COLECCIÓN

DE

EXÁMENES

ESTADÍSTICA I

Unión de Estudiantes de Ciencias Económicas | AECUC3M

LEC/LADE/LEC-DER/LADE-DER
CURSO 1996/97
EXAMEN DE ESTADÍSTICA I
1 de Febrero de 1997

PROBLEMA 2. .

La proporción en que dos componentes aparecen en un producto final viene dada por la función de densidad:

$$f(x,y) = \begin{cases} 4xy & \text{si } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{en caso contrario} \end{cases},$$

donde **X** e **Y** son las proporciones de cada producto. Se pide:

a) Hallar las funciones de densidad marginales y la esperanza de **Y** condicionada a **X**.

b) ?‘Son independientes **X** e **Y**?, ?‘por qué?.

c) ¿Cuál es la probabilidad de que la proporción del primer componente, **X**, sea inferior a 0.3, cuando la del segundo, **Y**, es 0.8?.

PROBLEMA 3. .

En un importante proyecto espacial de la N.A.S.A. se ha descubierto que la variación de la temperatura cada hora en una determinada zona del planeta Krypton se distribuye según una variable aleatoria cuya función de densidad es:

$$f(x) = \frac{1}{\sqrt{\theta+1}\sqrt{2\pi}} e^{-\frac{1}{2}\frac{x^2}{(\theta+1)}}, \text{ donde se supone que } \theta > -1.$$

Se desea estimar el parámetro θ. Se toma una muestra de n variaciones de temperatura independientes:

a) Calcular $\widehat{\theta}_n$, el estimador de máxima verosimilitud de θ.

b) Demostrar que la varianza asintótica (estimada) del estimador obtenido en el apartado **a)** es $\frac{2\left(\widehat{\theta}_n+1\right)^2}{n}$.

c) Construir un intervalo de confianza asintótico para θ tomando $\alpha = 0.10$. Para una muestra de tamaño $n = 100$ de la que se sabe que $\sum_{i=1}^{100} x_i^2 = 400$, ¿cuál sería el intervalo obtenido?.

LEC/LADE/LEC-DER/LADE-DER/LEC-PERIODISMO
CURSO 2005/06
EXAMEN DE ESTADÍSTICA I
9 de Febrero de 2006
TIEMPO MÁXIMO = 3 h

1. **(2,5 puntos).** La entrada en vigor de la Ley 28/2005 por la que se prohíbe fumar en lugares públicos ha suscitado opiniones muy encontradas. Con la aplicación de dicha Ley, un porcentaje de fumadores ha decidido dejar el tabaco y cierto organismo establece en el 20% este porcentaje.

   (a) Deducir la distribución de la proporción muestral de fumadores cuya intención es dejar el tabaco (en caso de muestras grandes).

   (b) ¿Es insesgado el estimador proporción muestral de fumadores con intención de dejar el tabaco? Razonar la respuesta.

   (c) El Gobierno asegura que el porcentaje de fumadores que van a decidir dejar de fumar con la aplicación de esta Ley, va a ser al menos el 28%. Para contrastar dicha hipótesis, se encuesta a 800 fumadores de los que 150 manifiestan su intención de dejar de fumar desde la entrada en vigor de la Ley. ¿Hay evidencia estadística suficiente para rechazar la hipótesis del Gobierno?

2. **(2,5 puntos).** Sea $X_1, ..., X_n$ una muestra aleatoria simple de una distribución Normal con media μ y desviación típica σ. Definimos la varianza muestral $\hat{\sigma}^2$ y la varianza muestral corregida (o cuasivarianza) s^2 como:

$$\hat{\sigma}^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n} \qquad\qquad s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$$

   donde $\bar{x}$ es la media muestral. Si sabemos que una distribución χ^2_n tiene su media igual a los grados de libertad ($E(\chi^2_n) = n$) y su varianza igual al doble que la media ($Var(\chi^2_n) = 2n$),

   (a) Demostrar que $\hat{\sigma}^2$ es un estimador sesgado de la varianza poblacional σ^2. Obtener su sesgo.

   (b) Demostrar que s^2 es un estimador insesgado de la varianza poblacional σ^2.

   (c) Obtener el error cuadrático medio de $\hat{\sigma}^2$.

   (d) Obtener el error cuadrático medio de s^2.

   (e) Basándonos en el error cuadrático medio, ¿qué estimador es preferible?

3. **(2,5 puntos).** Un fabricante de pasta indica en el envoltorio que la cantidad de pasta que contiene cada paquete es de 500 gr. Se cree que la desviación típica del peso de los paquetes es de 3 gr. Se toma una muestra de 20 paquetes y se obtiene un peso medio de 498 gr con una desviación típica corregida de 4 gr. Especificando las hipótesis necesarias, responder:

   (a) ¿Hay evidencia de que por término medio los paquetes contienen menos pasta de lo indicado? Según tu conclusión, ¿qué tipo de error podrías estar cometiendo y con qué probabilidad?

   (b) ¿Es la desviación típica mayor de lo que se cree? Utiliza un nivel de significación del 5%.

   (c) Calcular la potencia del contraste de la media del apartado (a) para un nivel de significación del 5% y un valor alternativo de la media de 497.

4. **(2.5 puntos).** Se desea evaluar la efectividad de una nueva vacuna antigripal. Para ello se suministra de manera voluntaria y gratuita, en una pequeña comunidad. La vacuna se administra en dos dosis, separadas por un período de dos semanas, de forma que algunas personas han recibido una sola dosis, otras han recibido las dos, y otras personas no han recibido ninguna. La siguiente tabla indica los resultados que se registraron durante la siguiente primavera en 1000 habitantes de la comunidad elegidos al azar.

	No vacunados (0 dosis)	Vacunados (1 dosis)	Vacunados (2 dosis)
Gripe	24	9	13
No Gripe	289	100	565

(a) ¿Proporcionan estos datos suficiente evidencia estadística (al nivel 0.05) para indicar dependencia entre el número de dosis recibidas y la protección frente a la gripe?

(b) Si consideramos vacunados a los que han recibido una ó dos dosis, ¿hay evidencia estadística para afirmar (al nivel 0.05) que la vacuna es efectiva frente a la gripe?

En ambos apartados escriba claramente las hipótesis que se contrastan, la región de rechazo y la conclusión.

NOTA: Las notas de cada grupo serán publicadas por cada profesor a última hora del miércoles 15 de febrero en Aula Global con las correspondientes instrucciones para solicitar revisión.

LEC/LADE/LEC-DER/LADE-DER/LEC-PERIODISMO
CURSO 2005/06
EXAMEN DE ESTADÍSTICA I
1 de Septiembre de 2006
TIEMPO MÁXIMO = 3 h

1. **(2,5 puntos).** El número X de defectos que presenta determinado tipo de piezas de una producción industrial sigue una distribución de Poisson con parámetro θ desconocido. Se toma una muestra aleatoria de 40 piezas y se registra el número de defectos que presenta cada pieza, obteniendo la siguiente información:

Número de defectos	0	1	2	3	4	5	6
Cantidad de piezas	4	10	11	8	4	2	1

(a) Determinar el estimador de máxima verosimilitud (EMV) de θ.

(b) Calcular la estimación máximo verosímil con los datos de la tabla.

(c) Calcular un intervalo de confianza asintótico para θ en la muestra obtenida, tomando $\alpha = 0.05$.

Nota: La función de masa de una distribución de Poisson de parámetro θ es:

$$P(X = x) = \frac{e^{-\theta}\theta^x}{x!}, \qquad x = 0, 1, 2, \ldots$$

2. **(2,5 puntos).** Se recogen medidas de polución atmosférica en 10 sitios de una ciudad:

3.3; 1.7; 3.7; 4.6; 2.3; 3.9; 4.3; 1.4; 1.6; 3.6

(a) Hallar un intervalo de confianza al 95% para la varianza poblacional, mencionando las hipótesis estadísticas que hay que asumir.

(b) La polución atmosférica media de estos 10 sitios medida un año antes fue 2.27. Contrastar al 5% la hipótesis de que el nivel medio de polución no ha variado. Especificar claramente las hipótesis nula y alternativa del contraste, así como las hipótesis estadísticas que hay que asumir para realizarlo.

(c) Una organización ecologista sospecha que la polución atmosférica media está aumentando cada año. Contrastar al 5% la hipótesis de que el nivel medio de polución no ha aumentado (tómese como H_0). Especificar claramente las hipótesis nula y alternativa del contraste, así como las hipótesis estadísticas que hay que asumir para realizarlo. Comentar el resultado de los contrastes realizados en (b) y (c).

3. **(2,5 puntos).** Los defensores de un nuevo molino de viento afirman que la potencia diaria que puede generar es una variable aleatoria (X) cuya media es al menos 800 Kw. Se asume que X es normal con desviación típica 120 Kw. Para realizar un contraste de la hipótesis mantenida por los defensores del nuevo molino, se toma una muestra aleatoria simple de tamaño 100 y se obtiene una media muestral igual a 776 Kw.

   (a) Definir y calcular el p-valor del contraste.

   (b) Definir y calcular la potencia del contraste tomando como hipótesis alternativa $\mu = 770$ y nivel de significación $\alpha = 0.05$.

4. **(2.5 puntos).** Se tomaron cinco muestras de plantas en cinco lugares distintos, A, B, C, D y E observando en cada una de ellas, el número de plantas infectadas con un cierto parásito. Las dos primeras muestras A y B se obtuvieron en una zona más húmeda que las tres restantes C, D y E. Los resultados fueron los siguientes:

Muestras	Infectados	No Infectados	Total
A	80	184	264
B	58	147	205
C	114	276	390
D	55	210	265
E	79	296	375
Total	386	1113	1499

(a) Contrastar la hipótesis (al nivel 0.05) de que la infección de parásitos es independiente de la zona donde se extrajeron las muestras.

(b) Consideramos ahora las muestras procedentes sólo de dos zonas: húmedas y secas. ¿Hay evidencia suficiente para afirmar que la infección es distinta en ambas zonas?

En ambos casos ((a) y (b)) especificar las hipótesis contrastadas, la región de rechazo y la conclusión.

NOTA: Las notas de cada grupo serán publicadas por cada profesor el en Aula Global con las correspondientes instrucciones para solicitar revisión.

3. **(2,5 puntos).** Un fabricante de pasta indica en el envoltorio que la cantidad de pasta que contiene cada paquete es de 500 gr. Se cree que la desviación típica del peso de los paquetes es de 3 gr. Se toma una muestra de 20 paquetes y se obtiene un peso medio de 498 gr con una desviación típica corregida de 4 gr. Especificando las hipótesis necesarias, responder:

   (a) ¿Hay evidencia de que por término medio los paquetes contienen menos pasta de lo indicado? Según tu conclusión, ¿qué tipo de error podrías estar cometiendo y con qué probabilidad?

   (b) ¿Es la desviación típica mayor de lo que se cree? Utiliza un nivel de significación del 5%.

   (c) Calcular la potencia del contraste de la media del apartado (a) para un nivel de significación del 5% y un valor alternativo de la media de 497.

4. **(2.5 puntos).** Se desea evaluar la efectividad de una nueva vacuna antigripal. Para ello se suministra de manera voluntaria y gratuita, en una pequeña comunidad. La vacuna se administra en dos dosis, separadas por un período de dos semanas, de forma que algunas personas han recibido una sola dosis, otras han recibido las dos, y otras personas no han recibido ninguna. La siguiente tabla indica los resultados que se registraron durante la siguiente primavera en 1000 habitantes de la comunidad elegidos al azar.

	No vacunados (0 dosis)	Vacunados (1 dosis)	Vacunados (2 dosis)
Gripe	24	9	13
No Gripe	289	100	565

(a) ¿Proporcionan estos datos suficiente evidencia estadística (al nivel 0.05) para indicar dependencia entre el número de dosis recibidas y la protección frente a la gripe?

(b) Si consideramos vacunados a los que han recibido una ó dos dosis, ¿hay evidencia estadística para afirmar (al nivel 0.05) que la vacuna es efectiva frente a la gripe?

En ambos apartados escriba claramente las hipótesis que se contrastan, la región de rechazo y la conclusión.

NOTA: Las notas de cada grupo serán publicadas por cada profesor a última hora del miércoles 15 de febrero en Aula Global con las correspondientes instrucciones para solicitar revisión.

PROBLEMA 4. .

Sea X la variable "rentabilidad de cierto tipo de fondos de inversión tras una apreciación fuerte del marco con respecto al dolar". Se considera que la media de esta variable es 15. Un economista afirma que dicha rentabilidad media ha variado, por lo que lleva a cabo un estudio en las condiciones reseñadas anteriormente sobre una muestra de 9 fondos cuya media muestral resulta ser de 15.308 y cuya varianza muestral corregida (cuasivarianza) es 0.193.

a) Especificando las hipótesis necesarias, contrastar la afirmación del economista al 5%.

b) A partir del resultado de **a)**, razonar si el intervalo de confianza para la media (centrado en $\overline{x}$) al 95% contendrá o no al valor 15.

c) Acotar el p-valor. Si el contraste se hubiera realizado al 10%, ¿aceptaríamos la hipótesis de que la media de la rentabilidad es 15 tras una apreciación fuerte del marco con respecto al dolar?. Razonar brevemente la respuesta.

LEC/LADE/LEC-DER/LADE-DER
CURSO 1996/97
EXAMEN DE ESTADÍSTICA I
Septiembre de 1997

PROBLEMA 2. .

Una máquina está compuesta por 2 componentes básicas **X** e **Y.** La componente **X** tiene 3 posibles tipos de defectos y la componente **Y**, 4. Ordenados ambos de menor a mayor gravedad, la función de probabilidad conjunta viene dada por:

X/Y	**0**	**1**	**2**	**3**
0	0.02	0.10	0	0.08
1	0.11	**k**	0.20	0.12
2	0.08	0.07	0.10	0.05

a) Calcular **k** y determinar las distribuciones marginales de **X** e **Y**.

b) Calcular la probabilidad de que la componente **X** tenga un defecto de tipo 1 sabiendo que el defecto que tiene la componente **Y** es de tipo 2.

c) Obtener el vector de medias **(X, Y)**.

d) ?'Son **X** e **Y** independientes?. ?'Por qué?.

e) Se sabe que $Var(\mathbf{X}) = 0.49$ y que $Var(\mathbf{Y}) = 0.1619$. ¿Se puede concluir que $Var(\mathbf{X} - \mathbf{Y}) = 1.6519$?. Razona la respuesta.

PROBLEMA 3.

Se ha recogido una muestra aleatoria de la previsión de inflación para el año 1997 en siete países de la OCDE. Las previsiones para estos países son:

1.5   2.1   1.9   2.3   2.5   3.2   3.0

a) Utilizando estos datos, construye un intervalo de confianza al 99% para la media de la previsión de inflación de estos siete países de la OCDE. Indica los supuestos que necesitas hacer.

b) Construye un intervalo de confianza al 99% para la desviación típica de la previsión de inflación de estos siete países de la OCDE. Indica qué supuestos necesitas hacer.

c) Los expertos consultados opinan que el intervalo de confianza para la media de la previsión de inflación del conjunto de países de la OCDE calculado en **a)** es demasiado amplio. Sugieren construir un intervalo de confianza cuya longitud total sea de 1 punto. Acotar el nivel de confianza para este nuevo intervalo.

PROBLEMA 4. .

Una compañía farmacéutica se dedica a envasar bolsas de vitamina C en polvo. Se sabe con certeza que el peso de las bolsas sigue una distribución normal. También se sabe que cuando el proceso está operando correctamente los parámetros de esta distribución normal son $\mu = 16$ y $\sigma^2 = 0.25$. Un determinado día se toma una muestra aleatoria de 16 bolsas obteniéndose los siguientes resultados:

$\overline{x} = 16.29 \quad s = 0.86$, donde $\overline{x}$ es la media muestral y $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{16} (x_i - \overline{x})^2}$.

a) ?'Se puede concluir a partir de la muestra que el proceso está operando correctamente ese día?. Acotar el p-valor de cada uno de los contrastes necesarios.

b) ¿Cómo cambiarían las conclusiones de **a)** si en lugar de tomar una muestra aleatoria de 16 bolsas se seleccionan las primeras 16 bolsas de la mañana?.

LEC/LADE/LEC-DER/LADE-DER
CURSO 1997/98
EXAMEN DE ESTADÍSTICA I
31 de Enero de 1998

PROBLEMA 2. .

a) Definir estimador insesgado de un cierto parámetro θ.

b) Definir el Error Cuadrático Medio (E. C. M.) de un estimador del parámetro θ y obtener la expresión del E. C. M. en función del sesgo y la varianza del estimador.

c) Para una muestra $X_1, ..., X_4$ de una población de media μ y varianza $k\mu^2$ considérense los siguientes estimadores de μ:

$$T_1 = \frac{X_1+4X_2}{5} \quad T_2 = \frac{X_1+X_2+X_3+X_4}{3}$$

i) Calcular el sesgo de T_1 y T_2.

ii) Calcular el E. C. M. de T_1 y T_2.

iii) ¿Para qué valores de k es el estimador T_2 mejor que T_1 de acuerdo al criterio del E. C. M.?.

PROBLEMA 3. .

Una empresa se dispone a comercializar un nuevo producto y estudia la conveniencia de lanzar una campaña publicitaria previa. Para averiguar si el porcentaje de personas que comprarían el producto aumentaría con esta campaña se llevaron a cabo dos encuestas distintas. La primera encuesta se realizó sobre 100 personas que no habían visto la campaña publicitaria, de las cuales 25 se mostraron interesadas en la compra del producto. En la segunda encuesta, las 100 personas visualizaron previamente la publicidad antes de responder si comprarían el producto, resultando que un total de 30 personas afirmaron su intención de adquirir el producto.

Resolver las siguientes cuestiones:

a) Construir un intervalo de confianza al 90% para la verdadera proporción de personas que comprarían el producto tras haber visto la publicidad.

b) ¿A cuántas personas habrá de encuestar la empresa si quiere asegurarse una longitud del intervalo de confianza para la anterior proporción inferior a 0.1?

c) Construir el intervalo de confianza al 90% para la diferencia de proporciones de compra entre personas que han visto la publicidad y las que no la han visto. En base a este intervalo, ¿podemos afirmar que la campaña es efectiva?.

d) Si las dos encuestas se hubieran realizado al mismo grupo de personas (antes y después de la visualización de la publicidad), ¿sería válido el intervalo que se propone en **c)**?, ¿por qué?. Razonar sin realizar operaciones.

PROBLEMA 4. .

A menudo, la rentabilidad de una empresa se mide por sus activos de retorno (el cociente de los ingresos netos respecto de los activos totales). En la siguiente tabla se presentan los activos de retorno de veinte empresas canadienses: diez del sector servicios y diez de otros sectores.

Servicios	4.28	8.95	9.33	1.25	0.66	0.62	-3.64	-1.01	-3.87	-1.73
Otros	1.33	0.76	0.69	0.72	9.68	0.59	0.89	0.94	1.13	1.01

La variabilidad de los activos de retorno en ambos grupos de empresas se considera análoga.

Indicando las hipótesis asumidas, contestar a las siguientes preguntas:

a) Construir un intervalo de confianza al 95% para la diferencia de medias. ¿Se puede afirmar a la vista del intervalo obtenido que la media de activos de retorno de las empresas del sector servicios es superior a la media de activos de retorno de otros sectores?.

b) Contrastar al 5% la hipótesis nula de que la media en activos de retorno de las empresas del sector servicios es menor que la media en activos de retorno de las empresas de otros sectores.

c) Acotar el $p - valor$ del contraste realizado en **b)**.

LEC/LADE/LEC-DER/LADE-DER
CURSO 1997/98
EXAMEN DE ESTADÍSTICA I
Septiembre de 1998

PROBLEMA 2.

Una compañía de automóviles tiene datos sobre el número de coches nuevos vendidos **X** y el número de vendedores **Y** en una muestra de 200 días.

Los datos son los siguientes:

X/Y	**1**	**2**	**3**
0	18	16	12
1	10	14	16
2	16	12	26
3	20	24	16

a) Encontrar las distribuciones marginales de **X** e **Y** y representar sus funciones de masa de probabilidad. **(1 punto)**

b) Encontrar las probabilidades condicionales $P(X = 2/Y = 1)$ y $P(Y = 2/X = 1)$.

c) Calcular $E(X)$ y $E(X/Y = 1)$.

d) ¿Son **X** e **Y** independientes?. Razonar la respuesta.

PROBLEMA 3.

Un fabricante de turrón afirma que una nueva presentación que consiste en una fotografía de un deportista muy famoso incrementará las ventas del producto en los supermercados en una media de 50 cajas semanales. Para una muestra de 20 supermercados, la media de los incrementos muestrales en las ventas fue de 41.3 cajas, con una desviación típica muestral corregida $\left(s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2}\right)$ de 12.2 cajas.

a) Contrastar al nivel 5% la hipótesis nula de que la media poblacional del incremento en las ventas es al menos de 50 cajas, indicando todos los supuestos que se hagan.

b) Acotar e interpretar el $p - valor$.

PROBLEMA 4.

El cambio en la coyuntura económica de un país lleva a un economista a investigar la rentabilidad de un determinado tipo de inversión en la actual situación. Para ello toma una muestra aleatoria de 9 fondos, encontrándose que la rentabilidad anual de los mismos en términos porcentuales es:

$$10.3 \quad 8.6 \quad 9.4 \quad 8.4 \quad 9.8 \quad 7.5 \quad 12 \quad 11.2 \quad 8$$

a) Especificando las hipótesis necesarias, determinar un intervalo de confianza al 90% para la media de la rentabilidad.

b) Un informe de la OCDE hace suponer que la rentabilidad iguala el 10%. Razonar brevemente si rechazarías esta afirmación mediante un contraste de hipótesis con nivel de significación del 10%.

c) Supongamos que el economista puede asumir que la varianza poblacional de la variable rentabilidad es 4. ¿Qué intervalo de confianza obtendría para la media basándose en las hipótesis anteriores?.

d) El estimador de máxima verosimilitud para la varianza viene dado por $\frac{\sum_{i=1}^{n} x_i{}^2}{n-2} - 9$ y su varianza asintótica es $\frac{\sum_{i=1}^{n} x_i{}^4}{n-2} - 81$. Se realiza un nuevo muestreo, esta vez con 200 fondos para verificar la hipótesis que asume el investigador en el apartado **c)**. Para esta nueva muestra se sabe que:

$$\sum_{i=1}^{200} x_i^2 = 2376 \quad \sum_{i=1}^{200} x_i^4 = 16236$$

¿Qué podemos afirmar de dicha hipótesis?.

LEC/LADE/LEC-DER/LADE-DER
CURSO 1998/99
EXAMEN DE ESTADÍSTICA I
1 de Febrero de 1999

PROBLEMA 2.

a) Sea la variable aleatoria $Y = X^2 + U$, donde X es una variable aleatoria simétrica y las variables aleatorias X y U son independientes, continuas y tienen varianza finita mayor que cero. ?'Cuáles de las siguientes opciones son correctas?. Razonar las respuestas.

i) X e Y son independientes.
ii) X e Y son incorreladas.
iii) U e Y son independientes.
iv) U e Y son incorreladas.

NOTA: Obsérvese que si X es simétrica se verifica que $E[X] = E[X^3] = 0$.

b) De una población con media μ y varianza σ^2 se extraen dos muestras aleatorias simples de tamaños n y m e independientes entre sí. Demuéstrese que la estimación de μ obtenida como la media muestral de las $(n+m)$ observaciones es más eficiente que la estimación que se obtendría como la media muestral correspondiente a cualquiera de las dos muestras.

PROBLEMA 3.

En un departamento de una facultad de medicina se está probando la eficacia de un bloqueante (fentolamina) del sistema nervioso simpático a nivel de la circulación cerebral. Para ello, se toma un lote de animales de experimentación y se miden una serie de parámetros, entre ellos el flujo sanguíneo cerebral. Si el bloqueante es eficaz aumentará el flujo sanguíneo, pues el sistema nervioso simpático actúa como vaso constrictor. Se pretende estudiar la evolución del efecto del bloqueante y para ello se mide el flujo antes, inmediatamente después, 24 horas más tarde y pasados 4 ó 5 días de la inyección del bloqueante. Los datos obtenidos con 5 animales y medidos en mililitros por minuto son los siguientes:

Animal	Antes	Inmediato	24 horas	4-5 días
1	57	75	68	60
2	54	78	65	57
3	56	71	62	55
4	60	80	70	67
5	55	73	66	58

Se pide responder a las siguientes preguntas:

a) Construir un intervalo de confianza al 90% para la diferencia de medias de las observaciones inmediatas y anteriores.

b) Contrastar al 5% de nivel de significación si existe evidencia en la muestra de que el bloqueante es eficaz a las 24 horas. Establecer las hipótesis necesarias.

c) Acotar el p-valor del contraste de hipótesis del apartado **b)**.

d) Si se comparan las observaciones inmediatas y a los 4-5 días, se puede comprobar la evolución del efecto del bloqueante. Sea $D = (\textbf{Inmediato}) - (\textbf{4-5 días})$. Suponiendo $\sigma_D^2 = 9$ para este caso, contrastar la hipótesis de que la media de las diferencias entre datos inmediatos y a los 4-5 días es igual a 15 unidades, frente a la hipótesis de un valor medio inferior a 15 unidades.

e) Calcular la potencia del contraste del apartado **d)** suponiendo un valor medio alternativo de 10 unidades.

PROBLEMA 4.

En una encuesta aparecida el día 17 de Noviembre de 1996 en el diario ABC, ante la pregunta: ?'Está usted de acuerdo con la afirmación de que "en los colegios de enseñanza primaria debería haber mucha más disciplina con los alumnos"?, se obtuvieron los siguientes porcentajes:

Respuesta	**Porcentaje**
Nada de acuerdo	18%
Poco de acuerdo	29%
Bastante de acuerdo	33%
Muy de acuerdo	19%
NS/NC	1%

Para verificar si estos datos siguen siendo ciertos hoy día, un grupo de estudiantes hace la misma pregunta a 100 personas elegidas en la calle al azar, y obtiene los siguientes resultados:

Respuesta	**No. de personas**
Nada de acuerdo	10
Poco de acuerdo	30
Bastante de acuerdo	25
Muy de acuerdo	20
NS/NC	15

a) Teniendo en cuenta los resultados de la muestra de 100 personas, ¿se pueden considerar admisibles en la actualidad los resultados del año 1996?.

b) ¿Qué supuesto teórico no se verifica estrictamente en el contraste que has realizado en el apartado **a)**?.

LEC/LADE/LEC-DER/LADE-DER
CURSO 1998/99
EXAMEN DE ESTADÍSTICA I
Septiembre de 1999

PROBLEMA 2.

a) Sea $X_1, ..., X_n$ una muestra aleatoria de una población de media μ y varianza σ^2. Considérense los siguientes estimadores de μ:

$$\widehat{\mu}^{(1)} = \frac{X_1}{4} + \frac{1}{2}\left[\frac{1}{(n-2)} \sum_{i=2}^{n-1} X_i\right] + \frac{X_n}{4} \qquad \widehat{\mu}^{(2)} = \frac{1}{n}\left[\sum_{i=1}^{n} X_i\right]$$

i) Demostrar que los dos estimadores son insesgados.
ii) Determinar la eficiencia relativa de $\widehat{\mu}^{(2)}$ respecto de $\widehat{\mu}^{(1)}$.
iii) Calcular el error cuadrático medio de los dos estimadores.

b) Un grupo de ecologistas se adentran en alta mar para recoger bidones con desechos tóxicos. La distribución de probabilidad para la variable aleatoria $X \equiv$ "cantidad de bidones recogidos el primer día en esa región de alta mar" es:

x	3	4	5
$P(X = x)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

Las probabilidades condicionadas $P(Y/X)$, siendo Y la variable aleatoria "cantidad de bidones recogidos el segundo día en la misma región de alta mar", son las que a continuación se muestran:

$$\begin{array}{lll} P(Y=0/X=3)=\frac{2}{4} & P(Y=0/X=4)=\frac{1}{2} & P(Y=0/X=5)=\frac{3}{4} \\ P(Y=1/X=3)=\frac{1}{4} & P(Y=1/X=4)=\frac{1}{2} & P(Y=1/X=5)=\frac{1}{4} \\ P(Y=2/X=3)=\frac{1}{4} & P(Y=2/X=4)=0 & P(Y=2/X=5)=0 \end{array}$$

i) ¿Son X e Y independientes?. Razonar la respuesta.
ii) Un millonario ecologista paga 50 euros por cada día y 10 euros por bidón, ¿cuál es el beneficio esperado por los ecologistas después de dos días en una región nueva de similares características?.

PROBLEMA 3.

Un constructor está considerando dos lugares alternativos para la construcción de un centro comercial regional. Como los ingresos de los hogares de la comunidad son una consideración a tener en cuenta en esa selección, desea probar la hipótesis nula de que no existe diferencia significativa entre el ingreso promedio por hogar de las dos comunidades. Consistente con esta hipótesis, supone que la varianza del ingreso por hogar es también igual en las dos comunidades. Para una muestra de 13 hogares de la primera comunidad, obtiene que el ingreso promedio es de 35500 dólares, con una desviación típica muestral corregida de 1900 dólares. Para una muestra de 16 hogares de la segunda comunidad, obtiene que el ingreso promedio es de 34600 dólares, con una desviación típica muestral corregida de 2400 dólares.

a) Contrastar al nivel 5% la hipótesis de igualdad de varianzas. Precisar las hipótesis asumidas.

b) Contrastar al nivel 5% la hipótesis de igualdad de ingresos promedio en las dos comunidades. Precisar las hipótesis asumidas.

c) Contrastar al nivel 1% la hipótesis de que el promedio de ingresos en la primera comunidad es superior al de la segunda (tómese como hipótesis alternativa $H_1 \equiv \mu_1 > \mu_2$). Precisar las hipótesis asumidas.

d) Acotar el $p - valor$ del contraste realizado en **c**).

PROBLEMA 4.

Para estimar el coste medio de alquiler en una ciudad, cada uno de los 60 entrevistadores contratados visita 100 pisos y calcula un intervalo de confianza al 95% para el coste medio de alquiler en la población.

a) ¿Qué hipótesis han tenido que asumir los entrevistadores para poder calcular el intervalo de confianza?.

b) ¿Cuántos de estos intervalos deben contener la media poblacional?.

c) Uno de los intervalos obtenidos es (85000,110000). Obtener la media muestral y la desviación típica muestral corregida de la muestra correspondiente a dicho intervalo.

d) Uno de los entrevistadores concluye que menos del 30% de los alquileres son superiores a 100000 pesetas. Si de los 6000 pisos analizados, 1680 alquileres son superiores a 100000 pesetas, ¿podemos aceptar la conclusión a la que ha llegado este entrevistador?.

LEC/LADE/LEC-DER/LADE-DER
CURSO 1999/2000
EXAMEN DE ESTADÍSTICA I
29 de Enero de 2000

PROBLEMA 2:

Dos hinchas, uno del TAUG y otro del NIX discuten sobre el promedio de goles por partido que marcará su equipo en la liga. Sólo disponen de datos de la primera fase de la liga. El TAUG estuvo ausente una semana, por lo que jugó un partido menos que el NIX. Los hinchas finalmente aceptan terminar su discusión con un CONTRASTE DE HIPÓTESIS al nivel de significación del 10%.

El seguidor del TAUG sostiene que el promedio de goles por partido de su equipo será como poco, un gol más que el promedio del NIX (tómese como H_0). Los datos son

n_{NIX} =16, $\overline{x}_{TAUG} = 2.25$ goles, $\overline{x}_{NIX} = 1.75$ goles, $S_{TAUG} \approx 1.8$ goles y $S_{NIX} \approx 1.6$ goles (donde S^2 denota la cuasi-varianza muestral)

a) Se ha realizado un contraste bilateral con $H_0 : \sigma^2_{TAUG} = \sigma^2_{NIX}$, obteniéndose un p-valor de 0.65, ¿ cuál es la conclusión de este contraste?.¿ Qué hipótesis se han asumido para realizar el contraste?.

b) Plantear el Contraste de hipótesis que resuelve la discusión de los hinchas, especificando las hipótesis, el estadístico y la región de contraste (región de rechazo).

c) Realizar el contraste de b), interpretar el resultado y acotar el p-valor.

PROBLEMA 3:

Se dispone de una muestra aleatoria simple de tamaño n de la variable aleatoria cuya densidad es

$$f(x) = \tfrac{1}{2}\theta^3 x^2 e^{-\theta x},\ x > 0$$

siendo θ un parámetro positivo desconocido. Se pide:

a) Determinar el estimador de máxima verosimilitud $\widehat{\theta}$ de θ.

b) Demostrar que la varianza asintótica estimada de $\widehat{\theta}$ es $\widehat{V[\widehat{\theta}]} = \frac{\widehat{\theta}^2}{6n}$.

c) Calcular $\widehat{\theta}$ y $\widehat{V[\widehat{\theta}]}$ para una muestra de tamaño n=250 y $\sum_{i=1}^{250} x_i = 432$.

d) Calcular un intervalo de confianza asintótico al nivel de confianza del 95% para la muestra del apartado (c).

PROBLEMA 4:

a) Calcula de forma teórica la media y la varianza de la variable aleatoria media muestral.

.

b) Una ONG dedicada al cuidado de animales en libertad ha recibido una donación de un socio, de cables para sujetar las casetas de pájaros a los árboles. El cable resiste un peso no superior a 350 gramos.

El peso de cada caseta X, se distribuye como una normal de media 250 gramos y desviación típica 25.5 gramos. El peso de cada pájaro, Y, se distribuye normalmente con media 30 gramos y desviación típica 2.5 gramos. Cada caseta alberga como máximo a tres pájaros.

i) Obtener la distribución del peso total de las casetas que albergan exactamente tres pájaros. Indicar las hipótesis asumidas.

ii) La ONG decide hacer un estudio previo sobre la utilidad del cable donado. Toma para ello una muestra aleatoria de 100 pesos totales de casetas que albergan exactamente tres pájaros. La media muestral de los 100 pesos totales observados, ¿es menor que 350 gramos en más del 95% de las casetas? ¿te parece útil la donación del cable?

LEC/LADE/LEC-DER/LADE-DER
CURSO 1999/2000
EXAMEN DE ESTADÍSTICA I
Septiembre de 2000

PROBLEMA 2.

Sean X e Y dos variables normales con igual media μ y desviación típica σ. Se pide:

a) Probar que para cualquier constante C se verifica que $Cov(X, -X + C) < 0$.

b) Suponiendo que se verifica que $Cov(X - Y, Y - X + 3) = -2\sigma^2$, demostrar que X e Y son independientes.

c) Razonar si es verdadera o falsa la siguiente afirmación, en el caso de que X e Y estuvieran incorreladas:

$$\Pr ob(X + Y - 2\mu \geq \sigma) = \Pr ob(2X - 2\mu \geq \sigma)$$

PROBLEMA 3.

Se quiere estudiar el porcentaje de estudiantes que abandonan una carrera universitaria en el primer curso.

a) ¿ A cuántos estudiantes habrá que entrevistar para que la longitud del intervalo de confianza al 99% para la proporción de estudiantes que abandonan los estudios no sea superior a 0.2 ?

b) Para comprobar si existen diferencias entre la proporción de estudiantes masculinos que abandonan los estudios y la proporción de estudiantes femeninas, se escoge una muestra de 85 hombres de los cuales 7 abandonaron la carrera y otra muestra de 115 mujeres de las cuales 10 abandonaron. Construir un intervalo de confianza al 95% para la diferencia entre las dos proporciones y decidir si existe una diferencia significativa entre las dos proporciones observadas.

c) Escribir las hipótesis necesarias para realizar los dos apartados anteriores.

PROBLEMA 4.

En un laboratorio químico se está investigando un nuevo producto para la conservación del agua en las piscinas. Se eligen 16 piscinas aleatoriamente y se mezcla el agua con dicho producto. Se mide el número de días que permanece el agua en buen estado, obteniéndose:

$\overline{X} = 28\ días \qquad S = 3\ días$ (siendo S^2 la cuasivarianza)

Si se admite que la variable respuesta sigue una distribución normal,

a) Calcular intervalos de confianza para la media y para la varianza de la variable respuesta al 99% especificando las hipótesis asumidas.

b) Qué tamaño de muestra se necesita para que al nivel de confianza del 95%, la longitud del intervalo de confianza para la media sea de 2 días (utilizar aproximación normal).

c) Una vez lanzado al mercado el producto, el laboratorio afirma que el agua permanecerá en buen estado durante 30 días exactamente. Contrastar esta afirmación al 5% de significación. Indicar las hipótesis asumidas.

d) Acotar el p-valor del contraste anterior interpretando el resultado.

LEC/LADE/LEC-DER/LADE-DER
CURSO 2000/01
EXAMEN DE ESTADÍSTICA I
31 de Enero de 2001

PROBLEMA 2:

Sea $(X_1, X_2, ..., X_{n_1})$ una muestra aleatoria simple de la variable aleatoria X, que se distribuye como una $N(\mu_1, \sigma^2)$. Sea $(Y_1, Y_2, ..., Y_{n_2})$ una muestra aleatoria simple de la variable aleatoria Y, que se distribuye como una $N(\mu_2, \sigma^2)$. X e Y son independientes.

Se quiere estimar σ^2 y para ello se consideran combinaciones lineales de las cuasi-varianzas muestrales S_1^2 y S_2^2, es decir,

$$\widehat{\sigma}^2 = \lambda S_1^2 + (1-\lambda)S_2^2, \text{ donde } 0 \square \lambda \square 1$$

a) Demostrar que este estimador es centrado para cualquier valor de λ

b) Determinar qué valor de λ define el estimador $\widehat{\sigma}^2$ más eficiente

(Indicación: $Var(\chi_n^2) = 2n$)

PROBLEMA 3:

El número de cuentas corrientes, X, y de ahorro, Y, que posee una familia elegida al azar, en un Banco, está representada en la siguiente tabla:

X/Y	0	1	2	3
0	.30	.10	.03	.02
1	.08	.15	.10	.02
2	.02	.05	.07	.06

a) ¿ Son X e Y variables independientes?. Razonar la respuesta.

b) Calcular el número medio de cuentas que una familia tiene en este Banco y su desviación típica.

c) ¿Qué proporción de familias posee el mismo número de cuentas corrientes que de ahorro?

d) Obtener la distribución del número de cuentas corrientes para las familias con 2 cuentas de ahorro.

e) Se eligen 3 familias al azar. Calcular el número medio de cuentas corrientes que poseen entre las 3, así como su desviación típica.

PROBLEMA 4:

Se ha obtenido el intervalo de confianza (3,5) para la media de una variable aleatoria X, con varianza poblacional igual a 4, a partir de una muestra aleatoria simple de tamaño 16.

a) Determinar la estimación puntual de la media poblacional de X, así como el nivel de confianza del intervalo anterior.

Especificar las hipótesis asumidas.

b) Hallar el p-valor obtenido al contrastar que la media poblacional es $4{,}5$ frente a la hipótesis alternativa de que es menor que dicho valor. Dar la distribución del estadístico del contraste.

Especificar claramente las hipótesis asumidas.

c) Para contrastar la normalidad de la variable aleatoria X, se ha tomado otra muestra de mayor tamaño y se han agrupado los datos en 8 clases.

Se sabe que el estadístico χ^2 del contraste ha tomado el valor 6. Para un nivel de significación del 5% y estimando la media de la distribución normal mediante los datos, ¿ es apropiado aceptar la normalidad de X?

LEC/LADE/LEC-DER/LADE-DER
CURSO 2000/01
EXAMEN DE ESTADÍSTICA I
Septiembre de 2001

PROBLEMA 2:

Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de piezas independientes y no identicamente distribuidas tomadas de un proceso de fabricación, donde la variable X_i sigue una ley exponencial, $Exp(1/i\theta)$, siendo i el orden de elección de la muestra; esto es, para X_1, $i = 1$, para X_2, i=2,...

Se pide:

a) Encontrar el estimador máximo verosímil para θ

b) Encontrar el estimador máximo verosímil para $\delta = \sqrt{\theta}$

Indicación: La función de densidad de una variable aleatoria exponencial(λ) es $f(x) = \lambda e^{-\lambda x}$, $x > 0$

PROBLEMA 3:

El gerente de operaciones del periódico de una gran ciudad quiere determinar la proporción de periódicos impresos con defectos, como demasiada tinta, configuración de páginas incorrecta, páginas duplicadas, etc.

El gerente decide tomar una muestra aleatoria de 100 periódicos y encuentra que 35 contienen algún tipo de defecto.

a) Si el gerente quiere un 90% de confianza en estimar la proporción verdadera de periódicos impresos con defectos, construye el intervalo de confianza.

b) Utilizando la información muestral, determinar el tamaño de la muestra para que el error de estimación no sea superior al 5%, con un nivel de confianza del 90%.

c) Si no se dispone de la información muestral, ni de información histórica fiable (caso más desfavorable), plantear el cálculo de n para el supuesto del apartado b).

PROBLEMA 4:

La llegada de aviones a cierto aeropuerto sigue una distribución Poisson (λ). Se realiza un estudio para determinar si es conveniente ampliar el aeropuerto. Para ello se analiza la distribución de la variable X: " número de llegadas al aeropuerto cada hora ". Tras observar una muestra aleatoria simple de 60 períodos de una hora, estos son los resultados obtenidos:

Valor de X	0	1	2	3	4	5	6	7	8	9
Frecuencia	1	2	13	7	10	16	8	2	0	1

Se pide:

a) Sabiendo que $V[X] = E[X] = \lambda$, y que el estimador de máxima verosimilitud para la media de la distribución es la media muestral, hallar una estimación puntual del parámetro basándonos en las observaciones obtenidas. Determinar la distribución de dicho estimador y decir si es exacta o aproximada.

b) Teniendo en cuenta las anteriores estimaciones, contrastar "el número medio de llegadas a la hora es superior a 5 ", (tómese como hipótesis alternativa). Especificar las hipótesis asumidas.

c) La siguiente tabla nos ofrece los valores esperados que se hubieran obtenido si las observaciones siguiesen realmente una distribución de Poisson de parámetro 4:

Valor de X	Menos de 2	2	3	4	5	6	7 o más
Frecuencia esperada	5,49	8,79	11,72	11,72	9,37	6,25	6,64

Determinar si existe evidencia significativa (al 1%) de que la muestra sigue una distribución de Poisson.

LEC/LADE/LEC-DER/LADE-DER
CURSO 2001/02
EXAMEN DE ESTADÍSTICA I
31 de Enero de 2002

PROBLEMA 2:

Sea $(X_1, X_2, ..., X_n)$ una muestra aleatoria simple de una variable aleatoria X con $E(X) = \mu$ y $Var(X) = \sigma^2$. Calcular el error cuadrático medio para los siguientes estimadores del parámetro μ.

a) $T_1 = X_1$

b) $\mathrm{T}_2 = (3X_1 - 2X_2 + X_3)/6$

PROBLEMA 3:

Cierto supermercado tiene una caja general de salida y una caja rápida. Sea X el número de clientes que están en espera en la caja general en un momento particular del día. Sea Y el número de clientes en espera en la caja rápida al mismo tiempo. Supongamos que la distribución de probabilidad conjunta de X e Y es como se indica en la tabla siguiente:

X\Y	0	1	2	3
0	0,08	0,07	0,04	0
1	0,06	0,15	0,05	0,04
2	0,05	0,04	0,1	0,06
3	0	0,03	0,04	0,07
4	0	0,01	0,05	0,06

a) ¿ Cuál es la probabilidad de que haya exactamente un cliente en cada línea de espera ?

b) ¿ Cuál es la probabilidad de que los números de clientes en las dos líneas de espera sean iguales ?

c) ¿ Cuál es la probabilidad de que haya por lo menos dos clientes más en una línea de espera que en la otra ?

d) ¿ Cuál es la probabilidad de que el número total de clientes entre las dos líneas de espera sea exactamente 4 ?

e) Determinar la distribución marginal de X y calcular el número esperado de clientes de la línea de espera en la caja general.

f) ¿ Cuál es la probabilidad de que haya por lo menos dos clientes en espera en la caja general si en la caja rápida hay 3 clientes ?

g) ¿ Son las variables aleatorias X e Y independientes ?

PROBLEMA 4:

El euro ha traído a nuestra vida diaria muchos datos nuevos para analizar. Vamos a fijarnos en cómo se desarrollan los pagos y devoluciones en los primeros días de implantación del euro. Se dispone de una muestra aleatoria simple de compras realizadas en distintos puntos de España y se obtienen los siguientes datos

	Grandes Superficies	Pequeño comercio
Se paga y se devuelve en euros	150	50
Se paga en pesetas y se devuelve en euros	200	150
Se paga en pesetas y se devuelve en pesetas	50	100

a) Estimar la "proporción de compras con pago y devolución en euros".

b) Se desea estimar la misma proporción que en a) con una confianza del 95%. Si el error de estimación (semilongitud del intervalo) no debe ser superior a 0.03, utilizando la información muestral ¿ qué tamaño muestral garantiza este supuesto ?, ¿ se consigue este supuesto con nuestra muestra de 700 compras ?

c) La prensa afirmaba que al menos el 75% de las compras se realizaba en euros (pagos y devoluciones). ¿ Existe evidencia muestral para no aceptar esta hipótesis con nuestros datos ?

d) Mucha gente se dirigió a los establecimientos que al pagar en pesetas devolvía en euros. Considerando las proporciones de estas compras para cada tipo de establecimientos, ¿es significativa la diferencia entre grandes superficies y pequeños comercios ?

e) Fijándonos ahora en las compras cuyos pagos y devoluciones ignoraban la nueva moneda, hemos podido calcular un intervalo de confianza para "la diferencia de las proporciones de compras" entre grandes superficies y pequeños comercios. Este es $(-0,2742, -0,1424)$. Calcular el nivel de confianza para este intervalo. ¿ Es significativa la diferencia entre ambos tipos de establecimientos ?.

LEC/LADE/LEC-DER/LADE-DER

CURSO 2001/02
EXAMEN DE ESTADÍSTICA I
3 de Septiembre de 2002

PROBLEMA 2:

La Universidad posee un considerable volumen de ascensores y elevadores en sus tres campus, cuyo mantenimiento realiza una empresa que afirma que al menos el 99% de los días lectivos se encuentran funcionando todos los ascensores y elevadores. La Universidad ha seleccionado aleatoriamente 41 días lectivos. Sólo en 38 de ellos el funcionamiento de ascensores y elevadores era total.

a) ¿ es aceptable la hipótesis que afirma la empresa de mantenimiento ?

b) Lo que la Universidad quiere contrastar es que al menos el 90% de los días lectivos el funcionamiento de ascensores y elevadores sea total. ¿ debe la Universidad, a la vista de sus datos, cambiar de empresa de mantenimiento ?

PROBLEMA 3:

Un fabricante de cigarrillos desea anunciar el contenido medio de nicotina de sus cigarrillos, para ello el laboratorio realizó 15 determinaciones de contenido. Los datos obtenidos fueron

$\overline{X} = 25.6\ mg, s = 2.3\ mg(S^2$ es la cuasivarianza muestral)

a) Precisando todas las hipótesis que debes asumir, emplea estos datos para construir un intervalo de confianza al 98%.

b) ¿ Es posible que este intervalo no contenga el valor medio de la muestra ?

c) Suponiendo $\sigma = 2.5$ mg, ¿ cuál sería el tamaño mínimo de muestra que debes tomar para obtener un I.C. al 95% de longitud 4 ?

PROBLEMA 4:

El tiempo de germinación de cierta planta se distribuye según una v.a. cuya función de densidad es

$$f(x) = \begin{cases} \lambda e^{-\lambda(x-20)}, x \geq 20 \\ 0, resto \end{cases}$$

a) Se toma una m.a.s. de tamaño n. Calcular el estimador de M.V. para λ.

b) Si se toma una m.a.s. de 100 plantas y se obtiene $\sum x_i = 2200$

i) ¿ Cuál es la estimación máximo verosímil de λ ?

ii) ¿ Cuál es la estimación de la Varianza Asintótica del estimador de máxima verosimilitud de λ ?

c) Con la muestra del apartado anterior, contrastar al 1% y al 5% la hipótesis nula de que $\lambda = 0.6$. Comentar el resultado obtenido.

LEC/LADE/LEC-DER/LADE-DER
CURSO 2002/03
EXAMEN DE ESTADÍSTICA I
8 de Febrero de 2003

PROBLEMA 2:

Un centro comercial quiere estimar la edad media de los clientes que visitan la tienda de productos contemporáneos. El gerente asume que la edad de estos clientes sigue una distribución Normal con una desviación típica igual a 5.

a) Si el gerente quiere estimar la edad media de estos clientes con un error de estimación no superior a 2 años con un intervalo de confianza del 95%, ¿ cuál sería el tamaño muestral necesario ?

b) Se han elegido aleatoriamente 40 de estos clientes y se ha obtenido una edad media de 29 años. Calcular un intervalo de confianza del 99% para la edad media poblacional.

c) En una muestra aleatoria de 10 clientes se ha obtenido una edad media de 29 años con una cuasivarianza de 24. Analizar si la desviación típica igual a 5 asumida por el gerente es aceptable o no al 10% de significación.

d) En la muestra de 40 clientes, 18 eran menores de 30 años. Construir un intervalo de confianza del 85% para la proporción poblacional de clientes menores de 30 años. ¿ Podemos concluir que el 50% de los clientes son menores de 30 años ?

(2.5 puntos)
PROBLEMA 3:

El consumo de un cierto producto en una familia de cuatro miembros durante los meses de verano, es una variable aleatoria con distribución uniforme en el intervalo $(\alpha, \alpha+1)$.
Sea $(X_1, ..., X_n)$ una muestra aleatoria de consumos de distintas familias.

a) Demostrar que la media muestral es un estimador sesgado de α y que sus sesgo es $\frac{1}{2}$.

b) Calcular el error cuadrático medio de $\overline{x}$.

c) Obtener un estimador insesgado de α (a partir de $\overline{x}$).

Nota.- La función de densidad de una $U(a, b)$ es $f(x) = \frac{1}{b-a}$ y sus momentos son E(X)=$\frac{a+b}{2}$ y V(X)=$\frac{(b-a)^2}{12}$.

(2.5 puntos)

PROBLEMA 4:

El tiempo en minutos que tarda un autobús en llegar a una parada se distribuye según una variable aleatoria cuya función de densidad es:

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}}, & si\ x > 0 \\ 0, en\ el\ resto \end{cases}$$

con $\theta \neq 0$.

(Recordar que $E(X) = \theta$ y V(X)=θ^2)

a) verificar que el estimador de máxima verosimilitud $\widehat{\theta}_n$ es la media muestral $\overline{X}$.

b) demostrar que la varianza asintótica estimada es $\frac{\overline{X}^2}{n}$.

c) verificar que $\widehat{\theta}_n$ es insesgado y consistente.

d) seleccionando una muestra de tamaño 50 obtenemos que la suma de los tiempos de tardanza de la totalidad de los autobuses es $\sum_{i=1}^{50} x_i = 500$. Calcular un intervalo de confianza aproximado del 98% para el parámetro poblacional θ.

(2.5 puntos)

Nota:-

Las notas se publicarán a partir del **Miércoles 12 a última hora.**

La revisión del examen se podrá solicitar hasta el **Lunes 17 a las 13:00 en el despacho 10.1.29**

LEC/LADE/LEC-DER/LADE-DER
CURSO 2003/04
EXAMEN DE ESTADÍSTICA I
9 de Febrero de 2004
TIEMPO MÁXIMO= 3 h.

PROBLEMA 1. (2,5 puntos)

a) Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria simple extraida de una población con media μ y varianza σ^2.

Demostrar que $E[\bar{X}] = \mu$ y $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.

b)Examinados los incrementos salariales de los altos ejecutivos de un amplio grupo de empresas se observa que tienen una distribución normal de media 12,1% y desviación típica 3,5%. Se toma una muestra aleatoria de 16 observaciones de la población de incrementos salariales.

b.1)¿Cuál es el error estándar (desviación típica) de la media muestral de incrementos salariales?

b.2)¿Cuál es la probabilidad de que la media muestral sea igual o inferior al 10%?

b.3)¿Cuál es la probabilidad de que la desviación típica muestral sea superior a 4,52 %?

PROBLEMA 2. (2,5 puntos)

En un sondeo de opinión llevado a cabo en el año 2002 con 1000 individuos en las ciudades de Murcia y Cartagena se encontró que el 55% de ellos eran favorables a la construcción de un nuevo aeropuerto. Dicho sondeo se repitió en el año 2003 tras una intensa campaña publicitaria, encontrándose esta vez un 60% de individuos favorables a esta opción, entre 1500 encuestados. Se pide:

a) Construir un intervalo de confianza de nivel 95% para la proporción p_1 de individuos favorables a la construcción del nuevo aeropuerto, de entre los encuestados en el año 2002. ¿Podría afirmarse al nivel de significación del 5% que sólo la mitad de los encuestados preferían un nuevo aeropuerto?

b) Razonar sin resolver numéricamente si la respuesta a la última pregunta sería o no la misma si el nivel de significación fuera igual a 10%.

c) Planteando los supuestos necesarios, construir un intervalo de confianza al 95% para la diferencia $p_2 - p_1$ de las proporciones de individuos favorables al nuevo aeropuerto en los años 2003 y 2002. ¿Podría afirmarse al nivel de significación del 5% que la campaña publicitaria en favor de la construcción del nuevo aeropuerto pudo tener efecto?

PROBLEMA 3. (2,5 puntos)

Sea $X_{1,\ldots,}X_n$ una muestra aleatoria de tamaño n de la función de distribución $Exponencial\ \ Truncada(\mu,\theta)$; donde $\mu > 0$ y $\theta > 0$. La función de distribución anterior tiene función de densidad

$$g(x) = \begin{cases} 0 & si \quad x \leq \mu \\ \frac{1}{\theta}\exp(-\frac{(x-\mu)}{\theta}) & si \quad x > \mu \end{cases}$$

a) Suponiendo conocido el valor del parámetro μ , deducir la expresión del estimador de máxima verosimilitud de θ.

b) Si $\mu = 10$ y la media muestral de un determinado conjunto de datos es 15.5, ¿cuál es el valor de la estimación por máxima verosimilitud de θ?

PROBLEMA 4. (2,5 puntos)

La gerente de un hospital decide realizar un análisis para estimar el modelo de probabilidad que sigue el tiempo de espera en urgencias. Para ello toma una muestra aleatoria de 44 datos que porporcionan un tiempo medio de espera de 19 minutos y medio. En primer lugar comprueba si el tiempo de espera se distribuye como una exponencial (parámetro desconocido). Ésta es la salida obtenida mediante el programa de statgraphics :

```
Goodness-of-Fit Tests for Tiempo de Espera en Urgencias

                              Chi-Square Test
----------------------------------------------------------------------------
               Lower         Upper      Observed      Expected
               Limit         Limit     Frequency     Frequency    Chi-Square
----------------------------------------------------------------------------
         at or below           5,0             5          9,91          2,43
                 5,0          10,0             6          7,68          0,37
                10,0          15,0             7          5,95          0,19
                15,0          25,0            12          8,18          1,78
                25,0          40,0            13          6,57          6,29
above           40,0                           1          5,71          3,89
----------------------------------------------------------------------------
Chi-Square =    with d.f.     P-Value =
```

a) Completar los datos que faltan en este contraste (valor del estadístico, grados de libertad, acotar p-valor) y establecer qué hipótesis hemos tenido que asumir.

b) ¿Podemos afirmar con un nivel de significación del 10% que la variable tiempo de espera sigue una distribución exponencial? Razonar brevemente la respuesta apoyándose en el apartado anterior.

c) Si hubiéramos partido de que los 19 minutos y medio no era una información muestral sino la esperanza de la variable bajo estudio (es decir, la inversa de su parámetro),¿Cómo afectaría este hecho al contraste (nivel de significación del 10%) ?

La gerente lleva a cabo un segundo análisis para determinar si una distribución normal es una mejor o peor aproximación al modelo de probabilidad del tiempo de espera:

Goodness-of Fit Tests for Tiempo de Espera en Urgencias
Chi-Square = 4,51446 with 3 d.g. p-value = 0,211001
Estimated Kolmogorov statistic DPLUS = 0,107434
Approximate p-Value = 0,690107

d) A la vista de todos los resultados, ¿qué decisión tomarías? Justifica la respuesta utilizando la información facilitada por todos los contrastes.

NOTA:

1) Las notas del examen saldrán a última hora de la tarde del jueves 12 de febrero.

2) Las normas para solicitar revisión de examen se pondrán en los tablones el mismo día 12 febrero.

LEC/LADE/LEC-DER/LADE-DER
CURSO 2003/04
EXAMEN DE ESTADÍSTICA I
1 de Septiembre de 2004
TIEMPO MÁXIMO= 3 h.

PROBLEMA 1. (3 puntos)

Supongamos que la variable X sigue una distribución de Poisson de parámetro λ,

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, 2, \ldots$$

sabiendo que $E(X) = Var(X) = \lambda$. Supongamos que disponemos de una muestra aleatoria simple, $X_1, X_2, \ldots, X_n$, de la variable X.

a) Construir el estimador de máxima verosimilitud de λ.

b) Dar la expresión del intervalo de confianza asintótico (aproximado) para λ con un nivel de confianza del $(1 - \alpha)\%$

c) A partir de una muestra de 100 observaciones se tienen los siguientes resultados (obtenidos en Statgraphics):

Summary Statistics for DATOS EJERCICIO
Count = 100
Average = 3,15
Variance = 4,00758
Standard deviation = 2,00189
Minimum = 0,0
Maximum = 10,0
Range = 10,0
Stnd. skewness = 2,19304
Stnd. kurtosis = 0,607952

A partir de estos datos, obtener la estimación puntual, $\widehat{\lambda}$, así como el correspondiente intervalo de confianza para λ al 95%.

d) Utilizando el intervalo obtenido, ¿se puede rechazar la hipótesis nula $H_0 : \lambda = 3$ contra $H_1 : \lambda \neq 3$?

¿Y la hipótesis nula $H_0 : \lambda = 4$ contra $H_1 : \lambda \neq 4$?¿A qué nivel de significación se han realizado estos contrastes?.

PROBLEMA 2. (2,5 puntos)

Para tratar cierta enfermedad se pueden utilizar varios medicamentos. Un vendedor de una compañía farmacéutica está interesado en el número de recetas de un nuevo medicamento para tratar dicha enfermedad. Se tomó una muestra aleatoria de 10 zonas de una gran ciudad y se recogieron los siguientes números de recetas por zona:

210-240-190-275-290-265-312-284-261-243

a) Calcular un estimador puntual de la media de la distribución de la media muestral del número de recetas por zona.

b) Calcular un estimador insesgado de la varianza de la media muestral del número de recetas por zona.

c) Estimar la media del número de recetas mediante un intervalo de confianza al 90%. Indicar las hipótesis necesarias.

d) Estimar la varianza del número de recetas mediante un intervalo de confianza al 95%. Indicar las hipótesis necesarias.

e) ¿Cómo se comportaría la amplitud de los intervalos calculados en c) y en d) si el nivel de confianza pasara a ser el 99%?

PROBLEMA 3. (2,5 puntos)

Un artículo de la revista "Consumer Reports" de noviembre de 1983 comparó varios tipos de baterías. El promedio de duración de baterías AA alcalinas marca Duracell y de baterías AA alcalinas marca Eveready Energizer se denotan como μ_1 y μ_2, respectivamente. Supongamos éstos los promedios de duración poblacionales.

a) Se toma una muestra de 19 baterías marca Duracell cuya duración media es de 4,4 horas y cuasivarianza (varianza corregida) es de $(1,2)^2$. Contrastar la hipótesis nula de que la desviación típica poblacional de las baterías marca Duracell es de 1,8 horas a un nivel de significación de 5% indicando las hipótesis necesarias para realizar dicho contraste.

b) Un mayorista juguetero está interesado en demostrar que el promedio de duración de las baterías marca Duracell supera en más de una hora a la duración media de las baterías marca Eveready, ya que le cuestan más baratas las primeras. Para ello, toma dos muestras de 100 baterías cada una de ambas marcas obteniendo los siguientes resultados:

Duración media de las 100 baterías Duracell=4,3 horas.

Desviación típica (corregida) de las 100 baterías Duracell=1,5 horas.

Duración media de las 100 baterías Eveready=4,1 horas.

Desviación típica (corregida) de las 100 baterías Eveready=1,7 horas.

¿Hay evidencia suficiente para aceptar la hipótesis del juguetero?

PROBLEMA 4. (2 puntos)

En un área metropolitana se selecciona una muestra de 250 personas al azar y se les pregunta cuántas veces habían comprado un cierto producto durante el mes anterior.

Los resultados son los de la tabla adjunta.

Veces que compran	Nº de personas
0	85
1	80
2	45
3	20
4	15
5	5

Contrastar con un 95% de confianza la hipótesis de que las observaciones constituyen una muestra de una población que se distribuye como una Poisson.

Nota: La función de masa de probabilidad de una variable de Poisson de parámetro λ aparece en forma explícita en el enunciado del ejercicio 1.

NOTA :

Las notas se publicarán a última hora del lunes 6 de septiembre.

Las normas para solicitar revisión de examen aparecerán en los tablones junto con las notas de la asignatura.

El plazo de solicitud de revisión finaliza el miércoles 8 de septiembre a las 10 horas.

LEC/LADE/LEC-DER/LADE-DER/LEC-PERIODISMO
CURSO 2004/05
EXAMEN DE ESTADÍSTICA I
15 de Febrero de 2005
TIEMPO MÁXIMO= 3 h

1. **(2,5 puntos).** El tiempo en minutos que dura un viaje en tren entre dos ciudades A y B, es una variable aleatoria normal con varianza 16. Hemos extraído una muestra aleatoria simple de 10 viajes, obteniendo los siguientes tiempos:

50	70	55	60	62	69	60	72	58	62

(a) Deducir la expresión de la desviación típica de la media muestral, y calcularla para esta muestra.

(b) ¿Entre qué valores se encontrará la verdadera media, μ, del tiempo que dura el viaje en tren de A a B, a un nivel de confianza al 95%?

(c) Se desea realizar el contraste

$$H_0 : \mu = 60$$
$$H_1 : \mu \neq 60$$

Utilizando el resultado obtenido en el apartado (b), ¿se rechaza la hipótesis nula con $\alpha = 0,05$? Razona la respuesta.

(d) Sin realizar ningún cálculo, ¿podríamos contestar al apartado (c) pero ahora para $\alpha = 0,01$? Razona la respuesta.

2. **(2 puntos).** Sea $X_1, ..., X_n$ una muestra aleatoria simple de tamaño n de una variable aleatoria con función de densidad:

$$f_\theta(x) = \frac{\theta}{(1+x)^{1+\theta}} \qquad x > 0, \;\; \theta > 0.$$

(a) Calcular el estimador de máxima verosimilitud de θ.

(b) Hallar la varianza asintótica del estimador de máxima verosimilitud de θ.

3. **(2,5 puntos).** Los dirigentes de una empresa agroalimentaria piensan que el éxito de venta de su producto en Andalucía es el mismo que el obtenido en la comunidad gallega. Para verificarlo realizaron una encuesta en Andalucia a 100 personas, de las que 49 mostraron la intención de compra del producto y a otras 200 personas, en Galicia, de las que 33 personas estuvieron interesadas en la compra del mismo producto.

(a) Construir un intervalo de confianza al 90% para la proporción de personas que comprarían el producto en Andalucía.

(b) Utilizando el resultado del apartado (a) y al 5% de nivel de significación, ¿se rechaza la hipótesis nula de que la proporción de personas que comprarían el producto en Andalucia es el 50%?.

(c) Construir un intervalo de confianza al 95% para la diferencia de proporciones en la intención de compra del producto en las dos Comunidades. ¿Podemos afirmar que los dirigentes de la empresa tienen razón?.

4. **(3 puntos).** El gerente de un hospital quiere reducir el tiempo de espera de los pacientes en el servicio de urgencias. Para ello, realiza un estudio del tipo y número de pacientes que llega a este servicio cada hora. Concluye que los pacientes se pueden clasificar en dos grandes grupos:

- Grupo L: pacientes con dolencias leves o moderadas.
- Grupo G: pacientes con dolencias graves.

También concluye que el número, X, de pacientes del grupo L que llega en una hora sigue una distribución $Poisson(\lambda_L)$ y que el número, Y, de pacientes del grupo G que llega en una hora sigue una distribución $Poisson(\lambda_G)$.

Posteriormente, toma 100 observaciones de X e Y en horas seleccionadas en diferentes días. Los estadísticos resumen de estas muestras son:

```
Summary Statistics

                    X                   Y
---------------------------------------------------
Count               100                 100
Average             10.27               5.0
Variance            10.6637             4.60606
Standard deviation  3.26554             2.14617
Minimum             5.0                 1.0
Maximum             21.0                10.0
Range               16.0                9.0
Stnd. skewness      2.8468              0.561879
Stnd. kurtosis      1.56713             -1.15913
---------------------------------------------------
```

(a) Calcular un intervalo de confianza al 95% para el parámetro λ_L.

(i) ¿Bajo qué hipótesis es válido el intervalo calculado?.

(ii) ¿Puede aceptarse la hipótesis de que la media del número de pacientes de tipo L es igual a 11 con un nivel de significación $\alpha = 0.1$?. Justificar la respuesta.

(b) Contrastar, con un nivel 0.05, la hipótesis (nula) de que el número medio de pacientes del grupo G es menor o igual que 4. Escribir claramente las hipótesis del contraste, el estadístico de contraste y la región de rechazo del mismo.

(c) El gerente cree que la suma de los pacientes de tipo L y de tipo G que llegan en una hora, $Z = X + Y$, sigue una distribución $Poisson(\lambda_Z)$. Completar la siguiente salida de `Statgraphics` para contrastar la hipótesis $H_0 : Z \sim Poisson$ a un nivel, $\alpha = 0.05$. En el caso del p-valor es suficiente con acotar su valor.

```
Goodness-of-Fit Tests for Z

                               Chi-Square Test
----------------------------------------------------------------
               Lower          Upper       Observed      Expected
               Limit          Limit      Frequency     Frequency
----------------------------------------------------------------
         at or below            9.5              4          6.15
                 9.5           11.5             17         10.60
                11.5           13.5             18         17.04
                13.5           15.5             13         20.26
                15.5           17.5             21         18.51
                17.5           19.5              9         13.40
                19.5           21.5             14          7.88
above           21.5                             4          6.16
----------------------------------------------------------------
Chi-Square = ? with ? d.f.   P-Value = ?
```

NOTA: Las notas de cada grupo serán publicadas por cada profesor el lunes 21 de febrero en Aula Global con las correspondientes instrucciones para solicitar revisión.

LEC/LADE/LEC-DER/LADE-DER/LEC-PERIODISMO
CURSO 2004/05
EXAMEN DE ESTADÍSTICA I
1 de Septiembre de 2005
TIEMPO MÁXIMO= 3 h

1. **(2,5 puntos).** Se toma una muestra del crecimiento de la oferta hotelera (número de camas), en términos porcentuales, de ocho países mediterráneos en el período 2000-2004. Se obtienen los siguientes datos:

$$4,6 \quad 0,1 \quad 3,4 \quad 9 \quad 7,3 \quad 0,8 \quad 1,6 \quad 4,6$$

(a) Especificando las hipótesis necesarias, determinar un intervalo de confianza al 90% para el valor medio y la desviación típica.

(b) Un informe anterior de Eurostat (Oficina de Estadística de la Unión Europea), preveía que el crecimiento medio en la zona no superaría el $3,5\%$ (tómese como H_0). Razonar brevemente si esta previsión fue acertada al 95% de nivel de confianza.

(c) Para el mismo periodo se tomó una muestra del crecimiento de la oferta hotelera en países del Caribe; los datos obtenidos fueron los siguientes:

$$3,1 \quad 2,8 \quad 3,9 \quad 7,1 \quad 6,2 \quad 4,3 \quad 1,3$$

Especificando las hipótesis necesarias, determinar si es aceptable la afirmación de que el crecimiento medio de la oferta hotelera es igual en ambas zonas turísticas al 99% de nivel de confianza, suponiendo que la varianza poblacional es la misma en ambas zonas.

(d) Acotar el p-valor del contraste realizado en (c).

2. **(2,5 puntos).** Se desea analizar la diferencia en el precio medio (en euros/m^2) de la vivienda de nueva construcción en las ciudades de Madrid y Barcelona. Suponga que el precio de vivienda en Madrid (X) sigue una distribución normal con media μ_x y varianza σ_x^2 y que el precio de la vivienda en Barcelona (Y) sigue también una distribución normal, en este caso, con media μ_y y varianza σ_y^2.

Para analizar la diferencia de medias $(\mu_x - \mu_y)$ se dispone de dos muestras aleatorias independientes de tamaño n_x y n_y, respectivamente. Se pide:

(a) Obtener razonadamente la media, la varianza y la distribución de $\overline{X} - \overline{Y}$.

(b) ¿Cuál es la probabilidad de que, dadas las correspondientes muestras, la diferencia de medias muestrales sea mayor que la diferencia $\mu_x - \mu_y$? Justifique los pasos que dé.

3. **(2,5 puntos).** La nota final en la asignatura de Estadística 1 es una variable aleatoria cuya distribución es Normal con desviación típica poblacional igual a 3. Para estimar la media poblacional se toma una muestra aleatoria simple de tamaño 5 y se proponen los siguientes estimadores:

$$\hat{\mu}_1 = \frac{X_1 - 2X_2 + 5X_4}{3}$$

$$\hat{\mu}_2 = X_2 - 2X_3 + 2X_5$$

(a) ¿Cuál de los dos estimadores propuestos tiene menor error cuadrático medio?

(b) Calcular el estimador de máxima verosimilitud de la media poblacional.

(c) Calcular el error cuadrático medio del estimador obtenido en (b).

Nota: la función de densidad de una variable aleatoria $N(\mu, \sigma^2)$ es la siguiente:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} \qquad -\infty < x < \infty$$

4. **(2,5 puntos).** El valor de cotización de ciertas acciones es una variable aleatoria con función de densidad

$$f(x) = \theta\left(\frac{1}{x}\right)^{\theta-1} \qquad \text{para } x > 1 \text{ y } \theta > 1.$$

(a) Obtener el estimador de máxima verosimilitud de θ y su varianza asintótica.

(b) Se toma una muestra de tamaño 100 de los valores de cotización de estas acciones y se obtiene que, la media de los logaritmos de las cotizaciones es 0,4. Calcular un intervalo de confianza para θ, con $\alpha = 0,05$.

NOTA: Las notas de cada grupo serán publicadas por cada profesor el 5 de septiembre, lunes, a última hora del día en Aula Global con las correspondientes instrucciones para solicitar revisión.

LEC/LADE/LEC-DER/LADE-DER/LEC-PERIODISMO
CURSO 2005/06
EXAMEN DE ESTADÍSTICA I
9 de Febrero de 2006
TIEMPO MÁXIMO = 3 h

1. **(2,5 puntos).** La entrada en vigor de la Ley 28/2005 por la que se prohíbe fumar en lugares públicos ha suscitado opiniones muy encontradas. Con la aplicación de dicha Ley, un porcentaje de fumadores ha decidido dejar el tabaco y cierto organismo establece en el 20% este porcentaje.

   (a) Deducir la distribución de la proporción muestral de fumadores cuya intención es dejar el tabaco (en caso de muestras grandes).

   (b) ¿Es insesgado el estimador proporción muestral de fumadores con intención de dejar el tabaco? Razonar la respuesta.

   (c) El Gobierno asegura que el porcentaje de fumadores que van a decidir dejar de fumar con la aplicación de esta Ley, va a ser al menos el 28%. Para contrastar dicha hipótesis, se encuesta a 800 fumadores de los que 150 manifiestan su intención de dejar de fumar desde la entrada en vigor de la Ley. ¿Hay evidencia estadística suficiente para rechazar la hipótesis del Gobierno?

2. **(2,5 puntos).** Sea $X_1, ..., X_n$ una muestra aleatoria simple de una distribución Normal con media μ y desviación típica σ. Definimos la varianza muestral $\hat{\sigma}^2$ y la varianza muestral corregida (o cuasivarianza) s^2 como:

$$\hat{\sigma}^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n} \qquad\qquad s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$$

   donde $\bar{x}$ es la media muestral. Si sabemos que una distribución χ^2_n tiene su media igual a los grados de libertad ($E(\chi^2_n) = n$) y su varianza igual al doble que la media ($Var(\chi^2_n) = 2n$),

   (a) Demostrar que $\hat{\sigma}^2$ es un estimador sesgado de la varianza poblacional σ^2. Obtener su sesgo.

   (b) Demostrar que s^2 es un estimador insesgado de la varianza poblacional σ^2.

   (c) Obtener el error cuadrático medio de $\hat{\sigma}^2$.

   (d) Obtener el error cuadrático medio de s^2.

   (e) Basándonos en el error cuadrático medio, ¿qué estimador es preferible?

3. **(2,5 puntos).** Un fabricante de pasta indica en el envoltorio que la cantidad de pasta que contiene cada paquete es de 500 gr. Se cree que la desviación típica del peso de los paquetes es de 3 gr. Se toma una muestra de 20 paquetes y se obtiene un peso medio de 498 gr con una desviación típica corregida de 4 gr. Especificando las hipótesis necesarias, responder:

   (a) ¿Hay evidencia de que por término medio los paquetes contienen menos pasta de lo indicado? Según tu conclusión, ¿qué tipo de error podrías estar cometiendo y con qué probabilidad?

   (b) ¿Es la desviación típica mayor de lo que se cree? Utiliza un nivel de significación del 5%.

   (c) Calcular la potencia del contraste de la media del apartado (a) para un nivel de significación del 5% y un valor alternativo de la media de 497.

4. **(2.5 puntos).** Se desea evaluar la efectividad de una nueva vacuna antigripal. Para ello se suministra de manera voluntaria y gratuita, en una pequeña comunidad. La vacuna se administra en dos dosis, separadas por un período de dos semanas, de forma que algunas personas han recibido una sola dosis, otras han recibido las dos, y otras personas no han recibido ninguna. La siguiente tabla indica los resultados que se registraron durante la siguiente primavera en 1000 habitantes de la comunidad elegidos al azar.

	No vacunados (0 dosis)	Vacunados (1 dosis)	Vacunados (2 dosis)
Gripe	24	9	13
No Gripe	289	100	565

(a) ¿Proporcionan estos datos suficiente evidencia estadística (al nivel 0.05) para indicar dependencia entre el número de dosis recibidas y la protección frente a la gripe?

(b) Si consideramos vacunados a los que han recibido una ó dos dosis, ¿hay evidencia estadística para afirmar (al nivel 0.05) que la vacuna es efectiva frente a la gripe?

En ambos apartados escriba claramente las hipótesis que se contrastan, la región de rechazo y la conclusión.

NOTA: Las notas de cada grupo serán publicadas por cada profesor a última hora del miércoles 15 de febrero en Aula Global con las correspondientes instrucciones para solicitar revisión.

LEC/LADE/LEC-DER/LADE-DER/LEC-PERIODISMO
CURSO 2005/06
EXAMEN DE ESTADÍSTICA I
1 de Septiembre de 2006
TIEMPO MÁXIMO = 3 h

1. **(2,5 puntos).** El número X de defectos que presenta determinado tipo de piezas de una producción industrial sigue una distribución de Poisson con parámetro θ desconocido. Se toma una muestra aleatoria de 40 piezas y se registra el número de defectos que presenta cada pieza, obteniendo la siguiente información:

Número de defectos	0	1	2	3	4	5	6
Cantidad de piezas	4	10	11	8	4	2	1

(a) Determinar el estimador de máxima verosimilitud (EMV) de θ.

(b) Calcular la estimación máximo verosímil con los datos de la tabla.

(c) Calcular un intervalo de confianza asintótico para θ en la muestra obtenida, tomando $\alpha = 0.05$.

Nota: La función de masa de una distribución de Poisson de parámetro θ es:

$$P(X = x) = \frac{e^{-\theta}\theta^x}{x!}, \qquad x = 0, 1, 2, \ldots$$

2. **(2,5 puntos).** Se recogen medidas de polución atmosférica en 10 sitios de una ciudad:

3.3; 1.7; 3.7; 4.6; 2.3; 3.9; 4.3; 1.4; 1.6; 3.6

(a) Hallar un intervalo de confianza al 95% para la varianza poblacional, mencionando las hipótesis estadísticas que hay que asumir.

(b) La polución atmosférica media de estos 10 sitios medida un año antes fue 2.27. Contrastar al 5% la hipótesis de que el nivel medio de polución no ha variado. Especificar claramente las hipótesis nula y alternativa del contraste, así como las hipótesis estadísticas que hay que asumir para realizarlo.

(c) Una organización ecologista sospecha que la polución atmosférica media está aumentando cada año. Contrastar al 5% la hipótesis de que el nivel medio de polución no ha aumentado (tómese como H_0). Especificar claramente las hipótesis nula y alternativa del contraste, así como las hipótesis estadísticas que hay que asumir para realizarlo. Comentar el resultado de los contrastes realizados en (b) y (c).

3. **(2,5 puntos).** Los defensores de un nuevo molino de viento afirman que la potencia diaria que puede generar es una variable aleatoria (X) cuya media es al menos 800 Kw. Se asume que X es normal con desviación típica 120 Kw. Para realizar un contraste de la hipótesis mantenida por los defensores del nuevo molino, se toma una muestra aleatoria simple de tamaño 100 y se obtiene una media muestral igual a 776 Kw.

   (a) Definir y calcular el p-valor del contraste.

   (b) Definir y calcular la potencia del contraste tomando como hipótesis alternativa $\mu = 770$ y nivel de significación $\alpha = 0.05$.

4. **(2.5 puntos).** Se tomaron cinco muestras de plantas en cinco lugares distintos, A, B, C, D y E observando en cada una de ellas, el número de plantas infectadas con un cierto parásito. Las dos primeras muestras A y B se obtuvieron en una zona más húmeda que las tres restantes C, D y E. Los resultados fueron los siguientes:

Muestras	Infectados	No Infectados	Total
A	80	184	264
B	58	147	205
C	114	276	390
D	55	210	265
E	79	296	375
Total	386	1113	1499

(a) Contrastar la hipótesis (al nivel 0.05) de que la infección de parásitos es independiente de la zona donde se extrajeron las muestras.

(b) Consideramos ahora las muestras procedentes sólo de dos zonas: húmedas y secas. ¿Hay evidencia suficiente para afirmar que la infección es distinta en ambas zonas?

En ambos casos ((a) y (b)) especificar las hipótesis contrastadas, la región de rechazo y la conclusión.

NOTA: Las notas de cada grupo serán publicadas por cada profesor el en Aula Global con las correspondientes instrucciones para solicitar revisión.

LEC/LADE/LEC-DER/LADE-DER/LEC-PERIODISMO
CURSO 2006/07
EXAMEN DE ESTADÍSTICA I
15 de Febrero de 2007
TIEMPO MÁXIMO = 3 h

1. **(3 puntos)** Un grupo de científicos está estudiando los diferentes hábitos de alimentación entre dos colectivos de Madrid y Tenerife . Sospechan que estas diferencias pueden influir de forma significativa en el peso de los recién nacidos. Por ello se han tomado dos muestras de pesos de recién nacidos, una en Madrid y otra en Tenerife, obteniéndose los siguientes resultados:

   a) Teniendo en cuenta estas dos muestras, y suponiendo normalidad en las poblaciones, determinar si existe suficiente evidencia al 5% de significación para afirmar que las varianzas poblacionales son distintas. Responder en base a la siguiente información que ofrece Statgraphics:

   b) ¿Hay evidencia estadística con un 95% de confianza, de que la media poblacional de los pesos de recién nacidos madrileños es mayor que la media poblacional de los pesos de los recién nacidos tinerfeños? Acotar el p-valor del contraste.

   c) Sin hacer ningún cálculo, ¿cómo cambiaría la respuesta del apartado anterior si hubiéramos utilizado un nivel de confianza de un 90%? ¿Y si hubiera sido de un 99%?

   d) Explicar el significado estadístico de nivel de confianza $(1-\alpha)100\%$.

2. **(2 puntos)** En un servicio de urgencias resulta crucial el tiempo de espera hasta que un ciudadano es atendido. Se sospecha que la variable "Tiempo de espera" puede ajustarse mediante una distribución normal. Por ello, se toma una muestra aleatoria simple de 50 datos y se obtiene el siguiente resultado al realizar un test Chi-cuadrado de bondad de ajuste:

   a) Completar los datos que faltan en la tabla anterior.

   b) Plantear el contraste de hipótesis y acotar el p-valor. ¿Hay evidencia estadística para rechazar la hipótesis de normalidad?

c) Para reforzar la decisión se realiza también el contraste de Kolmogorov-Smirnov obteniéndose $D_n = 0.0937677$. ¿Qué decisión tomaría con un nivel de significación del 5%? Usar la siguiente tabla:

3. **(3 puntos)** El encargado de una sala de cine está interesado en conocer la edad de su clientela. Para ello, decide preguntar las edades a 10 clientes seleccionados al azar. Suponiendo que la muestra es aleatoria simple y que la edad de los clientes (expresada en años) es una variable aleatoria, X, Normal, con media $\mu = 22$ y desviación típica $\sigma = 3$, se pide:

    a) Deducir las expresiones del valor esperado y la varianza de la media de las edades de los 10 clientes de la muestra, es decir, calcular $E(\bar{X})$ y $V(\bar{X})$.

    b) Calcular valor esperado de la varianza ($\hat{\sigma}^2$) y la cuasivarianza (o varianza corregida) (s^2) de las edades de los clientes de la muestra, es decir, calcular $E(\hat{\sigma}^2)$ y $E(s^2)$. ¿Es posible obtener el mismo valor?

    c) Supongamos que el encargado decide aumentar el tamaño de la muestra a 100 personas. Determinar,

        (i) La distribución de probabilidad de la media de las edades de los 100 clientes, razonando la respuesta. ¿Cuál es la probabilidad de que dicha media sea inferior a 25 años?

        (ii) La probabilidad de que la edad de un determinado cliente no llegue a 25 años.

        (iii) La probabilidad de que la cuasidesviación típica (o desviación típica corregida), s, supere los 4 años.

4. **(2 puntos)** En una determinada encuesta realizada en Getafe en 2006, se preguntó a 400 personas elegidas al azar si tomaban fruta al menos 2 veces a la semana. El 43% de los encuestados respondió afirmativamente.

    a) Construir un intervalo de confianza al 95% para p, proporción de ciudadanos de Getafe que toman fruta al menos 2 veces a la semana, especificando claramente las hipótesis necesarias.

    b) Si alguien afirmara que hay una probabilidad de 0.9 de que p esté en el intervalo calculado en a), ¿crees que es correcto? Razona tu respuesta.

    c) Se quiere reducir la longitud del intervalo calculado en a) a la mitad. Dada la información del enunciado, ¿cuál será el tamaño muestral necesario?

NOTA: Las notas de cada grupo serán publicadas por cada profesor a última hora del miércoles 21 de febrero en Aula Global con las correspondientes instrucciones para solicitar revisión.

LEC/LADE/LEC-DER/LADE-DER/LEC-PERIODISMO
CURSO 2006/07
EXAMEN DE ESTADÍSTICA I
22 de Septiembre de 2007
TIEMPO MÁXIMO = 3 h

1. **(2.5 puntos)** Se sabe que la función de densidad de una población es:

$$f(x) = \frac{\theta k^{\theta}}{x^{1+\theta}} \qquad x > k, \theta > 0$$

donde k es un valor fijo. Habiendo extraído una muestra de tamaño n, se pide:

a) Obtener el estimador de máxima verosimilitud de θ.

b) Hallar la varianza asintótica del estimador de máxima verosimilitud de θ.

c) Suponiendo que $k = exp(1) = 2.7183$, calcular el valor del estimador de máxima verosimilitud de θ para una muestra que toma los valores de los números enteros hasta 10 inclusive.

2. **(2.5 puntos)** La siguiente tabla muestra los resultados de una encuesta realizada a 300 personas sobre el periódico que suelen leer y su condición social:

Condición Social \ Periódico	A	B	C
Baja	31	11	12
Media	49	59	51
Alta	16	24	29
Muy Alta	4	6	8

a) ¿Proporcionan los datos suficiente evidencia estadística para indicar dependencia entre el periódico que se lee y la condición social?

b) Para la Condición Social Media, ¿se puede concluir que los tres periódicos se leen con igual probabilidad?

En los dos apartados indicar claramente la hipótesis a contrastar, la región de rechazo con nivel de significación 0.05, los grados de libertad y la conclusión.

3. **(2.5 puntos)** Se sabe que la variable aleatoria X = edad en la audiencia de un programa de TV, sigue una distribución normal de media $\mu = 20$ y desviación típica $\sigma = 3$. Se toma una muestra aleatoria simple de tamaño 100 de personas que ven el programa y se pide:

   a) Probabilidad de que la media de edad de los encuestados supere los 25 años. ¿Cuál es la probabilidad de que la media de los encuestados sea exactamente igual a 21 años?

   b) Se sabe que el 10% de las personas que ven el programa son de Madrid. Calcular la probabilidad de que haya menos de un 8% de madrileños en la muestra escogida.

   c) En el supuesto de que la media de la distribución fuese desconocida, se consi dera el estimador $\hat{\mu} = \frac{X_1+X_{100}}{2}$ para la media. Calcular la eficiencia relativa de la media muestral, $\bar{X}$, respecto del estimador propuesto, $\hat{\mu}$. ¿Qué conclusión se puede obtener?

4. **(2.5 puntos)** En cierto proceso industrial se considera la variable aleatoria X = coste unitario de producción (en Euros). Se sabe que esta variable aleatoria tiene una distribución normal de media μ y desviación típica $\sigma = 20$. Se toma una muestra aletoria de 25 observaciones obteniéndose una media muestral $\bar{x} = 120$.

   a) A partir de esta información muestral, calcular un intervalo de confianza al 95% para el coste unitario medio de producción.

   b) ¿Qué tamaño muestral será necesario tomar para que un intervalo, del 95% de confianza, garantice una precisión en la estimación, dada por una amplitud máxima de 10 Euros?. (Amplitud del intervalo = diferencia entre sus extremos).

   c) Ante la sospecha de que la varianza de X está aumentando se calcula la cuasi-varianza (o varianza corregida) muestral, s^2, de la muestra anterior y se obtiene un valor igual a 512.12. ¿Existe evidencia muestral al 99% para rechazar la hipótesis de que $\sigma = 20$?

NOTA: Las notas de cada grupo serán publicadas por cada profesor a última hora del sábado 29 de septiembre en Aula Global con las correspondientes instrucciones para solicitar revisión.

Estadística I
Examen Final - 28 Mayo de 2009
Tiempo: 2.5h - Total: 40 puntos

Nombre:..
Grupo:.................

Realizar los cálculos intermedios con 4 decimales y redondear el resultado final a 2 decimales.

1. (10 puntos) A 16 estudiantes de *Filosofía* se les preguntó cuántas clases de esta asignatura habían perdido durante el cuatrimestre. Las respuestas obtenidas fueron las siguientes:

$$3 \quad 1 \quad 6 \quad 1 \quad 9 \quad 1 \quad 7 \quad 2 \quad 3 \quad 0 \quad 4 \quad 4 \quad 5 \quad 0 \quad 2 \quad 0$$

   (a) (3 puntos) Calcula la distribución de frecuencias (absolutas y relativas) de los datos anteriores, completando la siguiente tabla.

Intervalo	Frecuencia absoluta	Frecuencia relativa
$[0, 2)$		
$[2, 4)$		
$[4, 6)$		
$[6, 8)$		
$[8, 10)$		

   (b) (2 puntos) Observando los resultados obtenidos en el apartado (a), ¿cuántos estudiantes han perdido al menos seis clases? ¿Qué porcentaje de estudiantes han perdido menos de cuatro clases?

   (c) (2.5 puntos) Utilizando los resultados del apartado (a), dibuja el histograma de los datos (usando frecuencias absolutas) y describe la forma que tiene la distribución (utiliza la rejilla de la cara posterior de esta página).

   (d) (1.5 puntos) Calcula la media muestral, la mediana y la moda.

   (e) (1 punto) ¿Sorprende que la media sea mayor que la mediana? Justifica brevemente tu respuesta.

2. (10 puntos) Una compañía decide realizar un test de aptitud a los nuevos representantes de ventas. La siguiente tabla muestra el promedio de ventas semanales, Y, y la puntuación del test de aptitud (en puntos), X, para una muestra de $n = 8$ representantes.

Promedio de ventas semanales (Y)	10	12	28	24	18	16	15	12
Puntuación (X)	55	60	85	75	80	85	65	60

(a) (4 puntos) Dibuja el diagrama de dispersión para el promedio de ventas semanales respecto de la puntuación (utiliza la rejilla de la cara posterior de esta página).

(b) (2 puntos) Se calculó la recta de regresión para el promedio de ventas mensual frente a la puntuación, usando R. Usa la salida obtenida para identificar la ordenada en el origen y la pendiente de la recta ajustada, y escribe la ecuación de la misma.

```
Call:
lm(formula = Y ~ X, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-6.6514 -1.1147 -0.6009  1.6261  5.3670

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -11.5046     9.5745  -1.202   0.2748
X             0.4018     0.1339   3.002   0.0240 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.279 on 6 degrees of freedom
Multiple R-squared: 0.6003,    Adjusted R-squared: 0.5337
F-statistic: 9.011 on 1 and 6 DF,    p-value: 0.02395
```

(c) (2 puntos) Dibuja la recta de regresión ajustada, en el diagrama de dispersión realizado en el apartado (a).

(d) (2 puntos) ¿Cuál es el valor predicho del promedio de ventas mensuales, para un representante que ha recibido 62 puntos en el test de aptitud?

3. (10 puntos) Debido a la alarma originada por la gripe de la cepa *N1H1*, el 65% de las llamadas recibidas por el teléfono nacional de emergencias están relacionadas con dudas acerca de dicha enfermedad. En un grupo de ocho llamadas elegidas al azar, halla la probabilidad de que estén relacionadas con la nueva gripe:

   (a) (2 puntos) Alguna de las ocho llamadas.

   (b) (4 puntos) Más de seis llamadas entre las ocho llamadas.

   Sea X la variable aleatoria que modeliza el número de llamadas relacionadas con la gripe, cuando se considera un total de 8 llamadas.

   (c) (2 puntos) Calcula la esperanza de X.

   (d) (2 puntos) Calcula la desviación típica de X.

4. (10 puntos) Se sabe que la demanda diaria de un producto fabricado por cierta empresa sigue una distribución normal. Se toma una muestra aleatoria simple de la demanda registrada en 15 días distintos, obteniéndose una media muestral igual a 3.56 y una desviación típica muestral (también llamada cuasidesviación típica) igual a 0.92.

   (a) (5 puntos) Construye un intervalo de confianza al 95% para la media de la demanda diaria de dicho producto, especificando las hipótesis necesarias.

   (b) (5 puntos) El plan de producción de la empresa está basado en la suposición de que la demanda media diaria del citado producto es igual a 4. Contrasta esta hipótesis, al nivel de significación de 0.05, e indica claramente la conclusión obtenida. Indica quiénes son las hipótesis nula y alternativa.

Estadística I
Examen Final - 19 de junio de 2009

Nombre:...
Grupo:.................

Realizar los cálculos intermedios con 4 decimales y redondear el resultado final a 2 decimales.

1. La siguiente tabla muestra las distribuciones de frecuencias absolutas de la variable `altura` (en metros) de $n = 500$ estudiantes de un centro educativo.

Intervalo	Frecuencias absolutas
$[1.59, 1.61)$	11
$[1.61, 1.63)$	71
$[1.63, 1.65)$	159
$[1.65, 1.67)$	177
$[1.67, 1.69)$	67
$[1.69, 1.71)$	15

(a) (4.5 puntos) Dibuja el histograma de la variable, utilizando la información de la tabla anterior, y describe la forma de la distribución (usa la rejilla de la cara posterior de esta página).

(b) (1.2 puntos) Calcula las frecuencias relativas.

(c) (2 puntos) ¿Cuántos estudiantes miden más de 1.67m, inclusive?
¿Qué porcentaje de estudiantes miden menos de 1.63m?

d) (2.3 puntos) Se analizó con `R` la variable `altura`, obteniéndose los siguientes resultados,

```
  mean     sd     0%    25%    50%    75%   100%   n
1.6505 0.0202 1.5990 1.6368 1.6508 1.6633 1.7018 500
```

Usa los resultados anteriores para identificar, o calcular, las siguientes medidas: media, mediana, desviación estándar, varianza, valor mínimo, valor máximo, rango, primer cuartil, tercer cuartil, rango intercuartílico y coeficiente de variación (completa en la cara posterior de esta página).

media = ...

mediana = ...

desviación estándar = ...

varianza = ...

mínimo = ...

máximo = ...

rango = ...

primer cuartil = ...

tercer cuartil = ...

rango intercuartílico = ...

coeficiente de variación = ...

2. Un mayorista de máquinas fotocopiadoras quiere realizar un análisis sobre la posible relación entre el precio de venta, X, y el número de ventas, Y, utilizando la información obtenida de ocho proveedores de dicha marca.

Precio (X)	550	600	640	600	500	650	450	525
Ventas (Y)	42	39	35	40	44	38	45	41

A continuación, se muestra la salida de `R` con el modelo de regresión ajustado para estas dos variables.

```
Call: lm(formula = Y ~ X, data = Datos)

Residuals:
    Min      1Q  Median      3Q     Max
-2.2571 -0.6006  0.3836  0.9196  1.1717

Coefficients:
             Estimate    Std.Error    t value Pr(>|t|)
(Intercept)  64.70135    4.02861      16.060  3.7e-06 ***
X            -0.04288    0.00709      -6.048 0.000925 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 1.318 on 6 degrees of freedom.
 Multiple R-squared: 0.8591, Adjusted R-squared: 0.8356
F-statistic: 36.58 on 1 and 6 DF,  p-value: 0.000925
```

(a) Dibuja el diagrama de dispersión correspondiente a estas variables (utiliza la rejilla de la cara posterior de esta página).

(b) Calcula el coeficiente de correlación, explicando de manera detallada dicho cálculo. ¿Existe relación entre el precio y el número de ventas de máquinas fotocopiadoras? Indica el tipo de relación y justifica la respuesta.

(c) Indica y justifica cuál es la ecuación de la recta de regresión de Y sobre X. Dibújala sobre el diagrama de dispersión anterior (utiliza la rejilla de la cara posterior de esta página).

3. Una compañía especializada en instalación de calderas estima que las probabilidades de instalar X calderas al mes son las siguientes:

X	0	1	2	3	4	5
probabilidad	0.10	0.13	0.25	0.29	0.16	0.07

   (a) Calcula la probabilidad de que tengan que instalar al menos 3 calderas en un mes.

   (b) Si en un mes ya se han instalado al menos dos calderas, calcula la probabilidad de que, como mucho, se instalen 4 calderas.

   (c) Si se eligen 6 meses al azar, ¿cuál es la probabilidad de que en al menos 2 de esos meses se hayan realizado al menos 3 instalaciones?

4. En un centro telefónico de atención al cliente se quiere estimar el número medio de llamadas que llegan en una hora. Para ello, mediante muestreo aleatorio simple, se cuenta el número de llamadas recibidas en 50 periodos de una hora de duración, obteniéndose una media muestral igual a 35 y una desviación típica muestral (también llamada cuasidesviación típica) igual a 6.

   (a) Construye un intervalo de confianza al 90% para el número medio de llamadas por hora, especificando las hipótesis necesarias para construirlo.

   (b) La dirección del centro estima que el número medio de llamadas que llegan por hora es 30. Sin embargo, en los últimos meses los trabajadores se quejan de un aumento considerable en el número de llamadas que atienden. Plantea y resuelve un contraste de hipótesis al nivel 0.05 para determinar si los trabajadores tienen razón y justifica la conclusión obtenida.

Examen Final de Estadística I, 26 de mayo de 2010.

Grados ADE, ADE-DER, ADE-INF, ECO, ECO-DER.

NORMAS: 1) Entregar cada problema en un cuadernillo distinto, aunque esté en blanco.
2) Realizar los cálculos con al menos dos cifras decimales significativas.
3) No se podrá abandonar el examen hasta transcurridos 30 minutos depués de haber empezado.
4) No está permitido salir del aula sin entregar el examen, aunque esté en blanco.

Problema 1 (10 puntos) Se realiza un experimento mediante el cual personas que han sido condenadas por pequeñas faltas deben realizar pagos mensuales a las víctimas según el criterio de un juez. La siguiente tabla contiene el salario mensual (x) de 10 condenados y los pagos (en euros) realizados a las víctimas (y):

Salarios	300	880	1000	1540	1560	1600	1600	2200	3200	6000
Pagos	200	380	400	200	800	600	800	1000	1600	2700

Se dispone además de la siguiente información: $\overline{x} = 1988$, $\overline{y} = 868$, $s_x^2 = 2594240$, $s_y^2 = 593351.1$ y $s_{xy} = 1206062$.

a) (2 puntos) Representa el diagrama de dispersión (manteniendo la escala) que muestre los pagos como función de los salarios. Calcula e interpreta el coeficiente de correlación lineal de Pearson. Razona si crees que, aparentemente, existe relación lineal entre ambas variables.

b) (2 puntos) Calcula la recta de regresión de y sobre x. Determina cuál es la pendiente de la recta e interpreta su valor. Representa la recta de regresión sobre el gráfico del apartado a) explicando detalladamente cómo hacer esta representación.

c) (2 puntos) Calcula el coeficiente de determinación y el porcentaje de variabilidad no explicado por el modelo de la recta de regresión. Justifica si consideras adecuado este modelo.

d) (2 puntos) Si el salario de una persona culpable de cierto delito es de 1400 euros, ¿cuál será el pago en este caso según la recta de regresión? Justifica si es razonable realizar dicha predicción.

e) (2 puntos) Representa un diagrama de caja para los datos de la variable *Salarios* y determina si existen datos atípicos. Justifica la respuesta.

Problema 2 (10 puntos) Los autobuses urbanos de cierta ciudad tardan en realizar sus trayectos un tiempo X que puede modelizarse según una distribución uniforme en el intervalo $(30, 40)$.

a) (1 punto) Dibuja la función de densidad de la variable aleatoria X. Especifica en los ejes de coordenadas tanto los valores de X con densidad positiva, como los valores que toma la función de densidad.

b) (2 puntos) ¿Cuál es la probabilidad de que un autobús tarde entre 30 y 37.5 minutos en realizar su trayecto?

c) (3 puntos) Se seleccionan aleatoriamente 100 autobuses urbanos de esta ciudad y se cuenta cuántos de ellos tardarán entre 30 y 37.5 minutos en realizar sus trayectos. Especifica el nombre de la distribución de esta nueva variable aleatoria, que llamaremos Y, e indica los valores de sus parámetros. Calcula la esperanza y desviación típica de Y.

d) (4 puntos) ¿Cuál es la probabilidad (aproximada) de que haya menos de 64 autobuses que tarden entre 30 y 37.5 minutos en realizar sus trayectos?

Problema 3 (10 puntos) Según un estudio sobre los hábitos de consumo de las familias de la Comunidad de Madrid, el gasto mensual en productos de marcas blancas de las familias puede modelizarse según una variable aleatoria X con ley Normal de desviación típica 50 euros. Los datos siguientes corresponden a una muestra aleatoria simple de tamaño $n = 10$ familias:

Familia	1	2	3	4	5	6	7	8	9	10
Gasto	325	193	203	305	251	240	233	238	354	311

a) (2.5 puntos) Define qué es un estimador insesgado (o centrado) y propón uno para el gasto medio mensual en productos de marcas blancas de las familias de la Comunidad de Madrid, razonando tu respuesta. ¿Qué estimación del gasto medio mensual se obtiene para el estimador propuesto?

b) (2.5 puntos) Compara el estimador propuesto en el apartado a) con el siguiente:

$$T = \frac{X_1 + X_n}{2},$$

donde $X_1, X_2, \ldots, X_n$ es una muestra aleatoria simple de X. Justifica con alguno de los criterios que conoces con cuál de ellos te quedarías. ¿Qué estimación se obtiene para el estimador T?

c) (2.5 puntos) Calcula un intervalo de confianza al 95% para el gasto medio mensual en productos de marcas blancas de las familias de la Comunidad de Madrid, especificando las hipótesis necesarias para calcularlo.

d) (2.5 puntos) Si queremos trabajar con niveles de confianza del 90%, ¿obtendríamos un intervalo más o menos amplio que el calculado anteriormente? Razona tu respuesta sin calcular el nuevo intervalo de confianza.

Examen Final de Estadística I, 23 de junio de 2010.
Grados ADE, ADE-DER, ADE-INF, ECO, ECO-DER.

NORMAS: 1) Entregar cada problema en un cuadernillo distinto, aunque esté en blanco.
2) Realizar los cálculos con al menos dos cifras decimales significativas.
3) No se podrá abandonar el examen hasta transcurridos 30 minutos depués de haber empezado.
4) Una vez empezado el examen, no estará permitido salir del aula.

Problema 1 (10 puntos) La siguiente tabla contiene información sobre la edad y la antigüedad, ambas medidas en años, de una plantilla de 75 trabajadores de una empresa (atención: se han dejado en blanco las casillas que contienen un cero):

	Antigüedad										
Edad	1	2	3	4	5	7	9	10	12	15	20
[20,30)	3	6		4							
[30,40)		1	4	2	3						
[40,50)				1	4	13	2	3	1		
[50,60)							2	8	7	4	
[60,70]										2	5

a) (1 punto) Indica el tipo de las variables **Edad** y **Antigüedad**.

b) (1.5 puntos) Especifica la distribución marginal de la variable **Antigüedad**. ¿Cual es la antigüedad media para los trabajadores de esta empresa?

c) (2 puntos) Compara la dispersión de las variables **Edad** y **Antigüedad**, ayudándote de la información siguiente:

```
numSummary(Datos[,c(Edad)], statistics=c("mean","sd","quantiles"), quantiles=c(0,.25,.5,.75,1))
        mean        sd      0%   25%  50%  75%   100%   n
Edad  44.891892   11.730498   20   38   46   53.0   67   74
```

d) (1.5 puntos) Especifica la distribución de la variable **Antigüedad** para los trabajadores con edades comprendidas entre 40 y 50 años. ¿Qué puedes decir acerca de la asimetría de esta distribución?

e) (2 puntos) A partir de la información siguiente, ¿qué puedes decir sobre la relación entre ambas variables? Explica en qué te basas.

```
Call:lm(formula=ANTIG\"UEDAD ~ EDAD, data=Datos)

Residuals:
  Min       1Q       Median     3Q      Max
-4.8303   -1.5464   -0.3001   1.4345   6.0651

Coefficients:
             Estimate   Std. Error  t value   Pr(>|t|)
(Intercept)   -8.45427    1.12128     -7.54    1.10e-10 ***
edad           0.37315    0.02418     15.44     < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.423 on 72 degrees of freedom
Multiple R-squared: 0.8679, Adjusted R-squared: 0.7647
F-statistic: 238.2 on 1 and 72 DF,  p-value: < 2.2e-16
```

f) (2 puntos) Utilizando la información del apartdado e), encuentra el modelo lineal que relaciona la variable antigüedad en función de la edad. Interpreta los coeficientes y usa este modelo para predecir cuanto tiempo durará en la empresa una persona que tiene 35 años. Justifica si crees razonable realizar dicha predicción.

Problema 2 (10 puntos) En un departamento de una empresa aseguradora se procesan comunicaciones de clientes de seguros de hogar. Dichas comunicaciones están relacionadas con incidencias de dos tipos (robos, tipo R y daños, tipo D). En promedio en un mes se reciben 86 comunicaciones de los dos tipos, de las que 24 incluyen incidencias de tipo R y 68 incidencias de tipo D. Una vez estudiadas, se aprueba compensar (C) un 65% de las reclamaciones de tipo R y un 80% de las reclamaciones de tipo D.
Partiendo de los datos anteriores, se pide que contestes las siguientes preguntas:

a) (3 puntos) ¿Cuál es la probabilidad de que una reclamación incluya incidencias de los dos tipos, R y D?

b) (3 puntos) ¿Cuál es la probabilidad de que se apruebe compensar una comunicación cualquiera?

c) (4 puntos) Si te envían para revisar una comunicación para la que se ha aprobado compensación (C), ¿cuál es la probabilidad de que sea de tipo D?

Problema 3 (10 puntos) Según un estudio realizado en cierta entidad bancaria durante el primer trimestre de 2010, en cada una de sus sucursales el número de cheques sin fondo que se reciben sigue una distribución de Poisson, con una media de 10 cheques sin fondo al día. Se elige una muestra de 200 sucursales mediante muestreo aleatorio simple y se registra el número de cheques sin fondo recibidos en cada una de ellas en un determinado día.

a) (4 puntos) ¿Cuál es la probabilidad de que el promedio de cheques sin fondo recibidos en las 200 sucursales sea mayor que 12?

b) (3 puntos) ¿Cuál es la probabilidad de que en una de las sucursales el número de cheques sin fondo recibidos un determinado día sea menor que 3?

c) (3 puntos) Suponiendo ahora que desconocemos el número medio de cheques sin fondo que llegan al día a una sucursal, es decir, sólo sabemos que la distribución del número de cheques sin fondo por día y por sucursal es Poisson de parámetro λ, propón un estimador insesgado para λ, basándote en una muestra aleatoria simple de tamaño 200 y razonando la respuesta.