Tema 9. Contrastes no paramétricos

Cuadernillo de ejercicios resueltos con contexto económico

Estadística II · Grado en Economía · URJC

En este documento se mantiene la notación del curso para este tema: en los contrastes χ^2 de bondad del ajuste se usan O_i para las frecuencias observadas, E_i para las frecuencias esperadas y $c_i = (O_i-E_i)^2/E_i$ para la contribución de la categoría i al estadístico. En independencia y homogeneidad se usan O_{ij} y E_{ij}. En Kolmogorov–Smirnov se usa D_n para la mayor discrepancia entre $F_n(x)$ y F(x). Como norma de trabajo, en la variable aleatoria y en los cuantiles críticos se redondea a 2 decimales, y en las probabilidades a 4 decimales. En el caso de KS, la constante c_α se toma de la tabla del curso y el valor crítico se calcula como K ≈ c_α/√n.

1. Criterio de trabajo del cuadernillo

Antes de calcular conviene automatizar siempre la misma secuencia de decisión. La mayoría de los errores en contrastes no paramétricos no nacen de un error de cálculo, sino de elegir mal el contraste o de interpretar mal el resultado.

2. Lista de contenidos del cuadernillo

Bloque	Contenido	Ejercicios
I	Elección del contraste y formulación de hipótesis	1-2
II	Contrastes χ^2 de bondad del ajuste	3-5
III	Contrastes χ^2 de independencia	6-8
IV	Contrastes χ^2 de homogeneidad	9-11
V	Contraste de Kolmogorov–Smirnov	12-14
VI	Otros contrastes no paramétricos	15-16
VII	Ejercicios de integración	17-18

3. Síntesis inicial: qué contraste usar

Situación	Pregunta básica	Procedimiento	Distribución / regla bajo H_0
Una muestra; categorías o clases	¿La distribución observada coincide con una distribución teórica?	χ^2 de bondad del ajuste	χ^2 con gl = r−1−k
Una muestra; dos variables cualitativas	¿Las variables son independientes?	χ^2 de independencia	χ^2 con gl = (f−1)(c−1)
Varias poblaciones; una variable cualitativa	¿Las poblaciones tienen la misma distribución?	χ^2 de homogeneidad	χ^2 con gl = (f−1)(c−1)
Una muestra; distribución continua	¿La muestra es compatible con una F(x) teórica?	Kolmogorov–Smirnov	D_n y K ≈ (c_α)/√n

4. Ejercicios resueltos

Bloque I. Elección del contraste y formulación de hipótesis

Ejercicio 1. Elegir correctamente el contraste

Enunciado. Para cada situación, indíquese qué contraste no paramétrico corresponde. a) Una empresa quiere comprobar si el reparto de compras por días laborables sigue una distribución uniforme. b) Se estudia si el tipo de jornada y el uso habitual del transporte público están relacionados en una muestra de personas trabajadoras. c) Se comparan tres grupos de edad para ver si presentan la misma distribución del canal de compra. d) Se quiere comprobar si una muestra de importes tipificados puede proceder de una distribución continua teórica F(x).

Solución

Paso 1. Identificación de la estructura de datos. En el apartado a) hay una sola muestra clasificada en categorías y se compara con una distribución teórica. En b) hay una sola muestra con dos variables cualitativas observadas conjuntamente. En c) se comparan varias poblaciones respecto de una sola variable cualitativa. En d) se compara una distribución empírica con una función de distribución teórica continua.

Paso 2. Selección del procedimiento. En a) corresponde un χ^2 de bondad del ajuste. En b) corresponde un χ^2 de independencia. En c) corresponde un χ^2 de homogeneidad. En d) corresponde un contraste de Kolmogorov–Smirnov de una muestra.

Paso 3. Justificación mínima. El criterio de elección no depende de gustos formales, sino del tipo de pregunta y de la estructura de los datos. El bloque χ^2 trabaja con frecuencias por categorías o por celdas; KS compara funciones de distribución acumulada.

Observación. Este ejercicio debe automatizar la elección del contraste antes de empezar cualquier cuenta.

Ejercicio 2. Formular H_0 y H_1 en contexto económico

CORE

Enunciado. Para cada situación, escríbanse la hipótesis nula y la alternativa con el sentido correcto. a) Una plataforma quiere comprobar si el reparto de suscripciones por plan tarifario sigue la estructura histórica 0,50; 0,30; 0,20. b) Se estudia si nivel educativo y tipo de contrato son independientes. c) Tres ciudades se comparan para saber si la distribución del medio de pago es la misma. d) Se sospecha que una muestra no procede de la distribución continua teórica $F(x)=x^2$ en [0;1].

Solución

Paso 1. Bondad del ajuste. En a), H_0 afirma que la población sigue la distribución teórica dada; H_1 afirma que no la sigue.

Paso 2. Independencia. En b), H_0: las dos variables son independientes. H_1: las dos variables no son independientes.

Paso 3. Homogeneidad. En c), H_0: las poblaciones son homogéneas respecto del medio de pago. H_1: al menos una distribución difiere.

Paso 4. Kolmogorov–Smirnov. En d), H_0: la muestra procede de una población con función de distribución $F(x)=x^2$. H_1: la muestra no procede de esa población.

Observación. En estos contrastes la formulación correcta de H_0 depende del tipo de pregunta: misma distribución; independencia; homogeneidad; o compatibilidad con una F(x) teórica.

Bloque II. Contrastes χ^2 de bondad del ajuste

Ejercicio 3. Bondad del ajuste con distribución uniforme

Enunciado. Una plataforma de contenidos clasifica 200 accesos de pago según la franja horaria en que se realizaron: 08–10; 10–12; 12–14; 14–16; 16–18. Quiere comprobar si los accesos se reparten uniformemente entre las 5 franjas. Las frecuencias observadas son 30; 46; 40; 50; 34. Contrástese al 5%.

Solución

Paso 1. Identificación del problema. Hay una sola muestra clasificada en 5 categorías y se quiere comparar la distribución observada con una distribución teórica uniforme. Corresponde un χ^2 de bondad del ajuste.

Paso 2. Hipótesis. H_0: $p_1 = p_2 = p_3 = p_4 = p_5 = 0{,}20$. H_1: la distribución no es uniforme.

Paso 3. Frecuencias esperadas. Si H_0 es cierta, en cada día debería esperarse $E_i = n \cdot p_i = 200 \cdot 0{,}20 = 40{,}00$.

Día	L	M	X	J	V
O_i	30	46	40	50	34
E_i	40,00	40,00	40,00	40,00	40,00
$c_i=(O_i-E_i)^2/E_i$	2,50	0,90	0,00	2,50	0,90

Paso 4. Estadístico. Sumamos las contribuciones: $\chi^2 = 2{,}50 + 0{,}90 + 0{,}00 + 2{,}50 + 0{,}90 = 6{,}80$.

Paso 5. Grados de libertad y valor crítico. Como hay r = 5 categorías y no se estima ningún parámetro, gl = r−1 = 4. Con α = 0,05 y la tabla χ^2 del curso, el valor crítico es $\chi^2\_\{1-\alpha;4\} = \chi^2\_\{0{,}95;4\} = 9{,}49$.

Paso 6. Decisión. Como 6,80 < 9,49, no rechazamos H_0.

Paso 7. Conclusión económica. La muestra no aporta evidencia suficiente para afirmar que los accesos de pago se distribuyan de forma no uniforme entre las cinco franjas horarias.

Observación. En bondad del ajuste el paso decisivo es construir bien las frecuencias esperadas antes de calcular el estadístico.

Ejercicio 4. Bondad del ajuste con distribución no uniforme

CORE

Enunciado. Un banco digital clasifica 300 nuevas cuentas abiertas según el canal de captación: oficina; recomendación; web; app; convenio. La estructura histórica es 0,10; 0,20; 0,40; 0,20; 0,10. En una muestra reciente se observan 18; 42; 150; 66; 24 altas. Contrástese al 5% si el patrón reciente coincide con la estructura histórica.

Solución

Paso 1. Identificación del problema. Hay una sola muestra clasificada en categorías y se compara con unas probabilidades teóricas conocidas. Corresponde un χ^2 de bondad del ajuste.

Paso 2. Hipótesis. H_0: la distribución de captación sigue 0,10; 0,20; 0,40; 0,20; 0,10. H_1: la distribución no sigue ese patrón.

Paso 3. Frecuencias esperadas. Multiplicamos cada probabilidad por n = 300: E_i = 30,00; 60,00; 120,00; 60,00; 30,00.

Canal	Oficina	Recomendación	Web	App	Convenio
O_i	18	42	150	66	24
E_i	30,00	60,00	120,00	60,00	30,00
c_i	4,80	5,40	7,50	0,60	1,20

Paso 4. Estadístico. La suma de contribuciones es $\chi^2 = 4{,}80 + 5{,}40 + 7{,}50 + 0{,}60 + 1{,}20 = 19{,}50$.

Paso 5. Grados de libertad y valor crítico. Hay $r = 5$ categorías y no se estima ningún parámetro, así que $gl = 4$. Con $\alpha = 0{,}05$, el valor crítico es 9,49.

Paso 6. Decisión. Como $19{,}50 > 9{,}49$, rechazamos H_0.

Paso 7. Conclusión económica. La muestra reciente indica que el patrón de captación ha cambiado respecto de la estructura histórica.

Observación. Conviene elegir ejemplos en los que la distribución teórica no sea uniforme para que el alumnado no confunda “esperada” con “igual para todos los grupos”.

Ejercicio 5. Comprobar si las frecuencias esperadas son razonables

PLUS

Enunciado. Una aseguradora clasifica 40 siniestros en 5 categorías de importe. Bajo H_0, las probabilidades teóricas son 0,50; 0,25; 0,15; 0,07; 0,03. Antes de resolver el contraste χ^2 de bondad del ajuste, revísese si las frecuencias esperadas son razonables y explíquese qué habría que hacer si no lo fueran.

Solución

Paso 1. Frecuencias esperadas. Multiplicamos $n = 40$ por cada probabilidad: E_i = 20,00; 10,00; 6,00; 2,80; 1,20.

Paso 2. Revisión de la condición práctica. Las dos últimas frecuencias esperadas son menores que 5. Por tanto, la aproximación χ^2 no es buena si se mantienen esas categorías tal como están.

Paso 3. Medida correctora. Lo habitual es reagrupar categorías contiguas hasta conseguir frecuencias esperadas suficientes. Por ejemplo, podría unirse la cuarta y la quinta categoría con la tercera si el contexto económico lo permite.

Paso 4. Conclusión operativa. Antes de aplicar el contraste, hay que rehacer la clasificación y recalcular las esperadas. No conviene seguir adelante con el estadístico original.

Observación. En χ^2 no basta con saber la fórmula; también hay que verificar que las frecuencias esperadas permitan usarla con sentido.

Bloque III. Contrastes χ^2 de independencia

Ejercicio 6. Independencia en una tabla 2×2

Enunciado. Se estudia si existe relación entre la situación laboral —empleo indefinido o temporal— y el uso habitual de banca móvil —sí o no— en una muestra de 100 personas. Los datos observados son los siguientes.

Solución

	Usa transporte público	No usa transporte público	Total
Jornada completa	**36**	**24**	**60**
Jornada parcial	**14**	**26**	**40**
Total	**50**	**50**	**100**

Paso 1. Identificación del problema. Hay una sola muestra y dos variables cualitativas observadas conjuntamente. Corresponde un χ^2 de independencia.

Paso 2. Hipótesis. H_0: situación laboral y uso habitual de banca móvil son independientes. H_1: no son independientes.

Paso 3. Frecuencias esperadas. Usamos E_{ij} = (total de fila · total de columna)/n. Así: E_{11} = (60·50)/100 = 30,00; E_{12} = 30,00; E_{21} = 20,00; E_{22} = 20,00.

Celda	O_{ij}	E_{ij}	$c_{ij}=(O_{ij}-E_{ij})^2/E_{ij}$
Completa / Usa	**36**	**30,00**	**1,20**
Completa / No usa	**24**	**30,00**	**1,20**
Parcial / Usa	**14**	**20,00**	**1,80**
Parcial / No usa	**26**	**20,00**	**1,80**

Paso 4. Estadístico. χ^2 = 1,20 + 1,20 + 1,80 + 1,80 = 6,00.

Paso 5. Grados de libertad y valor crítico. Como la tabla es 2×2, gl = (2−1)(2−1) = 1. Con α = 0,05, el valor crítico es χ^2_{0,95;1} = 3,84.

Paso 6. Decisión. Como 6,00 > 3,84, rechazamos H_0.

Paso 7. Conclusión económica. La muestra aporta evidencia de asociación entre la situación laboral y el uso habitual de banca móvil.

Observación. En una tabla 2×2 conviene mostrar celda a celda cómo se obtienen las frecuencias esperadas.

Ejercicio 7. Independencia en una tabla 2×3

CORE

Enunciado. Una empresa analiza si el canal habitual de compra depende del nivel de renta en una muestra de 150 clientes. Los datos observados son los siguientes.

Solución

	Renta baja	Renta media	Renta alta	Total
Compra web	**18**	**24**	**38**	**80**
Compra en tienda	**32**	**26**	**12**	**70**
Total	**50**	**50**	**50**	**150**

Paso 1. Identificación del problema. Hay una sola muestra y dos variables cualitativas: canal habitual y nivel de renta. Corresponde un χ^2 de independencia.

Paso 2. Hipótesis. H_0: ambas variables son independientes. H_1: existe asociación entre ellas.

Paso 3. Frecuencias esperadas. Para la fila “Compra web”, cada esperada vale E = (80·50)/150 = 26,67. Para la fila “Compra en tienda”, cada esperada vale E = (70·50)/150 = 23,33.

Celda	O_{ij}	E_{ij}	c_{ij}
Web / baja	**18**	**26,67**	**2,82**
Web / media	**24**	**26,67**	**0,27**
Web / alta	**38**	**26,67**	**4,81**
Tienda / baja	**32**	**23,33**	**3,22**
Tienda / media	**26**	**23,33**	**0,30**
Tienda / alta	**12**	**23,33**	**5,50**

Paso 4. Estadístico. $\chi^2 \approx 2,82 + 0,27 + 4,81 + 3,22 + 0,30 + 5,50 = 16,92$.

Paso 5. Grados de libertad y valor crítico. La tabla es 2×3, así que gl = (2−1)(3−1) = 2. Con α = 0,05, el valor crítico es χ^2_{0,95;2} = 5,99.

Paso 6. Decisión. Como 16,92 > 5,99, rechazamos H_0.

Paso 7. Conclusión económica. El canal habitual de compra no parece independiente del nivel de renta; la distribución observada sugiere una relación entre ambas variables.

Observación. Cuando la tabla no es 2×2, los grados de libertad cambian y conviene escribirlos explícitamente.

Ejercicio 8. Interpretación correcta del resultado en independencia

PLUS

Enunciado. En un contraste χ^2 de independencia se ha obtenido $\chi^{2*} = 7,40$ con gl = 2 y α = 0,05. Interprétese el resultado en lenguaje estadístico y en lenguaje económico. Después explíquese por qué este resultado no permite afirmar causalidad.

Solución

Paso 1. Comparación con el valor crítico. Para gl = 2 y α = 0,05, el valor crítico es 5,99. Como 7,40 > 5,99, se rechaza H_0.

Paso 2. Interpretación estadística. La muestra no es compatible con la hipótesis de independencia entre las variables estudiadas.

Paso 3. Interpretación económica. En el contexto económico, hay evidencia de asociación entre ambas variables; por tanto, conocer una de ellas ayuda a describir la distribución de la otra.

Paso 4. Límite interpretativo. El contraste detecta relación estadística, pero no demuestra una relación causal. Pueden existir variables omitidas o mecanismos de selección que expliquen la asociación.

Observación. En este bloque es tan importante decidir bien como saber interpretar qué significa y qué no significa rechazar la independencia.

Bloque IV. Contrastes χ^2 de homogeneidad

Ejercicio 9. Homogeneidad en una tabla 3×3

Enunciado. Se comparan tres regiones —Norte; Centro; Sur— según el canal en el que se contrató por última vez un seguro: oficina; web; o app. Los datos observados son los siguientes.

Solución

Canal	Jóvenes	Adultos	Mayores	Total
Tienda física	**18**	**30**	**42**	**90**
Web	**20**	**24**	**12**	**56**
App	**22**	**16**	**6**	**44**
Total	**60**	**70**	**60**	**190**

Paso 1. **Identificación del problema**. Se comparan varias poblaciones —Norte; Centro; Sur— respecto de una variable cualitativa: el canal de contratación. Corresponde un χ^2 de homogeneidad.

Paso 2. Hipótesis. H_0: las tres regiones son homogéneas respecto del canal de contratación. H_1: al menos una distribución difiere.

Paso 3. Frecuencias esperadas. Aplicamos E_{ij} = (total de fila · total de columna)/n. Se obtiene: para "Tienda física", 28,42; 33,16; 28,42. Para "Web", 17,68; 20,63; 17,68. Para "App", 13,89; 16,21; 13,89.

Paso 4. Contribuciones. Las contribuciones aproximadas de las nueve celdas son: 3,82; 0,30; 6,49; 0,30; 0,55; 1,82; 4,74; 0,00; 4,48.

Paso 5. Estadístico. $\chi^2 \approx 3{,}82 + 0{,}30 + 6{,}49 + 0{,}30 + 0{,}55 + 1{,}82 + 4{,}74 + 0{,}00 + 4{,}48 = 22{,}50$.

Paso 6. Grados de libertad y valor crítico. La tabla es 3×3, así que gl = (3−1)(3−1) = 4. Con α = 0,05, el valor crítico es χ^2_{0,95;4} = 9,49.

Paso 7. Decisión. Como 22,50 > 9,49, rechazamos H_0.

Paso 8. Conclusión económica. La distribución del canal de contratación no es la misma en las tres regiones.

Observación. Aunque el estadístico tiene la misma forma que en independencia, aquí la interpretación es distinta: comparamos distribuciones entre poblaciones.

Ejercicio 10. Homogeneidad en una tabla 3×2

Enunciado. Tres regiones se comparan para estudiar si la distribución entre dos marcas de ahorro es la misma. Los datos observados son los siguientes.

Solución

Marca	Norte	Centro	Sur	Total
Marca A	**48**	**42**	**30**	**120**
Marca B	**22**	**28**	**40**	**90**
Total	**70**	**70**	**70**	**210**

Paso 1. Identificación del problema. Se comparan varias poblaciones —las tres regiones— respecto de una variable cualitativa con dos categorías —marca elegida—. Corresponde un χ^2 de homogeneidad.

Paso 2. Hipótesis. H_0: las tres regiones presentan la misma distribución entre las dos marcas. H_1: al menos una distribución difiere.

Paso 3. Frecuencias esperadas. Para la fila "Marca A", cada frecuencia esperada es E = (120·70)/210 = 40,00. Para la fila "Marca B", cada frecuencia esperada es E = (90·70)/210 = 30,00.

Celda	O_{ij}	E_{ij}	c_{ij}
A / Norte	48	40,00	1,60
A / Centro	42	40,00	0,10
A / Sur	30	40,00	2,50
B / Norte	22	30,00	2,13
B / Centro	28	30,00	0,13
B / Sur	40	30,00	3,33

Paso 4. Estadístico. $\chi^2 \approx 1{,}60 + 0{,}10 + 2{,}50 + 2{,}13 + 0{,}13 + 3{,}33 = 9{,}79$.

Paso 5. Grados de libertad y valor crítico. La tabla es 2×3, así que gl = (2−1)(3−1) = 2. Con $\alpha = 0{,}05$, el valor crítico es 5,99.

Paso 6. Decisión. Como 9,79 > 5,99, rechazamos H_0.

Paso 7. Conclusión económica. La distribución entre las dos marcas no es homogénea en las tres regiones.

Observación. En homogeneidad conviene insistir en que la pregunta es si las distribuciones son las mismas entre poblaciones; no si dos variables son independientes dentro de una sola muestra.

Ejercicio 11. Distinguir homogeneidad e independencia

PLUS

Enunciado. Explíquese por qué los siguientes dos enunciados no responden a la misma pregunta: a) “¿El tipo de contrato y el uso del transporte público son independientes?” b) “¿La distribución del medio de transporte es la misma en trabajadores con contrato fijo y temporal?”

Solución

Paso 1. Primer enunciado. En a) se observa una sola muestra con dos variables cualitativas y se pregunta por la asociación entre ellas. Eso es un problema de independencia.

Paso 2. Segundo enunciado. En b) se comparan dos poblaciones —contrato fijo y temporal— respecto de una variable cualitativa —medio de transporte—. Eso es un problema de homogeneidad.

Paso 3. Relación entre ambos. Los dos contrastes usan el mismo estadístico χ^2 y los mismos grados de libertad, pero la interpretación económica cambia porque la pregunta estadística es distinta.

Observación. Buena parte de los errores del tema nacen de no distinguir bien estas dos preguntas.

Bloque V. Contraste de Kolmogorov–Smirnov

Ejercicio 12. Kolmogorov–Smirnov con $F(x)=x^2$

Enunciado. Se quiere contrastar, con nivel de significación α = 0,02, si una muestra de tamaño n = 50 sobre un índice de intensidad de uso, reescalado al intervalo [0;1], puede proceder de una población continua con función de distribución $F(x)=x^2$. La muestra, agrupada en intervalos, ha dado las frecuencias 1; 7; 10; 18; 14 para [0,00;0,20); [0,20;0,40); [0,40;0,60); [0,60;0,80); [0,80;1,00].

Solución

Paso 1. Hipótesis. H_0: la muestra procede de una población con función de distribución $F(x)=x^2$. H_1: la muestra no procede de esa población.

Paso 2. Frecuencias acumuladas. A partir de 1; 7; 10; 18; 14, las frecuencias acumuladas son N_i = 1; 8; 18; 36; 50.

Paso 3. Distribución empírica acumulada. Dividimos cada acumulada entre n = 50: $F_n(x)$ = 0,02; 0,16; 0,36; 0,72; 1,00.

Paso 4. Distribución teórica en los extremos derechos. F(0,20)=0,04; F(0,40)=0,16; F(0,60)=0,36; F(0,80)=0,64; F(1,00)=1,00.

Extremo derecho	0,20	0,40	0,60	0,80	1,00
$F_n(x)$	0,02	0,16	0,36	0,72	1,00
$F(x)$	0,04	0,16	0,36	0,64	1,00
$\lvert F_n(x)-F(x)\rvert$	0,02	0,00	0,00	0,08	0,00

Paso 5. Estadístico. La mayor discrepancia es D_n = 0,08.

Paso 6. Valor crítico. Con α = 0,02, la tabla KS da c_α = 1,517. Luego K ≈ c_α/√n = 1,517/√50 ≈ 0,21.

Paso 7. Decisión. Como 0,08 < 0,21, no rechazamos H_0.

Paso 8. Conclusión económica. La muestra es compatible con una población cuya distribución acumulada es $F(x)=x^2$ para ese índice reescalado.

Observación. En KS el trabajo no consiste en sumar discrepancias, sino en quedarse con la mayor de ellas.

Ejercicio 13. Kolmogorov–Smirnov con distribución uniforme

Enunciado. Se quiere contrastar, con nivel de significación α = 0,05, si una muestra de tamaño n = 40 puede proceder de una distribución uniforme en [0;1], es decir, F(x)=x. Las frecuencias por intervalos son 20; 8; 6; 6 en [0,00;0,25); [0,25;0,50); [0,50;0,75); [0,75;1,00].

Solución

Paso 1. Hipótesis. H_0: la muestra procede de una uniforme en [0;1]. H_1: la muestra no procede de esa distribución.

Paso 2. Frecuencias acumuladas. Las acumuladas son N_i = 20; 28; 34; 40.

Paso 3. Distribución empírica acumulada. $F_n(x)$ = 20/40; 28/40; 34/40; 40/40 = 0,50; 0,70; 0,85; 1,00.

Paso 4. Distribución teórica. En los extremos derechos se tiene F(0,25)=0,25; F(0,50)=0,50; F(0,75)=0,75; F(1,00)=1,00.

Extremo derecho	**0,25**	**0,50**	**0,75**	**1,00**
$F_n(x)$	**0,50**	**0,70**	**0,85**	**1,00**
F(x)	**0,25**	**0,50**	**0,75**	**1,00**
$\lvert F_n(x)–F(x)\rvert$	**0,25**	**0,20**	**0,10**	**0,00**

Paso 5. Estadístico. La mayor discrepancia es D_n = 0,25.

Paso 6. Valor crítico. Con α = 0,05, la tabla KS da c_α = 1,358. Así, K ≈ 1,358/√40 ≈ 0,21.

Paso 7. Decisión. Como 0,25 > 0,21, rechazamos H_0.

Paso 8. Conclusión económica. La muestra no es compatible con una distribución uniforme en [0;1].

Observación. Es útil disponer de un ejemplo en el que se rechaza H_0 para que el alumnado vea los dos posibles desenlaces del contraste.

Ejercicio 14. Lectura e interpretación del estadístico D_n

PLUS

Enunciado. En un contraste de Kolmogorov–Smirnov de una muestra se ha obtenido D_n = 0,12. Para α = 0,05 y n = 60, la tabla del curso da c_α = 1,358. Interprétese el resultado y tómese la decisión adecuada.

Solución

Paso 1. Valor crítico. Calculamos K ≈ c_α/√n = 1,358/√60 ≈ 1,358/7,75 ≈ 0,18.

Paso 2. Decisión. Como D_n = 0,12 < 0,18, no rechazamos H_0.

Paso 3. Interpretación. La mayor discrepancia observada entre $F_n(x)$ y F(x) no es suficientemente grande, con ese tamaño muestral y ese nivel de significación, como para considerar que la muestra contradice la distribución teórica.

Observación. La lectura correcta del valor crítico de KS debe apoyarse en n y en la constante c_α de la tabla.

Bloque VI. Ejercicios de integración

Ejercicio 14. Elegir bien el contraste a partir del enunciado

Enunciado. Para cada situación, indíquese el contraste adecuado y justifíquese brevemente la elección. a) Se quiere comprobar si una distribución por categorías sigue el patrón histórico de mercado. b) Se estudia si dos variables cualitativas observadas en una misma muestra están relacionadas. c) Se comparan varias poblaciones respecto de una variable cualitativa. d) Se quiere contrastar si una muestra continua es compatible con una F(x) teórica dada.

Solución

Paso 1. Apartado a). Corresponde un χ^2 de bondad del ajuste, porque se compara una distribución observada con una distribución teórica por categorías.

Paso 2. Apartado b). Corresponde un χ^2 de independencia, porque se estudia la asociación entre dos variables cualitativas en una misma muestra.

Paso 3. Apartado c). Corresponde un χ^2 de homogeneidad, porque se comparan varias poblaciones respecto de una sola variable cualitativa.

Paso 4. Apartado d). Corresponde Kolmogorov–Smirnov de una muestra, porque se compara $F_n(x)$ con una distribución teórica continua F(x).

Observación. Antes de calcular, el alumnado debe poder etiquetar correctamente el problema estadístico.

Ejercicio 18. Conclusión económica final

Enunciado. En un estudio sobre hábitos de compra se obtiene $\chi^{2*} = 12{,}40$ en un contraste de homogeneidad con gl = 4 y α = 0,05. Redáctese la conclusión estadística y la conclusión económica adecuadas.

Solución

Paso 1. Comparación con el valor crítico. Con gl = 4 y α = 0,05, la tabla χ^2 del curso da un valor crítico de 9,49. Como 12,40 > 9,49, se rechaza H_0.

Paso 2. Conclusión estadística. La muestra no es compatible con la hipótesis de homogeneidad; al menos una de las distribuciones comparadas difiere.

Paso 3. Conclusión económica. Los hábitos de compra no se distribuyen del mismo modo en todas las poblaciones estudiadas. Por tanto, el patrón de compra cambia entre grupos y no conviene describirlos como si fueran homogéneos.

Observación. Un buen ejercicio de cierre obliga a pasar de la decisión numérica a una interpretación económica limpia y prudente.

5. Errores frecuentes que conviene evitar

- Confundir χ^2 de independencia con χ^2 de homogeneidad.
- Olvidar calcular correctamente las frecuencias esperadas antes de construir el estadístico χ^2.
- Usar mal los grados de libertad en tablas de contingencia.
- Interpretar asociación como causalidad.
- Olvidar que en χ^2 la región crítica está en la cola derecha.
- En KS, sumar discrepancias en lugar de quedarse con la discrepancia máxima D_n.
- En KS, olvidar que el valor crítico depende de α y del tamaño muestral n a través de $K \approx c\_α/\sqrt{n}$.
- Redactar la conclusión sin sentido económico o sin referirse a la pregunta planteada.

6. Bibliografía recomendada

- Ruiz-Maya, L.; Martín-Pliego López, F. J. Fundamentos de inferencia estadística. 3.ª ed. Thomson Paraninfo.
- Tablas de probabilidad y cuantiles críticos elaboradas para la asignatura.
- Material docente complementario del tema 9 de Estadística II.