

Atividade Avaliativa Semana 9

Lista de exercícios

Número de Matrícula: 18111679

1. Faça essa questão manualmente. Você pode digitar as contas (faça de uma forma que eu entenda), não é necessário fazer a mão.

Um analista mediu o desempenho de um banco de dados com o tempo decorrido em função da complexidade de uma consulta em um banco de dados. A complexidade foi medida pelo número de palavras-chave na consulta. O número de operações de leitura de disco também foi medido, conforme mostrado na Tabela a seguir. Para esses dados, prepare o modelo de regressão para prever o tempo decorrido em função do número de palavras-chave e interpretar os resultados respondendo as questões a seguir:

- a. Determine os valores dos parâmetros do seu modelo.

$$b_1 = \frac{\sum xy - n\bar{x}\bar{y}}{\sum x^2 - n\bar{x}^2}$$

- b. Qual a porcentagem da variação é explicada pela regressão? Quais parâmetros são significativos, com uma confiança de 90%? E de 95% ?

Number of Keywords	Elapsed Time	Number of Disk Reads
1	0.75	3
2	0.70	6
4	0.80	7
8	1.28	78
16	1.60	92

2. Nessa questão você deve usar uma ferramenta de programação para fazer a regressão linear. Indique bibliotecas e funções que você usou. Anexe o link do repositório do seu código.

O tempo para criptografar um registro de k-bytes usando uma técnica de criptografia é mostrado na Tabela a seguir. Ajuste um modelo de regressão linear a esses dados respondendo às questões a seguir.

Record Size	Observations		
	1	2	3
128	386	375	393
256	850	805	824
384	1,544	1,644	1,553
512	3,035	3,123	3,235
640	6,650	6,839	6,768
768	13,887	14,567	13,456
896	28,059	27,439	27,659
1,024	50,916	52,129	51,360

- Faça os testes visuais para verificar se a regressão é adequada a esses dados.
 - Ajuste um modelo de regressão linear a esses dados.
 - Qual a porcentagem da variação é explicada pela regressão? Você está satisfeito com seu modelo? Se não, qual seria o seu próximo passo
 - Quais parâmetros são significativos, com uma confiança de 90%?
 - Qual o tempo esperado para criptografar um registro de 2^{20} kbits? Quais limites você colocaria para esta estimativa se você aceita um erro máximo de 10% para uma única medida futura?
3. Explique e exemplifique os seguintes conceitos encontrados na regressão linear múltipla: Analysis of Variance (ANOVA), F-Test e multicolinearidade. O seu exemplo deve ser diferente do livro. Você não precisa fazer os cálculos, apenas mostrar com o exemplo como esses conceitos são aplicados na regressão linear múltipla.
4. Os resultados de uma regressão múltipla baseada em nove observações são mostrados na Tabela a seguir. Com base nesses resultados, responda às seguintes perguntas. Justifique todas as respostas.
- Qual porcentagem de variância é explicada pela regressão?
 - A regressão é significativa no nível de confiança de 90%?
 - Qual variável tem o coeficiente mais alto?
 - Qual variável é mais significativa?
 - Quais parâmetros não são significativos em 90%?
 - Qual é o problema com essa regressão?
 - O que você tentaria a seguir para resolver o problema?

j	b_j	s_{bj}
1	1.3	3.6
2	2.7	1.8
3	0.5	0.6
4	5.0	8.3

Intercept = 75.3
 Coefficient of multiple correlation = 0.95
 Standard deviation of errors = 12.0
 F -value = 14.1