Metodi statistici per lo studio dei fenomeni biologici

Alla fine di questa lezione dovreste essere in grado di:

- descrivere la distribuzione di campionamento della differenza di due proporzioni in campioni indipendenti
- costruire gli intervalli di confidenza delle differenza di due proporzioni in campioni indipendenti
- condurre un test di significatività per confrontare due proporzioni in campioni indipendenti
- utilizzare il test del chi-quadrato per valutare l'associazione di due variabili

Esempio

	Trattamento (%)		
Morte	**Alotano**	**Morfina**	**Totale**
SI	8 (13,1)	10 (14,9)	18 (14,1)
NO	53 (86,9)	57 (85,1)	110 (85,9)
Totale	61	67	128

C'è differenza nella mortalità post-operatoria fra i due anestetici?

Confronto di due proporzioni

1 Ipotesi nulla

H_0: $\pi_a = \pi_m$ oppure $\pi_a - \pi_m = 0$

Indipendenza!

Ipotesi alternativa

H_1: $\pi_a \neq \pi_m$ oppure $\pi_a - \pi_m \neq 0$

2 Distribuzione di campionamento

Se la numerosità è sufficiente, la distribuzione delle differenze campionarie di due proporzioni, subordinatamente all'ipotesi nulla:

- È approssimativamente gaussiana
- La media della distribuzione è 0
- L'errore standard è uguale a:

$$\sqrt{\frac{p(1-p)}{n_a}+\frac{p(1-p)}{n_m}}$$

$np>5$

$n(1-p) > 5$

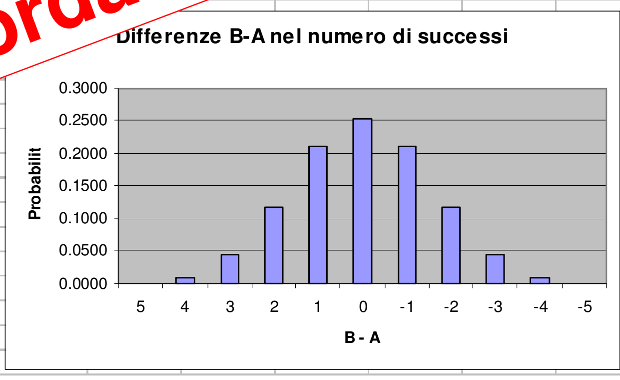
Altrimenti: binomiale

Ricordate?

n = 5; pT = 0,45; pC = 0,45

successi	A	B	A	B: 0	B: 1	B: 2	B: 3	B: 4	B: 5
0	0.05	0.05	0	0.0025	0.0105	0.0170	0.0140	0.0055	0.0010
1	0.21	0.21	1	0.0105	0.0441	0.0714	0.0588	0.0231	0.0042
2	0.34	0.34	2	0.0170	[illegible]	0.1156	0.0952	0.0374	0.0068
3	0.28	0.28	3	[illegible]	[illegible]	0.0952	0.0784	0.0308	0.0056
4	0.11	0.11	4	[illegible]	[illegible]	0.0374	0.0308	0.0121	0.0022
5	0.02	0.02	[illegible]	[illegible]	[illegible]	0.0068	0.0056	0.0022	0.0004

B-A	
5	[illegible]
4	[illegible]
3	[illegible]
2	0.1188
1	0.2101
0	0.2531
-1	0.2101
-2	0.1188
-3	0.0439
-4	0.0097
-5	0.0010



Confronto di due proporzioni

$\boxed{3}$ Test statistico

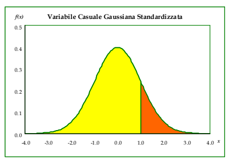
$$z = \frac{|p_a - p_m| - (\overset{H_0}{\cancel{\pi_a - \pi_m}})}{\sqrt{\frac{p(1-p)}{n_a} + \frac{p(1-p)}{n_m}}}$$

$$z = \frac{|0{,}131 - 0{,}149|}{\sqrt{\frac{0{,}141 \cdot 0{,}859}{61} + \frac{0{,}141 \cdot 0{,}859}{67}}} = 0{,}292$$

P (|z| > 0,292) = P(|p_a-p_m|>0,018) = 0.770

La tavola della distribuzione Gaussiana Standardizzata

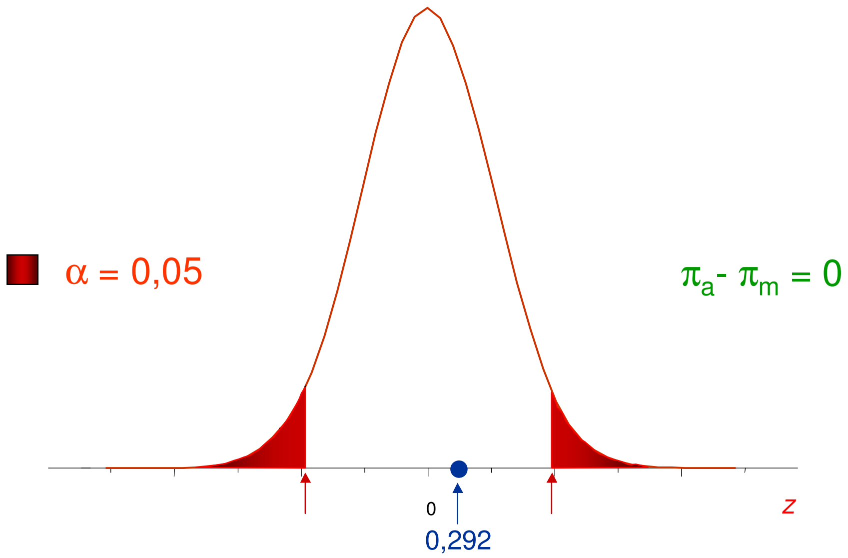
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500	0.496	0.492	0.488	0.484	0.480	0.476	0.472	0.468	0.464
0.1	0.460	0.456	0.452	0.448	0.444	0.440	0.436	0.433	0.429	0.425
0.2	0.421	0.417	0.413	0.409	0.405	0.401	0.397	0.394	0.390	0.386
0.3	0.382	0.378	0.374	0.371	0.367	0.363	0.359	0.356	0.352	0.348
0.4	0.345	0.341	0.337	0.334	0.330	0.326	0.323	0.319	0.316	0.312
0.5	0.309	0.305	0.302	0.298	0.295	0.291	0.288	0.284	0.281	0.278
0.6	0.274	0.271	0.268	0.264	0.261	0.258	0.255	0.251	0.248	0.245
0.7	0.242	0.239	0.236	0.233	0.230	0.227	0.224	0.221	0.218	0.215
0.8	0.212	0.209	0.206	0.203	0.200	0.198	0.195	0.192	0.189	0.187
0.9	0.184	0.181	0.179	0.176	0.174	0.171	0.169	0.166	0.164	0.161
1.0	0.159	0.156	0.154	0.152	0.149	0.147	0.145	0.142	0.140	0.138
1.1	0.136	0.133	0.131	0.129	0.127	0.125	0.123	0.121	0.119	0.117
1.2	0.115	0.113	0.111	0.109	0.107	0.106	0.104	0.102	0.100	0.099
1.3	0.097	0.095	0.093	0.092	0.090	0.089	0.087	0.085	0.084	0.082
1.4	0.081	0.079	0.078	0.076	0.075	0.074	0.072	0.071	0.069	0.068
1.5	0.067	0.066	0.064	0.063	0.062	0.061	0.059	0.058	0.057	0.056
1.6	0.055	0.054	0.053	0.052	0.051	0.049	0.048	0.048	0.046	0.046
1.7	0.045	0.044	0.043	0.042	0.041	0.040	0.039	0.038	0.037	0.037
1.8	0.036	0.035	0.034	0.034	0.033	0.032	0.031	0.030	0.029	0.029
1.9	0.029	0.028	0.027	0.027	0.026	0.026	0.025	0.024	0.024	0.023
2.0	0.023	0.022	0.022	0.021	0.021	0.020	0.020	0.019	0.019	0.018
2.1	0.018	0.017	0.017	0.017	0.016	0.016	0.015	0.015	0.015	0.014
2.2	0.014	0.014	0.013	0.013	0.013	0.012	0.012	0.012	0.011	0.011
2.3	0.011	0.010	0.010	0.010	0.010	0.009	0.009	0.009	0.009	0.008
2.4	0.008	0.008	0.008	0.008	0.007	0.007	0.007	0.007	0.007	0.006
2.5	0.006	0.006	0.006	0.006	0.006	0.005	0.005	0.005	0.005	0.005
2.6	0.005	0.005	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
2.7	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
2.8	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
2.9	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.001	0.001	0.001



P(p_a-p_m<-0,018) = 0.385

P(|p_a-p_m|>0,018) = 0.770

Statisticamente non significativo



Confronto di due proporzioni

3 Test statistico

$$z = \frac{|p_a - p_m| - \overset{H_0}{\cancel{(\pi_a - \pi_m)}}}{\sqrt{\frac{p(1-p)}{n_a} + \frac{p(1-p)}{n_m}}}$$

$$z = \frac{|0{,}131 - 0{,}149|}{\sqrt{\frac{0{,}141 \cdot 0{,}859}{61} + \frac{0{,}141 \cdot 0{,}859}{67}}} = 0{,}292$$

P $(|z| > 0{,}292) = P(|p_a\text{-}p_m| > 0{,}018) = 0.770$ *(p = 0.770)*

4 Conclusioni

Il test non è statisticamente significativo

Non posso concludere che i due trattamenti danno mortalità diverse

Esempio: intervallo di confidenza 95%

$$p_a - p_m \pm z_{0,95}\sqrt{\frac{p_a(1-p_a)}{n_a}+\frac{p_m(1-p_m)}{n_m}}$$

$$-0{,}018 \pm 1{,}96\sqrt{\frac{0{,}131\cdot 0{,}869}{61}+\frac{0{,}149\cdot 0{,}851}{67}}=$$

= -7,9%; +10,2%

Il test Chi-quadrato

C'è differenza fra i sessi nella distribuzione dei gruppi sanguigni?

Gruppo	Sesso: Femminile	Sesso: Maschile	Totale
O	52	49	101
A	30	20	50
B	5	4	9
AB	16	12	28
Totale	103	85	188

1 **H_o : il sesso ed il gruppo sanguigno sono indipendenti**

Il test Chi-quadrato

2

Se sesso e gruppo sanguigno fossero indipendenti, quanti studenti dovremmo aspettarci per ogni combinazione di caratteri? → frequenze attese

Gruppo	Sesso		
	Femminile	Maschile	Totale
O	55,3	45,7	101
A	27,4	22,6	50
B	4,9	4,1	9
AB	15,3	12,7	28
Totale	103	85	188

Es: $P(M \cap 0) = (101/188) * (103/188) * 188 = 45{,}7$

Il test Chi-quadrato

3

Si confrontano le frequenze osservate e attese in ciascuna categoria della tabella di contingenza

(O - E)

Gruppo	Sesso		
	Femminile	Maschile	Totale
O	-3.34	+3.34	
A	+2.60	-2.60	
B	+0.07	-0.07	
AB	+0.66	-0.66	
Totale			

$(O - E)^2$

Gruppo	Sesso		
	Femminile	Maschile	Totale
O	11.16	11.16	
A	6.76	6.76	
B	.005	.005	
AB	.44	.44	
Totale			

$$\chi^2 = \sum \frac{\boxed{(O-E)^2}}{E}$$

Il test Chi-quadrato

3 **Si confrontano le frequenze osservate e attese in ciascuna categoria della tabella di contingenza**

Gruppo	Sesso: Femminile	Sesso: Maschile	Totale
O	0,20	0,24	
A	0,25	0,30	
B	0,00	0,00	
AB	0,03	0,03	
Totale			

$$\chi^2 = \sum \boxed{\frac{(O-E)^2}{E}}$$

$\chi^2 = 1{,}06$

Gradi di libertà = 3

P =

Il test Chi-quadrato

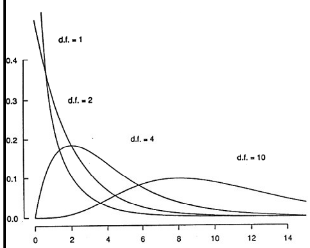


Figure 4.13. Chi-square distributions.

Probabilità di Chi-Quadrato

	Probabilità									
df	0,95	0,90	0,70	0,50	0,30	0,20	0,10	0,05	0,01	0,001
1	0,004	0,016	0,15	0,46	1,07	1,64	2,71	3,84	6,64	10,83
2	0,10	0,21	0,71	1,39	2,41	3,22	4,61	5,99	9,21	13,82
3	0,35	0,58	1,42	2,37	3,67	4,64	6,25	7,82	11,35	16,27
4	0,71	1,06	2,20	3,36	4,88	5,99	7,78	9,49	13,28	18,47
5	1,15	1,61	3,00	4,35	6,06	7,29	9,24	11,07	15,09	20,52
6	1,64	2,20	3,83	5,35	7,23	8,56	10,65	12,59	16,81	22,46
7	2,17	2,83	4,67	6,35	8,38	9,80	12,02	14,07	18,48	24,32
8	2,73	3,49	5,53	7,34	9,52	11,03	13,36	15,51	20,09	26,13
9	3,33	4,17	6,39	8,34	10,66	12,24	14,68	16,92	21,67	27,88
10	3,94	4,87	7,27	9,34	11,78	13,44	15,99	18,31	23,21	29,59
11	4,58	5,58	8,15	10,34	12,90	14,63	17,28	19,68	24,73	31,26
12	5,23	6,30	9,03	11,34	14,01	15,81	18,55	21,03	26,22	32,91
13	5,89	7,04	9,93	12,34	15,12	16,99	19,81	22,36	27,69	34,53
14	6,57	7,79	10,82	13,34	16,22	18,15	21,06	23,69	29,14	36,12
15	7,26	8,55	11,72	14,34	17,32	19,31	22,31	25,00	30,58	37,70
20	10,85	12,44	16,27	19,34	22,78	25,04	28,41	31,41	37,57	45,32
25	14,61	16,47	20,87	24,34	28,17	30,68	34,38	37,65	44,31	52,62
30	18,49	20,60	25,51	29,34	33,53	36,25	40,26	43,77	50,89	59,70
50	34,76	37,69	44,31	49,34	54,72	58,16	63,17	67,51	76,15	86,66

← Accettare | Rifiutare →
al livello di 0,05

Il test Chi-quadrato

3 **Si confrontano le frequenze osservate e attese in ciascuna categoria della tabella di contingenza**

$\chi^2 = 1{,}06$ Gradi di libertà = 3 $0{,}70 < P < 0{,}90$

Test statisticamente non significativo

4 **Si conclude che non ci sono prove sufficienti per confutare l'ipotesi nulla che le due variabili siano indipendenti**

Esempio

	Trattamento (%)		
Morte	Alotano	Morfina	Totale
SI	8 (13,1)	10 (14,9)	18 (14,1)
NO	53 (86,9)	57 (85,1)	110 (85,9)
Totale	61	67	128

C'è differenza nella mortalità post-operatoria fra i due anestetici?

$\chi_1^2 = 1{,}39$ Gradi di libertà = 1 $0{,}70 < P < 0{,}90$

Il test Chi-quadrato

Probabilità di Chi-Quadrato

	Probabilità									
df	0,95	0,90	0,70	0,50	0,30	0,20	0,10	0,05	0,01	0,001
1	0,004	0,016	0,15	0,46	1,07	1,64	2,71	3,84	6,64	10,83
2	0,10	0,21	0,71	1,39	2,41	3,22	4,61	5,99	9,21	13,82
3	0,35	0,58	1,42	2,37	3,67	4,64	6,25	7,82	11,35	16,27
4	0,71	1,06	2,20	3,36	4,88	5,99	7,78	9,49	13,28	18,47
5	1,15	1,61	3,00	4,35	6,06	7,29	9,24	11,07	15,09	20,52
6	1,64	2,20	3,83	5,35	7,23	8,56	10,65	12,59	16,81	22,46
7	2,17	2,83	4,67	6,35	8,38	9,80	12,02	14,07	18,48	24,32
8	2,73	3,49	5,53	7,34	9,52	11,03	13,36	15,51	20,09	26,13
9	3,33	4,17	6,39	8,34	10,66	12,24	14,68	16,92	21,67	27,88
10	3,94	4,87	7,27	9,34	11,78	13,44	15,99	18,31	23,21	29,59
11	4,58	5,58	8,15	10,34	12,90	14,63	17,28	19,68	24,73	31,26
12	5,23	6,30	9,03	11,34	14,01	15,81	18,55	21,03	26,22	32,91
13	5,89	7,04	9,93	12,34	15,12	16,99	19,81	22,36	27,69	34,53
14	6,57	7,79	10,82	13,34	16,22	18,15	21,06	23,69	29,14	36,12
15	7,26	8,55	11,72	14,34	17,32	19,31	22,31	25,00	30,58	37,70
20	10,85	12,44	16,27	19,34	22,78	25,04	28,41	31,41	37,57	45,32
25	14,61	16,47	20,87	24,34	28,17	30,68	34,38	37,65	44,31	52,62
30	18,49	20,60	25,51	29,34	33,53	36,25	40,26	43,77	50,89	59,70
50	34,76	37,69	44,31	49,34	54,72	58,16	63,17	67,51	76,15	86,66

Accettare | Rifiutare
al livello di 0,05

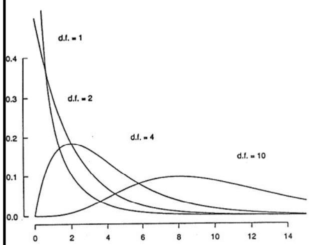


Figure 4.13. Chi-square distributions.