Fórmulas de Estadística

Estadística Descriptiva

Tamaño muestral n número de individuos de la muestra.

Frecuencias

Frecuencia absoluta n_i.
Frecuencia relativa $f_i = n_i/n$.
Frec. absoluta acumulada $N_i = \sum_{k=0}^{i} n_i$.
Frec. relativa acumulada $F_i = N_i/n$.

Medidas de Representatividad

Media $\overline{x} = \frac{\sum x_i n_i}{n}$.
Mediana me Es el valor que ocupa el centro de la distribución. Tiene frecuencia acumulada $N_{me} = n/2$ y frecuencia relativa acumulada $F_{me} = 0{,}5$.
Moda mo Es el valor con mayor frecuencia absoluta.

Medidas de Posición

Cuartiles c_1, c_2, c_3 dividen la muestra en 4 partes iguales. Tienen frecuencias relativas acumuladas $F_{c_1} = 0{,}25$, $F_{c_2} = 0{,}5$ y $F_{c_3} = 0{,}75$ respectivamente.
Percenciltes $p_1, p_2, \cdots, p_{99}$ dividen la distribución en 100 partes iguales. La frecuencia relativa acumulada correspondiente a el percentil i es $F_{p_i} = i/100$.

Medidas de Dispersión

Rango intercuartílico $RI = c_3 - c_1$.
Varianza $s^2 = \frac{\sum x_i^2 n_i}{n} - \overline{x}^2$
Desviación típica $s = +\sqrt{s^2}$.
Coeficiente de variación $cv = \frac{s}{|\overline{x}|}$.

Medidas de Forma

Coeficiente de asimetría

$$g_1 = \frac{\sum (x_i - \overline{x})^3 f_i}{s^3}.$$

- $g_1 = 0$ distribución simétrica.
- $g_1 < 0$ distribución asimétrica a la izquierda.
- $g_1 > 0$ distribución asimétrica a la derecha.

Coeficiente de apuntamiento

$$g_2 = \frac{\sum (x_i - \overline{x})^4 f_i}{s^4} - 3.$$

- $g_2 = 0$ apuntamiento normal.
- $g_2 < 0$ apuntamiento menor de lo normal.
- $g_2 > 0$ apuntamiento mayor de lo normal.

Transformaciones Lineales

Propiedades de la transformación $y = a + bx$

$$\overline{y} = a + b\overline{x} \qquad s_y = bs_x$$

Transformación de tipificación $y = \frac{x - \overline{x}}{s_x}$.

Regresión y Correlación

Regresión

Covarianza $s_{xy} = \frac{\sum x_i y_j n_{ij}}{n} - \overline{x}\overline{y}$.
Rectas de regresión

$$(y \text{ sobre } x)\ y = \overline{y} + \frac{s_{xy}}{s_x^2}(x - \overline{x}),$$

$$(x \text{ sobre } y)\ x = \overline{x} + \frac{s_{xy}}{s_y^2}(y - \overline{y}).$$

Coeficientes de regresión

$$(y \text{ sobre } x)\ b_{yx} = \frac{s_{xy}}{s_x^2} \qquad (x \text{ sobre } y)\ b_{xy} = \frac{s_{xy}}{s_y^2}$$

Correlación

Coeficiente de determinación lineal

$$r^2 = \frac{s_{xy}^2}{s_x^2 s_y^2} \qquad 0 \leq r^2 \leq 1$$

- $r^2 = 0$ no hay relación lineal.
- $r^2 = 1$ relación lineal perfecta.

Coeficiente de correlación lineal

$$r = \frac{s_{xy}}{s_x s_y}. \qquad -1 \leq r \leq 1$$

- $r = 0$ no hay relación lineal.
- $r = -1$ relación lineal perfecta decreciente.
- $r = 1$ relación lineal perfecta creciente.

Regresión no lineal

Modelo exponencial $y = e^{a+bx}$
Aplicar el logaritmo a la variable dependiente y calcular la recta $\log y = a + bx$.
Modelo logarítmico $y = a + b \log x$
Aplicar el logaritmo a la variable independiente y calcular la recta $y = a + b \log x$.
Modelo potencial $y = ax^b$
Aplicar el logaritmo a la variable dependiente e independiente y calcular la recta $\log y = a + b \log x$.

Relaciones entre atributos

Coeficiente de correlación de spearman Es el coeficiente de correlación lineal aplicado a los órdenes de los valores de la variable.
Coeficiente chi-cuadrado

$$\chi^2 = \sum \frac{\left(n_{ij} - \frac{n_{x_i} n_{y_j}}{n}\right)^2}{\frac{n_{x_i} n_{y_j}}{n}},$$

Coeficiente de contingencia

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} \qquad 0 \leq C < 1$$

Probabilidad

Espacio muestral E Es el conjunto de posibles resultados de un experimento.

Operaciones entre Sucesos

Unión $A \cup B$ Elementos de A más elementos de B.
Intersección $A \cap B$ Elementos comunes en A y B.
Contrario $\overline{A}$ Elementos de E menos los de A.
Diferencia $A - B$ Elementos de A menos los de B.
$A - B = A \cap \overline{B} = A - (A \cap B)$.
Sucesos incompatibles $A \cap B = \emptyset$.

Propiedades de las operaciones entre sucesos

Conmutativa $A \cup B = B \cup A$ y $A \cap B = B \cap A$.
Asociativa $A \cup (B \cup C) = (A \cup B) \cup C$ y $A \cup (B \cap C) = (A \cap B) \cap C$.
Elemento Neutro $A \cup \emptyset = A$ y $A \cap E = A$.
Elemento Antisimétrico $A \cup \overline{A} = E$ y $A \cap \overline{A} = \emptyset$.
Elemento Absorbente $A \cup E = E$ y $A \cap \emptyset = \emptyset$.
Leyes de Morgan $\overline{A \cup B} = \overline{A} \cap \overline{B}$ y $\overline{A \cap B} = \overline{A} \cup \overline{B}$

Probabilidad

Regla de Laplace

$$P(A) = \frac{n_A}{n_E} \quad \left(\frac{\text{casos favorables}}{\text{casos posibles}}\right).$$

Sólo se aplica cuando hay equiprobabilidad.
Probabilidad del contrario

$$P(\overline{A}) = 1 - P(A).$$

Probabilidad de la unión

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Probabilidad condicionada

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

Probabilidad de la intersección

$$P(A \cap B) = P(A)P(B/A).$$

Sucesos independientes $P(A/B) = P(A)$.
Sistema completo de sucesos $A_1, \ldots, A_n$ deben cumplir
$A_1 \cup \cdots \cup A_n = E$.
$A_i \cap A_j = \emptyset$ si $i \neq j$.
Teorema de la probabilidad total

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

Teorema de Bayes

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

Variables Aleatorias

Tamaño poblacional N número de individuos de la población.

Variables Aleatorias Discretas

Función de probabilidad $f(x_i) = P(X = x_i)$.
Función de distribución $F(x_i) = P(X \leq x_i)$.
Media o esperanza $E[X] = \mu = \sum x_i f(x_i)$.
Varianza $V[X] = \sigma^2 = \sum x_i^2 f(x_i) - \mu^2$.
Desviación típica $\sigma = +\sqrt{\sigma^2}$.

Modelos de Distribución Discretos

Uniforme $U(k)$

$$f(x) = 1/k.$$

Binomial $B(n, p)$

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

$$\mu = n \cdot p \quad \sigma = \sqrt{n \cdot p \cdot (1-p)}.$$

Poisson $P(\lambda)$

$$f(x) = e^{-\lambda} \frac{\lambda^x}{x!}$$

$$\mu = \lambda \quad \sigma = \sqrt{\lambda}.$$

Variables Aleatorias Continuas

Función de densidad $f(x)$ Debe cumplir

- $f(x) \geq 0$.
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

Función de distribución

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)\,dx.$$

Probabilidad de un intervalo

$$P(a \leq X \leq b) = \int_{a}^{b} f(x)\,dx = F(b) - F(a).$$

Media $E[X] = \mu = \int_{-\infty}^{\infty} x f(x)\,dx$.
Varianza $V[X] = \sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2$.

Modelos de Distribución Continuos

Uniforme $U(a, b)$

$$f(x) = \frac{1}{b-a}.$$

Normal $N(\mu, \sigma)$.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$