CBS
Colegio Bautista Shalom

Estadística I
Cuarto BACO PFS
Tercer Bimestre

Contenidos

MEDIDAS DE POSICIÓN NO CENTRAL

- ✓ CUARTILES.
- ✓ DECILES.
- ✓ PERCENTILES.
- ✓ SESGO.
- ✓ LA CURTOSIS.
  - EXCESO DE CURTOSIS.

NOTA: conforme avances en tu aprendizaje tu catedrático(a) te indicará la actividad o ejercicio a realizar. Sigue sus instrucciones.

MEDIDAS DE POSICIÓN NO CENTRAL

Las medidas de posición no centrales permiten conocer otros puntos característicos de la distribución que no son los valores centrales. Entre otros indicadores, se suelen utilizar una serie de valores que dividen la muestra en tramos iguales:

1. **Cuartiles:** son 3 valores que distribuyen la serie de datos, ordenada de forma creciente o decreciente, en cuatro tramos iguales, en los que cada uno de ellos concentra el 25% de los resultados.

2. **Deciles:** son 9 valores que distribuyen la serie de datos, ordenada de forma creciente o decreciente, en diez tramos iguales, en los que cada uno de ellos concentra el 10% de los resultados.

3. **Percentiles:** son 99 valores que distribuyen la serie de datos, ordenada de forma creciente o decreciente, en cien tramos iguales, en los que cada uno de ellos concentra el 1% de los resultados.

CUARTILES

Los cuartiles son los tres valores de la variable que dividen a un conjunto de datos ordenados en cuatro partes iguales.

Q_1, Q_2 y Q_3 determinan los valores correspondientes al 25%, al 50% y al 75% de los datos.

Q_2 coincide con la mediana.

Cálculo de los cuartiles:

1. Ordenamos los datos de menor a mayor.
2. Buscamos el lugar que ocupa cada cuartil mediante la expresión $\frac{k \cdot N}{4},\ k = 1, 2, 3$.

Número impar de datos: 2, 5, 3, 6, 7, 4, 9

$$\begin{array}{ccccccc} 2, & 3, & 4, & 5, & 6, & 7, & 9 \\ & \downarrow & & \downarrow & & \downarrow & \\ & Q_1 & & Q_2 & & Q_3 & \end{array}$$

Número par de datos: 2, 5, 3, 4, 6, 7, 1, 9

$$\begin{array}{ccccc} 1, & \underline{2,\ 3}, & \underline{4,\ 5}, & \underline{6,\ 7}, & 9 \\ & 2.5 & 4.5 & 6.5 & \\ & \downarrow & \downarrow & \downarrow & \\ & Q_1 & Q_2 & Q_3 & \end{array}$$

EJERCICIO DE CUARTILES.

Cálculo de los cuartiles para datos agrupados:

	f_i	F_i
[50, 60)	**8**	**8**
[60, 70)	**10**	**18**
[70, 80)	**16**	**34**
[80, 90)	**14**	**48**
[90, 100)	**10**	**58**
[100, 110)	**5**	**63**
[110, 120)	**2**	**65**
	65	

En primer lugar, buscamos la clase donde se encuentra $\frac{k \cdot N}{4}$, $k = 1, 2, 3$, en la tabla de las frecuencias acumuladas.

$$Q_k = L_i + \frac{\frac{k \cdot N}{4} - F_{i-1}}{f_i} \cdot a_i \qquad k = 1, 2, 3$$

Calcular los cuartiles de la distribución de la tabla:

Cálculo del primer cuartil:

$$\frac{65 \cdot 1}{4} = 16.25 \qquad Q_1 = 60 + \frac{16.25 - 8}{10} \cdot 10 = 68.25$$

Cálculo del segundo cuartil:

$$\frac{65 \cdot 2}{4} = 32.5 \qquad Q_2 = 70 + \frac{32.5 - 18}{16} \cdot 10 = 79.0625$$

Cálculo del tercer cuartil:

$$\frac{65 \cdot 3}{4} = 48.75 \qquad Q_3 = 90 + \frac{48.75 - 48}{10} \cdot 10 = 90.75$$

DECILES

Los deciles son los nueve valores que dividen la serie de datos en diez partes iguales.

Los deciles dan los valores correspondientes al 10%, al 20%... y al 90% de los datos.

D_5 coincide con la mediana.

Cálculo de los deciles:

En primer lugar, buscamos la clase donde se encuentra $\frac{k \cdot N}{10}$, $k = 1, 2, ..9$, en la tabla de las frecuencias acumuladas.

$$D_k = L_i + \frac{\frac{k \cdot N}{10} - F_{i-1}}{f_i} \cdot a_i \qquad k = 1, 2, ..9$$

EJERCICIO DE DECILES.

Calcular los deciles de la distribución de la tabla:

	F_i	F_i
[50, 60)	**8**	**8**
[60, 70)	**10**	**18**
[70, 80)	**16**	**34**
[80, 90)	**14**	**48**
[90, 100)	**10**	**58**
[100, 110)	**5**	**63**
[110, 120)	**2**	**65**
	65	

Cálculo del primer decil:

$$\frac{65 \cdot 1}{10} = 6.5 \qquad D_1 = 50 + \frac{6.5 - 0}{8} \cdot 10 = 58.12$$

Cálculo del segundo decil:

$$\frac{65 \cdot 2}{10} = 13 \qquad D_2 = 60 + \frac{13 - 8}{10} \cdot 10 = 65$$

Cálculo del tercer decil:

$$\frac{65 \cdot 3}{10} = 19.5 \qquad D_3 = 70 + \frac{19.5 - 18}{16} \cdot 10 = 70.94$$

Cálculo del cuarto decil:

$$\frac{65 \cdot 4}{10} = 26 \qquad D_4 = 70 + \frac{26 - 18}{16} \cdot 10 = 75$$

Cálculo del quinto decil:

$$\frac{65 \cdot 5}{10} = 32.5 \qquad D_5 = 70 + \frac{32.5 - 18}{16} \cdot 10 = 79.06$$

Cálculo del sexto decil:

$$\frac{65 \cdot 6}{10} = 39 \qquad D_6 = 80 + \frac{39 - 34}{14} \cdot 10 = 83.57$$

Cálculo del séptimo decil:

$$\frac{65 \cdot 7}{10} = 45.5 \qquad D_7 = 80 + \frac{45.5 - 34}{14} \cdot 10 = 88.21$$

Cálculo del octavo decil:

$$\frac{65 \cdot 8}{10} = 52 \qquad D_8 = 90 + \frac{52 - 48}{10} \cdot 10 = 94$$

Cálculo del noveno decil:

$$\frac{65 \cdot 9}{10} = 58.5 \qquad D_9 = 100 + \frac{58.5 - 58}{5} \cdot 10 = 101$$

PERCENTILES

Los **percentiles** son los **99 valores** que **dividen** la serie de **datos** en **100 partes iguales**.

Los **percentiles** dan los valores correspondientes al 1%, al 2%... y al 99% de los datos.

P_{50} coincide con la **mediana**.

Cálculo de los percentiles:

En primer lugar, buscamos la clase donde se encuentra $\frac{k \cdot N}{100}, \quad k = 1, 2, ..99$, en la tabla de las frecuencias acumuladas.

$$P_k = L_i + \frac{\frac{k \cdot N}{100} - F_{i-1}}{f_i} \cdot a_i \qquad\qquad k = 1, 2, ..99$$

Ejercicio de percentiles:

Calcular el percentil 35 y 60 de la distribución de la tabla:

	f_i	F_i
[50, 60)	**8**	**8**
[60, 70)	**10**	**18**
[70, 80)	**16**	**34**
[80, 90)	**14**	**48**
[90, 100)	**10**	**58**
[100, 110)	**5**	**63**
[110, 120)	**2**	**65**
	65	

Percentil 35:

$$\frac{65 \cdot 35}{100} = 22.75 \qquad P_{35} = 70 + \frac{22.75 - 18}{16} \cdot 10 = 72.97$$

Percentil 60:

$$\frac{65 \cdot 60}{100} = 39 \qquad P_{60} = 80 + \frac{39 - 34}{14} \cdot 10 = 83.57$$

SESGO

Se puede decir que es como un error que aparece en dicho resultado de alguna investigación esto puede deberse a los factores que dependen de la recolección de datos que nos podrían conducir a conclusiones que pueden ser verdaderas o falsas de lo podríamos llamar la realidad.

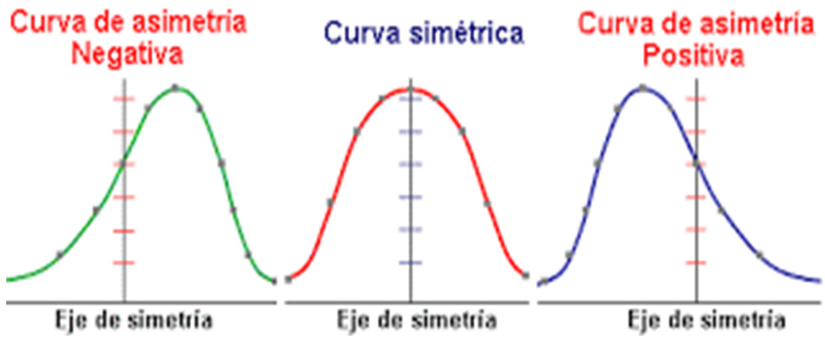
Asimetría: esta medida nos permite identificar si los datos que se están analizando o investigando se distribuyen de alguna forma uniforme o con cabalidad, existen tres tipos de estado las cuales pueden ser:

1. **Asimetría positiva:** se dice que esta ocurre cuando la mayoría de los datos recolectados se encuentran por encima del valor de la media aritmética.

2. **Simetría:** esta sucede cuando los datos recolectados se distribuyen de una forma igual de ambos lados, ósea que aproximadamente quedan con los mismos datos de los dos lados con respecto a la media.

3. **Asimetría Negativa:** en este caso es cuando la mayoría de los datos recopilados se juntan o aglomeran en los valores menores que la media.

Ejemplo de los 3 tipos de Asimetría:



LA CURTOSIS

También conocida como medida de apuntamiento. es una medida estadística, que determina el grado de concentración que presentan los valores de una variable alrededor de la zona central de la distribución de frecuencias.

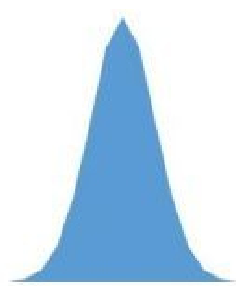
Cuando medimos una variable aleatoria, por lo general, los resultados que tienen una mayor frecuencia son los que se sitúan en torno a la media de la distribución. Imaginemos la altura de los alumnos de una clase. Si la altura media de la clase es 1,72, lo más normal es que las alturas del resto de los alumnos estén en torno a este valor (con cierto grado de variabilidad, pero sin ser esta demasiado grande). Si esto sucede, se considera que la distribución de la variable aleatoria se distribuye con normalidad. Pero dada la infinidad de variables que se pueden medir, esto no siempre sucede así.

Existen algunas variables que presentan un mayor grado de concentración (menor dispersión) de los valores en torno a su media y otras, por el contrario, presentan un menor grado de concentración (mayor dispersión) de sus valores en torno a su valor central. Por tanto, la curtosis nos informa de lo apuntada (mayor concentración) o lo achatada (menor concentración) que es una distribución.

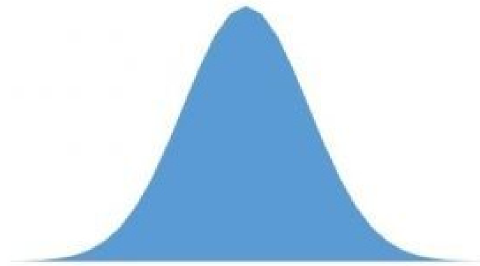
Tipos de curtosis:

Dependiendo del grado de curtosis, tenemos tres tipos de distribuciones:

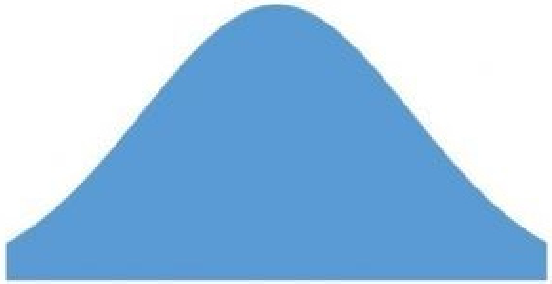
1. Leptocúrtica. Existe una gran concentración de los valores en torno a su media ($g^2>3$)



2. Mesocúrtica. Existe una concentración normal de los valores en torno a su media ($g^2=3$).



3. Platicúrtica. Existe una baja concentración de los valores en torno a su media ($g^2<3$).



Medidas de curtosis según los datos:

Dependiendo de la agrupación o no de los datos, se utiliza una fórmula u otra.

Datos sin agrupar:

$$g_2 = \frac{1}{N}\frac{\sum(x_i - \bar{x})^4}{6^4}$$

Datos agrupados en tablas de frecuencias:

$$g_2 = \frac{1}{N}\frac{\sum f_i(x_i - \bar{x})^4}{6^4}$$

Datos agrupados en intervalos:

$$g_2 = \frac{1}{N}\frac{\sum f_i(mx_i - \bar{x})^4}{6^4}$$

g_2 = Coeficiente de curtosis

N = Número de datos

x_i = Valor i-ésimo de las observaciones

$\bar{x}$ = Media aritmética de la distribución

6 = Desviación típica de la distribución

f_i = Frecuencia absoluta del dato i-ésimo

mx_i = Marca de clase

Ejemplo de cálculo de curtosis para datos sin agrupar:

Supongamos que queremos calcular la curtosis de la siguiente distribución:

8,5,9,10,12,7,2,6,8,9,10,7,7.

Primero calculamos la media aritmética (μ), que sería 7,69.

A continuación, calculamos la desviación típica, que sería 2,43.

Tras tener estos datos y para comodidad en el cálculo, se puede realizar una tabla para calcular la parte del numerador (cuarto momento de la distribución). Para el primer cálculo sería: $(Xi-\mu)^4 = (8-7{,}69)^4 = 0{,}009$.

Datos	$(Xi-\mu)^4$
8	0,0090
5	52,5411
9	2,9243
10	28,3604
12	344,3330
7	0,2297
2	1049,9134
6	8,2020
8	0,0090
9	2,9243
10	28,3604
7	0,2297
7	0,2297
N = 13	∑ = 1.518,27

Una vez tenemos esta tabla hecha, simplemente tendríamos que aplicar la formula expuesta con anterioridad para tener la curtosis.

$g_2 = 1.518{,}27/13*(2{,}43)^4 = 3{,}34$

En este caso dado que g_2 es mayor que 3, la distribución sería leptocurtica, presentando un mayor apuntamiento que la distribución normal.

EXCESO DE CURTOSIS

En algunos manuales la curtosis se presenta como exceso de curtosis. En este caso esta se compara directamente con la de la distribución normal. Dado que la distribución normal, tiene curtosis 3, para obtener el exceso, solo habría restarle 3 a nuestro resultado.

Exceso de curtosis = $g_2-3 = 3{,}34-3 = 0{,}34$.

La interpretación del resultado en este caso sería la siguiente:

$g_2-3 > 0$ -> distribución leptocúrtica.
$g_2-3 = 0$ -> distribución mesocúrtica (o normal).
$g_2-3 < 0$ -> distribución platicúrtica.

Formulas:

Sesgo	$S = \frac{[(Q_3 - \tilde{X}) - (\tilde{X} - Q_1)]}{Q_3 - Q_1}$ $S = \frac{[(83.82 - 81.17) - (81.17 - 75)]}{83.33 - 75}$ $S = 0.399$
Curtosis	$C = \left[\frac{\frac{(Q_3 - Q_1)}{2}}{(P90 - P10)}\right]$ $C = \left[\frac{\frac{(83.82 - 75)}{2}}{(90 - 67.55)}\right]$ $C = 0.20$

Ejemplo:

Notas ordenadas de forma ascendente:

25	37	38	39	39
42	46	47	48	51
54	55	56	58	58
58	60	61	63	63
66	66	67	67	
68	70	73	73	
74	76	76	77	
78	78	78	79	

Cálculos:

79-25 R= 54

Clases o categorías
i= √35 5.92≈ 6

Amplitud
A= R ÷i 9.12≈ 9

Tabla de frecuencias:

X		f	fa	fc	fr	Fr%	LRi-LRs	Xi	F*Xi	Xi- $\bar{X}$	|Xi- $\bar{X}$ |	(Xi- $\bar{X}$) 2	f(Xi- $\bar{X}$) 2
25	33	1	1	35	0.03	2.78	24.5- 33.5	29	29	-30.25	30.25	915.06	915.06
34	42	5	6	30	0.14	13.89	33.5-42.5	38	190	-21.25	21.25	451.56	2257.81
43	51	4	10	26	0.11	11.11	42.5-51.5	47	188	-12.25	12.25	150.06	600.25
52	60	7	17	19	0.19	19.44	51.5-60.5	56	392	-3.25	3.25	10.56	73.94
61	69	8	25	11	0.22	22.22	60.5-69.5	65	520	5.75	5.75	33.06	264.50
70	78	11	36	0	0.31	30.56	69.5-78.5	74	814	14.75	14.75	217.56	2393.19
		36			1.00	100.00			2133		87.5		6504.75

Proceso de fórmulas estadísticas:

Media Aritmética $\bar{X}$	$\bar{X}= \frac{\sum fxXi}{36}$ $\bar{X}$=59.25
Mediana $\tilde{X}$	$\tilde{X} = LR\ inf + \left[\frac{\frac{n}{2} - Fa\ ant}{f}\right] * i$ $Posición = \frac{n}{2} = \frac{36}{2} = 18 \quad "52 - 60"$ $\tilde{X} = 51.5 + \left[\frac{18 - 10}{7}\right] * 9$ $\tilde{X} = 61.79$
Moda $\hat{X}$	$\hat{X} = LR\ inf + \left(\frac{\Delta 1}{\Delta 1 + \Delta 2}\right) * i$ 8, 11, 0 $\quad \Delta 1 = 3, \Delta 2 = 11$ $\hat{X} = 69.5 + \left(\frac{3}{3 + 11}\right) * 9$ $\hat{X} = 71.43$
Cuartiles	$Q_1 = LR \inf + \left[\frac{\left(\frac{nx}{4}\right) - fa\ ant)}{f}\right] * i$ $\frac{nx}{4} = \frac{36 * 1}{4} = 9 \quad "43 - 51"$ $Q_1 = 43.5 + \left[\frac{(9 - 6)}{4}\right] * 9$ Q_1=50.25 $Q_3 = LR \inf + \left[\frac{\left(\frac{nx}{4}\right) - fa\ ant)}{f}\right] * i$ $\frac{nx}{4} = \frac{36 * 3}{4} = 27 \quad "70 - 78"$ $Q_3 = 69.5 + \left[\frac{(27 - 25)}{11}\right] * 9$ Q_3= 71.14

Deciles	$D_5 = LR \inf + \left[\frac{\left(\frac{nx}{10} - fa\ ant\right)}{f}\right] * i$ $\frac{nx}{4} = \frac{(36*5)}{10} = 18 \quad "61-69"$ $D_5 = 60.5 + \left[\frac{(18-17)}{11}\right] * 9$ D_5=61.32
Percentiles	$P_{90} = LR\ inf + \left[\frac{\left(\frac{nx}{100} - fa\ ant\right)}{f}\right] * i$ $\frac{nx}{4} = \frac{(36*90)}{100} = 32.4 \ \backsimeq 32 \quad "70-78"$ $P_{90} = 69.5 + \left[\frac{(32.4-25)}{11}\right] * 9$ P_{90}= 75.55 $P_{10} = LR\ inf + \left[\frac{\left(\frac{nx}{100} - fa\ ant\right)}{f}\right] * i$ $\frac{nx}{4} = \frac{(36*10)}{100} = 3.6 \ \backsimeq 4 \quad "32-42"$ $P_{10} = 33.5 + \left[\frac{(3.6-1)}{5}\right] * 9$ P_{10}= 38.18
Desviación media	$D_M = \sum f \frac{\lvert Xi - \bar{X}\rvert}{n}$ D_M= 10.98
Desviación típica	$D_T = \sqrt{\sum f \frac{(Xi-\bar{X})^2}{n}}$ $D_T = \sqrt{\frac{1803.75}{37}}$ D_T= 6.98

Coeficiente de variación	$\surd = \sum f \frac{(Xi - \bar{X})^2}{n}$ $\surd = \frac{1803.75}{37}$ √= 48.75
Sesgo	$S = \frac{[(Q_3 - \tilde{X}) - (\tilde{X} - Q_1)]}{Q_3 - Q_1}$ $S = \frac{[(83.82 - 81.17) - (81.17 - 75)]}{83.33 - 75}$ $S = 0.399$
Curtosis	$C = \left[\frac{\frac{(Q_3 - Q_1)}{2}}{(P90 - P10)} \right]$ $C = \left[\frac{\frac{(83.82 - 75)}{2}}{(90 - 67.55)} \right]$ $C = 0.20$

ÁREA PRÁCTICA. Deberás realizar los ejercicios que se le presentan de forma digital. La tabla estadística debe ir trabajada en Excel con sus respectivas formulas en cada columna completo como el del ejemplo y enviar al correo que te indique tu catedrático(a).

1.

15	73	1	65	16	3	42
36	42	3	61	19	36	47
30	45	29	73	69	34	23
22	21	33	27	55	58	17
4	17	48	25	36	11	4
54	70	51	3	34	26	10

2.

Calificaciones de 50 estudiantes en un examen				
57	84	68	57	73
76	77	37	80	70
64	29	65	65	54
70	57	42	59	61
62	64	82	51	62
70	87	47	80	59
64	85	80	51	82
60	63	36	56	50
62	73	72	70	59
55	61	65	66	76

3.

Estatura en cm de 50 estudiantes				
137	145	150	156	162
137	146	151	156	162
138	146	152	156	163
140	147	152	157	167
142	147	152	158	167
142	147	153	158	168
143	147	153	159	170
144	147	154	160	155
144	148	155	161	156
145	149	155	161	150

4. Las estaturas en centímetros de 50 estudiantes mujeres, un grupo se registraron. Los datos son:

157	155	171	150	163	150	172	161	154	174
163	148	152	163	149	158	176	164	157	153
169	161	160	164	155	162	151	167	167	167
170	158	163	175	169	169	158	150	156	157
174	162	150	151	165	170	156	170	153	154

5. Se registran los tiempos de las llamadas recibidas en un call center, y se obtiene la siguiente tabla de frecuencias con datos agrupados. Construir un histograma de frecuencias.

Tiempo de llamadas	Marcas de clase	Frecuencia absoluta	Frecuencia acumulada	Frecuencia porcentual
[0 - 10)	5	2	2	5%
[10 - 20)	15	6	8	15%
[20 - 30)	25	12	20	30%
[30 - 40)	35	10	30	25%
[40 - 50)	45	6	36	15%
[50 - 60]	55	4	40	10%
Total		**40**		**100%**