

Primeira Lista de ME613 - Primeiro Semestre de 2004

Obs: Nos exercícios 3 e 5, exiba os cálculos na forma de uma tabela, ou seja, não utilize um aplicativo computacional para fazê-los. Nos outros exercícios, não coloque as saídas dos aplicativos diretamente na solução: traduza os termos para português e coloque somente as informações que considerar **relevantes** e que sejam analisadas ou pedidas.

1 Sabemos que $\sum_{i=1}^n e_i = 0$. O que se pode dizer sobre $\sum_{i=1}^n \epsilon_i = 0$? Justifique.

2 (*) Suponha o seguinte modelo de regressão:

$$Y_i = \beta x_i + \epsilon_i,$$

onde $\beta \in \mathbb{R}$, $x' = (x_1, x_2, \dots, x_n)$ é um vetor de constantes e $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ são variáveis aleatórias independentes com esperança 0 e variância σ^2 .

- Ache o estimador de mínimos quadrados de β .
- Ache os resíduos para uma amostra (x_i, y_i) , $i = 1, 2, \dots, n$.
- Compare as propriedades da reta assim estimada com as propriedades 1-6 das páginas 47-49.
- Interprete os resultados.

3 Vinte alunos de uma universidade são aleatoriamente selecionados e recebem um teste no início de seu primeiro semestre letivo. Esse teste tenta prever o desempenho do(a) aluno(a) durante seu primeiro ano na universidade, expresso por seu coeficiente de rendimento (CR). Chamam-se respectivamente de X o resultado do teste e Y o CR ao final do primeiro ano, com os dados expostos na tabela abaixo:

i	1	2	3	4	5	6	7	8	9	10
X_i	5,5	4,8	4,7	3,9	4,5	6,2	6,0	5,2	4,7	4,3
Y_i	3,1	2,3	3,0	1,9	2,5	3,7	3,4	2,6	2,8	1,6
i	11	12	13	14	15	16	17	18	19	20
X_i	4,9	5,4	5,0	6,3	4,6	4,3	5,0	5,9	4,1	4,7
Y_i	2,0	2,9	2,3	3,2	1,8	1,4	2,0	3,8	2,2	1,5

- Obtenha os estimadores por mínimos quadrados de β_0 e β_1 .
- Faça o gráfico de dispersão dos dados e da reta de regressão estimada. Quão bom lhe parece ser o ajuste?
- Obtenha a estimativa pontual para o CR de um aluno cuja nota no teste seja 5,0.
- Qual a estimativa de mudança no CR quando houver uma melhora de 1,0 ponto na nota do teste?

4 Utilize o conjunto de dados SMSA no apêndice B.2. Espera-se que o número de médicos ativos (Y) seja relacionado à população total, dimensão territorial e renda pessoal total. Assuma um modelo de regressão de primeira ordem em cada uma das três variáveis.

- Ajuste o respectivo modelo de regressão linear simples para cada uma das três variáveis.
- Faça os três gráficos de dispersão e o da regressão estimada. Como você classificaria a relação linear para cada um dos três casos?

- (c) Calcule o EQM para cada uma das três regressões. Qual variável regressora apresenta menor variabilidade em torno da reta estimada? Compare suas conclusões nos itens (b) e (c).
- (d) Para cada região geográfica, estime a regressão linear simples do número total de crimes violentos na população total. Qual a interpretação dos interceptos estimados? E das estimativas dos coeficientes angulares?
- (e) As retas estimadas são similares para as quatro regiões? Discuta.
- (f) Calcule o EQM para cada uma das quatro regiões. Como se comparam as respectivas variabilidades em torno das retas estimadas? Compare as conclusões em (e) e (f).

5 Utilize o conjunto de dados sobre massa muscular (Problema 2.25).

- (a) Construa e realize um teste para decidir sobre uma possível associação linear negativa entre quantidade de massa muscular e idade. Utilize o erro do tipo I de 0,05. Seja preciso na definição das hipóteses envolvidas, regra de decisão e conclusões. Qual o p-valor do teste?
- (b) Calcule o p-valor do teste bicaudal para $H_0 : \beta_0 = 0$ e mostre que H_0 deve ser rejeitada num teste razoável. Podemos então concluir que β_0 traz informação relevante para a massa muscular de uma recém-nascida? Desenvolva.
- (c) Estime, com 95% de confiança, a diferença na massa muscular esperada para mulheres cujas idades difiram por um ano. Por que não é necessário saber exatamente a quais idades referimo-nos? Como podemos interpretar o resultado quanto à diferença de massa muscular?
- (d) Obtenha um intervalo de 95% de confiança para a massa muscular média de mulheres de 60 anos. Interprete-o.
- (e) Obtenha um intervalo de 95% de confiança para a massa muscular de Dona Maria, que tem 60 anos. Interprete-o. Compare (d) e (e).

6 Um estagiário numa firma de estudos econômicos recebe como tarefa estudar a relação entre as vendas de um determinado produto (Y , em milhões de dólares americanos) e a população (x , em milhões de habitantes) em cinquenta distritos de venda da firma. Utilizando-se do modelo de regressão linear simples normal, o estudante deseja testar a associação linear entre x e Y . Para tanto, ele recorre a um aplicativo computacional cujos resultados são resumidos na tabela seguinte:

Parâmetro	Estimativa	I.C. (95%)	
Intercepto	7,43119	-1,18518	16,0476
Inclinação	0,755048	0,452886	1,05721

- (a) O estudante conclui que há uma associação linear entre x e Y . Você concorda? Justifique. Qual o nível de significância do teste?
- (b) Alguém questiona o limite inferior para o intercepto, sob o ponto de vista de que vendas negativas não aconteceriam mesmo que a população no distrito fosse nula. Discuta.