ESTADÍSTICA EMPRESARIAL - Segundo Curso (L.A.D.E.) Curso 2.005-06
Convocatoria de Junio. 14-6-06

INSTRUCCIONES

1. El examen consta de cuestiones, que se responden sobre la hoja de codificación proporcionada, y problemas, que se responden en papel aparte.
2. Para escoger una respuesta, basta efectuar una marca **rellenando debidamente el rectángulo sobre el que está la letra escogida** en la hoja de codificación. Piénsalo antes; aunque puedes borrar si escribes con lápiz (número 2 o similar), marcas que no estén perfectamente borradas pueden ser leídas. Te aconsejamos que señales sobre el formulario de examen las respuestas que te parezcan adecuadas, y emplees los últimos diez minutos del tiempo asignado en transcribirlas a la hoja de codificación.
3. Hay siempre, en las preguntas de elección múltiple, una **única** respuesta correcta. Todas las cuestiones correctamente resueltas valen 1 punto mientras que las fallidas suponen una penalización de 0.2 puntos. Las preguntas no contestadas no suponen penalización.
4. Cada uno de los problemas, A, B y C debe responderse en una hoja de papel diferente. La recogida se producirá escalonadamente, en los momentos que constarán en la pizarra; primero, la hoja de codificación, y luego los problemas A, B, y C en este orden.
5. El formulario de examen tiene siete hojas numeradas correlativamente al pie (del 0.1 al 0.7). Cerciórate de recibirlas todas, y reclama si tu formulario fuera incompleto. Hay distintos tipos de examen. Este es del tipo 0; marca un 0 en la columna I de tu hoja de codificación, como en el ejemplo.
6. Los puntos obtenibles en cuestiones y problemas son 30 y 30 respectivamente. Son precisos 15 y 15 para superar el examen. Cuestionarios con puntuaciones iguales o superiores a 14 pueden en algún caso ser compensados por una buena nota en los problemas.
7. Rellena tus datos en la hoja de codificación y pliegos de papel suministrados. En "Convocatorias" (columna II) pondrás el número de convocatorias consumidas *incluyendo ésta.*

Ejemplo:

$\underbrace{\text{12545}}$ PEREZ, Ernesto Examen tipo $\underbrace{\text{0}}$ Convocatorias

CUESTIONES (Duración: 1 hora 30 minutos)

1. PREGUNTA-REGALO. La capital de España es:

   (A) París (B) Sebastopol (C) Madrid (D) Londres (E) Pekín

2. Sea $\{X_n\}_{n\in\mathcal{N}}$ una sucesión de variables aleatorias con la siguiente función característica:

$$\psi_n(u) = \left(1 - \frac{iu}{n^3}\right)^{-1}$$

   La sucesión convergerá:

   (A) Sólo en distribución a $X = 0$

   (B) Sólo en distribución a $X = 1$

   (C) En distribución y probabilidad a $X = 1$

   (D) En distribución y probabilidad a $X = 0$

   (E) Todo falso

3. Sea $\{X_n\}_{n\in\mathcal{N}}$ una sucesión de variables aleatorias con la siguiente función de cuantía:

$$P_n(x) = \begin{cases} \frac{1}{2} & \text{si } x = -\frac{1}{n} \\ \frac{1}{2} & \text{si } x = \frac{1}{n} \end{cases}$$

   La sucesión convergerá:

   (A) En distribución, probabilidad y media cuadrática a $X = \frac{1}{2}$

   (B) En distribución, probabilidad y media cuadrática a $X = 0$

   (C) Sólo en distribución a $X = 0$

   (D) Sólo en distribución a $X = \frac{1}{2}$

   (E) Todo falso

Las cuestiones 4 a 6 hacen referencia al siguiente enunciado:

La probabilidad de que al entrar en una tienda de electrodomésticos un cliente realice una compra es 0.20. Se supone independencia entre las entradas de clientes.

4. Si entran 20 clientes en la tienda, la probabilidad de que cinco de ellos realicen una compra es:

   (A) 0.8042 (B) 0.2182 (C) 0.1091 (D) 0.6296 (E) 0.1746

5. La probabilidad de que, entre los 20 clientes, al menos diez de ellos realicen una compra es:

   (A) 0.0074 (B) 0.0026 (C) 0.9974 (D) 0.0006 (E) 0.9994

6. Si ahora son 225 los clientes que entran en la tienda, la probabilidad aproximada de que no más de 41 clientes realicen una compra es:

   (A) 0.3546 (B) 0.7190 (C) 0.2810 (D) 0.5398 (E) 0.4602

Las cuestiones 7 y 8 hacen referencia al siguiente enunciado:

El número de coches que pasa **por minuto** por una determinada intersección de calles sigue una distribución de Poisson de media 6. Se asume independencia entre los coches que pasan por la mencionada intersección de calles.

7. La probabilidad de que en un minuto pasen exactamente seis coches por la mencionada intersección de calles es:

(A) 0.6063 (B) 0.7440 (C) 0.4457 (D) 0.1606 (E) 0.1377

8. La probabilidad de que en **6 minutos** pasen exactamente 42 coches por la mencionada intersección de coches es, aproximadamente:

(A) 0.8599 (B) 0.0387 (C) 0.1788 (D) 0.1401 (E) 0.8212

Las cuestiones 9 a 11 hacen referencia al siguiente enunciado:

Sean X_1, X_2 y X_3 variables aleatorias independientes entre sí con distribuciones normales de medias respectivas -1, 0 y 1 y varianzas 9, 1 y 4, respectivamente.

9. La probabilidad de que la v.a. $W = \dfrac{(X_1+1)^2}{9} + X_2^2$ tome valores menores que 0.103 es:

(A) 0.95 (B) 0.10 (C) 0.05 (D) 0.90 (E) 0.25

10. La probabilidad de que la v.a. $Y = \dfrac{2X_2}{(X_3-1)}$ tome valores menores que 1.376 es:

(A) 0.40 (B) 0.60 (C) 0.10 (D) 0.80 (E) 0.20

11. La probabilidad de que la v.a. Y^2 tome valores mayores que 161 es:

(A) 0.05 (B) 0.10 (C) 0.01 (D) 0.95 (E) 0.90

Las cuestiones 12 y 13 hacen referencia al siguiente enunciado:

Sean X e Y variables aleatorias independientes entre sí y tales que $X \in \gamma(a,r)$ e $Y \in \gamma(b,s)$.

12. Si definimos la v.a. $W = 2aX + 2bY$, la distribución de la variable aleatoria W es:

(A) $\gamma(2, r+s)$ (B) $\gamma(1/2, 2(r+s))$ (C) $\gamma(1/2, r+s)$ (D) $\exp(r+s)$ (E) $\gamma(2, 2(r+s))$

13. Si definimos la v.a. $Z = (X/b) + (Y/a)$, la distribución de la variable aleatoria Z es:

(A) $\gamma(ab, r+s)$ (B) $\gamma(a/b, r+s)$ (C) $\gamma(1/ab, r+s)$ (D) $\gamma(b/a, r+s)$ (E) χ^2_{r+s}

14. Sea X una variable aleatoria con función de densidad de probabilidad:

$$f(x,\theta) = \frac{1}{2^\theta \Gamma(\theta)}\ x^{\theta-1}\ e^{-x/2},\ x>0,\ \theta>0$$

Para estimar el parámetro θ se ha tomado una m.a.s. de tamaño n, $X_1, \ldots, X_n$. Un estadístico suficiente para el parámetro θ es:

(A) $\sum_{i=1}^n X_i$ (B) $\prod_{i=1}^n X_i$ (C) $\prod_{i=1}^n \ln X_i$ (D) $\prod_{i=1}^n \left(\frac{1}{X_i}\right)$ (E) $\sum_{i=1}^n X_i^2$

Las cuestiones 15 y 16 hacen referencia al siguiente enunciado:

Sea X una v.a. con función de cuantía dada por:

$$P(X=0)=\frac{1}{\theta};\ \ P(X=1)=\frac{5}{2\theta}$$

$$P(X=2)=\frac{(\theta-4)}{\theta};\ \ P(X=3)=\frac{1}{2\theta}$$

Para estimar el parámetro θ se ha tomado una m.a.s. de tamaño $n=20$, proporcionando los valores muestrales siguientes:

X	0	1	2	3
frecuencia	8	5	3	4

15. La estimación de θ por el método de máxima verosimilitud es:

(A) $\frac{40}{17}$ (B) $\frac{17}{80}$ (C) $\frac{80}{17}$ (D) $\frac{17}{40}$ (E) $\frac{60}{17}$

16. La estimación de θ por el método de momentos es:

(A) $\frac{80}{17}$ (B) $\frac{17}{80}$ (C) $\frac{60}{17}$ (D) $\frac{17}{40}$ (E) $\frac{40}{17}$

17. Sea X una variable aleatoria con función de densidad de probabilidad dada por:

$$f(x;\theta)=\theta\ e^{-\theta x},\qquad x>0,\qquad \theta>0$$

Se obtiene una m.a.s. de tamaño n: $X_1,\ldots,X_n$. El estimador máximo verosímil de θ es:

(A) $\frac{1}{\overline{X}-1}$ (B) $1-\overline{X}$ (C) $\overline{X}$ (D) $\overline{X}-1$ (E) $\frac{1}{\overline{X}}$

Las cuestiones 18 a 20 hacen referencia al siguiente enunciado:

Sea Z una v.a. tal que $Z\in b(p,n)$. Para estimar el parámetro p, se proponen los dos estimadores siguientes:

$$T_1=\frac{Z}{n};\qquad T_2=\frac{Z+1}{n+2}$$

18. Se verifica que:

(A) Ambos estimadores son insesgados (B) Sólo T_2 es insesgado (C) -

(D) Ambos estimadores son sesgados (E) Sólo T_1 es insesgado

19. Se verifica que:

(A) Sólo T_1 es consistente (B) Sólo T_2 es consistente (C) Ambos estimadores son consistentes

(D) - (E) -

20. El error cuadrático medio (ECM) del estimador T_1, para $p=1/2$, es:

(A) $\frac{1}{4}$ (B) $\frac{1}{4n}$ (C) $\frac{1}{2}$ (D) $\frac{4}{n}$ (E) $\frac{1}{2n}$

21. Sea X una variable aleatoria con la siguiente función de densidad de probabilidad:

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

Para contrastar la hipótesis de que $\theta = \frac{1}{2}$ frente a la alternativa de que $\theta = 1$, se toma una m.a.s. de tamaño $n = 1$. La región crítica más potente para dicha observación, para un determinado nivel de significación, es de la forma:

(A) $X \in (C_1, C_2)$ (B) $X \in (C_1, C_2)^c$ (C) $X \leq C$ (D) $X \geq C$ (E) Todo falso

Las cuestiones 22 y 23 hacen referencia al siguiente enunciado:

La variable aleatoria X sigue una distribución con la siguiente función de cuantía:

$$P(X = -1) = 3\theta^3 \qquad P(X = 0) = 1 - 6\theta^3 \qquad P(X = 1) = 3\theta^3$$

Para contrastar la hipótesis nula de que $\theta = 0.50$ frente a la alternativa de que $\theta = 0.20$ se toma una m.a.s. de tamaño $n = 1$ y se decide rechazar la hipótesis nula si el valor obtenido es $X = 0$.

22. El nivel de significación de la prueba es:

(A) 0.375 (B) 0.024 (C) 0.625 (D) 0.250 (E) 0.750

23. La probabilidad de error tipo II de la prueba es:

(A) 0.048 (B) 0.952 (C) 0.375 (D) 0.250 (E) 0.750

Las cuestiones 24 y 25 hacen referencia al siguiente enunciado:

El número de clientes que acude cada hora a una sucursal bancaria sigue una distribución de Poisson. El director de la sucursal considera que debe abrir una nueva ventanilla de atención al público si el promedio de personas que acuden cada hora a solicitar atención es al menos 7. Para contrastar la necesidad de abrir otra ventanilla, es decir $H_0 : \lambda \geq 7$ frente a la hipótesis alternativa $H_1 : \lambda < 7$, dispone de la información de una hora de atención.

24. La regla de decisión más potente al nivel de significación 0.10 será rechazar H_0 si el número de clientes que acude cada hora a solicitar atención en la mencionada sucursal bancaria es:

(A) $X \geq 3$ (B) $X \geq 4$ (C) $X \leq 7$ (D) $X \leq 4$ (E) $X \leq 3$

25. Para este contraste y para un valor $\lambda = 5$, la probabilidad de error tipo II será:

(A) 0.8753 (B) 0.2650 (C) 0.1247 (D) 0.7350 (E) 0.1404

Las cuestiones 26 y 27 hacen referencia al siguiente enunciado:

Un individuo está interesado en comprar un determinado modelo de lectora de DVD. Antes de hacerlo pregunta el precio en 31 comercios, obteniendo un precio medio muestral de 105 euros con una desviación típica muestral de 20 euros. Se supone normalidad.

26. Con una confianza del 90% se puede afirmar que el precio medio de la lectora de DVD se encuentra en el intervalo:

(A) (97.55, 112.45) (B) (100.22, 109.78) (C) (99.25, 110.75)
(D) (98.79, 111.21) (E) (101.35, 108.65)

27. Con una confianza del 90% se puede afirmar que la varianza de la lectora de DVD se encuentra en el intervalo:

(A) (307.69, 601.94) (B) (14.16, 33.51) (C) (283.11, 670.27)

(D) (255.64, 706.34) (E) (326.44, 585.45)

Las cuestiones 28 y 29 hacen referencia al siguiente enunciado:

Un investigador de la Facultad de Biología de la Universidad del País Vasco pretende estimar la proporción de ratas que, habiendo sido expuestas a un cierto factor de riesgo, desarrollarán un cáncer pulmonar. El investigador selecciona aleatoriamente 150 ratas de entre las que han sido expuestas al factor de riesgo durante los últimos años y de las que se tuviese información sobre el posible desarrollo de cáncer pulmonar. Los datos indicaron que de las 150 ratas seleccionadas, 57 habían desarrollado cáncer pulmonar.

28. El intervalo del 95% de confianza para la proporción de ratas que han sido expuestas al factor de riesgo y que desarrollarían cáncer pulmonar es, aproximadamente:

(A) (0.34, 0.42) (B) (0.32, 0.44) (C) (0.25, 0.51) (D) (0.28, 0.48) (E) (0.30, 0.46)

29. Al nivel de significación 5% se desea contratar la hipótesis nula de que la proporción de ratas expuestas al factor de riesgo que desarrollan cáncer pulmonar es **mayor o igual** que 0.40. La decisión será:

(A) No rechazar la hipótesis nula (B) No se puede decidir (C) Rechazar la hipótesis nula

(D) - (E) -

30. Se desea contrastar si el tipo de asistencia sanitaria que un paciente recibe en una emergencia está relacionado con el entorno en donde se encuentra su residencia habitual. Para ello se toma una m.a.s. y se clasifica a los pacientes de acuerdo con su residencia habitual (entorno urbano o entorno rural) y el tipo de asistencia sanitaria recibida en caso de emergencia (asistencia domiciliaria, ambulatorio, o traslado a hospital más cercano). El tipo de contraste que se ha de realizar es:

(A) Homogeneidad

(B) Independencia

(C) Comparación de medias

(D) Ajuste a una distribución totalmente especificada

(E) Ajuste a una distribución parcialmente especificada

PROBLEMAS (Duración: 75 minutos)

A. (10 puntos, 25 minutos)

Sea X una variable aleatoria continua con la siguiente función de densidad de probabilidad:

$$f(x,\theta) = \begin{cases} \dfrac{2\theta}{(1-\theta)}\, x^{(3\theta-1)/(1-\theta)} & \text{si } 0 \le x \le 1 \\ 0 & \text{en otro caso} \end{cases}$$

Para estimar el parámetro θ se ha tomado una m.a.s. de tamaño n, $X_1, \ldots, X_n$. **Ayuda**: Se sabe $\mathrm{E}(X) = \frac{2\theta}{\theta+1}$.

a) Obtén **detalladamente** el estimador de θ por el método de los momentos.

b) Obtén **detalladamente** el estimador máximo verosímil de θ.

B. (10 puntos, 25 minutos)

La siguiente tabla recoge la función de cuantía de la v.a. discreta X bajo la hipótesis nula ($P_0(x)$) y bajo la hipótesis alternativa ($P_1(x)$).

X	1	2	3	4	5	6
$P_0(x)$	0	0	0.05	0.10	0.40	0.45
$P_1(x)$	0.35	0.30	0.15	0.20	0	0

Tenemos una muestra aleatoria simple de tamaño $n = 1$ para contrastar la hipótesis nula $H_0 : P(x) = P_0(x)$ frente a la hipótesis alternativa $H_1 : P(x) = P_1(x)$.

a) ¿Incluirías los puntos $X = \{1, 2\}$ en la región crítica? Explica.

b) ¿Incluirías los puntos $X = \{5, 6\}$ en la región crítica? Explica.

c) Al nivel de significación del 10% y proporcionando todos los detalles utilizados para obtener la respuesta solicitada, obtén la región crítica más potente para este contraste. **Nota**: Antes de responder a este apartado recuerda lo que has respondido en los apartados anteriores.

C. (10 puntos, 25 minutos)

Una empresa quiere comercializar tres tipos de reproductores MP3 y tiene la siguiente información:

Tipo	1	2	3
Probabilidades	$(1-4\theta)$	2θ	2θ

a) Para estimar estas probabilidades se ha tomado una m.a.s. de 50 personas que ha arrojado los siguientes resultados: 20 personas compraron reproductores MP3 del tipo 1; 20 personas compraron reproductores del tipo 2 y 10 personas compraron reproductores del tipo 3. Estimar por el método de máxima verosimilitud el parámetro θ.

b) Contrastar al nivel de significación del 5% si la distribución de probabilidades que tiene la empresa es la correcta.

SOLUCIONES DEL CUESTIONARIO

1: C	11: A	21: D
2: D	12: C	22: D
3: B	13: A	23: A
4: E	14: B	24: E
5: B	15: C	25: D
6: C	16: A	26: D
7: D	17: E	27: C
8: B	18: E	28: E
9: C	19: C	29: A
10: D	20: B	30: B

SOLUCIONES DE LOS PROBLEMAS

A)

La función de densidad de probabilidad de la v.a. X es

$$f(x,\theta) = \begin{cases} \dfrac{2\theta}{(1-\theta)}\; x^{(3\theta-1)/(1-\theta)} & \text{si } 0 \le x \le 1 \\ \\ 0 & \text{en otro caso} \end{cases}$$

Para estimar el parámetro θ se ha tomado una m.a.s. de tamaño n, $X_1, \ldots, X_n$, y se sabe que $\mathrm{E}(X) = \frac{2\theta}{\theta+1}$.

a) Estimador por el método de momentos

Igualamos el primer momento poblacional al primer momento muestral. Es decir,

$$\alpha_1 = \mathrm{E}(X) = a_1 = \frac{\sum_{i=1}^n X_i}{n} = \overline{X}$$

Por tanto, como sabemos que $\mathrm{E}(X) = \frac{2\theta}{\theta+1}$, se tiene que:

$$\alpha_1 = \mathrm{E}(X) = \frac{2\theta}{\theta+1} = a_1 = \frac{\sum_{i=1}^n X_i}{n} = \overline{X} \Longrightarrow 2\theta = \overline{X}\theta + \overline{X}$$

$$\Longrightarrow \theta(2-\overline{X}) = \overline{X} \Longrightarrow \hat{\theta}_{MM} = \frac{\overline{X}}{(2-\overline{X})}$$

b) Estimador máximo verosímil

La función de verosimilitud de la muestra es:

$$\begin{aligned} L(\theta) &= f(x_1;\theta)\ \ldots\ f(x_n;\theta) \\ &= \left[\frac{2\theta}{(1-\theta)}\; x_1^{(3\theta-1)/(1-\theta)}\right] \cdots \left[\frac{2\theta}{(1-\theta)}\; x_n^{(3\theta-1)/(1-\theta)}\right] \\ &= \frac{2^n\theta^n}{(1-\theta)^n}\; \left[\prod_{i=1}^n x_i\right]^{(3\theta-1)/(1-\theta)} \end{aligned}$$

Calculamos su logaritmo neperiano:

$$\ln L(\theta) = n\ln 2 + n\ln\theta - n\ln(1-\theta) + \left(\frac{3\theta-1}{1-\theta}\right)\; \ln\left[\prod_{i=1}^n x_i\right]$$

Si derivamos respecto de θ e igualamos a cero tenemos que:

$$\frac{\partial \ln L(\theta)}{\partial\theta} = \frac{n}{\theta} + \frac{n}{(1-\theta)} + \ln\left[\prod_{i=1}^n x_i\right]\left[\frac{(1-\theta)(3)-(3\theta-1)(-1)}{(1-\theta)^2}\right] = 0$$

$$\Longrightarrow \frac{n}{\theta} + \frac{n}{(1-\theta)} + \left[\frac{2}{(1-\theta)^2}\right]\ln\left[\prod_{i=1}^n x_i\right] = 0,$$

$$\Longrightarrow \left[\frac{2}{(1-\theta)^2}\right] \ln\left[\prod_{i=1}^{n} x_i\right] = -\frac{n}{\theta} - \frac{n}{(1-\theta)} = -\left[\frac{n(1-\theta)+n\theta}{\theta(1-\theta)}\right] = -\frac{n}{\theta(1-\theta)}$$

$$\Longrightarrow \left[\frac{2}{(1-\theta)}\right] \ln\left[\prod_{i=1}^{n} x_i\right] = -\frac{n}{\theta}$$

$$\Longrightarrow 2\theta \ln\left[\prod_{i=1}^{n} x_i\right] = -n + n\theta \Longrightarrow n = \theta\left\{n - 2\ln\left[\prod_{i=1}^{n} x_i\right]\right\}$$

de donde,

$$\hat{\theta}_{MV} = \frac{n}{\{n - 2\ln\left[\prod_{i=1}^{n} X_i\right]\}} = \frac{n}{[n - 2\sum_{i=1}^{n}\ln(X_i)]}$$

B)

Queremos contrastar la hipótesis nula de que X es una v.a. discreta con función de cuantía $P_0(x)$ frente a la alternativa de que la función de cuantía es $P_1(x)$:

X	1	2	3	4	5	6
$P_0(x)$	0	0	0.05	0.10	0.40	0.45
$P_1(x)$	0.35	0.30	0.15	0.20	0	0

Hemos tomado una m.a.s. de tamaño $n = 1$; es decir, observamos X.

a) ¿Incluirías los puntos $X = \{1, 2\}$ en la región crítica?

Dado que, bajo la distribución de probabilidad de la hipótesis nula $P_0(x)$, estos puntos tienen probabilidad cero, la v.a. no puede tomar estos valores bajo la hipótesis nula. Por tanto, $X = \{1, 2\}$ son puntos de rechazo de H_0 y, por tanto, **siempre** deben incluirse en la región criítica.

b) ¿Incluirías los puntos $X = \{5, 6\}$ en la región crítica?

Dado que, bajo la distribución de probabilidad de la hipótesis alternativa $P_1(x)$, estos puntos tienen probabilidad cero, la v.a. no puede tomar estos valores bajo la hipótesis alternativa, pero sí bajo la hipótesis nula. Por tanto, $X = \{5, 6\}$ son puntos de no rechazo de H_0 y, por tanto, **nunca** deben incluirse en la región criítica.

c) Al nivel de significación $\alpha = 0.10$ y recordando lo respondido en los apartados anteriores, tenemos que las posibles regiones críticas para este contraste son $\mathrm{RC}_1 = \{1, 2, 3\}$ y $\mathrm{RC}_2 = \{1, 2, 4\}$. Esto se debe a que:

$$\alpha_1 = P(X \in \mathrm{RC}_1|P_0) = P(X = 1, 2, 3|P_0) = 0 + 0 + 0.05 = 0.05 \leq \alpha = 0.10$$
$$\alpha_2 = P(X \in \mathrm{RC}_2|P_0) = P(X = 1, 2, 4|P_0) = 0 + 0 + 0.10 = 0.10 \leq \alpha = 0.10$$

Para ver cuál de estas dos regiones críticas es la más potente para este contraste, calculamos las respectivas potencias:

$$\text{Potencia}_1 = P(X \in \mathrm{RC}_1|P_1) = P(X = 1, 2, 3|P_1) = 0.35 + 0.30 + 0.15 = 0.80$$
$$\text{Potencia}_2 = P(X \in \mathrm{RC}_2|P_1) = P(X = 1, 2, 4|P_1) = 0.35 + 0.30 + 0.20 = 0.85$$

De lo anterior, concluimos que, al nivel de significación $\alpha = 0.10$, la región crítica RC_2 es la más potente para este contraste.

C)

Se trata de un **contraste de bondad de ajuste a una distribución parcialmente especificada.** Los datos son los que aparecen a continuación:

Tipo	1	2	3
Probabilidades	$(1-4\theta)$	2θ	2θ

a) Se ha tomado una m.a.s. de 50 personas que ha arrojado los siguientes resultados: 20 personas compraron reproductores MP3 del tipo 1; 20 personas compraron reproductores del tipo 2 y 10 personas compraron reproductores del tipo 3 ($n = 50$). Para estimar el parámetro θ por el método de máxima verosimilitud, tenemos que la función de verosimilitud será:

$$L(\theta) = (1-4\theta)^{20}(2\theta)^{20}(2\theta)^{10} = 2^{30}\theta^{30}(1-4\theta)^{20}$$

Tomando logaritmos neperianos, tendremos que:

$$\ln L(\theta) = 30\ln 2 + 30\ln(\theta) + 20\ln(1-4\theta)$$

Tomando derivadas e igualando a cero para maximizar la función de verosimilitud:

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \frac{30}{\theta} - \frac{80}{(1-4\theta)} = 0 \Longrightarrow 30(1-4\theta) = 80\theta \Longrightarrow 30 - 120\theta = 80\theta$$

$$\Longrightarrow 30 = 200\theta \Longrightarrow \hat{\theta}_{MV} = \frac{30}{200} = 0.15$$

b) Contraste de de bondad de ajuste a una distribución parcialmente especificada.

En primer lugar, tenemos que las probabilidades estimadas, p_i de cada uno de los tipos de reproductores MP3 serán:

$$P(\text{Tipo 1}) = (1-4(0.15)) = 0.40; \qquad P(\text{Tipo 2}) = P(\text{Tipo 3}) = 2(0.15) = 0.30$$

Dado que hemos estimado el parámetro θ, tenemos que $h = 1$. Además, como tenemos tres tipos de reproductores MP3, $k = 3$. Con estos datos y para realizar el contraste, construimos la tabla:

	n_i	$\hat{p}_i$	$n\hat{p}_i$	$\frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i}$
Tipo 1	20	0.40	20	0
Tipo 2	20	0.30	15	1.67
Tipo 3	10	0.30	15	1.67
	$n = 50$	1	$n = 50$	$z \simeq 3.34$

Bajo la hipótesis nula de que la distribución de probabilidades que tiene la empresa es la correcta, el estadístico $\sum_i \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i} \sim \chi^2_{k-h-1}$, donde k es el número de tipos de reproductores MP3 ($k = 3$) y h es el número de parámetros estimados ($h = 1$).

La regla de desisión es rechazar la hipótesis nula al nivel de significación aproximado del $\alpha = 5\%$ si:

$$z > \chi^2_{k-h-1\,,\,0.05} = \chi^2_{1,0.05}$$

En este caso:

$$z = 3.34 < 3.84 = \chi^2_{1,0.05}$$

por lo que, a un nivel de significación aproximado del $\alpha = 5\%$, no se rechaza la hipótesis nula de que la distribución de probabilidades de los tipos de reproductores MP3 de la empresa es la correcta.