

## 8ª Lista de Exercícios - SME-803 Análise Exploratória de Dados

Profª Cibele Russo

1. Sejam  $X$  e  $Y$  variáveis quantitativas. Considere o coeficiente de correlação linear de Pearson, denotado por  $r$ . Apresente exemplos de gráficos de dispersão com as seguintes características:

(a)  $r > 0$  e associação linear moderada, (b)  $r < 0$  e associação linear forte, (c)  $r > 0$  e associação forte porém não linear, (d)  $r \cong 0$ .

2. Para avaliar a associação entre atividades físicas e depressão, um estudo prospectivo foi realizado com 10.000 mulheres, das quais 20% não praticavam mais que 30 minutos de atividade física semanal no início do estudo. Um total de 1757 mulheres apresentou ao menos 5 episódios depressivos ao longo de 3 anos, das quais 568 pertenciam ao grupo sedentário. Construa uma tabela de frequências, avalie os resultados e conclua se a prática da atividade física pode reduzir o risco de depressão.

3. Afirma-se que o consumo de tomate cozido duas vezes ao dia pode reduzir o risco da recorrência de câncer de próstata em no mínimo 40%. De 98 pacientes que não tiveram recorrência de câncer de próstata em 2 anos, 70 haviam consumido o vegetal na forma indicada e de 79 pacientes que tiveram recorrência de câncer de próstata no mesmo período, 20 haviam feito uso do tomate cozido. A afirmação é verdadeira? Justifique.

4. Os dados abaixo correspondem a índices de publicação dos docentes e pesquisadores do Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa em Engenharia (COPPE/UFRJ) (adaptado). Utilize a análise de agrupamentos para agrupar áreas similares em relação à pesquisa.

Tabela 1: Dados de publicações

	CI	CN	LI	LN	RI	RN	EX
Biomédica	16,2	47	0	0,2	7	4,4	2
Civil	79,2	63,6	0	0,6	22,6	6,6	4,6
Elétrica	76,4	62,4	0,4	0	27	6	3,6
Mecânica	28,2	41,8	0,2	0	18	3,4	2,8
Nuclear	26,4	51,6	0	0	20	1,8	1
Produção	20,6	59,6	0,2	2	9,6	14,6	15,4
Química	32,6	76,4	0	0,6	47,2	6,6	3
Sistemas	48,2	58,4	0,2	0,2	24,4	6,2	18,2
Transportes	13,2	29	0	0,4	2,4	2,4	4,4

Onde:

CI congresso internacional

CN congresso nacional

LI livro internacional

LN livro nacional

RI revista internacional

RN revista nacional

EX outras publicações

Matriz de distâncias (reduzida):

	(Bio)	(Civ)	(Ele)	(Mec)	(Nuc)	(Pro)	(Qui)	(Sis)
(Civ)	3.19							
(Ele)	4.16	2.94						
(Mec)	1.77	3.23	2.97					
(Nuc)	1.36	2.86	3.73	1.64				
(Pro)	4.78	4.58	5.25	5.01	5.46			
(Qui)	3.86	2.81	3.87	3.82	3.18	5		
(Sis)	3.61	2.97	2.99	3.02	3.49	3.97	3.65	
(Tra)	1.53	4.12	4.96	2.24	2.33	5.21	5.01	4.15

5. Os dados representados na Figura 1 são de estatísticas criminais norte-americanas contidas no conjunto de dados *USArrests* (disponível no R). Interprete os gráficos e correlações entre as variáveis. Que variável você escolheria para explicar a quantidade de assaltos em um modelo de regressão linear simples? Ajuste e obtenha o coeficiente de determinação desse modelo. Interprete as estimativas dos parâmetros da regressão. Dados:  $\text{Var}(\text{Assassinato}) = 18.97$ ,  $\text{Var}(\text{Assalto}) = 6945.16$ ,  $\text{Var}(\text{Estupro}) = 209.52$ ,  $\text{Var}(\text{População Urbana}) = 87.73$ ,  $\overline{\text{Assassinato}} = 7.79$ ,  $\overline{\text{Assalto}} = 170.76$ ,  $\overline{\text{Estupro}} = 65.54$ ,  $\overline{\text{População Urbana}} = 21.23$ .

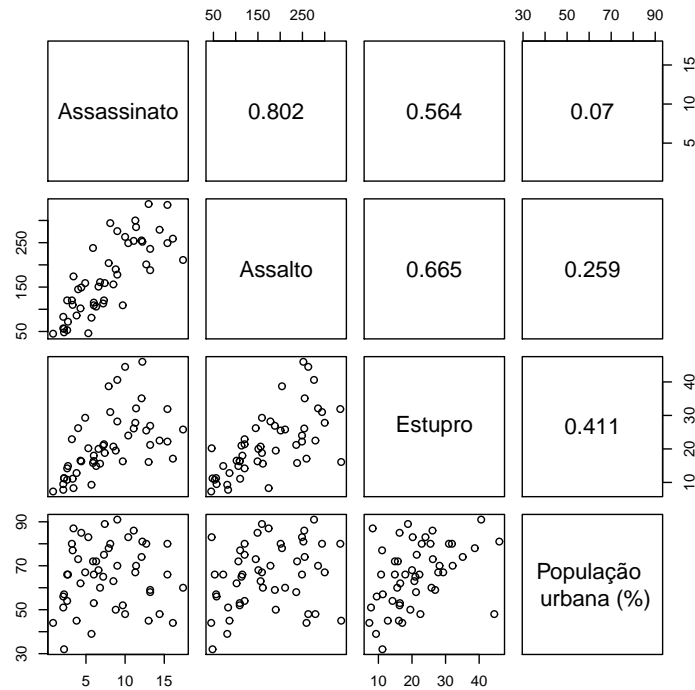


Figura 1: Dados de estatísticas criminais norte-americanas

6. A variável  $X$  é qualitativa e a variável  $Y$  é quantitativa.

- Qual(is) gráfico(s) você utilizaria para avaliar a associação entre as duas variáveis?
- Em relação ao item (a), apresente um exemplo de gráfico ilustrando uma situação em que há associação entre as duas variáveis.
- Em relação ao item (a), apresente um exemplo de gráfico ilustrando uma situação em que não há associação entre as duas variáveis.

7. Uma operadora de Seguro Saúde, com o objetivo de verificar a existência de associação entre faixa etária do cliente e tipo de exame solicitado, fez um levantamento de 1110 guias de pedido de exames, obtendo os resultados da Tabela 2. Verifique se há associação entre as variáveis faixa etária e exame solicitado.

Tabela 2: Dados de faixa etária e exames solicitados.

Exame	Faixa Etária				Total
	Até 18 anos	18 a 30 anos	30 a 45 anos	45 anos ou mais	
Sangue	60	100	150	350	660
Endoscopia	10	24	100	80	214
Eletrocardiograma	6	6	20	60	92
Outros	24	30	40	50	144
Total	100	160	310	540	1110