UTN
Facultad Regional Córdoba

TUP

Tecnicatura Universitaria en Programación

ESTADISTICA

Unidad Temática N° 2: Datos Estadísticos

Material Teórico

1° Año – 2° Cuatrimestre

Índice

DATOS ESTADÍSTICOS

Concepto de datos estadísticos

Es la variable bajo estudio.

Son aquellos datos que pueden ser analizados, es posible estudiar su comportamiento, permiten obtener conclusiones.

Siempre presentan variabilidad, es por ello que se los denomina variables. Sus valores varían en el tiempo y o difieren entre los distintos individuos u objetos bajo análisis.

Por ejemplo, la nota de un estudiante no es un dato estadístico, mientras que las notas obtenidas por los estudiantes en cierto examen si constituyen datos estadísticos.

Los valores posibles correspondientes a un dato estadístico son todos los valores que puede asumir una variable.

Los valores observados correspondientes a determinado dato estadístico son aquellos que se presentaron al recolectar los datos, y no siempre abarcan todos los posibles valores.

Clasificación de datos estadísticos

Variables cuantitativas: Están expresadas numéricamente. Surgen de respuestas expresadas en forma numérica.

Variables cuantitativas discretas: Surgen de un proceso de conteo. Ejemplo: cantidad de hijos en un grupo de familias, cantidad de materias aprobadas en un grupo de estudiantes.

No admiten valores con cifras decimales y no tienen unidad medida

Variables cuantitativas continuas: Surgen de un proceso de medición. Ejemplo: peso, estatura

Pueden admitir cifras decimales y/o llevan unidad de medida

Ejemplo:

Estatura de un niño:

1 metro

1,20 mts.

Cantidad de hijos promedio de 5 familias encuestadas: Variable cuantitativa discreta

2 3 2 1 1

Promedio: 9/5= 1,8

2 3 2 1 2

Promedio: 10/5= 2

Todo promedio es una variable cuantitativa continua

Variables categóricas o variables cualitativas o atributos:

Nominales: Las categorías no se diferencias por jerarquías. No existe una categoría mejor o superior que otra. Ejemplo: nacionalidad, género, estado civil

Ordinales: Admiten la clasificación de acuerdo a la jerarquía. Ejemplo: nivel de satisfacción con cierto servicio, nivel de estudios.

Organización y presentación de los datos estadísticos

Formas de presentación de los Datos Estadísticos:

Una vez recolectados, los datos estadísticos deben ser presentados adecuadamente para facilitar su análisis e interpretación.

Series Simples:

Representan los valores observados sin ser agrupados, en consecuencia esta forma de presentación resulta adecuada solo cuando tenemos pocos valores observados.

En una Serie Simple cada valor observado se representa como xi.

i= 1… n

n simboliza a la cantidad total de valores observados

Ejemplo:

La cantidad de hijos de 5 familias encuestadas:

$x_1= 2$ $x_2= 3$ $x_3= 1$ $x_4= 5$ $x_5= 2$

n=5

La primera familia encuestada tiene dos hijos.

La segunda familia encuestada tiene tres hijos.

De manera similar interpretamos los restantes valores observados.

Cuando tenemos gran cantidad de observaciones resulta conveniente agrupar los datos y presentarlos en forma de Tablas (cuadros) y/o a través de Gráficos.

Estructura de los cuadros y gráficos estadísticos

Ejemplo:

En la siguiente serie se representa la cantidad de hijos de 5 familias observadas.

$x_1 = 1$; $x_2 = 2$; $x_3 = 2$; $x_4 = 3$; $x_5 = 4$

Cuando el número de observaciones es grande, resulta conveniente agrupar los valores observados para facilitar su análisis. Estos valores pueden ser agrupados y presentados de manera resumida a través de tablas y/o gráficos

Tablas

Tabla de datos agrupados en lista: Resulta adecuada para agrupar datos cuando al recolectar la información se obtuvieron pocos valores y con mucha repetición.

i	y_i	n_i	$h_i = n_i /n$	$N_i = n_1 + n_2 + \ldots + n_i$	$H_i = N_i /n$
1	y_1	n_1	$h_1 = n_1 /n$	$N_1 = n_1$	$H_1 = N_1 /n$
2	y_2	n_2	$h_2 = n_2 /n$	$N_2 = n_1 + n_2$	$H_2 = N_2 /n$
3	y_3	n_3	$h_3 = n_3 /n$	$N_3 = n_1 + n_2 + n_3$	$H_3 = N_3 /n$
…	…	…	…	…	…
m	y_m	n_m	$h_m = n_m /n$	$N_m = n_1 + n_2 + \ldots + n_m$	$H_m = N_m /n$

i: En la primera columna se enumeran los intervalos desde 1 hasta m.

yi: Clases: Son los diferentes valores observados.

Frecuencia Absoluta Simple (ni):

Cantidad de observaciones que hay en cada clase i.

Frecuencia Relativa Simple (hi):

Proporción de observaciones que hay en cada clase i.

Se obtiene mediante el cociente entre la **Frecuencia Absoluta Simple (ni)** y la cantidad total de observaciones: $h_i = n_i / n$

Frecuencia Absoluta Acumulada (Ni):

Cantidad de observaciones **menores o iguales** que la clase i

Se obtiene sumando las frecuencias absolutas simples desde el primero hasta un intervalo i: $N_i = n_1 + n_2 + \dots + n_i$

Frecuencia Relativa Acumulada (Hi):

Proporción de observaciones **menores o iguales** que la clase i: $H_i = Ni / n$

Tablas de datos agrupados en intervalos: Resulta adecuada para agrupar datos cuando al recolectar la información se obtuvieron muchos valores y con poca o mucha repetición

i	y'_{i-1} y'_i	n_i	$h_i = n_i / n$	$N_i = n_1 + n_2 + \dots + n_i$	$H_i = N_i / n$
1	y'_0 y'_1	n_1	$h_1 = n_1 / n$	$N_1 = n_1$	$H_1 = N_1 / n$
2	y'_1 y'_2	n_2	$h_2 = n_2 / n$	$N_2 = n_1 + n_2$	$H_2 = N_2 / n$
3	y'_2 y'_3	n_3	$h_3 = n_3 / n$	$N_3 = n_1 + n_2 + n_3$	$H_3 = N_3 / n$
…	… …	…	…	…	…
m	y'_{m-1} y'_m	n_m	$h_m = n_m / n$	$N_m = n_1 + n_2 + \dots + n_m$	$H_m = N_m / n$

i: En la primera columna se enumeran los intervalos desde 1 hasta m.

y'_{i-1}: límite inferior del intervalo i

y'_i: límite superior del intervalo i

En la tabla con intervalos también se representarán las siguientes frecuencias:

Frecuencia Absoluta Simple (n_i):

Cantidad de observaciones que hay en cada intervalo i.

Frecuencia Relativa Simple (h_i):

Proporción de observaciones que hay en cada intervalo i

Se obtiene mediante el cociente entre la **Frecuencia Absoluta Simple (ni)** y la cantidad total de observaciones: $h_i = n_i / n$

Frecuencia absoluta acumulada (N_i):

Cantidad de observaciones acumuladas hasta el intervalo i sin incluir el límite superior.

Se obtiene sumando las frecuencias absolutas simples desde el primero hasta un intervalo i: $N_i = n_1 + n_2 + \dots + n_i$

Frecuencia relativa acumulada (Hi):

Proporción de observaciones acumuladas hasta el intervalo i sin incluir el límite superior.

Se obtiene mediante el cociente entre la **Frecuencia absoluta acumulada (N_i)** y la cantidad total de observaciones: $H_i = N_i / n$

A modo de ejemplo se presenta el siguiente caso:
El 29 de Junio de 2021 se encuestaron a los 105 estudiantes de la carrera Licenciatura en Kinesiología de la Universidad de Monterrey, y se obtuvo la siguiente información:

Nro. encuesta	Género	Edad (años)	Gastos generales en educación (pesos mxc)	Cantidad de materias aprobadas	Tiempo semanal dedicado al estudio (horas)	Trabaja	Nivel de Satisf. con la carrera	Nro. de hnos.
1	F	34	30320	4	10	NO	2	2
2	F	19	28760	5	3	SI	1	3
3	F	20	29540	2	15	NO	2	6
4	F	28	11000	3	5	NO	1	4
5	F	25	15400	9	18	NO	1	3
6	M	28	17500	6	12	SI	1	2
7	F	21	10000	2	4	SI	2	6
8	F	26	14900	8	20	NO	1	3
9	F	26	12860	7	7	NO	2	5
10	F	23	19530	8	10	NO	1	4
11	F	24	15590	2	2	SI	1	2
12	M	25	13610	5	11	NO	3	3
13	F	22	18500	9	5	NO	1	2
14	M	31	23410	2	14	NO	3	3
15	M	26	14761	6	3	SI	2	2
16	M	23	15500	8	6	SI	1	2
17	M	24	10152	1	5	SI	1	3
18	M	25	20790	5	4	SI	2	2

19	M	22	21734	8	10	NO	4	2
20	M	25	20200	9	20	NO	1	4
21	F	19	12180	5	7	NO	1	2
22	F	35	13652	3	2	NO	3	2
23	F	23	11370	5	7	NO	3	3
24	F	30	21323	8	12	NO	1	3
25	F	24	19820	6	4	NO	1	3
26	F	20	16920	3	15	NO	1	5
27	F	23	14210	2	2	SI	1	2
28	M	23	28723	5	4	SI	2	3
29	F	25	14829	7	21	NO	1	2
30	F	31	15212	5	2	NO	1	4
31	M	24	26870	5	7	NO	2	6
32	M	24	14213	3	5	SI	2	2
33	M	22	22505	8	18	NO	2	3
34	M	35	10543	2	1	SI	1	4
35	M	24	10050	1	1	SI	2	3
36	M	28	23411	4	5	SI	1	3
37	M	23	21254	4	3	SI	1	5
38	F	26	14402	2	2	SI	1	7
39	F	24	13543	6	7	NO	3	3
40	F	35	24832	5	6	NO	1	2
41	M	26	13458	7	5	NO	1	5
42	F	24	19229	8	10	SI	1	2
43	M	36	32000	4	11	NO	3	3
44	F	20	13290	3	3	NO	1	2
45	F	37	30546	2	2	NO	2	3
46	F	23	10160	2	1	NO	2	7
47	F	19	21457	6	9	SI	2	4
48	F	38	32520	8	15	SI	1	3
49	F	19	10033	5	12	SI	1	3
50	F	20	19915	5	4	SI	1	3

51	M	23	17803	4	3	SI	2	6
52	F	31	21410	4	3	SI	2	4
53	F	26	20838	3	2	SI	4	3
54	F	24	12385	3	4	SI	1	2
55	M	25	12376	2	2	SI	1	4
56	M	21	22554	1	2	SI	2	3
57	F	23	10129	5	5	NO	2	
58	F	29	13413	8	11	NO	1	3
59	F	27	26543	8	19	NO	1	2
60	M	26	12890	5	7	NO	4	2
61	F	25	15193	4	7	NO	2	3
62	M	23	14152	3	4	SI	2	
63	M	28	23215	1	1	NO	1	2
64	M	24	12539	5	3	NO	1	4
65	M	29	24643	6	12	NO	1	2
66	F	19	21659	7	10	NO	1	5
67	F	19	22430	5	3	SI	1	3
68	F	23	29540	2	15	NO	2	6
69	F	39	34000	3	5	NO	1	4
70	F	36	15409	9	18	NO	1	3
71	F	27	10000	2	4	SI	2	6
72	F	29	14904	8	21	NO	1	3
73	F	26	12868	7	7	NO	2	5
74	F	35	29533	8	10	NO	1	4
75	F	24	25432	9	5	NO	1	2
76	M	30	15504	8	6	SI	1	2
77	M	21	21739	8	10	NO	4	2
78	M	25	20201	9	20	NO	1	4
79	F	29	21324	8	12	NO	1	3
80	M	23	18676	5	4	SI	2	3
81	F	25	14829	7	12	NO	1	2
82	M	49	12503	8	18	NO	2	3

83	M	23	26432	2	1	SI	1	4
84	M	23	21253	4	3	SI	1	5
85	F	22	14839	5	6	NO	1	2
86	F	24	19596	2	2	NO	2	3
87	F	23	21416	4	3	SI	2	4
88	F	26	20454	3	2	SI	4	3
89	F	24	12380	3	4	SI	1	2
90	M	23	11872	7	10	NO	1	4
91	M	23	11610	3	2	NO	1	3
92	F	27	14855	4	4	SI	2	2
93	M	26	12374	2	2	SI	1	4
94	M	25	22550	1	2	SI	2	3
95	M	25	17632	6	5	SI	1	3
96	F	23	10125	5	5	NO	2	
97	F	24	13412	8	11	NO	1	3
98	F	29	23167	8	19	NO	1	2
99	M	36	32890	5	7	NO	4	2
100	F	25	15190	4	7	NO	2	3
101	M	23	27415	3	4	SI	2	
102	M	28	21076	1	1	NO	1	2
103	M	24	24253	5	3	NO	1	4
104	M	29	32021	6	12	NO	1	2
105	F	23	11560	7	10	NO	1	5

A continuación, en una tabla de Datos Agrupados en Lista se agruparán la cantidad de materias aprobadas de los 101 estudiantes que tienen al menos una materia aprobada.

i	y_i	n_i	$h_i = n_i/n$	N_i	$H_i = N_i/n$
1	2	32	32/101	32	32/101
2	3	36	36/101	32+36= 68	68/101
3	4	17	17/101	32+36+17= 85	85/101

4	5	8	8/101	32+36+17+8= 93	93/101
5	6	6	6/101	32+36+17+8+6= 99	99/101
6	7	2	2/101	32+36+17+8+6+2= 101	101/101

A modo de ejemplo se interpretarán las frecuencias de la tercera clase:

n_3=17 Hay 17 estudiantes que tienen 4 materias aprobadas

N_3= 85 Hay 85 estudiantes que tienen 4 o menos materias aprobadas

h_3: 17/101=0,17 Es la proporción de estudiantes que tienen 4 materias aprobadas

H_3: 85/101=0,84 Es la proporción de estudiantes que tienen 4 o menos materias aprobadas

A continuación se agruparán las edades de los 105 estudiantes en una tabla con 6 intervalos, el procedimiento para obtener dicha tabla comprende los siguientes pasos:

Paso 1

En la Serie Simple se identifica el valor mínimo y el valor máximo (la Serie Simple es la columna que contiene las edades de los 105 estudiantes):

Valor mínimo= 19
Valor máximo= 49

Paso 2: Se determina el recorrido:

R = Valor máximo – Valor mínimo = 49 – 19 = 30

Paso 3: Se obtiene la amplitud de cada intervalo i (ci)

La amplitud es la diferencia entre el límite superior y el límite inferior de cada intervalo.
Se obtiene a través del siguiente cociente:

ci= R/m Siendo "m" la cantidad de intervalos que contiene la tabla.
ci= 30/6
ci= 5

Paso 4: Se determinan los valores extremos de la tabla:

Límite inferior del primer intervalo = 19

(Valor mínimo identificado en el Paso 1)

Límite superior del primer intervalo = 49
(Valor máximo identificado en el Paso 1)

Paso 5: Se obtienen los límites en cada intervalo:

El límite superior de cada intervalo se obtiene a través de la siguiente fórmula:

límite inferior + amplitud= límite superior

Simbólicamente: $y'_{i-1} + c_i = y'_i$

Por ejemplo, el límite superior del primer intervalo se obtiene de la siguiente manera:

19+ 5 = 24

El límite superior de cada intervalo pasa a ser el límite inferior del siguiente intervalo.

y'_{i-1}	y'_i
19	24
24	29
29	34
34	39
39	44
44	49

Paso 6: Se obtienen las frecuencias en cada intervalo

Frecuencia Absoluta Simple (ni):

Para calcular la Frecuencia Absoluta Simple (ni) deben contarse en la Serie Simple (columna correspondiente a Edad) las observaciones comprendidas en cada intervalo, para ello se tendrá en cuenta que **cada intervalo comprende los valores mayores o iguales que el límite inferior y menores que el límite superior**.

Cada intervalo es cerrado a izquierda (es decir que incluye al límite inferior) y es abierto a derecha (es decir que no incluye al límite superior)
Por ejemplo, el primer intervalo comprende los valores mayores o iguales que 19 y menores que 24.

El último intervalo es el único que incluye el límite inferior y el límite superior.

Frecuencia Absoluta Acumulada (Ni):
Se obtiene sumando las **Frecuencias Absolutas Simples (ni)** desde el primer intervalo hasta el intervalo i.

Frecuencia Relativa Simple (hi):
Se obtiene mediante el cociente entre la **Frecuencia Absoluta Simple (ni)** y el total de observaciones (n).

Frecuencia Relativa Acumulada (Hi):
Se obtiene mediante el cociente entre la **Frecuencia Absoluta Acumulada (Ni)** y el total de observaciones (n).

Edad en años y'_{i-1}	y'_i	ni	Ni= $n_1 + n_2 + n_3 + \ldots + n_i$	hi= ni /n	H_i = Ni /n
19	24	37	37	37/105= 0,35	37/105= 0,35
24	29	45	37+ 45 = 82	45/105= 0,43	82/105= 0,78
29	34	11	37+ 45+ 11 = 93	11/105= 0,10	93/105= 0,88
34	39	10	37+ 45+ 11+ 10 = 103	10/105= 0,10	103/105= 0,98
39	44	1	37+ 45+ 11+ 10+ 1 = 104	1/105= 0,01	104/105= 0,99
44	49	1	37+ 45+ 11+ 10+ 1+ 1= 105	1/105= 0,01	105/105= 1
		∑ni=105		∑hi= 1	

Gráficos

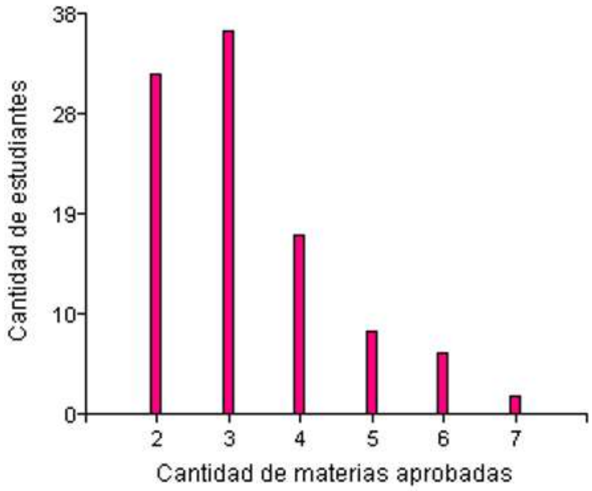
Gráficos para Datos cualitativos	- Círculo Radiado (Gráfico circular) - Barra Porcentual - Gráfico de barras	
Gráficos para	**frecuencias absolutas simples (ni)** frecuencias relativas simples (hi= **ni**/n)	**frecuencias absolutas acumuladas (Ni)** frecuencias relativas acumuladas (Hi= **Ni**/n)
Datos agrupados en lista	❖ Diagrama de bastones	❖ Gráfico acumulativo de frecuencias
Datos agrupados en intervalos	❖ Histograma ❖ Polígono de frecuencias	❖ Diagrama escalonado ❖ Ojiva

Gráficos para datos agrupados en lista:

Gráfico para representar frecuencias absolutas simples o frecuencias relativas simples:

Diagrama de Bastones:

Cantidad de materias aprobadas de 101 estudiantes.



Nombre del gráfico: Diagrama de Bastones.

Fuente: Encuesta realizada a 105 estudiantes de Licenciatura en Kinesiología

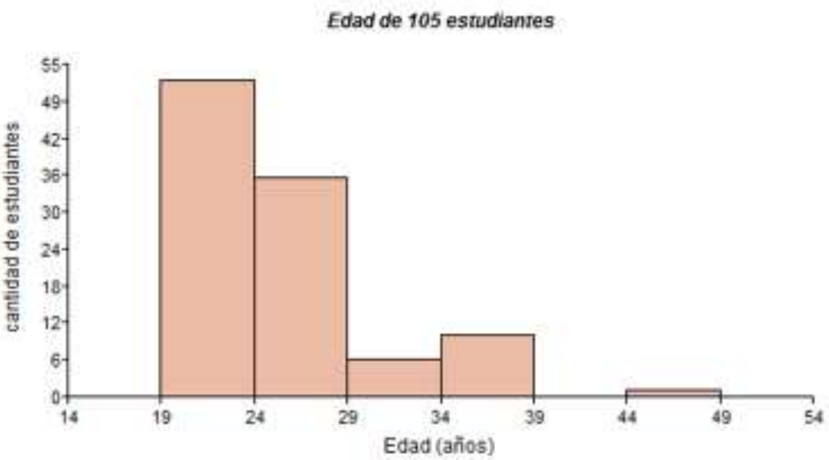
Universidad de Monterrey el 29 de Mayo de 2021.

Gráficos para datos agrupados en intervalos:

Gráficos para representar frecuencias absolutas simples o frecuencias relativas simples:

- ❖Histograma
- ❖Polígono de Frecuencias

Histograma:



Nombre del gráfico: Histograma.
Fuente: Encuesta realizada a 105 estudiantes de Licenciatura en Kinesiología-
Universidad de Monterrey el 29 de Mayo de 2021

A medida que se incrementan las edades, aumenta la cantidad de estudiantes, en consecuencia predominan los estudiantes de menor edad.

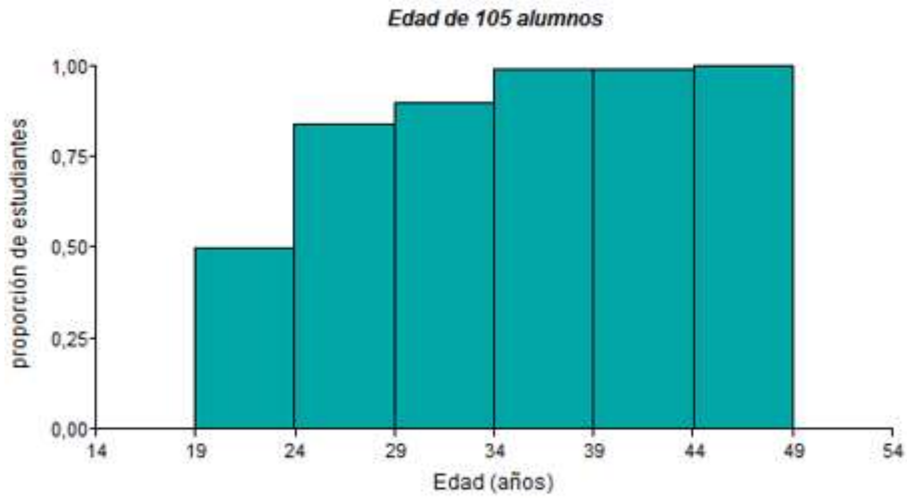
Polígono de Frecuencias:

Nombre del gráfico: Polígono de Frecuencias.
Fuente: Encuesta realizada a 105 estudiantes de Licenciatura en Kinesiología- Universidad de Monterrey el 29 de Mayo de 2021

Gráficos para representar frecuencias absolutas acumuladas o frecuencias acumuladas:

- ❖ Diagrama Escalonado
- ❖ Ojiva

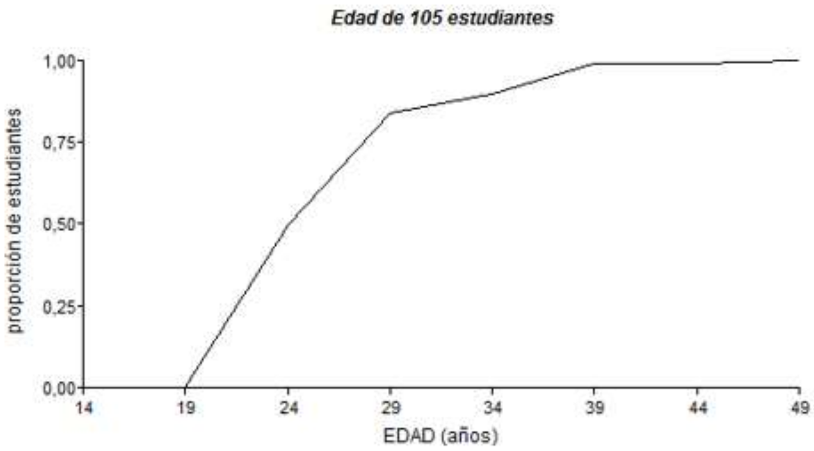
Diagrama Escalonado



Nombre del gráfico: Diagrama Escalonado.
Fuente: Encuesta realizada a 105 estudiantes de Licenciatura en Kinesiología- Universidad de Monterrey el 29 de Mayo de 2021

En los dos primeros intervalos se registra un incremento abrupto en la proporción acumulada de estudiantes, mientras que en los últimos intervalos no se visualiza un crecimiento en dicha proporción.

Ojiva



Nombre del gráfico: Ojiva.
Fuente: Encuesta realizada a 105 estudiantes de Licenciatura en Kinesiología-Universidad de Monterrey el 29 de Mayo de 2021

El Histograma, el Polígono de Frecuencias, el Diagrama Escalonado y la Ojiva solo se utilizan para representar datos agrupados en intervalos. No se utilizan para representar datos agrupados en lista.

Descripción de los Datos Estadísticos

La descripción de los Datos Estadísticos se realiza a través de las Medidas Descriptivas.

- **Medidas de Posición**

Medidas de Posición Centrales:

Media Aritmética

Mediana

Moda

Medidas de Posición no Centrales (Fractiles)

Cuartiles

Deciles

Percentiles

- **Medidas de Dispersión**

Varianza

Desviación Estándar

Coeficiente de Variación

Medidas de Posición Centrales:

Media Aritmética: Promedio

Suma de todos los valores observados divido la cantidad total de observaciones.

Media poblacional:

$$\mu = \frac{\sum X}{N}$$

Media muestral:

$$\bar{x} = \frac{\sum x}{n}$$

Ejemplo:

Cantidad de hijos de 7 familias que fueron seleccionadas para ser encuestadas:

4 2 1 0 6 5 3 n=7

Media= 21/7= 3

Interpretación: En promedio cada familia tiene 3 hijos (Cantidad de hijos promedio)

Mediana:

Es un valor tal que el 50% de las observaciones son menores que la mediana y el 50% de las observaciones son mayores que la mediana.

La Mediana divide al conjunto de observaciones en dos partes porcentualmente iguales.

Ejemplo:

Cantidad de hijos de 7 familias seleccionadas para ser encuestadas:

n= 7

Se ordenan los valores:

1°	2°	3°	4°	5°	6°	7°
0	1	2	3	4	5	6

n= 7 (cantidad total de observaciones)

Se debe encontrar la ubicación de la mediana a través de la siguiente fórmula:

$\left(\frac{n+1}{2}\right)^{0}$ Ubicación de la Mediana, no es el valor

$\left(\frac{7+1}{2}\right)^{0} = 4°$ La Mediana es el cuarto valor de la serie, no es el valor de Mediana

Mediana= 3

Interpretación: La mitad de las familias tienen 3 o menos hijos y la otra mitad de las familias tienen 3 o más hijos.

4 2 1 0 6 5 3 1

n= 8

Se ordenan los valores:

1°	2°	3°	4°	5°	6°	7°	8°	
0	1	1	2	3	4	5	6	n= 8

Se determina la posición de la Mediana:

$\left(\frac{n+1}{2}\right)^{0} = \left(\frac{8+1}{2}\right)^{0} = \mathbf{4,5}$ **No es el valor de la Mediana**

por lo tanto el valor de la Mediana es el promedio entre el **cuarto** y el **quinto** valor de la serie:

1°	2°	3°	**4°**	**4,5**	**5°**	6°	7°	8°
0	1	1	**2**	**2,5**	**3**	4	5	6

Mediana: (2 + 3) / 2= **2,5**

La mitad de las familias tienen menos de 2,5 hijos y otra la mitad de las familias tienen más de 2,5 hijos.

Moda:

Valor más frecuente

Ejemplos:

Cantidad de hijos de 8 familias encuestadas en una Provincia “C”:

4 2 1 0 6 5 3 1

Moda= 1 → Distribución Unimodal.

Interpretación: Cantidad de hijos más frecuente

Fue encuestado otro grupo de familias y se obtuvo la siguiente información:

1 **1** 2 **3** **3** 4 5 6 7

Dos Modas: **1** y **3** → Distribución Bimodal.

Cantidad de hijos de otro grupo de familias encuestadas:

1 **2** **1** **2** **3** **3** 0 4 5 6 7 8

Tres modas: **1**, **2** y **3**

La cantidad de hijos de otro grupo de familias:

1 2 3 4 5 6 7

No hay Moda debido a que no existe un valor que se presente mayor cantidad de veces, por lo tanto es una distribución Amodal.

Medidas de Posición Centrales para Datos agrupados en lista

Moda para datos agrupados en lista:

En la tabla de datos agrupados en lista, la Moda es el valor de la clase (yi) al cual le corresponde la mayor frecuencia absoluta simple (mayor ni)

Ejemplo:

En la tabla de datos agrupados en lista, la Moda es el valor de la clase (yi) al cual le corresponde la mayor frecuencia absoluta simple (mayor ni)

Cantidad de hijos de 50 empleados de una empresa:

yi: cantidad de hijos	ni frecuencia absoluta simple
1	8
2	10
Moda=3	mayor ni:15
4	13
5	4

Moda= 3 Cantidad de hijos más frecuente.

Media Aritmética para datos agrupados en lista:

$$M(y) = \frac{\sum yi * ni}{n}$$

Por cada clase se obtiene el producto: yi* **ni**

i	y_i	n_i	$y_i\, n_i$

1	y_1	n_1	$y_1\ n_1$
2	y_2	n_2	$y_2\ n_2$
3	y_3	n_3	$y_3\ n_3$
4	y_4	n_4	$y_4\ n_4$
…	…	…	…
m	y_m	n_m	$y_m\ n_m$
		$\sum n_i = n$	$\sum y_i\ n_i$

Ejemplo:

Cantidad de hijos de 50 empleados:

yi	ni frecuencia absoluta simple	yi*ni
1	8	1*8= 8
2	10	2*10= 20
3	15	3*15= 45
4	13	4*13= 52
5	4	5*4= 20
	n=50	∑ yi*ni= 145

M(y) = $\frac{\sum yi * ni}{n}$

M(y) = $\frac{145}{50}$

Media = 2,9

Interpretación: En promedio cada empleado tiene 2,9 hijos

(Cantidad de hijos promedio)

Otros ejemplos.

Cantidad de materias aprobadas de 50 estudiantes:

y i	**ni**	**y i*ni**
1	**2**	**1*2= 2**
2	**10**	**2*10=20**
Moda=3	**Mayor ni= 19**	**3*19= 57**
4	**12**	**4*12= 48**
5	**7**	**5*7= 35**
	n=50	**∑y i*ni= 162**

Cálculo e interpretación de la Media Aritmética:

Media= ∑y i*ni / n

= 162 / 50

Media= 3,24

Interpretación: En promedio cada estudiante tiene 3,24 materias aprobadas.

La interpretación de la Media, también puede expresarse de la siguiente manera: Cantidad promedio de materias aprobadas.

Cálculo e interpretación de la Moda:

Moda=3 Interpretación: Cantidad de materias aprobadas más frecuente.

Mediana para datos agrupados en lista

Ejemplos:

yi: cantidad de hijos	ni	Ni: frecuencia absoluta acumulada
1	**2**	**2**
yj-1=2	**10**	**Nj-1=12**
y j=3	**19**	**1era. Frec. Abs. Acum. que supera a n/2: N j=31**
4	**12**	**43**
5	**7**	**n=50**
	n=50	

1) Se obtiene la **Frecuencia absoluta acumulada: Ni**
2) Se obtiene **n**/2= **50**/2   **n**/2 = 25
3) N j= Primera **frecuencia absoluta acumulada que supera a n/2**

N j= 31 a continuación identificamos la clase (yi) a la cual le corresponde esta frecuencia y la simbolizamos como y j. En este caso yj= 3

4) Se identifica el valor de la **frecuencia absoluta acumulada de la clase anterior: Nj-1= 12**

5) **n/2 > Nj-1**   →   **yj=3 es el valor de la Mediana**
   25 **> 12**

Interpretación: La mitad de los estudiantes tiene 3 o menos materias aprobadas y la otra mitad de los estudiantes tienen 3 o más materias aprobadas.

Cantidad de hijos de 20 empleados que trabajan en una empresa "B":

Cantidad de hijos (yi)	Cantidad de empleados (**ni**)	Ni: frecuencia Absoluta Acumulada
0	**2**	**2**
1	**4**	**6**
yj-1= 2	**4**	**Nj-1= 10**
y j=3	**7**	**1era. Frec. Abs. Acum. que supera a n/2: Nj=17**
4	**3**	**20**

Se realizará el cálculo y la interpretación de la Mediana.

1) Calculamos **Ni** para cada clase
2) **n/2= 20/2= 10**

3) N j= 17 Primera frecuencia absoluta acumulada que supera a n/2
4) Nj-1= 10 Frecuencia anterior a N j
5) (n/2=10) = (Nj-1=10)

Si **n/2 = Nj-1**

entonces

Mediana= (y_{j-1}**+** y_j) / 2 Valor de la Mediana

y j-1: Clase a la cual le corresponde la frecuencia N j-1

y j: Clase a la cual le corresponde la frecuencia N j

n/2 = Nj-1 ⟶ Mediana= (yj-1+ y j) / 2

= (2 + 3) / 2

Mediana= 2,5

Interpretación: La mitad de los empleados tienen menos de 2,5 hijos y la otra mitad de los empleados tienen más de 2,5 hijos.

Medidas de Posición Centrales para Datos agrupados en intervalos

Media Aritmética para datos agrupados en intervalos:

Ejemplo:

i	y'_{i-1}	y'_i	ni	yi	yi * ni
1	100	200	4	150	150*4= 600
2	200	300	7	250	250*7= 1750
3	300	400	10	350	350*10= 3500
4	400	500	15	450	450*15= 6750
5	500	600	20	550	550*20= 11000
6	600	700	44	650	650*44= 28600

∑yi * ni= 52200

Primer paso: Por cada intervalo se obtiene el punto medio de cada intervalo

yi: Punto medio de cada intervalo

$$yi = \frac{y'_{i-1} + y'_i}{2}$$

Segundo paso: Por cada intervalo se obtiene el producto: yi * ni

Tercer paso: Se suman los productos obtenidos por cada intervalo:

∑yi * ni

Cuarto paso: Se divide la sumatoria de los productos obtenidos por la cantidad total de observaciones.

Media Aritmética:

$$M(y) = \frac{\sum yi * ni}{n}$$

$$M(y) = \frac{52200}{100}$$

M(y)= 522

Interpretación: Duración Promedio, es decir que en promedio cada lámpara tiene una duración de 522 días.

Mediana: Es un valor tal que el 50% de las observaciones son menores que la Mediana y el otro 50% son mayores que la Mediana.

$$Me = (y_{i-1})_j + c_j \left(\frac{\frac{n}{2} - N_{j-1}}{n_j} \right)$$

i	y'i-1		y'i	ni:Frecuencia Absoluta Simple	Ni: Frecuencia Absoluta Acumulada
1	100		200	4	4
2	200		300	7	11
3	300		400	10	21
4	400		500	15	N j-1 = 36
j= 5	500	Me	600	ni = 20	Nj= 56
6	600		700	44	100

1) n/2= 100/2

= 50

2) N_j = primera **Frecuencia Absoluta Acumulada** que supera a n/2

El intervalo al cual le corresponde esta frecuencia es el que contiene a la Mediana.

N_j = 36 por lo tanto la Mediana está comprendida entre **500** y 600

$(y_{i-1})_j$ = 500 límite inferior del intervalo de Mediana

3) c j = amplitud del intervalo de Mediana

= 600 - 500

c j = 100

4) N_{j-1}= **Frecuencia Absoluta Acumulada** del intervalo anterior al de Mediana.

N_{j-1} = 36

5) n j = **Frecuencia Absoluta Simple** del intervalo de Mediana.

n_j = 20

Reemplazamos en la fórmula:

$$Me = (y_{i-1})_j + c_j \left(\frac{\frac{n}{2} - N_{j-1}}{n_j} \right)$$

$$= 500 + 100 \left(\frac{50 - 36}{20} \right)$$

$$= 500 + 100 \left(\frac{14}{20} \right)$$

= 500 + 70

Me = 570

Interpretación: **La mitad de las lámparas duran menos de 570 días y la otra mitad de las lámparas duran más de 570 días.**

Moda: Valor más frecuente

$$Md = (y_{i-1})_j + c_j \left(\frac{d_1}{d_1+d_2}\right)$$

La moda se encuentra en aquel intervalo al cual le corresponde la mayor **frecuencia absoluta simple.**

$$d_1 = n_j - n_{j-1}$$

$$d_2 = n_j - n_{j+1}$$

n_{j-1}= ni del intervalo anterior al Modal

n_j = mayor **Frecuencia Absoluta Simple**

n_{j+1}= ni del intervalo posterior al Modal

Ejemplo:

i	y'$_{i-1}$	y'$_i$	**ni: Frecuencia Absoluta Simple**
1	100	200	4
2	200	300	7
3	300	400	10
4	400	500	15
5	500	600	20
6	**600**	700	n_j = **44**

n_j = **44** por lo tanto el Intervalo comprendido entre 600 y 700 es el Intervalo Modal

Límite inferior del intervalo modal: $(y_{i-1})_j$ = **600**

c_j: amplitud del intervalo modal

c_j = 700 – 600

= 100

El intervalo al cual le corresponde esta frecuencia es el que contiene la Moda.

Se identifican las frecuencias absolutas simples que intervienen en el cálculo de las diferencias:

n_{j-1}= 20 ni del intervalo anterior al Modal

$\boldsymbol{n_j}$ = **44** mayor ni

n_{j+1} = 0 ni del intervalo posterior al Modal

Se obtienen las diferencias d_1 y d_2, pero previamente deben calcularse:

$$d_1 = \boldsymbol{n_j} - n_{j-1}$$

= **44** – 20

d_1 = 24

$$d_2 = \boldsymbol{n_j} - n_{j+1}$$

= **44** - 0

$$d_2 = 44$$

Se reemplazan los valores en la fórmula:

$$Md = (\boldsymbol{y_{i-1}})_j + c_j \left(\frac{d_1}{d_1+d_2}\right)$$

$$= \mathbf{600} + 100 \left(\frac{24}{24+44}\right)$$

$$= \mathbf{600} + 100 \left(\frac{24}{68}\right)$$

= 600 + 35,29

Md = 635,29

Interpretación: Duración más frecuente

Medidas de Posición no centrales (Fractiles)

Son medidas que dividen al conjunto de datos en partes porcentualmente iguales.

Cuartiles:

Dividen al conjunto de datos en cuatro partes porcentualmente iguales, cada parte contiene el 25% de las observaciones.

Deciles:

Dividen al conjunto de datos en diez partes porcentualmente iguales, cada parte contiene el 10% de las observaciones.

Percentiles:

Dividen al conjunto de datos en cien partes porcentualmente iguales, cada parte contiene el 1% de las observaciones.

Fractiles para datos no agrupados (Series Simples)

Cuartiles:

Dividen al conjunto de datos en cuatro partes porcentualmente iguales, cada parte contiene el 25% de las observaciones.

```
  25%    25%    25%    25%
|_____|_____|_____|_____|
      Q1    Q2    Q3
```

```
  25%    25%    25%    25%
|_____|_____|_____|_____|
      Q1    Q2    Q3
```

La Mediana es el Segundo Cuartil (Q2)

Primer paso: Ordenamos los valores (de menor a mayor, o de mayor a menor)

Segundo paso: Se determina la ubicación del cuartil en la serie.

Fórmula para determinar la ubicación o posición de los cuartiles:

$[k(n + 1) / 4]^{\circ}$

k es el número de cuartil.

Ejemplos:

Para el **Primer** cuartil **k** vale **1**.

Para el **Segundo** cuartil **k** vale **2**.

Para el **Tercer** cuartil **k** vale **3**

Tercer paso: Determinamos el valor del cuartil.

Primer Cuartil o Cuartil 1(Q1)

Definición:

Es un valor tal que el 25% de las observaciones son menores que el primer cuartil y el 75% de las observaciones son mayores que el primer cuartil.

El primer cuartil supera al 25% de las observaciones y es superado por el 75% de las observaciones.

25% 25% 25% 25%

|_____|_____|_____|_____|

Q_1

Ejemplo:

Dada la cantidad de materias aprobadas de los estudiantes de cierta carrera:

9 7 5 1 6 2 6 9 4

Calcular el Primer Cuartil.

Se determinará el **Primer** Cuartil (Cuartil **uno**)

Paso 1: Se ordenan los valores

1 2 4 5 6 6 7 9 9

n: cantidad total de observaciones n= 9

Paso 2: Se determina la posición o ubicación del **Primer Cuartil**

k= **1**

[**k**(n + 1) / **4**]° = [**1**(9 + 1) / **4**]°

= 2,5 No es valor del Primer Cuartil, este resultado indica que el Primer Cuartil estará comprendido entre el segundo y el tercer valor.

Paso 3: Se Obtiene el valor del Primer Cuartil

El primer cuartil estará comprendido entre el segundo y el tercer valor.

1° 2° 2,5 3° 4°

1 2 4 5 6 6 7 9 9

El Primer Cuartil será el promedio entre el segundo y el tercer valor:

Q_1= (2 + 4)/ 2

Q_1= 3 Valor del Primer Cuartil

Q_1=3

Interpretación:

Para realizar la interpretación debe tenerse en cuenta la definición del cuartil y la característica bajo estudio.

Si el fractil <u>no</u> coincide con algún valor de la serie entonces se expresa la interpretación de la siguiente manera:

Interpretación en el caso planteado:

Aproximadamente el 25% de los estudiantes tienen menos de 3 materias aprobadas, y aproximadamente el 75% de los estudiantes tienen más de 3 materias aprobadas.

Tercer Cuartil (Q3):

Es un valor tal que el 75% de las observaciones son menores que el tercer cuartil y el 25% de las observaciones son mayores que el tercer cuartil.

El Tercer Cuartil supera al 75% de las observaciones y es superado por el 25% de las observaciones.

25% 25% 25% 25%

|____|____|____|____|

Q3

Ejemplo:

Dada la cantidad de materias aprobadas de 9 estudiantes:

1 2 4 5 6 6 7 9 9

Calcular el Tercer Cuartil

Primer paso:

Se ordenan los valores (En este caso ya se encuentran ordenados)

Segundo paso:

Se determina la posición o ubicación del **Tercer Cuartil** a través de la siguiente fórmula:

El **Tercer Cuartil** es el cuartil **3** por lo tanto **k=3**

[k(n + 1) / 4]° = [3(9 + 1) / 4]°

= [30 / 4]°

[30 / 4]° = 7,5 No es el valor del tercer cuartil

7,5 indica que el tercer cuartil será un valor comprendido entre el séptimo y el octavo valor.

1°	2°	3°	4°	5°	6°	7°	7,5	8°	9°
1	2	4	5	6	6	**7**		**9**	9

Tercer paso:

Se determina el valor del **Tercer Cuartil**

Si al calcular **[k(n + 1) / 4]°** se obtiene un resultado con cifras decimales, entonces se determina el valor del Cuartil mediante un promedio.

Tercer cuartil: Q3= (7 + 9)/2

Q3= 8 Valor del Tercer Cuartil

Interpretación: Aproximadamente el 75% de los estudiantes tienen menos de 8 materias aprobadas, y aproximadamente el 25% de los estudiantes tienen más de 8 materias aprobadas.

25% 25% 25% 25%

|____|_____|_____|_____|

Q3=8

Otros ejemplos:

Dada la cantidad de cuentas por cobrar en un conjunto de empresas muestreadas:

1 8 8 7 5 6 8 5 4 3 4

Calcular e interpretar Primer Cuartil y Tercer Cuartil

Primer paso: Se ordenan los valores observados

1 3 4 4 5 5 6 7 8 8 8

n= 11

Segundo paso: Se identifica la **ubicación** del **Primer Cuartil** (**Cuartil 4**):

[k(n + 1) / 4]°

k=1 n=11

[k(n + 1) / 4]°= [1(11 + 1) / 4]°

= (12 / 4)°

= 3°

3 no es el valor del primer cuartil, este resultado indica que el primer cuartil es el tercer valor de la serie

1°	2°	3°	4°	5°	6°	7°	8°	9°	10°	11°
1	3	4	4	5	5	6	7	8	8	8

El tercer valor de la serie es 4, por lo tanto el primer cuartil es igual a 4

Q1= **4**

25% 25% 25% 25%

|____|_____|_____|_____|

Q1= 4

Interpretación del Primer Cuartil:

Para realizar la interpretación se tiene en cuenta cuál es la variable bajo estudio, en este caso es la cantidad de cuentas por cobrar de 11 empresas.

Si el cuartil coincide con algún valor de la serie entonces se expresa la interpretación de la siguiente manera:

Aproximadamente el 25% de las observaciones son menores o iguales que el valor del cuartil y el 75% de las observaciones son mayores o iguales que el valor del fractil.

Interpretación en el caso planteado:

Aproximadamente el 25% de las empresas tienen 4 o menos cuentas por cobrar, y aproximadamente el 75% de las empresas tienen 4 o más cuentas por cobrar.

Tercer Cuartil:

1°	2°	3°	4°	5°	6°	7°	8°	9°	10°	11°
1	3	4	4	5	5	6	7	8	8	8

Tercer Cuartil (Cuartil 3)

[k(n + 1) / 4]°= **[3(11 + 1) / 4]°**

= (36 /4)°

= 9°

9 no es el valor del tercer cuartil, este resultado indica que el tercer cuartil es el noveno valor de la serie

Si al calcular **[k(n + 1) / 4]°** se obtiene un resultado exacto, entonces se busca en la serie el valor del Cuartil al cual le corresponda la posición calculada.

1°	2°	3°	4°	5°	6°	7°	8°	**9°**	10°	11°
1	3	4	4	5	5	6	7	**8**	8	8

El noveno valor de la serie es 8, por lo tanto el tercer cuartil es igual a 8

25% 25% 25% 25%

I____I_____I_____I____I

Q1 Q2 Q3=8 cuentas por cobrar

Característica bajo estudio: Cantidad de cuentas por cobrar de 11 empresas.

Interpretación:

Aproximadamente el 75% de las empresas tienen 8 o menos cuentas por cobrar y aproximadamente el 25% restante de las empresas tienen 8 o más cuentas por cobrar.

Deciles (Dk):

Dividen al conjunto de observaciones en diez partes porcentualmente iguales, y en consecuencia cada parte contiene el 10% de las observaciones.

10%	10%	10%	10%	10%	10%	10%	10%	10%	10%

D1 D2 D3 D4 D5 D6 D7 D8 D9

A modo de ejemplo, se definirán algunos deciles:

Primer Decil (D1)**:**

Es un valor tal que el 10% de las observaciones son menores que el Primer Decil y el 90% de las observaciones son mayores que el Primer Decil.

El Primer Decil supera al 10% de las observaciones y es superado por el 90% de las observaciones.

10% 10% 10% 10% 10% 10% 10% 10% 10% 10%

|____|____|____|____|____|____|____|____|____|____|

D1

Séptimo Decil (D7):

Es un valor tal que el 70% de las observaciones son menores que el Séptimo Decil y el 30% de las observaciones son mayores que el Séptimo Decil.

El Séptimo Decil supera al 70% de las observaciones y es superado por el 30% de las observaciones.

10% 10% 10% 10% 10% 10% 10% 10% 10% 10%

|____|____|____|____|____|____|____|____|____|____|

D7

La Mediana es el Quinto Decil: Es un valor tal que el 50% de las observaciones son menores que el Quinto Decil y el 50% de las observaciones son mayores que el Quinto Decil

10% 10% 10% 10% 10% 10% 10% 10% 10% 10%

|____|____|____|____|____|____|____|____|____|____|

D1 D2 D3 D4 D5 D6 D7 D8 D9

Fórmula para determinar la ubicación de los Deciles:

[k(n + 1) / 10]°

Cantidad de materias aprobadas de los estudiantes de cierta carrera:

1 2 4 5 6 6 8 9 10 n= 9

Séptimo Decil (Decil siete)

Fórmula para determinar la ubicación del Séptimo Decil:

[k(n + 1) /10]°= [7(9 + 1) / 10]°

= 70/10

[7(9 + 1) / 10]°= 7°

Se obtuvo un resultado exacto, por lo tanto se busca en la serie el valor del decil (no debe calcularse mediante promedio)

El Séptimo Decil es el séptimo valor de la serie.

1°	2°	3°	4°	5°	6°	7°	8°	9°	
1	3	4	5	6	6	**8**	9	10	n=9

El séptimo decil es el séptimo valor de la serie que es 8. El valor del séptimo decil es 8

10% 10% 10% 10% 10% 10% 10% 10% 10% 10%

|____|____|____|____|____|____|____|____|____|____|

D7= 8

Interpretación:

Aproximadamente el 70% de los alumnos tienen 8 o menos materias aprobadas, y aproximadamente el 30% de los alumnos tienen 8 o más materias aprobadas.

Calcular **Segundo Decil**.

Procedimiento:

Se determina la posición del Segundo Decil:

[2(9 + 1) / 10]°= 2° El Segundo Decil es el valor de la serie.

Se obtuvo un resultado exacto (2), por lo tanto se busca en la serie el valor del decil.

1° 2°

1 3 4 5 6 6 7 9 9

El Segundo Decil es el segundo valor de la serie que es 3. El valor del segundo decil es **3**: **D_2= 3**

10% 10% 10% 10% 10% 10% 10% 10% 10% 10%

|____|____|____|____|____|____|____|____|____|____|

D2=3

Interpretación: Aproximadamente el 20% de los estudiantes tienen 3 o menos materias aprobadas y aproximadamente el 80% de los estudiantes tienen 3 o más materias aprobadas.

Si el valor del fractil coincide con alguna observación, entonces se incluye en ambos porcentajes (3 o menos, 3 o más)

Ejemplo:

Cantidad de cuentas por cobrar de 11 empresas:

1 3 4 4 5 5 6 7 8 8 8

n=11

Calcular Sexto Decil

Procedimiento:

Se determina la ubicación del decil:

[k(n + 1) /10]°= [6(11 + 1) / 10]°

= 6*12/10

= 72/10

= 7,2

1° 2° 3° 4° 5° 6° 7° 7,2 8°

1 3 4 4 5 5 6 7 8 8 8 n=11

D6= (6 + 7) /2

D6 = 6,5 Valor del **sexto** decil para cuentas por cobrar de 11 empresas.

10% 10% 10% 10% 10% 10% 10% 10% 10% 10%

|____|____|____|____|____|____|____|____|____|____|

D6=6,5

Interpretación del sexto decil:

El 60% de las empresas tienen menos de 6,5 cuentas por cobrar, y el 40% de las empresas tienen más de 6,5 cuentas por cobrar

Al aplicar esta fórmula obtuvimos un resultado

[k(n + 1) /10]° con decimales, por lo tanto:

El decil debe calcularse mediante promedio

En la interpretación no debe incluirse en los porcentajes el valor del decil (es decir no utilizamos la palabra igual)

Percentiles (Pk)

Son valores que dividen al conjunto de observaciones en 100 partes porcentualmente iguales, por ello cada parte contiene el 1% de las observaciones.

1% 1% 1% 1% 1% 1% 1% 1% 1% 1% 1% 1% 1%

|____|____|____|____|____|____|____|____|____|____|____|____|____|____|____|

P1 P2 P3 P4 .... …….. P99

Fórmula para determinar la ubicación o posición de un Percentil:

[k (n + 1) /100]°

Ejemplos de Percentiles:

Cuarto Percentil:

Es un valor tal que el 4% de las observaciones son menores, y en consecuencia el 96% restante de las observaciones son mayores que el **Cuarto** Percentil (Percentil 4)

Percentil (P15)

Es un valor tal que y el 15% de las observaciones son menores, y en consecuencia el 85% de las observaciones son mayores que el percentil 15.

La Mediana es el Percentil 50: Es un valor tal que el 50% de las observaciones son menores, y por lo tanto el 50% de las observaciones son mayores que el percentil 50.

Ejemplo de percentil:

Cantidad de consultas que ingresaron a un centro de atención al cliente durante 12 días muestreados:

15 17 16 16 15 17 15 18 14 16 15 11

n=12

Calcular e interpretar el percentil 32 (P32)

Primer paso: Se ordenan los valores de la serie:

11 14 15 15 15 15 16 16 16 17 17 18

n=12

Segundo paso: Se determina la ubicación del Percentil 32

[32(12 + 1) /100]°= 4,16

El Percentil 32 será un valor comprendido entre el cuarto y el quinto de la serie.

1°	2°	3°	4°	5°							
11	14	15	15	15	15	16	16	16	17	17	18

Se obtiene el valor del Percentil 32

P32= (15 + 15) / 2

= 15

Interpretación:

Aproximadamente el 32% de los días se recibieron 15 o menos consultas y aproximadamente el 68% de los días se recibieron 15 o más consultas.

Fractiles para datos agrupados en intervalos:

Cuartiles para datos agrupados:

$$\mathbf{Qk} = (y_{i-1})_j + c_j\left(\frac{(\mathbf{k}*\mathbf{n}\ /4)-\ N_{j-1}}{n_j}\right)$$

Deciles para datos agrupados:

$$\mathbf{Dk} = (y_{i-1})_j + c_j\left(\frac{(\mathbf{k}*\mathbf{n}\ /10)-\ N_{j-1}}{n_j}\right)$$

Percentiles para datos agrupados:

$$\mathbf{Pk} = (y_{i-1})_j + c_j\left(\frac{(\mathbf{k}*\mathbf{n}\ /100)-\ N_{j-1}}{n_j}\right)$$

k: número de fractil. Ejemplo: Para el tercer cuartil k es igual a 3

$(y_{i-1})_j$: límite inferior del intervalo que contiene al fractil

cj: amplitud del intervalo que contiene el fractil

N_j: primera frecuencia absoluta acumulada que supera a [(k*n /cantidad)

Cantidad: Puede ser igual a **4, 10** o **100** según corresponda

Intervalo j: Intervalo que contiene el fractil (cuartil, decil o percentil)

N_{j-1}: frecuencia absoluta acumulada del intervalo **anterior** al que contiene al fractil

n_j: frecuencia absoluta simple del intervalo que contiene al fractil

Ejemplos:

Estatura en cm. de 100 clientes

i	Estatura en cm. y_{i-1}	y_i	Cantidad de clientes (ni)	Ni
1	140	150	40	40
2	150	160	19	59
3	160	170	16	75
4	170	180	12	87
5	180	190	9	96
6	190	200	4	n=100

Fuente: Registro de mediciones realizadas el 20 de Julio de 2021 a 100 clientes pertenecientes a un comercio "A" dedicado a la venta de indumentaria deportiva:

Calcular e interpretar el Tercer Cuartil:

i	Estatura en cm. y_{i-1}		y_i	Cantidad de clientes(ni)	Ni
1	140		150	40	40
2	150		160	19	59
j-1=3	160		170	16	N_{j-1}=**75**
j=4	**170**	**Q_3**	180	n_j= **12**	N_j=87
5	180		190	9	96
6	190		200	4	n=100

$$\mathbf{Qk} = (y_{i-1})_j + c_j + \left(\frac{(\mathbf{k}*\mathbf{n}\ /4) - N_{j-1}}{n_j}\right)$$

$$\mathbf{Q_3} = (y_{i-1})_j + c_j + \left(\frac{(\mathbf{3}*\mathbf{n}\ /4) - N_{j-1}}{n_j}\right)$$

n=100

k*n /4= 3*100 /4

= 75

N_j: Primera frecuencia absoluta acumulada que supera 3*n /4. Esta frecuencia determina el intervalo en el cual se encuentra el cuartil.

Nj=87 j=4 por lo tanto el Tercer Cuartil está comprendido entre 170 y 180

Tercer paso:

$(y_{i-1})_j$: límite inferior del intervalo que contiene al tercer cuartil

$(y_{i-1})_j$= **170**

c_j: amplitud del intervalo que contiene el tercer cuartil

c_j= 180 -170

=10

N_{j-1}: frecuencia absoluta acumulada del intervalo **anterior** al que contiene el tercer cuartil

N_{j-1}= **75**

n_j: frecuencia absoluta simple del intervalo que contiene el tercer cuartil

n_j= **12**

$$\mathbf{Q_3} = (y_{i-1})_j + c_j \left(\frac{(\mathbf{3}*\mathbf{n}\ /4) - N_{j-1}}{n_j}\right)$$

$$= \mathbf{170} + 10 \left(\frac{(\mathbf{3}*\mathbf{100}\ /4) - \mathbf{75}}{\mathbf{12}}\right)$$

$$= \mathbf{170} + 10 \left(\frac{\mathbf{75} - \mathbf{75}}{\mathbf{12}}\right)$$

$$= 170 + 10\left(\frac{0}{12}\right)$$

= 170 + 0

$Q_3 = 170$

El valor del fractil debe incluirse en los porcentajes cuando coincide con el límite inferior del intervalo, entonces en la interpretación debe incluirse en los porcentajes.

Interpretación:

Aproximadamente el 75% de los clientes tienen una estatura igual o menor que 170 cm. y el 25% restante tienen una estatura mayor o igual que 170 cm.

Primer cuartil (Q_1): Ejemplo

$$Q_k = (y_{i-1})_j + c_j + \left(\frac{(k*n/4) - N_{j-1}}{n_j}\right)$$

i	Estatura en cm. y_{i-1}	y_i	Cantidad de clientes(ni)	Ni
0	130	140	0	N_{j-1} =0
1	140 Q_1	150	n_j= 40	**Nj=40**
2	150	160	19	59
3	160	170	16	75
4	170	180	12	87
5	180	190	9	96
6	190	200	4	n=100

$$Q_1 = (y_{i-1})_j + c_j + \left(\frac{(1*n/4) - N_{j-1}}{n_j}\right)$$

n=100

k*n /4= 1*100 /4

= 25

Nj: Primera frecuencia absoluta acumulada que supera 1*n /4. Esta frecuencia determina el intervalo en el cual se encuentra el cuartil.

Nj=40 j=1 El primer cuartil se ubica en el primer intervalo, por lo tanto su valor está comprendido entre 140 y 150

$(y_{i-1})_j$: límite inferior del intervalo que contiene al tercer cuartil

$(y_{i-1})_j = 140$

c_j: amplitud del intervalo que contiene el tercer cuartil

c_j= 150 -140

=10

N_{j-1}: frecuencia absoluta acumulada del intervalo **anterior** al que contiene el tercer cuartil

Si el fractil se ubica en el primer intervalo, entonces Nj-1= 0

n_j: frecuencia absoluta simple del intervalo que contiene el tercer cuartil

n_j= 40

$$\mathbf{Q_1} = (y_{i-1})_j + c_j \left(\frac{(\mathbf{1} * \mathbf{n} / \mathbf{4}) - N_{j-1}}{n_j} \right)$$

$$= 140 + 10 \left(\frac{(\mathbf{1} * \mathbf{100} / \mathbf{4}) - \mathbf{0}}{\mathbf{40}} \right)$$

$$= 140 + 10 \left(\frac{\mathbf{25} - \mathbf{0}}{\mathbf{40}} \right)$$

$$= 140 + 10 \left(\frac{\mathbf{25}}{\mathbf{40}} \right)$$

= 140 + 6,25

$\mathbf{Q_1} =$ 146,25

Interpretación:

Aproximadamente el 25% de los clientes tienen una estatura menor que 146,25cm. y el 75% restante tienen una estatura mayor que 146,25 cm.

Octavo decil (D_8): Ejemplo

$$D_k = (y_{i-1})_j + c_j \left(\frac{(\mathbf{k} * \mathbf{n} / \mathbf{10}) - N_{j-1}}{n_j} \right)$$

i	Estatura en cm. y_{i-1}	y_i	Cantidad de Clientes(ni)	Ni
1	140	150	40	40
2	150	160	19	59
3	160	170	16	N_{j-1}=75
j=4	170 D8	180	n_j=12	Nj=87
5	180	190	9	96
6	190	200	4	n=100

$$D_8 = (y_{i-1})_j + c_j \left(\frac{(8*100/10) - N_{j-1}}{n_j} \right)$$

n = 100

1) 8*n /10= 8*100 /10

= 80

2) Nj= Primera frecuencia absoluta acumulada que supera a 8*n /10

Nj= 87 → j=4 Por lo tanto el Octavo Decil está comprendido entre 170 y 180

3) $(y_{i-1})_j$ = **170** límite inferior del intervalo que contiene al octavo decil

4) cj = 180 - 170

= 10

5) N_{j-1}: Frecuencia absoluta acumulada del intervalo anterior al que contiene el octavo decil

N_{j-1}=**75**

6) nj= Frecuencia absoluta simple del intervalo que contiene el octavo decil

n_j= 12

$$D_8 = 170 + 10\left(\frac{\left(8*\frac{100}{10}\right)-75}{12}\right)$$

$$= 170 + 10\left(\frac{80-75}{12}\right)$$

$$= 170 + 10\left(\frac{5}{12}\right)$$

$$= 170 + \frac{50}{12}$$

= 170 + 4,166666

D_8= 174,17 cm.

Interpretación:

Aproximadamente el 80% de los clientes tienen una estatura menor que 174,17 cm. y el 20% restante tienen una estatura mayor que 174,17 cm.

Percentil 32 (P_{32})

Ejemplo:

i	Estatura en cm. y_{i-1}		y i	Cantidad de Clientes(ni)	Ni
0	130		140	0	N_{j-1}= 0
1	140	P_{32}	150	n_j=40	N_j=40
2	150		160	19	59
3	160		170	16	75
4	170		180	12	87
5	180		190	9	96
6	190		200	4	n=100

$$P_{k} = \left(y_{i-1}\right)_j + c_j\left(\frac{(\mathbf{k}*\mathbf{100}/\mathbf{100}) - N_{j-1}}{n_j}\right)$$

$$P_{32} = \left(y_{i-1}\right)_j + c_j\left(\frac{(\mathbf{32}*\mathbf{100}/\mathbf{100}) - N_{j-1}}{n_j}\right)$$

1) 32*n /100= 32*100 /100

= 32

2) N_j: Primera frecuencia absoluta acumulada que supera a 32*n /100

N_j= **40** → j=1 Por lo tanto el percentil 32 está comprendido entre 140 y 150

3) $(y_{i-1})_j$ = **140**

4) cj= 150 -140

= 10

5) N_{j-1}: Frecuencia absoluta acumulada del intervalo anterior al que contiene el Percentil 32

N_{j-1}= **0**

Si el fractil se ubica en el primer intervalo, entonces Nj-1= 0

6) n_j: Frecuencia absoluta simple del intervalo que contiene el Percentil 32

n_j=**40**

$$P_{32} = \left(y_{i-1}\right)_j + c_j\left(\frac{\left(\mathbf{32}*\frac{\mathbf{100}}{\mathbf{100}}\right) - N_{j-1}}{n_j}\right)$$

$$= \mathbf{140} + 10\left(\frac{(\mathbf{32}*\mathbf{100}/\mathbf{100}) - \mathbf{0}}{\mathbf{40}}\right)$$

$$= 140 + 10\left(\frac{32 - 0}{40}\right)$$

$$= 140 + 10 * 8$$

= 140 + 8

P_{32} = 148 cm.

Al hacer la interpretación, si el fractil no coincide con el límite inferior del intervalo entonces no debe incluirse en el porcentaje.

Interpretación del Percentil 32:

Aproximadamente el 32% de los clientes tienen una estatura menor que 148 cm. y el 68% restante tienen una estatura mayor que 148 cm.

Medidas de Dispersión

Permiten comparar dos o más grupos de datos para determinar en cuál de ellos los valores son más homogéneos, es decir en cuál grupo los valores presentan menor variabilidad. En dicho grupo la media es más representativa de los valores.

- **Recorrido**
- **Varianza**
- **Desviación Estándar**
- **Coeficiente de Variación**

Se compararán las notas de dos cursos:

x: notas en puntos

Curso “A” (N=5)	Curso “B” (N=5)
4 4 7 10 10	6 6 7 8 8
Media(A)= 7 puntos	Media (B)= 7 puntos

En el curso “B” las notas son más homogéneas (más similares entre sí)

En el curso “B” las notas se encuentran a menor distancia con respecto a la Media, por lo tanto en dicho curso las notas son más homogéneas.

En el curso “B” la media es más representativa de los valores observados.

A través de la siguiente diferencia: (x- Media) calculamos la distancia que hay entre cada valor y la Media. A esta diferencia le llamamos Desvío con respecto a la Media.

[(4-7) + (4-7) + (7-7) + (10-7) + (10-7)] / 5 = 0

$\sum$ (x- Media) =0

Para evitar que el resultado sea igual a cero, elevamos **al cuadrado** cada desvío, y así surge la Varianza:

Varianza= $\sum(x - \text{Media})^2 / N$

Definición de la Varianza:

Es la sumatoria del cuadrado de los desvíos con respecto a la media aritmética sobre la cantidad total de observaciones.

También puede definirse la Varianza como el promedio de los desvíos al cuadrado con respecto a la media aritmética.

En el caso planteado es el promedio del cuadrado de los desvíos de las notas con respecto a la nota promedio.

Varianza en curso “B”:

$V(B) = [(6-7)^2 + (6-7)^2 + (7-7)^2 + (8-7)^2 + (8-7)^2] / 5$
$V(B) = 0{,}8$

Varianza en el curso “A”:

$V(A)= [(4-7)^2 + (4-7)^2 + (7-7)^2 + (10-7)^2 + (10-7)^2] / 5$
$V(A)= 7{,}2 \text{ (puntos)}^2$

La Varianza no refleja correctamente la distancia que hay entre los valores observados y la Media, debido a que la varianza no está expresada en la misma unidad de medida que posee la variable bajo estudio (notas en puntos y varianza expresada en puntos al cuadrado).

Si se calcula la raíz cuadrada de la Varianza entonces se obtendrá una medida expresada en la misma unidad que posee la variable bajo estudio, y así llegamos a la Desviación Standar.

Desviación Standar en el curso “A”:

DS(A)= $\sqrt{\text{Varianza}}$

DS(A) = $\sqrt{7{,}2\ \text{puntos}^2}$

DS(A)= 2,68 puntos

La Desviación Estándar indica la distancia promedio que hay entre los valores observados y la media.

En promedio las notas del curso “A” se encuentran a 2,68 puntos con respecto a la nota promedio.

Curso “A” (N=5)	Curso “B” (N=5)
4 4 7 10 10	6 6 7 8 8
Media(A)= 7 puntos	Media (B)= 7 puntos
DS(A)= 2,68	**DS(B)= 0,89**

En el curso “B” el rendimiento es más parejo (las notas son más homogéneas, es decir que presentan menor variabilidad)

Aquel grupo con menor Desviación Standar, es el grupo en el cual los valores se encuentran a menor distancia con respecto a la media, y en consecuencia es el que presenta valores más homogéneos, lo que equivale a decir que los valores presentan menor variabilidad, por lo tanto en dicho grupo la media es más representativa.

Dos condiciones para utilizar la Desviación Estándar:

Los grupos a comparar deben tener:

- Igual Media

  y

- La variable bajo estudio expresada en la misma unidad de medida.

Si no se cumplen **ambas** condiciones entonces debemos utilizar el Coeficiente de Variación para efectuar la comparación entre dos o más grupos de datos.

La Desviación Estándar permite comparar dos o más grupos de datos si poseen la misma Media Aritmética **y** tienen la variable bajo estudio expresada en la misma unidad de medida. Si no se cumplen ambas condiciones, entonces debemos utilizar el Coeficiente de Variación para comparar los grupos de datos. Aquel grupo en el cual se obtiene el menor Coeficiente de Variación, es el que posee valores más homogéneos.

Coeficiente de Variación: Cociente entre Desviación Standar (DS) y Media Aritmética (Promedio)

CV= (DS / Media)*100 %

Por ejemplo:

CV= 13%

La valor de la Desviación Standar equivale al 13% del valor de la Media

Aquel grupo de datos con menor Coeficiente de Variación es el más homogéneo.

0% < CV < 11% Datos son muy homogéneos

11% ≤ CV < 16% Datos homogéneos

16% ≤ CV < 26% Datos heterogéneos

CV ≥ 26% Datos muy heterogéneos (presentan mucha variabilidad, son muy desparejos, muy diferentes entre si)

A medida que se incrementa el Coeficiente de Variación los valores incrementan su variabilidad se vuelven más desparejos, son menos homogéneos, es decir son más heterogéneos y la media será menos representativa de los valores observados.

Resumen de fórmulas:

Varianza Poblacional: σ^2

$$\sigma^2 = \frac{\sum(\mathbf{x} - \boldsymbol{\mu})^2}{N}$$

Desviación Poblacional: σ

$$\sigma = \sqrt{\frac{\sum(\mathbf{x} - \boldsymbol{\mu})^2}{N}}$$

Media poblacional: **μ : Media o promedio calculado a partir de una población.**

Varianza muestral:

$$\hat{s}^2 = \frac{\sum(\mathbf{x} - \overline{x})^2}{n-1}$$

Fórmula alternativa de cálculo:

$$\hat{s}^2 = \frac{\sum x^2 - n\overline{x}^2}{n-1}$$

$$\overline{x} = \frac{\sum x}{n}$$

Media muestral: Media o promedio calculado a partir de una muestra.

Ejemplo:

Fueron registrados los precios de 6 artículos seleccionados en un comercio "A" (La palabra "seleccionados" indica que los valores provienen de una muestra).

3 6 5 9 2 9

Calcular medidas de dispersión.

Se obtendrán las medidas de dispersión para los precios en una muestra de 6 artículos de un comercio “A”

n=6

$$\sum x = 3 + 6 + 5 + 9 + 2 + 9$$

$$\sum x = 34$$

$\bar{x}$ = **34 /** 6

$$\bar{x} = 5,667$$

$$\sum x^2 = 3^2 + 6^2 + 5^2 + 9^2 + 2^2 + 9^2$$

$$\sum x^2 = 236$$

$$\hat{S}^2 = \frac{236 - 6 * 5,667^2}{6-1}$$

$\hat{S}^2$ = 8,66 Varianza a partir de una muestra.

Desviación Standar: Raíz cuadrada de la varianza:

$$\hat{S} = \sqrt{\frac{\sum x^2 - n\bar{x}^2}{n-1}}$$

$$\hat{S} = \sqrt{8,66}$$

$\hat{S}$ = 2,94

$\hat{S}$ simboliza Desviación Standar Muestral

Coeficiente de Variación: Cociente entre Desviación Standar y Media Aritmética (Promedio):

$$CV = \frac{\hat{S}}{\bar{x}} * 100$$

$$CV = \frac{2,94}{5,667} * 100$$

CV=51,87%

Los precios en la muestra de 6 artículos son muy heterogéneos es decir que presentan mucha variabilidad, son muy diferentes entre sí.

El valor de la **Desviación Standar** correspondiente al comercio "A" equivale al **51,87%** del valor de la **Media**

Si en otro comercio "B" la Desviación Standar es de $4,75 y la Media es de $12,54. ¿En cuál comercio los precios son más homogéneos?

¿Debe utilizarse la Desviación Estándar o el Coeficiente de Variación?

Los precios tienen distinta Media por lo tanto se utilizará el **Coeficiente de Variación** para efectuar la comparación. Aquel grupo con menor Coeficiente de Variación será el más homogéneo.

$$CV(B) = \frac{DS(B)}{Media(B)} * 100$$

$$= \frac{\$4,75}{\$12,54} * 100$$

CV(B)= 37,88%

Comercio "A"	Comercio "B"
Media = 5,66 Varianza = 8,66 DE = $\sqrt{8,66}$ DE = 2,94 Coeficiente de variación = $\frac{2,94}{5,66} * 100$	Media = 12,54 DE = 4,75 Coeficiente de variación = $\frac{4,75}{12,54}$*100
CV(A)= 51,87% > CV(B)= 37,88%	

En el comercio "B" se obtuvo menor Coeficiente de Variación, por lo tanto en dicho comercio los precios se encuentran a menor distancia con respecto al precio promedio, en consecuencia en el comercio "B" el precio promedio es más representativo y los precios son más homogéneos, es decir presentan menor variabilidad que en el comercio "A".

Medidas de Dispersión para datos agrupados en intervalos:

Ejemplo:

Dada la siguiente distribución correspondiente a las calificaciones en Estadística obtenidas por los estudiantes de un Curso "A" en cierto examen final:

y'_{i-1} - y'_i	n_i
20 - 30	3
30 - 40	6
40 - 50	5
50 - 60	7
60 - 70	10

70	-	80	29
80	-	90	12
90	-	100	8
			n=80

Se obtendrá la Varianza:

y'_{i-1}	-	y'_i	y_i	N_i	n_i	$y_i n_i$	$y^2_i n_i$
20	-	30	25	3	3	75	1875
30	-	40	35	9	6	210	7350
40	-	50	45	14	5	225	10125
50	-	60	55	21	7	385	21175
60	-	70	65	31	10	650	42250
70	-	80	75	60	29	2175	163125
80	-	90	85	72	12	1020	86700
90	-	100	95	80	8	760	72200
					n= 80	$\sum yi.ni$ =5500	$\sum yi^2 ni$ = **404800**

V(y)= $\frac{\sum yi^2 ni}{n}$ - $[M(y)]^2$

1) Por cada intervalo se obtiene: $yi^2 ni$

2) Se suman los productos: $\sum yi^2 ni$

3) A la sumatoria se la divide por n: $\frac{\sum yi^2 .ni}{n}$

4) Al cociente resultante se le resta el cuadrado de la Media:

V(y)= $\frac{\sum yi^2 .ni}{n}$ - $[M(y)]^2$

$M(y)$= $\frac{\sum yi.ni}{n}$

$M(y)$= $\frac{5500}{80}$

$M(y) = 68,75$

$$V(y)_A = \frac{404800}{80} - (68{,}75)^2$$

$$= 5060 - 4726 = 333{,}44$$

$$DS(y)_A = \sqrt{\frac{\Sigma yi^2 . ni}{n} - [M(y)]^2}$$

$$DS(y)_A = \sqrt{\frac{404800}{80} - (68,75)^2}$$

$$= \sqrt{333{,}44}$$

$$DS(y)_A = 18{,}26$$

$$CV(y)_A = \frac{18{,}26}{68{,}75} * 100$$

$$= 0{,}27*100$$

$$CV(y)_A = 27\%$$

En el curso "A" la Desviación Estándar equivale al 27% del valor de la Media

En el curso "A" las notas son muy heterogéneas debido a que el Coeficiente de Variación es mayor que el 26%.

Si en otro Curso "B" la media fue de 65,4 y la Desviación Estándar de 15,3 ¿En cuál Instituto las calificaciones fueron más homogéneas?

Las notas de ambos cursos tienen distinta Media Aritmética, por lo tanto se utilizará el Coeficiente de Variación para realizar la comparación.

Curso "B"

$DS(y)_B = 15,3$

$M(y)_B = 65,4$

$CV(y)_B = \frac{DS(y)B}{M(y)B}$ *100

$CV(y)_B = \frac{15,3}{65,4}$ *100

= 0,23 *100

$CV(y)_B$ = 23%

En el curso "B" la Desviación Estándar equivale al 23% del valor de la Media

En el curso "B" las notas son heterogéneas debido a que el Coeficiente de Variación es está comprendido entre el 16% y el 26%.

$CV(y)_A > CV(y)_B$

En el curso B se obtuvo menor Coeficiente de Variación, por lo tanto en dicho curso las notas se encuentran a menor distancia con respecto a la Media, en consecuencia las notas son más homogéneas y la Media es más representativa de las notas.

BIBLIOGRAFÍA

Libros:

Espejo, I. Fernández, F. López, M. Muñoz, M. Rodríguez, A. Sánchez, A. Valero, C. (2009) Estadística Descriptiva y Probabilidad: (Teoría y problemas). Editorial: Cádiz Universidad de Cádiz, 2009. https://libros.metabiblioteca.org/handle/001/140

Ronald E. Walpole (2012) Probabilidad y estadística para ingeniería y ciencias, 9na. Edición Pearson Educación, México. ISBN: 978-607-32-1417-9.