

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

**Modelo de regressão chances de sobrevivência proporcionais
para dados discretos com presença de censura**

Marcílio Ramos Pereira Cardial

Tese de Doutorado do Programa Interinstitucional de Pós-Graduação em
Estatística (PIPGEs)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Marcílio Ramos Pereira Cardial

**Modelo de regressão chances de sobrevivência
proporcionais para dados discretos com presença de
censura**

Tese apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP e ao Departamento de Estatística – DEs-UFSCar, como parte dos requisitos para obtenção do título de Doutor em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística. *VERSÃO REVISADA*

Área de Concentração: Estatística

Orientadora: Profa. Dra. Juliana Cobre

Coorientador: Prof. Dr. Eduardo Yoshio Nakano

**USP – São Carlos
Junho de 2023**

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

C267m Cardial, Marcílio Ramos Pereira
 Modelo de regressão chances de sobrevivência
proporcionais para dados discretos com presença de
censura / Marcílio Ramos Pereira Cardial;
orientadora Juliana Cobre; coorientador Eduardo
Yoshio Nakano. -- São Carlos, 2023.
 115 p.

 Tese (Doutorado - Programa Interinstitucional de
Pós-graduação em Estatística) -- Instituto de Ciências
Matemáticas e de Computação, Universidade de São
Paulo, 2023.

 1. Tempos de sobrevivência discretos. 2. Modelo
de chances proporcionais. 3. Modelo de regressão.
I. Cobre, Juliana, orient. II. Nakano, Eduardo
Yoshio, coorient. III. Título.

Marcílio Ramos Pereira Cardial

Odds proportional survival regression model for censored
discrete data

Thesis submitted to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP and to the Department of Statistics – DEs-UFSCar – in accordance with the requirements of the Statistics Interagency Graduate Program, for the degree of Doctor in Statistics. *FINAL VERSION*

Concentration Area: Statistics

Advisor: Profa. Dra. Juliana Cobre

Co-advisor: Prof. Dr. Eduardo Yoshio Nakano

USP – São Carlos

June 2023

À minha família

AGRADECIMENTOS

Primeiramente, a Deus, por todas as conquistas que já me fez alcançar não deixando em nenhum momento me faltar saúde, paz e disposição para realizar as minhas tarefas diárias. Certamente sem sua presença em minha vida nada seria possível, muito menos a execução deste trabalho.

Agradeço a minha família, aos meus pais Domingos e Maria e a minha esposa Sirlene que sempre me apoiaram em minhas decisões. Também agradeço a minha filha Yasmin, que é o maior presente que Deus me concedeu e a motivação para enfrentar com disposição momentos difíceis.

Aos professores Juliana Cobre e Eduardo Yoshio Nakano pela ótima orientação e por dedicar bastante tempo compartilhando seus conhecimentos comigo, sempre com muita paciência e disposição.

Aos meus amigos e colegas de Pós-graduação em especial ao Alex Mota, Luís Felipe, Alex de La Cruz, Isaac, Thiago, Michele, Sarah e Josimara pela amizade e companherismo indispensáveis.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Por fim, à todos que direta ou indiretamente contribuíram para que esse objetivo fosse alcançado.

RESUMO

CARDIAL, M. R. P. **Modelo de regressão chances de sobrevivência proporcionais para dados discretos com presença de censura**. 2023. 115 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2023.

Os modelos de sobrevivência, em sua maioria, consideram tempos de sobrevivência contínuos. Apesar disso, em vários estudos estes tempos são discretos, sendo desaconselhável, em algumas ocasiões, a utilização de um modelo contínuo para análise de dados discretos. Um dos modelos de regressão mais populares na análise de dados de sobrevivência é o modelo de riscos proporcionais de Cox, cuja principal característica é considerar que as covariáveis têm um efeito multiplicativo na função de risco. No entanto, essa característica não pode ser satisfeita quando os tempos de sobrevivência são discretos, devido ao fato da função de risco ser limitada no intervalo $(0,1)$. Para resolver esse problema, Cox sugeriu uma alternativa discreta de seu modelo. Uma outra alternativa de modelo de regressão foi apresentada por Bennett, que assume que as covariáveis têm um efeito multiplicativo na chance (*odds*) de sobrevivência. Esses modelos são denominados como modelos de chances (de sobrevivência) proporcionais. Neste contexto, o presente trabalho tem como objetivo considerar a modelagem de chances proporcionais como uma alternativa para construção de modelos de regressão para dados de sobrevivência discretos. Mais especificamente, os objetivos são: (a) o estudo do modelo de chances de sobrevivência proporcionais para tempos contínuos; (b) a construção do modelo de regressão para dados com chances de sobrevivência proporcionais e tempos discretos; (c) obtenção das estimativas pontuais e intervalares dos parâmetros do modelo; (d) propor procedimentos para verificação da suposição de chances proporcionais e da qualidade do ajuste do modelo; (e) ilustração do modelo e procedimentos propostos em um conjunto de dados reais. Os resultados obtidos em dados simulados indicaram evidências das propriedades assintóticas dos estimadores e o modelo proposto apresentou um bom ajuste ao conjunto de dados reais, provando ser uma boa alternativa para a modelagem de dados de sobrevivência discretos com covariáveis.

Palavras-chave: Tempos de sobrevivência discretos, Modelo de chances proporcionais, Modelo de regressão.

ABSTRACT

CARDIAL, M. R. P. **Odds proportional survival regression model for censored discrete data.** 2023. 115 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2023.

Survival models, in their majority, consider continuous survival times. However, in several studies these times are discrete, and in some occasions, it is not advisable to use a continuous model to analyze discrete data. One of the most popular regression models in the analysis of survival data is the Cox proportional hazards model, whose main characteristic is to consider that the covariates have a multiplicative effect on the hazard function. However, this feature cannot be satisfied when survival times are discrete, due to the hazard function being bounded in the interval $(0, 1)$. To solve this problem, Cox suggested a discrete alternative of his model. Another alternative regression model was presented by Bennett, which assumes that covariates have a multiplicative effect on the odds of survival. These models are referred to as proportional odds (survival) models. In this context, the present paper aims to consider proportional odds modeling as an alternative for building regression models for discrete survival data. More specifically, the objectives are: (a) to study the proportional odds model for continuous time; (b) to build the regression model for data with proportional odds of survival and discrete time; (c) to obtain point and interval estimates of the model parameters; (d) to propose procedures to verify the proportional odds assumption and the quality of the model fit; (e) to illustrate the model and proposed procedures on a real data set. The results obtained on simulated data indicated evidence of the asymptotic properties of the estimators and the proposed model showed a good fit to the real data set, proving to be a good alternative for modeling discrete survival data with covariates.

Keywords: Discrete survival times, Proportional odds model, Regression model.

LISTA DE ILUSTRAÇÕES

Figura 1 – Taxa de falha da distribuição Weibull Discreta para diferentes valores dos parâmetros.	34
Figura 2 – Taxas de falhas e respectivas razões entre os riscos para diversas distribuições base e valores da razão de chances de sobrevivência θ (Caso contínuo). . . .	39
Figura 3 – Taxas de falhas e respectivas razões entre os riscos para diversas distribuições base e valores da razão de chances de sobrevivência θ (Caso discreto). . . .	41
Figura 4 – Exemplo 1 (chances de sobrevivência proporcionais).	46
Figura 5 – Exemplo 2 (chances de sobrevivência proporcionais).	47
Figura 6 – Exemplo 3 (chances de sobrevivência não proporcionais).	48
Figura 7 – Média das estimativas, EQM e PC referentes ao parâmetro q	51
Figura 8 – Distribuições empíricas de \hat{q} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas.	51
Figura 9 – Distribuições empíricas de \hat{u} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas.	52
Figura 10 – Média das estimativas, EQM e PC referentes ao parâmetro η	53
Figura 11 – Distribuições empíricas de $\hat{\eta}$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas.	53
Figura 12 – Distribuições empíricas de \hat{v} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas.	54
Figura 13 – Média das estimativas, EQM e PC referentes ao parâmetro β_1	55
Figura 14 – Distribuições empíricas de $\hat{\beta}_1$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas.	55
Figura 15 – Média das estimativas, EQM e PC referentes ao parâmetro β_2	56
Figura 16 – Distribuições empíricas de $\hat{\beta}_2$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas.	56
Figura 17 – Média das estimativas, EQM e PC referentes ao parâmetro q por percentual de censura.	59
Figura 18 – Distribuições empíricas de \hat{q} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 5% de censuras.	60
Figura 19 – Distribuições empíricas de \hat{q} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 10% de censuras.	60
Figura 20 – Distribuições empíricas de \hat{q} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 15% de censuras.	61

Figura 21 – Distribuições empíricas de \hat{q} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 25% de censuras.	61
Figura 22 – Média das estimativas, EQM e PC referentes ao parâmetro η por percentual de censura.	62
Figura 23 – Média das estimativas, EQM e PC referentes ao parâmetro β_1 por percentual de censura.	63
Figura 24 – Média das estimativas, EQM e PC referentes ao parâmetro β_2 por percentual de censura.	64
Figura 25 – Função de sobrevivência estimada pelo método de Kaplan-Meier do conjunto dos dados da Tabela 3.	66
Figura 26 – Função de sobrevivência estimada para cada nível das covariáveis do conjunto de dados da Tabela 3.	67
Figura 27 – Verificação de suposição de chances de sobrevivência proporcionais para a covariável idade (Dados de Leucemia).	68
Figura 28 – Ajuste dos modelos aos dados de Leucemia por nível da covariável idade.	71
Figura 29 – Sobrevivência via Estimador de Kaplan-Meier do conjunto dos dados da Tabela 9.	73
Figura 30 – Verificação de suposição de chances de sobrevivência proporcionais para as covariáveis Bun (z_1) e Proteína (z_2).	76
Figura 31 – Ajuste dos modelos MRCSP-WD e MRWD aos dados de Mieloma por nível das covariáveis.	79
Figura 32 – Distribuições empíricas de \hat{u} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 5% de censuras.	105
Figura 33 – Distribuições empíricas de \hat{u} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 10% de censuras.	106
Figura 34 – Distribuições empíricas de \hat{u} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 15% de censuras.	106
Figura 35 – Distribuições empíricas de \hat{u} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 25% de censuras.	107
Figura 36 – Distribuições empíricas de $\hat{\eta}$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 5% de censuras.	107
Figura 37 – Distribuições empíricas de $\hat{\eta}$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 10% de censuras.	108
Figura 38 – Distribuições empíricas de $\hat{\eta}$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 15% de censuras.	108
Figura 39 – Distribuições empíricas de $\hat{\eta}$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 25% de censuras.	109
Figura 40 – Distribuições empíricas de \hat{v} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 5% de censuras.	109

- Figura 41 – Distribuições empíricas de \hat{v} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 10% de censuras. 110
- Figura 42 – Distribuições empíricas de \hat{v} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 15% de censuras. 110
- Figura 43 – Distribuições empíricas de \hat{v} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 25% de censuras. 111
- Figura 44 – Distribuições empíricas de $\hat{\beta}_1$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 5% de censuras. 111
- Figura 45 – Distribuições empíricas de $\hat{\beta}_1$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 10% de censuras. 112
- Figura 46 – Distribuições empíricas de $\hat{\beta}_1$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 15% de censuras. 112
- Figura 47 – Distribuições empíricas de $\hat{\beta}_1$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 25% de censuras. 113
- Figura 48 – Distribuições empíricas de $\hat{\beta}_2$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 5% de censuras. 113
- Figura 49 – Distribuições empíricas de $\hat{\beta}_2$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 10% de censuras. 114
- Figura 50 – Distribuições empíricas de $\hat{\beta}_2$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 15% de censuras. 114
- Figura 51 – Distribuições empíricas de $\hat{\beta}_2$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 25% de censuras. 115

LISTA DE TABELAS

Tabela 1 – Análise gráfica da validade de chances de sobrevivência proporcionais.	45
Tabela 2 – Cenários das simulações.	50
Tabela 3 – Tempo de sobrevivência e dados para dois fatores prognósticos possíveis de 30 pacientes com AML.	66
Tabela 4 – Teste de significância das covariáveis para MRCSP-WD e MRWD múltiplo (Dados de Leucemia).	67
Tabela 5 – Teste de significância da covariável idade para MRCSP-WD e MRWD (Dados de Leucemia).	68
Tabela 6 – Estimativas dos parâmetros do MRCSP-WD (Dados de Leucemia).	69
Tabela 7 – Estimativas dos parâmetros do MRWD (Dados de Leucemia).	70
Tabela 8 – Erros máximos oriundos da estimação dos modelos (Dados de Leucemia).	70
Tabela 9 – Tempo de sobrevivência de pacientes em um estudo sobre Mieloma múltiplo.	72
Tabela 10 – Características dos pacientes dos dados de Mieloma Múltiplo.	74
Tabela 11 – Teste de significância das covariáveis para MRCSP-WD e MRWD múltiplos (Dados de Mieloma).	74
Tabela 12 – Teste de significância das covariáveis para MRCSP-WD e MRWD para as covariáveis significativas (Dados de Mieloma).	75
Tabela 13 – Estimativas dos parâmetros do MRCSP-WD conjunto - covariáveis significativas (Dados mieloma).	77
Tabela 14 – Estimativas dos parâmetros do MRWD conjunto - covariáveis significativas (Dados mieloma).	78
Tabela 15 – Erros máximos oriundos da estimação dos modelos (Dados de Mieloma).	78
Tabela 16 – Média das estimativas, EQM e PC dos parâmetros MRCSP-WD considerando os cenários da simulação e diversos tamanhos amostrais.	103
Tabela 17 – Média das estimativas, EQM e Probabilidade de cobertura (PC) dos parâmetros MRCSP-WD considerando os cenários da simulação e diversos tamanhos amostrais e percentuais de censura.	104

LISTA DE ABREVIATURAS E SIGLAS

AML	Acute Myeloid Leukemia
Bun	Nitrogênio ureico no sangue
DTPBT	<i>Discrete two-parameter bathtub distribution</i>
EQM	erro quadrático médio
FDA	Função Distribuição Acumulada
K-M	Kaplan-Meier
MCP	Modelo de Chances Proporcionais
MCSP	Modelo de Chances de Sobrevivência Proporcionais
MLG	Modelo Linear Generalizado
MO	Marshall–Olkin
MRCSP	Modelo de Regressão Chances de Sobrevivência Proporcionais
MRCSP-WD	Modelo de Regressão Chances de Sobrevivência Proporcionais Weibull Discreto
MRWD	Modelo de Regressão Weibull Discreto
MTL	modelo de transformação linear
PC	probabilidade de cobertura
TRV	teste da razão de verossimilhanças
WD	Weibull discreta
WDE	Weibull discreta exponenciada

SUMÁRIO

1	INTRODUÇÃO	23
2	CONCEITOS BÁSICOS	25
2.1	Análise de Sobrevida	25
2.2	Inferência Frequentista	27
2.2.1	<i>Método de Máxima Verossimilhança</i>	27
2.2.2	<i>Intervalo de Confiança</i>	28
2.2.3	<i>Teste de Hipóteses</i>	29
2.3	Tempos Discretos	30
2.3.1	<i>Discretização de distribuições contínuas</i>	31
2.3.2	<i>Distribuição Weibull discreta</i>	32
3	MODELO DE CHANCES DE SOBREVIVÊNCIA PROPORCIONAIS	35
3.1	Revisão de Literatura	35
3.2	MCSP para tempos discretos	40
3.2.1	<i>Modelo de Regressão Chances de Sobrevida Proporcional Weibull Discreto</i>	42
3.3	Verificação da Suposição de chances de sobrevivência proporcionais	44
4	SIMULAÇÕES COMPUTACIONAIS	49
4.1	Tempos discretos sem a presença de censura	50
4.2	Tempos discretos com a presença de censura	57
5	APLICAÇÃO EM DADOS REAIS	65
5.1	Dados de Leucemia	65
5.1.1	<i>Verificação da suposição de chances de sobrevivência proporcionais</i>	68
5.1.2	<i>Inferência dos parâmetros do MRCSP-WD</i>	69
5.1.3	<i>Verificação do ajuste do MRCSP-WD e comparação com o Modelo de Regressão Weibull Discreto (MRWD)</i>	69
5.2	Dados de Mieloma	71
5.2.1	<i>Verificação de suposição de chances de sobrevivência proporcionais (dados Mieloma)</i>	75
5.2.2	<i>Modelo de regressão chances de sobrevivência proporcionais (MRCSP)</i>	77

5.2.3	<i>Verificação do ajuste do MRCSP-WD e comparação com o Modelo de Regressão Weibull Discreto (MRWD)</i>	78
6	CONSIDERAÇÕES FINAIS	81
	REFERÊNCIAS	83
APÊNDICE A	MCSP COMO MTL	87
APÊNDICE B	FUNÇÕES EM ANÁLISE DE SOBREVIVÊNCIA DO MCSP	89
B.1	Função de sobrevivência	89
B.2	Função de risco (caso contínuo)	89
B.3	Função de probabilidade (caso discreto)	90
B.4	Função de risco (caso discreto)	91
APÊNDICE C	SCRIPTS DESENVOLVIDOS	93
APÊNDICE D	INTERVALO DE CONFIANÇA DOS PARÂMETROS DA DISTRIBUIÇÃO WD	99
ANEXO A	RESULTADOS DAS SIMULAÇÕES	103
ANEXO B	DISTRIBUIÇÕES EMPIRICAS DAS ESTIMATIVAS COM AS CURVAS GAUSSIANAS SOBREPOSTAS	105

INTRODUÇÃO

A análise de sobrevivência é uma das áreas da estatística responsável pela análise de experimentos por meio de um conjunto de técnicas e modelos estatísticos cuja variável resposta é o tempo até a ocorrência do evento de interesse (falha), podendo ser atribuído, por exemplo, ao óbito de um indivíduo. Uma das características desse ramo da estatística é a presença de observações incompletas ou parciais da variável resposta denominados dados censurados que são importantes e úteis nas análises, uma vez que a omissão destes pode acarretar conclusões viciadas.

Embora haja vários estudos em que os tempos de sobrevivência sejam discretos (intrinsecamente discretos ou agrupados em intervalos), nos modelos de sobrevivência, em sua maioria, o tempo de sobrevivência é representado por uma variável aleatória contínua. Segundo [Tutz e Schmid \(2016\)](#), a princípio, as abordagens de modelagem tradicionais para tempos contínuos podem ser aplicadas a dados discretos (pelo menos se os tempos discretos estiverem agrupados em intervalos que são igualmente espaçados). Porém, os métodos estatísticos que são especialmente concebidos para tempos discretos têm uma série de vantagens. [Nakano e Carrasco \(2006\)](#), por meio de estudos de simulação e aplicação em dados reais, ilustraram ser desaconselhável, a utilização de um modelo contínuo para análise de dados discretos.

Assim, nos últimos anos, um grande número de artigos de pesquisa lidando com distribuições discretas oriundas da discretização de distribuições de variáveis aleatórias contínuas apareceram na literatura estatística em um contexto de análise de sobrevivência. Entre estes estão: distribuição Weibull discreta (WD) em [Nakagawa e Osaki \(1975\)](#) e [Vila, Nakano e Saulo \(2018\)](#), Weibull Geométrica discreta em [Jayakumar e Babu \(2018\)](#), Weibull discreta exponenciada (WDE) em [Cardial, Fachini-Gomes e Nakano \(2020\)](#), Banheira de dois parâmetros discreta [*Discrete two-parameter bathtub distribution (DTPBT)*] em [Sarhan \(2017\)](#) e Gumbel discreta em [Chakraborty \(2015\)](#).

Um aspecto comum dos estudos de sobrevivência é a presença de covariáveis ou variáveis

explicativas que representam a heterogeneidade existente, tais como, idade, sexo e tipo de tratamento. A forma mais eficiente de acomodar o efeito dessas covariáveis, propondo uma relação entre os tempos discretos de sobrevivência e as variáveis explicativas de interesse, é por meio de um modelo de regressão adequado para dados discretos de sobrevivência.

Um dos modelos de regressão mais populares na análise de dados de sobrevivência é o modelo de riscos proporcionais de [Cox \(1972\)](#), que no referido artigo introduziu o modelo para tempos discretos, em que as covariáveis têm um efeito multiplicativo na chance (*odds*) de risco. Um tratamento abrangente da metodologia estatística foi feito por [Tutz e Schmid \(2016\)](#) e para regressão semiparamétrica em [Berger e Schmid \(2018\)](#). Recentes extensões têm sido propostas em trabalhos como [Vallejos e Steel \(2017\)](#) e [Berger et al. \(2019\)](#).

No modelo de riscos proporcionais, a função de risco base em grande parte dos estudos é tratada como um parâmetro incômodo de alta dimensão e é altamente errático podendo ser estimada por diversos métodos (versão semiparamétrica). No entanto, o comportamento da função de risco é de potencial interesse médico porque está diretamente relacionado ao curso de uma doença. Para estimar esse risco suavemente, algum tipo de modelo paramétrico pode ser apropriado ([ROYSTON; PARMAR, 2002](#)). Desse modo, um modelo de regressão em sua versão paramétrica foi proposto em [Vieira \(2018\)](#).

Apesar do modelo de riscos proporcionais ser bem conhecido, o Modelo de Chances Proporcionais (MCP), em que as covariáveis têm um efeito multiplicativo na chance (*odds*) de sobrevivência, sendo denominado como Modelo de Chances de Sobrevivência Proporcional (MCSP), é potencialmente um concorrente do modelo de Cox, e também tem uma história bastante longa com sua utilização atrelada a dados contínuos. Sendo apresentado pela primeira vez em uma estrutura paramétrica por [Bennett \(1983\)](#), um extensivo estudo para demonstração das propriedades do modelo foi realizado por [Murphy, Rossini e Vaart \(1997\)](#), posteriormente foi proposto por [Yang e Prentice \(1999\)](#) a inferência semiparamétrica e aplicações fazendo uso do referido modelo realizadas em [Royston e Parmar \(2002\)](#), [Wang e Wang \(2020\)](#) e [Zhou, Zhang e Lu \(2022\)](#). Portanto, há uma ampla literatura do MCSP para dados contínuos. No entanto, há uma literatura limitada quando se trata de dados discretos.

Diante desse panorama e considerando que, até o momento, não foram observados trabalhos similares considerando modelos para dados discretos, o objetivo do presente trabalho consiste na inferência paramétrica para o Modelo de Regressão Chances de Sobrevivência Proporcional (MRCSP) para dados discretos com a presença de censura. Com a proposição do modelo, serão propostos procedimentos para a verificação de suas suposições, assim como formas de verificar sua qualidade de ajuste. No Capítulo 2 são apresentados conceitos estatísticos básicos com enfoque em tempos discretos essenciais para a construção do MRCSP apresentado no Capítulo 3. Por fim, os Capítulos 4 e 5 referem-se a ilustração do modelo por meio de dados simulados e por intermédio de dados obtidos na literatura respectivamente.

CONCEITOS BÁSICOS

No presente capítulo, serão apresentados conceitos básicos. Na Seção 2.1, são apresentados os principais conceitos de análise de sobrevivência, adicionalmente serão apresentadas as funções e relações que caracterizam essa área da estatística considerando tempos discretos. A Seção 2.2 refere-se à inferência de dados de sobrevivência a partir de uma abordagem frequentista. Por fim, a Seção 2.3 trata-se da caracterização do tempo até a ocorrência do evento de interesse discreto tal como a sua tratativa tendo em vista distribuições de probabilidade.

2.1 Análise de Sobrevivência

A análise de sobrevivência é definida como um conjunto de procedimentos estatísticos para análise de dados relacionados à variável resposta tempo até a ocorrência de um determinado evento de interesse. Tendo sua principal aplicação na área médica, na qual indivíduos são acompanhados até a ocorrência de um evento de interesse, sendo assim, anotado o tempo de sobrevivência.

Segundo Colosimo e Giolo (2006), o estudo nesse ramo da estatística tem como principal característica a presença de dados censurados, que é a existência de observações incompletas ou parciais, ou seja, quando não se observa o evento de interesse durante o tempo de acompanhamento, uma vez que, tal observação para todos os elementos de um determinado estudo nem sempre é possível. Sendo assim, há a necessidade da introdução de uma variável que indique se o tempo de vida foi ou não observado. Essa variável é definida na literatura como variável indicadora de censura ou falha. É válido lembrar que, mesmo censurados, todos os resultados provenientes de um estudo devem ser usados na análise estatística, pois a omissão da censura nos cálculos estatísticos poderá acarretar em estimativas viciadas.

Quanto à classificação dos tipos de censura, pode-se dividi-las em três grupos que são a censura à direita, à esquerda e intervalar. Na censura à direita, o tempo de interesse está à direita

do tempo registrado, enquanto que na censura à esquerda o evento de interesse está à esquerda do tempo registrado. Na censura intervalar, quando os indivíduos são acompanhados em visitas periódicas, sabe-se apenas que o evento de interesse ocorreu em algum momento entre duas visitas. Ainda, pode-se dividir a censura à direita em três subgrupos:

- Censura do Tipo I: o estudo termina após um período pré-estabelecido de tempo.
- Censura do Tipo II: o estudo termina após ter ocorrido o evento de interesse em um número pré-estabelecido de indivíduos estudados.
- Censura do Tipo Aleatório: ocorre quando um indivíduo é retirado no decorrer do estudo sem ter ocorrido a falha. Também ocorre, por exemplo, se o indivíduo falhar por uma razão diferente da estudada.

No presente trabalho, será considerada apenas a censura à direita aleatória, que é a mais utilizada em trabalhos de análise de sobrevivência. Seja \tilde{T} uma variável aleatória que representa o tempo até a falha e C uma variável aleatória, que representa o tempo até a censura. Considere, $T = \min(\tilde{T}, C)$, que denota o tempo observado do indivíduo com $\delta = \mathbb{I}_{[0, C]}(\tilde{T})$ representando a variável indicadora de censura ou falha que será igual a 0 se o tempo de sobrevida for censurado à direita e 1 se for de falha.

Nesse contexto, ao considerar T uma variável aleatória que assume valores inteiros não negativos, é possível especificar as funções que a representam em análise de sobrevivência. Assim, a função de probabilidade, que atribui probabilidade a cada um dos possíveis valores assumidos pela variável, é definida por $p(t) = P(T = t)$ e deve satisfazer as seguintes condições (MAGALHÃES, 2006):

$$i) 0 \leq p(t) \leq 1, \quad \forall t = 0, 1, 2, \dots;$$

$$ii) \sum_{t=0}^{\infty} p(t) = 1.$$

Por sua vez, a função de sobrevivência no tempo t , é a capacidade de um indivíduo sobreviver até o tempo t , e é definida como:

$$\begin{aligned} S(t) &= P(T > t) \\ &= \sum_{k=t+1}^{\infty} P(T = k), \quad t = 0, 1, 2, \dots \end{aligned}$$

No caso em que as distribuições de probabilidade são discretas, a função de risco é igual a zero, exceto nos pontos em que pode ocorrer uma falha. A função de risco para dados discretos, denotada por $h(t)$ e definida no intervalo $0 < h(t) < 1$, pode ser expressa como $h(t) = P(T = t | T \geq t)$ e por meio dessa definição é possível estabelecer a seguinte relação:

$$h(t) = \frac{p(t)}{S(t) + p(t)}, \quad t = 0, 1, 2, \dots \quad (2.1)$$

É possível estabelecer outras relações a partir das funções citadas, como:

$$F(t) = P(T \leq t) = \sum_{k=0}^t p(k) = 1 - S(t), \quad t = 0, 1, 2, \dots, \quad (2.2)$$

$$S(t) = \prod_{k=0}^t [1 - h(k)], \quad t = 0, 1, 2, \dots \quad (2.3)$$

e

$$p(t) = \begin{cases} 1 - S(0), & \text{se } t = 0 \\ S(t-1) - S(t), & \text{se } t = 1, 2, \dots, \end{cases} \quad (2.4)$$

em que $F(\cdot)$ representa a função distribuição acumulada. Mais detalhes sobre as funções e relações apresentadas podem ser encontrados em [Tutz e Schmid \(2016\)](#).

Segundo [Colosimo e Giolo \(2006\)](#), o uso de técnicas não paramétricas são importantes para descrever os dados de sobrevivência pela sua simplicidade e facilidade de aplicação, sendo recomendado para uma análise inicial dos dados. Nesse sentido, o estimador de Kaplan-Meier (K-M) ([1958](#)) é a ferramenta mais utilizada para estimar a função de sobrevivência na presença de censuras e será adotado no presente trabalho. Vale ainda ressaltar que, tais técnicas não envolvem nenhuma estrutura paramétrica, o que inviabiliza uma análise mais elaborada incluindo covariáveis.

Pelo fato de os dados de sobrevivência terem em sua estrutura a ocorrência de censuras, é necessário incorporar essa informação no processo inferencial. Na próxima seção são apresentadas as ferramentas para estimação pontual e intervalar e testes de hipóteses no contexto de análise de sobrevivência para tempos de sobrevivência discretos na abordagem frequentista.

2.2 Inferência Frequentista

O processo inferencial segundo a abordagem frequentista é realizado a partir da função de verossimilhança. A estimação pontual dos parâmetros é obtida por meio do método de máxima verossimilhança apresentado na Subseção 2.2.1. Por sua vez, a estimação intervalar dos parâmetros é mostrada na Subseção 2.2.2 e teste de hipótese em relação aos parâmetros na Subseção 2.2.3.

2.2.1 Método de Máxima Verossimilhança

Este é atualmente o método de estimação mais utilizado na inferência frequentista. Além disso, há uma boa base teórica desenvolvida em uma variedade de situações ([MIGON; GAMERMAN; LOUZADA, 2014](#)).

Seja t_1, t_2, \dots, t_n uma amostra aleatória observada de uma variável aleatória discreta T ; $p(t; \boldsymbol{\phi})$ representa a função de probabilidade da população e $\boldsymbol{\phi} = (\phi_1, \dots, \phi_k)$ é o vetor de

parâmetros da função de probabilidade. A função de verossimilhança para ϕ , na ausência de censuras, é dada por:

$$L(\phi; \mathbf{t}) = \prod_{i=1}^n p(t_i; \phi). \quad (2.5)$$

No entanto, quando há a presença de censuras (principal característica em análise de sobrevivência), os dados censurados devem ser separados dos elementos que sofreram o evento de interesse, frequentemente chamados de dados não censurados.

Dessa forma, as observações são divididas em dois conjuntos, as r primeiras são as não censuradas $(1, 2, \dots, r)$ e sua contribuição para a função de verossimilhança é dada por $p(t_i; \phi)$, e as $n - r$ seguintes são as censuradas $(r + 1, r + 2, \dots, n)$, sendo sua contribuição para função de verossimilhança dada pela função de sobrevivência $S(t_i; \phi)$. Assim, a função de verossimilhança assume a seguinte forma:

$$L(\phi; \mathbf{t}) \propto \prod_{i=1}^r p(t_i; \phi) \prod_{i=r+1}^n S(t_i; \phi). \quad (2.6)$$

A expressão (2.6) é equivalente a:

$$L(\phi; \mathbf{t}, \boldsymbol{\delta}) \propto \prod_{i=1}^n [p(t_i; \phi)]^{\delta_i} [S(t_i; \phi)]^{1-\delta_i}, \quad (2.7)$$

em que δ_i é a variável indicadora de falha, que assume valor 1 se o tempo t_i for de falha e 0 se for censura à direita; $p(\cdot; \phi)$ e $S(\cdot; \phi)$ são, respectivamente, a distribuição de probabilidade e função de sobrevivência do modelo considerado, $i = 1, 2, \dots, n$.

Ao aplicar o logaritmo na função de verossimilhança na equação (2.7), tem-se:

$$\ell(\phi; \mathbf{t}, \boldsymbol{\delta}) = \sum_{i=1}^n [\delta_i \log [p(t_i; \phi)] + (1 - \delta_i) \log [S(t_i; \phi)]] + c, \quad (2.8)$$

em que c é uma constante que não depende de ϕ .

Os estimadores de máxima verossimilhança são os valores de ϕ que maximizam $L(\phi; \mathbf{t}, \boldsymbol{\delta})$ ou equivalentemente $\ell(\phi; \mathbf{t}, \boldsymbol{\delta})$, normalmente representado por $\hat{\phi}$, e são obtidos resolvendo o sistema de equações:

$$\frac{\partial \ell(\phi; \mathbf{t}, \boldsymbol{\delta})}{\partial \phi} = 0. \quad (2.9)$$

2.2.2 Intervalo de Confiança

Os intervalos de confiança fornecem um método para adicionar mais informações a um estimador $\hat{\phi}$, quando deseja-se estimar um parâmetro desconhecido ϕ . Assim, pode-se encontrar um intervalo em que almeja-se ter alta probabilidade de conter ϕ .

A construção de intervalos de confiança é feita a partir da matriz informação de Fisher observada definida por:

$$I_f(\boldsymbol{\phi}) = -\frac{\partial^2 \ell(\mathbf{t}, \boldsymbol{\delta} | \boldsymbol{\phi})}{\partial \boldsymbol{\phi}^2}.$$

Sob certas condições de regularidade tem-se que:

$$\widehat{Var}(\widehat{\boldsymbol{\phi}}) \approx [I_f(\widehat{\boldsymbol{\phi}})]^{-1},$$

em que $I_f(\cdot)$ é a informação de Fisher observada da amostra.

Utiliza-se, então, a distribuição assintótica do estimador de máxima verossimilhança. Para grandes amostras, essa propriedade estabelece que:

$$\widehat{\boldsymbol{\phi}} \stackrel{a}{\sim} N_k(\boldsymbol{\phi}, \widehat{Var}(\widehat{\boldsymbol{\phi}})),$$

ou seja, $\widehat{\boldsymbol{\phi}}$ converge assintoticamente para uma distribuição normal multivariada com média $\boldsymbol{\phi}$ e matriz de variância e covariância $\widehat{Var}(\widehat{\boldsymbol{\phi}})$, em que k é a dimensão de $\widehat{\boldsymbol{\phi}}$.

Assim, um intervalo aproximado de $(1 - \alpha) \times 100\%$ de confiança para o parâmetro ϕ_j , $j = 1, 2, \dots, k$, é dado por:

$$\left[\widehat{\phi}_j - Z_{(1-\alpha/2)} \sqrt{\widehat{Var}(\widehat{\phi}_j)}; \widehat{\phi}_j + Z_{(1-\alpha/2)} \sqrt{\widehat{Var}(\widehat{\phi}_j)} \right], \quad j = 1, 2, \dots, k, \quad (2.10)$$

em que $\widehat{\phi}_j$ é o estimador de máxima verossimilhança de ϕ_j , $Z_{(1-\alpha/2)}$ é o quantil $(1 - \alpha/2)$ de uma distribuição normal padrão e $\widehat{Var}(\widehat{\phi}_j)$ é a estimativa da variância de $\widehat{\phi}_j$.

De forma análoga, um intervalo de confiança para uma função $g(\cdot)$, diferenciável em $\boldsymbol{\phi}$, pode ser obtido por meio do método delta que estabelece que:

$$\widehat{g}(\widehat{\boldsymbol{\phi}}) \stackrel{a}{\sim} N\left(g(\boldsymbol{\phi}), [g'(\widehat{\boldsymbol{\phi}})]^2 \widehat{Var}(\widehat{\boldsymbol{\phi}})\right), \quad (2.11)$$

em que $\widehat{g}(\widehat{\boldsymbol{\phi}}) = g(\widehat{\boldsymbol{\phi}})$ é o estimador de máxima verossimilhança de $g(\boldsymbol{\phi})$ e $g'(\widehat{\boldsymbol{\phi}})$ é a derivada de primeira ordem de $g(\widehat{\boldsymbol{\phi}})$.

A expressão (2.11) é utilizada quando se há o interesse em calcular intervalos de confiança para funções dos parâmetros de uma determinada distribuição. Mais detalhes sobre o intervalo de confiança e o método delta podem ser encontrados em [Lehmann e Casella \(2006\)](#).

2.2.3 Teste de Hipóteses

Para um modelo com um vetor de parâmetros $\boldsymbol{\phi} = (\phi_1, \dots, \phi_k)'$, em algumas situações, deseja-se testar hipóteses relacionadas a este vetor ou a um subconjunto dele.

O teste da razão de verossimilhanças (TRV) é baseado na função de verossimilhança e envolve a comparação dos valores do logaritmo da função de verossimilhança maximizada e sob a hipótese nula $H_0 : \phi = \phi_0$. A estatística do teste é dada por:

$$TRV = -2 \log \left[\frac{L(\phi_0; \mathbf{t}, \boldsymbol{\delta})}{L(\hat{\phi}; \mathbf{t}, \boldsymbol{\delta})} \right] = 2[\log L(\hat{\phi}; \mathbf{t}, \boldsymbol{\delta}) - \log L(\phi_0; \mathbf{t}, \boldsymbol{\delta})], \quad (2.12)$$

que, sob H_0 , segue aproximadamente uma distribuição qui-quadrado com ρ graus de liberdade que representa a diferença do número de parâmetros dos modelos (modelo completo e modelo sob H_0). Para amostras grandes, H_0 é rejeitada, a um nível de $\alpha \times 100\%$ de significância (α é um valor estabelecido previamente), se $TRV > \chi_{\rho, 1-\alpha}^2$, em que, $\chi_{\rho, 1-\alpha}^2$ é o quantil $(1 - \alpha)$ de uma distribuição qui-quadrado com ρ graus de liberdade.

Alternativamente, o teste pode ser realizado por meio do nível descritivo (valor-p) do teste, que neste caso é definido por:

$$\text{valor-p} = P(\chi_{\rho}^2 > TRV),$$

em que χ_{ρ}^2 é uma variável aleatória com distribuição qui-quadrado com ρ graus de liberdade. Assim, a hipótese H_0 é rejeitada se $\text{valor-p} < \alpha$.

2.3 Tempos Discretos

O objetivo de muitas análises estatísticas, em particular da análise de sobrevivência, é modelar o tempo até a ocorrência do evento de interesse denominado como *time-to-event*. Conforme, [Berger e Schmid \(2018\)](#), na maioria das vezes, é assumido nessas análises que o tempo de sobrevivência é dado por uma variável aleatória medida em uma escala contínua, havendo uma extensiva literatura para esse caso. No entanto, na prática, as medições de tempo costumam ser discretas, havendo situações em que a hora exata do evento pode não ser conhecida, mas apenas o intervalo durante o qual ocorreu o evento de interesse.

De maneira geral, o tempo até ocorrência do evento de interesse discreto, conforme [Tutz e Schmid \(2016\)](#), ocorre como:

- medições intrinsecamente/genuinamente discretas;
- dados agrupados.

Os dados agrupados representam eventos em intervalos de tempo subjacentes, e a variável resposta refere-se a um intervalo, que pode ter tamanhos iguais ou diferentes, há exemplos utilizados em trabalhos como: tempo em meses até a morte de homens diagnosticados com Síndrome de Imunodeficiência Adquirida (AIDS) ([BRUNELLO; NAKANO, 2015](#)) e tempo em meses de pacientes com câncer de pescoço e cabeça ([CARDIAL; FACHINI-GOMES; NAKANO, 2020](#)).

A respeito dos tempos genuinamente discretos, em que as medições representam números naturais, há diversas aplicações como: tempo até a gravidez que em estudos clínicos geralmente é medido pelo número de ciclos menstruais (BERGER; SCHMID, 2018), tempo até a evasão em universidade medido em semestres (VALLEJOS; STEEL, 2017) e tempo até a degeneração macular relacionada à idade entre idosos que foram monitorados por visitas de estudo anuais (BERGER *et al.*, 2019).

De acordo com Tutz e Schmid (2016), métodos estatísticos desenvolvidos para tempo discreto utilizados para esse fim, possuem uma série de vantagens:

- A consideração de modelos para tempos discretos tem a vantagem de que os riscos podem ser formuladas como probabilidades condicionais. Portanto, eles são muito mais acessíveis para interpretação do que funções de risco contínuas;
- Na prática, muitos tempos de evento são intrinsecamente discretos ou são observados em um escala discreta. Consequentemente, usar modelos para tempos discretos é mais apropriado do que a aproximação dos dados observados por um modelo de sobrevivência contínuo;
- Em contraste com os modelos de sobrevivência para tempo contínuo, modelos para eventos discretos não causam problemas com empates ¹.
- Modelos para tempos discretos podem ser incorporados a estrutura de um Modelo Linear Generalizado (MLG). Consequentemente, a estimativa é facilmente obtida usando um *software* padrão para a estimativa de MLG's.
- A incorporação na estrutura de MLG's permite usar a metodologia também para modelos avançados, por exemplo, ao incluir parâmetros subjetivos específicos nos chamados modelos de fragilidade.

Nakano e Carrasco (2006), utilizando tempos discretos aplicados ao modelo exponencial e ao análogo discreto deste modelo (modelo geométrico) mostraram haver um melhor ajuste dos dados ao utilizar o modelo discreto. O mesmo raciocínio foi seguido para simulações computacionais, observando aspectos como variabilidade dos dados, tamanho da amostra e percentual de censura, constatando-se que nem sempre é aceitável a utilização de um modelo contínuo para a análise de dados discretos.

2.3.1 Discretização de distribuições contínuas

Em diversas aplicações, variáveis originais podem ser contínuas por natureza, mas discretas por observação e, portanto, é razoável e conveniente modelar a situação por uma

¹ Se duas observações distintas têm o mesmo valor, recebendo assim a mesma classificação, são consideradas empatadas

distribuição discreta apropriada gerada a partir dos modelos contínuos subjacentes preservando um ou mais traços importantes da distribuição contínua. [Chakraborty \(2015\)](#), apresenta um levantamento completo dos métodos e construções para gerar análogos discretos de distribuições de probabilidade contínuas. Conforme [Jayakumar e Babu \(2018\)](#), entre as diferentes metodologias de discretização de distribuições contínuas, destacam-se:

- Discretizar a função de distribuição acumulada contínua;
- Discretizar a função densidade de probabilidade contínua;
- Discretizar a função de risco contínua;
- Obter distribuição discreta de tempo de vida a partir da taxa de falha alternativa.

A primeira metodologia citada acima, a única que será detalhada neste trabalho, é apresentada a seguir. Considere X uma variável aleatória contínua que assume valores não negativos. A variável aleatória discreta é dada por $T = \lfloor X \rfloor$, em que $\lfloor X \rfloor$ representa “a parte inteira de X ”, isto é, T denota o maior inteiro menor ou igual a X . Se $F_X(x)$ é a função de distribuição acumulada de X , a distribuição de probabilidade de T , representada por $p(t)$, pode ser escrita como:

$$\begin{aligned} p(t) &= P(T = t) = P(t \leq X < t + 1), \\ &= F_X(t + 1) - F_X(t) \quad t = 0, 1, 2, \dots \end{aligned} \quad (2.13)$$

Nesse contexto, algumas publicações em análise de sobrevivência lidam com a utilização dessa metodologia para determinar o análogo discreto das distribuições contínuas e consequentemente obter novas distribuições de probabilidade discretas ou atuar com os análogos discretos oriundos da referida metodologia em aplicações como: [Nakagawa e Osaki \(1975\)](#) que obtiveram a distribuição Weibull discreta (WD), [Jayakumar e Babu \(2018\)](#) que obtiveram a distribuição Weibull Geométrica discreta, [Vieira \(2018\)](#) que atua com a distribuição Log-logística discreta, [Cardial, Fachini-Gomes e Nakano \(2020\)](#) que trabalham com a distribuição Weibull discreta exponenciada (WDE), [Sarhan \(2017\)](#) que obteve a distribuição banheira de dois parâmetros discreta (DTPBT), [Chakraborty \(2015\)](#) que obteve a distribuição Gumbel discreta e [Brunello e Nakano \(2015\)](#) que atuam com a distribuição WD. A seguir será apresentada a distribuição Weibull discreta que será utilizada posteriormente.

2.3.2 Distribuição Weibull discreta

A distribuição Weibull é o modelo probabilístico mais aceito e utilizado na modelagem de dados de sobrevivência contínuos devido à flexibilidade que o mesmo apresenta. Essa versatilidade se deve aos dois parâmetros que a distribuição apresenta: o parâmetro de escala e o

parâmetro de forma, que proporcionam uma variedade de formas e devido a sua função taxa de falha ser monótona, ou seja, ela é crescente, decrescente ou constante.

Por ter essas vantagens, o caso discreto dessa distribuição vem sendo estudado ao longo dos anos, e há, conforme Vila, Nakano e Saulo (2018), pelo menos três versões conhecidas da distribuição Weibull discreta (WD): (i) distribuição Weibull discreta de tipo I, que discretiza a função de distribuição acumulada contínua (NAKAGAWA; OSAKI, 1975); (ii) distribuição Weibull discreta de tipo II, que discretiza a função de risco contínua (STEIN; DATTERO, 1984); e (iii) distribuição Weibull discreta de tipo III, que obtém a distribuição discreta de tempo de vida a partir de uma taxa de falha alternativa (PADGETT; SPURRIER, 1985).

No presente trabalho será utilizada a versão do tipo I, que foi introduzida por Nakagawa e Osaki (1975). Seja X , uma variável aleatória contínua que segue a distribuição Weibull, denotada por $X \sim W(\gamma, \eta)$, com função distribuição acumulada dada por:

$$F_X(x|\gamma, \eta) = 1 - S_X(x|\gamma, \eta) = \left[1 - \exp \left\{ - \left(\frac{x}{\gamma} \right)^\eta \right\} \right] \mathbb{I}_{\{x \geq 0\}}, \quad (2.14)$$

em que $\eta > 0$ é o parâmetro de forma e $\gamma > 0$ é o parâmetro de escala.

Ao obter a variável aleatória discreta correspondente denotada por $T = \lfloor X \rfloor$, considerando (2.14) em (2.13) e admitindo $q = \exp \left\{ \frac{-1}{\gamma^\eta} \right\}$. Então, T segue uma distribuição Weibull discreta (WD) com parâmetros $\eta > 0$ e $q \in (0, 1)$, denotado por $T \sim WD(q, \eta)$, se sua função de probabilidade é dada por:

$$p(t|q, \eta) = (q^{t^\eta} - q^{(t+1)^\eta}) \mathbb{I}_{\{t=0,1,2,\dots\}}, \quad (2.15)$$

A função distribuição acumulada e a função de sobrevivência da distribuição WD, obtidas por meio da relação (2.2), são expressas, respectivamente, por:

$$F(t|q, \eta) = \left(1 - q^{(\lfloor t \rfloor + 1)^\eta} \right) \mathbb{I}_{\{t \geq 0\}} \quad (2.16)$$

e

$$S(t|q, \eta) = \left(q^{(\lfloor t \rfloor + 1)^\eta} \right) \mathbb{I}_{\{t \geq 0\}}, \quad (2.17)$$

em que $\lfloor t \rfloor$ denota o maior inteiro menor ou igual a t .

A taxa de falha da distribuição WD, encontrada a partir das funções de probabilidade e sobrevivência por meio da relação (2.1), é dada por:

$$h(t|q, \eta) = \left(\frac{q^{t^\eta} - q^{(t+1)^\eta}}{q^{t^\eta}} \right) \mathbb{I}_{\{t=0,1,2,\dots\}}. \quad (2.18)$$

A Figura 1 apresenta gráfico da função de risco da WD para diferentes valores de q e η .

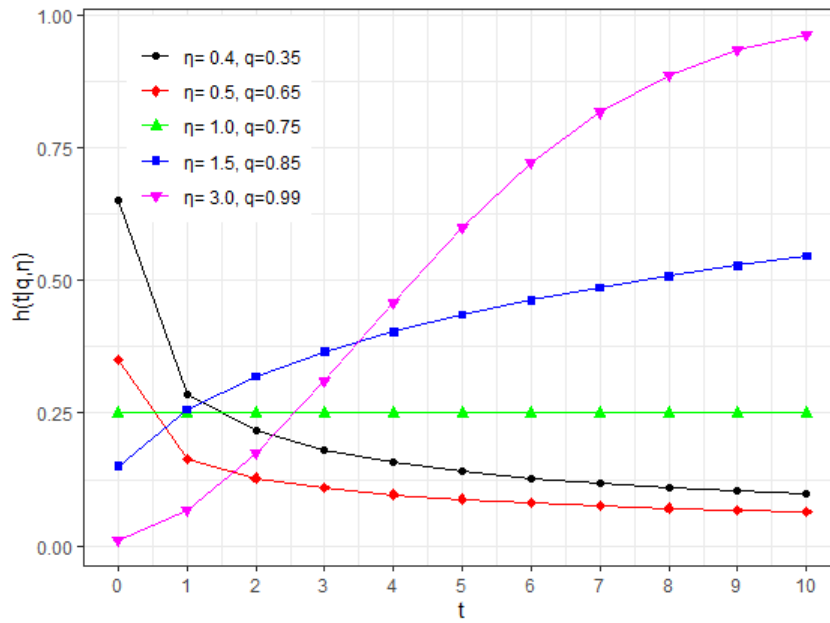


Figura 1 – Taxa de falha da distribuição Weibull Discreta para diferentes valores dos parâmetros.

Fonte: Elaborada pelo autor.

Nota-se, pela Figura 1, que a taxa de falha da distribuição WD apresenta diferentes formas que estão ligadas diretamente ao seu parâmetro de forma η , isto é:

- Quando $\eta > 1$, a taxa de falha é estritamente crescente;
- Quando $\eta = 1$, a taxa de falha é constante (a distribuição WD se reduz a distribuição geométrica);
- Quando $\eta < 1$, a taxa de falha é estritamente decrescente.

MODELO DE CHANCES DE SOBREVIVÊNCIA PROPORCIONAIS

Este capítulo tem como enfoque a construção do modelo de regressão paramétrico para tempos discretos utilizando a estrutura semiparamétrica do Modelo de Chances Proporcionais (MCP) de [Bennett](#), como uma alternativa para tratamento de tempos de sobrevivência discretos que incluam covariáveis. A Seção 3.1 trata-se de uma revisão de literatura que retrata os modelos de regressão para tempos até ocorrência do evento de interesse discretos utilizados atualmente e a caracterização do Modelo de Chances de Sobrevivência Proporcional (MCSP). Por sua vez, nas Seções 3.2 e 3.3 são apresentadas as contribuições do trabalho do ponto de vista teórico em que foram desenvolvidos o MCSP ao considerar tempos de sobrevivência até a ocorrência do evento de interesse discretos e a verificação das premissas necessárias para uso de tal modelo.

3.1 Revisão de Literatura

Na literatura estatística em análise de sobrevivência, existem diferentes abordagens para modelos de regressão, cujo o objetivo primordial é estabelecer uma relação entre a variável resposta e as covariáveis. Desse modo, os modelos mais populares podem ser classificados como modelo de falha acelerado e modelo de riscos proporcionais de [Cox](#).

Segundo [Eleuteri et al. \(2007\)](#), modelo de falha acelerada baseados em famílias de distribuições podem envolver suposições muito estritas sobre os tempos de falha que geralmente simplificam extremamente as evidências experimentais, particularmente no caso de dados clínicos. Por outro lado, modelo de riscos proporcionais em geral não fazem suposições sobre a distribuições de falha, mas suposições sobre como os recursos do sistema influenciam o tempo de sobrevivência (o pressuposto usual quando os dados são contínuos é a proporcionalidade dos riscos).

Tratando-se do procedimento para dados discretos de sobrevivência, em seu artigo, [Cox](#)

(1972, p. 192) sugere a utilização de uma estrutura semiparamétrica para a análise em tempos discretos.

A popularização de tal modelo se deve em parte pelo fato dos usuários não investirem esforços em informar a taxa de falha base, na qual recebe uma atenção menor nesses estudos, como em: Berger e Schmid (2018) e Berger *et al.* (2019). Sendo assim, no modelo de Cox a função risco base é tratada como um parâmetro incômodo de alta dimensão e é altamente errático.

Contudo, conforme Royston e Parmar (2002), o comportamento da função de risco é de potencial interesse médico porque está diretamente relacionado ao curso de uma doença. Para estimar esse risco suavemente, algum tipo de modelo paramétrico pode ser apropriado. Outra questão, surge na validação da sobrevivência do modelo, uma vez que, a função de risco base em um modelo de Cox pode ser considerada como altamente dimensional, o modelo ajustado é muito estreitamente adaptado aos dados em mãos, não ficando claro como validar tais modelos em um conjunto de dados independente. Assim, a especificação probabilística a ser utilizada do conjunto de dados requer uma aproximação paramétrica parcimoniosa para o processo de geração de dados. Desse modo, um modelo que utiliza a estrutura semiparamétrica sugerida por Cox assumindo diversas distribuições base discretas foi proposto por Vieira (2018).

Um concorrente do modelo de Cox é o Modelo de Chances Proporcionais (MCP) proposto por Bennett (1983), que introduziu a estrutura do modelo semiparamétrico em um contexto de análise de sobrevivência utilizando como distribuição base a distribuição Log-logística. Murphy, Rossini e Vaart (1997) demonstraram as propriedades deste modelo, entre elas que o estimador do coeficiente de regressão é assintoticamente normal com variância eficiente. A inferência semiparamétrica foi desenvolvida por Yang e Prentice (1999) e aplicações em tempos de sobrevivência contínuos são apresentados em Royston e Parmar (2002), Wang e Wang (2020) e Zhou, Zhang e Lu (2022).

Nesse modelo, as covariáveis têm um efeito multiplicativo na chance (*odds*) de sobrevivência. Dessa maneira, esse será denominado por Modelo de Chances de Sobrevivência Proporcionais (MCSP). Ao longo dos anos esse modelo que tem em sua proposta inicial a sua utilização como modelo de regressão, é utilizado em uma escala bem menor se comparado ao modelo de riscos proporcionais de Cox.

Na presença de um vetor de covariáveis $\mathbf{Z}' = (Z_1, \dots, Z_p)$, assume-se que o tempo até a ocorrência do evento de interesse, segue o MCSP dado por:

$$\frac{S(t|\mathbf{Z})}{1 - S(t|\mathbf{Z})} = g(\mathbf{Z}'\boldsymbol{\beta}) \frac{S_0(t)}{1 - S_0(t)}, \quad (3.1)$$

em que $S(\cdot)$ é a função de sobrevivência do tratamento teste, $S_0(\cdot)$ é a função de sobrevivência base e $g(\cdot)$ é uma função de ligação não negativa que assume valor 1 quando seu argumento é nulo.

Ao assumir que $\boldsymbol{\beta}' = (\beta_1, \dots, \beta_p)$ é o vetor de coeficientes associados ao vetor de covariáveis e $g(\mathbf{Z}'\boldsymbol{\beta}) = \exp\{\mathbf{Z}'\boldsymbol{\beta}\}$ (função de ligação logarítmica), a chance de sobrevivência do indivíduo que recebe o tratamento padrão ou o novo tratamento é expressa por:

$$\frac{S(t; \boldsymbol{\beta} | \mathbf{Z})}{1 - S(t; \boldsymbol{\beta} | \mathbf{Z})} = \exp\{\mathbf{Z}'\boldsymbol{\beta}\} \frac{S_0(t)}{1 - S_0(t)} \quad (3.2)$$

$$\log \left[\frac{S(t; \boldsymbol{\beta} | \mathbf{Z})}{1 - S(t; \boldsymbol{\beta} | \mathbf{Z})} \right] = G(t) + \mathbf{Z}'\boldsymbol{\beta} \quad (3.3)$$

em que $G(t) = \log \left[\frac{S_0(t)}{1 - S_0(t)} \right]$ é o logito de $S_0(t)$.

Note que, o intercepto β_0 não aparece no preditor linear pois a função de sobrevivência base, $S_0(t)$, absorve este termo constante. Segundo, [Yang e Prentice \(1999\)](#), o MCSP pode ser reformulado como um modelo de transformação linear (MTL)¹:

$$G(T) = \mathbf{Z}'\boldsymbol{\beta} + \varepsilon, \quad (3.4)$$

em que $G(\cdot)$ é uma transformação desconhecida e é dada como em (3.3), e ε tem uma distribuição logística padrão (Apêndice A). Nesta formulação, o MCSP pode ser visto como um modelo de riscos proporcionais com heterogeneidade não observável.

Note que, a partir da expressão (3.1), é possível estabelecer a função de sobrevivência do MCSP (Apêndice B.1):

$$S(t|z) = \frac{g(\mathbf{z}'\boldsymbol{\beta})S_0(t)}{1 + (g(\mathbf{z}'\boldsymbol{\beta}) - 1)S_0(t)}. \quad (3.5)$$

Conforme demonstrado por [Teixeira \(2017, p. 35\)](#), a razão das funções de sobrevivência (tratamento teste e base) converge de 1 para $g(\mathbf{z}'\boldsymbol{\beta})$ com o tempo.

[Marshall e Olkin \(1997\)](#), propuseram um método interessante de adicionar um novo parâmetro a uma distribuição base produzindo a chamada família Marshall–Olkin (MO) de distribuições. Essa família inclui a distribuição base como um exemplo básico e oferece mais flexibilidade a algumas distribuições para modelar vários tipos de dados. A família de distribuições MO também é conhecida como família de chances proporcionais ou família com parâmetro de inclinação. ([CORDEIRO; LEMONTE; ORTEGA, 2014](#))

Desse modo, tratando $g(\mathbf{z}'\boldsymbol{\beta})$ como uma constante θ em (3.5), essa expressão representa a função de sobrevivência da família de distribuições MO. Portanto, a família MO está intimamente relacionada ao MCSP na análise de sobrevivência ([KRISHNA *et al.*, 2013](#)).

Em análise de sobrevivência, existem as relações entre função de risco, sobrevivência, risco acumulado e densidade de probabilidade com enfoque em tempos contínuos de sobrevivência. Mais detalhes podem ser encontrados em [Colosimo e Giolo \(2006\)](#). Assim, é possível

¹ MTL é um tipo particular de função entre dois espaços vetoriais que preserva as operações de adição vetorial e multiplicação por escalar.

determinar a função de risco para o MCSP (Apêndice B.2):

$$h_c(t|\mathbf{z}) = \left[\frac{h_0(t)}{1 + (g(\mathbf{z}'\boldsymbol{\beta}) - 1)S_0(t)} \right] \mathbb{I}_{\{t \geq 0\}}, \quad (3.6)$$

em que $h_0(t)$ é a função de risco base.

Bennett traça um paralelo do MCSP ao modelo de riscos proporcionais, comparando a razão das funções de risco para o caso contínuo. Ele afirma que no modelo de **Cox** as taxas de risco para grupos separados de pacientes têm uma relação constante entre si. Por outro lado, no MCSP estas taxas de risco convergem com o tempo.

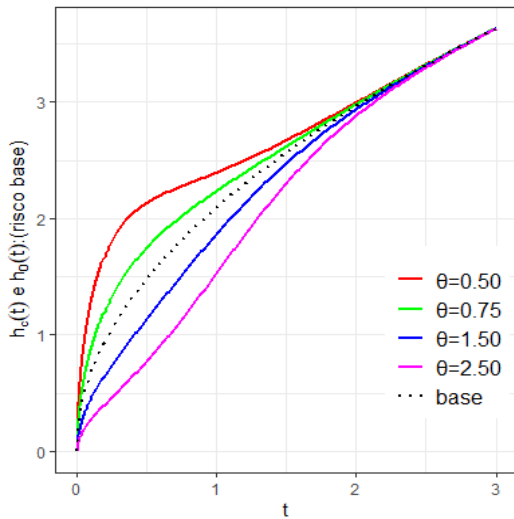
Para ilustração, seja $\theta = g(\mathbf{z}'\boldsymbol{\beta})$ representando a razão de chances de sobrevivência constantes ao longo do tempo em (3.6), e que também caracteriza a função de risco da família de distribuições MO. Foram consideradas distribuições base Weibull contínua, cujas taxas de falha são:

- (i) crescente: $W(\gamma = 0,8; \eta = 1,5)$;
- (ii) constante: $W(\gamma = 0,75; \eta = 1,0)$;
- (iii) decrescente: $W(\gamma = 0,75; \eta = 0,5)$.

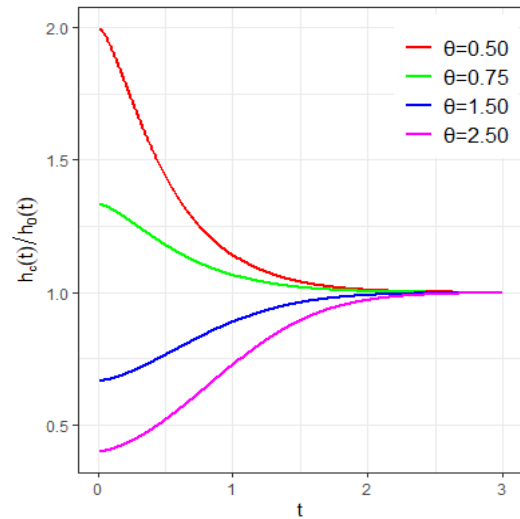
Na Figura 2, são apresentadas as taxas de falha do MCSP, (3.6), com diferentes valores para θ e distribuição base que possui risco: (a) crescente, (c) constante e (e) decrescente. E as respectivas razões entre as funções de risco do MCSP contínuo e base ((b) i, (d) ii e (f) iii).

Por meio da Figura 2 (a, c, e), verifica-se que os riscos (taxa de falha base e do MCSP assumindo diversos valores para θ) não são proporcionais, mas sim convergentes, ou seja, conforme o tempo aumenta, independente do valor de θ e do risco base ser crescente, constante ou decrescente, ele converge para um único ponto. No que se refere à razão entre os respectivos riscos do MCSP e risco base, Figura 2 (b, d, f), observa-se que quando $t \rightarrow 0$, a razão das funções de risco converge monotonicamente para θ^{-1} , e quando $t \rightarrow \infty$ converge para 1. Portanto, para todos os casos em estudo, o efeito inicial do novo tratamento sobre o risco desaparece ao longo do tempo.

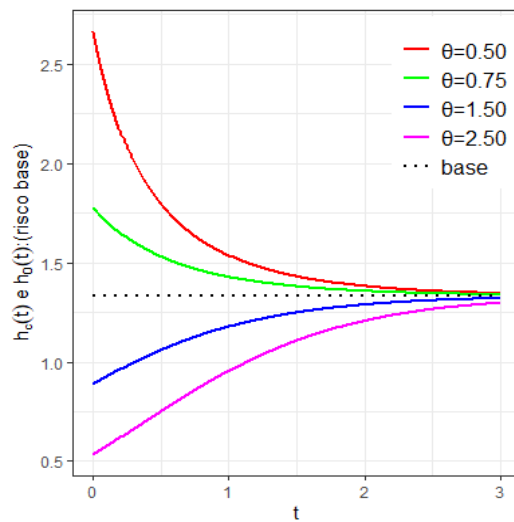
Murphy, Rossini e Vaart (1997) afirmam que, **Bennett** sugeriu usar o MCSP para demonstrar uma cura eficaz. Neste caso, a taxa de mortalidade de um grupo de pessoas doentes se aproximaria da mortalidade de um grupo de controle conforme o tempo passa. Pode-se tomar como exemplo, a taxa de morbidade para um grupo interrompendo um hábito, como fumar, pode ser comparado com a taxa de morbidade de pessoas semelhantes que nunca fumaram; este modelo pode ser usado para demonstrar que as taxas de morbidade convergem com o tempo.



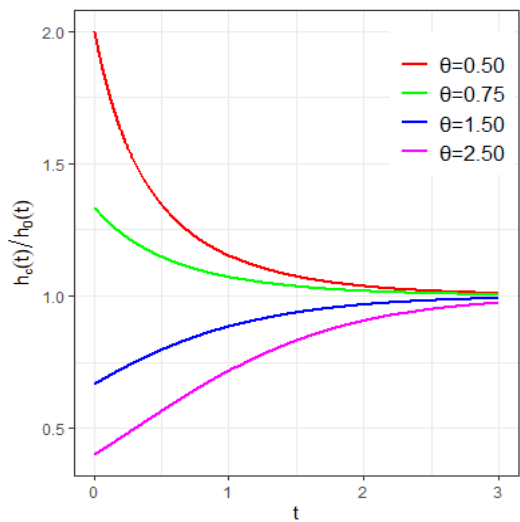
(a) Riscos do MCSP (risco base crescente).



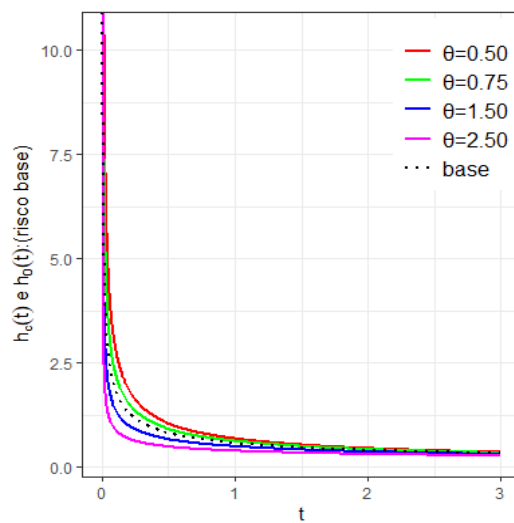
(b) Razão entre os riscos (risco base crescente).



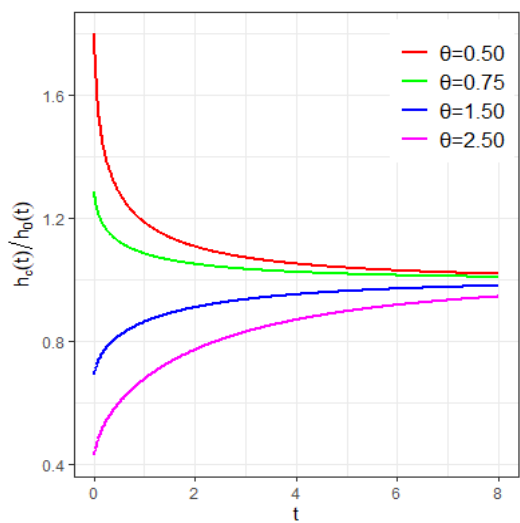
(c) Riscos do MCSP (risco base constante).



(d) Razões entre os riscos (risco base constante).



(e) Riscos do MCSP (risco base decrescente).



(f) Razões entre os riscos (risco base decrescente).

Figura 2 – Taxas de falhas e respectivas razões entre os riscos para diversas distribuições base e valores da razão de chances de sobrevivência θ (Caso contínuo).

Fonte: Elaborada pelo autor.

3.2 MCSP para tempos discretos

A função de probabilidade do MCSP para o caso discreto é dada por (Apêndice B.3):

$$p(t|\mathbf{z}) = \begin{cases} \frac{1 - S_0(0)}{1 + (g(\mathbf{z}'\boldsymbol{\beta}) - 1)S_0(0)}, & \text{se } t = 0 \\ g(\mathbf{z}'\boldsymbol{\beta}) \left(\frac{S_0(t-1)}{1 + (g(\mathbf{z}'\boldsymbol{\beta}) - 1)S_0(t-1)} - \frac{S_0(t)}{1 + (g(\mathbf{z}'\boldsymbol{\beta}) - 1)S_0(t)} \right), & \text{se } t = 1, 2, \dots \end{cases} \quad (3.7)$$

Por sua vez, a função de risco, é expressa por (Apêndice B.4):

$$h(t|\mathbf{z}) = \begin{cases} \frac{1 - S_0(0)}{1 + (g(\mathbf{z}'\boldsymbol{\beta}) - 1)S_0(0)}, & \text{se } t = 0 \\ 1 - \frac{S_0(t)(1 + (g(\mathbf{z}'\boldsymbol{\beta}) - 1)S_0(t-1))}{S_0(t-1)(1 + (g(\mathbf{z}'\boldsymbol{\beta}) - 1)S_0(t))}, & \text{se } t = 1, 2, \dots, \end{cases} \quad (3.8)$$

em que $S_0(\cdot)$ é a função de sobrevivência base e $g(\cdot)$ é uma função de ligação não negativa que assume o valor 1 quando o seu argumento é nulo.

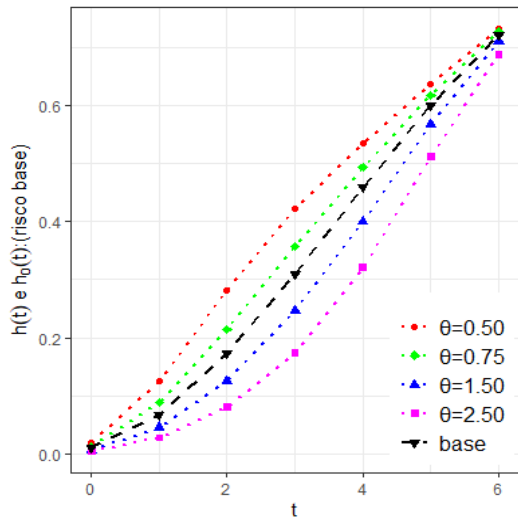
Para verificar o comportamento da função de risco do MCSP no caso discreto e também a razão entre esta e a função de risco base, a fim de fazer um paralelo em relação ao caso contínuo, seja $\theta = g(\mathbf{z}'\boldsymbol{\beta})$ representando a razão de chances de sobrevivência constantes ao longo do tempo em (3.8). Foram consideradas distribuições base Weibull discreta (WD), cujas taxas de falha são:

- (i) crescente: $WD(q = 0,99; \eta = 3,0)$;
- (ii) constante: $WD(q = 0,5; \eta = 1)$;
- (iii) decrescente: $WD(q = 0,5; \eta = 0,7)$.

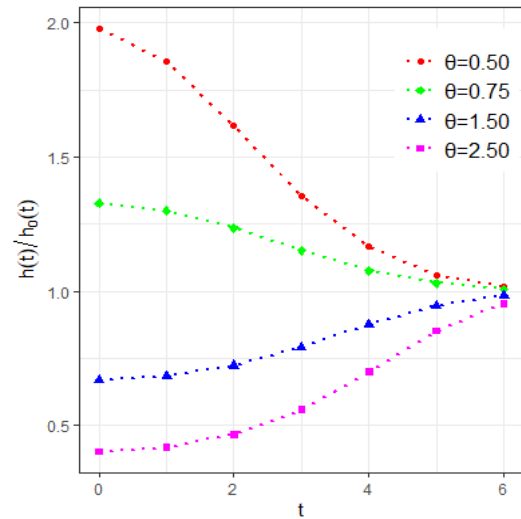
Na Figura 3, são apresentadas as taxas de falha do MCSP (caso discreto), (3.8), com diferentes valores para θ e distribuição base que possui risco: (a) crescente, (c) constante e (e) decrescente. E as respectivas razões entre as funções de risco do MCSP discreto e base ((b) i, (d) ii e (f) iii).

Observa-se, pela Figura 3, que a função de risco do MCSP para o caso discreto também possui a propriedade de riscos convergentes independente da forma função de risco base (a, c, e), isto é, convergem para o mesmo ponto conforme o tempo aumenta. E pela Figura 3 (b, d, f), nota-se que, quando $t \rightarrow \infty$, a razão entre os diversos riscos do MCSP e o risco base, classificados como: crescente, decrescente e constante, converge para 1.

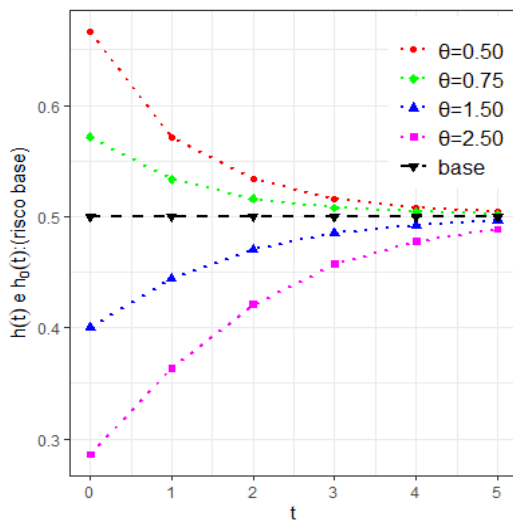
Portanto, para o caso discreto, preserva-se a ideia de cura eficaz inicialmente defendida por Bennett, para o MCSP para dados contínuos de sobrevivência. A fim de permitir a inclusão de covariáveis no MCSP definindo uma relação entre as covariáveis e os tempos de sobrevivência discretos, a seguir é apresentado o Modelo de Regressão Chances de Sobrevivência Proporcional Weibull Discreto (MRCSP-WD).



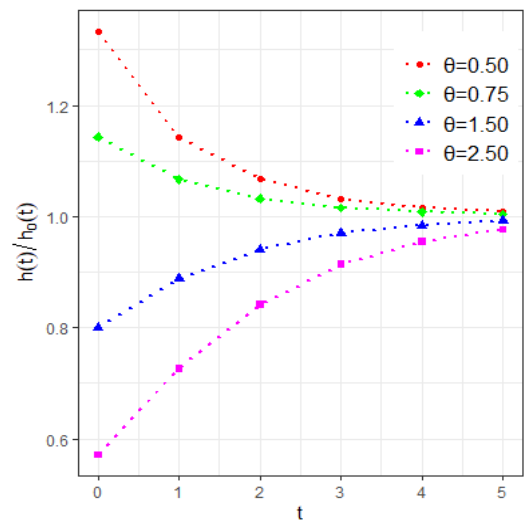
(a) Riscos do MCSP (risco base crescente).



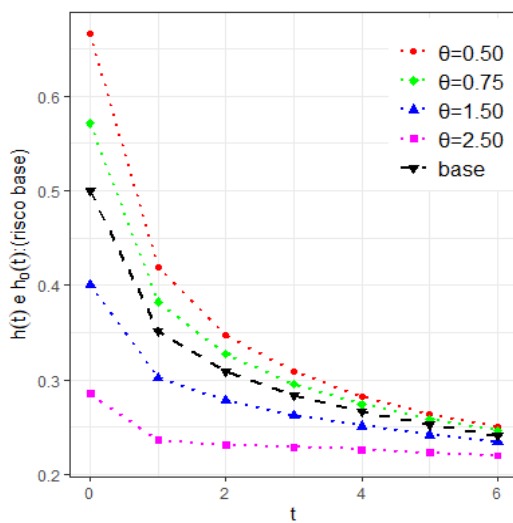
(b) Razão entre os riscos (risco base crescente).



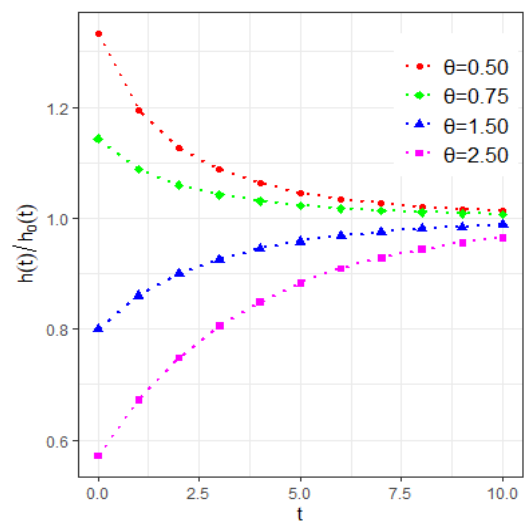
(c) Riscos do MCSP (risco base constante).



(d) Razões entre os riscos (risco base constante).



(e) Riscos do MCSP (risco base decrescente).



(f) Razões entre os riscos (risco base decrescente).

Figura 3 – Taxas de falhas e respectivas razões entre os riscos para diversas distribuições base e valores da razão de chances de sobrevivência θ (Caso discreto).

Fonte: Elaborada pelo autor.

3.2.1 Modelo de Regressão Chances de Sobrevivência Proporcionalis Weibull Discreto

Uma vez definida as funções em análise de sobrevivência que caracterizam o MCSP para o caso discreto, pode-se incluir covariáveis ao modelo atribuindo-se uma função a $g(\mathbf{z}'\boldsymbol{\beta})$, conforme apresentado em (3.2).

Para a versão paramétrica, outro requisito para obter o modelo de regressão desejado é a suposição de uma distribuição base, por se tratar de tempos de sobrevivência discretos, é necessário uma distribuição de probabilidade que modele de maneira adequada esses tempos. Nesse sentido a distribuição Weibull discreta (WD), que exhibe grandes vantagens conforme apresentado na Subseção 2.3.2, será utilizada.

Assim, a função de probabilidade e sobrevivência do Modelo de Regressão Chances de Sobrevivência Proporcionalis Weibull Discreto (MRCSP-WD) são dadas, respectivamente, por:

$$p(t; q, \eta, \boldsymbol{\beta} | \mathbf{z}) = \left[\frac{1 - q}{1 + (\exp \{ \mathbf{z}' \boldsymbol{\beta} \} - 1) q} \right]^{\mathbb{I}_{\{t=0\}}} \quad (3.9)$$

$$\times \left[\frac{\exp \{ \mathbf{z}' \boldsymbol{\beta} \} q^{t^\eta}}{1 + (\exp \{ \mathbf{z}' \boldsymbol{\beta} \} - 1) q^{t^\eta}} - \frac{\exp \{ \mathbf{z}' \boldsymbol{\beta} \} q^{(t+1)^\eta}}{1 + (\exp \{ \mathbf{z}' \boldsymbol{\beta} \} - 1) q^{(t+1)^\eta}} \right]^{\mathbb{I}_{\{t=1,2,\dots\}}}$$

e

$$S(t; q, \eta, \boldsymbol{\beta} | \mathbf{z}) = \left[\frac{\exp \{ \mathbf{z}' \boldsymbol{\beta} \} q^{(\lfloor t \rfloor + 1)^\eta}}{1 + (\exp \{ \mathbf{z}' \boldsymbol{\beta} \} - 1) q^{(\lfloor t \rfloor + 1)^\eta}} \right]^{\mathbb{I}_{\{t \geq 0\}}}, \quad (3.10)$$

em que $\lfloor t \rfloor$ é o maior inteiro menor ou igual a t .

Obtidas as funções descritas acima, o próximo passo consiste na estimação dos parâmetros do MRCSP-WD. Para essa tarefa será utilizado o método de máxima verossimilhança, conforme descrito na Subseção 2.2.1.

Seja uma amostra aleatória observada t_1, t_2, \dots, t_n de uma população com função de probabilidade e função de sobrevivência do MRCSP-WD definidas na equações (3.9) e (3.10), com δ_i denotando a variável indicadora de censura ou falha do indivíduo i , no qual é igual a 0 se o tempo t_i for censurado à direita e 1 se for de falha, e $\mathbf{z}'_i = (z_{i1}, z_{i2}, \dots, z_{ip})$ o vetor de covariáveis do indivíduo i , $i = 1, 2, \dots, n$. A função de verossimilhança do MRCSP-WD é dada por:

$$L(q, \eta, \boldsymbol{\beta}; \mathbf{t}, \boldsymbol{\delta}, \mathbf{z}) \propto \prod_{i=1}^n \left[\frac{\exp \{ \mathbf{z}'_i \boldsymbol{\beta} \} q^{t_i^\eta}}{1 + (\exp \{ \mathbf{z}'_i \boldsymbol{\beta} \} - 1) q^{t_i^\eta}} - \frac{\exp \{ \mathbf{z}'_i \boldsymbol{\beta} \} q^{(t_i+1)^\eta}}{1 + (\exp \{ \mathbf{z}'_i \boldsymbol{\beta} \} - 1) q^{(t_i+1)^\eta}} \right]^{(1 - \mathbb{I}_{\{t_i=0\}}) \delta_i}$$

$$\times \left[\frac{1 - q}{1 + (\exp \{ \mathbf{z}'_i \boldsymbol{\beta} \} - 1) q} \right]^{\mathbb{I}_{\{t_i=0\}} \delta_i} \left[\frac{\exp \{ \mathbf{z}'_i \boldsymbol{\beta} \} q^{(t_i+1)^\eta}}{1 + (\exp \{ \mathbf{z}'_i \boldsymbol{\beta} \} - 1) q^{(t_i+1)^\eta}} \right]^{(1 - \delta_i)} \quad (3.11)$$

Aplicando o logaritmo na função de verossimilhança (3.11), tem-se:

$$\begin{aligned} \ell(q, \eta, \boldsymbol{\beta}; \mathbf{t}, \boldsymbol{\delta}, \mathbf{z}) &= \sum_{i=1}^n \left\{ (1 - \mathbb{I}_{\{t_i=0\}}) \delta_i \log \left[\frac{\exp\{\mathbf{z}'_i \boldsymbol{\beta}\} q^{t_i^\eta}}{1 + (\exp\{\mathbf{z}'_i \boldsymbol{\beta}\} - 1) q^{t_i^\eta}} - \frac{\exp\{\mathbf{z}'_i \boldsymbol{\beta}\} q^{(t_i+1)^\eta}}{1 + (\exp\{\mathbf{z}'_i \boldsymbol{\beta}\} - 1) q^{(t_i+1)^\eta}} \right] \right\} + \\ &+ \sum_{i=1}^n \left\{ \mathbb{I}_{\{t_i=0\}} \delta_i \log \left[\frac{1 - q}{1 + (\exp\{\mathbf{z}'_i \boldsymbol{\beta}\} - 1) q} \right] \right\} \\ &+ \sum_{i=1}^n \left\{ (1 - \delta_i) \log \left[\frac{\exp\{\mathbf{z}'_i \boldsymbol{\beta}\} q^{(t_i+1)^\eta}}{1 + (\exp\{\mathbf{z}'_i \boldsymbol{\beta}\} - 1) q^{(t_i+1)^\eta}} \right] \right\} + c, \end{aligned} \quad (3.12)$$

em que c é uma constante que não depende de q , η e $\boldsymbol{\beta}$.

Seja $\boldsymbol{\vartheta} = (q, \eta, \boldsymbol{\beta})$, a primeira derivada do logaritmo da função de verossimilhança, é chamada de função escore, e igualando a zero obtém-se a equação de verossimilhança representada por:

$$\mathbf{U}(\boldsymbol{\vartheta}) = \frac{\partial \ell(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} = 0. \quad (3.13)$$

Assumindo \mathbf{Z} fixado para os procedimentos inferenciais desenvolvidos, o valor de $\widehat{\boldsymbol{\vartheta}} = (\widehat{q}, \widehat{\eta}, \widehat{\boldsymbol{\beta}})$, que satisfaz a equação (3.13), é o estimador de máxima verossimilhança do MRCSP-WD, que tem matriz de variância e covariância dada por:

$$\boldsymbol{\Sigma}(\widehat{\boldsymbol{\vartheta}}) = \left[- \frac{\partial^2 \ell(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^T} \Big|_{\boldsymbol{\vartheta}=\widehat{\boldsymbol{\vartheta}}} \right]^{-1} = \left[-J(\boldsymbol{\vartheta}) \Big|_{\boldsymbol{\vartheta}=\widehat{\boldsymbol{\vartheta}}} \right]^{-1} \quad (3.14)$$

Os elementos da matriz observada $J(\boldsymbol{\vartheta})$ e do estimador de máxima verossimilhança $\widehat{\boldsymbol{\vartheta}}$ são obtidos numericamente, por meio de métodos computacionais de otimização com o auxílio da função *optim()* pelo *software* R (R Core Team, 2019) utilizando o método *Nelder-Mead*.

Dessa maneira, a respeito dos intervalos de confiança para os parâmetros. Para $\boldsymbol{\beta}$ o intervalo é obtido como descrito em (2.10). No que se refere aos parâmetros da distribuição WD, estes têm limitação no espaço paramétrico, ou seja, $\eta > 0$ e $0 < q < 1$. Portanto, é necessário fazer uma transformação para deixá-los irrestritos, e então construir intervalos de confiança em que os limites são definidos considerando seus respectivos espaços paramétricos. Assim, para o parâmetro q foi considerada a transformação log-log e para o parâmetro η , a transformação logarítmica (Apêndice D).

Desta forma, os intervalos de confiança aproximados de $(1 - \alpha) \times 100\%$ de confiança para os parâmetros q e η , são dados respectivamente por:

$$\left[(\widehat{q}) e^{Z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}(\widehat{u})}}; (\widehat{q}) e^{-Z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}(\widehat{u})}} \right], \quad (3.15)$$

em que $\widehat{u} = \log[-\log \widehat{q}]$, e

$$\left[\widehat{\eta} e^{-Z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}(\widehat{v})}}; \widehat{\eta} e^{Z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}(\widehat{v})}} \right], \quad (3.16)$$

em que $\widehat{v} = \log \widehat{\eta}$ e $Z_{1-\alpha/2}$ é o quantil $(1 - \alpha/2)$ de uma distribuição normal padrão. Como definido na Subseção 2.2.2 os valores de $\widehat{\text{Var}}(\widehat{u})$ e $\widehat{\text{Var}}(\widehat{v})$ também são obtidos numericamente, e são calculados via método delta.

3.3 Verificação da Suposição de chances de sobrevivência proporcionais

Com o objetivo de investigar a suposição de chances sobrevivências proporcionais, na presente seção foram desenvolvidas ferramentas para este fim.

O modelo proposto em (3.1) pressupõe que as chances de sobrevivência para dois indivíduos são proporcionais. Considerando uma covariável z dicotômica que assume os valores 0 e 1, o modelo supõe que:

$$\frac{S(t|z=1)}{1-S(t|z=1)} = \theta \frac{S(t|z=0)}{1-S(t|z=0)}, \quad (3.17)$$

em que $S(\cdot)$ é a função de sobrevivência e θ é a constante de proporcionalidade que não depende de t .

Seja $o_i(t)$, a função chances (*odds*) de sobrevivência de um indivíduo com covariável $z = i$; $i = 0, 1$, expressa por:

$$o_i(t) = \frac{S(t|z=i)}{1-S(t|z=i)}, \quad i = 0, 1. \quad (3.18)$$

A função $O_i(\cdot)$ é, por sua vez, a função chances de sobrevivência acumulada dada por:

$$O_i(t) = \sum_{u=0}^t o_i(u) = \sum_{u=0}^t \frac{S(u|z=i)}{1-S(u|z=i)}, \quad i = 0, 1. \quad (3.19)$$

Note que, sob a suposição de chances de sobrevivência proporcionais, a partir das expressões (3.17) e (3.19), tem-se que:

$$O_1(t) = \theta O_0(t). \quad (3.20)$$

Ao aplicar o logaritmo em ambos os lados da igualdade em (3.20), tem-se:

$$\log(O_1(t)) = \log(\theta) + \log(O_0(t)). \quad (3.21)$$

Portanto, a relação entre $\log(O_1(t))$ e $\log(O_0(t))$ é uma reta com coeficiente angular, m_1 , igual a 1 e o coeficiente linear $m_0 = \log(\theta)$, ou seja:

$$\log(O_1(t)) = m_0 + m_1 \log(O_0(t)). \quad (3.22)$$

Assim, a suposição de chances de sobrevivência proporcionais pode ser verificada graficamente, ajustando-se uma reta de regressão simples com coeficiente angular, m_1 , igual a 1 (fixo).

É importante destacar que tal procedimento pode ser utilizado para covariáveis categóricas com três ou mais níveis, sendo necessário, nesse caso, comparar cada nível da covariável

duas-a-duas. Outra circunstância em que é possível a utilização do método é quando há covariáveis numéricas. Nesse cenário é preciso categorizar os valores das covariáveis e compará-las duas-a-duas.

Para a análise gráfica de verificação de chances de sobrevivência proporcionais, além do método discutido acima, outros dois gráficos (que podem ser esboçados utilizando as funções apresentadas nessa seção) podem auxiliar na verificação, tendo papéis coadjuvantes em tal avaliação. Na Tabela 1, estão especificados os dois gráficos auxiliares junto ao principal, com seus respectivos comportamentos para que a suposição seja atendida.

Tabela 1 – Análise gráfica da validade de chances de sobrevivência proporcionais.

Gráfico (eixo das abcissas, eixo das ordenadas)	Comportamento para que a suposição verificada seja verdadeira
$(t, o_i(t)), \quad i = 0, 1$	curvas proporcionais
$(t, \log o_i(t)), \quad i = 0, 1$	curvas paralelas
$(\log(O_0(t)), \log(O_1(t)))$	os pontos formados pelas coordenadas devem estar próximos a reta de regressão com coeficiente angular igual a um

Fonte: Elaborada pelo autor.

Para ilustração, serão apresentados e discutidos alguns exemplos, em que são consideradas variáveis categóricas com dois níveis.

O primeiro exemplo contempla uma situação em que a chances de sobrevivência proporcionais é válida. Neste caso, Figura 4, tem-se que a sobrevivência de um indivíduo com covariável $z = 0$ comporta-se segundo uma distribuição $WD(0,75; 0,5)$, e que a chance de sobrevivência de um indivíduo com covariável $z = 1$, é 1,5 vezes a chance de sobrevivência do indivíduo $z = 0$.

Note, pela Figura 4a, que as sobrevivências (base e do MCSP) não são proporcionais, pelo fato de as curvas não serem proporcionais ao longo do tempo. Por outro lado, na Figura 4b, verifica-se que as chances de sobrevivência são proporcionais por apresentar comportamento contrário ao da figura anterior. Portanto, as chances de sobrevivência proporcionais não implicam sobrevivências proporcionais (este mesmo comportamento é observado no próximo exemplo).

O gráfico $t \times \log(o_i(t))$, em que as curvas são paralelas, ratifica a informação de que as curvas de chances de sobrevivência ($t \times o_i(t)$) são de fato proporcionais. Por fim, o gráfico $\log(O_0(t)) \times \log(O_1(t))$ esboça os pontos que estão muito próximos da reta de regressão ajustada, cujo o coeficiente linear estimado é $\widehat{m}_0 = \log 1,5 = 0,4054$ (quadrado destacado na figura).

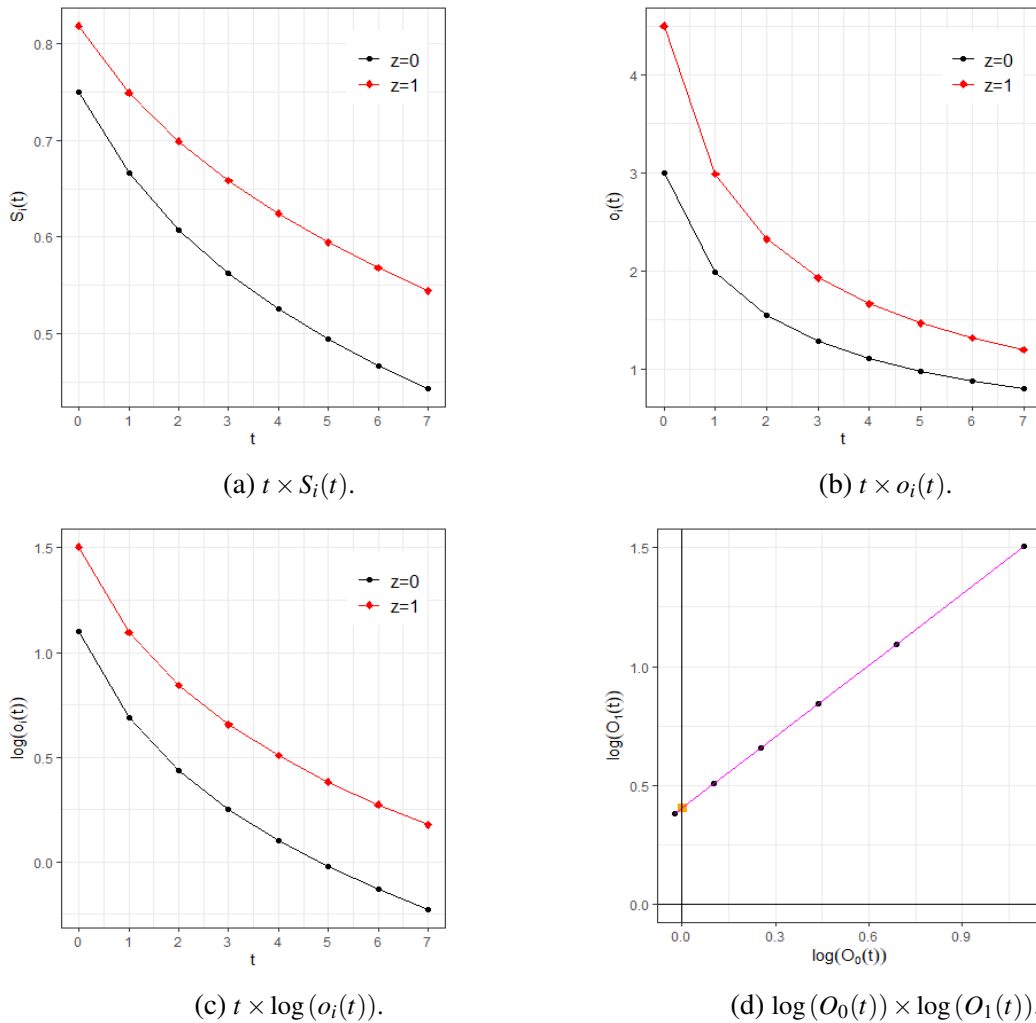


Figura 4 – Exemplo 1 (chances de sobrevivência proporcionais).

Fonte: Elaborada pelo autor.

Tal como o Exemplo 1, o segundo exemplo ilustra uma situação em que a chances de sobrevivência são proporcionais. Deste modo, ambos os exemplos têm comportamentos similares. No Exemplo 2, Figura 5, considerou-se $S(t|z=0)$ como sendo a função de sobrevivência de uma distribuição (base) $WD(0,8;1,5)$ e $o_1(t) = 2 \times o_0(t)$.

A Figura 5a, indica a inexistência de sobrevivência proporcionais. Destaca-se que, para uma coerente avaliação dos gráficos auxiliares ($t \times o_i(t)$ e $t \times \log(o_i(t))$), é essencial que a análise de ambos ocorram em conjunto. Note que a análise da Figura 5b isoladamente pode gerar interpretações confusas, uma vez que alguns desses pontos são aproximadamente iguais. Em complemento a este primeiro gráfico, pela Figura 5c, verifica-se de fato que as chances de sobrevivência são proporcionais, pois as curvas são claramente paralelas. Finalmente, o gráfico $\log(O_0(t)) \times \log(O_1(t))$, atesta a existência de chances de sobrevivência proporcionais, devido ao fato dos pontos formados pelas coordenadas estarem bastante próximos reta de regressão ajustada, cujo o coeficiente linear estimado é $\widehat{m}_0 = \log 2 = 0,6931$ (quadrado destacado na figura).

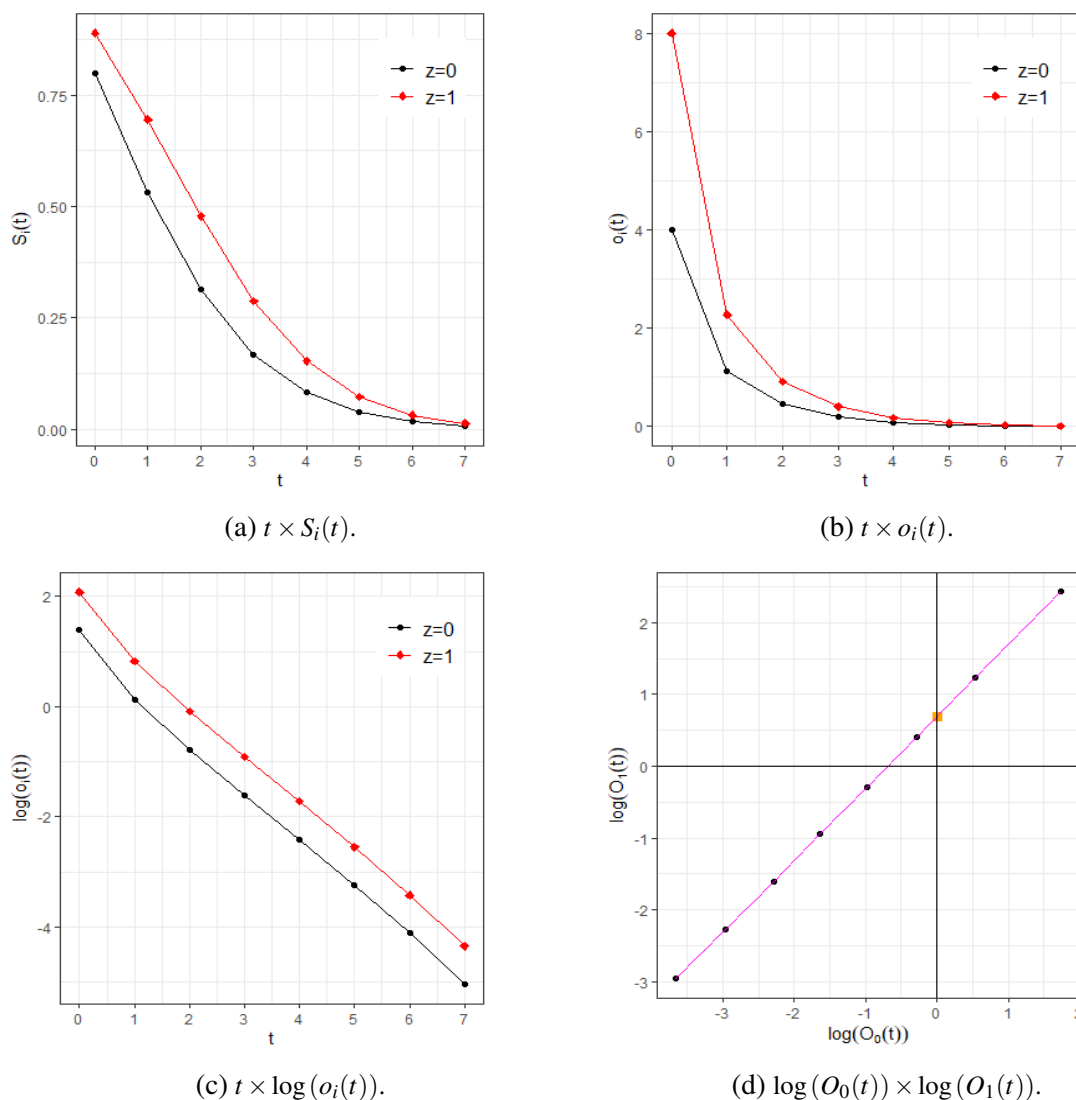


Figura 5 – Exemplo 2 (chances de sobrevivência proporcionais).

Fonte: Elaborada pelo autor.

Ao contrário dos casos anteriores, o Exemplo 3 ilustra uma situação em que a suposição de chances de sobrevivência proporcionais não é atendida. Para esse caso, foi considerado que a sobrevivência de um indivíduo com covariável $z = 0$ comporta-se de acordo uma distribuição $WD(0,35;0,4)$ e a sobrevivência de um indivíduo $z = 1$ tem comportamento segundo uma distribuição $WD(0,99;3,0)$ (Figura 6).

As chances de sobrevivência não são proporcionais, uma vez que no gráfico $t \times o_i(t)$ as curvas não são proporcionais e na Figura 6b as curvas não são paralelas. Além disso, no gráfico $\log(O_0(t)) \times \log(O_1(t))$, os pontos formados não passam próximos a reta de regressão ajustada, cujo o coeficiente linear estimado é $\widehat{m}_0 = 1,247$ (quadrado destacado na figura). Mais um aspecto diferente dos demais exemplos que também é observado, é que as curvas de sobrevivência se cruzam (Figura 6a).

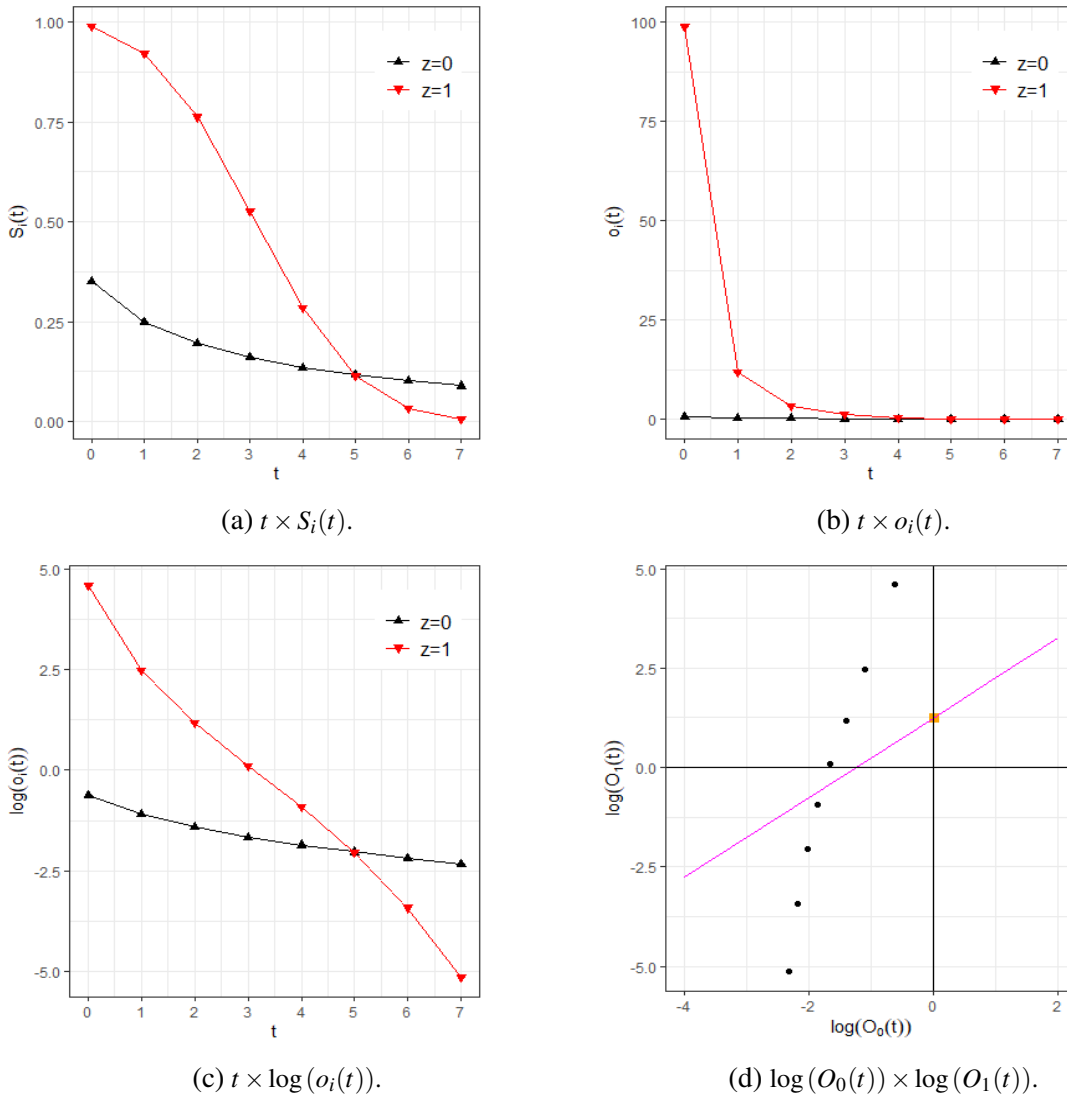


Figura 6 – Exemplo 3 (chances de sobrevivência não proporcionais).

Fonte: Elaborada pelo autor.

A análise gráfica é bastante informativa, no entanto, para que determinada avaliação para tomada de decisão seja completa, é aconselhável que haja uma medida de evidência. Assim, ao considerar a expressão (3.22), pode-se utilizar um teste de hipótese, a fim de verificar se as chances de sobrevivência apresentam valores proporcionais entre si. Deste modo, seja $t_{(j)}$, com $j = 1, 2, \dots, J$, o j -ésimo tempo distinto observado (censurado ou não censurado). A verificação pode ser conduzida, ao testar a hipótese do coeficiente angular da reta ser diferente de um ($m_1 \neq 1$). Assim, as hipóteses de interesse são descritas, em termos estatísticos, por:

$$H_0 : m_1 = 1 \quad \text{vs.} \quad H_1 : m_1 \neq 1. \tag{3.23}$$

É importante destacar que ao ocorrer a situação em que $m_1 = 1$ e $m_0 = 0$, e o referido teste de hipótese for realizado nessas condições e a hipótese nula não for rejeitada, essa é uma circunstância em que a covariável em estudo do MCSP não é significativa.

SIMULAÇÕES COMPUTACIONAIS

Devido à complexidade do Modelo de Regressão Chances de Sobrevivência Proporcional Weibull Discreto (MRCSP-WD), as condições de regularidade não são fáceis de verificar analiticamente. Neste caso, estudos de simulação são necessários; veja, por exemplo, [Ha e MacKenzie \(2010\)](#), [Ortega *et al.* \(2015\)](#) e [Barriga *et al.* \(2019\)](#). Seguindo essa ideia, no presente capítulo, é descrito um estudo de simulação realizado para investigar se as propriedades assintóticas usuais dos estimadores de máxima verossimilhança se mantêm. Também foi avaliado o comportamento do modelo proposto na presença de dados censurados.

O estudo do MRCSP-WD foi conduzido, considerando dados simulados no *software* R ([R Core Team, 2019](#)). Os códigos estão disponíveis no Apêndice C. Foram simulados tempos de sobrevivência do MRCSP-WD pelo método da transformação inversa, conforme o procedimento descrito a seguir. Inicialmente foram gerados tempos de sobrevivência contínuos por meio do Modelo Chances de Sobrevivência Proporcional Weibull (contínuo), ou seja (3.5) com:

$$S_0(t) = \left[\exp \left\{ - \left(\frac{t}{\gamma} \right)^\eta \right\} \right] \mathbb{I}_{\{t \geq 0\}}, \quad (4.1)$$

em que $\eta > 0$ é o parâmetro de forma e $\gamma > 0$ é o parâmetro de escala.

Sejam os parâmetros η e q da distribuição Weibull discreta, o parâmetro de escala, γ , da distribuição Weibull contínua é obtida por $\gamma = \left\{ -\frac{1}{\log(q)} \right\}^{1/\eta}$. Ademais, o parâmetro de forma, η da distribuição Weibull contínua é o mesmo da distribuição Weibull discreta. Os tempos contínuos gerados foram então discretizados considerando-se sua parte inteira, isto é, o maior inteiro menor ou igual ao valor gerado. Seguindo o procedimento descrito em [Oliveira \(2021\)](#), a censura foi incorporada independentemente do tempo de sobrevivência através da variável indicadora de censura gerada por uma distribuição Bernoulli, com os percentuais de censura que estão especificados na Seção 4.2. Acrescenta-se ainda que, as covariáveis, que serão descritas na próxima seção, foram incluídas em (3.5), levando em conta uma função de ligação logarítmica, ou seja, $g(\mathbf{z}'\boldsymbol{\beta}) = \exp\{\mathbf{z}'\boldsymbol{\beta}\}$.

4.1 Tempos discretos sem a presença de censura

As amostras dos tempos de sobrevivência foram simuladas, considerando duas covariáveis e diversos parâmetros do MRCSP-WD, tendo em vista os diferentes comportamentos das taxas de falha base conforme a Tabela 2.

Tabela 2 – Cenários das simulações.

Cenário	q	η	β_1	β_2	Taxa de Falha base
C_1	0,9	1,50	2,0	1,0	Crescente
C_2	0,9	1,00	2,0	1,0	Constante
C_3	0,9	0,75	2,0	1,0	Decrescente

Fonte: Elaborada pelo autor.

Os resultados das simulações apresentados neste trabalho consideraram no modelo uma covariável dicotômica, Z_1 , simulada a partir de uma distribuição Bernoulli com probabilidade de sucesso $p = 0,5$ e uma covariável, Z_2 , com distribuição normal padrão para cada cenário descrito.

As médias das estimativas, erro quadrático médio (EQM) e a probabilidade de cobertura (PC) dos estimadores do MRCSP-WD foram calculados a partir de 10.000 réplicas de Monte Carlo, considerando amostras de tamanho $n = 30, 50, 100, 250$ e 500 . Para a construção dos intervalos de confiança para o cálculo da PC foi considerado o grau de confiança de $0,95$.

A apresentação gráfica dos resultados está disposta por parâmetros (os resultados numéricos estão disponíveis na Tabela 16 no Anexo A), a fim de verificar o comportamento destes em relação ao tamanho amostral. Os parâmetros q e η são provenientes da distribuição base WD e este segundo é responsável pelo comportamento da taxa de falha e, por este motivo, são considerados diferentes valores do mesmo frente a esses comportamentos. Ademais, os parâmetros β_1 e β_2 são os coeficientes de regressão associados às covariáveis Z_1 e Z_2 , respectivamente.

Os resultados que se referem ao estimador q são apresentados nas Figuras 7, 8 e 9. Nota-se que as médias das estimativas são aproximadamente iguais ao verdadeiro valor do parâmetro (os valores mais distantes do verdadeiro valor do parâmetro são $0,8992$ e $0,8994$), independente do cenário e tamanho amostral, e que à medida que o tamanho aumenta há uma tendência das estimativas serem cada vez menos enviesadas (Figura 7a). Essa tendência é confirmada, observando-se o EQM, Figura 7b, que se torna cada vez menor ao passo que o tamanho amostral aumenta. Destaca-se ainda que, todas as estimativas de EQM são próximas de zero ($EQM(q) \leq 0,0034$). Além disso, as probabilidades de cobertura, Figura 7c, não diferem mais de $0,016$ do nível de confiança adotado, mesmo para $n = 30$ (menor tamanho amostral considerado nesse estudo), para todos os cenários e tamanhos amostrais em estudo indicando uma ótima precisão do estimador mesmo para tamanhos amostrais menores.

A distribuição empírica dos dados (histograma) do estimador \hat{q} com a curva normal esti-

mada sobreposta, Figura 8, fornecem indícios claros de normalidade do estimador independente do cenário ou tamanho amostral adotado.

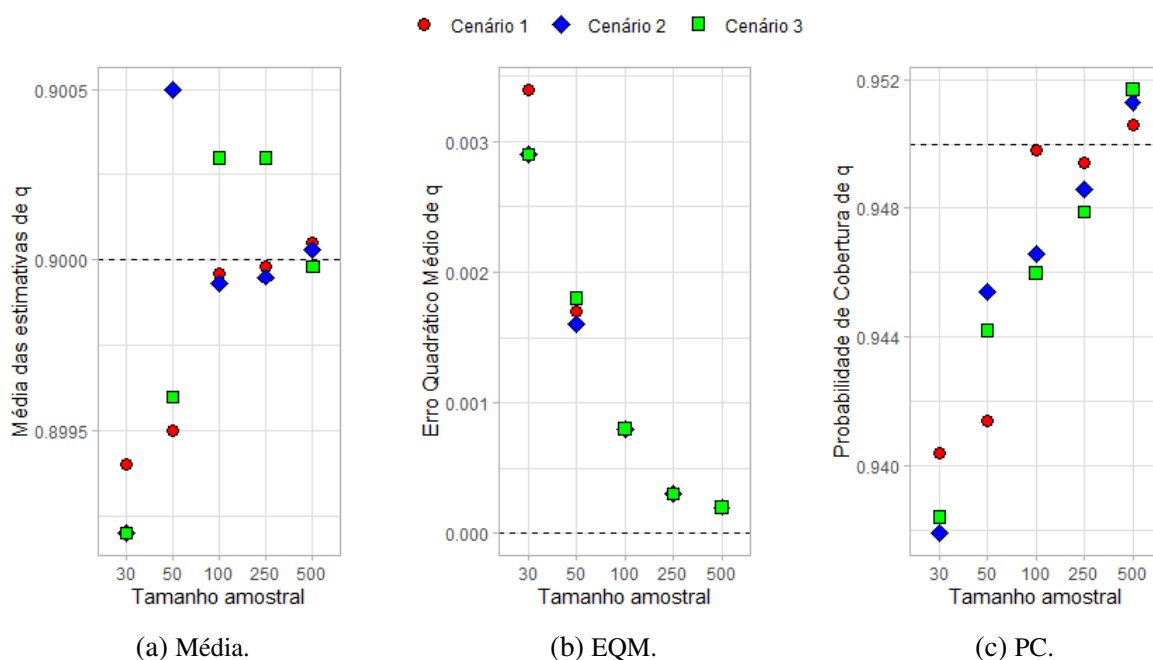


Figura 7 – Média das estimativas, EQM e PC referentes ao parâmetro q .

Fonte: Elaborada pelo autor.

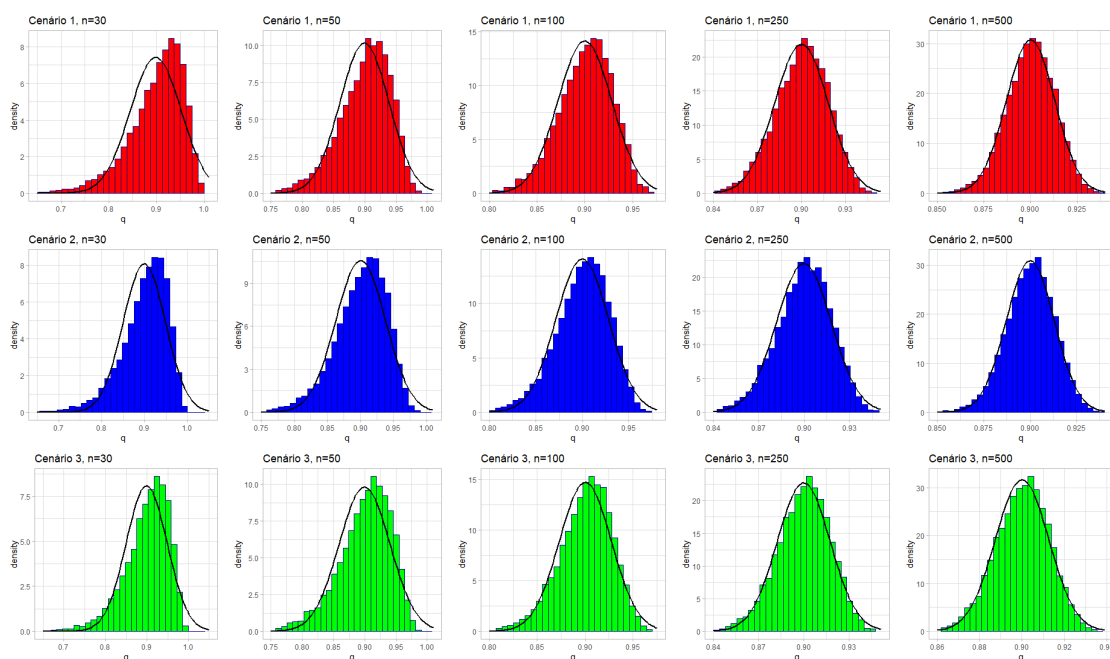


Figura 8 – Distribuições empíricas de \hat{q} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas.

Fonte: Elaborada pelo autor.

Além disso, na Figura 9, foi realizado a mesma análise da Figura descrita anteriormente para o estimador $\hat{u} = \log[-\log \hat{q}]$ (transformação que torna o parâmetro em estudo irrestrito), sendo este utilizado para a construção do intervalo de confiança (3.15). Por meio da referida Figura, pode-se verificar a normalidade assintótica do estimador e um melhor ajuste da distribuição empírica a distribuição Normal para tamanhos amostrais menores ($n = 30$ e 50), retratando a convergência assintótica mais rápida de \hat{u} a essa distribuição, quando comparado a \hat{q} (limitado).

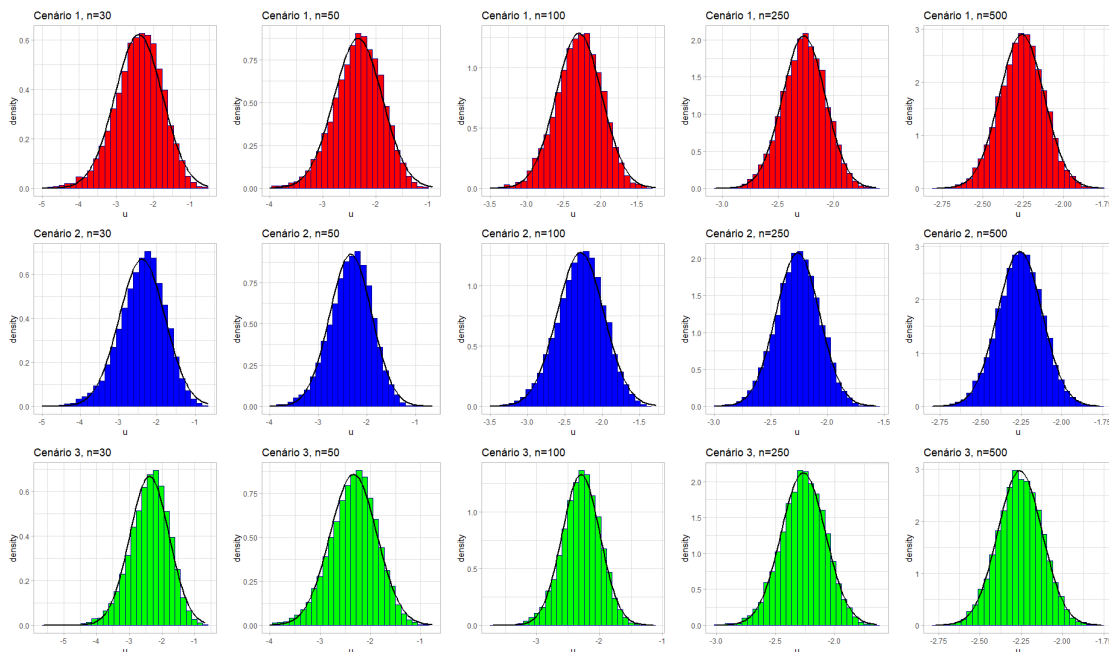


Figura 9 – Distribuições empíricas de \hat{u} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas.

Fonte: Elaborada pelo autor.

Os resultados que se referem ao estimador η são apresentados nas Figuras 10, 11 e 12, em que se constata que as médias das estimativas são aproximadamente iguais ao verdadeiro valor do parâmetro independente do cenário e tamanho amostral e maior precisão do estimador conforme o tamanho da amostra aumenta (Figura 10a).

Destaca-se que, na análise dos resultados das estimativas de EQM de η , Figura 10b, a comparação dos cenários é pertinente, devido a serem adotados diferentes valores de η , os maiores valores de estimativa de EQM são referentes ao Cenário 1, que possui maior valor do verdadeiro parâmetro entre os cenários estudados, o comportamento contrário ocorre para o Cenário 3, em que o verdadeiro valor do parâmetro é 0,75, indicando que o EQM se torna maior à proporção que o parâmetro é maior. No entanto, ressalta-se que as estimativas do Cenário 1, são próximas a zero mesmo para o menor tamanho amostral ($\leq 0,0736$). A ótima precisão do estimador é confirmada pela PC, Figura 10c, que não difere por mais de 0,0299 da confiança fixada para o menor tamanho amostral adotado e que tem um comportamento crescente e convergente para o verdadeiro grau de confiança adotado conforme aumento do tamanho amostral.

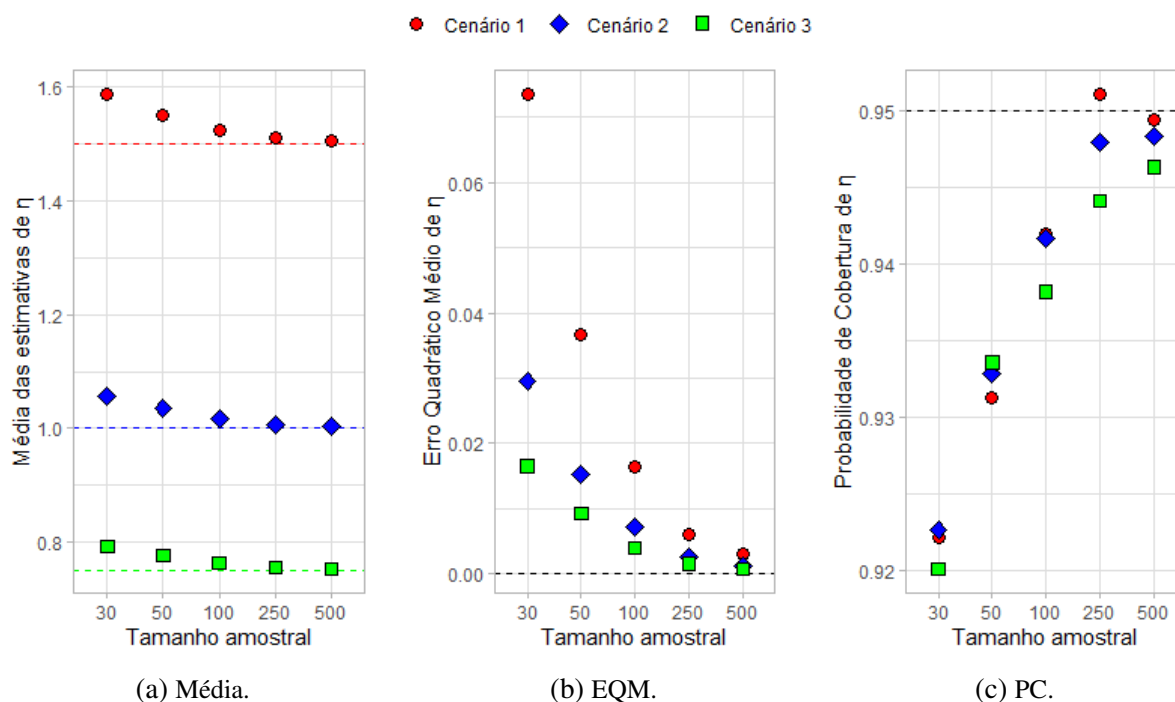


Figura 10 – Média das estimativas, EQM e PC referentes ao parâmetro η .

Fonte: Elaborada pelo autor.

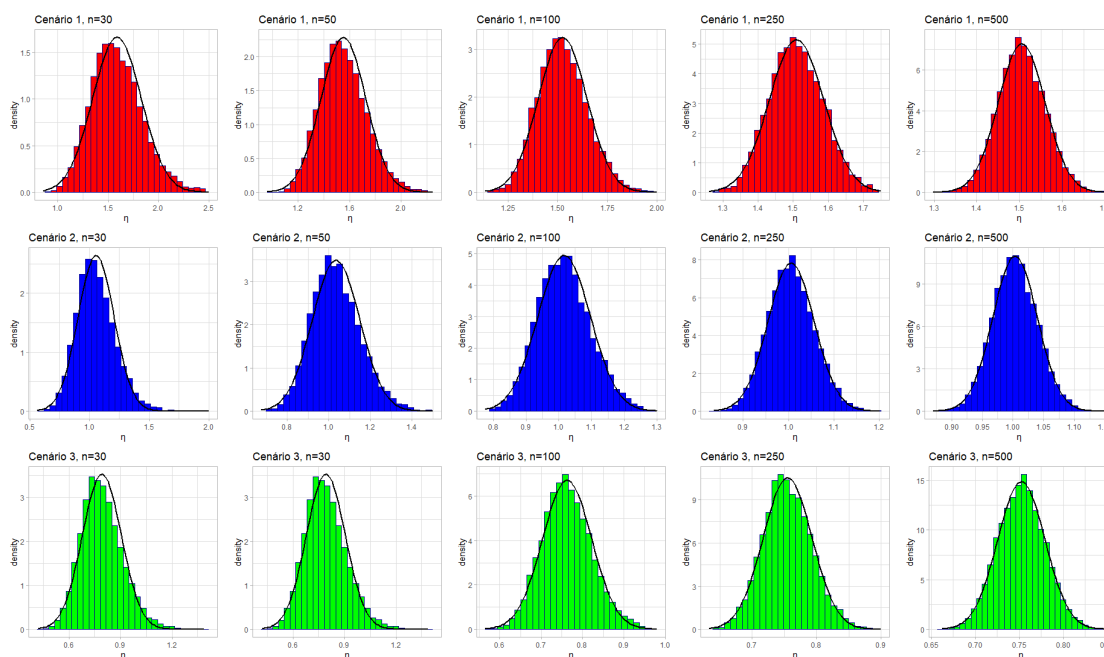


Figura 11 – Distribuições empíricas de $\hat{\eta}$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas.

Fonte: Elaborada pelo autor.

A distribuição empírica dos dados (histograma) do estimador $\hat{\eta}$ com a curva normal estimada sobreposta, Figura 11, fornecem indícios claros de normalidade do estimador independente do cenário ou tamanho amostral adotado. Ademais, na Figura 12, foi realizada a mesma

análise da Figura descrita anteriormente para o estimador $\hat{v} = \log \hat{\eta}$ (transformação que torna o parâmetro em estudo irrestrito), sendo este utilizado para a construção do intervalo de confiança, como em (3.16), em que a normalidade assintótica é novamente evidenciada.

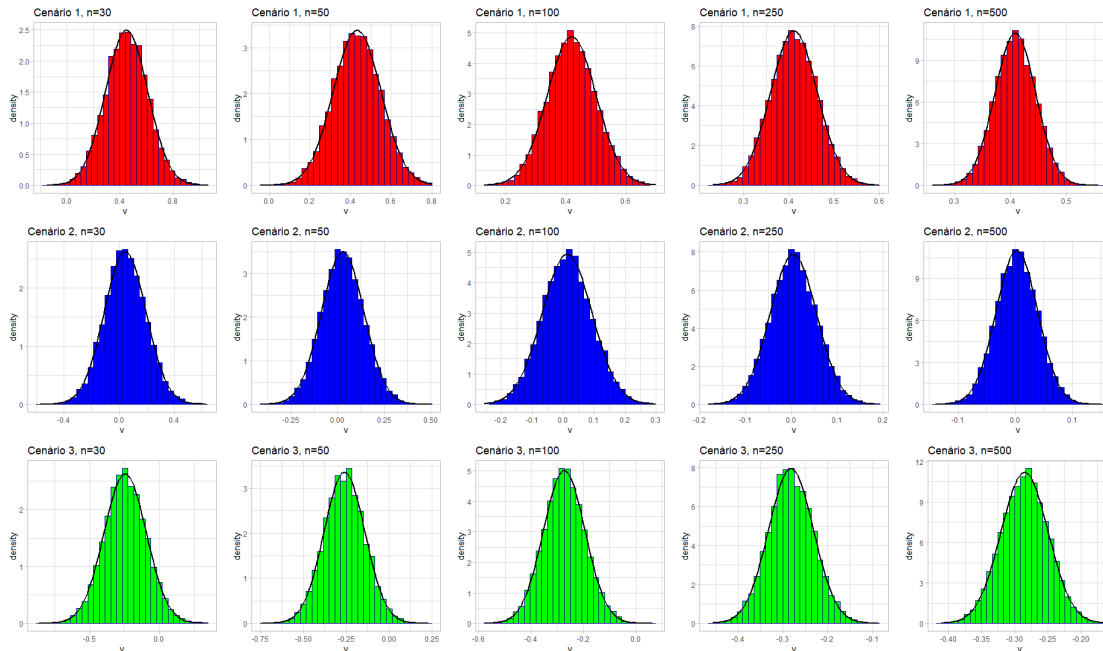


Figura 12 – Distribuições empíricas de \hat{v} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas.

Fonte: Elaborada pelo autor.

Nas Figuras 13 e 14, são apresentados os resultados relacionados ao estimador β_1 . Observa-se que as médias das estimativas, Figura 13a, tendem a ser superestimadas (estimativa maior que o verdadeiro valor do parâmetro) para tamanhos amostrais menores. No entanto, é possível reparar que estas não estão tão distantes do verdadeiro valor do parâmetro, pois as médias das estimativas são menores que 2,15. Quando o tamanho amostral se torna maior as médias das estimativas se aproximam do verdadeiro valor do parâmetro, cujo o valor é 2.

A tendência de precisão das estimativas se confirma pelas estimativas de EQM e PC, Figuras 13b e 13c, que possuem valores próximos a zero e ao verdadeiro grau de confiança adotado para $n = 30$ e convergente aos referidos valores conforme o aumento do tamanho amostral. A normalidade do referido estimador, independente do cenário e tamanho amostral adotado, é evidenciada na Figura 14.

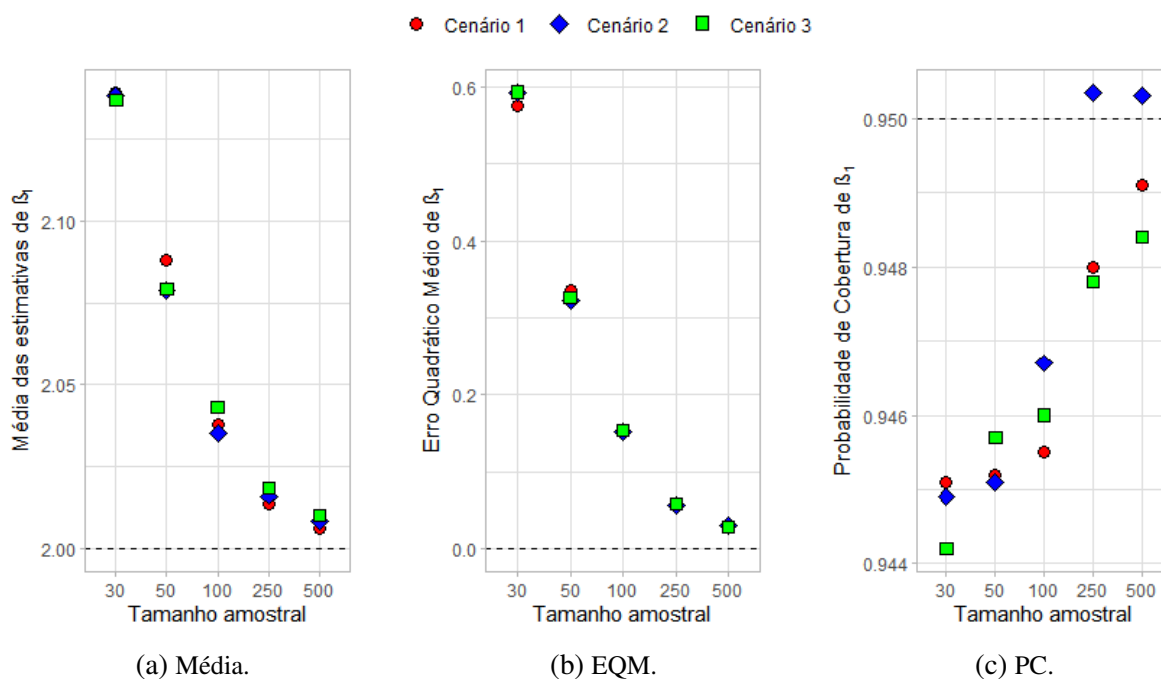


Figura 13 – Média das estimativas, EQM e PC referentes ao parâmetro β_1 .

Fonte: Elaborada pelo autor.

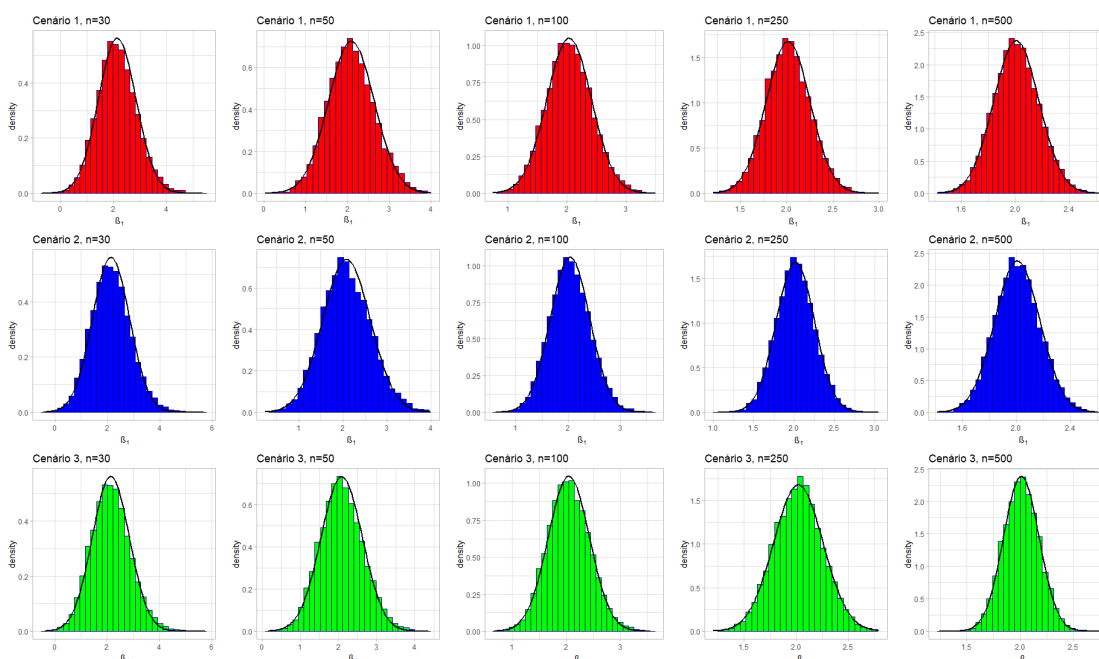


Figura 14 – Distribuições empíricas de $\hat{\beta}_1$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas.

Fonte: Elaborada pelo autor.

Por fim, nas Figuras 15 e 16 são apresentados os resultados pertencentes ao estimador β_2 . Tais resultados são semelhantes para o estimador de β_1 , com a convergência da PC ao verdadeiro grau de confiança adotado proporcionalmente associado ao aumento do tamanho amostral ainda

mais evidente.

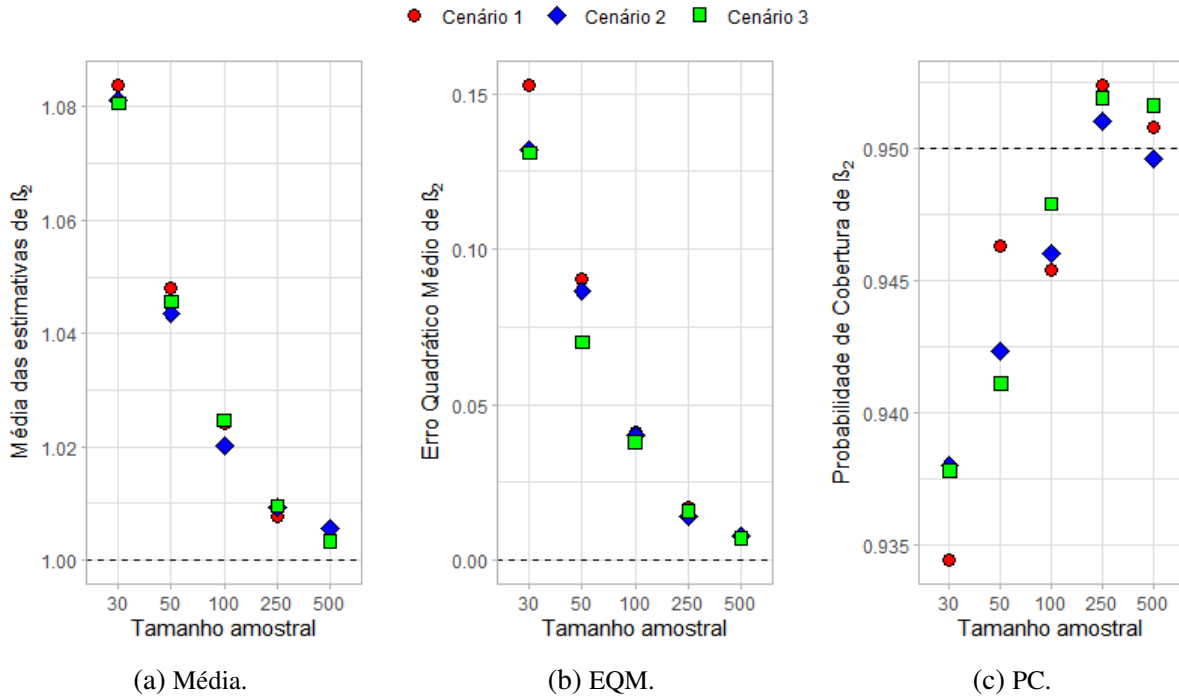


Figura 15 – Média das estimativas, EQM e PC referentes ao parâmetro β_2 .

Fonte: Elaborada pelo autor.

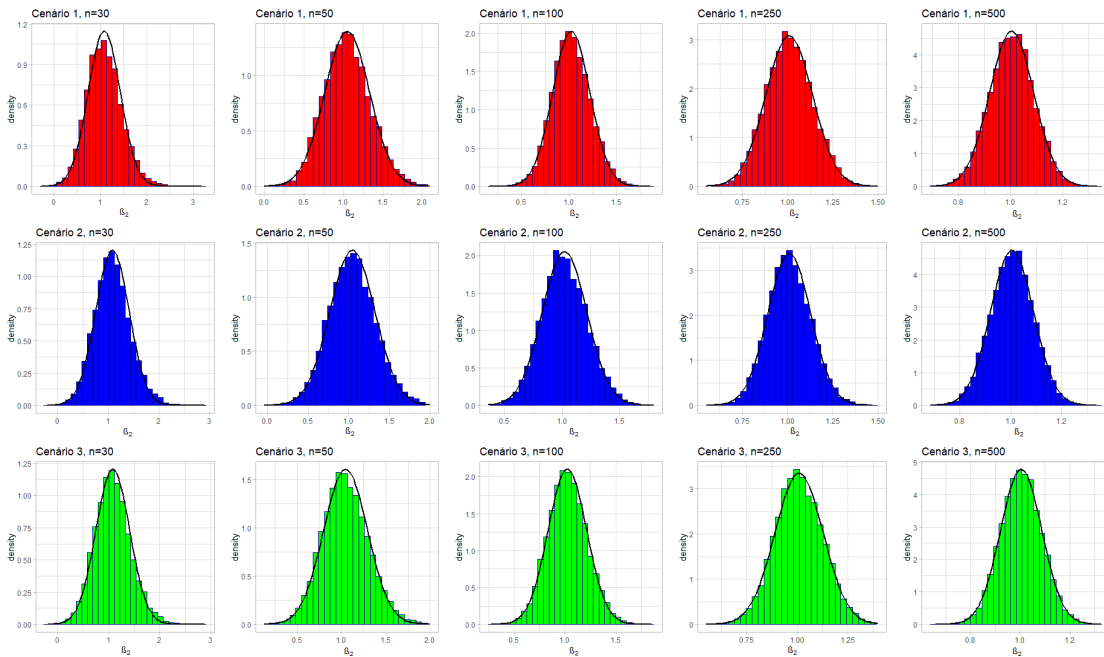


Figura 16 – Distribuições empíricas de $\hat{\beta}_2$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas.

Fonte: Elaborada pelo autor.

Analisando os resultados de maneira geral, nota-se que as simulações apresentaram ótimos resultados, com a média das estimativas dos parâmetros estudados próximas de seus verdadeiros valores, EQM's próximos a zero e PC's próximas ao grau de confiança estipulado mesmo para tamanhos amostrais menores e conforme o tamanho amostral se torna maior, os resultados das respectivas análises convergem para os seus respectivos verdadeiros valores. Além disso, a aguardada normalidade assintótica dos estimadores do modelo é evidenciada.

Assim, pode-se inferir que os estimadores são aproximadamente não viesados. Além disso, pelo fato do aumento do tamanho da amostra conduzir a menores estimativas de EQM e maiores PC, isto é, a menor a variância do estimador, infere-se que os estimadores são consistentes e ainda, pelo que foi evidenciado, os estimadores são assintoticamente normais. A avaliação do comportamento das estimativas para tempos discretos de sobrevivência com presença de censuras será realizada na próxima seção.

4.2 Tempos discretos com a presença de censura

Do ponto de vista teórico, quando há a presença de censuras em determinado conjunto de dados é esperado que as estimativas obtidas se comportem de forma diferente do que foi apresentado na Seção 4.1 em relação às diferenças entre estimativas e verdadeiro valor do parâmetro, devido ao fato do verdadeiro valor de o parâmetro não ser mais o que foi proposto nos Cenários da simulações, visto que as simulações são baseadas na Função Distribuição Acumulada (FDA) do modelo que, por sua vez, não leva em conta a parte censurada.

Visando avaliar o comportamento das estimativas do MRCSP-WD na presença de censura, nessa seção, serão avaliados os resultados (Médias das estimativas, EQM e PC) dispostos por parâmetros seguindo o mesmo racional estabelecido na seção anterior considerando percentuais de censura iguais a 5, 10, 15 e 25% (os resultados numéricos estão disponíveis na Tabela 17 no Anexo A). Para devida sumarização e interpretação dos resultados, gráficos de pontos a respeito de cada um dos parâmetros, serão apresentados considerando os diversos Cenários (Tabela 2) e tamanhos amostrais ($n=30, 50, 100, 250$ e 500) separados pelos diversos percentuais de censura citados.

Nas Figuras 17, 18, 19, 20 e 21 são apresentados os resultados referentes ao estimador q . Pode-se observar que quanto maior o percentual de censura maiores são os desvios das estimativas em relação ao verdadeiro valor do parâmetro (Figura 17a), comportamento natural para o estimador, visto que a maior quantidade de observações censuradas, alteram o valor das estimativas a respeito do verdadeiro valor do parâmetro consideravelmente. A probabilidade de cobertura de q apresentada na Figura 17c, que tem como grau de confiança 0,95, reforça essa afirmação. Note que, a maior quantidade de censuras ($n = 500$ e percentual de censura=25%), conduzem aos piores resultados de probabilidade de cobertura e maiores diferenças a respeito do grau de confiança estipulado para o estudo. É importante destacar que, quando a probabilidade

de cobertura estimada difere do grau de confiança adotado, não se pode afirmar que o processo de estimação produz estimativas não condizentes e sim de que as estimativas se alteram consideravelmente na presença de censura e o verdadeiro valor parâmetro não deve ser utilizado como referência, pois a FDA utilizada para gerar os dados não considera a parte censurada. Consequentemente o verdadeiro valor do parâmetro não é o proposto e sim um valor que se aproxima das estimativas apresentadas na Figura 17a.

As Figuras 18, 19, 20 e 21 reforçam o que foi descrito e permitem observar que, frente a esse comportamento nos diferentes percentuais de censura, o estimador tem a propriedade de normalidade assintótica a exemplo do que ocorre para o cenário sem censuras. As Figuras 18, 19, 20 e 21 apresentam a distribuição empírica dos dados (histograma) do estimador \hat{q} com a curva normal estimada sobreposta para os diferentes tamanhos amostrais e percentuais de censura igual a 5, 10, 15 e 25% de censuras respectivamente, em que comprova-se a normalidade assintótica dos estimadores de q . Note que, à medida que o tamanho da amostra se torna maior a normalidade torna-se ainda mais evidente. Tal comportamento quanto a normalidade assintótica é similar para u (transformação proposta para q), para o parâmetro η e sua transformação v e ainda para os parâmetros β_1 e β_2 (Anexo B). As Figuras 22, 23 e 24 apresentam as médias das estimativas, EQM's e PC dos parâmetros η , β_1 e β_2 , respectivamente.

De modo geral, fundamentado nos resultados das simulações da presente seção, nota-se que o comportamento dos estimadores na presença de observações censuradas conduz a desvios das estimativas em relação ao verdadeiro valor do parâmetro que são maiores à medida que o percentual de censura aumenta, consequentemente a probabilidade de cobertura se torna cada vez mais distante a respeito do verdadeiro grau de confiança.

No entanto, é importante ressaltar que esse comportamento é natural quando há dados censurados e que a normalidade dos estimadores foi atendida. Sendo assim, a construção de intervalos de confiança aproximados por uma distribuição Normal para os parâmetros da distribuição WD e suas respectivas transformações apresentadas, tal como para os parâmetros associados as covariáveis, é válida podendo ser utilizados para a estimação intervalar dos parâmetros. Em virtude disso, teste de hipóteses aproximados por uma distribuição Normal para verificar a significância da covariável também podem ser utilizados nas aplicações.

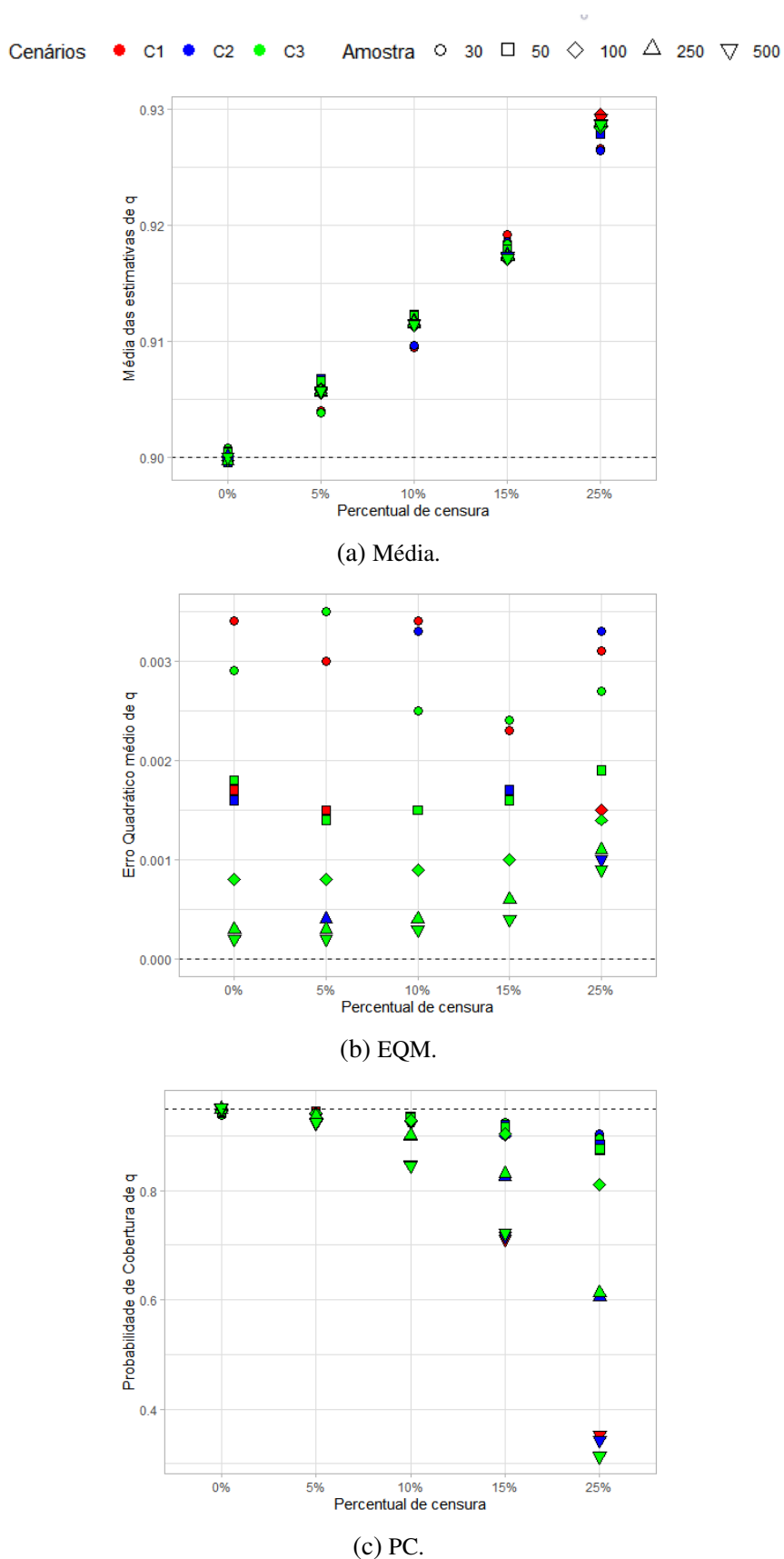


Figura 17 – Média das estimativas, EQM e PC referentes ao parâmetro q por percentual de censura.

Fonte: Elaborada pelo autor.

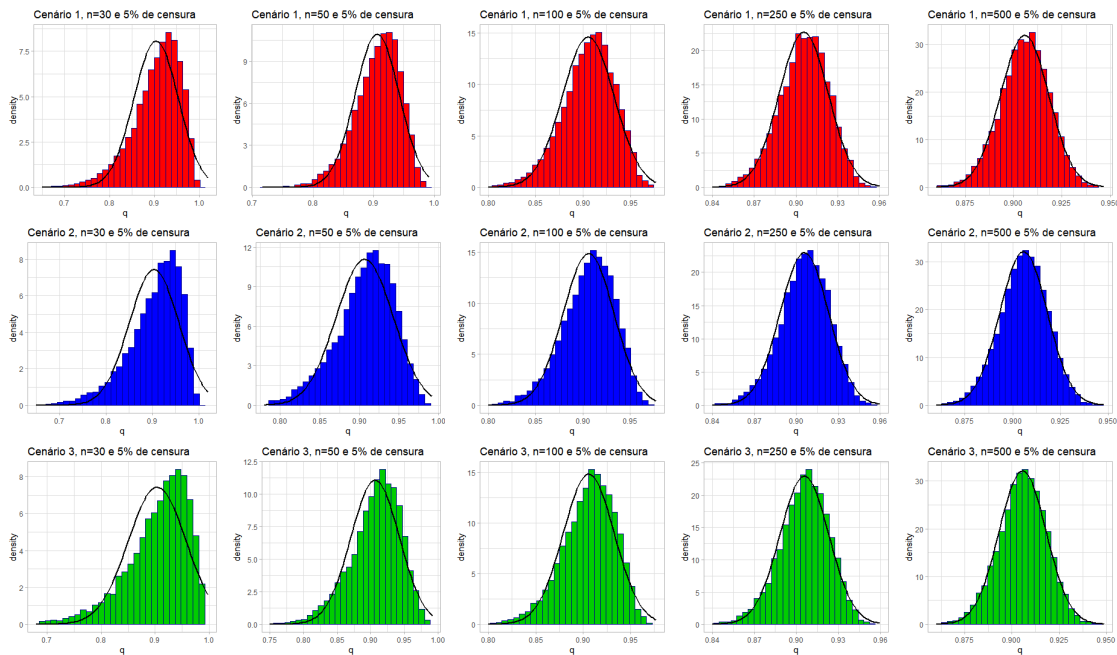


Figura 18 – Distribuições empíricas de \hat{q} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 5% de censuras.

Fonte: Elaborada pelo autor.

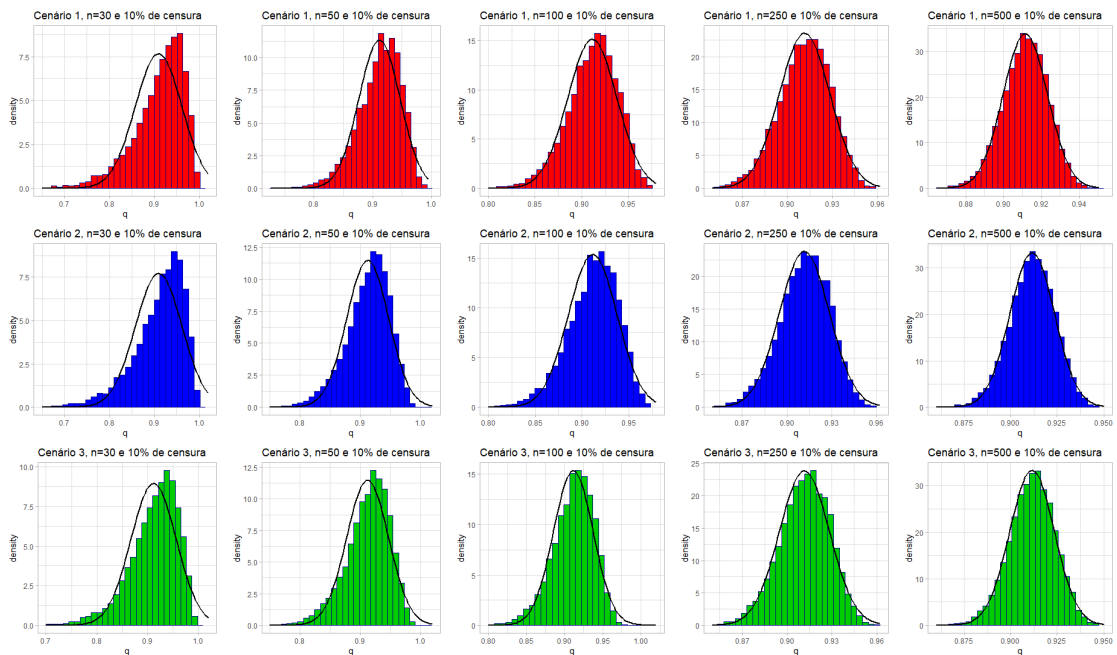


Figura 19 – Distribuições empíricas de \hat{q} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 10% de censuras.

Fonte: Elaborada pelo autor.

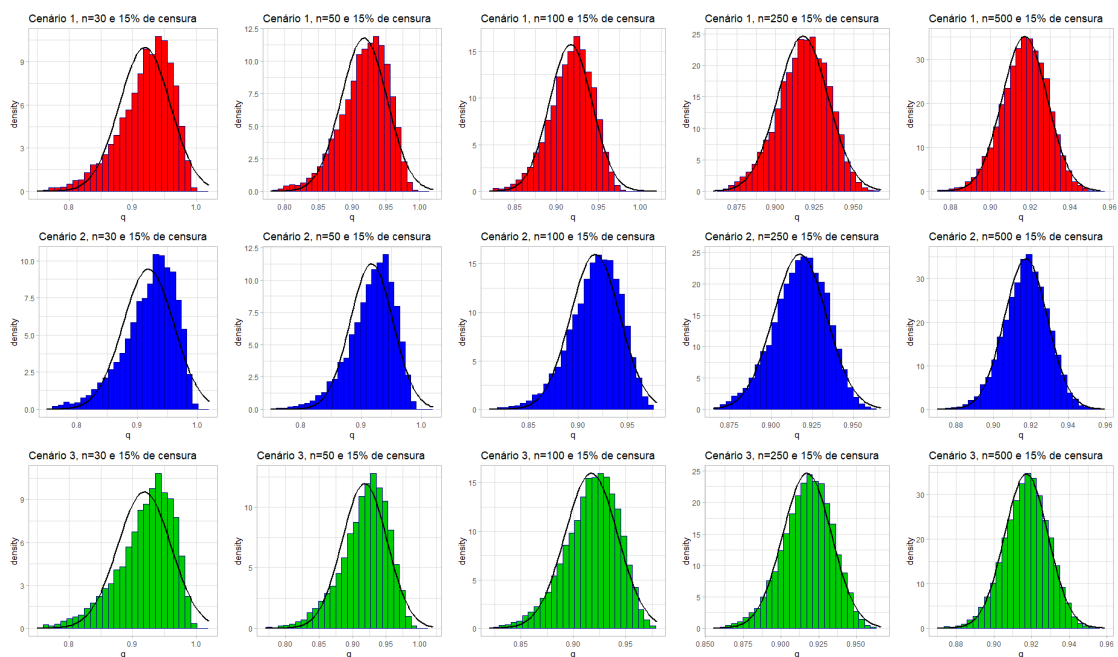


Figura 20 – Distribuições empíricas de \hat{q} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 15% de censuras.

Fonte: Elaborada pelo autor.

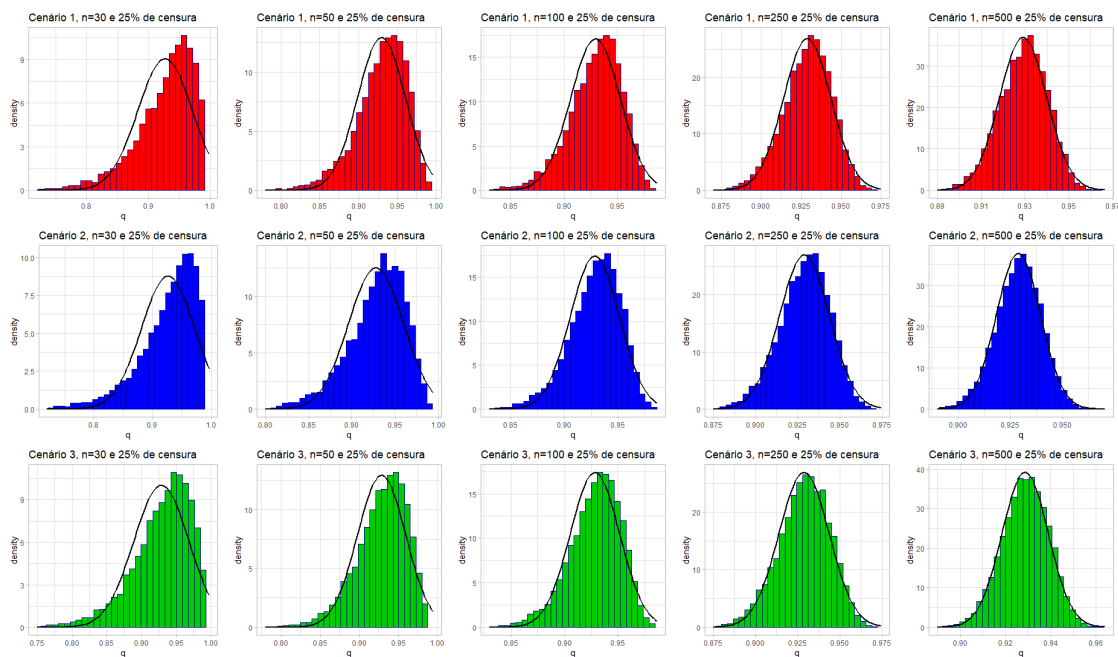
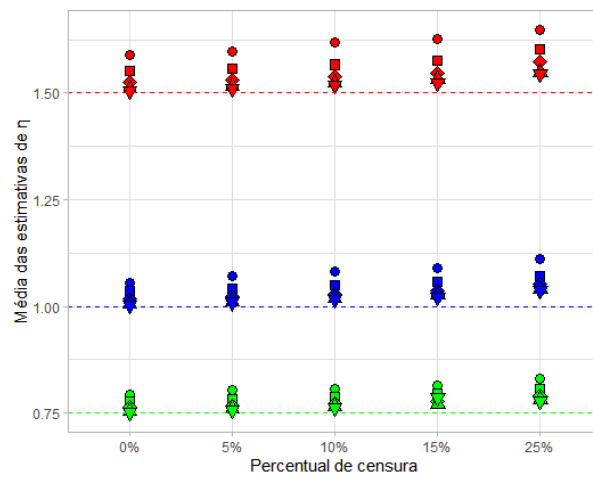


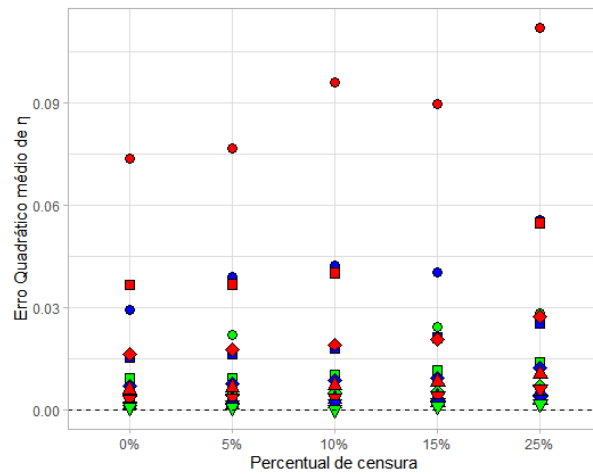
Figura 21 – Distribuições empíricas de \hat{q} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 25% de censuras.

Fonte: Elaborada pelo autor.

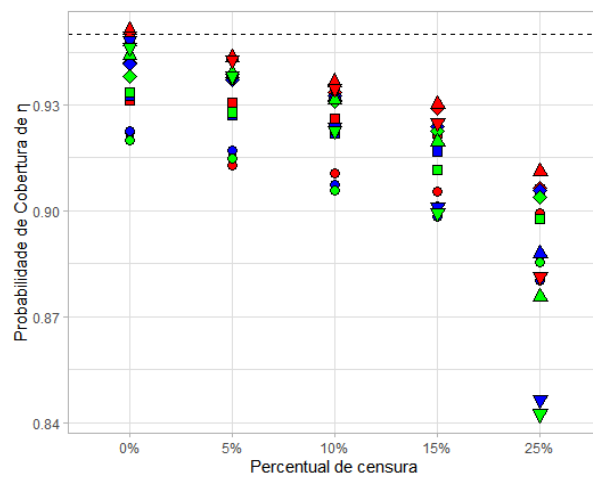
Cenários ● C1 ● C2 ● C3 Amostra ○ 30 □ 50 ◇ 100 △ 250 ▽ 500



(a) Média.



(b) EQM.

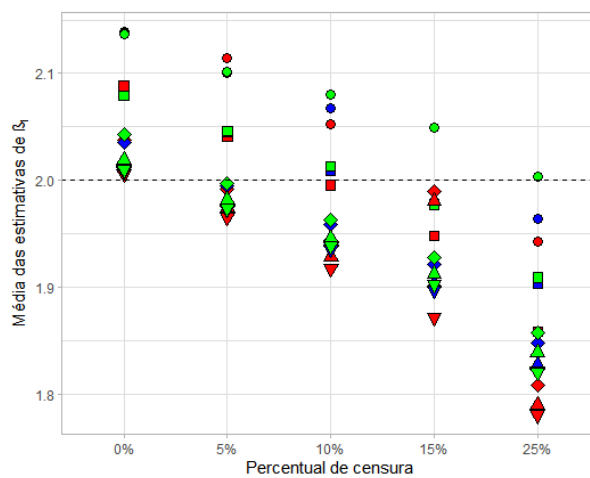


(c) PC.

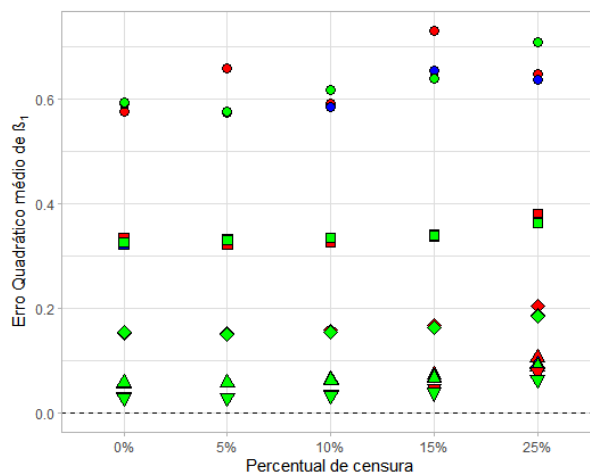
Figura 22 – Média das estimativas, EQM e PC referentes ao parâmetro η por percentual de censura.

Fonte: Elaborada pelo autor.

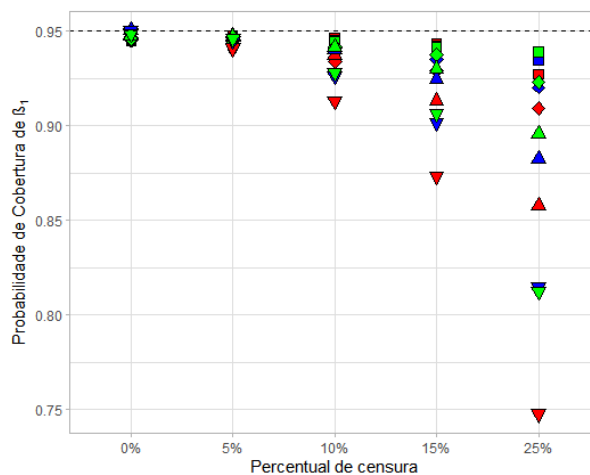
Cenários ● C1 ● C2 ● C3 Amostra ○ 30 □ 50 ◇ 100 △ 250 ▽ 500



(a) Média.



(b) EQM.



(c) PC.

Figura 23 – Média das estimativas, EQM e PC referentes ao parâmetro β_1 por percentual de censura.

Fonte: Elaborada pelo autor.

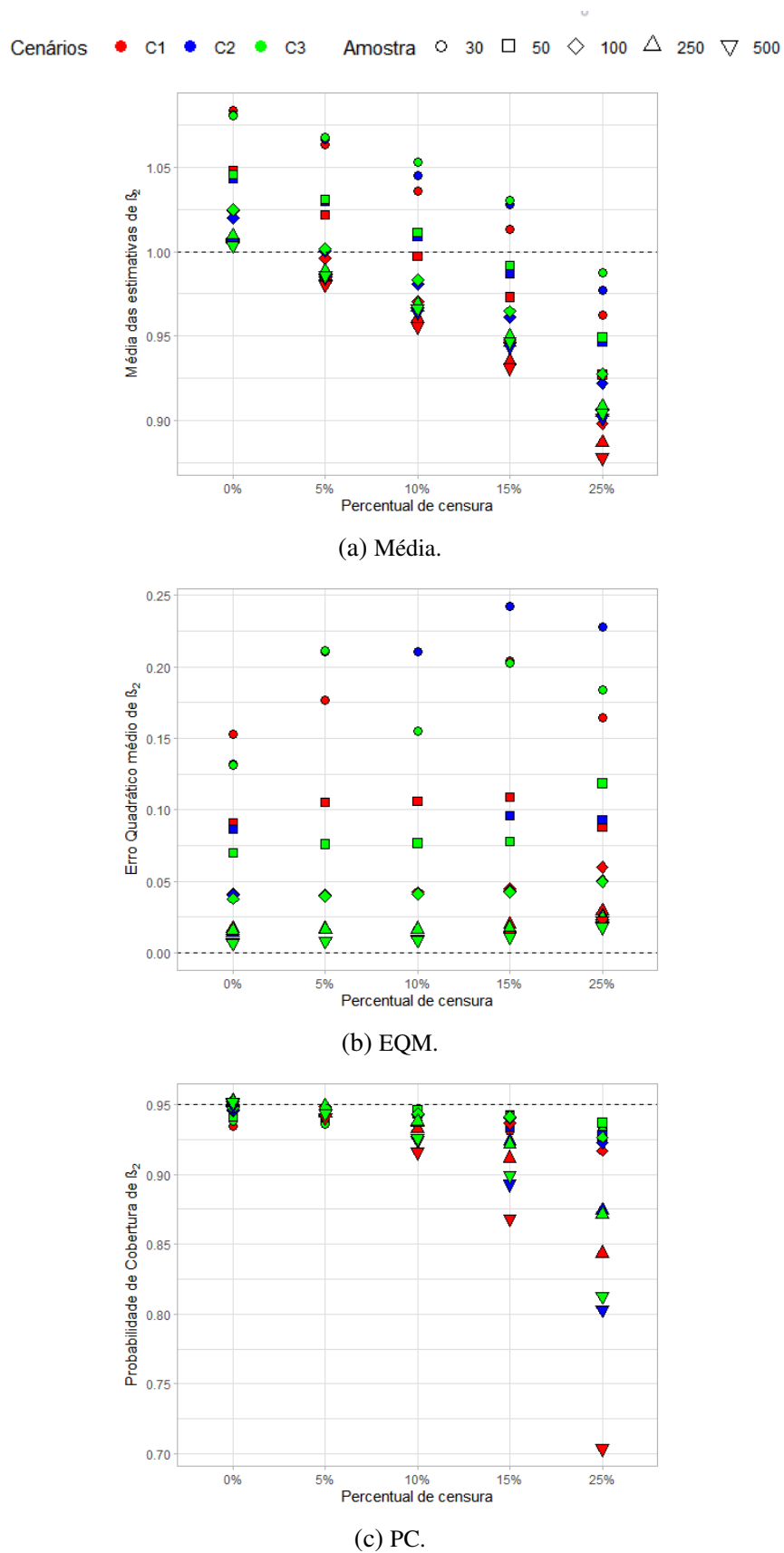


Figura 24 – Média das estimativas, EQM e PC referentes ao parâmetro β_2 por percentual de censura.

Fonte: Elaborada pelo autor.

APLICAÇÃO EM DADOS REAIS

5.1 Dados de Leucemia

A Leucemia Mieloide Aguda [Acute Myeloid Leukemia (AML)] é um tipo de câncer no qual a medula óssea produz um grande número de células sanguíneas anormais. É o tipo mais comum de leucemia aguda em adultos e piora rapidamente se não for tratado.

O conjunto de dados da presente aplicação é proveniente do livro do [Lee e Wang \(2003\)](#) e se refere ao tempo de sobrevivência em semanas de 30 pacientes com AML. Dois fatores prognósticos possíveis (ou covariáveis) são: a idade e estado de celularidade. Neste estudo essas duas covariáveis foram consideradas de forma dicotômica:

$$z_1 = \begin{cases} 1, & \text{se o paciente tem 50 anos ou mais.} \\ 0, & \text{caso contrário.} \end{cases}$$

$$z_2 = \begin{cases} 1, & \text{se a celularidade do corte do coágulo da medula for 100\%.} \\ 0, & \text{caso contrário.} \end{cases}$$

Os dados da aplicação são apresentados na Tabela 3 e a Figura 25 apresenta a estimativa da função de sobrevivência (sem considerar as covariáveis) obtida pelo estimador de Kaplan-Meier.

Por meio da Tabela 3, nota-se que o tempo máximo de sobrevivência é igual a 45 semanas e que o conjunto de dados possui $7/30 = 23,33\%$ de observações censuradas. Repare ainda como resultados obtidos da Tabela 3, que $56,67\%$ dos pacientes têm 50 anos ou mais. A mesma porcentagem corresponde a pacientes que têm a celularidade do corte do coágulo da medula igual a 100%. Portanto, o número de pacientes em ambos fatores prognósticos possíveis são iguais. No entanto, a forma pelo qual se distribuem perante ao tempo de sobrevivência são diferentes, o que pode ser observado por meio da Figura 26, em que são apresentadas as funções de sobrevivência de cada uma das covariáveis separadas por categorias.

Tabela 3 – Tempo de sobrevivência e dados para dois fatores prognósticos possíveis de 30 pacientes com AML.

Tempo de sobrevivência	z_1	z_2	Tempo de sobrevivência	z_1	z_2
18	0	0	8	1	0
9	0	1	2	1	1
28+	0	0	26+	1	0
31	0	1	10	1	1
39+	0	1	4	1	0
19+	0	1	3	1	0
45+	0	1	4	1	0
6	0	1	18	1	1
8	0	1	8	1	1
15	0	1	3	1	1
23	0	0	14	1	1
28+	0	0	3	1	0
7	0	1	13	1	1
12	1	0	13	1	1
9	1	0	35+	1	0

Nota: “+” indica as observações censuradas.

Fonte: Lee e Wang (2003).

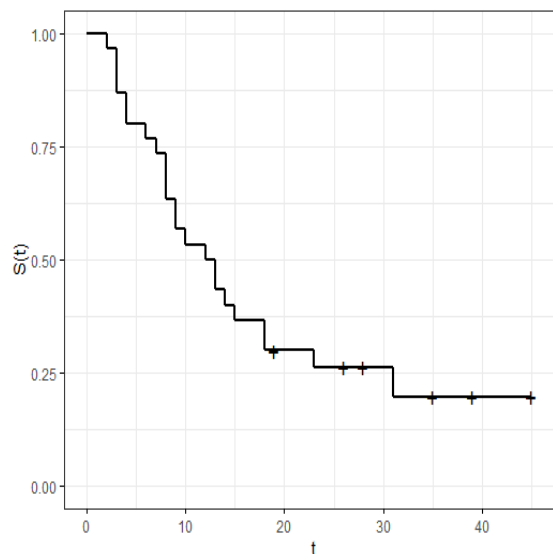


Figura 25 – Função de sobrevivência estimada pelo método de Kaplan-Meier do conjunto dos dados da Tabela 3.

Fonte: Elaborada pelo autor.

Por meio da Figura 26, observa-se que pacientes com menos de 50 anos tem maior probabilidade de sobrevivência visto que o Estimador de Kaplan-Meier para a referida categoria é maior em relação a função de sobrevivência estimada dos pacientes com 50 anos ou mais ($z_1 = 1$) para qualquer tempo t . Esta distinção não é possível de ser realizada para o outro fator prognóstico, visto que as curvas de sobrevivência sempre estão bem próximas, chegando ao ponto de se cruzarem.

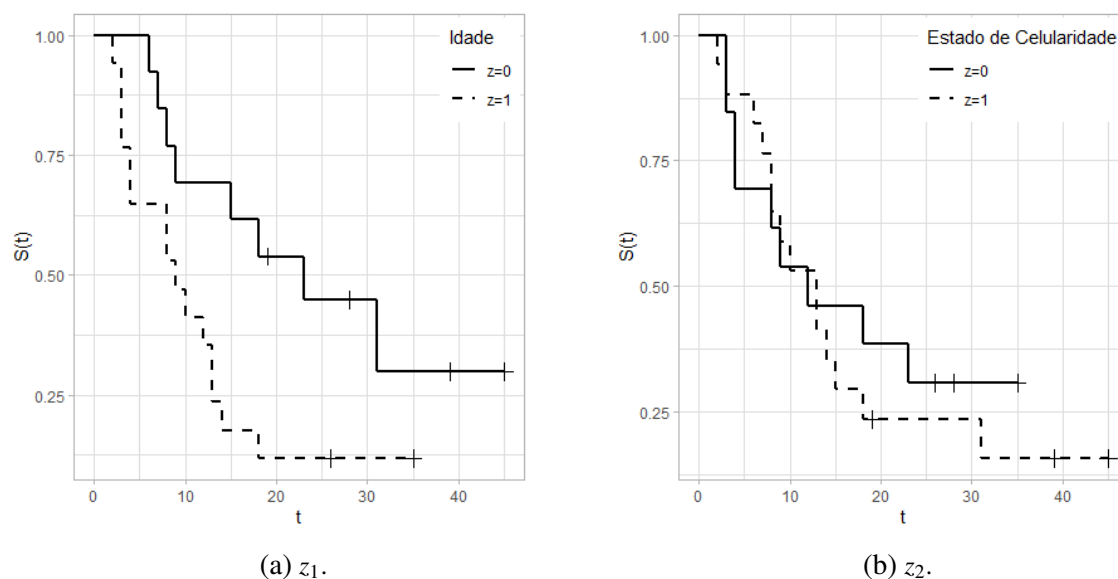


Figura 26 – Função de sobrevivência estimada para cada nível das covariáveis do conjunto de dados da Tabela 3.

Fonte: Elaborada pelo autor.

Dando prosseguimento ao estudo, foi realizado ajuste dos Modelos de Regressão: Chances de Sobrevivência Proporcionais Weibull Discreto (MRCSP-WD) e Weibull Discreto (MRWD). Inicialmente foram ajustados os modelos MRCSP-WD e MRWD múltiplos com o intuito de verificar a significância de suas covariáveis ($H_0 : \beta_1 = 0$ e $H_0 : \beta_2 = 0$). Os resultados de valor-p dos modelos múltiplos são apresentados na Tabela 4.

Tabela 4 – Teste de significância das covariáveis para MRCSP-WD e MRWD múltiplo (Dados de Leucemia).

MRCSP-WD			MRWD		
Variável	Parâmetro	valor-p	Variável	Parâmetro	valor-p
Idade	β_1	0,00014	Idade	β_1	0,00607
Estado de Celularidade	β_2	0,37572	Estado de Celularidade	β_2	0,26385

Nível de referência das variáveis: Idade = O paciente tem menos de 50 anos ($z_1 = 0$);

Estado de Celularidade = A celularidade do corte do coágulo da medula é menor que 100% ($z_2 = 0$).

Fonte: Elaborada pelo autor.

De acordo com os resultados da Tabela 4, a covariável estado de celularidade não estaria influenciando na sobrevivência dos pacientes. Por outro lado, a covariável idade é significativa (ao nível de significância de 5%) em ambos os modelos. Desse modo, excluiu-se a covariável z_2 de ambos modelos e foi realizado o teste de significância ajustando-se apenas a covariável idade e os resultados são apresentados na Tabela 5.

Tabela 5 – Teste de significância da covariável idade para MRCSP-WD e MRWD (Dados de Leucemia).

MRCSP-WD			MRWD		
Variável	Parâmetro	valor-p	Variável	Parâmetro	valor-p
Idade	β_1	0,00012	Idade	β_1	0,01039

Nível de referência: O paciente tem menos de 50 anos ($z_1 = 0$).

Fonte: Elaborada pelo autor.

Os resultados da Tabela 5 comprovam que a covariável idade influencia a sobrevivência dos pacientes com AML. Desse modo, levando em conta a covariável idade, na Subseção 5.1.1 será realizado o estudo de verificação de suposição de chances de sobrevivência proporcionais utilizando os métodos apresentados na Seção 3.3. Em seguida, serão apresentadas as estimativas do MRCSP-WD na Subseção 5.1.2, tal como o ajuste do referido modelo aos dados na Subseção 5.1.3.

5.1.1 Verificação da suposição de chances de sobrevivência proporcionais

A suposição de chances de sobrevivência proporcionais foi verificada para os dados da Tabela 3, observando a referida proporcionalidade para cada um dos níveis da covariável idade, por meio dos gráficos: $t \times \log(o_i(t))$, $i = 0, 1$; $\log(O_0(t)) \times \log(O_1(t))$ e o teste de hipótese proposto em 3.23. Os resultados são apresentados na Figura 27:

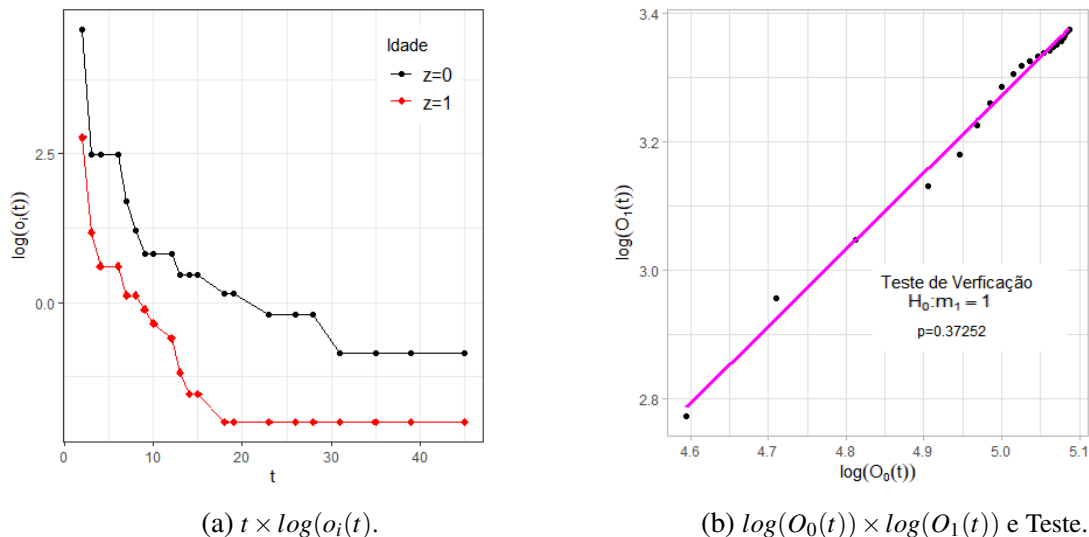


Figura 27 – Verificação de suposição de chances de sobrevivência proporcionais para a covariável idade (Dados de Leucemia).

Fonte: Elaborada pelo autor.

Por meio dos resultados do teste apresentado na Figura 27b nota-se que, a suposição de chances de sobrevivência proporcionais não foi rejeitada (tendo em vista um nível de significância de 5%), o gráfico apresentado $\log(O_0(t)) \times \log(O_1(t))$ corrobora para tal afirmação, visto que

os pontos formados pelas coordenadas estão próximos à reta de regressão ajustada. O gráfico $t \times \log(o_i(t))$, $i = 0, 1$, apresentado na Figura 27a, confirma a afirmação que a chances de sobrevivência proporcionais é válida visto que, o logaritmo da função de chances de sobrevivência das diferentes categorias são aproximadamente paralelas entre si.

5.1.2 Inferência dos parâmetros do MRCSP-WD

A verificação da suposição de chances de sobrevivência proporcionais realizada na subseção anterior indicou haver evidências que o modelo de regressão com tal característica mostra-se adequado para estimação dos parâmetros do conjunto de dados em estudo. Assim, foram calculadas as estimativas pontuais e intervalares dos parâmetros para o MRCSP considerando como distribuição base a distribuição Weibull Discreta (WD). As estimativas dos parâmetros do Modelo de Regressão Chances de Sobrevivência Proporcional - Weibull Discreto (MRCSP-WD), levando em conta a covariável idade, são apresentadas na Tabela 6.

Tabela 6 – Estimativas dos parâmetros do MRCSP-WD (Dados de Leucemia).

Variável	Parâmetro	Estimativa	Erro Padrão	IC (95%)	valor-p
	q	0,9963	0,0017	[0,9928; 0,9981]	-
Idade	η	1,6038	0,1555	[1,2991; 1,9086]	-
	β_1	-2,0152	0,5248	[-3,0483; -0,9867]	0,00012

Nível de referência: O paciente tem menos de 50 anos ($z_1 = 0$).

Fonte: Elaborada pelo autor.

A partir das estimativas apresentadas na Tabela 6, pode-se obter uma interpretação, quanto a chance de sobrevivência para as diferentes categorias da covariável idade. Uma vez que $\exp\{\beta_1\}$ representa a razão das chances de sobrevivência dos diferentes grupos, constante ao longo do tempo, admitindo o grupo de pacientes que tenha 50 anos ou mais ($z_1 = 1$). Nesse contexto, a chance de sobrevivência do paciente que tenha 50 anos ou mais é $\exp\{-2,0152\} = 0,1333$ vezes a chance de sobrevivência dos pacientes com menos de 50 anos.

Portanto, a chance de sobrevivência do paciente que tenha 50 anos ou mais é 0,8667 vezes menor do que a chance de sobrevivência do paciente que tenha menos de 50 anos ($z_1 = 0$). A título de ilustração do que foi descrito, observe o comportamento das Figuras 26a e 27a.

5.1.3 Verificação do ajuste do MRCSP-WD e comparação com o Modelo de Regressão Weibull Discreto (MRWD)

Para análise diagnóstica do modelo, a título de comparação, foi ajustado um modelo de regressão mais simples levando em conta a distribuição WD com o parâmetro q , recebendo o preditor linear considerando a função de ligação logito, isto é:

$$q = \frac{\exp\{\mathbf{Z}'\boldsymbol{\beta}\}}{1 + \exp\{\mathbf{Z}'\boldsymbol{\beta}\}}.$$

Tal modelo será denominado como Modelo de Regressão Weibull Discreto (MRWD), a fim de compará-lo com o modelo em estudo, e se a estrutura do modelo de chances de sobrevivência proporcionais contribuirá para estimativas precisas.

As estimativas dos parâmetros do MRWD, levando em conta a covariável idade, são apresentadas na Tabela 7.

Tabela 7 – Estimativas dos parâmetros do MRWD (Dados de Leucemia).

Parâmetro	Estimativa	Erro Padrão	IC(95%)	valor-p
η	1,2606	0,2106	[0,8479; 1,6733]	-
β_0	4,2125	0,7864	[2,8798; 5,9626]	-
β_1	-1,1611	0,4531	[-2,0492; -0,2729]	0,01039

Nível de referência: O paciente tem menos de 50 anos.

Fonte: Elaborada pelo autor.

Desse modo, para avaliação de ajuste dos modelos aos dados a título de comparação com o modelo mais simples, foram esboçados os gráficos de sobrevivência do estimador de Kaplan-Meier (K-M) e das sobrevivências dos modelos em estudo para cada um dos níveis da covariável, com intuito de analisar o conjunto de gráficos e interpretar o seu ajuste de maneira global (Figura 28). Além disso, foram calculados os erros máximos cometidos na estimação, denotados por ε , para cada um dos referidos níveis, para os dois modelos em estudo, e são apresentados na Tabela 8. Este erro é baseado na diferença máxima entre as estimativas da sobrevivência dos modelos em estudo ($\widehat{S}(t)$) e as estimativas empíricas (K-M) representadas por $\widehat{S}_{km}(t)$, e é definido por (NAKANO; CARRASCO, 2006):

$$\varepsilon = \max|\widehat{S}(t) - \widehat{S}_{km}(t)|. \quad (5.1)$$

Tabela 8 – Erros máximos oriundos da estimação dos modelos (Dados de Leucemia).

Nível da Covariável	$\varepsilon_{MRCSP-WD}$	ε_{MRWD}
$z_1 = 0$	0,1909	0,1477
$z_1 = 1$	0,1341	0,1738

Fonte: Elaborada pelo autor.

Note, por meio da Figura 28a aliada aos resultados da Tabela 8, que ambos os modelos possuem um bom ajuste aos dados, estando as estimativas de sobrevivência dos referidos modelos sempre próximas às estimativas empíricas.

Para comprovação do bom ajuste do MRCSP-WD aos dados, na Figura 28b são apresentadas as estimativas da função de sobrevivência com os respectivos Intervalos de Confiança (IC) com 95% de confiança para as funções de sobrevivência estimados via método delta para os dois níveis da covariável idade, com as estimativas de sobrevivência via estimador de Kaplan-Meier. Note que a maioria das estimativas empíricas encontram-se dentro dos ICs, corroborando o ótimo ajuste do MRCSP-WD aos dados da presente aplicação.

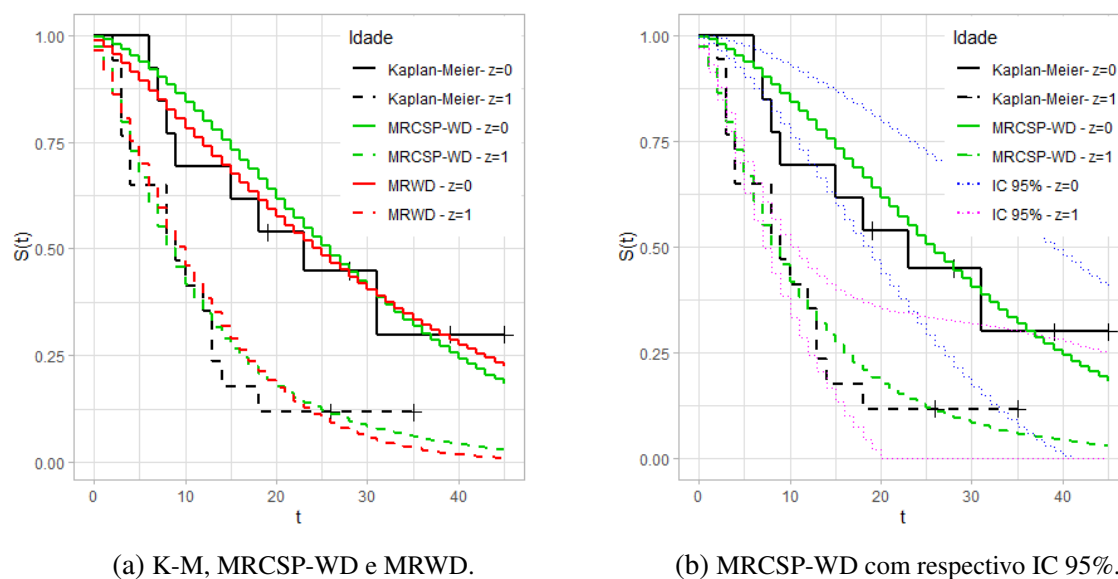


Figura 28 – Ajuste dos modelos aos dados de Leucemia por nível da covariável idade.

Fonte: Elaborada pelo autor.

5.2 Dados de Mieloma

O Mieloma múltiplo é uma doença maligna caracterizada pelo acúmulo de células plasmáticas anormais, um tipo de glóbulo branco, na medula óssea. A proliferação das células plasmáticas anormais dentro do osso causa dor e a destruição do tecido ósseo. Pacientes com mieloma múltiplo também apresentam anemia, hemorragias, infecções recorrentes e fraqueza. A não ser que tratada, a condição é invariavelmente fatal.

O conjunto de dados da presente aplicação é proveniente do livro do [Collett \(2015\)](#). O estudo foi realizado no Centro Médico da Universidade de West Virginia, EUA, visando examinar a associação entre os valores de certas variáveis explicativas ou numéricas e o tempo de sobrevivência dos pacientes. A variável resposta foi o tempo, em meses, desde o diagnóstico até a morte por mieloma.

Assim, os dados da Tabela 9 referem-se a 48 pacientes, sendo que a codificação do *status* de sobrevivência de um indivíduo na Tabela é 0 que denota uma observação censurada, isto é, o paciente não veio a óbito no momento em que o estudo foi concluído, e 1 denota a morte por Mieloma múltiplo. No momento do diagnóstico, os valores de uma série de variáveis explicativas foram registrados para cada paciente. Estes incluíram a idade do paciente em anos, sexo (1 = masculino, 2 = feminino), os níveis de Nitrogênio ureico no sangue (Bun), Cálcio sérico (Ca) e Hemoglobina (Hb), a porcentagem de células plasmáticas em medula óssea (Pcells) e uma variável indicadora (Proteína) que denota se a proteína de Bence-Jones estava ou não presente na urina (0 = ausente, 1 = presente).

Tabela 9 – Tempo de sobrevivência de pacientes em um estudo sobre Mieloma múltiplo.

Número do Paciente	Tempo de Sobrevivência	Status	Idade	Sexo	Bun	Ca	Hb	Pcells	Proteína
1	13	1	66	1	25	10	14,6	18	1
2	52	0	66	1	13	11	12,0	100	0
3	6	1	53	2	15	13	11,4	33	1
4	40	1	69	1	10	10	10,2	30	1
5	10	1	65	1	20	10	13,2	66	0
6	7	0	57	2	12	8	9,9	45	0
7	66	1	52	1	21	10	12,8	11	1
8	10	0	60	1	41	9	14,0	70	1
9	10	1	70	1	37	12	7,5	47	0
10	14	1	70	1	40	11	10,6	27	0
11	16	1	68	1	39	10	11,2	41	0
12	4	1	50	2	172	9	10,1	46	1
13	65	1	59	1	28	9	6,6	66	0
14	5	1	60	1	13	10	9,7	25	0
15	11	0	66	2	25	9	8,8	23	0
16	10	1	51	2	12	9	9,6	80	0
17	15	0	55	1	14	9	13,0	8	0
18	5	1	67	2	26	8	10,4	49	0
19	76	0	60	1	12	12	14,0	9	0
20	56	0	66	1	18	11	12,5	90	0
21	88	1	63	1	21	9	14,0	42	1
22	24	1	67	1	10	10	12,4	44	0
23	51	1	60	2	10	10	10,1	45	1
24	4	1	74	1	48	9	6,5	54	0
25	40	0	72	1	57	9	12,8	28	1
26	8	1	55	1	53	12	8,2	55	0
27	18	1	51	1	12	15	14,4	100	0
28	5	1	70	2	130	8	10,2	23	0
29	16	1	53	1	17	9	10,0	28	0
30	50	1	74	1	37	13	7,7	11	1
31	40	1	70	2	14	9	5,0	22	0
32	1	1	67	1	165	10	9,4	90	0
33	36	1	63	1	40	9	11,0	16	1
34	5	1	77	1	23	8	9,0	29	0
35	10	1	61	1	13	10	14,0	19	0
36	91	1	58	2	27	11	11,0	26	1
37	18	0	69	2	21	10	10,8	33	0
38	1	1	57	1	20	9	5,1	100	1
39	18	0	59	2	21	10	13,0	100	0
40	6	1	61	2	11	10	5,1	100	0
41	1	1	75	1	56	12	11,3	18	0
42	23	1	56	2	20	9	14,6	3	0
43	15	1	62	2	21	10	8,8	5	0
44	18	1	60	2	18	9	7,5	85	1
45	12	0	71	2	46	9	4,9	62	0
46	12	1	60	2	6	10	5,5	25	0
47	17	1	65	2	28	8	7,5	8	0
48	3	0	59	1	90	10	10,2	6	1

Fonte: Collett (2015).

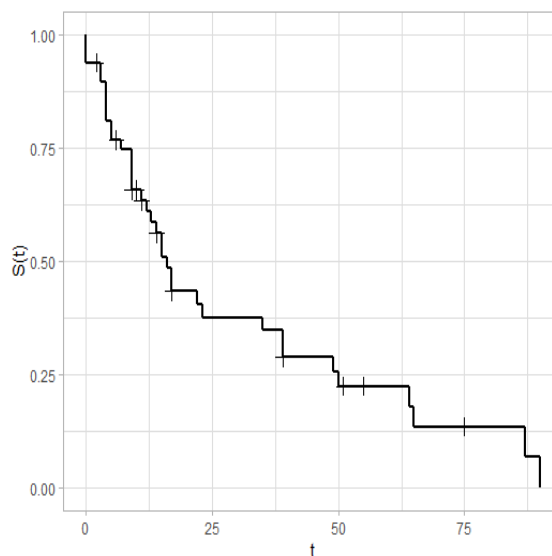


Figura 29 – Sobrevivência via Estimador de Kaplan-Meier do conjunto dos dados da Tabela 9.

Fonte: Elaborada pelo autor.

Dos dados da Tabela 9, foi realizado uma adequação nos tempos de sobrevivência que passaram a ser o tempo referido na Tabela menos 1, com o intuito de indicar meses completos. Neste contexto, o valor 0 representa aquele paciente que não completou nenhum mês de estudo, o valor 1 representa que o paciente tem um mês completo (mesmo que tenha iniciado o seu segundo mês) e assim sucessivamente. Outra adequação realizada foi a respeito da variável sexo que passou a ter as seguintes categorias: sexo (1 = masculino, 0 = feminino).

Por meio da Tabela 9 e adequações supracitadas, nota-se que o tempo máximo de sobrevivência é igual a 90 meses completos e que o conjunto de dados possui $12/48 = 25,00\%$ de observações censuradas. Na Figura 29, é apresentada a função de sobrevivência via estimador de Kaplan-Meier para os dados da Tabela 9.

Foi realizado um resumo em relação às características dos pacientes apresentados na Tabela 10. Em relação às variáveis qualitativas, foram identificados o número e a porcentagem de pacientes inseridos em cada categoria. A respeito das variáveis quantitativas são explicitadas as estatísticas de mínimo, máximo e média de tais variáveis.

A partir da Tabela 10, identifica-se a presença de 7 variáveis, sendo 2 qualitativas e 5 quantitativas. Sendo possível traçar um perfil do paciente com Mieloma múltiplo inserido no estudo. Os pacientes têm idade entre 50 e 77 anos, são predominantemente do sexo masculino e têm proteína de Bence-Jones presente na urina. Além disso o grupo pacientes, possui média: 34 mg/dL de níveis de Nitrogênio ureico no sangue; 10 mg/dL de Cálcio sérico; 10,3 g/dL de Hemoglobina e 43% de células plasmáticas em medula óssea.

Com o objetivo de investigar o efeito dos fatores de risco: Nitrogênio ureico, Cálcio sérico, Hemoglobina, Células plasmáticas e Proteína no tempo de sobrevivência dos pacientes

Tabela 10 – Características dos pacientes dos dados de Mieloma Múltiplo.

Variáveis Qualitativas			
Variável	Categoria	Quantidade	%
Sexo	Feminino	19	39,58
	Masculino	29	60,42
Proteína	Ausente	33	68,75
	Presente	15	31,25
Variáveis Quantitativas			
Variável	Mínimo	Máximo	Média
Idade	50	77	63
Nitrogênio ureico (Bun)	6	172	34
Cálcio sérico (Ca)	8	15	10
Hemoglobina (Hb)	4,9	14,6	10,3
Células plasmáticas (Pcells)	3	100	43

Fonte: Elaborada pelo autor.

com mieloma múltiplo, sendo que os efeitos desses fatores de risco podem ser modificados pela idade ou sexo de um paciente a significância das variáveis da presente aplicação deverá ser avaliada.

Assim, foi realizado ajuste dos Modelos de Regressão: Chances de Sobrevivência Proporcionais Weibull Discreto (MRCSP-WD) e Weibull Discreto (MRWD) múltiplos aos dados, com o intuito de verificar a significância de suas covariáveis. Os resultados de valor-p dos modelos múltiplos são apresentados na Tabela 11.

Tabela 11 – Teste de significância das covariáveis para MRCSP-WD e MRWD múltiplos (Dados de Mieloma).

MRCSP-WD			MRWD		
Variável	Parâmetro	valor-p	Variável	Parâmetro	valor-p
Idade	β_1	0,62201	Idade	β_1	0,54852
Sexo	β_2	0,68698	Sexo	β_2	0,95182
Nitrogênio ureico	β_3	0,00055	Nitrogênio ureico	β_3	0,00055
Cálcio sérico	β_4	0,63145	Cálcio sérico	β_4	0,81627
Hemoglobina	β_5	0,12828	Hemoglobina	β_5	0,11161
Células plasmáticas	β_6	0,44708	Células plasmáticas	β_6	0,86859
Proteína	β_7	0,04921	Proteína	β_7	0,09728

Nível de referência das variáveis qualitativas: Sexo = Feminino; Proteína = Ausente.

Fonte: Elaborada pelo autor.

De acordo com os dados da Tabela 11, a covariável Nitrogênio ureico é significativa (ao nível de significância de 5%) em ambos os modelos. A covariável Proteína por sua vez, é significativa levando em conta o nível de significância supracitado para o MRCSP-WD e significativa ao nível de 10% para o MRWD.

Por outro lado, as demais covariáveis não são significativas e não estariam influenciando

na sobrevivência dos pacientes. Desse modo, foi realizado o teste de significância ajustando-se apenas as covariáveis significativas e os resultados são apresentados na Tabela 12.

Tabela 12 – Teste de significância das covariáveis para MRCSP-WD e MRWD para as covariáveis significativas (Dados de Mieloma).

MRCSP-WD			MRWD		
Variável	Parâmetro	valor-p	Variável	Parâmetro	valor-p
Nitrogênio ureico	β_1	0,00079	Nitrogênio ureico	β_1	0,00024
Proteína	β_2	0,02315	Proteína	β_2	0,03729

Nível de referência da variável qualitativa: Proteína = Ausente.

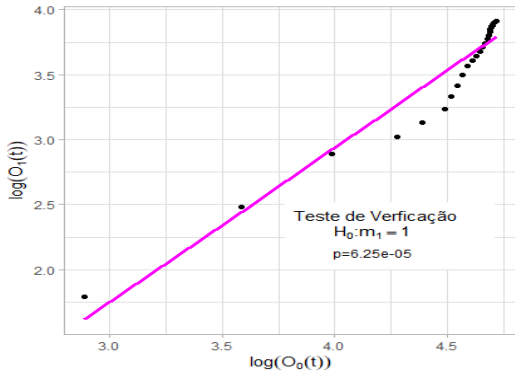
Fonte: Elaborada pelo autor.

Por meio da Tabela 12, de fato comprova-se que as covariáveis Nitrogênio ureico e Proteína influenciam a sobrevivência dos pacientes com Mieloma múltiplo (valor-p < nível de significância = 0,05) em ambos os modelos. Desse modo, levando em conta as referidas covariáveis, na Subseção 5.2.1 será realizado o estudo de verificação de suposição de chances de sobrevivência proporcionais. Em seguida, será apresentada as estimativas do MRCSP-WD na Subseção 5.2.2, tal como o ajuste do referido modelo aos dados na Subseção 5.2.3.

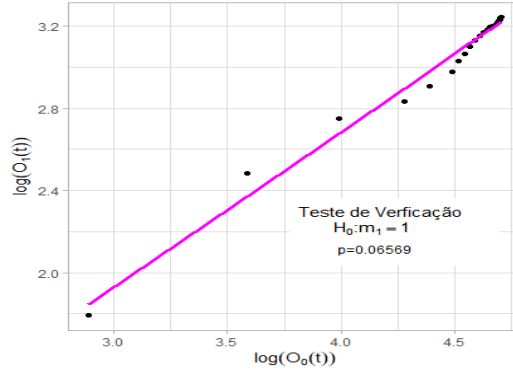
5.2.1 Verificação de suposição de chances de sobrevivência proporcionais (dados Mieloma)

A fim de verificar a suposição de chances de sobrevivência proporcionais nos diferentes níveis das covariáveis e visto que a covariável Nitrogênio ureico é quantitativa, é necessário realizar uma categorização dessa covariável. O Nitrogênio ureico no sangue (Bun) é um exame que determina a quantidade de nitrogênio no sangue sob a forma de ureia, sendo um preditor útil de hemorragia digestiva alta. Seguindo Tomizawa *et al.* (2015), os resultados de Bun dos pacientes para a verificação de suposição de chances de sobrevivência proporcionais e ajuste tratado na Subseção 5.2.3, foram categorizados em dois grupos (Menor ou igual a 21 mg/dL = 0; Maior que 21 mg/dL = 1).

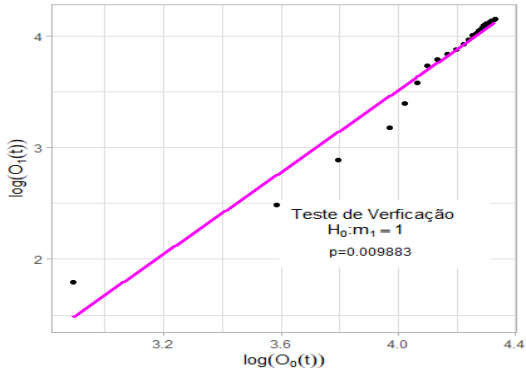
Desse modo, a suposição de chances de sobrevivência proporcionais foi verificada para o conjunto de dados, observando a referida proporcionalidade entre os níveis da covariável Bun e da covariável Proteína e para cada um dos níveis dessas duas covariáveis, por meio do gráfico: $\log(O_0(t)) \times \log(O_1(t))$ e o teste de hipótese proposto em 3.23. Visto que foram realizados consecutivos testes, será utilizado a correção de Bonferroni, mais detalhes podem ser encontrados em Montgomery, Peck e Vining (2021), visando corrigir a probabilidade de rejeitar incorretamente a hipótese nula, uma vez que serão realizados 8 testes ao todo, o nível de significância, passa a ser $0,05/8 = 0,00625$. Os resultados são apresentados na Figura 30, admitindo-se: $z_1 =$ nível da covariável Bun ($z_1 = 0$: Nitrogênio ureico menor ou igual a 21 mg/dL; $z_1 = 1$: Nitrogênio ureico maior que 21 mg/dL) e $z_2 =$ nível da covariável Proteína ($z_2 = 0$: Proteína de Bence-Jones ausente; $z_2 = 1$: Proteína de Bence-Jones presente).



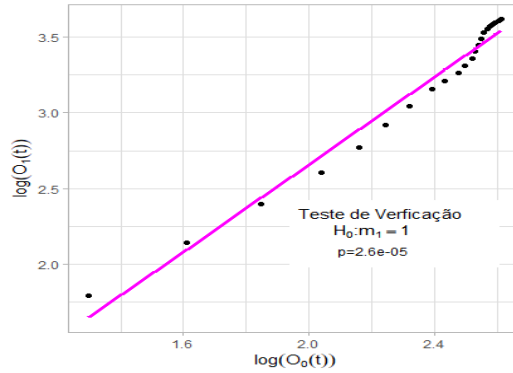
(a) $(z_1 = 0; z_2 = 0) \times (z_1 = 0; z_2 = 1)$.



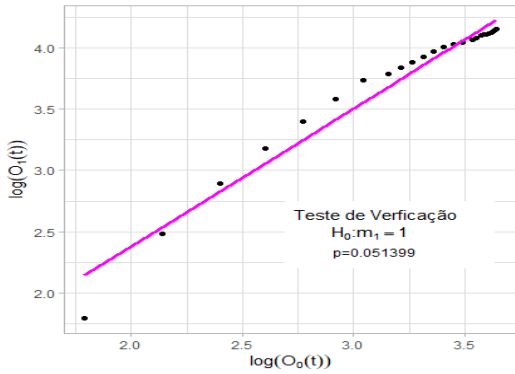
(b) $(z_1 = 0; z_2 = 0) \times (z_1 = 1; z_2 = 0)$.



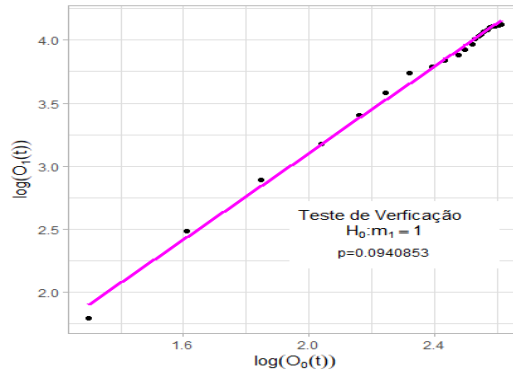
(c) $(z_1 = 0; z_2 = 0) \times (z_1 = 1; z_2 = 1)$.



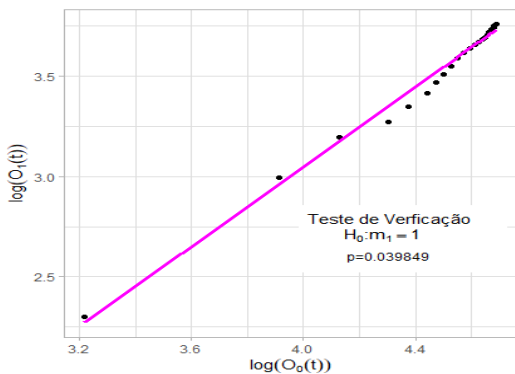
(d) $(z_1 = 0; z_2 = 1) \times (z_1 = 1; z_2 = 0)$.



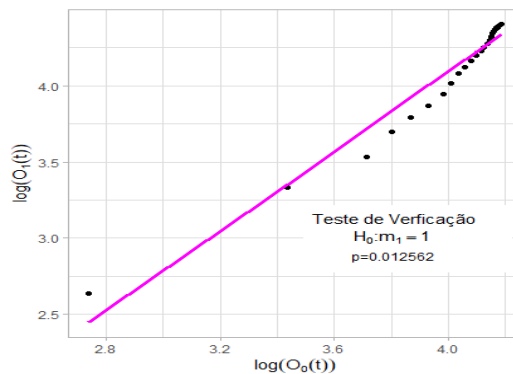
(e) $(z_1 = 0; z_2 = 1) \times (z_1 = 1; z_2 = 1)$.



(f) $(z_1 = 1; z_2 = 0) \times (z_1 = 1; z_2 = 1)$.



(g) Nitrogênio ureico.



(h) Proteína.

Figura 30 – Verificação de suposição de chances de sobrevivência proporcionais para as covariáveis Bun (z_1) e Proteína (z_2).

Fonte: Elaborada pelo autor.

Por meio dos resultados dos testes apresentados na Figura 30 nota-se que a suposição de chances de sobrevivência proporcionais não foi rejeitada para 06 dos 08 níveis de covariáveis consideradas neste estudo (tendo em vista um nível de significância de 0,625%). Os gráficos $\log(O_0(t)) \times \log(O_1(t))$ dos níveis das covariáveis dois a dois (Figura 30a a 30f) e entre os níveis da covariável Nitrogênio ureico (Figura 30g) e da covariável Proteína (Figura 30h), corroboram para tal afirmação. Note que nas Figuras pelas quais a suposição de chances de sobrevivência proporcionais não foi rejeitada nos testes de hipóteses os pontos se aproximam da reta de regressão ajustada.

Devido ao fato da maioria dos níveis estudados dois-a-dois possuírem chances de sobrevivência proporcionais, é um forte indicativo que os dados em estudo apresentam chances de sobrevivência proporcionais o que justifica a utilização de tal metodologia nesse conjunto de dados, que é apresentada na próxima subseção.

5.2.2 Modelo de regressão chances de sobrevivência proporcionais (MRCSP)

A verificação da suposição de chances de sobrevivência proporcionais realizada na subseção anterior indicou haver evidências que o modelo de regressão com tal característica mostra-se adequado para estimação dos parâmetros do conjunto de dados em estudo. Assim, foram calculadas as estimativas pontuais e intervalares dos parâmetros para o MRCSP considerando como distribuição base a distribuição Weibull Discreta (WD).

Desta forma, para o modelo MRCSP-WD conjunto, considerando as duas covariáveis significativas expostas nas subseções anteriores, foram calculadas as estimativas pontuais e intervalares de seus parâmetros que são apresentados na Tabela 13.

Tabela 13 – Estimativas dos parâmetros do MRCSP-WD conjunto - covariáveis significativas (Dados mieloma).

Variável	Parâmetro	Estimativa	Erro Padrão	IC (95%)	valor-p
-	q	0,9834	0,0104	[0,9631; 1,0037]	-
-	η	1,1481	0,1526	[0,8487; 1,4469]	-
Nitrogênio ureico	β_1	-0,0248	0,0074	[-0,0393; -0,0103]	0,0008
Proteína	β_2	1,4210	0,6258	[0,1945; 2,6475]	0,0232

Nível de referência das variáveis: Nitrogênio ureico= Menor ou igual a 21 mg/dL; Proteína = Ausente.

Fonte: Elaborada pelo autor.

Por meio das estimativas da Tabela 13 comprova-se a significância das covariáveis, uma vez que o intervalo de confiança dos parâmetros relacionados às covariáveis (β_1 e β_2) não contém o valor zero.

5.2.3 Verificação do ajuste do MRCSP-WD e comparação com o Modelo de Regressão Weibull Discreto (MRWD)

Por sua vez, para o modelo MRWD conjunto, considerando as duas covariáveis significativas expostas nas subseções anteriores, foram calculadas as estimativas pontuais e intervalares de seus parâmetros que são apresentados na Tabela 14.

Tabela 14 – Estimativas dos parâmetros do MRWD conjunto - covariáveis significativas (Dados mieloma).

Variável	Parâmetro	Estimativa	Erro Padrão	IC (95%)	valor-p
-	η	1,0903	0,1495	[0,7973; 1,3832]	-
-	β_0	4,0137	0,6199	[2,7987; 5,2287]	-
Nitrogênio ureico	β_1	-0,0218	0,0059	[-0,0334; -0,0102]	0,0002
Proteína	β_2	0,80005	0,3842	[0,0471; 1,5530]	0,0373

Nível de referência das variáveis: Nitrogênio ureico= Menor ou igual a 21 mg/dL; Proteína = Ausente.

Fonte: Elaborada pelo autor.

A exemplo do que ocorre para o MRCSP-WD, por meio das estimativas da Tabela 14 comprova-se a significância das covariáveis para o MRWD, uma vez que os intervalos de confiança dos parâmetros relacionados às covariáveis (β_1 e β_2) não contêm o valor zero.

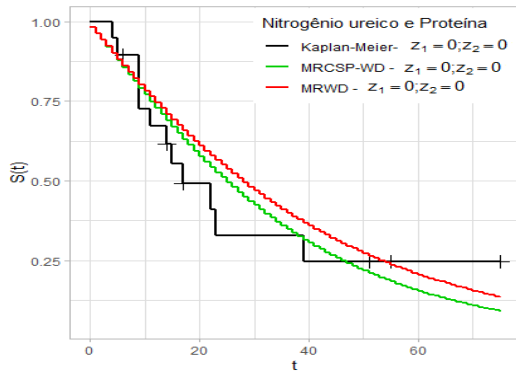
Desse modo, levando em conta a categorização das covariável Nitrogênio ureico no sangue (Bun) realizado na Subseção 5.2.1, para avaliação de ajuste dos modelos aos dados a título de comparação com o modelo mais simples, foram esboçados os gráficos de sobrevivência do estimador de Kaplan-Meier (K-M) e das sobrevivências dos modelos em estudo para cada um dos níveis das covariáveis, com intuito de analisar o conjunto de gráficos e interpretar o seu ajuste de maneira global (Figura 31). Além disso, foram calculados os erros máximos cometidos na estimação, para cada um dos níveis de covariáveis para os dois modelos em estudo que são apresentados na Tabela 15. Assim como na Subseção 5.2.1, os resultados são apresentados admitindo-se: $z_1 =$ nível da covariável Bun ($z_1 = 0$: Nitrogênio ureico menor ou igual a 21 mg/dL; $z_1 = 1$: Nitrogênio ureico maior que 21 mg/dL) e $z_2 =$ nível da covariável Proteína ($z_2 = 0$: Proteína de Bence-Jones ausente; $z_2 = 1$: Proteína de Bence-Jones presente).

Tabela 15 – Erros máximos oriundos da estimação dos modelos (Dados de Mieloma).

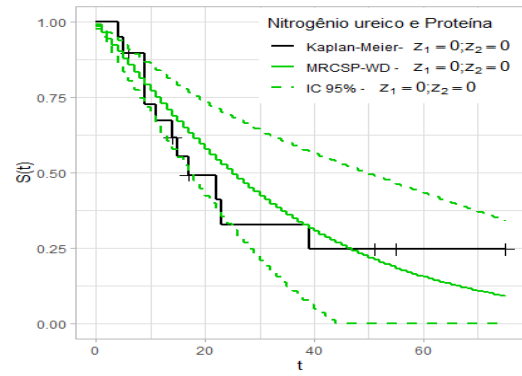
Nível da Covariável	$\epsilon_{MRCSP-WD}$	ϵ_{MRWD}
$z_1=0 ; z_2=0$	0,1982	0,2360
$z_1=0 ; z_2=1$	0,3045	0,2719
$z_1=1 ; z_2=0$	0,1306	0,2482
$z_1=1 ; z_2=1$	0,2891	0,3053

Fonte: Elaborada pelo autor.

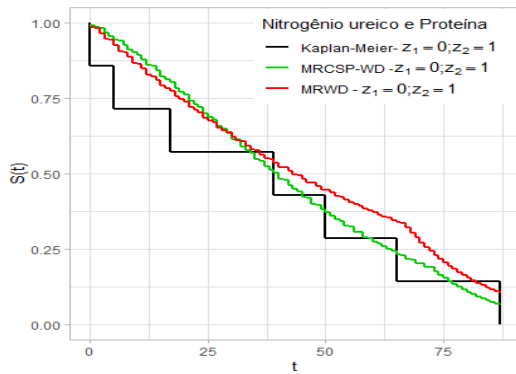
Note, por meio das Figuras 31a, 31c, 31e e 31g aliadas aos resultados da Tabela 15, que ambos os modelos possuem um bom ajuste aos dados, estando as estimativas de sobrevivência dos referidos modelos sempre próximas às estimativas empíricas.



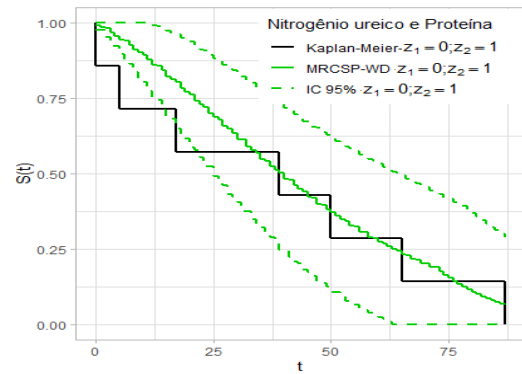
(a) MRCSP-WD e MRWD: $z_1=0 ; z_2=0$.



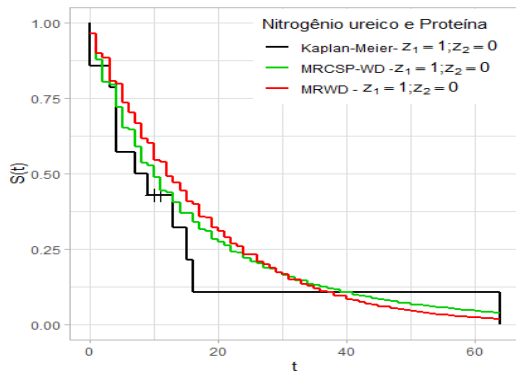
(b) MRCSP-WD com respectivo IC 95%: $z_1=0 ; z_2=0$.



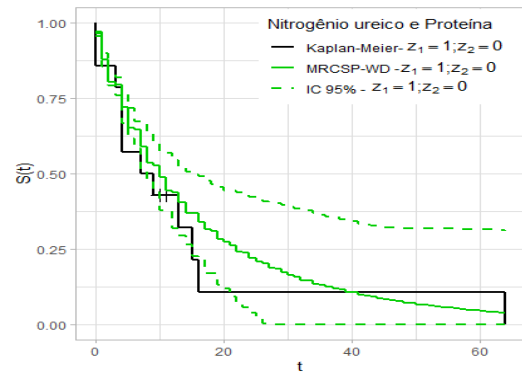
(c) MRCSP-WD e MRWD: $z_1=0 ; z_2=1$.



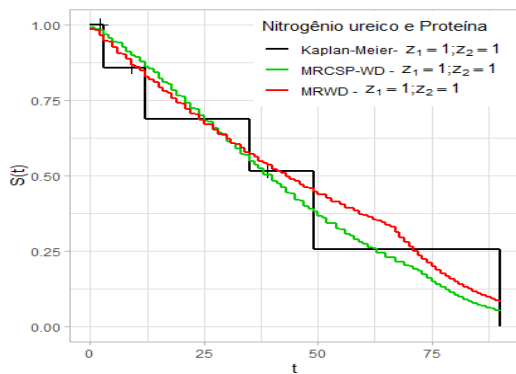
(d) MRCSP-WD com respectivo IC 95%: $z_1=0 ; z_2=1$.



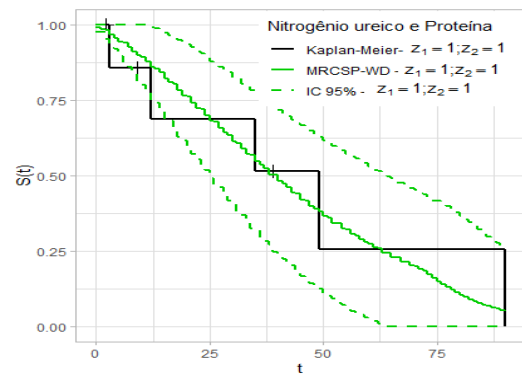
(e) MRCSP-WD e MRWD: $z_1=1 ; z_2=0$.



(f) MRCSP-WD com respectivo IC 95%: $z_1=1 ; z_2=0$.



(g) MRCSP-WD e MRWD: $z_1=1 ; z_2=1$.



(h) MRCSP-WD com respectivo IC 95%: $z_1=1 ; z_2=1$.

Figura 31 – Ajuste dos modelos MRCSP-WD e MRWD aos dados de Mieloma por nível das covariáveis.

Fonte: Elaborada pelo autor.

Para comprovação do bom ajuste do MRCSP-WD aos dados, nas Figuras 31b, 31d, 31f e 31h, são apresentadas as estimativas da função de sobrevivência com os respectivos IC com 95% de confiança para as sobrevivências de cada um dos níveis (z_1 = nível da covariável Bun; z_2 = nível da covariável Proteína) estimados via método delta com as estimativas de sobrevivência via estimador de Kaplan-Meier. Note que a maioria das estimativas empíricas encontram-se dentro do IC corroborando o ótimo ajuste do MRCSP-WD aos dados da aplicação dos dados de Mieloma.

CONSIDERAÇÕES FINAIS

O Modelo de Chances de Sobrevivência Proporcional (MCSP) proposto por [Bennett](#), no qual as covariáveis têm um efeito multiplicativo na chance de sobrevivência, é um modelo de regressão em que as funções de risco separadas de pacientes convergem com o tempo, sendo um concorrente em potencial do popular Modelo de riscos proporcionais de [Cox](#) em que as funções de risco possuem uma razão constante entre si.

Desse modo, o MCSP possui a interessante propriedade de cura eficaz em que a taxa de falha de um grupo de pessoas doentes se aproximaria da taxa de falha de um grupo de controle conforme o tempo passa. Apesar do extensivo uso desse modelo em aplicações, a sua utilização em um contexto de análise de sobrevivência geralmente ocorre para variável resposta (tempo até ocorrência do evento de interesse) contínua.

No entanto, na prática, as medições de tempo costumam ser discretas, havendo situações em que a hora exata do evento pode não ser conhecida, mas apenas o intervalo durante o qual ocorreu o evento de interesse. A utilização de métodos estatísticos que são especialmente concebidos para tempos discretos possui vantagens do ponto de vista teórico e prático.

Assim, a motivação deste trabalho surgiu do interesse de utilizar o MCSP que contém propriedades atrativas no contexto de análise de sobrevivência e que permite a presença de covariáveis por meio de uma relação entre estas e os tempos discretos de sobrevivência. Nesse contexto, considerou-se um processo inferencial paramétrico, tendo como distribuição base a WD.

O processo inferencial do Modelo de Regressão Chances de Sobrevivência Proporcional Weibull Discreto (MRCSP-WD), considerando-se tempos de sobrevivência discretos, foi desenvolvido a partir do método de máxima verossimilhança utilizando métodos computacionais de otimização para a obtenção das estimativas. Outra importante contribuição do ponto de vista teórico desse trabalho, consistiu em estabelecer critérios para a utilização de tal modelo em situações práticas.

No Capítulo 4, foram descritas as simulações computacionais realizadas via *software* R, os tempos de sobrevivência foram gerados considerando três cenários definidos a partir da taxa de falha da distribuição base a depender de seu parâmetro de forma e considerando duas covariáveis (uma dicotômica e uma numérica). O viés, EQM e PC para cada parâmetro da distribuição e dos coeficientes de regressão associados às referidas covariáveis para os três cenários foram calculados para diferentes tamanhos de amostra e percentuais de censura.

Na ausência de censuras atestou-se que os estimadores eram aproximadamente não viesados e consistentes. Enquanto que na presença de observações censuradas, à medida que essa proporção aumenta, há maiores desvios das estimativas em relação ao verdadeiro valor do parâmetro, sendo uma característica intrínseca destas estimativas, visto que o verdadeiro valor do parâmetro não deve ser utilizado como referência, pois a função distribuição acumulada utilizada para gerar os dados não considera a parte censurada. É importante destacar que independente do percentual de censura e tamanho amostral a normalidade dos estimadores estudados foi atendida.

As aplicações apresentadas no Capítulo 5 permitiram ilustrar o uso do MRCSP-WD em dados reais. Em ambas as aplicações foram realizados testes de significâncias das covariáveis de cada aplicação, a fim de examinar quais destas influenciavam a sobrevivência dos pacientes. Aliado aos testes supracitados, foi realizada a verificação de suposição de chances de sobrevivência proporcionais das covariáveis significativas, como critério das covariáveis que seriam utilizadas no modelo em estudo.

Para a primeira aplicação referente ao tempo de sobrevivência em semanas de pacientes diagnosticados com AML, foi considerada apenas covariável idade sendo esta dicotômica (pacientes com: 50 anos ou mais; menos de 50 anos), o MRCSP-WD apresentou um bom ajuste ao dados, sendo possível a interpretação de que a chance de sobrevivência do paciente com 50 anos ou mais é menor do que a chance de sobrevivência do paciente com menos de 50 anos.

A segunda aplicação trata-se do tempo de sobrevivência em meses de pacientes com mieloma múltiplo, que contava com duas covariáveis significativas: Nitrogênio ureico no sangue (numérica) e Proteína (dicotômica). A validade de chances de sobrevivência proporcionais foi comprovada avaliando os níveis de covariáveis dois-a-dois e as estimativas do MRCSP-WD apresentadas demonstram ser possível a estimação dos coeficientes de regressão tanto para variáveis qualitativas quanto para variáveis quantitativas, devido a um bom ajuste do modelo aos dados, após ser analisado o ajuste em cada nível de covariável.

De maneira geral, o Modelo de Regressão de Chances de Sobrevivência Proporcional para tempos de sobrevivência discretos mediante ao processo inferencial e a verificação viável de seu uso desenvolvidos no presente trabalho aliado as simulações computacionais que comprovaram a existência de normalidade assintótica dos estimadores possibilitaram explorar tal modelo em situações práticas em que foi possível verificar seu bom ajuste e fazer interpretações das covariáveis em relação aos tempos discretos de sobrevivência.

REFERÊNCIAS

- BARRIGA, G. D.; CANCHO, V. G.; GARIBAY, D. V.; CORDEIRO, G. M.; ORTEGA, E. M. A new survival model with surviving fraction: An application to colorectal cancer data. **Statistical methods in medical research**, SAGE Publications Sage UK: London, England, v. 28, n. 9, p. 2665–2680, 2019. Citado na página 49.
- BENNETT, S. Analysis of survival data by the proportional odds model. **Statistics in medicine**, Wiley Online Library, v. 2, n. 2, p. 273–277, 1983. Citado nas páginas 24, 35, 36, 38, 40 e 81.
- BERGER, M.; SCHMID, M. Semiparametric regression for discrete time-to-event data. **Statistical Modelling**, SAGE Publications Sage India: New Delhi, India, v. 18, n. 3-4, p. 322–345, 2018. Citado nas páginas 24, 30, 31 e 36.
- BERGER, M.; WELCHOWSKI, T.; SCHMITZ-VALCKENBERG, S.; SCHMID, M. A classification tree approach for the modeling of competing risks in discrete time. **Advances in Data Analysis and Classification**, Springer, v. 13, n. 4, p. 965–990, 2019. Citado nas páginas 24, 31 e 36.
- BRUNELLO, G. H. V.; NAKANO, E. Y. Inferência Bayesiana no Modelo Weibull Discreto em dados com Presença de Censura. **TEMA (São Carlos)**, SciELO Brasil, v. 16, p. 97–110, 2015. Citado nas páginas 30 e 32.
- CARDIAL, M. R. P.; FACHINI-GOMES, J. B.; NAKANO, E. Y. Exponentiated discrete Weibull distribution for censored data. **Brazilian Journal of Biometrics**, v. 38, n. 1, p. 35–56, 2020. Citado nas páginas 23, 30 e 32.
- CHAKRABORTY, S. Generating discrete analogues of continuous probability distributions-A survey of methods and constructions. **Journal of Statistical Distributions and Applications**, Springer, v. 2, n. 1, p. 1–30, 2015. Citado nas páginas 23 e 32.
- COLLETT, D. **Modelling Survival Data in Medical Research**. [S.l.]: CRC press, 2015. Citado nas páginas 71 e 72.
- COLOSIMO, E. A.; GIOLO, S. R. **Análise de sobrevivência aplicada**. [S.l.]: Editora Blucher, 2006. Citado nas páginas 25, 27 e 37.
- CORDEIRO, G. M.; LEMONTE, A. J.; ORTEGA, E. M. The Marshall-Olkin Family of Distributions: Mathematical Properties and New Models. **Journal of Statistical Theory and Practice**, Springer, v. 8, n. 2, p. 343–366, 2014. Citado na página 37.
- COX, D. R. Regression Models and Life-Tables. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 34, n. 2, p. 187–202, 1972. Citado nas páginas 24, 35, 36, 38 e 81.
- ELEUTERI, A.; AUNG, M.; TAKTAK, A.; DAMATO, B.; LISBOA, P. Continuous and Discrete Time Survival Analysis: Neural Network Approaches. In: IEEE. **2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society**. [S.l.], 2007. p. 5420–5423. Citado na página 35.

- HA, I. D.; MACKENZIE, G. Robust frailty modelling using non-proportional hazards models. **Statistical modelling**, SAGE Publications Sage India: New Delhi, India, v. 10, n. 3, p. 315–332, 2010. Citado na página 49.
- JAYAKUMAR, K.; BABU, M. G. Discrete Weibull geometric distribution and its properties. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 47, n. 7, p. 1767–1783, 2018. Citado nas páginas 23 e 32.
- KAPLAN, E. L.; MEIER, P. Nonparametric Estimation from Incomplete Observations. **Journal of the American statistical association**, Taylor & Francis, v. 53, n. 282, p. 457–481, 1958. Citado na página 27.
- KRISHNA, E.; JOSE, K.; ALICE, T.; RISTIĆ, M. M. The Marshall-Olkin Fréchet Distribution. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 42, n. 22, p. 4091–4107, 2013. Citado na página 37.
- LEE, E. T.; WANG, J. **Statistical Methods for Survival Data Analysis**. [S.l.]: John Wiley & Sons, 2003. v. 476. Citado nas páginas 65 e 66.
- LEHMANN, E. L.; CASELLA, G. **Theory of Point Estimation**. [S.l.]: Springer Science & Business Media, 2006. Citado na página 29.
- MAGALHÃES, M. N. **Probabilidade e Variáveis Aleatórias**. [S.l.]: Edusp, 2006. Citado na página 26.
- MARSHALL, A. W.; OLKIN, I. A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. **Biometrika**, Oxford University Press, v. 84, n. 3, p. 641–652, 1997. Citado na página 37.
- MIGON, H. S.; GAMERMAN, D.; LOUZADA, F. **Statistical Inference: An Integrated Approach**. [S.l.]: CRC press, 2014. Citado na página 27.
- MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to linear regression analysis**. [S.l.]: John Wiley & Sons, 2021. Citado na página 75.
- MURPHY, S.; ROSSINI, A.; VAART, A. W. van der. Maximum Likelihood Estimation in the Proportional Odds Model. **Journal of the American Statistical Association**, Taylor & Francis Group, v. 92, n. 439, p. 968–976, 1997. Citado nas páginas 24, 36 e 38.
- NAKAGAWA, T.; OSAKI, S. The Discrete Weibull Distribution. **IEEE transactions on reliability**, IEEE, v. 24, n. 5, p. 300–301, 1975. Citado nas páginas 23, 32 e 33.
- NAKANO, E. Y.; CARRASCO, C. G. Uma Avaliação do Uso de um Modelo Contínuo na Análise de Dados Discretos de Sobrevida. **TEMA (São Carlos)**, v. 7, n. 1, p. 91–100, 2006. Citado nas páginas 23, 31 e 70.
- OLIVEIRA, F. A. P. d. **Procedimentos de geração de dados de sobrevivência com censura à direita**. Dissertação (Mestrado) — Universidade de Brasília, 2021. Citado na página 49.
- ORTEGA, E. M.; CORDEIRO, G. M.; CAMPELO, A. K.; KATTAN, M. W.; CANCHO, V. G. A power series beta Weibull regression model for predicting breast carcinoma. **Statistics in medicine**, Wiley Online Library, v. 34, n. 8, p. 1366–1388, 2015. Citado na página 49.

PADGETT, W.; SPURRIER, J. D. On discrete failure models. **IEEE Transactions on Reliability**, IEEE, v. 34, n. 3, p. 253–256, 1985. Citado na página 33.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2019. Disponível em: <<https://www.R-project.org/>>. Citado nas páginas 43 e 49.

ROYSTON, P.; PARMAR, M. K. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. **Statistics in medicine**, Wiley Online Library, v. 21, n. 15, p. 2175–2197, 2002. Citado nas páginas 24 e 36.

SARHAN, A. M. A two-parameter discrete distribution with a bathtub hazard shape. **Communications for Statistical Applications and Methods**, The Korean Statistical Society, v. 24, n. 1, p. 15–27, 2017. Citado nas páginas 23 e 32.

STEIN, W. E.; DATTERO, R. A new discrete Weibull distribution. **IEEE Transactions on Reliability**, IEEE, v. 33, n. 2, p. 196–197, 1984. Citado na página 33.

TEIXEIRA, J. C. d. S. **Testes de superioridade para modelos de chances proporcionais com e sem fração de cura**. Dissertação (Mestrado) — Universidade Federal de São Carlos, 2017. Citado na página 37.

TOMIZAWA, M.; SHINOZAKI, F.; HASEGAWA, R.; SHIRAI, Y.; MOTOYOSHI, Y.; SUGIYAMA, T.; YAMAMOTO, S.; ISHIGE, N. Patient characteristics with high or low blood urea nitrogen in upper gastrointestinal bleeding. **World Journal of Gastroenterology: WJG**, Baishideng Publishing Group Inc, v. 21, n. 24, p. 7500, 2015. Citado na página 75.

TUTZ, G.; SCHMID, M. **Modeling Discrete Time-to-Event Data**. [S.l.]: Springer, 2016. Citado nas páginas 23, 24, 27, 30 e 31.

VALLEJOS, C. A.; STEEL, M. F. Bayesian survival modelling of university outcomes. **Journal of the Royal Statistical Society: Series A (Statistics in Society)**, Wiley Online Library, v. 180, n. 2, p. 613–631, 2017. Citado nas páginas 24 e 31.

VIEIRA, M. G. F. **Modelo de regressão odds-riscos proporcionais para dados de sobrevivência discretos**. Dissertação (Mestrado) — Universidade de Brasília, 2018. Citado nas páginas 24, 32 e 36.

VILA, R.; NAKANO, E. Y.; SAULO, H. Theoretical results on the discrete Weibull distribution of Nakagawa and Osaki. **Statistics**, Taylor & Francis, v. 53, n. 2, p. 339–363, 2018. Citado nas páginas 23 e 33.

WANG, L.; WANG, L. An EM algorithm for analyzing right-censored survival data under the semiparametric proportional odds model. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 51, n. 15, p. 5284–5297, 2020. Citado nas páginas 24 e 36.

YANG, S.; PRENTICE, R. L. Semiparametric Inference in the Proportional Odds Regression Model. **Journal of the American Statistical Association**, Taylor & Francis Group, v. 94, n. 445, p. 125–136, 1999. Citado nas páginas 24, 36 e 37.

ZHOU, J.; ZHANG, J.; LU, W. TransModel: An R Package for Linear Transformation Model with Censored Data. **Journal of Statistical Software**, v. 101, p. 1–12, 2022. Citado nas páginas 24 e 36.

MCSP COMO MTL

Seja o Modelo de Transformação Linear dado como em (3.4), tem-se que:

$$-\varepsilon = -G(t) + \mathbf{z}'\boldsymbol{\beta}.$$

Ao aplicar $\exp\{\cdot\}$ em ambos os lados da igualdade, fica:

$$\exp\{-\varepsilon\} = \exp\{-G(t)\} \exp\{\mathbf{z}'\boldsymbol{\beta}\}.$$

Ao assumir que ε segue uma distribuição Logística padrão. Portanto:

$$F(\varepsilon) = \frac{1}{1 + \exp\{-\varepsilon\}} \Rightarrow S(\varepsilon) = \frac{\exp\{-\varepsilon\}}{1 + \exp\{-\varepsilon\}}. \quad (\text{A.1})$$

Ao considerar, z uma covariável dicotômica que assume os valores 1 e 0, para os tratamentos teste e base respectivamente, a partir de (A.1), tem-se:

$$\begin{aligned} S(t; \boldsymbol{\beta} | \mathbf{z}) &= \frac{\exp\{-G(t)\} \exp\{\mathbf{z}'\boldsymbol{\beta}\}}{1 + \exp\{-G(t)\} \exp\{\mathbf{z}'\boldsymbol{\beta}\}} \\ \exp\{-G(t)\} &= \frac{S(t; \boldsymbol{\beta} | \mathbf{z})}{(1 - S(t; \boldsymbol{\beta} | \mathbf{z})) \exp\{\mathbf{z}'\boldsymbol{\beta}\}} \end{aligned} \quad (\text{A.2})$$

e

$$\begin{aligned} S_0(t) &= \frac{\exp\{-G(t)\}}{1 + \exp\{-G(t)\}} \\ \exp\{-G(t)\} &= \frac{S_0(t)}{(1 - S_0(t))}. \end{aligned} \quad (\text{A.3})$$

Ao igualar (A.2) e (A.3), tem-se:

$$\frac{S(t; \boldsymbol{\beta} | \mathbf{z})}{(1 - S(t; \boldsymbol{\beta} | \mathbf{z}))} = \exp\{\mathbf{z}'\boldsymbol{\beta}\} \frac{S_0(t)}{(1 - S_0(t))}. \quad (\text{A.4})$$

FUNÇÕES EM ANÁLISE DE SOBREVIVÊNCIA DO MCSP

B.1 Função de sobrevivência

A partir da expressão (3.1), tem-se que:

$$S(t|\mathbf{z})[1 - S_0(t)] = g(\mathbf{z}'\boldsymbol{\beta})S_0(t) - g(\mathbf{z}'\boldsymbol{\beta})S_0(t)S(t|\mathbf{z}).$$

Assim:

$$S(t|\mathbf{z})[1 + S_0(t)(g(\mathbf{z}'\boldsymbol{\beta}) - 1)] = g(\mathbf{z}'\boldsymbol{\beta})S_0(t).$$

Logo:

$$S(t|\mathbf{z}) = \frac{g(\mathbf{z}'\boldsymbol{\beta})S_0(t)}{1 + (g(\mathbf{z}'\boldsymbol{\beta}) - 1)S_0(t)}.$$

B.2 Função de risco (caso contínuo)

Sejam $S(\cdot)$ e $h(\cdot)$, a função de sobrevivência e de risco respectivamente, para T uma variável aleatória contínua e não negativa, uma relação matemática em análise de sobrevivência que associa as referidas funções é dada por:

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

A partir da expressão (3.5), tem-se que:

$$\log S(t|\mathbf{z}) = \log(g(\mathbf{z}'\boldsymbol{\beta})) + \log S_0(t) - \log[1 + (g(\mathbf{z}'\boldsymbol{\beta}) - 1)S_0(t)].$$

Assim:

$$\begin{aligned}
-\frac{d}{dt}(\log S(t|\mathbf{z})) &= -\left(\frac{\frac{d}{dt}S_0(t)}{S_0(t)} - \frac{(\frac{d}{dt}S_0(t)) [g(\mathbf{z}'\boldsymbol{\beta}) - 1]}{1 + (g(\mathbf{z}'\boldsymbol{\beta}) - 1)S_0(t)}\right) \\
&= -\frac{\frac{d}{dt}S_0(t)}{S_0(t)} \left(1 - \frac{S_0(t)[g(\mathbf{z}'\boldsymbol{\beta}) - 1]}{1 + (g(\mathbf{z}'\boldsymbol{\beta}) - 1)S_0(t)}\right) \\
&= -\frac{\frac{d}{dt}S_0(t)}{S_0(t)} \left(\frac{1}{1 + (g(\mathbf{z}'\boldsymbol{\beta}) - 1)S_0(t)}\right). \tag{B.2}
\end{aligned}$$

A saber, $\frac{\frac{d}{dt}S_0(t)}{S_0(t)} = \frac{d}{dt}(\log S_0(t))$. Portanto, pela relação (B.1), estabelece-se que:

$$h_0(t) = -\frac{d}{dt}(\log S_0(t)) \tag{B.3}$$

e

$$h_c(t|\mathbf{z}) = -\frac{d}{dt}(\log S(t|\mathbf{z})), \tag{B.4}$$

em que, $h_0(t)$ é a função de risco da distribuição base.

Logo, de (B.3) e (B.4) em (B.2), tem-se:

$$h_c(t|\mathbf{z}) = \left[\frac{h_0(t)}{1 + (g(\mathbf{z}'\boldsymbol{\beta}) - 1)S_0(t)}\right] \mathbb{I}_{\{t \geq 0\}}.$$

B.3 Função de probabilidade (caso discreto)

A partir função definida em (3.5) e da relação (2.4), é possível obter a função de probabilidade do MCSP para o caso discreto dada por:

$$p(t|\mathbf{z}) = \begin{cases} 1 - \frac{g(\mathbf{z}'\boldsymbol{\beta})S_0(0)}{1 + (g(\mathbf{z}'\boldsymbol{\beta}) - 1)S_0(0)}, & \text{se } t = 0 \\ \frac{g(\mathbf{z}'\boldsymbol{\beta})S_0(t-1)}{1 + (g(\mathbf{z}'\boldsymbol{\beta}) - 1)S_0(t-1)} - \frac{g(\mathbf{z}'\boldsymbol{\beta})S_0(t)}{1 + (g(\mathbf{z}'\boldsymbol{\beta}) - 1)S_0(t)}, & \text{se } t = 1, 2, \dots \end{cases}$$

Assim:

$$p(t|\mathbf{z}) = \begin{cases} \frac{1 + g(\mathbf{z}'\boldsymbol{\beta})S_0(0) - S_0(0) - g(\mathbf{z}'\boldsymbol{\beta})S_0(0)}{1 + (g(\mathbf{z}'\boldsymbol{\beta}) - 1)S_0(0)}, & \text{se } t = 0 \\ g(\mathbf{z}'\boldsymbol{\beta}) \left(\frac{S_0(t-1)}{1 + (g(\mathbf{z}'\boldsymbol{\beta}) - 1)S_0(t-1)} - \frac{S_0(t)}{1 + (g(\mathbf{z}'\boldsymbol{\beta}) - 1)S_0(t)} \right), & \text{se } t = 1, 2, \dots \end{cases}$$

Portanto:

$$p(t|\mathbf{z}) = \begin{cases} \frac{1 - S_0(0)}{1 + (g(\mathbf{z}'\boldsymbol{\beta}) - 1)S_0(0)}, & \text{se } t = 0 \\ g(\mathbf{z}'\boldsymbol{\beta}) \left(\frac{S_0(t-1)}{1 + (g(\mathbf{z}'\boldsymbol{\beta}) - 1)S_0(t-1)} - \frac{S_0(t)}{1 + (g(\mathbf{z}'\boldsymbol{\beta}) - 1)S_0(t)} \right), & \text{se } t = 1, 2, \dots \end{cases}$$

B.4 Função de risco (caso discreto)

A partir das funções de probabilidade (3.7) e sobrevivência (3.5), por meio da relação (2.1), a função de risco é expressa por:

$$h(t|\mathbf{z}) = \begin{cases} \frac{p(0|\mathbf{z})}{S(0|\mathbf{z}) + p(0|\mathbf{z})}, & \text{se } t = 0 \\ \frac{\frac{g(\mathbf{z}'\boldsymbol{\beta})S_0(t-1)}{1+(g(\mathbf{z}'\boldsymbol{\beta})-1)S_0(t-1)} - \frac{g(\mathbf{z}'\boldsymbol{\beta})S_0(t)}{1+(g(\mathbf{z}'\boldsymbol{\beta})-1)S_0(t)}}{\frac{g(\mathbf{z}'\boldsymbol{\beta})S_0(t)}{1+(g(\mathbf{z}'\boldsymbol{\beta})-1)S_0(t)} + \left(\frac{g(\mathbf{z}'\boldsymbol{\beta})S_0(t-1)}{1+(g(\mathbf{z}'\boldsymbol{\beta})-1)S_0(t-1)} - \frac{g(\mathbf{z}'\boldsymbol{\beta})S_0(t)}{1+(g(\mathbf{z}'\boldsymbol{\beta})-1)S_0(t)} \right)}, & \text{se } t = 1, 2, \dots \end{cases}$$

Considerando, $p(0|\mathbf{z})$ como obtido no Apêndice B.3 e $S(0|\mathbf{z})$ como obtido no Apêndice B.1, ambos para $t = 0$. A função $h(t|\mathbf{z})$ é reescrita como:

$$h(t|\mathbf{z}) = \begin{cases} \frac{\frac{1-S_0(0)}{1+(g(\mathbf{z}'\boldsymbol{\beta})-1)S_0(0)}}{\frac{g(\mathbf{z}'\boldsymbol{\beta})S_0(0)}{1+(g(\mathbf{z}'\boldsymbol{\beta})-1)S_0(0)} + \frac{1-S_0(0)}{1+(g(\mathbf{z}'\boldsymbol{\beta})-1)S_0(0)}}, & \text{se } t = 0 \\ 1 - \frac{g(\mathbf{z}'\boldsymbol{\beta})S_0(t)}{1+(g(\mathbf{z}'\boldsymbol{\beta})-1)S_0(t)} \frac{1+(g(\mathbf{z}'\boldsymbol{\beta})-1)S_0(t-1)}{g(\mathbf{z}'\boldsymbol{\beta})S_0(t-1)}, & \text{se } t = 1, 2, \dots \end{cases}$$

Assim:

$$h(t|\mathbf{z}) = \begin{cases} \frac{1-S_0(0)}{1+(g(\mathbf{z}'\boldsymbol{\beta})-1)S_0(0)}, & \text{se } t = 0 \\ 1 - \frac{S_0(t)(1+(g(\mathbf{z}'\boldsymbol{\beta})-1)S_0(t-1))}{S_0(t-1)(1+(g(\mathbf{z}'\boldsymbol{\beta})-1)S_0(t))}, & \text{se } t = 1, 2, \dots \end{cases}$$

Portanto:

$$h(t|\mathbf{z}) = \begin{cases} \frac{1-S_0(0)}{1+(g(\mathbf{z}'\boldsymbol{\beta})-1)S_0(0)}, & \text{se } t = 0 \\ 1 - \frac{S_0(t)(1+(g(\mathbf{z}'\boldsymbol{\beta})-1)S_0(t-1))}{S_0(t-1)(1+(g(\mathbf{z}'\boldsymbol{\beta})-1)S_0(t))}, & \text{se } t = 1, 2, \dots \end{cases}$$

SCRIPTS DESENVOLVIDOS

```

1: n=500                                # tamanho amostral, n=50,100,250,500
2: z1<-rbinom(n,1,0.5)                  # covariável 1
3: z2<-rnorm(n,0,1)                    # covariável 2
4: ##### Verdadeiros Valores dos Parâmetros
5: q.true=0.9
6: eta.true=1.5                         #C1=1.5; C2=1; C3=0.75
7: beta1.true<-2
8: beta2.true<-1
9:
10: ## Estimar os tempos relacionados com as covariáveis
11: ## determinar o parâmetro da Weibull Contínua (gamma)
12: gamma<-function(qwd,etawd){(-log(qwd))^(1/etawd)}
13: set.seed(2021)
14: rmrcspdweibull<-function(n, gamma, eta,beta1,beta2,z1,z2){
15:   u <- runif(n)
16:   s <- floor(gamma*(log1p((exp(beta1*z1+beta2*z2)-1)*u)-
17:     log(1 - u))^(1/eta) )
18:   return(s)}
19:
20: t<- rmrcspdweibull(n, gamma=gamma(q.true,eta.true), eta=eta.true,
21:   beta1=beta1.true,beta2=beta2.true,z1=z1,z2=z2)
22:
23: p.censura=0.00 # percentual de censura
24: censura<-rbinom(n,1,1-p.censura)
25:
26: #####
27: sob.base<-function(tempo,q,eta){q^((tempo+1)^eta)} # Sob. base WD
28:
29: sob.base.0<-function(tempo,q,eta){q^((tempo)^eta)} #Sob.base WD em (t-1)
30:
31: # função de ligação logaritmica
32: theta<-function(beta1,beta2){exp(beta1*z1 + beta2*z2)}

```

```

33:
34: # Função de Sobrevivência do MRCSP - WD
35: sob.mrcsp<-function(tempo,q,eta,theta){(sob.base(tempo,q,eta)*theta )/
36:   (1+(theta-1)*sob.base(tempo,q,eta))}
37:
38: # Função de Probabilidade do MRCSP - WD
39: prob.mrcsp<-function(tempo,q,eta,theta){
40:   ifelse(tempo > 0, (((sob.base.0(tempo,q,eta)*theta)
41:     /(1+(theta-1)*sob.base.0(tempo,q,eta)))-
42:     ((sob.base(tempo,q,eta)*theta )/
43:       (1+(theta-1)*sob.base(tempo,q,eta))))),
44:     (1-sob.base(tempo,q,eta))/(1+(theta-1)*sob.base(t,q,eta))) }
45:
46: # Função de log verossimilhança
47: log.vero.mrcsp <- function(parametros,tempo,z1,z2,censura){
48:   q<-parametros[1]
49:   eta <- parametros[2]
50:   beta1 <- parametros[3]
51:   beta2 <- parametros[4]
52:
53:   if((q > 0) && (q<1) && (eta >0))
54:
55:     return(-1*(sum(censura*log(prob.mrcsp(tempo,q,eta,theta(beta1,beta2)
56:   ))+
57:   (1-censura)* log(sob.mrcsp(tempo,q,eta,theta(beta1,beta2)))) ) ) )
58:   else return(-Inf)}
59: #####
60: M<-10500 # réplicas de monte carlo
61:
62: q.est<-eta.est<-b1.est<-b2.est<-conv<-desvpad1<-
63: desvpad2<-desvpad3<-desvpad4<-matrix(0,M,1) # matrizes para guardar
64: estim.
65: library(plyr)
66: progress_bar_text <- create_progress_bar("text")
67: progress_bar_text$init(M)
68: set.seed(2021)
69: for (i in 1:M){
70:   censura<-rbinom(n,1,1-p.censura)
71:   t<- rmrpcsdweibull(n, lambda=lambda(q.true,eta.true), eta=eta.true,
72:     beta1=beta1.true,beta2=beta2.true,z1=z1,z2=z2)
73:   v4 <- optim(c(q.true,eta.true,beta1.true,beta2.true),
74:     tempo=t,censura=censura,z1=z1,
75:     log.vero.mrcsp,NULL,hessian=T,method="Nelder-Mead")
76:   q.est[i,1]<-v4$par[1]
77:

```



```

78: eta.est[i,1] <-v4$par[2]
79:
80: b1.est[i,1] <-v4$par[3]
81:
82: b2.est[i,1] <-v4$par[4]
83:
84: conv[i,1] <-v4$convergence
85:
86: var.sim<-solve(v4$hessian)
87:
88: desvpad1[i,1] <-sqrt(var.sim[1,1])
89:
90: desvpad2[i,1] <-sqrt(var.sim[2,2])
91:
92: desvpad3[i,1] <-sqrt(var.sim[3,3])
93:
94: desvpad4[i,1] <-sqrt(var.sim[4,4])
95:
96: progress_bar_text$step()}
97:
98: resultados<-matrix(c(q.est,eta.est,b1.est,b2.est,conv,desvpad1,
99: desvpad2,desvpad3,desvpad4),ncol=9)
100: dados=as.data.frame(resultados)
101:
102: #####
103: conv<-dados$V5           # convergência
104: q.est<-dados$V1         # estimativas q
105: eta.est<-dados$V2       # estimativas eta
106: b1.est<-dados$V3        #estimativas beta 1
107: b2.est<-dados$V4        #estimativas beta 2
108: desvpad1<-dados$V6      # desvio padrão assintotico q
109: desvpad2<-dados$V7      # desvio padrão assintotico eta
110: desvpad3<-dados$V8      # desvio padrão assintotico beta 1
111: desvpad4<-dados$V9      # desvio padrão assintotico beta 2
112:
113: ##### q #####
114: M<-10000
115: entra<-which(conv==0)
116: entra<-entra[1:M]
117: est.q<-q.est[entra]
118: qchapeu<-round(mean(est.q),4) # Estimativa média q
119: EQM.q<-round(mean((est.q-q.true)^2),4) # EQM q
120:
121: # Intervalo de confiança q
122: require(emdbook)
123: desvpad1.val<-desvpad1[entra]
124: var.u<-matrix(0,length(desvpad1.val),1)

```

```

125: for (i in 1:length(entra)){var.u[i]<-deltavar(fun=log(-log(est.q)),
126: Sigma=(desvpad1.val[i])^2, meanval=c(est.q = est.q[i]))}
127:
128: LI.q<-est.q^(exp(qnorm(0.975)*sqrt(var.u))) # limite inferior
129:
130: LS.q<-est.q^(exp(-qnorm(0.975)*sqrt(var.u))) # limite superior
131:
132: PC.q<-mean(LI.q< q.true & q.true< LS.q) # Prob. de cobertura q
133:
134: ##### eta #####
135: est.eta<-eta.est[entra]
136: etachapeu<-round(mean(est.eta),4) # Estimativa média eta
137: EQM.eta<-round(mean((est.eta-eta.true)^2),4) # EQM eta
138:
139: # Intervalo de confiança eta
140: desvpad2.val<-desvpad2[entra]
141: var.v<-matrix(0,length(desvpad2.val),1)
142: for (i in 1:length(entra)){var.v[i]<-deltavar(fun = log(est.eta),
143: Sigma=(desvpad2.val[i])^2,meanval=c(est.eta = est.eta[i]))}
144:
145: LI.eta<-est.eta*exp((-qnorm(0.975)*sqrt(var.v))) # limite inferior
146: LS.eta<-est.eta*exp((qnorm(0.975)*sqrt(var.v))) # limite superior
147:
148: PC.eta<-mean(LI.eta< eta.true & eta.true< LS.eta) # Prob. de cob. eta
149:
150: ##### beta1 #####
151: est.b1<-b1.est[entra]
152: beta1chapeu<-round(mean(est.b1),4) # Estimativa média beta1
153: EQM.beta1<-round(mean((est.b1-beta1.true)^2),4) # EQM beta1
154:
155: # Intervalo de confiança beta1
156: desvpad3.val<-desvpad3[entra]
157: LI.b1 <-est.b1 -qnorm(0.975)*desvpad3.val
158:
159: LS.b1 <-est.b1 +qnorm(0.975)*desvpad3.val
160:
161: # Prob. de cob. beta1
162: PC.b1<-mean(LI.b1< beta1.true & beta1.true< LS.b1)
163:
164: ##### beta2 #####
165: est.b2<-b2.est[entra]
166: beta2chapeu<-round(mean(est.b2),4) # Estimativa média beta2
167: EQM.beta2<-round(mean((est.b2-beta2.true)^2),4) #EQM beta2
168:
169: # Intervalo de confiança beta2
170: desvpad4.val<-desvpad4[entra]
171: LI.b2 <-est.b2 -qnorm(0.975)*desvpad4.val

```

```
172:
173: LS.b2 <- est.b2 + qnorm(0.975) * desvpad4.val
174:
175: # Prob. de cob. beta2
176: PC.b2 <- mean(LI.b2 < beta2.true & beta2.true < LS.b2)
```


INTERVALO DE CONFIANÇA DOS PARÂMETROS DA DISTRIBUIÇÃO WD

Parâmetro q

Ao considerar a seguinte transformação

$$u = \log[-\log q] \implies q = e^{-e^u},$$

e que:

$$\hat{u} = \log[-\log \hat{q}] \implies \hat{q} = e^{-e^{\hat{u}}}.$$

Assim:

$$P\left(-Z_{1-\alpha/2} < \frac{\hat{u} - u}{\sqrt{\widehat{\text{Var}}(\hat{u})}} < Z_{1-\alpha/2}\right) = 1 - \alpha$$

$$\begin{aligned} -Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{u})} &< \hat{u} - u < Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{u})} \\ -\hat{u} - Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{u})} &< -u < -\hat{u} + Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{u})} \\ \hat{u} + Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{u})} &> u > \hat{u} - Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{u})} \\ \hat{u} - Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{u})} &< u < \hat{u} + Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{u})}. \end{aligned}$$

Aplicando e^{-e} , na desigualdade, fica:

$$e^{-e^{\hat{u} - Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{u})}}} < \underbrace{e^{-e^u}}_q < e^{-e^{\hat{u} + Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{u})}}}.$$

Tem-se:

$$\begin{aligned}
 e^{-e^{\hat{u}+Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{u})}} &< q < e^{-e^{\hat{u}-Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{u})}} \\
 e^{-e^{\hat{u}}}e^{Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{u})}} &< q < e^{-e^{\hat{u}}}e^{-Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{u})}} \\
 \left[e^{-e^{\hat{u}}} \right] e^{Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{u})}} &< q < \left[e^{-e^{\hat{u}}} \right] e^{-Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{u})}} \\
 (\hat{q})e^{Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{u})}} &< q < (\hat{q})e^{-Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{u})}}.
 \end{aligned}$$

Logo, o intervalo de confiança para o parâmetro q é dado por:

$$\left[(\hat{q})e^{Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{u})}}; (\hat{q})e^{-Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{u})}} \right].$$

Parâmetro η

Ao considerar a seguinte transformação

$$v = \log \eta \quad \Longrightarrow \quad \eta = e^v,$$

e que:

$$\hat{v} = \log \hat{\eta} \quad \Longrightarrow \quad \hat{\eta} = e^{\hat{v}}.$$

Assim:

$$\begin{aligned}
 P \left(-Z_{1-\alpha/2} < \frac{\hat{v} - v}{\sqrt{\widehat{\text{Var}}(\hat{v})}} < Z_{1-\alpha/2} \right) &= 1 - \alpha \\
 -Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{v})} < \hat{v} - v < Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{v})} \\
 -\hat{v} - Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{v})} < -v < -\hat{v} + Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{v})} \\
 \hat{v} + Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{v})} > v > \hat{v} - Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{v})} \\
 \hat{v} - Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{v})} < v < \hat{v} + Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{v})}.
 \end{aligned}$$

Aplicando e , na desigualdade, fica:

$$e^{\hat{v}-Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{v})}} < \underbrace{e^v}_{\eta} < e^{\hat{v}+Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{v})}}.$$

Tem-se:

$$\begin{aligned}
 e^{\hat{v}-Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{v})}} &< \eta < e^{\hat{v}+Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{v})}} \\
 e^{\hat{v}}}e^{-Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{v})}} &< \eta < e^{\hat{v}}}e^{Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{v})}} \\
 \left[e^{\hat{v}}} \right] e^{-Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{v})}} &< \eta < \left[e^{\hat{v}}} \right] e^{Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{v})}} \\
 \hat{\eta}e^{-Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{v})}} &< \eta < \hat{\eta}e^{Z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\hat{v})}}.
 \end{aligned}$$

Logo, o intervalo de confiança para o parâmetro η é dado por:

$$\left[\hat{\eta} e^{-Z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}(\hat{\nu})}}; \hat{\eta} e^{Z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}(\hat{\nu})}} \right].$$

RESULTADOS DAS SIMULAÇÕES

Tabela 16 – Média das estimativas, EQM e PC dos parâmetros MRCSP-WD considerando os cenários da simulação e diversos tamanhos amostrais.

n	Cenário	ϱ			η			β_1			β_2		
		Média	EQM	PC	Média	EQM	PC	Média	EQM	PC	Média	EQM	PC
30	C ₁	0,8984	0,0034	0,9404	1,5884	0,0736	0,9222	2,1392	0,5752	0,9451	1,0839	0,1529	0,9344
	C ₂	0,8992	0,0029	0,9379	1,0557	0,0295	0,9227	2,1380	0,5921	0,9449	1,0811	0,1318	0,9380
	C ₃	0,8992	0,0029	0,9384	0,7918	0,0165	0,9201	2,1369	0,5934	0,9442	1,0806	0,1310	0,9378
50	C ₁	0,8995	0,0017	0,9414	1,5512	0,0366	0,9313	2,0880	0,3352	0,9452	1,0479	0,0904	0,9463
	C ₂	0,9005	0,0016	0,9454	1,0344	0,0153	0,9328	2,0789	0,3226	0,9451	1,0435	0,0864	0,9423
	C ₃	0,8996	0,0018	0,9442	0,7764	0,0093	0,9336	2,0791	0,3267	0,9457	1,0456	0,0701	0,9411
100	C ₁	0,8999	0,0008	0,9498	1,5252	0,0163	0,9420	2,0380	0,1526	0,9455	1,0242	0,0410	0,9454
	C ₂	0,8999	0,0008	0,9466	1,0162	0,0071	0,9417	2,0352	0,1516	0,9467	1,0202	0,0404	0,9460
	C ₃	0,9003	0,0008	0,9460	0,7628	0,0039	0,9382	2,0431	0,1542	0,9460	1,0247	0,0379	0,9479
250	C ₁	0,8999	0,0003	0,9494	1,5097	0,0061	0,9511	2,0137	0,0579	0,9480	1,0077	0,0169	0,9524
	C ₂	0,8998	0,0003	0,9486	1,0057	0,0026	0,9479	2,0161	0,0568	0,9503	1,0093	0,0142	0,9510
	C ₃	0,9003	0,0003	0,9479	0,7550	0,0015	0,9441	2,0185	0,0584	0,9478	1,0096	0,0157	0,9519
500	C ₁	0,9001	0,0002	0,9506	1,5051	0,0030	0,9494	2,0067	0,0288	0,9491	1,0042	0,0071	0,9508
	C ₂	0,9001	0,0002	0,9513	1,0035	0,0013	0,9483	2,0068	0,0308	0,9503	1,0056	0,0077	0,9496
	C ₃	0,8999	0,0002	0,9517	0,7523	0,0007	0,9463	2,0101	0,0287	0,9484	1,0034	0,0070	0,9516

Fonte: Elaborada pelo autor.

Tabela 17 – Média das estimativas, EQM e Probabilidade de cobertura (PC) dos parâmetros MRCSP-WD considerando os cenários da simulação e diversos tamanhos amostrais e percentuais de censura.

n	Cenário	Percentual de Censura	q			η			β_1			β_2			
			Média	EQM	PC	Média	EQM	PC	Média	EQM	PC	Média	EQM	PC	
30	C ₁	5%	0,9040	0,0030	0,9327	1,5974	0,0765	0,9128	2,1140	0,6585	0,9403	1,0632	0,1763	0,9387	
			C ₂	0,9038	0,0035	0,9357	1,0719	0,0391	0,9171	2,1007	0,5741	0,9428	1,0664	0,2108	0,9369
			C ₃	0,9038	0,0035	0,9361	0,8040	0,0219	0,9149	2,1019	0,5756	0,9428	1,0680	0,2110	0,9361
	C ₁	10%	0,9094	0,0034	0,9309	1,6172	0,0960	0,9105	2,0523	0,5905	0,9408	1,0357	0,2108	0,9362	
			C ₂	0,9096	0,0033	0,9311	1,0813	0,0425	0,9075	2,0670	0,5857	0,9425	1,0448	0,2104	0,9370
			C ₃	0,9118	0,0025	0,9229	0,8075	0,0191	0,9057	2,0797	0,6166	0,9441	1,0529	0,1547	0,9384
	C ₁	15%	0,9192	0,0023	0,9233	1,6260	0,0896	0,9054	2,0491	0,7313	0,9386	1,0134	0,2043	0,9304	
			C ₂	0,9186	0,0024	0,9221	1,0892	0,0404	0,8983	2,0489	0,6543	0,9377	1,0280	0,2425	0,9366
			C ₃	0,9184	0,0024	0,9237	0,8155	0,0244	0,9012	2,0495	0,6385	0,9412	1,0306	0,2029	0,9385
	C ₁	25%	0,9266	0,0031	0,9023	1,6487	0,1121	0,8894	1,9425	0,6484	0,9382	0,9621	0,1642	0,9275	
			C ₂	0,9264	0,0033	0,9038	1,1106	0,0556	0,8803	1,9639	0,6380	0,9374	0,9720	0,2280	0,9330
			C ₃	0,9287	0,0027	0,8960	0,8313	0,0282	0,8856	2,0030	0,7098	0,9367	0,9872	0,1837	0,9314
50	C ₁	5%	0,9060	0,0015	0,9457	1,5576	0,0369	0,9306	2,0413	0,3223	0,9474	1,0217	0,1051	0,9455	
			C ₂	0,9067	0,0014	0,9414	1,0422	0,0165	0,9273	2,0447	0,3317	0,9452	1,0299	0,0763	0,9467
			C ₃	0,9066	0,0014	0,9417	0,7817	0,0093	0,9279	2,0463	0,3308	0,9455	1,0309	0,0761	0,9465
	C ₁	10%	0,9117	0,0015	0,9339	1,5660	0,0402	0,9261	1,9955	0,3267	0,9463	0,9971	0,1059	0,9433	
			C ₂	0,9123	0,0015	0,9346	1,0497	0,0182	0,9221	2,0086	0,3346	0,9453	1,0091	0,0768	0,9450
			C ₃	0,9122	0,0015	0,9340	0,7875	0,0103	0,9231	2,0127	0,3347	0,9448	1,0114	0,0768	0,9458
	C ₁	15%	0,9175	0,0016	0,9192	1,5759	0,0043	0,9221	1,9479	0,3374	0,9433	0,9731	0,1089	0,9423	
			C ₂	0,9172	0,0017	0,9206	1,0579	0,0212	0,9169	1,9777	0,3421	0,9419	0,9872	0,0959	0,9336
			C ₃	0,9179	0,0016	0,9171	0,7940	0,0115	0,9117	1,9764	0,3399	0,9416	0,9918	0,0781	0,9418
	C ₁	25%	0,9294	0,0019	0,8732	1,6018	0,0549	0,9052	1,8580	0,3809	0,9272	0,9270	0,0880	0,9249	
			C ₂	0,9279	0,0019	0,8850	1,0705	0,0255	0,9062	1,9030	0,3643	0,9343	0,9464	0,0932	0,9285
			C ₃	0,9286	0,0019	0,8764	0,8060	0,0141	0,8978	1,9088	0,3637	0,9391	0,9492	0,1187	0,9370
100	C ₁	5%	0,9055	0,0008	0,9403	1,5301	0,0177	0,9372	1,9911	0,1530	0,9468	0,9962	0,0406	0,9482	
			C ₂	0,9059	0,0008	0,9420	1,0226	0,0078	0,9371	1,9944	0,1495	0,9453	1,0005	0,0398	0,9418
			C ₃	0,9058	0,0008	0,9424	0,7672	0,0044	0,9387	1,9965	0,1494	0,9447	1,0016	0,0398	0,9426
	C ₁	10%	0,9113	0,0009	0,9286	1,5373	0,0192	0,9336	1,9433	0,1579	0,9338	0,9704	0,0423	0,9432	
			C ₂	0,9114	0,0009	0,9271	1,0287	0,0086	0,9326	1,9589	0,1550	0,9408	0,9808	0,0410	0,9425
			C ₃	0,9114	0,0009	0,9281	0,7719	0,0049	0,9311	1,9625	0,1549	0,9419	0,9830	0,0409	0,9434
	C ₁	15%	0,9170	0,0010	0,9038	1,5451	0,0208	0,9291	1,8995	0,1680	0,9289	0,9466	0,0450	0,9369	
			C ₂	0,9172	0,0010	0,9007	1,0358	0,0095	0,9238	1,9216	0,1635	0,9353	0,9613	0,0432	0,9403
			C ₃	0,9171	0,0010	0,9029	0,7774	0,0054	0,9227	1,9275	0,1627	0,9374	0,9645	0,0429	0,9410
	C ₁	25%	0,9296	0,0015	0,8104	1,5740	0,0275	0,9066	1,8081	0,2040	0,9091	0,8979	0,0596	0,9170	
			C ₂	0,9285	0,0014	0,8111	1,0519	0,0123	0,9057	1,8474	0,1881	0,9203	0,9220	0,0503	0,9222
			C ₃	0,9284	0,0014	0,8112	0,7824	0,0070	0,9040	1,8568	0,1857	0,9232	0,9273	0,0496	0,9266
250	C ₁	5%	0,9056	0,0004	0,9363	1,5158	0,0066	0,9435	1,9737	0,0590	0,9473	0,9855	0,0149	0,9438	
			C ₂	0,9057	0,0003	0,9362	1,0124	0,0030	0,9388	1,9798	0,0580	0,9471	0,9874	0,0155	0,9487
			C ₃	0,9057	0,0003	0,9373	0,7595	0,0017	0,9390	1,9818	0,0577	0,9469	0,9884	0,0155	0,9486
	C ₁	10%	0,9115	0,0004	0,8995	1,5235	0,0073	0,9365	1,9288	0,0639	0,9371	0,9602	0,0164	0,9328	
			C ₂	0,9117	0,0004	0,9008	1,0193	0,0033	0,9326	1,9420	0,0622	0,9403	0,9670	0,0168	0,9369
			C ₃	0,9116	0,0004	0,9003	0,7648	0,0019	0,9314	1,9458	0,0617	0,9417	0,9687	0,0166	0,9377
	C ₁	15%	0,9174	0,0006	0,8315	1,5311	0,0082	0,9305	1,8808	0,0743	0,9131	0,9350	0,0202	0,9116	
			C ₂	0,9175	0,0006	0,8262	1,0264	0,0039	0,9195	1,9037	0,0695	0,9246	0,9463	0,0188	0,9236
			C ₃	0,9173	0,0006	0,8316	0,7703	0,0022	0,9196	1,9123	0,0667	0,9301	0,9498	0,0175	0,9218
	C ₁	25%	0,9288	0,0011	0,6134	1,5481	0,0105	0,9111	1,7899	0,1065	0,8575	0,8866	0,0295	0,8436	
			C ₂	0,9289	0,0011	0,6055	1,0414	0,0053	0,8878	1,8268	0,0926	0,8822	0,9046	0,0258	0,8741
			C ₃	0,9288	0,0011	0,6132	0,7824	0,0031	0,8758	1,8385	0,0882	0,8955	0,9084	0,0242	0,8714
500	C ₁	5%	0,9056	0,0002	0,9238	1,5106	0,034	0,9427	1,9649	0,0300	0,9410	0,9806	0,0083	0,9407	
			C ₂	0,9057	0,0002	0,9244	1,0090	0,0015	0,9376	1,9724	0,0294	0,9449	0,9847	0,0080	0,9436
			C ₃	0,9057	0,0002	0,9256	0,7569	0,0008	0,9382	1,9743	0,0292	0,9459	0,9857	0,0080	0,9436
	C ₁	10%	0,9115	0,0003	0,8472	1,5170	0,0036	0,9347	1,9172	0,0366	0,9129	0,9555	0,0095	0,9158	
			C ₂	0,9115	0,0003	0,8452	1,0154	0,0018	0,9237	1,9353	0,0334	0,9265	0,9641	0,0093	0,9246
			C ₃	0,9114	0,0003	0,8457	0,7619	0,0001	0,9226	1,9390	0,0329	0,9280	0,9660	0,0092	0,9255
	C ₁	15%	0,9172	0,0004	0,7109	1,5238	0,0041	0,9248	1,8713	0,0466	0,8730	0,9310	0,0124	0,8679	
			C ₂	0,9173	0,0004	0,7174	1,0222	0,0021	0,9012	1,8974	0,0402	0,9017	0,9436	0,0114	0,8929
			C ₃	0,9172	0,0004	0,7229	0,7671	0,0012	0,8995	1,9029	0,0391	0,9061	0,9465	0,0110	0,8994
	C ₁	25%	0,9292	0,0010	0,3520	1,5450	0,0062	0,8815	1,7803	0,0788	0,7476	0,8777	0,0230	0,7037	
			C ₂	0,9288	0,0010	0,3443	1,0372	0,0032	0,8466	1,8209	0,0636	0,8148	0,9012	0,0183	0,8031
			C ₃	0,9287	0,0009	0,3136	0,7779	0,0018	0,8425	1,8198	0,0641	0,8123	0,9042	0,0180	0,8124

Fonte: Elaborada pelo autor.

DISTRIBUIÇÕES EMPIRICAS DAS ESTIMATIVAS COM AS CURVAS GAUSSIANAS SOBREPOSTAS

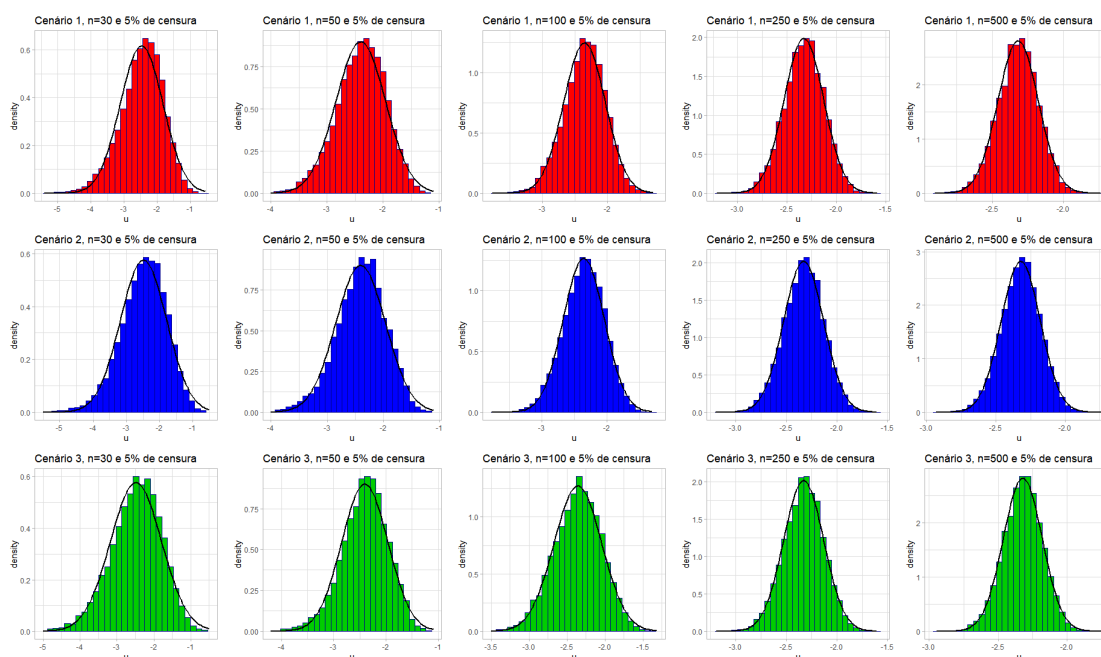


Figura 32 – Distribuições empíricas de \hat{u} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 5% de censuras.

Fonte: Elaborada pelo autor.

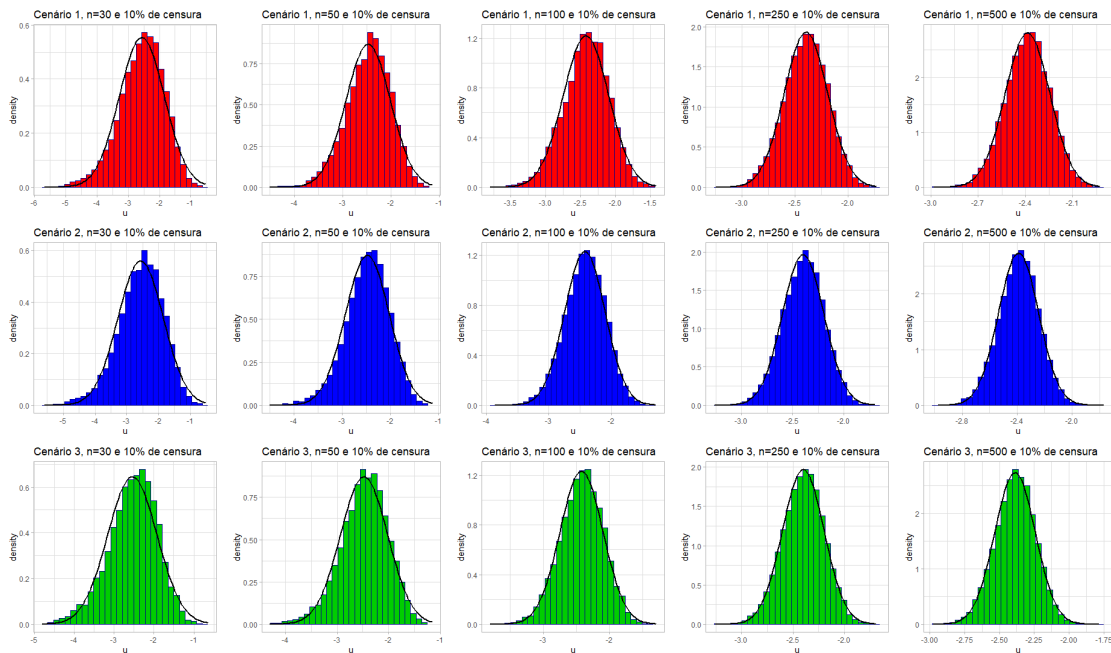


Figura 33 – Distribuições empíricas de \hat{u} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 10% de censuras.

Fonte: Elaborada pelo autor.

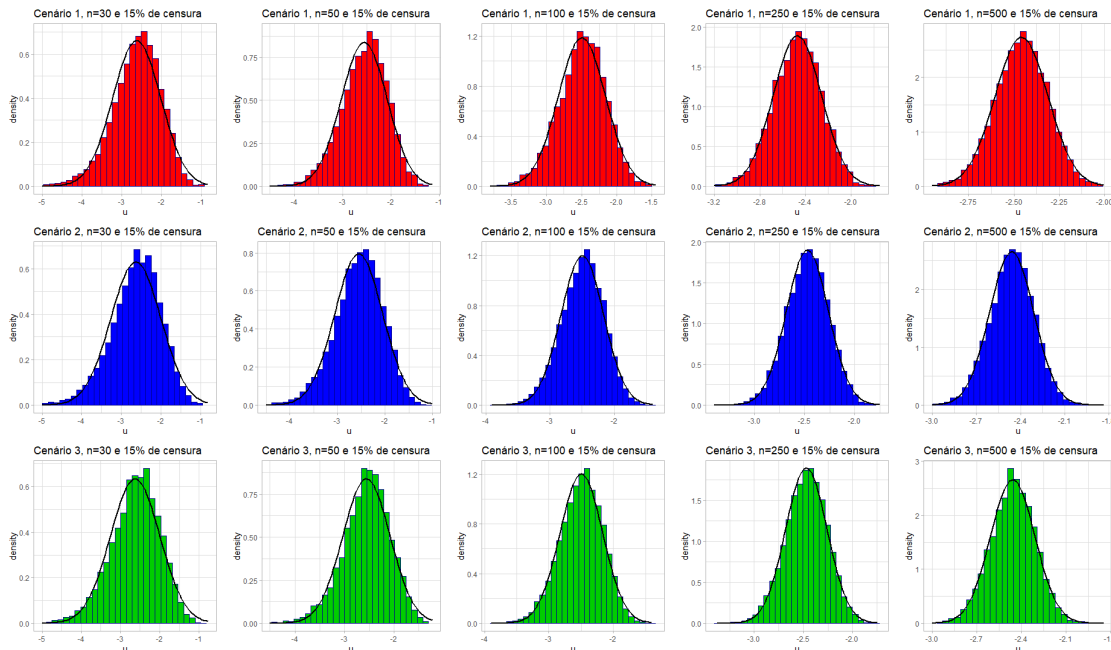


Figura 34 – Distribuições empíricas de \hat{u} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 15% de censuras.

Fonte: Elaborada pelo autor.

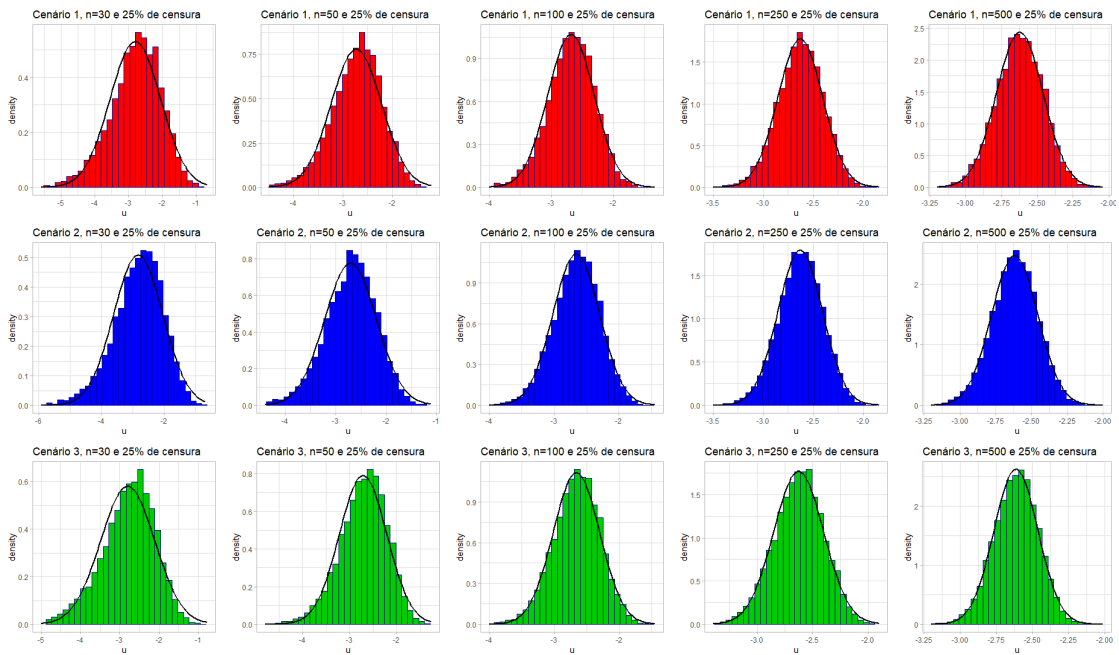


Figura 35 – Distribuições empíricas de \hat{u} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 25% de censuras.

Fonte: Elaborada pelo autor.

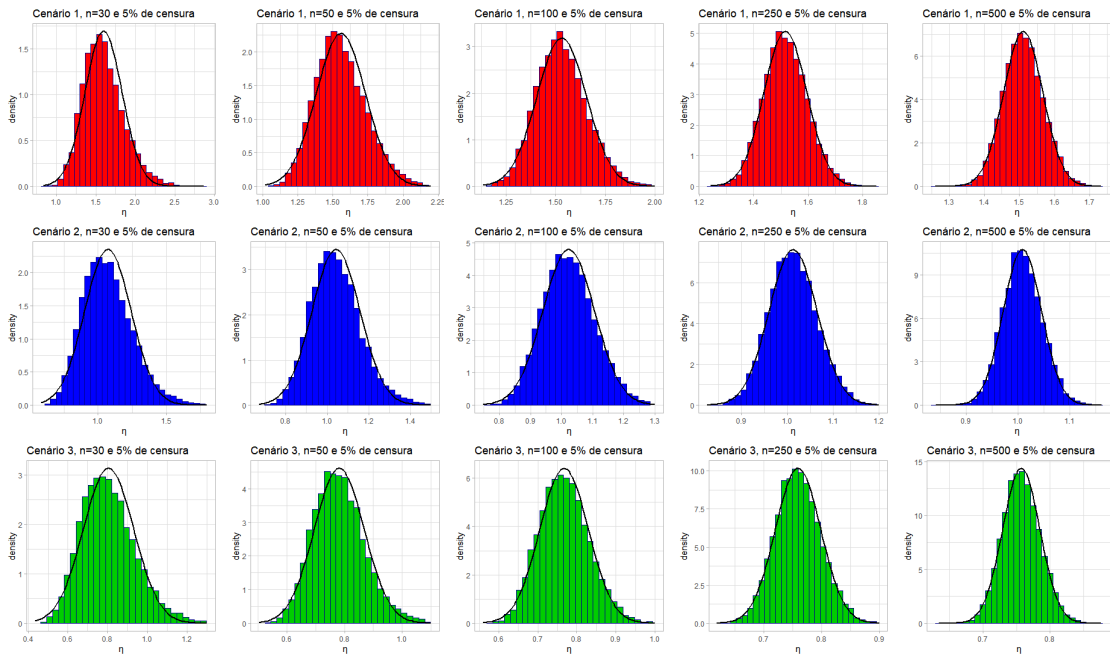


Figura 36 – Distribuições empíricas de $\hat{\eta}$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 5% de censuras.

Fonte: Elaborada pelo autor.

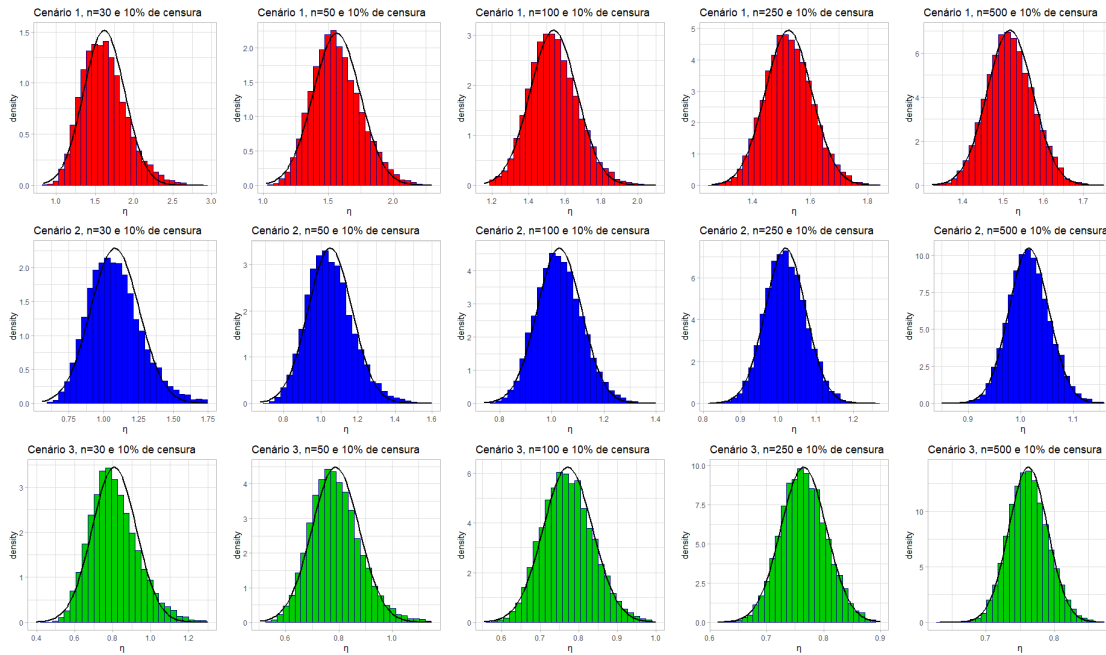


Figura 37 – Distribuições empíricas de $\hat{\eta}$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 10% de censuras.

Fonte: Elaborada pelo autor.

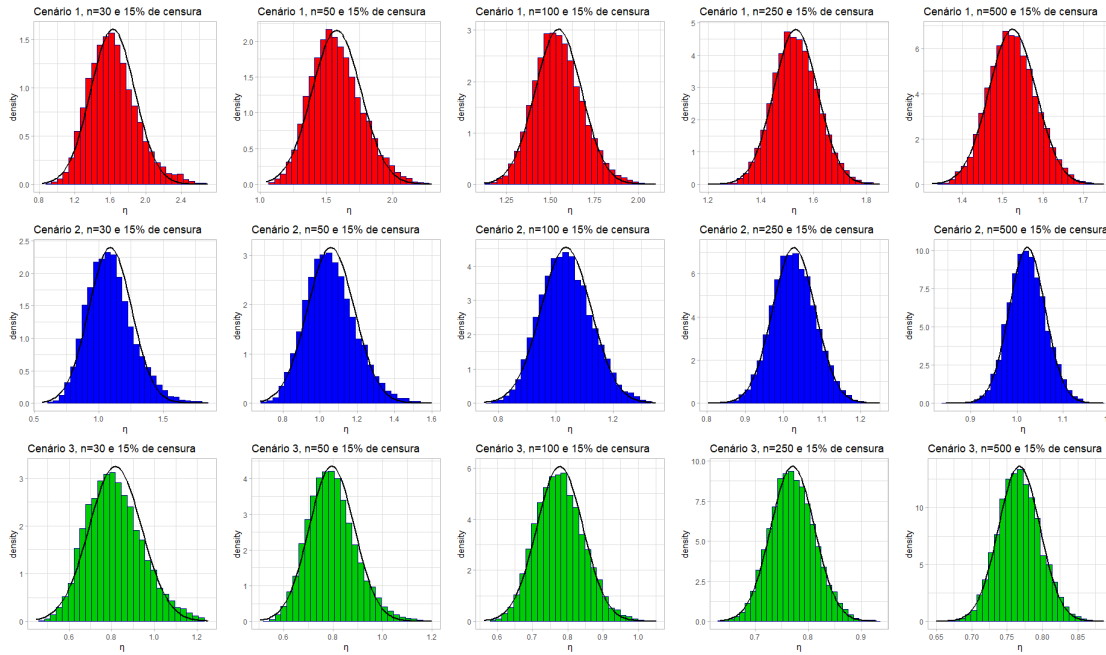


Figura 38 – Distribuições empíricas de $\hat{\eta}$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 15% de censuras.

Fonte: Elaborada pelo autor.

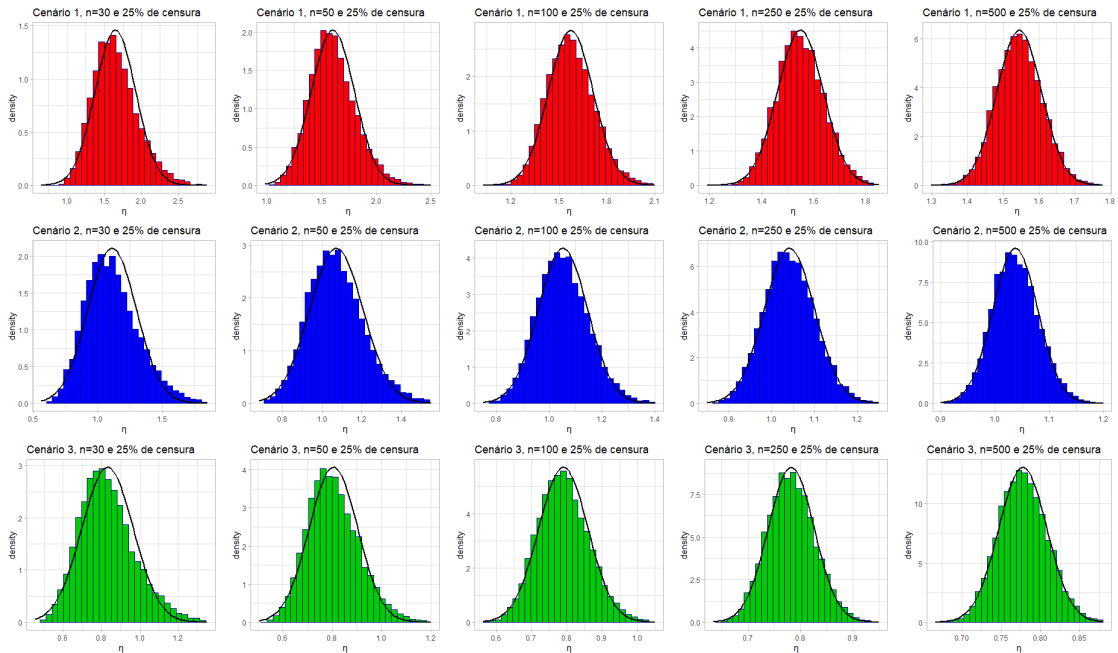


Figura 39 – Distribuições empíricas de $\hat{\eta}$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 25% de censuras.

Fonte: Elaborada pelo autor.

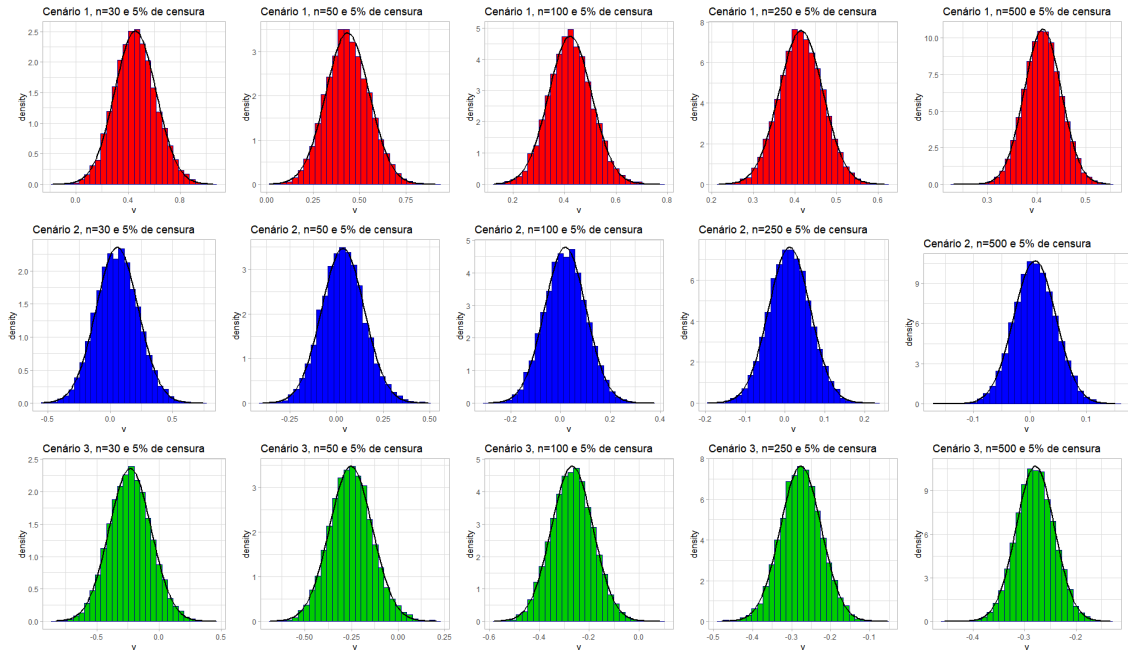


Figura 40 – Distribuições empíricas de \hat{v} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 5% de censuras.

Fonte: Elaborada pelo autor.

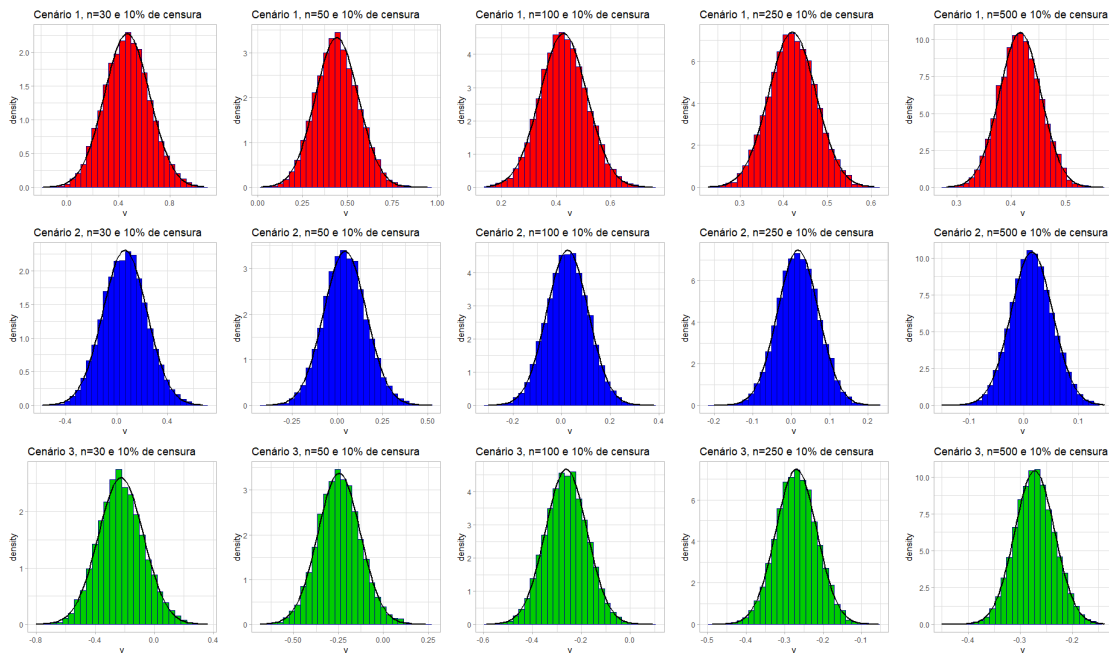


Figura 41 – Distribuições empíricas de \hat{v} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 10% de censuras.

Fonte: Elaborada pelo autor.

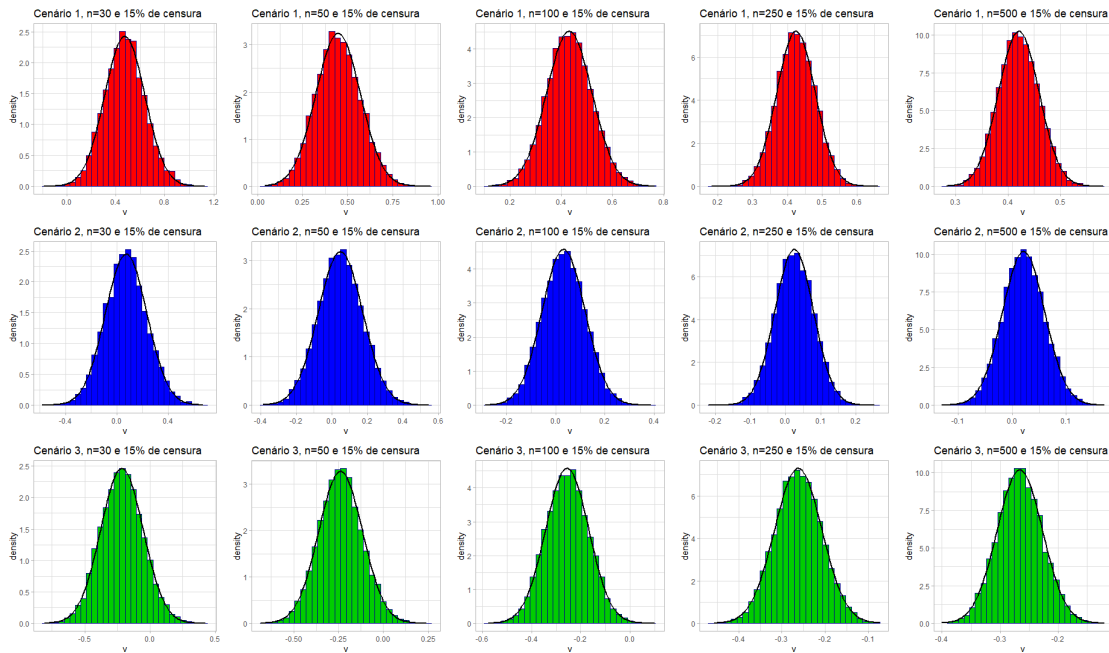


Figura 42 – Distribuições empíricas de \hat{v} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 15% de censuras.

Fonte: Elaborada pelo autor.

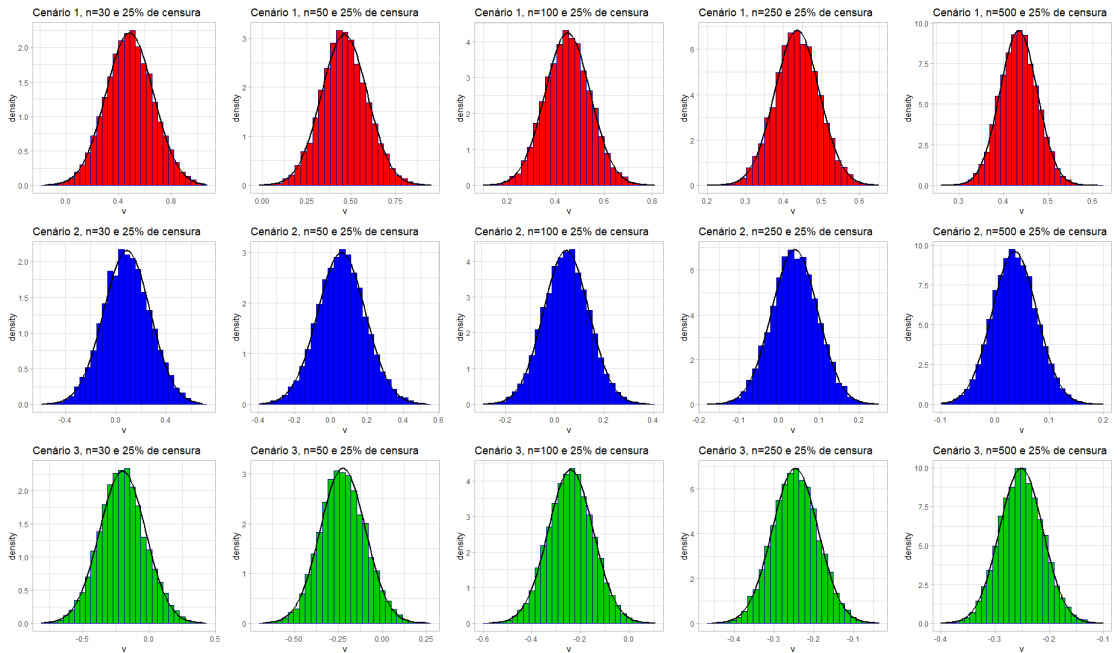


Figura 43 – Distribuições empíricas de \hat{v} separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 25% de censuras.

Fonte: Elaborada pelo autor.

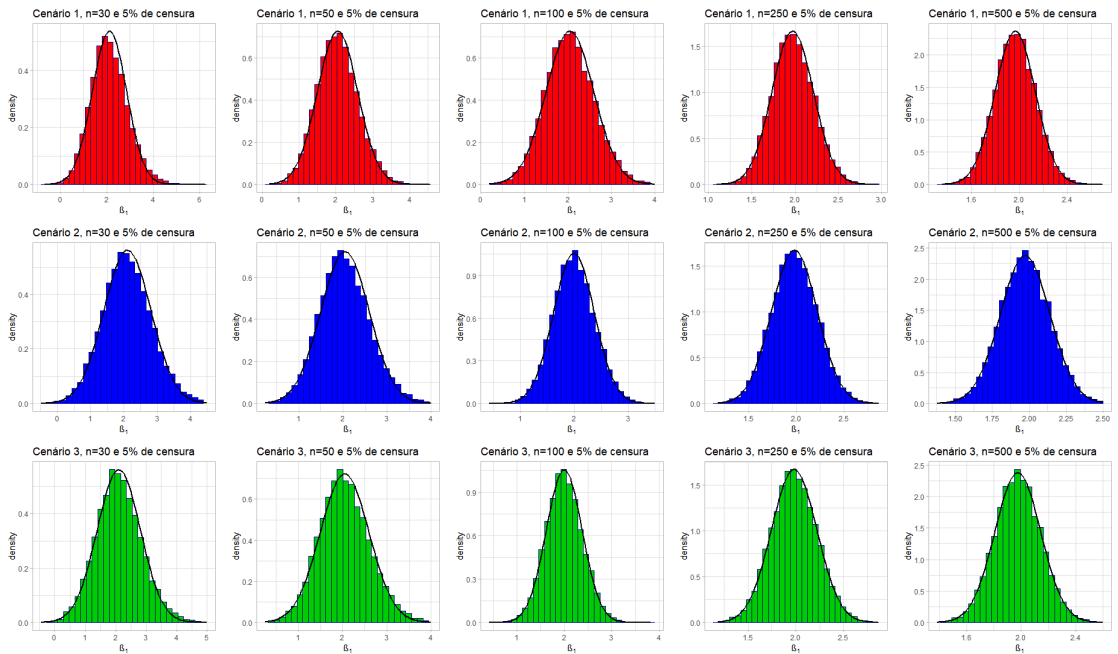


Figura 44 – Distribuições empíricas de $\hat{\beta}_1$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 5% de censuras.

Fonte: Elaborada pelo autor.

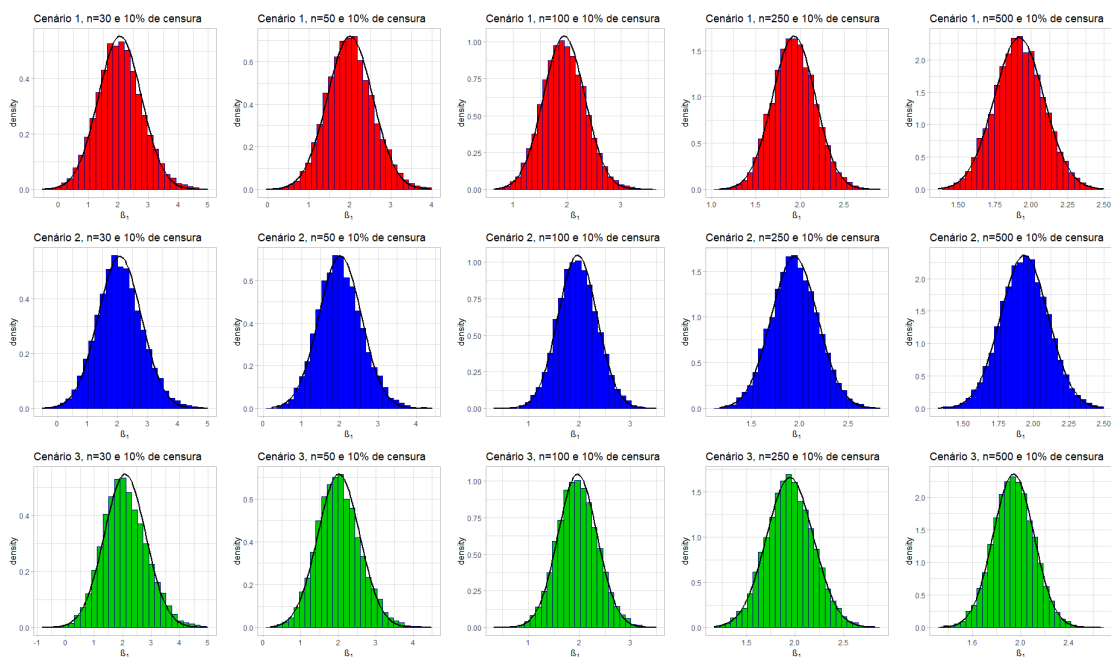


Figura 45 – Distribuições empíricas de $\hat{\beta}_1$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 10% de censuras.

Fonte: Elaborada pelo autor.

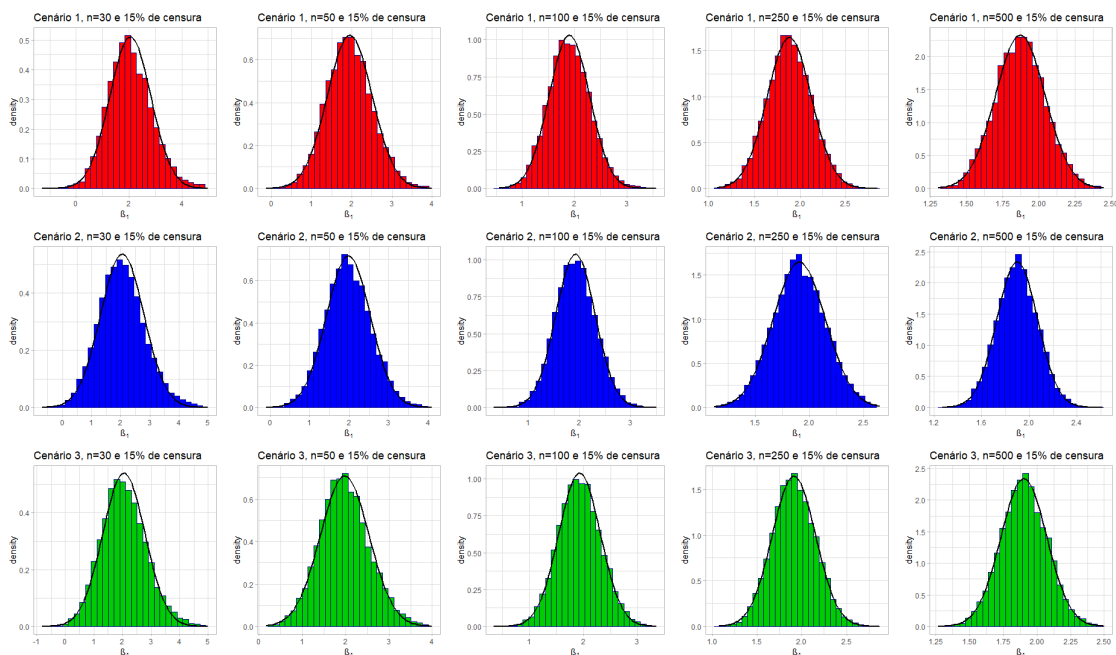


Figura 46 – Distribuições empíricas de $\hat{\beta}_1$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 15% de censuras.

Fonte: Elaborada pelo autor.

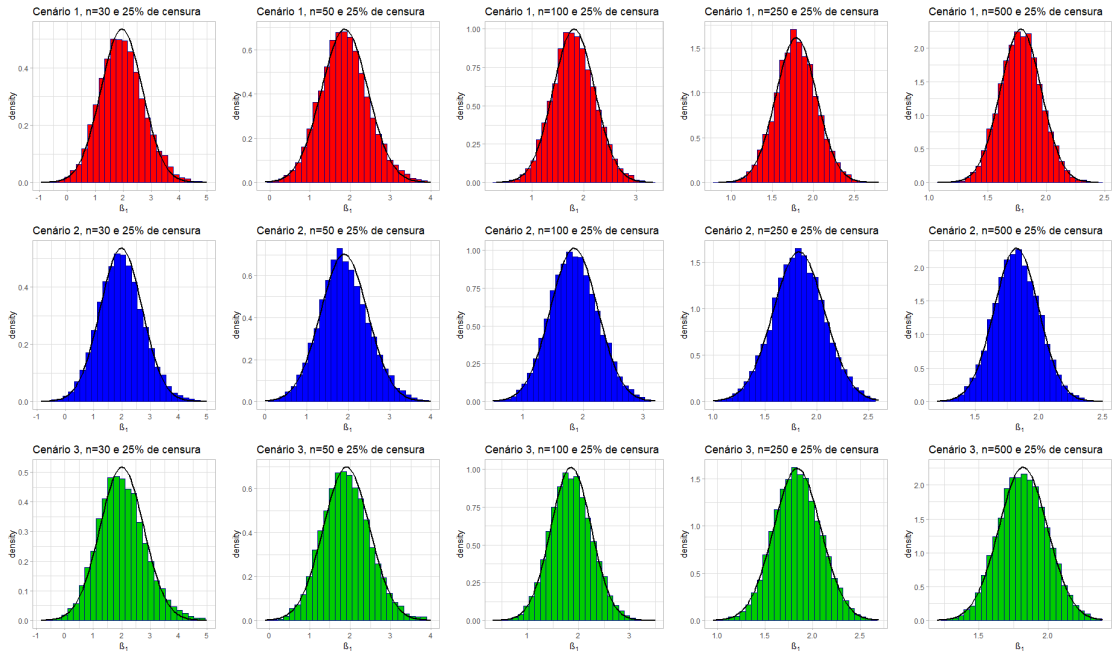


Figura 47 – Distribuições empíricas de $\hat{\beta}_1$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 25% de censuras.

Fonte: Elaborada pelo autor.

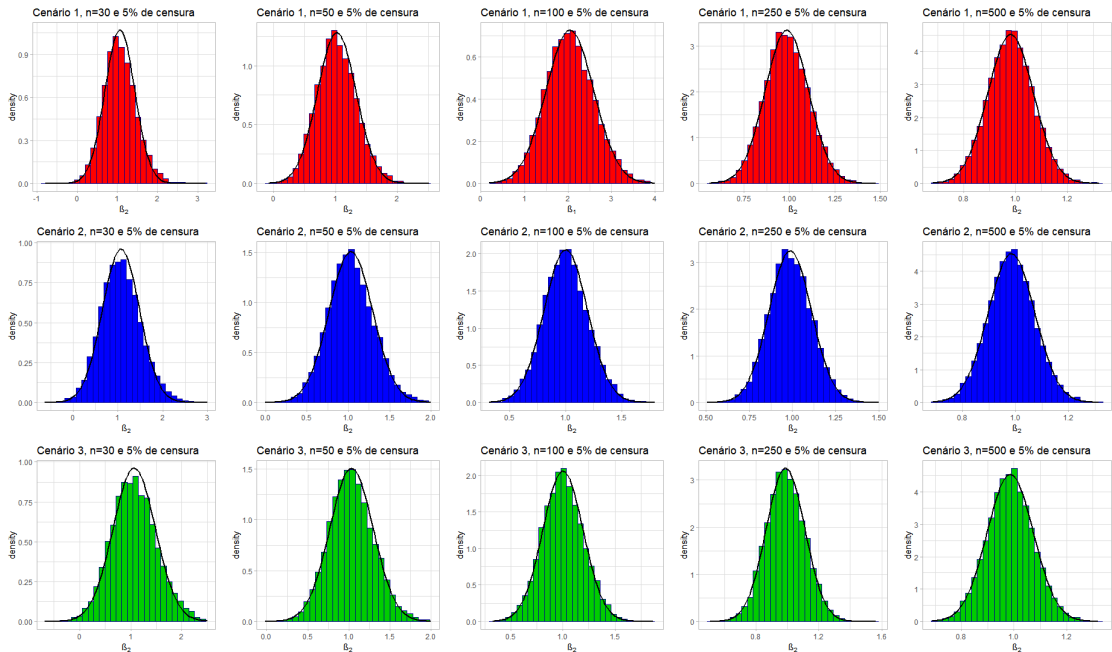


Figura 48 – Distribuições empíricas de $\hat{\beta}_2$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 5% de censuras.

Fonte: Elaborada pelo autor.

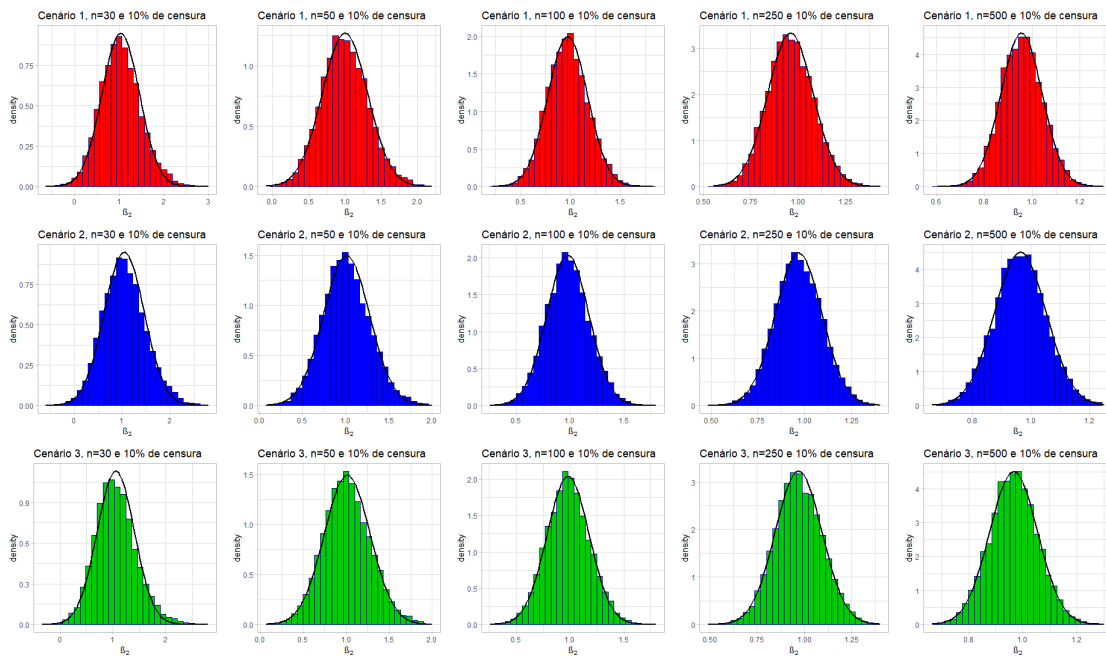


Figura 49 – Distribuições empíricas de $\hat{\beta}_2$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 10% de censuras.

Fonte: Elaborada pelo autor.

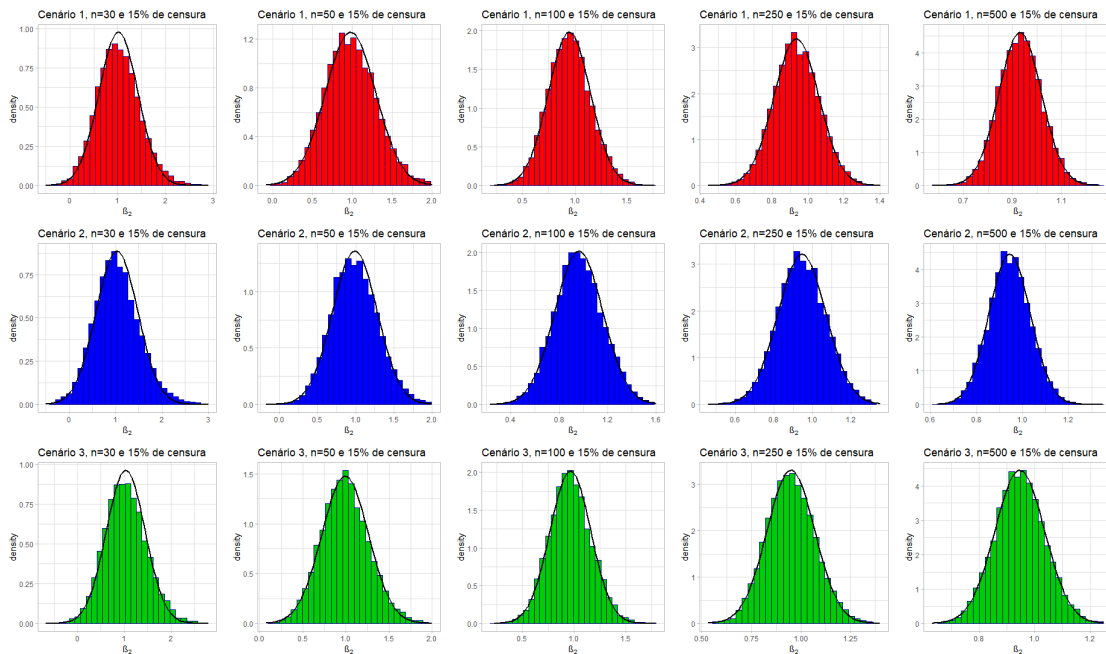


Figura 50 – Distribuições empíricas de $\hat{\beta}_2$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 15% de censuras.

Fonte: Elaborada pelo autor.

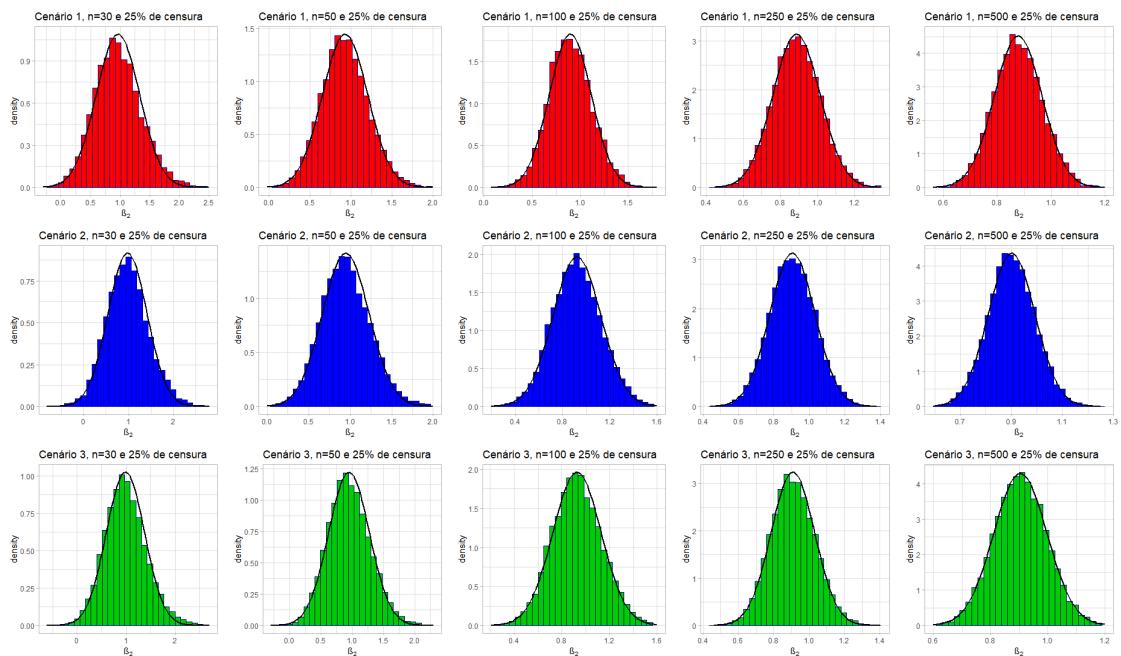


Figura 51 – Distribuições empíricas de $\hat{\beta}_2$ separadas por cenário e tamanho amostral com as respectivas curvas gaussianas estimadas sobrepostas para 25% de censuras.

Fonte: Elaborada pelo autor.

