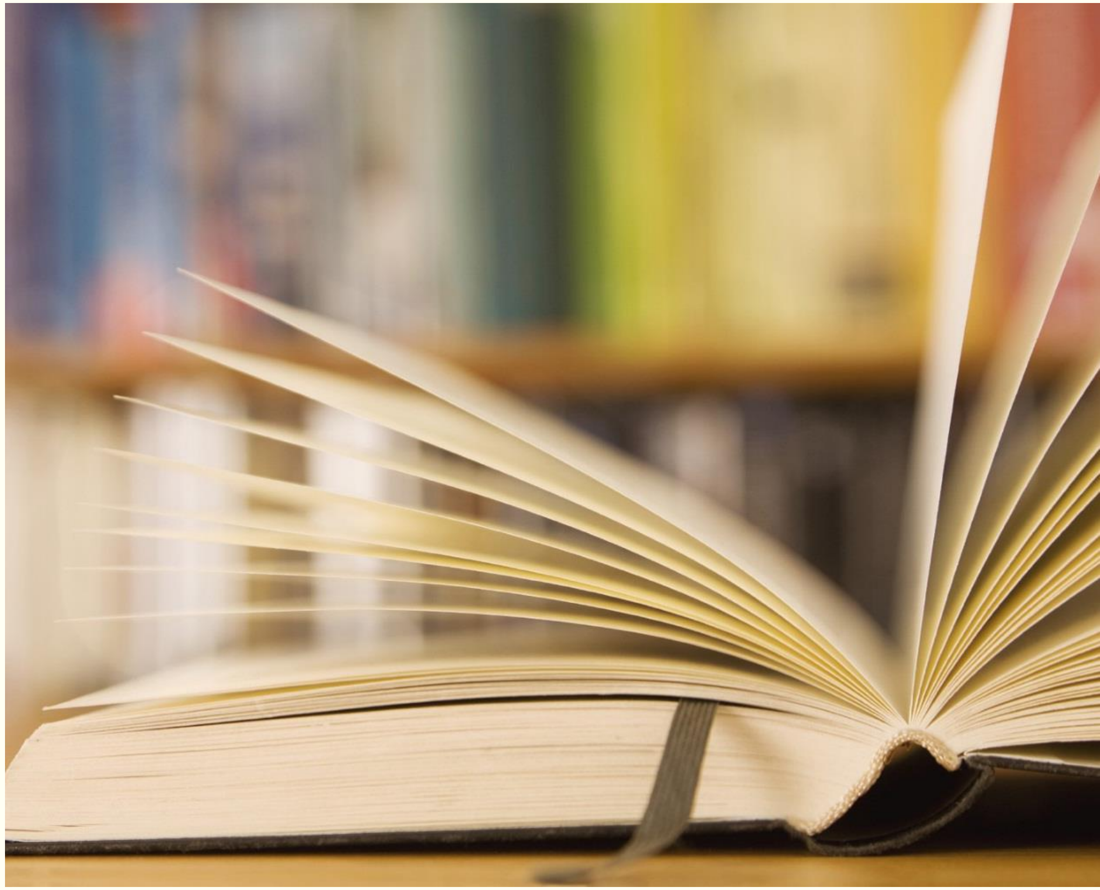
UNIVERSIDAD NACIONAL DE CHIMBORAZO

FACULTAD DE INGENIERÍA

CARRERA DE INGENIERÍA CIVIL



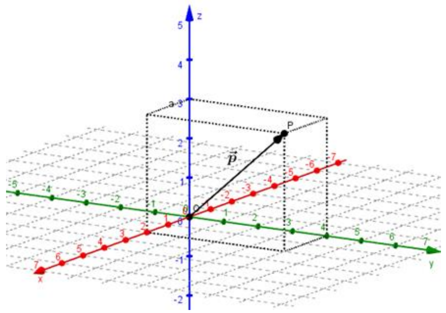
ASIGNATURA: Estadística

DOCENTE: Ing. Lidia Castro M.Sc

UNIDAD 1 → ESTADÍSTICA DESCRIPTIVA

OBJETIVO DE LA UNIDAD

Conocer y comprender el análisis de datos que intervienen en la estadística descriptiva.



1. Introducción
2. Definiciones fundamentales
3. Población y muestra
4. Frecuencia y rango
5. Distribución de frecuencias
6. Medidas de tendencia central
7. Medidas de dispersion

MEDIDAS DE FORMA

Las medidas de forma son indicadores que ofrecen información acerca de la manera en que los datos se encuentran contenidos dentro de una distribución. Se clasifican en dos grupos:

- *Medidas de sesgo o asimetría.*
- *Medidas de apuntamiento o curtosis.*

Sesgo - Asimetría

El sesgo es una medida que indica que tan simétrica o asimétrica es una distribución.

SIMÉTRICA: Es una distribución simétrica si los datos que se encuentran contenidos en la misma se ubican repartidos de forma semejante en ambos lados de la media.

ASIMÉTRICA: Una distribución asimétrica es aquella en la que los datos con las frecuencias más bajas se ubican a la derecha o a la izquierda de la media; por tanto, existen dos tipos de asimetría: la izquierda y la derecha.

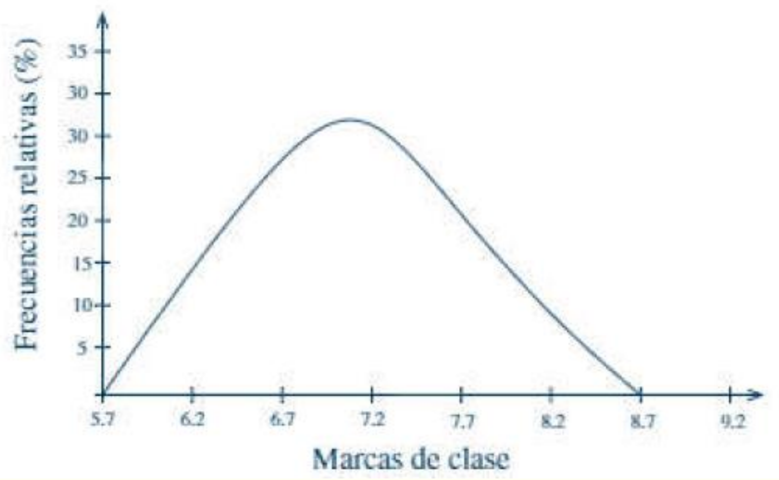
Asimetría

- **Asimetría a la izquierda:** También llamado **sesgo negativo,** es aquella en la que los datos con las frecuencias más bajas se ubican a la izquierda de la media, mientras que los datos con mayores frecuencias tienden a aglomerarse a la derecha.

- **Asimetría a la derecha:** También llamado **sesgo positivo,** es aquella en la que los daros con las menores frecuencias están ubicados a la derecha de la media y los dato con mayores frecuencias se encuentran localizados a la izquierda.

Ejemplo 1

Intervalo de clase	Marca de clase	Frecuencia de clase	Frecuencia relativa de clase (FRC %)
6.0 - 6.4	6.2	7	16%
6.5 - 6.9	6.7	11	24%
7.0 - 7.4	7.2	13	29%
7.5 - 7.9	7.7	8	18%
8.0 - 8.4	8.2	5	11%
8.5 - 8.9	8.7	1	2%



$$\bar{x} = 7,2$$

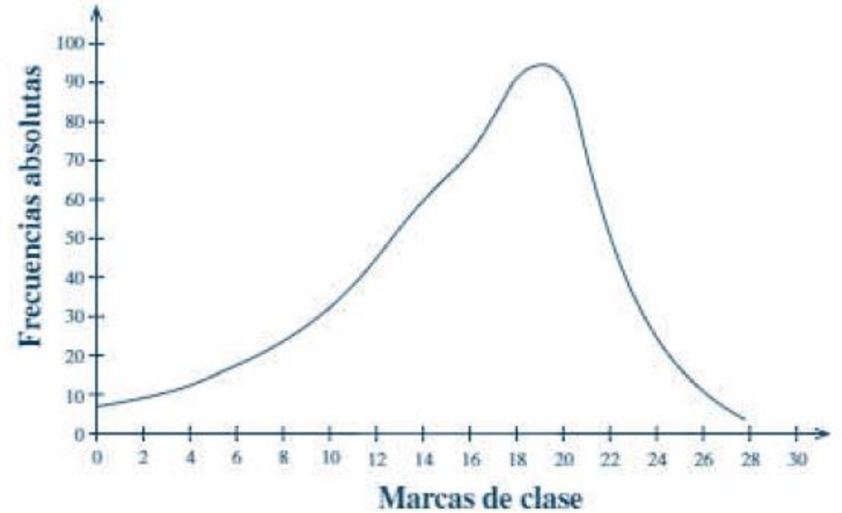
$$\tilde{x} = 7,2$$

$$x_m = 7,2$$

Si la media, mediana y moda son iguales los datos son simétricos

Ejemplo 2

Número de horas de estudio a la semana	Frecuencias absolutas
0	7
2	9
4	12
6	17
8	23
10	32
12	44
14	59
16	71
18	90
20	92
22	54
24	25
26	12
28	3



$$\bar{x} = 16$$

$$\tilde{x} = 18$$

$$x_m = 20$$

El valor de la media es 16, es posible observar los valores con menores frecuencias en el lado izquierdo, por debajo de la media. *La distribución es asimétrica a la izquierda o con sesgo negativo.*

Coeficiente de Fisher

Es una herramienta que sirve para determinar la simetría o la asimetría de una distribución:

$$S_f = \frac{\left(\frac{1}{N}\right)\sum_{i=1}^{k} n_i \times (x_i - \bar{x})^3}{S^3}$$

Donde

N: Número total de datos

$\bar{x}$: La media aritmética

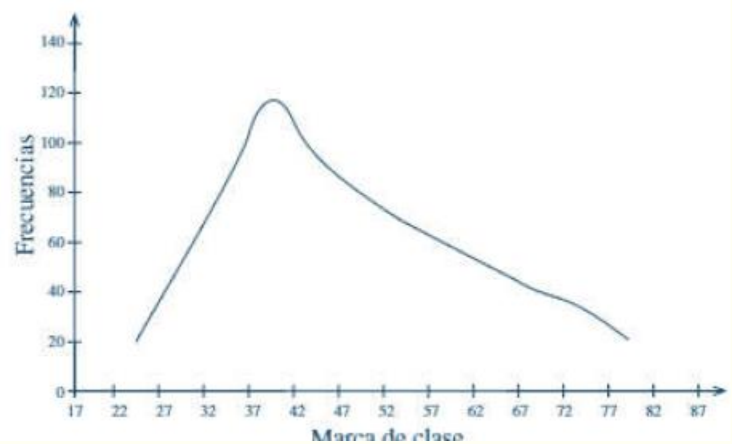
S: Desviación estándar

El coeficiente de Fisher cumple con las siguientes características.

1. Si $S_f = 0$ significa que la distribución es simétrica.
2. Si $S_f > 0$ significa que la distribución tiene sesgo positivo.
3. Si $S_f < 0$ significa que la distribución tiene sesgo negativo.

Ejemplo 3

Edad	Marca de clase	Frecuencia
20 - 24	22	21
25 - 29	27	53
30 - 34	32	84
35 - 39	37	117
40 - 44	42	95
45 - 49	47	80
50 - 54	52	68
55 - 59	57	59
60 - 64	62	49
65 - 69	67	40
70 - 74	72	33
75 - 79	77	21



$$\bar{x} = 46,21$$

$$\tilde{x} = 44,7$$

$$x_m = 37$$

El valor de la media es 42,208 es posible observar los valores con menores frecuencias en el lado derecho, por encima de la media. ***La distribución es asimétrica a la derecha o con sesgo positivo.***

Ejemplo 3

Intervalo	Marca de clase	Frecuencia absoluta	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	$(x_i-\bar{x})^3$	$f(x_i-\bar{x})^3$
20 - 24	22	21	–24.208	586.043	–14 187.134	–297 929.814
25 - 29	27	53	–19.208	368.960	–7 087.108	–375 616.722
30 - 34	32	84	–14.208	201.877	–2 868.332	–240 939.883
35 - 39	37	117	–9.208	84.793	–780.806	–91 354.291
40 - 44	42	95	–4.208	17.710	–74.530	–7 080.338
45 - 49	47	80	0.792	0.627	0.496	39.693
50 - 54	52	68	5.792	33.543	194.272	13 210.510
55 - 59	57	59	10.792	116.460	1 256.798	74 151.097
60 - 64	62	49	15.792	249.377	3 938.074	192 965.641
65 - 69	67	40	20.792	432.293	8 988.100	359 524.015
70 - 74	72	33	25.792	665.210	17 156.876	566 176.923
75 - 79	77	21	30.792	948.127	29 194.403	613 082.453
Suma				3 705.021		806 229.285

$$\bar{x} = 46,21$$

$$S = \sqrt{\frac{\sum(x-\bar{x})^2}{N-1}} = 2,27$$

$$S_f = \frac{\left(\frac{1}{N}\right)\sum_{i=1}^{k} n_i \times (x_i-\bar{x})^3}{S^3}$$

$$S_f = \frac{\left(\frac{1}{720}\right)(806\,229,285)}{2,27^3}$$

$$S_f = 95,73$$

El coeficiente de Fisher cumple con las siguientes características.

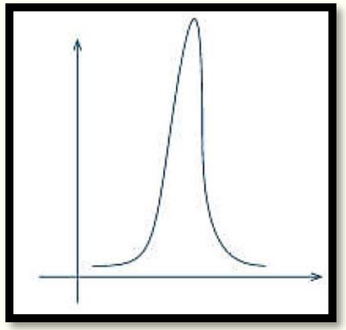
1. Si $S_f = 0$ significa que la distribución es simétrica.
2. Si $S_f > 0$ significa que la distribución tiene sesgo positivo.
3. Si $S_f < 0$ significa que la distribución tiene sesgo negativo.

Apuntalamiento o curtosis

Mide qué tan achatada o tan puntiaguda es una distribución. Así la curtosis indica cuan conglomerados se encuentran los datos en la zona central de la distribución.

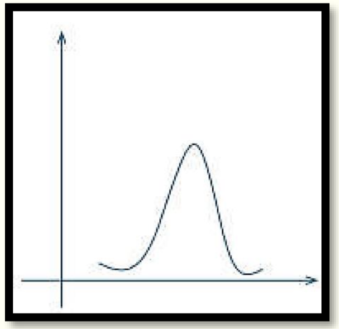
Distribución Leptocúrtica

La forma de su polígono es muy puntiaguda.



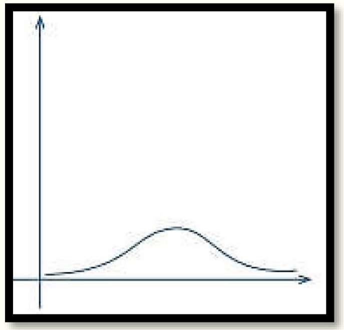
Distribución Mesocúrtica

La forma de su polígono de frecuencias es más achatada que la anterior.



Distribución Platicúrtica

Un polígono de frecuencias muy achatado.



Apuntalamiento o curtosis

$$A_f = \frac{\left(\frac{1}{N}\right)\sum_{i=1}^{k} n_i \times (x-\bar{x})^4}{S^4}$$

Las características de este coeficiente son las siguientes.

1. Si $A_f = 3$ significa que la distribución es mesocúrtica
2. Si $A_f < 3$ significa que la distribución es platicúrtica
3. Si $A_f > 3$ significa que la distribución es leptocúrtica.

Ejemplo 4

Intervalo	Marca de clase	Frecuencia absoluta	$(x_i - \bar{x})^4$	$f(x_i - \bar{x})^4$
20 - 24	22	21	343 446.868	7 212 384.229
25 - 29	27	53	136 131.532	7 214 971.191
30 - 34	32	84	40 754.216	3 423 354.161
35 - 39	37	117	7 189.921	841 220.7629
40 - 44	42	95	313.647	29 796.42223
45 - 49	47	80	0.393	31.42385753
50 - 54	52	68	1 125.160	76 510.87292
55 - 59	57	59	13 562.948	800 213.9286
60 - 64	62	49	62 188.757	3 047 249.095
65 - 69	67	40	186 877.587	7 475 103.491
70 - 74	72	33	442 504.439	14 602 646.48
75 - 79	77	21	898 944.312	18 877 830.54
Suma				63 601 312.6

$$A_f = \frac{\left(\frac{1}{N}\right)\sum_{i=1}^{k} n_i \times (x - \bar{x})^4}{S^4}$$

$$A_f = \frac{\left(\frac{1}{720}\right) 63\,601\,312{,}6}{2{,}27^4}$$

$$\boldsymbol{A_f = 3336,27}$$

Las características de este coeficiente son las siguientes.

1. Si $A_f = 3$ significa que la distribución es mesocúrtica
2. Si $A_f < 3$ significa que la distribución es platicúrtica
3. $\boxed{\text{Si } A_f > 3 \text{ significa que la distribución es leptocúrtica.}}$

¡GRACIAS!