LM−88 SOCIOLOGIA E RICERCA SOCIALE

Metodi Statistici per la Ricerca Sociale

Richiami di Inferenza Statistica

1. Si è a conoscenza che il fatturato annuo (in migliaia di euro) della piccole imprese di un paese si distribuisce come una Normale di media $\mu = 100$ e varianza $\sigma^2 = 625$.

   (a) Estratto un campione di $n = 9$ imprese, è più probabile osservare valori della media campionaria superiori a 115 o inferiori a 90?

   (b) Se si aumenta la dimensione campionaria portandola a $n = 25$ imprese, la probabilità di osservare valori della media campionaria superiori a 115 aumenta o diminuisce rispetto alla corrispondente probabilità calcolata al punto precedente?

   (c) Per quale dimensione campionare la deviazione standard (errore standard) della media campionaria è pari a 2.5?

   Soluzione:

   (a) È più probabile osservare valori della media campionaria inferiori a 90: il valore 90 è più vicino alla media $\mu = 100$ del valore 115: $P(\overline{Y} < 90) = 0.11507$ e $P(\overline{Y} > 115) = 0.03593$

   (b) Diminuisce: $P(\overline{Y} > 115) = 0.00135$

   (c) $n = 100$

2. In una città il 35% dei cittadini prende la metropolitana durante la mattina. Si estrae da tale popolazione un campione casuale di $n = 50$ cittadini.

   (a) Calcolare valore atteso e errore standard della proporzione campionaria.

   (b) Utilizzando l'approssimazione normale, calcolare la probabilità che la proporzione campinaria dei cittadini che prendono la metropolitana durante la mattina sia superiore a 1/3.

   Soluzione:

   (a) $\mu_{\hat{\pi}} = 0.35$ e $\sigma_{\hat{\pi}} = 0.067$

   (b) $P(\hat{\pi} > 1/3) = 0.59758$

3. In un piccolo paese con 5000 occupati il salario medio è $\mu = 1000$ euro con una varianza pari a $\sigma^2 = 40000$. Considerando un campione casuale di $n = 150$ occupati estratti da tale popolazione

   (a) Calcolare il valore atteso e errore standard della media campionaria dei salari.

   (b) La media campionaria ha distribuzione Normale?

   Soluzione:

   (a) $\mu_{\overline{Y}} = 1000$ e $\sigma_{\overline{Y}} = 16.33$

   (b) La dimensione campionaria è sufficiente grande da poter approssimare la distribuzione campionaria della media campionaria con una distribuzione Normale

4. In una sessione di esami di un corso di statistica a cui hanno partecipato diverse centinaia di studenti, sono state rilevate in modo casuale le votazioni di 10 studenti:

$$25 \quad 30 \quad 27 \quad 22 \quad 30 \quad 28 \quad 26 \quad 20 \quad 28 \quad 26$$

(a) Stimare, mediante un opportuno stimatore corretto, il voto medio.

(b) Stimare, mediante un opportuno stimatore corretto, la varianza del voto.

(c) Stimare la proporzione di 30 nella popolazione.

Soluzione:

(a) $\overline{Y} = 26.2$

(b) $s^2 = 10.4$

(c) Stima della proporzione di 30 nella popolazione $= 0.2$

5. In una popolazione di adulti, la varianza della statura è pari a 49 cm^2. Costruire l'intervallo di confidenza al 98% per la statura media della popolazione, supponendo di aver osservato un campione di 100 individui con media campionaria pari a 170 cm.

Soluzione:

$$z_{0.01} = 2.33 \qquad se_{\overline{Y}} = 0.7 \qquad IC_{0.98}(\mu) = [168.369; 171.631]$$

6. In una popolazione di studenti, il punteggio conseguito in una prova di matematica si distribuisce secondo una Normale. Un campione casuale di 25 studenti presenta media 98.4 e varianza campionaria (corretta) 201.64. Costruire l'intervallo di confidenza al 95% per il punteggio medio μ della popolazione.

Soluzione:

$$t_{0.025,24} = 2.06 \qquad \hat{se}_{\overline{Y}} = 2.84 \qquad IC_{0.95}(\mu) = [92.5496; 104.2504]$$

7. Una grande azienda di software vuole stimare la proporzione di ragazzi di età compresa tra i 10 e i 15 anni che utilizzano la rete internet. Dei 370 ragazzi intervistati 214 dichiarano di utilizzare la rete internet. Si calcoli l'intervallo di confidenza al 95% per la proporzione di ragazzi che utilizzano internet.

Soluzione:

$$\hat{\pi} = 0.578 \qquad z_{0.025} = 1.96 \qquad \hat{se}_{\hat{\pi}} = 0.026 \qquad IC_{0.95}(\pi) = [0.528; 0.629]$$

8. Un laboratorio analizza un composto farmaceutico per determinare la concentrazione di principio attivo. Vengono effettuate con un'opportuna strumentazione diverse misurazioni. È noto che i valori di concentrazione misurati nelle varie prove seguono una distribuzione Normale con media ignota e deviazione standard (che dipende dalla strumentazione) pari a $\sigma = 0.2$. Il laboratorio esegue 4 misurazioni del composto ottenendo i seguenti risultati

$$2.1 \quad 1.9 \quad 1.8 \quad 2$$

(a) Costruire un intervallo di confidenza al 95% per la concentrazione media del principio attivo.

(b) La casa farmaceutica stabilisce che l'ampiezza dell'intervallo di confidenza al 95% non debba superare 0.248. Quale dimensione campionaria bisognerà scegliere?

Soluzione:

(a) $\overline{y} = 1.95$ $\quad z_{0.025} = 1.96$ $\quad \hat{se}_{\overline{Y}} = 0.1$ $\quad IC_{0.95}(\mu) = [1.754; 2.146]$

(b) $n \geq (2 \cdot 1.96 \cdot 0.2/0.248)^2 = 9.994$ Quindi $n = 10$.

9. Si vuole studiare la temperatura massima a Marrakech (Marocco) nel periodo Marzo - Aprile. In tale periodo le temperature massime osservate in 7 giorni sono state le seguenti

$$22 \quad 26 \quad 25 \quad 25 \quad 28 \quad 23 \quad 24$$

Supponendo che le temperature massime giornaliere seguono una distribuzione Normale, costruire l'intervallo di confidenza al 90% per la media della temperatura massima.

Soluzione:

$\overline{y} = 24.714$ $\quad s^2 = 3.9048$ $\quad t_{0.05,6} = 1.94$ $\quad \hat{se}_{\overline{Y}} = 0.7469$ $\quad IC_{0.90}(\mu) = [23.26535; 26.16322]$

10. Sull'etichetta di un fustino di detersivo si dichiara un contenuto pari a 5 kg. Un'associazione di consumatori ritiene che il contenuto sia mediamente inferiore al peso dichiarato. Si ipotizza che il peso Y del prodotto sia una variabile aleatoria Normale con varianza nota $\sigma^2 = 0.25$. Dall'esame del contenuto di un campione di $n = 50$ fustini, risulta una media campionaria $\bar{Y} = 4.8$ kg.

(a) Definire l'ipotesi nulla e l'ipotesi alternativa

(b) Definire la statistica test, costruire la regione critica al livello di significatività del 5%.

(c) Determinare il valore della statistica test e prendere una decisione considerando un livello di significatività del 5%.

(d) Calcolare il $p-$value e determinare la conseguente decisione. Fino a quale livello di significatività si mantiene la stessa decisione?

Soluzione:

(a) $H_0 : \mu = 5$ versus $H_a : \mu < 5$

(b) $Z = \dfrac{\overline{Y} - 5}{\sqrt{0.25/50}}$ e $RC_{0.05} : Z \leq -1.645$

(c) $z^{oss} = -2.82$. Si rifiuta H_0 al livello di significatività del 5% perché $z^{oss} = -2.82 < -1.645$

(d) $p - valore = P(Z \leq -2.82) = 0.00234$. Il $p-$valore mostra evidenza contro l'ipotesi nulla (in particolare è più piccolo di 0.05). Si rifiuta H_0 per tutti i livelli di significatività maggiori di 0.00234.

11. Se il valore del $p-$value é uguale a 0.07, a quali livelli di significatività α saremo portati a rifiutare l'ipotesi nulla?

Soluzione: Per tutti i livelli di significatività maggiori di 0.07

12. Il direttore di un'azienda che produce pasta ha il sospetto che il peso medio dei pacchi di pasta confezionati sia superiore a 500 grammi. Un campione di 25 pacchi dell'ultima partita prodotta ha fatto registrare un peso medio pari a 510 grammi con una stima (corretta) della varianza pari a 100. Supposto che il peso dei pacchi sia distribuito normalmente, si verifichi l'ipotesi che il peso medio sia effettivamente di 500 grammi contro l'ipotesi che sia maggiore al livello di significatività $\alpha = 0.01$.

   Soluzione: Ipotesi: $H_0 : \mu = 500$ versus $H_a : \mu > 500$

   Statistica test e regione critica: $T = \dfrac{\overline{Y} - 500}{\sqrt{100/25}}$ e $RC_{0.01} : T \geq t_{0.01,24} = 2.492$

   Valore osservato della statistica test: $t^{oss} = 5$.

   Decisione: Si rifiuta H_0 al livello di significatività del 1% perché $t^{oss} = 5 > 2.492$.

13. La concentrazione media di sostanti inquinanti osservata nelle acque di un fiume in un determinato anno è stata del 7.1%. L'anno seguente sono state effettuate 5 rilevazioni e la concentrazione di sostanze inquinanti é risultata pari a $\bar{y} = 7.07\%$ con $s = 0.0265$. Al livello di significatività $\alpha = 0.05$, possiamo affermare che il valore medio della concentrazione è diversa da quello riscontrato nell'anno precedente? (Si supponga che la concentrazione segua una distribuzione Normale)

   Soluzione: Ipotesi: $H_0 : \mu = 7.1\%$ versus $H_a : \mu \neq 7.1\%$

   Statistica test $T = \dfrac{\overline{Y} - 7.1}{0.0265/\sqrt{5}}$

   Regione critica: $RC_{0.05} : T \leq -t_{0.025,4} = -2.776$ oppure $T \geq t_{0.025,4} = 2.776$

   Valore osservato della statistica test: $t^{oss} = -2.531$.

   Decisione: Non si rifiuta H_0 al livello di significatività del 5% perché $t^{oss} = -2.531$ non appartiene alla regione critica.

14. Si vuole stabilire se la percentuale di fumatori tra gli impiegati è diversa rispetto a quella dell'intero paese. Si è a conoscenza che nel paese il 25% delle persone fuma. In un ufficio con $n = 80$ impiegati il 29.8% sono fumatori. La differenza osservata è dovuta al caso? (Verificare l'ipotesi al livello di significatività $\alpha = 0.05$)

   Soluzione: Ipotesi: $H_0 : \pi = 0.25$ versus $H_a : \pi \neq 0.25\%$

   Statistica test $Z = \dfrac{\hat{\pi} - 0.25}{\sqrt{0.25 \cdot (1 - 0.25)/80}}$

   Regione critica: $RC_{0.05} : Z \leq 1.96$ oppure $Z \geq 1.96$

   Valore osservato della statistica test: $z^{oss} = 0.9915$.

   Decisione: Non si rifiuta H_0 al livello di significatività del 5% perché $z^{oss} = 0.9915$ non appartiene alla regione critica.

   $P-$valore $= 2 \cdot [1 - P(Z \leq 0.9915)] = 0.32145$ (il $p-$valore è grande)

15. In uno studio sugli effetti nocivi delle polveri respirate dagli operai di un'industria, si è trovato che 16 operai su 120 visitati presentano forti disturbi respiratori. È noto che il 10% della popolazione è affetta

da tali disturbi. Verificare se la proporzione tra gli operai è maggiore della proporzione nella popolazione al livello di significatività $\alpha = 0.05$.

Soluzione: Ipotesi: $H_0 : \pi = 0.10$ versus $H_a : \pi > 0.10\%$

Statistica test $Z = \dfrac{\hat{\pi} - 0.10}{\sqrt{0.10 \cdot (1 - 0.10)/120}}$

Regione critica: $RC_{0.05} : Z \geq 1.645$

Valore osservato della statistica test: $\hat{\pi} = 0.133$ quindi $z^{oss} = 1.217$.

Decisione: Non si rifiuta H_0 al livello di significatività del 5% perché $z^{oss} = 1.217$ non appartiene alla regione critica.

$P-$valore $= P(Z \geq 1.217) = 0.1118$ (il $p-$valore è maggiore di 0.05)

Esercizi dal libro di testo:

Capitolo 4.	–	4.18	4.27	4.29	4.30	4.32	4.33	4.35	4.50	4.51	4.52	4.53	
Capitolo 5.	–	5.14	5.24	5.22	5.26	5.33(b,c)	5.37	5.38	5.66	5.67	5.68	5.69	
Capitolo 6.	–	6.7	6.8	6.17	6.18	6.21	6.33	6.52	6.53	6.54	6.55	6.56	6.57