aproximación del fenómeno que se estudia, así como el marco de referencia que servirá para interpretar y discutir los hallazgos de la investigación.



Las hipótesis conceptuales, así como las de trabajo que se plantearán en el siguiente apartado, deberán ser una consecuencia lógica y más o menos directa del marco de referencia empleado y los resultados de los estudios reseñados en la sección de antecedentes.

#### **13.4.4. Metodología**

En esta sección se habrán de señalar todos los pasos seguidos en el desarrollo de la investigación, hasta la consecución de los resultados, en forma tal, que el lector que desee replicarla, pueda hacerlo con base en lo establecido en esta sección. Esta sección está constituida por diversas subsecciones, y debido a su importancia para el logro del objetivo que se planteó al inicio de este apartado, los mencionaremos, aunque sea en forma breve, indicando lo que habrá de incluirse en cada una de ellas.

1.Planteamiento del problema.- Este deberá derivarse directamente de la información contenida en las dos secciones anteriores (antecedentes y Marco teórico de referencia)

2.Planteamiento de hipótesis.- Conceptuales y de trabajo

3.Variables.- Clasificación; definición ( conceptual y operacional); control de variables ( experimental y/o estadístico)

4.Diseño de Investigación

5.Muestra.- Procedimiento de selección y características más importantes de la misma, así como especificación de su tamaño.

6.Instrumentos.- Si son aparatos conocidos, modelos, marca y nombre; si son nuevos, descripción amplia de su estructura, funcionamiento y manejo. Si son cuestionarios y/o pruebas conocidas, sólo mencionados; si fueron elaborados expresamente para este estudio: su procedimiento de construcción, aplicación, calificación, reestructuración después del piloteo; características psicométricas (confiabilidad y validez); versión final del instrumento.

7.Análisis de datos.- Se habrá de señalar cuáles análisis se llevaron a cabo, para qué y porqué

8.Resultados.- Se habrán de presentar los resultados obtenidos, por medio de Cuadros, Tablas, Gráficas o Figuras, de manera que la información no se repita innecesariamente.

Se deberá hacer una breve descripción de los resultados, señalando si las hipótesis fueron o no confirmadas.

#### **13.4.5 Interpretación y Discusión de Resultados**

En esta sección se deberá interpretar los resultados obtenidos a la luz del marco teórico de referencia, y se deberá discutir, contrastándolos con resultados de los estudios reseñados en la sección de antecedentes, indicando con claridad cuándo concuerdan

con los obtenidos por otros autores y cuándo no, señalando para cada caso, la razones que el investigador tiene para explicar la concordancia o falta de ella. Lo reportado en esta sección incluye también las limitaciones del estudio, así como la sugerencia de futuras investigaciones. En términos generales, lo que se señaló para Discusión en el reporte para publicación, también se aplica aquí, pero en este caso, la discusión puede ampliarse bastante, pues no existe límite de espacio preestablecido por nadie.

#### **13.4.6 Referencias Bibliográficas**

En esta sección, se seguirán los mismos lineamientos establecidos en la sección correspondiente en el reporte de investigación para publicación.

Este tipo de reportes, puede tener al final, tantos apéndices como sean necesarios, para garantizar la adecuada posibilidad de replicación por otro investigador interesado. Es decir los apéndices pueden contener los instrumentos iniciales, y los definitivos; tablas de estadísticas descriptivas de la muestra muy amplias; diagramas de aparatos, etc.