

MAE-5725 - MODELOS LINEARES

4ª Lista de Exercícios

Profa. Silvia N. Elia

- 1) Considere o modelo $Y_i = \beta x_i + \epsilon_i$ onde $\epsilon_i, i = 1, 2, \dots, n$ são variáveis aleatórias independentes, $\epsilon_i \sim N(0, \sigma_i^2)$, com $\sigma_i^2 = x_i^2 \sigma^2$.
- Obtenha o estimador não viciado uniformemente de variância mínima de β .
 - Determine a variância deste estimador.
 - Derive um intervalo de confiança com coeficiente $1 - \alpha$ para β .

- 2) Deseja-se comparar dois métodos de ensino A e B e com este objetivo, aplicou-se uma prova a n_1 alunos sujeitos ao método de ensino A e n_2 sujeitos ao método de ensino B. Admitindo o modelo

$$Y_i = \beta_0 + \beta_1 W_i + \epsilon_i, \epsilon_i, i = 1, 2, \dots, n \text{ variáveis aleatórias}$$

independentes $\epsilon_i \sim N(0, \sigma^2)$, com $W = \begin{cases} 0 & \text{para o método A} \\ 1 & \text{para o método B} \end{cases}$ e Y - nota da prova,

- Qual o significado prático de β_0 e β_1 ?
- Obtenha o estimador NVUVM da diferença das notas médias para alunos sujeitos aos métodos A e B, justificando a resposta.
- Construa o teste de hipóteses para a igualdade das notas médias segundo os dois métodos, contra a alternativa de que as médias são diferentes, usando a forma geral $C\beta = m$.
- Prove que o teste do item (c) é equivalente ao teste t de igualdade de médias para amostras independentes.

- 3) Considere o modelo $Y_i = \alpha + \beta x_i + \epsilon_i, \epsilon_i \sim N(0, \sigma^2 x_i^2), \epsilon_i, \epsilon_j$ v.a. independentes, $i \neq j$. Dados os valores de X e Y para uma amostra de 5 observações

X	1	2	5	5	10
Y	13	10	20	15	50

- Obtenha os estimadores NVUVM para α e β e as correspondentes estimativas.
- Teste, ao nível de significância 0,10 a hipótese $H_0: \beta = 1$ contra $H_a: \beta > 1$.

- 4) Para um grupo de n alunos sujeitos a três diferentes métodos de ensino, observou-se a nota num teste (Y), conhecendo-se o número de horas (X) que cada aluno estudou para o referido teste. Se o modelo utilizado para o problema foi

$$Y_i = \beta_0 + \beta_1 x_i + \alpha_1 z_{1i} + \alpha_2 z_{2i} + \epsilon_i, \epsilon_i \sim N(0, \sigma^2), \text{ independentes}$$

onde

$$(z_1, z_2) = \begin{cases} (1, 0) & \text{para o método de ensino A} \\ (0, 1) & \text{para o método de ensino B} \\ (0, 0) & \text{para o método de ensino C} \end{cases}$$

- a) Qual a expressão de $E(Y/x)$ para cada um dos métodos de ensino ?
 b) Qual o significado prático dos parâmetros α_1 , α_2 , e β ?

5) Seja

$$\begin{aligned} Y_1 &= \alpha_1 + \epsilon_1 \\ Y_2 &= 2\alpha_1 - \alpha_2 + \epsilon_2 \\ Y_3 &= \alpha_1 + 2\alpha_2 + \epsilon_3 \end{aligned}$$

onde $\epsilon_i \sim N(0, \sigma^2 I)$. Derive o teste F para $H_0: \alpha_1 = \alpha_2$ contra $H_a: \alpha_1 \neq \alpha_2$.

6) Obtenha os estimadores de mínimos quadrados ponderados para os parâmetros do modelo

$$Y = \beta_0 + \beta_1 X + \epsilon \text{ com base nos dados}$$

X	1	1	2	2	3	3	4	4	5	5
Y	2,5	3,5	3,5	4,5	4,5	5,5	5,0	7,0	6,0	8,0

sabendo que $\text{Var}(Y/x) = \begin{cases} \sigma^2 & \text{para } x = 1, 2, 3 \\ 2\sigma^2 & \text{para } x = 4, 5 \end{cases}$, e que ϵ_i, ϵ_j são variáveis

aleatórias independentes, $i \neq j$.

Admitindo normalidade para os ϵ_i , teste a hipótese $H_0: \beta_1 = 0$.

7) Considere o modelo de locação $Y_i = \mu + \epsilon_i$, $i = 1, 2, \dots, n$ com $\epsilon_i \sim N(0, \Sigma)$, onde

$$\Sigma = \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{bmatrix}$$

a) Mostre que $\hat{\beta} = \hat{\beta}_G$.

b) Determine a matriz F tal que $\Sigma X = XF$.

8) Os dados abaixo correspondem a uma amostra de $n = 20$ terrenos localizados em três diferentes regiões de uma cidade. sendo que X corresponde à área do terreno (em m^2) e Y ao preço.

X	200	200	250	300	300	400	400	450
Y	16	12	20	24	20	22	33	34
região	C	A	B	B	A	A	C	B

X	500	500	550	550	600	600	600	700
Y	40	41	43	42	50	50	48	55
região	C	C	B	A	C	C	B	B

X	700	800	800	800
Y	54	60	61	62
região	A	A	A	A

- Construa o modelo de regressão $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$.
- Construa o gráfico dos resíduos contra região. O que podemos concluir sobre a influência da região no preço dos terrenos?
- Admitindo um modelo de três retas paralelas para as regiões A, B e C, com o auxílio de variáveis dummy, obtenha as equações estimadas destas três retas. Qual a restrição que está sendo imposta quando admitimos o paralelismo das três retas?
- Teste a hipótese de que as retas para as regiões B e C são coincidentes.
- Teste a hipótese de que o preço independe da região.

9) Considere o modelo linear geral $\underline{Y} = \underline{X}\underline{\beta} + \underline{\epsilon}$, $\underline{\epsilon} \sim N(0, \sigma^2 \underline{I})$ e $\hat{\beta}_0$ o estimador de máxima verossimilhança de $\underline{\beta}$ sob a hipótese $C\underline{\beta} = \underline{m}$. Prove que $C\hat{\beta}_0 = \underline{m}$.

10) Obtenha o teste da razão de verossimilhança generalizada para $H_0: \underline{\beta} = \underline{b}$, $\underline{b}$ vetor de constantes especificadas, no modelo $\underline{Y} = \underline{X}\underline{\beta} + \underline{\epsilon}$, $\underline{\epsilon} \sim N(0, \sigma^2 \underline{I})$, $\underline{\beta}$ vetor de parâmetros $k + 1$ dimensional. Determine o estimador de mínimos quadrados de $\underline{\beta}$ sob H_0 .

11) Considere R^2 , o coeficiente de explicação após o ajuste de um modelo linear geral com intercepto, com k variáveis explicativas, com base em uma amostra de n observações.

- Suponha que uma nova variável explicativa foi adicionada ao modelo. Nessas condições, o valor do coeficiente de explicação aumenta? Justifique sua resposta.

Uma crítica que é feita ao coeficiente de explicação é que ele não utiliza o tamanho de amostra e o número de variáveis explicativas no modelo. Por isso, define-se o coeficiente de explicação ajustado para graus de liberdade:

$$R_a^2 = 1 - (n-1)/(n-p) (1 - R^2),$$

sendo p o número de variáveis explicativas no modelo mais 1. Nessas condições:

- Prove que $R_a^2 \leq R^2$.
- Prove que se Y é a variável resposta e $X_1, X_2, \dots, X_k$ é um conjunto de variáveis explicativas, a equação de regressão com maior R_a^2 é aquela com menor quadrado médio do resíduo.
- Prove que R_a^2 pode diminuir com o acréscimo de uma nova variável explicativa no modelo. (pontos)