

# Probabilidades y estadística descriptiva e inferencial

TITULO DE LA UNIDAD 1: ¿Qué dicen los gráficos? Análisis crítico de la información

OA 1. Argumentar y comunicar decisiones a partir del análisis crítico de información presente en histogramas, polígonos de frecuencia, frecuencia acumulada, diagramas de cajón y nube de puntos, incluyendo el uso de herramientas digitales.

# Estadística

- ▶ La **estadística** es una ciencia encargada de reunir, clasificar, describir, analizar y presentar datos numéricos a partir de hechos reales que permitirán inferir y tomar, a partir de ellos, conclusiones y decisiones lógicamente aceptables.
- ▶ Se clasifica en descriptiva e inductiva

# Estadística Descriptiva

- ▶ En la **estadística descriptiva**, el objetivo principal es la recopilación y tabulación de datos que utilizando un gráfico adecuado permita publicar los resultados

# Estadística inductiva o inferencial

- ▶ En la **estadística inductiva o inferencial** se estudia, analiza con espíritu científico, todos los datos, tablas y gráficos obtenidos a partir de una muestra para extraer conclusiones **probablemente válidas** para la **población**

# CONCEPTOS BASICOS DE ESTADISTICA

- ▶ **Población o universo:** Es el conjunto de todos los individuos u objetos que tienen alguna característica común observable (Formado Por todos los elementos cuyo conocimiento nos interesa. Cada elemento se denomina individuo)
- ▶ **Muestra:** Es un subconjunto finito de la población en estudio. Se usa una muestra cuando es imposible ( o poco práctico) estudiar a todos los individuos de una población.
- ▶ **Variable estadística:** Es la **característica o atributo** a observar
- ▶ **Dato:** Es el **conjunto de valores** asignados a la variable. Pueden ser **cualitativos** ( color, profesión, sexo, raza, etc) o **cuantitativos** ( altura, peso, edad, etc).

# Ejemplo.

## Ejemplo 1:

- ▶ Población: Animales del zoológico de Santiago
- ▶ Muestra: Aves
- ▶ Variable: Color del plumaje
- ▶ Datos: rojo, amarillo, negro, azul, etc

## Ejemplo 2:

- ▶ Población: Alumnas LMDE
- ▶ Muestra: Enseñanza Media
- ▶ Variable: Edad
- ▶ Datos: 15,16,17,18

# Tipos de Datos

- ▶ **Cualitativos:** son datos que solo toman valores asociados a las cualidades o atributos, clasificándolos en una de varias categorías, es decir, no son valores numéricos. Ej:
  - ▶ Sexo: f/m.
  - ▶ Hábito de fumar: Fumador/No fumador
  - ▶ Color de ojos: negro, azul, marrón, ...
  - ▶ Religión: católica, evangélica, ...
  - ▶ Estado civil: soltero, casado, divorciado,...

# Datos Cuantitativos

- ▶ Un **dato cuantitativo** se denomina **discreto** si sólo puede tomar **determinados valores** Ej: número de personas, cantidad de manzanas etc. o **continuo** cuando puede tomar **cualquier valor dentro de un intervalo**. Ej: altura, peso.
- ▶ Se llama **estadígrafo** al conjunto de **datos**.
- ▶ Los datos pueden obtenerse en forma **directa** por medio de encuestas, censos etc. O **indirecta** a través de periódicos, revistas especializadas, oficinas de estadísticas (I.N.E.) instituciones internacionales (O.N.U., UNESCO, etc)

# ORGANIZACIÓN DE DATOS

- ▶ En una encuesta se obtienen muchos datos, generalmente desordenados. Es necesario ordenarlos de menor a mayor o vice-versa. La diferencia entre el mayor y el menor de los datos de un conjunto se llama **rango**
- ▶ A veces se prefiere agruparlos según su magnitud en cierto número de intervalos de clase. Esta agrupación se conoce como **distribución de frecuencias**.

# Ejemplo:

- ▶ Los siguientes datos corresponden a los pesos en kilogramos de 33 niños

35	26	62	38	44	35	62	38	52	28	33
54	33	26	44	62	54	50	60	28	34	50
60	38	62	52	62	38	44	54	60	33	50

- ▶ Si las ordenamos de menor a mayor anotando la frecuencia de cada valor, tenemos:

Variable Est. peso(Kg)	Frecuencia absoluta	Frecuencia acumulada
26	2	2
28	2	4
33	3	7
35	2	9
38	4	13
44	3	16
50	3	19
52	2	21
54	4	25
60	3	28
62	5	33
<b>n</b>	<b>33</b>	

▶ **Frecuencia absoluta ( $f_i$ ):**

- ▶ Indica el número de veces que se repite el valor de una variable

▶ **Frecuencia acumulada ( $f_a$ ):**

- ▶ Como su nombre lo indica acumula los valores de la frecuencia absoluta (sumando el anterior). Indica la cantidad de observaciones menor o igual al valor considerado. En la tabla anterior podemos decir que 7 alumnas tienen un peso Menor o igual a 33 Kg

▶ **Frecuencia relativa ( $f_r$ ):**

- ▶ Es la comparación entre la frecuencia absoluta y el número (n) total de individuos de la población. La suma de las frecuencias relativas es 1. Se puede expresar en porcentajes y será **la frecuencia relativa porcentual**. Se obtiene **multiplicando por 100** la frecuencia relativa. La suma de las frecuencias relativas porcentuales es 100.

# Ejemplo

- ▶ Si anotamos los resultados de la última prueba de geometría obtenida por el curso (aproximando al entero)tenemos:

4 6 3 2 3 4 3 5 6 4 3 5 4 3 5 6 6 4 3 4 6 7 7 = 23 alumnas

Ordenando de menor a mayor y agrupando tenemos:

variable notas	Frecuencia absoluta ( $f_i$ )	F. acumulada ( $F_a$ )	F. relativa ( $f_r$ )	F. relativa porcentual
2	1	1	1/23	4,34 %
3	6	7	6/23	26,08 %
4	6	13	6/23	26,08 %
5	3	16	3/23	13,04 %
6	5	21	5/23	21,73 %
7	2	23	2/23	8,69 %

**TOTAL n = 23**

**1**

**100%**

# DATOS AGRUPADOS EN INTERVALOS

- ▶ Hay algunas variables que proporcionan datos muy distintos, entonces conviene agruparlos en **intervalos o clases**.
- ▶ En el ejemplo de las notas, si no se hubiera aproximado podríamos obtener 18 datos diferentes, entonces es conveniente hacer un intervalo de notas. Ej entre:

1-2 Se incluyen a todas las alumnas que obtienen entre 1,0 a 1,9

2-3 Se incluyen a todas las alumnas que obtienen entre 2,0 a 2,9

3-4 Se incluyen a todas las alumnas que obtienen entre 3,0 a 3,9 etc

- ▶ Se llama **marca de clase** al **promedio** de los **límites** de un **intervalo de clase**
- ▶ Así en nuestro ejemplo de las notas la marca de clase del primer intervalo será  $\frac{1+2}{2} = 1,5$

# Ejemplo Tabla de frecuencia con intervalos

Intervalos	Frecuencia absoluta $f_i$	Frecuencia absoluta acumulada $F_i$	Frecuencia relativa $f_r$
[5 - 5,5)	1	1	0,04
[5,5 - 6)	2	3	0,08
[6 - 6,5)	3	6	0,13
[6,5 - 7)	4	10	0,17
[7 - 7,5)	8	18	0,33
[7,5 - 8)	1	19	0,04
[8 - 8,5)	5	24	0,21
<b>Total</b>	<b>24</b>		<b>1,00</b>

# MEDIDAS DE TENDENCIA CENTRAL

- ▶ Son valores que caracterizan a una muestra. Indican la **tendencia** de los datos a concentrarse en torno a un valor central. Las medidas de tendencia central son: **media aritmética, mediana y moda.**

# MEDIA ARITMÉTICA

- ▶ Es el **promedio** entre los valores de la variable (Se suman todos los valores y se divide por el total) Como Ud. calcula su promedio de notas.

# MEDIANA

- ▶ **Valor que ocupa el lugar central**, es decir que antes de él y después de él hay igual número de valores siempre que se ordenen de menor a mayor o vice-versa

# MODA

- ▶ Es el valor de la variable que tiene **mas frecuencia absoluta** o sea el que mas se repite.
- ▶ Existen casos en que se pueden encontrar mas de una moda, otros en que es *indeterminada* y en aquellos datos de intervalos de frecuencia la moda es la marca de clase del intervalo de mayor frecuencia. La moda sirve para describir una distribución si se desea obtener información sobre el punto donde se encuentra la mayor concentración de observaciones.

# Recordemos:

## Medidas de tendencia Central

Media Aritmética	Moda	Mediana
<p>Promedio. Se calcula sumando todos los datos y luego dividiendo por la cantidad total.</p>	<p>Dato que más se repite.</p>	<p>Dato que ocupa el lugar central. Se calcula ordenando de forma creciente o decreciente, dividiendo por dos y observando cual es el dato que esta ocupando la posición central.</p>

# Recordemos:

## Medidas de tendencia Central

Media Aritmética	Moda	Mediana
Multiplicamos el dato por su frecuencia absoluta. Y luego dividimos por el total de datos	Dato con mayor frecuencia absoluta	Dividimos la cantidad de datos por dos y luego observamos en que frecuencia acumulada esta contenido la frecuencia acumulada de cada dato

# Ejemplo:

DATO $x_i$	$f_i$	$F_i$
1	6	6
2	5	11
3	4	15
4	5	20
Total	20	

Media:

$$\bar{x} = \frac{1 \cdot 6 + 2 \cdot 5 + 3 \cdot 4 + 4 \cdot 5}{20}$$
$$\bar{x} = 2,4$$

Moda:

El que posee mayor frecuencia absoluta, en este caso 1

Mediana:

$N=20$  , por la tanto la mitad de los datos es 10, si buscamos ese dato 10 en la frecuencia acumulada concluimos que la mediana es 2

# Media Aritmética Para Datos Agrupados en Intervalos

- ▶ Primero debemos calcular la marca de clase de cada intervalo.
- ▶ Luego multiplicamos la marca de clase por su **frecuencia absoluta**
- ▶ Posteriormente debemos sumar cada producto
- ▶ Para finalmente dividir dicha suma por la cantidad total de datos

# Media Aritmética Para Datos Agrupados en Intervalos

- ▶ Observa la siguiente tabla correspondiente a la estatura de 25 alumnos y completa la tabla

Intervalo (cm)	$X_i$ (Marca de clase)	$f_i$	$f_i * X_{mc}$
150 - 155		3	
156 - 161		7	
162 - 167		6	
168 - 173		4	
174 - 179		5	

Responde:

- ▶ ¿Cuál es la variable de estudio?
- ▶ ¿Qué intervalo tiene mayor frecuencia?
- ▶ ¿Cuántos estudiantes miden más de 1.62 m.?
- ▶ ¿Cuántos estudiantes miden menos de 1.62 m.?
- ▶ ¿Qué porcentaje de estudiantes miden entre 1.62 m y 1.73 m?
- ▶ Calcular la media aritmética

# Mediana Para Datos Agrupados En Intervalos

$$Me = Li + \frac{\frac{N}{2} - F_{i-1}}{fi} \cdot a$$

- ▶ Donde  $Li$ : limite inferior del intervalo donde se encuentra la mediana.
- ▶  $a$ : Amplitud
- ▶  $N$ : Número total de datos
- ▶  $fi$ : Frecuencia absoluta del intervalo donde se encuentra la mediana.
- ▶  $F_{i-1}$ : Frecuencia absoluta acumulada del intervalo anterior al que se encuentra la mediana

# Mediana para datos agrupados en intervalos

$$Me = Li + \frac{\frac{N}{2} - F_{i-1}}{f_i} \cdot a$$

Intervalo (cm)	$X_i$ (Marca de clase)	$f_i$	$F_i$
150 - 155	152,5	3	3
156 - 161	158,5	7	10
162 - 167	164,5	6	16
168 - 173	170,5	4	20
174 - 179	176,5	5	25

$$Me = 162 + \frac{12,5 - 10}{6} \cdot 5$$

$$Me = 162 + 2,1$$

$$Me = 164,1$$

## Calcular Media Aritmética y Mediana

Edades	X Mc	fi	Fa	fi * X mc
16 - 20		18		
21 - 25		10		
26 - 30		15		
31 - 35		13		
36 - 40		12		
41 - 45		23		
46 - 50		5		

# Moda para datos agrupados en intervalos

$$Mo = Li + \frac{(f_i - f_{i-1})}{(f_i - f_{i-1}) + (f_{i+1} - f_i)} \cdot a$$

► Donde:

Li: Limite inferior del intervalo modal

fi: Frecuencia absoluta del intervalo modal

fi-1: Frecuencia absoluta del intervalo anterior al intervalo modal

fi+1: Frecuencia absoluta del intervalo siguiente al intervalo modal

a: Amplitud.

# Actividad:

- ▶ Completa la siguiente tabla y luego responde: La siguiente tabla muestra la distribución de frecuencias de las edades de 20 personas (agrupadas en intervalos).

Edades de personas en años	$X_{mc}$	$f_i$	$F_i$	$f_r$	$f_r \%$	$X_{mc} \cdot f_i$
10 - 20		2				
20 - 30		6				
30 - 40		4				
40 - 50		1				
50 - 60		1				
60 - 70		4				
70 - 80		2				

- ▶ Calcular la media aritmética, la mediana y moda.
- ▶ ¿Cuántas personas tienen al menos 40 Años?
- ▶ ¿Qué porcentaje de personas tiene entre 30 y 60 años?
- ▶ ¿Cuántas personas tienen a lo más 40 años?

# En resumen...

- ▶ Medidas de tendencia central: media aritmética, mediana y moda.

Media Aritmética	Mediana	Moda
<ul style="list-style-type: none"> <li>✓ Primero debemos calcular la marca de clase de cada intervalo.</li> <li>✓ Luego multiplicamos la marca de clase con su <b>frecuencia absoluta</b></li> <li>✓ Posteriormente debemos sumar cada producto</li> <li>✓ Para finalmente dividir dicha suma por la cantidad total de datos</li> </ul>	$Me = Li + \frac{\frac{N}{2} - F_{i-1}}{f_i} \cdot a$ <p>           Li: limite inferior del intervalo donde se encuentra la mediana.            a: Amplitud            N: Número total de datos            fi: Frecuencia absoluta del intervalo donde se encuentra la mediana.            F i-1: Frecuencia absoluta acumulada del intervalo anterior al que se encuentra la mediana         </p>	$Mo = Li + \frac{(f_i - f_{i-1})}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \cdot a$ <p>           Li: Limite inferior del intervalo modal            fi: Frecuencia absoluta del intervalo modal            fi-1: Frecuencia absoluta del intervalo anterior al intervalo modal            fi+1: Frecuencia absoluta del intervalo siguiente al intervalo modal            a: Amplitud.         </p>