Describir la Data desde una variable

Dr. Joel De La Cruz Oller
2-2025-2026

Importante

Al final de esta presentación hay un enlace a una prueba corta de *Microsoft Forms*. Debe obtener al menos un 80% en la prueba para aprobarla.

Le exhortamos a que estudie el módulo con detenimiento antes de tomar la prueba.

Objetivos

Después de estudiar este módulo, los estudiantes podrán:

- Al finalizar el módulo, el estudiante podrá calcular e interpretar medidas de tendencia central (media, mediana, moda y media recortada) a partir de conjuntos de datos ordenados y explicar cómo los valores atípicos afectan cada medida.
- El estudiante será capaz de seleccionar y justificar la medida de tendencia central más apropiada según la forma de la distribución (simétrica, sesgada a la izquierda o derecha) y comparar media y mediana en ejemplos concretos.
- El estudiante construirá tablas de distribución de frecuencia para datos cuantitativos (definir rango, número y tamaño de clases), representará la información con histogramas, gráficas de barras o de pastel según corresponda, y leerá e interpretará frecuencias relativas y acumuladas.

Introducción

En casi todos los análisis de data, la distribución de frecuencia, gráficas de barras, gráficas de bizcocho o piza e histogramas pueden ser unas herramientas visuales informativas para examinar la imagen completa.

Pero no son muy exactas en el leguaje que usamos para describir las gráficas.

Si dijéramos una data es más compacta que otra, entonces: ¿Cuánto más compacta que a otra es?



Estadística

Medidas (análisis) que son aplicadas a data de la muestra son llamadas estadística.

Parámetro

Medidas (análisis) que son aplicadas a data de la población son llamadas parámetro.

Estadística Inferencial

Se refiere a realizar un buen acierto (inferencia) sobre los parámetros de la población usando estadística en la muestra.

CUADRO 1. Número y porcentaje de muertes, y tasas de mortalidad (por 100.000 habitantes) para las 10 principales causas de muerte, por sexo, Puerto Rico, 2008

Causas	Ambos Sexos			Hombres		Mujeres	
	Número	Porcentaje	Tasa	Número	Tasa	Número	Tasa
Todas	29.100	100,0	735,9	15.906	838,2	13.194	641,5
Enfermedades del corazón	5.356	18.4	135,4	2.883	151,9	2.473	120,2
Tumores malignos	5.008	17.2	126,6	2.842	149,8	2.166	105,3
Diabetes sacarina	2.852	9.8	72,1	1.401	73,8	1.451	70,5
Enfermedad de Alzheimer	1.591	5.5	40,2	537	28,3	1.054	51,2
Enfermedades cerebrovasculares	1.529	5.3	38,7	699	36,8	830	40,4
Enfermedades crónicas de las vías respiratorias inferiores	1.200	4.1	30,3	588	31,0	612	29,8
Accidentes	1.097	3.8	27,7	842	44,4	255	12,4
Nefritis, nefrosis y síndrome nefrótico	1.058	3.6	26,8	601	31,7	457	22,2
Influenza y neumonía	949	3.3	24,0	488	25,7	461	22,4
Septicemia	861	3,0	21,8	450	23,7	411	20,0

Fuente: Referencia (*13*).

Media aritmética (Arithmetic Mean)

Supongamos que hay n observaciones en un conjunto de datos, estos consisten de las observaciones $x_1, x_2, \ldots, x_n$;

Entonces la media aritmética está definida por

$$\frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

Muestra de la Media aritmética (Arithmetic Mean)

$$\bar{X} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

Para la población se usa :

$$\mu = \frac{1}{N}(x_1 + x_2 + \cdots + x_n)$$

Practiquemos ejercicios

Saque la media de la muestra de

1) 7, 2, 4, 9, 2

$$\overline{X} = \frac{1}{5}(7 + 2 + 4 + 9 + 2) = \frac{1}{5}(24) = 4.8$$

2) 8, 7, 9,3, 5

$$\overline{X} = \frac{1}{5}(8 + 7 + 9 + 3 + 5) = \frac{1}{5}(32) = 6.4$$

Practiquemos ejercicios

Saque la media de la población de:

1) 7, 2, 4, 9, 2, 8, 7, 9,3, 5

$$\mu = \frac{1}{10}(7 + 2 + 4 + 9 + 2 + 8 + 7 + 9 + 3 + 5) = \frac{1}{10}(56) = 5.6$$

Repaso de la Media aritmética

La media, llamada también mediana aritmética, es la medida de tendencia central conocida popularmente como “promedio”. Se define como la suma de todos esos valores dividida por el número de ellos.

Repaso de la Media aritmética

Media

Sean $x_1, x_2, x_3, \ldots, x_{n-1}, x_n$ los n valores observados para una variable cuantitativa X. Entonces la media aritmética o promedio de la variable X, que se denota con una barra encima de $\overline{X}$, es:

$$\overline{X} = \frac{x_1 + x_2 + x_3 + \cdots + x_{n-1} + x_n}{n}$$

En notación de sumatoria, la media aritmética se escribe:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n}(x_i)$$

Desviación

Dado que algún punto A y unos datos con punto x, entiendes x – A representa cuán lejos x se desvía de A.
Esta diferencia se llama la desviación.

Ejercicio

Tabla 1

Desviación de la mediana (xi) media = 5

Datos x_i	Desviación de la media $(x_i - 5)$
2	-3
11	6
4	-1
3	-2
5	0
	$\sum (x_i - 5) = 0$

Resistente

Medidas estadísticas la cuales no son afectada por los bigotes (“outliers”) se dice que son resistentes.
Estos son resistentes a las medidas

Mediana

La mediana de un grupo de observaciones son los datos arreglados en un orden que miden el valor del medio.
El mismo número de datos a cada lado de la mediana.

Cómo se saca la mediana

Mediana

- Se organizan los datos en orden y se cuenta la cantidad de datos.
- Si cuando cuentas te da un número par entonces la mediana es la suma de los valores del medio dividido por dos.
- Por lo contrario, si cuando sumas te da un número impar el número que está en el medio es la mediana.

Practiquemos para determinar la mediana

Veamos once observaciones 2,4,0,3,0,1, 8,5,1,5,9 encuentra la mediana

Ordene los datos de menor a mayor 0, 0, 1, 1, 2, 3, 4, 5, 5, 8, 9 = # 11

Como 11 es un número impar entonces la mediana esta en la posición 6

El número que está en la posición 6 es el número 3 entonces la mediana es 3

Practiquemos para determinar la mediana

Aquí tenemos diez notas 65, 98, 76, 83, 94, 79, 88, 72, 90, 85. Encuentra la mediana

Ordene los datos de menor a mayor 65, 72, 76, 79, 83, 85, 88, 90, 94, 98 = # 10

Como 10 es un número par entonces la mediana está en el promedio de los dos valores que están a la mitad del conjunto de datos en este caso 83 y 85.

Por lo tanto, la mediana es $\frac{83+85}{2} = \frac{168}{2} = 84$

Media recortada

La media recortada es una modificación de la media aritmética que ignora de los datos un porcentaje igual que son los valores más altos y el más bajo al calcular la media.

Cómo se saca la media recortada

Media recortada para un 10%

- Se organizan los datos en orden ascendente y se cuenta la cantidad de datos.
- Se elimina el 10 % con el valor más bajo.
- Se elimina el 10% con el valor más alto.
- Se calcula el valor de la media aritmética con el 80 % de la data restante.

Practiquemos para a determinar la media recortada a 10%

Consideremos los siguientes datos

16 18 20 21 23 23 24 32 36 42

media = 25.5 mediana = 23

18, 20,21,23,23,24,32,36

Media 24.6

La moda

La moda son los datos que con más frecuencia encontramos.

Practiquemos para determinar la moda

Encuentra la moda en los siguientes datos

5, 4, 3, 5, 2, 9, 9, 3, 2, 5, 7, 5

2,2,3,3,4,5,5,5,5,7,9,9

La Moda es 5

Resumen de medidas de tendencia central

Media:

Es la medida más común de tendencia central para datos discretos y continuos.

También llamada como promedio.

Es calculada sumando todas las observaciones y dividiendo entre el total de observaciones.

La media es sensible a los valores extremos o valores atípicos.

Resumen de medidas de tendencia central

Mediana:

Es determinada tomando el número del medio en una lista de valores ordenados de menor a mayor.

La mediana es una medida de tendencia central que no es sensible a valores atípicos comparado a la media.

Puede ser usado para sintetizar datos continuos o discretos

Resumen de medidas de tendencia central

Moda:

Es el valor mas frecuente en un set de datos.

Puede ser usada en la síntesis estadística para datos categóricos y ordinales, pero usualmente no es informativa para datos continuos o discretos ya que los valores únicos pueden ocurrir con baja frecuencia.

Relación entre la media y la mediana

- La medida de tendencia central más apropiada su uso depende en la distribución de los valores.
- Si la distribución de los datos es simétrico y unimodal, la media, mediana y moda serán iguales, la media es la mejor medida de tendencia central.
- Cuando los datos no son simétricos, la mediana es la mejor medida de tendencia central.
- En la distribución unimodal, cuando los datos están sesgados a la derecha, la media tiende a estar a la derecha de la mediana.

Relación entre la media y la mediana

- Cuando los datos están sesgados a la izquierda, la media tiende a estar a la izquierda de la mediana.

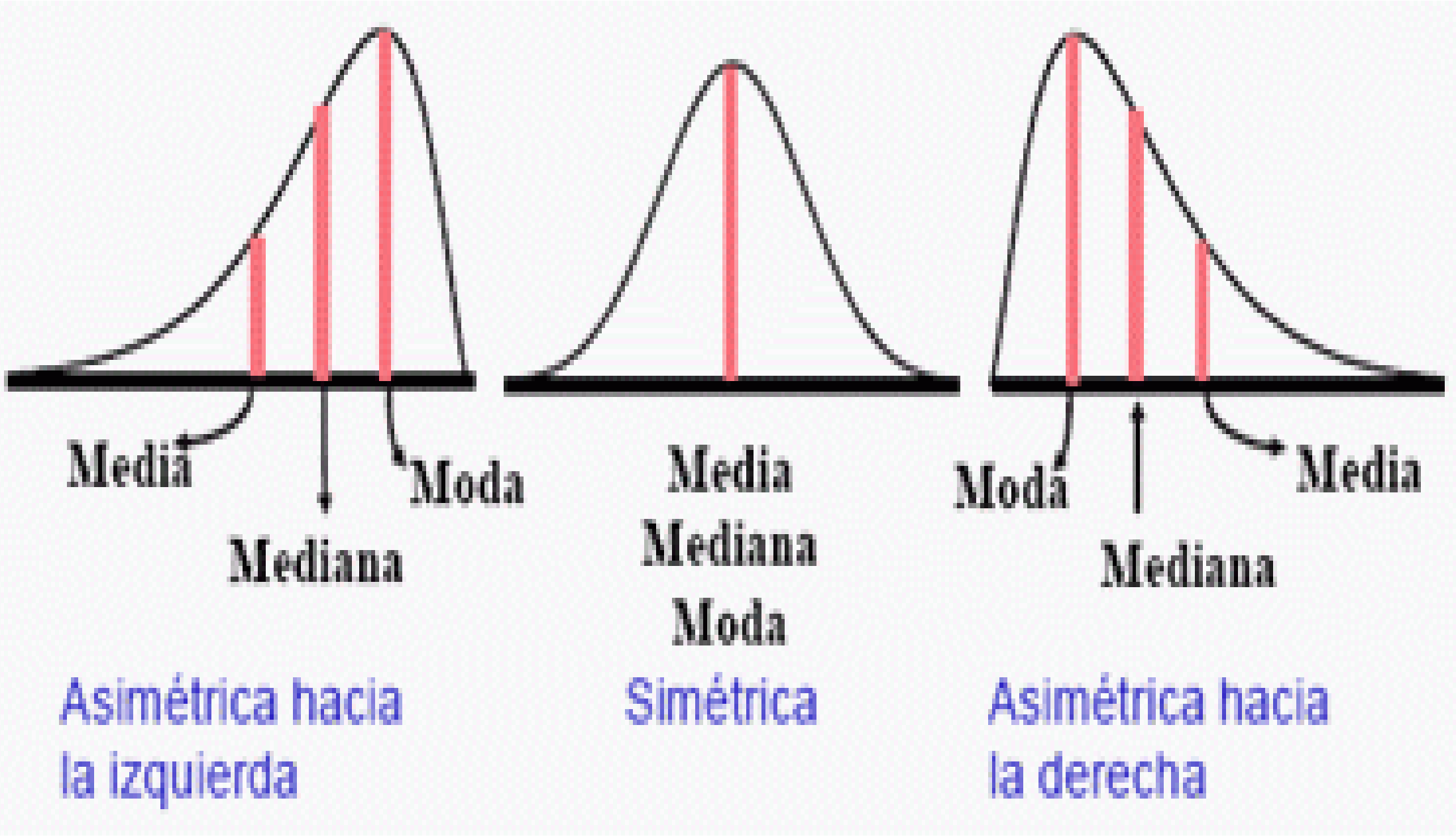

Media
Moda
Mediana
Media
Mediana
Moda
Moda
Mediana
Media
Asimétrica hacia la izquierda
Simétrica
Asimétrica hacia la derecha

Tabla 2 Aplicación de los niveles de medidas				
Medida de localización	Aplicable a los niveles de medida			
	Cualitativas		Cuantitativas	
	Nominal	Ordinal	Intervalo	Razón
Media			x	x
Mediana		x	x	x
Moda	x	x	x	x

Contenido

1. Introducción a la estadística descriptive
2. Estadística, parámetro y estadística inferencial
3. Medidas de tendencia central:
   - Media aritmética
   - Mediana
   - Moda
   - Media recortada
4. Desviación y valores atípicos (*outliers*)

Contenido

5. Distribución de los datos y relación entre la media y la mediana
6. Tablas de distribución de frecuencias
7. Representación gráfica e interpretación de los datos
8. Aplicación de las medidas según el nivel de medición
9. Ejercicios prácticos y repaso

Referencias

Illowsky, B., & Dean, S. (2023). Introductory statistics 2e. OpenStax. https://openstax.org/details/books/introductory-statistics-2e

National Institute of Standards and Technology. (s. f.). Measures of location. En NIST/SEMATECH e-Handbook of Statistical Methods. https://www.itl.nist.gov/div898/handbook/eda/section3/eda351.htm

Illowsky, B., & Dean, S. (s. f.). Statistics. OpenStax. https://openstax.org/books/statistics/pages/2-introduction

Para finalizar, haga clic sobre el enlace para tomar la prueba y enviar el informe a su profesor(a):

Describir la data desde una variable – Fill out form

¡Gracias por utilizar los recursos del CDCL !

Centro para el Desarrollo de Competencias Lingüísticas
Universidad de Puerto Rico, Recinto de Río Piedras

Conozca más sobre nuestros servicios:
http://generales.uprrp.edu/competencias-linguisticas/

Para revisiones: https://forms.office.com/r/f2pYYSQnHp