ESTADISTICA II

INGENIERIA INFORMATICA, 3^{ER} Curso

26 - Junio - 2.008 **Primera Parte - Test**

Apellidos y Nombre: ....

D.N.I. : ....

Nota : En la realización de este examen sólo esta permitido utilizar calculadoras que, a lo sumo, tengan funciones estadísticas básicas. **No se pueden utilizar calculadoras programables.**

Existe una sóla respuesta correcta por pregunta.

Cada respuesta correcta se valorará con 1 punto y cada incorrecta con -1/3. Las preguntas no contestadas no se valoran. Si se marcan varias respuestas a la vez se considerará la pregunta no contestada.

El valor de esta primera parte del examen es de CINCO PUNTOS sobre diez.

Responder con letras mayúsculas y bolígrafo.

Las respuestas elegidas que se considerarán válidas son las que se consignen en el cuadro que se adjunta a continuación.

Pregunta	**1**	**2**	**3**	**4**	**5**	**6**
Respuesta	**A**	**B**	**D**	**A**	**A**	**C**
Pregunta	**7**	**8**	**9**	**10**	**11**	**12**
Respuesta	**C**	**D**	**C**	**C**	**A**	**B**

CUESTIONES

1. Se desea contrastar al 90% la hipótesis "$\mu_X = 100$", esto es, la media de la variable X es igual a 100. Se utilizan dos contrastes: d_1 y d_2, el contraste d_2 es **más potente** que d_1.Entonces

A. P(aceptar "$\mu_X = 100$"/ $\mu_X = 90$) es mayor con d_1 que con d_2. **Solución**

B. P(rechazar "$\mu_X = 100$"/ $\mu_X = 100$) es mayor con d_1 que con d_2.

C. P(error tipo I) es mayor con d_1 que con d_2.

D. Ninguna de las otras respuestas.

2. En un diseño de experimentos de dos factores fijos con interacción y replicación se supone que

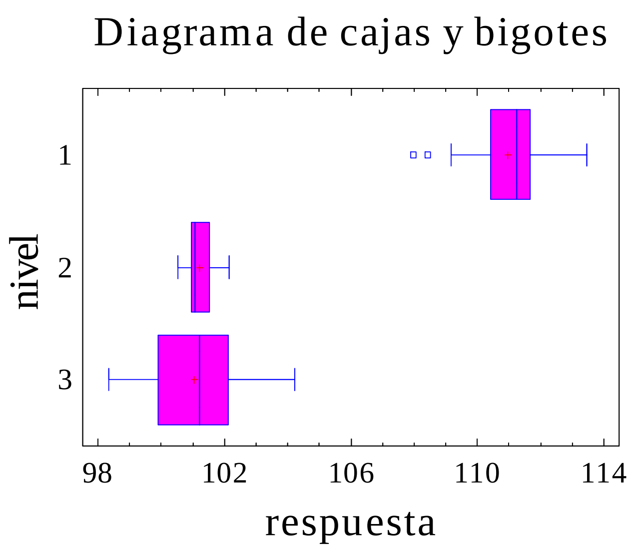
A. $\hat{Y}_{ij} \sim N(\mu + \alpha_i + \beta_j;\ \sigma^2)$.

B. $Y_{ijk} \sim N(\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij};\ \sigma^2)$. **Solución**

C. $\hat{Y}_{ij} \sim t_{g.l.}$

D. $Y_{ijk} \sim N(0;\ \sigma^2)$

3. En un ANOVA de una vía, se representa, para cada nivel del factor, un gráfico de cajas y bigotes de la respuesta. De él se deduce



A. Que el factor no es significativo

B. Que se verifica la hipótesis de normalidad.

C. Que es necesario introducir otro factor explicativo en el modelo.

D. Ninguna de las otras respuestas. **Solución**

4. Al ajustar un diseño de experimentos con un factor **aleatorio**: $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, $i = 1, 2, 3, 4, \quad j = 1, 2, \ldots, 10$. Se obtiene $\hat{s}_Y^2 = 3$ (cuasivarianza muestral de la variable Y) y $\hat{s}_R^2 = 1$ (varianza residual). Entonces la estimación de $Var(\tau_i) = \sigma_\tau^2$ es

A. 2.6 **Solución**

B. 5.4

C. 7.8

D. Ninguna de las otras tres respuestas.

Solución:

$$
\begin{aligned}
\hat{s}_R^2 &= 1 = \frac{scR}{40-4} \Rightarrow scR = 36 \\
\hat{s}_Y^2 &= 3 = \frac{scT}{40-1} \Rightarrow scT = 117 \\
scE &= scT - scR = 117 - 36 = 81 \\
\frac{81}{4-1} &= 27 = 10 \cdot \hat{\sigma}_\tau^2 + 1 \Rightarrow \hat{\sigma}_\tau^2 = 2.6
\end{aligned}
$$

5. En un cuadrado latino con cinco niveles en cada factor se ha obtenido que $scR = \sum_t e_t^2 = 100$. Un intervalo de confianza al 90% para la varianza del modelo es

A. $(4.76; 19.14)$ **Solución**

B. $(2.16; 31.35)$

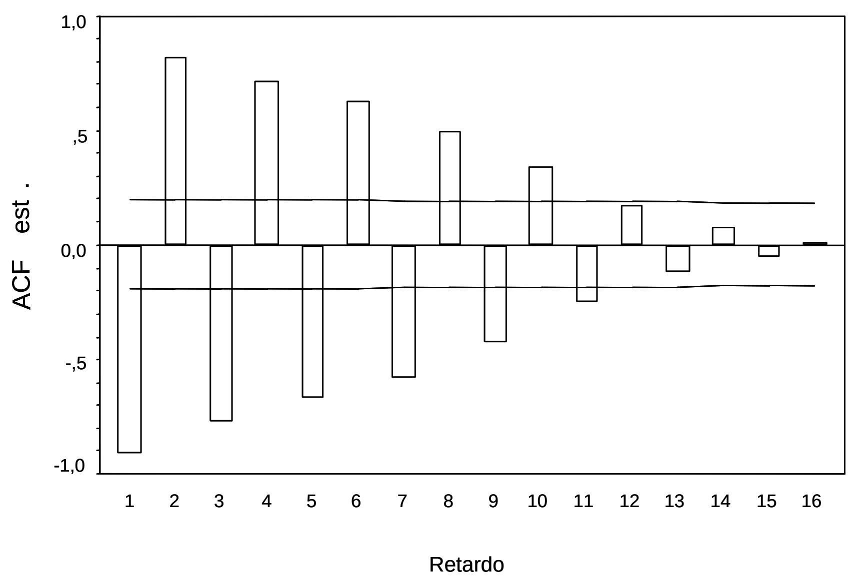
C. $(11.12; 33.37)$

D. Ninguna de las otras tres respuestas.

Solución:

$$\frac{scR}{\sigma^2} \sim \chi^2_{g.l} \Rightarrow \frac{100}{\sigma^2} \sim \chi^2_{12} \Rightarrow 5.226 < \frac{100}{\sigma^2} < 21.026$$
$$4.756 = \frac{100}{21.026} < \sigma^2 < \frac{100}{5.226} = 19.135$$

6. El correlograma de los residuos de un diseño de experimentos con dos factores es el de la figura adjunta. Se deduce que :



A. Se verifica la hipótesis de independencia.

B. Existe dependencia positiva.

C. Existe dependencia negativa. **Solución**

D. No se puede deducir nada de la hipótesis de independencia porque habría que trabajar con los residuos tipificados

7. Al ajustar un modelo de regresión múltiple el contraste de Durbin-Watson proporciona el valor $\hat{d} = 0.8$. Por tanto

A. Se verifica la hipótesis de independencia.

B. La primera autocorrelación de los residuos es aproximadamente $\hat{r}_1 \approx +0.3$.

C. Existe dependencia positiva en la variable de interés. **Solución.**

D. La primera autocorrelación de los residuos es aproximadamente $\hat{r}_1 \approx -0.8$.

Solución.

$$\hat{d} = 0.8 \approx 2 \cdot (1 - \hat{r}_1) \Rightarrow \hat{r}_1 \approx 1 - \frac{0.8}{2} = 0.6$$

8. En el ajuste de un modelo de regresión lineal múltiple $\tilde{\mathbf{Y}} = \mathbf{X}\tilde{\boldsymbol{\alpha}} + \tilde{\boldsymbol{\varepsilon}}$. Del teorema de Gauss-Markov se deduce que

A. Los estimadores por mínimos cuadrados de $\tilde{\boldsymbol{\alpha}}$ son de menor error cuadrático medio que los estimadores por máxima verosimilitud.

B. Los estimadores por mínimos cuadrados de $\tilde{\boldsymbol{\alpha}}$ son los de menor error cuadrático medio.

C. Los estimadores por mínimos cuadrados de $\tilde{\boldsymbol{\alpha}}$ son los de menor varianza.

D. Ninguna de las otras tres respuestas. **Solución.**

9. A partir de la muestra $\{(X_i, Y_i)\}_{i=1}^{200}$ se ha ajustado la recta de regresión de Y sobre X, obteniendo un coeficiente de determinación R^2_{XY}. Con la misma muestra se ha ajustado la recta de regresión de X sobre Y, obteniendo un coeficiente de determinación R^2_{YX}. Entonces:

A. En general, $R^2_{XY} \neq R^2_{YX}$.

B. No existe relación entre R^2_{XY} y R^2_{YX}.

C. Siempre, $R^2_{XY} = R^2_{YX}$. **Solución.**

D. Ninguna de las otras tres respuestas.

10. Al ajustar varios modelos de regresión múltiple a una nube de observaciones. Entre los posibles criterios para seleccionar el mejor modelo se verifica

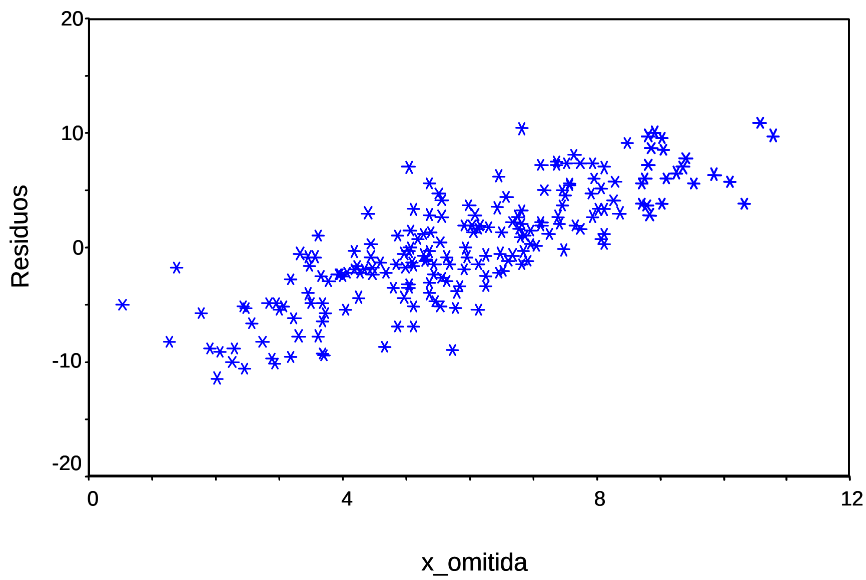
A. Maximizar $\bar{R}^2$ es equivalente a maximizar R^2.

B. Maximizar R^2 es equivalente a minimizar $\hat{s}^2_R$.

C. Maximizar $\bar{R}^2$ es equivalente a minimizar $\hat{s}^2_R$. **Solución.**

D. Ninguna de las otras tres respuestas.

11. Al ajustar un modelo de regresión lineal múltiple con cuatro regresoras se obtiene el siguiente gráfico de residuos frente a una variable no introducida en el modelo. Se deduce que



A. La variable omitida debe introducirse en el modelo. **Solución**

B. La variable omitida no debe introducirse en el modelo por el problema de multicolinealidad.

C. El modelo ajustado es adecuado porque la media de los residuos es cero.

D. Esta variable no debe introducirse en el modelo porque disminuiría el $\bar{R}^2$.

12. Al ajustar un modelo de regresión lineal múltiple con cuatro regresoras. La regresora X_4 es ortogonal a las otras tres regresoras. Entonces

A. $FIV(X_4) < 1$.

B. $FIV(X_4) = 1$. **Solución**

C. $1 < FIV(X_4) \leq 10$.

D. $10 \leq FIV(X_4) \leq 30$.

ESTADISTICA II, Ingeniería Informática,
Problemas, 26 - Junio - 2.008

Cada una de los dos problemas tiene una valoración de 2.5 puntos sobre diez.

Para aprobar el examen es necesario obtener una puntuación igual o superior a 1 punto en cada uno de los dos problemas.

Problema 1.

En un estudio (e.g. Brems y Whitten, 1987) sobre las ventajas de emplear iconos en el desarrollo de interfaces gráficas, se consideraron tres tipos de interfaz (Sólo iconos, Iconos y comandos, sólo comandos) y se anotaron las preferencias (valoraciones de 1 a 10) de cinco usuarios (variable bloque). Los resultados obtenidos fueron los siguientes:

	Interfaz		
Usuario	Iconos solo	Iconos y comandos	Comandos solo
1	4	6	5
2	3	5	6
3	3	5	5
4	4	4	6
5	4	7	7

$\Sigma y_{ij} = 74$
$\Sigma y_{ij}^2 = 388$

P.1. Formular el modelo de diseño de experimentos asociado a este problema y obtener las estimaciones puntuales de (todos) los parámetros del modelo.

Solución:

Se trata de un diseño en bloques completamente aleatorizado sin replicación, con $I = 3$ niveles del factor y $J = 5$ bloques. Se supone que no hay interacción entre el factor y la variable bloque. La formulación matemática del modelo es la siguiente:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

con ε_{ij} v.a. independientes con distribución $N\left(0, \sigma^2\right)$.

Para estimar los efectos, calculamos las medias por filas y por columnas y restamos la media global:

$\overline{y}_{ij.}$	I_1	I_2	I_3	$\overline{y}_{.j}$	$\widehat{\beta}_j = \overline{y}_{.j} - \overline{y}_{..}$
U_1	4	6	5	5	0.07
U_2	3	5	6	4.67	-0.26
U_3	3	5	5	4.33	-0.6
U_4	4	4	6	4.67	-0.26
U_5	4	7	7	6	1.07
$\overline{y}_{i.}$	3.6	5.4	5.8	$\overline{y}_{...} = 4.933$	
$\widehat{\alpha}_i = \overline{y}_{i.} - \overline{y}_{..}$	-1.33	0.47	0.87		

Para estimar la varianza es preferible calcular las sumas de cuadrados:

$$
\begin{aligned}
scT_\alpha &= 5\sum_{i=1}^{3}\widehat{\alpha}_i^2 = 5\times\left(1.33^2+0.47^2+0.87^2\right) = 13.734 \\
scT_\beta &= 3\sum_{j=1}^{2}\widehat{\beta}_j^2 = 3\times(0.07^2+0.26^2+0.6^2+0.26^2+1.07^2) = 4.935 \\
scG &= \sum_{ij}(y_{ij}-\overline{y}_{..})^2 = \sum_{ij}y_{ij}^2 - n\overline{y}_{..}^2 = \\
&= 388 - 15\times 4.933^2 = 22.983 \\
scR &= \sum_{ij}e_{ij}^2 = scG - scT_\alpha - scT_\beta \\
&= 22.983 - 13.734 - 4.935 = 4.314
\end{aligned}
$$

La estimación puntual de la varianza es:

$$\hat{\sigma}^2 = \hat{S}_R^2 = \frac{scR}{(I-1)(J-1)} = \frac{4.314}{8} = 0.539$$

P.2. Obtener la tabla ANOVA, realizar los correspondientes contrastes (con un nivel de significación del 5%) y calcular los porcentajes de variabilidad de las valoraciones explicadas por la diferencias entre interfaces y entre usuarios.

Solución:
A partir de las sumas de cuadrados obtenemos la tabla ANOVA:

F. var.	sc	gl	SCM	F	p-valor
Interfaz	13.734	2	6.867	12.74	$p < P(F_{2,8} > 11.04) = 0.005$
Usuario	4.935	4	1.234	2.29	$p > P(F_{4,8} > 3.84) = 0.05$
Residual	4.314	8	0.539		
Global	22.983	14	1.642		

Hay diferencias significativas entre las valoraciones medias de los tipos de interfaz pero no entre usuarios.

Los coeficientes de determinación son:

$$
\begin{aligned}
R^2\,(\text{"Interfaz"}) &= \frac{13.734}{22.983} = 0.598 \\
R^2\,(\text{"Usuario"}) &= \frac{4.935}{22.983} = 0.215 \\
R^2\,(\text{"Total"}) &= 1-\frac{4.314}{22.983} = 0.812
\end{aligned}
$$

Un 81.2% de la variabilidad de la respuesta es debida a diferencias entre interfaces y entre usuarios.

P.3. Contrastar si hay diferencias entre las valoraciones medias de las diferentes interfaces (con un nivel de significación del 5%). ¿Que ocurre en el caso de las valoraciones medias de los distintos usuarios?

Solución:
Se trata de realizar contrastes del tipo:

$$\begin{cases} H_0 : \theta_{ik} = \alpha_i - \alpha_k = \mu_{i.} - \mu_{k.} = 0 \\ H_1 : \theta_{ik} = \alpha_i - \alpha_k = \mu_{i.} - \mu_{k.} \neq 0 \end{cases}$$

La estimación puntual será:

$$\hat{\theta}_{ik} = \hat{\mu}_{i.} - \hat{\mu}_{k.} = \overline{y}_{i..} - \overline{y}_{k..}.$$

Como se trata de un diseño equilibrado en todos los casos el error estandar es el mismo:

$$\sigma\left(\hat{\theta}_{ik}\right) = \hat{S}_R\sqrt{\frac{2}{4}} = \sqrt{\frac{0.037}{2}} = 0.136.$$

Se rechaza la hipótesis nula si el valor observado del estadistico del contraste en valor absoluto $\left|\hat{d}_{ik}\right| = \left|\hat{\theta}_{ik}\right| / \sigma\left(\hat{\theta}_{ik}\right)$ es mayor que el correspondiente punto crítico. Si se utiliza el método DMS (LSD) el valor crítico sería:

$$t_{8,0.975} = 2.306.$$

- **Contraste 1** $\begin{cases} H_0^{12} : \mu_{1.} = \mu_{2.} \\ H_1^{12} : \mu_{1.} \neq \mu_{2.} \end{cases}$

$$\begin{aligned} \left|\hat{d}_{12}\right| &= \frac{|3.6 - 5.4|}{0.136} = 13.235 > 2.306 \\ &\Rightarrow \text{ Se rechaza } H_0^{12} \\ p - valor &= 2 \times P(t_8 > 13.235) < 2 \times 0.005 \end{aligned}$$

- **Contraste 2** $\begin{cases} H_0^{13} : \mu_{1.} = \mu_{3.} \\ H_1^{13} : \mu_{1.} \neq \mu_{3.} \end{cases}$

$$\begin{aligned} \left|\hat{d}_{13}\right| &= \frac{|3.6 - 5.8|}{0.136} = 16.176 > 2.306 \\ &\Rightarrow \text{ Se rechaza } H_0^{13} \\ p - valor &= 2 \times P(t_8 > 16.176) < 2 \times 0.005 \end{aligned}$$

- **Contraste 3** $\begin{cases} H_0^{23} : \mu_{2.} = \mu_{3.} \\ H_1^{23} : \mu_{2.} \neq \mu_{3.} \end{cases}$

$$\begin{aligned} \left|\hat{d}_{23}\right| &= \frac{|5.4 - 5.8|}{0.136} = 2.941 > 2.306 \\ &\Rightarrow \text{ Se rechaza } H_0^{23} \\ 2 \times 0.01 &< p - valor = 2 \times P(t_8 > 2.941) < 2 \times 0.005 \end{aligned}$$

NOTA: Si se utiliza el método de Scheffé el valor crítico sería:

$$\omega_S = \sqrt{(I-1)\, F_{I-1,IJ(K-1),(1-\alpha)}} = \sqrt{2 \times F_{2,8,0.95}} = \sqrt{2 \times 4.46} = 2.987.$$

Las conclusiones serían las mismas salvo que en el contraste 3 se aceptaría $\mu_{2.} = \mu_{3.}$.

En el caso de las valoraciones medias de los distintos usuarios no tiene sentido hacer las comparaciones múltiples porque ya aceptamos que no hay diferencias entre ellas.

P.4. Supongamos que sólo se considera el factor Interfaz y que los datos obtenidos se corresponden con tres modelos de interfaz seleccionados al azar (cada uno evaluado por 5 usuarios diferentes). Estudiar si la interfaz tiene un efecto significativo sobre la valoración (formular el modelo y las hipótesis del contraste) y obtener una estimación de la componente de la varianza debida al factor y de la varianza de la respuesta.

Solución:

Se trata de un diseño con un factor de efectos aleatorios:

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

con α_i *i.i.d.* $N\left(0, \sigma_\alpha^2\right)$ y ε_{ij} *i.i.d.* $N\left(0, \sigma^2\right)$ independientes. En este caso el contraste principal es:

$$\begin{cases} H_0 : \sigma_\alpha^2 = 0 \\ H_1 : \sigma_\alpha^2 \neq 0 \end{cases}$$

La tabla ANOVA correspondiente se obtendría agregando la SC y los GL del bloque Usuario al error:

F. var.	**sc**	**gl**	**SCM**	**F**	**p-valor**
Interfaz	13.734	2	6.867	8.91	$p = P\left(F_{2,12} > 8.91\right) \approx 0.005$
Residual	4.314	12	0.771		
Global	22.983	14	1.642		

Aceptamos claramente que $\sigma_\alpha^2 \neq 0$.
NOTA: $F_{2,10,0.995} = 9.43, F_{2,15,0.995} = 7.7$

Estimaciones puntuales:
Se trata de un diseño equilibrado por tanto:

$$\hat{\sigma}_\alpha^2 = \frac{\hat{S}_\alpha^2 - \hat{S}_R^2}{5} = \frac{6.867 - 0.771}{5} = 1.22,$$

es la estimación de la componente de la varianza debida al factor, y como $\hat{\sigma}^2 = \hat{S}_R^2$ entonces.

$$\hat{\sigma}_Y^2 = \hat{\sigma}^2 + \hat{\sigma}_\alpha^2 = 1.22 + 0.771 = 1.991.$$

Un $\frac{1.22}{1.991} \times 100 = 61.3\,\%$ de la variabilidad de la respuesta es debida a la variabilidad del factor.

ESTADISTICA II, Ingeniería Informática,
Problemas, 26 - Junio - 2.008

Problema 2:

En un estudio para determinar la relación en 24 países entre la variable de interés $Y =$ "Número de coches por cada 100 habitantes" y las variables explicativas $X_1 =$ "Ingresos per cápita en dólares" y $X_2 =$ "Precio de la gasolina en centavos por litro" se han obtenido los siguientes datos:

$$X^tX = \begin{pmatrix} 24 & 176.7 & 1136 \\ 176.7 & 1557.41 & 8440.8 \\ 1136 & 8440.8 & 57756 \end{pmatrix} \qquad X^tY = \begin{pmatrix} 654.4 \\ 5500.48 \\ 29765.4 \end{pmatrix}$$

$$\sum_{i=1}^{24} y_i = 654.4 \qquad \sum_{i=1}^{24} {y_i}^2 = 21268.96$$

$$\left(X^tX\right)^{-1} = \begin{pmatrix} 0.7669 & -2.5289 \times 10^{-2} & -1.1388 \times 10^{-2} \\ -2.5289 \times 10^{-2} & 3.9221 \times 10^{-3} & -7.5777 \times 10^{-5} \\ -1.1388 \times 10^{-2} & -7.5777 \times 10^{-5} & 2.5238 \times 10^{-4} \end{pmatrix}$$

Considerando el nivel de significación $\alpha = 0.05$, responde a las siguientes cuestiones:

P.5. Estima los coeficientes y obtén el modelo de regresión lineal.

Solución:

$$\widehat{\alpha} = \left(X^tX\right)^{-1} X^tY = \begin{pmatrix} 23.789 \\ 2.769 \\ -0.3569 \end{pmatrix}$$

$$\boxed{\widehat{y} = 23.789 + 2.769x_1 - 0.3569x_2}$$

P.6. Haz el contraste de regresión del modelo obtenido (tabla ANOVA) y calcula el coeficiente de determinación corregido.

Solución:

$$scR = \sum_{i=1}^{24} e_i^2 = Y^tY - \widehat{\alpha}^t X^tY = 21268.96 - 20173.33 = 1095.6302$$

$$scG = \sum_{i=1}^{24} \left(y_i - \bar{y}\right)^2 = \sum_{i=1}^{24} {y_i}^2 - 24\bar{y}^2 = 21268.96 - \frac{654.4^2}{24} = 3434.3776$$

Fuentes de variación	Suma de cuadrados	Grados de libertad	Varianzas
Modelo	2338.7473	2	$\widehat{S}_E^2 = \frac{scE}{2} = 1169.374$
Residual	1095.6302	21	$\widehat{S}_R^2 = \frac{scR}{21} = 52.173$
Global	3434.3776	23	$\widehat{S}_G^2 = \frac{scG}{23} = 149.321$

Contraste de regresión: $\begin{cases} H_0: & \text{El modelo no es significativo, i.e, } \alpha_1 = \alpha_2 = 0 \\ H_1: & \text{El modelo es significativo} \end{cases}$

Calculamos el estadístico asociado al contraste de regresión:

$$\widehat{F}_M = \frac{\widehat{S}_E^2}{\widehat{S}_R^2} = 22.413 \sim F_{2,21} \text{ (si } H_0 \text{ es cierta)}$$

$$p - valor = p(F_{2,21} \geq 22.413) \simeq 0 \Rightarrow \text{Rechazamos } H_0\text{, i.e., el modelo es significativo.}$$

El coeficiente de determinación corregido viene dado por:

$$\bar{R}^2 = 1 - \frac{\widehat{S}_R^2}{\widehat{S}_G^2} = 0.65.$$

P.7. Estudia si la información que disponemos de Estados Unidos ($y_i = 53$, $x_{1,i} = 9.7$, $x_{2,i} = 17$ y valor de influencia $h_{ii} = 0.306064$) nos indica que se trata de un dato atípico. Comprueba también si es influyente a priori.

Solución:

- No es un dato atípico, ya que el residuo estandarizado $r_i = \dfrac{e_i}{\widehat{S}_R\sqrt{1-h_{ii}}} = 1.4$, y, por tanto, $|r_i| < 2$.
- Es influyente a priori, ya que $h_{ii} = 0.306064 > 0.25 = 2\dfrac{3}{24} = 2\dfrac{k+1}{n}$.

P.8. Obtén el intervalo de predicción para el número de coches en un país en el que el ingreso per cápita es de 4 y el precio de la gasolina es de 40 centavos por litro (el valor de influencia en este caso es 0.061006).

Solución:
El intervalo de predicción para y_i viene dado por

$$\left[\widehat{y}_i - \widehat{S}_R\sqrt{1+h_{ii}}t_{21}(0.975), \widehat{y}_i + \widehat{S}_R\sqrt{1+h_{ii}}t_{21}(0.975)\right].$$

$$\widehat{y}_i = \widehat{\alpha}_0 + \widehat{\alpha}_1 4 + \widehat{\alpha}_2 40 = 20.587 \text{ y } t_{21}(0.975) = 2.0796.$$

Haciendo los cálculos correspondientes se obtiene que el intervalo de predicción es $[5.112, 36.06]$.

P.9. Supongamos ahora que añadimos las variables regresoras X_3 =“Toneladas de gasolina consumida por coche al año”, X_4 =“Densidad de población”, X_5 = “Miles de pasajeros-kilómetros por persona que usan bus o tren”y X_6 = “Población en millones de personas”. A la vista de la siguiente tabla, estudia qué variables son significativas en este nuevo modelo.

i	0	1	2	3	4	5	6
$\widehat{\alpha}_i$	46.973	3.224	-0.405	-10.76	-0.0107	-7.148	0.059
$\sigma(\widehat{\alpha}_i)$	9.094	0.3445	0.127	2.648	0.0109	1.748	0.027

Solución:

$\widehat{\alpha}_i$	46.973	3.224	-0.405	-10.76	-0.0107	-7.148	0.059
$\sigma(\widehat{\alpha}_i)$	9.094	0.3445	0.127	2.648	0.0109	1.748	0.027
$\widehat{t}_i = \dfrac{\widehat{\alpha}_i}{\sigma(\widehat{\alpha}_i)}$	5.191	9.358	-3.189	-4.063	-0.982	-4.089	2.185

Para cada coeficiente α_i consideramos el siguiente contraste de hipótesis: $\begin{cases} H_0 : \alpha_i = 0 \\ H_1 : \alpha_i \neq 0 \end{cases}$

Sabemos que $\widehat{t}_i \sim t_{17}$ (si H_0 es cierta). Sabemos además que $t_{17}(0.975) = 2.1098$. Por tanto:

- Para $i = 1, 2, 3, 5, 6, |\widehat{t}_i| > 2.1098 \Rightarrow p{-}valor = 2p(t_{17} > |\widehat{t}_i|) < 0.05 \Rightarrow$ Las variables son significativas.
- Para $i = 4, |\widehat{t}_i| < 2.1098 \Rightarrow p{-}valor = 2p(t_{17} > |\widehat{t}_i|) > 0.05 \Rightarrow$ La variable X_4 no es significativa.