

EXÁMENES DE MÉTODOS ESTADÍSTICOS I DE CURSOS ANTERIORES

EXAMEN DE JULIO DE 2009, CON SOLUCIÓN	1
EXAMEN DE ENERO DE 2009, CON SOLUCIÓN	13
EXAMEN DE JULIO 2008, CON SOLUCIÓN	31
EXAMEN DE ENERO 2008, CON SOLUCIÓN	44
EXAMEN DE JULIO 2007, CON SOLUCIÓN	59
EXAMEN DE ENERO 2007, CON SOLUCIÓN	70
EXAMEN DE JUNIO 2006, CON SOLUCIÓN	84
EXAMEN DE FEBRERO 2006, CON SOLUCIÓN.....	97

EXAMEN DE JULIO DE 2009, CON SOLUCIÓN

APELLIDOS, NOMBRE: _____

Firma: _____

GRUPO: _____

BLOQUE I. CUESTIONES (75 MINUTOS)

Todas las cuestiones deben ser debidamente justificadas. Aquellas cuestiones contestadas correctamente pero no justificadas no serán puntuadas. Cada cuestión son 10 puntos. La nota del bloque es la media de las notas de las cinco cuestiones.

1.- Para aceptar una partida de bobinas de tela para confección se examinan 4 metros lineales cortados en una bobina elegida al azar y se determina el número de defectos que hay. Si dicho número es menor de 3, se acepta la partida, en caso contrario se rechaza. Se pide:

- a) Calcular la probabilidad de aceptar, con este procedimiento, una partida que contenga por término medio 1,5 defectos por metro lineal.
- b) ¿Cuántos metros lineales se deberán examinar para aumentar dicha probabilidad al 10%?

2.- Un fabricante de coches compra motores a otra compañía. El fabricante recibe lotes de 40 motores y su plan de control de calidad consiste en seleccionar 8 de manera aleatoria y someterlos a prueba. Acepta el lote si ninguno de ellos presenta defectos, y de lo contrario lo rechaza.

- a) Si el lote contiene 2 motores defectuosos, ¿con qué probabilidad puede ser aceptado?
- b) ¿Cuál sería la probabilidad de aceptar el lote si éste fuera de 2000 motores y el proveedor asumiera un 5% de defectuosos?

3.- Queremos verificar si el tiempo de espera en una consulta de un centro médico se ha reducido por debajo de los 30 minutos habituales después de un cambio organizativo en el servicio. Para ello seleccionamos 10 pacientes al azar y verificamos su tiempo de espera, obteniendo los siguientes valores: $\bar{x} = 26,5$ minutos, $s_{n-1}^2 = 10$. Plantear el test que debemos contrastar y determinar si el cambio organizativo ha sido efectivo, utilizando un nivel de significación $\alpha = 0,10$.

4.- El gasto mensual en el uso del móvil de un alumno es una variable aleatoria con distribución Normal de media 20 euros y desviación típica 5. Calcular la “provisión” de fondos que tendrá que hacer el alumno para sufragar los gastos del tercer trimestre del año con una probabilidad superior al 85%.

5.- Dos cajas contienen piezas del proveedor A y piezas del proveedor B. Supongamos que una caja tiene 70 piezas de A y 30 de B, y que la otra caja tiene 10 de A y 25 de B. Seleccionamos una caja al azar y extraemos una pieza de la misma. Calcular la probabilidad de que la pieza sea del proveedor A.

APELLIDOS, NOMBRE: _____

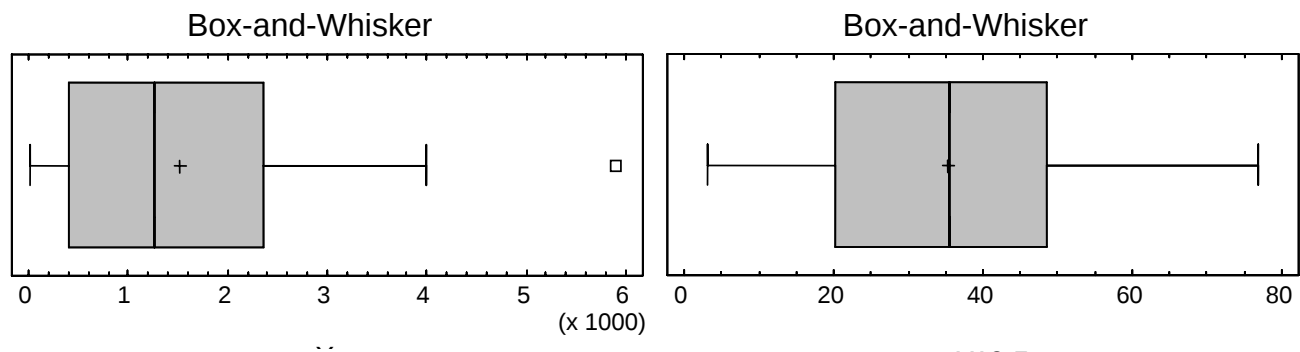
Firma: _____

Profesor de prácticas: _____

BLOQUE II. PRÁCTICAS DE ORDENADOR (45 MINUTOS)

Todas las cuestiones deben ser debidamente justificadas. Cada problema son 10 puntos. La nota del bloque es la media de las notas de los tres problemas.

1.- En una empresa que elabora alimentos congelados, uno de los parámetros de calidad es el número de microorganismos por gramo de producto (variable X). La empresa ha analizado dicho parámetro en un total de 50 partidas, y con los 50 datos resultantes se construye un diagrama Box-Whisker (gráfico inferior, izquierda). El gráfico de la derecha se ha construido con la raíz cuadrada de los datos. A la vista de los diagramas, responder a las siguientes preguntas justificando convenientemente las respuestas.



- La empresa está interesada en estudiar la pauta de distribución de la variable X . Para ello, ¿cree que hay valores de X anormalmente grandes que deberían eliminarse del estudio?
- ¿Cuál de las dos distribuciones, X o bien $\sqrt{X}$, tiene un menor coeficiente de asimetría?
- Se considera que la calidad de una partida es inadecuada si $X > 2400$. Calcular el porcentaje de partidas de mala calidad que se fabrican en la empresa.
- Se desea realizar un contraste de hipótesis para determinar si puede considerarse que el número medio de microorganismos por gramo de producto, en promedio, es de 1600. Para ello, ¿cuál de las siguientes respuestas es verdadera? ¿Por qué?

1.- Habría que trabajar con los datos de X y la hipótesis nula a plantear sería $H_0: m = 1600$

2.- Sería preferible trabajar con $\sqrt{X}$ dado que ésta tiende a distribuirse normalmente.

3.- Los apartados 1.) y 2.) son equivalentes, daría igual trabajar con X o con $\sqrt{X}$.

4.- Ninguna de las anteriores, necesitaríamos conocer σ para estimar la media poblacional.

2.- Una empresa de bebidas elabora refrescos en envases de 300 ml. La máquina que adiciona la bebida está calibrada para que el valor medio poblacional del volumen realmente adicionado sea 310 ml. Para saber si la máquina funciona correctamente o si por el contrario requiere ser ajustada, se toman 100 refrescos al azar y se determina el volumen de líquido que contienen. Los resultados, analizados con Statgraphics, son los siguientes:

Hypothesis Tests for volumen

Sample mean = 311,062

Sample median = 310,776

t-test

Null hypothesis: mean = 310,0

Alternative: not equal

Computed t statistic = 2,52085

P-Value = 0,0000000

Confidence Intervals for volumen

95,0% confidence interval for mean: 311,062 +/- 0,836233 [310,226;311,899]

95,0% confidence interval for standard deviation: [3,70029;4,89578]

a) ¿Qué se concluye del estudio, considerando un nivel de significación del 5%?

b) Justificar numéricamente el valor 2,52085 del estadístico de contraste, teniendo en cuenta que la cuasidesviación típica de los 100 datos vale 4,214.

c) Calcular el p-valor.

d) ¿Cuánto valdría el p-valor si la hipótesis alternativa fuese $H_1: m > 310$?

APELLIDOS, NOMBRE: _____**Firma:** _____**GRUPO:** _____**BLOQUE III. PROBLEMAS (90 MINUTOS)**

Cada problema vale 10 puntos. La nota del bloque es la media de las notas de los tres problemas.

1.- La proporción de cemento, X, que se añade al hormigón y el tiempo de secado, Y (horas), se sabe que siguen una distribución Normal bidimensional de vector de medias y matriz de varianzas-covarianzas,

$$\bar{\mu} = \begin{Bmatrix} 0,3 \\ 9 \end{Bmatrix} \quad \mathbf{V} = \begin{bmatrix} 0,05 & 0,25 \\ 0,25 & 2 \end{bmatrix}$$

- a) Si la proporción añadida de cemento es de 0,25, ¿cuál es la probabilidad de que el tiempo de secado dure más de 10 horas? (3 puntos)
- b) ¿Qué proporción de cemento debemos añadir para que el tiempo de secado sea inferior a 10 horas con una probabilidad del 97%? (4 puntos)
- c) Si el tiempo de secado fuese de 10 horas, ¿cuál es la probabilidad de que la proporción de cemento sea menor de 0,6 ? (3 puntos)

2.- El tiempo de espera en el andén de “Los Naranjos” para coger el tranvía de la línea 4 podemos admitir que es una variable X con distribución Normal de desviación típica 2 minutos. Un alumno estaría dispuesto a coger el tranvía si el tiempo medio de espera real fuera menor o igual a 7 minutos. En caso contrario prefiere coger el coche. Para tomar una decisión hace la siguiente prueba: determina el tiempo de espera en 9 viajes al azar y si el promedio del tiempo de espera es inferior a 8 minutos coge el tranvía y en caso contrario coge el coche.

- a) ¿Cuál es el contraste de hipótesis que se plantea? (1 punto)
- b) Determinar la probabilidad de error de tipo I de la prueba. (3 puntos)
- c) Determinar la probabilidad de error de tipo II si en realidad el tiempo medio de espera fuera de 10 minutos. (3 puntos)
- d) ¿En cuantos viajes (n) debería evaluar el tiempo de espera para que la diferencia entre el tiempo medio real y el tiempo medio de esos n viajes fuera menor que ± 1 minuto con una probabilidad mayor o igual al 95%?. (3 puntos)

APELLIDOS, NOMBRE: _____

Firma: _____

GRUPO: _____

BLOQUE I. CUESTIONES (75 MINUTOS)

Todas las cuestiones deben ser debidamente justificadas. Aquellas cuestiones contestadas correctamente pero no justificadas no serán puntuadas. Cada cuestión son 10 puntos. La nota del bloque es la media de las notas de las cinco cuestiones.

1.- Para aceptar una partida de bobinas de tela para confección se examinan 4 metros lineales cortados en una bobina elegida al azar y se determina el número de defectos que hay. Si dicho número es menor de 3, se acepta la partida, en caso contrario se rechaza. Se pide:

- Calcular la probabilidad de aceptar, con este procedimiento, una partida que contenga por término medio 1,5 defectos por metro lineal.
- ¿Cuántos metros lineales se deberán examinar para aumentar dicha probabilidad al 10%?

SOLUCIÓN:

v.a. X_1 : nº de defectos en un metro lineal de bobina $\rightarrow X_1 \approx Ps(\lambda = 1,5) \rightarrow E(X_1) = 1,5$

v.a. X_4 : nº de defectos en 4 metros lineales $\rightarrow X_4 \approx Ps(\lambda = 4 \cdot 1,5 = 6)$

a) $P(\text{aceptar}) = P(X_4 < 3) = P[Ps(\lambda = 6) \leq 2] = (\text{tablas}) = \mathbf{0,062}$

b) v.a. X_n : nº de defectos en “n” metros lineales $\rightarrow X_n \approx Ps(\lambda = 1,5n)$

$$P(X_n < 3) = 0,1; \quad P[Ps(\lambda = 1,5n) \leq 2] = 0,1 \xrightarrow{\text{tabla}} \lambda = 5,3 = 1,5n \rightarrow n = \mathbf{3,53 \text{ metros}}$$

2.- Un fabricante de coches compra motores a otra compañía. El fabricante recibe lotes de 40 motores y su plan de control de calidad consiste en seleccionar 8 de manera aleatoria y someterlos a prueba. Acepta el lote si ninguno de ellos presenta defectos, y de lo contrario lo rechaza.

- Si el lote contiene 2 motores defectuosos, ¿con qué probabilidad puede ser aceptado?
- ¿Cuál sería la probabilidad de aceptar el lote si éste fuera de 2000 motores y el proveedor asumiera un 5% de defectuosos?

SOLUCIÓN:

a) v.a. X : nº de motores defectuosos en la muestra $\rightarrow X \approx H(N = 40, n = 8, p = 2/40)$

$$P(\text{aceptar}) = P(X = 0) = \frac{\binom{40 \cdot 2 / 40}{0} \cdot \binom{40 \cdot 38 / 40}{8 - 0}}{\binom{40}{8}} = \frac{1 \cdot \binom{38}{8}}{\binom{40}{8}} = \frac{32 \cdot 31}{40 \cdot 39} = \mathbf{0,6359}$$

b) v.a. Y : nº de motores defectuosos en la muestra $\rightarrow Y \approx B(n = 8, p = 0,05)$

$$P(\text{aceptar}) = P(X = 0) = \binom{8}{0} \cdot 0,05^0 \cdot 0,95^8 = 0,95^8 = \mathbf{0,6634}$$

3.- Queremos verificar si el tiempo de espera en una consulta de un centro médico se ha reducido por debajo de los 30 minutos habituales después de un cambio organizativo en el servicio. Para ello seleccionamos 10 pacientes al azar y verificamos su tiempo de espera, obteniendo los siguientes valores: $\bar{x} = 26,5$ minutos, $s_{n-1}^2 = 10$. Plantear el test que debemos contrastar y determinar si el cambio organizativo ha sido efectivo, utilizando un nivel de significación $\alpha = 0,10$.

SOLUCIÓN:

Test de hipótesis: $H_0 : m = 30$; $H_1 : m < 30$

Aceptaremos H_0 si: $\bar{x} > m_0 - t_{n-1}^\alpha s_{n-1} / \sqrt{n}$; $t_{n-1}^\alpha = t_9^{0,1} = 1,383$; $26,5 > 30 - 1,383 \cdot \sqrt{10} / \sqrt{10}$

Aceptaremos H_0 si $26,5 > 28,617 \rightarrow$ No se cumple la condición, por lo que se rechaza H_0

4.- El gasto mensual en el uso del móvil de un alumno es una variable aleatoria con distribución Normal de media 20 euros y desviación típica 5. Calcular la “provisión” de fondos que tendrá que hacer el alumno para sufragar los gastos del tercer trimestre del año con una probabilidad superior al 85%.

SOLUCIÓN:

v.a. X: gasto mensual (euros)

v.a. Z: gasto trimestral $\rightarrow Z = X_1 + X_2 + X_3$

$E(Z) = E(X_1) + E(X_2) + E(X_3) = 3 \cdot 20 = 60$ Asumiendo independencia entre X_i :

$Var(Z) = Var(X_1 + X_2 + X_3) = Var(X_1) + Var(X_2) + Var(X_3) = 3 \cdot Var(X) = 3 \cdot 5^2 = 75$

$P(Z < n_{\min}) > 0,85$; $P[N(60; \sqrt{75}) < n_{\min}] > 0,85$; $P[N(0;1) < (n_{\min} - 60) / \sqrt{75}] > 0,85$

$(n_{\min} - 60) / \sqrt{75} > 1,04$; $n_{\min} > 69,01$; **$n_{\min} = 70$ euros**

5.- Dos cajas contienen piezas del proveedor A y piezas del proveedor B. Supongamos que una caja tiene 70 piezas de A y 30 de B, y que la otra caja tiene 10 de A y 25 de B. Seleccionamos una caja al azar y extraemos una pieza de la misma. Calcular la probabilidad de que la pieza sea del proveedor A.

SOLUCIÓN:

Suceso C_1 : la pieza pertenece a la caja 1; C_2 : la pieza pertenece a la caja 2

A: la pieza pertenece al proveedor A ; B: la pieza pertenece al proveedor B

$P(A/C_1) = 70/100$; $P(A/C_2) = 10/35$

$P(A) = P(A \cap C_1 \cup A \cap C_2) = P(A \cap C_1) + P(A \cap C_2) = P(C_1) \cdot P(A/C_1) + P(C_2) \cdot P(A/C_2) =$
 $= 0,5 \cdot 0,7 + 0,5 \cdot 10/35 = \mathbf{0,493}$

APELLIDOS, NOMBRE: _____

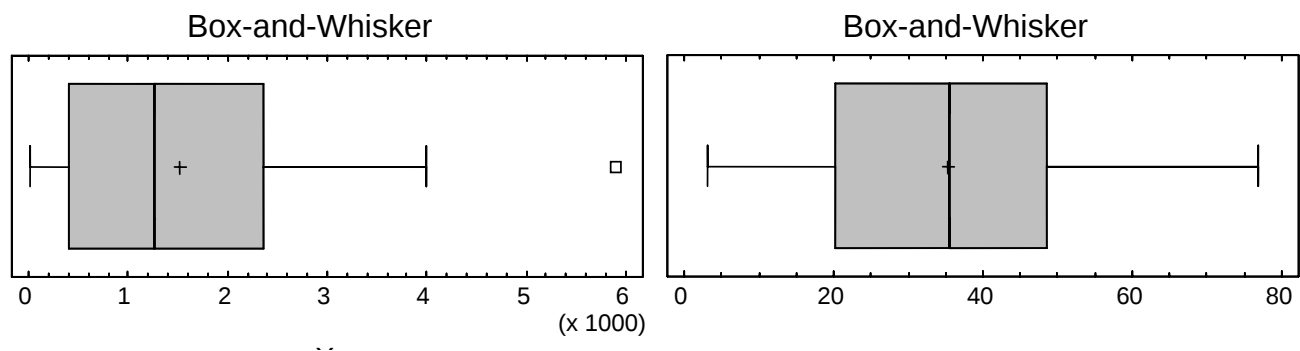
Firma: _____

Profesor de prácticas: _____

BLOQUE II. PRÁCTICAS DE ORDENADOR (45 MINUTOS)

Todas las cuestiones deben ser debidamente justificadas. Cada problema son 10 puntos. La nota del bloque es la media de las notas de los tres problemas.

1.- En una empresa que elabora alimentos congelados, uno de los parámetros de calidad es el número de microorganismos por gramo de producto (variable X). La empresa ha analizado dicho parámetro en un total de 50 partidas, y con los 50 datos resultantes se construye un diagrama Box-Whisker (gráfico inferior, izquierda). El gráfico de la derecha se ha construido con la raíz cuadrada de los datos. A la vista de los diagramas, responder a las siguientes preguntas justificando convenientemente las respuestas.



- a) La empresa está interesada en estudiar la pauta de distribución de la variable X . Para ello, ¿crees que hay valores de X anormalmente grandes que deberían eliminarse del estudio?

Solución: La distribución de X es asimétrica positiva, pues el bigote derecho es más largo que el izquierdo, la mediana está desplazada hacia la izquierda de la caja y la media es superior a la mediana. Se observa un punto aislado en el lado derecho, lo cual es habitual en distribuciones asimétricas positivas, por lo que no se puede considerar como punto anómalo que deba eliminarse del estudio. Además, el hecho de que la raíz cuadrada de la variable tienda a distribuirse normalmente es coherente con este criterio.

- b) ¿Cuál de las dos distribuciones, X o bien $\sqrt{X}$, tiene un menor coeficiente de asimetría?

Solución: la distribución de X es asimétrica positiva, por lo que su coeficiente de asimetría será positivo ($CA > 0$). En cambio, la de $\sqrt{X}$ es bastante simétrica, por lo que $CA \approx 0$. Por tanto, es menor el CA de $\sqrt{X}$.

- c) Se considera que la calidad de una partida es inadecuada si $X > 2400$. Calcular el porcentaje de partidas de mala calidad que se fabrican en la empresa.

Solución: 2400 es aproximadamente el tercer cuartil de X (extremo derecho de la caja). Por tanto, $P(X > 2400) \approx 0,25$.

d) Se desea realizar un contraste de hipótesis para determinar si puede considerarse que el número medio de microorganismos por gramo de producto, en promedio, es de 1600. Para ello, ¿cuál de las siguientes respuestas es verdadera? ¿Por qué?

1.- Habría que trabajar con los datos de X y la hipótesis nula a plantear sería $H_0: m = 1600$

2.- Sería preferible trabajar con $\sqrt{X}$ dado que ésta tiende a distribuirse normalmente.

3.- Los apartados 1.) y 2.) son equivalentes, daría igual trabajar con X o con $\sqrt{X}$.

4.- Ninguna de las anteriores, necesitaríamos conocer σ para estimar la media poblacional.

Solución: las fórmulas para estimar la media poblacional asumen que X sigue una distribución normal. En este caso X es asimétrica positiva, pero $\sqrt{X}$ tiende a distribuirse normalmente. Por tanto, el test de inferencia deberá realizarse con $\sqrt{X}$ y consecuentemente la respuesta correcta es la 2.).

2.- Una empresa de bebidas elabora refrescos en envases de 300 ml. La máquina que adiciona la bebida está calibrada para que el valor medio poblacional del volumen realmente adicionado sea 310 ml. Para saber si la máquina funciona correctamente o si por el contrario requiere ser ajustada, se toman 100 refrescos al azar y se determina el volumen de líquido que contienen. Los resultados, analizados con Statgraphics, son los siguientes:

Hypothesis Tests for volumen

Sample mean = 311,062

Sample median = 310,776

t-test

Null hypothesis: mean = 310,0

Alternative: not equal

Computed t statistic = 2,52085

P-Value = 0,000000

Confidence Intervals for volumen

95,0% confidence interval for mean: 311,062 +/- 0,836233 [310,226;311,899]

95,0% confidence interval for standard deviation: [3,70029;4,89578]

a) ¿Qué se concluye del estudio, considerando un nivel de significación del 5%?

Solución: como el valor de 310 no está dentro del intervalo [310,226; 311,899], se rechaza la hipótesis nula de que la media poblacional sea 310 ml. Por tanto, la máquina requiere ser ajustada.

- b) Justificar numéricamente el valor 2,52085 del estadístico de contraste, teniendo en cuenta que la cuasidesviación típica de los 100 datos vale 4,214.

Solución: $t = \frac{\bar{x} - m}{s'/\sqrt{n}} = \frac{311,062 - 310}{4,214/\sqrt{100}} = 2,520$

- c) Calcular el p-valor.

Solución: $p\text{-valor} = P(|t_{99}| > 2,52) = 2 \cdot P(t_{99} > 2,52) \approx 2 \cdot P[N(0;1) > 2,52] \approx 2 \cdot 0,006 \approx 0,012$

(el valor obtenido con Statgraphics es muy similar: 0,0133)

- d) ¿Cuánto valdría el p-valor si la hipótesis alternativa fuese $H_1: m > 310$?

Solución: valdría la mitad que en el apartado anterior, ya que en este caso,

$$p\text{-valor} = P(t_{99} > 2,52) \approx P[N(0;1) > 2,52] \approx 0,006$$

APELLIDOS, NOMBRE: _____

Firma: _____

GRUPO: _____

BLOQUE III. PROBLEMAS (90 MINUTOS)

Cada problema vale 10 puntos. La nota del bloque es la media de las notas de los tres problemas.

1.- La proporción de cemento, X , que se añade al hormigón y el tiempo de secado, Y (horas), se sabe que siguen una distribución Normal bidimensional de vector de medias y matriz de varianzas-covarianzas,

$$\bar{\mu} = \begin{Bmatrix} 0,3 \\ 9 \end{Bmatrix} \quad \mathbf{V} = \begin{bmatrix} 0,05 & 0,25 \\ 0,25 & 2 \end{bmatrix}$$

- Si la proporción añadida de cemento es de 0,25, ¿cuál es la probabilidad de que el tiempo de secado dure más de 10 horas? (3 puntos)
- ¿Qué proporción de cemento debemos añadir para que el tiempo de secado sea inferior a 10 horas con una probabilidad del 97%? (4 puntos)
- Si el tiempo de secado fuese de 10 horas, ¿cuál es la probabilidad de que la proporción de cemento sea menor de 0,6? (3 puntos)

SOLUCIÓN:

a) X : proporción de cemento; Y : tiempo de secado

$$\hat{Y} = m_Y + \frac{\text{cov}(X, Y)}{\sigma_X^2} (X - m_X) = 9 + \frac{0,25}{0,05} (X - 0,3) = 5X + 7,5$$

$$E(Y / X = 0,25) = 5 \cdot 0,25 + 7,5 = 8,75; \quad \rho = \text{cov}(X, Y) / (\sigma_X \cdot \sigma_Y) = 0,25 / (\sqrt{0,05} \cdot \sqrt{2}) = 0,791$$

$$D^2(U) = D^2(Y) \cdot (1 - \rho^2) = 2 \cdot (1 - 0,791^2) = 0,75 \rightarrow D(U) = 0,866$$

$$P(Y > 10 / X = 0,25) = P[N(8,75; 0,866) > 10] = 1 - \phi[(10 - 8,75) / 0,866] = 1 - \phi(1,44) = \mathbf{0,075}$$

b) $P(Y < 10 / X = x) = 0,97$; $P[N(m; 0,866) < 10] = 0,97$; $P[N(0;1) < (10 - m) / 0,866] = 0,97$
 $(10 - m) / 0,866 = 1,88 \rightarrow m = 8,372$; $m = E(Y / X = x) = 8,372 = 5 \cdot x + 7,5 \rightarrow \mathbf{x = 0,174}$

c) En este caso, Y : proporción de cemento; X : tiempo de secado

$$\hat{Y} = m_Y + \frac{\text{cov}(X, Y)}{\sigma_X^2} (X - m_X) = 0,3 + \frac{0,25}{2} (X - 9) = 0,125X - 0,825$$

$$E(Y / X = 10) = 0,125 \cdot 10 - 0,825 = 0,425$$

$$D^2(U) = D^2(Y) \cdot (1 - \rho^2) = 0,05 \cdot (1 - 0,791^2) = 0,0187 \rightarrow D(U) = 0,137$$

$$P(Y < 0,6 / X = 10) = P[N(0,425; 0,137) < 0,6] = \phi[(0,6 - 0,425) / 0,137] = \phi(1,276) = \mathbf{0,899}$$

2.- El tiempo de espera en el andén de “Los Naranjos” para coger el tranvía de la línea 4 podemos admitir que es una variable X con distribución Normal de desviación típica 2 minutos. Un alumno estaría dispuesto a coger el tranvía si el tiempo medio de espera real fuera menor o igual a 7 minutos. En caso contrario prefiere coger el coche. Para tomar una decisión hace la siguiente prueba: determina el tiempo de espera en 9 viajes al azar y si el promedio del tiempo de espera es inferior a 8 minutos coge el tranvía y en caso contrario coge el coche.

- a) ¿Cuál es el contraste de hipótesis que se plantea? (1 punto)
- b) Determinar la probabilidad de error de tipo I de la prueba. (3 puntos)
- c) Determinar la probabilidad de error de tipo II si en realidad el tiempo medio de espera fuera de 10 minutos. (3 puntos)
- d) ¿En cuantos viajes (n) debería evaluar el tiempo de espera para que la diferencia entre el tiempo medio real y el tiempo medio de esos n viajes fuera menor que ± 1 minuto con una probabilidad mayor o igual al 95%?. (3 puntos)

SOLUCIÓN:

- a) $H_0 : m \leq 7$; $H_1 : m > 7$ ya que el “igual” siempre se incluye en la hipótesis nula.
- b) $\alpha = P(\text{rechazar } H_0 / H_0) = P(\bar{X}_9 > 8 / H_0) = P[N(7; 2/\sqrt{9}) > 8] = 1 - P[N(7; 0,667) \leq 8] =$
 $= 1 - \phi[(8 - 7)/0,667] = 1 - \phi(1,5) = \mathbf{0,067}$
- c) $\beta = P(\text{aceptar } H_0 / H_1) = P(\bar{X}_9 < 8 / H_1) = P[N(10; 0,667) < 8] =$
 $= \phi[(8 - 10)/0,667] = \phi(-3) = \mathbf{0,0013}$
- d) $n \geq (z_{\alpha/2} \cdot \sigma / \varepsilon)^2 = (1,96 \cdot 2/1)^2 = 15,36 = \mathbf{16}$ viajes
 Otra forma: $\bar{X} \approx N(m; 2/\sqrt{n})$; $P[\bar{X} < (m - 1)] = 0,025$; $P[N(m; 2/\sqrt{n}) < (m - 1)] = 0,025$
 $\phi\left(\frac{m - 1 - m}{2/\sqrt{n}}\right) = 0,025$; $\frac{-1}{2/\sqrt{n}} = -1,96$; $n = (2 \cdot 1,96)^2 = 15,36 = \mathbf{16}$

EXAMEN DE ENERO DE 2009, CON SOLUCIÓN

APELLIDOS, NOMBRE: _____

Firma: _____

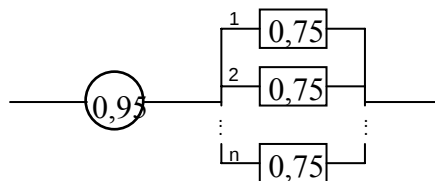
GRUPO: _____

BLOQUE I. CUESTIONES Y PROBLEMAS TIPO TEST (60 MINUTOS)

Todas las cuestiones deben ser debidamente justificadas. Aquellas cuestiones contestadas correctamente pero no justificadas no serán puntuadas.

1.- El tiempo de funcionamiento de una determinada pieza mecánica hasta su rotura se asume que sigue una distribución exponencial. Si la mediana de la distribución del tiempo vale 120 horas, calcular la probabilidad de que, al escoger al azar dos piezas, al menos una de ellas esté funcionando al cabo de 200 horas. *(2 puntos)*

2.- Dos tipos de componentes electrónicos con probabilidades de no fallo de 0,95 y 0,75 están conectados según el esquema indicado. ¿Cuál es el número n de elementos del segundo tipo que debemos colocar redundantes (ver figura) en paralelo para que la probabilidad de no fallo de todo el conjunto sea como mínimo de 0,90? ¿Cuál será la máxima probabilidad de no fallo del conjunto, que podríamos conseguir? *(1,5 puntos)*



3.- La resistencia de un tornillo en gr/mm^2 es una variable aleatoria con función de densidad:

$$f(x) = \begin{cases} 1 - k \cdot x & \text{si } 0 \leq x \leq 2 \\ 0 & \text{en el resto de casos} \end{cases}$$

a) Determinar el valor de k . *(0,5 puntos)*

b) ¿Cuál es el valor medio de la resistencia? *(0,5 puntos)*

c) Determinar la probabilidad de que un tornillo aguante más de $1,5 \text{ gr/mm}^2$ si para $1,0 \text{ gr/mm}^2$ aún resiste. *(1 punto)*

4.- El tiempo que tarda un horno en alcanzar los 1200°C es una variable uniforme que varía entre 15 y 25 minutos. El horno se ha puesto en marcha 25 veces en un mes. ¿Cuál es la probabilidad de que el tiempo promedio de esas 25 puestas en marcha supere los 24 minutos?

Nota: se asume que existe independencia en los valores de estos tiempos.

(1,5 puntos)

- a) 0,0344
- b) 0,0000
- c) 0,1264
- d) 0,1057

5.- De una población $X \sim N(10, 3)$ sacamos una muestra al azar de 28 valores. ¿Cuál es la probabilidad de que la cuasivarianza muestral sea mayor que 6? *(1,5 puntos)*

- a) 0,050
- b) 0,025
- c) 0,900
- d) 0,100

6.- Supongamos una muestra aleatoria $X_1, X_2, \dots, X_7$, de tamaño 7, tomada de una población que tiene media m y desviación típica d . Estudiamos un nuevo estadístico para estimar la media poblacional calculado según la fórmula:

$$\theta = \frac{2 \cdot X_1 - X_6 + X_4}{2}$$

¿Cuál es el mejor estimador de m : la media muestral o el estadístico θ ?

Nota: el mejor estimador es aquel que sea insesgado y de menor varianza

(1,5 puntos)

APELLIDOS, NOMBRE: _____

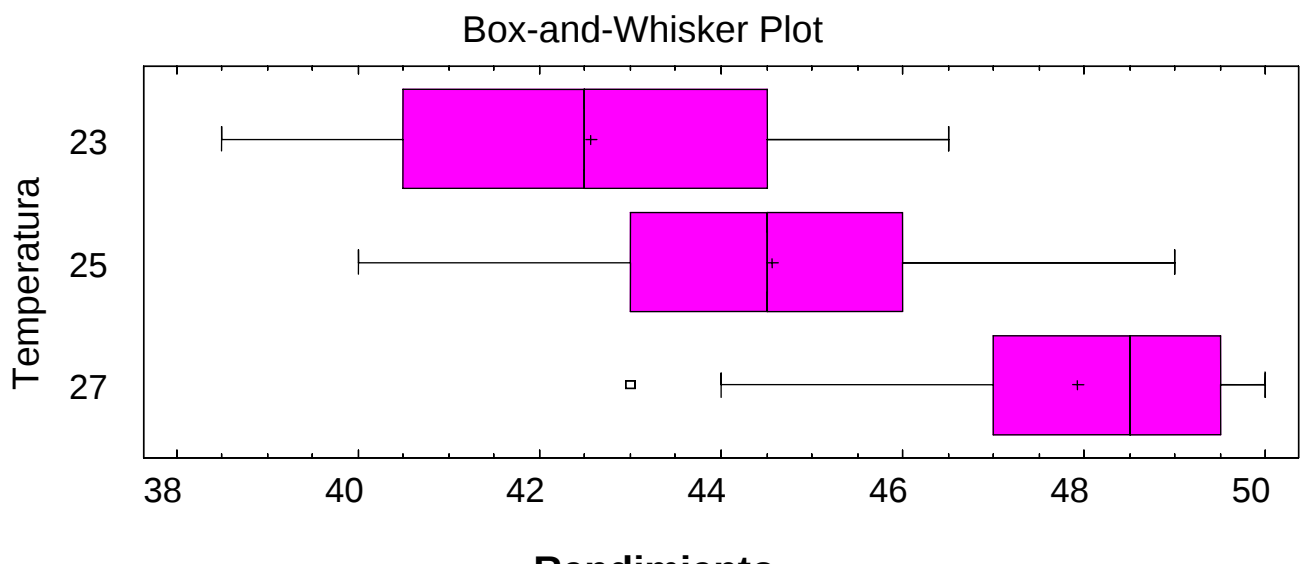
Firma: _____

Profesor de prácticas: _____

BLOQUE II. PRÁCTICAS DE ORDENADOR (45 MINUTOS)

Todas las cuestiones deben ser debidamente justificadas. Cada problema son 10 puntos. La nota del bloque es la media de las notas de los tres problemas.

1.- Una determinada empresa elabora una cierta materia activa por medio de un proceso de fermentación que se desarrolla a 23 °C. Tras obtener 30 lotes en estas condiciones, se decide aumentar la temperatura a 25 °C. Finalmente, tras otros 30 lotes, la temperatura se incrementa a 27 °C y se producen otros 30 lotes. Con los 90 datos de rendimiento se realiza el diagrama box-whisker múltiple que se muestra a continuación:



Responder a las siguientes preguntas justificando convenientemente las respuestas.

- ¿El valor del tercer cuartil del rendimiento obtenido a 23 °C es mayor que el segundo cuartil del rendimiento obtenido a 25 °C? (2,5 puntos)
- Comparando los diagramas de $T=23^{\circ}\text{C}$ y $T=25^{\circ}\text{C}$, ¿en cuál de los dos casos es mayor la dispersión de los datos? (2,5 puntos)
- ¿En cuál de los tres casos es menor el coeficiente de asimetría? (2,5 puntos)
- Si se realiza una regresión lineal simple entre temperatura y rendimiento, ¿cuál crees que sería el coeficiente de correlación? (2,5 puntos)
 - $r = -0.5$
 - próximo a cero (no existe correlación)
 - $r = 0.7$

2.- Una cierta universidad española realiza una encuesta a 300 de sus alumnos sobre el proceso de convergencia europea de las titulaciones académicas (plan Bolonia). Los resultados son los siguientes: 125 encuestados están a favor, 152 están en contra y 23 eligen la opción "no sabe/no contesta". La universidad realiza una campaña informativa sobre el plan Bolonia. Al terminar esta campaña se repite la misma encuesta con 200 alumnos, de los cuales 97 responden estar a favor, 86 responden que están en contra y 17 responden "no sabe/no contesta". La universidad concluye que la campaña informativa ha sido un éxito, pues inicialmente el porcentaje de encuestados en contra era superior al porcentaje de encuestados a favor, y tras realizar la campaña informativa los resultados son inversos. Para analizar el efecto de la campaña informativa, se analizan los datos con Statgraphics, cuyos resultados se muestran a continuación.

Frequency Table

	ns/nc	a favor	en contra	Row Total
antes campaña	23	125	152	300
	24,00	133,20	142,80	60,00%
	0,04	0,50	0,59	
despues camp.	17	97	86	200
	16,00	88,80	95,20	40,00%
	0,06	0,76	0,89	
Column	40	222	238	500
Total	8,00%	44,40%	47,60%	100,00%

Cell contents:

Observed frequency

Expected frequency

Contribution to chi-squared

- a) ¿Cuál es la hipótesis nula que se plantea? (2,5 puntos)
- b) Calcular el valor del estadístico de contraste. (2,5 puntos)
- c) ¿Cuál es la distribución teórica del estadístico de contraste si la hipótesis nula fuese cierta? (2,5 puntos)
- d) ¿Puede afirmarse con un nivel de confianza del 95% que la campaña informativa realmente ha sido un éxito? (2,5 puntos)

3.- En un determinado proceso químico se elabora un cierto producto líquido. La viscosidad resultante es el principal parámetro de calidad. Los técnicos sospechan que la viscosidad puede depender de la temperatura de reacción y de la cantidad de catalizador. Para estudiar esta hipótesis, se toman los datos de viscosidad, temperatura y cantidad de catalizador correspondientes a 50 lotes del producto y se realiza un análisis de regresión lineal múltiple cuyos resultados se muestran a continuación.

Multiple Regression Analysis

Dependent variable: viscosidad

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	-24,8334	19,7837	-1,25525	0,2156
temperat	3,32293	0,306009	10,8589	0,0000
cataliz	0,0272425	0,010429	2,61218	0,0120

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	6029,79	2	3014,89	65,15	0,0000
Residual	2175,11	47	46,279		
Total (Corr.)	8204,9	49			

R-squared = 73,4901 percent

R-squared (adjusted for d.f.) = 72,362 percent

Standard Error of Est. = 6,80287

Mean absolute error = 5,24693

Durbin-Watson statistic = 1,81172 (P=0,2549)

Lag 1 residual autocorrelation = 0,0658777

- Calcular el valor del coeficiente de determinación. ¿Cómo se interpreta en la práctica este parámetro? (2,5 puntos)
- Obtener la ecuación matemática que se recomendaría para predecir la viscosidad en función de las variables que ejercen un efecto estadísticamente significativo (considerar $\alpha=0,05$). (2,5 punt.)
- Interpretar el significado práctico del coeficiente asociado a la variable temperatura. (2 puntos)
- Se sospecha que pueda existir un efecto cuadrático de la temperatura. ¿Cómo se puede verificar esta hipótesis? ¿Cuál sería H_0 y H_1 del contraste de hipótesis a plantear? (3 puntos)

BLOQUE III. PROBLEMAS (90 MINUTOS)

Cada problema vale 10 puntos. La nota del bloque es la media de las notas de los tres problemas.

1.- En la elaboración de un determinado medicamento en forma de comprimido, uno de los parámetros de calidad críticos es el peso de cada comprimido. Esta variable sigue aproximadamente una distribución Normal de media 3 grs. y de desviación típica 0,05 grs.

- a) Calcular la probabilidad de que un comprimido pese más de 3,025 grs. (2,5 puntos)
- b) Un comprimido se considera defectuoso cuando su peso difiere de la media en más de 0,075 grs. Calcular la proporción de comprimidos defectuosos que se fabrican. (2,5 puntos)
- c) Estos comprimidos se envasan en cajas de 10 unidades. Si un envase contiene 2 o más comprimidos defectuosos se elimina del mercado. Determinar el porcentaje de cajas que se retiran del mercado. (2,5 puntos)
- d) Una farmacia hace un pedido de 100 cajas de dicho medicamento. Calcular la probabilidad de que en el pedido haya más de 40 cajas no aptas para la venta. (2,5 puntos)

2.- Una cierta empresa fabrica componentes electrónicos con un porcentaje de unidades defectuosas del 10%. El estado (correcto o defectuoso) de cada uno de los componentes fabricados se comprueba mediante un ensayo que cuando el componente funciona, la probabilidad de que el ensayo diga lo contrario es de 0,05. Si el componente es defectuoso, el ensayo no se equivoca.

El coste de fabricación del componente es de 3 € y se vende a 10 €, pero si al instalarlo resulta defectuoso, hay que devolver al cliente la cantidad cobrada más una indemnización de 11 €.

Se pide:

- a) Si el ensayo dice que el componente electrónico es defectuoso, ¿cuál es la probabilidad de que realmente lo sea? (5 puntos)
- b) ¿Cuál será el precio máximo que se podrá pedir por el ensayo? (5 puntos)

3.- Una empresa de productos alimentarios quiere lanzar al mercado un nuevo derivado lácteo con bífidos activo y ácidos grasos omega-3. Con el producto actual tiene una cuota habitual de mercado del 20% de los consumidores de derivados lácteos, y sólo lanzará el nuevo producto si un estudio de mercado demuestra que la cuota va a aumentar (no es previsible en absoluto que disminuya). En el estudio, que consistió en entrevistar a 150 consumidores de lácteos, se constató que 50 de los entrevistados estarían dispuestos a comprar el nuevo producto por sus características y precio.

- a) Con un nivel de significación $\alpha=0,05$, ¿qué decisión deberá tomar la empresa? (Plantear el test de hipótesis apropiado y justificar la decisión). (3 puntos)
- b) ¿Cuál sería la probabilidad de error de segunda especie β si la nueva proporción de consumidores fuera realmente del 22%? (3 puntos)
- c) Establecer un intervalo de confianza bilateral con un nivel de confianza del 95% para la proporción de compradores del nuevo producto. (2 puntos)
- d) Si queremos estimar la proporción de nuevos compradores con un error menor de $\pm 0,025$, un nivel de confianza del 90% y consideramos como valor previo y aproximado de P el 20% ($P \approx 0,20$), ¿a cuantos consumidores de lácteos deberíamos entrevistar para conseguir el objetivo previsto? (2 puntos)

APELLIDOS, NOMBRE: _____

Firma: _____

GRUPO: _____

BLOQUE I. CUESTIONES Y PROBLEMAS TIPO TEST (60 MINUTOS)

Todas las cuestiones deben ser debidamente justificadas. Aquellas cuestiones contestadas correctamente pero no justificadas no serán puntuadas.

1.- El tiempo de funcionamiento de una determinada pieza mecánica hasta su rotura se asume que sigue una distribución exponencial. Si la mediana de la distribución del tiempo vale 120 horas, calcular la probabilidad de que, al escoger al azar dos piezas, al menos una de ellas esté funcionando al cabo de 200 horas. (2 puntos)

Solución:

La función de distribución de la exponencial es: $F(x) = 1 - e^{-\lambda x}$

Si la mediana es 120, sabemos que: $F(120) = 0,5 = 1 - e^{-\lambda \cdot 120}$

Y despejando λ tenemos: $e^{-\lambda \cdot 120} = 0,5 \quad -\lambda \cdot 120 = \ln(0,5) \quad \lambda = \frac{-\ln(0,5)}{120} = 0,0058$

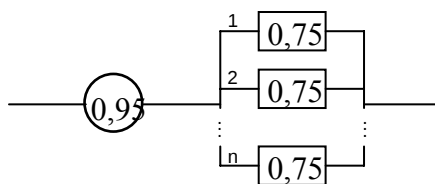
La probabilidad de que esté funcionando al cabo de 200 horas es,

$$P(X > 200) = e^{-0,0058 \cdot 200} = 0,3135$$

Y de que de 2 al menos una de ellas esté funcionando es,

$$\binom{2}{1} \cdot 0,3135 \cdot (1 - 0,3135) + \binom{2}{2} \cdot 0,3135^2 = 0,53$$

2.- Dos tipos de componentes electrónicos con probabilidades de no fallo de 0,95 y 0,75 están conectados según el esquema indicado. ¿Cuál es el número n de elementos del segundo tipo que debemos colocar redundantes (ver figura) en paralelo para que la probabilidad de no fallo de todo el conjunto sea como mínimo de 0,90? ¿Cuál será la máxima probabilidad de no fallo del conjunto, que podríamos conseguir? (1,5 puntos)



Solución:

$P(\text{no fallo del conjunto}) = P(\text{no fallo del subconjunto 1}) \cdot P(\text{no fallo del 2})$

$P(\text{no fallo subconjunto 2}) = 1 - P(\text{fallo subc. 2}) = 1 - P(\text{fallo 2}) \cdot P(\text{fallo 2}) \dots P(\text{fallo 2}) = 1 - 0,25^n$

La probabilidad de no fallo del conjunto es: $P(\text{no fallo}) = 0,95 \cdot (1 - 0,25^n) \geq 0,90$

Despejando n se tiene: $(1 - 0,25^n) \geq \frac{0,90}{0,95} \quad n \geq \frac{\ln[1 - (0,9/0,95)]}{\ln(0,25)} = 2,124$

Y por lo tanto $n=3$, se deben colocar 3 componentes en paralelo.

La máxima fiabilidad que se puede conseguir (a medida que aumenta el número de elementos en paralelo) es el valor de la mínima probabilidad de los elementos en serie, es decir, 0,95. Se cumple el dicho de que “una cadena es tan fuerte como el eslabón más débil”.

3.- La resistencia de un tornillo en gr/mm^2 es una variable aleatoria con función de densidad:

$$f(x) = \begin{cases} 1 - k \cdot x & \text{si } 0 \leq x \leq 2 \\ 0 & \text{en el resto de casos} \end{cases}$$

- Determinar el valor de k . (0,5 puntos)
- ¿Cuál es el valor medio de la resistencia? (0,5 puntos)
- Determinar la probabilidad de que un tornillo aguante más de 1,5 gr/mm^2 si para 1,0 gr/mm^2 aún resiste. (1 punto)

Solución:

$$a) \int_0^2 f(x) \cdot dx = 1 \quad ; \quad \int_0^2 (1 - k \cdot x) \cdot dx = 1 \quad ; \quad \left[x - 0,5k \cdot x^2 \right]_0^2 = 1 \quad ; \quad 2 - 0,5 \cdot k \cdot 4 = 1 \quad ; \quad k = 0,5$$

$$b) E(x) = \int f(x) \cdot x \cdot dx = \int_0^2 x \cdot (1 - 0,5x) dx = \int_0^2 (x - 0,5x^2) dx = \left[\frac{1}{2}x^2 - \frac{1}{6}x^3 \right]_0^2 = \frac{4}{2} - \frac{8}{6} = \frac{2}{3}$$

$$c) P(x > 1,5 / x > 1) = \frac{P[(x > 1,5) \cap (x > 1)]}{P(x > 1)} = \frac{P(x > 1,5)}{P(x > 1)} = \frac{0,0625}{0,25} = \frac{1}{4} \quad \text{ya que:}$$

$$P(x > 1,5) = \int_{1,5}^2 (1 - 0,5x) dx = \left[x - 0,25x^2 \right]_{1,5}^2 = (2 - 0,25 \cdot 4) - (1,5 - 0,25 \cdot 1,5^2) = 0,0625$$

$$P(x > 1) = \int_1^2 (1 - 0,5x) dx = \left[x - 0,25x^2 \right]_1^2 = (2 - 0,25 \cdot 4) - (1 - 0,25 \cdot 1^2) = 0,25$$

4.- El tiempo que tarda un horno en alcanzar los 1200°C es una variable uniforme que varía entre 15 y 25 minutos. El horno se ha puesto en marcha 25 veces en un mes. ¿Cuál es la probabilidad de que el tiempo promedio de esas 25 puestas en marcha supere los 24 minutos?

Nota: se asume que existe independencia en los valores de estos tiempos.

(1,5 puntos)

- 0,0344
- 0,0000
- 0,1264
- 0,1057

Solución:

v.a. X : tiempo en alcanzar $1200^\circ\text{C} \rightarrow X \sim U(15, 25) \rightarrow$

$$\rightarrow E(X) = 20 \quad ; \quad \sigma^2(X) = (b - a)^2 / 12 = (25 - 15)^2 / 12 = 8,333 \rightarrow \sigma(X) = \sqrt{8,333} = 2,887$$

Aunque X es una distribución uniforme, por el teorema central del límite (y asumiendo independencia entre los valores de tiempo) se cumple que la media muestral tiende a distribuirse según un modelo normal:

$$\bar{x} \approx N(\mu; \sigma / \sqrt{n}) \approx N(20; 2,887 / \sqrt{25}) \approx N(20; 0,577)$$

$$P(\bar{x} > 24) = P[N(20; 0,577) > 24] = P[N(0;1) > (24 - 20)/0,577] = P[N(0;1) > 6,93] = 0$$

5.- De una población $X \sim N(10, 3)$ sacamos una muestra al azar de 28 valores. ¿Cuál es la probabilidad de que la cuasivarianza muestral sea mayor que 6? (1,5 puntos)

- a) 0,050
- b) 0,025
- c) 0,900
- d) 0,100

Solución: $(n-1) \cdot \frac{s_{n-1}^2}{\sigma^2} \approx \chi_{n-1}^2$; $P(S_{n-1}^2 > 6) = P\left(\frac{27S_{n-1}^2}{9} > \frac{27 \cdot 6}{9}\right) = P(\chi_{27}^2 > 18) \cong 0,90$

6.- Supongamos una muestra aleatoria $X_1, X_2, \dots, X_7$, de tamaño 7, tomada de una población que tiene media m y desviación típica d . Estudiamos un nuevo estadístico para estimar la media poblacional calculado según la fórmula:

$$\theta = \frac{2 \cdot X_1 - X_6 + X_4}{2}$$

¿Cuál es el mejor estimador de m : la media muestral o el estadístico θ ?

Nota: el mejor estimador es aquel que sea insesgado y de menor varianza

(1,5 puntos)

Solución: la media muestral es un estimador insesgado ya que $E(\bar{x}) = m$

$$E(\theta) = E\left(\frac{2 \cdot X_1 - X_6 + X_4}{2}\right) = \frac{2 \cdot E(X_1) - E(X_6) + E(X_4)}{2} = \frac{2 \cdot m - m + m}{2} = m$$

$$\sigma^2(\theta) = \sigma^2\left(\frac{2 \cdot X_1 - X_6 + X_4}{2}\right) = \frac{2^2 \cdot \sigma^2(X_1) + \sigma^2(X_6) + \sigma^2(X_4)}{2^2} = \frac{4d^2 + d^2 + d^2}{4} = 1,5d^2$$

$$\sigma^2(\bar{x}) = \frac{\sigma^2(X)}{n} = \frac{d^2}{7} = 0,143d^2$$

Los dos estimadores son insesgados pero la media muestral tiene menor varianza. Por tanto, la media muestral será mejor estimador que θ .

APELLIDOS, NOMBRE: _____

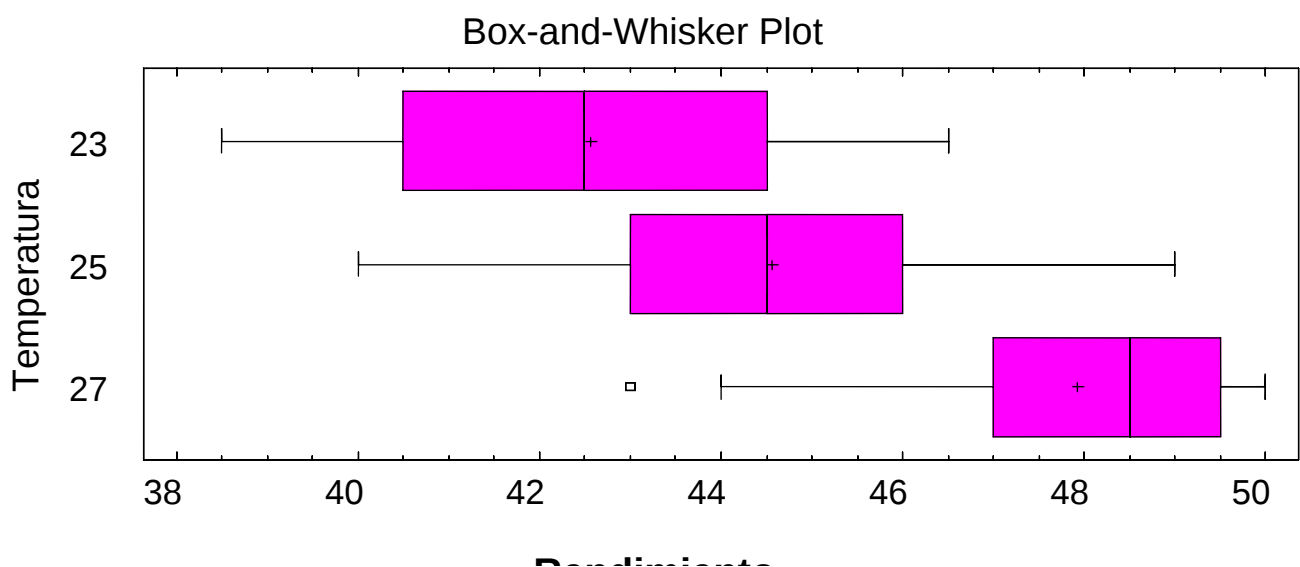
Firma: _____

Profesor de prácticas: _____

BLOQUE II. PRÁCTICAS DE ORDENADOR (45 MINUTOS)

Todas las cuestiones deben ser debidamente justificadas. Cada problema son 10 puntos. La nota del bloque es la media de las notas de los tres problemas.

1.- Una determinada empresa elabora una cierta materia activa por medio de un proceso de fermentación que se desarrolla a 23 °C. Tras obtener 30 lotes en estas condiciones, se decide aumentar la temperatura a 25 °C. Finalmente, tras otros 30 lotes, la temperatura se incrementa a 27 °C y se producen otros 30 lotes. Con los 90 datos de rendimiento se realiza el diagrama box-whisker múltiple que se muestra a continuación:



A la vista del diagrama, responder a las siguientes preguntas justificando convenientemente las respuestas.

- a) ¿El valor del tercer cuartil del rendimiento obtenido a 23 °C es mayor que el segundo cuartil del rendimiento obtenido a 25 °C? (2,5 puntos)

Solución: el tercer cuartil (extremo derecho de la caja) a $T=23$ vale 44,5. El segundo cuartil (mediana, línea central de la caja) a $T=25$ vale 44,5. Por tanto los dos valores son iguales.

- b) Comparando los diagramas de $T=23^{\circ}\text{C}$ y $T=25^{\circ}\text{C}$, ¿en cuál de los dos casos es mayor la dispersión de los datos? (2,5 puntos)

Solución: tanto el rango como el intervalo intercuartílico (IIC) son indicadores de la dispersión.

$$\text{Rango}_{T=23} = 46,5 - 38,5 = 8 ; \quad \text{Rango}_{T=25} = 49 - 40 = 9$$

$$\text{IIC}_{T=23} = 44,5 - 40,5 = 4 ; \quad \text{IIC}_{T=25} = 46 - 43 = 3$$

Resulta que a $T=23$ el rango es menor pero el IIC es mayor. Teniendo en cuenta que el IIC es un estimador más robusto de la dispersión, se concluye que la dispersión es mayor a $T=23$.

c) ¿En cuál de los tres casos es menor el coeficiente de asimetría? (2,5 puntos)

Solución: El diagrama Box-Whisker para T=23 y T=25 es bastante simétrico, por lo que su coeficiente de asimetría será próximo a cero. En cambio, para T=27 la forma del diagrama indica una asimetría negativa (bigote izquierdo más largo que el derecho, y mediana mayor que la media), de modo que su coeficiente de asimetría será negativo. Por tanto, el valor menor de los tres coeficientes corresponde a T=27.

d) Si se realiza una regresión lineal simple entre temperatura y rendimiento, ¿cuál crees que sería el coeficiente de correlación? (2,5 puntos)

a) $r = -0.5$

b) próximo a cero (no existe correlación)

c) $r = 0.7$

Solución: El gráfico indica una clara correlación (el rendimiento varía al variar la temperatura) por lo que se descarta la opción b). Teniendo en cuenta que el rendimiento aumenta al aumentar la temperatura, la correlación será positiva, y la solución es por tanto la c).

2.- Una cierta universidad española realiza una encuesta a 300 de sus alumnos sobre el proceso de convergencia europea de las titulaciones académicas (plan Bolonia). Los resultados son los siguientes: 125 encuestados están a favor, 152 están en contra y 23 eligen la opción "no sabe/no contesta". La universidad realiza una campaña informativa sobre el plan Bolonia. Al terminar esta campaña se repite la misma encuesta con 200 alumnos, de los cuales 97 responden estar a favor, 86 responden que están en contra y 17 responden "no sabe/no contesta". La universidad concluye que la campaña informativa ha sido un éxito, pues inicialmente el porcentaje de encuestados en contra era superior al porcentaje de encuestados a favor, y tras realizar la campaña informativa los resultados son inversos. Para analizar el efecto de la campaña informativa, se analizan los datos con Statgraphics, cuyos resultados se muestran a continuación.

Frequency Table

	ns/nc	a favor	en contra	Row Total
antes campaña	23	125	152	300
	24,00	133,20	142,80	60,00%
	0,04	0,50	0,59	
despues camp.	17	97	86	200
	16,00	88,80	95,20	40,00%
	0,06	0,76	0,89	
Column	40	222	238	500
Total	8,00%	44,40%	47,60%	100,00%

Cell contents:

*Observed frequency**Expected frequency**Contribution to chi-squared*

a) ¿Cuál es la hipótesis nula que se plantea? (2,5 puntos)

Solución: el análisis realizado es un contraste de independencia. Por tanto, la hipótesis nula es que las filas y columnas de la tabla son independientes. Es decir, que la opinión de los encuestados es la misma antes o después de la campaña informativa.

b) Calcular el valor del estadístico de contraste. (2,5 puntos)

$$D = \sum \frac{(O_i - E_i)^2}{E_i} = \sum \text{contribution to chi-square} = 0,04 + 0,5 + 0,59 + 0,06 + 0,76 + 0,89 = 2,85$$

c) ¿Cuál es la distribución teórica del estadístico de contraste si la hipótesis nula fuese cierta? (2,5 puntos)

Solución: $D \approx \chi^2_{(h-1)(k-1)} = \chi^2_{(3-1)(2-1)} = \chi^2_2$

La distribución teórica es una chi-cuadrado con 2 grados de libertad: (3 columnas–1) · (2 filas–1)

d) ¿Puede afirmarse con un nivel de confianza del 95% que la campaña informativa realmente ha sido un éxito? (2,5 puntos)

Solución: el valor crítico de una chi-cuadrado con 2 grados de libertad considerando $\alpha=0,05$ vale 5,99. Como el valor del estadístico de contraste es menor que el valor crítico, se acepta la hipótesis nula. Por tanto, no hay evidencia suficiente para concluir que la campaña informativa haya afectado significativamente a la opinión de los alumnos.

3.- En un determinado proceso químico se elabora un cierto producto líquido. La viscosidad resultante es el principal parámetro de calidad. Los técnicos sospechan que la viscosidad puede depender de la temperatura de reacción y de la cantidad de catalizador. Para estudiar esta hipótesis, se toman los datos de viscosidad, temperatura y cantidad de catalizador correspondientes a 50 lotes del producto y se realiza un análisis de regresión lineal múltiple cuyos resultados se muestran a continuación.

 Dependent variable: viscosidad

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	-24,8334	19,7837	-1,25525	0,2156
temperat	3,32293	0,306009	10,8589	0,0000
cataliz	0,0272425	0,010429	2,61218	0,0120

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	6029,79	2	3014,89	65,15	0,0000
Residual	2175,11	47	46,279		
Total (Corr.)	8204,9	49			

R-squared = 73,4901 percent

R-squared (adjusted for d.f.) = 72,362 percent

Standard Error of Est. = 6,80287

Mean absolute error = 5,24693

Durbin-Watson statistic = 1,81172 (P=0,2549)

Lag 1 residual autocorrelation = 0,0658777

a) Calcular el valor del coeficiente de determinación. ¿Cómo se interpreta en la práctica este parámetro? (2,5 puntos)

Solución: coeficiente de determinación = R-squared = 73,49%. Este parámetro indica que el modelo explica el 73,49% de la varianza de la viscosidad.

b) Obtener la ecuación matemática que se recomendaría para predecir la viscosidad en función de las variables que ejercen un efecto estadísticamente significativo (considerar $\alpha=0,05$). (2,5 punt.)

Solución: las variables que ejercen un efecto estadísticamente significativo son temperatura y catalizador, dado que su correspondiente p-valor es menor que 0,05. Por tanto, ambas variables deberán estar en el modelo. La constante no es estadísticamente significativa y podría eliminarse, con lo cual habría que volver a ajustar el modelo para estimar los coeficientes de las variables. Pero dado que no se dispone de esta información, conviene mantener la constante en el modelo, de modo que la ecuación sería: Viscosidad = -24,83 + 3,323·Temperat+0,02724·cataliz

c) Interpretar el significado práctico del coeficiente asociado a la variable temperatura. (2 puntos)

Solución: dicho coeficiente vale 3,32. Interpretación: si la temperatura aumenta en un grado centígrado, la viscosidad aumentará en promedio en 3,32 unidades.

d) Se sospecha que pueda existir un efecto cuadrático de la temperatura. ¿Cómo se puede verificar esta hipótesis? ¿Cuál sería H_0 y H_1 del contraste de hipótesis a plantear? (3 puntos)

Solución: introducir en el modelo el término cuadrático: temperatura^2 . Es decir, ajustar el modelo: $\text{viscosidad} = a + b \cdot \text{Temp} + c \cdot \text{Temp}^2 + d \cdot \text{cataliz}$

El contraste de hipótesis a plantear será: $H_0: c=0$ $H_1: c \neq 0$

Si el p-valor asociado a este contraste es menor que α se rechazará H_0 , concluyéndose que el efecto cuadrático es estadísticamente significativo.

APELLIDOS, NOMBRE: _____

Firma: _____

GRUPO: _____

BLOQUE III. PROBLEMAS (90 MINUTOS)

Cada problema vale 10 puntos. La nota del bloque es la media de las notas de los tres problemas.

1.- En la elaboración de un determinado medicamento en forma de comprimido, uno de los parámetros de calidad críticos es el peso de cada comprimido. Esta variable sigue aproximadamente una distribución Normal de media 3 grs. y de desviación típica 0,05 grs.

- Calcular la probabilidad de que un comprimido pese más de 3,025 grs. (2,5 puntos)
- Un comprimido se considera defectuoso cuando su peso difiere de la media en más de 0,075 grs. Calcular la proporción de comprimidos defectuosos que se fabrican. (2,5 puntos)
- Estos comprimidos se envasan en cajas de 10 unidades. Si un envase contiene 2 o más comprimidos defectuosos se elimina del mercado. Determinar el porcentaje de cajas que se retiran del mercado. (2,5 puntos)
- Una farmacia hace un pedido de 100 cajas de dicho medicamento. Calcular la probabilidad de que en el pedido haya más de 40 cajas no aptas para la venta. (2,5 puntos)

Solución:

$$a) P(X > 3,025) = P[N(3; 0,05) > 3,025] = P[N(0; 1) > (3,025 - 3)/0,05] = P[N(0; 1) > 0,5] = \mathbf{0,3085}$$

$$b) P(|x - 3| > 0,075) = 2 \cdot P(x > 3,075) = 2 \cdot P[N(3; 0,05) > 3,075] = 2 \cdot P[N(0; 1) > (3,075 - 3)/0,05] = 2 \cdot P[N(0; 1) > 1,5] = 2 \cdot 0,0668 = \mathbf{0,1336}$$

c) v.a. Y: número de comprimidos defectuosos en una caja de 10 unidades $\rightarrow Y \approx Bi(10; 0,1336)$

$$P(Y \geq 2) = 1 - P(Y = 0) - P(Y = 1) = 1 - \binom{10}{0} \cdot 0,134^0 \cdot (1 - 0,134)^{10} - \binom{10}{1} \cdot 0,134^1 \cdot (1 - 0,134)^9 = 1 - 0,2383 - 0,3675 = \mathbf{0,394}$$

d) v.a. Z: número de cajas no aptas en un pedido de 100 cajas $\rightarrow Z \approx Bi(100; 0,396)$

$$P(Z > 40) = P[Bi(100; 0,394) > 40] \approx P\left[N\left(100 \cdot 0,394; \sqrt{100 \cdot 0,394 \cdot (1 - 0,394)}\right) \geq 40,5\right] = P[N(39,4; 4,886) > 40,5] = P[N(0; 1) > (40,5 - 39,4)/4,886] = P[N(0; 1) > 0,225] = \mathbf{0,41}$$

2.- Una cierta empresa fabrica componentes electrónicos con un porcentaje de unidades defectuosas del 10%. El estado (correcto o defectuoso) de cada uno de los componentes fabricados se comprueba mediante un ensayo que cuando el componente funciona, la probabilidad de que el ensayo diga lo contrario es de 0,05. Si el componente es defectuoso, el ensayo no se equivoca.

El coste de fabricación del componente es de 3€ y se vende a 10€, pero si al instalarlo resulta defectuoso, hay que devolver al cliente la cantidad cobrada más una indemnización de 11€.

Se pide:

- Si el ensayo dice que el componente electrónico es defectuoso, ¿cuál es la probabilidad de que realmente lo sea? (5 puntos)
- ¿Cuál será el precio máximo que se podrá pedir por el ensayo? (5 puntos)

Solución:

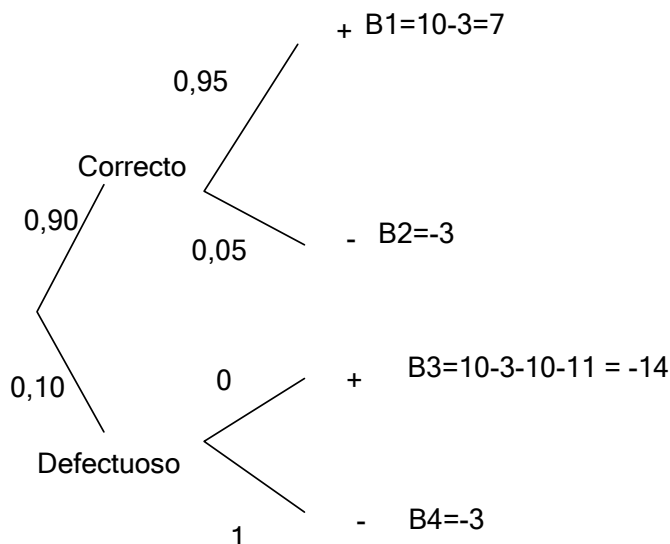
a) Suceso D: el componente es defectuoso $\rightarrow P(D)=0,1$

Suceso E: el ensayo dice que el componente es defectuoso $\rightarrow P(E/\bar{D})=0,05$; $P(E/D)=1$

$$P(E) = P(D) \cdot P(E/D) + P(\bar{D}) \cdot P(E/\bar{D}) = 0,1 \cdot 1 + 0,9 \cdot 0,05 = 0,145$$

$$P(D/E) = \frac{P(D \cap E)}{P(E)} = \frac{P(D) \cdot P(E/D)}{P(E)} = \frac{0,1 \cdot 1}{0,145} = 0,69$$

b) El árbol de probabilidades es:



Con el ensayo, el valor medio del beneficio viene dado por:

$$E(B) = 7 \cdot 0,90 \cdot 0,95 + (-3) \cdot 0,90 \cdot 0,05 + 0 + (-3) \cdot 0,10 = 5,55$$

Cuando no hay ensayo e instalamos el componente, por término medio el beneficio es:

$$E(B) = 7 \cdot 0,90 + (-14) \cdot 0,10 = 4,9$$

Lo máximo que se puede pedir por el ensayo es la diferencia del beneficio medio cuando está el ensayo y cuando no está: **Max= 5,55-4,9 = 0,65 €/ud**

b) Otra forma de resolver el apartado sin usar el árbol de probabilidad es usando la siguiente tabla. En primer lugar se calcula cuál sería el beneficio medio si no se realizase el ensayo (dos primeras columnas de la tabla, hay dos situaciones: que la pieza sea defectuosa o correcta). A continuación se calcula el beneficio medio si se realiza el ensayo (en este caso hay 4 posibilidades).

	Sin ensayo		Con ensayo			
	D	$\bar{D}$	$D \cap E$	$D \cap \bar{E}$	$\bar{D} \cap E$	$\bar{D} \cap \bar{E}$
A: coste	3	3	3	3	3	3
B: ingreso por venta	10	10	0	10	0	10
C: devolución +indemnización	21	0	0	21	0	0
D: coste prueba	0	0	k	k	k	k
Beneficio (-A+B-C-D)	-14	7	-3-k	-14-k	-3-k	7-k
Probabilidad	0,1	0,9	$P(D) \cdot P(E/D)$ 0,1·1=0,1	$P(D) \cdot P(\bar{E}/D)$ 0,1·0=0	$P(\bar{D}) \cdot P(E/\bar{D})$ 0,9·0,05	$P(\bar{D}) \cdot P(\bar{E}/\bar{D})$ 0,9·0,95

Sin ensayo: si llamamos X al beneficio, $E(X) = \sum X_i \cdot P(X = x_i) = -14 \cdot 0,1 + 7 \cdot 0,9 = 4,9 \text{ €/ud}$

Con ensayo: $E(X) = (-3-k) \cdot 0,1 - (3+k) \cdot 0,045 + (7-k) \cdot 0,855 = 5,55-k \text{ €/ud}$

El precio máximo a pagar por el ensayo será aquel que produzca el mismo beneficio medio que si no se hubiese realizado el ensayo: $5,55-k > 4,9 \rightarrow k < 0,65 \text{ €/ud}$

3.- Una empresa de productos alimentarios quiere lanzar al mercado un nuevo derivado lácteo con bífidos activo y ácidos grasos omega-3. Con el producto actual tiene una cuota habitual de mercado del 20% de los consumidores de derivados lácteos, y sólo lanzará el nuevo producto si un estudio de mercado demuestra que la cuota va a aumentar (no es previsible en absoluto que disminuya). En el estudio, que consistió en entrevistar a 150 consumidores de lácteos, se constató que 50 de los entrevistados estarían dispuestos a comprar el nuevo producto por sus características y precio.

- Con un nivel de significación $\alpha=0,05$, ¿qué decisión deberá tomar la empresa? (Plantear el test de hipótesis apropiado y justificar la decisión). (3 puntos)
- ¿Cuál sería el la probabilidad de error de segunda especie β si la nueva proporción de consumidores fuera realmente del 22%? (3 puntos)
- Establecer un intervalo de confianza bilateral con un nivel de confianza del 95% para la proporción de compradores del nuevo producto. (2 puntos)
- Si queremos estimar la proporción de nuevos compradores con un error menor de $\pm 0,025$, un nivel de confianza del 90% y consideramos como valor previo y aproximado de P el 20% ($P \approx 0,20$), ¿a cuantos consumidores de lácteos deberíamos entrevistar para conseguir el objetivo previsto? (2 puntos)

Solución:

a) El test de hipótesis a plantear es: $H_0: P=P_0=0,20$; $H_1: P>0,20$

Región de aceptación: Aceptar H_0 si $p \leq P_0 + z_{\alpha} \sqrt{\frac{P_0(1-P_0)}{n}} = 0,20 + 1,645 \sqrt{\frac{0,20 \cdot 0,80}{150}} = 0,2537$

Como $p=50/150=0,333 > 0,2537$ deberemos rechazar H_0 y admitir que la proporción ha aumentado y que por lo tanto se puede lanzar el nuevo producto.

b) $\beta = P(\text{aceptar } H_0 / H_1) = P(p \leq 0,2537 / H_1) = P\left[N\left(P, \sqrt{P(1-P)/n}\right) \leq 0,2537 / P = 0,22\right] =$
 $= P\left[N(0,22; 0,0338) \leq 0,2537\right] = \phi\left[(0,2537 - 0,22) / 0,0338\right] = \phi(1) = 0,8413$

c) Intervalo de confianza: $p \pm z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} = 0,333 \pm 1,96 \sqrt{\frac{0,333 \cdot 0,666}{150}} \Rightarrow [0,2576; 0,4083]$

d) El tamaño de muestra para estimar la proporción P con un error absoluto menor que e y un nivel de confianza $1-\alpha$ es

$$n \geq \frac{\hat{P}(1-\hat{P}) \cdot z_{\alpha/2}^2}{e^2} = \frac{0,2 \cdot 0,8 \cdot z_{0,10/2}^2}{0,025^2} = \frac{0,16 \cdot 1,645^2}{0,025^2} = 693$$

EXAMEN DE JULIO 2008, CON SOLUCIÓN

APELLIDOS, NOMBRE: _____

Firma: _____

GRUPO: _____

BLOQUE I. CUESTIONES Y TIPO TEST (60 MINUTOS)

Todas las cuestiones deben ser debidamente justificadas. Aquellas cuestiones contestadas correctamente pero no justificadas no serán puntuadas.

1.- De una baraja de 40 cartas extraemos 6 al azar. Se pide:

- a) ¿Cuál es la probabilidad de tener en esas 6 cartas 3oros y 2 copas? (1 punto)
- b) ¿Y si cada carta la hubiéramos puesto en el montón y vuelto a barajar antes de extraer la siguiente carta? (1 punto)

2.- El voltaje medido en una red eléctrica tiene una distribución normal con media 220 voltios y desviación típica de 4 voltios. Si se toman tres medidas independientes de voltaje:

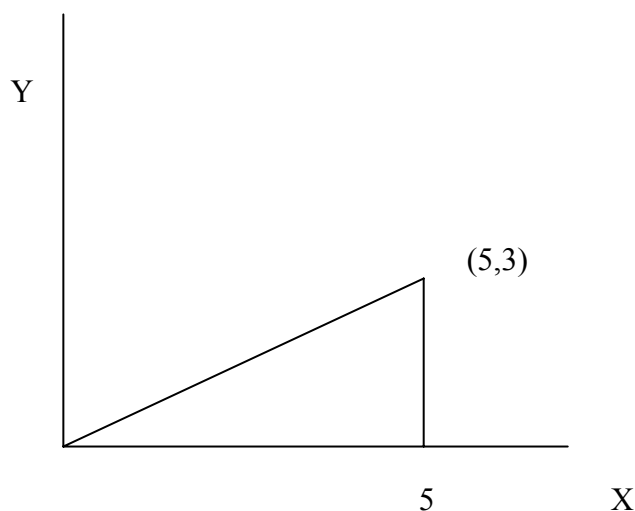
- a) ¿Cuál es la probabilidad de que las tres estén entre 208 y 216 voltios? (1 punto)
- b) ¿Y de que la media de las tres medidas sea superior a 210 voltios? (1 punto)

3.- Por información estadística de años anteriores se conoce que en cierta ciudad y durante el mes de julio la variable aleatoria $X = \text{número de cortes parciales de suministro eléctrico producidos en dicho mes}$ sigue la distribución de probabilidades:

$$P(X = k) = \frac{e^{-10} \cdot 10^k}{k!} \quad k = 0, 1, 2, \dots$$

- a) Obtener la probabilidad de que determinado mes se produzcan más de 2 cortes. (1 punto)
- b) Obtener la probabilidad de que se hayan producido exactamente 2 cortes en un mes sabiendo que se han producido menos de 4. (1 punto)

4.- La variable aleatoria bidimensional (X,Y) se distribuye uniformemente en el recinto que se indica en la figura.



Se pide:

- Calcular la función de distribución para $X=4$ e $Y=2$. (1 punto)
- Calcular la función de densidad marginal de y . (0,5 puntos)
- Calcular la función de densidad condicional de x cuando y vale 2. (0,5 puntos)

5.- Admitamos que el tiempo de funcionamiento hasta el fallo de unos relés obedece a una distribución exponencial. Para decidir sobre la duración media de los mismos planteamos las siguientes hipótesis:

$$H_0: \text{media} \geq 200 \text{ horas}$$

$$H_1: \text{media} < 200 \text{ horas}$$

Y decidimos con la siguiente regla:

- Tomamos dos relés y si los dos fallan antes de 150 horas rechazamos H_0 .
- En cualquier otro caso aceptamos H_0 .

Admitiendo que la duración de cada relé es independiente de la duración de los demás, se pide:

- Calcular el valor de α de la prueba (1 punto)
- Calcular el valor de β si la media fuera realmente 190 horas (1 punto)

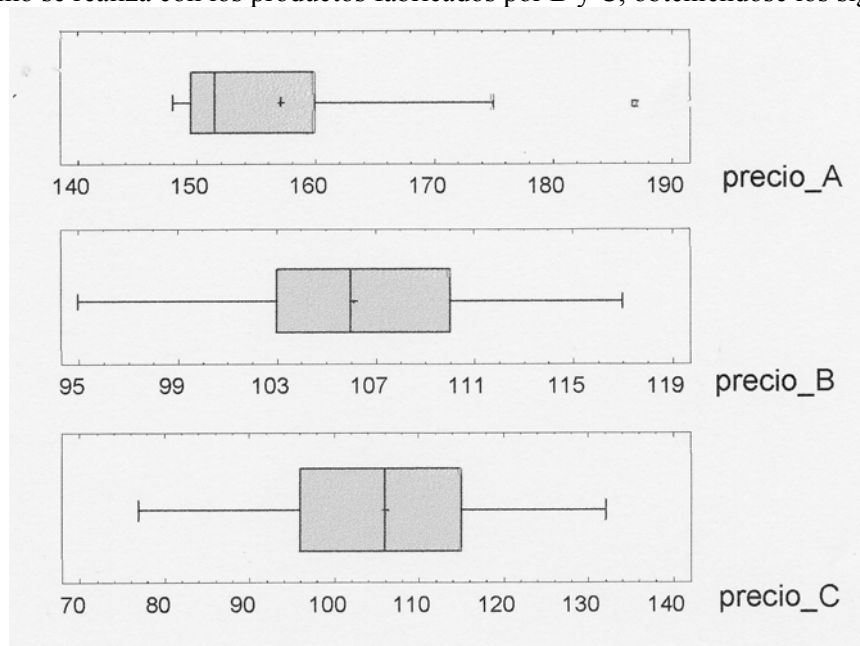
APELLIDOS, NOMBRE: _____

Firma: _____

GRUPO: _____

BLOQUE II. PRÁCTICAS DE ORDENADOR (30 MINUTOS)

1.- Tres empresas (A, B y C) fabrican un producto de características muy similares que se vende en pequeños comercios. El gerente de la empresa A desea saber si su producto se vende a un precio significativamente más caro o más barato que sus competidores. Para averiguarlo, se seleccionan al azar 15 comercios, se obtiene el precio del producto fabricado por A y con esos datos se realiza un diagrama box-whisker. Lo mismo se realiza con los productos fabricados por B y C, obteniéndose los siguientes gráficos:



A la vista de estos diagramas, responder a las siguientes preguntas justificando convenientemente las respuestas.

- ¿El valor del primer cuartil de A es menor que el del tercer cuartil de C? (2 puntos)
- ¿Cuál de los tres precios tiene mayor dispersión? (3 puntos)
- ¿En cuál de los tres casos es mayor el coeficiente de asimetría? (2 puntos)
- ¿Crees que hay evidencia suficiente para afirmar si el precio de alguna de las tres empresas es significativamente más barato o caro que las demás? (3 puntos)

2.- Una industria elabora piezas cerámicas por medio de un proceso de cocción en continuo. Uno de los parámetros de calidad de dichas piezas es la resistencia a la rotura. La industria desea aumentar la resistencia media de las piezas, pero los ingenieros desconocen cuáles son las variables del proceso de fabricación cuyo efecto en la calidad final es estadísticamente significativo (considerando $\alpha=0,05$). Para ello, se recopilan los datos de resistencia correspondientes a 40 ensayos de calidad realizados en los últimos meses, resultando una resistencia media de 74,675 y una cuasivarianza de 181,763. Cada uno de estos 40 datos corresponde a condiciones de producción ligeramente distintas, y se conoce el valor de las siguientes variables:

- tiempo de cocción (variable "tiempo", medida en minutos)
- temperatura de cocción (variable "temperatura", medida en °C)
- cantidad de esmalte (variable "esmalte", medida en gr/m^2)
- humedad del producto a la entrada del horno (variable "humedad", medida en gr/m^3)

Con estos datos se realiza un análisis de regresión lineal múltiple, cuyos resultados se muestran a continuación

Multiple Regression Analysis					
Dependent variable: resistencia					
Parameter	Estimate	Standard Error	T Statistic	P-Value	
CONSTANT	30,373	55,4213	0,548038	0,5871	
tiempo	0,0688825	0,0626572	1,09935	0,2791	
temperatura	-0,0148774	0,0101842	-1,46084	0,1530	
esmalte	0,73615	0,786056	0,936511	0,3554	
humedad	0,103824	0,159121	0,652486	0,5184	
Analysis of Variance					
Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	803,49	4	200,872	1,12	0,3634
Residual	6285,29	35	179,58		
Total (Corr.)	7088,77	39			
R-squared = percent					
Standard Error of Est. = 13,4007					
Mean absolute error = 10,1928					
Durbin-Watson statistic = 2,39409 (P=0,0916)					
Lag 1 residual autocorrelation = -0,205561					

A la vista de estos resultados, responder a las siguientes preguntas justificando convenientemente las respuestas:

- a) ¿Qué modelo matemático recomendarías a los ingenieros de proceso para predecir la resistencia de las piezas cerámicas? (4 puntos)
- b) Calcular el valor del coeficiente de determinación R^2 . ¿Cómo se interpreta en la práctica este parámetro? (3 puntos)
- c) Se considera que un lote de piezas es de mala calidad si se toma una muestra y su resistencia es inferior a 60 unidades. Calcular la probabilidad de que un lote sea de mala calidad si se ha fabricado en las siguientes condiciones: tiempo = 300 min, temperatura = 2000 °C, esmalte = 30 gr/m^2 , humedad = 330 gr/m^3 . (3 puntos)

3.- Una empresa que se dedica a la fabricación de perfumes está desarrollando una nueva fragancia que pretende sacar al mercado en gran escala. La empresa no está segura de si dicho perfume debería destinarse a hombres o a mujeres. Para decidir la mejor opción, realiza una encuesta a 100 hombres y a 100 mujeres. Cada uno de los 200 encuestados huele el perfume y elige una de las siguientes opciones:

- Opción A: el perfume debería ser destinado a hombres
- Opción B: el perfume debería ser destinado a mujeres
- Opción C: no está claro a qué sexo debería destinarse el perfume.

Los resultados se muestran en la siguiente tabla, indicándose las respuestas de cada encuestado así como las frecuencias esperadas (en cursiva).

Frequency Table				
	Opc A(hombr)	Opc B (mujer)	Opc C (??)	Row Total
encuestado_hombre	18	70	12	100
	<i>20,50</i>	<i>69,50</i>	<i>10,00</i>	<i>50,00%</i>
encuestado_mujer	23	69	8	100
	<i>20,50</i>	<i>69,50</i>	<i>10,00</i>	<i>50,00%</i>
Column Total	41	139	20	200
	<i>20,50%</i>	<i>69,50%</i>	<i>10,00%</i>	<i>100,00%</i>

Cell contents:
 Observed frequency
Expected frequency

Responder a las siguientes preguntas, justificando convenientemente la respuesta:

- a) Calcular el valor del estadístico de contraste asociado al test chi-cuadrado. (3 puntos)
- b) Calcular el p-valor asociado al contraste de hipótesis. (2 puntos)
- c) Teniendo en cuenta el p-valor y considerando un riesgo de primera especie del 5%, ¿cuál es la conclusión del test? (2 puntos)
- d) Teniendo en cuenta que el objetivo de la empresa es determinar si el perfume debería destinarse a hombres o a mujeres, ¿qué sugerirías a la vista de los resultados de la encuesta? (3 puntos)

BLOQUE III. PROBLEMAS (90 MINUTOS)

Cada problema son 10 puntos. La nota del bloque es la media de las notas de los problemas.

1.- Una empresa de componentes electrónicos fabrica un determinado módulo con un 6% de defectuosos, poniéndolos a la venta en paquetes de 6 unidades. Una empresa competidora vende imitaciones indistinguibles con un porcentaje de defectuosos del 40%, y los comercializa en el mismo envase de 6 unidades. Teniendo en cuenta que el 10% de los módulos que se venden en el mercado son de la empresa competidora:

- ¿Cuál es la probabilidad de que un paquete que contiene 2 módulos defectuosos sea de la empresa competidora? (5 puntos)
- ¿Cuál es la probabilidad de encontrar en el mercado un paquete que sea defectuoso por tener al menos un módulo defectuoso? (3 puntos)
- Si compramos 300 paquetes, ¿cuántos de ellos cabe esperar que no sean defectuosos? (2 puntos)

2.- Para las viviendas usadas de la ciudad de Valencia, la distancia al centro y el precio por m² configuran una variable con distribución normal bidimensional de parámetros

$$(X, Y) = N \left[\mu = \begin{pmatrix} 4 \\ 3000 \end{pmatrix}, V = \begin{bmatrix} 1,70 & \sigma_{1,2} \\ \sigma_{2,1} & 1 \cdot 10^6 \end{bmatrix} \right] \text{ y } \rho = -0,90 \quad \text{Expresados X en Km, Y en €/m}^2$$

- Calcular la proporción de viviendas usadas en la ciudad de Valencia con un valor inferior a 2000 €/m² (3 puntos)
- Calcular la proporción de viviendas que, costando 3000 €/m² están a menos de 3 Km del centro (3 puntos)
- ¿A cuántos Kms del centro nos tendremos que alejar para que el 75% de las viviendas cuesten menos de 2000 €/m²? (4 puntos)

3.- Los gestores del metro de Valencia están realizando estudios para mejorar el servicio que presta a sus usuarios este medio de transporte. Habitualmente, el valor medio del retraso sobre la hora inicialmente prevista en las paradas de la línea 1 es de 15 segundos, con una desviación típica de 10 segundos. Dado que estamos en verano, los técnicos creen que puede haber variado el tiempo medio de retraso. Para verificar este supuesto controlan el retraso en 20 llegadas en estaciones de la línea 1, obteniendo los siguientes valores expresados en segundos.

(10, 0, 3, -2, -4, 14, 20, 4, 30, 9, 3, 3, 6, 23, -10, 21, 2, 5, 23, -10)

Admitiendo que la variable X= al retraso del convoy sobre la hora inicialmente prevista, es una v.a. con distribución normal, se pide:

- Obtener un intervalo de confianza con un nivel de confianza del 90% para el tiempo medio de retraso de esa línea, utilizando únicamente la información dada por la muestra. (2 puntos)
- Con un nivel de significación $\alpha = 0,01$ plantear y contrastar el test apropiado que permita responder a la duda de los técnicos del metro. (4 puntos)
- Si en realidad el tiempo medio de retraso se hubiese reducido en 5 segundos ¿qué probabilidad habría de concluir, con la muestra obtenida y el test aplicado, que no se ha producido un cambio en el tiempo medio de retraso? (2 puntos)
- ¿En cuántas llegadas habría que controlar el retraso, para estimar el tiempo medio de retraso a partir de la media muestral con un error menor de 2 segundos y un nivel de confianza del 95%? (2 puntos)

SOLUCIONES AL EXAMEN**BLOQUE I****SOLUCIÓN CUESTIÓN 1)**

Sin reemplazamiento la probabilidad es:

$$P(3 \text{ oros}, 2 \text{ copas}) = \frac{\binom{10}{3} \binom{10}{2} \binom{20}{1}}{\binom{40}{6}} = 0,0281$$

Con reemplazamiento la probabilidad es:

$$P(3 \text{ oros}, 2 \text{ copas}) = \frac{6!}{3! 2! 1!} \left(\frac{1}{4}\right)^3 \left(\frac{1}{4}\right)^2 \left(\frac{1}{2}\right)^1 = 0,0293$$

SOLUCIÓN CUESTIÓN 2):

$$P(208 \leq X \leq 216) = \Phi\left(\frac{216 - 220}{4}\right) - \Phi\left(\frac{208 - 220}{4}\right) = \Phi(-1) - \Phi(-3) = 0,1587 - 0,0013 = 0,1574$$

$$a) p = P(X_1 \cap X_2 \cap X_3) = 0,1574^3 = 0,0039$$

$$b) x \equiv N(220; \frac{4}{\sqrt{3}}) \quad P(\bar{x} \geq 210) = 1 - \Phi\left(\frac{210 - 220}{4/\sqrt{3}}\right) = 1 - 0 = 1$$

SOLUCIÓN CUESTIÓN 3):

a)

$$P(X > 2) = 1 - P(X \leq 2) = 1 - P(Ps(10) \leq 2) = 1 - 0,0028 = 0,9972$$

b)

$$P(X = 2 / X < 4) = \frac{P(X = 2 \cap X < 4)}{P(X \leq 3)} = \frac{P(X = 2)}{P(X \leq 3)} = \frac{0,0028 - 0,0005}{0,0103} = 0,223$$

SOLUCIÓN CUESTIÓN 4):

$$\text{La recta es } y = \frac{3}{5}x$$

La función de densidad conjunta es

$$f(x, y) = \frac{1}{\text{Área}} = \frac{1}{5 \cdot 3/2} = \frac{2}{15}$$

a) La función de distribución en el punto (4, 2) es,

$$F(4, 2) = \int_0^2 dy \int_{5y/3}^4 \frac{2}{15} dx = \int_0^2 \frac{2}{15} (4 - \frac{5}{3}y) dy = \int_0^2 (\frac{8}{15} - \frac{10}{45}y) dy = \left[\frac{8}{15}y - \frac{10}{45 \cdot 2}y^2 \right]_0^2 = \frac{16}{15} - \frac{10 \cdot 4}{45 \cdot 2} = 0,6222$$

b) la función de densidad marginal de y es,

$$f(y) = \int f(x, y) dx = \int_{5y/3}^5 \frac{2}{15} dx = \frac{2}{15} (5 - \frac{5}{3}y) = \frac{10}{15} - \frac{10}{45}y$$

c) Para $y=2$

$$f(x/y=2) = \frac{f(x, y)}{f(y=2)} = \frac{2/15}{(\frac{10}{15} - \frac{10}{45} \cdot 2)} = \frac{2/15}{0,222} = 0,60$$

SOLUCIÓN CUESTIÓN 5):

Recordando que una variable exponencial de media m es una $EXP(1/m)$ y que $P(X \leq x) = 1 - e^{-x/m}$ tendremos:

$$\alpha = P(\text{rechazar } H_0/H_0 \text{ cierta}) = P(\text{fallo de los dos relés}/H_0) =$$

$$P(\text{fallo de los dos relés si } X = EXP(1/200)) =$$

$$P(EXP(1/200) \leq 150) \cdot P(EXP(1/200) \leq 150) = (P(EXP(1/200) \leq 150))^2 =$$

$$(1 - e^{-150/200})^2 = 0,278$$

$$\beta = P(\text{aceptar } H_0/H_1 \text{ cierta}) = 1 - P(\text{rechazar } H_0/H_1 \text{ cierta}) =$$

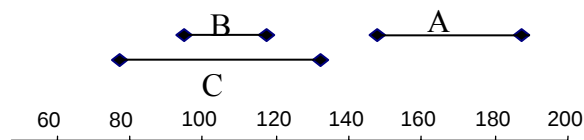
$$1 - P(\text{fallo de los dos relés si } X = EXP(1/190)) =$$

$$1 - (P(EXP(1/190) \leq 150))^2 =$$

$$1 - (1 - e^{-150/190})^2 = 0,7019$$

BLOQUE II**SOLUCIÓN CUESTIÓN 1):**

- a) No, ya que el primer cuartil de A vale 149,5 (extremo izquierdo de la caja), que es superior al tercer cuartil de C que vale 115 (extremo derecho de la caja).
- b) El intervalo intercuartílico es un buen indicador del grado de dispersión:
 $IIC_A = 160 - 149,5 = 10,5$; $IIC_B = 110 - 103 = 7$; $IIC_C = 115 - 96 = 19$
 Por tanto, el precio de C tiene mayor dispersión que los otros dos. Se llega a la misma conclusión calculando los rangos ($R_A = 39$; $R_B = 22$; $R_C = 55$).
- c) La forma del diagrama en el caso de A indica una distribución asimétrica positiva (bigote derecho más largo que el izquierdo, mediana desplazada hacia la izquierda), correspondiendo por tanto un coeficiente de asimetría positivo. En cambio en los otros dos casos el coeficiente de asimetría será próximo a cero, pues la forma del diagrama es casi simétrica. Por tanto, el coeficiente de asimetría es mayor en el caso A.
- d) Realizando un sencillo esquema se observa que los datos de B se superponen con los de C. En cambio, ninguno de los 15 datos de A está dentro del intervalo de variación de B ni de C. Por tanto, es razonable suponer que el precio de A será significativamente mayor que los otros dos, aunque para verificar esta suposición sería necesario realizar el correspondiente test de hipótesis.

**SOLUCIÓN CUESTIÓN 2):**

- a) A partir de los datos podemos asumir que la distribución marginal de la resistencia es una normal de media 74,675 y desviación típica $\sqrt{181,763} = 13,48$. Ninguno de los coeficientes del modelo es estadísticamente significativo ya que todos ellos tienen un $p\text{-value} > 0,05$. Por esta razón, la desviación típica residual (standard error of estimation) vale 13,41, que prácticamente coincide con la desviación típica marginal de la resistencia. Por tanto, el modelo matemático que se recomendaría para predecir la resistencia sería una Normal (74,675 ; 13,48).

$$b) R^2 = 1 - \frac{s_{res}^2}{s_Y^2} = 1 - \frac{13,4007^2}{181,763} = 0,012$$

R^2 también se define como el cociente entre la suma de cuadrados del modelo y la suma de cuadrados total. En nuestro ejemplo es,

$$R^2 = \frac{803,49}{7088,47} = 0,0134$$

Interpretación práctica: el 1,2% de la varianza de la variable Y (resistencia) está explicada por el modelo de regresión lineal múltiple con las 4 variables. Este valor es prácticamente igual a cero porque ninguna de las variables del modelo tiene capacidad estadísticamente significativa de predecir la resistencia.

- c) Dado que ninguna variable tiene un efecto estadísticamente significativo, el modelo a utilizar es el indicado en el apartado a):

$$P(R < 60) = P[N(74,675; 13,48) < 60] = P[N(0;1) < (60 - 74,675)/13,48] = P[N(0;1) < -1,09] = 0,14$$

SOLUCIÓN CUESTIÓN 3):

a)

$$D = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{(18 - 20,5)^2}{20,5} + \frac{(70 - 69,5)^2}{69,5} + \frac{(12 - 10)^2}{10} + \frac{(23 - 20,5)^2}{20,5} + \frac{(69 - 69,5)^2}{69,5} + \frac{(8 - 10)^2}{10} = 1,42$$

$$b) D \approx \chi^2_{(h-1)(k-1)} \approx \chi^2_{(3-1)(2-1)} \approx \chi^2_2$$

Son dos grados de libertad: $(3 \text{ columnas} - 1) \cdot (2 \text{ filas} - 1) = 2$

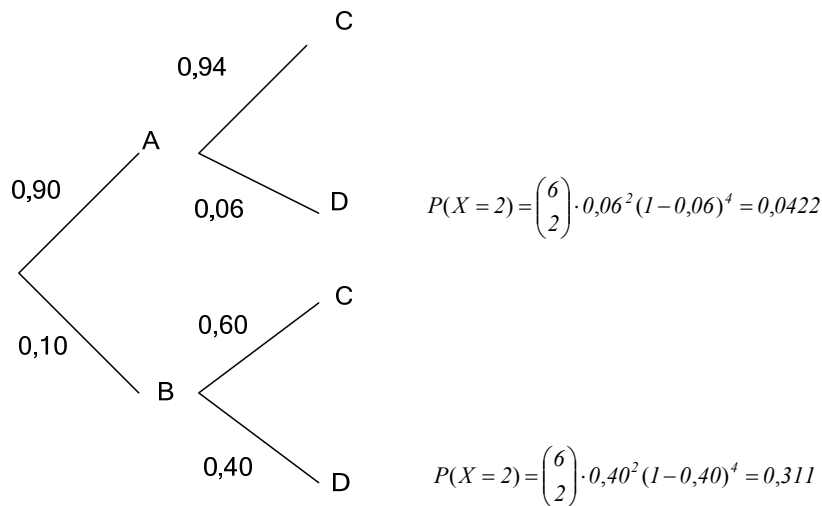
$p\text{-valor} = P(\chi^2_2 > 1,42) \approx 0,5$ (leyendo en la tabla de la chi-cuadrado)

- c) Como $p\text{-valor} > 0,05$ se acepta la hipótesis nula de independencia. Es decir, el hecho de que el encuestado sea hombre o mujer no influye significativamente en que el perfume sea percibido como masculino o femenino.
- d) Teniendo en cuenta que el test anterior no resulta significativo, pueden juntarse los datos correspondientes a hombres y mujeres. Descartando, por simplificar, las respuestas ambiguas, resulta la siguiente tabla. Si se realiza un nuevo test chi-cuadrado, se obtiene $D = (41-90)^2/90 + (139-90)^2/90 = 53,3$ que es muy superior al valor crítico de una chi-cuadrado con un grado de libertad ($\chi^2_1(\alpha=0.01) = 6,6$). Dado que las diferencias son estadísticamente significativas y el perfume es percibido con mayor frecuencia como femenino, debería destinarse a mujeres.

	Opc. A	Opc. B	(total)
Frec. observada	41	139	180
Frec. esperada	90	90	180

BLOQUE III**SOLUCIÓN PROBLEMA 1):**

El árbol de probabilidades es:



$$a) P(B/x=2) = \frac{0,10 \cdot 0,311}{0,90 \cdot 0,0422 + 0,10 \cdot 0,311} = 0,4502$$

b) La probabilidad de que al menos haya uno defectuoso es:

$$P(X \geq 1) = 1 - P(X = 0)$$

$$P(X \geq 1 / A) = 1 - 0,94^6 = 0,3101$$

$$P(X \geq 1 / B) = 1 - 0,60^6 = 0,9533$$

$$P(X \geq 1) = 0,90 \cdot 0,3101 + 0,10 \cdot 0,9533 = 0,374$$

c) El valor medio del número de paquetes correctos es:

$$E(\text{correctos}) = np = 300 \cdot (1 - 0,374) = 187,8$$

SOLUCIÓN PROBLEMA 2):

a) La variable valor de las viviendas usadas es:

$$Y = N(3000, \sqrt{1 \cdot 10^6}) = N(3000, 1000)$$

$$P(Y < 2000) = \Phi\left(\frac{2000 - 3000}{1000}\right) = \Phi(-1) = 0,1587 = 15,87\%$$

b) La distribución de la distancia X cuando el precio $Y=3000$, es

$$(X / y = 3000) = N\left(4 - 0,9 \frac{\sqrt{1,70}}{1000} (3000 - 3000), (1,7(1 - 0,90^2))^{1/2}\right) = N(4; 0,57)$$

Y la probabilidad pedida

$$P((X / y = 3000) \leq 3) = \Phi\left(\frac{3 - 4}{0,57}\right) = 0,0392 = 3,92\%$$

c) La variable precio condicionado por la distancia es

$$Y/x_0 = N(3000 - 0,90 \frac{1000}{\sqrt{1,7}}(x_0 - 4), (10^6(1 - 0,90^2))^{1/2}) = N(5761,07 - 690,27 x_0, 435,88) =$$

$$= N(m, 435,88)$$

Por el enunciado se deberá cumplir

$$P(Y/x_0 \leq 2000) = 0,75 \text{ es decir}$$

$$\Phi\left(\frac{2000 - m}{435,88}\right) = 0,75 \Rightarrow \text{tomando inversas } \frac{2000 - m}{435,88} = 0,675$$

despejando $m=1705$ e igualando

$$1705,77 = 5761,07 - 690,27 x_0 \text{ de donde } x_0 = 5,87 \text{ Km}$$

SOLUCIÓN PROBLEMA 3):

a) El intervalo de confianza para la media de una población normal con varianza desconocida es

$$\bar{x} \pm t_{n-1}^{\alpha/2} \frac{S'}{\sqrt{n}}$$

$$7,5 \pm 1,729 \frac{11,19}{\sqrt{20}} \rightarrow [3,17 ; 11,83]$$

b) Como no se especifica el cambio posible el test será bilateral

$$H_0: m = m_0 = 15$$

$$H_1: m \neq 15$$

La regla de decisión, con $\sigma=10$ conocida es, aceptar H_0 si:

$$\bar{x} \in \left[m_0 \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right] = \left[15 \pm 2,57 \frac{10}{\sqrt{20}} \right] = [9,24 ; 20,76]$$

Como $\bar{x} = 7,5 \notin [9,24 ; 20,76]$ debemos rechazar H_0 y admitir que ha habido cambios en la media de los retrasos con un nivel de significación del 1%

c) $\beta = P(\text{aceptar } H_0/H_1) =$

$$P(\bar{x} \in [9,24 ; 20,76] / m = 10) = P(N(10, \frac{10}{\sqrt{20}}) \in [9,24 ; 20,76]) =$$

$$= \Phi\left(\frac{20,76 - 10}{10/\sqrt{20}}\right) - \Phi\left(\frac{9,24 - 10}{10/\sqrt{20}}\right) = 1 - 0,367 = 0,633$$

d) Según las expresiones del cálculo del tamaño muestral

$$n \geq \left(\frac{\sigma z_{\alpha/2}}{e}\right)^2 = \left(\frac{10 \cdot 1,96}{2}\right)^2 = 96,04 \Rightarrow 97 \text{ llegadas}$$

EXAMEN DE ENERO 2008, CON SOLUCIÓN

APELLIDOS, NOMBRE: _____

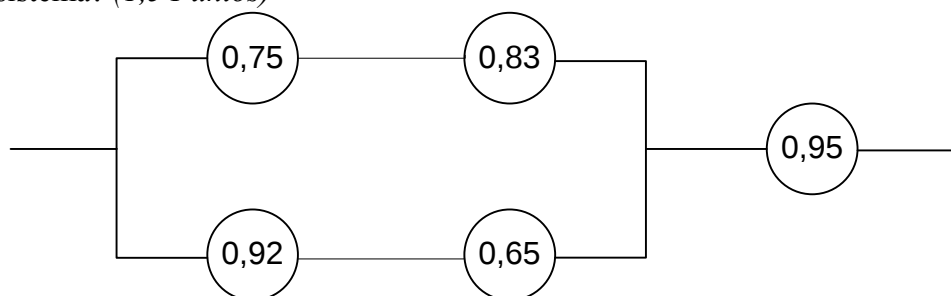
Firma: _____

GRUPO: _____

BLOQUE I. CUESTIONES Y TIPO TEST (60 MINUTOS)

Todas las cuestiones deben ser debidamente justificadas. Aquellas cuestiones contestadas correctamente pero no justificadas no serán puntuadas.

1.- En el siguiente esquema de fiabilidad, en el cual se indica el esquema de montaje de distintos elementos electrónicos y la probabilidad de no fallo de cada elemento, ¿cuál es la probabilidad de que falle el sistema? (1,5 Puntos)



2.- La duración en minutos de una llamada telefónica de larga distancia, se asimila a una variable aleatoria X cuya función de distribución es:

$$F(X) = \begin{cases} 0 & \text{para } x < 0 \\ 1 - \frac{2}{3}e^{-2x/3} - \frac{1}{3}e^{-x/3} & \text{para } x \geq 0 \end{cases}$$

Determinar:

- Función de densidad de probabilidad. (0,5 puntos)
- Probabilidad de que la duración de una llamada esté comprendida entre 3 y 6 minutos, (0,5 puntos)
- Una llamada lleva 3 minutos, probabilidad de que no pase de los 6 minutos. (0,5 puntos)

3.- Al corregir la primera edición de un libro el autor revisó las 111 primeras páginas y encontró la siguiente distribución de erratas por página

Nº de erratas:	0	1	2	3	4
Frecuencia:	72	27	7	4	1

Asumiendo que esta distribución sigue un modelo de Poisson, ¿cuál es la probabilidad de que el autor encuentre más de un error en una página elegida al azar en el libro? (1,5 puntos)

4.- En el proceso de inspección de unos tornillos se inspeccionan lotes de 50 unidades, se miden las longitudes de cada uno y si la longitud media está fuera del intervalo $[2,05 ; 2,2]$ cm se rechaza el lote. Si la longitud (cm) es una variable aleatoria con distribución normal $N[2,1 ; 0,083]$ y los tornillos se fabrican de tal modo que su longitud es independiente en cada uno, ¿cuál es la probabilidad de rechazar un lote? (1,5 puntos)

5.- Para comprobar si dos líneas de fabricación, que producen un mismo producto, generan la misma proporción de piezas defectuosas, tomaremos dos muestras de tamaño n_A de la línea A; para cada una de las muestras determinaremos la proporción muestral de piezas defectuosas y a continuación calcularemos la media: $(\hat{p}_{A1} + \hat{p}_{A2})/2$. A continuación repetiremos la misma operación en la línea B, tomando dos muestras de tamaño n_B , calculando la proporción muestral de piezas defectuosas y la media: $(\hat{p}_{B1} + \hat{p}_{B2})/2$. Si tomamos como estadístico:

$\left(\frac{\hat{p}_{A1} + \hat{p}_{A2}}{2} - \frac{\hat{p}_{B1} + \hat{p}_{B2}}{2} \right)$, ¿cuál será la distribución del estadístico obtenido? Razonar la respuesta. (1,5 puntos)

NOTA: la proporción de piezas defectuosas, en cada una de las dos muestras, se distribuye como

$$\hat{p} \equiv N(p; \sqrt{\frac{pq}{n}})$$

6.- En una serie de ensayos llevados a cabo para conocer la resistencia media de unas probetas de hormigón fabricado con unas especificaciones concretas hemos obtenido:

5 probetas con resistencia entre 340 y 345 Kg/cm²

12 probetas con resistencia entre 345 y 350 Kg/cm²

10 probetas entre 350 y 360 Kg/cm²

(Tomar como valor representativo de cada intervalo su valor central)

Asumiendo que la resistencia de las probetas sigue una distribución normal:

a) Obtener un intervalo de confianza para la media poblacional con un nivel de confianza del 99% (1,5 puntos)

b) ¿Cuántas probetas tendríamos que comprobar si queremos estimar la media poblacional mediante la media muestral de n probetas cometiendo un error absoluto menor que 5 Kg/cm² y un nivel de confianza del 95%, si asumimos que la dispersión es $\sigma=5$ Kg/cm²? (1 punto)

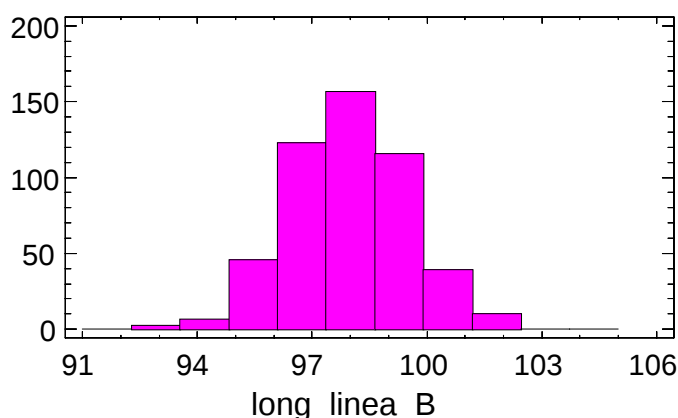
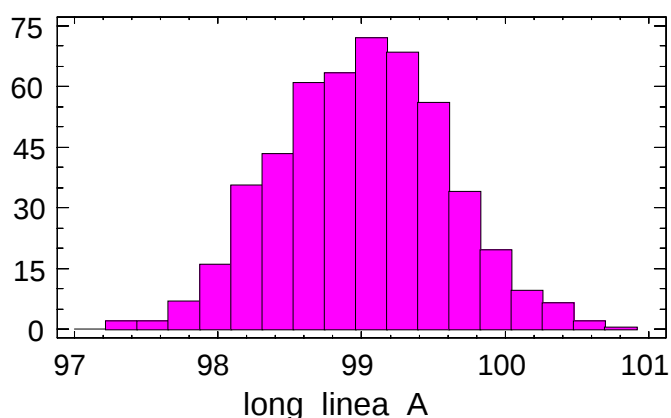
APELLIDOS, NOMBRE: _____

Firma: _____

GRUPO: _____

BLOQUE II. PRÁCTICAS DE ORDENADOR (30 MINUTOS)

1.- Una industria mecánica elabora piezas metálicas por medio de dos líneas de producción. El parámetro de calidad crítico de estas piezas es la longitud, la cual se considera correcta si está comprendida entre 98 y 102 mm. Los ingenieros han detectado una excesiva proporción de piezas fuera de estos límites de tolerancia, pero desconocen qué línea es la principal responsable de este problema. Con el objetivo de investigar esta cuestión, se toman 500 piezas elaboradas en la línea A y otras 500 obtenidas de la línea B. Tras medir la longitud de estas piezas, se obtienen los siguientes histogramas:



A la vista de estos histogramas, responder a las siguientes preguntas justificando convenientemente las respuestas.

- ¿Qué indica la escala vertical? ¿Por qué es tan diferente en los dos casos? (2 puntos)
- ¿Cuál de los siguientes valores crees que es la cuasivarianza muestral de la longitud de las piezas de la línea A? (2 puntos)
0,035
0,352
3,517
35,17
- ¿Cuál de las dos líneas tiene mayores parámetros de dispersión? (2 puntos)
- ¿Crees que la técnica utilizada es adecuada para detectar datos anómalos? ¿Qué otras técnicas utilizarías? (2 puntos)
- ¿Qué recomendarías a los técnicos del proceso para disminuir la proporción de piezas fuera de tolerancia? (2 puntos)

2.- La materia activa de un determinado medicamento se obtiene por fermentación con microorganismos modificados genéticamente. La concentración de la materia al terminar la fermentación (mg/l) es un índice del rendimiento del proceso. Con el objetivo de determinar qué variables son las que afectan al rendimiento, se recopila información de 30 lotes de fermentación obtenidos en el último mes. De cada uno de ellos se dispone de los siguientes datos: temperatura media (variable “temperatura” medida en °C), pH medio (variable “pH”), concentración inicial de azúcares (variable “azúcar”) y concentración inicial de proteínas (variable “proteína”), ambas medidas en gramos/litro. Con estos datos se realiza un análisis de regresión lineal múltiple, cuyos resultados se muestran a continuación.

Multiple Regression Analysis

Dependent variable: rendimiento

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	156,827	36,5803	4,28722	0,0002
temperatura	2,73502	0,709492	3,85489	0,0007
pH	-27,1323	4,14406	-6,54728	0,0000
azúcar	1,91988	0,212637	9,02891	0,0000
proteína	3,22501	0,76551	4,2129	0,0003

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	5372,01	4	1343,0	44,64	0,0000
Residual	752,181	25	30,0873		
Total (Corr.)	6124,2	29			

R-squared = 87,7179 percent

R-squared (adjusted for d.f.) = 85,7527 percent

Standard Error of Est. = 5,48519

Mean absolute error = 4,13418

Durbin-Watson statistic = 1,68767 (P=0,2086)

Lag 1 residual autocorrelation = 0,140306

A la vista de estos resultados, responder a las siguientes preguntas:

- Escribir la ecuación del modelo que se debería utilizar para predecir el rendimiento obtenido al finalizar la fermentación en función de las variables que ejercen un efecto estadísticamente significativo. (5 puntos)
Nota: justificar convenientemente cuáles son las variables con un efecto estadísticamente significativo, considerando un riesgo de primera especie del 5%.
- Interpretar qué significado práctico tiene el valor 156,827 y 2,73502 que aparecen en la columna Estimate. (5 puntos)

3.- Una industria necesita adquirir ciertos componentes electrónicos. En el mercado existen tres proveedores que ofrecen el mismo tipo de componente. El ingeniero de control de calidad no está seguro de qué proveedor ofrece mayor calidad y convence a su jefe para que se realice un ensayo. Para ello, se compran 500 componentes de cada proveedor y se determina cuántos de ellos son defectuosos. Las unidades defectuosas pueden clasificarse en dos categorías, según sea el tipo de defecto (defecto tipo 1 o tipo 2). Los resultados del ensayo se muestran en la siguiente tabla, indicándose el número de componentes defectuosos y correctos de cada distribuidor, así como las frecuencias esperadas (en cursiva). A la vista de los resultados, la proporción de unidades defectuosas es menor en el proveedor 2, y por ello el jefe concluye que éste es el que más interesa. El ingeniero quiere verificar si la hipótesis de su jefe es correcta, y decide realizar un análisis estadístico de los datos.

Frequency Table

	def_tipo1	def_tipo2	correctos	Row Total
proveedor 1	7	21	472	500
	<i>10,00</i>	<i>17,33</i>	<i>472,67</i>	<i>33,33%</i>
	0,90	0,78	0,00	
proveedor 2	9	11	480	500
	<i>10,00</i>	<i>17,33</i>	<i>472,67</i>	<i>33,33%</i>
	0,10	2,31	0,11	
proveedor 3	14	20	466	500
	<i>10,00</i>	<i>17,33</i>	<i>472,67</i>	<i>33,33%</i>
	1,60	0,41	0,09	
Column Total	30 2,00%	52 3,47%	1418 94,53%	1500 100,00%

Cell contents:

Observed frequency

Expected frequency

Contribution to chi-squared

Responder a las siguientes preguntas, justificando convenientemente la respuesta:

- ¿Cuál es la hipótesis nula del contraste de hipótesis que debería formular el ingeniero? ¿Cuál es la hipótesis alternativa? (2,5 puntos)
- Calcular el valor del estadístico de contraste asociado a dicho test de hipótesis. (2,5 puntos)
- ¿Cuál sería la distribución teórica de dicho estadístico de contraste si la hipótesis nula fuera cierta? (2,5 puntos)
- ¿Cuál es la conclusión obtenida del análisis estadístico de los datos? (considerar $\alpha=0.05$) (2,5 puntos)

APELLIDOS, NOMBRE: _____

Firma: _____

GRUPO: _____

BLOQUE III. PROBLEMAS (90 MINUTOS)**Cada problema son 10 puntos. La nota del bloque es la media de las notas de los problemas.**

1.- Un fabricante de un determinado tipo de motor eléctrico dispone de 3 proveedores para una de sus piezas mecánicas. Se sabe que el proveedor A tiene un 5% de piezas defectuosas, el B un 3% y el C un 1%. Un 70% de las piezas corresponde al proveedor A y un 20% al B. Se pide:

- Para un lote de piezas recibidas, de las cuales no se sabe a qué proveedor corresponden, se toman 4 unidades y 1 de ellas resulta ser defectuosa. Calcular la probabilidad de que el lote de piezas corresponda al proveedor B. (4 puntos)
- Si en una semana fabricamos 150 motores, con piezas de los tres proveedores, ¿cuántos motores serán defectuosos por término medio debido a dicha pieza? (3 puntos)
- La pieza en cuestión se monta en cada uno de los motores que salen a la venta. Cuando la pieza es correcta, el beneficio por la venta del motor es de 50 €, pero cuando la pieza es defectuosa, la pérdida es de 60 €. Calcular el valor medio del beneficio obtenido en una semana por la venta de 150 motores. (3 puntos)

2. La función de distribución conjunta de una variable aleatoria bidimensional continua es

$$F_{XY}(x, y) = (1 - e^{-2x}) \cdot (1 - e^{-y}) \quad \text{con } 0 < x < +\infty, 0 < y < +\infty$$

Obtener:

- Función de densidad conjunta (2 puntos)
- Funciones de distribución marginales. (3 puntos)
- ¿Son X e Y variables aleatorias independientes? (2 puntos)
- Calcular las probabilidades $P(X < 1, Y < 2)$, $P(X < 1/Y < 2)$ y $P(X < 1/Y = 2)$ (3 puntos)

3.- La duración de las llamadas telefónicas realizadas a través de una operadora de telefonía móvil podemos asumir que tiene una distribución normal en la que la media habitual es de 42 segundos. Ante una reciente subida de precios, el departamento comercial quiere saber si esta subida ha tenido repercusión o no sobre la duración promedio de las llamadas. Para ello controla la duración de 20 llamadas al azar obteniendo los siguientes valores:

32, 43, 56, 23, 31, 22, 66, 31, 43, 21, 12, 3, 12, 54, 34, 34, 10, 2, 20, 6

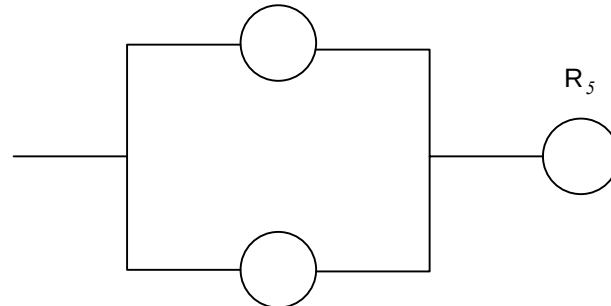
Responder a los siguientes apartados:

- a) Plantear el test que permita resolver la duda del departamento comercial (2 puntos)
- b) Con un nivel de significación del 1% y atendiendo a los valores muestrales obtenidos ¿qué respuesta daríamos al departamento comercial ante su duda? (3 puntos)
- c) Con la región de aceptación del test propuesto ¿con qué probabilidad cometeríamos el error de segunda especie si la media de la duración de las llamadas hubiese bajado realmente a 30 segundos? (3 puntos)
- d) Asumiendo que la desviación típica de la duración de las llamadas es de 20 segundos ¿con qué tamaño de muestra (número n de llamadas controladas) deberíamos realizar la prueba si queremos cometer el error de primera especie con una probabilidad del 5% y que, cuando la media real sea de 35 segundos, la probabilidad de cometer el error de segunda especie sea del 10%? (2 puntos)

SOLUCIONES**SOLUCIÓN CUESTIÓN 1:**

El sistema se reduce a

$$R_A = R_1 R_2 = 0,75 \cdot 0,83 = 0,6225$$

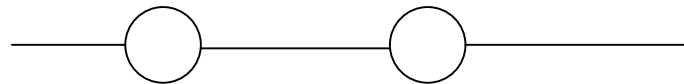


$$R_B = R_3 R_4 = 0,92 \cdot 0,65 = 0,598$$

Y este se reduce a

$$R_C = 0,8482$$

$$R_5 = 0,95$$



Y la fiabilidad del sistema es,

$$R_{\text{sys}} = R_C R_5 = [1 - (1 - R_1 R_2)(1 - R_3 R_4)] R_5 = [1 - (1 - 0,75 \cdot 0,83)(1 - 0,92 \cdot 0,65)] \cdot 0,95 = 0,8058$$

La probabilidad de fallo del sistema es,

$$P(\text{fallo del sistema}) = 1 - 0,8058 = 0,1942$$

SOLUCIÓN CUESTIÓN 2:

a) *La función de densidad viene dada por:*

$$f(x) = \frac{dF(x)}{dx} = \begin{cases} 0 & x < 0 \\ \frac{4}{9} e^{-2x/3} + \frac{1}{9} e^{-x/3} & x \geq 0 \end{cases}$$

b) *La probabilidad viene dada por:*

$$P(3 < X < 6) = F(6) - F(3) = 0,943 - 0,787 = 0,156$$

$$c) \quad P(X < 6 / x \geq 3) = \frac{\int_3^6 f(x) dx}{\int_3^{\infty} f(x) dx} = \frac{0,156}{1 - 0,787} = 0,732$$

SOLUCIÓN CUESTIÓN 3:

En primer lugar determinamos el valor de λ asociado a la distribución de Poisson

$$E[X] = 0 \cdot \frac{72}{111} + 1 \cdot \frac{27}{111} + 2 \cdot \frac{7}{111} + 3 \cdot \frac{4}{111} + 4 \cdot \frac{1}{111} = 0,5135$$

La probabilidad de que el autor encuentre más de una errata viene dada por:

$$P(X > 1) = 1 - [P(X = 0) + P(X = 1)] = 1 - \frac{e^{-0,5135} \cdot 0,5135^0}{0!} - \frac{e^{-0,5135} \cdot 0,5135}{1!} = 0,0943$$

SOLUCIÓN CUESTIÓN 4:

$X = \{\text{Longitud de una pieza}\} = N[2,1; 0,083]$

La probabilidad de aceptar el lote vendrá dada por:

$$P(2,05 < \bar{x} < 2,2) = P(\bar{x} < 2,2) - P(\bar{x} < 2,05)$$

Donde la media muestral sigue una distribución $N(2,1; 0,083/\sqrt{50})$ por tanto

$$P(2,05 < \bar{x} < 2,2) = P(\bar{x} < 2,2) - P(\bar{x} < 2,05) = P\left(Z < \frac{2,2 - 2,1}{0,083/\sqrt{50}}\right) - P\left(Z < \frac{2,05 - 2,1}{0,083/\sqrt{50}}\right) = P(Z < 8.51) - (P(Z < -4,25) = 1$$

La probabilidad de rechazar el lote es 0.

SOLUCIÓN CUESTIÓN 5:

$$\hat{p}_A \approx N(m = p_A, \sigma = \sqrt{p_A \cdot q_A / n_A})$$

$$\hat{p}_{A1} + \hat{p}_{A2} \approx N(m = 2p_A, \sigma = \sqrt{2p_A q_A / n_A})$$

$$\frac{\hat{p}_{A1} + \hat{p}_{A2}}{2} \approx N\left(m = p_A, \sigma = \sqrt{\frac{1}{2} \frac{p_A q_A}{n_A}}\right) \quad \text{Lo mismo resultará en el caso de B:}$$

$$\frac{\hat{p}_{B1} + \hat{p}_{B2}}{2} \approx N\left(m = p_B, \sigma = \sqrt{\frac{1}{2} \frac{p_B q_B}{n_B}}\right)$$

Teniendo en cuenta que si dos variables X, Y son independientes: $\sigma^2(X - Y) = \sigma^2(X) + \sigma^2(Y)$

$$\frac{\hat{p}_{A1} + \hat{p}_{A2}}{2} - \frac{\hat{p}_{B1} + \hat{p}_{B2}}{2} \approx N\left(m = p_A - p_B, \sigma = \sqrt{\frac{p_A q_A}{2n_A} + \frac{p_B q_B}{2n_B}}\right)$$

SOLUCIÓN CUESTIÓN 6:

a) El intervalo de confianza cuando σ es desconocido es

$$\bar{x} \pm t_{n-1}^{\alpha/2} \frac{s'}{\sqrt{n}}$$

Como tenemos los valores agrupados por frecuencias f_i en cada intervalo i cuya marca de clase o punto medio es I_i será:

$$\bar{x} = \frac{\sum f_i I_i}{\sum f_i} = 349,35 \quad y \quad s' = \sqrt{\frac{\sum f_i (I_i - \bar{x})^2}{(\sum f_i) - 1}} = 4,783$$

Y el intervalo

$$I = \left(349,35 \pm t_{27-1}^{0,01/2} \frac{s'}{\sqrt{27}} \right) = \left(349,35 \pm 2,779 \frac{4,783}{\sqrt{27}} \right) = (346,79; \quad 351,91)$$

b) El tamaño de muestra necesario será

$$n = \left(\frac{z_{\alpha/2} \sigma}{e} \right)^2 = \left(\frac{1,96 \cdot 5}{5} \right)^2 = 3,84 \rightarrow 4 \text{ probetas}$$

SOLUCIÓN PRÁCTICA 1:

a) La escala vertical es frecuencia absoluta: número de datos contenido en cada intervalo del histograma. Esta escala es mucho mayor en el histograma de la línea B porque éste tiene muchos menos intervalos (menos barras). Teniendo en cuenta que ambos histogramas se han construido con 500 datos, al dividir el rango de variación de la longitud en un menor número de intervalos aparecen más datos en cada uno de ellos, aumentando por tanto la frecuencia absoluta.

b) La forma del histograma se parece a la campana de Gauss, lo que sugiere que los datos siguen aproximadamente una distribución normal. En este tipo de distribución, el intervalo $m \pm 3s$ comprende el 99,7% de los datos, es decir prácticamente todos. En este caso todos los datos varían entre 97 y 101, aproximadamente, resultando un intervalo de variación de $101-97=4\text{mm}$ que corresponderá por tanto a 6 veces la desviación típica.

Por tanto, $s \approx (101-97)/6 = 0,67 \Rightarrow s^2 = 0,67^2 = 0,44$

De los valores que se indican, el más parecido es 0,352, que será la cuasivarianza.

c) Rango de A $\approx 101 - 97 = 4 \text{ mm}$

Rango de B $\approx 102,5 - 92,5 = 10 \text{ mm}$

El hecho que los rangos sean tan distintos y teniendo en cuenta que en ambos casos el modelo normal parece adecuado, la línea B tendrá mayores parámetros de dispersión que la A (es decir, mayor desviación típica, varianza e intervalo intercuartílico).

d) El histograma es en general una técnica poco adecuada para detectar datos anómalos, ya que un solo dato bastante extremo daría lugar a una barra de altura unitaria, que fácilmente puede pasar desapercibida. Para la detección de datos anómalos es más conveniente el uso del diagrama box-whisker o el papel probabilístico normal.

e) En la línea A la proporción de piezas fuera del intervalo $[98, 102]$ es bastante pequeña. En cambio, en la línea B la proporción de piezas inferiores a 98 es excesiva. Por tanto, para

disminuir la proporción de piezas fuera de tolerancias sobre todo hay que actuar sobre la línea B. La primera recomendación sería centrar el proceso, es decir realizar los ajustes en las máquinas pertinentes para que la media del proceso sea de 100 y no de 98 aproximadamente como es ahora. La segunda recomendación es investigar por qué la variabilidad es mucho mayor en la línea B que en la A. Una vez conocidas estas causas, es posible que se pueda actuar sobre ellas para disminuir la variabilidad. Otra posible medida sería centrar el proceso en el caso de la línea A, aunque esta medida sólo reduciría ligeramente la proporción de piezas fuera de tolerancia.

SOLUCIÓN PRÁCTICA 2:

a) Las cuatro variables del modelo tienen un efecto estadísticamente significativo, ya que su p-valor es mucho menor a 0.05 (riesgo de primera especie). Por tanto, el modelo de predicción deberá utilizar la información de las cuatro variables. A partir de los valores estimados de los coeficientes que aparecen en la tabla, el modelo será:

$$\text{Rendimiento} = 156,83 + 2,735 \cdot \text{temperatura} - 27,13 \cdot \text{pH} + 1,92 \cdot \text{azúcar} + 3,22 \cdot \text{proteína}$$

b) El valor 156,83 es la constante del modelo y se interpreta como el valor medio del rendimiento que cabe esperar si el valor de las cuatro variables del modelo fuese nulo. El valor 2,73502 es el coeficiente asociado a la variable temperatura y se interpreta como el incremento medio de rendimiento que cabe esperar si la temperatura media durante la fermentación se aumentase en 1° C y el resto de variables permanecieran constantes.

SOLUCIÓN PRÁCTICA 3:

a) H_0 : existe independencia entre los dos factores (proveedor y clasificación de los componentes en función de su calidad).

H_1 : no existe independencia (la calidad de las piezas de los tres proveedores no es la misma).

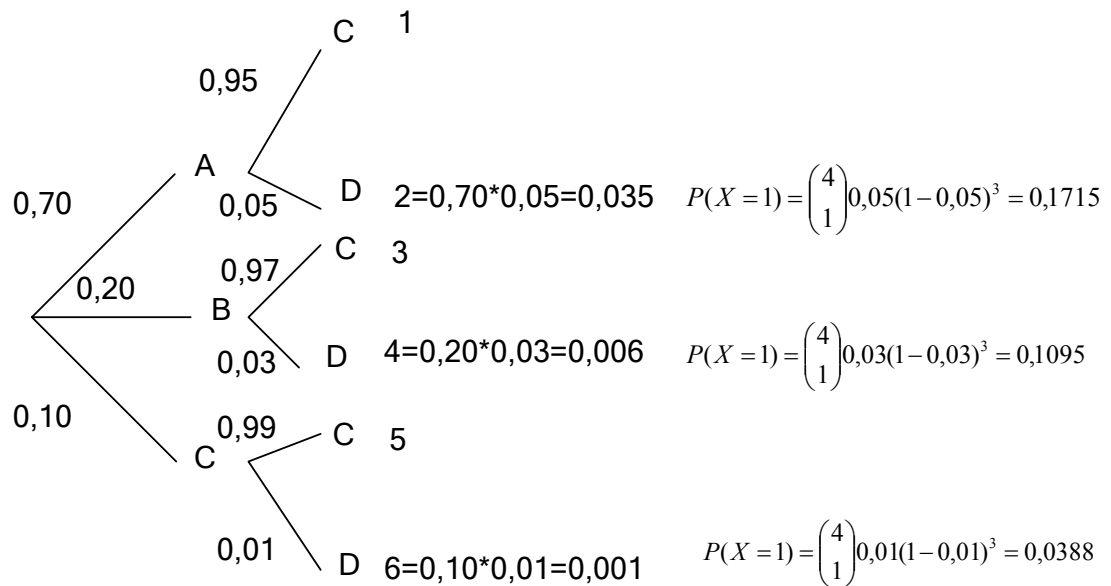
b) El estadístico de contraste se obtiene sumando los valores de “contribution to chi-squared”:
 $D = 0,90 + 0,78 + 0 + 0,10 + 2,31 + 0,11 + 1,60 + 0,41 + 0,09 = 6,3$

c) Si H_0 es cierta, D sigue una distribución chi-cuadrado con 4 grados de libertad (3 distribuidores menos uno) · (tres categorías de componentes menos una).

d) Considerando $\alpha=0.05$, el valor crítico de una chi-cuadrado con 4 grados de libertad vale 9,49. Como el valor observado de este estadístico es 6,31 que resulta menor de 9,49, la conclusión del test es aceptar la hipótesis nula. Por tanto, no hay evidencia en los datos para asegurar que el proveedor 2 ofrezca una proporción de piezas defectuosas significativamente menor que los otros dos proveedores. Es decir, no hay suficiente evidencia para corroborar la hipótesis que plantea el jefe.

SOLUCIÓN PROBLEMA 1:

a) El árbol de Bayes es.



La probabilidad de que siendo $x=1$, provenga del proveedor B es,

$$P(B/x=1) = \frac{0,20 \cdot 0,1095}{0,70 \cdot 0,1715 + 0,20 \cdot 0,1095 + 0,10 \cdot 0,0388} = 0,1502$$

b) La probabilidad de que un motor sea defectuoso es la suma de las probabilidades de los caminos 2, 4 y 6:

$$P(D) = 0,035 + 0,006 + 0,001 = 0,042$$

Y el valor medio del número de motores defectuosos es el valor medio de una binomial de parámetros $n=150$ y $p=0,042$, esto es:

$$E(x) = np = 150 \cdot 0,042 = 6,3 \text{ defectuosos/semana}$$

c) El valor medio del beneficio será:

$$E(B) = 50 \cdot P(C) + (-60) \cdot P(D) = 50 \cdot (1 - 0,042) - 60 \cdot 0,042 = 45,38 \text{€ / motor}$$

Para 150 motores, el valor medio del beneficio será de

$$E(B) = 150 \cdot 45,38 = 6807 \text{€}$$

SOLUCIÓN PROBLEMA 2:

a) La función de densidad conjunta será:

$$f_{XY}(x, y) = \frac{\partial F_{XY}(x, y)}{\partial x \partial y} = \frac{\partial((1 - e^{-2x})(1 - e^{-y}))}{\partial x \partial y} = \frac{\partial(2e^{-2x}(1 - e^{-y}))}{\partial y} = 2e^{-2x}e^{-y} \quad \text{con } 0 < x, y < +\infty$$

b) Las funciones de distribución marginales son:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dy = \int_0^{+\infty} 2e^{-2x} e^{-y} dy = 2e^{-2x} \quad \text{con } 0 < x < +\infty$$

$$F_X(x) = \int_0^x 2e^{-2x} dx = \left[-e^{-2x} \right]_0^x = 1 - e^{-2x}$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dx = \int_0^{+\infty} 2e^{-2x} e^{-y} dx = e^{-y} \quad \text{con } 0 < y < +\infty$$

$$F_Y(y) = \int_0^y e^{-y} dy = \left[-e^{-y} \right]_0^y = 1 - e^{-y}$$

c) Sí, porque se cumple que

$$f(x, y) = f(x) \cdot f(y) = 2e^{-2x} e^{-y}$$

y, por lo tanto, que

$$f(x/y) = \frac{f(x, y)}{f(y)} = f(x) = 2e^{-2x}$$

d) Las probabilidades pedidas son:

$$P(X < 1, Y < 2) = \int_0^1 \int_0^2 2e^{-2x} e^{-y} dy dx = 0,748$$

$$P(X < 1 / Y < 2) = \frac{P(X < 1, Y < 2)}{P(Y < 2)} = \frac{0,748}{\int_0^2 e^{-y} dy} = \frac{0,748}{0,865} = 0,8647$$

$$P(X < 1 / Y = 2) = \int_0^1 f_{X/Y}(x/2) dx = \int_0^1 2 \cdot e^{-2x} dx = 0,8647$$

SOLUCIÓN PROBLEMA 3:

a) Las hipótesis a contratar son:

$$H_0: m \geq m_0 = 42 \text{ contra}$$

$$H_1: m < m_0 = 42$$

b) La regla de decisión será

$$\text{Aceptar si } \bar{x} \geq m_0 - t_{n-1}^{\alpha} \frac{s'}{\sqrt{n}} = 42 - 2,539 \frac{18,14}{\sqrt{20}} = 31,70$$

(Dado que de la muestra se obtiene $\bar{x} = 27,75$ y $s' = 18,14$)

$$(\text{La } t_c \text{ calculada es } t_c = \frac{\bar{x} - m_0}{s' / \sqrt{n}} = \frac{27,75 - 42}{18,14 / \sqrt{20}} = -3,51 < -2,538, \text{ y cae en la zona de rechazo})$$

Como $27,75 < 31,70$ rechazamos H_0 y concluimos que sí ha habido una reducción significativa de la duración promedio de las llamadas

c) $\beta = P(\text{aceptar } H_0 / H_1)$

como conocemos todos $\frac{(\bar{x} - m)}{s' / \sqrt{n}} = t_{n-1}$ por lo que

$$\begin{aligned} P(\bar{x} \geq 31,70 / m = 30) &= P\left(\frac{\bar{x} - m}{s' / \sqrt{n}} \geq \left(\frac{31,70 - m}{s' / \sqrt{n}}\right)\right) = P(t_{n-1} \geq \frac{31,70 - m}{s' / \sqrt{n}}) = \\ &= P(t_{20-1} \geq \frac{31,70 - 30}{18,14 / \sqrt{20}}) = P(t_{19} \geq 0,419) = 0,34 \end{aligned}$$

Nota: Dado que el tamaño de la muestra $n=30$ es muy grande se podría haber hecho directamente como $\bar{x} \equiv N(m, \sigma / \sqrt{n})$ $P(z \geq 0,413) = 0,3398$

d) El tamaño requerido es

$$n = \left(\frac{z_{\alpha} + z_{\beta}}{|m - m_0|} \sigma \right)^2 = \left(\frac{1,645 + 1,28}{42 - 35} 20 \right)^2 = 69,8 \rightarrow 70 \text{ llamadas}$$

EXAMEN DE JULIO 2007, CON SOLUCIÓN

APELLIDOS, NOMBRE: _____

Firma: _____

GRUPO: _____

BLOQUE I. CUESTIONES Y TIPO TEST (60 MINUTOS)

Todas las cuestiones deben ser debidamente justificadas. Aquellas cuestiones contestadas correctamente pero no justificadas no serán puntuadas.

1. El rendimiento de una máquina fresadora en un día determinado de producción se distribuye como una Normal de media 80% y desviación típica del 5%. Si la máquina está en funcionamiento 5 días a la semana y los rendimientos, en días diferentes, son independientes entre sí ¿cuál es la probabilidad de que el rendimiento de la máquina sea mayor que el 85% en un día cualquiera? ¿Cuál es la probabilidad de que el rendimiento sea mayor que 85% al menos en 4 de los 5 días de la próxima semana? (1,5 puntos)

2. Los defectos en la carrocería de los coches fabricados en el turno de la mañana siguen una distribución $Ps(\lambda=1 \text{ defectos/carrocería})$ y los fabricados en el turno de tarde una $Ps(\lambda=5 \text{ defectos/carrocería})$.

Para conocer si un coche incógnita es del turno de la mañana planteamos el test:

$$H_0: \lambda=1 \text{ defecto}$$

$$H_1: \lambda=5 \text{ defectos}$$

y decidimos aceptar H_0 si encontramos en su carrocería menos de 3 defectos, calcular los valores de α y β del test propuesto. (1,5 puntos)

3. Para contrastar que la desviación típica del espesor de una plancha de aluminio es de 0,1 mm frente a la hipótesis alternativa de que ha aumentado, se toma una muestra de $n=15$ unidades y se calcula la cuasidesviación típica muestral. Para un nivel de significación del 5%, ¿a partir de qué valor de la cuasidesviación típica muestral hay evidencia suficiente para afirmar que la desviación típica poblacional es mayor que la supuesta? (2 puntos)

4. Si admitimos que la estatura de los alumnos de la UPV tiene una distribución Normal y que hay un 5% de alumnos más bajitos que García que mide 156 cm y un 27% de alumnos más altos que López que mide 178 cm

- ¿Qué proporción de alumnos son más altos que el prof. Clemente que mide 165 cm? (1 punto)
- Si medimos a los 64 alumnos que se han presentado a este examen y determinamos su estatura media, ¿Cuál es la probabilidad de que este promedio esté por arriba de la estatura del prof. Clemente, asumiendo que estos alumnos son una muestra representativa del conjunto de alumnos de la UPV? (1 punto)

5. La resistencia a la flexión de unas barras de acero se modeliza según la siguiente función de densidad:

$$f(y) = \begin{cases} \frac{1}{60}(y-35) & 35 \leq y \leq 41 \\ k(55-y) & 41 \leq y \leq 55 \\ 0 & \text{otros} \end{cases}$$

- Calcular el valor de la constante k . (0,5 puntos)
- Determinar la función de distribución de la variable resistencia. (0,5 puntos)
- Calcular el valor medio de la resistencia a la flexión. (0,5 puntos)

6. Uno de los cuadros de control de la Eh MeTeTe incorpora unos componentes cuya duración es exponencial de vida media 30 días. Hoy, día 10 de julio de 2007, se han reemplazados todos los componentes del cuadro colocando componentes nuevos. El funcionamiento del cuadro no se afecta sensiblemente mientras funcione al menos la mitad de los componentes, por lo tanto, ¿para qué fecha habrá que programar un mantenimiento preventivo si queremos que en el momento de la revisión la proporción de componentes en funcionamiento sea igual o superior al 50%? (1,5 puntos)

- a) El 30 de julio
- b) El 9 de agosto
- c) El 19 de agosto
- d) El día de San Valentín

APELLIDOS, NOMBRE: _____

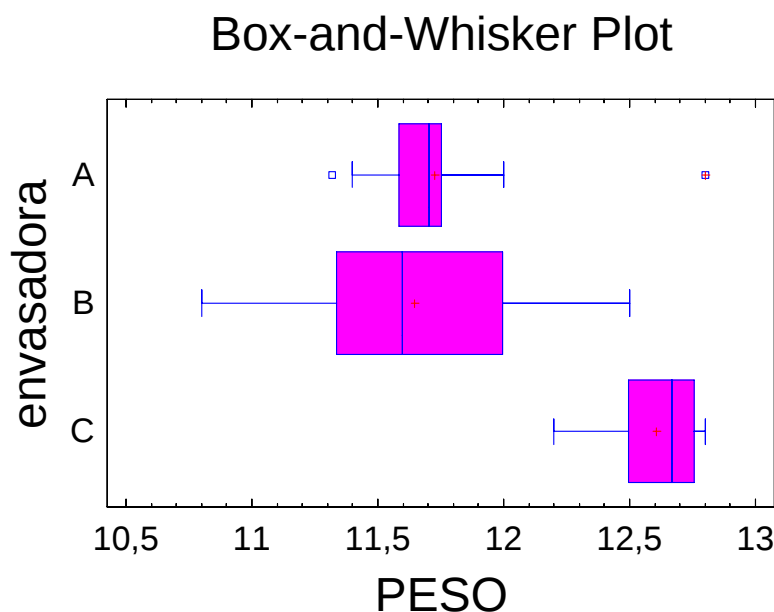
Firma: _____

GRUPO: _____

BLOQUE II. PRÁCTICAS DE ORDENADOR (30 MINUTOS)

PROBLEMA “A” - PRÁCTICA 1

Una industria elabora un producto químico líquido que se envasa en recipientes de 10 litros, que una vez llenos tienen un peso nominal de 11,20 kilogramos. El proceso de envasado se lleva a cabo por medio de tres máquinas envasadoras (A, B y C) que funcionan en paralelo. Las tres máquinas se calibraron hace seis meses para adicionar el mismo peso, pero el ingeniero de control sospecha que alguna de ellas necesita volver a ser calibrada. Para comprobarlo, se tomaron al azar 20 recipientes llenados por la máquina A y se obtuvo su peso, y se repitió lo mismo con las otras dos envasadoras. Con los datos de peso de cada una de las máquinas se ha obtenido el siguiente diagrama Box-Whisker múltiple:



Responder a las siguientes preguntas (con la precisión que permite el gráfico) justificando la respuesta:

- 1) Calcular el rango de la envasadora A. (0,5 puntos)

2) Calcular el rango intercuartílico de la envasadora B. (0,5 puntos)

3) ¿Qué modelo teórico crees que sería conveniente para modelizar la distribución de los datos de la envasadora B? (0,5 puntos)

0,12
0,25
1,49
0,85

5) ¿En cuál de los tres casos es menor el coeficiente de asimetría? (0,5 puntos)

PRÁCTICA 2. En una granja ovina se realiza un estudio para determinar cómo afecta la alimentación al desarrollo de las ovejas. Para ello se seleccionan al azar 50 ovejas de edades comprendidas entre 6 y 18 meses. Para cada una de ellas se determina su peso en kilogramos (variable “peso”), la cantidad media de pienso ingerido diariamente, en kilogramos (variable “pienso”), y la cantidad de agua ingerida diariamente en litros (variable “agua”). Con los datos obtenidos se realiza un análisis de regresión lineal múltiple cuyos resultados se muestran a continuación.

Multiple Regression Analysis

Dependent variable: peso

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	-138,965	31,6985	-4,38397	0,0001
edad	21,6735	4,40903	4,91569	0,0000
edad^2	-0,617246	0,18435	-3,34823	0,0017
agua	1,83546	2,62653	0,698813	0,4883
pienso	11,167	3,51715	3,17501	0,0027

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	37362,2	4	9340,55	39,03	0,0000
Residual	10769,5	45	239,321		
Total (Corr.)	48131,6	49			

R-squared = 77,625 percent

R-squared (adjusted for d.f.) = 75,6361 percent

Standard Error of Est. = 15,47
Mean absolute error = 11,778
Durbin-Watson statistic = 1,64689 (P=0,0778)
Lag 1 residual autocorrelation = 0,152771

A la vista de estos resultados responder a las siguientes preguntas justificando convenientemente las respuestas:

- a) Escribir la ecuación matemática que utilizarías para predecir el peso en función de las variables que ejercen un efecto estadísticamente significativo (tomar un 5% como nivel de significación). (0,5 puntos)
- b) En el modelo obtenido, ¿qué expresa el coeficiente asociado a la variable “pienso”? (0,5 puntos)
- c) Estimar el peso medio que cabría esperar para una oveja de tres meses de edad que consume 4 kg de pienso y 6 litros de agua en promedio cada día. (0,5 puntos)
- d) ¿Qué expresa el coeficiente R^2 ? (0,5 puntos)
- e) Se considera que una oveja tiene un desarrollo normal si alcanza un peso superior a 70 kg cuando tiene un año de edad. Calcular la proporción de ovejas en dicha granja con un desarrollo atípico. (Asumir que todas las ovejas consumen 4 kg de pienso y 4 litros de agua en promedio cada día). (1,5 puntos)

PRÁCTICA 3.- Hemos planteado el siguiente test de hipótesis sobre la media del espesor de unas pletinas mecanizadas en una fresadora (asumimos que el espesor sigue una distribución Normal de $\sigma=1\text{mm}$)

$$H_0: \mu=5\text{mm}$$

$$H_1: \mu>5\text{mm}$$

Tomamos una muestra de 10 pletinas y los espesores medidos los procesamos con el STATGRAPHICS. El ordenador nos proporciona como resultado el valor de la media muestral y el p-valor=0,08

- a) Interpretar el p-valor obtenido en el análisis anterior. (1 punto)
- b) En un test de hipótesis, ¿cómo se toman las decisiones en función de los valores del p-valor y el nivel de significación o probabilidad de error de primera especie α adoptado? (1 punto)
- c) Calcular el p-valor si la hipótesis alternativa fuera $H_1: \mu \neq 5\text{mm}$. (1 punto)

APELLIDOS, NOMBRE: _____

Firma: _____

GRUPO: _____

BLOQUE III. PROBLEMAS (90 MINUTOS)**Cada problema son 10 puntos. La nota del bloque es la media de las notas de los problemas.**

1. En la clínica ginecológica de la Dra. Cesárea Segura, se aplica a las presuntas embarazadas el test del “predictor” que acierta en el 98% de los casos en que es aplicado y falla en el 1%. Cuando el resultado de la prueba presenta dudas (1% restante de los casos) se aplica la “prueba de la rana” que acierta siempre.



Si admitimos que sólo el 75% de las pacientes sometidas a las pruebas están realmente embarazadas, calcular:

- La probabilidad de que no esté realmente embarazada una paciente a la que la doctora le ha dicho que sí lo está (¡vaya susto tonto!). (3 puntos)
 - La probabilidad de que habiéndole dicho que no está embarazada y que lo suyo “sólo son gases”, esté realmente embarazada (¡menuda jugada!). (3 puntos)
 - La proporción de aciertos de las pruebas conjuntamente realizadas. (2 puntos)
 - La proporción de errores de las pruebas conjuntamente realizadas. (2 puntos)
2. Entre los alumnos de la UPV aficionados al ciclismo, el tiempo que están pedaleando en cada salida es una variable aleatoria X con distribución $N=(m_x=100 \text{ min}; \sigma_x=20 \text{ min})$ y las kcalorías que queman en cada salida también es otra variable Y con distribución $N=(m_y=400 \text{ kcal}; \sigma_y=50 \text{ kcal})$. El coeficiente de correlación entre ellas es $\rho=0,90$. Se pide:
- ¿Qué consumo promedio de kcalorías cabe esperar que haya consumido un ciclista de la UPV que ha estado una hora pedaleando? (3 puntos)
 - Entre los ciclistas que han quemado 500 kcalorías ¿Qué proporción de ellos han pedaleado menos de dos horas? (3 puntos)
 - ¿Cuántos minutos han de pedalear los de la peña ciclista de la UPV para que el 90% de ellos hayan quemado más de 500 kcalorías? (4 puntos)
3. En las pasadas elecciones municipales, la empresa “Dermostropia” hizo una encuesta a pie de urna a un total de 500 valencianos que ya habían votado. De ellos 280 dijeron haber votado al PP, 200 al PSOE y el resto a otras formaciones.
- Con un nivel de confianza del 95%, calcular un intervalo de confianza para la proporción P de votantes del PP con los datos disponibles en el momento de la encuesta. (3 puntos)
 - ¿Cuál es el error cometido en la estimación de la proporción anterior si utilizamos un nivel de confianza del 99%? (3 puntos)
 - En ese momento de la encuesta, la empresa “Tri- σ ” entrevistó a 300 alicantinos y de ellos 165 habían votado al PP. Con un nivel de significación de $\alpha=0,05$ ¿podía admitirse que las proporciones de votantes del PP en Valencia y en Alicante eran las mismas? (4 puntos)

SOLUCIONES**BLOQUE I****Solución Cuestión 1.***SOLUCIÓN:*

Llamando a X =rendimiento de la máquina, ésta variable sigue una distribución $X \sim N(80; 5)$. La probabilidad de que el rendimiento de la máquina sea superior al 85% es:

$$P(X > 85) = P\left(Z \geq \frac{85 - 80}{5}\right) = P(Z \geq 1) = 0,1587$$

Si llamamos a $Y = n^\circ$ de días, de entre 5 consecutivos, en los que el rendimiento ha sido mayor de 85, ésta Y sigue una distribución binomial de parámetros $Y \sim B(n=5; p=P(X>85)=0,1587)$.

La probabilidad de que ocurra en al menos 4 de los 5 días es:

$$\begin{aligned} P(Y \geq 4) &= P(Y = 4) + P(Y = 5) = \binom{5}{4} 0,1587^4 (1 - 0,1587)^1 + \binom{5}{5} 0,1587^5 (1 - 0,1587)^0 = \\ &= 0,00267 + 0,00101 = 0,00277 \end{aligned}$$

Solución Cuestión 2.*SOLUCIÓN:*

N° defectos en una carrocería $X \sim Ps(\lambda=1)$ si H_0 cierta y $X \sim Ps(\lambda=5)$ si H_1 cierta

$$\alpha = P(\text{rechazar } H_0 \mid H_0 \text{ cierta}) = P(Ps(\lambda=1) \geq 3) = 1 - P(Ps(\lambda=1) \leq 2) = 0,0803$$

$$\beta = P(\text{aceptar } H_0 \mid H_1 \text{ cierta}) = P(Ps(\lambda=5) \leq 2) = 0,1247$$

Solución Cuestión 3.*SOLUCIÓN:* El valor que nos piden debe cumplir que:

$$P(s > c) = 0,05$$

Multiplicando y dividiendo a ambos lados de la desigualdad, se tiene,

$$P\left((n-1) \frac{s^2}{\sigma_0^2} > (n-1) \frac{c^2}{\sigma_0^2}\right) = 0,05$$

El valor de $\chi_{14;0,05}^2 = 23,68$ de donde $(n-1) \frac{c^2}{\sigma_0^2} = 23,68$. Despejando c queda:

$$c = \sqrt{\frac{23,68 \cdot 0,01}{14}} = 0,13$$

Solución Cuestión 4.*SOLUCIÓN:*

$$\Phi\left(\frac{156 - \mu}{\sigma}\right) = 0,05 \quad y \quad 1 - \Phi\left(\frac{178 - \mu}{\sigma}\right) = 0,27 \quad \text{o bien} \quad \Phi\left(\frac{178 - \mu}{\sigma}\right) = 0,73$$

$$\text{Tomando inversas} \quad \left(\frac{156 - \mu}{\sigma}\right) = -1,645 \quad y \quad \left(\frac{178 - \mu}{\sigma}\right) = 0,615$$

Resolviendo el sistema $\mu=172$ y $\sigma=9,73$. Llamando a X = altura de los alumnos de la UPV, ésta sigue una distribución normal $N(172;9,73)$.

$$a) P(x \geq 165) = P(z \geq \frac{165-172}{9,73}) = P(z \geq -0,7194) = 0,7641 = 76,41\%$$

$$b) P(\bar{x} \geq 165) = P(z \geq \frac{165-172}{9,73/\sqrt{64}}) = P(z \geq -5,76) = 1$$

Solución Cuestión 5.

SOLUCIÓN:

a)

$$k=1/140$$

$$f(y) = \begin{cases} \frac{1}{60}(y-35) & 35 \leq y \leq 41 \\ k(55-y) & 41 \leq y \leq 55 \\ 0 & \text{otros} \end{cases}$$

b) La función de distribución es.

$$P(Y \leq y_1) = \int_{-\infty}^{y_1} f_Y(y) dy = \frac{(y_1 - 35)^2}{120} \quad 35 \leq y_1 \leq 41$$

$$P(Y \leq y_2) = \int_{35}^{45} \frac{1}{60}(y-35) dy +$$

$$\int_{41}^{y_2} \frac{1}{140}(55-y) dy = \frac{1}{60} \left(\frac{45^2 - 35^2}{2} - 35 \cdot 10 \right) + \frac{55(y_2 - 41) - \frac{1}{2}(y_2^2 - 41^2)}{140} \quad 41 \leq y_2 \leq 55$$

c) El valor medio es:

$$E(y) = \int_{-\infty}^{+\infty} y f(y) dy = \int_{35}^{41} y \frac{1}{60}(y-35) dy + \int_{41}^{55} y \frac{1}{140}(55-y) dy =$$

$$\frac{1}{60} \left[\frac{y^3}{3} - 35 \frac{y^2}{2} \right]_{35}^{41} + \frac{1}{140} \left[55 \frac{y^2}{2} - \frac{y^3}{3} \right]_{41}^{55} = 11,7 + 31,967 = 43,667$$

Solución Cuestión 6.

SOLUCIÓN:

La duración de las componentes es una $X=EXP(\lambda=1/30=0,0333)$

La $P(X \geq x) = \exp(-\lambda x)$

Debemos buscar un $x_0/P(X \geq x_0) = 0,50$

$$\exp(-0,0333 x_0) = 0,50$$

Tomando logaritmos neperianos $x_0=20,79$ (tomamos 20 para más seguridad)

Habrà que revisar el $10+20= 30$ de julio

BLOQUE II

Solución práctica 1.

1) Calcular el rango de la envasadora A

$$\text{Rango} = \text{máximo} - \text{mínimo} = 12,8 - 11,3 = 1,5$$

2) Calcular el rango intercuartílico de la envasadora B

$$\text{Rango intercuartílico} = \text{tercer cuartil (extremo derecho de la caja)} - \text{primer cuartil (extremo izquierdo de la caja)} = 12 - 11,32 = 0,68$$

3) ¿Qué modelo teórico crees que sería conveniente para modelizar la distribución de los datos de la envasadora B?

La distribución es bastante simétrica y dado que estamos en el caso de una variable continua, el modelo normal parece razonable. El teorema central del límite justifica por qué en la vida real el modelo normal es capaz de modelizar frecuentemente las variables continuas.

4) ¿Cuál de los siguientes valores crees que es la desviación típica muestral del peso de los recipientes llenados por la envasadora B?

- 0,12
- 0,25
- 1,49
- 0,85

Rango de los datos de B: $12,5 - 10,8 = 1,7$. Asumiendo que la distribución de los datos de B sigue un modelo normal, el 99,7% de los datos estarán entre $m \pm 3s$. Por tanto, la desviación típica será similar al rango dividido entre 6: $s = \text{rango} / 6 = 1,7 / 6 = 0,28$. La solución correcta será por tanto 0,25.

5) ¿En cuál de los tres casos es menor el coeficiente de asimetría?

En la envasadora A y B, el gráfico sugiere que la distribución de los datos es bastante simétrica, por lo que el coeficiente de asimetría será próximo a cero. En cambio en la envasadora C los datos siguen una distribución asimétrica negativa: el bigote izquierdo es claramente más largo que el derecho y la mediana (línea interior de la caja) está desplazada hacia la derecha. Por tanto, al ser asimétrica negativa, su coeficiente de asimetría será negativo y dado que los demás son próximos a cero, éste será el menor de los tres. Solución: el coeficiente de asimetría de C es el menor de los tres.

Solución práctica 2.

- a) Escribir la ecuación matemática que utilizarías para predecir el peso en función de las variables que ejercen un efecto estadísticamente significativo.

La ecuación del modelo que se deduce de la tabla de resultados es:

$$\text{Peso} = -138,96 + 21,67 \cdot \text{edad} - 0,617 \cdot \text{edad}^2 + 1,83 \cdot \text{agua} + 11,17 \cdot \text{pienso}$$

En esta ecuación todos los coeficientes son estadísticamente significativos excepto el de agua (p -valor = 0,49). No tiene sentido utilizar una ecuación con un coeficiente no significativo para realizar predicciones, así que deberá considerarse como cero. Por tanto, la ecuación que se utilizará para predecir el peso sería:

$$\text{Peso} = -138,96 + 21,67 \cdot \text{edad} - 0,617 \cdot \text{edad}^2 + 11,17 \cdot \text{pienso}$$

(aunque sería aconsejable eliminar el término “agua” del modelo y recalcular los coeficientes para obtener una mejor estimación).

- b) En el modelo obtenido, ¿qué expresa el coeficiente asociado a la variable “pienso”?

Es lo que aumenta en promedio el peso de una oveja (en kg) por cada kilogramo de aumento del pienso, si el resto de variables permanecen constantes. Es decir, si el consumo de pienso se incrementa en 1 kg, el peso en promedio se incrementará 11,17 kg.

- c) Estimar el peso medio que cabría esperar para una oveja de tres meses de edad que consume 4 kg de pienso y 6 litros de agua en promedio cada día.

$$\text{Peso} = -138,96 + 21,67 \cdot \text{edad} - 0,617 \cdot \text{edad}^2 + 11,17 \cdot \text{pienso}$$

Sustituyendo en la ecuación: edad=3, pienso=4 se obtiene:

$$\text{peso} = -34,8 \text{ kg.}$$

Evidentemente este resultado no tiene sentido y la razón es muy sencilla: el modelo se ha ajustado con los datos de 50 ovejas con edades comprendidas entre 6 y 18 meses. El modelo no es capaz de predecir correctamente qué ocurrirá fuera de este intervalo, y por tanto no se puede utilizar el modelo para estimar cuál sería el peso esperado para una oveja de tres meses.

- d) ¿Qué expresa el coeficiente R^2 ?

Expresa el porcentaje de la varianza del peso explicado por el modelo. Es decir, el modelo explica el 77,6% de la varianza del peso.

- e) Se considera que una oveja tiene un desarrollo normal si alcanza un peso superior a 70 kg cuando tiene un año de edad. Calcular la proporción de ovejas en dicha granja con un desarrollo atípico (asumir que todas las ovejas consumen 4 kg de pienso y 4 litros de agua en promedio cada día).

$$\text{Peso} = -138,96 + 21,67 \cdot \text{edad} - 0,617 \cdot \text{edad}^2 + 11,17 \cdot \text{pienso}$$

Si edad=12 meses, pienso=4, sustituyendo en la ecuación: peso medio estimado = 76,91 kg.

Desviación típica residual = Standard Error of Est. = 15,47

$$P(\text{desarrollo atípico}) = P[N(m=76,91, s=15,47) < 70] = P[N(0,1) < (70-76,91)/15,47] = P[N(0,1) < -0,45] \\ = (\text{tabla}) = 0,326.$$

Solución práctica 3.

SOLUCIÓN:

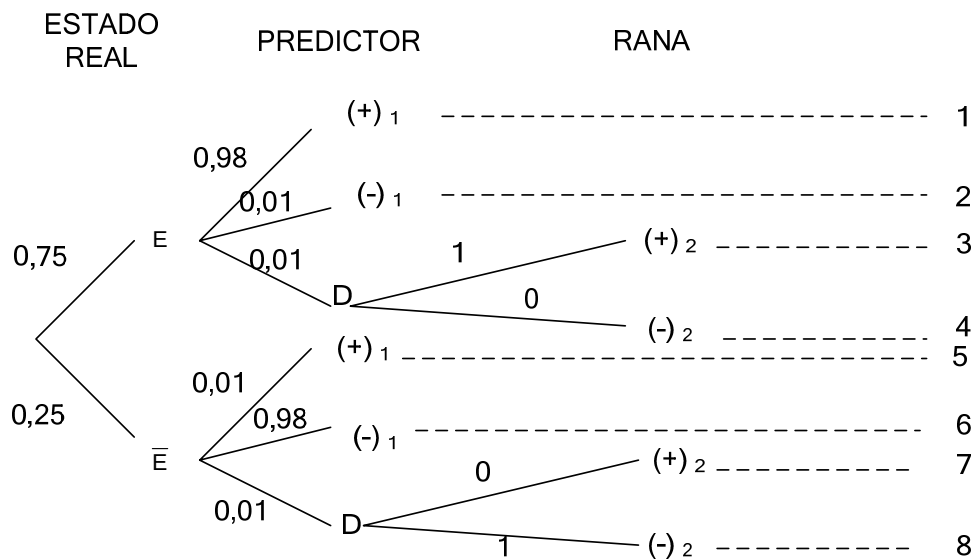
- a) Es la probabilidad de obtener una muestra que dé una media mayor o igual que la media muestral obtenida cuando la media poblacional μ sea realmente 5 mm.
- b) Se rechaza la hipótesis nula cuando el p -valor es menor que el nivel de significación α propuesto

- c) Cuando la hipótesis alternativa es a ambos lados de la media, el p-valor es la probabilidad de obtener en módulo un valor superior al observado. Con los mismos datos, si antes por encima es 0,08, por simetría, por debajo también será 0,08, y por ello el p-valor será la suma de estas dos probabilidades, esto es, 0,16.

BLOQUE III

Solución problema 1.

SOLUCIÓN:



$$a) P(\bar{E} | ((+)_1 \cup (+)_2)) = \frac{5+7}{1+3+5+7} = \frac{0,25 \cdot 0,01 + 0}{0,75 \cdot 0,98 + 0,75 \cdot 0,01 \cdot 1 + 0,25 \cdot 0,01 + 0} = 0,0033 \Rightarrow 3,3\%_{00}$$

$$b) P(E | ((-)_1 \cup (-)_2)) = \frac{2+4}{2+4+6+8} = \frac{0,75 \cdot 0,01 + 0}{0,75 \cdot 0,01 + 0 + 0,25 \cdot 0,98 + 0,25 \cdot 0,01 \cdot 1} = 0,029 \Rightarrow 2,9\%$$

$$c) P(\text{acertar}) = 1 + 3 + 6 + 8 = 0,75 \cdot 0,98 + 0,75 \cdot 0,01 \cdot 1 + 0,25 \cdot 0,98 + 0,25 \cdot 0,01 \cdot 1 = 0,99 \Rightarrow 99\%$$

$$d) P(\text{error}) = 1 - P(\text{aciertos}) = 1 - 0,99 = 0,01 \Rightarrow 1\%$$

Solución problema 2.

SOLUCIÓN:

$$E(\text{kcal} / t_i) = E(\text{kcal}) + \frac{\text{cov}(\text{kcal}, t_i)}{\sigma_{t_i}^2} (t_i - E(t_i)) = E(\text{kcal}) + \rho \frac{\sigma_{\text{kcal}}}{\sigma_{t_i}} (t_i - E(t_i))$$

$$D^2(\text{kcal} / t_i) = \sigma_{\text{kcal}}^2 (1 - \rho^2)$$

Con los datos del problema obtenemos:

$$a) E(\text{kcal}/t_i=60) = 310 \text{ kcal} \quad \text{Además obtenemos } D^2(\text{kcal}/t_i) = 475 \text{ kcal}^2 \rightarrow D(\text{kcal}/t_i) = 21,79$$

b) Con las fórmulas iniciales, intercambiando las variables obtenemos:

$$E(ti/kcal=500)=136 \text{ min} \quad D(ti/kcal=500)=8,72 \text{ por lo que}$$

$$P((ti/kcal = 500) < 120) = \Phi\left(\frac{120 - 136}{8,72}\right) = 0,0336 \rightarrow 3,36\%$$

c) Habrá que buscar un $ti_0/P(kcal/ti_0 > 500) = 0,90$ o bien que $P(kcal/ti_0 \leq 500) = 0,10$

$$\Phi\left(\frac{500 - \mu_c}{21,79}\right) = 0,10 \Rightarrow \mu_c = 527,9$$

$$\text{Como } E(kcal/ti) = E(kcal) - \rho \frac{\sigma_{kcal}}{\sigma_{ti}} (ti_0 - E(ti)) = 527,9 = 400 + 0,9 \frac{50}{20} (ti_0 - 100)$$

Despejando $ti_0 = 156,8 \text{ min}$

Solución problema 3.

SOLUCIÓN

$$a) p \pm z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \rightarrow \frac{280}{500} \pm 1,96 \sqrt{\frac{(280/500)(1-280/500)}{500}} \rightarrow [0,516; 0,604]$$

b) Considerando la precisión como la semiamplitud del intervalo de confianza

$$e = \pm 2,57 \sqrt{\frac{0,4(1-0,4)}{500}} \Rightarrow \pm 5,63\%$$

c) Contrastamos el test

$$H_0: P_1 = P_2$$

$$H_1: P_1 \neq P_2 \text{ deberemos aceptar si}$$

$$|p_1 - p_2| \leq z_{\alpha/2} \sqrt{P(1-P)} \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \quad \text{siendo} \quad P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

$$\left|\frac{280}{500} - \frac{165}{300}\right| \leq 1,96 \sqrt{0,5563(1-0,5563)} \sqrt{\left(\frac{1}{500} + \frac{1}{300}\right)} \quad \text{siendo} \quad P = \frac{280 + 165}{500 + 300} = 0,5563$$

$|0,56 - 0,55| = 0,01 < 0,071$ por lo que no rechazamos la hipótesis nula y admitimos que la proporción en Alicante es igual que en Valencia

EXAMEN DE ENERO 2007, CON SOLUCIÓN

APELLIDOS, NOMBRE: _____

Firma: _____

GRUPO: _____

BLOQUE I. CUESTIONES Y TIPO TEST (60 MINUTOS)

Todas las cuestiones deben ser debidamente justificadas. Aquellas cuestiones contestadas correctamente pero no justificadas no serán puntuadas.

1.- Un embarque de sustancias químicas llega en 15 contenedores. Se eligen tres al azar, sin reemplazamiento, para hacer una inspección de la pureza del producto. Si dos de los 15 contenedores no cumplen con los requisitos de pureza ¿cuál es la probabilidad de que uno de ellos esté en la muestra? (1,5 puntos)

2.- En una pequeña red de computadoras, el tiempo que transcurre entre accesos de los usuarios se distribuye como una exponencial de media 0,5 minutos. ¿Cuál es la probabilidad de que el primer acceso se haya realizado entre 0,5 y 1 minutos de haber encendido el ordenador? (1,5 puntos)

3.- Para estimar la proporción de bujías defectuosas que produce una empresa ¿qué tamaño de muestra debemos emplear para cometer un error superior al 3% con una probabilidad del 15%, si hemos tomado una muestra preliminar y en ella hemos observado que hay un 2,5% de unidades defectuosas? (1,5 puntos)

4.- La resistencia eléctrica de un componente electrónico sigue una distribución normal y su desviación típica es de 1,5 ohmios. Para contrastar que la media de dicha resistencia es de 20 ohmios, se toma una muestra de 9 unidades y se calcula la media muestral. Si la media de las resistencias fuera de 22 ohmios, y considerando un nivel de significación del 1% ¿cuánto vale la probabilidad de cometer un error de segunda especie? (2 puntos)

5.- La variable aleatoria X tiene una función de densidad de probabilidad

$$f(x) = \begin{cases} k(7 - 2x) & 0 \leq x \leq 3 \\ 0 & \text{en otro caso} \end{cases}$$

Si se obtienen tres observaciones independientes de la variable ¿Cuál es la probabilidad de que todas ellas sean menores que la media? (2 puntos)

- a) 0,167
- b) 0,531
- c) 0,723
- d) 0,833

6.- En un proceso de fabricación de la industria textil se definen los siguientes sucesos:

D={el telar está desengrasado}

E={El suministro eléctrico es insuficiente}

T={El tejido presenta defectos}

Además se sabe:

$$\begin{array}{ll} P(D)=0,08 & P(T|D)=0,68 \\ P(E)=0,62 & P(T \cup E)=0,72 \\ P(T)=0,27 & \end{array}$$

Las probabilidades $P(D \cup T)$ y $P(T|E)$ son respectivamente: (1,5 punto)

- a) 0,3284; 0,2700
- b) 0,2956; 0,2742
- c) 0,3284; 0,2742
- d) 0,2956; 0,2700

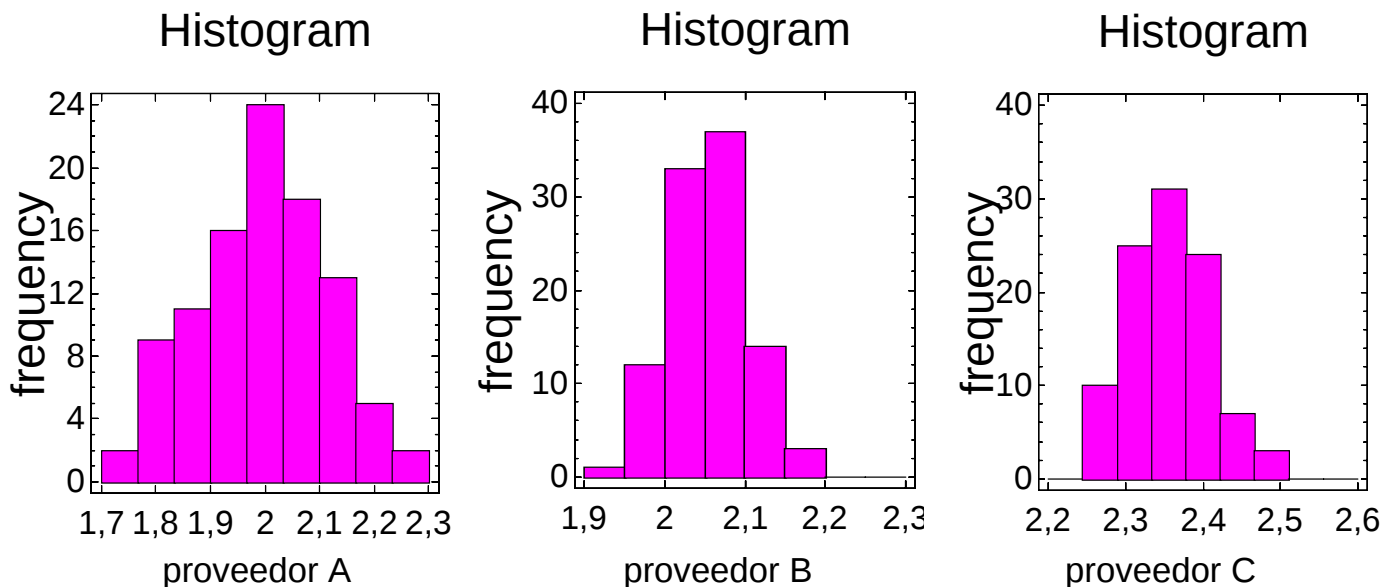
APELLIDOS, NOMBRE: _____

Firma: _____

GRUPO: _____

BLOQUE II. PRÁCTICAS DE ORDENADOR (30 MINUTOS)

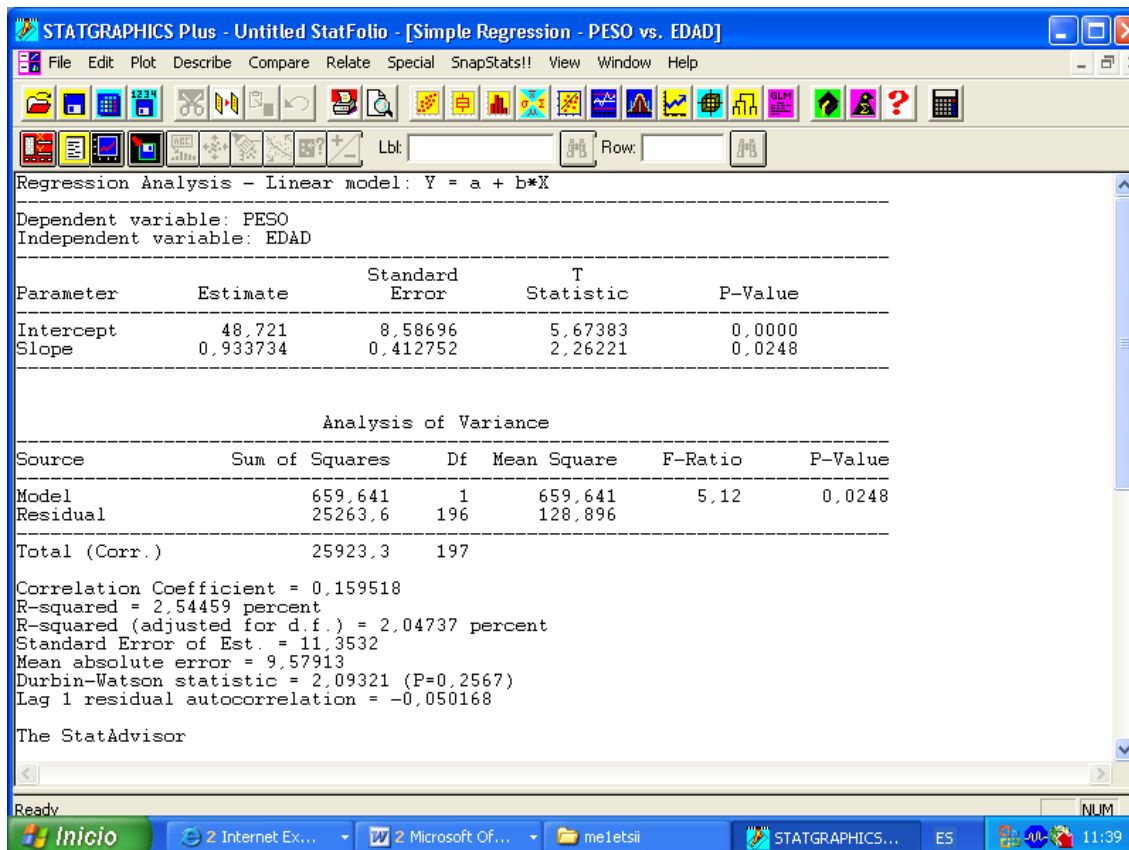
1.- Una empresa metalúrgica utiliza cadenas que le suministran tres proveedores (A, B y C). Con el objetivo de determinar la resistencia de las mismas, se tomó una muestra de 100 cadenas del proveedor A y se calculó para cada una de ellas el esfuerzo necesario hasta su rotura (medido en toneladas). Se tomó otra muestra aleatoria del proveedor B y se repitió el ensayo, y lo mismo con el proveedor C. A continuación se construyó un histograma de frecuencias con los datos de cada proveedor, los cuales se muestran a continuación:



A la vista de la figura, responder a las siguientes preguntas justificando la respuesta:

- Indicar qué proveedor tiene mayor rango, y calcular aproximadamente su valor. (1 punto)
- Indicar qué proveedor tiene mayores parámetros de posición. (0,5 puntos)
- ¿Qué modelo teórico crees que sería conveniente para modelizar la distribución de los datos del proveedor A? (0,5 puntos)
- ¿Por qué la escala vertical de frecuencias alcanza mayores valores en el proveedor B que en el A? (1 punto)
- Los proveedores A y B suministran las cadenas a igual precio. A la vista de la figura, ¿sugerirías que la empresa prescindiera de uno de los dos proveedores? (1 punto)

2.- Queremos realizar un estudio para analizar si existe una relación entre el peso de los hombres y su edad. Tras seleccionar una muestra representativa de un conjunto de individuos de una edad entre 19 y 30 años hemos realizado un análisis cuyo resultado es el siguiente:



A la vista de los resultados:

- ¿Consideras que existe una correlación estadísticamente significativa entre la edad de un hombre y su peso? Justifica razonadamente tu respuesta utilizando lo que te muestra el Statgraphics para realizar tus argumentos. Utilizar un α del 5% (1 punto)
- ¿Cuál es la interpretación y el valor del coeficiente de determinación? (1 punto)
- ¿Cuál es el valor de la pendiente de la recta y su interpretación práctica? (1 punto)

3.- A partir de la información bursátil que proporciona el mercado financiero secundario español se ha obtenido la siguiente información sobre 75 empresas:

Frequency Table

Rentabilidad*	PER MENOR 20	PER 20 A 40	PER MAS DE 40**	Row Total
Menos del 2%	5 6,67%	6 8,00%	10 13,33%	21 28,00%
Entre 2 y 3%	12 16,00%	8 10,67%	5 6,67%	25 33,33%
Más del 3%	7 9,33%	10 13,33%	12 16,00%	29 38,67%
Column Total	24 32,00%	24 32,00%	27 36,00%	75 100,00%

Cell contents:

Observed frequency

Percentage of table

Chi-Square Test

Chi-Square	Df	P-Value
5,93		0,2043

*La rentabilidad se mide en términos del porcentaje que representan los dividendos repartidos por la empresa respecto al valor de sus acciones.

** El PER (price earning ratio) mide la relación entre el precio y el dividendo por acción, o dicho de otra forma, el número de años que, con el reparto de dividendos actual, tardaríamos en recuperar la inversión realizada.

Para esta tabla de contingencia se pide:

- ¿Cuántos grados de libertad corresponden al valor de la distribución χ^2 obtenida? Razonar la respuesta. (1,5 puntos)
- ¿Existe alguna relación entre el PER y la rentabilidad a la vista de los resultados? Razonar la respuesta. (1,5 puntos)

APELLIDOS, NOMBRE: _____

Firma: _____

GRUPO: _____

BLOQUE III. PROBLEMAS (90 MINUTOS)**Cada problema son 10 puntos. La nota del bloque es la media de las notas de los problemas.**

1.- Uno de los procesos clave a la hora de recolectar las naranjas de un campo, por medio de un robot, consiste en que éste las clasifique inicialmente en dos grupos bien diferenciados: naranja de primera o de baja calidad. El robot establece dicha clasificación, calculando la característica L del espacio de color CIELAB, que nos da una idea de la luminosidad que emite dicha naranja. Se dispone de dos cintas: “cinta 1” y “cinta 2”, de manera que las naranjas clasificadas por el robot como de primera serán enviadas a la “cinta 1” y las clasificadas como de baja calidad serán enviadas a la “cinta 2”. Se sabe que la proporción de naranjas en la “cinta 1” es del 70% sabiendo además que en la “cinta 1” hay un 20% de naranjas de baja calidad y en la “cinta 2” hay un 90% de naranjas de baja calidad.

- a) ¿Cuál es la probabilidad a priori de ser naranja de primera? (2 puntos)
- b) ¿Cuál es la probabilidad de que una naranja de baja calidad caiga en la “cinta 1”? (2 puntos)
- c) Si nos vamos a la “cinta 1” y cogemos 6 naranjas, ¿cuál es la probabilidad de obtener más de 3 naranjas de primera? (3 puntos)
- d) Si tomamos una muestra de 3 naranjas de la cinta 1, ¿cuál será el valor esperado de naranjas de primera en dicha muestra? (3 puntos)

2.- La comunidad de propietarios de un edificio de Valencia ha decidido instalar unos paneles solares en su tejado para generar la energía eléctrica que consumen los vecinos. El presidente de la comunidad, que tiene amplios conocimientos de estadística, ha llevado a cabo algunas investigaciones y ha concluido que el número de horas diarias de sol (X) y la producción de energía (Y, medida en kilovatios hora) siguen una distribución normal bidimensional con los siguientes parámetros:

$$\bar{\mu} = \begin{pmatrix} 10 \\ 7 \end{pmatrix} \quad \mathbf{V} = \begin{pmatrix} 4 & \text{COV}(x, y) \\ \text{COV}(x, y) & 2,5 \end{pmatrix} \quad \rho = 0,92$$

- a) Si en un día de diciembre hay 5 horas diarias de sol, ¿cuál será la producción de energía eléctrica esperada para ese día? (2 puntos)
- b) El presidente pretende abastecer a la comunidad únicamente con la energía eléctrica generada por los paneles. Si dicha comunidad requiere de una producción diaria de al menos 3,5 kilovatios hora, ¿cuál es la probabilidad de que en un día de diciembre con 5 horas de sol no se cumpla este requerimiento? (4 puntos)
- c) ¿Para qué duración de horas diarias de sol se suministrará una producción de energía diaria de al menos 5 kilovatios hora con una probabilidad del 90%? (4 puntos)

3.- Se quiere comparar la exactitud (igualdad de medias) y la precisión (igualdad de varianzas) de dos métodos de análisis. Para ello se analiza repetidamente una misma muestra de un sustrato obteniéndose los siguientes valores en mg/100cc.

Método A: 24; 25; 24; 26; 27; 25; 24; 26; 27

Método B: 24; 24; 25; 22; 26; 22

Se pide:

- a) ¿Qué conclusiones podemos obtener respecto de la exactitud y de la precisión con un nivel de significación $\alpha=0,01$? (6 puntos)
- b) Si en el test para comparar la exactitud tomamos un nivel de significación $\alpha=0,40$ ¿Qué decisión habría que tomar? ¿Han cambiado las conclusiones respecto al apartado a)? ¿Por qué? (4 puntos)

SOLUCIONES DEL EXAMEN DE ME1 DE ENERO DE 2007**SOLUCIÓN BLOQUE I PREGUNTA 1:**

$X = n^\circ$ de contenedores que no cumplen $= H(15, 3, 2/15)$

$$P(X=1) = \frac{\binom{15 \cdot 2 / 15}{1} \cdot \binom{15 \cdot 13 / 15}{2}}{\binom{15}{3}} = \frac{\binom{2}{1} \cdot \binom{13}{2}}{\binom{15}{3}} = \frac{2 \cdot 78}{455} = 0,343$$

SOLUCIÓN BLOQUE I PREGUNTA 2:

$X =$ Tiempo entre accesos $= EXP(\lambda = 1/0,5 = 2) \text{ min.}$

$$P(0,5 < X < 1) = e^{-2 \cdot 0,5} - e^{-2 \cdot 1} = 0,233$$

SOLUCIÓN BLOQUE I PREGUNTA 3:

Al estimar p el error que se produce es,

$$\varepsilon = z_{\alpha/2} \sqrt{\frac{pq}{n}}$$

Despejando n se tiene

$$n = \frac{z_{\alpha/2}^2 pq}{\varepsilon^2}$$

Para $\alpha = 0,15$, $z_{\alpha/2} = 1,44$. Si conocemos que $p = 0,025$, el tamaño muestral es,

$$n = \frac{1,44^2 \cdot 0,025 \cdot 0,975}{0,03^2} = 56,2$$

Esto es, $n = 57$.

SOLUCIÓN BLOQUE I PREGUNTA 4:

La zona de aceptación del contraste es:

$$X_0(s) = \left\{ \bar{x} / \mu_0 - z_{\alpha/2} \frac{\sigma_0}{\sqrt{n}} \leq \bar{x} \leq \mu_0 + z_{\alpha/2} \frac{\sigma_0}{\sqrt{n}} \right\}$$

Para los valores del enunciado, la zona de aceptación es:

$$Z.A. = \left\{ 20 - 2,58 \frac{1,5}{\sqrt{9}} \leq \bar{x} \leq 20 + 2,58 \frac{1,5}{\sqrt{9}} \right\} = \{18,71 \leq \bar{x} \leq 21,29\}$$

El valor de β es:

$$\beta = P(18,71 \leq \bar{x} \leq 21,29 / \mu = 22) = \Phi\left(\frac{21,29 - 22}{1,51\sqrt{9}}\right) - \Phi\left(\frac{18,71 - 22}{1,51\sqrt{9}}\right) = \Phi(-1,42) - \Phi(-6,58) = 0,078$$

SOLUCIÓN BLOQUE I PREGUNTA 5:

$$\int_0^3 (7 - 2x) dx = k \left[7x - x^2 \right]_0^3 = 12k \quad \Rightarrow I = 12k \Rightarrow k = \frac{I}{12}$$

La media viene dada por:

$$E[X] = \int_0^3 x \cdot \frac{(7 - 2x)}{12} dx = 1,125$$

Para determinar la probabilidad de que tres observaciones independientes sean menores que la media, calculamos la probabilidad de que una observación sea menor que la media.

$$P(X < 1,125) = \int_0^{1,125} \frac{(7 - 2x)}{12} dx = 0,5508$$

Puesto que las tres observaciones son independientes la probabilidad pedida es igual a $(0,5508)^3 = 0,167$

SOLUCIÓN BLOQUE I PREGUNTA 6:

$$\begin{aligned} a) P(D \cup T) &= P(D) + P(T) - P(D \cap T) = P(D) + P(T) - P(D)P(T|D) = \\ &= 0,08 + 0,27 - 0,08 \times 0,68 = 0,2956 \end{aligned}$$

$$b) P(T|E) = \frac{P(T \cap E)}{P(E)}$$

puesto que:

$$P(T \cap E) = P(T) + P(E) - P(T \cup E) = 0,27 + 0,62 - 0,72 = 0,17$$

Tenemos

$$P(T|E) = \frac{P(T \cap E)}{P(E)} = \frac{0,17}{0,62} = 0,2742$$

SOLUCIONES BLOQUE II PREGUNTA 1:

a) *Cálculo aproximado del rango:*

$$\text{Proveedor A} = 2,3 - 1,7 = 0,6$$

$$\text{Proveedor B} = 2,2 - 1,9 = 0,3$$

$$\text{Proveedor C} = 2,5 - 2,25 = 0,25$$

Se deduce claramente que el proveedor A es el de mayor rango.

b) *A la vista de la figura, la distribución es aproximadamente normal, por lo que los parámetros de posición habituales (media y mediana) coinciden. El proveedor C es el que tiene mayor media (aproximadamente, 2,35), por lo que éste es el que tiene los mayores parámetros de posición.*

c) *La distribución normal parece adecuada para modelizar la distribución de los datos, ya que los tres histogramas son aproximadamente simétricos y tienen forma de una campana de Gauss, es decir, se asemeja a la función de densidad de la distribución normal.*

d) *Porque el histograma del proveedor A se ha construido con 9 intervalos, mientras que en el caso del proveedor B éste se ha construido con un número de intervalos menor.*

e) *La resistencia media en los proveedores A y B es bastante similar. En cambio, la desviación típica es claramente menor en el proveedor B. Es preferible utilizar cadenas de B ya que de este modo se reduce la proporción de cadenas de baja resistencia, que es lo que conviene.*

SOLUCIÓN BLOQUE II PREGUNTA 2:

a) *Podemos considerar que sí que existe una correlación significativa entre la edad de un hombre y su peso porque el p-valor ($= 0,0248$) es menor que el valor de α ($=0,05$).*

b) *El R^2 se interpreta como el porcentaje de variabilidad determinada por el modelo. En este caso, a pesar de que el modelo es significativo, éste sólo explica el 2,54459% de la variabilidad.*

c) *La **pendiente (Slope)** indica el incremento medio que se produce en el peso por unidad incrementada en la edad. El signo indica la relación que existe entre edad y peso. En este caso es una relación positiva. Así pues, en este caso por cada unidad en que se incremente la edad el peso se incrementará en 0,933734 unidades.*

SOLUCIÓN BLOQUE II PREGUNTA 3:

a) *Los grados de libertad son el producto del número de filas menos uno por el número de columnas menos uno, y de aquí que sean 4 grados de libertad.*

b) *Los resultados obtenidos no nos permiten rechazar la hipótesis de independencia entre el PER y la rentabilidad de las acciones. La razón estriba en que existen otras formas de retribución al accionista a parte de los dividendos, como pueden ser las ampliaciones gratuitas de capital o la retribución indirecta al accionista mediante el aumento del valor de las acciones que cotizan en el mercado secundario (Bolsa).*

SOLUCIÓN BLOQUE III PROBLEMA 1:

a) ¿Cuál es la probabilidad a priori de ser naranja de primera?

$$C1 = \{\text{Cinta 1}\}$$

$$C2 = \{\text{Cinta 2}\}$$

$$P = \{\text{Naranja de primera}\}$$

$$M = \{\text{Naranja de baja calidad}\}$$

$$P(C1) = 0,7 \rightarrow P(C2) = 0,3$$

$$P(M | C1) = 0,2 \rightarrow P(P | C1) = 0,8$$

$$P(M | C2) = 0,9 \rightarrow P(P | C2) = 0,1$$

Teorema de la probabilidad total

$$P(P) = P(P | C1) \cdot P(C1) + P(P | C2) \cdot P(C2) = 0,8 \cdot 0,7 + 0,1 \cdot 0,3 = 0,59$$

b) ¿Cuál es la probabilidad de que una naranja de baja calidad caiga en la “cinta 1”?

Teorema de Bayes

$$P(C1 | M) = \frac{P(M | C1) \cdot P(C1)}{P(M)} = \frac{P(M | C1) \cdot P(C1)}{1 - P(P)} = \frac{0,2 \cdot 0,7}{1 - 0,59} = \frac{0,2 \cdot 0,7}{0,41} = 0,3415$$

c) Si nos vamos a la “cinta 1” y cogemos 6 naranjas. ¿Cuál es la probabilidad de obtener más de 3 naranjas de primera?

$$\begin{aligned} X &\sim B(n = 6, p = P(P | C1)) = B(n = 6, p = 0,8) \\ P(X > 3) &= 1 - P(X \leq 3) = 1 - (P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)) = \\ &= 1 - \left(\binom{6}{0} 0,8^0 0,2^6 + \binom{6}{1} 0,8^1 0,2^5 + \binom{6}{2} 0,8^2 0,2^4 + \binom{6}{3} 0,8^3 0,2^3 \right) = 0,901 \end{aligned}$$

d) ¿Si ahora las muestras son de 3 naranjas, cuál será el valor esperado de naranjas de primera en la “cinta 1” en cada muestra de 3 naranjas?

$$\begin{aligned} X &\sim B(n = 3; p = P(P | C1)) = B(n = 3; p = 0,8) \\ E(X) &= \sum_{0 \leq i \leq 3} i P(X = i) = \\ &= 0 \cdot P(X = 0) + 1 \cdot P(X = 1) + 2 \cdot P(X = 2) + 3 \cdot P(X = 3) = \\ &= 0 \cdot \binom{3}{0} 0,8^0 0,2^3 + 1 \cdot \binom{3}{1} 0,8^1 0,2^2 + 2 \cdot \binom{3}{2} 0,8^2 0,2^1 + 3 \cdot \binom{3}{3} 0,8^3 0,2^0 = \\ &= 0 \cdot \frac{3!}{0!(3-0)!} 0,8^0 0,2^3 + 1 \cdot \frac{3!}{1!(3-1)!} 0,8^1 0,2^2 + 2 \cdot \frac{3!}{2!(3-2)!} 0,8^2 0,2^1 + 3 \cdot \frac{3!}{3!(3-3)!} 0,8^3 0,2^0 = \\ &= 0 \cdot 1 \cdot 0,8^0 0,2^3 + 1 \cdot \frac{3}{1} \cdot 0,8^1 0,2^2 + 2 \cdot \frac{6}{2} 0,8^2 0,2^1 + 3 \cdot 1 \cdot 0,8^3 0,2^0 = \\ &= 0 + 3 \cdot 0,8^1 0,2^2 + 6 \cdot 0,8^2 0,2^1 + 3 \cdot 0,8^3 0,2^0 = 2,4 \text{ naranjas de primera} \end{aligned}$$

SOLUCIÓN BLOQUE III PROBLEMA 2:

a) La producción de energía eléctrica esperada para un día con 5 horas de sol es de:

$$E(Y|X=5) = 7 + 0,92\sqrt{\frac{2,5}{4}}(5-10) = 3,363 \text{ kWh.}$$

b) $Y|X=5 \equiv N(E(Y|X=5)=3,363; \sqrt{D^2(Y|X=5)} = \sqrt{2,5(1-0,92^2)} = 0,62)$

$$P(Y < 3,5 | X=5) = P(Z < \frac{3,5 - 3,363}{0,62}) = P(Z < 0,221) = 0,587$$

La probabilidad de no producir los 3,5 kWh requeridos, en un día con 5 horas de sol, es de 0,587.

$$c) E(y/x) = 7 + 0,92\sqrt{\frac{2,5}{4}}(x-10) = 0,727x - 0,273$$

Lo que nos piden es que $P(y > 5/x) = 0,90$. Tipificando se tiene,

$$P(z > \frac{5 - (0,727x - 0,273)}{0,62}) = P(z > \frac{5,273 - 0,727x}{0,62}) = 0,90$$

Para $z_{0,90} = -1,282$, de donde

$$\frac{5,273 - 0,727x}{0,62} = -1,282$$

Despejando, se tiene que $x = 8,346$ horas diarias de sol.

SOLUCIÓN BLOQUE III PROBLEMA 3:

Con los datos del problema obtenemos

$$X_A = 25,33 \quad S_{n_A-1}^2 = 1,50$$

$$X_B = 23,83 \quad S_{n_B-1}^2 = 2,57$$

El test planteado para comparar la exactitudes $H_0: \mu_A = \mu_B$ contra $H_1: \mu_A \neq \mu_B$

Y la regla de decisión es aceptar si

$$|\bar{x}_A - \bar{x}_B| \leq t_{n_A+n_B-2}^{\alpha/2} S \sqrt{\frac{1}{n_A} + \frac{1}{n_B}} \quad \text{con} \quad S = \sqrt{\frac{(n_A-1)S_{n_A-1}^2 + (n_B-1)S_{n_B-1}^2}{n_A + n_B - 2}}$$

Siendo la $t_{13}^{0,005} = 3,012$ y la $t_{13}^{0,20} = 0,87$.

Haciendo operaciones

$1,5 < 2,1946$ por lo que debemos aceptar la igualdad de medias poblacionales y por ello la igualdad de exactitud de ambos métodos con el nivel de significación previsto.

El test para comprobar a igualdad de la precisión es $H_0: \sigma_A^2 = \sigma_B^2$ contra $H_1: \sigma_A^2 \neq \sigma_B^2$

Y la regla de decisión es aceptar la hipótesis nula si

$$\frac{S_i^2}{S_j^2} \leq F_{n_i-1, n_j-1}^\alpha \text{ siendo } i \text{ la muestra de mayor } S^2$$

Como $2,57/1,5 = 1,71 < F_{5,8}^{0,01} = 6,63$ debemos aceptar la igualdad de varianzas y por la tanto los dos métodos presentan la misma precisión

En el apartado b) cuando $\alpha = 0,40$ el valor de $t_{n_A+n_B-2}^{\alpha/2} S \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}$ es de 0,634

Como la diferencia de medias muestrales es $1,5 > 0,634$ debemos rechazar la hipótesis de igualdad de medias.

Al fijar un α mucho mayor en este caso, estamos dispuestos a cometer un error de primera especie mayor, es decir, a dar como significativas diferencias más pequeñas. Por eso en este caso salen diferentes las medias poblacionales

EXAMEN DE JUNIO 2006, CON SOLUCIÓN

APELLIDOS, NOMBRE: _____ FIRMA: _____

PROFESOR: _____

Cada pregunta es 1 punto. Si una respuesta no está justificada, no se puntuará.

TIEMPO: 1 HORA Y 30 MINUTOS

CUESTIONES (40% nota)

C1.- Para adaptar la frecuencia de paso de los autobuses de la línea 90 a la demanda real de los viajeros en horas punta, los técnicos de la E.M.T. necesitan estimar con un error máximo de ± 5 minutos el tiempo medio que un autobús de dicha línea tarda en hacer una ronda completa en horario de 8 a 9 y de 13 a 15 horas. Para ello deciden medir el tiempo de n recorridos elegidos al azar en dicho horario y obtener la media aritmética que será la estimación de la "media poblacional" que utilizarán en la planificación. Si admitimos que la dispersión de los tiempos de recorrido de una ronda es $\sigma = 12$ minutos, y consideramos un nivel de confianza para la estimación de 0,99 ¿Cuántos recorridos deberemos medir?

- a) 24
- b) 128
- c) 8
- d) 38

C2.- El precio de venta de un determinado artículo es de 3€/unidad. El número de unidades vendidas en un día es una variable aleatoria con función de densidad:

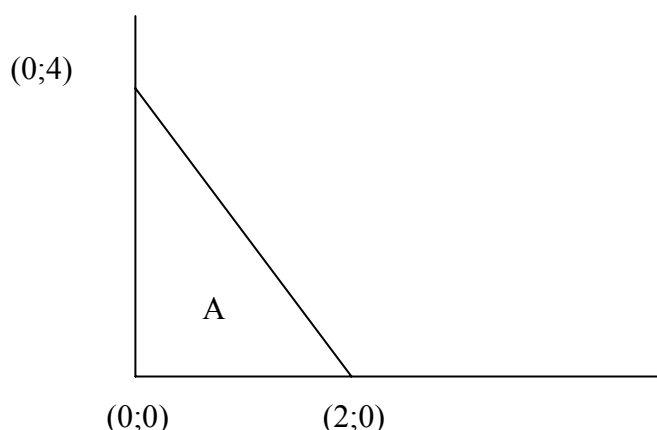
$$f(x) = 10^{-6} \left(\frac{4000}{3} - \frac{2}{3}x \right) \quad \text{para } 0 \leq x \leq 1000$$
$$f(x) = 0 \quad \text{en otro caso}$$

Si se venden menos de 500 unidades, el coste por unidad es 2€, pero a partir de 500 unidades el coste por unidad viene en función del número de unidades vendidas según la expresión de

$$\text{Coste}(x) = 3 - \frac{x}{500}.$$

Calcular el valor medio del beneficio diario.

C3.- La variable aleatoria bidimensional (x,y) se encuentra distribuida uniformemente en el recinto A que se indica en la figura de abajo.



La función de distribución en el punto $(1,5; 3)$ vale:

- a) 0
- b) $3/4$
- c) $1/2$
- d) $14/16$

C4.- Para estimar la proporción P de alumnos que se van a presentar al examen de una asignatura en la que hay matriculados muchísimos alumnos y reservar las aulas necesarias, un profesor escribe un correo a n_1 alumnos elegidos al azar de la lista de alumnos y averigua el número de ellos v_1 que se van a presentar. Otro profesor, que no estaba al tanto de lo que hacía su colega, manda otro correo a otros n_2 alumnos y determina el número de ellos v_2 que se presentarán al examen (cada alumno puede ser elegido indistintamente por uno o por ambos profesores). Finalmente, y para aprovechar toda la información obtenida, deciden estimar P mediante la expresión

$$\hat{P} = \frac{\frac{v_1}{n_1} + \frac{v_2}{n_2}}{2} = \frac{p_1 + p_2}{2}$$

El estimador $\hat{P}$ utilizado es

- a) Inssegado y con varianza igual a $\frac{P(1-P)}{2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$
- b) Sesgado y con varianza igual a $\frac{P(1-P)}{2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
- c) Inssegado y con varianza igual a $\frac{(n_1 + n_2)P(1-P)}{4n_1n_2}$
- d) Inssegado y con varianza igual a $P(1-P) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$

C5.- Una empresa envasadora de zumo de tomate indica en la etiqueta del bote que el contenido del mismo es de 300 cc. Si representamos por X el volumen real de zumo de un bote escogido al azar lo deseable es (razonar la respuesta):

- a) $\mu_x=300$ y σ_x lo mas grande posible
- b) $\mu_x=300$ y σ_x lo mas pequeña posible
- c) $\mu_x=300$ y σ_x negativo
- d) Da igual el valor de σ_x lo importante es que $\mu_x=300$.

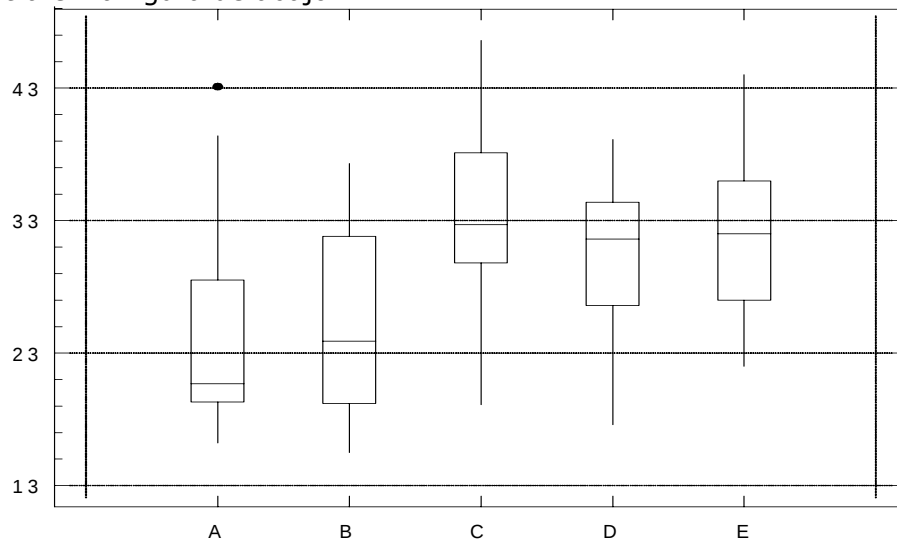
C6.- La ganancia de un concesionario de coches en unidades de 1000€ en la venta de un nuevo automóvil es $Y=X^2$ donde X es una variable aleatoria de función de densidad

$$f(x) = \begin{cases} 2(1-x) & 0 < x < 1 \\ 0 & \text{otros} \end{cases}$$

La función de densidad de la variable Y viene dada por:

- a) $y^{-\frac{1}{2}} \quad -1 < y < 1$
- b) $y^{-\frac{1}{2}} - 1 \quad 0 < y < 1$
- c) $\sqrt{y} \quad -1 < y < 1$
- d) $y^{\frac{1}{2}} \quad 0 < y < 1$

C7.- Se ha obtenido la concentración del colesterol HDL en sangre en cinco grupos de pacientes, representándose en un diagrama de Caja y Bigotes Múltiples los resultados, tal y como se muestra en la figura de abajo.



Si estudiamos la dispersión

- a) El grupo B son los pacientes que presentan con mayor dispersión
- b) Todos los grupos de pacientes tienen la misma dispersión
- c) Los grupos C o D son los que tienen mayor dispersión
- d) Con éste diagrama no podemos estudiar la dispersión de los datos

C8.- Se ha procedido a extraer de un lote grande de bolsas de pipas de 100 grs. una muestra aleatoria de 100 bolsas.

A la vista de los resultados:

Summary Statistics for GRS100

Count = 100
Average = 102,634
Median = 102,82
Mode = 104,88
Geometric mean = 102,539
Variance = 19,6604
Standard deviation = 4,43401
Standard error = 0,443401
Minimum = 92,35
Maximum = 115,0
Range = 22,65
Lower quartile = 99,775
Upper quartile = 105,325
Interquartile range = 5,55
Skewness = 0,0939556
Std. skewness = 0,383572
Kurtosis = 0,220815
Std. kurtosis = 0,450736
Coeff. of variation = 4,32023%
Sum = 10263,4

95,0% confidence interval for standard deviation: [3,89309;5,15088]

t-test

Null hypothesis: mean = 101,0
Alternative: not equal

Computed t statistic = 3,68448
P-Value = 0,000374054
alpha = 0,05.

¿Cuál de las siguientes afirmaciones son verdaderas?

- I) Con los resultados anteriores podemos saber si los datos proceden de una distribución normal.
II) Podemos asumir una media poblacional de 101 grs. con un nivel de confianza del 95%.
III) Podemos asumir una varianza de 16 grs². con un nivel de confianza del 95%.
- a) La I y II
 - b) La II y la III.
 - c) La III.
 - d) Todas son falsas.

APELLIDOS, NOMBRE: _____ FIRMA: _____

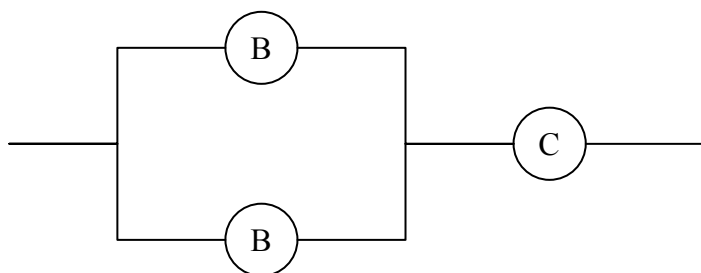
PROFESOR: _____

Cada problema son 10 puntos.**TIEMPO: 2 HORAS****PROBLEMAS (60% de la nota)**

P1.- Para el montaje de un determinado sistema con bombillas se dispone de 3 proveedores, A, B y C. Se sabe que el proveedor A produce con un 5% de bombillas defectuosas, el B un 3% y el C un 1%. El 10 % de las bombillas son del proveedor A, el 20 del B y el 70% del C. Todas las bombillas van a un almacén común.

a) Si cogemos 3 bombillas de un mismo proveedor, sin saber exactamente cuál es, y resultan que hay 2 defectuosas, ¿de qué proveedor es más probable que procedan?

b) Si las bombillas del proveedor B sabemos que su duración sigue una exponencial de media 100 horas, y que las del proveedor C siguen una normal de media 130 y sigma 20, y si cogemos dos bombillas en paralelo del proveedor B y las situamos junto a una bombilla en serie del proveedor C, tal y como se muestra en el esquema siguiente



¿Cuál es la probabilidad de que haya luz en este sistema al cabo de 150 horas de funcionamiento?

P2.- Una determinada empresa del sector eléctrico gestiona un parque eólico situado en la Comunidad Valenciana. Dicho parque está constituido por 50 torres eólicas. Sean $X_1, X_2, \dots, X_{50}$ las potencias producidas por cada una de las torres en un momento determinado. Además sabemos que cada X_i sigue una distribución normal con media 1000 Kw y varianza 2500 Kw²

- ¿Cuál es la probabilidad de que la potencia producida por una torre sea mayor que 1050 (unidades)Kw?
- Suponiendo que las variables $X_1, \dots, X_{50}$ son independientes, ¿cuál es la probabilidad de que haya exactamente 8 torres que estén produciendo más de 1050 Kw?
- Calcular la probabilidad de que más de 10 torres estén produciendo más de 1050 Kw.
- Si P es la potencia total producida por el parque en un momento dado, y si seguimos suponiendo que las variables $X_1, \dots, X_{50}$ son independientes, ¿Cuánto vale la media y la varianza de P ? ¿cuál es su distribución?

P3.- La empresa de seguridad que realiza la vigilancia en el campus de la UPV quiere comprobar si un nuevo pegamento *aumenta* la adhesividad de las "pegatinas" que colocan los agentes en los cristales de los coches mal aparcados. Para ello hace el siguiente experimento: Coloca 4 pegatinas de las actuales y 4 con el nuevo pegamento sobre un cristal. Después de dos horas de secado se arrancan las pegatinas midiendo el esfuerzo a tracción en condiciones normalizadas y sin que se rompan los papeles. Los resultados obtenidos, expresados en Newton, han sido:

Pegatinas actuales 4.2, 4.1, 4.4, 4.6

Pegatinas con nuevo pegamento 4.1, 4.8, 4.6, 5.0

Se asume que la resistencia del pegado tiene en los dos casos una distribución Normal con la misma desviación típica $\sigma = 0,2$ Newton

- Plantear las hipótesis que se deberán contrastar
- ¿Cuál sería el resultado del contraste empleando los resultados del experimento y considerando un nivel de significación de $\alpha = 0,05$?
- Si la media de la resistencia del nuevo pegamento fuera en realidad 0,5 Newton superior a la media del pegamento actual ¿Cuál sería la probabilidad β de cometer el error de segunda especie del contraste anterior?
- En el apartado b) y con los datos del experimento actual hemos tomado una decisión ¿Podemos estar cometiendo un error de tipo II o de segunda especie con la decisión que hemos tomado? ¿por qué?

P4.- Se han estudiado los gastos del departamento comercial (en millones de euros) de una muestra de 8 empresas de un determinado sector (X) y los beneficios netos obtenidos (en millones de euros) por dichas empresas (Y), obteniéndose la siguiente información:

$$\sum x_i = 24; \sum x_i y_i = 64; \sum y_i = 40; S_x^2 = 12; S_y^2 = 6$$

Supongamos que ambas variables siguen una distribución normal bivalente:

- Calcular el coeficiente de correlación e interpretar el resultado.
- Obtener $E(Y|X=x)$
- ¿Cuál sería el beneficio neto esperado para una empresa cuyo departamento comercial ha gastado 4 millones de euros?

SOLUCIONES AL EXAMEN DE ME1 DE JUNIO 2006**SOLUCIÓN C1:**

Se trata de la estimación de la media de una población Normal de varianza conocida ($\sigma^2=22$)

El error absoluto de la estimación en este caso es

$$e = z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \text{ de donde}$$

$$n = \left(\frac{z_{\alpha/2} \sigma}{e} \right)^2 = \left(\frac{2,57 \cdot 12}{5} \right)^2 = 38 \text{ recorridos}$$

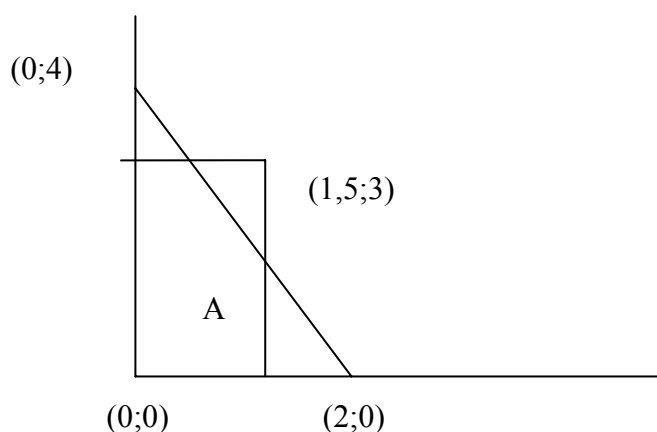
SOLUCIÓN C2:

Llamando a x el número de unidades vendidas en un periodo determinado, se cumple que:

- Si $x \leq 500$ el beneficio es $B_1 = (3-2)x = x$
- Si $x \geq 500$ el beneficio es $B_2 = (3 - (3 - \frac{x}{500})) \cdot x = \frac{x^2}{500}$

El valor medio del beneficio total viene dado por:

$$\begin{aligned} E[B(x)] &= \int_0^{500} x \cdot 10^{-6} \left(\frac{4000}{3} - \frac{2}{3}x \right) dx + \int_{500}^{1000} \frac{x^2}{500} \cdot 10^{-6} \left(\frac{4000}{3} - \frac{2}{3}x \right) dx = \\ &= 10^{-6} \left[\frac{4000}{3} \frac{x^2}{2} - \frac{2}{3} \frac{x^3}{3} \right]_0^{500} + \frac{10^{-6}}{500} \left[\frac{4000}{3} \frac{x^3}{3} - \frac{2}{3} \frac{x^4}{4} \right]_{500}^{1000} = \\ &= 10^{-6} \left[\frac{4000}{3} \frac{500^2}{2} - \frac{2}{3} \frac{500^3}{3} \right] + \frac{10^{-6}}{500} \left[\frac{4000}{3} \frac{1000^3}{3} - \frac{2}{3} \frac{1000^4}{4} - \frac{4000}{3} \frac{500^3}{3} + \frac{2}{3} \frac{500^4}{4} \right] = \\ &= 604,167 \end{aligned}$$

SOLUCIÓN C3

La función de distribución en el punto $(1,5;3)$ viene dada por la cantidad de masa que hay por debajo y a la izquierda de este punto.

La función de densidad, bajo la condición de que todo el volumen es la unidad vale $f(x,y) = 1/4$.

La función de distribución viene dada por:

$$F_{x,y}(1,5;3) = 1 - V_1 - V_2 = 1 - \frac{1 \cdot 0,5}{2} \frac{1}{4} - \frac{1 \cdot 0,5}{2} \frac{1}{4} = \frac{14}{16}$$

SOLUCIÓN C4:

Sabemos que, en general, $E(p)=P$ y que $D^2(p)=P(1-P)/n$ luego, como las dos muestras son de la misma población y en ella la proporción de alumnos que se van a presentar es P se cumple:

$$E(p_1)=P \text{ y que } D^2(p_1)=P(1-P)/n_1$$

$$E(p_2)=P \text{ y que } D^2(p_2)=P(1-P)/n_2$$

Además, p_1 y p_2 son independientes según las premisas del enunciado, por lo tanto

$$E(\hat{P}) = E\left(\frac{p_1 + p_2}{2}\right) = \frac{1}{2}(E(p_1) + E(p_2)) = \frac{1}{2}(P + P) = P \text{ (estimador insesgado)}$$

$$D^2(\hat{P}) = D^2\left(\frac{p_1 + p_2}{2}\right) = \frac{1}{2^2} D^2(p_1 + p_2) = \frac{1}{4}(D^2(p_1) + D^2(p_2)) = \frac{1}{4}\left[\frac{P(1-P)}{n_1} + \frac{P(1-P)}{n_2}\right] = \frac{P(1-P)}{4}\left(\frac{1}{n_1} + \frac{1}{n_2}\right) = \frac{(n_1 + n_2)P(1-P)}{4n_1n_2}$$

SOLUCIÓN C5:

La b)

SOLUCIÓN C6:

En el rango de valores de X ($0 < X < 1$) la relación definida entre X e Y es biunívoca, por lo tanto la función de densidad de la transformada será

$$f_y(y) = f_x(x(y)) \left| \frac{dx}{dy} \right|$$

Como $Y=X^2$ tendremos la función inversa $x(y)$ de la forma $X=+Y^{1/2}$ por ser $X>0$ siempre, luego

$$X(y)=y^{1/2}$$

$$\frac{dx}{dy} = \frac{1}{2} y^{-1/2}$$

Además, por la relación establecida entre las variables la y variará entre $0 < y < 1$

Sustituyendo

$$f_y(y) = 2(1 - y^{1/2}) \frac{1}{2} y^{-1/2} = y^{-1/2} - 1 \text{ para } 0 < y < 1$$

SOLUCIÓN C7:

La a). La caja del grupo B es la mayor de todos los grupos, ta y como se aprecia e la figura de arriba.

SOLUCIÓN C8:

La c)

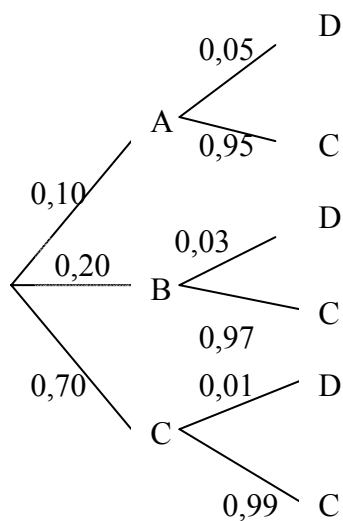
La I) es falsa. No podemos con estos resultados si los datos proceden de una distribución. Debemos aplicar un test de Chi-cuadrado o semejante.

La II) es falsa ya que el estadístico t es muy grande, 3,68, y su p-valor es menor del 0,05, lo cual indica que debemos rechazar la hipótesis nula de que la media poblacional es de 101 grs., con un nivel de significación del 5%.

La III) es cierta, ya que si la varianza es 16, entonces la desviación típica es 4, y éste valor queda dentro del intervalo de confianza al 95% para la desviación típica, [3,89;5,15], por lo que aceptamos la hipótesis nula de que la desviación típica es 4.

SOLUCIÓN P1:

a) El árbol de probabilidades es.



Si x es el número de defectuosas de 3 bombillas tomadas del mismo proveedor, esta variable aleatoria sigue una distribución binomial, de manera que:

$$P(x = 2 / A) = 3 \cdot 0,05^2 \cdot 0,95$$

$$P(x = 2 / B) = 3 \cdot 0,03^2 \cdot 0,97$$

$$P(x = 2 / C) = 3 \cdot 0,01^2 \cdot 0,99$$

Por el teorema de Bayes:

$$P(A/x=2) = \frac{0,10 \cdot 3 \cdot 0,05^2 \cdot 0,95}{0,10 \cdot 3 \cdot 0,05^2 \cdot 0,95 + 0,20 \cdot 3 \cdot 0,03^2 \cdot 0,97 + 0,70 \cdot 3 \cdot 0,01^2 \cdot 0,99} = 0,481$$

$$P(B/x=2) = \frac{0,20 \cdot 3 \cdot 0,03^2 \cdot 0,97}{0,10 \cdot 3 \cdot 0,05^2 \cdot 0,95 + 0,20 \cdot 3 \cdot 0,03^2 \cdot 0,97 + 0,70 \cdot 3 \cdot 0,01^2 \cdot 0,99} = 0,354$$

$$P(C/x=2) = \frac{0,70 \cdot 3 \cdot 0,01^2 \cdot 0,99}{0,10 \cdot 3 \cdot 0,05^2 \cdot 0,95 + 0,20 \cdot 3 \cdot 0,03^2 \cdot 0,97 + 0,70 \cdot 3 \cdot 0,01^2 \cdot 0,99} = 0,14$$

Lo más probable es que sean del proveedor A.

b) La probabilidad de que una bombilla de B dure más de 150 horas es:

$$p_B = P(B > 150) = e^{-\frac{150}{100}} = 0,223$$

Y la probabilidad que una bombilla de C dure más de 150 horas viene dada por:

$$p_C = P(C > 150) = P(N(0,1) \geq \frac{150-130}{20}) = P(z \geq 1) = 0,159$$

La probabilidad de que haya luz es:

$$p = (1 - (1 - p_B)^2) \cdot p_C = (1 - (1 - 0,223)^2) \cdot 0,159 = 0,063$$

Para que haya luz en el sistema se debe cumplir que haya luz en la parte de las B y que haya luz en la parte de C. Que haya luz en la parte de B equivale a 1- que no haya luz. Y no habrá luz en la parte de B cuando se cumpla que no haya luz en ninguna de las dos bombillas.

SOLUCIÓN P2:

a) $X \equiv N(1000;50)$

$$P(X > 1050) = P(N(1000;50) > 1050) = P\left(Z > \frac{1050-1000}{50}\right) = P(Z > 1) = 0,1587$$

b) $Y \equiv B(50;0,1587)$

$$P(Y = 8) = \binom{50}{8} \cdot 0,1587^8 \cdot (1 - 0,1587)^{42} = 0,1522$$

c) Nos piden $P(Y > 10)$ mediante la aproximación de la binomial a la normal tenemos $Y \approx N(7,935;2,584)$ por tanto:

$$P(Y > 10) = P(N(7,935;2,584) > 10) = P\left(Z > \frac{10-7,935}{2,584}\right) = P(Z > 0,799) = 0,212$$

d) $P_{\text{total}} = X_1 + X_2 + \dots + X_{50}$

Aplicando el teorema central del límite tenemos:

$$\mu_{P_{\text{total}}} = n \cdot \mu = 50 \cdot 1000 = 50000 \text{ Kw}$$

$$\sigma_{P_{\text{total}}} = \sqrt{n} \cdot \sigma = \sqrt{50} \cdot 50 = 353,5 \text{ Kw}$$

Por tanto la potencia total se distribuirá como una normal de parámetros $P_{\text{total}} \equiv N(50000;353,5)$.

SOLUCIÓN P3:

- a) Se trata de un test unilateral sobre las medias de dos poblaciones Normales

$H_0: \mu_1 = \mu_2$ (hipótesis nula de que la **diferencia es nula**)

$H_1: \mu_1 < \mu_2$ (hipótesis alternativa de que la media del nuevo pegamento es mayor)

- b) El contraste es

$$\text{Aceptar } H_0 \text{ si } -z_\alpha \leq \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \text{ según las tablas o también}$$

$$\text{Aceptar } H_0 \text{ si } \bar{x}_2 - \bar{x}_1 \leq z_\alpha \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 1,645 \sqrt{\frac{0,04}{4} + \frac{0,04}{4}} = 0,2326$$

Como $\bar{x}_2 - \bar{x}_1 = 4,625 - 4,325 = 0,30 > 0,2326 \Rightarrow$ **Rechazamos la hipótesis de igualdad de medias y admitimos que el nuevo pegamento es mas fuerte**

- c) En un contraste de hipótesis, $\beta = P(\text{Aceptar } H_0/H_1 \text{ cierta})$ y en nuestro caso concreto será

$$\beta = P((\bar{x}_2 - \bar{x}_1) \leq 0,2326 /_{(\mu_2 - \mu_1)=0,5}) = P(N(0,5, \sqrt{\frac{0,04}{4} + \frac{0,04}{4}}) \leq 0,2326) = 0,029$$

- d) El error de tipo II **solo se comete cuando se acepta la hipótesis nula**. Como en el caso del apartado b) **se rechaza la hipótesis nula**, no puede cometerse el error de tipo II dado que no aceptamos dicha hipótesis nula.

SOLUCIÓN P4:

- a) $r = -0,825$

La covarianza viene dada por

$$\text{cov}(x, y) = E(xy) - E(x)E(y) = \frac{1}{n} \sum_i x_i y_i - \frac{\sum x}{n} \frac{\sum y}{n} = \frac{64}{8} - \frac{24}{8} \cdot \frac{40}{8} = -7$$

Y el coeficiente de correlación es:

$$\rho(x, y) = \frac{\text{cov}(x, y)}{D(x)D(y)} = \frac{-7}{\sqrt{12 \cdot 6}} = -0,825$$

- b) $E(Y|X=x) = 6,75 - 0,583x$. Viene dado por la expresión

$$E(y/x) = \mu_y + \rho(x, y) \frac{D(y)}{D(x)} (x - \mu_x) = 5 - 0,852 \frac{\sqrt{6}}{\sqrt{12}} (x - 3)$$

- c) B^0 neto esperado = 4,42

EXAMEN DE FEBRERO 2006, CON SOLUCIÓN

APELLIDOS, NOMBRE: _____

PROFESOR: _____

INSTRUCCIONES PARA EL EXAMEN

En el cuadernillo se ha dejado espacio en blanco después de cada pregunta para realizar los cálculos. Aparte de este espacio en blanco puede utilizar la parte de atrás de cada hoja de este cuadernillo.

Hay distintos tipos de examen. A continuación aparece la parte superior de su hoja de respuestas. Por ejemplo, si su examen es del tipo E, observe cómo ya aparece indicado su tipo de examen. Observe también cómo se ha indicado el DNI de un alumno que supuestamente tiene el "99.724.106". Haga Usted lo mismo ahora.



**UNIVERSIDAD POLITÉCNICA
DE VALENCIA**

TIPO E

1.º APELLIDO	2.º APELLIDO	NOMBRE
CENTRO	ASIGNATURA	
CURSO	FIRMA	FECHA
GRUPO	CALIFICACION	

EXPEDIENTE o D.N.I.

9	9	7	2	4	1	0	6
0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9

HOJA DE RESPUESTAS

– MARQUE CORRECTAMENTE.

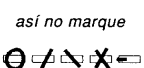
– BORRE BIEN EN CASO DE ERROR.

– ESCRIBA CON LAPIZ UNICAMENTE

marque así



así no marque



CODIGO

a	b	c	d	e	f	g	h	i	j	k	l	m
0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9	9	9	9	9

EDAD

0	0
1	1
2	2
3	3
4	4
5	5
6	6
SEXO	
V	8
M	9

SEXO

7
8
9

Las preguntas acertadas suman un punto. Las que no se contesten suman 0 puntos. Las mal contestadas restan 0.33 puntos. Sólo hay una respuesta correcta.

TODAS LAS CUESTIONES CONTESTADAS DEBERÁN SER DEBIDAMENTE JUSTIFICADAS EN LAS HOJAS DEL EXAMEN.

RECUERDE: PONGA AHORA SU NOMBRE EN LA HOJA DE RESPUESTAS E INDIQUE SU NUMERO DE DNI.

SE RETIRARÁ TODO FORMULARIO QUE TENGA ALGO ESCRITO POR EL ALUMNO.

NO SE PERMITE "PRESTAR" EL FORMULARIO NI LA CALCULADORA.

CUESTIONES (40% nota)

Un estudiante de Ingeniería Industrial tiene un examen tipo test en el que cada pregunta tiene 5 posibles respuestas y sólo 1 es correcta. Este alumno se sabe la respuesta al 70% de las preguntas y cuando no se la sabe, elige una al azar.

- I) ¿Cuál es la probabilidad de que este alumno responda correctamente a una pregunta cualquiera elegida aleatoriamente?
- II) Otro alumno que se presenta a ese mismo examen no ha estudiado absolutamente nada y piensa responder al azar a todas las preguntas. Si el examen tiene 20 preguntas en total y para aprobar hay que responder correctamente, al menos, la mitad de ellas. ¿Cuál es la probabilidad de que apruebe?

- a) **0.76, 0.003**
- b) 0.76, 0.008
- c) 0.15, 0.003
- d) 0.15, 0.008

$$\begin{aligned} \text{I)} \quad P(\text{acertar}) &= P(\text{acertar}|\text{saber respuesta}) \cdot P(\text{saber respuesta}) + \\ &\quad + P(\text{acertar}|\text{no saber respuesta}) \cdot P(\text{no saber respuesta}) = \\ &= 1 \cdot 0.7 + (1/5) \cdot 0.3 = 0.76 \end{aligned}$$

$$\begin{aligned} \text{II)} \quad X &= \text{n}^\circ \text{ preguntas que responde correctamente} \\ X &= \text{Bi}(n=20, p=1/5) \end{aligned}$$

$$P(\text{aprobar}) = P(X \geq 10) = 1 - P(X \leq 9) = 1 - 0.9970 = 0.003$$

Queremos estimar por muestreo, la previsión del gasto medio por familia μ en las rebajas del 2006. Para ello decidimos preguntar la previsión de gasto a 200 familias españolas elegidas al azar y dar como estimación de μ el valor que obtengamos de la media muestral. Si admitimos que el gasto por familia tiene una distribución $N(\mu, 50)$ ¿Cuál es la probabilidad de que el error cometido en la estimación sea, en valor absoluto, menor de 10 euros?

- a) 0.0046
- b) 0.1586
- c) **0.9954**
- d) 0.9977

Sol:

El gasto por familia es una variable con distribución $X \equiv N(\mu, 50)$

El gasto medio de 200 familias será otra variable $\bar{X} \equiv N(\mu, \frac{50}{\sqrt{200}}) \equiv N(\mu, 3,536)$

$$\text{La } P(|\bar{X} - \mu| \leq 10) = P(\bar{X} \leq \mu + 10) - P(\bar{X} \leq \mu - 10) = \Phi\left(\frac{(\mu + 10) - \mu}{3,536}\right) - \Phi\left(\frac{(\mu - 10) - \mu}{3,536}\right) = 0,9954$$

En una factoría de automóviles se desea verificar la hipótesis de que el promedio de defectos/carrocería se mantiene igual o inferior a 1 defecto/carrocería. Para realizar dicha verificación se inspeccionan 5 carrocerías y si el total de defectos es menor o igual que 8 se acepta la hipótesis establecida, rechazándola si dicho número es mayor que 8.

Admitiendo que el número de defectos por carrocería es una variable con distribución de Poisson

- Determinar la probabilidad α de cometer el error de primera especie con el test propuesto.
- Calcular la probabilidad β de cometer el error de segunda especie con el test anterior si en realidad se están produciendo, en promedio, 2,6 defectos por carrocería.

- 0.00005, 0.998
- 0.068, 0.10**
- 0.068, 0.90
- 0.00005, 0.10

Sol:

Las hipótesis planteadas son:

$$H_0: \lambda \leq 1$$

$$H_1: \lambda > 1$$

La regla de decisión es:

Aceptar H_0 si el número de defectos en $n=5$ carrocerías es ≤ 8

Número de defectos en una carrocería $X=Ps(\lambda)$

Número de defectos en 5 carrocerías $Y = X_1 + X_2 + \dots + X_5 = Ps(5\lambda)$

Luego deberemos aceptar H_0 si $Y \leq 8$

- $\alpha = \max P(\text{rechazar } H_0/H_0 \text{ cierta}) = P(Y > 8/\lambda = 1)$
 $\alpha = P(Ps(5) > 8) = 1 - P(Ps(5) \leq 8) = \underline{\mathbf{0,068}}$
- $\beta = P(\text{aceptar } H_0/H_1 \text{ cierta}) = P(Y \leq 8/\lambda = 2,6)$
 $\beta = P(Ps(5 \times 2,6) \leq 8) = P(Ps(13) \leq 8) = \underline{\mathbf{0,10}}$

La probabilidad de error de primera especie de un contraste bilateral es 0.03 entonces:

- Se aceptará la hipótesis nula para el 97% de las muestras observadas
- Si la hipótesis nula es cierta, sólo el 3% de las veces se obtendrá una discrepancia mayor que la observada
- Si la hipótesis nula es cierta, sólo el 3% de las veces se rechazará**
- La probabilidad de rechazar la hipótesis nula cuando es cierta es de 0.015.

Para comprobar si el tiempo de fallo de unos componentes electrónicos son diferentes según el proveedor que los haya fabricado, se tomaron 76 unidades del proveedor A y 225 unidades del proveedor B, sometiéndose a pruebas de fallo, y con los valores obtenidos, analizados mediante el *statgraphics*, han proporcionado los siguientes resultados:

Comparación de Medias

95,0% intervalo de confianza para la media de A: 1237,53 +/- 36,1139 [1201,42,1273,65]
 95,0% intervalo de confianza para la media de B: 1370,6 +/- 26,797 [1343,8,1397,39]
 95,0% intervalos de confianza para la diferencia de medias:
 suponiendo varianzas iguales: -133,063 +/- 50,5163 [-183,58, -82,547]

contrastes t de comparación de medias

Hipótesis nula: media1 = media2
 Hipótesis alt.: media1 <> media2
 suponiendo varianzas iguales: t = -5,18366

Comparación de Desviaciones Típicas

	A	B
Desviación Típica	158,04	203,975
Varianza	24976,8	41605,7
GL	75	224

Cociente de varianzas = 0,600322

95,0% Intervalos de Confianza

Desviación Típica deA: [136,296;188,104]
 Desviación Típica deB: [186,708;224,789]
 Cociente de varianzas: [0,421105;0,886152]

Contrastes F para comparar varianzas

Hipótesis nula: sigma1 = sigma2
 (1) Hipótesis alt.: sigma1 <> sigma2
 F = 0,600322 P-Valor = 0,010796

¿Cuál de las siguientes afirmaciones es verdadera?

- a) Las varianzas de los dos proveedores es la misma.
- b) Las medias de los dos proveedores son iguales
- c) **Los dos proveedores difieren en la varianza**
- d) Los dos proveedores tiene la misma varianza y la misma media.

Se desea conocer la relación que hay entre el Y =rendimiento de un producto y las variables X_1 = temperatura del proceso, X_2 = catalizador 1, X_3 = velocidad de agitación. Para ello se han tomado 15 pruebas y con los valores observados se ha ajustado una regresión lineal múltiple, que mediante el statgraphics proporciona los siguientes resultados:

Análisis de Regresión Múltiple

Variable dependiente: Y

Parámetro	Estimación	Error estándar	Estadístico T	P-Valor
CONSTANTE	210,936	16,7648	12,5821	0,0000
X1	0,374701	1,62103	0,23115	0,8214
X2	0,406432	0,13054	3,11347	0,0099
X3	0,51967	0,144574	3,59449	0,0042

Análisis de Varianza

Fuente	Suma de cuadrados	GL	Cuadrado medio	Cociente-F	P-Valor
Modelo	46523,7	3	15507,9	34,05	0,0000
Residuo	5009,61	11	455,42		
Total (Corr.)	51533,3	14			

R-cuadrado = 90,2789 porcentaje

R-cuadrado (ajustado para g.l.) = 87,6277 porcentaje

Error estándar de est. = 21,3406

Error absoluto medio = 12,2235

Estadístico de Durbin-Watson = 1,72174 (P=0,2812)

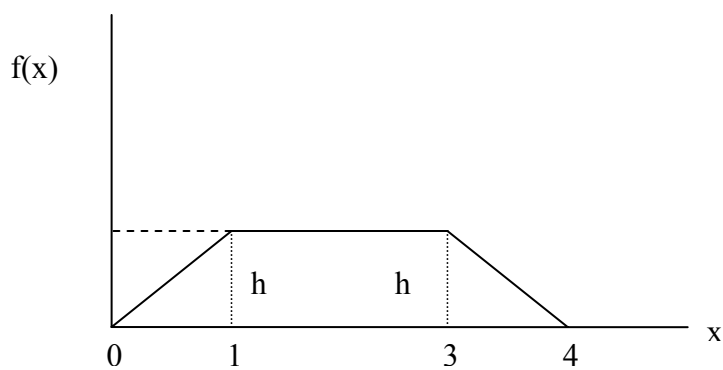
Autocorrelación residual en Lag 1 = 0,125857

A la vista de estos resultados, ¿cuáles son las afirmaciones correctas?

- I) El modelo de regresión no es correcto porque el porcentaje explicado de la varianza total no supera el 95%.
- II) La variable independiente que más influye en la explicación de la variable dependiente es la temperatura, puesto que su coeficiente de regresión es el de mayor valor.
- III) La temperatura es significativa por tener un p-valor mayor del 5%.
- IV) Las variables X_2 y X_3 son significativas ya que su p-valor son menores de 0,05

- a) La III y la I
- b) La II y la III
- c) La I y la II
- d) La IV

La v.a. X tiene la función de densidad que se muestra en la figura de abajo. La $P(x \geq 1,75)$ es:



- a) 0,416
- b) 0,25
- c) 0,583**
- d) 0,75

Para que $f(x)$ sea una función de densidad, el área encerrada bajo dicha curva y el eje de las x 's debe de ser la unidad. Para que se cumpla esta la altura h que hay en la figura debe ser:

$$\frac{h \cdot 1}{2} + 2 \cdot h + \frac{h \cdot 1}{2} = 1$$

De donde $h = 1/3$.

La probabilidad pedida es $P(X = 1,75) = (3 - 1,75) \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{2} = 0,583$

El control de calidad de un chip se realiza mediante dos pruebas que dan como resultado dos variables continuas, X e Y . El chip pasa el control de calidad si en la prueba $|X - Y| \leq 4$. Las variables X e Y siguen una distribución normal de parámetros $X = N(10, 2)$ e $Y = N(10, 1)$ siendo $\text{cov}(X, Y) = 0.5$. Bajo estas condiciones la proporción de chips defectuosos que se fabrican es igual:

- a) 0.0455**
- b) 0.0726
- c) 0.1010
- d) 0.8024

Sea un sistema eléctrico de tres componentes en paralelo. El funcionamiento de cada componente es independiente de los demás siendo su probabilidad de avería del 20% . Para que el sistema funcione adecuadamente deben funcionar al menos dos componentes. ¿Cuál es la probabilidad de que el sistema funcione adecuadamente?

- a) 0.104
- b) 0.640
- c) 0.896
- d) 0.512

PROBLEMAS (60% de la nota)

Una determinada fábrica produce relés con un 1,5% de defectuosos. Para comprobar su funcionamiento se realiza una verificación con la que se admiten como correctos aquellos que lo son con una probabilidad del 95%, y los que no lo son con una probabilidad del 3 %. ¿Cuál será la probabilidad de que un relé que en dos verificaciones sucesivas ha sido dado por bueno, realmente lo sea?

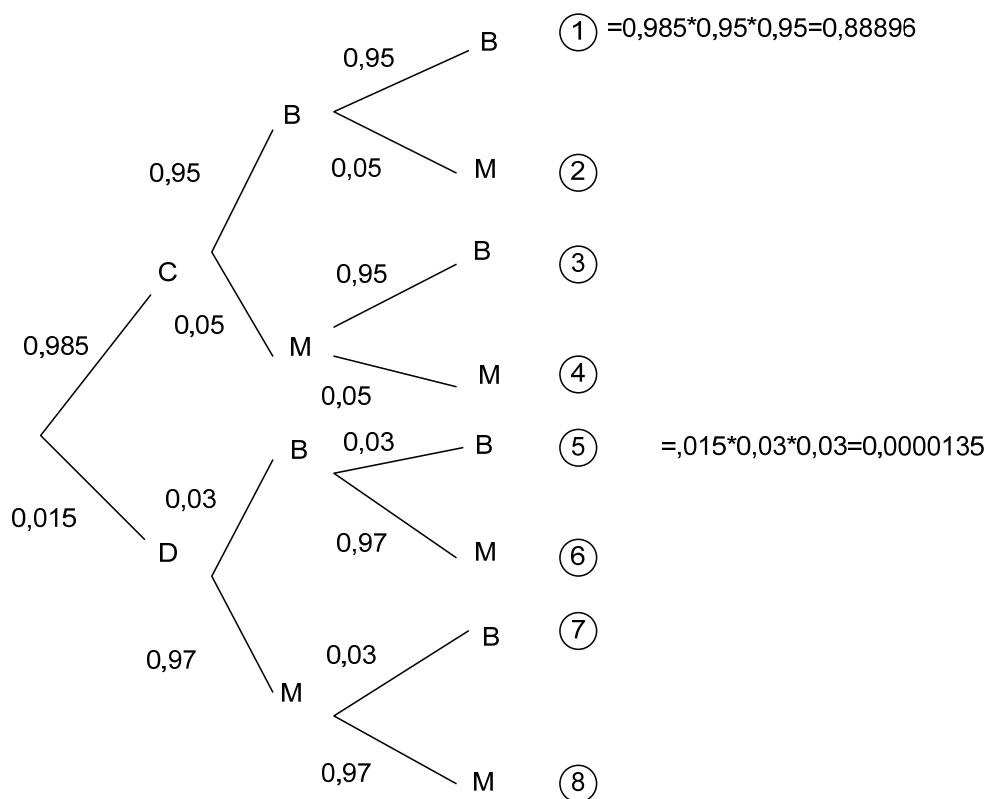
Sol:

Llamando a:

C = pieza correcta; D = pieza defectuosa, y

B = La verificación dice que la pieza es correcta; M = la verificación dice que la pieza es defectuosa.

El árbol de probabilidad es,



La probabilidad pedida viene dada por:

$$P(C/BB) = \frac{C1}{C1 + C5} = \frac{0,88896}{0,88896 + 0,0000135} = 0,99998 = 99,998\%$$

Una conocida marca láctea afirma en sus spots publicitarios que el contenido medio en calcio de la leche que comercializa es de 2 gramos/litro, es decir, por brik de leche. Un consumidor ha encargado a un laboratorio que analice el contenido en calcio de un brik de leche y afirma que denunciará a la compañía por publicidad engañosa si el laboratorio le indica que dicho contenido es menor o igual a 1,3 gramos/litro. Por experiencias anteriores, este laboratorio sabe que el contenido en calcio por litro de leche sigue una distribución normal y que su desviación típica es de 0,5 gramos/litro.

- ¿Qué contraste de hipótesis plantearías para decidir si la compañía está engañando en su publicidad?
- ¿En qué caso aceptará este consumidor la hipótesis de que la compañía no engaña en su spot publicitario?
- ¿Qué probabilidad de error de tipo I, o de primera especie, está cometiendo el consumidor?
- Si el verdadero valor de la media de la distribución del contenido en calcio de la leche es de 1,8 gramos/litro, ¿cuál sería la probabilidad de error de tipo II de la regla de decisión de este consumidor?
- La compañía, por su cuenta, ha decidido analizar el contenido en calcio de 100 brik de leche. El contenido medio en calcio de esta muestra ha resultado ser de 1,7 gramos/litro. Con estos datos, calcula un intervalo de confianza al 95% para el contenido medio de calcio.

$$H_0: \mu=2$$

$$H_1: \mu<2$$

$$X=\text{contenido en calcio del brik de leche analizado} = N(\mu; 0,5)$$

Si $X \leq 1,3$, rechazaremos H_0

En caso de que el contenido en calcio del brik analizado sea mayor que 1,3 gramos/litro.

$$a) \alpha = P(X \leq 1,3 \mid \mu=2) = P(Z \leq (1,3-2)/0,5) = \Phi(-1,4) = 0,0808$$

$$b) \beta = P(X > 1,3 \mid \mu=1,8) = P(Z > (1,3-1,8)/0,5) = 1 - \Phi(-1) = 1 - 0,1587 = 0,8413$$

$$c) n=100 \quad \bar{x}=1,7 \quad s=0,5 \quad \alpha=0,05$$

$$\text{Intervalo Confianza } 95\% = 1,7 \pm 1,96 \cdot 0,5/\sqrt{100} = (1,602; 1,798)$$

Queremos contrastar mediante muestreo nuestra suposición de que, como máximo, hay un 2% de alumnos matriculados en la UPV (en cualquier ciclo) que proceden de otras universidades. Para ello elegiremos al azar en el Servicio de Alumnado de la UPV los expedientes de n alumnos y aceptaremos nuestra propuesta si entre ellos hay como máximo c expedientes de alumnos procedentes de otras universidades. Queremos que si hay el 2% de alumnos procedentes del exterior, el test nos lleve a aceptar nuestra suposición con una probabilidad del 98% y si en realidad hubiera un 7,2% de alumnos solo se aceptara con una probabilidad del 30%. Determinar los valores de n y c para que se cumplan las condiciones expresadas en el enunciado.

Sol:

Las hipótesis a contrastar son:

$$H_0: P \leq 0,02$$

$$H_1: P > 0,02$$

Siendo P la proporción de estudiantes procedentes de otras universidades.

La regla de decisión o plan de muestreo será:

Aceptar H_0 si en una muestra aleatoria de n estudiantes hay C o menos del exterior

Queremos que se cumpla que cuando $P=0,02$ la probabilidad de aceptar H_0 sea $(1-\alpha) = 0,98$ y que cuando sea $P=0,072$ la probabilidad de aceptar H_0 sea $\beta \leq 0,30$

Necesitamos calcular el tamaño de muestra o número de alumnos a encuestar n , y el número máximo de ellos C de dicha muestra que procedan de otras universidades para aceptar nuestra H_0 de forma que se cumplan las condiciones del punto anterior.

Admitimos (luego comprobaremos el supuesto) que la variable que describe el número de alumnos procedentes de otras universidades en una muestra de tamaño n , es una variable $Ps(n,P)$ siendo P la proporción real de alumnos procedentes del exterior.

Como se explicó en clase, procedemos por tanteos creando la siguiente tabla: (*)

C	(nP)	$n=(nP)/0,02$	$P_{aceptar}(P=0,072)=P(Ps(n,0,072) \leq C)$
0	0,02	1	
1	0,20	10	$P(Ps(10 \times 0,072) \leq 1) = 0,83 >>> 0,30$ no
2	0,55	28	$P(Ps(28 \times 0,072) \leq 2) = 0,67 >>> 0,30$ no
3			
4			
5	2,10	105	$P(Ps(105 \times 0,072) \leq 5) = 0,24 < 0,30$ O.K.

Luego $n=105$ y $C=5$

La aproximación prevista al principio es válida dado que $n > 50$ y $nP < 15$

(*) Comenzamos fijando un valor de C y determinando en tablas o en el ábaco de Piosson el valor del parámetro (nP) de una variable da Poisson tal que $P(Ps(nP) \leq C) = 0,98$

Con el valor de nP calculado determinamos n según se indica en la tabla

Después verificamos si se cumplen la condición de β

Procedemos incrementando C y repitiendo lo anterior hasta que se verifique la condición de β . Los valores de n y de C cuando se cumplen las dos condiciones son el resultado del problema

Para la refrigeración de una máquina se utiliza un fluido que se recircula con un caudal variable X expresado en litros/s. La temperatura de régimen de dicha máquina también es una variable aleatoria Y . Ambas se distribuyen conjuntamente según una Normal bivalente de parámetros

$$(X, Y) = N\left(\vec{\mu} = \begin{Bmatrix} 20 \text{ l/s} \\ 80^\circ \text{C} \end{Bmatrix}, \vec{V} = \begin{bmatrix} 9 & -5,5 \\ -5,5 & 4 \end{bmatrix}\right)$$

- ¿Con qué caudal de refrigerante se garantiza que la máquina funcione el 75% del tiempo por debajo de 79°C ?
- ¿Cuál es la función de densidad marginal de la variable caudal de refrigerante?
- Escribir la ecuación de la curva de regresión condicional de Y/X .
- ¿A qué temperatura cabe esperar que funcione la máquina si se el refrigerante está circulando un caudal de 25 litros/s?

Sol:

a) La variable temperatura de funcionamiento condicionada al caudal de refrigerante Y/X sabemos que tiene una distribución

$$Y/X \equiv N\left(\mu_y + \frac{\sigma_{x,y}^2}{\sigma_x^2}(x - \mu_x), \sigma_y^2(1 - \rho^2)^{1/2}\right) \equiv N\left(80 + \frac{-5,5}{9}(x - 20), 2(1 - \frac{5,5^2}{9 \cdot 4})^{1/2}\right)$$

$$Y/X = N(92,22 - 0,611x, 0,7993)$$

Según el enunciado, habrá que determinar el valor de X tal que

$$P(Y/X \leq 79) = 0,75 \Rightarrow P(N(\mu_c, 0,7993) \leq 79) = 0,75 \Rightarrow \Phi\left(\frac{79 - \mu_c}{0,7993}\right) = 0,75$$

$$\text{De tablas } \frac{79 - \mu_c}{0,7993} = 0,675 \Rightarrow \mu_c = 78,46$$

Igualando $\mu_c = 78,46 = 92,22 - 0,611x$ obtenemos $x = 22,52$ litros/seg

b) La distribución marginal de la variable caudal es $X = N(20, 3)$ y por lo tanto su función de

$$\text{densidad es } f(x) = \frac{1}{3\sqrt{2\pi}} \text{EXP}\left(-\frac{(x-20)^2}{9}\right) \quad -\infty < X < +\infty$$

c) La $\text{crc } Y/X = E(Y/X) = \left(\mu_y + \frac{\sigma_{x,y}^2}{\sigma_x^2}(x - \mu_x)\right) = 80 - (5,5/9)(x - 20) = 92,22 - 0,611X$ que es también la $\text{rrmc } Y/X$

d) Para $x = 25$ litros/seg, el valor esperado de la temperatura será

$$E(Y/x=25) = 92,22 - 0,611 \cdot 25 = 76,945^\circ \text{C}$$