CORSO DI STATISTICA (parte 2) - ESERCITAZIONE 7

Dott.ssa Antonella Costanzo

a.costanzo@unicas.it

Esercizio 1. Test delle ipotesi per la media (varianza nota), p-value del test

Il manager di un fast-food è interessato a valutare se il tempo di attesa per effettuare un ordine da parte dei clienti è cambiato negli ultimi 6 mesi rispetto allo scorso anno. E' noto, dall'esperienza passata, che il tempo medio di attesa prima di ordinare è di 4.5 minuti con una deviazione standard di 1.2 minuti. Inoltre il manager può ragionevolmente supporre che la distribuzione del tempo sia normale.
A tale proposito, viene selezionato a caso un campione di 25 ordini effettuati in un'ora da cui risulta un tempo medio di attesa di 5.1 minuti. Il manager può affermare con un livello di significatività del 5% che il tempo medio di attesa per ordinare è significativamente cambiato rispetto al passato?
Calcolare inoltre il livello di significatività osservato (p-value) del test.

Sol.

1. Definizione del sistema di ipotesi:

$H_0: \mu = 4.5$

$H_1: \mu \neq 4.5$

2. Livello di significatività

$\alpha = 0.05$

3. Costruzione della statistica test

Siccome la varianza della popolazione è nota, allora:

$$Z_{stat} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Il livello di significatività è $\alpha = 0.05$, il test è bidirezionale, quindi i quantili della distribuzione Normale standard da individuare sono: $\pm z_{(1-\frac{\alpha}{2})}$ ossia:

$-z_{0.975} = -1.96$
$+z_{0.975} = +1.96$

4. Definizione della regola di decisione (regione di rifiuto)

Se $Z_{stat} \leq -1.96$ oppure $Z_{stat} \geq +1.96$ allora rifiutiamo l'ipotesi nulla

5. A partire dal campione calcolo il valore della statistica sotto l'ipotesi nulla

$$Z_{stat} = \frac{5.1 - 4.5}{\frac{1.2}{\sqrt{25}}} = 2.50$$

6. Decisione

Poichè $Z_{stat} = 2.50 > 1.96$ allora rifiutiamo l'ipotesi nulla. C'è evidenza sufficiente per affermare che il tempo medio di attesa per ordinare è significativamente cambiato nel tempo: il manager conclude che oggi, in media, il cliente aspetta più a lungo per ordinare rispetto al passato

6.bis Decisione in base al p-value del test

Definizione:

Il p-value o livello di significatività osservato è la probabilità di osservare valori della statistica test meno favorevoli ad H_0del valore effettivamente ottenuto. In altre parole, il p-value è il più piccolo livello di significatività tale da indurre al rifiuto dell'ipotesi nulla sulla base dei dati osservati. Esso 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. Se il p-value<α, rifiuto quindi l'ipotesi nulla.

Nel nostro caso, essendo il test bidirezionale:

$$2 * P\left(|Z_{stat}| > Z_{stat|H_o}\right) = 2 * P(|Z_{stat}| > 2.50) = 2(1 - 0.9938) = 0.0124$$

Siccome il p-value < α=0.05, allora rifiuto l'ipotesi nulla.

Esercizio 2. Test delle ipotesi per la media, varianza ignota

E' noto che su un certo tratto stradale del Comune di Roma la velocità media delle auto è di 60 km/h. Tuttavia, la polizia municipale sospetta che lungo quel tratto, la velocità media dei veicoli sia diventata superiore. A partire da un campione di 25 autovetture si è registrata effettivamente una velocità media di 64.8 Km/h con uno scarto quadratico medio campionario di 23.2 km/h.

a)Sottoporre a verifica delle ipotesi con un livello di significatività del 5% l'ipotesi nulla (ossia che la velocità media su quel tratto di strada sia di 60 Km/h) contro l' ipotesi alternativa (ossia la velocità media sia superiore a 60 km/h).

Sol.

a)

1.*Definizione del sistema di ipotesi:*

$$H_0 : \mu = 60$$
$$H_1 : \mu > 60$$

2. Livello di significatività

$$\alpha = 0.05$$

3. Costruzione della statistica test

Siccome la varianza della popolazione è ignota, allora:

$$T_{stat} = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} \sim t_{n-1}^{\alpha}$$

Il livello di significatività è $\alpha = 0.05$, il test è a una coda, quindi i quantili della distribuzione t di student da individuare sono:

$$t_{24,0.05} = 1.711$$

4. *Definizione della regola di decisione (regione di rifiuto)*

Se $T_{stat} \geq 1.711$ allora rifiutiamo l'ipotesi nulla

5. A partire dal campione calcolo il valore della statistica sotto l'ipotesi nulla

$$T = \frac{64.8 - 60}{23.2 / \sqrt{25}} = 1.0344$$

6. Decisione

Poichè $T_{stat} = 1.0344 < 1.711$ allora non rifiutiamo l'ipotesi nulla.

Esercizio 3. Test delle ipotesi per la proporzione

Da un'indagine condotta sulla composizione del management delle aziende italiane è risultato che il 35% delle aziende italiane è gestito da donne. Inoltre, in un campione di n = 100 aziende localizzate nel sud Italia è risultato che 15 sono gestite da donne. Verificare, al livello α = 0.1, se il campione può ritenersi rappresentativo della realtà aziendale italiana. Determinare il livello di significatività (p-value) del test

Sol

1.Definizione del sistema di ipotesi:

$H_0 : p = 0.35$
$H_1 : p \neq 0.35$

2. Livello di significatività

$\alpha = 0.01$

3. Costruzione della statistica test

Sfrutto il TLC, infatti n è sufficientemente grande per cui la statistica test è:

$$Z_{stat} = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \rightarrow N(0,1)$$

Il livello di significatività è $\alpha = 0.01$, il test è a due code, quindi il quantile della distribuzione Normale standard da individuare è:

$z_{0.995} = 2.575$

4. Definizione della regola di decisione (regione di rifiuto)

Se $Z_{stat} \geq 2.575$ oppure $Z_{stat} \leq -2.575$ allora rifiutiamo l'ipotesi nulla

Oppure, direttamente $|Z_{stat}| > 2.575$

5. A partire dal campione calcolo il valore della statistica sotto l'ipotesi nulla

$$Z_{stat} = \frac{0.15 - 0.35}{\sqrt{\frac{0.35(1-0.35)}{100}}} = -4.19$$

6. Decisione

Poichè $|Z_{stat}| = 4.19 > 2.575$ si rifiuta l'ipotesi che il campione di aziende osservato sia rappresentativo della realtà aziendale italiana.

Esercizio 4. Test sulla proporzione, errore di I e II tipo, potenza del test

Un produttore di sementi afferma che almeno il 90% dei semi germoglia. Si sottopone a test l'affermazione del produttore al livello di significatività dell' 1%:

a) determinare l'errore di I tipo

b) determinare la potenza del test quando l'ipotesi alternativa afferma che la percentuale di semi che germogliano è 85% e la numerosità campionaria è 220.

d) come varierebbe la potenza del test se l'ipotesi alternativa assumesse che la percentuale di semi che germogliano è 80%?

Sol.

a) L'errore di I tipo (rifiutare l'ipotesi nulla quando è vera) è pari al livello di significatività fissato per il test quindi α=0.01

b) Per calcolare la potenza del test occorre ricordare che l'errore di II tipo è la probabilità di accettare l'ipotesi nulla quando è falsa. Esso è quantificabile solo fissando un valore per il parametro su cui si intende eseguire il test, ossia fissando una distribuzione alternativa differente da quella specificata sotto l'ipotesi nulla.

1.Definizione del sistema di ipotesi: (iniziale)

$$H_0 : p = 0.90$$
$$H_1 : p < 0.90$$

2. Livello di significatività

$$\alpha = 0.01$$

3. Costruzione della statistica test (sotto l'ipotesi nulla)

Sfrutto il TLC, infatti n è sufficientemente grande per cui la statistica test è:

$$Z_{stat} = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \to N(0,1)$$

Il livello di significatività è $\alpha = 0.01$, il test è unidirezionale, quindi il quantile della distribuzione Normale standard è: $z_{0.99} = -2.33$, per cui se $Z_{stat} < -2.33$ rifiuto l'ipotesi nulla

A partire da questo si determina la potenza del test intesa come probabilità di rifiutare correttamente H_0, in pratica la probabilità di prendere la "decisione giusta":

$$P\left(Z_{(\mathrm{Lim})|H_1} < z\right) = 1 - P\left(Z_{(\mathrm{Lim})|H_1} > z\right) = 1 - \beta$$

dapprima determiniamo il valore di $\bar{p}$ (valore campionario) estremo che avvalora l'ipotesi nulla, ossia $\bar{p}_0$. Esso corrisponde al valore "limite" tra la regione di accettazione e la regione di rifiuto. In particolare:

$$Z_{(limite)} = \frac{\bar{p}_0 - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = z_{0.99} = -2.33$$

da cui:

$$\bar{p}_0 = p_0 - 2.33\sqrt{\frac{p_0(1-p_0)}{n}} = 0.90 - 2.33(0.02) = 0.853$$

Dunque, H_0 è falsa a partire dal valore della statistica uguale a 0.853.

Considerando l'ipotesi alternativa proposta per p, ossia $p_A = 0.85$, il valore della statistica test sotto l'ipotesi alternativa è:

$$Z_{(lim1)} = \frac{\bar{p}_0 - p_A}{\sqrt{\frac{p_A(1-p_A)}{n}}} = \frac{0.853 - 0.85}{\sqrt{\frac{0.85(1-0.85)}{220}}} = 0.125$$

La probabilità di accettare l'ipotesi nulla quando è falsa (errore di II specie) è dunque:

$\beta = P\left(Z_{(\mathrm{lim}\ 1)} > 0.125\right) = 0.4522$

Per cui il complemento a 1 dell'errore di II specie (1-β) è la potenza del test :

$$(1-\beta) = 1 - P\left(Z_{(lim1)} > 0.125\right) = P\left(Z_{(lim1)} < 0.125\right) = 0.5478$$

c) Seguendo lo stesso tipo di ragionamento, con l'ipotesi alternativa $p_A = 0.80$ il valore della statistica test è:

$$Z_{(lim2)} = \frac{\bar{p}_0 - p_A}{\sqrt{\frac{p_A(1-p_A)}{n}}} = \frac{0.853 - 0.80}{\sqrt{\frac{0.80(1-0.80)}{220}}} = 1.96$$

$\beta = P(Z_{(lim2)} > 1.96) = 0.025$

$(1 - \beta) = 1 - 0.025 = 0.975$

Assumendo $p_A = 0.80$ la potenza del test aumenta. Quindi aumenta la probabilità di prendere la "decisione giusta".

Note utili:

- Maggiore è la distanza tra ipotesi alternativa (che stiamo considerando vera) e l'ipotesi nulla, maggiore sarà la potenza del test.
- Minore è la dispersione della variabile, minore sarà la varianza della proporzione campionaria, più strette saranno le corrispondenti distribuzioni, e maggiore sarà la potenza del test
- Maggiore è l'α prescelto, maggiore sarà la potenza del test. Questo è logico perchè α e β sono tra loro interdipendenti
- La potenza del test è influenzata dalla dimensione campionaria. A parità di α, la deviazione standard della proporzione campionaria è inversamente proporzionale alla numerosità.

Esercizio 5. Test della differenza tra proporzioni

Un'azienda automobilistica, prima di immettere sul mercato un nuovo modello di un'auto già in commercio, realizza un sondaggio di opinioni. In particolare, l'indagine rivela che su un campione di n1 = 100 donne il 36% preferisce il nuovo modello di auto rispetto a quello già in commercio mentre, su un campione di n2 = 100 uomini solo il 25% preferisce il nuovo modello. Verificare, al livello α = 0.01, l'ipotesi che non ci sia differenza nelle preferenze in base al sesso dei potenziali acquirenti.

Sol.

1. Sistema di ipotesi da verificare:

$H_0: p_1 - p_2 = 0$

$H_1: p_1 - p_2 \neq 0$

2. Livello di significatività

α=0.01

3. *Costruzione della statistica test (sotto l'ipotesi nulla)*

$$Z = \frac{(\bar{p}_1 - \bar{p}_2)}{\sqrt{p(1-p)\frac{1}{n_1} + \frac{1}{n_2}}} \rightarrow N(0,1)$$

4. *Definizione della regola di decisione*

Il test è bidirezionale, dalle tavole della normale standard risulta

$z_{0.995} = \pm 2.57$

Se $|Z| > 2.575$ rifiutiamo l'ipotesi nulla

5. *A partire dal campione calcolo il valore della statistica test sotto l'ipotesi nulla*

La stima della proporzione, comune ad entrambe le popolazioni, di soggetti che preferiscono il nuovo modello è ottenuta da:

$$p = \frac{\bar{p}_1 n_1 + \bar{p}_2 n_2}{n_1 + n_2} = \frac{0.36 * 100 + 0.25 * 100}{100 + 100} = 0.305$$

Per cui il valore della statistica campionaria risulta:

$$Z = \frac{(0.36 - 0.25)}{\sqrt{0.305(1 - 0.305)\frac{1}{100} + \frac{1}{100}}} = 1.68$$

6. Decisione

Siccome $|Z| = 1.68 < 2.575$ non rifiutiamo l'ipotesi nulla. Esiste indipendenza dal sesso rispetto alle preferenze dei potenziali acquirenti