Solución de algunos ejercicios de autocomprobación

2. *a*) Ajustamos la recta de regresión:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i = \bar{Y} + \hat{\beta}_1 (X_i - \bar{X})$$

Hacemos algunos cálculos a partir de los datos que nos dan:

$$\bar{x} = \frac{\sum x_i}{n} = 69.3 \quad V(X) = \frac{\sum x_i^2}{n} - \bar{x}^2 = 11.61$$
$$\bar{y} = \frac{\sum y_i}{n} = 157.5 \quad V(Y) = \frac{\sum y_i^2}{n} - \bar{y}^2 = 618.25$$
$$Cov(X,Y) = \frac{\sum x_i y_i}{n} - \bar{x}\bar{y} = 80.05$$

$$\hat{\beta}_1 = \frac{Cov(X,Y)}{V(X)} = 6.8949$$

con lo que la Recta de Regresión será:

$$\hat{Y}_i = 157.5 + 6.8949(X_i - 69.3) = 6.8949 X_i - 320.31$$

El coeficiente de correlación lineal es

$$r = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}} = 0.9448$$

b) La pendiente del modelo es β_1, así que calculamos un IC al 90 % para el parámetro β_1:

$$\left(\hat{\beta}_1 \pm t_{n-2,\frac{\alpha}{2}} \sqrt{\frac{\hat{S_R^2}}{nV(X)}} \right)$$

Buscando en las tablas $t_{8,0.05} = 1.86$. Falta calcular la varianza residual $\hat{S_R^2}$.

$$\hat{S_R^2} = \frac{(1-r^2)nV(Y)}{n-2} = 82.8896$$

También se puede calcular la varianza residual a partir de los elementos de la Tabla ANOVA.

El IC que se obtiene es:

$$(5.3233; 8.4665)$$

Como el intervalo de confianza obtenido está a la derecha del valor 5 **no podemos rechazar** la afirmación de la agencia de publicidad y, por tanto, podemos admitir que mil euros gastados en publicidad aporta más de cinco mil euros de beneficios (en base a la muestra observada).

4. *a*) Ajustamos la recta de regresión:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i = \bar{Y} + \hat{\beta}_1(X_i - \bar{X})$$

Hacemos algunos cálculos a partir de los datos que nos dan:

$$\bar{x} = \frac{\sum x_i}{n} = 5.375 \quad V(X) = \frac{\sum x_i^2}{n} - \bar{x}^2 = 2.484375$$
$$\bar{y} = \frac{\sum y_i}{n} = 4 \quad V(Y) = \frac{\sum y_i^2}{n} - \bar{y}^2 = 2$$
$$Cov(X,Y) = \frac{\sum x_i y_i}{n} - \bar{x}\bar{y} = 1$$

$$\hat{\beta}_1 = \frac{Cov(X,Y)}{V(X)} = 0.4025157$$

con lo que la Recta de Regresión será:

$$\hat{Y}_i = 4 + 0.4025(X_i - 5.375) = 0.4025X_i + 1.8365$$

b) La pendiente del modelo es β_1, así que calculamos un IC al 90 % para el parámetro β_1:

$$\left(\hat{\beta}_1 \pm t_{n-2,\frac{\alpha}{2}} \sqrt{\frac{\hat{S_R^2}}{nV(X)}}\right)$$

Buscando en las tablas $t_{6,0.05} = 1.943$. Falta calcular la varianza residual $\hat{S_R^2}$. Para ello necesitamos calcular primero el coeficiente de correlación r.

$$r = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}} = 0.4486$$

con lo que

$$\hat{S_R^2} = \frac{(1-r^2)nV(Y)}{n-2} = 2.13$$

También se puede calcular la varianza residual a partir de los elementos de la Tabla ANOVA.

El IC que se obtiene es:

$$(-0.233576, 1.038576)$$

5. a) Para calcular la recta de regresión con $X =$*precio del billete* e $Y =$*miles de viajeros*, primero realizamos algunos cálculos:

$$\sum x_i = 340 \qquad \sum y_i = 5500$$

$$\bar{x} = 42.5 \qquad \bar{y} = 687.5$$

$$\sum x_i^2 = 15500 \qquad \sum y_i^2 = 3830800$$

$$V(X) = 131.25 \qquad V(Y) = 6193.75$$

$$\sum x_i y_i = 227200$$

$$Cov(X,Y) = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{n} = -818.75$$

Calculamos los estimadores de la pendiente y del término independiente:

$$\hat{\beta}_1 = \frac{cov(X,Y)}{V(X)} = \frac{-818.75}{131.25} = -6.238$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X} = 687.5 + (6.238)(42.5) = 952.62$$

con lo que la Recta de Regresión es:

$$\hat{Y}_i = 952.62 - 6.238X_i$$

Para el coeficiente de correlación:

$$r = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{-818.75}{\sqrt{(131.25)(6193.75)}} = -0.908$$

b) Debemos realizar el contraste:

$$\begin{cases} H_0 & : & \beta_1 & = & -5 \\ H_1 & : & \beta_1 & < & -5 \end{cases}$$

El signo negativo del 5 se debe a que nos piden comprobar si el número de pasajeros que se **pierden** al aumentar en 1 céntimo el precio del billete puede ser 5. Sabemos que:

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{S_R^2}}{nV(X)}}} \sim t_{n-1}$$

Calculamos $\hat{d}$:

$$\hat{d} = \frac{-6.238 - (-5)}{\sqrt{\frac{1448.4}{8\times 131.25}}} = \frac{-1.238}{1.174} = -1.05$$

Como es un contraste unilateral por la izquierda, debemos buscar el percentil $t_{n-2,\alpha} = t_{6,0.05} = 1.94$ y comparar el valor de $\hat{d}$ con el -1.94. Como $-1.05 > -1.94$ no podemos rechazar H_0 con un 95 % de confianza.

Para calcular el p-valor interpolamos entre los percentiles $t_{6,0.1} = 1.44$ y $t_{6,0.2} = 0.906$[1], obteniendo:

$$P(t_6 > 1.05) = P(t_6 < -1.05) = 0.173$$

que está en la región de duda.

Observación: También podíamos haber realizado el contraste unilateral por la derecha:

$$\begin{cases} H_0 & : & \beta_1 & = & 5 \\ H_1 & : & \beta_1 & > & 5 \end{cases}$$

En ese caso,

$$\hat{d} = \frac{6.238 - 5}{\sqrt{\frac{1448.4}{8 \times 131.25}}} = \frac{1.238}{1.174} = 1.05$$

y debíamos comparar este valor con el percentil positivo $t_{n-2,\alpha} = t_{6,0.05} = 1.94$. Como $1.05 < 1.94$ no podemos rechazar H_0.

[1]Si utilizamos la tabla que está actualmente en la web, interpolaríamos entre $t_{6,0.1} = 1.44$ y $t_{6,0.5} = 0$, en cuyo caso obtendrímos un p-valor igual a 0.21 y no rechazarímos H_0