Universidade Federal Fluminense (UFF)

Instituto de Matemática e Estatística (IME)

Departamento de Estatística (GET)

Disciplina: Modelos Lineares I

Professor: José Rodrigo de Moraes

3ª Lista de Exercícios – Data: 07/10/2019

Assunto: *Inferência no modelo de RLM, Teste F parcial e análise dos resíduos*

1ª Questão: Considerando o modelo de RLM da forma $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, onde $\boldsymbol{\varepsilon} \sim N\left(0\,,\, \sigma^2 \cdot \boldsymbol{I}_n\right)$, demonstre usando o método de MQ e MV que $\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}\grave{}\boldsymbol{X}\right)^{-1}\boldsymbol{X}\grave{}\boldsymbol{Y}$.

2ª Questão: Para um conjunto de dados foi ajustado um modelo de regressão linear simples com um intercepto e uma variável explicativa. Considerando a matriz abaixo:

$$\left(\boldsymbol{X}\grave{}\boldsymbol{X}\right)^{-1} = \begin{bmatrix} 31/177 & -3/177 \\ -3/177 & 6/177 \end{bmatrix}$$

a) Determine o tamanho da amostra n.

b) Encontre o valor de $\sum_{i=1}^{n} X_i^2$.

c) Encontre a média da variável explicativa X.

d) Se $\hat{\beta} = \begin{bmatrix} 0{,}5 \\ 2{,}5 \end{bmatrix}$ determine $\boldsymbol{X}\grave{}\boldsymbol{Y}$.

3ª Questão: Mostre que $\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}\grave{}\boldsymbol{X}\right)^{-1}\boldsymbol{X}\grave{}\boldsymbol{Y}$ pode ser expresso por $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \left(\boldsymbol{X}\grave{}\boldsymbol{X}\right)^{-1}\boldsymbol{X}\grave{}\boldsymbol{\varepsilon}$, e que $\hat{\boldsymbol{\beta}}$ é um estimador não viciado para $\boldsymbol{\beta}$ e que tem variância dada por $\boldsymbol{VAR}\left(\hat{\boldsymbol{\beta}}\right) = \sigma^2\left(\boldsymbol{X}\grave{}\boldsymbol{X}\right)^{-1}$.

4ª Questão: Mostre no modelo de RLM a soma dos quadrados dos resíduos pode ser representada por meio de notação matricial do seguinte modo:

$$SQRes = SQT - SQM = \boldsymbol{Y}\grave{}\boldsymbol{Y} - \hat{\boldsymbol{\beta}}\grave{}\boldsymbol{X}\grave{}\boldsymbol{Y}$$

5ª Questão: Considere os dados sobre porcentagem de total de calorias (Y) obtidas do complexo de carboidratos para 20 homens diabéticos dependentes de insulina que foram submetidos a uma alta dieta de carboidrato por um período de seis meses. Suspeita-se que o regime está relacionado com a idade, peso e outros componentes da dieta, tais como a porcentagem de calorias como proteínas. Os dados se encontram na tabela a seguir:

Aluno	% de carboidrato (Y)	% de proteína (X_1)	Peso (X_2)	Idade (X_3)
1	33	14	100	33
2	40	15	92	47
3	37	18	135	49
4	27	12	144	35
5	30	15	140	46
6	43	15	101	52
7	34	14	95	62
8	48	17	101	23
9	30	15	98	32
10	38	14	105	42
11	50	17	108	31
12	51	19	85	61
13	30	19	130	63
14	36	20	127	40
15	41	15	109	50
16	42	16	107	64
17	46	18	117	56
18	24	13	100	61
19	35	18	118	48
20	37	14	102	28

Ajuste modelos de regressão linear aos dados observados, representando as suas equações e descrevendo seus componentes/variáveis. Faça a escolha (seleção) do modelo, mas mostre todas as etapas de teste até a escolha do

modelo final. Use o teste de comparabilidade de modelos (Teste F parcial). Avalie por método gráfico e por meio de alguma medida estatística se os valores estimados da variável resposta diferem muito dos seus respectivos valores observados. Avalie também, por método gráfico, se a hipótese de normalidade dos erros é atendida, e caso negativo explique quais as implicações da violação dessa hipótese. **OBS:** *Use o programa R.*

6ª Questão: Um estudo no condado de Alachua, Flórida, investigou o relacionamento entre certos índices de saúde mental e diversas variáveis explicativas, tais como o escore dos eventos vividos (X_1) e posição socioeconômica (X_2). O interesse principal do estudo estava focado no índice de distúrbio mental (Y) que incorporou dimensões de sintomas psiquiátricos, incluindo aspectos de ansiedade e depressão. Escores maiores altos desse índice indicavam maior distúrbio mental. Com relação às duas variáveis explicativas mencionadas, os escores dos eventos vividos é uma medida composta da severidade dos principais eventos vividos que o indivíduo experimentou nos últimos três anos. Esses eventos variavam de transtornos pessoais graves, como uma morte na família para eventos menos graves, como mudar-se de local de moradia. Assim essa medida, variou de 3 a 97 na amostra, sendo que um escore alto é indicativo de uma maior gravidade nos eventos vividos. Quanto a variável "posição sócio-econômica", é um índice composto baseado na ocupação, renda e nível educacional do indivíduo, mensurado numa escala que varia de 0 a 100, sendo que quanto maior o escore, maior o nível socioeconômico do indivíduo. Os dados do referido estudo são fornecidos na tabela a seguir:

Indiv.	Distúrbio mental (Y)	Eventos vividos (X_1)	PSE (X_2)	Indiv.	Distúrbio mental (Y)	Eventos vividos (X_1)	PSE (X_2)
1	17	46	84	**21**	27	60	70
2	19	39	97	**22**	28	97	89
3	20	27	24	**23**	28	37	50
4	20	3	85	**24**	28	30	90
5	20	10	15	**25**	28	13	56
6	21	44	55	**26**	28	40	56
7	21	37	78	**27**	29	5	40
8	22	35	91	**28**	30	59	72
9	22	78	60	**29**	30	44	53
10	23	32	74	**30**	31	35	38
11	24	33	67	**31**	31	95	29
12	24	18	39	**32**	31	63	53
13	25	81	87	**33**	31	42	7
14	26	22	95	**34**	32	38	32
15	26	50	40	**35**	33	45	55
16	26	48	52	**36**	34	70	58
17	26	45	61	**37**	34	57	16
18	27	21	45	**38**	34	40	29
19	27	55	88	**39**	41	49	3
20	27	45	56	**40**	41	89	75

a) Faça gráficos apropriados para avaliar a relação das variáveis explicativas "eventos vividos" e "pse" com a variável resposta "distúrbio mental".

b) Ajuste modelos de regressão normal aos dados observados, representando as suas equações, e descrevendo seus termos e variáveis. Faça a escolha (seleção) do modelo, mas mostre todas as etapas de teste até a escolha do modelo final. Entre os testes estatísticos utilizados, considere para a seleção do modelo final o teste de comparabilidade de modelos (Teste F parcial). Teste também a existência do efeito de X_1*X_2 (interação) no modelo.

c) Interprete as estimativas dos parâmetros do modelo selecionado, no contexto do problema, e obtenha alguma medida de qualidade do ajuste.

d) Use o modelo selecionado para estimar o índice médio de distúrbio mental para indivíduos com um escore de 45 para eventos vividos e um escore de 56 para posição socioeconômica, e construa o respectivo intervalo de confiança de 95%.

e) Use o modelo selecionado para prever o índice de distúrbio mental de um indivíduo com um escore de 45 para eventos vividos e um escore de 56 para posição socioeconômica, e construa o respectivo intervalo de predição de 95%.

7ª Questão: Considere um estudo com o objetivo de identificar as variáveis que afetam os gastos em academias de ginástica. Foram considerados três variáveis potenciais: consumo de energia (em Kilowatts), horas de mão-de-obra e número de alunos matriculados.

Academia	Gastos ($)	Consumo de energia (KW)	Horas mão de obra	Nº de alunos
1	350	6	10	100
2	400	8	14	110
3	470	12	16	110
4	550	10	26	98
5	620	15	24	112
6	380	7	12	95
7	290	6	13	75
8	490	9	21	124
9	580	11	20	126
10	610	13	24	116
11	560	12	23	99
12	420	14	12	104
13	450	11	19	108
14	510	12	19	108
15	380	9	11	89

a) Usando o *Programa R*, ajuste um modelo de regressão linear considerando as três variáveis simultaneamente. Escreva a equação do modelo ajustado, descrevendo os seus componentes e variáveis, no contexto do problema.

b) Avalie a significância dos parâmetros do modelo usando o Teste F de significância geral, construído com base na tabela de Análise de Variância (ANOVA). Utilize também o teste de significância individual, considerando o nível de 5%.

c) Interprete as estimativas dos parâmetros do modelo no contexto do problema.

d) Obtenha pelo menos duas medidas de qualidade do ajuste, e interprete-as no contexto do problema.

e) Avalie se a hipótese de normalidade dos erros é satisfeita, usando algum método gráfico (como, por exemplo, o *QQ Plot*) e algum teste estatístico (como por exemplo, o *Teste de Shapiro-Wilk e/ou Kolmogorov-Smirnov*).

f) Utilizando os resíduos estudentizados do modelo, verifique a hipótese de homocedasticidade dos erros e avalie se há ou não valores discrepantes (*outliers*)?

g) Estime a variância do erro do modelo, usando o método de mínimos quadrados (MQO) e de máxima verossimilhança (MV). As estimativas são diferentes ou iguais? Caso afirmativa, qual é a diferença relativa (em %)?

8ª Questão: Num estudo sobre manutenção de máquinas de bebidas, deseja-se explicar o tempo gasto (Y) para execução dos serviços (em minutos), em função da quantidade de bebida estocada (X_1) em máquinas acionadas por moeda (em unidades) e da distância percorrida (X_2) pelo profissional responsável pelos serviços (em pés). Os dados referentes a cada máquina, se encontram na tabela a seguir.

a) Faça gráficos apropriados para avaliar a relação das variáveis explicativas com a variável resposta do modelo. Complemente a análise gráfica, calculando alguma medida estatística.

b) Ajuste um modelo de regressão linear considerando as duas variáveis simultaneamente. Interprete as estimativas dos parâmetros do modelo e avalie sua significância estatística usando o teste significância individual (método do p-valor). Calcule alguma medida de qualidade do ajuste e interprete-a. Qual a sua conclusão?

c) Cheque se a hipótese de normalidade dos erros é satisfeita, usando a análise gráfica dos resíduos estudentizados (*QQ Plot*) e testes estatísticos (*Teste de Kolmogorov-Smirnov* e/ou *Teste de Shapiro-Wilk*) usando o nível de significância de 5%.

d) Utilizando os resíduos estudentizados, verifique a hipótese de homocedasticidade dos erros e avalie se há ou não valores discrepantes ou atípicos (*outliers*). Em caso de existência, quantos outliers?

Máquina	Tempo gasto de serviço (Y)	Quantidade de bebida (X_1)	Distância percorrida (X_2)
1	16,68	7	560
2	11,50	3	220
3	12,03	3	340
4	14,88	4	80
5	13,75	6	150
6	18,11	7	330
7	8,00	2	110
8	17,83	7	210
9	79,24	30	1460
10	21,50	5	605
11	40,33	16	688
12	21,00	10	215
13	13,50	4	255
14	19,75	6	462
15	24,00	9	448
16	29,00	10	776
17	15,35	6	200
18	19,00	7	132
19	9,50	3	36
20	35,10	17	770
21	17,90	10	140
22	52,32	26	810
23	18,75	9	450
24	19,83	8	635
25	10,75	4	150

9ª Questão: Alguns clientes procuraram um setor da empresa de plano de saúde para resolver problemas relacionados ao pagamento mensal dos seus planos e conhecer a carência do plano para seus recém-dependentes, em função de diferentes serviços de saúde. Os consultores de plantão atendem os clientes por ordem de chegada, incluindo a distribuição de senhas. A fim de avaliar a opinião dos clientes sobre o sistema de atendimento, um pesquisador registrou o tempo de espera (X_1), em minutos, e o tempo de atendimento efetivo (X_2), em minutos, para todos os clientes que procuraram tal setor, num dado período do dia. Após o atendimento, o pesquisador levantou algumas perguntas adicionais aos clientes, que o permitiu criar um índice de satisfação global (escala de 0 a 100, onde 100 indica máxima satisfação) de cada cliente sobre o sistema de atendimento utilizado por esta empresa de plano de saúde.

Usando as saídas do *programa SPSS* fornecidas a seguir, pede-se:

a) Defina o objetivo do pesquisador ao coletar essas informações.

b) Determine o número de clientes para os quais o pesquisador levantou tais informações.

c) Escreva a equação do modelo ajustado e avalie a significância dos parâmetros do modelo, considerando o nível de significância de 5%. Justifique.

d) Calcule o coeficiente de determinação do modelo, e interprete-o no contexto.

e) Construa ICs de 95% para os respectivos parâmetros β_0 , β_1 e β_2 do modelo.

f) Estime o índice médio de satisfação de clientes quando os tempos de espera e de atendimento são de 10 minutos.

g) Faça a previsão do índice de satisfação de um cliente quando o tempo de espera é de 50 minutos e o tempo de atendimento é de 15 minutos.

h) Interprete a estatística F da tabela ANOVA (contexto), e diga qual a sua conclusão extraída a partir desta estatística, considerando o nível de significância de 5%. E ao nível de 1%?

i) Obtenha uma estimativa para o desvio-padrão do erro do modelo.

OBS: *É necessário usar as notações apropriadas e mostrar todos os cálculos adicionais quando necessário.*

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1566,033	2	783,017	20,534	,000[a]
	Residual	838,912	22	38,132		
	Total	2404,945	24			

a. Predictors: (Constant), Tempo_atend, Tempo_espera

b. Dependent Variable: Escore_satisfacao_global

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	84,039	3,534		23,779	,000
	Tempo_espera	-,457	,096	-,596	-4,737	,000
	Tempo_atend	,377	,087	,544	4,316	,000

a. Dependent Variable: Escore_satisfacao_global

10ª Questão: Um experimento foi realizado com objetivo de avaliar se o teor de manganês (em partes por mil) e a espessura (em mm) afetam a resistência à tração (em kg/mm^2) em chapas de aço.

a) Analise a relação entre ambas as variáveis com a resistência à tração da chapa, com base nas Figuras 1 e 2:

Figura 1: Gráfico de dispersão entre o teor de manganês e a resistência das chapas.

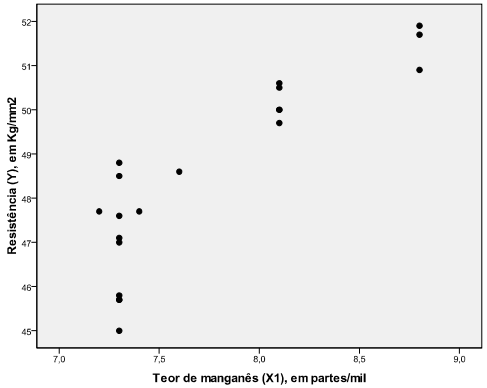
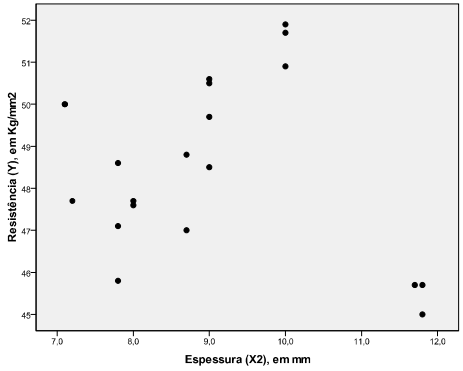


Figura 2: Gráfico de dispersão entre a espessura e a resistência das chapas.



b) Usando as *Saídas do SPSS*, analise os resultados do ajuste do modelo de RLM:

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	,929[a]	,862	,846	,8228	,862	53,188	2	17	,000

a. Predictors: (Constant), Espessura, Teor_manganes

b. Dependent Variable: Resistencia

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	72,010	2	36,005	53,188	,000[a]
	Residual	11,508	17	,677		
	Total	83,518	19			

a. Predictors: (Constant), Espessura, Teor_manganes

b. Dependent Variable: Resistencia

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	26,641	2,723		9,782	,000	20,895	32,387
	Teor_manganes	3,320	,332	,904	10,001	,000	2,620	4,021
	Espessura	-,425	,126	-,305	-3,371	,004	-,691	-,159

a. Dependent Variable: Resistencia

c) Usando a análise gráfica dos resíduos estudentizados do modelo (Figuras de 3 a 6), avalie as hipóteses de *homocedasticidade*, *independência* e *normalidade* dos erros. Há evidências de outliers? Escreva um pequeno relatório, apresentando as suas principais conclusões, e no último parágrafo informe se o modelo é adequado ou não para os dados observados.

Figura 3: Gráfico de dispersão entre os valores estimados da resistência das chapas e os resíduos estudentizados do modelo.

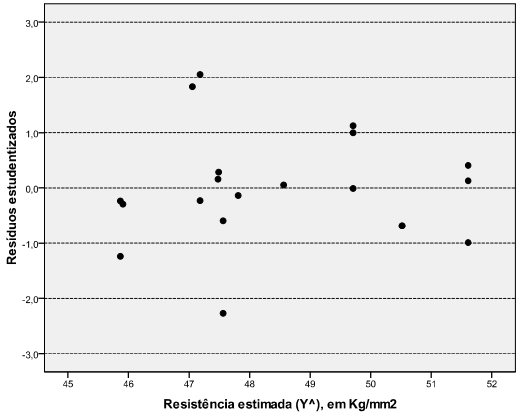


Figura 4: Gráfico de dispersão entre o teor de manganês e os resíduos estudentizados do modelo.

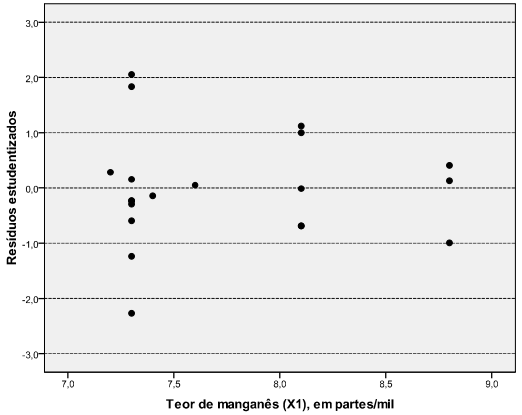


Figura 5: Gráfico de dispersão entre a espessura e os resíduos estudentizados do modelo.

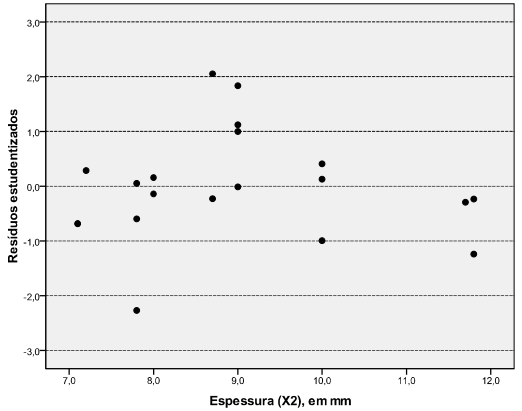
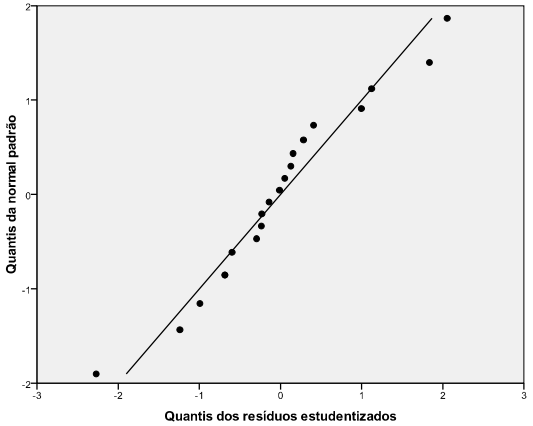


Figura 6: *QQ Plot* (normalidade) dos resíduos estudentizados do modelo.



Respostas da 3ª Lista de Exercícios:

"Modelos Lineares I"

1ª Questão:

Consultar notas de Aula do Prof. Dr. José Rodrigo, e/ou referências indicadas pelo professor.

2ª Questão:

a) $n = 6$

b) $\sum_{i=1}^{n} X_i^2 = 31.$

c) $\overline{X} = 0,5$

d) $(X'X) = \begin{bmatrix} 6 & 3 \\ 3 & 31 \end{bmatrix}$ e $det(X'X) = 177$

e) $X'Y = \begin{bmatrix} 10,5 \\ 79 \end{bmatrix}$

3ª Questão:

Consultar notas de Aula do Prof. Dr. José Rodrigo e/ou referências indicadas pelo professor.

4ª Questão:

Consultar notas de Aula do Prof. Dr. José Rodrigo.

5ª Questão:

1º Teste – *Hipóteses a serem testadas:*

H_0: Modelo reduzido: modelo com "proteína" e "peso"

H_1: Modelo completo: modelo com "proteína", "peso" e "idade"

Estatística de teste F:

$$f_{obs} = \frac{(SQ\mathrm{Re}s_0 - SQ\mathrm{Re}s_1)/(p-q)}{SQ\mathrm{Re}s_1/(n-p)} = \frac{(606,022 - 567,663)/(4-3)}{567,663/(20-4)} \cong 1,081$$

f_{obs} =1,081 ; $f_{1,16;\ 0,05}$ = 4,49 (ou p-valor=0,314 com auxílio computacional). Não rejeita-se H_0 ao nível de significância de 5%, ou seja, o modelo reduzido é tão adequado quanto ao modelo completo.

2º Teste – *Hipóteses a serem testadas:*

H_0: Modelo reduzido: modelo com "proteína"

H_1: Modelo completo: modelo com "proteína" e "peso"

Estatística de teste F:

$$f_{obs} = \frac{(SQ\mathrm{Re}s_0 - SQ\mathrm{Re}s_1)/(p-q)}{SQ\mathrm{Re}s_1/(n-p)} = \frac{(858,650 - 606,022)/(3-2)}{606,022/(20-3)} \cong 7,087$$

f_{obs} =7,087 ; $f_{1,17;\ 0,05}$ = 4,45 (ou p-valor=0,016). Rejeita-se H_0 ao nível de significância de 5%, ou seja, o modelo reduzido não é tão adequado quanto ao modelo completo.

OBS: *Outra(s) comparação(ões) é(são) necessária(s) até a seleção do modelo final !!!*

Equação do modelo final:

$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} = 33,130 + 1,824\, X_{i1} - 0,222\, X_{i2}$; $\forall \quad i = 1, 2, ..., 20$

6ª Questão:

a) Sugestão: *Fazer gráficos de dispersão entre X_1 e Y; e entre X_2 e Y.*

b) Aplicar o teste de comparabilidade de modelos (Teste F parcial):

1º Teste – *Hipóteses a serem testadas:*

H_0: Modelo reduzido: modelo com "eventos vividos" e "pse".

H_1: Modelo completo: modelo com "eventos vividos", "pse" e "eventos vividos*pse".

OBS: *Tem que representar o modelo !!!*

Estatística de teste F:

$$f_{obs} = \frac{\left(SQ\,\mathrm{Re}\,s_0 - SQ\,\mathrm{Re}\,s_1\right)\Big/\left(p-q\right)}{SQ\,\mathrm{Re}\,s_1\Big/\left(n-p\right)} = \frac{\left(768{,}162 - 758{,}769\right)\Big/\left(4-3\right)}{758{,}769\Big/\left(40-4\right)} \cong 0{,}446$$

f_{obs} =0,446 ; $f_{1,36;\,0,05}$ = 4,11 (ou p-valor=0,509 com auxílio computacional). Não rejeita-se H_0 ao nível de significância de 5%, ou seja, o modelo reduzido é tão adequado quanto ao modelo completo. Escolhe-se o modelo reduzido: $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$.

OBS: *Outras comparações são necessárias até a seleção do modelo final !!!*

Equação do modelo selecionado:

$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} = 28{,}230 + 0{,}103\, X_{i1} - 0{,}097\, X_{i2}$; $\forall \quad i = 1, 2, ..., 40$

c) R^2=33,9%

d) Índice médio estimado de distúrbio mental ≅ 27,42

Intervalo de confiança de 95% para o índice médio de distúrbio mental (X_1=45 e X_2=56) ≅ [25,96; 28,88]

e) Índice previsto de distúrbio mental ≅ 27,42

Intervalo de predição de 95% para o índice de distúrbio mental (X_1=45 e X_2=56) ≅ [18,07; 36,07]

7ª Questão:

a)

$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \hat{\beta}_3 X_{i3} = -28{,}422 + 10{,}920\, x_{i1} + 11{,}467\, X_{i2} + 1{,}758\, X_{i3}$;

$\forall \quad i = 1, 2, ..., 15$

b) Com exceção do intercepto do modelo, todos os demais parâmetros são significativamente diferentes de zero, ao nível de 5%, indicando que ...

f) A hipótese de homocedasticidade é satisfeita. Parece que não existe outliers.

OBS: *As demais letras por conta do aluno !!!*

8ª Questão:

a) Sugestão: *Fazer gráficos de dispersão entre X_1 e Y; e entre X_2 e Y. Calcular o coeficiente de correlação linear de Pearson entre X_1 e Y; e entre X_2 e Y.*

b) $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} = 2{,}341 + 1{,}616\, X_{i1} + 0{,}014\, X_{i2}$; $\forall \quad i = 1, 2, ..., 25$

R^2=96,0% → 96,0% da variação dos tempos gastos para a execução do serviço é explicado pelo modelo ajustado.

c) A hipótese de normalidade dos erros é satisfeita usando os resíduos estudentizados do modelo. Tanto o método gráfico quanto o método formal (ambos os testes) levam a mesma conclusão.

d) Existe 1 *outlier* (máquina 9) no conjunto de dados ($r_9^S = 3{,}214$).

9ª Questão:

a) *O que acha ???*

b) n=25 clientes

c) $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} = 84{,}039 - 0{,}457\, X_{i1} + 0{,}377\, X_{i2}$; $\forall \quad i = 1, 2, ..., 25$

Todos os parâmetros são significativos (inclusive o intercepto).

d) R^2=0,651. *Tem que interpretar no contexto!!!*

e) $IC_{\beta_0,95\%}$ =[76,710;91,369]; $IC_{\beta_1,95\%}$ =[-0,657;-0,257]; $IC_{\beta_2,95\%}$ =[0,196 ; 0,558];

f) Índice médio estimado de satisfação dos clientes: 83,24

g) Índice previsto de satisfação de um cliente: 66,84

h) p-valor<0,001 → Pelo menos uma das variáveis explicativas tem efeito significativo ao nível de 5% (e também ao nível de 1%).

i) $\hat{\sigma}$ = 6,175 ("*std. error of the estimate*")

10ª Questão:

a) Analise o sentido e o grau da relação entre X_1 e Y, e entre X_2 e Y.

b) Escreva a equação do modelo, avalie a significância individual e geral dos parâmetros do modelo, interprete as estimativas dos parâmetros e alguma medida de qualidade do ajuste, etc.

c) Analise cada um dos gráficos dos resíduos.