

Exercício 1 : O Pesquisador, Parte I

Um pesquisador deseja verificar se um instrumento para medir concentração de ácido láctico no sangue está bem calibrado. Para isto ele tomou 20 amostras de concentrações conhecidas e determinou a respectiva concentração através do instrumento. Como uma análise de regressão poderia auxiliar o pesquisador? Modele o problema acima, especificando as variáveis independentes e dependente e as hipóteses de interesse.

Exercício 2 : Modelo de Segundo Grau, Parte I

Encontre os estimadores de mínimos quadrados para o modelo de 2º grau $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon$.

Exercício 3 : Modelo de Segundo Grau, Parte II

Sejam

Tabela 1: Tabela dos dados para o Exercício 3.

X	0	$3/2$	2	$5/2$	3	$7/2$
Y	$1/10$	$3/10$	$9/10$	$11/5$	4	6

Item (a) Construa o modelo $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ e determine o coeficiente de explicação.

Item (b) Ajuste uma parábola para os dados e escreva a ANOVA.

Item (c) Determine o coeficiente de explicação para o modelo do item (b) e compare com o obtido no item (a). Justifique a diferença obtida entre os dois valores.

Exercício 4 : Introdução ao TikZ

No modelo $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ com $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ e $\epsilon_i \perp \epsilon_j$, $i \neq j$, suponha que $\beta_0 = 210$, $\beta_1 = 4$ e $\sigma = 5$.

Item (a) Qual a distribuição de y dado $x \in \{10, 20, 40\}$?

Item (b) Construa um desenho que esquematize este modelo.

Exercício 5 : Umidade $\times$ Densidade

Acredita-se que a umidade de um produto influencia a densidade final do produto. Num experimento, a umidade foi controlada e a densidade final foi medida resultando nos seguintes dados (codificados):

Tabela 2: Tabela dos dados para o Exercício 5.

X (umidade)	4,7	5	5,2	5,2	5,9	4,7	5,9	5,2	5,2	5,3	5,9	5,6	5
Y (densidade)	3	3	4	5	10	2	9	3	3	7	6	6	4

Item (a) Se adotarmos o modelo $y = \beta_0 + \beta_1 x + \epsilon$, qual o significado prático de β_1 ?

Item (b) Ajuste o modelo $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ e construa a ANOVA.

Item (c) Teste as hipóteses $\mathcal{H}_0 : \beta_1 = 0$ contra $\mathcal{H}_a : \beta_1 \neq 0$. Qual dos modelos você adotaria para estes dados?

Exercício 6 : Teste de hipóteses

Com base em 10 pares de valores das variáveis X e Y obteve-se a equação de regressão $\hat{y} = 20 - x$, com coeficiente de explicação $R^2 = 0,64$. Se a estimativa não viciada de σ^2 é 4, teste ao nível $\alpha = 0,05$ a hipótese $\mathcal{H}_0 : \beta_1 = 0$ contra a alternativa $\mathcal{H}_a : \beta_1 \neq 0$.

Exercício 7 : Conjuntos de retas

Dado um conjunto de pares de valores $\{(x_{ij}, Y_{ij})\}, i \in \{1, \dots, m\}$ e $j \in \{1, \dots, n_i\}$, ajusta-se um conjunto de retas paralelas como

$$\hat{y}_{ij} = a_i + b x_{ij}.$$

Mostre que as estimativas dos parâmetros, de acordo com o método dos mínimos quadrados, são dadas por:

$$\begin{aligned} \hat{b} &= \frac{\sum_{i=1}^m \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(y_{ij} - \bar{y}_i)}{\sum_{i=1}^m \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}, \\ \hat{a}_i &= \bar{y}_i - \hat{b} \bar{x}_i, \quad \forall i \in \{1, \dots, m\}, \\ \text{onde } \bar{x}_i &= \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \text{ e } \bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}, \quad \forall i \in \{1, \dots, m\}. \end{aligned}$$

Exercício 8 : Ponto médio

Prove que a reta de regressão $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ passa pelo ponto $(\bar{x}, \bar{y})$.

Exercício 9 : Resíduo

Prove que se $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ e $e_i = y_i - \hat{y}_i$, $i \in \{1, \dots, n\}$, então:

Item (a)

$$\sum_{i=1}^n e_i = 0.$$

Item (b)

$$\sum_{i=1}^n x_i e_i = 0.$$

Item (c)

$$\sum_{i=1}^n \hat{y}_i e_i = 0.$$

Item (d)

$$\sum_{i=1}^n \hat{y}_i = \sum_{i=1}^n y_i.$$

Exercício 10 : Modelo Inverso

Com base em n observações $\{(x_i, Y_i)\}_{i \in \{1, \dots, n\}}$, encontre os estimadores de mínimos quadrados para o modelo:

$$\mathbb{E}[y|x] = \beta_0 + \beta_1 \frac{1}{x}.$$

Calcule a variância de $\hat{\beta}_1$.

Exercício 11 : Modelo sem Intercepto

Adotou-se o modelo $\mathbb{E}[y|x] = \beta x$, na situação em que o verdadeiro modelo era $\mathbb{E}[y|x] = \beta_0 + \beta_1 x$. Prove que o estimador de mínimos quadrados de β é viciado para β_1 e determine seu vício.

Exercício 12 : Modelo com apenas Intercepto

Com base numa amostra de n observações da variável y , determine o estimador de mínimos quadrados de β_0 no modelo $y_i = \beta_0 + \epsilon_i$. A qual característica populacional corresponde β_0 neste modelo?

Exercício 13 : Uma Afirmação

O que você acha da afirmação: “Para que o método de mínimos quadrados seja válido, é necessário que a distribuição da variável y seja normal”?

Exercício 14 : Propriedades de F_{obs}

No modelo de regressão linear simples $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$, sejam $F_{\text{obs}} = \text{SSR}/\text{MSE}$ e R^2 o coeficiente de explicação do modelo.

Item (a) Prove que $F_{\text{obs}} = \frac{(n-2)R^2}{1-R^2}$.

Item (b) Analise o comportamento de F para os possíveis valores de R^2 .

Exercício 15 : O Pesquisador, Parte II

Um pesquisador desejava estimar o nível de altitude acima do mar através do ponto da ebulição da água. Ele sabia que a altitude poderia ser determinada através da pressão atmosférica, medida através de um barômetro. No seu experimento, realizado em 1850, ele registrou a relação entre pressão e o ponto de ebulição. Seu interesse foi motivado pela dificuldade do transporte dos frágeis barômetros da época. As medidas do ponto de ebulição lhe forneceriam uma previsão da pressão e posteriormente uma rápida maneira de prever altitudes. Os dados a seguir correspondem a uma amostra de 30 pares de medidas as variáveis TEMP - ponto de evulição (graus Fahrenheit) e PRES - pressão barométrica (em polegadas de mercúrio).

Tabela 3: Tabela dos dados para o Exercício 15.

Temp	Pres	Temp	Pres
210,8	29,211	189,5	18,869
210,2	28,559	188,8	18,356
208,4	27,972	188,5	18,507
202,5	24,697	185,7	17,267
200,6	23,726	186,0	17,221
200,1	23,369	185,6	17,062
199,5	23,030	184,1	16,959
197,0	21,892	184,6	16,881
196,4	21,928	184,1	16,817
196,3	21,654	183,2	16,385
195,6	21,605	182,4	16,235
193,4	20,480	181,9	16,106
193,6	20,212	181,9	15,928
191,4	19,758	181,0	15,919
191,1	19,490	180,6	15,376
190,6	19,386		

Item **(a)** Construa o diagrama de dispersão da variável Pres contra Temp. Uma reta se ajusta bem aos dados?

Item **(b)** Construa um diagrama de dispersão da variável $100 \cdot \log_{10} \text{Pres}$ contra Temp e compare com o do item **(a)** (como os valores de $\log \text{Pres}$ não variam muito, eles foram multiplicados por 100. Isto evita que trabalhem com valores muito baixos e com pouca variabilidade, sem comprometer os principais aspectos da análise). Este diagrama de dispersão está mais próximo de uma reta que o do item **(a)**?

Item **(c)** Ajuste o modelo de regressão linear simples $y = 100 \cdot \log_{10} \text{Pres}$ em $X = \text{Temp}$ e calcule as estatísticas relevantes (estimativas dos parâmetros, testes e tabela de Análise de Variância, R^2).

Exercício 16 : Modelo segmentado

Com base em n observações $\{(x_i, Y_i)\}_{i \in \{1, \dots, n\}}$:

Item **(a)** Encontre os estimadores de mínimos quadrados para o seguinte modelo

$$\mathbb{E}[y|x] = \beta_0 + \beta_1(x - c)_+,$$

onde $c \in \mathbb{R}$ é um valor fixo (não deve ser estimado), e a função $(x)_+$ é denominada função “positiva truncada”, tal que

$$(x)_+ = \begin{cases} x & \text{se } x \geq 0, \\ 0 & \text{se } x < 0. \end{cases}$$

Item **(b)** Considere o seguinte experimento: um roedor é injetado com uma substância química, e oito semanas após a injeção é iniciado um processo semanal onde ele é levado a se exercitar através do uso de uma barra rotatória, de forma que sua performance semanal é baseada no tempo máximo (em segundos) que o roedor permaneceu na barra rotatória. São feitas medições da semana 8 até a semana 21. É obtida a seguinte tabela

Tabela 4: Tabela dos dados para o Exercício 16.

Semana	Tempo
8	130
9	180
10	180
11	180
12	176
13	177
14	175
15	167
16	106
17	103
18	107
19	52
20	18
21	0

Construa o gráfico de dispersão para este conjunto de dados. É possível observar uma mudança na tendência linear dos dados? Se sim, dê um “palpite” de onde ocorre esta mudança.

Item **(c)** Construa uma estimativa para o modelo linear $\mathbb{E}[y|x] = \beta_0 + \beta_1 x$ a partir dos dados do item **(b)** usando as observações das semanas 8 até (e incluindo) a semana 15. Teste $\mathcal{H}_0 : \beta_1 = 0$ a nível de significância $\alpha = 0,05$.

Item **(d)** Refaça o item **(c)** utilizando apenas as observações da semana 16 até (e incluindo) a semana 21.

Item **(e)** Construa uma estimativa para os parâmetros β_0 e β_1 vistos no modelo do item **(a)** a partir de todas as observações dos dados do item **(b)**, fixando $c = 15,5$. Interprete os resultados.

Item **(f)** Reescreva o modelo estimado no item **(e)** como segue

$$\hat{y} = \begin{cases} \hat{\alpha}_0 & \text{se } x \leq 15,5 \\ \hat{\alpha}_1 + \hat{\alpha}_2 x & \text{se } x > 15,5 \end{cases},$$

em que $(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2)$ são combinações lineares de $(\hat{\beta}_0, \hat{\beta}_1)$.

Item **(g)** Teste no modelo estimado em **(e)** a hipótese $\mathcal{H}_0 : \beta_1 = 0$.

Exercício 17 : Dados padronizados

Sabe-se que para um conjunto de dados $\sum_{i=1}^n y_i = 0$, $\sum_{i=1}^n y_i^2 = 1$, $\sum_{i=1}^n x_i = 0$ e $\sum_{i=1}^n x_i^2 = 1$. Adotou-se o modelo $\mathbb{E}[y|x] = \beta_0 + \beta_1 x$ para realização da regressão. Mostre que, para este conjunto de dados, os estimadores de mínimos quadrados são $\hat{\beta}_0 = 0$ e $\hat{\beta}_1 = r$, onde r indica a correlação amostral entre X e Y .

Exercício 18 : Correlação

Supõe-se que a quantidade de vapor (em 1000 libras) gasto mensalmente em uma determinada usina está correlacionada à temperatura ambiente mensal na mesma usina (em graus Fahrenheit). A seguinte tabela contém os dados associados para um ano qualquer

Tabela 5: Tabela dos dados para o Exercício 18.

Mês	Temperatura	Vapor gasto
Janeiro	21	185,79
Fevereiro	24	214,47
Março	32	288,03
Abril	47	424,84
Mai	50	454,68
Junho	59	539,03
Julho	68	621,55
Agosto	74	675,06
Setembro	62	562,03
Outubro	50	452,93
Novembro	41	369,95
Dezembro	30	273,98

Item **(a)** Estime a correlação linear ρ .

Observação:

Nota-se que, para amostras suficientemente grandes, para testarmos a hipótese nula $\mathcal{H}_0 : \rho = \rho_0$ a nível de significância α podemos usar da seguinte estatística teste

$$z_{\text{obs}} = \sqrt{n-3} \{ \text{arctanh}(\hat{\rho}) - \text{arctanh}(\rho_0) \},$$

de forma que rejeitamos $\mathcal{H}_0$ se $z_c < |z_{\text{obs}}|$, onde z_c é calculado como

$$\mathbb{P}(Z \geq z_c) = \frac{\alpha}{2},$$

onde $Z \sim \mathcal{N}(0, 1)$.

Item **(b)** Supondo que a amostra dada é suficientemente grande, teste a hipótese $\mathcal{H}_0 : \rho = 0$ e a hipótese $\mathcal{H}_0 : \rho = 1/2$, a nível de significância $\alpha = 0,05$.

Observação:

Ainda no contexto de amostras suficientemente grandes, é possível construir um intervalo de confiança a nível γ para ρ como segue

$$\text{IC}(\rho, \gamma) = \left[\tanh \left(\text{arctanh}(\hat{\rho}) - \frac{|z_c|}{\sqrt{n-3}} \right); \tanh \left(\text{arctanh}(\hat{\rho}) + \frac{|z_c|}{\sqrt{n-3}} \right) \right]$$

onde, para $Z \sim \mathcal{N}(0, 1)$, z_c é obtido como segue

$$\mathbb{P}(Z \geq z_c) = \frac{1-\gamma}{2},$$

Item **(c)** Supondo que a amostra dada é suficientemente grande, calcule um intervalo de confiança a nível $\gamma = 99\%$ para ρ .