Colegio Dr. B. A Houssay Educación Secundaria

Unidad N° 1: Medidas de Posición y Dispersión

Medidas de Posición y Dispersión

La tabulación y representación gráfica de los datos observados, es sin dudas el primer paso a seguir en el análisis e interpretación de un fenómeno. Ahora pasamos a otra etapa, que consiste en el **análisis y medición de datos**.

La comparación de distribuciones de frecuencias se hace relativamente fácil mediante los gráficos.

Pero cuando se trata de comparar tablas con muy abundantes datos, la tarea es compleja y poco clara.

Se observa generalmente una tendencia a agruparse alrededor de ciertos valores centrales llamados **medidas o parámetros de posición**. Los más conocidos son:

- La media aritmética (promedio)
- La mediana
- La moda

Estos parámetros caracterizan a la distribución y representan a los valores de la serie.

Pero para tener una idea más compleja de una distribución de frecuencias, además de los valores centrales, es necesario conocer la forma de **dispersión** de los datos, es decir, la desviación con respecto a los valores centrales.

Como el valor central más usado es el promedio los desvíos se miden con respecto a él.

Los parámetros de dispersión que analizaremos son:

- Varianza
- Desviación estándar

Medidas de Posición

a) Media aritmética (promedio)

Es el promedio de las observaciones. Esta medida se simboliza $\bar{x}$.

Para calcularla, pueden utilizarse las siguientes expresiones:

- ✓ Para distribuciones de *variables discretas*

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i \cdot f_i}{n}$$

f_i: es la frecuencia absoluta del intervalo

n: es el número total de observaciones

x_i : cada uno de los distintos valores que toma la variable

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + x_4 + \ldots + x_n$$

- ✓ Para distribuciones de ***variables continuas separadas en intervalos de clase***

$$\bar{x}_i = \frac{\sum_{i=1}^{k} x_{mi} \cdot f_i}{n}$$

x_{mi}: es la marca de clase

f_i: es la frecuencia absoluta del intervalo

k: la cantidad de intervalos de la distribución

n: total de observaciones

La marca de clase se calcula: $x_{mi} = \dfrac{\text{límite superior} + \text{límite inferior}}{2}$

b) La Moda

Es el valor de la variable que presenta la mayor frecuencia. Se simboliza $\mathbf{M_o}$

- ✓ Para distribuciones de *variables discretas*

  Es el dato que posee la mayor frecuencia.

- ✓ Para distribuciones de *variables continuas separadas en intervalos de clase.*

  Es la marca de clase correspondiente al intervalo de mayor frecuencia absoluta.

c) La Mediana

Es el valor correspondiente a la posición central de la distribución, cuyos datos están ordenados en forma creciente (de menor a mayor). Es el valor que divide a la distribución en dos partes de igual cantidad de observaciones. Se simboliza $\mathbf{M_e}$

- ✓ Para distribuciones de *variables discretas*

  - si ***n*** es impar, es el dato numérico central;
  - si ***n*** es par, es el promedio de los datos centrales.

- ✓ Para distribuciones de *variables continuas o con datos agrupados*.

$$Me = l_i + \left(\frac{\frac{n}{2} - f_{a\,m-1}}{f_m}\right) \cdot a$$

Donde:
l_i: es el límite inferior del intervalo de clase de la mediana.
n: tamaño de la muestra.
$f_{a\,m-1}$: es la frecuencia absoluta acumulada del intervalo anterior a la mediana.
f_m: es la frecuencia absoluta del intervalo correspondiente a la mediana
a: amplitud del intervalo

Ejemplo:

1) Sea la cantidad de asignaturas aprobadas por los estudiantes que ingresan a 2° año de una carrera universitaria. La variable es cuantitativa discreta.

x_i	f_i	f_r	f_{ai}	f_{ar}	$x_i \cdot f_i$
0	21				$0 \cdot 21 = 0$
1	29				$1 \cdot 29 = 29$
2	36				$2 \cdot 36 = 72$
3	30				
4	23				
5	13				
6	8				
7	13				
Total	173				

a) Media aritmética:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i \cdot f_i}{n} = \frac{487}{173} = 2{,}81$$

El promedio de asignaturas aprobadas es 2,81, esto es aproximadamente 3 materias aprobadas

b) Moda:

La mayor frecuencia absoluta es 36, por lo tanto el valor de la moda es: $M_o = 2$

Significa que la mayor cantidad de alumnos poseen 2 asignaturas aprobadas.

c) Mediana:

Como $n = 173$ (número impar) es el valor numérico central, es decir $\frac{n}{2} = \frac{173}{2} = 86{,}5$; por lo tanto la $M_e = 3$

2) Se le pregunto a 90 estudiantes elegidos en forma aleatoria, sobre el gasto realizado el día anterior (transporte, fotocopias, artículos de librería). Las respuestas fueron las siguientes:

8	8	8	12	12	12	20	20	19	25	25	27	44	33	17	17	7	6	18	18	18
24	25	15	15	15	16	16	16	10	10	9	9	19	19	8	20	22	23	23	40	32
30	29	28	17	17	13	13	13	16	16	15	15	15	23	23	21	14	14	21	10	10
12	11	12	12	21	21	21	17	28	28	17	17	18	18	18	14	8	15	15	18	20
27	25	30	40	11	14															

La tabla de frecuencia es:

intervalos	x_{mi}	f_i	fa_i	f_r	fa_r	$x_i \cdot f_i$
[6 ; 12)	9	15				9·15=135
[12 ; 18)	15	33				15·33=495
[18 ; 24)						
[24 ; 30)						
[30 ; 36)						
[36 ; 42)						
[42 , 48]						
Total						

Media aritmética (promedio)

$$\bar{x} = \frac{\sum_{i=1}^{7} x_{mi} \cdot f_i}{n} = \frac{1686}{90} = 18{,}73$$

el gasto promedio de los 90 estudiantes es de $ 18,73

La Mediana

1. Determinamos el intervalo que contiene la mediana, para ello se efectúa:

$$\frac{n}{2} = \frac{90}{2} = 45$$

   Se ubica en la columna de las frecuencias acumuladas fa_i quien es el primero contiene ese valor (45) . el intervalo buscado es el [12; 18).
2. Se obtienen el valor de la mediana

$$Me = l_i + \left(\frac{\frac{n}{2} - f_{a\,m-1}}{f_m} \right) \cdot a = 12 + \left(\frac{\frac{90}{2} - 15}{33} \right) \cdot 6 = 12 + \left(\frac{30}{33} \right) \cdot 6 = 17{,}45$$

$$Me = 17{,}45$$

Significa que a la izquierda de 17,45 se encuentran el 50% de las observaciones y hacia la derecha de ese valor el 50% de las observaciones restantes.

La Moda

1. Se obtiene el intervalo modal; el que presenta mayor frecuencia absoluta [12; 18).
2. El valor de la moda será la marca de clase de dicho intervalo

$$Mo = 16$$

ACTIVIDADES

1) La siguiente tabla muestra los valores de una encuesta realizada para conocer el número de veces que un estudiante de quinto año, revisa su celular en un día.

xi	f_i	f_{ai}	f_r	f_{rp}	f_{ar}	$x_i \cdot f_i$
24	5					
25	6					
26	9					
27	10					
28	2					
Total	32					

a) Completar la tabla.
b) ¿Cuál es la mayor cantidad de veces que un estudiante revisa su celular? ¿ Cómo se llama este parámetro?
c) Calcula la media aritmética
d) ¿Cuál es la mediana?

2) Se realizó un estudio acerca de los montos de ventas diarias, en pesos, de 25 comercios elegidos al azar. Los datos recopilados son los siguientes:

180	321	142	180	100
230	420	190	180	130
240	430	230	230	135
275	513	250	280	230
350	520	251	600	303

a) Completar la tabla.

intervalos	x_i	f_i	fa_i	f_r	fa_r	$x_i \cdot f_i$
[100; 184)		7				
[184; 268)		8				
[268; 352)		5				
[352; 436)		2				
[436; 520)		1				
[520; 604]		2				
Total		25				

b) ¿Cuál es la mayor cantidad de ventas diarias? ¿Cómo se llama este parámetro?
c) Calcula la media aritmética
d) ¿Cuál es la mediana?

Medidas de Dispersión

El uso de las medidas de dispersión, permite analizar un grupo de datos y obtener información en dos aspectos:

a) Descubrir las irregularidades que existen y resumirlas por medio de un valor, como puede ser el promedio, la moda y la mediana.
b) Establecer en qué medida, en un grupo de datos, estos se concentran o se dispersan alrededor de un valor típico, por ejemplo, el promedio.

La dispersión de un grupo de datos, con base en el promedio (valor típico) indica la confiabilidad de este valor, para inferir en relación con el resto de la población.

Por ejemplo, si la edad promedio en Argentina de la niñez que ingresa a primer grado de escuela, es de seis años y su dispersión es poca, esta información permite a las autoridades del Ministerio de Educación, diseñar contenidos temáticos dirigidos a niños de esta edad.

Por otro lado, si la dispersión es alta, tendríamos niños con edades diferentes al promedio de seis años, incluso, se puede tener niños con edades de 10 o más, lo que implicaría que el diseño de los contenidos temáticos iniciales, no se adaptaría de forma adecuada a los niños mayores.

Es importante señalar, que podemos encontrar diferentes grupos de datos con el mismo promedio, sin embargo, con diferente dispersión o variabilidad con respecto al valor típico (por ejemplo, al promedio).

Dentro de las medidas de dispersión más comunes, se tiene el recorrido, varianza, desviación estándar, dispersión relativa o coeficiente de variación.

1. Recorrido o amplitud total (R)

El recorrido es la diferencia entre el valor mayor y el valor menor del conjunto de datos.

Ejemplo 1: si tenemos los siguientes valores, que representan una muestra del tiempo en minutos que duran en ser atendidos seis clientes en un banco: 10, 12, 14, 15, 20, 25, el recorrido será:

R = 25 - 10 = 15 minutos

El recorrido presenta ciertas limitaciones que no hacen muy popular su uso, entre ellas, el no tomar en cuenta todos los datos, sea de la población o de la muestra, ya que su cálculo se realiza solamente tomando el valor mayor y el menor.

2. Varianza

La varianza es el promedio de la suma del cuadrado de las desviaciones de cada observación respecto de su media. Sus fórmulas para datos agrupados y datos no agrupados, son las siguientes:

Para la muestra
Datos no agrupados
$\boldsymbol{\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$
Datos agrupados
$\boldsymbol{\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot f_i}{n}}$

Para datos no agrupados

Se ilustra el procedimiento de cálculo, con los siguientes valores.

Ejemplo 2: Suponga que los datos se refieren a una muestra de los días que pernoctan los turistas que ingresan por el Aeropuerto Internacional de Ezeiza, de la ciudad de Buenos Aires, Argentina.

Días	f_i	$x_i \cdot f_i$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
2	1	2	-4,11	16,90
3	1	3	-3,11	9,68
4	1	4	-2,11	4,46
5	3	15	-1,11	1,23
8	1	9	1,89	3,57
10	1	10	3,89	15,12
12	1	12	5,89	34,68
Total	9	55		85,64

$$\bar{x} = \frac{\sum_{i=1}^{7} x_i \cdot f_i}{n} = \frac{55}{9} = 6,\hat{1}$$

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = \frac{85{,}64}{9} = 9{,}51 = 10$$

Interpretación: El promedio del cuadrado de las distancias entre cada observación y el promedio de muestra, es de 10. Si las observaciones se expresan en días, la varianza queda en términos de días al cuadrado.

3. Desviación Estándar

La desviación estándar es el promedio de las desviaciones de cada observación respecto de su media. Indica cuánto se alejan en promedio las observaciones, de la media del conjunto. Es la medida de dispersión más usada en estadística. Su fórmula es la raíz cuadrada de la varianza:

Para la muestra
Datos no agrupados
$\boldsymbol{\sigma = \sqrt[2]{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}}$
Datos agrupados
$\boldsymbol{\sigma = \sqrt[2]{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot f_i}{n}}}$

Ejemplo: El siguiente conjunto de datos: 2, 3, 4, 5, 5, 5, 8, 10, 12, del ejemplo 2 tiene varianza de 10 y su desviación estándar es de:

$$\sigma = \sqrt[2]{10} = 3{,}16$$

En promedio, el número de las noches que pernoctan los turistas que llegan al aeropuerto, se aleja 3,16 noches, a partir de la media.

Variancia en datos agrupados

Los datos u observaciones cuando se encuentran agrupados en una distribución de frecuencia, pierden su individualidad. Para calcular la varianza se requiere del cálculo previo de los puntos medios de cada clase, que vienen a ser el valor representativo de los datos u observaciones contenidas en cada clase. La frecuencia de cada clase representa el número de datos que se encuentran en cada una de las diferentes clases.

***Ejemplo 3*:** Una empresa de auditoría es contratada por el banco BNG para determinar la dispersión en el monto de los cheques que se cambian diariamente en la oficina principal. Al director de operaciones del banco, le preocupa tener una desviación estándar en el monto de cambio mayor a $200 ¿Debe preocuparse el director?

La siguiente distribución de frecuencia se presenta por la empresa auditora.

Obtenga la variancia y la desviación estándar.

Clase (monto del cheque) miles de dólares	Frecuencia absoluta f_i	Punto medio x_i	$x_i \cdot f_i$	$(x_i - \bar{x})^2 \cdot f_i$
[0 -199,5)	10	99,5	997,5	2 398 795,5
[199,5- 399,5)	13	299,5	3.893,5	1 093 488,5
[399,5 – 599,5)	17	499,5	8.491,5	137 776,5
[599,5 – 799,5)	42	699,5	29.379,0	507 969,0
[799,5 – 999,5]	18	899,5	16.191,0	1 729 521,0
Total	**100**		**58 952,5**	**5 867 550,5**

Calculamos el promedio:

$$\bar{x} = \frac{\sum_{i=1}^{5} x_i \cdot f_i}{n} = \frac{58952{,}5}{100} = 589{,}53$$

Calculamos la varianza:

$$\sigma^2 = \frac{\sum_{i=1}^{5} (x_i - \bar{x})^2 \cdot f_i}{n} = \frac{5867550{,}5}{100} = 58675{,}51$$

Desviación Estándar:

$$\sigma = \sqrt{58675{,}51} = 242{,}23$$

4. Dispersión Relativa o Coeficiente de variación (CV)

La dispersión relativa, permite comparar diferentes grupos de datos, en cuanto a su variabilidad, ya sea que estén expresados en la misma unidad de medida o en unidades distintas, o bien, exista una gran diferencia entre los promedios de los conjuntos de datos. Es el desvío promedio de las observaciones respecto del promedio. Se calcula dividiendo la desviación estándar entre el promedio.
Su fórmula es:

Para la muestra
$\boldsymbol{Cv = \frac{\sigma}{\bar{x}} \cdot 100\%}$

***Ejemplo 4*:** Una empresa dedicada al empaque de comida rápida le interesa medir la variabilidad de los diferentes turnos de trabajo. En una muestra de cinco horas para los turnos de trabajo diurno y nocturno, se obtuvo el siguiente conteo por hora:

Turno diurno: 700, 715, 699, 720, 718.

Turno nocturno 675, 695, 725, 740, 750.

Calcule el coeficiente de variación para cada turno de trabajo.

Turno Diurno	
x_i	$(x_i - \bar{x})^2$
699	129,96
700	108,16
715	21,16
718	57,76
720	92,16
Total: 3552	409,2

Turno Nocturno	
x_i	$(x_i - \bar{x})^2$
675	1764
695	484
725	64
740	529
750	1089
Total: 3585	3930

Turno Diurno	
Promedio	$\bar{x} = \frac{\sum_{i=1}^{5} x_i \cdot f_i}{n} = \frac{3552}{5} = 710,4$
Varianza	$\sigma^2 = \frac{\sum_{i=1}^{5}(x_i - \bar{x})^2}{n} = \frac{409,2}{5} = 81,84$
Desviación estándar	$\sigma = \sqrt{81,84} = 9,05$
Coeficiente de variación	$Cv = \frac{\sigma}{\bar{x}} \cdot 100\% = \frac{9,05}{710,4} \cdot 100\% = 1,27$

Turno Nocturno	
Promedio	$\bar{x} = \frac{\sum_{i=1}^{5} x_i \cdot f_i}{n} = \frac{3585}{5} = 717$
Varianza	$\sigma^2 = \frac{\sum_{i=1}^{5}(x_i - \bar{x})^2}{n} = \frac{3930}{5} = 786$
Desviación estándar	$\sigma = \sqrt{786} = 28,04$
Coeficiente de variación	$Cv = \frac{\sigma}{\bar{x}} \cdot 100\% = \frac{28,04}{717} \cdot 100\% = 3,91$

En el turno nocturno se presenta una mayor variabilidad porcentual.

Ejercicios

Ejercicio 1. Se realiza una encuesta de salida en el Aeropuerto Internacional Ezeiza, de la ciudad de Buenos Aires, con el objetivo de conocer el gasto realizado en dólares de los turistas con permanencia en Argentina. Se obtiene la siguiente información.

Clase (gastos en dólares)	f_i	x_i	$x_i \cdot f_i$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot f_i$
[500; 550)	12				
[550; 600)	14				
[600; 650)	18				
[650; 700)	10				
[700; 750]	6				
Total					

Calcular la varianza, desviación estándar y coeficiente de variación.

4Ejercicio 2. Según una muestra realizada por el Ministerio de Comercio Exterior, las importaciones de bienes y servicios en millones de dólares, por producto, para el año 2019, se obtuvo la siguiente distribución de frecuencias:

Importaciones en millones de dólares	Cantidad de productos f_i	x_i	$x_i \cdot f_i$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot f_i$
[5; 10)	10				
[10; 15)	12				
[15; 20)	18				
[20; 25)	22				
[25; 30)	24				
[30; 35]	26				
Total					

Calcular la varianza, desviación estándar y coeficiente de variación.

Ejercicio 3. Los datos siguientes corresponden a una muestra de las comisiones ganadas por corredores de bolsa de valores durante el mes de diciembre del 2019, en miles de dólares.

Miles de dólares	f_i	x_i	$x_i \cdot f_i$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot f_i$
[5; 10)	3				
[10; 15)	4				
[15; 20)	6				
[20; 25)	5				
[25; 30]	3				
Total					

Calcular la varianza, desviación estándar y coeficiente de variación