Probabilità e Statistica

DOCENTE: Prof. Elisabetta Scoppola

FOGLIO DI ESERCIZI 7

TA: Matteo Quattropani
matteo.quattropani@uniroma3.it;

22 gennaio 2018

INTERVALLI DI CONFIDENZA E TEST

Esercizio 14.1 Sia dato un campione di taglia 10 da una popolazione normale di deviazione standard 4. La media campionaria risulta uguale a 48.

(1) Dare un intervallo di confidenza del 95% per la media della popolazione.
(2) Testare l'ipotesi $H_0 : \mu = 45$ contro l'ipotesti $H_1 : \mu \neq 45$ al 5% di significatività.
(3) Qual è il livello di significatività dell'ipotesi $H_0 : \mu \leq 45$ contro $H_1 = \mu > 45$?

Soluzione.

(1) Sappiamo che
$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$
e vorremmo trovare $x > 0$ t.c.
$$\mathbb{P}(|\bar{X} - \mu| < x) = 0.95$$
Il membro di sinistra dell'ultima uguaglianza è uguale a
$$\mathbb{P}(\frac{|\bar{X} - \mu|}{\sigma/\sqrt{n}} < x\sqrt{n}/\sigma) = 0.95.$$
ovvero, detta $Z \sim \mathcal{N}(0, 1)$
$$\mathbb{P}(|Z| < x\sqrt{n}/\sigma) = 0.95$$
$$x\sqrt{n}/\sigma = 1.96$$
$$x = 1.96\sigma/\sqrt{n} \approx 2.47.$$
Dunque l'intervallo è $[48 - 2.47, 48 + 2.47]$.

(2) Poichè
$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$
sotto H_0
$$\mathbb{P}(|Z| \geq \frac{\bar{X} - 45}{\sigma/\sqrt{n}}) = \mathbb{P}(|Z| \geq \frac{3}{4/\sqrt{10}} = 2.37) \leq \mathbb{P}(|Z| \geq 1.96) = 0.05.$$
Quindi rifiuto H_0.

(3) Se valesse H_0 allora $\exists \varepsilon \geq 0$ t.c. $\mu = 45 - \varepsilon$. In quel caso, la probabilità di osservare un campione con le caratteristiche estreme almeno come quelle di quello in esame è data da
$$\mathbb{P}(Z \geq \frac{\bar{X} - 45 + \varepsilon}{\sigma/\sqrt{n}}) \leq \mathbb{P}(Z \geq \frac{\bar{X} - 45}{\sigma/\sqrt{n}}) =: p_{H_0}$$
dunque
$$p_{H_0} = 1 - \Phi(2.37) \approx 0.0089.$$

Esercizio 14.2 Sia p la probabilità che una moneta dia testa. Su n lanci ha dato k volte testa.

(1) In funzione di n e k, dare un intervallo di confidenza del 95% per la probabilità di testa assumendo n grande.
(2) In funzione di n e k esprimere il livello di significatività dell'ipotesi H_0 : "la moneta è equa" contro H_1 : "la moneta non è equa". Dare il risultato in forma esatta (n piccolo) e in forma approssimata (n grande).
(3) Supponiamo di aver osservato $k = (1-2\delta)\frac{n}{2}$. Per quali valori di $\delta > 0$ viene rifiutata l'ipotesi nulla H_0 :"Testa è favorita", contro H_1 :"Croce è favorita", se viene richiesta una significatività del 5%? Dare il risultato in forma esatta (n piccolo) e in forma approssimata (n grande).

Soluzione.

(1) Detto $\hat{p} := k/n$

$$\frac{n\hat{p} - np}{\sqrt{n\hat{p}(1-\hat{p})}} \xrightarrow{d} Z \sim \mathcal{N}(0,1)$$

dunque, detto $z \in \mathbb{R}^+$ il valore per cui

$$\mathbb{P}(|\hat{p} - p| \leq z) = 0.95.$$

Ne segue che

$$\mathbb{P}(\frac{n|\hat{p}-p|}{\sqrt{n\hat{p}(1-\hat{p})}} \leq zn/\sqrt{n\hat{p}(1-\hat{p})}) \to \mathbb{P}(|Z| \leq z\sqrt{n}/\sqrt{\hat{p}(1-\hat{p})}) = 0.95.$$

per tanto

$$z\sqrt{n}/\sqrt{n\hat{p}(1-\hat{p})} = 1.96 \Longrightarrow z = 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96\sqrt{\frac{k(n-k)}{n^3}}.$$

Pertanto l'intevallo di confidenza è

$$\left[\frac{k}{n} - \frac{1.96}{n}\sqrt{\frac{k(n-k)}{n}}, \frac{k}{n} + \frac{1.96}{n}\sqrt{\frac{k(n-k)}{n}}\right].$$

(2) In altri termini $H_0 : p = 1/2$ e $H_1 : p \neq 1/2$. Se vale H_0 allora sarebbe vero che l'outcome di una variabile aleatoria (approssimativamente) normale standard abbia assunto il valore

$$\frac{\hat{p} - 1/2}{\sqrt{1/2(1-1/2)/n}} = 2\sqrt{n}\left(\frac{k}{n} - \frac{1}{2}\right)$$

La probabilità di un evento "di tale portata" è (approssimativamente)

$$\mathbb{P}\left(|Z| \geq \sqrt{2n}\left(\frac{k}{n} - \frac{1}{2}\right)\right) = 1 - 2\Phi\left(2\sqrt{n}\left(\frac{k}{n} - \frac{1}{2}\right)\right) =: p_{H_0}$$

in altre parole rifiutiamo l'ipotesi nulla per ogni soglia di significatività maggiore di p_{H_0}. Se n è piccolo, allora sotto H_0

$$k \sim Bin(n, 1/2).$$

Sia $X \sim Bin(n, 1/2)$. Noi vorremmo che, fissato un qualche α,

$$\mathbb{P}(X \geq k) = 1 - \sum_{i=0}^{k} \binom{n}{i} \frac{1}{2^n} < \frac{\alpha}{2}$$

o che

$$\mathbb{P}(X \leq k) = \sum_{i=0}^{k} \binom{n}{i} \frac{1}{2^n} < \frac{\alpha}{2}.$$

Se $k > n/2$ il p-value è quindi dato da

$$p_{H_0} := 2\mathbb{P}(X \geq k)$$

altrimenti
$$p_{H_0} := 2\mathbb{P}(X \leq k).$$
(3) Sia $H_0 : p \geq 1/2$ e $H_1 : p < 1/2$. Se vale H_0 allora $\exists \varepsilon \geq 0$ t.c. $p = 1/2 + \varepsilon$. Quindi
$$\sqrt{n}\frac{\hat{p} - \frac{1}{2} - \varepsilon}{\sqrt{\left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)}} \sim \mathcal{N}(0,1).$$
Ho osservato k (e quindi $\hat{p}$) e vorrei trovare per quali valori di $\delta > 0$ ho
$$\mathbb{P}\left(Z \leq -\sqrt{n}\frac{\delta+\varepsilon}{\sqrt{\left(\frac{1}{2}-\varepsilon\right)\left(\frac{1}{2}+\varepsilon\right)}}\right) \leq 0.05.$$
Iniziamo col notare che
$$\mathbb{P}\left(Z \leq -\sqrt{n}\frac{\delta+\varepsilon}{\sqrt{\left(\frac{1}{2}-\varepsilon\right)\left(\frac{1}{2}+\varepsilon\right)}}\right) \leq \mathbb{P}\left(Z \leq -\sqrt{n}\frac{\delta}{\sqrt{\left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)}}\right)$$
notiamo anche che, $\forall \varepsilon \in (0, 1/2)$
$$\sqrt{\left(\frac{1}{2}-\varepsilon\right)\left(\frac{1}{2}+\varepsilon\right)}^{-1} \geq 2$$
e dunque
$$\mathbb{P}\left(Z \leq -\sqrt{n}\frac{\delta}{\sqrt{\left(\frac{1}{2}+\varepsilon\right)\left(\frac{1}{2}-\varepsilon\right)}}\right) \leq \mathbb{P}\left(Z \leq -2\sqrt{n}\delta\right)$$
e il membro di destra è minore di 0.05 se
$$2\sqrt{n}\delta \geq 1.645 \Longrightarrow \delta \geq \frac{1.645}{2\sqrt{n}}.$$
Supponiamo ora n piccolo. Se vale H_0, i.e. $p = 1/2 + \varepsilon$ e $X \sim Bin(n, p)$ allora
$$\begin{aligned}\mathbb{P}(X \leq k = (1-2\delta)n/2) &= \sum_{i=0}^{(1-2\delta)n/2} \binom{n}{i}\left(\frac{1}{2}+\varepsilon\right)^i\left(\frac{1}{2}-\varepsilon\right)^{n-i} \\ &\leq \sum_{i=0}^{(1-2\delta)n/2} \binom{n}{i}\left(\frac{1}{2}\right)^i\left(\frac{1}{2}\right)^{n-i} \\ &= \sum_{i=0}^{(1-2\delta)n/2} \binom{n}{i}\left(\frac{1}{2}\right)^n\end{aligned}$$
Quest'ultimo è minore di 0.05 se
$$\delta \geq \max\{\delta \in (0, 1/2) \mid \sum_{i=0}^{(1-2\delta)n/2} \binom{n}{i}\left(\frac{1}{2}\right)^n \geq 0.05\}.$$

Esercizio 14.3 Un'azienda che produce componenti meccaniche mette sul mercato un nuovo prodotto che afferma essere migliore del precedente in termini di durata. Gli stessi produttori affermano che però la deviazione standard della durata del nuovo modello è raddoppiata, da 1 a 2 giorni. Un campione di prova di 10 pezzi viene inviato ad un'azienda cliente che aveva già acquistato 100 pezzi del modello precedente. La durata media del campione di 100 pezzi è 10 giorni mentre la durata media del campione del nuovo modello è di 11 giorni. Assumendo che la durata di vita di un prodotto sia approssimativamente gaussiana, ed assumendo vera l'informazione fornita dal produttore sulla varianza dei due prodotti testare l'ipotesi H_0 :"La durata media dei due modelli è uguale", contro H_1 :"La durata media dei due modelli non è uguale" e poi contro H_2 : "Il nuovo modello ha durata media maggiore".

Soluzione. Siano $n_X = 100$, $n_Y = 10$, $\bar{X} = 10$, $\bar{Y} = 11$, $\sigma_X = 1$, $\sigma_Y = 2$. Se vale H_0 allora

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \sim \mathcal{N}(0, 1).$$

Poichè

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} = -1.56,$$

rifiutiamo H_0 per ogni soglia maggiore o uguale a

$$p_{H_0} := \mathbb{P}(|Z| \geq 1.56) = 0.1189.$$

Se testiamo H_0 contro H_2 la nostra statistica test rimarrà la stessa, tuttavia rifuteremo l'ipotesi nulla per ogni soglia maggiore o uguale a

$$p_{H_0} := \mathbb{P}(Z \leq -1.56) = 0.0593.$$

Esercizio 14.4 Supponiamo che il salario medio degli alumni dell'università di Roma Tre sia distribuito come una v.a. gaussiana. Su un campione di 40 alumni il salario medio è di 1600 euro al mese e la deviazione standard campionaria è di 200 euro al mese.

(1) Dare un'intervallo di confidenza del 95% per la media.
(2) Qual è il livello di significatività con cui possiamo rifiutare l'ipotesi H_0 :"La media dei salari degli alumni è maggiore di 1700" contro l'ipotesi H_1 :"La media dei salari degli alumni è minore di 1700"?

Soluzione.

(1) Chiamiamo $n = 40$, $\bar{X} = 1600$ e $S = 200$. Sappiamo che

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

dunque

$$\mathbb{P}\left(-t_{\alpha/2,n-1} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t_{\alpha/2,n-1}\right) = 1 - \alpha$$

ne sugue che

$$\mathbb{P}\left(\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

ponendo $\alpha = 0.05$ e sapendo che

$$t_{0.025,39} \approx 2.023$$

otteniamo

$$t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} \approx 63.97$$

e quindi l'intervallo richiesto è

$$[1536.03, 1663.97]\,.$$

(2) Se è vera H_0 allora esiste $\varepsilon \geq 0$ t.c. $\mu = 1700 + \varepsilon$, dunque l'evento verificatosi è

$$T = \frac{-100 - \varepsilon}{200/\sqrt{40}}$$

con $T \sim t_{39}$. La probabilità di osservare un valore così negativo è data da

$$\mathbb{P}(T \leq \frac{-100-\varepsilon}{200/\sqrt{40}}) \leq \mathbb{P}(T \leq \frac{-100}{200/\sqrt{40}}) \approx \mathbb{P}(T \leq -3.16) \approx 0.0015.$$

In altre parole la significativà con cui possiamo rifiutare H_0 è 0.0015.