

# Lista 3 - MAE0217

Guilherme Navarro - N<sup>o</sup>USP: 8943160 E Victor Ribeiro Baião Decanini N<sup>o</sup> USP: 9790502

## Exercício 1

a)

Estado civil	Escolaridade			
	Ensino Fundamental	Ensino Médio	Ensino Superior	Total
Casado	5	12	3	20
Solteiro	7	6	3	16
Total	12	18	6	36

b)

Fazendo uma nova tabela com as frequências relativas, podemos notar que o grau de escolaridade não tem muita dependência com o estado civil, no sentido que os casados tem um grau de escolaridade muito próximo dos solteiros, como podemos ver na tabela a seguir:

Estado civil	Escolaridade			
	Ensino Fundamental	Ensino Médio	Ensino Superior	Total
Casado	13.89%	33.33%	8.33%	56.56%
Solteiro	19.44%	16.67%	8.33%	44.44%
Total	33.33%	50%	16.67%	100%

c)

A estatística de qui-quadrado de Pearson:  $X^2 = \sum_{j=1}^n \frac{(o_j - e_j)^2}{e_j} \Rightarrow X^2 = 1.9125$

Coefficiente de Contingência:  $C = \sqrt{\frac{X^2}{X^2 + n}} \Rightarrow C = 0.2245999$

Coefficiente de Tschuprov:  $T = \sqrt{\frac{X^2/n}{\sqrt{(r-1)(c-1)}}} \Rightarrow T = 0.193817$

Analisando os coeficientes calculados, podemos notar que há uma certa independência nos dados, como já era esperado apenas com uma análise empírica dos dados.

d)

**Medidas resumo de de pessoas do interior:**

Número de pessoas: 12

Média interior =  $\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i \Rightarrow \bar{X} = 11.55$

Mediana = 10.645

Desvio Padrão =  $DP(X) = \sqrt{Var(x)} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \Rightarrow DP(X) = 5.2960551$

Máximo = 23.3

Mínimo = 4

### Medidas resumo de de pessoas da capital:

Número de pessoas: 11

Média capital =  $\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i \Rightarrow \bar{X} = 11.4554545$

Mediana = 9.77

Desvio Padrão =  $DP(X) = \sqrt{Var(x)} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \Rightarrow DP(X) = 5.4766529$

Máximo = 19.4

Mínimo = 4.56

### Medidas resumo de de pessoas de ‘outra’:

Número de pessoas: 13

Média “outra” =  $\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i \Rightarrow \bar{X} = 10.4453846$

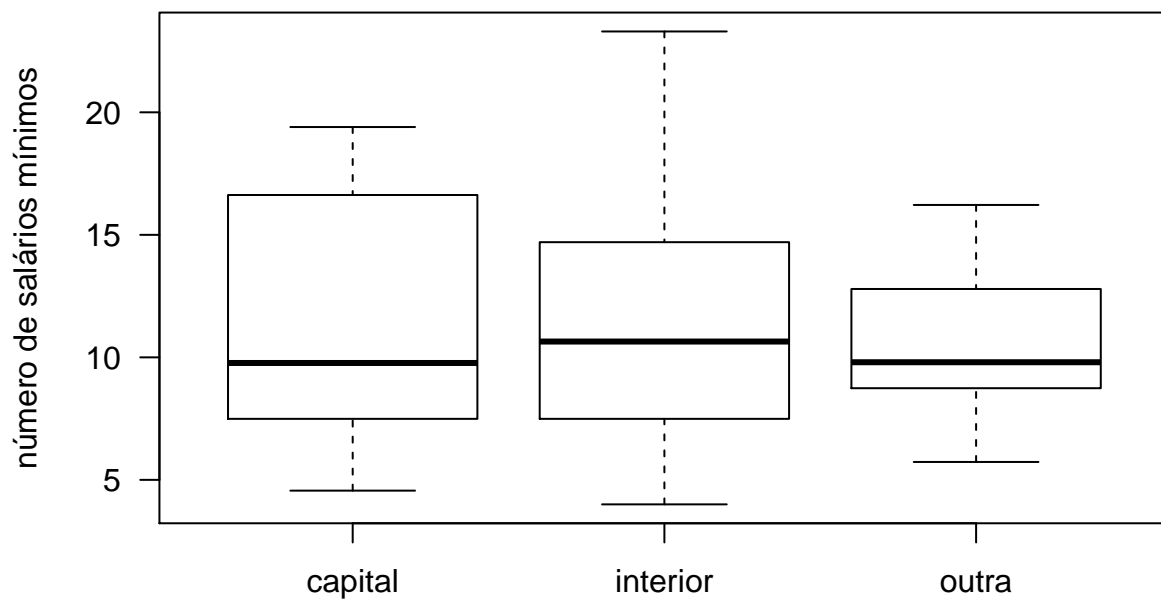
Mediana = 9.8

Desvio Padrão =  $DP(X) = \sqrt{Var(x)} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \Rightarrow DP(X) = 3.1454534$

Máximo = 16.22

Mínimo = 5.73

### Gráficos Box plots



Fazendo uma análise das medidas resumo dos funcionários do interior, capital e “outra” em conjunto o gráfico Box-Plot, chegamos a conclusão que a procedência não tem dependência com os salários.

## Exercício 2

Dose de radiação gama (cGy)	Frequência de células com múltiplos micronúcleos	Total de células examinadas
0	1	2373
20	6	2662
50	25	1991
100	47	2047
200	82	2611
300	207	2442
400	254	2838
500	285	1746

a)

$$\text{razão de chances}(0-20): \frac{\pi_1}{\pi_0} = \frac{6/2662}{1/2373} = 5.3486101$$

$$\text{razão de chances}(0-50): \frac{\pi_1}{\pi_0} = \frac{25/1991}{1/2373} = 29.7965846$$

$$\text{razão de chances}(0-100): \frac{\pi_1}{\pi_0} = \frac{47/2047}{1/2373} = 54.4851001$$

$$\text{razão de chances}(0-200): \frac{\pi_1}{\pi_0} = \frac{82/2611}{1/2373} = 74.5254692$$

$$\text{razão de chances}(0-300): \frac{\pi_1}{\pi_0} = \frac{207/2442}{1/2373} = 201.1511057$$

$$\text{razão de chances}(0-400): \frac{\pi_1}{\pi_0} = \frac{254/2838}{1/2373} = 212.3826638$$

$$\text{razão de chances}(0-500): \frac{\pi_1}{\pi_0} = \frac{285/1746}{1/2373} = 387.3453608$$

b)

$$\text{razão de chances}(0-20): \frac{\pi_1*(1-\pi_0)}{\pi_0*(1-\pi_1)} = \frac{(6/2662)*(2372/2373)}{(1/2373)*(2656/2662)} = 5.3584337$$

$$\text{razão de chances}(0-50): \frac{\pi_1*(1-\pi_0)}{\pi_0*(1-\pi_1)} = \frac{(25/1991)*(2372/2373)}{(1/2373)*(1966/1991)} = 30.162767$$

$$\text{razão de chances}(0-100): \frac{\pi_1*(1-\pi_0)}{\pi_0*(1-\pi_1)} = \frac{(47/2047)*(2372/2373)}{(1/2373)*(2000/2047)} = 55.742$$

$$\text{razão de chances}(0-200): \frac{\pi_1*(1-\pi_0)}{\pi_0*(1-\pi_1)} = \frac{(82/2611)*(2372/2373)}{(1/2373)*(2529/2611)} = 76.9094504$$

$$\text{razão de chances}(0-300): \frac{\pi_1*(1-\pi_0)}{\pi_0*(1-\pi_1)} = \frac{(207/2442)*(2372/2373)}{(1/2373)*(2235/2442)} = 219.6885906$$

$$\text{razão de chances}(0-400): \frac{\pi_1*(1-\pi_0)}{\pi_0*(1-\pi_1)} = \frac{(254/2838)*(2372/2373)}{(1/2373)*(2584/2838)} = 233.1609907$$

$$\text{razão de chances(0-500): } \frac{\pi_1 * (1 - \pi_0)}{\pi_0 * (1 - \pi_1)} = \frac{(285/1746) * (2372/2373)}{(1/2373) * (1461/1746)} = 462.7104723$$

c)

Fazendo uma análise dos riscos relativos e as razões de chances concluímos que, conforme aumenta a dose de radiação, aumenta também a quantidade de células com múltiplos micronúcleos. Além disso, as razões são bem próximas, com a razão de chances sendo sempre maior.

## Exercício 3

a)

Tabela

	HIV +(D)	HIV -(ND)	Total
Teste +(T+)	50	25	75
Teste -(T-)	50	75	125
Total	100	100	200

b)

$$\text{Sensibilidade : } \hat{S} = P(T+|D) = \frac{50}{100} = 0.5$$

$$\text{Especificidade : } \hat{E} = P(T-|ND) = \frac{75}{100} = 0.75$$

c)

$$\text{Valor Preditivo Positivo (VPP) : } V\hat{P}P = P(D|T+) = \frac{50}{75}$$

$$\text{Valor Preditivo Negativo (VPN) : } V\hat{P}N = P(ND|T-) = \frac{75}{125} = 0.6$$

d)

$$\text{Acurácia do teste (AC) : } \hat{A}C = P[(ND \cap T-) \cup (D \cap T+)] = \frac{50+75}{200} = 0.625$$

## Exercício 4

Doses da Substância A	Dose da Substância B		
	Baixa	Média	Alta
Baixa	10.4	8.9	4.8
Baixa	12.8	9.1	4.5
Baixa	14.6	8.5	4.4
Baixa	10.5	9.0	4.6
Média	5.8	8.9	9.1
Média	5.2	9.1	9.3
Média	5.5	8.7	8.7
Média	5.3	9.0	9.4

a)

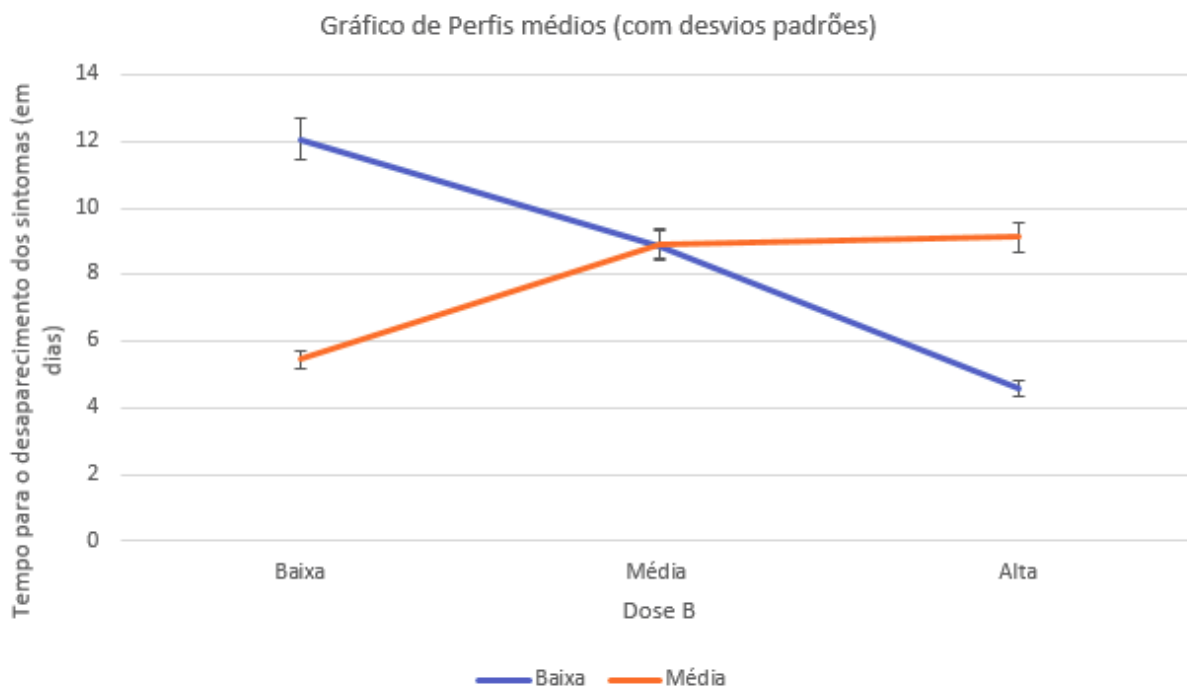
Doses da Substância A	Doses da Substância B	n	Média	Desvio Padrão
Baixa	Alta	4	4.57	0.148
	Média	4	8.87	0.228
	Baixa	4	12.75	1.74
Média	Alta	4	9.12	0.268
	Média	4	8.92	0.148
	Baixa	4	5.45	0.23

Logo, fazendo uma análise descritiva com base na tabela acima concluimos que a dose da substância A tem que ser BAIXA e a dose da substância B tem que ser ALTA, observando o intervalo da média mais ou menos o desvio padrão.

b)

Para fazer a comparação de 6 tratamentos quanto ao tempo para desaparecimento dos sintomas o modelo a ser utilizado é o ANOVA, sendo os seus fatores: Doses da substância A, e dose da substância B, com 2 e 3 níveis respectivamente.

c)



Levando em conta médias e desvios padrões, vemos que: Há uma interação entre os fatores pois eles se cruzam, mesmo que tenham pequenos desvios padrões.

## Exercício 5

a)

Coefficiente de correlação de Pearson = 0.7530873

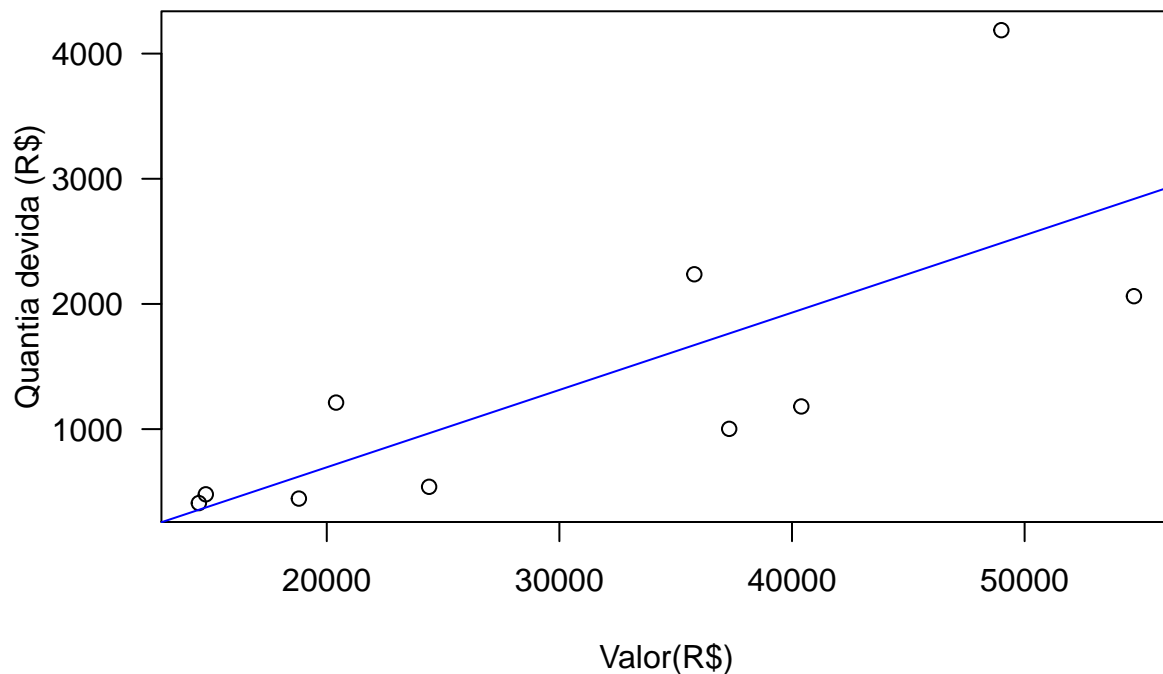
Coefficiente de correlação de Spearman = 0.7939394

O coeficiente de correlação de Pearson nos diz que entre as variáveis “Valor do imóvel” e “quantia devida” tem uma correlação forte e positiva, ou seja, conforme aumenta o valor do imóvel maior é sua dívida. E o coeficiente de Spearman confirma que a correlação é mesmo linear, pois seu valor foi muito próximo da de Pearson.

b)

Analisando o coeficiente de correlação linear e o gráfico de dispersão de forma intuitiva vemos que o modelo de regressão linear simples melhor se adequa a esta situação, logo:

$$Y = \hat{\beta}X + \hat{\alpha} \text{ com } \hat{\alpha} = -540.40 \text{ e } \hat{\beta} = 0.06178 \Rightarrow Y = 0.06178X - 540.40$$



c)

Usando a equação para prever a taxa média devida cujo valor é igual a R\$42.000,00 é:

$$Y = 0.06178 * 42400 - 540.40 \Rightarrow Y = 2079.7. \text{ Ou seja, o valor é aproximadamente: R\$2080,00}$$

d)

o coeficiente  $R^2 = 0.567$ , ou seja, o modelo ajustado com o valor da propriedade explica 57% da variação total dos valores da dívida com a prefeitura.