ESTADISTICA II

INGENIERIA INFORMATICA, 3ER Curso

29 - Junio - 2.007 **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**	**D**	**D**	**A**	**D**	**D**
Pregunta	**7**	**8**	**9**	**10**	**11**	**12**
Respuesta	**C**	**C**	**B**	**D**	**C**	**C**

CUESTIONES

1. En un contraste unilateral al 90% el contraste d_2 es más potente que el d_1 entonces

- **A.** La probabilidad de error de tipo II de d_1 es mayor que la de d_2. **Solución**
- **B.** La región de rechazo de d_2 es menor que la de d_1.
- **C.** La probabilidad de error de tipo I de d_1 es mayor que la de d_2.
- **D.** Ninguna de las otras respuestas.

2. Se ha medido, durante 30 días, el número de altas en una web que ofrece un cierto servicio en Internet, con los siguientes resultados:

Altas	0	1	2	3	≥ 4
Días	10	13	5	1	1

Entonces, aplicando el test de chi-cuadrado (utilizar tres grupos) para comprobar si el número de altas sigue una distribución de Poisson con $\alpha = 0.05$:

A. $Q = 0.55$ y se acepta porque $P\left(\chi_2^2 \geq 0.55\right) > 0.1$

B. $Q = 0.51$ y se acepta que sigue una Poisson porque $P\left(\chi_3^2 \geq 0.51\right) > 0.05$.

C. $Q = 0.51$ y se acepta que sigue una Poisson porque $P\left(\chi_4^2 \geq 0.51\right) > 0.05$

D. $Q = 0.55$ y se acepta porque $P\left(\chi_1^2 \geq 0.55\right) > 0.05$. **Solución**

Se estima el parámetro λ por máxima verosimilitud:

$$\widehat{\lambda} = \frac{1}{\overline{x}} = 1$$

y las probabilidades asociadas a cada valor de la variable:

$$P\left(X = k\right) = e^{-\lambda}\frac{\lambda^k}{k!}$$

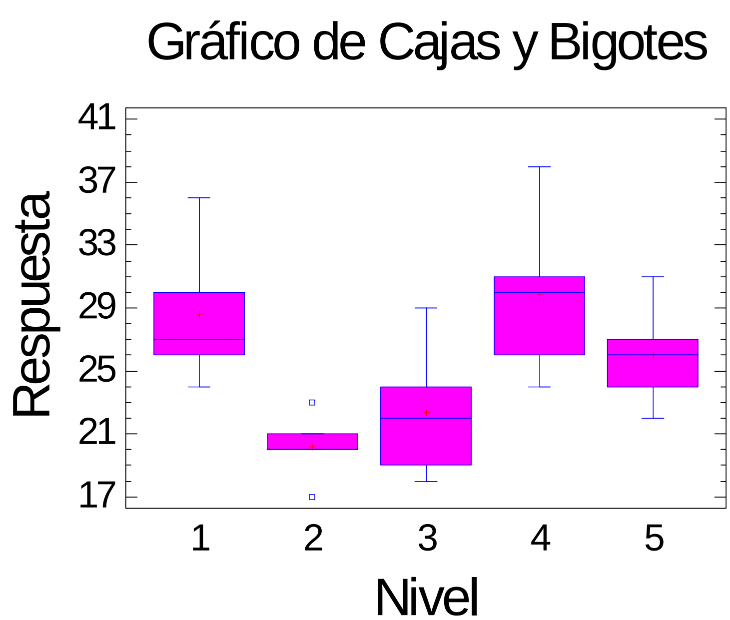
Ahora se construye la tabla de frecuencias observadas y esperadas:

x	O_i	p_i	$E_i = np_i$	$\frac{(O_i - E_i)^2}{E_i}$
0	10	0.3679	11.037	0.09744
1	13	0.3679	11.037	0.349
≥ 2	7	0.2642	7.9272	0.10845
	$n = 30$	1	30	$Q = 0.555$

El $p-$valor, según una chi-cuadrado con $k - r - 1 = 3 - 1 - 1 = 1$ grados de libertad es

$$p - valor = P\left(\chi_1^2 \geq 0.55\right) > 0.1 \Rightarrow \textit{ Se acepta } H_0.$$

3. En un diseño de experimentos de un factor con cinco niveles, se representa un box-plot de la respuesta para cada nivel del factor. De él se deduce:



A. Presencia clara de homocedasticidad.

B. El factor no es significativo.

C. Se observa una estructura de dependencia positiva en los errores.

D. La varianza de los residuos no es constante para cada nivel del factor. **Solución**

4. Con respecto al análisis de la varianza con un factor de efectos fijos o efectos aleatorios:

A. Al contrario que en el modelo de efectos fijos, en el modelo de efectos aleatorios las observaciones pertenecientes a un mismo tratamiento son correladas. **Solución**

B. A diferencia que en el modelo de efectos fijos, en el modelo de efectos aleatorios la media de la respuesta es cero.

C. En el modelo de efectos fijos el número de niveles es inferior al caso de efectos aleatorios.

D. En el modelo de efectos fijos, se busca conocer el efecto del factor sobre la variabilidad de la respuesta.

5. El test de Barlett se emplea en el análisis de la varianza para verificar

A. El supuesto de homogeneidad de varianzas. **Solución**

B. La normalidad de los residuos

C. Es la alternativa no paramétrica al ANOVA.

D. La hipótesis de independencia de las observaciones.

6. Los parámetros de un modelo de diseño de experimentos estimados por el método de máxima-verosimilitud coinciden con los obtenidos por el método de mínimos cuadrados:

A. Siempre.

B. Bajo la hipótesis de independencia.

C. Bajo la hipótesis de homocedasticidad.

D. Bajo la hipótesis de normalidad. **Solución**

7. En un modelo de regresión lineal múltiple, el estimador máximo verosímil de la varianza del modelo $\left(\hat{\sigma}^2_{MV}\right)$ verifica que

A. $Sesgo(\hat{\sigma}^2_{MV}) = 0 \quad \text{y} \quad \lim_{n\to\infty} var(\hat{\sigma}^2_{MV}) = 0$

B. $\lim_{n\to\infty} Sesgo(\hat{\sigma}^2_{MV}) = 0 \quad y \quad var(\hat{\sigma}^2_{MV}) = 0$

C. $\lim_{n\to\infty} Sesgo(\hat{\sigma}^2_{MV}) = 0 \quad y \quad \lim_{n\to\infty} var(\hat{\sigma}^2_{MV}) = 0$ **Solución**

D. Ninguna de las otras tres respuestas.

8. En el ajuste de un modelo de regresión lineal múltiple con cuatro regresoras se ha obtenido que el $FIV(x_4) = 8$. Por tanto

A. Como $FIV(x_4) < 10$ no hay indicios de multicolinealidad.

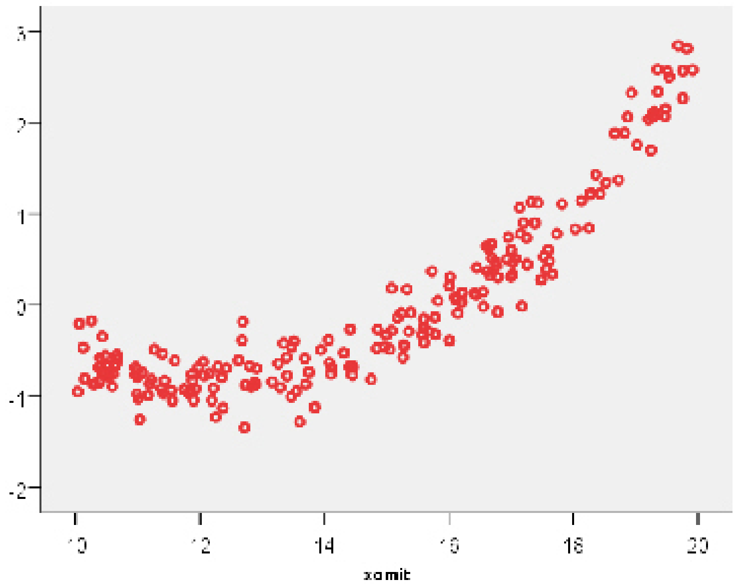
B. Como $FIV(x_4) < 10$ no hay observaciones influyentes a priori pero puede haber a posteriori.

C. El coeficiente de correlación múltiple de x_4 con las otras tres regresoras es $|r_{4.otras}| > 0.90$, existe multicolinealidad. **Solución**

D. Como $FIV(x_4)$ es grande se deben introducir en el modelo todas las regresoras.

$$\begin{aligned} FIV(x_4) &= 8 = \frac{1}{1 - r^2_{4.otras}} \Rightarrow 1 - r^2_{4.otras} = 0.125 \\ r^2_{4.otras} &= 0.875 \Rightarrow |r_{4.otras}| = 0.935\,4 > 0.90 \end{aligned}$$

9. Calculado el ajuste de un modelo de regresión lineal múltiple, en la gráfica adjunta se representan los residuos estandarizados frente a una regresora omitida (x_{omit}, esta variable no está en el modelo). A la vista de esta gráfica



A. La variable x_{omit} no se debe introducir en el modelo.

B. Se deben introducir en el modelo como regresoras las variables: x_{omit} y x_{omit}^2 **Solución**

C. Existe multicolinealidad de la variable x_{omit} con las otras regresoras.

D. Ninguna de las otras tres respuestas.

10. Se ha ajustado un modelo de regresión lineal múltiple con tres regresoras. Para unos valores de la regresora $\tilde{\mathbf{x}}_0$ se ha obtenido que el valor de influencia asociado es $h_0 = 0.05$.El intervalo de confianza al 90% para $E(Y/\tilde{\mathbf{x}}_0) = m_0$ es 10 ± 0.822. En base a estos datos se obtiene que el intervalo de predicción para $Y/\tilde{\mathbf{x}}_0$ viene dado por

A. Con esos datos no se puede calcular.

B. (10 ± 1.533)

C. (10 ± 8.852)

D. Ninguna de las otras tres respuestas. **Solución** (10 ± 3.767)

$$\begin{aligned}
\hat{s}_R \cdot \sqrt{h_0} \cdot t &= 0.822 \Rightarrow \hat{s}_R \cdot t = \frac{0.822}{\sqrt{0.05}} = 3.676\,1 \\
\hat{s}_R \cdot \sqrt{1 + h_0} \cdot t &= 3.676\,1 \cdot \sqrt{1.05} = 3.767
\end{aligned}$$

11. En el ajuste de un modelo de regresión lineal múltiple con cinco regresoras y cien observaciones se ha obtenido que el estadístico de Durbin-Watson es $\hat{d} = 3.4$. Se deduce que

A. El contraste no es concluyente y no se puede afirmar nada acerca de la dependencia o no de los residuos.

B. Los residuos son independientes.

C. Los residuos son dependientes con autocorrelación negativa. **Solución**

D. Los residuos son dependientes con autocorrelación positiva.

12. Al ajustar una recta de regresión de Y sobre X se ha obtenido

$$\hat{y} = 1 + 1.5x$$

Se conocen los siguientes datos muestrales: $s_X^2 = 4, \quad \bar{y} = 2.5, \quad s_Y^2 = 10.$

En base a ello la recta de regresión de X sobre Y es

A. $\hat{x} = -\frac{0.2}{3} + \frac{0.2}{3}x$

B. $\hat{x} = 2.5 + 3.3y$

C. $\hat{x} = -0.5 + 0.6y$ **Solución**

D. Ninguna de las otras tres respuestas.

$$\begin{aligned}
1.5 &= \frac{s_{XY}}{4} \Rightarrow s_{XY} = 6 \\
\hat{\beta}_1 &= \frac{s_{XY}}{10} = \frac{6}{10} = 0.6
\end{aligned}$$

$$\begin{aligned}
2.5 &= 1 + 1.5 \cdot \bar{x} \Rightarrow \bar{x} = 1 \\
\hat{\beta}_0 &= 1 - 0.6 \cdot 2.5 = -0.5
\end{aligned}$$

ESTADISTICA II, Ingeniería Informática,

Problemas, 29 - Junio - 2.007

Apellidos, Nombre: ....

D.N.I.: ....

Responder de forma concisa y justificada a las siguientes cuestiones.
Las respuestas se escribirán con bolígrafo a continuación de las preguntas.
Cada una de las preguntas tiene una valoración máxima de 0.5 puntos sobre cinco.
Para aprobar el examen es necesario obtener una puntuación igual o superior a 1 punto en cada uno de los dos problemas.

Problema 1. *Se desea comparar la duración de tres tipos de baterías recargables (A, B y C) en teléfonos móviles de similares características. Dado que los móviles que utilizan estas baterías no son exactamente iguales se han elegido cinco teléfonos y se ha observado la duración de cada una de las baterías para estudiar el efecto de las mismas en la variable "duración en horas". Los resultados son los siguientes:*

Teléfono	Batería A	Batería B	Batería C
T1	20	22	35
T2	35	42	40
T3	50	30	60
T4	40	35	50
T5	30	22	45

$\Sigma\Sigma y_{ij} = 556$
$\Sigma\Sigma y_{ij}^2 = 22432$

P.1. Formular el modelo de diseño de experimentos asociado a este problema. Detallar todas las suposiciones que se hacen en él. Calcular las estimaciones de los efectos del factor "batería".

Solución:
Se trata de un diseño en **bloques completamente aleatorizado** con un factor tratamiento (tipo de batería) con $I = 3$ niveles y un factor bloque (teléfono) con $J = 5$ niveles.

Se supone que las interacciones entre ambos factores son nulas. 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 $N\left(0, \sigma^2\right)$.

Se obtienen las siguientes estimaciones:

Teléfono	Batería A	Batería B	Batería C	$\overline{y}_{.j}$	$\widehat{\beta}_j$
T1	20	22	35	25.667	-11.4
T2	35	42	40	39	1.933
T3	50	30	60	46.667	9.6
T4	40	35	50	41.667	4.6
T5	30	22	45	32.333	-4.734
$\overline{y}_{i.}$	35	30.2	46	$\overline{y}_{..} = 37.067$	
$\widehat{\alpha}_i$	-2.067	-6.867	8.933		

P.2. Completar la tabla ANOVA e indicar qué efectos son significativos y cuáles no (nivel de significación 5%). Calcular los coeficientes de determinación.

Solución:
Las sumas de cuadrados son:

$$\begin{aligned}
VE(\alpha) &= \Sigma\Sigma\widehat{\alpha}_i^2 = 5 \times \left(2.067^2 + 6.867^2 + 8.933^2\right) = 656.13 \\
VE(\beta) &= \Sigma\Sigma\widehat{\beta}_j^2 = 3 \times \left(11.4^2 + 1.933^2 + 9.6^2 + 4.6^2 + 4.734^2\right) = 808.28 \\
VGlobal &= \Sigma\left(y_{ij} - \overline{y}_{...}\right)^2 = \Sigma y_{ijk}^2 - n\overline{y}_{..}^2 = 22432 - 15 \times 37.067^2 = 1822.93 \\
VR &= \Sigma e_{ij}^2 = VG - VE(\alpha) - VE(\beta) \\
&= 1822.93 - 656.13 - 808.28 = 358.52
\end{aligned}$$

Fuentes variación.	sc	gl	SCM	F	p-valor	¿significativo?
Factor Batería (α)	656.13	2	328.067	7.32	$p < P(F_{2,8} \geq 6.06) = 0.025$	**SÍ**
Bloque Teléfono (β)	808.28	4	202.067	4.51	$p < P(F_{4,8} \geq 3.84) = 0.05$	**SÍ**
Residual	358.52	8	44.8167			
Global	1822.93	14				

Los coeficientes de determinación son:

$$\begin{aligned}
R^2(\text{"Batería"}) &= \frac{656.13}{1822.93} = 0.36 \\
R^2(\text{"Teléfono"}) &= \frac{808.28}{1822.93} = 0.443 \\
R^2(\text{"Total"}) &= 1 - \frac{358.52}{1822.93} = 0.803
\end{aligned}$$

P.3. Calcular un intervalo de confianza al 90% para la diferencia de medias entre la duración de las baterías A y B. En base a dicho intervalo, ¿existen diferencias significativas entre ambas baterías? Realiza el contraste $H_0 : \mu_{batA} = \mu_{batB}$ frente a $H_1 : \mu_{batA} \neq \mu_{batB}$. Siendo μ_{batA} (respectivamente μ_{batB}) la duración media de la batería A (respectivamente batería B)). Explicar por qué el resultado obtenido es coherente con el resultado del intervalo de confianza.

Solución:
Para calcular intervalos de confianza para $\mu_{batA} - \mu_{batB}$ usamos el estadístico

$$\frac{\left(\overline{Y}_{1\cdot} - \overline{Y}_{2\cdot}\right) - \left(\mu_{batA} - \mu_{batB}\right)}{\widehat{s}_R\sqrt{\dfrac{2}{J}}} \in t_{(I-1)(J-1)}$$

de manera que

$$\begin{aligned}
IC_{90\%}\left(\mu_{batA} - \mu_{batB}\right) &= \left(\overline{y}_{1\cdot} - \overline{y}_{2\cdot} \pm t_{(I-1)(J-1),1-\frac{\alpha}{2}}\widehat{s}_R\sqrt{\frac{2}{J}}\right) \\
&= \left((35 - 30.2) \pm 1.8595\sqrt{44.8167}\sqrt{\frac{2}{5}}\right) \\
&= (4.8 \pm 7.873\,1) = (-3.073\,1, 12.673)
\end{aligned}$$

No hay diferencias significativas, al $\alpha = 0.1$ de significación, entre los efectos de las baterías A y B. Se acepta H_0 a este nivel.

Haciendo el contraste directamente, se obtiene como estadístico del contraste

$$\hat{d} = \frac{\overline{Y}_{1\cdot} - \overline{Y}_{2\cdot}}{\widehat{s}_R\sqrt{\dfrac{2}{J}}} = \frac{35 - 30.2}{\sqrt{44.8167}\sqrt{\dfrac{2}{5}}} = 0.2677$$

que bajo H_0 debe pertenecer a una $t_{(I-1)(J-1)} = t_8$, obteniendo un $p-valor > 2 \times 0.250 = 0.500$ y se acepta H_0 para todo $\alpha < 0.50$, en particular, para $\alpha = 0.1$

P.4. Obtener los grupos homogéneos, al 95%, para el factor Batería, según el método de Scheffé.

Solución:
Para estudiar los grupos homogéneos, realizamos los siguientes contrastes:

$$\textbf{Contraste 1}\left\{\begin{array}{l} H_0^{12} : \alpha_1 = \alpha_2 \\ H_1^{12} : \alpha_1 \neq \alpha_2 \end{array}\right.$$
$$\textbf{Contraste 2}\left\{\begin{array}{l} H_0^{13} : \alpha_1 = \alpha_3 \\ H_1^{13} : \alpha_1 \neq \alpha_3 \end{array}\right.$$
$$\textbf{Contraste 3}\left\{\begin{array}{l} H_0^{23} : \alpha_2 = \alpha_3 \\ H_1^{23} : \alpha_2 \neq \alpha_3 \end{array}\right.$$

El valor crítico es

$$\omega_S = \sqrt{(I-1)\,F_{I-1,(I-1)(J-1),\alpha}} = \sqrt{2 \times F_{2,8,0.05}} = \sqrt{2 \times 4.46} = 2.9866$$

$$\begin{aligned} &\sqrt{(I-1)\,F_{I-1,(I-1)(J-1),\alpha}} \times \widehat{s}_R \times \sqrt{\sum \frac{b_i^2}{n_i}} \\ = \quad & 2.9866 \times \sqrt{44.8167} \times \sqrt{\frac{1}{5}+\frac{1}{5}} = 12.645 \end{aligned}$$

- **Constraste 1**: $|\Sigma b_i \overline{y}_{i.}| = |\overline{y}_{1.} - \overline{y}_{2.}| = |35 - 30.2| = 4.8 < 12.645 \Rightarrow$ Se acepta H_0^{12}
- **Constraste 2**: $|\overline{y}_{1.} - \overline{y}_{3.}| = |35 - 46| = 11 < 12.645 \ \Rightarrow$ Se acepta H_0^{13}
- **Constraste 3**: $|\overline{y}_{2.} - \overline{y}_{3.}| = |30.2 - 46| = 15.8 > 12.645 \ \Rightarrow$ Se rechaza H_0^{23}

Los grupos homogéneos son:

- (B1,B2) y B3

o bien

- (B1,B3) y B2.

P.5. ¿Valió la pena fijar 5 teléfonos móviles diferentes y aplicar sobre cada uno los tres tipos de baterías?
Si los datos proviniesen de 15 teléfonos móviles diferentes ¿Sería significativo en este caso el factor Batería ?

Solución:
El nuevo modelo sería un ANOVA de un sólo factor de efectos fijos, para el que obtendríamos la siguiente tabla:

F. var.	sc	gl	SCM	F	p-valor	¿significativo?
Factor-Batería (α)	656.13	2	328.067	3.37	$p = P(F_{2,12} \geq 3.37) > 0.05$[1]	**NO**
Residual	1166.8	12	97.233			
Global	1822.93	14				

[1] No se tienen datos en las tablas para una $F_{2,12}$. Calculamos el p-valor a partir de la $F_{2,10}$ y la $F_{2,15}$:
$$\left\{ \begin{array}{l} P(F_{2,10} \geq 3.37) > P(F_{2,10} \geq 4.10) = 0.05 \\ P(F_{2,15} \geq 3.37) > P(F_{2,15} \geq 3.68) = 0.05 \end{array} \right. \Rightarrow P(F_{2,12} \geq 3.37) > 0.05$$

ESTADISTICA II, Ingeniería Informática,

Problemas, 29 - Junio - 2.007

Apellidos, Nombre: ....

D.N.I.: ....

Problema 2:

Una empresa de ventas por internet de productos informáticos está interesada en estudiar que variables influyen en sus costes mensuales (variable de interés). Para ello recogieron los costes de distribución (en miles de euros), las ventas (en cientos de miles de euros) y el número de órdenes de compras (en miles) de los últimos 24 meses. A partir de dicha muestra se obtuvieron los siguientes datos:

$$X^tX = \begin{pmatrix} 24 & 109.56 & 105.432 \\ 109.56 & 515.4306 & 492.3585 \\ 105.432 & 492.3585 & 475.6584 \end{pmatrix}, \quad X^tY = \begin{pmatrix} 171.029 \\ 801.1657 \\ 771.4703 \end{pmatrix}$$

$$Y^tY = 1257.24, \qquad \sum_{i=1}^{24} y_i = 171.029$$

$$\left(X^tX\right)^{-1} = \begin{pmatrix} 1.6692607 & -0.1230062 & -0.2426748 \\ -0.1230062 & 0.1819062 & -0.1610280 \\ -0.2426748 & -0.1610280 & 0.2225740 \end{pmatrix}$$

P.6. Obtener las estimaciones de los parámetros del modelo de regresión con dos regresoras (coeficientes y varianza del modelo)

Solución:

$$\begin{aligned} \hat{\alpha} &= \left(X^tX\right)^{-1} X^tY = (-0.2728247, 0.4711387, 1.1946926)^t \\ scR &= Y^tY - \hat{\alpha}^t X^tY = 1257.24 - 1252.469 = 4.771 \\ \hat{s}_R^2 &= \frac{4.771}{24-3} = 0.2272 \end{aligned}$$

P.7. Contrastar si existe una relación lineal significativa entre la respuesta y las regresoras (tabla ANOVA). En caso afirmativo calcular e interpretar el coeficiente de determinación corregido por grados de libertad.

Solución:

$$\begin{aligned} scG &= 1257.24 - \frac{171.029^2}{24} = 38.4517 \\ scE &= 38.4517 - 4.771 = 33.6807 \\ \hat{s}_Y^2 &= \frac{38.452}{23} = 1.67183 \end{aligned}$$

Fuente variación	**SC**	**gl**	**MSC**	**F**	**p-valor**
Regresión	$scE = 33.681$	2	$\hat{s}_E^2 = 16.8405$	$\hat{F} = 74.122$	$<< 0.005$
Residual	$scR = 4.771$	21	$\hat{s}_R^2 = 0.2272$		
Total	$scG = 38.452$	23	$\hat{s}_Y^2 = 1.67183$		

El contraste de regresión es significativo y el modelo de regresión es significativo

$$\begin{aligned} R^2 &= \frac{33.681}{38.452} = 0.875923 \\ \bar{R}^2 &= 1 - \frac{0.2272}{1.67183} = 0.864101 \Rightarrow \textit{Buen ajuste ...} \end{aligned}$$

P.8. Estudiar si la primera observación $(y_1 = 5.3,\ x_{1,1} = 3.86,\ x_{1,2} = 4.02)$ **es influyente a priori. Obtener el correspondiente intervalo de predicción al 95%.**

Solución:

$$\begin{aligned} h_{11} &= x_1^t \left(X^t X\right)^{-1} x_1 = 0.0780562 << 2 \times \frac{3}{24} = 0.25 \\ &\Rightarrow \textit{No es influyente a priori} \end{aligned}$$

$$\begin{aligned} \hat{y}_1 &= 6.342462 \\ \sigma^2(\hat{y}_1) &= \hat{s}_R^2 (1 + h_{11}) = 0.2272 \times 1.0780562 = 0.24493 \\ \sigma(\hat{y}_1) &= 0.4949 \\ t_{21,0.975} &= 2.0796 \end{aligned}$$

$$\begin{aligned} \textit{Int. Pred.} &= (6.342462 \pm 2.0796 \times 0.4949) = \\ &= (6.342462 \pm 1.0292) = (\mathbf{5.3133},\ \mathbf{7.3717}) \end{aligned}$$

P.9. Contrastar, considerando un nivel de significación del 5%, si las variables regresoras del modelo son significativas (contrastes individuales). Calcular la $Cov(\hat{\alpha}_1, \hat{\alpha}_2)$ ¿Son ortogonales estas dos variables?

Solución:

$$\begin{aligned}
Var(\hat{\boldsymbol{\alpha}}) &= \hat{s}_R^2 \cdot \left(X^t X\right)^{-1} \Rightarrow \\
Var(\hat{\alpha}_1) &= \hat{s}_R^2 \cdot q_{11} = 0.2272 \cdot 0.1819062 = 4.1329 \times 10^{-2} \\
Var(\hat{\alpha}_2) &= \hat{s}_R^2 \cdot q_{22} = 0.2272 \cdot 0.2225740 = 5.0569 \times 10^{-2} \\
Cov(\hat{\alpha}_1, \hat{\alpha}_2) &= \hat{s}_R^2 \cdot q_{12} = 0.2272 \cdot (-0.1610280) = -3.6586 \times 10^{-2}
\end{aligned}$$

Dado que $P(t_{18} > 2.10) = 0.025$ se obtiene finalmente

Parám (α_i)	Est ($\hat{\alpha}_i$)	$\sigma(\hat{\alpha}_i)$	$\hat{t}_i = \frac{\hat{\alpha}_i}{\sigma(\hat{\alpha}_i)}$	**¿es significativa** X_i?
α_1	0.4711387	0.20330	$\hat{t}_1 = \frac{0.4711387}{0.20330} = 2.31751$	**SI**
α_2	1.1946926	0.22488	$\hat{t}_2 = \frac{1.1946926}{0.22488} = 5.31269$	**SI**

Las regresoras x_1 y x_2 *no son ortogonales* ya que en ese caso $Cov(\hat{\alpha}_1, \hat{\alpha}_2) = 0$. Además

$$r(\hat{\alpha}_1, \hat{\alpha}_2) = \frac{-0.1610280}{\sqrt{0.1819062 \cdot 0.2225740}} = -0.80028 \text{ es claramente significativa}$$

P.10. Calcular la correlación parcial y simple entre las variables Y y X_1.

Solución:

Coeficiente de correlación simple entre las variables $\mathbf{Y}$ *y* $\mathbf{X}_1$

$$r_{Y1} = \frac{s_{Y1}}{s_Y \cdot s_1} = \frac{0.85076}{\sqrt{1.67183} \cdot \sqrt{0.63705}} = \mathbf{0.82437}$$

$$\begin{aligned}
s_{Y1} &= \frac{801.1657}{24} - \frac{171.029}{24} \cdot \frac{109.56}{24} = 0.85076 \\
s_1^2 &= \frac{515.4306}{24} - \left(\frac{109.56}{24}\right)^2 = 0.63705
\end{aligned}$$

Coeficiente de correlación parcial entre las variables $\mathbf{Y}$ *y* $\mathbf{X}_1$

$$\begin{aligned}
r_{y1.2}^2 &= \frac{\hat{t}_1^2}{\hat{t}_1^2 + n - (k+1)} \\
&= \frac{2.31751^2}{2.31751^2 + 21} = 0.20367 \\
&\Rightarrow \; r_{y1.2} = \sqrt{0.20367} = \mathbf{0.4513}
\end{aligned}$$