

Universidade Estadual de Santa Cruz -
UESC Departamento de Ciências Exatas e Tecnológicas - DCET
CET083 - Probabilidade e Estatística
Curso de Ciência da Computação
Prof. José Cláudio Faria
2019.1

Nome: João Henrique dos Santos Queiroz	Matrícula: 201710737
Nome: Luís Carlos Santos Câmara	Matrícula: 201610337
Nome: Mateus Reis dos Santos	Matrícula: 201720304

As 3 matrículas acima foram usadas para gerar os dados.

1 AED: Apresentações tabulares e gráficas

1.1 Diagrama de caixa (boxplot) para Y1 e Y2

1.1.1 – Y1 e Y2: Antes e após a eliminação de possíveis outliers - sem distinção de sexo.

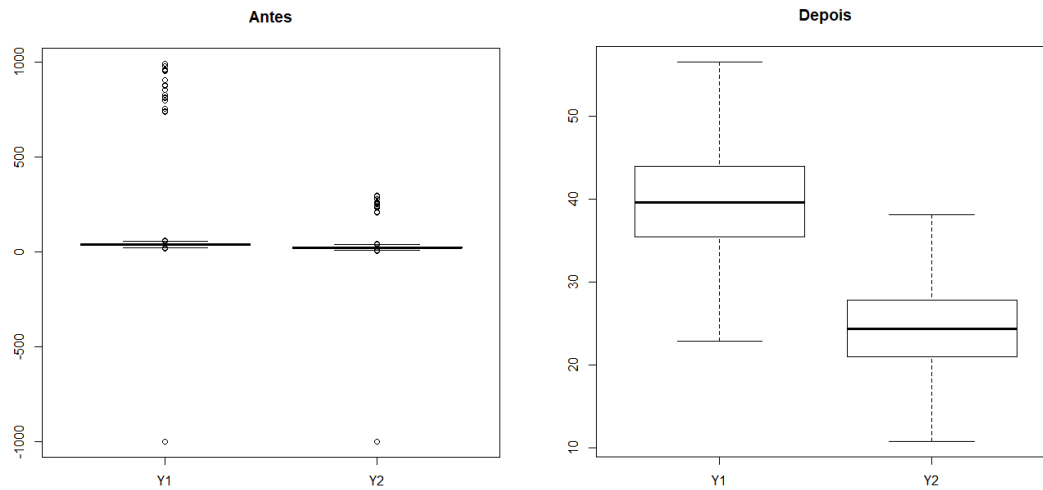


Figura 1 – Diagrama de caixa de Y1 (un) e Y2 (un) antes e após a eliminação de outliers, UESC/BA - 2019.

1.1.2 – Y1 e Y2: Antes e após a eliminação de possíveis outliers – com distinção de sexo.

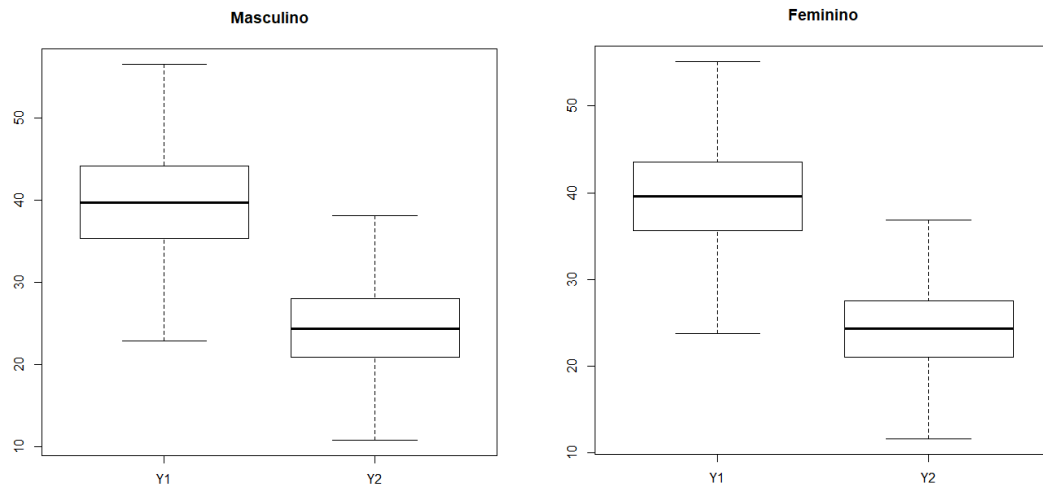


Figura 2 – Diagrama de caixa de Y1 (un) e Y2 (un) (sexo masculino e feminino, respectivamente), UESC/BA - 2019.

1.2. Para Y1

1.2.1 Apresentações tabulares para cada sexo.

Tabela 1 – Tabela de distribuição de frequência de Y1 (un) (sexo masculino), UESC/BA – 2019

Class limits	f	rf(%)	cf(%)
22.64 † 25.52	11	0.83	0.83
25.52 † 28.40	48	3.63	4.47
28.40 † 31.28	79	5.98	10.45
31.28 † 34.16	125	9.46	19.91
34.16 † 37.04	196	14.84	34.75
37.04 † 39.91	219	16.58	51.32
39.91 † 42.79	227	17.18	68.51
42.79 † 45.67	170	12.87	81.38
45.67 † 48.55	136	10.30	91.67
48.55 † 51.43	74	5.60	97.27
51.43 † 54.30	29	2.20	99.47
54.30 † 57.19	7	0.53	100

Tabela 2 – Tabela de distribuição de frequência de Y1 (un) (sexo feminino), UESC/BA – 2019

Class limits	f	rf(%)	cf(%)
22.74 † 25.79	7	1.27	1.27
25.79 † 28.84	19	3.44	4.71
28.84 † 31.89	37	6.70	11.41
31.89 † 34.94	55	9.96	21.38
34.94 † 37.98	105	19.02	40.40
37.98 † 41.03	110	19.93	60.33
41.03 † 44.08	93	16.85	77.17
44.08 † 47.13	58	10.51	87.68
47.13 † 50.18	51	9.24	96.92
50.18 † 53.23	13	2.36	99.28
53.23 † 56.28	4	0.72	100

1.1.2 Histograma e polígono de frequência acumulada

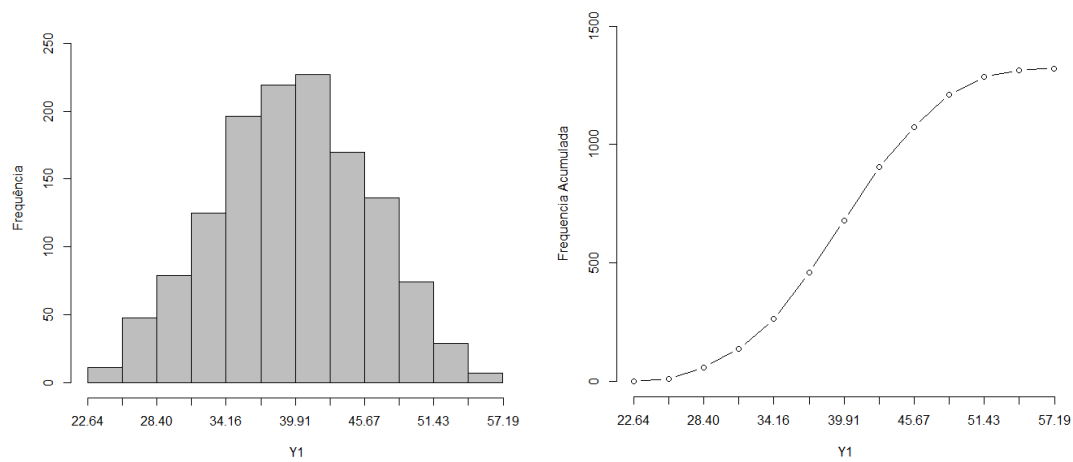


Figura 3 – Histograma e polígono de frequência acumulada de Y1 (un) (sexo masculino), UESC/BA - 2019.

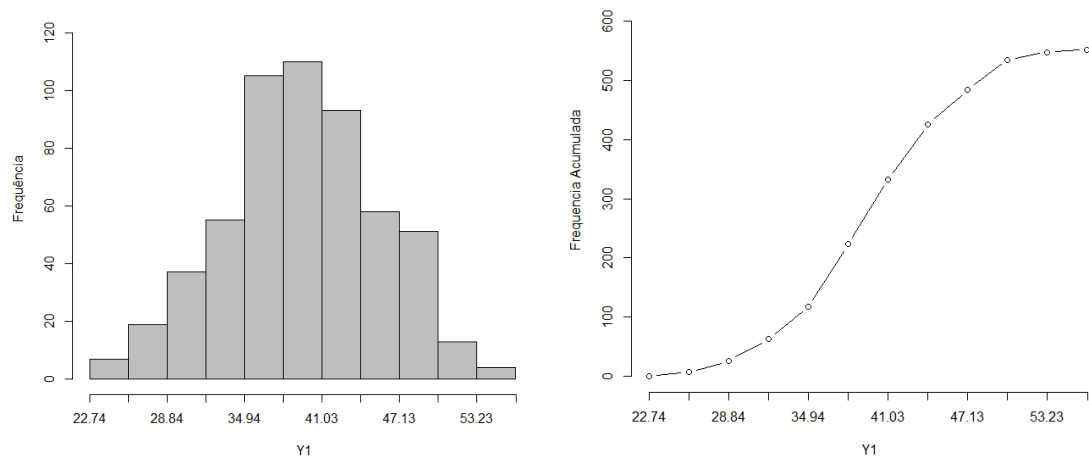


Figura 4 – Histograma e polígono de frequência acumulada de Y1 (un) (sexo feminino), UESC/BA - 2019.

2 AED: medidas estatísticas básicas

2.1 AED: Medidas determinadas a partir dos vetores

2.1.1 Tendência central

Tabela 3 – Medidas de tendência central (sexo masculino), UESC/BA – 2019

	media	mediana	moda
Y1	39.64	39.72	40.91
Y2	24.34	24.40	25.37

Tabela 4 – Medidas de tendência central (sexo feminino), UESC/BA – 2019

	media	mediana	moda
Y1	39.48	39.59	37.07
Y2	24.29	24.29	26.25

2.1.2 Posição

Tabela 5 – Quartis dos usuários (sexo masculino), UESC/BA - 2019

	25%	50%	75%
Y1	35.41	39.72	44.20
Y2	20.92	24.40	28.04

Tabela 6 – Quartis dos usuários (sexo feminino), UESC/BA – 2019

	25%	50%	75%
Y1	35.60	39.60	43.56
Y2	21.06	24.29	27.51

Tabela 7 – Decis dos usuários (sexo masculino), UESC/BA - 2019

	10%	20%	30%	40%	50%	60%	70%	80%	90%
Y1	31.17	34.17	36.33	37.95	39.72	41.36	43.01	45.41	47.83
Y2	17.48	19.91	21.66	22.98	24.40	25.74	27.08	29.01	30.98

Tabela 8 – Decis dos usuários (sexo feminino), UESC/BA - 2019

	10%	20%	30%	40%	50%	60%	70%	80%	90%
Y1	31.63	34.55	36.29	37.90	39.60	40.88	42.34	44.72	47.83
Y2	17.94	20.24	21.58	23.12	24.29	25.53	26.78	28.47	30.32

2.1.3 Dispersão

Tabela 9 – Dispersão dos usuários (sexo masculino), UESC/BA – 2019

	a.t	variância	d.padrão	c.v
Y1	33.75	40.06	6.33	15.97
Y2	27.34	26.28	5.13	21.06

Tabela 10 – Dispersão dos usuários (sexo feminino), UESC/BA – 2019

	a.t	variância	d.padrão	c.v
Y1	32.75	36.96	6.08	15.40
Y2	26.36	23.39	4.83	19.91

2.2 AED: Medidas determinadas a partir de apresentações tabulares

Tabela 11 – Tabela de distribuição de frequência reconstruída de publicação, UESC/BA – 2019

Class limits	f	rf(%)	cf(%)
10.00 f 20.00	7	3.04	3.04
20.00 f 30.00	19	8.26	11.30
30.00 f 40.00	28	12.17	23.48
40.00 f 50.00	38	16.52	40.00
50.00 f 60.00	47	20.43	60.43
60.00 f 70.00	38	16.52	76.96
70.00 f 80.00	27	11.74	88.70
80.00 f 90.00	19	8.26	96.96
90.00 f 100.00	7	3.04	100.00

Tabela 12 – Medidas de tendência central, UESC/BA – 2019

	m	md	mo
medida	54.91	54.89	55.00

Tabela 13 – Medidas de posição: quartis, UESC/BA – 2019

	10%	20%	30%
quartil	40.92	54.89	68.82

Tabela 14 – Medidas de posição: decis, UESC/BA – 2019

	10%	20%	30%	40%	50%	60%	70%	80%	90%
decil	28.42	37.14	43.95	50.00	39.60	59.79	65.79	72.59	81.58

Tabela 15 – Medidas de dispersão, UESC/BA – 2019

	a.t	variância	d.padrão	c.v
medida	230	376.41	19.40	35.33

3 AED: Medidas estatísticas de associação e regressão linear

3.1 Associação

3.1.1 Estimativas: covariância e correlação linear simples

Tabela 16 – Matriz de variâncias e covariâncias (sexo masculino), UESC/BA – 2019

	Y1	Y2
Y1	40.06	32.45
Y2	32.45	26.28

Tabela 17 – Matriz de variâncias e covariâncias (sexo feminino), UESC/BA – 2019

	Y1	Y2
Y1	36.96	-20.03
Y2	-20.03	23.39

Tabela 18 – Matriz de correlações lineares simples (sexo masculino), UESC/BA – 2019

	Y1	Y2
Y1	1.00	0.99
Y2	0.99	1.00

Tabela 19 – Matriz de correlações lineares simples (sexo feminino), UESC/BA – 2019

	Y1	Y2
Y1	1.00	-0.68
Y2	-0.68	1.00

3.1.2 Diagrama de dispersão dos dados

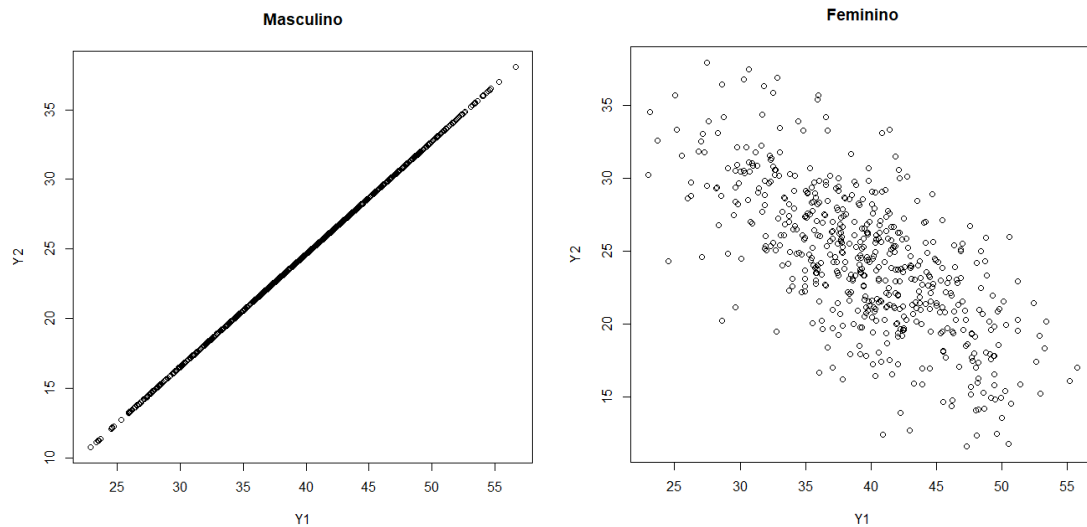


Figura 5 – Diagrama de dispersão de Y1 (un) e Y2 (un) (sexo masculino e feminino, respectivamente), UESC/BA – 2019.

3.1.3 Comparação de estudos semelhantes

Visto que tal estudo se trata de unidade de medidas distintas, a medida estatística mais recomendada para o caso seria a correlação. Neste caso deseja-se uma medida na qual não é influenciada pela unidade de medida, que é o caso da correlação, que também não é influenciada pelo tamanho da amostra.

3.2 Regressão linear

3.2.1 Ajustamento

Tabela 20 – Polinômio grau I, UESC/BA – 2019

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.57	0.64	5.58	0.00
X	0.99	0.11	9.21	0.00

Tabela 21 – Polinômio grau II, UESC/BA - 2019

	Estimate	Std. Error	T value	Pr(> t)
(Intercept)	2.49	0.68	3.68	0.00
X	1.73	0.31	5.50	0.00
I(X^2)	-0.07	0.03	-2.43	0.05

3.2.2. Qual modelo melhor explica o fenômeno em estudo?

Para verificarmos qual modelo melhor explica o fenômeno em estudo, deve-se verificar quais são os coeficientes de determinação R^2 , que explica o grau de ajuste do modelo, ou o percentual de variação. Seu valor varia de 0 a 1, sendo o mais melhor modelo aquele que tem o coeficiente de determinação mais próximo de 1. Este coeficiente vai indicar quanto o modelo foi capaz de explicar os dados coletados.

Além disso podemos ver na figura 6 como o segundo modelo comporta melhor os valores de x e y em comparação com o primeiro modelo.

Tabela 22 – Coeficiente de determinação R^2 dos modelos grau I e II, UESC/BA - 2019

Coeficiente de determinação	
Grau 1	0.91
Grau 2	0.95

3.2.3 Diagrama de dispersão do melhor modelo

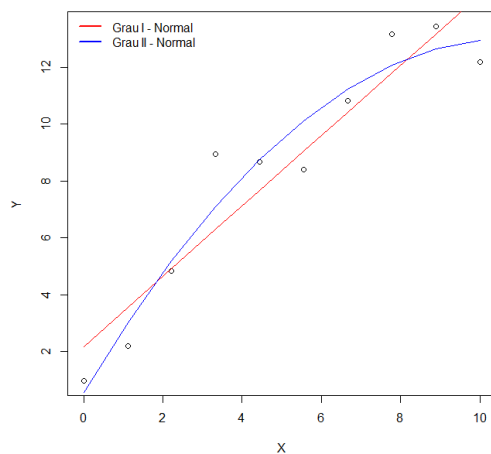


Figura 6 – Diagrama de dispersão dos dados, UESC/BA - 2019.

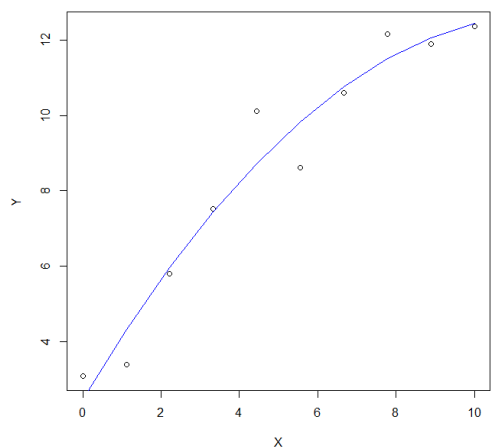


Figura 7 – Diagrama de dispersão dos dados com o melhor modelo, modelo de regressão linear de segundo grau, UESC/BA – 2019.

3.2.4. Pelos critérios de ajustamento e escolha de modelos vistos em aula, os coeficientes de determinação (r^2) de modelos lineares ajustados (forçados e não forçados para a origem) são comparáveis? Justifique com fundamentação estatística.

O R^2 é uma medida descritiva da qualidade do ajuste obtido do modelo, tal valor depende do número de observações e deve ser usado com precaução, visto que é possível torna-lo maior pela adição de um número suficiente em termos no modelo, assim quando não há dados repetidos um polinômio de grau $n-1$ dará um ajuste perfeito, onde $R^2 = 1$, mas quando houver valores repetidos isso nunca irá acontecer, pois não será possível explicar a variabilidade devido ao erro.

Os coeficientes de determinação R^2 são comparáveis, entretanto um valor grande de R^2 , por exemplo, poderá ser grande apenas porque x variou em uma amplitude muito grande, de forma análoga, R^2 poderá ser pequeno porque a amplitude de x foi muito pequena para permitir que uma relação y fosse detectada.

4 Contextualização

O artigo em questão propõe um estudo sobre a ansiedade causada pela estatística na população dos alunos do curso de psicologia, para realização deste artigo foi utilizada uma amostra de 397 estudantes de psicologia, sendo estes de 4 universidades diferentes, destes 307 (77,3%) do sexo feminino e 90 (22,7%), do masculino.

Os dados foram digitados e analisados por meio do SPSS 20, Amos 18 e do programa de Análise Paralela de Patil, Singh, Mishra e Donovan (2007). Foram utilizadas técnicas estatísticas descritivas, Análises Fatoriais Exploratória (AFE) e confirmatória (AFC), Análise Paralela para retenção de fatores, Coeficiente de Correlação de Pearson e Análise de Variância.

Na tentativa de gerar evidências de validade de critério, Oliver et al. (2014) correlacionaram os fatores da SAS com o rendimento acadêmico e o interesse e a satisfação com a estatística. Em geral, nenhum dos fatores apresentou correlação estatística significativa com os critérios.

Um estudo da variância foi fundamental no estudo de Williams para indicar que o imediatismo é bastante relacionado com os fatores da ansiedade estatística, onde ele confirma que 6 a 20% da variância é explicada por essa relação, ou seja, uma distância de 6 a 20% da média dos participantes, o que indica o quão relacionado o imediatismo está à ansiedade. A análise da variância também possibilitou avaliar a diferença entre homens e mulheres nas três dimensões da ansiedade estatística, onde em todas as comparações as mulheres apresentaram médias maiores, mostrando que os transtornos de ansiedade são mais frequentes em mulheres.

A média, juntamente com o desvio padrão foi fundamental na coleta de amostras para a avaliação da SAS (Statistical Anxiety Scale), onde avaliaram 159 indivíduos estudantes por meios de escores com idade média de 21,6 anos (sabendo que a média é uma das medidas de tendência central mais importantes pois engloba todos os dados, porém é influenciada pelos extremos) e com desvio padrão de 3,5 anos, ou seja, com uma margem de erro em torno da média de 3,5 anos, ou 26,5 \pm 3,5.

O artigo possui tanto apresentações tabulares e como uma apresentação gráfica. As apresentações tabulares não se encontram dentro de todas as normas, pois apesar de conter título, cabeçalho, coluna indicadora e os traços obrigatórios, as apresentações tabulares contêm alguns traços que não devem estar presentes, além da ausência da data de referência. Algumas tabelas também possuem um ponto “.” ao fim do título, o que não deve haver. A apresentação gráfica presente também não se encontra de acordo com as normas, pois além de não possuir data de referência, não é o título no local obrigatório.