

INGENIERIA TECNICA INDUSTRIAL-MECANICA

Formulario de Estadística. Curso 2009-2010

ESTADÍSTICA DESCRIPTIVA

Datos sin agrupar
media

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Datos agrupados

$$\bar{x} = \frac{\sum_{i=1}^m n_i c_i}{n}$$

m : n° clases
 n_i : n° datos por clase
 c_i : marca de clase

desviación típica de la población

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^m (c_i - \mu)^2 n_i}{N}}$$

desviación típica muestral

$$s_{n-1} = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{n-1}}, s_n = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{n}}$$

$$s_{n-1} = \sqrt{\frac{\sum_{i=1}^m (c_i - \bar{x})^2 n_i}{n-1}}, s_{n-1} = \sqrt{\frac{\sum_{i=1}^m (c_i - \bar{x})^2 n_i}{n-1}}$$

coeficiente de asimetría

$$C.A. = \frac{\sum_{i=1}^N (x_i - \mu)^3}{N\sigma^3}, C.A. = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{(n-1)s_{n-1}^3}$$

$$C.A. = \frac{\sum_{i=1}^m (c_i - \bar{x})^3 n_i}{(n-1)s_{n-1}^3}$$

coeficiente de curtosis

$$C.Ap. = \frac{\sum_{i=1}^N (x_i - \mu)^4}{N\sigma^4} - 3, C.Ap. = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{(n-1)s_{n-1}^4} - 3$$

$$C.Ap. = \frac{\sum_{i=1}^m (c_i - \bar{x})^4 n_i}{(n-1)s_{n-1}^4} - 3$$

Diagrama de cajas

$$L.I. = Q_1 - 1.5 \frac{Q_3 - Q_1}{2},$$

$$L.S. = Q_3 + 1.5 \frac{Q_3 - Q_1}{2}$$

Transformaciones lineales

$$y_i = a + bx_i$$

$$s_y = |b|s_x$$

Teorema de Chebyshev

Independientemente de la forma de la distribución de frecuencias, o de probabilidad, de un conjunto de datos:

$$p(\mu - k\sigma \leq X \leq \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$

Distribuciones bivariantes de frecuencias

distribución conjunta $f_r(x_i, y_j) = \frac{n_{ij}}{n}, \quad \sum_{i,j} f_r(x_i, y_j) = 1$

distribuciones marginales $f_r(x_i) = \sum_j f_r(x_i, y_j), \quad f_r(y_j) = \sum_i f_r(x_i, y_j)$

distribuciones condicionales $f_r(x_i|y_j) = \frac{f_r(x_i, y_j)}{f_r(y_j)}, \quad f_r(y_j|x_i) = \frac{f_r(x_i, y_j)}{f_r(x_i)}$

Covarianza $Cov(X, Y) = \sum_j \sum_i f_r(x_i, y_j)(x_i - \bar{x})(y_j - \bar{y})$

Coefficiente de correlación $\rho(X, Y) = \frac{Cov(X, Y)}{s_x s_y}$

recta de regresión $Y = a + bX, \quad a = \bar{y} - b\bar{x}, \quad b = \frac{Cov(X, Y)}{s_x^2}$

PROBABILIDAD

$$p(A) + p(\bar{A}) = 1$$

$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

sucesos incompatibles $p(A \cup B) = p(A) + p(B), \quad A \cap B = \emptyset$

sucesos condicionados $p(A|B) = \frac{p(A \cap B)}{p(B)}$

sucesos independientes $p(A \cap B) = p(A) \cdot p(B)$

Teorema de Bayes:

Si el espacio muestral lo podemos dividir en n sucesos A_i disjuntos dos a dos, de los que conocemos su probabilidad y la probabilidad condicionada $p(B|A_i)$, entonces

$$p(A_i|B) = \frac{p(B|A_i)p(A_i)}{\sum_{j=1}^n p(B|A_j)p(A_j)}$$

Fórmulas de combinatoria

$$V_{n,k} = \frac{n!}{(n-k)!} \quad C_{n,m} = \frac{n!}{m!(n-m)!} \quad CR_{n,m} = C_{n+m-1,m} = \binom{n+m-1}{m}$$

$$VR_{n,m} = n^m \quad P_n = n! \quad PR_n^{n_1 n_2 \dots} = \frac{n!}{n_1! n_2! \dots} \quad n_1 + n_2 + \dots = n$$

DISTRIBUCIONES DISCRETAS DE PROBABILIDAD

Distribución de Bernoulli $X = \begin{cases} 0 & \text{si es fallo} \\ 1 & \text{si es acierto} \end{cases}$

$$f(x) = \begin{cases} p^x q^{1-x} & x = 0,1 \\ 0 & \text{en otro caso} \end{cases} \quad E(X) = p \quad \text{Var}(X) = p(1-p)$$

Distribución binomial $X = \text{n}^\circ \text{ de aciertos}$

$$f(x) = \begin{cases} C_{n,x} p^x q^{n-x} & x = 0,1,2,\dots,n \\ 0 & \text{en otro caso} \end{cases} \quad E(X) = np \quad \text{Var}(X) = npq$$

Distribución de Poisson $X = \text{n}^\circ \text{ de aciertos}$

$$f(x) = \begin{cases} \frac{e^{-\lambda} \lambda^x}{x!} & x = 0,1,2,\dots,n \\ 0 & \text{en otro caso} \end{cases} \quad E(X) = \text{Var}(X) = \lambda$$

Distribución hipergeométrica $X = \text{n}^\circ \text{ de aciertos}$

$$f(x) = \begin{cases} \frac{C_{k,x} C_{N-k,n-x}}{C_{N,n}} & x = 0,1,2,\dots,n \\ 0 & \text{en otro caso} \end{cases} \quad \begin{array}{l} k : \text{n}^\circ \text{ aciertos población} \\ E(X) = \frac{n \cdot k}{N} \quad \text{Var}(X) = npq \frac{N-n}{N-1} \end{array}$$

Distribución binomial negativa $X: \text{n}^\circ \text{ fracasos antes de obtener } k \text{ éxitos}$

$$p(x; k, p) = \binom{k+x-1}{x} p^k q^x \quad x = 0,1,2,\dots,n$$

$$E(X) = \frac{k(1-p)}{p} \quad \text{Var}(X) = \frac{k(1-p)}{p^2}$$

Valor medio de X: $E(X) = \mu = \sum_x x \cdot f(x)$

Mediana: $p(X \leq m) \leq 1/2 \quad p(X \geq m) \geq 1/2$
 $p(X \leq m) \geq p(X > m)$
 $p(X < m) \leq p(X \geq m)$

Varianza: $\text{Var}(X) = E[(X - \mu)^2] = \sum_x (x - \mu)^2 \cdot f(x)$

Propiedades

Si X_1, X_2, \dots, X_n es una muestra aleatoria de una población cualquiera, y,

$$E(X_i) = \mu \quad \text{Var}(X_i) = \sigma^2 \Rightarrow \begin{cases} \sum_{i=1}^n X_i \approx N(n\mu, n\sigma^2) \\ \bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right) \\ \sqrt{n}\left(\frac{\bar{X} - \mu}{\sigma}\right) \approx N(0,1) \end{cases}$$

Distribución chi-cuadrado

$$\chi_n^2 \rightarrow f(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & x > 0 \\ 0 & x < 0 \end{cases} \quad \begin{matrix} E(X) = n \\ \text{Var}(X) = 2n \end{matrix}$$

Si X_1, X_2, \dots, X_n son v.a. que siguen una distribución normal tipificada $N(0,1)$, entonces $X_1^2 + X_2^2 + \dots + X_n^2$ es una distribución χ_n^2 , con n grados de libertad.

Si X_1, X_2, \dots, X_n son una muestra aleatoria cuya media es \bar{x} , entonces $\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2}$ sigue una una distribución χ_{n-1}^2 , con (n-1) grados de libertad

Distribución t de Student

Si X e Y son v.a. independientes, donde X sigue una dist. normal $N(0, 1)$, e Y una χ_n^2 , entonces, la v.a. $t = \frac{X}{\sqrt{\frac{Y}{n}}}$ sigue una distribución t de Student con n grados de libertad

Si X_1, X_2, \dots, X_n son una muestra aleatoria de una población normal $N(\mu, \sigma^2)$, entonces la variable $\frac{\bar{x} - \mu}{s_{n-1}} \sqrt{n-1}$ sigue una una distribución t_{n-1} , con (n-1) grados de libertad.

Función de distribución: $F(x) = p(X \leq x)$

DISTRIBUCIONES MULTIVARIANTES

Función de probabilidad $f(x, y) = p(X = x, Y = y)$
 $\sum_{x,y} f(x, y) = 1$ $\int f(x, y) dx dy = 1$

Función de distribución
 $F(x, y) = p(X \leq x, Y \leq y)$ $f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$

Distribuciones marginales

$$f_1(x) = \sum_y f(x, y) \quad f_1(x) = \int f(x, y) dy$$

$$f_2(y) = \sum_x f(x, y) \quad f_2(y) = \int f(x, y) dx$$

Distribuciones condicionadas

$$p(X = x | Y = y) = \frac{p(X = x, Y = y)}{p(Y = y)} \quad p(Y = y | X = x) = \frac{p(X = x, Y = y)}{p(X = x)}$$

$$g_1(x|y) = \frac{f(x, y)}{f_2(y)} \quad g_2(y|x) = \frac{f(x, y)}{f_1(x)}$$

VARIABLES ALEATORIAS INDEPENDIENTES $f(x, y) = f_1(x)f_2(y)$

Covarianza $Cov(X, Y) = E[(X - \mu_x)(Y - \mu_y)] = E(XY) - E(X)E(Y)$

Correlación $\rho(X, Y) = \frac{Cov(X, Y)}{\sigma_x \sigma_y}$

Propiedades:

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$$

$$Var(aX + bY + c) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y)$$

ESTIMACION

θ es insesgado si: $E(\hat{\theta}) = \theta$

Función de verosimilitud $L(x_1, \dots, x_n | \theta) = f(x_1 | \theta) \dots f(x_n | \theta)$

Intervalos de confianza:

para la media de una población cualquiera (σ^2 conocida)

$$\mu \in \left[\bar{x} - \frac{\sigma}{\sqrt{n\alpha}}, \bar{x} + \frac{\sigma}{\sqrt{n\alpha}} \right]$$

para la media de una población normal (σ^2 conocida)

$$\mu \in \left[\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$$

para la media de una población normal (σ^2 desconocida, $n < 30$)

$$\mu \in \left[\bar{x} - t_{1-\frac{\alpha}{2}} \frac{S_{n-1}}{\sqrt{n}}, \bar{x} + t_{1-\frac{\alpha}{2}} \frac{S_{n-1}}{\sqrt{n}} \right]$$

para la media de una población normal (σ^2 desconocida, $n \geq 30$)

$$\mu \in \left[\bar{x} - z_{1-\frac{\alpha}{2}} \frac{S_{n-1}}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \frac{S_{n-1}}{\sqrt{n}} \right]$$

para la varianza de una población normal (μ y σ^2 desconocidas)

$$\sigma^2 \in \left[\frac{(n-1)s_{n-1}^2}{b}, \frac{(n-1)s_{n-1}^2}{a} \right]$$

siendo a y b constantes dependientes de la distribución Ji-cuadrado.