

LABORATÓRIO 5: Regressão Linear

Fernado Bispo, Jeff Caponero

Sumário

Introdução	2
Parte 1: Regressão Linear Simples - Diagnóstico do modelo	3
Metodologia	3
Resultados	4
Ajuste do Modelo	4
Significância do Modelo	4
Análise de Resíduos	6
Parte 2: Regressão Linear Múltipla - Estimação pontual	7

Introdução

O laboratório desta semana está subdividido em duas partes com análises de dois conjuntos de dados distintos que visa a continuidade da aplicação das técnicas de Regressão Linear Simples com a aplicabilidade das técnicas de análise de resíduos e transformação de variáveis inclusive. Para melhor desenvolvimento do processo de análise, este relatório foi dividido em duas partes contendo as análises de cada um dos conjuntos de dados e contando com suas respectivas apresentações sobre o contexto a ser analisado.

Parte 1: Regressão Linear Simples - Diagnóstico do modelo

Metodologia

O conjunto de dados *trees*, disponível no pacote *datasets*, contém informações de 31 cerejeiras (*Black cherry*) da Floresta Nacional de Allegheny, relativas a três características:

- Volume de madeira útil (em metros cúbicos (m^3));
- Altura (em metros (m));
- Circunferência (em metros(m)) a 1,37 de altura.

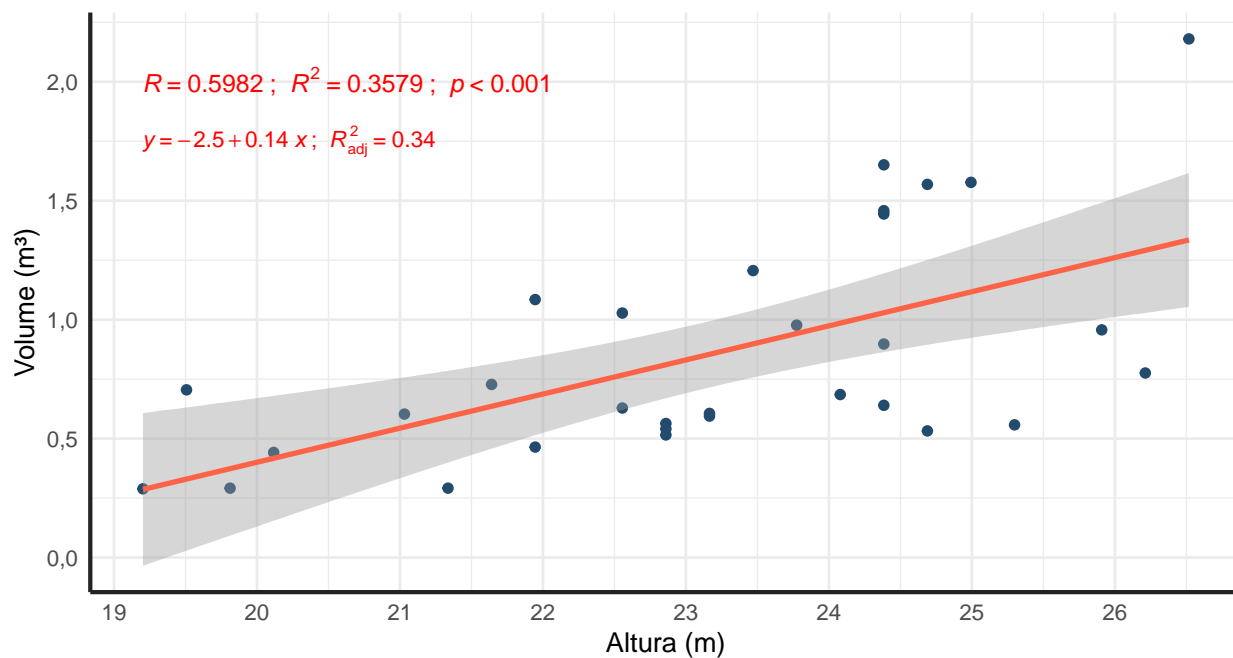
Para esta atividade **serão considerados apenas as informações referentes ao volume e altura das árvores**. Com base nestes dados se desenvolverá:

- (a) Ajuste um modelo linear simples para volume como função da altura da árvore;
- (b) Avaliação gráfica dos resíduos Jackknife para diagnóstico do modelo ajustado;
- (c) Transformações das características;
- (d) Avaliação da transformação mais apropriada dentro da família proposta por Box e Cox;
- (e) Indicação da melhor transformação analisada.

Resultados

Ajuste do Modelo

Figura 1: Modelo Ajustado entre o Volume e Altura



Significância do Modelo

Após o ajuste do modelo existe a necessidade de se avaliar a significância do mesmo, o teste de hipótese para tal situação será realizado, contendo as seguintes hipóteses:

$$H_0 : \hat{\beta}_1 = 0$$

$$H_1 : \hat{\beta}_1 \neq 0.$$

As Tabelas 4 e 5 trazem os principais resultados da tabela ANOVA e do Intervalo de Confiança para os parâmetros, possibilitando assim inferir sobre o modelo ajustado.

Tabela 1: Análise de Variância (ANOVA)

	GL	Soma de Quadrados	Quadrado Médio	Estatística F-Snedecor	p-valor
Regressão	1	2,326	2,326	16,164	0
Resíduos	29	4,174	0,144	NA	NA

Legenda:

¹ GL: Graus de Liberdade

Tabela 2: Intervalos de Confiança para os parâmetros estimados no MRLS.

	LI	LS
Beta 0	-4,162	-0,772
Beta 1	0,070	0,216

Legenda:

¹ LI: Limite Inferior (alpha = 2,5%)

² LS: Limite Superior (alpha = 97,5%)

* Nível de Significância de 5%.

Com base na Tabela 1, avaliando o p-valor é possível afirmar que o modelo é significativo rejeitando assim H_0 que tem como pressuposto $\hat{\beta}_1 = 0$.

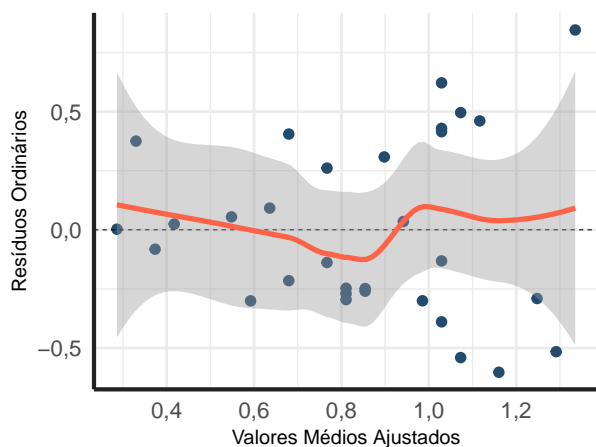
Através dos Intervalos de Confiança calculados é possível afirmar **com 95% de confiança é possível afirmar que o verdadeiro valor de β_0 está entre (-4,1624; -0,7717) e que o verdadeiro valor de β_1 está entre (0,0704; 0,2163).**

Análise de Resíduos

Figura 5: Análise de resíduos do modelo ajustado

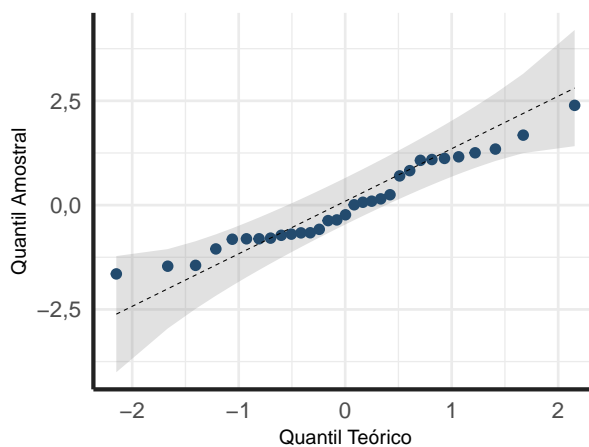
Sub Fig. A:

Gráfico de Resíduos vs. Valores Ajustados



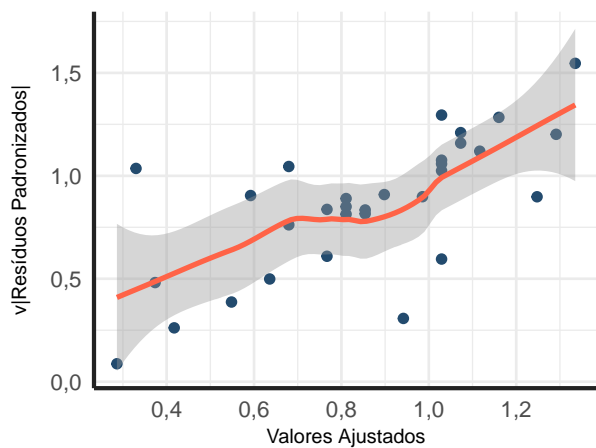
Sub Fig. B:

Gráfico quantil-quantil normal



Sub Fig. C:

Homogeneidade de Variâncias (Locação-Escala)



A Figura que trata da Homogeneidade de Variâncias (Locação-Escala) mostra que há um problema na variabilidade dos dados, ou seja, há uma mudança na variabilidade dos dados, caracterizando uma Heterocedasticidade dos dados.

Parte 2: Regressão Linear Múltipla - Estimação pontual