

**Grado en Ingeniería Informática**

**Estadística**

**SEGUNDO PARCIAL**

4 de Junio de 2012

Apellidos, nombre	
Grupo:	Firma:

**Instrucciones**

1. **Rellenar** la cabecera del examen: **nombre, grupo y firma**.
2. Responder a cada pregunta en la hoja correspondiente.
3. Justificar todas las respuestas.
4. No se permiten anotaciones personales en el formulario. Sobre la mesa sólo se permite el DNI, calculadora, útiles de escritura, las tablas y el formulario.
5. **No desgrapar** las hojas.
6. Las **preguntas NO puntúan lo mismo**.
7. Se debe **firmar** en las hojas que hay en la mesa del profesor **al entregar el examen**. Esta firma es el justificante de la entrega del mismo.
8. Tiempo disponible: **2 horas**

**1. (10 puntos)** Una empresa tiene una página web a la que la gente accede para informarse de los nuevos productos que ofrece. Esta empresa ha lanzado un nuevo producto y se desea conocer el interés despertado en los usuarios debido al lanzamiento de dicho producto. Para ello se ha medido el número medio diario de accesos a dicha página web en los 10 días posteriores al lanzamiento del producto, obteniéndose los siguientes resultados:

**Frecuencia = 10**

**Media = 145,0**

**Varianza = 1072,22**

**Asimetría tipificada = 1,14907**

**Curtosis tipificada = 0,0464325**

A la vista de estos resultados responde a las siguientes preguntas:

**a) ¿Es admisible que nuestra muestra proceda de una distribución normal? (2 ptos).**

Sí puesto que nuestro coeficiente de asimetría tipificada y curtosis tipificada están ambos entre -2 y 2.

**b) Con un riesgo de 1ª especie  $\alpha=0,05$  ¿Es admisible una media poblacional en el número medio de accesos diarios de 120 accesos? (4 puntos).**

95,0% intervalo de confianza para la media: 145,0 +/- 23,4243 [121,576;168,424]

contraste t

-----

Hipótesis nula: media = 120,0

Alternativa: no igual

Estadístico t = 2,41434

P-valor = 0,0389713

Se rechaza la hipótesis nula para  $\alpha = 0,05$ .

$$t_{\text{tabla}} = t_{N-1}^{\alpha/2} = t_{10-1}^{0,05/2} = t_9^{0,025} = 2,262$$

$[t_{\text{calculada}}=2,41434] > [t_{\text{tabla}}=2,262] \rightarrow \text{Se Rechaza } H_0$

**c) Con un riesgo de 1ª especie  $\alpha=0,05$  ¿Es admisible una desviación típica de 25 en el número medio de accesos en la población? (4 puntos).**

$$\left[ \sqrt{\frac{(N-1)S^2}{g_2}}, \sqrt{\frac{(N-1)S^2}{g_1}} \right] \quad g_1 / P(\chi_{10-1}^2 > g_1) = 1 - 0,025 = 0,975 \quad \text{y} \quad g_2 / P(\chi_{10}^2 > g_2) = 0,025$$

$g_1 = 19,023$  y  $g_2 = 2,7$

95,0% intervalo de confianza para la desviación típica: [22,523;59,7792]

Como 25 está dentro del intervalo de confianza, admitimos los 25 accesos como desviación típica poblacional.

**2. (10 puntos)** Se desea estudiar la influencia que la configuración (dos posibles A y B) tiene sobre la media y la dispersión del rendimiento de un sistema informático. Para ello se ha diseñado un experimento en el que cada prueba consiste en la ejecución de un programa *test* bajo cada una de las configuraciones. La configuración A se ensayó 8 veces ( $N_1$ ), la B a su vez 10 veces ( $N_2$ ) y el rendimiento se midió mediante el parámetro *Tiempo de Reacción del Sistema* (tiempo desde que el usuario presiona “enter” hasta que se empieza a dar el servicio solicitado). Tras realizar las pruebas y recoger los datos, se obtuvieron los siguientes resultados:

Configuración A    media= 6,61    varianza= 4,45

Configuración B    media= 12,43    varianza=14,47

Indica, justificando detalladamente las respuestas, cuáles de las siguientes afirmaciones son VERDADERAS:

**a)** No hay diferencia significativa entre los rendimientos medios con un riesgo de primera especie  $\alpha=0,01$ . **(2,5 puntos) Falsa**

El intervalo de confianza para la diferencia de medias con  $\alpha=1\%$  es  $[-10,219, -1,421]$ , que no contiene el cero, por lo que hay diferencia significativa entre las medias.

**b)** Con un nivel de confianza del 95% hay diferencias significativas entre los rendimientos medios. **(2,5 puntos) Verdadera**

Con  $\alpha=0,05$  el intervalo para la diferencia de medias resulta  $[-9,013, -2,627]$  que no contiene el cero por lo que las medias difieren significativamente.

**c)** El p-value para la comparación de medias es menor que 0,01. **(2,5 puntos) Verdadera** ver justificación del apartado a.

**d)** Las desviaciones típicas para los dos tipos de procesadores difieren significativamente considerando un riesgo de primera especie  $\alpha=5\%$ . **(2,5 puntos) Falsa**

El intervalo del ratio de varianzas con  $\alpha=5\%$  resulta  $[0,073, 1,46]$  que contiene el valor uno por lo que las varianzas no difieren significativamente.

**3. (10 puntos)** Se desea estudiar el efecto que tienen el procesador y la carga de trabajo sobre el tiempo de utilización de la CPU, en la ejecución de cierto tipo de procedimientos. Para ello se han probado tres procesadores (A, B y C), combinados con tres cargas (10, 20 y 30). Cada tratamiento se repitió dos veces, y con los resultados se obtuvo la siguiente tabla del ANOVA, que está incompleta:

Analysis of Variance for T_CPU					
Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
MAIN EFFECTS					
A:PROCESADOR	515,111				
B:CARGA	85,7778				
INTERACTIONS					
AB	96,8889				
RESIDUAL	44,5				
TOTAL (CORRECTED)					

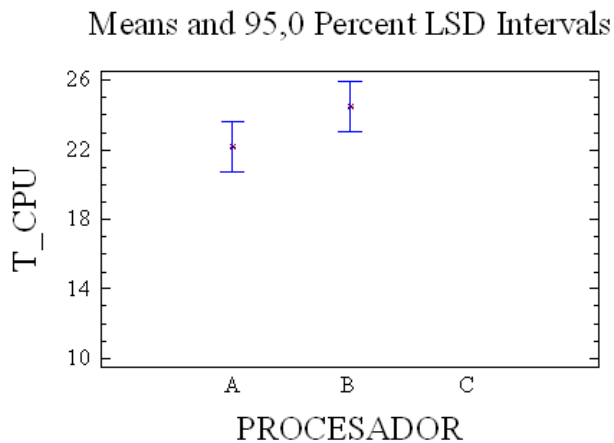
**a) Completa la tabla del ANOVA e indica qué efectos son significativos, utilizando un nivel de significación del 5%. Justifica la respuesta y los cálculos realizados. (4 puntos)**

Analysis of Variance for T_CPU					
Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
MAIN EFFECTS					
A:PROCESADOR	515,111	2	257,556	52,09	<0,05
B:CARGA	85,7778	2	42,8889	8,67	<0,05
INTERACTIONS					
AB	96,8889	4	24,2222	4,9	<0,05
RESIDUAL	44,5	9	4,94444		
TOTAL (CORRECTED)					
	742,278	17			

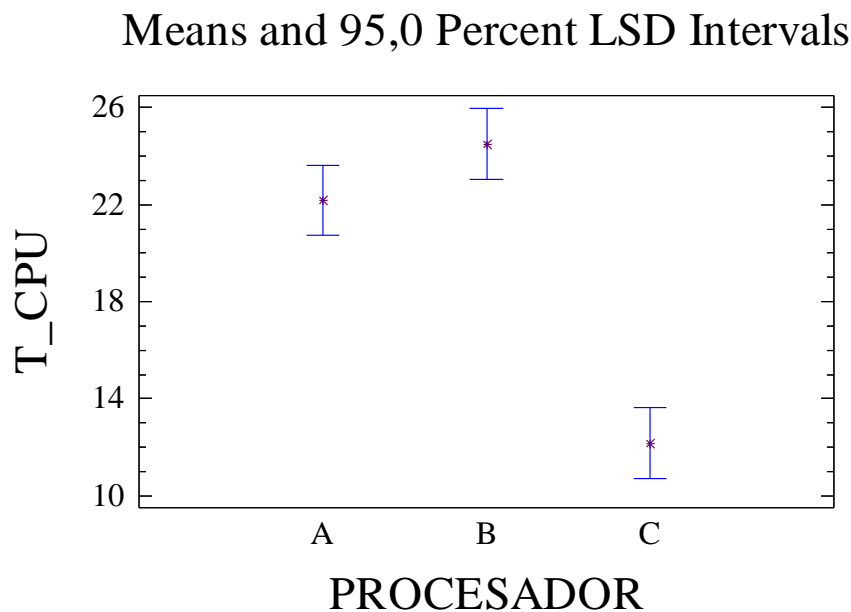
La  $SC_{total}$  es la suma de todas las demás SC. Los grados de libertad totales son  $18-1=17$ . Los grados de libertad del efecto de procesador son  $3-1=2$ . Los grados de libertad de carga son  $3-1=2$ . Los grados de libertad de la interacción entre estos dos factores  $2 \times 2=4$ . Los grados de libertad residuales serán  $17-2-2-4=9$ . Los cuadrados medios se obtienen dividiendo las sumas de cuadrados por los grados de libertad. Las F-ratio se calculan dividiendo el cuadrado medio de cada efecto por el cuadrado medio residual.

La F tablas para los efectos de carga y procesador vale 4,26, Como las dos F-ratios son mayores que la F de tabla los dos efectos son significativos. La F de tabla para la interacción vale 3,63, menor que la F-ratio de este efecto por lo que el dicha interacción es también significativa.

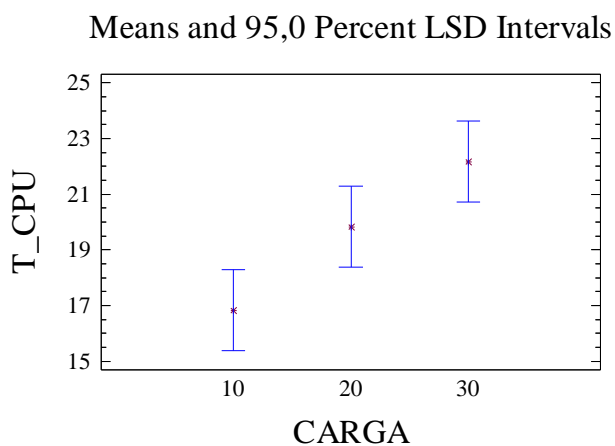
b) Dado el siguiente gráfico, dibuja **aproximadamente** el intervalo LSD para el procesador C de acuerdo al resultado obtenido en la tabla del ANOVA y al gráfico del apartado d). (2 puntos)



Con C habrá diferencia significativa respecto a A y B, y dará menor media. El gráfico resultante

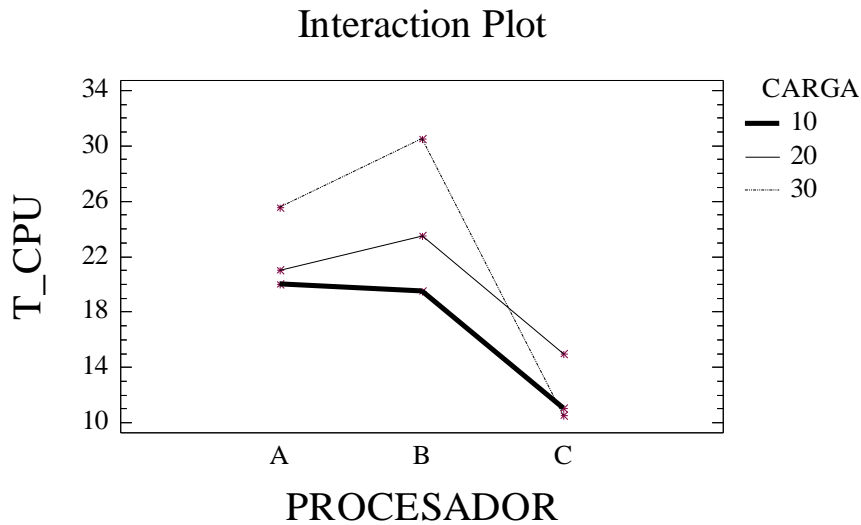


c) A partir del siguiente gráfico, indica a nivel descriptivo la naturaleza del efecto de la carga utilizada. (2 puntos)



Naturaleza lineal positiva.

**d)** A partir del siguiente gráfico, indica a nivel descriptivo la naturaleza del efecto de la carga utilizada. (2 puntos)

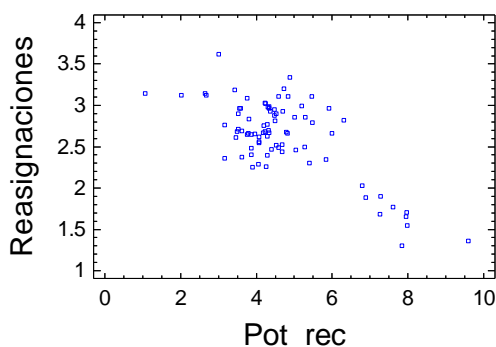


La interacción ha salido significativa, y vemos por ejemplo que la carga 30 pasa de tener los valores más elevados de T\_CPU con los procesadores A y B al más bajo con el procesador C. Con las otras cargas pasa algo parecido pero las diferencias entre procesadores son distintas. Con carga 20 la diferencia entre los procesadores A y B es menor que con carga 30, y esta diferencia es aún menor con carga 10. La mayor diferencia entre procesadores B y C se da con carga 30. Con carga 10 y 20 la diferencia entre procesadores B y C es similar.

El tratamiento óptimo para minimizar T\_CPU es Procesador C y Carga 30.

**4. (5 puntos)** Una compañía de telefonía móvil ha venido detectando problemas de funcionamiento en una estación concreta de su red. Para tratar de resolverlos ha estudiado para esta estación un indicador de calidad de la llamada: el número de veces que se reasigna el canal de transmisión una vez la llamada está en curso. Se analiza si se trata de un problema de cobertura y para ello se han recopilado los niveles de potencia recibidos (en mW) por los terminales móviles servidos por esta estación cada vez que inician una llamada. Con el objeto de analizar la posible relación entre ambas variables, en el informe correspondiente se recopilan sus datos medios diarios para todas las llamadas registradas en dicha estación. De ese informe se ha extraído la siguiente información:

Gráfico de Reasignaciones frente a Pot\_rec



Resumen Estadístico

	Pot_rec	Reasignaciones
-----		
Frecuencia	82	82
Media	4.65893	2.63361
Mediana	4.332	2.67762
Desv. típica	1.43172	0.450234
Mínimo	1.072	1.30222
Máximo	9.592	3.62142
Rango	8.52	2.3192
Primer cuartil	3.864	2.46312
Tercer cuartil	5.032	2.96102
Rango interc.	1.168	0.4979
-----		
Coef. correl. lineal rxy	= -0.7229	

A partir de estos datos, se pide:

**a)** Describir la naturaleza de la relación entre las variables aleatorias estudiadas. **(1,5 puntos)**

La relación es lineal negativa intermedia

**b)** Calcular la matriz de varianzas y covarianzas entre ambas variables. **(3,5 puntos)**

Matriz de varianzas y covarianzas:

	Pot_rec	Reasignaciones
-----		
Pot_rec	2.04982	-0.466009
Reasignaciones	-0.466009	0.202711
-----		

**5. (10 puntos)** Un informático que trabaja en una empresa dedicada al montaje y venta de ordenadores estándar de sobremesa necesita implementar un modelo para predecir el tiempo de entrega de los pedidos de ordenadores nuevos. Se entiende como tiempo de entrega el número de días transcurridos entre el pedido de un ordenador nuevo y la entrega real del mismo. El informático piensa que entre el número de extras pedidos respecto de la configuración básica del equipo pedido y el tiempo de entrega del mismo puede haber una relación lineal. Para formular el modelo selecciona aleatoriamente una muestra de 16 pedidos de los que recoge la información relativa al número de extras y el tiempo de entrega de cada uno.

Tras analizar los datos mediante un modelo de regresión lineal, obtiene los resultados que se presentan a continuación:

**Regression Analysis - Linear model:  $Y = a + b \cdot X$**

Dependent variable: tiempo

Independent variable: n. de extras

Parameter	Estimate	Standard Error	T Statistic	P-Value
Intercept		1,59084	13,7823	0,0000
Slope	2,06871	0,116411	17,7707	

**Analysis of Variance**

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	2927,23	1	2927,23	315,80	0,0000
Residual	129,77	14	9,26932		
Total (Corr.)	3057,0	15			

Correlation Coefficient = 0,978545

R-squared = percent

Standard Error of Est. =

**a)** Plantea y estima el modelo de regresión del *tiempo de entrega* en función del *número de extras*, indicando qué parámetros son significativos y por qué. Utiliza  $\alpha=5\%$ . **(3 puntos)**

Variable dependiente  $Y$  = tiempo de entrega (TE)

Variable independiente  $X$  = N° de extras (NE)

$$E(TE/NE) = \beta_0 + \beta_1 NE$$

$$\text{Estimación de } \beta_0 : t_{\text{statistic}} \times S_{b_0} = 13,7823 \times 1,59084 = 21,9254$$

$$\text{Estimación de } \beta_1 : 2,06871$$

$$\text{Modelo: } E(TE/NE) = 21,9254 + 2,06871xNE$$

Significación de la estimación de  $\beta_0$  :  $p\text{-value} < 0,05 \rightarrow$  **significativo**

Significación de la estimación de  $\beta_1$  :  $[t \text{ calculada}=17,7707] > [t_{14}=0,145] \rightarrow$  **significativo**



**b)** ¿Es el modelo planteado significativo globalmente?. Utiliza un Riesgo de 1ª especie  $\alpha$  de 0,01. **(2 puntos)**

Para ello miramos el resultado del **ANOVA**, y como  $p\text{-value} < 0,01$ , podemos decir que sí, como cabía esperar al haber resultado los dos parámetros significativos en el *test t*.

**c)** ¿Qué porcentaje de la variabilidad del tiempo de entrega viene explicada por el número de extras? Indica el parámetro que cuantifica dicho porcentaje. **(2 puntos)**

El  $R^2$  es el parámetro que indica el porcentaje de la variabilidad del tiempo de entrega explicada por el número de extras. Se puede calcular en este caso elevando al cuadrado el coeficientes de correlación, lo que da  $(0,978545)^2 \times 100 = 95,75\%$ . También se puede calcular dividiendo la suma de cuadrados del modelo por la total.

**d)** Si se recibe un pedido de un equipo con 16 extras, ¿cuántos días, en promedio, predice el modelo para la entrega? **(1 puntos)**

$$(TE/NE=16) = 21,9254 + 2,06871 \times 16 = 55,02 \rightarrow \mathbf{56 \text{ días mínimo}}$$

**e)** ¿Entre que valores aproximadamente estará el tiempo de entrega, en promedio, en el 95% de los pedidos en los que se solicitan 16 extras? **(2 puntos)**

$$(TE/NE=16) \sim N(m=55,02; \sigma=\sigma_{\text{Residual}})$$

$$\sigma_{\text{Residual}} = \text{Raíz}(\text{CMR}) = \text{Raíz}(9,26932) = 3,04456 \quad (6)$$

$$95\% \text{ de los datos} \in [m - 2\sigma, m + 2\sigma] \rightarrow [55,02 - 2 \times 3,04456, 55,02 + 2 \times 3,04456] \rightarrow \mathbf{[48,93, 61,12]}$$