



Estadística descriptiva unidimensional

Tabla de frecuencias									
I_i	L_{i-1}	L_i	x_i	n_i	N_i	f_i	F_i	α_i	d_i
				$\sum n_i = N$		$\sum f_i = 1$			

- I_i = intervalos
- L_{i-1} = límite inferior del intervalo
- L_i = límite superior del intervalo
- x_i = dato o marca de clase en datos agrupados en intervalos $\left(\frac{L_{i-1}+L_i}{2}\right)$
- n_i = frecuencia absoluta. N es la población
- N_i = frecuencia absoluta acumulada
- f_i = frecuencia relativa. $f_i = \frac{n_i}{N}$
- F_i = frecuencia relativa acumulada
- c_i = amplitud del intervalo. $c_i = L_i - L_{i-1}$
- d_i = densidad. Solo cuando las amplitudes c_i no son constantes. $d_i = \frac{n_i}{c_i}$

Media aritmética:

$$\bar{x} = \frac{x_1 n_1 + x_2 n_2 + \dots + x_n n_n}{N} = \frac{\sum_{i=1}^n x_i \cdot n_i}{N}$$

Mediana: es el valor central tras ordenar de menor a mayor.

$$\text{Discretos} \begin{cases} \text{IMPARES: } \frac{N-1}{2} + 1 \\ \text{PARES: dos valores} \rightarrow \frac{N}{2} ; \frac{N}{2} + 1 \end{cases}$$

$$\text{Intervalos} \left\{ \frac{N}{2} \rightarrow \text{Buscamos superior en } N_i \rightarrow Me = L_{i-1} + \frac{\frac{N}{2} - N_{i-1}}{n_i} \cdot \alpha_i \right.$$

Moda: valor x_i más repetido, con mayor n_i o d_i (o valores, pueden ser varios).

$$\text{Intervalos de } c_i = \text{cte: } Mo = L_{i-1} + \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} \cdot c_i$$

$$\text{Intervalos de } c_i \neq \text{cte: } Mo = L_{i-1} + \frac{d_i - d_{i-1}}{(d_i - d_{i-1}) + (d_i - d_{i+1})} \cdot c_i$$



Cuartiles: valores que separan la población al 25% cada trozo.

$$\text{Intervalos: } \begin{cases} \frac{N \cdot 1}{4} \rightarrow \text{Buscamos superior en } N_i \rightarrow Q_1 = L_{i-1} + \frac{\frac{N \cdot 1}{4} - N_{i-1}}{n_i} \cdot c_i \\ \frac{N \cdot 2}{4} \rightarrow \text{Buscamos superior en } N_i \rightarrow Q_2 = L_{i-1} + \frac{\frac{N \cdot 2}{4} - N_{i-1}}{n_i} \cdot c_i \\ \frac{N \cdot 3}{4} \rightarrow \text{Buscamos superior en } N_i \rightarrow Q_3 = L_{i-1} + \frac{\frac{N \cdot 3}{4} - N_{i-1}}{n_i} \cdot c_i \end{cases}$$

Momentos:

$$\text{Respecto a origen: } a_r = \sum_{i=1}^n x_i^r \cdot \frac{n_i}{N} \quad ; \quad \text{Respecto a media: } m_r = \sum_{i=1}^n (x_i - \bar{x})^r \cdot \frac{n_i}{N}$$

Varianza:

$$\sigma^2 = \sum_{i=1}^n \frac{x_i^2 \cdot n_i}{N} - \bar{x}^2 \quad ; \quad S^2 = m_2 = a_2 - a_1^2$$

Cuasi-Varianza ó varianza muestral:

$$S^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2 \cdot n_i}{N - 1}$$

Desviación típica: raíz cuadrada de varianza.

$$\sigma = +\sqrt{\sigma^2} = +\sqrt{\sum_{i=1}^n \frac{x_i^2 \cdot n_i}{N} - \bar{x}^2}$$

Coefficiente de variación de Pearson: a mayor V, mayor dispersión, menor representatividad de la media.

$$V = \frac{\sigma}{\bar{x}}$$

Coefficiente de asimetría de Fisher:

$$\gamma_1 = \frac{m_3}{S^3} = \frac{\sum_{i=1}^n (x_i - \bar{x})^3 \cdot \frac{n_i}{N}}{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \frac{n_i}{N} \right]^{\frac{3}{2}}} \rightarrow \begin{cases} \gamma_1 > 0: \text{Asimetría positiva: } Mo < \bar{x} \\ \gamma_1 = 0: \text{Simétrica: } Mo = \bar{x} \\ \gamma_1 < 0: \text{Asimetría negativa: } Mo > \bar{x} \end{cases}$$

Coefficiente de apuntamiento: curtosis.

$$\gamma_2 = \frac{m_4}{S^4} - 3 = \frac{\sum_{i=1}^n (x_i - \bar{x})^4 \cdot \frac{n_i}{N}}{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \frac{n_i}{N} \right]^2} - 3 \rightarrow \begin{cases} \gamma_2 > 0: \text{Leptocúrtica (puntiaguda)} \\ \gamma_2 = 0: \text{Mesocúrtica (normal)} \\ \gamma_2 < 0: \text{Platicúrtica (plana)} \end{cases}$$



Estadística descriptiva bidimensional

Tabla de frecuencias bidimensional							
X/Y	y_1	y_2	...	y_j	...	y_k	Marginal X
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1k}	$n_{1\bullet}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2k}	$n_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{ik}	$n_{i\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_h	n_{h1}	n_{h2}	...	n_{hj}	...	n_{hk}	$n_{h\bullet}$
Marginal Y	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet j}$	...	$n_{\bullet k}$	N

Momento respecto a origen:

$$a_{rs} = \sum_{i=1}^h \sum_{j=1}^k x_i^r \cdot y_j^s \cdot \frac{n_{ij}}{N} \quad ; \quad \begin{cases} a_{10} = \bar{x} \\ a_{01} = \bar{y} \end{cases}$$

Momento respecto a la media:

$$m_{rs} = \sum_{i=1}^h \sum_{j=1}^k (x_i - \bar{x})^r \cdot (y_j - \bar{y})^s \cdot \frac{n_{ij}}{N} \quad ; \quad \begin{cases} m_{20} = a_{20} - a_{10}^2 = S_x^2 \\ m_{02} = a_{02} - a_{01}^2 = S_y^2 \end{cases}$$

Covarianza:

$$\sigma_{xy} = \sum_{i=1}^h \sum_{j=1}^k \frac{x_i \cdot y_j \cdot n_{ij}}{N} - \bar{x} \cdot \bar{y} = m_{11} = a_{11} - a_{10} \cdot a_{01}$$

Regresión lineal:

$$y \text{ sobre } x \rightarrow \hat{y} = \hat{a} + \hat{b} \cdot x \rightarrow \hat{b} = \frac{\sigma_{xy}}{\sigma_x^2} \quad ; \quad \hat{a} = \bar{y} - \frac{\sigma_{xy}}{\sigma_x^2} \cdot \bar{x}$$

$$x \text{ sobre } y \rightarrow \hat{x} = \hat{a} + \hat{b} \cdot y \rightarrow \hat{b} = \frac{\sigma_{xy}}{\sigma_y^2} \quad ; \quad \hat{a} = \bar{x} - \frac{\sigma_{xy}}{\sigma_y^2} \cdot \bar{y}$$

Coefficiente de determinación lineal (bondad del ajuste):

$$R^2 = \frac{\sigma_{xy}^2}{\sigma_x^2 \cdot \sigma_y^2} \quad 0 \leq R^2 \leq 1$$

Coefficiente de correlación lineal:

$$r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} \quad -1 \leq r \leq 1$$



Probabilidad de sucesos:

Ley de Laplace

$$\text{Probabilidad de } A = \frac{\text{casos favorables de } A}{\text{casos posibles}}$$

Unión de sucesos $A \cup B$:

$$\text{Sucesos compatibles: } P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$\text{Sucesos incompatibles: } P(A \cup B) = P(A) + P(B)$$

Intersección de sucesos $A \cap B$:

$$\text{Sucesos independientes: } P(A \cap B) = P(A) \cdot P(B)$$

$$\text{Sucesos dependientes: } \begin{cases} \text{Probabilidad condicionada: } P(A|B) = \frac{P(A \cap B)}{P(B)} \\ \text{despejando ... } P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A) \end{cases}$$

Sucesos contrarios $\bar{A}$:

$$P(\bar{A}) = 1 - P(A)$$

Leyes de Morgan:

$$P(\bar{A} \cup \bar{B}) = P(\overline{A \cap B})$$

$$P(\overline{A \cup B}) = P(\bar{A} \cap \bar{B})$$

Diferencia de sucesos $A - B$:

$$P(A - B) = P(A \cap \bar{B}) = P(A) - P(A \cap B)$$

$$P(B - A) = P(B \cap \bar{A}) = P(B) - P(A \cap B)$$

Teorema de probabilidad total:

$$P(B) = P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2) + \dots + P(A_n) \cdot P(B|A_n)$$

Teorema Bayes:

$$P(A_i|B) = \frac{P(A_i) \cdot P(B|A_i)}{P(A_1) \cdot P(B|A_1) + \dots + P(A_n) \cdot P(B|A_n)} = \frac{P(A_i) \cdot P(B|A_i)}{P(B)}$$



Distribuciones de probabilidad:

Función de probabilidad en variable discreta (ejem: un dado):

$$\text{Función de masa o cuantía: } f(x) = \begin{cases} 1/6 & \text{si } x = 1 \\ 1/6 & \text{si } x = 2 \\ 1/6 & \text{si } x = 3 \\ 1/6 & \text{si } x = 4 \\ 1/6 & \text{si } x = 5 \\ 1/6 & \text{si } x = 6 \end{cases}$$

Función de distribución en variable discreta (ejem: un dado):

$$\text{Función de distribución: } F(x) = \begin{cases} 0 & \text{si } x < 1 \\ 1/6 & \text{si } 1 \leq x < 2 \\ 2/6 & \text{si } 2 \leq x < 3 \\ 3/6 & \text{si } 3 \leq x < 4 \\ 4/6 & \text{si } 4 \leq x < 5 \\ 5/6 & \text{si } 5 \leq x < 6 \\ 1 & \text{si } x \geq 6 \end{cases}$$

Función de probabilidad en variable continua (ejem):

$$\text{Función de densidad: } f(x) = \begin{cases} \frac{x+1}{8} & \text{si } 2 < x < 4 \\ 0 & \text{en el resto} \end{cases}$$

$$\text{Propiedades: } \begin{cases} \int_{-\infty}^{+\infty} f(x)dx = 1 \\ \text{Esperanza: } E(X) = \int_{-\infty}^{+\infty} x \cdot f(x)dx \\ \text{Varianza: } Var(X) = \int_{-\infty}^{+\infty} x^2 \cdot f(x)dx - [E(X)]^2 \end{cases}$$

Función de distribución en variable continua (mismo ejem anterior):

$$\text{Función de distribución: } F(x) = \begin{cases} 0 & \text{si } x \leq 2 \\ \int_2^x \frac{t+1}{8} dt & 2 < x < 4 \\ 1 & \text{si } x \geq 4 \end{cases} = \begin{cases} 0 & \text{si } x \leq 2 \\ \frac{x^2}{16} + \frac{x}{8} - \frac{1}{2} & 2 < x < 4 \\ 1 & \text{si } x \geq 4 \end{cases}$$

Probabilidades en variables continua sobre funciones de probabilidad (ejem anterior):

$$P(X < 3,2) = \int_{-\infty}^{3,2} f(x)dx = \int_2^{3,2} \frac{x+1}{8} dx = \frac{1}{8} \cdot \left(\frac{x^2}{2} + x \right)_2^{3,2} = 0,54$$

$$\text{Otra forma: } P(X < 3,2) = F(3,2) = \frac{3,2^2}{16} + \frac{3,2}{8} - \frac{1}{2} = 0,54$$

$$P(2,9 < X < 3,2) = \int_{2,9}^{3,2} f(x)dx = \int_{2,9}^{3,2} \frac{x+1}{8} dx = \frac{1}{8} \cdot \left(\frac{x^2}{2} + x \right)_{2,9}^{3,2} = 0,1519$$



Otra forma: $P(2,9 < X < 3,2) = F(3,2) - F(2,9) = 0,54 - \left(\frac{2,9^2}{16} + \frac{2,9}{8} - \frac{1}{2}\right) = 0,1519$

Función de probabilidad en variable continua bidimensional (ejem):

Función de masa: $f(x, y) = \begin{cases} 1/2 & \text{si } x = 0, y = 1 \\ 1/2 & \text{si } x = 1, y = 0 \\ 0 & \text{en el resto} \end{cases}$

Función de densidad: $f(x, y) = \begin{cases} 4xy & \text{si } 0 < x < 1, 0 < y < 1 \\ 0 & \text{en el resto} \end{cases}$

Función marginal X: $f(x) = \begin{cases} \int_0^1 4xy \, dy & \text{si } 0 < x < 1 \\ 0 & \text{en el resto} \end{cases} = \begin{cases} 2x & \text{si } 0 < x < 1 \\ 0 & \text{en el resto} \end{cases}$

Función marginal Y: $f(y) = \begin{cases} \int_0^1 4xy \, dx & \text{si } 0 < y < 1 \\ 0 & \text{en el resto} \end{cases} = \begin{cases} 2y & \text{si } 0 < y < 1 \\ 0 & \text{en el resto} \end{cases}$

Función de distribución: $F(x, y) = \begin{cases} 0 & \text{si } x \leq 0 \cap y \leq 0 \\ \int_0^x \int_0^y 4ts \, ds \, dt & \text{si } 0 < x < 1, 0 < y < 1 \\ 1 & \text{si } x \geq 1 \cap y \geq 1 \end{cases}$

El cálculo de probabilidades se hace igual, pero con integrales dobles, o con la función de distribución.

Propiedades: $\left\{ \begin{array}{l} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \, dx \, dy = 1 \\ \text{Esperanza: } E(X) = \int_{-\infty}^{+\infty} x \cdot f(x) \, dx \quad E(Y) = \int_{-\infty}^{+\infty} y \cdot f(y) \, dy \\ \text{Varianza: } \begin{cases} \text{Var}(X) = \int_{-\infty}^{+\infty} x^2 \cdot f(x) \, dx - [E(X)]^2 \\ \text{Var}(Y) = \int_{-\infty}^{+\infty} y^2 \cdot f(y) \, dy - [E(Y)]^2 \end{cases} \\ \text{Covarianza: } \text{Cov}(XY) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x \cdot y \cdot f(x, y) \, dx \, dy - E(X) \cdot E(Y) \end{array} \right.$



Distribuciones de probabilidad COMUNES:

Distribuciones de variable DISCRETA		
Binomial $B(n, p)$	$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$	$E(X) = n \cdot p$ $Var(X) = n \cdot p \cdot (1 - p)$
Geométrica $G(p)$	$P(X = k) = p (1 - p)^{k-1}$ Probabilidad de en las primeras "k" pruebas esté el primer error. "p" es la probabilidad de error.	$E(X) = 1/p$ $Var(X) = \frac{(1 - p)}{p^2}$
Binomial negativa $BN(r, p)$	$P(X = k) = \binom{k+r-1}{k} p^r (1 - p)^k$ Probabilidad de que pasen "k" aciertos antes de "r" errores. "p" es la probabilidad de error.	$E(X) = \frac{r(1 - p)}{p}$ $Var(X) = \frac{r(1 - p)}{p^2}$
Hipergeométrica $H(N, M, n)$	$P(X = k) = \frac{\binom{N-M}{n-k} \cdot \binom{M}{k}}{\binom{N}{n}}$ Ejem: en un pedido de "N" motores donde se sabe que hay "M" motores defectuosos ¿Cuál es la probabilidad de obtener "k" valores defectuosos si se hace una prueba de "n" motores?	$E(X) = n \cdot \frac{M}{N}$ $Var(X) = \frac{M(N-M)n(N-n)}{N^2(N-1)}$
Poisson $Po(\lambda)$	$P(X = k) = \frac{e^{-\lambda} \cdot \lambda^k}{k!}$	$E(X) = \lambda$ $Var(X) = \lambda$
Distribuciones de variable CONTINUA		
Uniforme $U(a, b)$	$f(x) = \frac{1}{b-a} \quad a < x < b$ $F(x) = \frac{x-a}{b-a}$	$E(X) = \frac{a+b}{2}$ $Var(X) = \frac{(b-a)^2}{12}$
Exponencial $Exp(\beta)$	$f(x) = \beta e^{-\beta x} \quad 0 < x$ $F(x) = 1 - e^{-\beta x}$	$E(X) = \frac{1}{\beta}$ $Var(X) = \frac{1}{\beta^2}$
Normal $N(\mu, \sigma)$	Tipificar $N(\mu, \sigma) \rightarrow z = \frac{k - \mu}{\sigma} \rightarrow N(0,1)$ $P(X > k) = P(Z > z) \rightarrow \text{Tabla}$	$P(Z > +z) \rightarrow \text{Tabla}$ $P(Z < +z) = 1 - P(Z > +z)$ $P(Z > -z) = 1 - P(Z > +z)$ $P(Z < -z) = P(Z > +z)$
Una muestra de variable discreta muy grande podemos considerarla continua		
<i>Binomial:</i> Si $n \cdot p \geq 10$: $B(n, p) \rightarrow \left\{ \begin{matrix} \mu = n \cdot p \\ \sigma = \sqrt{n \cdot p \cdot (1 - p)} \end{matrix} \right\} \rightarrow N(\mu, \sigma)$ <i>Poisson:</i> Si $n \geq 100$: $Po(\lambda) \rightarrow \left\{ \begin{matrix} \mu = \lambda \\ \sigma = \sqrt{\lambda} \end{matrix} \right\} \rightarrow N(\lambda, \sqrt{\lambda})$		



Intervalos de confianza:

Contraste de hipótesis:

ANOVA: