

CE071 - ANÁLISE DE REGRESSÃO LINEAR
Prof^a Suely Ruiz Giolo

LISTA 1 - EXERCÍCIOS

- 1) Para o modelo de regressão linear simples normal, encontre os estimadores de máxima verossimilhança dos parâmetros β_0 , β_1 e σ^2 .
- 2) Mostre que os estimadores de MQ $\hat{\beta}_0$ e $\hat{\beta}_1$ são não-viciados.
- 3) Encontre as variâncias de $\hat{\beta}_0$ e $\hat{\beta}_1$.
- 4) Mostre que $\sum_{i=1}^n y_i = \sum_{i=1}^n \hat{y}_i$.
- 5) Mostre que a soma dos resíduos é igual a zero, isto é, $\sum_{i=1}^n e_i = 0$.
- 6) Mostre que $\sum_{i=1}^n x_i e_i = 0$.
- 7) Mostre que $\sum_{i=1}^n \hat{y}_i e_i = 0$.
- 8) Mostre que para $x = \bar{x}$ tem-se $\hat{y} = \bar{y}$.
- 9) Mostre que $\sum_{i=1}^n (x_i - \bar{x}) = 0$.
- 10) Mostre que $\sum_{i=1}^n y_i (x_i - \bar{x}) = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$.
- 11) Mostre que $\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2$.

CE071 - ANÁLISE DE REGRESSÃO LINEAR

Prof^a Suely Ruiz Giolo

LISTA 2 - EXERCÍCIOS

1) Para os dados na Tabela 1 encontre (se possível):

- Modelo de regressão linear que apresente boa adequação aos dados e que atenda aos pressupostos assumidos para os erros.
- Intervalo de confiança para os parâmetros do modelo com suas respectivas interpretações.
- Intervalo de confiança e de predição para dois valores possíveis de x e interpretações.
- Conclusões gerais do experimento.

Tabela 1. Dados de 53 clientes para o mês de agosto de 1995 fornecidos por uma empresa de eletricidade interessada (por questões de planejamento) em desenvolver um modelo relacionando a demanda em horário de pico (y) com o total de energia utilizada (x) no mês.

Cliente	Total de energia (kwh)	Demanda horário de pico (kw)
1	679	0.79
2	292	0.44
3	1012	0.56
4	493	0.79
5	582	2.70
6	1156	3.64
7	997	4.73
8	2189	9.50
9	1097	5.34
10	2078	6.85
11	1818	5.84
12	1700	5.21
13	747	3.25
14	2030	4.43
15	1643	3.16
16	414	0.50
17	354	0.17
18	1276	1.88
19	745	0.77
20	435	1.39
21	540	0.56
22	874	1.56
23	1543	5.28
24	1029	0.64
25	710	4.00
26	1434	0.31
27	837	4.20

Cliente	Total de energia (kwh)	Demanda horário de pico (kw)
28	1748	4.88
29	1381	3.48
30	1428	7.58
31	1255	2.63
32	1777	4.99
33	370	0.59
34	2316	8.19
35	1130	4.79
36	463	0.51
37	770	1.74
38	724	4.10
39	808	3.94
40	790	0.96
41	783	3.29
42	406	0.44
43	1242	3.24
44	658	2.14
45	1746	5.71
46	468	0.64
47	1114	1.90
48	413	0.51
49	1787	8.33
50	3560	14.94
51	1495	5.11
52	2221	3.85
53	1526	3.93

Fonte: Montgomery & Peck (1992).

2) O motor de um foguete é fabricado pela ligação de um propulsor de ignição e um propulsor sustentador dentro de uma caixa de metal. A resistência da ligação entre os dois tipos de propulsor é uma característica de qualidade importante. Suspeita-se que essa resistência esteja relacionada com a idade (em semanas) do lote do propulsor sustentador. Vinte observações sobre a resistência e a correspondente idade do lote do propulsor sustentador foram coletadas e são mostradas a seguir.

Observação i	Resistência (psi) y_i	Idade (semanas) x_i
1	2158.70	15.50
2	1678.15	23.75
3	2316.00	8.00
4	2061.30	17.00
5	2207.50	5.50
6	1708.30	19.00
7	1784.70	24.00
8	2575.00	2.50
9	2357.90	7.50
10	2256.70	11.00
11	2165.20	13.00
12	2399.55	3.75
13	1779.80	25.00
14	2336.75	9.75
15	1765.30	22.00
16	2053.50	18.00
17	2414.40	6.00
18	2200.50	12.50
19	2654.20	2.00
20	1753.70	21.50

Fonte: Montgomery & Peck (1992).

- Ajuste um modelo de regressão linear que apresente boa adequação aos dados e que atenda aos pressupostos assumidos para os erros.
- Construa intervalos de confiança para os parâmetros do modelo e interprete-os.
- Obtenha intervalos de confiança e predição para 4 valores de x e interprete-os.
- Verifique se os resultados de 4 a 8 da Lista 1 são satisfeitos.
- Forneça conclusões gerais para esse experimento.

3) Com o objetivo de estudar a relação entre a temperatura e o tempo de reação química, certo experimento foi realizado 35 vezes para 7 temperaturas diferentes. A seguir são fornecidas as temperaturas em °C e os tempos de reação obtidos em segundos.

Temperatura (X)	Tempo (Y)				
20	12.3	11.8	11.5	12.1	11.7
30	11.8	11.5	11.4	11.7	11.2
40	10.9	11.2	10.8	10.6	10.3
50	10.4	9.8	9.5	9.9	9.2
60	9.6	9.0	8.7	8.3	9.1
70	9.1	9.3	8.5	8.6	8.3
80	8.4	8.1	8.1	7.7	7.9

Fonte: Charmet et al. (2008).

- Analise os dados desse experimento fazendo uso de um MRLS.
- Tire conclusões para esse experimento.

4) Para se conhecer o efeito do carbono (em %), contido em fios de aço, em resistências (em μ ohms cm a 20°), foi realizado um experimento o qual forneceu os resultados a seguir.

Carbono (%) X	Resistência (μ ohms cm a 20°) Y			
0.05	11.7	12.4	13.2	
0.10	12.9	13.8	14.0	14.9
0.15	14.6	15.8		
0.20	15.0	15.9	16.4	
0.25	16.1	16.6	16.9	
0.30	16.5	17.3		
0.40	17.7	18.2	18.9	
0.50	19.2	19.5	20.4	
0.60	20.6	21.2	22.0	
0.70	22.0	22.5	23.1	23.7
0.75	22.4	23.6	24.0	
0.80	23.7	24.4	24.9	25.3
0.85	24.8	25.6	26.9	
0.90	26.4	27.0	27.8	28.0
0.95	27.1	27.8	28.5	

Fonte: Charnet et al. (2008).

(a) Analise os dados desse experimento e tire conclusões.

CE071 - ANÁLISE DE REGRESSÃO LINEAR
Prof^a Suely Ruiz Giolo

LISTA 3 - EXERCÍCIOS

1. Um experimento foi realizado para verificar o efeito da temperatura (X_1) e da concentração (X_2) na produção (Y) de certo produto químico. Os dados obtidos foram:

Y	X_1	X_2
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

Fonte: Charnet et al (2008).

- (a) Ajuste os modelos:
- 1) Y em função de X_1
 - 2) Y em função de X_2
 - 3) Y em função de X_1 e X_2
 - 4) Y em função de X_2 e X_1
 - 5) Y em função de X_1 , X_2 e $X_1 \cdot X_2$
 - 6) Y em função de X_2 , X_1 e $X_1 \cdot X_2$

e, para cada um desses modelos, registre em uma tabela: os coeficientes de regressão estimados, a SQres, o QMres e os coeficientes de determinação R^2 e R_a^2

(b) Apenas observando os resultados obtidos em (a), qual modelo selecionaria como sendo o *melhor candidato* para explicar a produção Y ?

(c) Obtenha medidas para detectar a presença de multicolinearidade e tire conclusões.

(d) Obtenha os resíduos do modelo selecionado e faça uma análise gráfica dos mesmos a fim de avaliar os pressupostos assumidos para os erros.

(e) Proceda ao diagnóstico de influência do modelo selecionado.

- (f) Represente matricialmente o modelo selecionado.
- (g) Teste as hipóteses: $H_0: \beta_1 = \beta_2 = 0$ versus H_a : pelo menos um β_j ($j = 1, 2$) difere de zero.
- (h) Teste as hipóteses: $H_0: \beta_j = 0$ versus $H_a: \beta_j \neq 0$ ($j = 0, 1, 2$) fazendo uso:
h₁) dos intervalos de confiança dos β_j ($j = 0, 1, 2$)
h₂) do teste t^* parcial
h₃) do teste F^* parcial
- (i) Obtenha e interprete os coeficientes de determinação simples e parcial.
- (j) Obtenha e interprete os coeficientes de correlação simples e parcial.
- (k) Qual a estimativa obtida para σ^2 ?
- (l) Interprete os parâmetros do modelo selecionado no contexto dos dados.
- (m) Obtenha, usando o modelo selecionado, os valores preditos da produção Y .
- (n) **n₁**) Fixe $x_2 = 10$ e obtenha $E(Y | x_1, x_2 = 10)$.
n₂) Fixe $x_2 = 25$ e obtenha $E(Y | x_1, x_2 = 25)$.
n₃) Represente graficamente os resultados obtidos em n_1) e n_2).
- (o) **o₁**) Fixe $x_1 = 100$ e obtenha $E(Y | x_1 = 100, x_2)$.
o₂) Fixe $x_1 = 130$ e obtenha $E(Y | x_1 = 130, x_2)$.
o₃) Represente graficamente os resultados obtidos em o_1) e o_2).
- (p) Obtenha o intervalo de 95% de confiança para a esperança de Y em $x_1 = 80$ e $x_2 = 10$.
- (q) Obtenha o gráfico dos dados observados e do plano ajustado.
- (r) Obtenha o gráfico somente do plano ajustado.

2. Considerando os dados apresentados a seguir provenientes de um experimento realizado para avaliar diversos tipos de ferramentas em que: Y = tempo de vida efetivo da ferramenta (em horas), X_1 = velocidade do torno (rpm) e X_2 = tipo da ferramenta.

Y	X1	X2	Y	X1	X2	Y	X1	X2	Y	X1	X2
18.73	610	A	30.16	670	B	13.51	615	C	24.51	615	D
14.52	950	A	27.09	770	B	4.03	900	C	4.03	900	D
17.43	720	A	25.40	880	B	6.54	810	C	6.54	810	D
14.54	840	A	26.05	1000	B	8.59	750	C	15.59	750	D
13.44	980	A	33.49	760	B	15.72	530	C	31.22	530	D
24.39	530	A	35.62	590	B	10.08	670	C	18.08	670	D
13.34	680	A	26.07	910	B	5.38	920	C	5.38	920	D
22.71	540	A	36.78	650	B	7.52	860	C	7.52	860	D
12.68	890	A	34.95	810	B	14.07	590	C	24.07	590	D
19.32	730	A	43.67	500	B	10.48	630	C	19.48	630	D

- (a) Analise os dados desse experimento considerando as ferramentas A, B e C nas análises.
(b) Faça o mesmo considerando agora as ferramentas A, B e D.
(c) Idem considerando agora as ferramentas A, B, C e D.

3. Na avaliação de candidatos a um trabalho foram registrados os dados a seguir em que: Y = performance para o trabalho sob avaliação, X_1 = teste para verificar as habilidades relacionadas ao trabalho e X_2 = raça (branca = 0 e outras = 1)

	X1	X2	Y
1	0.28	1	0.83
2	0.97	1	2.59
3	1.25	1	2.97
4	2.46	1	8.14
5	2.51	1	8.00
6	1.17	1	3.30
7	1.78	1	6.53
8	1.21	1	2.03
9	1.63	1	5.00
10	1.98	1	5.04
11	2.36	0	4.25
12	2.11	0	5.30
13	0.45	0	2.39
14	1.76	0	3.69
15	2.09	0	5.56
16	1.50	0	3.00
17	1.25	0	3.85
18	0.72	0	2.90
19	0.42	0	2.85
20	1.53	0	2.95

(a) Analise os dados e tire conclusões.

4. Para estudar a relação entre anos de experiência = EXP, nível educacional = EDUC e função administrativa = MGT (sim ou não) com o salário anual em dólares = SAL = Y em que: EDUC: 1 = high school, 2 = college e 3 = advanced degree e MGT: 1 = sim (exerce função administrativa) e 0 = não, foram coletados os dados a seguir.

	EXP	EDUC	MGT	SAL		EXP	EDUC	MGT	SAL
1	1	1	1	13876	24	6	2	1	22884
2	1	3	0	11608	25	7	1	1	16978
3	1	3	1	18701	26	8	2	0	14803
4	1	2	0	11283	27	8	1	1	17404
5	1	3	0	11767	28	8	3	1	22184
6	2	2	1	20872	29	8	1	0	13548
7	2	2	0	11772	30	10	1	0	14467
8	2	1	0	10535	31	10	2	0	15942
9	2	3	0	12195	32	10	3	1	23174
10	3	2	0	12313	33	10	2	1	23780
11	3	1	1	14975	34	11	2	1	25410
12	3	2	1	21371	35	11	1	0	14861
13	3	3	1	19800	36	12	2	0	16882
14	4	1	0	11417	37	12	3	1	24170
15	4	3	1	20263	38	13	1	0	15990
16	4	3	0	13231	39	13	2	1	26330
17	4	2	0	12884	40	14	2	0	17949
18	5	2	0	13245	41	15	3	1	25685
19	5	3	0	13677	42	16	2	1	27837
20	5	1	1	15965	43	16	2	0	18838
21	6	1	0	12336	44	16	1	0	17483
22	6	3	1	21352	45	17	2	0	19207
23	6	2	0	13839	46	20	1	0	19346

(a) Analise os dados obtendo um modelo que se ajuste bem aos dados.

(b) Usando um IC de 95% estime o quanto, em média, aumenta o salário anual a cada acréscimo de um ano de experiência.

(c) Verifique o quanto as variáveis qualitativas afetam as diferenças salariais. Apresente os resultados na tabela a seguir.

EDUC	MGT	Coefficientes	Estimativa do salário base
1	1	$\beta_0 + \beta_4$	$9458,378 + 3988,817 = 13447,20$
1	0		
2	1		
2	0		
3	1		
3	0		

5. Analise os dados a seguir fazendo uso de regressão polinomial.

	X	Y		X	Y
1	0.0	8.33	22	10.5	14.92
2	0.5	8.23	23	11.0	14.37
3	1.0	7.17	24	11.5	14.63
4	1.5	7.14	25	12.0	15.18
5	2.0	7.31	26	12.5	14.51
6	2.5	7.60	27	13.0	14.34
7	3.0	7.94	28	13.5	13.81
8	3.5	8.30	29	14.0	13.79
9	4.0	8.76	30	14.5	13.05
10	4.5	8.71	31	15.0	13.04
11	5.0	9.71	32	15.5	12.60
12	5.5	10.26	33	16.0	12.05
13	6.0	10.91	34	16.5	11.15
14	6.5	11.67	35	17.0	11.15
15	7.0	11.76	36	17.5	10.14
16	7.5	12.81	37	18.0	10.08
17	8.0	13.30	38	18.5	9.78
18	8.5	13.88	39	19.0	9.80
19	9.0	14.59	40	19.5	9.95
20	9.5	14.05	41	20.0	9.51
21	10.0	14.48			

6. Para avaliar a demanda de supervisores em indústrias, foram registradas as seguintes informações em 27 estabelecimentos industriais: $Y = n^o$ de supervisores e $X = n^o$ de trabalhadores supervisionados.

	X	Y		X	Y
1	294	30	15	615	100
2	247	32	16	999	109
3	267	37	17	1022	114
4	358	44	18	1015	117
5	423	47	19	700	106
6	311	49	20	850	128
7	450	56	21	980	130
8	534	62	22	1025	160
9	438	68	23	1021	97
10	697	78	24	1200	180
11	688	80	25	1250	112
12	630	84	26	1500	210
13	709	88	27	1650	135
14	627	97			

- (a) Ajuste e avalie a adequação do modelos:
1. Y em X
 2. $(Y)^{1/2}$ em X
 3. Y em X usando MQP com pesos $w_i = 1/x_i^2$
 4. $\log(Y)$ em X e X^2 (centrar X e X^2 na média se necessário)
- (b) Com base nos resultados obtidos, qual o melhor modelo em sua opinião?