

ESTADÍSTICA E INTRODUCCIÓN A LA ECONOMETRÍA
Curso 2022/23
Problemas del Tema 5

OBSERVACIÓN: Los problemas marcados con asterisco deben resolverse utilizando Excel. En todos los problemas, supondremos que la muestra que se utiliza es una muestra aleatoria simple.

1.- Una empresa de componentes electrónicos afirma en su último informe que el 93% de los componentes que fabrican no contienen defectos, el 5% tienen un defecto, y el 2% tienen dos defectos. En una muestra de 500 componentes, se han encontrado 454 sin defectos, 30 con un defecto y 16 con dos defectos. Con estos datos, queremos analizar si hay evidencia suficiente para rechazar la afirmación que hace la compañía en su informe, con nivel de significación aproximado 0.10.

a) Indica cuáles son las hipótesis nula y alternativa que debemos utilizar, qué valor toma el estadístico de contraste, cuál es la región crítica y cuál es la conclusión.

b) Calcula el p-valor aproximado del contraste realizado en el apartado anterior, e indica si habríamos obtenido la misma conclusión en nuestro análisis si hubiéramos utilizado 0.05 como nivel de significación aproximado.

2.- Un profesor afirma que sus alumnos no estudian nada y que, por tanto, si realiza un examen tipo test con 4 preguntas de verdadero/falso, el número de preguntas acertadas por un alumno es una variable aleatoria (v.a.) con distribución binomial $Bi(4; \frac{1}{2})$. Para comprobar su afirmación, el profesor hace un examen con estas características a 240 alumnos, indicándoles que deben contestar a todas las preguntas. Los resultados del examen se recogen en la siguiente tabla:

Respuestas acertadas	0	1	2	3	4
Número de alumnos	8	48	90	70	24

Con estos datos, queremos analizar si hay evidencia suficiente para decir que la afirmación del profesor es falsa, con nivel de significación aproximado 0.05.

a) Indica cuáles son las hipótesis nula y alternativa que debemos utilizar, qué valor toma el estadístico de contraste, cuál es la región crítica y cuál es la conclusión.

b) A la vista de los cálculos realizados en el apartado anterior, ¿piensas que los resultados obtenidos son mejores o peores que los resultados que cabría esperar si los alumnos escogiesen las respuestas al azar?

3.- Basándose en datos históricos, una fábrica chocolatera considera que el 50% de sus clientes prefiere chocolate con leche, el 45% prefiere chocolate negro, el 4% prefiere chocolate con almendras y el 1% prefiere chocolate con avellanas. Sin embargo, se sospecha que en los últimos años ha habido un cambio en las preferencias de los clientes. Para analizar si ha sido así, se ha encuestado a 200 clientes de la compañía, obteniéndose los resultados siguientes:

Chocolate Preferido	Con Leche	Negro	Con Almendras	Con Avellanas
Número de clientes	98	83	16	3

Con estos datos, la empresa quiere analizar si hay evidencia suficiente para pensar que las preferencias de los clientes han cambiado.

a) ¿Crees que es razonable efectuar este análisis con el contraste de bondad de ajuste chi-cuadrado, utilizando cuatro categorías diferentes? Si tu respuesta es afirmativa, calcula el valor que toma el estadístico de contraste; si tu respuesta es negativa, indica qué cambios tendrían que hacerse para poder efectuar este análisis con el contraste de bondad de ajuste chi-cuadrado y, con estos cambios, calcula el valor que toma el estadístico de contraste.

b) Si tomamos como nivel de significación aproximado 0.05, ¿hay evidencia suficiente para pensar que las preferencias de los clientes han cambiado?

4.- Se dispone de una muestra con los rendimientos de 1480 activos financieros en el último periodo. Llamando $x_1, x_2, \dots, x_{1480}$ a estos rendimientos, se sabe que:

$$\begin{aligned} \sum_{i=1}^{1480} x_i &= 101.22 & \sum_{i=1}^{1480} (x_i - \bar{x})^2 &= 531.72 \\ \sum_{i=1}^{1480} (x_i - \bar{x})^3 &= -24.62 & \sum_{i=1}^{1480} (x_i - \bar{x})^4 &= 688.15 \end{aligned}$$

a) Con estos datos, ¿crees que la distribución de los rendimientos es asimétrica a la derecha, asimétrica a la izquierda, o aproximadamente simétrica?

b) Con estos datos, ¿crees que hay una presencia de rendimientos extremos similar, inferior o superior a la que cabría esperar si la distribución de los rendimientos fuera normal?

c) Contrasta con nivel de significación aproximado 0.01 la hipótesis de que la distribución de los rendimientos es normal, utilizando el contraste de normalidad Jarque-Bera y, en caso de que se rechaze la hipótesis de normalidad, indica cuál crees que es la causa de este rechazo.

*5.- Una gran cadena comercial quiere analizar si las ventas anuales de sus establecimientos siguen una distribución normal. En la hoja “Problema 5” del archivo Excel adjunto pueden verse las ventas (en miles de euros) en el año 2018 de 40 establecimientos de la cadena.

a) Representa el histograma agrupando las observaciones en 6 clases y, teniendo en cuenta el histograma obtenido, indica si puede pensarse que las ventas anuales siguen una distribución normal.

b) Contrasta con nivel de significación 0.05 la hipótesis nula de que las ventas anuales tienen distribución normal utilizando el contraste de normalidad Jarque-Bera. (NOTA: La relación entre el coeficiente de exceso curtosis muestral $CUR_X - 3$ y el valor que da la función CURTOSIS de Excel es:

$$CUR_X - 3 = \frac{(n - 2)(n - 3)}{(n - 1)(n + 1)} CURTOSIS(A1:A_n) - \frac{6}{n + 1}$$

siendo n el tamaño muestral. Además, ten en cuenta que cuando se utiliza este contraste con $n = 40$ observaciones y nivel de significación 0.05, en la región crítica debe reemplazarse el valor $\chi_{2;0.05}^2$ por 4.7407).

6.- Se ha realizado una encuesta a 400 individuos elegidos al azar de una determinada población con el fin de estudiar la posible relación entre nivel educativo (clasificado en tres categorías: estudios obligatorios/estudios medios/estudios universitarios) y consumo de tabaco (clasificado en dos categorías: consumo habitual/no consumo habitual). De los 232 individuos encuestados con menor nivel de estudios, 70 consumen tabaco habitualmente. De los 116 individuos encuestados con nivel de estudios medio, 21 consumen tabaco habitualmente. De los 52 individuos encuestados con mayor nivel de estudios, 9 consumen tabaco habitualmente. Con estos datos, queremos analizar si hay evidencia de que hay dependencia entre nivel educativo y consumo de tabaco, con nivel de significación aproximado 0.05.

a) Indica cuáles son las hipótesis nula y alternativa que debemos utilizar, qué valor toma el estadístico de contraste, cuál es la región crítica y cuál es la conclusión.

b) Calcula el p-valor aproximado del contraste realizado en el apartado anterior, e indica si habríamos obtenido la misma conclusión en nuestro análisis si hubiéramos utilizado 0.01 como nivel de significación aproximado.

7.- Se desea saber si el grado de satisfacción con la sanidad pública es el mismo en cada una de las tres provincias que forman una comunidad autónoma. Se ha encuestado a 400 personas de esta comunidad y se les ha pedido que indiquen la provincia donde residen y si están muy satisfechos con la sanidad pública, medianamente satisfechos o poco satisfechos. La tabla siguiente muestra algunos de los resultados obtenidos:

Grado de satisfacción \ Provincia	Provincia A	Provincia B	Provincia C	Total Fila
Muy satisfechos	?	27	46	?
Medianamente satisfechos	59	?	?	?
Poco satisfechos	31	15	?	101
Total Columna	122	84	?	400

a) Completa la tabla, calculando los valores señalados con “?”.

b) ¿Hay evidencia suficiente para decir que el grado de satisfacción con la sanidad pública no es independiente de la provincia de residencia, tomando 0.05 como nivel de significación aproximado?

c) Un informe afirma que en toda la población de esta comunidad autónoma hay un 30% de personas muy satisfechas con la sanidad pública, un 50% de personas medianamente satisfechas y un 20% de personas poco satisfechas. A la vista de los resultados obtenidos con esta muestra, analiza si existen motivos para considerar que los porcentajes presentados en este informe son inadmisibles, con nivel de significación aproximado 0.05.

***8.-** Se ha realizado una encuesta a un grupo de 100 estudiantes de grado de una universidad. Se les ha preguntado qué tipo de estudios seguían (Ciencias/Ciencias Sociales/Letras), en qué curso estaban (Primero/Segundo/Tercero/Cuarto) y si utilizaban los servicios de biblioteca (SÍ/NO). Los resultados obtenidos pueden verse en la hoja “Problema 8” del archivo Excel adjunto.

a) Contrasta la hipótesis nula de independencia entre el tipo de estudios y la frecuencia de uso de los servicios de biblioteca, con nivel de significación aproximado 0.05. Si se rechaza la hipótesis nula, indica a qué crees que se debe el rechazo.

b) Contrasta la hipótesis nula de independencia entre el curso en que está el estudiante y la frecuencia de uso de los servicios de biblioteca, con nivel de significación aproximado 0.05. Si se rechaza la hipótesis nula, indica a qué crees que se debe el rechazo.