

Contenidos

1. Introducción

2. Estadística Descriptiva:

- Univariante

- Bivariante

3.- Probabilidades

4.- Distribución de probabilidades

Bibliografía

1. Murray y Espieges “ Estadística”
2. Paul Newbold “ Estadística para los negocios”
3. Mario Triola “ Probabilidad y Estadística”
4. L. Chao “ Estadística para la ciencia Administrativa “
5. R. Levin “ Estadística para Administradores “
6. Montgomery “ Probabilidad y Estadística para Ingeniería
7. Apuntes Prof.

Programación

Unidad 1: Estadística Descriptiva

- Población y muestra.
- Variables y atributos.
- Técnicas de muestreo: muestreo aleatorio simple, conglomerados, doble, estratificado, sistemático.
- Estadígrafos de tendencia central.
- Estadígrafos de dispersión.
- Análisis Exploratorio de Datos: recolección y presentación de datos., clasificación, representación y métodos gráficos.

Unidad 2: Probabilidades

- Experimentos aleatorios.
- Espacio muestral.
- Eventos.
- Sucesos.
- Probabilidad como frecuencia relativa y como función matemática.
- Axiomas.
- Probabilidad condicional e independencia.
- Probabilidad total.
- Teorema de Bayes.

Unidad 3: Variables Aleatorias Unidimensionales

- Variables discretas. - Funciones de probabilidad y de distribución acumulada.
- Variables continuas. - Funciones de densidad y de distribución acumulada.
- Esperanza y Varianza.
- Función generatriz de momentos.
- Distribuciones de probabilidad discretas: Bernoulli; Binomial; Geométrica; Hipergeométrica; Poisson.
- Distribuciones de probabilidad continuas: Uniforme; Exponencial; Normal, Beta, Gamma.

Modalidad del Curso

- Pruebas

- 33% prueba 1 27 Abril
- 33% prueba 2 27 Mayo
- 34% prueba 3 29 Junio
- Recuperativa 6 Julio

- Observación: La recuperativa la dan los alumnos que han faltado una de las prueba, y/o quienes necesiten nota para aprobar. Estos últimos podrán optar como nota máxima, la nota necesaria para obtener un 55 final.

ESTADISTICA

CIENCIA QUE NOS PERMITE TOMAR
DECISIONES BAJO CIERTA
INCERTIDUMBRE

Términos Comunes

Experimento

Proceso de realizar una observación o una medición.

Variable

Característica o fenómeno, que puede tomar distintos valores.

Dato

Resultado de la observación de una variable.

Población

Conjunto total de elementos o individuos, que poseen una característica común, acerca de la cual se quiere información

Muestra

Subconjunto de la población, seleccionada de acuerdo a una regla o un plan.

Censo

Obtención de todos los datos de interés que posee la población

Estadístico

Función o formula que depende de los datos de la muestra,
Su valor es conocido, pero variable.

Parámetro

Función o formula que depende de los datos de la población.

Inferencia

Su valor es fijo, único, pero en general, desconocido.

Estimación del parámetro a través del estadístico

Método Científico

1. Planteamiento del Problema
2. Diseño del Experimento
3. Experimentación y Recolección
4. Organización y Descripción de Resultados
5. Inferencia Estadística

Muestreo

Notación:

N : Tamaño de la Población

n : Tamaño de la Muestra

Tipos de Muestreos Aleatorios

- Aleatorio Simple
- Estratificado
- Sistemático
- Por Conglomerado
- Muestreo en Etapas

Muestreo Aleatorio Simple

Todos los elementos de la población, tienen la misma oportunidad de estar contenidos en la muestra.

Se enumeran todos los elementos, y luego se extraen de uno en uno, aleatoriamente, hasta completar el tamaño de la muestra.

Cuando la población es muy grande, se puede recurrir a los números aleatorios.

Muestreo Estratificado

Método de selección, utilizado cuando la población, está dividida en grupos llamados estratos, cada uno, formado por una gran cantidad de elementos homogéneos.

Se toma una muestra aleatoria simple, en cada estrato.

Los estratos, pueden ser de igual o distinto tamaño, si son distintos, una manera posible de determinar el tamaño de la muestra al interior de cada estrato, es que esta sea proporcional al tamaño del mismo, a este tipo de asignación, se le conoce como **Afijación Proporcional**, que no siempre resulta la mejor, debido al costo de muestreo en cada estrato.

Muestreo Sistemático

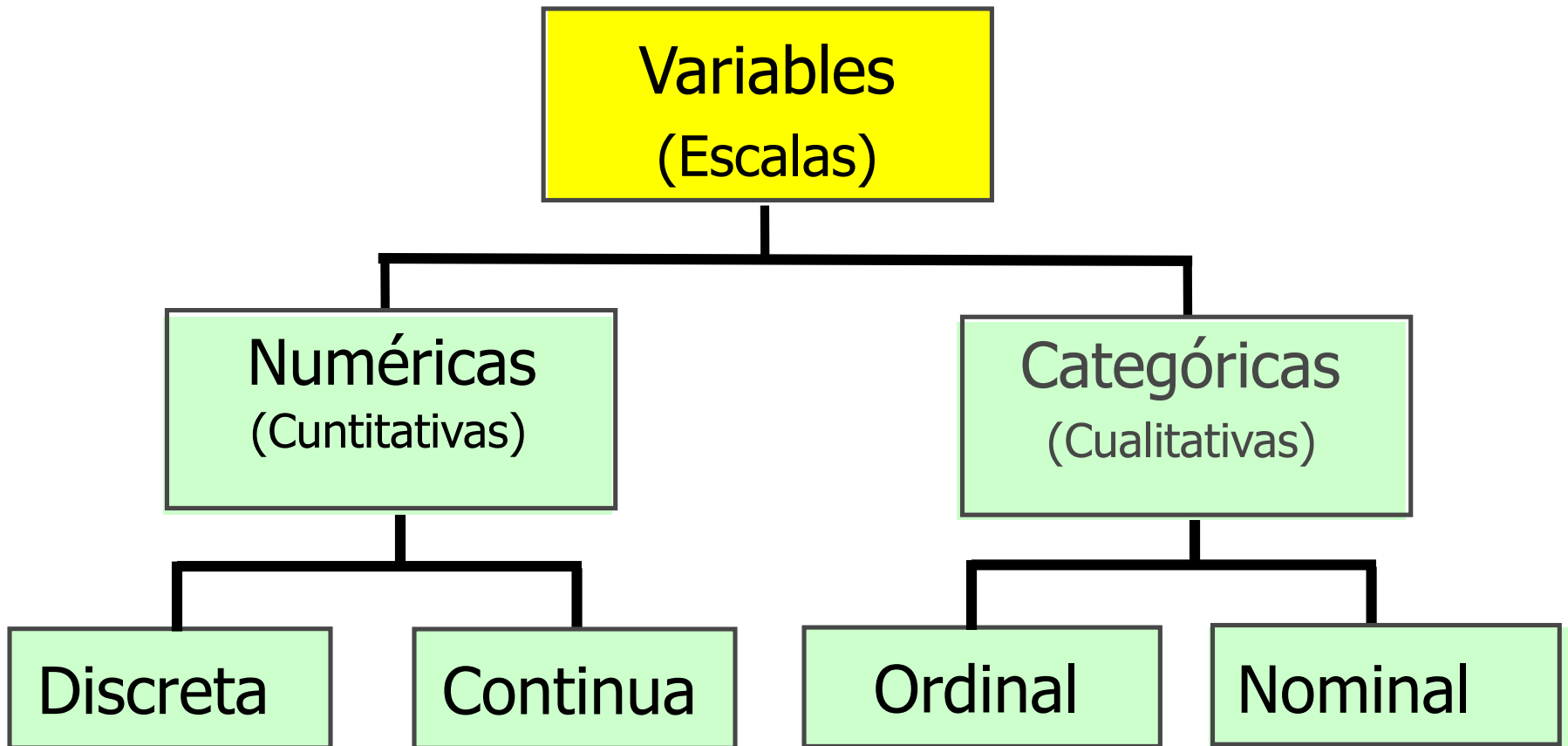
Este método, se utiliza cuando la población tiene sus elementos ordenados. Se divide la población (de tamaño N) en tantas sub poblaciones, como sea el tamaño de la muestra (n), todas de igual tamaño ($k = N/n$).

Se selecciona al azar un elemento de la primera sub población, y de ahí en adelante, de las sub poblaciones siguientes, se extrae el elemento correspondiente.

Muestreo por Conglomerado

Se utiliza, cuando la población, está dividida en una gran cantidad de pequeños grupos, llamados Conglomerados, cada uno formado por elementos heterogéneos. Se toma una muestra aleatoria de Conglomerados, y luego se censan todos los conglomerados seleccionados.

Tipos (género) de Variables (Datos)



Ejemplo

Variable	Género (Tipo)
Nombre	Categórica - Nominal
Dirección	Categórica - Nominal
Edad	Numérica - Continua
No. Teléfono	Categórica - Nominal
Remuneración	Numérica - Discreta
Color de pelo	Categórica - Nominal
Sonido	Categórica - Ordinal
Intensidad del Sonido	Numérica - Continua
Calidad	Categórica - Ordinal
Velocidad	Numérica - Continua
Hora	No es variable (Unidad de medida de tiempo)

Escala Nominal

- Usa números como una manera de separar los elementos de la población en diferentes clases o categorías.
- *El número asignado a la observación sólo sirve como un nombre o identificador,* para distinguir la categoría a la cual pertenece la observación.
- Ejemplo alumnos por...
 - Sexo : 1: masculino, 2: femenino;
 - Colegio: 1: Trinity College; 2: Scuola Andrea Doria, 3: Lycee du París.

Escala Ordinal

- Existe un orden implícito entre las mediciones.
- El *valor numérico es usado sólo como una manera de arreglar los elementos de acuerdo al orden establecido.*
 - existe una relación de orden total entre las clases.
 - no es posible cuantificar la diferencia entre los individuos pertenecientes a las distintas clases.

■ Ejemplo: Calificaciones

1 (muy bueno), 2 (bueno), 3 (satisfactorio), 4 (admisible), 5 (deficiente)

MB

B

S

A

D

A

B

C

D

E

Organización de Datos

Consideremos que la variable X , se divide
En k clases o categorías, denominadas:

$$C_1, C_2, \dots, C_i, \dots, C_k$$

Dando origen a la Siguiete tabla de Frecuencia

Tablas de Frecuencias

		Absoluta	Relativa	Acumulada Absoluta	Acumulada Relativa
i	X	n_i	f_i	N_i	F_i
1	C_1	n_1	f_1	N_1	F_1
2	C_2	n_2	f_2	N_2	F_2
:	:	:	:	:	:
i	C_i	n_i	$f_i = n_i/n$	$N_i = n_1 + n_2 + \dots + n_i$	$F_i = N_i/n$
:	:	:	:	:	:
k	C_k	n_k	f_k	$N_k = n$	$F_k = 1$
T o t a l		n	1	///	///

Tablas de Frecuencias

- Frecuencia Absoluta:

$$\left((n_i) \Rightarrow n = \sum_{i=1}^k n_i \right)$$

- Frecuencia Relativa :

$$\left(\left(f_i = \frac{n_i}{n} \right) \Rightarrow 1 = \sum_{i=1}^k f_i \right)$$

- Frecuencia Acumulada Absoluta:

$$\left((N_i) \Rightarrow \begin{aligned} N_i &= n_1 + n_2 + \dots + n_i \\ &= \sum_{j=1}^i n_j \end{aligned} \right)$$

- Frecuencia Acumulada Relativa :

$$\left((F_i) \Rightarrow F_i = \frac{N_i}{n} \right)$$

Ejemplo 1: Opinión de alumnos por una bebida nueva

M MB B B R M MM R MB MB R B B M R
B B R MB B B B R M MM B R R B R
B R

Determine: a.- La variable x: **Opinión de los alumnos.....**
 b.- El tipo de Variable: **Categórica Ordinal**
 c.- Una tabla de frecuencia

Tabla					
i	X	ni	f _i	N _i	F _i
1	MM	2	.06	2	.06
2	M	4	.13	6	.19
3	R	10	.31	16	.50
4	B	12	.38	28	.88
5	MB	4	.12	32	1.0
Total		32	1.0	//	//

d.- Determine e Interprete

k= **5** f₃= **.31**

n₃= **10** N₃= **16**

n= **32** F₃= **.50**

e.- ¿Cuántos alumnos consideran que la bebida es al menos regular?

26 alumnos

Ejemplo 2:

Número de artículos defectuosos por día.

4 3 5 8 3 6 3 2 2 4 4 3 3 5 8 6 8 3
6 6 4 3 5 2 2 4 6 6 3 5

Determine:

a.- La variable X:

Número de artic. Defec. Por día

b.- El tipo de Variable:

Numérica - Discreta

c.- Una tabla de frecuencia

Tabla

i	X	n _i	f _i	N _i	F _i
1	2	4	.13	4	.13
2	3	8	.27	12	.40
3	4	5	.17	17	.57
4	5	4	.13	21	.70
5	6	6	.20	27	.90
6	7	0	.00	27	.90
7	8	3	.10	30	1.0
Total		30	1.0	//	//

	*						
	*						
	*			*			
	*	*		*			
*	*	*	*	*			
*	*	*	*	*		*	
*	*	*	*	*		*	
*	*	*	*	*		*	
2	3	4	5	6	7	8	

d.- Determine e Interprete

k= 7 f₄= .13

n₃= 5 N₄= 21

n= 30 F₃= .57

e.- ¿En cuantos días el número de artículos defectuosos fue de al menos 4?

18 días

f.- ¿Cuál fué el número máximo de artículos def. por día, en los 15 días que hubieron Menos artic. def.?

4 artículos

g.- ¿Cuántos artíc. Defec. en total, se juntaron en los 6 días en que hubieron más def.?

42 artículos

Tabla de Frecuencia con Intervalos de Clase

1. Rango: $R = | \text{Valor Máx.} - \text{Valor mín.} | + 1_u$ (1u: Una unidad de medida)
2. Cantidad de Intervalos, Según Sturges:
 $k = 1 + 3.3 \log(n)$ $k \in \mathbb{IN}$
3. Amplitud: $a = R \div k$ ($a \approx$ valor superior cuando no es exacto)
Ej. Si $1_u = 0.01$
y $a = R/k = 4.571 \approx 4.58$ (se expresa en la unidad de medida)
4. Adicionales: $p = a * k - R = p^* + p^{**}$
5. Intervalos : ♦Aparentes (X_A) ♦Reales (X_R)
6. Marcas de Clases (X_i)

$$X_i = (Ls + Li) \div 2$$

Ejemplo

Sea Valor mínimo = 4.7
 Valor Máximo = 12.6
 $n = 42$

Determine los intervalos aparentes y reales con sus respectivas marcas de clase.

$$1. \text{ Rango } R = | 12.6 - 4.7 | + 0.1 = 8.0$$

$$2. K = 1 + 3.3 \log(42) = 6.4 \approx 6$$

$$3. a = 8.0 \div 6 = 1.33 \approx 1.4$$

$$4. P = 1.4 * 6 - 8.0 = 0.4 \quad (= 0.2 + 0.2)$$

5. =Intervalos :

i	X_A	X_R	X_i
1	4.5 – 5.8	4.45 – 5.85	5.15
2	5.9 – 7.2	5.85 – 7.25	6.55
3	7.3 – 8.6	7.25 – 8.65	7.95
4	8.7 – 10.0	8.65 – 10.05	9.35
5	10.1 – 11.4	10.05 – 11.45	10.75
6	11.5 – 12.8	11.45 – 12.85	12.15

Ejemplo 2:

Considere los siguientes datos:

0.94 1.05 0.86 0.94 0.96 1.03 1.01

0.78 0.84 0.86 1.04 0.76 0.65 0.70

Confeccione una tabla de frecuencia.

$$1. \text{ Rango } R = | 1.05 - 0.65 | + 0.01 = 0.41$$

$$2. K = 1 + 3.3 \log(14) = 4.8 \approx 5$$

$$3. a = 0.41/5 = 0.082 \approx 0.09$$

$$4. P = 0.09 * 5 - 0.41 = 0.04 \quad (=0.02 + 0.02)$$

5. =Intervalos :

i	X_A	X_R	X_i	n_i	f_i	N_i	F_i
1	0.63 – 0.71	0.625 – 0.715	0.67	2	0.14	2	0.14
2	0.72 – 0.80	0.715 – 0.805	0.76	2	0.14	4	0.28
3	0.81 – 0.89	0.805 – 0.895	0.85	3	0.21	7	0.50
4	0.90 – 0.98	0.895 – 0.985	0.94	3	0.21	10	0.70
5	0.99 – 1.07	0.985 – 1.075	1.03	4	0.29	14	1.00
				14	1.00		

Complete la siguiente Tabla

Luego como $a = 2.9$

i X_R n_i f_i N_i F_i X_i

1	8.25 - 11.15	12	0.1	12	0.1	9.7
2	11.15 - 14.05	24	0.20	36	0.3	12.6
3	14.05 - 16.95	12	0.1	48	0.4	15.5
4	16.95 - 19.85	48	0.4	96	0.80	18.4
5	19.85 - 22.75	24	0.2	120	1.0	21.3
	Total	120		////	/////	/////

Tarea No. 1 (Excel)

- Determine una muestra aleatoria de tamaño $n = 1500$
- Construya una tabla de frecuencia 9 intervalos.
- Ilustre los datos a través de un:

Histograma

Ojiva

Nota: Consulte videos de youtube.

TIPOS de GRÁFICOS

Principalmente en variables :

- Gráfico de Barras → Categóricas y Numéricas Discretas
- Gráfico de Sectores → Categóricas
- Histograma → Numéricas Continuas
- Polígono → Numéricas Continuas
- Ojiva → Numéricas
- Píctograma → Numéricas y Categóricas
- Diagrama de Caja y Bigote (Tarea)

Gráfico de Barras

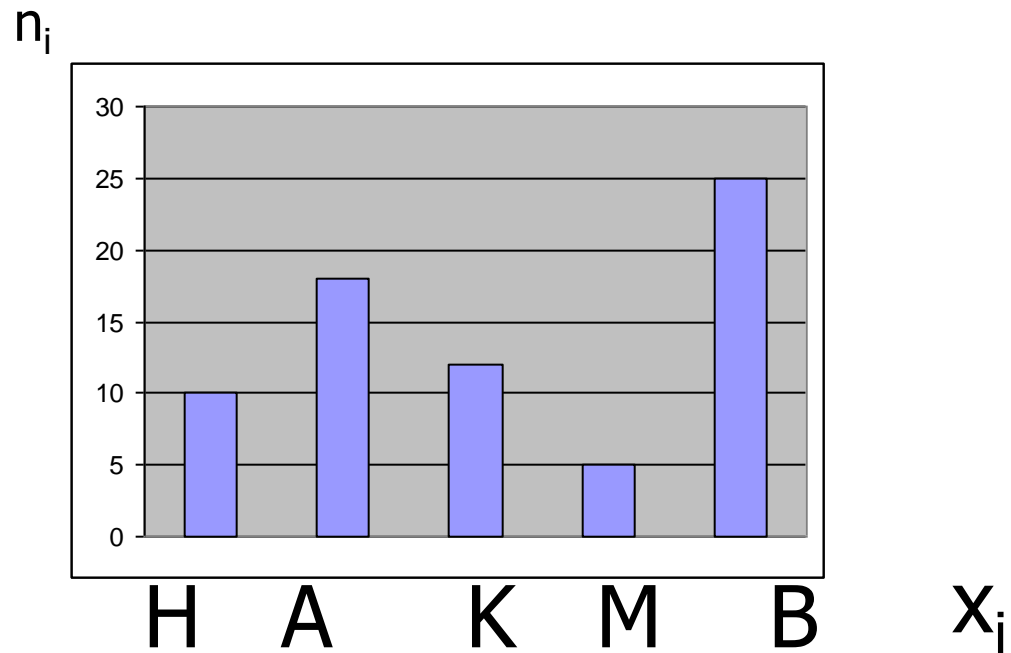
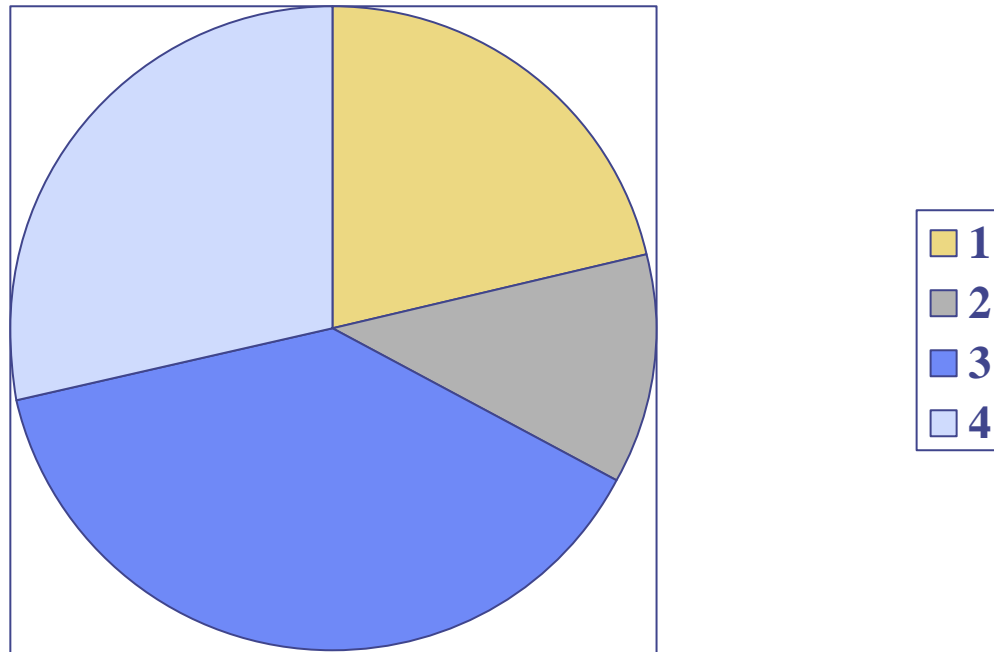
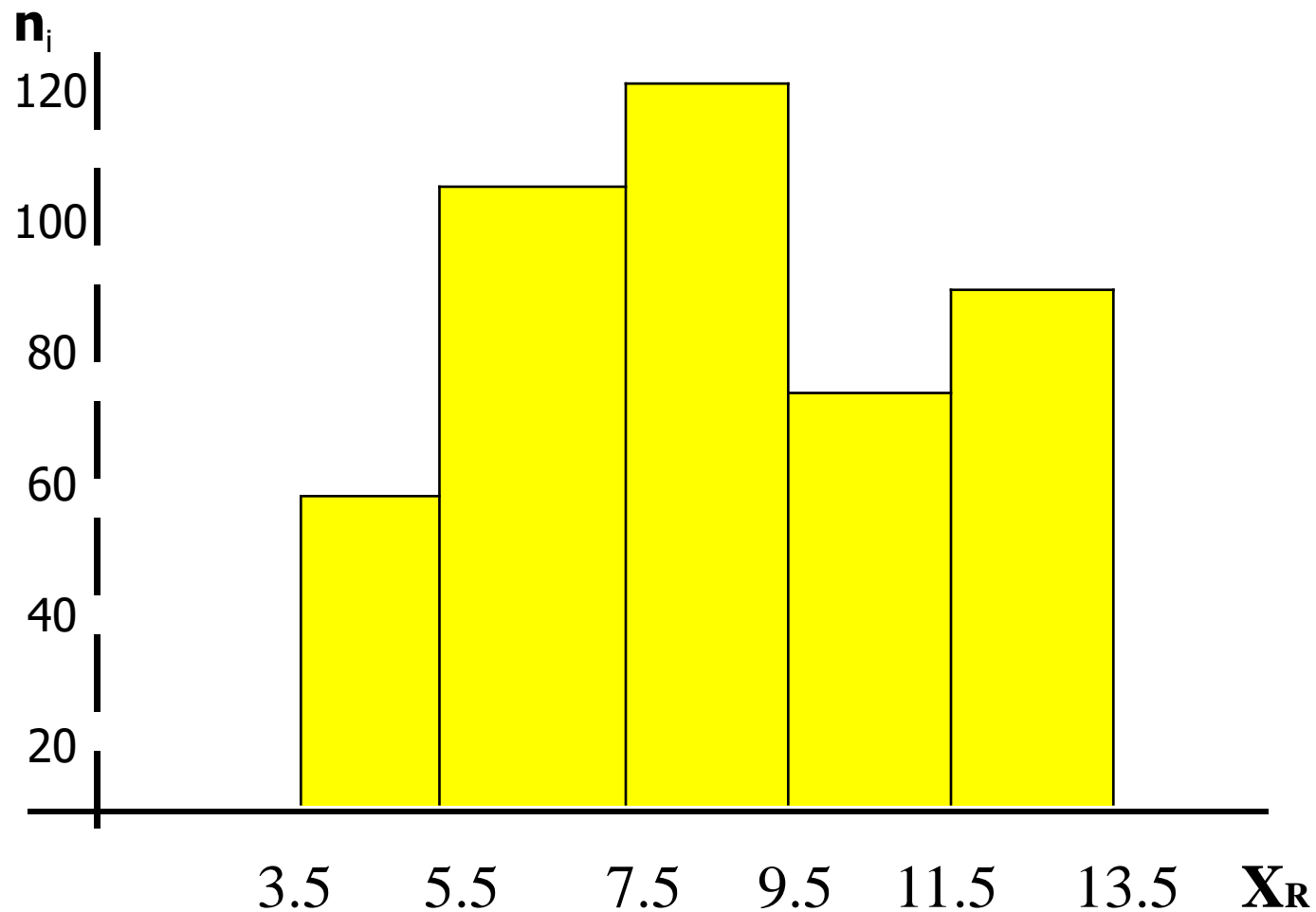


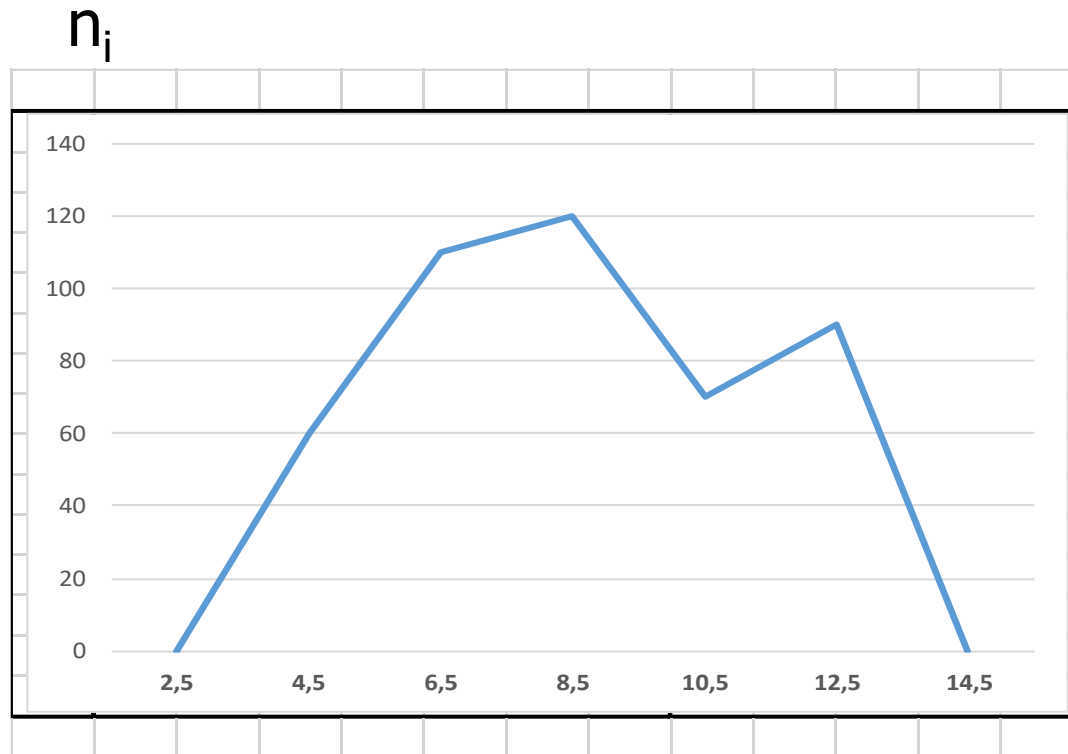
Gráfico de Sectores (ó Circular)



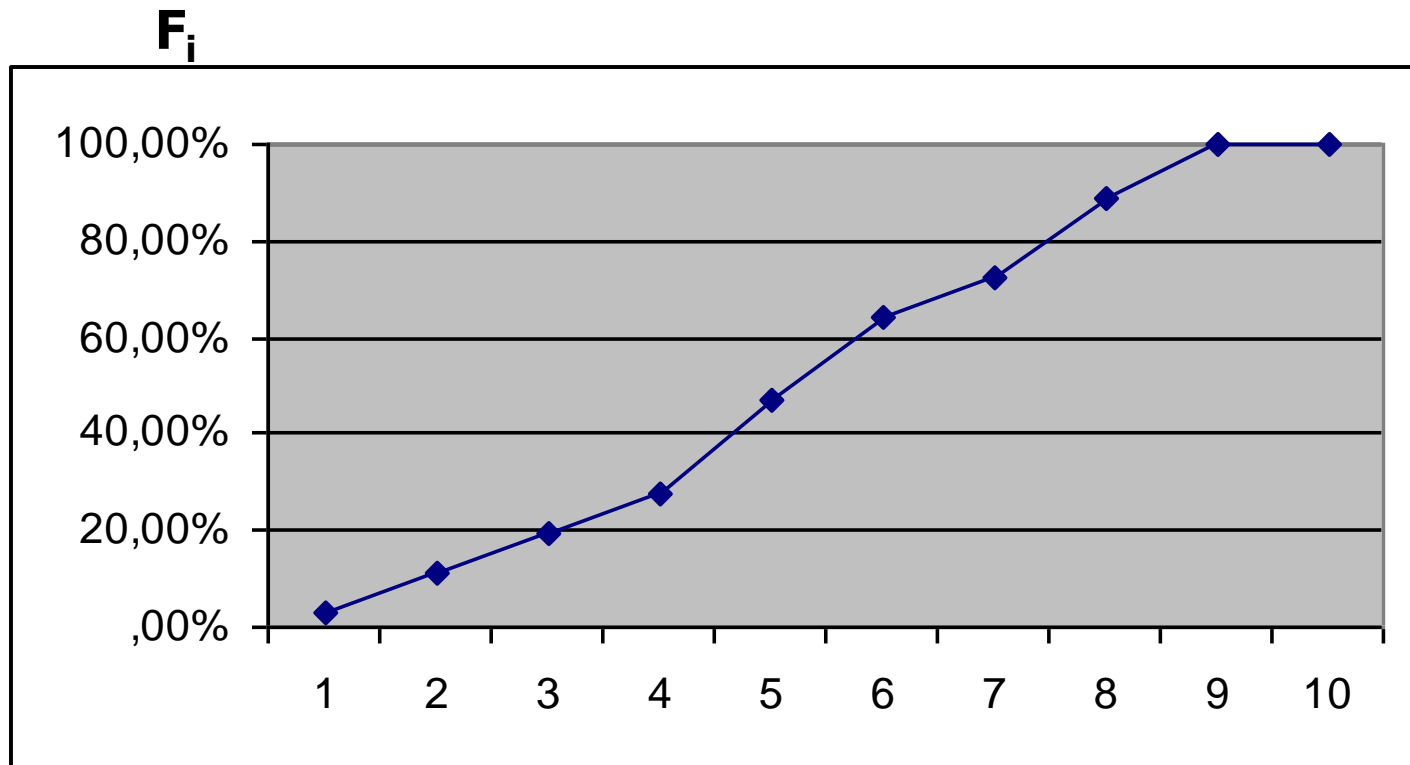
Histograma



Polígono

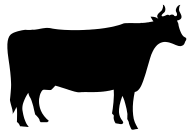


Ojiva

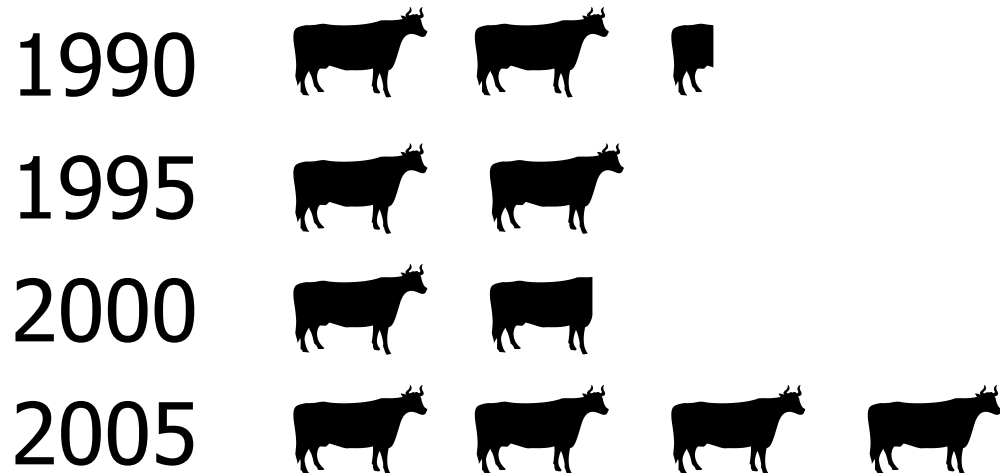


Pictograma

(Consumo de carne de vacuno)



1000 ejemplares



Ejemplo 1

i	x_i	n_i	f_i	$f_i \times 360$	($^\circ$)ac
1	H	10	.143	51.4	51.4
2	A	18	.257	92.5	143.9
3	K	12	.171	61.6	205.5
4	M	5	.071	25.6	231.1
5	B	25	.357	128.5	360.0
To	tal	70	1.00	360	/////

Confeccione un gráfico de Barras y uno de Sectores

Ejercicio

Supongamos que los datos siguientes representan los tiempos, que demoran unos atletas en terminar una maratón.

1:35 1:28 1:45 1:52 1:40 1:30 1:38 1:47 1:37
1:30 1:40 1:36 1:29 1:35 1:37 1:36 1:40 1:36 1:48
2:05 1:32 2:28 1:45 1:50 1:47 2:29 1:44 1:49

1. Identifique la variable
2. Indique el género (tipo)
3. Construya una tabla de frecuencia de 5 intervalos
4. Confeccione un; Histograma, Ojiva y un gráfico de sectores
5. Calcule de la tabla:
 - a.- Media, Moda, Mediana, P_{35} .
 - b.- Si clasifican los 10 primeros, cual es el tiempo máximo para clasificar?
 - c. Construya un grafico de caja y bigote.

ESTADISTICOS (o Estimadores)

- CUANTILES
- MEDIDAS DE TENDENCIA CENTRAL
- MEDIDAS DE DISPERSIÓN

CUANTILES: (ESTADÍSTICOS DE ORDEN)

Cuartíles : (K_k)

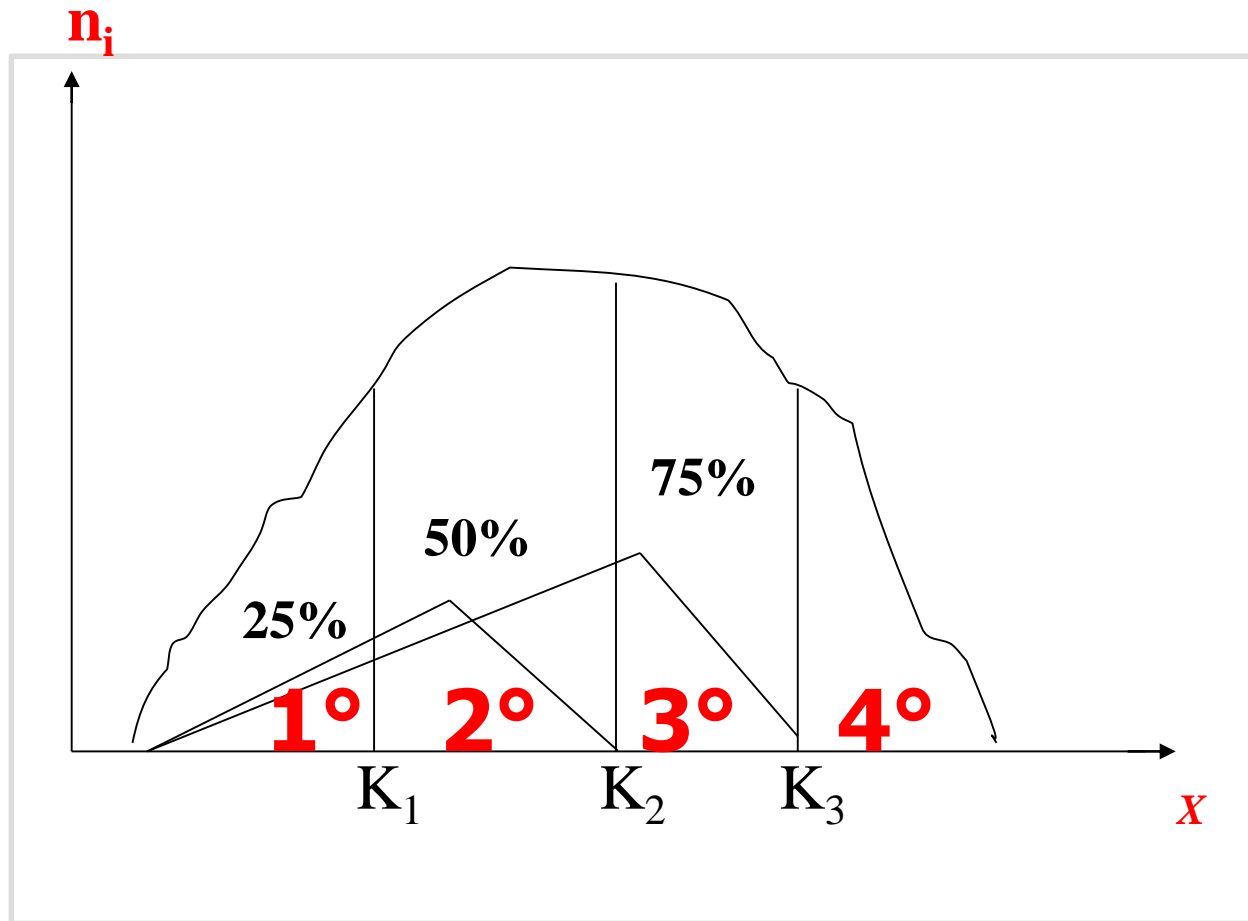
Quintíles : (Q_q)

Decíles : (D_d)

Percentíles : (P_p)

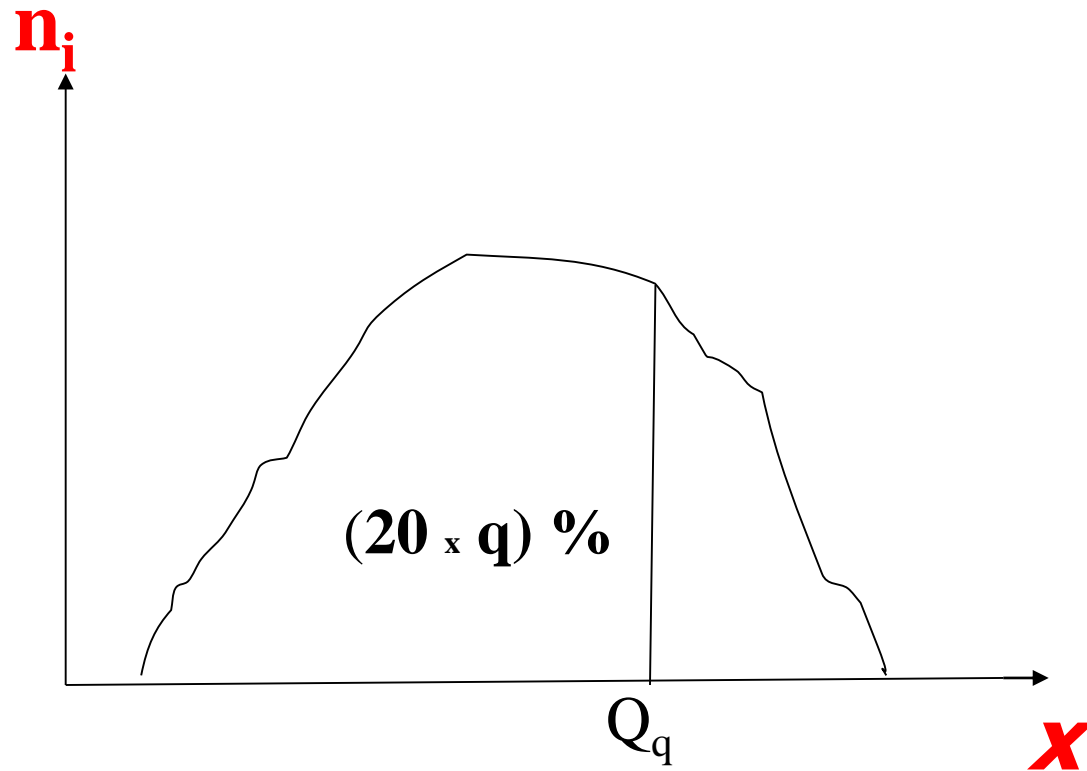
Cuartiles:

$$k = 1, 2, 3$$



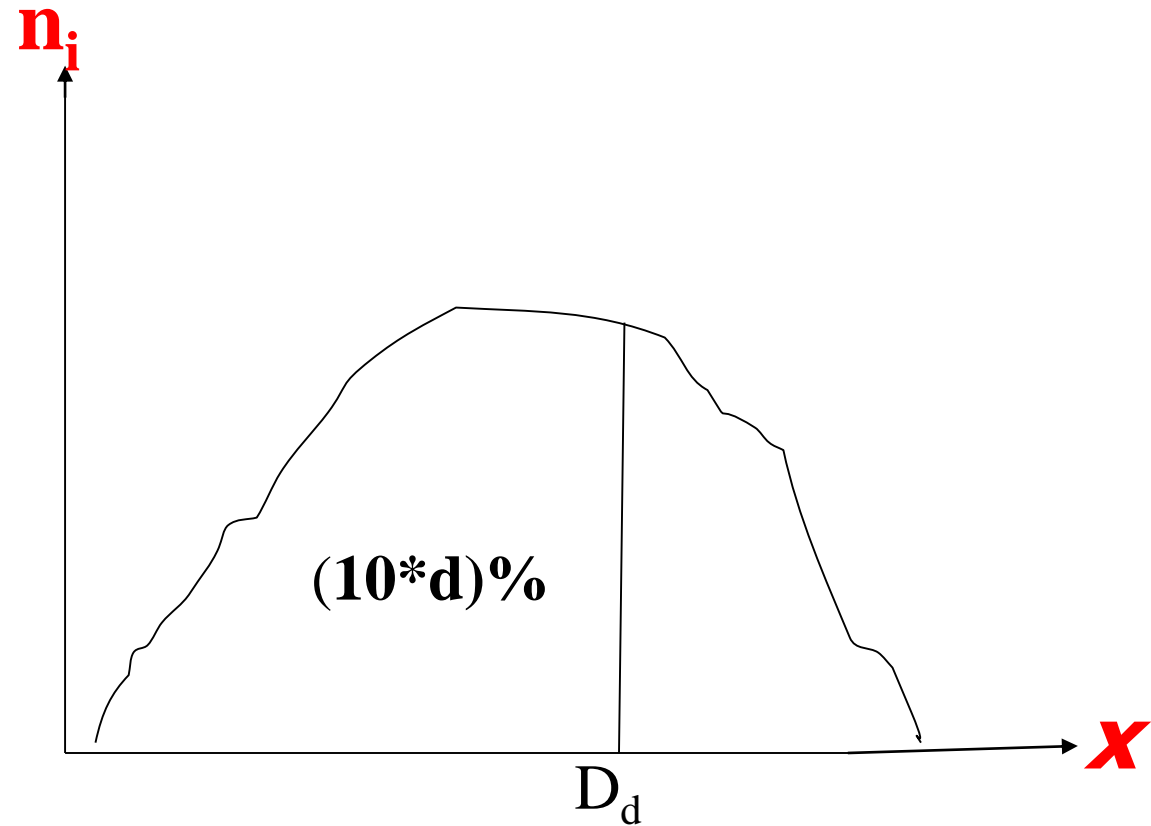
Quintiles:

$$q = 1, 2, 3, 4$$



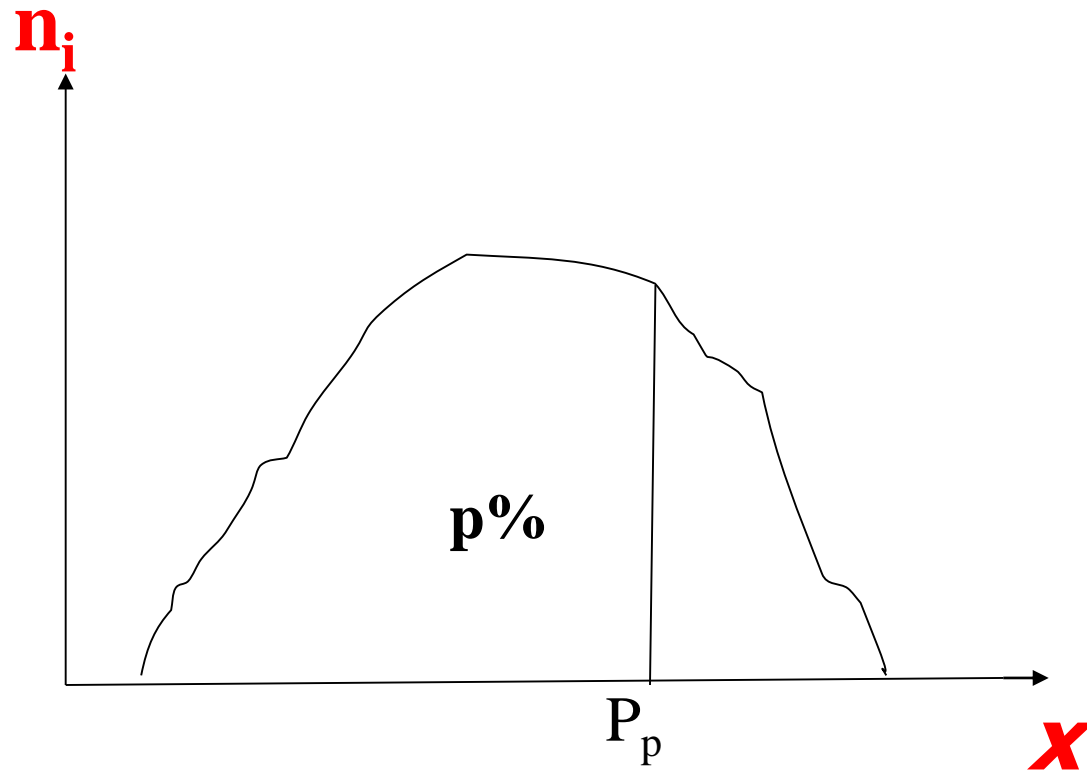
Deciles:

$$d = 1, 2, \dots, 9$$



Percentíles:

$$p = 1, 2, \dots, 99$$



Ejercicio

La tabla siguiente, representa los años de Servicio de los trabajadores de una empresa.

i	X_i	n_i	f_i	N_i	F_i
1	3	15	0.09	15	0.09
2	6	18	0.13	33	0.22
3	9	26	0.17	59	0.39
4	10	38	0.25	97	0.64
5	12	30	0.20	127	0.84
6	18	25	0.16	152	1.00
Total		152	1.00		

1.- Calcular : K_1 , Q_3 , D_2 , P_{45}

$$\blacklozenge K_1 = 9 \qquad \blacklozenge Q_3 = 10$$

$$\blacklozenge D_2 = 6 \qquad \blacklozenge P_{45} = 10$$

2.- ¿Cuál es el año de servicio mínimo de los 30 trabajadores más antiguos

12 años

Cuantiles:

t : 1,2,3

1,2,3,4

1,2.....9

1,2,.....99

Cuartiles

Quintiles

Deciles

Percentiles

**Cuando los datos están ordenados en una tabla de frecuencia
Con intervalos de clase, se puede utilizar la siguiente expresión**

$$C_t = Li + \left[\frac{t * n}{S} - N_{g-1} \right] * \frac{a}{n_g}$$

Li : Limite real inferior del intervalo que contiene el cuantil **t**, que es aquel donde por primera vez la frecuencia acumulada pasa el $(t/S \times 100)\%$.

S : 4, 5, 10, 100 (Cuartil, Quintil, Decíl, Percentil , respectivamente)

N_{g-1}: frec. Acumulada anterior al intervalo que contiene el **C_t**.

n_g : frec. Absoluta del intervalo que contiene el **C_t**.

a : Amplitud del intervalo.

n : Tamaño de Muestra (Total de Datos)

Ejemplo

i	X	n _i	f _i	N _i	F _i
1	2.4 - 2.9	4	.08	4	.08
2	3.0 - 3.5	8	.16	12	.24
3	3.6 - 4.1	15	.30	27	.54
4	4.2 - 4.7	10	.20	37	.74
5	4.8 - 5.3	5	.10	42	.84
6	5.4 - 5.9	8	.16	50	1.0
Total		50	1.0		

$$K_3 = 4.75 + \left[\frac{3 * 50}{4} - 37 \right] * \frac{0.6}{5} = 4.81$$



$$D_1 = 2.95 + \left[\frac{1 * 50}{10} - 4 \right] * \frac{0.6}{8} = 3.025$$



$$P_{35} = 3.55 + \left[\frac{35 * 50}{100} - 12 \right] * \frac{0.6}{15} = 3.77$$



$$P_{74} = 4.75$$

$$Q_3 = 4.33$$



Determine : K_3 , D_1 , P_{35} , y P_{74} , Q_3

Si la variable X representa la utilidad en M\$ por día determine:

i.- ¿Cuál fue la utilidad máxima de los 15 días que ganó menos?

→ M\$ 3.67

ii. ¿En cuántos días la utilidad fue de al menos M\$ 5.0?

→ 11 días

iii.- Que porcentaje de los días tiene una utilidad no superior a 5.0 ni inferior a 4.5 M\$?

MEDIDAS DE TENDENCIA CENTRAL

- MODA
- MEDIANA
- MEDIA ARITMÉTICA
- MEDIA ARMÓNICA
- MEDIA GEOMÉTRICA

MODA (ó MODO)

La denotaremos por : Mo

- a.- Está dada por la observación que más se repite ó la de mayor frecuencia.
- b.- Es posible determinar para cualquier tipo de variable.
- c.- Pueden existir muchas o ninguna.

MODA

Ejemplo 1: Sean los siguientes datos:

F D R F T D R U D U U D

<u>i</u>	<u>X_i</u>	<u>n_i</u>
----------	----------------------	----------------------

1	F	2
---	---	---

2	D	4
---	---	---

3	R	2
---	---	---

4	T	1
---	---	---

5	U	3
---	---	---

Total 12

Luego; en este caso la

Moda es:

Mo = D

MODA

Ejemplo 2 :

2 3 5 1 1 5 2 5 4 2 4 2 5 1 6

<u>i</u>	<u>X_i</u>	<u>n_i</u>
1	1	3
2	2	4
3	3	1
4	4	2
5	5	4
6	6	1
Total		15

En este caso tenemos dos Mo:

$Mo_1 = 2$

$Mo_2 = 5$

MODA

Interpolada

La moda para datos tabulados, se obtiene a partir de la siguiente expresión:

$$M_o = Li + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2} \right] * a$$

Donde: **Li** : Limite real inferior del intervalo que contiene la **Mo** que es aquel que tiene mayor frecuencia

Δ_1 : Diferencia entre las frecuencias absolutas del intervalo modal y el intervalo anterior.

Δ_2 : Diferencia entre las frecuencias absolutas del intervalo modal y el intervalo posterior.

a : Amplitud del intervalo.

Ejemplo a :

<u>i</u>	X	n_i
1	4 - 5	42
2	6 - 7	68
3	8 - 9	102
<u>4</u>	10 - 11	38
T otal		250

$$L_i = 7.5$$
$$\Delta_1 = (102 - 68) = 34$$
$$\Delta_2 = (102 - 38) = 64$$

Calcule la Moda

$$M_o = Li + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2} \right] * a$$

$$M_o = 7.5 + \left[\frac{34}{34 + 64} \right] * 2 = \quad \mathbf{8.2}$$

MODA

Interpolada

La moda interpolada, cuando existen **distintas amplitudes**, se obtiene a partir de la siguiente expresión:

$$M_o = Li + \left[\frac{h_{i+1}}{h_{i-1} + h_{i+1}} \right] * a_i$$

Donde: $h_i = f_i / a_i$: se denomina densidad del intervalo

El intervalo modal, será el que tiene mayor densidad

a_i : Amplitud del intervalo i

Ejemplo b :

Calcule la Moda (intervalos de distinta amplitud)

<u>i</u>	X	n_i	f_i	a_i	h_i
1	0 - 4	42	0.168	4	0.042
2	4 - 5	68	0.272	1	0.272
3	5 - 7	102	0.408	2	0.204
4	7 - 11	38	0.152	4	0.038
T otal		250			

$$M_o = Li + \left[\frac{h_{i+1}}{h_{i-1} + h_{i+1}} \right] * a_i$$

$$M_o = 4 + \frac{0.204}{0.042+0.204} \times 1 = \mathbf{4.83}$$

Obs: $h_i = f_i / a_i$: se denomina densidad del intervalo

MEDIANA

- La denotaremos como : Md
- Puntuación que divide la distribución de los datos (ó la muestra) en dos partes iguales.
- Es decir nos indica el punto hasta donde se tiene acumulado el 50% de las observaciones.
- Nota: - Para su determinación, los datos se ordenan previamente.
- No tiene sentido cuando la variable es categórica nominal.

MEDIANA

- Si el número de observaciones impar, entonces la Md estará dada por la observación central.
- Si el número de observaciones par, entonces la Md estará dada por el promedio de las dos observaciones centrales.

MEDIANA

Ejemplo: 1

2 3 5 5 7 9 10 27 29 30 38 40 980 $n=13$

En este caso la $Md = 10$

Ejemplo: 2

2,1 2,3 2,6 3,6 5,8 6,8 7,9 108,3 $n=8$

Luego ; $Md = (3,6+5,8)/2 = 4,7$

MEDIANA

Ejemplo: dado los siguientes datos:

i	X_i	n_i	f_i	N_i	F_i
1	4	3	.18	3	.18
2	5	2	.11	5	.29
3	6	2	.12	7	.41
4	7	3	.18	10	.59
5	8	4	.23	14	.82
6	9	3	1.0	17	1.0
Total		17			

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
4	4	4	5	5	6	6	7	7	7	8	8	8	8	9	9	9

$$Md = 7$$

Determine; Md

MEDIANA

La mediana para datos tabulados, se obtiene a partir de la siguiente expresión:

$$Obs: M_d = K_2 = D_5 = P_{50}$$

$$M_d = Li + \left[\frac{n}{2} - N_{d-1} \right] * \frac{a}{n_d}$$

Donde: **Li** : Limite real inferior del intervalo que contiene la Md, que es aquel donde por primera vez la frecuencia acumulada pasa el 50%.

n : Tamaño de la muestra

N_{d-1} : Frec. acumulada anterior al intervalo que contiene la Mediana.

n_d : Frec. Absoluta del intervalo que contiene la Md

a : Amplitud del intervalo.

Ejemplo de Mediana

i	X	n _i	f _i	N _i	F _i
1	4 - 5	4	.08	4	.08
2	6 - 7	8	.16	12	.24
3	8 - 9	15	.30	27	.54
4	10 - 11	18	.36	45	.90
5	12 - 13	2	.04	47	.94
6	14 - 15	3	.06	50	1.0
Total		50	1.0		

Determine; M_d , M_o

$$M_d = Li + \left[\frac{n}{2} - N_{d-1} \right] * \frac{a}{n_d}$$

$$M_d = 7.5 + \left[\frac{50}{2} - 12 \right] * \frac{2}{15} = \mathbf{9.23}$$

$$M_o = Li + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2} \right] * a$$

$$M_o = 9.5 + \left[\frac{3}{3 + 16} \right] * 2 = \mathbf{9.82}$$

MEDIA ARITMÉTICA

La media Aritmética también llamada Promedio ó simplemente Media, y esta dada para datos no tabulados por : (Se calcula solamente en variables numéricas)

Muestral	Poblacional
$\overline{X} = \frac{\sum_{i=1}^n X_i}{n}$	$\mu = \frac{\sum_{i=1}^N X_i}{N}$

Donde: X_i : Corresponde a las Observaciones

n : Tamaño de la Muestra

N : Tamaño de la Población

MEDIA ARITMÉTICA

Para datos Tabulados la Media está dada por :

$$\overline{X} = \frac{\sum_{i=1}^k n_i X_i}{n} = \sum_{i=1}^k f_i x_i$$

Donde: X_i : Observación ó Marca de Clase

n : Tamaño de la muestra

n_i : Frecuencia Absoluta de la observación o del intervalo

k : Número de intervalos

Ejemplo: La tabla siguiente representa una muestra de los años de servicio de trabajadores.

i	Xi	ni	xi*ni	fi	fi_xXi
1	5	4	20	0.167	0.835
2	10	7	70	0.292	2.920
3	12	8	96	0.333	3.996
4	15	3	45	0.125	1.875
5	20	2	40	0.083	1.660
Total		24	271	1.000	11.292

Determine la Media:

$$\therefore \bar{X} = \frac{\sum_{i=1}^5 n_i x_i}{n}$$

$$\therefore \bar{X} = \frac{271}{24} = 11.292$$

$$\therefore \bar{X} = \sum_{i=1}^5 f_i x_i = 11.292$$

Ejemplo 2: Se toma una muestra de 21 observaciones, calcule la Media.

i	X	n_i	x_i	n_i*x_i	u_i	n_i*u_i
1	2.3 - 2.4	4	2.35	9.40	-3	-12
2	2.5 - 2.6	5	2.55	12.75	-2	-10
3	2.7 - 2.8	3	2.75	8.25	-1	-3
4	2.9 - 3.0	2	2.95	5.90	0	0
5	3.1 - 3.2	4	3.15	12.60	1	4
6	3.3 - 3.4	3	3.35	10.05	2	6
Total		21		58.95		-15

Luego como $\therefore \bar{X} = \frac{\sum_{i=1}^6 n_i x_i}{n} = \frac{58.95}{21} = 2.807$

$$\bar{X} = A + \frac{\sum_{i=1}^k n_i u_i}{n} a = 2.95 + \frac{-15}{21} 0.2 = 2.807$$

MEDIA PONDERADA

$$\overline{X}_P = \overline{X}_T = \frac{n_1 * \overline{x}_1 + n_2 \overline{x}_2 + \dots + n_k \overline{x}_k}{n_1 + n_2 + \dots + n_k} = \sum_{i=1}^k \frac{n_i * \overline{x}_i}{n}$$

Ejemplo: En una empresa donde se distinguen tres tipos de trabajadores, el salario medio es de **290**. Los **100** trabajadores de la categoría **A** tienen un salario medio de **480**, los **300** de la categoría **B** tienen un salario medio de **320**, ¿ Cuánto es el salario medio de los **400** de la categoría **C**?

$$290 = \frac{100 \times 480 + 300 \times 320 + 400 \times \overline{X}_c}{100 + 300 + 400}$$

$$\text{Luego } \overline{X}_c = 220$$

Tarea N°

- **Ventajas y Desventajas de:** Moda, Mediana y Media.
- **Defina y de un ejemplo de :**
- Media Geométrica : (G)
- Media Armónica : (H)
- Sesgo
- Curtosis
- Ocupando los datos de la tarea anterior;
 - a) Calcule; **Moda, Mediana y Media**, con los datos; no tabulados y tabulados.
 - b) Construya un grafico de caja y bigotes

Media Geométrica:

$$G = \sqrt[n]{\prod_{i=1}^n x_i} = \sqrt[n]{x_1 * x_2 * \dots * x_n}$$

Media Armónica:

$$H = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

$$H = \frac{n}{\sum_{i=1}^n \frac{1}{x_i} * n_i}$$

MEDIDAS DE DISPERSIÓN

Rango: (R)

Desviación Intercuartílica: (K)

Desviación Media: (DM)

Varianza: (s^2)

Desviación Típica: (s)

MEDIDAS DE DISPERSIÓN

Rango: (R)

$$\mathbf{R} = \left| \text{V. Máximo} - \text{V. Mínimo} \right| + \mathbf{1} \text{ Unidad de Medida}$$

MEDIDAS DE DISPERSIÓN

Desviación Intercuartílica: (K)

$$K = \frac{K_3 - K_1}{2}$$

MEDIDAS DE DISPERSIÓN

Desviación Media: (DM)

Para datos No Tabulados

$$DM = \frac{\sum_{i=1}^n |X_i - \bar{X}|}{n}$$

Para datos Tabulados

$$DM = \frac{\sum_{i=1}^k n_i |X_i - \bar{X}|}{n}$$

MEDIDAS DE DISPERSIÓN

Varianza: (Muestral S^2 , Poblacional σ^2)

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N} = \frac{\sum_{i=1}^N x_i^2}{N} - \mu^2$$

Muestral : Para datos No Tabulados

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} = \frac{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}{n(n - 1)}$$

MEDIDAS DE DISPERSIÓN

Muestral : Para datos Tabulados

$$S^2 = \frac{\sum_{i=1}^k n_i (X_i - \bar{X})^2}{n - 1} = \frac{n \sum_{i=1}^k n_i x_i^2 - \left(\sum_{i=1}^k n_i x_i \right)^2}{n(n - 1)}$$

Ejemplo

Sean las edades (en año) de 5 niños

Obs. :	x_1	x_2	x_3	x_4	x_5	Total	Determinar: Desv Media Varianza
X_i :	2	5	6	8	9	30	
$x_i - \bar{x}$:	- 4	-1	0	2	3	0	$\bar{X} = \frac{30}{5} = 6$
$x_i - \bar{x}$:	4	1	0	2	3	10	$\therefore DM = \frac{\sum_{i=1}^n X_i - \bar{X} }{n} = \frac{10}{5} = 2$
$(x_i - \bar{x})^2$:	16	1	0	4	9	30	
X^2_i :	4	25	36	64	81	210	

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{30}{4} = 7.5$$

$$S^2 = \frac{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}{n(n-1)} = \frac{5 * 210 - (30)^2}{5 * 4} = 7.5$$

MEDIDAS DE DISPERSIÓN

Desviación Típica: (S , σ)

Poblacional: $\rightarrow \sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}} = \sqrt{\frac{\sum_{i=1}^N x_i^2}{N} - \mu^2}$

Muestral : Para datos No Tabulados

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} = \sqrt{\frac{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}{n(n - 1)}}$$

MEDIDAS DE DISPERSIÓN

Para datos Tabulados

$$S = \sqrt{\frac{\sum_{i=1}^k n_i (X_i - \bar{X})^2}{n - 1}} = \sqrt{\frac{n \sum_{i=1}^k n_i x_i^2 - \left(\sum_{i=1}^k n_i x_i \right)^2}{n(n - 1)}}$$

COEFICIENTE DE VARIACIÓN

$$CV = \left| \frac{S}{\bar{X}} \right| \quad \left[= \left| \frac{\sigma}{\mu} \right| \right]$$

Mide la **homogeneidad** que existe en los datos respecto a la variable en estudio.

Mientras más pequeño , (cercano a cero), más homogéneo.

-

Ejemplo 2

Calcule: la Media , Desv. Típica, CV

i	X	n _i	x _i	n _i ·x _i	n _i ·x _i ²
1	2.3 - 2.4	4	2.35	9.40	22.09
2	2.5 - 2.6	5	2.55	12.75	32.51
3	2.7 - 2.8	3	2.75	8.25	22.69
4	2.9 - 3.0	2	2.95	5.90	17.41
5	3.1 - 3.2	4	3.15	12.60	39.69
6	3.3 - 3.4	1	3.35	3.35	11.22
		19		52.25	145.61

i	Y	ni	yi	niyi	niyi ²
1	2.2 - 2.4	10			
2	2.4 - 2.6	25			
3	2.6 - 2.8	48			
4	2.8 - 3.0	64			
5	3.0 - 3.2	43			
6	3.2 - 3.4	22			
7	3.4 - 3.6	8			
	Total				

Cuál de las variables X e Y es más homogénea?

Sabemos que: $\bar{X} = \frac{\sum_{i=1}^k n_i x_i}{n} = \frac{52.25}{19} = 2.75$

$$S = \sqrt{\frac{n \sum_{i=1}^n n_i x_i^2 - \left(\sum_{i=1}^n n_i x_i \right)^2}{n(n-1)}} = \sqrt{\frac{19(145.61) - (52.25)^2}{19(19-1)}} = \mathbf{0.327}$$

$$CV = \frac{s}{x} = \frac{0.327}{2.75} = 0.119$$

Coeficiente de Asimetría (Sesgo)

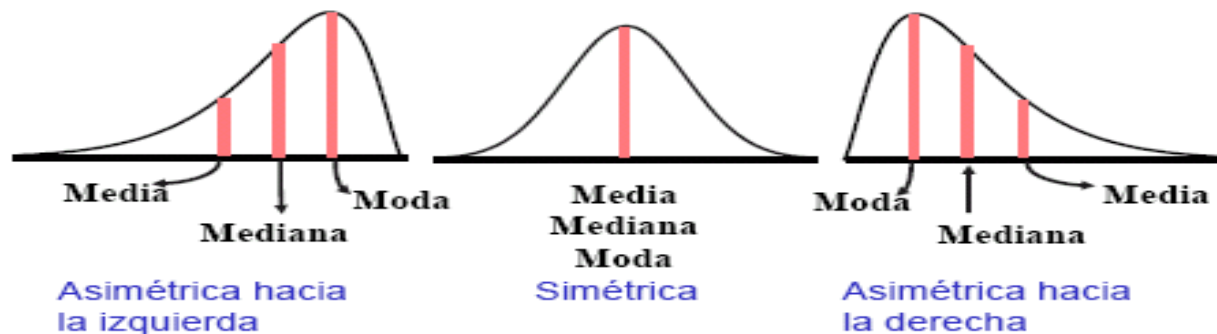
$$A_s = \frac{\bar{x} - Mo}{S}$$

de Pearson

$$\gamma_1 = \frac{1}{S^3} \sum_{i=1}^k n_i (x_i - \bar{x})^3 = \frac{\mu_3}{S^3}$$

de Fisher

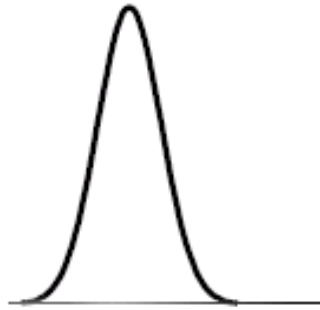
Su valor es cero cuando la distribución es simétrica, positivo cuando existe asimetría a la derecha y negativo cuando existe asimetría a la izquierda.



Coeficiente de Curtosis:

$$\gamma_2 = \frac{1/n}{S^4} \sum_{i=1}^k n_i (x_i - \bar{x})^4 - 3 = \frac{\mu_4}{S^4} - 3$$

$\gamma_2 > 0$: *Leptocúrtica*
 $\gamma_2 = 0$: *Mesocúrtica*
 $\gamma_2 < 0$: *Platicúrtica*



Leptocúrtica



Mesocúrtica



Platicúrtica

Estadística Descriptiva Bivariante

Estadística Descriptiva Bivariante

Sean las siguientes variables:

$X : A_1, A_2, \dots, A_f$

$Y : B_1, B_2, \dots, B_c$

Estas variables, se pueden ordenar en una tabla de doble entrada llamada **Tabla de Contingencia**:

Tabla de Contingencia

X \ Y	B₁	B₂	..B_j..	B_c	Total
A₁	n₁₁	n₁₂	:		n_{1.}
A₂	n₂₁	n₂₂	:		n_{2.}
A_i	n_{ij}	n_{i.}
:			:		:
A_f			:		n_{f.}
Total	n.₁	n.₂	... n._j	n._c	n

Frecuencia Absoluta (Conjunta)

Se denota por n_{ij} y se define como: Cantidad de elementos que cuentan simultáneamente con la característica A_i de la variable X y B_j de la variable Y

Frecuencia Absoluta Marginal

$$\mathbf{n_{i\bullet}}: \text{Total de la fila } \mathbf{i} = \sum_{j=1}^c n_{ij}$$

$$\mathbf{n_{\bullet j}}: \text{Total de la Columna } \mathbf{j} = \sum_{i=1}^f n_{ij}$$

$$\mathbf{n_{\bullet\bullet}}: \text{Total General} = \sum_{i=1}^f \sum_{j=1}^c n_{ij} = \mathbf{n} \quad (\text{Tamaño muestra})$$

Frecuencia Relativa Conjunta

Se denota por $\mathbf{f}_{ij} = \frac{n_{ij}}{n}$

Frecuencia Marginal Relativa

$$\mathbf{f_{i\cdot}} = \sum_{j=1}^c f_{ij} = \frac{n_{i\cdot}}{n}$$

$$\mathbf{f_{\cdot j}} = \sum_{i=1}^f f_{ij} = \frac{n_{\cdot j}}{n}$$

$$\mathbf{f_{..}} = \sum_{i=1}^f \sum_{j=1}^c f_{ij} = 1$$

Frecuencia Condicional

De x dado y (x/y):

$$f_{i/j} = \frac{f_{ij}}{f_{.j}} = \frac{n_{ij}}{n_{.j}}$$

De y dado x (y/x):

$$f_{j/i} = \frac{f_{ij}}{f_{i.}} = \frac{n_{ij}}{n_{i.}}$$

Independencia de Variables

Diremos que dos variables X e Y son independientes si y solo si, la conjunta es igual al producto de las marginales, para todo i , y para todo j . Es decir;

$$f_{ij} = f_{i.} * f_{.j} \quad \forall_i, \forall_j$$

$$o \quad n_{ij} = \frac{n_{i.} * n_{.j}}{n}$$

Estadística Descriptiva Bivariante

Ejemplo: Considere una muestra de contenedores en un recinto portuario.

Sea X : Peso (toneladas) Y : País de origen

$X \setminus Y$	Francia	Alemania	Japón	España	Total
2 – 4	4	5	4	10	23
4 – 6	8	7	10	2	27
6 – 8	3	9	3	10	25
8 – 10	1	10	7	6	24
10 - 12	8	2	0	12	22
Total	24	33	24	40	121

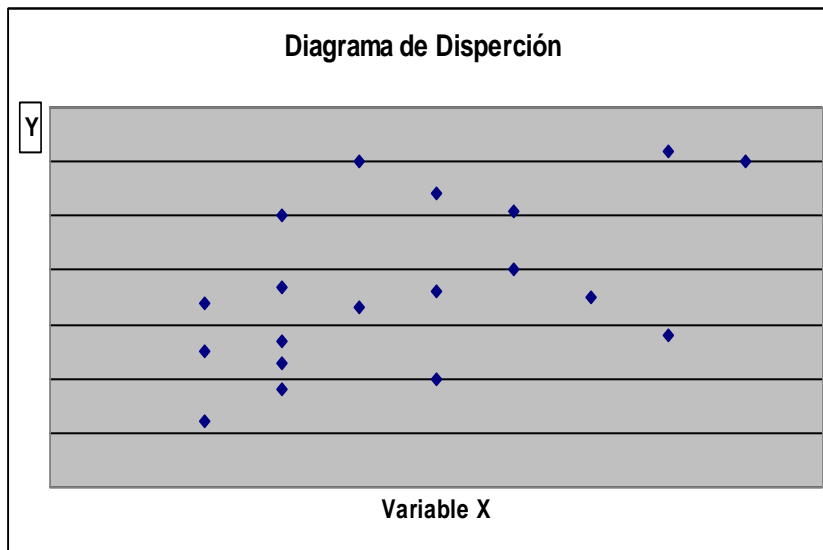
Determine e Interprete : $f = 5$ $c = 4$ $n_{32} = 9$ $n_{2.} = 27$ $n_{.3} = 24$ $n_{..} = 121$

$$\begin{array}{ccccc}
 f_{34} = \mathbf{10/121} & f_{4.} = \mathbf{24/121} & f_{.2} = \mathbf{33/121} & f_{i=4/j=3} = \mathbf{7/24} & f_{j=4/i=3} = \mathbf{10/25} \\
 \mathbf{0.083} & \mathbf{0.20} & \mathbf{0.27} & \mathbf{0.29} & \mathbf{0.40}
 \end{array}$$

Podemos decir que X e Y son independientes?

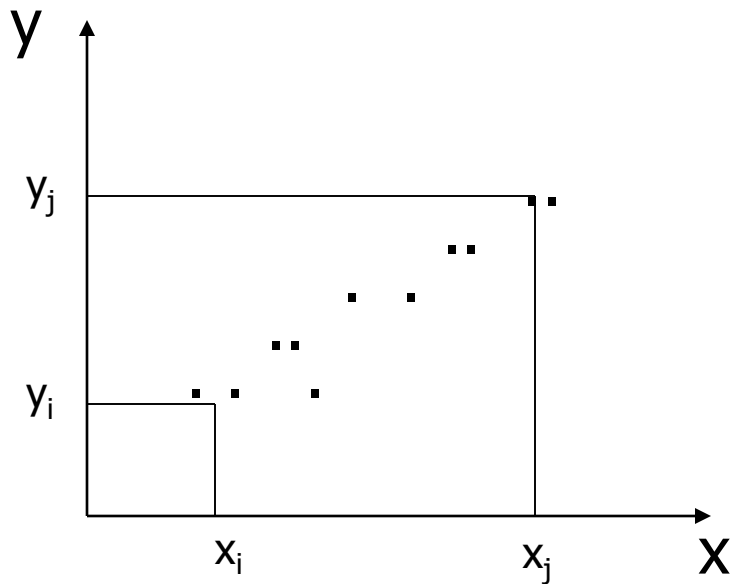
Asociación de Variables Numéricas

i	X	Y
1	x_1	y_1
2	x_2	y_2
.	.	.
.	.	.
.	.	.
n	x_n	y_n

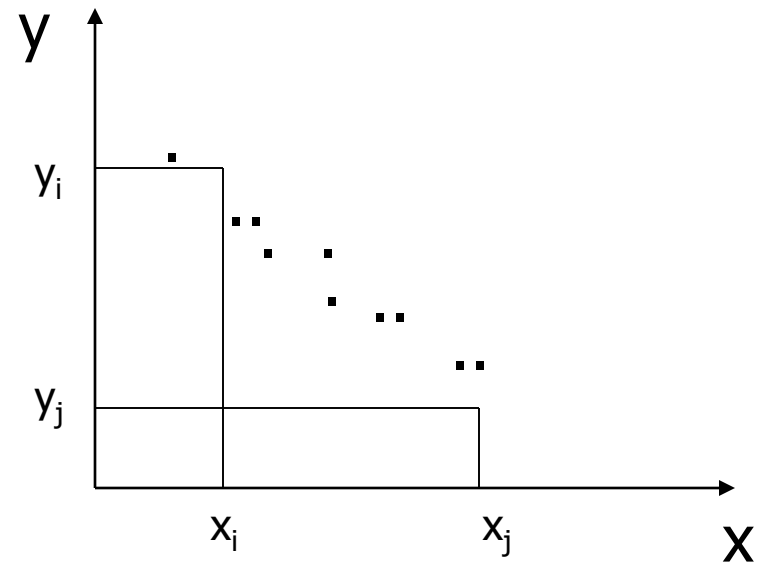


Tipos de Asociación

Favorable :



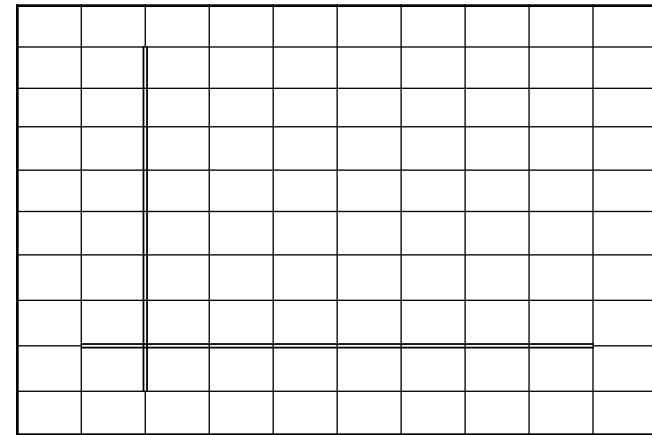
Inversa:



Ejemplo:

Sea X: Remuneración

Y: Cargas Familiares



i	x	y	$(x - \bar{x})$	$(y - \bar{y})$	$(x - \bar{x})(y - \bar{y})$
1	1	2	-2	-1.5	3
2	1	3	-2	-0.5	1
3	3	3	0	-0.5	0
4	4	5	1	1.5	1.5
5	4	2	1	-1.5	-1.5
6	5	6	2	2.5	5
Total	18	21	0	0	9
Media	3.0	3.5	-----		1.5

Covarianza (cov (x,y))

$$Cov(x, y) = \frac{\sum (x - \bar{x})(y - \bar{y})}{n}$$

$$Cov(x, y) = \frac{9}{6} = 1.5$$

$$-\infty \leq Cov(x, y) \leq \infty$$

Coeficiente de Correlación

(Pearson)

Mide el grado de asociación
entre las variables

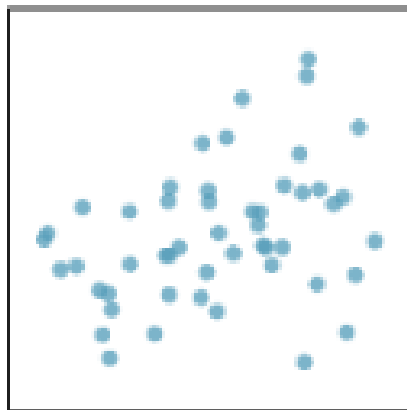
$$r = \frac{Cov(x, y)}{s_x * s_y}$$

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

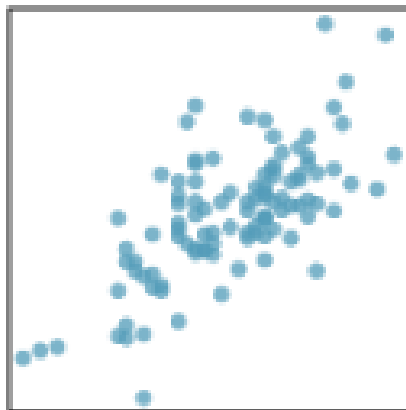
$$-1 \leq r \leq 1$$

r : Se expresa en porcentaje

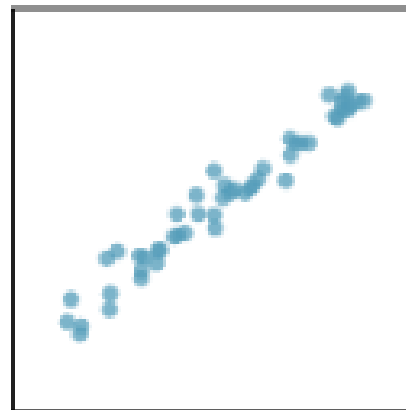
Coeficiente de Correlación



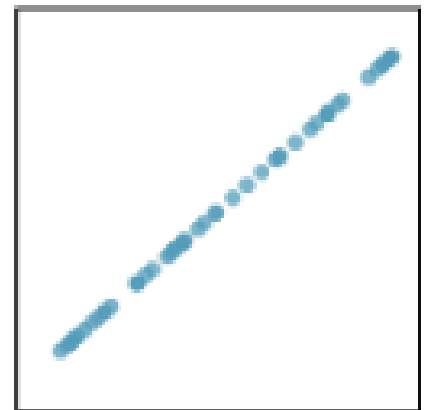
$R = 0.33$



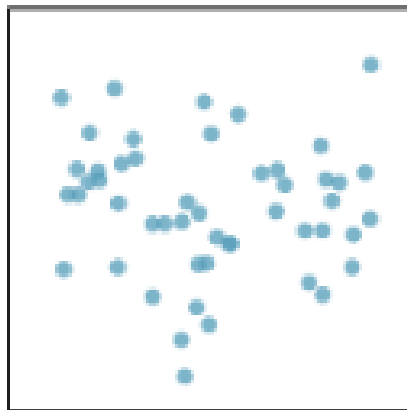
$R = 0.69$



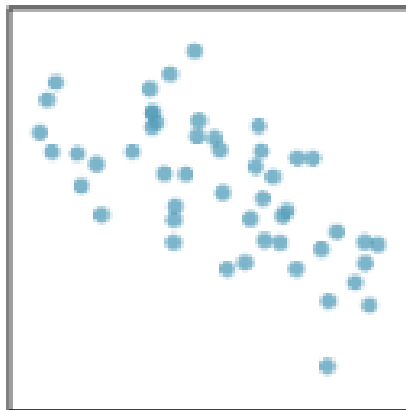
$R = 0.98$



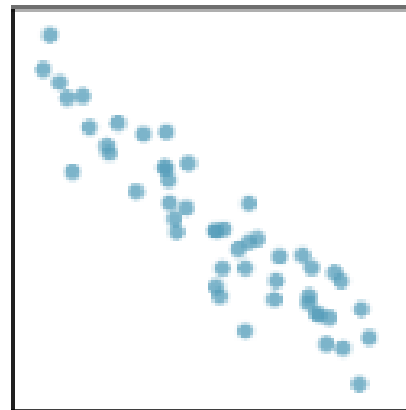
$R = 1.00$



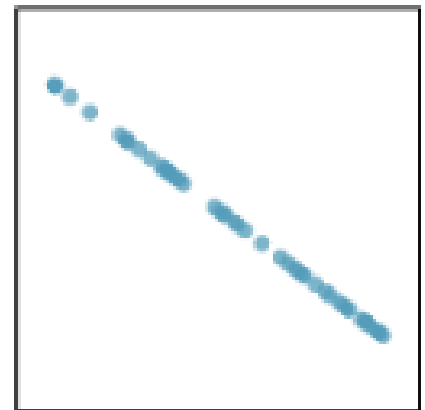
$R = -0.08$



$R = -0.64$



$R = -0.92$



$R = -1.00$

$r = 1,2??$ o $-1,2 ???$

En nuestro caso tenemos:

i	X	Y	XY	X ²	Y ²
1	1	2	2	1	4
2	1	3	3	1	9
3	3	3	9	9	9
4	4	5	20	16	25
5	4	2	8	16	4
6	5	6	30	25	36
Total	18	21	72	68	87
	$\sum x$	$\sum y$	$\sum xy$	$\sum x^2$	$\sum y^2$

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{(n \sum x^2 - (\sum x)^2)(n \sum y^2 - (\sum y)^2)}} = \frac{6 * 72 - 18 * 21}{\sqrt{(6 * 68 - 18^2)(6 * 87 - 21^2)}} = \mathbf{0.6547}$$

Coeficiente de Correlación



Tarea 4

- 1.- Construir una tabla de contingencia de 5 x 4, donde las variables sean independientes.
- 2.- Ejercicio de aplicación del coeficiente de correlación de Sperman. (Variables Categóricas) considerando una muestra de tamaño $n = 50$.

Regresión Lineal

Consiste en ajustar a los datos (representados en el diagrama de Dispersión, una línea, que puede ser recta o curvilínea .

En esta oportunidad analizaremos el caso de la línea recta. Esta recta, también sirve para marcar la tendencia De los datos, para hacer proyecciones, y para estimar algún valor de **y** dado un valor de **x**.

$$\hat{Y} = a + bX$$

Error:

$$e_i = y_i - \hat{y}_i$$

$$\sum e_i = \sum (y_i - \hat{y}_i) = 0$$

Varianza del Error: (σ^2)

cuyo estimador está dado por:

$$S_e^2 = S_{y/x}^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n - 2} \longrightarrow 0$$

Debemos minimizar la varianza del Error
Para tal efecto debemos minimizar:

$$A = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - a - bx_i)^2$$

Es decir derivar A e igualar a cero

Sistema de Ecuaciones Normales

$$\frac{\partial A}{\partial a} = \mathbf{O}$$
$$\frac{\partial A}{\partial b} = \mathbf{O}$$

Luego tenemos:

$$\begin{aligned}\frac{\partial A}{\partial a} &= 2 \sum (y_i - a - bx_i)(-1) = 0 \\ &= na + b \sum x_i = \sum y_i\end{aligned}$$

$$\begin{aligned}\frac{\partial A}{\partial b} &= 2 \sum (y_i - a - bx_i)(-x_i) = 0 \\ &a \sum x_i + b \sum x_i^2 = \sum x_i y_i\end{aligned}$$

Por lo tanto, el sistema de **Ecuaciones Normales** queda de la siguiente forma:

$$na + b \sum x_i = \sum y$$

$$a \sum x_i + b \sum x_i^2 = \sum xy$$

Por determinante tenemos;

$$\Delta = \begin{vmatrix} n & \sum x \\ \sum x & \sum x^2 \end{vmatrix} = n \sum x^2 - (\sum x)^2$$

$$\Delta_a = \begin{vmatrix} \sum y & \sum x \\ \sum xy & \sum x^2 \end{vmatrix} = \sum y \sum x^2 - \sum x \sum xy$$

$$a = \frac{\Delta_a}{\Delta}$$

$$\Delta_b = \begin{vmatrix} n & \sum y \\ \sum x & \sum xy \end{vmatrix} = n \sum xy - \sum x \sum y$$

$$b = \frac{\Delta_b}{\Delta}$$

Estimadores Mínimos Cuadrado

Los que dan origen a la recta de los mínimos cuadrados, o de varianza mínima.

$$a = \frac{\sum y \sum x^2 - \sum x \sum xy}{n \sum x^2 - (\sum x)^2} = \bar{Y} - b\bar{X}$$

$$b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

En el caso que estamos analizando tenemos:

						(e)	(e ²)	
i	X	Y	XY	X ²	Y ²	\hat{Y}	$Y - \hat{Y}$	$(Y - \hat{Y})^2$
1	1	2	2	1	4	2.2	-0.2	0.04
2	1	3	3	1	9	2.2	0.8	0.64
3	3	3	9	9	9	3.5	-0.5	0.25
4	4	5	20	16	25	4.1	0.9	0.81
5	4	2	8	16	4	4.1	-2.1	4.41
6	5	6	30	25	36	4.8	1.2	1.44
Total	18	21	72	68	87		0.0	7.59

Luego tenemos que;

$$a = \frac{(21)(68) - (18)(72)}{n(68) - (18)^2} = 1.57 \quad b = \frac{6(72) - (18)(21)}{n(68) - (18)^2} = 0.64$$

y la varianza del error es

$$S_{y/x}^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n - 2} = \frac{7.59}{4} = 1.898$$

$$\hat{y} = 1.57 + 0.64x$$



Error Típico: (σ)

$$S_{y/x} = \sqrt{S_e^2} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}} = \sqrt{\frac{7.59}{4}} = \sqrt{1.898} = 1.38$$

$$\hat{y} = a + bx \pm S_{y/x} \qquad \hat{y} = 1.57 + 0.64x \pm 1,38$$

También se puede obtener a partir de:

$$S_e = S_{y/x} = \sqrt{\frac{\sum y^2 - a \sum y - b \sum xy}{n-2}}$$



Coeficiente de Determinación

El coeficiente de Determinación, nos indica la variabilidad explicada por la recta de regresión lineal, es decir que tan bueno es el ajuste de la recta.

Esta dado por: r^2 $0 \leq r^2 \leq 1$

En nuestro caso $r = 0.6547$ (65.47%)

Luego, el r^2 , será 0.4286 (42.86%), es decir esta será la variabilidad explicada por la recta de regresión.

Nota: Referencia para el ajuste, también lo da el error típico $S_{y/x}$



Ejemplo:

Supongamos que tenemos dos variables: **X : Años de servicio de vendedores.**
Y : Ventas en M\$

1. **Confeccione un diagrama de dispersión**
2. **Determine el grado de asociación entre las variables**
3. **Estime a través de una recta de m. c. ¿cuanto debiera vender un vendedor con siete años de servicio?.**
4. **Estime a través de una recta de m. c. ¿cuanto años de servicio debería tener, un vendedor que vende m\$ 4.0?.**
5. **Que tan bueno es el ajuste de la regresión?**

Vendedor	X	Y	XY	X²	Y²
1	2	1.2	2.4	4	1.44
2	2	2.4	4.8	4	5.76
3	3	4.1	12.3	9	16.81
4	4	3.1	12.4	16	9.61
5	4	2.4	9.6	16	5.76
6	5	3.8	19.0	25	14.44
Total	20	17.0	60.5	74	53.82



Luego:

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{(n \sum x^2 - (\sum x)^2)(n \sum y^2 - (\sum y)^2)}} = .5954$$

$$a = \frac{\sum y \sum x^2 - \sum x \sum xy}{n \sum x^2 - (\sum x)^2} = \frac{(17)(74) - (20)(60.5)}{6(74) - (20)^2} = 1.09$$

$$b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} = \frac{6(60.5) - (20)(17)}{6(74) - (20)^2} = .523$$

$$\hat{y} = 1.09 + 0.523x \pm s_{y/x}$$



Varianza del Error:

$$S_{y/x}^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n-2} = \frac{\sum y^2 - a \sum y - b \sum xy}{n-2} = .912$$

Error Típico:

$$S_{y/x} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}} = \sqrt{\frac{\sum y^2 - a \sum y - b \sum xy}{n-2}} = .955$$

$$\hat{y} = 1.09 + 0.523x \pm 0.955$$



Luego:

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{(n \sum x^2 - (\sum x)^2)(n \sum y^2 - (\sum y)^2)}} = 0.5954$$

$$c = \frac{\sum x \sum y^2 - \sum y \sum xy}{n \sum y^2 - (\sum y)^2} = 1.41$$

$$d = \frac{n \sum xy - \sum x \sum y}{n \sum y^2 - (\sum y)^2} = 0.68$$

$$S_{x/y} = \sqrt{\frac{\sum x^2 - c \sum x - d \sum xy}{n - 2}} = 1.079$$



- **Las propiedades de la media son las siguientes:**
 - - **La Media** de una constante es la propia constante.
 - - La media de la suma o diferencia de variables es igual a la suma o diferencia de las medias de dichas variables.
 - - La media del producto de una constante por una variable, es igual a la constante por la media de la variable.
 - - La media de una combinación lineal de dos o más variables es igual a la combinación lineal de las medias de dichas variables.
 - - La media es el centro de gravedad de la distribución, ya que las desviaciones respecto a la media suman 0.
- - **Mediana:** La mediana es el valor del elemento que ocupa el lugar central, si los datos están ordenados, bien de forma creciente o de forma decreciente.
- - **Moda:** La moda es el valor más frecuente, es decir es el valor de la variable que se repite un mayor número de veces.
- En el caso de una distribución totalmente simétrica, la media y la mediana coinciden. Si la media y la mediana difieren mucho significa que hay heterogeneidad entre los datos y que la distribución, por tanto será asimétrica.

◆ **Las propiedades de la varianza son:**

- ◆- La varianza es siempre positiva o cero.
- ◆- La varianza de una constante es cero.
- ◆- La varianza de la suma o diferencia de una variable y una constante es igual a la varianza de la variable.
- ◆- La varianza de un producto de una constante por una variable es igual al cuadrado de la constante por la varianza de la variable.

◆ **Las propiedades de la desviación típica son:**

- ◆- La desviación típica es siempre positiva o cero.
- ◆- La desviación típica de una constante es cero.
- ◆- La desviación típica de una constante por una variable es igual a la constante por la desviación típica de la variable.
- ◆- La desviación típica de la suma o diferencia de una variable y una constante es igual a la desviación típica de la variable.

Ejercicio: Se toma una muestra de 112 turistas registrando el peso de equipaje y nacionalidad.

X \ Y	Chilena	Argentina	Peruana	Brasileña	Total
5 – 10	4	2	4	10	20
10 – 15	8	3	1	12	24
15 – 20	3	9	3	10	25
20 – 25	0	10	1	10	21
25 - 30	5	2	0	15	22
Total	20	26	9	57	112

Determine:

- 1.- Cuántos turistas chilenos app llevan un peso no superior a 12kr.
- 2.- ¿Podemos decir que el comportamiento del peso que llevan los turistas argentinos , es mas homogéneo del que llevan los brasileiros?
- 3.- Determine e interprete: $f_{j=1/i=2} =$
- 4.- Determine el grado de asociación respecto al peso de equipaje, entre los turistas chilenos y argentinos.
- 5.- A través de una recta de m. c. estime cuántos turistas deberían llevar un peso 27.5kr

Cuartiles

$$K_k = Li + \left[\frac{k * n}{4} - N_{g-1} \right] * \frac{a}{n_g}$$



Quintiles

$$Q_q = Li + \left[\frac{q^* n}{5} - N_{g-1} \right] * \frac{a}{n_g}$$



Deciles

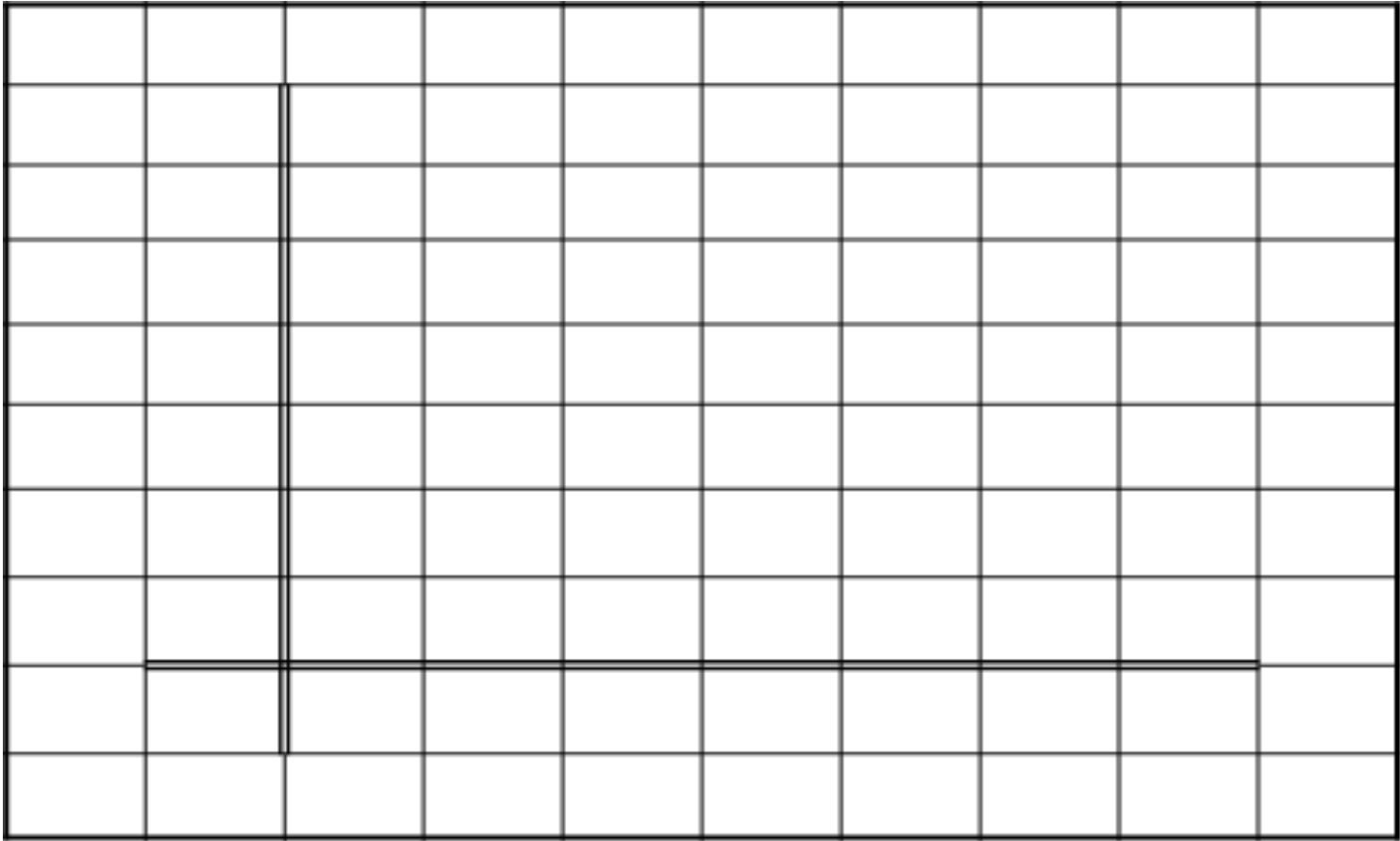
$$D_d = Li + \left[\frac{d * n}{10} - N_{g-1} \right] * \frac{a}{n_g}$$



Percentiles

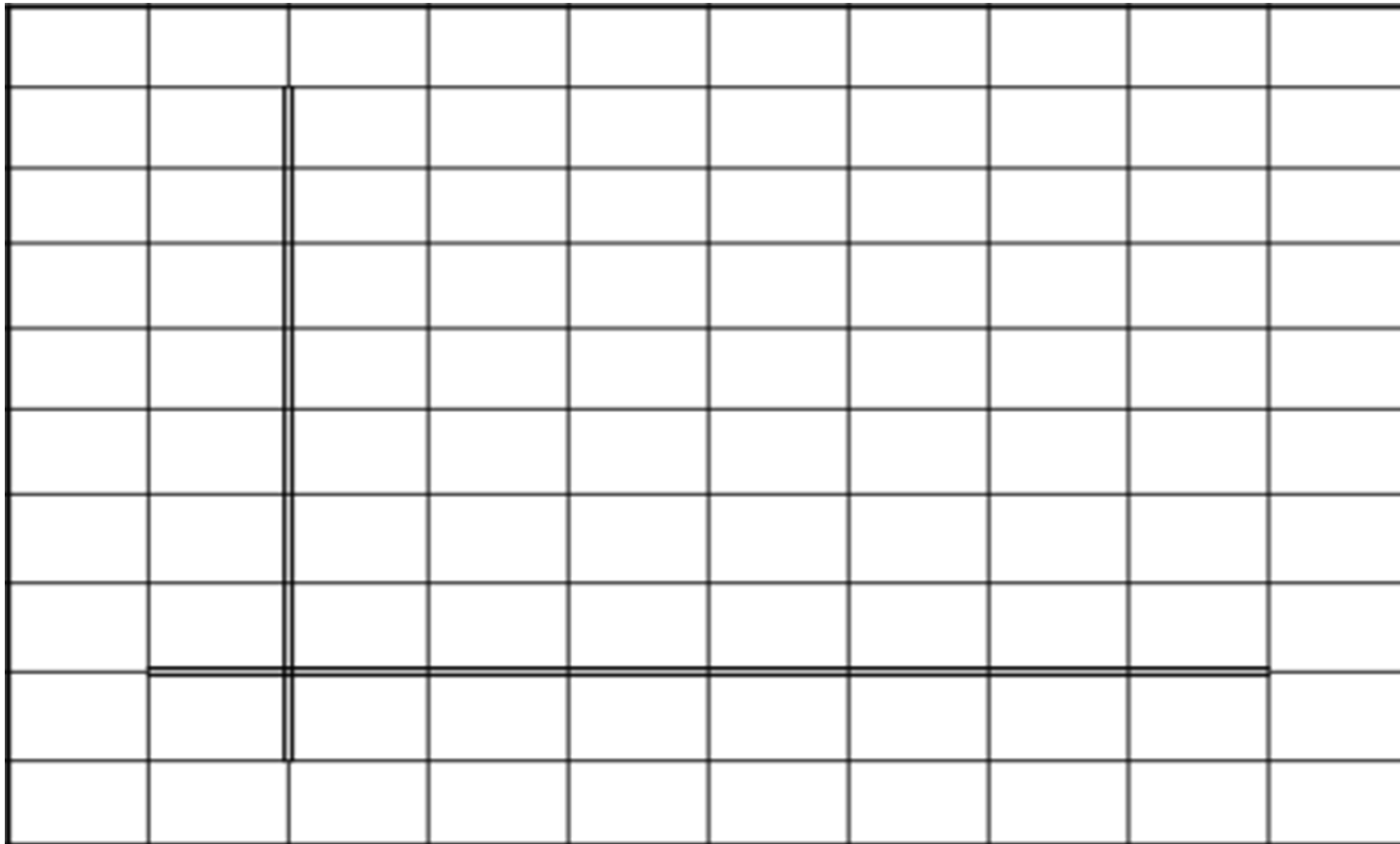
$$P_p = Li + \left[\frac{p * n}{100} - N_{g-1} \right] * \frac{a}{n_g}$$





X	\hat{Y}
1	2.2
1	2.2
3	3.5
4	4.1
4	4.1
5	4.8





X	Y
2	1.2
2	2.4
3	4.1
4	3.1
4	2.4
5	3.8

2017.0



UNIVERSIDAD TECNICA FEDERICO SANTA MARIA
COMPROBANTE FERIADO

Yo **David Becerra Rojas** Rut 7.242.213 - 9 Unidad **Electro e Inform** declaro haber hecho uso del derecho a feriado anual remunerado correspondiente al período 2020 -2021

DETALLE:

Total, día base 25 hábiles

Feriado Progresivo 00 día (s) hábiles

Total, días 25 hábil (es)

HACE USO DEL FERIADO COLECTIVO:

Del _01_ de Febrero de 2021

Al _26_ de Febrero de 2021

Hará uso del Feriado colectivo: 12 al 19 de Julio de 2021.

Total, días de feriado efectivo 25.

Días pendientes a fijar con la autoridad _____ días hábiles

Días Fer. Prog. A fijar con la autoridad: 00 días hábile

Firma Empleador

Firma trabajador

Organización xxxxxx

Observación:

Corrección Prueba 1

Dos compañías A y B fabrican el mismo tipo de cable. Un distribuidor desea conocer la diferencia promedio de la resistencia a la rotura de estos, para lo cual toma muestras de 100 cables de A y 50 cables de B.

La muestra de los cables de la compañía A arroja una resistencia promedio a la rotura de 4500 Kilogramos, mientras que los cables de la compañía B arroja una resistencia promedio a la rotura de 4000 kilogramos.

Se sabe, por experiencia, que la desviación típica de la resistencia a la rotura es de 300 kilogramos para la compañía A y de 200 kilogramos para la compañía B.

Se pide estimar, con un nivel de confianza del 95% el intervalo de confianza de la diferencia de medias de la resistencia a la rotura entre los dos cables y concluir cuál de las medias es mayor.

Sabiendo que la resistencia a la rotura se distribuye normalmente para ambas compañías.

Una empresa de fibra óptica evalúa dos formas para proveer internet a zonas rurales respecto al tiempo que demora en minutos en llegar la señal a cada casa, consideradas muestras respectivas, los resultados de las pruebas son los siguientes:

Determine un intervalo de confianza al 90% para comparar la variabilidad de los tiempos de demora.

F1 : n1: 31 s1: 7.071

F2: n2 : 25 s2: 4.899

Se realizó una encuesta nacional a 902 trabajadores de la construcción, para saber si en sus empresas se cumplen las medidas de seguridad laboral. En la encuesta se encontró que 397 trabajadores, estuvieron satisfechas con las medidas de seguridad dispuestas por sus empresas.

El Banco Prestapoco, que cuenta con 64 sucursales en todo el país, ha decidido tomar una muestra de sucursales, tamaño 25, para determinar el monto promedio de la morosidad. De la muestra se obtuvieron los siguientes estadísticos.

- Con un 95% de confianza, podemos decir que el monto promedio de morosidad de banco es \$ 345,000?

Media:354500

Desv.típica:24250

- Con qué confianza aproximadamente, podríamos decir que la media poblacional, se encuentra en el intervalo (346,600 ; 362,400)?

Los siguientes datos corresponden a la tasa de interés de 8 financieras de las 18 que componen el mercado de una cierta ciudad A.

2,05 2,49 1,18 1,95 2,05 2,12 2,04 2,22

Determine un intervalo de confianza del 90% para la desviación típica de la población.

Dos compañías A y B fabrican el mismo tipo de cable.
Un distribuidor desea conocer la diferencia promedio de la resistencia a la rotura de estos, para lo cual toma muestras de 100 cables de A y 120 cables de B.
La muestra de los cables de la compañía A arroja una resistencia promedio a la rotura de 1480 Kilogramos, mientras que los cables de la compañía B arroja una resistencia promedio a la rotura de 1500 kilogramos.
Se sabe, por experiencia, que la desviación típica de la resistencia a la rotura es de 120 kilogramos para la compañía A y de 100 kilogramos para la compañía B.
Se pide estimar, con un nivel de confianza del 90%, el intervalo de confianza de la diferencia de medias de la resistencia a la rotura entre los dos cables y concluir cuál de las medias es mayor.

sabiendo que la resistencia a la rotura se distribuye normalmente para ambas compañías.

Los siguientes datos corresponden a la tasa de interés de 10 financieras de las 20 que componen el mercado de una cierta ciudad A.

1,05 2,12 2,04 4,05 3,98 1,95 1,18 2,43 2,22 3,75

Determine un intervalo de confianza del 90% para la **desviación típica** de la población.

El consumo de combustible por faena, se ha medido en una muestra de 144 motores. Los datos, se presentan en la siguiente tabla.

i	X	ni
1	2,3 - 2,6	10
2	2,6 - 2,9	18
3	2,9 - 3,2	29
4	3,2 - 3,5	45
5	3,5 - 3,8	27
6	3,8 - 4,1	15
Total		144

Determine:

Determine el coeficiente de Asimetría de Pearson.

Respuesta 1

Cuántos motores aproximadamente tuvieron un consumo de al menos 3,3?

Respuesta 2

La tabla siguiente representa la distribución de las rentas, en Miles de pesos, obtenidos de una muestra de trabajadores de la minería.

i	X	ni
1	501 - 550	45
2	551 - 600	89
3	601 - 650	120
4	651 - 700	105
5	701 - 750	72
6	751 - 800	44
Total		475

Determine:

Cuántos Trabajadores aproximadamente, tienen renta igual o superior a 615?

Elegir...

El coeficiente de Asimetría de Pearson

Elegir...

Una muestra de 321 vehículos es considerada para determinar medir el kilometraje recorrido en un día, según su marca, en las alrededores de un aeropuerto.

A/B	Ford	Chevrolet	Mazda	Citroen	Total
20 - 30	12	18	24	30	84
30 - 40	9	6	4	1	20
40 - 50	14	22	15	10	61
50 - 60	28	10	18	8	64
60 - 70	26	31	20	15	92
Total	89	87	81	64	321

Determine e Interprete;

$f_{j=2/i=4}$

Elegir...

n_{34}

Elegir...

f_{43}

Elegir...

Para realizar un estudio del consumo de electricidad en una determinada región, se toma una muestra de 12 viviendas, registrando el consumo (kilowatt/hora), y

Los datos obtenidos, se presentan en la siguiente tabla.

	kilopwattrra	Integrantes
	Hora	Vivienda
1	456	3
2	582	3
3	430	4
4	659	6
5	850	5
6	734	6
7	449	2
8	790	5
9	830	5
10	480	4
11	845	6
12	736	5
Total	7841	54

Determine;

El grado de asociación entre las variables.

0,7591 (75,91%) ↕

Cual es el erro típico de estimación?

+/- : 113,57 ↕

A través de una recta de mínimos cuadrados (MC), estime, Cuanto consumo debería tener una vivienda con 8 integrantes?

896,7 kilowatt/hora ↕