

Revisão 1a - 2025.1 - CD

Nome e Sobrenome: _____

Disciplina: Ciência dos Dados Semestre: 2025.1

Professor: Dr. Diego Pinheiro

Observações:

- **NÃO ESQUEÇA** de apresentar todos os cálculos e/ou justificativas necessários para se chegar às respostas (NÃO serão aceitas questões apenas com respostas);
- As respostas deverão ser escritas à **CANETA** Os cálculos podem ser feitos à lápis.

Questão 1

(Pontuação máxima: 1 pontos)

Considere dois vetores, \mathbf{x} e \mathbf{y} , com tamanho n cada um. Como representar vetorialmente a operação $f(\mathbf{x}, \mathbf{y})$ a seguir.

$$f(\mathbf{x}, \mathbf{y}) = (x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2$$

$f(\mathbf{x}, \mathbf{y}) =$

Questão 2

(Pontuação máxima: 2 pontos)

Muitos conceitos de Álgebra Linear são fundamentais para Ciência dos Dados. Considere uma matrix X de dados, com n linhas e m colunas, e um vetor de pesos

\mathbf{w} com m linhas e 1 coluna. Como representar, por exemplo, o vetor $\mathbf{y} = X\mathbf{w}$ como uma combinação linear do dados.

$\mathbf{y} =$

Questão 3

(Pontuação máxima: 3 pontos)

Sobre a construção de um intervalo de confiança assinala (V) verdadeiro ou (F) falso nas afirmações a seguir.

- () Quanto maior um intervalo de confiança, maior a precisão que ele fornece
- () O intervalo de confiança é uma probabilidade observada nos dados
- () O nível de significância α de um intervalo de confiança com nível de confiança igual a 99% é maior que o nível de significância de um intervalo de confiança com nível de confiança igual a 95%
- () Um intervalo de confiança construído para um estimador θ com nível de confiança igual a 99% indica que o intervalo captura o valor verdadeiro de θ com probabilidade igual à 99%
- () O método Bootstrapping realiza um número de B reamostragens com repetição e pode ajudar na construção de intervalos de confiança, quanto maior o valor de B , maior o nível de confiança
- () Seja X uma variável aleatória contínua e $F(x)$ a sua Função Distribuição Acumulada (CDF, do inglês *Cumulative Distribution Function*), então $F(x) = Pr(X > x)$
- () Seja X uma variável aleatória contínua e $f(x)$ a sua Função de Densidade de Probabilidade (PDF, do inglês *Probability Density Function*), então $f(x) = Pr(X \leq x) - Pr(X < x)$
- () Seja $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$ uma variável aleatória contínua calculada em função de n variáveis aleatórias X_i identicamente distribuídas, quanto maior o valor de n

maior o intervalo de confiança de \bar{X} construído com o mesmo nível de significância α

() Seja X uma variável aleatória contínua e $F(x)$ a sua Função Distribuição Acumulada (CDF, do inglês *Cumulative Distribution Function*), então $\int_{-\infty}^{+\infty} F(x) = 1$

Questão 4

(Pontuação máxima: 4 pontos)

Considere uma amostra de dados (\mathcal{X}) cujo tamanho é 20. Considere que foi realizado um procedimento de 30 reamostragens independentes de 20 valores da amostra com um número $B = 9,999$ repetições utilizando um estimador θ . Ao analisar a distribuição de todos os 9,999 valores, você percebeu as seguintes propriedades estatísticas.

$Pr(\theta < 6.55)$	0.001
$Pr(\theta < 7.95)$	0.005
$Pr(\theta < 7.95)$	0.01
$Pr(\theta < 7.95)$	0.025
$Pr(\theta < 8.40)$	0.05
$Pr(\theta < 10.5)$.5
$Pr(\theta < 12.6)$	0.95
$Pr(\theta < 13.05)$	0.975
$Pr(\theta < 13.75)$	0.995
$Pr(\theta < 14.35)$	0.999

Construa um intervalo de confiança (CI) com 5% de significância estatística para esse estimador θ utilizando as informações fornecidas. Apresente o limite inferior (l) e o limite superior (u).

(u,l) =