

Q

Friday, 16 June 2023

13:32

1) Abbiamo un campione casuale di 100 assenze di studenti:

Giorno	Lun	Mar	Mer	Giov	Ven
Freq	27	19	13	15	26

Verificare l'ipotesi che un'assenza abbia la stessa probabilità di capitare in uno qualunque dei giorni della settimana.

Si usa tquadro di adattamento alla distribuzione $\pi = (t(1), \dots, t(5)) = (\frac{1}{5}, \dots, \frac{1}{5})$

Aka uniforme discreta nei cinque giorni

f_1, \dots, f_5 = frequenze assolute attese

$n = 100$ dati

$$f_1 = n * t(i) = 100 * \frac{1}{5} = 20 = \pi$$

h_0 : popolazione distrib. π

$$Q = \sum \frac{(N_j - f_j)^2}{f_j} \sim \chi^2(5 - 1)$$

$$RC \rightarrow q = \sum \frac{(n_k - f_j)^2}{f_k} > \chi^2_{4,\alpha}$$

$$\alpha = 5\% = 0.05$$

Si fanno i conti

$$q = \frac{(27 - 20)^2}{20} + \frac{(19 - 20)^2}{20} + \dots + \frac{(26 - 20)^2}{20} = 8$$

$$\chi^2_{4,0.05} = 9.49$$

$$8 > 9.49?$$

No quindi non posso rifiutare a livello 5%

Quindi i dati non sono in contraddizione significativa tra di loro

Quindi accetto h_0

Troviamo $\bar{\alpha}$

$$\chi^2_{\bar{\alpha}} = 8 \rightarrow 0.05 < \bar{\alpha} < 0.1$$

- 2) Il numero di difetti in un macchinario prodotto segue una Poisson $\lambda = 1.76$ per 50 macchinari.

Difetti	0	1	2	3	4	5+
Freq ass.	10	14	15	5	3	3

E' ancora affidabile la stima?

Dobbiamo prima calcolare densità discreta

$$P(Y = k) = \frac{e^{-\lambda}(\lambda)^k}{k!}$$

$$n = 50$$

$$\pi = (\pi(0) \dots \pi(6))$$

$$\pi(0) = e^{-1.76} * \frac{(1.76)^0}{0!} = 0.172$$

$$\pi(1) = e^{-1.76} * \frac{(1.76)^1}{1!} = 0.303$$

$$\pi(2) = 0.266$$

$$\pi(3) = 0.156$$

$$\pi(4) = 0.069$$

$$\pi(5) = 1 - P(X \leq 4) = 0.034$$

Ora calcoliamo frequenze assolute attese

$$f_i = n * \pi(i) = 50 * 0.172 = 8.6$$

$$f_1 = 15.15$$

$$f_2 = 13.3$$

$$f_3 = 7.8$$

$$f_4 = 3.45$$

$$f_5 = 1.7$$

Qui non sono tutte ≥ 5

Quindi accorpriamo f_4 e f_5

$$\pi(4) = P(Y \geq 4) = 0.103$$

$$f_4 = 5.15$$

Quindi ora possiamo applicare la formula

$$q = \sum \frac{(n_j - f_j)^2}{f_j} > X_{4,\alpha}^2$$

$$X_{4,0.05}^2 = 9.49$$

$$q = \frac{(10 - 8.6)^2}{8.6} + \dots + \frac{(6 - 5.15)^2}{5.15}$$

$$1.677 < 9.49$$

... ..

Quindi non posso rifiutare h_0

Quindi i dati non sono in contraddizione significativa in h_0

Quindi non possiamo escludere che non segua poisson

3) Una ditta vuole misurare un modello.

Abbiamo 100 posti

Tempo intercorso	F.a.
[0, 300)	55
[300, 600)	25
[600, 900)	10
[900, 1200)	4
[1200, 1500)	3
[1500, 1800)	2
[1800, 1825]	1

Il tempo segue una legge esponenziale?

La stima di una esponenziale è

$$\frac{1}{\bar{x}}$$

E la media è

$$\bar{x} = \frac{1}{100} \left(55 * \frac{300 + 0}{2} + 25 * \frac{600 + 300}{2} + \dots \right) \simeq 403.63$$

$$\frac{1}{403.63} \simeq 0.002477$$

$$Y \sim \exp(0.002477) \rightarrow P(Y \leq y) = 1 - e^{-\lambda y}$$

$$\pi(1) = P(Y < 300) = 1 - e^{-0.002477 * 300} \simeq 0.524$$

$$\pi(2) = P(300 < Y < 600) = 1 - e^{\lambda 600} - (1 - e^{\lambda * 300}) \simeq 0.249$$

$$\pi(3) = 0.119$$

$$\pi(4) = 0.056$$

$$\pi(5) = 0.027$$

$$\pi(6) = 0.013$$

$$\pi(7) = 0.001$$

$$\pi(8) = 0.011$$

$$f_1 = n * \pi(1) = 100 * 0.524 = 52.4$$

$$f_2 = 24.9$$

$$f_3 = 11.9$$

$$f_4 = 5.6$$

$$F_5 = 2.7$$

$$f_6 = 1.3$$

$$f_7 = 0.1$$

$$\hat{f}_8 = 1.1$$

Accorpriamo 5 \rightarrow 8

$$f_6 = 5.2$$

$$n_6 = 6$$

Ed ora possiamo applicare sempre la stessa formula

$$Q = \sum \frac{(N_j - f_j)^2}{f_j} \sim \chi^2(5 - 1 - 1)$$

Togliamo un altro 1 siccome abbiamo stimato λ

In generale per ogni parametro stimato si toglie 1

$$\frac{(N_j - f_j)^2}{f_j} > \chi^2_{3,\alpha}$$

$$\alpha = 5\% = 0.05$$

$$2.672 \nlessgtr 7.81$$

Quindi non posso escludere che i dati seguono la legge esponenziale.

4) $n = 150$

classe	intervallo	frequenze osservate
1	$[0, 1/4)$	30
2	$[1/4, 1/2)$	37
3	$[1/2, 3/4)$	52
4	$[3/4, 1]$	31

Vogliamo verificare che $X \sim U[0, 1]$

a. Siccome è uniforme, la probabilità è

$$\frac{1}{4}$$

(Io e Camilla stavamo calcolando la frequenza attesa come la media, si vede che ci siamo appena svegliati)

$$b. \quad Q = \sum \frac{(N_j - f_j)^2}{f_j} = \frac{1}{f_j} * \sum (N_j - f_j)^2$$

N = Frequenza osservata

j = Frequenza attesa

$$f_j = n * p = 150 * \frac{1}{4} = 37.5$$

$$\frac{1}{37.5} * ((30 - 37.5)^2 + (37 - 37.5)^2 + \dots) = \text{calcolatrice} = \frac{206}{25} = 8.24$$

c. $\alpha = 5\%$

$$2.672 < 7.81 \quad \chi^2$$

$$8.24 > \Lambda_{k-1, \alpha}$$

$k = 4 \rightarrow$ Quante classi abbiamo

$$8.24 >? X_{3,0.05}^2 \rightarrow 8.24 > 7.915$$

I nostri dati sono in contraddizione

$$\alpha = 1\%$$

$$8.24 >? X_{3,0.01}^2 \rightarrow 8.24 > 11.345$$

I nostri non sono in contraddizione significativa con $\alpha = 1\%$

E quindi il nostro $\bar{\alpha}$ è tra $0.01 < \bar{\alpha} < 0.05$

5) Approssimazioni:

$$\bar{x}_n = 112.85$$

$$s_n = 20.80$$

$$n = 12$$

Noi abbiamo media e varianza incognite

a. $\alpha = 95\% = 0.05$

Calcoliamo intervallo di confidenza

$$\left(\bar{x}_n \pm \frac{t_{n-1, \frac{\alpha}{2}} s_n}{\sqrt{n}} \right)$$

$$\left(112.85 \pm \frac{t_{11, 0.025} * s_n}{2\sqrt{3}} \right) = \left(112.85 \pm \frac{2.201 * 20.80}{2\sqrt{3}} \right) = (99.63, 126.07)$$

b. Usiamo formula t con media e varianza incognita

$$h_0: m = 120$$

$$h_1: m \neq 120$$

Ci manca solo la varianza

$$\left| \frac{\bar{x}_n - m_0}{s_n} \sqrt{n} \right| > t_{n-1, \frac{\alpha}{2}}$$

$$\left| \frac{112.95 - 120}{20.8} 2\sqrt{3} \right| >? 2.201$$

$$1.19 \nless 2.201$$

Quindi non posso rifiutare a livello 5%

c. Stimare p value

$$1.19 = t_{11, \frac{\bar{\alpha}}{2}}$$

$$0.2 < \bar{\alpha} < 0.1$$

Il suo duoppio

$$0.4 < \bar{\alpha} < 0.2$$

6) F

X	Y
A	0.87
B	0.09

C	0.03
D	0.01

Frequenze assolute:

X	Y
A	1188
B	91
C	47
D	10

- a. Trovare frequenze attese

$$f_1 = 1162.32$$

$$f_2 = 120.24$$

$$f_3 = 40.08$$

$$f_4 = 13.36$$

- b. $\alpha = 0.05$

$$q = \sum \frac{(n_j - f_j)^2}{f_j} > X_{k-1, \alpha}^2$$

$$= \left(\frac{(1188 - 1162.32)^2}{1162.32} + \frac{(91 - 120.24)^2}{12.24} + \dots \right) > X_{3, 0.05}^2$$

$$X^2 = 7.815$$

$$9.71 > 7.815$$

Vero quindi rifiuto h_0

Quindi il modello non è conforme

- c. $9.71 > X_{3, \bar{\alpha}}^2$

$$0.025 < \bar{\alpha} < 0.1$$

