

METODO DELLE TABELLE DI CONTINGENZA

Se dobbiamo confrontare più di due campioni, o i campioni possono assumere più di due possibili valori, il test Z è insufficiente, e bisogna usare il metodo delle tabelle di contingenza.

Vediamo nel caso 2×2 (2 campioni $\times$ 2 possibili risultati).
con l'esempio albatro / mafino. Raccogliamo i dati osservati in una tabella:

	mati	non mati	totali
A	8 (9)	53 (52)	61
M	10 (9)	57 (58)	67
	<u>18</u>	110	<u>128</u>

tabelle dei dati osservati:

$$O_{ij} \quad \begin{matrix} i = 1, 2 \\ j = 1, 2 \end{matrix}$$

Se l'albatro e la mafino avessero $p_A = p_M$,
che tabella ci attenderemmo in media?

$$\hat{p} = \frac{8 + 10}{61 + 67} = \frac{18}{128} = 0,141$$

$$\begin{aligned} \text{mati attesi } \times \text{ albatro: } & \hat{p} \times n_1 = 61 \times 0,141 = 8,6 \simeq 9 \\ & \times \text{ mafino: } \hat{p} \times n_2 = 67 \times 0,141 = 9,4 \simeq 9 \end{aligned}$$

tabella dei valori attesi

$$E_{ij} \quad , \quad i = 1, 2 \quad , \quad j = 1, 2$$

mati no tot.

	matr.	no	tot.
A	9	52	61
M	9	58	67
	18	110	128

Se vale $H_0: p_A = p_M$, le due tabelle "sono simili". Costruiamo

$$\sum_{i=1}^2 \sum_{j=1}^2 \frac{(O_{ij} - E_{ij})^2}{E_{ij}} =: \chi^2 \quad (\text{chi quadro})$$

che legge ha?

O_{ij} sono v.al. binomiali, come E_{ij} (x arrotondando agli interi)

Se n_1 e n_2 "grandi", allora O_{ij} e E_{ij} sono approssimabili con normali ... e χ^2 ?

" n_1 grande": $n_1 \hat{p}_x \simeq n_1 \cdot \hat{p} = E_{11} > 5$
 $n_1(1 - \hat{p}_x) \simeq n_1(1 - \hat{p}) = E_{12} > 5$

stesso per " n_2 grande", quindi " n_1 e n_2 grandi" $\leftrightarrow E_{ij} > 5$

$\forall i, j$. Se è così, $O_{ij} - E_{ij}$ è ^(appross.) gaussiana, con media 0.

Dalla teoria della probabilità, se $X_1, \dots, X_n$ i.i.d. $\sim N(0, 1)$ allora la legge di

$$\chi^2 := \sum_{i=1}^n X_i^2$$

si chiama legge del chi quadro con n gradi di libertà.

si chiama legge del quadrato con n gradi di libertà.

La nostra statistica

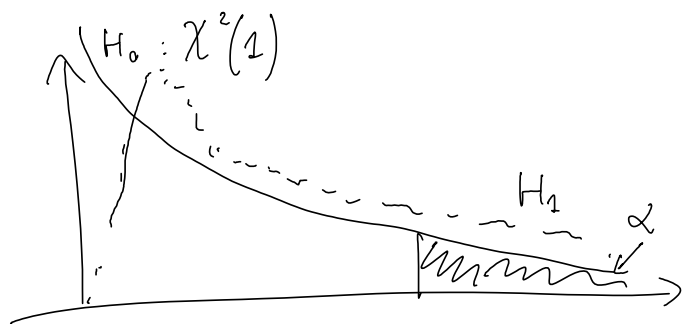
$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$r = \text{n}^\circ \text{ righe } (= 2)$$

$$c = \text{n}^\circ \text{ colonne } (= 2)$$

se $E_{ij} > 5 \quad \forall i, j$, si può approssimare con una χ^2 con
gradi di libertà $\nu = (r-1)(c-1) \quad (= (2-1)(2-1) = 1)$

sotto $H_0: p_X \neq p_Y$, $O_{ij} - E_{ij}$ ha la media zero,
e χ^2 tenderà ad assumere valori "grandi"



regime critica: $[\chi_{1-\alpha}^2(\nu), +\infty)$

regole:

- se $\chi^2 < \chi_{1-\alpha}^2(\nu)$, accetta H_0

- se $\chi^2 > \chi_{1-\alpha}^2(\nu)$, rifiuta H_0 e accetta H_1

nota: il test Z e il test χ^2 producono sempre lo stesso
risultato, e $Z^2 = \chi^2$, e $(q_{1-\frac{\alpha}{2}})^2 = \chi_{1-\alpha}^2(1)$

8.9 Quantili della distribuzione χ^2

$$\mathbb{P}\{\chi^2 \leq x\} = \alpha, \quad \chi^2 \sim \chi^2(\nu)$$

Nella tabella sono riportati vari ν significativi e $\alpha = 0.95, 0.975, 0.99$ e 0.995 .
 Nell'intersezione tra α e ν si legge $x = \chi_\alpha(\nu)$.

ν	0.95	0.975	0.99	0.995	0.999
1	3.841	5.024	6.635	7.879	10.828
2	5.991	7.378	9.210	10.597	13.816
3	7.815	9.348	11.345	12.838	16.266
4	9.488	11.143	13.277	14.860	18.467
5	11.070	12.833	15.086	16.750	20.515
6	12.592	14.449	16.812	18.548	22.458
7	14.067	16.013	18.475	20.278	24.322
8	15.507	17.535	20.090	21.955	26.124
9	16.919	19.023	21.666	23.589	27.877
10	18.307	20.483	23.209	25.188	29.588
11	19.675	21.920	24.725	26.757	31.264
12	21.026	23.337	26.217	28.300	32.909
13	22.362	24.736	27.688	29.819	34.528
14	23.685	26.119	29.141	31.319	36.123
15	24.996	27.488	30.578	32.801	37.697
16	26.296	28.845	32.000	34.267	39.252
17	27.587	30.191	33.409	35.718	40.790
18	28.869	31.526	34.805	37.156	42.312
19	30.144	32.852	36.191	38.582	43.820
20	31.410	34.170	37.566	39.997	45.315
21	32.671	35.479	38.932	41.401	46.797
22	33.924	36.781	40.289	42.796	48.268
23	35.172	38.076	41.638	44.181	49.728
24	36.415	39.364	42.980	45.559	51.179
25	37.652	40.646	44.314	46.928	52.620
26	38.885	41.923	45.642	48.290	54.052
27	40.113	43.195	46.963	49.615	55.476
28	41.337	44.461	48.278	50.993	56.892
29	42.557	45.722	49.588	52.336	58.301
30	43.773	46.979	50.892	53.672	59.703
31	44.985	48.232	52.191	55.003	61.098
32	46.194	49.480	53.486	56.328	62.487
33	47.400	50.725	54.776	57.648	63.870
34	48.602	51.966	56.061	58.964	65.247
35	49.802	53.203	57.342	60.275	66.619
36	50.998	54.437	58.619	61.581	67.985
37	52.192	55.668	59.893	62.883	69.346
38	53.384	56.896	61.162	64.181	70.703
39	54.572	58.120	62.428	65.476	72.055
40	55.758	59.342	63.691	66.766	73.402
41	56.942	60.561	64.950	68.053	74.745
42	58.124	61.777	66.206	69.336	76.084
43	59.304	62.990	67.459	70.616	77.419
44	60.481	64.201	68.610	71.893	78.750
45	61.656	65.410	69.857	73.166	80.077
46	62.830	66.617	71.201	74.437	81.400
47	64.001	67.821	72.443	75.704	82.720
48	65.171	69.023	73.683	76.969	84.037
49	66.339	70.222	74.919	78.231	85.351
50	67.505	71.420	76.154	79.490	86.661

Nel nostro caso:

$$\begin{array}{ll} 8 (8,6) & 53 (52,4) \\ 10 (9,4) & 52 (52,6) \end{array}$$

$$E_{ij} > 5 \quad \forall i, j \Rightarrow \text{test } \chi^2$$

$$\chi^2 = \frac{(8 - 8,6)^2}{8,6} + \frac{(53 - 52,4)^2}{52,4} + \frac{(10 - 9,4)^2}{9,4} + \frac{(52 - 52,6)^2}{52,6}$$

$$= 0,6^2 \left(\frac{1}{8,6} + \frac{1}{52,4} + \frac{1}{9,4} + \frac{1}{52,6} \right) = 0,09$$

no rifiutiamo $\alpha = 0.05$ il valore critico è $\chi_{0.05}^2(1) = 3,841$

$$\chi^2 < 3,841 \Rightarrow \text{accetta } H_0: p_X = p_Y$$

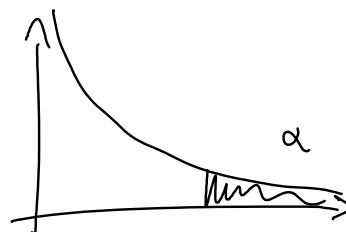
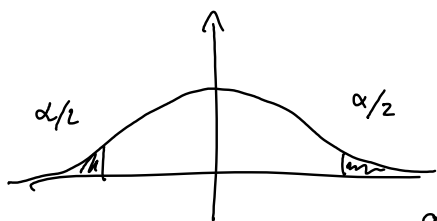
Possiamo anche ripartire $P > 0,05$

nota: nel caso 2×2 , il test Z e il test χ^2 sono equivalenti:
e producono gli stessi risultati. Difatti:

$$Z^2 = \chi^2 \approx \chi^2(1), \text{ che è il quadrato di una } N(0,1)$$

n.c. di test Z

n.c. di test χ^2



$$\begin{aligned} 1 - \alpha &= P\{|Z| < \text{val. c.}\} \\ &= P\{|Z|^2 < (\text{val. c.})^2\} \\ &= P\{Z^2 \leq q_{1-\frac{\alpha}{2}}^2\} \end{aligned}$$

$$\longrightarrow = P\{\chi^2 < \chi_{1-\alpha}^2(1)\} = 1 - \alpha$$

e i valori critici sono b.c. $q_{1-\frac{\alpha}{2}}^2 = \chi_{1-\alpha}^2(1)$