

Probabilidad y Estadística

Unidad 1: Introducción a la Estadística

¿Qué es la Estadística?

Es la ciencia que se encarga de la recolección, ordenamiento, representación, análisis e interpretación de un conjunto de datos sobre hechos, individuos o grupos de los mismos, para deducir de ello conclusiones precisas o estimaciones futuras.

La Estadística no son sólo los resultados de encuestas, ni el cálculo de unos porcentajes, sino que se basa en una metodología científica que pretende sacar conclusiones a partir de unas observaciones hechas.



¿Para qué sirve la Estadística?

- Facilitar la toma de decisiones.
 - Sirve como base para realizar evaluación de situaciones.
 - Permite hacer diagnósticos de los problemas económicos y sociales.
 - Ayuda a establecer parámetros de posibles resultados.
 - Calcular la cantidad de afectados en un desastre; medir la densidad de población; calcular las edades y géneros de las personas.
 - Medir los recursos.
 - Medir los efectos de un fármaco.
 - Medir las preferencias de la gente.
 - Medir los mercados dentro de las agencias de publicidad.
- Realizar estudios de mercado.
 - Conocer o calcular las preferencias electorales.
 - Calcular la aceptación de un producto.




Ingeniería en Sistemas de Información



Un administrador de red está interesado en comprobar si el servidor de páginas *web* influye en el tiempo de respuesta a distintas peticiones. Para ello instala tres servidores diferentes en la misma máquina y, desde distintos puestos, encarga a varios usuarios de la *web* que visiten diferentes páginas en un orden totalmente aleatorio. ¿**Afecta el factor servidor *web* al tiempo de respuesta** de la red?

Orígenes de la Estadística

La Estadística actual es el resultado de la unión de dos disciplinas que evolucionaron de forma independiente hasta confluir en el siglo XIX:



La “Estadística”, o ciencia del Estado, que estudia la descripción de datos.

El **Cálculo de Probabilidades**, que nació en el siglo XVII como la teoría matemática de los juegos de azar.

La **palabra Estadística** procede del vocablo “**Estado**”, pues era función principal de los Gobiernos de los Estados establecer registros de población, nacimientos, defunciones, impuestos, cosechas... La necesidad de poseer datos cifrados sobre la población y sus condiciones materiales de existencia han debido hacerse sentir desde que se establecieron sociedades humanas organizadas.

Tipos de Estadística

Estadística Descriptiva

Es la parte de la estadística que permite analizar todo un conjunto de datos, de los cuales se extraen conclusiones verdaderas, únicamente para ese conjunto. Para realizar este análisis se procede a la recolección, organización, representación e interpretación de la información obtenida.



Estadística Inferencial

Es la parte de la estadística que permite realizar generalizaciones hacia una población de estudio a partir de los resultados obtenidos en una muestra, tomada de ella.

Las conclusiones arribadas fundamentan la toma de decisiones para que las mismas sean las acertadas y minimizar así los riesgos de equivocarse

Lenguaje de la Estadística

Población

Conjunto de todos los elementos que cumplen una determinada característica, que deseamos medir o estudiar



Muestra

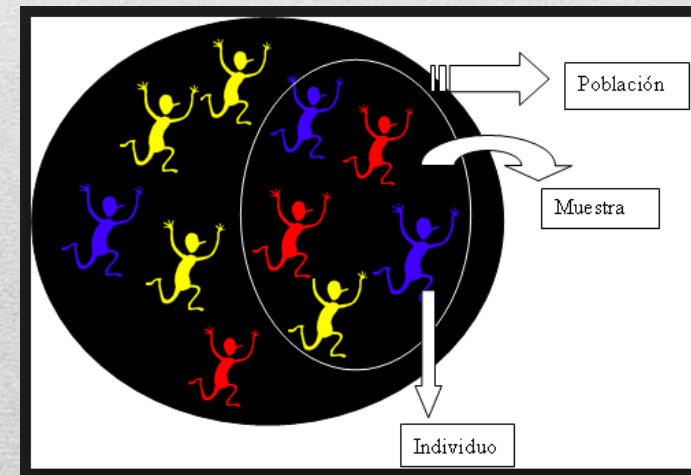
Subconjunto representativo de la población.

Una buena muestra es aquella que refleja las características esenciales de la población de la cual se obtuvo.



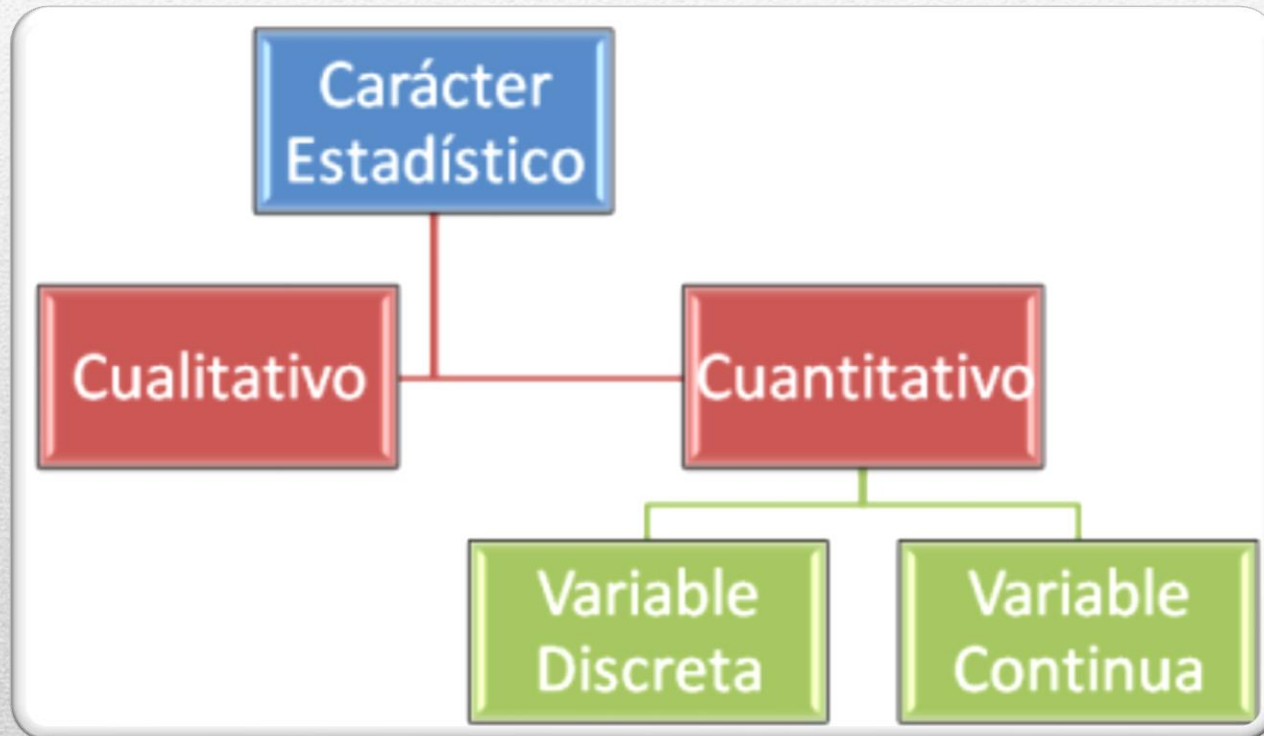
Unidad Estadística

Es cada elemento de la población.



Carácter Estadístico

Cada una de las propiedades o aspectos que pueden estudiarse en los individuos o elementos de una población



Serie o distribución de frecuencias

Es la correspondencia que hay entre cada valor de la variable (que se denomina solamente variable para la variable discreta y clase para variables continuas) y su respectivo número de observaciones o frecuencias.

EDADES	x	f	f _r	F
13-15	14	4	0,2	4
15-17	16	9	0,45	13
17-19	18	3	0,15	16
19-21	20	3	0,15	19
21-23	22	1	0,5	20
		20	1	

PROGRAMAS	N° DE TELEVIDENTES
NOTICIEROS	17
SERIES	15
NOVELAS	18
DIBUJOS	16
CULTURALES	14
TOTAL	80

$\bar{x} = 1,9 / 2,1 / 1,7 / 1,5 / 1,8 / 1,1 / 1,7 / 1,3 / 1,9 / 1,7 / 1,1$

x	f _i	x _i · f _i
0	3	0
1	5	5
2	5	10
3	4	12
4	2	8
5	1	5
	<u>n=20</u>	<u>40</u>

Serie de frecuencia – Carácter estadística cualitativo

Categorías	Frecuencia Absoluta	Frecuencia Relativa	Frecuencia relativa porcentual
Cat. 1	f_1	$f_{ri}=f_1/n$	$f_{r\%}=f_{ri}.100\%$
Cat. 2	f_2	$f_{ri}=f_2/n$	$f_{r\%}=f_{r2}.100\%$
Cat.3	f_3	$f_{ri}=f_3/n$	$f_{r\%}=f_{r3}.100\%$
.....
.....
.....
Total			

Serie de frecuencia – Carácter estadística cuantitativo discreto

Valores de la variable	Frecuencia Absoluta	Frecuencia Relativa	Frecuencia relativa porcentual	Frecuencia Acumulada	Frecuencia acumulada relativa	Frecuencia acumulada relativa
X_1	f_1	$f_{ri}=f_1/n$	$f_{r\%}=f_{r1}.100\%$	$F_1=f_1$	$F_{rk}=F_1/n$	$F_{r\%}=F_{r1}.100\%$
X_2	f_2	$f_{ri}=f_2/n$	$f_{r\%}=f_{r2}.100\%$	$F_2= f_1+ f_2$	$F_{rk}=F_2/n$	$F_{r\%}=F_{r2}.100\%$
X_3	f_3	$f_{ri}=f_3/n$	$f_{r\%}=f_{r3}.100\%$	$F_3= f_1+ f_2+ f_3$	$F_{rk}=f_3/n$	$F_{r\%}=F_{r3}.100\%$
.....
.....
.....
Total						

Serie de frecuencia – Carácter estadística cuantitativo continuo

Intervalos de clases	Frecuencia Absoluta	Frecuencia Relativa	Frecuencia relativa porcentual	Frecuencia Acumulada	Frecuencia acumulada relativa	Frecuencia acumulada relativa
LRI-LRS	f_1	$f_{ri}=f_1/n$	$f_{r\%}=f_{r1}.100\%$	$F_1=f_1$	$F_{rk}=F_1/n$	$F_{r\%}=F_{r1}.100\%$
.....	f_2	$f_{ri}=f_2/n$	$f_{r\%}=f_{r2}.100\%$	$F_2= f_1+ f_2$	$F_{rk}=F_2/n$	$F_{r\%}=F_{r2}.100\%$
.....	f_3	$f_{ri}=f_3/n$	$f_{r\%}=f_{r3}.100\%$	$F_3= f_1+ f_2+ f_3$	$F_{rk}=f_3/n$	$F_{r\%}=F_{r3}.100\%$
.....
.....
.....
Total						

Gráficos Estadísticos

Un gráfico brinda un resumen de la información recopilada y es más fácil de comprender que un cuadro.

La desventaja que presentan es que al realizar el gráfico, muchas veces se pierde la información original y por ende se trabaja con una aproximación de la misma.

Existen diferentes gráficos, los cuales se realizan en relación al tipo de carácter estadístico que es-

Gráfico de Barras

Participación en el mercado de las cuatro empresas principales en las ventas totales de productos alimentarios y bebidas ultraprocesados en América Latina, 2005 y 2013

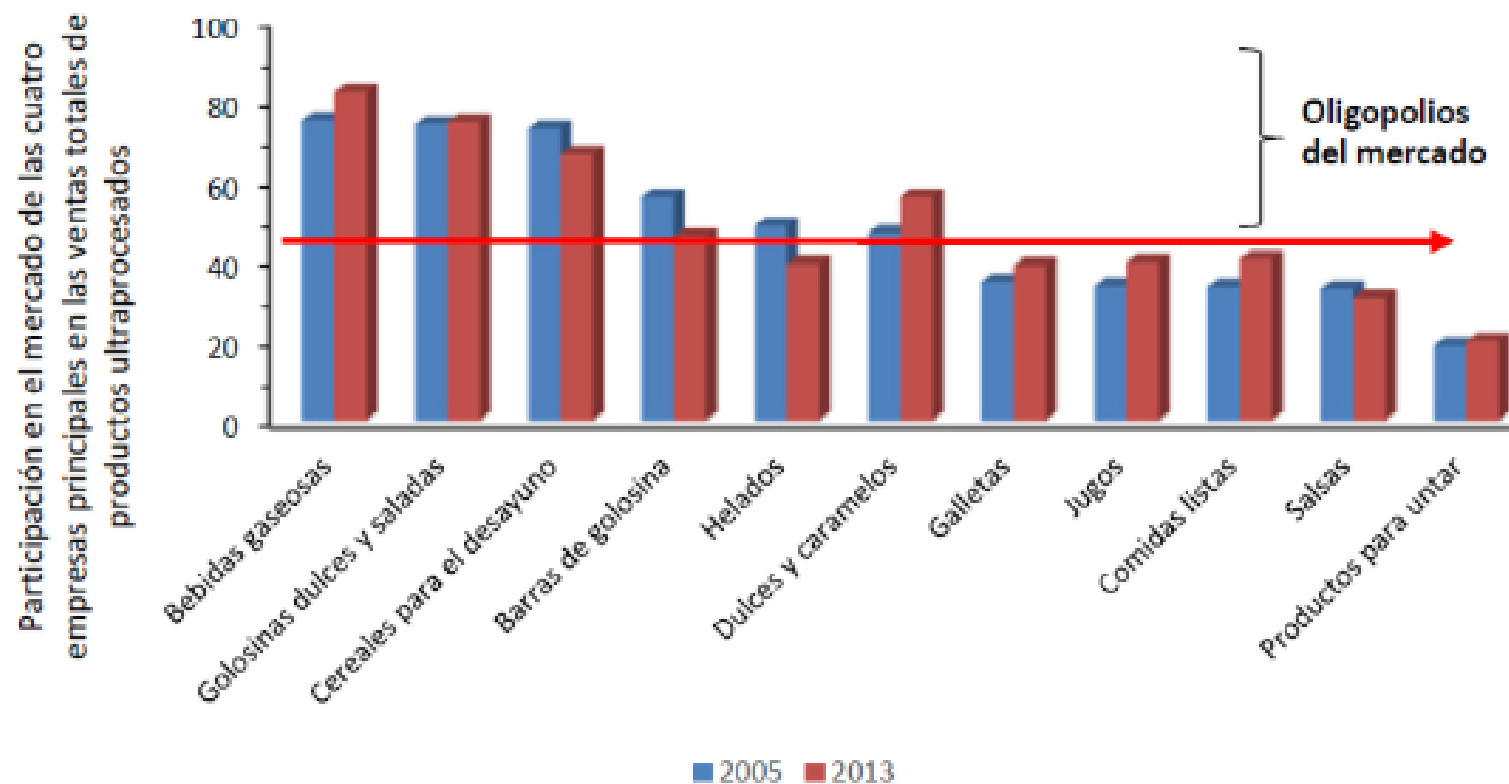
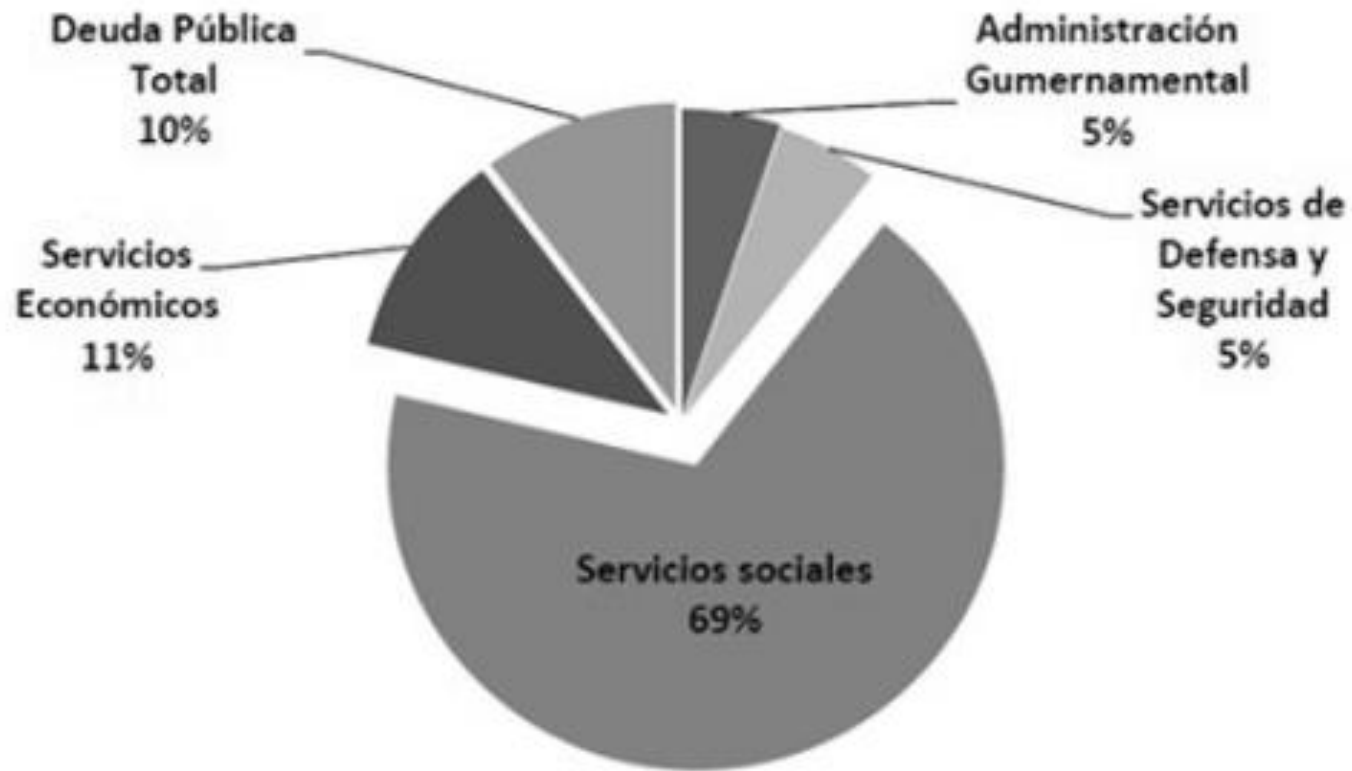


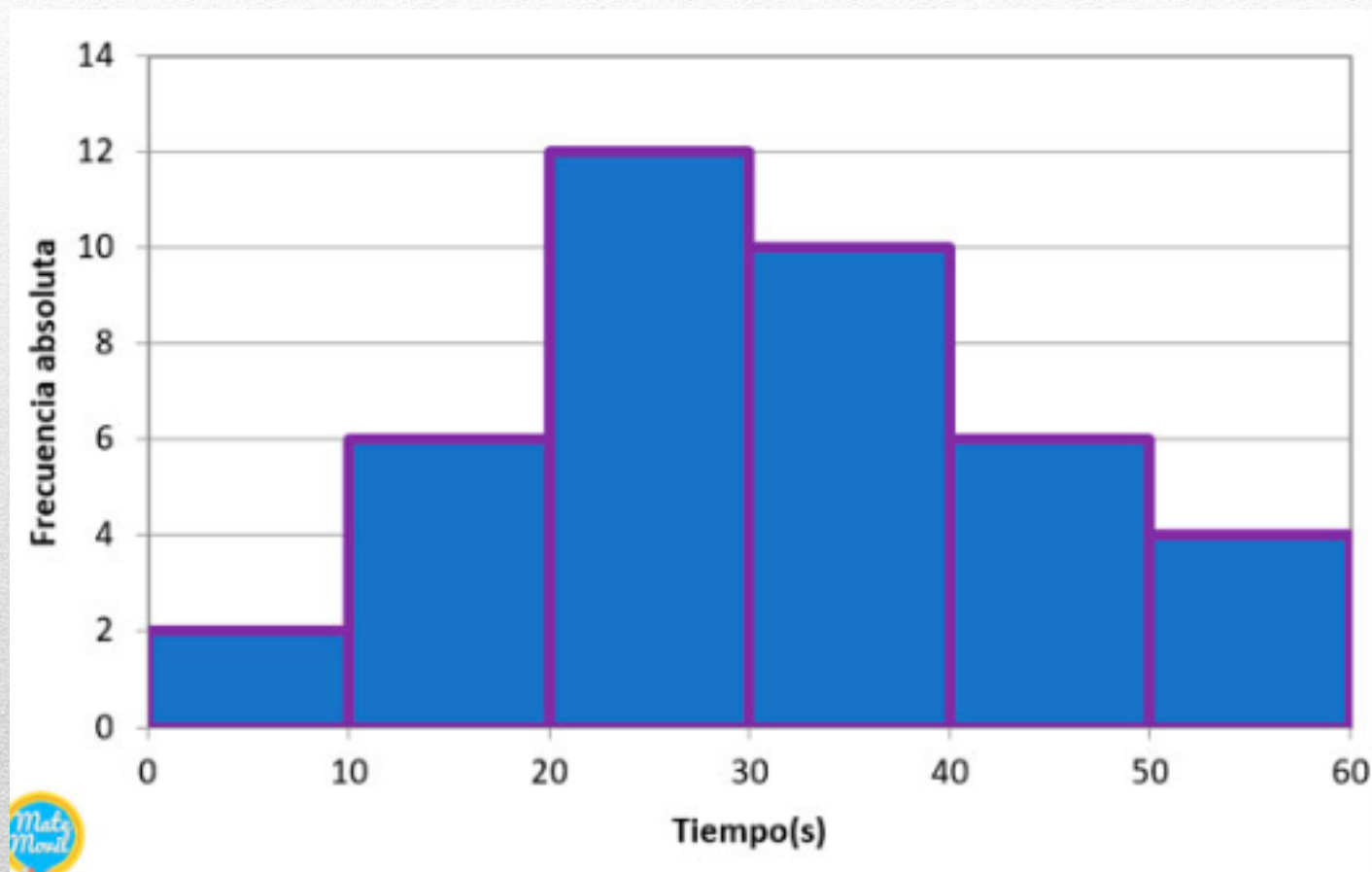
Gráfico Circular

Gráfico N° 1. Distribución del Gasto Público en el Presupuesto 2018



Fuente: *CEPA en base a MECON*

Histograma



Polígono de Frecuencias

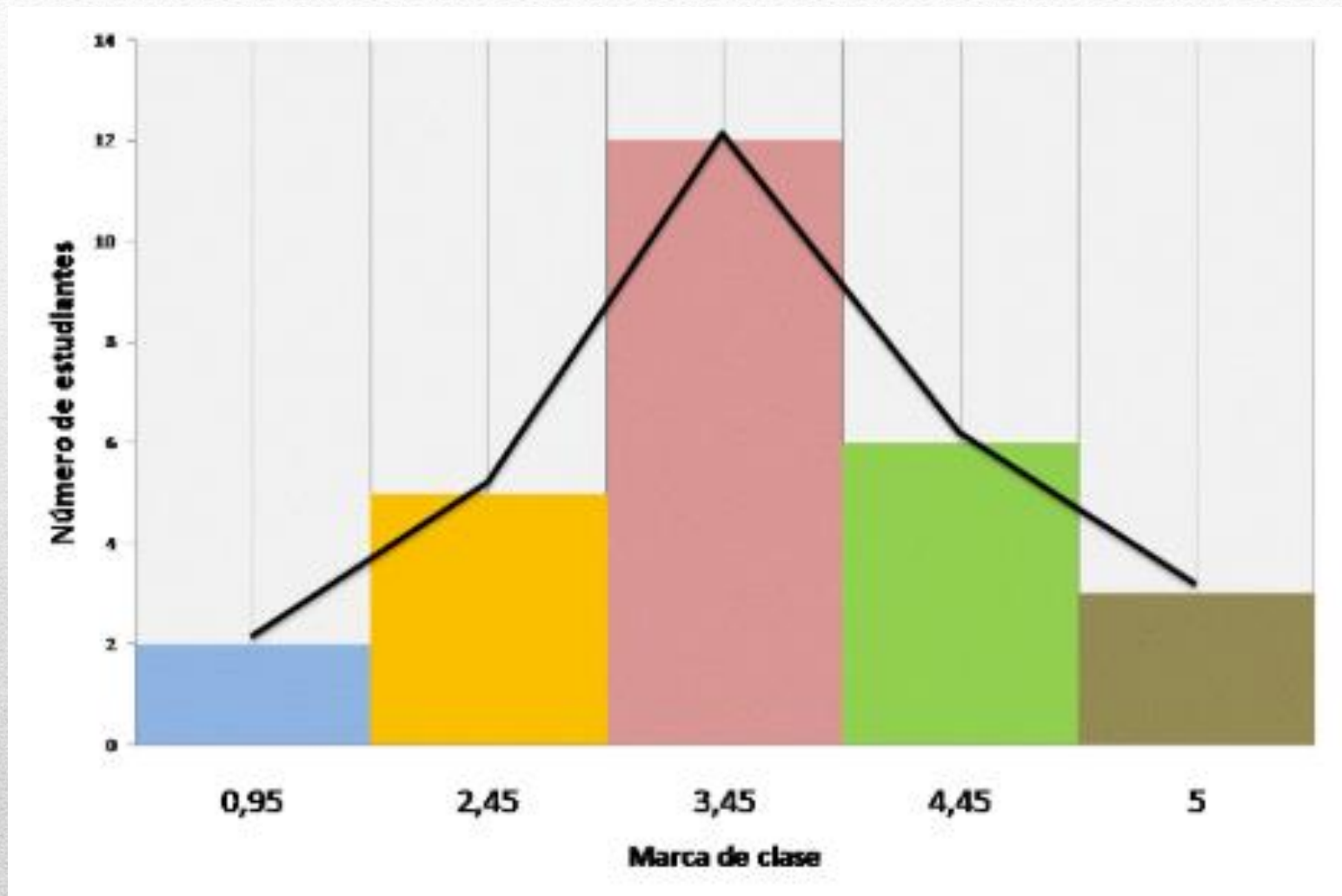
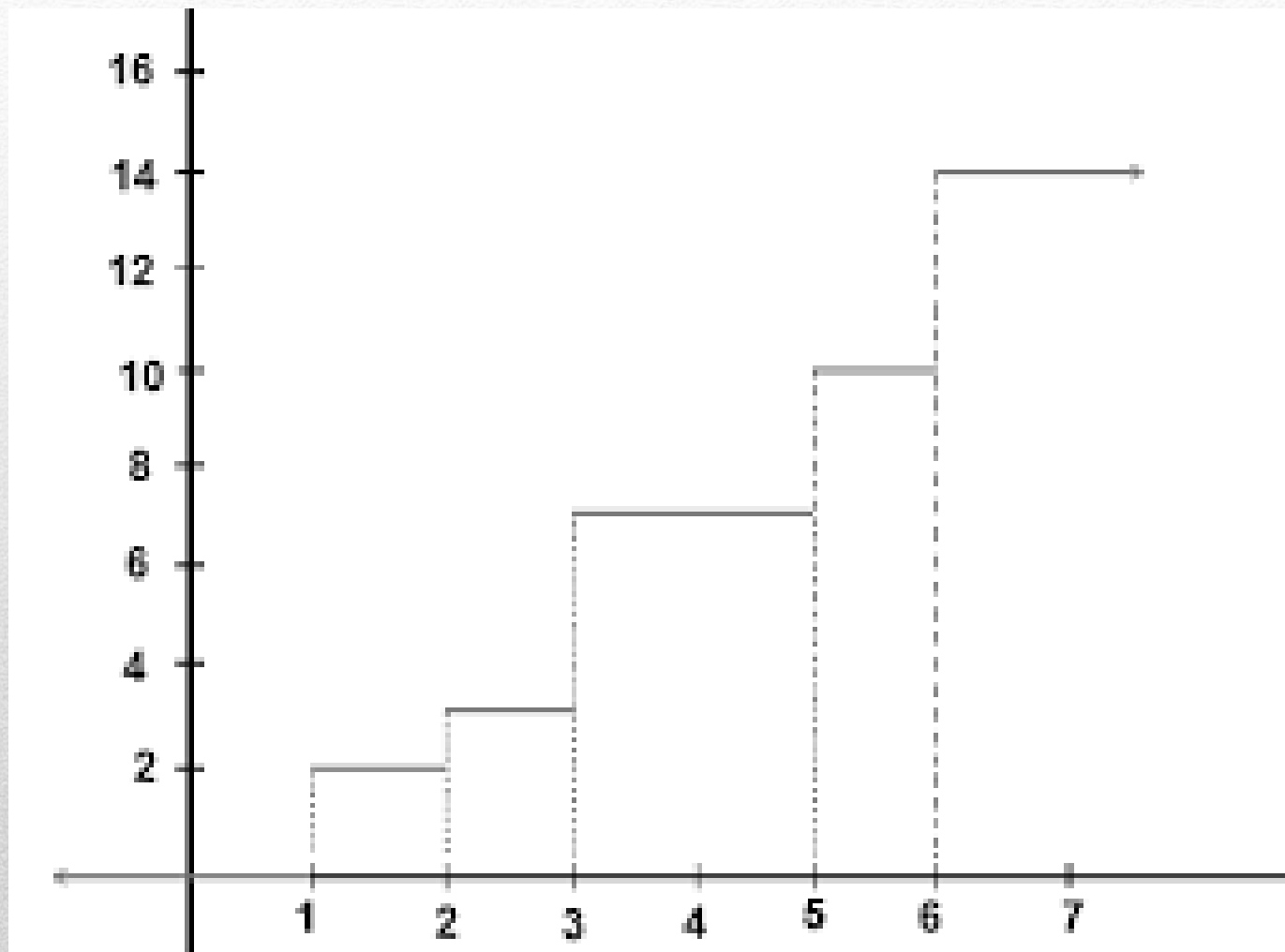
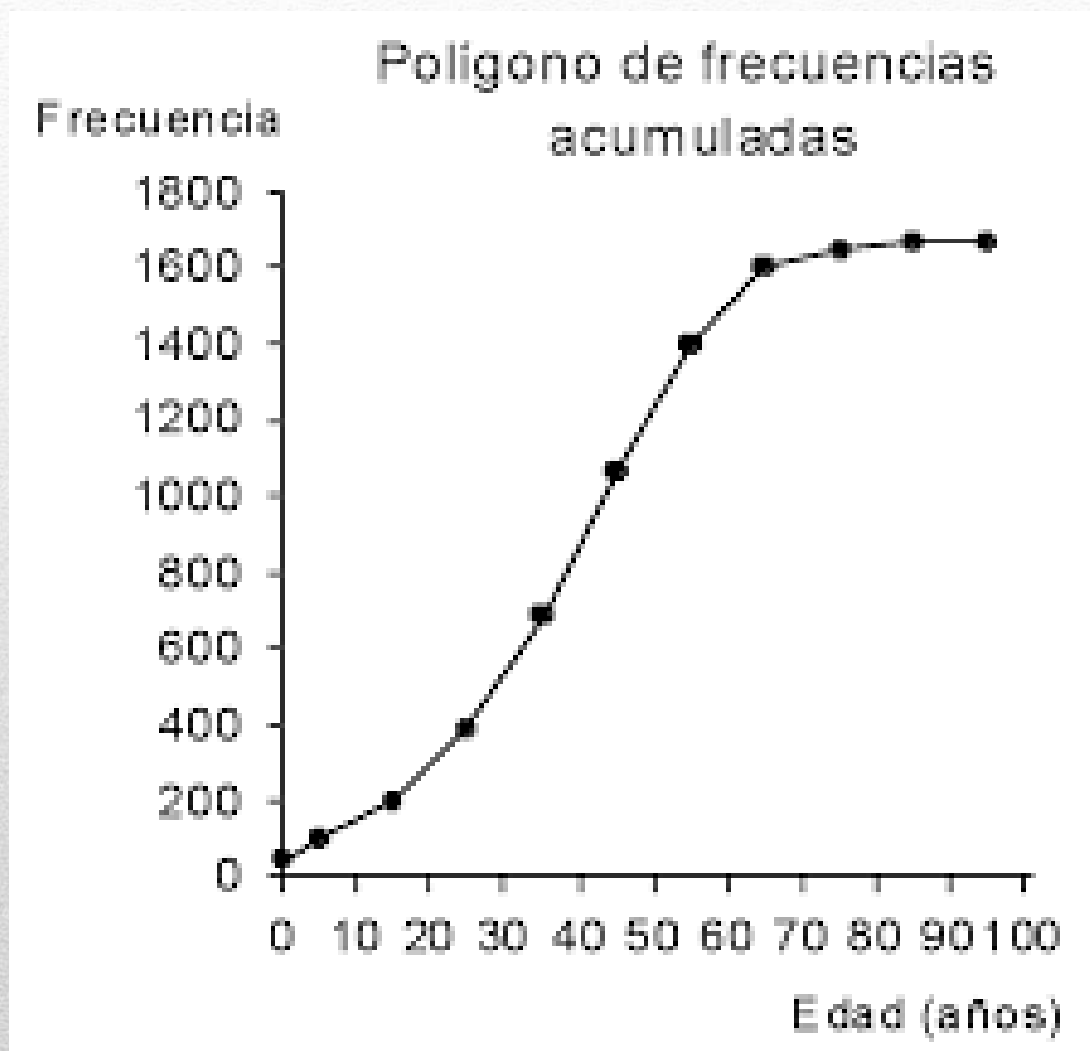


Diagrama de frecuencias acumuladas



Ojiva



Medidas de Centralización

Situación 1:

Los siguientes datos corresponden al número de interrupciones semanales en el sistema de red de una prestigiosa Universidad registrado el último año.

Nº interrupciones	fi
0	20
1	15
2	8
3	6
4	2
5	2

- a)¿Cuál es el número promedio de interrupciones semanales?
 - b)¿Cuál es el número de interrupciones que más predomina?
 - c)¿Cuál es el número de interrupciones del 50% de las semanas?
-

Situación 2

Una sucursal de una cadena de minisupers tiene siete empleados, los cuales se quejan a la gerencia asegurando que el salario promedio semanal en esa sucursal es de \$90000. La gerencia responde que el salario promedio es de \$113750. La siguiente tabla resume los datos:

Cargo	Salario	N° de empleados
Gerente	250000	1
Subgerente	190000	1
Cajero	110000	2
Vendedor	95000	1
Auxiliar de venta	90000	5
Mantenimiento	80000	2

¿Quién tiene razón los empleados o la gerencia? ¿cuál de las dos respuesta representa mejor al conjunto de los valores?

Medidas de Dispersión

Situación 1

Un equipo de baloncesto necesita un pívot. Se han seleccionado dos jugadores que en sus últimos 10 partidos han anotado estos puntos. ¿Cuál de ellos elegirías?

Jugador A	26	17	8	21	15	7	9	10	28	9
Jugador B	12	15	16	18	17	14	16	16	12	14

Situación 2

Anabel Miranda utiliza dos máquinas diferentes para producir papeleras para las fotocopadoras Mita. Una muestra de las papeleras de la primera máquina midieron 12.9, 12.2, 12.7, 12.3, 11.8, 11.5 y 11.2 pulgadas. Las bandejas elaboradas con la segunda máquina midieron 12.9, 12.8, 12.6, 11.1, 11.5, 11.8, y 11.6 pulgadas. Anabel debe utilizar la máquina con la mayor consistencia en los tamaños de las papeleras. ¿Cuál máquina debe utilizar?

Medidas de concentración

La siguiente tabla muestra la antigüedad en años, del personal dentro de una empresa:

Años	0-1	1-3	3-5	5-8	8-15	15-20	20-30
Frecuencia	850	380	210	110	100	80	30

Al firmar el contrato colectivo, se conviene en darle un bono por antigüedad equivalente al 80% de su sueldo a los empleados que estén en el grupo del 10% más antiguos, al 50% de su sueldo , al grupo que esté en el siguiente 30% más antiguo y al 30% del sueldo, al grupo del 40% siguiente en antigüedad.

Medidas de forma

Una distribución es *simétrica* cuando valores de las variables equidistantes, a uno y otro, del valor central tienen la misma frecuencia. Es decir, si la dispersión es igual o similar a ambos lados.

Una distribución es *asimétrica* cuando sus datos tienden agruparse hacia uno de los extremos de la distribución.

Tomando como eje referencia el valor de la moda, es tipo de sesgo se define por el grado de dispersión de los datos a ambos lados (colas) de ese eje. La cola más dispersa en el lado de los valores altos de la variable caracteriza a la simetría positiva; si en el lado de los más bajos, asimetría negativa.

