- Fecha esperada de terminación:

---

1. Considere el modelo de regresión lineal simple que pasa por el origen

$$y = \beta x + \epsilon$$

donde $\epsilon \sim \mathrm{N}(0, \sigma^2)$.

a) Muestre que el estimador de mínimos cuadrados de β está dado por

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}$$

b) Muestre que $\hat{\beta}$ es insesgado y encuentre $\mathrm{E}[\hat{y}]$.

c) Considere el estimador insesgado de la varianza s^2 dado por

$$s^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n-1} \sim \frac{\sigma^2}{n-1}\,\chi^2_{n-1}$$

y construya intervalos de confianza para: el coeficiente β; la respuesta media al nivel x, $\hat{y}(x)$; y para una nueva observación en x, $\dot{y}(x)$.

2. Considere el problema de regresión lineal simple

$$y = \beta_0 + \beta_1 x + \epsilon$$

y suponga ahora que el regresor X es también una variable aleatoria de manera que (X, Y) sigue una distribución normal bivariada con función de densidad conjunta

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\,\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}$$

donde μ_* y σ_*^2 son las medias y varianzas correspondientes y ρ el coeficiente de correlación entre X y Y. Esto es,

$$\rho = \mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$$

Muestre que dado $X = x$, la *v. a.* respuesta y se distribuye normalmente con media $(\beta_0 + \beta_1 x)$ y varianza $\sigma_X(1-\rho^2)$. Abusando de la notación,

$$(Y|X = x) \sim \mathrm{N}\left(\beta_0 + \beta_1 x,\ \sigma_Y^2(1-\rho^2)\right)$$

y donde

$$\beta_0 = \mu_Y - \rho\,\frac{\sigma_Y}{\sigma_X}\,\mu_X, \quad \text{y} \quad \beta_1 = \rho\,\frac{\sigma_Y}{\sigma_X}$$

Esto es,

$$\mathrm{E}[Y|X = x] = \beta_0 + \beta_1 x$$

$\mathrm{E}[Y|X = x]$ se conoce como la *ecuación de regresión.*

3. Considere el modelo de regresión lineal simple

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

para $i = 1, \ldots, n$. Tome los estimadores de mínimos cuadrados

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}; \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

donde $S_{xy} = \sum(x_i - \bar{x})(y_i - \bar{y})$ y $S_{xx} = \sum(x_i - \bar{x})^2$ y considere la respuesta ajustada por el modelo al nivel x_i por

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Muestre entonces que:

a) La *suma de cuadrados totales sin corregir* se puede descomponer por

$$\sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2 + n\bar{y}^2$$

b) La *suma de cuadrados totales corregida* ($\mathrm{SC_{Tot}}$) se puede descomponer como la *suma de cuadrados debida a la regresión* ($\mathrm{SC_{Mod}}$) mas la *suma de cuadrados de los residuales* ($\mathrm{SC_{Res}}$). Esto es,

$$\begin{aligned} \mathrm{SC_{Tot}} &= \mathrm{SC_{Mod}} + \mathrm{SC_{Res}} \\ \sum_{i=1}^{n}(y_i - \bar{y})^2 &= \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \end{aligned}$$

c) Concluya que el *coeficiente de determinación* R^2

$$R^2 = \frac{\mathrm{SC_{Mod}}}{\mathrm{SC_{Tot}}} = 1 - \frac{\mathrm{SC_{Res}}}{\mathrm{SC_{Tot}}}$$