

# Estadística y Probabilidad

Editorial Pearson



# Unidad 1: Introducción a la Estadística y la Probabilidad

## Aprendizajes Esperados:

- ✓ - Población, muestra, datos
- ✓ - Variables: cualitativas y cuantitativas; discretas y continuas
- Medidas de resumen
  - Medidas de tendencia central (media, mediana y modo)
  - Medidas de variación (rango, varianza, desviación estándar)
  - Cuantiles (cuartiles, deciles y percentiles)
- Métodos gráficos: gráfico de barras, gráfico de pastel, diagrama de puntos, histogramas, caja y bigote, ojiva
- Espacio muestral: finitos e infinitos discretos y continuos
- Tipos de eventos: seguros, imposibles, simples, compuestos, excluyentes
- Probabilidad y enfoques de probabilidad
- Los axiomas de Kolmogorov
- Leyes de la probabilidad
- Probabilidad condicional
- Independencia
- Teorema de Bayes
- Aplicaciones

## Conceptos Básicos

**Población:** Es el conjunto sobre el que estamos interesados en obtener conclusiones (hacer inferencia). Normalmente es demasiado grande para poder abarcarlo.

**Muestra:** Es un subconjunto de la población al que tenemos acceso y sobre el que realmente hacemos las observaciones (mediciones). Debería ser "representativo".



pues está formado por miembros "seleccionados" de la población (individuos, unidades experimentales)

Las conclusiones basadas en una muestra aleatoria son confiables

**Variable:** Característica observable que varía entre los diferentes individuos de una población

**Dato:** Valor particular de la variable

**Parámetro:** Cantidad numérica calculada sobre una población (como la altura media de los individuos de un país)

De este modo, la estadística sistematiza, ordena y presenta los datos referentes a un fenómeno que presenta variabilidad o incertidumbre para su estudio metódico

**Clasificación de Variables**

Las variables estadísticas se clasifican así:

- **Variables cualitativas:** Expresan características o cualidades y no pueden ser medidas con números

- **Variables cualitativas ordinales:** Presentan valores no numéricos pero existe un orden, como las medallas: oro, plata, bronce, o un estado de satisfacción: muy satisfecho, regular, insatisfecho

- **Variables cualitativas nominales:** Presenta valores no numéricos y no existe un orden, como el estado civil o el país de nacimiento

- **Variables cuantitativas:** Se expresan mediante un número

- **Variables cuantitativas discretas:** Puede asumir un número contable de valores, como el número de hijos en una familia, puede ser 0, 1, 2, 3, 4

- **Variables cuantitativas continuas:** Puede asumir un número incontable de valores, como la estatura de



una persona puede ser 1.784596 metros o 1.589617325 metros, existen infinitos, o incontables

## Medidas de Resumen

Sirven para describir en forma resumida un conjunto de datos que constituyen una muestra tomada de alguna población

### Medidas de tendencia central

#### Media

A grandes rasgos es el promedio de los datos. Cuando se tienen  $n$  datos, una fórmula que describe a la media es la siguiente:

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

Lo anterior es un promedio simple, en éste, se le dan el mismo valor a todos los elementos.

En un promedio ponderado, no todos los elementos valen lo mismo, por lo que se recurre a esta herramienta. Su fórmula es:

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

donde  $x$  son los elementos,  $w$  es el peso del elemento y  $\bar{x}$  es el promedio.

Por ejemplo, se tienen las siguientes materias con una calificación dada, pero cada materia vale ciertos créditos



Materia	Calificación	Número de créditos
Inglés	9	6
Estadística	7	6
Literatura	9	6
Psicología	10	5
Física	7	7
Química	6	7
Lógica	9	4

Bajo la fórmula, se realiza:

$$\bar{x} = \frac{54 + 42 + 54 + 50 + 49 + 42 + 36}{6 + 6 + 6 + 5 + 7 + 7 + 4}$$

$$\bar{x} = 7.97$$

A su vez, también se pueden trabajar promedios con datos agrupados. Para ello, se usa la siguiente fórmula:

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

donde  $f$  es la frecuencia de cada grupo,  $n$  es el total de datos y  $x$  es la marca de clase, es decir, definida por la fórmula:

$$x = \frac{Li + Ls}{2}$$

donde  $Li$  es el límite inferior de la clase y  $Ls$  es el límite superior de la clase.

A continuación se presenta un ejemplo del cálculo de una media de edades con datos agrupados:



Edad	f	x	f <sub>x</sub>
21-30	28	25.5	714
31-40	30	35.5	1,065
41-50	12	45.5	546
51-60	2	55.5	111
61-70	2	65.5	131
71-80	2	75.5	151

$$\Sigma f = 76$$

$$\Sigma f_x = 2,718$$

$$\bar{x} = \frac{2,718}{76} \rightarrow \bar{x} = 35.76$$

Mediana

Valor que se encuentra en medio. Divide a la población en dos. Es un valor posicional

Datos no agrupados

Se ordenan los datos de manera ascendente y la mediana es el valor que se encuentra en medio:

6, 7, 7, 9, 9, 9, 10

Si la cantidad de datos es par, se toman dos valores medios y se obtiene el promedio de ellos:

6, 7, 7, 8, 9, 9, 9, 10

8.5

Datos agrupados

Se obtiene por la fórmula:

$$\tilde{x} = L_i + \frac{\frac{n}{2} - f_{a(i-1)}}{f_i} \cdot a$$



donde:  $L_i$  es el límite inferior de la clase,  
 $n$  es el total de datos,  $f_{ai}$  es la frecuencia  
 acumulada de la clase anterior,  $f_i$  es la frecuencia  
 simple de la clase donde está la mediana y  $a$   
 es el ancho de la clase.

Tomando los datos del ejemplo del cálculo de  
 una media de edades con datos agrupados, el  
 cálculo de la mediana es el siguiente:

Edad	$f_i$	$f_{ai}$
21-30	28	28
31-40	30	58
41-50	12	70
51-60	2	72
61-70	2	74
71-80	2	76

$$n = 76 \therefore \frac{n}{2} = 38$$

Dato que estaría  
 en la clase 2

$\rightarrow i = \text{ renglón } 2$

$$L_i = 31, \quad f_i = 30, \quad a = 10$$

$$f_{ai-1} = 28$$

$$\tilde{x} = 31 + \frac{38 - 28}{30} \cdot 10$$

$$\tilde{x} = 34.333$$

**Moda**

Dato que se repite en mayor frecuencia:

6 7 7 8 9 9 9 10

1      2      1      3      1

Puede existir una población bimodal

Pueden existir poblaciones sin modas

**Datos no agrupados**

Se contabiliza por separado la frecuencia de  
 cada dato en la muestra y la frecuencia  
 mayor corresponde al dato que es la moda de  
 la población



Dato	Frecuencia
6	1
7	2
8	1
9	3
10	1

→ Moda

Datos agrupados

Se obtiene por la fórmula:

$$\hat{x} = Li + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot a$$

donde  $Li$  es el límite inferior de la clase donde se encuentra el dato de la moda,  $a$  es el ancho de clase,  $\Delta_1 = f_i - f_{i-1}$ ,  $\Delta_2 = f_i - f_{i+1}$  y  $f_i$  es la frecuencia de la clase donde está la moda

Ojo: Tratamos la fórmula para problemas con frecuencias simples

Retomando nuevamente los datos del ejemplo del cálculo de una media de edades con datos agrupados, el cálculo de la moda es el siguiente

Edad	Frecuencia
21-30	28
31-40	30
41-50	12
51-60	2
61-70	2
71-80	2