

AMSB 7/1/2019 AA1819

Es 1 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 nel numero opportuno di classi, giustificando la scelta
- 2) Si stimino frequenze assoluta e relativa
- 3) Stimare media e varianze pesate
- 4) 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)

La scelta del numero di classi può essere fatta con la regola di Sturges.

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 Si vuole verificare se l'effetto sulla durata del sonno di due farmaci sia lo stesso. In 10 soggetti si monitorano le ore medie giornaliere dopo l'assunzione dei due farmaci (1 mese di osservazione per ogni farmaco). Nella tabella sono mostrate le differenze delle ore medie giornaliere di sonno per ogni soggetto. Si usi una significatività pari a $\alpha = 0.05$.

| Soggetto | Farmaco 1 | Farmaco 2 |
|----------|-----------|-----------|
| 1 | 1.9 | 0.9 |
| 2 | -1.6 | 0.8 |
| 3 | -0.2 | 1.1 |
| 4 | -1.2 | 0.1 |
| 5 | -0.1 | -0.1 |
| 6 | 3.4 | 4.6 |
| 7 | 3.7 | 3.9 |
| 8 | 0.8 | 1.6 |
| 9 | 0.0 | 4.6 |
| 10 | 2.0 | 2.4 |

Bisogna usare il test di Wilcoxon, quindi dei ranghi con segno. Infatti non abbiamo indicazioni circa la gaussianità dei dati e i dati sono appaiati. Potremmo utilizzare un test t per dati appaiati, ma assumendoci dei rischi circa la validità del test. Nell'effettuare il test di Wilcoxon una differenza è nulla quindi i gradi di libertà sono 9.

La statistica risultante è 37, mentre il valore critico è 33, quindi si rifiuta l'ipotesi nulla.

Es 3 Per verificare l'affidabilità della misura di una bilancia, vengono eseguite quattro misure usando un peso certificato di un grammo. Le misure ottenute sono 0.98, 1.02, 1.01 e 0.95 grammi. Creare l'intervallo di confidenza del peso misurato al 95%. Il risultato ottenuto di può dire qualcosa sull'affidabilità della bilancia? Quali operazioni potremmo fare se volessimo ridurre l'ampiezza dell'intervallo di confidenza?

L'intervallo di confidenza al 95% risulta essere (0.94, 1.04)

L'affidabilità della bilancia può essere discussa usando due valutazioni:

a priori osservando l'errore standard, proporzionale all'ampiezza dell'intervallo, quindi dovremmo decidere se questo parametro può essere considerato sufficiente o meno, a posteriori, considerando il fatto che in realtà conosciamo il valore vero, e il fatto che questo ricada nell'intervallo di confidenza ci porta a pensare che il bias, ovvero l'errore sistematico della bilancia, sia limitato.

Es 4 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 uniformi 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 Nella produzione di un composto vengono testati 4 differenti catalizzatori. Si vuole verificare se i 4 differiscano riguardo alla concentrazione finale di un determinato composto, e nel caso di verificare quale/i siano responsabili delle eventuali differenze. Nella tabella seguente sono espresse le concentrazioni del composto in funzione del catalizzatore, in un numero di prove effettuate. Si considerino i dati a distribuzione gaussiana e si utilizzi un livello di significatività pari a 0.05. Specificare le ipotesi nulla e alternativa.

| Cat 1 | Cat 2 | Cat 3 | Cat 4 |
|-------|-------|-------|-------|
| 58.2 | 56.3 | 50.1 | 52.9 |
| 57.2 | 54.5 | 54.2 | 49.9 |
| 58.4 | 57.0 | 55.4 | 50.0 |
| 55.8 | 55.3 | | 51.7 |
| 54.9 | | | |

Test Anova.

| Gruppi | Conteggio | Somma | Media | Varianza |
|-----------|-----------|--------|-------|----------|
| Colonna 1 | 5 | 284.50 | 56.90 | 2.31 |
| Colonna 2 | 4 | 223.10 | 55.78 | 1.21 |
| Colonna 3 | 3 | 159.70 | 53.23 | 7.72 |
| Colonna 4 | 4 | 204.50 | 51.13 | 2.08 |

ANALISI VARIANZA

| <i>Origine della variazione</i> | <i>SQ</i> | <i>gdl</i> | <i>MQ</i> | <i>F</i> | <i>F crit</i> |
|---------------------------------|-----------------|------------|-----------|----------|---------------|
| Tra gruppi | 85.68 | 3.00 | 28.56 | 9.92 | 3.49 |
| In gruppi | 34.56 | 12.00 | 2.88 | | |
| Totale | 120.2375 | 15 | | | |

Riguardo al test di Bonferroni si ricorda di porre attenzione ai gradi di libertà e alla riduzione della significatività dividendola per il numero di test (in questo caso 6)

Es 6 Si vuole creare una commissione di 10 persone selezionate tra 100. Considerando che nel gruppo di 100 persone sono presenti 40 uomini e 60 donne, si dica:

- 1) In quanti modi diversi può essere scelta la commissione
- 2) In quanti modi diversi può essere scelta la commissione se deve essere formata da 5 donne e 5 uomini
- 3) Qual è la probabilità che una commissione scelta a caso sia formata da 5 uomini e 5 donne

- 1) $C_{100,10}$
- 2) $C_{60,5}C_{40,5}$
- 3) $C_{60,5}C_{40,5} / C_{100,10}$

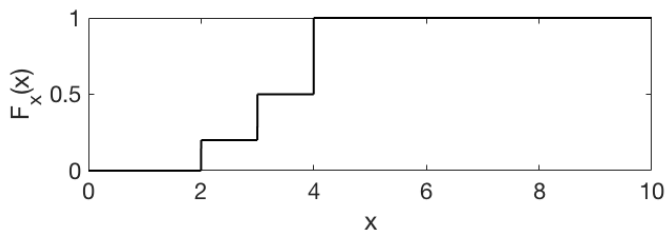
D1 Descrivere uso e forma della distribuzione binomiale. Specificare in quali condizioni sia possibile approssimarla con una variabile gaussiana: indicarne in questo caso valore atteso e varianza.

Risolvere il seguente esercizio: si considerano 6 testimoni, che devono identificare un assassino tra 4 sospetti (uno di questi siete voi). Supponendo che la scelta dei testimoni sia in realtà casuale, calcolare le seguenti probabilità:

- 1) Probabilità di non essere identificato
- 2) Probabilità di essere identificato una volta
- 3) Probabilità di essere identificato più di una volta

D2 Si discuta il significato di probabilità condizionata, di probabilità a priori e a posteriori. Si descriva e discuta il Teorema di Bayes. Si descriva il teorema delle probabilità totali e si faccia un esempio che ne dimostri l'utilità.

D3 Nella figura è mostrata la funzione di distribuzione di una variabile aleatoria x . Si faccia il grafico della densità di probabilità corrispondente e si discuta di quale tipo di variabile aleatoria si tratti. Inoltre si calcolino le probabilità che x assuma valori nei seguenti intervalli: a) $X < 5$ b) $3 < X < 5$



Questa distribuzione è relativa ad una variabile aleatoria discreta che può assumere valori 2, 3 e 4 con masse di probabilità pari a 0.2, 0.3 e 0.5, rispettivamente.

La probabilità a) è pari a 1. La probabilità b) è pari alla probabilità che la variabile assuma valori pari a 4 ed è quindi 0.5.

D4 Sia data la variabile aleatoria così definita

$$\begin{cases} y = \alpha x & \text{per } x \geq 0 \\ y = 0 & \text{per } x < 0 \end{cases}$$

con $\alpha > 0$ e x variabile aleatoria con densità di probabilità nota. Descrivere le relazioni tra le densità di probabilità e le funzioni di distribuzione di x e y .

La densità di probabilità di y in funzione di quella di x è

$$f_y(y) = \begin{cases} \frac{f_x(y/\alpha)}{\alpha} & \text{per } y > 0 \\ c\delta(y) & \text{per } y = 0 \text{ dove } c = \int_{-\infty}^0 f_x(x)dx \\ 0 & \text{altrove} \end{cases}$$

La funzione di distribuzione di y in funzione di quella di x è

0 per $y < 0$

$c + F_x(y/\alpha)$ per $y \geq 0$

Questo esercizio è stato valutato positivamente chi ha indicato la soluzione della densità di y per $y > 0$ e chi almeno si è posto il problema relativamente ai valori della densità per $y \leq 0$