Solución preguntas 3-6 Guía 1 Control 3

P3)

a) Para probar lo anterior, solo basta con desarrollar las expresiones:

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{\sum [x_i y_i - \bar{y} x_i - \bar{x} y_i + \bar{x}\bar{y}]}{\sum [x_i^2 - 2\bar{x} x_i + \bar{x}^2]} = \frac{\sum x_i y_i - \bar{y} \sum x_i - \bar{x} \sum y_i + n\bar{x}\bar{y}}{\sum x_i^2 - 2\bar{x} \sum x_i + n\bar{x}^2}$$

$$= \frac{\sum x_i y_i - \bar{y} n\bar{x} - \bar{x} n\bar{y} + n\bar{x}\bar{y}}{\sum x_i^2 - 2\bar{x} n\bar{x} + n\bar{x}^2} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

Por otro lado:

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x}) y_i - \sum (x_i - \bar{x})\bar{y}}{\sum (x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2}$$

Porque:

$$\sum_{i=1}^{n} (x_i - \bar{x})\bar{y} = \sum_{i=1}^{n} (x_i \bar{y} - \bar{x}\bar{y}) = \bar{y} \sum_{i=1}^{n} x_i - n\bar{x}\bar{y} = \bar{y} n\bar{x} - n\bar{x}\bar{y} = 0$$

b) Se tiene que:

$$\mathrm{Cov}(\bar{Y}, \hat{\beta}_1) = \mathrm{Cov}\left(\bar{Y}, \frac{\sum (x_i - \bar{x}) Y_i}{\sum (x_i - \bar{x})^2}\right) = \frac{1}{\sum (x_i - \bar{x})^2} \mathrm{Cov}\left(\bar{Y}, \sum_{i=1}^{n} (x_i - \bar{x}) Y_i\right)$$

$$= \frac{1}{\sum (x_i - \bar{x})^2} \sum_{i=1}^{n} (x_i - \bar{x}) \mathrm{Cov}(\bar{Y}, Y_i)$$

Pero:

$$\mathrm{Cov}(\bar{Y}, Y_i) = \mathrm{Cov}\left(\frac{1}{n} \sum_{i=1}^{n} Y_i, Y_i\right) = \frac{1}{n} \sum_{j=1}^{n} \mathrm{Cov}(Y_j, Y_i) = \frac{\sigma^2}{n}$$

Luego:

$$\mathrm{Cov}(\bar{Y}, \hat{\beta}_1) = \frac{1}{\sum (x_i - \bar{x})^2} \sum_{i=1}^{n} (x_i - \bar{x}) \mathrm{Cov}(\bar{Y}, Y_i) = \frac{\sum (x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \frac{\sigma^2}{n} = 0$$

c) Se tiene:

$$\mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_0) = \mathrm{Cov}(\hat{\beta}_1, \bar{Y} - \hat{\beta}_1 \bar{x}) = \mathrm{Cov}(\hat{\beta}_1, \bar{Y}) - \bar{x}\, \mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_1) = 0 - \bar{x}\, \mathrm{Var}(\hat{\beta}_1) = \frac{-\bar{x}\sigma^2}{\sum (x_i - \bar{x})^2}$$

d)

Manera larga

Desarrollar cada expresión del lado derecho y reemplazar luego por $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ en cada caso, para después reemplazar las fórmulas de $\hat{\beta}_0$ y $\hat{\beta}_1$ respectivamente.

Manera corta

Notar que:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i = (\bar{y} - \hat{\beta}_1 \bar{x}) + \hat{\beta}_1 x_i \Rightarrow \hat{y}_i - \bar{y} = \hat{\beta}_1 (x_i - \bar{x})$$

Luego:

$$\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = \sum_{i=1}^{n} \hat{\beta}_1^2 (x_i - \bar{x})^2$$

y:

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - (\bar{y} + \hat{\beta}_1(x_i - \bar{x}))^2 = \sum_{i=1}^{n} \left[(y_i - \bar{y}) - \hat{\beta}_1(x_i - \bar{x})\right]^2$$

Luego sumando ambas expresiones:

$$\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \hat{\beta}_1^2(x_i - \bar{x})^2 + \sum_{i=1}^{n} \left[(y_i - \bar{y}) - \hat{\beta}_1(x_i - \bar{x})\right]^2$$

$$= \sum_{i=1}^{n} (y_i - \bar{y})^2 - 2\hat{\beta}_1 \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) + 2\hat{\beta}_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$= \sum_{i=1}^{n} (y_i - \bar{y})^2 - 2\hat{\beta}_1 \left[\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) - \hat{\beta}_1 \sum_{i=1}^{n} (x_i - \bar{x})^2\right]$$

Pero hay que recordar que:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Luego el segundo término es 0, de donde se obtiene lo pedido.

e) Para obtener el resultado pedido basta recordar dos cosas:

$$\tilde{\sigma}^2 = \frac{\text{SSR}}{n-2} \quad \text{y que} \quad R^2 = 1 - \frac{\text{SSR}}{\text{SST}}$$

Combinando ambas expresiones se tiene que:

$$\tilde{\sigma}^2 = \frac{\text{SST}(1 - R^2)}{n-2} = \frac{n S_y^2(1 - R^2)}{n-2}$$

P4)

De acuerdo al enunciado del problema, la situación se puede modelar como:

$$Y_i = \mu + \varepsilon_i \quad (i = 1, ..., n)$$

En que Y_i es el peso corporal observado, μ es el peso real del individuo y ε_i es el error, tal que $\varepsilon_i \sim N(0, \sigma^2)$. La solución MCO para μ es:

$$Q = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \mu)^2 \text{ , } \frac{\partial Q}{\partial \mu} = 0 \Rightarrow \hat{\mu} = \bar{y}$$

Desde el punto de vista clásico, el problema es equivalente a encontrar un estimador para el parámetro μ de una v.a. $Y \sim N(\mu, \sigma^2)$. De donde, al aplicar la técnica de Máxima Verosimilitud se obtiene $\hat{\mu} = \bar{y}$, que es idéntico al resultado anterior. Los resultados son idénticos ya que bajo los supuestos usuales del modelo lineal, la estimación por MCO coincide con la estimación Máximo Verosímil. Si el término del error tuviera otra distribución y se usaran otros métodos de estimación para ambos casos, los resultados podrían diferir.

P5)

a)

En la salida de un modelo de regresión se tienen las siguientes relaciones:

$$T = \frac{\text{Coeficiente}}{\text{Desv. Est.}}$$

$$g.l.\,\text{Total} = g.l.\text{Regresion} + g.l.\text{Residuos}$$

$$\text{Suma de Cuadrados Total} = \text{Suma de Cuadrados Regresion} + \text{Suma Cuadrados Residuos}$$

$$\text{Cuadrado Medio} = \frac{\text{Suma de Cuadrados}}{g.l.}$$

$$F = \frac{\text{Cuadrados Medios Regresion}}{\text{Cuadrados Medios Residuos}}$$

Usando estas relaciones se tienen los siguientes resultados:

$$\text{Coef. de } X\,(\hat{\beta}_1) = T \cdot \text{Desv.Est} = 0,05012 \cdot 18,03 = \; 0,90364$$

$$\text{Desv.Est. Cte}\,(\sigma_{\hat{\beta}_0}) = \frac{\text{Coef. de Cte.}}{T} = \frac{3,830}{2,17} = 1,768$$

$$g.l.\text{Residuos} = g.l.\text{Total} - g.l.\text{Regresion} = 32 - 1 = 31$$

$$P - \text{Valor}\,(\hat{\beta}_1) \approx 0$$

$$\text{Suma Cuadrados Regresion} = \text{Suma Cuadrados Total} - \text{Suma Cuadrados Residuos} = 3390,6$$

$$\text{CM Regresion} = \frac{\text{SC Regresion}}{g.l.\text{Regresion}} = \frac{3390,6}{1} = 3390,6 \quad \text{CM Residuos} = \frac{323,3}{31} = 10,4$$

$$F = \frac{\text{CM Regresion}}{\text{CM Residuos}} = \frac{3390,6}{10,4} = 325,08$$

Como el valor de F es muy grande, se tiene que $P - \text{Valor} \approx 0$.

Los resultados de las tablas muestran que tanto las variables por separado como el modelo conjunto son significativos.

b) Una estimación insesgada para σ^2 es:

$$\tilde{\sigma}^2 = \frac{\text{SSR}}{n-2} = \frac{\text{Suma Cuadrados Residuos}}{33-2} = \frac{323,3}{31} = 10,4$$

c) Un intervalo de confianza para β_1 es

$\text{IC}_{\beta_1} = [\hat{\beta}_1 - t_{n-2}^{(\alpha/2)} \cdot \sigma_{\hat{\beta}_1}; \hat{\beta}_1 + t_{n-2}^{(\alpha/2)} \cdot \sigma_{\hat{\beta}_1}] = [0,90364 - 2,045 \cdot 0,05012; 0,90364 + 2,045 \cdot 0,05012] =$ $[0,8011; 1,0061]$

d) Se tiene el siguiente conjunto de hipótesis:

$$\begin{aligned} H_0 &: \beta_1 = 1 \\ H_1 &: \beta_1 < 1 \end{aligned}$$

El estadístico de contraste es:

$$t = \frac{\hat{\beta}_1 - \beta_1^{(0)}}{\sigma_{\hat{\beta}_1}} = \frac{0,90364 - 1,0}{0,05012} = -1,92$$

Se rechazará la hipótesis $\hat{\beta}_1 < C$. Se tiene que, fijando un nivel $\alpha = 0,05$:

$$\Pr(\hat{\beta}_1 < C | \beta_1 = 1) = 0,05$$

$$\Pr(\hat{\beta}_1 < C | \beta_1 = 1) = \Pr\left(\frac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} < \frac{C - \beta_1}{\sigma_{\hat{\beta}_1}} \right) = \Pr\left(t < \frac{C - \beta_1}{\sigma_{\hat{\beta}_1}} \right) = 0,05 \Rightarrow \frac{C - \beta_1}{\sigma_{\hat{\beta}_1}} = -1,697$$

De donde $C \approx 0,91$, y como $\hat{\beta}_1 = 0,90364 < C = 0,91 \Rightarrow$ Se rechaza H_0.

e) Una predicción para x=20:

$$y = 3,830 + 0,90364 \cdot 20 \approx 21,9$$

P6)

Este es un problema que se resuelve a través de tablas de contingencia. Debemos encontrar las frecuencias teóricas esperadas de defectos para los tres turnos, y para ello necesitamos conocer los totales de cada fila y columna, los que son:

Total Matutino (M)= 45+905=950

Total Vespertino (V)= 55+890=945

Total Nocturno (N)=70+870=940

Total Defectuosos (D)= 45+55+70=170

Total No defectuosos (ND)= 905+890+870=2665

Total General = 2835

$$e_{M,D} = \frac{950 \cdot 170}{2835} = 57,0 \quad e_{V,D} = \frac{945 \cdot 170}{2835} = 56,7 \qquad e_{N,D} = 170 - 57,0 - 56,7 = 56,3$$

$$e_{M,ND} = 950 - 57 = 893 \quad e_{V,ND} = 945 - 56,7 = 888,3 \quad e_{N,ND} = 940 - 56,3 = 883,7$$

El valor de Q es entonces:

$$Q = \frac{(45 - 57)^2}{57} + \frac{(55 - 56,7)^2}{56,7} + \ldots + \frac{(870 - 883,7)^2}{883,7} = 6,29$$

Lo cual tiene asociado un P-valor de $\approx$ 0,04, por lo que se rechaza la hipótesis de que los defectos no dependen del turno.