Exercícios Vários:

1. Uma empresa que repara computadores, pretende estudar a relação entre a duração de uma chamada telefónica e o número de componentes reparadas. Os dados encontram-se no ficheiro P027.dat.

a) Representa os dados graficamente.

b) Determine o coeficiente de correlação.

c) Estime a recta de regressão linear.

d) Utiliza essa equação para prever a duração de uma chamada na qual 4 componentes têm que ser reparados.

e) Determine um intervalo de confiança a 95% para o verdadeiro declive da recta de regressão.

f) Determine um intervalo de confiança a 95% para a ordenada da recta de regressão.

g) Teste a hipótese de o declive ser igual a zero, supondo que $\alpha = 0.05$.

Exercícios Vários:

Exercício 1:

Suponha que foi realizado um ensaio para avaliar o crescimento radicular de uma certa cultivar de uma espécie agrícola. Para o efeito, foi medido o comprimento (em mm) da raíz principal (Y), decorridos x dias. Obtiveram-se os seguintes resultados:

x	1	7	13	20	27	34	62
y	5	10	12	29	36	83	102

Utilize o programa R para responder às seguintes questões.

a) Introduza os dados.

b) Construa um diagrama de dispersão para visualizar a relação entre as variáveis.

c) Utilizando um modelo de regressão linear simples, exprima os comprimentos da raiz principal como função dos dias decorridos. Interprete os valores obtidos.

d) Obtenha estimativas das variâncias e dos desvios padrões associados às estimativas dos parâmetros do modelo de regressão linear.

e) Obtenha uma estimativa da variância dos erros.

f) Obtenha um intervalo de de confiança a 95% para os coeficientes de regressão.

g) Teste se a ordenada na origem é significativamente diferente de 0, ao nível de significância 1%.

h) Utilize um teste de hipóteses sobre o declive da recta de regressão para validar a seguinte afirmação: "não existe uma relação linear significativa entre os dias e o comprimento da raíz, para a referida cultivar".

i) Valide de novo a afirmação anterior mas agora utilizando um teste F.

j) Utilize um teste de hipótese para validar a seguinte afirmação: " por cada dia a mais, a raíz da cultivar cresce, em média, 2mm.

k) Comente a qualidade da recta obtida, calculando o coeficiente de correlação e interpretando o valor obtido.

l) Determine a soma dos quadrados totais a partir do cálculo da variância amostral de Y.

m) Indique o valor da soma dos quadrados dos resíduos.

n) Suponha agora que a relação entre as variáveis é dada pelo modelo de regressão: $Y^{-1} = \beta_0 + \beta_1 X + \varepsilon$. Estime o novo modelo de regressão.

Exercício 2:

Um conjunto de n=23 dados bidimensionais $\{(x_i, y_i)\}_{i=1}^{23}$ tem centro de gravidade
$(\overline{x}, \overline{y})$ = (12.5,-116.826087). Foi ajustada a recta de regressão de y sobre x. O resíduo associado ao ponto (9.50,-48.0) é e_i=3.93

(a) Qual é a equação da recta de regressão?

(b) Sabendo que a soma dos quadrados devidos à regressão é SQR = 124742.0703 e que a variância de y é s_y^2=6071.882798, calcule (justificando as suas respostas):

i. s_x^2

ii. cov_{xy}

iii. o coeficiente de determinação

iv. a soma dos quadrados dos resíduos, SQE

v. o coeficiente de correlação.

Fórmula de cálculo:

$$SQRE = SQT - SQM = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 - \hat{\beta}_1^2 \sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Exercício 3:

Os encargos diários com o consumo de gás propano (Y) de uma empresa dependem da temperatura ambiente (X). A tabela seguinte apresenta o valor desses encargos em função da temperatura exterior:

Temperatura (_C)	5	10	15	20	25
Encargos (euro)	20	17	13	11	9

(a) Ajuste um modelo de regressão linear simples aos dados.

(b) Diga como interpreta o valor de $\hat{\beta}_1$ obtido.

(c) Quantifique a qualidade do ajuste obtido e interprete.

(d) Determine um intervalo de confiança a 95% para os encargos médios com gás propano num dia em que a temperatura ambiente é de 17_C.

(e) Determine o respectivo coeficiente de correlação; com base no valor obtido, que pode concluir quanto ao grau de associação das duas variáveis?

(f) Determine um intervalo de confiança a 95% para o verdadeiro declive da recta de regressão.

(g) Determine um intervalo de confiança a 95% para a ordenada da recta de regressão.

Exercício 4:

Os dados em *diabetes.sav* representam os valores de PH (X) e de iões de hidrogénio (Y) na urina de 9 doentes diabéticos.

a) Estime um modelo de regressão linear.

b) Construa intervalos de confiança a 95% para cada um dos coeficientes de regressão.

c) Teste a hipótese de o declive ser igual a zero, supondo que $\alpha = 0.05$.

d) Determine os valores estimados da variável dependente.

e) Represente graficamente os valores observados e estimados da variável dependente.

f) Estime E(Y) para os doentes diabéticos de valor de PH na urina de 6.0. Determine o intervalo de confiança relativo ao número médio de iões de hidrogénio na urina desses doentes diabéticos

g) Supondo que um dado doente apresentava valor de PH na urina de 6.0, qual o valor de $\hat{y}$. Preveja, com um grau de confiança de 95% o número de iões de hidrogénio na urina desse doente.

h) Indique:

i) qual a a percentagem de variância de Y explicada pela recta de regressão.

ii) a tabela ANOVA associada à regressão estimada no exercício 1 e conclua se o modelo de regressão é significativo?

Exercício 5:

Considere X: a altura do atleta (em metros) e Y: a melhor marca em salto em altura (em metros). Para 20 atletas obteve-se:

$\sum x_i = 37.36 \qquad \sum y_i = 47.42 \qquad \sum x_i y_i = 88.618 \qquad \sum x_i^2 = 69.8978 \qquad \sum y_i^2 = 112.4638$

a) Estimar a recta de regressão de Y sobre X.

b) Qual a percentagem de variância de Y explicada pela recta de regressão?

c) Estimar a variância dos erros.

d) Testar $H_0 : \beta_1 = 0$ ao nível de significância de 5%.

e) Estimar E(Y) para os atletas que medem 2 metros.

f) Determine o intervalo de confiança relativo ao número médio da melhor marca em salto em altura dos atletas de 2 metros de altura, com um nível de confiança de 99%.

g) Estabeleça a tabela Anova associada a esta regressão.

Exercício 5:

Seja $\hat{y}_i = 3 - 5x_i$ e $R^2 = 60.84\%$ e n=50.

a) Determine o coeficiente de correlação r_{XY} .

b) Testar $H_0 : \beta_1 = 0$ ao nível de significância de 5%.