ESTADISTICA EMPRESARIAL - Segundo Curso — Curso 2.009-10

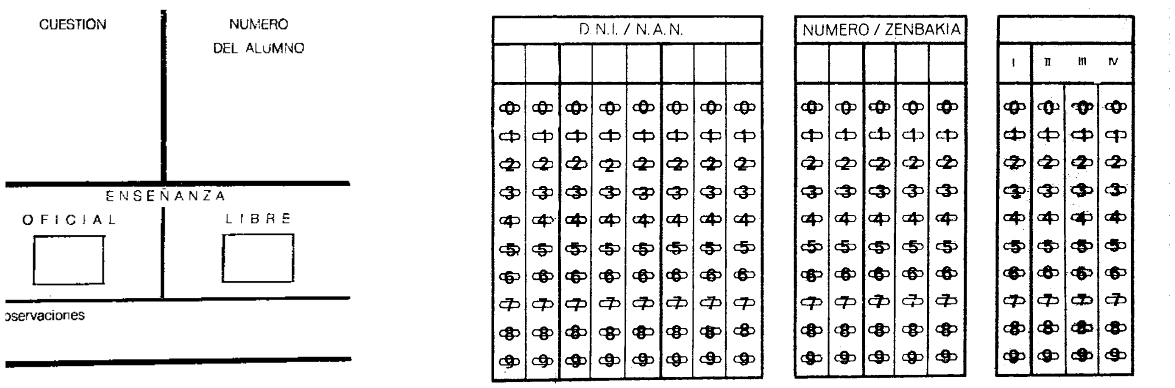
Convocatoria de Junio. 16-06-10

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 seis hojas numeradas correlativamente al pie (del 0.1 al 0.6). 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:

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



CUESTIONES (Duración: 1 hora 40 minutos)

1. **PREGUNTA–REGALO.** La capital de España es:

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

Las cuestiones 2 a 6 hacen referencia al siguiente enunciado:

El porcentaje de clientes de un determinado banco que muestra interés por un nuevo producto tras una campaña publicitaria es del 40%. Se supone independencia entre los clientes.

2. Si se seleccionan al azar 7 clientes, la probabilidad de que exactamente 2 de ellos muestren interés por el producto es:

   (A) 0.0124 (B) 0.0036 (C) 0.0940 (D) 0.0774 (E) 0.2613

3. Si se seleccionan al azar 20 clientes, la probabilidad de que muestren interés por el producto menos de 10 de ellos es:

   (A) 0.8725 (B) 0.2403 (C) 0.1597 (D) 0.7553 (E) 0.1275

4. Si se seleccionan al azar 20 clientes, la probabilidad de que menos de 15 **no** muestren interés por el producto es:

   (A) 0.1256 (B) 0.9997 (C) 0.0016 (D) 0.8744 (E) 0.7984

5. Si el banco tiene 36000 clientes, ¿cuántos clientes se espera que muestren interés por el nuevo producto?

   (A) 144 (B) 86.4 (C) 8640 (D) 14400 (E) 5760

6. Si el banco tiene 36000 clientes, la probabilidad aproximada de que más de 14500 muestren interés por el producto es:

   (A) 0.8599 (B) 0.8632 (C) 0.1703 (D) 0.1401 (E) 0.1368

Las cuestiones 7 a 10 hacen referencia al siguiente enunciado:

El número de bicicletas prestadas en un cuarto de hora por un determinado ayuntamiento sigue una distribución de Poisson con varianza igual a 1.75. Se supone independencia entre los préstamos en diferentes momentos.

7. La probabilidad de que en un cuarto de hora sean prestadas exactamente 2 bicicletas es:

   (A) 0.1308 (B) 0.4145 (C) 0.5322 (D) 0.2661 (E) 0.2072

8. El número más probable de bicicletas prestadas en una hora es:

   (A) 6 (B) 7 y 8 (C) 7 (D) 8 (E) 6 y 7

9. La probabilidad aproximada de que en una hora sean prestadas menos de 5 bicicletas es:

   (A) 0.8088 (B) 0.1730 (C) 0.1912 (D) 0.8270 (E) 0.3134

10. La probabilidad aproximada de que en el transcurso de 5 horas sean prestadas al menos 35 bicicletas es:

(A) 0.5319 (B) 0.4681 (C) 0.1701 (D) 0.4361 (E) 0.5639

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

$$P_n(x) = \begin{cases} 1 - \frac{2+5n}{10n^2} & \text{si } x = 1 \\ \frac{1}{2n} & \text{si } x = 3 \\ \frac{1}{5n^2} & \text{si } x = 5 \end{cases}$$

La sucesión converge:

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

(B) Sólo en distribución a $X = 3$

(C) Sólo en probabilidad a $X = 3$

(D) En distribución y probabilidad a $X = 3$

(E) Sólo en distribución a $X = 1$

12. Sea $\{X_n\}_{n\in\mathcal{N}}$ una sucesión de variables aleatorias con función característica dada por $\psi_n(u) = \frac{n^2}{n^2-iu}$. Esta sucesión de v.a. convergerá:

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

(B) Sólo en probabilidad a $X = 0$

(C) Sólo en distribución a $X = 0$

(D) Sólo en distribución a $X = 4$

(E) En distribución y probabilidad a $X = 4$

Las cuestiones 13 a 15 hacen referencia al siguiente enunciado:

La duración, en minutos, de las llamadas telefónicas de un individuo es una variable aleatoria, X, con distribución uniforme en el intervalo $(0, 10)$

13. La probabilidad de que una llamada dure entre 2 y 5 minutos es:

(A) 0.3 (B) 0.5 (C) 0.7 (D) 0.4 (E) 0.6

14. Si el individuo realiza 80 llamadas independientes, en un mes, la distribución aproximada de la duración total de las llamadas es:

(A) $U(0, 800)$ (B) exp(1/400) (C) P(400) (D) $N(400,\ 666.67)$ (E) $U(0, 80)$

15. Si cada llamada le cuesta 0.2 euros de establecimiento de llamada y 0.1 euros por cada minuto hablado, el precio medio, en euros, de las 80 llamadas será:

(A) 56 (B) 40.2 (C) 8.2 (D) 28 (E) 48

16. Si X es una variable aleatoria con distribución $\gamma(4,2)$, su función característica es:

(A) $(1-4iu)^{-2}$ (B) $\left(1-\frac{iu}{2}\right)^{-4}$ (C) Todo falso (D) $\left(1-\frac{iu}{4}\right)^{-2}$ (E) $(1-2iu)^{-4}$

17. Sea X una variable aleatoria con distribución $\gamma(1,1)$. La $P(X>2)$ es:

(A) 0.2706 (B) 0.1353 (C) 0.8647 (D) 0.4212 (E) 0.5788

Las cuestiones 18 y 19 hacen referencia al siguiente enunciado:

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

$$P(X=-2)=\theta; \qquad P(X=0)=1-2\theta; \qquad P(X=2)=\theta.$$

Para estimar el parámetro θ se ha tomado una m.a.s. de tamaño $n=10$, en la que se ha obtenido 5 veces el valor -2, 3 veces el 0 y 2 veces el 2.

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

(A) 0.64 (B) 0.39 (C) 0.35 (D) 0.72 (E) 0.41

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

(A) 0.41 (B) 0.39 (C) 0.64 (D) 0.72 (E) 0.35

Las cuestiones 20 y 21 hacen referencia al siguiente enunciado:

Sea X una v.a. con función de densidad:

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

cuya media es 2θ. Para estimar el parámetro θ se ha tomado una m.a.s. de tamaño n, $X_1,\ldots,X_n$.

20. El estimador de θ por el método de máxima verosimilitud es:

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

21. El estimador de θ por el método de momentos es:

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

Las cuestiones 22 y 23 hacen referencia al siguiente enunciado:

Sea X una v.a. con distribución $N(m,\sigma^2=9)$ y una m.a.s. de tamaño 100, $X_1,\ldots,X_{100}$. Se proponen dos estimadores para m: $\hat{m}_1=\frac{X_1+\ldots+X_{100}}{100}=\overline{X}$ y $\hat{m}_2=3\overline{X}$.

22. ¿Son estimadores insesgados de m?

(A) Sólo $\hat{m}_1$ (B) Ambos (C) - (D) Ninguno (E) Sólo $\hat{m}_2$

23. Las varianzas de $\hat{m}_1$ y $\hat{m}_2$ son, respectivamente:

(A) 0.09 y 0.81 (B) 9 y 81 (C) 0.09 y 0.27 (D) 9 y 27 (E) Todo falso

24. Sea X una variable aleatoria con distribución $F_{3,5}$. La $P(X > 0.1883)$ es:

(A) 0.90 (B) 0.80 (C) 0.10 (D) 0.95 (E) 0.05

Las cuestiones 25 y 26 hacen referencia al siguiente enunciado:

Sea X una v.a. con distribución de Poisson de parámetro θ. Se quiere contrastar la hipótesis nula $H_0 : \theta = 3$ frente a la hipótesis alternativa $H_1 : \theta = 4$. Para ello se ha tomado una m.a.s. de tamaño 2, X_1, X_2. Se decide rechazar la hipótesis nula si $X_1 + X_2 < 3$.

25. La probabilidad de error de tipo I de la prueba es:

(A) 0.4232 (B) 0.9380 (C) 0.0620 (D) 0.1512 (E) 0.5768

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

(A) 0.9862 (B) 0.2381 (C) 0.7619 (D) 0.0138 (E) 0.4335

27. Se desea estimar la proporción de individuos que han visto un determinado anuncio publicitario. Para ello se toma una m.a.s. de 2000 individuos en la que se observa que el 72% han visto el anuncio. Un intervalo de confianza 0.95 para dicha proporción es, aproximadamente:

(A) $(0.70, 0.74)$ (B) $(0.028, 0.044)$ (C) $(0.32, 0.82)$ (D) $(0.13, 0.16)$ (E) $(0.61, 0.83)$

28. Sea X una variable aleatoria con distribución $N(m, \sigma^2)$. Se toma una m.a.s. de tamaño 15 de la que se obtiene una varianza muestral igual a 3.4. Un intervalo de confianza 0.95 para la varianza poblacional es:

(A) (1.95, 9.06) (B) (1.98, 11.96) (C) (1.57, 10.72) (D) (2.35, 7.96) (E) (2.04, 7.02)

Las cuestiones 29 y 30 hacen referencia al siguiente enunciado:

Se desea saber si el tipo de juego de azar en que un individuo gasta su dinero está relacionado con la edad del mismo. Para ello se toma **una m.a.s.** y se clasifica a los individuos de acuerdo con su edad $(\leq 30, 30 - 50, \geq 50)$ y el tipo de juego (lotería primitiva, lotería nacional, quiniela, bingo).

29. El tipo de contraste que se ha de realizar es:

(A) Independencia

(B) Ajuste a una distribución parcialmente especificada

(C) Homogeneidad

(D) Ajuste a una distribución totalmente especificada

(E) -

30. El número de grados de libertad de la distribución del estadístico de contraste de la prueba es:

(A) 6 (B) 4 (C) 12 (D) 3 (E) 11

PROBLEMAS (Duración: 75 minutos)

A. (10 puntos, 25 minutos)

Sea $X_1, \ldots, X_n$ una m.a.s. de tamaño n tomada de una v.a. X con distribución $\gamma(\frac{1}{\theta}, 3)$. Es decir, su función de densidad de probabilidad viene dada por:

$$f(x;\theta) = \frac{1}{2\,\theta^3}\, x^2\, e^{-\frac{x}{\theta}}, \qquad x > 0, \quad \theta > 0$$

i) Obtén el estimador para el parámetro θ por el método de máxima verosimilitud.

ii) Obtén el estimador para el parámetro θ por el método de momentos.

iii) ¿Es el estimador por el método de momentos insesgado?, ¿consistente?, ¿eficiente?

Nota: Recuerda que la cota de Cramer-Rao para un estimador insesgado de θ, a partir de una m.a.s. de tamaño n es:

$$L_C = \frac{1}{nE\left[\frac{\partial \ln f(x,\theta)}{\partial \theta}\right]^2}$$

B. (10 puntos, 25 minutos)

El 2 de enero de 2010 se publica un artículo en el periódico sobre los índices de audiencia de las campanadas de fin de año. En ese artículo se afirma que, de aquellos que vieron las campanadas en televisión, el 50.8% las vieron en TVE1, el 21.2% en Tele 5, el 12.9% en televisiones autonómicas, el 3.3% en TVE2, el 3% en Antena 3, el 2% en La sexta, el 1.3% en Cuatro y el 5.5% en otras cadenas.

Para contrastar si los datos del artículo son ciertos se realiza una encuesta con personas seleccionadas de forma aleatoria obteniendose que, de los 500 individuos que afirman que vieron las campanadas en televisión, 200 las vieron en TVE1, 110 en Tele 5, 70 en televisiones autonómicas, 15 en TVE2, 20 en Antena 3, 15 en La sexta, 50 en Cuatro y 20 en otras cadenas.

Contrasta, al nivel de significación del 5% la hipótesis de que el artículo informa sobre los índices de audiencia reales.

C. (10 puntos, 25 minutos)

Un conductor acostumbra a controlar semanalmente el consumo medio de gasolina en litros cada 100 Km. Un día realiza un curso sobre conducción eficiente y quiere saber si realmente su consumo de gasolina ha disminuido, por lo que continúa controlando el consumo. Al cabo del tiempo dispone de datos de 25 semanas anteriores al curso realizado con un consumo medio de 8.4 litros cada 100 Km y una desviación típica de 2.5 litros y de las 16 semanas posteriores al curso con un consumo medio de 6.0 litros a los 100 y una desviación típica de 2.1 litros.

Nota: Se supone que antes y después del curso los datos semanales son independientes y la distribución del consumo es normal.

i) Contrasta la igualdad de varianzas del consumo medio antes y después del curso. Utiliza para ello un nivel de significación del 10%.

ii) Contrasta la hipótesis nula de que el consumo medio de gasolina se ha mantenido sin cambios frente a la hipótesis alternativa de que se ha reducido después del curso. Utiliza para ello un nivel de significación del 10%.

SOLUCIONES DEL CUESTIONARIO (tipo 0)

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

SOLUCIONES DE LOS PROBLEMAS

Problema A

Sea $X_1, \ldots, X_n$ una m.a.s. de tamaño n tomada de una v.a. X con distribución $\gamma(\frac{1}{\theta}, 3)$. Es decir, su función de densidad de probabilidad viene dada por:

$$f(x;\theta) = \frac{1}{2\,\theta^3}\; x^2\; e^{-\frac{x}{\theta}}, \qquad x > 0, \quad \theta > 0$$

i) Así, la función de verosimilitud estará dada por

$$L(\vec{x};\theta) = f(x_1;\theta) f(x_2;\theta) \cdots f(x_n;\theta)$$

$$L(\vec{x};\theta) = \left(\frac{1}{2\,\theta^3}\; x_1^2\; e^{-\frac{x_1}{\theta}}\right)\left(\frac{1}{2\,\theta^3}\; x_2^2\; e^{-\frac{x_2}{\theta}}\right)\;\cdots\;\left(\frac{1}{2\,\theta^3}\; x_n^2\; e^{-\frac{x_n}{\theta}}\right)$$

$$L(\vec{x};\theta) = \frac{1}{2^n\,\theta^{3n}}\;\left(\prod_{i=1}^{n} x_i^2\right)\; e^{-\frac{\sum_{i=1}^n x_i}{\theta}}$$

El estimador máximo verosímil de θ es el valor que maximiza la función de verosimilitud o, equivalentemente, su logaritmo neperiano:

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

Derivando respecto de θ, tenemos que:

$$\frac{\partial \ln L(\vec{x};\theta)}{\partial\theta} = -\frac{3n}{\theta} + \frac{\sum_{i=1}^n x_i}{\theta^2} = 0$$

Por lo tanto,

$$\frac{3n\theta}{\theta^2} = \frac{\sum_{i=1}^n x_i}{\theta^2}$$

$$\Rightarrow \qquad \hat{\theta}_{\text{MV}} = \frac{\sum_{i=1}^n X_i}{3n} = \frac{\overline{X}}{3}$$

ii) Para calcular el estimador de θ por el método de momentos, debemos encontrar $\alpha_1 = \mathrm{E}(X)$. En este caso, por ser X una v.a. con distribución gamma: $\gamma(a = \frac{1}{\theta}, r = 3)$

$$\alpha_1 = \mathrm{E}(X) = \frac{r}{a} = \frac{3}{(1/\theta)} = 3\theta$$

De forma que, después de igualar los momentos poblacionales y muestrales; es decir, $\alpha_1 = a_1$; tenemos que:

$$\alpha_1 = \mathrm{E}(X) = 3\theta = \overline{x} = a_1,$$

$$\Rightarrow \hat{\theta}_{\text{MM}} = \frac{\overline{X}}{3}$$

iii) Insesgadez:

Para ver si el estimador por el método de momentos es insesgado hay que comprobar si $\mathrm{E}(\hat{\theta}_{\mathrm{MM}}) = \theta$.

En este caso,

$$\mathrm{E}(\hat{\theta}_{\mathrm{MM}}) \;=\; \mathrm{E}\left(\frac{\overline{X}}{3}\right) \;=\; \frac{1}{3}\,\mathrm{E}(\overline{X}) \;=\; \frac{1}{3}\,\mathrm{E}(X) \;=\; \frac{1}{3}\,(3\,\theta) \;=\; \theta$$

Por lo tanto, sí es insesgado.

Consistencia

Para ver si el estimador es consistente calculamos la varianza del estimador.

$$\mathrm{Var}(\hat{\theta}_{\mathrm{MM}}) \;=\; \mathrm{Var}\left(\frac{\overline{X}}{3}\right) \;=\; \frac{1}{9}\,\mathrm{Var}(\overline{X}) \;=\; \frac{1}{9}\left[\frac{\mathrm{Var}(X)}{n}\right]$$

Necesitamos $\mathrm{Var}(X)$. En este caso, por ser X una v.a. con distribución gamma:

$$\mathrm{Var}(X) = \frac{r}{a^2} = \frac{3}{(1/\theta^2)} = 3\theta^2$$

Por lo tanto,

$$\mathrm{Var}(\hat{\theta}_{\mathrm{MM}}) \;=\; \frac{3\theta^2}{9n} \;=\; \frac{\theta^2}{3n}$$

Dado que se trata de un estimador insesgado y cuya varianza tiende a 0 cuando n tiende a infinito, se cumplen las condiciones suficientes para la consistencia. Por tanto, podemos afirmar que se trata de un estimador consistente para θ.

Eficiencia:

Para probar si es eficiente hay que comprobar si su varianza coincide con la cota de Cramer-Rao. Calculamos la cota de Cramer-Rao para estimadores regulares e insesgados:

$$L_c = \frac{1}{n\mathrm{E}\left(\frac{\partial \ln f(x;\theta)}{\partial\theta}\right)^2}$$

En nuestro caso,

$$\ln f(x;\theta) = -\ln 2 - 3\ln\theta + 2\ln x - \frac{x}{\theta}$$

$$\frac{\partial f(x;\theta)}{\partial\theta} = -\frac{3}{\theta} + \frac{x}{\theta^2} \;=\; \frac{1}{\theta^2}\,(x - 3\theta)$$

$$\mathrm{E}\left(\frac{\partial \ln f(x;\theta)}{\partial\theta}\right)^2 = \mathrm{E}\left(\frac{X-3\theta}{\theta^2}\right)^2 = \frac{\mathrm{E}\,(X-3\theta)^2}{\theta^4} = \frac{\mathrm{Var}(X)}{\theta^4} \;=\frac{1}{\theta^4}\;\left(3\theta^2\right) = \frac{3}{\theta^2}$$

Sustituyendo en la cota de Cramer-Rao esta esperanza se llega a:

$$L_c = \frac{1}{n\left(\frac{3}{\theta^2}\right)} = \frac{\theta^2}{3n} = \mathrm{Var}(\hat{\theta}_{\mathrm{MM}})$$

Como la varianza del estimador es igual a la cota de Cramer-Rao, el estimador por el método de los momentos es un estimador eficiente.

Problema B

Se trata de un ajuste a una distribución totalmente especificada.

Construimos la tabla:

	p_i	n_i	np_i	$(n_i - np_i)$	$\frac{(n_i-np_i)^2}{np_i}$
TVE1	0.508	200	254	-54	11.4803
Tele 5	0.212	110	106	+4	0.1509
Tel. Auto.	0.129	70	64.5	+5.5	0.4690
TVE2	0.033	15	16.5	-1.5	0.1364
Antena 3	0.030	20	15	+5	1.6667
La Sexta	0.020	15	10	+5	2.5000
Cuatro	0.013	50	6.5	+43.5	291.1154
Otras Cad.	0.055	20	27.5	-7.5	2.0455
	1	$n = 500$	1	0	$z = 309.5642$

Bajo la hipótesis nula de ajuste a una distribución totalmente especificada, el estadístico $\sum_i \frac{(n_i-np_i)^2}{np_i} \sim \chi^2_{(K-1)}$, donde K es el número de categorías en que ha sido dividida la muestra ($K = 8$).

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

$$z > \chi^2_{(8-1)\,,\,0.05} = \chi^2_{7,0.05}.$$

En este caso:

$$z = 309.5642 > 14.1 = \chi^2_{7,0.05},$$

por lo que se rechaza la hipótesis nula de ajuste a la distribución especificada a un nivel de significación del 5%. Es decir, se puede afirmar que la muestra obtenida no se ajusta a los índices de audiencia sobre los que informa el artículo.

Problema C

Sean X y Y las dos v.a. que representan el consumo de gasolina antes y después del curso, respectivamente. Las dos v.a. son independientes $X \sim N(m_X, \sigma_X^2)$ e $Y \sim N(m_Y, \sigma_Y^2)$. Se han tomado dos m.a.s. de tamaño n_1 y n_2, $\vec{X} = (X_1, X_2, \ldots, X_{n_1})$ e $\vec{Y} = (Y_1, Y_2, \ldots, Y_{n_2})$.

Los datos muestrales son:

$$n_1 = 25 \qquad n_2 = 16$$

$$\overline{x} = 8.4 \qquad \overline{y} = 6$$

$$s_X = 2.5 \qquad s_Y = 2.1$$

i) Contrasta la igualdad de varianzas del consumo medio antes y después del curso. Utiliza para ello un nivel de significación del 10%.

$H_0 : \sigma_X^2 = \sigma_Y^2 \quad \Rightarrow \quad \frac{\sigma_X^2}{\sigma_Y^2} = 1$

$H_1 : \sigma_X^2 \neq \sigma_Y^2 \quad \Rightarrow \quad \frac{\sigma_X^2}{\sigma_Y^2} \neq 1$

Se sabe que:

$$\frac{n_1 S_X^2}{\sigma_X^2} \in \chi^2_{n_1-1} \qquad \frac{n_2 S_Y^2}{\sigma_Y^2} \in \chi^2_{n_2-1}$$

De aquí, es inmediato que:

$$\frac{\left(n_1 S_X^2/\sigma_X^2\right)/(n_1-1)}{\left(n_2 S_Y^2/\sigma_Y^2\right)/(n_2-1)} = \frac{n_1(n_2-1)S_X^2}{n_2(n_1-1)S_Y^2} \frac{1}{\underbrace{\left(\sigma_X^2/\sigma_Y^2\right)}_{=1 \text{ bajo } H_0}} \in \mathcal{F}_{n_1-1,n_2-1}$$

$$\text{RC} = \left\{ \frac{s_X^2}{s_Y^2} \quad \text{t.q.} \quad \frac{n_1(n_2-1)s_X^2}{n_2(n_1-1)s_Y^2} \leq F_{\overline{n_1-1,n_2-1}|1-\alpha/2} \quad \text{ó} \quad \frac{n_1(n_2-1)s_X^2}{n_2(n_1-1)s_Y^2} \geq F_{\overline{n_1-1,n_2-1}|\alpha/2} \right\}$$

En nuestro caso:

$F_{\overline{n_1-1,n_2-1}|\alpha/2} = F_{\overline{25-1,16-1}|0.05} = 2.29$

$F_{\overline{n_1-1,n_2-1}|1-\alpha/2} = F_{\overline{25-1,16-1}|0.95} = \frac{1}{F_{\overline{15,24}|0.05}} = \frac{1}{2.11} = 0.4739$

$\frac{n_1(n_2-1)s_X^2}{n_2(n_1-1)s_Y^2} = \frac{25\,(16-1)\,(2.5)^2}{16\,(25-1)\,(2.1)^2} = \frac{2343.75}{1693.44} = 1.3840$

Dado que:

$0.4379 \leq 1.3840 \leq 2.29,$

a un nivel de significación del 10%, no se rechaza la hipótesis nula de igualdad de varianzas antes y después del curso.

ii) Contrasta la hipótesis nula de que el consumo medio de gasolina se ha mantenido sin cambios frente a la hipótesis alternativa de que se ha reducido después del curso. Utiliza para ello un nivel de significación del 10%.

Dado el resultado del apartado anterior estamos en la situación de distribuciones normales e independientes, y varianzas desconocidas pero iguales, tenemos que:

$$\frac{(\bar{X}-\bar{Y})-(m_X-m_Y)}{\sqrt{\frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}}} \in N(0,1)$$

$$\frac{n_1 S_X^2}{\sigma^2} \in \chi^2_{n_1-1} \qquad \frac{n_2 S_Y^2}{\sigma^2} \in \chi^2_{n_2-1} \qquad \Rightarrow \quad \left(\frac{n_1 S_X^2 + n_2 S_Y^2}{\sigma^2}\right) \in \chi^2_{n_1+n_2-2}$$

$$\frac{\frac{(\bar{X}-\bar{Y})-(m_X-m_Y)}{\sqrt{\frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}}}}{\sqrt{\frac{\left(\frac{n_1 S_X^2+n_2 S_Y^2}{\sigma^2}\right)}{n_1+n_2-2}}} = \frac{(\bar{X}-\bar{Y})-(m_X-m_Y)}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\sqrt{\frac{n_1 S_X^2+n_2 S_Y^2}{n_1+n_2-2}}} \in t_{n_1+n_2-2}$$

$H_0 : m_X = m_Y \qquad \equiv \qquad m_X - m_Y = 0$

$H_1 : m_X > m_Y \qquad \equiv \qquad m_X - m_Y > 0$

Bajo H_0 :

$$\frac{(\bar{X}-\bar{Y})-0}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\sqrt{\frac{n_1 S_X^2+n_2 S_Y^2}{n_1+n_2-2}}} \in t_{n_1+n_2-2}$$

$$\text{RC} = \left\{ \bar{x} \text{ e } \bar{y} \quad \text{tales que} \quad \frac{(\bar{x}-\bar{y})-0}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\sqrt{\frac{n_1 s_X^2+n_2 s_Y^2}{n_1+n_2-2}}} \geq t_{\overline{n_1+n_2-2}|\alpha} \right\}$$

En este caso

$t_{\overline{n_1+n_2-2}|\alpha} \;=\; t_{\overline{25+16-2}|0.10} \;=\; t_{\overline{39}|0.10} = 1.30$, y

$\frac{(8.4-6)-0}{\sqrt{\frac{1}{25}+\frac{1}{16}}\sqrt{\frac{25\,(2.5)^2+16\,(2.1)^2}{25+16-2}}} = 3.1085$

Dado que:

$3.1085 \geq 1.30$,

a un nivel de significación del 10%, se rechaza la hipótesis nula de igualdad de medias antes y después del curso y, por tanto, se concluye que el curso ha reducido el consumo medio de gasolina.