2ª Lista de Exercícios - SME0820 Análise de Regressão - 21/03/2012

Exercício 1. Para calibrar um termopar, um engenheiro mecânico escolheu 16 níveis de temperatura entre 100^{o}C e 400^{o}C. A temperatura real (medida por um termômetro de precisão conhecida) e a leitura observada em um termopar de precisão desconhecida estão apresentadas na Tabela 1.

TABELA 1. Temperatura real (X, em ^{o}C) e temperatura observada (Y, em ^{o}C).

X	Y	X	Y
100	88,8	260	245,1
120	108,7	280	257,7
140	129,8	300	277
160	146,2	320	298,1
180	161,6	340	318,8
200	179,9	360	334,6
220	202,4	380	355,2
240	224,5	400	377

(a) Ajuste um modelo de regressão linear simples do tipo

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i,\ i = 1, \ldots, n,$$

assumindo que $\mathrm{E}(\epsilon_i) = 0$ e $\mathrm{Var}(\epsilon_i) = \sigma^2$ e ϵ_i é não correlacionado com ϵ_j para $i \neq j$; $i, j = 1, \ldots, n$ e verifique a significância da regressão. Obtenha uma estimativa não viesada para σ^2.

(b) Obtenha um intervalo de confiança para a temperatura medida, se a temperatura real for de 200^{o}C.

(c) Um modelo sem intercepto faria sentido?

Cálculos fornecidos:
$\sum_{i=1}^{n} X_i = 4000$; $\sum_{i=1}^{n} Y_i = 3705,4$; $\sum_{i=1}^{n} X_i Y_i = 1055962$; $\sum_{i=1}^{n} X_i^2 = 1136000$;
$\sum_{i=1}^{n} Y_i^2 = 981730,74$;

Exercício 2. A nota final da turma de Análise de Regressão é composta por uma prova e um trabalho, na qual ambos tem valor mínimo zero e máximo dez. O professor resolveu estudar a relação entre a nota na prova e no trabalho por meio de um modelo de regressão linear simples, com a suposição usual de normalidade e independência dos erros. Para isso, utilizou uma amostra de 50 alunos, cujos resultados são apresentados na Tabela 2:

TABELA 2. Notas dos 50 alunos na prova e no trabalho.

Prova (Y)	**Trabalho (X)**				
0	0,0	0,4	0,3	1,0	0,8
1	0,6	0,8	1,0	0,7	1,6
2	1,5	1,8	1,5	2,2	2,9
3	1,3	2,2	2,8	2,5	3,5
4	2,9	3,6	3,3	4,1	4,8
5	4,2	2,7	5,0	5,3	5,8
6	5,5	5,8	6,1	6,3	6,6
7	5,9	6,4	6,6	6,9	7,3
8	7,1	7,7	7,9	8,1	8,4
9	8,5	8,4	8,8	9,0	9,1

(a) Encontre as distribuições de $\hat{\beta}_0$ e $\hat{\beta}_1$.

(b) Apresente as estatísticas de teste e o critério de rejeição para avaliar se $\beta_0 = \beta_0^\star$ e se $\beta_1 = \beta_1^\star$.

(c) Estime β_0 e β_1 utilizando algum software estatístico. Interprete os resultados.

(d) Construa um intervalo de confiança com nível 95% para a nota média de prova para um aluno que obteve nota 5,0 no trabalho. Construa um intervalo de predição com nível 95% para a nota de prova de um aluno que obteve nota 6,0 no trabalho. Interprete os resultados.

Exercício 3. Deseja-se explicar o tempo de entrega de um certo produto (em unidades de tempo), utilizando o número de itens solicitados como variável explicativa. Para isso, foram coletadas 25 observações (X_i, Y_i) com $i = 1, \ldots, 25$ (Tabela 3). Um analista ajustou um modelo de regressão linear simples usual. Entretanto, o dono da empresa alegou que o modelo era inválido, já que se nenhum item fosse solicitado o tempo de entrega seria zero, ou seja, que o modelo correto seria um modelo sem intercepto (Montgomery, Peck, and Vining, 2001).

TABELA 3. Tempo de entrega (Y) e número de itens (X).

Y	X	Y	X	Y	X	Y	X	Y	X
16.68	7	17.83	7	24.00	9	52.32	26	17.90	10
11.50	3	79.24	30	29.00	10	18.75	9	19.75	6
12.03	3	21.50	5	15.35	6	19.83	8	8.00	2
14.88	4	40.33	16	19.00	7	10.75	4	35.10	17
13.75	6	21.00	10	9.50	3	13.50	4	18.11	7

(a) Ajuste um modelo **sem intercepto** aos dados de tempo de entrega e avalie a significância da regressão $(\alpha = 0,05)$.

(b) Ajuste um modelo **com intercepto** apresentando os cálculos intermediários e avalie a significância da regressão $(\alpha = 0,05)$. Avalie se $\beta_0 = 0$ nesse modelo ao nível de significância de $\alpha = 0,05$. Sua conclusão mudaria para $\alpha = 0,01$?

(c) Escolha e interprete o melhor modelo para indicar ao empresário, justificando sua escolha.

Cálculos fornecidos:

$\sum_{i=1}^{n} X_i = 219$; $\sum_{i=1}^{n} Y_i = 559,60$; $\sum_{i=1}^{n} X_i Y_i = 7375,44$; $\sum_{i=1}^{n} X_i^2 = 3055$; $\sum_{i=1}^{n} Y_i^2 = 18310,63$.

Exercício 4. Considere o coeficiente de correlação amostral entre duas variáveis U e W

$$r_{UW} = \frac{\sum_{i=1}^{n}(U_i - \overline{U})(W_i - \overline{W})}{\sqrt{\sum_{i=1}^{n}(U_i - \overline{U})^2}\sqrt{\sum_{i=1}^{n}(W_i - \overline{W})^2}}.$$

Considerando o modelo de regressão linear simples em que Y é a variável resposta e X é a variável explicativa, mostre que $r_{XY}^2 = R^2$, em que R^2 é o coeficiente de determinação do modelo.

Exercício 5. Considere o coeficiente de correlação amostral r definido no Exercício 4 e o modelo de regressão linear simples $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i,\ i = 1, \ldots, n$ com $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$.

(a) Mostre que $\widehat{Y}_i - \overline{Y} = \widehat{\beta}_1(X_i - \overline{X})$, em que $\widehat{\beta}_1$ é o estimador de mínimos quadrados de β_1.

(b) Mostre que $r_{XY}^2 = R^2$ no modelo de regressão linear simples.

(c) Mostre que $r_{Y\widehat{Y}} = r_{XY}$.