SIMULAZIONE ESAME PSICOMETRIA 1

Esercizio 1

Nel 2017 è stata condotta una ricerca sull'uso dei trasporti pubblici. Ad un campione casuale di 1400 studenti dell'Università di Trieste è stato chiesto se preferivano che venisse ridotto il costo del biglietto dell'autobus a fronte di una diminuzione dell'orario di circolazione dei mezzi pubblici, oppure aumentato il costo del biglietto dell'autobus a fronte di un'estensione dell'orario di circolazione. Il 53.2% ha risposto "ridurre il costo del biglietto", il 46.8% ha risposto "aumentare il costo del biglietto".

a) verifica se la proporzione della popolazione è a favore di una o dell'altra soluzione (con livello di significatività α=0.05); esplicita tutti i passaggi e interpreta il risultato
b) riporta e interpreta il p-valore per H_a: $\prod < 0.50$ (con livello di significatività α=0.05)
c) riporta e interpreta il p-valore per H_a: $\prod > 0.50$ (con livello di significatività α=0.05)

Soluzione

a) test di significatività per una proporzione

notazione

$\prod$ = proporzione di popolazione che ha risposto "ridurre il costo del biglietto"

$\prod$-cappello (proporzione campionaria) = 0.532

n = 1400

assunzioni

- casualità → campione casuale
- dimensione campionaria sufficientemente ampia in modo che la distribuzione della proporzione campionaria sia approssimativamente normale

ipotesi

H_0: $\prod = 0.50$

H_a: $\prod \neq 0.50$

test

z = ($\prod$-cappello - 0.50)/se_0

errore standard sotto l'ipotesi nulla, $se_0 = \sqrt{\prod_0 (1 - \prod_0)/n} = \sqrt{(0.50 \times 0.50) / 1400} = 0.013$

z = (0.532 - 0.50)/0.013 = 2.46

p-valore

per ipotesi alternativa bidirezionale → p = 2 x (0.0069) = 0.0138

conclusioni e interpretazione risultato

α=0.05

p < α → rifiuto H_0 in favore di H_a

La proporzione della popolazione è a favore di una delle due soluzioni.

b) con H_a: $\prod < 0.50$

z = 2.46, p = 1 - 0.0069 = 0.9931, p > α → <u>non</u> rifiuto H_0

c) con H_a: $\prod > 0.50$

z = 2.46, p = 0.0069, p < α → rifiuto H_0

Esercizio 2
Nella tabella sono riportati i dati ottenuti confrontando due gruppi di partecipanti di età 21-30 e 31-40 rispetto al numero di numero di ore al giorno in cui i partecipanti usano il cellulare:

gruppo	N	media	s
21-30	1200	4.20	1.60
31-40	945	3.40	2.40

a) costruisci l'intervallo di confidenza al 95% per il confronto di medie; cosa puoi dedurre?
b) conduci un test di significatività per analizzare se le medie di popolazione differiscono tra i partecipanti di età 21-30 e 31-40. Riporta le conclusioni per il livello di significatività α=0.05

Soluzione
notazione
μ_1 = media giornaliera del numero di ore di utilizzo del cellulare per la popolazione 21-30 anni
μ_2 = media giornaliera del numero di ore di utilizzo del cellulare per la popolazione 31-40 anni
$\bar{y}_1 = 4.20$ $n_1 = 1200$ $s_1 = 1.60$
$\bar{y}_2 = 3.40$ $n_2 = 945$ $s_2 = 2.40$

a) 95% IC per $\mu_1 - \mu_2$ ha forma stima puntuale ± margine d'errore: $(\mu_1 - \mu_2) \pm t(se)$
$\bar{y}_1 - \bar{y}_2 = 4.20 - 3.40 = 0.80$
$se = \sqrt{s_1^2 / n_1 + s_2^2 / n_2} = \sqrt{(1.60)^2/1200 + (2.40)^2/945} = 0.09$
essendo i campioni molto grandi il t-score per il margine d'errore può essere sostituito dall'uso dello z-score (per un IC al 95%: z=1.96)
95% IC per $\mu_1 - \mu_2 = 0.80 \pm 1.96(0.09) = 0.80 \pm 0.18$ (0.62, 0.98)
Nell'intervallo di confidenza al 95% IC per $\mu_1 - \mu_2$ sono presenti solamente valori positivi, quindi è plausibile che $\mu_1 > \mu_2$ → con una fiducia del 95% possiamo ritenere che la media giornaliera del numero di ore di utilizzo del cellulare sia più elevata per la popolazione 21-30 anni che per la popolazione 31-40 e che la differenza sia compresa tra 0.62 e 0.98 ore.

b) test di significatività per due medie - campioni indipendenti
assunzioni
- casualità → campione casuale
- distribuzione normale nella popolazione

ipotesi
H_0: $\mu_1 - \mu_2 = 0$
H_a: $\mu_1 - \mu_2 \neq 0$
test
$t = (\bar{y}_2 - \bar{y}_1)/se = 0.80/0.09 = 8.89$
In grandi campioni, la distribuzione t è essenzialmente uguale a quella normale, t=8.89 è un valore molto grande perciò il **p-valore** sarà uno 0 seguito da molti 0 (il valore 8.89 non lo troviamo nella tavola)
conclusioni e interpretazione risultato
α=0.05, sicuramente $p < \alpha$ → rifiuto H_0 in favore di H_a

Quesito teorico

a) riporta la definizione di "intervallo di confidenza per un parametro" e definisci che cos'è il "livello di fiducia".

b) l'ampiezza di un intervallo di confidenza si modifica al variare del livello di fiducia? se si, come?

c) in un sondaggio è stato chiesto agli intervistati "Per quante ore al giorno guardi la tv?". I dati analizzati tramite software statistico hanno riportato i seguenti risultati:

Variabile	N	media	SE media	IC 95%
ore tv	892	2.76	0.08	(2.60, 2.93)

l'interpretazione "in una lunga serie di osservazioni, il 95% delle volte i soggetti hanno guardato la tv per un intervallo da 2.60 a 2.93 ore al giorno" è corretta? se no, fornisci l'interpretazione corretta

Soluzione

a) un intervallo di confidenza per un parametro è un intervallo di valori entro cui si ritiene ricada il valore di detto parametro. La probabilità associata al fatto che l'intervallo contenga il parametro è denominata livello di fiducia (solitamente è un numero prossimo a 1, come 0.95 o 0.99).

b) l'ampiezza di un intervallo di confidenza si modifica al variare del livello di fiducia: al crescere del livello di fiducia, cresce anche l'ampiezza di un intervallo di confidenza (l'ampiezza di un intervallo di confidenza invece decresce al crescere della dimensione campionaria).

c) no. La corretta interpretazione è: possiamo avere una fiducia al 95% che la media della popolazione sia compresa tra 2.60 e 2.93 ore/giorno (l'intervallo di confidenza è una previsione della media della popolazione e non dei valori che vengono assunti da y per i singoli soggetti).

Domanda breve 1

Una ricercatrice sottopone un partecipante a due compiti atti a valutare l'attenzione: il test TMT ($\mu = 22$, $\sigma = 3$) ed il test delle matrici attentive ($\mu = 70$, $\sigma = 7$). In entrambi i compiti si è più bravi se si ottiene un punteggio più alto. Il partecipante ottiene un punteggio pari a 23.2 al test TMT e pari a 74.9 al test delle matrici attentive.
È possibile stabilire in quale dei due compiti è più bravo il partecipante? se sì, spiega come e trai le conclusioni.

Soluzione

Sì, confrontando a quante deviazioni standard dalla rispettiva media ricade ciascun punteggio.

TMT: $\mu_{TMT} = 22$

$\sigma_{TMT} = 3$

$y_{TMT} = 23.2$

MAT: $\mu_{TMT} = 70$

$\sigma_{TMT} = 7$

$y_{MAT} = 74.9$

$z_{TMT} = (y_{TMT} - \mu)/\sigma = (23.2 - 22)/3 = 0.4$

$z_{MAT} = (y_{TMT} - \mu)/\sigma = (74.9 - 70)/7 = 0.7$

Il partecipante è più bravo al test delle matrici attentive, poiché il suo punteggio ricade a 0.7 deviazioni standard dalla media, mentre il punteggio ottenuto al test TMT ricade a 0.4 deviazioni standard dalla media.

Domanda breve 2

È stato stimato che nel mese di aprile la probabilità che piova a Gorizia è 0.5. Assumendo che la piovosità nei 30 giorni che compongono il mese sia indipendente e considerando la prima settimana del mese (7 giorni), calcolare la probabilità che:

a) piova almeno un giorno
b) piova almeno un giorno si e uno no

Soluzione

Sia X la variabile aleatoria che definisce il numero di giorni in cui piove

a) piova almeno un giorno su 7

$r \geq 1$, $n = 7$, $p = 0.5$, $q = 1- p = 0.5$

$P(X\geq1) = 1 - P(X=0) = 1 - (0.0078) = 0.9922$

con $P(X=0) = (0.5)^7 = 0.0078$

La probabilità che piova almeno un giorno è pari a 0.9922.

b) piova almeno un giorno si e uno no

$r \geq 1 \cup r < 7$, $n = 7$, $p = 0.5$, $q = 0.5$

$P(X\geq1 \cup X<7) = 1 - P(X=0) - P(X=7) = 0.9922 - 0.0078 = 0.9844$

con $P(X=7) = (0.5)^7 = 0.0078$

La probabilità che piova almeno un giorno si e uno no è pari a 0.9844.