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.

Solução: $\boldsymbol{r}_{xy} = 0.9936$

c) Estime a recta de regressão linear.

Solução: $\hat{\boldsymbol{Min}} = 4.162 + 15.509\boldsymbol{Units}$

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

Solução: $\hat{\boldsymbol{Min}} = 66.20$

e) Mostre que $\boldsymbol{r}_{Min,Units} = \boldsymbol{r}_{Min,\hat{Min}} = 0.9936$

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

Solução: $(14.408;16.610)$

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

Solução: $(-3.148;11.471)$

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

Solução: p_value= 8.92e-13, Rejeitar Ho

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.

Solução: a) $\hat{y} = 153.574 - 21.632x$

b) i) 12.12 **ii)** -262.1479 **iii)** 0.934 **iv)** 8839.3513 **v)** -0.966

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.

Solução: $\hat{y} = 22.4 - 0.56x$

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

Solução: Aumentando um grau a temperatura ambiente, o valor esperado dos encargos diários com o consumo de gás propano diminui aproximadamente 0.56 euros.

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

Solução: $R^2 = 0.98$ ou seja, 98% da variação total dos encargos diários com o consumo de gás propano é explicada pelo modelo de regressão linear simples com a temperatura ambiente como variável explicativa.

(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.

Solução: (11.79984;13.96016)

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

Solução: $r_{xy} = -0.9899$ **;** Associação linear forte negativa

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

Solução: (-0.7069909;-0.4130091)

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

Solução: (19.9624317;24.8375683)

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:

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

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

Solução: $\hat{y} = 1.731246 + 0.34248x$

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

Solução: 41.39%

c) Estimar a variância dos erros.

Solução: $\hat{\sigma}^2 = 0.001008751$

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

Solução: ET=3.565283; pvalue= 0.002211517 Rejeitar H0

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

Solução: $\hat{E}(Y) = 2.4162$

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%.

Solução: (2.374374 2.458041)

g) Estabeleça a tabela Anova associada a esta regressão.

Solução: SSE= 0.01815752; MSE= 0.001008751; SST= 0.03098; MST= 0.001630526; SSR=MSR= 0.01282248; F= 12.71124; pvalue= 0.002211517

Exercício 6:

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

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

Solução: $\boldsymbol{r_{xy}} = -0.78$

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

Solução: ET=8.635627; pvalue= 2.453349e-11. Rejeitar H0

NOTA: Seria mais fácil usar a fórmula: $\boldsymbol{F} = \frac{\boldsymbol{n} - \boldsymbol{p} - 1}{\boldsymbol{p}} \frac{\boldsymbol{R}^2}{1 - \boldsymbol{R}^2}$

Fórmula de cálculo:

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