11MIAR–Herramientas de Estadística (Complemento Formativo-3ECTS)

Walter A. Ortiz

Máster Universitario en Inteligencia Artificial
Universidad Internacional de Valencia

July 24, 2024

Medidas de tendencia central

- Media arítmetica $\bar{X} = \mu$

$$\bar{X} = \frac{x_1 + x_2 + x_3 \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\bar{X} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 \cdots + x_n f_n}{n} = \frac{\sum_{i=1}^{n} x_i f_i}{n}$$
(Distribución de frecuencias)

Medidas de tendencia central

- Mediana $\tilde{X}$, se debe tener en cuenta el tamaño de la muestra n, y organizar los datos de menor a mayor
  Caso 1: n impar $\frac{n}{2}$ es la posición.
  Caso 2: n par $\frac{n+1}{2}$ es la posición y se calcula la semi-suma $\frac{x_{i-1}+x_i}{2}$.
  Datos agrupados: Inicialmente se debe emplear la columna de la distribución de frecuencias absolutas acumuladas. Se tiene,

$$M_e = L_{i-1} + \frac{\frac{n}{2} - F_{i-1}}{f_i} c$$

Medidas de tendencia central

- Moda $\hat{X}$, es el dato o característica con mayor frecuencia absoluta.
  Datos agrupados: Se debe emplear la columna de la distribución de frecuencias absolutas se tiene,
  $M_o = L_{i-1} + \frac{f_{i+1}}{f_{i+1} + f_{i-1}} c$

Medidas de tendencia central

Relación entre las medidas de tendencia central

- $\bar{X} = M_e = M_0$ la distribución es simétrica.
- $\bar{X} > M_e > M_0$ la distribución es asimétrica positiva.
- $\bar{X} < M_e < M_0$ la distribución es asimétrica negativa.

Medidas de dispersión

Varianza: Es la media aritmética de los cuadrados de las desviaciones.

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{N}; \quad s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1}$$

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2 f_i}{N}; \quad s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2 f_i}{n-1}$$

Desviación típica: Es la raíz cuadrada de la varianza $\sigma = \sqrt{\sigma^2}$; $s = \sqrt{s^2}$

Medidas de posición

Cuartiles: $Q_k = \frac{nk}{4}$, $k = 1, 2, 3$. En su defecto para datos agrupados,

$$Q_k = L_{i-1} + \frac{\frac{nk}{4} - F_{i-1}}{n_i} c$$

También $Q_2 = M_e$

Rango intercuartiíico: $R_C = Q_3 - Q_1$

Universidad Internacional de Valencia

Medidas de Asimetria y Apuntamiento (Curtosis)

Asimetría: $A_s = \dfrac{\bar{X} - M_o}{s}$ o $A_s = \dfrac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$, Donde

$A_S = 0$ Es simétrica

$A_s < 0$ Asimétrica negativa, $A_s > 0$ Asimétrica Positiva.

Curtosis: Es una medida que representa la altura de la curva de modo que,

$$A_p = \frac{M_4}{(s^2)^2} = \frac{\dfrac{\sum_{i=1}^{n}(x_i - \bar{X})^4}{n-1}}{s^4}$$

donde :

$A_p = 3$ entonces la distribución es Mesocurtica.

$A_p > 3$ entonces la distribución es Leptocurtica.

$A_p < 3$ entonces la distribución es Platicurtica.

Propiedades de la probabilidad:

- Sean los eventos A, B y $A \cap B = \emptyset$ entonces $P(A \cup B) = P(A) + P(B)$.
- $P(A^c) = 1 - P(A)$.
- $P(\emptyset) = 0$
- Si $A \subset B$ entonces $P(A) \leq P(B)$ y $P(B - A) = P(B) - P(A)$. En particular se tiene que$P(A) \leq 1$ para todo $A \in \mathcal{F}$.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Sean los eventos y $P(A) > 0$ entonces entonces se define la probabilidad del evento B bajo la condición A como sigue:

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

Universidad Internacional de Valencia

Teorema: (teorema de probabilidad total). Sean los eventos $A_1, A_2, \ldots, A_n$, $A_i \cap A_j = \emptyset$ para todo $i \neq j$j y $\bigcup_{n=1}^{\infty} A_i = \Omega$, tal que $P(A_i) > 0$ para todo i. Entonces para cualquier $B \in \mathcal{F}$ se satisface:

$$P(B) = \sum_{n} P(B/A_n)P(A_n)$$

Teorema de Bayes

Teorema: Sea $A_1, A_2, \ldots, A_n$ una partición finita o numerable de Ω es decir, $A_i \cap A_j = \emptyset$ para todo $i \neq j$j y $\bigcup_{n=1}^{\infty} A_i = \Omega$, tal que $P(A_i) > 0$ para todo i. Entonces para cualquier $B \in \mathcal{F}$ con $P(B) > 0$ se satisface:

$$P(A_n/B) = \frac{P(B/A_n)P(A_n)}{\sum_n P(B/A_n)P(A_n)}$$

Distribuciones de probabilidad

Discretas

- Bernoulli $P(X = X) = p^x(1-p)^{1-x}$
  $E(X) = p$, $V(x) = p(1-p)$
- Binomial $P(X = X) = \binom{n}{x} p^x(1-p)^{n-x}$
  $E(X) = np$, $V(x) = np(1-p)$
- Hipergeométrica $P(X = x) = \dfrac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$; $E(X) = \dfrac{nk}{N}$
- Poisson $P(X = x) = \dfrac{\mu^x e^{-\mu}}{x!}$; $E(X) = \mu = V(X)$

Distribuciones de probabilidad

Continuas Distribución normal $N(\mu, \sigma)$

$$Z = \frac{x - \mu}{\sigma}$$

Inferencia Estadística

Estimación puntual $\mu = E(x)$, $\sigma^2 = \frac{\sigma^2}{n}$
Intervalo de confianza $N(\mu, \sigma)$ conocido σ

$$\mu \in \bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$N(\mu, \sigma)$ desconocido σ

$$\mu \in \bar{X} \pm t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}$$