Formulario

Estadística descriptiva univariante

Notación

- $X, Y, ...$: Variables.
- x_i:
  - En datos individuales: Cada uno de los valores observados de la variable X
  - En datos agrupados: Cada uno de los k posibles valores de la variable X.
- n: Número total de observaciones en la muestra.
- N: Número total de observaciones en la población.
- n_i: Número de observaciones en la clase i.
- c_i: Marca de clase en datos agrupados por intervalos.
- $L_i, i = 0, ..., k$: Límites de los intervalos $(L_{i-1}, L_i]$.

Tablas de frecuencias

- n_i: Frecuencia absoluta, número de observaciones en la clase i.
- f_i: Frecuencia relativa. $f_i = \frac{n_i}{n}$
- N_i: Frecuencia absoluta acumulada. $N_i = \sum\limits_{j=1}^{i} n_j$
- N_i: Frecuencia relativa acumulada. $F_i = \sum\limits_{j=1}^{i} f_j = \frac{N_i}{n}$
- Número de intervalos en variables continuas:
  - Si $n \leq 100, k \approx \sqrt{n}$
  - Si $n > 100, k \approx 1 + \log_2 n$
- A: amplitud de la variable. $A = x_{max} - x_{min}$
- a_i: Amplitud de la clase i. $a_i = A/k$
- c_i: Marca de clase. $c_i = \frac{L_{i-1}+L_i}{2}$

Medidas de tendencia central

- Moda: clase más frecuente. $x_i : n_i = \max_{j=1,...k}\{n_j\}$
- Media aritmética: $\bar{x} = \frac{\sum\limits_{i=1}^{n} x_i}{n}$.
  - Propiedad: $Y = a + bX \implies \bar{y} = a + b\bar{x}$
  - En variables discretas agrupadas: $\bar{x} = \frac{1}{n}\sum\limits_{i=1}^{k} n_i x_i = \sum\limits_{i=1}^{k} f_i x_i$,
  - En variables agrupadas en intervalos: $\bar{x} = \frac{1}{n}\sum\limits_{i=1}^{k} n_i c_i = \sum\limits_{i=1}^{k} f_i c_i$
- Mediana: $\min\limits_{i=1,...n} x_i : F_i \geq 0{,}5$
- Media geométrica: $m_g = (\Pi_{i=1}^{n} x_i)^{\frac{1}{n}}$

- Media armónica: $H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$

Medidas de posición

- Percentil de orden p: $P_{p\%} = \min_{i=1,\ldots n} x_i : F_i \geq p/100$
- Cuartiles: $Q_1 = P_{25}$; $Q_3 = P_{75}$

Medidas de dispersión

- Rango o recorrido: $R = \max_i x_i - \min_i x_i$
- Desviación media absoluta: $DMA = \frac{1}{n} \sum_{i=1}^{n} |x_i - \bar{x}|$.
- Desviación absoluta mediana: $DAM = Me|x_i - Me_x|, \; i = 1, \ldots, n$.
- Varianza muestral o cuasivarianza: $s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)$
- Varianza poblacional: $\sigma^2 = \frac{\sum_{i=1}^{N}(X_i-\mu)^2}{N} = \frac{1}{N}\sum_{i=1}^{n} X_i^2 - \mu^2$
- Desviación típica muestral o cuasidesviación típica: $s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$.
- Propiedad de la varianza: $Y = a + bX \implies s_y^2 = b^2 s_X^2$
- Tipificación: $Z = \frac{X-\bar{x}}{s} \implies \bar{z} = 0; s^2 = 1$
- Coeficiente de variación: $CV = \frac{s}{|\bar{x}|}$
- Rango intercuartílico: $IQR = Q_3 - Q_1$

Medidas de forma

- Coeficiente de asimetría: $\gamma_1 = \frac{m_3}{s^3}$
  - $m_3 = \frac{1}{n}\sum_{i=1}^{n}(x-\bar{x})^3$
- Coeficiente de curtosis (apuntamiento): $\gamma_2 = \frac{m_4}{s^4} - 3$
  - $m_4 = \frac{1}{n}\sum_{i=1}^{n}(x-\bar{x})^4$

Estadística descriptiva bivariante

Notación

- $X, Y, ...$: Variables.
- x_i, y_j: Cada uno de los k posibles valores de la variable X.
- (x_i, y_i): Cada uno de los n pares de valores observados.
- n: Número total de observaciones en la muestra.
- n_i: Número de clases de la variable X.
- n_j: Número de clases de la variable Y.

- n_{ij}: Número de observaciones en la clase i de la variable X **y** en la clase j de la variable Y.

Tablas de frecuencias

- n_{ij}: Frecuencia absoluta conjunta, número de observaciones en la clase i de la variable X **y** en la clase j de la variable Y.
- f_{ij}: Frecuencia relativa conjunta. $f_{ij} = \frac{n_{ij}}{n}$
- Frecuencias marginales de X:
  - Absolutas: $n_{i\cdot} = \sum_{j=1}^{n_j} n_{ij}$
  - Relativas: $f_{i\cdot} = \sum_{j=1}^{n_j} f_{ij}$
- Frecuencias marginales de Y:
  - $n_{\cdot j} = \sum_{i=1}^{n_i} n_{ij}$
  - $f_{\cdot j} = \sum_{i=1}^{n_i} f_{ij}$
- Frecuencias condicionadas:
  - $f_{x_i|y=y_j} = \frac{n_{ij}}{n_{\text{ů}j}}$.
  - $f_{y_j|x=x_y} = \frac{n_{ij}}{n_{i\text{ů}}}$.
- Independencia: Si $f_{ij} = f_{i\cdot} \cdot f_{\cdot j}$ $\forall i, j$, entonces las variables X e Y son independientes.

Covarianza y correlación

- Covarianza poblacional:
  - Definición: $\sigma_{xy} = \frac{1}{N} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$
  - Cálculo abreviado: $\sigma_{xy} = \frac{1}{N} \sum_{i=1}^{N} (X_i \cdot Y_i) - \bar{X} \cdot \bar{Y}$
- Covarianza muestral:
  - Definición: $s_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$
  - Cálculo abreviado: $s_{xy} = \frac{1}{n-1} \left(\sum_{i=1}^{n} (x_i \cdot y_i) - n \cdot \bar{x} \cdot \bar{y} \right)$
- Coeficiente de correlación lineal: $r_{xy} = \frac{s_{xy}}{s_x \cdot s_y}$
- Matriz de covarianzas (caso bivariante):

$$\mathbf{S} = \begin{bmatrix} s_x^2 & s_{xy} \\ s_{xy} & s_y^2 \end{bmatrix}$$

Regresión lineal simple

- Recta de regresión: $y = a + bx$
  - $b = \frac{s_{xy}}{s_x^2}$
  - $a = \bar{y} - b\bar{x}$
  - $b = \frac{s_y}{s_x} r_{xy}$

- Predicción de nuevos valores: $\hat{y}_{n+1} = a + bx_{n+1}$
- Residuos: $\varepsilon_i = y_i - \hat{y}_i = y_i - (a + bx_i)$
- Varianza residual: $s_\varepsilon^2 = \frac{1}{n}\sum\limits_{i=1}^{n} \varepsilon_i^2$
    - $\frac{s_\varepsilon^2}{s_y^2} = (1 - r_{xy}^2)$
- Coeficiente de determinación: $R^2 = 1 - \frac{s_\varepsilon^2}{s_y^2} = r_{xy}^2$

Probabilidad

Notación

- $A, B, \ldots$: Sucesos
- ω: Suceso elemental
- Ω: Espacio muestral
- $\emptyset$: Suceso imposible
- A^c: Suceso complementario del suceso A

Definiciones

- Unión de sucesos: $A \cup B$: Ocurre A **o** Ocurre B, o los dos
- Intersección de sucesos: $A \cap B$: Ocurre A **y** Ocurre B
- Sucesos disjuntos o mutuamente excluyentes, o incompatibles: $A \cap B = \emptyset$
- Partición del espacio muestral: Colección de sucesos $A_1, A_2, \ldots \in \Omega$ que cumplen:
    - $A_1, A_2, \ldots : \quad A_i \subset \Omega \;\forall i$
    - $A_i \cap A_j = \emptyset \;\forall i \neq j$,
    - $\bigcup\limits_i A_i = \Omega$.
- Sigma álgebra de sucesos $\aleph$ (*aleph*). conjunto de sucesos que:
    - Pertenecen a $\aleph$,
    - Si $A \in \aleph \implies A^c \in \aleph$
    - Si $\{A_i\} \in \aleph \;\; \forall i$, entonces $\bigcup\limits_i A_i \in \aleph$ y $\bigcap\limits_i A_i \in \aleph$

Propiedades

- **Conmutativa**:
    - $A \cup B = B \cup A$.
    - $A \cap B = B \cap A$.
- **Asociativa**:
    - $A \cup (B \cup C) = (A \cup B) \cup C$.
    - $A \cap (B \cap C) = (A \cap B) \cap C$.
- **Distributiva**:
    - $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
    - $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
- **Leyes de De Morgan**:
    - $(A \cup B)^c = A^c \cap B^c$.
    - $(A \cap B)^c = A^c \cup B^c$.
- $A \cup A = A \cap A = A \cup \emptyset = A \cap \Omega = A$.
- $A \cup \Omega = \Omega$.

Definiciones de probabilidad

- Definición de Laplace: $P(A) = \frac{\text{casos favorables a } A}{\text{casos posibles}}$

- Definición frecuentista: $P(A) = \lim_{n\to\infty} \frac{n(A)}{n}$
- Definición axiomática:
  - **Primer axioma**: $\forall A \in \aleph \ \exists \ P(A) \geq 0$.
  - **Segundo axioma**: $P(\Omega) = 1$.
  - **Tercer axioma**: Dada la sucesión $A_1, \ldots, A_i, \ldots : A_i \in \aleph \ \forall \, i, A_i \cap A_j = \emptyset \ \forall i \neq j$, se cumple:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Teoremas derivados

- Dados n sucesos disjuntos dos a dos $A_1, \ldots, A_n : A_i \cap A_j = \emptyset \ \forall i \neq j$:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i).$$

* $P(A^c) = 1 - P(A)$.

- $P(\emptyset) = 0$.
- Dados $A_1, A_2 : A_1 \subset A_2 \implies P(A_1) \leq P(A_2)$.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) -$
  $- \ldots + (-1)^{n-1} P\left(\bigcap_{i=1}^{n} A_i\right)$.

$$\boxed{0 \leq P(A) \leq 1}.$$

Probabilidad condicionada e independencia

- Probabilidad de A condicionada a B: $P(A|B) = \frac{P(A\cap B)}{P(B)}$
- Probabilidad de la interesección: $\boxed{P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)}$
- Regla de la cadena:

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \ldots \cdot P\left(A_n | \bigcap_{i=1}^{n-1} S_i\right)$$

- A y B independientes $\iff P(A|B) = P(A)$ y $P(B|A) = P(B)$
  - $\boxed{P(A \cap B) = P(A) \cdot P(B)}$ (solo si son independientes)
  - $P(A^c|B) = 1 - P(A|B)$

Probabilidad total y fórmula de Bayes

- Probabilidad total:

$$\boxed{P(B) = \sum_{i=1}^{n} P(B/A_i) \cdot P(A_i)}$$

- Fórmula de Bayes:

$$\boxed{P(A_i|B) = \frac{P(B|A_i) \cdot P(A_i)}{\sum_{i=1}^{n} P(B/A_i) \cdot P(A_i)}}$$

Variable aleatoria

- Función de distribución: $F(x) = P[X \leq x]$
- Probabilidad en un intervalo: $P[a < X \leq b] = F(b) - F(a)$
- Probabilidad del intervalo complementario: $P[X > a] = 1 - F(a)$
- Función de masa de probabilidad (VA discreta):
  - $p(x_i) = P[X = x_i] = P[x_{i-1} < X \leq x_i] = F(x_i) - F(x_{i-1})$
  - Condiciones:
    * $p(x_i) \geq 0 \; \forall i.$
    * $\sum_{i=1}^{\infty} p(x_i) = 1.$
  - Función de distribución: $F(x_i) = \sum_{j=1}^{i} p(x_j)$
- Función de densidad (VA continua):
  - $f(x) = \frac{dF(x)}{dx}$
  - $F(x) = \int_{-\infty}^{x} f(t)dt = P[X \leq x]$
  - Condiciones:
    * $f(x) \geq 0$
    * $\int_{-\infty}^{\infty} f(x)dx = 1$
  - Probabilidad en un intervalo: $P[a < X \leq b] = \int_a^b f(x)dx$
  - Consecuencia: $P[X = x] = 0$
- Características:
  - Media: $\mu = E[X]$
    * VA discreta: $\mu = E[X] = \sum_i x_i p(x_i)$
    * VA continua: $\mu = E[X] = \int_{-\infty}^{\infty} x f(x)dx$
    * Propiedad: $E[a + bX] = a + bE[X]$
  - Varianza: $V[X] = \sigma^2 = E[(X - \mu)^2] = E[X^2] - (E[X])^2$
    * VA discreta: $\alpha_2 = E[X^2] = \sum_i x_i^2 p(x_i)$
    * VA continua: $\alpha_2 = E[X^2] = \int_{-\infty}^{\infty} x^2 f(x)dx$
    * Propiedad: $V[a + bX] = b^2 V[X]$
  - Desviación típica: $\sigma = +\sqrt{\sigma^2}$
  - Coeficiente de variación: $CV = \frac{\sigma}{\mu}$
- Tipificación de variables aleatorias:

$$Z = \frac{X - \mu}{\sigma} \implies \mu_Z = 0; \; \sigma_Z = 1$$

Modelos de distribución de probabilidad

Distribuciones discretas más importantes

- Bernoulli: $Ber(p)$
  - $X = \begin{cases} 1 & \text{con probabilidad } p \\ 0 & \text{con probabilidad } 1-p \end{cases}$
  - Media: $\mu = E[X] = p$.
  - Varianza: $\sigma^2 = V[X] = p \cdot (1-p)$.
- Binomial: $X \sim Bin(n;\ p);\ n > 0,\ 0 < p < 1$
  - $P[X = x] = \binom{n}{x} \cdot p^x \cdot (1-p)^{(n-x)}$
  - Media: $\mu = E[X] = n \cdot p$.
  - Varianza: $\sigma^2 = V[X] = n \cdot p \cdot (1-p)$.
  - Aditiva: $Y = \sum\limits_{j=1}^{m} X_j,\ X_j \sim Bin(n_j;\ p) \implies Y \sim Bin\left(\sum\limits_{j=1}^{m} n_j;\ p\right)$
- Geométrica
  - Media:
  - Varianza:
- Binomial negativa
  - Media:
  - Varianza:
- Poisson
  - Media:
  - Varianza:
  - Aditiva:
- Hipergeométrica
  - Media:
  - Varianza:
  - Aditiva:

Distribuciones discretas más importantes

- Uniforme
  - Media:
  - Varianza:
- Exponencial
  - Media:
  - Varianza:
- Normal
  - Media:
  - Varianza: