

Lista de Exercícios de Estatística

Professor: Rodrigo Lambert

Questão 1. *Sejam X e Y duas amostras de tamanho n dadas por:*

$$X = x_1, x_2, \dots, x_n$$

$$Y = y_1, y_2, \dots, y_n$$

Se para duas constantes α e β , e para todo $i = 1, 2, \dots, n$ escrevemos

$$y_i = \alpha + \beta x_i + \epsilon_i,$$

onde ϵ_i é a distância (erro) entre o ponto amostral (x_i, y_i) e o ponto da reta $y_i = \alpha + \beta x_i$ dado por $(x_i, \alpha + \beta x_i)$. Use o método dos mínimos quadrados para provar que

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}; \quad \hat{\beta} = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}$$

Questão 2. *Considerando o coeficiente de correlação de Pearson dado por*

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Dê uma interpretação geométrica para ρ como sendo o cosseno do ângulo entre os vetores $X = (x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_n - \bar{x})$ e $Y = (y_1 - \bar{y}, y_2 - \bar{y}, \dots, y_n - \bar{y})$.

Questão 3. *Em uma cidade no litoral de Portugal, foram medidos os seguintes dados sobre ninhos de cegonha encontrados (em milhares de unidades), e taxa de natalidade bruta (por mil habitantes) durante um período de 6 anos:*

Ano	2001	2002	2003	2004	2005	2006
Ninhos (x_i)	7	8	8	10	9	11
Tr. natal. (y_i)	10	11	11	11	13	14

(a) Use o método dos mínimos quadrados para traçar uma reta que melhor se ajuste a esses dados.

(b) Calcule o coeficiente de correlação de Pearson entre o número de ninhos de cegonha e a taxa de natalidade na região. Qual a sua conclusão sobre o resultado obtido?

Questão 4. (Retirado e adaptado de [1]) Uma amostra retirada ao acaso de alunos de uma escola produziu as seguintes notas em duas componentes curriculares:

Matemática (x_i)	56	50	72	67	31	50	65	40	80	61
Português (y_i)	60	50	67	75	44	56	72	48	76	62

(a) Calcule o coeficiente de correlação de Pearson entre x e y .

(b) Use o método dos mínimos quadrados para traçar uma reta que explique a relação entre as notas.

Questão 5. (Retirado e adaptado de [1]) Seja a seguinte tabela de uma amostra de alturas de pais e seus respectivos filhos.

Mãe	165	169	166	158	172	160	167	155	169	166
Filha	168	170	168	160	172	164	165	158	172	164

(a) Diga quem é a variável independente (x_i), e quem é a variável dependente (y_i).

(b) Calcule o coeficiente de correlação de Pearson entre x e y .

(c) Determine e faça um esboço de uma reta de regressão linear simples para esses dados. Dica: use o método dos mínimos quadrados.

(d) Com base nesses dados, qual seria a altura esperada de uma filha cuja mãe mede 171 cm? E qual seria a altura esperada de uma filha cuja mãe tivesse 166 cm? Discuta o último resultado quando comparado ao conjunto de dados.

Questão 6. (Retirado de [2]) Um pesquisador deseja verificar se um instrumento para medir a concentração de determinada substância no sangue está bem calibrado. Para isto, ele tomou 15 amostras de concentrações conhecidas (X) e determinou a respectiva concentração através do instrumento (Y), obtendo:

<i>X</i>	2,0	2,0	2,0	4,0	4,0	4,0	6,0	6,0	6,0	8,0	8,0	8,0	10,0	10,0	10,0
<i>Y</i>	2,1	1,8	1,9	4,5	4,2	4,0	6,2	6,0	6,5	8,2	7,8	7,7	9,6	10,0	10,1

(a) Construa o diagrama de dispersão para esses dados (gráfico dos pontos no plano cartesiano).

(b) Trace no gráfico a reta com 45° de inclinação passando pela origem. Como essa reta pode ser útil na avaliação do instrumento?

(c) Calcule o coeficiente de correlação entre as variáveis *X* e *Y*.

(d) Obtenha a reta de regressão da variável *Y* em função de *X*.

Questão 7. Por resultado obtido em sala, sabemos que o estimador de mínimos quadrados de $\beta = (\beta_0, \beta_1, \dots, \beta_k)$ na notação matricial no modelo de regressão linear múltipla é dado por

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} .$$

Mostre que o modelo de regressão linear simples que leva a $(\hat{\alpha}, \hat{\beta})$ é obtido como um caso particular da expressão acima.

Questão 8. (Retirado e adaptado de [3]) Uma empresa que vende por correio componentes de computadores pessoais, software e hardware possui um depósito geral para a distribuição dos produtos. Atualmente, a administração se encontra examinando o processo de distribuição deste depósito e está interessado em estudar os fatores que afetam os custos de distribuição do depósito. Atualmente um pequeno cargo de manipulação se adiciona ao pedido, independentemente da quantidade pela que se fizeram. Foram coletados dados correspondentes de 3 meses e respeito aos custos de distribuição de depósito, as vendas e número de pedidos. A continuação apresenta os resultados:

<i>Mês</i>	1	2	3
<i>X</i> ₁	386	446	512
<i>X</i> ₂	4015	3806	5309
<i>Y</i>	52,95	71,66	85,56

Onde: *Y* = Custo de distribuição (em milhares de dólares); *X*₁ = Vendas (em milhares de dólares) e *X*₂ = Número de pedidos.

(a) Ajuste os dados a um modelo de regressão com duas variáveis regressoras e interprete as estimativas dos parâmetros do modelo.

(b) Estime o custo de distribuição do depósito mensal da empresa quando as vendas são 400.000 dólares e o número de pedidos é de 4.500.

(c) Supondo que as duas variáveis explicativas (independentes) contribuem no modelo, qual delas é mais importante? Por quê?

Referências

- [1] *Exercícios de regressão linear*. Disponível online em: http://lodi.est.ips.pt/probest/diversos/exercicios/exercpe0910_regressao_linear.pdf
- [2] V.H. Lachos. *Lista de exercícios: Estatística para experimentalistas*. Disponível online em: <https://www.ime.unicamp.br/~hlachos/ExerciciosRegre.doc>
- [3] V.H. Lachos. *Segunda Lista de Exercícios - Regressão linear Multipla. I-2008s*. Disponível online em: https://www.ime.unicamp.br/~hlachos/ME175_lis2.pdf