

4ª Lista de ME613 - Primeiro semestre de 2004

1. Dê os graus de liberdade associados a cada uma das soma de quadrados extras a seguir: (1) $SQR(x_1|x_2)$; (2) $SQR(x_2|x_1, x_3)$; (3) $SQR(x_1, x_2|x_3, x_4)$; (4) $SQR(x_1, x_2, x_3|x_4, x_5)$; e (5) $SQR(x_1, x_2, x_3|x_4, x_5, x_6, x_7, x_8, x_9, x_{10})$.

2. Utilize o conjunto de dados do problema 7.8 do livro, *Brand preference*.

(a) Obtenha a tabela ANOVA com a decomposição da SQR nas somas de quadrados extras associadas a x_1 e a x_2 dada x_1 .

(b) Teste se x_2 pode ser retirada do modelo quando x_1 não o for. Use a estatística de teste F^* e nível de significância 0,001 (não se esqueça de reportar o p-valor e interpretá-lo. Exiba as hipóteses, interprete-as, assim como a regra de decisão e suas conclusões.

(c) Ajuste um modelo de primeira ordem para Y no conteúdo líquido (x_1). Qual a função de regressão estimada?

(d) Faça a regressão em x_1 e x_2 conjuntamente. Compare as ANOVAS obtidas. A $SQR(x_1)$ da RLS é igual à $SQR(x_1|x_2)$ da regressão em x_1 e x_2 ? Interprete o resultado.

(e) Obtenha a matriz de correlações das regressoras. Reinterprete os resultados obtidos nos itens anteriores em razão dessa matriz.

3. Utilize o conjunto de dados S₁SA. Para a predição do número de médicos ativos em uma área, decidiu-se pela utilização das regressoras: população total (x_1); e renda pessoal total (x_2). A questão que se coloca agora é sobre a eventual inclusão de outras regressoras e, se tal for a ação, qual seria a variável cuja contribuição seria maior. Assume-se um modelo múltiplo de primeira ordem.

(a) Para cada uma das seguintes candidatas, calcule o coeficiente de determinação parcial, dado que x_1 e x_2 estão no modelo: área total (x_3); percentual populacional com 65 anos ou mais (x_4); número de leitos hospitalares (x_5); e total de crimes violentos (x_6).

(b) Com base nos resultados em (a), qual das quatro candidatas é a melhor? A SQ extra dessa variável maior do que a das outras três? Interprete o resultado.

(c) Usando F^* , teste se a variável selecionada em (b) útil quando x_1 e x_2 já estão no modelo. Seja cuidadoso(a) na especificação das hipóteses, nos modelos que elas representam, na regra de decisão e na conclusão. O que poder-se-ia concluir sobre os valores de F^* (sem quaisquer cálculos) das outras três candidatas? Justifique.

4. Utilizando a situação do exercício 7.29 do livro-texto, encontre $SQR(R) - SQR(F)$ (em termos matriciais) para testar se $\beta_1 = \beta_2$.

5. Suponha o teste linear geral, em que H_0 afirma $\mathbf{C}\beta = \mathbf{h}$. Ache os estimadores de mínimos quadrados no modelo reduzido. Dica: Monte a função que você deseja minimizar e utilize o método lagrangiano com a restrição dada pela hipótese nula.

6. Uma nova tecnologia de injeção para redução de consumo de combustível foi testada em doze caminhonetes. Nos dados a seguir, x_i indica a veloci-

dade constante (milhas/hora) na i -ésima observação e Y_i o desempenho (milhas/galão).

i	1	2	3	4	5	6	7	8	9	10	11	12
x_i	35	35	40	40	45	45	50	50	55	55	60	60
Y_i	22	20	28	31	37	38	41	39	34	37	27	30

Assume-se que o modelo de segunda ordem com erros independentes normais seja adequado.

- Ajuste o modelo de segunda ordem. Grafie a reta de regressão estimada e os dados. Parece a função de regressão quadrática ter um bom ajuste? Qual o coeficiente de determinação?
- Teste se há relação de regressão. Use $\alpha = 0,05$ e explicita as hipótese, regra de decisão envolvidas e sua conclusão.
- Estime o desempenho esperado em milhas/galão num próximo teste, feito a 48 milhas/hora e construa um I.P. de 95% para sua medida. Interprete.
- Preveja o desempenho em milhas/galão num próximo teste, feito a 48 milhas/hora e construa um I.P. de 95% para sua medida. Interprete.
- Teste se o termo quadrático pode ser retirado do modelo. Explicita as hipóteses, regra de decisão, conclusão e p-valor. Interprete.
- Escreva a função de regressão estimada em termos da regressora original x .
- Calcule os respectivos coeficientes de correlação entre x e x^2 e X e X^2 . Há necessidade neste caso de utilizar desvios com relação à média? Explique.
- Qual o valor da velocidade para o qual espera-se melhor desempenho?
- Esse valor acontece no suporte dos dados? Quais suas considerações sobre sua utilidade, de possíveis aplicações desse resultado?

7. Utilize o conjunto de dados *Calculator Maintenance*, do problema 2.18. Os usuários de calculadoras eletrônicas são divididos em instituições de terinamento, em que se utiliza um modelo de estudante, e firmas comerciais, em que se utiliza um modelo comercial. um analista gostaria de relacionar o tipo de modelo da calculadora ao tempo de serviço despendido no caso. A seguir, apresentam-se os dezoito tipos de atendimento:

i	1	2	3	4	5	6	7	8	9
Modelo	C	E	C	C	C	E	E	C	C
i	10	11	12	13	14	15	16	17	18
Modelo	C	C	E	E	C	E	C	C	C

Assuma-se que o modelo dado pela equação (10.4) seja válido, sendo $x_2 = 1$, para um modelo de estudante, e zero, caso contrário.

- (a) Interprete os coeficientes do modelo.
- (b) Ajuste-o
- (c) Estime o efeito do modelo de calculadora sobre o tempo médio de atendimento com 95% de confiança. Interprete seu resultado.
- (d) Por que o analista gostaria de incluir x_1 , o número de calculadoras, no modelo de regressão, mesmo sendo seu objetivo o efeito do tipo de calculadora?
- (e) Obtenha os resíduos do ajuste em (b) e grafе-os contra x_1x_2 . há qualquer indicação de que uma interação seria benéfica ao modelo?
- (f) Ajuste o modelo (10.6)
- (g) Teste se a interação pode ser retirada do modelo. Seja preciso(a) quanto às hipóteses, regra de decisão, p-valor e conclusão. Compare seu resultado inferencial com o intuitivo, do item (e). Caso seu resíduo não possa ser retirado do modelo, descreva a natureza da interação.

8. Utilize o conjunto de dados *SENIC*, para estudar a regressão de risco de infecção (Y) no tempo de permanência (x_1), idade (x_2), proporção de exames torácicos rotineiros (x_3) e afiliação da Faculdade de Medicina (x_4).

- (a) Ajuste um modelo de primeira ordem, sendo $x_4 = 1$ se o hospital tem afiliação a uma Faculdade de Medicina e zero, caso contrário.
- (b) Estime o efeito da afiliação sobre o risco de infecção com 98% de confiança. Interprete-o.
- (c) Há sugestão de que o efeito da afiliação pode interagir com efeitos da idade e dos testes rotineiros. introduza as interações apropriadas ao modelo. Teste se os termos de interação são úteis, especificando as hipóteses, a regra de decisão, o p-valor e a conclusão.

Para os itens (d)-(f), utilize as regressoras idade como x_1 , proporção de exames laboratoriais (x_2), número médio diário de pacientes (x_3), infraestrutura (x_4) e região (x_5 , x_6 e x_7).

- (d) Ajuste um modelo de primeira ordem, sendo $x_5 = 1$ se NE, $x_6 = 1$ se NC e $x_7 = 1$ se S.
- (e) Teste o efeito dos exames laboratoriais de rotina. Seja específico nas hipóteses, na regra de decisão, no p-valor e na conclusão.
- (f) Examine se o efeito do tempo de permanência em hospitais localizados na região oeste difere daquele em hospitais das outras três regiões, usando intervalos

de confiança apropriados para cada comparação pareada. use o procedimento de Bonferroni com 95% de confiança conjunta. Qual sua conclusão?