

# CAPÍTULO 9

## ANÁLISIS DE DATOS

### INTRODUCCIÓN

**Y**a estamos cerca de la meta. Ya codificamos los datos (organizarlos en una matriz y archivarlos), ahora, procedamos a analizarlos. Recordemos que en el enfoque cualitativo el análisis es el resultado de las discusiones en el grupo y en el enfoque cuantitativo nos ayudamos de la estadística, pero de todas maneras, los dos enfoques pueden valerse de las dos formas según la necesidad real del análisis. Precisemos la ayuda estadística. En esta etapa avanzamos tres pasos:

1. La descripción de los datos, para lo cual nos valemos de la estadística descriptiva.
2. La relación entre variables, mediante la estadística inferencial.
3. Interpretación de resultados.

Antes de partir, recordemos lo siguiente: este análisis se puede hacer con enfoques cualitativo o cuantitativo, según el propósito del investigador al analizar su objeto de estudio. El enfoque cualitativo realiza el análisis a través de la discusión grupal y una de las técnicas más usadas es la historia de vida. El enfoque cuantitativo utiliza el análisis estadístico. Sin embargo, la experiencia nos enseña que los dos enfoques pueden complementarse, por ejemplo, el cualitativo puede utilizar cifras cuando el investigador lo estime necesario para relacionar datos sobre clase social, profesión, sexo, etc., y el cuantitativo puede valerse de la discusión grupal para profundizar el análisis de sus datos.

Según lo anterior, presentamos a continuación una síntesis del análisis estadístico por cuanto brinda luces al enfoque cuantitativo y al cualitativo. (sobre la historia de vida como técnica. (Ver 8.1.1.6.4).

Inicialmente, recordemos que los métodos de análisis estadístico son variados y que cada uno tiene su finalidad, por lo tanto, no debemos hacer más análisis de los necesarios pues la estadística es una ayuda, no un fin. Lo primero que hacemos es seleccionar la prueba estadística que necesitamos; luego, elaboramos y ejecutamos el programa en computadora para obtener el análisis de los datos.

Actualmente, el análisis de los datos se realiza por computadora. Las empresas tienen centros de investigación con sistemas de cómputo para archivar y analizar los datos. Por esta razón, en este trabajo nos detenemos en la interpretación de métodos cuantitativos (ver cuadro 8) y no en los procedimientos de cálculo. Quien esté interesado en este último puede consultar cualquier texto de estadística social.

## **9.1 DESCRIPCIÓN DE LOS DATOS**

El primer paso es describir los datos obtenidos para cada variable, esto se hace con la ayuda de la estadística descriptiva y... ¿Cómo describimos los datos obtenidos? Para esto seleccionamos la prueba estadística necesaria, por ejemplo, para la variable “altura” en un estudio hipotético sobre las características físicas de 100 estudiantes de la UPTC, lo podemos hacer por medio de la distribución de frecuencias y mediante las medidas de tendencia central. ¿Cómo se aplica la distribución de frecuencias?. Veámoslo:

### **9.1.1 La distribución de frecuencias:**

La frecuencia es el número de veces que aparece un elemento dado en una entidad concreta. La distribución es el agrupamiento de los datos, es un conjunto de puntuaciones ordenadas en sus respectivas categorías. Las frecuencias pueden ser:

- **ABSOLUTAS:** corresponden al número real de veces que aparece un elemento en un conjunto de datos.
  
- **RELATIVAS:** llamadas también proporcionales o porcentajes, son la proporción de casos respecto de un total en que aparece un elemento. Se hallan por medio de esta fórmula:

$$\text{Porcentaje} = \frac{N_c}{N_t} (100)$$

Donde:

$N_c$  = frecuencia absoluta.

$N_t$  = total de casos.

Retomando el ejemplo 2 de la Práctica 1 podemos establecer la siguiente frecuencia relativa para el número de estudiantes que manejan la página web, 40 de 596:

$$\text{Porcentaje} = \frac{40}{596} (100) = 6.7 \% (7\% \text{ aproximado})$$

– **ACUMULADAS**, son la suma de la frecuencia de cada categoría con el dato de la categoría siguiente; se puede acumular en números arábigos o en porcentajes (ver Tabla 7). Cada ciencia utiliza las tres clases o una sola, según su necesidad, por ejemplo, en lingüística se trabaja con los tres tipos pero a veces sólo se usan las relativas.

La distribución de frecuencias se puede representar gráficamente mediante **tablas** y **gráficos**. Recordemos que se debe hacer una distribución de frecuencias para cada variable.

#### 9.1.1.1 Las tablas:

Para elaborar una tabla ordenamos los elementos (datos psicológicos, contextuales o lingüísticos) en clases o categorías mutuamente excluyentes. Los datos que aparecen en las tablas corresponden a las categorías con sus respectivas frecuencias (es conveniente dar cuenta de las tres). Entre más variables se registre en una tabla, más compleja es ésta. A veces las categorías (clases) de las distribuciones de frecuencias son muy numerosas y hay necesidad de resumirlas, tal como se demuestra en la tabla siguiente que relaciona la altura de 100 estudiantes de la UPTC.



CATEGORÍAS	FRECUENCIA ABSOLUTAS	FRECUENCIA RELATIVAS	FRECUENCIA ACUMULADA
ALTURA (cm)	N° DE ESTUDIANTES		
150 - 160	13	13%	13
162 - 170	18	18%	31
171 - 180	42	42%	73
181 - 190	27	27%	100
TOTAL	100	100%	

**TABLA 7.** DISTRIBUCIÓN DE FRECUENCIAS HIPOTÉTICAS SOBRE LA VARIABLE ALTURA DE 100 ESTUDIANTES DE LA UPTC.

Un posible comentario sería: la altura más repetida está entre 171 y 180 centímetros que corresponde a un 42% de la muestra. Solamente el 13% de la muestra posee la estatura más baja (150-160 centímetros) y un 27% posee la estatura más alta (181-190 centímetros).

Las frecuencias acumuladas también pueden expresarse en porcentajes. Veamos un ejemplo de utilización de estas ayudas en un estudio sobre el uso de “tu” y “usted” en una comunidad española rural realizado por Francisco Moreno (1986).

INTERLOCUTORES	FRECUENCIAS RELATIVAS	
	Hombres	Mujeres
Ancianos	80%	100%
Desconocidos	100%	60%
Personas de respeto	40%	60%
Personas de edad	20%	10%
Personas con carrera	0%	0%

**TABLA 8.** FRECUENCIAS DEL USO DE “TU” Y “USTED”. Fuente: Moreno 1986.

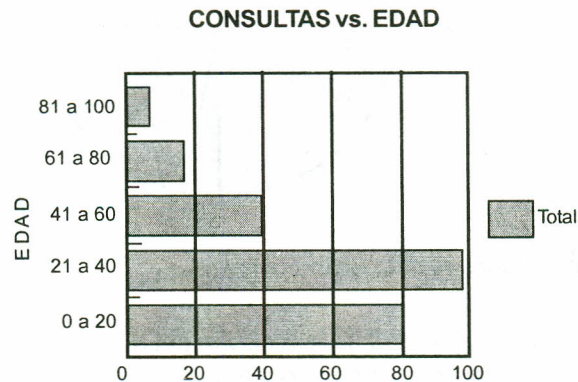


### 9.1.1.2 Los gráficos

Entre estos encontramos los diagramas de barras, los gráficos circulares, los histogramas y los polígonos de frecuencia. Su utilidad se basa en que permiten comparar de un vistazo frecuencias pertenecientes a categorías distintas pues, por lo menos, se representan dos parámetros (variables): uno, en el eje vertical y el otro, en el eje horizontal. Podemos representar gráficamente lo anterior, mediante la aplicación de los datos tomados de un estudio titulado “ Consultas Odontológicas vs. Sexo” realizado en la Escuela de Administración de Negocios por Olga Báez en el año 2000.

#### 9.1.1.2.1 Los diagramas de barras

Tienen el mismo valor y la misma significación estadística, la elección de uno u otro se basa en la facilidad para la comprensión del lector y en la habilidad del investigador para adaptarlos a su estudio.



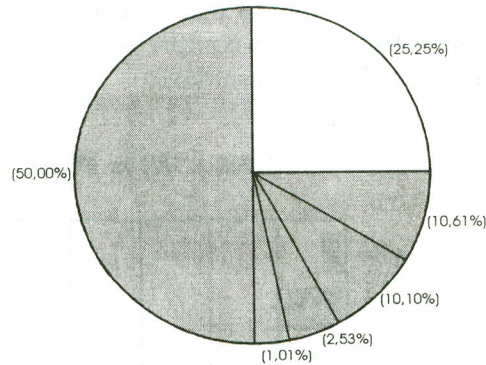
**GRÁFICO 6.** DIAGRAMA DE BARRAS. FUENTE: BÁEZ O. 2000.

Posible comentario: La población entre los 21 y 40 años realiza el mayor número de consultas y la población entre 81 y 100 años de edad la menor frecuencia de consulta. Los pacientes entre 41 y 60 años realizan un poco menos de la mitad del total de la consulta y los menores de 20 años un 80% del total de la consulta.

### 9.1.1.2.2 Los gráficos circulares.

Actualmente, con el uso de la computadora se dispone de gran variedad de programas y paquetes que elaboran cualquier gráfica, con colores, efectos de movimiento y en tercera dimensión. Siguiendo el ejemplo del trabajo de Báez, obtenemos el siguiente gráfico circular para la consulta odontológica realizada en el mes de enero de 2000, con los siguientes datos de frecuencias relativas: Urgencias, 50%; Control, 25.25%; Ortodoncia, 10,61%; Restauración, 10.10%; Exodoncia, 2.53%; Periodoncia, 1.01%.

#### CONSULTA Vs. ESPECIALIDAD



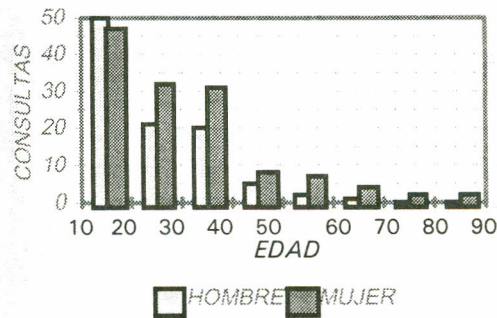
**GRÁFICO 7.** GRÁFICO CIRCULAR. FUENTE: BÁEZ O. 2000.

Posible comentario: la mitad de la consulta del mes correspondió a urgencias; la cuarta parte de la misma, obedeció a control, y solo una mínima parte fue solicitud por pacientes de periodoncia. Este resultado nos brinda un indicativo para incrementar el número de odontólogos generales.

### 9.1.1.2.3 Los histogramas.

Se utilizan cuando los parámetros (i.e. consulta- sexo) son escalas graduales cuantitativas. Los grados de las escalas (intervalos) deben estar dispuestos de tal forma que su valor cuantitativo sea inferior al de los grados superiores y superior al de los grados inferiores. Los contenidos de los histogramas se representan por curvas o combinando diversas técnicas: curvas y diagramas de barras. En el ejemplo Báez 2000, respondamos esta pregunta: ¿Cuántas consultas odontológicas realizaron las mujeres y cuántas los hombres de 35 años de edad, durante el mes de febrero de 2000?

## CONSULTA Vs. SEXO



**GRÁFICO 8.** CONSULTA ODONTOLÓGICA FEBRERO DE 2000.

Con base en el gráfico 8 podemos responder: mujeres = 30 consultas y hombres = 20 consultas.

### 9.1.1.2.4 El polígono de frecuencia

El Polígono de frecuencias relaciona las puntuaciones con sus frecuencias respectivas y es característico de la medición por intervalos. Se construye así: en el eje horizontal (X) ubicamos las categorías o intervalos y en el eje vertical (Y) colocamos las frecuencias y determinamos los puntos medios de cada categoría o intervalo, por ejemplo, en el intervalo 0 –20 el punto medio sería 10. Luego miramos cuántas frecuencias tiene cada categoría y trazamos un punto en la intersección de las frecuencias y los puntos medios de los intervalos. Finalmente, unimos los puntos trazados en las intersecciones. Retomando el ejemplo de Báez, se obtiene el siguiente polígono: para marzo, con la siguiente distribución:

CATEGORÍA O INTERVALO	FRECUENCIA	
	Hombres	Mujeres
0 – 10	47	47
11 – 20	10	80
21 – 30	20	30
31 – 40	70	60
41 – 50	20	80
51 – 60	10	10
61 – 70	60	20
71 – 80	25	60



## CONSULAS Vs. EDAD

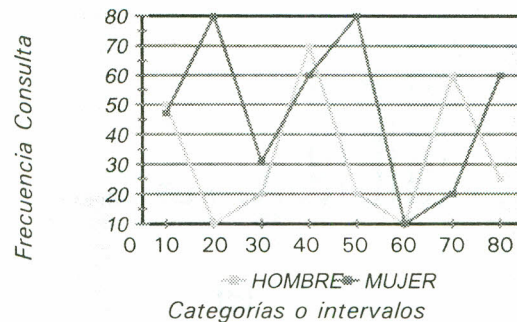


GRÁFICO 9. POLÍGONO DE FRECUENCIAS. FUENTE: BÁEZ 2000.

En conclusión, para describir y resumir cada una de las variables de una investigación se obtiene la distribución de frecuencias, se gráfica y se traza el polígono de frecuencias correspondiente. Para el análisis se puede emplear encabezamientos como: la mitad de los entrevistados; poco más de la mitad; casi la quinta parte; diez de cada cien; sólo uno de cada diez; etc.

### 9.1.2 Medidas de tendencia central.

Las medidas de tendencia central son puntos determinados en una distribución de datos, es decir, corresponden a los valores medios o centrales y ayudan a ubicar la medida dentro de la escala de medición. **La distribución**, es la clasificación tabular de los datos en clases con las respectivas frecuencias en las escalas establecidas. Las medidas de tendencia central más importantes son: La media, la mediana, la moda y la desviación típica o varianza.

#### 9.1.2.1 Media aritmética.

La media es el promedio aritmético de una distribución. Se representa

$\bar{X}$  (se lee equis barra), se halla sumando las frecuencias de una serie de elementos y dividiendo el resultado por el número de elementos estudiados. Este es el concepto que se aplica cuando se promedia las notas de un estudiante obtenidas durante un período académico para obtener la nota definitiva. Para evitar equivocaciones debe acompañarse de las medidas de variabilidad (de la desviación típica o de la varianza). Por ejemplo, un estudiante obtiene las

siguientes notas en el mes de mayo: 4.0, 3.0, 2.8. Para aplicar la media o promedio aplicamos la siguiente fórmula:

$$\bar{X} = \frac{\Sigma X}{N}$$

Donde :

$\Sigma$  = Señala que debemos realizar una sumatoria.

$\bar{X}$  = Representa los datos.

N = Representa el número total de casos.

Remplazamos.

$$\bar{X} = \frac{4.0+3.0+2.8}{3} = \frac{9.8}{3} = 3.26$$

La distribución de los datos a lo largo de una escala frecuentemente forma una figura simétrica representada por una curva llamada normal y que muestra acumulación de elementos en los estadios intermedios llamada curva normal o campana de Gauss.

#### 9.1.2.2 Mediana.

Es el punto intermedio de una escala de intervalos. Para hallar la mediana es necesario que los datos estén ordenados en forma ascendente o descendente. Su cálculo es útil para la interpretación de análisis porque da un punto de referencia o un dato típico para valorar el comportamiento cuantitativo de un grupo de elementos. Media y mediana no necesariamente coinciden, Por ejemplo, si los datos obtenidos fueran 4, 5, 6, 8 y 12. la mediana corresponde a 6 porque deja dos datos por debajo y dos por encima. La media corresponde a 7.

$$\bar{X} = \frac{4+5+6+8+12}{5} = \frac{35}{5} = 7 \quad \text{Mediana} = 6 \text{ (punto intermedio).}$$

Como vemos la mediana parte la distribución de los datos ordenados en dos mitades. Para hallarla aplicamos la fórmula.

$$\bar{X} = \frac{N+1}{2}$$

En el ejemplo anterior sería:

$$\text{Mediana} = \frac{5+1}{2} = 3$$

Esto indica que el tercer número de la secuencia corresponde a la mediana. La aplicación de la fórmula no da el valor de la mediana sino su ubicación.

### 9.1.2.3 Moda.

La moda es la categoría que se presenta con mayor frecuencia en un conjunto de datos, es decir, es la frecuencia más repetida y se utiliza con cualquier nivel de medición. En el ejemplo anterior equivale a 12 y en la tabla 7, la frecuencia absoluta es 42.

## 9.1.3 Medidas de variabilidad.

Las medidas de variabilidad indican la dispersión de los datos en la escala de medición e indican donde están diseminados los valores obtenidos; son intervalos y designan distancias o número de unidades en la escala de medición. Las principales son:

### 9.1.3.1 Varianza.

Dos poblaciones pueden presentar una media idéntica de una variable determinada pero, entre ellas puede haber diferencias importantes: una población puede ser homogénea, entonces, la variabilidad sería pequeña y en la población heterogénea, podría ser muy grande; es decir, en la heterogénea existirá mayor dispersión que en la homogénea, por ejemplo, hipotéticamente tenemos dos muestras de 10 hablantes, una de hombres y otra de mujeres tomada de la población de 90 estudiantes del primer semestre de idiomas de la UPTC. De esas muestras recogemos el número de emisiones del vocablo “gomelo” que aparece en diálogos con los estudiantes. En primer lugar, establecemos la siguiente tabla de frecuencias absolutas (X) ordenadas de mayor a menor. Al final de cada columna tenemos la sumatoria respectiva y, por último, la media de  $\bar{X}$ .



HOMBRES			MUJERES		
X	d	d <sup>2</sup>	X	d	d <sup>2</sup>
12	7	49	22	17	289
8	3	9	18	13	169
7	2	4	2	-3	9
6	1	1	2	-3	9
6	1	1	1	-4	16
4	-1	1	1	-4	16
3	-2	4	1	-4	16
2	-3	9	1	-4	16
1	-4	16	1	-4	16
1	-4	16	1	-4	16
$\Sigma_x = 50$	$\Sigma_d = 0$	$\Sigma_d^2 = 110$	$\Sigma_x = 50$	$\Sigma_d = 0$	$\Sigma_d^2 = 572$
X = 5			X = 5		

**TABLA 9.** FRECUENCIAS DE APARICIÓN HIPOTÉTICA DEL VOCABLO "GOMELO"

En la primera columna de los cuadros aparecen las frecuencias absolutas de producción de los individuos en cada muestra, la suma total ( $\Sigma_x$ ) y la media (X). Obsérvese que la media de pronunciación es de **5** en cada muestra. Esta coincidencia puede ser engañosa si no valoramos qué dispersión tiene el fenómeno entre hombres y mujeres.

Para calcular la dispersión nos valemos de la varianza (V) y de la desviación típica (S) que informan cómo se distribuyen los elementos alrededor de la media, es decir, se cuantifica la variación y la desviación de los datos respecto de la media.

El cálculo de la varianza consta de tres actividades:

- Ver la diferencia respecto de la media, para lo cual se realiza una resta ( $X - \bar{X}$ ). Para comprobar si los cálculos están bien hechos, realizadas las restas, se suman todos los resultados entre sí ( $\Sigma d$ ) y debe obtenerse cero (**0**) (ver la 2ª columna de cada muestra).

- Calcular el cuadrado de la diferencia obtenida en cada resta ( $d^2$ ) y sumar los resultados ( $\Sigma d^2$ ) (columna 3ª).
- Dividir  $\Sigma d^2$  entre el número X analizadas ( $n = 10$ ) menos 1 y se obtiene la varianza.

	<b>HOMBRES</b>	<b>MUJERES</b>
$V = \frac{\Sigma d^2}{n-1}$	$V = \frac{110}{10-1} = 12.2$	$V = \frac{572}{10-1} = 63.6$

Al comparar las varianzas observamos que la homogeneidad es diferente (hombres = 12.2 y mujeres = 66.6).

En la fase interpretativa se señalan los por qué.

#### 9.1.3.2. Desviación típica o estándar.

Sólo requiere aplicar la fórmula:

$S = \sqrt{V}$	$S = \sqrt{\frac{12.2}{3.5}}$	$S = \sqrt{\frac{63.6}{8}}$
	<b>HOMBRES</b>	<b>MUJERES</b>

Así queda clara la mayor homogeneidad de la primera respecto a la segunda.

- **la probabilidad.** Su fin es cuantificar la relación ente las características de una muestra y la respectiva población. En sociolingüística se aplica en el modelo de la regla variable. Cada reglas incluyen un coeficiente que cuantifica la probabilidad de aplicación de una regla según ciertas condiciones lingüísticas y sociales.

Los coeficientes se calculan a partir de las frecuencias de los materiales recogidos pues una probabilidad es una frecuencia relativa esperada. El modelo de regla variable o cálculo de coeficientes de probabilidad de una regla fue propuesta por David Sankoff y Pascale Rousseaw en 1978, su nombre es **modelo lógico** y se presenta así:

$$\frac{P}{1-P} = \frac{P_0}{1-P_0} \times \frac{P_a}{1-P_a} \times \frac{P_b}{1-P_b} \times \dots \times \frac{P_n}{1-P_n}$$

Donde **P** es probabilidad; **P<sub>0</sub>** es la probabilidad de que la regla se aplique en el contexto más favorable; **a, b, ...n** son cada uno de los rasgos del contexto en que puede darse la regla.

La aplicación del modelo lógico sobre materiales reales se hace a través de un programa de ordenador llamado **VARBRUL**. Tanto el modelo como el programa pertenecen a la estadística del análisis multivariable.

### 9.1.3.3 Rango o recorrido.

Es la diferencia entre la puntuación mayor y la puntuación menor, indica el número de unidades necesario para incluir el valor máximo y el mínimo. Se calcula así:

Puntuación Mayor menos Puntuación menor.

$$X_M - X_m$$

Por ejemplo, si se tienen los siguientes valores:

15 18 21 28 31 35 40

El rango será:  $40 - 15 = 25$ . Cuanto más grande sea el rango, mayor será la dispersión de los datos de una distribución..

## 9.2. ANÁLISIS ESTADÍSTICO INFERENCIAL.

El propósito de una investigación no solo es describir la distribución de las variables, sino generalizar los resultados alcanzados en la muestra a la población y por esto se da el nombre de estadística inferencial por que los datos aplicados a la población son inferidos de los datos obtenidos en la muestra.

El segundo fin del análisis estadístico es hacer estimaciones de **significación** y de **fiabilidad** mediante pruebas estadísticas: **análisis paramétrico** (análisis de varianza, t de student y análisis de covarianza); **análisis no paramétrico** (chi cuadrado, coeficiente de correlación); **análisis multivariado** (regresión múltiple, análisis lineal path, análisis de factores, análisis multivariado de varianza, análisis discriminante).



La estadística inferencial se utiliza para probar hipótesis y para estimar parámetros.

La prueba de la hipótesis consiste en que el investigador, a través de la prueba de hipótesis determina si esta es congruente con los datos obtenidos en la muestra o no. Si es congruente se acepta y si no se rechaza.

Cuando los datos han sido descritos y se han corregido valores estadísticos importantes como media y desviación típica, se puede, a partir de las frecuencias profundizar el estudio planteando un conjunto de hipótesis que los datos deben aceptar o rechazar ofreciendo probabilidades de error el más bajo posible. Las hipótesis pueden referirse a la existencia de diversas variables y plantearse en forma negativa o positiva.

La hipótesis negativa o nula consiste en presentar una hipótesis que debe ser rechazada al aplicar la estadística sobre los datos, de tal manera que se compruebe lo contrario, tal como se explicó en 5.4.3.2. Hacia la hipótesis nula están orientadas varias pruebas estadísticas: **varianza, T y  $X^2$** .

Al plantear o comprobar hipótesis es importante tener en cuenta la preparación y los tipos de datos que se van a comprobar. La formulación de la hipótesis no puede ser ambigua y los datos deben estar ordenados en escalas y ver si las variables son cuantitativas o cualitativas.

## **9.2 LA RELACIÓN ENTRE VARIABLES.**

Se realiza por medio de la estadística inferencial que comprende las siguientes técnicas:

### **9.2.1 Análisis paramétrico**

#### **9.2.1.1 Análisis de varianza (ANOVA).**

Se aplica cuando las variables dependientes están organizadas en escalas de intervalos. Su finalidad es comparar los medios y la forma como se distribuyen los datos alrededor de esos medios.

Hay dos tipos de **ANOVA**: el análisis **simple** y el análisis **compuesto**. El simple consiste en comprobar si muestras diferentes de una misma variable tienen comportamiento distinto, respecto de otra variable. El compuesto permite hacer lo mismo, pero atendiendo a más de una variable.

### 9.2.1.2 La prueba T de Student.

Su fin coincide con el del análisis simple anterior. Para aplicar la prueba T necesitamos contar con dos variables, ejemplo, sexo y abertura de la vocal final tras la caída de una consonante implosiva final, así la primera se cuantifica, sería la variable independiente y la segunda la dependiente (lingüística). Aplicando la prueba T se averigua si la conducta de hombres y mujeres es significativamente distinta en relación con la abertura de la vocal final. La diferencia entre las muestras se da por el cálculo de F (F.—) dividiendo la varianza más grande entre las más pequeñas. La prueba T dice si la diferencia se debe al azar o no. Para calcular el valor de t se aplica:

$$T = \frac{X_1 - X_2}{\sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}}$$

**n** = tamaño de cada muestra.

(1 y 2)

**S** = Desviación típica.

**X** = Media.

La diferencia entre la prueba t y el análisis simple está en que t solo trabaja con 2 muestras y el simple con más, siempre que sena estados distintos de una misma variable (variantes) ejemplo, edad ~ 20; 21 -30; 31 - 40. La ventaja del análisis compuesto de la varianza es que maneja más de una variable (sexo, edad) dentro de las cuales se distinguen niveles, con los cuales se trabaja realmente.

Esta prueba se llama así porque su descubridor Cosset publicó trabajos con el seudónimo **Student** a principios del siglo XX.

### 9.2.1.3 Análisis de covarianza (ANCOVA).

Analiza la relación entre una variable dependiente y dos o más independientes, eliminando y controlando el efecto de por lo menos una de las variables independientes.

## 9.2.2 Análisis no paramétrico.

### 9.2.2.1 Chi cuadrado ( $\chi^2$ ).

Su objeto es comprobar si la distribución de dos datos variables es independiente. Su cálculo trabaja con valores esperados teóricamente (e) y en valores observados (o). Su formula es:

$$\chi^2 = \sum \frac{(o - e)^2}{e}$$

$\chi$  corresponde a la letra griega "Chi".

Esta prueba se puede aplicar sobre cualquier número de variable y trabaja con frecuencias absolutas. El cálculo de los valores esperados se hace sobre la tabla de frecuencias observadas y ordenadas en líneas (horizontal) y columnas (vertical) ejemplo:

	Columna A	Columna B	Columna C	Total
Línea 1	62	58	60	180
Línea 2	108	96	102	306
Línea 3	46	31	49	126
Línea 4	33	27	37	97
TOTAL	249	212	248	709

**TABLA 10 . FRECUENCIAS OBSERVADAS Y ORDENADAS EN LÍNEA.**

Los valores esperados de las frecuencias de cada columna se obtienen por medio de una regla de tres sobre los totales:

$$\begin{array}{l} \text{total absoluto (709)} \quad \text{-----} \quad \text{total línea (180)} \\ \text{total columna A (249)} \quad \text{-----} \quad \text{(e) línea 1 Col. A} \end{array}$$

$$e = \frac{249 \times 180}{709} = 63.2$$

**Respuesta:** El valor observado 62 (línea 1 col A) le corresponde un valor teórico o esperado de 63.2.

#### 9.2.2.2 Coeficiente de correlación.

También se aplican los conceptos de **correlaciones, covarianza, regresión** y el **análisis multivariable** en investigaciones de tipo cualitativo.



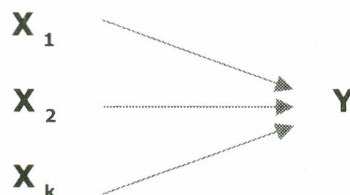
En la aplicación de estas técnicas tiene gran importancia la informática la cual tienen programas que facilitan el cálculo estadístico. Ejemplo: **SYMPHONY** u hoja de cálculo y gráficos; **STATGRAPHICS Y SPSS**. Cada día se hace más fácil su aplicación y los resultados serán más fiables. Ver más información en 9.3.1.4.

### 9.2.3 Análisis multivariado

Analiza la relación entre variables independientes y al menos una dependiente. Es un método complejo que requiere el uso de computador para efectuar los cálculos. Los principales métodos son:

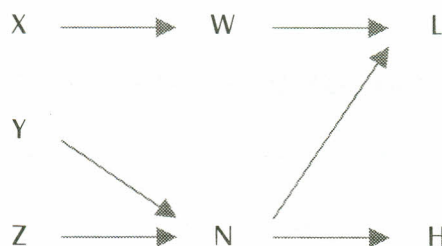
#### 9.2.3.1 Regresión múltiple

Analiza el efecto de dos o más variables independientes sobre una dependiente. Predice el valor de una variable dependiente conociendo el valor y la influencia de las variables independientes. Tiene especial aplicación en las hipótesis de tipo correlacional / causal.



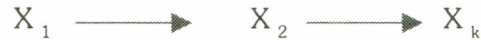
#### 9.2.3.2 Análisis lineal path.

Sirve para representar interrelaciones entre variables. Analiza la influencia directa o indirecta de unas variables sobre otras. Cuantifica efectos. Tiene aplicación, principalmente, en el tipo de hipótesis correlacional / causal.



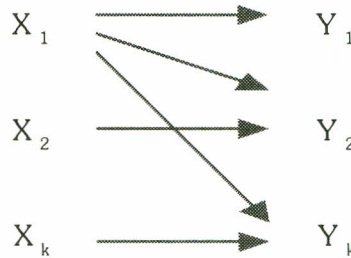
### 9.2.3.3 Análisis de factores.

Determina el número y naturaleza de un grupo de constructos implícito en un conjunto de mediciones, un constructo es un atributo que explica un fenómeno, es decir explica un fenómeno complejo en función de varias variables. Su especial aplicación está en la hipótesis correlacional / causal.



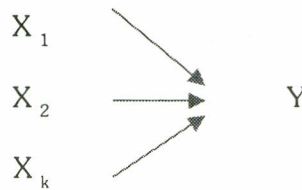
### 9.2.3.4 Análisis multivariado de varianza. (MANOVA).

Analiza la relación entre dos o más variables independientes y dos o más variables dependientes.



### 9.2.3.5 Análisis discriminante.

Se utiliza cuando las variables independientes son medidas por intervalos y la variable dependiente es una categoría. Sirve para predecir la pertenencia de un caso a una de las categorías de la variable dependiente sobre la base de varias independientes.



## 9.3 INTERPRETACIÓN DE RESULTADOS.

La interpretación o explicación de los hechos es la fase culminante del proceso de investigación, debe tener en cuenta los pasos anteriores, la finalidad del estudio y el análisis realizado. Se puede formular las siguientes reglas para la interpretación.

### **9.3.1 Finalidad del estudio y el análisis de los datos.**

El análisis de los datos determina la naturaleza de la interpretación. La relación entre el análisis y la interpretación de los resultados es diferente de una investigación a otra. Estudiar estas relaciones es más complicado en los trabajos exploratorios que en los experimentales.

A través de la finalidad del estudio y del tipo de análisis en la interpretación, se relacionan los datos analizados con otros aspectos conocidos y se abordan los factores internos, así toda interpretación debe ajustarse a las siguientes reglas:

#### **9.3.1.1 Continuidad en el proceso investigativo general.**

Esta continuidad se logra relacionando los resultados de nuestro estudio con los resultados de otros estudios para establecer la posibilidad de continuarlo.

#### **9.3.1.2 Conceptos aclaratorios.**

Con las normas anteriores se garantiza la homogeneidad y la coherencia del proceso y la relación con investigaciones a fines; sólo así los pasos de una persona pueden ser útiles para otra.

#### **9.3.1.3 La interpretación de los análisis.**

Es la última fase del proceso metodológico. Su importancia está en que permite al investigador hacer sus aportes personales y dar sentido a lo analizado. Las posibilidades de interpretación son ilimitadas. En esta etapa se decide si las hipótesis nulas son rechazadas y las positivas aceptadas o cual es la significación teórica de los resultados, además se valoran las técnicas utilizadas. Se debe dar cuenta de los antecedentes del estudio para aprovechar experiencias anteriores. También se determina la relación causal entre variables, sobre ellas el método impone la formulación de hipótesis.

#### **9.3.1.4 Cómo se lleva a cabo el análisis estadístico.**

Actualmente, el análisis estadístico se realiza mediante programas de computador, utilizando paquetes estadísticos. Cada paquete tiene su propio formato, instrucciones, procedimiento y característica. El procedimiento para analizar los datos consiste en crear o desarrollar un programa basándonos en el manual de un paquete. Este programa incluye la matriz de datos y las pruebas estadísticas seleccionadas. Luego se corre el programa, se obtienen los resultados y se interpretan.



Los principales paquetes estadísticos, actualmente son:

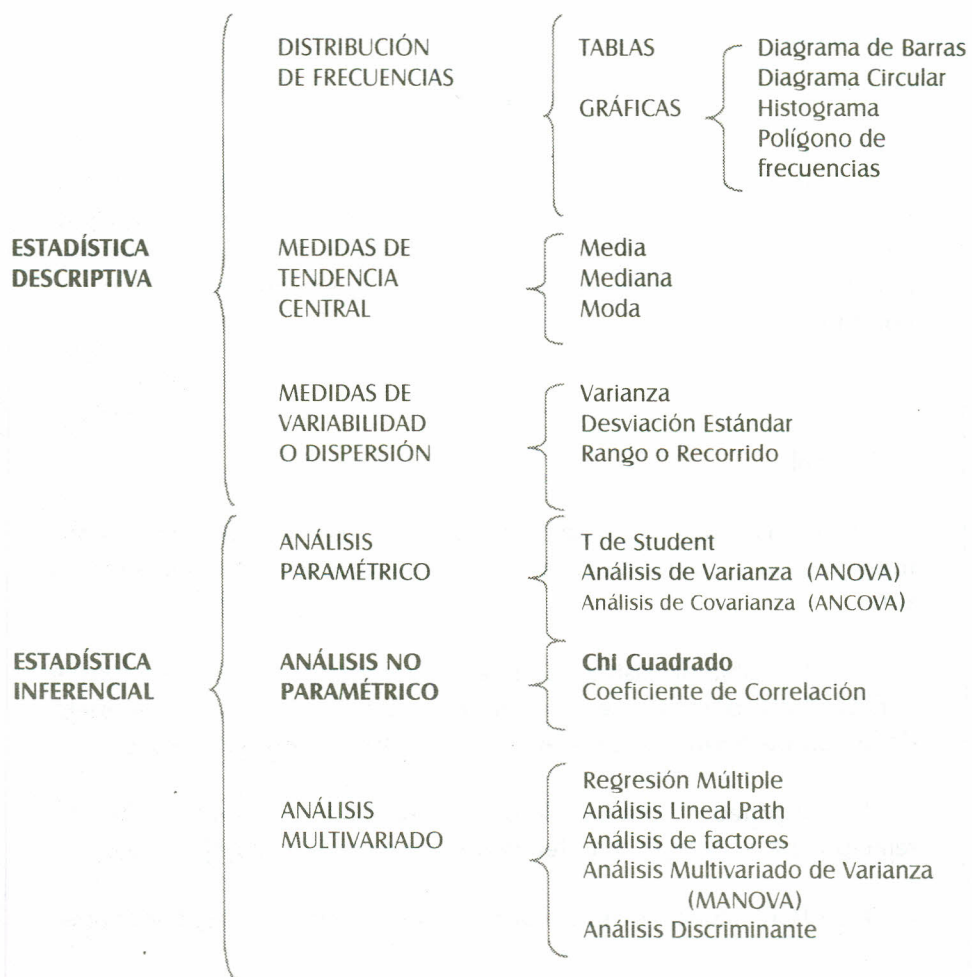
- **SPSS** es un paquete para Ciencias Sociales desarrollado en la Universidad de Chicago.
- **MINITAB** su costo es bajo, incluye numerosas pruebas estadísticas y una guía tutorial para aprenderlo a utilizar.
- **BMDP** Programa Biomédico Computarizado, aunque está diseñado para el área médica tiene aplicación en Ciencias Sociales. Desarrollado por la Universidad de California.
- **ESP** Paquete Econométrico de Software, se utiliza para análisis estadísticos de series cronológicas.
- **OSIRIS** conjunto organizado de rutinas integradas para la investigación con estadísticas. Desarrollado por la Universidad de Michigan.
- **SAS** Sistema de Análisis Estadístico, desarrollado por la Universidad de Carolina del Norte.

#### 9.3.1.5 Relación entre pruebas estadísticas y tipos de hipótesis.

PRUEBA ESTADÍSTICA PARAMÉTRICA	TIPO DE HIPÓTESIS
Coefficiente de Correlación	Correlacional
Regresión Lineal	Correlacional Causal
Prueba T	Diferencia de grupos
Análisis de varianza (ANOVA)	Diferencia de grupos / causal
Análisis de Covarianza (ANCOVA)	Correlacional / Causal
NO PARAMÉTRICA	
Chi Cuadrado	Diferencia de grupos / Correlación
Coefficiente de Correlación	Correlacional
MULTIVARIADO	
Regresión Múltiple	Correlacional / Causal
Análisis Lineal	Correlacional / Causal
Análisis de Factores	Correlacional / Causal
Análisis de Varianza (MANOVA)	Correlacional / Causal
Análisis Discriminante	Correlacional / Causal

**TABLA 11.** RELACIÓN PRUEBA ESTADÍSTICA – TIPO DE HIPÓTESIS.

En conclusión, el siguiente cuadro resume las técnicas del análisis estadístico:



CUADRO 8. TÉCNICAS ESTADÍSTICAS PARA EL ANÁLISIS DE DATOS.

## PRÁCTICA 9

Escoja las técnicas estadísticas que considere necesarias para realizar el análisis de los datos recogidos en el paso anterior, en el estudio del ejemplo que viene realizando con el presente trabajo. Se sugiere elaborar una tabla y un gráfico circular. Utilice el programa de la computadora.

Ejemplo:

En el caso que se viene trabajando en este estudio, acerca del manejo de la computadora por el niño, las estadísticas que necesitamos son:

1. Una tabla de distribución de frecuencias y un gráfico de barras con los datos obtenidos en las variables “tiempo dedicado al manejo de la computadora” y “programas preferidos”. (ver 8.2.1.4.)
2. La prueba T para evaluar la diferencia entre niños y niñas en relación con el tiempo que dedican al manejo de la computadora.
3. El Chi-cuadrado para relacionar las variables, sexo y beneficios.



## BIBLIOGRAFÍA

1. BÁEZ, Olga. Consultas odontológicas vs. Sexo, edad. Escuela de Administración de negocios, EAN. Bogotá, 2000.
2. CAMACHO DE BÁEZ, Briceida. (1985) La variabilidad lingüística; sus logros y limitaciones. En Revista Educación y Ciencia. Quinta época No. 1, Tunja, UPTC.
3. DURKHEIM, E. (1985). Les règles de la méthode sociologique, París. Versión española consultada, las reglas del método sociológico, Barcelona, Orbis.
4. GUMPERZ, J.J. y HYMES, D. (eds) (1972). Directions in Sociolinguistics, Nwe York, Holt Rinehart and Winston.
5. FASOLD, R. (1984). The Sociolinguistics of Society, Oxford, Blackwell.
6. KIBRIK, A. The Methodology of field investigation in linguistics, the Hague Mouton (1977)
7. LAVANDERA, B. (1984). Variación y significado. Buenos Aires, Hachette.
8. MILROY, L. (1987). Language and social Networks, 2ª de. Oxford, Blackwell.
9. MORENO, Francisco. (1986). Sociolingüística de los tratamientos. Estudios sobre una comunidad rural. Anuario de letras XXIV.  
\_\_\_\_\_, (1990) Estudios sobre variaciones lingüísticas. Alcalá de Henaos. Universidad de Alcalá.
10. MURRAY, R. Spiegel. (1980) Estadística. Serie Shaum.
11. NAGHI, M.N. (1984). Metodología de la investigación en Administración, Contaduría y Economía. México, Ed. LIMUSA.
12. SELLTIZ, C. y otros. (1980). Métodos de investigación en las relaciones sociales, 9ª ed. Madrid, Rialp.

13. SIEGEL, S. (1982). Estadística no paramétrica aplicada a la ciencia de la conducta. México, Editorial Trillas.
14. SILVA - CORVALÁN, Carmen. (1988). Sociolingüística: Teoría y Análisis, Madrid, Alambra.
15. WIERSMA (1986). Research methods in education: an introduction. Boston Mass. Allyn and Bacon, Inc. Capítulo 12.