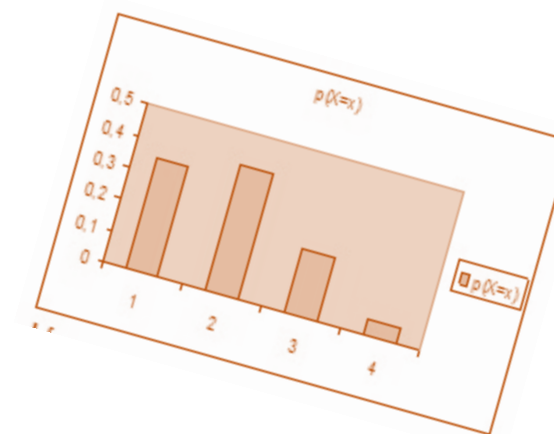
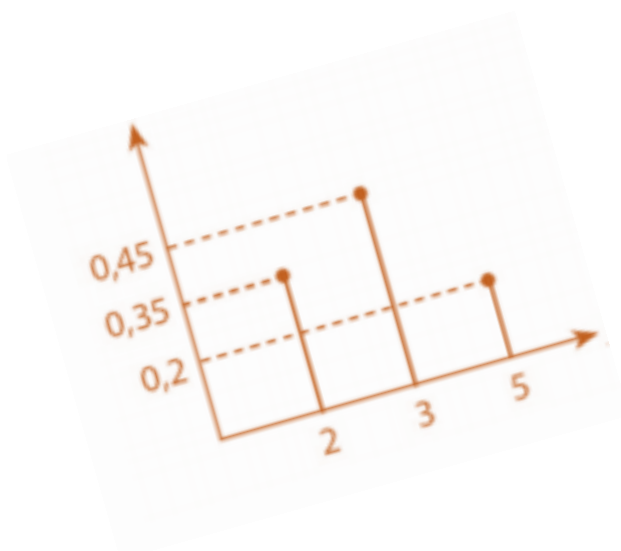


Medidas de Tendencia Central

Electivo Probabilidad y Estadística

Montserrat I. Guerrero Barra

3ero y 4to Medio



Definiciones Previas

Variable Estadística: Característica o atributo observable que se mide en una muestra.

Distribución de datos: Corresponden a los datos o valores de la variable de estudio.

Por ejemplo, si la variable de estudio es *Notas obtenidas por un estudiante en matemática*. La distribución de datos es: 2,0 – 3,4 – 7,0 – 2,7 – 1,0

Valor atípico: Dato numéricamente distante del resto de los valores de la distribución. Puede ser excesivamente más pequeño o excesivamente más grande que los demás datos.

Por ejemplo, si la distribución de datos es: 2,0 – 2,4 – 7,0 – 2,3 – 2,0. El valor atípico es el dato **7,0** porque es evidentemente más grande en comparación a los demás datos.

Datos no agrupados: Datos que generalmente no presentan un orden previo y se organizan en una tabla de frecuencias (*sin intervalos*). Estos datos se presentan con cualquier tipo de variable.

Datos agrupados: Datos que únicamente pueden ser organizados en una tabla de frecuencias para datos agrupados (*con intervalos*). Estos datos se presentan con variables cuantitativas.

Ver guía de Contenido N° 2: Conceptos básicos de Estadística

Medidas de Tendencia Central (MTC)

Valor numérico cuyo objetivo es resumir los datos de una distribución

Media Aritmética (\bar{X})

Promedio

Cociente (*División*) entre la suma de todos los datos y la cantidad total de datos

Moda (Mo)

Dato con mayor frecuencia absoluta dentro de la distribución.

Mediana (Me)

Dato central en una distribución ordenada creciente o decrecientemente.

Características de la Media Aritmética \bar{X}

1. *Es sensible a valores atípicos*

Ejemplo: Un estudiante en la asignatura de matemática obtuvo las siguientes calificaciones: 6,7 6,8 6,9 6,8 1,0 ¿Cuál es su promedio?

$$\bar{X} = \frac{6,7 + 6,8 + 6,9 + 6,8 + 1,0}{5} = 5,64$$

En esta distribución de notas, la calificación 1,0 es un valor atípico y provocó que el promedio del estudiante disminuyera.

2. *Es dudosa su aplicabilidad en variables cuantitativas discretas*

Se desprende por obviedad que el promedio no puede obtenerse al trabajar con variables cualitativas.

Ejemplo con variable cuantitativa discreta:

En una campo hay 5 aves, 3 patos, 8 perros y 3 gatos ¿Cuál es en promedio la cantidad de animales dentro del campo?

$$\text{Promedio} = \frac{5+3+8+3}{4} = 4,75$$

En esta distribución de cantidades de animales, no existen **4,75 animales** y por lo tanto el valor del promedio no tiene coherencia con esta variable

Características de la Moda (M_o)

Ejemplo: En un colegio se realizó una encuesta para determinar la cantidad de estudiantes afectados por bullying en cada curso y realizar una intervención inmediata con el curso con más afectado. La investigación arrojó los siguientes resultados:

1. *Una distribución de datos puede no tener moda (Distribución Amodal)*

Casos de <i>bullying</i>	
Curso	N.º de alumnos afectados (f)
7.º básico	2
8.º básico	2
1.º medio	2
2.º medio	2
3.º medio	2
4.º medio	2

Como las frecuencias absolutas de los cursos son iguales, no existe moda. Es decir, que como todos los cursos tienen la misma cantidad de estudiantes afectados por bullying, no existe un curso con mayor cantidad de afectados.

2. *Una distribución de datos puede tener dos modas o más (Distribución Bimodal o multimodal)*

Casos de <i>bullying</i>	
Curso	N.º de alumnos afectados (f)
7.º básico	9
8.º básico	9
1.º medio	2
2.º medio	0
3.º medio	3
4.º medio	6

Como las frecuencias absolutas en 7º y 8º básico son mayores e iguales, la distribución es bimodal. Es decir, que los cursos con mayor cantidad de afectados son 7º y 8º básico por lo tanto en ellos se hará la intervención.

Características de la Mediana (Me)

1. *Se necesita que los datos se ordenen de forma creciente o decreciente para determinarla*

Como la mediana es el dato central en una distribución, es necesario que los datos se organicen a partir de su valor para poder identificarlo. No importa si los datos se ordenan de forma creciente o decreciente, ya que el dato central será siempre el mismo. Por ejemplo, sea la siguiente distribución de datos: 10, 15, 12, 15, 10, 16, 10, 10, 11, 14, 15 ¿Cuál es la mediana?

Orden Creciente

10, 10, 10, 10, 11, **12**, 14, 15, 15, 15, 16

Orden Decreciente

16, 15, 15, 15, 14, **12**, 11, 10, 10, 10, 10

2. *Varía si la cantidad total de datos en par o impar*

Si la distribución tiene una cantidad **impar** de datos, existe un único dato central (*Situación anterior*). Sin embargo, cuando la distribución tiene una cantidad **par** de datos existen dos datos centrales y la mediana se obtiene a partir de promedio de ambos. Por ejemplo, sea la siguiente distribución de datos: 10, 15, 12, 15, 13, 10, 16, 10, 10, 11, 14, 15 ¿Cuál es la mediana?

Orden Creciente

10, 10, 10, 10, 11, **12, 13**, 14, 15, 15, 15, 16

$$\text{Mediana} = \frac{12 + 13}{2} = 12,5$$

Tablas de Frecuencia

N° de Mascotas	Frecuencia Absoluta f_n	Frecuencia Absoluta Acumulada	Frecuencia Relativa	Frecuencia Relativa Acumulada
2	3	3	12,5%	12,5%
3	10	13	41,7%	54,2%
4	7	20	29,2%	83,4%
5	2	22	8,3%	91,7%
6	2	24	8,3%	100%
Total	24		100%	

Tabla de Frecuencia para datos no agrupados

Estatura de los Estudiantes	Marca de Clase C_n	Frecuencia Absoluta f_n	Frecuencia Absoluta Acumulada	Frecuencia Relativa	Frecuencia Relativa Acumulada
[1,30 - 1,35)	1,325	12	12	24%	24%
[1,35 - 1,40)	1,375	16	28	32%	56%
[1,40 - 1,45)	1,425	12	40	24%	80%
[1,45 - 1,50)	1,475	10	50	20%	100%
Total		50		100%	

Tabla de Frecuencia para datos agrupados

Tipos de Frecuencias

Frecuencia Absoluta f_i
Corresponde a la cantidad de veces que se repite el dato en la variable.

Frecuencia Absoluta Acumulada
Es la suma de la Frecuencia Absoluta de cada dato con la Frecuencia Absoluta Acumulada del dato anterior.

Frecuencia Relativa
Cociente entre la frecuencia absoluta de cada dato y el tamaño de la muestra. Se puede expresar como decimal, fracción o porcentaje.

Frecuencia Relativa Acumulada
Es la suma de la Frecuencia Relativa de cada dato con la Frecuencia Relativa Acumulada del dato anterior.

Variable Estadística
Aquí van los datos de la variable: 2 libros, 3 libros, etc.

Nº de Libros	Frecuencia Absoluta f_i	Frecuencia Absoluta Acumulada	Frecuencia Relativa	Frecuencia Relativa Acumulada
2	3	3	$\frac{3}{24} = 0,124 = 12,5\%$	12,5%
3	10	3 + 10 = 13	$\frac{10}{24} = 0,417 = 41,7\%$	54,2%
4	7	13 + 7 = 20	$\frac{7}{24} = 0,292 = 29,2\%$	83,4%
5	2	20 + 2 = 22	$\frac{2}{24} = 0,083 = 8,3\%$	91,7%
6	2	22 + 2 = 24	$\frac{2}{24} = 0,083 = 8,3\%$	100%

Significa que:
"20 estudiantes tienen como máximo 4 libros"

Significa que:
"2 estudiantes tienen 5 libros"

Significa que:
"El tamaño de la muestra/ total de datos es 24"

Significa que:
"El 41,7% de los estudiantes tienen 3 libros"

Significa que:
"El 91,7% de los estudiantes tienen a lo más 5 libros"

Cálculo de las MTC en tabla de frecuencia para datos **no** agrupados

La siguiente tabla presenta el tiempo que tardan 84 estudiantes en llegar a sus casas desde el liceo. Con la información proporcionada, calcula e interpreta las MTC.

Tiempo <i>minutos</i>	Frecuencia Absoluta	Frec. Abs. Acumulada	Frecuencia Relativa	Frec. Rel. Acumulada
14	8	8	9,5%	9,5%
15	45	53	53,6%	63,1%
16	23	76	27,4%	90,5%
17	8	84	9,5%	100%

Mediana

Como el dato central se ubica en la mitad de la distribución, basta con **identificar el dato** cuya Frecuencia Relativa Acumulada iguala o sobrepasa primero el 50%

En la tabla, el porcentaje que sobrepasa primero al 50% de los datos es 63,1% que corresponde al dato *15 minutos*.

Interpretación: El 50% de los estudiantes tardan como máximo *15 minutos* en llegar a sus casas desde el liceo.

Moda

El dato con mayor frecuencia absoluta es *15 minutos*, Interpretación: La moda es que los estudiantes tarden *15 minutos* en llegar a sus casas desde el liceo.

Media Aritmética

Si los datos se ordenan en tablas de frecuencia para datos no agrupados, la media aritmética se obtiene a partir de la fórmula:

$$\bar{X} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \dots}{n}$$

Simbología:
 x_i = Datos de la variable.
 f_i = Frec. Abs. de cada dato.
 n = Número total de datos

$$\bar{X} = \frac{14 \cdot 8 + 15 \cdot 45 + 16 \cdot 23 + 17 \cdot 8}{84} = 15,37$$

Interpretación: Los estudiantes tardar en promedio *15,37 minutos* en regresar a sus casas desde el liceo

Características de la tabla de frecuencia para datos agrupados

Marca de Clase (C):

Se obtiene al promediar los extremos de cada intervalo

En el I_1 la marca de clase es $C_1 = \frac{1,30 + 1,35}{2} = 1,325$

N° de Intervalo	Estatura de los Estudiantes	Marca de Clase C_i	Frecuencia Absoluta f_i	Frecuencia Absoluta Acumulada	Frecuencia Relativa	Frecuencia Relativa Acumulada
$I_1 \rightarrow$	[1,30 - 1,35)	1,325	12	12	24%	24%
$I_2 \rightarrow$	[1,35 - 1,40)	1,375	16	28	32%	56%
$I_3 \rightarrow$	[1,40 - 1,45)	1,425	12	40	24%	80%
$I_4 \rightarrow$	[1,45 - 1,50)	1,475	10	50	20%	100%
	Total		50		100%	

Límites del Intervalo

Para el I_1 el dato 1,35 corresponde a su *Límite Superior* L_{S1} y el dato 1,30 corresponde a su *Límite Inferior* L_{I1}

Amplitud de Intervalo (A)

Corresponde a la diferencia entre el límite superior e inferior de cada intervalo y coincide para todos.

En I_1 su amplitud es $1,35 - 1,30 = 0,05$

En I_2 su amplitud es $1,40 - 1,35 = 0,05$

Cálculo de la Media Aritmética en tablas de frecuencias de datos agrupados

Media Aritmética

$$\bar{X} = \frac{C_1 \cdot f_1 + C_2 \cdot f_2 + \dots}{n}$$

Simbología

C_i : Marca de Clase

f_i : Frecuencia Absoluta

n : Cantidad total de datos (*Tamaño de la muestra*)

Estatura de los Estudiantes	Marca de Clase	Frecuencia Absoluta	Frecuencia Absoluta Acumulada	Frecuencia Relativa	Frecuencia Relativa Acumulada
[1,30 - 1,35)	1,325	12	12	24%	24%
[1,35 - 1,40)	1,375	16	28	32%	56%
[1,40 - 1,45)	1,425	12	40	24%	80%
[1,45 - 1,50)	1,475	10	50	20%	100%
Total		50		100%	

En esta tabla de frecuencias hay 4 intervalos y por lo tanto existen 4 marcas de clase, cada una con su propia frecuencia absoluta. Para calcular la media aritmética, basta solo con identificar las marcas de clase, sus frecuencias absolutas y la cantidad total de datos, para luego reemplazarlos en la fórmula y operar.

$$\bar{X} = \frac{1,325 \cdot 12 + 1,375 \cdot 16 + 1,425 \cdot 12 + 1,475 \cdot 10}{50} = 1,395$$

Interpretación: En promedio la estatura de los estudiantes que se midieron corresponde a 1,395 metros.

Cálculo de la Moda en tablas de frecuencias de datos agrupados

En primera instancia, es necesario tener presente la diferencia entre la moda y el intervalo modal.

- **Moda:** Corresponde al dato con mayor frecuencia absoluta (*El que más se repite*) y que se aplica con datos no agrupados.
- **Intervalo Modal:** Corresponde a aquel intervalo con mayor frecuencia absoluta y que tiene aplicabilidad con tablas de frecuencia para datos agrupados.

Estatura de los Estudiantes	Marca de Clase	Frecuencia Absoluta	Frecuencia Absoluta Acumulada	Frecuencia Relativa	Frecuencia Relativa Acumulada
[1,30 - 1,35)	1,325	12	12	24%	24%
[1,35 - 1,40)	1,375	16	28	32%	56%
[1,40 - 1,45)	1,425	12	40	24%	80%
[1,45 - 1,50)	1,475	10	50	20%	100%
Total		50		100%	

En esta tabla de frecuencias, existe solo un intervalo con mayor frecuencia absoluta y es el segundo (I_2).

Interpretación: La moda es que la estatura de los estudiantes oscile entre los 1,35 metros y 1,40 metros.
(*Distribución Unimodal*)

Cálculo de la Mediana en tablas de frecuencias de datos agrupados

Mediana

$$Me = L_{Ii} + \frac{\frac{n}{2} - F_{i-1}}{f_i} \cdot A$$

El Intervalo donde se encuentra la mediana es aquel cuya Frec. Relativa Acumulada iguala o sobrepasa primero el 50%. En este caso, es el segundo I_2 : [1,35 - 1,40) ya que su frecuencia absoluta acumulada es 56%

$$Me = 1,40 + \frac{50 - 12}{16} \cdot 0,05$$

$$Me = 1,40 + \frac{13}{16} \cdot 0,05 = 1,44$$

Simbología

A: Amplitud de Intervalo

n: Cantidad total de datos (*Tamaño de la muestra*)

i: Intervalo donde se encuentra la mediana (Intervalo i-ésimo)

L_{Ii} : Límite inferior del intervalo i-ésimo.

f_i : Frecuencia absoluta del intervalo i-ésimo.

F_{i-1} : Frecuencia absoluta acumulada del intervalo anterior al i-ésimo

Estatura de los Estudiantes	Marca de Clase	Frecuencia Absoluta	Frecuencia Absoluta Acumulada	Frecuencia Relativa	Frecuencia Relativa Acumulada
[1,30 - 1,35)	1,325	12	12	24%	24%
[1,35 - 1,40)	1,375	16	28	32%	56%
[1,40 - 1,45)	1,425	12	40	24%	80%
[1,45 - 1,50)	1,475	10	50	20%	100%
Total		50		100%	

Interpretación: La estatura del 50% de los estudiantes es menor o igual a 1,44 metros.