



**Universidade Estadual de Santa Cruz**  
**Departamento de Ciências Exatas e Tecnológicas - DCET**  
**Curso: Engenharia Elétrica**  
**Discente: José Cláudio Faria**  
**Docente: Sarah Modesto Sanches(202012319) e Kevin Moura(201811019).**

## **Análise Exploratória de dados: CET173 - Probabilidade e Estatística**

**Matriculas: 202012319 201811019 201811019**  
**Os dados foram simulados computacionalmente e tem finalidades apenas acadêmicas e didáticas para fins de avaliação**

12 de novembro de 2021 ILHÉUS-BA

# 1 AED: Apresentações tabulares e gráficas (2.0)

## 1.1 (1.0) Diagrama de caixa (boxplot) para Y1 e Y2:

### 1.1.1 (0.5) Antes e após a eliminação de possíveis outliers - sem distinção de sexo:

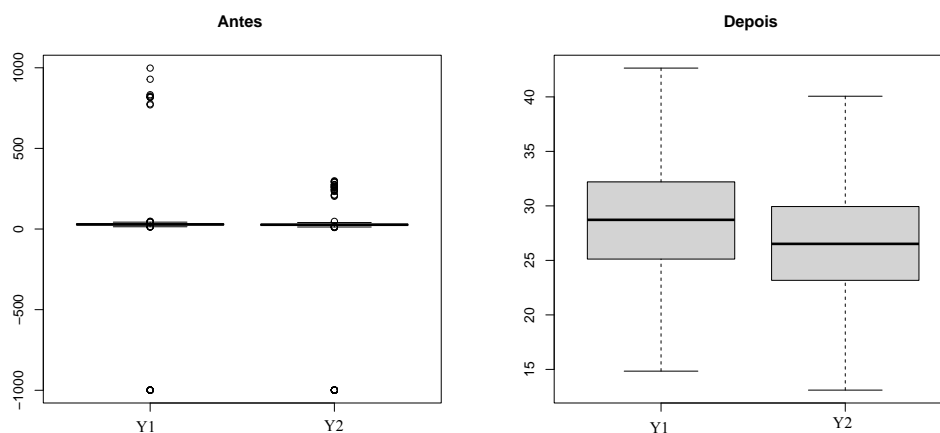


Figura 1: Diagrama de caixa de Y1 (un) e Y2 (un) antes e após a eliminação de outliers, UESC/BA - 2021.

### 1.1.2. (0.5) Após a eliminação de possíveis outliers para cada sexo:

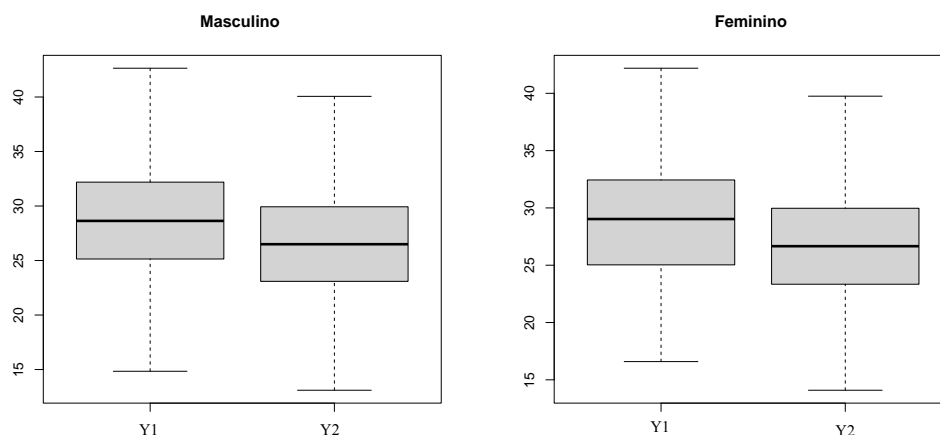


Figura 2: Diagrama de caixa de Y1 (un) e Y2 (un) (sexo masculino e feminino, respectivamente, UESC/BA - 2021).

## 1.2 Para Y1

### 1.2.1 (0.5)Apresentações tabulares

Tabela 1: Tabela de distribuição de frequência de Y1 (sexo masculino), UESC/BA - 2021

Class limits	f	rf	rf(%)	cf	cf(%)
[14.692,17.056)	14.00	0.01	0.94	14.00	0.94
[17.056,19.421)	32.00	0.02	2.14	46.00	3.08
[19.421,21.785)	95.00	0.06	6.36	141.00	9.44
[21.785,24.15)	163.00	0.11	10.91	304.00	20.35
[24.15,26.514)	211.00	0.14	14.12	515.00	34.47
[26.514,28.879)	263.00	0.18	17.60	778.00	52.07
[28.879,31.244)	247.00	0.17	16.53	1025.00	68.61
[31.244,33.608)	206.00	0.14	13.79	1231.00	82.40
[33.608,35.973)	146.00	0.10	9.77	1377.00	92.17
[35.973,38.337)	85.00	0.06	5.69	1462.00	97.86
[38.337,40.702)	22.00	0.01	1.47	1484.00	99.33
[40.702,43.066)	10.00	0.01	0.67	1494.00	100.00

Tabela 2: Tabela de distribuição de frequência de Y1 (sexo feminino), UESC/BA - 2021

Class limits	f	rf	rf(%)	cf	cf(%)
[16.424,19.043)	12.00	0.03	3.17	12.00	3.17
[19.043,21.662)	25.00	0.07	6.61	37.00	9.79
[21.662,24.28)	41.00	0.11	10.85	78.00	20.63
[24.28,26.899)	62.00	0.16	16.40	140.00	37.04
[26.899,29.518)	62.00	0.16	16.40	202.00	53.44
[29.518,32.137)	76.00	0.20	20.11	278.00	73.54
[32.137,34.756)	46.00	0.12	12.17	324.00	85.71
[34.756,37.374)	35.00	0.09	9.26	359.00	94.97
[37.374,39.993)	16.00	0.04	4.23	375.00	99.21
[39.993,42.612)	3.00	0.01	0.79	378.00	100.00

## 1.2.2 (0.5) Histograma e o polígono de frequência acumulada

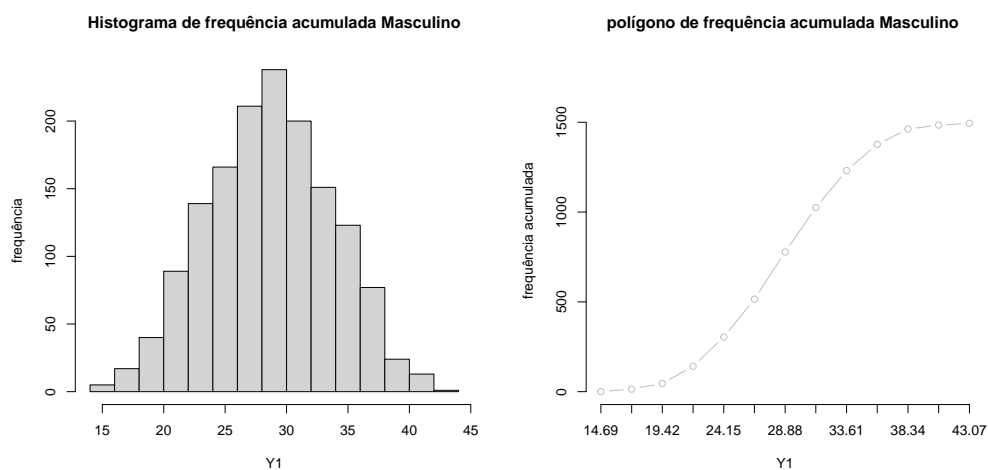


Figura 3: Histograma e polígono de frequência acumulada de Y1 (un) (sexo masculino), UESC/BA - 2021.

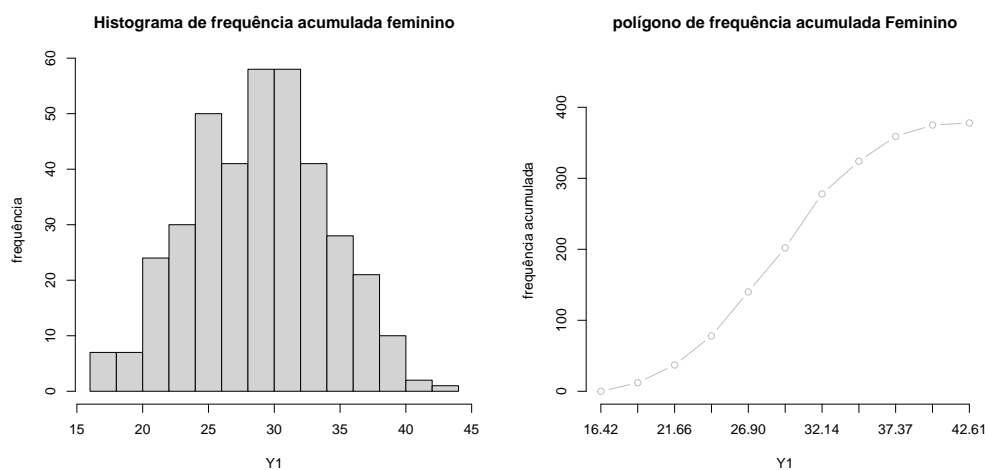


Figura 4: Histograma e polígono de frequência acumulada de Y1 (un) (sexo feminino), UESC/BA - 2021.

## 2 AED: medidas estatísticas básicas(3.0).

### 2.1 (1.5) AED: Medidas determinadas a partir dos vetores

#### 2.1.1 (0.5) Tendência central

Tabela 3: Medidas de tendência central (sexo masculino), UESC/BA - 2021

	n	m	md
Y1	28.66	27.81	28.63
Y2	26.43	31.01	26.49

Tabela 4: Medidas de tendência central (sexo feminino), UESC/BA - 2021

	n	m	md
Y1	28.86	33.15	29.03
Y2	26.52	27.14	26.66

#### 2.1.2 (0.5) Posição

Tabela 5: Quartis dos usuários (sexo masculino), UESC/BA - 2021

	25%	50%	75%
Y1	25.14	28.63	32.19
Y2	23.09	26.49	29.93

Tabela 6: Quartis dos usuários (sexo feminino), UESC/BA - 2021

	25%	50%	75%
Y1	25.04	29.03	32.43
Y2	23.36	26.66	29.99

Tabela 7: Decis dos usuários (sexo Masculino), UESC/BA - 2021

	10%	20%	30%	40%	50%	60%	70%	80%	90%
Y1	21.97	24.09	25.91	27.37	28.63	29.96	31.41	33.10	35.46
Y2	20.00	22.10	23.77	25.26	26.49	27.71	29.12	30.79	32.64

Tabela 8: Decis dos usuários (sexo Feminino), UESC/BA - 2021

	10%	20%	30%	40%	50%	60%	70%	80%	90%
Y1	22.08	24.23	25.69	27.60	29.03	30.30	31.69	33.08	35.84
Y2	19.27	22.51	24.06	25.49	26.66	28.02	29.21	31.11	35.27

### 2.1.3 (0.5) Dispersão

Tabela 9: Dispersão dos usuários (sexo masculino), UESC/BA - 2020

	a.t	variância	d.padrão	c.v
Y1	27.80	25.56	5.06	17.64
Y2	26.96	23.34	4.83	18.27

Tabela 10: Dispersão dos usuários (sexo feminino), UESC/BA - 2020

	a.t	variância	d.padrão	c.v
Y1	25.60	26.79	5.18	17.94
Y2	25.66	25.53	5.05	19.05

## 2.2 AED: Medidas determinadas a partir de apresentações tabulares (1.5)

Tabela 11: Tabela de distribuição de frequência reconstruída de publicação, UESC/BA - 2020

Class limits	f	rf	rf(%)	cf	cf(%)
[10,20)	9.00	0.04	3.90	9.00	3.90
[20,30)	18.00	0.08	7.79	27.00	11.69
[30,40)	28.00	0.12	12.12	55.00	23.81
[40,50)	39.00	0.17	16.88	94.00	40.69
[50,60)	47.00	0.20	20.35	141.00	61.04
[60,70)	39.00	0.17	16.88	180.00	77.92
[70,80)	27.00	0.12	11.69	207.00	89.61
[80,90)	17.00	0.07	7.36	224.00	96.97
[90,100)	7.00	0.03	3.03	231.00	100.00

## 2.3 (0.5) Tendência central

Tabela 12: Medidas de tendência central, UESC/BA - 2021

	n	m	md
medida	54.44	55.00	54.57

### 2.1.2 (0.5) Posição

Tabela 13: Quartis dos usuários, UESC/BA - 2021

	10%	20%	30%
medida	40.71	54.57	68.27

Tabela 14: Decis dos usuários, UESC/BA - 2021

	10%	20%	30%	40%	50%	60%	70%	80%	90%
medida	27.83	36.86	43.67	49.59	54.57	59.49	65.31	71.78	80.53

## 2.4 (0.5) Dispersão

Tabela 15: Dispersão dos usuários, UESC/BA - 2020

	variância	d.padrão	c.v
Y1	377.51	19.43	2.80



### 3 AED: Medidas estatísticas de associação e regressão linear (4.0)

#### 3.1 (1.5) Associação

##### 3.1.1 (0.5) Estimativas: covariância e correlação linear simples

Tabela 16: Matriz de variâncias e covariâncias (sexo masculino), UESC/BA - 2021

	Y1	Y2
Y1	25.56	23.00
Y2	23.00	23.34

Tabela 17: Matriz de variâncias e covariâncias (sexo feminino), UESC/BA - 2021

	Y1	Y2
Y1	26.79	-20.63
Y2	-20.63	25.53

Tabela 18: Matriz de correlações lineares simples (sexo masculino), UESC/BA - 2021

	Y1	Y2
Y1	1.00	0.94
Y2	0.94	1.00

Tabela 19: Matriz de correlações lineares simples (sexo feminino), UESC/BA - 2021

	Y1	Y2
Y1	1.00	-0.79
Y2	-0.79	1.00

### 3.1.2 (0.5) Diagrama de dispersão dos dados

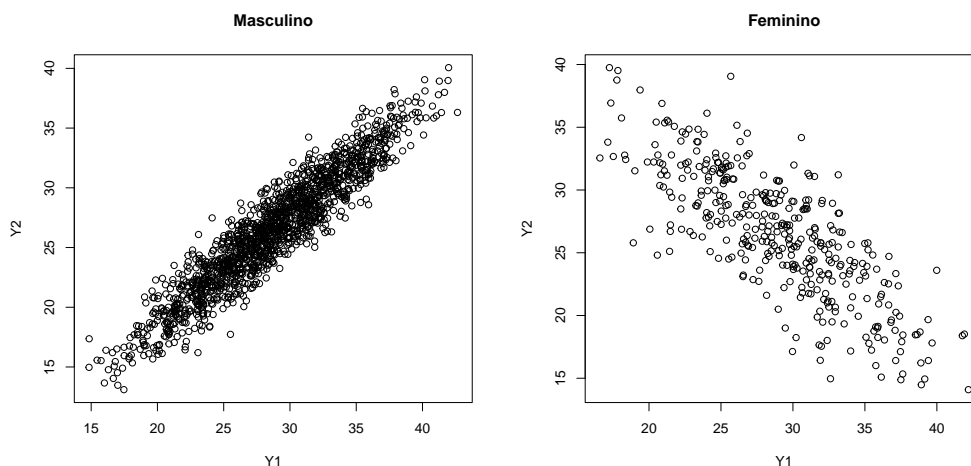


Figura 5: Diagrama de dispersão de Y1 (un) e Y2 (un) (sexo masculino e feminino, respectivamente), UESC/BA - 2021.

### 3.1.3 (0.5) Comparação de estudos semelhantes

Ao procurar entender como uma variável se comporta enquanto outra está variando, a correlação linear - medida de associação que não é influenciada pela unidade de medida das variáveis nem pelo tamanho da amostra- é a mais recomendada para comparar as associações entre as variáveis tanto no estudo que utiliza unidades (un) quanto no que utiliza unidades (100 \* un).

## 3.2 3.2 Regressão linear (2.5)

### 3.2.1 (1.0) Ajustamento

Tabela 20: Polinômio grau I, UESC/BA - 2021

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.7291	0.6237	2.77	0.0242
x	1.3072	0.1052	12.43	0.0000

Tabela 21: Polinômio grau II, UESC/BA - 2021

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.6429	0.6431	1.00	0.3508
x	2.0405	0.2995	6.81	0.0003
I(x <sup>2</sup> )	-0.0733	0.0288	-2.54	0.0385

### 3.2.2 (0.5) Diagrama de dispersão com modelos ajustados

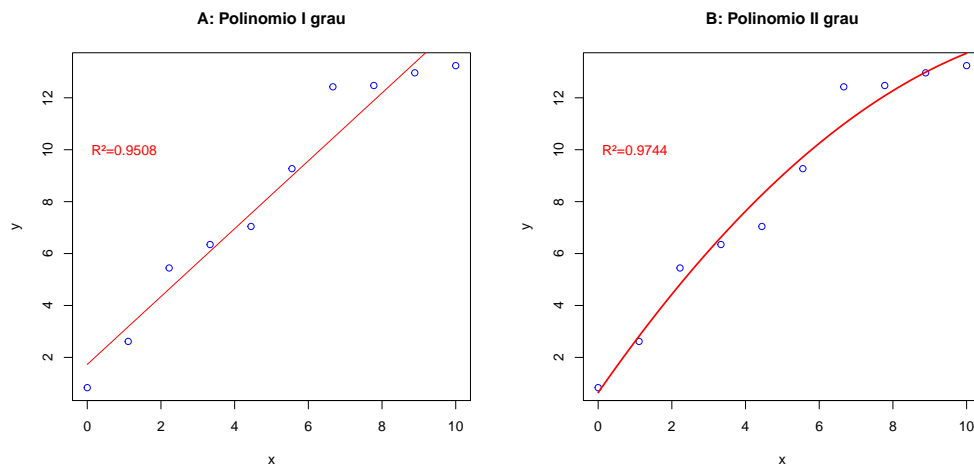


Figura 6: Diagrama de dispersão dos dados, modelos ajustados e respectivos  $r^2$ , UESC/BA - 2021.

### 3.2.3 (0.5) Qual modelo melhor explica o fenômeno em estudo?

A priori conseguimos observar através da análise dos gráficos maior proximidade dos pontos da curva do polinômio de segundo grau. Ademais O modelo linear de 2º grau possui  $R^2 = 0.9744$ , maior que o  $R^2$  do modelo linear de 1º grau. Como o  $R^2$  é uma medida de ajuste de um modelo estatístico linear generalizado - como a regressão linear simples ou múltipla- aos valores observados de uma variável aleatória ele estabelece uma relação de proporcionalidade com o nível de ajuste á amostra. Dessa forma, o modelo linear de 2º grau explica de forma mais clara o fenômeno estudado.

### 3.2.4 (0.5) Critérios de ajustamento e escolha de modelos

Segundo Levine (2014), o coeficiente de determinação múltipla é igual à soma dos quadrados da regressão (SQReg) dividida pela soma total dos quadrados. Além disso, o  $R^2$  é o índice que indica a porcentagem de explicação da variável dependente em função das variáveis independentes com variação de 1% a 100% e que depende do número de observações. Haja vista a possibilidade de torna-lo maior pela adição de números suficiente em termos de modelo, o mesmo deve ser utilizado com precaução. Dessa forma, infere-se que os coeficientes de determinação são comparáveis, levando em consideração que um valor grande de  $R^2$  pode ser fruto de uma amplitude  $\times$  muito grande e um valor pequeno de  $R^2$  podera ser fruto de uma amplitude  $\times$  muito pequena.

### 3.3 Contextualização (1.0))

O artigo selecionado busca analisar a morbimortalidade pela Covid-19 segundo raça, cor e etnia em países como Brasil e Estados Unidos através de dados divulgados tais como; boletins epidemiológicos divulgados pelo Ministério da Saúde, resultados parciais da Pesquisa Nacional por Amostragem de Domicílio (PNAD) COVID-19 realizada pelo IBGE no Brasil e estudos na área da saúde sobre os impactos da pandemia nos Estados Unidos, sob a perspectiva de raça/cor/etnia.

Dos dados constituintes do estudo descritivo e exploratório temos:

- 21 boletins epidemiológicos divulgados pelo Ministério da saúde que investigam aspectos da produção e divulgação dos dados desagregados por raça/cor ao longo do período de janeiro a julho de 2020.

- A Pesquisa Nacional por Amostra de Domicílios, PNAD-COVID-19 a qual estima o número de pessoas com sintomas referidos associados à síndrome gripal. A coleta desta pesquisa teve início em 4 de maio de 2020 e entrevistou por telefone 48 mil domicílios por semana, totalizando 193 mil domicílios por mês. Posto que os dados de hospitalizações de SRAG por COVID-19 aparecem somente em gráficos com porcentagens e com uma nota de rodapé explicando que 14.698 casos foram excluídos da análise devido ao campo raça/cor ter sido registrado como ignorado. A partir dos gráficos observamos que 43,1% dos óbitos ocorreram entre pessoas de raça/cor branca, seguido da raça/cor parda (47,3%) e preta (7,5%). Dentre estes óbitos, 4.425 tiveram a raça/cor ignorada e não foram incluídos na análise.

Ademais, o artigo possui tanto apresentações gráficas como uma apresentação tabular, nas quais fica evidente a divergência da norma verificada no ponto “.” ao fim do título, o que não deve haver. Em contrapartida as apresentações gráficas presentes possuem data de referência, fonte e local obrigatório.

Não obstante, o estudo do coeficiente de variação, das frequências absolutas e relativas apresentados na Tabela 1, evidenciam o maior acometimento da população negra pela COVID-19 e maior utilização do SUS por esta população para o cuidado à saúde.

### 3.4 Referências

LEVINE, D. M. et al.. Estatística: teoria e aplicações usando o Microsoft Excel em português. 6.ed. Rio de Janeiro: LTC, 2014

MORBIMORTALIDADE pela Covid-19 segundo raça/cor/etnia: a experiência do Brasil e dos Estados Unidos. Saúde em debate, [S. l.], p. 0-22, 23 ago. 2021.