Associação linear. Coeficiente de correlação linear

1. Descrição

A tarefa permite estudar a relação entre o sinal do coeficiente de correlação e o declive da reta de mínimos quadrados, assim como associação linear entre as variáveis estatísticas e o coeficiente de correlação. Também se pretende com a presente tarefa analisar a reduzida resistência do coeficiente de correlação linear.

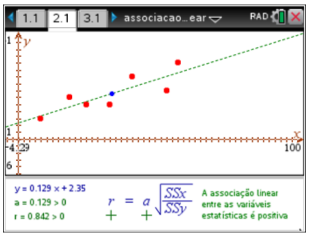
Ficheiro: associação_linear.tns

2. Metas Curriculares

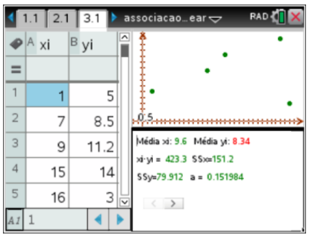
Estatística 11 – EST11

1.9 Designar, dado um número natural n e uma amostra de dados bivariados quantitativos $\underset{\sim}{(x,y)}$, por «coeficiente de correlação linear» o quociente $\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{SS_x SS_y}}$, representá-lo por «$r$» , reconhecer que $r = a\sqrt{\frac{SS_x}{SS_y}}$ onde a é o declive da reta de mínimos quadrados, justificar que r e a têm o mesmo sinal e saber que $|r|$ é sempre menor ou igual a 1, tomando o valor 1 unicamente nos casos em que todos os pontos $P_i(x_i, y_i)$, $1 \leq i \leq n$, estão alinhados e referir que a «associação linear entre as variáveis estatísticas» é positiva (respetivamente negativa) se $r > 0$ (respetivamente se $r < 0$) e que é tão mais «forte» quanto mais perto de 1 estiver $|r|$.

4. Guia de utilização e de exploração



A nuvem de pontos representada na página 2.1 pode ser alterada movendo cada um dos pontos, clicando e arrastando-os. A visibilidade da reta de mínimos quadrados correspondente à nuvem de pontos é controlada pelo seletor disponível.



Na página 3.1, não é possível mover os pontos. Apenas podem ser alterados através da tabela definida na folha de cálculo da mesma página. A página 3.2, contém em execução a aplicação calculadora para que o aluno possa efetuar os cálculos respeitantes à questão 2.

Exercício 1

Neste exercício pretende-se explorar a relação entre os sinal do coeficiente de correlação linear e o sinal do declive da reta de mínimos quadrado, assim como a intensidade da associação linear de acordo com o valor do coeficiente de correlação.

Autores: Jorge Santos e José Martins

Exercício 1.1

Pretende concluir-se que, na situação referida, o sinal do declive da reta de mínimos quadrados e o sinal do coeficiente de correlação são iguais.

Exercício 1.2

Na situação em que a reta de mínimos quadrados tem declive negativo, o sinal do coeficiente de correlação também é negativo. O sinal do declive da reta de mínimos quadrados é igual ao sinal do coeficiente de correlação, ou seja, a conclusão da alínea anterior mantém-se.

Exercício 1.3

*Pelas alíneas 1.1. e 1.2. podemos concluir que o coeficiente de correlação linear e o declive da reta de mínimos quadrados têm sinais **iguais**.*

Exercício 1.4

Sabe-se que $r = a\sqrt{\frac{SS_x}{SS_y}}$. Logo, o sinal do coeficiente de correlação depende do valor de a, declive da reta de mínimos quadrados e de $\sqrt{\frac{SS_x}{SS_y}}$. Como estamos perante um produto entre a e $\sqrt{\frac{SS_x}{SS_y}}$ cujo sinal é sempre positivo por definição de raiz quadrada, o sinal do coeficiente de correlação igual ao sinal do declive da reta de mínimos quadrados.

Exercício 1.5

Pretende concluir-se que quanto mais forte for a associação linear, ou seja, quanto mais próximos os pontos estiverem da reta de mínimos quadrados, mais perto $|r|$ está de 1.

Exercício 1.6

Para que $|r| = 1$, os pontos têm que estar alinhados segundo uma reta não horizontal e não vertical. O valor $|r|$ está próximo de zero quando a nuvem de pontos é dispersa ou existem outliers que proporcionam essa situação. No caso em que os pontos estão alinhados horizontalmente, o valor $|r|$ é igual a zero.

Exercício 1.7

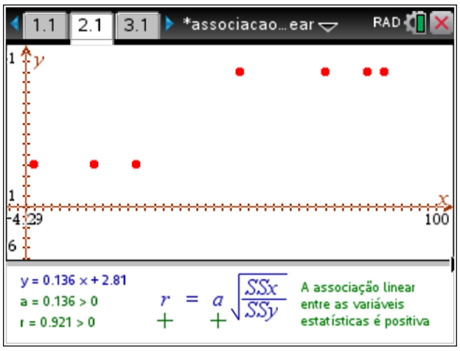
Sintetizado as alíneas 1.6 e 1.7 temos:

*$|r|$ é sempre **menor** ou igual a 1, tomando o valor 1 unicamente nos casos em que todos os pontos estão **alinhados**. A associação linear entre as variáveis é tão mais forte quanto mais **perto** de 1 estiver $|r|$.*

Autores: Jorge Santos e José Martins

Exercício 2

Neste exercício estuda-se a resistência do coeficiente de correlação e a importância de representar a nuvem de pontos antes de interpretar o coeficiente de correlação entre duas variáveis. Poderão ser exploradas situações em que o coeficiente de correlação é próximo de 1 ou -1 , no entanto, a representação gráfica não evidencia uma associação linear. É importante ter presente que o coeficiente de correlação mede o grau de associação linear e não outro tipo de associação.



Exercício 2.1

Para a resolução deste exercício, a reta de mínimos quadrados deve estar oculta. Para tal, utilize o seletor disponibilizado debaixo da representação gráfica. Apenas por observação, poder-se-á deduzir que a associação linear é forte, dado que todos os pontos estão todos patricamente alinhados exceto um deles.

Exercício 2.2

Deve-se determinar o pedido usando as fórmulas correspondentes e não obter os resultados diretos através da calculadora, para que seja percetível o uso da média nos cálculos, que é uma medida pouco resistente.

Exercício 2.2.1

$$\bar{x} = \frac{\sum_{i=1}^{6} x_i}{6} = 6{,}6 \qquad \bar{y} = \frac{\sum_{i=1}^{6} y_i}{6} = 3{,}03333 \qquad \sum_{i=1}^{6} x_i y_i = 123{,}3$$

$$a = \frac{\sum_{i=1}^{6} x_i y_i - 6\bar{x}\,\bar{y}}{SS_x} \approx \frac{123{,}3 - 6 \times 6{,}6 \times 3{,}0333}{90{,}28} \approx 0{,}0352$$

$$b = \bar{y} - a\bar{x} \approx 3{,}0333 - 0{,}0352 \times 6{,}6 \approx 2{,}801$$

A equação da reta de mínimos quadrados com aproximação à décima de milésima é

$$y = 0{,}0352x + 2{,}801$$

Autores: Jorge Santos e José Martins

Exercício 2.2.2

$$r \approx 0{,}0352\sqrt{\frac{90{,}28}{26{,}4723}} = 0{,}065$$

Exercício 2.3

O valor do coeficiente de correlação não confirma a intensidade da associação linear deduzida na questão 2.1.. Esperar-se-ia que o coeficiente de correlação fosse próximo de 1 uma vez que aparentemente, por observação, a associação linear das duas variáveis é forte.

Exercício 2.4

$$a = \frac{\sum_{i=1}^{5} x_i y_i - 5\bar{x}\,\bar{y}}{SS_x} \approx \frac{122 - 5 \times 5{,}32 \times 3{,}62}{41{,}128} \approx 0{,}6251$$

$r \approx 0{,}6251\sqrt{\frac{41{,}128}{16{,}148}} \approx 0{,}9976$, donde se conclui que a associação linear entre as variáveis é forte, pois o coeficiente de correlação é próximo de 1 e as variáveis evidenciam uma associação linear.

Exercício 2.5

O coeficiente de correlação linear não é uma medida resistente, pois é muito influênciada por *outliers*, valores que se distinguem dos restantes. A razão pela qual o coeficiente é pouco resistente deve-se à sua fórmula de cálculo. Para o seu cálculo é necessário a média que é uma medida não resistente, ou seja, é muito influenciada por valores muito pequenos ou muito elevados relativamente aos restantes.

Autores: Jorge Santos e José Martins