Análisis exploratorio de datos

Media muestral

$$\bar{x} = \frac{x_1 + \cdots + x_n}{n}$$

(Cuasi)varianza muestral

$$S^2 = \frac{(x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n-1} = \frac{n}{n-1}\left(\frac{x_1^2 + \cdots + x_n^2}{n} - \bar{x}^2\right)$$

Covarianza

$$S_{xy} = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + \cdots + (x_n - \bar{x})(y_n - \bar{y})}{n-1} = \frac{n}{n-1}\left(\frac{x_1y_1 + \cdots + x_ny_n}{n} - \bar{x}\bar{y}\right)$$

Correlación

$$r_{xy} = \frac{S_{xy}}{S_xS_y}$$

Intervalos de confianza

Media de una pob. normal (σ conocida): $\left[\bar{x} \mp z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$

Media de una pob. normal (σ desconocida): $\left[\bar{x} \mp t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}\right]$

Media de una pob. (σ desconocida): $\left[\bar{x} \mp z_{\alpha/2}\frac{s}{\sqrt{n}}\right]$ (n grande)

Proporción: $\left[\hat{p} \mp z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$ (n grande)

Contrastes para la media

$H_0 : \mu \leq \mu_0 \quad R = \left\{\frac{\bar{x}-\mu_0}{s/\sqrt{n}} > t_{n-1,\alpha}\right\}.$

$H_0 : \mu \geq \mu_0 \quad R = \left\{\frac{\bar{x}-\mu_0}{s/\sqrt{n}} < -t_{n-1,\alpha}\right\}.$

$H_0 : \mu = \mu_0 \quad R = \left\{\frac{|\bar{x}-\mu_0|}{s/\sqrt{n}} > t_{n-1,\alpha/2}\right\}.$

Contrastes para dos medias (muestras independientes, varianzas iguales)

$H_0: \ \mu_1 = \mu_2 \quad R = \left\{ \frac{|\bar{x}-\bar{y}|}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} > t_{n_1+n_2-2,\alpha/2} \right\}$, donde $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$

$H_0: \ \mu_1 \leq \mu_2 \quad R = \left\{ \frac{\bar{x}-\bar{y}}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} > t_{n_1+n_2-2,\alpha} \right\}$

Contrastes para dos medias (datos emparejados)

$H_0: \ \mu_1 = \mu_2 \quad R = \left\{ \frac{|\bar{d}|}{S_d/\sqrt{n}} > t_{n-1;\alpha/2} \right\}$, donde $d_i = x_i - y_i$.

$H_0: \ \mu_1 \leq \mu_2 \quad R = \left\{ \frac{\bar{d}}{S_d/\sqrt{n}} > t_{n-1;\alpha} \right\}$.

Contrastes para una proporción (n grande):

$H_0 : p \leq p_0 \quad R = \left\{ \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} > z_\alpha \right\}$

$H_0 : p \geq p_0 \quad R = \left\{ \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} < -z_\alpha \right\}$

$H_0 : p = p_0 \quad R = \left\{ \frac{|\hat{p}-p_0|}{\sqrt{\frac{p_0(1-p_0)}{n}}} > z_{\alpha/2} \right\}$

Contrastes para dos proporciones (n_1 y n_2 grandes):

$H_0: \ p_1 \leq p_2 \quad R = \left\{ \frac{\hat{p}_1-\hat{p}_2}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} > z_\alpha \right\}$, donde $\bar{p} = \frac{n_1\hat{p}_1+n_2\hat{p}_2}{n_1+n_2}$.

$H_0: \ p_1 = p_2 \quad R = \left\{ \frac{|\hat{p}_1-\hat{p}_2|}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} > z_{\alpha/2} \right\}$

Regresión lineal simple:

Recta de mínimos cuadrados

$$y - \bar{y} = \frac{S_{xy}}{S_x^2}(x - \bar{x}) \quad \text{ó} \quad y - \bar{y} = r_{xy}\frac{S_y}{S_x}(x - \bar{x})$$

Intervalos de confianza: $\left[\hat{\beta}_i \mp t_{n-2,\alpha/2} \times \text{E.T.}(\hat{\beta}_i)\right]$

Contrastes:

$H_0 : \beta_i \leq \beta_i^* \quad R = \left\{ \frac{\hat{\beta}_i - \beta_i^*}{\text{E.T.}(\hat{\beta}_i)} > t_{n-2,\alpha} \right\}.$

$H_0 : \beta_i \geq \beta_i^* \quad R = \left\{ \frac{\hat{\beta}_i - \beta_i^*}{\text{E.T.}(\hat{\beta}_i)} < -t_{n-2,\alpha} \right\}.$

$H_0 : \beta_i = \beta_i^* \quad R = \left\{ \frac{|\hat{\beta}_i - \beta_i^*|}{\text{E.T.}(\hat{\beta}_i)} > t_{n-2,\alpha/2} \right\}.$