

# ESEMPIO DOMANDE CHIUSE

**10 domande:**

risposta giusta = 1

risposta sbagliata =  $-1/3$

risposta non data = 0

**Tempo previsto:** 20 minuti

**Range voto:** 0 -10

**Sbarramento:** passa con voto  $> 4.5$

1. Se  $X$ ,  $Y$  e  $Z$  sono variabili aleatorie indipendenti, e  $a$  e  $b$  sono delle quantità deterministiche diverse da zero, allora la covarianza  $Cov(X, aY + bZ)$  è:
- A. un valore diverso da 0 pari a:  $a \cdot Cov(X, Y) + b \cdot Cov(X, Z)$
  - B. uguale a 0
  - C. un valore diverso da 0 pari a:  $a \cdot b \cdot Cov(X, Y + Z)$
  - D. un valore diverso da 0 pari a:  $a \cdot b \cdot Cov(X \cdot Y, X \cdot Z)$

2. Dato il seguente insieme di valori osservati per la variabile X  
{71 76 80 80 90 100 110 125 130 135}

la mediana risulta:

- A. Mediana = 99.7
- B. Mediana = 95
- C. Mediana = 90
- D. Mediana = 100

3. Il campo di variazione o range di un insieme di dati:
- A. è un valore sempre negativo
  - B. è un indice di posizione
  - C. è pari alla differenza tra il valore massimo e il valore minimo
  - D. è un indice insensibile alla presenza di outlier.

4. Sia  $X$  una variabile aleatoria discreta che assume  $n$  valori distinti  $x_1, x_2, \dots, x_n$ . Siano  $f(x)$  la densità discreta di  $X$  e  $F(x)$  la funzione di ripartizione di  $X$ . Quale delle seguenti condizioni deve essere soddisfatta?
- A.  $f(x_i) > 1 \forall i = 1, \dots, n$
  - B.  $F(x_n) = +\infty$
  - C.  $\sum_{i=1}^n F(x_i) = 1$
  - D.  $\sum_{i=1}^n f(x_i) = 1$

5. L'indice di asimmetria o skewness:
- A. È definito come rapporto tra il momento centrale di ordine 3 e il cubo della deviazione standard.
  - B. È definito come rapporto tra il momento centrale di ordine 4 e il quadrato della varianza.
  - C. È pari a 3 per una variabile aleatoria normale.
  - D. È positivo se la distribuzione è simmetrica.

6. Consideriamo lo stimatore ai minimi quadrati lineari  $\hat{\beta} = (X^T X)^{-1} X^T Y$ , dove come di consueto  $Y$  è il vettore dei valori dell'outcome da predire e  $X$  è la matrice delle variabili indipendenti. Sia  $\hat{\sigma}^2$  il valore della varianza dell'errore stimato a posteriori. La matrice di covarianza di  $\hat{\beta}$  può essere calcolata come:
- A.  $Cov(\hat{\beta}) = \hat{\sigma}^2 X^T X$
  - B.  $Cov(\hat{\beta}) = (X^T X)^{-1}$
  - C.  $Cov(\hat{\beta}) = \hat{\sigma}^2 (X^T X)^{-1}$
  - D. Nessuna delle precedenti

7. Dati due campioni estratti da popolazioni aventi distribuzione normale a media e varianza incognite e potenzialmente diverse, vogliamo testare l'ipotesi nulla  $H_0$ : le medie delle due popolazioni sono uguali. Che test statistico è più opportuno usare?
- A. Il t-test di Welch
  - B. Il test di Wilcoxon Mann-Whitney
  - C. Il test dei segni
  - D. Il log-rank test

8. Data una matrice di dati  $X$  in cui ogni colonna rappresenta una variabile, quali delle seguenti operazioni è necessario fare prima di effettuare l'analisi delle componenti principali (PCA)?
- A. Standardizzare le colonne di  $X$
  - B. Sottrarre la media a ciascuna colonna di  $X$
  - C. Scegliere il numero di componenti principali
  - D. Tutte le precedenti

9. Si consideri l'indice Akaike Information Criterion (AIC) dato dall'equazione:

$$AIC = n \cdot \log \left( \frac{SSE}{n} \right) + 2 \cdot p$$

dove  $n$  è il numero di osservazioni,  $p$  il numero di parametri e  $SSE$  la somma dei quadrati dei residui. All'aumentare della complessità del modello, tipicamente:

- A. Il primo addendo aumenta, il secondo diminuisce
- B. Il primo addendo diminuisce, il secondo aumenta
- C. Entrambi gli addendi aumentano
- D. Entrambi gli addendi diminuiscono

10. La regolarizzazione elastic net:

- A. Azzerà sempre almeno uno dei coefficienti del modello di regressione.
- B. Non azzerà mai nessuno dei coefficienti del modello di regressione.
- C. Combina le proprietà delle regolarizzazioni Ridge e LASSO.
- D. Non ha nessuna delle precedenti proprietà.