

AMSB 17/9/2018 AA1718 e precedenti

Si ricorda nell'effettuare un test di specificare ogni volta le ipotesi nulla e alternativa.

Es 1 (tutti) Si considerino i seguenti dati

-3.7 4.3 -4.6 -1.4 3.5 2.2 5.3 -0.2 1.2 3.2 3.5 -2.1

- 1) Si costruisca l'istogramma suddividendo in 4 classi $[-6, -3[$, $[-3, 0[$, $[0, 3[$, $[3, 6[$. Se ne faccia il grafico. Si stimino la media pesata e la varianza pesata
- 2) Si stimino due parametri che potrebbero essere usati come indicatori per verificare la gaussianità della distribuzione dei dati (per questo punto non è necessario fare riferimento ai dati suddivisi in classi)

Nel calcolo della media e delle varianze pesate devono essere presi in considerazione i valori centrali come rappresentativi di ogni classe.

Alcuni parametri che potrebbero essere utili per verificare la gaussianità, in prima approssimazione, possono essere la curtosi e la skewness.

Es 2 (tutti) Si vuole verificare se i valori di connettività funzionale tra due aree cerebrali (PDC), durante due compiti cognitivi, siano significativamente diversi. Si esegue un esperimento su due gruppi di soggetti di 6 e 7 persone. A ciascuno dei due gruppi si fa eseguire uno dei test. Si ipotizza che i valori osservati seguano una distribuzione di tipo Chi quadro con 10 gdl. Si adotti un livello di significatività pari allo 0.05

PDC gruppo 1

0.90 0.82 0.74 0.35 0.65 0.55

PDC gruppo 2

0.96 0.95 0.44 0.57 0.25 0.72 0.81

Bisogna utilizzare un test non parametrico. In questo caso abbiamo dati indipendenti, quindi il test è di Mann-Whitney. La somma dei ranghi è pari a 41, che ricade all'interno della regione di accettazione (27,57). Quindi si accetta l'ipotesi nulla con significatività pari a 0.05

Es 3 (tutti) Si vuole stimare il tempo di assemblaggio di un dispositivo. Si misura il tempo impiegato da 120 lavoratori, osservando un tempo medio pari a 16.2 minuti. Sapendo che la deviazione standard è pari a 3.6 minuti, si fornisca una stima al 92% del tempo medio di assemblaggio.

Quanto dovrebbe essere grande il campione per avere un intervallo di confidenza al 92% ampio 30 secondi?

La deviazione standard è nota, quindi si usa una statistica z. L'intervallo di confidenza al 92% è (15.6, 16.8). Per avere un'ampiezza dell'intervallo pari a 30 secondi, il numero di lavoratori utilizzati per avere la stima deve essere superiore a 207.

Es 4 (tutti) In un test per la valutazione della piacevolezza del parlato vengono fatti valutare 5 voci a 800 persone. Nella tabella sono indicate le percentuali dei soggetti che hanno indicato la voce i-esima come più piacevole. È possibile dire che l'indicazione di piacevolezza sia stata uniforme? ($\alpha = 0.05$)

Campione Vocale	%
1	15
2	25
3	20
4	12
5	28

Test del chi quadro. Le frequenze attese sono quelle nel caso di distribuzioni uniforme e pari allo 20%.

Importante: non si deve lavorare con le percentuali ma con le frequenze, in termini di numerosità delle varie classi. Quindi, la frequenza osservata per la classe 1 è $0.15 \cdot 800 = 120$, la frequenza attesa è 160. La statistica finale, senza correzione di Yates (questa sarebbe opportuna, ma non è stato penalizzato chi non l'ha fatta), è pari a 71.2.

I gradi di libertà sono 3 per cui il valore critico è pari a 6. Quindi rifiutiamo l'ipotesi di nulla di distribuzione uniforme con un $\alpha = 0.05$.

Es 5 (tutti) Si vuole verificare se la frequenza cardiaca in un soggetto impegnato in un compito stressogeno, sia modificata dalla presenza di animali domestici o amici. Il test viene eseguito su 45 persone divisi in tre gruppi di uguale numerosità (controlli-da soli, con animale domestico, con un amico). Si utilizzi un livello di significatività pari a 0.05. Si studi nel dettaglio l'origine di eventuali differenze

Gruppo	Freq. Card. Media	Dev. Std.
Controllo	82.52	9.24
Animali	73.48	9.97
Amici	91.33	8.34

Bisogna usare un test anova. La statistica F è pari a 14.09. Il valore critico di F per $\alpha = 0.05$ è pari a 3.23 (approssimato con gdl denominatore pari a 40). Quindi si rifiuta l'ipotesi nulla.

Riguardo al test di Bonferroni si ricorda di porre attenzione ai gradi di libertà (42) e alla riduzione della significatività dividendola per il numero di test (in questo caso 3)

Es 6 (3 cfu) Si vuole verificare se esista una relazione di tipo lineare tra due variabili x e y. Nella tabella sono riportate le misure di y in funzione dei valori di x. Fare il grafico del modello, includendo le osservazioni, e determinarne la significatività $\alpha = 0.05$. Stimare inoltre il coefficiente di correlazione a partire dal coefficiente di determinazione e dall'analisi della relazione.

x	y		
2	12.2	13.0	
3	10.9	11.6	10.0
4	10.3	9.2	

Statistica della regressione

<i>R multiplo</i>	<i>0.878</i>
<i>R al quadrato</i>	<i>0.771</i>
<i>R al quadrato corretto</i>	<i>0.725</i>
<i>Errore standard</i>	<i>0.695</i>
<i>Osservazioni</i>	<i>7.000</i>

ANALISI VARIANZA

	<i>gdl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>Significatività F</i>
<i>Regressione</i>	<i>1</i>	<i>8.1225</i>	<i>8.1225</i>	<i>16.8392</i>	<i>0.0093</i>
<i>Residuo</i>	<i>5</i>	<i>2.4118</i>	<i>0.4824</i>		
<i>Totale</i>	<i>6</i>	<i>10.5343</i>			

	<i>Coefficienti</i>	<i>Errore standard</i>	<i>Stat t</i>
<i>Intercetta</i>	<i>15.304</i>	<i>1.074</i>	<i>14.245</i>
<i>Variabile X 1</i>	<i>-1.425</i>	<i>0.347</i>	<i>-4.104</i>

Importante riportare il segno corretto del coefficiente di correlazione, quest'ultimo determinato a partire dallo r^2 , andando a vedere il segno della pendenza della retta di regressione.

D1 (2cfu)

Si descrivano le ipotesi che devono essere verificate per la validità del test anova.

D1 (solo per coloro che hanno seguito nel AA 2017_2018) Si consideri un'urna contenente 10 palline rosse e 5 nere. Si utilizzi la distribuzione binomiale per calcolare la probabilità che 3 palline siano nere, effettuando una estrazione con reintroduzione di 5 palline

D2 (tutti)

Si discuta un metodo per la stima della gaussianità dei dati. Si sottolinei la rilevanza di tale tipo di test nell'ambito dell'analisi statistica.

D3 (3cfu)

Si definisca la funzione di distribuzione per una variabile aleatoria gaussiana. Si faccia il grafico della funzione di distribuzione gaussiana con valore medio pari a 3 e deviazione standard pari a 5. Usando le tabelle a disposizione di stimi la probabilità che tale variabile assuma valori:

- a) maggiori di 10.5 b) compresi tra 10.5 e 13

La probabilità a) è pari a 0.6%

La probabilità b) è pari a 1.7%

D4 (3 cfu)

Si discuta cosa si intende per stimatore non polarizzato e consistente. Si descriva uno stimatore che sia polarizzato, ma consistente.