

Modelos Lineares e Aplicações

Exercícios Vários:

1. A partir de uma amostra de 22 observações de uma variável aleatória bidimensional obteve-se:

$$\sum_{i=1}^n x_i = 124 \quad \sum_{i=1}^n x_i^2 = 772 \quad \sum_{i=1}^n y_i = 1248 \quad \sum_{i=1}^n y_i^2 = 78624 \quad \sum_{i=1}^n x_i y_i = 7504$$

- a) Indica uma estimativa da correlação de X e Y
- b) Indica uma estimativa da variância de $Z=Y-4X$
- c) Indica uma estimativa da covariância entre X e Z.

2. Mostre que:

a) $\sum_{i=1}^n (x_i - \bar{x}) = 0$

b) $(n-1)\text{cov}(X, Y) = \sum_{i=1}^n (x_i - \bar{x})y_i$

3. O Quadro seguinte apresenta a taxa de desemprego para os quinze países com um valor mais elevado do Índice de Desenvolvimento Humano(IDH). Pretende-se aferir se existe associação entre o desenvolvimento humano e a taxa de desemprego. O que conclui?

Países	TD
Canadá	9.3
Noruega	4.1
EUA	5.0
Japão	3.5
Bélgica	9.0
Suécia	8.0
Austrália	8.5
Holanda	5.6
Islândia	3.8
UK	7.1
França	12.4
Suíça	4.2
Finlândia	14.5
Alemanha	9.8
Dinamarca	5.4

4. No SPSS foi obtida a seguinte tabela de correlações:

Correlations		Colesterol HDL	Colesterol HDL Após Água	Colesterol HDL Após Chá
Colesterol HDL	Pearson Correlation	1	,757**	,774**
	Sig. (2-tailed)		,000	,000
	N	37	37	37
Colesterol HDL Após Água	Pearson Correlation	,757**	1	,935**
	Sig. (2-tailed)	,000		,000
	N	37	37	37
Colesterol HDL Após Chá	Pearson Correlation	,774**	,935**	1
	Sig. (2-tailed)	,000	,000	
	N	37	37	37

** . Correlation is significant at the 0.01 level (2-tailed).

Diga, justificando, se existe uma associação significativa entre os vários pares de variáveis em causa, ao nível de significância 1%. Qual o par que apresenta uma associação mais forte?

5. Considere os dados apresentados em *exercicio7.sav*.

- Calcular o coeficiente de correlação r_{xy} .
- Represente as observações num diagrama de dispersão
- Indique uma transformação adequada aos dados.
- Calcule o novo coeficiente de correlação r_{xy} .

6. Considere a amostra que se encontra no package datasets, com o nome de faithful.

- Quantas variáveis temos nesta amostra? De que tipo são?
- Faça uma representação gráfica dos dados.
- Calcule o coeficiente de correlação entre a duração da erupção e o tempo decorrido até à erupção seguinte.
- Teste, ao nível de 5%, a hipótese de não existir correlação.

7. O programa R tem vários conjuntos de dados disponíveis. Um desses conjuntos de dados designa-se anscombe e pode ser visto apenas escrevendo o nome do objecto:

```
> anscombe
```

Considere os dados anscombe disponíveis no R. Determine, e comente os valores obtidos para:

(a) As médias de cada variável x_i e y_i ($i = 1 : 4$).

(b) As variâncias de cada variável x_i e y_i ($i = 1 : 4$).

(c) Os coeficientes de correlação

Após comentar os resultados obtidos, construa os repetitivos diagramas de dispersão.

Comente esses gráficos, à luz dos valores anteriormente obtidos.

Coeficiente de correlação por “grupos”

8. Considere os dados apresentados em *crabs.xls*.

a) Calcular o coeficiente de correlação entre as variáveis.

b) Selecionar as observações pertencentes à espécie “B”.

Os comandos necessários são:

Data → Select Cases → If condition is satisfied → If... **Especie="B"** → Continue →

Unselect cases are → Filtered

c) Representar estas observações num diagrama de dispersão, representando os caranguejos fêmeas e machos de diferentes cores.

Os comandos necessários são:

Graphs → Scatter → Matrix → Define

Matrix Variables: **CL, CT, CLMC, CMC, CC**

Set Marks by : **Sexo**

Options..... Titles.....

d) Calcular o novo coeficiente de correlação entre as variáveis.

e) Calcular o novo coeficiente de correlação entre as variáveis mas separadamente para os caranguejos fêmeas e para os caranguejos machos.

Os comandos necessários são:

Data → Split File → Compare Groups → Groups Based on: **Sexo**

Analyse → Correlate → Bivariate

Variables: **CL, CT, CLMC, CMC, CC**

SPSS

Em SPSS, os comandos necessários a obter o diagrama de dispersão são:

Graphs → Legacy dialogs → Scatter/dot → Simple Scatter → Define
YAxis XAxis Options..... Titles.....

Clica-se em cima do gráfico e abre-se um menu designado SPSS Chart
Editor....

Em SPSS, os comandos necessários a obter o coeficiente de correlação são:

Analyse → Correlate → Bivariate
Variables:
Options.....

Nota: podem-se seleccionar mais do que duas variáveis, caso em que o SPSS fornece uma tabela de correlações para todas as combinações de pares de variáveis. O SPSS fornece também o p-value dos testes ao significado dos coeficientes, para cada par de variáveis.

R

`cov(x,y)`

`cor(x,y)`

`cor.test(x,y)`

`cor(..., method= "spearman")`

`cor(..., method= "kendall")`