# INTELIGENCIA DE NEGOCIOS

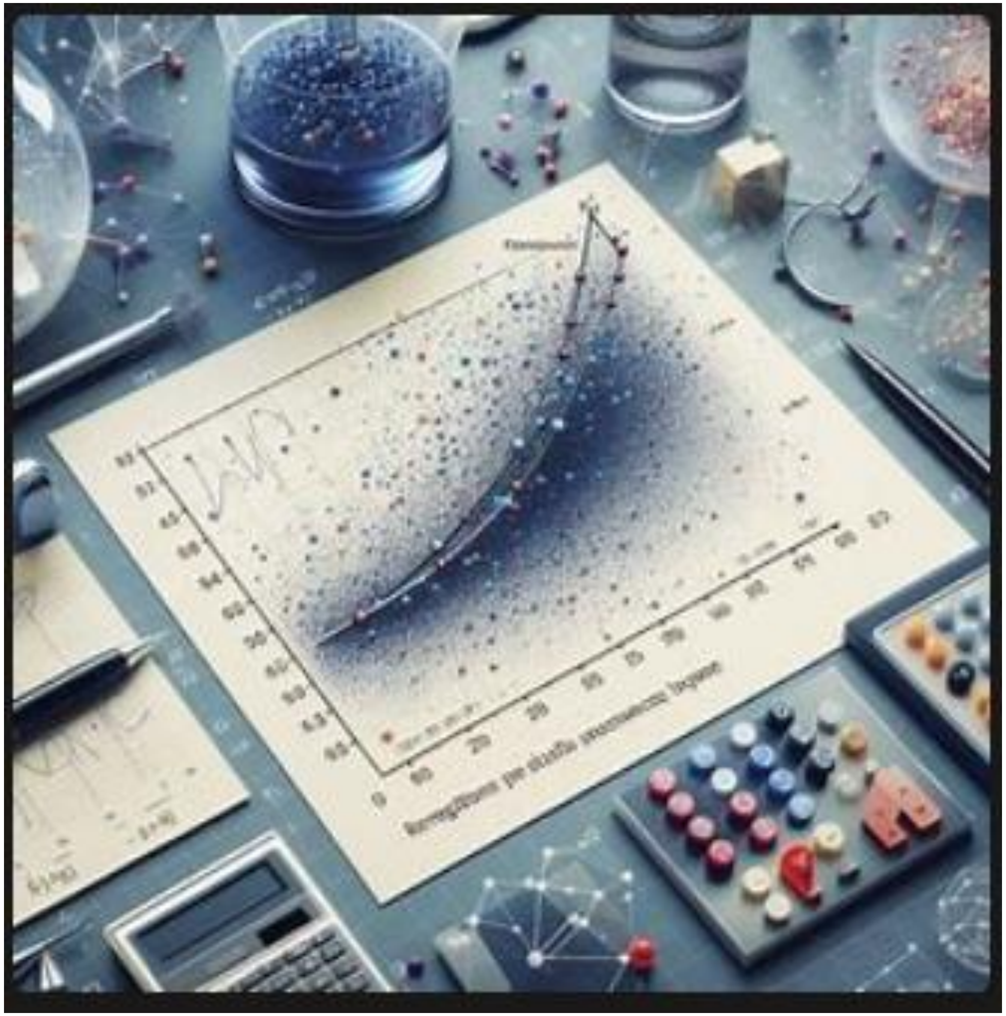
## Tema de la sesión: Medidas de tendencia central, Medidas de dispersión

Carrera: Diseño y Desarrollo de Software

# Capacidades Terminales

- Conocer e implementar estructura de datos bajo DataFrames y series.
- Conocer y aplicar las medidas de tendencia central
- Conocer y aplicar las Medidas de dispersión

# Motivación



# Contenido

- Introducción
- Estructura de datos en Pandas
- Medidas de tendencia central
- Medidas de dispersión
- Conclusiones
- Referencias Bibliográficas

# Introducción

- Es importante gestionar datos bajo una estructura apropiada que permita ser procesada para su posterior análisis y para ello es posible trabajar bajo la estructura DataFrame, Serie y Panel.
- La estructura de datos es fundamental para automatizar el proceso de manipular datos con Python.
- DataFrame es una estructura de datos tabular que se compone de columnas y filas ordenadas. Ejemplo, un diccionario de listas.
- Series es un objeto unidimensional (1D) similar a la columna de una tabla.
- Panel es una estructura en 3D.

# Estructura de datos en Pandas

- La estructura de datos básica de pandas se denomina DataFrame, que es una colección ordenada de columnas con nombres y tipos, similar a una tabla de base de datos, donde una sola fila representa un registro y las columnas representan atributos particulares o campos, sin embargo, existe también la estructura de Series que es un objeto unidimensional (1D) similar a la columna de una tabla y por último la estructura Panel es en 3D y la mayor estructura de datos.

Tecsup

# Medidas de tendencia central

- Se llama medidas de posición, tendencia central o centralización a unos valores numéricos en torno a los cuales se agrupan, en mayor o menor medida, los valores de una variable estadística. Estas medidas se conocen también como promedios.

- Para que un valor pueda ser considerado promedio, debe cumplirse que esté situado entre el menor y el mayor de la serie y que su cálculo y utilización resulten sencillos en términos matemáticos.

# Medidas de tendencia central

Las medidas de centralización son parámetros representativos de distribuciones de frecuencia como:

- **Media aritmética:** Se define media aritmética de una serie de valores como el resultado producido al sumar todos ellos y dividir la suma por el número total de valores.

# Ejemplo media

- Valores:

**10, 12, 8, 15, 9, 13, 10**

$$\text{Media} = \frac{(10 + 12 + 8 + 15 + 9 + 13 + 10)}{7}$$

Media = 11

# Medidas de tendencia central

Las medidas de centralización son parámetros representativos de distribuciones de frecuencia como:

- **Mediana:** La media aritmética no siempre es representativa de una serie estadística. Para complementarla, se utiliza un valor numérico conocido como mediana o valor central.

- **Moda:** En una serie de valores a los que se asocia una frecuencia, se define moda como el valor de la variable que posee una frecuencia mayor que los restantes.

Tecsup

# Ejemplo mediana

Mediana cuando la cantidad de números son impares:

Valores:

**10 12 8 15 9 13 10**

Mediana:

~~8~~ ~~9~~ ~~10~~ 10 ~~12~~ ~~13~~ ~~15~~

# Ejemplo mediana

Mediana cuando la cantidad de números son impares:

Valores: **10 12 8 15 9 13 10 14**

Mediana:

~~8~~ ~~9~~ ~~10~~ 10 12 ~~13~~ ~~14~~ ~~15~~

$$\text{Mediana} = \frac{(10 + 12)}{2}$$

# Ejemplo Moda

- Valores:

**10 12 8 15 9 13 10**

Moda:

**10 12 8 15 9 13 10**

Moda: 10

# Medidas de dispersión

Las medidas de dispersión son parámetros estadísticos que indican como se alejan los datos respecto de la media aritmética. Sirven como indicador de la variabilidad de los datos. Las medidas de dispersión más utilizadas son el rango, la desviación estándar y la varianza:

# Medidas de dispersión

- **Rango:** Indica la dispersión entre los valores extremos de una variable. se calcula como la diferencia entre el mayor y el menor valor de la variable. Se denota como R.

  Se calcula como:

  $$\mathbf{R = x_{(n)} - x_{(1)}}$$

  Donde:

  $x_{(n)}$: Es el mayor valor de la variable.

  $x_{(1)}$: Es el menor valor de la variable.

Tecsup

# Ejemplo Rango

- Valores:

**10 12 <u>8</u> <u>15</u> 9 13 10**

Rango:

Mayor = 15

Menor = 8

R = 15 – 8

R = 7

# Medidas de dispersión

- **Varianza:** La varianza es una medida de dispersión que representa la variabilidad de una serie de datos con respecto a su media.
- Es la suma de los cuadrados de los residuos dividida por las observaciones totales:

Varianza población

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{N}$$

Varianza muestra

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

# Ejemplo Varianza

- Valores:

**10 12 9 13 10**

$$\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{N}$$

- Se calcula la media:

$$\bar{x} = \frac{(10 + 12 + 9 + 13 + 10)}{5}$$

$$\bar{x} = 10.8$$

- Se calcula la varianza:

$$\sigma^2 = \frac{(10-10.8)^2 + (12-10.8)^2 + (9-10.8)^2 + (13-10.8)^2 + (10-10.8)^2}{5}$$

$$\sigma^2 = 2.16$$

# Medidas de dispersión

- **Desviación estándar:** La desviación estándar mide el grado de dispersión de los datos con respecto a la media, se denota como s para una muestra o como σ para la población. Se define como la raíz cuadrada de la varianza según la expresión:

$$\sigma = \sqrt{\frac{\sum_{i}^{N}(X_i - \bar{X})^2}{N}}$$

# Ejemplo Varianza

- Valores:

$$\sigma = \sqrt{\frac{\sum(x_i - \bar{x}\ )^2}{N}}$$

**10 12 9 13 10**

- Se calcula la varianza:

$$\sigma^2 = \frac{(10-10.8)^{\,2} + (12-10.8)^{\,2} + (9-10.8)^{\,2} + (13-10.8)^{\,2} + (10-10.8)^{\,2}}{5}$$

$$\sigma^2 = 2.16$$

- Se calcula la desviación estándar

$$\sigma = \sqrt{2.16}$$

# Conclusiones

- La estructura de datos más empleada para el análisis de datos es DataFrame y en casos especiales Series.
- El DataFrame y/o Series sirve como fuente de datos para realizar operaciones estadísticas como medidas de tendencia central y de dispersión.

# Referencias Bibliográficas


- Business Intelligence Roadmap Addison-
Information Technology Series Moss, Larissa T. 
Shaku (2003)

- The Data Warehouse Toolkit, Wiley Computer Pub
[Kim92] Kimball, Ralph (1992)

- The Data Warehouse Lifecycle Toolkit, Wiley Co
Publishing. [Kim98] Kimball, Ralph et all (1998)

Tecsup®

TECNOLOGÍA
CON SENTIDO