

Universidade Federal Fluminense (UFF)
 Instituto de Matemática e Estatística (IME)
 Departamento de Estatística (GET)

Projeto de Monitoria 2015

Exercícios de análise de regressão linear aplicados: uma abordagem usando o programa R

Monitora: Rosana Gayer Carvalho
Professor: Dr. José Rodrigo de Moraes

Universidade Federal Fluminense (UFF)

Instituto de Matemática e Estatística (IME)

Departamento de Estatística (GET)

Disciplina: Modelos Lineares I

Professor: Dr. José Rodrigo de Moraes

Monitora: Rosana Gayer Carvalho (Graduação de Estatística)

Exercícios de monitoria de Modelos Lineares I

Assuntos: Modelos de regressão linear simples e múltipla

Modelo de regressão logística binária

Exercício 1: Considere um estudo com o objetivo de avaliar a associação entre a região brasileira (Norte, Nordeste, Sudeste, Sul e Centro-oeste) e a taxa de municípios com serviço de abastecimento de água ligados a rede geral de distribuição para os estados brasileiros. Considerando a saída do R fornecida abaixo, responda as seguintes questões:

```
> summary(modelo)
```

```
call:
lm(formula = base$Taxaabast ~ base$Reg_Norte + base$Reg_Nordeste +
    base$Reg_Sudeste + base$Reg_Sul)

Residuals:
    Min       1Q   Median       3Q      Max
-0.063000 -0.000333  0.002667  0.009667  0.014000

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.996500   0.009125  109.203  <2e-16 ***
base$Reg_Norte -0.010500   0.011439  -0.918   0.369
base$Reg_Nordeste -0.006167  0.010967  -0.562   0.580
base$Reg_Sudeste  0.003500   0.012905   0.271   0.789
base$Reg_Sul     0.001167   0.013939   0.084   0.934
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01825 on 22 degrees of freedom
Multiple R-squared:  0.08795, Adjusted R-squared:  -0.07788
F-statistic: 0.5304 on 4 and 22 DF, p-value: 0.7147
```

- a) Qual modelo foi utilizado para estabelecer a associação entre essas variáveis? Qual a variável explicativa e a variável resposta do modelo? Qual o número de estados considerados na modelagem?
- b) Represente a equação do modelo teórico, descrevendo todos os seus termos e variáveis, no contexto do problema.
- c) Interprete as estimativas dos parâmetros do modelo no contexto do problema.
- d) Avalie a significância da associação entre a região e a taxa municipal com serviços de abastecimento de água ligados a rede geral de distribuição. Justifique a sua resposta.

Exercício 2: Considerando os dados do IBGE avalie a existência de relação entre a densidade demográfica (hab/km²) e o indicador de desenvolvimento humano (IDH) dos 92 municípios do Estado do Rio de Janeiro. Os dados se encontram na Tabela 1:

Tabela 1: Dados sobre o IDH, População (em habitantes) em 2010, Área (em km²) dos municípios do Estado do Rio de Janeiro.

Id Município	IDH	Pop.	Área	Id Município	IDH	Pop.	Área
1 Angra dos Reis	0,724	169511	825,1	47 Nilópolis	0,753	157425	19,4
2 Aperibé	0,692	10213	94,6	48 Niterói	0,837	487562	133,9
3 Araruama	0,718	112008	638	49 Nova Friburgo	0,745	182082	933,4
4 Areal	0,684	11423	110,9	50 Nova Iguaçu	0,713	796257	521,3
5 Armação dos Búzios	0,728	27560	70,3	51 Paracambi	0,720	47124	179,7
6 Arraial do Cabo	0,733	27715	160,3	52 Paraiba do Sul	0,702	41084	580,5
7 Barra do Pirai	0,733	94778	579	53 Paraty	0,693	37533	925,4
8 Barra Mansa	0,729	177813	547,2	54 Paty do Alferes	0,671	26359	318,8
9 Belford Roxo	0,684	469332	77,8	55 Petrópolis	0,745	295917	795,8
10 Bom Jardim	0,660	25333	384,6	56 Pinheiral	0,715	22719	76,5
11 Bom Jesus do Itabapoana	0,732	35411	598,8	57 Pirai	0,708	26314	505,4
12 Cabo Frio	0,735	186227	410,4	58 Porciúncula	0,697	17760	302
13 Cachoeiras de Macacu	0,700	54273	953,8	59 Porto Real	0,713	16592	50,8
14 Cambuci	0,691	14827	561,7	60 Quatis	0,690	12793	286,1
15 Campos dos Goytacazes	0,716	463731	4026,7	61 Queimados	0,680	137962	75,7
16 Cantagalo	0,709	19830	749,3	62 Quissamã	0,704	20242	712,9
17 Carapebus	0,713	13359	308,1	63 Resende	0,768	119769	1094,4
18 Cardoso Moreira	0,648	12600	524,6	64 Rio Bonito	0,710	55551	456,5
19 Carmo	0,696	17434	324,7	65 Rio Claro	0,683	17425	837,3
20 Casimiro de Abreu	0,726	35347	460,8	66 Rio das Flores	0,680	8561	478,3

(Continua)

Tabela 1 (Continuação): Dados sobre o IDH, População (em habitantes) em 2010, Área (em km²) dos municípios do Estado do Rio de Janeiro.

Id Município	IDH	Pop.	Área	Id Município	IDH	Pop.	Área
21 Comendador Levy Gasparian	0,685	8180	106,9	67 Rio das Ostras	0,773	105676	229
22 Conceição de Macabu	0,712	21211	347,3	68 Rio de Janeiro	0,799	6320446	1197,5
23 Cordeiro	0,729	20430	116,4	69 Santa Maria Madalena	0,668	10321	814,8
24 Duas Barras	0,659	10930	375,1	70 Santo Antônio de Pádua	0,718	40589	603,4
25 Duque de Caxias	0,711	855048	467,6	71 São Fidélis	0,691	37543	1031,6
26 Engenheiro Paulo de Frontin	0,722	13237	132,9	72 São Francisco Itabapoana	0,639	41354	1122,4
27 Guapimirim	0,698	51483	360,8	73 São Gonçalo	0,739	999728	247,7
28 Iguaba Grande	0,761	22851	52	74 São João da Barra	0,671	32747	455
29 Itaboraí	0,693	218008	430,4	75 São João de Meriti	0,719	458673	35,2
30 Itaguaí	0,715	109091	275,9	76 São José de Ubá	0,652	7003	250,3
31 Italva	0,688	14063	293,8	77 São José do Vale do R.Preto	0,660	20251	220,3
32 Itaocara	0,713	22899	431,3	78 São Pedro da Aldeia	0,712	87875	332,8
33 Itaperuna	0,730	95841	1105,3	79 São Sebastião do Alto	0,646	8895	397,9
34 Itatiaia	0,737	28783	245,2	80 Sapucaia	0,675	17525	541,2
35 Japeri	0,659	95492	81,9	81 Saquarema	0,709	74234	353,6
36 Laje do Muriaé	0,668	7487	250	82 Seropédica	0,713	78186	283,8
37 Macaé	0,764	206728	1216,9	83 Silva Jardim	0,654	21349	937,6
38 Macuco	0,703	5269	77,7	84 Sumidouro	0,611	14900	395,5
39 Magé	0,709	227322	388,5	85 Tanguá	0,654	30732	145,5
40 Mangaratiba	0,753	36456	356,4	86 Teresópolis	0,730	163746	770,6
41 Maricá	0,765	127461	362,6	87 Trajano de Moraes	0,667	10289	589,8
42 Mendes	0,736	17935	97	88 Três Rios	0,725	77432	326,8
43 Mesquita	0,737	168376	39,1	89 Valença	0,738	71843	1304,8
44 Miguel Pereira	0,745	24642	289,2	90 Varre-Sai	0,659	9475	190,1
45 Miracema	0,713	26843	304,5	91 Vassouras	0,714	34410	538,1
46 Natividade	0,730	15082	386,7	92 Volta Redonda	0,771	257803	182,5

Pede-se:

a) Calcule a densidade demográfica para cada município, construa e analise o gráfico de dispersão entre a densidade demográfica e o IDH, e calcule o coeficiente de correlação linear de Pearson.

b) Represente e ajuste um modelo de regressão linear simples para avaliar se existe relação significativa entre a densidade demográfica e o IDH. Use o teste F ou o Teste T, considerando o nível de significância de $\alpha=5\%$. **OBS:** Para o teste adotado por você, defina: 1) Hipóteses a serem testadas; 2) Estatística de teste; 3) Região crítica (RC); e 4) Tomada de decisão (inclusive no contexto do problema).

c) Interprete as estimativas pontuais dos parâmetros do modelo ajustado na letra (b) no contexto do problema. Pode-se concluir que municípios com alta densidade demográfica tendem a possuir maior grau de desenvolvimento humano? Justifique a sua resposta.

d) Avalie se as hipóteses de normalidade e homocedasticidade dos erros são satisfeitas usando métodos gráficos.

e) Aplique o logaritmo (neperiano) da variável “densidade demográfica”, construa o gráfico de dispersão entre o logaritmo da densidade demográfica e o IDH, e calcule o coeficiente de correlação linear de Pearson.

f) Ajuste um novo modelo de regressão linear simples para avaliar se existe relação significativa entre o logaritmo (neperiano) da densidade demográfica e o IDH, e verifique agora se ambas as hipóteses de normalidade e homocedasticidade dos erros foram satisfeitas.

g) Qual modelo você escolheria? Justifique a sua resposta.

Exercício 3: Os dados sobre o valor da produção (em R\$ 1.000) de banana (cacho) e a área destinada a colheita (em hectares) para 26 estados brasileiros, são apresentados na Tabela 2.

Tabela 2: Produção (em R\$ 1.000) de banana (cacho) e área destinada à colheita (em hectares).

Estado*	Área	Produção
Rondônia	8277	80211
Acre	7672	39009
Amazonas	8567	86956
Roraima	9714	114606
Pará	43510	411926
Amapá	2012	18151
Tocantins	3461	23795
Maranhão	9921	76284
Piauí	1957	36430
Ceará	49255	301883
Rio Grande do Norte	5200	86678
Paraíba	12507	112914
Pernambuco	36783	225959
Alagoas	3927	16461
Sergipe	3177	40591
Bahia	74077	797503
Minas Gerais	41430	845351

(Continua)

Tabela 2 (Continuação): Produção (em R\$ 1.000) de banana (cacho) e área destinada à colheita (em hectares).

Estado*	Área	Produção
Espírito Santo	21793	169145
Rio de Janeiro	22365	99973
São Paulo	51434	687484
Paraná	11707	136961
Santa Catarina	29261	312335
Rio Grande do Sul	12208	102100
Mato Grosso do Sul	1402	10870
Mato Grosso	6300	131611
Goiás	12521	146261

*Exceto o Distrito Federal

Fonte: IBGE, Diretoria de Pesquisas, Coordenação de Agropecuária, Produção Agrícola Municipal 2013.

a) Faça o gráfico de dispersão e calcule o coeficiente de correlação linear de Pearson entre o valor da produção e a área destinada a colheita.

b) Ajuste um modelo de regressão linear normal para explicar a variabilidade dos valores da produção de banana, a partir da área destinada a colheita. Avalie a significância dos parâmetros do modelo e as hipóteses de normalidade e homocedasticidade dos erros. Qual a sua conclusão?

c) Aplique o teste de White ($\alpha=5\%$), para confirmar se há violação ou não da hipótese de homocedasticidade para o modelo ajustado na letra (b). **OBS:** Para o teste adotado por você, defina: 1) Hipóteses a serem testadas; 2) Estatística de teste; 3) Região crítica (RC); e 4) Tomada de decisão (inclusive no contexto do problema).

d) No caso de presença de heterocedasticidade, faça alguma correção para contornar tal violação e interprete as estimativas pontuais dos parâmetros do modelo em termos das variáveis originais. Mostre por meio de métodos gráficos que a hipótese de homocedasticidade e normalidade são aproximadamente satisfeitas.

Exercício 4: Considerando os dados do Censo Agropecuário 2006 do IBGE sobre as despesas totais realizadas pelos estabelecimentos (em R\$ 1.000) e a quantidade de estabelecimentos agropecuários nos municípios fluminenses (Tabela 3). Na despesa total estão incluídas aquelas despesas provenientes: do arrendamento de terras, da contratação de serviços, dos salários pagos em dinheiro ou produtos para pessoas da família (inclusive 13º, férias e encargos), salários pagos em dinheiro ou produtos para empregados (inclusive 13º, férias e encargos), adubos, corretivos do solo, sementes e mudas, agrotóxicos, armazenamento da produção,

transporte da produção, compras de animais, medicamentos para animais, sal e rações, entre outros tipos de despesas.

a) Construa um gráfico apropriado para representar a relação entre a despesa total e o número de estabelecimentos agropecuários nos municípios fluminenses. Faça a análise gráfica.

b) Ajuste e represente o modelo de regressão linear normal para explicar a variabilidade das despesas totais em cada município, a partir do número de estabelecimentos agropecuários. Avalie a significância individual dos parâmetros do modelo, a qualidade global do ajuste e as hipóteses básicas do modelo.

c) Caso exista violação de alguma das hipóteses, identifique-a e proponha um novo modelo estatístico a ser ajustado aos dados observados. **Sugestão:** Utilize $X^* = \log(\text{Estab})$ e $Y^* = \log(\text{Despesa})$.

d) Ajuste e represente o modelo proposto na letra (c), avalie a significância individual dos parâmetros do modelo e a qualidade global do ajuste; e verifique se todas as hipóteses básicas do modelo foram satisfeitas.

e) Com base no modelo ajustado na letra (d), escreva o modelo em função das variáveis originais.

Tabela 3: Despesa total (em R\$1.000) e número de estabelecimentos agropecuários nos municípios fluminenses.

Id	Município*	Estab	Despesa	Id	Município*	Estab	Despesa
1	Bom Jesus do Itabapoana	1029	11171,62	47	Saquarema	312	2963,49
2	Italva	414	3710,49	48	Barra Mansa	734	13853,84
3	Itaperuna	1174	15291,23	49	Itatiaia	43	1420,47
4	Laje do Muriaé	394	3533,28	50	Pinheiral	107	2204,41
5	Natividade	443	6310,33	51	Pirai	417	5768,92
6	Porciúncula	1263	9025,93	52	Porto Real	42	1373,55
7	Varre-Sai	641	7819,2	53	Quatis	289	3698,52
8	Aperibé	224	1400,34	54	Resende	498	12782,68
9	Cambuci	1121	9930,62	55	Rio Claro	609	26067,37
10	Itaocara	1575	10886,67	56	Volta Redonda	205	3935,11
11	Miracema	372	5555,18	57	Barra do Pirai	234	19172,12
12	Santo Antônio de Pádua	942	7041,13	58	Rio das Flores	206	6282,88
13	São José de Ubá	427	4063,93	59	Valença	1050	16000,72

(Continua)

Tabela 3 (Continuação): Despesa total (em R\$1.000) e número de estabelecimentos agropecuários nos municípios fluminenses.

Id	Município*	Estab	Despesa	Id	Município*	Estab	Despesa
14	Campos dos Goytacazes	7507	71372,39	60	Angra dos Reis	269	1159,08
15	Cardoso Moreira	626	4943,73	61	Parati	429	1963,4
16	São Fidélis	3151	13746,38	62	Engenheiro P. de Frontin	43	2405,02
17	São Francisco de Itabapoana	3190	29463,71	63	Mendes	18	389,81
18	São João da Barra	673	3033,98	64	Miguel Pereira	102	2441,2
19	Carapebus	152	1591,57	65	Paracambi	203	1419,51
20	Conceição de Macabu	202	2555,94	66	Paty do Alferes	406	8219,26
21	Macaré	554	7792,96	67	Vassouras	322	12482,56
22	Quissamã	247	5211,3	68	Petrópolis	489	21323,06
23	Areal	103	2244,74	69	São José do V do Rio Preto	479	23940,59
24	Comendador Levy Gasparian	72	1563,39	70	Teresópolis	2801	34444,76
25	Paraíba do Sul	355	14799,05	71	Cachoeiras de Macacu	1534	45748,01
26	Sapucaia	674	12019,31	72	Rio Bonito	601	10807
27	Três Rios	222	9080,15	73	Itaguaí	360	2833,65
28	Cantagalo	853	9738,84	74	Mangaratiba	206	1389,05
29	Carmo	470	5681,62	75	Seropédica	390	3648,04
30	Cordeiro	175	2588,68	76	Belford Roxo	27	144,4
31	Macuco	44	535,95	77	Duque de Caxias	264	2453,54
32	Bom Jardim	753	9361,84	78	Guapimirim	156	6032,45
33	Duas Barras	419	8221,97	79	Itaboraí	247	8705,65
34	Nova Friburgo	1641	23702,96	80	Japeri	269	1081,42
35	Sumidouro	2503	29465,49	81	Magé	425	3420,57
36	Santa Maria Madalena	715	7473,86	82	Maricá	151	3295,66
37	São Sebastião do Alto	753	10051,58	83	Mesquita	70	276,29
38	Trajano de Moraes	864	5636,32	84	Niterói	16	632,66
39	Casimiro de Abreu	225	4486,65	85	Nova Iguaçu	390	2308,11
40	Rio das Ostras	182	3010,96	86	Queimados	118	553,39
41	Silva Jardim	368	12629,82	87	Rio de Janeiro	1020	15681,14
42	Araruama	705	6470,95	88	São Gonçalo	217	1097,24
43	Armação dos Búzios	10	353,08	89	Tanguá	299	2614,53
44	Cabo Frio	173	5681,92				
45	Iguaba Grande	41	437,62				
46	São Pedro da Aldeia	247	3238,56				

* Nilópolis, São João de Meriti e Arraial do Cabo não apresentaram informações para pelo menos uma das variáveis. **Fonte:** IBGE, Censo Agropecuário 2006.

Exercício 5: Considerando os dados do Censo Agropecuário 2006 do IBGE (Tabela 4) sobre a quantidade vendida (1.000 frutos), o número de pés colhidos (1.000 pés) e a área colhida (em hectares - ha) de coco-da-baía para cada município fluminense, responda os itens a seguir:

- a) Faça os gráficos de dispersão para avaliar a relação entre as variáveis explicativas “pés colhidos” e “área colhida” com a variável resposta “Q_vendida”.
- b) Ajuste um modelo de regressão linear considerando as duas variáveis simultaneamente e avalie a existência de multicolinearidade, usando métodos gráficos e medidas estatísticas. Para o modelo selecionado usando o teste T de significância individual ($\alpha=5\%$), verifique se a hipótese de normalidade é violada através do teste de Kolmogorov-Smirnov ou teste de Shapiro-Wilk e o *QQ-Plot* dos resíduos estudentizados do modelo.
- c) Caso ocorra violação da hipótese de normalidade, faça transformações nos dados e ajuste e represente um novo modelo. **Sugestão:** Use a transformação logarítmica (base 10) para todas as variáveis consideradas no estudo. Avalie a significância dos parâmetros do modelo usando o Teste T de significância individual ($\alpha=5\%$) e interprete as estimativas dos parâmetros do modelo.
- d) Verifique se a hipótese de normalidade do novo modelo é satisfeita através do teste de Kolmogorov-Smirnov ou teste de Shapiro-Wilk e o *QQ-Plot* dos resíduos estudentizados do novo modelo. Justifique sua resposta. Calcule alguma medida de qualidade do ajuste, e interprete-a no contexto do problema.

Tabela 4: Dados sobre quantidade e colheita de coco-da-baía para n=39 municípios do Estado do Rio de Janeiro que possuíam estabelecimentos com mais de 50 pés existentes de coco-da-baía.

Id	Municípios	Q_Vendida (1.000 frutos)	Pés colhidos (1.000 pés)	Área colhida (ha)
1	Bom Jesus do Itabapoana	48	1	5
2	Natividade	8	2	3
3	Cambuci	31	2	5
4	Itaocara	78	2	11
5	Miracema	30	1	5
6	Santo Antônio de Pádua	115	42	10
7	Campos dos Goytacazes	1354	55	237
8	São Fidélis	4	1	4
9	São Francisco de Itabapoana	1283	54	175
10	São João da Barra	143	5	16

(Continua)

Tabela 4 (Continuação): Dados sobre quantidade e colheita de coco-da-baía para n=39 municípios do Estado do Rio de Janeiro que possuíam estabelecimentos com mais de 50 pés existentes de coco-da-baía.

Id	Municípios	Q_Vendida (1.000 frutos)	Pés colhidos (1.000 pés)	Área colhida (ha)
11	Carapebus	97	5	21
12	Conceição de Macabu	292	11	50
13	Macaé	23	2	7
14	Quissamã	1746	55	172
15	Santa Maria Madalena	117	6	12
16	Casimiro de Abreu	6	1	6
17	Rio das Ostras	84	5	22
18	Silva Jardim	1213	28	128
19	Araruama	902	30	105
20	Cabo Frio	102	6	22
21	São Pedro da Aldeia	13	1	4
22	Saquarema	624	33	151
23	Parati	19	2	7
24	Paracambi	18	4	1
25	Cachoeiras de Macacu	1262	30	136
26	Rio Bonito	350	12	34
27	Itaguaí	2376	87	300
28	Mangaratiba	11	2	5
29	Seropédica	1352	100	156
30	Duque de Caxias	71	4	15
31	Guapimirim	13	3	15
32	Itaboraí	115	9	26
33	Japeri	13	2	9
34	Magé	76	6	21
35	Maricá	107	3	5
36	Nova Iguaçu	138	7	26
37	Rio de Janeiro	2082	47	212
38	São Gonçalo	61	9	33
39	Tanguá	539	23	75

Fonte: IBGE, Censo Agropecuário 2006

Exercício 6: Os dados disponibilizados neste exercício são provenientes da Diretoria de Pesquisas, Coordenação de Agropecuária, Produção Agrícola Municipal 2007 do IBGE (Tabela 5). Tais dados se referem as seguintes variáveis em nível municipal: área colhida (ha), quantidade produzida (em toneladas) e valor da produção (em R\$ 1.000) de milho (em grãos):

Tabela 5: Dados sobre área colhida (hectares - ha), quantidade produzida (em toneladas) e valor da produção de milho (em grãos) para 55 municípios fluminense com produção de milho.

Id	Município	Área colhida (hectares)	Quantidade produzida (toneladas)	Valor (R\$ 1.000)
1	Aperibé	45	90	45
2	Araruama	112	173	52
3	Areal	40	80	28
4	Barra Mansa	40	72	22
5	Bom Jardim	92	312	112
6	Bom Jesus do Itabapoana	1000	2000	1000
7	Cabo Frio	8	8	4
8	Cambuci	560	1120	560
9	Campos dos Goytacazes	558	837	167
10	Cantagalo	150	540	180
11	Carapebus	6	10	4
12	Cardoso Moreira	368	625	156
13	Carmo	225	675	243
14	Casimiro de Abreu	50	116	46
15	Comendador Levy Gasparian	15	36	21
16	Cordeiro	80	160	58
17	Duas Barras	240	720	269
18	Itaboraí	45	70	34
19	Italva	320	730	438
20	Itaocara	200	400	200
21	Itaperuna	1800	3600	1260
22	Itatiaia	25	70	39
23	Laje do Muriaé	40	72	24
24	Macaé	200	600	180
25	Macuco	12	36	12
26	Magé	30	60	9
27	Miracema	280	812	406
28	Natividade	234	561	292
29	Nova Friburgo	35	87	32
30	Paraíba do Sul	115	325	114
31	Porciúncula	180	450	189
32	Porto Real	30	84	46
33	Quatis	20	50	15
34	Quissamã	100	150	53
35	Resende	160	448	233
36	Rio Bonito	4	6	2
37	Rio Claro	150	240	72
38	Rio das Flores	20	48	38
39	Rio das Ostras	30	60	30
40	Santa Maria Madalena	140	182	62
41	Santo Antônio de Pádua	202	444	222
42	São Fidélis	231	300	125
43	São Francisco de Itabapoana	100	150	62
44	São José de Ubá	400	1400	700
45	São Pedro da Aldeia	5	6	3

(Continua)

Tabela 5 (Continuação): Dados sobre área colhida (hectares - ha), quantidade produzida (em toneladas) e valor da produção de milho (em grãos) para 55 municípios fluminense com produção de milho.

Id	Município	Área colhida (hectares)	Quantidade produzida (toneladas)	Valor (R\$ 1.000)
46	São Sebastião do Alto	450	1125	383
47	Sapucaia	75	180	63
48	Saquarema	15	16	7
49	Silva Jardim	18	36	11
50	Sumidouro	190	646	240
51	Trajano de Moraes	150	375	131
52	Três Rios	30	66	20
53	Varre-Sai	520	1092	546
54	Vassouras	25	60	33
55	Volta Redonda	10	20	6

a) Calcule o rendimento médio de milho (kg/ha) para cada município usando a expressão abaixo:

$$Rend_{medio} = \frac{(Qde\ produzida * 1000)}{\text{Área colhida}}$$

b) Construa um gráfico de dispersão entre o rendimento médio de milho (Kg/ha) e o valor da produção de milho (em R\$ 1.000). Faça a análise gráfica.

c) Ajuste um modelo de regressão linear para explicar a variabilidade dos valores da produção de milho. Avalie a significância dos parâmetros do modelo usando o teste T ($\alpha=5\%$), e a hipótese de normalidade usando método gráfico.

d) Caso ocorra violação da hipótese de normalidade, faça alguma transformação na variável resposta e ajuste um novo modelo. **Sugestão:** Use a transformação logarítmica (base e). Avalie a significância dos parâmetros desse modelo e interprete as estimativas pontuais.

e) Calcule e interprete alguma medida de qualidade do ajuste, e verifique se as hipóteses básicas do modelo estão satisfeitas.

Exercício 7: Os dados disponibilizados neste exercício são provenientes da Diretoria de Pesquisas, Coordenação de Agropecuária, Produção Agrícola Municipal 2007 do IBGE. Tais dados se referem as seguintes variáveis referentes a municípios selecionados do Estado do Rio de Janeiro: rendimento médio de feijão (kg/ha) e valor da produção (em R\$1.000) de feijão (em grãos). Considerando a saída do R-Studio fornecida abaixo, responda as seguintes questões:

```
> summary(modelo)
```

```
Call:
lm(formula = base$Valor ~ base$Rend)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-59.44 -46.79   4.71  21.94  89.09
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1261.5770   176.7910  -7.136 0.000188 ***
base$Rend      1.8673     0.2223   8.402 6.66e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 54.53 on 7 degrees of freedom
Multiple R-squared:  0.9098, Adjusted R-squared:  0.8969
F-statistic: 70.59 on 1 and 7 DF, p-value: 6.658e-05
```

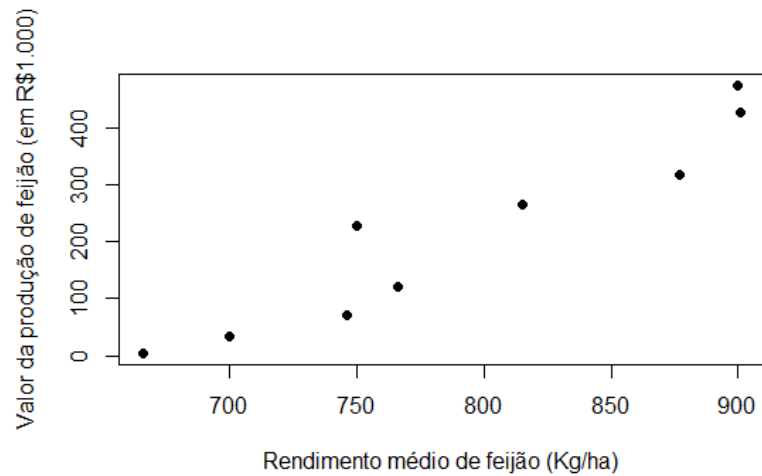


Figura. Gráfico de dispersão entre o rendimento médio de feijão (em Kg/ha) e o valor da produção de feijão (em R\$1.000).

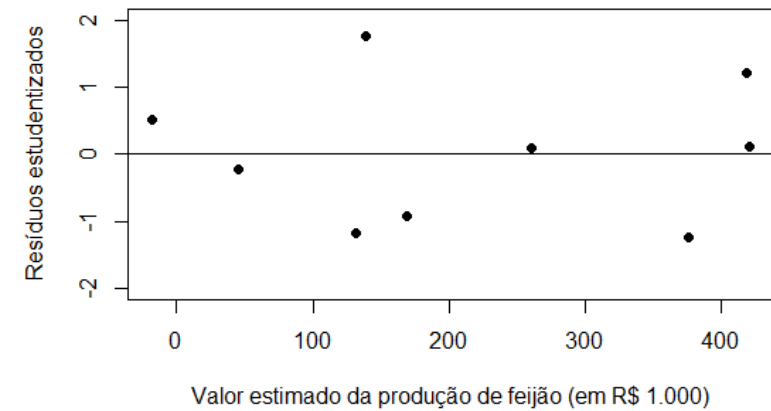


Figura. Gráfico de dispersão entre o valor estimado da produção de feijão (em R\$1.000) e os resíduos estudantizados do modelo.

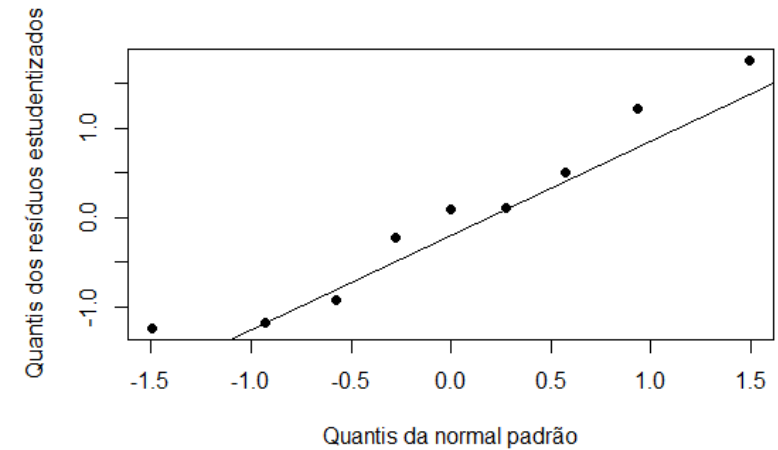


Figura. QQ plot dos resíduos estudantizados do modelo.

```
> shapiro.test(res_student)
```

```
Shapiro-wilk normality test
```

```
data: res_student
W = 0.9412, p-value = 0.5946
```

- a) Analise o gráfico de dispersão entre o rendimento médio de feijão (Kg/ha) e o valor da produção de feijão (em R\$ 1.000).
- b) Represente o modelo de regressão linear ajustado para explicar a variabilidade dos valores da produção de feijão. Obtenha e interprete o IC de 95% para os parâmetros do modelo e avalie a significância dos parâmetros do modelo usando o teste T ($\alpha=5\%$) descrevendo todas as etapas do teste. **OBS:** 1) Hipóteses a serem testadas; 2) Estatística de teste; 3) Região crítica (RC); e 4) Tomada de decisão (inclusive no contexto do problema).
- c) Construa a tabela ANOVA e interprete todos os seus componentes.
- d) Cheque se as hipóteses básicas do modelo são satisfeitas. Justifique com base nos métodos gráficos. Para avaliar a hipótese de normalidade utilize também o teste de Shapiro-Wilk.
- e) Interprete as estimativas pontuais dos parâmetros do modelo ajustado e calcule e interprete o coeficiente de determinação do modelo.

Exercício 8: Um estudo dirigido aos municípios brasileiros levantou informação sobre a ocorrência ou não de desastres naturais ocasionados por escorregamentos ou deslizamentos de encostas. Essa informação sobre ocorrência de desastres, bem como as grandes regiões brasileiras e o tamanho populacional para 5.298 municípios brasileiros se encontram no arquivo “deslizamento.sav”.

Variáveis	Categorias
Ocorrência de escorregamentos ou deslizamentos de encostas	0=Não 1=Sim
Grandes regiões brasileiras	1=Norte 2=Nordeste 3=Sudeste 4=Sul 5=Centro-Oeste (referência)
Tamanho populacional	1=Até 5.000 hab. 2=5.001 até 50.000 hab. 3=Mais do que 50.000 hab. (referência)

- a) Qual modelo você utilizaria para estimar a chance de ocorrência de desastres oriundos de escorregamentos ou deslizamentos de encostas, a partir das variáveis grandes regiões brasileiras e tamanho populacional? Justifique a sua resposta e represente o modelo descrevendo todos os seus termos e variáveis, no contexto do problema.
- b) Ajuste o modelo representado no item a) e avalie a significância individual dos parâmetros do modelo, usando o teste de Wald, considerando o nível de significância de 5%. Interprete as estimativas dos parâmetros do modelo.
- c) Avalie a capacidade preditiva do modelo. Calcule a taxa global de classificações corretas e os valores de sensibilidade e especificidade (considerando ponte de corte 0,5).

Exercício 9: Para 5.426 municípios brasileiros levantou-se informação sobre a ocorrência ou não de desastres naturais ocasionados por enchentes ou inundações nos últimos 5 anos. Além da informação sobre ocorrências de enchentes, encontra-se no arquivo “inundações.sav” a identificação da região geográfica a qual o município pertence e o tamanho populacional.

Variáveis	Categorias
Ocorrência de enchentes ou inundações	0=Não 1=Sim
Grandes regiões brasileiras	1=Norte 2=Nordeste 3=Sudeste 4=Sul 5=Centro-Oeste (referência)
Tamanho populacional	1=Até 5.000 hab. 2=Mais do que 5.000 até 50.000 hab. 3=Mais do que 50.000 hab. (referência)

- a) Qual modelo você utilizaria para estimar a chance de ocorrência de desastres ocasionados por enchentes ou inundações, a partir das variáveis grandes regiões brasileiras e tamanho populacional? Justifique a sua resposta e represente o modelo descrevendo todos os seus termos e variáveis, no contexto do problema.

- b) Avalie a significância da relação entre chance de ocorrência de desastres dessa natureza e as variáveis grandes regiões brasileiras e tamanho populacional, com base no teste de Wald de significância individual dos parâmetros do modelo.
- c) Represente o modelo ajustado (selecionado) e interprete as estimativas pontuais no contexto do problema.
- d) Calcule alguma medida de qualidade do ajuste do modelo selecionado e interprete-a (contexto).
- e) Represente graficamente a relação entre a probabilidade estimada de ocorrência de enchentes ou inundações e as grandes regiões brasileiras e tamanho populacional.

Exercício 10: Com base nos dados de 92 municípios do Estado do Rio de Janeiro (MUNIC 2009), deseja-se avaliar se as características sócio demográficas e tamanho populacional estão significativamente associadas com o conhecimento do gestor municipal sobre a agenda de compromissos dos objetivos do milênio. Os dados referentes as variáveis do gestor e do município são fornecidos no arquivo “*conhecimento.sav*”.

Variáveis	Categorias
Conhecimento do gestor municipal sobre a agenda de compromissos dos objetivos do milênio	0=Não 1=Sim
Sexo	1=Feminino 0=Masculino (referência)
Faixa etária	1=De 21 até 39 anos 2=De 40 até 49 anos 3=De 50 até 59 anos 4=De 60 anos ou mais (referência)
Escolaridade	1= Ensino fundamental 2= Ensino médio 3= Ensino superior ou mais (referência)
Tamanho populacional	1=De 5.000 até 50.000 hab. 0=Mais do que 50.000 hab. (referência)

- a) Ajuste um modelo para explicar a chance de conhecimento do gestor sobre a agenda de compromissos dos objetivos do milênio, considerando como variáveis explicativas as suas características sociodemográficas e o tamanho populacional.

- b) Avalie a significância dos parâmetros do modelo, com base no teste de Wald de significância individual ($\alpha=5\%$).
- c) Represente o modelo ajustado (selecionado) e interprete as estimativas pontuais no contexto do problema.
- d) Calcule alguma medida de qualidade do ajuste do modelo selecionado e interprete-a (contexto).
- e) Faça a representação gráfica das probabilidades estimadas de conhecimento do gestor sobre a agenda considerado a(s) variável(eis) com efeito significativo.