

1

Anemia

emoglobina	n
9 - 12	46
12 - 13	19
13 - 14	59
14 - 17	16
tot	140

Data la distribuzione dell'emoglobina in un gruppo di pazienti maschi sottoposti a trattamento:

- Circa un paziente su 3 era fortemente anemico (emogl. meno di 12)

- La mediana era fra 13 e 14

- Il primo terzile era fra 9 e 12

- La media aritmetica era fra 13 e 14

- La distribuzione era fortemente simmetrica

- Nel verificare che il trattamento induce anemia ($Hgb < 13$), la media è risultata statisticamente significativa (assumendo che la varianza fosse nota e pari a 5)

Anemia - percentuali

emoglobina	n	%
9 - 12	46	33
12 - 13	19	14
13 - 14	59	42
14 - 17	16	11
tot	140	

Data la distribuzione dell'emoglobina in un gruppo di pazienti maschi sottoposti a trattamento:

- Circa un paziente su 3 era fortemente anemico (emogl. meno di 12)

- La mediana era fra 13 e 14

- Il primo terzile era fra 9 e 12

- La media aritmetica era fra 13 e 14

- La distribuzione era fortemente simmetrica

- Nel verificare che il trattamento induce anemia ($Hgb < 13$), la media è risultata statisticamente significativa (assumendo che la varianza fosse nota e pari a 5)

Anemia - quantili

emoglobina	n	%
9 - 12	46	33
12 - 13	19	14
13 - 14	59	42
14 - 17	16	11
tot	140	

Data la distribuzione dell'emoglobina in un gruppo di pazienti maschi sottoposti a trattamento:

- Circa un paziente su 3 era fortemente anemico (emogl. meno di 12)

- La mediana era fra 13 e 14

- Il primo terzile era fra 9 e 12

- La media aritmetica era fra 13 e 14

- La distribuzione era fortemente simmetrica

- Nel verificare che il trattamento induce anemia ($Hgb < 13$), la media è risultata statisticamente significativa (assumendo che la varianza fosse nota e pari a 5)

Anemia - quantili

emoglobina	n	%	% cum
9 - 12	46	33	33
12 - 13	19	14	47
13 - 14	59	42	89
14 - 17	16	11	100
tot	140		



Anemia

emoglobina	n
9 - 12	46
12 - 13	19
13 - 14	59
14 - 17	16
tot	140

Data la distribuzione dell'emoglobina in un gruppo di pazienti maschi sottoposti a trattamento:

- Circa un paziente su 3 era fortemente anemico (emogl. meno di 12)

- La mediana era fra 13 e 14

- Il primo terzile era fra 9 e 12

- La media aritmetica era fra 13 e 14

- La distribuzione era fortemente simmetrica

- Nel verificare che il trattamento induce anemia ($Hgb < 13$), la media è risultata statisticamente significativa (assumendo che la varianza fosse nota e pari a 5)

Anemia – calcolo della media

emoglobina	n	valore centrale (x)	x·n
9 - 12	46	10.5	483
12 - 13	19	12.5	237.5
13 - 14	59	13.5	796.5
14 - 17	16	15.5	248
tot	140		1765

$$1765 / 140 = 12.6$$

Anemia – test sulla media

emoglobina	n
9 - 12	46
12 - 13	19
13 - 14	59
14 - 17	16
tot	140

- Nel verificare che il trattamento induce anemia ($Hgb < 13$), la media è risultata statisticamente significativa (assumendo che la varianza fosse nota e pari a 5)

- media del campione = 12.6

$H_0: \mu = 13$ $H_1: \mu < 13$

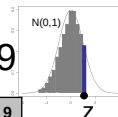
T-test su una media

-Campione grande, varianza nota: ok

$$t = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{12.6 - 13}{\sqrt{5} / \sqrt{140}} = -2.117$$

Cerchiamo $2.117 \approx 2.12$ sulle

Tavole della Normale (0,1), z da 2.00 a 3.99



z	0	1	2	3	4	5	6	7	8	9
2.0	0.977	0.978	0.978	0.979	0.979	0.980	0.980	0.981	0.981	0.982
2.1	0.982	0.983	0.983	0.983	0.984	0.984	0.985	0.985	0.985	0.986
2.2	0.986	0.986	0.987	0.987	0.987	0.988	0.988	0.988	0.989	0.989
2.3	0.989	0.990	0.990	0.990	0.990	0.991	0.991	0.991	0.991	0.992
2.4	0.992	0.992	0.992	0.992	0.993	0.993	0.993	0.993	0.993	0.994
2.5	0.994	0.994	0.994	0.994	0.994	0.994	0.994	0.994	0.994	0.994
2.6	0.995	0.995	0.995	0.995	0.995	0.995	0.995	0.995	0.995	0.995
2.7	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997
2.8	0.997	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998
2.9	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.999	0.999	0.999
3.0	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999
3.1	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999
3.2	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999
3.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3.4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3.6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3.7	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3.8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3.9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

$\Phi(2.117) = 0.983$

p - value = $(1 - 0.983) = 0.017$

Anemia

emoglobina	n
9 - 12	46
12 - 13	19
13 - 14	59
14 - 17	16
tot	140

Data la distribuzione dell'emoglobina in un gruppo di pazienti maschi sottoposti a trattamento:

- ✓ - Circa un paziente su 3 era fortemente anemico (emogl. meno di 12)
- ✓ - La mediana era fra 13 e 14
- ✓ - Il primo terzile era fra 9 e 12
- La media aritmetica era fra 13 e 14
- La distribuzione era fortemente simmetrica
- ✓ - Nel verificare che il trattamento induce anemia (Hgb<13), la media è risultata statisticamente significativa (assumendo che la varianza fosse nota e pari a 5)

Anemia

- Nel verificare che il trattamento induce anemia (Hgb<13), la media è risultata statisticamente significativa (assumendo che la varianza fosse nota e pari a 5): **VERO**

Se il test viene fatto a due code ...

- la significatività aumenta
- la media osservata rimane significativa al livello del 5% ✓

$$\Phi(2.12) = 0.983$$

$$p\text{-value} = 2 \cdot (1 - 0.983) = 2 \cdot 0.017 = 0.034$$

Anemia

- Nel verificare che il trattamento induce anemia con un test a due code, la media (12.6) è risultata statisticamente significativamente diversa da 13 al livello del 5%

L'intervallo di confidenza al 95% ... $\left(z_{\frac{\alpha}{2}} = 1.96 \right)$

- contiene il valore 13

- contiene il valore 12.6 ✓

$$\begin{aligned} 95\%CI &= (\bar{x} - 1.96 \cdot \sigma / \sqrt{n}, \bar{x} + 1.96 \cdot \sigma / \sqrt{n}) \\ &= (12.6 - 1.96 \cdot \sqrt{5} / \sqrt{140}, 12.6 + 1.96 \cdot \sqrt{5} / \sqrt{140}) \\ &= (12.23, 12.97) \end{aligned}$$

L'IC ottenuto vuol dire che:

- nella popolazione l'emoglobina è compresa fra 12.23 e 12.97
- preso a caso un individuo della popolazione, ci aspettiamo un valore dell'emoglobina compreso fra 12.23 e 12.97 ✓
- ... con una variabilità espressa dalla deviazione standard della popolazione (varianza=5)

2

Risposta

In un ospedale, un gruppo di 40 pazienti prende il farmaco A, e un gruppo di 35 pazienti prende il farmaco B. Si osservano rispettivamente 32 Risposte con A e 19 con B.

- A è più efficace di B di quasi il 70% ($32/19 = 1.68$)
- A è più efficace di B: prob(Risposta) è 80% contro 54.3% ✓
- A è più efficace di B di quasi il 50% ($0.8/0.54 = 1.47$) ✓

- La differenza osservata risulta significativa al livello del 5%
- La differenza osservata risulta significativa al livello del 1%

$$P_A = 32/40 = 0.8$$

$$P_B = 19/35 = 0.543$$

Test del chi-Quadrato

H0: $X^2 = 0$ (valore di riferimento: assenza di associazione)

H1: $X^2 > 0$ (siccome X^2 non può valere meno di 0, il test è sempre unilaterale)

Risposta: il test del Chi-Quadrato

Mettiamo i dati in tabella

farmaco	Risposta		tot
	No	Si	
A	8	32	40
B	16	19	35
tot	24	51	75

Dobbiamo calcolare le frequenze attese sotto H_0 : tot riga x tot col / tot

12.80	27.20
11.20	23.80

Calcolo della statistica test:

Somma delle quantità: $(\text{osservata} - \text{attesa})^2 / \text{attesa}$

$$\chi^2 = 5.67$$

La regione di rifiuto cambia a seconda del livello di significatività scelto:

- la soglia per il test al livello $\alpha=5\%$ è 3.841 $\rightarrow$ Rifiutiamo H_0
- al livello (più "cauto") del 1% è 6.635 $\rightarrow$ Accettiamo H_0 . L'evidenza di una associazione c'è, ma non è fortissima ...

Risposta

In un ospedale, un gruppo di 40 pazienti prende il farmaco A, e un gruppo di 35 pazienti prende il farmaco B. Si osservano rispettivamente 32 Risposte con A e 19 con B.

- A è più efficace di B di quasi il 70% ($32/19 = 1.68$)
- A è più efficace di B: prob(Risposta) è 80% contro 54.3% ✓
- A è più efficace di B di quasi il 50% ($0.8/0.54 = 1.47$) ✓
- La differenza osservata risulta significativa al livello del 5% ✓
- La differenza osservata risulta significativa al livello del 1%

$$P_A = 32/40 = 0.8$$

$$P_B = 19/35 = 0.543$$

3

Relazione dose – colesterolo (ed età)

In un ospedale, si vuole verificare se aumentando la dose somministrata di un certo farmaco in pazienti con colesterolo alto, si riesce a ridurlo più efficacemente. Si investiga quindi la relazione (lineare) fra dose (mg) e livello del colesterolo (mg/dl). Risulta $\hat{\beta} = -0.4$ ($p = 0.003$)

Qualcuno osserva anche che è noto che all'aumentare dell'età il colesterolo tende ad aumentare.

Come interpretiamo la nostra analisi?

Relazione dose – colesterolo (ed età)

$\hat{\beta} = -0.4$ ($p = 0.003$) all'aumentare dell'età il colesterolo tende ad aumentare

- Sul campione, per ogni mg in più di farmaco si è osservata una riduzione del colesterolo di 0.4 mg/dl ✓
- La relazione lineare fra dose e colesterolo è negativa (decrescente) e non forte (non è vicina a -1)
- Quando la dose non influenza il colesterolo, c'è una prob. dello 0.3% di osservare nel campione una relazione decrescente con una pendenza della retta pari a -0.4 ✓
- Osservando nel campione una relazione decrescente con una pendenza della retta pari a -0.4 abbiamo una prob. dello 0.3% che tuttavia nella popolazione la dose non influenzi il colesterolo
- Osservando nel campione una relazione decrescente con una pendenza della retta pari a -0.4 abbiamo una prob. quasi del 100% (100% meno 0.3%) che nella popolazione un aumento di 1 mg di dose faccia diminuire il colesterolo di 0.4 mg/dl
- L'età è sicuramente un confondente della relazione dose – colesterolo, quindi questa analisi è distorta (biased)
- Se i soggetti più anziani hanno ricevuto dosi inferiori di farmaco, questa analisi potrebbe essere distorta (biased) a causa del confondimento ✓
- Sapendo che le dosi non risultano associate statisticamente con le età dei soggetti, l'analisi non è affetta da bias da confondimento ✓
- Sapendo che i soggetti più anziani hanno ricevuto dosi superiori di farmaco, l'analisi andrebbe "aggiustata" per età per avere una stima più precisa del β , che potrebbe essere inferiore a -0.4 ✓

4

Anemia: confronto fra trattamenti

In un ospedale, un gruppo di 40 pazienti prende il farmaco A, e un gruppo di 35 pazienti prende il farmaco B. Si è visto che nell'induzione di Risposta A è più efficace di B di quasi il 50% (RR = 1.47). Si vuole ora guardare il livello di emoglobina post-trattamento. I dati campionari sono:

$$\bar{x}_A = 10.2 \quad s_A = 3.1$$

$$\bar{x}_B = 9.8 \quad s_B = 3.0$$

- La media complessiva è esattamente pari a 10
- La variabilità è leggermente superiore nel gruppo A
- La variabilità è leggermente superiore nel gruppo B ✓
- Per valutare la significatività della differenza di emoglobina fra A e B, bisogna usare un T-test per dati appaiati
- Sapendo che risulta p-value=0.57: diciamo che siamo al limite della significatività statistica convenzionale

Attenzione!

Media complessiva: fare la media ponderata

Confronto della variabilità: meglio in termini relativi, cioè tramite il coefficiente di variazione

Dati appaiati?: No, i due gruppi sono distinti, il fatto che l'emoglobina è post-trattamento non ha rilevanza (sarebbero appaiati se stessimo valutando la differenza fra il valore prima (A) e il valore dopo (B) il trattamento sullo stesso gruppo di pazienti).

gruppo	n	medie	std	Ammontari totali: media x n	cv = std / media (x100)
A	40	10.2	3.1	408	30.39
B	35	9.8	3	343	30.61
				751	

$$\text{Media} = 751 / 75 = 10.1$$

Svolgiamo il test

Ipotesi e tipo di Test: $H_0: \mu_A - \mu_B = 0$ vs $H_1: \mu_A - \mu_B \neq 0$; T-test

Ipotesi del test: assumiamo varianze uguali nelle popolazioni, campioni grandi: soddisfatte.

Calcolo della statistica test:

$$s = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{(40-1) \cdot 3.1 + (35-1) \cdot 3}{40 + 35 - 2}} = 3.05$$

$$t = \frac{\bar{x}_A - \bar{x}_B}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{0.4}{3.05 \sqrt{\frac{1}{40} + \frac{1}{35}}} = 0.57$$

Se vogliamo usare il metodo delle regioni di rifiuto: le soglie per un test bilaterale al livello $\alpha=5\%$ sono ± 1.96 . La nostra t cade all'interno cioè in zona di accettazione.

Calcoliamo anche il p-value:

$$\Phi(0.57) = 0.716 \rightarrow p\text{-value} = 2 \cdot (1 - 0.716) = 0.570$$

Conclusione: il p-value è molto alto, accettiamo l'ipotesi nulla: non c'è differenza fra i due gruppi.