

PROJETO FINAL

Rhaíra Keitianne da Silva

19 de Dezembro de 2019

Análise bidimensional - variáveis qualitativas

Foram escolhidas as variáveis B e C para variáveis qualitativas.

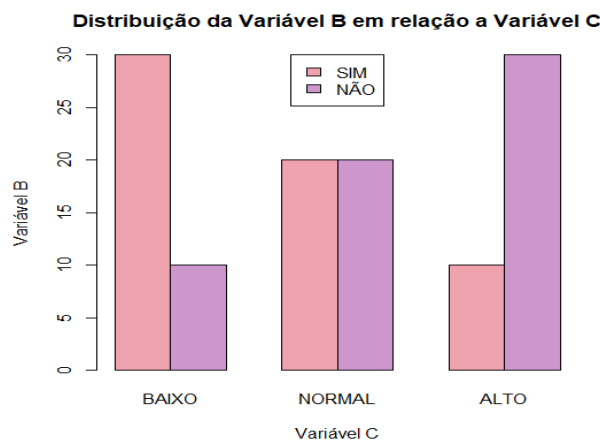


Tabela de contingência:

Mostram quantos dados se encaixam em cada categoria e servem para avaliar a semelhança das distribuições de cada caso com a distribuição marginal.

Frequências Absolutas			
	BAIXO	NORMAL	ALTO
SIM	30	20	10
NÃO	10	20	30

Frequências Relativas (linhas)			
	BAIXO	NORMAL	ALTO
SIM	0.50	0.33	0.17
NÃO	0.17	0.33	0.50

Qui-quadrado:

Esta medida avalia a relação entre duas variáveis qualitativas, através do desvio entre as frequências esperadas e frequências observadas, e valores altos desta medida indicam associação entre as variáveis. Esta distribuição é muito utilizada principalmente para realizar o Teste de Qui-quadrado. Tal teste tem por objetivo verificar se a frequência observada de uma variável é significativamente diferente da frequência absoluta esperada e nele há duas hipóteses: H0 (existe independência entre as variáveis) e H1 (existe independência entre as variáveis). O p-valor é a probabilidade de que a estatística do teste tenha valor extremo em relação ao valor observado quando a hipótese H0 é verdadeira. Para as variáveis B e C, o qui-quadrado obtido foi 20 e o p-valor obtido foi de 4.54e-05.

Variáveis Quantitativas

Foram escolhidas as variáveis 1, 2 e 3 para as variáveis quantitativas.

O coeficiente de correlação mede o grau de associação entre duas variáveis. Para as variáveis 1 e 2, o coeficiente de correlação foi igual a -0.07836227 e para as variáveis 1 e 3, foi igual a -0.08652531. Como os coeficientes foram muito próximos de zero, pode-se concluir que há uma associação linear fraca, sendo também possível chegar a esta conclusão analisando os gráficos de dispersão para as respectivas variáveis. Além disso, os valores foram negativos, indicando

Gráfico de dispersão para as variáveis 1 e 2

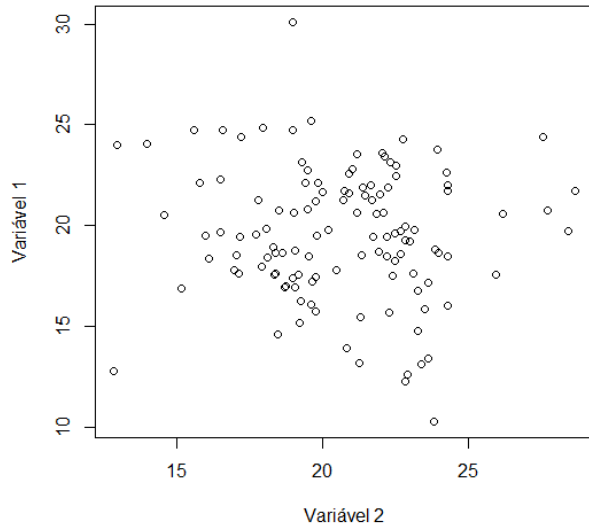


Gráfico de dispersão para as variáveis 2 e 3

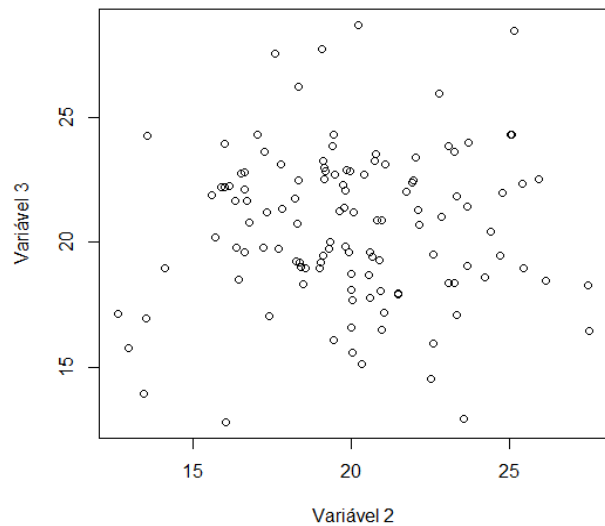
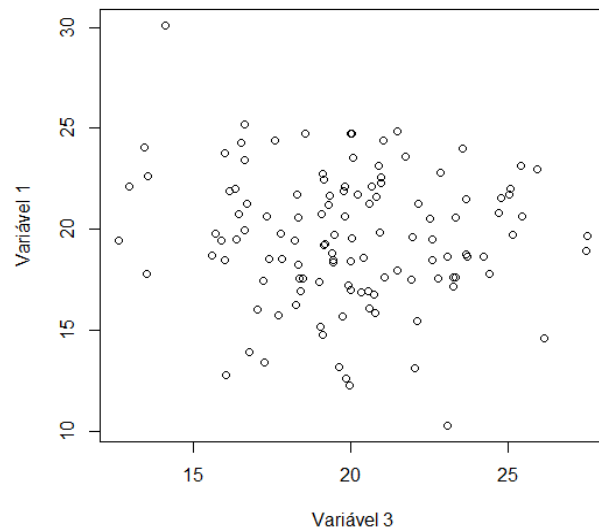


Gráfico de dispersão para as variáveis 1 e 3



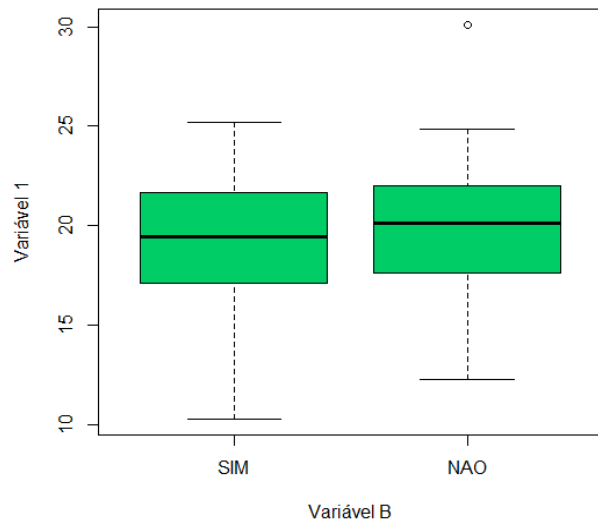
foi de 0.03164116, também sendo próximo de zero, mas positivo, indicando uma relação direta. Para o teste de correlação com essas variáveis, foi obtido t de Student de 0.34388 e p-valor de 0.7315. Já o Teste T serve para comparar as médias de dois grupos. Para o Teste T das variáveis 1 e 2 foram obtidos t de Student igual a -2.596 e p-valor de 0.01002; para as variáveis 1 e 3, t de Student igual a -0.95026 e p-valor de 0.3429; e para as variáveis 2 e 3, t de Student de 1.6636 e p-valor de 0.09752.

Qualitativa e quantitativa:

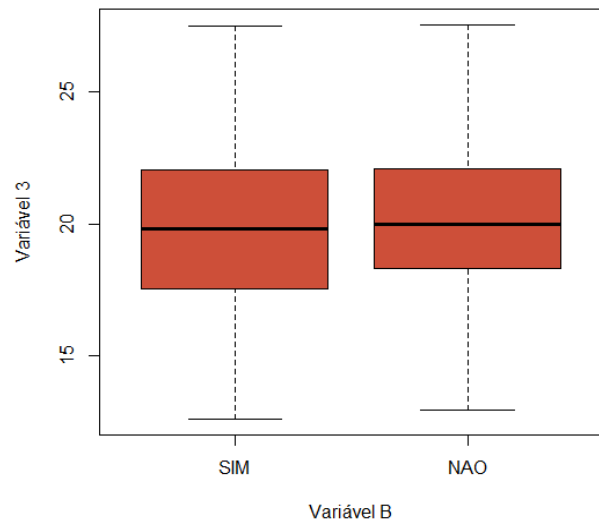
Em um box plot são apresentados o mínimo, o primeiro quartil, a mediana, o terceiro quartil e o máximo. É possível usá-lo para comparar graficamente mais de um conjunto de medidas com respeito à média, à dispersão e à distribuição, como foi o caso deste projeto.

relação inversa. Para o teste de correlação com as variáveis 1 e 2, foi obtido t de Student igual a -0.85386 e p-valor igual a 0.3949 e, para as variáveis 1 e 3, t de Student foi igual a -0.94344 e o p-valor igual a 0.3474. Já para as variáveis 2 e 3, o coeficiente de correlação

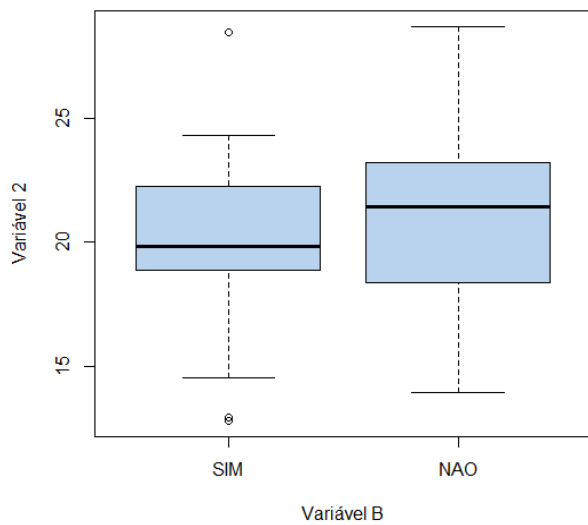
Box plot da Variável 1 segundo a Variável B



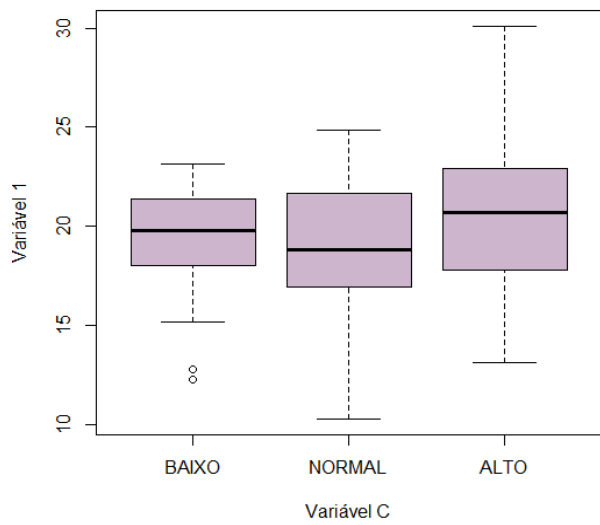
Box plot da Variável 3 segundo a Variável B



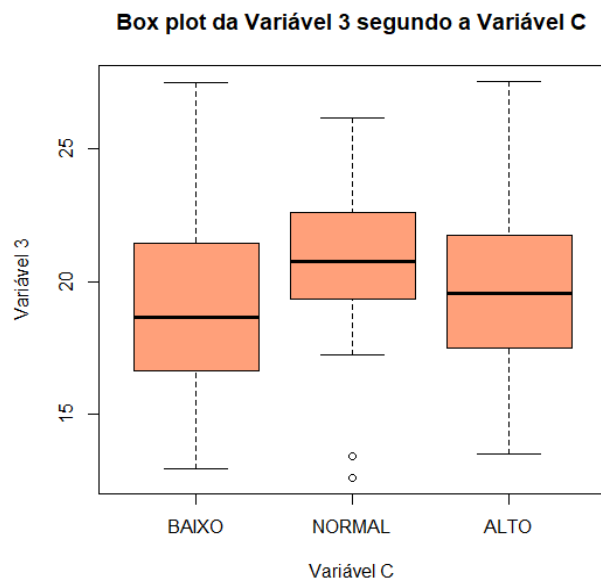
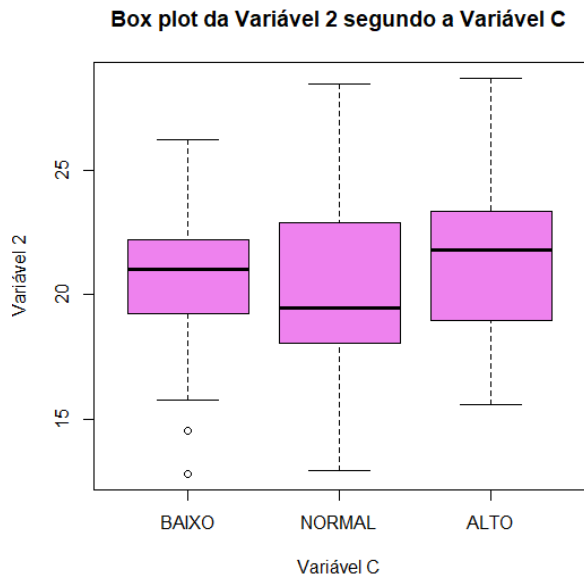
Box plot da Variável 2 segundo a Variável B



Box plot da Variável 1 segundo a Variável C



A descrição por grupo tem como objetivo avaliar o comportamento da variável quantitativa nos grupos da variável qualitativa.



Variáveis 1 e B	Grupos		Variáveis 2 e B	Grupos	
	SIM	NÃO		SIM	NÃO
Média	19.15	19.99	Média	20.29	20.97
Desvio Padrão	3.22	3.29	Desvio Padrão	2.71	3.31
Mediana	19.45	20.1	Mediana	19.84	21.43
Mínimo	10.28	12.24	Mínimo	12.84	13.98
Máximo	25.17	30.09	Máximo	28.43	28.67
Curtose	-0.15	0.41	Curtose	1.11	-0.54

Variáveis 3 e B	Grupos	
	SIM	NÃO
Média	19.86	20.07
Desvio Padrão	3.29	03.02
Mediana	19.8	20
Mínimo	12.63	12.94
Máximo	27.47	27.53
Curtose	-0.51	-0.08

Variáveis 1 e C	Grupos		
	BAIXO	NORMAL	ALTO
Média	19.34	19.02	20.36
Desvio Padrão	2.61	3.41	3.62
Mediana	19.79	18.8	20.69
Mínimo	12.24	10.28	13.14
Máximo	23.13	24.84	30.09
Curtose	0.27	-0.29	-0.08

Variáveis 2 e C	Grupos		
	BAIXO	NORMAL	ALTO
Média	20.48	19.98	21.43
Desvio Padrão	2.66	3.16	3.15
Mediana	21.04	19.46	21.81
Mínimo	12.84	12.95	15.58
Máximo	26.2	28.43	28.67
Curtose	0.7	-0.09	-0.45

Variáveis 3 e C	Grupos		
	BAIXO	NORMAL	ALTO
Média	19.29	20.82	19.78
Desvio Padrão	3.38	2.82	03.09
Mediana	18.68	20.76	19.55
Mínimo	12.94	12.63	13.54
Máximo	27.47	26.16	27.53
Curtose	-0.48	01.02	-0.22

dependente e as variáveis 2 e 3 as independentes.

$$V1 = 22.99877 - 0.08153V2 - 0.08746V3 \quad (1)$$

Modelo de Regressão

Foi criado o seguinte modelo, sendo a variável 1 a

Neste modelo, V1 é a variável 1; V2, a variável 2; e V3, a variável 3. O R^2 múltiplo ou coeficiente de determinação é usado para determinar quão bem o modelo se ajusta aos dados, informando quanto

da variância da variável dependente pode ser explicada pelas variáveis independentes. Quanto mais próximo de 1, maior a qualidade deste modelo. Para o modelo apresentado em (1), o R^2 múltiplo foi de 0.01321. O R^2 ajustado fornece as mesmas informações, mas ajusta o número de termos no modelo, e para este modelo o R^2 ajustado foi de -0.003657.

Os intervalos de confiança para os coeficientes foram:

- Coeficiente A (equivalente a 22.99877 no modelo): de 17.5223728 a 28.4751739;
- Coeficiente B (equivalente a -0.08153 no modelo): de -0.2774977 a 0.1144471;
- Coeficiente C (equivalente a -0.08746 no modelo): de -0.2766475 a 0.1017181.