

第 13 章 卡方檢定·速查小抄

Chi-Square Tests — 開書考速查

核心：一個統計量、三種檢定、一律右尾

$\chi^2 = \sum \frac{(O-E)^2}{E} \sim \chi_{df}^2$ 。 χ^2 越大 \Rightarrow 觀察 O 與期望 E 差越多 \Rightarrow 越該拒絕 H_0 ；故三種檢定都是**右尾**，拒絕域 $R = \{\chi^2 \geq \chi_{\alpha,df}^2\}$ 。統計量只用**次數** (O, E)，不是比例或百分比。

七步 SOP

認題型 \rightarrow 算期望次數 $E \rightarrow$ 算 $\chi^2 = \sum \frac{(O-E)^2}{E} \rightarrow$ 定自由度 $df \rightarrow$ 查 $\chi_{\alpha,df}^2 \rightarrow \chi^2 \geq$ 臨界值則**拒絕** $H_0 \rightarrow$ 用情境語言下結論。

決策兩法 (等價；卡方一律右尾)

臨界值法： $\chi_{obs}^2 \geq \chi_{\alpha,df}^2$ 就拒絕 H_0 。
p-value 法： $p = P(\chi_{df}^2 \geq \chi_{obs}^2)$ (右尾面積)， $p \leq \alpha$ 就拒絕、 p 愈小愈容易拒絕。兩法同結論： $\chi_{obs}^2 \geq \chi_{\alpha,df}^2 \Leftrightarrow p \leq \alpha$ 。手算 χ^2 算不出精確 p ，用 χ^2 表把 p **夾範圍** (例： χ_{obs}^2 介於 $\chi_{.05,df}^2$ 與 $\chi_{.025,df}^2$ 之間 $\Rightarrow 0.025 < p < 0.05$)。

三種檢定

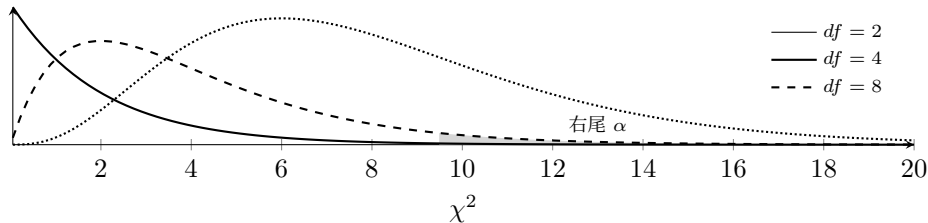
檢定	用途	期望次數 E	自由度 df
適合度	單一母體某類別變數是否符合特定分配	$E_i = n p_{i0}$	$k - 1 - m$
獨立性	同一母體兩屬性是否獨立 (列聯表)	$E_{ij} = \frac{R_i C_j}{n}$	$(r - 1)(c - 1)$
齊一性	多母體分配是否相同 (列聯表)	$E_{ij} = \frac{R_i C_j}{n}$	$(r - 1)(c - 1)$

$m =$ 由樣本估計的參數個數 (機率全給定時 $m = 0$ ，適合度 $df = k - 1$)。獨立性與齊一性算法**完全相同**，只差抽樣設計與結論措辭。

五的規則 & χ^2 分配性質

每格 $E \geq 5$ 卡方近似才可靠；若有格子 $E < 5$ ，把**相鄰類別合併**，合併後 k (或 r, c) 變小、 df 跟著減。 χ^2 分配：只取**非負值**、**形狀右偏**；自由度越大越接近常態；期望值 $E(\chi^2) = df$ 。

χ^2 分配形狀 (示意) 與右尾拒絕域



右偏； df 越大峰越右移、越對稱。檢定恆右尾： χ^2 落入陰影 ($\geq \chi_{\alpha,df}^2$) 即拒絕 H_0 。圖以 $df = 4, \alpha = .05, \chi_{.05,4}^2 = 9.488$ 示意。

χ^2 上尾臨界值表 $\chi_{\alpha,df}^2$ ($P(\chi^2 > \text{值}) = \alpha$ ； $\chi^2 \geq$ 此值即拒絕 H_0)

df	.10	.05	.025	.01	df	.10	.05	.025	.01
1	2.706	3.841	5.024	6.635	16	23.542	26.296	28.845	32.000
2	4.605	5.991	7.378	9.210	17	24.769	27.587	30.191	33.409
3	6.251	7.815	9.348	11.345	18	25.989	28.869	31.526	34.805
4	7.779	9.488	11.143	13.277	19	27.204	30.144	32.852	36.191
5	9.236	11.070	12.833	15.086	20	28.412	31.410	34.170	37.566
6	10.645	12.592	14.449	16.812	21	29.615	32.671	35.479	38.932
7	12.017	14.067	16.013	18.475	22	30.813	33.924	36.781	40.289
8	13.362	15.507	17.535	20.090	23	32.007	35.172	38.076	41.638
9	14.684	16.919	19.023	21.666	24	33.196	36.415	39.364	42.980
10	15.987	18.307	20.483	23.209	25	34.382	37.652	40.646	44.314
11	17.275	19.675	21.920	24.725	26	35.563	38.885	41.923	45.642
12	18.549	21.026	23.337	26.217	27	36.741	40.113	43.195	46.963
13	19.812	22.362	24.736	27.688	28	37.916	41.337	44.461	48.278
14	21.064	23.685	26.119	29.141	29	39.087	42.557	45.722	49.588
15	22.307	24.996	27.488	30.578	30	40.256	43.773	46.979	50.892

常見易錯

- 一律**右尾**，沒有雙尾、左尾。
- df 別記錯：適合度 $k - 1 - m$ 、列聯表 $(r - 1)(c - 1)$ ，不是 $rc - 1$ 。
- 期望次數來源：適合度 $E = n p_{i0}$ ；列聯表 $E_{ij} = R_i C_j / n$ (列總 \times 行總 \div 總和)。
- 統計量用**次數** (O, E)，不是比例。
- 獨立性與齊一性算法相同，但 H_0 措辭不同 (獨立 vs 分配相同)。