Universidade Federal Fluminense (UFF)

Instituto de Matemática e Estatística (IME)

Departamento de Estatística (GET)

Disciplina: Modelos Lineares I

Professor: José Rodrigo de Moraes

6ª Lista de Exercícios – Data: 18/11/2019 (2ª feira)

Assunto: *Transformação para linearização do modelo, Regressão polinomial.*

1ª Questão: Considerando os dados sobre a produção brasileira de automóveis no período de 1957 a 1973, pede-se:

Tabela 1: Produção brasileira de automóveis.

Ano	Produção de automóveis
1957	30.542
1958	60.983
1959	96.114
1960	133.041
1961	145.584
1962	191.194
1963	174.191
1964	183.707
1965	185.187
1966	224.574
1967	225.362
1968	278.473
1969	349.519
1970	416.047
1971	516.038
1972	608.985
1973	729.135

Fonte: *Associação Nacional dos Frabricantes de Veículos Automotores (Anfavea)*

a) Ajuste o três modelos estatísticos abaixo usando o método de mínimos quadrados ordinários (MQO):

Modelo 1: $Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i$

Modelo 2: $Y_i = \beta_0 \; \beta_1^{X_{i1}} \varepsilon_i$

Modelo 3: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i1}^2 + \varepsilon_i$

OBS: Nos ajustes dos modelos denote 1957 como 1, 1958 como 2, e assim sucessivamente.

b) Ao nível de significância de 5%, avalie a significância dos parâmetros e calcule o coeficiente de determinação ajustado para cada modelo.

c) Quais dos três modelos você escolheria? Justifique a sua resposta.

2ª Questão: Considere novamente os dados coletados num hospital universitário por uma equipe multiprofissional (Tabela 2) sobre o volume expiratório forçado (VEF) e a altura (cm) de garotos na faixa etária de 10 a 14 anos. Estes

Tabela 2: Dados sobre n=12 garotos.

Garoto	Altura	VEF
1	134	1,70
2	138	1,90
3	142	2,00
4	146	2,10
5	150	2,20
6	154	2,50
7	158	2,70
8	162	3,00
9	166	3,10
10	170	3,40
11	174	3,80
12	178	3,90

Pede-se:

a) Construa o gráfico de dispersão entre altura e VEF, e analise-o.

b) Ajuste um modelo de RLS normal para explicar a variabilidade dos valores do VEF, e avalie a significância individual dos parâmetros do modelo e a hipótese de linearidade usando a análise gráfica dos resíduos estudentizados.

c) Caso exista violação da hipótese de linearidade, proponha e ajuste um novo modelo estatístico aos dados observados. Com este novo modelo, a hipótese de linearidade foi satisfeita? Justifique a sua resposta.

d) Avalie se a qualidade do ajuste melhorou empregando o novo modelo. Justifique a sua resposta.

3ª Questão: Um experimento foi realizado para verificar o efeito da temperatura e da concentração sobre a produção de certo produto químico. Os dados obtidos são apresentados na tabela 3.

a) Ajuste um modelo de regressão linear considerando as duas variáveis explicativas simultaneamente para explicar a produção do produto químico; e avalie a significância individual dos parâmetros do modelo e se as hipóteses básicas foram atendidas. Use a análise gráfica dos resíduos estudentizados.

b) Você detectou a violação de alguma hipótese? Proponha um novo modelo para resolver tal violação. Demonstre por meio da análise gráfica dos resíduos estudentizados que a violação foi de fato resolvida. As outras hipóteses do modelo (*normalidade*, *homocedasticidade* e *independência*) também foram atendidas?

Tabela 3: Dados de um experimento com certo produto químico.

Produção	Temperatura	Concentração
189	80	10
203	100	10
222	120	10
234	140	10
261	160	10
204	80	15
212	100	15
223	120	15
246	140	15
273	160	15
220	80	20
228	100	20
252	120	20
263	140	20
291	160	20
226	80	25
232	100	25
259	120	25
268	140	25
294	160	25

4ª Questão: Com o objetivo de avaliar o efeito do tempo de exposição (em minutos) de produtos alimentícios a uma temperatura de 300^{o}F sobre o número de bactérias sobreviventes, pede-se:

Tabela 4: Exposição de produtos alimentícios a uma temperatura de 300ºF

Tempo de exposição	Nº de bactérias
175	1
108	2
95	3
82	4
71	5
50	6
49	7
31	8
28	9
17	10
16	11
11	12

a) Ajuste um modelo de regressão linear para os dados observados; e avalie a significância individual dos parâmetros, a qualidade global do ajuste e se as hipóteses básicas do modelo foram atendidas. **Sugestão:** *Use a análise gráfica dos resíduos estudentizados*.

b) Utilize alguma transformação para os valores da variável resposta do modelo, e verifique se agora o modelo de regressão linear é apropriado. Avalie os mesmos aspectos pedidos na letra (a).

c) Caso o modelo ajustado na letra (b) ainda não seja apropriado, ajuste o modelo de regressão polinomial (modelo não linear) para os dados observados. Cheque as hipóteses básicas do modelo.

d) Para o modelo de regressão polinomial, avalie se existe alguma observação fortemente influente no conjunto de dados. Se excluí-la, o modelo polinomial continua sendo apropriado para os dados observados?

Respostas da 6ª Lista de Exercícios:

"Modelos Lineares I"

1ª Questão:

a)

Modelo 1: $\hat{Y}_i = -54.051{,}074 + 35.735{,}583 X_{i1}$

Modelo 2: $\hat{Y}_i = 10^{4{,}710 + 0{,}067 X_{i1}} = 51.286 \cdot (1{,}166)^{X_{i1}}$

Modelo 3: $Y_i = 109.179{,}456 - 15.810{,}900 X_{i1} + 2.863{,}693 X_{i1}^2$

b)

Modelo 1: R^2ajust=83,9%

Intercepto: t_{obs}=-1,357; p-valor = 0,195 > 0,05

Ano: t_{obs}=9,195; p-valor < 0,001 < 0,05

Modelo 2: R^2ajust=90,3%

Intercepto: t_{obs}=84,555; p-valor < 0,001 < 0,05

Ano: t_{obs}=12,263; p-valor < 0,001 < 0,05

Modelo 3: R^2ajust=94,6%

Intercepto: t_{obs}=2,924; p-valor = 0,011 < 0,05

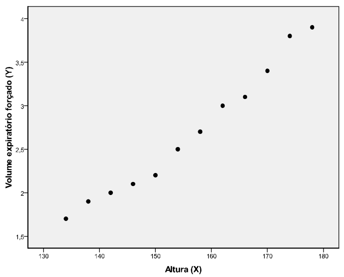
Ano: t_{obs}=-1,655; p-valor = 0,120 > 0,05

Ano^2: t_{obs}=5,553; p-valor < 0,001 < 0,05

c) *Por conta do aluno!!!* **Sugestão:** Plote num único gráfico de dispersão o ano (t) e os valores ajustados de Y obtidos para cada modelo.

2ª Questão:

a) Figura 1: Gráfico de dispersão entre altura e volume expiratório forçado de 12 garotos.



b) $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i = -5{,}313 + 0{,}051 X_i; \quad i = 1, 2, \ldots, 12$

Intercepto: t_{obs}=-13,465; p-valor < 0,001

Ano: t_{obs}=20,366; p-valor < 0,001

Linearidade não atendida → Fazer gráfico de dispersão entre X e resíduos estudentizados.

c) $Y_i = 7{,}765 - 0{,}118 X_{i1} + 0{,}001 X_{i1}^2; \quad i = 1, 2, \ldots, 12$

OBS: *Avalie também a significância dos parâmetros do novo modelo.*

A lineariadade foi satisfeita.

d) A qualidade do ajuste melhorou como o modelo de regressão polinomial.

3ª Questão:

a) $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} = 93{,}180 + 0{,}870 X_{i1} + 2{,}424 X_{i2}; \quad i = 1, 2, \ldots, 20$

Intercepto: p-valor < 0,001

Temperatura: p-valor < 0,001

Concentração: p-valor < 0,001

Gráfico de dispersão entre X_1 e os resíduos estudentizados → existe uma relação curvilínea (*violação da hipótese de linearidade*). **OBS:** Verifique também o gráfico dos valores ajustados *versus* resíduos estudentizados.

b) $Y_i = 164{,}823 - 0{,}394 X_{i1} + 2{,}424 X_{i2} + 0{,}005 X_{i1}^2; \quad i = 1, 2, \ldots, 20$

Intercepto: p-valor < 0,001

Temperatura: p-valor=0,326

Concentração: p-valor < 0,001

Temperatura^2: p-valor = 0,005

Gráfico de dispersão entre X_1 e os resíduos estudentizados → *sugere que a violação da hipótese de linearidade foi contornada*. **OBS:** Verifique também o gráfico dos valores ajustados *versus* resíduos estudentizados.

QQ Plot → normalidade aproximadamente satisfeita.

4ª Questão:

a)

$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i = 10{,}755 - 0{,}070\, X_i; \quad i = 1, 2, \ldots, 12$

Intercepto: p-valor < 0,001 < 0,05

Tempo de exposição: p-valor < 0,001 < 0,05

R^2=86,9% e R^2ajust=85,6%

Linearidade não atendida → Fazer gráfico de dispersão entre X e resíduos estudentizados.

Normalidade não atendida → *QQ Plot* dos resíduos estudentizados.

b) Contagens → usar $Y_i^* = \sqrt{Y_i}$

$\hat{Y}_i^* = \hat{\beta}_0 + \hat{\beta}_1 X_i = 3{,}398 - 0{,}016\, X_i; \quad i = 1, 2, \ldots, 12$

Intercepto: p-valor < 0,001 < 0,05

Tempo de exposição: p-valor < 0,001 < 0,05

R^2=94,4% e R^2ajust=93,9%

Linearidade não atendida → Fazer gráfico de dispersão entre X e resíduos estudentizados.

Normalidade atendida → *QQ Plot* dos resíduos estudentizados.

c) $Y_i^* = 3{,}688 - 0{,}027 X_{i1} + 0{,}000063\, X_{i1}^2 \; ; \quad i = 1, 2, \ldots, 12$

Intercepto: p-valor < 0,001

Temperatura: p-valor < 0,001

Temperatura^2: p-valor < 0,001

R^2=98,8% e R^2ajust=98,6%

Gráfico de dispersão entre X_1 e os resíduos estudentizados → *sugere que a violação da hipótese de linearidade foi contornada*. **OBS:** *Verifique também o gráfico dos valores ajustados versus resíduos estudentizados.*

QQ Plot → normalidade aproximadamente satisfeita.

d) Observação fortemente influente: obs 1

Com a exclusão da observação 1 o modelo polinomial continua apropriado ?

$Y_i^* = 3{,}608 - 0{,}022 X_{i1} + 0{,}000021\, X_{i1}^2 \; ; \quad i = 1, 2, \ldots, 11$

Intercepto: p-valor < 0,001

Temperatura: p-valor < 0,001

Temperatura^2: p-valor = 0,529