

## Metodi Statistici per la Bioingegneria 24/25

### Altri esempi di domande d'esame

#### Quiz

1. Dato il seguente insieme di dati:

[1 4 4 4 5 7 7 8 8 10]

La moda risulta:

- A. 4
  - B. 10
  - C. 6
  - D. 7
2. Dato il seguente insieme di dati:
- [1 4 4 4 5 7 7 8 8 10]
- Il 30° percentile è:
- A. 4.5
  - B. 4
  - C. 5
  - D. Nessuna delle precedenti
3. Sia  $X$  una variabile aleatoria avente varianza pari a  $10 \text{ m}^2$ . Allora la variabile aleatoria  $Y = 3X + 3$  ha varianza pari a:
- A.  $13 \text{ m}^2$
  - B.  $30 \text{ m}^2$
  - C.  $90 \text{ m}^2$
  - D.  $93 \text{ m}^2$
4. Sia data una variabile aleatoria  $X$  avente valore atteso  $E[X]$  e varianza  $\text{var}(X)$ . Per standardizzare la variabile aleatoria  $X$  dobbiamo calcolare:
- A.  $\frac{X - \text{var}(X)}{E[X]}$
  - B.  $\frac{X - E[X]}{\text{var}(X)}$
  - C.  $\frac{X - E[X]}{\sqrt{\text{var}(X)}}$
  - D.  $\frac{X - \text{var}(X)}{\sqrt{E[X]}}$
5. Siano  $X$  e  $Y$  due variabili aleatorie aventi valore atteso, rispettivamente,  $E[X]$  e  $E[Y]$  diverso da 0. Se  $X$  e  $Y$  sono indipendenti, allora il valore atteso del loro prodotto  $E[XY]$  è pari a:
- A.  $E[X] \cdot E[Y]$
  - B.  $E[X] + E[Y]$
  - C. 0
  - D. Nessuna delle precedenti
6. Abbiamo a disposizione i dati di due campioni estratti da due popolazioni normali aventi medie incognite e varianze incognite ma uguali. Applichiamo un t-test a due campioni bilaterale per testare l'ipotesi  $H_0$ : le due medie sono uguali. Se il test non rifiuta  $H_0$  cosa possiamo concludere?
- A. Le due medie sono uguali

- B. Le due medie sono significativamente diverse  
 C. Le due medie sono diverse ma la differenza non è statisticamente significativa  
 D. Non possiamo concludere nulla sulle due medie
7. Dato un campione estratto da una popolazione normale avente media nulla e varianza incognita, applichiamo il t-test a singolo campione per verificare il seguente sistema di ipotesi:  
 H0: la media è 0  
 H1: la media è diversa da 0  
 Utilizzando un livello di significatività al 5%, rifiutiamo H0 quando il p-value risulta:  
 A.  $> 0.05$   
 B.  $< 0.05$   
 C.  $> 0.025$   
 D.  $< 0.025$
8. Dato un campione estratto da una popolazione avente distribuzione continua incognita, che test statistico possiamo utilizzare per verificare il seguente sistema di ipotesi:  
 H0: la mediana della popolazione è pari a 10  
 H1: la mediana della popolazione è diversa da 10  
 A. Il t test a singolo campione  
 B. Il test dei segni  
 C. Il test di Wilcoxon Mann-Whitney  
 D. Il test dei ranghi con segno
9. Sia M1 un modello di regressione lineare contenente 10 variabili indipendenti. Sia M2 modello ridotto, ottenuto scartando dal modello M1 due variabili indipendenti. Che metrica possiamo utilizzare per confrontare M1 e M2 e scegliere il modello più adatto a descrivere i dati in maniera parsimoniosa?  
 A. L'indice di condizionamento  
 B. L'indice Variance Inflation Factor (VIF)  
 C. L'indice Akaike Information Criterion (AIC)  
 D. Il coefficiente di determinazione  $R^2$
10. Consideriamo un modello di regressione logistica che realizza una classificazione a due classi. Siano TP, TN, FP e FN rispettivamente il numero di veri positivi, veri negativi, falsi positivi e falsi negativi. La precisione del classificatore è pari a:  
 A.  $\frac{TP}{TP+FP}$   
 B.  $\frac{TP}{TP+FN}$   
 C.  $\frac{TN}{TN+FN}$   
 D.  $\frac{TP+TN}{TP+TN+FP+FN}$
11. Se in un'analisi di sopravvivenza un soggetto è censurato al tempo  $t_0$ , allora il suo tempo di sopravvivenza è:  
 A. Pari a  $t_0$   
 B. Almeno pari a  $t_0$   
 C. Minore di  $t_0$   
 D. Minore o uguale a  $t_0$

12. Il concordance index di un modello di Cox rappresenta:
- A. La probabilità che il modello assegni risk score maggiori agli individui aventi tempi di sopravvivenza minori
  - B. La probabilità che il modello assegni risk score maggiori agli individui aventi tempi di sopravvivenza maggiori
  - C. La probabilità che il modello assegni risk score minori agli individui aventi tempi di sopravvivenza minori
  - D. La probabilità che il modello assegni risk score pari a 0 agli individui censurati
13. L'algoritmo di clustering k-means cerca la partizione dei dati che:
- A. Minimizza la somma delle distanze tra i cluster
  - B. Massimizza la somma delle variabilità intra-cluster
  - C. Minimizza la somma delle variabilità intra-cluster
  - D. Minimizza la somma delle variabilità tra i cluster
14. Nel clustering gerarchico agglomerativo, il criterio di linkage è il criterio utilizzato per:
- A. Quantificare la distanza tra diverse osservazioni nello stesso cluster
  - B. Quantificare la distanza tra due cluster distinti
  - C. Decidere il numero di cluster
  - D. Quantificare quanto fedelmente il dendrogramma rappresenta le distanze tra coppie di osservazioni nel dataset
15. Nella Principal Component Analysis (PCA), quale delle seguenti proprietà devono avere i vettori dei loadings relativi alle componenti principali?
- A. Norma pari a 1
  - B. Norma pari a 0
  - C. Media pari a 0
  - D. Media 0 e deviazione standard pari a 1

### **Domande aperte**

1. Descrivere i principali indici di dispersione utilizzati nella statistica descrittiva.
2. Descrivere i principali step per l'applicazione di un test statistico parametrico per la verifica di un'ipotesi statistica  $H_0$ .
3. Descrivere i principali strumenti che possono essere impiegati per verificare se un campione proviene da una distribuzione normale.