

**6ª Lista de Exercícios - SME0803 Análise Exploratória de Dados Data: 10/05/2012**  
**Entregar exercícios 2 e 10 no dia 18/05/2012 às 19h. Pode ser feito em duplas.**

Para os exercícios de 1 a 6, considere  $n$  pares de observações  $(x_i, y_i)$ ,  $i = 1, \dots, n$  e a covariância e correlação amostrais entre  $x$  e  $y$  como

$$\text{Cov}(x, y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad \text{e}$$

$$\text{Cor}(x, y) = \frac{\text{Cov}(x, y)}{\sqrt{\text{Var}(x)}\sqrt{\text{Var}(y)}} = \frac{\text{Cov}(x, y)}{s_x s_y}$$

**Exercício 1.** Mostre que  $\text{Cor}(x, x) = 1$ .

**Exercício 2.** Mostre que se  $a_1, a_2 \in \mathbb{R}$  e  $b_1, b_2 > 0$ , então  $\text{Cor}(a_1 + b_1x, a_2 + b_2y) = \text{Cor}(x, y)$  (invariância sob transformação linear).

**Exercício 3.** Prove que se  $y = a + bx$  e  $b > 0$ , então  $\text{Cor}(x, y) = 1$ .

**Exercício 4.** Prove que se  $y = a + bx$  e  $b < 0$ , então  $\text{Cor}(x, y) = -1$ .

**Exercício 5.** Mostre que se  $x$  ou  $y$  é simétrica em torno de  $x_0$  ou  $y_0$ , então  $\text{Cor}(x, y) = 0$ .

**Exercício 6.** Mostre que se  $u_i = \frac{(x_i - x_0)}{h}$  e  $v_i = \frac{(y_i - y_0)}{k}$  então  $r_{uv} = r_{xy}$ .

**Exercício 7.** Observe os dados de índice de inflação entre os anos 1967 e 1979.

Inflação (y)	Ano (x)
128	1967
192	1969
277	1971
373	1973
613	1975
1236	1977
2639	1979

- (a) Faça o gráfico de dispersão de  $x$  e  $y$ . Interprete o resultado.
- (b) Ajuste o modelo  $y = \alpha + \beta x + \epsilon$  aos dados. Encontre as estimativas de mínimos quadrados de  $\alpha$  e  $\beta$ . Interprete-as.
- (c) Quanto deve ser a inflação em 1981? Justifique.
- (d) Você teria alguma restrição em adotar o modelo linear nesse caso? Obtenha e interprete  $R^2$ .

**Exercício 8.** Considere o conjunto de dados `Ilocos` (`data(Ilocos)`) do pacote `ineq` (`library(ineq)`) em R. Os dados foram coletados em domicílios nas Filipinas. O comando `?Ilocos` fornece diversas informações sobre as variáveis. Considere as variáveis `income`, `sex`, `family.size`, `urbanity` e `province`. É de interesse verificar se a primeira variável (`income`) está relacionada às demais.

- (a) Apresente as variáveis utilizando medidas resumo (todas que você considerar pertinentes) e recursos gráficos.
- (b) Procure relações entre a variável `family.size` e outras que você julgar relevantes.

**Exercício 9.** Um estudo foi conduzido para verificar se as pessoas declaram seus pesos corretamente. No experimento realizado, 15 pessoas foram selecionadas ao acaso, a cada uma delas perguntou-se os pesos, que depois foram aferidos em balanças devidamente calibradas. Os resultados encontram-se na tabela a seguir

Indivíduo	Declarado (y)	Real (x)	Indivíduo	Declarado (y)	Real (x)
1	82	83	9	45	44
2	58	57	10	81	82
3	69	73	11	78	76
4	70	76	12	65	67
5	54	55	13	56	54
6	62	60	14	63	60
7	92	98	15	70	71
8	75	74			

- Construa o diagrama de dispersão de  $x$  e  $y$ .
- Trace em um gráfico a reta  $y = x$ . Como essa reta pode ser útil na análise desejada?
- Calcule o coeficiente de correlação entre  $x$  e  $y$ .
- Com base nos itens anteriores, conclua sobre a eficiência com que as pessoas declaram seus próprios pesos.

**Exercício 10.** Em um estudo sobre plantação de trigo, deseja-se avaliar a relação entre a disponibilidade de nitrogênio no solo e a quantidade de nitrogênio na planta. Os dados coletados contêm informação sobre a disponibilidade de nitrogênio no solo onde a planta estava plantada ( $x$ ) e também a quantidade de nitrogênio existente na planta ( $y$ ). Os dados estão apresentados na tabela a seguir:

Nitrogênio no solo (ppm*)	Nitrogênio na planta (ppm*)
0,42	0,13
0,45	0,15
0,50	0,16
0,55	0,17
0,68	0,18
0,69	0,18
0,70	0,19
0,73	0,20
0,80	0,20
0,90	0,21
0,92	0,22
0,94	0,23

\* partes por milhão

- Construa o diagrama de dispersão e interprete-o.
- Calcule e interprete o coeficiente de correlação  $r$  de Pearson entre  $x$  e  $y$ .
- Ajuste uma reta de regressão para a relação entre as variáveis  $y$ : quantidade de nitrogênio na planta (dependente) e  $x$ : nitrogênio no solo (independente). Interprete o valor do coeficiente angular obtido.
- Considerando a reta ajustada dada no item (c), estime a quantidade de nitrogênio que se espera encontrar em uma planta encontrada em um solo com 0,96ppm.
- Obtenha o coeficiente de determinação do modelo,  $R^2$ . Interprete os resultados.