

Algunas distribuciones de probabilidad

Distribución	Notación	Soporte R_X	Función de probabilidad	$E(X)$	$Var(X)$
Uniforme discreta	$U\{x_1, \dots, x_K\}$	$\{x_1, \dots, x_K\}$	$p(x) = \begin{cases} \frac{1}{K} & x \in R_X \\ 0 & \text{otro caso} \end{cases}$	$\frac{1}{K} \sum_{i=1}^K x_k$	$\frac{1}{K} \sum_{i=1}^K (x_k - E(X))^2$
Bernoulli	$Be(p)$	$\{0, 1\}$	$p(x) = p^x(1-p)^{1-x}$	p	$p(1-p)$
Binomial	$\mathbf{Bin}(n, p)$	$\{0, 1, \dots, n\}$	$p(x) = \binom{n}{p} p^x(1-p)^{n-x}$	np	$np(1-p)$
Poisson	$\mathcal{P}(\lambda)$	$\{0\} \cup \mathbb{N}$	$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$	λ	λ
Uniforme continua	$U(a, b)$	$[a, b]$	$f(x) = \frac{1}{b-a} I_{[a,b]}(x)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normal	$\mathcal{N}(n, p)$	$\mathbb{R}$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$	μ	σ^2
Exponencial	$\exp(\lambda)$	$\mathbb{R}^+$	$f(x) = \frac{1}{\lambda} \exp\left\{-\frac{x}{\lambda}\right\}$	λ	λ^2

Estimación puntual

Parámetro	Estimador
Media	$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$
Varianza	$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^n (X_i^2 - n\bar{X}^2)}{n-1}$
Correlación	$r = \frac{S_{XY}}{S_x S_y}, \quad S_{XY} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n-1} = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{n-1}$

Concepto	Fórmula
Sesgo	$B(\hat{\theta}) = E(\hat{\theta}) - \theta$
Error de estimación	$ \hat{\theta} - \theta $
Error cuadrático medio	$ECM(\hat{\theta}) = E((\hat{\theta} - \theta)^2) = Var(\hat{\theta}) + B(\hat{\theta})^2$

Regresión Lineal Simple (RLS)

Modelo y Estimadores de Mínimos Cuadrados

El modelo de Regresión Lineal Simple es:

$$Y_i = E(Y|X_i) + \varepsilon_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, \dots, n$$

Cuadro 1: Fórmulas de Regresión Lineal Simple

Característica	Fórmula	Descripción / notas
Pendiente (β_1) Estimador ($\hat{\beta}_1$)	$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$	Coefficiente de la Pendiente
Intercepto (β_0) Estimador ($\hat{\beta}_0$)	$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$	Coefficiente de Intercepto
Ecuación de Regresión Estimada	$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$	Línea de Mejor Ajuste
Varianza del Error (σ^2) Estimador (s^2)	$s^2 = \frac{SSE}{n-2} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2}$	Estimador de la Varianza del Error
Distribución de los estimadores	$\hat{\beta}_i \sim t_{n-2}, \quad i = 1, 2$	
Intervalos de confianza	$\hat{\beta}_i \pm t_{n-2} \cdot ee(\hat{\beta}_i)$	$se(\hat{\beta}_i)$ o $ee(\hat{\beta}_i)$ es el error estándar de $\hat{\beta}_i$
Estadístico t	$t = \frac{\hat{\beta}_i - \beta_i}{ee(\hat{\beta}_i)} \sim t_{n-2}$	
Residuales	$e_i = y_i - \hat{y}_i$	
Coefficiente de determinación (R^2)	$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$	Proporción de Variación Explicada SSR = Suma de cuadrados de la regresión SSE = Suma de cuadrados del error SST = Suma de cuadrados total
Estadístico F	$F = \frac{SSR}{SSE/(n-2)} \sim F_{(1, n-2)}$	

Medidas de Asociación y Varianza

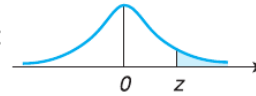
Cuadro 2: Medidas de Asociación y Varianza en RLS

Medida	Fórmula	Descripción
Coefficiente de Correlación (r)	$r = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}}$	Mide la Fuerza y Dirección Lineal
Suma de Cuadrados Total (SST)	$SST = \sum_{i=1}^n (y_i - \bar{y})^2$	Variación Total de Y
Suma de Cuadrados de Regresión (SSR)	$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	Variación Explicada por la Regresión
Suma de Cuadrados del Error (SSE)	$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$	Variación No Explicada (Residual)

Caso	Distribución de Muestreo	Estadístico y su Distribución	Intervalo de Confianza
Media, σ^2 conocida	$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$	$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$	$\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$
Media, σ^2 desconocida	$T \sim t_{(n-1)}$	$T = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{(n-1)}$	$\bar{x} \pm t_{n-1, \alpha/2} \frac{s}{\sqrt{n}}$
Proporción	$P \sim \mathcal{N}\left(p, \frac{pq}{n}\right), \quad q = 1 - p$	$Z^* = \frac{\hat{p} - p}{\sqrt{\hat{p}\hat{q}/n}} \sim \mathcal{N}(0, 1)$	$\hat{p} \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$
Varianza	$J = \frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$	$J = \frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$	$\left(\frac{(n-1)s^2}{\chi_{n-1, \alpha/2}^2}, \frac{(n-1)s^2}{\chi_{n-1, 1-\alpha/2}^2}\right)$
Diferencia de medias, σ_1^2, σ_2^2 conocidas	$\bar{X}_1 - \bar{X}_2 \sim \mathcal{N}\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$	$Z^* = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$	$(\bar{X}_1 - \bar{X}_2) \pm Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
Diferencia de medias, σ_1^2, σ_2^2 desconocidas	$T^* \sim t_{(n_1+n_2-2)}$	$T^* = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$ con $s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$	$(\bar{X}_1 - \bar{X}_2) \pm t_{(n_1+n_2-2), \alpha/2} \cdot s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
Diferencia de proporciones	$\hat{P}_1 - \hat{P}_2 \sim \mathcal{N}\left(p_1 - p_2, \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}\right)$	$Z^{**} = \frac{(\hat{P}_1 - \hat{P}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}} \sim \mathcal{N}(0, 1)$	$(\hat{p}_1 - \hat{p}_2) \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$
Cociente de varianzas	$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1, n_2-1}$	$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1, n_2-1}$	$\left(\frac{S_1^2}{S_2^2} \frac{1}{F_{n_1-1, n_2-1, \alpha/2}}, \frac{S_1^2}{S_2^2} F_{n_2-1, n_1-1, \alpha/2}\right)$
Diferencia de medias p/ muestras dependientes	$\bar{D} \sim t_{(n-1)} \quad s_D^2 = \frac{\sum (d_i - \bar{d})^2}{n-1}$ $d_i = y_i - x_i \quad \bar{d} = \bar{y} - \bar{x}$	$T = \frac{\bar{d} - (\mu_x - \mu_y)}{\frac{s_D}{\sqrt{n}}} \sim t_{(n-1)}$	$\bar{d} \pm t_{(n-1, \alpha/2)} \frac{s_D}{\sqrt{n}}$
Coeficiente de correlación	$T^{**} \sim t_{n-2}$	$T^{**} = \frac{r\sqrt{(n-2)}}{\sqrt{(1-r^2)}} \sim t_{n-2}$	

Pruebas de hipótesis

Notación: z_α corresponde al cuantíl que deja área α a la derecha:



Pruebas de hipótesis para una muestra

Parámetro	Condición	Nula H_0	Alternativa H_a	Estadístico de Prueba	Región Rechazo
Media (μ)	σ^2 conocida	$\mu = \mu_0$	$\mu > \mu_0$ $\mu < \mu_0$ $\mu \neq \mu_0$	$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$	$z \geq z_\alpha$ $z \leq -z_\alpha$ $z \leq -z_{\alpha/2} \text{ o } z \geq z_{\alpha/2}$
Media (μ)	σ^2 desconocida	$\mu = \mu_0$	$\mu > \mu_0$ $\mu < \mu_0$ $\mu \neq \mu_0$	$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim t_{n-1}$	$t \geq t_\alpha$ $t \leq -t_\alpha$ $t \leq -t_{\alpha/2} \text{ o } t \geq t_{\alpha/2}$
Proporción (p)	$n < 25$	$p = p_0$	$p > p_0$ $p < p_0$ $p \neq p_0$	$Y \sim \text{Bin}(n, p_0)$	$Y \geq C_1, C_1 \text{ tal que } P(Y \geq C_1) \approx \alpha$ $Y \leq C_2, C_2 \text{ tal que } P(Y \leq C_2) \approx \alpha$ $Y \leq C_3 \text{ o } Y \geq C_4 \text{ con } C_3 \text{ y } C_4 \text{ tales que } P(Y \leq C_3) \approx P(Y \geq C_4) \approx \alpha/2$
Proporción (p)	$n \geq 25$	$p = p_0$	$p > p_0$ $p < p_0$ $p \neq p_0$	$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim \mathcal{N}(0, 1)$	$z \geq z_\alpha$ $z \leq -z_\alpha$ $z \leq -z_{\alpha/2} \text{ o } z \geq z_{\alpha/2}$
Varianza (σ^2)	—	$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$ $\sigma^2 < \sigma_0^2$ $\sigma^2 \neq \sigma_0^2$	$J = \frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$	$J \geq \chi_{n-1, \alpha}^2$ $J \leq \chi_{n-1, 1-\alpha}^2$ $J \leq \chi_{n-1, \alpha/2}^2 \text{ o } J \geq \chi_{n-1, 1-\alpha/2}^2$

Pruebas de hipótesis para dos muestras

Parámetro	Condición	Nula H_0	Alternativa H_a	Estadístico de Prueba	Región Rechazo
$\mu_1 - \mu_2$	Muestras independientes σ_1^2, σ_2^2 conocidas	$\mu_1 - \mu_2 = \delta$	$\mu_1 - \mu_2 > \delta$ $\mu_1 - \mu_2 < \delta$ $\mu_1 - \mu_2 \neq \delta$	$z^{**} = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0, 1)$	$z^{**} \geq z_\alpha$ $z^{**} \leq -z_\alpha$ $z^{**} \leq -z_{\alpha/2} \text{ o } z^{**} \geq z_{\alpha/2}$
$\mu_1 - \mu_2$	Muestras independientes σ_1^2, σ_2^2 desconocidas pero iguales	$\mu_1 - \mu_2 = \delta$	$\mu_1 - \mu_2 > \delta$ $\mu_1 - \mu_2 < \delta$ $\mu_1 - \mu_2 \neq \delta$	$t^* = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$ $s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$	$t^* \geq t_{\alpha, n_1+n_2-2}$ $t^* \leq -t_{\alpha, n_1+n_2-2}$ $t^* \leq -t_{\alpha/2, n_1+n_2-2} \text{ o } t^* \geq t_{\alpha/2, n_1+n_2-2}$
$\mu_1 - \mu_2$	Muestras independientes σ_1^2, σ_2^2 desconocidas	$\mu_1 - \mu_2 = \delta$	$\mu_1 - \mu_2 > \delta$ $\mu_1 - \mu_2 < \delta$ $\mu_1 - \mu_2 \neq \delta$	$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_\nu$ $\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{s_1^2}{n_1}\right)^2 + \left(\frac{s_2^2}{n_2}\right)^2}$	$t^* \geq t_{\alpha, \nu}$ $t^* \leq -t_{\alpha, \nu}$ $t^* \leq -t_{\alpha/2, \nu} \text{ o } t^* \geq t_{\alpha/2, \nu}$
$p_1 - p_2$	Muestras independientes	$p_1 - p_2 = \delta$	$p_1 - p_2 > \delta$ $p_1 - p_2 < \delta$ $p_1 - p_2 \neq \delta$	$z^{***} = \frac{(\hat{p}_1 - \hat{p}_2) - \delta}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}} \sim \mathcal{N}(0, 1)$	$z^{***} \geq z_\alpha$ $z^{***} \leq -z_\alpha$ $z^{***} \leq -z_{\alpha/2} \text{ o } z^{***} \geq z_{\alpha/2}$
$\frac{\sigma_1^2}{\sigma_2^2}$	Muestras normales	$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$ $\sigma_1^2 < \sigma_2^2$ $\sigma_1^2 \neq \sigma_2^2$	$F = \frac{\max\{s_1^2, s_2^2\}}{\min\{s_1^2, s_2^2\}} \sim F_{n-1, m-1}$ donde n es el tamaño de la muestra con la mayor varianza	$F \geq F_{n-1, m-1, \alpha}$ $F \leq F_{n-1, m-1, 1-\alpha}$ $F \leq F_{n-1, m-1, 1-\alpha/2} \text{ o } F \geq F_{n-1, m-1, \alpha/2}$
$\mu_1 - \mu_2$	Muestras pareadas	$\mu_1 - \mu_2 = \delta$	$\mu_1 - \mu_2 > \delta$ $\mu_1 - \mu_2 < \delta$ $\mu_1 - \mu_2 \neq \delta$	$t^{**} = \frac{\bar{d} - \delta}{s_D / \sqrt{n}} \sim t_{n-1}$	$t^{**} \geq t_{\alpha, n-1}$ $t^{**} \leq -t_{\alpha, n-1}$ $t^{**} \leq -t_{\alpha/2, n-1} \text{ o } t^{**} \geq t_{\alpha/2, n-1}$
ρ	Muestras normales	$\rho = 0$	$\rho > 0$ $\rho < 0$ $\rho \neq 0$	$t^{***} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2}$	$t^{***} \geq t_{\alpha, n-2}$ $t^{***} \leq -t_{\alpha, n-2}$ $t^{***} \leq -t_{\alpha/2, n-2} \text{ o } t^{***} \geq t_{\alpha/2, n-2}$