

Nivel Básico – Explorador

Análisis de datos Misión 2

ESTADÍSTICA DESCRIPTIVA BÁSICA

Distribuciones de Probabilidad, Medidas de Tendencia Central y Varianza

INTRODUCCIÓN A LA ESTADÍSTICA DESCRIPTIVA

¿Qué es la estadística descriptiva?

¿Para qué las distribuciones de probabilidad, medidas de tendencia central y varianza?

DISTRIBUCIONES DE PROBABILIDAD

¿Qué es una Distribución de Probabilidad?

Una distribución de probabilidad describe cómo se distribuyen los valores de una variable aleatoria. Es una función que asigna probabilidades a cada posible valor.

Función de densidad de probabilidad

X =Cantidad de azúcar en gramos en 100 mL de cierto jugo

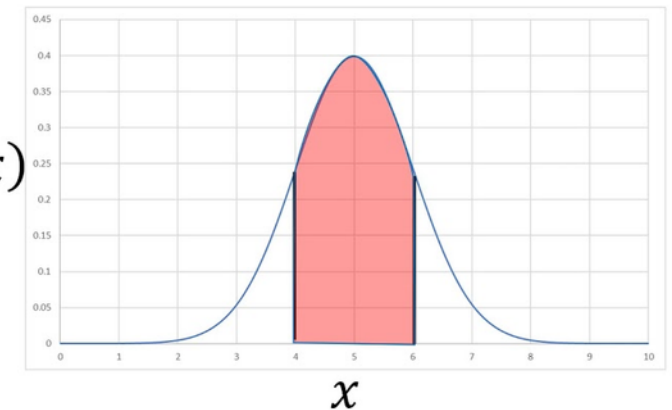
$x=0$ $x=5$ $x=2.3$

$x=4.7886$ $x=3.000001$

$$P(X = x) = 0$$

$$P(a \leq X \leq b) = p$$

$f(x)$

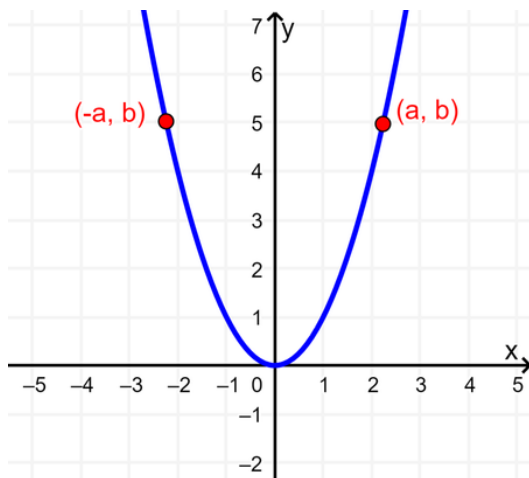


$$p = \int_a^b f(x)dx$$

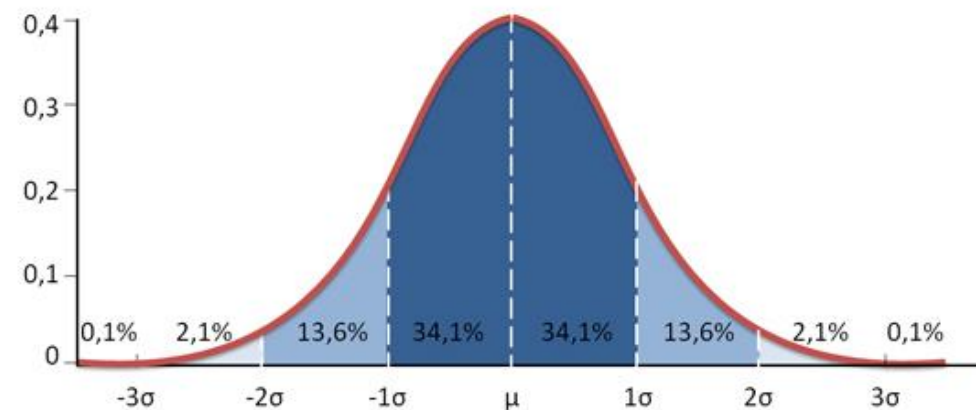
DISTRIBUCIÓN NORMAL

La distribución normal, también conocida como distribución gaussiana, es una distribución continua que se caracteriza por su forma de campana.

Simetría: La media, mediana y moda son iguales.



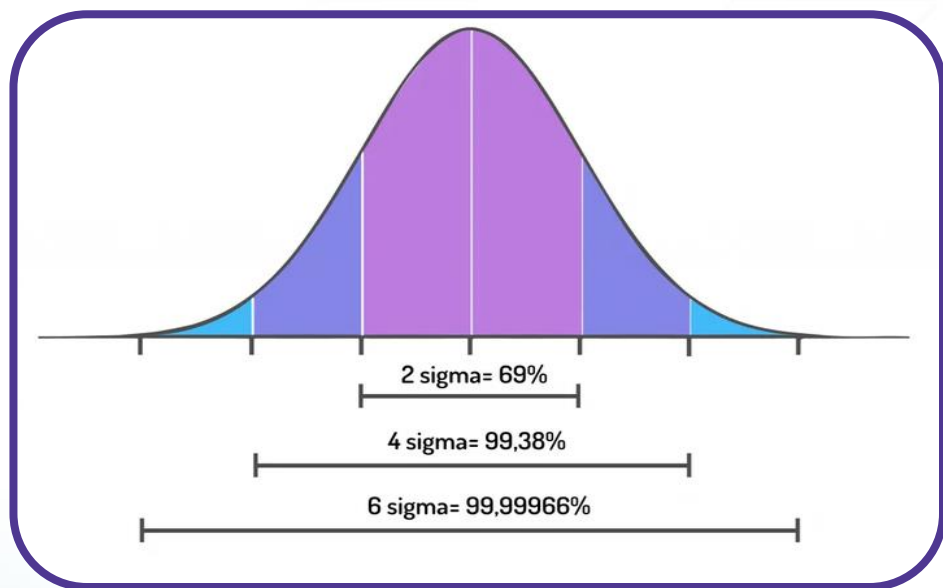
Campana de Gauss: La mayoría de los datos se concentran alrededor de la media.



DISTRIBUCIÓN NORMAL

Propiedades

- **Media (μ):** Punto medio de la distribución.
- **Desviación estándar (σ):** Medida de la dispersión.
- **68-95-99.7 Regla:** Aproximadamente el 68% de los datos están dentro de 1 desviación estándar de la media, el 95% dentro de 2, y el 99.7% dentro de 3.



MEDIDAS DE TENDENCIA CENTRAL

¿Qué son las medidas de tendencia central?

Las medidas de tendencia central describen el centro o el punto de concentración de un conjunto de datos.

Principales Medidas:

- **Media:** Promedio de los datos.
- **Mediana:** Valor central cuando los datos están ordenados.
- **Moda:** Valor que aparece con mayor frecuencia.

CÁLCULO DE LA MEDIA

Suma de todos los valores y dividiéndolo entre el número total de valores.

$$\text{Media} = \frac{\sum x_i}{n}$$

Ejemplo

Para los datos {2, 4, 6, 8}, la media es

$$\frac{2+4+6+8}{4} = 5.$$

CÁLCULO DE LA MEDIANA

Valor que se encuentra en el medio de un conjunto de datos ordenados.

Pasos:

- Ordenar los datos.
- Encontrar el valor central.
- Si hay un número par de datos, tomar el promedio de los dos valores centrales.

Ejemplo

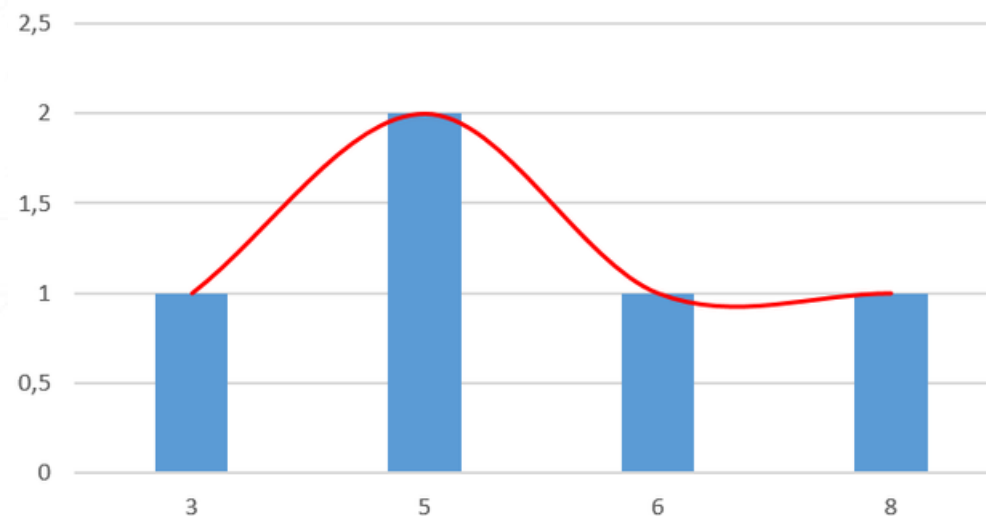
Para los datos {3, 5, 7}, la mediana es 5.

CÁLCULO DE LA MODA

Valor que aparece con mayor frecuencia en un conjunto de datos.

Ejemplo

Para los datos {3, 5, 5, 6, 8}, la moda es 2.



VARIANZA

¿Qué es la Varianza?

La varianza mide la dispersión de un conjunto de datos respecto a su media.
Es una medida de la variabilidad.

Fórmula

$$\text{Varianza} = \frac{\sum (x_i - \text{Media})^2}{n}$$

CÓMO CALCULAR LA VARIANZA

- Calcular la media.
- Restar la media de cada valor y elevar al cuadrado.
- Sumar los cuadrados y dividir por el número total de valores.

Ejemplo

Para los datos {2,4,6}, media = 4.

$$\text{Varianza} = \frac{(2-4)^2 + (4-4)^2 + (6-4)^2}{3} = \frac{8}{3} \approx 2.67$$

RELACIÓN ENTRE VARIANZA Y DESVIACIÓN ESTÁNDAR

La desviación estándar es la raíz cuadrada de la varianza. Proporciona una medida de dispersión en las mismas unidades que los datos originales.

Fórmula de la Desviación Estándar:

$$\text{Desviación Estándar} = \sqrt{\text{Varianza}}$$

Ejemplo

Para una varianza de 2,67, la desviación estándar es $\sqrt{2.67} \approx 1.63$