



unam – ents

Universidad Nacional Autónoma de México Escuela Nacional de Trabajo Social

Estadística aplicada a la Investigación Social I

Lic. Ciro López Mendoza

Área: Metodología y práctica
de trabajo social

Semestre: 3

Créditos: 5

Carácter: Obligatorio

Sistema Universidad Abierta

Contenido

	Pág.
Presentación	3
Introducción	4
Objetivo general	6
Perfil de egreso	7
Temario general	10
Unidad 1 La investigación social y la estadística	17
Unidad 2 Población y muestreo	23
Unidad 3 Método Estadístico	55
Unidad 4 Recolección de datos	60
Unidad 5 Recuento y representación de datos	65
Unidad 6 Descripción de datos	77
Unidad 7 Correlación y regresión lineal	102
Glosario	116
Preguntas frecuentes	129
Bibliografía	137

Presentación

La Escuela Nacional de Trabajo Social inició sus estudios de *Licenciatura en Sistema Universidad Abierta*, en el año escolar 2003, con el Plan de Estudios aprobado por el H. Consejo Universitario el 10 de julio de 1996. Fue reestructurado en el año 2002 con aprobación del Consejo Académico del Área de las Ciencias Sociales, en su sesión del 26 de noviembre de 2002.

En el Sistema Universidad Abierta, la relación entre asesores, estudiantes y material didáctico es fundamental. En este sentido, en la Escuela se puso especial atención para lograr mayor calidad en los materiales.

De ésta manera, el material que ahora te presentamos debe constituirse en una herramienta fundamental para tú aprendizaje independiente, cada uno de los componentes que lo integran guardan una congruencia con el fin de que el estudiante pueda alcanzar los objetivos académicos de la asignatura.

El material pretende desarrollar al máximo los contenidos académicos, temas y subtemas que son considerados en el programa de estudio de la asignatura. Esto no pretende soslayar el papel y responsabilidad preponderante del estudiante, que debe profundizar en la búsqueda de conocimientos en todas aquellas fuentes que tenga a su alcance hasta hacer realidad los objetivos y el perfil de egreso propuesto.

Este material es perfectible, por ello, con el apoyo de las experiencias de los estudiantes y otros profesores, serán revisados y actualizados de manera permanente por el asesor, cuyos aportes sin duda, contribuirán para su mejora y enriquecimiento.

Te damos la más cordial bienvenida y te deseamos toda clase de éxitos en los estudios que inicias en esta, tu Escuela, la **Escuela Nacional de Trabajo Social** de la **Universidad Nacional Autónoma de México**.

Introducción

El presente material aborda los aspectos de mayor relevancia de la estadística aplicada a la investigación social; parte de la definición de ambas disciplinas y de su relación e importancia en el análisis de los problemas de carácter social.



Este material presenta aspectos vinculados a población y muestreo, en donde estudiaras la forma en cómo se puede obtener un tamaño de muestra y los distintos métodos para hacerse de dicho subconjunto de la población.



con el objeto de proporcionarte la capacidad para seleccionar y aplicar las medidas más adecuadas en el análisis de los fenómenos sociales y con ello, puedas tomar decisiones con la menor incertidumbre posible.

El texto te brinda un método sencillo y práctico, para entender la estadística desde un punto de vista lógico más que matemático. Es decir, te proporciona las herramientas estadísticas para el estudio cuantitativo y cualitativo de los datos procedentes de un proceso de investigación científica,

En la unidad I estudiarás los conceptos fundamentales de la investigación y la estadística con objeto de que determines su relación e importancia.

En la unidad II aprenderás conceptos como población, muestreo y muestra para generar en ti la capacidad de calcular un tamaño de muestra así como elegir el tipo de muestreo más adecuado según el estudio que pretendas realizar.

El método estadístico lo encontrarás en la unidad III lo cual te permitirá conocer una secuencia lógica de pasos para analizar los datos procedentes de una investigación.

La unidad IV te proporciona el concepto más importante del que se alimenta la estadística, es decir, las variables así como su nivel de medición, elementos que integran el primer paso del método estadístico, la recolección.

En la unidad V encontrarás el segundo y tercer paso del método estadístico que se refieren al conteo

y presentación de datos respectivamente. Aquí podrás definir la frecuencia de los datos así como el tipo de gráfico o cuadro que le corresponde a cada variable según su nivel de medición.

El paso número cuatro del método estadístico lo estudiarás en la unidad VI, el cual se refiere a la descripción de datos y en donde podrás conocer, calcular e interpretar medidas de resumen para variables cualitativas y cuantitativas.

Finalmente la unidad VII te brinda los elementos básicos para realizar procesos de correlación y regresión lineal, es decir, por un lado determinar la existencia o no de una relación entre dos variables, así como su magnitud y dirección y por el otro, a partir de dicho procedimiento efectuar una predicción.

Objetivo general

Al finalizar el curso, **emplearás** elementos de muestreo y diversas pruebas estadísticas de tipo descriptivo, así como de correlación y regresión con el propósito de que **selecciones y apliques** dichos aspectos dentro de un proceso de investigación social.



Perfil de egreso

Al terminar el curso serás capaz de seleccionar y aplicar las medidas estadísticas de nivel descriptivo, elementos de muestreo así como de correlación y regresión en el estudio de los fenómenos sociales. Contarás con los siguientes conocimientos, habilidades y actitudes:

Conocimientos sobre:

- El proceso de investigación científica y el papel de la estadística dentro del mismo.
- La relación entre la investigación y la estadística como elementos indisolubles e invariablemente complementarios.
- Elementos generales de población y muestreo.
- El método estadístico, etapas y elementos que lo constituyen.
- Variables, su conceptualización así como sus niveles de medición.
- Las formas de presentación de las variables en tablas y gráficos, según su nivel de medición.
- Las medidas de resumen para variables cualitativas y cuantitativas.
- Elementos de correlación y regresión lineal.



Habilidades:

- Identificarás el papel que juega la estadística en un proceso de investigación científica.
- Determinarás el tipo de muestreo más adecuado en el estudio de los fenómenos sociales.
- Reconocerás las escalas de medición de las variables y aplicar las técnicas estadísticas que le corresponda a cada una de ellas.
- Distinguirás la forma más adecuada de cómo se deben presentar las variables en consideración a sus niveles de medición.
- Emplearás las medidas de resumen que mejor representen a un conjunto de datos.



- Reconocerás las características de un conjunto de datos a partir de la forma y tendencia de su distribución.
- Determinarás la asociación entre variables, dirección y magnitud así como desarrollar predicciones a partir de la relación entre ellas.

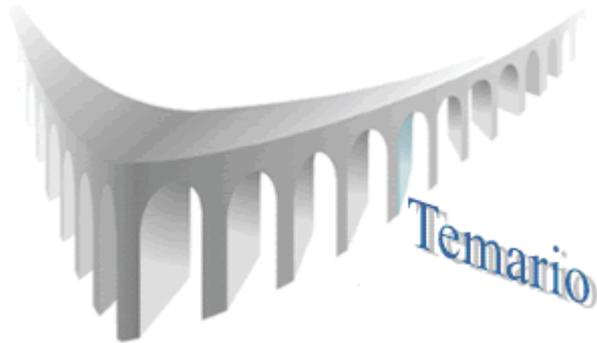
Actitudes:

- Reconocerás a la estadística como un elemento fundamental al desarrollar un proceso de investigación científica
- Aplicarás el método estadístico para el estudio de los fenómenos sociales y la toma de decisiones
- Adoptarás un proceso de enseñanza aprendizaje continuo basado en una secuencia sistemática de pasos.
- Aplicarás el método estadístico en el ámbito personal y profesional.



Temario general

(Horas totales: 64)



I. LA INVESTIGACIÓN SOCIAL Y LA ESTADÍSTICA (2 horas)

- 1.1 Concepto de investigación social
- 1.2 Concepto de estadística
- 1.3 Objeto y clasificación de la estadística
- 1.4 Relación e importancia de la investigación social y la estadística

II. POBLACIÓN Y MUESTREO (16 horas)

- 2.1 Conceptos fundamentales
- 2.2 Cualidades de una buena muestra
- 2.3 Tamaño de la muestra



- 2.4 Tipos de muestreo
 - 2.4.1 Probabilístico
 - 2.4.1.1 Aleatorio simple
 - 2.4.1.2 Sistemático
 - 2.4.1.3 Aleatorio estratificado

2.4.1.4 Estratificado y por racimos

2.4.2 No probabilística

2.4.2.1 De juicio o criterio

2.4.2.2 De diseño de bola de nieve

2.4.2.3 Por accidente

2.4.2.4 Por conveniencia

2.4.2.5 Por cuota

2.4.2.6 Por sujetos-tipo

2.4.2.7 Por sujetos voluntarios

2.4.2.8 Por expertos

III. MÉTODO ESTADÍSTICO (1 hora)

3.1 Concepto y ubicación en el contexto de la investigación social

3.2 Etapas

3.2.1 Recolectar

3.2.2 Contar

3.2.3 Presentar

3.2.4 Describir

3.2.5 Analizar

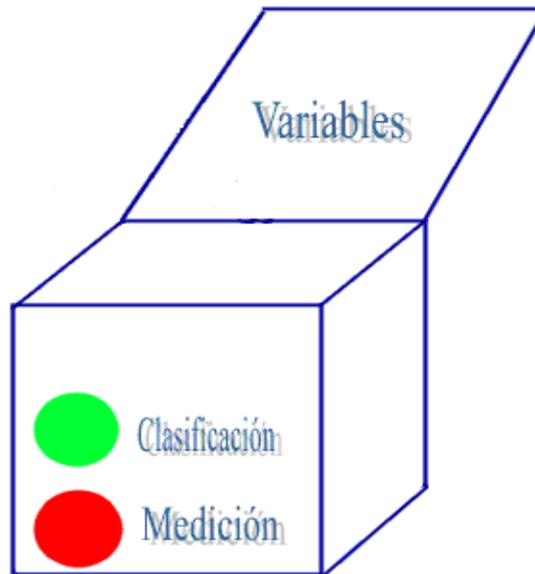


IV. RECOLECCIÓN DE DATOS (2 horas)

4.1 Variables

4.1.1 Clasificación

4.1.2 Niveles de medición



V. RECuento Y PRESENTACIÓN DE DATOS (5 horas)

5.1 Conteo de datos

5.2 Presentación de datos



5.2.1 Características mínimas para la creación de tablas o cuadros

5.2.2 Clasificación de datos

5.2.2.1 Intervalo cerrado

5.2.2.2 Intervalo abierto

5.2.2.3 Tamaño del intervalo

5.2.2.4 Limite inferior y superior

5.2.2.5 Marca de clase

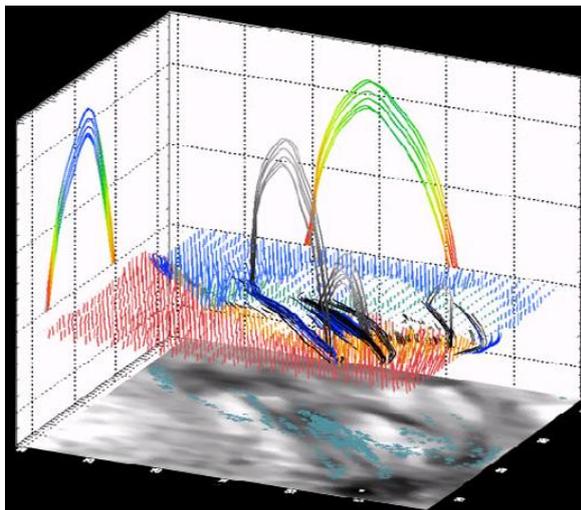
5.2.2.6 Limite inferior y superior reales

5.2.2.7 Frecuencia Absoluta

5.2.2.8 Frecuencia Relativa

5.2.2.9 Frecuencia absoluta acumulada

5.2.2.10 Frecuencia relativa acumulada



5.2.3 Características mínimas para la creación de gráficos

5.2.3.1 Diagrama barras separadas

5.2.3.2 Circular, de sectores o de pastel

5.2.3.3 Pictograma

5.2.3.4 Histograma

5.2.3.5 Polígono de frecuencias

5.2.3.6 Polígono de frecuencias acumuladas u ojiva o diagrama de Galton

5.2.3.7 Diagrama de caja

VI. DESCRIPCIÓN DE DATOS (24 horas)



6.1 Para variables cualitativas

6.1.1 Proporciones

6.1.2 Porcentajes

6.1.3 Razones

6.1.4 Tasas

6.2 Para variables cuantitativas

6.2.1 Medidas de tendencia central

6.2.1.1 Moda

6.2.1.2 Mediana

6.2.1.3 Media aritmética

6.2.1.4 Media ponderada

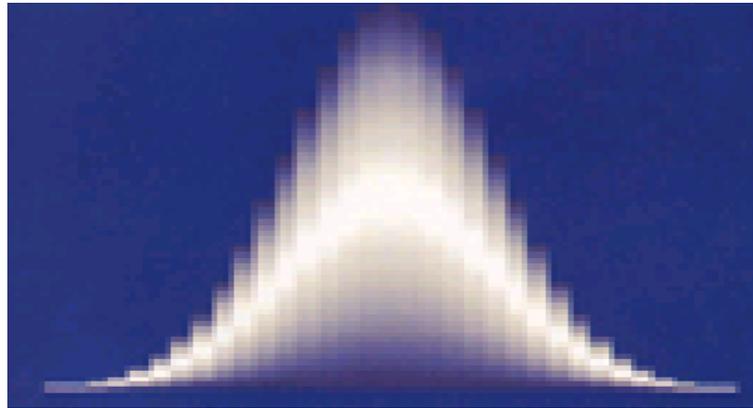


6.2.2 Medidas de posición

6.2.2.1 Centil

6.2.2.2 Cuartil

6.2.2.3 Decil



6.2.3 Medidas de dispersión

6.2.3.1 Rango

6.2.3.2 Rango intercuartílico

6.2.3.3 Desviación media

6.2.3.4 Varianza

6.2.3.5 Desviación estándar

6.2.3.6 Coeficiente de variación

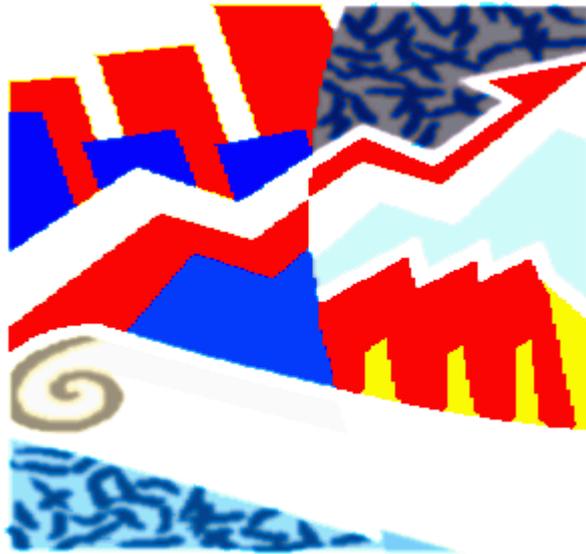
6.2.4 Medidas de distribución

6.2.4.1 Sesgo

6.2.4.2 Curtosis

6.2.4.3 Calculo de sesgo y curtosis

VII. CORRELACIÓN Y REGRESIÓN LINEAL (14 horas)



7.1 Relaciones lineales

7.1.1 Tipos

7.2 Correlación

7.2.1 Coeficientes de correlación

7.2.1 r de Pearson

7.2.2 Spearman r_s

7.3 Regresión

7.3.1 Recta de regresión por mínimos cuadrados

Unidad I

La investigación social y la estadística

Introducción

La presente unidad aborda los conceptos de investigación social y estadística, establece el objetivo y clasificación de esta última para con ello enlazar o encadenar ambas ideas y encontrar su relación e importancia en el estudio de la problemática social que enfrentan los sujetos en lo individual y en lo colectivo.

Objetivo

Al finalizar la presente unidad el alumno empleará los elementos fundamentales de la investigación social y de la estadística con el propósito de establecer su relación e importancia para el estudio de los fenómenos sociales.

Temario

- 1.1 Concepto de investigación social
- 1.2 Concepto de estadística
- 1.3 Objeto y clasificación de la estadística
- 1.4 Relación e importancia de la investigación social y la estadística

Diagrama conceptual



1.1 Concepto de investigación social

Elorza (2000) señala atinadamente que la ciencia se basa en la contrastación empírica de las teorías con la evidencia, a su vez las teorías se comprueban tratando de demostrar que son falsas; si no se logra esto, se retiene la teoría. El método de la ciencia es el de las conjeturas audaces e ingeniosas seguidas por intentos rigurosos de refutarlas.

Así, las teorías tratan de dar sentido a los hechos de la realidad y explicarlos, por su parte la investigación es el elemento creativo de la ciencia, donde se procura establecer la relación entre variables con el objetivo de expandir el conocimiento y la comprensión de la realidad.

La investigación científica -señala Hernández, Fernández y Baptista (2003) al retomar a Kelinger (1975)- es el proceso sistemático, controlado, empírico y crítico, de proposiciones hipotéticas sobre las presumidas relaciones entre fenómenos naturales. Es sistemática y controlada porque implica la existencia de una disciplina constante para hacer investigación científica y no se deja los hechos a la casualidad. Empírica significa que se basa en fenómenos observables de la realidad. Y crítica quiere decir que se juzga constantemente de manera objetiva y se eliminan las preferencias personales y los juicios de valor.

La investigación científica es el proceso mediante el cual se obtienen conocimientos. Investigar es “seguir sistemáticamente la huella”, seguir el rastro de los hechos para explicarlos. La definición etimológica sugiere que investigar es volver a buscar. De *in* y *vestigium*: huella pista; hacia la pista, seguir la pista.

Reynolds, G.S. (1973) apunta que la mayor parte de la investigación da como resultado un redescubrimiento y, por lo tanto, una confirmación de principios y de hechos conocidos, o bien representan un intento cuidadoso de responder en forma objetiva y reiterada a una pregunta no contestada hasta entonces. Por

otra parte, la investigación significa la búsqueda y descubrimiento de hechos y principios que anteriormente eran mal entendidos o no se concebían. Es un proceso en el que la única constante es el cambio.

Para Hernández, Fernández y Baptista (2003), Cozby (2004), Selltiz, Wrightsman y Stuart (1980), Castañeda, De la Torre, Morány Lara (2002), entre otros, la investigación puede cumplir dos propósitos fundamentales: a) producir conocimiento y teorías (investigación básica) y b) resolver problemas prácticos (investigación aplicada).

Cuando la motivación que lleva a realizar una investigación consiste en acrecentar el conocimiento se dice que se trata de ciencia pura o básica; en cambio, cuando se investiga con fines prácticos se habla de ciencia aplicada.

La investigación es una característica esencial de los últimos siglos, representa la herencia más clara y de mayor trascendencia hecha por el hombre.

Así, la **investigación social** puede ser definida como un proceso sistemático, controlado, empírico y crítico de aseveraciones hipotéticas sobre las posibles relaciones sociales que presentan los sujetos en lo individual y/o en lo colectivo.

1.2 Concepto de estadística

La estadística se emplea actualmente como un valioso auxiliar en los más diversos campos del conocimiento y en las más variadas de las ciencias fundamentales y aplicadas; difícilmente podría encontrarse un campo de la actividad cognitiva en el que el instrumental estadístico no tenga aplicación.

Para definir qué es la estadística es indispensable partir de lo que no es, es decir:

- No es un conjunto de técnicas con las que uno pueda probar todo aquello que uno desee.
- No es una mera colección de datos.
- No sólo es aplicable a una gran colección de datos.
- No es un instrumento de medición.
- No establece los pasos a seguir en la construcción de un instrumento de recolección de datos.

Así, la estadística puede ser definida como aquella que se ocupa de los métodos y procedimientos para recoger, clasificar, resumir, hallar regularidades y analizar los *datos*, siempre y cuando la *variabilidad* e *incertidumbre* sea una causa intrínseca de los mismos; así como de realizar *inferencias* a partir de ellos, con la finalidad de ayudar a la toma de *decisiones* y en su caso formular *predicciones*.

La estadística es aquella que mediante métodos científicos, recopila, organiza, presenta, resume, y analiza datos para obtener conclusiones válidas y tomar decisiones razonables con base en dicho análisis.

Es un conjunto de técnicas diseñadas para cubrir dos funciones, a saber: describir e inferir.

Finalmente la estadística se puede asumir como una herramienta matemática de apoyo a la investigación social que recopila, cuenta, presenta, describe y analiza un conjunto de datos variables asumiendo un margen de error o incertidumbre.

1.3 Objeto y clasificación de la estadística

Un conjunto de datos se obtiene a partir de observaciones numéricas de conjuntos que se caracterizan por la variación que muestran sus componentes. Estos datos permiten el estudio de fenómenos que se distinguen por su variación.

La estadística tiene por objeto –según establece Holguín (1981)- resumir los datos más destacados de los elementos que componen un conjunto, logrando así aprehender más fácilmente su contenido.

Al tener un conjunto de datos sumamente extenso y por tanto complejo; es conveniente resumirlos, reducirlos, hasta que la masa caótica y desordenada de los datos tome forma mediante la obtención de medidas estadísticas. Esto permitirá describir sus características más preponderantes y poner de relieve las relaciones existentes entre sus componentes en un momento o a lo largo de un tiempo determinado. A partir de dicha descripción es posible construir inferencia estadística.

Es posible por tanto clasificar la estadística en descriptiva, cuando los resultados del tratamiento estadístico no pretenden ir más allá del conjunto de datos, e inferencial cuando el objetivo del estudio es derivar las conclusiones obtenidas a un conjunto de datos más amplio.

Estadística descriptiva: es la rama de la estadística que recolecta, recuenta, presenta y describe un conjunto de datos.

Estadística inferencial o analítica es aquella que proporciona los métodos para estimar las características de un grupo total (población), basándose en datos de un conjunto pequeño (muestra) de observaciones.

En este sentido la estadística descriptiva establece las características generales de un grupo de datos utilizando métodos numéricos y gráficos que resumen y presentan la información contenida en ellos.

La estadística inferencial por su parte se apoya en el cálculo de probabilidades y a partir de datos muestrales, efectúa estimaciones, decisiones, predicciones u otras generalizaciones sobre un conjunto mayor de datos.

1.4 Relación e importancia de la investigación social y la estadística

La investigación social y la estadística convergen a partir de las siguientes premisas:

- Son procesos de constante exploración y descubrimiento.
- Son medio para examinar y entender la operación de los fenómenos sociales.
- Brindan puntos de vista y procedimientos técnicos que revelan detalles que de otra forma escaparían a nuestra conciencia.
- Tienen un carácter universal.
- Generan conocimiento.
- Tienen una metodología.

Resumen de la unidad

La presente unidad temática aborda el concepto de investigación científica para puntualizar la noción de investigación social, como elemento independiente respecto a la idea de estadística, objeto y clasificación con la finalidad de entrelazarlos y establecer así su relación e importancia.

La investigación y la estadística se estudian en esta unidad como procesos de constante exploración y descubrimiento; y como medios para examinar y/o entender la operación de los fenómenos sociales.

Unidad II

Población y muestreo

Introducción

La presente unidad aborda los conceptos básicos de muestreo así como los elementos generales para calcular un tamaño de muestra acorde con las exigencias del investigador. Por otra parte, presenta una clasificación general de los tipos de muestreo que existen en sus dos ramas fundamentales: lo probabilístico y lo no probabilística o determinístico.

Objetivo

Al finalizar la presente unidad aplicarás los aspectos fundamentales del muestreo y serás capaz de calcular un tamaño de muestra con el propósito de que los emplees en el estudio de los fenómenos sociales.

Temario

2.1 Conceptos fundamentales

2.2 Cualidades de una buena muestra

2.3 Tamaño de la muestra

2.4 Tipos de muestreo

 2.4.1 Probabilístico

 2.4.1.1 Aleatorio simple

 2.4.1.2 Sistemático

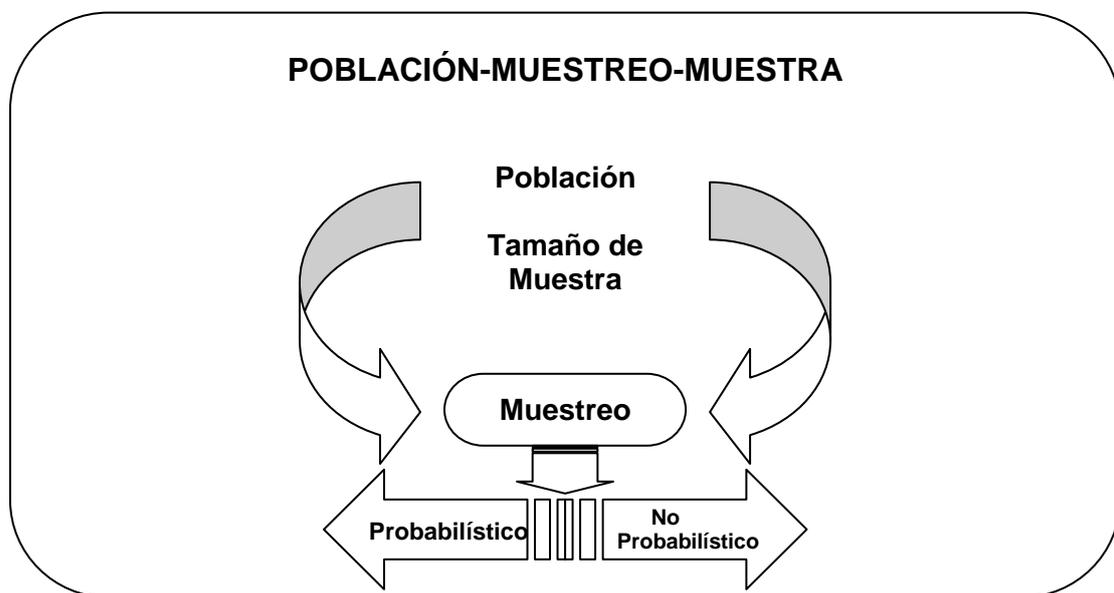
 2.4.1.3 Aleatorio estratificado

 2.4.1.4 Estratificado y por racimos

 2.4.2 No probabilística

- 2.4.2.1 De juicio o criterio
- 2.4.2.2 De diseño de bola de nieve
- 2.4.2.3 Por accidente
- 2.4.2.4 Por conveniencia
- 2.4.2.5 Por cuota
- 2.4.2.6 Por sujetos-tipo
- 2.4.2.7 Por sujetos voluntarios
- 2.4.2.8 Por expertos

Diagrama conceptual



2.1 Conceptos fundamentales

Pretender generar un examen de todos los elementos de una determinada población implicaría demasiados recursos y un tiempo desmedido, es por ello que se recurre al análisis de una pequeña parte, es decir, se realiza un muestreo con el objetivo de establecer generalizaciones con respecto a un grupo total sin tener que examinarlos uno por uno.

La parte del grupo de elementos que se examina recibe el nombre de muestra, y el grupo total a partir del cual se seleccionó la muestra se conoce como población o universo.

Al momento de concebir una idea y entrar en el proceso de construcción de una investigación, invariablemente se tendrá que abordar el punto relativo a la selección de la muestra, ello implica –dice Hernández, Fernández y Baptista (2003)– dos aspectos esenciales: *definir la unidad de análisis* que se refiere a los individuos, instituciones, objetos, agrupaciones, situaciones, eventos, etcétera, es decir, el sobre qué o quiénes se van a recolectar datos, y *delimitar una población*; esto es, establecer las características específicas que tendrá que cumplir una unidad de análisis, respecto a tres prioridades: 1) contenido, 2) lugar y 3) tiempo, esto indica qué va a ser medido, dónde y cuándo.

Conceptos básicos

Población Es la totalidad de un conjunto de elementos, seres u objetos que se pretende investigar.	Muestreo Conjunto de procedimientos que se ejecutan para seleccionar una muestra.	Muestra Es un subconjunto de la totalidad de un conjunto de elementos, seres u objetos que se pretende investigar. Es una fracción o subgrupo de una población.
--	---	---

Al obtener una parte o fracción de una población, el problema principal consiste en asegurar que el subconjunto sea representativo de la población, de manera que permita generalizar a la población los resultados obtenidos de la muestra, de ello depende la forma en que sea extraída la muestra.

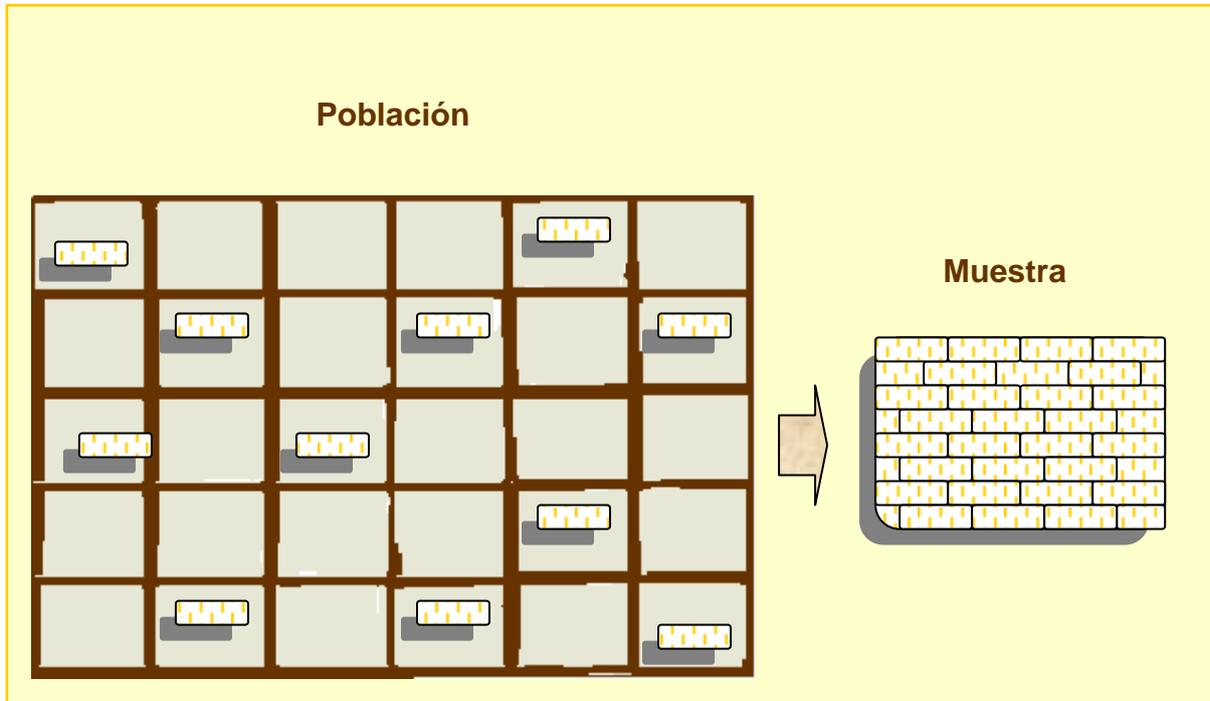
2.2 Cualidades de una buena muestra

Ander Egg (1995) señala que una muestra tiene validez técnico-estadística, si cumple con los siguientes requisitos:

- 1) Ser representativa o reflejo general del conjunto o universo estudiado, reproduciendo lo más exactamente, las características del mismo.
- 2) Que su tamaño sea estadísticamente proporcionado a la magnitud del universo.

- 3) Que el error muestral se mantenga dentro de los límites adoptados como permitidos.

Relación: Población-muestra



En este sentido, el esquema de relación: población-muestra recupera la esencia del concepto de muestra, ya que no se trata sólo de obtener un subconjunto de la población, sino de que tal fracción sea la suma de diversos elementos contenidos en el total de aspectos o características de la población.

2.3 Tamaño de muestra

Cuando un investigador decide analizar un fenómeno social generalmente se encuentra ante la disyuntiva de determinar cuántos sujetos deben conformar su muestra, para ello existe una extensa gama de fórmulas que consideran distintos parámetros, sin embargo, lejos de retomar todas ellas, es preferible abordar en este trabajo un esquema accesible para obtener un subconjunto de una población llamado muestra.

Primer paso

Se calcula un tamaño provisional de la muestra (n') a través de la división de la varianza de la muestra (S^2) entre la varianza de la población (V^2).

n' = Tamaño provisional de la muestra

$$n' = \frac{S^2}{V^2}$$

La **varianza de la muestra (S^2)**: se define como el producto de la probabilidad de ocurrencia y no ocurrencia del fenómeno a medir. Es decir, el multiplicar la proporción del fenómeno que cumple con las características establecidas por el investigador para conformar la muestra por aquella fracción que no las cumple.

Para obtenerla se emplea la siguiente ecuación:

$$S^2 = p (1 - p)$$

La **varianza de la población (V^2)**: se define como el cuadrado del error estándar (Se)². El error estándar es la diferencia entre la media de la población y la media de la muestra. Esta diferencia no debe ser mayor a 0.05, es decir, que de 100 casos, 95 veces la predicción establecida sea correcta. Para obtenerla se emplea la siguiente ecuación:

$$V^2 = (Se)^2$$

Segundo paso

Una vez calculada una muestra provisional es necesario ajustarla o corregirla retomando el tamaño de la población (**N**), mediante la siguiente fórmula:

$$n = \frac{n'}{1 + \frac{n'}{N}}$$

Ejemplo

El Grupo 2426 de Prácticas Comunitarias de la Escuela Nacional de Trabajo Social quiere hacer un estudio sobre violencia familiar hacia la mujer en la colonia “Antonia M. M.”. Establece como población: todas aquellas mujeres mayores de 18 años, solteras, casadas, en unión libre y viudas que pertenezcan a dicha colonia. El número total de mujeres que cumple tales características es de $N= 950$. ¿Cuál es el número de mujeres n que se tienen que entrevistar, si se asume un error estándar de 0.015?

Primer paso

Calcular el tamaño provisional de la muestra (**n'**):

$$n' = \frac{S^2}{v^2}$$

Varianza de la muestra (**S²**):

Para obtener la varianza de la muestra es necesario definir la proporción de la población que cumple con las características establecidas para la misma. En este caso, la proporción de mujeres mayores de 18 años, solteras, casadas y en unión libre que pertenezcan a dicha colonia, es de: $0.65 (65\%)^1$.

¹ Este valor se establece a partir de una estimación que el investigador hace respecto a las características específicas de la población. Se apoya en datos obtenidos por organismos especializados.

$$S^2 = p (1 - p)$$

$$= 0.65 (1 - 0.65) = 0.2275$$

Varianza de la población (V^2):

Su definición es el cuadrado del error estándar $(Se)^2$. En este caso, se asume un error estándar de 0.015.

$$V^2 = (Se)^2$$

$$= (0.015)^2 = 0.000225$$

Sustituyendo:

$$n' = \frac{S^2}{V^2} = \frac{0.2275}{0.000225} = 1011$$

Segundo paso

Ajustar la muestra provisional.

Datos: Tamaño de la población (N): 950 mujeres

Muestra sin ajustar (n'): 1011

$$n = \frac{n'}{1 + \frac{n'}{N}} = \frac{1011}{1 + \frac{1011}{950}} = \frac{1011}{2.0642} = 489.77 = 490$$

Conclusión

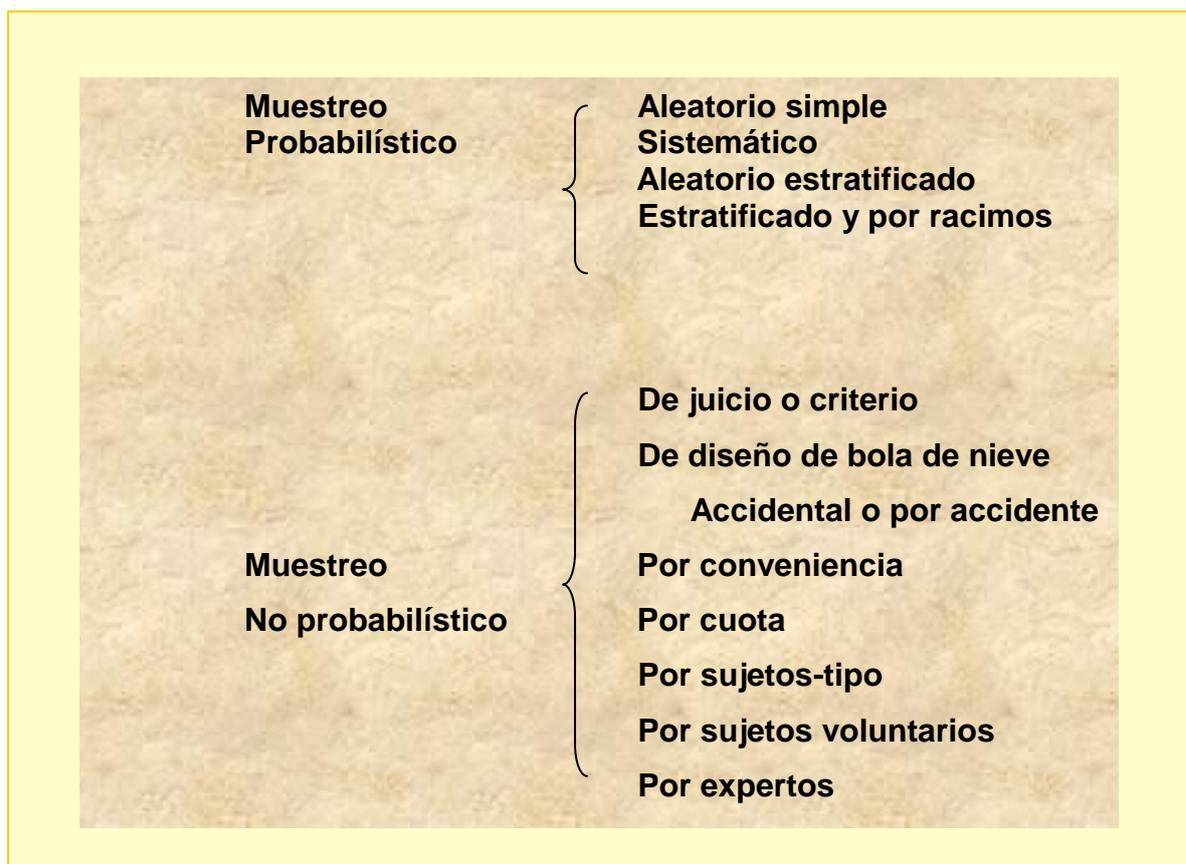
Se necesita una muestra de 490 mujeres.

Una vez que se obtiene el tamaño de la muestra es preciso determinar cómo y dónde seleccionar los sujetos o elementos que deberán integrar ese subgrupo de la población. Para ello, es indispensable analizar los tipos de muestra, objeto del siguiente apartado.

2.4 Tipos de muestreo

Uno de los principales problemas que enfrenta el investigador consiste en elegir un tipo de muestreo adecuado para el estudio de una problemática social; considerando la estructura y los procedimientos de selección, los muestreos se clasifican en: probabilísticos y no probabilísticos.

Tipos de muestreo



2.4.1 Muestreo probabilístico

En este tipo de muestreo, cada unidad muestral de la población en estudio tendrá la misma probabilidad de ser incluida en la investigación. Permite generalizar (con determinado margen de error) en la población total los resultados obtenidos. Selltiz, Wrightsman y Stuart (1980) establecen que la característica esencial del muestreo de probabilidad es que puede especificarse para cada elemento de la población la probabilidad de que irá incluido en la muestra.

Una de las ventajas que presenta este tipo de muestreo, señala Elorza (2000), es que cuenta con modelos para calcular el tamaño de la muestra con una confiabilidad y un grado de error. Establece que en el muestreo probabilístico es viable calcular parámetros insesgados referentes al universo en estudio a partir de poca información sobre el mismo.

Selltiz, Wrightsman y Stuart (1980) y Pick y López (1998) consideran que este tipo de muestreo es el más adecuado e incluso la única modalidad que hace posible planes de muestreo representativo, puesto que reduce al máximo los prejuicios de selección que el investigador pueda tener.

Entre las principales desventajas que presenta este tipo de muestreo se encuentra el alto costo del mismo y la exigencia de mayor tiempo; para algunos investigadores –dice Elorza (2000)- les resulta complejo, ya que no cuentan con conocimientos de estadística inferencial.

Muestreo probabilístico

Es aquél en el que todos los elementos de una población tienen la misma probabilidad de ser elegidos.

2.4.1.1 Aleatorio simple

Pagano (1999) señala que una muestra aleatoria es elegida de una población mediante un proceso que asegura: 1) cada muestra aleatoria de un tamaño dado tenga una misma probabilidad de ser elegida; 2) todos los miembros de la población tengan la misma probabilidad de quedar en la muestra.

Gonick y Smith (1999) establecen al respecto que un muestreo aleatorio simple presenta dos propiedades que lo convierten en un estándar frente al que se miden todos los otros métodos de muestreo:

- 1) Representativa: cada unidad tiene las mismas posibilidades de ser escogida. Dicho en términos estadísticos, es representativa dado que presenta ausencia de sesgo.
- 2) Independencia: la selección de una unidad no influye en la selección de otras unidades.

Un método para obtener este tipo de muestra es colocar en un recipiente todas las unidades de análisis; ya sea, fichas o papeles con nombres o números que correspondan a cada una de ellas y extraer aleatoriamente hasta completar el tamaño de la muestra. Para este tipo de muestreo es necesario conocer la población en estudio, la cual no debe ser muy grande.

Muestreo aleatorio simple

Es el procedimiento en el que un tamaño de muestra n es seleccionado de una población de tamaño N , de tal manera que cada muestra posible de tamaño n tiene la misma posibilidad de ser seleccionada.

Cuando la población es muy grande, este método de selección aleatorio es difícil de aplicar. Para atender este problema es necesario numerar en forma progresiva cada uno de los sujetos en estudio y después seleccionarlos mediante el empleo de una tabla de números aleatorios.

Ejemplo:

Para ilustrar el uso de la tabla de número aleatorios y el procedimiento para obtener una muestra aleatoria se empleará el siguiente ejemplo:

Problema: Se tiene una población de 45 alumnos y se desea obtener una muestra de $n= 10$, para conocer su opinión acerca de las enfermedades de transmisión sexual.

Primer paso

Elaborar una lista de los alumnos enumerándolos del 1 al 45.

Segundo paso

De modo arbitrario se selecciona una parte de la tabla de números aleatorios, una columna y un renglón.

Supóngase que se seleccionó la columna 3 y el renglón 2, lo que presenta como resultado que se iniciará a seleccionar desde el número 27 (se puede comenzar a buscar valores que integran la muestra, ya sea, hacia abajo, hacia arriba, al lado izquierdo o derecho e incluso de manera diagonal).

En este caso, se considera los valores que se encuentran hacia abajo es decir, 27, 47, 04, 36, 27, 18, 24, 87, 41, 37, 41, 09, 65, 68, 62, 55, 93, 06, 37, 75, y 08.

De esta serie de números se omiten los valores mayores a la población es decir, mayores de 45; así como, los valores repetidos.

Tercer paso

Obtener las unidades muestrales resultantes del paso anterior, es decir, hay que entrevistar a los alumnos que ocupan los lugares: 27, 04, 36, 18, 24, 41, 37, 09, 06 y 08, que conforman el total de la muestra $n= 10$.

Tabla de números aleatorios (segmento)

Renglón	Columnas	1	2	3	4	5
1		53	95	67	80	79
2		62	12	27	41	05
3		90	16	47	72	20
4		10	59	04	76	80
5		32	17	36	64	08
6		54	71	27	89	41
7		10	60	18	77	34
8		42	20	24	36	78
9		73	55	87	48	49
10		21	56	41	23	58
11		09	60	37	99	06
12		63	26	41	08	21
13		98	72	09	45	69
14		87	89	65	22	98
15		05	91	68	44	67
16		75	93	62	49	95
17		76	15	55	38	29
18		26	76	93	84	08
19		08	35	06	83	76
20		59	73	37	06	26
21		87	94	75	45	72
22		05	74	08	91	37
23		49	82	39	40	51
24		02	25	92	97	41
25		59	41	49	100	13

2.4.1.2 Sistemático

Consiste en la selección a intervalos de los individuos. Una vez que se tiene el tamaño de la población (N) y el tamaño de la muestra (n) el siguiente paso es determinar mediante una división de los dos valores anteriores, el número de selección sistemática (k), es decir, el “tamaño del salto o paso” a considerar para obtener el tamaño de muestra que se desea.

Muestreo sistemático

Es la selección de cada *k*ésimo, nombre de una lista de población después de seleccionar aleatoriamente el punto de inicio.

El muestreo sistemático se obtiene mediante la siguiente fórmula:

$$k = \frac{\text{Tamaño de la población}}{\text{Tamaño de la muestra}}$$

Una vez identificado el valor k (tamaño del paso o “tamaño del salto”), de un listado se elige un “caso inicial” según lo indique el valor de k , a partir del cual se contará según el tamaño del paso.

Ejemplo:

En la Escuela Nacional de Trabajo Social se quiere hacer un estudio a los alumnos de tercer semestre respecto a sus necesidades de formación extracurricular. Se sabe que se tiene una población (N) de 720 y al calcular la muestra (n) resultó igual a 286. De tal manera que al calcular k se tiene:

$$k = \frac{\text{Tamaño de la población}}{\text{Tamaño de la muestra}}$$
$$k = \frac{720}{286} = 2.51, \text{ redondeado} = 3$$

De la lista numerada de los 720 alumnos se cuenta tres (dado que es el “tamaño del paso o salto”) y se considera al estudiante que ocupa el lugar 3 para aplicarle un instrumento de investigación; se cuenta nuevamente tres y se considera al número 6, después al 9, luego al 12, posteriormente al 15, etcétera, hasta obtener los 286 alumnos que considera la muestra.

2.4.1.3 Aleatorio estratificado

Cuando se conocen las propiedades de una población, y el objetivo es mejorar la representatividad de una muestra es posible aplicar un muestreo estratificado, es decir, generar estratos o categorías que se presenten en la población y que sean relevantes para los objetivos del estudio, el número de estratos no debe ser muy grande y si se necesitan estimaciones de las subdivisiones, entonces se requiere un mayor número de estratos.

Muestreo aleatorio estratificado

Es la selección unidades de muestra a partir de dividir en subgrupos o estratos a la población de estudio y extraer proporcionalmente de cada uno de ellos, los elementos que integrarán la muestra.

Para obtener un tamaño de muestra por cada estrato, se tienen que seguir dos pasos específicos:

Primer paso

Obtener la proporción definida por el tamaño de la muestra (n) entre total de la población (N); mediante la siguiente formula:

$$\text{Proporción} = \frac{n}{N}$$

Segundo paso

- a) Obtener el tamaño de muestra en cada uno de los estratos; multiplicando la proporción obtenida en el primer paso por el total de la población por estrato, mediante la siguiente fórmula:

$$\text{Tamaño de la muestra por estrato} = \frac{\text{Población total por estrato}}{\text{Población total}} \times \text{Proporción}$$

Ejemplo:

Para explicar este tipo de muestreo se retoma el ejemplo utilizado para obtener el tamaño de muestra. De esta manera se tiene el siguiente problema:

El Grupo 2426 de Prácticas Comunitarias de la Escuela Nacional de Trabajo Social quiere hacer un estudio sobre violencia familiar hacia la mujer en la colonia "Antonia M. M.". Establece como población: todas aquellas mujeres, mayores de 18 años, solteras, casadas, en unión libre y viudas que pertenezcan a dicha colonia. El número total de mujeres que cumple tales características es de $N= 950$. ¿Cuál es el número de mujeres n que se tiene que entrevistar, si se asume un error estándar de 0.015?

Como puede corroborarse en el apartado de *tamaño de la muestra* el resultado de dicho cálculo fue $n= 490$. Ante ello, el Grupo 2426 decide aplicar un muestreo estratificado, partiendo de los siguientes datos, ya recabados por ellos mismos:

Tabla de referencia de datos

No. de estrato	Nombre del estrato, según estado civil de las mujeres	Población total por estrato	Tamaño de muestra por estrato
1	Solteras	420	?
2	Casadas	242	?
3	Unión libre	183	?
4	Viudas	105	?
		N = 950	n = 490

Primer paso

Obtener la proporción definida por el tamaño de la muestra (n) entre el total de la población (N); mediante la siguiente formula:

$$\text{Proporción} = \frac{n}{N} = \frac{490}{950} = 0.5157$$

Segundo paso

Obtener el tamaño de muestra por estrato, en cada uno de ellos; multiplicando la proporción obtenida en el primer paso por el total de la población por estrato, mediante la siguiente formula:

Tamaño de la

muestra por estrato	Población total = por estrato	X	Proporción
	= 420	X	0.5157 = 217
	= 242	X	0.5157 = 125
	= 183	X	0.5157 = 94
	= 105	X	0.5157 = 54

En resumen, se tiene una muestra estratificada con los siguientes datos:

Tabla de resultados

No. de estrato	Nombre del estrato, según estado civil de las mujeres	Población total por estrato	Tamaño de muestra por estrato
1	Solteras	420	217
2	Casadas	242	125
3	Unión libre	183	94
4	Viudas	105	54
		N = 950	n = 490

Para definir a los sujetos a quienes se les tiene que aplicar un instrumento, es conveniente recurrir a un muestreo aleatorio simple en cada tamaño de muestra por estrato, es decir, aplicar un primer muestreo simple a 217 mujeres solteras, un segundo, a 125 mujeres casadas, así sucesivamente.

2.4.1.4 Estratificado y por racimos

Hernández, Fernández y Baptista (2003) señalan que cuando el investigador se ve limitado por recursos financieros, por tiempo, distancias geográficas o por una combinación de éstos y otros obstáculos, se recurre al muestreo por racimos.

Muestreo estratificado y por racimos

Es la selección de sujetos o elementos muestrales inmersos en conjuntos, grupos o estratos definidos física o geográficamente.

Entre las ventajas que se pueden atribuir al muestreo estratificado y por racimos se encuentran su bajo costo y reducido empleo de tiempo y esfuerzo.

Este tipo de muestreo, basa su desarrollo en la extracción de unidades de análisis que se encuentran encapsuladas, agrupadas o encerradas en determinados lugares físicos o geográficos denominados racimos.

El muestreo estratificado y por racimos –señala Hernández, Fernández y Baptista (2003)– implica diferenciar entre la unidad de análisis (es decir, quiénes van a ser medidos o a quién se le va aplicar el instrumento de medición) y la unidad muestral (que se refiere al racimo a través del cual se logra el acceso a la unidad de análisis).

Para obtener un muestreo estratificado y por racimos, es necesario cumplir lo siguiente:

Primer paso

Determinar el tamaño de la muestra a partir del número de racimos (en este caso cuadras).

Segundo paso

Seleccionar los racimos y aplicar muestreo estratificado.

Tercer paso

Dentro de los racimos seleccionados, elegir los sujetos de estudio que van a medirse.

Ejemplo:

El Grupo 2426 de Prácticas Comunitarias de la Escuela Nacional de Trabajo Social quiere hacer un estudio en la colonia “Antonia M. M.”, sobre el manejo de la basura que se genera en casa. Establece como población: todas aquellas personas mayores de 18 años que pertenezcan a dicha colonia.

El grupo en referencia desconoce el número de personas con las características señaladas. Sin embargo, deciden generar una muestra que abarque todos los sujetos mayores de edad, de dicha colonia.

Dadas las condiciones a las que se hace mención, el Grupo 2426 recurre a la selección de una muestra estratificada y por racimos, para ello, se hace llegar de un mapa de la Colonia “Antonia M. M.” el cual indica que hay 145 cuadras.

En este caso, las cuadras se utilizan como racimos, es decir, como unidades muestrales, a partir de las cuales se obtienen las unidades de análisis o sujetos de estudio.

Primer paso

Determinar el tamaño de la muestra a partir del número de racimos (en este caso cuadras).

¿Cuántas cuadras se necesitan tomar como muestra, de una población total de 145, si se quiere asumir un error estándar de 0.015 y una probabilidad de ocurrencia de 0.50?

Fórmula para calcular el tamaño de la muestra provisional o sin ajustar (n’):

$$n' = \frac{S^2}{V^2}$$

Varianza de la muestra (S^2):

$$S^2 = p (1 - p)$$

$$= 0.50 (1 - 0.50) = 0.25$$

Varianza de la población (V^2):

$$V^2 = (Se)^2$$

$$= (0.01)^2 = 0.0001$$

Sustituyendo:

$$n' = \frac{S^2}{V^2} = \frac{0.25}{0.0001} = 2500$$

Ajuste de la muestra provisional.

Datos: Tamaño de la población (N): 145 cuadras

Muestra sin ajustar (n'): 2500

$$n = \frac{n'}{1 + \frac{n'}{N}} = \frac{2500}{1 + \frac{2500}{145}} = \frac{2500}{18.2413} = 137.05 = 137$$

Conclusión:

Se necesita una muestra de 137 cuadras.

Segundo paso

Seleccionar los estratos y aplicar muestreo estratificado.

El Grupo 2426 de Prácticas Comunitarias de la Escuela Nacional de Trabajo Social, conoce de acuerdo a estudios previos que la colonia “Antonia M. M.” que ha sido dividida de acuerdo con cuatro estratos por nivel educativo, que agrupan a la población según el número de años cursados en un sistema educativo. La relación estrato-número de cuadras se presenta a continuación:

Relación estrato-número de cuadras

Número de Estratos	Número de cuadras o racimos
1	56
2	42
3	29
4	18
Total	145

Estratificación de la muestra

Datos: Tamaño de la muestra: $n = 137$

Tamaño de la población: $N = 145$

Fórmula para estratificar la muestra:

$$\text{Proporción} = \frac{n}{N} = \frac{137}{145} = 0.9448$$

Fórmula para obtener el tamaño de muestra por estrato:

Tamaño de la

muestra por	Número de			
estrato y racimo	= racimos por estrato	X	Proporción	
	= 56	X	0.9448	= 53
	= 42	X	0.9448	= 40
	= 29	X	0.9448	= 27
	= 18	X	0.9448	= 17

Tercer paso

Dentro de los racimos seleccionados, elegir las unidades de análisis que van a medirse. En este apartado se determina en primera instancia el número de hogares-sujetos de estudio por cada racimo (o cuadra)² y se multiplica por el tamaño de la muestra por estrato y racimo para obtener el total de hogares o unidades de análisis por estrato a quienes se les aplicará el instrumento.

Tabla de resultados

Número de estratos	Número de cuadras o racimos por estrato	Tamaño de la muestra por estrato y racimo	Número de hogares-sujetos de estudio por cuadra o racimo	Total de hogares por estrato
1	56	53	23	1219
2	42	40	15	600
3	29	27	19	513
4	18	17	25	425
TOTAL	N = 145	n = 137	60	2757

³ Este valor se obtiene de por medio de la numeración y conteo de los hogares-sujetos de estudio por cuadra o racimo, de acuerdo al mapa con que se cuenta de la Colonia "Antonia M. M."

Las unidades de análisis u hogares por estrato se deben seleccionar aleatoriamente para asegurar que cada uno de ellos tenga la misma probabilidad de ser elegido. Es decir, se enumeran los hogares del primer estrato y se aplica un muestreo aleatorio simple, a la misma situación se deben someter los demás estratos.

2.4.2 Muestreo no probabilístico o determinístico

Selltiz, Wrightsman y Stuart (1980) establecen que en el muestreo de no probabilidad, no probabilístico o determinístico no existe forma de estimar la probabilidad que cada elemento tiene de ser incluido en la muestra, y no hay tampoco seguridad de que cada elemento tenga alguna posibilidad de ser incluido.

Muestreo no probabilístico o determinístico

Es aquél en el que la elección de los elementos no depende de la probabilidad, sino de las características determinadas por el investigador o por quien hace la muestra.

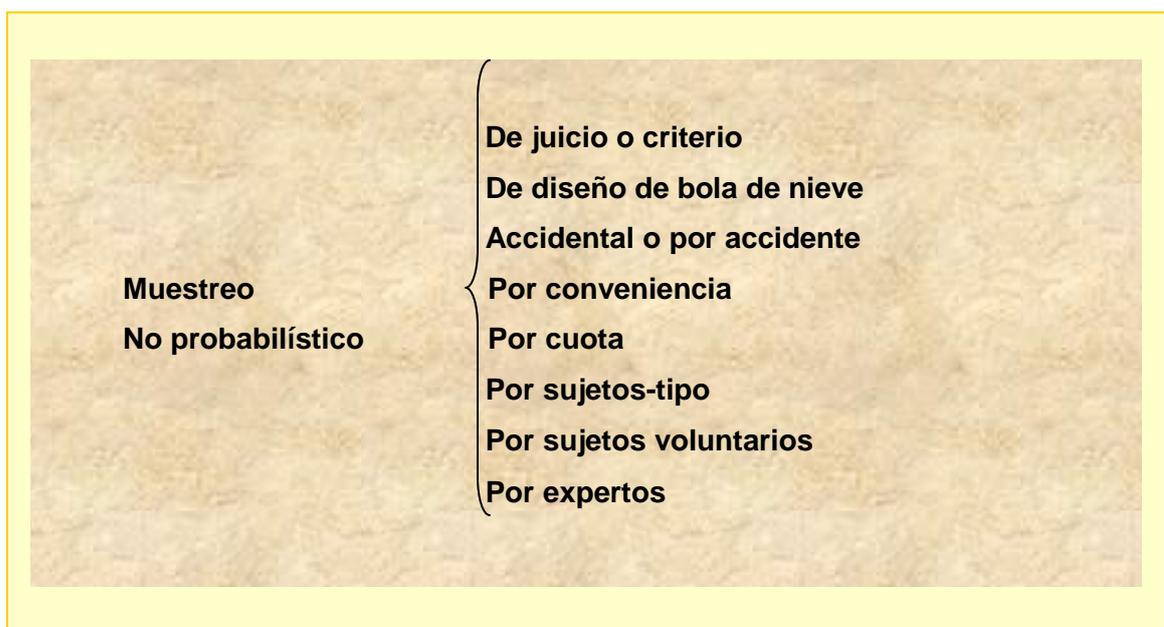
Hernández, Fernández y Baptista (2003) respaldan tal aseveración al afirmar que al ser muestras no probabilísticas, no es posible calcular con precisión el error estándar, es decir, no se puede calcular con qué nivel de confianza se hace una estimación.

El muestreo de tipo no probabilístico presenta como principal desventaja el tener un valor limitado y relativo para la muestra en sí, y nulo para la población. Es decir, no existe sustento estadístico para poder hacer generalizaciones a una población, dado que, en las muestras de este tipo, los sujetos de una población determinada no tienen la misma probabilidad de ser elegidos sino que dependen de las características o condiciones que establezca el investigador o grupo de encuestadores, para generar una muestra.

Una de las ventajas que presenta este tipo de muestreo es su bajo costo y escaso empleo de tiempo. Por otra parte, en este tipo de muestreo el investigador no requiere conocimientos avanzados de estadística.

Por su parte, Hernández, Fernández y Baptista (2003) establecen que una muestra no probabilística es útil para determinado diseño de estudio que requiere no tanto de una “representatividad” de elementos de una población, sino una cuidadosa y controlada elección de los sujetos con ciertas características según se estipule en el planteamiento del problema.

En este sentido, –continúan– al no interesar tanto la posibilidad de generalizar los resultados, las muestras no probabilísticas o dirigidas son de gran valor, pues logran –si se procede cuidadosamente y con una profunda inmersión inicial en el campo– obtener los casos (personas, contextos, situaciones) que interesan al investigador y que llegan a ofrecer una gran riqueza para la recolección y el análisis de datos.



2.4.2.1 De juicio o criterio

En este tipo de muestreo tanto el tamaño de la muestra como la elección de los elementos que la integran están sujetos al juicio del investigador basado generalmente en su experiencia. Haroldo Elorza (2000) señala que este tipo de muestreo es útil y aún aconsejable: cuando el muestreo probabilístico no es factible o resulta costoso.

Muestreo de juicio o criterio

Es aquél en el que la conformación de la muestra depende del criterio, juicio o discernimiento del investigador quien se apoya generalmente en su praxis profesional.

Ejemplo

Un estudiante de la ENTS desea conocer el tipo de adicción a drogas tanto legales como ilegales que presentan los niños de la calle. Resulta difícil obtener una lista de niños de la calle y plantear una muestra aleatoria simple. Por ello, es conveniente una muestra de juicio y acudir a aplicar un instrumento a los niños de la calle que se reúnen en la estación del Metro Observatorio en la ciudad de México.

2.4.2.2 De diseño de bola de nieve

Este tipo de diseño resulta conveniente cuando se necesita alcanzar poblaciones pequeñas o datos especializados.

Earl Babbie (2000) establece que para llevar a cabo este tipo de muestreo, se reúnen los datos de los pocos miembros que se conozcan de una población objetivo que se pueda localizar y se le pide la información necesaria para ubicar a otros miembros que conozcan de esa población.

Muestreo de diseño de bola de nieve

Es aquél en el que la configuración de la muestra se construye a partir de la referencia de un sujeto a otro, es decir, de la acumulación de elementos a través de la remisión del investigador de un individuo a otro y de este a otro.

La denominación de bola de nieve se refiere a la acumulación que resulta de que cada sujeto localizado proponga a otro u otros. Debido a que este procedimiento da por resultado muestras de representatividad cuestionable, se usa sobre todo con fines exploratorios.

Ejemplo:

Un grupo de estudiantes de Trabajo Social de la UNAM desean realizar un estudio sobre las condiciones socio-laborales de los inmigrantes mexicanos a EU que pertenecen al municipio de Atotonilco de Tula en el Estado de Hidalgo, una vez que han regresado a su lugar de origen.

Como es difícil contar con una lista de personas que cumpla con tales características y aplicar un procedimiento de muestreo de tipo probabilístico se recurre a un muestreo no probabilístico de carácter bola de nieve, es decir, se localiza a un inmigrante que cumpla con las características para aplicarle un instrumento, al mismo se le pregunta de otra persona que cumpla con el perfil y a este último se le cuestiona sobre otro sujeto de estudio, etcétera. En este sentido, la bola de nieve crecería en la medida que los encuestados señalen a otra persona y ésta a otra.

2.4.2.3 Accidental o por accidente

Castañeda, De la Torre, Morán y Lara (2002) señalan que este tipo de muestreo consiste en seleccionar de manera arbitraria a los individuos que

conformarán la muestra. Radica en aprovechar o utilizar para el estudio a las personas disponibles en un momento dado según lo que interese estudiar.

Muestreo accidental o por accidente

Es aquél en el que el investigador decide por voluntad propia qué elementos integran la muestra aprovechando los sujetos de que dispone.

Selltiz, Wrightsman y Stuart (1980) señalan que en este tipo de muestra, el investigador sólo puede desear que la equivocación no sea demasiado grande.

Ejemplo

Un alumno de la ENTS-UNAM desea conocer cuál es la opinión de sus compañeros respecto a la forma de impartir la clase de sus profesores de asignatura. En este sentido, aplica un instrumento a los compañeros que asistieron en un día determinado y obtiene resultados.

2.4.2.4 Por conveniencia

Haroldo Elorza (2000) señala que este tipo de muestreo consiste sencillamente en ponerse en contacto con la unidad muestral apropiada. Este tipo de muestreo se utiliza para pruebas piloto o para probar un cuestionario.

Muestreo por conveniencia

Es aquél en el que el investigador decide de acuerdo con los objetivos del estudio los elementos integrarán la muestra.

Ejemplo

Un profesor universitario desea introducir un nuevo método de enseñanza de la asignatura de Estadística. Con base en su experiencia decide aplicar un instrumento a estudiantes de las licenciaturas en Trabajo Social, Psicología y Medicina. Todo ello a conveniencia del investigador.

2.4.2.5 Por cuota

En el muestreo por cuota se determina una cantidad (cuota) de individuos de una población para que sean miembros de la muestra. El investigador selecciona una muestra considerando algunos fenómenos o variables a estudiar como sexo, raza, religión, etcétera.

Muestreo por cuota

Es aquél en el que el investigador establece una cantidad denominada cuota que obtiene de una categoría o estrato para conformar la muestra.

Pick y López (1998) señalan que para utilizar el muestreo de cuota primero se tiene que conocer la población que se pretende estudiar y hacer una clasificación por estratos de acuerdo con los objetivos del estudio. Consideran que no es necesario que se tome una proporción representativa de cada estrato (aunque se tiene que recoger un porcentaje mínimo de cada estrato). Una vez que se deciden los estratos, el investigador elige arbitrariamente los sujetos de estudio que integrarán cada uno de los estratos.

Por lo tanto, –dice Pick y López (1998) – el muestreo de cuota constituye un método de muestreo estratificado en el cual la selección dentro de los estratos no es al azar, sino accidental. Lo anterior –establecen– representa la principal desventaja.

Ejemplo

Un profesor de la ENTS-UNAM desea investigar cuáles son las necesidades de formación extracurricular de los estudiantes de Trabajo Social de la UNAM. Decide generar estratos por semestre par y considerar un 25% de cada uno de ellos; obtiene los siguientes resultados:

Semestre	Población	Porcentaje a extraer	Muestra
Segundo	100	25	25
Cuarto	96	25	24
Sexto	80	25	20
Octavo	76	25	19
Total	352	100	88

El total de estudiantes que integran la muestra es de 88.

2.4.2.6 Por sujetos-tipo

Hernández, Fernández y Baptista (2003) apuntan que el muestreo por sujetos-tipo es aquél cuyo objetivo es la riqueza, profundidad y calidad de la información, no la cantidad ni la estandarización. Se utiliza en estudios de tipo exploratorio así como en investigaciones de carácter cualitativo como es el caso de los estudios con perspectiva fenomenológica.

Muestreo por sujetos-tipo

Es aquél en el que el investigador elige elementos de muestra que comparten las mismas características (políticas, económicas, sociales, académicas, etc.) para generar información de tipo cualitativo más que cuantitativo.

Ejemplo

Un investigador está interesado en conocer de manera directa cuál es la percepción de los estudiantes respecto al desempeño de los profesores de la ENTS-UNAM en el salón de clase; para ello, forma quince grupos de diez alumnos, al mismo tiempo programa una sesión por cada grupo, con el objetivo de que expresen sus puntos de vista respecto a la variable a medir bajo indicadores de: actitudes, motivación, inconformidad, liderazgo, etcétera.

El investigador sistematiza la información y obtiene resultados.

2.4.2.7 Por sujetos voluntarios

Se trata de muestras ocasionales donde los sujetos voluntariamente acceden a participar en un estudio. Se procura que los sujetos sean homogéneos en variables tales como edad, sexo, ocupación, de manera que los resultados o efectos no obedezcan a diferencias individuales, sino a condiciones a las que fueron sometidos.

Muestreo por sujetos voluntarios

Es aquél en el que las unidades muestrales o elementos a considerar deciden libremente formar parte o no de la muestra.

Ejemplo

Un profesor desea probar dos métodos de enseñanza en estadística. Para ello requiere integrar dos grupos. El académico decide convocar de manera abierta a estudiantes universitarios.

Quienes estén interesados y acudan al llamado formarán parte de la muestra. En este caso, la elección de los individuos sujetos de estudio depende de circunstancias fortuitas.

2.4.2.8 Por expertos

Se trata de estudios basados en la opinión de sujetos expertos en un tema determinado. Estas muestras son frecuentes en estudios de corte cualitativo y exploratorio con el objetivo de generar hipótesis más precisas o como base para la construcción de instrumentos de medición.

Muestreo por expertos

Es aquél en el que la selección de los elementos muestrales depende de la experiencia, reconocimiento o prestigio que una persona pueda tener respecto a un tema determinado.

Ejemplo

Un investigador de la Escuela Nacional de Trabajo Social desarrolla un estudio sobre pobreza en México y recurre a una muestra de $n=15$ expertos. Todos, sujetos idóneos y de reconocido prestigio para abordar y discutir sobre tipología, índices, factores, organismos que la combaten, etcétera.

Resumen de la unidad

La presente unidad parte del estudio de conceptos básicos relacionados con población y muestra para con ello profundizar en la obtención de un tamaño de muestra y en la aplicación de métodos de muestreo bajo dos vertientes: lo probabilístico y lo no probabilístico.

El muestreo probabilístico parte de la idea que todos los elementos muestrales tienen la misma probabilidad de ser elegidos, el concepto de no muestreo probabilístico en cambio asume que las características de la muestra serán determinados de acuerdo a criterios del investigador.

Unidad III

Método estadístico

Introducción

En esta unidad se abordan los conceptos básicos del método estadístico principalmente lo relacionado con sus etapas y la importancia de su seguimiento secuencial en relación con el proceso de investigación social.

El método estadístico aporta una referencia lógica para el tratamiento de los datos apoyada en cinco pasos fundamentales y secuenciales con un grado de complejidad mayor a medida que avanza el proceso.

Objetivo

Al finalizar la presente unidad emplearás el método estadístico como base para el estudio de los problemas sociales con el propósito de descubrir las características esenciales que los determinan.

Temario

3.1 Concepto y ubicación en el contexto de la investigación social

3.2 Etapas

3.2.1 Recolectar

3.2.2 Contar

3.2.3 Presentar

3.2.4 Describir

3.2.5 Analizar

Diagrama conceptual



3.1 Concepto y ubicación en el contexto de la investigación social

En el estudio de los fenómenos sociales la investigación y la estadística convergen a través de un método específico y claro que permiten asumir una secuencia sistemática de pasos para ir más allá de simplemente conocer los problemas, es decir, tener bases firmes para la toma de decisiones, en este sentido, es importante generar una relación y secuencia lógica entre ambas disciplinas.

Por otra parte, es conveniente establecer que en esta relación, la estadística cumple una función fundamental al analizar datos procedentes de la investigación social específicamente si se considera el siguiente proceso:

- 1) Concebir la idea a investigar.
- 2) Plantear el problema de investigación: establecer objetivos; preguntas y justificación.
- 3) Elaborar el marco teórico: revisión, obtención y consulta de la literatura; extracción y recopilación de la información de interés y construcción del marco teórico.

- 4) Definir si la investigación se inicia como exploratoria, descriptiva correlacional o explicativa y hasta que nivel se llegará.
- 5) Establecer hipótesis: detectar las variables; definir conceptualmente y operacionalmente las variables.
- 6) Seleccionar el diseño apropiado de investigación: diseño experimental; preexperimental o cuasiexperimental.
- 7) Selección de la muestra: determinar el universo y extraer la muestra.
- 8) Recolección de los datos: elaborar el instrumento de medición y aplicarlo; calcular validez y confiabilidad del instrumento; codificar los datos.
- 9) Analizar los datos: seleccionar las pruebas estadísticas; elaborar el problema de análisis y realizar el mismo.
- 10) Presentar los resultados: elaborar el reporte de investigación y presentar el reporte de investigación.

Como se puede apreciar la estadística se enmarca dentro de un proceso de investigación social, específicamente una vez que se han obtenido los datos.

La estadística proporciona elementos teóricos y prácticos que permiten estudiar los fenómenos sociales, ayuda a organizar los esfuerzos por comprender la problemática social. Su principal apoyo es el método estadístico, es decir, la secuencia sistemática de procedimientos para el manejo de los datos derivados de la investigación social con el propósito de comprobar hipótesis.

El método estadístico es el conjunto de procedimientos que recolecta, cuenta, presenta, describe y analiza los datos, a fin de extraerles el máximo de información, base para la toma de decisiones.

3.2 Etapas

El método estadístico como conjunto de procedimientos enmarca cinco acciones secuenciales: 1) recolectar, 2) contar, 3) presentar, 4) describir y 5) analizar.

3.2.1 Recolectar

En esta etapa se estudia un concepto fundamental para la estadística y la investigación social tal es el caso de las variables desde su tipología hasta sus niveles de medición, es este sentido, se entiende por **recolectar** a la etapa donde se acopia la información cualitativa y cuantitativa a través de variables estableciendo su nivel de medición.

3.2.2 Contar

En esta etapa se contabiliza los valores obtenidos de cada una de las variables para determinar su frecuencia. Se entiende por **contar** al procedimiento donde los datos son sometidos a revisión, clasificación y cómputo numérico.

3.2.3 Presentar

Al generarse el conteo es preciso elaboran cuadros o tablas, así como gráficos que permitan una inspección precisa y rápida de los datos, ello implica **presentar**.

El presentar implica establecer la diversidad de frecuencias, las normas mínimas para la presentación de tablas o cuadros y gráficas así como su tipología que deberá coincidir con el nivel de medición de la o las variables a mostrar.

3.2.4 Describir

En esta etapa la información es resumida en forma de medidas que permiten expresar de forma sintética las principales propiedades numéricas de los datos.

Describir implica el cálculo de medidas estadísticas que representan las características generales de un estudio basado en una muestra; incluye

medidas de tendencia central, de posición, de dispersión, de correlación y de distribución.

3.2.5 Analizar

En esta etapa, mediante fórmulas estadísticas y el uso de tablas específicamente diseñadas, se efectúa la comparación de las medidas de resumen previamente calculadas.

Analizar o inferir implica proporcionar los métodos para estimar las características de un grupo total (población), basándose en datos de un conjunto pequeño (muestra) de observaciones.

Resumen de la unidad

En la presente unidad se estudia el concepto de método estadístico como un elemento indispensable para estudio de la estadística y su relación con la investigación social, dicho método incluye cinco pasos lógicos y secuenciales, a saber: 1) contar, donde los datos son sometidos a revisión, clasificación y cómputo numérico; 2) recolectar, donde se acopia la información a través de variables; 3) presentar, que es la generación de tablas o cuadros permite mostrar la distribución de los datos; 4) describir, donde los datos se resumen en forma de medidas que permiten expresar sus propiedades principales y 5) analizar o inferir, donde se determina si los resultados de datos muestrales pueden ser generalizados a una población.

Unidad IV

Recolección de datos

Introducción

En la presente unidad se aborda la materia prima de la estadística, es decir, las variables, se parte de su conceptualización, clasificación y niveles de medición, ello te permitirá valorar por un lado, el tipo de variable que quieres obtener y por el otro cual será el proceso para poder medirla.

Objetivo

Al finalizar la presente unidad el alumno profundizará sobre el concepto de variable, clasificación y niveles de medición con el propósito de generar en él la capacidad de identificar las características estadísticas de la misma.

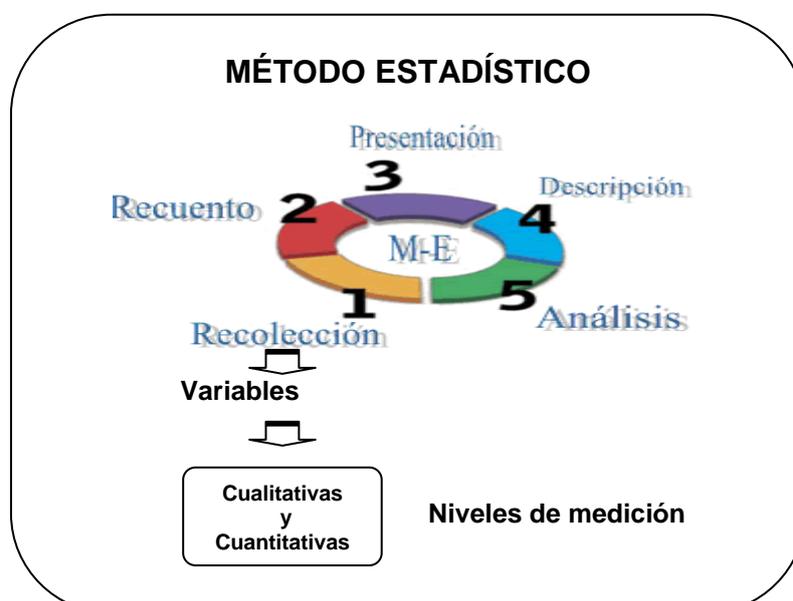
Temario

4.1 Variables

4.1.1 Clasificación

4.1.2 Niveles de medición

Diagrama conceptual



4.1 Variables

Una variable es una propiedad que puede variar y cuya variación es susceptible de medirse. Son características, cualidades, propiedades o atributos que pueden adoptar diferentes valores, magnitudes o intensidades en los diversos sujetos en que se miden.

Una variable es medida utilizando una escala de medición. La elección de la(s) escala(s) de medición a utilizar depende, en primer lugar, del tipo de variable en estudio, y, además, del manejo estadístico a la que se someterá la información. En términos prácticos, existe una correspondencia directa entre el concepto de variable y escala de medición.

4.1.1 Clasificación

La naturaleza de los datos es de gran importancia a la hora de elegir el método estadístico más apropiado para abordar su análisis. Con este fin, las variables se clasifican estadística y metodológicamente. Las primeras en consideración a su nivel de medición, las segundas en razón de un orden de precedencia.

Estadísticamente o por su nivel de medición, las variables se clasifican en: cuantitativas y cualitativas.

Variables cualitativas. Este tipo de variables representan una cualidad o atributo que clasifica a cada caso en una de varias categorías. Estas a su vez se clasifican en nominal u ordinal.

El nivel cualitativo implica la asignación de una característica o categoría que representa una cualidad determinada o asignada a una variable.

Variables cuantitativas. Son las variables que pueden medirse, cuantificarse o expresarse numéricamente.

En el nivel cuantitativo, medir significa además de asignar un atributo a una unidad de análisis saber “cuánto” mayor o menor esta una escala de otra, es decir, especifica la distancia o intervalo entre valores (el valor 70 es el doble del valor de 35).

Metodológicamente o por orden de precedencia, las variables se clasifican en: independiente y dependiente.

Variable independiente: es la variable manipulada (el predictor) para determinar sus efectos (predicciones) sobre la variable dependiente. Variable de un experimento que es controlada en forma sistemática por el investigador.

Variable dependiente: es el resultado o variable criterio que está relacionada con cambios en la variable independiente. Variable en un experimento, medida por un investigador, para determinar el efecto de una variable independiente.

4.1.2 Niveles de medición

Las variables se clasifican en cualitativas o cuantitativas. Las escalas de las variables cualitativas reciben el nombre de “modalidad”, las escalas de las variables cuantitativas reciben el nombre de “valor” o “clase”. En este sentido, una variable no es sino el conjunto de las distintas modalidades o valores o clases definidas por una escala.

Según su nivel de medición o clasificación estadística, las variables cualitativas se dividen en:

Nominales: son aquellas en las que los datos se ajustan por categorías que no mantienen una relación de orden entre sí. Significa simplemente asignar un atributo o característica a una unidad de análisis sin importar jerarquía (color de los ojos, sexo, profesión).

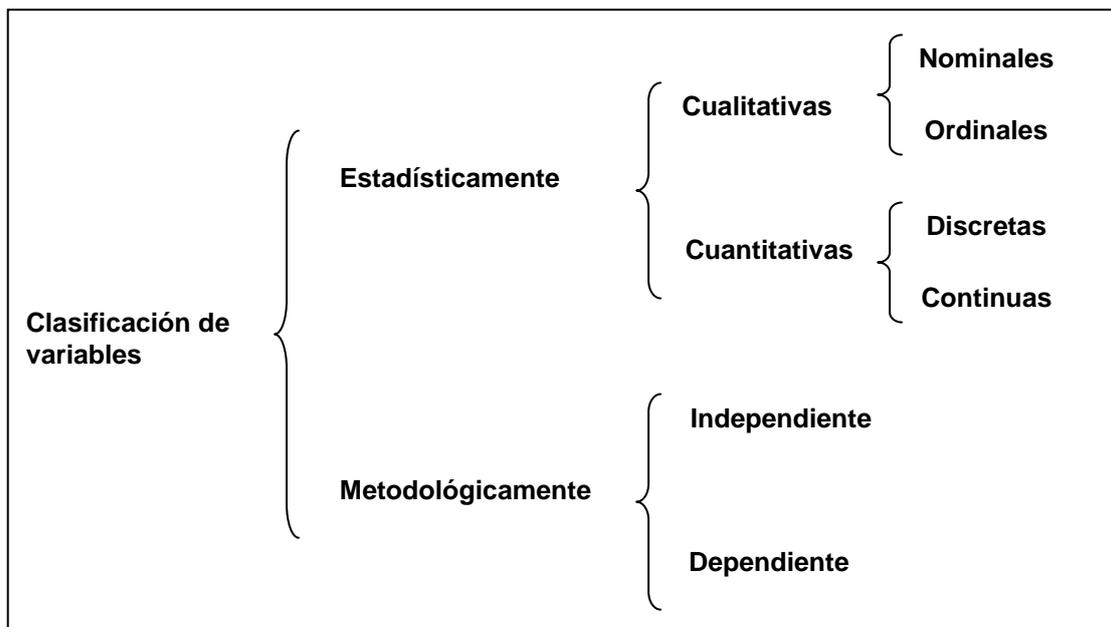
Ordinales: Son aquellas en las que existe un orden o jerarquía entre las categorías. Significa asignar un atributo a una unidad de análisis cuyas categorías pueden ser ordenadas en una serie creciente o decreciente (la categoría ‘secundaria completa’ puede ordenarse en una serie, pues está entre ‘secundaria incompleta’ y ‘universitaria incompleta’). Otros ejemplos son: grados de desnutrición, respuesta a un tratamiento, nivel socioeconómico.

Según su nivel de medición o clasificación estadística, las variables cuantitativas se dividen en:

Discretas: son aquellas que no admiten todos los valores intermedios en un rango. Son el resultado de contar, se suelen tomar solamente valores enteros (número de hijos, número de partos, número de hermanos, etc).

Continuas: son aquellas que admiten cualquier valor dentro de un rango numérico determinado. Son el resultado de medir, y pueden contener decimales (edad, peso, talla). Se pueden subdividir a voluntad. Pueden tomar, entonces, cualquier valor de un determinado intervalo.

CLASIFICACIÓN DE VARIABLES



Resumen de la unidad

En la presente unidad se aborda el concepto de variable así como su clasificación estadística y metodológica con el propósito de establecer los niveles de medición de las mismas.

El nivel de medición de una variable representa la base que define la forma de presentación de una variable así como las medidas estadísticas que se le pueden aplicar.

Unidad V

Recuento y presentación de datos

Introducción

En esta unidad se estudia la forma de conteo de los datos así como los distintos esquemas de presentación de los mismos, es decir, la construcción de cuadros y gráficos según normas mínimas y nivel de medición de las variables se trate de tipo cuantitativo o cualitativo. Ello implica el estudio de conceptos como frecuencia en todas sus modalidades, intervalo, marca de clase y límite.

Objetivo

Al finalizar la presente unidad el alumno empleará métodos de conteo así como formas de presentación de datos con el propósito de contar y presentar variables en razón de su nivel de medición.

Temario

5.1 Conteo de datos

5.2 Presentación de datos

5.2.1 Características mínimas para la creación de tablas o cuadros

5.2.2 Clasificación de datos

5.2.2.1 Intervalo cerrado

5.2.2.2 Intervalo abierto

5.2.2.3 Tamaño del intervalo

5.2.2.4 Límite inferior y superior

5.2.2.5 Marca de clase

5.2.2.6 Límite inferior y superior reales

5.2.2.7 Frecuencia Absoluta

5.2.2.8 Frecuencia Relativa

5.2.2.9 Frecuencia absoluta acumulada

5.2.2.10 Frecuencia relativa acumulada

5.2.3 Características mínimas para la creación de gráficos

5.2.3.1 Diagrama barras separadas

5.2.3.2 Circular, de sectores o de pastel

5.2.3.3 Pictograma

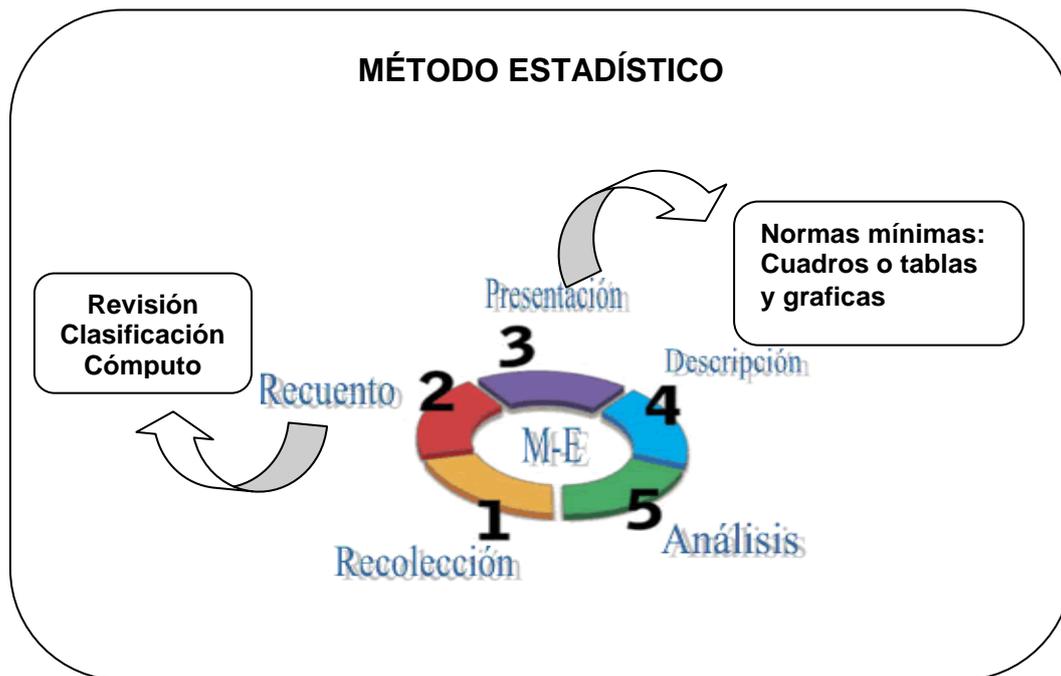
5.2.3.4 Histograma

5.2.3.5 Polígono de frecuencias

5.2.3.6 Polígono de frecuencias acumuladas u ojiva o diagrama de Galton

5.2.3.7 Diagrama de caja

Diagrama conceptual



5.1 Conteo de datos

Este procedimiento implica establecer que frecuencia tienen los datos respecto a cada variable. El conjunto de datos se someten a un proceso de revisión, clasificación y cómputo numérico.

5.2 Presentación de datos

El presentar implica establecer la diversidad de frecuencias, las normas mínimas para crear tablas o cuadros y gráficas así como la tipología que esta en correspondencia con el nivel de medición de la variable a mostrar.

5.2.1 Características mínimas para la creación de tablas o cuadros

Cada cuadro debe tener un título completo que responda a las preguntas:

¿Qué se esta presentando?

¿Cómo se esta presentando?

¿De dónde provienen los datos?

¿De cuándo son los datos?

Las variables deben tener un nombre exacto que este acorde con la escala de clases o modalidades que se hayan definido para dicha variable.

La escala de cada variable debe tener modalidades o clases que sean exhaustivas y mutuamente excluyentes.

En vista de que los cuadros tienen por propósito presentar las características numéricas de todos los datos, estos deben ser exactos.

Ejemplo:

Diagram illustrating the relationship between a title, variable name, and a frequency distribution table.

TÍTULO COMPLETO (PROFESORES SEGÚN ESTADO CIVIL) is linked to the title of the table.

NOMBRE EXACTO DE LA VARIABLE (ESTADO CIVIL) is linked to the variable name in the table.

MODALIDADES MUTUAMENTE EXCLUYENTES Y EXHAUSTIVAS (SOLTEROS, CASADOS, DIVORCIADOS, VIUDOS, UNIÓN LIBRE) are linked to the categories in the table.

The table is titled **PROFESORES SEGÚN ESTADO CIVIL** and is labeled **UNIVERSIDAD NUEVO MUNDO 1986**.

ESTADO CIVIL	NÚMERO	%
SOLTEROS	384	25.5
CASADOS	542	35.9
DIVORCIADOS	266	17.6
VIUDOS	194	12.9
UNIÓN LIBRE	122	8.1
TOTAL:	1.508	100.0

The table is labeled **TABLA DE DISTRIBUCIÓN DE FRECUENCIAS**.

The table shows **FRECUENCIAS RELATIVAS** (Relative Frequencies) and **FRECUENCIAS ABSOLUTAS** (Absolute Frequencies).

5.2.2 Clasificación de datos

La clasificación de datos permite organizar y ordenar lo recabado a partir de un proceso de aplicación de un instrumento de medición.

5.2.2.1 Intervalo cerrado

En el que se conoce tanto su límite inferior como su límite superior.

Ejemplo: Personas de 15 a 20 años de antigüedad y con estatura de 1.60 a 1.75 metros.

5.2.2.2 Intervalo abierto

En el que uno de sus límites se desconoce o se excluye.

Ejemplo: Personas menores de 35 años y estatura mayor de 1.60 metros.

5.2.2.3 Tamaño del intervalo

Es la distancia que presenta las escalas de clases de las variables cualitativas.

5.2.2.4 Limite inferior y superior

Son los límites establecidos según el tamaño del intervalo de cada una de las clases que conforman las variables de tipo cuantitativo discretas o continuas.

5.2.2.5 Marca de clase

La marca de clase o punto de medio de clase se obtiene mediante la siguiente formula:

$$X = \frac{Li + Ls}{2}$$

Donde:

X= Marca de clase

Li= Límite inferior

Ls= Limite superior

5.2.2.6 Límite inferior y superior reales

Son límites calculados de mayor amplitud que los límites superiores e inferiores, con objeto de incluir en ellos ciertos datos que pudiera presentar la muestra.

5.2.2.7 Frecuencia Absoluta

La frecuencia absoluta se determina como el número de veces que se repite un dato.

5.2.2.8 Frecuencia Relativa

Las frecuencias relativas representan el porcentaje de veces en que ocurre un dato.

5.2.2.9 Frecuencia absoluta acumulada

La frecuencia absoluta acumulada para un determinado valor se considera como la frecuencia de cada dato más la suma de los valores anteriores a dicha suma.

5.2.2.10 Frecuencia relativa acumulada

Es la suma de los valores acumulados de la frecuencia relativa. La suma de las frecuencias relativas que se han acumulado, incluyendo la clase sobre la que se está calculando la frecuencia relativa.

Ejemplo de clasificación de datos:

**Estudiantes, según edad.
Escuela Primaria "Revolución". Enero, 2004.**

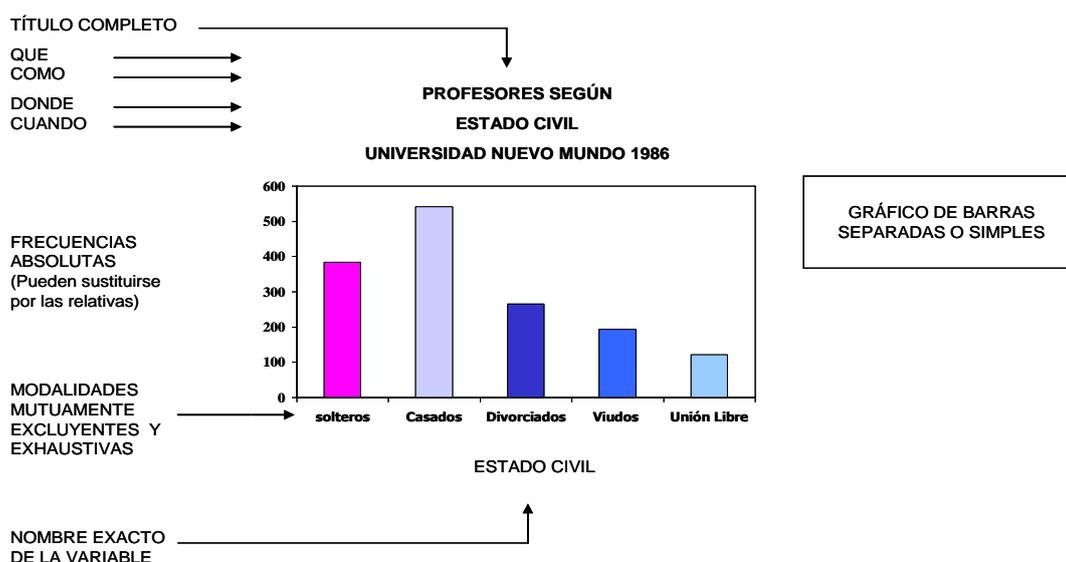
Edad en años	Frecuencia	Porcentaje	Frecuencia Acumulada	Porcentaje Acumulado
6 años	58	10.76%	58	10.76%
7 años	56	10.39%	114	21.15%
8 años	78	14.47%	192	35.62%
9 años	95	17.63%	287	53.25%
10 años	89	16.51%	376	69.76%
11 años	65	12.06%	441	81.82%
12 años	98	18.18%	539	100%
Total	539	100%		

5.2.3 Características mínimas para la creación de gráficos

Cada gráfico debe tener un título completo que responda a las preguntas: ¿Qué se está presentando?, ¿Cómo se está presentando? ¿De dónde provienen los datos?, ¿De cuándo son los datos?.

La escala de cada variable debe tener modalidades o clases que sean exhaustivas y mutuamente excluyentes. En vista de que los gráficos tienen por propósito agilizar la captación visual de las características esenciales de los datos, debe evitarse utilizar efectos que limiten su comprensión visual.

Ejemplo:



5.2.3.1 Diagrama barras separadas

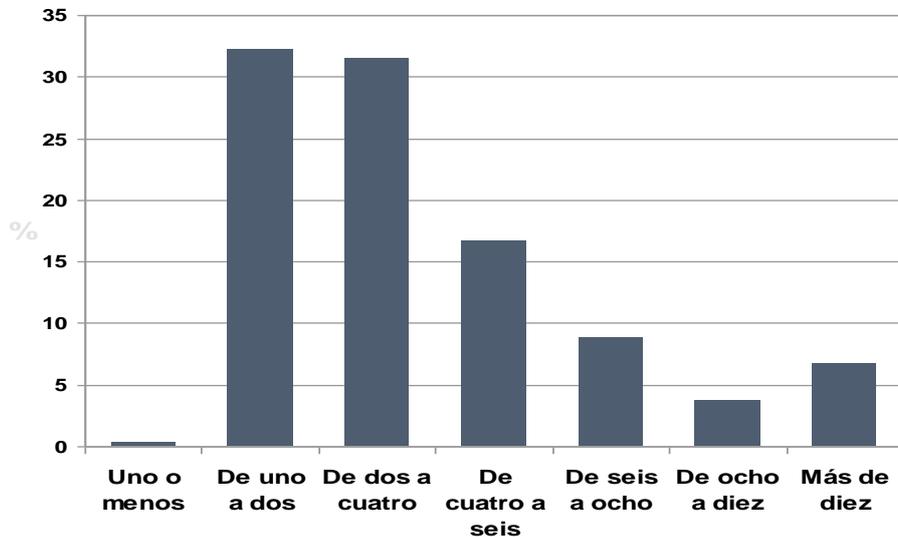
Es aquel que se compone de una serie de barras verticales u horizontales, donde la longitud de la barra representa la frecuencia proporcional (o relativa) de una categoría de una variable nominal u ordinal. La barra pueda también representar la frecuencia absoluta de una categoría.

Consiste en un conjunto de columnas separadas que representan la frecuencia o el porcentaje de cada uno de los valores o categorías de la variable de interés.

Los gráficos de barra separadas son especialmente eficaces para ilustrar la competencia entre categorías.

Ejemplo:

**Alumnos según ingresos familiares mensuales
UNAM, 1999 - 2000**



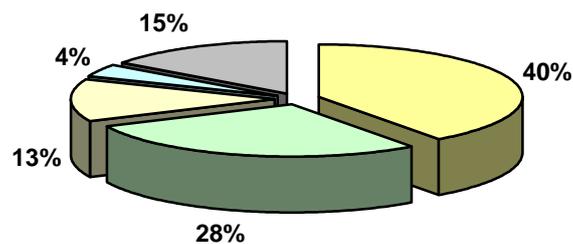
5.2.3.2 Circular, de sectores o de pastel

Es un círculo que se divide (o rebana) desde su punto central, donde cada espacio representa la frecuencia proporcional (o relativa) de determinada categoría de una variable nominal u ordinal

Es especialmente útil para transmitir un sentido de equidad, tamaño relativo o desigualdad entre categorías.

Ejemplo:

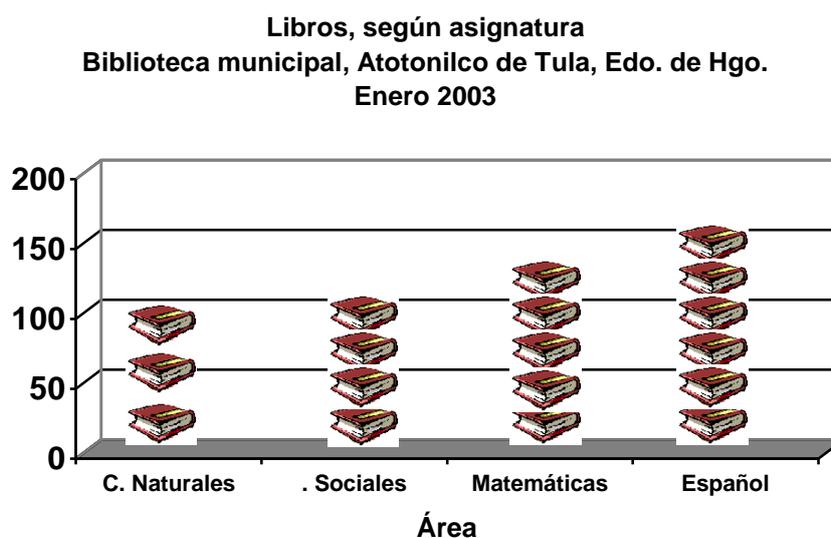
**Residentes, según estado civil.
San Miguel Ameyalco, Edo. de México
Enero, 2003**



5.2.3.3 Pictograma

Son figuras con formas de objetos que se colocan de manera que el número o el tamaño de los objetos representen la frecuencia proporcional (o relativa) de una categoría. El número o el tamaño de los objetos también pueden representar la frecuencia absoluta.

Ejemplo:

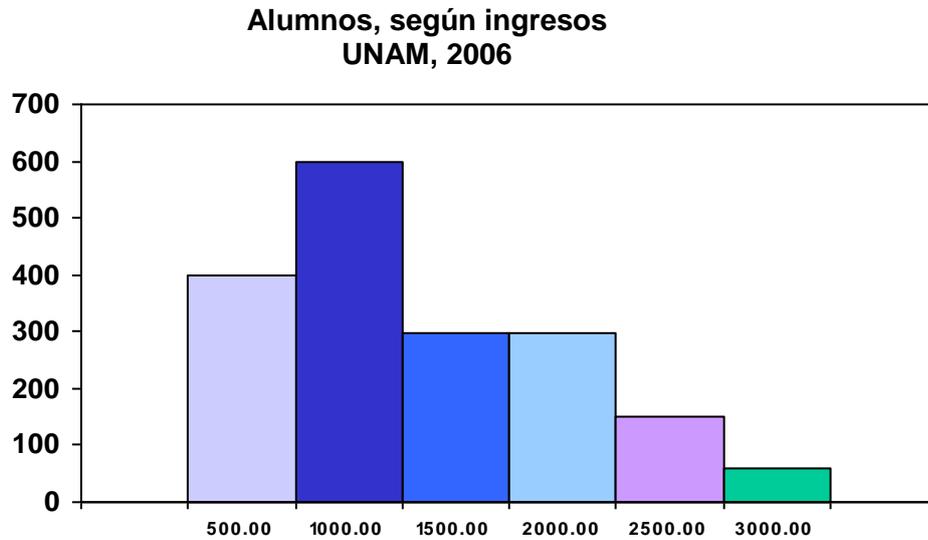


5.2.3.4 Histograma

Es un diagrama de 90 grados que presenta las puntuaciones de una variable cuantitativa (discreta o continua) o de intervalo o razón a lo largo del eje horizontal, y la frecuencia de cada puntuación, en una columna al eje vertical.

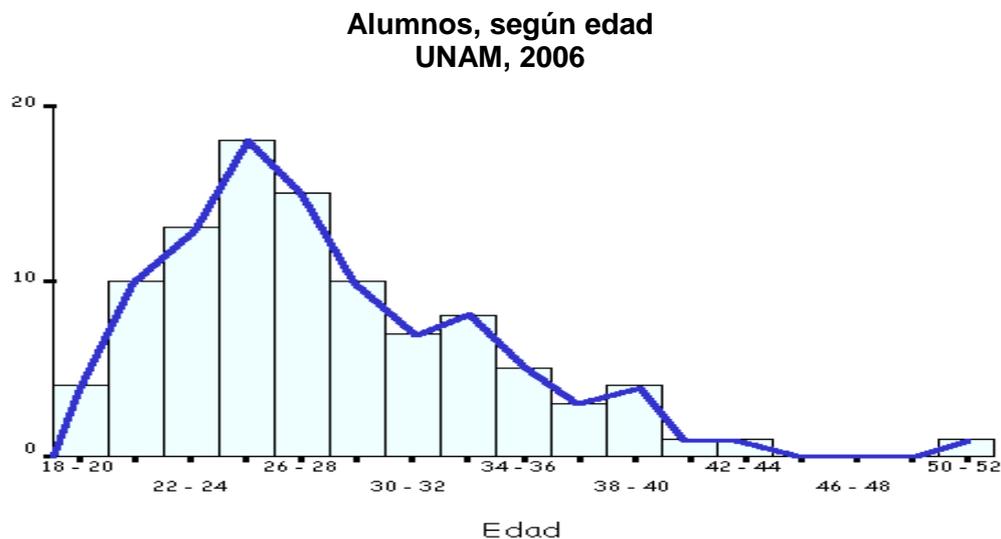
Es útil cuando se trata de representar distribución de frecuencias cuya variable es continua y viene dada por intervalo o clase.

Ejemplo:



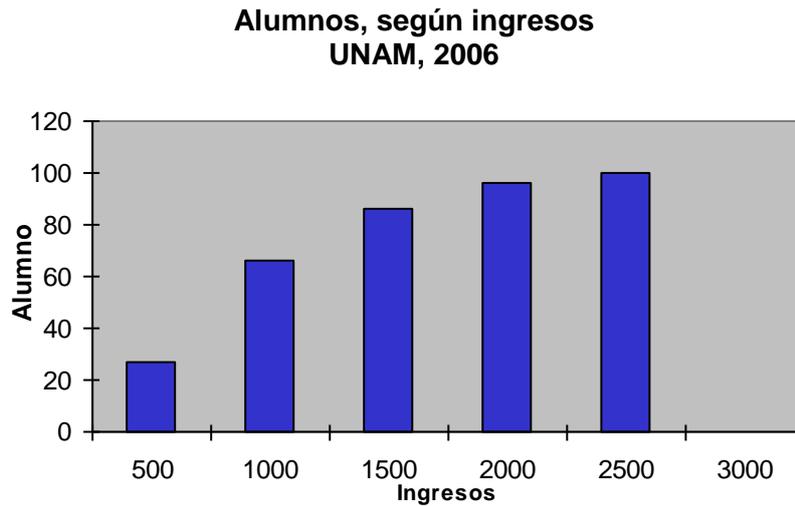
5.2.3.5 Polígono de frecuencias

Es un diagrama de 90 grados con puntuaciones de intervalo o razón (discreta o continua) señaladas en el eje horizontal y las frecuencias de las puntuaciones están representadas por las alturas de puntos localizados sobre las puntuaciones y conectados mediante líneas rectas.



5.2.3.6 Polígono de frecuencias acumuladas u ojiva o diagrama de Galton

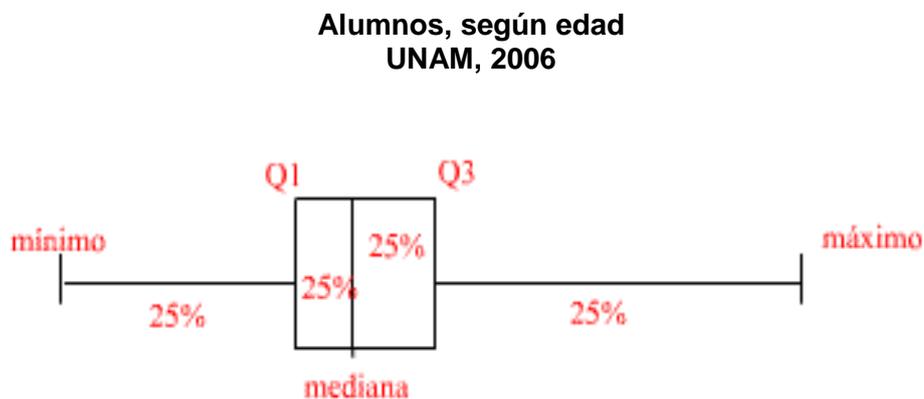
Es la representación gráfica de las frecuencias relativas acumuladas o frecuencias porcentuales acumuladas de una variable.



5.2.3.7 Diagrama de caja

Es la representación gráfica que despliega cuartiles, así como puntuaciones mínimas y máximas de una distribución.

Es una técnica útil para detectar valores extremos de una distribución.



Resumen de la unidad

Esta unidad aborda las principales formas de mostrar los datos respecto a cuadros y gráficos así como las condiciones o normas para su aplicación en consideración al nivel de medición de las variables. Se establecen las normas mínimas que deben cumplir un gráfico y un cuadro a efecto de facilitar la comprensión de los datos que se muestran.

Unidad VI

Descripción de datos

Introducción

En la presente unidad se estudian medidas de resumen aplicables a variables de tipo cualitativo y cuantitativo. En el primer caso se abordan conceptos como proporciones, porcentajes, razones y tasas. Respecto al segundo caso se enfatiza en estadísticas de tendencia central, de posición, dispersión y distribución.

Objetivo

Al finalizar la presente unidad emplearás medidas de resumen de tipo cuantitativo y cualitativo en el estudio de los fenómenos sociales, con el propósito establecer los cimientos de la estadística descriptiva con miras a la generalización.

Temario

6.1 Para variables cualitativas

6.1.1 Proporciones

6.1.2 Porcentajes

6.1.3 Razones

6.1.4 Tasas

6.2 Para variables cuantitativas

6.2.1 Medidas de tendencia central

6.2.1.1 Moda

6.2.1.2 Mediana

6.2.1.3 Media aritmética

6.2.1.4 Media ponderada

6.2.2 Medidas de posición

6.2.2.1 Centil

6.2.2.2 Cuartil

6.2.2.3 Decil

6.2.3 Medidas de dispersión

6.2.3.1 Rango

6.2.3.2 Rango intercuartílico

6.2.3.3 Desviación media

6.2.3.4 Varianza

6.2.3.5 Desviación estándar

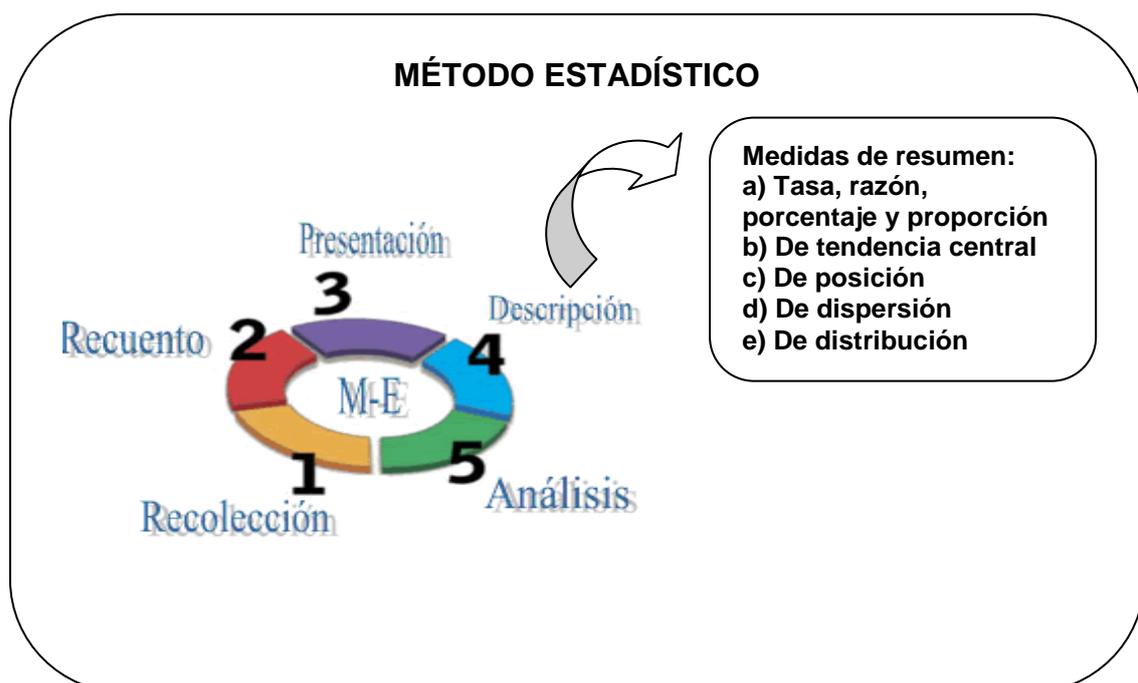
6.2.3.6 Coeficiente de variación

6.2.4 Medidas de distribución

6.2.4.1 Sesgo

6.2.4.2 Curtosis

Diagrama conceptual



6.1 Para variables cualitativas

Las medidas estadísticas aplicables para variables de tipo cualitativo son: proporciones, porcentajes, razones y tasas. La moda también es aplicable a este nivel de medición aunque se incluye en la parte de variables cuantitativas.

6.1.1 Proporciones

Definición: Una proporción es una medida de resumen para variables cualitativas que consiste en la comparación, a través de una división, entre un subconjunto y el conjunto al que pertenece. Es la relación que se establece entre una parte con respecto al otro.

Ejemplo: En un país de Centroamérica el año de 1999 hubo 54,345 nacimientos en total; de ellos 36,562 fueron niños.

Por otra parte, en un país de Norteamérica ocurrieron un total de 21,215 nacimientos en total de los cuales 6,876 fueron niños.

Procedimiento: de acuerdo a la definición de proporción deben dividirse un subconjunto entre el conjunto al que pertenece.

Así, para el país de Centroamérica:

$$\text{Proporción de nacimientos de niños en un país de Centroamérica} = \frac{36,562}{54,345} = 0.67$$

Por otra parte, para el país de Norteamérica:

$$\text{Proporción de nacimientos de niños en un país de Norteamérica} = \frac{6,876}{21,215} = 0.32$$

INTERPRETACIÓN:

Para el caso del país de Centroamérica: la importancia del subconjunto de nacimientos de niños, en relación con el conjunto total es de 0.67, lo que equivale a más de dos terceras partes de todos los nacimientos.

Para el caso del país del Norteamérica: la importancia del subconjunto de nacimientos de niños, en relación con el conjunto total es de 0.32, lo que equivale a aproximadamente una tercera parte del total de nacimientos.

En conclusión: la importancia de los nacimientos de niños es mucho mayor en el país de Centroamérica que en el país del Norteamérica.

6.1.2 Porcentajes

Es la proporción multiplicada “por cien”. Es la relación que se establece entre una de las parte con respecto al todo o total multiplicado por cien.

De acuerdo al ejercicio anterior los porcentajes de defunciones de menores de un año en el país de Centroamérica y en el país de Norteamérica son: 67% y 32%, respectivamente.

6.1.3 Razones

Definición: Una razón es una medida de resumen para variables cualitativas que consiste en la comparación, a través de una división, entre dos conjuntos o grupos de elementos de diferente o igual naturaleza. Es la relación que se da entre dos subconjuntos o dos conjuntos.

Ejemplo: En una comunidad rural existen 2,250 personas dependientes de 450 mujeres jefas de familia; por otra parte, en una comunidad urbana existen 2385 personas dependientes de 795 jefas de familia.

Procedimiento: De acuerdo a la definición de razón deben dividirse dos grupos de elementos. En este caso, para la comunidad rural se divide el número de personas dependientes entre el número de jefas de familia:

$$\text{Razón} \quad \frac{\text{Personas}}{\text{Jefas de familia}} \text{ en la comunidad rural} = \frac{2250}{450} = 5$$

En cuanto a la comunidad urbana, se divide el número de personas dependientes entre el número de jefas de familia:

$$\text{Razón} \quad \frac{\text{Personas}}{\text{Jefas de familia}} \text{ en la comunidad urbana} = \frac{2385}{795} = 3$$

INTERPRETACIÓN:

Para el caso de la comunidad rural: Existen cinco personas dependientes por cada jefa de familia.

Para el caso de la comunidad urbana: Existen 3 personas dependientes por cada jefa de familia.

En conclusión: parece haber mayor cantidad de personas dependientes para las jefas de familia de la comunidad rural.

6.1.4 Tasas

Definición: Una tasa o coeficiente es una medida de resumen para variables cualitativas que consiste en la comparación, a través de una división, entre el número de veces que ocurre un cierto tipo de eventos y la población en la que *puede ocurrir* dicho tipo de eventos. Usualmente el resultado de tal división consiste en una cifra fraccionaria menor a uno; por ello, el resultado de la división suele ser multiplicado por alguna constante que sea múltiplo del número 10.

Ejemplo: En Atotonilco de Tula, Estado de Hidalgo, a lo largo del año 2000, ingresaron al nivel medio superior 945 estudiantes. De ellos. 350 eran de sexo masculino. Se desea condensar la información de tal manera que los cálculos

produzcan una medida de resumen que permita imaginar o evocar la magnitud o probabilidad de que ingrese una persona de sexo masculino al nivel medio superior.

Procedimiento: De acuerdo a la definición de tasa debe dividirse el evento entre la población en la cual dicho evento puede ocurrir. El resultado debe multiplicarse por un múltiplo del número 10.

$$\text{Tasa de ingreso de una persona de sexo masculino a nivel superior} = \frac{350}{945} \times 100 = 37$$

INTERPRETACIÓN:

En Atotonilco de Tula, Estado de Hidalgo, a lo largo del año de referencia, 37 de cada 100 estudiantes que ingresaron al nivel medio superior eran de sexo masculino.

6.2 Para variables cuantitativas

Las medidas estadísticas aplicables para variables de tipo cuantitativo son: Moda, mediana, media aritmética, media ponderada, rango, centil, cuartil, rango intercuartílico, decil, desviación media, varianza, desviación estándar, coeficiente de variación, sesgo y curtosis.

6.2.1 Medidas de tendencia central

Las medidas de tendencia central son: Moda, mediana, media aritmética y media ponderada.

6.2.1.1 Moda

Definición: Es el valor que en una serie se repite con mayor frecuencia. Es el valor de mayor frecuencia u ocurrencia.

Ejemplo:

Un grupo de prácticas de especialización de la Escuela Nacional de Trabajo Social acudió a la Escuela Primaria "Revolución" a impartir un curso sobre educación ambiental dirigido a 20 estudiantes. Una vez impartidas las sesiones y realizadas las tareas, este grupo deseaba conocer cuál había sido el impacto del curso, para ello generó un instrumento de tipo cuestionario, una vez recabada la información, inicio la aplicación de pruebas estadísticas para responder el objetivo planteado.

Estudiantes, según producción semanal individual de basura en casa. Escuela Primaria "Revolución". Enero, 2004.

Número de alumnos	Producción individual aproximada de basura Kgrs.	Frecuencia Relativa	Frecuencia Absoluta Acumulada	Frecuencia Relativa Acumulada
1	12.5	4.18%	12.50	4.18%
2	12.7	4.25%	25.20	8.43%
3	12.9	4.32%	38.10	12.75%
4	13.2	4.42%	51.30	17.17%
5	13.7	4.59%	65.00	21.75%
6	14.3	4.79%	79.30	26.54%
7	14.3	4.79%	93.60	31.33%
8	14.3	4.79%	107.90	36.11%
9	14.3	4.79%	122.20	40.90%
10	14.9	4.99%	137.10	45.88%
11	15.2	5.09%	152.30	50.97%
12	15.5	5.19%	167.80	56.16%
13	15.7	5.25%	183.50	61.41%
14	15.9	5.32%	199.40	66.73%
15	15.9	5.32%	215.30	72.05%
16	16.1	5.39%	231.40	77.44%
17	16.4	5.49%	247.80	82.93%
18	16.6	5.56%	264.40	88.49%
19	16.9	5.66%	281.30	94.14%
20	17.5	5.86%	298.80	100.00%
Total	298.8	100.00%		

Procedimiento: Buscar si existe un valor que se repita con mayor frecuencia que los demás.

En el ejemplo: El valor 14.3 se repite en cuatro ocasiones.

INTERPRETACIÓN:

En el ejemplo: “La producción individual de basura más frecuente en el grupo de estudiantes es de 14.3 kgrs.”

6.2.1.2 Mediana

Definición: En una serie de valores ordenados de mayor a menor o viceversa, es aquel valor que divide en dos partes de igual tamaño a toda la serie. Es decir, por debajo de la mediana se localiza la mitad de los datos, el 50% , y la otra mitad, el otro 50%, por encima de la mediana.

Procedimiento: Ordenar la serie y localizar el valor que la divide en dos partes de igual tamaño, de tal manera que en una parte quede el 50% de los datos y en la otra el 50% restante.

Fórmula:	para serie impar	$N+1 / 2$. Establece el lugar donde se encuentra la mediana
	para serie par	Promedio de los valores centrales

En el ejemplo: El promedio de los valores centrales 14.9 kgrs. y 15.2 kgrs. equivale a 14.94 kgrs.

INTERPRETACIÓN:

“La mitad de los estudiantes produce 14.94 kgrs. de basura o menos y la otra mitad produce 14.94 kgrs. o más”.

6.2.1.3 Media aritmética

Definición: Es el valor que tendrían todos los datos de una serie numérica si ellos fueran de igual valor.

Procedimiento: Sumar todos los valores y dividir tal suma entre el número de valores que componen a la serie simple.

Fórmula: $\bar{X} = \Sigma x / n$

En el ejemplo: $\bar{X} = 298.8 / 20 = 14.94$

INTERPRETACIÓN:

“Si todos los estudiantes produjeran la misma cantidad de basura esta sería de 14.94 kgrs.”

6.2.1.4 Media ponderada

Definición: Es el valor que se obtendría si se asignase un peso o una ponderación a cada uno de ellos.

Ejemplo: En un curso de Estadística Aplicada a la Investigación Social I, se aplican dos exámenes parciales y se le asignan los siguientes “pesos”: $P_1= 1$ $P_2= 2$. Si un estudiante obtiene en el primer examen 8 y el segundo 6. ¿Cuál es el promedio ponderado?

Procedimiento: Sumar el producto de todos los valores por el número de su ponderación y dividir tal suma entre la suma del número de ponderaciones.

Fórmula: $\bar{X}_p = \Sigma Px / \Sigma P$

En el ejemplo: $\bar{X}_p = (1) 8 + (2) 6 / 1+ 2 = 20 / 3 = 6.66$

INTERPRETACIÓN: “Si se le asignase un peso o ponderación a los valores obtenidos la calificación sería de 6.66.”

6.2.2 Medidas de posición

Las medias estadísticas de posición son: centil, cuartil y decil.

6.2.2.1 Centil

Definición: En una serie de valores ordenados de mayor a menor o viceversa, es aquel valor que divide en dos partes porcentualmente complementarias a toda la serie. Se refiere a un dato en una distribución de puntuaciones o valores, por debajo y por encima del cual queda un porcentaje determinado de de casos. Un centil se representa con C_n .

Procedimiento: Ordenar la serie y localizar el valor que la divide en los porcentajes complementaria deseados.

En el ejemplo: Para encontrar el valor del centil 50 (C_{50}) debe localizarse aquel valor que deje la mitad con menor o igual magnitud a él y a la otra mitad de los valores con magnitud más grande o igual a él.

INTERPRETACIÓN:

(Válida para el Centil 50 o C_{50}): “El 50% de los estudiantes produjeron 15.05 kgrs. de basura o menor y el 50% restante produjo 15.05 kgrs. o mayor”.

6.2.2.2 Cuartil

Definición: En un conjunto de valores ordenados de mayor a menor o viceversa, son los valores que dividen a una serie de datos en cuatro partes iguales. Se expresan con la letra Q_1 , Q_2 , Q_3 y Q_4 , cada uno representa el 25%, de tal manera que el valor del cuartil dos (Q_2) es el mismo que el de la mediana.

Procedimiento: Ordenar la serie y localizar el valor que la divide en cuatro partes porcentualmente complementarias.

En el ejemplo: Para encontrar el valor del cuartil dos Q_2 debe localizarse a aquel valor que deje dos cuartas partes con menor o igual magnitud a él y a las

dos cuartas partes restantes de los valores con magnitudes más grandes o iguales a él.

INTERPRETACIÓN:

(Válida para el Cuartil dos o Q_2): “El 50% de los estudiantes produjeron 15.05 kgrs. de basura o menor y el 50% restante produjo 15.05 kgrs. o mayor”.

6.2.2.3 Decil

Son los valores que dividen a una serie de datos en diez partes iguales. Se expresan con la letra D_1, D_2, D_3 y D_4, D_5, D_6, D_7 y D_8, D_9, D_{10} , cada uno representa el 10%, de tal manera que el valor del decil cinco, cuartil dos (Q_2) es el mismo que el de la mediana.

Procedimiento: Ordenar la serie y localizar el valor que la divide en diez partes porcentualmente complementarias.

En el ejemplo: Para encontrar el valor del decil cinco D_5 debe localizarse a aquel valor que deje cinco cuartas partes con menor o igual magnitud a él y a las cinco cuartas partes restantes de los valores con magnitudes más grandes o iguales a él.

INTERPRETACIÓN:

(Válida para el Decil cinco o D_5): “El 50% de los estudiantes produjeron 15.05 kgrs. de basura o menor y el 50% restante produjo 15.05 kgrs. o mayor”.

6.2.3 Medidas de dispersión

Las medias estadísticas de dispersión son: rango, rango intercuartílico, desviación media, varianza, desviación estándar y coeficiente de variación.

6.2.3.1 Rango

Definición: Es la diferencia entre el valor mayor y el valor menor de una serie.

Procedimiento: Encontrar por sustracción o resta, la diferencia entre el valor más grande de la serie (X_{\max}) y el valor más pequeño (X_{\min}).

En el ejemplo: $(X_{\max}) 17.5 - (X_{\min}) 12.5 = 5$.

INTERPRETACIÓN:

La diferencia entre el estudiante que más produce basura y el que menos lo hace es de 5 kgrs.”

6.2.3.2 Rango intercuartílico

Definición: Es la diferencia entre el cuartil tres Q_3 y el cuartil uno Q_1 , de una serie de datos.

Procedimiento: Ordenar la serie y localizar el valor que la divide en cuatro partes porcentualmente complementarias.

En el ejemplo: Para encontrar el valor del rango intercuartílico debe localizarse a aquel valor que deje una cuarta parte con menor o igual magnitud a él y a las tres cuartas partes restantes de los valores con magnitudes más grandes o iguales a él. Y aquel valor que deje una tres cuartas partes con menor o igual magnitud a él y a una cuarta parte restante de los valores con magnitudes más grandes o iguales a él.

Fórmula: $RIC = (Q_3) 15.95 - (Q_1) 14.15 = 1.8$.

INTERPRETACIÓN:

“La diferencia del 50% de los valores centrales es de 1.8 kgrs. de producción de basura por parte de los estudiantes”.

6.2.3.3 Desviación media

Definición: Es la magnitud o distancia entre cada observación o dato y la media aritmética. El promedio de estas distancias da una medida racional de la dispersión de los datos.

Fórmula:

$$D_m = \frac{\sum |X - \bar{X}|}{n}$$

Las rayas verticales indican valor absoluto, es decir, hay que considerar a todos los datos negativos como positivos.

Procedimiento:

- 1) Obtener la media aritmética: Fórmula: $\bar{X} = \sum x / n$.
- 2) Calcular la desviación real y desviación absoluta de cada valor con respecto a la media aritmética.
- 3) Sumar la distancia o desviación absoluta de las puntuaciones respecto a la media aritmética.
- 4) Dividir la sumatoria de las desviaciones absolutas entre el número total de valores.

Estudiantes, según producción semanal individual de basura en casa. Escuela Primaria "Revolución". Enero, 2004.

Número de alumnos	Producción individual aproximada de basura Kgrs.	Desviación real de cada dato con respecto a la media aritmética $X - \bar{X}$	Desviación absoluta de cada dato con respecto a la media aritmética $ X - \bar{X} $
1	12.5	-2.44	2.44
2	12.7	-2.24	2.24
3	12.9	-2.04	2.04
4	13.2	-1.74	1.74
5	13.7	-1.24	1.24
6	14.3	-0.64	0.64
7	14.3	-0.64	0.64

8	14.3	-0.64	0.64
9	14.3	-0.64	0.64
10	14.9	-0.04	0.04
11	15.2	0.26	0.26
12	15.5	0.56	0.56
13	15.7	0.76	0.76
14	15.9	0.96	0.96
15	15.9	0.96	0.96
16	16.1	1.16	1.16
17	16.4	1.46	1.46
18	16.6	1.66	1.66
19	16.9	1.96	1.96
20	17.5	2.56	2.56
Total	298.8		$\Sigma 24.6 $

Fórmula:

$$D_m = \frac{\Sigma | X - \bar{X} |}{n}$$

$$D_m = \frac{\Sigma | 24.6 |}{20}$$

$$D_m = \underline{1.23}$$

INTERPRETACIÓN:

El promedio de desviaciones de cada dato es de 1.23 con respecto a la media aritmética.

6.2.3.4 Varianza

Definición: Es el cuadrado de la desviación estándar. Es el promedio de desviaciones o diferencias cuadráticas de cada valor de una serie con respecto al promedio de dicha serie.

6.2.3.5 Desviación estándar

Definición: Es la raíz cuadrada de la varianza. Es el promedio de desviación de las puntuaciones con respecto a la media.

Fórmula:
Series simples

$$s = \sqrt{\frac{\sum (X - \bar{X})^2}{N}}$$

Procedimiento:

1. Obtener el promedio de la serie de valores: Fórmula: $\bar{X} = \Sigma x / n$
2. Calcular la desviación de cada valor con respecto a la media y elevar al cuadrado cada una de ellas para obtener la sumatoria de las desviaciones cuadráticas.
3. Dividir la sumatoria de las desviaciones cuadráticas entre el número de valores menos uno.

Obtener la raíz cuadrada.

Estudiantes, según producción semanal individual de basura en casa. Escuela Primaria "Revolución". Enero, 2004.

Número de alumnos	Producción individual aproximada de basura Kgrs.	Desviación real de cada dato con respecto a la media aritmética	Desviación real cuadrática de cada dato con respecto a la media aritmética
		$X - \bar{X}$	$(X - \bar{X})^2$
1	12.5	-2.44	5.9536
2	12.7	-2.24	5.0176
3	12.9	-2.04	4.1616
4	13.2	-1.74	3.0276
5	13.7	-1.24	1.5376
6	14.3	-0.64	0.4096
7	14.3	-0.64	0.4096
8	14.3	-0.64	0.4096
9	14.3	-0.64	0.4096
10	14.9	-0.04	0.0016
11	15.2	0.26	0.0676
12	15.5	0.56	0.3136
13	15.7	0.76	0.5776
14	15.9	0.96	0.9216
15	15.9	0.96	0.9216
16	16.1	1.16	1.3456
17	16.4	1.46	2.1316
18	16.6	1.66	2.7556

19	16.9	1.96	3.8416
20	17.5	2.56	6.5536
Total		298.8	$\Sigma (X - \bar{X})^2 = 40.768$

Fórmula:
Series simples

$$s = \sqrt{\frac{\Sigma (X - \bar{X})^2}{N}}$$

Fórmula:
Series simples

$$s = \sqrt{\frac{40.768}{20}}$$

s = 1.42

INTERPRETACIÓN 1:

La producción de basura se desvía en promedio 1.42 kgrs de la media aritmética.

Para poder realizar una interpretación más adecuada de la desviación estándar es imprescindible considerar las propiedades de la curva normal.

La distribución normal presenta como principales características:

- a) La curva normal es un polígono de frecuencias en forma de campana para el que están calculadas sus áreas en función de los diversos valores del eje horizontal o del eje de las X o abscisas.
- b) En el eje de las X o abscisas se encuentran valores de tipo cuantitativo continuo, genéricamente denominados puntuaciones "Z", cuyas magnitudes teóricamente pueden ir, de izquierda a derecha desde menos infinito a más infinito.
- c) La media de todos los valores z de la abscisa equivale a cero, pues la mitad son negativos y la mitad son positivos. En el sitio de la abscisa que corresponde al cero, es decir la media, se encuentra la parte más

alta de la curva. En este sitio también se encuentra la mediana de todos los valores z de la abscisa, pues el 50% de ellos está antes del cero y el 50% restante se encuentra después.

- d)** La curva es simétrica alrededor de la media; esto es, hay una mitad izquierda que es reflejo de la mitad derecha. Es decir la asimetría es cero, la mitad de la curva es exactamente igual a la otra mitad. La distancia entre $\mu + 3\sigma$ y $\mu - 3\sigma$ es la misma.
- e)** En la abscisa existen segmentos unitarios de igual longitud y de tamaño 1. Los segmentos a la izquierda de la media tienen signo negativo y los segmentos a la derecha de la media tienen signo positivo. Tales segmentos, denominados desviaciones estándar (σ) pueden dividirse en fracciones infinitamente pequeñas y continuas.
- f)** La curva es asintótica; es decir, sus extremos teóricamente nunca tocan la abscisa. Por ello, la longitud de la abscisa podría ser infinitamente larga; sin embargo se acostumbra graficar solo hasta la distancia de tres segmentos a la izquierda y a la derecha de la media.
- g)** Toda el área bajo la curva equivale a 1 o a 100%. Por lo anterior el área a la izquierda de la media equivale a 0.5 ó 50% y el área a la derecha de la media equivale también a 0.5 ó 50%.
- h)** Es unimodal, una sola moda
- i)** Es una función particular entre desviaciones con respecto a la media de una distribución y la probabilidad de que éstas ocurran.
- j)** El área que se encuentra sobre el segmento de la abscisa que va desde la media hasta el valor z de +1 equivale a 0.3413 o 34.13%; por simetría,

el área que se encuentra sobre el segmento que va desde la media hasta el valor z de -1 de la abscisa también equivale a 0.3413 o 34.13%.

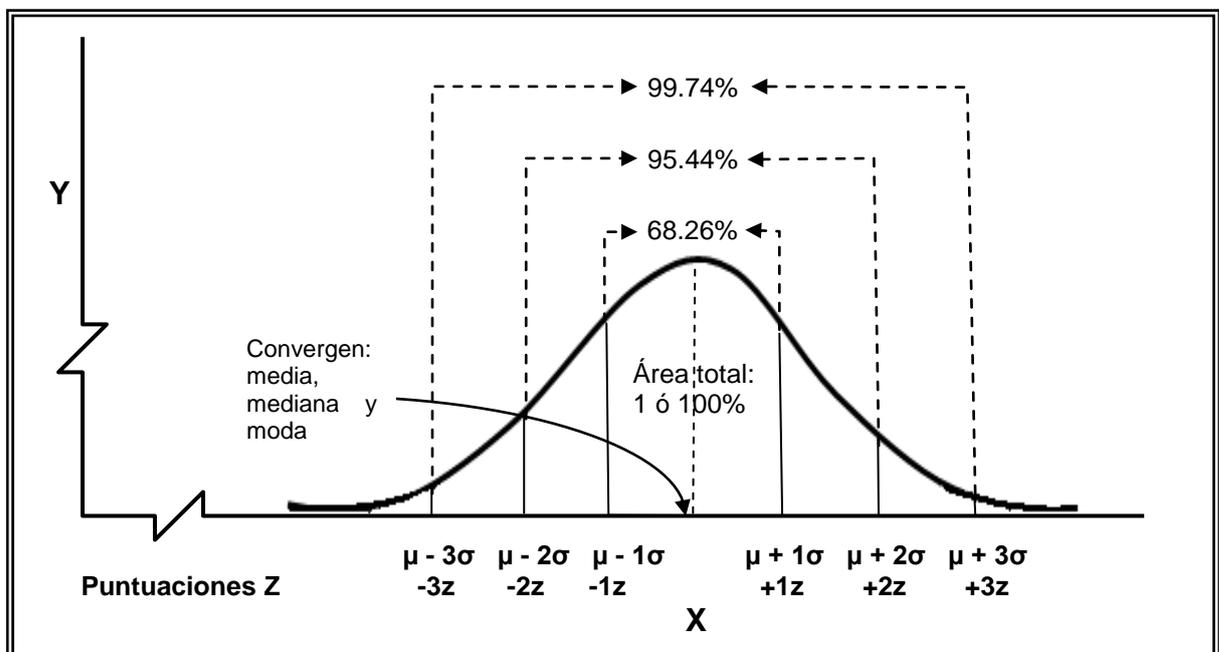
k) El área que se encuentra sobre el segmento de la abscisa que va más allá del valor z de +1 equivale a 0.1587 o 15.87%; por simetría, el área que se encuentra sobre el segmento que va más allá (hacia menos infinito) del valor z de -1 de la abscisa también equivale a 0.1587 o 15.87%.

l) Es mesocúrtica. El valor de su curtosis equivale a cero.

m) La media, la mediana y la moda coinciden en el mismo punto.

n) Para cualquier segmento de la abscisa, y aún para fracciones de segmento, se encuentran calculadas las áreas correspondientes en una tabla específicamente diseñada para tal efecto.

Características de la curva normal



En las distribuciones con forma normal, existe una relación especial entre la media y la desviación estándar con respecto al área en que se encuentra por debajo de la curva.

En este sentido es conveniente regresar a la forma de interpretar la desviación estándar a partir del dato calculado, es decir, 1.42 kgrs. de producción de basura. La preguntas son: ¿Cuántos kilogramos de basura produce el 68.26% central de los estudiantes de la Escuela Primaria “Revolución”? ¿Cuántos kilogramos de basura produce el 15.87% superior de los estudiantes?, ¿Cuántos kilogramos de basura produce el 15.87% inferior de los estudiantes?

Las respuestas se basan en las propiedades de la curva normal y los datos calculados.

Interpretación 2:

¿Cuántos kilogramos de basura produce el 68.26% central de los estudiantes de la Escuela Primaria “Revolución”?

Considerando las propiedades de la curva normal si sumamos a la media aritmética de la serie una desviación estándar obtendremos el valor de:
 $14.94+1.42= 16.36$ kgrs.

Considerando las propiedades de la curva normal si restamos a la media aritmética de la serie una desviación estándar obtendremos el valor de:
 $14.94-1.42= 13.52$ kgrs.

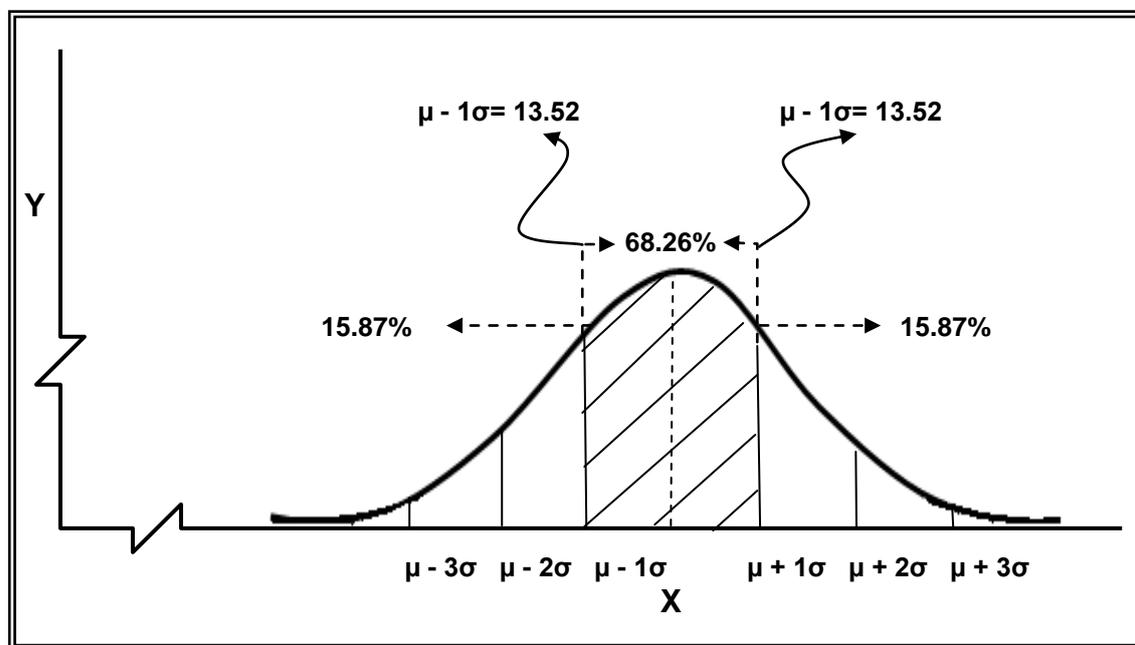
Entonces el 68.26% central de los estudiantes de la Escuela Primaria “Revolución” producen entre 13.52 kgrs. y 16.36 kgrs.

A partir del dato anterior se deducen las dos preguntas respecto al 15.87% superior e inferior, es decir:

El 15.87% superior de los estudiantes produce 16.36 kgrs. o más.

El 15.87% inferior de los estudiantes produce 13.52 kgrs. o menos.

**Diagrama de la curva normal considerando estudiantes, según producción semanal individual de basura en casa.
Escuela Primaria "Revolución". Enero, 2004, el ejemplo**



6.2.3.6 Coeficiente de variación

Es una relación o razón estadística entre la desviación estándar y la media aritmética multiplicada por 100, es decir, se lleva a términos de porcentaje.

Fórmula:

$$CV = \frac{S}{\bar{X}}$$

$$CV = \frac{1.42}{14.94}$$

$$CV = 0.095 \times 100 = \underline{\underline{9.50\%}}$$

INTERPRETACIÓN: La producción de basura de estudiantes de la Escuela Primaria "Revolución" varía en un 9.50%.

6.2.4 Medidas de distribución

6.2.4.1 Sesgo

El sesgo se define como la falta de simetría en una distribución. Cuando una curva esta equilibrada con relación a su eje vertical, se dice que es simétrica; cuando no observa esta situación, se dice que es asimétrica.

6.2.4.2 Curtosis

La curtosis se describe como el grado en que las proporciones observadas difieren de las de la curva normal. Distribuciones con una proporción mayor de valores extremos tienen curtosis positiva (leptocúrtica); las que tienen menos valores extremos tienen curtosis negativa (platicúrticas).

6.2.4.3 Cálculo de sesgo y curtosis

La distribución de los valores cuantitativos continuos tienen semejanza a la curva normal si su sesgo, calculado a través del método de momentos, vale entre -0.5 y +0.5, lo cual se simboliza de la siguiente forma:

$$-0.5 < a_3 < +0.5$$

y su curtosis, también calculada a través del método de momentos, vale entre 2 y 4, lo cual se simboliza de la siguiente forma:

$$2 < a_4 < 4$$

Las fórmulas para calcular el sesgo y la curtosis, a través del método de momentos, son los siguientes:

<p>SESGO: $a_3 = \frac{m_3}{\sqrt[3]{m_2}}$</p>	<p>CURTOSIS: $a_4 = \frac{m_4}{m_2^2}$</p>
--	---

El cálculo de momentos para *series simples* de datos cuantitativos continuos se hace con las fórmulas siguientes:

Momento 2:

$$m_2 = \frac{\sum \langle X - \bar{X} \rangle^2}{n}$$

Momento 3:

$$m_3 = \frac{\sum \langle X - \bar{X} \rangle^3}{n}$$

Momento 4:

$$m_4 = \frac{\sum \langle X - \bar{X} \rangle^4}{n}$$

En el caso de las series simples de valores, conviene efectuar el cálculo de los momentos a través de una tabla auxiliar de trabajo como la del siguiente ejemplo:

**20 Niños de un año de edad según peso.
HIM "Federico Gómez", Enero 2004**

9.1	9.4	8.9	9.6	10.5	8.8	9.4	9.2	9.0	8.1
9.3	8.8	9.5	9.7	9.2	9.4	9.6	9.0	9.4	9.8

El promedio equivale a: 9.285 kgrs

Cada uno de los valores X	Desviación de cada valor con respecto al promedio (X - \bar{X})	Elevación al cuadrado de cada una de las desviaciones (X - \bar{X}) ²	Elevación al cubo de cada una de las desviaciones (X - \bar{X}) ³	Elevación a la cuarta de cada una de las desviaciones (X - \bar{X}) ⁴
9.1	-0.185	0.034	-0.006	0.001
9.4	0.115	0.013	0.002	0.000
8.9	-0.385	0.148	-0.057	0.022
9.6	0.315	0.099	0.031	0.010
10.5	1.215	1.476	1.794	2.179
8.8	-0.485	0.235	-0.114	0.055
9.4	0.115	0.013	0.002	0.000
9.2	-0.085	0.007	-0.001	0.000
9	-0.285	0.081	-0.023	0.007
8.1	-1.185	1.404	-1.664	1.972
9.3	0.015	0.000	0.000	0.000
8.8	-0.485	0.235	-0.114	0.055
9.5	0.215	0.046	0.010	0.002
9.7	0.415	0.172	0.071	0.030
9.2	-0.085	0.007	-0.001	0.000
9.4	0.115	0.013	0.002	0.000
9.6	0.315	0.099	0.031	0.010
9	-0.285	0.081	-0.023	0.007
9.4	0.115	0.013	0.002	0.000
9.8	0.515	0.265	0.137	0.070
SUMATORIAS	0.000	4.441	0.079	4.421

Sustituyendo en las fórmulas para el cálculo de momentos en series simples se tiene:

Momento 2:

$$m_2 = \frac{\sum \langle x - \bar{x} \rangle^2}{n}$$

Momento 3:

$$m_3 = \frac{\sum \langle x - \bar{x} \rangle^3}{n}$$

Momento 4:

$$m_4 = \frac{\sum \langle x - \bar{x} \rangle^4}{n}$$

$$m_2 = \frac{4.441}{20} = 0.222 \quad m_3 = \frac{0.079}{20} = 0.004 \quad m_4 = \frac{4.421}{20} = 0.221$$

Finalmente, usando los valores calculados para los momentos y sustituyendo para las fórmulas de sesgo y curtosis en series simples se tiene:

SESGO: $a_3 = \frac{m_3}{\left(\sqrt{m_2}\right)^3}$ CURTOSIS: $a_4 = \frac{m_4}{m_2^2}$

$$a_3 = \frac{0.004}{\left(\sqrt{0.222}\right)^3} = \frac{0.004}{(0.471)^2} = \frac{0.004}{0.105} = 0.038 \quad a_4 = \frac{0.221}{(0.222)^2} = \frac{0.221}{0.049} = 4.484$$

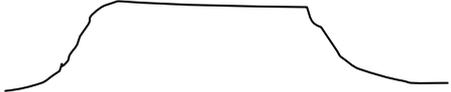
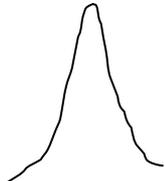
INTERPRETACIÓN:

En vista de que el sesgo calculado se encuentra en el intervalo que va desde -0.5 hasta +0.5 puede decirse que la distribución de los pesos de los niños se asemeja en asimetría a la curva normal.

Sin embargo, en vista de que la curtosis calculada está fuera del intervalo que va desde 2 hasta 4 no puede decirse que el grado de apuntamiento o aplanamiento de los pesos de los 20 niños sea semejante a la de la curva normal.

En resumen, la serie es simétrica como la curva normal pero más apuntada o elevada que ella.

Ejemplo de curvas o distribuciones y su interpretación

 <p>Distribución simétrica (asimetría=0), con curtosis positiva y una desviación estándar y varianza medias.</p>	 <p>Distribución simétrica positiva, curtosis positiva y una desviación estándar y varianza mayores.</p>
 <p>Distribución asimétrica positiva, curtosis negativa y una desviación estándar y varianza considerables.</p>	 <p>Distribución asimétrica negativa, curtosis positiva y una desviación estándar y varianza menores.</p>
 <p>Distribución simétrica, curtosis positiva y una desviación estándar y varianza bajas.</p>	 <p>Curva normal, curtosis=0, asimetría=0 y una desviación estándar y varianza promedios</p>

Resumen de la unidad

En la presente unidad emplearás las medidas de resumen para variables cualitativas y cuantitativas a fin de identificar su definición, establecer su forma de cálculo y sobre todo determinar la forma de interpretación de cada una de ellas.

Unidad VII

Correlación y regresión lineal

Introducción

En la presente unidad se aborda el concepto de relaciones lineales como base de elementos de correlación y regresión con la finalidad de determinar la relación o no de variables y sobre esa base establecer predicciones. Para ello se utiliza la graficación de dichos elementos que permiten visualizar su comportamiento.

Objetivo

Al finalizar la presente unidad emplearás medidas de regresión y correlación en el estudio de los fenómenos sociales con el propósito establecer la relación o no entre variables y hacer predicciones.

Temario

7.1 Relaciones lineales

7.1.1 Tipos

7.2 Correlación

7.2.1 Coeficientes de correlación

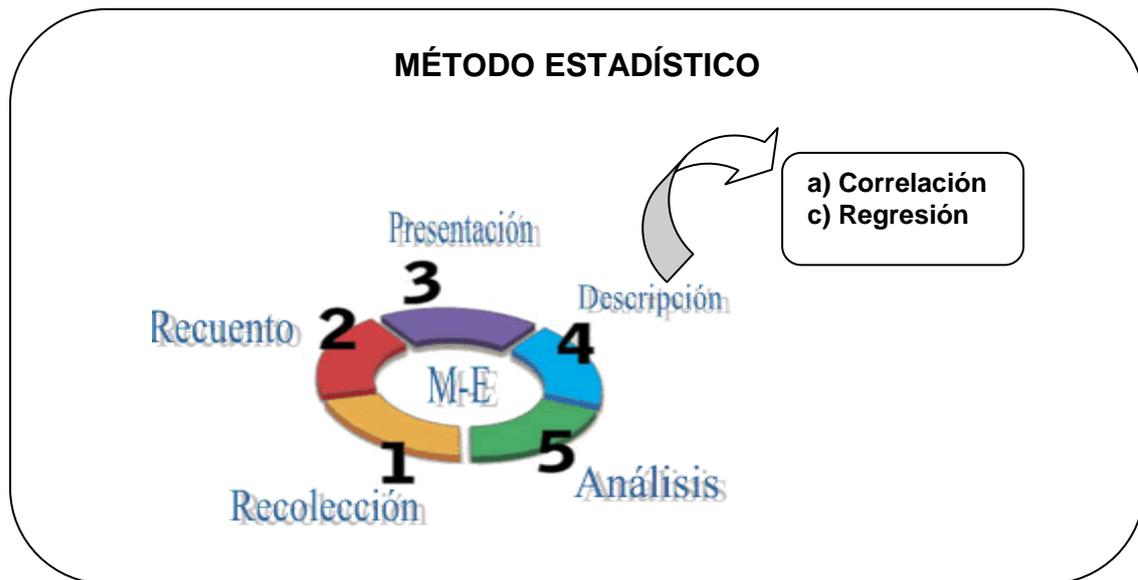
7.2.1.1 r de Pearson

7.2.1.2 Spearman r_s

7.3 Regresión

7.3.1 Recta de regresión por mínimos cuadrados

Diagrama conceptual

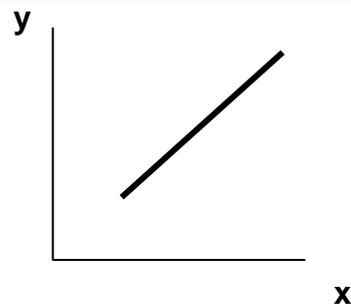


7.1 Relaciones lineales

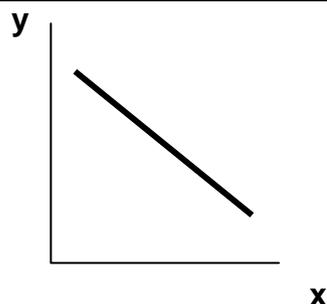
Una **relación lineal** entre dos variables es aquella que puede representarse con la mejor exactitud mediante una línea recta. Dicha relación puede ser positiva o negativa y a la vez perfecta e imperfecta.

7.1.1 Tipos

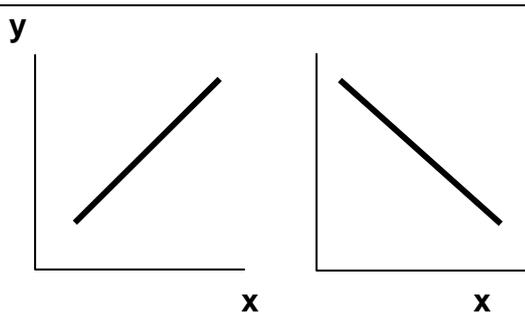
Una **relación positiva** indica que existe una relación directa entre las variables.



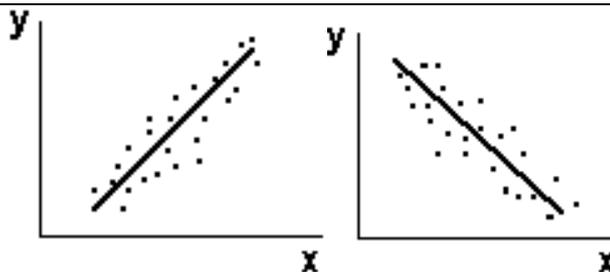
Una **relación negativa** existe cuando hay una relación inversa entre X e Y.



Relación perfecta: Es aquella en la que existe una relación positiva o negativa para la cual todos los puntos caen sobre la recta.



Relación imperfecta: Es aquella en la que existe una relación, pero no todos los puntos caen sobre la recta.



Relación curvilínea: Es aquella en la que existe una relación, pero los puntos forman una curva.



Sin relación: Es aquella en la que no existe una relación, porque todos los puntos están dispersos.



7.2 Correlación

La correlación se ocupa de establecer si existe una relación entre dos o más variables, así como su magnitud y dirección.

La dirección de la relación se refiere a si ésta es lineal positiva o negativa, perfecta o imperfecta, curvilínea, o inexistente.

El grado de la relación se refiere a su magnitud o fuerza a partir de la siguiente escala:

- 1 Correlación positiva perfecta
- 0.90 Correlación positiva muy fuerte
- 0.75 Correlación positiva considerable
- 0.50 Correlación positiva media
- 0.10 Correlación positiva débil
- 0.00 No existe correlación alguna
- 0.10 Correlación negativa débil
- 0.50 Correlación negativa media
- 0.75 Correlación negativa considerable
- 0.90 Correlación negativa muy fuerte
- 1 Correlación negativa perfecta

7.2.1 Coeficientes de correlación

Los principales coeficientes de correlación son:

a) Coeficiente de correlación de Pearson: Determina si existe relación lineal entre dos variables a nivel intervalar.

b) Coeficiente de correlación de Spearman: Determina si existe relación lineal entre dos variables a nivel ordinal.

c) Coeficiente de correlación Phi: Determina si existe relación lineal entre dos variables.

d) Coeficiente de correlación Biserial Puntual: Determina si existe relación lineal entre dos variables a nivel nominal con dos valores cada una (dicotómicas).

7.2.1.1 r de Pearson

Es el coeficiente de correlación que determina si existe relación lineal entre dos variables a nivel intervalar

Fórmula:

$$r = \frac{n \sum xy - [(\sum x)(\sum y)]}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$$

Puntuaciones de un grupo de 10 individuos en pruebas de ansiedad y de depresión

Individuo	Puntuación en prueba de ansiedad (x)	Puntuación en prueba de depresión (y)	X ²	Y ²	XY
1	9	3	81	9	27
2	10	4	100	16	40
3	10	5	100	25	50
4	11	4	121	16	44
5	12	4	144	16	48
6	12	6	144	36	72
7	12	7	144	49	84
8	13	6	169	36	78
9	14	6	196	36	84
10	15	6	225	36	90
11	16	7	256	49	112
12	16	8	256	64	128
Sumatorias	150	66	1936	388	857

Procedimiento:

Primer paso: identificación de las variables a medir. Se trata de un solo grupo de individuos a quienes se midieron un par de variables, ambas de tipo cuantitativo continuo.

Segundo paso: planteamiento de hipótesis estadísticas:

Hipótesis estadística nula: $H_0. r = 0$. Ello implica que no hay relación.

Hipótesis estadística alternativa: $H_a. r \neq 0$. Ello implica que hay relación.

Tercer paso: Cálculo del coeficiente de Pearson con la fórmula establecida.

Fórmula:

$$r = \frac{n \sum xy - [(\sum x)(\sum y)]}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$$

$$r = \frac{12(857) - [(150)(66)]}{\sqrt{[12(1936) - (150)^2][12(388) - (66)^2]}}$$

$$r = 0.82$$

Cuarto paso:

De acuerdo a la escala arriba señalada se puede establecer la siguiente conclusión: al menos para el grupo estudiado puede considerarse que existe asociación positiva muy fuerte entre las puntuaciones obtenidas por los sujetos en ansiedad y depresión.

7.2.1.2 Spearman r_s

Determina si existe relación lineal entre dos variables a nivel ordinal.

Primero: Identificación de las variables. En el ejemplo Se trata de un solo grupo de individuos a quienes se midieron un par de variables ambas de tipo cuantitativo discreto. 10 estudiantes de la ENTS a quienes se les aplicó una prueba de matemáticas (x) y estadística (y).

Segundo: Planteamiento de la hipótesis estadística.

Hipótesis nula: $H_0 = r_s = 0$. Ello implica que no hay relación.

Hipótesis Alternativa: $H_a = r_s \neq 0$. Ello implica que hay relación.

Tercero: Cálculo del coeficiente de Spearman

- Jerarquizar puntuaciones asignando su posición general o rango $R(x)$ y $R(y)$ en cada una de las series.
- Al encontrar puntuaciones “empatadas” o con el mismo valor se establece la misma posición al promediar sus lugares.
- Calcular las diferencias y elevarlas al cuadrado para obtener su sumatoria.

10 estudiantes

según puntuaciones en prueba de matemáticas y estadística

ENTS-UNAM, 2002

Estudiante	Matemáticas	Estadística	R(x)	R(y)	Diferencia $R(x)-R(y)$	$(R(x)-R(y))^2$
1	84	52	3.5	1.0	2.5	6.25
2	75	39	8.0	6.0	2.0	4.0
3	98	48	1.0	3.0	-2.0	4.0
4	70	32	10.0	10.0	0.0	0.0
5	75	40	8.0	5.0	3.0	9.0
6	80	36	6.0	9.0	-3.0	9.0
7	83	38	5.0	7.0	-2.0	4.0
8	75	37	8.0	8.0	0.0	0.0
9	84	50	3.5	2.0	1.5	2.25
10	90	46	2.0	4.0	-2.0	4.0
Sumatoria						42.5

Fórmula para calcular el coeficiente de Spearman

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

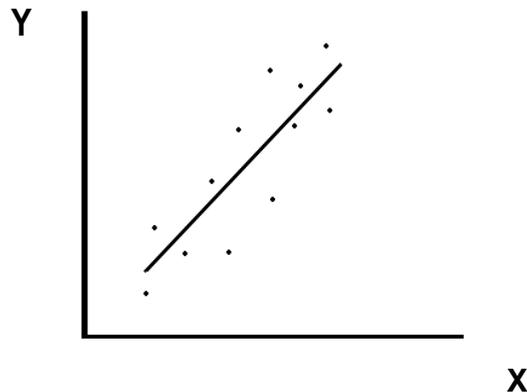
$$r_s = 1 - \frac{6 (42.5)}{10 (10^2 - 1)} = 0.74$$

Cuarto Paso:

De acuerdo a la escala arriba señalada se puede establecer la siguiente conclusión: **al menos para el grupo estudiado, puede considerarse que existe asociación considerable entre las calificaciones obtenidas por los estudiantes de la ENTS-UNAM.**

7.3 Regresión

Regresión: es una prueba estadística que analiza la relación entre dos o más variables para determinar **una predicción.**



7.3.1 Recta de regresión por mínimos cuadrados

La ecuación de la línea de regresión por mínimos cuadrados está dada por:

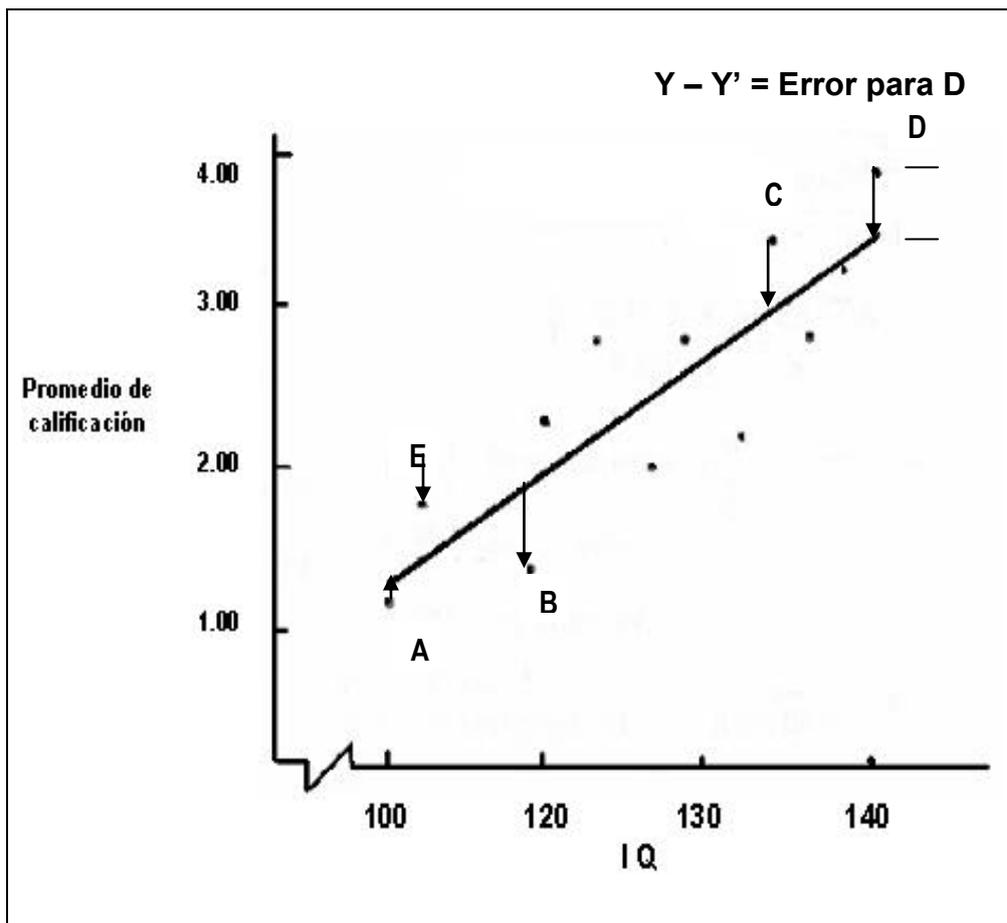
$$Y' = b_Y X + a_Y$$

Ecuación de regresión lineal para predecir Y
dado X

donde Y' = el valor predicho o estimado de.

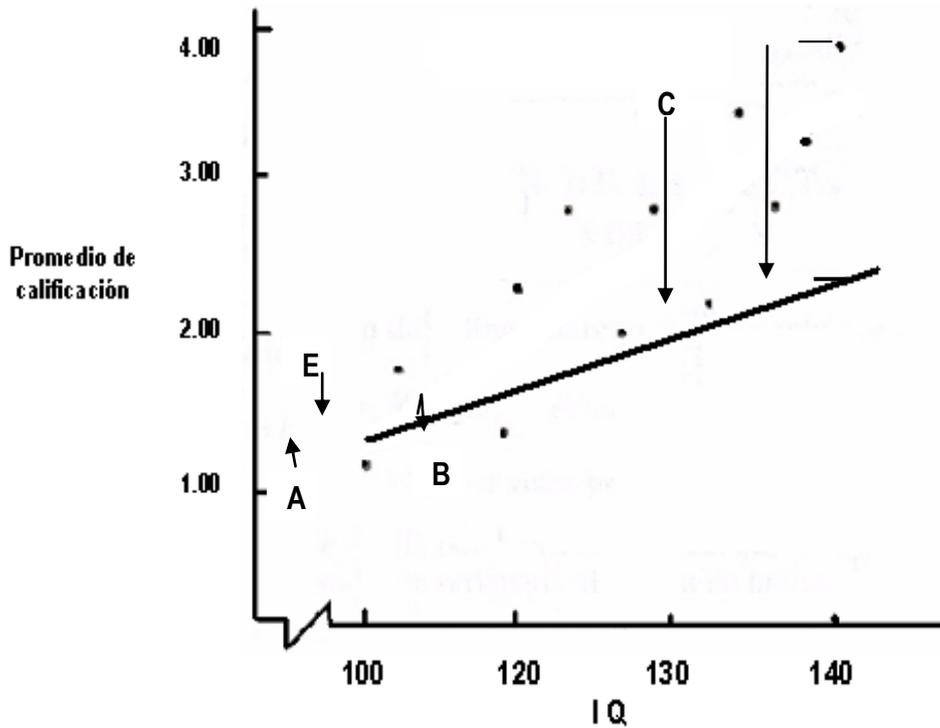
b_Y = la pendiente de la recta que minimiza los errores de predicción de Y

a_Y = la ordenada al origen de la recta que minimiza los errores de predicción de Y .



$Y - Y' = \text{Error para D}$

D



La constante de regresión b_Y es igual a:

$$b_Y = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{N}}{\sum X^2 - \frac{(\sum X)^2}{N}}$$

Que es la ecuación para calcular la constante de regresión b para predecir, Y dado X .

La constante de regresión a_Y está dada por

$$a_Y = \bar{Y} - b_Y \bar{X}$$

Que es la ecuación para calcular la constante de regresión a para predecir Y dado X

Ejemplo: Un trabajador social con especialización en psicología social esta interesado en determinar si es posible utilizar el IQ (Coeficiente Intelectual) de estudiantes universitarios para predecir su posible promedio de calificación a obtener en una prueba de estadística.

Aspectos a resolver:

- a) Elaborar un grafico de dispersión para los datos obtenidos.
- b) Deducir la línea de regresión

Finalmente, **la pregunta es:** Si un estudiante tiene un IQ de 124; al emplear la línea de regresión deducida, ¿Cuál es la predicción del promedio de las calificaciones?

Primer Paso: Calcular los valores necesarios para obtener la constante de regresión b_y

**Estudiantes según IQ y promedio de calificaciones.
ENTS-UNAM. Enero 2004**

Niño	IQ X	Promedio de calificaciones Y	X Y	X ²
1	110	1.0	110.0	12,100
2	112	1.6	179.2	12,544
3	118	1.2	141.6	13,924
4	119	2.1	249.9	14,161
5	122	2.6	317.2	14,884
6	125	1.8	225.0	15,625
7	127	2.6	330.2	16,129
8	130	2.0	260.0	16,900
9	132	3.2	422.4	17,424
10	134	2.6	348.4	17,956
11	136	3.0	408.0	18,496
12	138	3.6	496.8	19,044
SUMATORIAS	Σ X= 1503	Σ Y= 27.3	Σ XY= 3488.7	Σ X²= 189,187

Segundo Paso: Sustituir los valores en la fórmula de la constante de regresión

b_y

$$\begin{aligned}
 b_Y &= \frac{\sum XY - \frac{(\sum X)(\sum Y)}{N}}{\sum X^2 - \frac{(\sum X)^2}{N}} \\
 &= \frac{3488.7 - \frac{1503(7.3)}{12}}{189,187 - \frac{503^2}{12}} \\
 &= \frac{69.375}{936.25} = 0.074
 \end{aligned}$$

Tercer Paso: Sustituir los valores en la fórmula de la constante de regresión:

$$\begin{aligned}
 a_Y &= \bar{Y} - b_Y \bar{X} \\
 &= 2.275 - 0.0741(25.25) \\
 &= -7.006
 \end{aligned}$$

Cuarto Paso:

Construir la ecuación de regresión lineal para predecir Y dado X

$$Y' = b_Y X + a_Y$$

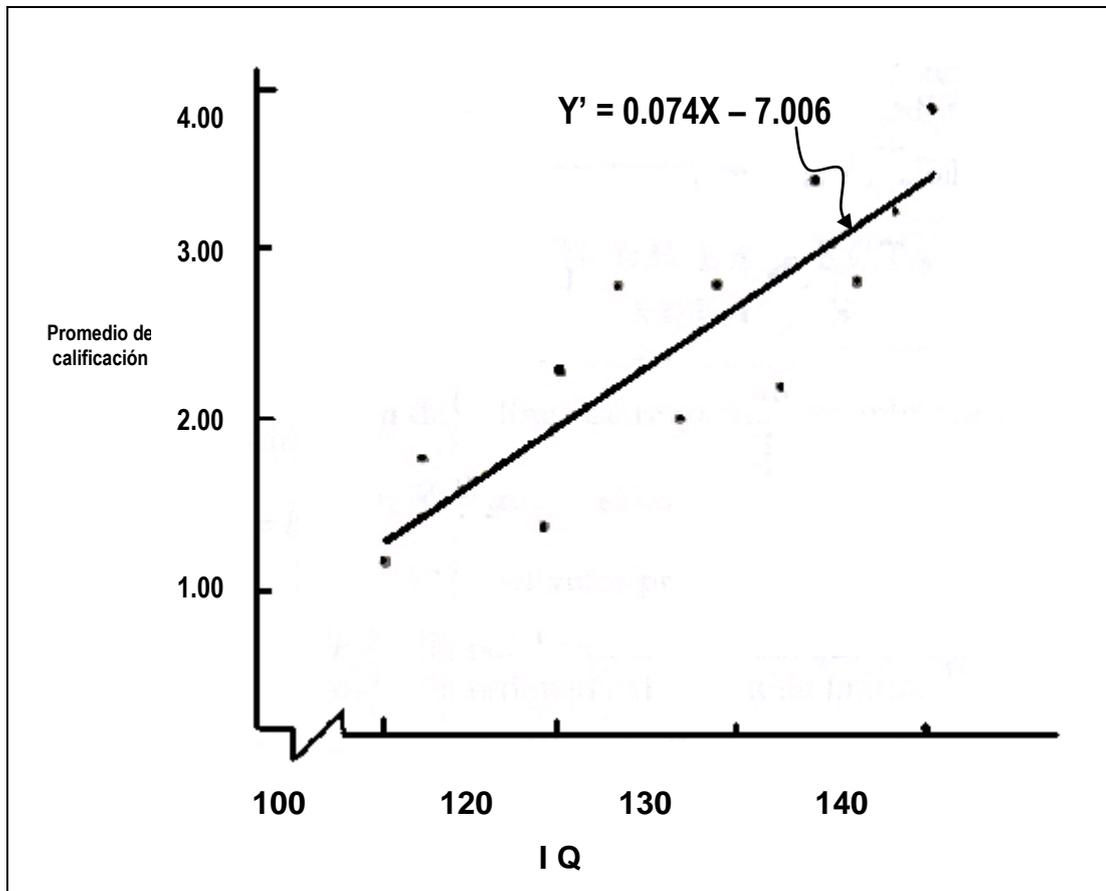
Ecuación de regresión lineal deducida para predecir Y dado X

$$Y' = 0.074 X - 7.006$$

Gráfica de línea de regresión deducida:

Línea de regresión para estudiantes según IQ y promedio

ENTS-UNAM. Enero 2004



Quinto Paso: Realizar una predicción a partir de la ecuación de regresión lineal deducida.

Si un estudiante tiene un IQ de 124; al emplear la línea de regresión deducida, ¿Cuál es la predicción del promedio de las calificaciones? Es decir, ¿Cuál es el promedio que podría obtener?

Valores: $X = 124$ de IQ

$$Y' = b_Y X + a_Y \quad \text{equivale a} \quad Y' = 0.074X - 7.006$$

Sustituyendo:

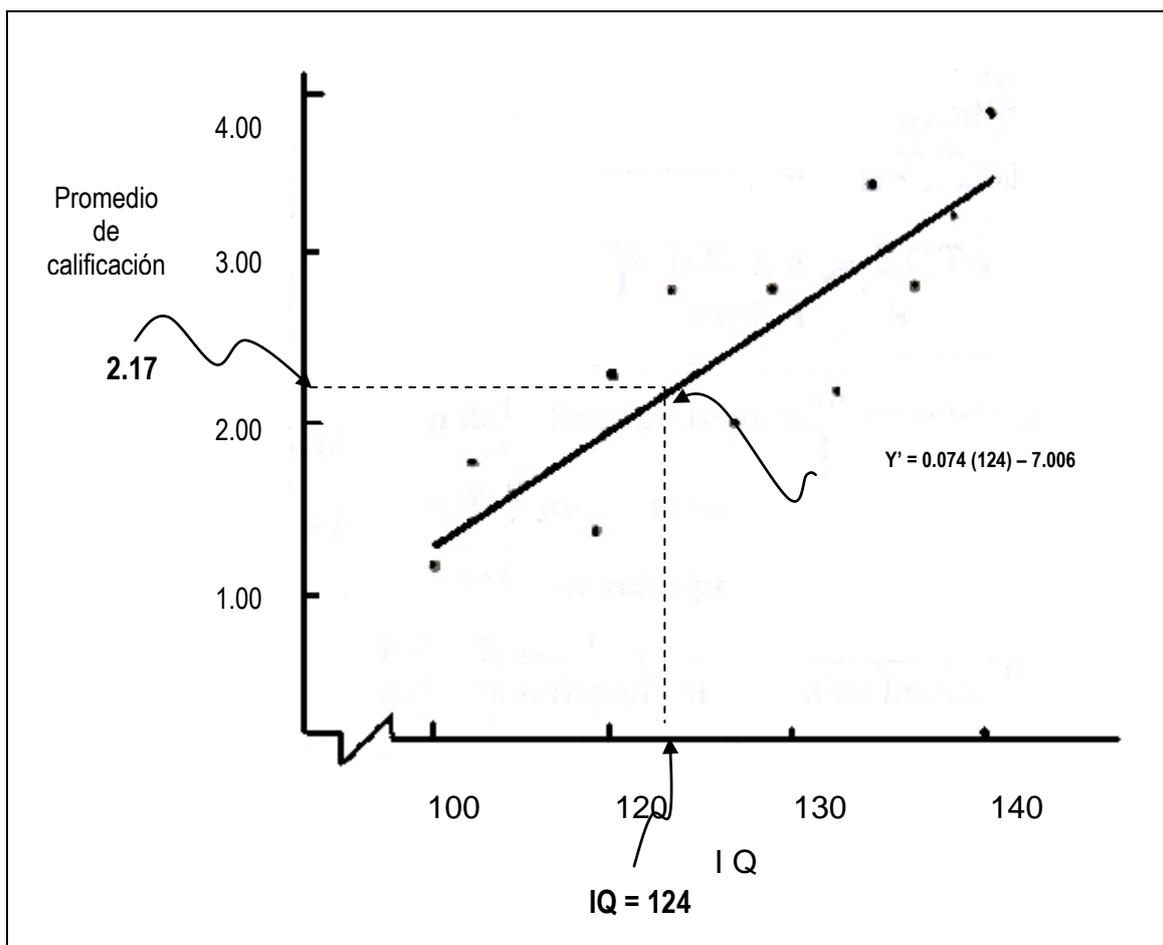
$$Y' = 0.074 (124) - 7.006 = 2.17$$

El promedio de calificación que obtendría un niño de PRO-EDUCACIÓN en una prueba de estimulación temprana con un coeficiente intelectual de 124 sería de 2.17.

Sexto Paso:

Referencia gráfica del valor obtenido:

**Línea de regresión para estudiantes según IQ y promedio
ENTS-UNAM. Enero 2004.**



Resumen de la unidad

En la presente unidad se aborda dos conceptos fundamentales que determinan por un lado la relación o no entre variables y por el otro los procedimientos para hacer predicciones, es decir, correlación y regresión respectivamente. Se estudia la tipología de coeficientes así como la forma de construir una línea de regresión con objeto de predecir el comportamiento de una variable.

Glosario

Unidad 1

Análisis de datos: conjunto de operaciones lógicas o numéricas que se aplican a la información obtenida por medio de los instrumentos de recolección.

Dato: Medidas que se realizan sobre los sujetos de un experimento.

Estadística: Es aquella que se ocupa de los métodos y procedimientos para recoger, clasificar, resumir, hallar regularidades y analizar los *datos*, siempre y cuando la variabilidad e *incertidumbre* sea una causa intrínseca de los mismos; así como de realizar *inferencias* a partir de ellos, con la finalidad de ayudar a la toma de *decisiones* y en su caso formular *predicciones*.

Estadística descriptiva: es la rama de la estadística que recolecta, recuenta, presenta y describe un conjunto de datos.

Estadística inferencial: es aquella rama de la estadística que hace planteamientos acerca de los atributos de la población utilizando probabilidades basadas en muestras aleatorias.

Investigación científica: es el proceso sistemático, controlado, empírico y crítico, de proposiciones hipotéticas sobre las presumidas relaciones entre fenómenos naturales.

Unidad 2

Dato: Medidas que se realizan sobre los sujetos de un experimento.

Estrato: Parte de una muestra que conforma un subgrupo de sujetos con características específicas compartidas entre sí. Una muestra puede conformarla dos o más estratos.

Muestra representativa: es aquella cuyas características y atributos responden cercanamente con los de la población correspondiente.

Muestra: Es un subconjunto de la totalidad de un conjunto de elementos, seres u objetos que se pretende investigar. Es una fracción o subgrupo de una población.

Muestreo accidental o por accidente: Es aquél en el que el investigador decide por voluntad propia qué elementos integran la muestra aprovechando los sujetos de que dispone.

Muestreo aleatorio estratificado: Es la selección unidades de muestra a partir de dividir en subgrupos o estratos a la población de estudio y extraer proporcionalmente de cada uno de ellos, los elementos que integrarán la muestra.

Muestreo aleatorio simple: Es el procedimiento en el que un tamaño de muestra n es seleccionado de una población de tamaño N , de tal manera que cada muestra posible de tamaño n tiene la misma posibilidad de ser seleccionada.

Muestreo con reemplazo: método de muestreo en el cual cada miembro de la población elegida para la muestra se regresa a la población antes de elegir al siguiente miembro.

Muestreo de diseño de bola de nieve: es aquél en el que la configuración de la muestra se construye a partir de la referencia de un sujeto a otro, es decir, de la acumulación de elementos a través de la remisión del investigador de un individuo a otro y de este a otro.

Muestreo de juicio o criterio: Es aquél en el que la conformación de la muestra depende del criterio, juicio o discernimiento del investigador quien se apoya generalmente en su praxis profesional.

Muestreo estratificado y por racimos: Es la selección de sujetos o elementos muestrales inmersos en conjuntos, grupos o estratos definidos física o geográficamente.

Muestreo no probabilístico o determinístico: Es aquél en el que la elección de los elementos no depende de la probabilidad, sino de las características determinadas por el investigador o por quien hace la muestra.

Muestreo por conveniencia: Es aquél en el que el investigador decide de acuerdo con los objetivos del estudio los elementos integrarán la muestra.

Muestreo por cuota: Es aquél en el que el investigador establece una cantidad denominada cuota que obtiene de una categoría o estrato para conformar la muestra.

Muestreo por expertos: Es aquél en el que la selección de los elementos muestrales depende de la experiencia, reconocimiento o prestigio que una persona pueda tener respecto a un tema determinado.

Muestreo por sujetos voluntarios: Es aquél en el que las unidades muestrales o elementos a considerar deciden libremente formar parte o no de la muestra.

Muestreo por sujetos-tipo: Es aquél en el que el investigador elige elementos de muestra que comparten las mismas características (políticas, económicas, sociales, académicas, etc.) para generar información de tipo cualitativo más que cuantitativo.

Muestreo probabilístico: Es aquél en el que todos los elementos de una población tienen la misma probabilidad de ser elegidos.

Muestreo sin reemplazo: método de muestreo en el cual los miembros de la muestra no se regresan a la población antes de elegir a los miembros siguientes.

Muestreo sistemático: Es la selección de cada *késimo*, nombre de una lista de población después de seleccionar aleatoriamente el punto de inicio.

Muestreo: Conjunto de procedimientos que se ejecutan para seleccionar una muestra.

Población: todos los miembros, elementos, observaciones o valores que se ajustan a un criterio específico. Conjunto completo de individuos, objetos o datos en cuyo estudio está interesado un investigador.

Unidad 3

Analizar o inferir: etapa del método estadístico que proporciona los métodos para estimar las características de un grupo total (población), basándose en datos de un conjunto pequeño (muestra) de observaciones.

Contar: etapa del método estadístico donde los datos son sometidos a revisión, clasificación y cómputo numérico.

Dato: Medidas que se realizan sobre los sujetos de un experimento.

Describir: etapa del método estadístico donde los datos se resumen en forma de medidas que permiten expresar las principales propiedades numéricas de los datos.

Estadística descriptiva: es la rama de la estadística que recolecta, recuenta, presenta y describe un conjunto de datos.

Estadística inferencial: es aquella rama de la estadística que hace planteamientos acerca de los atributos de la población utilizando probabilidades basadas en muestras aleatorias.

Estadística: Es aquella que se ocupa de los métodos y procedimientos para recoger, clasificar, resumir, hallar regularidades y analizar los *datos*, siempre y cuando la variabilidad e *incertidumbre* sea una causa intrínseca de los mismos; así como de realizar *inferencias* a partir de ellos, con la finalidad de ayudar a la toma de *decisiones* y en su caso formular *predicciones*.

Estudio de datos: conjunto de operaciones lógicas o numéricas que se aplican a la información obtenida por medio de los instrumentos de recolección.

Método estadístico: Es el conjunto de procedimientos que recolecta, cuenta, presenta, describe y analiza los datos, a fin de extraerles el máximo de información, base para la toma de decisiones

Presentar etapa del método estadístico que mediante el uso de tablas o cuadros y graficas permite una inspección precisa y rápida de los datos.

Recolectar: etapa del método estadístico donde se acopia la información cualitativa y cuantitativa a través de variables estableciendo su nivel de medición.

Unidad 4

Cualitativa: la medición ocurre cuando los numerales asignados se usan como etiquetas o nombres más que para una cuantificación.

Cuantitativa: medición que asigna números a las observaciones reflejando la cantidad o grado que posee el atributo.

Dato: medidas que se realizan sobre los sujetos de un experimento.

Datos categóricos (o nominales): comprenden variables en las que las observaciones no tienen un rango u orden inherente o un continuo fundamental, por ejemplo, género, raza y trabajo son variables categóricas.

Datos independientes: se producen cuando cada observación no resulta afectada por y no está relacionada con cualquier otra observación en el conjunto de datos.

Escala de medición de intervalo: se aplica a datos que, además de clasificarse y ordenarse (como los de la escala ordinal), se puede saber con exactitud el tamaño (la cantidad) de la diferencia entre ellos.

Escala de medición de razón: datos de admiten el cero absoluto o verdadero. Esta escala posee todas las características de la escala de intervalo y, además, proporciona la certeza de que existe una concordancia entre el dato y el hecho real.

Escala nominal de medición: se usan valores como etiquetas o nombres. las variables categóricas representan escalas nominales.

Escala ordinal: sus mediciones presuponen un continuo fundamental y proporcionan datos en la forma de rangos. Esto implica que un número mayor

indica una cantidad o grado más grande del atributo medido que lo que indica un número más bajo, pero las diferencias entre rangos pueden no ser iguales.

Estudio de datos: conjunto de operaciones lógicas o numéricas que se aplican a la información obtenida por medio de los instrumentos de recolección.

Variable continua: variable que, teóricamente, puede asumir un número infinito de valores entre las unidades adyacentes de una escala. Son aquellas que adquieren valores numéricos decimales o fraccionados.

Variable dependiente: es el resultado o variable criterio que está relacionada con cambios en la variable independiente. Variable en un experimento, medida por un investigador, para determinar el efecto de una variable independiente.

Variable discreta: variable para la cual no existen valores posibles entre las unidades adyacentes en una escala. Son aquellas que adquieren valores numéricos enteros.

Variable independiente: es la variable manipulada (el predictor) para determinar sus efectos (predicciones) sobre la variable dependiente. Variable de un experimento que es controlada en forma sistemática por el investigador.

Variables dicotómicas: son aquellas que están compuestas de sólo dos categorías distintas.

Variables: son características o atributos que dan las observaciones que difieren. Cualquier propiedad o característica de algún evento, objeto o persona, que puede tener diversos valores en diversos instantes, según las condiciones.

Unidad 5

Abscisa: se refiere al eje horizontal o eje **X** de una gráfica.

Distribución de frecuencias acumuladas: número de datos que caen por debajo del límite superior real de cada intervalo.

Distribución de frecuencias relativas: proporción del número total de datos que aparecen en cada intervalo.

Distribución de frecuencias: lista de valores de datos y su frecuencia de aparición.

Distribución de porcentajes acumulados: porcentaje de datos que caen por debajo del límite superior real de cada intervalo.

Eventos mutuamente excluyentes: dos eventos que no pueden ocurrir al mismo tiempo; es decir, la ocurrencia de uno impide la ocurrencia de otro.

Frecuencia: Número de veces que se repite un elemento en una unidad de registro.

Gráfica de dispersión: es un conjunto de puntos en un plano x y, cada uno de los cuales indica simultáneamente el desempeño de un sujeto tanto en la variable x u horizontal como en la variable y o vertical.

Tablas de contingencia: son arreglos bidimensionales que muestra las frecuencias de la celda, es decir, el número de observaciones que caen en las categorías de los subgrupos formadas al cruzar los niveles de la variable de fila con los niveles de la variable de columna.

Unidad 6

Centil: valor sobre la escala de medida, debajo del cual cae un porcentaje dado de los datos en la distribución. Es aquel valor que divide a una serie de datos en partes porcentualmente complementarias.

Coeficiente de variación: es una relación o razón estadística entre la desviación estándar y la media aritmética multiplicada por 100.

Cuartil: es uno de los tres puntos (Q_1 , Q_2 , Q_3) que parten la distribución en cuatro segmentos iguales Q_1 es el punto que divide el cuarto inferior de la distribución de los tres cuartos superiores; $Q_1=P_{25}$, $Q_2=P_{50}$, $Q_3=P_{75}$.

Curtosis: describe el grado en que las proporciones observadas difieren de las de la curva normal. Distribuciones con una proporción mayor de valores extremos tienen curtosis positiva (leptocúrtica); las que tienen menos valores extremos tienen curtosis negativa (platicúrticas).

Desviación estándar: es la raíz cuadrada de la varianza. Es el promedio de desviación de las puntuaciones con respecto a la media.

Desviación media: es la magnitud o distancia entre cada observación o dato y la media aritmética.

Media (o media aritmética): es el promedio aritmético de un conjunto de valores. Suma de los datos dividida entre el número de los mismos. Es aquel valor que tendrían todos los de una serie si estos fueran de igual valor.

Mediana: es el punto medio de una distribución de valores; precisamente la mitad de valores cae arriba de la mediana; también se le llama percentil 50 o Q_2 . Es aquel valor que divide a una serie de datos en dos partes de igual tamaño.

Medición: es un proceso por el que se asignan números (o cuantificaciones) a las observaciones.

Moda: es el valor (o categoría) con la mayor frecuencia de ocurrencia.

Porcentaje: es una proporción multiplicada por 100.

Proporción: es la división entre un subconjunto y el conjunto al que pertenece.

Rango: es la diferencia entre el valor máximo y el valor mínimo de una serie de datos.

Razón: es la división entre dos elementos de igual o de distinta naturaleza.

Sesgo negativo: describe distribuciones simétricas en las que la mediana excede a la media; la cola de la distribución es hacia los valores bajos.

Sesgo positivo: describe distribuciones asimétricas en las que la media excede la mediana; los valores “se alargan” hacia los valores altos.

Sesgo: describe la falta de simetría en una distribución. Es una tendencia sistemática para una estadística inferencial (por ejemplo, s) a ser consistentemente más grande o más pequeña que el parámetro de población correspondiente (por ejemplo, σ).

Tasa: es la división entre los eventos que ocurren de un determinado fenómeno entre la población en donde puede ocurrir dicho evento por un múltiplo de 10.

Varianza: es el cuadrado de la desviación estándar. Es el promedio de desviaciones o diferencias cuadráticas de cada valor de una serie con respecto al promedio de dicha serie.

Unidad 7

Coefficiente de correlación: son mediciones descriptivas que muestran la dirección y grado de la relación entre dos variables. Expresión cuantitativa de la magnitud y dirección de una relación.

Confiabilidad: capacidad que tiene un instrumento de arrojar los resultados equivalentes entre los respondedores, independientemente de quien lo aplique.

Correlación negativa: se indica cuando observaciones por arriba de la media de una variable tienden a asociarse con observaciones por debajo de la media en una segunda variable, y viceversa.

Correlación positiva: describe una relación bivariada entre dos variables en las que los valores de los sujetos tienden a ir juntos (a algún punto); aquellos que se clasifican arriba de la media en una variable probablemente también se clasificarán por arriba de la media en la segunda variable.

Efecto de regresión: se refiere al fenómeno de que los sujetos que se desvían marcadamente de la media, cuando se vuelven a medir, tiende a regresar o clasificar más cerca de la media del grupo.

Estudio correlacional: no mide variables sino la relación que se establece entre ellas.

Gráfica de dispersión: es un conjunto de puntos en un plano xy , cada uno de los cuales indica simultáneamente el desempeño de un sujeto tanto en la variable x u horizontal como en la variable y o vertical.

Hipótesis direccional: especifica a priori la dirección de una diferencia en un parámetro. En las pruebas de una cola se emplean hipótesis direccionales.

Hipótesis no direccional: se utilizan dos pruebas inferenciales de dos colas en las que la hipótesis nula puede ser rechazada por cualquier resultado no aleatorio en cualquiera de las dos direcciones, (para una prueba direccional, la hipótesis nula sólo puede ser rechazada si la diferencia está en la dirección especificada con anterioridad).

Hipótesis nula o hipótesis estadística: es un planteamiento que especifica un valor numérico para un parámetro de población.

Hipótesis: es una afirmación que especifica un valor numérico para un parámetro. Suposición sujeta a prueba. Enunciado que intenta captar lo que rige el desarrollo de un hecho que se pretende comprender.

H_0 (hipótesis estadística nula): es una afirmación que especifica un valor numérico para un parámetro de población.

Intercepción de la regresión: es la constante aditiva en la ecuación de regresión para predecir y a partir de x .

Intervalo de confianza .95 (o IC 0.95): especifica un rango de valores dentro del cual el parámetro objetivo reside en 95% de las aplicaciones.

Límites de confianza: consiste de un límite inferior y un límite superior entre los que se presume que cae el parámetro objetivo. Valores que establecen la frontera del intervalo de confianza.

Línea de regresión: es la línea recta de “mejor ajuste” para predecir valores de criterio (y), bisecta el enjambre de puntos que componen la gráfica de dispersión, conectando de esta forma las medias del criterio predichas para todos los valores de x .

r de Pearson: medida de la forma en que una pareja de datos ocupa posiciones iguales o opuestas dentro de sus propias distribuciones.

Regresión: es un procedimiento estadístico para predecir el desempeño en variables de criterio de una o más variables predictoras.

Regresión múltiple: es un procedimiento estadístico para predecir el desempeño en variables de criterio a partir de dos o más variables predictoras.

Relaciones curvilíneas entre pares de variables: se indican cuando los enjambres de puntos que componen la gráfica de dispersión tienden a seguir una curva, más que una línea recta.

Preguntas frecuentes

Unidad 1

¿Qué es la investigación científica?

Es el proceso sistemático, controlado, empírico y crítico, de proposiciones hipotéticas sobre las presumidas relaciones entre fenómenos naturales.

¿Qué es la investigación social?

Es un proceso sistemático, controlado, empírico y crítico de aseveraciones hipotéticas sobre las posibles relaciones sociales que presentan los sujetos en lo individual y/o en lo colectivo.

¿Qué es la estadística?

Es aquella que se ocupa de los métodos y procedimientos para recoger, clasificar, resumir, hallar regularidades y analizar los *datos*, siempre y cuando la variabilidad e *incertidumbre* sea una causa intrínseca de los mismos; así como de realizar *inferencias* a partir de ellos, con la finalidad de ayudar a la toma de *decisiones* y en su caso formular *predicciones*.

¿Cuál es el objeto de la estadística?

Resumir los datos más destacados de los elementos que componen un conjunto, logrando así aprehender más fácilmente su contenido.

¿Qué es la estadística descriptiva?

Es la rama de la estadística que recolecta, recuenta, presenta y describe un conjunto de datos.

¿Qué es la estadística inferencial?

Es aquella que proporciona los métodos para estimar las características de un grupo total (población), basándose en datos de un conjunto pequeño (muestra) de observaciones.

¿En que se relacionan la investigación social y la estadística?

Son proceso de constante exploración y descubrimiento, de carácter universal, con esquemas metodológicos que permiten el estudio minucioso de los fenómenos sociales.

Unidad 2

¿Qué es una población?

Todos lo miembros, elementos, observaciones o valores que se ajustan a un criterio específico. Conjunto completo de individuos, objetos o datos en cuyo estudio esta interesado un investigador.

¿Qué es el muestreo?

Conjunto de procedimientos que se ejecutan para seleccionar una muestra.

¿Qué es una muestra?

Es un subconjunto de la totalidad de un conjunto de elementos, seres u objetos que se pretende investigar. Es una fracción o subgrupo de una población.

¿Qué relación hay entre población muestreo y muestra?

La población es el conjunto total de datos a los cuales se les aplica un procedimiento llamado muestreo para obtener un tamaño de muestra o muestra previamente calculada.

¿Qué es el muestreo probabilístico?

Es aquél en el que todos los elementos de una población tienen la misma probabilidad de ser elegidos.

¿Qué es el muestreo no probabilístico o determinístico?

Es aquél en el que la elección de los elementos no depende de la probabilidad, sino de las características determinadas por el investigador o por quien diseña la muestra.

¿Qué diferencia hay entre el muestreo probabilística y el no probabilística?

El muestreo probabilístico parte de la idea que todos los elementos muestrales tienen la misma probabilidad de ser elegidos, el concepto de no muestreo probabilístico en cambio asume que las características de la muestra serán determinados de acuerdo a criterios del investigador.

Unidad 3

¿Qué es el método estadístico?

Es el conjunto de procedimientos que recolecta, cuenta, presenta, describe y analiza los datos, a fin de extraerles el máximo de información, base para la toma de decisiones.

¿Cuántas y cuáles son las etapas del método estadístico?

Cinco y son: 1) recolectar, 2) contar, 3) presentar, 4) describir y 5) analizar.

¿En qué consiste la etapa de recolección?

Es la etapa del método estadístico donde se acopia la información cualitativa y cuantitativa a través de variables estableciendo su nivel de medición.

¿En qué consiste la etapa de recuento?

Es la etapa del método estadístico donde los datos son sometidos a revisión, clasificación y cómputo numérico.

¿En qué consiste la etapa de presentación?

Es la etapa del método estadístico que mediante el uso de tablas o cuadros y graficas permite una inspección precisa y rápida de los datos.

¿En qué consiste la etapa de descripción?

Es la etapa del método estadístico donde los datos se resumen en forma de medidas que permiten expresar las principales propiedades numéricas de los datos.

¿En qué consiste la etapa de análisis?

Es la etapa del método estadístico que proporciona los métodos para estimar las características de un grupo total (población), basándose en datos de un conjunto pequeño (muestra) de observaciones.

Unidad 4

¿Qué es una variable?

Es una característica susceptible de medirse.

¿Cómo se clasifican las variables?

Se clasifican en cualitativas (nominal y ordinal) y cuantitativas (discretas y continuas).

¿Cuáles son los niveles de medición de las variables?

Nominal, ordinal, discretas o continuas.

¿Cuál es la diferencia entre el nivel de medición nominal y ordinal?

El nivel nominal asigna características o atributos a la variable sin importar orden ni jerarquía en el caso del nivel ordinal se asigna características o atributos considerando jerarquía u orden.

Unidad 5

¿Qué es una frecuencia absoluta?

La frecuencia absoluta se determina como el número de veces que se repite un dato.

¿Qué es una frecuencia relativa?

Las frecuencias relativas representan el porcentaje de veces en que ocurre un dato.

¿Qué es una frecuencia absoluta acumulada?

La frecuencia absoluta acumulada para un determinado valor se considera como la frecuencia de cada dato más la suma de los valores anteriores a dicha suma.

¿Qué es una frecuencia relativa acumulada?

Es la suma de los valores acumulados de la frecuencia relativa. La suma de las frecuencias relativas que se han acumulado, incluyendo la clase sobre la que se está calculando la frecuencia relativa.

¿Cuáles son las características mínimas para la presentación de cuadros o tablas?

Cada cuadro debe tener un título completo que responda a las preguntas: ¿Qué se está presentando?, ¿Cómo se está presentando?, ¿De dónde provienen los datos? y ¿De cuándo son los datos?.

Las variables deben tener un nombre exacto que este acorde con la escala de clases o modalidades. La escala de cada variable debe tener modalidades o clases que sean exhaustivas y mutuamente excluyentes.

¿Cuáles son las gráficas más adecuadas para variables cualitativas?

Gráficos de barras separadas, circular o pictograma.

¿Cuáles son las graficas más adecuadas para variables cuantitativas?

Histograma, polígono de frecuencias, polígono de frecuencias acumuladas, diagrama de caja.

Unidad 6

¿Qué es una proporción?

Es la división entre un subconjunto y el conjunto al que pertenece.

¿Qué es un porcentaje?

Es una proporción multiplicada por 100.

¿Qué es una razón?

Es la división entre dos elementos de igual o de distinta naturaleza.

¿Qué es una tasa?

Es la división entre los eventos que ocurren de un determinado fenómeno entre la población en donde puede ocurrir dicho evento por un múltiplo de 10.

¿Qué es la moda?

Es el valor que más se repite en una serie de datos.

¿Qué es la media?

Es el valor que tendrían todos los datos si estos fueran de igual valor.

¿Qué es la mediana?

Es el valor que divide a una serie de datos en dos partes de igual tamaño.

¿Qué es un centil?

Es aquel valor que divide a una serie de datos en partes porcentualmente complementarias.

¿Qué es el rango?

Es la diferencia entre el valor máximo y el valor mínimo de una serie de datos.

¿Qué es la varianza?

Es el cuadrado de la desviación estándar. Es el promedio de desviaciones o diferencias cuadráticas de cada valor de una serie con respecto al promedio de dicha serie.

¿Qué es la desviación media?

Es la magnitud o distancia entre cada observación o dato y la media aritmética.

¿Qué es la desviación estándar?

Es la raíz cuadrada de la varianza. Es el promedio de desviación de las puntuaciones con respecto a la media.

¿Qué es el coeficiente de variación?

Es una relación o razón estadística entre la desviación estándar y la media aritmética multiplicada por 100.

¿Qué es el sesgo?

El sesgo se define como la falta de simetría en una distribución.

¿Qué es la curtosis?

La curtosis se describe como el grado en que las proporciones observadas difieren de la curva normal.

Unidad 7

¿Qué es el coeficiente de correlación?

Son mediciones descriptivas que muestran la dirección y grado de la relación entre dos variables. Expresión cuantitativa de la magnitud y dirección de una relación.

¿Qué es el coeficiente de Pearson?

Determina si existe relación lineal entre dos variables a nivel intervalar.

¿Qué es el coeficiente de correlación Phi?

Determina si existe relación lineal entre dos variables.

¿Qué es el coeficiente de correlación Biserial Puntual?

Determina si existe relación lineal entre dos variables a nivel nominal con dos valores cada una (dicotómicas).

¿Qué es el coeficiente de correlación de Spearman?

Determina si existe relación lineal entre dos variables a nivel ordinal.

¿Qué es la correlación?

Se ocupa de establecer si existe una relación entre dos o más variables, así como su magnitud y dirección.

¿Qué es la regresión?

Es un procedimiento estadístico para predecir el desempeño en variables de criterio de una o más variables predictoras.

Material complementario

ANDER-EGG, Ezequiel, *Técnicas de investigación social*, Lumen, Argentina, 1995.

BABBIE, Earl, *Fundamentos de la investigación social*, Thomson Internacional Editores, México, 1999.

CASTAÑEDA, Juan, DE LA TORRE, María Olivia, MORÁN, José Manuel y LARA, Luz Patricia, *Metodología de la investigación*, McGraw-Hill, México, 2002.

COZBY, Paul C., *Métodos de investigación del comportamiento*, Mc Graw Hill, México, 2004.

ELORZA, Haroldo, *Estadística para las ciencias sociales y del comportamiento*, Oxford University Press, México, 2000.

GARRIDO LUQUE, Alicia, ALVARADO ESTRAMINA, José Luis. *Técnicas de análisis estadístico en ciencias sociales*. Servicio de publicaciones. Universidad Complutense. España. 1995.

GLASS, G. *Métodos estadísticos aplicados a las ciencias sociales*, México, Prentice Hall, 1986.

GONICK, Larry y SMITH, Woolcott, *La estadística en cómic*, Zendera Zariquiey, España, 1999.

GUERRERO, V. *Estadística básica para estudiantes de economía y otras ciencias sociales*, México, Fondo de Cultura Económica, 1989.

HERNÁNDEZ, Roberto, FERNÁNDEZ, Carlos y BAPTISTA, Pilar, *Metodología de la investigación*, McGraw-Hill, México, 2003.

HOLGUÍN, F. *Estadística descriptiva aplicada a las ciencias sociales*, México, Facultad de Ciencias Políticas y Sociales-UNAM, 1981.

LEACH C. *Fundamentos de estadística; enfoque no paramétrico para ciencias sociales*, México, Limusa, 1982.

LEVIN, J. *Fundamentos de estadística en la investigación social*. México, Haria, 1992.

MATEOS J. *Estadística en investigación social: ejercicios resultados*, España, Paraninfo, 1989.

Mc GUIGAN, F. *Psicología experimental: enfoque metodológico*, México, Trillas, 1983.

NÚÑEZ DEL PRADO A. *Estadística básica para planificación*, México, Siglo XXI, 1990.

PAGANO, R. Robert, *Estadística para las ciencias del comportamiento*, Thomson Internacional Editores, México, 1999.

PÉREZ, B. *Estadística para las ciencias sociales*, México, UAM-Iztapalapa, 1992.

PICK, Susan y LÓPEZ, Ana Luisa, *Cómo investigar en ciencias sociales*, Trillas, México, 1998.

PORTILLA, E. *Estadística primer curso*, México, Mc Graw Hill, 1992.

REYNAGA O., J. DE GARAY G. B. y GARCÍA R. J. *Módulo Preparatorio*. Unidad de Bioestadística, Depto. de Medicina Social, Preventiva y Salud Pública. Facultad de medicina UNAM, México. 1980.

RITCHEY Ferris J. *Estadística para las ciencias sociales, El potencial de la imaginación estadística*, México. Mc Graw Hill. 2004.

SELLTIZ Claire, WRIGHTSMAN Lawrence S. y COOK Stuart W., *Métodos de investigación en ciencias sociales*, RIALP, España, 1980.

SIERRA B. Restituto. *Diccionario Practico de Estadística*, Paraninfo. España,1991.

Universidad Nacional Autónoma de México

Dr. Juan Ramón de la Fuente
Rector

Lic. Enrique del Val Blanco
Secretario General

Mtro. Daniel Barrera Pérez
Secretario Administrativo

Dra. Rosaura Ruiz Gutiérrez
Secretaria de Desarrollo Institucional

Mtro. José Antonio Vela Capdevila
Secretario de Servicios a la Comunidad Universitaria

Mtro. Jorge Islas López
Abogado General

Escuela Nacional de Trabajo Social

Mtro. Carlos Arteaga Basurto
Director

Dr. Guillermo Campos y Covarrubias
Secretario General

Mtro. Salvador Alvarado Garibaldi
Secretario Académico

Mtra. Laura Ortega García
Jefa de la División de Estudios Profesionales

Mtro. Jorge Hernández Valdes
Jefe de la División de Estudios de Posgrado

Mtra. Juana Leticia Cano Soriano
Secretaria de Apoyo y Desarrollo Escolar

Mtra. Margarita Pérez Durán
Coordinadora del Sistema Universidad Abierta y
Educación a Distancia

Estadística Aplicada a la Investigación Social I

Lic. Ciro López Mendoza

1ª impresión, marzo 2006

Mtra. Margarita Pérez Durán
Coordinadora del Sistema de Universidad Abierta y Educación a Distancia

Diseño de portada: David Díaz Vázquez

D.R. 2003, Universidad Nacional Autónoma de México
Circuito Exterior s/n, Ciudad Universitaria
C.P. 04510, Coyoacán, México, D.F.
Escuela Nacional de Trabajo Social
www.trabajosocial.unam.mx
suad_ents@yahoo.com.mx
ISBN 000-00-0000-0

Impreso y hecho en México
Printed and made in México