

Universidade Federal de Campina Grande  
Centro de Ciências e Tecnologia  
Programa de Pós-Graduação em Matemática  
Curso de Mestrado em Matemática

# Estimação Robusta para o Modelo de Tempo de Vida Acelerado Birnbaum-Saunders<sup>†</sup>

por

Deyvison Guimarães Barbosa

sob orientação da

Prof<sup>a</sup>. Dr<sup>a</sup>. Michelli Karinne Barros da Silva

Dissertação apresentada ao Corpo Docente do Programa de Pós-Graduação em Matemática - CCT - UFCG, como requisito parcial para obtenção do título de Mestre em Matemática.

---

<sup>†</sup>Este trabalho contou com apoio financeiro do CNPq

B238e

Barbosa, Deyvison Guimarães.

Estimação robusta para o modelo de tempo de vida acelerado Birnbaum-Saunders / Deyvison Guimarães Barbosa. - Campina Grande, 2019.

123 f. : il.

Dissertação (Mestrado em Matemática) - Universidade Federal de Campina Grande, Centro de Ciências e Tecnologia, 2019.

"Orientação: Profa. Dra. Michelli Karinne Barros da Silva  
Referências.

1. Estimação Robusta. 2. Modelo de Tempo de Vida Acelerado. 3. Distribuição Birnbaum-Saunders. I. Silva, Michelli Karinne Barros da. II. Título.

CDU 519.246.8(043)

# Estimação Robusta para o Modelo de Tempo de Vida Acelerado Birnbaum-Saunders

por

Deyvison Guimarães Barbosa

Dissertação apresentada ao Corpo Docente do Programa de Pós-Graduação em Matemática - CCT - UFCG, como requisito parcial para obtenção do título de Mestre em Matemática.

Área de Concentração: Probabilidade e Estatística

Aprovada por:

---

Prof<sup>fa</sup>. Dr<sup>a</sup>. Aline Barbosa Tsuyuguchi - UFPE

---

Prof<sup>o</sup>. Dr. Joelson da Cruz Campos - UFCG

---

Prof<sup>a</sup>. Dr<sup>a</sup>. Michelli Karinne Barros da Silva - UFCG

Orientadora

Universidade Federal de Campina Grande  
Centro de Ciências e Tecnologia  
Programa de Pós-Graduação em Matemática  
Curso de Mestrado em Matemática

Novembro / 2019



# Resumo

O método de estimação robusto é uma importante ferramenta para limitar a influência de *outliers* tanto na variável resposta quanto nas covariáveis de um modelo de tempo de vida acelerado. Neste trabalho utilizamos a distribuição Birnbaum-Saunders, a qual tem sido amplamente utilizada em estudos de análise de sobrevivência. Apresentamos a forma log-linear do modelo de tempo de vida acelerado, realizamos a estimação robusta e obtivemos as variâncias dos estimadores robustos por uma função de variância e covariância do tipo sanduíche. Além disso, usando diferentes cenários, realizamos estudos de simulação para avaliar o desempenho do estimador robusto proposto. Por fim, apresentamos duas aplicações utilizando conjuntos de dados reais para ilustrar a metodologia apresentada.

**Palavras-chave.** Estimação robusta; Modelo de tempo de vida acelerado; Distribuição Birnbaum-Saunders.



# Abstract

The robust estimation method is an important tool to limit the influence of *outliers* on both the response variable and the covariates of an accelerated life model. In this work we used the Birnbaum-Saunders distribution, which has been extensively applied in survival analysis studies. We presented the log-linear form of the accelerated life model, we performed the robust estimation and approximated the variances of the robust estimators by a function of variance and covariance of the sandwich type. In addition, using different scenarios, we performed simulation studies to evaluate the performance of the proposed robust estimator. Finally, we present two applications in real data sets that illustrate the methodology presented.

**Keywords.** Robust estimation; Accelerated Lifetime Model; Birnbaum-Saunders Distribution.





# Agradecimentos

Muitas das pessoas de quem mais gosto estiveram, direta ou indiretamente, presentes durante o desenvolvimento deste trabalho, e cada uma delas me deu incentivos valiosos. Agradeço sinceramente a João Salvino, Ana Beatriz, Ana Gabriela, Klebia, Kennedy, Giovana, Thalytta, Ioneris, Vera, Felipe, Marcelo, Amanda e Larissa.

Também devo muito a Terezinha, a pessoa mais consistentemente entusiasta do universo, que sempre foi minha incentivadora, certamente não estaria onde estou sem sua motivação. Amo e agradeço.

Meus agradecimentos também vão para o professor Joelson, por todo seu conhecimento e bom humor. Suas sugestões valiosas contribuíram muito para o desenvolvimento deste trabalho.

Mais agradecimentos à equipe incrivelmente dedicada da UAEst, professores e funcionários, pelo suporte e contribuição. À professora Aline Tsuyuguchi por ter aceitado o convite para participar da banca de avaliação deste trabalho e pelas sugestões e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo financiamento deste projeto.

Sou especialmente grato a brilhante professora, Michelli Barros, pela excelente orientação, cordialidade, compreensão, paciência, dedicação e por ter me guiado graciosamente em todas as etapas desde minha aceitação no programa até a conclusão desta dissertação. E ao seu esposo Horácio com que tem uma linda família.

Finalmente, muito amor e consideração por minha mãe, Normanda Guimarães. Sempre encorajadora e apoiadora. Obrigado por ser tão boa pra mim e por acreditar nos momentos em que até eu duvidei. Me torno uma pessoa melhor só por ter você, minha melhor companhia.

Serei sempre grato.



# Dedicatória

Para Ana Beatriz e Ana Gabriela,  
sobrinhas e guardiãs do meu coração.



# Sumário

<b>Introdução</b>	<b>1</b>
Organização da Dissertação . . . . .	3
Suporte Computacional . . . . .	4
<b>1 Conceitos Básicos</b>	<b>5</b>
1.1 Análise de Sobrevivência . . . . .	5
1.1.1 Caracterização dos dados de sobrevivência . . . . .	5
1.1.2 Representação dos dados de sobrevivência . . . . .	6
1.1.3 Funções importantes na análise de sobrevivência . . . . .	6
1.1.4 Modelo de tempo de vida acelerado . . . . .	13
1.2 Robustez . . . . .	17
1.2.1 Estatísticas robustas . . . . .	18
1.2.2 M-Estimadores . . . . .	22
1.2.3 Estimação robusta . . . . .	24
<b>2 Distribuição Birnbaum-Saunders</b>	<b>29</b>
2.1 Origem . . . . .	29
2.2 Função de sobrevivência e função de risco . . . . .	32
2.3 Propriedades . . . . .	34
2.4 Estimação . . . . .	36
2.5 Distribuição Senh-Normal . . . . .	37
2.5.1 Propriedades da distribuição Senh-Normal . . . . .	39
2.6 Modelo de tempo de vida acelerado Birnbaum-Saunders . . . . .	41

<b>3</b>	<b>Estimação robusta para o modelo de tempo de vida acelerado Birnbaum-Saunders</b>	<b>45</b>
3.1	Estudo de simulação . . . . .	50
3.2	Exemplo simulado . . . . .	63
<b>4</b>	<b>Aplicações</b>	<b>71</b>
4.1	Aplicação 1 . . . . .	71
4.1.1	Estudo sobre a vida útil de uma peça de metal . . . . .	71
4.1.2	Ajuste do modelo log-BS . . . . .	73
4.2	Aplicação 2 . . . . .	75
4.2.1	Estudo sobre o tempo de vida de pacientes com câncer de pulmão	75
4.2.2	Ajuste do modelo log-BS . . . . .	77
<b>5</b>	<b>Conclusões</b>	<b>89</b>
	<b>Apêndice</b>	<b>90</b>
<b>A</b>	<b>Expressões do Capítulo 3</b>	<b>91</b>
A.1	Funções para $\epsilon_i \sim \text{SN}(\alpha, 0, 2)$ . . . . .	91
A.2	Funções escore . . . . .	92
A.3	Funções $\Upsilon'(\boldsymbol{\theta})$ . . . . .	95
	<b>Referências Bibliográficas</b>	<b>104</b>

# Lista de Figuras

2.1	Função densidade da distribuição BS para $\beta = 1$ com diferentes valores de $\alpha$ (a) e $\alpha = 0,5$ com diferentes valores de $\beta$ (b). . . . .	32
2.2	Função de sobrevivência da distribuição BS para $\beta = 1$ com diferentes valores de $\alpha$ (a) e $\alpha = 0,5$ com diferentes valores de $\beta$ (b). . . . .	33
2.3	Função de risco da distribuição BS para $\beta = 1$ com diferentes valores de $\alpha$ (a) e $\alpha = 1$ com diferentes valores de $\beta$ (b). . . . .	34
2.4	Função densidade da distribuição senh-normal para $\gamma = 0$ e $\sigma = 1$ com diferentes valores de $\alpha$ (a) e $\alpha = 1$ , $\gamma = 0$ com diferentes valores de $\sigma$ (b). . . . .	38
3.1	Vieses empíricos da estimativa do $\alpha$ para diferentes combinações de tempos de sobrevivência censurados: 0% (a), 10% (b) e 30% (c). . . . .	57
3.2	Vieses empíricos da estimativa do $\beta_0$ com três diferentes combinações de tempos de sobrevivência censurados: 0% (a), 10% (b) e 30% (c). . . . .	57
3.3	Vieses empíricos da estimativa do $\beta_1$ com três diferentes combinações de tempos de sobrevivência censurados: 0% (a), 10% (b) e 30% (c). . . . .	57
3.4	Erros quadráticos médios (EQM) da estimativa do $\alpha$ para diferentes combinações de tempos de sobrevivência censurados: 0% (a), 10% (b) e 30% (c). . . . .	58
3.5	Erros quadráticos médios (EQM) da estimativa do $\beta_0$ com três diferentes combinações de tempos de sobrevivência censurados: 0% (a), 10% (b) e 30% (c). . . . .	58
3.6	Erros quadráticos médios (EQM) da estimativa do $\beta_1$ com três diferentes combinações de tempos de sobrevivência censurados: 0% (a), 10% (b) e 30% (c). . . . .	58

3.7	Probabilidades de cobertura (PCs) da estimativa do $\alpha$ para diferentes combinações de tempos de sobrevivência censurados: 0% (a), 10% (b) e 30% (c). . . . .	59
3.8	Probabilidades de cobertura (PCs) da estimativa do $\beta_0$ com três diferentes combinações de tempos de sobrevivência censurados: 0% (a), 10% (b) e 30% (c). . . . .	59
3.9	Probabilidades de cobertura (PCs) da estimativa do $\beta_1$ com diferentes combinações de tempos de sobrevivência censurados: 0% (a), 10% (b) e 30% (c). . . . .	59
3.10	Vieses empíricos da estimativa do $\alpha$ para diferentes combinações de tamanhos amostrais: $n = 100$ (a) e $n = 200$ (b). . . . .	60
3.11	Vieses empíricos da estimativa do $\beta_0$ para diferentes combinações de tamanhos amostrais: $n = 100$ (a) e $n = 200$ (b). . . . .	60
3.12	Vieses empíricos da estimativa do $\beta_1$ para diferentes combinações de tamanhos amostrais: $n = 100$ (a) e $n = 200$ (b). . . . .	60
3.13	Erros quadráticos médios (EQM) da estimativa do $\alpha$ para diferentes combinações de tamanhos amostrais: $n = 100$ (a) e $n = 200$ (b). . . . .	61
3.14	Erros quadráticos médios (EQM) da estimativa do $\beta_0$ para diferentes combinações de tamanhos amostrais: $n = 100$ (a) e $n = 200$ (b). . . . .	61
3.15	Erros quadráticos médios (EQM) da estimativa do $\beta_1$ para diferentes combinações de tamanhos amostrais: $n = 100$ (a) e $n = 200$ (b). . . . .	61
3.16	Probabilidades de cobertura (PCs) da estimativa do $\alpha$ para diferentes combinações de tamanhos amostrais: $n = 100$ (a) e $n = 200$ (b). . . . .	62
3.17	Probabilidades de cobertura (PCs) da estimativa do $\beta_0$ para diferentes combinações de tamanhos amostrais: $n = 100$ (a) e $n = 200$ (b). . . . .	62
3.18	Probabilidades de cobertura (PCs) da estimativa do $\beta_1$ para diferentes combinações de tamanhos amostrais: $n = 100$ (a) e $n = 200$ (b). . . . .	62
3.19	Gráficos dos tempos de sobrevivência contra as covariáveis $x_1$ e $x_2$ para o conjuntos de dados simulados sem <i>outlier</i> . . . . .	63
3.20	Gráficos dos resíduos $r_i$ baseados no ERM e EMV ajustados aos dados simulados sem <i>outliers</i> . . . . .	64



3.21	Gráficos normais de probabilidade para os resíduos $r_{q_i}$ baseados no ERM e EMV ajustados aos dados simulados sem <i>outliers</i> . . . . .	65
3.22	Gráficos dos tempos de sobrevivência contra as covariáveis $x_1$ e $x_2$ para o conjuntos de dados simulados com <i>outliers</i> . . . . .	66
3.23	Gráficos dos resíduos $r_i$ baseados no ERM e EMV ajustados aos dados simulados com <i>outliers</i> . . . . .	67
3.24	Gráficos dos pesos $w(r_i)$ utilizado no ERM para ajustar os dados simulados com <i>outliers</i> . . . . .	67
3.25	Gráficos normais de probabilidade para os resíduos $r_{q_i}$ baseados no ERM e EMV para os dados simulados com <i>outliers</i> . . . . .	68
4.1	Histograma e boxplot para o número de ciclos até a ocorrência da falha.	72
4.2	Gráficos dos resíduos $r_i$ baseados no ERM e EMV ajustados aos dados sobre a vida útil de uma peça de metal. . . . .	74
4.3	Gráficos dos resíduos $r_{q_i}$ baseados no ERM e EMV ajustados aos dados sobre a vida útil de uma peça de metal. . . . .	74
4.4	Gráficos normais de probabilidade para os resíduos $r_{q_i}$ baseados no ERM e EMV ajustados aos dados sobre a vida útil de uma peça de metal. . .	75
4.5	Histograma e boxplot dos tempos de sobrevivência dos pacientes com câncer de pulmão. . . . .	76
4.6	Gráficos dos resíduos $r_i$ baseados no ERM e EMV ajustados aos dados de câncer de pulmão. . . . .	79
4.7	Gráficos dos resíduos $r_{q_i}$ baseados no ERM e EMV ajustados aos dados de câncer de pulmão. . . . .	79
4.8	Gráficos normais de probabilidade para os resíduos $r_{q_i}$ baseados no ERM e EMV ajustados aos dados de câncer de pulmão. . . . .	80
4.9	Gráficos dos pesos $w(r_i)$ utilizado no ERM para ajustar os dados de câncer de pulmão. . . . .	84
4.10	Gráfico dos resíduos $r_i$ baseados no ERM e EMV ajustados ao modelo reduzido dos dados de câncer de pulmão. . . . .	85
4.11	Gráfico dos resíduos $r_{q_i}$ baseados no ERM e EMV ajustados ao modelo reduzido dos dados de câncer de pulmão. . . . .	85

4.12 Gráficos normais de probabilidade para os resíduos $r_{q_i}$ baseados no ERM e EMV ajustados ao modelo reduzido dos dados de câncer de pulmão. .	88
---	----

# Lista de Tabelas

1.1	Estimativas clássicas e robustas para o teor de cobre na farinha integral para diferentes quantidades de observações. . . . .	20
3.1	Vieses empíricos, erros quadráticos médios (EQM), probabilidades de cobertura (PCs) e comprimentos médios dos intervalos de confiança de 95% para estimação do modelo Birnbaum-Saunders com os métodos: robusto (ERM) e clássico (EMV). (0% de censura). . . . .	54
3.2	Vieses empíricos, erros quadráticos médios (EQM), probabilidades de cobertura (PCs) e comprimentos médios dos intervalos de confiança de 95% para estimação do modelo Birnbaum-Saunders com os métodos: robusto (ERM) e clássico (EMV). (10% de censura). . . . .	55
3.3	Vieses empíricos, erros quadráticos médios (EQM), probabilidades de cobertura (PCs) e comprimentos médios dos intervalos de confiança de 95% para estimação do modelo Birnbaum-Saunders com os métodos: robusto (ERM) e clássico (EMV). (30% de censura). . . . .	56
3.4	Estimativa dos parâmetros do modelo log-BS para os dados simulados sem <i>outliers</i> utilizando o ERM e o EMV. . . . .	64
3.5	Estimativa dos parâmetros do modelo log-BS para os dados simulados com <i>outliers</i> utilizando o ERM e o EMV. . . . .	66
3.6	Comparação das estimativa dos parâmetros do modelo log-BS. O ERM para os dados com <i>outliers</i> e o EMV para os dados após retirada dos <i>outliers</i> . . . . .	69
4.1	Dados de fadiga biaxial de Brown e Miller . . . . .	71
4.2	Medidas resumo da variável $N$ . . . . .	72

4.3	Estimativa dos parâmetros do modelo log-BS para os dados sobre a vida útil de uma peça de metal utilizando o ERM e o EMV. . . . .	73
4.4	Medidas resumo da variável $T$ . . . . .	76
4.5	Estimativa dos parâmetros do modelo log-BS para os dados de câncer de pulmão utilizando o ERM e o EMV. . . . .	78
4.6	Estimativas robustas, mudanças relativas (-MR- em %) e os correspondentes níveis descritivos para os casos indicados. . . . .	81
4.7	Estimativas por máxima verossimilhança, mudanças relativas (-MR- em %) e os correspondentes níveis descritivos para os casos indicados. . . . .	82
4.8	Estimativa dos parâmetros do modelo reduzido log-BS para os dados de câncer de pulmão utilizando o ERM e o EMV. . . . .	83
4.9	Estimativas robustas, mudanças relativas (-MR- em %) e os correspondentes níveis descritivos para os casos indicados do modelo reduzido. . . . .	86
4.10	Estimativas por máxima verossimilhança, mudanças relativas (-MR- em %) e os correspondentes níveis descritivos para os casos indicados do modelo reduzido. . . . .	87

# Lista de Abreviações

ERM	Estimador robusto de máxima verossimilhança
EMV	Estimador de máxima verossimilhança
CV	Coefficiente de variação
BS	Birnbaum-Saunders
BSG	Birnbaum-Saunders Generalizada
SN	Senh-Normal
EQM	Erro quadrático médio
PC	Probabilidade de cobertura
Comp.	Comprimento
MR	Mudança relativa
fda	Função de distribuição acumulada
fdp	Função densidade de probabilidade



# Introdução

Nas últimas décadas, o modelo de regressão de Cox tem sido o mais utilizado na análise de dados de sobrevivência, dada a sua flexibilidade. Porém, nem sempre as hipóteses subjacentes ao modelo são verificadas tornando seu uso não plausível. O modelo de tempo de vida acelerado, embora menos utilizado, constitui frequentemente uma boa alternativa, pois tem interpretação física intuitiva e uma capacidade de abranger uma ampla gama de distribuições de vida. Além disso, a função das covariáveis é acelerar ou desacelerar o tempo de vida. Uma boa revisão desses modelos e suas aplicações na Análise de Sobrevivência podem ser encontradas em [Wei \(1992\)](#), [Kalbfleisch e Prentice \(2011\)](#) e [Collett \(2015\)](#).

Um modelo, seja ele paramétrico ou semiparamétrico, é usado para descrever os dados de sobrevivência no qual é necessário realizar o procedimento de estimação. No entanto, a validade e o bom desempenho desse procedimento geralmente exige uma adesão estrita às premissas do modelo, uma condição que está em contraste com a experiência obtida em trabalhos de campo ([Heritier et al., 2009](#)). Sob essas premissas um dos métodos de estimação favoritos da estatística clássica é o de máxima verossimilhança. Valores atípicos, os chamados *outliers* podem violar tais hipóteses provocando a instabilidade dos métodos clássicos de estimação. Além de tais observações atípicas, existem na prática numerosos exemplos de desvios que afetam sensivelmente o rendimento das técnicas clássicas. O problema dos *outliers* é uma questão importante na Análise de Sobrevivência pois gera falta de confiabilidade dos pressupostos exatos do modelo e causa problemas no cálculo das estimativas dos parâmetros. Como uma alternativa para o método de estimação de máxima verossimilhança, é sugerido um método robusto, o qual traz uma pequena perda de eficiência, mas permite lidar com desvios moderados dos modelos “ideais”. Nos concentraremos principalmente na robustez em

relação aos *outliers*. A estatística robusta é uma extensão da estatística clássica que leva em conta especificamente o fato de que os modelos usados são apenas aproximados. A filosofia básica é produzir procedimentos estatísticos que sejam estáveis em relação a pequenas mudanças nos dados ou a outros desvios. Métodos de detecção de *outliers* em bancos de dados são bastante discutidos em muitos livros didáticos sobre métodos estatísticos clássicos, exemplo disso foi o trabalho realizado por [Pinto et al. \(2015\)](#) que propuseram ferramentas para identificar *outliers* em dados de sobrevivência multivariada.

Os primeiros grandes avanços dos métodos robustos ocorreram nos anos 60, podemos citar, como os mais relevantes, os trabalhos de: [Tukey \(1960\)](#) que aparentemente foi o primeiro a reconhecer a extrema sensibilidade para desvios dos pressupostos de alguns procedimentos estatísticos convencionais; [Huber \(1964\)](#) que iniciou uma busca de estimadores robustos segundo um ponto de vista formal e introduziu uma classe de estimadores (M-estimadores) que veremos com detalhes mais adiante; [Huber et al. \(1967\)](#) que provaram a consistência e normalidade assintótica dos M-estimadores sob condições muito gerais, servindo de base para análises do comportamento assintótico dessa classe e [Hampel \(1968\)](#) que definiu formalmente o conceito de robustez qualitativa, definiu também outros conceitos como curva de influência e ponto de ruptura. Daí, o campo da estatística robusta experimentou um crescimento substancial como área de pesquisa, o que pode ser evidenciado pelo grande número de livros e artigos publicados dentre os quais destacamos [Yohai \(1974\)](#), [Huber e Ronchetti \(1981\)](#), [Hampel et al. \(1986\)](#), [Rousseeuw e Leroy \(1987\)](#) e [Staudte e Sheather \(1990\)](#). Como obras mais recentes temos [Heritier et al. \(2009\)](#), [Andrews e Hampel \(2015\)](#) e [Maronna et al. \(2018\)](#). Para lidar especificamente com a presença de *outliers*, foram desenvolvidos extensos trabalhos na literatura para a análise robusta de dados de sobrevivência, mas a maioria focado no modelo de riscos proporcionais de Cox. A análise robusta voltada para os modelos paramétricos de tempo de vida acelerados foi utilizada anteriormente por [Locatelli et al. \(2011\)](#) em que estimadores robustos para modelos de tempo de falha acelerado com distribuição de erro assimétrica (ou simétrica) e observações censuradas foram propostos. Um estudo mais recente foi feito por [Sinha \(2018\)](#) no qual foi desenvolvido um método robusto na estrutura do modelo de regressão de tempo de vida acelerado, que é capaz de lidar com *outliers* tanto na variável resposta quanto nas co-



variáveis. Tal método utiliza a forma log-linear do modelo de tempo de vida acelerado e introduz uma versão robusta das equações de verossimilhança usando a função psi de Huber ([Huber et al., 1973](#)), essa função é bem-comportada e fornece estimadores robustos únicos, que são quase tão eficientes, no sentido de atingir a menor variância assintótica possível, quanto os estimadores de máxima verossimilhança quando não há valores discrepantes nos dados.

Neste trabalho propomos estimação robusta para os parâmetros do modelo de vida acelerado Birnbaum-Saunders (BS). A distribuição BS vem ganhando destaque nas últimas décadas, inclusive com algumas generalizações e extensões. Essa distribuição possui vantagens comparada a outras distribuições de vida principalmente quando há interesse no ajuste dos percentis mais baixos ou mais altos da distribuição. Para mais detalhes podemos citar [Leiva \(2016\)](#) que descreve os desenvolvimentos teóricos mais recentes desse modelo, [Birnbaum e Saunders \(1969\)](#) que definem e obtêm originalmente os estimadores de máxima verossimilhança e [Rieck e Nedelman \(1991\)](#) que propuseram um modelo na forma log-linear da distribuição e desenvolveram métodos de estimação. Em [Barros et al. \(2008\)](#) foi discutido que o processo de estimação por máxima verossimilhança para o modelo BS é sensível a observações discrepantes, como alternativa para a modelagem dos dados, os autores propuseram uma nova classe de modelos de regressão baseados na distribuição Birnbaum-Saunders Generalizada (BSG), mais especificamente na BS- $t$  (modelo log-BSG cujo erro tem distribuição BSG gerada pela distribuição  $t$  de Student). Para essa classe o processo de estimação por máxima verossimilhança é robusto. Aqui, seguimos por outro caminho, utilizaremos um método de estimação robusto para os parâmetros do modelo de tempo de vida acelerado BS de acordo com a proposta de [Sinha \(2018\)](#).

## Organização da Dissertação

A presente dissertação está dividida em cinco capítulos como descrito à seguir. No primeiro capítulo apresentamos os conceitos básicos de Análise de Sobrevida que foram fundamentais para o desenvolvimento desse trabalho. Além disso, abordamos noções gerais de robustez e os principais conceitos relacionados a estimação robusta. No segundo capítulo, estudamos as distribuições Birnbaum-Saunders e Senh-Normal, bem

como as principais propriedades referentes à essas distribuições. No terceiro capítulo realizamos a estimação robusta dos parâmetros do modelo log-linear de tempo de vida acelerado BS juntamente com a aproximação das variâncias dos estimadores robustos por meio de uma função de variância e covariância do tipo sanduíche. Neste capítulo, apresentamos também os resultados obtidos através dos estudos de simulação realizados no intuito de avaliar o desempenho do estimador robusto proposto. No quarto capítulo aplicamos a teoria estudada em duas situações reais de estudo. Por fim, no quinto capítulo relatamos as conclusões do trabalho.

## Suporte Computacional

A linguagem de programação R constitui a plataforma computacional utilizada no desenvolvimento deste trabalho. As apresentações gráficas, os estudos de simulação, o exemplo simulado e a aplicação foram feitas utilizando o *software* R em sua versão 3.4.2 para Windows que se encontra disponível em <http://www.r-project.org> gratuitamente. Esta linguagem caracteriza-se pela flexibilidade e a conveniência dos tradicionais pacotes estatísticos possibilitando facilmente a manipulação de dados e modelagem estatística em geral. Maiores detalhes podem ser encontrados em Ihaka e Gentleman (1996) e em Cribari-Neto e Zarkos (1999).

# Capítulo 1

## Conceitos Básicos

No decorrer desse capítulo realizaremos uma breve revisão de alguns conceitos da área de Análise de Sobrevivência. Abordaremos noções gerais de robustez e os principais conceitos relacionados a estimação robusta. Tal revisão será de fundamental importância para os estudos a serem realizados neste trabalho.

### 1.1 Análise de Sobrevivência

A Análise de Sobrevivência consiste no estudo de dados em que a variável resposta corresponde ao tempo até a ocorrência de um determinado evento de interesse. Segundo [Colosimo e Giolo \(2006\)](#) este evento é denominado como *falha*, que pode corresponder, por exemplo, ao tempo decorrido do diagnóstico até a cura ou morte de um paciente ou ao tempo até a queima de um componente eletrônico. Assume-se que a variável resposta é não-negativa e medida em uma escala contínua. Na Análise de Sobrevivência o objetivo é estimar o tempo de falha (ou tempo de vida).

#### 1.1.1 Caracterização dos dados de sobrevivência

Em muitos experimentos o acompanhamento das unidades é frequentemente interrompido, seja porque o estudo terminou para a análise ou a falha ocorreu devido uma causa diferente daquela sob estudo. Neste caso, a informação correspondente implica em uma resposta incompleta ou parcial chamada de censura. [Colosimo e Giolo \(2006\)](#) apontam a censura como uma das principais características dos dados da Análise de

Sobrevivência possuindo mecanismos em três diferentes tipos: tipo I, tipo II e aleatória. A censura tipo I ocorre quando o estudo é concebido com uma data pré-estabelecida para acabar e em seu término há indivíduos que não apresentaram o evento de interesse. A Censura tipo II acontece quando o estudo termina após a ocorrência de um número de falhas predefinidas e a censura aleatória ocorre quando um indivíduo deixa o estudo sem que tenha ocorrido a falha, ou seja, por motivos que diferem do evento de interesse.

Os tipos de censura ainda podem ser classificadas como: à direita, à esquerda e intervalar. A censura à direita é quando o tempo de ocorrência do evento de interesse está à direita do tempo estabelecido para a realização do estudo, já a censura à esquerda é quando o tempo registrado é maior que o tempo de falha, ou seja, o evento de interesse já aconteceu antes da observação e a censura intervalar é aquela que acontece em estudos periódicos e é conhecido somente que o evento de interesse ocorreu em um certo intervalo de tempo.

### 1.1.2 Representação dos dados de sobrevivência

Considere  $T$  uma variável aleatória representando o tempo de falha. Os dados de sobrevivência para o indivíduo  $i$  ( $i = 1, 2, \dots, n$ ) de uma amostra aleatória de  $T$  ( $t_1, \dots, t_n$ ) são geralmente representados pelo par  $(t_i, \delta_i)$  sendo  $t_i$  o tempo de falha ou censura e  $\delta_i$  a variável indicadora de falha, isto é,

$$\delta_i = \begin{cases} 0 & \text{se } t_i \text{ é um tempo censurado} \\ 1 & \text{se } t_i \text{ é um tempo de falha.} \end{cases}$$

Quando no estudo há presença de covariáveis  $\mathbf{x}_i$ , os dados ficam representados por  $(t_i, \delta_i, \mathbf{x}_i)$ .

### 1.1.3 Funções importantes na análise de sobrevivência

Seja  $T$  uma variável aleatória contínua não-negativa, representando o tempo de falha. Em Análise de Sobrevivência essa variável é geralmente especificada pelas funções de sobrevivência e taxa de falha (Colosimo e Giolo, 2006). A seguir definiremos essas e algumas outras funções básicas.

## Função de sobrevivência

A função de sobrevivência é a probabilidade de um indivíduo não sofrer determinado evento durante um intervalo  $[0, t]$ , ou seja, a probabilidade de que o indivíduo sobreviva mais que o tempo  $t$  é definida como

$$S(t) = P(T > t).$$

A função de sobrevivência se relaciona com a função de distribuição acumulada da seguinte forma,

$$\begin{aligned} S(t) &= P(T > t) \\ &= 1 - P(T \leq t) \\ &= 1 - F(t). \end{aligned}$$

A função de sobrevivência é uma função não crescente no tempo com as propriedades de que a probabilidade de sobreviver pelo menos ao tempo zero é 1 ( $S(0) = 1$ ), à medida que o tempo passa,  $S(t)$  decresce ou permanece constante e no limite, quando  $t \rightarrow \infty$ , a probabilidade de sobreviver indefinidamente no tempo tende a zero ( $S(\infty) = 0$ ).

## Função de taxa de falha

A função taxa de falha, também chamada de função de risco, é uma outra função importante na análise de dados de sobrevivência. Por definição, a taxa de falha em um intervalo de tempo  $[t_1, t_2)$  é a razão entre a probabilidade de um evento ocorrer neste intervalo, sabendo que o evento não ocorreu antes de  $t_1$ , e o comprimento do intervalo  $[t_1, t_2)$ . Em outras palavras, a taxa de falha é dada por

$$\frac{P(t_1 \leq T < t_2 | T \geq t_1)}{t_2 - t_1}. \quad (1.1)$$

Usando o fato que

$$P(t_1 \leq T < t_2 | T \geq t_1) = \frac{P(t_1 \leq T < t_2)}{P(T \geq t_1)} = \frac{S(t_1) - S(t_2)}{S(t_1)}$$

a expressão (1.1) pode ser reescrita como

$$\frac{S(t_1) - S(t_2)}{(t_2 - t_1)S(t_1)}. \quad (1.2)$$

De forma geral, redefinindo o intervalo como  $[t, t + \Delta t)$ , a expressão (1.2) assume a forma

$$\frac{S(t) - S(t + \Delta t)}{\Delta t S(t)}.$$

Definimos então, a função taxa de falha instantânea da variável aleatória  $T$ , indicada por  $h$ , como sendo o limite

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{S(t) - S(t + \Delta t)}{\Delta t S(t)} = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}. \quad (1.3)$$

Conforme Colosimo e Giolo (2006) a função taxa de falha é mais informativa do que a função de sobrevivência. De fato, diferentes distribuições podem apresentar formas das funções taxa de falha muito distintas enquanto que as respectivas funções de sobrevivência podem ser bem semelhantes.

### Função de taxa de falha acumulada

Outra função útil na caracterização do tempo de sobrevivência é a taxa de falha acumulada ou risco acumulado, indicada por  $H$  e definida por

$$H(t) = \int_0^t h(u) du, \quad (1.4)$$

essa função, como o próprio nome diz, fornece a taxa de falha acumulada do indivíduo. Segundo Colosimo e Giolo (2006) esta função não possui interpretação direta, porém é bastante útil na avaliação da função de maior interesse que é a de taxa de falha instantânea,  $h(t)$ . Como na estimação não paramétrica  $h(t)$  é difícil de ser estimada, podemos utilizar  $H(t)$  que possui propriedades ótimas.

## Tempo médio e vida média residual

O tempo médio de vida é dado por

$$t_m = E(T),$$

mas como  $T$  é uma variável aleatória que assume apenas valores não-negativos, o tempo médio de vida pode ser obtido através da equação

$$t_m = \int_0^{\infty} S(t)dt,$$

pois

$$\begin{aligned} \int_0^{\infty} S(t)dt &= \int_0^{\infty} P(T > t)dt \\ &= \int_0^{\infty} \int_t^{\infty} f_T(x)dxdt \\ &= \int_0^{\infty} \int_0^x f_T(x)dt dx \\ &= \int_0^{\infty} [tf_T(x) \big|_0^x] dx \\ &= \int_0^{\infty} xf_T(x)dx \\ &= E(T). \end{aligned}$$

A vida média residual corresponde ao tempo médio restante de vida de um indivíduo condicionada a um tempo  $t$  e é definida por

$$\text{vmr}(t) = \frac{\int_t^{\infty} (u-t)f(u)du}{S(t)} \stackrel{*}{=} \frac{\int_t^{\infty} S(u)du}{S(t)}. \quad (1.5)$$

Podemos justificar (\*) da seguinte forma, note que

$$\int_t^{\infty} (u-t)f(u)du = \int_t^{\infty} uf(u)du - t \int_t^{\infty} f(u)du,$$

assim, integrando por partes, fazendo  $w = u$  e  $dv = f(u)du = -\frac{d}{du}S(u)$ , teremos

$dw = du$  e  $v = -S(u)$ , uma vez que  $\int_t^\infty f(u)du = S(t)$ , temos

$$\begin{aligned}
\int_t^\infty (u-t)f(u)du &= \lim_{x \rightarrow \infty} -uS(u)|_t^x + \int_t^\infty S(u)du - tS(t) \\
&= -\lim_{x \rightarrow \infty} xS(x) + tS(t) + \int_t^\infty S(u)du - tS(t) \\
&= -\lim_{x \rightarrow \infty} xS(x) + \int_t^\infty S(u)du \\
&\stackrel{**}{=} \int_t^\infty S(u)du,
\end{aligned}$$

em (\*\*) usamos o fato de  $E(T) < \infty$  implica que  $\lim_{x \rightarrow \infty} xS(x) = 0$  (Kalbfleisch e Prentice, 2011). Portanto, a vida média residual pode ser simplificada da seguinte forma

$$\text{vmr}(t) = \frac{\int_t^\infty S(u)du}{S(t)}.$$

Observe que a vida média residual está relacionada com o tempo médio de vida, uma vez que  $\text{vmr}(0) = t_m$ .

## Relações entre as funções

Através das funções abordadas até agora, apresentaremos algumas relações importantes, as quais mostram que o conhecimento de uma das funções, implica no conhecimento das demais. Considerando  $T$  uma variável aleatória contínua e não-negativa, temos que:

$$1. \ h(t) = -\frac{d}{dt} (\log S(t)).$$

*Demonstração:* Usando (1.3) temos

$$\begin{aligned}
h(t) &= \lim_{\Delta t \rightarrow 0} \frac{S(t) - S(t + \Delta t)}{\Delta t S(t)} \\
&= -\frac{1}{S(t)} \lim_{\Delta t \rightarrow 0} \frac{S(t + \Delta t) - S(t)}{\Delta t} \\
&= -\frac{1}{S(t)} \frac{d}{dt} S(t) \\
&= -\frac{d}{dt} (\log S(t)).
\end{aligned}$$



$$2. h(t) = \frac{f(t)}{S(t)}.$$

*Demonstração:* Usando o fato de que  $h(t) = -\frac{d}{dt}(\log S(t))$  temos

$$\begin{aligned} h(t) &= -\frac{d}{dt}(\log S(t)) \\ &= -\frac{1}{S(t)} \frac{d}{dt} S(t) \\ &= -\frac{1}{S(t)} \frac{d}{dt} \{1 - F(t)\} \\ &= -\frac{1}{S(t)} (-f(t)) \\ &= \frac{f(t)}{S(t)}. \end{aligned}$$

$$3. H(t) = -\log S(t).$$

*Demonstração:* Usando o fato de que  $h(t) = \frac{f(t)}{S(t)}$  na equação (1.4), obtemos que

$$\begin{aligned} H(t) &= \int_0^t h(x) dx \\ &= \int_0^t \frac{f(x)}{S(x)} dx \\ &= \int_0^t \frac{f(x)}{1 - F(x)} dx, \end{aligned}$$

fazendo  $u = 1 - F(x)$ , temos que  $du = -f(x)dx$  e portanto, podemos reescrever a integral em questão do seguinte modo

$$\begin{aligned} H(t) &= \int_1^{1-F(t)} -\frac{1}{u} du \\ &= -\int_1^{1-F(t)} \frac{1}{u} du \\ &= -\log(u) \Big|_1^{1-F(t)} \\ &= -\log\{1 - F(t)\} + \log(1) \\ &= -\log S(t). \end{aligned}$$

4.  $S(t) = \exp\{-H(t)\}.$

*Demonstração:* Sabemos que  $H(t) = \int_0^t h(u)du$ , uma vez que  $H(t) = -\log S(t)$ , temos que  $-\log S(t) = \int_0^t h(u)du$ , assim  $\log S(t) = -\int_0^t h(u)du$  e consequentemente

$$\begin{aligned} S(t) &= \exp\left\{-\int_0^t h(u)du\right\} \\ &= \exp\{-H(t)\}. \end{aligned}$$

5.  $S(t) = \frac{\text{vmr}(0)}{\text{vmr}(t)} \exp\left\{-\int_0^t \frac{du}{\text{vmr}(u)}\right\}.$

*Demonstração:* Sabemos pela equação (1.5) que  $\frac{1}{\text{vmr}(t)} = \frac{S(t)}{\int_t^\infty S(u)du}$ , então

$$\begin{aligned} \frac{1}{\text{vmr}(t)} &= -\frac{d}{dt} \log\left\{\int_t^\infty S(u)du\right\} \\ &= -\frac{d}{dt} \log\{\text{vmr}(t)S(t)\}, \end{aligned}$$

logo

$$\begin{aligned} \int_0^t \frac{1}{\text{vmr}(u)}du &= \int_0^t -\frac{d}{dt} \log\{\text{vmr}(u)S(u)\} du \\ &= -\log\{\text{vmr}(t)S(t)\} + \log\{\text{vmr}(0)S(0)\} \\ &= -\log\{\text{vmr}(t)S(t)\} + \log\{\text{vmr}(0)\} \\ &= -\log\left\{\frac{\text{vmr}(t)S(t)}{\text{vmr}(0)}\right\}, \end{aligned}$$

daí

$$\exp\left\{-\int_0^t \frac{du}{\text{vmr}(u)}\right\} = \frac{\text{vmr}(t)S(t)}{\text{vmr}(0)}$$

e consequentemente

$$S(t) = \frac{\text{vmr}(0)}{\text{vmr}(t)} \exp\left\{-\int_0^t \frac{du}{\text{vmr}(u)}\right\}.$$

$$6. \ h(t) = \frac{1}{\text{vmr}(t)} \left( \frac{d}{dt} \text{vmr}(t) + 1 \right).$$

*Demonstração:* Sabendo que  $h(t) = -\frac{d}{dt} (\log S(t))$  e usando o fato de que  $S(t) = \frac{\text{vmr}(0)}{\text{vmr}(t)} \exp \left\{ -\int_0^t \frac{du}{\text{vmr}(u)} \right\}$  temos

$$\begin{aligned} h(t) &= -\frac{d}{dt} \left[ \log \left( \frac{\text{vmr}(0)}{\text{vmr}(t)} \right) - \int_0^t \frac{du}{\text{vmr}(u)} \right] \\ &= -\frac{d}{dt} \log\{\text{vmr}(0)\} + \frac{d}{dt} \log\{\text{vmr}(t)\} + \frac{1}{\text{vmr}(t)} \\ &= \frac{1}{\text{vmr}(t)} \frac{d}{dt} \text{vmr}(t) + \frac{1}{\text{vmr}(t)} \\ &= \frac{1}{\text{vmr}(t)} \left( \frac{d}{dt} \text{vmr}(t) + 1 \right). \end{aligned}$$

#### 1.1.4 Modelo de tempo de vida acelerado

Em algumas situações experimentais, o comportamento da variável resposta, pode ser melhor explicado por meio de uma estrutura com covariáveis e parâmetros desconhecidos. Este é o contexto de um modelo de regressão cujo objetivo é estimar o efeito de covariáveis sobre uma variável resposta. Segundo [Colosimo e Giolo \(2006\)](#), em Análise de Sobrevida existem duas grandes classes de modelos: os paramétricos e os semiparamétricos. Os modelos paramétricos são mais eficientes, porém menos flexíveis do que os modelos semiparamétricos, já os modelos semiparamétricos, também denominados de modelo de regressão de Cox, são mais flexíveis e permite incorporar covariáveis dependentes do tempo. O modelo de tempo de vida acelerado é um modelo paramétrico.

Sejam  $T$  a variável resposta,  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^\top$  os coeficientes desconhecidos das  $p$  variáveis explicativas  $\mathbf{x}_i^\top = (1, \mathbf{x}_{1i}, \mathbf{x}_{2i}, \dots, \mathbf{x}_{pi})$  e  $\sigma$  o parâmetro de dispersão. Supondo que o termo de  $\epsilon_i$ , erro aleatório, segue uma certa distribuição, o modelo que pode ser escrito na forma

$$T = \exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\} \exp\{\sigma \epsilon_i\} \quad (1.6)$$

ou, equivalentemente, na forma linear

$$\log(T) = \mathbf{x}_i^\top \boldsymbol{\beta} + \sigma \epsilon_i, \quad (1.7)$$

é denominado modelo de tempo de vida acelerado. Esse modelo é frequentemente usado pois engloba uma ampla gama de distribuições de tempo de vida. As extensões desse modelo podem ser obtidas considerando-se várias distribuições para  $\epsilon_i$ , por exemplo:

- Se  $\epsilon_i$  é Valor extremo ( $T$  é Weibull);
- Se  $\epsilon_i$  é Normal ( $T$  é lognormal);
- Se  $\epsilon_i$  é Log-gama ( $T$  é gama);
- Se  $\epsilon_i$  é Logística ( $T$  é log-logística);
- Se  $\epsilon_i$  é Senh-Normal ( $T$  é Birnbaum-Saunders).

No modelo de tempo de vida acelerado, as covariáveis medidas em um indivíduo atuam multiplicativamente no tempo e tem efeito de acelerar ou desacelerar o tempo de vida, ou seja, o modelo pode ser usado, por exemplo, para interpretar a velocidade de progressão de uma doença, por isso esse modelo é bastante atraente na Análise de Sobrevivência.

De acordo com [Sinha \(2018\)](#) o componente  $\sigma \epsilon_i$ , no modelo linearizado (1.7), corresponde a distribuição de referência quando  $\mathbf{x}_i = \mathbf{0}$  e o componente  $T_0 = \exp\{\sigma \epsilon_i\}$  é a distribuição de referência no modelo com escala original (1.6). Assim, a sobrevivência associada a  $T_0$ , denotada por  $S_0(t)$ , é dada por

$$\begin{aligned} S_0(t) &= P(T_0 > t) \\ &= P(\exp\{\sigma \epsilon_i\} > t) \\ &= P(\sigma \epsilon_i > \log(t)) \\ &= P\left(\epsilon_i > \frac{\log(t)}{\sigma}\right). \end{aligned}$$

Considerando o efeito das covariáveis ( $T = T_0 \exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}$ ), a função de sobrevivência correspondente para um indivíduo no tempo  $t$  é dada por

$$S(t|\mathbf{x}_i) = P(T > t)$$

$$\begin{aligned}
&= P(T_0 \exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\} > t) \\
&= P\left(T_0 > \frac{t}{\exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}}\right) \\
&= S_0\left(\frac{t}{\exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}}\right),
\end{aligned}$$

em que  $S_0(t)$  é função de sobrevivência da distribuição de referência no tempo  $t$  quando todas as  $p$  covariáveis forem iguais a zero. Segue que

$$\begin{aligned}
F(t|\mathbf{x}_i) &= P(T \leq t) \\
&= P(T_0 \exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\} \leq t) \\
&= P\left(T_0 \leq \frac{t}{\exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}}\right) \\
&= F_0\left(\frac{t}{\exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}}\right).
\end{aligned}$$

Daí,

$$\begin{aligned}
f(t|\mathbf{x}_i) &= F'_0\left(\frac{t}{\exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}}\right) \\
&= f_0\left(\frac{t}{\exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}}\right) \frac{1}{\exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}} \\
&= \exp\{-\mathbf{x}_i^\top \boldsymbol{\beta}\} f_0\left(\frac{t}{\exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}}\right).
\end{aligned}$$

Portanto, a função de risco para um determinado indivíduo é dado por

$$\begin{aligned}
h(t|\mathbf{x}_i) &= \frac{f(t|\mathbf{x}_i)}{S(t|\mathbf{x}_i)} \\
&= \frac{\exp\{-\mathbf{x}_i^\top \boldsymbol{\beta}\} f_0\left(\frac{t}{\exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}}\right)}{S_0\left(\frac{t}{\exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}}\right)} \\
&= \exp\{-\mathbf{x}_i^\top \boldsymbol{\beta}\} \frac{f_0\left(\frac{t}{\exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}}\right)}{S_0\left(\frac{t}{\exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}}\right)} \\
&= \exp\{-\mathbf{x}_i^\top \boldsymbol{\beta}\} h_0\left(\frac{t}{\exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}}\right),
\end{aligned}$$

em que  $h_0(t)$  é função de risco da distribuição de referência no tempo  $t$ .

Para a forma log-linear do modelo de tempo de vida acelerado (1.7), supõe-se

que o termo  $\epsilon_i$  do componente aleatório segue uma certa distribuição e é usado para modelar a discrepância entre  $\log(t_i)$  e o componente linear do modelo (Sinha, 2018). As funções de distribuição acumulada e densidade são dadas, respectivamente por

$$\begin{aligned}
F(t|\mathbf{x}_i) &= P(\log(T) \leq \log(t)) \\
&= P(\exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\} \exp\{\sigma \epsilon_i\} \leq t) \\
&= P(\mathbf{x}_i^\top \boldsymbol{\beta} + \sigma \epsilon_i \leq \log(t)) \\
&= P\left(\epsilon_i \leq \frac{\log(t) - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma}\right) \\
&= F_{\epsilon_i}\left(\frac{\log(t) - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma}\right)
\end{aligned}$$

e

$$\begin{aligned}
f(t|\mathbf{x}_i) &= \left[ F_{\epsilon_i}\left(\frac{\log(t) - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma}\right) \right]' \\
&= f_{\epsilon_i}\left(\frac{\log(t) - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma}\right) \frac{1}{\sigma} \frac{1}{t} \\
&= \frac{1}{\sigma t} f_{\epsilon_i}\left(\frac{\log(t) - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma}\right).
\end{aligned}$$

A função de sobrevivência  $S(t|\mathbf{x}_i)$  pode ser obtida a partir da função de sobrevivência  $S_{\epsilon_i}(t)$ , do erro  $\epsilon_i$ , uma vez que

$$\begin{aligned}
S(t|\mathbf{x}_i) &= P(\log(T) > \log(t)) \\
&= P(\exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\} \exp\{\sigma \epsilon_i\} > t) \\
&= P(\mathbf{x}_i^\top \boldsymbol{\beta} + \sigma \epsilon_i > \log(t)) \\
&= P\left(\epsilon_i > \frac{\log(t) - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma}\right) \\
&= S_{\epsilon_i}\left(\frac{\log(t) - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma}\right).
\end{aligned}$$

Similarmente, a função de risco,  $h(t|\mathbf{x}_i)$ , do indivíduo no momento  $t$  pode ser obtida a partir da função de risco  $h_{\epsilon_i}(t)$  de  $\epsilon_i$  por

$$h(t|\mathbf{x}_i) = \frac{f(t|\mathbf{x}_i)}{S(t|\mathbf{x}_i)}$$

$$\begin{aligned}
&= \frac{\frac{1}{\sigma t} f_{\epsilon_i} \left( \frac{\log(t) - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma} \right)}{S_{\epsilon_i} \left( \frac{\log(t) - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma} \right)} \\
&= \frac{1}{\sigma t} h_{\epsilon_i} \left( \frac{\log(t) - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma} \right).
\end{aligned}$$

Por exemplo, se os tempos de falha ou censura,  $T$ , seguem uma distribuição BS, então o erro  $\epsilon_i$  segue distribuição senh-normal na qual  $h_{\epsilon_i}(\epsilon_i) = \frac{\frac{1}{2} \xi_{i1} \phi(\xi_{i2})}{1 - \Phi(\xi_{i2})}$  e  $S_{\epsilon_i}(\epsilon_i) = 1 - \Phi(\xi_{i2})$  em que  $\xi_{i1} = \frac{2}{\alpha} \cosh\left(\frac{\epsilon_i}{2}\right)$ ,  $\xi_{i2} = \frac{2}{\alpha} \sinh\left(\frac{\epsilon_i}{2}\right)$  e  $\epsilon_i = \log(t_i) - \mathbf{x}_i^\top \boldsymbol{\beta}$ . Aqui nosso objetivo será estimar os parâmetros do modelo,  $\boldsymbol{\beta}$  e  $\alpha$ , por um método robusto adequado.

## 1.2 Robustez

A expressão “Robustez” será utilizada em um sentido relativamente restrito conforme [Huber \(1996\)](#), no qual significa que um determinado método de análise estatística é não sensível a pequenas violações (ou desvios) das suposições. A situação mais comum refere-se ao afastamento da forma da distribuição de probabilidade postulada, mas também pode está associada a outros tipos de exigências ou suposições. Uma forma relativamente típica de violação da distribuição de probabilidade assumida é a contaminação da amostra com valores extremos (*outliers*). De acordo com [Barnett e Lewis \(1978\)](#), *outliers* são observações inconsistentes com o restante de um conjunto de dados, ou seja, observações atípicas, aquelas “bem separadas” da maioria e que são muito comuns de se observar em dados reais. Dependendo da quantidade de *outliers* podemos ter um alongamento nas caudas da distribuição que consequentemente acabam inflacionando as estimativas dos parâmetros do modelo proposto.

Embora os métodos robustos tenham começado a surgir na década de 60, com o objetivo de minimizar os impactos causados por valores extremos nas estimativas dos parâmetros, eles ainda são muito negligenciados. É comum, por exemplo, a decisão de eliminar parte das observações, isto é, excluir os valores atípicos e, em seguida, aplicar procedimentos inferenciais clássicos aos dados restantes. No entanto, essa prática pode ser falha em muitos aspectos, principalmente em configurações que *outliers* possam ser mascarados e muito difíceis de detectar. Seria enganoso pensar sempre em *outliers* como

dados “ruins”, eles podem conter informações relevantes e como dito por [Farcomeni e Ventura \(2012\)](#), em muitos casos é mais eficiente reduzir o peso dessas observações no processo de estimação em vez de descartá-las.

Atualmente, o desenvolvimento de técnicas de estimação robusta tem crescido rapidamente nas mais diversas áreas do estudo. Estamos cada vez mais cientes que os modelos e muitas suposições comumente feitas são, na melhor das hipóteses, aproximações da realidade e as estatísticas devem estar “em algum sentido” preparadas para desvios moderados das premissas do modelo. O objetivo, portanto, é encontrar estimativas que sejam “quase ideais” quando as premissas são exatas e também “quase ótimas” quando as premissas são próximas da realidade. [Farcomeni e Ventura \(2012\)](#) comentam o fato dos métodos robustos serem considerados uma boa escolha na presença de dados com valores extremos, pois as observações podem ser calibradas para ter uma pequena perda de eficiência em relação as técnicas paramétricas clássicas, porém conseguindo ser mais resistentes a algumas violações de suposições. A estatística robusta segundo [Heritier et al. \(2009\)](#) busca produzir estimadores que possam ser considerados consistentes e razoavelmente eficientes produzindo estatísticas de teste com níveis estáveis quando o modelo não é bem especificado.

### 1.2.1 Estatísticas robustas

Estimação é um processo muito utilizado na inferência estatística com o objetivo de atribuir um valor a um parâmetro, para o qual não se conhece o valor absoluto. Este processo dar-se a partir de uma amostra de indivíduos de uma determinada população e o estimador é uma estatística da amostra utilizada para estimar tal parâmetro. Estatísticas clássicas como a média, variância, covariância, correlação, ajuste de mínimos quadrados, entre outras, podem ser influenciadas por *outliers* e muitas vezes não fornecem um bom ajuste para a maior parte dos dados. Nesse contexto surgem as estimativas robustas que fornecem um bom ajuste para a maior parte dos dados quando estes contêm *outliers*, bem como para os dados sem *outliers*.

A seguir, veremos um exemplo apresentado por [Maronna et al. \(2018\)](#) de estimativas clássicas e robustas para a média e o desvio padrão de um banco de dados contendo *outliers*. Através desse exemplo podemos observar o quanto de influência um



único *outlier* pode ter sobre estas estimativas clássicas.

**Exemplo 1.1** Considere as seguintes 24 determinações do teor de cobre na farinha integral (em partes por milhão), classificadas em ordem crescente ([Committee et al., 1989](#)):

2, 20   2, 20   2, 40   2, 40   2, 50   2, 70   2, 80   2, 90  
 3, 03   3, 03   3, 10   3, 37   3, 40   3, 40   3, 40   3, 50  
 3, 60   3, 70   3, 70   3, 70   3, 70   3, 77   5, 28   28, 95

O valor 28, 95 imediatamente se destaca do resto dos valores e seria considerado um *outlier* por quase todo mundo. Sabe-se que para um conjunto de valores observados, os estimadores clássicos para a média da amostra e para o desvio padrão da amostra (DP) são definidos por

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{e} \quad \text{DP} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2}.$$

A média da amostra é apenas a média aritmética das observações e, como tal, pode-se esperar que ela forneça uma boa estimativa do centro ou da localização dos dados. Da mesma forma, espera-se que o DP da amostra forneça uma boa estimativa da dispersão dos dados.

Conforme a Tabela 1.1 temos que os valores da média e do DP para o conjunto de dados acima é  $\bar{x} = 4,28$  e  $\text{DP} = 5,30$ . Note que  $\bar{x} = 4,28$  é maior que todos os valores de dados, com exceção de dois (5,25 e 28,95), e não está entre a maior parte das observações, ou seja, não representa uma boa estimativa do centro dos dados. Ao excluir o valor 28, 95, os valores da média e do DP são alterados para  $\bar{x} = 3,21$  e  $\text{DP} = 0,69$ . Temos agora uma média que fornece uma boa estimativa do centro dos dados e um DP sete vezes menor do que era com a presença do *outlier*.

Suponha, por exemplo, que o valor 28, 95 seja substituído por um valor arbitrário  $x_{24}$ , a partir da definição de média da amostra fica claro que variando  $x_{24}$  de  $-\infty$  para  $+\infty$ , o valor da média muda de  $-\infty$  para  $+\infty$ . É fácil verificar também pela definição de DP que variar  $x_{24}$  de  $-\infty$  para  $+\infty$  leva o mesmo a variar de um valor positivo menor do que aquele baseado nas primeiras 23 observações para  $+\infty$ . Assim, podemos

Tabela 1.1: Estimativas clássicas e robustas para o teor de cobre na farinha integral para diferentes quantidades de observações.

	As 23 observações regulares	Todas as 24 observações
$\bar{x}_n$	3,21	4,28
Med	3,37	3,38
DP	0,69	5,30
DAMN	0,50	0,53

dizer que um único *outlier* tem uma influência ilimitada nessas estatísticas clássicas.

Embora esse exemplo sugira descartar *outliers*, já comentamos sobre possíveis problemas em fazer uso desta prática. Podemos, portanto, pensar em melhores possibilidades.

Um método muito antigo também usado para estimar o “meio” dos dados é a mediana da amostra definida por

$$\text{Med}(X_n) = \begin{cases} x_{(\frac{n+1}{2})} & \text{se } n \text{ é ímpar} \\ \frac{1}{2} \left( x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right) & \text{se } n \text{ é par,} \end{cases}$$

em que  $x_{(1)} \leq x_{(2)} \leq \dots x_{(n)}$  são as observações ordenadas.

Nesse exemplo, conforme a Tabela 1.1, a mediana da amostra total é de 3,38, enquanto a mediana sem o maior valor é de 3,37, mostrando que a mediana não é muito afetada pela presença desse valor. Supondo novamente que o valor 28,95 seja substituído por um valor arbitrário  $x_{24}$ , fica claro a partir da definição da mediana que, quando  $x_{24}$  varia de  $-\infty$  a  $+\infty$ , o valor da mediana da amostra não muda como no caso da média. Em vez disso, sofre apenas uma pequena alteração. Observe que, para essa amostra, o valor da mediana com o *outlier* é relativamente próximo do valor médio da amostra de 3,21 com o *outlier* excluído. Portanto, a mediana estima melhor o centro dos dados com ou sem a presença de *outliers* e não é muito influenciada por observações atípicas, ou seja, a mediana é uma boa alternativa robusta para a média.

Da mesma forma, uma alternativa robusta ao DP é o desvio absoluto mediano da mediana (DAM), definido por

$$\text{DAM}(\mathbf{x}) = \text{Med}\{|\mathbf{x} - \text{Med}(\mathbf{x})|\},$$

este estimador usa a mediana da amostra duas vezes, primeiro para obter uma estimativa do centro dos dados, a fim de formar o conjunto de resíduos absolutos sobre a mediana da amostra,  $\{|\mathbf{x} - \text{Med}(\mathbf{x})|\}$ , e depois para calcular a mediana da amostra desses resíduos absolutos. Para tornar o DAM comparável ao DP, definimos o DAM normalizado (“DAMN”) por

$$\text{DAMN}(\mathbf{x}) = \frac{\text{DAM}(\mathbf{x})}{0,6745},$$

em que a razão para essa definição é que 0,6745 é o DAM de uma variável aleatória normal padrão e, portanto, uma variável  $N(\mu, \sigma^2)$  possui  $\text{DAMN} = \sigma$ .

Neste exemplo, de acordo com a Tabela 1.1 obtemos  $\text{DAMN} = 0,53$ , em comparação com  $\text{DP} = 5,30$ . A exclusão do *outlier* gera  $\text{DAMN} = 0,50$ , em comparação com o DP da amostra que é um pouco maior de 0,69. O DAM claramente não é muito influenciado pela presença de um *outlier* e, como tal, fornece uma boa alternativa robusta ao DP da amostra.

Portanto, os estimadores robustos são menos influenciados pelo *outlier* e a estimativa robusta calculada a partir de todas as observações é comparável a estimativa clássica aplicada aos dados sem o *outlier*.

Surge portanto a pergunta: Porque não usar sempre a mediana e o desvio absoluto mediano? Uma explicação informal dada por Maronna et al. (2018) é que, se os dados não contiverem *outliers*, estas estimativas terão um desempenho estatístico mais fraco do que o das estimativas clássicas. As estimativas clássicas são, em certo sentido, “ótimas” quando os dados são exatamente distribuídos de acordo com o modelo assumido, mas podem ser muito abaixo do ideal quando a distribuição dos dados difere do modelo assumido por uma “pequena” quantidade. Estimativas robustas, por outro lado, mantêm um desempenho aproximadamente ótimo, não apenas sob o modelo assumido, mas também sob pequenas perturbações.

*Outliers* podem causar problemas no ajuste de várias distribuições paramétricas, como por exemplo, nas distribuições Normal, Exponencial, Weibull, Gama ou Birnbaum-Saunders, onde a abordagem clássica é usar o estimador de máxima verossimilhança (EMV) que muitas vezes fornecem estimativas não robustas. Nesses casos, é necessário ter alternativas robustas ao EMV, a fim de obter um bom ajuste para a

maior parte dos dados.

### 1.2.2 M-Estimadores

De acordo [Heritier et al. \(2009\)](#), um estimador é frequentemente escolhido como um membro de uma classe geral de estimadores que é ideal em algum sentido ou preenche um conjunto de boas propriedades. Nessa seção, apresentamos a classe de M-estimadores, porém, não apresentaremos todos os M-estimadores robustos disponíveis na literatura, todavia nos concentraremos apenas no estimador de Huber que será utilizado ao longo deste trabalho.

Suponha que temos  $p$  grupos de observações  $\{(\mathbf{x}_{1i}, \dots, \mathbf{x}_{pi}); i = 1, \dots, n\}$  independentes e identicamente distribuídas (i.i.d.) de um modelo comum com função de distribuição  $F(\mathbf{x}_i|\boldsymbol{\theta})$  e densidade correspondente  $f(\mathbf{x}_i|\boldsymbol{\theta})$ . Um M-estimador de  $\boldsymbol{\theta}$  é dado pela solução do problema de minimização dado por

$$\hat{\boldsymbol{\theta}} = \min_{\boldsymbol{\theta}} \sum_{i=1}^n \rho(\mathbf{x}_i|\boldsymbol{\theta}), \quad (1.8)$$

ou alternativamente, pela solução para  $\boldsymbol{\theta}$  de

$$\sum_{i=1}^n \psi(\mathbf{x}_i|\boldsymbol{\theta}) = \mathbf{0}, \quad (1.9)$$

em que  $\rho$  e  $\psi$  são funções adequadas e  $\psi = \rho'$ . Observe que os M-estimadores podem ser vistos como uma generalização da estimação por máxima verosimilhança, pois se configurarmos  $\rho(\mathbf{x}_i|\boldsymbol{\theta}) = -\log(f(\mathbf{x}_i|\boldsymbol{\theta}))$  em (1.8) obtemos o EMV de  $\boldsymbol{\theta}$  o qual seria “ótimo” no sentido de atingir a menor variância assintótica possível se soubéssemos exatamente a distribuição  $F(\mathbf{x}_i|\boldsymbol{\theta})$ , mas sabemos que na maioria das vezes essa distribuição é apenas aproximada, neste caso, faz-se necessário optar por um M-estimador robusto que é obtido através de uma escolha oportuna para  $\psi(\cdot)$  em (1.9). Basta então escolher uma função  $\psi$  limitada, pois se for ilimitada em seu argumento  $\mathbf{x}_i$ , o estimador correspondente não é robusto para o parâmetro de interesse. O objetivo da utilização da função  $\rho(\mathbf{x}_i|\boldsymbol{\theta})$  em (1.8) consiste em limitar no processo de estimação a influência dos resíduos elevados do modelo. Os conhecidos estimadores robustos de locação e escala, a mediana e o desvio absoluto mediano, respectivamente, pertencem à classe

dos M-estimadores.

Os M-estimadores incluem o chamado estimador robusto de máxima verossimilhança (ERM), visto como uma média ponderada da função escore e definido como a solução para  $\boldsymbol{\theta}$  de

$$\sum_{i=1}^n s(\mathbf{x}_i|\boldsymbol{\theta})w(\mathbf{x}_i|\boldsymbol{\theta}) - a_{\boldsymbol{\theta}} = \mathbf{0}, \quad (1.10)$$

em que  $s(\mathbf{x}_i|\boldsymbol{\theta})$  é a função escore,  $w(\mathbf{x}_i|\boldsymbol{\theta})$  é uma função de ponderação e  $a_{\boldsymbol{\theta}}$  é um fator de correção de consistência, isto é, uma constante de sintonia na qual possui a função de retirar o viés introduzido pela robustez. Observe que (1.10) será correspondente ao EMV quando a função de ponderação  $w(\mathbf{x}_i|\boldsymbol{\theta}) = 1$  para todo  $i$  e, conseqüentemente,  $a_{\boldsymbol{\theta}} = 0$ . Para construir um ERM, basta escolher pesos que façam  $s(\mathbf{x}_i|\boldsymbol{\theta})w(\mathbf{x}_i|\boldsymbol{\theta})$  em (1.10) ser limitado. Os pesos podem depender apenas das observações, por exemplo, uma abordagem clássica é escolher a função de peso em função da distância robusta de Mahalanobis, que é expressa por

$$d(\mathbf{x}_i) = \sqrt{(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^\top \hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}_i - \hat{\boldsymbol{\mu}})},$$

em que  $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$  são estimativas robustas de locação e dispersão multivariada ([Rousseeuw e Van Zomeren, 1990](#)). Uma escolha popular são os pesos de Huber representados por

$$w_c(\mathbf{x}_i) = \min \left\{ 1, \frac{c}{d(\mathbf{x}_i)} \right\}, \quad (1.11)$$

ou seja, o peso é igual a um para todos os valores (pequenos) de  $\mathbf{x}_i$  satisfazendo  $d(\mathbf{x}_i) < c$  e é igual a  $c/d(\mathbf{x}_i)$  caso contrário, em que  $c$  é uma constante de afinação e seu valor é escolhido para assegurar uma dada variância assintótica.

A função  $\psi$  de Huber é definida por

$$\psi_c = \max(-c, \min(\mathbf{x}_i, c)),$$

com a função  $\rho$  correspondente

$$\rho_c = \begin{cases} \frac{1}{2}\mathbf{x}_i^2 & \text{se } |\mathbf{x}_i| \leq c, \\ c|\mathbf{x}_i| - \frac{1}{2}c^2 & \text{se } |\mathbf{x}_i| > c. \end{cases}$$

Os pesos de Huber podem ser usados em diferentes modelos, contanto que o argumento  $\mathbf{x}_i$  da função de peso (1.11), seja uma observação, um resíduo ou outra quantidade, tenha um valor desproporcional, pois se os pesos forem iguais a um nenhuma redução é realizada. A observação terá o peso reduzido apenas se o argumento exceder algum valor limite  $c$ , o qual é escolhido com base em argumentos de robustez, quanto menor o  $c$ , menor o peso e mais robusto será o estimador, quanto maior o  $c$ , mais as observações deixarão de ser ponderadas e menos robusto será o estimador.

### 1.2.3 Estimação robusta

Nesta seção apresentaremos como foi desenvolvida a construção do método de estimação robusta proposto por [Sinha \(2018\)](#).

Primeiro, descreveremos a estimação dos parâmetros do modelo (1.7) pelo método padrão de máxima verossimilhança, assumindo que os tempos de sobrevivência seguem uma certa distribuição de probabilidade. Segundo [Colosimo e Giolo \(2006\)](#), para um conjunto de dados representados por  $\{(t_i, \delta_i, \mathbf{x}_i); i = 1, \dots, n\}$ , a função de verossimilhança para o vetor de parâmetros  $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma)^\top$  é dada por

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n [h(t_i|\mathbf{x}_i)]^{\delta_i} S(t_i|\mathbf{x}_i).$$

O logaritmo da função de verossimilhança correspondente fica dado por

$$\begin{aligned} l(\boldsymbol{\theta}) &= \sum_{i=1}^n [\delta_i \log\{h(t_i|\mathbf{x}_i)\} + \log\{S(t_i|\mathbf{x}_i)\}] \\ &= \sum_{i=1}^n \left[ \delta_i \log \left\{ \frac{1}{\sigma t_i} h_{\epsilon_i} \left( \frac{\log(t_i) - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma} \right) \right\} + \log \left\{ S_{\epsilon_i} \left( \frac{\log(t_i) - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma} \right) \right\} \right] \\ &= \sum_{i=1}^n [\delta_i \{-\log(\sigma t_i) + \log\{h_{\epsilon_i}(\epsilon_i)\}\} + \log\{S_{\epsilon_i}(\epsilon_i)\}], \end{aligned}$$

em que  $\epsilon_i = (\log(t_i) - \mathbf{x}_i^\top \boldsymbol{\beta})/\sigma$ .

As funções escores de  $\beta$  e  $\sigma$  são dadas, respectivamente, por

$$\begin{aligned}\frac{\partial l(\boldsymbol{\theta})}{\partial \beta} &= \sum_{i=1}^n \left[ \delta_i \frac{\partial \log\{h_{\epsilon_i}(\epsilon_i)\}}{\partial \epsilon_i} + \frac{\partial \log\{S_{\epsilon_i}(\epsilon_i)\}}{\partial \epsilon_i} \right] \frac{\partial \epsilon_i}{\partial \beta}, \\ \frac{\partial l(\boldsymbol{\theta})}{\partial \sigma} &= \sum_{i=1}^n \left[ \delta_i \frac{\partial \log\{h_{\epsilon_i}(\epsilon_i)\}}{\partial \epsilon_i} + \frac{\partial \log\{S_{\epsilon_i}(\epsilon_i)\}}{\partial \epsilon_i} \right] \frac{\partial \epsilon_i}{\partial \sigma} - \sum_{i=1}^n \frac{\delta_i}{\sigma}.\end{aligned}$$

que levam as seguintes equações de verossimilhança

$$\begin{cases} \sum_{i=1}^n \left[ \delta_i \frac{\partial \log\{h_{\epsilon_i}(\epsilon_i)\}}{\partial \epsilon_i} + \frac{\partial \log\{S_{\epsilon_i}(\epsilon_i)\}}{\partial \epsilon_i} \right] \mathbf{x}_i = \mathbf{0}, \\ \sum_{i=1}^n \left[ \epsilon_i \left\{ \delta_i \frac{\partial \log\{h_{\epsilon_i}(\epsilon_i)\}}{\partial \epsilon_i} + \frac{\partial \log\{S_{\epsilon_i}(\epsilon_i)\}}{\partial \epsilon_i} \right\} + \delta_i \right] = 0. \end{cases} \quad (1.12)$$

O sistema de equações (1.12) possui solução numérica em termos de  $\beta$  e  $\sigma$  e correspondem as estimativas de máxima verossimilhança dos respectivos parâmetros. Como o vetor escore é uma função do erro  $\epsilon_i$  e das covariáveis  $\mathbf{x}_i$ , que são ilimitadas para o tempo de sobrevivência  $t_i$  e covariáveis  $\mathbf{x}_i$ , o EMV é, geralmente, não robustos na presença de *outliers* nos tempos de sobrevivência e/ou covariáveis. Como solução, usaremos um procedimento robusto (Sinha, 2018) usando uma função  $\psi$  monotônica do tipo Huber (Huber et al., 1973) em  $\epsilon_i$  e uma função de peso  $w(\mathbf{x}_i)$  para limitar a influência de *outliers* nos tempos de sobrevivência. Especificamente, consideramos encontrar estimadores robustos do modelo de tempo de falha acelerado resolvendo as equações de estimação

$$\begin{cases} \sum_{i=1}^n \left[ \left\{ \delta_i \frac{\partial \log\{h_{\epsilon_i}(z_i)\}}{\partial z_i} + \frac{\partial \log\{S_{\epsilon_i}(z_i)\}}{\partial z_i} \right\} w(\mathbf{x}_i) \mathbf{x}_i - \mathbf{a}_\beta \right] = \mathbf{0}, \\ \sum_{i=1}^n \left[ z_i \left\{ \delta_i \frac{\partial \log\{h_{\epsilon_i}(z_i)\}}{\partial z_i} + \frac{\partial \log\{S_{\epsilon_i}(z_i)\}}{\partial z_i} \right\} + \delta_i - a_\sigma \right] = 0, \end{cases} \quad (1.13)$$

em que  $z_i = E(\epsilon_i) + \sqrt{Var(\epsilon_i)}\psi_c(r_i)$  com  $r_i = (\epsilon_i - E(\epsilon_i))/\sqrt{Var(\epsilon_i)}$  e  $\psi_c$  é a função  $\psi$  de Huber:  $\psi_c(r_i) = \max(-c, \min(r_i, c))$ , dependendo da constante de afinação  $c$ , que é tipicamente escolhida para fornecer um certo nível de eficiência assintótica na distribuição subjacente. A função de peso  $w(\mathbf{x})$  é escolhida em função da distância de

Mahalanobis  $d(\mathbf{x}_i) = d(\mathbf{x}_i, \hat{\mu}, \hat{\Sigma}) = \sqrt{(\mathbf{x}_i - \hat{\mu})^\top \hat{\Sigma}^{-1}(\mathbf{x}_i - \hat{\mu})}$  na forma

$$w_c(\mathbf{x}_i) = \min \left\{ 1, \frac{c}{d(\mathbf{x}_i)} \right\},$$

com  $\hat{\mu}$  e  $\hat{\Sigma}$  sendo estimadores robustos de locação e escala de  $\mathbf{x}$ , como por exemplo, os estimadores de elipsóide de volume mínimo, proposto por [Rousseeuw e Van Zomeren \(1990\)](#). As constantes de sintonia  $\mathbf{a}_\beta$  e  $a_\sigma$  são escolhidas de modo a fornecer estimativas não-viesadas para os parâmetros do modelo  $\beta$  e  $\sigma$ , que são dadas por

$$\begin{aligned} \mathbf{a}_\beta &= \frac{1}{n} \sum_{i=1}^n \delta_i E \left\{ \frac{\partial \log\{h_{\epsilon_i}(z_i)\}}{\partial z_i} + \frac{\partial \log\{S_{\epsilon_i}(z_i)\}}{\partial z_i} \right\} w(\mathbf{x}_i) \mathbf{x}_i \\ a_\sigma &= \frac{1}{n} \sum_{i=1}^n \delta_i E \left[ z_i \left\{ \frac{\partial \log\{h_{\epsilon_i}(z_i)\}}{\partial z_i} + \frac{\partial \log\{S_{\epsilon_i}(z_i)\}}{\partial z_i} \right\} + 1 \right], \end{aligned}$$

em que as esperanças são tomadas com relação à distribuição do erro aleatório  $\epsilon_i$ . É importante destacar que quando o  $c \rightarrow \infty$  na função  $\psi_c$  de Huber, temos que  $w(\mathbf{x}) = 1$  o que implica que as constantes de sintonia serão iguais a zero e, conseqüentemente, os estimadores robustos tornam-se os estimadores de máxima verossimilhança.

O sistema de equações (1.13) podem ser resolvidas numericamente para obter os estimadores robustos usando um método iterativo como, por exemplo, o método de Newton-Raphson ou escore de Fisher (veja mais em [Paula \(2013\)](#)). Observe que a função de risco  $h_{\epsilon_i}(\epsilon_i)$  e a função de sobrevivência  $S_{\epsilon_i}(\epsilon_i)$  do erro  $\epsilon_i$  assumem diferentes formas, dependendo da distribuição e tempo de sobrevivência  $t_i$ .

## Distribuição Assintótica

Aqui apresentaremos as propriedades assintóticas dadas em [Sinha \(2018\)](#) do ERM proposto para o modelo de tempo de falha acelerado. Lembre-se de que os estimadores robustos,  $\hat{\theta} = (\hat{\beta}^\top, \hat{\sigma})^\top$  de  $\theta = (\beta^\top, \sigma)^\top$ , são obtidos resolvendo o sistema de equações (1.13), que podem ser reescritas na forma

$$\Psi_n(\mathbf{t}, \theta) = \frac{1}{n} \sum_{i=1}^n \varphi(t_i, \theta) = \mathbf{0},$$



onde  $\boldsymbol{\varphi}(t_i, \boldsymbol{\beta}) = \{\varphi_\sigma(t_i, \boldsymbol{\theta})\}^\top$  com

$$\begin{aligned}\boldsymbol{\varphi}'_\beta(t_i, \boldsymbol{\theta}) &= \left\{ \delta_i \frac{\partial \log\{h_{\epsilon_i}(z_i)\}}{\partial z_i} + \frac{\partial \log\{S_{\epsilon_i}(z_i)\}}{\partial z_i} \right\} w(\mathbf{x}_i) \mathbf{x}_i - \mathbf{a}_\beta \\ \varphi_\sigma(t_i, \boldsymbol{\theta}) &= z_i \left\{ \delta_i \frac{\partial \log\{h_{\epsilon_i}(z_i)\}}{\partial z_i} + \frac{\partial \log\{S_{\epsilon_i}(z_i)\}}{\partial z_i} \right\} + \delta_i - a_\sigma.\end{aligned}$$

Considere que o valor verdadeiro  $\boldsymbol{\theta}_0$  de  $\boldsymbol{\theta}$  é obtido resolvendo as equações

$$\bar{\boldsymbol{\Psi}}_n(\boldsymbol{\theta}) = E[\boldsymbol{\Psi}_n(\mathbf{t}, \boldsymbol{\theta})] = \mathbf{0},$$

em que a esperança é tomada em relação à distribuição dos tempos de sobrevivência.

Usando o Teorema do Valor Médio, podemos escrever

$$\begin{aligned}\boldsymbol{\Psi}'_n(\mathbf{t}, \tilde{\boldsymbol{\theta}}) &= \frac{\boldsymbol{\Psi}_n(\mathbf{t}, \hat{\boldsymbol{\theta}}) - \boldsymbol{\Psi}_n(\mathbf{t}, \boldsymbol{\theta}_0)}{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0} \\ \boldsymbol{\Psi}'_n(\mathbf{t}, \tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) &= \boldsymbol{\Psi}_n(\mathbf{t}, \hat{\boldsymbol{\theta}}) - \boldsymbol{\Psi}_n(\mathbf{t}, \boldsymbol{\theta}_0) \\ \boldsymbol{\Psi}_n(\mathbf{t}, \hat{\boldsymbol{\theta}}) &= \boldsymbol{\Psi}_n(\mathbf{t}, \boldsymbol{\theta}_0) + \boldsymbol{\Psi}'_n(\mathbf{t}, \tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0),\end{aligned}\tag{1.14}$$

em que  $\tilde{\boldsymbol{\theta}}$  está entre  $\hat{\boldsymbol{\theta}}$  e  $\boldsymbol{\theta}_0$ , e  $\boldsymbol{\Psi}'_n(\mathbf{t}, \tilde{\boldsymbol{\theta}})$  é a primeira derivada de  $\boldsymbol{\Psi}_n(\mathbf{t}, \boldsymbol{\theta})$  em relação a  $\boldsymbol{\theta}$  avaliada em  $\tilde{\boldsymbol{\theta}}$ . Como  $\boldsymbol{\Psi}_n(\mathbf{t}, \boldsymbol{\theta})$  é zero em  $\hat{\boldsymbol{\theta}}$ , a equação (1.14) fica

$$\begin{aligned}0 &= \boldsymbol{\Psi}_n(\mathbf{t}, \boldsymbol{\theta}_0) + \boldsymbol{\Psi}'_n(\mathbf{t}, \tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ \boldsymbol{\Psi}_n(\mathbf{t}, \boldsymbol{\theta}_0) &= -\boldsymbol{\Psi}'_n(\mathbf{t}, \tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) &= \frac{\boldsymbol{\Psi}_n(\mathbf{t}, \boldsymbol{\theta}_0)}{-\boldsymbol{\Psi}'_n(\mathbf{t}, \tilde{\boldsymbol{\theta}})} \\ \sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) &= \{-\boldsymbol{\Psi}'_n(\mathbf{t}, \tilde{\boldsymbol{\theta}})\}^{-1} \{\sqrt{n}\boldsymbol{\Psi}_n(\mathbf{t}, \boldsymbol{\theta}_0)\}.\end{aligned}\tag{1.15}$$

A normalidade assintótica de  $\hat{\boldsymbol{\theta}}$  é assegurada se a matriz  $\boldsymbol{\Psi}'_n(\mathbf{t}, \tilde{\boldsymbol{\theta}})$  estiver adequadamente comportada, isto é, se convergir, e se  $\sqrt{n}\boldsymbol{\Psi}_n(\mathbf{t}, \boldsymbol{\theta}_0)$  tiver a propriedade do limite central.

Sob condições de regularidade (ver [Sinha \(2004\)](#)), pode ser demonstrado que  $\hat{\boldsymbol{\theta}} \rightarrow \boldsymbol{\theta}_0$  e  $\|\boldsymbol{\Psi}'_n(\mathbf{t}, \tilde{\boldsymbol{\theta}}) - \bar{\boldsymbol{\Psi}}'_n(\boldsymbol{\theta}_0)\| \rightarrow 0$  quando  $n \rightarrow \infty$ . Além disso,  $\sqrt{n}\boldsymbol{\Psi}_n(\mathbf{t}, \boldsymbol{\theta}_0)$  é assintoticamente  $N(\mathbf{0}, \mathbf{Q}_n(\boldsymbol{\theta}_0))$ , em que  $\mathbf{Q}_n(\boldsymbol{\theta}_0) = \text{var}(\sqrt{n}\boldsymbol{\Psi}_n(\mathbf{t}, \boldsymbol{\theta}_0))$ .

Note que

$$\begin{aligned}
\mathbf{Q}_n(\boldsymbol{\theta}_0) &= \text{var}(\sqrt{n}\boldsymbol{\Psi}_n(\mathbf{t}, \boldsymbol{\theta}_0)) \\
&= n \text{var}(\boldsymbol{\Psi}_n(\mathbf{t}, \boldsymbol{\theta}_0)) \\
&= n E[\boldsymbol{\Psi}_n(\mathbf{t}, \boldsymbol{\theta})\boldsymbol{\Psi}_n(\mathbf{t}, \boldsymbol{\theta})^\top],
\end{aligned}$$

uma vez que  $E[\boldsymbol{\Psi}_n(\mathbf{t}, \boldsymbol{\theta})] = \mathbf{0}$ . Como  $\frac{1}{n} \sum_{i=1}^n \boldsymbol{\Psi}_n(\mathbf{t}, \boldsymbol{\theta})\boldsymbol{\Psi}_n(\mathbf{t}, \boldsymbol{\theta})^\top \xrightarrow{p} E[\boldsymbol{\Psi}_n(\mathbf{t}, \boldsymbol{\theta})\boldsymbol{\Psi}_n(\mathbf{t}, \boldsymbol{\theta})^\top]$ , podemos estimar  $\mathbf{Q}_n(\boldsymbol{\theta}_0)$  por

$$\sum_{i=1}^n \boldsymbol{\Psi}_n(\mathbf{t}, \boldsymbol{\theta})\boldsymbol{\Psi}_n(\mathbf{t}, \boldsymbol{\theta})^\top.$$

Então, a partir de (1.15), podemos afirmar que

$$\frac{\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\mathbf{M}_n(\boldsymbol{\theta}_0)}{\sqrt{\mathbf{Q}_n(\boldsymbol{\theta}_0)}} \rightsquigarrow N(\mathbf{0}, \mathbf{I}),$$

em que  $\mathbf{M}_n(\boldsymbol{\theta}_0) = -E[\boldsymbol{\Psi}'_n(\mathbf{t}, \boldsymbol{\theta})]$ . A matriz de covariância assintótica do ERM,  $\hat{\boldsymbol{\theta}}$ , é obtida a partir da função de variância e covariância do tipo sanduíche dada por

$$\mathbf{V}_n(\boldsymbol{\theta}_0) = \mathbf{M}_n^{-1}(\boldsymbol{\theta}_0)\mathbf{Q}_n(\boldsymbol{\theta}_0)\{\mathbf{M}_n^{-1}(\boldsymbol{\theta}_0)\}^\top. \quad (1.16)$$

Uma estimativa da matriz de covariância acima pode ser obtida avaliando  $\mathbf{V}_n(\boldsymbol{\theta}_0)$  nas estimativas robustas de  $\hat{\boldsymbol{\theta}}$ .

# Capítulo 2

## Distribuição Birnbaum-Saunders

O desgaste de um certo tipo de material ou equipamento que ocorre após um uso contínuo ou exposição a situações de estresse e tensão, é chamado de fadiga. Um dos interesses na Análise de Sobrevivência, por exemplo, é avaliar a durabilidade de um material ou equipamento que está sob situações que levam a fadiga. Geralmente, a maioria das distribuições bidimensionais paramétrica se ajustam razoavelmente bem a esses dados, alguns exemplos são as distribuições gama, lognormal e Weibull, pois apresentam um bom ajuste na região central da distribuição. Entretanto, esses modelos não se adequam muito bem aos percentis mais extremos, regiões onde geralmente possuem poucos dados. Diante disso, [Birnbaum e Saunders \(1969\)](#) visando encontrar uma alternativa para solucionar estes problemas, construíram através de um determinado tipo de processo de fadiga, uma nova distribuição chamada de distribuição Birnbaum-Saunders (BS). Vejamos a seguir como foi desenvolvido a construção dessa distribuição.

### 2.1 Origem

[Birnbaum e Saunders \(1969\)](#) sob a suposição de que um material é sujeito a um padrão cíclico <sup>1</sup> de tensão e força. A sequência de  $m$  cargas aplicadas em um ciclo será sempre a mesma em cada ciclo de cargas. Note que esta sequência levará até a falha desse material após o crescimento de uma fissura dominante dentro do material até sua ruptura, que ocorrerá quando tal fissura exceder um certo nível de resistência,

---

<sup>1</sup>Um ciclo é definido como  $m$  oscilações de carga e cada aplicação da  $i$ -ésima oscilação em um ciclo resulta em uma extensão incremental da fissura, definido como  $X_i$ .

denotado por  $\omega$ . A distribuição desta variável aleatória dependerá somente da fissura atual causada pela oscilação de carga neste ciclo, ou seja, as extensões incrementais das fissuras em cada ciclo são supostamente independentes. A extensão da fissura devido ao  $j$ -ésimo ciclo é dada por

$$Y_j = \sum_{i=1}^m X_i$$

em que  $Y_j$  corresponde a uma sequência de variáveis aleatórias independente e identicamente distribuídas com média  $\mu$  e variância  $\sigma^2$ , para todo  $j = 1, 2, \dots$ . Desta forma pode-se expressar a extensão total da fissura, após  $n$  ciclos, pela seguinte variável aleatória,

$$W_n = \sum_{i=1}^n Y_i$$

cuja função de distribuição acumulada (fda) é dada por

$$H_n(\omega) = P(W_n \leq \omega), \quad \text{para } n = 1, 2, \dots$$

Consideremos  $C$  como sendo o número de ciclos necessários para que ocorra a falha do material, ou seja, o número de ciclos que devem ocorrer para que a extensão total da fissura ( $W_n$ ) ultrapasse o valor crítico  $\omega$  ( $W_n > \omega$ ). Assim, a fda de  $C$  é

$$\begin{aligned} P(C \leq n) &= P\left(\sum_{i=1}^n Y_i > \omega\right) \\ &= P(W_n > \omega) \\ &= 1 - H_n(\omega). \end{aligned}$$

Além disso,

$$\begin{aligned} P(C \leq n) &= 1 - P\left(\sum_{i=1}^n Y_i \leq \omega\right) \\ &= 1 - P\left(\sum_{i=1}^n \frac{Y_i - \mu}{\sigma\sqrt{n}} \leq \frac{\omega - n\mu}{\sigma\sqrt{n}}\right). \end{aligned}$$

Mediante a suposição que os  $Y_j$ 's são independentes e identicamente distribuídos, pelo Teorema Central do Limite a fda de  $C$  pode ser expressa de forma aproximada como,

$$\begin{aligned} P(C \leq n) &\cong 1 - \Phi\left(\frac{\omega - n\mu}{\sigma\sqrt{n}}\right) \\ &= \Phi\left(-\frac{\omega - n\mu}{\sigma\sqrt{n}}\right) \\ &= \Phi\left(\frac{\mu\sqrt{n}}{\sigma} - \frac{\omega}{\sigma\sqrt{n}}\right), \end{aligned} \quad (2.1)$$

em que  $\Phi(\cdot)$  representa a fda da distribuição normal padrão.

Segundo [Birnbaum e Saunders \(1969\)](#), é possível substituir  $n$  em (2.1) por uma variável real não negativa  $t$ , tal que  $T$  é a extensão contínua da variável discreta  $C$  e representa o tempo total até a ocorrência de falha. Neste caso, se tomarmos uma reparametrização da forma

$$\alpha = \frac{\sigma}{\sqrt{\mu\omega}} > 0 \quad \text{e} \quad \beta = \frac{\omega}{\mu} > 0,$$

a fda de  $T$  fica expressa por

$$F_T(t) = \Phi\left[\frac{1}{\alpha}\left(\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right)\right], \quad \text{para } t > 0. \quad (2.2)$$

No que segue, para dizer que  $T$  segue uma distribuição BS, usaremos  $T \sim \text{BS}(\alpha, \beta)$ , em que  $\alpha$  e  $\beta$  correspondem aos parâmetros de forma e escala, respectivamente. A distribuição BS também é definida a partir da distribuição normal através da seguinte relação

$$T = \beta \left[ \frac{\alpha Z}{2} + \sqrt{\left(\frac{\alpha Z}{2}\right)^2 + 1} \right]^2, \quad (2.3)$$

em que  $Z \sim N(0, 1)$ . Mais detalhes ver, por exemplo, [Leiva \(2016\)](#).

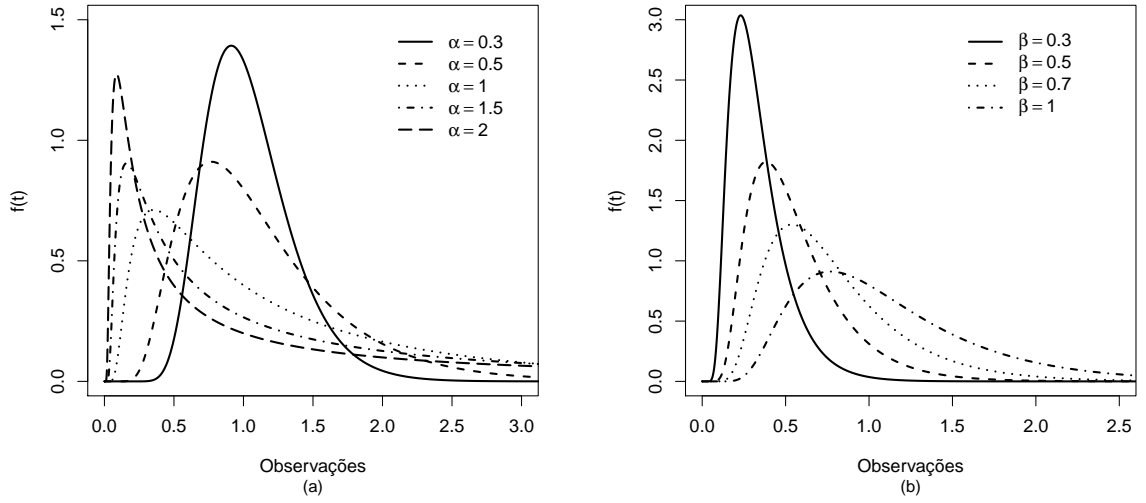
Derivando (2.2) com relação à  $t$ , obtemos a função densidade de probabilidade

(fdp) da variável aleatória  $T$ , ou seja,

$$f_T(t) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2\alpha^2} \left( \frac{t}{\beta} + \frac{\beta}{t} - 2 \right) \right\} \frac{t^{-3/2}(\beta + t)}{2\alpha\sqrt{\beta}}, \quad (2.4)$$

com  $t, \alpha, \beta > 0$ .

Figura 2.1: Função densidade da distribuição BS para  $\beta = 1$  com diferentes valores de  $\alpha$  (a) e  $\alpha = 0,5$  com diferentes valores de  $\beta$  (b).



Fonte: Autoria própria.

A Figura 2.1 ilustra o comportamento da função densidade da distribuição BS para diferentes valores dos parâmetros  $\alpha$  e  $\beta$ . Podemos observar na Figura 2.1 (a) que o parâmetro  $\alpha$  controla a forma da distribuição, a distribuição BS torna-se assimétrica à medida que  $\alpha$  cresce. Enquanto que o aumento do valor de  $\beta$  provoca mudanças apenas na escala, o aumento do valor de  $\beta$  provoca um crescimento na variabilidade das observações, comportamento este que pode ser observado na Figura 2.1 (b).

## 2.2 Função de sobrevivência e função de risco

Para uma variável aleatória  $T$  com distribuição BS a função de sobrevivência é dada por

$$S_T(t) = 1 - F_T(t)$$

$$\begin{aligned}
&= 1 - \Phi \left[ \frac{1}{\alpha} \left( \sqrt{\frac{\beta}{t}} - \sqrt{\frac{t}{\beta}} \right) \right] \\
&= \Phi \left[ \frac{1}{\alpha} \left( \sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}} \right) \right].
\end{aligned}$$

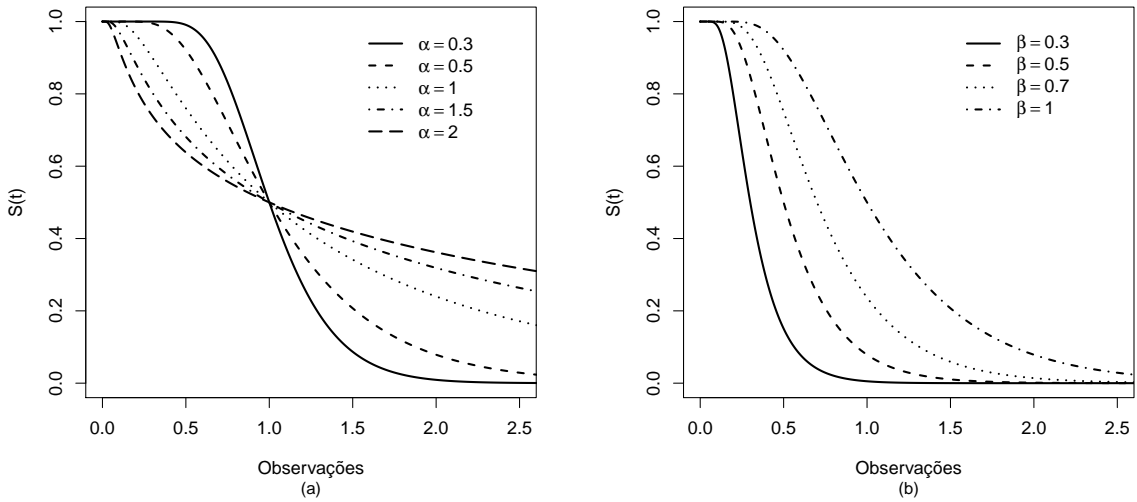
Já a função de risco é obtida da seguinte forma

$$\begin{aligned}
h_T(t) &= \frac{f_T(t)}{S_T(t)} = \frac{\frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2\alpha^2} \left( \frac{t}{\beta} + \frac{\beta}{t} - 2 \right) \right\} \frac{t^{-3/2}(\beta+t)}{2\alpha\sqrt{\beta}}}{\Phi \left[ \frac{1}{\alpha} \left( \sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}} \right) \right]} \\
&= \frac{t^{-3/2}(\beta+t) \exp \left\{ -\frac{1}{2\alpha^2} \left( \frac{t}{\beta} + \frac{\beta}{t} - 2 \right) \right\}}{2\alpha\sqrt{2\pi\beta} \Phi \left[ \frac{1}{\alpha} \left( \sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}} \right) \right]}.
\end{aligned}$$

As Figuras 2.2 e 2.3 ilustram as respectivas curvas das funções de sobrevivência e risco da distribuição BS para diferentes valores dos parâmetros  $\alpha$  e  $\beta$ .

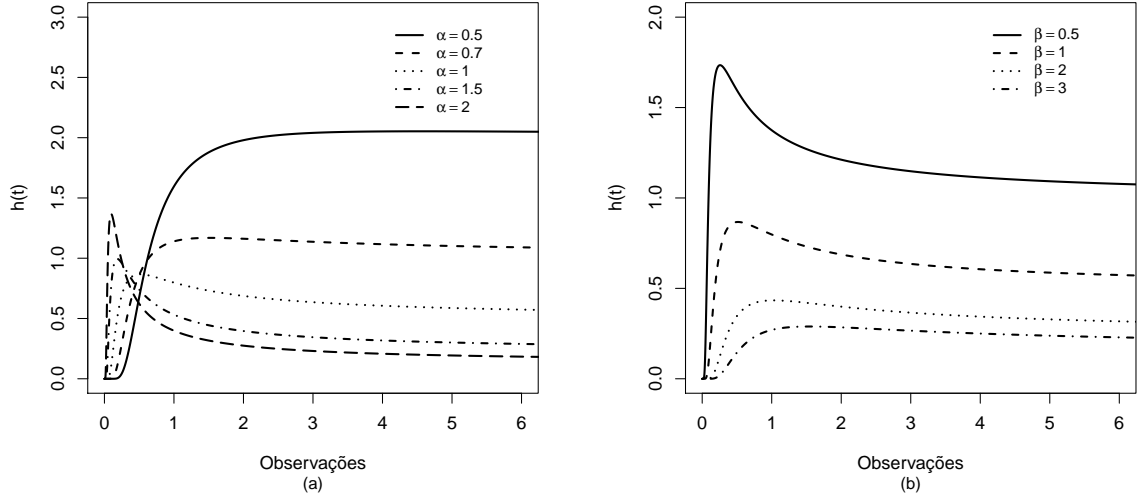
Na Figura 2.2 (a) observamos que, valores maiores de  $\alpha$  fazem com que a função de sobrevivência decresça de forma mais lenta, além disso conforme diminuimos os valores de  $\alpha$  a função de sobrevivência assume valores maiores até  $t = 1$ . Após  $t = 1$ ,

Figura 2.2: Função de sobrevivência da distribuição BS para  $\beta = 1$  com diferentes valores de  $\alpha$  (a) e  $\alpha = 0,5$  com diferentes valores de  $\beta$  (b).



Fonte: Autoria própria.

Figura 2.3: Função de risco da distribuição BS para  $\beta = 1$  com diferentes valores de  $\alpha$  (a) e  $\alpha = 1$  com diferentes valores de  $\beta$  (b).



Fonte: Autoria própria.

valores maiores de  $\alpha$  levam a maiores valores assumidos pela função de sobrevivência. Já na Figura 2.2 (b), nota-se que aumentando o valor de  $\beta$  faz com que o valor assumido pela função de sobrevivência seja sempre superior.

A Figura 2.3 (a) indica que valores menores de  $\alpha$  fazem com que a função de risco assumia maiores valores, além disso, nota-se que as curvas crescem até um certo  $t$ , e em seguida aparentam se estabilizar para algum valor. Na Figura 2.3 (b) os valores menores de  $\beta$  levam a maiores valores da função de risco, aqui as curvas também crescem até um certo  $t$  se estabilizando para algum valor em seguida.

## 2.3 Propriedades

Considere  $T \sim \text{BS}(\alpha, \beta)$ , então  $T$  é provida das seguintes propriedades:

**Propriedade 2.1.** A média e a variância de  $T$  são dadas, respectivamente, por

$$E(T) = \beta \left( 1 + \frac{\alpha^2}{2} \right) \quad e \quad \text{Var}(T) = (\alpha\beta)^2 \left( 1 + \frac{5\alpha^2}{4} \right).$$

*Demonstração:* Ver [Birnbaum e Saunders \(1969\)](#) ou [Leiva \(2016\)](#).



**Propriedade 2.2.** Se  $T \sim BS(\alpha, \beta)$ , então  $T^{-1} \sim BS(\alpha, \beta^{-1})$ .

*Demosntração:* Definamos  $Y = T^{-1}$ , consequentemente

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P(T^{-1} \leq y) \\
&= P(T > y^{-1}) \\
&= 1 - F_T(y^{-1}) \\
&= 1 - \Phi \left[ \frac{1}{\alpha} \left( \sqrt{\frac{y^{-1}}{\beta}} - \sqrt{\frac{\beta}{y^{-1}}} \right) \right] \\
&= 1 - \Phi \left[ -\frac{1}{\alpha} \left( \sqrt{\frac{y}{\beta^{-1}}} - \sqrt{\frac{\beta^{-1}}{y}} \right) \right] \\
&= \Phi \left[ \frac{1}{\alpha} \left( \sqrt{\frac{y}{\beta^{-1}}} - \sqrt{\frac{\beta^{-1}}{y}} \right) \right],
\end{aligned}$$

ou seja,  $Y \sim BS(\alpha, \beta^{-1})$ .

**Propriedade 2.3.** Se  $T \sim BS(\alpha, \beta)$  e  $a > 0$ , então  $aT \sim BS(\alpha, a\beta)$ .

*Demosntração:* Definamos, agora, a variável  $Y$  como sendo  $Y = aT$ , assim

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P(aT \leq y) \\
&= P(T \leq y/a) \\
&= F_T(y/a) \\
&= \Phi \left[ \frac{1}{\alpha} \left( \sqrt{\frac{y/a}{\beta}} - \sqrt{\frac{\beta}{y/a}} \right) \right] \\
&= \Phi \left[ \frac{1}{\alpha} \left( \sqrt{\frac{y}{a\beta}} - \sqrt{\frac{a\beta}{y}} \right) \right],
\end{aligned}$$

ou seja,  $Y \sim BS(\alpha, a\beta)$ .

**Propriedade 2.4.** O parâmetro  $\beta$  é a mediana da distribuição  $T \sim BS(\alpha, \beta)$ .

*Demonstração:*

$$F_T(\beta) = \Phi \left[ \frac{1}{\alpha} \left( \sqrt{\frac{\beta}{\beta}} - \sqrt{\frac{\beta}{\beta}} \right) \right] = \Phi \left[ \frac{1}{\alpha} (1 - 1) \right] = \Phi(0) = 0,5.$$

## 2.4 Estimação

Considere  $T_1, \dots, T_n$  uma amostra aleatória i.i.d. de tamanho  $n$  da distribuição BS com parâmetros  $\alpha$  e  $\beta$ . Utilizaremos o método de máxima verossimilhança, para observações não censuradas, para obter um estimador para  $\boldsymbol{\theta} = (\alpha, \beta)^\top$ , neste caso, a função de verossimilhança fica dada por

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n f(t_i|\boldsymbol{\theta}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2\alpha^2} \left( \frac{t_i}{\beta} + \frac{\beta}{t_i} - 2 \right) \right\} \frac{t_i^{-3/2}(\beta + t_i)}{2\alpha\sqrt{\beta}}. \quad (2.5)$$

Daí, o logaritmo da função de verossimilhança tem a seguinte forma

$$\begin{aligned} l(\boldsymbol{\theta}) &= \log(L(\boldsymbol{\theta})) \\ &= \sum_{i=1}^n \log \left[ \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2\alpha^2} \left( \frac{t_i}{\beta} + \frac{\beta}{t_i} - 2 \right) \right\} \frac{t_i^{-3/2}(\beta + t_i)}{2\alpha\sqrt{\beta}} \right] \\ &= \sum_{i=1}^n \left\{ \log \left( \frac{1}{\sqrt{2\pi}} \right) - \frac{1}{2\alpha^2} \left( \frac{t_i}{\beta} + \frac{\beta}{t_i} - 2 \right) - \frac{3}{2} \log(t_i) + \log(\beta + t_i) \right. \\ &\quad \left. - \log(2) - \log(\alpha) - \frac{1}{2} \log(\beta) \right\} \end{aligned} \quad (2.6)$$

$$\begin{aligned} &= n \log \left( \frac{1}{\sqrt{2\pi}} \right) - \frac{1}{2\alpha^2} \sum_{i=1}^n \left( \frac{t_i}{\beta} + \frac{\beta}{t_i} - 2 \right) - \frac{3}{2} \sum_{i=1}^n \log(t_i) \\ &\quad + \sum_{i=1}^n \log(\beta + t_i) - n \log(2) - n \log(\alpha) - \frac{n}{2} \log(\beta) \\ &= n \left( \log \left( \frac{1}{\sqrt{2\pi}} \right) - \log(2) \right) - \frac{1}{2\alpha^2} \sum_{i=1}^n \left( \frac{t_i}{\beta} + \frac{\beta}{t_i} - 2 \right) - \frac{3}{2} \sum_{i=1}^n \log(t_i) \\ &\quad + \sum_{i=1}^n \log(\beta + t_i) - n \log(\alpha) - \frac{n}{2} \log(\beta). \end{aligned} \quad (2.7)$$

Os estimadores de máxima verossimilhança de  $\boldsymbol{\theta}$ ,  $\hat{\alpha}$  e  $\hat{\beta}$  respectivamente, serão obtidos pela solução do sistema de equações formado pelas derivadas parciais de pri-

meira ordem do logaritmo da função de verossimilhança expressa em (2.7), em relação a cada um dos parâmetros, igualadas a zero, isto é, pela solução de  $U(\hat{\boldsymbol{\theta}}) = \mathbf{0}$ , ou seja,

$$U(\hat{\alpha}) = \frac{1}{\hat{\alpha}^3} \sum_{i=1}^n \left( \frac{t_i}{\hat{\beta}} + \frac{\hat{\beta}}{t_i} - 2 \right) - \frac{n}{\hat{\alpha}} = 0 \quad (2.8)$$

e

$$U(\hat{\beta}) = -\frac{1}{2\hat{\alpha}^2} \sum_{i=1}^n \left( -\frac{t_i}{\hat{\beta}^2} + \frac{1}{t_i} - 2 \right) + \sum_{i=1}^n \frac{1}{\hat{\beta} + t_i} - \frac{n}{2\hat{\beta}} = 0. \quad (2.9)$$

Da equação (2.8) obtemos que

$$\begin{aligned} \frac{n}{\hat{\alpha}} &= \frac{1}{\hat{\alpha}^3} \sum_{i=1}^n \left( \frac{t_i}{\hat{\beta}} + \frac{\hat{\beta}}{t_i} - 2 \right) \\ \hat{\alpha}^2 &= \frac{1}{n} \sum_{i=1}^n \left( \frac{t_i}{\hat{\beta}} + \frac{\hat{\beta}}{t_i} - 2 \right) \\ \hat{\alpha} &= \sqrt{\frac{1}{n} \sum_{i=1}^n \left( \frac{t_i}{\hat{\beta}} + \frac{\hat{\beta}}{t_i} - 2 \right)}, \end{aligned} \quad (2.10)$$

porém, não conseguimos obter solução analítica para o estimador de máxima verossimilhança  $\hat{\beta}$ , que deve ser obtido via algum método numérico, por exemplo, o método de Newton-Raphson ou um algoritmo quase Newton (BFGS, por exemplo).

## 2.5 Distribuição Senh-Normal

Nesta seção apresentaremos a distribuição seno hiperbólico normal, ou simplesmente senh-normal (SN), que foi desenvolvida por [Rieck e Nedelman \(1991\)](#) mediante uma transformação da distribuição normal baseada na seguinte relação

$$Y = \operatorname{arcsenh} \left( \frac{\alpha Z}{2} \right) \sigma + \gamma$$

em que  $Z \sim N(0, 1)$ ,  $\alpha > 0$  é o parâmetro de forma,  $\gamma \in \mathbb{R}$  é o parâmetro de locação e  $\sigma > 0$  é o parâmetro de escala. Dizemos que uma variável segue distribuição senh-normal, denotado simplesmente por  $Y \sim \text{SN}(\alpha, \gamma, \sigma)$ , se sua função de distribuição

acumulada for dada por

$$F_Y(y) = \Phi \left[ \frac{2}{\alpha} \sinh \left( \frac{y - \gamma}{\sigma} \right) \right], \quad y \in \mathbb{R},$$

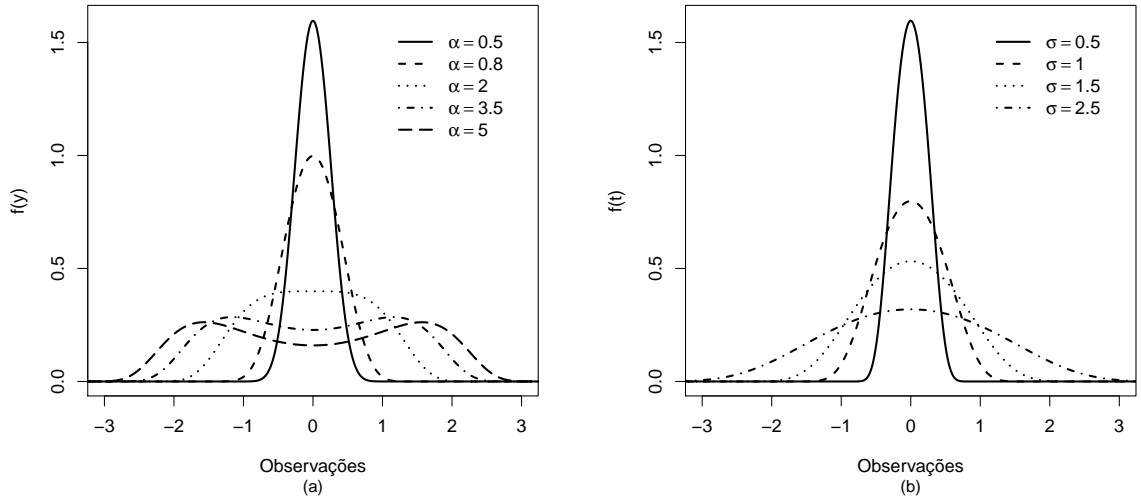
em que  $\Phi(\cdot)$  representa a fda da distribuição normal padrão.

A função densidade de  $Y \sim \text{SN}(\alpha, \gamma, \sigma)$  é dada por

$$f_Y(y) = \phi \left[ \frac{2}{\alpha} \sinh \left( \frac{y - \gamma}{\sigma} \right) \right] \frac{2}{\alpha \sigma} \cosh \left( \frac{y - \gamma}{\sigma} \right), \quad y \in \mathbb{R},$$

com  $\phi(\cdot)$  representando a fdp da distribuição normal padrão.

Figura 2.4: Função densidade da distribuição senh-normal para  $\gamma = 0$  e  $\sigma = 1$  com diferentes valores de  $\alpha$  (a) e  $\alpha = 1$ ,  $\gamma = 0$  com diferentes valores de  $\sigma$  (b).



Fonte: Autoria própria.

A Figura 2.4 ilustra o comportamento da função densidade da distribuição senh-normal para diferentes valores dos parâmetros  $\alpha$  e  $\sigma$ . Podemos observar na Figura 2.4 (a) que o parâmetro  $\alpha$  controla a forma da distribuição, a mesma permanece simétrica e que torna-se bimodal à medida que  $\alpha$  cresce. Por outro lado, um aumento no valor de  $\sigma$  provoca mudanças apenas na escala, ou seja, o aumento de  $\sigma$  provoca um crescimento na variabilidade das observações, comportamento este que pode ser observado na Figura 2.4 (b).

A função de sobrevivência e a função de risco para a variável aleatória  $Y \sim$

$SN(\alpha, \gamma, \sigma)$  são dadas, respectivamente, por

$$\begin{aligned} S_Y(y) &= 1 - F_Y(y) \\ &= \Phi \left[ -\frac{2}{\alpha} \sinh \left( \frac{y - \gamma}{\sigma} \right) \right] \end{aligned}$$

e

$$\begin{aligned} h_Y(y) &= \frac{f_Y(y)}{S_Y(y)} \\ &= \frac{\phi \left[ \frac{2}{\alpha} \sinh \left( \frac{y - \gamma}{\sigma} \right) \right] \frac{2}{\alpha \sigma} \cosh \left( \frac{y - \gamma}{\sigma} \right)}{\Phi \left[ -\frac{2}{\alpha} \sinh \left( \frac{y - \gamma}{\sigma} \right) \right]}, \end{aligned}$$

em que  $\phi(\cdot)$  e  $\Phi(\cdot)$  são, respectivamente, a fdp e fda da distribuição normal padrão.

### 2.5.1 Propriedades da distribuição Senh-Normal

Exploraremos aqui algumas propriedades que serão de extrema importância para construção do modelo de tempo de vida acelerado BS e a primeira delas relaciona as distribuições BS e SN.

**Propriedade 2.5.** *Uma variável aleatória  $Y = \log(T)$  possui distribuição SN com parâmetros  $\alpha$ ,  $\log(\beta)$  e 2 se, e somente se, a variável aleatória  $T$  possui distribuição BS de parâmetros  $\alpha$  e  $\beta$ , ou seja*

$$T \sim BS(\alpha, \beta) \iff Y = \log(T) \sim SN(\alpha, \log(\beta), 2).$$

*Demonstração:* Suponhamos inicialmente que  $T \sim BS(\alpha, \beta)$ . Assim,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(\log(T) \leq y) \\ &= P(T \leq e^y) \\ &= F_T(e^y) \\ &= \Phi \left[ \frac{1}{\alpha} \left( \sqrt{\frac{e^y}{\beta}} - \sqrt{\frac{\beta}{e^y}} \right) \right] \\ &= \Phi \left[ \frac{1}{\alpha} \left( e^{\frac{y}{2}} \beta^{-\frac{1}{2}} - \beta^{\frac{1}{2}} e^{-\frac{y}{2}} \right) \right] \end{aligned}$$

$$\begin{aligned}
&= \Phi \left[ \frac{1}{\alpha} \left( \exp \left\{ \frac{y - \log(\beta)}{2} \right\} - \exp \left\{ -\frac{y - \log(\beta)}{2} \right\} \right) \right] \\
&= \Phi \left[ \frac{2}{\alpha} \sinh \left( \frac{y - \log(\beta)}{2} \right) \right],
\end{aligned}$$

isto é,  $Y = \log(T) \sim \text{SN}(\alpha, \log(\beta), 2)$ . Suponhamos, agora, que  $Y = \log(T) \sim \text{SN}(\alpha, \log(\beta), 2)$ . Assim,

$$\begin{aligned}
F_T(t) &= P(T \leq t) \\
&= P(\log(T) \leq \log(t)) \\
&= P(Y \leq \log(t)) \\
&= F_Y(\log(t)) \\
&= \Phi \left[ \frac{2}{\alpha} \sinh \left( \frac{\log(t) - \log(\beta)}{2} \right) \right] \\
&= \Phi \left[ \frac{1}{\alpha} \left( \exp \left\{ \frac{\log(t) - \log(\beta)}{2} \right\} - \exp \left\{ -\frac{\log(t) - \log(\beta)}{2} \right\} \right) \right] \\
&= \Phi \left[ \frac{1}{\alpha} \left( t^{\frac{1}{2}} \beta^{-\frac{1}{2}} - \beta^{\frac{1}{2}} t^{-\frac{1}{2}} \right) \right] \\
&= \Phi \left[ \frac{1}{\alpha} \left( \sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}} \right) \right],
\end{aligned}$$

isto é,  $T \sim \text{BS}(\alpha, \beta)$ .

Devido a Propriedade 2.5 a distribuição de  $Y = \log(T)$  é também chamada de distribuição log-Birnbaum-Saunders (log-BS). Outra propriedade nos permite examinar qual é o efeito de uma transformação linear sobre a distribuição SN.

**Propriedade 2.6.** Se  $Y \sim \text{SN}(\alpha, \gamma, \sigma)$ , então  $a + bY \sim \text{SN}(\alpha, a + b\gamma, |b|\sigma)$

*Demonstração:* Ver [Silva \(2013\)](#).

**Propriedade 2.7.** A esperança e a variância de uma variável aleatória  $Y \sim \text{SN}(\alpha, \gamma, \sigma)$ , são dadas, respectivamente, por

$$E(Y) = \gamma \quad e \quad \text{Var}(Y) = \sigma^2 g(\alpha),$$

em que  $g(\alpha)$  é uma função baseada na função de Bessel modificada do terceiro tipo e que depende apenas de  $\alpha$ .

*Demonstração:* Ver [Villegas et al. \(2011\)](#).

Usando expansões assintóticas da função geradora de momentos, [Rieck \(1989\)](#) obteve aproximações para a variância da distribuição SN para  $\alpha$  pequeno e para  $\alpha$  grande. Para pequenos valores de  $\alpha$ , a variância da distribuição SN pode ser obtidas por

$$Var(Y) = \frac{\alpha^2 \sigma^2}{4} - \frac{\alpha^4 \sigma^2}{16},$$

já para grandes valores de  $\alpha$ , a variância da distribuição SN podem ser obtidas por

$$Var(Y) = \sigma^2 \{2,197543455 - 1,963510026 \log(\alpha \sqrt{(2)}) + [\log(\alpha \sqrt{(2)})]^2\}.$$

**Propriedade 2.8.** *O parâmetro  $\gamma$  é a mediana da distribuição SN.*

*Demonstração:*

$$F_Y(\gamma) = \Phi \left[ \frac{2}{\alpha} \sinh \left( \frac{\gamma - \gamma}{\sigma} \right) \right] = \Phi[0] = 0,5.$$

## 2.6 Modelo de tempo de vida acelerado Birnbaum-Saunders

O modelo de regressão BS adota um componente determinístico não-linear da forma  $\exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}$  combinado com um componente aleatório na forma  $\exp\{\epsilon_i\}$  gerando o modelo

$$T = \exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\} \exp\{\epsilon_i\}, \quad (2.11)$$

em que  $\mathbf{x}_i^\top = (1, x_{1i}, x_{2i}, \dots, x_{pi})$  e  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^\top$ , esse modelo é um caso particular dos modelos de tempo de vida acelerado visto anteriormente, pois consideramos fixo o parâmetro extra de escala, ou seja,  $\sigma = 1$ . Linearizando ([2.11](#)) obtemos o modelo

log-BS, dado por

$$\log(T) = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \quad (2.12)$$

em que  $\epsilon_i \sim \text{SN}(\alpha, 0, 2)$ .

Para a forma log-linear deste modelo de tempo de vida acelerado, as funções de distribuição acumulada e densidade serão dadas, respectivamente, por

$$\begin{aligned} F(t|\mathbf{x}_i) &= P(T \leq t) \\ &= P(\exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\} \exp\{\epsilon_i\} \leq t) \\ &= P(\mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i \leq \log(t)) \\ &= P(\epsilon_i \leq \log(t) - \mathbf{x}_i^\top \boldsymbol{\beta}) \\ &= F_{\epsilon_i}(\log(t) - \mathbf{x}_i^\top \boldsymbol{\beta}) \end{aligned}$$

e

$$\begin{aligned} f(t|\mathbf{x}_i) &= [F_{\epsilon_i}(\log(t) - \mathbf{x}_i^\top \boldsymbol{\beta})]' \\ &= f_{\epsilon_i}(\log(t) - \mathbf{x}_i^\top \boldsymbol{\beta}) \frac{1}{t} \\ &= \frac{1}{t} f_{\epsilon_i}(\log(t) - \mathbf{x}_i^\top \boldsymbol{\beta}), \end{aligned}$$

portanto, para o modelo log-BS, como  $\epsilon_i \sim \text{SN}(\alpha, 0, 2)$ , temos que  $\log(T) \sim \text{SN}(\alpha, \mathbf{x}_i^\top \boldsymbol{\beta}, 2)$  com

$$F(t|\mathbf{x}_i) = \Phi\left(\frac{2}{\alpha} \sinh\left(\frac{\log(t) - \mathbf{x}_i^\top \boldsymbol{\beta}}{2}\right)\right)$$

e

$$f(t|\mathbf{x}_i) = \frac{1}{\alpha t} \phi\left(\frac{2}{\alpha} \sinh\left(\frac{\log(t) - \mathbf{x}_i^\top \boldsymbol{\beta}}{2}\right)\right) \cosh\left(\frac{\log(t) - \mathbf{x}_i^\top \boldsymbol{\beta}}{2}\right).$$

Já as funções de sobrevivência e risco,  $S(t|\mathbf{x}_i)$  e  $h(t|\mathbf{x}_i)$ , podem ser obtidas, respectivamente, a partir da função de sobrevivência e da função de risco do erro  $\epsilon_i$ , ou



seja,

$$\begin{aligned}
S(t|\mathbf{x}_i) &= P(T > t) \\
&= P(\exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\} \exp\{\epsilon_i\} > t) \\
&= P(\mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i > \log(t)) \\
&= P(\epsilon_i > \log(t) - \mathbf{x}_i^\top \boldsymbol{\beta}) \\
&= S_{\epsilon_i}(\log(t) - \mathbf{x}_i^\top \boldsymbol{\beta}) .
\end{aligned}$$

e

$$\begin{aligned}
h(t|\mathbf{x}_i) &= \frac{f(t|\mathbf{x}_i)}{S(t|\mathbf{x}_i)} \\
&= \frac{\frac{1}{t} f_{\epsilon_i}(\log(t) - \mathbf{x}_i^\top \boldsymbol{\beta})}{S_{\epsilon_i}(\log(t) - \mathbf{x}_i^\top \boldsymbol{\beta})} \\
&= \frac{1}{t} h_{\epsilon_i}(\log(t) - \mathbf{x}_i^\top \boldsymbol{\beta}) .
\end{aligned}$$



## Capítulo 3

# Estimação robusta para o modelo de tempo de vida acelerado Birnbaum-Saunders

Inicialmente descreveremos a estimação dos parâmetros do modelo pelo método de máxima verossimilhança. Para os dados  $\{(t_i, \delta_i, \mathbf{x}_i); i = 1, \dots, n\}$ , a função de verossimilhança para o parâmetro  $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}^\top)^\top$  é dada por

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n [h(t_i|\mathbf{x}_i)]^{\delta_i} S(t_i|\mathbf{x}_i).$$

O logaritmo da função de verossimilhança correspondente é dado por

$$l(\boldsymbol{\theta}) = \sum_{i=1}^n [\delta_i \{-\log(t_i) + \log\{h_{\epsilon_i}(\epsilon_i)\}\} + \log\{S_{\epsilon_i}(\epsilon_i)\}],$$

em que  $\epsilon_i = \log(t_i) - \mathbf{x}_i^\top \boldsymbol{\beta}$ .

As funções score para estimar  $\alpha$  e  $\boldsymbol{\beta}$  podem ser escritas na forma

$$\begin{aligned} \frac{\partial l(\boldsymbol{\theta})}{\partial \alpha} &= \sum_{i=1}^n \left[ \delta_i \frac{\partial \log\{h_{\epsilon_i}(\epsilon_i)\}}{\partial \alpha} + \frac{\partial \log\{S_{\epsilon_i}(\epsilon_i)\}}{\partial \alpha} \right], \\ \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^n \left[ \delta_i \frac{\partial \log\{h_{\epsilon_i}(\epsilon_i)\}}{\partial \epsilon_i} + \frac{\partial \log\{S_{\epsilon_i}(\epsilon_i)\}}{\partial \epsilon_i} \right] \frac{\partial \epsilon_i}{\partial \boldsymbol{\beta}}. \end{aligned}$$

que levam às equações de verossimilhança

$$\begin{cases} \sum_{i=1}^n \left[ \delta_i \frac{\partial \log\{h_{\epsilon_i}(\epsilon_i)\}}{\partial \alpha} + \frac{\partial \log\{S_{\epsilon_i}(\epsilon_i)\}}{\partial \alpha} \right] = 0, \\ \sum_{i=1}^n \left[ \delta_i \frac{\partial \log\{h_{\epsilon_i}(\epsilon_i)\}}{\partial \epsilon_i} + \frac{\partial \log\{S_{\epsilon_i}(\epsilon_i)\}}{\partial \epsilon_i} \right] \mathbf{x}_i = \mathbf{0}. \end{cases} \quad (3.1)$$

e assim as equações robustas são escritas como

$$\begin{cases} \sum_{i=1}^n \left[ \left\{ \delta_i \frac{\partial \log\{h_{\epsilon_i}(z_i)\}}{\partial \alpha} + \frac{\partial \log\{S_{\epsilon_i}(z_i)\}}{\partial \alpha} \right\} - a_\alpha \right] = 0, \\ \sum_{i=1}^n \left[ \left\{ \delta_i \frac{\partial \log\{h_{\epsilon_i}(z_i)\}}{\partial z_i} + \frac{\partial \log\{S_{\epsilon_i}(z_i)\}}{\partial z_i} \right\} w(\mathbf{x}_i) \mathbf{x}_i - \mathbf{a}_\beta \right] = \mathbf{0}, \end{cases} \quad (3.2)$$

em que  $z_i = E(\epsilon_i) + \sqrt{Var(\epsilon_i)}\psi_c(r_i)$  com  $r_i = (\epsilon_i - E(\epsilon_i))/\sqrt{Var(\epsilon_i)}$  e  $\psi_c$  é a função psi de Huber:  $\psi_c(r_i) = \max(-c, \min(r_i, c))$ .

As constantes de sintonia  $a_\alpha$  e  $\mathbf{a}_\beta$  que são escolhidas de modo a fornecer estimativas não viesadas para os parâmetros do modelo  $\alpha$  e  $\beta$  são dadas por

$$\begin{aligned} a_\alpha &= \frac{1}{n} \sum_{i=1}^n \delta_i E \left\{ \frac{\partial \log\{h_{\epsilon_i}(z_i)\}}{\partial \alpha} + \frac{\partial \log\{S_{\epsilon_i}(z_i)\}}{\partial \alpha} \right\} \\ \mathbf{a}_\beta &= \frac{1}{n} \sum_{i=1}^n \delta_i E \left\{ \frac{\partial \log\{h_{\epsilon_i}(z_i)\}}{\partial z_i} + \frac{\partial \log\{S_{\epsilon_i}(z_i)\}}{\partial z_i} \right\} w(\mathbf{x}_i) \mathbf{x}_i. \end{aligned}$$

Portanto, em termos da forma log-linear do modelo, se  $T \sim \text{BS}(\exp\{\mathbf{x}_i^\top \beta\}, \alpha)$ , então  $\epsilon_i \sim \text{SN}(\alpha, 0, 2)$ , assim

$$f_{\epsilon_i}(\epsilon_i) = \frac{1}{2} \xi_{i1} \phi(\xi_{i2}) \quad \text{e} \quad F_{\epsilon_i}(\epsilon_i) = \Phi(\xi_{i2}),$$

em que  $\xi_{i1} = \frac{2}{\alpha} \cosh\left(\frac{\epsilon_i}{2}\right)$  e  $\xi_{i2} = \frac{2}{\alpha} \sinh\left(\frac{\epsilon_i}{2}\right)$ , com média  $E(\epsilon_i) = 0$  e variância  $Var(\epsilon_i) = \sigma^2 g(\alpha)$ , em que  $g(\alpha)$  é uma função baseada na função de Bessel modificada do terceiro tipo. A função de sobrevivência  $S_{\epsilon_i}(\epsilon_i)$  e a função de risco  $h_{\epsilon_i}(\epsilon_i)$  de  $\epsilon_i$  são dadas, respectivamente, como

$$S_{\epsilon_i}(\epsilon_i) = 1 - \Phi(\xi_{i2}) \quad \text{e} \quad h_{\epsilon_i}(\epsilon_i) = \frac{\frac{1}{2} \xi_{i1} \phi(\xi_{i2})}{1 - \Phi(\xi_{i2})},$$

(ver apêndice A.1), para o qual as derivadas parciais de  $\log\{S_{\epsilon_i}(\epsilon_i)\}$  e  $\log\{h_{\epsilon_i}(\epsilon_i)\}$  em

relação a  $\alpha$  são dadas por

$$\frac{\partial \log\{S_{\epsilon_i}(\epsilon_i)\}}{\partial \alpha} = \frac{\xi_{i2}}{\alpha} K(\xi_{i2}) \quad \text{e} \quad \frac{\partial \log\{h_{\epsilon_i}(\epsilon_i)\}}{\partial \alpha} = -\frac{1}{\alpha} + \frac{\xi_{i2}^2}{\alpha} - \frac{\xi_{i2}}{\alpha} K(\xi_{i2})$$

em que  $K(\xi_{i2}) = \frac{\phi(\xi_{i2})}{1-\Phi(\xi_{i2})}$  e as derivadas parciais de  $\log\{h_{\epsilon_i}(\epsilon_i)\}$  e  $\log\{S_{\epsilon_i}(\epsilon_i)\}$  em relação a  $\epsilon_i$  são

$$\frac{\partial \log\{S_{\epsilon_i}(\epsilon_i)\}}{\partial \epsilon_i} = -\frac{\xi_{i1}}{2} K(\xi_{i2}) \quad \text{e} \quad \frac{\partial \log\{h_{\epsilon_i}(\epsilon_i)\}}{\partial \epsilon_i} = \frac{1}{2} \tanh\left(\frac{\epsilon_i}{2}\right) - \frac{1}{\alpha^2} \sinh(\epsilon_i) + \frac{\xi_{i1}}{2} K(\xi_{i2}).$$

Substituindo estas derivadas em (3.1), as equações de verossimilhança para o modelo de tempo de falha BS são dadas por

$$\begin{cases} \sum_{i=1}^n \left[ \delta_i \left( \frac{\xi_{i2}^2}{\alpha} - \frac{1}{\alpha} \right) + (1 - \delta_i) \frac{\xi_{i2}}{\alpha} K(\xi_{i2}) \right] & = 0, \quad (\text{para } \alpha) \\ \sum_{i=1}^n \left[ \delta_i \left( \frac{1}{\alpha^2} \sinh(\epsilon_i) - \frac{1}{2} \tanh\left(\frac{\epsilon_i}{2}\right) \right) + (1 - \delta_i) \frac{\xi_{i1}}{2} K(\xi_{i2}) \right] \mathbf{x}_i & = \mathbf{0}, \quad (\text{para } \boldsymbol{\beta}) \end{cases}$$

(ver apêndice A.2).

E as equações robustas ficam dadas por

$$\begin{cases} \sum_{i=1}^n \left[ \left\{ \delta_i \left( \frac{\xi_{i2z}^2}{\alpha} - \frac{1}{\alpha} \right) + (1 - \delta_i) \frac{\xi_{i2z}}{\alpha} K(\xi_{i2z}) \right\} - a_\alpha \right] & = 0. \\ \sum_{i=1}^n \left[ \left\{ \delta_i \left( \frac{1}{\alpha^2} \sinh(z_i) - \frac{1}{2} \tanh\left(\frac{z_i}{2}\right) \right) + (1 - \delta_i) \frac{\xi_{i1z}}{2} K(\xi_{i2z}) \right\} w(\mathbf{x}_i) \mathbf{x}_i - \mathbf{a}_\beta \right] & = \mathbf{0}, \end{cases} \quad (3.3)$$

em que  $\xi_{i1z} = \frac{2}{\alpha} \cosh\left(\frac{z_i}{2}\right)$ ,  $\xi_{i2z} = \frac{2}{\alpha} \sinh\left(\frac{z_i}{2}\right)$ ,  $z_i = E(\epsilon_i) + \sqrt{\text{Var}(\epsilon_i)} \psi_c(r_i)$  com  $r_i = (\epsilon_i - E(\epsilon_i))/\sqrt{\text{Var}(\epsilon_i)}$ ,  $E(\epsilon_i) = 0$  e  $\text{Var}(\epsilon_i) = \sigma^2 g(\alpha)$ , em que  $g(\alpha)$  é uma função baseada na função de Bessel modificada do terceiro tipo. As constantes de sintonia são dadas por

$$\begin{aligned} a_\alpha &= \frac{1}{n} \sum_{i=1}^n \delta_i E \left\{ \frac{\xi_{i2z}^2}{\alpha} - \frac{1}{\alpha} \right\} \\ \mathbf{a}_\beta &= \frac{1}{n} \sum_{i=1}^n \delta_i E \left\{ \frac{1}{\alpha^2} \sinh(z_i) - \frac{1}{2} \tanh\left(\frac{z_i}{2}\right) \right\} w(\mathbf{x}_i) \mathbf{x}_i. \end{aligned}$$

O sistema de equações (3.3) podem ser resolvidas numericamente para obtenção

dos estimadores robustos de  $\alpha$  e  $\beta$  e para resolvê-las usaremos o método iterativo do escore de Fisher no qual tais equações para  $\boldsymbol{\theta} = (\alpha, \beta^\top)^\top$  são reescritas na forma  $\Upsilon(\boldsymbol{\theta}) = \mathbf{0}$ , onde a função objetivo  $\Upsilon(\boldsymbol{\theta}) = (\Upsilon_\alpha, \Upsilon_\beta^\top)^\top$  possui os componentes

$$\begin{aligned}\Upsilon_\alpha &= \sum_{i=1}^n \left[ \left\{ \delta_i \left( \frac{\xi_{i2z}^2}{\alpha} - \frac{1}{\alpha} \right) + (1 - \delta_i) \frac{\xi_{i2z}}{\alpha} K(\xi_{i2z}) \right\} - a_\alpha \right], \\ \Upsilon_\beta &= \sum_{i=1}^n \left[ \left\{ \delta_i \left( \frac{1}{\alpha^2} \sinh(z_i) - \frac{1}{2} \tanh\left(\frac{z_i}{2}\right) \right) + (1 - \delta_i) \frac{\xi_{i1z}}{2} K(\xi_{i2z}) \right\} w(\mathbf{x}_i) \mathbf{x}_i - \mathbf{a}_\beta \right].\end{aligned}$$

Usando uma aproximação de Taylor de primeira ordem de  $\Upsilon(\boldsymbol{\theta})$  em torno de alguns valores iniciais  $\boldsymbol{\theta}^{(0)}$ , podemos chegar ao processo iterativo de Newton-Raphson

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} - \left\{ \Upsilon'(\boldsymbol{\theta}^{(k)}) \right\}^{-1} \Upsilon(\boldsymbol{\theta}^{(k)}), \quad (3.4)$$

para  $k = 0, 1, 2, \dots$ , onde  $\Upsilon'(\boldsymbol{\theta}^{(k)})$  é a primeira derivada de  $\Upsilon(\boldsymbol{\theta})$  em relação a  $\boldsymbol{\theta}$  avaliada em  $\boldsymbol{\theta}^{(k)}$ . A matriz  $\Upsilon'(\boldsymbol{\theta})$  é de ordem  $(p+2) \times (p+2)$ , e pode ser particionada como

$$\Upsilon'(\boldsymbol{\theta}) = \begin{pmatrix} \Upsilon'_{\alpha\alpha} & \Upsilon'_{\alpha\beta} \\ \Upsilon'_{\beta\alpha} & \Upsilon'_{\beta\beta} \end{pmatrix},$$

em que  $\Upsilon'_{\alpha\alpha}$  é de ordem  $1 \times 1$ ,  $\Upsilon'_{\alpha\beta}$  é de ordem  $1 \times (p+1)$ ,  $\Upsilon'_{\beta\alpha}$  é de ordem  $(p+1) \times 1$  e  $\Upsilon'_{\beta\beta}$  é de ordem  $(p+1) \times (p+1)$  onde os componentes da matriz particionada são dados por

$$\begin{aligned}\Upsilon'_{\alpha\alpha} &= \sum_{i=1}^n \left[ \delta_i \left( \frac{1 - 3\xi_{i2z}^2}{\alpha^2} \right) - (1 - \delta_i) \left( \frac{2\xi_{i2z}}{\alpha^2} K(\xi_{i2z}) + \frac{\xi_{i2z}^2}{\alpha^2} K'(\xi_{i2z}) \right) \right], \\ \Upsilon'_{\alpha\beta} &= \sum_{i=1}^n \left[ \delta_i \left( -\frac{2}{\alpha^3} \sinh(z_i) \right) - (1 - \delta_i) \left( \frac{\xi_{i1z}}{2\alpha} K(\xi_{i2z}) + \frac{1}{\alpha^3} \sinh(z_i) K'(\xi_{i2z}) \right) \right] \psi'_c(r_i) \mathbf{x}_i, \\ \Upsilon'_{\beta\alpha} &= \sum_{i=1}^n \left[ \delta_i \left( -\frac{2}{\alpha^3} \sinh(z_i) \right) - (1 - \delta_i) \left( \frac{\xi_{i1z}}{2\alpha} K(\xi_{i2z}) + \frac{1}{\alpha^3} \sinh(z_i) K'(\xi_{i2z}) \right) \right] w(\mathbf{x}_i) \mathbf{x}_i, \\ \Upsilon'_{\beta\beta} &= \sum_{i=1}^n \left[ \delta_i \left( \frac{1}{4} \operatorname{sech}^2\left(\frac{z_i}{2}\right) - \frac{1}{\alpha^2} \cosh(z_i) \right) \right. \\ &\quad \left. - (1 - \delta_i) \left( \frac{\xi_{i2z}}{4} K(\xi_{i2z}) + \frac{\xi_{i1z}^2}{4} K'(\xi_{i2z}) \right) \right] \psi'_c(r_i) w(\mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^\top,\end{aligned}$$

em que  $K'(\xi_{i2z}) = \frac{\phi'(\xi_{i2z})}{1 - \Phi(\xi_{i2z})} + \left( \frac{\phi(\xi_{i2z})}{1 - \Phi(\xi_{i2z})} \right)^2$ , (ver apêndice A.3).

Como a matriz  $\Upsilon'(\boldsymbol{\theta})$  pode não ser positiva definida, a aplicação do método escore de Fisher descrito em Paula (2013) pode ser mais conveniente. O método consiste em substituir a matriz  $\Upsilon'(\boldsymbol{\theta})$  pelo negativo do correspondente valor esperado  $-E[\Upsilon'(\boldsymbol{\theta})]$ . Assim, o processo iterativo em (3.4) resulta em

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \left\{ E[\Upsilon'(\boldsymbol{\theta}^{(k)})] \right\}^{-1} \Upsilon(\boldsymbol{\theta}^{(k)}), \quad (3.5)$$

A partir das estimativas iniciais  $\boldsymbol{\theta}^{(0)}$ , continuamos as iterações (3.5) até que uma convergência seja alcançada. Nesse trabalho, o critério de convergência foi dado pela condição da  $\sum_{j=1}^p |\boldsymbol{\theta}_j^{(k)} - \boldsymbol{\theta}_j^{(k+1)}| < 0,015$ , sendo  $p$  o número de parâmetros do modelo. As estimativas iniciais dos parâmetros do modelo podem ser escolhidas como as de máxima verossimilhança e as estimativas que convergirem são as estimativas robustas  $\hat{\boldsymbol{\theta}}$  de  $\boldsymbol{\theta}$ .

Seguindo (1.16), a variância aproximada do ERM  $\hat{\boldsymbol{\theta}}$  pode ser obtida da matriz de variância-covariância do tipo sanduíche.

$$V(\hat{\boldsymbol{\theta}}) = \mathbf{M}^{-1} \mathbf{Q} \{\mathbf{M}^{-1}\}^{\top},$$

em que  $\mathbf{M}$  é uma matriz  $(p+2) \times (p+2)$  é definida por  $\mathbf{M} = -\Upsilon'(\boldsymbol{\theta})$  e  $\mathbf{Q}$  corresponde a uma matriz  $(p+2) \times (p+2)$ , que pode ser particionada como

$$\mathbf{Q} = \begin{pmatrix} Q_{\alpha\alpha} & Q_{\alpha\beta} \\ Q_{\beta\alpha} & Q_{\beta\beta} \end{pmatrix},$$

em que  $Q_{\alpha\alpha}$  é de ordem  $1 \times 1$ ,  $Q_{\alpha\beta}$  é de ordem  $1 \times (p+1)$ ,  $Q_{\beta\alpha}$  é de ordem  $(p+1) \times 1$  e  $Q_{\beta\beta}$  é de ordem  $(p+1) \times (p+1)$  cujos componentes são dados por

$$\begin{aligned} Q_{\alpha\alpha} &= \sum_{i=1}^n \left[ \left\{ \delta_i \left( \frac{\xi_{i2z}^2}{\alpha} - \frac{1}{\alpha} \right) + (1 - \delta_i) \frac{\xi_{i2z}}{\alpha} K(\xi_{i2z}) \right\} - a_{\alpha} \right]^2, \\ Q_{\beta\beta} &= \sum_{i=1}^n \left[ \left\{ \delta_i \left( \frac{1}{\alpha^2} \sinh(z_i) - \frac{1}{2} \tanh\left(\frac{z_i}{2}\right) \right) + (1 - \delta_i) \frac{\xi_{i1z}}{2} K(\xi_{i2z}) \right\} w(\mathbf{x}_i) \mathbf{x}_i - \mathbf{a}_{\beta} \right] \times \\ &\quad \left[ \left\{ \delta_i \left( \frac{1}{\alpha^2} \sinh(z_i) - \frac{1}{2} \tanh\left(\frac{z_i}{2}\right) \right) + (1 - \delta_i) \frac{\xi_{i1z}}{2} K(\xi_{i2z}) \right\} w(\mathbf{x}_i) \mathbf{x}_i - \mathbf{a}_{\beta} \right]^{\top}, \end{aligned}$$

$$Q_{\alpha\beta} = Q_{\beta\alpha}^\top = \sum_{i=1}^n \left[ \left\{ \delta_i \left( \frac{\xi_{i2z}^2}{\alpha} - \frac{1}{\alpha} \right) + (1 - \delta_i) \frac{\xi_{i2z}}{\alpha} K(\xi_{i2z}) \right\} - a_\alpha \right] \times \\ \left[ \left\{ \delta_i \left( \frac{1}{\alpha^2} \sinh(z_i) - \frac{1}{2} \tanh\left(\frac{z_i}{2}\right) \right) + (1 - \delta_i) \frac{\xi_{i1z}}{2} K(\xi_{i2z}) \right\} w(\mathbf{x}_i) \mathbf{x}_i - \mathbf{a}_\beta \right].$$

Ambas as matrizes  $\mathbf{M}$  e  $\mathbf{Q}$  são avaliadas no ERM  $\hat{\boldsymbol{\theta}}$  para uma estimativa da matriz de variância-covariância  $\hat{V}(\hat{\boldsymbol{\theta}})$ .

### 3.1 Estudo de simulação

Para avaliar o desempenho do estimador robusto proposto, realizamos um conjunto de simulações no qual investigamos primeiro o comportamento do estimador quando nenhum *outlier* foi considerado e em seguida quando os dados foram contaminados com *outliers*.

Os tempos de sobrevivência de  $t_i$  ( $i = 1, \dots, n$ ) foram gerados a partir do modelo de tempo de vida acelerado BS através de observações normais usando a expressão (2.3) com parâmetro de escala  $\exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}$  em que  $\mathbf{x}_i^\top \boldsymbol{\beta} = \beta_0 + \beta_1 x_{i1}$ . A variável explicativa  $x_1$  foi gerada a partir de uma distribuição uniforme  $U(0, 1)$ . Os parâmetros foram fixados em  $\alpha = 1$ ,  $\beta_0 = 4$  e  $\beta_1 = -2$  e os conjuntos de dados foram gerados para duas combinações de tamanhos de amostras  $n = 100, 200$  e com três combinações de tempos de sobrevivência censurados, sendo elas 0%, 10% e 30% de censura aleatória. As porcentagens de censura ( $pc$ ) foram obtidas seguindo procedimentos apresentados por Wang et al. (2006) no qual geramos as variáveis censuradas  $C_i$  da distribuição log-normal com fda  $\Phi\left(\frac{\log(u) - \mu}{\sigma}\right)$ , em que  $\Phi(\cdot)$  é a fda da distribuição normal padrão. Fixamos  $\sigma = 1$  e variamos o valor de  $\mu$  até obter a proporção de censura desejada. Assim, os tempos observados foram  $\min(T_i, C_i)$  e o indicador de falha é  $\delta_i = I_{(T_i \leq C_i)}$ , sendo  $I_{(\cdot)}$  a função indicadora. Para cada combinação de  $n$  e  $pc$  foram geradas 5000 amostras aleatórias e para cada amostra ajustamos o modelo de tempo de vida acelerado BS dado em (2.12), em que  $Y_i = \log\{\min(T_i, C_i)\}$ .

Para estudar os estimadores na presença de dados contaminados, criamos alguns *outliers* movendo uma pequena porção de pontos. Especificamente, para produzir esses *outliers*, nós substituímos aleatoriamente um valor da covariável  $x_i$  por  $x_i + 2$  e o logaritmo do tempo de sobrevivência correspondente  $y_i$  por  $y_i + 5$ . O EMV pode ser



fortemente influenciado por esse tipo de desvio e nosso objetivo é avaliar o quão eficaz é o método robusto proposto para limitar a influencia desses *outliers*.

Para encontrar as estimativas pelo método robusto, escolhemos a constante de afinação  $c = 1,4$  para a função  $\psi_c$  de Huber. Esse valor foi escolhido com base em [Sinha \(2018\)](#). Realizamos algumas simulações com valores próximos, mas o valor de  $c = 1,4$  foi finalmente escolhido uma vez que forneceu as melhores estimativas.

As Tabelas de [3.1-3.3](#) apresentam vieses empíricos, erros quadráticos médios, probabilidades de cobertura e comprimentos médios dos intervalos com 95% de confiança para estimação dos parâmetros do modelo BS pelos métodos robusto e por máxima verossimilhança para tamanhos de amostra  $n = 100$  e  $n = 200$  com 0%, 10% e 30% de censura. As simulações foram realizadas para dados sem *outliers*, assim como, com 1, 2 e 4 *outliers*. Está claro na Tabela [3.1](#) que quando não há *outliers*, o método de estimação robusto e o método de estimação por máxima verossimilhança geralmente fornecem estimadores não viesados dos parâmetros do modelo. O ERM, no entanto, apresenta-se como menos eficientes em termos de maiores erros quadráticos médios das estimativas robustas, em comparação com o EMV quando não há *outliers*. Por exemplo, ao estimar o parâmetro de regressão  $\beta_1$  com  $n = 100$ , o ERM produz um erro quadrático médio de 0,1303 sendo maior do que o EMV que produz um erro quadrático médio de 0,0973, como mostrado na Tabela [3.1](#). Esperamos uma pequena perda de eficiência ao usar um método robusto quando, de fato, não há discrepâncias nos dados. No entanto, nosso foco está nos dados contaminados o qual se espera que o método robusto tenha um desempenho melhor que o método de máxima verossimilhança. Através da Tabela [3.1](#) observamos que, no caso de valores discrepantes, o ERM geralmente supera o EMV clássico, onde o EMV produz vieses e erros quadráticos médios maiores e probabilidades de cobertura muito mais baixas (praticamente iguais a zero em quase todos os cenários considerados) em comparação com o ERM. As estimativas por máxima verossimilhança parecem ser fortemente influenciados pelos valores discrepantes. Por exemplo, no caso de um único *outlier* e  $n = 100$ , na Tabela [3.1](#), o EMV de  $\beta_1$  fornece viés de 1,0923, enquanto que o ERM fornece um viés muito menor, de 0,1754. Resultados similares podem ser observados para os dados censurados (Tabelas [3.2](#) e [3.3](#)). Em geral, na ausência de *outliers* quando aumenta-se a porcentagem de censura, o viés e o erro quadrático médio de ambos estimadores aumentam. No entanto, na

presença de *outliers*, o EMV fornece estimativas com vieses e erros quadráticos médios muito maiores do que o ERM, e esses, em geral, aumentam com o aumento da porcentagem de censura.

O resultado das simulações também são apresentadas de forma gráfica (Figuras de 3.1-3.18). Na Figura 3.1 temos os vieses empíricos da estimativa do  $\alpha$  para as três diferentes combinações de tempos de sobrevivência censurados, podemos observar que na ausência de *outliers* o ERM e o EMV fornecem vieses próximos e próximos a zero, na presença de censura e sem *outliers* o EMV apresenta viés um pouco menor comparado ao ERM, porém para conjuntos de dados contendo *outliers*, o EMV fornece vieses bem maiores que o ERM independentemente da porcentagem de censura. Resultados semelhantes podem ser observados para os vieses empíricos das estimativas do  $\beta_0$  e  $\beta_1$  nas Figuras 3.2 e 3.3, respectivamente.

Os EQMs da estimativa do  $\alpha$  para diferentes combinações de tempos de sobrevivência censurados são apresentados na Figura 3.4. Assim como os vieses, o ERM e o EMV fornecem viéses próximos e próximos a zero quando não há *outliers*, porém na presença de *outliers*, o EMV fornece EQMs bem maiores que o ERM. Além disso, para amostras censuradas (Figuras 3.4 (b) e 3.4 (c)), o EMV apresenta EQMs muito maiores comparado a amostras sem censura (Figura 3.4 (a)). Enquanto que o ERM fornece EQMs próximos a zero independentemente da quantidade de *outliers* e/ou porcentagens de censura. Resultados semelhantes podem ser observados para os EQMs das estimativas do  $\beta_0$  e  $\beta_1$  nas Figuras 3.5 e 3.6, respectivamente.

Na Figura 3.7 temos as PCs dos intervalos com 95% de confiança da estimativa do  $\alpha$  para as três diferentes combinações de tempos de sobrevivência censurados. Observe que o EMV, independente da porcentagem de censura, fornece PC próximo a 0,95 quando não há contaminação da amostra com *outliers*. O ERM, devido a pequena perda de eficiência fornece, nesse cenário, PC um pouco menor, porém avaliando os métodos de estimação na presença dos *outliers*, o ERM passa a fornecer melhores PCs comparado ao EMV. Na presença de dados censurados (Figuras 3.7 (b) e 3.7 (c)), temos que um único *outlier* leva a PC do EMV para próximo de zero e igualando-se a zero conforme aumentamos a quantidade de *outliers*. Resultados semelhantes podem ser observados para as PCs das estimativas do  $\beta_0$  e  $\beta_1$  nas Figuras 3.8 e 3.9, respectivamente.

Na Figura 3.10 temos os vieses empíricos da estimativa do  $\alpha$  para as duas combinações de tamanhos amostrais, podemos observar que, para ambos os métodos de estimação, os vieses aumentam conforme aumenta-se a porcentagem de censura, porém para o EMV, esse crescimento é muito maior, na presença dos *outliers*, quando comparado com o ERM. Resultados semelhantes podem ser observados para os vieses empíricos das estimativas do  $\beta_0$  e  $\beta_1$  nas Figuras 3.11 e 3.12, respectivamente. Para a estimativa desses dois parâmetros, o maior tamanho de amostra  $n = 200$ , apresenta vieses um pouco menores.

Os EQMs da estimativa do  $\alpha$  para os tamanhos amostrais  $n = 100$  e  $n = 200$  são apresentados na Figura 3.13. A partir desses gráficos, fica claro que o ERM fornece EQM próximo a zero independentemente da quantidade de *outliers* e/ou de censura, enquanto que o EMV apresenta um aumento do valor dos EQMs conforme aumentamos o número de *outliers* e/ou a porcentagem de censura. Resultados semelhantes podem ser observados para os EQMs das estimativas do  $\beta_0$  e  $\beta_1$  nas Figuras 3.14 e 3.15, respectivamente. Para a estimativa desses dois parâmetros, o maior tamanho de amostra, apresenta EQMs um pouco menores.

Na Figura 3.16 temos as PCs dos intervalos com 95% de confiança da estimativa do  $\alpha$  para as duas combinações de tamanhos amostrais. Novamente podemos observar que na presença dos *outliers* o ERM fornece melhores PCs, independentemente da porcentagem de censura, enquanto que o EMV, principalmente quando as amostras são censuradas, fornece PCs bem próximos a zero. Resultados semelhantes podem ser observados para as PCs das estimativas do  $\beta_0$  e  $\beta_1$  nas Figuras 3.17 e 3.18, respectivamente.

De modo geral, podemos dizer que sempre que um conjunto de dados possuir *outliers*, a inferência baseada no método de estimação robusto será mais confiável em comparação com o método de estimação clássico por máxima verossimilhança. No caso de nenhum *outlier*, o método robusto ainda é tão eficiente quanto o por máxima verossimilhança. Ou seja, o ERM pode ser usado em todos os cenários.

Tabela 3.1: Vieses empíricos, erros quadráticos médios (EQM), probabilidades de cobertura (PCs) e comprimentos médios dos intervalos de confiança de 95% para estimação do modelo Birnbaum-Saunders com os métodos: robusto (ERM) e clássico (EMV). (0% de censura).

$n$	Parâmetros	ERM				EMV			
		Viés	EQM	PC	Comp.	Viés	EQM	PC	Comp.
100	Sem outlier								
	$\alpha$	-0,0025	0,0081	0,8192	0,2426	-0,0117	0,0052	0,9326	0,2740
	$\beta_0$	0,1526	0,0809	0,8180	0,2426	-0,0072	0,0424	0,9442	0,8056
	$\beta_1$	0,0119	0,1303	0,8926	1,1801	0,0138	0,0973	0,9444	1,2184
	1 outlier								
	$\alpha$	0,0153	0,0084	0,8326	0,2503	0,1139	0,0217	0,7074	0,3089
	$\beta_0$	0,0766	0,0548	0,8524	0,7000	-0,5371	0,3222	0,0790	0,5693
	$\beta_1$	0,1754	0,1762	0,8612	1,1999	1,0923	1,2830	0,0040	0,6788
	2 outliers								
	$\alpha$	0,0369	0,0106	0,8132	0,2605	0,1740	0,0404	0,4648	0,3256
	$\beta_0$	-0,0323	0,0786	0,8238	0,7884	-0,7688	0,6243	0,0040	0,6170
	$\beta_1$	0,3839	0,3692	0,7532	1,3220	1,4639	2,2175	0,0000	0,6934
	4 outliers								
	$\alpha$	0,1102	0,0253	0,6268	0,2887	0,2606	0,0788	0,1716	0,3497
	$\beta_0$	-0,2749	0,1712	0,5286	0,7116	-0,9146	0,8644	0,0000	0,6412
	$\beta_1$	0,9522	1,2709	0,3872	1,2507	1,8652	3,5355	0,0000	0,7056
200	Sem outlier								
	$\alpha$	0,0034	0,0040	0,8230	0,1708	-0,0069	0,0025	0,9416	0,1946
	$\beta_0$	0,1572	0,0450	0,7032	0,4754	-0,0017	0,0153	0,9484	0,4843
	$\beta_1$	0,0077	0,0585	0,9020	0,7993	0,0064	0,0437	0,9484	0,8130
	1 outlier								
	$\alpha$	0,0162	0,0043	0,8228	0,1745	0,1091	0,0172	0,5140	0,2176
	$\beta_0$	0,1169	0,0337	0,8052	0,4921	-0,4627	0,2448	0,0778	0,4339
	$\beta_1$	0,0983	0,0628	0,9000	0,8391	0,9840	1,0803	0,0102	0,5756
	2 outliers								
	$\alpha$	0,0228	0,0047	0,8170	0,1767	0,1363	0,0238	0,3656	0,2229
	$\beta_0$	0,0812	0,0341	0,8504	0,5419	-0,6270	0,4155	0,0020	0,4274
	$\beta_1$	0,1735	0,1167	0,8508	0,9695	1,3073	1,7796	0,0000	0,5434
	4 outliers								
	$\alpha$	0,0459	0,0068	0,7406	0,1846	0,1883	0,0407	0,1416	0,2331
	$\beta_0$	-0,0062	0,0317	0,8608	0,5320	-0,7086	0,5173	0,0000	0,4202
	$\beta_1$	0,3861	0,2663	0,6816	1,0216	1,5520	2,4525	0,0000	0,5059

Parâmetros fixados em  $\alpha = 1$ ,  $\beta_0 = 4$  e  $\beta_1 = -2$ .

Tabela 3.2: Vieses empíricos, erros quadráticos médios (EQM), probabilidades de cobertura (PCs) e comprimentos médios dos intervalos de confiança de 95% para estimação do modelo Birnbaum-Saunders com os métodos: robusto (ERM) e clássico (EMV). (10% de censura).

$n$	Parâmetros	ERM				EMV			
		Viés	EQM	PC	Comp.	Viés	EQM	PC	Comp.
100	Sem outlier								
	$\alpha$	0,0375	0,0108	0,8890	0,3123	-0,0124	0,0054	0,9384	0,2921
	$\beta_0$	0,2330	0,0906	0,7828	0,7582	0,0036	0,0272	0,9642	0,6994
	$\beta_1$	-0,0797	0,1322	0,9046	1,2325	-0,0027	0,0912	0,9536	1,2234
	1 outlier								
	$\alpha$	0,0686	0,0151	0,8546	0,3310	0,4332	0,2063	0,0224	0,4268
	$\beta_0$	0,2221	0,0916	0,8182	0,8228	-1,1847	1,4396	0,0000	0,6780
	$\beta_1$	-0,0096	0,1663	0,9086	1,4002	2,6646	7,2217	0,0000	0,7262
	2 outliers								
	$\alpha$	0,0936	0,0196	0,8192	0,3434	0,5818	0,3565	0,0008	0,4719
	$\beta_0$	0,2113	0,0875	0,8402	0,8374	-1,3736	1,9156	0,0000	0,7308
	$\beta_1$	0,0591	0,1421	0,9178	1,3202	2,8954	8,4524	0,0000	0,6985
	4 outliers								
	$\alpha$	0,1578	0,0460	0,7090	0,4055	0,6949	0,5012	0,0000	0,5049
	$\beta_0$	0,0618	0,0646	0,9130	0,8541	-1,5624	2,4678	0,0000	0,7898
	$\beta_1$	0,4368	0,5594	0,8050	1,5801	3,6222	13,1808	0,0000	0,8147
200	Sem outlier								
	$\alpha$	0,0511	0,0075	0,8252	0,2213	-0,0021	0,0027	0,9482	0,2087
	$\beta_0$	0,2601	0,0879	0,5778	0,5789	0,0212	0,0162	0,9582	0,5276
	$\beta_1$	-0,1095	0,0816	0,8890	0,9269	-0,0255	0,0526	0,9528	0,9040
	1 outlier								
	$\alpha$	0,0641	0,0092	0,7872	0,2286	0,3415	0,1276	0,0096	0,2823
	$\beta_0$	0,2341	0,0732	0,6316	0,5583	-0,9634	0,9527	0,0000	0,4441
	$\beta_1$	-0,0490	0,0684	0,9142	0,9077	2,1318	4,6377	0,0000	0,4629
	2 outliers								
	$\alpha$	0,0774	0,0112	0,7306	0,2343	0,7009	0,5095	0,0022	0,3717
	$\beta_0$	0,2447	0,0793	0,6108	0,5690	-1,3449	1,8390	0,0018	0,5839
	$\beta_1$	-0,0462	0,0666	0,9112	0,8834	2,9734	8,9336	0,0004	0,5850
	4 outliers								
	$\alpha$	0,1111	0,0181	0,5778	0,2465	0,6686	0,4642	0,0022	0,3604
	$\beta_0$	0,2418	0,0852	0,7098	0,6650	-1,5046	2,2912	0,0038	0,5713
	$\beta_1$	0,0251	0,0830	0,9214	1,0112	3,0945	9,6507	0,0000	0,5662

Parâmetros fixados em  $\alpha = 1$ ,  $\beta_0 = 4$  e  $\beta_1 = -2$ .

Tabela 3.3: Vieses empíricos, erros quadráticos médios (EQM), probabilidades de cobertura (PCs) e comprimentos médios dos intervalos de confiança de 95% para estimação do modelo Birnbaum-Saunders com os métodos: robusto (ERM) e clássico (EMV). (30% de censura).

$n$	Parâmetros	ERM				EMV			
		Viés	EQM	PC	Comp.	Viés	EQM	PC	Comp.
100	Sem outlier								
	$\alpha$	0,0766	0,0227	0,8776	0,4046	0,0053	0,0071	0,9560	0,3447
	$\beta_0$	0,4000	0,2210	0,7928	1,1617	0,0779	0,0476	0,9612	0,8799
	$\beta_1$	-0,2311	0,2278	0,9000	1,5555	-0,0585	0,1218	0,9478	1,3591
	1 outlier								
	$\alpha$	0,1139	0,0314	0,9076	0,4706	0,5376	0,3127	0,0148	0,5501
	$\beta_0$	0,3364	0,1681	0,8632	1,1527	-1,1131	1,2686	0,0000	0,8558
	$\beta_1$	-0,1117	0,1974	0,9222	1,6046	2,6095	6,9102	0,0000	0,8276
	2 outliers								
	$\alpha$	0,1677	0,0539	0,9170	0,9563	0,6969	0,5115	0,0004	0,6151
	$\beta_0$	0,2923	0,1453	0,9102	1,7144	-1,2583	1,6093	0,0000	0,9706
	$\beta_1$	0,0351	0,2084	0,9354	1,7870	3,0853	9,6020	0,0000	0,9393
	4 outliers								
	$\alpha$	0,2320	0,0914	0,9076	2,1706	0,6903	0,4944	0,0000	0,6113
	$\beta_0$	0,2748	0,1315	0,9490	3,3800	-1,2149	1,4959	0,0000	0,8901
	$\beta_1$	0,2022	0,2236	0,9164	2,0422	2,9364	8,6597	0,0000	0,7545
200	Sem outlier								
	$\alpha$	0,0723	0,0132	0,8046	0,2748	0,0021	0,0035	0,9574	0,2424
	$\beta_0$	0,3367	0,1386	0,6086	0,7545	0,0370	0,0187	0,9698	0,5803
	$\beta_1$	-0,1862	0,1154	0,8860	1,0698	-0,0289	0,0557	0,9540	0,9485
	1 outlier								
	$\alpha$	0,1349	0,0288	0,6402	0,3285	0,6111	0,4005	0,0268	0,4601
	$\beta_0$	0,4453	0,2323	0,5168	0,9052	-1,1111	1,2685	0,0116	0,7433
	$\beta_1$	-0,2234	0,1440	0,8836	1,1717	2,5094	6,4318	0,0054	0,7699
	2 outliers								
	$\alpha$	0,1191	0,0236	0,7044	0,3188	0,5251	0,2892	0,0006	0,3944
	$\beta_0$	0,3251	0,1327	0,6808	0,7949	-1,0850	1,1924	0,0006	0,5908
	$\beta_1$	-0,1069	0,1056	0,9260	1,1439	2,5858	6,7422	0,0002	0,5673
	4 outliers								
	$\alpha$	0,1696	0,0412	0,5672	0,4995	0,7547	0,5854	0,0002	0,4561
	$\beta_0$	0,3428	0,1451	0,6858	1,0357	-1,1932	1,4368	0,0004	0,6952
	$\beta_1$	-0,0495	0,0877	0,9356	1,1582	2,8577	8,2039	0,0000	0,6091

Parâmetros fixados em  $\alpha = 1$ ,  $\beta_0 = 4$  e  $\beta_1 = -2$ .

Figura 3.1: Vieses empíricos da estimativa do  $\alpha$  para diferentes combinações de tempos de sobrevivência censurados: 0% (a), 10% (b) e 30% (c).

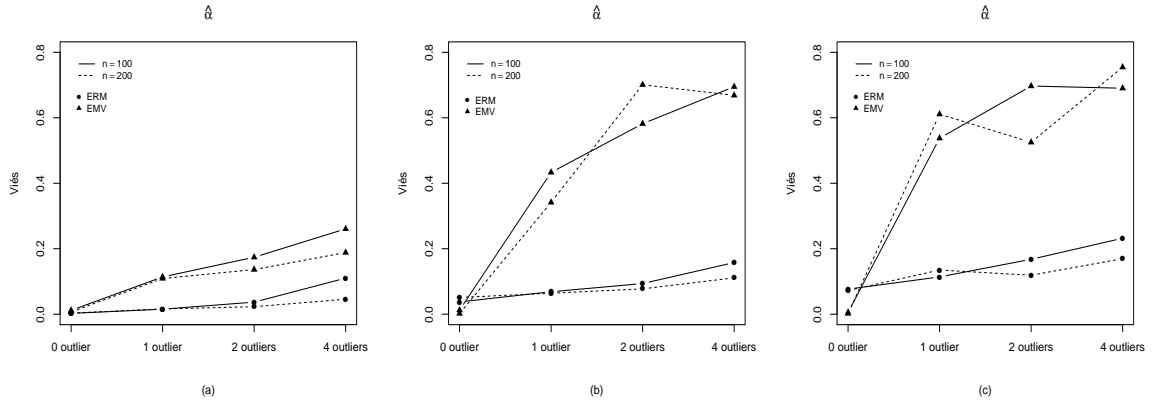


Figura 3.2: Vieses empíricos da estimativa do  $\beta_0$  com três diferentes combinações de tempos de sobrevivência censurados: 0% (a), 10% (b) e 30% (c).

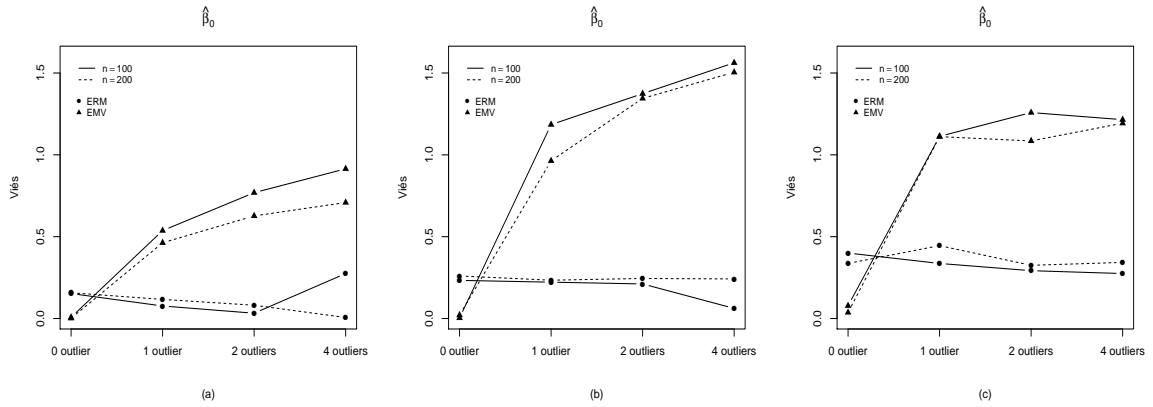
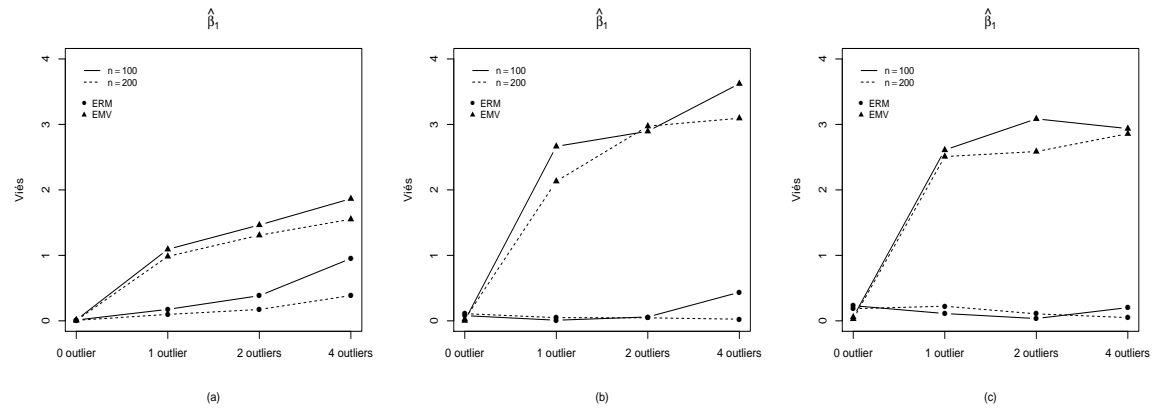


Figura 3.3: Vieses empíricos da estimativa do  $\beta_1$  com três diferentes combinações de tempos de sobrevivência censurados: 0% (a), 10% (b) e 30% (c).



Fonte: Autoria própria.

Figura 3.4: Erros quadráticos médios (EQM) da estimativa do  $\alpha$  para diferentes combinações de tempos de sobrevivência censurados: 0% (a), 10% (b) e 30% (c).

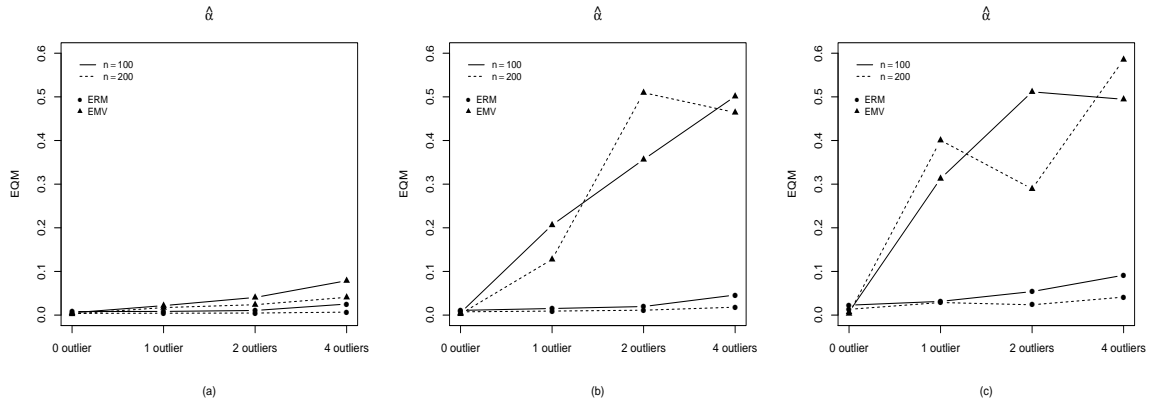


Figura 3.5: Erros quadráticos médios (EQM) da estimativa do  $\beta_0$  com três diferentes combinações de tempos de sobrevivência censurados: 0% (a), 10% (b) e 30% (c).

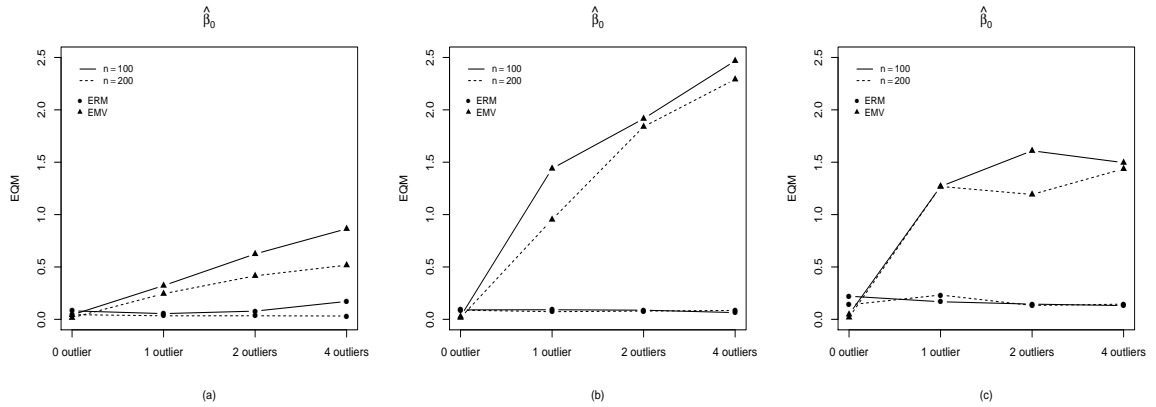
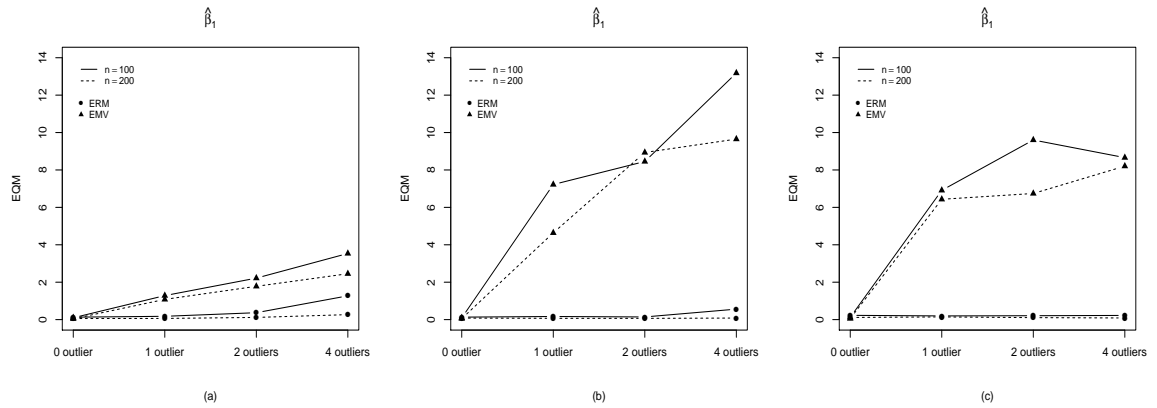


Figura 3.6: Erros quadráticos médios (EQM) da estimativa do  $\beta_1$  com três diferentes combinações de tempos de sobrevivência censurados: 0% (a), 10% (b) e 30% (c).



Fonte: Autoria própria.



Figura 3.7: Probabilidades de cobertura (PCs) da estimativa do  $\alpha$  para diferentes combinações de tempos de sobrevivência censurados: 0% (a), 10% (b) e 30% (c).

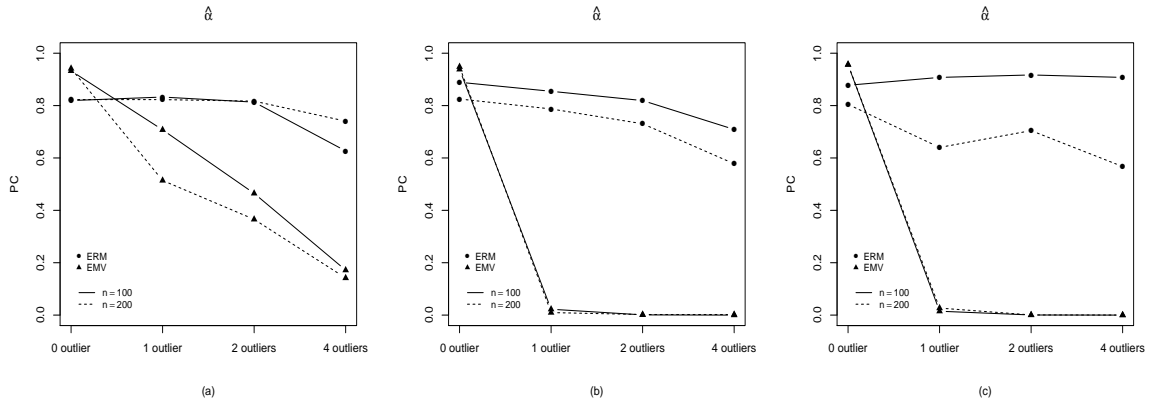


Figura 3.8: Probabilidades de cobertura (PCs) da estimativa do  $\beta_0$  com três diferentes combinações de tempos de sobrevivência censurados: 0% (a), 10% (b) e 30% (c).

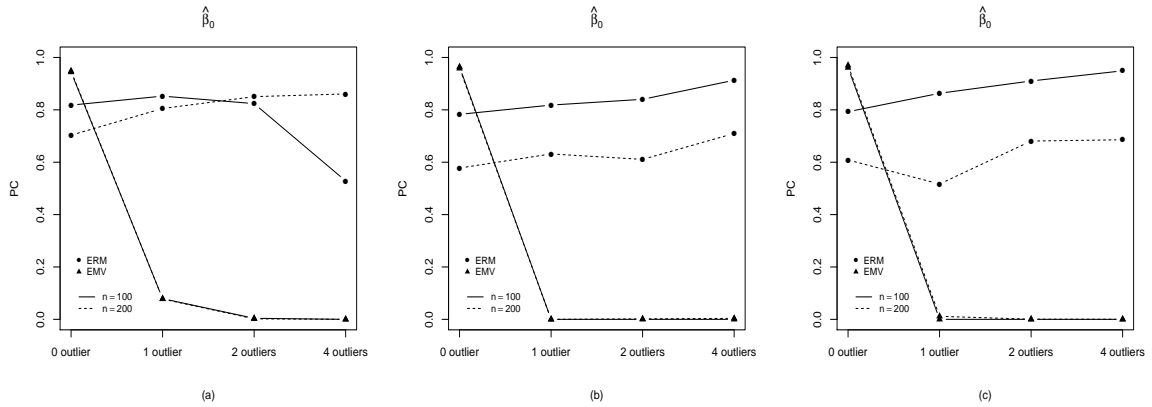
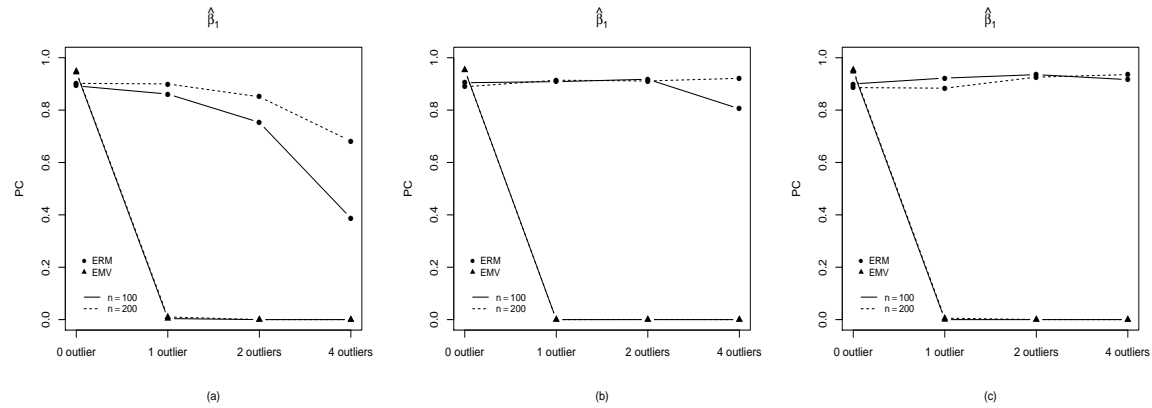


Figura 3.9: Probabilidades de cobertura (PCs) da estimativa do  $\beta_1$  com diferentes combinações de tempos de sobrevivência censurados: 0% (a), 10% (b) e 30% (c).



Fonte: Autoria própria.

Figura 3.10: Vieses empíricos da estimativa do  $\alpha$  para diferentes combinações de tamanhos amostrais:  $n = 100$  (a) e  $n = 200$  (b).

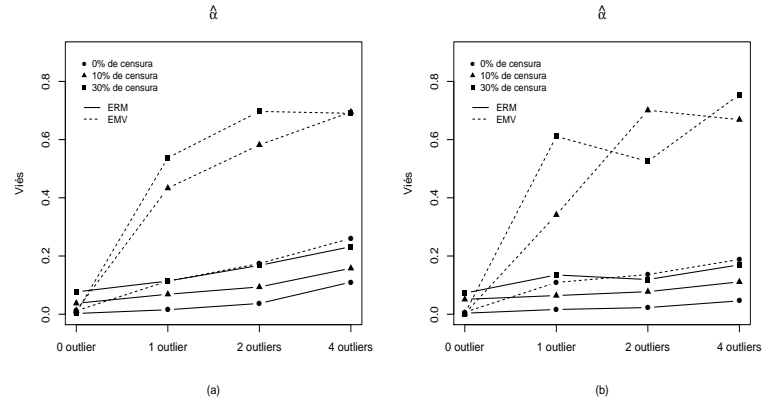


Figura 3.11: Vieses empíricos da estimativa do  $\beta_0$  para diferentes combinações de tamanhos amostrais:  $n = 100$  (a) e  $n = 200$  (b).

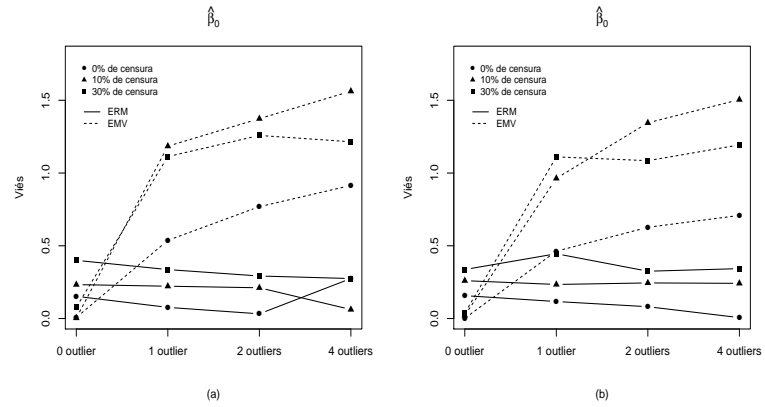
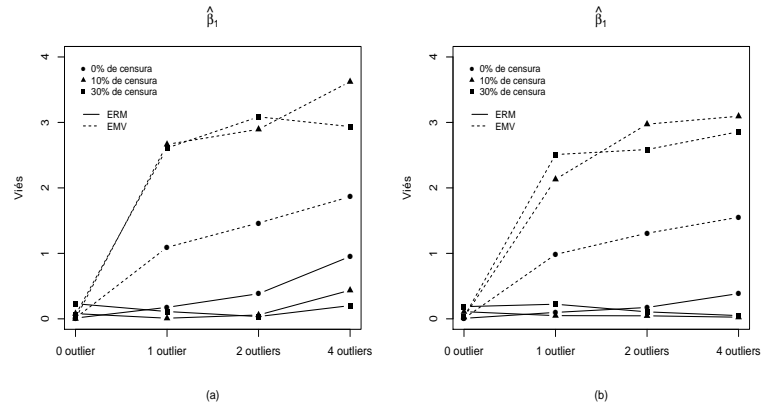


Figura 3.12: Vieses empíricos da estimativa do  $\beta_1$  para diferentes combinações de tamanhos amostrais:  $n = 100$  (a) e  $n = 200$  (b).



Fonte: Autoria própria.

Figura 3.13: Erros quadráticos médios (EQM) da estimativa do  $\alpha$  para diferentes combinações de tamanhos amostrais:  $n = 100$  (a) e  $n = 200$  (b).

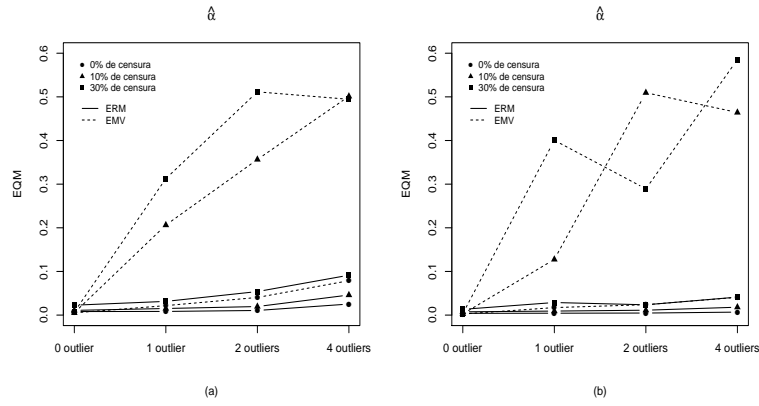


Figura 3.14: Erros quadráticos médios (EQM) da estimativa do  $\beta_0$  para diferentes combinações de tamanhos amostrais:  $n = 100$  (a) e  $n = 200$  (b).

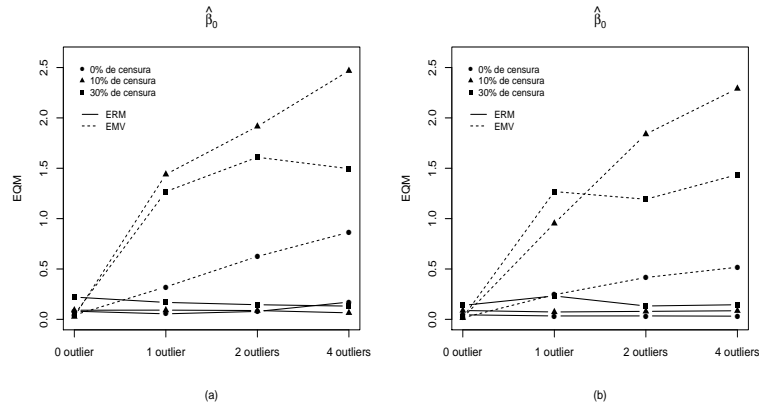
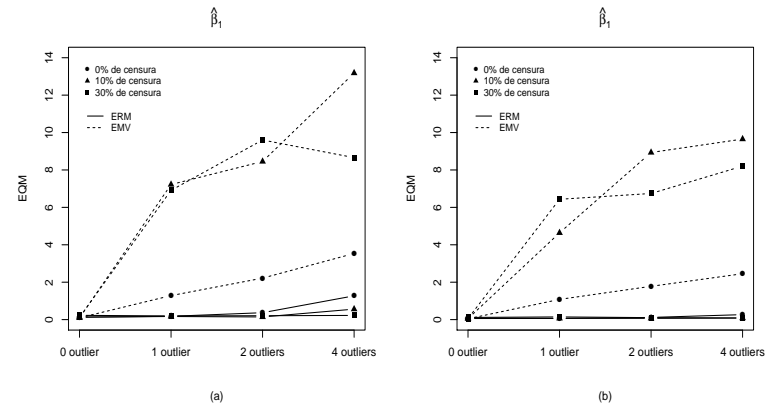


Figura 3.15: Erros quadráticos médios (EQM) da estimativa do  $\beta_1$  para diferentes combinações de tamanhos amostrais:  $n = 100$  (a) e  $n = 200$  (b).



Fonte: Autoria própria.

Figura 3.16: Probabilidades de cobertura (PCs) da estimativa do  $\alpha$  para diferentes combinações de tamanhos amostrais:  $n = 100$  (a) e  $n = 200$  (b).

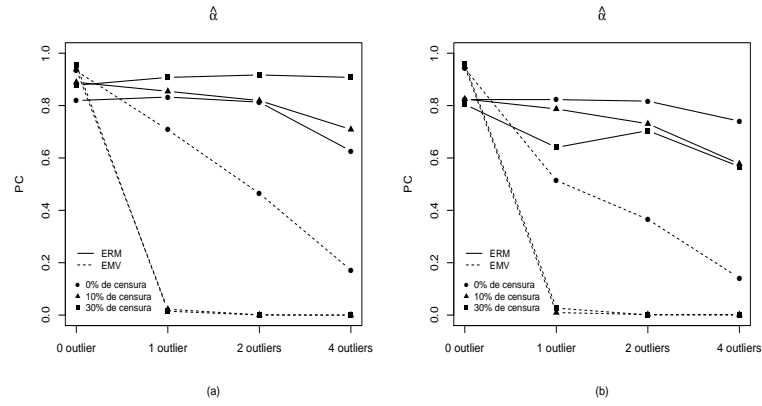


Figura 3.17: Probabilidades de cobertura (PCs) da estimativa do  $\beta_0$  para diferentes combinações de tamanhos amostrais:  $n = 100$  (a) e  $n = 200$  (b).

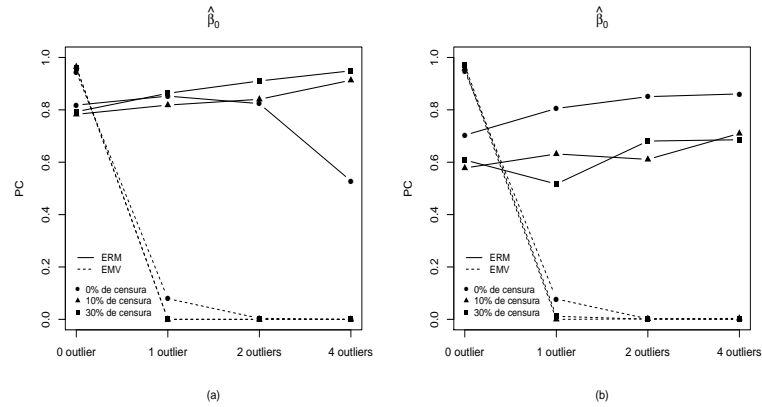
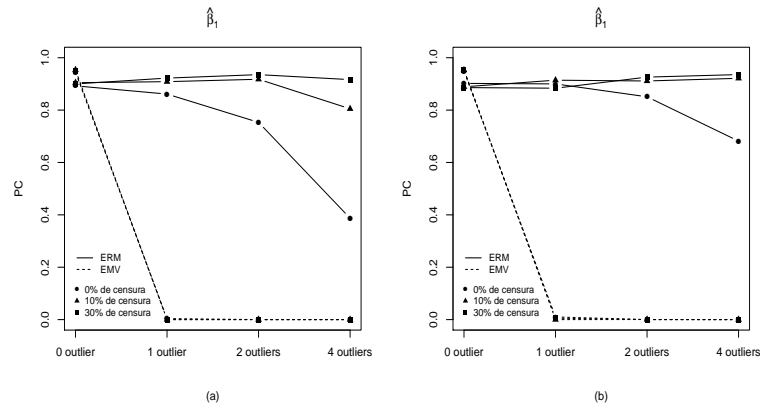


Figura 3.18: Probabilidades de cobertura (PCs) da estimativa do  $\beta_1$  para diferentes combinações de tamanhos amostrais:  $n = 100$  (a) e  $n = 200$  (b).



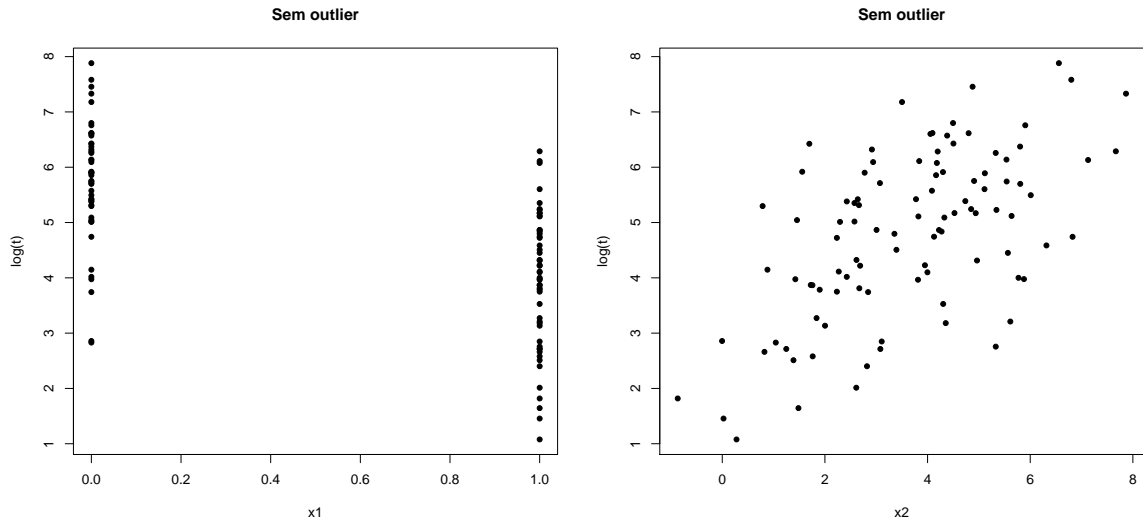
Fonte: Autoria própria.

## 3.2 Exemplo simulado

Para fins ilustrativos, usaremos um exemplo simulado no qual também poderemos avaliar o desempenho do ERM. Neste exemplo, utilizaremos uma covariável binária  $x_1$  usada como indicador de grupo e uma covariável contínua  $x_2$  que segue distribuição normal com média 4 e variância 4. Os tempos de sobrevivência  $t_i$  foram gerados a partir do modelo BS com parâmetros fixados em  $\alpha = 1$ ,  $\beta_0 = 4$ ,  $\beta_1 = -2$  e  $\beta_2 = 0,5$ , tamanho de amostra  $n = 100$  e 10% de censura.

Os gráficos da Figura 3.19 exibem o conjunto de dados simulados sem pontos discrepantes.

Figura 3.19: Gráficos dos tempos de sobrevivência contra as covariáveis  $x_1$  e  $x_2$  para o conjunto de dados simulados sem *outlier*.



Fonte: Autoria própria.

O conjunto de dados não contaminados mencionado acima foi ajustado usando o modelo de tempo de falha acelerado log-BS, ou seja,

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i, \quad i = 1, \dots, 100, \quad (3.6)$$

em que  $\epsilon_i \sim \text{SN}(\alpha, 0, 2)$  e  $y_i = \log(t_i)$ . A Tabela 3.4 apresenta as estimativas do modelo e os respectivos erros padrão tanto para o método robusto quanto para o método clássico. A constante de afinação escolhida para o método robusto foi  $c = 1,4$  seguindo

Tabela 3.4: Estimativa dos parâmetros do modelo log-BS para os dados simulados sem *outliers* utilizando o ERM e o EMV.

Parâmetros	ERM		EMV	
	Estimativa	Erro Padrão	Estimativa	Erro Padrão
$\alpha$	0,978	0,072	0,975	0,073
$\beta_0$	4,306	0,302	4,124	0,227
$\beta_1$	-1,870	0,205	-1,805	0,181
$\beta_2$	0,477	0,062	0,471	0,052

Parâmetros fixados em  $\alpha = 1$ ,  $\beta_0 = 4$ ,  $\beta_1 = -2$  e  $\beta_2 = 0,5$ .

a sugestão de [Sinha \(2018\)](#) no qual afirma garantir pelo menos 90% de eficiência nas distribuições subjacentes. Como esperado, os métodos fornecem valores de estimativas e erros padrão muito próximos.

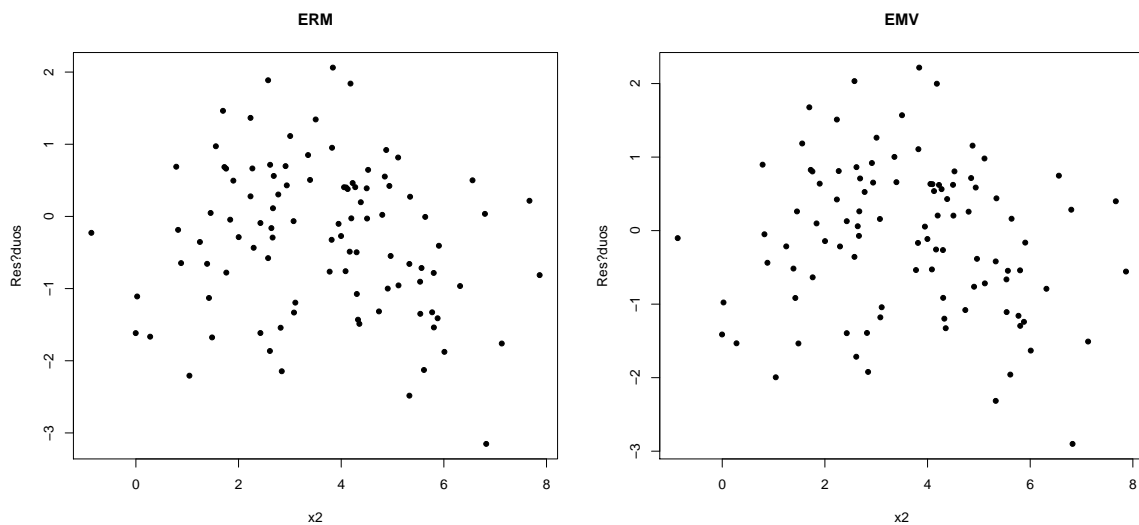
Na Figura 3.20 temos os gráficos dos resíduos padronizados,  $r_i$ , definidos por

$$r_i = \frac{\hat{\epsilon}_i - E(\hat{\epsilon}_i)}{\sqrt{Var(\hat{\epsilon}_i)}} \quad (3.7)$$

baseados nas estimativas robustas e de máxima verossimilhança contra a variável  $x_2$ . A aleatoriedade desses resíduos indica o bom desempenho de ambos os ajustes.

Em seguida obtivemos os resíduos quantílicos discutido em [Barros e Paula \(2019\)](#) úteis para a construção dos gráficos normais de probabilidade no qual utilizaremos para

Figura 3.20: Gráficos dos resíduos  $r_i$  baseados no ERM e EMV ajustados aos dados simulados sem *outliers*.



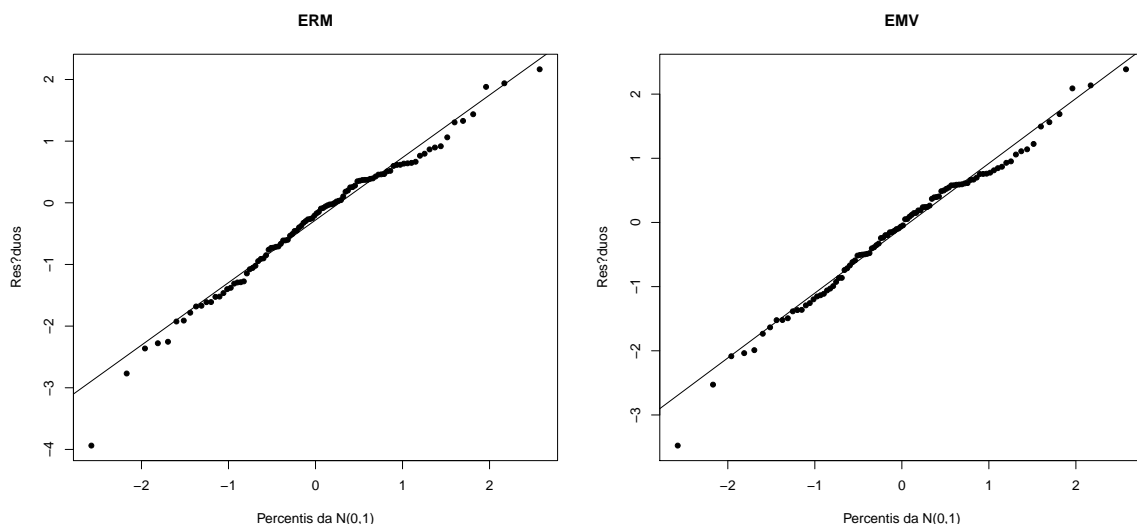
Fonte: Autoria própria.

avaliar a adequação do modelo postulado ao ajustar o conjunto de dados. O resíduo quantílico é definido por  $r_{q_i} = \Phi^{-1}\{F(t_i|\hat{\theta})\}$  em que  $\Phi(\cdot)$  denota a fda da  $N(0, 1)$ , com  $i = 1, \dots, n$ . Para o modelo de regressão log-BS, os resíduos quantílicos ficam dados por

$$r_{q_i} = \frac{2}{\hat{\alpha}} \sinh\left(\frac{\hat{\epsilon}_i}{2}\right). \quad (3.8)$$

Com esses resíduos construímos os gráficos normais de probabilidade exibidos na Figura 3.21. Nota-se claramente a semelhança destes gráficos para ambos os métodos de estimação, além de não indicarem afastamento da suposição de que os erros tem distribuição log-BS.

Figura 3.21: Gráficos normais de probabilidade para os resíduos  $r_{q_i}$  baseados no ERM e EMV ajustados aos dados simulados sem *outliers*.

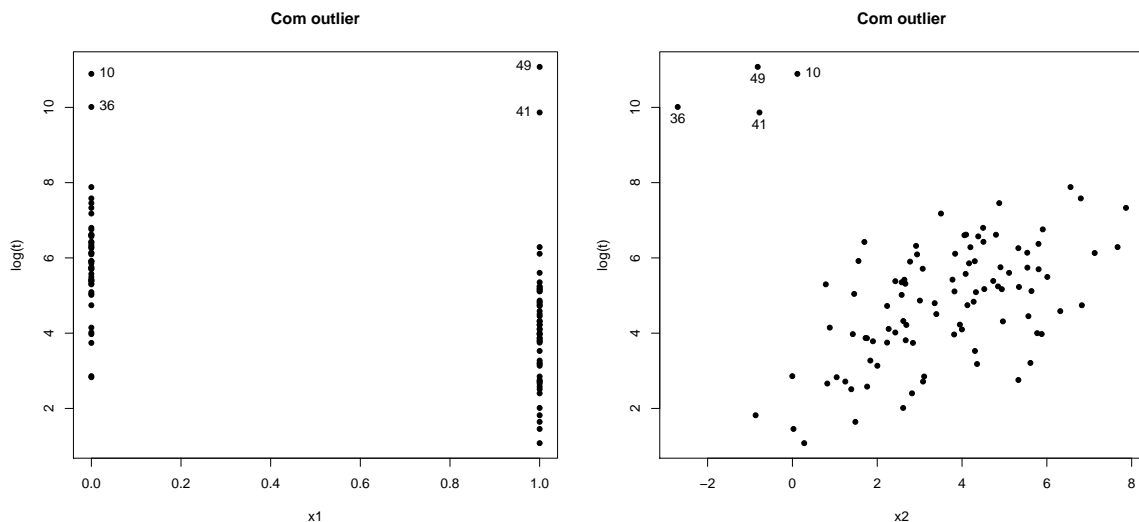


Fonte: Autoria própria.

No intuito de avaliarmos os métodos de estimação na presença de valores extremos, criamos alguns *outliers* no conjunto de dados simulados da Figura 3.19. Para produzirmos esses *outliers*, substituímos aleatoriamente quatro valores da covariável  $x_2$  por  $x_2 - 5$  e os correspondentes logaritmos dos tempos de sobrevivência  $y_i$  por  $y_i + 5$ .

Os gráficos da Figura 3.22 exibem os dados contaminados com quatro valores extremos que estão distantes da maior parte dos dados. A Tabela 3.5 apresenta as estimativas obtidas pelo método robusto e pelo método de máxima verossimilhança

Figura 3.22: Gráficos dos tempos de sobrevivência contra as covariáveis  $x_1$  e  $x_2$  para o conjunto de dados simulados com *outliers*.



Fonte: Autoria própria.

com os respectivos erros padrão dos parâmetros do modelo para o conjunto de dados contaminado. Como esperado, as estimativas são bastante diferentes, uma vez que as estimativas por máxima verossimilhança foram fortemente afetadas pelos *outliers* considerados nos dados.

Tabela 3.5: Estimativa dos parâmetros do modelo log-BS para os dados simulados com *outliers* utilizando o ERM e o EMV.

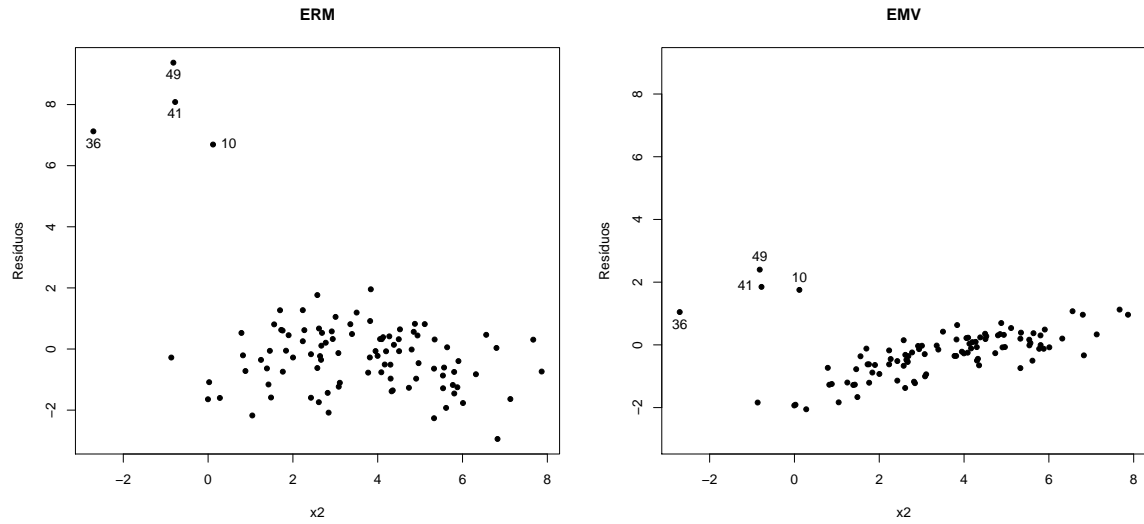
Parâmetros	ERM		EMV	
	Estimativa	Erro Padrão	Estimativa	Erro Padrão
$\alpha$	1,055	0,082	3,181	0,257
$\beta_0$	4,433	0,300	7,083	0,316
$\beta_1$	-1,950	0,199	-1,448	0,372
$\beta_2$	0,458	0,062	-0,236	0,130

Parâmetros fixados em  $\alpha = 1$ ,  $\beta_0 = 4$ ,  $\beta_1 = -2$  e  $\beta_2 = 0,5$ .

A Figura 3.23 mostra os gráficos dos resíduos  $r_i$  obtidos através do ERM e do EMV contra a variável  $x_2$ , cada um deles exibe quatro resíduos altos, que correspondem aos *outliers* nos dados contaminados. É interessante notar que embora os *outliers* sejam claramente representados em ambos os gráficos, esses valores são ampliados pelo método robusto em maior escala comparado ao de máxima verossimilhança. Além disso, o EMV exibe claramente uma tendência positiva na parcela residual em relação

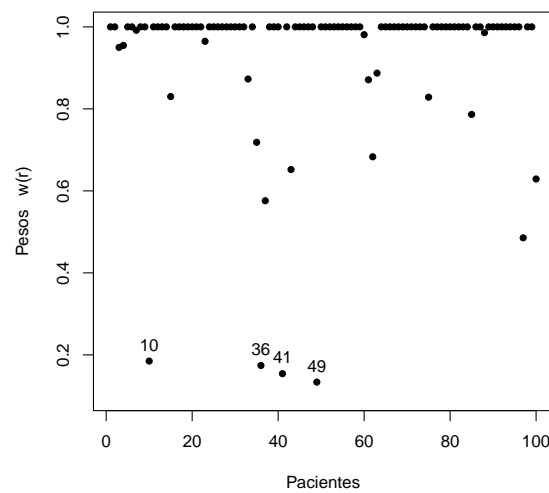


Figura 3.23: Gráficos dos resíduos  $r_i$  baseados no ERM e EMV ajustados aos dados simulados com *outliers*.



Fonte: Autoria própria.

Figura 3.24: Gráficos dos pesos  $w(r_i)$  utilizado no ERM para ajustar os dados simulados com *outliers*.



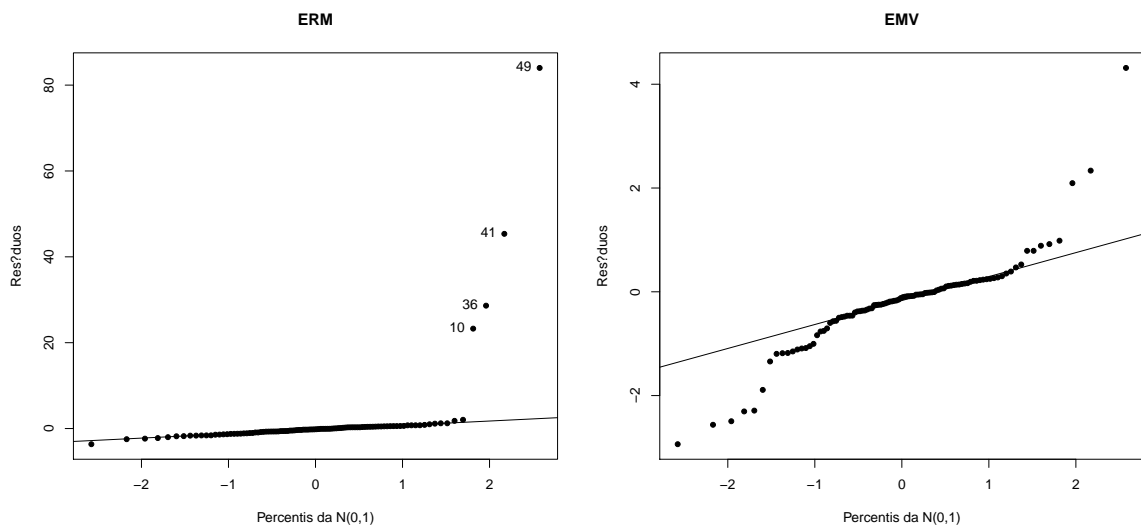
Fonte: Autoria própria.

a covariável  $x_2$ , o que indica claramente uma deficiência do ajuste de máxima verossimilhança sobre o ajuste robusto proposto. A Figura 3.24 mostra a função de peso,  $w(r_i)$ , obtida a partir do ajuste usando o ERM, o qual podemos observar claramente que as quatro observações destacadas com resíduo alto e que correspondem aos *outliers*

criados, recebem pesos inferiores a 0,3.

Na Figura 3.25 temos os gráficos normais de probabilidade construídos com os resíduos  $r_{q_i}$ . Para o ajuste do conjunto de dados contaminados, os métodos de estimação utilizados apresentam esses gráficos de forma bastante diferente. A Figura 3.25 (a) referente ao método robusto mostra um bom ajuste para a maior parte dos dados, como evidenciado pelo comportamento linear na parte inferior do gráfico. Ao mesmo tempo, revela claramente os valores discrepantes que foram evidenciados no gráfico dos resíduos  $r_i$  na Figura 3.23. Por outro lado, a Figura 3.25 (b) referente ao EMV mostra um pequeno comportamento linear apenas no meio do gráfico e com várias discrepâncias nos valores não centrais levando a uma conclusão de que o modelo não está bem ajustado ao conjunto de dados e que seria necessário a utilização de uma distribuição com caudas mais pesadas para os erros.

Figura 3.25: Gráficos normais de probabilidade para os resíduos  $r_{q_i}$  baseados no ERM e EMV para os dados simulados com *outliers*.



Fonte: Autoria própria.

Por fim, realizamos o ajuste do modelo (3.6) pelo método de máxima verossimilhança para os dados simulados retirando os quatro *outliers* que foram criados para que possamos comparar as estimativas obtidas com o método robusto com os dados completos, cujas estimativas e respectivos erros padrão se encontram na Tabela 3.6. Nota-se que ambos os métodos de estimação apresentam valores próximos, ou seja, o ERM proposto ajusta bem o conjunto de dados sem a necessidade de excluir observações

discrepantes.

Tabela 3.6: Comparação das estimativa dos parâmetros do modelo log-BS. O ERM para os dados com *outliers* e o EMV para os dados após retirada dos *outliers*.

Parâmetros	ERM com <i>outliers</i>		EMV sem <i>outliers</i>	
	Estimativa	Erro Padrão	Estimativa	Erro Padrão
$\alpha$	1,055	0,082	0,962	0,073
$\beta_0$	4,433	0,300	4,127	0,225
$\beta_1$	-1,950	0,199	-1,846	0,182
$\beta_2$	0,458	0,062	0,469	0,051

Parâmetros fixados em  $\alpha = 1$ ,  $\beta_0 = 4$ ,  $\beta_1 = -2$  e  $\beta_2 = 0,5$ .



# Capítulo 4

## Aplicações

Neste capítulo ilustraremos a teoria estudada aplicando-a a duas situações reais de estudo.

### 4.1 Aplicação 1

#### 4.1.1 Estudo sobre a vida útil de uma peça de metal

Para esta aplicação consideramos o conjunto de dados não censurados de fadiga biaxial relatado por Brown e Miller apresentado em [Rieck e Nedelman \(1991\)](#) e exposto na Tabela 4.1 sobre a vida útil de uma peça de metal. A variável resposta,  $N$ , é o número de ciclos até a ocorrência da falha e a variável explicativa,  $W$ , é o trabalho

Tabela 4.1: Dados de fadiga biaxial de Brown e Miller

Peça	$N$	$W$	Peça	$N$	$W$	Peça	$N$	$W$	Peça	$N$	$W$
1	3280	11,5	13	804	24,0	25	750	40,1	37	212	60,5
2	5046	13,0	14	1093	24,6	26	316	40,1	38	327	62,1
3	1563	14,3	15	1125	25,2	27	456	43,0	39	373	62,8
4	4707	15,6	16	884	25,5	28	552	44,1	40	125	66,5
5	977	16,0	17	1300	26,3	29	355	46,5	41	187	67,0
6	2834	17,3	18	852	27,9	30	242	47,3	42	135	67,1
7	2266	19,3	19	580	28,3	31	190	48,7	43	245	67,9
8	2208	21,1	20	1066	28,4	32	127	52,9	44	137	68,8
9	1040	21,5	21	1114	28,6	33	185	56,9	45	200	75,4
10	700	22,6	22	386	30,9	34	255	59,9	46	190	100,5
11	1583	22,6	23	745	31,9	35	195	60,2			
12	482	24,0	24	736	34,5	36	283	60,3			

por ciclo expresso na unidade de medida  $\text{mJ}/\text{m}^3$ . Quarenta e seis observações foram consideradas.

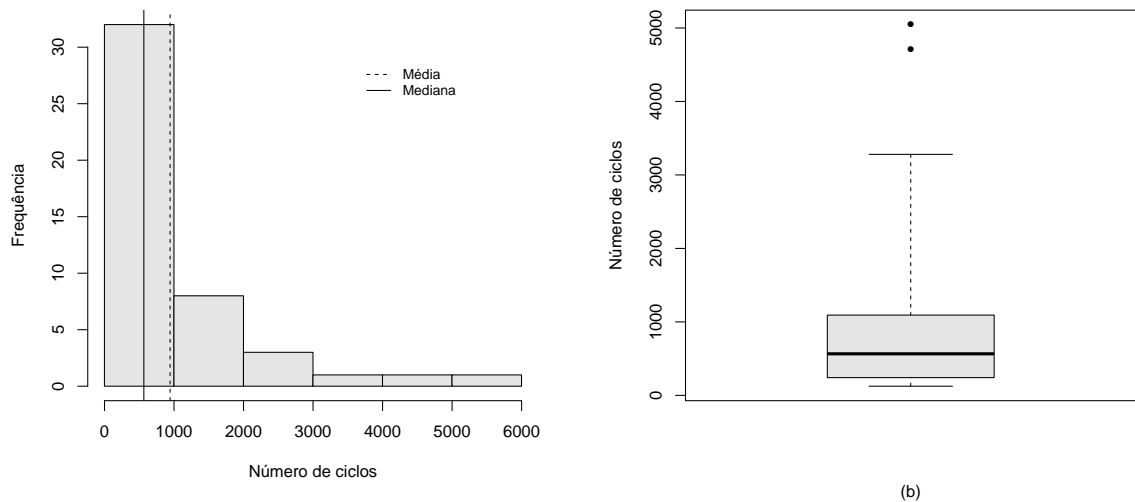
O objetivo do estudo foi modelar o número de ciclos até que ocorra a falha da peça de metal. Inicialmente, realizamos uma análise descritiva da variável resposta. Na Tabela 4.5 encontram-se algumas medidas resumo relacionadas ao número de ciclos. Nota-se que  $N$  assume valores entre 125 e 5046. O coeficiente de variação (CV) é igual a 117,76% indicando que a média de 943,7 não é representativa. O terceiro quartil,  $Q_3$ , revela que 75% dos números de ciclos são inferiores a 1086,2.

Tabela 4.2: Medidas resumo da variável  $N$ .

Mínimo	1° Quartil	Mediana	Média	3° Quartil	Máximo	Desvio Padrão	CV
125	242,8	566	943,7	1086,2	5046	1111,3	117,76%

Na Figura 4.1 tem-se o histograma de frequência e o boxplot da variável resposta. A Figura 4.1 (a) mostra um comportamento assimétrico em que a maior parte das observações se encontram com o número de ciclos no intervalo de 0 a 1000. Destacamos que a média dos dados, por pouco, não sai do intervalo onde estão a maior parte das observações confirmando que a média é pouco representativa, ao contrário da mediana. Isso ocorre devido a presença de alguns valores extremos. O comportamento assimétrico

Figura 4.1: Histograma e boxplot para o número de ciclos até a ocorrência da falha.



Fonte: Autoria própria.

também é observado na figura 4.1 (b) além de identificar as observações extremas que possuem números de ciclos superiores a 4700, a saber: observações #2 e #4.

### 4.1.2 Ajuste do modelo log-BS

Com base na natureza dos dados e na discussão de Rieck e Nedelman (1991), consideraremos o o modelo log-BS na forma

$$\log(N_i) = \beta_0 + \beta_1 \log(W_i) + \epsilon_i, \quad i = 1, \dots, 46, \quad (4.1)$$

em que  $\epsilon_i$  são variáveis aleatórias i.i.d. tais que  $\epsilon_i \sim \text{SN}(\alpha, 0, 2)$ . Os valores das estimativas dos parâmetros do modelo, bem como o os erros padrão e os respectivos  $p$ -valores obtidos do ajuste utilizando-se dos métodos de estimação robusto e por máxima verossimilhança estão dispostos na Tabela 4.3. Temos para esse estudo, que o ERM possui desempenho semelhante ao EMV, isso porque para esse banco de dados não há presença de observações com resíduo alto como veremos a seguir.

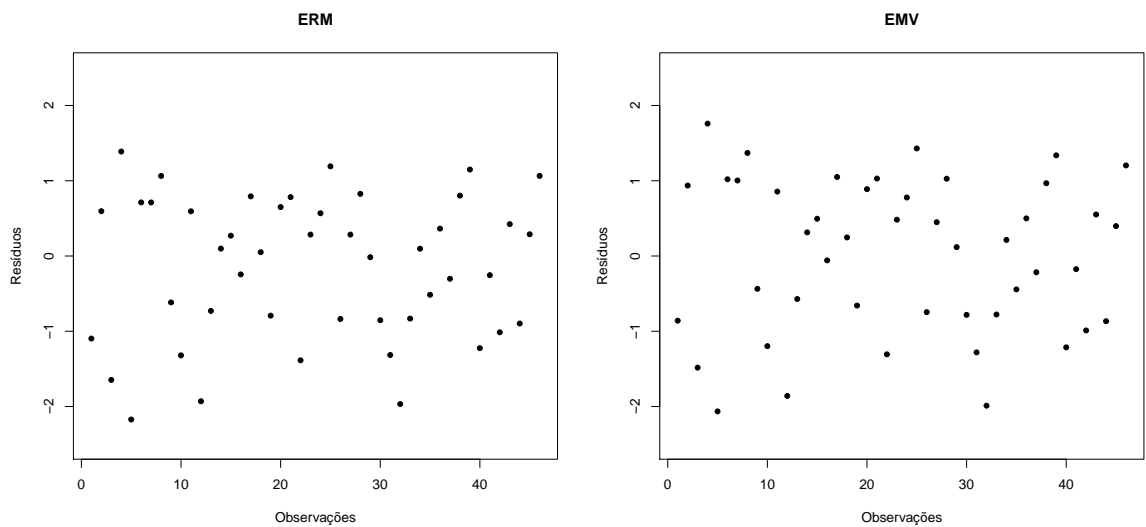
Tabela 4.3: Estimativa dos parâmetros do modelo log-BS para os dados sobre a vida útil de uma peça de metal utilizando o ERM e o EMV.

Parâmetros	ERM			EMV		
	Estimativa	Erro Padrão	$p$ -valor	Estimativa	Erro Padrão	$p$ -valor
$\alpha$	0,438	0,037	0,000	0,410	0,043	0,000
$\beta_0$	12,458	0,440	0,000	12,279	0,389	0,000
$\beta_1$	-1,704	0,121	0,000	-1,671	0,108	0,000

Nas Figuras 4.2 e 4.3 são apresentados os gráficos dos resíduos  $r_i$  na forma (3.7) e  $r_{q_i}$  na forma (3.8), respectivamente, baseados no ERM e no EMV. É possível notar, em ambas as figuras, que embora identificados valores extremos na análise descritiva desse estudo, nenhuma das observações foi destacada com alto resíduo para esse conjunto de dados. Além disso, esses resíduos possuem um comportamento aleatório em ambos os métodos de estimação. Na Figura 4.4 temos os gráficos normais de probabilidade construídos com os resíduos  $r_{q_i}$  obtidos através de ambos métodos de estimação. Nota-se que independente do método utilizado, os gráficos fornecem de forma semelhante um bom ajuste do modelo indicando que não há afastamento da suposição de que os erros tem distribuição log-BS.

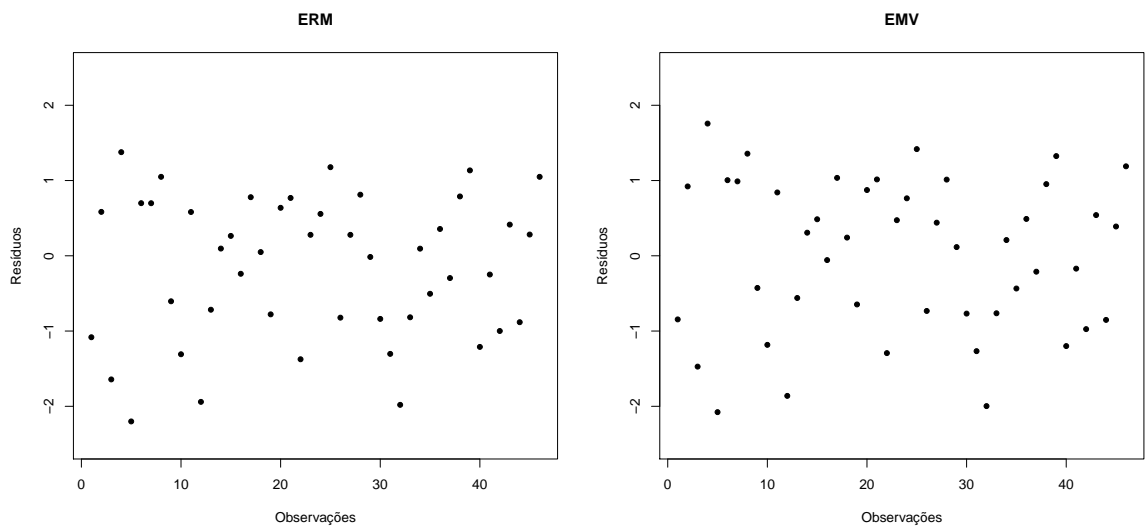
Portanto, embora o uso do método de estimação robusto seja bem justificado para uma análise dos dados na presença de *outliers* esse método demonstra um desempenho tão satisfatório quanto o método por máxima verossimilhança quando não há evidência de observações extremas nos resíduos.

Figura 4.2: Gráficos dos resíduos  $r_i$  baseados no ERM e EMV ajustados aos dados sobre a vida útil de uma peça de metal.



Fonte: Autoria própria.

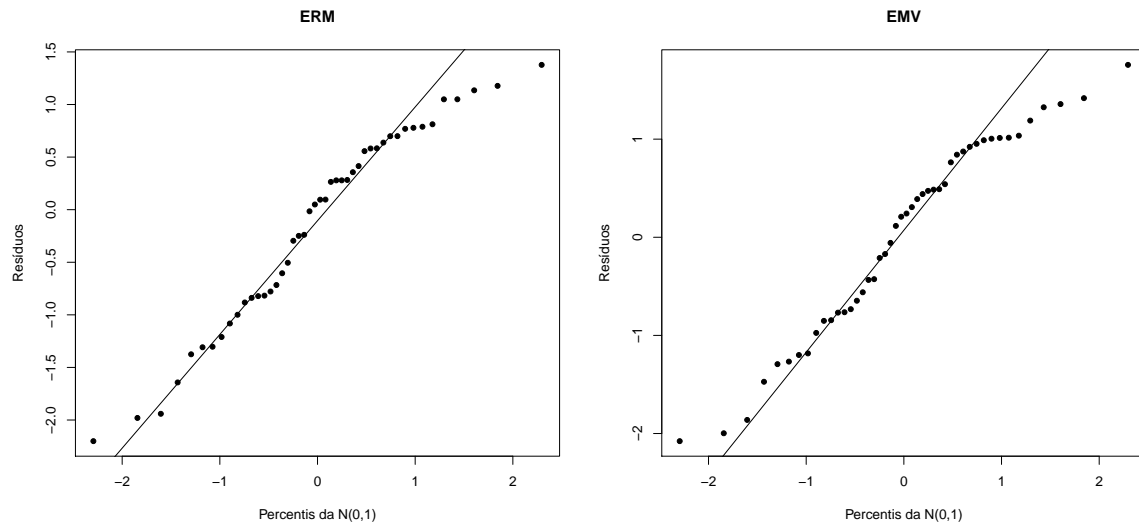
Figura 4.3: Gráficos dos resíduos  $r_{qi}$  baseados no ERM e EMV ajustados aos dados sobre a vida útil de uma peça de metal.



Fonte: Autoria própria.



Figura 4.4: Gráficos normais de probabilidade para os resíduos  $r_{q_i}$  baseados no ERM e EMV ajustados aos dados sobre a vida útil de uma peça de metal.



Fonte: Autoria própria.

## 4.2 Aplicação 2

### 4.2.1 Estudo sobre o tempo de vida de pacientes com câncer de pulmão

Consideremos o conjunto de dados apresentado em [Kalbfleisch e Prentice \(2011\)](#), em que homens com câncer de pulmão avançado foram aleatorizados entre um tratamento de quimioterapia padrão e um teste. A variável resposta é o tempo de sobrevivência ( $T$ , em dias) de 137 pacientes com câncer de pulmão (nove deles censurados) e as variáveis explicativas são:

- ( $x_1$ ) uma medida de aleatorização do estado do paciente (escala *Karnofsky*), sendo
  - 10 – 30 para completamente hospitalizado,
  - 40 – 60 para parcialmente confinado
  - 70 – 90 para capacidade de cuidar-se sozinho;
- ( $x_2$ ) tempo em meses do diagnóstico até a aleatorização;
- ( $x_3$ ) idade em anos, no instante da aleatorização;
- ( $x_4$ ) tratamento anterior, dicotomizada em 10 para sim e 0 para não;

( ) tipo histológico do tumor: *squamous*, *small cell*, *adeno*, *large cell*. Para essa variável consideramos três variáveis auxiliares  $x_5 = 1$ ,  $x_6 = 1$  e  $x_7 = 1$  se o tipo de célula do câncer é *squamous*, *small cell* e *adeno*, respectivamente, e 0 caso contrário.;

( $x_8$ ) tipo de tratamento: 0 para padrão e 1 para teste.

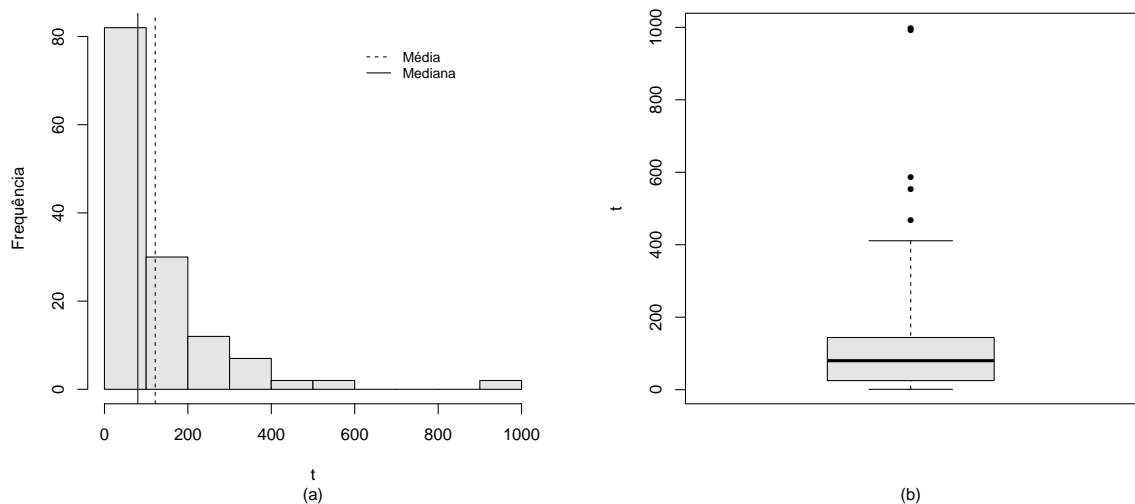
O objetivo do estudo é tentar explicar o tempo de sobrevivência dos pacientes com câncer de pulmão através de uma estrutura de regressão com as covariáveis mencionadas. Inicialmente, fizemos uma análise descritiva da variável resposta. Na Tabela 4.4 encontram-se os resultados de algumas medidas resumo relacionadas aos tempos de sobrevivência. Nota-se que  $T$  assume valores entre 1 e 999. O coeficiente de variação (CV) igual a 129,75% indica que a média de 121,6 não é representativa. O terceiro quartil,  $Q_3$ , revela que 75% dos tempos de sobrevivência são inferiores a 144 dias.

Tabela 4.4: Medidas resumo da variável  $T$ .

Mínimo	1° Quartil	Mediana	Média	3° Quartil	Máximo	Desvio Padrão	CV
1	25	80	121,6	144	999	157,8	129,75%

Na Figura 4.5 (a) tem-se o histograma de frequência da variável resposta. Este

Figura 4.5: Histograma e boxplot dos tempos de sobrevivência dos pacientes com câncer de pulmão.



Fonte: Autoria própria.

gráfico mostra um comportamento assimétrico em que a maior parte das observações se encontram no intervalo de 0 a 100, destacamos que a média dos dados não encontra-se entre a maior parte das observações confirmando uma média pouco representativa ao contrário da mediana. Na Figura 4.5 (b) também observa-se esse comportamento assimétrico, além disso, nota-se a presença de valores extremos, a saber: os dois maiores tempos de sobrevivência são iguais 999 e 991 dias, observações #70 e #75, respectivamente.

### 4.2.2 Ajuste do modelo log-BS

Esse conjunto de dados tem sido analisado através de modelos de regressão usando diferentes distribuições para os erros, a saber: exponencial, log-normal, gama generalizada, log-logística e Weibull (ver, por exemplo, Lee e Wang (2003)). Porém, dado a natureza do comportamento da variável resposta, Barros (2007) baseando-se no argumento de tempo de fadiga propõe para este conjunto de dados o modelo de tempo de vida acelerado BS na forma log-linear dado em (2.12). Mais especificamente,

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_8 x_{i8} + \epsilon_i, \quad i = 1, \dots, 137, \quad (4.2)$$

em que  $\epsilon_i$  são variáveis aleatórias i.i.d. tais que  $\epsilon_i \sim \text{SN}(\alpha, 0, 2)$  e a resposta  $y_i = \log(t_i)$  denota o logaritmo do tempo de sobrevivência ou do tempo de censura. Barros (2007) analisou este conjunto de dados através dos modelo log-BS e expôs o problema da falta de robustez do método de máxima verossimilhança contra observações extremas. O caminho utilizado para contornar esse problema foi considerar um modelo de caudas mais pesadas para ajustar os dados. O modelo utilizado foi o  $t$  de Student log-BS (modelo log-BSG cujo erro tem distribuição BSG gerada pela distribuição  $t$  de Student). Aqui, nosso objetivo não é considerar outra distribuição para os erros e sim considerar um método de estimação robusto para estimar os parâmetros do modelo log-BS na presença de *outliers*. Portanto, voltaremos a analisar esse conjunto dados através do modelo log-BS (4.2) para comparar os métodos de estimação robusto e por máxima verossimilhança. Os valores das estimativas, bem como os erros padrão e os respectivos  $p$ -valores estão dispostos na Tabela 4.5.

Fixando o nível de significância em 5%, temos que as variáveis explicativas  $x_2, x_4,$

Tabela 4.5: Estimativa dos parâmetros do modelo log-BS para os dados de câncer de pulmão utilizando o ERM e o EMV.

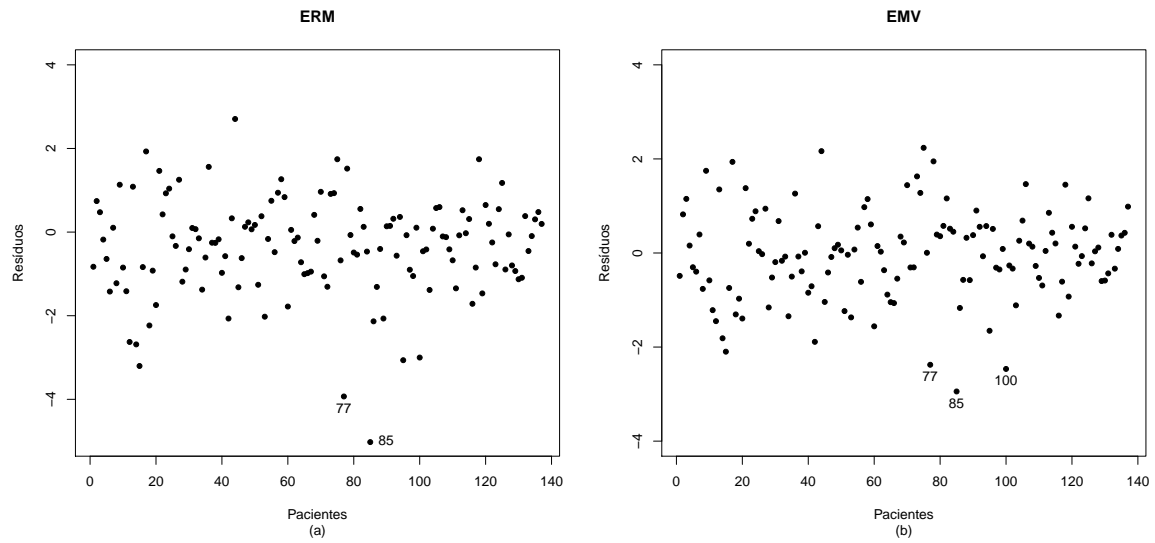
Parâmetros	ERM			EMV		
	Estimativa	Erro Padrão	$p$ -valor	Estimativa	Erro Padrão	$p$ -valor
$\alpha$	1,008	0,068	0,000	1,262	0,079	0,000
$\beta_0$	2,647	0,633	0,000	1,141	0,638	0,074
$\beta_1$	0,035	0,004	0,000	0,041	0,005	0,000
$\beta_2$	0,004	0,007	0,568	-0,003	0,009	0,739
$\beta_3$	0,001	0,009	0,912	0,022	0,008	0,006
$\beta_4$	-0,004	0,017	0,814	-0,002	0,023	0,931
$\beta_5$	0,195	0,227	0,390	-0,280	0,297	0,346
$\beta_6$	-0,649	0,197	0,001	-0,705	0,294	0,016
$\beta_7$	-0,792	0,198	0,000	-0,685	0,358	0,056
$\beta_8$	-0,034	0,147	0,817	-0,382	0,192	0,047

$x_5$  e  $x_7$  são marginalmente não significativas usando o EMV. Ou seja, esta análise inicial indica que o tempo da doença em meses e ter realizado ou não algum tipo de tratamento anteriormente, não possui dependência aparente com os tempos de sobrevivência. Por outro lado, usando o ERM e fixando o nível de significância também em 5%, as variáveis explicativas com efeito não significativo são  $x_2$ ,  $x_3$ ,  $x_4$ ,  $x_5$  e  $x_8$ . Ou seja, apenas o estado do paciente e o tipo do tumor possuem dependência aparente com os tempos de sobrevivência nesta análise inicial.

Uma análise de diagnóstico baseada nos resíduos  $r_i$ , dados por (3.7), de ambos os métodos de estimação é mostrada na Figura 4.6. A Figura 4.6 (a) referente ao ERM nos mostra um comportamento aleatórios destes resíduos, no entanto revela forte destaque para as observações #77 e #85 com resíduo alto. A Figura 4.6(b) referente ao EMV também mostra um comportamento aleatório dos resíduos e destaca três observações como pontos influentes, são elas: #77, #85 e #100. Porém, note que os valores destacados pelo ERM são ampliados em maior escala comparado ao EMV. Outra análise de diagnóstico baseando-se nos resíduos  $r_{q_i}$ , dados por (3.8), foi realizada para os métodos de estimação e estão representadas na Figura 4.7. Nota-se mais uma vez que não há nenhum comportamento sistemático desses resíduos, mas que existem observações atípicas que correspondem as mesmas que foram identificadas na Figura 4.6. Observe que o diagnóstico baseado nos resíduos  $r_{q_i}$  amplia em maior escala os valores destacados quando comparado com o diagnóstico baseado nos resíduos  $r_i$ .

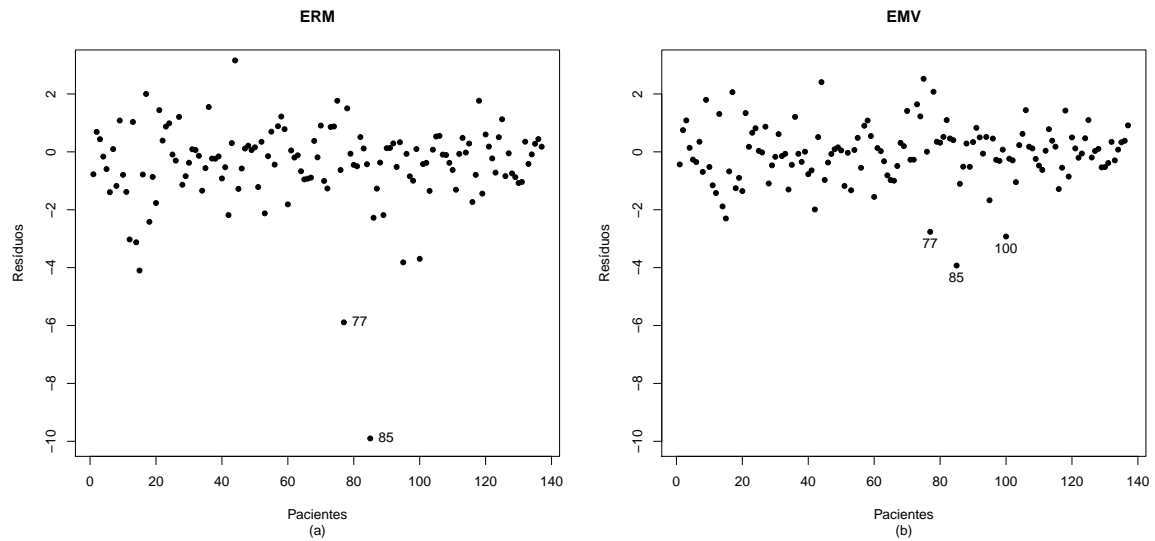
Os gráficos normais de probabilidade gerados para os resíduos quantílicos  $r_{q_i}$

Figura 4.6: Gráficos dos resíduos  $r_i$  baseados no ERM e EMV ajustados aos dados de câncer de pulmão.



Fonte: Autoria própria.

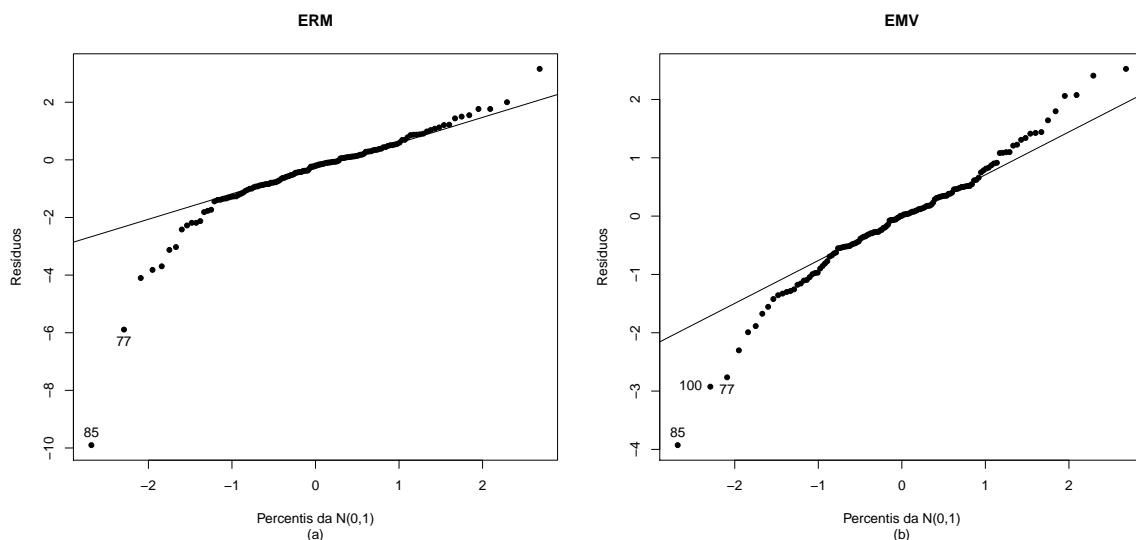
Figura 4.7: Gráficos dos resíduos  $r_{q_i}$  baseados no ERM e EMV ajustados aos dados de câncer de pulmão.



Fonte: Autoria própria.

baseados no ERM e no EMV são apresentados na Figura 4.8. Temos pela Figura 4.8 (a) que o ajuste do modelo pelo ERM evidencia claramente os valores discrepantes que foram revelados nas Figuras 4.6 (a) e 4.7 (a) e ajusta muito bem a maior parte dos dados, como é possível observar pelo comportamento linear da parte superior do

Figura 4.8: Gráficos normais de probabilidade para os resíduos  $r_{q_i}$  baseados no ERM e EMV ajustados aos dados de câncer de pulmão.



Fonte: Autoria própria.

gráfico. Por outro lado, a Figura 4.8 referente ao EMV, mostra uma indicação de que uma distribuição com caudas mais pesadas para os erros seria mais apropriada, já que evidencia várias outras observações discrepantes além das observações destacadas nos gráficos das Figuras 4.6 (b) e 4.7 (b).

Nas Tabelas 4.6 e 4.7 encontram-se, respectivamente, as estimativas robustas e por máxima verossimilhança dos parâmetros do modelo quando retiramos individualmente e conjuntamente cada observação considerada como influente. Também foram calculadas as mudanças relativas, em porcentagem, de cada estimativa, definida por  $MR_j = [(\hat{\theta}_j - \hat{\theta}_{j(i)})/\hat{\theta}_j] \times 100$ , para  $j = 0, 1, \dots, 9$  e  $i = 1, \dots, 137$ , em que  $\hat{\theta}_{j(i)}$  denota o estimador para  $\hat{\theta}_j$  quando a  $i$ -ésima observação é excluída.

A Tabela 4.6 mostra as estimativas robustas dos parâmetros do modelo, as variações percentuais nessas estimativas quando eliminamos, individualmente, as observações #77 e #85 ou conjuntamente (representada pelo conjunto  $I = \{\#77 \text{ e } \#85\}$ ) e os respectivos  $p$ -valores para as novas estimativas. Podemos perceber, como esperado, que as variações percentuais são mínimas, sendo algumas iguais a zero inclusive. As estimativas robustas permanecem próximas e não há mudança inferencial para nenhum dos coeficientes. Na Tabela 4.7 temos as estimativas por máxima verossimilhança dos parâmetros do modelo, as variações percentuais nessas estimativas quando eliminamos,

Tabela 4.6: Estimativas robustas, mudanças relativas (-MR- em %) e os correspondentes níveis descritivos para os casos indicados.

Parâmetros		Observação eliminada			
		Nenhuma	77	85	Conjunto I
$\alpha$	$\hat{\alpha}$	1,008	0,988	0,990	0,973
	MR	-	2	-2	3
	$p$ -valor	0,000	0,000	0,000	0,000
$\beta_0$	$\hat{\beta}_0$	2,647	2,670	2,729	2,750
	MR	-	-1	-3	-4
	$p$ -valor	0,000	0,000	0,000	0,000
$\beta_1$	$\hat{\beta}_1$	0,035	0,034	0,035	0,034
	MR	-	3	0	3
	$p$ -valor	0,000	0,000	0,000	0,000
$\beta_2$	$\hat{\beta}_2$	0,004	0,005	0,004	0,005
	MR	-	-25	0	-25
	$p$ -valor	0,568	0,476	0,568	0,476
$\beta_3$	$\hat{\beta}_3$	0,001	0,001	0,000	0,000
	MR	-	0	100	100
	$p$ -valor	0,912	0,912	1,000	1,000
$\beta_4$	$\hat{\beta}_4$	-0,004	-0,003	-0,005	-0,004
	MR	-	25	-25	0
	$p$ -valor	0,814	0,860	0,768	0,814
$\beta_5$	$\hat{\beta}_5$	0,195	0,215	0,219	0,240
	MR	-	-10	-12	-23
	$p$ -valor	0,390	0,350	0,338	0,286
$\beta_6$	$\hat{\beta}_6$	-0,649	-0,660	-0,648	-0,659
	MR	-	-2	0	-2
	$p$ -valor	0,001	0,000	0,002	0,000
$\beta_7$	$\hat{\beta}_7$	-0,792	-0,790	-0,792	-0,791
	MR	-	0	0	0
	$p$ -valor	0,000	0,000	0,000	0,000
$\beta_8$	$\hat{\beta}_8$	-0,034	-0,032	-0,030	-0,028
	MR	-	6	12	18
	$p$ -valor	0,817	0,834	0,842	0,850

individualmente, as observações #77, #85 e #100 ou conjuntamente (representada pelo conjunto  $J = \{\#77, \#85 \text{ e } \#100\}$ ) e os respectivos  $p$ -valores para as novas estimativas. Podemos observar que, diferentemente do ERM, as variações percentuais são grandes,

Tabela 4.7: Estimativas por máxima verossimilhança, mudanças relativas (-MR- em %) e os correspondentes níveis descritivos para os casos indicados.

Parâmetros		Observação eliminada				
		Nenhuma	77	85	100	Conjunto J
$\alpha$	$\hat{\alpha}$	1,262	1,218	1,154	1,223	1,052
	MR	-	3	9	3	17
	$p$ -valor	0,000	0,000	0,000	0,000	0,000
$\beta_0$	$\hat{\beta}_0$	1,141	1,174	2,324	0,739	2,062
	MR	-	-3	-104	35	-81
	$p$ -valor	0,074	0,056	0,000	0,240	0,000
$\beta_1$	$\hat{\beta}_1$	0,041	0,036	0,037	0,044	0,035
	MR	-	12	10	-7	15
	$p$ -valor	0,000	0,000	0,000	0,000	0,000
$\beta_2$	$\hat{\beta}_2$	-0,003	-0,004	-0,003	-0,004	-0,006
	MR	-	-33	0	-33	-100
	$p$ -valor	0,738	0,618	0,708	0,618	0,392
$\beta_3$	$\hat{\beta}_3$	0,022	0,025	0,005	0,025	0,011
	MR	-	-14	77	-14	50
	$p$ -valor	0,006	0,002	0,532	0,002	0,170
$\beta_4$	$\hat{\beta}_4$	-0,002	0,009	-0,028	-0,005	-0,018
	MR	-	550	-1300	-150	-800
	$p$ -valor	0,930	0,696	0,204	0,820	0,368
$\beta_5$	$\hat{\beta}_5$	-0,280	-0,201	-0,064	-0,285	0,054
	MR	-	28	77	-2	119
	$p$ -valor	0,346	0,488	0,822	0,322	0,834
$\beta_6$	$\hat{\beta}_6$	-0,705	-0,734	-0,717	-0,566	-0,627
	MR	-	-4	-2	20	11
	$p$ -valor	0,016	0,010	0,008	0,050	0,012
$\beta_7$	$\hat{\beta}_7$	-0,685	-0,744	-0,800	-0,710	-0,882
	MR	-	-9	-17	-4	-29
	$p$ -valor	0,056	0,032	0,016	0,044	0,002
$\beta_8$	$\hat{\beta}_8$	-0,382	-0,313	-0,199	-0,242	0,005
	MR	-	18	48	37	101
	$p$ -valor	0,046	0,092	0,272	0,212	0,976

mesmo para variáveis que foram significativas a um nível de 5%. As estimativas por máxima verossimilhança mudam bastante em alguns coeficientes, nota-se também que há mudança inferencial. Em particular,  $\beta_3$  e  $\beta_8$  deixam de ser significativos a um nível



de 5% quando eliminamos as três observações consideradas influentes e o  $\beta_7$  passa a ser significativo.

Embora as estimativas por máxima verossimilhança apresentem falta de robustez na presença de observações extremas, com o exposto acima e por Barros (2007), é possível escolher um modelo reduzido que inclui as variáveis medida de aleatorização do estado do paciente e o tipo histológico do tumor. Especificamente, este modelo é dado por

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \epsilon_i, \quad i = 1, \dots, 137, \quad (4.3)$$

o qual já foi indicado pela análise inicial quando utilizamos o método robusto de estimação.

Agora, os métodos robusto e por máxima verossimilhança serão utilizados para ajustar o modelo reduzido 4.3, no qual inclui apenas as variáveis significativas. As estimativas, bem como os erros padrão e os respectivos  $p$ -valores estão dispostos na Tabela 4.8.

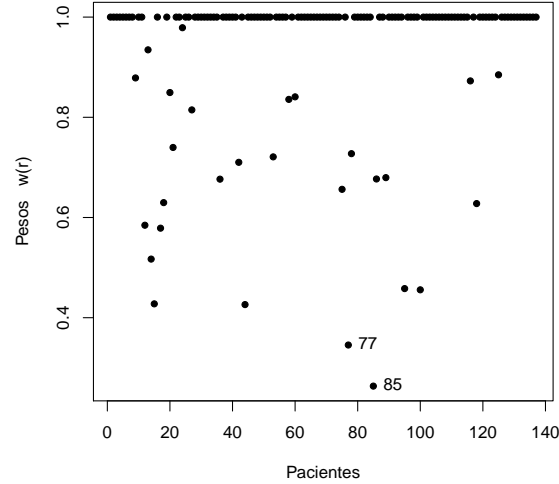
Tabela 4.8: Estimativa dos parâmetros do modelo reduzido log-BS para os dados de câncer de pulmão utilizando o ERM e o EMV.

Parâmetros	ERM			EMV		
	Estimativa	Erro Padrão	$p$ -valor	Estimativa	Erro Padrão	$p$ -valor
$\alpha$	1,021	0,069	0,000	1,322	0,083	0,000
$\beta_0$	2,705	0,282	0,000	2,227	0,415	0,000
$\beta_1$	0,035	0,004	0,000	0,040	0,005	0,000
$\beta_5$	0,202	0,208	0,332	-0,510	0,306	0,096
$\beta_6$	-0,634	0,190	0,000	-0,638	0,311	0,040
$\beta_7$	-0,806	0,204	0,000	-0,744	0,371	0,044

Observe que, embora  $\beta_5$  indique a não significância da variável  $x_5$ , lembremos que trata-se de uma variável auxiliar do tipo histológico do tumor que por sua vez é significativa devido as outras variáveis auxiliares  $x_6$  e  $x_7$  serem marginalmente significativas. A variável auxiliar  $x_5$  apresenta também a maior discrepância entre os dois conjuntos de estimativas, o ERM fornece uma estimativa de 0,202 com um erro padrão de 0,208, enquanto o EMV fornece uma estimativa de -0,510 com um erro padrão um pouco maior de 0,306.

Sabemos que uma certa quantidade de *outliers* causam discrepâncias entre as

Figura 4.9: Gráficos dos pesos  $w(r_i)$  utilizado no ERM para ajustar os dados de câncer de pulmão.

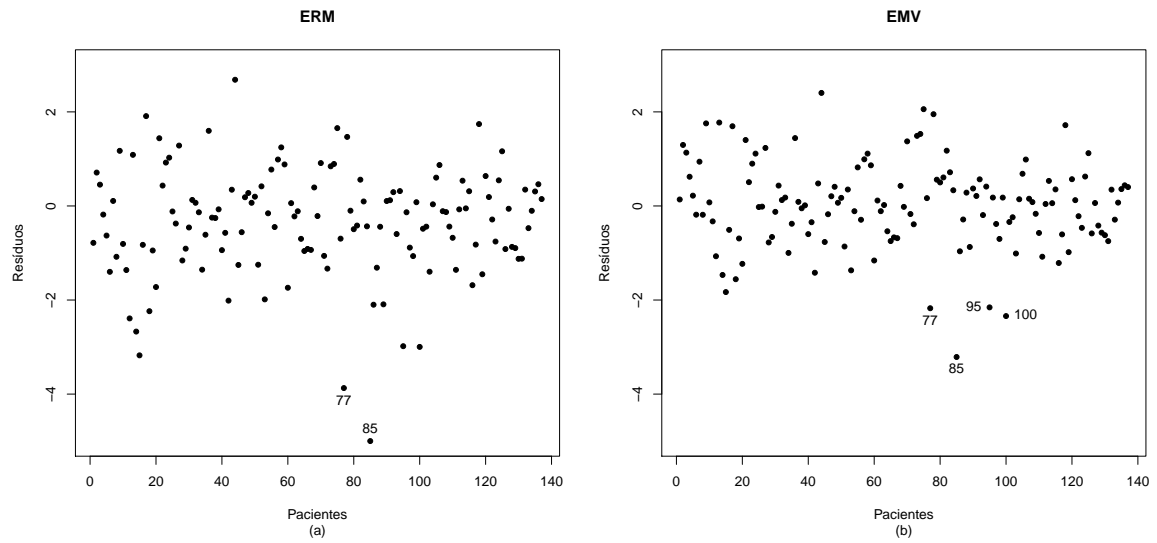


Fonte: Autoria própria.

estimativas robustas e as estimativas por máxima verossimilhança, esses *outliers* podem ser identificados a partir das parcelas residuais, bem como pela função de peso  $w(r_i)$ . A Figura 4.9 mostra a função de peso  $w(r_i)$  obtida a partir do ajuste pelo ERM. A função de peso indica claramente um considerável número de pontos com pesos menores que 1, dois deles recebem um peso inferior a 0,4, que correspondem exatamente aos dois valores extremos que foram detectados anteriormente, #77 e #85.

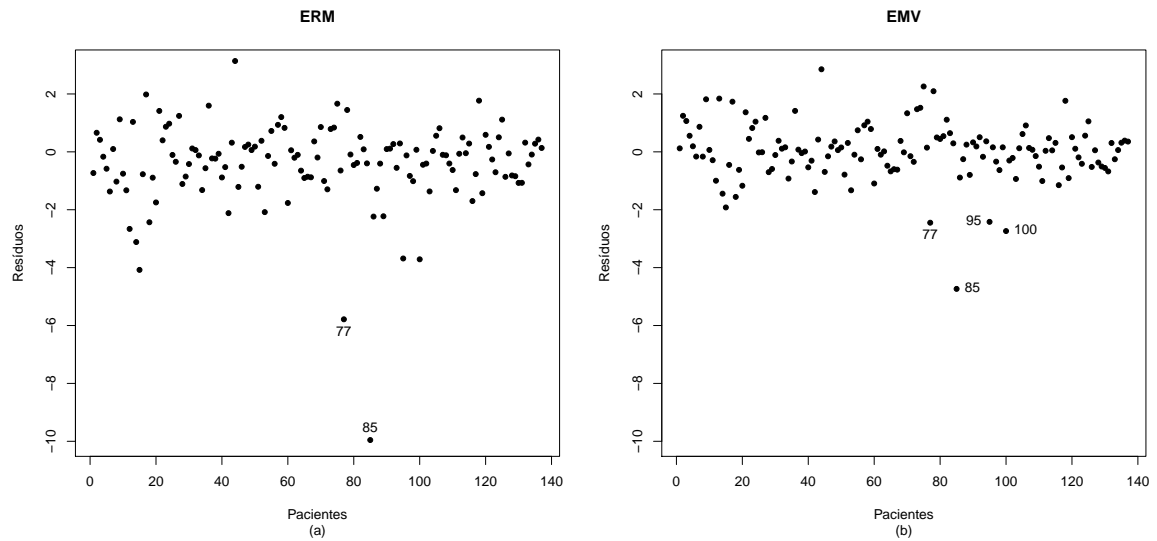
Na Figura 4.10 temos os gráficos dos resíduos  $r_i$  obtidos a partir dos ajustes utilizando o ERM e o EMV. Na Figura 4.10 (a) observa-se um comportamento aleatório, mas também notamos alguns valores discrepantes, sendo esses os mesmos que receberam os menores pesos no processo de estimação. O uso do ERM é, portanto, bem justificado aqui para uma análise válida dos dados na presença dos *outliers*. Na Figura 4.10 (b) temos agora quatro observações indicadas como influentes, são elas: #77, #85, #95 e #100. Novamente observa-se que o destaque para esses resíduos foram em menor escala comparado ao ERM. Esse comportamento se repete quando analisamos os gráficos dos resíduos quantílicos,  $r_{qi}$ , obtidos a partir dos ajustes utilizando o ERM e o EMV e que estão apresentados na Figura 4.11. Os resíduos  $r_{qi}$  mais uma vez destacam-se em maior escala comparados aos resíduos  $r_i$ .

Figura 4.10: Gráfico dos resíduos  $r_i$  baseados no ERM e EMV ajustados ao modelo reduzido dos dados de câncer de pulmão.



Fonte: Autoria própria.

Figura 4.11: Gráfico dos resíduos  $r_{q_i}$  baseados no ERM e EMV ajustados ao modelo reduzido dos dados de câncer de pulmão.



Fonte: Autoria própria.

As observações #77 e #85 destacadas como pontos influentes pelo ERM, assim como na análise do modelo 4.2, mais uma vez não causam mudanças inferenciais nas estimativas quando são retiradas dos dados, como pode ser observado na Tabela 4.9, isso ocorre devido a robustez do método que em seu processo de construção atribui pouco

Tabela 4.9: Estimativas robustas, mudanças relativas (-MR- em %) e os correspondentes níveis descritivos para os casos indicados do modelo reduzido.

Parâmetros		Observação eliminada			
		Nenhuma	77	85	Conjunto I
$\alpha$	$\hat{\alpha}$	1,021	0,999	1,000	1,000
	MR	-	2	2	2
	$p$ -valor	0,000	0,000	0,000	0,000
$\beta_0$	$\hat{\beta}_0$	2,705	2,749	2,716	2,716
	MR	-	-2	0	0
	$p$ -valor	0,000	0,000	0,000	0,000
$\beta_1$	$\hat{\beta}_1$	0,035	0,034	0,035	0,035
	MR	-	3	0	0
	$p$ -valor	0,000	0,000	0,000	0,000
$\beta_5$	$\hat{\beta}_5$	0,202	0,224	0,218	0,218
	MR	-	-11	-8	-8
	$p$ -valor	0,332	0,312	0,308	0,308
$\beta_6$	$\hat{\beta}_6$	-0,634	-0,645	-0,641	-0,641
	MR	-	-2	-1	-1
	$p$ -valor	0,000	0,000	0,000	0,000
$\beta_7$	$\hat{\beta}_7$	-0,806	-0,804	-0,806	-0,806
	MR	-	0	0	0
	$p$ -valor	0,000	0,000	0,000	0,000

peso a estes valores. A Tabela 4.10 mostra os impactos causados nas estimativas por máxima verossimilhança ao eliminar, individualmente ou conjuntamente (representada pelo conjunto  $K = \{\#77, \#85, \#95 \text{ e } \#100\}$ ) as observações destacadas pelo método. Percebemos que as mudanças relativas aqui são maiores quando comparadas ao ERM, além disso, há mudança inferencial no parâmetro  $\beta_6$ , isto é,  $\beta_6$  deixa de ser significativo a um nível de 5% quando eliminamos individualmente as observações #95 e #100.

Os gráficos normais de probabilidade gerados para os resíduos quantílicos  $r_{q_i}$  baseados no ERM e no EMV são apresentados na Figura 4.12. Temos pela Figura 4.12 (b) referente ao EMV, mais uma vez a indicação de que seria mais apropriado a escolha de uma distribuição com caudas mais pesadas para os erros, pois várias observações discrepantes são evidenciadas na parte extrema superior do gráfico, além das observações discrepantes na parte extrema inferior que são as mesmas destacadas no gráfico da Figura 4.10. Já a Figura 4.12 (a) mostra que o ajuste do modelo usando o

Tabela 4.10: Estimativas por máxima verossimilhança, mudanças relativas (-MR- em %) e os correspondentes níveis descritivos para os casos indicados do modelo reduzido.

Parâmetros		Observação eliminada					
		Nenhuma	77	85	95	100	Conjunto K
$\alpha$	$\hat{\alpha}$	1,322	1,289	1,177	1,296	1,288	1,031
	MR	-	2	11	2	3	22
	$p$ -valor	0,000	0,000	0,000	0,000	0,000	0,000
$\beta_0$	$\hat{\beta}_0$	2,227	2,445	2,330	2,303	2,028	2,622
	MR	-	-10	-5	-3	9	-18
	$p$ -valor	0,000	0,000	0,000	0,000	0,000	0,000
$\beta_1$	$\hat{\beta}_1$	0,040	0,037	0,038	0,039	0,043	0,034
	MR	-	8	5	3	-7	15
	$p$ -valor	0,000	0,000	0,000	0,000	0,000	0,000
$\beta_5$	$\hat{\beta}_5$	-0,510	-0,452	-0,164	-0,518	-0,479	0,012
	MR	-	11	68	-2	6	102
	$p$ -valor	0,096	0,146	0,564	0,086	0,116	0,962
$\beta_6$	$\hat{\beta}_6$	-0,638	-0,671	-0,656	-0,565	-0,492	-0,511
	MR	-	-5	-3	11	23	20
	$p$ -valor	0,040	0,034	0,018	0,068	0,118	0,034
$\beta_7$	$\hat{\beta}_7$	-0,744	-0,788	-0,765	-0,760	-0,701	-0,818
	MR	-	-6	-3	-2	6	-10
	$p$ -valor	0,044	0,036	0,020	0,038	0,054	0,004

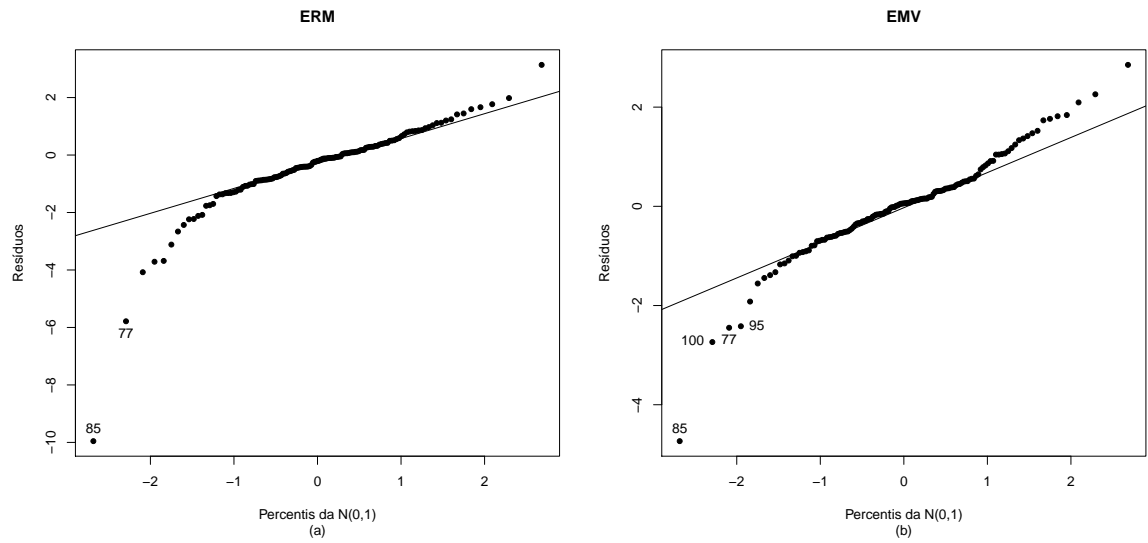
ERM ressalta claramente os valores discrepantes que foram revelados na Figura 4.9 (b) e ajusta muito bem a maior parte dos dados, como evidenciado pelo comportamento linear da parte superior do gráfico.

Portanto, ajustar o modelo log-BS para esse conjunto de dados pelo método de estimação robusto parece ser mais conveniente. Assim, o modelo final selecionado é dado pela equação (4.3) com as seguintes estimativas e os respectivos erros padrão aproximados entre parênteses

$$\hat{\alpha} = 1,021 (0,069), \hat{\beta}_0 = 2,705 (0,282), \hat{\beta}_1 = 0,035 (0,004), \\ \hat{\beta}_5 = 0,202 (0,208), \hat{\beta}_6 = -0,634 (0,190), \hat{\beta}_7 = -0,806 (0,204).$$

A interpretação do modelo é dado da seguinte maneira: É esperado um crescimento do tempo de sobrevivência para as maiores medidas de aleatorização do estado

Figura 4.12: Gráficos normais de probabilidade para os resíduos  $r_{q_i}$  baseados no ERM e EMV ajustados ao modelo reduzido dos dados de câncer de pulmão.



Fonte: Autoria própria.

do paciente (escala *Karnofsky*), mas se fixamos a medida de aleatorização, o tempo de sobrevivência esperado deverá decrescer em aproximadamente  $1,89 [e^{0,634}]$  e  $2,24 [e^{0,806}]$ , respectivamente, para o tipo de célula do câncer *small cell* e *adeno* com respeito ao *squamous*.

# Capítulo 5

## Conclusões

Neste trabalho observamos a influência que os *outliers* ocasionam nas estimativas de máxima verossimilhança para o modelo de tempo de vida acelerado Birnbaum-Saunders. Propomos um estimador robusto, que é uma generalização do estimador de máxima verossimilhança, e avaliamos esses estimadores através de estudos de simulação de Monte Carlo, através dos quais concluímos que o método robusto proposto é equivalente ao método de máxima verossimilhança quando não há valores discrepantes nos dados. No entanto, quando os dados contêm *outliers*, o método robusto é muito melhor que o método de máxima verossimilhança em termos de vieses, erros quadrados médios e probabilidades de cobertura de intervalos de confiança. Um exemplo simulado foi apresentado, o qual ilustra bem a superioridade do método robusto em relação ao de máxima verossimilhança. Por fim, duas aplicações usando dados reais indicaram um desempenho satisfatório do método proposto ao lidar com dados sem *outliers* bem como a sua utilidade na presença dos *outliers*.





# Apêndice A

## Expressões do Capítulo 3

### A.1 Funções para $\epsilon_i \sim \text{SN}(\alpha, 0, 2)$

Se o termo  $\epsilon_i$  tem uma distribuição Senh-normal, ou seja,  $\epsilon_i \sim \text{SN}(\alpha, 0, 2)$ , então a função densidade para  $\epsilon_i$  será dada por

$$\begin{aligned} f_{\epsilon_i}(\epsilon_i) &= \frac{1}{\alpha} \phi\left(\frac{2}{\alpha} \sinh\left(\frac{\epsilon_i}{2}\right)\right) \cosh\left(\frac{\epsilon_i}{2}\right) \\ &= \frac{2}{2\alpha} \cosh\left(\frac{\epsilon_i}{2}\right) \phi\left(\frac{2}{\alpha} \sinh\left(\frac{\epsilon_i}{2}\right)\right) \\ &= \frac{1}{2} \frac{2}{\alpha} \cosh\left(\frac{\epsilon_i}{2}\right) \phi\left(\frac{2}{\alpha} \sinh\left(\frac{\epsilon_i}{2}\right)\right) \\ &= \frac{1}{2} \xi_{i1} \phi(\xi_{i2}), \end{aligned}$$

em que  $\xi_{i1} = \frac{2}{\alpha} \cosh\left(\frac{\epsilon_i}{2}\right)$  e  $\xi_{i2} = \frac{2}{\alpha} \sinh\left(\frac{\epsilon_i}{2}\right)$ .

A função de distribuição acumulada para  $\epsilon_i$  é dada por

$$F_{\epsilon_i}(\epsilon_i) = \Phi\left(\frac{2}{\alpha} \sinh\left(\frac{\epsilon_i}{2}\right)\right) = \Phi(\xi_{i2}).$$

A função de sobrevivência  $S_{\epsilon_i}(\epsilon_i)$  e a função de risco  $h_{\epsilon_i}(\epsilon_i)$  de  $\epsilon_i$  são obtidas, respectivamente, por

$$S_{\epsilon_i}(\epsilon_i) = 1 - F_{\epsilon_i}(\epsilon_i) = 1 - \Phi(\xi_{i2}),$$

$$h_{\epsilon_i}(\epsilon_i) = \frac{f_{\epsilon_i}(\epsilon_i)}{S_{\epsilon_i}(\epsilon_i)} = \frac{\frac{1}{2} \xi_{i1} \phi(\xi_{i2})}{1 - \Phi(\xi_{i2})}.$$

## A.2 Funções escore

Sabemos que  $\xi_{i1} = \frac{2}{\alpha} \cosh\left(\frac{\epsilon_i}{2}\right)$  e  $\xi_{i2} = \frac{2}{\alpha} \sinh\left(\frac{\epsilon_i}{2}\right)$ , daí temos que

$$\frac{\partial \xi_{i1}}{\partial \alpha} = -\frac{2}{\alpha^2} \cosh\left(\frac{\epsilon_i}{2}\right) = -\frac{\xi_{i1}}{\alpha}, \quad (\text{A.1})$$

$$\frac{\partial \xi_{i2}}{\partial \alpha} = -\frac{2}{\alpha^2} \sinh\left(\frac{\epsilon_i}{2}\right) = -\frac{\xi_{i2}}{\alpha}, \quad (\text{A.2})$$

$$\frac{\partial \xi_{i1}}{\partial \epsilon_i} = \frac{2}{\alpha} \sinh\left(\frac{\epsilon_i}{2}\right) \left(\frac{1}{2}\right) = \frac{\xi_{i2}}{2}, \quad (\text{A.3})$$

$$\frac{\partial \xi_{i2}}{\partial \epsilon_i} = \frac{2}{\alpha} \cosh\left(\frac{\epsilon_i}{2}\right) \left(\frac{1}{2}\right) = \frac{\xi_{i1}}{2}, \quad (\text{A.4})$$

$$\frac{\phi'(\xi_{i2})}{\phi(\xi_{i2})} = \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{\xi_{i2}^2}{2}} (-\xi_{i2})}{\frac{1}{\sqrt{2\pi}} e^{-\frac{\xi_{i2}^2}{2}}} = -\xi_{i2}, \quad (\text{A.5})$$

$$\frac{\xi_{i2}}{\xi_{i1}} = \frac{\frac{2}{\alpha} \sinh\left(\frac{\epsilon_i}{2}\right)}{\frac{2}{\alpha} \cosh\left(\frac{\epsilon_i}{2}\right)} = \tanh\left(\frac{\epsilon_i}{2}\right) \quad (\text{A.6})$$

e

$$\begin{aligned} \xi_{i1} \xi_{i2} &= \frac{2}{\alpha} \cosh\left(\frac{\epsilon_i}{2}\right) \frac{2}{\alpha} \sinh\left(\frac{\epsilon_i}{2}\right) \\ &= \frac{4}{\alpha^2} \cosh\left(\frac{\epsilon_i}{2}\right) \sinh\left(\frac{\epsilon_i}{2}\right) \\ &= \frac{2}{\alpha^2} \sinh(\epsilon_i), \end{aligned} \quad (\text{A.7})$$

em que a última igualdade decorre do fato de que  $\cosh(x)\sinh(x) = \frac{1}{2}\sinh(2x)$ .

A função escore para  $\alpha$  é dada por

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \alpha} = \sum_{i=1}^n \left[ \delta_i \frac{\partial \log\{h_{\epsilon_i}(\epsilon_i)\}}{\partial \alpha} + \frac{\partial \log\{S_{\epsilon_i}(\epsilon_i)\}}{\partial \alpha} \right]. \quad (\text{A.8})$$

Temos que

$$\log\{S_{\epsilon_i}(\epsilon_i)\} = \log(1 - \Phi(\xi_{i2}))$$

e

$$\begin{aligned} \log\{h_{\epsilon_i}(\epsilon_i)\} &= \log\left(\frac{1}{2} \xi_{i1} \phi(\xi_{i2})\right) - \log(1 - \Phi(\xi_{i2})) \\ &= \log(0,5) + \log(\xi_{i1}) + \log(\phi(\xi_{i2})) - \log\{S_{\epsilon_i}(\epsilon_i)\}, \end{aligned}$$

assim, usando (A.2) temos

$$\begin{aligned} \frac{\partial \log\{S_{\epsilon_i}(\epsilon_i)\}}{\partial \alpha} &= -\frac{\phi(\xi_{i2})}{1 - \Phi(\xi_{i2})} \left( \frac{\partial \xi_{i2}}{\partial \alpha} \right) \\ &= -\frac{\phi(\xi_{i2})}{1 - \Phi(\xi_{i2})} \left( -\frac{\xi_{i2}}{\alpha} \right) \\ &= \frac{\xi_{i2}}{\alpha} K(\xi_{i2}), \end{aligned} \quad (\text{A.9})$$

em que  $K(\xi_{i2}) = \frac{\phi(\xi_{i2})}{1 - \Phi(\xi_{i2})}$ . E usando (A.1), (A.2) e (A.5) temos

$$\begin{aligned} \frac{\partial \log\{h_{\epsilon_i}(\epsilon_i)\}}{\partial \alpha} &= \frac{1}{\xi_{i1}} \left( \frac{\partial \xi_{i1}}{\partial \alpha} \right) + \frac{\phi'(\xi_{i2})}{\phi(\xi_{i2})} \left( \frac{\partial \xi_{i2}}{\partial \alpha} \right) - \frac{\partial \log\{S_{\epsilon_i}(\epsilon_i)\}}{\partial \alpha} \\ &= \frac{1}{\xi_{i1}} \left( -\frac{\xi_{i1}}{\alpha} \right) - \xi_{i2} \left( -\frac{\xi_{i2}}{\alpha} \right) - \frac{\xi_{i2}}{\alpha} K(\xi_{i2}) \\ &= -\frac{1}{\alpha} + \frac{\xi_{i2}^2}{\alpha} - \frac{\xi_{i2}}{\alpha} K(\xi_{i2}). \end{aligned} \quad (\text{A.10})$$

Portanto, substituindo (A.9) e (A.10) em (A.8) obtemos

$$\begin{aligned} \frac{\partial l(\boldsymbol{\theta})}{\partial \alpha} &= \sum_{i=1}^n \left[ \delta_i \left( -\frac{1}{\alpha} + \frac{\xi_{i2}^2}{\alpha} - \frac{\xi_{i2}}{\alpha} K(\xi_{i2}) \right) + \frac{\xi_{i2}}{\alpha} K(\xi_{i2}) \right] \\ &= \sum_{i=1}^n \left[ \delta_i \left( -\frac{1}{\alpha} + \frac{\xi_{i2}^2}{\alpha} \right) - \delta_i \frac{\xi_{i2}}{\alpha} K(\xi_{i2}) + \frac{\xi_{i2}}{\alpha} K(\xi_{i2}) \right] \end{aligned}$$

$$= \sum_{i=1}^n \left[ \delta_i \left( \frac{\xi_{i2}^2}{\alpha} - \frac{1}{\alpha} \right) + (1 - \delta_i) \frac{\xi_{i2}}{\alpha} K(\xi_{i2}) \right].$$

A função escore para  $\boldsymbol{\beta}$  é dada por

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \left[ \delta_i \frac{\partial \log\{h_{\epsilon_i}(\epsilon_i)\}}{\partial \epsilon_i} + \frac{\partial \log\{S_{\epsilon_i}(\epsilon_i)\}}{\partial \epsilon_i} \right] (-\mathbf{x}_i), \quad (\text{A.11})$$

usando (A.4) temos

$$\begin{aligned} \frac{\partial \log\{S_{\epsilon_i}(\epsilon_i)\}}{\partial \epsilon_i} &= -\frac{\phi(\xi_{i2})}{1 - \Phi(\xi_{i2})} \left( \frac{\partial \xi_{i2}}{\partial \epsilon_i} \right) \\ &= \frac{\phi(\xi_{i2})}{1 - \Phi(\xi_{i2})} \left( \frac{\xi_{i1}}{2} \right) \\ &= -\frac{\xi_{i1}}{2} K(\xi_{i2}), \end{aligned} \quad (\text{A.12})$$

e usando (A.3), (A.4), (A.5), (A.6) e (A.7) temos

$$\begin{aligned} \frac{\partial \log\{h_{\epsilon_i}(\epsilon_i)\}}{\partial \epsilon_i} &= \frac{1}{\xi_{i1}} \left( \frac{\partial \xi_{i1}}{\partial \epsilon_i} \right) + \frac{\phi'(\xi_{i2})}{\phi(\xi_{i2})} \left( \frac{\partial \xi_{i2}}{\partial \epsilon_i} \right) - \frac{\partial \log\{S_{\epsilon_i}(\epsilon_i)\}}{\partial \epsilon_i} \\ &= \frac{1}{\xi_{i1}} \left( \frac{\xi_{i2}}{2} \right) + \frac{\phi'(\xi_{i2})}{\phi(\xi_{i2})} \left( \frac{\xi_{i1}}{2} \right) + \frac{\xi_{i1}}{2} K(\xi_{i2}) \\ &= \frac{1}{2} \left( \frac{\xi_{i2}}{\xi_{i1}} \right) - \frac{1}{2} \xi_{i1} \xi_{i2} + \frac{\xi_{i1}}{2} K(\xi_{i2}) \\ &= \frac{1}{2} \tanh\left(\frac{\epsilon_i}{2}\right) - \frac{1}{\alpha^2} \sinh(\epsilon_i) + \frac{\xi_{i1}}{2} K(\xi_{i2}). \end{aligned} \quad (\text{A.13})$$

Assim, substituindo (A.12) e (A.13) em (A.11) temos

$$\begin{aligned} \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^n \left[ \delta_i \left( \frac{1}{2} \tanh\left(\frac{\epsilon_i}{2}\right) - \frac{1}{\alpha^2} \sinh(\epsilon_i) + \frac{\xi_{i1}}{2} K(\xi_{i2}) \right) - \frac{\xi_{i1}}{2} K(\xi_{i2}) \right] (-\mathbf{x}_i) \\ &= \sum_{i=1}^n \left[ \delta_i \left( \frac{1}{2} \tanh\left(\frac{\epsilon_i}{2}\right) - \frac{1}{\alpha^2} \sinh(\epsilon_i) \right) + \delta_i \frac{\xi_{i1}}{2} K(\xi_{i2}) - \frac{\xi_{i1}}{2} K(\xi_{i2}) \right] (-\mathbf{x}_i) \\ &= \sum_{i=1}^n \left[ \delta_i \left( \frac{1}{\alpha^2} \sinh(\epsilon_i) - \frac{1}{2} \tanh\left(\frac{\epsilon_i}{2}\right) \right) + (1 - \delta_i) \frac{\xi_{i1}}{2} K(\xi_{i2}) \right] \mathbf{x}_i. \end{aligned}$$

### A.3 Funções $\Upsilon'(\boldsymbol{\theta})$

Seja  $\xi_{i1z} = \frac{2}{\alpha} \cosh\left(\frac{z_i}{2}\right)$  e  $\xi_{i2z} = \frac{2}{\alpha} \sinh\left(\frac{z_i}{2}\right)$ , em que  $z_i = E(\epsilon_i) + \sqrt{\text{Var}(\epsilon_i)}\psi_c(r_i)$  com  $\psi_c(r_i) = \max(-c, \min(r_i, c))$  e  $r_i = (\epsilon_i - E(\epsilon_i))/\sqrt{\text{Var}(\epsilon_i)}$  e que  $\epsilon_i = \log(t_i) - \mathbf{x}_i^\top \boldsymbol{\beta}$ , daí temos

$$\frac{\partial \xi_{i1z}}{\partial \alpha} = -\frac{2}{\alpha^2} \cosh\left(\frac{z_i}{2}\right) = -\frac{\xi_{i1z}}{\alpha}, \quad (\text{A.14})$$

$$\frac{\partial \xi_{i2z}}{\partial \alpha} = -\frac{2}{\alpha^2} \sinh\left(\frac{z_i}{2}\right) = -\frac{\xi_{i2z}}{\alpha}, \quad (\text{A.15})$$

$$\begin{aligned} \frac{\partial \xi_{i1z}}{\partial \boldsymbol{\beta}} &= \frac{2}{\alpha} \sinh\left(\frac{z_i}{2}\right) \left( \frac{1}{2} \sqrt{\text{Var}(\epsilon_i)} \psi'_c(r_i) \frac{1}{\sqrt{\text{Var}(\epsilon_i)}} (-\mathbf{x}_i) \right) \\ &= -\frac{\xi_{i2z}}{2} \psi'_c(r_i) \mathbf{x}_i, \end{aligned} \quad (\text{A.16})$$

$$\begin{aligned} \frac{\partial \xi_{i2z}}{\partial \boldsymbol{\beta}} &= \frac{2}{\alpha} \cosh\left(\frac{z_i}{2}\right) \left( \frac{1}{2} \sqrt{\text{Var}(\epsilon_i)} \psi'_c(r_i) \frac{1}{\sqrt{\text{Var}(\epsilon_i)}} (-\mathbf{x}_i) \right) \\ &= -\frac{\xi_{i1z}}{2} \psi'_c(r_i) \mathbf{x}_i, \end{aligned} \quad (\text{A.17})$$

$$\begin{aligned} \frac{\partial K(\xi_{i2z})}{\partial \xi_{i2z}} &= \frac{\partial}{\partial \xi_{i2z}} \left[ \frac{\phi(\xi_{i2z})}{1 - \Phi(\xi_{i2z})} \right] \\ &= \frac{\phi'(\xi_{i2z})(1 - \Phi(\xi_{i2z})) - \phi(\xi_{i2z})(-\phi(\xi_{i2z}))}{[1 - \Phi(\xi_{i2z})]^2} \\ &= \frac{\phi'(\xi_{i2z})(1 - \Phi(\xi_{i2z})) + \phi^2(\xi_{i2z})}{[1 - \Phi(\xi_{i2z})]^2} \\ &= \frac{\phi'(\xi_{i2z})}{1 - \Phi(\xi_{i2z})} + \left( \frac{\phi(\xi_{i2z})}{1 - \Phi(\xi_{i2z})} \right)^2 = K'(\xi_{i2z}), \end{aligned} \quad (\text{A.18})$$

$$\begin{aligned} \frac{\partial}{\partial \alpha} \left[ \frac{\xi_{i2z}^2}{\alpha} \right] &= \frac{2 \xi_{i2z} \left( \frac{\partial \xi_{i2z}}{\partial \alpha} \right) \alpha - \xi_{i2z}^2}{\alpha^2} \\ &= \frac{2 \xi_{i2z} \left( -\frac{\xi_{i2z}}{\alpha} \right) \alpha - \xi_{i2z}^2}{\alpha^2} \end{aligned}$$

$$\begin{aligned}
&= \frac{-2\xi_{i2z}^2 - \xi_{i2z}^2}{\alpha^2} \\
&= -\frac{3\xi_{i2z}^2}{\alpha^2},
\end{aligned} \tag{A.19}$$

$$\begin{aligned}
\frac{\partial}{\partial\alpha} \left[ \frac{\xi_{i2z}}{\alpha} \right] &= \frac{\left( \frac{\partial\xi_{i2z}}{\partial\alpha} \right) \alpha - \xi_{i2z}}{\alpha^2} \\
&= \frac{\left( -\frac{\xi_{i2z}}{\alpha} \right) \alpha - \xi_{i2z}}{\alpha^2} \\
&= \frac{-\xi_{i2z} - \xi_{i2z}}{\alpha^2} \\
&= -\frac{2\xi_{i2z}}{\alpha^2},
\end{aligned} \tag{A.20}$$

$$\frac{\partial}{\partial\alpha} \left[ \frac{1}{\alpha^2} \sinh(z_i) \right] = -\frac{2}{\alpha^3} \sinh(z_i), \tag{A.21}$$

$$\frac{\partial}{\partial\alpha} \left[ \frac{\xi_{i1z}}{2} \right] = \frac{1}{2} \left( \frac{\partial\xi_{i1z}}{\partial\alpha} \right) = \frac{1}{2} \left( -\frac{\xi_{i1z}}{\alpha} \right) = -\frac{\xi_{i1z}}{2\alpha}, \tag{A.22}$$

$$\begin{aligned}
\xi_{i2z}\xi_{i1z} &= \frac{2}{\alpha} \sinh\left(\frac{z_i}{2}\right) \frac{2}{\alpha} \cosh\left(\frac{z_i}{2}\right) \\
&= \frac{4}{\alpha^2} \sinh\left(\frac{z_i}{2}\right) \cosh\left(\frac{z_i}{2}\right) \\
&= \frac{2}{\alpha^2} \sinh(z_i),
\end{aligned} \tag{A.23}$$

$$\begin{aligned}
\frac{\partial}{\partial\beta} \left[ \frac{\xi_{i2z}^2}{\alpha} \right] &= \frac{1}{\alpha} \left( 2\xi_{i2z} \left( \frac{\partial\xi_{i2z}}{\partial\beta} \right) \right) \\
&= \frac{2\xi_{i2z}}{\alpha} \left( -\frac{\xi_{i1z}}{2} \psi'_c(r_i) \mathbf{x}_i \right) \\
&= -\frac{\xi_{i2z}\xi_{i1z}}{\alpha} \psi'_c(r_i) \mathbf{x}_i \\
&= -\frac{2}{\alpha^3} \sinh(z_i) \psi'_c(r_i) \mathbf{x}_i,
\end{aligned} \tag{A.24}$$

$$\frac{\partial}{\partial\beta} \left[ \frac{\xi_{i2z}}{\alpha} \right] = \frac{1}{\alpha} \left( \frac{\partial\xi_{i2z}}{\partial\beta} \right) = \frac{1}{\alpha} \left( -\frac{\xi_{i1z}}{2} \psi'_c(r_i) \mathbf{x}_i \right) = -\frac{\xi_{i1z}}{2\alpha} \psi'_c(r_i) \mathbf{x}_i, \tag{A.25}$$

$$\begin{aligned}
\frac{\partial}{\partial \beta} \left[ -\frac{1}{2} \tanh \left( \frac{z_i}{2} \right) \right] &= -\frac{1}{2} \operatorname{sech}^2 \left( \frac{z_i}{2} \right) \left( \frac{1}{2} \sqrt{\operatorname{Var}(\epsilon_i)} \psi'_c(r_i) \frac{1}{\sqrt{\operatorname{Var}(\epsilon_i)}} (-\mathbf{x}_i) \right) \\
&= \frac{1}{4} \operatorname{sech}^2 \left( \frac{z_i}{2} \right) \psi'_c(r_i) \mathbf{x}_i,
\end{aligned} \tag{A.26}$$

$$\begin{aligned}
\frac{\partial}{\partial \beta} \left[ \frac{1}{\alpha^2} \sinh(z_i) \right] &= \frac{1}{\alpha^2} \cosh(z_i) \left( \sqrt{\operatorname{Var}(\epsilon_i)} \psi'_c(r_i) \frac{1}{\sqrt{\operatorname{Var}(\epsilon_i)}} (-\mathbf{x}_i) \right) \\
&= -\frac{1}{\alpha^2} \cosh(z_i) \psi'_c(r_i) \mathbf{x}_i
\end{aligned} \tag{A.27}$$

e

$$\frac{\partial}{\partial \beta} \left[ \frac{\xi_{i1z}}{2} \right] = \frac{1}{2} \left( \frac{\partial \xi_{i1z}}{\partial \beta} \right) = \frac{1}{2} \left( -\frac{\xi_{i2z}}{2} \psi'_c(r_i) \mathbf{x}_i \right) = -\frac{\xi_{i2z}}{4} \psi'_c(r_i) \mathbf{x}_i. \tag{A.28}$$

A função  $\Upsilon(\boldsymbol{\theta}) = (\Upsilon_\alpha, \Upsilon'_\beta)'$  é formada pelas funções escores robustas e possui os componentes dados por

$$\begin{aligned}
\Upsilon_\alpha &= \sum_{i=1}^n \left[ \left\{ \delta_i \left( \frac{\xi_{i2z}^2}{\alpha} - \frac{1}{\alpha} \right) + (1 - \delta_i) \frac{\xi_{i2z}}{\alpha} K(\xi_{i2z}) \right\} - a_\alpha \right], \\
\Upsilon_\beta &= \sum_{i=1}^n \left[ \left\{ \delta_i \left( \frac{1}{\alpha^2} \sinh(z_i) - \frac{1}{2} \tanh \left( \frac{z_i}{2} \right) \right) + (1 - \delta_i) \frac{\xi_{i1z}}{2} K(\xi_{i2z}) \right\} w(\mathbf{x}_i) \mathbf{x}_i - \mathbf{a}_\beta \right].
\end{aligned}$$

A matriz  $\Upsilon'(\boldsymbol{\theta})$  é obtida derivando a função  $\Upsilon(\boldsymbol{\theta})$  em relação a  $\boldsymbol{\theta}$ . Usando (A.15), (A.18), (A.19) e (A.20) a primeira derivada de  $\Upsilon_\alpha$  em relação a  $\alpha$  é obtida por

$$\begin{aligned}
\Upsilon'_{\alpha\alpha} &= \sum_{i=1}^n \left[ \delta_i \left( -\frac{3\xi_{i2z}^2}{\alpha^2} + \frac{1}{\alpha^2} \right) + (1 - \delta_i) \left( -\frac{2\xi_{i2z}}{\alpha^2} K(\xi_{i2z}) + \frac{\xi_{i2z}}{\alpha} K'(\xi_{i2z}) \left( -\frac{\xi_{i2z}}{\alpha} \right) \right) \right] \\
&= \sum_{i=1}^n \left[ \delta_i \left( \frac{1 - 3\xi_{i2z}^2}{\alpha^2} \right) + (1 - \delta_i) \left( -\frac{2\xi_{i2z}}{\alpha^2} K(\xi_{i2z}) - \frac{\xi_{i2z}^2}{\alpha^2} K'(\xi_{i2z}) \right) \right] \\
&= \sum_{i=1}^n \left[ \delta_i \left( \frac{1 - 3\xi_{i2z}^2}{\alpha^2} \right) - (1 - \delta_i) \left( \frac{2\xi_{i2z}}{\alpha^2} K(\xi_{i2z}) + \frac{\xi_{i2z}^2}{\alpha^2} K'(\xi_{i2z}) \right) \right].
\end{aligned}$$

Usando (A.14), (A.18), (A.21), (A.22) e (A.23) a primeira derivada de  $\Upsilon_\beta$  em relação a  $\alpha$  é obtida por

$$\Upsilon'_{\beta\alpha} = \sum_{i=1}^n \left[ \delta_i \left( -\frac{2}{\alpha^3} \sinh(z_i) \right) + (1 - \delta_i) \left( -\frac{\xi_{i1z}}{2\alpha} K(\xi_{i2z}) + \frac{\xi_{i1z}}{2} K'(\xi_{i2z}) \left( -\frac{\xi_{i2z}}{\alpha} \right) \right) \right]$$

$$\begin{aligned}
& \times w(\mathbf{x}_i) \mathbf{x}_i \\
& = \sum_{i=1}^n \left[ \delta_i \left( -\frac{2}{\alpha^3} \sinh(z_i) \right) + (1 - \delta_i) \left( -\frac{\xi_{i1z}}{2\alpha} K(\xi_{i2z}) - \frac{1}{2\alpha} \xi_{i1z} \xi_{i2z} K'(\xi_{i2z}) \right) \right] \\
& \times w(\mathbf{x}_i) \mathbf{x}_i \\
& = \sum_{i=1}^n \left[ \delta_i \left( -\frac{2}{\alpha^3} \sinh(z_i) \right) - (1 - \delta_i) \left( \frac{\xi_{i1z}}{2\alpha} K(\xi_{i2z}) + \frac{1}{\alpha^3} \sinh(z_i) K'(\xi_{i2z}) \right) \right] \\
& \times w(\mathbf{x}_i) \mathbf{x}_i.
\end{aligned}$$

Usando (A.17), (A.18), (A.23), (A.24) e (A.25) a primeira derivada de  $\Upsilon_\alpha$  em relação a  $\beta$  é obtida por

$$\begin{aligned}
\Upsilon'_{\alpha\beta} &= \sum_{i=1}^n \left[ \delta_i \left( -\frac{2}{\alpha^3} \sinh(z_i) \psi'_c(r_i) \mathbf{x}_i \right) \right. \\
& \quad \left. + (1 - \delta_i) \left( \left( -\frac{\xi_{i1z}}{2\alpha} \psi'_c(r_i) \mathbf{x}_i \right) K(\xi_{i2z}) + \frac{\xi_{i2z}}{\alpha} K'(\xi_{i2z}) \left( -\frac{\xi_{i1z}}{2} \psi'_c(r_i) \mathbf{x}_i \right) \right) \right] \\
&= \sum_{i=1}^n \left[ \delta_i \left( -\frac{2}{\alpha^3} \sinh(z_i) \psi'_c(r_i) \mathbf{x}_i \right) \right. \\
& \quad \left. + (1 - \delta_i) \left( -\frac{\xi_{i1z}}{2\alpha} K(\xi_{i2z}) \psi'_c(r_i) \mathbf{x}_i - \frac{1}{2\alpha} \xi_{i2z} \xi_{i1z} K'(\xi_{i2z}) \psi'_c(r_i) \mathbf{x}_i \right) \right] \\
&= \sum_{i=1}^n \left[ \delta_i \left( -\frac{2}{\alpha^3} \sinh(z_i) \psi'_c(r_i) \mathbf{x}_i \right) \right. \\
& \quad \left. - (1 - \delta_i) \left( \frac{\xi_{i1z}}{2\alpha} K(\xi_{i2z}) \psi'_c(r_i) \mathbf{x}_i + \frac{1}{2\alpha} \left( \frac{2}{\alpha^2} \sinh(z_i) \right) K'(\xi_{i2z}) \psi'_c(r_i) \mathbf{x}_i \right) \right] \\
&= \sum_{i=1}^n \left[ \delta_i \left( -\frac{2}{\alpha^3} \sinh(z_i) \psi'_c(r_i) \mathbf{x}_i \right) \right. \\
& \quad \left. - (1 - \delta_i) \left( \frac{\xi_{i1z}}{2\alpha} K(\xi_{i2z}) \psi'_c(r_i) \mathbf{x}_i + \frac{1}{\alpha^3} \sinh(z_i) K'(\xi_{i2z}) \psi'_c(r_i) \mathbf{x}_i \right) \right] \\
&= \sum_{i=1}^n \left[ \delta_i \left( -\frac{2}{\alpha^3} \sinh(z_i) \right) - (1 - \delta_i) \left( \frac{\xi_{i1z}}{2\alpha} K(\xi_{i2z}) + \frac{1}{\alpha^3} \sinh(z_i) K'(\xi_{i2z}) \right) \right] \\
& \quad \times \psi'_c(r_i) \mathbf{x}_i.
\end{aligned}$$

Por fim, usando (A.17), (A.18), (A.23), (A.26), (A.27) e (A.28) a primeira derivada de  $\Upsilon_\beta$  em relação a  $\beta$  é obtida por

$$\begin{aligned}
\Upsilon'_{\beta\beta} &= \sum_{i=1}^n \left[ \delta_i \left( -\frac{1}{\alpha^2} \cosh(z_i) \psi'_c(r_i) \mathbf{x}_i + \frac{1}{4} \operatorname{sech}^2\left(\frac{z_i}{2}\right) \psi'_c(r_i) \mathbf{x}_i \right) \right. \\
& \quad \left. + (1 - \delta_i) \left( -\frac{\xi_{i2z}}{4} \psi'_c(r_i) \mathbf{x}_i K(\xi_{i2z}) + \frac{\xi_{i1z}}{2} K'(\xi_{i2z}) \left( -\frac{\xi_{i1z}}{2} \psi'_c(r_i) \mathbf{x}_i \right) \right) \right] w(\mathbf{x}_i) \mathbf{x}_i
\end{aligned}$$



$$\begin{aligned}
&= \sum_{i=1}^n \left[ \delta_i \left( \frac{1}{4} \operatorname{sech}^2 \left( \frac{z_i}{2} \right) \psi'_c(r_i) \mathbf{x}_i - \frac{1}{\alpha^2} \cosh(z_i) \psi'_c(r_i) \mathbf{x}_i \right) \right. \\
&\quad \left. + (1 - \delta_i) \left( -\frac{\xi_{i2z}}{4} \psi'_c(r_i) \mathbf{x}_i K(\xi_{i2z}) - \frac{\xi_{i1z}^2}{4} K'(\xi_{i2z}) \psi'_c(r_i) \mathbf{x}_i \right) \right] w(\mathbf{x}_i) \mathbf{x}_i \\
&= \sum_{i=1}^n \left[ \delta_i \left( \frac{1}{4} \operatorname{sech}^2 \left( \frac{z_i}{2} \right) - \frac{1}{\alpha^2} \cosh(z_i) \right) \right. \\
&\quad \left. - (1 - \delta_i) \left( \frac{\xi_{i2z}}{4} K(\xi_{i2z}) + \frac{\xi_{i1z}^2}{4} K'(\xi_{i2z}) \right) \right] \psi'_c(r_i) w(\mathbf{x}_i) \mathbf{x}_i \mathbf{x}'_i.
\end{aligned}$$



# Referências Bibliográficas

- Andrews, D. F. e F. R. Hampel. *Robust estimates of location: Survey and advances*, Volume 1280. Princeton University Press 2015.
- Barnett, V. e T. Lewis. *Outliers in Statistical Data*. Wiley series in probability and mathematical statistics. John Wiley and Sons 1978.
- Barros, M. Modelos de regressao birnbaum-saunders generalizados. *Unpublished Ph. D. thesis, Sao Paulo University, Brazil (in Portuguese)* (2007).
- Barros, M. e G. A. Paula. Discussion of “birnbaum-saunders distributions: A review of models, analysis and applications”. *Applied Stochastic Models in Business and Industry* 35(1) (2019), 96–99.
- Barros, M., G. A. Paula, e V. Leiva. A new class of survival regression models with heavy-tailed errors: robustness and diagnostics. *Lifetime data analysis* 14(3) (2008), 316–332.
- Birnbaum, Z. W. e S. C. Saunders. A new family of life distributions. *Journal of applied probability* 6(2) (1969), 319–327.
- Collett, D. *Modelling survival data in medical research*. Chapman and Hall/CRC 2015.
- Colosimo, E. A. e S. R. Giolo. Análise de sobrevivência. *São Paulo: Edgard Blücher* (2006).
- Committee, A. M. et al. Robust statistics—how not to reject outliers. part 1. basic concepts. *Analyst* 114(12) (1989), 1693–1697.
- Cribari-Neto, F. e S. G. Zarkos. R: Yet another econometric programming environment. *Journal of Applied Econometrics* 14(3) (1999), 319–329.

- Farcomeni, A. e L. Ventura. An overview of robust methods in medical research. *Statistical Methods in Medical Research* 21(2) (2012), 111–133.
- Hampel, F. R. Contribution to the theory of robust estimation. *Ph. D. Thesis, University of California, Berkeley* (1968).
- Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, e W. A. Stahel. *Robust statistics*. Wiley Online Library 1986.
- Heritier, S., E. Cantoni, S. Copt, e M.-P. Victoria-Feser. *Robust methods in biostatistics*, Volume 825. John Wiley & Sons 2009.
- Huber, P. J. Robust estimation of a location parameter: *Annals mathematics statistics*, 35 (1964), 73–101.
- Huber, P. J. *Robust statistical procedures*, Volume 68. Siam 1996.
- Huber, P. J. et al. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Volume 1, pp. 221–233. University of California Press.
- Huber, P. J. et al. Robust regression: asymptotics, conjectures and monte carlo. *The Annals of Statistics* 1(5) (1973), 799–821.
- Huber, P. J. e E. M. Ronchetti. Robust statistics john wiley & sons. *New York* 1(1) (1981).
- Ihaka, R. e R. Gentleman. R: a language for data analysis and graphics. *Journal of computational and graphical statistics* 5(3) (1996), 299–314.
- Kalbfleisch, J. D. e R. L. Prentice. *The statistical analysis of failure time data*, Volume 360. John Wiley & Sons 2011.
- Lee, E. T. e J. Wang. *Statistical methods for survival data analysis*, Volume 476. John Wiley & Sons 2003.
- Leiva, V. *The Birnbaum-Saunders Distribution* (1 ed.). Elsevier, Academic Press 2016.

- Locatelli, I., A. Marazzi, e V. J. Yohai. Robust accelerated failure time regression. *Computational Statistics & Data Analysis* 55(1) (2011), 874–887.
- Maronna, R. A., R. D. Martin, V. J. Yohai, e M. Salibián-Barrera. *Robust Statistics: Theory and Methods (with R)*. Wiley 2018.
- Paula, G. A. *Modelos de regressão: com apoio computacional*. IME-USP São Paulo 2013.
- Pinto, J. D., A. M. Carvalho, e S. Vinga (2015). Outlier detection in cox proportional hazards models based on the concordance c-index. In *International Workshop on Machine Learning, Optimization and Big Data*, pp. 252–256. Springer.
- Rieck, J. *Statistical Analysis for the Birnbaum-Saunders Fatigue Life Distribution. unpublished Ph. D.* Ph. D. thesis, thesis, Clemson University, Dept. of Mathematical Statistics 1989.
- Rieck, J. R. e J. R. Nedelman. A log-linear model for the birnbaum-saunders distribution. *Technometrics* 33(1) (1991), 51–60.
- Rousseeuw, P. J. e A. M. Leroy. *Robust regression and outlier detection*, Volume 1. Wiley Online Library 1987.
- Rousseeuw, P. J. e B. C. Van Zomeren. Unmasking multivariate outliers and leverage points. *Journal of the American Statistical association* 85(411) (1990), 633–639.
- Silva, D. K. X. Análise de diagnóstico para o modelo de regressão log-birnbaum-saunders generalizado (2013). Dissertação de Mestrado. UAMat/UFCG, Campina Grande.
- Sinha, S. K. Robust analysis of generalized linear mixed models. *Journal of the American Statistical Association* 99(466) (2004), 451–460.
- Sinha, S. K. Robust estimation in accelerated failure time models. *Lifetime data analysis* (2018), 1–27.
- Staudte, R. G. e S. J. Sheather. *Robust estimation and testing*. Wiley Online Library 1990.

- Tukey, J. W. A survey of sampling from contaminated distributions. *Contributions to probability and statistics* (1960), 448–485.
- Villegas, C., G. A. Paula, e V. Leiva. Birnbaum-saunders mixed models for censored reliability data analysis. *IEEE Transactions on Reliability* 60(4) (2011), 748–758.
- Wang, Z., A. F. Desmond, e X. Lu. Modified censored moment estimation for the two-parameter birnbaum–saunders distribution. *Computational Statistics & Data Analysis* 50(4) (2006), 1033–1051.
- Wei, L.-J. The accelerated failure time model: a useful alternative to the cox regression model in survival analysis. *Statistics in medicine* 11(14-15) (1992), 1871–1879.
- Yohai, V. J. Robust estimation in the linear model. *The Annals of Statistics* (1974), 562–567.