



**Universidade de Brasília  
Departamento de Estatística**

**Estudo do Resíduo de Cox-Snell para Dados Censurados**

**Luíza Maria Veiga de Sant'Anna**

Monografia apresentada para obtenção do título de Bacharel em Estatística.

**Brasília  
2017**



Luíza Maria Veiga de Sant'Anna

**Estudo do Resíduo de Cox-Snell para Dados Censurados**

Orientadora:

Profa. Dra. **Juliana Betini Fachini Gomes**

Monografia apresentada para obtenção do título de Bacharel em Estatística.

**Brasília**  
**2017**



## DEDICATÓRIA

*Aos meus amados pais,*

***Lucimar** e **Silvio**, pessoas fundamentais nessa conquista. Por não medirem esforços para me dar uma boa educação, bons exemplos e pelo amor incondicional.*



## AGRADECIMENTOS

A Deus, por me dar forças, iluminar meus caminhos e me conceder uma família unida e solidária.

Aos meus pais, Lucimar e Silvio, pelo amor incondicional, pelo apoio em todos os momentos, por serem exemplos para mim e meu porto seguro em todos os momentos.

Às minhas avós, Antônia e Zoraide, por todo o amor e motivação.

À minha família, pelo apoio e por ser fonte de forças e motivação.

À professora Dra. Juliana Betini Fachini Gomes, pela orientação, compreensão, paciência e carinho desde o primeiro momento e por ser um exemplo de profissional e ser humano.

Aos funcionários do Departamento de Estatística da UnB, pelo auxílio durante todo o curso.

Aos meus amigos da escola, Bruna Urueña, Fernanda Yoshizaki, Gustavo Pereira, Larissa Berber, Ludmila Ulhoa, Mariana de Amorim e Sarah Almeida, por todo o amor, apoio e compreensão sempre e por me acompanharem há tanto tempo alegrando os meus dias.

Aos grandes amigos que a UnB me trouxe, Bárbara Santiago, Bruno Vilas Boas, Bruno Matos, Eduarda Leão, Felipe Martins, Isabella Cristine, Laura Teixeira, Leylanne Alencar, Lucas Rodrigues, Ludimila Nobre e Marina Macedo, pelo apoio, pelas incontáveis vezes em que me ajudaram nas mais diversas situações, por fazerem meus dias mais alegres e cheios de amor na universidade e por todo o apoio, paciência e compreensão principalmente na reta final.

Às amigas da PGFN, Bruna Costa, Francielle de Jesus, Jacqueline Fonseca e Jéssica Duarte, pela companhia, pelo amor e por tornarem meus dias de trabalho mais leves e felizes.

À minha primeira e inesquecível chefe, Patrícia Castilho, por todos os ensinamentos e por ser um exemplo e fonte de inspiração.

Aos amigos da CNI, Bianca Bassul, Cleiton Felinto, Flávia Samantha, Israel Azevedo, Nicholas Müller e Thalita Oliveira, pela amizade, compreensão, força, por serem fonte de amor e alegria e pela companhia todas as tardes nos últimos seis meses.

A todas as pessoas que contribuíram direta ou indiretamente para a realização

6

deste trabalho.



## SUMÁRIO

RESUMO . . . . .	9
1 INTRODUÇÃO . . . . .	11
2 REVISÃO DE LITERATURA . . . . .	13
2.1 Análise de Sobrevivência . . . . .	13
2.1.1 Função de Sobrevivência . . . . .	15
2.1.2 Função de Risco . . . . .	15
2.1.3 Estimador de Kaplan-Meier . . . . .	15
2.2 Distribuições de Probabilidade . . . . .	16
2.2.1 Distribuição Exponencial . . . . .	16
2.2.2 Distribuição Weibull . . . . .	17
2.2.3 Distribuição Log-normal . . . . .	17
2.2.4 Distribuição Log-logística . . . . .	18
2.3 MODELOS DE REGRESSÃO . . . . .	19
2.3.1 Modelo de Regressão Log-linear . . . . .	19
2.4 Estimação de Parâmetros . . . . .	21
2.5 Análise de Resíduos . . . . .	22
2.5.1 Resíduos de Cox-Snell . . . . .	23
2.6 Avaliação de Estimadores . . . . .	24
3 METODOLOGIA . . . . .	25
3.1 Material . . . . .	25
3.2 Métodos . . . . .	25
4 RESULTADOS E DISCUSSÃO . . . . .	31
4.1 Sem covariáveis . . . . .	31
4.1.1 Distribuição Exponencial . . . . .	31
4.1.2 Distribuição Weibull . . . . .	37
4.1.3 Distribuição Log-normal . . . . .	41
4.1.4 Distribuição Log-logística . . . . .	46
4.2 Com covariável . . . . .	51
4.2.1 Distribuição Exponencial . . . . .	51

4.2.2 Distribuição Weibull . . . . .	55
4.2.3 Distribuição Log-normal . . . . .	59
4.2.4 Distribuição Log-logística . . . . .	63
5 CONSIDERAÇÕES FINAIS . . . . .	67
REFERÊNCIAS . . . . .	69
ANEXOS . . . . .	71

## RESUMO

### Estudo do Resíduo de Cox-Snell para Dados Censurados

Neste trabalho são apresentados os resultados do comportamento dos resíduos de Cox-Snell para diferentes cenários de simulação. O objetivo principal é verificar o desempenho dos resíduos de Cox-Snell quando os dados seguem diferentes distribuições de probabilidade uma vez que um dos pressupostos do uso desse resíduo é que os dados seguem uma distribuição Exponencial se o modelo estiver bem ajustado. Métodos de simulação de Monte Carlo foram aplicados para estudar a distribuição empírica dos resíduos considerando diferentes cenários criados a partir da combinação de distribuições de probabilidade e porcentagens de censura e a qualidade dos estimadores dos parâmetros das distribuições e dos modelos de regressão foi avaliada pelo erro quadrático médio e vício.

Palavras-chave: Análise de Sobrevivência; Censura; Modelos de Regressão; Resíduo de Cox-Snell



## 1 INTRODUÇÃO

A Estatística pode ser amplamente aplicada em diversas áreas do conhecimento e apresenta várias ramificações. Uma delas, é a Análise de Sobrevivência, cujo emprego pode ser encontrado na Medicina, na área de Finanças e até mesmo nas Engenharias, motivo pelo qual também pode ser chamada de Análise de Confiabilidade.

Seu objeto de estudo é o tempo decorrido até à ocorrência de algum evento específico, o evento de interesse. A principal característica desse objeto de estudo é a censura, que ocorre quando, por algum motivo não relacionado com o evento de interesse, o mesmo não pode ser observado. No caso de uma aplicação médica, por exemplo, em que o evento de interesse é a morte devido à alguma doença específica, a censura pode se dar através da morte por algum outro motivo que não o previamente definido ou pela cura do paciente.

A variável resposta é composta pelos tempos de falha e de censura. Assim, considerando que a resposta é uma variável aleatória, pode-se utilizar métodos gráficos e funções para definir modelos que apresentam melhor adequabilidade às observações. Além disso, podem ser consideradas covariáveis, variáveis cujo efeito não se tem interesse direto, mas podem interferir na resposta, de forma que modelos de regressão podem ser construídos a fim de se estudar a relação entre elas e o tempo.

Uma vez definido o modelo, é necessária então a realização de uma análise de resíduos a fim de avaliar a adequabilidade do mesmo aos dados. O resíduo de Cox-Snell é umas das opções, visto que é muito útil para examinar o ajuste global de modelos paramétricos e semi-paramétricos de tal forma que, se o modelo apresentar um bom ajuste, pode-se tomar os resíduos,  $e_i$ 's, como uma amostra censurada de uma distribuição exponencial padrão.

No entanto, alguns autores questionam a utilização do resíduo de Cox-Snell para tal análise devido à particularidade citada anteriormente, a censura.

Sendo assim, o objetivo principal do trabalho é verificar o comportamento desse resíduo quando os dados seguem diferentes distribuições de probabilidade já que, segundo Lawless (2002), os resíduos devem seguir uma distribuição Exponencial padrão se o modelo for adequado. Tal objetivo será alcançado com o auxílio dos objetivos específicos que são criar diferentes cenários a partir da combinação de diferentes distribuições de probabilidade, porcentagens de censura e presença ou não de covariáveis.

Para a realização deste trabalho, será considerada a metodologia de simulação de Monte Carlo utilizando o software estatístico R (R Core Team, 2017) de forma a tornar possível a estimação de parâmetros e análise do comportamento dos resíduos dos modelos.

A organização do trabalho é feita de maneira que na Seção 2 são apresentados alguns conceitos de Análise de Sobrevida, distribuições de probabilidade, modelos de localização e escala, estimação de parâmetros e avaliação da qualidade deles e análise de resíduos, enquanto a Seção 3 traz explicações a respeito dos dados utilizados e da metodologia utilizada para alcançar os objetivos do trabalho bem como a revisão bibliográfica de conceitos e definições. Já a Seção 4 traz os resultados obtidos e a discussão dos mesmos e, por fim, a quinta e última seção apresenta as considerações finais do trabalho.

## 2 REVISÃO DE LITERATURA

Nesta seção são apresentados alguns conceitos de Análise de Sobrevida importantes para a realização do estudo, de forma que seja possível atingir o objetivo principal: estudar o comportamento do resíduo de Cox-Snell com dados que seguem diferentes distribuições de probabilidade.

### 2.1 Análise de Sobrevida

Análise de Sobrevida é uma das ramificações da Estatística, cuja aplicação pode ser visto desde à Medicina, passando pela área de Finanças e chegando às Engenharias, motivo pelo qual também pode ser chamada de Análise de Confiabilidade.

Seu objeto de estudo é o tempo decorrido até à ocorrência do evento de interesse, o tempo de falha. No entanto, em alguns casos, tem-se apenas observações parciais desse tempo, as quais são denominadas censuras.

A censura é uma particularidade da Análise de Sobrevida e se caracteriza pela não observação do evento de interesse por algum motivo não relacionado com o mesmo. Nesses casos não se tem o tempo decorrido até o evento de interesse, mas tem-se o tempo até a ocorrência da censura e, apesar de incompleta, essa informação é muito útil para a análise.

As censuras podem ser classificadas da seguinte maneira:

#### 1. Censura intervalar

A censura intervalar é um tipo mais geral de censura que acontece em estudos que os elementos da amostra são observados em ocasiões periódicas e sabe-se apenas que o evento de interesse ocorreu em um determinado intervalo de tempo. As censuras à esquerda e direita são casos particulares da censura intervalar.

#### 2. Censura à esquerda

Segundo Colosimo (2006), a censura à esquerda ocorre quando tempo registrado é maior do que o tempo de falha. Isto é, o evento de interesse já havia ocorrido quando o indivíduo foi observado. Como exemplo, pode-se citar um estudo para determinar a idade em que as crianças aprendem a ler em uma determinada comunidade. Quando os pesquisadores começaram a pesquisa, algumas crianças já sabiam ler e não lembravam com que idade aprenderam.

### 3. Censura à direita

A censura à direita ocorre quando o tempo registrado é menor do que o tempo de falha. Isto é, o evento de interesse ainda não havia ocorrido quando o indivíduo foi observado. Esse tipo de censura pode ser subdividido em três categorias:

Censura tipo I: Ocorre nos casos em que o tempo de duração do experimento é previamente fixado. Assim, o tempo de sobrevivência,  $T$ , é observado se o mesmo for menor que o tempo de censura,  $C$ ; caso contrário, sabe-se apenas que o tempo de falha foi maior que o de censura. Esse tipo de censura é predominante em práticas médicas.

Censura tipo II: Ocorre nos casos em que o estudo é conduzido até que um número pré-definido de elementos falhe. O pesquisador determina um número  $k$  ( $k < n$ ) de falhas e observa as unidades em estudo até que as  $k$  falhas aconteçam. Esse tipo de censura é predominante em experimentos industriais e da Engenharia.

Censura aleatória: Ocorre quando um ou mais indivíduos não podem ser acompanhados até o final do estudo ou ainda quando estes falharem por eventos alheios ao de interesse. Esse tipo de censura ocorre naturalmente, sem manipulação do pesquisador. Censuras por perda de acompanhamento, término do estudo e falha devido à outra causa são exemplos de censuras aleatórias.

Seja  $T$  o tempo de sobrevivência e  $C$  o tempo de censura. Utilizando o mecanismo de censura à direita aleatória, tem-se a variável resposta definida como

$$t = \min(T, C).$$

Dessa forma, faz-se necessária a introdução de uma variável dicotômica na análise que indique se o tempo de falha foi ou não observado. Tal variável é denominada variável indicadora de censura, ou apenas censura e é definida da seguinte maneira:

$$\delta = \begin{cases} 0, & \text{se } T > C, \\ 1, & \text{se } T \leq C. \end{cases}$$

No presente trabalho, será considerada a censura à direita aleatória.

### 2.1.1 Função de Sobrevivência

Seja  $T$  uma variável aleatória positiva referente ao tempo de falha e  $f(t)$  a função de densidade. Segundo Colosimo, a função de sobrevivência é definida como a probabilidade de uma observação não falhar até um certo tempo  $t$ . Denotada por  $S(t)$ , é matematicamente definida por:

$$S(t) = P[T > t] = 1 - F(t),$$

sendo  $F(t)$  a função de distribuição acumulada de  $T$ .

### 2.1.2 Função de Risco

Seja  $T$  uma variável aleatória positiva referente ao tempo de falha. A função de risco,  $h(t)$ , também chamada de função de taxa de falha, representa o risco instantâneo que o indivíduo tem de experimentar o evento de interesse em um determinado tempo  $t$ . Assim, a função de taxa de falha é definida por:

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

### 2.1.3 Estimador de Kaplan-Meier

O primeiro passo para analisar qualquer banco de dados é a análise exploratória das observações que o compõem a partir de gráficos e medidas descritivas. No entanto, devido à censura, não é possível utilizar determinadas técnicas tradicionais de exploração de dados, pois o tempo de censura informa somente que o tempo de falha do indivíduo em questão é maior do que o registrado (Colosimo e Giolo, 2006). Assim, as estimativas são feitas a partir de um estimador empírico, o estimador de Kaplan-Meier ou produto-limite.

Considerando um estudo envolvendo  $n$  indivíduos e os tempos de sobrevivência  $t_1, t_2, \dots, t_r$ , distintos, tem-se que a função de sobrevivência,  $S(t)$ , é estimada por:

$$\widehat{S}(t) = \prod_{j:t_j < t} \left(1 - \frac{d_j}{n_j}\right),$$

em que  $n_j$  é o número de indivíduos que estão sob risco no tempo  $t_j$  e  $d_j$  é o número de indivíduos que experimentaram o evento de interesse no tempo  $t_j, j = 1, 2, \dots, r$ .

Como principais propriedades desse estimador pode-se citar que ele é estimador de máxima verossimilhança de  $S(t)$ , é considerado não-viciado para amostras grandes, é

fracamente consistente e converge assintoticamente para um processo gaussiano.

Uma das importâncias práticas desse estimador é o auxílio no estudo da função de sobrevivência. Com base nessa função, decide-se o modelo paramétrico a ser utilizado, não em termos de distribuição de probabilidade, mas sim de classes de modelos com os quais podemos trabalhar. Por exemplo, nos casos em que os dados se distribuem de maneira contínua, espera-se que poucos indivíduos falhem ao mesmo tempo, isto é,  $d_j$  assumindo valores pequenos. Se isso não ocorre, pode ser indicativo de que:

- o modelo mais adequado seja o de fração de cura;
- os dados se distribuam discretamente;
- ou que seja necessário utilizar dados grupados, quando pode-se assumir que os dados se distribuem continuamente.

## 2.2 Distribuições de Probabilidade

Quando se pensa em fazer modelagem para dados de sobrevivência, deve-se considerar a distribuição de probabilidade que a variável resposta pode assumir.

Tradicionalmente, na Estatística o modelo normal é muito utilizado. No entanto, devido à particularidade da variável resposta, de assumir valores positivos, são consideradas apenas distribuições que assumem valores nesse espaço, como a Exponencial, Weibull, Log-logística e Log-normal.

### 2.2.1 Distribuição Exponencial

A distribuição exponencial é um dos modelos probabilísticos mais simples usados para descrever o tempo de sobrevivência. Apresenta um único parâmetro e é a única que apresenta uma taxa de falha constante, isto é, tanto uma unidade mais velha quanto uma nova, que ainda não apresentaram falha, apresentam a mesma taxa de falha em um intervalo futuro.

Seja  $T$  uma variável aleatória referente ao tempo de falha. Se  $T$  segue uma distribuição exponencial, sua função de densidade é dada por:

$$f(t) = \frac{1}{\alpha} \exp \left\{ -\frac{t}{\alpha} \right\}, \quad t \geq 0,$$

sendo  $\alpha > 0$  o tempo médio de vida.

Além disso, as funções de sobrevivência,  $S(t)$ , e de taxa de falha,  $h(t)$ , são dadas, respectivamente, por:

$$S(t) = \exp \left\{ -\frac{t}{\alpha} \right\}$$

e

$$h(t) = \frac{1}{\alpha},$$

para  $t \geq 0$ .

### 2.2.2 Distribuição Weibull

A distribuição de Weibull é frequentemente utilizada em estudos biomédicos e industriais. Isso se deve ao fato dela apresentar uma grande variedade de formas e uma função de taxa de falha monótona, isto é, ela é crescente, decrescente ou constante.

Seja  $T$  uma variável aleatória referente ao tempo de falha. Se  $T$  segue uma distribuição de Weibull, sua função de densidade de probabilidade é dada por:

$$f(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1} \exp \left\{ -\left(\frac{t}{\alpha}\right)^\gamma \right\}, \quad t \geq 0,$$

sendo  $\gamma$  o parâmetro de forma e  $\alpha$  o de escala, ambos positivos.

As funções de sobrevivência,  $S(t)$ , e de taxa de falha,  $h(t)$ , são dadas, respectivamente, por:

$$S(t) = \exp \left\{ -\left(\frac{t}{\alpha}\right)^\gamma \right\}$$

e

$$h(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1},$$

para  $t \geq 0$ ,  $\gamma > 0$  e  $\alpha > 0$ .

Quando o parâmetro de forma for igual a 1, tem-se como caso particular a distribuição Exponencial, com função de taxa de falha constante. Quando  $\gamma > 1$ , essa função é crescente e para valores menores do 1, ela é decrescente.

### 2.2.3 Distribuição Log-normal

A distribuição log-normal é muito utilizada para descrever o tempo de vida de produtos e indivíduos, assim como a distribuição de Weibull.

Seja  $T$  uma variável aleatória referente ao tempo de falha de um indivíduo. Se  $T$  segue uma distribuição log-normal, sua função de densidade é dada por:

$$f(t) = \frac{1}{\sqrt{2\pi t\sigma}} \exp \left\{ -\frac{1}{2} \left( \frac{\log(t) - \mu}{\sigma} \right)^2 \right\}, \quad t > 0,$$

sendo  $-\infty < \mu < \infty$  e  $\sigma > 0$ , respectivamente, a média e o desvio-padrão do logaritmo do tempo de falha.

As funções de sobrevivência,  $S(t)$ , e de risco,  $h(t)$ , não apresentam forma analítica explícita, sendo representadas, respectivamente, por:

$$S(t) = \Phi \left( \frac{-\log(t) + \mu}{\sigma} \right)$$

e

$$h(t) = \frac{f(t)}{S(t)} = \frac{\frac{1}{\sqrt{2\pi t\sigma}} \exp \left\{ -\frac{1}{2} \left( \frac{\log(t) - \mu}{\sigma} \right)^2 \right\}}{\Phi \left( \frac{-\log(t) + \mu}{\sigma} \right)},$$

sendo  $\Phi(\cdot)$  a função de distribuição acumulada de uma normal padrão.

Diferentemente da distribuição de Weibull, essa distribuição contempla funções de risco unimodais.

#### 2.2.4 Distribuição Log-logística

A distribuição log-logística se apresenta, em algumas situações práticas, como uma alternativa à de Weibull e à log-normal.

Seja  $T$  uma variável aleatória referente ao tempo de falha de um indivíduo. Se  $T$  segue uma distribuição log-logística, sua função de densidade de probabilidade é dada por:

$$f(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1} \left( 1 + \left( \frac{t}{\alpha} \right)^\gamma \right)^{-2}, \quad t > 0,$$

sendo  $\alpha > 0$  o parâmetro de escala e  $\gamma > 0$  o de forma.

As funções de sobrevivência,  $S(t)$ , e de risco,  $h(t)$ , são dadas, respectivamente, por:

$$S(t) = \frac{1}{1 + \left( \frac{t}{\alpha} \right)^\gamma}$$

e

$$h(t) = \frac{\gamma \left( \frac{t}{\alpha} \right)^{\gamma-1}}{\alpha \left[ 1 + \left( \frac{t}{\alpha} \right)^\gamma \right]}.$$

Assim como a distribuição log-normal, a log-logística contempla funções de risco unimodais. Além disso, uma vantagem da log-logística em relação à log-normal é que a primeira tem funções de sobrevivência e risco com formas definidas.

## 2.3 MODELOS DE REGRESSÃO

No estudo de dados de sobrevivência, é comum ocorrer situações em que existem variáveis associadas ao tempo de vida. Essas variáveis são chamadas covariáveis e podem representar tanto a heterogeneidade dos dados quanto a diferença de tratamento recebida pelos indivíduos. Na indústria, por exemplo, o tempo de falha de um equipamento pode ser afetado pela voltagem à qual o equipamento é submetido, enquanto na área médica o tempo de vida de um paciente pode ser influenciado pela idade ou tipo de tumor.

Uma das maneiras de se estudar o efeito das covariáveis no tempo de sobrevivência é fazer a reparametrização de um dos parâmetros do modelo de forma que a covariável seja incluída e, assim, seja possível avaliar sua influência.

Outra maneira de fazer isso é considerar duas classes de modelos de regressão: os modelos paramétricos, ou modelos de tempo de falha acelerado, e os semi-paramétricos, também chamados de modelos de riscos proporcionais ou modelo de regressão de Cox.

### 2.3.1 Modelo de Regressão Log-linear

Segundo Lawless (2002), um modelo log-linear é definido da seguinte maneira:

$$Y = \log(T) = \mu + \sigma W, \quad -\infty < Y < \infty.$$

Considerando-se que o parâmetro  $\mu$  depende de um vetor de covariáveis  $\mathbf{x}$  e, por isso, pode ser escrito como  $\mu(\mathbf{x}) = \mathbf{x}^T \boldsymbol{\beta}$ , em que  $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)^T$  é um vetor de parâmetros desconhecidos, pode-se escrever o modelo expresso acima da seguinte maneira:

$$Y = \mu(\mathbf{x}) + \sigma W,$$

em que  $Y = \log(T)$ ,  $W$  é o erro aleatório e  $\mu(\mathbf{x}) = \mathbf{x}^T \boldsymbol{\beta}$ .

Vale ressaltar que o modelo é log-linear para  $T$  e, portanto, é um modelo de regressão linear para  $Y$ . Além disso, uma característica do modelo é que o vetor de covariáveis tem efeito multiplicativo em  $T$ , isto é,  $T = \exp(\mathbf{x}^T \boldsymbol{\beta}) \exp(\sigma W)$  e, dessa forma, tem efeito linear em  $Y$ .

A variável  $Y$  faz parte de uma família de distribuições caracterizada pelo fato de que  $\mu$  ( $-\infty < \mu < \infty$ ) é um parâmetro de locação e,  $\sigma$  ( $0 < \sigma < \infty$ ), um parâmetro de escala. A função de densidade das distribuições da família de locação e escala tem a seguinte forma

$$f(y; \mu; \sigma) = \frac{1}{\sigma} g\left(\frac{y - \mu}{\sigma}\right) = \frac{1}{\sigma} g\left(\frac{y - \mathbf{x}^T \boldsymbol{\beta}}{\sigma}\right), \quad -\infty < y < \infty,$$

e a função de sobrevivência pode ser escrita como  $G\left(\frac{y - \mu}{\sigma}\right)$ .

As distribuições de probabilidade comentadas na seção anterior fazem parte da família de modelos de locação-escala e, ao considerar a transformação  $Y = \log(T)$  e algumas reparametrizações nos parâmetros, são obtidos os seguintes modelos de locação-escala:

- Modelo de regressão log-Weibull

Ao considerar  $Y = \log(T)$ ,  $\alpha = \exp(\mathbf{x}^T \boldsymbol{\beta})$  e  $\gamma = \frac{1}{\sigma}$ , as funções de densidade de probabilidade e sobrevivência de  $Y$  se  $T$  segue uma distribuição Weibull são dadas por:

$$f(y) = \frac{1}{\sigma} \exp\left\{-\left(\frac{y - \mathbf{x}^T \boldsymbol{\beta}}{\sigma}\right)^\gamma - \exp\left(-\left(\frac{y - \mathbf{x}^T \boldsymbol{\beta}}{\sigma}\right)^\gamma\right)\right\}$$

e

$$S(y) = \exp\left\{-\exp\left(-\left(\frac{y - \mathbf{x}^T \boldsymbol{\beta}}{\sigma}\right)^\gamma\right)\right\}.$$

Dessa forma, o modelo para  $Y$  é da seguinte forma:

$$Y = \mathbf{x}^T \boldsymbol{\beta} + \sigma W,$$

em que  $W$  segue uma distribuição do Valor Extremo padrão com funções de densidade de probabilidade e sobrevivência definidas como

$$f_0(w) = \exp(w - e^w) \quad \text{e} \quad S_0(w) = \exp(-e^w).$$

- Modelo de regressão normal

Considerando  $Y = \log(T)$  e os parâmetros  $\mu = \mathbf{x}^T \boldsymbol{\beta}$  e  $\sigma$ , as funções de densidade de probabilidade e sobrevivência de  $Y$  se  $T$  segue uma distribuição Normal são dadas por:

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{y - \mathbf{x}^T \boldsymbol{\beta}}{\sigma}\right)^2\right]$$

e

$$S(y) = 1 - \phi\left(\frac{y - \mathbf{x}^T \boldsymbol{\beta}}{\sigma}\right).$$

O modelo de regressão Normal, também conhecido como modelo de tempo de vida acelerado, pertence à classe de modelos de locação e escala, de forma que  $W$  segue uma distribuição Normal padrão. Da mesma maneira que o modelo anterior, considera-se o modelo  $Y = \mathbf{x}^T \boldsymbol{\beta} + \sigma W$  e, dessa forma, tem-se que as funções de densidade de probabilidade e sobrevivência de  $W$  são definidas da seguinte maneira:

$$f_0(w) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-w^2}{2}\right) \quad \text{e} \quad S_0(w) = 1 - \phi(w).$$

- Modelo de regressão logístico

Seja  $Y = \log(T)$ ,  $\alpha = \exp(\mathbf{x}^T \boldsymbol{\beta})$  e  $\gamma = \frac{1}{\sigma}$ . Se  $T$  segue uma distribuição log-logística, as funções de densidade de probabilidade e de sobrevivência de  $Y$  são definidas da seguinte maneira:

$$f(y) = \frac{1}{\sigma} \exp\left(\frac{y - \mathbf{x}^T \boldsymbol{\beta}}{\sigma}\right) \left[1 + \exp\left(\frac{y - \mathbf{x}^T \boldsymbol{\beta}}{\sigma}\right)\right]^{-2}$$

e

$$S(y) = \left[1 + \exp\left(\frac{y - \mathbf{x}^T \boldsymbol{\beta}}{\sigma}\right)\right]^{-1}.$$

O modelo de regressão logístico, também conhecido como modelo de tempo de vida acelerado, tem forma de locação-escala de maneira que  $W$  segue uma distribuição Logística padrão. Da mesma maneira que o modelo anterior, considera-se o modelo  $Y = \mathbf{x}^T \boldsymbol{\beta} + \sigma W$  e, dessa forma, tem-se que as funções de densidade de probabilidade e sobrevivência de  $W$  são definidas da seguinte maneira:

$$f_0(w) = \frac{e^w}{(1 + e^w)^2} \quad \text{e} \quad S_0(w) = \frac{1}{(1 + e^w)}.$$

## 2.4 Estimação de Parâmetros

Tendo determinado a distribuição de probabilidade, é necessário então realizar a estimação dos parâmetros do modelo. Há dois métodos de estimação que são muito conhecidos e adotados na literatura: o Método de Mínimos Quadrados e o Método de Máxima Verossimilhança.

Devido às censuras, é necessário um método que seja capaz de incorporar todas as informações disponíveis. Dessa forma, o Método de Mínimos Quadrados se torna inadequado, visto que não é possível incorporar a censura na função que deve ser maximizada.

Por sua vez, o Método da Máxima Verossimilhança, cujo objetivo é encontrar o valor do parâmetro que maximiza a probabilidade da amostra observada ocorrer, se mostra adequado, pois permite considerar a função de sobrevivência e contribuição das censuras na função de verossimilhança. A função de densidade delas corresponde aos tempos de falha.

Independente do mecanismo de censura à direita adotado, a expressão para a função de verossimilhança é a mesma, dada por (Colosimo e Giolo, 2006):

$$L(\theta) \propto \prod_{i=1}^n [f(t_i; \theta)]^{\delta_i} [S(t_i; \theta)]^{1-\delta_i},$$

sendo  $\delta_i$  a variável indicadora de censura.

Ao aplicar o logaritmo na função de verossimilhança na equação acima, tem-se:

$$l(\theta) = \sum_{i=1}^n [\delta_i \log [f(t_i; \theta)] + (1 - \delta_i) \log [S(t_i; \theta)]] + c,$$

em que  $c$  é uma constante que não depende de  $\theta$ .

Os estimadores de máxima verossimilhança são os valores que maximizam  $L(\theta)$ , ou equivalentemente  $l(\theta)$  e são obtidos resolvendo o sistema de equações

$$\frac{dl(\theta)}{d\theta} = 0$$

.

## 2.5 Análise de Resíduos

A avaliação do ajuste do modelo ajustado é de suma importância na análise dos dados. Esse passo tem como objetivo principal examinar a adequação da distribuição considerada para a variável resposta, verificar as suposições básicas do modelo assim como detectar a presença de pontos extremos, observar a relevância de um fator omitido e analisar a forma funcional do modelo.

Segundo Klein e Moeschberger (1997), essas técnicas devem ser aplicadas de maneira a rejeitar modelos claramente inapropriados e não para provar que um modelo específico está correto.

Em sobrevivência, devido à presença de observações censuradas, uma maneira de se fazer a análise é utilizando os resíduos de Cox-Snell.

### 2.5.1 Resíduos de Cox-Snell

Os resíduos de Cox-Snell auxiliam no exame do ajuste global do modelo e são determinados por:

$$\hat{e}_i = \hat{H}(t_i|\mathbf{x}_i) \quad \text{ou} \quad \hat{e}_i = \hat{H}(y_i|\mathbf{x}_i)$$

sendo  $\hat{H}(\cdot)$  a função de taxa de falha acumulada obtida a partir do modelo ajustado,  $t_i$  os tempos de falha dos indivíduos e  $y_i$  a variável originada a partir da transformação  $Y = \log(T)$ .

Os resíduos  $\hat{e}_i$  vêm de uma população homogênea e devem seguir uma distribuição exponencial padrão se o modelo for adequado (Lawless, 2002). Assim, o uso de técnicas gráficas para a análise da qualidade do modelo se torna possível da seguinte maneira:

- O gráfico da função de sobrevivência dos resíduos estimada por Kaplan-Meier versus a função de sobrevivência do modelo exponencial padrão deve ser aproximadamente uma reta com inclinação 1;
- Ou a curva de sobrevivência desse resíduo estimada por Kaplan-Meier e a curva de sobrevivência do modelo exponencial devem estar próximas.

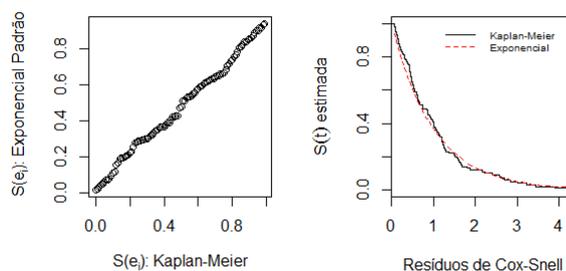


Figura 1 – Ilustração do resíduo de Cox-Snell para o modelo exponencial.

No gráfico acima são apresentados os resultados em uma situação ideal em que os dados se distribuem segundo uma Exponencial e não há censura entre eles.

## 2.6 Avaliação de Estimadores

Seja  $\theta$  um parâmetro numa população finita ou em um modelo de interesse formulado para descrever aspectos dessa população.

A qualidade de um estimador  $W$  para  $\theta$ , sob um plano amostral definido, é usualmente avaliada por duas medidas: o erro quadrático médio e o vício do estimador.

Segundo Casella (2011), o erro quadrático médio (EQM) mede a diferença quadrática média entre o estimador  $W$  e o parâmetro  $\theta$  e é calculado da seguinte maneira:

$$EQM_{\theta} = E_{\theta}(W - \theta)^2 = \text{Var}_{\theta}W + (E_{\theta}W - \theta)^2 = \text{Var}_{\theta}W + (\text{vício}_{\theta}W)^2,$$

em que  $\text{Var}_{\theta}W$  e  $E_{\theta}W$  são, respectivamente, a variância e valor esperado de  $W$ .

Casella (2011) ainda define o vício de um estimador pontual  $W$  de um parâmetro  $\theta$  como a diferença entre o valor esperado de  $W$  e  $\theta$ , isto é,

$$\text{vício}_{\theta} = E_{\theta}W - \theta.$$

Um estimador é dito não-viesado para  $\theta$  se

$$E_{\theta}W = \theta,$$

isto é, se  $\text{vício}_{\theta} = 0$  para todo  $\theta \in \Theta$ . Pode-se dizer ainda que, se  $\lim_{n \rightarrow \infty} \text{vício}_{\theta} = 0$ , o estimador  $W$  é assintoticamente não viciado para  $\theta$ . Quando  $W$  é não-viesado, tem-se que:

$$EQM_{\theta} = \text{Var}_{\theta}W.$$

É desejado que se observe as seguintes propriedades no estimador:

- i.  $\lim_{n \rightarrow \infty} \text{vício}_{\theta} = 0$ ;
- ii.  $\lim_{n \rightarrow \infty} EQM_{\theta} = 0$ .

Isto é, espera-se que o estimador seja assintoticamente não viciado e que tenha mínimo erro quadrático médio.

Dessa forma, ao verificar essas propriedades para um estimador, pode-se comprovar a acurácia do modelo proposto.

### 3 METODOLOGIA

#### 3.1 Material

Neste trabalho serão considerados dados simulados. Os diferentes cenários de simulação serão definidos ao utilizar diferentes distribuições de probabilidade, tamanhos de amostra e porcentagens de censura.

#### 3.2 Métodos

No trabalho em questão, a variável resposta será criada por meio de uma simulação de dados e adotando-se o mecanismo de censura à direita.

Pode-se descrever simulação de dados, dentre outras definições, como uma técnica para analisar a distribuição empírica das medidas de resíduo propostas quando os dados são submetidos a algum modelo. No processo de simulação, tem-se controle da distribuição dos dados e, dessa forma, é possível saber se as suposições utilizadas no modelo são exatas e, em caso positivo, pode-se utilizar o conhecimento a respeito do comportamento das medidas de resíduo para ratificar o uso dos modelos propostos e as suposições consideradas quando os mesmos são aplicados a dados reais.

Neste trabalho, a simulação de dados será utilizada para criar diferentes cenários a partir da combinação de diferentes tamanhos de amostras, distribuições de probabilidade, percentuais de censuras e presença ou ausência de covariáveis. Dessa forma, serão consideradas amostras de 30, 50, 100, 300 e 500 indivíduos com tempos de sobrevivência gerados a partir das distribuições Exponencial, Weibull, Log-normal e Log-logística, percentuais de censura de 0%, 10%, 30% e 50%. Além disso, vale ressaltar que, para cada combinação de distribuição de probabilidade, tamanho de amostra, percentual de censura e presença ou não de covariável, foram geradas 1.000 amostras ( $M = 1.000$ ).

Os tempos de vida e a censura foram gerados no software *R* (R Core Team, 2017) utilizando funções já definidas no programa baseadas no método de inversão (ou de transformação inversa) para geração de número aleatórios.

Esse método é útil quando o objetivo é gerar valores randômicos  $x_i$  de alguma população estatística particular, com função de distribuição  $F$ , ou seja, gerar uma variável aleatória  $X$  com a propriedade  $F_X(x) = P(X \leq x)$  para todo  $x$ . Para isso, primeiramente é gerada uma variável aleatória  $U$  tal que  $U \sim \text{Uniforme}[0, 1]$  e, em seguida, estabelece-se que

$X = F_X^{-1}(U)$ . Assim, tem-se que

$$P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x).$$

A censura foi gerada considerando-se uma distribuição Uniforme cujos parâmetros são modificados de acordo com a porcentagem de dados censurados.

Tendo definido a variável resposta, parte-se então para a modelagem. Nesse momento, serão simuladas covariáveis quantitativas de maneira que seja possível ajustar modelos sem covariáveis e com covariáveis quantitativas.

Em seguida, serão ajustados os modelos de regressão referentes às distribuições de probabilidade consideradas e, por fim, será feita a análise de resíduo do modelo ajustado com os dados simulados. Ao simular os dados de uma determinada distribuição de probabilidade e ajustar um modelo dessa distribuição a eles, espera-se que o modelo apresente um bom ajuste, ou seja, é esperado que os resíduos de Cox-Snell sigam uma distribuição Exponencial padrão. Ao fazer isso, este trabalho visa verificar a veracidade da informação de que, independentemente da distribuição de probabilidade dos dados, se o modelo se ajusta bem aos dados, o resíduo de Cox-Snell segue uma distribuição Exponencial padrão.

Uma sugestão deste trabalho é realizar a verificação comentada acima a partir do gráfico do quantil dos resíduos observados versus o quantil teórico da distribuição exponencial padrão, uma vez que o mesmo possibilita a constatação de afastamento ou não da suposição para a distribuição dos resíduos de Cox-Snell, da seguinte maneira:

- i. Gera-se uma amostra de  $n$  observações de determinada distribuição de probabilidade;
- ii. Ajusta-se um modelo aos dados gerados usando  $(\delta_i, \mathbf{x}_i)$  do conjunto de dados e calcula-se o valor dos resíduos;
- iii. Repete-se os dois primeiros passos  $m$  vezes, sendo  $m$  o número de amostras que se deseja gerar;
- iv. Ordena-se os valores dos resíduos de cada amostra, formando  $m$  conjuntos das  $n$  estatísticas de ordem. É importante ressaltar que, para diferenciar se a observação falhou ou foi censurada, considerou-se o resíduo como  $e_i$  se o tempo de falha foi observado e  $e_i + 1$  se o tempo foi censurada como sugere Colosimo e Giolo (2006).

- v. Para cada uma das estatísticas de ordem, calcula-se a média dos  $m$  conjuntos e, por fim, é feito o gráfico dessas médias contra as estatísticas de ordem da distribuição Exponencial padrão.

Dessa forma, tem-se o seguinte gráfico para a situação definida na Seção 2.5.1.

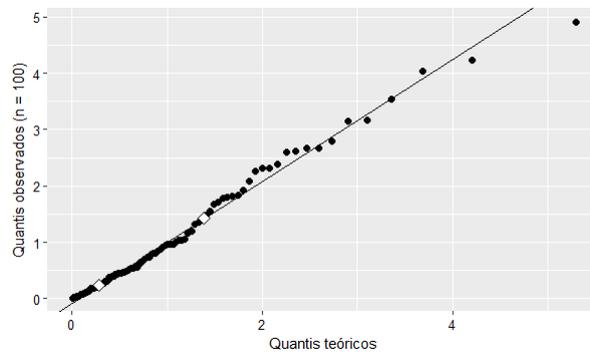


Figura 2 – Ilustração do gráfico exponencial de probabilidade para o resíduo Cox-Snell para o modelo exponencial.

Assim, se o modelo tiver um bom ajuste, espera-se que os gráficos não apontem nenhum afastamento sério da suposição de que os resíduos se distribuem segundo uma Exponencial padrão.

Agora, será ilustrado o passo-a-passo para a construção de todos os cenários de simulação dos resultados no software *R* (R Core Team, 2017). Considerando que os dados se distribuem segundo uma Exponencial Padrão, com  $\alpha = 0,5$ ,  $n = 30$ , 10% de censura e apenas uma replicação do processo ( $M = 1$ ), tem-se os seguintes passos:

- 1º) Para gerar os tempos, utiliza-se a função *rexp* do software *R* (R Core Team, 2017), em que o parâmetro  $n$  é o número de tempos que se deseja gerar e o parâmetro  $r$  é tal que

$$\alpha = \frac{1}{r}.$$

```
n <- 30
r <- 2
set.seed(12)
tempo <- rexp(100,2)
```

- 2º) Na geração das censuras, considera-se uma distribuição Uniforme cujos parâmetros variam de acordo com a porcentagem de censura desejada, de forma que, quanto maior a porcentagem de censura, menor deve ser o  $k$  escolhido.

```

k <- 6.8
r <- 2
set.seed(11)
censura <- runif(n,0,k*max(tempo))
delta <- ifelse(tempo < censura,1,0)

```

3º) Para estimar o modelo exponencial, utiliza-se a função *survreg* do pacote *survival* (Therneau T, 2015).

```

require(survival)
mod <- survreg(Surv(tempo,delta)~1,dist="exponential")
alpha <- exp(mod$coefficients[1])
surv <- exp(-(tempo/alpha2))

```

4º) O cálculo do resíduo de Cox-Snell e de seus quantis é feito da seguinte maneira:

```

ei <- -log(surv)
eit <- ifelse(delta1==1,ei,ei+1)
eio <- sort(eit)

```

5º) Para calcular os quantis teóricos e fazer o gráfico sugerido nesse trabalho, é utilizada a seguinte função:

```

qqq1 <- function(x, ref.line=T, distr=qexp, param=list(rate=1)){
x <- na.omit(x)
n <- length(x)
i <- seq_along(x)           # índices posicionais
pteo <- (i-0.5)/n           # probabilidades teóricas
qteo <- do.call(distr,      # quantis teóricos sob a distribuição
c(list(p=pteo), param))

if(ref.line){

qrto <- quantile(x, c(1,3)/4) # 1º e 3º quartis observados

qrtd <- do.call(distr,      # 1º e 3º quartis teóricos
c(list(p=c(1,3)/4), param))

}

require(ggplot2)

```

```
ggplot() +  
  geom_abline(intercept = 0, slope = 1) +  
  geom_point(aes(x = qteo, y = xo), size = 2, shape = 16) +  
  geom_point(aes(x = qrto, y = qrto), size = 3, shape = 23, fill = "white") +  
  labs(x = "Quantis teóricos", y = "Quantis observados (n = 30)") +  
  theme(axis.title.y = element_text(colour = "black",  
    size = 11.5, hjust = 0.5, angle = 90)) +  
  theme(axis.title.x = element_text(colour = "black",  
    size = 11.5, hjust = 0.5, angle = 0)) +  
  theme(axis.text = element_text(colour = "black", size = 9.5))  
}
```

Os resultados apresentados na seção a seguir são gerados seguindo os passos comentados acima.



## 4 RESULTADOS E DISCUSSÃO

Nesta seção são apresentados os resultados referentes ao estudo dos dados gerados a partir do método de simulação descrito na seção anterior.

Em um primeiro momento, será avaliada a qualidade do ajuste dos modelos para as distribuições consideradas no trabalho com ausência de covariável e, em seguida, será incluída uma covariável quantitativa gerada a partir de uma distribuição Uniforme(0, 1).

Vale ressaltar que, nos gráficos expostos a seguir, os dois pontos destacados em branco são referentes, respectivamente, ao primeiro e ao terceiro quartil.

### 4.1 Sem covariáveis

#### 4.1.1 Distribuição Exponencial

Para gerar os tempos de vida da distribuição Exponencial, o parâmetro considerado foi  $\alpha = 0,5$ . Além disso, a equação do resíduo para este cenário é definida da seguinte maneira:

$$\hat{e}_i = -\log(\hat{S}(t)) = \frac{t}{\hat{\alpha}}.$$

- Sem censura

Tabela 1 – Estimativas de  $\alpha$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\widehat{\alpha}$	EQM	Vício
30	0,5035	$8,37 \cdot 10^{-3}$	$3,48 \cdot 10^{-3}$
50	0,5003	$4,98 \cdot 10^{-3}$	$0,30 \cdot 10^{-3}$
100	0,5005	$1,60 \cdot 10^{-3}$	$0,50 \cdot 10^{-3}$
300	0,4999	$0,80 \cdot 10^{-3}$	$-0,03 \cdot 10^{-3}$
500	0,4992	$0,48 \cdot 10^{-3}$	$-0,82 \cdot 10^{-3}$

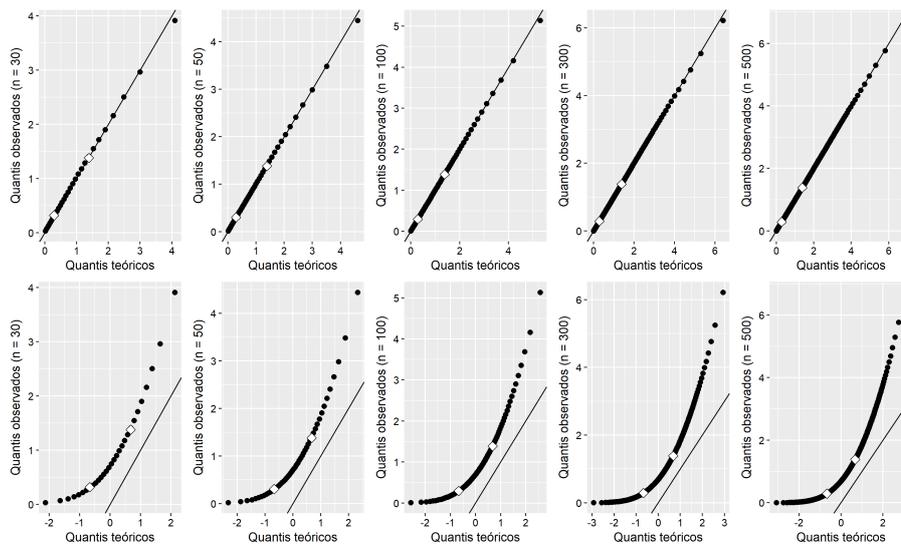


Figura 3 – Gráficos exponenciais e normais de probabilidade para os resíduos de Cox-Snell

- 10% censura

Tabela 2 – Estimativas de  $\alpha$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\widehat{\alpha}$	EQM	Vício
30	0,5574	$13,73 \cdot 10^{-3}$	$57,37 \cdot 10^{-3}$
50	0,5593	$10,42 \cdot 10^{-3}$	$59,30 \cdot 10^{-3}$
100	0,5556	$6,16 \cdot 10^{-3}$	$55,56 \cdot 10^{-3}$
300	0,5549	$4,08 \cdot 10^{-3}$	$55,49 \cdot 10^{-3}$
500	0,5555	$3,73 \cdot 10^{-3}$	$55,55 \cdot 10^{-3}$

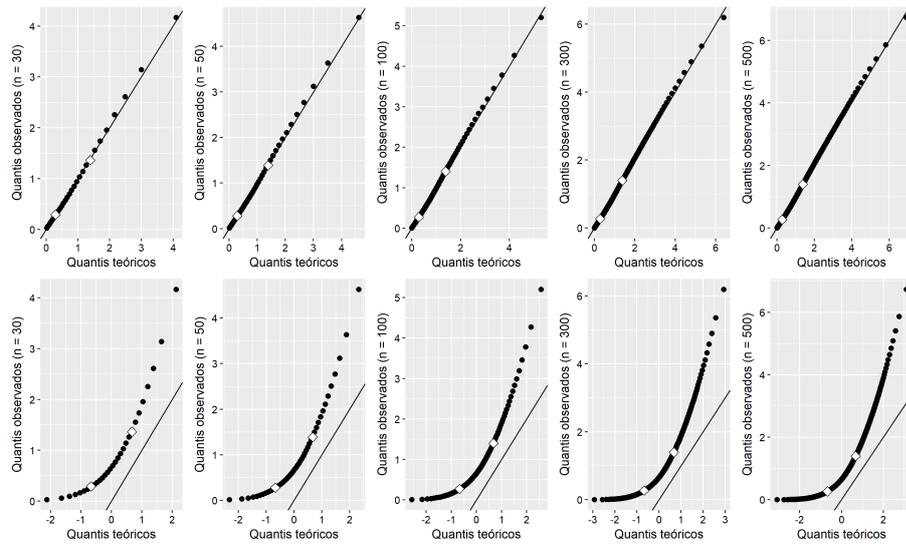


Figura 4 – Gráficos exponenciais e normais de probabilidade para os resíduos de Cox-Snell

- 30% censura

Tabela 3 – Estimativas de  $\alpha$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\widehat{\alpha}$	EQM	Vício
30	0,7157	$62,75 \cdot 10^{-3}$	$215,70 \cdot 10^{-3}$
50	0,7146	$56,76 \cdot 10^{-3}$	$214,62 \cdot 10^{-3}$
100	0,7130	$50,39 \cdot 10^{-3}$	$212,96 \cdot 10^{-3}$
300	0,7139	$47,51 \cdot 10^{-3}$	$213,87 \cdot 10^{-3}$
500	0,7135	$46,69 \cdot 10^{-3}$	$213,55 \cdot 10^{-2}$

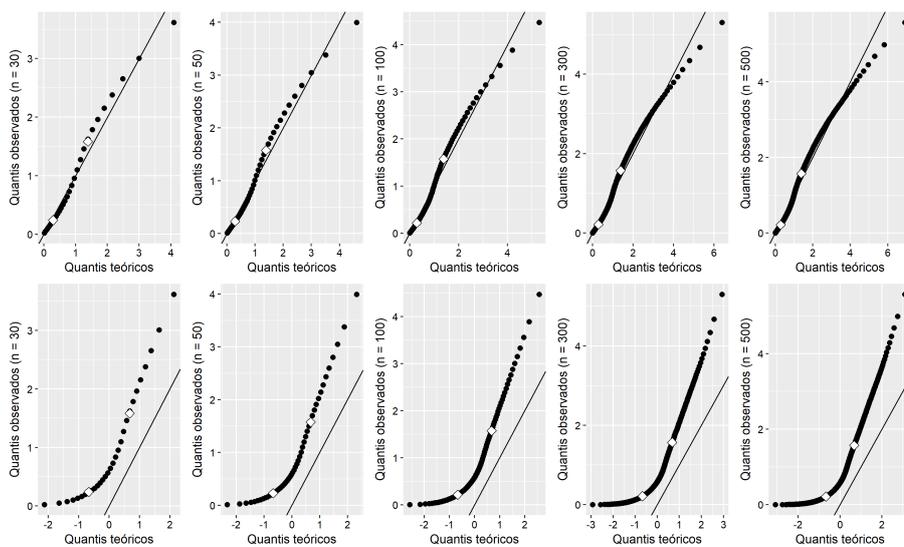


Figura 5 – Gráficos exponenciais e normais de probabilidade para os resíduos de Cox-Snell

- 50% censura

Tabela 4 – Estimativas de  $\alpha$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\widehat{\alpha}$	EQM	Vício
30	0,9969	278,79.10 <sup>-3</sup>	496,87.10 <sup>-3</sup>
50	1,0030	273,69.10 <sup>-3</sup>	503,02.10 <sup>-3</sup>
100	1,0064	265,57.10 <sup>-3</sup>	506,37.10 <sup>-3</sup>
300	0,9991	252,56.10 <sup>-3</sup>	499,14.10 <sup>-3</sup>
500	1,0012	253,12.10 <sup>-3</sup>	501,20.10 <sup>-3</sup>

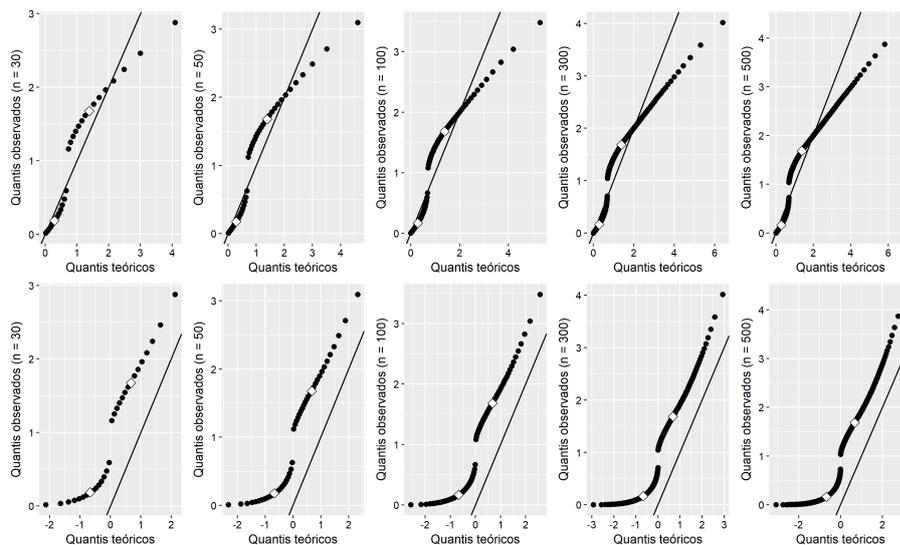


Figura 6 – Gráficos exponenciais e normais de probabilidade para os resíduos de Cox-Snell

Ao observar os valores de EQM e vício expostos nas Tabelas 1 a 4, pode-se perceber que o erro quadrático médio diminui com o aumento do tamanho da amostra, isto é, quanto maior a amostra, maior a qualidade do estimador. Por sua vez, o aumento do percentual de censura acarreta um aumento das medidas. Segundo Cardial (2017), este fato é esperado, pois as estimativas são naturalmente viesadas uma vez que a função de verossimilhança na presença de censuras conta com a distribuição da função de sobrevivência. Isto é, ao gerar os tempos de vida, a censura não é considerada, mas, ao iniciar o processo de modelagem, a mesma é incluída, causando o viés nas estimativas.

Os resultados comentados acima foram observados em todos os outros cenários e, por isso, serão omitidos nas próximas subseções.

Em relação aos gráficos dos resíduos de Cox-Snell, apresentados nas Figuras 3 a 6, os que estão na linha de cima, pode-se dizer que, nos cenários sem censura e com 10%

de dados censurados, é razoável assumir que a distribuição empírica dos resíduos apresenta concordância com uma distribuição Exponencial padrão, uma vez que os conjuntos de pontos não apresentam grandes desvios da reta de referência e nem cruzam a mesma. Nos cenários com 30 e 50% de censura, já são observados alguns desvios, como o cruzamento da reta de referência no primeiro cenário e um afastamento maior da suposição de que os dados se distribuem segundo uma Exponencial padrão no segundo.

A fim de comparação, foram construídos também os gráficos quantil-quantil da Normal padrão e dos resíduos (colocados na linha de baixo das figuras) e, ao observá-los, percebe-se que há indícios de afastamentos sérios da suposição de distribuição Normal padrão dos resíduos. Esses gráficos foram construídos apenas como uma curiosidade sobre a possibilidade de os resíduos também seguirem uma distribuição Normal padrão.

### 4.1.2 Distribuição Weibull

Para gerar os tempos de vida da distribuição Weibull, os parâmetros considerados foram  $\alpha = 2,5$  e  $\gamma = 2$ . Para o cálculo dos resíduos, a seguinte equação é utilizada:

$$\hat{e}_i = -\log(\hat{S}(t)) = \left(\frac{t}{\hat{\alpha}}\right)^{\hat{\gamma}}.$$

- Sem censura

Tabela 5 – Estimativas de  $\alpha, \gamma$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\hat{\gamma}$	EQM	Vício	$\hat{\alpha}$	EQM	Vício
30	2,1127	113,16.10 <sup>-3</sup>	112,67.10 <sup>-3</sup>	2,5055	60,79.10 <sup>-3</sup>	5,52.10 <sup>-3</sup>
50	2,0689	61,27.10 <sup>-3</sup>	68,96.10 <sup>-3</sup>	2,5051	37,98.10 <sup>-3</sup>	5,06.10 <sup>-3</sup>
100	2,0344	27,51.10 <sup>-3</sup>	34,40.10 <sup>-3</sup>	2,5008	17,69.10 <sup>-3</sup>	0,7.10 <sup>-3</sup>
300	2,0113	8,30.10 <sup>-3</sup>	11,27.10 <sup>-3</sup>	2,4995	5,65.10 <sup>-3</sup>	-0,51.10 <sup>-3</sup>
500	2,0061	4,95.10 <sup>-3</sup>	6,12.10 <sup>-3</sup>	2,4997	3,47.10 <sup>-3</sup>	-0,27.10 <sup>-3</sup>

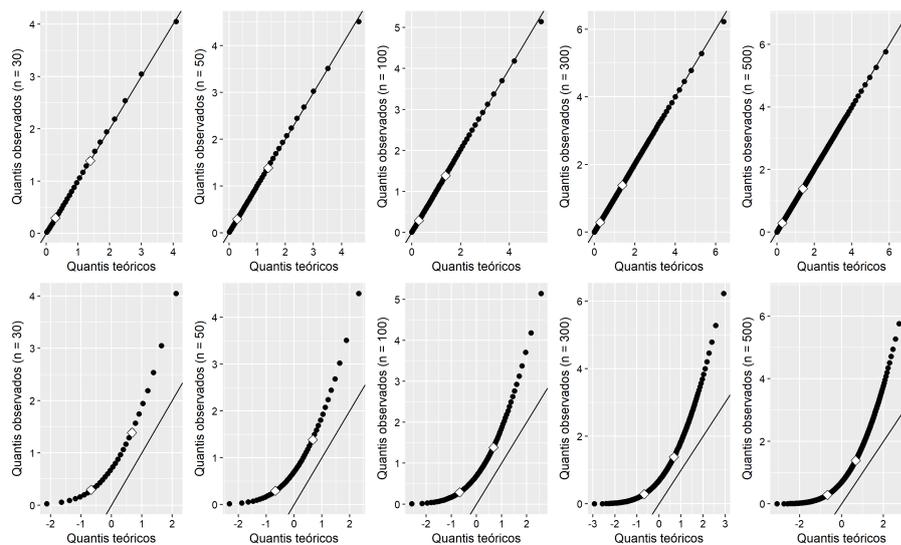


Figura 7 – Gráficos exponenciais e normais de probabilidade para os resíduos de Cox-Snell

- 10% de censura

Tabela 6 – Estimativas de  $\alpha, \gamma$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\widehat{\gamma}$	EQM	Vício	$\widehat{\alpha}$	EQM	Vício
30	2,0214	$98,65 \cdot 10^{-3}$	$21,37 \cdot 10^{-3}$	2,6089	$73,64 \cdot 10^{-3}$	$108,89 \cdot 10^{-3}$
50	1,9781	$57,35 \cdot 10^{-3}$	$-21,87 \cdot 10^{-3}$	2,6180	$51,02 \cdot 10^{-3}$	$118,02 \cdot 10^{-3}$
100	1,9173	$30,57 \cdot 10^{-3}$	$-82,68 \cdot 10^{-3}$	2,6051	$29,36 \cdot 10^{-3}$	$105,06 \cdot 10^{-3}$
300	1,9361	$13,06 \cdot 10^{-3}$	$-63,86 \cdot 10^{-3}$	2,6209	$20,45 \cdot 10^{-3}$	$120,81 \cdot 10^{-3}$
500	1,9275	$10,19 \cdot 10^{-3}$	$-72,53 \cdot 10^{-3}$	2,6198	$17,91 \cdot 10^{-3}$	$119,78 \cdot 10^{-3}$

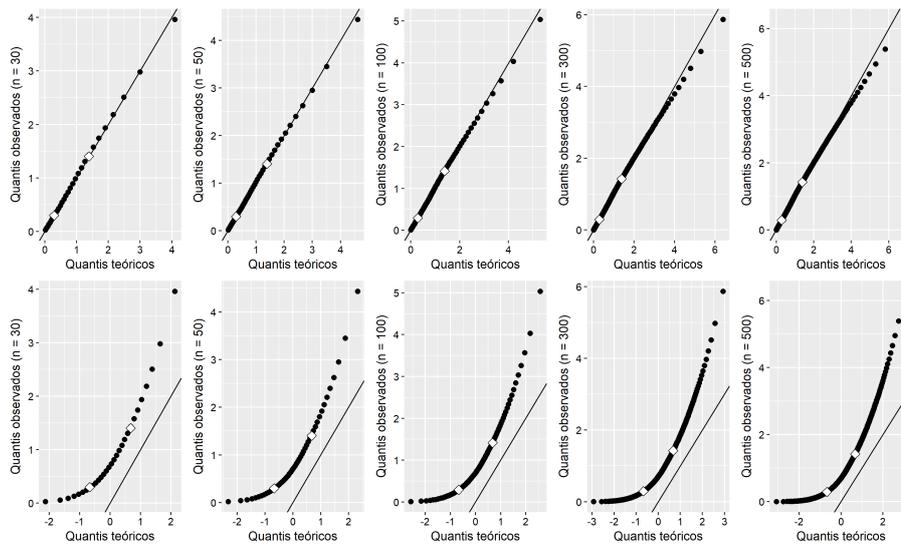


Figura 8 – Gráficos exponenciais e normais de probabilidade para os resíduos de Cox-Snell

- 30% de censura

Tabela 7 – Estimativas de  $\alpha, \gamma$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\widehat{\gamma}$	EQM	Vício	$\widehat{\alpha}$	EQM	Vício
30	1,8513	142,03.10 <sup>-3</sup>	-148,69.10 <sup>-3</sup>	2,9756	298,81.10 <sup>-3</sup>	475,56.10 <sup>-3</sup>
50	1,7869	107,45.10 <sup>-3</sup>	-213,14.10 <sup>-3</sup>	2,9797	276,04.10 <sup>-3</sup>	479,69.10 <sup>-3</sup>
100	1,7460	92,72.10 <sup>-3</sup>	-254,02.10 <sup>-3</sup>	2,9830	254,99.10 <sup>-3</sup>	482,96.10 <sup>-3</sup>
300	1,7330	80,82.10 <sup>-3</sup>	-266,96.10 <sup>-3</sup>	2,9820	239,58.10 <sup>-3</sup>	482,01.10 <sup>-3</sup>
500	1,7239	81,48.10 <sup>-3</sup>	-276,10.10 <sup>-3</sup>	2,9815	236,27.10 <sup>-3</sup>	481,52.10 <sup>-3</sup>

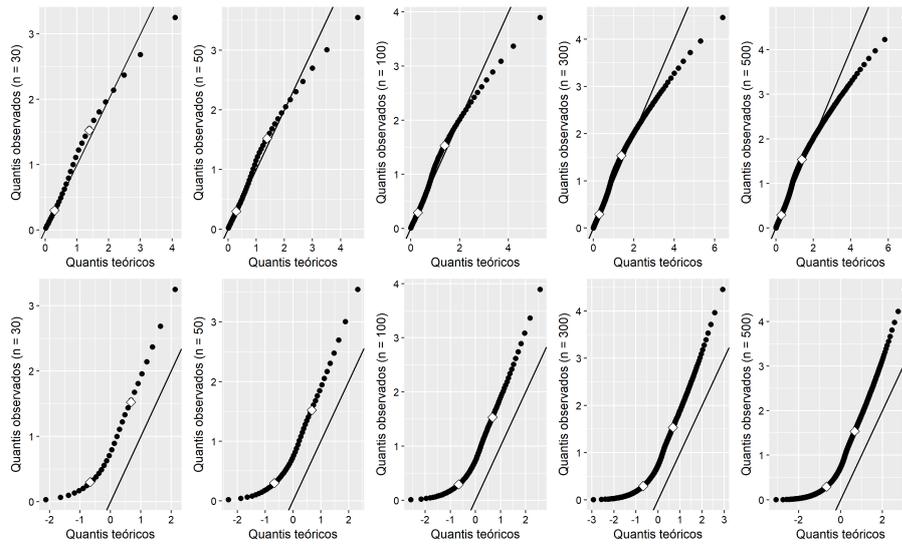


Figura 9 – Gráficos exponenciais e normais de probabilidade para os resíduos de Cox-Snell

- 50% de censura

Tabela 8 – Estimativas de  $\alpha, \gamma$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\hat{\gamma}$	EQM	Vício	$\hat{\alpha}$	EQM	Vício
30	1,5486	$308,72 \cdot 10^{-3}$	$-451,38 \cdot 10^{-3}$	3,7603	$1.742,66 \cdot 10^{-3}$	$1.277,02 \cdot 10^{-3}$
50	1,4948	$310,28 \cdot 10^{-3}$	$-505,17 \cdot 10^{-3}$	3,7770	$1.727,79 \cdot 10^{-3}$	$1.277,02 \cdot 10^{-3}$
100	1,4688	$307,67 \cdot 10^{-3}$	$-531,19 \cdot 10^{-3}$	3,7894	$1.711,59 \cdot 10^{-3}$	$1.289,43 \cdot 10^{-3}$
300	1,4448	$316,55 \cdot 10^{-3}$	$-555,19 \cdot 10^{-3}$	3,8036	$1.713,96 \cdot 10^{-3}$	$1.303,62 \cdot 10^{-3}$
500	1,4401	$318,27 \cdot 10^{-3}$	$-559,88 \cdot 10^{-3}$	3,7980	$1.694,03 \cdot 10^{-3}$	$1.298,01 \cdot 10^{-3}$

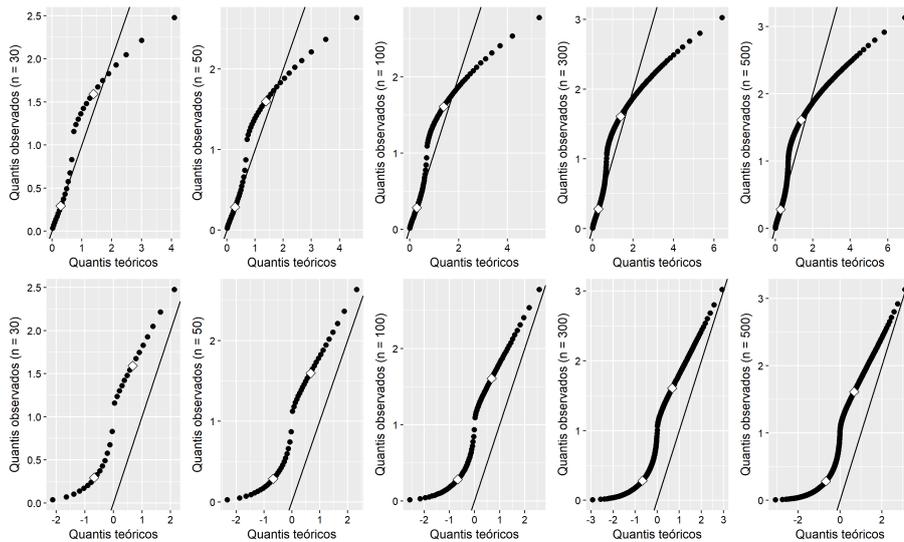


Figura 10 – Gráficos exponenciais e normais de probabilidade para os resíduos de Cox-Snell

Ao observar os gráficos dos resíduos de Cox-Snell, apresentados nas Figuras de 7 a 10, os que estão na linha de cima, pode-se dizer que, nos cenários sem censura e com 10% de dados censurados, é razoável assumir que a distribuição empírica dos resíduos apresenta concordância com uma distribuição Exponencial padrão, uma vez que os conjuntos de pontos não apresentam grandes desvios da reta de referência e nem cruzam a mesma. Nos cenários com 30 e 50% de censura, já são observados alguns desvios, como o cruzamento da reta de referência no primeiro cenário e um afastamento maior da suposição de que os dados se distribuem segundo uma Exponencial padrão no segundo.

A fim de comparação, foram construídos também os gráficos quantil-quantil da Normal padrão e dos resíduos (colocados na linha de baixo das figuras) e, ao observá-los, percebe-se que há indícios de afastamentos sérios da suposição de distribuição Normal padrão dos resíduos.

### 4.1.3 Distribuição Log-normal

Para gerar os tempos de vida da distribuição Log-normal, os parâmetros considerados foram  $\mu = 0,5$  e  $\sigma = 2$ . Para se obter os resíduos, foi utilizada a seguinte equação:

$$\hat{e}_i = -\log(\hat{S}(t)) = -\log \left[ \Phi \left( \frac{-\log(t) + \hat{\mu}}{\hat{\sigma}} \right) \right].$$

- Sem censura

Tabela 9 – Estimativas de  $\mu, \sigma$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\widehat{\mu}$	EQM	Vício	$\widehat{\sigma}$	EQM	Vício
30	0,4954	134,96.10 <sup>-3</sup>	-4,65.10 <sup>-3</sup>	1,9512	66,63.10 <sup>-3</sup>	-48,79.10 <sup>-3</sup>
50	0,4943	80,38.10 <sup>-3</sup>	-5,68.10 <sup>-3</sup>	1,9710	38,36.10 <sup>-3</sup>	-28,99.10 <sup>-3</sup>
100	0,4990	37,89.10 <sup>-3</sup>	-0,97.10 <sup>-3</sup>	1,9878	20,39.10 <sup>-3</sup>	-12,21.10 <sup>-3</sup>
300	0,5002	12,59.10 <sup>-3</sup>	0,16.10 <sup>-3</sup>	1,9972	6,88.10 <sup>-3</sup>	-2,83.10 <sup>-3</sup>
500	0,5014	7,84.10 <sup>-3</sup>	1,44.10 <sup>-3</sup>	1,9887	4,24.10 <sup>-3</sup>	-1,33.10 <sup>-3</sup>

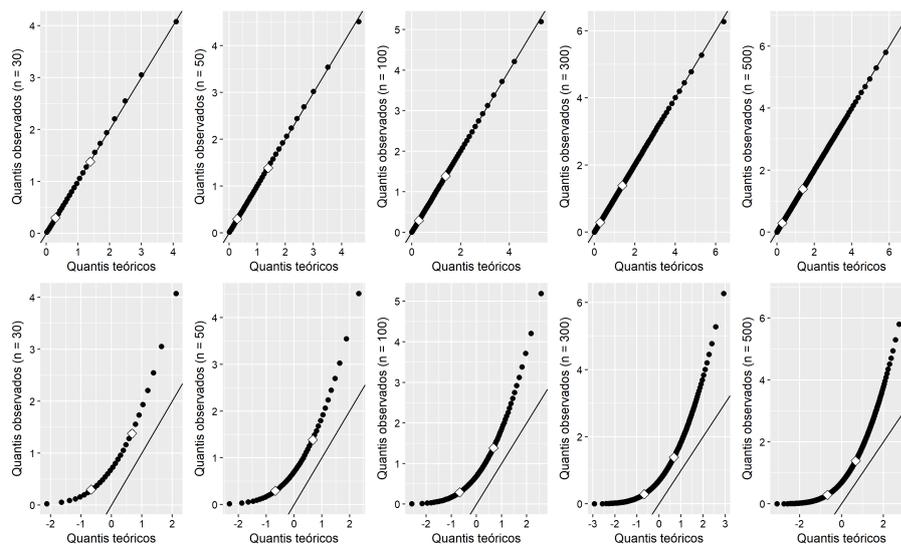


Figura 11 – Gráficos exponenciais e normais de probabilidade para os resíduos de Cox-Snell

- 10% de censura

Tabela 10 – Estimativas de  $\mu, \sigma$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\widehat{\mu}$	EQM	Vício	$\widehat{\sigma}$	EQM	Vício
30	0,6305	144,69.10 <sup>-3</sup>	130,15.10 <sup>-3</sup>	2,1189	96,45.10 <sup>-3</sup>	118,94.10 <sup>-3</sup>
50	0,6022	92,11.10 <sup>-3</sup>	102,18.10 <sup>-3</sup>	2,1180	59,98.10 <sup>-3</sup>	118,04.10 <sup>-3</sup>
100	0,6003	49,39.10 <sup>-3</sup>	100,33.10 <sup>-3</sup>	2,1482	45,55.10 <sup>-3</sup>	148,23.10 <sup>-3</sup>
300	0,6028	23,98.10 <sup>-3</sup>	102,77.10 <sup>-3</sup>	2,1609	34,89.10 <sup>-3</sup>	160,89.10 <sup>-3</sup>
500	0,6140	20,59.10 <sup>-3</sup>	113,95.10 <sup>-3</sup>	2,1721	34,54.10 <sup>-3</sup>	172,09.10 <sup>-3</sup>

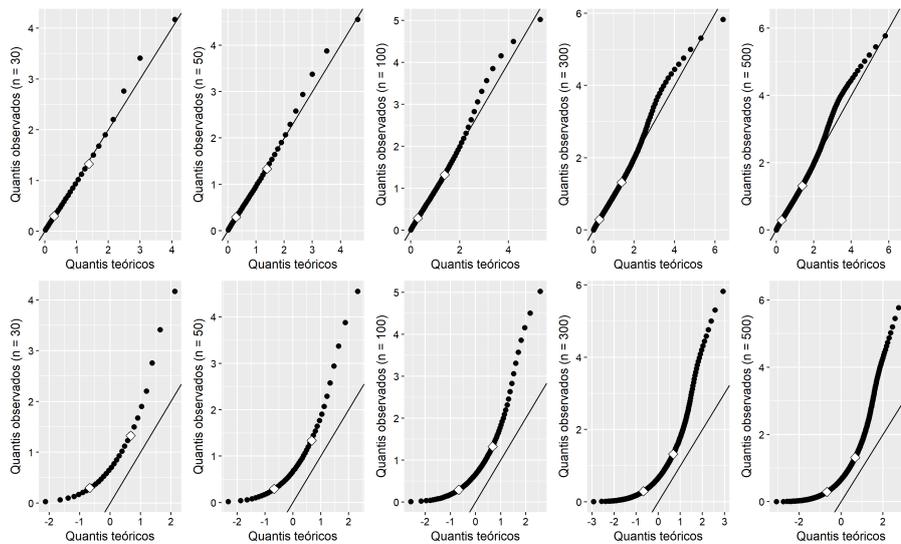


Figura 12 – Gráficos exponenciais e normais de probabilidade para os resíduos de Cox-Snell

- 30% de censura

Tabela 11 – Estimativas de  $\mu, \sigma$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\widehat{\mu}$	EQM	Vício	$\widehat{\sigma}$	EQM	Vício
30	1,0117	402,52.10 <sup>-3</sup>	511,67.10 <sup>-3</sup>	2,5027	371,76.10 <sup>-3</sup>	502,65.10 <sup>-3</sup>
50	1,0063	335,23.10 <sup>-3</sup>	506,31.10 <sup>-3</sup>	2,5415	363,63.10 <sup>-3</sup>	541,51.10 <sup>-3</sup>
100	0,9974	289,19.10 <sup>-3</sup>	497,38.10 <sup>-3</sup>	2,5894	382,66.10 <sup>-3</sup>	589,43.10 <sup>-3</sup>
300	1,0047	268,08.10 <sup>-3</sup>	504,65.10 <sup>-3</sup>	2,6241	401,99.10 <sup>-3</sup>	624,11.10 <sup>-3</sup>
500	1,0048	263,82.10 <sup>-3</sup>	504,77.10 <sup>-3</sup>	2,6342	409,51.10 <sup>-3</sup>	634,17.10 <sup>-3</sup>

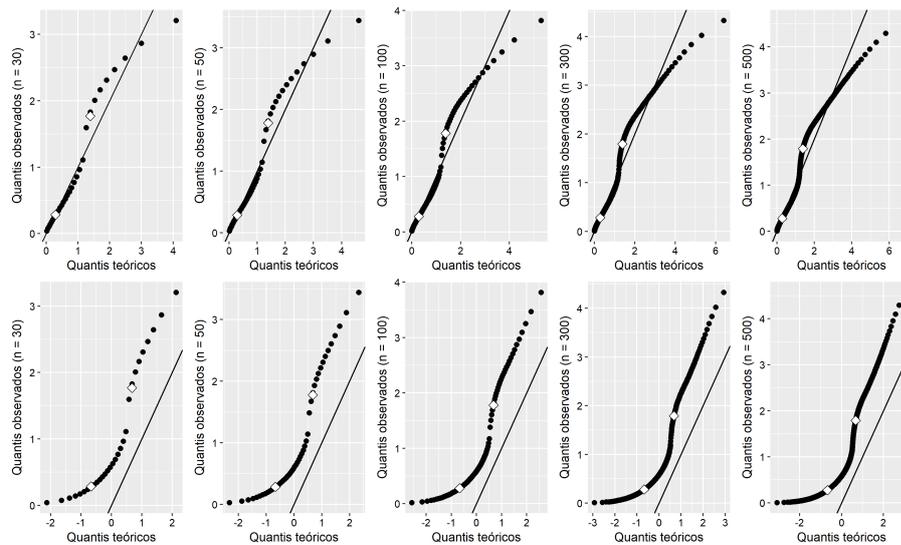


Figura 13 – Gráficos exponenciais e normais de probabilidade para os resíduos de Cox-Snell

- 50% de censura

Tabela 12 – Estimativas de  $\mu, \sigma$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\bar{\mu}$	EQM	Vício	$\bar{\sigma}$	EQM	Vício
30	1,7628	$1.760,92 \cdot 10^{-3}$	$1.262,83 \cdot 10^{-3}$	3,1522	$1.534,19 \cdot 10^{-3}$	$1.152,20 \cdot 10^{-3}$
50	1,7619	$1.685,90 \cdot 10^{-3}$	$1.281,86 \cdot 10^{-3}$	3,1967	$1.556,22 \cdot 10^{-3}$	$1.196,70 \cdot 10^{-3}$
100	1,8182	$1.790,99 \cdot 10^{-3}$	$1.318,23 \cdot 10^{-3}$	3,2738	$1.686,88 \cdot 10^{-3}$	$1,273,82 \cdot 10^{-3}$
300	1,8310	$1.788,10 \cdot 10^{-3}$	$1.331,01 \cdot 10^{-3}$	3,3160	$1.753,21 \cdot 10^{-3}$	$1.316,03 \cdot 10^{-3}$
500	1,8369	$1.797,32 \cdot 10^{-3}$	$1.336,89 \cdot 10^{-3}$	3,3198	$1.753,95 \cdot 10^{-3}$	$1.319,84 \cdot 10^{-3}$

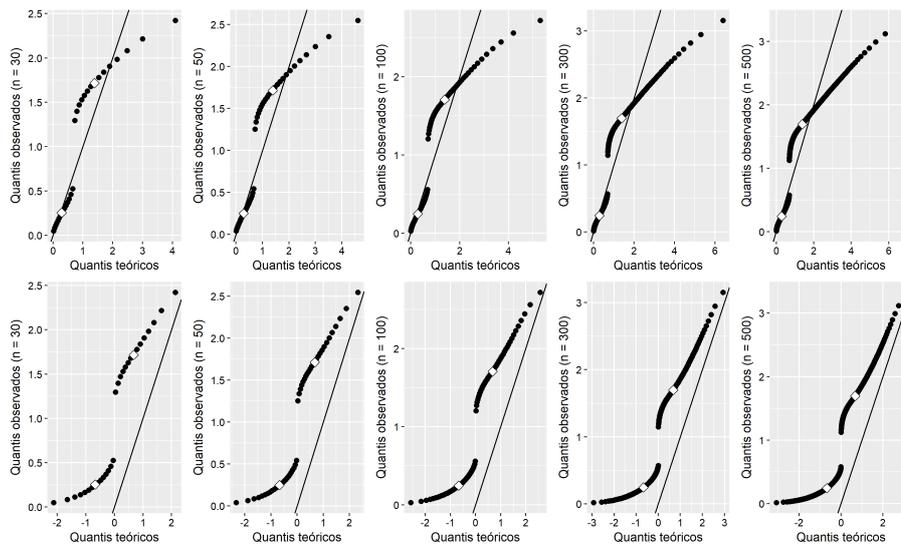


Figura 14 – Gráficos exponenciais e normais de probabilidade para os resíduos de Cox-Snell

Ao observar os gráficos dos resíduos de Cox-Snell, apresentados nas Figuras 11 a 14, observando os que estão na linha de cima, pode-se dizer que, no cenário sem censura, é razoável assumir que a distribuição empírica dos resíduos apresenta concordância com uma distribuição Exponencial padrão, uma vez que os conjuntos de pontos não apresentam grandes desvios da reta de referência e nem cruzam a mesma. No cenário com 10% de censura, alguns desvios já são observados nos últimos quantis, mas ainda é possível assumir que a distribuição dos dados apresenta concordância com a Exponencial padrão. Nos cenários com 30 e 50% de censura, já é observado um afastamento maior da suposição de que os dados se distribuem segundo uma Exponencial padrão.

A fim de comparação, foram construídos também os gráficos quantil-quantil da Normal padrão e dos resíduos (colocados na linha de baixo das figuras) e, ao observá-los,

percebe-se que há indícios de afastamentos sérios da suposição de distribuição Normal padrão dos resíduos.

#### 4.1.4 Distribuição Log-logística

Para gerar os tempos de vida da distribuição Log-logística, os parâmetros considerados foram  $\alpha = 1,5$  e  $\gamma = 0,98$ . Os resíduos, por sua vez, foram calculados a partir da seguinte equação:

$$\hat{e}_i = -\log(\hat{S}(t)) = -\log \left[ \frac{1}{1 + \left(\frac{t}{\hat{\alpha}}\right)^{\hat{\gamma}}} \right].$$

- Sem censura

Tabela 13 – Estimativas de  $\alpha, \gamma$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\hat{\alpha}$	EQM	Vício	$\hat{\gamma}$	EQM	Vício
30	1,4867	$100,92 \cdot 10^{-3}$	$-13,29 \cdot 10^{-3}$	0,9582	$22,78 \cdot 10^{-3}$	$-21,83 \cdot 10^{-3}$
50	1,4882	$62,59 \cdot 10^{-3}$	$-11,79 \cdot 10^{-3}$	0,9674	$13,96 \cdot 10^{-3}$	$-12,61 \cdot 10^{-3}$
100	1,4974	$29,86 \cdot 10^{-3}$	$-2,60 \cdot 10^{-3}$	0,9724	$6,57 \cdot 10^{-3}$	$-7,60 \cdot 10^{-3}$
300	1,5005	$9,38 \cdot 10^{-3}$	$0,50 \cdot 10^{-3}$	0,9771	$2,31 \cdot 10^{-3}$	$-2,91 \cdot 10^{-3}$
500	1,5004	$5,76 \cdot 10^{-3}$	$0,4 \cdot 10^{-3}$	0,9787	$1,40 \cdot 10^{-3}$	$-1,26 \cdot 10^{-3}$

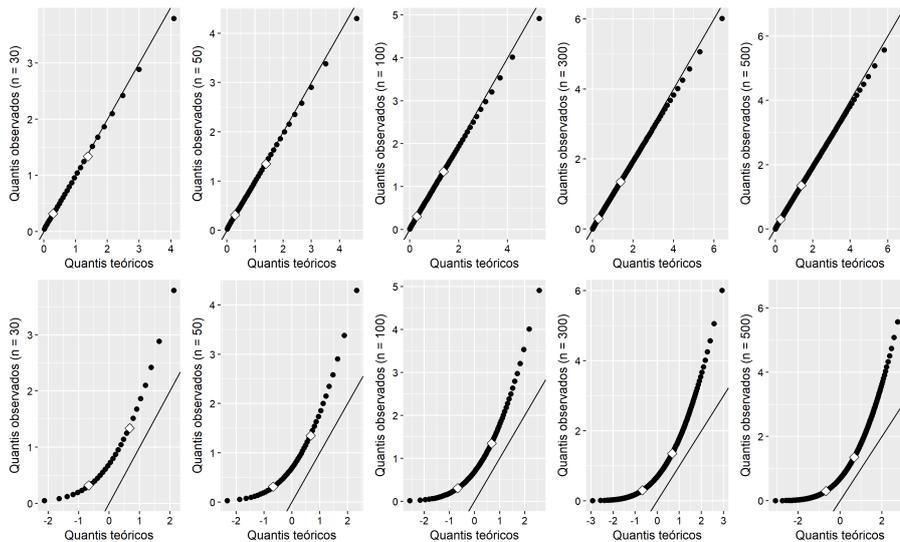


Figura 15 – Gráficos exponenciais e normais de probabilidade para os resíduos de Cox-Snell

- 10% de censura

Tabela 14 – Estimativas de  $\alpha, \gamma$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\widehat{\alpha}$	EQM	Vício	$\widehat{\gamma}$	EQM	Vício
30	1,5569	108,24.10 <sup>-3</sup>	56,86.10 <sup>-3</sup>	1,0344	32,80.10 <sup>-3</sup>	54,40.10 <sup>-3</sup>
50	1,5534	61,78.10 <sup>-3</sup>	53,44.10 <sup>-3</sup>	1,0357	19,99.10 <sup>-3</sup>	55,67.10 <sup>-3</sup>
100	1,5465	35,32.10 <sup>-3</sup>	46,47.10 <sup>-3</sup>	1,0449	13,11.10 <sup>-3</sup>	64,85.10 <sup>-3</sup>
300	1,5494	15,10.10 <sup>-3</sup>	49,41.10 <sup>-3</sup>	1,0533	8,93.10 <sup>-3</sup>	73,25.10 <sup>-3</sup>
500	1,5459	10,51.10 <sup>-3</sup>	45,90.10 <sup>-3</sup>	1,0583	8,82.10 <sup>-3</sup>	78,27.10 <sup>-3</sup>

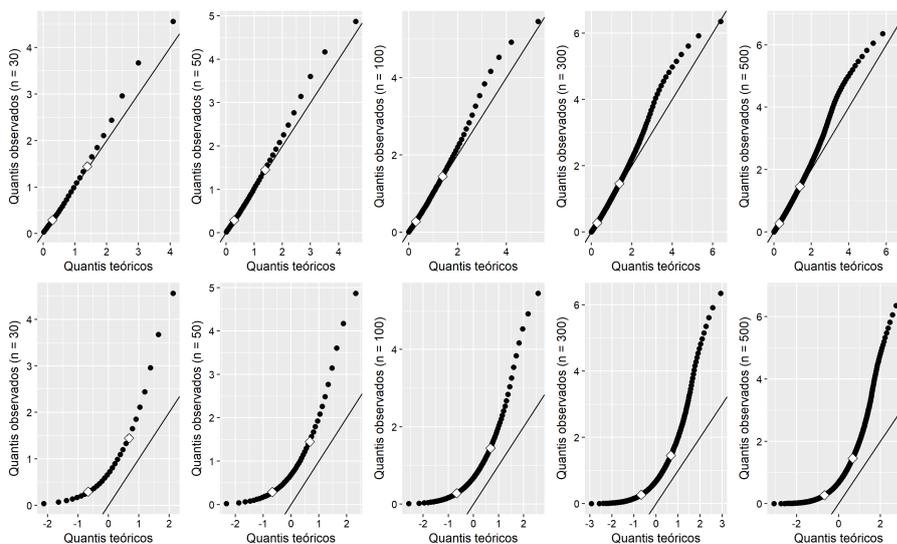


Figura 16 – Gráficos exponenciais e normais de probabilidade para os resíduos de Cox-Snell

- 30% de censura

Tabela 15 – Estimativas de  $\alpha, \gamma$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\widehat{\alpha}$	EQM	Vício	$\widehat{\gamma}$	EQM	Vício
30	1,7905	191,84.10 <sup>-3</sup>	290,55.10 <sup>-3</sup>	1,2349	108,44.10 <sup>-3</sup>	254,94.10 <sup>-3</sup>
50	1,7876	143,80.10 <sup>-3</sup>	287,55.10 <sup>-3</sup>	1,2622	107,74.10 <sup>-3</sup>	282,18.10 <sup>-3</sup>
100	1,8016	122,17.10 <sup>-3</sup>	301,56.10 <sup>-3</sup>	1,2679	96,74.10 <sup>-3</sup>	287,89.10 <sup>-3</sup>
300	1,7976	101,24.10 <sup>-3</sup>	297,63.10 <sup>-3</sup>	1,2787	94,91.10 <sup>-3</sup>	298,69.10 <sup>-3</sup>
500	1,8037	101,42.10 <sup>-3</sup>	303,65.10 <sup>-3</sup>	1,3037	109,13.10 <sup>-3</sup>	323,65.10 <sup>-3</sup>

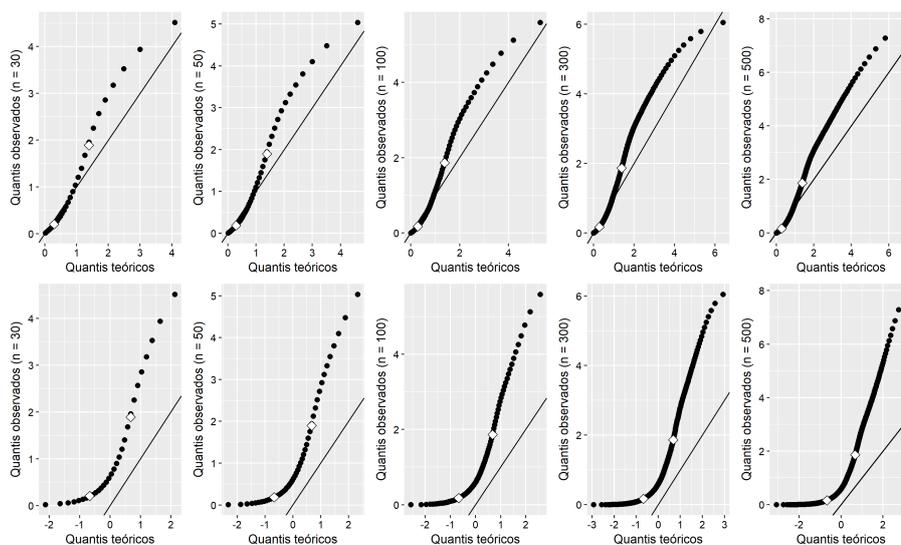


Figura 17 – Gráficos exponenciais e normais de probabilidade para os resíduos de Cox-Snell

- 50% de censura

Tabela 16 – Estimativas de  $\alpha, \gamma$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\widehat{\alpha}$	EQM	Vício	$\widehat{\gamma}$	EQM	Vício
30	2,4080	945,68.10 <sup>-3</sup>	907,92.10 <sup>-3</sup>	1,5818	429,31.10 <sup>-3</sup>	601,78.10 <sup>-3</sup>
50	2,4056	890,90.10 <sup>-3</sup>	905,57.10 <sup>-3</sup>	1,5968	421,96.10 <sup>-3</sup>	616,79.10 <sup>-3</sup>
100	2,4267	897,37.10 <sup>-3</sup>	926,73.10 <sup>-3</sup>	1,5939	398,45.10 <sup>-3</sup>	613,90.10 <sup>-3</sup>
300	2,4932	1.005,94.10 <sup>-3</sup>	993,22.10 <sup>-3</sup>	1,7017	531,73.10 <sup>-3</sup>	721,67.10 <sup>-3</sup>
500	2,4945	1.002,13.10 <sup>-3</sup>	994,54.10 <sup>-3</sup>	1,7087	530,08.10 <sup>-3</sup>	728,68.10 <sup>-3</sup>

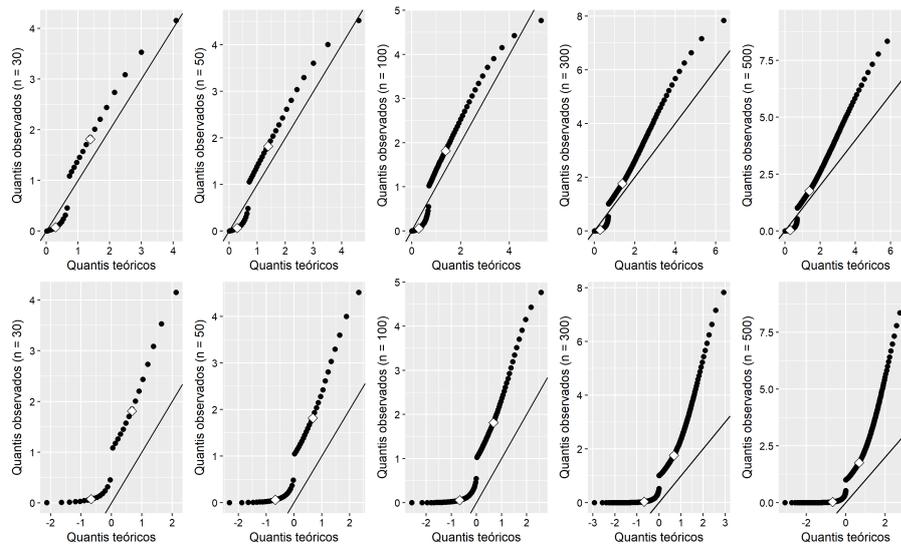


Figura 18 – Gráficos exponenciais e normais de probabilidade para os resíduos de Cox-Snell

Ao observar os gráficos dos resíduos de Cox-Snell, apresentados nas Figuras 15 a 18, os que estão na linha de cima, pode-se dizer que, no cenário sem censura, é razoável assumir que a distribuição empírica dos resíduos apresenta concordância com uma distribuição Exponencial padrão, uma vez que os conjuntos de pontos não apresentam grandes desvios da reta de referência e nem cruzam a mesma. No cenário com 10% de censura, alguns desvios já são observados nos últimos quantis, mas ainda é possível assumir que a distribuição dos dados apresenta concordância com a Exponencial padrão. Nos cenários com 30 e 50% de censura, já é observado um afastamento maior da suposição de que os dados se distribuem segundo uma Exponencial padrão.

A fim de comparação, foram construídos também os gráficos quantil-quantil da Normal padrão e dos resíduos (colocados na linha de baixo das figuras) e, ao observá-los,

percebe-se que há indícios de afastamentos sérios da suposição de distribuição Normal padrão dos resíduos.

## 4.2 Com covariável

### 4.2.1 Distribuição Exponencial

Para gerar os tempos de vida da distribuição Exponencial, os parâmetros considerados foram  $\beta_0 = 8,47$  e  $\beta_1 = -1,11$ . No cálculo dos resíduos, a seguinte equação foi utilizada:

$$\hat{e}_i = -\log(\hat{S}(y)) = \exp(y - \mathbf{x}^T \hat{\boldsymbol{\beta}}).$$

- Sem censura

Tabela 17 – Estimativas de  $\beta_0, \beta_1$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\hat{\beta}_0$	EQM	Vício	$\hat{\beta}_1$	EQM	Vício
30	8,441	$132,90 \cdot 10^{-3}$	$-29,24 \cdot 10^{-3}$	-1,132	$419,06 \cdot 10^{-3}$	$-22,30 \cdot 10^{-3}$
50	8,439	$82,88 \cdot 10^{-3}$	$-31,19 \cdot 10^{-3}$	-1,094	$254,92 \cdot 10^{-3}$	$16,88 \cdot 10^{-3}$
100	8,469	$42,65 \cdot 10^{-3}$	$-31,19 \cdot 10^{-3}$	-1,120	$125,32 \cdot 10^{-3}$	$-9,71 \cdot 10^{-3}$
300	8,462	$14,04 \cdot 10^{-3}$	$-8,08 \cdot 10^{-3}$	-1,105	$42,65 \cdot 10^{-3}$	$5,05 \cdot 10^{-3}$
500	8,464	$7,38 \cdot 10^{-3}$	$5,87 \cdot 10^{-3}$	-1,107	$22,63 \cdot 10^{-3}$	$3,49 \cdot 10^{-3}$

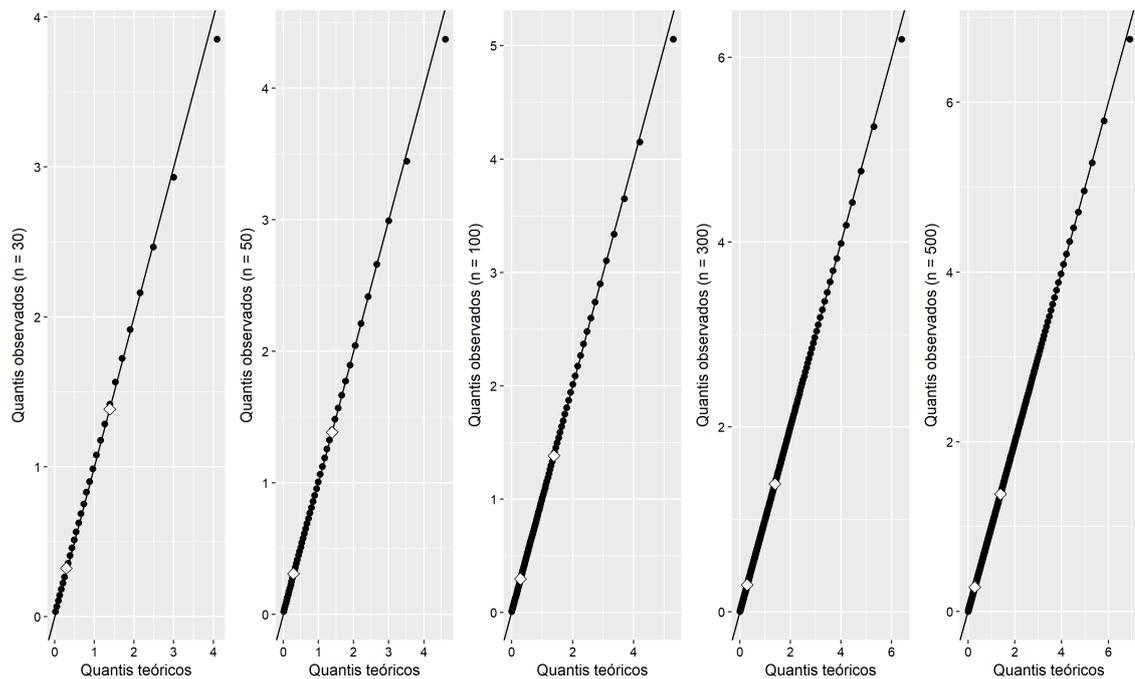


Figura 19 – Gráficos exponenciais de probabilidade para os resíduos de Cox-Snell

- 10% de censura

Tabela 18 – Estimativas de  $\beta_0, \beta_1$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\widehat{\beta}_0$	EQM	Vício	$\widehat{\beta}_1$	EQM	Vício
30	8,630	$204,99 \cdot 10^{-3}$	$159,50 \cdot 10^{-3}$	-1,256	$640,77 \cdot 10^{-3}$	$-164,62 \cdot 10^{-3}$
50	8,636	$130,65 \cdot 10^{-3}$	$165,86 \cdot 10^{-3}$	-1,246	$342,29 \cdot 10^{-3}$	$-135,52 \cdot 10^{-3}$
100	8,611	$65,77 \cdot 10^{-3}$	$140,81 \cdot 10^{-3}$	-1,201	$159,33 \cdot 10^{-3}$	$-90,73 \cdot 10^{-3}$
300	8,629	$41,31 \cdot 10^{-3}$	$158,47 \cdot 10^{-3}$	-1,223	$64,60 \cdot 10^{-3}$	$-112,75 \cdot 10^{-3}$
500	8,635	$37,46 \cdot 10^{-3}$	$165,11 \cdot 10^{-3}$	-1,231	$47,06 \cdot 10^{-3}$	$-121,31 \cdot 10^{-3}$

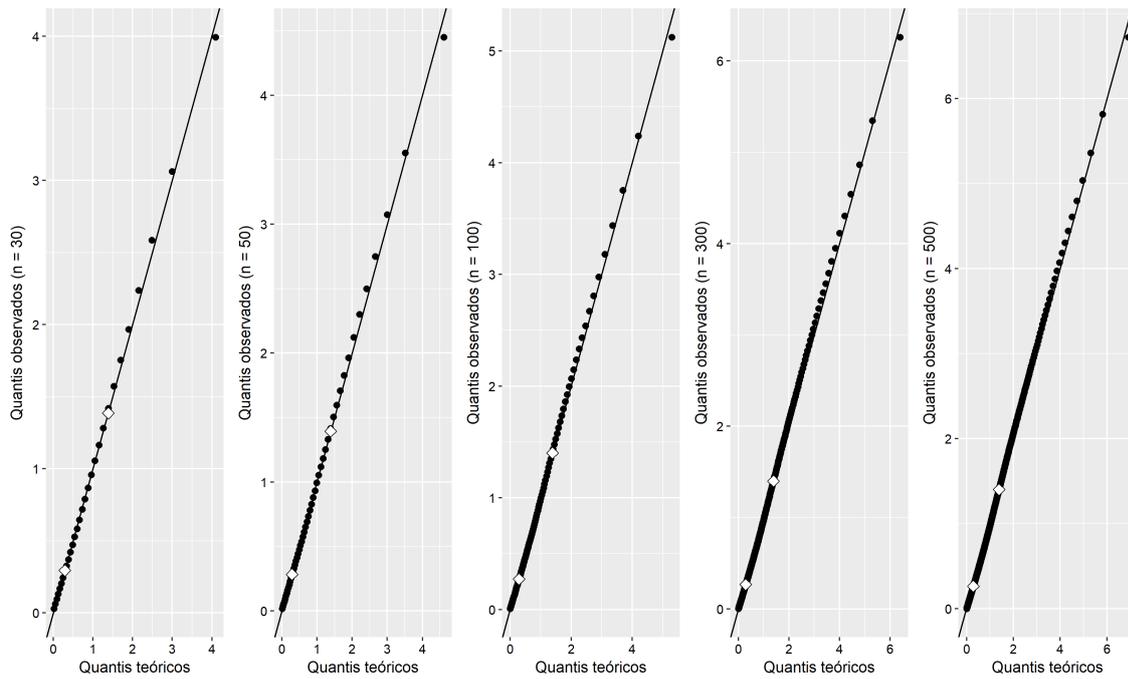


Figura 20 – Gráficos exponenciais de probabilidade para os resíduos de Cox-Snell

- 30% de censura

Tabela 19 – Estimativas de  $\beta_0, \beta_1$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\widehat{\beta}_0$	EQM	Vício	$\widehat{\beta}_1$	EQM	Vício
30	8,965	$538,81 \cdot 10^{-3}$	$494,43 \cdot 10^{-3}$	-1,441	$1.069,54 \cdot 10^{-3}$	$-331,26 \cdot 10^{-3}$
50	9,009	$450,11 \cdot 10^{-3}$	$538,99 \cdot 10^{-3}$	-1,502	$659,80 \cdot 10^{-3}$	$-391,50 \cdot 10^{-3}$
100	9,014	$375,54 \cdot 10^{-3}$	$544,28 \cdot 10^{-3}$	-1,487	$396,58 \cdot 10^{-3}$	$-374,63 \cdot 10^{-3}$
300	9,018	$324,41 \cdot 10^{-3}$	$548,11 \cdot 10^{-3}$	-1,489	$222,53 \cdot 10^{-3}$	$-379,25 \cdot 10^{-3}$
500	9,025	$324,61 \cdot 10^{-3}$	$555,12 \cdot 10^{-3}$	-1,499	$203,48 \cdot 10^{-3}$	$389,05 \cdot 10^{-3}$

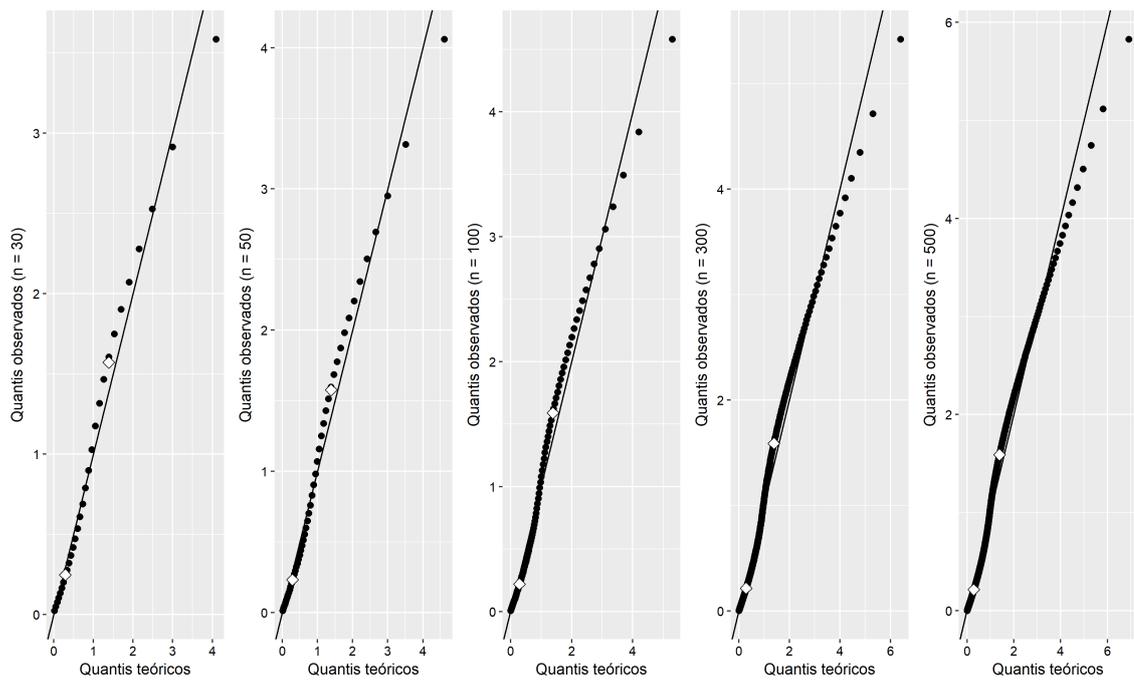


Figura 21 – Gráficos exponenciais de probabilidade para os resíduos de Cox-Snell

- 50% de censura

Tabela 20 – Estimativas de  $\beta_0, \beta_1$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\widehat{\beta}_0$	EQM	Vício	$\widehat{\beta}_1$	EQM	Vício
30	9,502	$1.567, 99.10^{-3}$	$1.032, 06.10^{-3}$	-1,802	$1.980, 32.10^{-3}$	$-692, 25.10^{-3}$
50	9,483	$1.381, 84.10^{-3}$	$1.012, 53.10^{-3}$	-1,754	$1.272, 29.10^{-3}$	$-643, 79.10^{-3}$
100	9,480	$1.225, 50.10^{-3}$	$1.010, 27.10^{-3}$	-1,747	$784, 97.10^{-3}$	$-636, 51.10^{-3}$
300	9,505	$1.204, 15.10^{-3}$	$1.035, 06.10^{-3}$	-1,777	$584, 57.10^{-3}$	$-667, 19.10^{-3}$
500	9,489	$1.153, 51.10^{-3}$	$1.019, 01.10^{-3}$	-1,746	$488, 66.10^{-3}$	$-635, 77.10^{-3}$

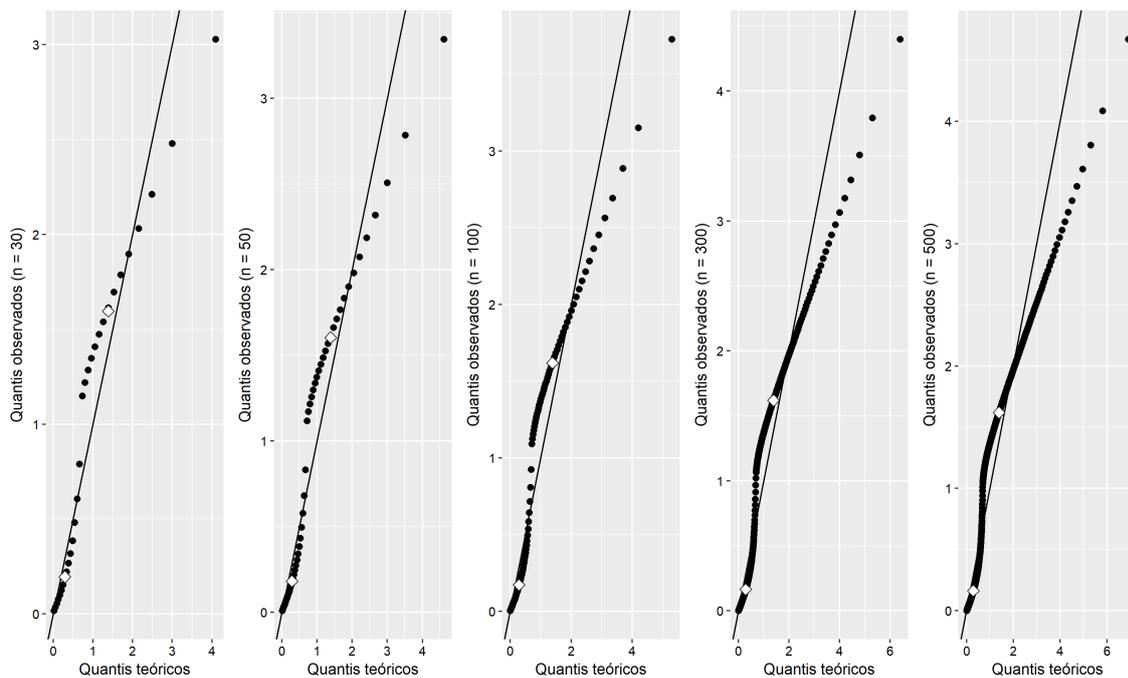


Figura 22 – Gráficos exponenciais de probabilidade para os resíduos de Cox-Snell

Ao observar os gráficos dos resíduos de Cox-Snell, apresentados nas Figuras 19 a 22, pode-se dizer que, nos cenários sem censura e com 10% de dados censurados, é razoável assumir que a distribuição empírica dos resíduos apresenta concordância com uma distribuição Exponencial padrão, uma vez que os conjuntos de pontos não apresentam grandes desvios da reta de referência e nem cruzam a mesma. Nos cenários com 30 e 50% de censura, já são observados alguns desvios, como o cruzamento da reta de referência no primeiro cenário e um afastamento maior da suposição de que os dados se distribuem segundo uma Exponencial padrão no segundo.

### 4.2.2 Distribuição Weibull

Para gerar os tempos de vida da distribuição Exponencial, os parâmetros considerados foram  $\beta_0 = 8,47$ ,  $\beta_1 = -1,11$  e  $\sigma = 0,98$ . Para calcular os resíduos, foi utilizada a seguinte equação:

$$\hat{e}_i = -\log(\hat{S}(y)) = \exp\left(\frac{y - \mathbf{x}^T \hat{\boldsymbol{\beta}}}{\hat{\sigma}}\right).$$

- Sem censura

Tabela 21 – Estimativas de  $\beta_0, \beta_1, \sigma$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\widehat{\beta}_0$	EQM	Vício	$\widehat{\beta}_1$	EQM	Vício	$\widehat{\sigma}$	EQM	Vício
30	8,448	$144,45 \cdot 10^{-3}$	$-22,41 \cdot 10^{-3}$	-1,105	$416,61 \cdot 10^{-3}$	$5,39 \cdot 10^{-3}$	0,934	$20,93 \cdot 10^{-3}$	$-46,26 \cdot 10^{-3}$
50	8,454	$86,24 \cdot 10^{-3}$	$-15,98 \cdot 10^{-3}$	-1,108	$252,56 \cdot 10^{-3}$	$2,48 \cdot 10^{-3}$	0,951	$12,88 \cdot 10^{-3}$	$-29,45 \cdot 10^{-3}$
100	8,470	$38,91 \cdot 10^{-3}$	$0,10 \cdot 10^{-3}$	-1,120	$118,91 \cdot 10^{-3}$	$-9,96 \cdot 10^{-3}$	0,968	$6,03 \cdot 10^{-3}$	$11,96 \cdot 10^{-3}$
300	8,467	$12,96 \cdot 10^{-3}$	$-2,83 \cdot 10^{-3}$	-1,106	$40,26 \cdot 10^{-3}$	$4,24 \cdot 10^{-3}$	0,976	$1,97 \cdot 10^{-3}$	$-3,64 \cdot 10^{-3}$
500	8,467	$8,13 \cdot 10^{-3}$	$-2,59 \cdot 10^{-3}$	-1,111	$23,31 \cdot 10^{-3}$	$-0,46 \cdot 10^{-3}$	0,978	$1,16 \cdot 10^{-3}$	$-1,97 \cdot 10^{-3}$

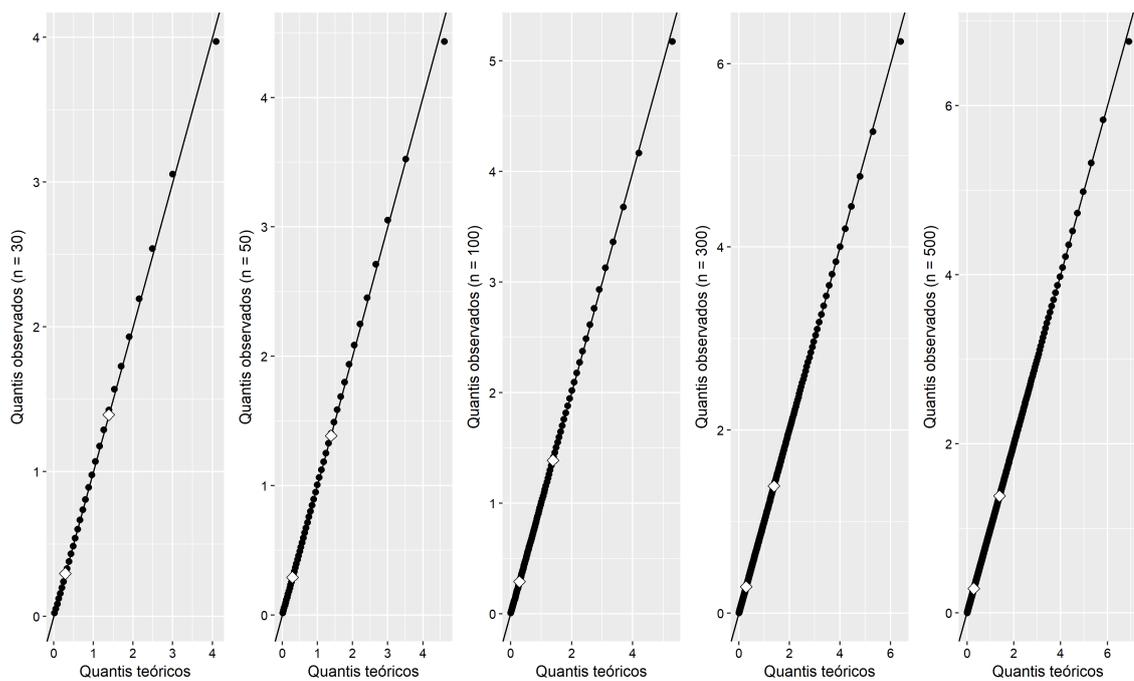


Figura 23 – Gráficos exponenciais de probabilidade para os resíduos de Cox-Snell

- 10% de censura

Tabela 22 – Estimativas de  $\beta_0, \beta_1, \sigma$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\widehat{\beta}_0$	EQM	Vício	$\widehat{\beta}_1$	EQM	Vício	$\widehat{\sigma}$	EQM	Vício
30	8,603	$202,50 \cdot 10^{-3}$	$133,30 \cdot 10^{-3}$	-1,298	$597,18 \cdot 10^{-3}$	$-158,80 \cdot 10^{-3}$	1,015	$23,97 \cdot 10^{-3}$	$34,67 \cdot 10^{-3}$
50	8,609	$131,10 \cdot 10^{-3}$	$-139,34 \cdot 10^{-3}$	-1,240	$371,10 \cdot 10^{-3}$	$-130,20 \cdot 10^{-3}$	1,020	$16,48 \cdot 10^{-3}$	$40,41 \cdot 10^{-3}$
100	8,613	$39,34 \cdot 10^{-3}$	$143,37 \cdot 10^{-3}$	-1,242	$188,65 \cdot 10^{-3}$	$-131,79 \cdot 10^{-3}$	1,029	$9,85 \cdot 10^{-3}$	$49,25 \cdot 10^{-3}$
300	8,611	$35,66 \cdot 10^{-3}$	$141,14 \cdot 10^{-3}$	-1,227	$61,70 \cdot 10^{-3}$	$-117,35 \cdot 10^{-3}$	1,039	$5,98 \cdot 10^{-3}$	$59,12 \cdot 10^{-3}$
500	8,614	$30,52 \cdot 10^{-3}$	$144,32 \cdot 10^{-3}$	-1,234	$45,97 \cdot 10^{-3}$	$-124,35 \cdot 10^{-3}$	1,041	$5,20 \cdot 10^{-3}$	$60,93 \cdot 10^{-3}$

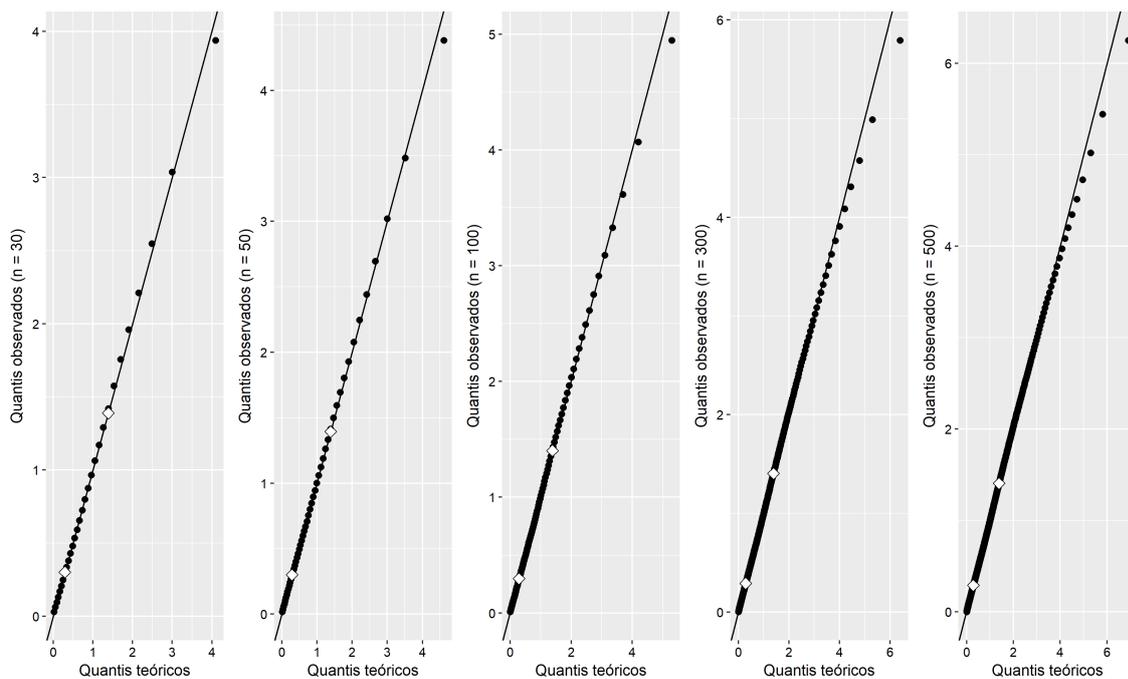


Figura 24 – Gráficos exponenciais de probabilidade para os resíduos de Cox-Snell

- 30% de censura

Tabela 23 – Estimativas de  $\beta_0, \beta_1, \sigma$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\widehat{\beta}_0$	EQM	Vício	$\widehat{\beta}_1$	EQM	Vício	$\widehat{\sigma}$	EQM	Vício
30	9,016	$612,97 \cdot 10^{-3}$	$545,46 \cdot 10^{-3}$	-1,542	$1.234,80 \cdot 10^{-3}$	$-439,24 \cdot 10^{-3}$	1,167	$73,80 \cdot 10^{-3}$	$186,82 \cdot 10^{-3}$
50	9,049	$514,98 \cdot 10^{-3}$	$579,30 \cdot 10^{-3}$	-1,574	$792,05 \cdot 10^{-3}$	$-463,85 \cdot 10^{-3}$	1,191	$67,60 \cdot 10^{-3}$	$210,56 \cdot 10^{-3}$
100	9,056	$361,56 \cdot 10^{-3}$	$585,46 \cdot 10^{-3}$	-1,570	$500,56 \cdot 10^{-3}$	$-459,76 \cdot 10^{-3}$	1,216	$67,93 \cdot 10^{-3}$	$235,83 \cdot 10^{-3}$
300	9,074	$393,66 \cdot 10^{-3}$	$604,35 \cdot 10^{-3}$	-1,606	$338,19 \cdot 10^{-3}$	$-496,34 \cdot 10^{-3}$	1,222	$62,47 \cdot 10^{-3}$	$241,76 \cdot 10^{-3}$
500	9,069	$376,76 \cdot 10^{-3}$	$598,52 \cdot 10^{-3}$	-1,592	$294,22 \cdot 10^{-3}$	$-482,06 \cdot 10^{-3}$	1,224	$61,90 \cdot 10^{-3}$	$243,65 \cdot 10^{-3}$

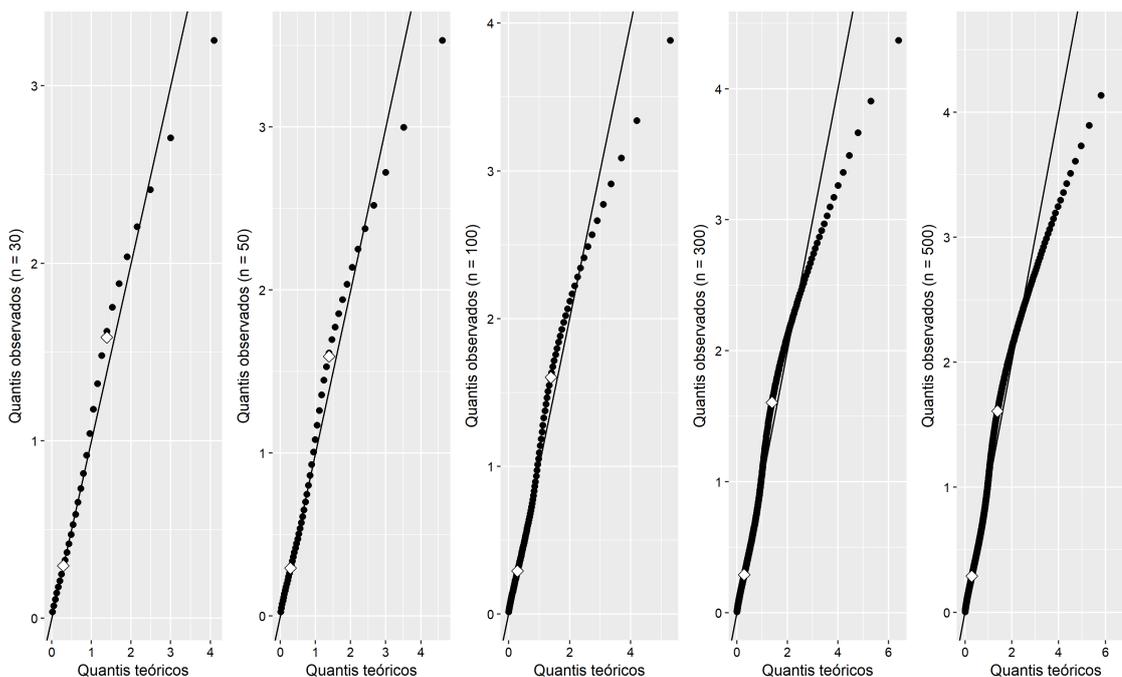


Figura 25 – Gráficos exponenciais de probabilidade para os resíduos de Cox-Snell

- 50% de censura

Tabela 24 – Estimativas de  $\beta_0, \beta_1, \sigma$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\widehat{\beta}_0$	EQM	Vício	$\widehat{\beta}_1$	EQM	Vício	$\widehat{\sigma}$	EQM	Vício
30	9,772	2.361, 73.10 <sup>-3</sup>	1.301, 71.10 <sup>-3</sup>	-1,973	2.932, 82.10 <sup>-3</sup>	-862, 72.10 <sup>-3</sup>	1,418	262, 26.10 <sup>-3</sup>	438, 05.10 <sup>-3</sup>
50	9,813	2.196, 13.10 <sup>-3</sup>	1.342, 54.10 <sup>-3</sup>	-2,012	2.077, 80.10 <sup>-3</sup>	-902, 03.10 <sup>-3</sup>	1,452	267, 07.10 <sup>-3</sup>	472, 43.10 <sup>-3</sup>
100	9,841	1.898, 67.10 <sup>-3</sup>	1.371, 09.10 <sup>-3</sup>	-2,058	1.495, 15.10 <sup>-3</sup>	-948, 17.10 <sup>-3</sup>	1,474	264, 94.10 <sup>-3</sup>	494, 25.10 <sup>-3</sup>
300	9,855	1.975, 31.10 <sup>-3</sup>	1.384, 56.10 <sup>-3</sup>	-2,080	1.132, 93.10 <sup>-3</sup>	-969, 75.10 <sup>-3</sup>	1,482	258, 77.10 <sup>-3</sup>	501, 71.10 <sup>-3</sup>
500	9,865	1.981, 22.10 <sup>-3</sup>	1.394, 47.10 <sup>-3</sup>	-2,092	1.085, 56.10 <sup>-3</sup>	-981, 61.10 <sup>-3</sup>	1,488	262, 59.10 <sup>-3</sup>	508, 11.10 <sup>-3</sup>

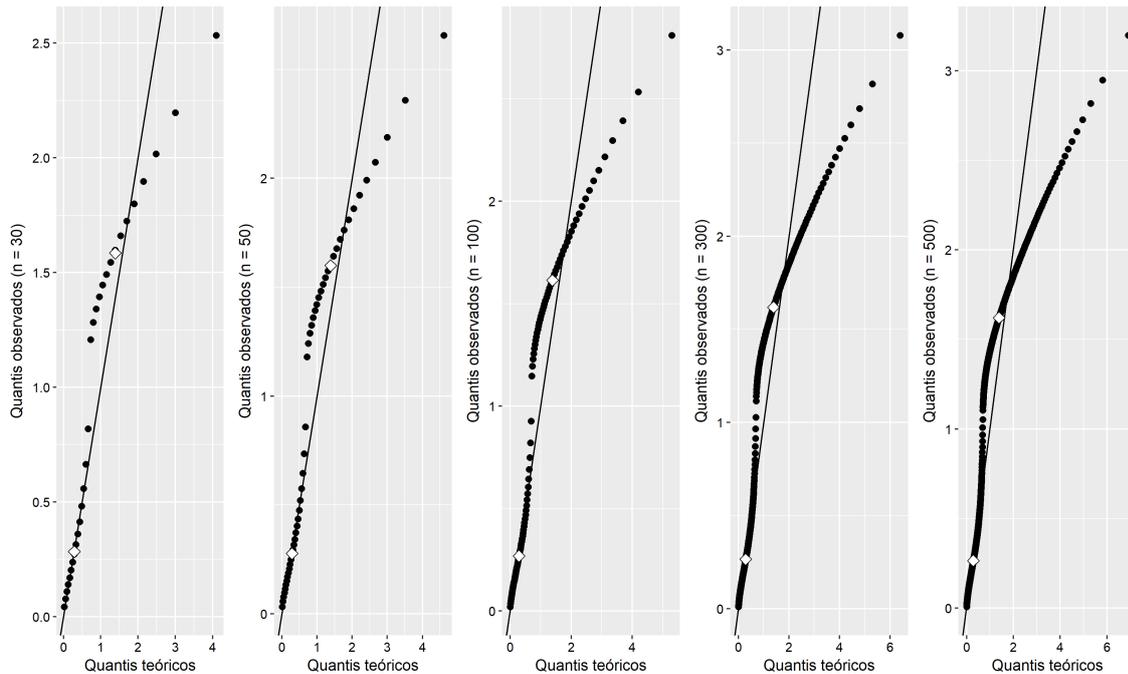


Figura 26 – Gráficos exponenciais de probabilidade para os resíduos de Cox-Snell

Ao observar os gráficos dos resíduos de Cox-Snell, apresentados nas Figuras 23 a 26, pode-se dizer que, nos cenários sem censura e com 10% de dados censurados, é razoável assumir que a distribuição empírica dos resíduos apresenta concordância com uma distribuição Exponencial padrão, uma vez que os conjuntos de pontos não apresentam grandes desvios da reta de referência e nem cruzam a mesma. Nos cenários com 30 e 50% de censura, já são observados alguns desvios, como o cruzamento da reta de referência no primeiro cenário e um afastamento maior da suposição de que os dados se distribuem segundo uma Exponencial padrão no segundo.

### 4.2.3 Distribuição Log-normal

Para gerar os tempos de vida da distribuição Exponencial, os parâmetros considerados foram  $\beta_0 = 8,47$ ,  $\beta_1 = -1,11$  e  $\sigma = 0,98$ . Além disso, a equação do resíduo para este cenário é definida da seguinte maneira:

$$\hat{e}_i = -\log(\hat{S}(y)) = -\log \left[ 1 - \Phi \left( \frac{y - \mathbf{x}^T \hat{\beta}}{\hat{\sigma}} \right) \right].$$

- Sem censura

Tabela 25 – Estimativas de  $\beta_0, \beta_1, \sigma$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\hat{\beta}_0$	EQM	Vício	$\hat{\beta}_1$	EQM	Vício	$\hat{\sigma}$	EQM	Vício
30	8,468	$138,27 \cdot 10^{-3}$	$-2,19 \cdot 10^{-3}$	-1,103	$426,02 \cdot 10^{-3}$	$7,27 \cdot 10^{-3}$	0,942	$16,61 \cdot 10^{-3}$	$-39,31 \cdot 10^{-3}$
50	8,474	$76,43 \cdot 10^{-3}$	$4,39 \cdot 10^{-3}$	-1,117	$244,56 \cdot 10^{-3}$	$-6,69 \cdot 10^{-3}$	0,953	$10,73 \cdot 10^{-3}$	$-27,06 \cdot 10^{-3}$
100	8,462	$36,66 \cdot 10^{-3}$	$8,42 \cdot 10^{-3}$	-1,088	$113,71 \cdot 10^{-3}$	$22,49 \cdot 10^{-3}$	0,969	$4,97 \cdot 10^{-3}$	$11,54 \cdot 10^{-3}$
300	8,466	$12,04 \cdot 10^{-3}$	$-4,14 \cdot 10^{-3}$	-1,099	$37,23 \cdot 10^{-3}$	$10,07 \cdot 10^{-3}$	0,977	$1,62 \cdot 10^{-3}$	$-2,54 \cdot 10^{-3}$
500	8,475	$6,95 \cdot 10^{-3}$	$4,51 \cdot 10^{-3}$	-1,116	$21,42 \cdot 10^{-3}$	$-5,62 \cdot 10^{-3}$	0,978	$0,98 \cdot 10^{-3}$	$-2,52 \cdot 10^{-3}$

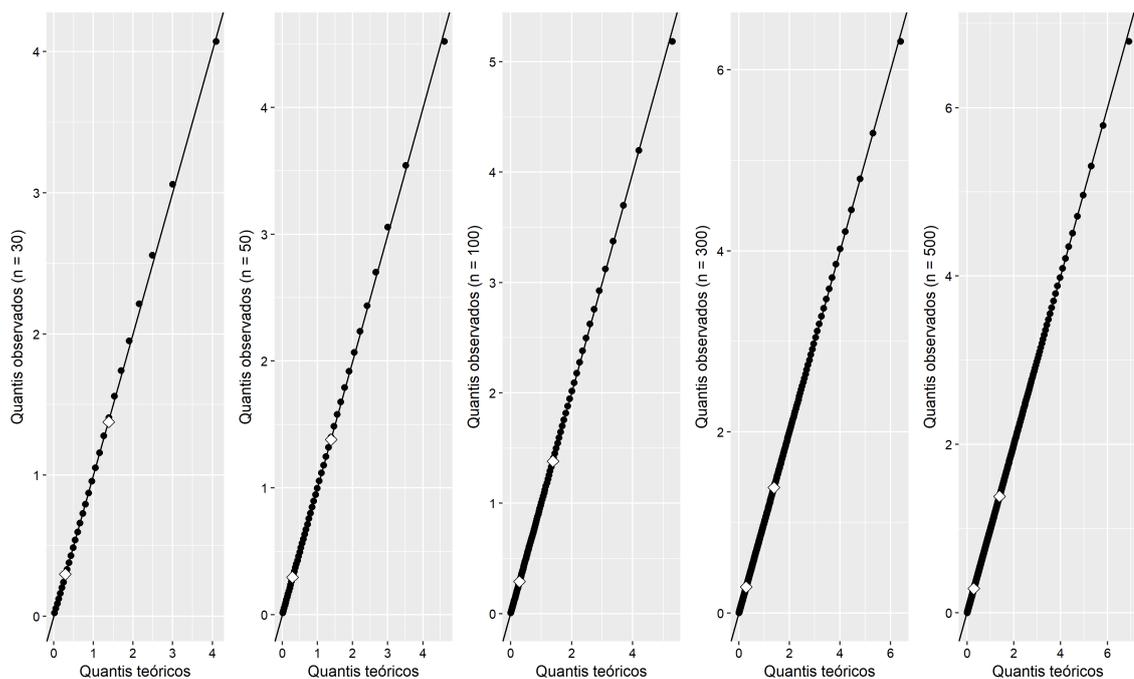


Figura 27 – Gráficos exponenciais de probabilidade para os resíduos de Cox-Snell

- 10% de censura

Tabela 26 – Estimativas de  $\beta_0, \beta_1, \sigma$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\widehat{\beta}_0$	EQM	Vício	$\widehat{\beta}_1$	EQM	Vício	$\widehat{\sigma}$	EQM	Vício
30	8,560	$160,23 \cdot 10^{-3}$	$89,71 \cdot 10^{-3}$	-1,158	$469,02 \cdot 10^{-3}$	$-47,99 \cdot 10^{-3}$	1,002	$20,90 \cdot 10^{-3}$	$22,21 \cdot 10^{-3}$
50	8,550	$96,14 \cdot 10^{-3}$	$80,19 \cdot 10^{-3}$	-1,144	$282,71 \cdot 10^{-3}$	$-34,32 \cdot 10^{-3}$	1,017	$12,53 \cdot 10^{-3}$	$36,61 \cdot 10^{-3}$
100	8,568	$49,65 \cdot 10^{-3}$	$97,77 \cdot 10^{-3}$	-1,180	$128,32 \cdot 10^{-3}$	$-70,12 \cdot 10^{-3}$	1,031	$8,43 \cdot 10^{-3}$	$51,15 \cdot 10^{-3}$
300	8,566	$23,71 \cdot 10^{-3}$	$96,36 \cdot 10^{-3}$	-1,169	$47,19 \cdot 10^{-3}$	$-59,04 \cdot 10^{-3}$	1,045	$6,21 \cdot 10^{-3}$	$65,18 \cdot 10^{-3}$
500	8,565	$17,15 \cdot 10^{-3}$	$94,71 \cdot 10^{-3}$	-1,173	$29,18 \cdot 10^{-3}$	$-63,03 \cdot 10^{-3}$	1,049	$5,99 \cdot 10^{-3}$	$69,27 \cdot 10^{-3}$

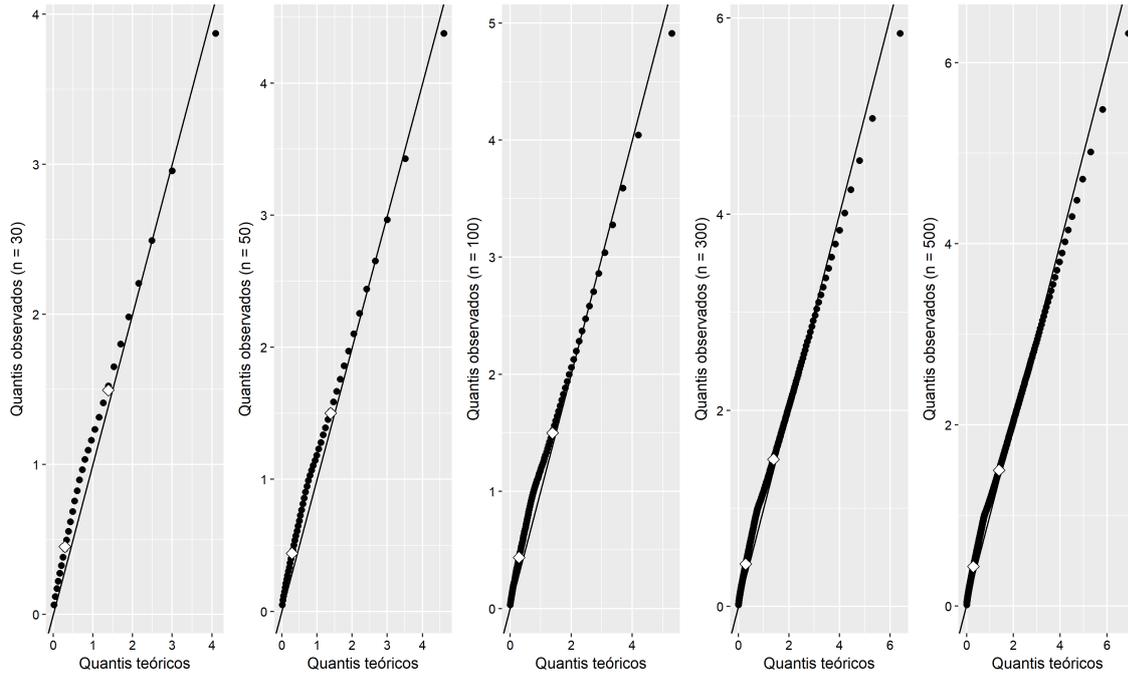


Figura 28 – Gráficos exponenciais de probabilidade para os resíduos de Cox-Snell

- 30% de censura

Tabela 27 – Estimativas de  $\beta_0, \beta_1, \sigma$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\widehat{\beta}_0$	EQM	Vício	$\widehat{\beta}_1$	EQM	Vício	$\widehat{\sigma}$	EQM	Vício
30	8,8450	$383,29 \cdot 10^{-3}$	$374,97 \cdot 10^{-3}$	-1,3276	$80,41 \cdot 10^{-3}$	$-217,57 \cdot 10^{-3}$	1,1733	$70,52 \cdot 10^{-3}$	$193,29 \cdot 10^{-3}$
50	8,850	$266,64 \cdot 10^{-3}$	$379,72 \cdot 10^{-3}$	-1,340	$434,60 \cdot 10^{-3}$	$-230,20 \cdot 10^{-3}$	1,200	$66,19 \cdot 10^{-3}$	$219,89 \cdot 10^{-3}$
100	8,864	$214,44 \cdot 10^{-3}$	$394,28 \cdot 10^{-3}$	-1,369	$258,32 \cdot 10^{-3}$	$-259,36 \cdot 10^{-3}$	1,219	$65,82 \cdot 10^{-3}$	$238,78 \cdot 10^{-3}$
300	8,871	$182,28 \cdot 10^{-3}$	$400,68 \cdot 10^{-3}$	-1,375	$141,03 \cdot 10^{-3}$	$-265,07 \cdot 10^{-3}$	1,241	$71,30 \cdot 10^{-3}$	$261,09 \cdot 10^{-3}$
500	8,878	$178,82 \cdot 10^{-3}$	$408,02 \cdot 10^{-3}$	-1,388	$116,16 \cdot 10^{-3}$	$-277,78 \cdot 10^{-3}$	1,242	$70,37 \cdot 10^{-3}$	$261,80 \cdot 10^{-3}$

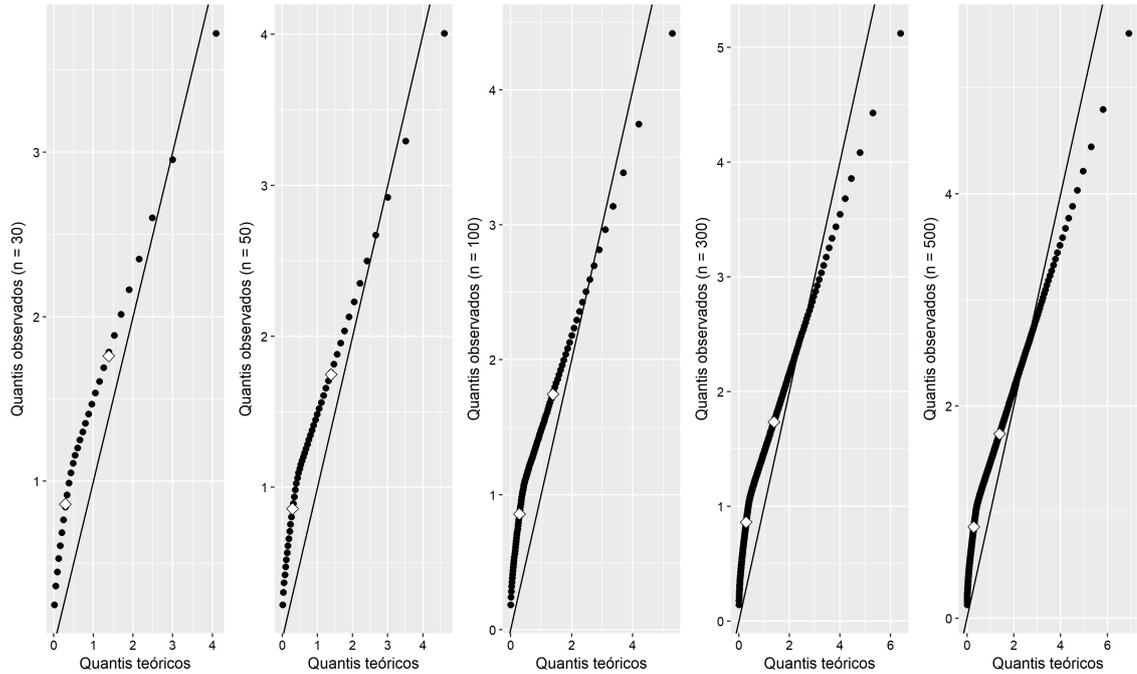


Figura 29 – Gráficos exponenciais de probabilidade para os resíduos de Cox-Snell

- 50% de censura

Tabela 28 – Estimativas de  $\beta_0, \beta_1, \sigma$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\widehat{\beta}_0$	EQM	Vício	$\widehat{\beta}_1$	EQM	Vício	$\widehat{\sigma}$	EQM	Vício
30	9,403	$1.261,84 \cdot 10^{-3}$	$933,07 \cdot 10^{-3}$	-1,657	$1.564,65 \cdot 10^{-3}$	$-547,07 \cdot 10^{-3}$	1,468	$293,31 \cdot 10^{-3}$	$488,21 \cdot 10^{-3}$
50	9,437	$1.149,61 \cdot 10^{-3}$	$966,74 \cdot 10^{-3}$	-1,740	$1.098,53 \cdot 10^{-3}$	$-630,28 \cdot 10^{-3}$	1,477	$278,72 \cdot 10^{-3}$	$497,18 \cdot 10^{-3}$
100	9,419	$1.016,48 \cdot 10^{-3}$	$949,01 \cdot 10^{-3}$	-1,701	$715,87 \cdot 10^{-3}$	$-590,91 \cdot 10^{-3}$	1,512	$297,01 \cdot 10^{-3}$	$531,98 \cdot 10^{-3}$
300	9,425	$947,72 \cdot 10^{-3}$	$954,92 \cdot 10^{-3}$	-1,704	$471,14 \cdot 10^{-3}$	$-594,27 \cdot 10^{-3}$	1,524	$300,73 \cdot 10^{-3}$	$543,85 \cdot 10^{-3}$
500	9,448	$978,74 \cdot 10^{-3}$	$978,31 \cdot 10^{-3}$	-1,731	$456,39 \cdot 10^{-3}$	$-621,24 \cdot 10^{-3}$	1,547	$325,02 \cdot 10^{-3}$	$567,32 \cdot 10^{-3}$

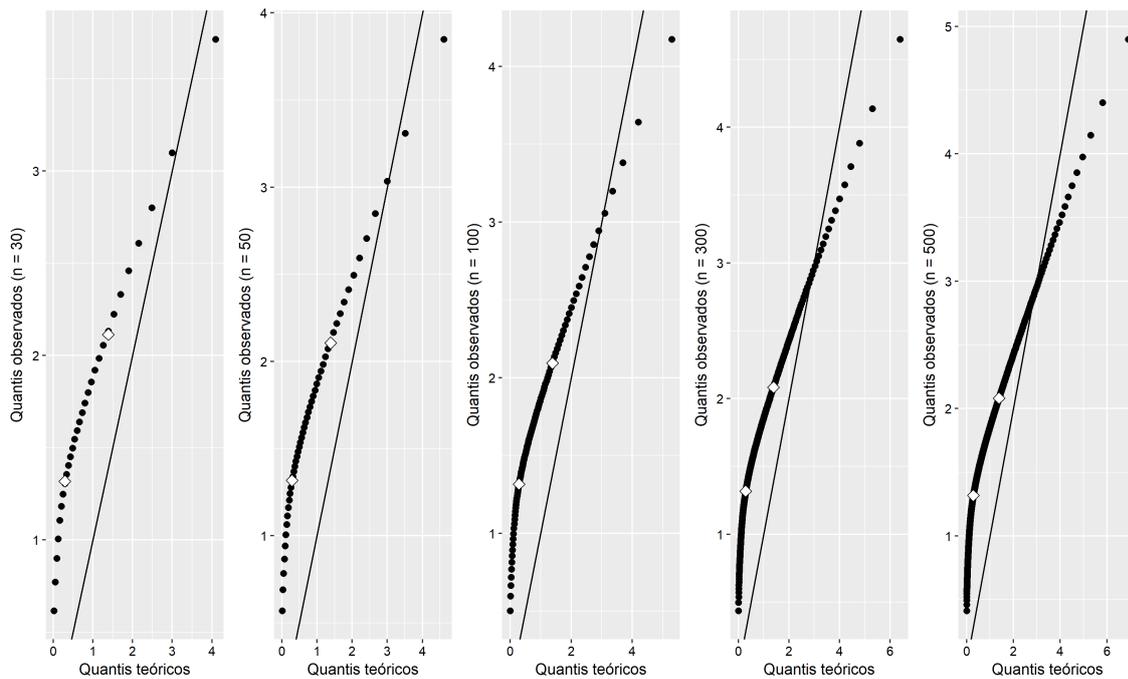


Figura 30 – Gráficos exponenciais de probabilidade para os resíduos de Cox-Snell

Ao observar os gráficos dos resíduos de Cox-Snell, apresentados nas Figuras 27 a 30, pode-se dizer que, no cenário sem censura, é razoável assumir que a distribuição empírica dos resíduos apresenta concordância com uma distribuição Exponencial padrão, uma vez que os conjuntos de pontos não apresentam grandes desvios da reta de referência e nem cruzam a mesma. No cenário com 10% de censura, alguns desvios já são observados nos últimos quantis, mas ainda é possível assumir que a distribuição dos dados apresenta concordância com a Exponencial padrão. Nos cenários com 30 e 50% de censura, já é observado um afastamento maior da suposição de que os dados se distribuem segundo uma Exponencial padrão.

#### 4.2.4 Distribuição Log-logística

Para gerar os tempos de vida da distribuição Exponencial, os parâmetros considerados foram  $\beta_0 = 8,47$ ,  $\beta_1 = -1,11$  e  $\sigma = 0,98$ . Os resíduos, por sua vez, foram calculados com base na seguinte equação:

$$\hat{e}_i = -\log(\hat{S}(y)) = -\log \left[ \frac{1}{1 + \exp \left( \frac{y - \mathbf{x}^T \hat{\beta}}{\hat{\sigma}} \right)} \right].$$

- Sem censura

Tabela 29 – Estimativas de  $\beta_0, \beta_1, \sigma$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\hat{\beta}_0$	EQM	Vício	$\hat{\beta}_1$	EQM	Vício	$\hat{\sigma}$	EQM	Vício
30	8,465	$404,86 \cdot 10^{-3}$	$-5,26 \cdot 10^{-3}$	-1,103	$1.207,03 \cdot 10^{-3}$	$6,91 \cdot 10^{-3}$	0,942	$23,43 \cdot 10^{-3}$	$-37,51 \cdot 10^{-3}$
50	8,468	$248,31 \cdot 10^{-3}$	$-1,85 \cdot 10^{-3}$	-1,096	$747,16 \cdot 10^{-3}$	$13,71 \cdot 10^{-3}$	0,954	$13,53 \cdot 10^{-3}$	$-25,58 \cdot 10^{-3}$
100	8,468	$114,05 \cdot 10^{-3}$	$-1,60 \cdot 10^{-3}$	-1,108	$360,99 \cdot 10^{-3}$	$2,21 \cdot 10^{-3}$	0,971	$6,99 \cdot 10^{-3}$	$-8,98 \cdot 10^{-3}$
300	8,470	$38,95 \cdot 10^{-3}$	$-0,42 \cdot 10^{-3}$	-1,113	$122,21 \cdot 10^{-3}$	$-2,53 \cdot 10^{-3}$	0,978	$2,35 \cdot 10^{-3}$	$-2,13 \cdot 10^{-3}$
500	8,476	$23,49 \cdot 10^{-3}$	$5,60 \cdot 10^{-3}$	-1,115	$69,71 \cdot 10^{-3}$	$-4,80 \cdot 10^{-3}$	0,977	$1,26 \cdot 10^{-3}$	$-2,80 \cdot 10^{-3}$

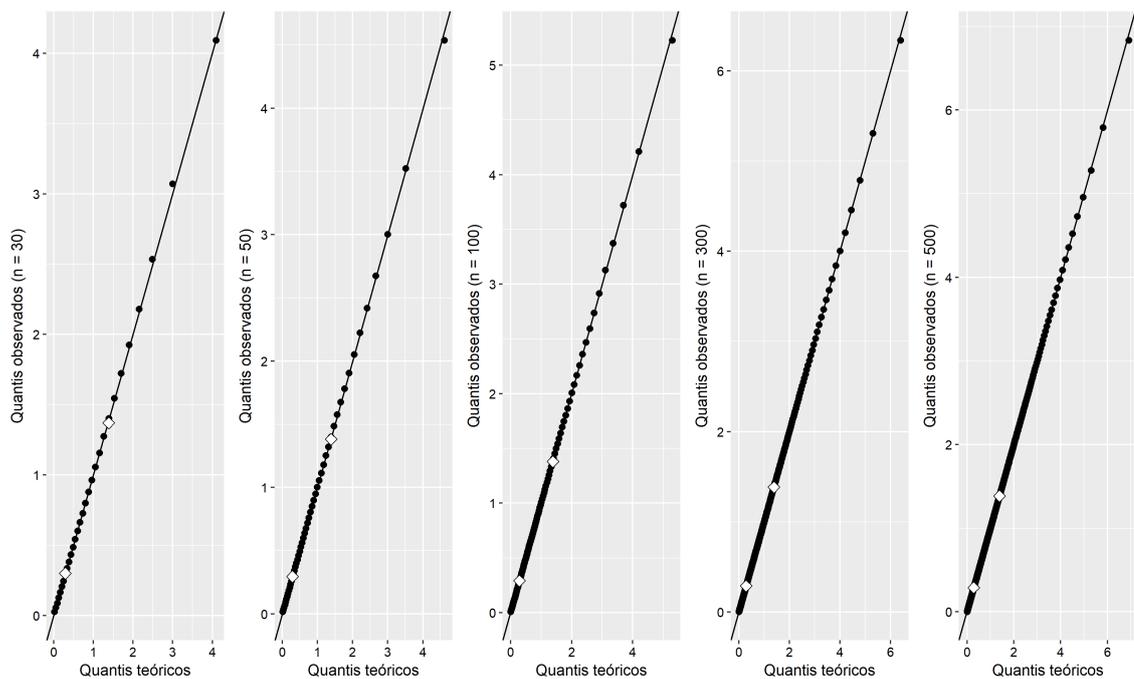


Figura 31 – Gráficos exponenciais de probabilidade para os resíduos de Cox-Snell

- 10% de censura

Tabela 30 – Estimativas de  $\beta_0, \beta_1, \sigma$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\widehat{\beta}_0$	EQM	Vício	$\widehat{\beta}_1$	EQM	Vício	$\widehat{\sigma}$	EQM	Vício
30	8,506	$458,61 \cdot 10^{-3}$	$35,76 \cdot 10^{-3}$	-1,102	$1.461,14 \cdot 10^{-3}$	$7,99 \cdot 10^{-3}$	1,008	$25,70 \cdot 10^{-3}$	$27,70 \cdot 10^{-3}$
50	8,539	$270,41 \cdot 10^{-3}$	$68,53 \cdot 10^{-3}$	-1,142	$826,22 \cdot 10^{-3}$	$-31,71 \cdot 10^{-3}$	1,020	$16,53 \cdot 10^{-3}$	$40,42 \cdot 10^{-3}$
100	8,554	$133,79 \cdot 10^{-3}$	$84,31 \cdot 10^{-3}$	-1,173	$393,01 \cdot 10^{-3}$	$-62,99 \cdot 10^{-3}$	1,039	$10,68 \cdot 10^{-3}$	$58,57 \cdot 10^{-3}$
300	8,539	$48,46 \cdot 10^{-3}$	$69,04 \cdot 10^{-3}$	-1,150	$132,76 \cdot 10^{-3}$	$-39,51 \cdot 10^{-3}$	1,048	$7,30 \cdot 10^{-3}$	$67,97 \cdot 10^{-3}$
500	8,544	$28,80 \cdot 10^{-3}$	$73,75 \cdot 10^{-3}$	-1,163	$80,67 \cdot 10^{-3}$	$-53,28 \cdot 10^{-3}$	1,052	$6,67 \cdot 10^{-3}$	$72,18 \cdot 10^{-3}$

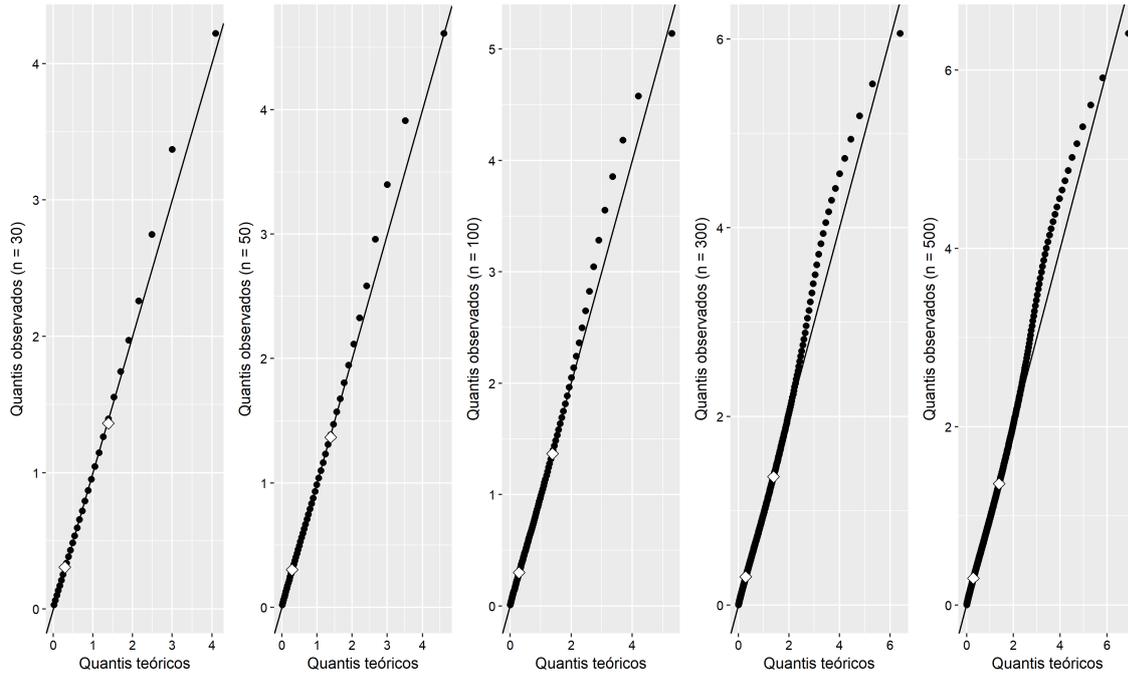


Figura 32 – Gráficos exponenciais de probabilidade para os resíduos de Cox-Snell

- 30% de censura

Tabela 31 – Estimativas de  $\beta_0, \beta_1, \sigma$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\widehat{\beta}_0$	EQM	Vício	$\widehat{\beta}_1$	EQM	Vício	$\widehat{\sigma}$	EQM	Vício
30	8,884	$778,25 \cdot 10^{-3}$	$413,56 \cdot 10^{-3}$	-1,340	$2.053,61 \cdot 10^{-3}$	$-230,19 \cdot 10^{-3}$	1,201	$88,38 \cdot 10^{-3}$	$221,40 \cdot 10^{-3}$
50	8,899	$555,21 \cdot 10^{-3}$	$429,42 \cdot 10^{-3}$	-1,358	$1.264,99 \cdot 10^{-3}$	$-247,54 \cdot 10^{-3}$	1,221	$80,10 \cdot 10^{-3}$	$240,76 \cdot 10^{-3}$
100	8,893	$382,58 \cdot 10^{-3}$	$423,35 \cdot 10^{-3}$	-1,352	$701,72 \cdot 10^{-3}$	$-241,92 \cdot 10^{-3}$	1,236	$76,66 \cdot 10^{-3}$	$255,61 \cdot 10^{-3}$
300	8,944	$291,29 \cdot 10^{-3}$	$473,85 \cdot 10^{-3}$	-1,434	$329,15 \cdot 10^{-3}$	$-324,41 \cdot 10^{-3}$	1,296	$104,36 \cdot 10^{-3}$	$316,28 \cdot 10^{-3}$
500	8,933	$252,10 \cdot 10^{-3}$	$462,72 \cdot 10^{-3}$	-1,410	$208,03 \cdot 10^{-3}$	$-300,01 \cdot 10^{-3}$	1,302	$105,99 \cdot 10^{-3}$	$321,52 \cdot 10^{-3}$

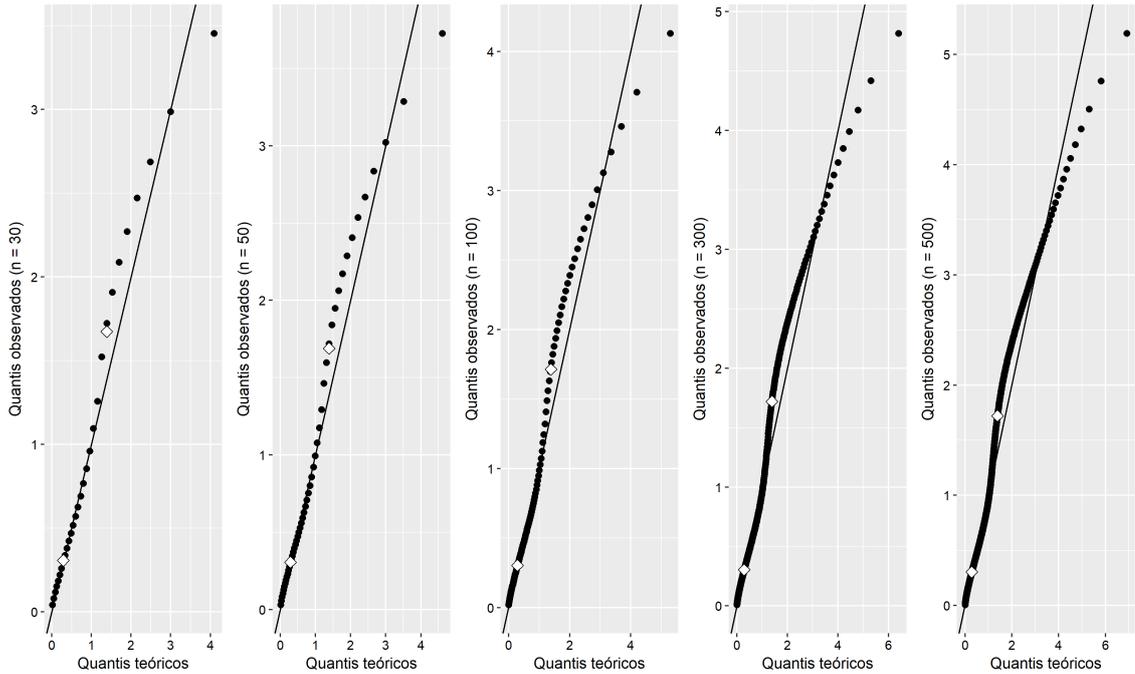


Figura 33 – Gráficos exponenciais de probabilidade para os resíduos de Cox-Snell

- 50% de censura

Tabela 32 – Estimativas de  $\beta_0, \beta_1, \sigma$ , EQM e vício segundo tamanhos de amostras

Tamanho da amostra	$\widehat{\beta}_0$	EQM	Vício	$\widehat{\beta}_1$	EQM	Vício	$\widehat{\sigma}$	EQM	Vício
30	9,809	$3.082, 20.10^{-3}$	$1.338, 58.10^{-3}$	-1,857	$4.870, 04.10^{-3}$	$-746, 52.10^{-3}$	1,564	$408, 86.10^{-3}$	$583, 50.10^{-3}$
50	9,803	$2.574, 69.10^{-3}$	$1.333, 43.10^{-3}$	-1,854	$3.161, 95.10^{-3}$	$-743, 90.10^{-3}$	1,597	$423, 82.10^{-3}$	$616, 41.10^{-3}$
100	9,868	$2.336, 55.10^{-3}$	$1.398, 25.10^{-3}$	-1,947	$1.978, 94.10^{-3}$	$-837, 30.10^{-3}$	1,646	$466, 01.10^{-3}$	$665, 58.10^{-3}$
300	9,911	$2.210, 28.10^{-3}$	$1.441, 06.10^{-3}$	-1,979	$1.195, 37.10^{-3}$	$-868, 51.10^{-3}$	1,674	$488, 58.10^{-3}$	$694, 03.10^{-3}$
500	9,895	$2.110, 81.10^{-3}$	$1.424, 53.10^{-3}$	-1,955	$980, 50.10^{-3}$	$-844, 74.10^{-3}$	1,675	$487, 70.10^{-3}$	$695, 30.10^{-3}$

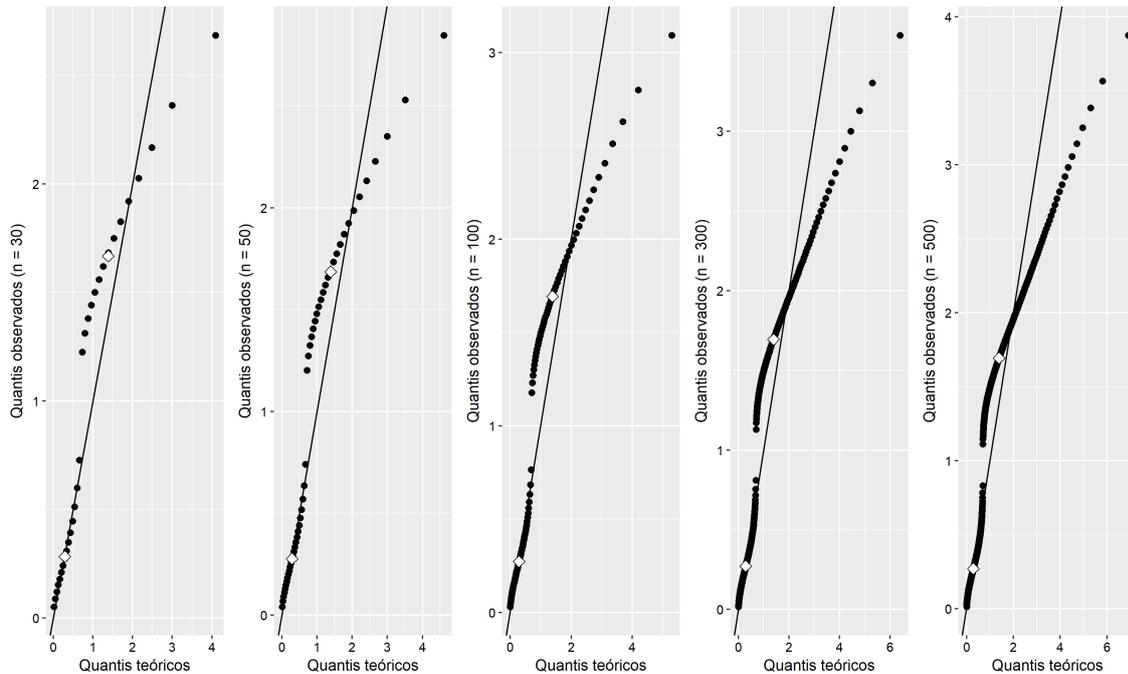


Figura 34 – Gráficos exponenciais de probabilidade para os resíduos de Cox-Snell

Ao observar os gráficos dos resíduos de Cox-Snell, apresentados nas Figuras 31 a 34, pode-se dizer que, no cenário sem censura, é razoável assumir que a distribuição empírica dos resíduos apresenta concordância com uma distribuição Exponencial padrão, uma vez que os conjuntos de pontos não apresentam grandes desvios da reta de referência e nem cruzam a mesma. Nos cenários com 10 e 30% de censura, alguns desvios já são observados nos últimos quantis, mas ainda é possível assumir que a distribuição dos dados apresenta concordância com a Exponencial padrão. Por fim, no cenários com 50% de censura, já é observado um afastamento maior da suposição de que os dados se distribuem segundo uma Exponencial padrão.

## 5 CONSIDERAÇÕES FINAIS

O objetivo inicial do trabalho era estudar o comportamento dos resíduos de Cox-Snell em diferentes cenários. Para tanto, foi realizado um estudo de simulação considerando diferentes distribuições de probabilidade, tamanhos de amostra, porcentagens de censura e presença ou ausência de covariáveis.

Por meio das simulações, foi possível observar que, nos cenários sem censura e com 10% de dados censurados, a distribuição empírica dos resíduos dos modelos apresenta concordância com a distribuição Exponencial padrão. Isto é, o pressuposto de utilização do resíduo de Cox-Snell é atendido. No entanto, ao aumentar a porcentagem de censura são observados alguns desvios da suposição de que os resíduos se distribuem segundo uma Exponencial padrão.

Para fins de comparação, foram construídos gráficos normais de probabilidade para os cenários em que não foi considerada covariável e, a partir deles, pôde-se ver que, em todos os casos, os resíduos apresentaram graves afastamentos da suposição de normalidade.

Foram apresentados também estimadores para os parâmetros das distribuições e calculadas medidas de qualidade dos mesmos. A partir dos resultados obtidos, foi possível concluir que o erro quadrático médio diminui à medida que o tamanho da amostra aumenta e que o aumento do percentual de censura acarreta um crescimento da medida, fato que, como dito anteriormente, é esperado uma vez que, segundo Cardial (2017) a função de verossimilhança na presença de censuras conta com a distribuição da função de sobrevivência.

Além disso, no decorrer do trabalho, pôde-se perceber também que determinadas escolhas de parâmetros geravam dados com um comportamento muito peculiar, fato que prejudica o ajuste dos mesmos a um modelo e interfere diretamente na distribuição dos resíduos.

De maneira geral, conclui-se que o resíduo de Cox-Snell pode ser utilizado para avaliar a qualidade do ajuste de modelos com diferentes distribuições de probabilidade, sendo necessário, porém, certo cuidado em cenários com grande quantidade de dados censurados.



**REFERÊNCIAS**

- COLOSIMO, E. A. e GIOLO, S. R. (2006). **Análise de Sobrevivência Aplicada** São Paulo: Edgard Blucher.
- LAWLESS, J.F. (2002). **Statistical models and methods for lifetime data** 2.ed. Waterloo, Ontario.
- GOMES, E. M. C. **Análise de sensibilidade e resíduos em modelos de regressão com respostas bivariadas por meio de cópulas**. 2007. Dissertação (Mestrado em Agronomia) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba
- SILVA, G. O. **Modelos de regressão quando a função de taxa de falha não é monótona e o modelo probabilístico beta Weibull modificada**. 2008. Dissertação (Doutorado em Agronomia) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba
- CARRASCO, J. M. F. **Modelo de regressão log-Weibull modificado e a nova distribuição Weibull modificada generalizada**. 2007. Dissertação (Mestrado em Agronomia) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba
- CARDIAL, M. R. P. **Distribuição Weibull Discreta Exponenciada para dados com presença de censura: uma abordagem clássica e bayesiana**. 2017. Dissertação (Mestrado em Estatística) - Departamento de Estatística do Instituto de Ciências Exatas, Universidade de Brasília, Brasília
- CASELLA, G. e BERGER, R. L. (2014). **Inferência Estatística**. 2.ed. São Paulo, Brasil
- KLEIN, J.P.; MOESCHBERGER, M.L. (1997). **Survival Analysis: Thechniques for Censored and Truncated Data**. New York: Springer-Verlang p.357.
- R Core Team (2017). R: A language and enviroment for statistical computing. Vienna, Austria. Disponível em: <http://www.R-project.org/>.



## ANEXOS



## Anexo A: Exemplo de programa em R utilizado para simulação de cenários sem covariável

A seguir, será apresentada a programação utilizada para simular tempos de falha com distribuição Exponencial, com censura e sem covariável.

```

r <- 2
c <- 0.1 #percentual censura
n <- 30

final1t <- matrix(0, ncol = (n + 5))

pb <- winProgressBar(title = "Progresso",
label = "0%", min = 0, max = 100, initial = 0)

simus <- 1001
i <- 1
while(nrow(final1t) < simus){

k <- sample(seq(0.02, 3, 0.01), 1, replace = FALSE)

n <- n

set.seed(i)

tempo <- rexp(n, r)
censura <- runif(n, 0, k*max(tempo))
delta <- ifelse(tempo <= censura, 1, 0)

if(((n - sum(delta))/n) == c){ # Se o percentual de censura for igual ao definido,
# a programação segue. Caso contrário, outro k é sorteado

mod1 <- survreg(Surv(tempo, delta1) ~ 1, dist = "exponential")
alphai <- exp(mod2$coefficients[1])
surv <- as.vector(exp(-(tempo/alphai)))
ei <- as.vector(-log((surv)))
eit <- ifelse(delta==1,ei,ei+1)
eio <- sort(eit)
temp <- c(n, i, k, (n - sum(delta1))/n, alphai, eio)

final1t <- rbind(final1t, temp)

}

i <- i + 1

```

74

```
setWinProgressBar(pb, (nrow(final1t) - 1)/1000*100,  
label = sprintf("%.1f%% das amostras geradas",  
round((nrow(final1t) - 1)/1000*100, 1)))
```

```
}
```

```
final1 <- final1t[-1,]
```



## Anexo B: Exemplo de programa em R utilizado para simulação de cenários com covariável

A seguir, será apresentada a programação utilizada para simular tempos de falha com distribuição Exponencial, com censura e com covariável.

```

beta0 <- 8.47
beta1 <- -1.11
c <- 0.1 #percentual censura
n <- 30

final2t <- matrix(0, ncol = (n + 8))
pb <- winProgressBar(title = "Progresso",
label = "0%", min = 0, max = 100, initial = 0)
simus <- 1001
i <- 1

while(nrow(final2t) < simus){

k <- sample(seq(0.01, 4, 0.01), 1, replace = FALSE)
set.seed(i)
covar <- runif(n,0,1)
alpha <- exp(beta0 + (beta1*covar))
tempo <- rexp(n,(1/alpha))
censura <- runif(n,0,k*max(tempo))
delta <- ifelse(tempo <= censura,1,0)

if(((n - sum(delta))/n) == c){
mod2 <- survreg(Surv(log(tempo),delta1)~covar,dist="extreme",scale=1)
b0 <- mod2$coefficients[1]
b1 <- mod2$coefficients[2]
mi <- (b0 + b1*covar)
surv <- as.vector(exp(-exp(log(tempo)-mi)))
ei <- as.vector(-log(surv))
eit <- ifelse(delta==1,ei,ei+1)
eio <- sort(eit)
temp <- c(n, i, k, (n - sum(delta))/n,b0,b1,eio)

final2t <- rbind(final2t, temp)
}

i <- i + 1

setWinProgressBar(pb, (nrow(final2t) - 1)/1000*100,
label = sprintf("%.1f%% das amostras geradas",
round((nrow(final2t) - 1)/1000*100, 1)))

```

```
}
```

```
final2 <- final2t[-1,]
```