

Diagnóstico em modelagem

Alfredo Rossi e Cayan

LAMFO

23 de Março de 2019

Diagnóstico em modelagem

Diagnóstico em
modelagem

Alfredo Rossi e
Cayan

Diagnóstico em
modelagem

Diagnóstico em modelagem

Diagnóstico em
modelagem

Alfredo Rossi e
Cayan

Diagnóstico em
modelagem

- ▶ Definição: Diagnóstico é o conjunto de técnicas para verificar a adequabilidade de um modelo
- ▶ Modelos tradicionais (Pressupostos)
- ▶ Random Forest, Redes neurais e similares

Vantagens

Diagnóstico em modelagem

Diagnóstico em
modelagem

Alfredo Rossi e
Cayan

Diagnóstico em
modelagem

- ▶ Desvios sistemáticos
 - ▶ função ligação
 - ▶ função variância
 - ▶ matriz do modelo
- ▶ Discrepâncias isoladas
 - ▶ pontos nos extremos da amplitude
 - ▶ observações erradas

Vantagens

Diagnóstico em modelagem

Diagnóstico em
modelagem

Alfredo Rossi e
Cayan

Diagnóstico em
modelagem

- ▶ Avaliar se pressupostos são atendidos
- ▶ Anomalia de observações
- ▶ Interpretabilidade e entendimento do modelo

Desvantagens

Diagnóstico em modelagem

Diagnóstico em
modelagem

Alfredo Rossi e
Cayan

Diagnóstico em
modelagem

- ▶ Pressupostos limitam a adequabilidade para qualquer base
- ▶ Pressupostos violados levam a resultados duvidosos

Métodos

Diagnóstico em modelagem

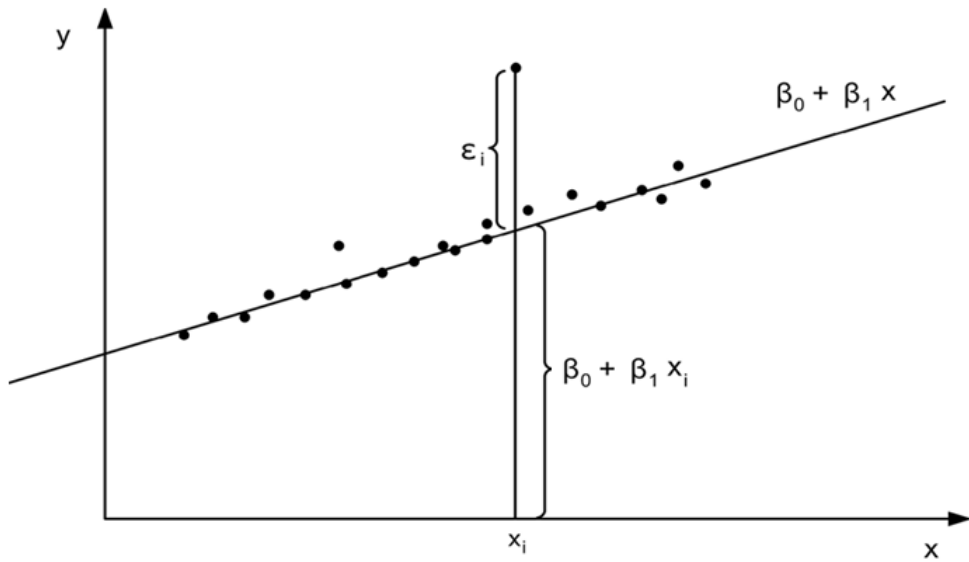
Diagnóstico em
modelagem

Alfredo Rossi e
Cayan

Diagnóstico em
modelagem

Métodos

- ▶ Resíduos
- ▶ Testes e métricas (usando resíduos)
- ▶ Gráficos



Métodos

- ▶ Alavancagem, levege
- ▶ Pode incluir parâmetros de dispersão e ligação
- ▶ $H = X(X^T X)^{-1} X^T$
- ▶ $\hat{y} = yH$

Resíduos

- ▶ $r_i = y_i - \hat{y}_i$
- ▶ $var(r_i) = \sigma^2(I - H)$
- ▶ Resíduo comum pode não ser adequado

Resíduos Estudentizados Internamente (var distinta)

- ▶ $rsi_i = \frac{r_i}{s \times \sqrt{1 - h_{ii}}}$
- ▶ Padronização

Resíduos Estudentizados Externamente (jackknifed)

- ▶ $rse_i = \frac{r_i}{s_{(i)} \times \sqrt{1 - h_{ii}}}$
- ▶ Sem a observação i

Outliers e Pontos influentes

- ▶ no ajuste geral do modelo
- ▶ no conjunto das estimativas dos parâmetros
- ▶ na estimativa de um determinado parâmetro

Outliers e Pontos influentes

- ▶ Medida de leverage: h_{ii} ;
- ▶ Medida de inconsistência: $rse(i)$;
- ▶ Medida de influência sobre o parâmetro β_j : $DFBetaS(i)$ para β_j ;

Outliers e Pontos influentes

- ▶ Medidas de influência geral:
 $DFFitS(i)$, $D(i)$ ou $C(i)$.

Critérios

- ▶ Ponto Inconsistente: $|rse_{(i)}|$ grande, $|rse_{(i)}| \geq t_{1\alpha/(2n);np1}$
- ▶ Ponto de alavanca: h_{ii} grande, $h_{ii} \geq 2p/n$

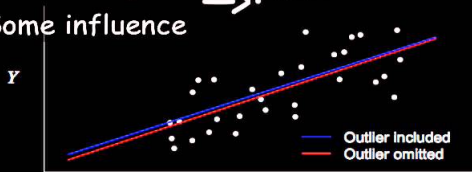
Critérios

- ▶ Outlier: $rse_{(i)}$ grande e h_{ii} pequeno
- ▶ Ponto influente:
 $DFFitS_{(i)}$, $C_{(i)}$, $D_{(i)}$ ou $DFBetaS_{(i)}$
grande ($>0,5$)



Low leverage
Some influence

→ Outlier 1

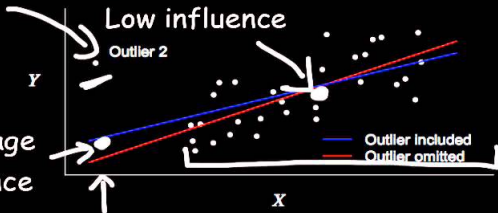


High leverage
High influence

Low leverage
Low influence

Some leverage
High influence

High leverage
Low influence



Testes

- ▶ Teste de Razão Verossimilhança, Goodness of fit, Lack of fit
- ▶ Testes de Normalidade (Shapiro-Wilk, KS, LL)

Testes

- ▶ Testes de homocedasticidade (White, Breuch-Pagan)
- ▶ Teste de Hosmer-Lemeshow
- ▶ Teste para parâmetro de Dispersão

Gráficos

Diagnóstico em modelagem

Diagnóstico em
modelagem

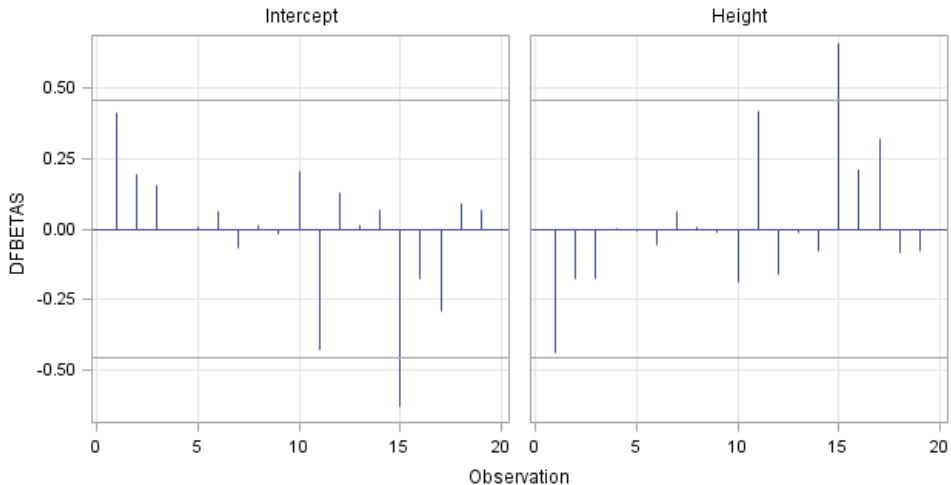
Alfredo Rossi e
Cayan

Diagnóstico em
modelagem

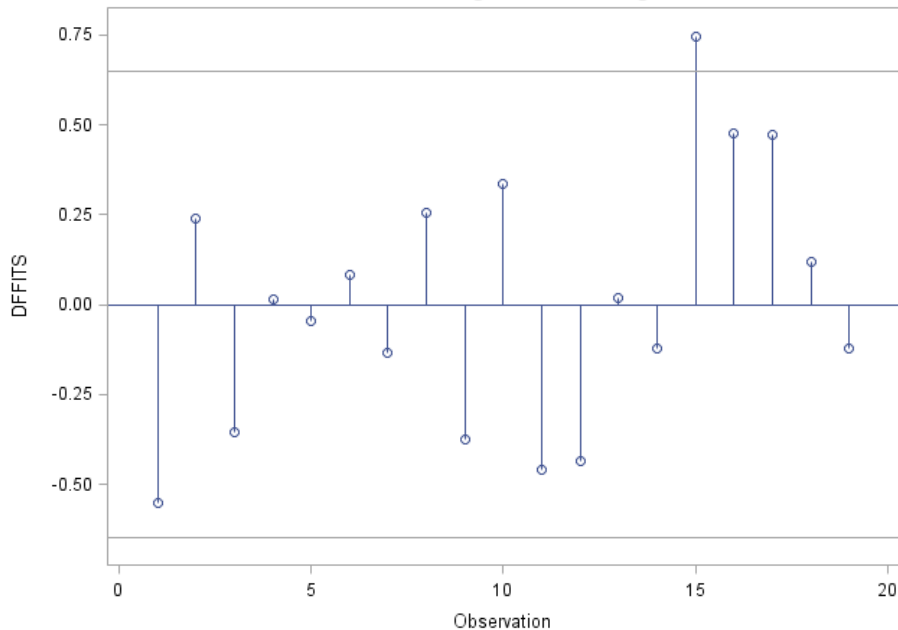
Gráficos

- ▶ Resíduos \times variáveis
- ▶ Resíduos \times valores ajustados
- ▶ Gráficos de índices (por obs)
- ▶ QQPLOT, Histograma e densidade dos resíduos

Influence Diagnostics for Weight



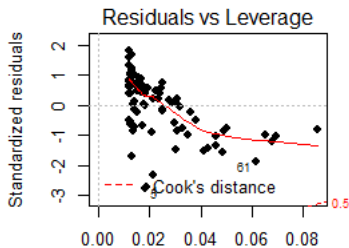
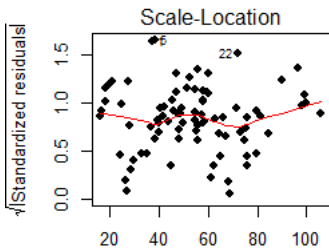
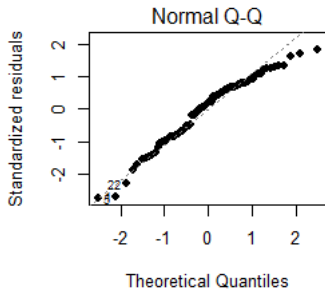
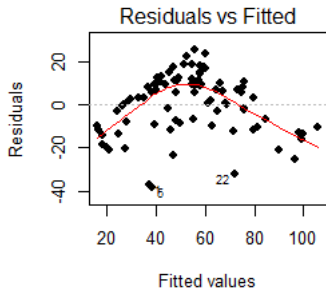
Influence Diagnostics for Weight

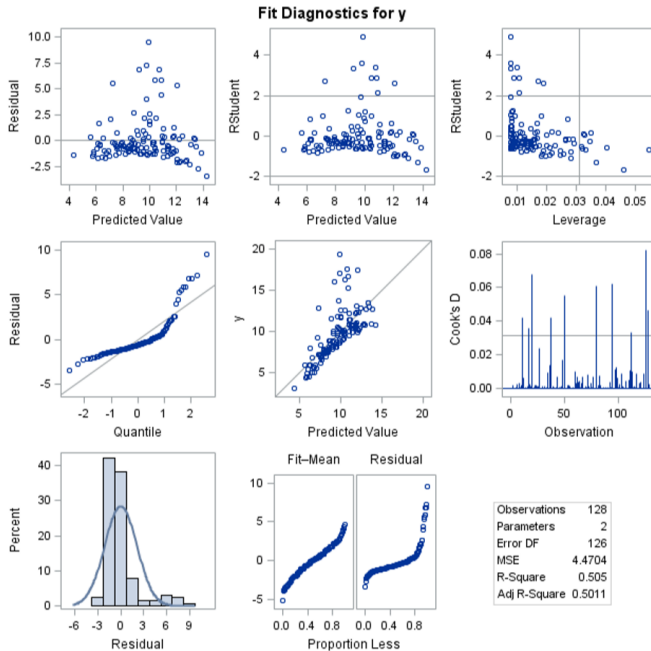


Diagnóstico em
modelagem

Alfredo Rossi e
Cayan

Diagnóstico em
modelagem





Aplicação no R



Diagnóstico em
modelagem

Alfredo Rossi e
Cayan

Diagnóstico em
modelagem