

2ª Lista de Regressão
Primeiro semestre de 2004

1. Mostre que o estimador por mínimos quadrados de β_0 é o estimador linear não-viciado de mínima variância.
2. Suponha que o modelo usual de regressão seja válido exceto pela suposição de que as variâncias dos erros sejam iguais. A situação real é que, quanto maior for X , maior será a variância do erro. O fato de $\beta_1 = 0$ implica, como no modelo usual, que não exista relação (linear ou de outro tipo) entre X e Y ? Justifique.
3. Suponha que cinco observações de Y sejam obtidas para $X = 1, 4, 10, 11, 14$. Assuma também que $\sigma = 0,6$, $\beta_0 = 5$ e $\beta_1 = 3$.
 - (a) Quais são os valores esperados do QME e do QMReg?
 - (b) Para determinar a existência de uma relação linear entre X e Y , qual seria sua recomendação para o desenho do experimento:
 - (I) $X = 1, 4, 10, 11, 14$ ou (II) $X = 6, 7, 8, 9, 10$. Justifique. Sua resposta mudaria se o objetivo fosse estimar a resposta média para $X = 8$? Por quê?
4. Suponha que o modelo usual de regressão seja válido.
 - (a) Caso queiramos utilizar um teste linear geral para $H_0 : \beta_1 = 5$ vs $H_1 : \beta_1 \neq 5$, especifique o modelo completo, modelo reduzido e gl_r .
 - (b) Caso queiramos utilizar um teste linear geral para $H_0 : \beta_0 = 2, \beta_1 = 5$ vs $H_1 : \beta_0 \neq 2$ ou $\beta_1 \neq 5$, especifique o modelo completo, modelo reduzido e gl_r .
 - (c) Interprete todas as hipóteses dos itens (a) e (b).
5. Utilizando-se do exercício 4 da Lista 1,
 - (a) Estabeleça pelo R^2 qual a variável resposta que reduz mais substancialmente a variabilidade do número de médicos ativos.
 - (b) Obtenha intervalos de 90% de confiança para β_1 em cada região. As retas de regressão lhe parecem ter os mesmos coeficientes angulares? Qual sua conclusão?
6. Utilize os dados do exercício 3 da Lista 1. Calcule os valores estimados e os resíduos.
 - (a) Faça um gráfico de caixa para a variável regressora. Analise-o.
 - (b) Faça um ramo-e-folhas dos resíduos. Analise-o.
 - (c) Faça um gráfico de resíduos vs. valores estimados. Quais os problemas do modelo que podem ser estudados? Quais suas conclusões?
 - (d) Faça um gráfico Q-Q para a normalidade dos resíduos. Estime a correlação entre os resíduos ordenados e os seus valores esperados sob normalidade. Teste a hipótese de normalidade com $\alpha = 0,05$. Quais são suas conclusões?
 - (e) Novas informações exclusas do modelo são apresentadas a seguir: teste de inteligência (X_2) e CR do nível médio (X_3). Utilize-se de gráficos para verificar a importância da inclusão de X_2 e X_3 no modelo. Quais suas conclusões?

i	1	2	3	4	5	6	7	8	9	10
X_2	105	113	118	107	110	125	115	121	117	111
X_3	2,9	2,8	3,1	2,4	3,0	2,4	3,5	3,1	3,1	2,9
i	11	12	13	14	15	16	17	18	19	20
X_2	123	114	120	132	122	110	119	109	116	108
X_3	3,2	3,3	3,4	2,6	3,0	2,8	3,3	3,4	2,6	2,7

7. Um estudante recebeu um conjunto de dados para o qual estimou uma regressão linear e obteve as informações a seguir:

i	1	2	3	4	5
Y_i	35	17	42	28	53
\hat{Y}_i	42	29	32	32	40
e_i	-7	-12	10	-4	13

O estudante fez o gráfico dos resíduos vs. Y_i e encontrou uma relação positiva mas, ao fazer o gráfico dos resíduos vs. valores estimados nenhuma relação foi encontrada. Qual sua interpretação para esse fato? Qual seria um gráfico tão ou mais relevante para a mesma pergunta? Justifique.

8. Mostre que, sob o modelo de regressão normal, o erro médio quadrático puro (EMQP) tem valor esperado σ^2 .

9. Um modelo de regressão com β_0 está sendo estudado. Dados com replicações foram obtidos. Apresente os modelos completo e reduzido para testar a validade da função de regressão considerada. Quais são os respectivos graus de liberdade se $n = 20$ e $c = 10$?

10. Utilize os dados do exercício 5.

(a) Para cada uma das três regressoras, obtenha os respectivos resíduos e analise-os em detalhes. Quais suas conclusões? Em particular, alguma regressora parece mais apta para a regressão linear do que as outras? Justifique.

(b) Para cada região, obtenha os respectivos resíduos e analise-os em detalhes. Quais suas conclusões? Em particular, como você observa a variabilidade de cada uma das regiões? Justifique.

11. Um certo produto, há dez anos no mercado, tem seu comportamento de vendas segundo a tabela abaixo, onde x representa seu tempo no mercado e Y as vendas anuais no seu $(x + 1)$ -ésimo ano no mercado.

i	1	2	3	4	5	6	7	8	9	10
x_i	0	1	2	3	4	5	6	7	8	9
Y_i	98	135	162	178	221	232	283	300	374	395

(a) Pelo gráfico de dispersão de Y vs x , a regressão linear simples lhe parece adequada? Justifique.

(b) Use o procedimento de Box-Cox, com padronização, para encontrar uma transformação de potência adequada de Y . Calcule a SQE para $\lambda = 0, 3; 0, 4; 0, 5; 0, 6$ e $0, 7$. Qual transformação de Y esses números sugerem? Justifique.

(c) Use a transformação $Y' = \sqrt{Y}$ e obtenha as estimativas de regressão linear simples para os dados transformados.

(d) Faça o gráfico da reta estimada e dos dados transformados. Parece-lhe ser a regressão linear simples um bom modelo? Justifique.

(e) Obtenha os resíduos e faça o seu gráfico vs. os valores estimados e o gráfico Q-Q para a normalidade. Qual sua análise?

(f) Expresse a função de regressão estimada nas unidades originais.