

Libera Università di Bolzano – Facoltà di Scienze e Tecnologie
Corso di Matematica e Statistica – Modulo Statistica
A.A. 2009/10 – Prof. Paolo Coletti
Esercizi

In preparazione dell'esame si consiglia di svolgere i Problemi Guida e gli Esercizi non risolti del libro consigliato. Per gli esercizi da svolgere con l'aiuto di Microsoft Excel e per gli esercizi della parte relativa alla dispensa, vengono proposti in questo documento degli esercizi guida.

DURANTE L'ESAME È OBBLIGATORIO PER TUTTI GLI ESERCIZI SCRIVERE, OPPURE SALVARE NEL FILE EXCEL, I CALCOLI EFFETTUATI.

Esercizi che richiedono l'aiuto di Microsoft Excel

Esercizio 1

La probabilità che una mela sia danneggiata è del 4%. Prelevando un campione di 30 mele, scrivi qual è la probabilità che le mele danneggiate siano 3 o meno.

Esercizio 2

Scrivi il numero di combinazioni (l'ordine non è importante) che si possono ottenere estraendo 12 oggetti da un insieme di 16 oggetti diversi.

Esercizio 3

Scrivi tutti i coefficienti binomiali per $n=10$.

Esercizio 4

Considera questi dati discreti: 3; 4; 4; 4; 12; 12; 13; 13; 13; 17; 17; 17; 17. Scrivi media, mediana, moda, varianza, deviazione standard, coefficiente di variazione in percentuale arrotondato a due cifre decimali. Scrivi perché la media differisce dalla mediana. Produci su Excel un grafico appropriato che rappresenti i dati.

Esercizio 5

Considera una distribuzione discreta uniforme con 10 valori (a tua scelta). Produci su Excel un grafico appropriato che rappresenti i dati.

Esercizio 6

Considera queste misure continue di altezze: 3,0; 4,2; 4,3; 4,4; 12,1; 12,5; 13,9; 13,9; 13,9; 17,0; 17,1; 17,3; 17,4. Scrivi media, mediana, varianza, deviazione standard, coefficiente di variazione in percentuale arrotondato a 4 cifre decimali. Produci su Excel un grafico appropriato che rappresenti i dati.

Esercizio 7

Su un foglio Excel produci una colonna di 300 numeri casuali tra 7 e 81. Supponi che siano continui. Produci su Excel un grafico appropriato che rappresenti i dati.

Esercizio 8

Considera questi dati discreti: 3; 4; 4; 4; 12; 12; 13; 13; 13; 17; 17; 17; 17. Calcola la media pesata usando un peso pari a 7 per i primi 10 numeri e pari a 2 per gli altri.

Esercizio 9

Considera i dati del foglio Excel. Produci su Excel uno scatterplot per i dati indicati con x e y usando pallini rossi e cambiando l'intervallo orizzontale e verticale in modo che i dati occupino tutto il grafico. Aggiungi la retta di regressione lineare con equazione e R-quadro. Scrivi il valore teorico per $x=19$ e per $x=287$.

Esercizio 10

Considera i dati del foglio Excel. Produci su Excel uno scatterplot con asse verticale logaritmico per i dati indicati con w e z e una funzione di regressione esponenziale con equazione e R-quadro. Scrivi il coefficiente di correlazione lineare.

Esercizio 11

Considera i dati del foglio Excel. Produci su Excel uno scatterplot per i dati indicati con x e y e una retta di regressione con equazione. Per ogni dato, calcola su Excel l'errore di interpolazione (residuo). Scrivi la media e la varianza degli errori di interpolazione.

Esercizio 12

Considera i dati del foglio Excel. Utilizzando la regressione multilineare con variabile dipendente z e variabili indipendenti x , w , y , scrivi l'equazione della funzione interpolante con gli opportuni coefficienti numerici, il valore di R quadro, e la deviazione standard degli errori di interpolazione (residui). Disegna su Excel lo scatterplot dei dati z , x , w , y .

Esercizio 13

Considera una distribuzione normale con $\mu=3$ e $\sigma^2=2$ e scrivi la probabilità di ottenere un valore:

- | | |
|-----------------------|------------------------|
| a) minore di 3 | d) compreso tra 3 e 1 |
| b) maggiore di 2 | e) compreso tra 5 e 8. |
| c) compreso tra 1 e 3 | |

Esercizio 14

Considera una distribuzione normale con $\mu=-2$ e $\sigma^2=1$ e scrivi il valore che ha:

- alla propria destra una probabilità del 45%
- alla propria sinistra una probabilità del 50%
- alla propria sinistra una probabilità del 95%
- alla propria sinistra una probabilità del 100%.

Esercizio 16

Considera la seguente variabile casuale discreta $x = \begin{cases} 1 & 20\% \\ 2 & 20\% \\ 3 & 40\% \\ 6 & 20\% \end{cases}$. Scrivi media, mediana, varianza. Scrivi la probabilità che la media di 81 valori di x sia minore di 2,5.

Esercizio 17

Considera una distribuzione esponenziale con $\mu=3$. Scrivi la probabilità che la media di 144 valori estratti dalla distribuzione sia maggiore di 4.

Esercizio 18

Considera i dati 3, 5, 6, 7. Prendendo come soglia $\alpha=5\%$, effettua un test statistico per verificare se la media è 4,5. Scrivi H_0 e H_1 , i prerequisiti del test, il valore della statistica, della significatività e scrivi se H_0 viene accettata o rifiutata. Spiega il significato della decisione.

Esercizio 18

Considera i dati del foglio Excel indicati con x . Prendendo come soglia $\alpha=5\%$, effettua un test statistico per verificare se la media è 6. Scrivi H_0 e H_1 , i prerequisiti del test, il valore della statistica, della significatività e scrivi se H_0 viene accettata o rifiutata. Spiega il significato della decisione.

Esercizio 19

Considera i dati del foglio Excel indicati con x e w . Prendendo come soglia $\alpha=5\%$ e supponendo le loro varianze uguali, effettua un test statistico per verificare se la media di x è uguale alla media di w . Scrivi H_0 e H_1 , i prerequisiti del test, il valore della statistica, della significatività e scrivi se H_0 viene accettata o rifiutata. Spiega il significato della decisione.

Esercizio 20

Considera i dati del foglio Excel indicati con x e w e i dati 5, 5, 5, 6, 6, 6, 7, 7, 7. Prendendo come soglia $\alpha=5\%$, effettua un test statistico per verificare se la media delle tre variabili sono uguali. Scrivi H_0 e H_1 , i prerequisiti del test, il valore della statistica, della significatività e scrivi se H_0 viene accettata o rifiutata. Spiega il significato della decisione.

Esercizio 21

Considera la seguente tabella di contingenza:

Ph	Minore di 5	Tra 5 e 5,5	Maggiore di 5,5	Totale riga
Mele integre	600	800	300	1700
Mele danneggiate	50	100	50	200
Totale colonna	650	900	350	1900

Scrivi delle apposite percentuali per evidenziare il fatto che le mele danneggiate aumentino quando aumenta il Ph. Prendendo come soglia $\alpha=5\%$, effettua un test statistico per verificare se il danneggiamento della mela dipende dal Ph. Scrivi H_0 e H_1 , i prerequisiti del test, il valore della statistica, della significatività e scrivi se H_0 viene accettata o rifiutata. Spiega il significato della decisione.

Esercizi svolti sulla probabilità condizionata**Esercizio 201**

In base alla tua esperienza degli anni precedente, la probabilità che un melo del tuo campo sia colpito da ticchiolatura (malattia causata dal fungo *Venturia Inaequalis*) è dello 1%.

Viene inventato un test che:

- se un melo ha la ticchiolatura, il test risulta positivo nel 97% dei casi e negativo nell'3% dei casi
- se un melo non ha la ticchiolatura, il test risulta positivo nel 5% dei casi e negativo nel 95% dei casi.



Se sottoponi un albero a questo test e risulta positivo, quale è la probabilità che abbia la ticchiolatura? Scrivi il risultato enunciando i passaggi effettuati.

Soluzione

La risposta non è, come potrebbe sembrare, 97%! Questo perché il 97% è la probabilità che un albero con la ticchiolatura dia un risultato positivo $P(\text{test positivo} \mid \text{ticchiolatura})$, mentre la domanda chiede la probabilità che albero con un risultato positivo abbia la ticchiolatura $P(\text{ticchiolatura} \mid \text{test positivo})$.



$$P(\text{ticchiolatura} \mid \text{test positivo}) = P(\text{ticchiolatura} \cap \text{test positivo}) / P(\text{test positivo})$$

Calcolare la probabilità che un albero abbia la ticchiolatura e fornisca un test positivo non è semplice con i dati in nostro possesso in quanto, evidentemente, i due eventi non sono indipendenti e non possiamo usare la relazione $P(A \cap B) = P(A) \cdot P(B)$. Per calcolarla applichiamo quindi il teorema di Bayes:

$$P(\text{ticchiolatura} \mid \text{test positivo}) = P(\text{test positivo} \mid \text{ticchiolatura}) \cdot P(\text{ticchiolatura}) / P(\text{test positivo}) = 0,97 \cdot 0,01 / P(\text{test positivo})$$

Calcoliamo adesso $P(\text{test positivo})$:

- un melo ha la ticchiolatura nel 1% dei casi e in questa situazione il test è positivo nel 97% dei casi, quindi un melo ha la ticchiolatura e un test positivo nello 0,97% dei casi
- un melo non ha la ticchiolatura nel 99% dei casi e in questa situazione il test è positivo nel 5% dei casi, quindi un melo non ha la ticchiolatura e ha comunque un test positivo nel 4,95% dei casi

$$\text{Quindi } P(\text{test positivo}) = 0,0097 + 0,0495 = 0,0592$$

Pertanto la risposta è $0,97 \cdot 0,01 / 0,0592 \approx 0,1639 = 16,39 \%$, il che vuol dire che il nostro test non è per nulla sicuro e ci darà parecchie segnalazioni positive anche per alberi sani.

Esercizio 202

In Sardegna viene compiuto un omicidio e gli investigatori dispongono soltanto di un capello del colpevole. Viene deciso di effettuare un test del DNA a tutta la popolazione adulta composta da 1.000.000 persone. Purtroppo essendo il capello danneggiato, il test non è estremamente affidabile. Se effettuato sul colpevole risulta sempre positivo, mentre se effettuato su un innocente risulta purtroppo comunque positivo in un caso su 100.000. Dopo un po' di giorni una persona risulta finalmente positiva. Qual è la probabilità che sia il colpevole? Scrivi il risultato enunciando i passaggi effettuati utilizzando la probabilità condizionata. C'è un ragionamento intuitivo che permette di arrivare approssimativamente allo stesso risultato?

Soluzione

Anche qui la risposta non è, come potrebbe sembrare $99.999/100.000=99,999\%$! La domanda chiede la probabilità che la persona sia colpevole sapendo che il test è positivo $P(\text{colpevole} \mid \text{test positivo})$.

$$P(\text{colpevole} \mid \text{test positivo}) = P(\text{colpevole} \cap \text{test positivo}) / P(\text{test positivo})$$

In questo caso, il numeratore è molto semplice da calcolare. Difatti, dato che un colpevole risulta sempre positivo, $P(\text{colpevole} \cap \text{test positivo}) = P(\text{colpevole}) = 1/1.000.000$. La probabilità che il test sia positivo si calcola invece nel solito modo:

- una persona è colpevole nel $1/1.000.000$ dei casi e in questa situazione il test è sempre positivo, quindi una persona è colpevole e un test positivo nel $1/1.000.000$ dei casi
- una persona è innocente nel $999.999/1.000.000$ dei casi e in questa situazione il test è positivo nel $1/100.000$ dei casi, quindi una persona è innocente e allo stesso tempo ha comunque un test positivo nel $999.999/100.000.000.000$ dei casi.

Quindi $P(\text{test positivo}) = (100.000+999.999)/100.000.000.000 = 1.099.999/100.000.000.000$

Pertanto la risposta è $\frac{1/1.000.000}{1.099.000/100.000.000.000} = \frac{1}{1.000.000} \cdot \frac{100.000.000.000}{1.099.000} = \frac{100}{1.099} \approx 9,099\%$.

Ad un risultato simile si poteva arrivare con il ragionamento intuitivo che il test sbaglia, trovando un innocente positivo, ogni 100.000 persone controllate. Essendo gli abitanti 1.000.000 si trovano comunque 10 persone positive che sono innocenti, più eventualmente il vero colpevole. Quindi per ognuno degli 11 positivi, la probabilità che sia il colpevole è $1/11$ che è circa 9,09%.