

Exercício 1

Uma indústria, desejando melhorar o desempenho de seus 30 funcionários da área administrativa, montou 3 cursos experimentais de inglês utilizando 3 metodologias distintas (*MA*, *MB* e *MC*). Os conceitos, *A*, *B*, *C*, obtidos pelos funcionários em cada curso são mostrados a seguir:

Método *A* (*MA*): *B A B B A A C B A A*

Método *B* (*MB*): *C B C A C C C C C*

Método *C* (*MC*): *C A B B C C C B C B C*

- Identifique as variáveis em estudo. Classifique-as.
- Construa uma tabela de contingência para as variáveis ‘Conceito’ e ‘Metodologia’.
- Podemos dizer que há indícios da existência de associação entre conceito e metodologia? Justifique.

(a) Método do curso: qualitativa nominal

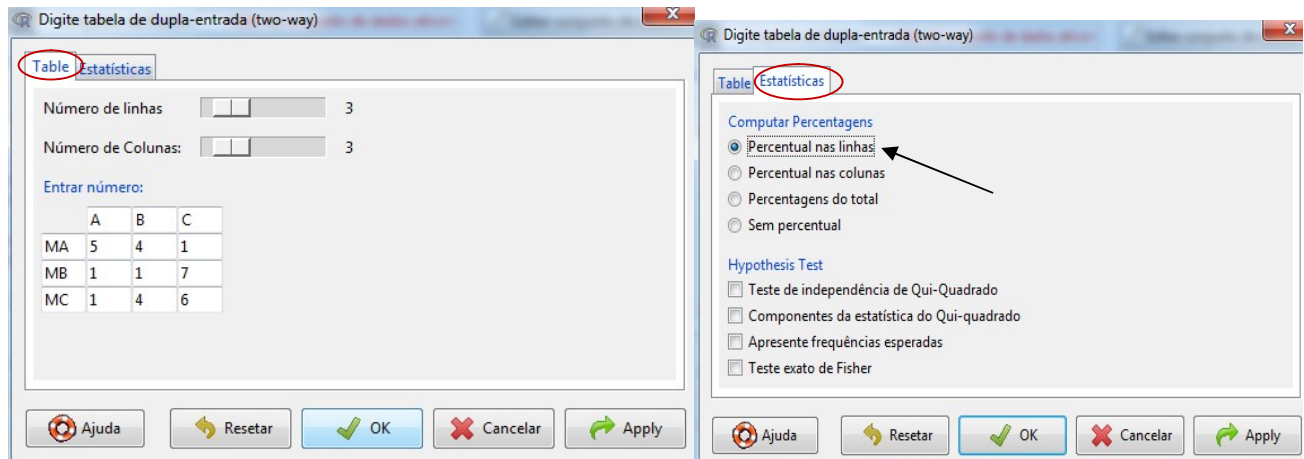
Conceito: qualitativa ordinal

(b) Tabela de Contingência entre Método e Conceito (% em linhas)

Método	Conceito			Total
	<i>A</i>	<i>B</i>	<i>C</i>	
<i>MA</i>	5 (50%)	4 (40%)	1 (10%)	10 (100%)
<i>MB</i>	1 (11,1%)	1 (11,1%)	7 (77,8%)	9 (100%)
<i>MC</i>	1 (9,1%)	4 (36,4%)	6 (54,5%)	11 (100%)
Total	7 (23,3%)	9 (30%)	14 (46,7%)	30 (100%)

(c) Obter % em linha a mão (na tabela acima) e via *Rcmdr* a partir da tabela de dupla entrada

Estatísticas -> Tabelas de Contingência -> Digite e analise tabela de dupla entrada



Saída Rcmdr

```
> .Table # Counts
```

```
  A B C
MA 5 4 1
MB 1 1 7
MC 1 4 6
```

```
> rowPercents(.Table) # Row Percentages
```

```
  A    B    C Total Count
MA 50.0 40.0 10.0    100    10
MB 11.1 11.1 77.8    100     9
MC  9.1 36.4 54.5    100    11
```

Sim, há indícios de associação entre o método do curso e o conceito alcançado: o método B apresenta uma proporção alta de funcionários com grau C (77,8%), enquanto que o Método A apresenta uma baixa proporção de funcionários com esse grau C (10%); o método C apresenta uma proporção baixa (9,1%) de funcionários com grau A; a maioria teve grau B ou C. O Método que aparenta ter melhor desempenho é o Método A.

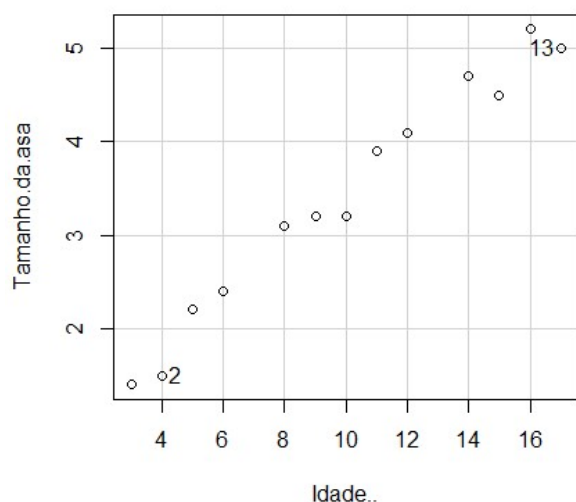
Exercício 2

Os dados abaixo correspondem aos valores do comprimento (em *cm*) da asa e a idade (em dias) de 13 pardais, com o objetivo de investigar possível associação entre essas medidas das aves, após a eclosão.

<i>X</i> : Idade (em dias)	<i>Y</i> : Tamanho da asa (em <i>cm</i>)
3	1,4
4	1,5
5	2,2
6	2,4
8	3,1
9	3,2
10	3,2
11	3,9
12	4,1
14	4,7
15	4,5
16	5,2
17	5,0

- Construa (a mão) o diagrama de dispersão entre as variáveis idade (*X*) e tamanho da asa (*Y*). O que o gráfico sugere?
- Calcule o coeficiente de correlação linear de Pearson entre *X* e *Y*.
- Ajuste uma reta de regressão para descrever a relação entre as variáveis *X* e *Y*. Qual é a interpretação do coeficiente angular estimado *b*?
- De quanto é a variação média no tamanho da asa entre pardais com diferença de idade de 4 dias no período inicial de vida após a eclosão?
- Considerando a reta estimada em (c), estime o tamanho médio da asa correspondente a pardais com 1 semana de vida e com 2 semanas de vida.

(a)



(b)

$$r = \frac{\sum_{i=1}^n X_i Y_i - n \bar{X} \bar{Y}}{(n-1) S_X S_Y} \quad \text{e} \quad S_X^2 = \frac{\sum_{i=1}^n X_i^2 - n \bar{X}^2}{n-1}$$

X (Idade em anos)	Y (Tamanho da Asa em cm)	X ²	Y ²	X.Y
3	1,4	9	1,96	4,2
4	1,5	16	2,25	6
5	2,2	25	4,84	11
6	2,4	36	5,76	14,4
8	3,1	64	9,61	24,8
9	3,2	81	10,24	28,8
10	3,2	100	10,24	32
11	3,9	121	15,21	42,9
12	4,1	144	16,81	49,2
14	4,7	196	22,09	65,8
15	4,5	225	20,25	67,5
16	5,2	256	27,04	83,2
17	5	289	25	85
130	44,4	1562	171,3	514,8

$$\bar{x} = \frac{130}{13} = 10 \quad \bar{y} = \frac{44,4}{13} = 3,42$$

$$S_X^2 = \frac{1562 - 13 \times 10^2}{13 - 1} = \frac{262}{12} = 21,8333 \Rightarrow S_X = \sqrt{21,8333} \cong 4,673$$

$$S_Y^2 = \frac{171,3 - 13 \times 3,42^2}{13 - 1} = \frac{19,2468}{12} = 1,604 \Rightarrow S_Y = \sqrt{1,604} \cong 1,266$$

Assim,
$$r = \frac{\sum_{i=1}^n X_i Y_i - n \bar{X} \bar{Y}}{(n-1) S_X S_Y} = \frac{514,8 - 13 \times 10 \times 3,421}{12 \times 4,673 \times 1,266} = \frac{70,2}{70,992} \cong 0,9888$$

(c) Reta ajustada: $\hat{Y} = a + bX$, sendo

$$b = \frac{\sum_{i=1}^n X_i Y_i - n \bar{X} \bar{Y}}{(n-1) S_X^2} \quad \text{e} \quad a = \bar{Y} - b \bar{X}$$

Assim, substituindo temos:

$$b = \frac{514,8 - 13 \times 10 \times 3,42}{12 \times 21,8333} = \frac{70,2}{261,9996} \cong 0,268$$

$$a = 3,42 - 0,268 \times 10 = 3,42 - 2,68 = 0,740$$

Logo, a reta ajustada é: $\hat{Y} = 0,740 + 0,268X$

Interpretação de b : A cada aumento de 1 dia na idade do pardal, o comprimento da asa aumenta, em média, 0,268 cm.

(d) Pardais com diferença de 4 dias, tem diferença no comprimento da asa de $4 \times b = 4 \times 0,268 = 1,072$ cm.

(e) Para $x = 7$, $\hat{Y} = 0,740 + 0,268 \times 7 = 2,616$ cm

Para $x = 14$, $\hat{Y} = 0,740 + 0,268 \times 14 = 4,492$ cm

Exercício 3

Um estudo foi realizado para comparar vinho *Pinot Noir* produzido em três diferentes regiões. Amostras de cada região foram tomadas e avaliadas por um painel de juízes em uma série de características. Quanto maior a pontuação, melhor o vinho. Os dados estão no arquivo **Wine.xls**.

- Construa o gráfico de dispersão para Aroma e Flavor (Sabor). Identifique no mesmo gráfico as amostras quanto à região. Comente sobre a relação entre estas variáveis em cada região.
- Suponha que a qualidade dos vinhos é classificada como “1” se 'Quality' < 11 , como “2” se 'Quality' estiver entre 11 e 12; e como “3” se for superior a 12. Construa uma tabela de contingência com a distribuição das amostras de acordo com a qualidade e a região. Há indicação de associação entre as variáveis? Justifique.

Leitura dos dados:

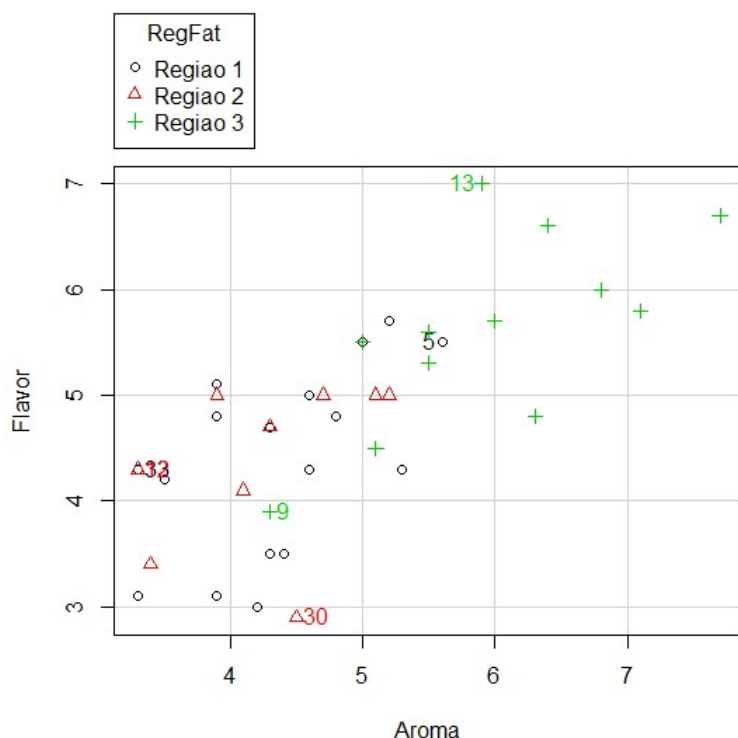
Dados -> Importar arquivos de dados -> do arquivo Excel

(a) Primeiro converter (Region) Região para Fator:

Dados -> Modificação de variáveis ...-> Converter variável numérica para fator-> Escolher Region e dar nome para a nova variável (exemplo: RegFat)->Ok-> Dar nomes para as regiões (Região 1, Região 2, Região 3)-> Ok

Fazer Diagrama, por grupo, com a nova Região;

Gráficos -> Diagrama de dispersão -> X= Aroma; Y=Flavor -> Gráficos por grupos -> RegFat



O gráfico sugere a existência de relação próxima de linear e crescente entre as notas de Aroma e Sabor nas três regiões; quanto maior a pontuação para Aroma, há tendência de maior pontuação para Sabor. Observa-se ainda que na Região 3 (“+”) a pontuação nessas características apresenta valores maiores quando comparada às das outras regiões

(b) Construir a nova variável chamada, por ex., QualidFat, com a categorização proposta para ‘Quality’

Dados -> Modificacao de var. no conj. de dados -> recodificar variáveis

Novo nome de variável = QualidFat

Definição p/ codificação

0:10.99 = 1

11:12 = 2

12.01:99 = 3

Construir tabela de frequência entre RegFat (linha) e QualidFat (coluna)

Estatísticas -> Tabelas de Contingência-> Tabela de dupla entrada- > RegFat e QualidFat (em Estatística, escolher Percentual nas linhas)

Saída do Rcmdr:

Frequency table:

	QualidFat		
RegiFat	1	2	3
Região 1	3	5	9
Região 2	6	2	1
Região 3	0	0	11

Row percentages:

	QualidFat				
RegFat	1	2	3	Total	Count
Região 1	17.6	29.4	52.9	99.9	17
Região 2	66.7	22.2	11.1	100.0	9
Região 3	0.0	0.0	100.0	100.0	11

Os resultados sugerem existência de associação entre Região e Qualidade do Vinho: 100% das amostras da Região 3 obtiveram pontuação na faixa mais alta de qualidade (3); A Região 2 apresentou uma grande proporção (66,7%) de amostras de vinho com pontuação baixa de qualidade (1) (e apenas 11% das amostras tiveram com pontuação alta). A Região 3 parece ser a que produz vinho de melhor qualidade, seguido da Região 1.

Exercício 4

No artigo intitulado “Estimativas do Valor Energético a partir de Características Químicas e Bromatológicas dos Alimentos” (Rev. Bras. Zootec., 30(6):1837-1856, 2001) estudou-se a disponibilidade de energia dos alimentos, considerando os nutrientes digestíveis totais (*NDT*) e também as análises químicas e metabólicas das dietas. Nele se apresentam os seguintes gráficos:

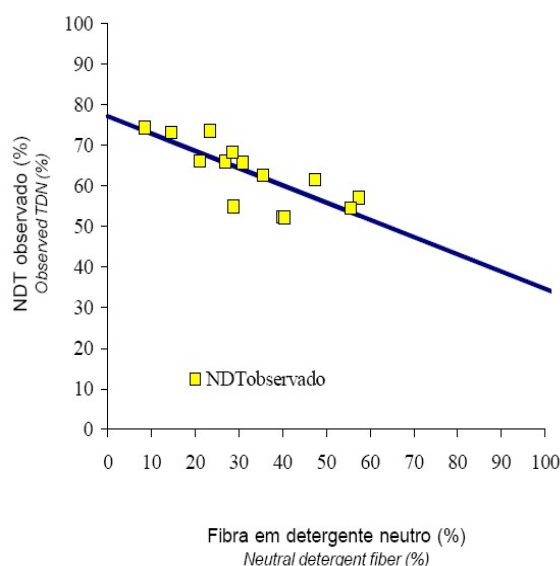


Figura 3 - Relação entre o NDT e a fibra em detergente ácido (FDA) nas dietas totais. A linha contínua representa a reta obtida pela equação de regressão $NDT = 77,13 - 0,4250FDA$ ($r^2 = 0,59$; $P < 0,01$).

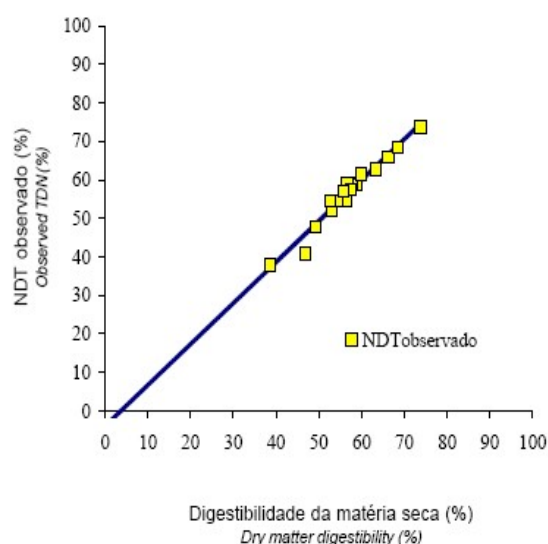


Figura 4 - Relação entre o NDT e a digestibilidade da matéria seca (DMS) nas dietas totais. A linha contínua representa a reta obtida pela equação de regressão $NDT = -3,84 + 1,064DMS$ ($r^2 = 0,96$; $P < 0,01$).

- (a) Para qual variável observou-se uma correlação linear maior com o *NDT*? Justifique.
- (b) Calcule o valor esperado de *NDT* se o valor de porcentagem da digestibilidade da matéria seca for 47.
- (a) DMS (digestibilidade da matéria seca), pq observou-se uma correlação maior entre NDT e DMS em relação à correlação entre NDT e FDA.
- FDN: $r^2 = 0,59 \Rightarrow r = -\sqrt{0,59} = -0,7681$
- DMS: $r^2 = 0,96 \Rightarrow r = \sqrt{0,96} = 0,9798$
- (b) $NDT^{\wedge} = -3,84 + 1,064 \times 47 = 46,168$, ou seja, aproximadamente 46%.