



Universidad Nacional de Educación
Enrique Guzmán y Valle
Alma Máter del Magisterio Nacional

Prefacio:

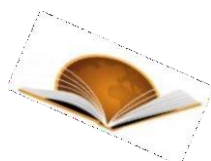
La asignatura es de naturaleza práctico - teórico, orientado a desarrollar en el estudiante habilidades superiores del pensamiento para el razonamiento lógico y creativo, solución de problemas y la toma de decisiones. El curso está orientado a proporcionar al estudiante conocimientos estadísticos fundamentales sobre las técnicas de investigación estadística para recoger, analizar y mostrar información confiable y de calidad necesaria para la toma de decisiones. La asignatura está diseñada para que el alumno al final de cada clase desarrolle casos prácticos en base a datos reales.



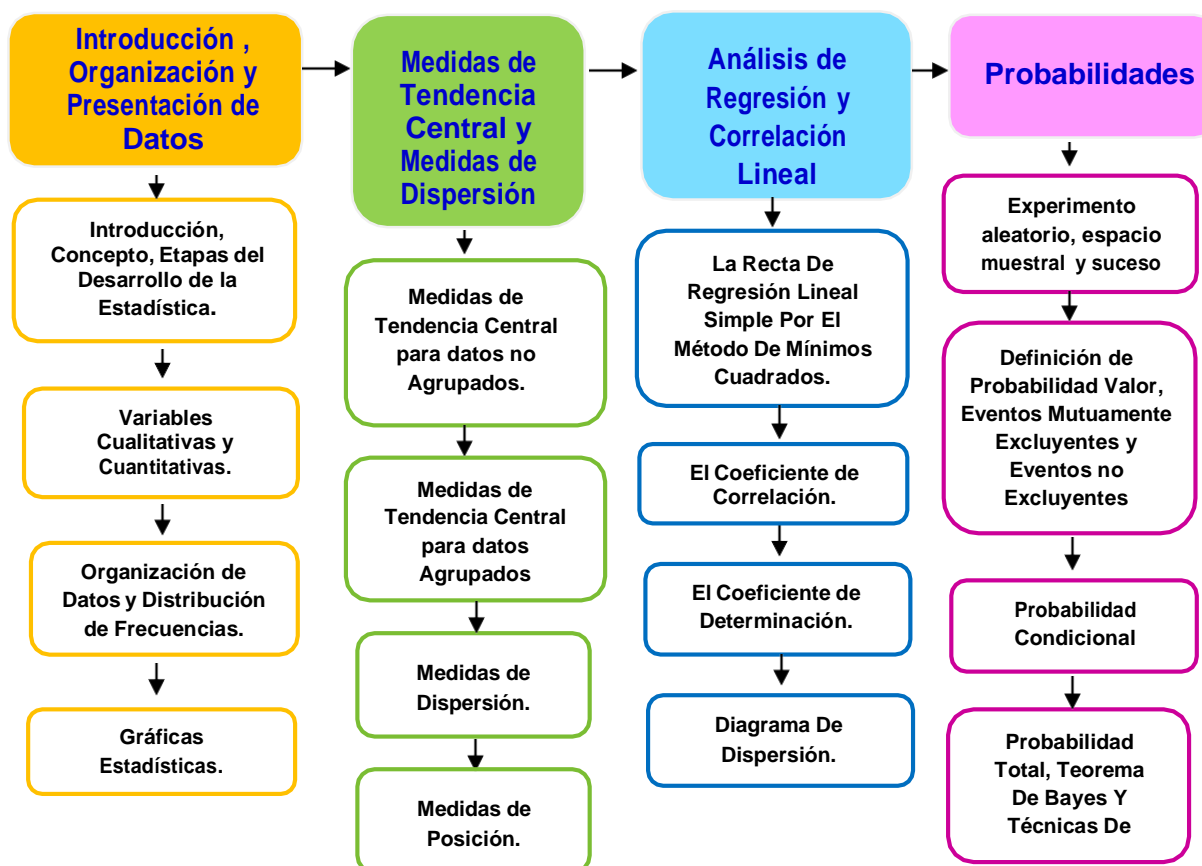
Comprende cuatro Unidades de Aprendizaje:

- Unidad I: **Introducción, Organización Y Presentación De Datos.**
- Unidad II: **Medidas De Tendencia Central Y Medidas De Dispersión.**
- Unidad III: **Análisis De Regresión Y Correlación Lineal.**
- Unidad IV: **Probabilidades.**





Estructura de los Contenidos



La competencia que el estudiante debe lograr al final de la asignatura es:

“Aplicar técnicas estadísticas para la recolección, presentación, análisis e interpretación de datos estadísticos.”



Índice del Contenido

I. PREFACIO	02
II. DESARROLLO DE LOS CONTENIDOS	03 - 151
UNIDAD DE APRENDIZAJE 1: INTRODUCCION , ORGANIZACION Y PRESENTACION DE DATOS	05-45
1. Introducción	06
a. Presentación y contextualización	06
b. Competencia (logro)	06
c. Capacidades	06
d. Actitudes	06
e. Ideas básicas y contenido	06
2. Desarrollo de los temas	07-39
a. Tema 01: Introducción, Concepto, Etapas del Desarrollo de la Estadística.	07
b. Tema 02: Variables Cualitativas y Cuantitativas.	14
c. Tema 03: Organización de Datos y Distribución de Frecuencias.	19
d. Tema 04: Gráficas Estadísticas.	28
3. Lecturas recomendadas	40
4. Actividades	40
5. Autoevaluación	41
6. Resumen	45
UNIDAD DE APRENDIZAJE 2: MEDIDAS DE TENDENCIA CENTRAL Y MEDIDAS DE DISPERSION	46-83
1. Introducción	47
a. Presentación y contextualización	47
b. Competencia (logro)	47
c. Capacidades	47
d. Actitudes	47
e. Ideas básicas y contenido	47
2. Desarrollo de los temas	48-77
a. Tema 01: Medidas de Tendencia Central para datos no Agrupados.	48
b. Tema 02: Medidas de Tendencia Central para datos Agrupados.	54
c. Tema 03: Medidas de Dispersión.	66
d. Tema 04: Medidas de Posición.	73
3. Lecturas recomendadas	78
4. Actividades	78
5. Autoevaluación	79
6. Resumen	83
UNIDAD DE APRENDIZAJE 3: ANALISIS DE REGRESION Y CORRELACION LINEAL	84-114
1. Introducción	85
a. Presentación y contextualización	85
b. Competencia (logro)	85
c. Capacidades	85
d. Actitudes	85
e. Ideas básicas y contenido	85
2. Desarrollo de los temas	86-105
a. Tema 01: La Recta De Regresión Lineal Simple Por El Método De Mínimos Cuadrados.	86
b. Tema 02: El Coeficiente de Correlación.	90
c. Tema 03: El Coeficiente de Determinación.	84
d. Tema 04: Diagrama De Dispersión.	95
3. Lecturas recomendadas	106
4. Actividades	107
5. Autoevaluación	108
6. Resumen	114
UNIDAD DE APRENDIZAJE 4: PROBABILIDADES	115-148
1. Introducción	116
a. Presentación y contextualización	116
b. Competencia	116
c. Capacidades	116
d. Actitudes	116
e. Ideas básicas y contenido	116
2. Desarrollo de los temas	117-143
a. Tema 01: Experimento aleatorio, espacio muestral, suceso	117
b. Tema 02: Definición De Probabilidad, Valor, Eventos Mutuamente Excluyentes Y Eventos No Excluyentes	122
c. Tema 03: Probabilidad Condicional.	127
d. Tema 04: Probabilidad Total, Teorema de Bayes y Tecnicas de Conteo.	121
3. Lecturas recomendadas	133
4. Actividades	144
5. Autoevaluación	145
6. Resumen	148
III. GLOSARIO	149
IV. FUENTES DE INFORMACION	150
V. SOLUCIONARIO	151

UNIDAD 1



Introducción, Organización y Presentación de Datos

Introducción

a) Presentación y contextualización

Los temas que se tratan en la presente unidad temática, tiene por finalidad que el estudiante comprenda, las nociones básicas de la estadística, tablas y gráficos estadísticos.

Para poder hacer que el alumno pueda usar estos conocimientos en su vida diaria, problemas simples, y dar un enfoque más analítico con respecto a los problemas.

b) Competencia

Recopila, organiza, sistematiza la información estadística, y representa mediante gráficos estadísticos.

c) Capacidades

1. Define y explica la importancia de la estadística y sus etapas.
2. Describe y aplica los diferentes tipos de variables en la estadística descriptiva.
3. Explica las maneras de cómo organizar datos y distribuir frecuencias.
4. Explica la estructura de cada uno de los gráficos usados en la estadística descriptiva.

d) Actitudes

- ✓ Toma iniciativa y lidera al equipo en el cumplimiento de las actividades asignadas a su vez promueve actividades y toma de decisiones pertinentes.
- ✓ Planifica y cumple oportunamente sus tareas o actividades diarias y presenta sus trabajos en forma organizada.

e) Presentación de Ideas básicas y contenido esenciales de la Unidad:

La Unidad de Aprendizaje 01: Introducción, Organización Y Presentación De Datos, comprende el desarrollo de los siguientes temas:

TEMA 01: Introducción, Concepto, Etapas del Desarrollo de la Estadística.

TEMA 02: Variables Cualitativas y Cuantitativas.

TEMA 03: Organización de Datos y Distribución de Frecuencias.

TEMA 04: Gráficas Estadísticas.

.Introducción

.Conceptos

SEMANA 1

.Clases

De estadística

.Tipos de Datos



Competencia:

Definir y explicar la importancia de la estadística y sus etapas.



Desarrollo de los Temas



Tema 01: Introducción, Concepto, Clases de estadística y Tipo de datos

A. Introducción:

La palabra "estadística" suele utilizarse bajo dos significados distintos, a saber:

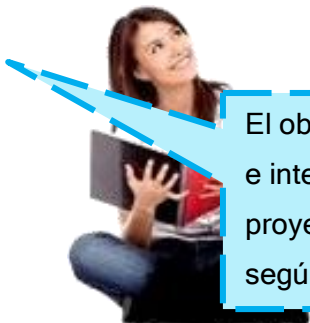
1º Como colección de datos numéricos.- Esto es el significado más vulgar de la palabra estadística. Se sobreentiende que dichos datos numéricos han de estar presentados de manera ordenada y sistemática.

Una información numérica cualquiera puede no constituir una estadística, para merecer este apelativo, los datos han de constituir un conjunto coherente, establecido de forma sistemática y siguiendo un criterio de ordenación.



2º Como ciencia.- En este significado, La Estadística estudia el comportamiento de los fenómenos de masas. Como todas las ciencias, busca las características generales de un colectivo y prescinde de las particulares de cada elemento. Al investigar el sexo de los nacimientos, iniciaremos el trabajo tomando un grupo numeroso de nacimientos y obtener después la proporción de varones.

Es muy frecuente enfrentarnos con fenómenos en los que es muy difícil predecir el resultado; así, no podemos dar una lista, con las personas que van a morir con una cierta edad, o el sexo de un nuevo ser hasta que transcurra un determinado tiempo de embarazo.



El objetivo de la estadística como ciencia es recopilar, e interpretar datos que en el futuro servirán para proyectar posibles problemáticas futuras, consiguiendo según estos datos, la solución más viable y rápida.

B. Concepto:

Es una ciencia aplicada que nos proporciona un conjunto de métodos o técnicas para:

- **Recopilar.**
- **Organizar.**
- **Presentar Datos.**
- **Analizar Datos.**



¿Quiénes Usan La Estadística?



Los métodos estadísticos han encontrado en la actualidad aplicación en el Gobierno, la administración de negocios, las Ciencias Sociales, la Psicología, las Ingenierías, las Ciencias Físicas y Naturales y en muchos otros campos de la actividad intelectual.

Algunos ejemplos:



En Los Organismos De Gobierno. Los diferentes indicadores económicos, tales como índices de precios al por mayor y al consumidor, las tasas de interés, las fluctuaciones del mercado bursátil y el índice de producción industrial, no solamente describen el estado actual de la economía, sino que proporcionan pistas acerca de sus futuras tendencias. Con el auxilio de tales indicadores, los encargados de las políticas de los distintos organismos serían capaces de tomar decisiones más racionales con respecto a sus operaciones.



En La Administración De Negocios. La creciente complejidad de la economía ha provocado un terrible grado de incertidumbre acerca de las operaciones futuras de cualquier empresa de negocios. Más y más compañías están usando el análisis estadístico como herramienta para la toma de decisiones, especialmente en áreas tales como investigación de mercados, predicciones y planeación a largo plazo en lo referente a recursos financieros y humanos.

En La Educación Y En La Psicología. La necesidad de analizar e interpretar datos numéricos ha hecho necesario para educadores y para psicólogos tener al menos alguna comprensión básica de los métodos estadísticos. La necesidad del psicólogo de herramientas estadísticas especiales ha llevado al desarrollo de nuevas técnicas estadísticas en las últimas décadas.

En las Ciencias Biológicas y en la Medicina.

En la agricultura, se utilizan para determinar los efectos de clases de semillas, de insecticidas y de fertilizantes en los campos. Se emplea también para determinar los posibles efectos laterales o la efectividad de las medicinas y para proporcionar mejores métodos para controlar la diseminación de enfermedades contagiosas.



En la Sociología, en la Antropología y en las Ciencias del Comportamiento.

Las técnicas estadísticas se han aplicado a una amplia variedad de proyectos de investigación que impliquen el estudio de individuos y de grupos.

En La Ingeniería. La aplicación de los principios estadísticos al control de calidad ha sido una práctica aceptada durante varias décadas.

C. Clases de estadística

La historia de la estadística está resumida en tres grandes Clases o etapas o fases:

1.- Los Censos: Desde el momento en que se constituye una autoridad política, la idea de inventariar de una forma más o menos regular la población y las riquezas existentes en el territorio está ligada la conciencia de soberanía y con los primeros esfuerzos administrativos.

Los comienzos de la estadística pueden ser hallados en el antiguo Egipto, cuyos faraones lograron recopilar, hacia el año 3050 antes de Cristo, prolijos datos relativos a la población y a las riquezas del país.



2.- De La Descripción A La Aritmética Política: Las ideas mercantilistas entrañan una intensificación de este tipo de investigación. Colbert multiplica las encuestas sobre artículos manufacturados, el comercio y la población. Vauban, quien hace la primera propuesta de un impuesto sobre los ingresos, se señala como el verdadero precursor de los sondeos. La escuela inglesa proporciona un nuevo progreso al superar la fase puramente descriptiva. Uno de sus principales exponentes Petty es autor de la famosa Aritmética Política. Chaptal, ministro del interior francés, publica, en 1801, el primer censo general de población y desarrolla estudios industriales, de las producciones y de los cambios, los cuales se hicieron sistemáticos durante las dos terceras partes del siglo XIX.

3.- Estadística Y Cálculo De Probabilidades: El cálculo de probabilidades se incorpora, rápidamente, como un instrumento de análisis extremadamente poderoso para el estudio de los fenómenos económicos y sociales y, en general, para el estudio de fenómenos "cuyas causas son demasiado complejas para conocerlas totalmente y hacer posible su análisis". Godofredo Achenwall, profesor de la Universidad de Gotinga, acuñó, en 1760, la palabra estadística, que extrajo del término italiano statista (estadista). Creía, y con sobrada razón, que los datos de la nueva ciencia serían el aliado más eficaz del gobernante consciente.

D. Población, Elementos Y Caracteres.

Es obvio que todo estudio estadístico ha de estar referido a un conjunto o colección de personas o cosas. Este conjunto de personas o cosas es lo que denominaremos población.

Población: conjunto de todos los individuos (personas, objetos, animales, etc.) que porten información sobre el fenómeno que se estudia.

Por ejemplo: si estudiamos el precio de la vivienda en una ciudad, la población será el total de las viviendas de dicha ciudad.



Individuo: cualquier elemento que porte información sobre el fenómeno que se estudia. Así, si estudiamos la altura de los niños de una clase, cada alumno es un individuo; si estudiamos el precio de la vivienda, cada vivienda es un individuo.

Las personas o cosas que forman parte de la población se denominan **elementos**. En sentido estadístico un elemento puede ser algo con existencia real, como un automóvil o una casa, o algo más abstracto como la temperatura, un voto, o un intervalo de tiempo.

Luego por tanto de cada elemento de la población podremos estudiar uno o más aspectos cualidades o caracteres.

La población puede ser según su tamaño de dos tipos:

Población Finita:

Cuando el número de elementos que lo forman, se pueden enumerar, por ejemplo el número de alumnos de un centro de enseñanza, o grupo clase.



Población Infinita:

Cuando la cantidad de elementos que la forman no es posible numerarlo. Como por ejemplo si se realizase un estudio sobre los productos que hay en el mercado. Hay tantos y de tantas calidades que esta población podría considerarse infinita.



Muestra:

Subconjunto que seleccionamos de la población. Así, si se estudia el precio de la vivienda de una ciudad, lo normal será no recoger información sobre todas las viviendas de la ciudad (sería una labor muy compleja), sino que se suele seleccionar un subgrupo muestra) que se entienda que es suficientemente representativo.



PRIMERA UNIDAD: Organización y presentación de datos

Semana 1. Conceptos básicos, clases de estadística, tipos de datos. Semana 2. Recolección de datos, escalas de medición.



Semana 3. Distribuciones de frecuencias para datos cualitativos, Tablas y gráficos estadísticos. Semana 4. Distribuciones de frecuencias para datos cuantitativos, Tablas y gráficos estadísticos. Semana 5. Práctica calificada y evaluación primera lectura.

Variables Cualitativas y

TEMA 2

Cuantitativas



Competencia:

Describir y aplicar los diferentes tipos de variables en la estadística descriptiva.



Las variables pueden ser clasificadas como cuantitativas (intervalares) o cualitativas (categóricas), dependiendo si los valores presentados tienen o no un orden de magnitud natural (cuantitativas), o simplemente un atributo no sometido a cuantificación (cualitativa).



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. Un atributo corresponde a un valor

específico de una variable, como es el caso de la variable sexo, la que posee dos atributos: varón o mujer.

En variables que exploran el grado de acuerdo o desacuerdo frente a una afirmación los atributos podrían ser:



1 = muy en desacuerdo

2 = en desacuerdo

3 = indiferente

4 = de acuerdo

5 = muy de acuerdo

Dependiendo de los valores que puede tener una variable cualitativa, ésta puede a su vez ser dicotómicas (cuando sólo pueden adoptar un sólo valor sin jerarquía entre sí; hombre - mujer, positivo-negativo, presente-ausente), o bien, poli o multicotómicas, si existe la posibilidad de que adopten múltiples valores (edad, talla, nivel socioeconómico, grupos sanguíneos, calificación previsional de usuarios).

1. Las escalas de medición de una variable cualitativa son:

Nominal.- Nominal, En este nivel de medición se establecen categorías distintivas que no implican un orden específico.

Ejemplo:

Nombres de personas, de establecimientos, raza, grupos sanguíneos, estado civil. Estas variables no tienen ningún orden inherente a ellas ni un orden de jerarquía.



Ordinal.- Cuando se establecen categorías con dos o mas niveles que implican un orden inherente entre sí.

Ejemplo:

Grados de desnutrición, respuesta a un tratamiento, nivel socioeconómico, intensidad de consumo de alcohol, días de la semana, meses del año.

A pesar de este orden jerárquico no es posible obtener

valoración numérica lógica entre dos valores.

1. Las variables de tipo cuantitativo pueden a su vez ser clasificadas como continuas o discretas. Las escalas cuantitativas son reconocidas también como escalas intervalares o numéricas.



Continua.- Cuando entre dos valores determinados existen infinitas posibilidades de valores. **Ejemplo:** El peso, la talla, la presión arterial o el nivel de colesterol sérico.

Discreta.- Cuando la variable a medir sólo puede adoptar un sólo valor numérico, entero, con valores intermedios que carecen de sentido **Ejemplo:** El número de hijos, de unidades vecinales del sector, número de exámenes de laboratorio o de pacientes atendidos.



Continuas

Ej. Presión arterial, peso, edad, talla, IMC (Índice de Masa Corporal)

Discretas

Ej. Número de hijos, episodios de infección urinaria

Categorías (cualitativas)



Dicotómicas

Ej. Sexo masculino y femenino; vivo/muerto

Politómicas

Ej. Grupo sanguíneo, raza, estado civil, grado de instrucción

En la práctica, salvo contadas excepciones no se dispone de métodos de medición sofisticados como para poder medir exactamente los valores. Tanto las variables discretas como las continuas pueden agruparse construyendo intervalos, entre cuyos valores extremos se ubicarán las diferentes observaciones registradas. Sin embargo, estrictamente hablando, sólo las variables continuas pueden ser objeto de categorización mediante intervalos.

Clasificación de Variables

NOTA:

Las variables también se pueden clasificar en:

Variables unidimensionales:

Sólo recogen información sobre una característica (por ejemplo: edad de los alumnos de una clase).

Variables Bidimensionales:

Recogen información sobre dos características de la población (por ejemplo: edad y altura de los alumnos de una clase).

Variables pluridimensionales:

Recogen información sobre tres o más características (por ejemplo: edad, altura y peso de los alumnos de una clase).

Ordenables:





Aquellas que sugieren una ordenación, por ejemplo la graduación militar, El nivel de estudios, etc.

No ordenables:

Aquellas que sólo admiten una mera ordenación alfabética, pero no establece orden por su naturaleza, por ejemplo el color de pelo, sexo, estado civil, etc.

Organización de Datos y Distribución de Frecuencias

TEMA 3



Competencia:

Explicar las maneras de cómo organizar datos y distribuir frecuencias.





Tema 03: Organización De Datos Y Distribución De Frecuencias

Organización de los Datos Obtenidos de una Muestra

Cuando se han recopilado datos mediante un muestreo o un censo, la primera inquietud que aparece es sobre la manera en la que se puede realizar un análisis descriptivo apropiado con la información recolectada de manera que resulte sencillo entender lo que ocurre en la población de la que se han captado las observaciones.

En este tema se proporcionan algunos procedimientos para la tabulación de datos que conducen a la formación de cuadros o tablas de frecuencias.



Organización De Los Datos Cualitativos

*MG.
NSHT*

Antes de iniciar el trabajo de organización de datos cualitativos, es necesario determinar si éstos corresponden a variables cualitativas nominales u ordinales. Si los datos son cualitativos nominales, se forman categorías que pueden ser presentadas en cualquier orden: por ejemplo los colores de preferencia de las personas. Si los datos son ordinales, entonces deben estar asociados a algún orden en su presentación.

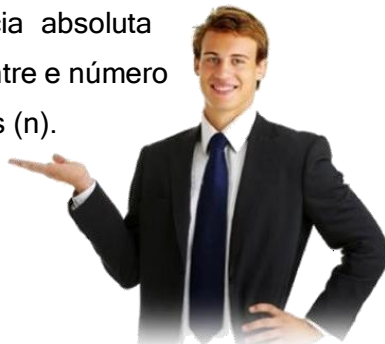
Una vez definido el tipo de variable, se obtiene mediante un proceso de conteo las frecuencias absolutas (número de veces que se repite cada respuesta), luego las frecuencias relativas (división de cada frecuencia absoluta entre el tamaño de muestra) y/o los porcentajes de cada respuesta (cada frecuencia relativa multiplicada por 100). También se puede encontrar las



frecuencias absolutas acumuladas (F_i)

Frecuencia relativa (h_i).

Es la **proporción** del total de observaciones que caen dentro de cada modalidad o valor. Se obtiene dividiendo la frecuencia absoluta (f_i) de la modalidad entre el número total de observaciones (n).



Frecuencia acumulada (F_i).

Para cada clase, valor o modalidad, la frecuencia acumulada equivale a la frecuencia absoluta (f_i) de la fila sumada a la frecuencia acumulada de la fila anterior. Para la primera fila, la frecuencia acumulada equivale simplemente a la frecuencia absoluta de la misma fila.

Ejemplo 1. Una revista conocida efectuó una encuesta respecto a lo adecuado de la protección policial en la ciudad. Se seleccionó un total de 419 personas. Las respuestas se presentan en la siguiente tabla de frecuencias:

Respuesta	Frecuencia absoluta	Frecuencia relativa	Porcentaje
Si	293	0.6993	69.93
No	80	0.1909	19.09
No sabe/ no responde	46	0.1098	10.98



Modalidad	Frecuencia absoluta	Frecuencia relativa	Porcentaje
Primaria completa	125	0.625	62.5
Secundaria completa	70	0.35	35
Educac.superior completa	5	0.025	2.5
	200	1	100

Organización De Datos Cuantitativos Discretos

Cuando se tienen datos cuantitativos discretos cuyo número de resultados es grande, la información puede ser clasificada y presentada directamente de la identidad de la misma. En estos casos, primero se ordenan los posibles valores de la variable según su magnitud, y a continuación se obtienen, mediante un proceso de conteo, las frecuencias absolutas asociadas a cada uno de dichos valores; las frecuencias relativas y porcentuales se obtienen de manera similar a lo descrito para las variables cualitativas.



Ejemplo. Consideremos la variable número de hijos y tomemos las observaciones de

MG.
NSHT

una muestra de 150 familias de zonas marginales de Lima Metropolitana. Los valores obtenidos se pueden agrupar en diferentes valores: 0 hijos, 1 hijo, 2 hijos, 3 hijos o 4 hijos. Para hacer un arreglo de estas observaciones, usaremos una tabla como la siguiente:

Organización De Datos Cuantitativos Continuos

Número de hijos	Frecuencia absoluta	Frecuencia relativa	Porcentaje
0	2	0.013	1.3
1	15	0.100	10.0
2	40	0.267	26.7
3	55	0.367	36.7
4	38	0.253	25.3
TOTAL	150	1	100

Ejemplo. Tomamos una muestra de 100 niños de 10 años de edad para estudiar su estatura.

Entonces la variable estatura que es cuantitativa continua se puede presentar en una tabla del siguiente tipo:

INTERVALOS DE CLASE	FRECUENCIA(fi)
[1.0mt.- 1.15mt]	3
[1.15mt.- 1.30mt]	39
[1.30mt.- 1.45mt]	55
[1.45mt.- 1.60mt]	3
TOTAL	100



A la organización de las observaciones de una muestra en una tabla para expresar la frecuencia de cada una de sus modalidades o valores se le conoce como **distribución de frecuencias**. En las distribuciones de frecuencia de las variables cuantitativas

continuas, también se acostumbra colocar otras columnas además de la frecuencia absoluta (**fi**), estas nos permitirán tener una mayor información sobre los datos y nos

facilitarán los cálculos de las medidas descriptivas o estadísticas de la muestra. Estas son la frecuencia relativa (h_i), la frecuencia acumulada (F_i y H_i).

La organización de los datos para el caso en que la variable estadística usada tenga muchos valores implica el arreglo de las observaciones en intervalos de clases. El proceso para hallar los intervalos de clase es el siguiente:

Debemos hallar, en primer lugar, en la muestra, el menor valor observado y el mayor valor observado.

El número de intervalos no deberá ser tan pequeño (menor que 5) o tan grande (mayor de 15) que la verdadera naturaleza de la distribución sea imposible de visualizar. La longitud del intervalo de clase deberá ser siempre la misma.

Si la longitud de cada intervalo no fuera exacta, se puede tomar por exceso asegurándonos de este modo que la reunión de todos los intervalos cubrirá a todos los valores observados.



Para construir los intervalos se usa los intervalos cerrados a la izquierda y abiertos a la derecha: $[L_i, L_{Si}]$, donde L_i es el límite inferior del intervalo y L_{Si} es su límite superior.

¿Cómo decidimos cuántos intervalos de clase tomar?



Existen varias reglas que se basan en el tamaño de nuestra población o muestra. Una de las reglas más usadas es la **Regla de Sturges**, regla empírica que funciona bastante bien para grupos de 30 a 300 observaciones.

Esta regla nos dice que el número de intervalos de clase para una muestra de tamaño n será k si este resulta un número entero o el siguiente número entero a k , si k resulta un número decimal.



La ecuación para hallar k es: $k = 1 + 3.3 * \log n$, donde n es el tamaño de la muestra.

La **marca de clase (xi)**, definida como el punto medio del intervalo de clase, deberá tener de preferencia el mismo número de decimales que los valores observados. **La marca de clase puede considerarse que es un representante de los datos que caen en el intervalo.**

$$Xi = \frac{LH + LS}{2}$$

Ejemplo 1. Suponga que los datos que se presentan a continuación corresponden a los valores de la inflación anual durante el año 2008 de un total de 20 ciudades de una región del país. Construir la distribución de frecuencias

8.2 12.8 10.5 9.3 12.7 10.2 9.1 10.7 8.2 12.8 8.5 11.6 8.4 10.1
10.2 13.1 9.8 12.1 13.6 11.7

Solución

- $R = 13.6 - 8.2 = 5.4$
- $K = 1 + 3.3 \log 20 = 1 + 3.3 (1.3010) = 1 + 4.29 = 5.29 = 5$ (redondeo por aproximación)
- $C = R/k = 5.4 / 5 = 1.08 = 1.1$ (redondeo por exceso; los datos tienen un decimal)
- Los límites de los intervalos se obtienen del siguiente modo:

$LI_1 = 8.2$	$LS_1 = LI_2 = 9.3$
$LI_2 = LI_1 + c = 8.2 + 1.1 = 9.3$	$LS_2 = LI_3 = 10.4$
$LI_3 = LI_2 + c = 9.3 + 1.1 = 10.4$	$LS_3 = LI_4 = 11.5$
$LI_4 = LI_3 + c = 10.4 + 1.1 = 11.5$	$LS_4 = LI_5 = 12.6$
$LI_5 = LI_4 + c = 11.5 + 1.1 = 12.6$	$LS_5 = LS_4 + c = 12.6 + 1.1 = 13.7$
- Las marcas de clase se calculan de la siguiente manera:
$$X1 = \frac{8.2+9.3}{2} = 8.75; X2 = \frac{9.3+10.4}{2} = 9.85 \text{ y así sucesivamente}$$
- Para determinar las frecuencias absolutas se procede como sigue: Se toma la primera observación 8.2 y se busca el intervalo de clase que pertenece, es el 8.2 - 9.3, luego se asigna una tarja en la intersección de la columna de conteo y la fila de ese intervalo. Se toma ahora la otra observación 12.8, la cual pertenece al intervalo 12.6 - 13.7, entonces se asigna una tarja en la intersección de la fila de este nuevo intervalo y la columna de conteo. Así sucesivamente hasta agotar la última observación. Sumando las tarjas se obtiene la frecuencia absoluta de cada clase.
- Para obtener las frecuencias acumuladas se procede de la siguiente forma:
$$F1 = f1 = 5 \quad F2 = F1 + f2 = 5 + 5 = 10$$

Con los resultados anteriores se obtiene el siguiente cuadro de distribución de frecuencias

Intervalos de clase	Marca de clase X_i	Tarjas	Frecuencia absoluta f_i	Frecuencia acumulada aF_i	Frecuencia relativa h_i
8.2 ; 9.3	8.75	////	5	5	5/20
9.3 ; 10.4	9.85	////	5	10	5/20
10.4 ; 11.5	10.95	//	2	12	2/20
11.5 ; 12.6	12.05	///	3	15	3/20
12.6 ; 13.7	13.15	////	5	20	5/20
			20		1



Ejemplo 2. A continuación, se presenta una lista ya ordenada de las observaciones hechas sobre el ingreso de las personas.

53	57	58	61	61	63	64	66	67	68
69	70	71	72	73	74	74	74	74	77
77	77	78	78	79	79	79	81	81	81
82	82	83	83	84	85	85	86	87	87
88	90	90	90	90	92	93	94	96	97



Para estos ingresos, el menor valor de la muestra es 53 dólares y el mayor valor de la muestra es 97 dólares. Luego, el rango de estos valores es: $97 - 53 = 44$ dólares

Al aplicar la regla de Sturges con $n = 50$, tendremos: $k = 1 + 3.3 \cdot (1.69897) = 6.6$, lo que equivale a tomar 7 intervalos s)

El tamaño o amplitud de cada intervalo de clase se determina así: $c=R /K= 44 / 7$
 $= 6.29 = 7$. (redondeo por exceso, al entero superior, considerando que los datos son
 entero. Si los datos tienen decimales el proceso es el mismo).

INTERVALOS	x_i	f_i	h_i	F_i
[53 ; 60]	56.5	3	3/ 50	3
[60 ; 67]	63.5	5	5/ 50	8
[67 ; 74]	70.5	7	7/50	15
[74 ; 81]	77.5	12	12 / 50	27
[81 ; 88]	84.5	13	13/50	40
[88 ; 95]	91.5	10	10 / 50	50
TOTAL		50	1	

Siguiendo el mismo proceso utilizado para el ejemplo 1, se tiene la tabla de
 distribución de frecuencias:



Gráficas

TEMA 4

Estadísticas



Competencia:

Explicar la estructura de cada uno de los gráficos usados en la estadística descriptiva.





Tema 04: Gráficas Estadísticas

Presentación De Los Datos Obtenidos De Una Muestra

Una vez realizada la organización de los datos observados, es necesario presentar estos de forma gráfica forma visual permitirá resaltar algunos hechos que muestran los datos. Se verán diversos tipos de gráficos catalogados según el tipo de variable a presentarse.



Para Una Variable Cualitativa

Para una variable cualitativa, ya sea nominal u ordinal, la presentación de la información obtenida organizada en una distribución de frecuencias puede ser presentada mediante dos gráficos: gráfico de barras y gráfico de sectores.

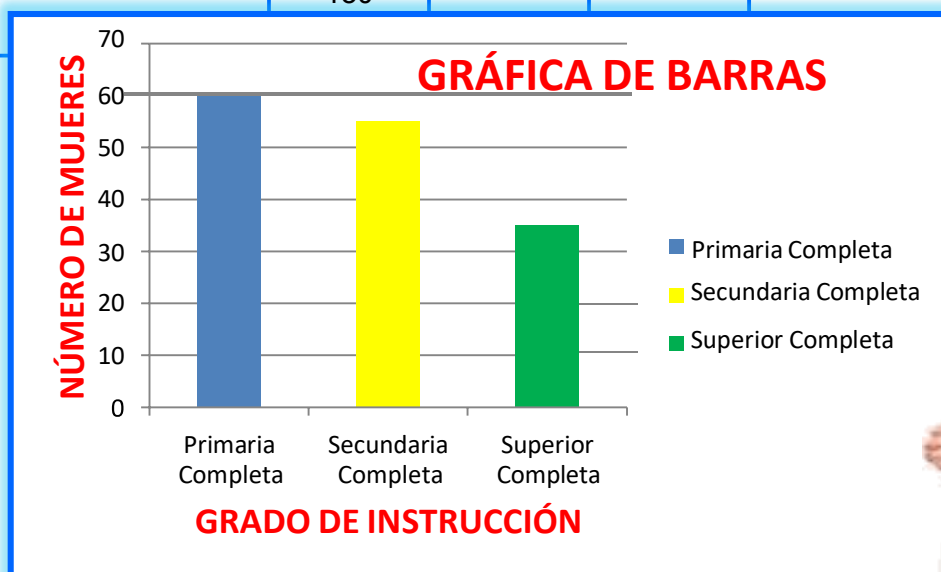
a) Gráfica De Barras

En este tipo de presentación, cada barra rectangular corresponde a una modalidad. Todas las barras tienen base de igual longitud y altura proporcional a la frecuencia (f_i) o frecuencia relativa (h_i) que presen modalidad. Tomemos la distribución de frecuencias de la variable grado de instrucción, obtenida de una muestra de 150 mujeres. Se considerará que cada mujer pertenece al mayor grado de instrucción que ha concluido.

MODALIDAD	f_i	h_i	P_i
Primaria Completa	60	0,40	40%
Secundaria Completa	55	0,37	37%

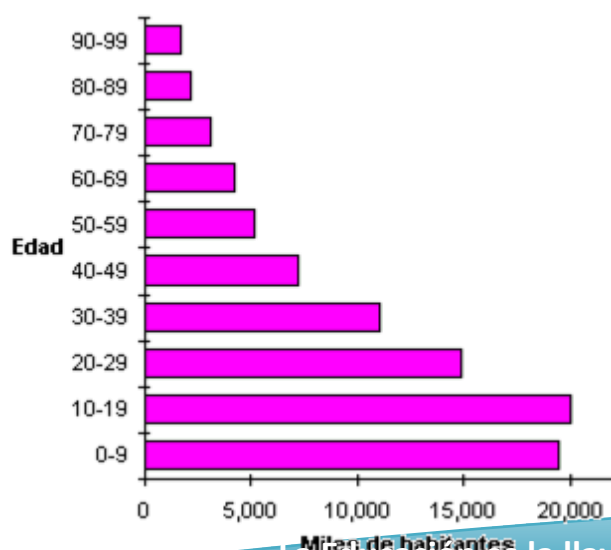


Superior Completa	35	0,23	23%
	150		



También es posible realizar **GRÁFICAS DE BARRAS HORIZONTALES**, los cuales se parecen mucho a las gráficas de columnas, con la salvedad importante de que la función de los ejes se intercambia y el eje horizontal queda destinado a las frecuencias y el eje vertical a las clases.

Es muy común que este tipo de gráficos se utilicen para ilustrar el tamaño de una población dividida en estratos como, por ejemplo, son sus edades. El ejemplo que se presenta es la población de un país ficticio llamado "Perulandia":



A este tipo de gráficos en particular se le llama **pirámide de edades** por su forma. Incluso, cuando se compara la población masculina y femenina por estratos de edades, se estila utilizar el lado izquierdo para la población de un sexo y el lado derecho para el otro, el resultado es una "pirámide" casi simétrica (dependerá de la población en particular).

b) Gráfica De Sectores

Otra forma de presentar la información de una variable cualitativa es utilizando una gráfica de sectores (también denominada gráfica tipo "pie" o "pastel").

La gráfica de sectores es un círculo dividido en varios sectores proporcionales en tamaño a las **frecuencias relativas (hi)** de las diferentes modalidades. En el caso anterior de la distribución de frecuencias, tendremos:

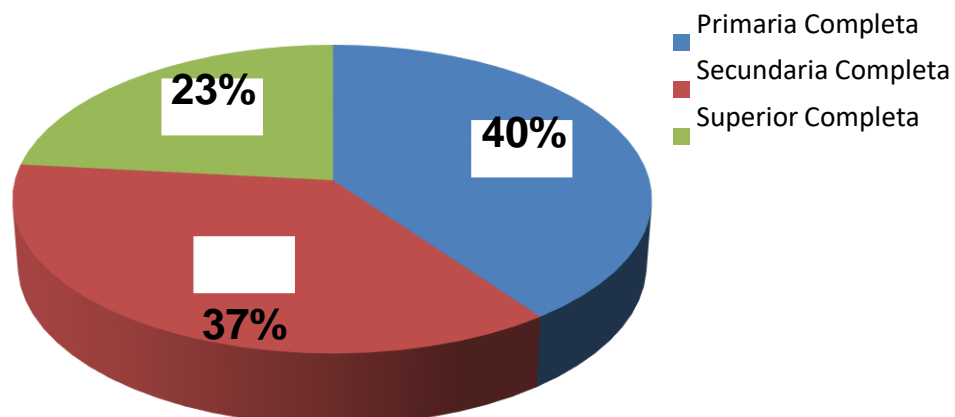
Primaria completa 40% de 360 grados = 144 grados

Secundaria completa 37% de 360 grados = 133.2 grados

Superior completa 23% de 360 grados = 82.8 grados



MODALIDAD	fi	hi	Pi
Primaria Completa	60	0,40	40%
Secundaria Completa	55	0,37	37%
Superior Completa	35	0,23	23%
	150		



c) Gráfico De Bastones

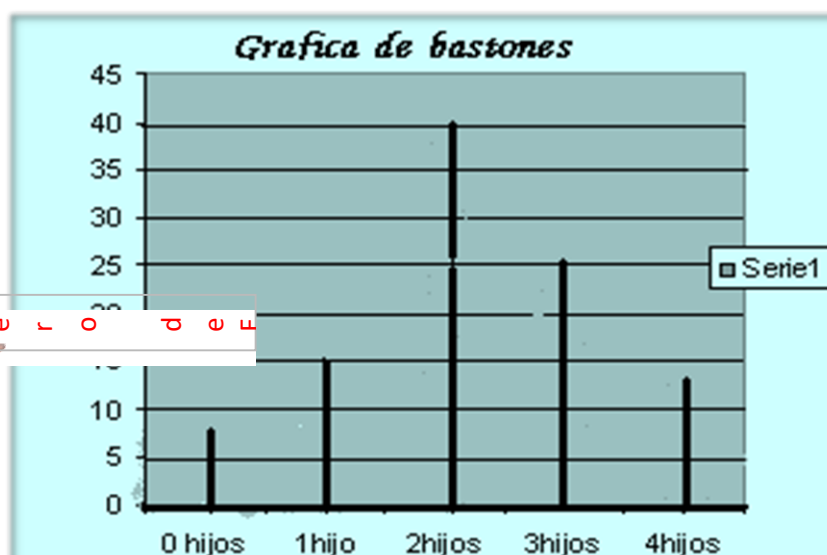
Para una variable cuantitativa discreta (con pocos valores) es posible usar los gráficos de barras. Pero existe otro gráfico, diseñado para este tipo de variables y es la **GRÁFICA DE BASTONES**. En esta gráfica, la frecuencia del valor de la variable es representada por un segmento de recta en vez de una barra.

Tomemos la variable número de hijos, los posibles valores de esta variable son 0 hijos, 1 hijo, 2 hijos, 3 hijos ó 4 hijos.

Luego, para una muestra de 100 datos tendremos la distribución de frecuencias:



VALOR	fi	Hi
0 Hijos	7	$7 / 100 = 0.07$
1 Hijo	15	$15 / 100 = 0.15$
2 Hijos	40	$40 / 100 = 0.40$
3 Hijos	25	$25 / 100 = 0.25$
4 Hijos	13	$13 / 100 = 0.13$
TOTAL	100	1.00



d) Histograma

Para Una Variable Cuantitativa Discreta (Con Muchos Valores) O Continua existe una gráfica equivalente a la gráfica de barras, se denomina **histograma**.

Esta forma de presentación también consiste en graficar barras, pero, a diferencia de la gráfica de barras, aquí

las barras están pegadas unas a otras.

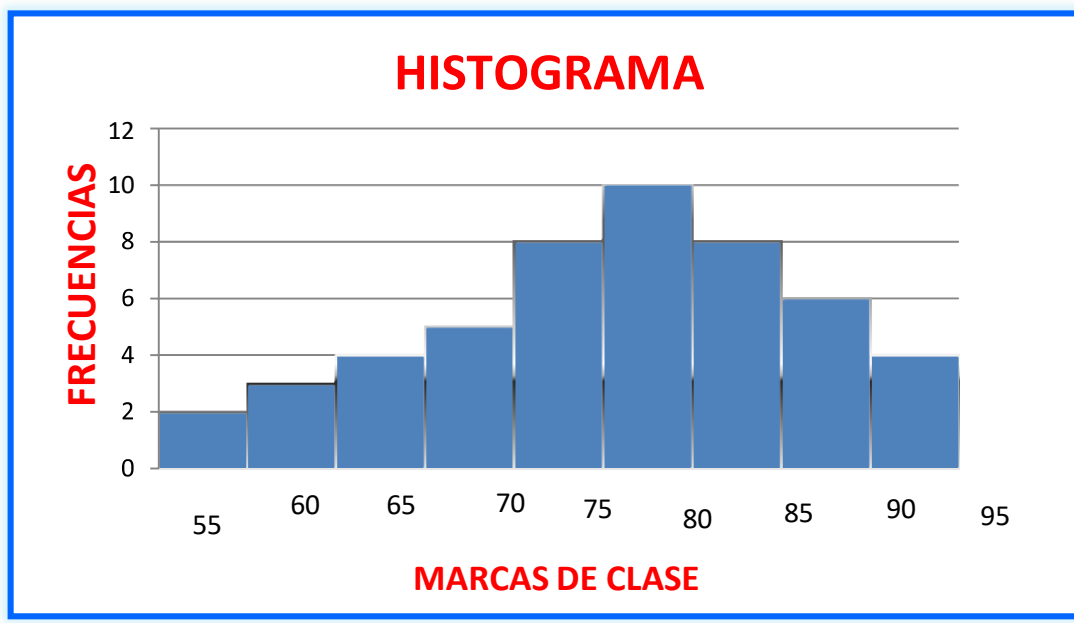


Cada barra corresponde a un intervalo de clase y se acostumbra a colocar el valor inicial y final de cada intervalo o la marca de clase para identificar cada barra. La altura de cada barra puede ser proporcional a la frecuencia (f_i) o la frecuencia relativa (h_i) del intervalo.

Intervalos	X_i	f_i	F_i	h_i	P_i
[52,5 ; 57,5]	55	2	2	0,04	4%
[57,5 ; 62,5]	60	3	5	0,06	6%
[62,5 ; 67,5]	65	4	9	0,08	8%
[67,5 ; 72,5]	70	5	14	0,10	10%
[72,5 ; 77,5]	75	8	22	0,16	16%
[77,5 ; 82,5]	80	10	32	0,20	20%
[82,5 ; 87,5]	85	8	40	0,16	16%
[87,5 ; 92,5]	90	6	46	0,12	12%
[92,5 ; 97,5]	95	4	50	0,08	8%



Total		50	1,00	100%
-------	--	----	------	------



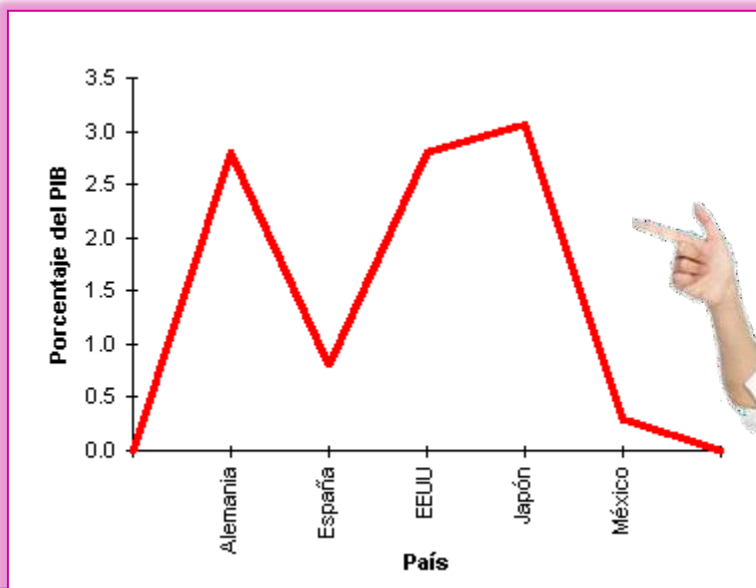
e) Polígono De Frecuencias

Uniendo los puntos medios de los lados superiores de cada barra rectangular del histograma se obtiene un gráfico llamado polígono de frecuencias. El conocimiento del polígono de frecuencias ayudará más adelante en la búsqueda del modelo teórico que mejor describa a los elementos de la población de acuerdo con la variable que se estudia.

Polígono De Frecuencias (Línea Negra)



Otro Ejemplo De Polígono De Frecuencias (Línea Negra)



f) Ojiva

Una gráfica similar al polígono de frecuencias es la **ojiva**, pero ésta se obtiene de aplicar parcialmente la misma técnica a una distribución acumulativa y de igual manera que éstas, existen las **ojivas mayores que** y las **ojivas menores que**.

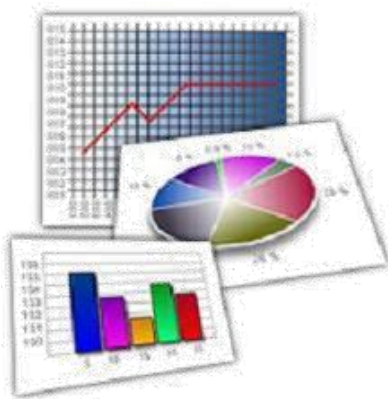


Existen dos diferencias fundamentales entre las ojivas y los polígonos de frecuencias (y por esto la aplicación de la técnica es parcial):

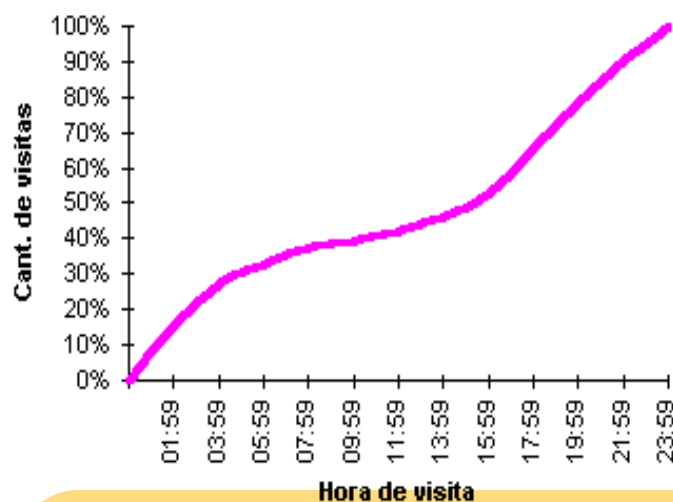
Un extremo de la ojiva no se "amarra" al eje horizontal, para la ojiva mayor que sucede con el extremo izquierdo; para la ojiva menor que, con el derecho. En el eje horizontal en lugar de colocar las marcas de clase se colocan las fronteras de clase.

Para el caso de la **ojiva mayor que**, es la frontera menor; para la **ojiva menor que**, la frontera mayor. La ojiva mayor que (izquierda) se le denomina de esta manera porque viendo el punto que está sobre la frontera de clase "4:00" se ven las visitas

que se realizaron en una hora mayor que las 4:00 horas (en cuestiones temporales se diría después de las 4:00 horas).
De forma análoga, en la ojiva menor que, la frecuencia que se representa en cada frontera de clase son el número de observaciones **menores que** la frontera señalada (en caso de tiempos sería el número de observaciones antes de la hora que señala la frontera).



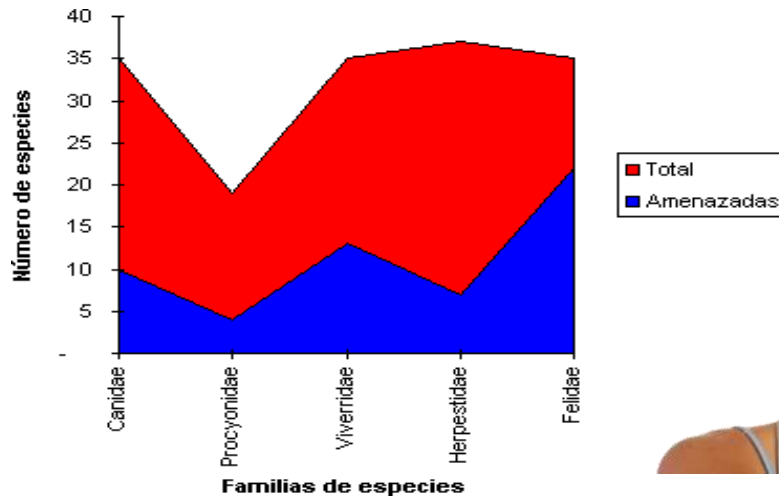
Si se utiliza una distribución porcentual acumulativa entonces se obtiene una ojiva (mayor que o menor que según sea el caso) cuyo eje vertical tiene una escala que va del 0% al 100%. El siguiente ejemplo es la misma **ojiva menor que** se acaba de usar, pero con una distribución porcentual:



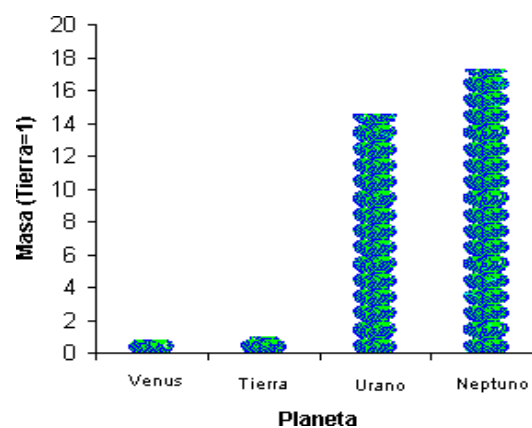
g) Gráfica De Áreas



En ocasiones, al comparar dos series de observaciones (o de datos) se utiliza una llamada **gráfica de áreas**, la cual consiste en rellenar el área que se encuentre debajo de las líneas que resultan de una gráfica de líneas. El ejemplo que se presenta es la comparación del total de las especies de las familias del orden Carnívora y las que están amenazadas, en México, (fuente: Revista "Ciencia y Desarrollo").



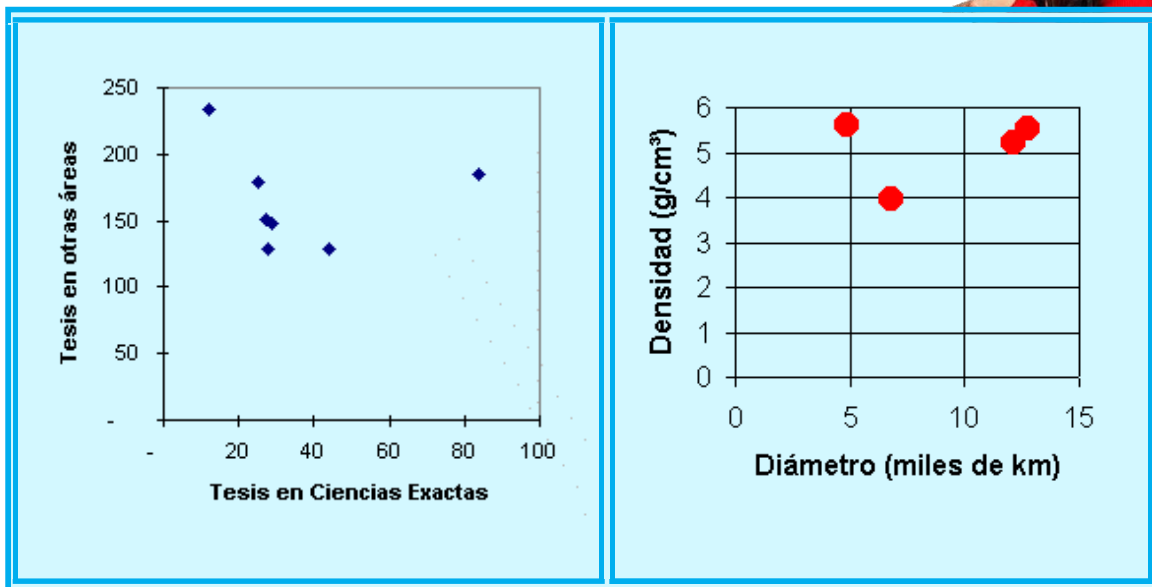
Actualmente, y mucho en los medios masivos de comunicación, se utilizan gráficos para ilustrar los datos o los resultados de alguna investigación. Regularmente se utilizan dibujos para representar dicha información, y el tamaño o el número de estos dibujos dentro de una gráfica queda determinado por la frecuencia correspondiente. A este tipo de gráfica se le llama **pictograma** y éstos son dos ejemplos:



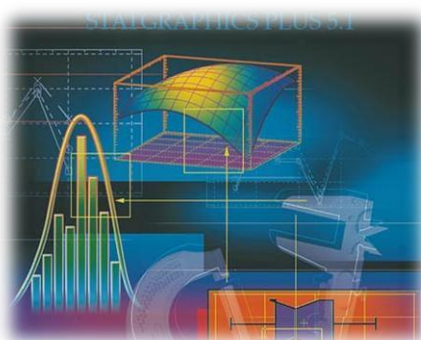
h) Gráfica De Dispersión

Cuando se pretende ilustrar la dispersión de las observaciones realizadas, y así trabajar algunas cosas como correlaciones se puede utilizar una **gráfica de dispersión**.

Por ejemplo, el ejemplo de la izquierda es la dispersión que se presenta al comparar el número de tesis doctorales en ciencias exactas contra el número de total de tesis doctorales (todo en México) en observaciones anuales entre 1984 y 1990 (fuente: Revista "Ciencia y Desarrollo", 1994, XIX (114):12):



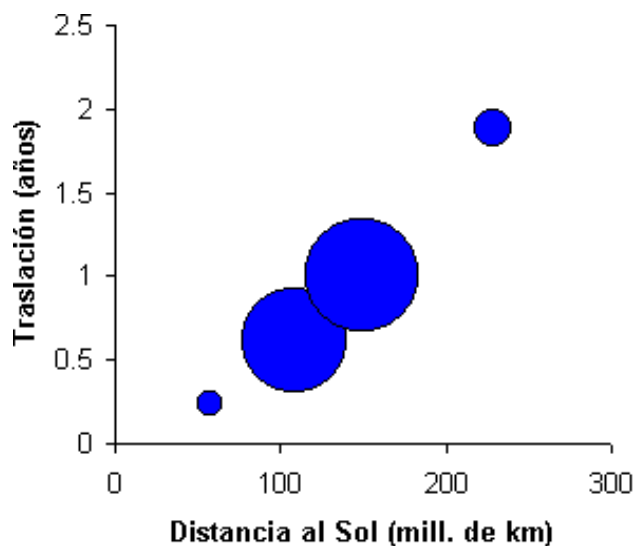
La gráfica de la derecha es resultado de comparar el diámetro (en miles de kilómetros) de los planetas interiores de nuestro sistema solar contra sus densidades (en gramos por centímetro cúbico). Es interesante observar que los puntos parecen "seguir" una línea imaginaria que se asemeja a una recta, con excepción de un caso atípico: Mercurio.



Uno de los usos de este tipo de gráficas es precisamente encontrar si las observaciones siguen algún patrón lineal (una línea de tendencia) o si existen valores atípicos. Para el caso del *Excel*, el programa es capaz de graficar las líneas de tendencias que siguen un conjunto de datos.

i) Gráfica De Burbujas

Un tipo de gráfico similar a las gráficas de dispersión son las **gráficas de burbujas**, en las cuales se presenta la dispersión de las observaciones de la misma forma que aquéllas, pero se le añade la posibilidad de visualizar otra variable representada en el tamaño del punto, pues éstos se convierten en círculos (burbujas) con radios proporcionales a las magnitudes que representan.



Este ejemplo compara la distancia que existe entre cada uno de los planetas interiores de nuestro sistema solar con respecto al Sol, contra el tiempo que necesitan para recorrer sus órbitas, y el tamaño de las burbujas que indica la masa de cada planeta. Además existen otros tipos de gráficos, cada uno con características particulares que les proporcionan cierta intencionalidad para su uso, como son las **gráficas de radar** y las **gráficas polares**.

Lecturas Recomendadas

- ❖ INTRODUCCIÓN, CONCEPTO, ETAPAS DEL DESARROLLO DE LA ESTADÍSTICA.

<http://www.monografias.com/trabajos15/estadistica/estadistica.shtml>

- ❖ ORGANIZACIÓN DE DATOS Y DISTRIBUCIÓN DE FRECUENCIAS

http://www.vitutor.net/2/11/distribucion_frecuencias.html

Actividades y Ejercicios

Ingresa al link *presentación de datos*, lee atentamente las indicaciones, desarrolla los ejercicios y envíalo por el mismo medio.

1. De los siguientes valores:

Ford	Toyota	Nissan	Hyundai	Hyundai	Ford
Nissan	Ford	Hyundai	Nissan	Hyundai	Toyota
Hyundai	Nissan	Toyota	Ford	Toyota	Hyundai
Ford	Hyundai				

- Hallar la frecuencia absoluta y relacional.
- Hallar la frecuencia acumulada absoluta y relacional.
- Realizar un grafico de barras.
- Dibujar un diagrama circular.

2. Suponga que en estudio socioeconómico se observó, entre otras variables, el número de trabajadores eventuales que tienen las empresas comerciales de una región de la ciudad de Trujillo. Mediante una muestra de 30 empresas se encontraron los siguientes resultados.

4	10	5	8	10	6	10	7	8	6	9	7	9	6	8
8	6	7	10	8	7	8	9	7	5	9	4	7	8	6

Construir el cuadro de distribución de frecuencias.

3. Suponga que se ha llevado a cabo una encuesta a 28 personas elegidas al azar para analizar su opinión sobre la calidad de una nueva conserva que se desea introducir en el mercado. Los resultados observados fueron los siguientes:

Bueno	Malo	Bueno	Excelente	Regular	Bueno	Regular
Regular	Regular	Excelente	Excelente	Bueno	Excelente	Bueno
Bueno	Excelente	Bueno	Malo	Bueno	Bueno	Malo
Bueno	Excelente	Bueno	Bueno	Excelente	Bueno	Excelente

Construir el cuadro de distribución de frecuencias! éxito... SsP!



Autoevaluación

1. Clasifique las variables referidas a la población de electores del Perú.

- Preferencia electoral.
- Edad del elector.
- Estado socio económico del elector.
- Número de integrantes en la familia del elector.
- Sexo del elector.
- Grado de instrucción del elector.
- Ingresos mensuales del elector.

- a) 4 Cualitativas y 3 Cuantitativas
- b) 3 Cualitativas y 4 Cuantitativas
- c) 2 Cualitativas y 5 Cuantitativas
- d) 5 Cualitativas y 2 Cuantitativas
- e) 1 Cualitativas y 6 Cuantitativas

2. El objetivo principal de la estadística descriptiva es:

- a) Describir una población.
- b) Hallar las regularidades que se encuentran en los fenómenos de masa.
- c) Inferir algo acerca de la población.
- d) Calcular un promedio.
- e) Hallar el promedio de acuerdo a la cantidad.

3. ¿Cuáles de las siguientes afirmaciones es cierto respecto a una muestra?

- a) Es parte de una población.
- b) Debe contener al menos cinco observaciones.
- c) Se refiere a estadística descriptiva.
- d) Se refiere a una variable no contable.
- e) Contiene dentro a la población.



4. Una variable cualitativa.

- a) Siempre se refiere a una cualidad.
- b) Es no numérica.
- c) Siempre tiene sólo dos resultados posibles.
- d) Todas las anteriores son correctas.
- e) Es numérica.

5. Una variable en escala nominal.

- a) Casi siempre es el resultado de contar algo.
- b) Tiene un punto cero significativo.
- c) Puede adquirir valores negativos.
- d) No puede tener más de dos categorías.
- e) Sólo sirve para nombrar su característica

6. En una empresa, se hizo el estudio sobre las edades de los empleados y se obtuvo la siguiente tabla:

EDADES	Nº DE EMPLEADOS
[20 - 25]	12
[25 - 30]	15
[30 - 35]	23
[35 - 40]	11
[40 - 45]	9

Total: 70

Donde **A** es el porcentaje de empleados con 30 años ó más.

B es el porcentaje de empleados entre 40 y 45 años. Señale **A - B** (aprox.)

- a) 65%
- b) 60%
- c) 63%
- d) 64%
- e) 62%

7. La tabla muestra la distribución del ingreso familiar correspondiente a 80



familias.

f_i : frecuencia absoluta simple

F_i : frecuencia absoluta acumulada

h_i : frecuencia relativa simple en tanto por uno

Intervalo de Ingreso S/.	f_i	F_i	h_i
[160 - 170]	12	12	
[170 - 180]	48	60	
[180 - 190]			0,125
[190 - 200]			0,075
[200 - 210]			

Determine el número de familias que ganan menos de 200 nuevos soles.

- a) 66
- b) 70
- c) 54
- d) 76
- e) 50

8. En una prueba de aptitud académica se evaluaron a “n” estudiantes y las notas obtenidas se clasificaron en una tabla de distribución de frecuencias como se muestra a continuación.

Marca de Clase	Frecuencias Relativas
45	$K/50$
55	$3k/100$
65	$2k/25$
75	$3k/50$
85	$K/100$

¿Qué porcentaje de estudiantes obtuvieron una nota menor que 65 puntos o igual que 85 puntos?

- a) 30%
- b) 40%

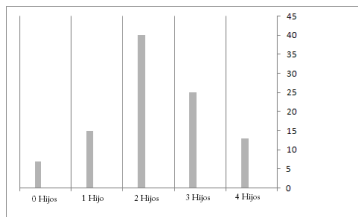
c) 50%

d) 60%

e) 70%

9. ¿Cuál de los siguientes diagramas es un histograma?

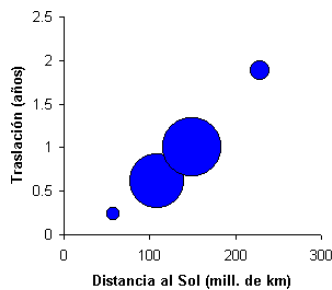
a)



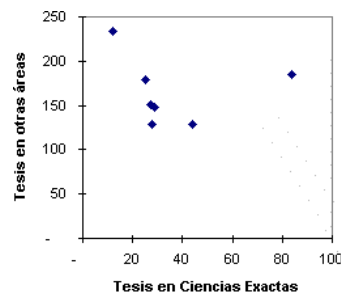
b)



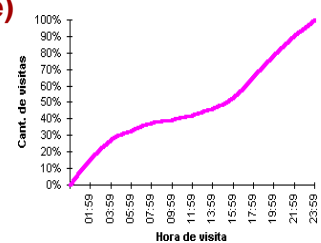
c)



d)

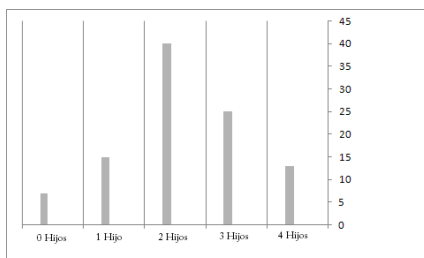


e)



10. ¿Cuál de los siguientes diagramas es una ojiva?

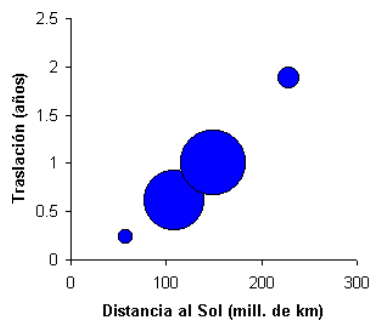
a)



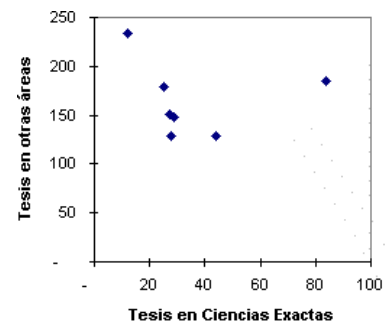
b)



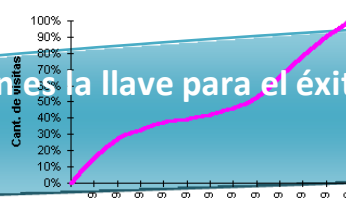
c)



d)



e)



La Educación es la llave para el éxito... SsP!

MG.
NSHT