Statistica

Esercitazione 14

Alfonso Iodice D'Enza
iodicede@unicas.it

Università degli studi di Cassino

Ex.1: Verifica Ipotesi sulla media (varianza nota)

Le funi prodotte da un certo macchinario hanno una resistenza media alla rottura pari a $\mu = 1800N$ con una variabilità (scarto quadratico medio) $\sigma = 100N$. In seguito ad una modifica nel processo produttivo, si ritiene che la resistenza alla rottura delle funi prodotte sia migliorata. Per verificare tale miglioramento è stato analizzato un campione di $n = 50$ funi di nuova produzione. La resistenza nuova osservata è stata $\bar{X} = 1850N$. Verificare ad un livello di significatività del $\alpha = 0.01$ che il vi sia un effettivo miglioramento nella resistenza delle funi.

Svolgimento

L'effetto del nuovo processo produttivo è positivo se si traduce in un aumento della resistenza media alla rottura delle funi prodotte. Formalmente,

- $H_0 : \mu = 1800$
- $H_1 : \mu > 1800$

Ex.1: Verifica Ipotesi sulla media (varianza nota)

Il test è unidirezionale, si è interessati a che la statistica test costruita in base alla media osservata. In particolare la statistica di interesse è

$$Z_{oss} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{1850 - 1800}{100/\sqrt{50}} = 3.55$$

Per verificare se l'ipotesi nulla sia da rifiutare o meno è necessario confrontare il valore della statistica osservata con il valore critico $Z_c = Z_{1-\alpha} = Z_{0.99} = 2.33$. Poichè risulta $Z_{oss} > Z_c (3.55 > 2.33)$, si rifiuta l'ipotesi nulla e si può concludere che il nuovo processo produttivo abbia effetto sulla resistenza delle funi.

Ex.1.1: Verifica Ipotesi sulla media

Si consideri di voler applicare la stessa modifica del processo produttivo ad un macchinario che produce funi caratterizzati da una resistenza media $\mu = 1830N$. Si supponga che la variabilità sia uguale a quella del macchinario precedente $\sigma = 100$. Calcolare il $p.value$ sapendo che anche in questo caso il campione di 50 unità osservato ha una resistenza media di $\bar{X} = 1850$.

nota

Il $p.value$ è il più piccolo livello di significatività tale da indurre al rifiuto dell'ipotesi H_0 sulla base dei dati osservati. In altre parole rappresenta la probabilità che i dati non compatibili con l'ipotesi nulla siano stati osservati quando in realtà H_0 era vera. Di conseguenza un $p.value$ molto piccolo è un forte indicatore del fatto che H_0 non è vera.

Ex.1.1: Verifica Ipotesi sulla media

Statistica

A. Iodice

Si consideri di voler applicare la stessa modifica del processo produttivo ad un macchinario che produce funi caratterizzati da una resistenza media $\mu = 1830N$. Si supponga che la variabilità sia uguale a quella del macchinario precedente $\sigma = 100$. Calcolare il $p.value$ sapendo che anche in questo caso il campione di 50 unità osservato ha una resistenza media di $\bar{X} = 1850$.

svolgimento

$$Z_{oss} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}}{\sigma}\left(\bar{X} - \mu\right) = \frac{\sqrt{50}}{100}1850 - 1830 = 1.41$$

si vuole $P(Z > 1.41) = 1 - P(Z < 1.41) = 1 - 0.9207 = 0.0793$ dunque il $p.value = .0793$: questo vuol dire che ad un livello di significatività del 5% H_0 non sarebbe da rifiutare($Z_{1-0.05} = 1.645$, dunque $1.41 < 1.645$) , mentre un test al 10% di significatività indurrebbe a rifiutare H_0($Z_{1-0.1} = 1.285$, dunque $1.41 > 1.285$) .

Ex.2: Verifica Ipotesi sulla media (varianza non nota)

Un gruppo di volontari tra i pazienti ricoverati in ospedale per colesterolo alto (almeno 240ml per ogni dl decilitro di sangue) è stato selezionato per testare l'efficacia di un farmaco anti-colesterolo. Dunque a 40 volontari è stato somministrato il farmaco per 60 giorni ed stato poi misurato nuovamente il livello di colesterolo. Il livello medio di colesterolo nel campione si è ridotto di 6.8, mentre la deviazione standard campionaria riscontrata è 12.1. Utilizzare un livello di significatività pari al 5%.

Svolgimento

Per testare che la variazione del livello di colesterolo riscontrata dopo il trattamento occorre considerare le seguenti ipotesi

- $H_0 : \mu = 0$
- $H_1 : \mu \neq 0$

dove μ rappresenta il *decremento* di colesterolo.

Ex.2: Verifica Ipotesi sulla media (varianza non nota)

Svolgimento

La statistica test da utilizzare in questo caso è

$$T_{oss} = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{\sqrt{n}}{S}(\bar{X} - \mu) = \frac{\sqrt{40}}{12.1} 6.8 = 3.554$$

il valore critico è $t_{n-1,\alpha/2} = t_{39,0.025} = 2.02$. Poichè $|T_{oss}| > |T_c|$ allora si rifiuta H_0. Esiste dunque un effetto non casuale sul livello di colesterolo, non si può tuttavia sostenere che il farmaco abbia avuto effetto: potrebbero essere state altre le cause del miglioramento dello stato dei pazienti (effetto placebo). E' dunque prassi operativa separare i pazienti in due gruppi: solo ad uno dei due gruppi si somministra realmente il farmaco, dopo di che si analizzano le eventuali variazioni nel livello di colesterolo tra i due gruppi.
Il $p.value = 2P(T_{39} > 3.554) = 0.0001$.

Ex.3: Verifica Ipotesi sulla proporzione

In un esperimento per verificare eventuali capacità extrasensoriali di un soggetto A si procede nel seguente modo: si consegna un mazzo di 50 carte di colore rosso e blu. La prova per il soggetto A consiste nell'indovinare il colore della carta scelta da un soggetto B che si trova in un'altra stanza. Se il soggetto A indovina 32 carte, è possibile sostenere che si tratta di un fenomeno paranormale, ad un livello di significatività del 5%?

Svolgimento

La probabilità che un soggetto indovini se la carta è blu o rossa è $p = 0.5$. L'ipotesi nulla è che il soggetto A provi semplicemente ad indovinare le scelte di B; l'ipotesi alternativa 'e che il soggetto A abbia effettivamente delle capacità extrasensoriali. Formalmente,

- $H_0 : p = 0.5$
- $H_1 : p > 0.5$

Il test è unidirezionale dal momento che si è interessati a che il soggetto A abbia prestazioni superiori al semplice 'tirare ad indovinare'. Sotto ipotesi nulla, vale a dire per $p = 0.5$, media e scarto quadratico medio valgono rispettivamente

$\mu = Np = 50 \times 0.5 = 25$ e $\sigma = \sqrt{Np(1-p)} = \sqrt{50 \times 0.5 \times 0.5} = \sqrt{12.5} = 3.54$.

Ex.3: Verifica Ipotesi sulla proporzione

Svolgimento

Ad un livello di significatività del 5% bisogna scegliere z_c tale che $P(Z \leq z_c) = 0.95$. Pertanto $z_c = 1.645$. In base ai risultati dell'esperimento, il valore osservato z_o corrisponde

$$z_{oss} = \frac{32-25}{3.54} = 1.98$$

poichè risulta essere $z_{oss} > z_c$ $(1.98 > 1.645)$, si rifiuta l'ipotesi nulla per la quale il soggetto A ha tirato ad indovinare.

Ex.4: Verifica Ipotesi sulla Varianza

Un macchinario produce batterie al lithio che hanno una durata di vita media di 3 anni con uno scarto quadratico medio di 1 anno. Estratto il seguente campione di $n = 5$ batterie caratterizzate da una durata pari a

$$1.9, 2.4, 3, 3.5, 4.2$$

si vuole verificare, ad un livello di significatività del 5%, se la varianza dichiarata per le batterie prodotte dal macchinario sia effettivamente pari ad 1.

Svolgimento

Le ipotesi possono essere formalizzate nel seguente modo

- $H_0 : \sigma^2 = 1$
- $H_1 : \sigma^2 \neq 1$

La statistica a cui fare riferimento è la seguente:

$$y = \frac{\hat{S}^2}{\sigma^2}(n-1) \text{ dove } \hat{S}^2 = \frac{\sum_{i=1}^{n} x_i - \bar{x}}{n-1}$$

Tale statistica si distribuisce come una variabile casuale χ^2 con $n-1$ gradi di libertà.

Ex.4: Verifica Ipotesi sulla Varianza

Svolgimento

Essendo $n-1=4$ e il test bidirezionale, gli estremi della regione di accettazione sono

$$\chi^2_{\frac{1-\alpha}{2},4} = \chi^2_{0.975,4} = 0.48 \text{ e } \chi^2_{\frac{\alpha}{2},4} = \chi^2_{0.025,4} = 11.14$$

Per decidere se rigettare o meno l'ipotesi nulla bisogna verificare se il valore osservato della statistica sia compreso o meno nei limiti della regione di accettazione. In particolare, dai dati campionari risulta essere

$$y_o = \frac{\hat{S}^2}{\sigma_o^2}(n-1) = \frac{0.9}{1} \times 4 = 3.6$$

Poichè il valore osservato è compreso nei limiti della regione di accettazione, ovvero $0.48 < 3.6 < 11.14$, si accetta l'ipotesi nulla e, di conseguenza, si attesta che la varianza della durata delle batterie prodotte non si discosta significativamente da 1.

Ex.5: Verifica Ipotesi sulla indipendenza tra mutabili

Per testare l'efficacia di un farmaco nella cura di una malattia si considera un campione di 200 pazienti, li si suddivide in due gruppi (A e B). Al gruppo A viene somministrato il farmaco, al gruppo B no; si osserva poi, per ciascun gruppo, il numero di pazienti guariti. I risultati della prova sono riassunti nella seguente tabella

	guariti	non guariti	tot.
gruppo A	75	25	100
gruppo B	65	35	100
tot	140	60	200

Table: Appartenenza ai gruppi vs. guariti (non guariti).

A partire da tale esperimento, si costruisca un test per valutare l'efficacia del farmaco nella cura della malattia.

Svolgimento

Per stabillire se il farmaco sia o meno curativo si può sottoporre la verifica di ipotesi che le due variabili *gruppo di appartenenza* e *guarigione dalla malattia* siano o meno indipendenti: se così dovesse risultare, ovvero che la guarigione dalla malattia non è influenzata dalla assunzione del farmaco, si potrebbe concludere che il farmaco è inefficace.

Ex.5: Verifica Ipotesi sulla indipendenza tra mutabili

Svolgimento

Formalmente

- H_0 : le variabili considerate sono indipendenti.
- H_1 : le variabili considerate non sono indipendenti.

Al fine di calcolare l'indice quadratico di connessione è necessario calcolare le frequenze che ci si attenderebbe se, fissati i marginali di tabella, le variabili fossero indipendenti.

	guariti	non guariti	tot.
gruppo A	70	30	100
gruppo B	70	30	100
tot	140	60	200

Table: Frequenze attese sotto l'ipotesi di indipendenza

Ex.5: Verifica Ipotesi sulla indipendenza tra mutabili

Statistica

A. Iodice

Svolgimento

Si passa a calcolare l'indice quadratico di connessione

$$\chi^2 = \sum_{i=1}^{h}\sum_{j=1}^{k} \frac{(n_{ij} - \hat{n_{ij}})^2}{\hat{n_{ij}}} = \frac{(75-70)^2}{70} + \frac{(25-30)^2}{30} + \frac{(65-70)^2}{70} + \frac{(35-30)^2}{30} = 2.38$$

La statistica appena calcolata si distribuisce secondo una v.c. *chi-quadro* χ^2 con $(h-1)\times(k-1) = 1$ grado di libertà (h e k sono rispettivamente il numero di modalità delle mutabili considerate). Ad un livello di significatività del 5% e con 1 g.d.l. il valore critico è $\chi^2_{0.95,1} = 3.84$. Poichè il valore osservato della statistica è minore del valore critico $(2.38 < 3.84)$, non si può rifiutare l'ipotesi nulla: di conseguenza non si può sostenere che il farmaco abbia effetti sulla cura della malattia.