Universidade Estadual de Campinas
Instituto de Matemática, Estatística e Computação Cientifica
Departamento de Estatística

Relatório - Parte I
Trabalho Final de ME613

Eliane Ramos de Siqueira RA:155233
Guilherme Pazian RA:160323
Henrique Capatto RA:146406
Murilo Salgado Razoli RA:150987

Professor: Caio Lucidius Naberezny Azevedo

Campinas-SP, 06 de Dezembro de 2016

1. Introdução

O conjunto de dados analisado corresponde a informações de homens e mulheres envolvidos em exercícios regulares e apresenta para cada indivíduo, o peso(em kg) e altura(em cm) medidos e informados pelo mesmo. Além disso, o sexo de cada indivíduo também foi coletado, sendo que 112 são do sexo feminino e 88 são do sexo masculino, totalizando 200 pessoas.Os dados podem ser encontrados no R no pacote car, sob o nome “Davis”.

O objetivo é estudar o impacto da altura no peso, levando em consideração o sexo.

Utilizamos a metodologia dos modelos normais lineares homocedásticos, metodologias da qualidade do ajuste e comparação de modelos apropriados com o suporte computacional do R.

2. Análise descritiva

Observando a tabela 1, podemos ver que em média, o peso e a altura dos homens são maiores que os das mulheres. Além disso vemos valores superiores em todas as estatísticas, para os homens, incluindo uma maior variação nos dados, variação essa, mostrada pelos valores de erro padrão.

Tabela 1: Estatísticas Descritivas

Sexo	Variável	Minímo	1ºQuartil	Mediana	Média	3ºQuartil	Máximo	Erro Padrão
F	Peso	39	52.50	56	56.89	62	78	6.890509
F	Altura	148	161.50	165	164.70	169	178	5.683460
M	Peso	54	67.75	75	75.90	83	119	11.890342
M	Altura	163	173.00	178	178.00	183	197	6.440701

A figura 1 mostra um boxplot dos valores de peso por gênero e um boxplot dos valores de altura também por gênero. Observando-os podemos confirmar os padrões já identificados pelas estatísticas descritivas.

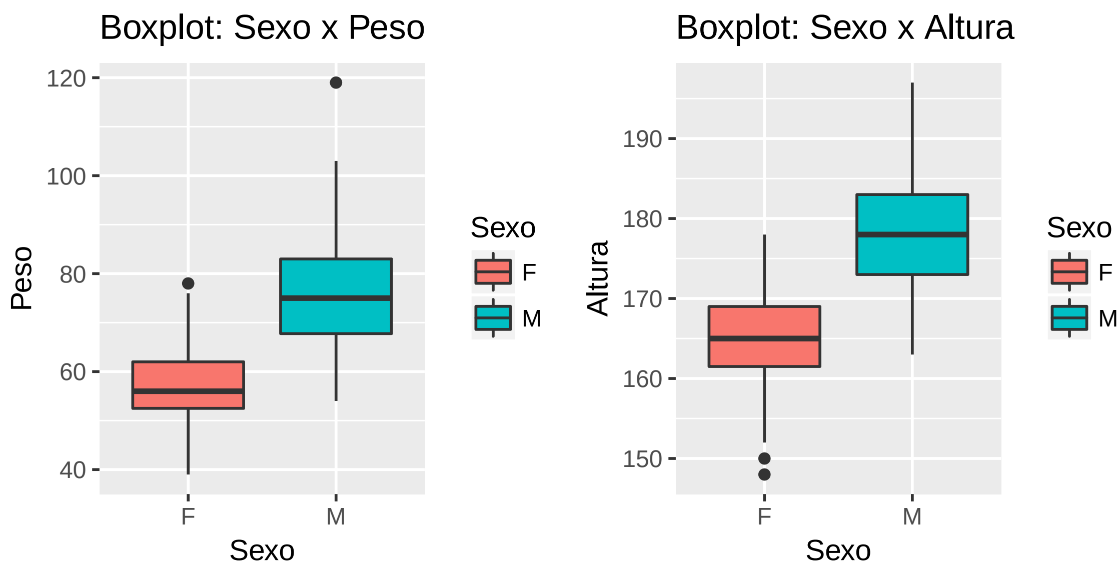

Figura 1: Boxplot Comparativo

Na figura 2 abaixo, temos um gráfico de dispersão do peso em relação a altura dos indivíduos, desconsiderando o gênero. Podemos perceber que há uma relação positiva entre a váriavel resposta(peso) e a variável explicativa(altura), isto é, quanto maior a altura, maior o peso do indivíduo. Tal relação pode ser razoavelmente representada por uma reta ou uma curva quadrática.

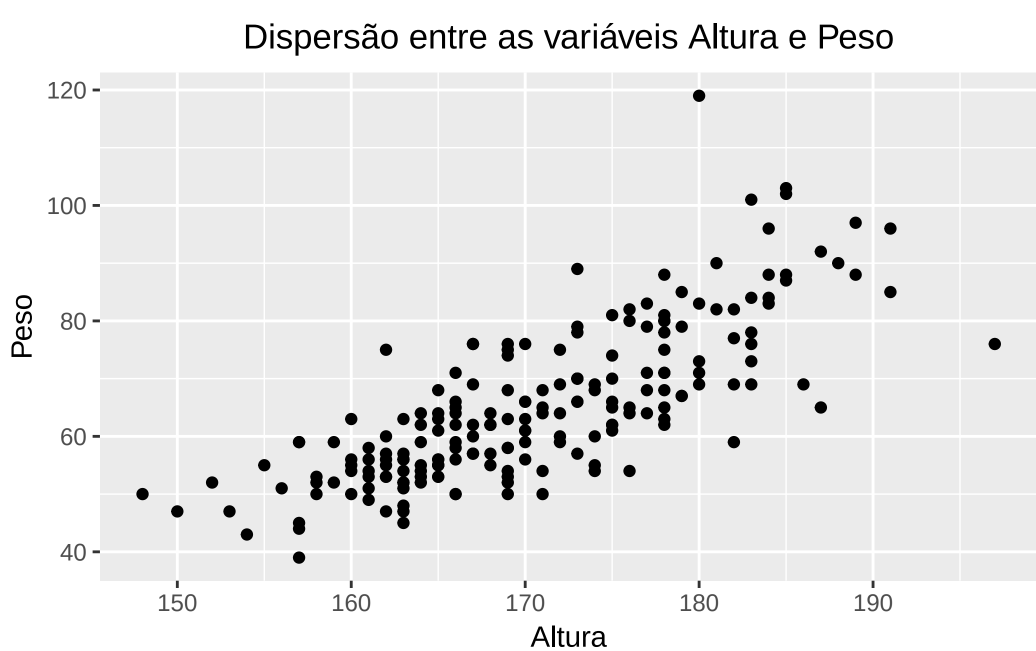

Figura 2: Dispersão entre as variáveis Altura e Peso

A figura 3 também apresenta os gráficos de dispersão entre a variável resposta e explicativa, desta vez considerando o sexo de cada indivíduo. Podemos ver que para ambos os sexos,há uma relação positiva, o peso cresce a medida que a altura cresce. Além disso é possível notar a presença de um outlier para o sexo masculino, enquanto que para o sexo feminino podemos ver em torno de três outliers. Tanto para o sexo feminino quanto o masculino podemos ainda razoavelmente representar esta relação entre peso e altura por uma reta ou uma curva quadrática.

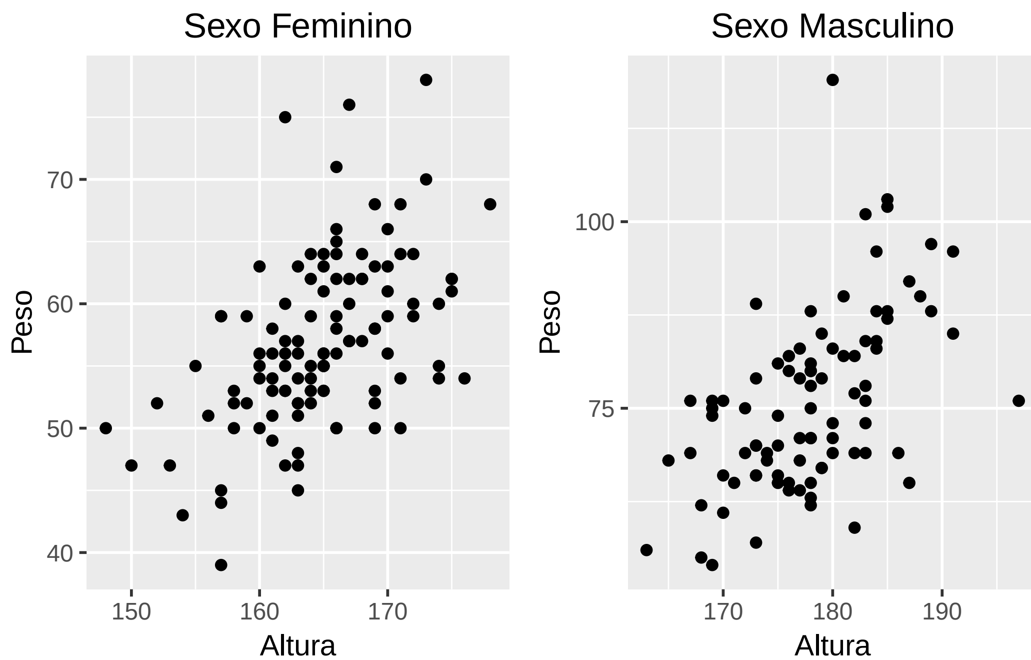

Figura 3: Dispersão entre as variáveis Altura e Peso considerando o sexo.

3. Análise Inferencial

Devido ao objetivo em questão e aos resultados da análise descritiva, vamos considerar quatro modelos e verificar qual deles é o mais reduzido e melhor se ajusta aos dados.

Para os dois primeiros, consideramos modelos com intercepto, o primeiro deles, é um modelo linear centrado na média, enquanto que o segundo é um modelo quadrático também centrado na média. Esses dois modelos nos permite identificar se o intercepto modifica as análises sobre a significância da covariável sexo.

Para os outros dois, consideramos o modelo sem intecepto, onde o primeiro deles é modelo linear e o segundo, um modelo quadrático. Tal sugestão foi feita, pois além dos objetivos descritos acima, desejamos identificar se esses modelos fornecem informação relevante, mesmo contendo menos paramêtros.

Dado a natureza dos dados, não faz sentido observarmos algum valor diferente de zero para a variavel resposta quando a variavel explicativa for é igual a zero, por isso, os modelos com intercepto sugeridos são todos centrados na média.

Modelo 1

$$Y_{ij} = \beta_{0i} + \beta_{1ij}(x_{ij} - \bar{x}) + \varepsilon_{ij} \begin{cases} i = 1,2 \\ j = 1,\ldots,199 \end{cases}$$

Onde: $\varepsilon_{ij} \sim N(0, \sigma^2)$.

- Y_{ij}: Peso do j-ésimo indíviduo do i-ésimo sexo.
- x_{ij}: Altura do j-ésimo indivíduo do i-ésimo sexo.
- β_{0i}: Peso esperado indivíduo do i-ésimo sexo, quando sua altura é igual a média.
- β_{1ij}: Incremento (positivo ou negativo) no peso esperado do j-ésimo indivíduo do i-ésimo sexo, para o aumento em uma unidade na altura.

Modelo 2

$$Y_{ij} = \beta_0 i + \beta_1 ij(x_{ij} - \bar{x}) + \beta_2 ij(x_{ij} - \bar{x})^2 + \varepsilon_{ij} \begin{cases} i = 1,2 \\ j = 1,\ldots,199 \end{cases}$$

Onde: $\varepsilon_{ij} \sim N(0, \sigma^2)$

- Y_{ij}: Peso do j-ésimo indíviduo do i-ésimo sexo.
- x_{ij}: Altura do j-ésimo indivíduo do i-ésimo sexo.
- β_{0i}: Peso esperado indivíduo do i-ésimo sexo, quando sua altura é igual a média.
- $\beta_1 i$: Incremento(positivo ou negativo) no peso quando se aumenta uma unidade de x_{ij}.
- O pode-se avaliar o significado $\frac{-\beta_1 i}{2\beta_2 i}$ como o valor máximo ou mínimo de peso

Modelo 3

$$Y_{ij} = \beta_1 i(x_{ij}) + \varepsilon_{ij} \begin{cases} i = 1,2 \\ j = 1,\ldots,199 \end{cases}$$

Com: $\varepsilon_{ij} \sim N(0, \sigma^2)$

- Y_{ij}: Peso do j-ésimo indíviduo do i-ésimo sexo.
- x_{ij}: Altura do j-ésimo indivíduo do i-ésimo sexo.

- $\beta_1 ij$: Incremento(positivo ou negativo) no peso quando se aumenta uma unidade de x_{ij}.

  Modelo 4

  Considere o seguinte modelo quadrático, ainda com o fator Sexo:

$$Y_{ij} = \beta_1 i(x_{ij}) + \beta_2 i(x_{ij})^2 + \varepsilon_{ij} \begin{cases} i = 1,2 \\ j = 1,\ldots,199 \end{cases}$$

  Com: $\varepsilon_{ij} \sim N(0, \sigma^2)$

- Y_{ij}: Peso do j-ésimo indívíduo do i-ésimo sexo.
- x_{ij}: Altura do j-ésimo indivíduo do i-ésimo sexo.
- $B_1 ij$: Incremento(positivo ou negativo) no peso quando se aumenta uma unidade de x_{ij}.

Para o modelo 1, como mostra a figura 4, podemos notar pelo gráfico A que uma observação se destaca das demais, porém não temos indícios que possa levar a uma conclusão sobre a dependência dos dados. Já o gráfico B, nota-se uma possível heterocedasticidade nos dados pois entre os valores ajustados de 50 a 80, os resíduos parecem se comportar de maneira mais discrepante em relação aos valores maiores que 80. Tanto pelo gráfico C, quanto pelo gráfico D, nota-se uma leve assimetria positiva nos dados. Além disso, no gráfico D, temos a presença de outliers, o que pode indicar que uma distribuição assimétrica com caudas mais pesadas que a distribuição Normal poderiam ser consideradas à análise. O gráfico de envelopes, na figura 5, mostra uma pequena tendência e alguns pontos fora dos **encontrar uma palavra**, o que sugere uma não normalidade dos dados.

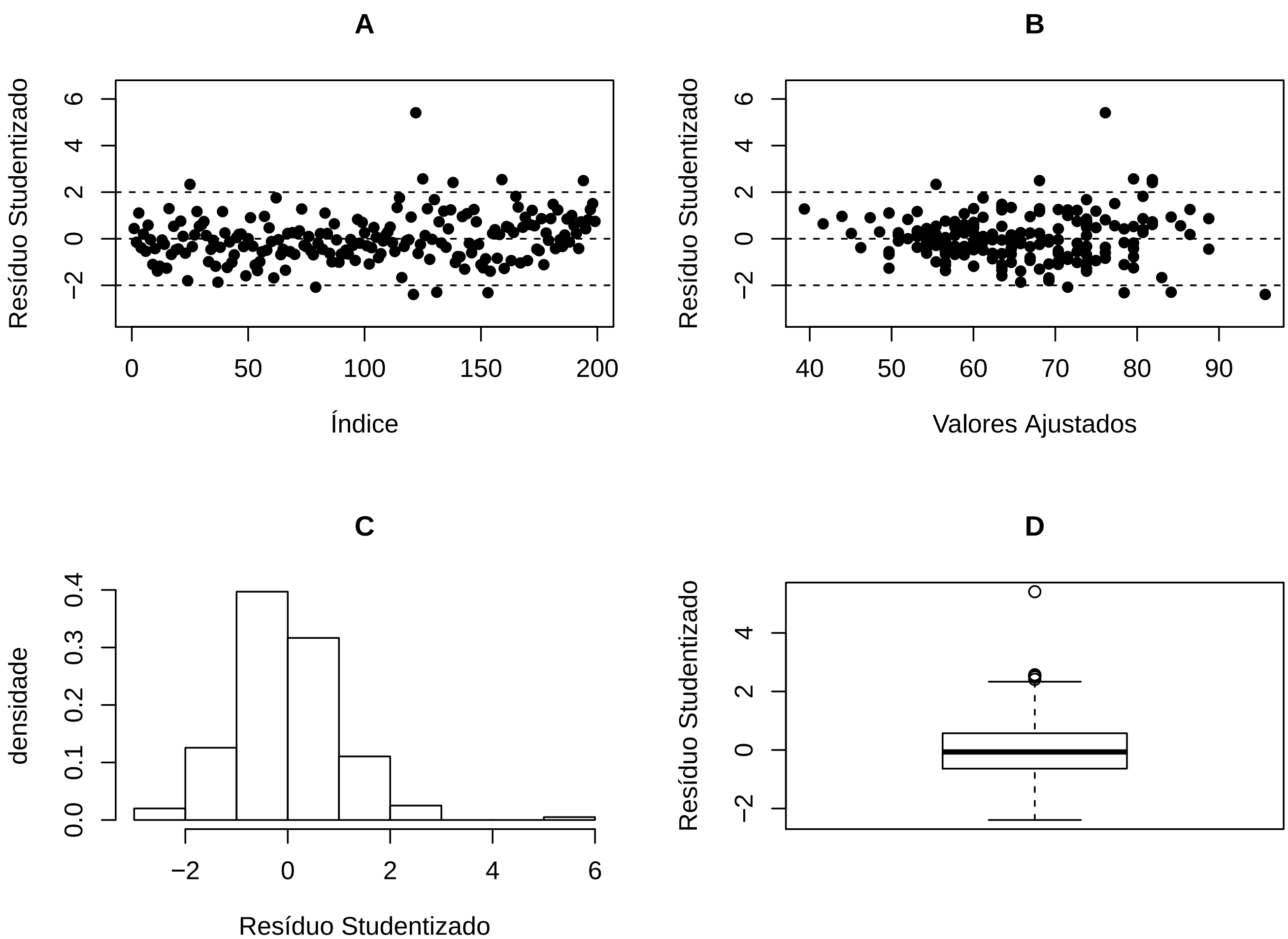

Figura 4: Análise residual para o modelo 1

envelope1

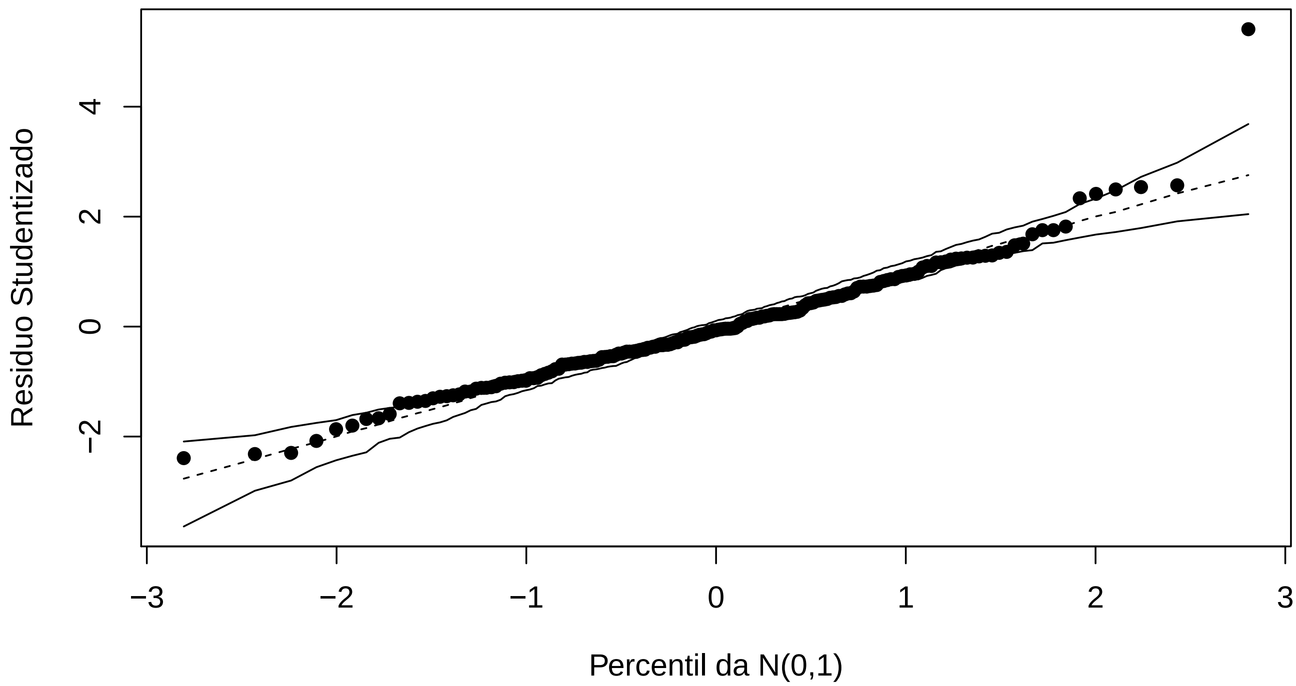

Figura 5: Gráfico de envelope para o resíduos studentizado para o modelo 1

Análise de resíduos para o modelo 2

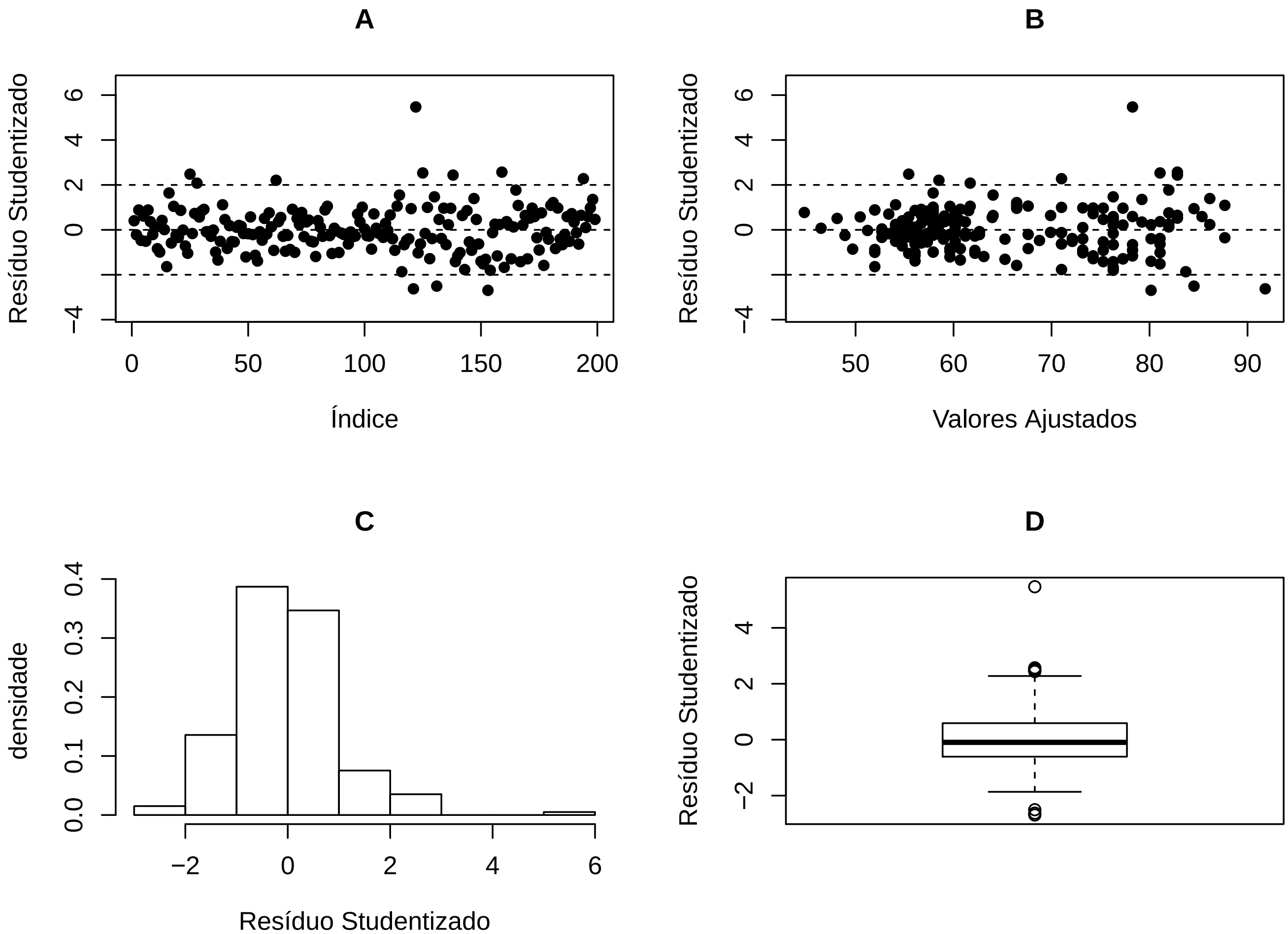

Figura 6: Análise residual para o modelo 2

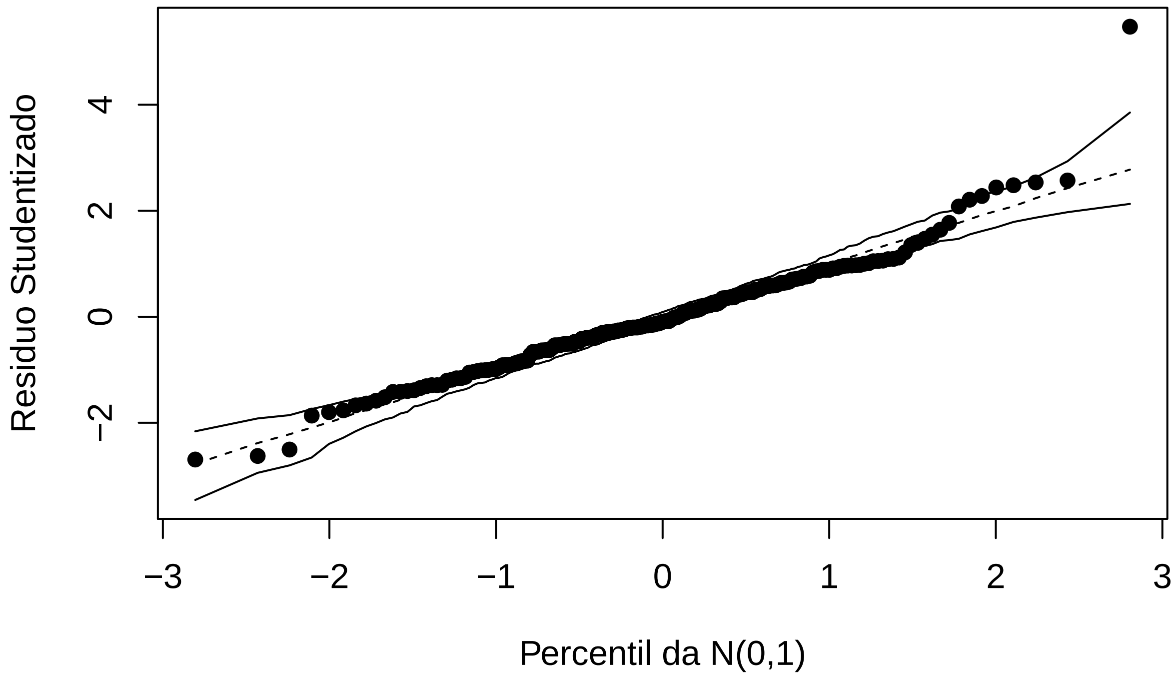

Figura 7: Gráfico de envelope para o resíduos studentizado para o modelo 2

Análise de resíduos para o modelo3

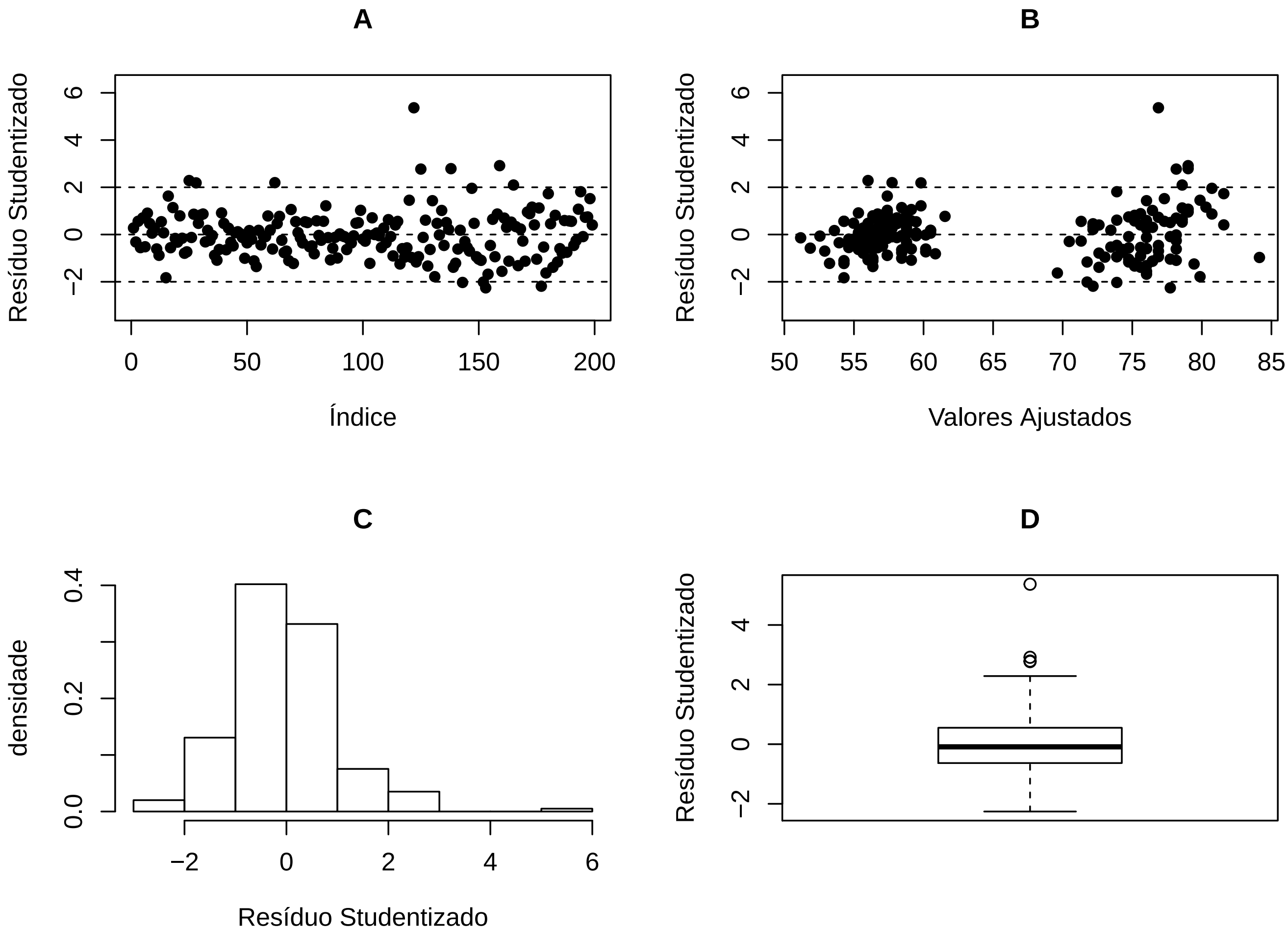

Figura 8: Análise residual para o modelo 3

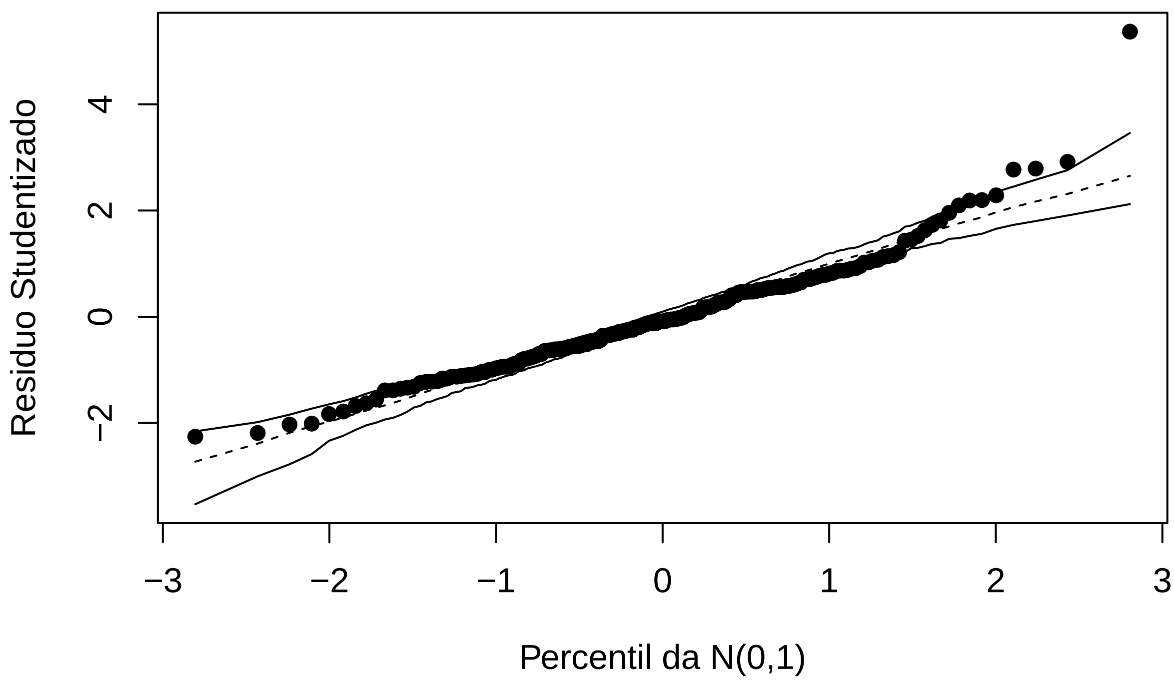

Figura 9: Gráfico de envelope para o resíduos studentizado para o modelo 3

Análise de resíduos para o modelo4

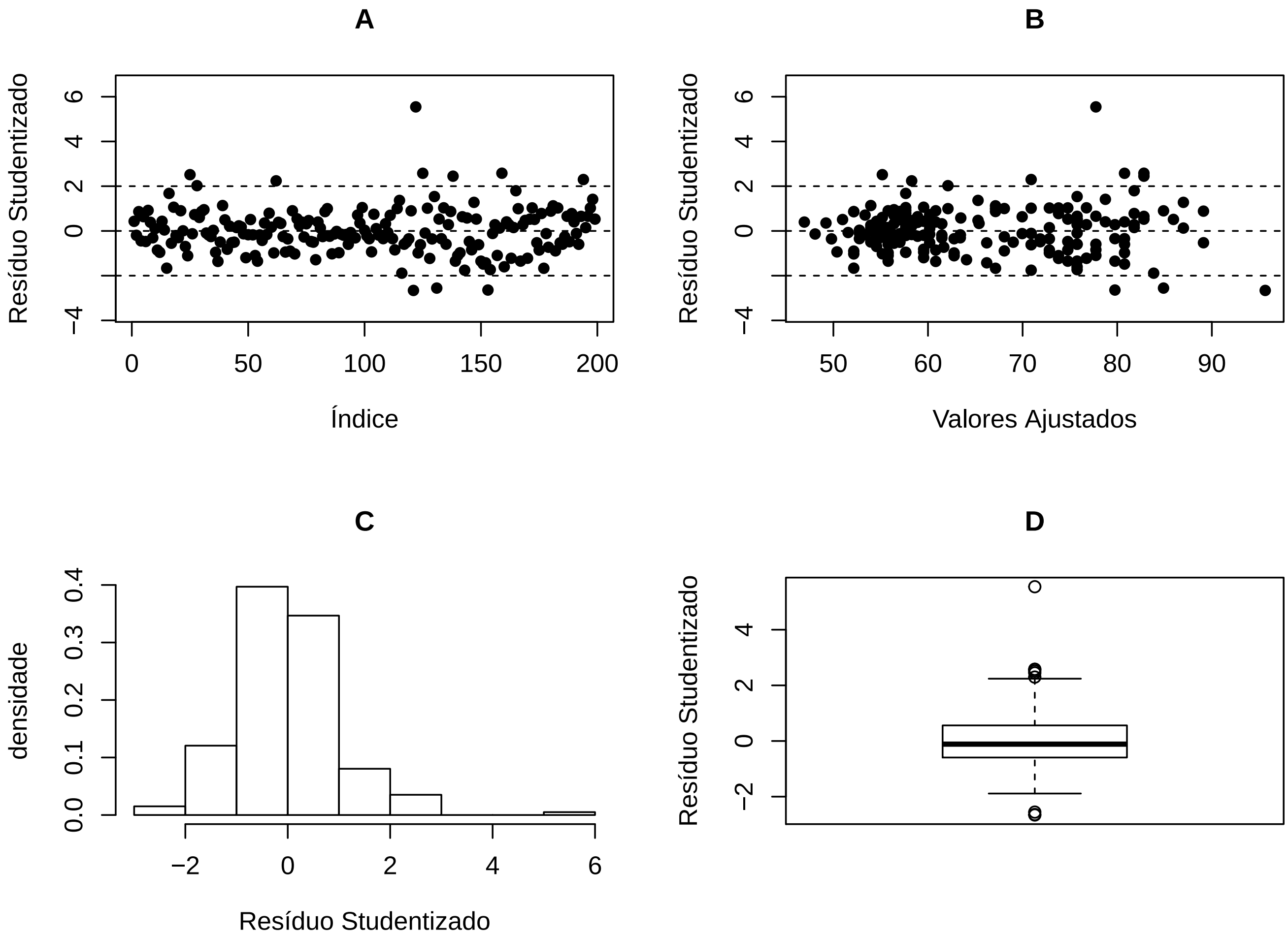

Figura 10: Análise residual para o modelo 4

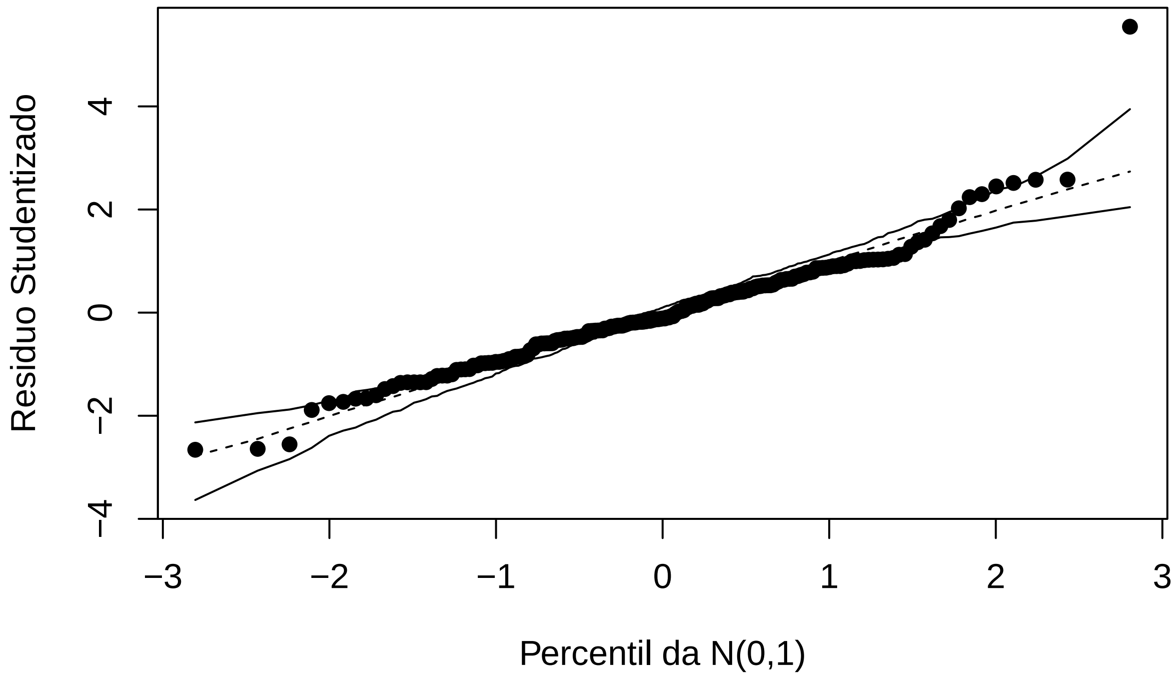

Figura 11: Gráfico de envelope para o resíduos studentizado para o modelo 4

Tabela 2: Estimativas dos parâmetros, intervalo de confiança e teste de nulidade

Modelo	Parâmetro	Estimativa	EP	Valor T	Valor p
Modelo 1	β_{01}	65,30	0,60	108,07	<0,0001
	β_{02}	1,15	0,07	16,98	<0,0001
Modelo 2	β_{01}	60,48	1,12	54,20	<0,0001
	β_{02}	68,30	1,35	50,50	<0,0001
	β_{11}	0,52	0,25	2,07	0,0397
	β_{12}	1,15	0,27	4,24	<0,0001
	β_{21}	-0,01	0,02	-0,46	0,6427
	β_{22}	-0,01	0,01	-0,66	0,5075
Modelo 3	β_{11}	65,30	0,60	108,07	1
	β_{12}	1,15	0,07	16,98	1
Modelo 4	β_{11}	0,07	0,14	0,51	1
	β_{12}	-0,14	0,13	-1,01	1
	β_{21}	0,00	0,00	2,03	1
	β_{22}	0,00	0,00	4,21	<0,0001

4. Conclusões

escrever conclusões

escrever conclusões

escrever conclusões

escrever conclusões escrever conclusões escrever conclusões

5. Referências Bibliográficas

- Azevedo, C. L. N (2016). Notas de aula sobre planejamento e análise de experimentos,http://www.ime.unicamp.br/~cnaber/Material_ME613_2S_2016.htm

- Faraway, J. J. (2014). Linear Models with R, Second Edition,Chapman e Hall/CRC Texts in Statistical Science

- Draper, N. R. and Smith, H. (1998). Applied regression analysis, third edition. New York, NY: John Wiley e Sons.

- Paula, G. A. (2013). Modelos de regressão com apoio computacional, versão pré-eliminar https://www.ime.usp.br/~giapaula/texto_2013.pdf

6. Apêndice