
Clase 178 — Test chi-cuadrado de independencia y bondad de ajuste

Parte: 3 — Estadística Inferencial y Causal · Fuente: ISLP, cap. 4 + Bruce & Bruce, cap. 3
Chi-Square Test. Duración estimada: 70 min.

Clase 178 — Test chi-cuadrado de independencia y bondad de ajuste

Parte: 3 — Estadística Inferencial y Causal · Fuente: ISLP, cap. 4 + Bruce & Bruce, cap. 3 Chi-Square Test. Duración estimada: 70 min.

Objetivo

Aplicar el test chi-cuadrado de Pearson en sus dos formas: (a) independencia en una tabla de contingencia de dos variables categóricas, y (b) bondad de ajuste entre una distribución observada y una teórica. Reconocer cuándo el test es válido (frecuencias esperadas ≥ 5 por celda) y cuándo hay que recurrir a Fisher exact o a la simulación de Monte Carlo.

Resultados de aprendizaje

Al finalizar, el estudiante podrá:

- Construir una tabla de contingencia con `pd.crosstab` y aplicar `scipy.stats.chi2_contingency` interpretando `chi2`, `dof`, `pvalue` y `expected`.
- Verificar el supuesto de frecuencias esperadas mínimas (regla de Cochran: ≥ 5 en $\geq 80\%$ de celdas).
- Decidir entre chi-cuadrado, Fisher exact (`scipy.stats.fisher_exact`, tablas 2×2 con conteos chicos) y chi-cuadrado con simulación (`lambda_='log-likelihood'` o `montecarlo`).
- Calcular Cramér's V como effect size para tablas $r \times c$ (análogo al Cohen's d categórico).
- Aplicar bondad de ajuste con `scipy.stats.chisquare` para validar datos, ruedas de roulette o conteos en bins.

Temas

#	Tema	Por qué importa
1	Tablas de contingencia	Estructura natural de datos categóricos cr
2	Estadístico $\chi^2 = \sum (O - E)^2 / E$	Lo que mide el test: distancia entre obser
3	Grados de libertad $(r-1) \cdot (c-1)$	Determinan la distribución de referencia.
4	Supuesto de $E \geq 5$ (Cochran)	Si se viola, el p-value asintótico es inco
5	Fisher exact para 2×2 con n chico	Alternativa exacta cuando χ^2 no sirve.
6	Cramér's V	Effect size — qué tan fuerte es la asociac
7	Bondad de ajuste vs independencia	Misma fórmula, distinto problema.

Definiciones y características

- Tabla de contingencia: matriz $r \times c$ donde $O_{\{ij\}}$ es la frecuencia observada del cruce (fila i , columna j).
- Frecuencia esperada bajo independencia: $E_{\{ij\}} = (\text{row}_i\text{ total} \cdot \text{col}_j\text{ total}) / n$. Es lo que esperaríamos si las dos variables fueran independientes.
- Estadístico χ^2 : $\sum (O_{\{ij\}} - E_{\{ij\}})^2 / E_{\{ij\}}$. Sigue χ^2 con $(r-1) \cdot (c-1)$ gl si las E son suficientemente grandes.
- Test de independencia: H_0 : las variables categóricas son independientes. Se aplica a una tabla

cruzada de dos variables.

- Test de bondad de ajuste: H_0 : la muestra proviene de una distribución teórica especificada. Las E vienen de la distribución teórica, no del producto de marginales.
- Cochran's rule: el test asintótico es válido si todas las $E \geq 1$ y al menos 80 % de las celdas tienen $E \geq 5$. Si no, usar Fisher (2×2) o Monte Carlo ($> 2 \times 2$).
- Fisher exact test: calcula el p-value exacto enumerando todas las tablas con las mismas marginales. Computacionalmente caro para n grande.
- Cramér's V: $V = \sqrt{\chi^2 / (n \cdot \min(r-1, c-1))}$. Va de 0 a 1. Interpretación cualitativa con $\min(r-1, c-1) = 1$: 0.1 small, 0.3 medium, 0.5 large (Cohen 1988).
- Test de homogeneidad: matemáticamente idéntico al de independencia, pero el diseño muestral es distinto (se fijan los marginales de una variable). El cálculo y la interpretación práctica son los mismos.

Dataset / recursos

- `seaborn.load_dataset('titanic')`: cruzar `survived` × `class` o `survived` × `sex`.
- `seaborn.load_dataset('tips')`: `smoker` × `day`.
- Bondad de ajuste: simular tiradas de un dado posiblemente cargado y testear contra distribución uniforme.
- Librerías: `pandas`, `scipy.stats`, `pingouin` (que tiene `pg.chi2_independence` con Cramér's V incluido).

Ejercicios

1. Tabla cruzada: `pd.crosstab(titanic.survived, titanic['class'])`. Aplicá `chi2`, `p`, `dof`, `expected = scipy.stats.chi2_contingency(tabla)`. Reportá los cuatro valores e interpretá.
2. Effect size: calculá Cramér's V manualmente: $V = \sqrt{\chi^2 / (n * \min(r-1, c-1))}$. Verificá contra `pingouin.chi2_independence(titanic, x='survived', y='class')`.
3. Cochran check: imprimí la matriz `expected` y contá cuántas celdas tienen $E < 5$. Si supera el 20 %, recalculá con `chi2_contingency(tabla, lambda_='log-likelihood')` (G-test, mejor para celdas chicas).
4. Fisher exact (2×2): tomá la subtabla `survived` × `sex` y aplicá `scipy.stats.fisher_exact(tabla_2x2)`. Comparalo con chi-cuadrado.
5. Bondad de ajuste: simulá `rng = np.random.default_rng(7)`; `tiros = rng.choice([1,2,3,4,5,6], size=600, p=[0.18, 0.16, 0.17, 0.17, 0.16, 0.16])`. Hipótesis: el dado es justo (p uniforme). Aplicá `scipy.stats.chisquare(observado, f_exp=esperado)` con `esperado = [100]*6`. ¿Rechazás H_0 ?

Homework verificable

Notebook que sobre `titanic`:

1. Cruza `survived` × `class` y `survived` × `sex` por separado.
2. Para cada cruce: χ^2 , `gl`, p-value, Cramér's V.
3. Identifica cuál de los dos tiene asociación más fuerte (mayor V) y cuál tiene evidencia estadística más fuerte (menor p).
4. En 3 líneas, explica por qué p y V pueden ordenar distinto cuando n cambia entre comparaciones.

Criterio de aceptación: ambos tests deben rechazar H_0 al 5 %; Cramér's V debe ser mayor para `sex` que para `class`; la conclusión debe distinguir tamaño de efecto de evidencia.

Errores comunes

Síntoma / mensaje	Causa y cómo arreglar
expected tiene celdas con valores < 5 y re	El p-value asintótico no es confiable. Fix
Aplico chi ² sobre una columna numérica con	Chi ² es solo para categóricas / conteos. F
Confundo "asociación" con "causalidad" por	Chi ² detecta dependencia estadística, no r
Reporto solo $p < 0.05$ con $n = 10^5$ y declar	Con n gigante, asociaciones triviales son
Bondad de ajuste con f_{exp} proporcional pe	chisquare(obs, f_exp) requiere que f_exp.s

Preguntas frecuentes

¿Cuándo Fisher exact y cuándo chi-cuadrado?

Fisher cuando alguna $E < 5$ en una tabla 2×2 . Para tablas mayores, Fisher es costoso; mejor usar Monte Carlo simulation (`scipy.stats.chi2_contingency(..., method='monte-carlo')` desde `scipy 1.11`) o G-test.

¿Qué es la corrección de Yates?

Para tablas 2×2 , resta 0.5 al $|O - E|$ antes de elevar al cuadrado. Hace el test más conservador. `scipy.stats.chi2_contingency` la aplica por default en 2×2 (`correction=True`). En la práctica, con n moderado, casi no cambia el p-value.

¿Por qué $dof = (r-1) \cdot (c-1)$?

Porque al fijar los totales marginales ($r+c-1$ restricciones), solo $(r-1) \cdot (c-1)$ celdas son libres de variar — las otras quedan determinadas por aritmética.

¿G-test es mejor que chi-cuadrado?

Asintóticamente equivalentes, pero G-test tiene mejor comportamiento con celdas chicas y se generaliza mejor (es el log-likelihood ratio test). En `scipy`: `chi2_contingency(tabla, lambda_='log-likelihood')`.

¿Puedo usar chi-cuadrado para validar la salida de un modelo de clasificación?

Sí — `crosstab(y_true, y_pred)` da la confusion matrix, y un chi² sobre esa tabla testea si la predicción es independiente de la verdad (H: clasificador trivial). Pero ojo: prefieres métricas específicas (accuracy, F1, Kappa de Cohen) — chi² no captura el balance de clases.

Referencias

- ISLP, cap. 4 — Classification, parte sobre datos categóricos.
- Bruce & Bruce, cap. 3 — sección Chi-Square Test.
- Cochran, W.G. (1954), Some Methods for Strengthening the Common χ^2 Tests, Biometrics.
- `scipy.stats.chi2_contingency` y `fisher_exact`.
- `pingouin.chi2_independence` — reporta Cramér's V automáticamente.

Siguiente clase

Clase 179 — ANOVA (one-way, two-way)

Apéndice: notebook (primer bloque)

Primera celda ejecutable del notebook de la clase.

```
# Imports y configuración inicial
```

Archivos complementarios

- notebook.ipynb