

**Conceptos básicos metodológicos y estadísticos; escalas de medición,
cálculos aritméticos y gráficos de frecuencia**

Estadística Aplicada a las Ciencias Sociales I

Licenciatura en Sociología

Notas de clase

Dra. Rosana Guevara Ramos



Casa abierta al tiempo

□ **Conceptos metodológicos básicos**

Marco empírico (N o n): Conjunto de unidades de análisis que componen el espacio empírico (u observable) a investigar y es delimitado desde el marco teórico. Este puede quedar constituido por una muestra o una población, de acuerdo a la capacidad de cobertura que se tenga en la investigación (Ejemplo: Todos los estudiantes de bachillerato en Tlaxcala -Población→Censo- o un grupo representativo de los estudiantes de bachillerato en Tlaxcala - Muestra→Encuesta).

Unidad de análisis (i): Elemento social concreto sobre el cual se aplica la observación o medición de la variable de interés y constituye la unidad "mínima" de análisis (Ejemplo: familia, institución, empresa, persona, grupo social, etc..).

Población (N): Conjunto total de las unidades de análisis posiblemente observables (Ejemplo: Todos los alumnos de bachillerato).

Muestra (n): Subconjunto del total de las unidades de análisis factiblemente observable, la selección puede hacerse conforme a diferentes técnicas generando así distintos tipos y clases de muestreo (Ejemplo: Sólo una parte de los alumnos de bachillerato)

Variable (x,y,z): Característica o atributo social cuyo valor cambia de unidad a unidad de análisis y dicha "variabilidad" es observable o registrable (Ejemplo: Edad, escolaridad de los padres, preferencias políticas, etc..).

Variable de interés: Atributo del objeto de estudio que interesa particularmente al investigador dentro de una amplia gama de variables posibles; y cuyo registro y evaluación contribuye al análisis y explicación del fenómeno social investigado. (Ejemplo: Escolaridad de los padres).

Variabes simples y complejas: Las variables simples se caracterizan por poder medirse u observarse directamente y/o mediante una sola medición. Mientras que las variables complejas -por representar dimensiones complejas o sumamente abstractas de la realidad social - deberán utilizar varios indicadores sintetizados en un índice o indicador resumen que permitirá medir sistemáticamente las diferentes características o atributos.(Ejemplo: Variable simple = sexo, edad, estado civil; Variable compleja = religiosidad, cultura política, estrato social, etc)

Indicador (x,y,z = a): Aspecto que forma parte del atributo o variable compleja analizada y constituye una medida "parcial"; a la cual se recurre cuando la variable no es de observación directa o simple -por el grado de complejidad o abstracción de esa característica evaluada. Estadísticamente se le da el mismo tratamiento que a una variable, aunque metodológicamente son distintas dado que la variable es una medición definitiva y completa; mientras el indicador es una medición parcial.. (Ejemplo: la variable religiosidad que se evalúa a través de a) creencia en dios; b) asistencia a reuniones religiosas; c) lectura de temas religiosos; d) prácticas religiosas proselitistas; etc..)

Índice (x,y,z): Indicador resumen que conjunta una serie de indicadores y permite una medición final única, -aunque de origen esta medición fue compuesta o conformada por varias mediciones o indicadores. Normalmente el índice o indicador resumen se mide a través de escalas ordinales. (Ejemplo: el indicador resumen o medida última y única seguramente se expresaría como "Grado de religiosidad" y se elaboraría con base en la sumatoria de los puntajes que haya reportado cada indicador a, b, c y d)

Valor (x): Registro numérico o simbólico que adquiere la variable en sus múltiples manifestaciones a través de las unidades de análisis (Ejemplo: Años (Xi) 18, 20, 2, 23, 24.....n)

Dato: Lugar que se le asigna a cada valor para ordenarlos sistemáticamente conforme a un criterio "ascendente" o "descendente"(Ejemplo: dato primero, segundo, tercero, etc.).

Caso: Entidad informativa que contiene todas las mediciones o valores realizados a la unidad de análisis (sujeto, familia, institución, empresa, etc.). La cual sirve para comparar posteriormente los diferentes valores observados caso por caso (Ejemplo: Todos los valores obtenidos de los alumnos de bachillerato a través del Caso 1; Caso 2; hasta el caso n o N.)

Manual de Códigos: Sistema de codificación que clasifica y resume la información mediante la "etiquetación" de variables y valores; tomando como referente las preguntas de un cuestionario. El manual o libro de códigos permite un manejo claro de las variables y los valores; lo cual facilita la captura y vaciado de estos en una base de datos. Es recomendable elaborarlo junto con el cuestionario.

Base de datos: Sistema de información que permite ordenar los valores con base a dos criterios: variables (columnas) y casos (renglones).

□ Conceptos estadísticos básicos

Frecuencia: Número de veces que se repite un mismo valor y constituye el análisis estadístico básico para ordenar una serie de datos dispersos.

- **Frecuencia absoluta (fa):** Expresada en números absolutos
- **Frecuencia relativa (fr):** Expresada en porcentajes
- **Frecuencia acumulada (FaAc, FrAc):** Agrupada sistemáticamente mediante una sumatoria de la frecuencia absoluta o relativa

Moda (x): Valor con mayor frecuencia o que se repite más veces dentro de un listado o conjunto de datos

Mediana (x) : Valor que se encuentra en medio del listado de los datos cuando estos fueron ordenados de manera ascendente o descendente.

Media (x): Promedio de los valores que se obtiene sumando todos los valores y dividiéndolos en el número total de datos que compone la muestra o la población.

Rango o recorrido (r): Distancia entre el dato menor y el mayor en un listado de valores ($Rango = Ls - Li$)

Varianza (s): Sumatoria de la diferencia de la media con respecto a cada dato (x) elevada al cuadrado entre n-1. Sirve para calcular el promedio de las distancias de cada dato con respecto a la media y se eleva al cuadrado para eliminar números negativos -puesto que estos no existen en mediciones de distancias. Es una medida útil para algunas pruebas estadísticas y constituye un cálculo previo para obtener la desviación estándar.

Desviación Estándar (s): Es la raíz cuadrada de la sumatoria de la diferencia de la media con respecto a cada dato (x) elevada al cuadrado entre n-1. Sirve para medir el promedio de las distancias de cada dato con respecto a la media

□ Escalas de medición

Definición de medición: Consiste en la asignación de un nombre, signo o número a las observaciones (valor) para clasificarlas y, así ordenadas, poder manipularlas matemáticamente (acumular, sumar, restar, multiplicar y/o dividir, etc.).

Tipo de Variable	Escala de medición	Descripción	Ejemplos	Medidas y pruebas estadísticas válidas
Cualitativa	Nominal o clasificatoria	La variable se expresa a través de símbolos, letras o números (simbólicos no reales) La clasificación o etiquetación permite diferenciar un conjunto de otro, por lo que las categorías de clasificación deberán ser mutuamente excluyentes bajo un principio de igualdad o equivalencia (=). Los códigos o clasificaciones expresados en números no aceptan tratamiento matemático. Sin embargo se pueden ordenar y contabilizar en frecuencias y modas	1) Variable: Percepción de ingreso Valores: (*) si, (&) no 2) Variable: Genero Valores: (M), masculino; (F), Femenino 3) Variable: Religión Valores: (1), Católico (2), Protestante (3), Otra 4) Variable: Filiación partidaria Valores: (1), PRI; (2), PAN y (3), PRD. 5) Variable: Profesión Valores: (P), Politólogo; (S), Sociólogo; (PS), Psicólogo Social; (A), Administrador	1) Frecuencias (absolutas y relativas) 2) Cálculos aritméticos con base en frecuencias (tasas, razones y proporciones) 3) Estadística descriptiva: Moda Estadística inferencial 4) Pruebas de proporciones: Binomial y de Chi cuadrada (análisis univariado) 5) Pruebas de correlación con Ji Cuadrada (análisis multivariado) con coeficientes Phi, C de contingencia y V de Cramer. 6) Pruebas de cambios o diferencias
	Ordinal o jerárquica	La variable se clasifica de manera jerárquica ya sea en orden ascendente o descendente de ahí que el criterio de agrupación manifieste valores diferentes entre los datos y los signos < o > adquieran un sentido, la distancia entre las distintas categorías no es conocida.	1) Variable: Nivel salarial por zona: (A) Zona A (alto) (B) Zona B (medio) (C) Zona C (bajo) Aunque se supone que la distancia exacta entre cada nivel no es conocida. 2) Variable: Estrato socioeconómico Valores: (3) Alto, (2) Medio, (1) Bajo. 3) Variable: Grados de aceptación: (A) Alto, (M) Medio y (B) Bajo 4) Variable: Aptitud (M) Mayor (D) Media (N) Nula	Todas las permitidas en la escala anterior más : 1) Estadística descriptiva: Mediana Estadística inferencial 2) Pruebas de correlación - Coeficiente R de Spearman - Coeficiente Gamma - Tau de Kendall - W de Kendall - Prueba de medianas 3) Pruebas de cambios o diferencias

Cuantitativa	Intervalar	Además de las propiedades de la escala ordinal tiene las siguientes característica: 1) La distancia o proporción entre las categorías son conocidas, 2) El punto cero y la unidad de medición son "arbitrarios"*. Los valores si aceptan un tratamiento matemático, admiten todas las manipulaciones matemáticas (+, x, -, etc.) (*)"arbitrarios" en la medida en que el investigador los construye y los justifica teórica, empírica y metodológicamente para su objeto de estudio..	1) Variable: Grados centígrados 2) Variable: Grados Fahrenheit 3) Variable: Coeficiente Intelectual (IQ) 4) Tiempo en calendario, siglos	Todas las medidas de las escalas anteriores más: 1) Estadística descriptiva Medias, desviación estándar y varianza Estadística inferencial 2) Pruebas de correlación - Coeficiente de Correlación de Pearson 3) Prueba t 4) Anova 3) Regresión múltiple: lineal y no lineal
	Razón o proporción	Reúne todas las características de las escalas nominal, ordinal, e intervalo, además de tener un punto cero (0) absoluto o verdadero, en tanto que significa ausencia de propiedad y la distancia entre los intervalos y valores son conocidos , admiten toda operación matemática.	1) Variable: Ingreso en pesos 2) Variable: Edades en número de años 3) Variable: Estaturas en metros 4) Variable: Población en número de habitantes	Admite todas las pruebas estadísticas anteriores

Ejemplo de una variable expresada en diferentes escalas de medición

ESCALA	SALARIOS
Nominal	Zonas A, B,C.
Ordinal	Niveles de ingreso: Alto, Medio, Bajo
Razón o proporción	\$600 - \$1,200 \$1,800 - \$2,400 \$3,000 - \$3,600

□ **Cálculos aritméticos**

Proporciones, porcentajes, razones y tasas

A través de operaciones aritméticas es posible que los datos resulten más significativos de ahí que mediante las proporciones, porcentajes, razones y tasas es posible comparar cantidades o establecer comparaciones que en ocasiones expresadas en números absolutos no resultan tan claras.

1. Proporción. Representa las cantidades o frecuencias parciales de los valores con respecto al total de esas frecuencias basados en la unidad 1. Se calcula

dividiendo el número parcial entre el número total y la sumatoria de todos los cocientes (resultados parciales) deberá dar al final 1.

$$\text{Fórmula: } P = \frac{np}{Nt}$$

Donde

np= es la frecuencia parcial del valor

Nt= es la suma de todas las frecuencias parciales o total de éstas

2. Porcentajes. Representa las cantidades o frecuencias parciales de los valores con respecto al total que toma como unidad el 100. Se calcula dividiendo cada frecuencia parcial entre el total de frecuencias y se multiplicándola por 100 o aplicando la regla de 3.

$$\text{Fórmula: } \% = \frac{np}{Nt} \cdot 100$$

Regla de 3

Frecuencia total ----- 100%

Frecuencia parcial----- ¿?

Nota: La regla de 3 permite despejar respectivamente cualquiera de los factores que se desconozcan siempre 1 a la vez.

3. Razones. Es el cociente que simboliza el resultado de comparar dos cantidades derivadas de las frecuencias de dos distintos valores de una variable; es un valor relativo que no depende de valores absolutos. Se calcula dividiendo el número mayor entre el número menor; los resultados se expresan anteponiendo dos puntos (:) entre los resultados obtenidos. Variable género: valores y sus respectivas frecuencias M=90 y H=30; el resultado es 3:1 (Existen 3 mujeres por 1 hombre).

$$R = \frac{N \text{ mayor}}{N \text{ menor}}$$

4. Tasas de incremento. Son indicadores numéricos que comparan cantidades distintas en series de tiempo con el fin de comprender el comportamiento de una variable a lo largo de cierto periodo. La tasa aritmética o de crecimiento simple (tis) sirve para evaluar sólo dos periodos; mientras que la tasa geométrica o de crecimiento promedio (tip) sirve para evaluar más de dos periodos.

Las fórmulas para obtener los cálculos son:

Tasa aritmética o de crecimiento simple (tis)

$$Tis = [(t_1/t_0) - 1] 100$$

Tasa geométrica o de crecimiento promedio (tip)

$$Tip = [(t_1/t_0)^{1/n-1} - 1] 100$$

Donde

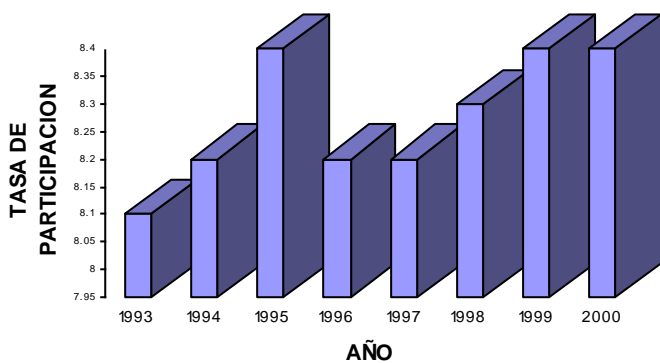
t₀= dato inicial y **t₁** = dato final

□ **Definiciones de los gráficos de frecuencias más comunes**

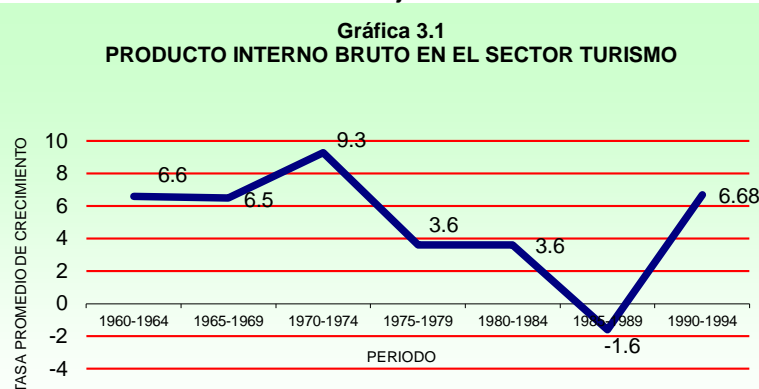
Los gráficos constituyen imágenes visuales que nos ayudan a comprender de una forma rápida y clara el comportamiento de una variable en función de la distribución de sus frecuencias. Existen diferentes diseños al respecto en función de las características de las variables y la claridad que busquemos.

1. Histograma. Representación gráfica de la distribución de frecuencias de una de las variables en forma de diagrama de barras en cuyas abscisas (eje de las x) se encuentran normalmente los valores de la variable y en cuyas ordenadas (eje de las y) se especifican las frecuencias correspondientes a cada valor. Este tipo de gráfico se usa en variables cuantitativas.

Gráfica 3.2
PARTICIPACION DEL PIB TURISTICO EN EL PIB NACIONAL (1993-2000)

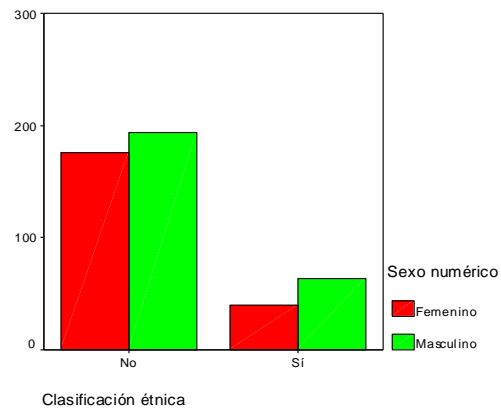


2. Polígono de frecuencia. Representación gráfica de la distribución de frecuencias a través de una línea quebrada que une los puntos medios de las barras correspondientes al histograma original. Es importante observar que los puntos extremos de la línea del polígono de frecuencia se deberán colocar al principio y al final sobre la línea base del eje de las X.

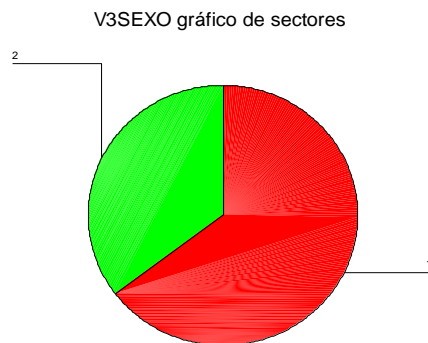


3. Gráfica de barras. Representación gráfica de la distribución de frecuencias a través de barras cuya base comprende los valores categóricos de la variable nominal u ordinal y la altura está determinada en función de la frecuencia de cada

uno de esos valores. Es una gráfica particularmente apropiada para variables que comprendan más de 4 valores.



4. Diagrama circular o gráfica de *pie*. Representación gráfica de la distribución de frecuencias mediante una imagen circular cuya área total comprende 360° y se encuentra dividida en función de las frecuencias parciales expresadas en grados. Se usa regularmente en variables nominales y ordinales cuando éstas tienen un máximo de 4 valores.



Bibliografía

- Babbie, E. (1988). *Métodos de investigación por encuesta*. Colección Biblioteca de la salud, Fondo de Cultura Económica; México.
- Babbie, E. (1992). *The Practice of Social Research*. Editorial Wadsworth. Part 2, 3 y 4; California.
- Berenson, Mark L. y D. M. Levine. (1996). *Estadística básica en Administración. Conceptos y aplicaciones*. Editorial Prentice Hall; México.
- Blalock, H. (1986). Segunda parte. "Estadística descriptiva univariada" en *Estadística Social*. Editorial Fondo de Cultura Económica; México.
- Boudon y Lazarsfeld. (1985). "De los conceptos a los índices empíricos" en *Metodología de las Ciencias Sociales*. Editorial LAIA. (Vol. I); Barcelona.
- Cea D'Ancona, M. A. (2001). *Metodología cuantitativa. Estrategias y técnicas de investigación social*. Síntesis Sociología.; Madrid.
- Güell, Anthony" M. (1985). "Hipótesis y variables" en *Metodología de las Ciencias Sociales*. Editorial LAIA. (Vol. I); Barcelona.
- Hans Zeisel. (1990). *Dígalo con números*. Fondo de Cultura Económica; México.
- R. Hernández, C. Fernández y P. Baptista (1998), *Metodología de la Investigación*, 2ª ed. Mc Graw Hill; México
- Mendenhall, W. (1987). *Introducción a la probabilidad y la estadística*. Grupo Editorial Iberoamérica; México.
- M.R. Spiegel, J. Schiller y R.A. Srinivasan (2009); *Probabilidad y estadística*. Mc Graw Hill
- Padua, J. e I. Ahman. (1982). "Escala para la medición de actitudes" en *Técnicas de investigación aplicada a las ciencias sociales*. Ed. FCE; México.
- Pardo, A. y Ruiz, M.A. (2002). *SPSS 11, Guía para el Análisis de datos*. Mac Graw Hill; México.
- Ritchey, Ferris. (2002). *Estadística para las Ciencias Sociales*. Mc Graw Hill; México.
- Spiegel, M.R. y Stephens, L. (2004). *Estadística*. Mc Graw Hill; México.
- Wright Mills. (1961). *La imaginación sociológica*. Fondo de Cultura Económica; México.