

Análise do ajuste do modelo de regressão L_1

Kévin Allan Sales Rodrigues

TESE APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO
DE
DOUTOR EM CIÊNCIAS

Programa: Probabilidade e Estatística
Orientadora: Prof^a. Dr^a. Silvia Nagib Elian

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da CAPES
e do CNPq

São Paulo, março de 2024

Análise do ajuste do modelo de regressão L_1

Esta versão da tese contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 14/03/2024. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof^ª. Dr^ª. Silvia Nagib Elian (orientadora) - IME-USP
- Prof. Dr. Gilberto Alvarenga Paula - IME-USP
- Prof. Dr. Gustavo Henrique de Araújo Pereira - UFSCAR
- Prof. Dr. Francisco José de Azevêdo Cysneiros - UFPE
- Prof^ª. Dr^ª. Reiko Aoki - ICMC-USP

Ficha catalográfica elaborada com dados inseridos pelo(a) autor(a)
Biblioteca Carlos Benjamin de Lyra
Instituto de Matemática e Estatística
Universidade de São Paulo

Rodrigues, Kévin Allan Sales

Análise do ajuste do modelo de regressão L1 / Kévin
Allan Sales Rodrigues; orientadora, Silvia Nagib
Elian. - São Paulo, 2024.
194 p.: il.

Tese (Doutorado) - Programa de Pós-Graduação em
Estatística / Instituto de Matemática e Estatística
/ Universidade de São Paulo.

Bibliografia

Versão corrigida

1. Regressão L1. 2. Regressão de mínima soma de
erros absolutos. 3. Medidas de Influência. 4. Métodos
de Diagnóstico. 5. Regressão Robusta. I. Elían,
Silvia Nagib. II. Título.

Bibliotecárias do Serviço de Informação e Biblioteca
Carlos Benjamin de Lyra do IME-USP, responsáveis pela
estrutura de catalogação da publicação de acordo com a AACR2:
Maria Lúcia Ribeiro CRB-8/2766; Stela do Nascimento Madruga CRB 8/7534.

Agradecimentos

À minha família, pelo apoio dado durante meu ensino básico e superior.

Os agradecimentos restantes serão divididos em três partes que correspondem às três etapas da minha formação: ensino técnico, ensino superior (graduação) e pós-graduação (mestrado e doutorado).

Parte 1 - Ensino Técnico

No período de 2010 a 2012, cursei ensino técnico na EEEP (Escola Estadual de Educação Profissional) Júlia Giffoni, onde tive a oportunidade de cursar o curso técnico integrado (inclui ensino médio). Lá tive os primeiros contatos com atividades extraclasse que acabaram por direcionar minha trajetória acadêmica, como: olimpíadas científicas, cursos extracurriculares, competições escolares de natureza acadêmica, estágio, etc.

Por isso agradeço a todos os docentes e funcionários EEEP Júlia Giffoni que tornaram isso possível. Destaco alguns profissionais da educação que me motivaram, e contribuíram fortemente para meu ingresso no ensino superior.

- Prof^a. Esp. Maria Cleiriane Bento dos Santos Monteiro: por ter sido uma excepcional professora de matemática e diretora de turma. Como professora de matemática, a Cleiriane incentivou-me a participar de olimpíadas científicas, em especial, a OBMEP. Esse incentivo foi fundamental para que meu interesse por matemática aumentasse consideravelmente. Como diretora de turma, senti muito empenho e dedicação emanados de sua pessoa (cada uma das 12 turmas possuía um(a) diretor(a) docente, esse papel demandava considerável esforço e o(a) diretor(a) de turma era o responsável imediato por todas as questões disciplinares). Mas, acima de tudo, por ter sido uma das primeiras pessoas a ver potencial em mim.
- Prof. Esp. André Wellington Soares de Lima: por ter me ensinado muito sobre a linguagem Java e programação orientada a objetos em geral, sobre banco de dados e também por ter sido excelente supervisor e orientador, respectivamente, durante meu estágio e TCC.
- Coordenadora Elisângela Alves Passos: por incentivar continuamente os estudantes a prestarem ENEM e vestibulares em geral, com a finalidade de ingressar no ensino superior. Naquela época isso foi muito importante, pois muitos estudantes não possuíam expectativas além de uma vaga no mercado de trabalho.

- Prof^a. Rejane Soares Silva: por incentivar os estudantes a cursarem graduação, por sempre demonstrar muito ânimo nas aulas e por fazer o melhor que podia, dadas as limitações do laboratório de ciências.

Além disso, agradeço aos docentes responsáveis pela pequena (e importante – para mim) biblioteca da escola e aos colegas de turma que tive.

Parte 2 - Ensino Superior

No período de 2013 a 2016, cursei bacharelado em estatística na UFC (Universidade Federal do Ceará). Na UFC aprendi probabilidade e estatística com os melhores docentes do Ceará.

Por isso agradeço a todos os docentes e funcionários do DEMA (Departamento de Estatística e Matemática Aplicada) da UFC, com destaque para algumas pessoas que são mencionadas a seguir.

Agradecimentos específicos:

- Ao meu orientador de TCC, Prof. Dr. José Ailton Alencar Andrade, pelo suporte no pouco tempo que lhe coube, pelas suas correções, incentivos e principalmente paciência durante a orientação do trabalho final de graduação.
- Ao Prof. Dr. Rafael Bráz Azevedo Farias, que me acompanhou durante minha trajetória na graduação, do primeiro ao quarto ano do curso, também foi o meu primeiro orientador do DEMA. E por todas as ajudas oferecidas. Até hoje, para mim, o Rafael é um modelo de docente a ser seguido e é uma das minhas principais referências de pessoa e profissional.
- Ao Prof. Dr. João Maurício Araújo Mota, não só por ser um excelente professor (o melhor na minha opinião) mas também por ser uma pessoa muito amigável, e em especial, pelo apoio, carinho, respeito e amizade, pelos puxões de orelha e principalmente pelos conselhos dados e pela paciência que teve comigo durante o curso.
- Ao Prof. Dr. Pushpa Narayan Rathie, por ter transmitido sua paixão pela pesquisa científica e por ter, efetivamente, ensinado-me as primeiras noções acerca de escrita científica, processo de pesquisa acadêmica na área de estatística, etc.
- À Erione, Márgeri e Luiza pelo excelente serviço prestado e por estarem sempre dispostas a me ajudar. Elas foram excelentes secretárias da coordenação do bacharelado em estatística e da secretaria do DEMA.

Adicionalmente, agradeço aos colegas de turma que tive ao decorrer dos quatro anos de graduação.

Parte 3 - Pós-Graduação

Agradeço à USP, e mais especificamente ao IME, pelo ambiente amigável e pela excelente estrutura oferecida. A todos os professores que participaram da minha formação a nível de pós-graduação e que me receberam tão bem neste programa.

Agradeço, especialmente, algumas pessoas que tiveram influência proeminente sobre minha trajetória pelo mestrado e doutorado. Tais personalidades são mencionadas a seguir.

Agradecimentos específicos:

- Um agradecimento extra-especial à minha orientadora, Prof^a. Dr^a. Silvia Nagib Elian, pelo conhecimento compartilhado, pelo suporte prestado, incentivos e principalmente pela paciência durante a orientação deste trabalho (e da minha dissertação). Com probabilidade 1, não teria sido possível concluir esta tese sem a presença da Silvia na minha formação a nível de doutorado.
- Ao Prof. Dr. Vanderlei da Costa Bueno, por ter sido meu orientador de programa durante o primeiro ano do doutorado, por ter sido supervisor dos meus estágios de docência. Mas principalmente, por ser uma grande referência, sobretudo, no campo da docência (fui monitor/estagiário de docência, ininterruptamente, do 2º semestre de 2017 ao 1º semestre de 2021 sob supervisão do Vanderlei – assim aprendi, acompanhando diversas turmas ao decorrer dos anos, como um professor organiza, prepara e executa suas atividades docentes).
- À Prof^a. Dr^a. Márcia D’Elia Branco, por ter sido minha primeira orientadora a nível de pós-graduação (orientadora de programa durante o primeiro ano do mestrado), pelas informações, conselhos e incentivos dados ao longo do ano de 2017.
- À Regiane pelo excelente serviço prestado na secretaria do programa e por estar sempre disposta a ajudar todos os alunos.

Agradeço também aos professores participantes da banca de qualificação de doutorado e das bancas examinadoras de mestrado e doutorado que proporcionaram-me o aprendizado necessário para progredir como pesquisador.

Ao CNPq (141836/2020-2 e 133167/2017-8) e à CAPES (88887.341938/2019-00) pelas bolsas concedidas durante a pós-graduação e graduação.

Enfim, agradeço a todos que contribuíram direta ou indiretamente para o meu desenvolvimento acadêmico, profissional e pessoal.

“A man who dares to waste one hour of time has not discovered the value of life.”

(Charles Darwin)

“Science is what we understand well enough to explain to a computer; art is everything else.”

(Donald E. Knuth)

“We must be careful not to confuse data with the abstractions we use to analyze them.”

(William James)

“An approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem.”

(John Tukey)

Resumo

RODRIGUES, K. A. S. **Análise do ajuste do modelo de regressão L_1** . 2024. 194 p. Tese (Doutorado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2024.

Este trabalho conta com um capítulo de revisão bibliográfica sobre regressão L_1 que inclui aspectos históricos, sua relação com outros métodos de regressão e seus resultados inferenciais: estimação pontual (e intervalar) e testes de hipóteses. O texto apresenta técnicas e métodos inéditos para análise do ajuste do modelo de regressão L_1 . São desenvolvidos, no contexto da regressão L_1 , testes paramétricos para verificar a hipótese de falta de ajuste e testes para verificar a hipótese de homogeneidade de variâncias. O texto detalha o desenvolvimento de uma medida de diagnóstico para regressão L_1 com base em verossimilhanças preditivas e divergência de Kullback-Leibler, detalhes técnicos para aplicação da medida diagnóstico, como processamento paralelo e ferramentas computacionais, são discutidos. A adaptação da técnica busca avante (*forward search*) ao contexto da regressão L_1 é concebida e apresentada. São realizadas aplicações a dados reais e estudos de simulação para ilustrar os métodos desenvolvidos ao decorrer do texto.

Palavras-chave: Regressão L_1 , Regressão de mínima soma de erros absolutos, Medidas de Influência, Métodos de Diagnóstico, Regressão Robusta.

Abstract

RODRIGUES, K. A. S. **Analysis of the adjustment of the L_1 regression model**. 2024. 194 p. Tese (Doutorado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2024.

This work has a literature review chapter on L_1 regression that includes historical aspects, its relationship with other regression methods and its inferential results: point (and interval) estimation and hypothesis tests. This text presents new techniques and methods for analyzing the fit of the L_1 regression model. In the context of L_1 regression, we developed parametric tests to verify the hypothesis of lack of fit and tests to verify the homogeneity of variances hypothesis. The text details the development of a diagnostic measure for L_1 regression based on predictive likelihoods and Kullback-Leibler divergence, and technical details for applying the diagnostic measure, such as parallel processing and computational tools, are discussed. The forward search technique adaptation to the context of L_1 regression is designed and presented. We made Applications to actual data and simulation studies to illustrate the methods developed throughout the text.

Keywords: L_1 Regression, LAD Regression, Influence Measures, Diagnostic methods, Robust Regression.

Sumário

Lista de Figuras	xiii
Lista de Tabelas	xvii
1 Regressão L_1	1
1.1 Introdução à Regressão L_1	3
1.2 Distribuição Laplace	5
1.3 Estimação dos parâmetros do modelo	7
1.3.1 Não unicidade e degeneração do estimador L_1	11
1.4 Inferência	13
1.4.1 Intervalos de confiança	13
1.4.2 Testes de hipóteses	15
1.5 Relações entre classes de modelos	15
1.6 Objetivo e organização do trabalho	19
2 Testes Paramétricos de Falta de Ajuste para Regressão L_1	21
2.1 Comentários sobre os testes	27
2.2 Estudo de simulação das estatísticas LF_1 e LF_2 em configurações com efeitos de grupo, polinomial e exponencial	28
2.3 Estudo de simulação das estatísticas LF_1 e LF_2 em configurações com efeito senoidal	31
2.4 Aplicação a dados reais	35
3 Teste de Homogeneidade de Variâncias para Regressão L_1	41
3.1 Estatística da Razão de Verossimilhanças	44
3.2 Estatísticas propostas para testar homogeneidade de variâncias	48
3.3 Detalhes da Simulação Balanceada	51
3.4 Detalhes da Simulação Desbalanceada	65
3.5 Detalhes da Simulação com diferentes números de grupos	72
4 Medidas de Diagnóstico com Base em Verossimilhanças Preditivas	81
4.1 Revisão bibliográfica sobre VP e DP	82
4.2 DPP no contexto da Regressão L_1	84

4.3	Função de Influência Preditiva a partir da DKL entre as DPPs	86
4.4	Não unicidade e degeneração do estimador L_1 e suas consequências na FVPP(z) numérica	89
4.5	Outras medidas de influência para comparação	90
4.6	Dados sobre bile litogênica	91
4.6.1	Aplicação da FIP aos dados sobre bile litogênica	95
4.6.2	Análise confirmatória	100
4.7	Dados sobre incêndios	101
4.7.1	Aplicação da FIP aos dados sobre incêndios	107
4.7.2	Análise confirmatória	112
4.8	Considerações sobre as aplicações	113
5	Técnica <i>Forward Search</i> para Regressão L_1	115
5.1	Técnica busca avante	116
5.2	Busca avante na regressão L_1	118
5.3	Aplicações	118
5.3.1	Dados sobre bile litogênica	118
5.3.2	Dados sobre incêndios	124
6	Conclusões e Trabalhos Futuros	131
A	Nomenclaturas para Regressão L_1	133
B	Norma L_k	135
C	Tabelas dos Estudos de Simulação do Capítulo 2	137
C.1	Tabelas do Estudo de simulação das estatísticas LF_1 e LF_2 em configurações com efeitos de grupo, polinomial e exponencial	137
C.2	Tabelas do Estudo de simulação das estatísticas LF_1 e LF_2 em configurações com efeito senoidal	146
D	Testes para Verificar a Suposição da Distribuição dos Erros	153
E	Ferramentas Computacionais e Processamento Paralelo	157
E.1	Processamento paralelo	157
E.2	Ferramentas computacionais	159
	Referências Bibliográficas	163

Lista de Figuras

1.1	Funções de distribuição da Laplace com parâmetros (0, 1), (1, 1) e (0, 2).	6
1.2	Gráfico com as distribuições Laplace(0, 1) e normal(0, 1).	7
1.3	Diagrama de Venn com relação entre as classes de modelos.	18
2.1	Gráfico dos dados simulados, a linha sólida em azul representa o modelo que gerou os dados e a linha em vermelho representa o modelo ajustado; réplica do tipo 1 com $\beta_2 = 0$.	31
3.1	Diagrama de dispersão entre coeficiente de variação (CV) e média da estatística LRT.	56
3.2	Diagrama de dispersão entre coeficiente de variação (CV) e taxa de rejeição da estatística LRT.	56
3.3	Histogramas LRT, Parte 1 da simulação, Cenários 1 a 12.	61
3.4	Histogramas LRT, Parte 1 da simulação, Cenários 13 a 24.	62
3.5	QQ-plots LRT, Parte 1 da simulação, Cenários 1 a 12.	63
3.6	QQ-plots LRT, Parte 1 da simulação, Cenários 13 a 24.	64
3.7	Histogramas LRT, Parte 2 da simulação, Cenários 1 a 12 (casos homocedásticos).	68
3.8	Histogramas LRT, Parte 2 da simulação, Cenários 13 a 24 (casos heterocedásticos).	69
3.9	QQ-plots LRT, Parte 2 da simulação, Cenários 1 a 12 (casos homocedásticos).	70
3.10	QQ-plots LRT, Parte 2 da simulação, Cenários 13 a 24 (casos heterocedásticos).	71
3.11	Histogramas LRT, Parte 3 da simulação, Cenários 1 a 12.	76
3.12	Histogramas LRT, Parte 3 da simulação, Cenários 13 a 24.	77
3.13	QQ-plots LRT, Parte 3 da simulação, Cenários 1 a 12.	78
3.14	QQ-plots LRT, Parte 3 da simulação, Cenários 13 a 24.	79
4.1	Gráfico da medida FIP (DKL entre $DPP(z)$ e $DPP_{(j)}(z)$) do modelo ajustado para variável resposta Concentração.	96

4.2 Gráfico da medida de influência baseada na norma L_1 do modelo ajustado para variável resposta Concentração. 97

4.3 Gráfico da medida de influência baseada na norma L_2 do modelo ajustado para variável resposta Concentração. 97

4.4 Gráfico da distância de Cook do modelo ajustado para variável resposta Concentração. 97

4.5 Gráfico da medida afastamento condicional de verossimilhanças do modelo ajustado para variável resposta Concentração. 98

4.6 Gráfico da medida afastamento de verossimilhanças do modelo ajustado para variável resposta Concentração. 98

4.7 Diagrama de dispersão entre as variáveis Concentração e Idade, a reta em vermelho foi construída a partir do modelo ajustado. 99

4.8 Gráfico da medida FIP (DKL entre $DPP(z)$ e $DPP_{(j)}(z)$) do modelo ajustado para variável resposta $\log(\text{Incêndio})$ por meio da Equação (4.4). 109

4.9 Gráfico da medida de influência baseada na norma L_1 do modelo ajustado para variável resposta $\log(\text{Incêndio})$ por meio da Equação (4.4). 109

4.10 Gráfico da medida de influência baseada na norma L_2 do modelo ajustado para variável resposta $\log(\text{Incêndio})$ por meio da Equação (4.4). 110

4.11 Gráfico da distância de Cook do modelo ajustado para variável resposta $\log(\text{Incêndio})$ por meio da Equação (4.4). 110

4.12 Gráfico da medida afastamento condicional de verossimilhanças do modelo ajustado para variável resposta $\log(\text{Incêndio})$ por meio da Equação (4.4). 110

4.13 Gráfico da medida afastamento de verossimilhanças do modelo ajustado para variável resposta $\log(\text{Incêndio})$ por meio da Equação (4.4). 111

4.14 Gráfico de dispersão entre as variáveis Roubo e Renda. 113

5.1 Gráfico de busca avante dos resíduos L_1 padronizados pela estimativa final $\hat{\lambda}_{MLE}$ para o modelo ajustado para os dados sobre bile por meio da Equação (4.2). 119

5.2 Gráfico de busca avante das estimativas dos parâmetros de posição para o modelo ajustado para os dados sobre bile por meio da Equação (4.2). 119

5.3 Gráfico de busca avante das estatísticas de teste de significância dos parâmetros de posição para o modelo ajustado para os dados sobre bile por meio da Equação (4.2). 122

5.4 Gráfico de busca avante da estimativa $\hat{\lambda}_{MLE}$ para o modelo ajustado para os dados sobre bile por meio da Equação (4.2). 122

5.5 Gráfico de busca avante dos resíduos L_1 padronizados pela estimativa final $\hat{\lambda}_{MLE}$ 124

5.6	Gráfico de busca avante das estimativas dos parâmetros de posição para o modelo ajustado para os dados sobre incêndios por meio da Equação (4.4).	126
5.7	Gráfico de busca avante das estatísticas de teste de significância dos parâmetros de posição para o modelo ajustado para os dados sobre incêndios por meio da Equação (4.4).	128
5.8	Gráfico de busca avante da estimativa $\hat{\lambda}_{MLE}$ para o modelo ajustado para os dados sobre incêndios por meio da Equação (4.4).	129

Lista de Tabelas

2.1	Configurações da simulação.	29
2.2	Taxas de rejeição da estatística $LF_1(\hat{\lambda}_{MLE})$ quando $\lambda = 10$ para as oito configurações.	30
2.3	Dados da simulação.	32
2.4	Taxas de rejeição da estatística $LF_1(\hat{\lambda}_{MLE})$ quando $\beta_2 = 0$ para os cinco tipos de réplicas.	34
2.5	Taxas de rejeição da estatística $LF_1(\hat{\lambda}_{MLE})$ quando $\beta_2 = 0,8$ para os cinco tipos de réplicas.	34
2.6	Taxas de rejeição da estatística $LF_1(\hat{\lambda}_{MLE})$ quando $\beta_2 = 1,6$ para os cinco tipos de réplicas.	35
2.7	Dados sobre imóveis já classificado por Su e Yang (2006)	36
2.8	Dados sobre imóveis de Su e Yang (2006) com agrupamento considerando cada variável explicativa isoladamente.	37
2.9	Médias das variáveis explicativas X_1 e X_2 dentro dos grupos formados pelos comandos apresentados.	38
3.1	Dados da simulação balanceada para 24 cenários, com 5 grupos contendo 20 observações por grupo, totalizando 100 observações ($N = 100$). Foram geradas 1000 réplicas por cenário.	52
3.2	LRT – Resultado da simulação balanceada para 24 cenários, com 5 grupos contendo 20 observações por grupo, totalizando 100 observações ($N = 100$).	53
3.3	$\max\text{-}\hat{\lambda}$ e $\max\text{-}\log\text{-}\hat{\lambda}$ – Resultado da simulação balanceada para 24 cenários, com 5 grupos contendo 20 observações por grupo, totalizando 100 observações ($N = 100$).	54
3.4	Resultado da simulação balanceada, relação entre coeficiente de variação (CV) e estatística LRT, incluindo média e taxa de rejeição. A taxa de rejeição foi calculada a partir de dois valores críticos: um obtido pela distribuição $\chi^2(4)$ (distribuição assintótica da estatística sugerida ao fim das simulações) e outro via <i>bootstrap</i>	55

3.5	Comparativo entre as taxas de rejeição das estatísticas LRT, $\max-\hat{\lambda}$ e $\max-\log-\hat{\lambda}$ na simulação balanceada.	58
3.6	Dados da simulação desbalanceada para 24 cenários, com 5 grupos de diferentes tamanhos amostrais, totalizando 100 observações ($N = 100$). Foram geradas 1000 réplicas por cenário.	66
3.7	LRT e $\max-\hat{\lambda}$ – Resultado da simulação desbalanceada para 24 cenários, com 5 grupos de diferentes tamanhos amostrais, totalizando 100 observações ($N = 100$).	67
3.8	Dados da simulação para 24 cenários com diferentes números de grupos. Cada grupo contém o mesmo número de observações, 120/número de grupos, totalizando 120 observações ($N = 120$). Foram geradas 1000 réplicas por cenário.	74
3.9	LRT e $\max-\hat{\lambda}$ – Resultado da simulação para 24 cenários com diferentes números de grupos. Cada grupo contém o mesmo número de observações, 120/número de grupos, totalizando 120 observações ($N = 120$).	75
4.1	Dados sobre bile litogênica.	92
4.2	Estatísticas descritivas dos dados sobre bile litogênica.	93
4.3	Estimativas, erros padrão das estimativas e valores-P no modelo ajustado para variável resposta Concentração.	93
4.4	Variável (resposta) Concentração, valores preditos e resíduos obtidos via regressão L_1	94
4.5	FIP aplicada aos dados sobre bile litogênica.	95
4.6	Observações influentes, por meio das demais medidas de influência, no modelo ajustado para a variável Concentração.	96
4.7	Observações influentes, pela FIP, no modelo ajustado para a variável Concentração.	100
4.8	Análise confirmatória para modelo ajustado com dados sobre bile litogênica.	101
4.9	Estatísticas descritivas dos dados sobre incêndios.	102
4.10	Matriz com as correlações entre as variáveis $\log(\text{Incêndio})$, Idade, Roubo e Renda dos dados sobre incêndios.	102
4.11	Dados sobre incêndios.	103
4.12	Estimativas, erros padrão das estimativas e valores-P no modelo ajustado para variável resposta $\log(\text{Incêndio})$ utilizando as variáveis Idade, Roubo e Renda.	104
4.13	Estimativas, erros padrão das estimativas e valores-P no modelo ajustado para variável resposta $\log(\text{Incêndio})$ utilizando as variáveis Roubo e Renda.	104
4.14	Variável (resposta) $\log(\text{Incêndio})$, valores preditos e resíduos obtidos via regressão L_1	106
4.15	FIP aplicada aos dados sobre incêndios.	107

4.16	Observações influentes, por meio das demais medidas de influência, no modelo ajustado para a variável $\log(\text{Incêndio})$ por meio da Equação (4.4).	108
4.17	Observações influentes, destacadas pela FIP e outras medidas de influência, no modelo ajustado para a variável $\log(\text{Incêndio})$ por meio da Equação (4.4).	111
4.18	Análise confirmatória para modelo ajustado com dados sobre incêndio.	113
5.1	Resultados da aplicação da técnica busca avante ao modelo ajustado aos dados sobre bile litogênica, contendo ordem de inclusão das observações no modelo, índices das observações incluídas e estimativas dos parâmetros de posição e de escala em cada inclusão.	120
5.2	Análise confirmatória por meio da busca avante para modelo ajustado com dados sobre bile litogênica.	123
5.3	Resultados da aplicação da técnica busca avante ao modelo ajustado aos dados sobre incêndios, contendo a ordem de inclusão das observações no modelo, índices das observações incluídas e estimativas dos parâmetros de posição e de escala em cada inclusão.	125
5.4	Análise confirmatória por meio da busca avante para modelo ajustado com dados sobre incêndios.	130
C.1	Simulação com $\lambda = 1, 3, 5, 10, 20$ considerando a configuração 1.	138
C.2	Simulação com $\lambda = 1, 3, 5, 10, 20$ considerando a configuração 2.	139
C.3	Simulação com $\lambda = 1, 3, 5, 10, 20$ considerando a configuração 3.	140
C.4	Simulação com $\lambda = 1, 3, 5, 10, 20$ considerando a configuração 4.	141
C.5	Simulação com $\lambda = 1, 3, 5, 10, 20$ considerando a configuração 5.	142
C.6	Simulação com $\lambda = 1, 3, 5, 10, 20$ considerando a configuração 6.	143
C.7	Simulação com $\lambda = 1, 3, 5, 10, 20$ considerando a configuração 7.	144
C.8	Simulação com $\lambda = 1, 3, 5, 10, 20$ considerando a configuração 8.	145
C.9	$\beta_2 = 0$ com 5 réplicas para cada valor da variável explicativa.	146
C.10	$\beta_2 = 0, 8$ com 5 réplicas para cada valor da variável explicativa.	146
C.11	$\beta_2 = 1, 6$ com 5 réplicas para cada valor da variável explicativa.	146
C.12	$\beta_2 = 2, 4$ com 5 réplicas para cada valor da variável explicativa.	147
C.13	$\beta_2 = 3, 2$ com 5 réplicas para cada valor da variável explicativa.	147
C.14	$\beta_2 = 0$ com 10 réplicas para cada valor da variável explicativa.	147
C.15	$\beta_2 = 0, 8$ com 10 réplicas para cada valor da variável explicativa.	147
C.16	$\beta_2 = 1, 6$ com 10 réplicas para cada valor da variável explicativa.	148
C.17	$\beta_2 = 2, 4$ com 10 réplicas para cada valor da variável explicativa.	148
C.18	$\beta_2 = 3, 2$ com 10 réplicas para cada valor da variável explicativa.	148
C.19	$\beta_2 = 0$ com 15 réplicas para cada valor da variável explicativa.	148
C.20	$\beta_2 = 0, 8$ com 15 réplicas para cada valor da variável explicativa.	149
C.21	$\beta_2 = 1, 6$ com 15 réplicas para cada valor da variável explicativa.	149

C.22 $\beta_2 = 2, 4$ com 15 réplicas para cada valor da variável explicativa.	149
C.23 $\beta_2 = 3, 2$ com 15 réplicas para cada valor da variável explicativa.	149
C.24 $\beta_2 = 0$ com 20 réplicas para cada valor da variável explicativa.	150
C.25 $\beta_2 = 0, 8$ com 20 réplicas para cada valor da variável explicativa.	150
C.26 $\beta_2 = 1, 6$ com 20 réplicas para cada valor da variável explicativa.	150
C.27 $\beta_2 = 2, 4$ com 20 réplicas para cada valor da variável explicativa.	150
C.28 $\beta_2 = 3, 2$ com 20 réplicas para cada valor da variável explicativa.	151
C.29 $\beta_2 = 0$ com 30 réplicas para cada valor da variável explicativa.	151
C.30 $\beta_2 = 0, 8$ com 30 réplicas para cada valor da variável explicativa.	151
C.31 $\beta_2 = 1, 6$ com 30 réplicas para cada valor da variável explicativa.	151
C.32 $\beta_2 = 2, 4$ com 30 réplicas para cada valor da variável explicativa.	152
C.33 $\beta_2 = 3, 2$ com 30 réplicas para cada valor da variável explicativa.	152
D.1 Pontos críticos da estatística U^2 (Watson).	155
D.2 Pontos críticos da estatística W^2 (Cramér-von Mises).	155
D.3 Pontos críticos da estatística A^2 (Anderson-Darling).	155
D.4 Pontos críticos da estatística $\sqrt{n}D$ (Kolmogorov-Smirnov).	156
D.5 Pontos críticos da estatística $\sqrt{n}V$ (Kuiper).	156

Capítulo 1

Regressão L_1

Neste capítulo descreveremos a regressão L_1 , incluindo alguns aspectos históricos, sua relação com outros métodos de regressão e seus resultados inferenciais: estimação intervalar e testes de hipóteses – enfim, trata-se de uma introdução para que o leitor habitue-se com os conceitos relativos à regressão L_1 .

Na extensa literatura estatística, encontramos vários estudos sobre regressão linear usando o método dos mínimos quadrados para estimativa de parâmetros conforme proposto por [Legendre \(1805\)](#). O método dos mínimos quadrados fornece resultados ótimos sob certas condições, como por exemplo: homocedasticidade, erros com caudas mesocúrticas, etc. No entanto, há casos em que essas condições não são atendidas. Por exemplo: se os erros não forem normalmente distribuídos, tiverem caudas pesadas ou valores discrepantes (*outliers*), a regressão de mínimos quadrados é insatisfatória nessas situações. Para este último caso, presença de *outliers*, precisamos de procedimentos robustos que levem em conta a presença desses *outliers* ou de técnicas que possam identificá-los. Segundo [Rousseeuw e Leroy \(2003\)](#), a regressão L_1 é uma boa alternativa robusta ao método de mínimos quadrados, mesmo quando as suposições necessárias à aplicação de mínimos quadrados são válidas.

A regressão L_1 é recomendada quando a distribuição de erros possui caudas pesadas ([Bloomfield e Steiger, 1983](#)). Vários autores ([Birkes e Dodge, 1993](#); [Bloomfield e Steiger, 1983](#); [Rousseeuw e Leroy, 2003](#)) consideram-na uma boa alternativa robusta à regressão via mínimos quadrados. Isso ocorre porque a regressão L_1 é menos sensível a observações com resíduos absolutos altos do que a regressão via mínimos quadrados.

Um fato histórico relevante é que o estimador L_1 foi proposto mais de 50 anos antes que o estimador de mínimos quadrados. Este método foi introduzido por [Boscovich \(1757\)](#) no século XVIII, mas evoluiu muito lentamente em comparação com a regressão de mínimos quadrados. Mais tarde, [Boscovich \(1760\)](#) desenvolveu um algoritmo geométrico para obter as estimativas dos coeficientes de regressão L_1 . Uma formulação algébrica do algoritmo foi desenvolvida por [Laplace \(1793\)](#). A regressão L_1 permaneceu adormecida por muitas décadas, por razões que serão explicadas mais adiante.

O trabalho de [Edgeworth \(1887\)](#) reformulou o problema proposto por [Boscovich \(1757\)](#), reintroduzindo o problema de regressão L_1 . [Edgeworth \(1888\)](#) discutiu sobre a não unicidade da solução da regressão L_1 , além de contribuir na demonstração de que os estimadores L_1 são equivalentes aos de máxima verossimilhança quando os erros têm distribuição Laplace, e na elaboração de um algoritmo para determinar a reta da regressão L_1 .

Além dos obstáculos computacionais, a falta de ferramentas de inferência estatística provou ser um problema para o progresso e aplicação da regressão L_1 em comparação com a regressão de mínimos quadrados. Essa diferença no ferramental inferencial é perceptível até hoje, pois enquanto a regressão via mínimos quadrados fornece resultados analíticos exatos a regressão L_1 limita-se a fornecer resultados inferenciais assintóticos. Contudo, há mais de 30 anos, diversos estudos de Monte Carlo foram desenvolvidos sobre propriedades dos estimadores para pequenas amostras; resultados assintóticos sobre a distribuição dos estimadores e procedimentos para inferência também foram apresentados; algumas aplicações da regressão L_1 são: estimação de custos no setor de transportes ([Wilson, 1979](#)), reconhecimento de padrão espectral de massa ([Fausett e Weber, 1978](#)), estimação de parâmetros farmacocinéticos ([Bezerra, 1990](#)), astronomia ([Dodge, 1997](#)), modelagem de diversos dados geofísicos ([Afifah et al., 2017](#)), estimação de parâmetros econômicos ([Yoo, 2001](#)), etc.

Até aqui tecemos breves comentários sobre a história da regressão L_1 , mas para uma abordagem mais completa da história e desenvolvimento da regressão L_1 recomendamos a leitura dos clássicos de [Harter \(1974a,b, 1975a,b,c,d, 1976\)](#) – que fornecem uma ampla revisão histórica e bibliográfica sobre regressão de mínimos quadrados, L_1 e L_∞ ([Farebrother, 1985](#)) incluindo suas similaridades e divergências. [Kuks e Olman \(1971, 1972\)](#) e [Fletcher et al. \(1974\)](#) também abordam regressão L_∞ . [Dodge \(2002\)](#) e [Farebrother \(1990, 1997\)](#) fornecem aspectos históricos mais recentes não cobertos por Harter. Uma revisão abrangente dos avanços mais recentes sobre o método de mínima soma dos erros absolutos é apresentada por [Dielman \(1984, 2005\)](#).

Os livros de [Bloomfield e Steiger \(1983\)](#) (*Least Absolute Deviations: Theory, Applications, and Algorithms*), [Birkes e Dodge \(1993\)](#) (*Alternative Methods of Regression*) e [Rousseeuw e Leroy \(2003\)](#) (*Robust Regression and Outlier Detection*) são ótimos para o estudo inicial da regressão L_1 pois apresentam as propriedades, inferência e algoritmos da regressão de mínima soma dos erros absolutos, além de fornecer o contexto histórico do desenvolvimento deste método de regressão. Em especial, a segunda referência, [Birkes e Dodge \(1993\)](#), apresenta uma comparação bem abrangente entre os mais diversos métodos de regressão, incluindo: mínimos quadrados, mínima soma de erros absolutos, regressão M, regressão não paramétrica, regressão *Ridge* e Regressão Bayesiana.

É oportuno comentar que na literatura a regressão L_1 é tratada por diversos nomes, para conhecer as várias nomenclaturas da regressão L_1 consulte o Apêndice A. Ao de-

correr deste capítulo faremos uma breve introdução à regressão L_1 na Seção 1.1, revisaremos as principais propriedades e características da distribuição Laplace na Seção 1.2 e na Seção 1.3 apresentaremos os estimadores pontuais do modelo de regressão L_1 – incluindo uma discussão abrangente sobre os casos em que a estimativa L_1 não é única ou é degenerada. A Seção 1.4 tratará de inferência no contexto da regressão L_1 abordando estimativas intervalares, via intervalos de confiança assintóticos, e testes de hipóteses. Para situar onde o modelo linear Laplace se encaixa com relação às classes de modelos quantílicos, lineares generalizados e lineares elípticos elaboramos a Seção 1.5. Por fim, encerraremos o capítulo com o objetivo desta tese e com uma descrição de cada um dos capítulos subsequentes.

1.1 Introdução à Regressão L_1

Um dos tipos de modelos de regressão mais difundidos é o modelo linear, neste modelo assume-se que há uma relação linear entre a variável resposta e as variáveis explicativas. A forma do modelo linear é dada por

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_p X_{pi} + \epsilon_i, \quad i = 1, \dots, n \text{ e } p + 1 \leq n \quad (1.1)$$

em que Y_i é a variável resposta relativa à i -ésima observação, X_{ki} denota o valor da k -ésima variável explicativa da i -ésima observação, β_0 é o intercepto do modelo, β_k é o parâmetro associado à k -ésima variável explicativa e ϵ_i é o erro relativo à i -ésima observação. Quando o modelo (1.1) só possui uma variável explicativa, isto é, $p = 1$, é denominado modelo de regressão linear simples.

Vale salientar que na regressão L_1 a matriz \mathbf{X} não precisa ter posto completo como no caso da estimação via mínimos quadrados. Podemos expressar este mesmo modelo na forma matricial, isto é,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

em que $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^\top$ é o vetor de variáveis resposta com dimensão $n \times 1$, $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top, \dots, \mathbf{X}_n^\top)^\top$, em que $\mathbf{X}_i = (1, X_{1i}, X_{2i}, \dots, X_{pi})$ é o vetor com as p variáveis explicativas associadas à i -ésima observação, é a matriz de especificação ou de planejamento do modelo, que contém os valores das variáveis explicativas para as n observações, com dimensão $n \times (p + 1)$ no modelo com intercepto (com dimensão $n \times p$ no modelo sem intercepto), $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^\top$ é o vetor com $p + 1$ parâmetros de posição do modelo de regressão e $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)^\top$ é o vetor $n \times 1$ dos erros aleatórios.

Geralmente também se supõe a homoscedasticidade dos erros (igualdade das variâncias dos erros). Além da homoscedasticidade, também adicionaremos a suposição de que os erros aleatórios são independentes e identicamente distribuídas com distribuição Laplace(0, λ). Note que λ é o parâmetro de escala do modelo de regressão, veja a Seção

1.5 para mais detalhes.

Existem várias formas de estimar os parâmetros de posição e escala de um modelo linear. Dentre os métodos de estimação se destacam o de mínimos quadrados, máxima verossimilhança e mínima soma dos erros absolutos. Em geral, estes estimadores são distintos e só coincidem em casos bem particulares. Por exemplo, os estimadores de mínimos quadrados coincidem com os estimadores de máxima verossimilhança quando os erros têm distribuição normal e os estimadores de mínima soma dos erros absolutos coincidem com os estimadores de máxima verossimilhança quando os erros têm distribuição exponencial dupla (Laplace).

Cada método de estimação tem suas peculiaridades e suposições próprias e consequentemente propriedades particulares. Por exemplo, o método de máxima verossimilhança fornece propriedades assintóticas úteis (sob as famosas condições de regularidade) mas requer a especificação da distribuição dos erros, ϵ_i , e além disso, requer a maximização da função de verossimilhança.

O método de mínimos quadrados tem boas propriedades: estimadores não viesados, fácil resolução (com a suposição de posto completo e homoscedasticidade), não requer a definição da distribuição dos erros e foi amplamente estudado. Mas em geral, a solução de mínimos quadrados é sensível a observações aberrantes. Observações aberrantes podem gerar grandes ou pequenos resíduos, em valor absoluto. No caso em que observações aberrantes geram resíduos absolutos pequenos, tais observações deslocam a reta de regressão para próximo de si, influenciando as estimativas dos parâmetros de posição do modelo e possivelmente alterando a inferência realizada a partir do modelo ajustado. Já as estimativas via método de mínima soma dos erros absolutos não sofrem desse problema mas em contrapartida, em geral, requerem solução via algoritmos de programação linear.

Além das diferenças nas suposições, no nível de dificuldade de cálculo e propriedades dos estimadores obtidos por estes métodos de estimação, existem as diferenças na interpretação das estimativas dos parâmetros do modelo. Por exemplo, é bem conhecida a interpretação dos $\hat{\beta}_i$'s no modelo ajustado via mínimos quadrados, estima-se que um incremento de uma unidade na variável explicativa X_i gera um incremento $\hat{\beta}_i$ (ou decremento se $\hat{\beta}_i$ for negativo) no valor esperado da variável resposta se as demais covariáveis forem mantidas fixadas. Isto é, $\hat{\beta}_i$ é a estimativa da taxa de variação na média da variável resposta por unidade da i -ésima variável explicativa. De modo análogo, utilizando o estimador de mínima soma de erros absolutos, $\hat{\beta}_i$ é a estimativa da taxa de variação na mediana da variável resposta por unidade da i -ésima variável explicativa, já que, conforme veremos posteriormente, o estimador de β na regressão L_1 coincide com o correspondente estimador no modelo de regressão quantílica para o quantil de ordem 0,5.

1.2 Distribuição Laplace

De acordo com [Narula e Wellington \(1977, 1985\)](#) o método de regressão L_1 oferece uma alternativa robusta ao procedimento de mínimos quadrados quando os erros têm distribuição Laplace. [Rodrigues \(2019\)](#) aborda testes para verificar se a suposição dos erros seguirem distribuição Laplace é atendida. O pacote *diagL1* ([Rodrigues e Elian, 2024](#)) contém a função `laplace.dist.test()`, desenvolvida com base na função `laplace.test()` do pacote *lawstat* ([Gastwirth et al., 2023](#)), que calcula cinco estatísticas para verificar se os erros possuem distribuição Laplace. Utilizando os resultados revisados por [Rodrigues \(2019\)](#) e suporte computacional fornecido pelo *diagL1* é possível checar a suposição acerca dos erros do modelo sem grande dificuldade.

A função densidade de probabilidade (FDP) da distribuição Laplace é dada por

$$f(x) = \frac{1}{2\lambda} e^{-\frac{|x-\mu|}{\lambda}} \mathbb{I}_{\mathbb{R}}(x), \quad \mu \in \mathbb{R} \text{ e } \lambda > 0. \quad (1.2)$$

A distribuição Laplace é também conhecida como exponencial dupla devido ao fato de que essa distribuição é ou pode ser obtida como a diferença entre duas variáveis aleatórias independentes com distribuição exponencial com parâmetro $1/\lambda$. Além disso, se W tem distribuição Laplace, $|W|$ tem distribuição exponencial. Dentre as principais características da Laplace destacamos: simetria em torno de μ , média, moda e mediana iguais a μ , variância dada por $2\lambda^2$ e curtose 6, ou seja, é leptocúrtica: distribuições leptocúrticas são distribuições cujas caudas são mais pesadas que as da distribuição normal. Sua função geradora de momentos e função distribuição acumulada são dadas, respectivamente, por

$$M_X(t) = \frac{e^{\mu t}}{1 - (\lambda t)^2}, \quad |t| < \frac{1}{\lambda}$$

e

$$F(x) = \begin{cases} \frac{1}{2} e^{\frac{x-\mu}{\lambda}}, & \text{se } x \leq \mu \\ 1 - \frac{1}{2} e^{-\frac{x-\mu}{\lambda}}, & \text{se } x > \mu \end{cases}.$$

Talvez a característica mais interessante da distribuição Laplace seja a interpretação especial de seu parâmetro λ , que é o desvio absoluto médio da distribuição – o desvio absoluto médio de uma variável aleatória X é definido por $\mathbb{E}[|X - \mathbb{E}[X]|]$. Os cálculos a seguir tornam esta afirmação mais clara.

$$\mathbb{E}[|X - \mathbb{E}[X]|] = \mathbb{E}[|X - \mu|] = \int_{-\infty}^{+\infty} |x - \mu| \frac{1}{2\lambda} e^{-\frac{|x-\mu|}{\lambda}} dx$$

fazendo a substituição $y = x - \mu$ ($dy = dx$) temos

$$\int_{-\infty}^{+\infty} |y| \frac{1}{2\lambda} e^{-\frac{|y|}{\lambda}} dy = \frac{1}{2\lambda} \int_{-\infty}^{+\infty} |y| e^{-\frac{|y|}{\lambda}} dy = \frac{2}{2\lambda} \int_0^{+\infty} y e^{-\frac{y}{\lambda}} dy,$$

a última igualdade decorre do fato do integrando ser função par, fazendo a substituição $w = \frac{y}{\lambda}$ ($dw = \frac{1}{\lambda} dy$) temos

$$\frac{1}{\lambda} \int_0^{+\infty} w \lambda e^{-w} \lambda dw = \frac{\lambda^2}{\lambda} \underbrace{\int_0^{+\infty} w e^{-w} dw}_{=1} = \lambda,$$

a última igualdade decorre do fato da integral da função densidade de probabilidade de uma variável aleatória com distribuição exponencial sobre seu suporte ser igual a 1. Assim, mostramos efetivamente que λ é o desvio absoluto médio da distribuição Laplace. Esse resultado, em algum sentido, lembra o resultado de que o parâmetro σ^2 da distribuição normal é interpretado como variância da distribuição normal. Isso ocorre pelo fato de tanto a variância quanto o desvio absoluto médio serem medidas de dispersão estatística.

A Figura 1.1 apresenta os gráficos das funções distribuição acumulada das distribuições Laplace(0,1), Laplace(1,1) e Laplace(0,2). O parâmetro de posição, μ , desloca o gráfico à direita (se $\mu > 0$) ou à esquerda (se $\mu < 0$).

Laplace(μ, λ) representa a distribuição Laplace com parâmetros μ e λ conforme Equação (1.2), esta notação é usada no texto e na Figura 1.1.

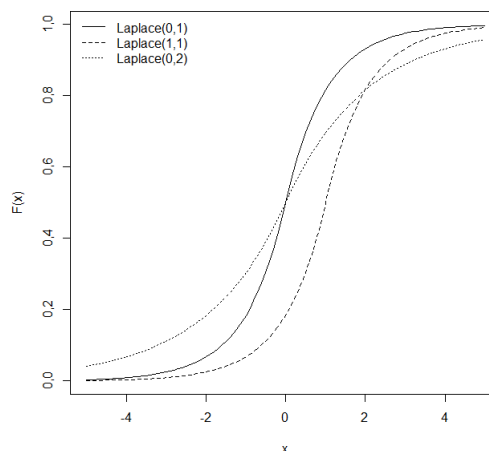


Figura 1.1: Funções de distribuição da Laplace com parâmetros (0, 1), (1, 1) e (0, 2).

A Figura 1.2 apresenta as funções densidade de probabilidade das distribuições Laplace e normal nas suas respectivas formas padrão. A diferença entre estas distribuições é nítida, principalmente no ponto zero, que é a média de ambas, sendo que a densidade da distribuição Laplace apresenta um pico e a densidade da normal uma curva suave. Também nota-se com clareza que as caudas da Laplace têm um decaimento mais lento que as caudas da normal, o que faz com que a distribuição Laplace tenha caudas mais pesadas que a distribuição normal.

Do ponto de vista inferencial, um fato importante é que os estimadores de máxima verossimilhança dos parâmetros de posição, μ , e escala, λ , da Laplace são dados, respectivamente, por

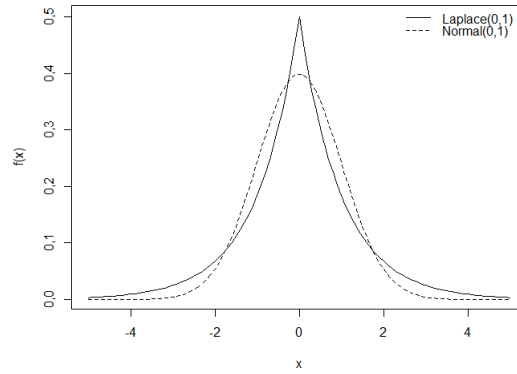


Figura 1.2: Gráfico com as distribuições $Laplace(0, 1)$ e $normal(0, 1)$.

$$\hat{\mu} = \begin{cases} X_{(\frac{n+1}{2})}, & \text{se } n \text{ é ímpar} \\ \frac{X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}}{2}, & \text{se } n \text{ é par} \end{cases}$$

e

$$\hat{\lambda} = \sum_{i=1}^n \frac{|X_i - \hat{\mu}|}{n},$$

em que $X_{(\frac{n+1}{2})}$, $X_{(\frac{n}{2})}$ e $X_{(\frac{n}{2}+1)}$ denotam as estatísticas de ordem $\frac{n+1}{2}$, $\frac{n}{2}$ e $\frac{n}{2} + 1$, respectivamente. Fica evidente que $\hat{\mu}$ é simplesmente a mediana amostral das observações da variável aleatória X .

1.3 Estimação dos parâmetros do modelo

Antes de efetivamente introduzirmos o problema de estimação dos parâmetros do modelo de regressão pelo procedimento de mínima soma dos erros absolutos, trataremos brevemente do modelo de regressão quantílico que foi proposto por [Koenker e Bassett \(1978\)](#). Conforme comentado anteriormente, o modelo de regressão L_1 pode ser visto como um caso particular do modelo quantílico pois é o caso quando o quantil de interesse é de ordem 0,5, isto é, a mediana.

Seja $Y_i, i = 1, \dots, n$ uma amostra aleatória da variável aleatória Y , com distribuição simétrica em torno de $\beta(\theta)$. Para $0 < \theta < 1$, o θ -ésimo quantil amostral pode ser definido como solução do problema de minimização dado por

$$\min_{\beta(\theta) \in \mathbb{R}} \left[\sum_{i \in \{i: Y_i \geq \beta(\theta)\}} \theta |Y_i - \beta(\theta)| + \sum_{i \in \{i: Y_i < \beta(\theta)\}} (1 - \theta) |Y_i - \beta(\theta)| \right], \theta \in (0, 1).$$

Seja $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$ o vetor de variáveis resposta do modelo linear, em que Y_i é dado pela Equação (1.1). O estimador de $\beta(\theta)$ no modelo de regressão quantílica de ordem θ é qualquer solução em $\hat{\beta}(\theta)$ do problema de minimização (1.3)

$$\min_{\beta(\theta) \in \mathbb{R}^{\rho+1}} \left[\sum_{i \in \{i: Y_i \geq \mathbf{X}_i \beta(\theta)\}} \theta |Y_i - \mathbf{X}_i \beta(\theta)| + \sum_{i \in \{i: Y_i < \mathbf{X}_i \beta(\theta)\}} (1 - \theta) |Y_i - \mathbf{X}_i \beta(\theta)| \right], \theta \in (0, 1), \quad (1.3)$$

em que \mathbf{X}_i é a i -ésima linha da matriz \mathbf{X} .

O parâmetro θ pondera a regressão quantílica, ou seja, é tal que pelo menos de $100\theta\%$ dos valores dos Y_i encontram-se acima do hiperplano de regressão quantílica e pelo menos $100(1 - \theta)\%$ abaixo.

Quando $\theta = 1/2$, $\hat{\beta}(\theta) = \hat{\beta}(0, 5)$ é o estimador da regressão L_1 , que é um caso particular importante da regressão quantílica, tendo como medida de posição a mediana amostral. Doravante usaremos as notações β e $\hat{\beta}$ para denotar, respectivamente, $\beta(0, 5)$ e $\hat{\beta}(0, 5)$ devido a importância deste caso particular. Note que neste caso o problema de minimização dado em (1.3) se reduz a

$$\min_{\beta \in \mathbb{R}^{\rho+1}} \left[\frac{1}{2} \sum_{i=1}^n |Y_i - \mathbf{X}_i \beta| \right] = \min_{\beta \in \mathbb{R}^{\rho+1}} \left[\sum_{i=1}^n |Y_i - \mathbf{X}_i \beta| \right],$$

que é o problema de minimização da regressão L_1 . Eis, portanto, o objeto do nosso estudo, determinar o estimador $\hat{\beta}$ que minimiza a soma dos erros absolutos. É interessante notar que esse problema é basicamente um problema de minimização de norma L_k com $k = 1$. Há vários tipos de normas e cada uma dessas normas pode dar origem a um método de estimação. Por exemplo o estimador de mínimos quadrados é baseado na norma L_k com $k = 2$. Consulte o Apêndice B para saber um pouco mais sobre normas.

Motivados pelo avanço da computação e sua acessibilidade, [Charnes et al. \(1955\)](#) reformularam o método de estimação via mínima soma dos erros absolutos como um problema de programação linear e o solucionaram usando o método simplex. De fato, o método de estimação é equivalente ao seguinte problema de programação linear:

$$\min_{\beta} [\mathbf{1}^\top \epsilon^+ + \mathbf{1}^\top \epsilon^-], \quad (1.4)$$

em que $\epsilon^+ = (\epsilon_1^+, \epsilon_2^+, \dots, \epsilon_n^+)^\top$ e $\epsilon^- = (\epsilon_1^-, \epsilon_2^-, \dots, \epsilon_n^-)^\top$, sendo

$$\epsilon_i^+ = \begin{cases} \epsilon_i, & \text{se } \epsilon_i > 0 \\ 0, & \text{se } \epsilon_i \leq 0 \end{cases}, \quad i = 1, 2, \dots, n$$

e

$$\epsilon_i^- = \begin{cases} |\epsilon_i|, & \text{se } \epsilon_i < 0 \\ 0, & \text{se } \epsilon_i \geq 0 \end{cases}, \quad i = 1, 2, \dots, n;$$

sujeito à condição

$$Y = X\beta + \epsilon^+ - \epsilon^-,$$

com ϵ^+ e ϵ^- sendo vetores cujos elementos são todos não negativos.

A solução deste problema é o estimador L_1 de β e ϵ^+, ϵ^- são os erros positivos e negativos, respectivamente e $\mathbf{1}^\top = (1, \dots, 1)$ com dimensão $1 \times n$.

É importante ressaltar que este problema não possui solução analítica e portanto é necessário utilizar um método iterativo, diferentemente do método de mínimos quadrados que possui solução analítica. Alguns algoritmos foram propostos para proceder a esta minimização, dentre os algoritmos propostos se destacam os de [Barrodale e Roberts \(1973, 1974\)](#) que se baseiam no método simplex da programação linear. Para mais detalhes sobre programação linear veja [Goldbarg e Luna \(2005\)](#).

Antes de concluirmos esta seção, mostraremos que de fato o estimador de máxima verossimilhança de β quando os erros têm distribuição exponencial dupla coincidem com os estimadores de mínima soma dos erros absolutos.

Partindo da FDP da distribuição $Laplace(\mu, \lambda)$, dada pela Equação (1.2), vemos que a função de verossimilhança é dada por

$$L(\beta, \lambda, \mathbf{Y}) = \prod_{i=1}^n \frac{1}{2\lambda} \exp\left(-\frac{|Y_i - \mathbf{X}_i\beta|}{\lambda}\right) = (2\lambda)^{-n} \exp\left(-\frac{1}{\lambda} \sum_{i=1}^n |Y_i - \mathbf{X}_i\beta|\right),$$

para obtermos os estimadores de máxima verossimilhança precisaremos maximizar esta função com relação aos parâmetros β e λ . Para facilitar este trabalho consideraremos o logaritmo natural da função de verossimilhança, perceba que como o logaritmo é uma função monótona crescente, então o ponto de máximo da função de verossimilhança coincidirá com o ponto de máximo do logaritmo da função de verossimilhança. Nessas condições

$$l(\beta, \lambda, \mathbf{Y}) = \ln(L(\beta, \lambda, \mathbf{Y})) = -n \ln(2\lambda) - \frac{1}{\lambda} \sum_{i=1}^n |Y_i - \mathbf{X}_i\beta|.$$

Maximizando esta função com relação a β , vemos que o termo $-n \ln(2\lambda)$ é constante, portanto

$$\max_{\beta} \left[-n \ln(2\lambda) - \frac{1}{\lambda} \sum_{i=1}^n |Y_i - \mathbf{X}_i\beta| \right] = \max_{\beta} \left[-\frac{1}{\lambda} \sum_{i=1}^n |Y_i - \mathbf{X}_i\beta| \right].$$

Como $\lambda > 0$, então devemos obter $\hat{\beta}$ de modo a maximizar

$$\left[-\sum_{i=1}^n |Y_i - \mathbf{X}_i\beta| \right].$$

É importante perceber que queremos maximizar uma função estritamente negativa, consequentemente, podemos reescrever esse problema de maximização como um problema de minimização, isto é,

$$\min_{\beta} \left[\sum_{i=1}^n |Y_i - \mathbf{X}_i \beta| \right].$$

Logo vemos que o estimador de máxima verossimilhança de β supondo distribuição exponencial dupla para os erros é equivalente ao estimador de mínima soma de erros absolutos de β .

Podemos seguir obtendo o estimador de máxima verossimilhança de λ , basta derivar a função $l(\beta, \lambda, \mathbf{Y})$ com respeito a λ , igualá-la a zero e resolver a equação em termos de λ . Ou seja,

$$\frac{\partial l(\beta, \lambda, \mathbf{Y})}{\partial \lambda} = -n\lambda^{-1} + \lambda^{-2} \sum_{i=1}^n |Y_i - \mathbf{X}_i \beta|,$$

igualando a zero,

$$\begin{aligned} \frac{\partial l(\beta, \lambda, \mathbf{Y})}{\partial \lambda} = 0 &\Leftrightarrow -n\lambda^{-1} + \lambda^{-2} \sum_{i=1}^n |Y_i - \mathbf{X}_i \beta| = 0 \\ &\Leftrightarrow -n + \lambda^{-1} \sum_{i=1}^n |Y_i - \mathbf{X}_i \beta| = 0 \Leftrightarrow \lambda = \frac{1}{n} \sum_{i=1}^n |Y_i - \mathbf{X}_i \beta|, \end{aligned}$$

que fornece o ponto crítico $\lambda^* = \frac{1}{n} \sum_{i=1}^n |Y_i - \mathbf{X}_i \beta|$, que é candidato a estimador de máxima verossimilhança de λ . Mas é necessário mostrar que a função $l(\beta, \lambda, \mathbf{Y})$ é côncava, isto é, mostrar que a segunda derivada do logaritmo da função de verossimilhança é uma função que assume valor negativo no ponto crítico. Isso é mostrado a seguir.

$$\begin{aligned} \frac{\partial^2 l(\beta, \lambda, \mathbf{Y})}{\partial \lambda^2} &= n\lambda^{-2} - 2\lambda^{-3} \sum_{i=1}^n |Y_i - \mathbf{X}_i \beta| = \lambda^{-2} \left(n - 2\lambda^{-1} \sum_{i=1}^n |Y_i - \mathbf{X}_i \beta| \right) \\ \Rightarrow \frac{\partial^2 l(\beta, \lambda, \mathbf{Y})}{\partial \lambda^2} \Big|_{\lambda=\lambda^*} &= (\lambda^*)^{-2} \left(n - 2(\lambda^*)^{-1} \sum_{i=1}^n |Y_i - \mathbf{X}_i \beta| \right) \\ &= (\lambda^*)^{-2} (n - 2n) = \underbrace{-n}_{<0} \underbrace{(\lambda^*)^{-2}}_{>0} < 0. \end{aligned}$$

Assim o estimador de máxima verossimilhança de λ fica dado por

$$\hat{\lambda} = \sum_{i=1}^n \frac{|Y_i - \mathbf{X}_i \hat{\beta}|}{n}. \quad (1.5)$$

Outro estimador para λ é o discutido por [André e Elian \(2022\)](#) e recomendado por [Birkes e Dodge \(1993\)](#) e [Tavares \(1998\)](#), que é resistente a *outliers*, dado por

$$\hat{\lambda} = \sqrt{n^*}(\hat{\epsilon}_{(n^*-m+1)} - \hat{\epsilon}_{(m)})/4, \quad (1.6)$$

em que n^* é o número de resíduos não nulos (em um modelo com intercepto e p variáveis explicativas $n^* = n - p - 1$), $m = (n^* + 1)/2 - \sqrt{n^*}$ e $\hat{\epsilon}_{(1)}, \dots, \hat{\epsilon}_{(n^*)}$ são as estatísticas de ordem construídas a partir dos n^* resíduos não nulos. Este estimador é fracamente consistente para λ , ou seja, ele converge em probabilidade para λ . André e Elia (2022) apresentam estudos de simulação para comparar o comportamento e desempenho dos estimadores apresentados nas Equações (1.5) e (1.6) nos contextos de pequenas amostras em que a distribuição do erro é Laplace ou pertence à família de distribuições elípticas. O principal resultado de André e Elia (2022) é que, quando os erros possuem distribuição Laplace, o estimador de máxima verossimilhança para λ é assintoticamente não viesado.

Ao decorrer deste texto utilizaremos principalmente o estimador de máxima verossimilhança de λ , que é dado na Equação (1.5), mas em algumas seções também faremos uso do estimador robusto de λ , que é apresentado na Equação (1.6). Por isso, quando estivermos trabalhando com mais de um estimador de λ , para fins de clareza, $\hat{\lambda}_{MLE}$ denotará o obtido via máxima verossimilhança e $\hat{\lambda}_{ROS}$ denotará o obtido via estatísticas de ordem dos resíduos.

Observação 1.1 As siglas das notações dos estimadores de λ vêm do inglês, sendo que MLE (*Maximum Likelihood Estimator*) significa estimador de máxima verossimilhança e ROS (*Residuals Order Statistics*) significa estatísticas de ordem dos resíduos.

Um resultado interessante acerca dos resíduos absolutos é que eles são assintoticamente não correlacionados, isto é, $\lim_{n \rightarrow \infty} \text{Cor}(|\hat{\epsilon}_i|, |\hat{\epsilon}_j|) = 0, \forall i \neq j \in \{1, \dots, n\}$; este resultado foi apresentado por Kendall (1994), basta ver que $|\hat{\epsilon}_i| = |y_i - \hat{\beta} \mathbf{X}_i|$ e aplicar o resultado dado na página 361 de Kendall (1994). Em aplicações é útil examinar os resíduos padronizados, que são definidos por $\hat{\epsilon}_i / \hat{\lambda}_{MLE}, i = 1, \dots, n$.

Na Subseção 1.3.1 discutiremos duas situações numéricas pouco usuais que podem ocorrer ao obter a estimativa L_1 : a não unicidade e a degeneração da estimativa L_1 .

1.3.1 Não unicidade e degeneração do estimador L_1

A não unicidade do estimador implica dizer, em cenários pouco usuais, é possível que haja infinitas soluções para o problema de mínima soma de erros absolutos. Um fato interessante é que quando há duas soluções (nos vértices do poliedro convexo definidos pelas restrições lineares do método simplex) ou mais para o problema de ajuste do hiperplano de regressão L_1 então, necessariamente, há um conjunto infinito não enumerável de soluções. Esses fatos decorrem imediatamente da natureza do processo de estimação L_1 , uma forma de obter a estimativa L_1 é pelo uso do método simplex. O estudo do método simplex explica a razão, no contexto da regressão L_1 , de que quando há mais de uma solução então há infinitas soluções.

O problema da não unicidade do estimador L_1 é conhecido há mais de 140 anos, de fato, [Turner \(1887\)](#) já havia notado as semelhanças entre o estimador de β na regressão L_1 e a mediana, principalmente com relação aos casos de não unicidade de estimativa pontual. Esse tópico também foi discutido por [Harter \(1977\)](#). Note que as propriedades numéricas do estimador L_1 eram conhecidas antes mesmo da popularização da área hoje conhecida como Programação Linear – esta área fornece métodos amplamente utilizados na Matemática aplicada, como o já mencionado método simplex.

O problema da não unicidade do estimador L_1 pode ser facilmente contornado realizando um duplo problema de minimização, submetendo as infinitas soluções ao problema de minimização da soma de erros quadráticos. Ou seja, basta escolhermos a solução que minimiza a norma L_2 dentre as múltiplas soluções fornecidas pela minimização da norma L_1 , quando existirem diversas soluções. Chamamos essa estimação em duas fases de estimação L_1L_2 , esse conceito é discutido em profundidade por [Planitz e Gates \(1991\)](#). [Planitz \(1985\)](#) apresenta um amplo estudo acerca de teoremas de existência e unicidade do ínfimo, no contexto das normas L_1 e L_∞ , com aplicações numéricas – o que é bem relevante, pois são apresentados resultados aplicáveis ao contexto da regressão L_1 .

Um ponto interessante é que a estimação L_1 usual é um caso particular da estimação L_1L_2 , pois há dois cenários possíveis quando aplicamos a estimação L_1 : primeiro, há uma única solução, logo a fase 2 do L_1L_2 é trivial, pois só há uma única solução para a fase 1; segundo, há infinitas soluções e qualquer uma delas satisfaz os requisitos para ser considerada estimação L_1 e portanto a estimativa L_1L_2 é uma das estimativas L_1 .

Mesmo no caso em que há múltiplas soluções para o problema de estimação L_1 isso nem sempre é um problema. Pois se há múltiplas soluções então elas fornecerão, necessariamente, a mesma soma de resíduos absolutos (SRA). E há casos nos quais não precisamos diretamente da estimativa de β , ao invés disso só precisamos da SRA. Um exemplo de medida em que não importa qual das soluções é escolhida, porque só depende da SRA, é o cálculo da função de verossimilhança preditiva perfilada (FVPP), que será tratada no Capítulo 4.

A degeneração na regressão L_1 pode ser identificada pelo fato de haver maior número de resíduos nulos (correspondentes às observações sobre o hiperplano da regressão L_1 ajustada) que o número de parâmetros de posição do modelo. Em outras palavras, o usual é que o número de resíduos nulos (associadas às observações definidoras) seja igual ao número de parâmetros de posição. [Appa e Smith \(1973\)](#) apresentam uma suposição (razoável para as aplicações práticas) acerca da regressão L_1 que contorna o inconveniente da degeneração da solução de mínima soma de erros absolutos.

Se a estimativa L_1 não for degenerada, que é o cenário usual, o número de resíduos nulos é $p + 1$ (modelo com $p + 1$ parâmetros, em que um deles é o intercepto e os demais estão associados às variáveis explicativas) porque no ajuste via regressão L_1 , exatamente $p + 1$ observações ficam sobre o hiperplano de regressão. Se a estimativa L_1

for degenerada então o número de resíduos nulos será superior a $p + 1$, que corresponde ao número de parâmetros de posição do modelo.

Para mais detalhes sobre os situações incomuns de estimação L_1 : degeneração e não unicidade consulte [Appa e Smith \(1973\)](#).

1.4 Inferência

Nesta seção, serão abordados os aspectos inferenciais da regressão L_1 tais como intervalos de confiança e testes de hipóteses para os parâmetros do modelo.

Vários autores analisaram, por meio de estudos de simulação, a eficiência relativa entre o estimador via mínima soma de erros absolutos e o estimador de mínimos quadrados, fazendo uso de diferentes distribuições. Dentre eles destacamos [Rosenberg e Carlson \(1977\)](#). Verificou-se a superioridade (eficiência) do estimador L_1 em relação ao estimador de mínimos quadrados, considerando inclusive tamanhos amostrais reduzidos, quando os erros têm distribuição Cauchy e Laplace. Os autores supracitados obtiveram alguns resultados, via simulação, relativos a propriedades dos estimadores L_1 :

1. O estimador L_1 tem menor desvio padrão que os estimadores de mínimos quadrados para regressão quando os erros têm curtose elevada.
2. Os estimadores L_1 são assintoticamente normais.
3. A diferença $(\hat{\beta} - \beta)$ em regressão L_1 é assintoticamente normal com média $\mathbf{0}$ e matriz de variâncias e covariâncias $\lambda^2(\mathbf{X}^T \mathbf{X})^{-1}$, em que λ^2/n é a variância da mediana em uma amostra de tamanho n .

Tais resultados mostraram que para as distribuições dos ϵ_i para as quais a mediana é assintoticamente mais eficiente do que a média, o estimador L_1 fornece um elipsóide de confiança para β cujo volume é estritamente menor que o volume do elipsóide de confiança obtido pelo método de mínimos quadrados. Por meio da distribuição assintótica dos estimadores L_1 , podemos obter intervalos de confiança e testes de hipóteses no contexto de regressão L_1 .

1.4.1 Intervalos de confiança

Denotemos por $\hat{\beta}$ o estimador de regressão L_1 de β no problema da regressão quantílica de ordem 0,5 definida em (1.3). Estamos interessados em um intervalo de confiança para o contraste $\mathbf{h}^T \beta$, em que \mathbf{h} é um vetor de constantes $(p + 1)$ -dimensional e β é o vetor de parâmetros, então o estimador de $\mathbf{h}^T \beta$ é $\mathbf{h}^T \hat{\beta}$. Baseado nos resultados assintóticos, podemos construir um intervalo de $(1 - \alpha)100\%$ de confiança para $\mathbf{h}^T \beta$, que é dado pela expressão (1.7):

$$IC_{1-\alpha}(\mathbf{h}^\top \boldsymbol{\beta}) = \left[\mathbf{h}^\top \hat{\boldsymbol{\beta}} - z_{\alpha/2} \lambda \left[\mathbf{h}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{h} \right]^{\frac{1}{2}} ; \mathbf{h}^\top \hat{\boldsymbol{\beta}} + z_{\alpha/2} \lambda \left[\mathbf{h}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{h} \right]^{\frac{1}{2}} \right], \quad (1.7)$$

em que $z_{\alpha/2}$ é o quantil de ordem $1 - \alpha/2$ da normal padrão e λ é o desvio padrão amostral da mediana dos resíduos.

Para um componente β_i de $\boldsymbol{\beta}$, o intervalo de $(1 - \alpha)100\%$ de confiança para β_i será dado por

$$IC_{1-\alpha}(\beta_i) = \left[\hat{\beta}_i - z_{\alpha/2} \lambda \left[(\mathbf{X}^\top \mathbf{X})_{ii}^{-1} \right]^{\frac{1}{2}} ; \hat{\beta}_i + z_{\alpha/2} \lambda \left[(\mathbf{X}^\top \mathbf{X})_{ii}^{-1} \right]^{\frac{1}{2}} \right], \quad (1.8)$$

em que o $(\mathbf{X}^\top \mathbf{X})_{ii}^{-1}$ representa o i -ésimo elemento da diagonal principal de $(\mathbf{X}^\top \mathbf{X})^{-1}$.

Se λ for desconhecido, que é o caso mais comum, deverá ser estimado. Temos que $\lambda = \theta/2$, em que $\theta = [f(\eta)]^{-1}$ em que $f(\eta)$ é a densidade dos erros aleatórios aplicada na mediana (η é a mediana, neste contexto $\eta = 0$ porque a mediana dos erros é zero por suposição). Para estimarmos λ será necessário estimar $[f(\eta)]^{-1}$. Uma estimativa consistente de $[f(\eta)]^{-1}$, dada por [Cox e Hinkley \(1974\)](#), é:

$$\hat{\theta} = \frac{\hat{\epsilon}_{(i)} - \hat{\epsilon}_{(r)}}{(i - r)/n}, \quad i = 1, \dots, n \text{ e } r = 1, \dots, n,$$

em que denotamos os resíduos observados por $\hat{\epsilon}_i, i = 1, \dots, n$ sendo que $\hat{\epsilon}_i = Y_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}$ e $\hat{\epsilon}_{(i)}$ e $\hat{\epsilon}_{(r)}$ são os resíduos ordenados de ordens i e r , respectivamente.

Além disso, $i = \lceil n/2 \rceil + v$ e $r = \lceil n/2 \rceil - v$, em que $\lceil \cdot \rceil$ indica a função teto (menor inteiro maior ou igual a) e v é escolhido de tal forma que i e r fiquem simétricos em torno da mediana amostral dos resíduos. Recomenda-se que v seja pequeno e $\hat{\epsilon}_{(i)}, \hat{\epsilon}_{(r)} \neq 0$ ([Bezerra, 1990](#)). Verifica-se que $\hat{\theta}$ é um estimador consistente de 2λ , logo uma estimativa consistente para λ é dada por $\hat{\lambda} = \hat{\theta}/2$.

A partir daí, obtém-se o intervalo de $(1 - \alpha)100\%$ de confiança para $\mathbf{h}^\top \boldsymbol{\beta}$ substituindo λ por seu estimador consistente, $\hat{\lambda}$, no intervalo de confiança (1.7).

Os intervalos de confiança apresentados são baseados na normalidade assintótica de $\hat{\boldsymbol{\beta}}$, suposição esta que depende do tamanho amostral e da distribuição dos erros do modelo. O método *bootstrap* tem sido sugerido na construção de intervalos de confiança e construção de testes de hipóteses para pequenas amostras ([Bezerra, 1990](#)). [Efron e Tibshirani \(1993\)](#) fornecem mais detalhes acerca do procedimento *bootstrap*.

1.4.2 Testes de hipóteses

Reproduziremos a seguir os testes de hipóteses assintóticos apresentados por [Bezerra \(1990\)](#).

1. $H_0 : \mathbf{h}^\top \boldsymbol{\beta} = c$ versus $H_1 : \mathbf{h}^\top \boldsymbol{\beta} \neq c$ com nível de significância α . Rejeita-se H_0 se $|Z^*| > z_{1-\alpha/2}$ em que Z^* é dado por

$$Z^* = \frac{\mathbf{h}^\top \hat{\boldsymbol{\beta}} - c}{\hat{\lambda} [\mathbf{h}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{h}]^{\frac{1}{2}}}.$$

2. $H_0 : \beta_i = 0$ versus $H_1 : \beta_i \neq 0$, $i = 0, \dots, p$ com nível de significância α . Rejeita-se H_0 se $|Z^*| > z_{1-\alpha/2}$ em que Z^* é dado por

$$Z^* = \frac{\hat{\beta}_i}{\hat{\lambda} [(\mathbf{X}^\top \mathbf{X})_{ii}^{-1}]^{\frac{1}{2}}}.$$

3. $H_0 : \boldsymbol{\beta}^* = \mathbf{0}$ versus $H_1 : \exists i \in \{1, \dots, p\}$ tal que $\beta_i \neq 0$ com nível de significância α . Rejeita-se H_0 se $W > \chi_{1-\alpha}^2(p)$ em que W é dado por

$$W = \frac{(\hat{\boldsymbol{\beta}}^*)^\top \mathbf{X}^{*\top} \mathbf{X}^* (\hat{\boldsymbol{\beta}}^*)}{\hat{\lambda}^2},$$

em que \mathbf{X}^* significa que a primeira coluna de \mathbf{X} , que é composta de 1's, foi retirada da matriz e no caso do $\boldsymbol{\beta}^*$ significa que o primeiro elemento, β_0 , foi retirado do vetor. A estatística do teste, W , possui distribuição assintótica Qui-quadrado com p graus de liberdade. Note que $\chi_{1-\alpha}^2(p)$ denota o quantil de ordem $(1 - \alpha)$ da distribuição χ_p^2 .

Segundo [Bassett e Koenker \(1978\)](#) os três testes apresentados são equivalentes ao teste de Neyman-Pearson – detalhes podem ser vistos na fonte original, [Neyman e Pearson \(1933\)](#) – e ao teste da razão de verossimilhanças.

1.5 Relações entre classes de modelos

Até agora, no contexto do modelo linear, comentamos que os estimadores de mínimos quadrados são idênticos aos estimadores de máxima verossimilhança quando se supõe distribuição normal para os erros. Também apresentamos o resultado de que os estimadores via mínima soma dos erros absolutos (L_1) são idênticos aos estimadores de máxima verossimilhança quando os erros têm distribuição Laplace. Ressaltamos ainda que a regressão L_1 é um caso particular da regressão quantílica, quando estamos interessados no quantil de ordem 0,5.

Agora apresentaremos mais uma relação entre modelos. Inicialmente, introduziremos a família de distribuições elípticas (no caso univariado também podem ser chamadas de distribuições simétricas) e então enquadraremos as distribuições normal e exponencial dupla como casos particulares desta família. [Chmielewski \(1981\)](#) apresenta uma ampla revisão bibliográfica sobre as distribuições elípticas e [Cysneiros et al. \(2005\)](#) apresentam modelos de regressão cujos erros possuem distribuições elípticas – generalizando o tradicional modelo linear normal. Outra característica interessante das distribuições normal e exponencial dupla é que, enquanto a distribuição normal pertence à família exponencial, a distribuição exponencial dupla não pertence à família exponencial. Como a demonstração desse fato é consideravelmente técnica não a transcreveremos aqui, mas recomendamos fortemente que o leitor veja a demonstração em ([Shao, 2005](#), p. 51–52).

A consequência direta da distribuição Laplace não pertencer à família exponencial é que não poderemos utilizar a grande variedade de resultados cuja principal premissa é que a distribuição pertença à família exponencial. Assim, resultados assintóticos acerca de estimação pontual, intervalar e testes de hipóteses da literatura clássica, como os fornecidos por [Lehmann e Casella \(2006\)](#) e [Lehmann e Romano \(2022\)](#), não amparam a distribuição Laplace. Por isso estudos sobre métodos associados à distribuição Laplace são particularmente relevantes, pois ajudam a preencher lacunas na literatura que, de outra maneira, permaneceriam inexploradas. Posteriormente apresentaremos brevemente as ideias do modelo linear elíptico e mostraremos que tanto o modelo com erros normais quanto o modelo com erros Laplace pertencem a esta classe.

As densidades da distribuição $N(\mu, \sigma^2)$ e da $Laplace(\mu, \lambda)$ são dadas, respectivamente, por

$$f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{e} \quad f(x, \mu, \lambda) = \frac{1}{2\lambda} e^{-\frac{|x-\mu|}{\lambda}}.$$

Definição 1.1 (*Família de distribuições elípticas*) Diz-se que a variável aleatória Y pertence à família de distribuições simétricas, com parâmetro de posição μ , e parâmetro de escala θ , se sua densidade é da forma:

$$f(y; \mu, \theta, g(\cdot)) = \frac{1}{\sqrt{\theta}} g(u) \quad , \quad \theta > 0, \quad \mu \in \mathbb{R}, \quad \text{em que } u = \frac{(y - \mu)^2}{\theta}, \quad (1.9)$$

$g(\cdot)$ é a função geradora de densidades satisfazendo $\int_0^{+\infty} u^{-1/2} g(u) du = 1$.

Algumas distribuições elípticas possuem parâmetros adicionais que não são parâmetros de posição nem de escala, esses parâmetros são os parâmetros de forma (um exemplo de distribuição simétrica com parâmetro de forma é t-Student, que tem um parâmetro denominado graus de liberdade que afeta sua curtose). Quando a densidade da distribuição simétrica possuir parâmetro de forma ν , o parâmetro afetará a função geradora de densidades, isto é, a função geradora de densidades dependerá do parâmetro de forma,

$$f(y; \mu, \theta, g(\cdot), \nu) = \frac{1}{\sqrt{\theta}} g(u, \nu) \quad , \quad \theta > 0, \quad \mu \in \mathbb{R} \text{ e } \nu > 0.$$

Note que as distribuições normal e Laplace pertencem à família de distribuições elípticas com funções geradoras dadas, respectivamente, por

$$g(u) = \frac{1}{\sqrt{2\pi}} e^{-u/2}, \quad \mu = \mu \text{ e } \theta = \sigma^2$$

e

$$g(u) = \frac{1}{2} e^{-\sqrt{|u|}}, \quad \mu = \mu \text{ e } \theta = \lambda^2.$$

Então as duas distribuições apresentadas pertencem à família de distribuições elípticas e conseqüentemente os dois modelos apresentados são casos particulares do modelo linear elíptico. Os parâmetros do modelo linear elíptico podem ser estimados via máxima verossimilhança, e a grande vantagem desta formulação mais geral de modelo linear é que não estamos mais restritos à distribuição normal, podemos por exemplo utilizar uma distribuição com caudas mais pesadas que a normal, como a Cauchy, para modelar os dados.

A abordagem de atribuir uma distribuição com caudas mais pesadas para os erros se torna particularmente útil quando ajustamos um modelo com suposição de normalidade mas os resíduos apresentam um alto grau de curtose. É sabidamente conhecido na literatura estatística que a distribuição normal é mesocúrtica e conseqüentemente a mesma não é adequada para modelar dados com caudas pesadas. Mais detalhes sobre o modelo linear elíptico podem ser vistos em [Cysneiros *et al.* \(2005\)](#) e para um aprofundamento nas técnicas de diagnóstico, veja [Villegas *et al.* \(2013\)](#).

A família de distribuições elípticas tem a importante propriedade de ser fechada por transformação linear, isto é, a transformação linear de uma distribuição elíptica também é elíptica. Ou seja, as distribuições elípticas possuem a seguinte propriedade:

se $Y \sim EI(\mu, \theta, g(\cdot), \alpha)$, $b \in \mathbb{R}$ e $a \neq 0$ então

$$W = aY + b \sim EI(a\mu + b, a^2\theta, g(\cdot), \alpha).$$

Portanto, no caso dos modelos lineares elípticos, a distribuição do erro e a distribuição da variável resposta condicionada às variáveis explicativas pertencem à mesma classe de distribuições. Note que no caso dos modelos lineares generalizados isto não ocorre, pois as distribuições do erro e da variável resposta condicionada às variáveis explicativas não pertencem necessariamente à mesma classe de distribuições. Por isso, no contexto dos modelos lineares generalizados se atribui a distribuição diretamente na variável resposta condicionada às variáveis explicativas. Basicamente, dentro da classe de modelos lineares elípticos, a inferência é realizada de forma análoga ao caso do modelo linear normal, isto é, só troca-se a distribuição normal por outra distribuição elíptica qualquer.

A Figura 1.3 ilustra e sintetiza, por meio de um Diagrama de Venn, a relação entre as classes de modelos vistas até aqui, em especial, é destacado o fato que os modelos lineares normal e Laplace pertencem à classe de modelos lineares elípticos, pois tanto a distribuição normal quanto a distribuição Laplace pertencem à família de distribuições elípticas. Além disso, destaca, simultaneamente, que o modelo linear normal é um caso particular de modelo linear generalizado e a regressão L_1 é um caso especial da regressão quantílica. Também ressalta que diferentemente da distribuição normal, a distribuição Laplace não pertence à família exponencial.

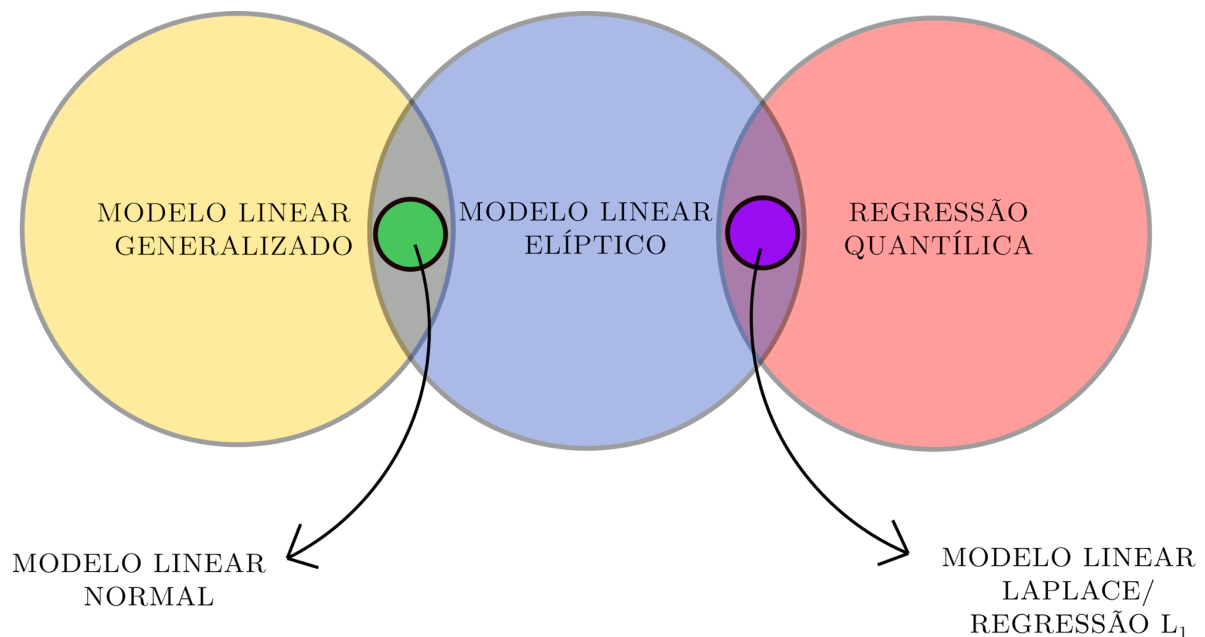


Figura 1.3: Diagrama de Venn com relação entre as classes de modelos.

A Figura 1.3 é a principal contribuição original desse capítulo e é útil para situar as similaridades e diferenças da regressão L_1 em relação a classes de modelos de regressão amplamente estudadas. Assim, concluímos a revisão e síntese dos tópicos fundamentais ligados à regressão L_1 . Na próxima seção descreveremos o objetivo deste texto e como está dividido.

1.6 Objetivo e organização do trabalho

O objetivo desta tese é desenvolver e apresentar técnicas de diagnóstico inéditas para análise do ajuste da regressão L_1 . Também é finalidade deste texto apresentar estudos analítico-matemático e simulações sobre as técnicas de diagnóstico desenvolvidas, bem como aplicá-las a dados reais.

O Capítulo 2 fornece testes de falta de ajuste no âmbito da regressão L_1 . O Capítulo 3 é voltado ao desenvolvimento de testes para verificar a hipótese de homogeneidade de variâncias, enquanto o Capítulo 4 tem como foco a construção de uma medida de influência preditiva para regressão L_1 . O Capítulo 5 fornece uma técnica de diagnóstico para regressão L_1 com base no método *Forward Search*, esta técnica permitiu desenvolver gráficos de diagnóstico para regressão L_1 . São sugeridos tópicos de pesquisa na área de regressão L_1 no Capítulo 6. Para tornar fácil e sistemática a aplicação dos métodos desenvolvidos e discutidos neste texto, criamos o pacote *diagL1* (Rodrigues e Elian, 2024), que permite utilizar as técnicas desenvolvidas de forma simples e intuitiva. O pacote pode ser acessado e baixado em <https://cran.r-project.org/package=diagL1>.

Este texto também possui alguns apêndices. O Apêndice A contém várias sinônimos presentes na literatura para referir-se a regressão L_1 , enquanto o Apêndice B apresenta conceitos básicos acerca de normas. Já o Apêndice C contém tabelas com resultados dos estudos de simulação realizados no Capítulo 2. O Apêndice D fornece testes para verificar se os erros possuem distribuição Laplace. Por fim, o Apêndice E discute tópicos auxiliares ao Capítulo 4, incluindo processamento paralelo e ferramentas computacionais.

Capítulo 2

Testes Paramétricos de Falta de Ajuste para Regressão L_1

Uma suposição imediata do modelo de regressão L_1 apresentado é a linearidade, isto é, a relação entre a variável resposta e as variáveis explicativas deve ser linear. Uma forma de verificar se esta suposição foi atendida é por meio de gráficos que envolvem a variável resposta e uma das variáveis explicativas. Este método tem duas desvantagens, primeira, é demasiadamente subjetivo, segunda, só podemos examinar um relacionamento entre variável resposta e uma variável explicativa por vez, o que não permite verificar diretamente o relacionamento global. Por isso, é necessário um método mais objetivo para examinar se a suposição de linearidade do modelo foi atendida.

No caso do ajuste por mínimos quadrados, vários testes de linearidade são encontrados na literatura. Pouco no entanto pode ser encontrado quando se utiliza a regressão L_1 . Proporemos a seguir três testes paramétricos de falta de ajuste com base nas somas de erros absolutos para verificar a suposição de falta de ajuste. Esses testes requerem medidas repetidas na variável Y para um ou mais níveis de X, ou seja, é necessário que réplicas componham a amostra. A seguir apresentamos o teste tanto para o modelo de regressão linear simples quanto para o múltiplo.

Avaliaremos a performance dos testes paramétricos de falta de ajuste propostos por meio de simulações. Posteriormente faremos uma aplicação aos dados presentes em [Su e Yang \(2006\)](#).

Testes paramétricos de falta de ajuste para o modelo de regressão linear simples L_1

Começaremos apresentando as hipóteses nula e alternativa dos testes, que são:

H_0 : A função de regressão linear é adequada, isto é, não há falta de ajuste,

H_a : A função de regressão linear não é adequada, ou seja, há falta de ajuste.

$$\begin{aligned}\mu_2 - \mu_1 &= \beta_0 + \beta_1 x_2 - (\beta_0 + \beta_1 x_1) = \beta_1(x_2 - x_1) \\ \mu_3 - \mu_1 &= \beta_0 + \beta_1 x_3 - (\beta_0 + \beta_1 x_1) = \beta_1(x_3 - x_1) \\ &\vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \\ \mu_k - \mu_1 &= \beta_0 + \beta_1 x_k - (\beta_0 + \beta_1 x_1) = \beta_1(x_k - x_1)\end{aligned}$$

logo

$$\beta_1 = \frac{\mu_2 - \mu_1}{x_2 - x_1} = \frac{\mu_3 - \mu_1}{x_3 - x_1} = \dots = \frac{\mu_k - \mu_1}{x_k - x_1}.$$

Na realidade, temos $\beta_1 = \frac{\mu_h - \mu_l}{x_h - x_l}$, para quaisquer h e l distintos em $1, \dots, k$. Daí temos também que

$$\beta_0 = \mu_i - \beta_1 x_i = \mu_i - \frac{\mu_h - \mu_l}{x_h - x_l} x_i, \text{ para quaisquer } h \text{ e } l \text{ distintos em } 1, \dots, k.$$

Note que, ainda sob H_0 ,

$$0 = \frac{\mu_3 - \mu_2}{x_3 - x_2} - \frac{\mu_2 - \mu_1}{x_2 - x_1} = \frac{\mu_4 - \mu_3}{x_4 - x_3} - \frac{\mu_3 - \mu_2}{x_3 - x_2} = \dots = \frac{\mu_k - \mu_{k-1}}{x_k - x_{k-1}} - \frac{\mu_{k-1} - \mu_{k-2}}{x_{k-1} - x_{k-2}} \Leftrightarrow \mathbf{A}^\top \boldsymbol{\mu} = \mathbf{0},$$

isto é, são $k - 2$ diferenças de frações que resultam em zero.

Assim, sob H_0 , temos que

$$\mathbf{A}^\top \boldsymbol{\mu} = \mathbf{0}_{(k-2) \times 1},$$

em que $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_k)^\top$ denota o vetor de parâmetros do modelo completo e \mathbf{A}^\top é definida a seguir.

$$\mathbf{A}^\top = \begin{bmatrix} \frac{1}{x_2 - x_1} & \frac{-1}{x_2 - x_1} + \frac{-1}{x_3 - x_2} & \frac{1}{x_3 - x_2} & 0 & 0 & \dots \\ 0 & \frac{1}{x_3 - x_2} & \frac{-1}{x_3 - x_2} + \frac{-1}{x_4 - x_3} & \frac{1}{x_4 - x_3} & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\ 0 & 0 & 0 & \frac{1}{x_{k-1} - x_{k-2}} & \frac{-1}{x_{k-1} - x_{k-2}} + \frac{-1}{x_k - x_{k-1}} & \frac{1}{x_k - x_{k-1}} \end{bmatrix},$$

perceba que \mathbf{A}^\top tem dimensão $(k - 2) \times k$ e tem posto $k - 2$.

A partir dessas observações é possível executar o teste de falta de ajuste como um teste linear geral cujas hipóteses são

$$\begin{cases} H_0 : \mathbf{A}^\top \boldsymbol{\mu} = \mathbf{0} & \text{(o modelo linear é adequado)} \\ H_a : \mathbf{A}^\top \boldsymbol{\mu} \neq \mathbf{0} & \text{(o modelo linear é inadequado)} \end{cases}.$$

Com esta reformulação das hipóteses dos testes para falta de ajuste, podemos aplicar os resultados de [Bai, Rao, e Yin \(1990\)](#).

Proporemos três testes paramétricos considerando a suposição de que os erros possuem distribuição Laplace.

Antes de calcular as estatísticas dos testes é necessário obter duas estimativas de $\boldsymbol{\mu}$,

uma restrita à hipótese $\mathbf{A}^\top \boldsymbol{\mu} = \mathbf{0}$ ($\tilde{\boldsymbol{\mu}}_n$) e outra sem restrições ($\hat{\boldsymbol{\mu}}_n$), sob o modelo completo. Neste caso, $\hat{\boldsymbol{\mu}}_n$ e $\tilde{\boldsymbol{\mu}}_n$ são dados, respectivamente, por

$$\hat{\boldsymbol{\mu}}_n = \arg \min_{\boldsymbol{\mu}} \sum_{i=1}^k \sum_{j=1}^{n_i} |Y_{ij} - \mathbf{c}_i^\top \boldsymbol{\mu}|$$

e

$$\tilde{\boldsymbol{\mu}}_n = \arg \min_{\boldsymbol{\mu}} \sum_{i=1}^k \sum_{j=1}^{n_i} |Y_{ij} - \mathbf{c}_i^\top \boldsymbol{\mu}|, \text{ sujeito a } \mathbf{A}^\top \boldsymbol{\mu} = \mathbf{0},$$

em que \mathbf{c}_i denota o i -ésimo vetor da base canônica de \mathbb{R}^k , ou seja, \mathbf{c}_i denota um vetor de dimensão k cujo i -ésimo componente é 1 e os demais componentes são iguais a 0.

Note ainda que

$$\sum_{i=1}^k \sum_{j=1}^{n_i} |Y_{ij} - \mathbf{c}_i^\top \tilde{\boldsymbol{\mu}}_n| = \sum_{i=1}^k \sum_{j=1}^{n_i} |Y_{ij} - \mathbf{X}_i \hat{\boldsymbol{\beta}}_n|,$$

em que

$$\boldsymbol{\beta} = (\beta_0, \beta_1)^\top \text{ e } \hat{\boldsymbol{\beta}}_n = \arg \min_{\boldsymbol{\beta}} \sum_{i=1}^k \sum_{j=1}^{n_i} |Y_{ij} - \mathbf{X}_i \boldsymbol{\beta}|.$$

O Corolário 2.1 é consequência imediata do Teorema 3 apresentado por [Bai, Rao, e Yin \(1990\)](#) ao contexto do modelo linear Laplace. Este corolário já está adaptado ao contexto dos testes paramétricos de falta de ajuste e à notação desta tese.

Corolário 2.1 (do Teorema 3 de [Bai, Rao, e Yin \(1990\)](#)) *Considere que as duas condições a seguir sejam válidas*

1 *os erros são independentes e identicamente distribuídos com distribuição Laplace(0, λ);*

2 $\lim_{n \rightarrow \infty} \max_{i \leq k} \left\| \mathbf{c}_i^\top \mathbf{S}_n^{-1/2} \right\| = 0$, em que $\|\cdot\|$ denota a norma euclidiana e $\mathbf{S}_n = \sum_{i=1}^k \sum_{j=1}^{n_i} \mathbf{c}_i \mathbf{c}_i^\top$ é uma matriz positiva definida para algum n ,

então

$$\sum_{i=1}^k \sum_{j=1}^{n_i} |Y_{ij} - \mathbf{c}_i^\top \tilde{\boldsymbol{\mu}}_n| - \sum_{i=1}^k \sum_{j=1}^{n_i} |Y_{ij} - \mathbf{c}_i^\top \hat{\boldsymbol{\mu}}_n| \xrightarrow{D} [4f(0)]^{-1} \chi_{k-2}^2, \quad (2.1)$$

em que $f(\cdot)$ denota a função densidade de probabilidade dos erros do modelo, que possuem distribuição Laplace com parâmetro de posição nulo, cuja função foi apresentada na Equação (1.2). Portanto

$$[4f(0)]^{-1} = \left[4 \frac{1}{2\lambda} \right]^{-1} = \frac{\lambda}{2}.$$

Demonstração. A demonstração do teorema que dá origem a este corolário pode ser conferida em [Bai, Rao, e Yin \(1990\)](#). Para mostrar que esse resultado é verdadeiro no

contexto do teste de falta de ajuste aqui desenvolvido basta verificar se as duas suposições são realmente válidas.

A suposição 1 é imediata e decorre das suposições do modelo, vamos verificar a suposição 2. Veja que \mathbf{S}_n é uma matriz diagonal de dimensão $k \times k$ cujos componentes são (n_1, n_2, \dots, n_k) e conseqüentemente $\mathbf{S}_n^{-1/2}$ é uma matriz diagonal cujos componentes são $(1/\sqrt{n_1}, 1/\sqrt{n_2}, \dots, 1/\sqrt{n_k})$. Perceba que $\left\| \mathbf{c}_i^\top \mathbf{S}_n^{-1/2} \right\| = 1/\sqrt{n_i} \forall i \in \{1, 2, \dots, k\}$. Note que

$$\max_{i \leq k} \left\| \mathbf{c}_i^\top \mathbf{S}_n^{-1/2} \right\| = \max_{i \leq k} 1/\sqrt{n_i} = 1/\sqrt{n_{\min}},$$

pois o valor máximo é atingindo quando n_i é mínimo; n_{\min} representa o menor dos $n_i \forall i \in \{1, 2, \dots, k\}$ e pode ser visto nas Tabelas 2.1 e 2.3, que apresentam as especificações das simulações.

Como a suposição 2 só é atendida se $\lim_{n \rightarrow \infty} \max_{i \leq k} \left\| \mathbf{c}_i^\top \mathbf{S}_n^{-1/2} \right\| = 0$ então é necessário que $\lim_{n \rightarrow \infty} 1/\sqrt{n_{\min}} = 0$, mas isso só ocorre quando $n_{\min} \rightarrow \infty$. Assim, a suposição 2 só é atendida se $\lim_{n \rightarrow \infty} n_i = \infty, \forall i \in \{1, 2, \dots, k\}$. Em outras palavras, como o limite dado na suposição 2 só é válido quando o limite do menor número de réplicas (n_{\min}) é infinito então isso equivale a todos os números de réplicas ($n_i \forall i \in \{1, 2, \dots, k\}$) tenderem ao infinito. ■

As estatísticas LF_1 e LF_2 , cujas expressões são dadas em (2.2) e (2.3) – bem como suas respectivas distribuições assintóticas sob H_0 – são as conseqüências mais úteis do Corolário 2.1.

Observação 2.1 As estatísticas desenvolvidas a partir do Corolário 2.1 terão desempenhos melhores à medida que o número mínimo de réplicas (n_{\min}) por valor da variável explicativa for maior.

Para que as distribuições das estatísticas LF_1 e LF_2 se aproximem adequadamente da distribuição χ^2 , é importante balancear bem o número de réplicas, pois é ineficaz, por exemplo, ter 1000 réplicas em um único valor de x e 2 réplicas nos demais valores e esperar que o teste funcione adequadamente. Para aplicações reais, é importante descobrir a partir de qual número mínimo de réplicas os testes terão boa efetividade. As estatísticas LF_1 e LF_2 são apresentadas nas Expressões (2.2) e (2.3). A primeira dada por

$$LF_1 = \frac{2}{\lambda} \left(\sum_{i=1}^k \sum_{j=1}^{n_i} |Y_{ij} - \mathbf{x}_i \hat{\beta}_n| - \sum_{i=1}^k \sum_{j=1}^{n_i} |Y_{ij} - \mathbf{c}_i^\top \hat{\mu}_n| \right) \xrightarrow{D} \chi_{k-2}^2, \quad (2.2)$$

em que λ é o parâmetro de escala que indica o desvio médio da distribuição Laplace.

Outra estatística para testar as mesmas hipóteses, sob as mesmas suposições da

estatística LF_1 presente em (2.2) é dada por

$$LF_2 = 2 \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} |Y_{ij} - \mathbf{x}_i \hat{\beta}_n| - \sum_{i=1}^k \sum_{j=1}^{n_i} |Y_{ij} - \mathbf{c}_i^\top \hat{\boldsymbol{\mu}}_n|}{\sum_{i=1}^k \sum_{j=1}^{n_i} |Y_{ij} - \mathbf{c}_i^\top \hat{\boldsymbol{\mu}}_n| / (n - k)} \xrightarrow{D} \chi_{k-2}^2. \quad (2.3)$$

O resultado assintótico acerca da estatística LF_1 , apresentado na Equação (2.2), é consequência imediata do Corolário 2.1. Enquanto a distribuição assintótica da estatística LF_2 é fundamentada na aplicação do Corolário 2.1 e na Equação (3.6) de Bai, Rao, e Yin (1990), que é outra consequência do Teorema 3 do trabalho dos autores previamente citados. De fato, o denominador da estatística LF_2 está fazendo o mesmo papel que $\hat{\lambda}$ na estatística LF_1 expressa na Equação (2.4). Esse papel em comum é o de estimador consistente para λ .

As estatísticas LF_1 e LF_2 não possuem distribuições exatas conhecidas, por isso obtivemos estes resultados assintóticos; o que é esperado no contexto da regressão L_1 , já que neste contexto toda a inferência é realizada de forma assintótica. Uma observação sobre a Equação (2.2), em aplicações a dados reais o parâmetro λ é desconhecido, portanto basta substituir λ por um estimador consistente, que denotaremos por $\hat{\lambda}$. Nessas condições, como consequência do Teorema de Slutsky, então

$$LF_1 = \frac{2}{\hat{\lambda}} \left(\sum_{i=1}^k \sum_{j=1}^{n_i} |Y_{ij} - \mathbf{x}_i \hat{\beta}_n| - \sum_{i=1}^k \sum_{j=1}^{n_i} |Y_{ij} - \mathbf{c}_i^\top \hat{\boldsymbol{\mu}}_n| \right) \xrightarrow{D} \chi_{k-2}^2. \quad (2.4)$$

Para efetivamente utilizar a estatística LF_1 , expressa na Equação (2.4), é necessário utilizar uma estimativa robusta de λ . No estudo de simulação deste capítulo aplicaremos dois estimadores consistentes para λ , o estimador de máxima verossimilhança para λ denotado por $\hat{\lambda}_{MLE}$, expresso na Equação (1.5), e outro baseado nas estatísticas de ordem dos resíduos denotado por $\hat{\lambda}_{ROS}$ (Residuals Order Statistics), expresso na Equação (1.6). Note que temos efetivamente três testes: LF_1 utilizando o estimador $\hat{\lambda}_{MLE}$, LF_1 utilizando o estimador $\hat{\lambda}_{ROS}$ e LF_2 .

Observação 2.2 A região crítica dos testes com base nas estatísticas LF_1 e LF_2 é unilateral à direita, isso fica mais claro ao observar as expressões dessas estatísticas dadas em (2.2) e (2.3).

Os testes propostos neste capítulo possuem as mesmas hipóteses nula e alternativa que o teste F para falta de ajuste com réplicas do modelo linear normal, que é baseada em somas de resíduos ao quadrado.

Uma similaridade importante entre os testes paramétricos propostos e o teste F é que em muitos cenários não teremos réplicas exatas, mas há estudos que apontam soluções para ajustar o teste F ao caso em que não temos réplicas exatas, como Green (1971) e

Shillington (1979). Para os testes propostos é possível aproveitar uma das metodologias do teste F quando não há réplicas exatas, que se trata de formar grupos de observações para proceder com a aplicação dos testes.

Há três diferenças fundamentais entre os testes paramétricos propostos para falta de ajuste na regressão L_1 e o tradicional teste F para falta de ajuste:

1. a mais óbvia, os testes paramétricos de falta de ajuste propostos usa soma de resíduos absolutos e não quadráticos;
2. enquanto a estatística do teste F é baseada em resultados exatos e fórmulas fechadas as estatísticas dos testes propostos são baseadas em resultados assintóticos e sem fórmulas fechadas;
3. enquanto a estatística do teste F possui distribuição exata as estatísticas dos testes propostos somente possuem distribuição assintótica conhecida.

2.1 Comentários sobre os testes

Uma vez que temos efetivamente três testes paramétricos, é necessário examinar as propriedades dos testes por meio de simulações e se possível destacar qual é o melhor teste paramétrico. A seguir estão alguns comentários sobre os testes desenvolvidos.

- 1) Se houver um número baixo de réplicas em alguns valores de x podemos montar uma estatística que não envolva os valores com poucas réplicas. Neste caso os resultados apresentados se mantêm, contudo perde-se parte da informação local fornecida pelos valores de x não utilizados no teste, isto é, é possível realizar uma extrapolação dos testes apresentados modificando a equação a seguir,

$$0 = \frac{\mu_3 - \mu_2}{x_3 - x_2} - \frac{\mu_2 - \mu_1}{x_2 - x_1} = \frac{\mu_4 - \mu_3}{x_4 - x_3} - \frac{\mu_3 - \mu_2}{x_3 - x_2} = \dots = \frac{\mu_k - \mu_{k-1}}{x_k - x_{k-1}} - \frac{\mu_{k-1} - \mu_{k-2}}{x_{k-1} - x_{k-2}}.$$

- 2) Não é necessário que haja réplicas para todos os valores de x para que o teste possa ser aplicado, basta que haja réplicas para pelo menos um valor de x .
- 3) Quando não há réplicas no conjunto de dados, um teste aproximado para falta de ajuste pode ser realizado considerando observações adjacentes em termos de x como sendo pseudo-réplicas (também conhecidas como réplicas aproximadas). Para elucidar melhor esse conceito de réplicas aproximadas considere a seguinte situação. Há apenas uma variável explicativa (X) com 100 observações assumindo os valores 1, 2, 3, ..., 98, 99, 100 para a variável explicativa. Uma ideia para aplicar os testes desenvolvidos neste capítulo é agrupar as observações com base em alguma medida de similaridade ou distância entre os valores da variável explicativa

(X). Como os valores da situação são os números de 1 a 100, é possível formar 10 grupos de observações sequenciais, isto é, grupo 1 = {1, 2, ..., 9, 10}, grupo 2 = {11, 12, ..., 19, 20}, ..., grupo 10 = {91, 92, ..., 99, 100}. Após formar os grupos, calcularemos a média dos valores da variável explicativa de cada uma das observações pertencente ao grupo, isto é, para fins de aplicação do teste consideraremos como variável explicativa para os dez (10) grupos: $\bar{x}_1 = 5,5$, $\bar{x}_2 = 15,5$, ..., $\bar{x}_{10} = 95,5$. Após isso, será possível aplicar o teste.

- 4) A estimativa $\tilde{\mu}$ pode ser obtida por meio de $\hat{\beta}$ utilizando a fórmula $\tilde{\mu}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, pois são duas formas de parametrizar o modelo linear associado à hipótese nula. No caso em que a solução L_1 não é única, essa equação não tem utilidade, pois há múltiplas soluções. Contudo, as estatísticas LF_1 e LF_2 não dependem diretamente das estimativas dos parâmetros, mas sim da soma de resíduos absolutos. Daí há duas possibilidades: primeira, a solução L_1 é única e a equação é válida, gerando a mesma soma de resíduos absolutos; segunda possibilidade, a equação não é válida, mas por haver múltiplas soluções todas elas geram a mesma soma de resíduos absolutos.
- 5) Os termos LF_1 e LF_2 para as estatísticas dos testes paramétricos são formados a partir de uma sigla que significa *Lack of Fit*.

2.2 Estudo de simulação das estatísticas LF_1 e LF_2 em configurações com efeitos de grupo, polinomial e exponencial

Nesta seção apresentaremos os resultados obtidos via simulação para os testes LF_1 e LF_2 . Definimos quatro tipos de modelos lineares a serem simulados: o modelo com efeitos de grupo, o modelo com termo quadrático, o modelo com termo cúbico e o modelo com termo exponencial. As funções de regressão dos quatro modelos são dadas a seguir:

$$\begin{aligned}
 Y_{ij} &= \mu_i + \epsilon_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i && \text{(grupo).} \\
 Y_{ij} &= \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i && \text{(quadrático).} \\
 Y_{ij} &= \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \epsilon_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i && \text{(cúbico).} \\
 Y_{ij} &= \beta_0 + \beta_1 e^{x_i} + \epsilon_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i && \text{(exponencial).}
 \end{aligned} \tag{2.5}$$

A princípio, considerando exatamente a mesma amostra, é esperado que os testes LF_1 e LF_2 tenham mais facilidade de rejeitar o modelo com termo exponencial do que o modelo com termo cúbico, também esperamos que os teste rejeitem mais facilmente o modelo com termo cúbico do que o com termo quadrático.

A seguir estão os valores numéricos dos parâmetros presentes nos modelos simulados.

$$\begin{aligned}
 Y_{ij} &= \mu_i + \epsilon_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i, && \text{(grupo).} \\
 Y_{ij} &= 5 + 2x_i + 1x_i^2 + \epsilon_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i && \text{(quadrático).} \\
 Y_{ij} &= 5 + 2x_i + 0,8x_i^2 + 0,2x_i^3 + \epsilon_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i && \text{(cúbico).} \\
 Y_{ij} &= 5 + 0,2e^{x_i} + \epsilon_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i && \text{(exponencial).}
 \end{aligned}
 \tag{2.6}$$

Simularemos cada um dos quatro modelos com cinco valores distintos para o parâmetro λ , isto é, $\lambda = 1, 3, 5, 10, 20$. Além disso, variaremos o número de réplicas para os valores das variáveis explicativas. Na Tabela 2.1 está a descrição da simulação dos modelos com relação aos valores assumidos pelas variáveis explicativas e a quantidade de réplicas para cada valor.

No caso do modelo com efeito de grupo, em conformidade com a Tabela 2.1, nas configurações de 1 a 4, $\mu_i \in \{10, 1, 9, 3, 7\}$ e nas configurações de 5 a 8, $\mu_i \in \{8, 2, 6, 0, 9, 10, 1, 9, 3, 7\}$.

config.	valores da variável explicativa	número de réplicas para cada valor	k	n	n_{min}
1	1, 2, 3, 4, 5	10, 10, 10, 10, 10	5	50	10
2	1, 2, 3, 4, 5	20, 20, 20, 20, 20	5	100	20
3	1, 2, 3, 4, 5	15, 20, 30, 20, 15	5	100	15
4	1, 2, 3, 4, 5	5, 30, 30, 30, 5	5	100	5
5	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	10, 10, 10, 10, 10, 10, 10, 10, 10, 10	10	100	10
6	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	5, 35, 5, 35, 5, 35, 5, 35, 5, 35	10	200	5
7	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	20, 20, 20, 20, 20, 20, 20, 20, 20, 20	10	200	20
8	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	30, 30, 30, 30, 30, 30, 30, 30, 30, 30	10	300	30

k : número de valores distintos da variável explicativa.

n : tamanho amostral.

n_{min} : menor número de réplicas.

Tabela 2.1: Configurações da simulação.

A estatística LF_1 será computada com os dois estimadores de λ apresentados. O objetivo é investigar o desempenho e o poder dos dois testes, no caso da medida LF_1 temos ainda o interesse em verificar se o estimador de λ utilizado tem impacto no desempenho do teste. As tabelas com os resultados detalhados das simulações estão na Seção C.1 do Apêndice C.

Resultados das simulações

Foram realizados 1000 ajustes de modelos para cada uma das situações a seguir, isto é, fixamos um valor para λ no conjunto $\{1, 3, 5, 10, 20\}$, uma das oito configurações de réplicas correspondendo a cada linha da Tabela 2.1 e finalmente ajustamos 1000

modelos com efeito nos grupos, 1000 com termo quadrático, 1000 com termo cúbico e 1000 com termo exponencial. Em todas as simulações consideramos o nível de 5% de significância. Em cada uma das tabelas é calculada a taxa de rejeição da hipótese nula, a média da estatística do teste e o valor crítico do teste, a comparação do valor médio da estatística do teste com o valor crítico permite uma rápida comparação das três estatísticas nos diferentes cenários simulados.

As conclusões das simulações, nas Tabelas C.1 a C.8, foram que o aumento do desvio absoluto médio λ piora o desempenho dos três testes para todos os tipos de modelos simulados (com efeito no grupo, com termo quadrático, com termo cúbico e com termo exponencial), pois percebe-se que à medida que o parâmetro de escala (λ) aumenta, a taxa de rejeição dos três testes cai. O desempenho dos testes melhora quando o tamanho da amostra (número de observações com réplicas) aumenta. O teste LF_1 utilizando $\hat{\lambda}_{MLE}$ se saiu melhor do que o mesmo teste utilizando $\hat{\lambda}_{ROS}$ em todas as configurações de simulação. Por pequena margem, o teste LF_1 utilizando $\hat{\lambda}_{MLE}$ também teve desempenho superior ao teste com base na estatística LF_2 . Assim, para aplicações futuras é mais vantajoso utilizar o teste LF_1 com estimador $\hat{\lambda}_{MLE}$.

Em posse destes primeiros resultados faremos uma análise minuciosa do poder do teste, o objetivo é verificar o impacto das especificações k , n e n_{min} no poder do teste e assim estabelecer diretrizes iniciais para a utilização do teste. Escolhemos estudar a taxa de rejeição em todos os cenários quando $\lambda = 10$, pois esse é um contexto desafiante mas não tão extremo como $\lambda = 20$. A estatística LF_1 ($\hat{\lambda}_{MLE}$) foi escolhida devido a sua superioridade verificada nas simulações. A Tabela 2.2 contém as taxas de rejeição da estatística LF_1 ($\hat{\lambda}_{MLE}$) quando $\lambda = 10$ para as oito configurações junto das suas respectivas especificações k , n e n_{min} .

Config.	Grupo	Quadrático	Cúbico	Exponencial	k	n	n_{min}	n/k
1	0,454	0,037	0,583	0,636	5	50	10	50/5 = 10
2	0,760	0,122	0,905	0,947	5	100	20	100/5 = 20
3	0,768	0,109	0,876	0,908	5	100	15	100/5 = 20
4	0,792	0,051	0,507	0,535	5	100	5	100/5 = 20
5	0,568	0,986	1,000	1,000	10	100	10	100/10 = 10
6	0,885	1,000	1,000	1,000	10	200	5	200/10 = 20
7	0,911	1,000	1,000	1,000	10	200	20	200/10 = 20
8	0,987	1,000	1,000	1,000	10	300	30	300/10 = 30

k : número de valores distintos da variável explicativa.

n : tamanho amostral.

n_{min} : menor número de réplicas.

Tabela 2.2: Taxas de rejeição da estatística LF_1 ($\hat{\lambda}_{MLE}$) quando $\lambda = 10$ para as oito configurações.

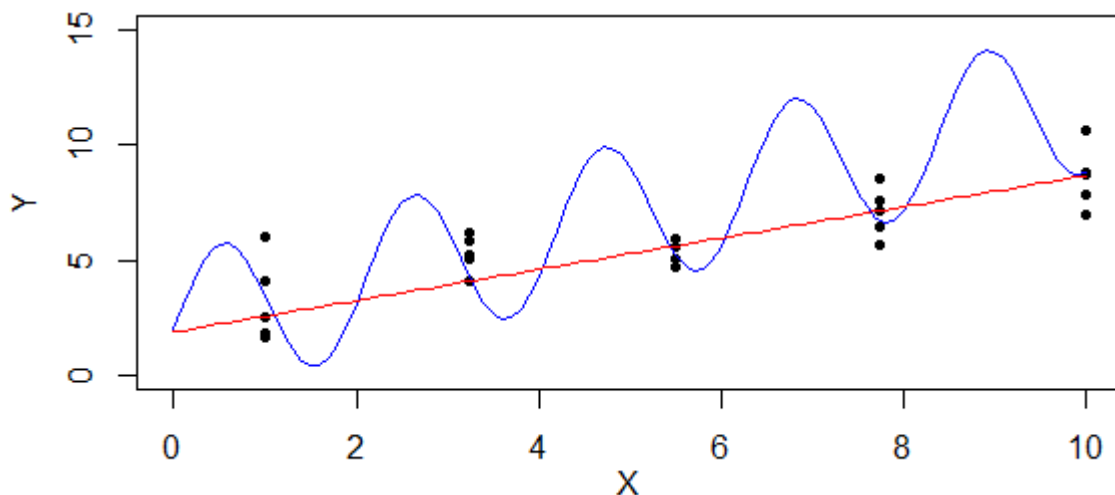
Os destaques em negrito da Tabela 2.2 apresentam as duas menores taxas de rejeição da estatística $LF_1(\hat{\lambda}_{MLE})$ para cada um dos quatro tipos de efeitos. Somente as configurações 1, 4 e 5 possuem as menores taxas de rejeição. As observações 1 e 5 possuem indicador $n/k = 10$, já a configuração 4 possui $n_{min} = 5$. A configuração 1 possui as menores taxas de rejeição para efeitos de grupo e com termo quadrático, o que sugere que $n \leq 50$ é inadequado para utilizar o teste desenvolvido. A configuração 4 possui as menores taxas de rejeição para efeitos cúbico e exponencial, o que sugere que $n_{min} \leq 5$ é inadequado para aplicar o teste.

Em resumo, com base nos resultados das simulações desta seção, é recomendável aplicar o teste com base na estatística $LF_1(\hat{\lambda}_{MLE})$ quando $n > 50$, $n_{min} > 5$ e $n/k > 10$.

2.3 Estudo de simulação das estatísticas LF_1 e LF_2 em configurações com efeito senoidal

O propósito da simulação senoidal é verificar o desempenho dos testes de falta de ajuste quando há presença de erros sistemáticos no modelo ajustado. Aqui adicionaremos a perturbação sistemática aos erros por meio da função $\text{seno}(\cdot)$, mas qualquer outra função periódica serviria a este objetivo. A Figura 2.1 ilustra graficamente o tipo de perturbação que simularemos.

Figura 2.1: Gráfico dos dados simulados, a linha sólida em azul representa o modelo que gerou os dados e a linha em vermelho representa o modelo ajustado; réplica do tipo 1 com $\beta_2 = 0$.



O modelo simulado será

$$Y_{ij} = \beta_0 + \beta_1 \cdot x_{ij} + \beta_2 \text{sen}(2x_{ij}) + \epsilon_{ij}, \quad i = 1, \dots, 5, \quad j = 1, \dots, n_c, \quad (2.7)$$

em que n_c denota o número de réplicas para cada valor da variável explicativa. Os 5 valores de x_{ij} são os 5 números de 1 a 10 criados a partir do comando “seq(1, 10, length.out

= 5)” do [R Core Team \(2023\)](#). Este modelo simulado foi inspirado no apresentado por [Su e Yang \(2006\)](#). Para simplificar a simulação, $x_{ij} = x_{ij'}$, $i = 1, \dots, 5$, isto é, os valores de x_{ij} independem do índice j e dependem somente do índice i .

Observação 2.3 A notação “seq($a, b, \text{length.out} = m$)”, que significa a sequência de m números equidistantes entre os números reais a e b (inclusive). Essa notação é utilizada no software [R Core Team \(2023\)](#) para criar sequências com as propriedades descritas. Por meio dessa notação podemos representar as sequências de números de forma concisa.

Para facilitar a compreensão ilustraremos a notação com um exemplo. Considere “seq(1, 10, length.out = 5)”, que significa a sequência de 5 números reais equidistantes entre os números 1 e 10, ou de forma explícita: 1, 3, 25, 5, 50, 7, 75 e 10; o que totaliza os cinco números da sequência. A praticidade dessa notação torna-se mais evidente ao utilizarmos “seq(1, 10, length.out = 30)” em vez de escrever os trinta números de forma explícita.

Nesta simulação $\lambda = \sqrt{5/2} \approx 1,58$ (esse valor foi escolhido para que a variância dos erros seja exatamente igual a 5), $\beta_0 = 2$, $\beta_1 = 1$ e utilizaremos cinco valores para β_2 , que serão 0, 0,8, 1,6, 2,4 e 3,2 (cada um dos cinco valores de β_2 corresponde a um dos cinco cenários de réplicas). O motivo de variar os valores de β_2 é que é esperado que à medida que os valores de β_2 crescem, o desempenho do modelo linear simples piore, devido ao efeito senoidal associado à variável explicativa (o efeito senoidal é uma perturbação sistemática que viola a suposição de linearidade do modelo). Assim, faz sentido esperar que a qualidade do ajuste do modelo linear piore à medida que β_2 aumenta, levando os testes a rejeitarem H_0 para valores de β_2 superiores a 0. A Tabela 2.3 descreve os tipos de réplicas utilizados na simulação. Dessa forma, combinando os cinco tipos de réplicas e os cinco valores de β_2 temos no total de $5 \cdot 5 = 25$ cenários. Cada um dos 25 cenários foi simulado 1000 vezes.

valores da variável explicativa	número de réplicas para cada valor	k	n	n_{min}
Réplica tipo 1	5	5	$5 \cdot 5 = 25$	5
Réplica tipo 2	10	5	$10 \cdot 5 = 50$	10
Réplica tipo 3	15	5	$15 \cdot 5 = 75$	15
Réplica tipo 4	20	5	$20 \cdot 5 = 100$	20
Réplica tipo 5	30	5	$30 \cdot 5 = 150$	30

k : número de valores distintos da variável explicativa.

n : tamanho amostral.

n_{min} : menor número de réplicas.

Tabela 2.3: Dados da simulação.

A medida LF_1 será computada com os dois estimadores de λ apresentados. O objetivo é investigar o poder e a taxa de cobertura dos dois testes, no caso da medida LF_1

temos interesse em verificar se o estimador de λ utilizado tem impacto para pequenas amostras. As tabelas com os resultados detalhados das simulações estão na Seção C.2 do Apêndice C.

Resultados das simulações

Ao analisar os resultados das simulações, nas Tabelas C.9 a C.33, notamos alguns padrões, que são comentados a seguir.

1. Para os cenários em que $\beta_2 = 0$, a média da estatística $LF_1(\hat{\lambda}_{MLE})$ calculada a partir das 1000 simulações costuma ser levemente superior às das estatísticas $LF_1(\hat{\lambda}_{ROS})$ e LF_2 sob H_0 .
2. Para os cenários em que $\beta_2 = 0, 8$, quanto maior o número de réplicas maior a taxa de rejeição dos três testes, indicando que o poder dos testes aumenta à medida que o número de réplicas aumenta. Ainda neste cenário, o teste com base na estatística $LF_1(\hat{\lambda}_{MLE})$ apresenta superioridade frente aos outros dois testes com relação a taxa de rejeição.
3. Nos cenários em que $\beta_2 = 1, 6$, nas simulações com 30 réplicas, a taxa de rejeição dos três testes é igual a 1. No caso em que temos apenas 5 réplicas a superioridade da estatística $LF_1(\hat{\lambda}_{MLE})$ é imediata, pois tem taxa de rejeição superior aos demais testes.
4. Nos cenários em que $\beta_2 = 2, 4$, nas simulações com 15 réplicas ou mais, a taxa de rejeição dos três testes é igual a 1, já na simulação com 5 réplicas somente o teste com base na estatística $LF_1(\hat{\lambda}_{MLE})$ apresenta taxa de rejeição superior a 0,9.
5. Nos cenários em que $\beta_2 = 3, 2$, nas simulações com 15 réplicas ou mais, a taxa de rejeição dos três testes é igual a 1, já na simulação com 5 réplicas o teste com base na estatística $LF_1(\hat{\lambda}_{MLE})$ apresenta taxa de rejeição igual a 0,98.

Podemos sintetizar os padrões notando que os desempenhos dos três testes melhoraram à medida que o número de réplicas dos valores de x ou a perturbação β_2 , aumenta. A explicação para o fato do desempenho dos testes melhorar com o aumento de β_2 é que um valor de β_2 maior implica em uma perturbação sistemática na linearidade do modelo de maior magnitude – uma violação mais grave da linearidade do modelo permite aos testes desenvolvidos discriminar com maior nitidez casos nos quais a hipótese de linearidade é plausível e casos em que não é. Além disso, nas simulações, o teste com base em $LF_1(\hat{\lambda}_{MLE})$ demonstrou superioridade em relação aos outros dois testes quando o número de réplicas é baixo ou quando a perturbação β_2 é pequena. Assim, embora os testes apresentem um bom desempenho em cenários diversos, em situações gerais é recomendável que seja utilizado o teste com base em $LF_1(\hat{\lambda}_{MLE})$.

A probabilidade do erro do tipo I dos testes parece sofrer pouca influência do número de réplicas, conforme pode ser notado na Tabela 2.4. Vale ressaltar que o α utilizado nas simulações é 5%.

Analisaremos com mais detalhes a probabilidade do erro do tipo II. O objetivo é examinar com mais detalhes o impacto do número de réplicas na taxa de rejeição do teste com base em $LF_1(\hat{\lambda}_{MLE})$. As Tabelas 2.5 e 2.6 contêm as taxas de rejeição da estatística $LF_1(\hat{\lambda}_{MLE})$ para os cinco tipos de réplicas, quando $\beta_2 = 0,8$ e $\beta_2 = 1,6$, respectivamente.

A Tabela 2.6 evidencia a discrepância entre a taxa de rejeição quando são utilizadas 5 ou 10 réplicas por observação, esse resultado corrobora com os resultados da análise minuciosa feita na Seção 2.2, reforçando a importância de $n_{min} > 5$. Já a Tabela 2.5, que considera um cenário mais desafiante, pois $\beta_2 = 0,8$ (o que torna a perturbação sistemática da suposição de linearidade menos intensa que o caso em que $\beta_2 = 1,6$), ilustra que à medida que o número de réplicas aumenta o erro do tipo II diminui. É notável a discrepância na taxa de rejeição quando se passa de 5 para 10 réplicas, pois a taxa de rejeição passa de 0,118 a 0,299. Assim com relação ao n_{min} , em aplicações reais é recomendável que $n_{min} \geq 10$ – sendo uma recomendação mais estrita que a vista na Seção 2.2.

Tipo de réplica	Taxa de rejeição $LF_1(\hat{\lambda}_{MLE})$	n_{min}
Réplica tipo 1	0,018	5
Réplica tipo 2	0,004	10
Réplica tipo 3	0,011	15
Réplica tipo 4	0,004	20
Réplica tipo 5	0,010	30

Tabela 2.4: Taxas de rejeição da estatística $LF_1(\hat{\lambda}_{MLE})$ quando $\beta_2 = 0$ para os cinco tipos de réplicas.

Tipo de réplica	Taxa de rejeição $LF_1(\hat{\lambda}_{MLE})$	n_{min}
Réplica tipo 1	0,118	5
Réplica tipo 2	0,299	10
Réplica tipo 3	0,489	15
Réplica tipo 4	0,668	20
Réplica tipo 5	0,871	30

Tabela 2.5: Taxas de rejeição da estatística $LF_1(\hat{\lambda}_{MLE})$ quando $\beta_2 = 0,8$ para os cinco tipos de réplicas.

Tipo de réplica	Taxa de rejeição $LF_1(\hat{\lambda}_{MLE})$	n_{min}
Réplica tipo 1	0,550	5
Réplica tipo 2	0,905	10
Réplica tipo 3	0,982	15
Réplica tipo 4	0,996	20
Réplica tipo 5	1,000	30

Tabela 2.6: Taxas de rejeição da estatística $LF_1(\hat{\lambda}_{MLE})$ quando $\beta_2 = 1,6$ para os cinco tipos de réplicas.

Considerando os resultados das simulações desta seção e da seção anterior, recomendamos aplicar o teste com base na estatística $LF_1(\hat{\lambda}_{MLE})$ quando pelo menos $n > 50$, $n_{min} \geq 10$ e $n/k > 10$.

2.4 Aplicação a dados reais

Aplicaremos os três testes desenvolvidos neste capítulo ao conjunto de dados de imóveis apresentado por [Groebner et al. \(2001\)](#) e analisado por [Su e Yang \(2006\)](#), que está transcrito na Tabela 2.7.

Optamos por ilustrar os testes nesse conjunto de dados por uma razão: é um conjunto de dados em que não há réplicas exatas, isso ilustra um cenário realista no qual o experimento não pode ser previamente planejado de modo a existirem réplicas exatas. [Su e Yang \(2006\)](#) já fizeram um agrupamento das réplicas em cinco *clusters* com base nas variáveis Área (em pé²) e Idade (em anos), mas como faremos aplicações a modelos lineares simples, é mais adequado agrupar as observações considerando cada uma das variáveis explicativas separadamente.

As variáveis utilizadas nos modelos são: Y (variável resposta) Preço de Venda (do imóvel em dólares), X_1 Área (do imóvel em pé²) e X_2 Idade (em anos). O agrupamento das observações será feito via *K-Means* pelos comandos do R a seguir.

```
set.seed(10)
kmeans(dados_Su_Yang$squared_feet, centers = 5, iter.max = 10000)
kmeans(dados_Su_Yang$age, centers = 5, iter.max = 10000)
```

Nos comandos apresentados, “dados_Su_Yang” contém os dados da Tabela 2.7, “centers” é o argumento que indica a quantidade de grupos formados, “iter.max” é o define o número máximo de iterações e “set.seed(10)” fixa a semente do gerador de números pseudo-aleatórios (isso serve para que o resultado possa ser reproduzido por qualquer pessoa). A Tabela 2.8 contém os dados com os agrupamentos feitos pelos comandos mencionados anteriormente.

Observações	Preço de Venda (em dólares)	Área (em pé ²)	Idade	Cluster (grupo)
1	200400	1661	1	1
2	159800	1665	1	1
3	190500	1672	6	1
4	173300	1680	9	1
5	175500	1689	1	1
6	193100	1700	1	1
7	152750	1700	22	1
8	188400	1728	4	1
9	96400	1730	12	1
10	159700	1738	4	1
11	272100	2018	1	2
12	110200	2022	36	2
13	242500	2037	6	2
14	194750	2050	3	2
15	199750	2072	29	2
16	142100	2073	44	2
17	127150	2080	28	2
18	195250	2112	78	2
19	224600	2144	23	2
20	201000	2160	32	2
21	106250	2168	2	2
22	123900	2172	1	2
23	72000	954	53	3
24	110000	1000	28	3
25	80350	1016	21	3
26	82700	1100	6	3
27	57600	1217	23	3
28	150750	1228	15	3
29	164750	1290	19	3
30	166100	1301	7	3
31	144750	1312	41	3
32	72950	1407	37	3
33	190600	1477	1	3
34	176300	1488	1	3
35	229250	2358	21	4
36	288900	2383	1	4
37	229700	2400	15	4
38	331000	2418	5	4
39	301700	2460	1	4
40	169900	2500	49	4
41	253750	2518	1	4
42	225400	2542	6	4
43	245750	2546	25	4
44	208900	2600	30	4
45	262750	2619	1	4
46	309300	2950	1	5
47	235500	3000	11	5
48	130100	3000	43	5
49	289900	3063	1	5
50	302900	3114	10	5
51	298600	3217	3	5
52	337300	3277	1	5
53	280300	3300	12	5
54	317800	3403	1	5
55	298500	3550	13	5

Tabela 2.7: *Dados sobre imóveis já classificado por Su e Yang (2006).*

Observações	Preço de Venda (em dólares)	Área (em pé ²)	Idade	Cluster (Área)	Cluster (Idade)
1	200400	1661	1	3	3
2	159800	1665	1	3	3
3	190500	1672	6	3	2
4	173300	1680	9	3	2
5	175500	1689	1	3	3
6	193100	1700	1	3	3
7	152750	1700	22	3	4
8	188400	1728	4	3	2
9	96400	1730	12	3	1
10	159700	1738	4	3	2
11	272100	2018	1	5	3
12	110200	2022	36	5	4
13	242500	2037	6	5	2
14	194750	2050	3	5	3
15	199750	2072	29	5	4
16	142100	2073	44	5	5
17	127150	2080	28	5	4
18	195250	2112	78	5	5
19	224600	2144	23	5	4
20	201000	2160	32	5	4
21	106250	2168	2	5	3
22	123900	2172	1	5	3
23	72000	954	53	2	5
24	110000	1000	28	2	4
25	80350	1016	21	2	4
26	82700	1100	6	2	2
27	57600	1217	23	2	4
28	150750	1228	15	2	1
29	164750	1290	19	2	1
30	166100	1301	7	2	2
31	144750	1312	41	2	5
32	72950	1407	37	2	4
33	190600	1477	1	3	3
34	176300	1488	1	3	3
35	229250	2358	21	1	4
36	288900	2383	1	1	3
37	229700	2400	15	1	1
38	331000	2418	5	1	2
39	301700	2460	1	1	3
40	169900	2500	49	1	5
41	253750	2518	1	1	3
42	225400	2542	6	1	2
43	245750	2546	25	1	4
44	208900	2600	30	1	4
45	262750	2619	1	1	3
46	309300	2950	1	4	3
47	235500	3000	11	4	1
48	130100	3000	43	4	5
49	289900	3063	1	4	3
50	302900	3114	10	4	1
51	298600	3217	3	4	3
52	337300	3277	1	4	3
53	280300	3300	12	4	1
54	317800	3403	1	4	3
55	298500	3550	13	4	1

Tabela 2.8: Dados sobre imóveis de *Su e Yang (2006)* com agrupamento considerando cada variável explicativa isoladamente.

Ajustaremos um modelo de regressão L_1 linear simples utilizando cada uma das duas variáveis explicativas. Como não temos réplicas exatas nos dados, usaremos a classificação das observações em *clusters* como réplicas aproximadas. Uma ideia inicial é usar as médias das variáveis explicativas dentro dos *clusters* como valores das variáveis explicativas para fins de aplicação dos testes desenvolvidos. Chamaremos tais médias de \bar{x}_{1i} e \bar{x}_{2i} , $i = 1, \dots, 5$. A Tabela 2.9 apresenta os valores de \bar{x}_{1i} e \bar{x}_{2i} para cada *cluster*. Note que os índices dos grupos não estão ordenados de forma crescente em relação à média de cada grupo, mas isso não afeta a aplicação, pois o índice do grupo é apenas

um rótulo.

Grupo	\bar{X}_1	\bar{X}_2
1	2485,818	13,375000
2	1182,500	5,888889
3	1660,667	1,263158
4	3187,400	27,307692
5	2092,333	51,333333

Tabela 2.9: Médias das variáveis explicativas X_1 e X_2 dentro dos grupos formados pelos comandos apresentados.

Teste de falta de ajuste no modelo de regressão L_1 linear simples utilizando X_1 (Área)

Usando a notação apresentada anteriormente podemos expressar os modelos sob as hipóteses nula e alternativa. Sob H_0 temos o modelo linear dado por

$$Y_{ij} = \beta_0 + \beta_1 \bar{x}_{1i} + \epsilon_{ij}, \quad i = 1, \dots, 5, \quad j = 1, \dots, n_i$$

e sob H_a temos o modelo completo

$$Y_{ij} = \mu_i + \epsilon_{ij}, \quad i = 1, \dots, 5, \quad j = 1, \dots, n_i.$$

É evidente que o modelo sob H_0 possui apenas dois parâmetros de posição, que são β_0 e β_1 , já o modelo sob H_a possui cinco parâmetros de posição, $\mu_1, \mu_2, \mu_3, \mu_4$ e μ_5 — correspondendo, respectivamente, à média de cada um dos cinco *clusters*.

Os modelos ajustados, sob H_0 e H_a , são, respectivamente,

$$\hat{Y}_{ij} = 185044,3 + 4,613696 \bar{x}_{1i}, \quad i = 1, \dots, 5, \quad j = 1, \dots, n_i$$

e

$$\hat{Y}_{ij} = \hat{\mu}_i, \quad i = 1, \dots, 5, \quad j = 1, \dots, n_i$$

em que $\hat{\mu}_1 = 245750$, $\hat{\mu}_2 = 110000$, $\hat{\mu}_3 = 176300$, $\hat{\mu}_4 = 298500$, e $\hat{\mu}_5 = 195250$.

A soma de resíduos absolutos do modelo ajustado sob H_0 é 3280067 e sob H_a é 1830000. Além disso, $\hat{\lambda}_{MLE} = 59637,58$ e $\hat{\lambda}_{ROS} = 81521,9$. Os valores das três estatísticas foram

1. $LF_1(\hat{\lambda}_{MLE}) = 48,6293$;
2. $LF_1(\hat{\lambda}_{ROS}) = 35,5749$;
3. $LF_2 = 79,23863$.

Lembrando que, sob H_0 , as estatísticas dos três testes possuem distribuição assintótica $\chi^2(k-2)$, em que $k = 5$ denota o número de *clusters*, se fixarmos o nível de significância em 5%, o valor crítico do teste será o percentil de ordem 0,95 da distribuição $\chi^2(3)$, que é igual a 7,814728. Mais precisamente, o valor-P associado ao teste $LF_1(\hat{\lambda}_{MLE})$ foi $1,564374 \times 10^{-10}$. Como as três estatísticas apresentaram valores acima do valor crítico, rejeitamos a hipótese de linearidade do modelo. Se houvesse discordância entre as três estatísticas optaríamos em seguir o teste $LF_1(\hat{\lambda}_{MLE})$, pois este demonstrou desempenho superior aos demais em diversos cenários simulados nas seções anteriores.

Teste de falta de ajuste no modelo de regressão L_1 linear simples utilizando X_2 (Idade)

Faremos o mesmo procedimento para a variável X_2 . Sob H_0 temos o modelo linear dado por

$$Y_{ij} = \beta_0^* + \beta_1^* \bar{X}_{1i} + \epsilon_{ij}^*, \quad i = 1, \dots, 5, \quad j = 1, \dots, n_i$$

e sob H_a temos o modelo completo

$$Y_{ij} = \mu_i^* + \epsilon_{ij}^*, \quad i = 1, \dots, 5, \quad j = 1, \dots, n_i.$$

Os modelos ajustados, sob H_0 e H_a , são, respectivamente,

$$\hat{Y}_{ij} = 211958,3974 - 819,3194\bar{x}_{1i}, \quad i = 1, \dots, 5, \quad j = 1, \dots, n_i$$

e

$$\hat{Y}_{ij} = \hat{\mu}_i^*, \quad i = 1, \dots, 5, \quad j = 1, \dots, n_i$$

em que $\hat{\mu}_1^* = 229700$, $\hat{\mu}_2^* = 188400$, $\hat{\mu}_3^* = 253750$, $\hat{\mu}_4^* = 152750$, e $\hat{\mu}_5^* = 144750$.

A soma de resíduos absolutos do modelo ajustado sob H_0 é 3253442 e sob H_a é 2957650. Além disso, $\hat{\lambda}_{MLE}^* = 59153,48$ e $\hat{\lambda}_{ROS}^* = 83842,27$. Os valores das três estatísticas foram

1. $LF_1(\hat{\lambda}_{MLE}^*) = 10,00082$;
2. $LF_1(\hat{\lambda}_{ROS}^*) = 7,055905$;
3. $LF_2 = 10,0009$.

Neste caso rejeitamos a hipótese de linearidade do modelo, pois as três estatísticas estão acima do valor crítico do teste, que é 7,814728. Mais precisamente, o valor-P associado ao teste $LF_1(\hat{\lambda}_{MLE}^*)$ foi 0,0185592. No caso em que a hipótese de linearidade é rejeitada, é necessário buscar alguma transformação na variável resposta ou variável explicativa e aplicar o teste de hipóteses novamente.

Como o propósito desta seção é ilustrar a utilização dos testes de hipóteses desenvolvidos neste capítulo, encerraremos a aplicação aqui.

Em pesquisas futuras pretendemos demonstrar um resultado mais geral, que permita aplicar um teste de falta de ajuste no caso em que o modelo possui mais de uma variável explicativa.

Capítulo 3

Teste de Homogeneidade de Variâncias para Regressão L_1

Uma suposição necessária para aplicar o modelo de regressão L_1 apresentado é a homogeneidade de variâncias, isto é, a variância da variável resposta não pode se modificar em função de nenhuma das variáveis explicativas nem da variável resposta.

Nosso objetivo neste capítulo é construir um teste de homocedasticidade para regressão L_1 a partir da razão de verossimilhanças. As hipóteses do teste são

$$H_0 : \mathbf{V} = \lambda \mathbf{I}_n \text{ contra } H_a : \mathbf{V} \neq \lambda \mathbf{I}_n, \quad (3.1)$$

em que \mathbf{V} denota a matriz de variâncias e covariâncias do modelo, isto é,

$$\mathbf{V} = \mathbf{E} [\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top].$$

Observação 3.1 A variância de uma variável aleatória W com distribuição Laplace, cuja função densidade de probabilidade pode ser vista na Equação (1.2), pode ser calculada como função de λ . Mais precisamente

$$\text{Var}(W) = 2\lambda^2.$$

É claro que para criarmos um teste desta natureza é conveniente especificar a forma da matriz \mathbf{V} sob a hipótese alternativa. A partir da especificação da forma de H_a podemos obter diversos testes de hipóteses. Como não temos interesse em trabalhar com dados correlacionados, isto é, as observações são independentes, então sob H_a , sem perda de generalidade podemos supor que

$$\mathbf{V} = \bigoplus_{i=1}^g \lambda_i \mathbf{I}_{n_i},$$

em que dividimos as N observações em g grupos com n_1, n_2, \dots, n_g observações em

cada grupo, sendo $N = \sum_{i=1}^g n_i$. O símbolo \oplus denota a soma direta de matrizes.

Um desafio a ser superado antes de aplicar o teste de homogeneidade de variâncias é como formar os grupos de observações. Se houver apenas uma variável explicativa, o agrupamento de observações pode ser feito por meio de grupos de tamanhos próximos, em cada grupo haverá observações com valores sequenciais na variável explicativa. Se houver mais de uma variável explicativa será necessário utilizar alguma técnica de análise de *cluster*. Alguns exemplos de tais técnicas são: *K-Means*, agrupamento hierárquico (*hierarchical clustering*) e DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*). Kaufman e Rousseeuw (2009) e Hennig *et al.* (2015) tratam destas técnicas de análise de *cluster*.

O algoritmo *K-Means* é uma abordagem em que os dados são organizados em grupos distintos com base na proximidade aos centróides designados para cada *cluster*. Isso implica em calcular a média dos dados de cada grupo e atribuir pontos ao *cluster* cujo centróide é o mais próximo. Por outro lado, o agrupamento hierárquico é um processo que estabelece uma estrutura de árvore de *clusters*, permitindo uma visão hierárquica das relações entre os dados. Começando com *clusters* individuais e combinando-os em *clusters* maiores, essa abordagem revela agrupamentos em diferentes níveis de detalhes. Já o DBSCAN é um método que considera a densidade de pontos para identificar agrupamentos. Ele atribui pontos a um *cluster* se eles estiverem próximos uns dos outros em relação à distância especificada e ao número mínimo de pontos requeridos. Ademais, o DBSCAN é eficaz na identificação de ruídos e é capaz de lidar com agrupamentos de formas e tamanhos variados de maneira adaptativa.

Neste caso, estaríamos supondo que a variância é função dos valores da variável explicativa. Se isso não ocorrer, os grupos deverão ser formados de modo que todos os elementos do grupo teriam variâncias iguais e diferentes dos demais.

Partindo da premissa que temos g grupos, podemos expressar as hipóteses apresentadas em (3.1) diretamente em termos dos parâmetros de escala correspondentes a cada um dos g grupos. Ou seja, temos um vetor $\lambda = (\lambda_1, \dots, \lambda_g)^\top$ de tamanho g cujos elementos são os parâmetros de escala dos grupos. Assim, podemos reescrever as hipóteses apresentadas em (3.1) como

$$H_0 : \lambda = \lambda \mathbf{1}_g \text{ contra } H_a : \lambda \in \mathbb{R}_+^g, \text{ em que } \mathbf{1}_g = \underbrace{(1, \dots, 1)}_{g \text{ vezes}}^\top, \quad (3.2)$$

portanto não é possível aplicar os resultados assintóticos de Bai, Rao, e Yin (1990), pois não é possível reescrever as hipóteses relacionadas ao vetor de parâmetros de escala da Equação (3.2) exclusivamente em termos dos parâmetros de posição como por exemplo

$$H_0 : \mathbf{A}^\top \beta = \mathbf{0} \text{ contra } H_a : \mathbf{A}^\top \beta \neq \mathbf{0}.$$

A menos que o vetor de parâmetros de escala possa ser visto como uma função linear

do vetor dos parâmetros de posição, mas neste caso o vetor de parâmetros de escala seria desnecessário.

Na maioria dos casos práticos, a heterocedasticidade se manifesta como uma função monótona de alguma variável explicativa ou da própria variável resposta. Portanto, podemos formar grupos de observações a partir da segmentação da variável associada à heterogeneidade das variâncias.

Seguindo com o raciocínio da construção do teste, o teste será desenvolvido por meio da razão de verossimilhanças tal como no teste da razão de verossimilhanças para verificar a homocedasticidade no modelo linear normal, que foi apresentado em [Johnston e DiNardo \(1996, pg. 200–201\)](#). No numerador maximizaremos a função de verossimilhança sob H_0 e no denominador maximizaremos a função de verossimilhança sob todo o espaço paramétrico. Como em outros casos em que a estatística da razão de verossimilhanças é utilizada, aplicaremos a função $\ln(\cdot)$ à razão de verossimilhanças propriamente dita. Assim teremos uma estatística que na realidade é a diferença entre o logaritmo da verossimilhança maximizada sob H_0 e o logaritmo da verossimilhança maximizada sob todo o espaço paramétrico multiplicada por -2 .

Observação 3.2 Veremos posteriormente que a estatística do teste da razão de verossimilhanças desenvolvida no contexto da regressão L_1 é similar à estatística análoga no caso da regressão linear normal, conforme Equações (3.5) e (3.6). O mais natural neste ponto é questionar a originalidade da contribuição do teste desenvolvido neste capítulo. A contribuição reside no fato de que os resultados acerca da razão de verossimilhanças que amparam a obtenção da distribuição da estatística do teste na distribuição normal não se aplicam ao mesmo teste no contexto da regressão L_1 , pois as condições de regularidade não são válidas para a regressão L_1 . Por isso conduzimos extensos estudos de simulação para encontrar empiricamente a distribuição assintótica da estatística da razão de verossimilhanças no contexto da regressão L_1 .

Nota-se que o problema de desenvolver um teste de homogeneidade de variâncias aplicável a dados sem rótulo de grupos (sem grupos definidos) envolve dois subproblemas:

- 1 definir quantos grupos existem e quais observações compõem cada um dos grupos,
- 2 construir o teste propriamente dito a partir do número de grupos (g) e das observações que compõem cada grupo. Isto é, a solução do problema pode ser modularizada em duas partes que podem ser resolvidas de forma independente.

Neste texto, não entraremos nos detalhes de análise de *cluster* no contexto da regressão L_1 por duas razões:

- 1 o problema de análise de *cluster* no âmbito da regressão L_1 não apresenta nenhuma peculiaridade em relação à análise de *cluster* estabelecida na literatura,

2 há vasta literatura especializada em análise de *cluster*, como por exemplo [Kaufman e Rousseeuw \(2009\)](#) e [Hennig *et al.* \(2015\)](#), livros que abordam análise de *cluster*.

Sem perda de generalidade, considerando o que foi exposto, seguiremos o texto considerando que os dados possuem rótulo de grupos (grupos já definidos). Assim, daremos ênfase ao desenvolvimento detalhado do teste de homogeneidade de variâncias em si. Em aplicação do teste, caso seja necessário para melhor compreensão da metodologia, serão comentados os aspectos relevantes acerca da formação dos grupos nos dados sem rótulo de grupos.

A seguir obteremos a estatística do teste da razão de verossimilhanças para verificar a hipótese de homogeneidade de variâncias na regressão L_1 , o teste é definido para número de grupos conhecidos.

3.1 Estatística da Razão de Verossimilhanças

Por praticidade, organizamos os cálculos em duas partes, uma que trata da maximização sob todo o espaço paramétrico e outra que efetivamente apresenta o cálculo da estatística do teste. A Seção 1.3 trata da maximização sob H_0 , por isso seguiremos para a maximização no espaço total.

Maximização da Função de Verossimilhança sob todo o espaço paramétrico

Neste caso

$$\begin{aligned} L(\beta, \lambda, \mathbf{Y}) &= \prod_{i=1}^g \prod_{j=1}^{n_i} \frac{1}{2\lambda_i} \exp\left(-\frac{|Y_{ij} - \mathbf{X}_{ij}\beta|}{\lambda_i}\right) = \prod_{i=1}^g (2\lambda_i)^{-n_i} \exp\left(-\frac{1}{\lambda_i} \sum_{j=1}^{n_i} |Y_{ij} - \mathbf{X}_{ij}\beta|\right), \\ &= 2^{-N} \prod_{i=1}^g (\lambda_i)^{-n_i} \cdot \exp\left(-\sum_{i=1}^g \frac{1}{\lambda_i} \sum_{j=1}^{n_i} |Y_{ij} - \mathbf{X}_{ij}\beta|\right). \end{aligned}$$

Diferentemente do logaritmo da função de verossimilhança sob H_0 , esta expressão requer o uso de algoritmos de otimização iterativos para sua maximização, pois não há como dissociar a estimação do vetor de parâmetros β da estimação do vetor de parâmetros λ . Doravante $\tilde{\beta}$ e $\tilde{\lambda}$ serão os estimadores de máxima verossimilhança de β e λ sob todo o espaço paramétrico.

Para obtermos estes estimadores precisaremos maximizar esta função com relação aos parâmetros β e $\lambda = (\lambda_1, \dots, \lambda_g)^\top$. Para facilitar este trabalho consideraremos o logaritmo natural da função de verossimilhança. Nessas condições

$$l(\beta, \lambda, \mathbf{Y}) = \ln(L(\beta, \lambda, \mathbf{Y})) = -N \ln(2) - \sum_{i=1}^g n_i \ln(\lambda_i) - \sum_{i=1}^g \frac{1}{\lambda_i} \sum_{j=1}^{n_i} |Y_{ij} - \mathbf{X}_{ij}\beta|. \quad (3.3)$$

A literatura da regressão L_1 aponta diversos algoritmos baseados em programação linear para obtenção de estimativas para os parâmetros β e λ sob H_0 (modelo homocedástico), como por exemplo os algoritmos de [Barrodale e Roberts \(1973, 1974\)](#). Entretanto, a literatura não aponta algoritmos específicos para obter os estimadores $\tilde{\lambda}$ e $\tilde{\beta}$. Por isso proporemos um procedimento iterativo para obtenção de tais estimadores baseado nos algoritmos para obtenção dos estimadores L_1 e nos algoritmos para estimação de mínimos quadrados ponderados.

Considere a Equação (3.3), embora não seja produtivo derivar o logaritmo da função de verossimilhança em relação a β (devido ao módulo) é relevante derivar em relação aos parâmetros λ_i , $i = 1, \dots, g$. Desta forma, teremos

$$\frac{\partial l(\beta, \lambda, \mathbf{Y})}{\partial \lambda_i} = -\frac{n_i}{\lambda_i} + \frac{1}{\lambda_i^2} \sum_{j=1}^{n_i} |Y_{ij} - \mathbf{X}_{ij}\beta|, \quad i = 1, \dots, g.$$

Se igualarmos esta derivada a zero e aplicarmos o estimador $\tilde{\beta}$ à derivada obteremos

$$\frac{\partial l(\beta, \lambda, \mathbf{Y})}{\partial \lambda_i} = 0 \Leftrightarrow -\frac{n_i}{\lambda_i} + \frac{1}{\lambda_i^2} \sum_{j=1}^{n_i} |Y_{ij} - \mathbf{X}_{ij}\tilde{\beta}| = 0 \Rightarrow \tilde{\lambda}_i = \frac{\sum_{j=1}^{n_i} |Y_{ij} - \mathbf{X}_{ij}\tilde{\beta}|}{n_i}. \quad (3.4)$$

Assim, o estimador $\tilde{\lambda}$ pode ser expresso em termos do estimador $\tilde{\beta}$ de forma similar à Equação (1.5).

A partir da Equação (3.3), para obter o estimador de máxima verossimilhança $\tilde{\beta}$ basta maximizar

$$\left[-\sum_{i=1}^g \frac{1}{\lambda_i} \sum_{j=1}^{n_i} |Y_{ij} - \mathbf{X}_{ij}\beta| \right], \quad \text{com relação a } \beta.$$

Isso mostra que $\tilde{\beta}$ é basicamente um estimador L_1 do vetor β ponderado pelos componentes do vetor de parâmetros de escala λ . Desta forma, para obtermos $\tilde{\beta}$ é necessário conhecermos $\tilde{\lambda}$ e vice-versa.

Partindo da relação entre os estimadores $\tilde{\lambda}$ e $\tilde{\beta}$ dada em (3.4), para concluirmos nosso algoritmo iterativo precisamos de um valor inicial como estimativa do parâmetro β . Esse valor inicial pode ser o próprio estimador de mínima soma dos erros absolutos ordinário de β , isto é, $\hat{\beta}$.

A seguir definimos o algoritmo para estimação de $\tilde{\beta}$ e $\tilde{\lambda}$.

1. Utilize o estimador $\hat{\beta}$, obtido a partir dos dados, como estimativa no passo 1 para β , $\tilde{\beta}_{(1)}$.
2. A partir da relação $\tilde{\lambda}_{i(1)} = \frac{\sum_{j=1}^{n_i} |Y_{ij} - \mathbf{X}_{ij}\tilde{\beta}_{(1)}|}{n_i}$, $i = 1, \dots, g$ obtenha o estimador de passo 1 para λ , $\tilde{\lambda}_{(1)}$.
3. Resolva o problema de maximização $\max_{\beta^{(k)}} \left[- \sum_{i=1}^g \frac{1}{\tilde{\lambda}_{i(k-1)}} \sum_{j=1}^{n_i} |Y_{ij} - \mathbf{X}_{ij}\beta^{(k)}| \right]$, a solução deste problema será o novo estimador $\tilde{\beta}_{(k)}$, o estimador de máxima verossimilhança de β no passo k .
4. A partir da relação $\tilde{\lambda}_{i(k)} = \frac{\sum_{j=1}^{n_i} |Y_{ij} - \mathbf{X}_{ij}\tilde{\beta}_{(k)}|}{n_i}$, $i = 1, \dots, g$, obtenha o novo estimador $\tilde{\lambda}_{(k)}$.
5. Repita os passos 3 e 4 até que um nível adequado de convergência seja atingido pelos estimadores $\tilde{\beta}_{(k)}$ e $\tilde{\lambda}_{(k)}$, assim, $\tilde{\beta} = \tilde{\beta}_{(k)}$ e $\tilde{\lambda} = \tilde{\lambda}_{(k)}$.

No algoritmo, optamos por não apresentar um critério específico de convergência, isso fica a critério de quem for estimar os parâmetros. Contudo, salientamos que o nosso código em [R Core Team \(2023\)](#) estabelece critérios razoáveis para convergência deste algoritmo.

Estatística do Teste da Razão de Verossimilhanças para testar homocedasticidade

Nesta seção apresentaremos a fórmula para o cálculo da estatística do teste da razão de verossimilhanças para testar homocedasticidade.

Lembrando que, no processo de estimação, após a convergência do algoritmo temos

$$\tilde{\lambda}_i = \frac{\sum_{j=1}^{n_i} |Y_{ij} - \mathbf{X}_{ij}\tilde{\beta}|}{n_i}, \quad i = 1, \dots, g \quad \text{e} \quad \hat{\lambda} = \frac{\sum_{j=1}^N |Y_j - \mathbf{X}_j\hat{\beta}|}{N},$$

podemos simplificar a estatística do teste.

Perceba que

$$\begin{aligned}
& l(\hat{\beta}, \hat{\lambda}, \mathbf{Y}) - l(\tilde{\beta}, \tilde{\lambda}, \mathbf{Y}) \\
&= -N \ln(2\hat{\lambda}) - \frac{1}{\hat{\lambda}} \sum_{j=1}^N |Y_j - \mathbf{X}_j \hat{\beta}| - \left[-N \ln(2) - \sum_{i=1}^g n_i \ln(\tilde{\lambda}_i) - \sum_{i=1}^g \frac{1}{\tilde{\lambda}_i} \sum_{j=1}^{n_i} |Y_{ij} - \mathbf{X}_{ij} \tilde{\beta}| \right] \\
&= -N \ln(\hat{\lambda}) - \frac{1}{\hat{\lambda}} \sum_{j=1}^N |Y_j - \mathbf{X}_j \hat{\beta}| + \sum_{i=1}^g n_i \ln(\tilde{\lambda}_i) + \sum_{i=1}^g \frac{1}{\tilde{\lambda}_i} \sum_{j=1}^{n_i} |Y_{ij} - \mathbf{X}_{ij} \tilde{\beta}| \\
&= -N \ln(\hat{\lambda}) - N + \sum_{i=1}^g n_i \ln(\tilde{\lambda}_i) + \sum_{i=1}^g n_i \\
&= -N \ln(\hat{\lambda}) - N + \sum_{i=1}^g n_i \ln(\tilde{\lambda}_i) + N \\
&= -N \ln(\hat{\lambda}) + \sum_{i=1}^g n_i \ln(\tilde{\lambda}_i) \\
&= \sum_{i=1}^g n_i \ln(\tilde{\lambda}_i / \hat{\lambda}).
\end{aligned}$$

Dessa forma, a estatística do teste de razão de verossimilhanças será

$$-2 (l(\hat{\beta}, \hat{\lambda}, \mathbf{Y}) - l(\tilde{\beta}, \tilde{\lambda}, \mathbf{Y})) = -2 \sum_{i=1}^g n_i \ln(\tilde{\lambda}_i / \hat{\lambda}), \quad (3.5)$$

adotaremos o nome LRT para a estatística dada na Expressão (3.5). LRT vem de *likelihood ratio test*, que significa teste da razão de verossimilhanças em Português.

É interessante notar que a estatística da Equação (3.5) é muito similar à estatística do teste da razão de verossimilhanças para verificar homogeneidade de variâncias no contexto da regressão linear com erros possuindo distribuição normal, cujo núcleo é apresentado em Johnston e DiNardo (1996, pg. 170), que está expresso na Equação (3.6).

$$-2 \left[N \cdot \ln(\hat{\sigma}^2) - \sum_{i=1}^g n_i \cdot \ln(\tilde{\sigma}_i^2) \right] = -2 \sum_{i=1}^g n_i \ln(\tilde{\sigma}_i / \hat{\sigma}) \stackrel{a}{\sim} \chi^2(g-1), \quad (3.6)$$

em que σ e σ_i , $i = 1, \dots, g$ são os parâmetros de escala da distribuição normal.

Um ponto importante sobre a estatística apresentada na Equação (3.5) é que ela não pode ser calculada de forma analítica, pois depende das estimativas $\hat{\beta}$, $\hat{\lambda}$, $\tilde{\beta}$ e $\tilde{\lambda}$. Entretanto, vale lembrar que também não é possível calcular a estatística do teste de homogeneidade de variâncias para regressão via mínimos quadrados de forma analítica, conforme mostrado por Johnston e DiNardo (1996, pg 170).

Outro ponto importante é que, sob H_0 , é esperado que o valor numérico da estatística

do teste proposto, expressa na Equação (3.5), se torne próximo de zero, pois neste caso $\lambda_1 = \lambda_2 = \dots = \lambda_g = \lambda$ e portanto é esperado que $\tilde{\lambda}_i \approx \hat{\lambda}$, $i = 1, \dots, g$, o que acarreta $\tilde{\lambda}_i/\hat{\lambda} \approx 1$ e conseqüentemente $\ln(\tilde{\lambda}_i/\hat{\lambda}) \approx 0$. Assim, quanto mais próximo de zero for o valor da estatística (3.5), mais forte é a evidência a favor de H_0 e a recíproca também é verdadeira.

Temos interesse em obter a distribuição da estatística do teste sob H_0 , mas como não sabemos a distribuição exata dos estimadores $\tilde{\beta}$ e $\tilde{\lambda}$, só podemos estudar a distribuição teórica assintótica ou a distribuição empírica da estatística em cenários controlados. Neste texto focaremos no estudo da distribuição empírica da estatística em cenários controlados.

As Seções 3.3, 3.4 e 3.5 apresentam resultados empíricos, via simulação, que dão suporte à ideia de que

$$-2 \sum_{i=1}^g n_i \ln(\tilde{\lambda}_i/\hat{\lambda}) \stackrel{a}{\approx} \chi^2(g-1), \quad (3.7)$$

isto é, a distribuição assintótica da estatística LRT sob H_0 é uma Qui-quadrado.

3.2 Estatísticas propostas para testar homogeneidade de variâncias

Nesta seção proporemos outras estatísticas para testar a hipótese de homogeneidade de variâncias e concorrer com a estatística obtida a partir da razão de verossimilhanças. As estatísticas propostas são a desenvolvida anteriormente, LRT, e outras duas, dadas nos itens 2 e 3 a seguir.

1. $LRT = -2 \sum_{i=1}^g n_i \ln(\tilde{\lambda}_i/\hat{\lambda})$.
2. $\max\text{-}\hat{\lambda} = \max_{i \in \{1, \dots, g\}} \tilde{\lambda}_i/\hat{\lambda}$;
3. $\max\text{-log-}\hat{\lambda} = \max_{i \in \{1, \dots, g\}} \ln(\tilde{\lambda}_i/\hat{\lambda})$;

As estatísticas “candidatas” 2 e 3 foram propostas de forma heurística, intuitivamente, sem formalismo estatístico.

Para estabelecer os valores críticos dos testes para homogeneidade de variâncias, fixaremos o nível de significância que chamaremos de α . Nos três casos, o teste terá região crítica unilateral à direita, pois grandes valores das estatísticas sugerem heterocedasticidade. Na estatística 1, o valor crítico será dado pelo quantil de ordem $0,95$ da distribuição Qui-quadrado com $g - 1$ graus de liberdade. Para as demais estatísticas os valores críticos serão estabelecidos via metodologia *bootstrap*, que será apresentada na Seção 3.2.

O que dificulta a obtenção das distribuições das estatísticas 2 e 3 sob H_0 é que os resultados aplicáveis à regressão L_1 são essencialmente assintóticos, isso inviabiliza obter as distribuições das estatísticas 2 e 3 sob H_0 de forma exata.

Bootstrap para as estatísticas propostas

O objetivo deste trecho é detalhar a metodologia para calcular valores-P com base em abordagens de reamostragem *bootstrap*, permitindo aplicar testes cuja distribuição exata ou assintótica sob H_0 são inacessíveis, como é o caso das estatísticas propostas já apresentadas. Vale lembrar que as condições de regularidade, para obtenção da distribuição assintótica da estatística LRT sob H_0 , não se aplicam no contexto do modelo linear Laplace, pois a derivada com relação aos parâmetros de posição na origem não está definida.

Desde sua introdução por Efron (1979), o método *bootstrap* se estabeleceu como uma ferramenta estatística fundamental, desempenhando papéis cruciais, como a construção de intervalos de confiança, a correção de vieses em estimativas e a realização de testes de hipóteses. A ideia central subjacente a esse método é tratar a amostra de dados como uma representação da própria população, obtendo, a partir dela, réplicas da amostra original. Isso pode ser feito por meio de duas abordagens principais: a reamostragem com reposição, denominada *bootstrap* não-paramétrico, ou o ajuste de um modelo aos dados, conhecido como *bootstrap* paramétrico.

No contexto em que conhecemos a distribuição assintótica da estatística do teste sob a hipótese nula, é conhecido que a distribuição limite da estatística do teste não é adequada para amostras pequenas (são necessários estudos de simulação para estabelecer o que delimita uma amostra “pequena” de uma amostra “não pequena”). Além disso, essa aproximação pode ser inadequada em amostras pequenas, levando a distorções na taxa de rejeição do teste sob H_0 . Uma alternativa é melhorar o desempenho do teste utilizando valores críticos obtidos via abordagem de reamostragem *bootstrap*. Assim, a regra de decisão do teste pode ser formulada em termos do valor-P *bootstrap*.

Quando o tamanho amostral é reduzido, a aplicação da distribuição assintótica da estatística de teste sob H_0 pode ser inapropriada, resultando em taxas distorcidas de rejeição da hipótese nula. Para contornar essa situação, a técnica de reamostragem *bootstrap* é empregada, permitindo a estimativa da distribuição nula da estatística de teste sem a necessidade de recorrer a uma aproximação assintótica.

Optamos por aplicar o *bootstrap* paramétrico em nossos estudos de simulação, pois conhecemos o modelo a ser gerado.

A seguir está o algoritmo *bootstrap* a ser utilizado para obter os quantis empíricos das estatísticas dos testes sob H_0 , e por consequência, os limitantes das regiões críticas dos testes.

- 1 Ajuste um modelo de regressão L_1 homocedástico ao conjunto de dados originais. Fixe um nível de significância α para o teste.

- 2 A partir das estimativas dos parâmetros do modelo de regressão L_1 homocedástico ajustado aos dados originais, simule aleatoriamente uma amostra *bootstrap* a partir da forma funcional do modelo, utilizando as estimativas dos parâmetros do modelo homocedástico como os valores dos parâmetros do modelo gerador da amostra *bootstrap*.
- 3 Refaça B vezes o passo 2, de forma a obter B amostras *bootstrap*.
- 4 Calcule a estatística do teste para cada uma das B amostras *bootstrap*, que satisfaz H_0 conforme descrito no passo 2.
- 5 Calcule o valor crítico para o teste por meio do quantil empírico de ordem α computado a partir da estatística do teste calculada B vezes, sob H_0 , no passo 4.
- 6 Calcule a estatística do teste a partir dos dados originais e verifique se é menor ou maior do que o quantil empírico de ordem α , calculado no passo anterior. Se for maior, rejeite a hipótese nula, se for menor, não rejeite a hipótese nula.

Note que para obter o valor crítico via *bootstrap* é necessário simular um modelo, que sabemos satisfazer H_0 . A partir dos quantis empíricos da estatística do teste calculados com base em reamostragens desse modelo calculamos o valor crítico para testar a hipótese H_0 com base nos dados que temos em mãos.

Observação 3.3 O valor B denota o número de réplicas *bootstrap* a serem utilizadas, quanto maior B mais preciso e confiável serão os resultados obtidos via *bootstrap*.

Simulações para verificar a distribuição empírica das estatísticas apresentadas

Encontram-se a seguir os resultados das simulações efetuadas para obtermos as distribuições empíricas das estatísticas apresentadas para testar a hipótese de homogeneidade de variâncias.

Utilizamos histogramas e QQ-plots para verificar se as distribuições empíricas encontradas são similares às distribuições “aproximadas” propostas. Para o mesmo propósito utilizamos medidas quantitativas que descrevem numericamente o grau de similaridade entre uma distribuição teórica e uma distribuição empírica obtida, como por exemplo, a função *fitdist()* do pacote *fitdistrplus* (Delignette-Muller *et al.*, 2023).

A função *fitdist()* permite ajustar uma distribuição teórica a um conjunto de dados por meio da estimação de máxima verossimilhança dos parâmetros da distribuição teórica a partir dos dados. Combinando graficamente o histograma gerado a partir da estatística do teste com a função densidade de probabilidade teórica, cujos parâmetros foram estimados via máxima verossimilhança, teremos uma abordagem gráfica para avaliar a similaridade entre a distribuição da estatística do teste e sua respectiva distribuição teórica

proposta. Essa combinação gráfica pode ser realizada por meio da função *denscomp()*, também pertencente ao pacote *fitdistrplus*. As Figuras 3.3, 3.4, 3.7, 3.8, 3.11 e 3.12 foram criadas por meio das funções *fitdist()* e *denscomp()*. Já as Figuras 3.5, 3.6, 3.9, 3.10, 3.13 e 3.14, que apresentam gráficos do tipo QQ-plot (gráfico que compara os quantis empíricos aos quantis teóricos), foram criadas por meio das funções *fitdist()*, *qqplot()* e *qq-line()*. Para mais detalhes consulte o manual do pacote *fitdistrplus* [Delignette-Muller et al. \(2023\)](#) ou o artigo original que apresenta o pacote [Delignette-Muller e Dutang \(2015\)](#).

Observação 3.4 Nas Figuras 3.3, 3.4, 3.7, 3.8, 3.11 e 3.12 optamos por examinar o histograma, mas há outras funções gráficas no pacote *fitdistrplus* que podem ser úteis, como:

- *qqcomp()* que gera o gráfico quantil-quantil (QQ-plot) entre os quantis empíricos contra os quantis teóricos;
- *cdfcomp()* que gera o gráfico da função de distribuição acumulada empírica contra a função de distribuição acumulada teórica;
- *ppcomp()* que gera o gráfico das probabilidades teóricas contra as probabilidades empíricas.

As simulações serão divididas em três partes:

1. simulação balanceada, na qual todos os cenários têm a mesma quantidade de grupos e os grupos contêm o mesmo número de observações;
2. simulação desbalanceada, na qual os cenários apresentados têm a mesma quantidade de grupos, mas os grupos contêm número de observações distintos;
3. simulação com diferentes números de grupos, em que os cenários apresentados têm número de grupos distintos, mas o número total de observações (somando as observações de todos os grupos) é o mesmo.

Essas simulações servirão para estudarmos o comportamento da distribuição de cada uma das estatísticas propostas, sob H_0 (modelo homocedástico), nos mais variados cenários que podem surgir em aplicações reais.

3.3 Detalhes da Simulação Balanceada

O modelo simulado será

$$Y_i = \beta_0 + \beta_1 \cdot X_i + \epsilon_i, \quad i = 1, 2, \dots, N,$$

em que N é o tamanho amostral. Dividiremos a amostra em grupos de mesmo tamanho.

Nas simulações, foram adotados $\beta_0 = 2$ e $\beta_1 = 3$ e $N = 100$ observações. Os 100 valores de x_i são os 100 números de 0 a 10 criados a partir do comando “seq(0,10, length.out =100)” do **R Core Team (2023)**. Os valores de λ para cada grupo serão variados e estão descritos na Tabela 3.1.

Observação 3.5 O comando “seq(0,10, length.out =100)” não gera os valores 0, 0,1 0,2 ... 10, pois para gerar esses valores seria necessário utilizar o comando “seq(0,10, length.out =101)”, mas neste último caso $N = 101$ e não $N = 100$. Optamos por simular $N = 100$ porque o número 100 tem como divisores 1, 2, 4, 5, 10, 20, 25, 50, e 100, enquanto 101 é um número primo e, conseqüentemente, só é divisível por 1 e 101. Então não seria possível dividir a amostra com 101 observações em grupos balanceados para realizar a primeira simulação.

Cenário	número de grupos	valores dos λ 's	Desvio Padrão λ	Coef. de Var. λ	Max Média λ	número de observações por grupo	N
1	5	2, 2, 2, 2, 2	0,000	0,000	1,000	20	100
2	5	5, 5, 5, 5, 5	0,000	0,000	1,000	20	100
3	5	10, 10, 10, 10, 10	0,000	0,000	1,000	20	100
4	5	100, 100, 100, 100, 100	0,000	0,000	1,000	20	100
5	5	1, 1,1, 1,2, 1,3, 1,4	0,158	0,132	1,167	20	100
6	5	1, 1,25, 1,5, 1,75, 2	0,395	0,264	1,333	20	100
7	5	1, 1,5, 2, 2,5, 3	0,791	0,395	1,500	20	100
8	5	2, 2,1, 2,2, 2,3, 2,4	0,158	0,072	1,091	20	100
9	5	2, 2,25, 2,5, 2,75, 3	0,395	0,158	1,200	20	100
10	5	2, 2,5, 3, 3,5, 4	0,791	0,264	1,333	20	100
11	5	5, 5,1, 5,2, 5,3, 5,4	0,158	0,030	1,038	20	100
12	5	5, 5,25, 5,5, 5,75, 6	0,395	0,072	1,091	20	100
13	5	5, 5,5, 6, 6,5, 7	0,791	0,132	1,167	20	100
14	5	10, 10,1, 10,2, 10,3, 10,4	0,158	0,016	1,020	20	100
15	5	10, 10,25, 10,5, 10,75, 11	0,395	0,038	1,048	20	100
16	5	10, 10,5, 11, 11,5, 12	0,791	0,072	1,091	20	100
17	5	100, 100,1, 100,2, 100,3, 100,4	0,158	0,002	1,002	20	100
18	5	100, 100,25, 100,5, 100,75, 101	0,395	0,004	1,005	20	100
19	5	100, 100,5, 101, 101,5, 102	0,791	0,008	1,010	20	100
20	5	1, 2, 3, 4, 5	1,581	0,527	1,667	20	100
21	5	2, 4, 6, 8, 10	3,162	0,527	1,667	20	100
22	5	3, 6, 9, 12, 15	4,743	0,527	1,667	20	100
23	5	1, 5, 7, 15, 30	11,480	0,990	2,586	20	100
24	5	1, 10, 25, 50, 70	28,543	0,915	2,244	20	100

Tabela 3.1: Dados da simulação balanceada para 24 cenários, com 5 grupos contendo 20 observações por grupo, totalizando 100 observações ($N = 100$). Foram geradas 1000 réplicas por cenário.

Cenário	Hipótese real	Média da Estatística e Valores Críticos ($\alpha = 5\%$)			Taxa de rejeição	
		LRT	crítico assintótico	crítico <i>bootstrap</i>	assintótico	<i>bootstrap</i>
1	H_0	4,152	9,488	9,925	0,065	0,052
2	H_0	4,143	9,488	9,912	0,053	0,049
3	H_0	4,016	9,488	9,909	0,050	0,043
4	H_0	4,269	9,488	9,931	0,071	0,056
5	H_a	5,560	9,488	9,912	0,129	0,109
6	H_a	10,365	9,488	9,912	0,474	0,450
7	H_a	18,286	9,488	9,923	0,876	0,856
8	H_a	4,672	9,488	9,930	0,079	0,073
9	H_a	6,141	9,488	9,937	0,175	0,159
10	H_a	10,134	9,488	9,921	0,479	0,458
11	H_a	4,299	9,488	9,905	0,077	0,065
12	H_a	4,516	9,488	9,917	0,076	0,057
13	H_a	5,631	9,488	9,916	0,161	0,143
14	H_a	4,211	9,488	9,922	0,060	0,049
15	H_a	4,191	9,488	9,937	0,057	0,051
16	H_a	4,479	9,488	9,935	0,083	0,075
17	H_a	4,278	9,488	9,918	0,065	0,050
18	H_a	4,239	9,488	9,919	0,056	0,041
19	H_a	4,350	9,488	9,932	0,072	0,062
20	H_a	31,797	9,488	9,909	0,997	0,997
21	H_a	32,199	9,488	9,913	0,999	0,999
22	H_a	32,180	9,488	9,909	0,999	0,998
23	H_a	108,479	9,488	9,933	1,000	1,000
24	H_a	146,308	9,488	9,916	1,000	1,000

Tabela 3.2: LRT – Resultado da simulação balanceada para 24 cenários, com 5 grupos contendo 20 observações por grupo, totalizando 100 observações ($N = 100$).

Observação 3.6 Informações complementares para Tabela 3.2: 1000 réplicas foram geradas por cenário. Para cada uma dessas réplicas, 1000 réplicas *bootstrap* foram geradas para obter os valores críticos *bootstrap* correspondentes. A média da estatística foi computada com base nessas 1000 réplicas. O valor crítico assintótico foi o quantil de ordem 0,95 de uma distribuição $\chi^2(4)$. O valor crítico *bootstrap* foi obtido pelo quantil de ordem 0,95 da estatística LRT calculada nas 1000 réplicas *bootstrap*, gerando um valor crítico *bootstrap* para cada uma das 1000 réplicas originais. Na coluna “crítico *bootstrap*”, foi calculada a média dos 1000 valores críticos *bootstrap* para cada cenário.

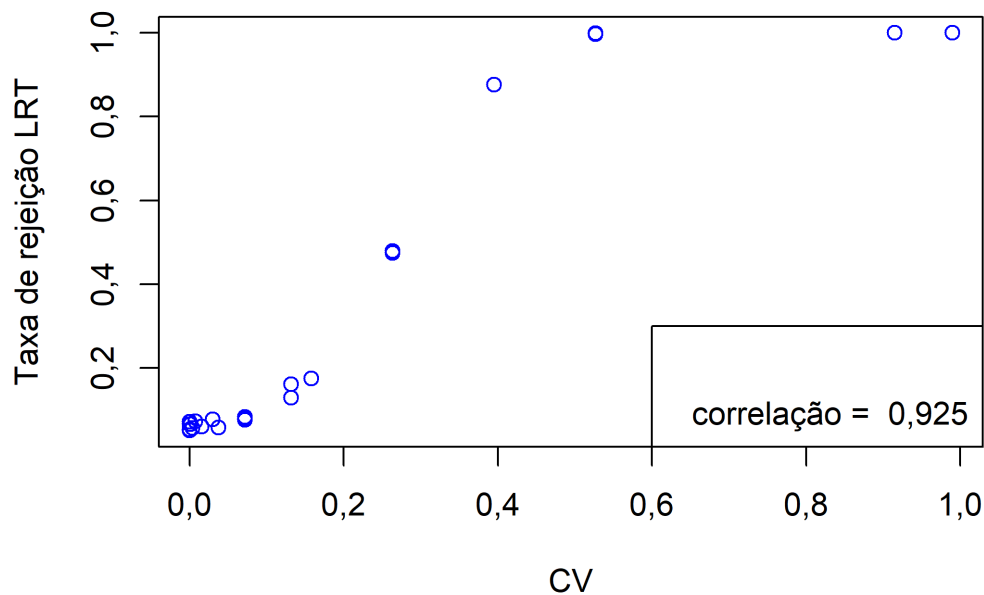
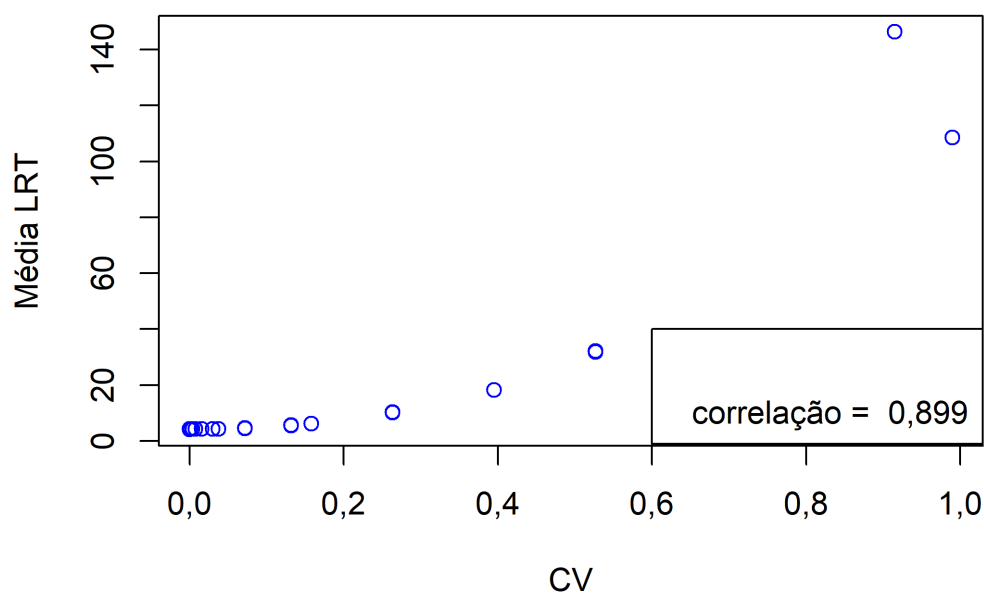
Cenário	Hipótese real	Média das Estatísticas e Valores Críticos ($\alpha = 5\%$)				Taxa de rejeição <i>bootstrap</i>	
		$\max-\hat{\lambda}$	crítico <i>bootstrap</i>	$\max-\log-\hat{\lambda}$	crítico <i>bootstrap</i>	$\max-\hat{\lambda}$	$\max-\log-\hat{\lambda}$
1	H_0	1,273	1,516	0,237	0,416	0,044	0,044
2	H_0	1,273	1,515	0,237	0,416	0,042	0,042
3	H_0	1,272	1,516	0,236	0,416	0,048	0,048
4	H_0	1,283	1,516	0,245	0,416	0,060	0,060
5	H_a	1,329	1,515	0,279	0,415	0,120	0,120
6	H_a	1,446	1,516	0,361	0,416	0,314	0,314
7	H_a	1,580	1,516	0,449	0,416	0,585	0,585
8	H_a	1,299	1,515	0,256	0,416	0,070	0,070
9	H_a	1,345	1,516	0,290	0,416	0,152	0,152
10	H_a	1,445	1,515	0,361	0,416	0,326	0,326
11	H_a	1,281	1,515	0,243	0,416	0,054	0,054
12	H_a	1,291	1,516	0,251	0,416	0,065	0,065
13	H_a	1,333	1,516	0,281	0,416	0,127	0,127
14	H_a	1,280	1,516	0,243	0,416	0,047	0,047
15	H_a	1,283	1,516	0,244	0,416	0,052	0,052
16	H_a	1,292	1,516	0,250	0,416	0,077	0,077
17	H_a	1,280	1,516	0,242	0,416	0,047	0,047
18	H_a	1,280	1,515	0,242	0,415	0,056	0,056
19	H_a	1,281	1,516	0,243	0,416	0,050	0,050
20	H_a	1,731	1,515	0,541	0,416	0,834	0,834
21	H_a	1,737	1,515	0,544	0,416	0,841	0,841
22	H_a	1,734	1,515	0,543	0,415	0,846	0,846
23	H_a	2,586	1,516	0,942	0,416	1,000	1,000
24	H_a	2,302	1,516	0,827	0,416	1,000	1,000

Tabela 3.3: $\max-\hat{\lambda}$ e $\max-\log-\hat{\lambda}$ – Resultado da simulação balanceada para 24 cenários, com 5 grupos contendo 20 observações por grupo, totalizando 100 observações ($N = 100$).

Observação 3.7 Informações complementares para Tabela 3.3: 1000 réplicas foram geradas por cenário. Para cada uma dessas réplicas, 1000 réplicas *bootstrap* foram geradas para obter os valores críticos *bootstrap* correspondentes. A média de cada estatística foi computada com base nessas 1000 réplicas. Os valores críticos *bootstrap* foram obtidos pelo quantil de ordem 0,95 das estatísticas $\max-\hat{\lambda}$ e $\max-\log-\hat{\lambda}$ calculadas nas 1000 réplicas *bootstrap*.

CV	Cenário	Hipótese real	média LRT	taxa de rejeição LRT (assint.)	taxa de rejeição LRT (boots.)
0,000	1	H_0	4,152	0,065	0,052
0,000	2	H_0	4,143	0,053	0,049
0,000	3	H_0	4,016	0,050	0,043
0,000	4	H_0	4,269	0,071	0,056
0,002	17	H_a	4,278	0,065	0,050
0,004	18	H_a	4,239	0,056	0,041
0,008	19	H_a	4,350	0,072	0,062
0,016	14	H_a	4,211	0,060	0,049
0,030	11	H_a	4,299	0,077	0,065
0,038	15	H_a	4,191	0,057	0,051
0,072	8	H_a	4,672	0,079	0,073
0,072	12	H_a	4,516	0,076	0,057
0,072	16	H_a	4,479	0,083	0,075
0,132	5	H_a	5,560	0,129	0,109
0,132	13	H_a	5,631	0,161	0,143
0,158	9	H_a	6,141	0,175	0,159
0,264	6	H_a	10,365	0,474	0,450
0,264	10	H_a	10,134	0,479	0,458
0,395	7	H_a	18,286	0,876	0,856
0,527	20	H_a	31,797	0,997	0,997
0,527	21	H_a	32,199	0,999	0,999
0,527	22	H_a	32,180	0,999	0,998
0,915	24	H_a	146,308	1,000	1,000
0,990	23	H_a	108,479	1,000	1,000

Tabela 3.4: Resultado da simulação balanceada, relação entre coeficiente de variação (CV) e estatística LRT, incluindo média e taxa de rejeição. A taxa de rejeição foi calculada a partir de dois valores críticos: um obtido pela distribuição $\chi^2(4)$ (distribuição assintótica da estatística sugerida ao fim das simulações) e outro via bootstrap.

Figura 3.1: Diagrama de dispersão entre coeficiente de variação (CV) e média da estatística LRT.**Figura 3.2:** Diagrama de dispersão entre coeficiente de variação (CV) e taxa de rejeição da estatística LRT.

As Tabelas 3.1 e 3.2 trazem duas valiosas informações acerca dos testes propostos. A primeira informação é que o desvio padrão (DP) dos λ 's é uma péssima medida

para mensurar o grau de heterocedasticidade dos cenários, enquanto o coeficiente de variação (CV) e o quociente $\frac{\text{Max}}{\text{Média}} \lambda$ (Max/Média) são boas medidas unidimensionais do grau de heterocedasticidade dos cenários. Isso ficará nítido mais adiante, quando veremos que de fato a estatística LRT tem capacidade de discriminar entre cenários homocedásticos (ou aproximadamente homocedásticos) e heterocedásticos, pois o teste baseado na estatística LRT apresenta desempenhos distintos para diferentes cenários que possuem o mesmo desvio padrão para os λ s. Por exemplo, os cenários 5, 8, 11 e 14 (desvio padrão = 0,158) possuem o mesmo desvio padrão, mas a média da estatística LRT com base nas réplicas, bem como a taxa de rejeição de H_0 , tende a cair conforme percorre-se os cenários 5, 8, 11, 14, nesta ordem. A mesma observação pode ser feita com relação ao grupo de cenários 6, 9, 12 e 15 (desvio padrão = 0,395) e ao grupo de cenários 7, 10, 13 e 16 (desvio padrão = 0,791). Comentário análogo é aplicável a Tabela 3.3.

Por outro lado, tanto CV quanto Max/Média são úteis como medida descritiva amostral da heterocedasticidade. Em um cenário prático, os valores populacionais dos λ 's são desconhecidos, então bastaria calcular CV e Max/Média a partir das estimativas dos λ 's. Pensando de forma mais ampla, em contextos fora da regressão L_1 , podemos usar CV e Max/Média como medidas de heterocedasticidade a partir de estimativas de parâmetros de escala. A utilidade da medida CV fica mais nítida na Tabela 3.4 e nas Figuras 3.1 e 3.2. Na Tabela 3.4 há uma relação monótona crescente entre CV e média da LRT, isso fica ainda mais claro na Figura 3.1. Os mesmos comentários se aplicam à relação entre CV e taxa de rejeição da LRT, basta consultar a Figura 3.2 e a Tabela 3.4. É notável que, em ambas as relações, a correlação fica em torno de 0,9, o que indica uma forte relação linear crescente. Por isso, nas próximas simulações não calcularemos o desvio padrão dos λ 's.

A segunda informação é que uma boa estatística para testar heterocedasticidade deve fornecer taxa de rejeição crescente em termos das medidas CV e Max/Média. Por isso prosseguiremos com a análise usando a medida CV como norteadora, examinaremos com detalhes as taxas de rejeição das estatísticas na Tabela 3.4: que apresenta os valores CV e taxa de rejeição.

Fica claro a partir das Tabelas 3.3 e 3.5 que as estatísticas $\max-\hat{\lambda}$ e $\max-\log-\hat{\lambda}$ têm a mesma taxa de rejeição para os vinte quatro cenários, o que é razoável considerando que $\max-\log-\hat{\lambda}$ é o logaritmo natural de $\max-\hat{\lambda}$. Assim, a partir deste ponto do capítulo seguiremos somente com uma destas duas estatísticas e optamos por escolher seguindo com $\max-\hat{\lambda}$. A razão para esta decisão é que $\max-\hat{\lambda}$ é uma estatística um pouco mais simples e de forte apelo intuitivo.

CV	Cenário	Hipótese real	Taxas de rejeição			
			LRT (assint.)	LRT (boots.)	max- $\hat{\lambda}$	max-log- $\hat{\lambda}$
0,000	1	H_0	0,065	0,052	0,044	0,044
0,000	2	H_0	0,053	0,049	0,042	0,042
0,000	3	H_0	0,050	0,043	0,048	0,048
0,000	4	H_0	0,071	0,056	0,060	0,060
0,002	17	H_a	0,065	0,050	0,047	0,047
0,004	18	H_a	0,056	0,041	0,056	0,056
0,008	19	H_a	0,072	0,062	0,050	0,050
0,016	14	H_a	0,060	0,049	0,047	0,047
0,030	11	H_a	0,077	0,065	0,054	0,054
0,038	15	H_a	0,057	0,051	0,052	0,052
0,072	8	H_a	0,079	0,073	0,070	0,070
0,072	12	H_a	0,076	0,057	0,065	0,065
0,072	16	H_a	0,083	0,075	0,077	0,077
0,132	5	H_a	0,129	0,109	0,120	0,120
0,132	13	H_a	0,161	0,143	0,127	0,127
0,158	9	H_a	0,175	0,159	0,152	0,152
0,264	6	H_a	0,474	0,450	0,314	0,314
0,264	10	H_a	0,479	0,458	0,326	0,326
0,395	7	H_a	0,876	0,856	0,585	0,585
0,527	20	H_a	0,997	0,997	0,834	0,834
0,527	21	H_a	0,999	0,999	0,841	0,841
0,527	22	H_a	0,999	0,998	0,846	0,846
0,915	24	H_a	1,000	1,000	1,000	1,000
0,990	23	H_a	1,000	1,000	1,000	1,000

Tabela 3.5: Comparativo entre as taxas de rejeição das estatísticas LRT, max- $\hat{\lambda}$ e max-log- $\hat{\lambda}$ na simulação balanceada.

O comportamento ideal para uma estatística que será utilizada para verificar a hipótese de homocedasticidade é uma taxa de rejeição próxima ao nível de significância fixado, neste caso $\alpha = 5\%$, quando H_0 é verdadeira e alta taxa de rejeição quando H_0 é falsa. A característica que destacará a “melhor” estatística é o aumento da taxa de rejeição à medida que CV aumenta.

Primeiramente, a Tabela 3.5 destaca para os cenários homocedásticos – cenários 1, 2, 3 e 4 – que todas as estatísticas apresentam taxa de rejeição próxima de 0,05. Mas a estatística LRT usando valor crítico fornecido pelo resultado assintótico forneceu, em três dos quatro cenários homocedásticos, taxa de rejeição superior ao valor 5% predefinido, sendo uma das taxas de rejeição superior a 7%. Isso se deve ao conhecido fato de que

resultados assintóticos não se adequam perfeitamente em contexto de amostras finitas, mas sim fornecem aproximações úteis para aplicações práticas. Sendo assim, parece aceitável para aplicações um erro do tipo I variando entre 0,050 e 0,071 quando se fixa o nível de significância em 0,05. Uma alternativa imediata é usar o valor crítico obtido via *bootstrap*, que fornece erro do tipo I variando entre 0,043 e 0,056 para a estatística LRT.

Esgotada a análise dos cenários homocedásticos, passamos a analisar os cenários heterocedásticos. Os cinco cenários com maiores CV – cenários 20, 21, 22, 23 e 24 – apresentaram as maiores taxas de rejeição. Vale destacar que a estatística LRT fornece taxas de rejeição que variam de 0,997 a 1 nestes cenários, enquanto a estatística $\max-\hat{\lambda}$ fornece taxas de rejeição que variam de 0,834 a 1, evidenciando que a estatística LRT é mais poderosa que $\max-\hat{\lambda}$ para contextos nos quais CV está entre 0,527 e 0,990. Para valor de CV maior ou igual a 1, a simulação sugere que os testes são igualmente poderosos, pois a taxa de rejeição ficará 1 para as estatísticas. Contudo, são necessários mais estudos de simulação, abrangendo diversos outros cenários, para confirmar a assertiva mencionada.

Para os cenários em que CV está entre 0,132 e 0,395 a estatística LRT apresenta taxa de rejeição consistentemente superior à estatística $\max-\hat{\lambda}$, ainda que com pequena margem para valores de CV próximos de 0,132. Já para os cenários em que CV está entre 0,002 e 0,072 a diferença entre as taxas de rejeição das estatísticas é muito pequena.

Um ponto importante a destacar é que a estatística LRT utilizando o valor crítico obtido assintoticamente é, de forma consistente, levemente mais poderoso que a mesma estatística utilizando o valor crítico obtido via *bootstrap*. A conclusão sobre usar o valor crítico obtido por meio da distribuição assintótica ou via *bootstrap* para a estatística LRT é a seguinte: usar o valor crítico assintótico aumenta o poder do teste em troca do aumento do erro do tipo I (além do valor de α fixado), já o valor crítico *bootstrap* não viola o nível de significância fixado e apresenta essencialmente o mesmo poder do teste quando $CV \geq 0,527$.

Como recomendação geral, é mais conservador utilizar o valor crítico *bootstrap* para o teste com base na estatística LRT, pela garantia de que o erro do tipo I prefixado será respeitado. Além disso, é esperado que quanto menor o tamanho amostral pior a qualidade da aproximação assintótica, assim o valor crítico *bootstrap* deve ser preferível para contextos em que há amostra pequena.

Assim, dentre todas as estatísticas apresentadas, a estatística LRT com valor crítico *bootstrap* deve ser utilizada em aplicações práticas em detrimento das estatísticas $\max-\hat{\lambda}$, $\max-\log-\hat{\lambda}$ e LRT com valor crítico assintótico.

As Figuras 3.3 e 3.4 mostram os histogramas da estatística LRT nos 24 cenários simulados bem como o ajuste da distribuição χ^2 com parâmetro estimado via máxima verossimilhança a partir dos valores da estatística LRT em cada cenário. As Figuras 3.5

e 3.6 mostram os QQ-plots correspondentes às Figuras 3.3 e 3.4. Percebemos que nos cenários homocedásticos a estimativa dos graus de liberdade fica próxima de 4, o que evidencia empiricamente que, assintoticamente, a distribuição da estatística do teste se aproxima da χ^2 com 4 graus de liberdade. Conforme dito antes da Seção 3.3, as Figuras 3.3 e 3.4, bem como as Figuras 3.7, 3.8, 3.11 e 3.12 foram criadas por meio das funções *fitdist()* e *denscomp()* do pacote *fitdistrplus* (Delignette-Muller *et al.*, 2023). Assim a curva em vermelho nestas figuras é a curva da função densidade de probabilidade da Qui-quadrado cujos graus de liberdade foram estimados via máxima verossimilhança a partir das 1000 estatísticas LRT computadas a partir das 1000 réplicas da simulação.

Esse resultado, destacado nas Figuras 3.3 e 3.4, é positivo, pois fornece evidências empíricas de que a distribuição da estatística do LRT é $\chi^2(g - 1)$, em que g denota o número de grupos. É bem conhecido que o teste da razão de verossimilhanças para testar a hipótese de homocedasticidade no modelo linear normal possui distribuição assintótica $\chi^2(g - 1)$, o que torna o resultado empírico obtido muito reconfortante.

Por outro lado, nos cenários em que há alto grau de heterocedasticidade, a estimativa de máxima verossimilhança do parâmetro correspondente aos graus de liberdade da χ^2 fica bem elevada. Como a distribuição Qui-quadrado converge para a distribuição normal quando os graus de liberdade tendem ao infinito então, sob a hipótese H_a , quando há alto nível de heterocedasticidade, conjecturamos que a distribuição da estatística LRT tenha distribuição normal. Entretanto, não sabemos qual a relação funcional entre o grau de heterocedasticidade e os parâmetros da distribuição normal, apenas percebemos que a média e a variância da distribuição normal são funções crescentes do grau de heterocedasticidade. Por enquanto não dispomos de meios analíticos para obter a distribuição assintótica da estatística LRT sob H_a , o que permitiria estudar analiticamente o poder do teste baseado na estatística LRT, mas esse é um tópico aberto para futuras pesquisas na área de regressão L_1 .

Nas próximas duas partes do estudo de simulação, considerando as informações da primeira parte da simulação discutidas anteriormente, procederemos computando somente as estatística LRT e $\max\hat{\lambda}$, examinaremos o desempenho destas estatísticas em cenários mais desafiadores.

Figura 3.3: Histogramas LRT, Parte 1 da simulação, Cenários 1 a 12.

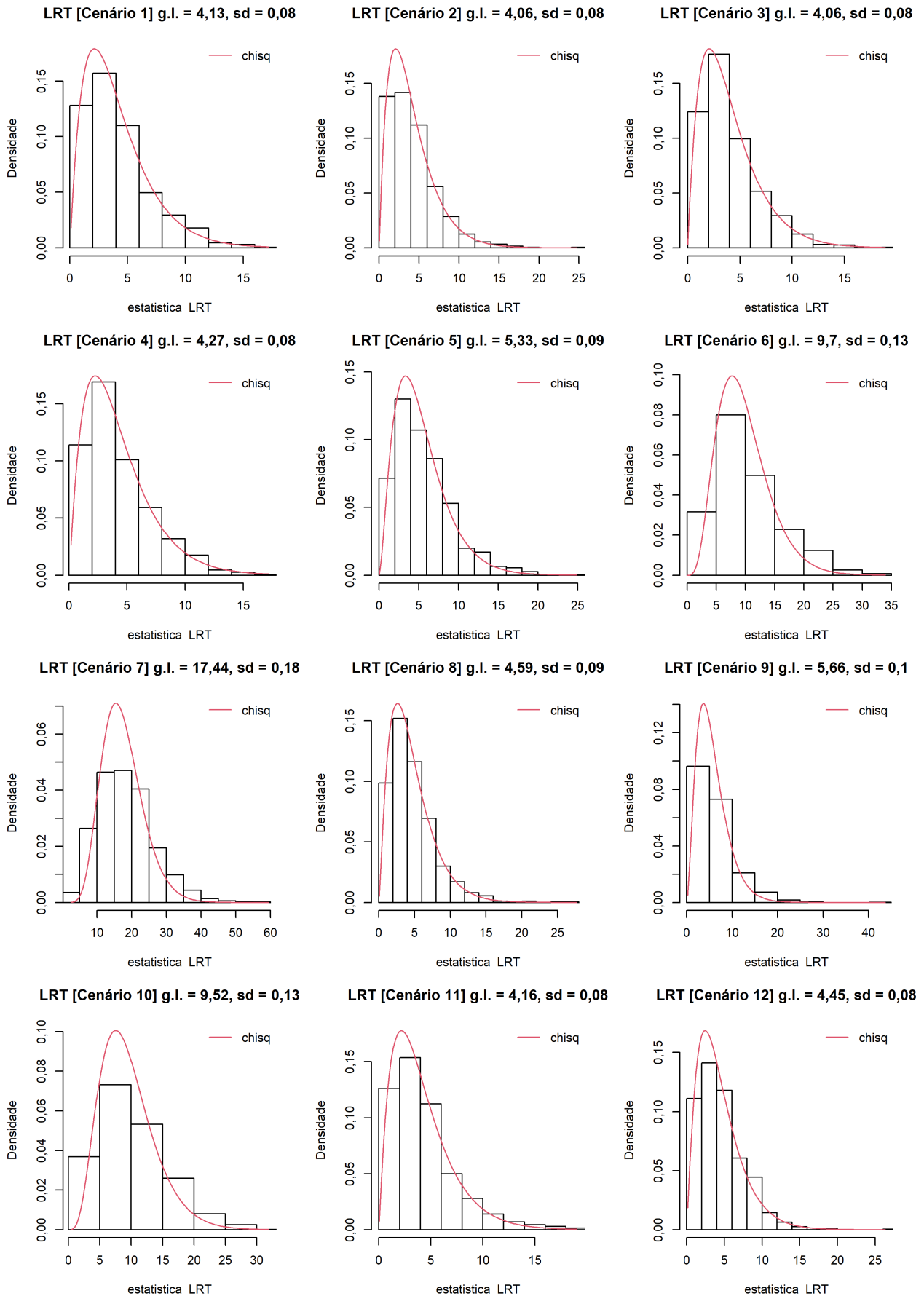


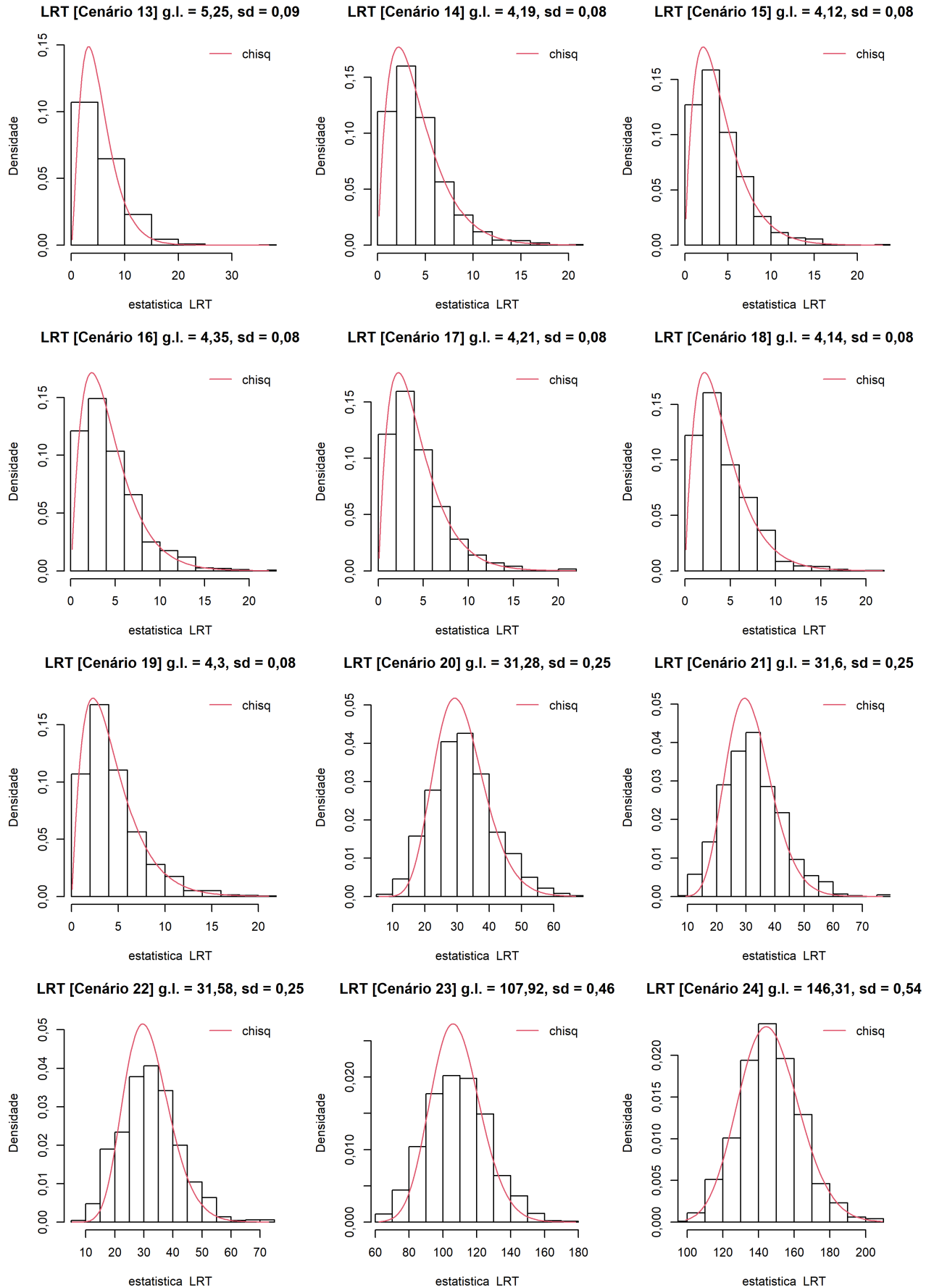
Figura 3.4: Histogramas LRT, Parte 1 da simulação, Cenários 13 a 24.

Figura 3.5: QQ-plots LRT, Parte 1 da simulação, Cenários 1 a 12.

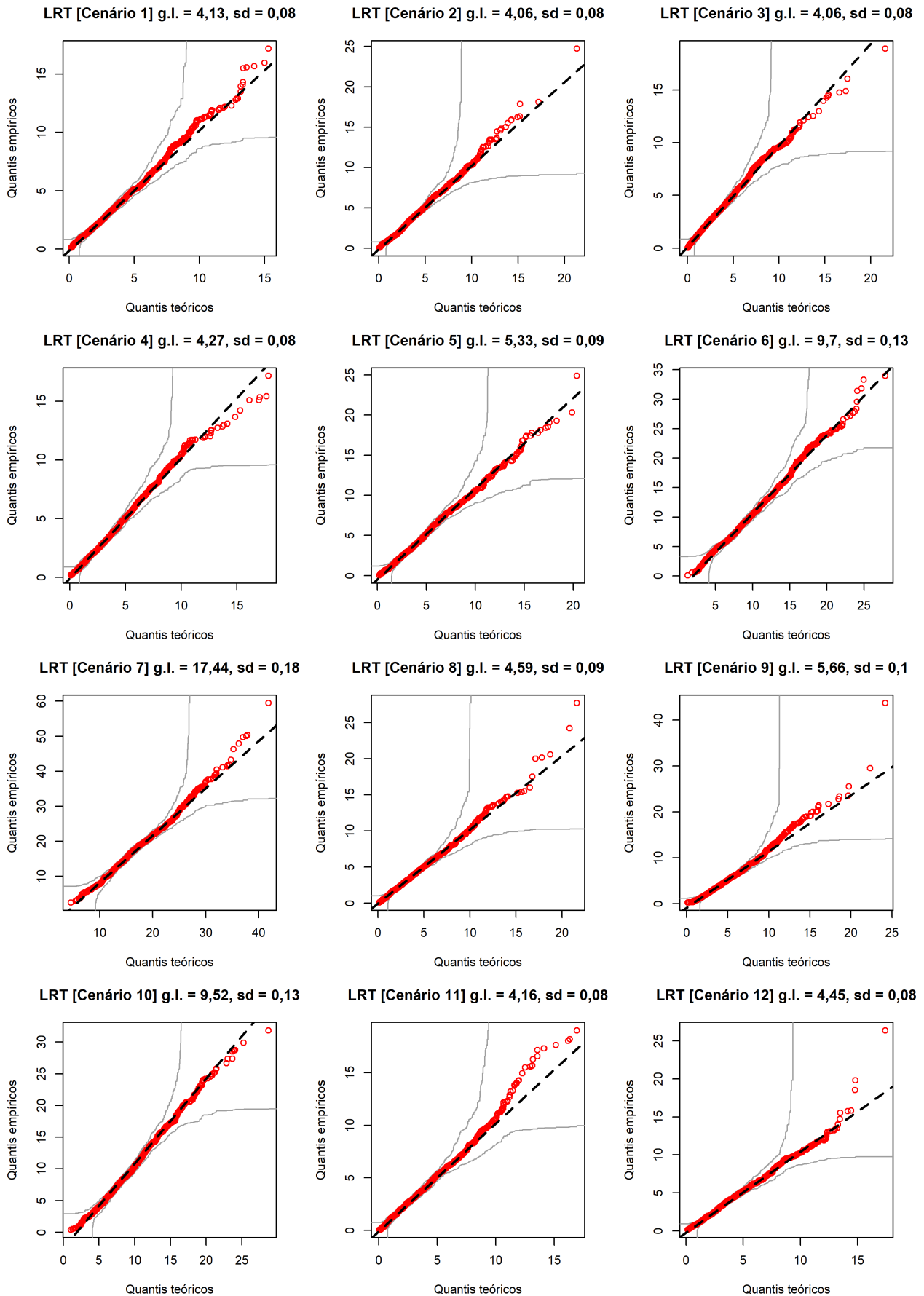
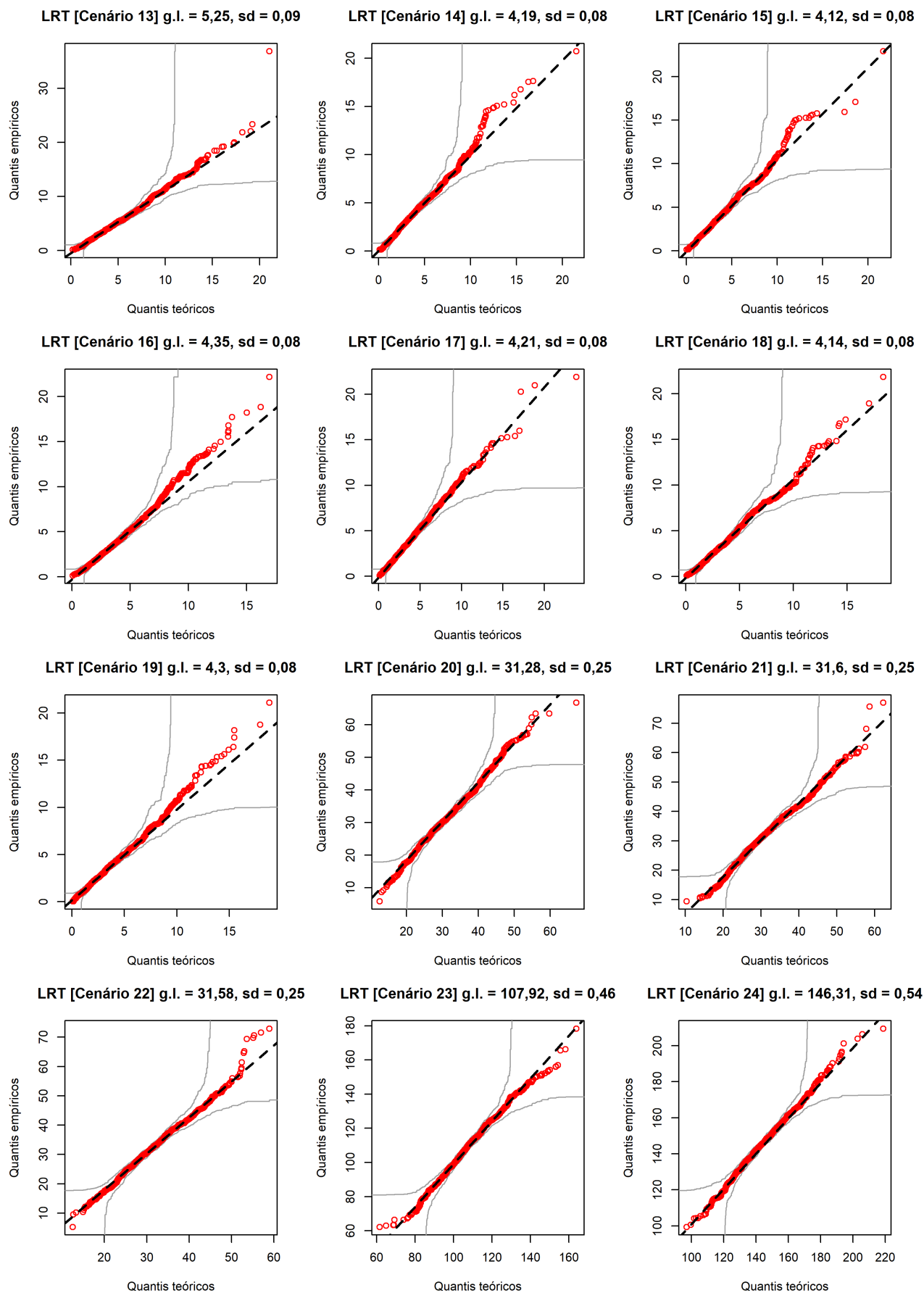


Figura 3.6: *QQ-plots LRT, Parte 1 da simulação, Cenários 13 a 24.*

3.4 Detalhes da Simulação Desbalanceada

Esta simulação utilizará grupos de tamanhos distintos para verificar o desempenho das estatísticas LRT e $\max\text{-}\hat{\lambda}$ em cenários desbalanceados, ou seja, com grupos contendo números distintos de observações. Os objetivos desta simulação são examinar o desempenho das estatísticas LRT e $\max\text{-}\hat{\lambda}$ em cenários desbalanceados e verificar o quanto a distribuição da estatística LRT é afetada em cenários desbalanceados, em especial se a distribuição $\chi^2(g - 1)$ continua adequada. Aqui relembramos a fórmula da estatística LRT.

$$\text{LRT} = -2 \sum_{i=1}^g n_i \ln(\tilde{\lambda}_i / \hat{\lambda}),$$

perceba que a própria estatística do teste leva em consideração o tamanho amostral dos grupos por meio das quantidades n_i , $i = 1, \dots, g$.

A Tabela 3.6 contém os cenários da parte 2 da simulação. Nessa tabela os 12 primeiros cenários são homocedásticos com crescente grau de desbalanceamento, os 12 cenários seguintes são heterocedásticos com crescente grau de desbalanceamento.

A Tabela 3.7 mostra que a existência de grupos com diferentes números de observações não afeta o desempenho do teste baseado na estatística LRT. Já as taxas de rejeição da estatística $\max\text{-}\hat{\lambda}$ para os cenários 13 a 24, heterocedásticos, são consideravelmente menores que as taxas de rejeição da estatística LRT.

A Figura 3.7 atesta que a distribuição empírica da estatística do teste sob H_0 continua próxima à distribuição Qui-quadrado dos cenários balanceados homocedásticos. A Figura 3.8 mostra os histogramas para os cenários heterocedásticos. As Figuras 3.9 e 3.10 mostram os QQ-plots correspondentes às Figuras 3.7 e 3.8. Considerando as tabelas e figuras da parte 2 da simulação, é notável que a existência de grupos desbalanceados não afeta a performance do teste baseado na estatística LRT e a distribuição empírica da estatística LRT nos cenários desbalanceados continua similar a dos casos balanceados. Já a estatística $\max\text{-}\hat{\lambda}$ apresentou poder bem menor que a estatística LRT. Em resumo, conclui-se que os desempenhos das estatísticas LRT e $\max\text{-}\hat{\lambda}$ não são afetados nem prejudicados por dados desbalanceados, e fica evidente que os testes baseados na estatística LRT também são mais poderosos que o teste com base na estatística $\max\text{-}\hat{\lambda}$ em contextos desbalanceados.

Cenário	observações por grupo	valores dos λ 's	CV λ	$\frac{\text{Max}}{\text{Média}} \lambda$	N
1	20, 20, 20, 20, 20	5, 5, 5, 5, 5	0	1	100
2	21, 19, 21, 19, 20	5, 5, 5, 5, 5	0	1	100
3	22, 18, 20, 20, 20	5, 5, 5, 5, 5	0	1	100
4	22, 18, 21, 19, 20	5, 5, 5, 5, 5	0	1	100
5	22, 18, 22, 18, 20	5, 5, 5, 5, 5	0	1	100
6	23, 17, 20, 20, 20	5, 5, 5, 5, 5	0	1	100
7	23, 17, 22, 18, 20	5, 5, 5, 5, 5	0	1	100
8	23, 17, 23, 17, 20	5, 5, 5, 5, 5	0	1	100
9	24, 16, 20, 20, 20	5, 5, 5, 5, 5	0	1	100
10	24, 16, 23, 17, 20	5, 5, 5, 5, 5	0	1	100
11	30, 10, 25, 15, 20	5, 5, 5, 5, 5	0	1	100
12	35, 5, 30, 10, 20	5, 5, 5, 5, 5	0	1	100
13	20, 20, 20, 20, 20	1, 2, 3, 4, 5	0,527	1,667	100
14	21, 19, 21, 19, 20	1, 2, 3, 4, 5	0,527	1,667	100
15	22, 18, 20, 20, 20	1, 2, 3, 4, 5	0,527	1,667	100
16	22, 18, 21, 19, 20	1, 2, 3, 4, 5	0,527	1,667	100
17	22, 18, 22, 18, 20	1, 2, 3, 4, 5	0,527	1,667	100
18	23, 17, 20, 20, 20	1, 2, 3, 4, 5	0,527	1,667	100
19	23, 17, 22, 18, 20	1, 2, 3, 4, 5	0,527	1,667	100
20	23, 17, 23, 17, 20	1, 2, 3, 4, 5	0,527	1,667	100
21	24, 16, 20, 20, 20	1, 2, 3, 4, 5	0,527	1,667	100
22	24, 16, 23, 17, 20	1, 2, 3, 4, 5	0,527	1,667	100
23	30, 10, 25, 15, 20	1, 2, 3, 4, 5	0,527	1,667	100
24	35, 5, 30, 10, 20	1, 2, 3, 4, 5	0,527	1,667	100

Tabela 3.6: Dados da simulação desbalanceada para 24 cenários, com 5 grupos de diferentes tamanhos amostrais, totalizando 100 observações ($N = 100$). Foram geradas 1000 réplicas por cenário.

Cenário	Hipótese real	Média da Estatística e Valores Críticos ($\alpha = 5\%$)					Taxa de rejeição			
		LRT	crít. assintótico	LRT	crít. <i>bootstrap</i>	LRT	max- $\hat{\lambda}$	crít. max- $\hat{\lambda}$	LRT assintótico	LRT <i>bootstrap</i>
1	H_0	4,152	9,488	9,926	9,926	1,273	1,516	0,065	0,057	0,043
2	H_0	4,127	9,488	9,895	9,895	1,272	1,517	0,054	0,046	0,043
3	H_0	3,994	9,488	9,907	9,907	1,271	1,518	0,049	0,040	0,046
4	H_0	4,275	9,488	9,906	9,906	1,285	1,519	0,070	0,058	0,055
5	H_0	4,058	9,488	9,906	9,906	1,274	1,521	0,050	0,041	0,042
6	H_0	4,130	9,488	9,904	9,904	1,281	1,521	0,053	0,047	0,049
7	H_0	4,313	9,488	9,910	9,910	1,283	1,524	0,067	0,056	0,053
8	H_0	4,207	9,488	9,913	9,913	1,282	1,528	0,053	0,044	0,049
9	H_0	4,131	9,488	9,905	9,905	1,277	1,526	0,069	0,058	0,056
10	H_0	4,296	9,488	9,921	9,921	1,287	1,532	0,052	0,047	0,059
11	H_0	4,280	9,488	9,895	9,895	1,308	1,613	0,074	0,063	0,064
12	H_0	4,145	9,488	10,030	10,030	1,354	1,867	0,056	0,048	0,037
13	H_a	32,432	9,488	9,931	9,931	1,745	1,516	0,999	0,998	0,830
14	H_a	33,518	9,488	9,915	9,915	1,758	1,517	0,999	0,998	0,851
15	H_a	33,169	9,488	9,915	9,915	1,744	1,519	0,999	0,997	0,838
16	H_a	33,792	9,488	9,906	9,906	1,752	1,518	0,999	0,999	0,858
17	H_a	33,392	9,488	9,904	9,904	1,766	1,520	0,998	0,997	0,863
18	H_a	34,322	9,488	9,917	9,917	1,752	1,522	0,996	0,996	0,848
19	H_a	34,580	9,488	9,909	9,909	1,773	1,524	0,999	0,999	0,852
20	H_a	33,580	9,488	9,917	9,917	1,769	1,527	0,998	0,997	0,849
21	H_a	35,055	9,488	9,924	9,924	1,760	1,527	1,000	1,000	0,860
22	H_a	34,709	9,488	9,903	9,903	1,777	1,532	0,999	0,999	0,860
23	H_a	39,288	9,488	9,932	9,932	1,855	1,614	1,000	1,000	0,825
24	H_a	42,046	9,488	10,039	10,039	1,939	1,869	1,000	1,000	0,556

Tabela 3.7: LRT e max- $\hat{\lambda}$ – Resultado da simulação desbalanceada para 24 cenários, com 5 grupos de diferentes tamanhos amostrais, totalizando 100 observações ($N = 100$).

Observação 3.8 Informações complementares para Tabela 3.7: 1000 réplicas foram geradas por cenário. Para cada uma dessas réplicas, 1000 réplicas *bootstrap* foram geradas para obter os valores críticos *bootstrap* correspondentes. A média de cada estatística foi computada com base nessas 1000 réplicas. O valor crítico assintótico para LRT foi o quantil de ordem 0,95 de uma distribuição $\chi^2(4)$. Os valores críticos *bootstrap* foram obtidos pelo quantil de ordem 0,95 das estatísticas LRT e max- $\hat{\lambda}$ calculadas nas 1000 réplicas *bootstrap*, gerando um valor crítico *bootstrap* para cada uma das 1000 réplicas originais. Na colunas “crít. *bootstrap* LRT” e “crít. max- $\hat{\lambda}$ ”, são apresentadas as médias dos 1000 valores críticos *bootstrap* para cada cenário.

Figura 3.7: Histogramas LRT, Parte 2 da simulação, Cenários 1 a 12 (casos homocedásticos).

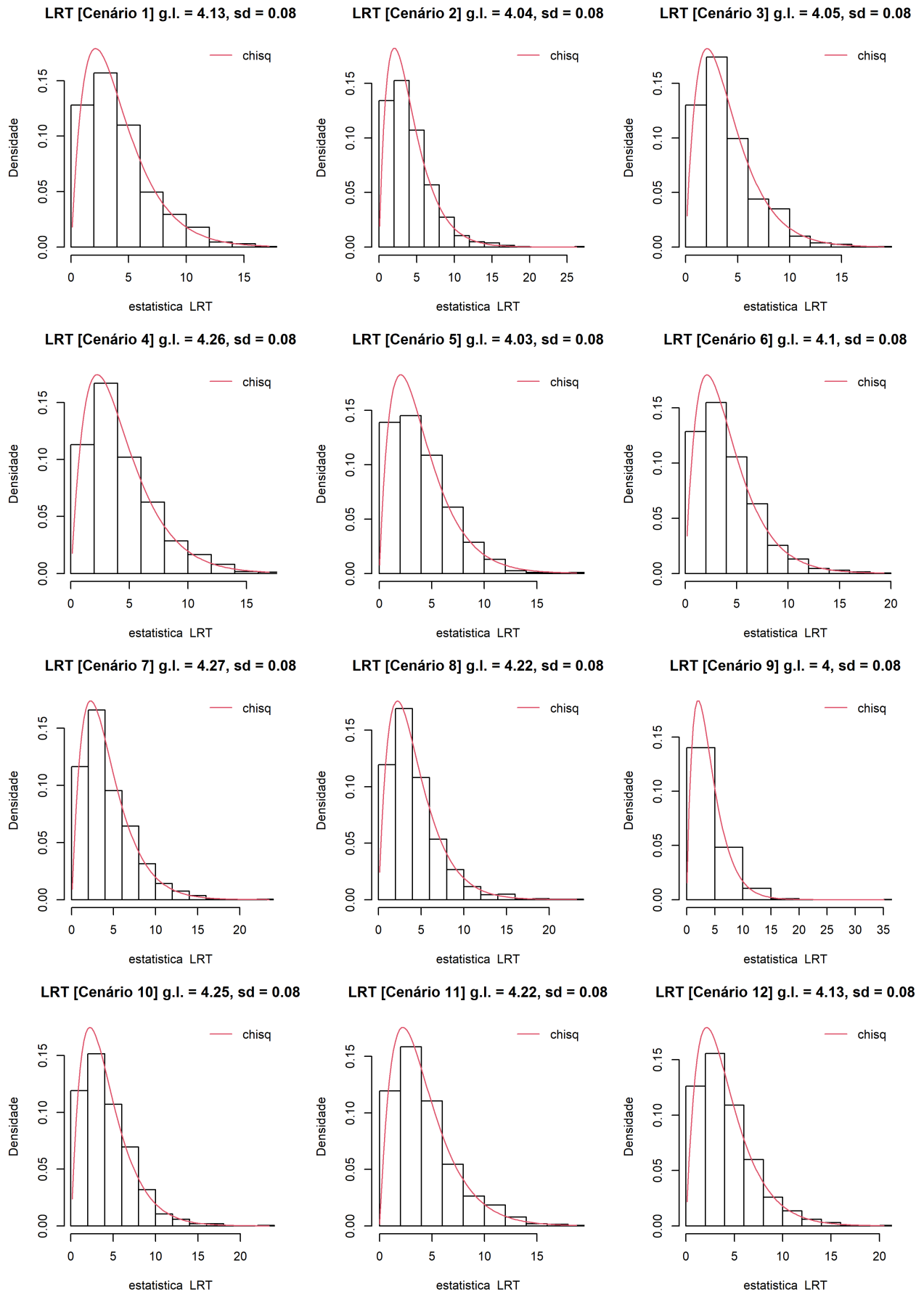


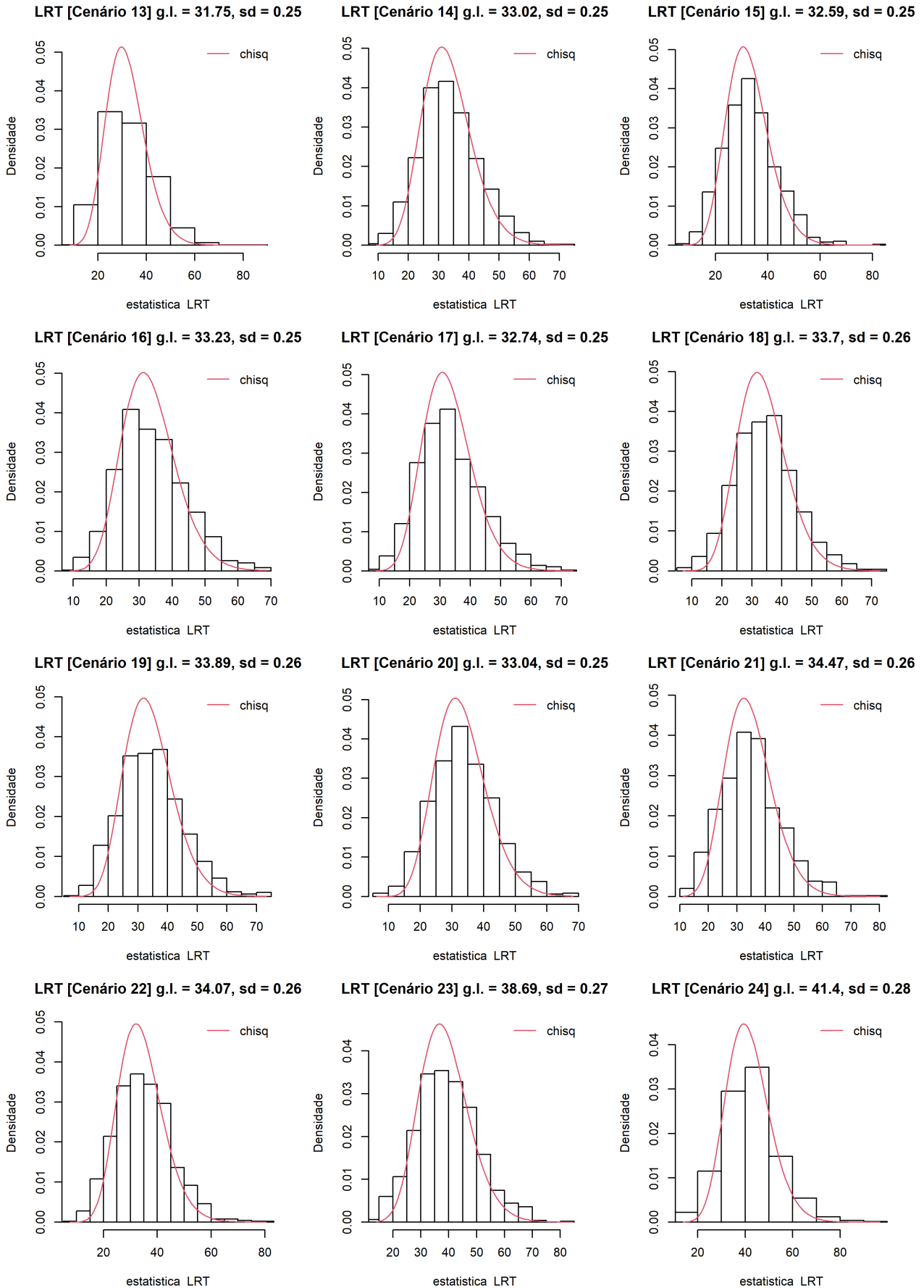
Figura 3.8: Histogramas LRT, Parte 2 da simulação, Cenários 13 a 24 (casos heterocedásticos).

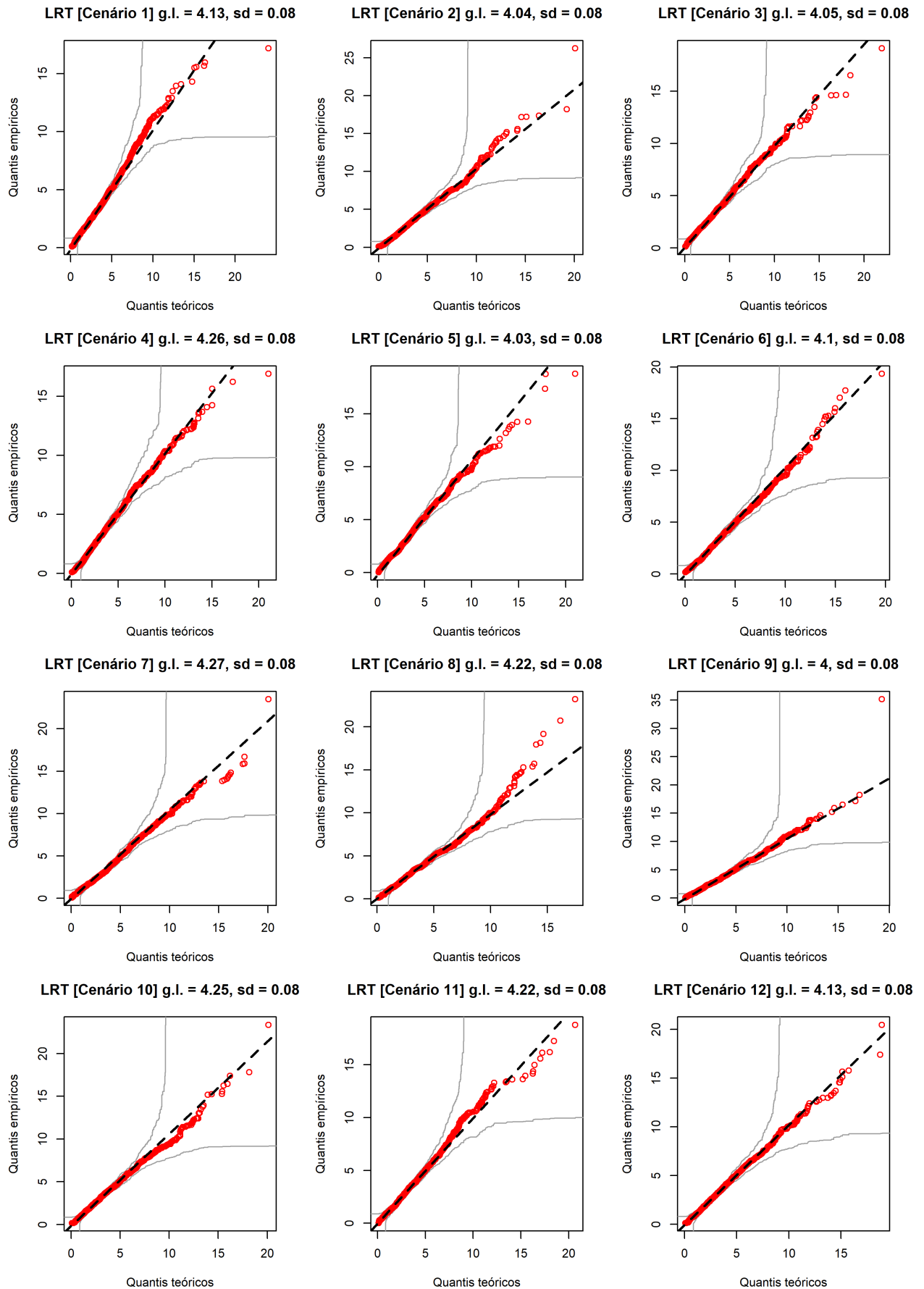
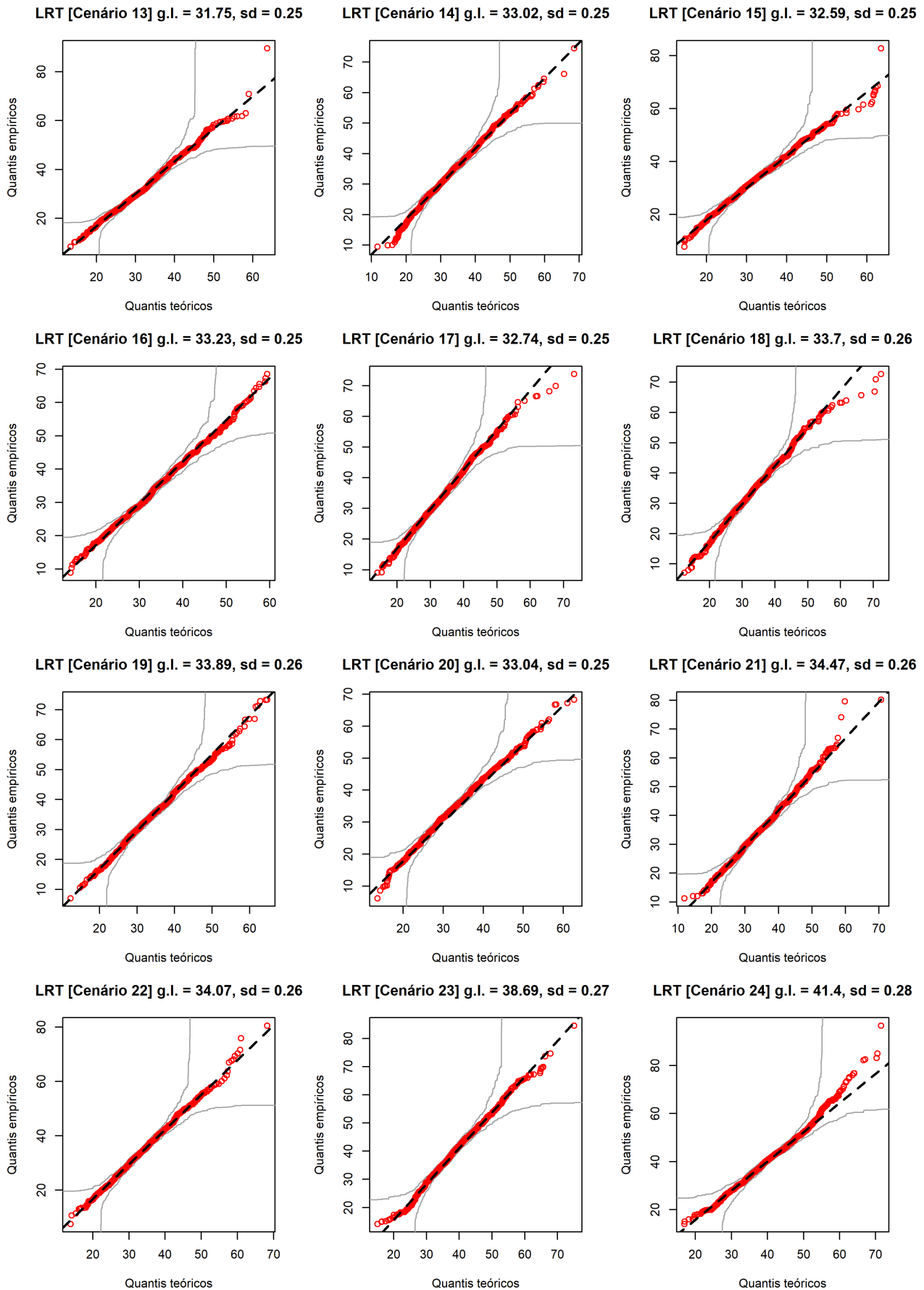
Figura 3.9: *QQ-plots LRT, Parte 2 da simulação, Cenários 1 a 12 (casos homocedásticos).*

Figura 3.10: *QQ-plots LRT, Parte 2 da simulação, Cenários 13 a 24 (casos heterocedásticos).*

3.5 Detalhes da Simulação com diferentes números de grupos

Esta simulação utilizará dados com diferentes números de grupos com mesmo número de observações para verificar o desempenho das estatísticas LRT e $\max\hat{\lambda}$ em cenários com poucas e com muitas observações por grupo, ou seja, com poucos ou muitos grupos. Os objetivos desta simulação são verificar o desempenho das estatísticas LRT e $\max\hat{\lambda}$ em cenários com diferentes números de grupos e pesquisar o quanto a distribuição da estatística LRT, sob H_0 , é afetada em cenários em que há muitos grupos (poucas observações por grupo) ou em que há poucos grupos (muitas observações por grupo), em especial se a aproximação pela distribuição $\chi^2(g - 1)$ continua adequada. Aqui relembramos a fórmula da estatística LRT.

$$\text{LRT} = -2 \sum_{i=1}^g n_i \ln(\tilde{\lambda}_i / \hat{\lambda}),$$

perceba que a própria estatística do teste leva em consideração o número de grupos por meio da quantidade g que denota o número de grupos.

A Tabela 3.8 contém os cenários da parte 3 da simulação. Os cenários são divididos em três tipos: homocedásticos (cenários de 1 a 8), levemente heterocedásticos (cenários de 9 a 16), heterocedásticos (cenários de 17 a 24). O número de grupos nos cenários de 1 a 8 são, respectivamente, 2, 3, 4, 8, 10, 15, 20, 30. Esse padrão de número de grupos é repetido nos cenários de 9 a 16 e nos cenários de 17 a 24.

Observação 3.9 Na Tabela 3.8 há a notação “seq($a, b, \text{length.out} = m$)”, que significa a sequência de m números equidistantes entre os números reais a e b (inclusive). Essa notação é utilizada no software *R Core Team (2023)* para criar sequências com as propriedades descritas. Por meio dessa notação podemos representar as sequências de números dentro da tabela, de forma que tal tabela caiba em uma única página.

Para facilitar a compreensão ilustraremos a notação com um exemplo. Considere “seq(4, 5, length.out = 10)”, que significa a sequência de 10 números equidistantes entre os números reais 4 e 5, ou de forma explícita: 4, 4,111111, 4,222222, 4,333333, 4,444444, 4,555556, 4,666667, 4,777778, 4,888889 e 5; o que totaliza os dez números da sequência. A praticidade dessa notação torna-se mais evidente ao utilizarmos “seq(1, 5, length.out = 30)” em vez de escrever os trinta números de forma explícita.

A Tabela 3.9 mostra que, quando o número de grupos se eleva e conseqüentemente há poucas observações por grupo, o teste tem sua performance prejudicada, aumentando a probabilidade do erro do tipo II (e a probabilidade do erro do tipo I no caso do teste com base na estatística LRT usando o valor crítico assintótico) dos testes.

Considerando o teste com base na estatística LRT usando o valor crítico assintótico,

analisando os cenários homocedásticos, em particular, os cenários 6 (15 grupos com 8 observações), 7 (20 grupos com 6 observações) e 8 (30 grupos com 4 observações) observamos que a probabilidade do erro do tipo I aumentou para além do fixado ($\alpha = 0,05$). Analisando os cenários 22 (15 grupos com 8 observações), 23 (20 grupos com 6 observações) e 24 (30 grupos com 4 observações), que são heterocedásticos, percebemos que o poder do teste cai a medida que o número de observações por grupo diminui.

Na simulação, quanto maior o número de grupos, menor o número de observações por grupo. Mas vale destacar que o aumento da probabilidade do erro tipo I e diminuição do poder do teste deve-se ao baixo número de observações por grupo e não ao número de grupos em si. Nos cenários 8, 16 e 24 estimamos o parâmetro de escala de cada modelo a partir de apenas 4 observações, o que caracteriza uma situação crítica (usar 4 observações para estimar dois parâmetros de posição e um parâmetro de escala) para qualquer teste de heterocedasticidade. Quando temos mais grupos, temos mais parâmetros λ_i a estimar e menos observações por grupo para obter essas estimativas.

O teste LRT usando o valor crítico assintótico é inadequado quando o número de observações por grupo fica abaixo de 8, pois nesta situação a taxa do erro do tipo I fica acima do nível de significância estabelecido. Examine as taxas de rejeição para LRT usando o valor crítico assintótico nos cenários 6 e 7 da Tabela 3.9, note que a probabilidade do erro do tipo I passa de 0,083 a 0,088.

Restam somente o teste LRT usando o valor crítico *bootstrap* e o teste utilizando $\max\hat{\lambda}$. Mas as taxas de rejeição dos testes LRT (*bootstrap*) e $\max\hat{\lambda}$ nos cenários 20, 21, 22, 23 e 24 da Tabela 3.9 indicam que o teste com base em $\max\hat{\lambda}$ é menos poderoso que seu concorrente, teste LRT (*bootstrap*).

As Figuras 3.11 e 3.12 atestam que a distribuição empírica da estatística do teste sob H_0 continua próxima à distribuição Qui-quadrado com $g - 1$ graus de liberdade. As Figuras 3.13 e 3.14 mostram os QQ-plots correspondentes às Figuras 3.11 e 3.12. Considerando as tabelas e figuras da parte 3 da simulação, notamos que desde que o número de observações por grupo não seja inferior a 8, o desempenho do teste LRT (assintótico) não será prejudicado.

Cenário	número de grupos	valores dos λ 's	CV λ	$\frac{\text{Max}}{\text{Média}} \lambda$	N
1	2	$\lambda = 5$ para todos os grupos	0	1	120
2	3	$\lambda = 5$ para todos os grupos	0	1	120
3	4	$\lambda = 5$ para todos os grupos	0	1	120
4	8	$\lambda = 5$ para todos os grupos	0	1	120
5	10	$\lambda = 5$ para todos os grupos	0	1	120
6	15	$\lambda = 5$ para todos os grupos	0	1	120
7	20	$\lambda = 5$ para todos os grupos	0	1	120
8	30	$\lambda = 5$ para todos os grupos	0	1	120
9	2	seq(4, 5, length.out = 2)	0,157	1,111	120
10	3	seq(4, 5, length.out = 3)	0,111	1,111	120
11	4	seq(4, 5, length.out = 4)	0,096	1,111	120
12	8	seq(4, 5, length.out = 8)	0,078	1,111	120
13	10	seq(4, 5, length.out = 10)	0,075	1,111	120
14	15	seq(4, 5, length.out = 15)	0,071	1,111	120
15	20	seq(4, 5, length.out = 20)	0,069	1,111	120
16	30	seq(4, 5, length.out = 30)	0,067	1,111	120
17	2	seq(1, 5, length.out = 2)	0,943	1,667	120
18	3	seq(1, 5, length.out = 3)	0,667	1,667	120
19	4	seq(1, 5, length.out = 4)	0,574	1,667	120
20	8	seq(1, 5, length.out = 8)	0,467	1,667	120
21	10	seq(1, 5, length.out = 10)	0,449	1,667	120
22	15	seq(1, 5, length.out = 15)	0,426	1,667	120
23	20	seq(1, 5, length.out = 20)	0,415	1,667	120
24	30	seq(1, 5, length.out = 30)	0,405	1,667	120

Tabela 3.8: Dados da simulação para 24 cenários com diferentes números de grupos. Cada grupo contém o mesmo número de observações, 120/número de grupos, totalizando 120 observações ($N = 120$). Foram geradas 1000 réplicas por cenário.

Cenário	Hipótese real	Média da Estatística e Valores Críticos ($\alpha = 5\%$)					Taxa de rejeição		
		LRT	crít. assintótico	LRT	crít. <i>bootstrap</i>	LRT	max- $\hat{\lambda}$	crít. max- $\hat{\lambda}$	LRT assintótico
1	H_0	0,991	3,841	3,910	1,073	1,180	0,054	0,049	0,049
2	H_0	1,872	5,991	6,145	1,130	1,286	0,042	0,037	0,037
3	H_0	3,023	7,815	8,064	1,194	1,380	0,051	0,044	0,047
4	H_0	7,357	14,067	14,745	1,408	1,696	0,061	0,049	0,056
5	H_0	9,509	16,919	17,816	1,503	1,836	0,057	0,045	0,052
6	H_0	15,009	23,685	25,225	1,714	2,160	0,083	0,052	0,052
7	H_0	20,657	30,144	32,528	1,913	2,463	0,088	0,045	0,045
8	H_0	32,576	42,557	47,200	2,326	3,035	0,126	0,057	0,054
9	H_a	2,520	3,841	3,902	1,120	1,179	0,233	0,231	0,231
10	H_a	3,127	5,991	6,138	1,171	1,287	0,133	0,123	0,108
11	H_a	3,894	7,815	8,071	1,224	1,380	0,103	0,095	0,091
12	H_a	8,075	14,067	14,723	1,431	1,696	0,095	0,072	0,075
13	H_a	10,223	16,919	17,805	1,521	1,835	0,102	0,078	0,071
14	H_a	15,357	23,685	25,214	1,735	2,162	0,092	0,064	0,067
15	H_a	21,047	30,144	32,556	1,942	2,465	0,086	0,056	0,059
16	H_a	32,318	42,557	47,201	2,323	3,037	0,116	0,059	0,061
17	H_a	71,382	3,841	3,906	1,671	1,179	1,000	1,000	1,000
18	H_a	49,218	5,991	6,153	1,667	1,287	1,000	1,000	0,997
19	H_a	41,034	7,815	8,067	1,679	1,381	1,000	1,000	0,956
20	H_a	35,421	14,067	14,715	1,860	1,696	0,992	0,992	0,698
21	H_a	36,470	16,919	17,798	1,973	1,836	0,985	0,978	0,637
22	H_a	39,007	23,685	25,210	2,194	2,159	0,933	0,901	0,480
23	H_a	43,972	30,144	32,494	2,446	2,464	0,913	0,867	0,413
24	H_a	54,695	42,557	47,183	2,883	3,036	0,856	0,719	0,335

Tabela 3.9: LRT e max- $\hat{\lambda}$ – Resultado da simulação para 24 cenários com diferentes números de grupos. Cada grupo contém o mesmo número de observações, 120/número de grupos, totalizando 120 observações ($N = 120$).

Observação 3.10 Informações complementares para Tabela 3.9: 1000 réplicas foram geradas por cenário. Para cada uma dessas réplicas, 1000 réplicas *bootstrap* foram geradas para obter os valores críticos *bootstrap* correspondentes. A média de cada estatística foi computada com base nessas 1000 réplicas. O valor crítico assintótico para LRT foi o quantil de ordem 0,95 de uma distribuição $\chi^2(4)$. Os valores críticos *bootstrap* foram obtidos pelo quantil de ordem 0,95 das estatísticas LRT e max- $\hat{\lambda}$ calculadas nas 1000 réplicas *bootstrap*, gerando um valor crítico *bootstrap* para cada uma das 1000 réplicas originais. Na colunas “crít. *bootstrap* LRT” e “crít. max- $\hat{\lambda}$ ”, são apresentadas as médias dos 1000 valores críticos *bootstrap* para cada cenário.

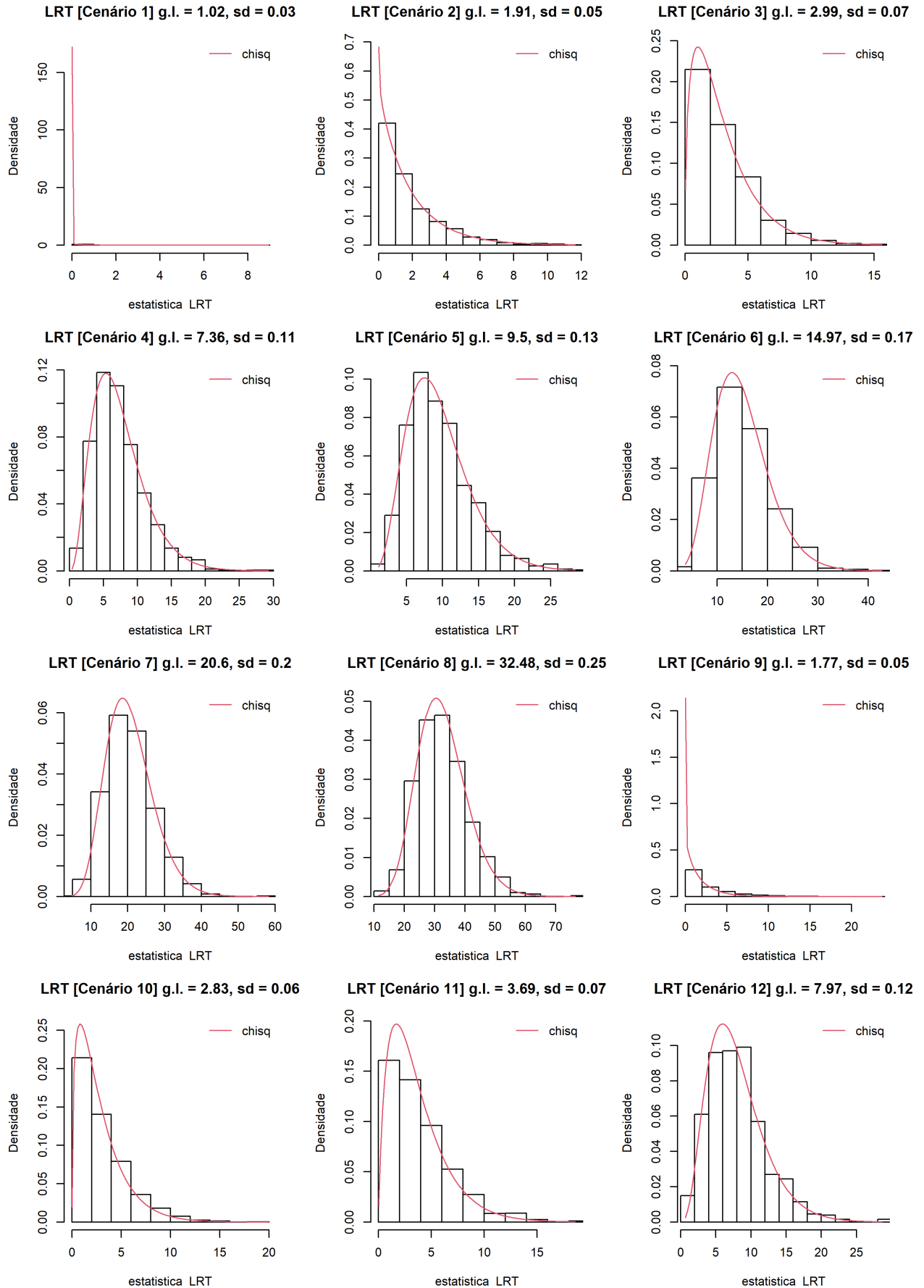
Figura 3.11: Histogramas LRT, Parte 3 da simulação, Cenários 1 a 12.

Figura 3.12: Histogramas LRT, Parte 3 da simulação, Cenários 13 a 24.

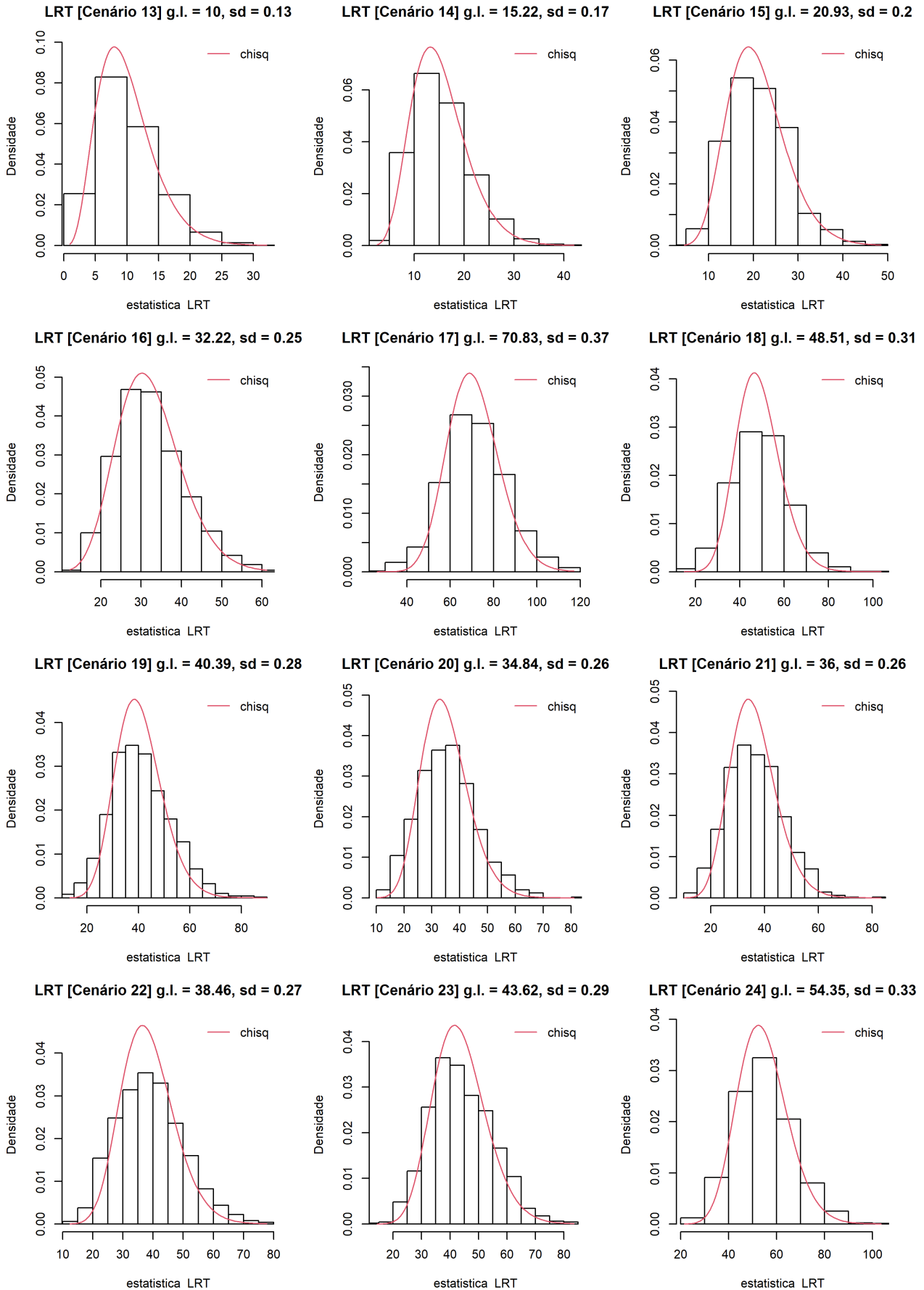


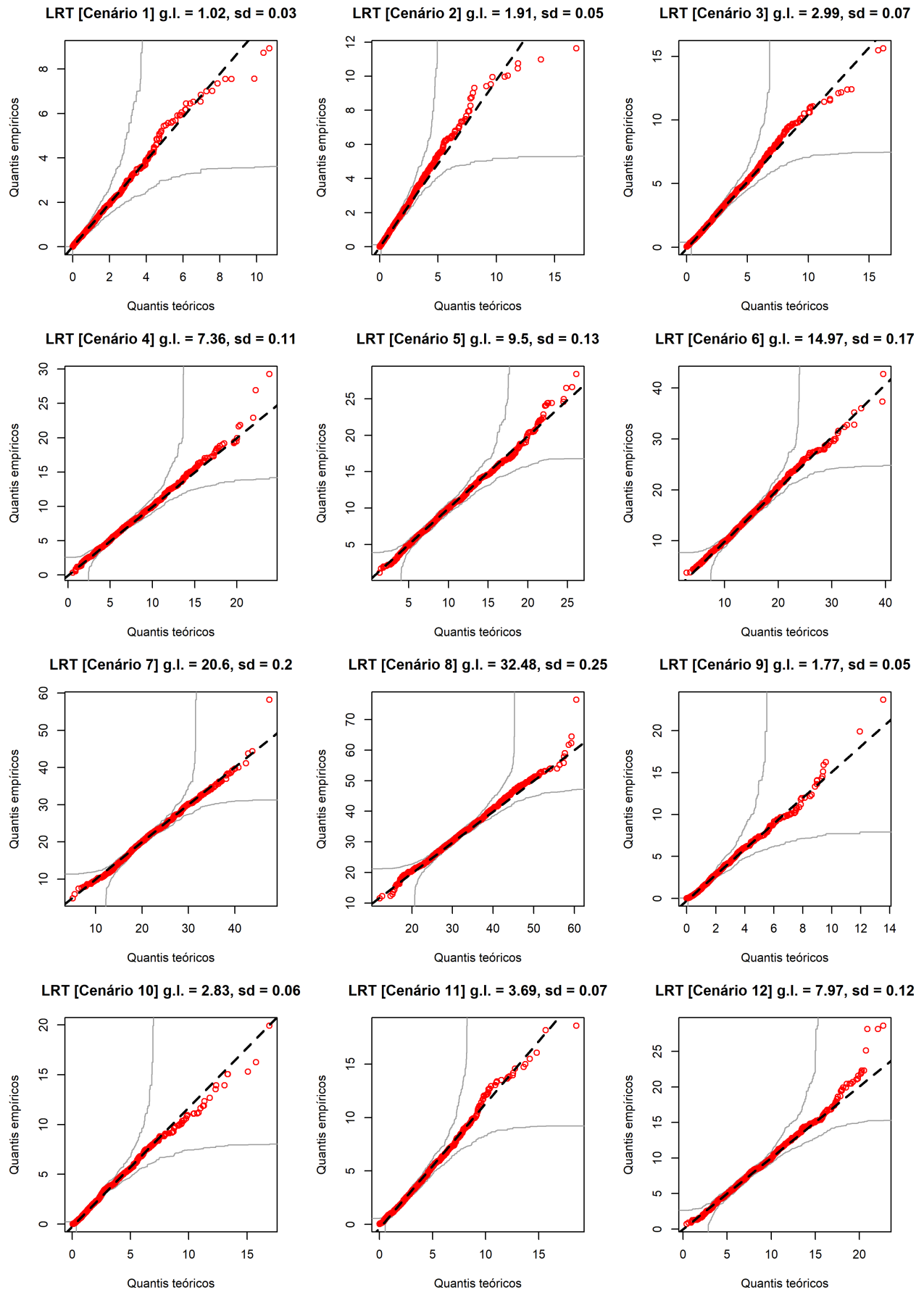
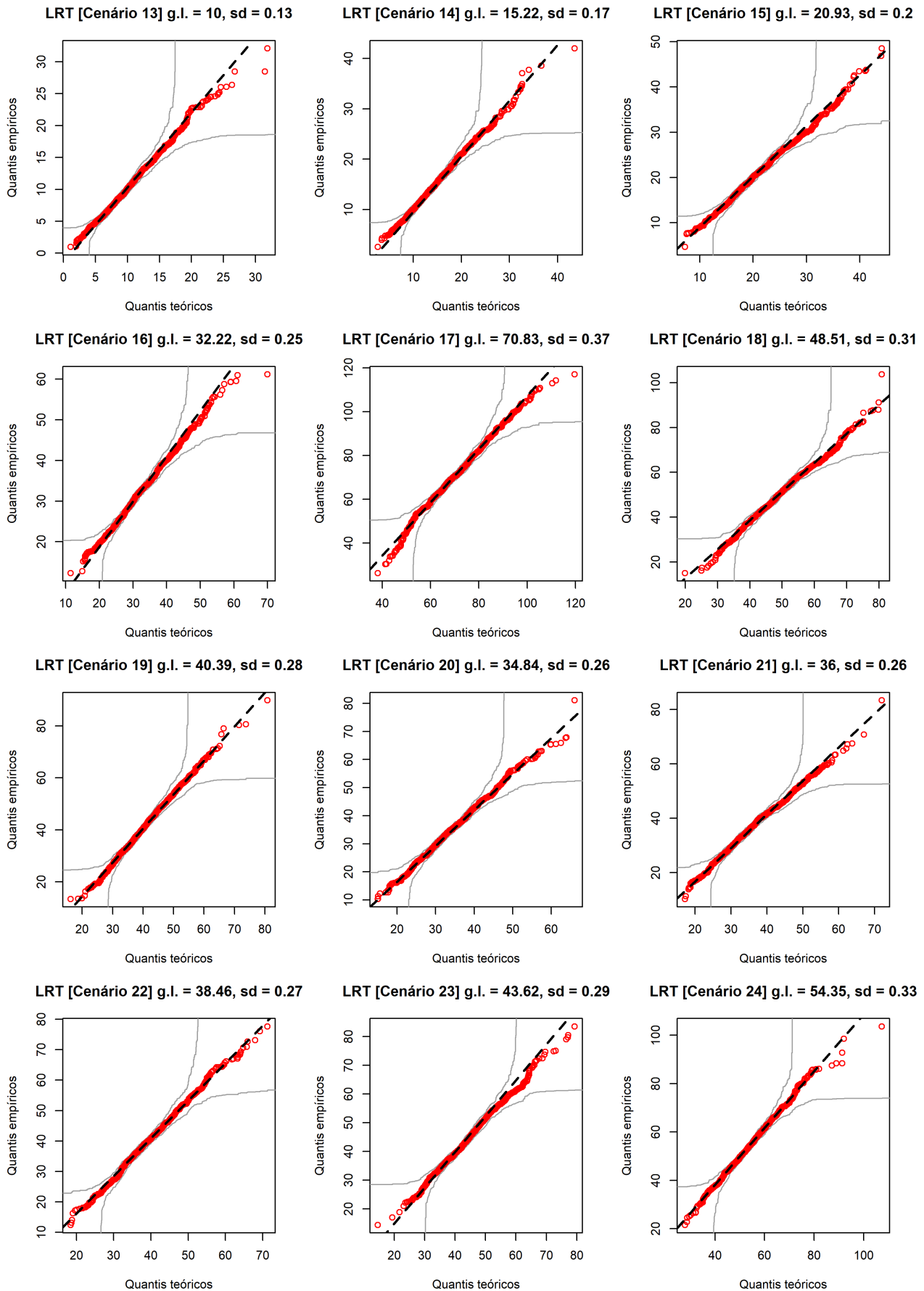
Figura 3.13: *QQ-plots LRT, Parte 3 da simulação, Cenários 1 a 12.*

Figura 3.14: QQ-plots LRT, Parte 3 da simulação, Cenários 13 a 24.



Devido à análise desenvolvida nesta seção e das análises apresentadas nas Seções 3.3 e 3.4, recomendamos em situações práticas que o teste LRT usando o valor crítico *bootstrap* seja aplicado, ao invés dos outros dois testes.

Capítulo 4

Medidas de Diagnóstico com Base em Verossimilhanças Preditivas

Este capítulo é dedicado ao desenvolvimento de Funções de Verossimilhança Preditivas, Densidades Preditivas e Funções de Influência Preditivas no contexto da regressão L_1 . O principal objetivo é obter novas medidas de diagnóstico.

Abordaremos os conceitos de verossimilhança preditiva (VP) e densidade preditiva (DP) e aplicaremos DP como ferramenta para detecção de observações influentes no âmbito da regressão L_1 . Um fato histórico relevante é que o termo “verossimilhança preditiva” foi cunhado por [Hinkley \(1979\)](#), que também iniciou um estudo sistemático das verossimilhanças preditivas. Esses conceitos estão associados a algo mais amplo, que se chama análise de predição estatística, para mais detalhes sobre análise de predição consulte [Aitchison e Dunsmore \(1975\)](#), pois neste capítulo não daremos foco na previsão propriamente dita, mas sim no desenvolvimento de ferramentas para detecção de observações influentes.

Na análise de predição estatística, como em todas as áreas da Estatística, os métodos podem ser divididos, por razões pedagógicas, em frequentistas e bayesianos. A abordagem bayesiana é bem direta e decorre da utilização da distribuição a posteriori, mas requer a aplicação de uma priori. Aqui focaremos e nos debruçaremos sobre os métodos frequentistas, que são baseados somente na função de verossimilhança, o que nos isenta da necessidade de definir uma distribuição a priori.

Na Seção 4.1 faremos uma breve revisão da literatura acerca de VP e DP com ênfase em resultados relevantes para aplicação desse ferramental à regressão L_1 . Na Seção 4.2 obteremos a expressão numérica da densidade preditiva perfilada (DPP) para a regressão L_1 . Na Seção 4.3 obteremos uma função de influência preditiva (FIP) por meio da aplicação da divergência de Kullback-Leibler às DPPs de forma numérica, bem como um algoritmo para sua computação. Na Seção 4.4 comentaremos sobre os impactos da não unicidade e degeneração do estimador L_1 no cálculo da FIP. Tratamos de assuntos auxiliares ao cálculo numérico da FIP, tais como noções de processamento paralelo e

ferramentas computacionais no Apêndice E. Na Seção 4.5 apresentamos medidas de influência para futura comparação com os resultados fornecidos pela aplicação da FIP na análise de dados reais. Próximo ao fim deste capítulo há as Seções 4.6 e 4.7 que apresentam aplicação da FIP a dados reais nas áreas de poluição e imóveis, respectivamente.

4.1 Revisão bibliográfica sobre VP e DP

A literatura é majoritariamente voltada para VPs supondo que os dados possuem distribuição pertencente à família exponencial, com ênfase na distribuição normal. Harris (1989) apresenta resultados para distribuições da família exponencial. Isso decorre do fato de ser muito mais fácil tratar analiticamente, incluindo obtenção de fórmulas fechadas para VPs, quando fazemos uso de suposições como normalidade dos dados ou de resultados restritos à família exponencial. Lauritzen (1974) foi um dos pioneiros no estudo de funções de verossimilhança preditivas não bayesianas, mas seu resultado se restringe aos casos em que existe estatística suficiente minimal. É evidente que tais resultados teóricos não são imediatamente aplicáveis ao contexto em que estamos trabalhando, lembre que a distribuição Laplace não pertence à família exponencial. Por isso nosso estudo é pioneiro, trabalhamos com VP fora do contexto da família exponencial. Isso exige um apoio computacional maior para cálculo numérico das DPs e também mais conhecimentos pragmáticos acerca de processamento de dados.

Bjørnstad (1990) apresenta 14 possibilidades de VPs, embora boa parte delas apresentem muitas similaridades. Para obter uma VP é necessário eliminar o parâmetro θ da verossimilhança, para isso há três possibilidades: condicionar em relação a uma estatística suficiente, integrar em relação ao parâmetro ou maximizar em relação ao parâmetro. Com os devidos ajustes, o mesmo se aplica no caso do parâmetro não ser unidimensional, isto é, no caso em que há um vetor de parâmetros ao invés de um único parâmetro escalar. De certa forma, o método de previsão bayesiano está contido no caminho em que integramos com relação ao parâmetro.

Para um primeiro contato com VP recomenda-se a leitura de Bjørnstad (1990), pois além desse artigo apresentar uma revisão bibliográfica completa – até o ano de 1990 – de VP e assuntos associados, há uma preocupação em sistematizar o estudo das diversas VPs propostas na literatura. Inclusive, a referência supracitada continua sendo relevante em publicações recentes e influentes como Lee e Nelder (2009). Para aqueles que desejam uma revisão bibliográfica com uma riqueza maior de detalhes, sobretudo no desenvolvimento das equações, em língua portuguesa, recomenda-se fortemente a leitura de Rojas (2000), pois esta referência apresenta uma revisão de literatura sobre Densidades Preditivas que cobre publicações até 1998, em Português.

Ao tratar com a questão da VP, é basilar lidar com duas quantidades desconheci-

das, θ e z , em que θ é o parâmetro e z é a observação futura a ser predita. É nítido que utilizaremos nosso conhecimento (estimativa) acerca de θ como meio para obter informações sobre z , assim θ tem caráter essencialmente secundário – essa noção de parâmetro de natureza auxiliar é expressa em Inglês como *nuisance parameter*. Assim, como desconhecemos θ , parece razoável substituir θ por sua estimativa pontual, isto é, usar $\hat{\theta}$ ao invés de θ em $f_{\theta}(z|y)$ – essa VP é conhecida na literatura como verossimilhança pseudo preditiva ou clássica. Tal ideia é equivocada, pois não fornece resultados adequados. Conforme salientado por Butler (1986), essa solução é inadequada, pois assume implicitamente que $\theta = \hat{\theta}$ e portanto não leva em conta a incerteza adicional devido à estimação de θ . Bjørnstad (1990) ilustra via simulação que o resultado de Butler (1986) é verdadeiro. Levando em conta que a solução trivial é inadequada, é razoável buscarmos alguma alternativa melhor de VP.

Mathiasen (1979), Lejeune e Faulkenberry (1982) e Levy e Perng (1984) estudaram em profundidade a Verossimilhança Preditiva Perfilada (VPP), que denotaremos de forma simplificada por L_p . L_p é bem definida tanto no caso em que Y e Z são independentes quanto no caso em que não são (aqui ainda não definimos Y , observação passada, e Z , observação futura, mas isso será feito na Seção 4.2, já no âmbito da regressão L_1). Leonard (1982) notou que, no âmbito da família exponencial, L_p fornece valores similares aos da densidade preditiva posteriori bayesiana com priori uniforme quando a amostra é pequena e que elas são assintoticamente equivalentes, considerando o mesmo contexto. Assim, L_p é uma excelente alternativa clássica no campo das funções preditivas. Conforme Bjørnstad (1990), há uma ideia clara, intuitiva e apelativa por trás da L_p . Considerando z como o “parâmetro de interesse” e θ como o parâmetro incômodo (*nuisance parameter*), o valor mais provável de θ , dado (y, z) – ou seja o “estimador de máxima verossimilhança” de θ com base em (y, z) – é determinado e $L_p(z|y)$ é a verossimilhança resultante. Isso corresponde à verossimilhança perfilada na inferência paramétrica e L_p é denominada verossimilhança preditiva perfilada. A vantagem dessa construção é que ela é pragmática e amplamente aplicável a diversos casos, inclusive casos externos à família exponencial – que é o nosso cenário com distribuição Laplace.

Observação 4.1 Um parâmetro incômodo é qualquer parâmetro desconhecido e de natureza secundária, mas que deve ser levado em conta em análises estatísticas. Se o foco da análise não está relacionado a determinado parâmetro então dizemos que tal parâmetro é de natureza secundária. Por essência, se um parâmetro incômodo passa a ser conhecido ou torna-se o foco da análise então o mesmo deixa de ser parâmetro incômodo.

A seguir, definimos a L_p de forma geral.

Definição 4.1 (Verossimilhança Preditiva Perfilada - L_p) A função de verossimilhança preditiva perfilada (VPP) para \mathbf{z} , em que $\mathbf{z} = g(\mathbf{v})$ com \mathbf{v} sendo o vetor de novas obser-

vações e \mathbf{x} sendo o vetor de observações passadas, é dada por

$$L_p(\mathbf{z}|\mathbf{x}) = \sup_{\theta \in \Theta} l_{\mathbf{x}}(\mathbf{z}, \theta) = \sup_{\theta \in \Theta} f(\mathbf{x}, \mathbf{z}; \theta),$$

note que a VPP é a função de verossimilhança conjunta das amostras de X e V , maximizada com relação aos parâmetros e vista como função de Z (lembre que Z é função de V), pois os valores de X são as observações já conhecidas.

Na Definição 4.1, é nítido que θ desempenha o papel de parâmetro incômodo. Um exemplo típico em que há parâmetro incômodo é no teste *t-Student* para média, em que os dados são extraídos de uma população normal e o parâmetro σ^2 é desconhecido. Nesse exemplo, o valor de σ^2 é desconhecido e o foco da análise está na média, portanto σ^2 é um parâmetro incômodo neste contexto.

A seguir apresentamos a definição de densidade preditiva perfilada (DPP).

Definição 4.2 (Densidade Preditiva Perfilada) A Densidade Preditiva Perfilada (DPP) é definida como:

$$f_p(\mathbf{z}|\mathbf{x}) = k(\mathbf{x})L_p(\mathbf{z}|\mathbf{x}),$$

em que $k(\mathbf{x})$ é uma constante de padronização (que leva \mathbf{x} em consideração) e tal que

$$\int_{-\infty}^{\infty} k(\mathbf{x})L_p(\mathbf{z}|\mathbf{x})d\mathbf{z} = 1.$$

A DPP é apenas uma VPP padronizada, isto é, uma função oriunda da VPP que satisfaz os requisitos necessários para ser uma função densidade de probabilidade. Inclusive, padronizar garante uma melhor comparabilidade entre VPs obtidas por métodos distintos. Por exemplo, se duas VPs são distintas mas se após a padronização ambas as VPs gerarem a mesma DP, então significa que ambas as VPs fornecem a mesma previsão. Além disso, alguns autores, tais como [Lejeune e Faulkenberry \(1982\)](#) e [Levy e Perng \(1986\)](#), interpretam DPP como uma estimativa de $f_{\theta}(z|y)$, distribuição de z condicionada a observação de y e ao conhecimento do parâmetro θ .

Com a Definição 4.2 concluída, aplicaremos o conceito de DPP no contexto da regressão L_1 a seguir.

4.2 DPP no contexto da Regressão L_1

Considere o modelo observável

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{u}$$

e o não observável

$$\mathbf{Z} = \mathbf{W}\beta + \mathbf{u}^*,$$

desta forma, \mathbf{X} e \mathbf{W} são matrizes conhecidas, \mathbf{Y} é um vetor n -dimensional de variáveis aleatórias (observadas) com distribuição Laplace, $\mathbf{Y} \sim \text{Laplace}(\mathbf{X}\beta, \lambda \mathbf{I}_n)$, \mathbf{Z} é um vetor m -dimensional de variáveis aleatórias (não observadas) com distribuição Laplace, $\mathbf{Z} \sim \text{Laplace}(\mathbf{W}\beta, \lambda \mathbf{I}_m)$, \mathbf{u} e \mathbf{u}^* são vetores de erros e β e λ são os parâmetros de posição e escala, respectivamente. Também considere que os vetores aleatórios \mathbf{u} e \mathbf{u}^* são independentes, o que necessariamente implica a independência entre \mathbf{Y} e \mathbf{Z} .

A função de verossimilhança de \mathbf{Y} é

$$L(\theta; \mathbf{Y}) = (2\lambda)^{-n} \exp\left(-\frac{1}{\lambda} \sum_{i=1}^n |Y_i - \mathbf{X}_i\beta|\right), \quad \theta = (\lambda, \beta),$$

e de forma análoga, a função de verossimilhança de \mathbf{Z} é

$$L(\theta; \mathbf{Z}) = (2\lambda)^{-m} \exp\left(-\frac{1}{\lambda} \sum_{i=1}^m |Z_i - \mathbf{W}_i\beta|\right),$$

em que $\theta = (\beta^\top, \lambda)^\top$ denota o vetor de parâmetros do modelo e \mathbf{W}_i é a i -ésima linha da matriz \mathbf{W} , $i = 1, 2, \dots, m$.

Como consequência, a função de verossimilhança conjunta \mathbf{Y} e \mathbf{Z} é

$$L(\theta; \mathbf{Y}, \mathbf{Z}) \stackrel{\text{ind}}{=} L(\theta; \mathbf{Y}) \cdot L(\theta; \mathbf{Z}) = (2\lambda)^{-n-m} \exp\left[-\frac{1}{\lambda} \left(\sum_{i=1}^n |Y_i - \mathbf{X}_i\beta| + \sum_{i=1}^m |Z_i - \mathbf{W}_i\beta|\right)\right].$$

Para obter a VPP, basta substituímos θ pelo seu respectivo estimador de máxima verossimilhança (EMV). No caso temos

$$\hat{\beta} = L_1(\mathbf{Y}, \mathbf{Z}, \mathbf{X}, \mathbf{W}) \quad (\text{essa notação peculiar serve para estabelecer } \hat{\beta} \text{ como função de } \mathbf{Y}, \mathbf{Z}, \mathbf{X} \text{ e } \mathbf{W})$$

e

$$\hat{\lambda} = \frac{1}{n+m} \left(\sum_{i=1}^n |Y_i - \mathbf{X}_i\hat{\beta}| + \sum_{i=1}^m |Z_i - \mathbf{W}_i\hat{\beta}| \right).$$

Assim a VPP é dada por

$$\begin{aligned} \text{FVPP}(\mathbf{z}|\mathbf{y}) &= L(\hat{\theta}; \mathbf{Y}, \mathbf{Z}) = (2\hat{\lambda})^{-n-m} \exp\left[-\frac{1}{\hat{\lambda}} \left(\sum_{i=1}^n |Y_i - \mathbf{X}_i\hat{\beta}| + \sum_{i=1}^m |Z_i - \mathbf{W}_i\hat{\beta}|\right)\right] \\ &= 2^{-n-m} \hat{\lambda}^{-n-m} \exp(-n-m) \\ &= 2^{-n-m} \left(\frac{1}{n+m} \left(\sum_{i=1}^n |y_i - \mathbf{X}_i\hat{\beta}| + \sum_{i=1}^m |z_i - \mathbf{W}_i\hat{\beta}|\right) \right)^{-n-m} \exp(-n-m) \\ &= C(n, m) \left(\sum_{i=1}^n |y_i - \mathbf{X}_i\hat{\beta}| + \sum_{i=1}^m |z_i - \mathbf{W}_i\hat{\beta}|\right)^{-n-m}, \end{aligned}$$

em que $C(n, m) = \left(\frac{2}{n+m}\right)^{-n-m} \exp(-n - m)$.

Para simplificar os cálculos da FVPP, considere que $m = 1$, isto é, o vetor aleatório \mathbf{Z} se reduz uma variável aleatória escalar Z . Assim temos

$$\text{FVPP}(z|\mathbf{y}) = C(n, 1) \left(\sum_{i=1}^n |y_i - \mathbf{X}_i \hat{\beta}| + |z - \mathbf{W} \hat{\beta}| \right)^{-n-1},$$

para transformar $\text{FVPP}(z|\mathbf{y})$ em uma DPP basta encontrar a função constante (em relação a z) de padronização. Basta resolver a integral a seguir:

$$C(n, 1) \int_{\mathbb{R}} \left(\sum_{i=1}^n |y_i - \mathbf{X}_i \hat{\beta}| + |z - \mathbf{W} \hat{\beta}| \right)^{-n-1} dz. \quad (4.1)$$

O grande problema da integral da Expressão (4.1) é que a mesma não pode ser resolvida de forma analítica. Isto decorre do fato de que $\hat{\beta}$ é função de Z , mas não temos como expressar $\hat{\beta}$ explicita e algebricamente em termos de Z , essa dificuldade nos impede de obter uma forma analítica para a constante de padronização.

Por outro lado, podemos resolver a integral da Equação (4.1) numericamente, basta levarmos em conta que desconhecemos a forma funcional do integrando, mas podemos calcular o valor do integrando ponto a ponto. O caminho da integração numérica é o que utilizaremos para obtermos a DPP.

Em posse da DPP, nosso objetivo é obter uma medida que avalie a influência de cada observação no ajuste geral do modelo e na previsão da nova observação, trataremos de uma medida dessa natureza na seção a seguir.

4.3 Função de Influência Preditiva a partir da DKL entre as DPPs

Podemos obter uma função de influência preditiva (FIP) a partir da aplicação da divergência de Kullback-Leibler às DPPs com todas as observações e sem a j -ésima observação. De fato, [Johnson e Geisser \(1982, 1983\)](#) são os pioneiros dessa metodologia e a aplicaram com sucesso na regressão linear normal – obtendo numerosos resultados analíticos.

Usaremos a notação $\text{DPP}(z)$ para a DPP com todas as observações e $\text{DPP}_{(j)}(z)$ para a DPP sem a j -ésima observação, analogamente, chamamos de $\text{FVPP}(z)$ a FVPP com todas as observações e $\text{FVPP}_{(j)}(z)$ a FVPP sem a j -ésima observação, $j = 1, \dots, n$.

Lembrando da forma da FVPP(z), podemos expressar $FVPP_{(j)}(z)$ como

$$FVPP_{(j)}(z) = FVPP_{(j)}(z|\mathbf{y}) = C(n-1, 1) \left(\sum_{i=1, i \neq j}^n |y_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_{(j)}| + |z - \mathbf{W} \hat{\boldsymbol{\beta}}_{(j)}| \right)^{-n+1-1},$$

em que $C(n, m) = \left(\frac{2}{n+m}\right)^{-n-m} \exp(-n-m)$, e $\hat{\boldsymbol{\beta}}_{(j)} = L_1(\mathbf{Y}_{(j)}, \mathbf{Z}, \mathbf{X}_{(j)}, \mathbf{W})$; sendo que $\mathbf{Y}_{(j)}$ e $\mathbf{X}_{(j)}$, indicam, respectivamente, a exclusão do j -ésimo componente do vetor \mathbf{Y} e a exclusão da j -ésima linha da matriz \mathbf{X} . A DPP $_{(j)}(z)$ é obtida imediatamente ao padronizarmos $FVPP_{(j)}(z)$.

Kullback e Leibler (1951) introduziram uma medida de divergência entre funções que é útil até hoje no estudo de probabilidade e estatística, tanto teóricas quanto aplicadas. A divergência de Kullback-Leibler (DKL) é uma ferramenta oriunda da Teoria da Informação desenvolvida para mensurar o grau de discrepância entre duas funções definidas no mesmo domínio. No contexto da probabilidade, essas funções podem ser ambas funções densidade de probabilidade ou ambas funções de probabilidade. Ainda no escopo da Teoria da informação, a DKL pode ser vista como uma medida de entropia relativa. Assim, com base em nossa perspectiva, a DKL constitui-se em valiosa ferramenta matemática para desenvolvimento de técnicas estatísticas voltadas para diagnóstico em modelos de regressão.

O valor da DKL é sempre não negativo. Com respeito à interpretação da DKL, uma DKL igual a 0 indica que as funções (ou distribuições) $f(\cdot)$ e $g(\cdot)$ são iguais ou muito similares, enquanto uma DKL de 1 indica que se comportam de maneiras distintas. Quanto maior o valor da DKL, maior a discrepância entre as duas funções utilizadas em seu cálculo.

Em outras palavras, ao mensurarmos a discrepância entre DPP(z) e DPP $_{(j)}(z)$ temos uma medida geral (independe de valores específicos de z) que indica o grau de influência da j -ésima observação no ajuste geral do modelo e na previsão da observação futura. A seguir definimos a DKL.

Definição 4.3 (DKL) A divergência de Kullback-Leibler entre duas funções, $g(\cdot)$ e $h(\cdot)$ é definida como

$$D_{KL}(g(\cdot), h(\cdot)) = \int_{-\infty}^{+\infty} g(x) \ln \left(\frac{g(x)}{h(x)} \right) dx.$$

Se $g(\cdot)$ for uma função densidade de probabilidade, DKL pode ser interpretada como a esperança de $\ln \left(\frac{g(\cdot)}{h(\cdot)} \right)$ com relação à variável aleatória cuja função densidade de probabilidade é $g(\cdot)$.

Como não conseguimos obter analiticamente DPP(z) e tampouco DPP $_{(j)}(z)$, é evidente que não há como calcular a DKL entre DPP(z) e DPP $_{(j)}(z)$ de forma analítica. Mas podemos, mais uma vez, calcular a integral que define a DKL entre as densidades preditivas de forma numérica, pois conhecemos o integrando ponto a ponto. Seguiremos com

o raciocínio de forma numérica.

Aplicaremos a medida DKL à $DPP(z)$ e à $DPP_{(j)}(z)$, isto é, na Definição 4.3 tomaremos $g(\cdot) = DPP(z)$ e $h(\cdot) = DPP_{(j)}(z)$. Lembre que $DPP(z) = k \cdot FVPP(z)$ e $DPP_{(j)}(z) = k_j \cdot FVPP_{(j)}(z)$, em que as constantes de integração k e k_j são:

$$k = \left(C(n, 1) \int_{-\infty}^{+\infty} \left(\sum_{i=1}^n |y_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}| + |z - \mathbf{W} \hat{\boldsymbol{\beta}}| \right)^{-n-1} dz \right)^{-1}$$

e

$$k_j = \left(C(n-1, 1) \int_{-\infty}^{+\infty} \left(\sum_{i=1, i \neq j}^n |y_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_{(j)}| + |z - \mathbf{W} \hat{\boldsymbol{\beta}}_{(j)}| \right)^{-n} dz \right)^{-1}, \quad j = 1, \dots, n.$$

Nestas condições, definimos a função de influência preditiva (FIP) associada à j -ésima observação, $FIP(j)$, como

$$\begin{aligned} FIP(j) &= D_{KL}(j) = D_{KL}(DPP(z), DPP_{(j)}(z)) \\ &= \int_{-\infty}^{+\infty} DPP(z) \ln \left(\frac{DPP(z)}{DPP_{(j)}(z)} \right) dz \\ &= \int_{-\infty}^{+\infty} k \cdot FVPP(z) \ln \left(\frac{k \cdot FVPP(z)}{k_j \cdot FVPP_{(j)}(z)} \right) dz \\ &= \int_{-\infty}^{+\infty} k \cdot FVPP(z) \left[\ln \left(\frac{k}{k_j} \right) + \ln \left(\frac{FVPP(z)}{FVPP_{(j)}(z)} \right) \right] dz \\ &= \int_{-\infty}^{+\infty} k \cdot FVPP(z) \cdot \ln \left(\frac{k}{k_j} \right) dz + \int_{-\infty}^{+\infty} k \cdot FVPP(z) \cdot \ln \left(\frac{FVPP(z)}{FVPP_{(j)}(z)} \right) dz \\ &= \ln \left(\frac{k}{k_j} \right) \underbrace{\int_{-\infty}^{+\infty} k \cdot FVPP(z) dz}_{=1} + k \int_{-\infty}^{+\infty} FVPP(z) \cdot \ln \left(\frac{FVPP(z)}{FVPP_{(j)}(z)} \right) dz \\ &= \ln \left(\frac{k}{k_j} \right) + k \int_{-\infty}^{+\infty} FVPP(z) \cdot \ln \left(\frac{FVPP(z)}{FVPP_{(j)}(z)} \right) dz. \end{aligned}$$

Só podemos resolver a integral de forma numérica, pois não conhecemos analiticamente $FVPP(z)$ e $FVPP_{(j)}(z)$, mas conhecemos essas funções de forma numérica. Mas ao trabalharmos com diversas operações de natureza numérica em sequência, é importante termos atenção à forma como os erros numéricos (de arredondamento e de truncamento) propagar-se-ão.

Os erros de arredondamento decorrem do fato dos computadores utilizarem um número finito (limitado) de dígitos para representarem números reais. Além disso, quando substitui-se um processo matemático infinito por um processo finito, há erro de truncamento. Por exemplo, quando calculamos uma integral numérica, estamos cometendo

erro de truncamento, pois computamos um somatório finito ao invés de um somatório infinito. O principal objetivo quando lidamos com sucessivas operações numéricas é minimizar o acúmulo de erros numéricos. Mais detalhes acerca da análise numérica e sobre como os erros numéricos acumulam-se podem ser vistos em [Higham \(2002\)](#).

Lembre que não temos expressões fechadas para FVPP(z) e FVPP $_{(j)}$ (z), por isso precisaremos ajustar um modelo de regressão para cada valor de z , sendo um modelo para os dados com todas observações e outro sem a j -ésima observação. É evidente que não teremos como ajustar modelos para todos os valores em \mathbb{R} , assim ajustaremos modelos para valores de z em um intervalo razoável, como $(-1000, 1000)$. Assim poderemos, por exemplo, considerar como sequência de valores para z de -1000 a 1000 com espaçamento de 20 unidades. É evidente que o intervalo “razoável” deve necessariamente englobar a região com maiores valores na função densidade de probabilidade.

Observação 4.2 É evidente que o intervalo razoável dependerá dos valores mais prováveis que z pode assumir. Se é mais provável que z assuma valores de forma simétrica em torno de zero, então o intervalo $(-1000, 1000)$ é razoável. Mas se, por exemplo, em determinado contexto, z habitualmente assume valores em torno de 50, então o intervalo razoável deve ser centrado em 50, ou seja, um intervalo razoável seria $(-950, 1050)$.

A seguir, construímos o algoritmo simplificado para o cálculo da DKL numérica.

Descrição do Algoritmo da DKL numérica

1. Calcule FVPP(z) e FVPP $_{(j)}$ (z), $j = 1, \dots, N$ numericamente.
2. Obtenha, numericamente, as constantes de padronização k e k_j , $j = 1, \dots, N$; para isto basta lembrar que $k = \left(\int_{-\infty}^{+\infty} \text{FVPP}(z) dz \right)^{-1}$ e $k_j = \left(\int_{-\infty}^{+\infty} \text{FVPP}_{(j)}(z) dz \right)^{-1}$, $j = 1, \dots, N$.
3. Calcule a $D_{KL}(j)$, $j = 1, \dots, N$, usando FVPP(z) e FVPP $_{(j)}$ (z), $j = 1, \dots, N$ e suas respectivas constantes de padronização.
4. (opcional) Faça um gráfico de dispersão dos valores de $D_{KL}(j)$, $j = 1, \dots, N$ em função do índice da observação.

Neste ponto recomendamos ao leitor ler o Apêndice E. No apêndice há uma breve explicação sobre processamento paralelo e sobre os pacotes computacionais utilizados para implementar processamento paralelo no cálculo da FIP.

4.4 Não unicidade e degeneração do estimador L_1 e suas consequências na FVPP(z) numérica

Não é possível expressar FVPP(z) de forma algébrica. Isso decorre do fato de termos que calcular a estimativa $\hat{\beta} = L_1(\mathbf{Y}, \mathbf{Z}, \mathbf{X}, \mathbf{W})$ para cada valor de z , este é um cenário

diferente da regressão linear normal, em que há fórmulas fechadas para o estimador $\hat{\beta}$.

O cálculo da estimativa $\hat{\beta}$ pode depender diretamente de z (se a observação z pertencer ao conjunto das observações definidoras, isto é, se o resíduo correspondente a z for nulo) ou não depender de z (se a observação z não pertencer ao conjunto das observações definidoras).

Por não haver uma relação analítica entre $\hat{\beta}$ e z , não será possível fazer transformações de variáveis nas integrais, o que inviabiliza um tratamento analítico da integral. O mesmo raciocínio é aplicável para $FVPP_{(j)}(z)$ e $\hat{\beta}_{(j)}$, que são calculadas após a exclusão da j -ésima observação. Da incapacidade de expressar as FVPPs, surge a impossibilidade de calcular, analiticamente, a integral correspondente à DKL entre as DPPs. Mas podemos calcular a DKL entre as DPPs de forma numérica.

Note que tanto a $FVPP(z)$ quanto a $FVPP_j(z)$ são únicas mesmo que o estimador L_1 de β não seja único. Isso decorre do fato de ambas as medidas serem baseadas em uma quantidade inversamente proporcional a soma de resíduos absolutos (SRA), a $FVPP(z)$ considera os resíduos de todas as observações e a $FVPP_j(z)$ também, com exceção da j -ésima observação, que é excluída. Então, mesmo que haja infinitas soluções para o problema da estimação L_1 , as múltiplas soluções fornecerão exatamente as mesmas $FVPP(z)$ e $FVPP_j(z)$, $j = 1, \dots, N$, porque a SRA será sempre a mesma, independente da solução escolhida.

Se algo não ficou claro acerca da não unicidade do estimador L_1 é recomendado ao leitor rever a Subseção 1.3.1.

4.5 Outras medidas de influência para comparação

Revisaremos as fórmulas das medidas de influência apresentadas em [Rodrigues \(2019\)](#) e suas respectivas interpretações de forma sucinta.

1. Afastamento de verossimilhanças:

$$LD_i(\theta) = 2 \left(n \ln \left(\hat{\lambda}_{(i)} / \hat{\lambda} \right) + |Y_i - \mathbf{X}_i \hat{\beta}_{(i)}| / \hat{\lambda}_{(i)} - 1 \right),$$

em que $\theta = (\beta^\top, \lambda)^\top$, λ é o parâmetro de escala da distribuição Laplace, $\hat{\beta}$ e $\hat{\lambda}$ são os estimadores de máxima verossimilhança de β e λ , $\hat{\beta}_{(i)}$ e $\hat{\lambda}_{(i)}$ denotam os estimadores de máxima verossimilhança de β e λ , excluindo-se a i -ésima observação da amostra. Se a i -ésima observação não é influente, a estimativa $\hat{\lambda}_{(i)}$ calculada sem a i -ésima observação será próxima de $\hat{\lambda}$, de modo que $\hat{\lambda}_{(i)} / \hat{\lambda} \approx 1$. Além disso, $|Y_i - \mathbf{X}_i \hat{\beta}_{(i)}|$ deve ser próximo da média dos erros absolutos, $\hat{\lambda}_{(i)}$, ou seja, $|Y_i - \mathbf{X}_i \hat{\beta}_{(i)}| \approx \hat{\lambda}_{(i)}$ e como consequência, LD_i deve ser próxima de zero.

2. **Afastamento condicional de verossimilhanças:**

$$LD_i(\beta|\lambda) = 2n \ln \left(\frac{\sum_{j=1}^n |Y_j - \mathbf{X}_j \hat{\beta}_{(i)}|}{\sum_{j=1}^n |Y_j - \mathbf{X}_j \hat{\beta}|} \right).$$

3. **Distância de Cook:**

$$D_i = \hat{\lambda}^{-2} (\hat{\beta} - \hat{\beta}_{(i)})^\top \mathbf{X}^\top \mathbf{X} (\hat{\beta} - \hat{\beta}_{(i)}).$$

4. **Medida de influência baseada na norma L_2 :**

$$FD_i = \hat{\lambda}^{-2} \|\mathbf{X}\hat{\beta} - \mathbf{X}\hat{\beta}_{(i)}\|_2.$$

5. **Medida de influência baseada na norma L_1 :**

$$FD_{1i} = \hat{\lambda}^{-1} \|\mathbf{X}\hat{\beta} - \mathbf{X}\hat{\beta}_{(i)}\|_1.$$

Todas as medidas de influência listadas já estão adequadas ao contexto da regressão L_1 . É útil lembrar que dentre as cinco medidas listadas, a única com propriedade de discernir observações conjuntamente influentes na estimação dos parâmetros de posição e de escala é a primeira medida, o afastamento de verossimilhanças (AV). As demais quatro medidas só detectam observações influentes na estimação dos parâmetros de posição. Assim, é esperado que os resultados das quatro últimas medidas sejam concordantes entre si e distintos do resultado fornecido pela medida AV. Mais detalhes sobre as medidas 1 e 2 encontram-se em [Cook et al. \(1988\)](#) e [Elian et al. \(2000\)](#), já, para a medida 3, basta consultar [Cook \(1977\)](#), e para as medidas 4 e 5 consultar [Sun e Wei \(2004\)](#).

4.6 Dados sobre bile litogênica

A Tabela 4.1 apresenta dados de concentrações de bile litogênica nas índias Pima, que foram extraídos de [Goodall \(1983\)](#). Os dados fornecem a idade e a saturação do colesterol biliar de 29 mulheres. As duas variáveis são Idade, que é mensurada em anos, e Concentração, que corresponde a saturação de colesterol biliar nas índias Pima.

A saturação de colesterol na bile é a porcentagem de colesterol em relação à quantidade máxima que a bile pode dissolver a uma determinada temperatura e pH. Quando a saturação de colesterol na bile atinge 100%, significa que a bile está saturada, mas ela pode continuar a dissolver colesterol até que atinja um ponto de supersaturação, ponto em que a bile não pode mais manter o colesterol em solução. Em outras palavras, valores

de saturação de colesterol na bile superiores a 100% indicam que a bile está supersaturada com colesterol, o que pode aumentar o risco de formação de cálculos biliares de colesterol.

A correlação entre as variáveis Concentração e Idade é 0,457, nosso objetivo é modelar a variável Concentração por meio da regressão L_1 usando Idade como variável explicativa. Posteriormente aplicaremos a técnica de diagnóstico desenvolvida neste capítulo.

Observações	Idade	Concentração
1	10,00	98
2	10,30	120
3	10,60	111
4	10,60	93
5	11,00	159
6	11,40	120
7	11,00	101
8	12,40	130
9	12,80	65
10	13,20	165
11	13,20	106
12	13,20	101
13	13,60	152
14	13,90	169
15	14,20	120
16	14,60	122
17	15,40	122
18	16,80	122
19	17,00	177
20	17,70	93
21	18,20	111
22	18,70	148
23	19,80	143
24	19,90	122
25	20,10	101
26	20,20	150
27	20,70	165
28	20,70	250
29	21,00	187

Fonte: [Goodall \(1983\)](#).

Tabela 4.1: *Dados sobre bile litogênica.*

A Tabela 4.2 contém as estatísticas descritivas mínimo, primeiro quartil (Q_1), mediana, média, terceiro quartil (Q_3) e máximo das variáveis Idade e Concentração.

Estatísticas	Idade	Concentração
Mínimo	10,00	65,0
Q_1	12,40	106,0
Mediana	14,20	122,0
Média	15,25	131,8
Q_3	18,70	152,0
Máximo	21,00	250,0

Tabela 4.2: Estatísticas descritivas dos dados sobre bile litogênica.

Iniciaremos a etapa de análise de regressão propriamente dita, utilizaremos um modelo linear com intercepto. A forma funcional do modelo que ajustaremos ao conjunto de dados sobre bile litogênica é dada pela Equação (4.2).

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad i = 1, \dots, 29. \quad (4.2)$$

As estimativas, os erros padrão das estimativas e os valores-P (níveis descritivos) do modelo ajustado através do método de regressão L_1 se encontram na Tabela 4.3, os valores preditos e resíduos estão na Tabela 4.4. Somente a estimativa do parâmetro β_0 , que representa o intercepto do modelo, apresentou valor-P superior a 0,05, mas continuaremos a análise com o intercepto no modelo.

Parâmetros	Estimativa	Erro Padrão	valor-P
β_0	49,0000	28,9606	0,1021
β_1	5,0000	1,8461	0,0115
λ (estimado via $\hat{\lambda}_{MLE}$)	24,0000		
λ (estimado via $\hat{\lambda}_{ROS}$)	27,9293		

Tabela 4.3: Estimativas, erros padrão das estimativas e valores-P no modelo ajustado para variável resposta Concentração.

Aplicaremos os testes para verificar se a suposição dos erros possuírem distribuição Laplace é razoável por meio dos cinco testes de hipóteses apresentados no Apêndice D, ao nível de 5% de significância. Os resultados dos testes são dados na lista a seguir, lembrando que o tamanho da amostra é 29 e que os valores críticos dos testes são dados nas Tabelas D.1 a D.5.

1. teste baseado na estatística de Watson: $U^2 = 0,0328885 < 0,083 = U_*^2$
2. teste baseado na estatística de Cramér-von Mises: $W^2 = 0,08988966 < 0,144 = W_*^2$

3. teste baseado na estatística de Anderson-Darling: $A^2 = 0,6133157 < 0,987 = A^2_*$
4. teste baseado na estatística de Kolmogorov-Smirnov: $\sqrt{29}D = 0,6190759 < 0,906 = \sqrt{29}D_*$
5. teste baseado na estatística de Kuiper: $\sqrt{29}V = 0,8976189 < 1,295 = \sqrt{29}V_*$

Para os cinco testes, a hipótese nula não foi rejeitada ao nível de 5% de significância, portanto não rejeitamos a hipótese de que os erros possuem distribuição Laplace.

Observações	Y_i	\hat{Y}_i	$Y_i - \hat{Y}_i$ (resíduo)	$\hat{\epsilon}_i / \hat{\lambda}_{MLE}$ (resíduo padronizado)
1	98,00	99,00	-1,00	-0,04
2	120,00	100,50	19,50	0,81
3	111,00	102,00	9,00	0,38
4	93,00	102,00	-9,00	-0,38
5	159,00	104,00	55,00	2,29
6	120,00	106,00	14,00	0,58
7	101,00	104,00	-3,00	-0,12
8	130,00	111,00	19,00	0,79
9	65,00	113,00	-48,00	-2,00
10	165,00	115,00	50,00	2,08
11	106,00	115,00	-9,00	-0,38
12	101,00	115,00	-14,00	-0,58
13	152,00	117,00	35,00	1,46
14	169,00	118,50	50,50	2,10
15	120,00	120,00	0,00	0,00
16	122,00	122,00	0,00	0,00
17	122,00	126,00	-4,00	-0,17
18	122,00	133,00	-11,00	-0,46
19	177,00	134,00	43,00	1,79
20	93,00	137,50	-44,50	-1,85
21	111,00	140,00	-29,00	-1,21
22	148,00	142,50	5,50	0,23
23	143,00	148,00	-5,00	-0,21
24	122,00	148,50	-26,50	-1,10
25	101,00	149,50	-48,50	-2,02
26	150,00	150,00	0,00	0,00
27	165,00	152,50	12,50	0,52
28	250,00	152,50	97,50	4,06
29	187,00	154,00	33,00	1,38

Tabela 4.4: Variável (resposta) Concentração, valores preditos e resíduos obtidos via regressão L_1 .

Observação 4.3 Para os cinco testes aplicados utilizamos o valor crítico considerando $n = 35$, pois nas Tabelas D.1 a D.5 não há valor crítico para $n = 29$. Neste tipo de situação há dois caminhos para aplicar os testes, escolher o valor de n mais próximo, ou interpolar. Os valores de n mais próximos presentes nas tabelas são 20 e 35, de forma que $n = 35$ é o valor mais adequado. Optamos por escolher o valor de n mais próximo.

4.6.1 Aplicação da FIP aos dados sobre bile litogênica

Segundo a Tabela 4.5 e a Figura 4.1, nota-se que a FIP indica que a observação 28 é influente. Em seguida, as observações 9, 7, 1, 11, 12 e 4, nessa ordem, são mais influentes que as demais, pois possuem valores de divergência entre $DPP(z)$ e $DPP_{(j)}(z)$ superior a 0,003. É natural que interpretemos que essas observações são influentes nas estimativas dos parâmetros, incluindo posição e escala, do modelo. Também é inevitável comparar os resultados apontados pela FIP com os resultados apontados por outras medidas de influência da regressão L_1 , como as expostas na Seção 4.5.

Obs	FIP	Obs	FIP	Obs	FIP
28	0,00754	5	0,00100	23	0,00039
9	0,00396	19	0,00086	18	0,00037
7	0,00355	26	0,00080	3	0,00025
1	0,00353	14	0,00072	6	0,00012
11	0,00352	22	0,00070	13	0,00011
12	0,00340	10	0,00069	8	0,00004
4	0,00335	25	0,00061	2	0,00003
15	0,00283	29	0,00051	21	0,00002
16	0,00237	20	0,00042	24	< 0,00001
17	0,00143	27	0,00041		

Tabela 4.5: FIP aplicada aos dados sobre bile litogênica.

A Tabela 4.6 apresenta as medidas de influência vistas na Seção 4.5 e aplicadas ao modelo ajustado via Equação (4.2) aos dados sobre bile litogênica. O detalhe mais importante destacado na Tabela 4.6 é que com exceção da medida baseada no afastamento de verossimilhanças, todas as demais medidas concordam entre si. Isso já era esperado, pois essas medidas mensuram a influência da observação somente na estimativa dos parâmetros de posição.

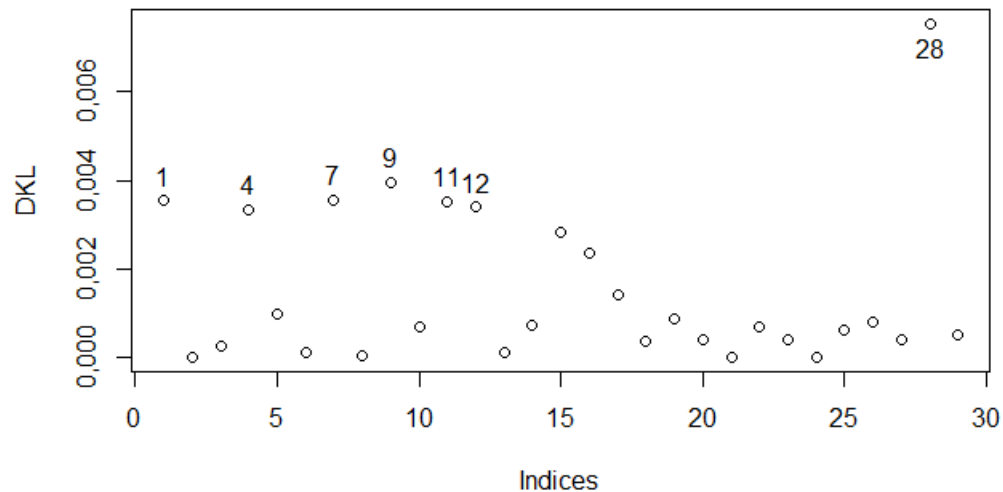


Figura 4.1: Gráfico da medida FIP (DKL entre $DPP(z)$ e $DPP_{(j)}(z)$) do modelo ajustado para variável resposta Concentração.

ordem	obs	FD_{1i}	obs	FD_i	obs	D_i	obs	$LD_i(\beta \lambda)$	obs	$LD_i(\theta)$
1	1	6,466	1	0,055	1	1,771	1	0,469	28	0,619
2	4	6,466	4	0,055	4	1,771	4	0,469	1	0,480
3	7	6,466	7	0,055	7	1,771	7	0,469	7	0,479
4	19	3,766	19	0,034	19	0,652	19	0,144	4	0,472
5	22	3,766	22	0,034	22	0,652	22	0,144	19	0,174
6	26	3,766	26	0,034	26	0,652	26	0,144	26	0,166
7	27	3,766	27	0,034	27	0,652	27	0,144	29	0,159
8	28	3,766	28	0,034	28	0,652	28	0,144	22	0,157
9	29	3,766	29	0,034	29	0,652	29	0,144	27	0,146
10	15	< 0,001	15	< 0,001	15	< 0,001	2	< 0,001	5	0,066
11	16	< 0,001	16	< 0,001	2	< 0,001	3	< 0,001	14	0,048
12	9	< 0,001	9	< 0,001	3	< 0,001	5	< 0,001	10	0,046
13	11	< 0,001	11	< 0,001	5	< 0,001	6	< 0,001	25	0,041
14	12	< 0,001	12	< 0,001	6	< 0,001	8	< 0,001	9	0,039
15	17	< 0,001	17	< 0,001	8	< 0,001	9	< 0,001	15	0,035
16	18	< 0,001	18	< 0,001	10	< 0,001	10	< 0,001	16	0,035
17	20	< 0,001	20	< 0,001	13	< 0,001	11	< 0,001	20	0,028
18	21	< 0,001	21	< 0,001	14	< 0,001	12	< 0,001	17	0,025
19	23	< 0,001	23	< 0,001	9	< 0,001	13	< 0,001	23	0,022
20	24	< 0,001	24	< 0,001	11	< 0,001	14	< 0,001	3	0,014
21	25	< 0,001	25	< 0,001	12	< 0,001	15	< 0,001	11	0,014
22	2	< 0,001	2	< 0,001	16	< 0,001	16	< 0,001	18	0,011
23	3	< 0,001	3	< 0,001	17	< 0,001	17	< 0,001	13	0,008
24	5	< 0,001	5	< 0,001	18	< 0,001	18	< 0,001	6	0,006
25	6	< 0,001	6	< 0,001	20	< 0,001	20	< 0,001	12	0,006
26	8	< 0,001	8	< 0,001	21	< 0,001	21	< 0,001	21	0,002
27	10	< 0,001	10	< 0,001	23	< 0,001	23	< 0,001	8	0,002
28	13	< 0,001	13	< 0,001	24	< 0,001	24	< 0,001	2	0,001
29	14	< 0,001	14	< 0,001	25	< 0,001	25	< 0,001	24	< 0,001

Tabela 4.6: Observações influentes, por meio das demais medidas de influência, no modelo ajustado para a variável Concentração.

As Figuras 4.2, 4.3, 4.4, 4.5 e 4.6 são os gráficos correspondentes às 5 medidas de influência utilizadas na construção da Tabela 4.6.

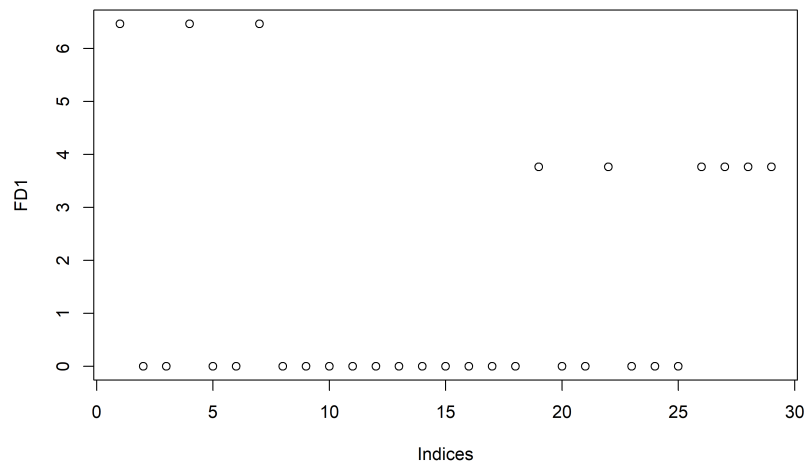


Figura 4.2: Gráfico da medida de influência baseada na norma L_1 do modelo ajustado para variável resposta Concentração.

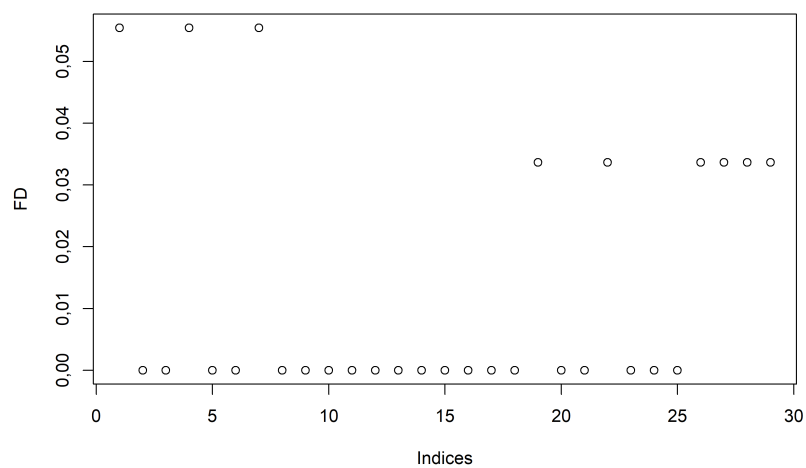


Figura 4.3: Gráfico da medida de influência baseada na norma L_2 do modelo ajustado para variável resposta Concentração.

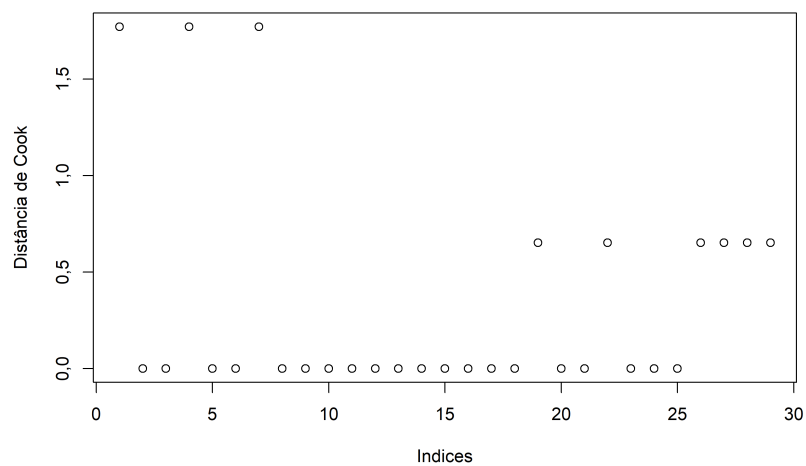


Figura 4.4: Gráfico da distância de Cook do modelo ajustado para variável resposta Concentração.

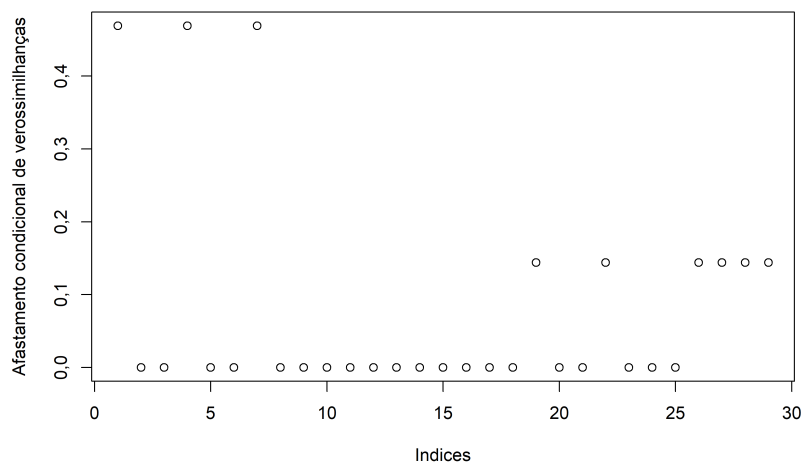


Figura 4.5: Gráfico da medida afastamento condicional de verossimilhanças do modelo ajustado para variável resposta Concentração.

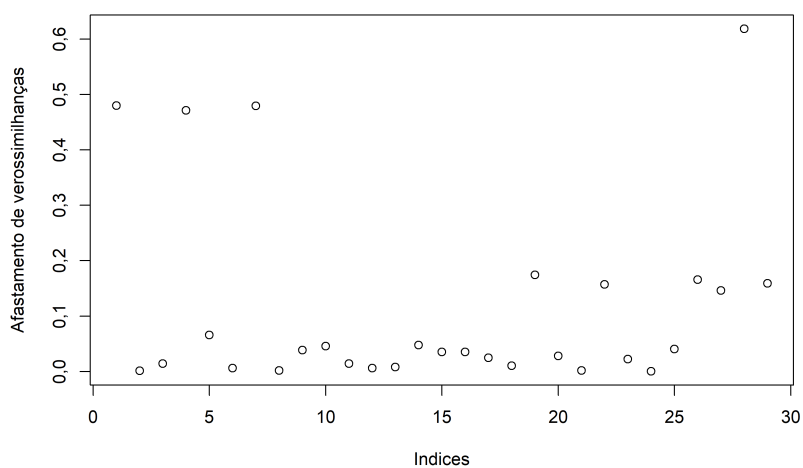


Figura 4.6: Gráfico da medida afastamento de verossimilhanças do modelo ajustado para variável resposta Concentração.

Analisando as Tabelas 4.5 e 4.6, é notável que a medida AV indica a observação 28 como mais influente que todas as demais, coincidindo neste aspecto com a FIP. Isto é um indicativo de que a FIP, da mesma forma que a medida AV, mensura a influência das observações na estimação dos parâmetros de escala e posição do modelo.

Se consideramos como influentes as observações com FIP superior a 0,003, teremos que investigar as observações 28, 9, 7, 1, 11, 12 e 4 – o que engloba as primeiras três observações mais influentes pelas quatro primeiras medidas da Tabela 4.6 e as quatro observações mais influentes da medida AV.

Para os dados sobre bile litogênica, a medida AV fornece como pontos influentes as observações 28, 1, 7 e 4, já as outras quatro medidas indicam que as observações 1, 4 e 7 são influentes. Então a observação 28 deve ser influente majoritariamente na estimação do parâmetro de escala enquanto as observações 1, 4 e 7 se destacam por serem influentes nos parâmetros de posição.

A Tabela 4.7 apresenta os dados das observações influentes destacadas pela Figura

4.1 e a Figura 4.7 exibe o diagrama de dispersão entre as variáveis Concentração e Idade com a reta em vermelho construída a partir do modelo ajustado. A seguir são ressaltados os aspectos mais relevantes acerca das observações 28, 9, 7, 1, 11, 12 e 4 considerando as Tabelas 4.7, 4.2 e 4.1.

- A observação 28 possui o maior valor de Concentração e o segundo maior valor de Idade. Além dos valores extremos nas variáveis resposta e explicativa, esta observação também não se comporta conforme a relação estabelecida pelo modelo ajustado, como ilustra a Figura 4.7.
- A observação 9 possui o menor valor Concentração.
- A observação 7 possui o quinto menor valor tanto para variável Idade quanto para variável Concentração, estando abaixo do primeiro quartil em ambas as variáveis.
- A observação 1 possui o menor valor para variável Idade e o quarto menor valor para variável Concentração, estando abaixo do primeiro quartil em ambas as variáveis.
- A observação 11 possui o oitavo menor valor para variável Concentração.
- A observação 12 possui o quinto menor valor para variável Concentração, tem o mesmo valor para variável Concentração que a observação 7.
- A observação 4 possui o segundo menor valor para variável Concentração e o terceiro menor valor para variável Idade, estando abaixo do primeiro quartil em ambas as variáveis.

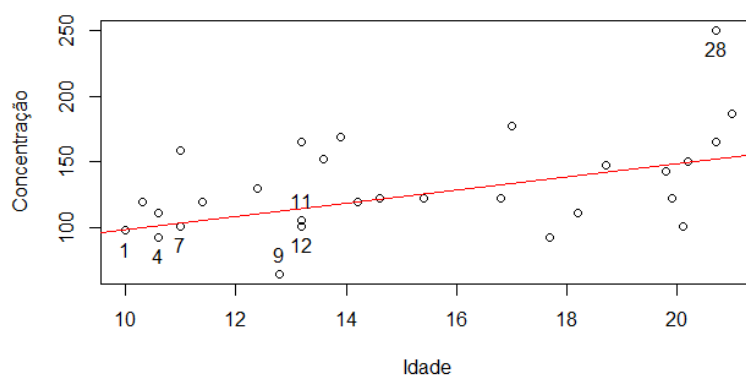


Figura 4.7: Diagrama de dispersão entre as variáveis Concentração e Idade, a reta em vermelho foi construída a partir do modelo ajustado.

Observações	Idade	Concentração
28	20,70	250
9	12,80	65
7	11	101
1	10	98
11	13,20	106
12	13,20	101
4	10,60	93

Tabela 4.7: Observações influentes, pela FIP, no modelo ajustado para a variável Concentração.

Na Subseção 4.6.2 faremos a análise confirmatória com relação às observações influentes obtidas. A análise confirmatória consistirá em ajustar o modelo com todas as observações (modelo completo) e com a exclusão, individual, das observações 28, 9, 7, 1, 11, 12, 4; após o ajuste são verificadas as variações absolutas e percentuais nas estimativas dos parâmetros de posição e escala do modelo. Caso a medida FIP seja um bom indicador para detecção de observações influentes então é esperado que quanto maior o valor FIP associado a uma observação então maior será a variação nas estimativas dos parâmetros de localização e escala.

4.6.2 Análise confirmatória

Faremos a seguir uma breve análise confirmatória, isto é, examinaremos a influência da exclusão das observações 28, 9, 7, 1, 11, 12 e 4 nas estimativas dos parâmetros do modelo. A Tabela 4.8 apresenta as estimativas dos parâmetros do modelo com a exclusão de cada uma das observações supracitadas, entre parêntesis é apresentada a porcentagem de variação entre a estimativa do parâmetro do respectivo modelo e do modelo ajustado com todas as observações.

A exclusão da observação 28 é responsável por mudanças significativas nas estimativas dos parâmetros, em especial, nota-se que a variação na estimativa do parâmetro β_0 chega a ser superior a 28%, sendo que o esperado de uma observação não influente seria algo em torno de 3,5%. Além disso, a estimativa de β_1 diminui em 19,2%. Adicionalmente, dentre as observações examinadas, a observação 28 é a que tem maior influência na estimativa de λ (considerando estimativa via máxima verossimilhança), com a exclusão desta observação a estimativa de λ cai em mais de 10%.

Já a exclusão isolada de cada uma das observações 7, 1 e 4 gera exatamente estimativas idênticas para os parâmetros do modelo. Nestes casos, a estimativa de β_0 aumenta em mais de 50% e a estimativa de β_1 diminui em mais de 30%. As observações 9, 11 e 12 não são influentes na estimativa dos parâmetros de posição, pois a exclusão de qualquer uma destas observações não afeta as estimativas de β_0 e β_1 . Porém, a exclusão de qualquer uma das observações 9, 11 e 12 gera variações na estimativa de λ da

ordem de $-3,6\%$ a $2,2\%$, o que é pouco. De forma que as observações 9, 11 e 12 têm o comportamento esperado de uma observação não influente.

Também é notável que a exclusão da observação 28 é responsável por diminuir a estimativa de λ , implicando que a amplitude dos intervalos de confiança para os parâmetros de posição diminuirá quando esta observação for excluída.

É interessante notar que as observações 7, 1 e 4 são mais influentes que a observação 28 na estimativa de todos os parâmetros de posição, enquanto a observação 28 é mais influente que as demais na estimativa do parâmetro λ . Isso deixa claro que a FIP é de fato uma medida de influência global, pois ela capta as observações mais influentes tanto na estimação dos parâmetros de posição quanto na estimação do parâmetro de escala.

Estimativas	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\lambda}$
completo	49	5	24
sem obs 28	63,038 (28,6%)	4,038 (-19,2%)	21,227 (-11,6%)
sem obs 9	49 (0%)	5 (0%)	23,143 (-3,6%)
sem obs 7	74,13 (51,3%)	3,478 (-30,4%)	24,652 (2,7%)
sem obs 1	74,13 (51,3%)	3,478 (-30,4%)	24,669 (2,8%)
sem obs 11	49 (0%)	5 (0%)	24,536 (2,2%)
sem obs 12	49 (0%)	5 (0%)	24,357 (1,5%)
sem obs 4	74,13 (51,3%)	3,478 (-30,4%)	24,416 (1,7%)

Tabela 4.8: Análise confirmatória para modelo ajustado com dados sobre bile litogênica.

Como conclusão, verificamos que a medida de influência proposta foi capaz de destacar observações cuja exclusão realmente afeta as estimativas de todos os parâmetros do modelo. No entanto, a medida de influência proposta detectou também observações que pouco afetam as estimativas dos parâmetros do modelo.

4.7 Dados sobre incêndios

A Tabela 4.11 apresenta o *Fire data* de Birkes e Dodge (1993). Este *dataset* contém dados de incidência de incêndios em 47 áreas residenciais em Chicago no ano de 1975. As variáveis são: Área, que corresponde ao código da área residencial; Incêndio, taxa de incêndios por 1000 residências; $\log(\text{Incêndio})$, logaritmo natural da variável Incêndio; Idade, porcentagem de unidades habitacionais construídas antes de 1940; Roubo, taxa de roubos por 1000 residentes; Renda, renda familiar média, em que cada unidade equivale a 1000 dólares.

A Tabela 4.9 contém as estatísticas descritivas mínimo, primeiro quartil (Q_1), mediana, média, terceiro quartil (Q_3) e máximo das variáveis $\log(\text{Incêndio})$, Idade, Roubo e Renda.

Estatísticas	$\log(\text{Incêndio})$	Idade	Roubo	Renda
Mínimo	0,693	0,0200	3,00	5,583
Q_1	1,732	0,4860	22,00	8,447
Mediana	2,342	0,6500	29,00	10,694
Média	2,235	0,6033	32,36	10,696
Q_3	2,775	0,7730	38,00	11,989
Máximo	3,681	0,9010	147,00	21,480

Tabela 4.9: Estatísticas descritivas dos dados sobre incêndios.

Nosso objetivo é modelar a variável $\log(\text{Incêndio})$ por meio da regressão L_1 utilizando Idade, Roubo e Renda como variáveis explicativas. A variável $\log(\text{Incêndio})$ será útil para podermos ajustar o modelo L_1 homocedástico, pois se utilizarmos a variável Incêndio como variável resposta violaremos a suposição de homocedasticidade do modelo. Para mais detalhes sobre a heterocedasticidade deste conjunto de dados e a justificativa da transformação empregada na variável resposta, consulte Birkes e Dodge (1993, pg. 66–68). A variável Área não terá utilidade em nossa modelagem, pois coincide com o índice da observação. Antes de ajustar o modelo, devemos calcular a correlação entre as variáveis.

A matriz de correlação das variáveis $\log(\text{Incêndio})$, Idade, Roubo e Renda é expressa na Tabela 4.10. As três variáveis explicativas apresentam valor absoluto da correlação com a variável $\log(\text{Incêndio})$ superior a 0,5. Dado que utilizaremos as variáveis Idade, Roubo e Renda como variáveis explicativas é importante observar a correlação existente entre elas. Neste sentido a correlação entre Idade e Renda chama atenção, pois há correlação em valor absoluto superior a 0,5 entre estas variáveis. Isto é um indicativo que uma dessas variáveis não poderá permanecer no modelo.

	$\log(\text{Incêndio})$	Idade	Roubo	Renda
$\log(\text{Incêndio})$	1,00	0,53	0,51	-0,63
Idade	0,53	1,00	0,32	-0,53
Roubo	0,51	0,32	1,00	-0,17
Renda	-0,63	-0,53	-0,17	1,00

Tabela 4.10: Matriz com as correlações entre as variáveis $\log(\text{Incêndio})$, Idade, Roubo e Renda dos dados sobre incêndios.

Observações	Área	Incêndio	log(Incêndio)	Idade	Roubo	Renda
1	1	6,20	1,82	0,60	29,00	11,74
2	2	9,50	2,25	0,76	44,00	9,32
3	3	10,50	2,35	0,73	36,00	9,95
4	4	77,00	2,04	0,67	37,00	10,66
5	5	8,60	2,15	0,81	53,00	9,73
6	6	34,10	3,53	0,53	68,00	8,23
7	7	11,00	2,40	0,43	75,00	21,48
8	8	6,90	1,93	0,78	18,00	11,10
9	9	7,30	1,99	0,90	31,00	10,69
10	10	15,10	2,71	0,90	25,00	9,63
11	11	29,10	3,37	0,83	34,00	8,00
12	12	2,20	0,79	0,40	14,00	13,72
13	13	5,70	1,74	0,28	11,00	16,25
14	14	2,00	0,69	0,08	11,00	13,69
15	15	2,50	0,92	0,64	22,00	12,40
16	16	3,00	1,10	0,51	17,00	12,20
17	17	5,40	1,69	0,85	27,00	11,60
18	18	2,20	0,79	0,44	9,00	12,77
19	19	7,20	1,97	0,84	29,00	11,08
20	20	15,10	2,71	0,90	30,00	10,51
21	21	16,50	2,80	0,73	40,00	9,78
22	22	18,40	2,91	0,73	32,00	7,34
23	23	36,20	3,59	0,63	41,00	6,57
24	24	39,70	3,68	0,83	147,00	7,46
25	25	18,50	2,92	0,78	22,00	8,01
26	26	23,30	3,15	0,79	29,00	8,18
27	27	12,20	2,50	0,48	46,00	8,21
28	28	5,60	1,72	0,71	23,00	11,23
29	29	21,80	3,08	0,73	4,00	8,33
30	30	21,60	3,07	0,65	31,00	5,58
31	31	9,00	2,20	0,75	39,00	8,56
32	32	3,60	1,28	0,21	15,00	12,10
33	33	5,00	1,61	0,62	32,00	11,88
34	34	28,60	3,35	0,78	27,00	9,74
35	35	17,40	2,86	0,69	32,00	7,52
36	36	11,30	2,42	0,73	34,00	7,39
37	37	3,40	1,22	0,02	17,00	13,84
38	38	11,90	2,48	0,57	46,00	11,04
39	39	10,50	2,35	0,56	42,00	10,33
40	40	10,70	2,37	0,68	43,00	10,91
41	41	10,80	2,38	0,58	34,00	11,16
42	42	4,80	1,57	0,15	19,00	13,32
43	43	10,40	2,34	0,41	25,00	12,96
44	44	15,60	2,75	0,58	28,00	11,26
45	45	7,00	1,95	0,11	3,00	10,08
46	46	71,00	1,96	0,49	23,00	11,43
47	47	4,90	1,59	0,47	27,00	13,73

Fonte: Página 67 de Birkes e Dodge (1993).

Tabela 4.11: *Dados sobre incêndios.*

Iniciaremos a etapa de análise de regressão propriamente dita, utilizaremos um modelo linear com intercepto. A forma funcional do modelo que ajustaremos ao conjunto de dados sobre incêndios é dada pela Equação (4.3).

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \epsilon_i, \quad i = 1, \dots, 47. \quad (4.3)$$

As estimativas, os erros padrão das estimativas e os valores-P (níveis descritivos) do modelo ajustado por meio da Equação (4.3) se encontram na Tabela 4.12. Observamos que o valor-P associado à $\hat{\beta}_1$, correspondente à variável Idade, é superior a 0,59. Por isso ajustaremos um novo modelo sem a variável Idade, a forma funcional deste novo modelo é dada pela Equação (4.4).

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i, \quad i = 1, \dots, 47. \quad (4.4)$$

Parâmetros	Estimativa	Erro Padrão	valor-P
β_0	3,9939	0,7619	< 0,0001
β_1 (Idade)	0,3188	0,5992	0,5975
β_2 (Roubo)	0,0067	0,0052	0,2094
β_3 (Renda)	-0,2087	0,0473	< 0,0001
λ (estimado via $\hat{\lambda}_{MLE}$)	0,3924		
λ (estimado via $\hat{\lambda}_{ROS}$)	0,6085		

Tabela 4.12: Estimativas, erros padrão das estimativas e valores-P no modelo ajustado para variável resposta $\log(\text{Incêndio})$ utilizando as variáveis Idade, Roubo e Renda.

As estimativas, os erros padrão das estimativas e os valores-P (níveis descritivos) do modelo ajustado por meio da Equação (4.4) se encontram na Tabela 4.13, os valores preditos e resíduos estão na Tabela 4.14. As variáveis Roubo e Renda são significativas ao nível de 5%, continuaremos a análise com este último modelo ajustado.

Parâmetros	Estimativa	Erro Padrão	valor-P
β_0	3,7572	0,4394	< 0,0001
β_1 (Roubo)	0,0116	0,0044	0,0117
β_2 (Renda)	-0,1831	0,0356	< 0,0001
λ (estimado via $\hat{\lambda}_{MLE}$)	0,3945		
λ (estimado via $\hat{\lambda}_{ROS}$)	0,4988		

Tabela 4.13: Estimativas, erros padrão das estimativas e valores-P no modelo ajustado para variável resposta $\log(\text{Incêndio})$ utilizando as variáveis Roubo e Renda.

Aplicaremos os testes para verificar se a suposição dos erros possuírem distribuição Laplace é razoável por meio dos cinco testes de hipóteses apresentados no Apêndice

D, ao nível de significância de 5%. Os resultados dos testes encontram-se a seguir, lembrando que o tamanho da amostra é 47 e que os valores críticos dos testes são dados nas Tabelas D.1 a D.5.

1. teste baseado na estatística de Watson: $U^2 = 0,06297133 < 0,083 = U_*^2$
2. teste baseado na estatística de Cramér-von Mises: $W^2 = 0,07195359 < 0,138 = W_*^2$
3. teste baseado na estatística de Anderson-Darling: $A^2 = 0,5592098 < 0,961 = A_*^2$
4. teste baseado na estatística de Kolmogorov-Smirnov: $\sqrt{47}D = 0,7134159 < 0,961 = \sqrt{47}D_*$
5. teste baseado na estatística de Kuiper: $\sqrt{47}V = 1,108183 < 1,305 = \sqrt{47}V_*$

Observação 4.4 Nos cinco testes aplicados utilizamos o valor crítico considerando $n = 50$, pois nas Tabelas D.1 a D.5 não há valor crítico para $n = 47$. Neste tipo de situação há dois caminhos para aplicar os testes, escolher o valor de n mais próximo, ou interpolar. Os valores de n mais próximos presentes nas tabelas são 35 e 50, de forma que $n = 50$ é o valor mais adequado. Optamos por escolher o valor de n mais próximo.

Para os cinco testes, a hipótese nula não foi rejeitada ao nível de 5% de significância, portanto não rejeitamos a hipótese de que os erros possuem distribuição Laplace. Na próxima subseção aplicaremos a técnica de diagnóstico desenvolvida neste capítulo.

Observações	Y_i	\hat{Y}_i	$Y_i - \hat{Y}_i$ (resíduo)	$\hat{\epsilon}_i / \hat{\lambda}_{MLE}$ (resíduo padronizado)
1	1,82	1,94	-0,12	-0,30
2	2,25	2,56	-0,31	-0,78
3	2,35	2,35	-0,00	-0,00
4	2,04	2,23	-0,19	-0,49
5	2,15	2,59	-0,44	-1,11
6	3,53	3,04	0,49	1,25
7	2,40	0,69	1,71	4,33
8	1,93	1,93	0,00	0,00
9	1,99	2,16	-0,17	-0,43
10	2,71	2,28	0,43	1,10
11	3,37	2,69	0,68	1,74
12	0,79	1,41	-0,62	-1,57
13	1,74	0,91	0,83	2,11
14	0,69	1,38	-0,69	-1,74
15	0,92	1,74	-0,82	-2,09
16	1,10	1,72	-0,62	-1,57
17	1,69	1,95	-0,26	-0,66
18	0,79	1,52	-0,74	-1,86
19	1,97	2,06	-0,09	-0,22
20	2,71	2,18	0,54	1,36
21	2,80	2,43	0,37	0,95
22	2,91	2,78	0,13	0,33
23	3,59	3,03	0,56	1,42
24	3,68	4,09	-0,41	-1,04
25	2,92	2,54	0,37	0,95
26	3,15	2,60	0,55	1,40
27	2,50	2,79	-0,28	-0,72
28	1,72	1,97	-0,24	-0,62
29	3,08	2,28	0,80	2,04
30	3,07	3,09	-0,02	-0,05
31	2,20	2,64	-0,44	-1,12
32	1,28	1,71	-0,43	-1,10
33	1,61	1,95	-0,34	-0,87
34	3,35	2,29	1,07	2,71
35	2,86	2,75	0,11	0,27
36	2,42	2,80	-0,37	-0,94
37	1,22	1,42	-0,20	-0,49
38	2,48	2,27	0,21	0,53
39	2,35	2,35	0,00	0,00
40	2,37	2,26	0,11	0,29
41	2,38	2,11	0,27	0,69
42	1,57	1,54	0,03	0,08
43	2,34	1,67	0,67	1,70
44	2,75	2,02	0,73	1,84
45	1,95	1,95	0,00	0,00
46	1,96	1,93	0,03	0,07
47	1,59	1,56	0,03	0,09

Tabela 4.14: Variável (resposta) $\log(\text{Incêndio})$, valores preditos e resíduos obtidos via regressão L_1 .

4.7.1 Aplicação da FIP aos dados sobre incêndios

Segundo a Tabela 4.15 e a Figura 4.8, nota-se que a FIP indica que a observação 7 é influente. Em seguida, as observações 13 e 24, nessa ordem, são mais influentes que as demais, pois possuem valores de divergência entre $DPP(z)$ e $DPP_{(j)}(z)$ superior a 0,004. É natural que interpretemos que essas observações são influentes nas estimativas dos parâmetros do modelo, incluindo posição e escala. Também é inevitável comparar os resultados apontados pela FIP com os resultados apontados por outras medidas de influência da regressão L_1 , como as expostas na Seção 4.5.

Obs	FIP	Obs	FIP	Obs	FIP
7	0,02401	45	0,00034	16	0,00010
13	0,00502	18	0,00029	9	0,00008
24	0,00461	15	0,00029	4	0,00006
47	0,00324	14	0,00025	23	0,00005
43	0,00283	29	0,00025	26	0,00004
42	0,00258	3	0,00022	32	0,00004
38	0,00184	20	0,00020	28	0,00004
46	0,00171	12	0,00017	17	0,00003
40	0,00147	5	0,00016	36	0,00003
44	0,00136	27	0,00014	2	0,00003
6	0,00133	19	0,00014	31	0,00002
8	0,00131	11	0,00013	21	0,00001
39	0,00097	35	0,00013	33	0,00001
41	0,00083	1	0,00012	25	0,00001
34	0,00068	37	0,00012	10	< 0,00001
30	0,00038	22	0,00011		

Tabela 4.15: FIP aplicada aos dados sobre incêndios.

A Tabela 4.16 apresenta as medidas de influência discutidas na Seção 4.5, aplicadas ao modelo ajustado via Equação (4.4) aos dados sobre incêndios. As medidas de influência baseadas nas normas L_1 e L_2 parecem concordar bastante com a distância de Cook, enquanto as duas medidas baseadas em afastamento de verossimilhanças concordam entre si.

Analisando as Tabelas 4.15 e 4.16, observa-se que as cinco medidas de influência da Tabela 4.16 indicam as observações 24, 6, 7 e 13 (este grupo de observações é constituído pelas três observações mais influentes segundo cada uma das cinco medidas de influência) como as mais influentes, diferindo da FIP que indica as observações 7, 13 e 24 como mais influentes. O único consenso que podemos tirar é que a observação 7 muito provavelmente será influente na estimação dos parâmetros de posição e escala do modelo.

ordem	obs	FD_{1i}	obs	FD_i	obs	D_i	obs	$LD_i(\beta \lambda)$	obs	$LD_i(\theta)$
1	24	19,8639	24	11,7043	24	21,3196	24	4,6852	24	4,9597
2	7	15,6381	7	8,3199	7	10,7727	7	0,9040	7	1,7180
3	13	12,1362	13	6,1367	13	5,8608	6	0,6533	6	0,6613
4	42	10,6750	45	5,6584	45	4,9829	45	0,4101	13	0,4709
5	43	10,6750	42	5,3985	47	4,5357	13	0,3884	45	0,4190
6	44	10,6750	43	5,3985	42	4,5357	42	0,2125	43	0,2363
7	47	10,6750	44	5,3985	43	4,5357	43	0,2125	44	0,2346
8	8	9,5007	47	5,3985	44	4,5357	44	0,2125	42	0,2192
9	46	9,5007	38	5,0717	39	4,0031	47	0,2125	47	0,2179
10	38	9,1004	39	5,0717	38	4,0031	14	0,1400	18	0,1606
11	39	9,1004	40	5,0717	40	4,0031	18	0,1400	14	0,1558
12	40	9,1004	41	5,0717	41	4,0031	46	0,1314	8	0,1493
13	41	9,1004	8	4,7229	8	3,4714	8	0,1314	46	0,1455
14	45	9,0537	46	4,7229	46	3,4714	41	0,1082	39	0,1265
15	3	7,3050	5	4,0055	3	2,4969	38	0,1082	40	0,1153
16	5	7,3050	3	4,0055	5	2,4969	39	0,1082	38	0,1101
17	27	7,3050	27	4,0055	27	2,4969	40	0,1082	41	0,1091
18	31	7,3050	31	4,0055	31	2,4969	12	0,0480	34	0,0699
19	30	7,3050	30	4,0055	30	2,4969	16	0,0480	15	0,0632
20	36	7,3050	36	4,0055	36	2,4969	37	0,0480	12	0,0566
21	2	7,3050	2	4,0055	2	2,4969	32	0,0480	16	0,0563
22	6	5,4473	6	3,4150	6	1,8150	15	0,0348	37	0,0526
23	14	2,2000	14	1,0645	14	0,1763	11	0,0321	3	0,0488
24	18	2,2000	18	1,0645	18	0,1763	23	0,0321	32	0,0484
25	12	1,2666	12	0,6116	12	0,0582	26	0,0321	11	0,0452
26	32	1,2666	32	0,6116	32	0,0582	22	0,0321	35	0,0428
27	16	1,2666	16	0,6116	16	0,0582	25	0,0321	22	0,0411
28	37	1,2666	37	0,6116	37	0,0582	35	0,0321	29	0,0410
29	15	1,1016	15	0,5500	15	0,0471	3	0,0294	23	0,0369
30	22	1,0768	22	0,5353	35	0,0446	5	0,0294	26	0,0361
31	25	1,0768	25	0,5353	22	0,0446	27	0,0294	30	0,0351
32	35	1,0768	35	0,5353	25	0,0446	31	0,0294	25	0,0321
33	23	1,0768	23	0,5353	11	0,0446	2	0,0294	31	0,0312
34	11	1,0768	11	0,5353	26	0,0446	30	0,0294	36	0,0305
35	26	1,0768	26	0,5353	23	0,0446	36	0,0294	2	0,0298
36	1	0,6522	1	0,3242	1	0,0164	1	0,0158	5	0,0298
37	33	0,6522	33	0,3242	33	0,0164	17	0,0158	27	0,0297
38	17	0,6522	17	0,3242	17	0,0164	33	0,0158	1	0,0263
39	29	0,6522	29	0,3242	29	0,0164	29	0,0158	17	0,0183
40	10	0,1255	10	0,0608	10	0,0006	21	0,0018	33	0,0161
41	20	0,1255	20	0,0608	20	0,0006	10	0,0018	19	0,0130
42	34	0,1255	34	0,0608	34	0,0006	20	0,0018	9	0,0071
43	21	0,1255	21	0,0608	21	0,0006	34	0,0018	4	0,0057
44	4	< 0,000	4	< 0,000	4	< 0,000	4	< 0,000	20	0,0047
45	9	< 0,000	9	< 0,000	9	< 0,000	9	< 0,000	28	0,0032
46	19	< 0,000	19	< 0,000	19	< 0,000	19	< 0,000	10	0,0020
47	28	< 0,000	28	< 0,000	28	< 0,000	28	< 0,000	21	0,0019

Tabela 4.16: Observações influentes, por meio das demais medidas de influência, no modelo ajustado para a variável log(Incêndio) por meio da Equação (4.4).

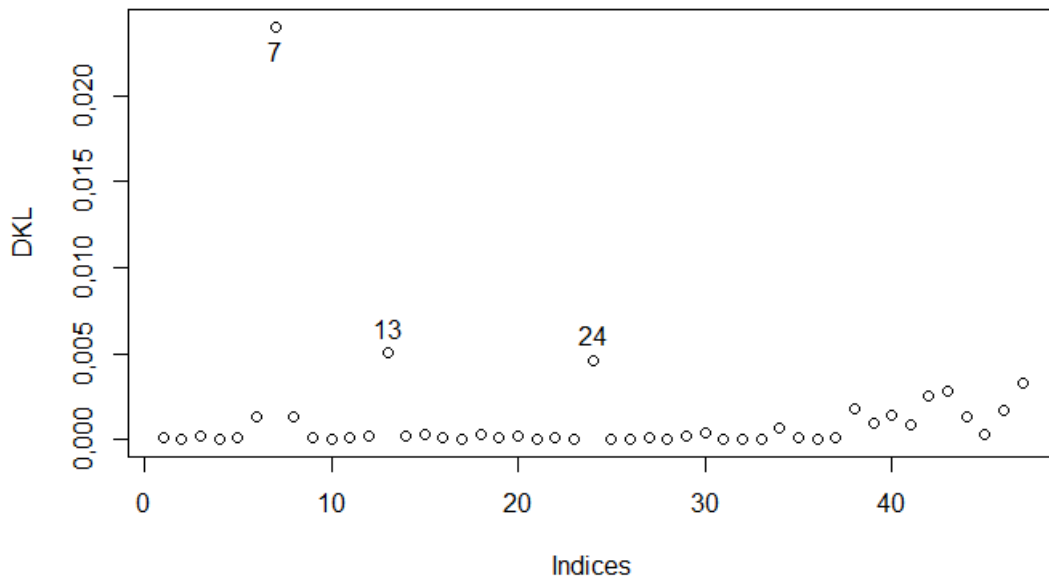


Figura 4.8: Gráfico da medida FIP (DKL entre $DPP(z)$ e $DPP_{(j)}(z)$) do modelo ajustado para variável resposta $\log(\text{Incêndio})$ por meio da Equação (4.4).

As Figuras 4.9, 4.10, 4.11, 4.12 e 4.13 são os gráficos correspondentes às 5 medidas de influência utilizadas na construção da Tabela 4.16.

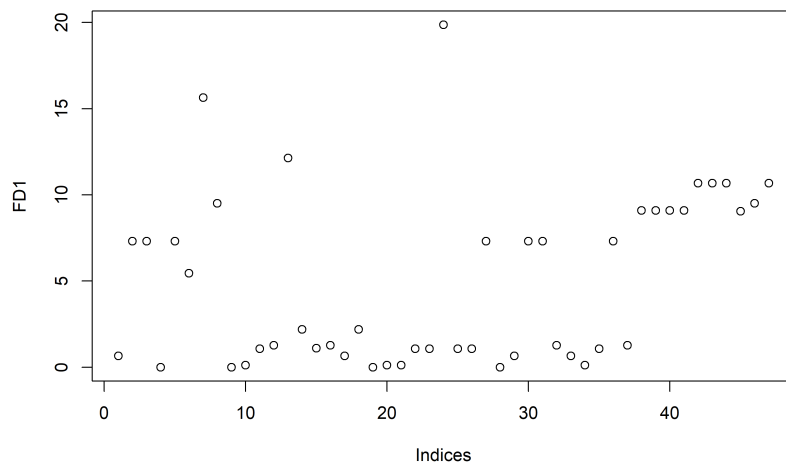


Figura 4.9: Gráfico da medida de influência baseada na norma L_1 do modelo ajustado para variável resposta $\log(\text{Incêndio})$ por meio da Equação (4.4).

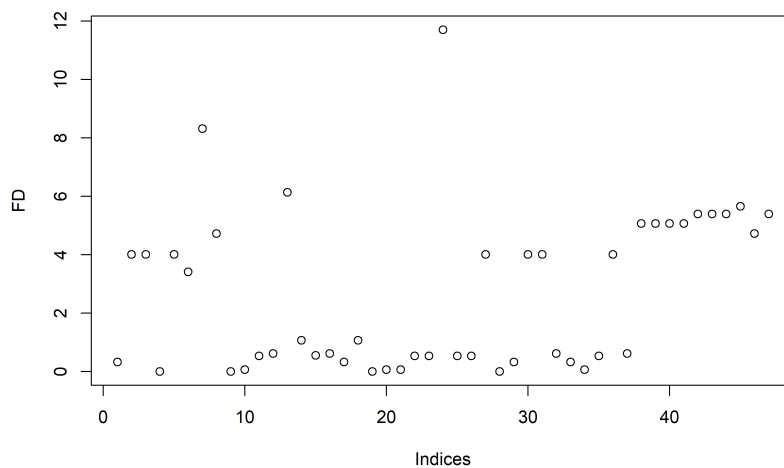


Figura 4.10: Gráfico da medida de influência baseada na norma L_2 do modelo ajustado para variável resposta $\log(\text{Incêndio})$ por meio da Equação (4.4).

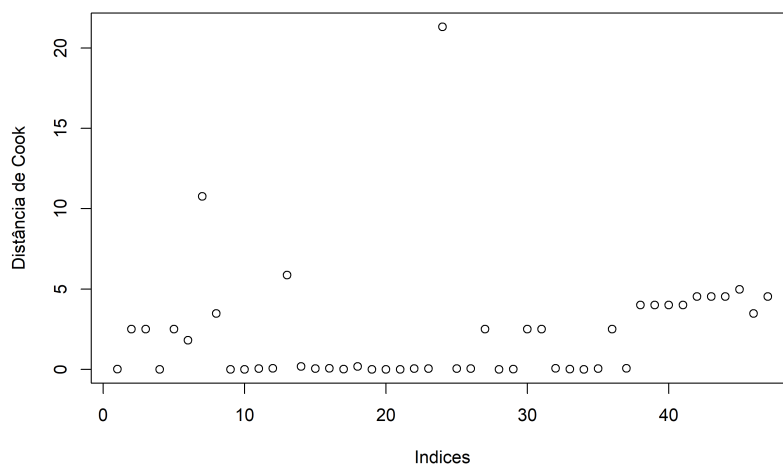


Figura 4.11: Gráfico da distância de Cook do modelo ajustado para variável resposta $\log(\text{Incêndio})$ por meio da Equação (4.4).

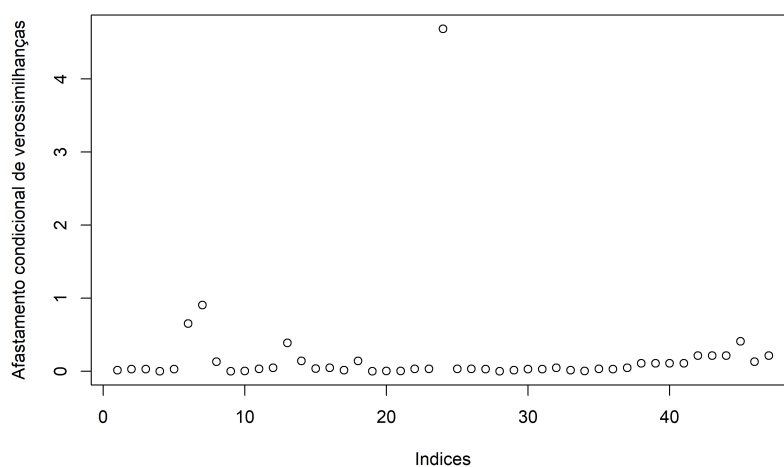


Figura 4.12: Gráfico da medida afastamento condicional de verossimilhanças do modelo ajustado para variável resposta $\log(\text{Incêndio})$ por meio da Equação (4.4).

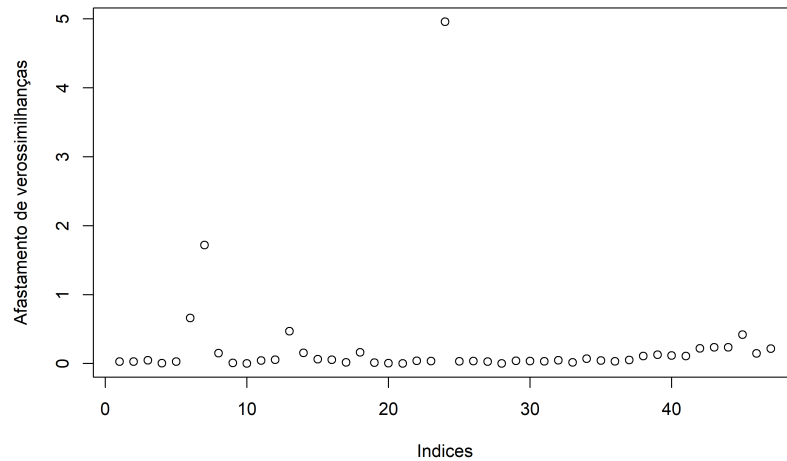


Figura 4.13: Gráfico da medida afastamento de verossimilhanças do modelo ajustado para variável resposta $\log(\text{Incêndio})$ por meio da Equação (4.4).

Se consideramos como influentes as observações com FIP superior a 0,004, teremos que investigar as observações 7, 13 e 24. Também investigaremos as observações mais influentes pelas cinco medidas presentes da Tabela 4.16, que são 24, 7, 6 e 13.

A Tabela 4.17 apresenta os dados das observações influentes destacadas pela Figura 4.8. A seguir são ressaltados os aspectos mais relevantes acerca das observações 7, 13, 24 e 6 considerando as Tabelas 4.17, 4.9 e 4.11.

- A observação 7 possui o maior valor de Renda e o segundo maior valor de Roubo.
- A observação 13 possui o segundo maior valor para variável Renda e o quarto menor valor para variável Roubo.
- A observação 24 possui o maior valor para variável $\log(\text{Incêndio})$, o maior valor para variável Roubo e o quinto menor valor para variável Renda.
- A observação 6 possui o terceiro maior valor para variável $\log(\text{Incêndio})$, o terceiro maior valor para variável Roubo e o décimo primeiro menor valor para variável Renda.

Observações	$\log(\text{Incêndio})$	Roubo	Renda
7	2,40	75,00	21,48
13	1,74	11,00	16,25
24	3,68	147,00	7,46
6	3,53	68,00	8,23

Tabela 4.17: Observações influentes, destacadas pela FIP e outras medidas de influência, no modelo ajustado para a variável $\log(\text{Incêndio})$ por meio da Equação (4.4).

Na Subseção 4.7.2 faremos a análise confirmatória com relação às observações influentes obtidas. A análise confirmatória consistirá em ajustar o modelo com todas as

observações (modelo completo) e com a exclusão, individual, das observações 7, 13, 24 e 6; após o ajuste são verificadas as variações absolutas e percentuais nas estimativas dos parâmetros de posição e escala do modelo. Caso a medida FIP seja um bom indicador para detecção de observações influentes então é esperado que quanto maior o valor FIP associado a uma observação então maior será a variação nas estimativas dos parâmetros de localização e escala.

4.7.2 Análise confirmatória

Faremos a seguir uma breve análise confirmatória, isto é, examinaremos a influência da exclusão das observações 7, 13, 24 e 6 nas estimativas dos parâmetros do modelo. A Tabela 4.18 apresenta as estimativas dos parâmetros do modelo com a exclusão de cada uma das observações supracitadas, entre parêntesis é apresentada a porcentagem de variação entre a estimativa do parâmetro do respectivo modelo e do modelo ajustado com todas as observações.

A exclusão da observação 7 é responsável por mudanças significativas nas estimativas dos parâmetros, em especial, nota-se que a variação na estimativa do parâmetro β_0 chega a ser superior a 18% e a variação na estimativa do parâmetro β_2 chega a ser superior a 32%. Sendo que o esperado de uma observação não influente seria uma variação em torno de 6,4%. Mas dentre as observações examinadas, a de número 7 é a que tem maior influência na estimativa de λ (considerando estimativa via máxima verossimilhança), com a exclusão desta observação, a estimativa de λ diminui em mais de 11%. Esta observação é um nítido ponto de alavanca, pois possui o maior valor de Renda e o segundo maior valor de Roubo (indicando que os valores das variáveis explicativas associadas à observação 7 estão conjuntamente bem distantes no subespaço gerado pelos vetores das variáveis explicativas), isso fica mais claro na Figura 4.14.

Já a exclusão da observação 24 é a que tem maior influência na estimativa do parâmetro β_1 chega a ser superior a 96%. Vale destacar que a observação 24 possui o maior valor para variável log(Incêndio) e o maior valor para variável Roubo. Assim, já era conhecido que esta observação era, simultaneamente, *outlier* na variável resposta e ponto de alavanca.

Além das observações 7 e 24, destacam-se a observação 6 pela influência na estimativa de β_1 , da ordem de -29%, e a observação 13 pela influência na estimativa de β_2 , da ordem de 27%.

Estimativas	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\lambda}$
completo	3,757	0,012	-0,183	0,394
sem obs 7	4,468 (18,9%)	0,007 (-40,1%)	-0,242 (32,3%)	0,35 (-11,2%)
sem obs 13	4,257 (13,3%)	0,011 (-2%)	-0,233 (27,1%)	0,38 (-3,7%)
sem obs 24	3,664 (-2,5%)	0,023 (96,5%)	-0,204 (11,5%)	0,385 (-2,5%)
sem obs 6	3,834 (2%)	0,008 (-29,3%)	-0,182 (-0,8%)	0,392 (-0,6%)

Tabela 4.18: Análise confirmatória para modelo ajustado com dados sobre incêndio.

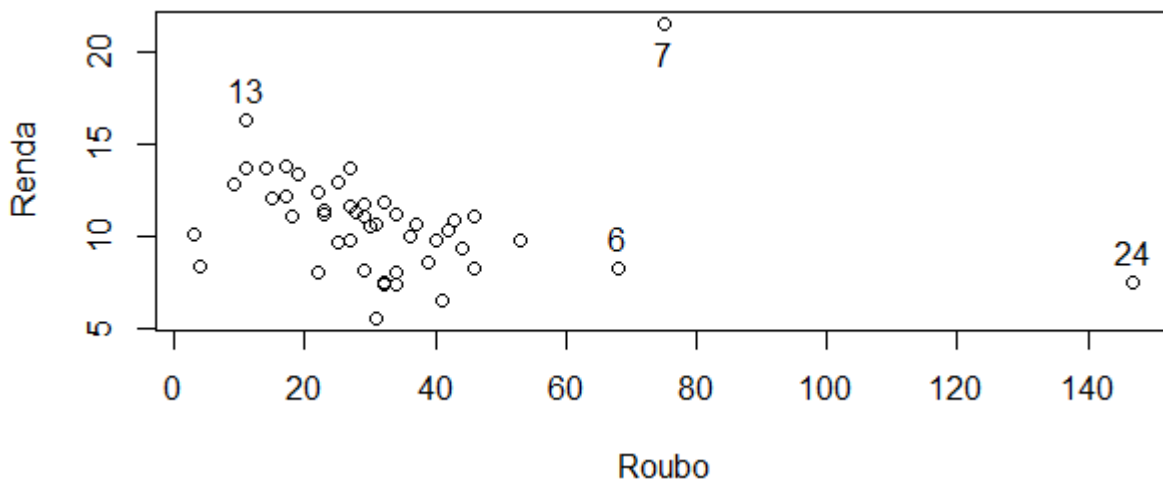


Figura 4.14: Gráfico de dispersão entre as variáveis Roubo e Renda.

Nesta aplicação, a medida AV (Afastamento de Verossimilhanças) conseguiu indicar as observações 24, 7, 6 e 13 como influentes, acertando completamente as observações efetivamente influentes na estimação dos parâmetros de posição e escala do modelo.

Como conclusão, vemos que nesta aplicação a FIP teve êxito em detectar as observações 7, 13 e 24, que são de fato muito influentes na estimação dos parâmetros de posição e escala do modelo. Além disso, a observação 6 ficou na décima primeira posição de ponto mais influente segundo a FIP.

4.8 Considerações sobre as aplicações

Aplicamos o modelo de regressão L_1 linear aos conjuntos de dados sobre bile litigênica e sobre incêndios. Em ambos os casos a hipótese dos erros terem distribuição Laplace foi adequada, o que justificou o uso da regressão L_1 pois o estimador de mínima soma dos erros absolutos é equivalente ao de máxima verossimilhança quando os erros têm distribuição Laplace. De forma que é desejável que os erros possuam distribuição Laplace, embora não seja obrigatório que os erros possuam distribuição Laplace para aplicação da regressão L_1 .

Como conclusão das aplicações realizadas nas Seções 4.6 e 4.7, notamos que a medida de influência proposta foi capaz de destacar observações cuja exclusão realmente afeta as estimativas de todos os parâmetros do modelo. Em especial, os resultados da medida de influência global que propomos é similar, em certos aspectos, com outra medida de influência já estabelecida na literatura, a medida AV. Mas a FIP nem sempre indica as mesmas observações que a medida AV.

Em ambas as aplicações, após a análise confirmatória, ficou ainda mais nítido que as observações destacadas com maior valor da medida FIP são realmente influentes no ajuste global do modelo. Mas vale lembrar que na primeira aplicação, a medida FIP destacou algumas observações que não são influentes, que foram as observações 9, 11 e 12. É interessante notar que a observação indicada pela FIP costuma ser a observação mais influente na estimativa de λ , como foi mostrado em ambas as aplicações. Outro ponto importante é que a ordenação das observações quanto ao grau de influência pela FIP é distinta da ordenação das observações fornecida pelas outras cinco medidas de influência expostas por [Rodrigues \(2019\)](#).

Em aplicações cotidianas faz mais sentido utilizar medidas de influência que detectem observações influentes no ajuste geral do modelo do que métodos que detectem observações influentes apenas nas estimativas dos parâmetros de posição. Isso se aplica, a menos que o interesse seja de fato exclusivamente na estimativa dos parâmetros de posição. Assim, FIP e AV são medidas úteis em contextos gerais e ACV (afastamento condicional de verossimilhanças), distância de Cook e as duas medidas de [Sun e Wei \(2004\)](#) são úteis quando o interesse está em detectar observações influentes nas estimativas dos parâmetros de posição.

No próximo capítulo desenvolveremos uma ferramenta gráfica e iterativa de diagnóstico exploratória, elaborada sob medida para a regressão L_1 .

Capítulo 5

Técnica *Forward Search* para Regressão L_1

Alguns autores, tais como [Bustamante \(2020\)](#) chamam a técnica *forward search* de “procura passo a frente”, mas essa tradução literal soa pouco natural. Por isso, neste trabalho utilizaremos o termo “busca avante” para se referir a técnica *forward search*. A palavra “avante” é sinônimo de “à frente” e “adiante”, sendo um termo adequado para traduzir a palavra “forward” da língua inglesa. Este termo que propomos soa mais natural em Português e transmite de forma objetiva e fidedigna o significado de *forward search*, tal qual este último termo transmite em língua inglesa.

Um livro que trata em profundidade o tópico de busca avante é [Atkinson e Riani \(2000\)](#). [Atkinson et al. \(2010\)](#) trazem uma excelente revisão bibliográfica sobre busca avante, abrangendo publicações feitas até 2010 e contendo várias aplicações da técnica a dados reais. Já [Aoki, Bustamante, e Paula \(2022\)](#) desenvolvem uma técnica de diagnóstico inovadora, que consegue detectar observações que são “mascaradas” por outras observações. Vale destacar que muitas técnicas de diagnóstico falham em identificar as observações influentes mascaradas.

A técnica desenvolvida por [Aoki, Bustamante, e Paula \(2022\)](#) combina a busca avante e a influência local, por isso a técnica é inovadora. Infelizmente, não será possível aplicar esta última técnica mencionada à regressão L_1 , pois a regressão L_1 impõe restrições quanto às derivadas em relação aos parâmetros de posição, inviabilizando o cálculo da influência local.

A seguir, construímos o algoritmo simplificado para a aplicação da técnica busca avante.

Descrição do Algoritmo da busca avante

1. Ajuste a regressão LMS a p observações iniciais, em que p denota o número de parâmetros de posição do modelo.
2. Haverá mais $n-p$ passos, de forma em que a cada passo o número de observações no modelo é incrementado em uma unidade. A observação incluída em cada passo

é a que apresenta menor resíduo absoluto padronizado, considerando o modelo ajustado no passo anterior.

3. Ao decorrer dos passos, monitore as mudanças em medidas importantes, como estimativas de parâmetros e estatística de teste de hipóteses.
4. Examine as variações nas medidas monitoradas, verifique quais observações foram incluídas quando houve as maiores variações nas medidas monitoradas. Aplique a análise confirmatória às observações responsáveis pelas maiores variações nas medidas monitoradas.

Na seção a seguir detalharemos o algoritmo apresentado.

5.1 Técnica busca avante

A busca avante é um procedimento iterativo utilizado na detecção de observações influentes em modelos de regressão. Esta técnica envolve a seleção incremental de observações com base nos resíduos.

Começamos com um ajuste do modelo para poucas observações e depois, a cada passo, ajustamos o modelo a um conjunto cujo número de observações é maior que o conjunto de observações utilizado no passo anterior em uma unidade. Esse processo é executado até que todas as observações estejam no modelo.

O ponto de partida é encontrado ajustando-se o modelo a um grande número de pequenos subconjuntos, usando métodos robustos para determinar qual subconjunto se ajusta melhor. Posteriormente, ordenamos todas as observações por proximidade a este modelo ajustado. O critério de proximidade mais natural são os resíduos absolutos padronizados. Isto é, as observações com menor resíduo absoluto padronizado entram no modelo antes das demais observações. Normalmente entra uma observação por passo da busca, mas às vezes duas ou mais entram no modelo enquanto uma ou mais saem. Como resultado da busca avante temos uma ordenação das observações por proximidade com o modelo ajustado com todas as observações.

Atkinson e Riani (2000) utilizam a regressão LMS (*Least Median of Squares*) como ponto de partida para selecionar as observações que serão incluídas no primeiro passo da busca avançada. As estimativas LMS dos parâmetros do modelo são obtidas a partir da minimização da mediana dos resíduos quadráticos, esta abordagem apresenta ponto de ruptura assintótico de 50%. A função `forwardSearch_regL1` presente no pacote `diagL1` segue esta recomendação e também obtém as observações iniciais para busca avante por meio da regressão LMS.

Observação 5.1 O conceito de ponto de ruptura é importante para comparar métodos robustos e verificar qual método é mais robusto em situações extremas. O valor mínimo

de ponto de ruptura é 0 e o valor máximo é 50%, quanto maior o ponto de ruptura mais robusto é o método. Há um outro conceito importante relacionado ao ponto de ruptura, que é o ponto de ajuste perfeito, Ellis e Morgenthaler (1992) demonstram resultados que relacionam ponto de ruptura (*breakdown point*) e ponto de ajuste perfeito (*exact fit point*). Os conceitos de ponto de ruptura e ponto de ajuste perfeito são abordados em detalhes em Rousseeuw e Leroy (2003). Rodrigues (2019) fornece uma discussão, em língua portuguesa, sobre ponto de ruptura, ponto de ajuste perfeito e regressão L_1 em sua Seção 4.1.

Observação 5.2 Neste ponto do texto é natural que o leitor se pergunte: “qual método de regressão é mais robusto? A regressão L_1 ou a regressão LMS?” A resposta é: a regressão LMS é mais robusta que a regressão L_1 , pois a regressão L_1 é sensível aos *outliers* em X . Neste texto, *outlier* em X é uma observação ou um conjunto de observações cujos vetores de variável explicativas estão “afastados” do subespaço gerado pela maioria dos vetores de variáveis explicativas. Outra pergunta natural é: “Considerando que a regressão LMS é mais robusta que a regressão L_1 , por que não usamos a regressão LMS ao invés da L_1 ?” A resposta a este questionamento tem dois aspectos: 1) a regressão L_1 é mais intuitiva e interpretável que a regressão LMS, 2) a regressão LMS é menos eficiente, no sentido dos estimadores LMS convergirem mais lentamente para os parâmetros (Rousseeuw, 1984).

Ao aplicar a técnica busca avante esperamos que as últimas observações a ingressarem no modelo sejam aquelas que estão mais distantes do modelo e, portanto, poderão causar alterações no ajuste uma vez incluídas. Ou seja, é esperado que as últimas observações a serem incluídas sejam *outliers* e possivelmente influentes na estimativa dos parâmetros de localização e escala.

Se o modelo e os dados concordarem, o ajuste da regressão LMS e o da regressão linear L_1 serão semelhantes, assim como as estimativas dos parâmetros e os resíduos dos dois ajustes. Mas as estimativas dos parâmetros e os resíduos do modelo L_1 ajustado podem oscilar durante a busca avante.

Conforme Atkinson e Riani (2000), o algoritmo de busca avante é composto por três etapas: a primeira diz respeito à escolha de um subconjunto inicial, a segunda refere-se à forma como avançamos na busca avante e a terceira diz respeito ao monitoramento das estatísticas durante as iterações da busca avante. Já explicamos as duas primeiras etapas, agora resta explicar o monitoramento das estatísticas.

Concomitantemente à busca avante, monitoraremos as mudanças em quantidades importantes, como estimativa dos parâmetros de posição e escala e estatísticas para testar hipóteses acerca dos parâmetros de posição. O monitoramento de quantidades chave durante a busca avante fornece uma riqueza de informações não apenas para a detecção de valores discrepantes, mas também sobre a influência de cada observação nos aspectos inferenciais do modelo.

Para mais detalhes sobre a busca avante, consulte a Seção 2.6 de [Atkinson e Riani \(2000\)](#). Na próxima seção explicaremos as alterações necessárias na busca avante discutida pelos autores supracitados, que foi pensada para a regressão via mínimos quadrados e adaptada para a família de modelos lineares generalizados.

5.2 Busca avante na regressão L_1

A literatura estabelecida não aborda a aplicação da busca avante para a regressão L_1 . Além disso, a busca avante foi uma técnica concebida para o modelo ajustado via mínimos quadrados e modelos lineares generalizados. Por exemplo, no contexto da regressão L_1 , faz mais sentido monitorar o resíduo absoluto médio, que é o estimador de máxima verossimilhança de λ , do que o estimador da variância.

Como só conhecemos assintoticamente a distribuição dos estimadores L_1 , então é necessário cautela ao interpretar os gráficos de monitoramento das estimativas dos parâmetros de posição e das estatísticas de teste de significância dos parâmetros de posição para o modelo ajustado – especialmente com relação aos primeiros passos da busca avante.

Em resumo, aplica-se o que foi discutido diversas vezes no Capítulo 1, os resultados inferenciais para regressão L_1 são assintóticos, pois não temos resultados inferenciais exatos para amostras finitas no contexto da regressão L_1 . Tendo isto em mente é possível fazer adaptações à técnica busca avante tradicional e conseqüentemente aplicá-la à regressão linear L_1 .

Na próxima seção apresentaremos aplicações da técnica busca avante aos conjuntos de dados sobre bile litogênica e incêndios. A busca avante será aplicada tendo o mesmo objetivo proposto por [Atkinson e Riani \(2000\)](#), isto é, a técnica será aplicada graficamente e de maneira informal, servindo como ferramenta de diagnóstico exploratória. É esperado que a aplicação da busca avante a dados reais torne sua metodologia mais clara.

5.3 Aplicações

A seguir reexaminaremos os dados sobre bile litogênica e sobre incêndios com objetivo de comparar os resultados obtidos por meio da técnica busca avante com os resultados obtidos com as medidas abordadas no Capítulo 4.

5.3.1 Dados sobre bile litogênica

Retornando aos dados sobre bile litogênica, considere o modelo expresso na Equação (4.2).

A Tabela 5.1 traz os resultados da aplicação da técnica busca avante ao modelo ajustado aos dados sobre bile litogênica. Esta tabela contém a ordem de inclusão das observações no modelo, índices das observações incluídas e estimativas dos parâmetros de posição e de escala ($\hat{\lambda}_{MLE}$) em cada inclusão.

A Figura 5.1 apresenta o gráfico de busca avante dos resíduos L_1 padronizados pela estimativa final $\hat{\lambda}_{MLE}$. As observações com os maiores resíduos positivos são 28, 5, 14 e 10; já as observações com os menores resíduos (maiores resíduos em valor absoluto dentre os resíduos negativos) são 25 e 9. Isso era esperado, visto que na Tabela 5.1 as observações 28, 5, 14, 10, 25 e 9 são as últimas seis observações a serem incluídas no modelo. Vale frisar que uma análise dos resíduos padronizados indicou que o modelo está bem ajustado. Para a análise confirmatória, que será feita posteriormente, examinaremos as observações 28 e 5, pois estas são as observações com resíduos mais extremos.

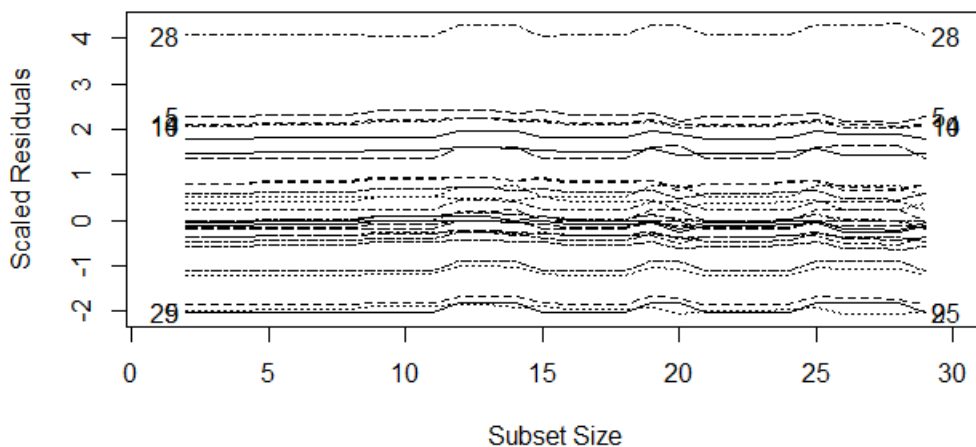


Figura 5.1: Gráfico de busca avante dos resíduos L_1 padronizados pela estimativa final $\hat{\lambda}_{MLE}$ para o modelo ajustado para os dados sobre bile por meio da Equação (4.2).

A Figura 5.2 apresenta o gráfico de busca avante das estimativas dos parâmetros de posição para o modelo ajustado para os dados sobre bile por meio da Equação (4.2).

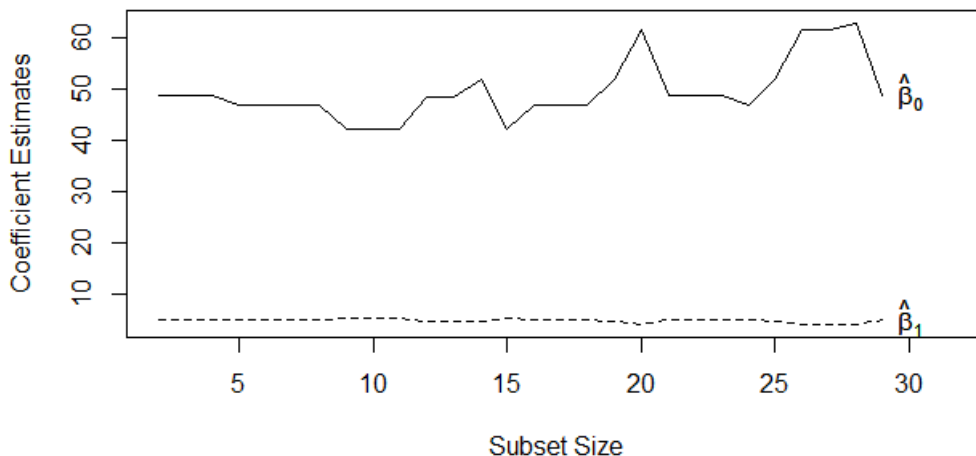


Figura 5.2: Gráfico de busca avante das estimativas dos parâmetros de posição para o modelo ajustado para os dados sobre bile por meio da Equação (4.2).

ordem	obs incluída	$\hat{\beta}_0$ (intercepto)	$\hat{\beta}_1$ (Idade)	$\hat{\lambda}_{MLE}$
2	15 e 16	49,00	5,00	0
3	26	49,00	5,00	0
4	1	49,00	5,00	0,25
5	7	47,02	5,10	0,65
6	17	47,02	5,10	1,13
7	23	47,02	5,10	1,68
8	22	47,02	5,10	2,17
9	4	42,41	5,33	2,81
10	11	42,41	5,33	3,20
11	18	42,41	5,33	3,81
12	12	48,50	4,77	4,45
13	3	48,50	4,77	5,03
14	6	52,08	4,59	5,88
15	27	42,41	5,33	6,33
16	8	47,02	5,10	7,17
17	2	47,02	5,10	7,95
18	24	47,02	5,10	8,98
19	21	52,08	4,59	9,89
20	13	61,68	4,11	11,24
21	29	49,00	5,00	12,33
22	19	49,00	5,00	13,73
23	20	49,00	5,00	15,07
24	9	47,02	5,10	16,41
25	25	52,08	4,59	17,61
26	10	61,68	4,11	18,93
27	14	61,68	4,11	20,09
28	5	63,04	4,04	21,23
29	28	49,00	5,00	24,00

Tabela 5.1: Resultados da aplicação da técnica busca avante ao modelo ajustado aos dados sobre bile litogênica, contendo ordem de inclusão das observações no modelo, índices das observações incluídas e estimativas dos parâmetros de posição e de escala em cada inclusão.

Examinaremos quais observações geraram tais oscilações após suas respectivas inclusões por meio da Figura 5.2 e da Tabela 5.1.

1. A observação 4 foi incluída no passo 9, fazendo com que $\hat{\beta}_0$ passasse de 47,02 para 42,41 (variação de $-9,8\%$), $\hat{\beta}_1$ passasse de 5,10 para 5,33 (variação de 4,5%) e $\hat{\lambda}_{MLE}$ passasse de 2,17 para 2,81 (variação de 29,5%).
2. A observação 12 foi incluída no passo 12, fazendo com que $\hat{\beta}_0$ passasse de 42,41

- para 48,50 (variação de 14,4%), $\hat{\beta}_1$ passasse de 5,33 para 4,77 (variação de -10,5%) e $\hat{\lambda}_{MLE}$ passasse de 3,81 para 4,45 (variação de 16,8%).
3. A observação 27 foi incluída no passo 15, fazendo com que $\hat{\beta}_0$ passasse de 52,08 para 42,41 (variação de -18,6%), $\hat{\beta}_1$ passasse de 4,59 para 5,33 (variação de 16,1%) e $\hat{\lambda}_{MLE}$ passasse de 5,88 para 6,33 (variação de 7,7%).
 4. A observação 21 foi incluída no passo 19, fazendo com que $\hat{\beta}_0$ passasse de 47,02 para 52,08 (variação de 10,8%), $\hat{\beta}_1$ passasse de 5,10 para 4,59 (variação de -10%) e $\hat{\lambda}_{MLE}$ passasse de 8,98 para 9,89 (variação de 10,1%).
 5. A observação 13 foi incluída no passo 20, fazendo com que $\hat{\beta}_0$ passasse de 52,08 para 61,68 (variação de 18,4%), $\hat{\beta}_1$ passasse de 4,59 para 4,11 (variação de -10,5%) e $\hat{\lambda}_{MLE}$ passasse de 9,89 para 11,24 (variação de 13,7%).
 6. A observação 29 foi incluída no passo 21, fazendo com que $\hat{\beta}_0$ passasse de 61,68 para 49,00 (variação de -20,6%), $\hat{\beta}_1$ passasse de 4,11 para 5,00 (variação de 21,7%) e $\hat{\lambda}_{MLE}$ passasse de 11,24 para 12,33 (variação de 9,7%).
 7. A observação 25 foi incluída no passo 25, fazendo com que $\hat{\beta}_0$ passasse de 47,02 para 52,08 (variação de 10,8%), $\hat{\beta}_1$ passasse de 5,10 para 4,59 (variação de -10%) e $\hat{\lambda}_{MLE}$ passasse de 16,41 para 17,61 (variação de 7,3%).
 8. A observação 10 foi incluída no passo 26, fazendo com que $\hat{\beta}_0$ passasse de 52,08 para 61,68 (variação de 18,4%), $\hat{\beta}_1$ passasse de 4,59 para 4,11 (variação de -10,5%) e $\hat{\lambda}_{MLE}$ passasse de 17,61 para 18,93 (variação de 7,5%).
 9. A observação 28 foi incluída no passo 29, fazendo com que $\hat{\beta}_0$ passasse de 63,04 para 49,00 (variação de -22,3%), $\hat{\beta}_1$ passasse de 4,04 para 5,00 (variação de 23,8%) e $\hat{\lambda}_{MLE}$ passasse de 21,23 para 24,00 (variação de 13%).

Assim, mais a frente faremos a análise confirmatória das observações 4, 28 e 29. Examinaremos a observação 4 devido a sua expressiva influência em $\hat{\lambda}_{MLE}$, com variação de quase 30%, já as observações 28 e 29, serão examinadas devido a ambas serem influentes na estimativa dos parâmetros de posição do modelo, com variação absoluta superior a 20% na estimativa de cada parâmetro.

A Figura 5.3 apresenta o gráfico de busca avante das estatísticas de teste de significância dos parâmetros de posição para o modelo ajustado para os dados sobre bile por meio da Equação (4.2). As faixas horizontais centrais na figura estão em $\pm 2,58$, o que cobre 99% de uma distribuição normal padrão. Os saltos na curva são causados pela inclusão de possíveis *outliers*. A curva da estatística de teste para $\hat{\beta}_0$ é suave na maior parte do gráfico, tendo uma mudança abrupta com a inclusão da observação 28. Já a respectiva curva $\hat{\beta}_1$ é notavelmente irregular em torno das inclusões 15 e 20.

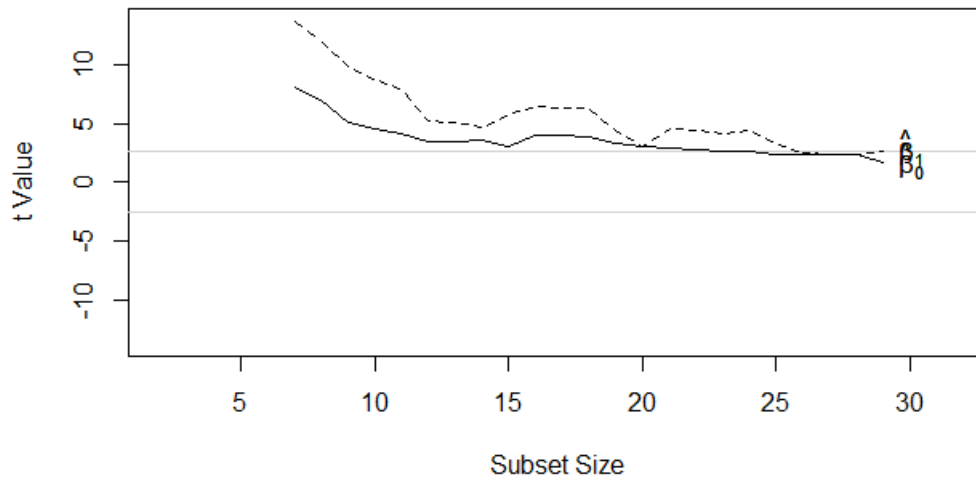


Figura 5.3: Gráfico de busca avante das estatísticas de teste de significância dos parâmetros de posição para o modelo ajustado para os dados sobre bile por meio da Equação (4.2).

Observação 5.3 As estatísticas de teste de significância dos parâmetros de posição presentes na Figura 5.3 foram construídas com base no teste de hipótese 2 da Subseção 1.4.2. Vale lembrar que a estatística do teste 2, da Subseção 1.4.2, sob H_0 possui distribuição assintótica normal padrão. Então é necessário cautela ao examinar os primeiros passos da busca avante, pois para amostras pequenas a aproximação assintótica se mostrará pouco adequada. Esta é uma característica da regressão L_1 e outros métodos de regressão que dependem de resultados inferenciais assintóticos, ou seja, essa cautela não se aplica à busca avante aplicada ao modelo linear normal, porque no caso do modelo linear normal há resultados inferenciais exatos para amostras finitas.

A Figura 5.4 apresenta o gráfico de busca avante da estimativa $\hat{\lambda}_{MLE}$ para o modelo ajustado para os dados sobre bile por meio da Equação (4.2). A curva monótona do gráfico indica que as observações foram ordenadas adequadamente pela busca avante. Para um modelo bem ajustado é esperado uma curva monótona e presença de saltos eventuais no final da curva.

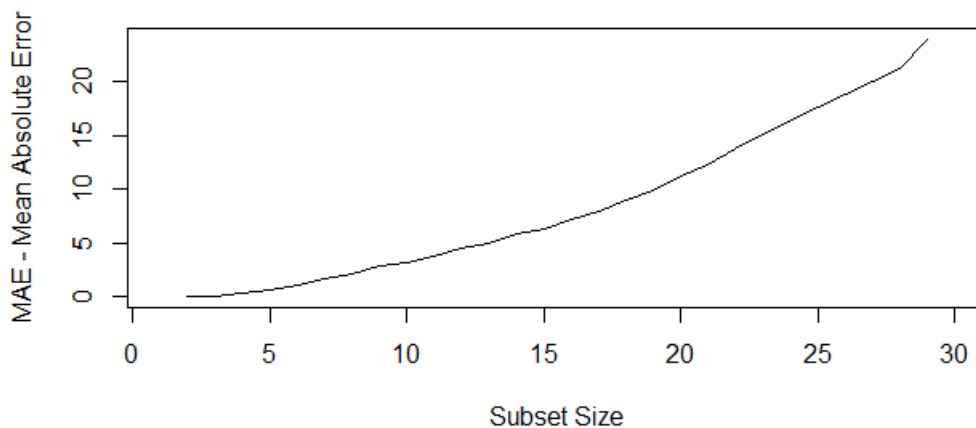


Figura 5.4: Gráfico de busca avante da estimativa $\hat{\lambda}_{MLE}$ para o modelo ajustado para os dados sobre bile por meio da Equação (4.2).

Com a análise da Tabela 5.1 e das Figuras 5.1 a 5.4 concluída partiremos para análise confirmatória. Para esta análise consideraremos as duas observações com os resíduos mais extremos, que estão destacadas na Figura 5.1 e listadas nas duas últimas linhas da Tabela 5.1; bem como as três observações destacadas pela Figura 5.2. Assim, efetivamente examinaremos as observações 4, 5, 28 e 29.

A Tabela 5.2 apresenta as estimativas dos parâmetros do modelo com a exclusão de cada uma das observações supracitadas, entre parêntesis é apresentada a porcentagem de variação entre a estimativa do parâmetro do respectivo modelo e do modelo com todas as observações.

A observação 28, que foi destacada tanto pela Figura 5.1 quanto pela Figura 5.2, mostrou-se a mais influente na estimativa de λ (considerando estimativa via máxima verossimilhança). Isto coincide com um dos resultados apontados pela Tabela 4.8.

A observação mais influente nas estimativas dos parâmetros de posição, destacada pela Tabela 5.2, foi a 4. Mas a busca avante falhou em detectar as observações 1 e 7 como influentes nas estimativas dos parâmetros de posição, conforme destacado pela Tabela 4.8.

Por outro lado, a busca avante conseguiu detectar uma observação tão influente nas estimativas dos parâmetros de posição quanto a observação 28, estamos nos referindo à observação 29. Esta observação não foi detectada pelas medidas de influência aplicadas no Capítulo 4.

Estimativas	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\lambda}$
completo	49	5	24
sem obs 4	74,13 (51,3%)	3,478 (-30,4%)	24,416 (1,7%)
sem obs 5	49 (0%)	5 (0%)	22,893 (-4,6%)
sem obs 28	63,038 (28,6%)	4,038 (-19,2%)	21,227 (-11,6%)
sem obs 29	63,038 (28,6%)	4,038 (-19,2%)	23,521 (-2%)

Tabela 5.2: Análise confirmatória por meio da busca avante para modelo ajustado com dados sobre bile litogênica.

Como conclusão, vemos que a busca avante foi capaz de identificar as observações 28 e 4 como influentes e não trouxe outras observações mais influentes que estas. Pode-se dizer que para a análise desse conjunto de dados a busca avante foi tão eficaz quanto as medidas de influência estudadas no Capítulo 4. Mas vale salientar que a técnica busca avante, por ser uma abordagem mais empírica do que as medidas de influência, exige uma análise mais detalhada e portanto tem aplicação mais complexa que as medidas de influência.

5.3.2 Dados sobre incêndios

Retornando aos dados sobre incêndios, considere o modelo expresso na Equação (4.4).

A Tabela 5.3 traz os resultados da aplicação da técnica busca avante ao modelo ajustado aos dados. Esta tabela contém a ordem de inclusão das observações no modelo, índices das observações incluídas e estimativas dos parâmetros de posição e de escala ($\hat{\lambda}_{MLE}$) em cada inclusão.

Note que duas observações entraram no modelo na inclusão 18, as observações 46 e 10. Na metodologia busca avante é raro que mais de uma observação ingresse no modelo em uma única inclusão, mas isso pode ocorrer. Quando isto ocorre, para cada observação que foi incluída de forma adicional é necessário retirar uma observação que já estava no modelo. Se duas observações foram incluídas no passo 18 da busca então uma observação deve ter sido removida, examinaremos isso agora.

No passo 17 da busca, as observações 1, 8, 11, 12, 14, 17, 18, 19, 21, 25, 26, 28, 32, 33, 38, 40 e 45 estavam incluídas no modelo. Já, no passo 18, da busca as observações 1, 8, 10, 11, 12, 14, 17, 19, 21, 25, 26, 28, 32, 33, 38, 40, 45 e 46 estavam no modelo. Assim, a observação 18, que entrou no passo 14 da busca, foi removida no passo 18, sendo reincluída no passo 19.

A Figura 5.5 apresenta o gráfico de busca avante dos resíduos L_1 padronizados pela estimativa final $\hat{\lambda}_{MLE}$. As observações com os dois maiores resíduos positivos são 7 e 13; dentre as observações com resíduos negativos, a de maior valor absoluto foi a 24. Isso era esperado, visto que na Tabela 5.3 as observações 7 e 13 são as últimas duas observações a serem incluídas o modelo. Além disso, a observação 24 está entre as 6 últimas a serem incluídas no modelo.

Vale frisar que, com exceção das observações 7, 13 e 24, o padrão dos resíduos padronizados é estável, o que indica que o modelo está bem ajustado. Para a análise confirmatória, que será feita posteriormente, examinaremos as observações 7, 13 e 24, pois estas são as observações com resíduos mais extremos.

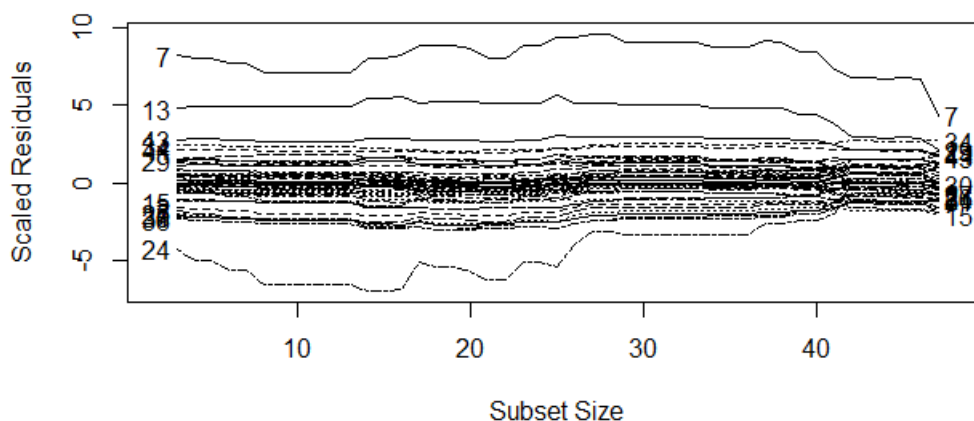


Figura 5.5: Gráfico de busca avante dos resíduos L_1 padronizados pela estimativa final $\hat{\lambda}_{MLE}$.

ordem	obs incluída	$\hat{\beta}_0$ (intercepto)	$\hat{\beta}_1$ (Roubo)	$\hat{\beta}_2$ (Renda)	$\hat{\lambda}_{MLE}$
3	25, 12 e 19	5,32	0,02	-0,35	0,00
4	14	5,31	0,02	-0,35	0,01
5	32	5,31	0,02	-0,35	0,02
6	40	5,23	0,02	-0,35	0,02
7	1	5,23	0,02	-0,35	0,03
8	38	5,12	0,02	-0,34	0,03
9	21	5,12	0,02	-0,34	0,03
10	26	5,12	0,02	-0,34	0,04
11	28	5,12	0,02	-0,34	0,05
12	17	5,12	0,02	-0,34	0,05
13	11	5,12	0,02	-0,34	0,06
14	18	5,51	0,03	-0,38	0,06
15	33	5,51	0,03	-0,38	0,07
16	8	5,58	0,03	-0,39	0,08
17	45	5,72	0,02	-0,38	0,08
18	46 e 10	5,83	0,02	-0,39	0,09
19	18	5,83	0,02	-0,39	0,09
20	41	5,76	0,02	-0,39	0,10
21	9	5,52	0,02	-0,37	0,11
22	16	5,52	0,02	-0,37	0,12
23	39	5,72	0,02	-0,38	0,12
24	3	5,72	0,02	-0,38	0,13
25	4	5,87	0,02	-0,40	0,13
26	6	5,76	0,02	-0,38	0,14
27	23	5,71	0,01	-0,38	0,14
28	15	5,71	0,01	-0,38	0,15
29	35	5,50	0,01	-0,36	0,16
30	22	5,50	0,01	-0,36	0,17
31	37	5,50	0,01	-0,36	0,18
32	2	5,50	0,01	-0,36	0,19
33	29	5,50	0,01	-0,36	0,20
34	20	5,41	0,01	-0,35	0,21
35	42	5,41	0,01	-0,35	0,22
36	47	5,41	0,01	-0,35	0,23
37	5	5,48	0,01	-0,36	0,24
38	27	5,48	0,01	-0,35	0,25
39	31	5,25	0,01	-0,33	0,27
40	30	5,25	0,01	-0,33	0,28
41	36	4,83	0,01	-0,29	0,29
42	24	4,49	0,01	-0,25	0,29
43	44	4,49	0,01	-0,25	0,30
44	43	4,48	0,01	-0,24	0,32
45	34	4,54	0,01	-0,25	0,33
46	13	4,47	0,01	-0,24	0,35
47	7	3,76	0,01	-0,18	0,39

Tabela 5.3: Resultados da aplicação da técnica busca avante ao modelo ajustado aos dados sobre incêndios, contendo a ordem de inclusão das observações no modelo, índices das observações incluídas e estimativas dos parâmetros de posição e de escala em cada inclusão.

A Figura 5.6 apresenta o gráfico de busca avante das estimativas dos parâmetros de posição. Tanto a estimativa $\hat{\beta}_1$ quanto a estimativa $\hat{\beta}_2$ permanecem estáveis durante toda a busca. Mas a estimativa do intercepto, $\hat{\beta}_0$, oscila após a entrada de algumas observações.

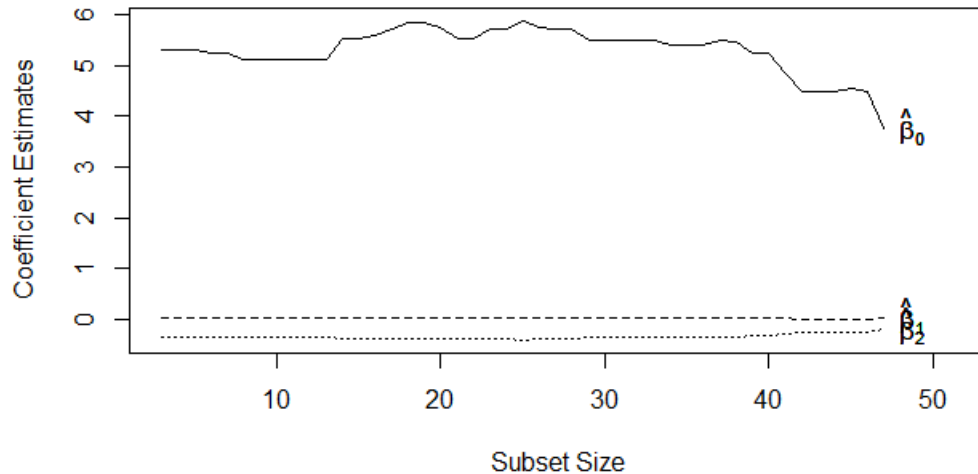


Figura 5.6: Gráfico de busca avante das estimativas dos parâmetros de posição para o modelo ajustado para os dados sobre incêndios por meio da Equação (4.4).

Examinaremos quais observações geraram tais oscilações após suas respectivas inclusões por meio da Figura 5.6 e da Tabela 5.3. Como as estimativas de β_1 e β_2 estão bastante estáveis, focaremos na variação das estimativas do intercepto do modelo, β_0 e na estimativa de λ .

1. A observação 18 foi incluída no passo 14, fazendo com que $\hat{\beta}_0$ passasse de 5,12 para 5,51 (variação de 7,6%) e $\hat{\lambda}_{MLE}$ passasse de 0,06 para 0,06 (variação de 0%).
2. A observação 8 foi incluída no passo 16, fazendo com que $\hat{\beta}_0$ passasse de 5,51 para 5,58 (variação de 1,3%) e $\hat{\lambda}_{MLE}$ passasse de 0,07 para 0,08 (variação de 14,3%).
3. A observação 45 foi incluída no passo 17, fazendo com que $\hat{\beta}_0$ passasse de 5,58 para 5,72 (variação de 2,5%) e $\hat{\lambda}_{MLE}$ passasse de 0,08 para 0,08 (variação de 0%).
4. As observações 46 e 10 foram incluídas no passo 18, fazendo com que $\hat{\beta}_0$ passasse de 5,72 para 5,83 (variação de 1,9%) e $\hat{\lambda}_{MLE}$ passasse de 0,08 para 0,09 (variação de 12,5%).
5. A observação 41 foi incluída no passo 20, fazendo com que $\hat{\beta}_0$ passasse de 5,83 para 5,76 (variação de -1,2%) e $\hat{\lambda}_{MLE}$ passasse de 0,09 para 0,10 (variação de 11,1%).

6. A observação 9 foi incluída no passo 21, fazendo com que $\hat{\beta}_0$ passasse de 5,76 para 5,52 (variação de $-4,2\%$) e $\hat{\lambda}_{MLE}$ passasse de 0,10 para 0,11 (variação de 10%).
7. A observação 39 foi incluída no passo 23, fazendo com que $\hat{\beta}_0$ passasse de 5,52 para 5,72 (variação de $3,6\%$) e $\hat{\lambda}_{MLE}$ passasse de 0,12 para 0,12 (variação de 0%).
8. A observação 4 foi incluída no passo 25, fazendo com que $\hat{\beta}_0$ passasse de 5,72 para 5,87 (variação de $2,6\%$) e $\hat{\lambda}_{MLE}$ passasse de 0,13 para 0,13 (variação de 0%).
9. A observação 35 foi incluída no passo 29, fazendo com que $\hat{\beta}_0$ passasse de 5,71 para 5,50 (variação de $-3,7\%$) e $\hat{\lambda}_{MLE}$ passasse de 0,15 para 0,16 (variação de $6,7\%$).
10. A observação 20 foi incluída no passo 34, fazendo com que $\hat{\beta}_0$ passasse de 5,50 para 5,41 (variação de $-1,6\%$) e $\hat{\lambda}_{MLE}$ passasse de 0,20 para 0,21 (variação de 5%).
11. A observação 5 foi incluída no passo 37, fazendo com que $\hat{\beta}_0$ passasse de 5,41 para 5,48 (variação de $1,3\%$) e $\hat{\lambda}_{MLE}$ passasse de 0,23 para 0,24 (variação de $4,3\%$).
12. A observação 31 foi incluída no passo 39, fazendo com que $\hat{\beta}_0$ passasse de 5,48 para 5,25 (variação de $-4,2\%$) e $\hat{\lambda}_{MLE}$ passasse de 0,25 para 0,27 (variação de 8%).
13. A observação 36 foi incluída no passo 41, fazendo com que $\hat{\beta}_0$ passasse de 5,25 para 4,83 (variação de -8%) e $\hat{\lambda}_{MLE}$ passasse de 0,28 para 0,29 (variação de $3,6\%$).
14. A observação 24 foi incluída no passo 42, fazendo com que $\hat{\beta}_0$ passasse de 4,83 para 4,49 (variação de -7%) e $\hat{\lambda}_{MLE}$ passasse de 0,29 para 0,29 (variação de 0%).
15. A observação 34 foi incluída no passo 45, fazendo com que $\hat{\beta}_0$ passasse de 4,48 para 4,54 (variação de $1,3\%$) e $\hat{\lambda}_{MLE}$ passasse de 0,32 para 0,33 (variação de $3,1\%$).
16. A observação 13 foi incluída no passo 46, fazendo com que $\hat{\beta}_0$ passasse de 4,54 para 4,47 (variação de $-1,5\%$) e $\hat{\lambda}_{MLE}$ passasse de 0,33 para 0,35 (variação de $6,1\%$).

17. A observação 7 foi incluída no passo 47, fazendo com que $\hat{\beta}_0$ passasse de 4,47 para 3,76 (variação de $-15,9\%$) e $\hat{\lambda}_{MLE}$ passasse de 0,35 para 0,39 (variação de $11,4\%$).

Assim, mais a frente faremos a análise confirmatória das observações 7, 18, 24 e 36. O critério de escolha destas observações foi o fato delas gerarem uma variação absoluta igual ou superior a 7% na estimativa de β_0 .

A Figura 5.7 apresenta o gráfico de busca avante das estatísticas de teste de significância dos parâmetros de posição. As faixas horizontais centrais na figura estão em $\pm 2,58$, o que cobre 99% de uma distribuição normal padrão. Os saltos na curva são causados pela inclusão de possíveis *outliers*. A curva da estatística de teste para β_1 é mais suave que a curva associada às demais estimativas. Já as curvas associadas a $\hat{\beta}_0$ e $\hat{\beta}_2$ são mais irregulares.

As oscilações mais relevantes na curva da estatística de teste para β_0 ocorrem em duas regiões: entre os passos 25 e 30 da busca e entre os passos 35 e 40 da busca. Esta mesma observação é válida para a curva da estatística de teste para β_2 . Entretanto, apesar das oscilações presentes nas curvas das estatísticas de teste para β_0 e β_2 , elas permanecem fora das faixas centrais durante toda a busca, isto é indicativo de que esses parâmetros são realmente significativos. Já a curva correspondente a estatística de teste para $\hat{\beta}_1$ permanece dentro das faixas centrais, que correspondem ao intervalo de 99% de uma distribuição normal padrão, entre os passos 39 e 46. Mas como estamos conduzindo esta análise com o nível de significância fixado em 5% então isto indica que o parâmetro β_1 também é significativo para o modelo.

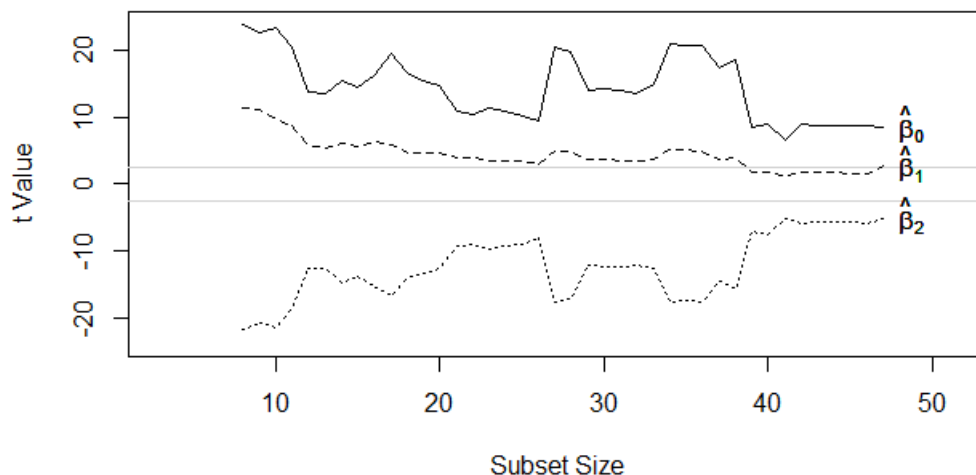


Figura 5.7: Gráfico de busca avante das estatísticas de teste de significância dos parâmetros de posição para o modelo ajustado para os dados sobre incêndios por meio da Equação (4.4).

Observação 5.4 As estatísticas de teste de significância dos parâmetros de posição presentes na Figura 5.7 foram construídas com base no teste de hipótese 2 da Subseção 1.4.2. Vale lembrar que a estatística do teste 2, da Subseção 1.4.2, sob H_0 possui

distribuição assintótica normal padrão. Então é necessário cautela ao examinar os primeiros passos da busca avante, pois para amostras pequenas a aproximação assintótica se mostrará pouco adequada. Esta é uma característica da regressão L_1 e outros métodos de regressão que dependem de resultados inferenciais assintóticos, ou seja, essa cautela não se aplica à busca avante aplicada ao modelo linear normal, porque no caso do modelo linear normal há resultados inferenciais exatos para amostras finitas.

A Figura 5.8 apresenta o gráfico de busca avante da estimativa $\hat{\lambda}_{MLE}$ para o modelo ajustado para os dados sobre incêndios por meio da Equação (4.4). A curva monótona do gráfico indica que as observações foram ordenadas adequadamente pela busca avante. Para um modelo bem ajustado é esperado uma curva monótona e eventualmente presença de saltos no final da curva.

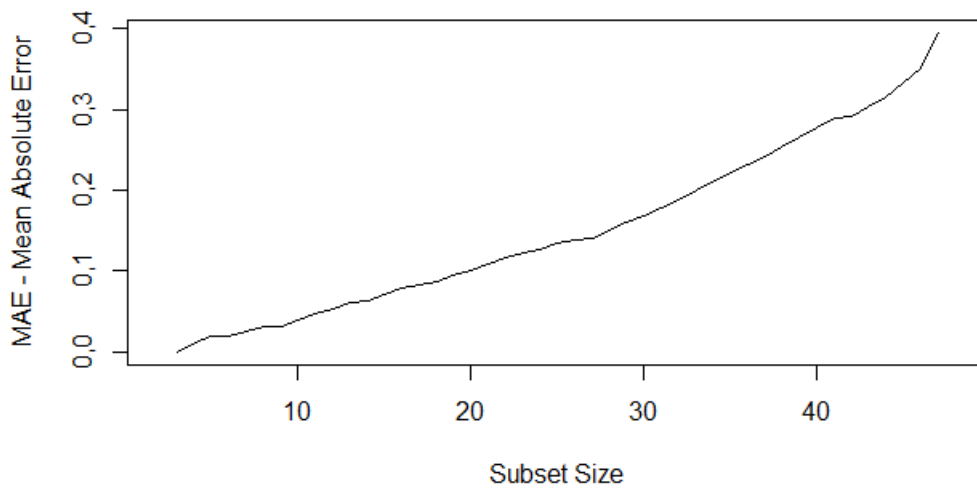


Figura 5.8: Gráfico de busca avante da estimativa $\hat{\lambda}_{MLE}$ para o modelo ajustado para os dados sobre incêndios por meio da Equação (4.4).

Concluídas as análises da Tabela 5.3 e das Figuras 5.5 a 5.8, partiremos para análise confirmatória. Para esta análise consideraremos as três observações com os resíduos mais extremos, que estão destacadas na Figura 5.5 e listadas entre as seis últimas linhas da Tabela 5.3; bem como as quatro observações destacadas pela Figura 5.6. Assim, efetivamente examinaremos as observações 7, 13, 18, 24 e 36.

A Tabela 5.4 apresenta as estimativas dos parâmetros do modelo com a exclusão de cada uma das observações supracitadas, entre parêntesis é apresentada a porcentagem de variação entre a estimativa do parâmetro do respectivo modelo e do modelo com todas as observações.

A observação 7, que foi destacada tanto pela Figura 5.5 quanto pela Figura 5.6, mostrou-se a mais influente na estimativa de λ (considerando estimativa via máxima verossimilhança) e na estimativa do intercepto. Isto coincide com um dos resultados apontados pela Tabela 4.18. A observação mais influente na estimativas do parâmetro β_1 pela Tabela 5.4 foi a 24.

Estimativas	$\hat{\beta}_0$ (intercepto)	$\hat{\beta}_1$ (Roubo)	$\hat{\beta}_2$ (Renda)	$\hat{\lambda}$
completo	3,757	0,012	-0,183	0,394
sem obs 7	4,468 (18,9%)	0,007 (-40,1%)	-0,242 (32,3%)	0,35 (-11,2%)
sem obs 13	4,257 (13,3%)	0,011 (-2%)	-0,233 (27,1%)	0,38 (-3,7%)
sem obs 18	3,724 (-0,9%)	0,011 (-5,5%)	-0,177 (-3,1%)	0,387 (-1,9%)
sem obs 24	3,664 (-2,5%)	0,023 (96,5%)	-0,204 (11,5%)	0,385 (-2,5%)
sem obs 36	4,151 (10,5%)	0,01 (-9,4%)	-0,217 (18,4%)	0,393 (-0,4%)

Tabela 5.4: Análise confirmatória por meio da busca avante para modelo ajustado com dados sobre incêndios.

Como conclusão, vemos que a busca avante foi capaz de identificar as observações 7 e 24 como influentes e não trouxe outras observações mais influentes que estas. Pode-se dizer que para a análise desse conjunto de dados a busca avante foi tão eficaz quanto as medidas de influência estudadas no Capítulo 4. Novamente aqui salientamos que a técnica busca avante, por ser uma abordagem mais empírica do que as medidas de influência, exige uma análise mais detalhada e portanto tem aplicação mais complexa que as medidas de influência.

Capítulo 6

Conclusões e Trabalhos Futuros

Nesta tese fizemos uma apresentação abrangente sobre a regressão linear L_1 . Além disso, desenvolvemos testes para verificar a hipótese de linearidade do modelo e a hipótese de homogeneidade de variâncias. Também desenvolvemos uma nova medida de influência com base em verossimilhanças preditivas e adaptamos a busca avante (*forward search*), um procedimento iterativo utilizado na detecção de observações influentes em modelos de regressão, ao contexto da regressão L_1 .

Um produto adicional desta pesquisa foi o desenvolvimento do pacote *diagL1* (Rodrigues e Elian, 2024), que permite utilizar as técnicas desenvolvidas e aplicadas nesta tese de forma simples e prática. Nos próximos parágrafos apresentaremos as funcionalidades do *diagL1*.

No pacote há quatro funções relacionadas à distribuição Laplace: *dlaplace* calcula a função de densidade de probabilidade, *plaplace* computa a função de distribuição acumulativa, *qlaplace* fornece o quantil correspondente a uma probabilidade e *rlaplace* gera números aleatórios a partir da distribuição Laplace.

A função *regL1* serve para ajustar modelos de regressão L_1 homocedásticos, enquanto a função *regL1_het* é utilizada para ajustar modelos de regressão L_1 heterocedásticos. A função *laplace.dist.test* contém os testes para verificar a suposição de que os erros possuem distribuição Laplace conforme o Apêndice D.

A função *LF_test* é utilizada para aplicar os três testes de falta de ajuste desenvolvidos no Capítulo 2, enquanto a função *HV_test* destina-se à aplicação dos testes de homogeneidade de variâncias elaborados no Capítulo 3.

Também há funções para o cálculo das medidas de influência para modelos de regressão L_1 apresentadas na Seção 4.5: *likelihoodD*, que calcula o afastamento de verossimilhanças; *likelihoodCD*, que calcula o afastamento condicional de verossimilhanças; *CookDistance*, que determina a distância de Cook; e *FD*, que computa uma medida de influência baseada na norma L_1 ou L_2 para um modelo L_1 ajustado. Além disso, há a função *PIF*, que aplica a função de influência preditiva, desenvolvida no Capítulo 4, ao modelo de regressão L_1 . O termo designado à função é *PIF*, uma vez que corresponde

às iniciais de “*predictive influence function*”, expressão em Inglês que traduzida significa função de influência preditiva.

A função *forwardSearch_regL1* aplica a técnica busca avante, em conformidade com o exposto no Capítulo 5, para modelos de regressão L_1 . O método *plot* da classe *forwardSearch_regL1* serve para construir os gráficos da técnica busca avante.

As funções *boot.regL1* e *boot_het.regL1* servem para obter, via *bootstrap*, erros padrão, intervalos de confiança e testes de hipóteses referentes a modelos de regressão linear L_1 , nos cenários homocedásticos e heterocedásticos, respectivamente.

Os conjuntos de dados examinados nesta tese, sobre bile litogênica e sobre incêndios, também estão incluídos no *diagL1*. Existem ainda algumas outras funções no pacote, estas servem como funções auxiliares para aquelas funções já mencionadas. Vale destacar que o pacote *diagL1* trata somente de modelos de regressão L_1 lineares.

Há alguns tópicos que podem ser explorados em estudos futuros. Listarei os tópicos de pesquisa organizados de acordo com os respectivos capítulos relacionados.

1. Para o Capítulo 2, uma sugestão de pesquisa é demonstrar um resultado mais geral, que permita aplicar um teste de falta de ajuste no caso em que o modelo possui mais de uma variável explicativa. Outra sugestão é desenvolver estatísticas não paramétricas para testar a hipótese de linearidade do modelo de regressão L_1 , posteriormente comparar tais testes com os testes LF_1 e LF_2 em estudos de simulação.
2. Para o Capítulo 3, uma sugestão é estudar de forma mais aprofundada, de preferência analiticamente, a distribuição das estatísticas propostas para testar a hipótese de homogeneidade de variâncias sob H_0 . Um outro ponto importante é desenvolver uma rotina computacional que determine o rótulo de grupos de forma “padrão” por meio da análise de *cluster* e aplique o teste de homogeneidade de variâncias em sequência, de forma que o usuário não tenha que se preocupar com a parte de análise de *cluster*.
3. Para o Capítulo 4, uma sugestão para trabalhos futuros é avaliar a eficiência das medidas de influência discutidas em casos nos quais a suposição de erros com distribuição Laplace não é satisfeita, especialmente com distribuições leptocúrticas (caudas pesadas).
4. Para o Capítulo 5, uma sugestão para trabalhos futuros é adaptar a técnica busca avante para o contexto da regressão quantílica.

Apêndice A

Nomenclaturas para Regressão L_1

É importante notar que, diferentemente do que ocorre com respeito a nomenclatura da regressão via mínimos quadrados, há dissenso em como a literatura denomina a regressão L_1 . Sendo assim, coexistem diversas nomenclaturas para o método de mínima soma de erros absolutos. Conhecer as diversas terminologias adotadas para regressão L_1 facilita a busca por artigos, evitando que literatura relevante seja ignorada. Por esta razão, [Bloomfield e Steiger \(1983\)](#) fizeram uma lista com os termos mais recorrentes na literatura e nós transcrevemos e traduzimos tais termos a seguir – para facilitar a organização, utilizamos o formato: nome em Inglês (sigla em Inglês) [nome em Português].

- *Absolute Deviation Curve-fitting* (AD) [Ajuste de curva via desvio absoluto];
- *Least Absolute Error* (LAE) [Mínimo erro absoluto];
- *Least Absolute Residual Regression* (LAR) [Regressão via mínimo resíduo absoluto];
- *Least Absolute Value Regression* (LAV) [Regressão via mínimo valor absoluto];
- *Least Deviations* (LD) [Desvios mínimos];
- *Least Total Deviation* (LTD) [Desvio total mínimo];
- *Minimum Absolute Deviation Regression* (MAD) [Regressão via desvio absoluto mínimo];
- *Minimum Deviation* (MD) [Desvio mínimo];
- *Minimum Sum of Absolute Errors* (MSAE) [Soma mínima de erros absolutos];
- *Sum of Absolute Deviations* (SAD) [Soma de desvios absolutos];
- *Sum of Absolute Errors* (SAE) [Soma de erros absolutos];
- *Sum of Absolute Value of Deviations* (SAV) [Soma de valores absolutos dos desvios].

Apêndice B

Norma L_k

Uma norma é uma função que associa cada vetor de um espaço vetorial a um número real não-negativo. O conceito de norma está relacionado à noção geométrica de comprimento. No contexto da regressão linear temos um vetor de variáveis respostas pertencente ao conjunto \mathbb{R}^n , em que n é o tamanho da amostra e estamos interessados em obter a reta que possua a menor “distância” possível do vetor de variáveis respostas. Para isso é necessário que esta “distância” a ser minimizada seja definida a princípio.

As normas L_k são usualmente utilizadas como medida de distância de um vetor com relação à origem e são definidas pela Equação (B.1),

$$\|\mathbf{Y}\|_k = \left(\sum_{i=1}^n |Y_i|^k \right)^{1/k}, \quad 1 \leq k, \quad (\text{B.1})$$

em que Y_i é a i -ésima componente do vetor \mathbf{Y} . Também é possível definir uma métrica (medida de distância entre dois vetores de mesma dimensão) a partir de uma norma, sendo esta uma das mais importantes aplicações das normas.

Dois casos particulares importantes são as normas L_2 e L_1 que são dadas, respectivamente, por

$$\|\mathbf{Y}\|_2 = \left(\sum_{i=1}^n (Y_i)^2 \right)^{1/2} \quad \text{e}$$
$$\|\mathbf{Y}\|_1 = \sum_{i=1}^n |Y_i|.$$

A norma L_2 é conhecida como distância euclidiana. É importante ressaltar que a reta de mínimos quadrados minimiza a distância euclidiana pois

$$\operatorname{argmin}_{\hat{\mathbf{y}}} \left(\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \right)^{1/2} = \operatorname{argmin}_{\hat{\mathbf{y}}} \left(\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \right)$$

devido a monotonicidade da função potência. Enquanto a reta de mínima soma dos desvios

absolutos minimiza a norma L_1 , daí é derivado o nome da regressão L_1 .

Também há a norma L_∞ dada na Equação (B.2), que é um caso limite da norma apresentada (B.1) quando $k \rightarrow \infty$.

$$\|\mathbf{Y}\|_\infty = \max_{i \in \{1, \dots, n\}} |Y_i|. \quad (\text{B.2})$$

É interessante notar, que da mesma forma que as normas L_1 e L_2 dão origem a métodos de estimação de parâmetros de modelos de regressão, a norma L_∞ dá origem ao método de estimação via mínimo erro absoluto máximo (ou estimador minimax) e à regressão L_∞ (ou regressão de Chebyshev). Existem outros aspectos intrigantes sobre a regressão L_∞ , como o fato dela ser o caso particular da regressão quantílica quando o quantil de interesse é $\theta = 1$, examine [Koenker \(2005\)](#) caso deseje saber mais sobre a regressão quantílica. Para mais propriedades, detalhes e resultados sobre a regressão L_∞ consulte a coleção de artigos de [Harter \(1974a,b, 1975a,b,c,d, 1976\)](#).

Assim, dependendo da norma escolhida como medida de distância, a reta ajustada e as propriedades inferenciais podem mudar. Mais detalhes sobre normas de vetores e matrizes podem ser vistos em [Harville \(1997\)](#), [Searle \(2006\)](#) ou [Seber \(2008\)](#).

Apêndice C

Tabelas dos Estudos de Simulação do Capítulo 2

Para manter a fluidez do texto no Capítulo 2 colocamos as tabelas com os resultados detalhados das simulações neste apêndice. As duas seções deste apêndice correspondem aos dois estudos de simulação do Capítulo 2.

C.1 Tabelas do Estudo de simulação das estatísticas LF_1 e LF_2 em configurações com efeitos de grupo, polinomial e exponencial

Configuração 1 e $\lambda = 1$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média $LF_1 (\hat{\lambda}_{ROS})$	33,026	29,606	40,789	78,894	7,815
média $LF_1 (\hat{\lambda}_{MLE})$	69,071	45,611	75,924	77,623	7,815
média LF_2	206,860	78,327	291,433	319,840	7,815
taxa de rejeição $LF_1 (\hat{\lambda}_{ROS})$	1,000	1,000	1,000	1,000	7,815
taxa de rejeição $LF_1 (\hat{\lambda}_{MLE})$	1,000	1,000	1,000	1,000	7,815
taxa de rejeição LF_2	1,000	1,000	1,000	1,000	7,815
Configuração 1 e $\lambda = 3$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média $LF_1 (\hat{\lambda}_{ROS})$	21,775	10,139	28,150	34,494	7,815
média $LF_1 (\hat{\lambda}_{MLE})$	32,201	12,501	40,898	45,628	7,815
média LF_2	44,394	13,392	64,684	78,379	7,815
taxa de rejeição $LF_1 (\hat{\lambda}_{ROS})$	0,997	0,665	1,000	1,000	7,815
taxa de rejeição $LF_1 (\hat{\lambda}_{MLE})$	0,999	0,762	1,000	1,000	7,815
taxa de rejeição LF_2	0,999	0,754	1,000	1,000	7,815
Configuração 1 e $\lambda = 5$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média $LF_1 (\hat{\lambda}_{ROS})$	13,801	5,042	17,839	20,350	7,815
média $LF_1 (\hat{\lambda}_{MLE})$	18,029	5,773	23,519	26,961	7,815
média LF_2	20,572	5,741	28,693	34,598	7,815
taxa de rejeição $LF_1 (\hat{\lambda}_{ROS})$	0,889	0,224	0,974	0,987	7,815
taxa de rejeição $LF_1 (\hat{\lambda}_{MLE})$	0,949	0,284	0,992	0,998	7,815
taxa de rejeição LF_2	0,944	0,271	0,991	0,998	7,815
Configuração 1 e $\lambda = 10$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média $LF_1 (\hat{\lambda}_{ROS})$	7,042	2,255	8,122	8,624	7,815
média $LF_1 (\hat{\lambda}_{MLE})$	7,949	2,381	9,468	10,317	7,815
média LF_2	8,070	2,265	9,784	10,772	7,815
taxa de rejeição $LF_1 (\hat{\lambda}_{ROS})$	0,367	0,035	0,483	0,537	7,815
taxa de rejeição $LF_1 (\hat{\lambda}_{MLE})$	0,454	0,037	0,583	0,636	7,815
taxa de rejeição LF_2	0,444	0,032	0,566	0,620	7,815
Configuração 1 e $\lambda = 20$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média $LF_1 (\hat{\lambda}_{ROS})$	4,354	1,349	4,041	3,563	7,815
média $LF_1 (\hat{\lambda}_{MLE})$	4,663	1,421	4,365	3,915	7,815
média LF_2	4,539	1,331	4,253	3,804	7,815
taxa de rejeição $LF_1 (\hat{\lambda}_{ROS})$	0,131	0,009	0,114	0,105	7,815
taxa de rejeição $LF_1 (\hat{\lambda}_{MLE})$	0,169	0,016	0,150	0,138	7,815
taxa de rejeição LF_2	0,157	0,014	0,146	0,129	7,815

Tabela C.1: Simulação com $\lambda = 1, 3, 5, 10, 20$ considerando a configuração 1.

Configuração 2 e $\lambda = 1$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	69,002	53,975	109,691	155,144	7,815
média LF_1 ($\hat{\lambda}_{MLE}$)	136,983	90,884	151,050	154,722	7,815
média LF_2	418,295	161,163	593,135	656,890	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	1,000	1,000	1,000	1,000	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	1,000	1,000	1,000	1,000	7,815
taxa de rejeição LF_2	1,000	1,000	1,000	1,000	7,815
Configuração 2 e $\lambda = 3$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	41,294	18,158	50,362	60,515	7,815
média LF_1 ($\hat{\lambda}_{MLE}$)	62,878	23,207	81,052	89,943	7,815
média LF_2	88,653	25,443	131,926	158,051	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	1,000	0,953	1,000	1,000	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	1,000	0,974	1,000	1,000	7,815
taxa de rejeição LF_2	1,000	0,974	1,000	1,000	7,815
Configuração 2 e $\lambda = 5$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	25,012	8,925	31,723	36,558	7,815
média LF_1 ($\hat{\lambda}_{MLE}$)	32,996	10,340	45,239	52,487	7,815
média LF_2	38,253	10,581	56,559	68,900	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	1,000	0,531	1,000	1,000	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	1,000	0,607	1,000	1,000	7,815
taxa de rejeição LF_2	1,000	0,600	1,000	1,000	7,815
Configuração 2 e $\lambda = 10$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	10,717	3,459	14,014	15,753	7,815
média LF_1 ($\hat{\lambda}_{MLE}$)	12,488	3,743	16,881	19,651	7,815
média LF_2	12,891	3,696	17,840	21,111	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	0,672	0,106	0,842	0,911	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	0,760	0,122	0,905	0,947	7,815
taxa de rejeição LF_2	0,753	0,118	0,902	0,945	7,815
Configuração 2 e $\lambda = 20$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	5,208	1,691	5,873	6,015	7,815
média LF_1 ($\hat{\lambda}_{MLE}$)	5,646	1,749	6,395	6,769	7,815
média LF_2	5,605	1,698	6,386	6,772	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	0,196	0,019	0,288	0,286	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	0,232	0,023	0,322	0,364	7,815
taxa de rejeição LF_2	0,224	0,022	0,310	0,361	7,815

Tabela C.2: Simulação com $\lambda = 1, 3, 5, 10, 20$ considerando a configuração 2.

Configuração 3 e $\lambda = 1$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média $LF_1 (\hat{\lambda}_{ROS})$	67,913	68,295	186,693	213,194	7,815
média $LF_1 (\hat{\lambda}_{MLE})$	137,554	84,227	144,446	147,603	7,815
média LF_2	424,378	140,398	500,318	541,425	7,815
taxa de rejeição $LF_1 (\hat{\lambda}_{ROS})$	1,000	1,000	1,000	1,000	7,815
taxa de rejeição $LF_1 (\hat{\lambda}_{MLE})$	1,000	1,000	1,000	1,000	7,815
taxa de rejeição LF_2	1,000	1,000	1,000	1,000	7,815
Configuração 3 e $\lambda = 3$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média $LF_1 (\hat{\lambda}_{ROS})$	40,609	17,984	59,264	72,896	7,815
média $LF_1 (\hat{\lambda}_{MLE})$	63,354	21,943	74,728	82,467	7,815
média LF_2	89,674	23,878	115,231	135,235	7,815
taxa de rejeição $LF_1 (\hat{\lambda}_{ROS})$	1,000	0,952	1,000	1,000	7,815
taxa de rejeição $LF_1 (\hat{\lambda}_{MLE})$	1,000	0,972	1,000	1,000	7,815
taxa de rejeição LF_2	1,000	0,971	1,000	1,000	7,815
Configuração 3 e $\lambda = 5$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média $LF_1 (\hat{\lambda}_{ROS})$	24,575	8,386	32,971	38,733	7,815
média $LF_1 (\hat{\lambda}_{MLE})$	33,442	9,481	42,425	48,454	7,815
média LF_2	38,876	9,638	52,151	61,845	7,815
taxa de rejeição $LF_1 (\hat{\lambda}_{ROS})$	0,999	0,514	1,000	1,000	7,815
taxa de rejeição $LF_1 (\hat{\lambda}_{MLE})$	1,000	0,570	1,000	1,000	7,815
taxa de rejeição LF_2	1,000	0,567	1,000	1,000	7,815
Configuração 3 e $\lambda = 10$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média $LF_1 (\hat{\lambda}_{ROS})$	10,944	3,310	13,495	15,036	7,815
média $LF_1 (\hat{\lambda}_{MLE})$	12,974	3,553	15,849	17,879	7,815
média LF_2	13,421	3,505	16,650	19,037	7,815
taxa de rejeição $LF_1 (\hat{\lambda}_{ROS})$	0,711	0,100	0,818	0,869	7,815
taxa de rejeição $LF_1 (\hat{\lambda}_{MLE})$	0,768	0,109	0,876	0,908	7,815
taxa de rejeição LF_2	0,764	0,102	0,873	0,906	7,815
Configuração 3 e $\lambda = 20$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média $LF_1 (\hat{\lambda}_{ROS})$	5,683	1,698	5,841	5,389	7,815
média $LF_1 (\hat{\lambda}_{MLE})$	6,240	1,777	6,310	5,915	7,815
média LF_2	6,214	1,727	6,296	5,895	7,815
taxa de rejeição $LF_1 (\hat{\lambda}_{ROS})$	0,250	0,019	0,257	0,243	7,815
taxa de rejeição $LF_1 (\hat{\lambda}_{MLE})$	0,303	0,029	0,319	0,293	7,815
taxa de rejeição LF_2	0,296	0,028	0,314	0,290	7,815

Tabela C.3: Simulação com $\lambda = 1, 3, 5, 10, 20$ considerando a configuração 3.

Configuração 4 e $\lambda = 1$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	98,941	44,944	115,919	123,776	7,815
média LF_1 ($\hat{\lambda}_{MLE}$)	138,588	48,845	109,114	108,297	7,815
média LF_2	434,349	62,245	230,699	227,066	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	1,000	1,000	1,000	1,000	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	1,000	1,000	1,000	1,000	7,815
taxa de rejeição LF_2	1,000	1,000	1,000	1,000	7,815
Configuração 4 e $\lambda = 3$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	43,015	9,911	38,550	43,179	7,815
média LF_1 ($\hat{\lambda}_{MLE}$)	65,101	10,777	41,941	43,632	7,815
média LF_2	93,282	11,005	51,222	53,699	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	1,000	0,617	0,999	1,000	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	1,000	0,686	0,999	1,000	7,815
taxa de rejeição LF_2	1,000	0,683	0,999	1,000	7,815
Configuração 4 e $\lambda = 5$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	25,981	4,827	20,656	22,961	7,815
média LF_1 ($\hat{\lambda}_{MLE}$)	34,935	5,178	22,503	24,050	7,815
média LF_2	40,972	5,141	24,439	26,352	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	0,998	0,193	0,974	0,973	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	1,000	0,234	0,984	0,984	7,815
taxa de rejeição LF_2	1,000	0,231	0,984	0,984	7,815
Configuração 4 e $\lambda = 10$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	11,387	2,200	7,940	8,236	7,815
média LF_1 ($\hat{\lambda}_{MLE}$)	13,415	2,309	8,545	8,723	7,815
média LF_2	13,920	2,254	8,614	8,811	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	0,722	0,040	0,455	0,459	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	0,792	0,051	0,507	0,535	7,815
taxa de rejeição LF_2	0,786	0,050	0,498	0,531	7,815
Configuração 4 e $\lambda = 20$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	6,187	1,417	4,020	3,364	7,815
média LF_1 ($\hat{\lambda}_{MLE}$)	6,644	1,465	4,164	3,521	7,815
média LF_2	6,627	1,420	4,097	3,458	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	0,295	0,014	0,119	0,096	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	0,331	0,016	0,137	0,114	7,815
taxa de rejeição LF_2	0,324	0,015	0,136	0,112	7,815

Tabela C.4: Simulação com $\lambda = 1, 3, 5, 10, 20$ considerando a configuração 4.

Configuração 5 e $\lambda = 1$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	68,730	114,039	130,774	412,646	15,507
média LF_1 ($\hat{\lambda}_{MLE}$)	141,488	168,302	192,098	199,657	15,507
média LF_2	441,880	966,984	4419,956	105766,869	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	1,000	1,000	1,000	1,000	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	1,000	1,000	1,000	1,000	15,507
taxa de rejeição LF_2	1,000	1,000	1,000	1,000	15,507
Configuração 5 e $\lambda = 3$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	42,868	74,048	129,167	415,125	15,507
média LF_1 ($\hat{\lambda}_{MLE}$)	67,280	114,388	176,764	198,968	15,507
média LF_2	92,916	244,280	1384,855	35051,755	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	1,000	1,000	1,000	1,000	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	1,000	1,000	1,000	1,000	15,507
taxa de rejeição LF_2	1,000	1,000	1,000	1,000	15,507
Configuração 5 e $\lambda = 5$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	27,796	53,147	105,949	418,383	15,507
média LF_1 ($\hat{\lambda}_{MLE}$)	37,619	77,461	161,686	198,262	15,507
média LF_2	42,482	115,830	768,561	20742,318	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	0,967	1,000	1,000	1,000	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	0,996	1,000	1,000	1,000	15,507
taxa de rejeição LF_2	0,996	1,000	1,000	1,000	15,507
Configuração 5 e $\lambda = 10$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	14,498	25,919	83,364	428,891	15,507
média LF_1 ($\hat{\lambda}_{MLE}$)	17,108	33,852	128,045	196,567	15,507
média LF_2	17,096	37,375	325,181	10414,338	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	0,397	0,935	1,000	1,000	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	0,568	0,986	1,000	1,000	15,507
taxa de rejeição LF_2	0,542	0,985	1,000	1,000	15,507
Configuração 5 e $\lambda = 20$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	9,479	9,843	55,088	421,086	15,507
média LF_1 ($\hat{\lambda}_{MLE}$)	10,209	11,396	79,365	193,107	15,507
média LF_2	9,815	11,077	120,559	5094,104	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	0,107	0,142	1,000	1,000	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	0,140	0,225	1,000	1,000	15,507
taxa de rejeição LF_2	0,132	0,207	1,000	1,000	15,507

Tabela C.5: Simulação com $\lambda = 1, 3, 5, 10, 20$ considerando a configuração 5.

Configuração 6 e $\lambda = 1$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	153,073	236,645	986,514	30244,479	15,507
média LF_1 ($\hat{\lambda}_{MLE}$)	266,964	334,861	384,383	399,470	15,507
média LF_2	767,770	1964,554	9401,073	288076,094	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	1,000	1,000	1,000	1,000	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	1,000	1,000	1,000	1,000	15,507
taxa de rejeição LF_2	1,000	1,000	1,000	1,000	15,507
Configuração 6 e $\lambda = 3$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	76,061	121,979	321,885	10130,128	15,507
média LF_1 ($\hat{\lambda}_{MLE}$)	116,876	224,885	354,035	398,415	15,507
média LF_2	158,307	491,680	2941,917	95998,570	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	1,000	1,000	1,000	1,000	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	1,000	1,000	1,000	1,000	15,507
taxa de rejeição LF_2	1,000	1,000	1,000	1,000	15,507
Configuração 6 e $\lambda = 5$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	44,496	86,728	222,793	6112,192	15,507
média LF_1 ($\hat{\lambda}_{MLE}$)	62,958	149,901	325,068	397,360	15,507
média LF_2	71,657	229,698	1658,186	57458,749	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	1,000	1,000	1,000	1,000	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	1,000	1,000	1,000	1,000	15,507
taxa de rejeição LF_2	1,000	1,000	1,000	1,000	15,507
Configuração 6 e $\lambda = 10$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	20,453	43,365	150,646	3049,934	15,507
média LF_1 ($\hat{\lambda}_{MLE}$)	25,551	63,337	259,270	394,737	15,507
média LF_2	26,149	72,181	705,112	28634,923	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	0,751	1,000	1,000	1,000	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	0,885	1,000	1,000	1,000	15,507
taxa de rejeição LF_2	0,878	1,000	1,000	1,000	15,507
Configuração 6 e $\lambda = 20$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	11,627	16,329	92,910	1531,975	15,507
média LF_1 ($\hat{\lambda}_{MLE}$)	13,395	20,813	161,915	389,491	15,507
média LF_2	13,255	21,065	260,594	14156,797	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	0,191	0,520	1,000	1,000	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	0,313	0,710	1,000	1,000	15,507
taxa de rejeição LF_2	0,306	0,706	1,000	1,000	15,507

Tabela C.6: Simulação com $\lambda = 1, 3, 5, 10, 20$ considerando a configuração 6.

Configuração 7 e $\lambda = 1$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	143,662	261,293	1074,737	25572,167	15,507
média LF_1 ($\hat{\lambda}_{MLE}$)	280,144	336,191	384,009	399,305	15,507
média LF_2	894,271	2014,298	9170,376	219313,269	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	1,000	1,000	1,000	1,000	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	1,000	1,000	1,000	1,000	15,507
taxa de rejeição LF_2	1,000	1,000	1,000	1,000	15,507
Configuração 7 e $\lambda = 3$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	75,902	135,003	347,926	8588,822	15,507
média LF_1 ($\hat{\lambda}_{MLE}$)	128,394	227,169	352,943	397,909	15,507
média LF_2	181,388	503,768	2865,475	72679,931	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	1,000	1,000	1,000	1,000	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	1,000	1,000	1,000	1,000	15,507
taxa de rejeição LF_2	1,000	1,000	1,000	1,000	15,507
Configuração 7 e $\lambda = 5$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	46,814	96,192	227,079	5125,238	15,507
média LF_1 ($\hat{\lambda}_{MLE}$)	69,105	152,673	322,468	396,519	15,507
média LF_2	80,120	236,599	1590,060	43506,241	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	1,000	1,000	1,000	1,000	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	1,000	1,000	1,000	1,000	15,507
taxa de rejeição LF_2	1,000	1,000	1,000	1,000	15,507
Configuração 7 e $\lambda = 10$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	21,560	46,590	156,522	2499,095	15,507
média LF_1 ($\hat{\lambda}_{MLE}$)	27,629	66,521	254,494	393,030	15,507
média LF_2	28,454	76,563	669,492	21532,356	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	0,807	1,000	1,000	1,000	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	0,911	1,000	1,000	1,000	15,507
taxa de rejeição LF_2	0,908	1,000	1,000	1,000	15,507
Configuração 7 e $\lambda = 20$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	11,256	17,654	98,335	1252,862	15,507
média LF_1 ($\hat{\lambda}_{MLE}$)	13,033	22,271	157,984	386,115	15,507
média LF_2	12,889	22,619	250,320	10622,418	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	0,187	0,591	1,000	1,000	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	0,273	0,789	1,000	1,000	15,507
taxa de rejeição LF_2	0,262	0,781	1,000	1,000	15,507

Tabela C.7: Simulação com $\lambda = 1, 3, 5, 10, 20$ considerando a configuração 7.

Configuração 8 e $\lambda = 1$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	234,990	450,951	1965,897	46525,204	15,507
média LF_1 ($\hat{\lambda}_{MLE}$)	418,412	503,407	576,042	598,951	15,507
média LF_2	1342,766	3034,436	13990,899	332051,221	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	1,000	1,000	1,000	1,000	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	1,000	1,000	1,000	1,000	15,507
taxa de rejeição LF_2	1,000	1,000	1,000	1,000	15,507
Configuração 8 e $\lambda = 3$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	114,810	208,444	608,636	15398,435	15,507
média LF_1 ($\hat{\lambda}_{MLE}$)	191,244	340,477	528,912	596,855	15,507
média LF_2	273,195	764,661	4332,065	110453,829	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	1,000	1,000	1,000	1,000	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	1,000	1,000	1,000	1,000	15,507
taxa de rejeição LF_2	1,000	1,000	1,000	1,000	15,507
Configuração 8 e $\lambda = 5$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	68,028	145,102	382,394	9184,016	15,507
média LF_1 ($\hat{\lambda}_{MLE}$)	101,032	230,362	483,844	594,765	15,507
média LF_2	118,203	363,565	2425,438	66110,204	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	1,000	1,000	1,000	1,000	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	1,000	1,000	1,000	1,000	15,507
taxa de rejeição LF_2	1,000	1,000	1,000	1,000	15,507
Configuração 8 e $\lambda = 10$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	28,756	70,193	239,907	4557,033	15,507
média LF_1 ($\hat{\lambda}_{MLE}$)	36,521	99,085	381,620	589,523	15,507
média LF_2	37,809	115,432	1018,363	32743,602	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	0,963	1,000	1,000	1,000	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	0,987	1,000	1,000	1,000	15,507
taxa de rejeição LF_2	0,987	1,000	1,000	1,000	15,507
Configuração 8 e $\lambda = 20$	Grupo	Quadrático	Cúbico	Exponencial	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	14,003	25,909	149,284	2285,009	15,507
média LF_1 ($\hat{\lambda}_{MLE}$)	16,013	32,319	235,289	579,100	15,507
média LF_2	15,984	33,227	376,339	16124,999	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	0,374	0,897	1,000	1,000	15,507
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	0,496	0,953	1,000	1,000	15,507
taxa de rejeição LF_2	0,486	0,953	1,000	1,000	15,507

Tabela C.8: Simulação com $\lambda = 1, 3, 5, 10, 20$ considerando a configuração 8.

C.2 Tabelas do Estudo de simulação das estatísticas LF_1 e LF_2 em configurações com efeito senoidal

	Linear ($\beta_2 = 0$)	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	1,172	7,815
média LF_1 ($\hat{\lambda}_{MLE}$)	1,460	7,815
média LF_2	1,271	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	0,007	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	0,018	7,815
taxa de rejeição LF_2	0,016	7,815

Tabela C.9: $\beta_2 = 0$ com 5 réplicas para cada valor da variável explicativa.

	Modelo Senoide	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	2,857	7,815
média LF_1 ($\hat{\lambda}_{MLE}$)	3,775	7,815
média LF_2	3,495	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	0,038	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	0,118	7,815
taxa de rejeição LF_2	0,099	7,815

Tabela C.10: $\beta_2 = 0,8$ com 5 réplicas para cada valor da variável explicativa.

	Modelo Senoide	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	6,145	7,815
média LF_1 ($\hat{\lambda}_{MLE}$)	8,679	7,815
média LF_2	9,076	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	0,293	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	0,550	7,815
taxa de rejeição LF_2	0,525	7,815

Tabela C.11: $\beta_2 = 1,6$ com 5 réplicas para cada valor da variável explicativa.

	Modelo Senoide	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	9,716	7,815
média LF_1 ($\hat{\lambda}_{MLE}$)	14,368	7,815
média LF_2	17,397	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	0,673	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	0,905	7,815
taxa de rejeição LF_2	0,893	7,815

Tabela C.12: $\beta_2 = 2,4$ com 5 réplicas para cada valor da variável explicativa.

	Modelo Senoide	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	12,866	7,815
média LF_1 ($\hat{\lambda}_{MLE}$)	19,120	7,815
média LF_2	26,649	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	0,891	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	0,980	7,815
taxa de rejeição LF_2	0,976	7,815

Tabela C.13: $\beta_2 = 3,2$ com 5 réplicas para cada valor da variável explicativa.

	Linear ($\beta_2 = 0$)	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	1,151	7,815
média LF_1 ($\hat{\lambda}_{MLE}$)	1,205	7,815
média LF_2	1,120	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	0,005	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	0,004	7,815
taxa de rejeição LF_2	0,004	7,815

Tabela C.14: $\beta_2 = 0$ com 10 réplicas para cada valor da variável explicativa.

	Modelo Senoide	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	5,356	7,815
média LF_1 ($\hat{\lambda}_{MLE}$)	6,076	7,815
média LF_2	6,047	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	0,226	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	0,299	7,815
taxa de rejeição LF_2	0,294	7,815

Tabela C.15: $\beta_2 = 0,8$ com 10 réplicas para cada valor da variável explicativa.

	Modelo Senoide	Crítico
média LF ₁ ($\hat{\lambda}_{ROS}$)	13,465	7,815
média LF ₁ ($\hat{\lambda}_{MLE}$)	16,463	7,815
média LF ₂	18,472	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{ROS}$)	0,834	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{MLE}$)	0,905	7,815
taxa de rejeição LF ₂	0,897	7,815

Tabela C.16: $\beta_2 = 1,6$ com 10 réplicas para cada valor da variável explicativa.

	Modelo Senoide	Crítico
média LF ₁ ($\hat{\lambda}_{ROS}$)	21,813	7,815
média LF ₁ ($\hat{\lambda}_{MLE}$)	27,154	7,815
média LF ₂	34,640	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{ROS}$)	0,999	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{MLE}$)	1,000	7,815
taxa de rejeição LF ₂	1,000	7,815

Tabela C.17: $\beta_2 = 2,4$ com 10 réplicas para cada valor da variável explicativa.

	Modelo Senoide	Crítico
média LF ₁ ($\hat{\lambda}_{ROS}$)	30,343	7,815
média LF ₁ ($\hat{\lambda}_{MLE}$)	36,615	7,815
média LF ₂	53,934	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{ROS}$)	1,000	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{MLE}$)	0,999	7,815
taxa de rejeição LF ₂	0,999	7,815

Tabela C.18: $\beta_2 = 3,2$ com 10 réplicas para cada valor da variável explicativa.

	Linear ($\beta_2 = 0$)	Crítico
média LF ₁ ($\hat{\lambda}_{ROS}$)	1,078	7,815
média LF ₁ ($\hat{\lambda}_{MLE}$)	1,243	7,815
média LF ₂	1,188	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{ROS}$)	0,006	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{MLE}$)	0,011	7,815
taxa de rejeição LF ₂	0,011	7,815

Tabela C.19: $\beta_2 = 0$ com 15 réplicas para cada valor da variável explicativa.

	Modelo Senoide	Crítico
média LF ₁ ($\hat{\lambda}_{ROS}$)	6,681	7,815
média LF ₁ ($\hat{\lambda}_{MLE}$)	8,439	7,815
média LF ₂	8,570	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{ROS}$)	0,349	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{MLE}$)	0,489	7,815
taxa de rejeição LF ₂	0,478	7,815

Tabela C.20: $\beta_2 = 0,8$ com 15 réplicas para cada valor da variável explicativa.

	Modelo Senoide	Crítico
média LF ₁ ($\hat{\lambda}_{ROS}$)	17,220	7,815
média LF ₁ ($\hat{\lambda}_{MLE}$)	23,919	7,815
média LF ₂	27,248	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{ROS}$)	0,957	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{MLE}$)	0,982	7,815
taxa de rejeição LF ₂	0,981	7,815

Tabela C.21: $\beta_2 = 1,6$ com 15 réplicas para cada valor da variável explicativa.

	Modelo Senoide	Crítico
média LF ₁ ($\hat{\lambda}_{ROS}$)	28,450	7,815
média LF ₁ ($\hat{\lambda}_{MLE}$)	40,267	7,815
média LF ₂	52,678	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{ROS}$)	1,000	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{MLE}$)	1,000	7,815
taxa de rejeição LF ₂	1,000	7,815

Tabela C.22: $\beta_2 = 2,4$ com 15 réplicas para cada valor da variável explicativa.

	Modelo Senoide	Crítico
média LF ₁ ($\hat{\lambda}_{ROS}$)	38,813	7,815
média LF ₁ ($\hat{\lambda}_{MLE}$)	54,304	7,815
média LF ₂	81,214	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{ROS}$)	1,000	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{MLE}$)	1,000	7,815
taxa de rejeição LF ₂	1,000	7,815

Tabela C.23: $\beta_2 = 3,2$ com 15 réplicas para cada valor da variável explicativa.

	Linear ($\beta_2 = 0$)	Crítico
média LF ₁ ($\hat{\lambda}_{ROS}$)	1,127	7,815
média LF ₁ ($\hat{\lambda}_{MLE}$)	1,173	7,815
média LF ₂	1,132	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{ROS}$)	0,005	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{MLE}$)	0,004	7,815
taxa de rejeição LF ₂	0,004	7,815

Tabela C.24: $\beta_2 = 0$ com 20 réplicas para cada valor da variável explicativa.

	Modelo Senoide	Crítico
média LF ₁ ($\hat{\lambda}_{ROS}$)	9,713	7,815
média LF ₁ ($\hat{\lambda}_{MLE}$)	11,145	7,815
média LF ₂	11,427	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{ROS}$)	0,602	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{MLE}$)	0,668	7,815
taxa de rejeição LF ₂	0,657	7,815

Tabela C.25: $\beta_2 = 0,8$ com 20 réplicas para cada valor da variável explicativa.

	Modelo Senoide	Crítico
média LF ₁ ($\hat{\lambda}_{ROS}$)	25,848	7,815
média LF ₁ ($\hat{\lambda}_{MLE}$)	31,528	7,815
média LF ₂	36,213	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{ROS}$)	0,991	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{MLE}$)	0,996	7,815
taxa de rejeição LF ₂	0,996	7,815

Tabela C.26: $\beta_2 = 1,6$ com 20 réplicas para cada valor da variável explicativa.

	Modelo Senoide	Crítico
média LF ₁ ($\hat{\lambda}_{ROS}$)	41,364	7,815
média LF ₁ ($\hat{\lambda}_{MLE}$)	52,756	7,815
média LF ₂	69,356	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{ROS}$)	1,000	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{MLE}$)	1,000	7,815
taxa de rejeição LF ₂	1,000	7,815

Tabela C.27: $\beta_2 = 2,4$ com 20 réplicas para cada valor da variável explicativa.

	Modelo Senoide	Crítico
média LF ₁ ($\hat{\lambda}_{ROS}$)	57,168	7,815
média LF ₁ ($\hat{\lambda}_{MLE}$)	72,260	7,815
média LF ₂	109,309	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{ROS}$)	1,000	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{MLE}$)	1,000	7,815
taxa de rejeição LF ₂	1,000	7,815

Tabela C.28: $\beta_2 = 3,2$ com 20 réplicas para cada valor da variável explicativa.

	Linear ($\beta_2 = 0$)	Crítico
média LF ₁ ($\hat{\lambda}_{ROS}$)	1,048	7,815
média LF ₁ ($\hat{\lambda}_{MLE}$)	1,152	7,815
média LF ₂	1,127	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{ROS}$)	0,007	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{MLE}$)	0,010	7,815
taxa de rejeição LF ₂	0,010	7,815

Tabela C.29: $\beta_2 = 0$ com 30 réplicas para cada valor da variável explicativa.

	Modelo Senoide	Crítico
média LF ₁ ($\hat{\lambda}_{ROS}$)	13,041	7,815
média LF ₁ ($\hat{\lambda}_{MLE}$)	16,154	7,815
média LF ₂	16,726	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{ROS}$)	0,802	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{MLE}$)	0,871	7,815
taxa de rejeição LF ₂	0,868	7,815

Tabela C.30: $\beta_2 = 0,8$ com 30 réplicas para cada valor da variável explicativa.

	Modelo Senoide	Crítico
média LF ₁ ($\hat{\lambda}_{ROS}$)	35,052	7,815
média LF ₁ ($\hat{\lambda}_{MLE}$)	47,014	7,815
média LF ₂	54,536	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{ROS}$)	1,000	7,815
taxa de rejeição LF ₁ ($\hat{\lambda}_{MLE}$)	1,000	7,815
taxa de rejeição LF ₂	1,000	7,815

Tabela C.31: $\beta_2 = 1,6$ com 30 réplicas para cada valor da variável explicativa.

	Modelo Senoide	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	58,471	7,815
média LF_1 ($\hat{\lambda}_{MLE}$)	80,185	7,815
média LF_2	107,007	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	1,000	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	1,000	7,815
taxa de rejeição LF_2	1,000	7,815

Tabela C.32: $\beta_2 = 2,4$ com 30 réplicas para cada valor da variável explicativa.

	Modelo Senoide	Crítico
média LF_1 ($\hat{\lambda}_{ROS}$)	79,764	7,815
média LF_1 ($\hat{\lambda}_{MLE}$)	108,611	7,815
média LF_2	166,340	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{ROS}$)	1,000	7,815
taxa de rejeição LF_1 ($\hat{\lambda}_{MLE}$)	1,000	7,815
taxa de rejeição LF_2	1,000	7,815

Tabela C.33: $\beta_2 = 3,2$ com 30 réplicas para cada valor da variável explicativa.

Apêndice D

Testes para Verificar a Suposição da Distribuição dos Erros

Este apêndice é essencialmente uma cópia da Seção 3.1 da dissertação de [Rodrigues \(2019\)](#). É fundamental para a aplicação da regressão L_1 ter técnicas adequadas para verificar se a hipótese dos erros seguirem distribuição Laplace é razoável, por isso introduzimos este apêndice.

[Puig e Stephens \(2000\)](#) apresentaram cinco testes para verificar se a distribuição dos dados é Laplace, todos baseados na FDE, que é definida por

$$F_n(k) = \frac{\#\{X_i \leq k\}}{n},$$

em que $\#\{X_i \leq k\}$ é o número de elementos amostrais menores ou iguais a k .

Os testes foram baseados nas estatísticas de qualidade do ajuste de: Cramér-von Mises, Watson, Anderson-Darling, Kolmogorov-Smirnov e Kuiper. Segundo os autores supracitados, o teste mais poderoso, nos restringindo às cinco estatísticas citadas, para testar se os dados têm distribuição Laplace contra a hipótese dos dados terem distribuição simétrica diversa é o baseado na estatística de Watson. Isto ocorre devido ao fato da estatística de Watson detectar melhor divergências com relação à variância da distribuição hipotética (é natural que isso ocorra, pois é esperado que o parâmetro de posição seja o mesmo para todas as distribuições simétricas consideradas, e então a comparação entre as distribuições recairá naturalmente sobre a variância).

O procedimento para realizar os cinco testes é apresentado em quatro passos a seguir:

- 1*) Calcular as estimativas pontuais de máxima verossimilhança dos parâmetros μ e λ da distribuição Laplace a partir da amostra X_1, \dots, X_n .
- 2*) Considerar as estatísticas de ordem $X_{(1)}, \dots, X_{(n)}$, em que $X_{(1)}$ é o mínimo e $X_{(n)}$ é o máximo e definir

$$z_{(i)} = F(x_{(i)}; \hat{\mu}, \hat{\lambda}),$$

em que a notação $F(x_{(i)}; \hat{\mu}, \hat{\lambda})$ simboliza a função de distribuição de uma variável aleatória com distribuição Laplace $(\hat{\mu}, \hat{\lambda})$ calculada no ponto $x_{(i)}$.

- 3*) Escolher um teste (estatística do teste) dentre: Cramér-von Mises (W^2), Watson (U^2), Anderson-Darling (A^2), Kolmogorov-Smirnov (D) e Kuiper (V) e calcular a respectiva estatística do teste:

$$W^2 = \sum_{i=1}^n \left[\left[z_{(i)} - (2i-1)/(2n) \right]^2 \right] + \frac{1}{12n},$$

$$U^2 = W^2 - n \left(\bar{z} - \frac{1}{2} \right)^2,$$

em que \bar{z} é a média dos $z_{(i)}$,

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n (2i-1) \left[\ln(z_{(i)}) + \ln(1 - z_{(n+1-i)}) \right],$$

$$D = \max(D^+, D^-) \text{ e}$$

$$V = D^+ + D^-,$$

em que

$$D^+ = \max_i (i/n - z_{(i)}) \text{ e } D^- = \max_i (z_{(i)} - (i-1)/n).$$

- 4*) Utilizando as Tabelas D.1 a D.5, fixando o nível de significância α , comparar a estatística calculada no passo anterior com o valor crítico fornecido pela respectiva tabela do teste escolhido. Caso o valor da estatística seja maior que o valor crítico, rejeita-se a hipótese nula, isto é, descarta-se a hipótese de que os dados são provenientes da distribuição Laplace.

Por exemplo, se escolhermos o teste baseado na estatística de Watson, U^2 , fixarmos o nível de significância em 5%, $\alpha = 0,05$, com o tamanho amostral igual a 20 e $U^2 = 0,085$, hipotéticos, obtemos $U^2 = 0,085 > 0,080 = U^2_*$ a partir da Tabela D.1; então rejeitamos a hipótese nula.

Se o tamanho da amostra n não estivesse na tabela, bastaria interpolarmos para obter o ponto de corte U^2_* . Recomenda-se que em caso de interpolação, sejam utilizados os tamanhos amostrais tabelados com a mesma paridade do seu n , isto é, se n é par, interpole usando os tamanhos amostrais pares tabelados, ignorando os elementos ímpares, caso n seja ímpar, utilize somente os elementos ímpares (Puig e Stephens, 2000). Segundo os autores, esse cuidado na interpolação decorre do fato das Tabelas de 3.1 a 3.5 terem sido obtidas via método de Monte Carlo (veja o artigo original, Metropolis e Ulam (1949), para mais detalhes) e a velocidade de convergência mudar dependendo da paridade do n , acarretando assim, diferenças para tamanhos amostrais pequenos ($n < 100$).

Tamanho Amostral n	Nível de Significância					
	0,5	0,25	0,1	0,05	0,025	0,01
10	0,038	0,051	0,066	0,078	0,090	0,105
15	0,039	0,053	0,069	0,081	0,092	0,108
20	0,039	0,052	0,068	0,080	0,092	0,108
35	0,039	0,053	0,070	0,083	0,095	0,112
50	0,039	0,053	0,070	0,083	0,094	0,111
75	0,039	0,054	0,071	0,083	0,096	0,113
100	0,039	0,054	0,071	0,083	0,096	0,113
$n > 100$	0,039	0,054	0,072	0,084	0,097	0,114

Fonte: Puig e Stephens (2000).

Tabela D.1: Pontos críticos da estatística U^2 (Watson).

Tamanho Amostral n	Nível de Significância					
	0,5	0,25	0,1	0,05	0,025	0,01
10	0,049	0,070	0,097	0,119	0,142	0,174
15	0,054	0,080	0,115	0,143	0,172	0,209
20	0,051	0,074	0,107	0,131	0,157	0,192
35	0,054	0,080	0,116	0,144	0,173	0,213
50	0,053	0,078	0,111	0,138	0,165	0,202
75	0,054	0,080	0,116	0,144	0,174	0,213
100	0,053	0,079	0,114	0,141	0,169	0,208
$n > 100$	0,054	0,080	0,116	0,144	0,173	0,214

Fonte: Puig e Stephens (2000).

Tabela D.2: Pontos críticos da estatística W^2 (Cramér-von Mises).

Tamanho Amostral n	Nível de Significância					
	0,5	0,25	0,1	0,05	0,025	0,01
10	0,366	0,513	0,714	0,869	1,023	1,234
15	0,397	0,570	0,807	0,991	1,160	1,415
20	0,381	0,541	0,760	0,930	1,103	1,336
35	0,396	0,568	0,797	0,987	1,179	1,438
50	0,390	0,554	0,783	0,961	1,137	1,373
75	0,395	0,567	0,797	0,984	1,178	1,442
100	0,394	0,561	0,792	0,972	1,156	1,408
$n > 100$	0,396	0,567	0,798	0,983	1,177	1,442

Fonte: Puig e Stephens (2000).

Tabela D.3: Pontos críticos da estatística A^2 (Anderson-Darling).

No *Software* livre *R* (R Core Team, 2023) o pacote *diagL1* (Rodrigues e Elian, 2024) contém uma função chamada *laplace.dist.test* que calcula as cinco estatísticas, portanto basta definir o nível de significância desejado, usar esse comando e comparar os valores das estatísticas retornados pelo *R* com os respectivos valores críticos apresentados nas Tabelas D.1 a D.5 para obter o resultado do teste de hipóteses.

Como temos interesse em verificar se os erros do modelo de regressão possuem

Tamanho Amostral n	Nível de Significância					
	0,5	0,25	0,1	0,05	0,025	0,01
10	0,548	0,645	0,747	0,813	0,871	0,944
15	0,589	0,699	0,812	0,884	0,949	1,026
20	0,582	0,688	0,797	0,868	0,933	1,006
35	0,604	0,716	0,830	0,906	0,973	1,058
50	0,604	0,714	0,828	0,900	0,967	1,048
75	0,613	0,724	0,844	0,920	0,987	1,074
100	0,614	0,725	0,841	0,918	0,985	1,065

Fonte: Puig e Stephens (2000).

Tabela D.4: Pontos críticos da estatística $\sqrt{n}D$ (Kolmogorov-Smirnov).

Tamanho Amostral n	Nível de Significância					
	0,5	0,25	0,1	0,05	0,025	0,01
10	0,897	1,016	1,133	1,206	1,282	1,373
15	0,921	1,045	1,170	1,251	1,322	1,411
20	0,929	1,054	1,179	1,261	1,333	1,423
35	0,952	1,080	1,210	1,295	1,374	1,467
50	0,959	1,089	1,222	1,305	1,383	1,473
75	0,970	1,101	1,235	1,319	1,395	1,489
100	0,974	1,106	1,240	1,324	1,404	1,501

Fonte: Puig e Stephens (2000).

Tabela D.5: Pontos críticos da estatística $\sqrt{n}V$ (Kuiper).

distribuição Laplace então podemos reescrever o procedimento para o nosso problema, os passos são apresentados a seguir.

- 1) Calcular os resíduos "studentizados" $\hat{\epsilon}_1^*, \dots, \hat{\epsilon}_n^*$, dados por

$$\hat{\epsilon}_i^* = \frac{\hat{\epsilon}_i}{\hat{\lambda}}, \quad i = 1, \dots, n,$$

em que $\hat{\epsilon}_1, \dots, \hat{\epsilon}_n$ são os resíduos da regressão L_1 e $\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n |\hat{\epsilon}_i|$.

- 2) Considerar as estatísticas de ordem $\hat{\epsilon}_{(1)}^*, \dots, \hat{\epsilon}_{(n)}^*$, em que $\hat{\epsilon}_{(1)}^*$ é o mínimo e $\hat{\epsilon}_{(n)}^*$ é o máximo e definir

$$z_{(i)} = F(\hat{\epsilon}_{(i)}^*; 0, 1).$$

- 3) Executar os passos 3* e 4* como apresentados anteriormente.

A obtenção da distribuição assintótica das estatísticas dos testes apresentados, sob a hipótese nula, é bastante complexa e pode ser encontrada na Seção 3 de Puig e Stephens (2000).

Apêndice E

Ferramentas Computacionais e Processamento Paralelo

Este apêndice trata de assuntos auxiliares para a aplicação da DKL ao contexto da regressão L_1 , ou seja, é um apêndice relacionado ao Capítulo 4. Embora coadjuvantes, ferramentas computacionais e conhecimento introdutório de computação paralela são importantes tanto em situações práticas quanto teóricas, pois quando se lida com métodos computacionalmente intensivos, é essencial utilizar meios para diminuição do tempo cronológico empregado. Este apêndice está dividido em duas partes, uma descrevendo aspectos básicos do processamento paralelo e a outra voltada à descrição dos pacotes computacionais utilizados.

E.1 Processamento paralelo

A ideia basilar que fundamenta o paralelismo é que, se é possível executar um procedimento computacional em um tempo T em um único processador, então será possível executar o mesmo procedimento, utilizando K processadores, em tempo T/K . Dessa forma é possível diminuir o tempo cronológico gasto de forma proporcional ao número de processadores. Contudo, essa ideia precisa de alguns refinamentos para tornar-se mais precisa, pois outras variáveis influenciam diretamente no tempo cronológico gasto. Os refinamentos são justamente fornecidos pela Lei de Amdahl ([Amdahl, 1967](#)) e pela Lei de Gustafson ([Gustafson, 1988](#)). A partir das Lei de Amdahl é possível verificar que quanto maior o percentual de código paralelizável mais eficiente será utilizar o processamento paralelo. Por outro lado, se o algoritmo tiver um baixo percentual paralelizável, usar mais processadores pode ser ineficaz, isto é, não diminuir o tempo cronológico gasto no processamento. Assim, simplesmente aumentar o número de processadores não necessariamente diminuirá o tempo gasto, pois isso dependerá do percentual do código paralelizável.

Adams *et al.* (1996) apresentam uma resenha sobre computação paralela aplicada à computação estatística. Kontoghiorghes (2005) trata de processamento paralelo por meio do software R. Rosenthal (2000) e Eddelbuettel (2021) são outras referências de computação paralela na Estatística.

Podemos classificar o processamento em dois tipos: sequencial e paralelo. Note que processamento e computação são sinônimos neste contexto. A computação sequencial, como o próprio nome sugere, é aquela feita em uma sequência pré-definida de operações computacionais. Já a computação paralela é caracterizada por operações computacionais feitas concorrentemente, isto é, são operações conduzidas simultaneamente. Outra forma de visualizar a dicotomia entre computação sequencial e paralela é observar que a computação sequencial é conduzida em um único núcleo, enquanto a computação paralela é realizada em múltiplos núcleos do processador.

Existe inclusive o conceito de processamento distribuído, que pode ser visto como um “nível mais avançado” de paralelismo. Na computação distribuída não nos restringimos aos núcleos de processamento de um único computador, em vez disso utilizamos os núcleos de processamento de mais de um computador. No R, há o pacote *future* que apresenta uma estrutura unificada que permite executar processamento paralelo e distribuído com relativa facilidade, veja mais detalhes acerca da teoria sobre o *future* em Bengtsson (2021). Como o custo computacional para os procedimentos desta tese exigiram menos de um dia computacional, optamos por não aplicar o processamento distribuído (até porque usamos apenas um único computador pessoal) neste texto. Entretanto, em estudos que imponham custo computacional em ordem de grandeza mais elevada, faz-se necessário empregar os meios fornecidos pela computação distribuída.

É comum que não existam disciplinas que abordem o conceito de computação paralela em cursos de graduação e pós-graduação na área de Estatística. Contudo, à medida que surgem modelos e técnicas estatísticas cada vez mais computacionalmente intensivas, mais clara é a necessidade do conhecimento de computação paralela. Inclusive a computação paralela é muito útil para reduzir o tempo necessário para executar simulações estatísticas. A vantagem do uso da computação paralela é a redução no tempo cronológico de execução de técnicas computacionalmente intensivas. A redução no tempo é aproximadamente proporcional ao número de núcleos de processamento utilizados, as Leis de Amdahl e Gustafson explicitam a relação entre o número de processadores e percentual do algoritmo que é paralelizável e o tempo cronológico, consulte Amdahl (1967) e Gustafson (1988) para mais detalhes. A seguir entenderemos como a Taxonomia de Flynn se aplica no contexto de paralelismo aplicado a problemas estatísticos.

Flynn (1972, 1966) fornece uma taxonomia para classificação dos processamentos paralelos. A rigor, taxonomia é o ramo da Biologia responsável por classificar seres vivos. Fora do âmbito da Biologia, “taxonomia” é um termo usado como sinônimo de sistema de classificação. A taxonomia de Flynn é útil para classificarmos os tipos de paralelismo existentes, pois para cada tipo de paralelismo há uma forma adequada de otimizar o

algoritmo. Há quatro classificações para o paralelismo em função da taxonomia de Flynn:

1. SISD (*Single Instruction Single Data*): nesse caso há um único conjunto de dados e um único código a ser aplicado aos dados, não há processamento paralelo, tudo é feito de forma sequencial;
2. SIMD (*Single Instruction Multiple Data*): nesse caso há mais de um conjunto de dados e um único código a ser aplicado aos dados, essa é a situação na qual trocar o processamento sequencial pelo paralelo é mais simples;
3. MISD (*Multiple Instruction Single Data*): nesse caso há apenas um conjunto de dados e mais de um código a ser aplicado ao conjunto de dados;
4. MIMD (*Multiple Instruction Multiple Data*): nesse caso há mais de um conjunto de dados e mais de um código a ser aplicado aos dados.

Do ponto de vista estatístico, SISD é o cenário mais natural, pois trata-se do processamento sequencial, que é o mais usual. O SIMD traz um contexto de fácil paralelização. Um exemplo de SIMD é ajustar o modelo de regressão L_1 a 10 conjuntos de dados distintos. Um exemplo de MISD é quando ajustamos vários modelos de regressão a um único conjunto de dados, almejando o melhor ajuste. O MIMD é o cenário mais complexo, pois há, simultaneamente, vários conjuntos de dados e várias instruções. Por exemplo, se ajustamos vários modelos de regressão a cinco conjuntos de dados, buscando o melhor ajuste em cada um dos conjuntos de dados, então trata-se do MIMD. Pela Taxonomia de Flynn, o problema do cálculo da DKL de um conjunto de dados se encaixa no SIMD, pois para $j = 1, \dots, N$, a j -ésima observação do conjunto de dados inicial é excluída, de forma que para efeitos práticos, trata-se de um “novo” conjunto de dados.

E.2 Ferramentas computacionais

Para o cálculo numérico foram necessárias algumas ferramentas computacionais, que são apresentadas a seguir.

O pacote *diagL1* (Rodrigues e Elian, 2024), que tem como alicerce o pacote *quantreg* (Koenker et al., 2023), traz a função *regL1()*, que permite-nos ajustar modelos de regressão L_1 de forma computacionalmente eficiente. A vantagem do pacote *diagL1* é que ele fornece diversas ferramentas de diagnóstico voltadas para regressão L_1 , incluindo testes para verificar a suposição de que os erros seguem distribuição Laplace, tudo em um único pacote.

O pacote *cubature* (Narasimhan et al., 2023) fornece um método de integração adaptativo chamado *hcubature*, baseado na biblioteca “cubature” da linguagem C escrita por

Steven G. Johnson. A função *hcubature()* é a versão h-adaptativa que particiona recursivamente o domínio de integração em subdomínios menores, aplicando a mesma regra de integração para cada um, até que a convergência seja alcançada.

A função *hcubature()* é particularmente útil, pois permite-nos realizar integração numérica de funções não-suaves, isto é, funções com muitas descontinuidades. Em especial, essa função permite calcular integrais cujo integrando não possui forma analítica, ao contrário da função *integrate()*, que fornece resultados incorretos para integrais cujos integrandos são não-suaves.

Tendo em vista que tanto o ajuste do modelo quanto o cálculo de algumas medidas de influência, como a DKL, são feitos de forma numérica, então é necessário pensar em meios de realizar o processamento dessas funções de forma mais rápida.

Pacotes como *parallel* (pacote básico do R, não é necessário instalá-lo), *doParallel* (Daniel *et al.*, 2022) e *foreach* (Microsoft e Weston, 2022) auxiliam com a tarefa de tornar códigos e processamentos sequenciais em códigos e processamentos paralelos.

O pacote *parallel* é o responsável por executar o processamento paralelo de fato, além de fornecer funções úteis como *detectCores()*, que permite detectar o número de núcleos do computador. A principal razão para usar o pacote *foreach* é que ele suporta execução paralela, isto é, pode executar essas operações repetidas em vários núcleos do computador. O pacote *doParallel* atua como uma interface entre *foreach* e o pacote *parallel*, uma das funções mais úteis do *doParallel* é *registerDoParallel()*, é usada para registrar o processo interno (*backend*) paralelo do pacote *foreach*. Em resumo, pode-se dizer que sem esses pacotes seria inviável aplicar computação paralela aos cálculos de medidas de influência, pois seria necessário muito conhecimento especializado antes de colher os frutos da computação paralela. A seguir apresentaremos alguns detalhes do *foreach*.

O pacote *foreach* (Microsoft e Weston, 2022) permite executar algoritmos em paralelo em uma arquitetura *multi-core*, ou seja, em um computador com múltiplos núcleos de processamento. Usando o *foreach*, é possível executar algoritmos via *loops* (ciclos), em outras palavras, o algoritmo pode ser executado repetidamente em uma arquitetura *multi-core* a partir de *loops* predefinidos. Também vale mencionar que o *foreach* usa métodos de paralelismo explícito.

O paralelismo pode ser explícito ou implícito. Classifica-se o paralelismo em explícito quando cabe ao programador definir as tarefas e trechos do código que serão executadas em paralelo. Classifica-se o paralelismo em implícito quando cabe ao sistema de execução ou ao compilador: detectar o paralelismo do código ou definir as tarefas aos processadores. É evidente que escrever um código em paralelismo implícito é mais fácil (menos trabalhoso) do que escrever o mesmo código em paralelismo explícito, mas os programas gerados utilizando paralelismo implícito geralmente são menos eficientes que os gerados usando paralelismo explícito, pois no caso explícito há a possibilidade de usar as especificidades do problema como meio de otimizar o código.

Além disso, o pacote *foreach* fornece uma sintaxe pragmática com blocos de código fáceis de construir e de ler. Para utilizar a principal função do pacote *foreach*, que é homônima ao próprio pacote, é primordial adaptar o algoritmo em forma de um ou vários *loops*. Para construir os blocos mencionados anteriormente é possível usar dois operadores binários, que são *%do%* e *%dopar%*. O primeiro operador, *%do%*, é usado para executar o algoritmo sequencialmente e o operador *%dopar%* é usado para executar o algoritmo em paralelo.

Assim, podemos dividir partes do algoritmo em duas categorias, paralelizado e não paralelizado. As partes classificadas como paralelizadas devem ser executadas por meio do operador *%dopar%*, desta forma o computador de fato executará o algoritmo de forma paralelizada utilizando os múltiplos núcleos de processamento. As demais partes, que são não paralelizadas, devem ser executadas usando o *%do%*.

Encerramos aqui as noções sobre paralelismo necessárias ao entendimento dos resultados deste texto. Recomendamos ao leitor interessado em se aprofundar em paralelismo a literatura especializada, como por exemplo, [Kontoghiorghes \(2005\)](#).

Referências Bibliográficas

- Adams, N. M., Kirby, S., Harris, P. e Clegg, D. (1996). A review of parallel processing for statistical computation. *Statistics and Computing*, **6**(1), 37–49. Citado na pág. [157](#)
- Afifah, R., Andriyana, Y. e Jaya, I. G. N. M. (2017). Robust geographically weighted regression with least absolute deviation method in case of poverty in java island. *AIP Conference Proceedings*, **1827**(1), 020023. doi: 10.1063/1.4979439. Citado na pág. [2](#)
- Aitchison, J. e Dunsmore, I. R. (1975). *Statistical Prediction Analysis*. Cambridge University Press. Citado na pág. [81](#)
- Amdahl, G. M. (1967). Validity of the single processor approach to achieving large scale computing capabilities. Em *Proceedings of the April 18-20, 1967, spring joint computer conference*, páginas 483–485. Citado na pág. [157](#), [158](#)
- André, C. D. S. e Elian, S. N. (2022). A comparison of the estimators of the scale parameter of the errors distribution in the L_1 regression. *Open Journal of Statistics*, **12**(2), 261–276. Citado na pág. [10](#), [11](#)
- Aoki, R., Bustamante, J. P. M. e Paula, G. A. (2022). Local influence diagnostics with forward search in regression analysis. *Statistical Papers*, **63**(5), 1477–1497. Citado na pág. [115](#)
- Appa, G. e Smith, C. (1973). On L_1 and Chebyshev estimation. *Mathematical Programming*, **5**(1), 73–87. Citado na pág. [12](#), [13](#)
- Atkinson, A. C. e Riani, M. (2000). *Robust diagnostic regression analysis*. Springer. Citado na pág. [115](#), [116](#), [117](#), [118](#)
- Atkinson, A. C., Riani, M. e Cerioli, A. (2010). The forward search: Theory and data analysis. *Journal of the Korean Statistical Society*, **39**(2), 117–134. ISSN 1226-3192. Citado na pág. [115](#)
- Bai, Z., Rao, C. R. e Yin, Y. (1990). Least absolute deviations analysis of variance. *Sankhyā: The Indian Journal of Statistics, Series A*, páginas 166–177. Citado na pág. [23](#), [24](#), [26](#), [42](#)
- Barrodale, I. e Roberts, F. D. K. (1973). An improved algorithm for discrete L_1 linear approximation. *SIAM Journal on Numerical Analysis*, **10**, 839–848. Citado na pág. [9](#), [45](#)
- Barrodale, I. e Roberts, F. D. K. (1974). Solution of an overdetermined system of equations in the L_1 norm. *Communications of the ACM*, **17**, 319–320. Citado na pág. [9](#), [45](#)
- Bassett, G. e Koenker, R. (1978). Asymptotic theory of least absolute error regression. *Journal of the American Statistical Association*, **73**, 618–622. Citado na pág. [15](#)

- Bengtsson, H. (2021). A unifying framework for parallel and distributed processing in R using futures. *The R Journal*, **13**(2), 208–227. doi: 10.32614/RJ-2021-048. URL <https://doi.org/10.32614/RJ-2021-048>. Citado na pág. 158
- Bezerra, M. I. S. (1990). Métodos para análise de regressão L_1 . Dissertação de Mestrado, IMECC/UNICAMP, Brasil. Citado na pág. 2, 14, 15
- Birkes, D. e Dodge, Y. (1993). *Alternative Methods of Regression*. New York, John Wiley & Sons. Citado na pág. 1, 2, 10, 101, 102, 103
- Bjørnstad, J. F. (1990). Predictive likelihood: A review. *Statistical Science*, **5**(2), 242–254. ISSN 08834237. Citado na pág. 82, 83
- Bloomfield, P. e Steiger, W. L. (1983). *Least Absolute Deviations: Theory, Applications, and Algorithms*. Birkhauser, Boston. Citado na pág. 1, 2, 133
- Boscovich, R. J. (1757). De litteraria expeditione per pontificiam ditioned, et synopsis amplioris operis, ac habentur plura ejus ex exemplaria etiam sensorum impressa. *Bononiensi Scientiarum et Artum Instituto Atque Academia Commentarii*, **4**, 353–396. Citado na pág. 1, 2
- Boscovich, R. J. (1760). De recentissimis graduum dimensionibus, et figura, ac magnitudine terrae inde derivanda. *Philosophiae Recentioris, a Benedicto Stay in Romano Archigynasis Publico Eloquentare Professore, vesibus traditae, Libri X, cum adnotianibus et Supplementis P. Rogerii Joseph Boscovich, S. J.*, **2**, 406–426. Citado na pág. 1
- Bustamante, J. P. M. (2020). Diagnóstico de influência local para a obtenção de dados mascarados influentes em modelos de regressão com erros nas variáveis e propriedades assintóticas do modelo de calibração ultraestrutural. Tese de Doutorado, Estatística Interinstitucional do ICMC/USP e UFSCar, Brasil. Citado na pág. 115
- Butler, R. W. (1986). Predictive likelihood inference with applications. *Journal of the Royal Statistical Society. Series B (Methodological)*, **48**(1), 1–38. Citado na pág. 83
- Charnes, A., Cooper, W. W. e Ferguson, R. O. (1955). Optimal estimation of executive compensation by linear programming. *Management Science*, **1**, 138–151. Citado na pág. 8
- Chmielewski, M. A. (1981). Elliptically symmetric distributions: A review and bibliography. *International Statistical Review*, **49**(1), 67–74. Citado na pág. 16
- Cook, R. D., Peña, D. e Weisberg, S. (1988). The likelihood displacement: a unifying principle for influence measures. *Communications in Statistics - Theory and Methods*, **17**, 623–640. Citado na pág. 91
- Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, **19**(1), 15–18. Citado na pág. 91
- Cox, R. D. e Hinkley, D. V. (1974). *Theoretical statistics*. London, Chapman and hall Ltd. Citado na pág. 14
- Cysneiros, F. J. A., Paula, G. A. e Galea, M. (2005). *Modelos Simétricos Aplicados*. São Paulo: ABE. 9a Escola de Modelos de Regressão. Citado na pág. 16, 17

- Daniel, F., Microsoft, Weston, S. e Tenenbaum, D. *doParallel: Foreach Parallel Adaptor for the 'parallel' Package*, 2022. URL <https://CRAN.R-project.org/package=doParallel>. R package version 1.0.17. Citado na pág. 160
- Delignette-Muller, M. L. e Dutang, C. (2015). *fitdistrplus: An R package for fitting distributions*. *Journal of Statistical Software*, **64**(4), 1–34. Citado na pág. 51
- Delignette-Muller, M.-L., Dutang, C. e Siberchicot, A. *fitdistrplus: Help to Fit of a Parametric Distribution to Non-Censored or Censored Data*, 2023. URL <https://CRAN.R-project.org/package=fitdistrplus>. R package version 1.1-11. Citado na pág. 50, 51, 60
- Dielman, T. E. (1984). Least absolute value estimation in regression models: An annotated bibliography. *Communications in Statistics- Theory and Methods*, **4**, 513–541. Citado na pág. 2
- Dielman, T. E. (2005). Least absolute value regression: recent contributions. *Journal of Statistical Computation and Simulation*, **75**(4), 263–286. Citado na pág. 2
- Dodge, Y. (2002). *Statistical Data Analysis Based on the L_1 -norm and Related Methods*. Birkhäuser. Citado na pág. 2
- Dodge, Y. (1997). Lad regression for detecting outliers in response and explanatory variables. *Journal of Multivariate Analysis*, **61**, 144–158. Citado na pág. 2
- Eddelbuettel, D. (2021). Parallel computing with r: A brief review. *Wiley Interdisciplinary Reviews: Computational Statistics*, **13**(2), e1515. Citado na pág. 158
- Edgeworth, F. Y. (1887). A new method of reducing observations relating to several quantities. *Philosophical Magazine (5th Series)*, **24**, 222–223. Citado na pág. 2
- Edgeworth, F. Y. (1888). On a new method of reducing observations relating to several quantities. *Philosophical Magazine (5th Series)*, **25**, 184–191. Citado na pág. 2
- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, **7**, 1–26. Citado na pág. 49
- Efron, B. e Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Monographs on statistics and applied probability, volume 57. Citado na pág. 14
- Elian, S. N., André, C. D. S. e Narula, S. C. (2000). Influence measure for the L_1 regression. *Communications in Statistics - Theory and Methods*, **29**, 837–849. Citado na pág. 91
- Ellis, S. P. e Morgenthaler, S. (1992). Leverage and breakdown in L_1 regression. *Journal of the American Statistical Association*, **87**, 143–148. Citado na pág. 117
- Farebrother, R. W. (1990). Studies in the history of probability and statistics XLII. Further details of contacts between Boscovich and Simpson in June 1760. *Biometrika*, **77**(2), 397–400. Citado na pág. 2
- Farebrother, R. W. (1997). Notes on the early history of elemental set methods. *L_1 -Statistical Procedures and Related Topics. IMS Lecture Notes-Monograph Series*, **31**, 161–170. Citado na pág. 2

- Farebrother, R. (1985). Unbiased L_1 and L_∞ estimation. *Communications in Statistics - Theory and Methods*, **14**(8), 1941–1962. Citado na pág. 2
- Fausett, D. W. e Weber, J. H. (1978). Mass spectral pattern recognition via techniques of mathematical programming. *Analytical Chemistry*, **50**(6), 722–731. Citado na pág. 2
- Fletcher, R., Grant, J. A. e Hebden, M. D. (1974). Linear minimax approximation as the limit of best L_p -approximation. *SIAM Journal on Numerical Analysis*, **11**(1), 123–136. Citado na pág. 2
- Flynn, M. J. (1972). Some computer organizations and their effectiveness. *IEEE Transactions on Computers*, **C-21**(9), 948–960. Citado na pág. 158
- Flynn, M. (1966). Very high-speed computing systems. *Proceedings of the IEEE*, **54**(12), 1901–1909. Citado na pág. 158
- Gastwirth, J. L., Gel, Y. R., Hui, W. L. W., Lyubchich, V., Miao, W. e Noguchi, K. (2023). *lawstat: Tools for Biostatistics, Public Policy, and Law*. URL <https://CRAN.R-project.org/package=lawstat>. R package version 3.6. Citado na pág. 5
- Goldbarg, M. C. e Luna, H. P. L. (2005). *Otimização combinatória e programação linear: modelos e algoritmos. 2. ed. rev. e atual.* Rio de Janeiro: Elsevier. Citado na pág. 9
- Goodall, C. (1983). Examining residuals. Em Hoaglin, D. C., Mosteller, F. e Tukey, J. W., editors, *Understanding robust and exploratory data analysis*, páginas 211–246. Wiley series in probability and mathematical statistics, New York. Citado na pág. 91, 92
- Green, J. R. (1971). Testing departure from a regression, without using replication. *Technometrics*, **13**(3), 609–615. Citado na pág. 26
- Groebner, D. F., Shannon, P. W., Fry, P. C. e Smith, K. D. (2001). *Business Statistics: A Decision-Making Approach*. Prentice-hall, 5. edição. Citado na pág. 35
- Gustafson, J. L. (1988). Reevaluating amdahl's law. *Commun. ACM*, **31**(5), 532–533. Citado na pág. 157, 158
- Harris, I. R. (1989). Predictive fit for natural exponential families. *Biometrika*, **76**(4), 675–684. Citado na pág. 82
- Harter, H. L. (1974). The method of least squares and some alternatives: Part I. *International Statistical Review/Revue Internationale de Statistique*, **42**(2), 147–174. Citado na pág. 2, 136
- Harter, H. L. (1974). The method of least squares and some alternatives: Part II. *International Statistical Review/Revue Internationale de Statistique*, **42**(3), 235–282. Citado na pág. 2, 136
- Harter, H. L. (1975). The method of least squares and some alternatives: Part III. *International Statistical Review/Revue Internationale de Statistique*, **43**(1), 1–44. Citado na pág. 2, 136
- Harter, H. L. (1975). The method of least squares and some alternatives: Part IV. *International Statistical Review/Revue Internationale de Statistique*, **43**(2), 125–190. Citado na pág. 2, 136

- Harter, H. L. (1975). The method of least squares and some alternatives: Addendum to Part IV. *International Statistical Review/Revue Internationale de Statistique*, **43**(3), 273–278. Citado na pág. 2, 136
- Harter, H. L. (1975). The method of least squares and some alternatives: Part V. *International Statistical Review/Revue Internationale de Statistique*, **43**(3), 269–272. Citado na pág. 2, 136
- Harter, H. L. (1976). The method of least squares and some alternatives: Part VI. Subject and author indexes. *International Statistical Review/Revue Internationale de Statistique*, **44**(1), 113–159. Citado na pág. 2, 136
- Harter, L. H. (1977). Nonuniqueness of least absolute values regression. *Communications in Statistics-Theory and Methods*, **6**(9), 829–838. Citado na pág. 12
- Harville, D. A. (1997). *Matrix algebra from a statistician's perspective*. New York: Springer. Citado na pág. 136
- Hennig, C., Meila, M., Murtagh, F. e Rocci, R. (2015). *Handbook of Cluster Analysis*. CRC Press. Citado na pág. 42, 44
- Higham, N. J. (2002). *Accuracy and stability of numerical algorithms*. SIAM. Citado na pág. 89
- Hinkley, D. (1979). Predictive likelihood. *The Annals of Statistics*, **7**(4), 718–728. Citado na pág. 81
- Johnson, W. e Geisser, S. (1982). Assessing the predictive influence of observations. Em Kallianpur, Krishnaiah e Ghosh, editors, *Essays in Honor of C. R. Rao*, páginas 343–358. North Holland, Amsterdam. Citado na pág. 86
- Johnson, W. e Geisser, S. (1983). A predictive view of the detection and characterization of influential observations in regression analysis. *Journal of the American Statistical Association*, **78**(381), 137–144. Citado na pág. 86
- Johnston, J. e DiNardo, J. (1996). *Econometric methods*. McGraw-Hill/Irwin, 4 edição. Citado na pág. 43, 47
- Kaufman, L. e Rousseeuw, P. J. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley. Citado na pág. 42, 44
- Kendall, M. (1994). *Kendall's Advanced Theory of Statistics (A. Stuart and J.K. Ord, Eds)*. Arnold, London, 6. edição. Citado na pág. 11
- Koenker, R. e Bassett, G. (1978). Regression quantiles. *Econometrica*, **46**, 33–50. Citado na pág. 7
- Koenker, R. (2005). *Quantile regression*. Cambridge University Press. Citado na pág. 136
- Koenker, R., Portnoy, S., Ng, P. T., Melly, B., Zeileis, A., Grosjean, P., Moler, C., Saad, Y., Chernozhukov, V., Fernandez-Val, I. e Ripley, B. D. *quantreg: Quantile Regression*, 2023. URL <https://CRAN.R-project.org/package=quantreg>. R package version 5.97. Citado na pág. 159

- Kontoghiorghes, E. J. (2005). *Handbook of parallel computing and statistics*. CRC Press. Citado na pág. 158, 161
- Kuks, J. e Olman, W. (1971). Minimax linear estimation of regression coefficients (I) (in Russian). *Iswestija Akademija Nauk Estonskoj SSR*, **20**, 480–482. Citado na pág. 2
- Kuks, J. e Olman, W. (1972). Minimax linear estimation of regression coefficients (II) (in Russian). *Iswestija Akademija Nauk Estonskoj SSR*, **21**, 66–72. Citado na pág. 2
- Kullback, S. e Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, **22**(1), 79 – 86. Citado na pág. 87
- Laplace, P. S. (1793). Sur quelques points du système du monde. mémoires de l'académie royale des sciences de paris. páginas 1–87. Reprinted in his Oeuvres, Vol. 11, Gauthier et Fils, Paris, 1895, pp. 477-558. Citado na pág. 1
- Lauritzen, S. L. (1974). Sufficiency, prediction and extreme models. *Scandinavian Journal of Statistics*, **1**(3), 128–134. Citado na pág. 82
- Lee, Y. e Nelder, J. A. (2009). Likelihood inference for models with unobservables: another view. *Statistical Science*, **24**(3), 255–269. Citado na pág. 82
- Legendre, A. M. (1805). *Nouvelles Méthodes pour la Détermination des Orbites des Comètes*. Firmin Didot. Citado na pág. 1
- Lehmann, E. L. e Casella, G. (2006). *Theory of point estimation*(2nd edition). Springer. ISBN 9780387227283. Citado na pág. 16
- Lehmann, E. L. e Romano, J. P. (2022). *Testing statistical hypotheses*.(4th edition). Springer Texts in Statistics. ISBN 9783030705787. Citado na pág. 16
- Lejeune, M. e Faulkenberry, G. D. (1982). A simple predictive density function. *Journal of the American Statistical Association*, **77**(379), 654–657. Citado na pág. 83, 84
- Leonard, T. (1982). Comment on “a simple predictive density function” by M. Lejeune and G. D. Faulkenberry. *Journal of the American Statistical Association*, **77**(379), 657–658. Citado na pág. 83
- Levy, M. S. e Perng, S. K. (1986). An optimal prediction function for the normal linear model. *Journal of the American Statistical Association*, **81**(393), 196–198. Citado na pág. 84
- Levy, M. S. e Perng, S. (1984). A maximum likelihood prediction function for the linear model with consistency results. *Communications in Statistics A – Theory and Methods*, **13**(10), 1257–1273. Citado na pág. 83
- Mathiasen, P. E. (1979). Prediction functions. *Scandinavian Journal of Statistics*, **6**(1), 1–21. Citado na pág. 83
- Metropolis, N. e Ulam, S. (1949). The monte carlo method. *Journal of the American Statistical Association*, **44**, 335–341. Citado na pág. 154
- Microsoft e Weston, S. *foreach: Provides Foreach Looping Construct*, 2022. URL <https://CRAN.R-project.org/package=foreach>. R package version 1.5.2. Citado na pág. 160

- Narasimhan, B., Koller, M., Johnson, S. G., Hahn, T., Bouvier, A., Kiêu, K. e Gaure, S. *cubature: Adaptive Multivariate Integration over Hypercubes*, 2023. URL <https://CRAN.R-project.org/package=cubature>. R package version 2.0.4.6. Citado na pág. 159
- Narula, S. C. e Wellington, J. F. (1977). Prediction, linear regression and minimum sum of absolute errors. *Technometrics*, **19**, 185–190. Citado na pág. 5
- Narula, S. C. e Wellington, J. F. (1985). Interior analysis for the minimum sum of absolute errors regression. *Technometrics*, **27**, 181–188. Citado na pág. 5
- Neyman, J. e Pearson, E. S. (1933). On the problems of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London*, **231A**, 289–338. Citado na pág. 15
- Planitz, M. (1985). Existence and uniqueness of best approximants, with numerical applications. Tese de Doutorado, Thames Polytechnic, UK. Citado na pág. 12
- Planitz, M. e Gates, J. (1991). Strict discrete approximation in the L_1 and L_∞ norms. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **40**(1), 113–122. Citado na pág. 12
- Puig, P. e Stephens, M. A. (2000). Tests of fit for the laplace distribution, with applications. *Technometrics*, **42**, 417–424. Citado na pág. 153, 154, 155, 156
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023. URL <http://www.R-project.org>. Citado na pág. 32, 46, 52, 72, 155
- Rodrigues, K. A. S. (2019). Diagnóstico em regressão L_1 . Dissertação de Mestrado, IME/USP, Brasil. Citado na pág. 5, 90, 114, 117, 153
- Rodrigues, K. A. S. e Elian, S. N. *diagL1: Routines for Fit, Inference and Diagnostics in Linear L1 and LAD Models. version 1.0.0*, 2024. Citado na pág. 5, 19, 131, 155, 159
- Rojas, M. D. I. (2000). Densidades preditivas no modelo de regressão. Dissertação de Mestrado, IME/USP, Brasil. Citado na pág. 82
- Rosenberg, B. e Carlson, D. (1977). A simplex approximation of the sampling distribution of least absolute residual regression estimates. *Communications in Statistics - Simulation and Computation*, **6**, 421–437. Citado na pág. 13
- Rosenthal, J. S. (2000). Parallel computing and monte carlo algorithms. *Far east journal of theoretical statistics*, **4**(2), 207–236. Citado na pág. 158
- Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association*, **79**(388), 871–880. Citado na pág. 117
- Rousseeuw, P. J. e Leroy, A. M. (2003). *Robust Regression and Outlier Detection*. Wiley. Citado na pág. 1, 2, 117
- Searle, S. R. (2006). *Matrix Algebra Useful for Statistics*. New York, John Wiley & Sons. Citado na pág. 136
- Seber, G. A. F. (2008). *A matrix handbook for statisticians*. Wiley. Citado na pág. 136

- Shao, J. (2005). *Mathematical Statistics: Exercises and Solutions*. Springer. ISBN 9780387249704. Citado na pág. [16](#)
- Shillington, E. R. (1979). Testing lack of fit in regression without replication. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, **7**(2), 137–146. Citado na pág. [27](#)
- Su, Z. e Yang, S.-S. (2006). A note on lack-of-fit tests for linear models without replication. *Journal of the American Statistical Association*, **101**(473), 205–210. Citado na pág. [xvii](#), [21](#), [32](#), [35](#), [36](#), [37](#)
- Sun, R.-B. e Wei, B.-C. (2004). On influence assessment for lad regression. *Statistics & Probability Letters*, **67**, 97–110. Citado na pág. [91](#), [114](#)
- Tavares, R. A. (1998). Seleção de variáveis em regressão L_1 . Dissertação de Mestrado, IME/USP, Brasil. Citado na pág. [10](#)
- Turner, H. H. (1887). On mr. edgeworth's method of reducing observations relating to several quantities. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **24**(151), 466–470. Citado na pág. [12](#)
- Villegas, C., Paula, G. A., Cysneiros, F. J. A. e Galea, M. (2013). Influence diagnostics in generalized symmetric linear models. *Computational Statistics & Data Analysis (Print)*, **59**, 161–170. Citado na pág. [17](#)
- Wilson, H. G. (1979). Upgrading transport costing methodology. *Transportation Journal*, **18**(3), 49–55. Citado na pág. [2](#)
- Yoo, S.-H. (2001). A robust estimation of hedonic price models: least absolute deviations estimation. *Applied Economics Letters*, **8**(1), 55–58. Citado na pág. [2](#)