



Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

**Análise do tempo até a re-hospitalização de pacientes
com esquizofrenia via técnicas de Análise de
Sobrevivência**

Marcia Araujo Maia

Trabalho de Conclusão de Curso apresentado ao Departamento de Estatística da Universidade de Brasília, como parte dos requisitos para a obtenção do título de Bacharel em Estatística.

**Brasília
2015**

Marcia Araujo Maia

**Análise do tempo até a re-hospitalização de pacientes
com esquizofrenia via técnicas de Análise de
Sobrevivência**

Trabalho de Conclusão de Curso apresentado ao
Departamento de Estatística da Universidade de
Brasília, como parte dos requisitos para a obtenção
do título de Bacharel em Estatística.

Orientador: Prof. Dr. **Eduardo Yoshio Nakano**

Brasília
2015

Agradecimentos

Agradeço a Deus por fazer-me sonhar e realizar os meus sonhos. Estou sendo muito abençoada nesse momento e sei que não é mérito meu, e sim dádiva d'Ele na minha vida.

Agradeço ao Renato Lopes Fagundes e à Thais Rodrigues por terem apresentado-me à estatística e pelo incentivo para ser estudante desse curso. Muito obrigada, estou apaixonada pela estatística.

Agradeço ao professor Eduardo Yoshio Nakano por orientar-me, por toda a prestatividade, por compartilhar seu conhecimento e por incentivar-me a continuar na vida acadêmica

Agradeço à tia Mônica por todo carinho, amor, apoio, por incentivar-me sempre a querer evoluir, e por ser um exemplo de pessoa bondosa e altruísta. Agradeço à Sandrinha por todo apoio, incentivo e por me deixar dormir em sua casa nas vésperas de provas.

Agradeço à minha prima Juliana por ensinar-me desde criança a tomar gosto pelos estudos. Obrigada por alfabetizar-me. E, por sempre estar comigo quando precisei.

Agradeço à Carol (secretaria do CIC) por me consolar num momento difícil e dizer que tudo ia dar certo. Obrigada, deu certo.

Agradeço ao professor Demerson André Polli por orientar-me no PIBIC, por ser amigo e ajudar-me nas dificuldades com os conteúdos.

Agradeço à professora Julina Betini Fachini Gomes por ser uma excelente professora, por passar seu conhecimento, colaborando para que esse trabalho fosse realizado. Obrigada por ser tão compreensiva e pelas palavras de incentivo enquanto coordenadora.

Agradeço à professora Maria Tereza por toda paciência, por ser tão prestativa, por todo ensinamento e por sempre fazer o possível para resolver os problemas que apareciam.

Agradeço aos meus colegas de trabalho da Gerência de Risco de Mercado (GERIM-Caixa Econômica Federal) por todo ensinamento, pela paciência e pela oportunidade. Vocês foram essenciais para o meu aprendizado. Em especial, obrigada à Cleusa por tratar-me como uma filha, à Juliana Puga por ser uma excelente estatista e por me passar parte da sua experiência.

Agradeço às minhas amigas que fizeram parte do grupo de sobrevivência (Rhayssa, Mariana, Maria Gabriela e Andressa), pois usamos o meu banco de dados em nosso trabalho. Obrigada por darem o ponto de vista de vocês sobre as técnicas que devíamos

usar e por deixarem eu usar o que já havíamos feito em meu trabalho.

Agradeço ao pessoal da secretaria da estatística, ao pessoal da portaria, em especial, à Verinha por ser amiga e por sempre nos incentivar nos momentos difíceis e por fazer com que a minha vida e dos meus amigos não fosse tão difícil.

Agradeço a todos meus amigos que estiveram comigo ao longo da graduação, em especial, Adolfo, Pablo, Rosi, Nahari, Bianca Dorneles, Bianca Souza, Hudson, Susana, Maria Júlia, Poliana, Vanessa Rezende, Patrícia Akemi, Cadu, Pedro Oliveira, Pedro Rangel, Japa (Lucas Nanami), Isabella e Lucas Alves, Fernanda, Alex Luís e Rachmine, por todos os momentos de estudo, de brincadeira e alegria.

Agradeço ao bonde da estatística (Geiziane, Mariana, Agda, Rodrigo, Mateus, Andressa, Claudia, Érica Rios, Maria Gabriela e Yuri) por todo tempo de estudo juntos, pela força que me dão quando acho que as coisas não vão dar certo. Vocês são muito importante para minha vida acadêmica e para minha vida pessoal. Em especial, o meu muito obrigada à Geiziane por ser um exemplo de dedicação e por ajudar-me sempre.

Resumo

O objetivo desse trabalho foi obter um modelo para estimar o tempo entre pacientes com esquizofrenia tratados por diferentes tipos de antipsicóticos receberem alta e voltarem a ser re-hospitalizados. Essa informação é de grande importância, pois pode ser utilizada como uma medida de efetividade do medicamento, então pode auxiliar o psiquiatra na escolha do melhor tratamento a ser utilizado. A análise considerou uma amostra de 132 pacientes com esquizofrenia e foi realizada utilizando o modelo paramétrico Log-Normal e o modelo de riscos proporcionais de Cox. Toda a análise foi realizada pelo software livre R.

Palavras-chave: Análise de sobrevivência, dados censurados, antipsicóticos, medicamentos.

Sumário

1	Introdução e Justificativa	1
	Introdução	1
2	Revisão Bibliográfica	3
2.1	Conceitos básicos de Análise de Sobrevivência	3
2.1.1	Censura	3
2.1.2	Descrição do Tempo de Sobrevivência	5
2.1.3	Função de Risco Acumulada	7
2.1.4	Relações entre as funções	7
2.1.5	Estimação Não Paramétrica da Função de sobrevivência	8
2.1.6	Comparação de curvas de sobrevivência	9
2.1.7	Determinação Empírica da Função de Risco	9
2.1.8	Principais Modelos Paramétricos em Análise de Sobrevivência . . .	10
2.1.9	Estimação dos Parâmetros dos Modelos	13
2.1.10	Adequação do Modelo Ajustado	15
2.2	Modelo de Regressão de Cox	16
2.2.1	Estimação dos parâmetros no modelo de Cox	17
2.2.2	Estimando funções relacionadas a $h_0(t)$	17
2.2.3	Adequação do modelo de Cox	18
3	Resultados e Discussões	21
3.1	Ajuste do modelo	28
3.1.1	Ajuste do modelo paramétrico	28
3.1.2	Ajuste do modelo de Cox	31
3.2	Interpretando as estimativas dos parâmetros	35
4	Conclusões	39
	Referências Bibliográficas	41

Lista de Figuras

2.1	Tipos de mecanismos de censura à direita. (Colosimo;Giolo, 2006).	5
2.2	Formas da curva do Tempo Total em Teste - TTT	10
3.1	Função de Sobrevivência do tempo até a reinternação do paciente	23
3.2	Função de Sobrevivência do tempo até a reinternação do paciente por grupo de medicação	24
3.3	Gráficos Kaplan-Meier para as covariáveis medidas no estudo do tempo até a re-hospitalização de pacientes com esquizofrenia	27
3.4	Curvas TTT dos dados do tempo até a re-hospitalização de pacientes com esquizofrenia	29
3.5	Função de Sobrevivência estimada por Kaplan-Meier e pelos modelos paramétricos Weibull, Exponencial, Log-Normal e Log-Logística	29
3.6	$\hat{S}(t)$ dos resíduos de Cox-Snell estimada pelo método de Kaplan-Meier e pelo modelo exponencial padrão e curva de $\hat{S}(t)$ estimada	30
3.7	Resíduos padronizados de Schoenfeld para as covariáveis do modelo de Cox final	33
3.8	Gráficos da função de sobrevivência estimada dos resíduos de Cox-Snell e Exponencial Padrão.	34

Lista de Tabelas

3.1	Descrição das covariáveis do estudo	21
3.2	Frequências absolutas e relativas das covariáveis	22
3.3	Testes log-rank e de Wilcoxon utilizados para testar a igualdade das curvas de sobrevivência obtidas para as covariáveis consideradas neste estudo . . .	24
3.4	Descrição das covariáveis do estudo com observações referentes ao segundo grupo de antipsicóticos e à Clozapina.	25
3.5	Frequências absolutas e relativas das covariáveis com observações referentes ao segundo grupo de antipsicóticos e à Clozapina.	26
3.6	Testes log-rank e de Wilcoxon utilizados para testar a igualdade das curvas de sobrevivência obtidas para as covariáveis consideradas neste estudo . . .	28
3.7	Critérios de seleção de variáveis AIC, AICc e BIC paos modelos Exponencial, de Weibull, Log-Normal e Log-Logística	30
3.8	Teste de proporcionalidade para as funções de risco do modelo final de Cox	32
3.9	Resultado do ajuste do modelo de Cox para os dados do tempo até a re-hospitalização de pacientes com esquizofrenia e correspondentes razões de riscos (RR).	35
3.10	Valor do Fator Inflacionário da Variância (FIV) para cada covariável selecionada para o modelo.	35
3.11	Estimativas da função de sobrevivência de base $\hat{S}_0(t)$	36
A.1	Modelo com apenas uma covariável	45
A.2	Modelo: $X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8$	45
A.3	Modelo: $X_1 + X_2 + X_3 + X_4 + X_6 + X_7 + X_8$	46
A.4	$X_1 + X_2 + X_3 + X_4 + X_6 + X_8$	46
A.5	Modelo: $X_1 + X_2 + X_4 + X_6 + X_8$	46
A.6	Modelo: $X_2 + X_4 + X_6 + X_8$	46
A.7	Modelo: $X_2 + X_6 + X_8$	47

Capítulo 1

Introdução e Justificativa

O termo esquizofrenia vem do grego e significa mente dividida. Essa doença faz com que haja uma separação entre a imaginação e a realidade. Segundo Silva (2006), a definição atual indica uma psicose crônica idiopática, aparentando ser um conjunto de diferentes doenças com sintomas que se assemelham e se sobrepõem. Alguns dos sintomas da esquizofrenia são alucinações, delírios, alterações do comportamento, do pensamento, dentre outros.

Os pacientes com esquizofrenia precisam receber tratamento por toda vida, e algumas maneiras conhecidas para tratar a doença são o uso de medicação, a prática de terapia, e diversos outros métodos. No caso do tratamento por medicação, o tempo entre as re-hospitalizações é um importante indicador de efetividade do medicamento, e ter um modelo que estime corretamente o tempo até a re-hospitalização dos pacientes permitirá avaliar o grau de efetividade do medicamento administrado no paciente e decidir se é necessária, ou não, a sua substituição. Os diferentes tipos de antipsicóticos podem apresentar efeitos colaterais diversos e uma estimativa precisa de sua eficiência permitirá a melhor escolha do medicamento para o paciente.

Em situações como essa, e em diversas outras áreas de conhecimento, é importante a realização de análises estatísticas. Além disso, uma técnica estatística que é bastante utilizada, principalmente em estudos relacionados à saúde, é a Análise de Sobrevida.

Nesse contexto, o objetivo geral deste trabalho é ajustar um modelo de sobrevivência que estime o tempo até a re-hospitalização de pacientes com esquizofrenia tratados por diferentes tipos de antipsicóticos. Além disso, tem-se como objetivos específicos: ajustar um modelo paramétrico e verificar a qualidade do ajuste, pois esses modelos são mais precisos; e ajustar o modelo semi-paramétrico de Cox, pois esse modelo é bastante flexível por possuir uma parte não paramétrica.

A possibilidade de se obter um modelo que estime com precisão o tempo entre as rehospitalizações de pacientes com esquizofrenia é de grande importância para medir a efetividade do tratamento, e pode auxiliar os psiquiatras na escolha do tratamento a ser utilizado no paciente com esquizofrenia.

Essa pesquisa foi realizada no Instituto de Psiquiatria (IPq) da Universidade de São Paulo (USP) no período de 1 de dezembro de 1997 a 31 de dezembro de 2004 e, o tempo nesse estudo foi observado em dias. Este estudo considerou uma amostra de 190 pacientes e oito covariáveis, dados que fazem parte do trabalho de Werneck et al. (2011). Todas as análises foram realizadas pelo software free R (R CORE TEAM, 2013).

Capítulo 2

Revisão Bibliográfica

2.1 Conceitos básicos de Análise de Sobrevivência

A Análise de Sobrevivência é uma técnica estatística que estuda dados relacionados ao tempo decorrido até a ocorrência do evento de interesse, o qual precisa ser claramente definido. Esse é definido como variável de interesse.

Os conjuntos de dados de sobrevivência são caracterizados pelos tempos de falha e, muito frequentemente, pelas censuras (Colosimo, E.A., et al, 2006). O tempo até a ocorrência de um evento de interesse é denominado tempo de falha. E, segundo Garcia (2013) a presença de observações incompletas ou parciais são denominadas censura.

Tempo de falha

Em Análise de sobrevivência, o tempo até a ocorrência do evento de interesse é chamado de tempo de falha (ou tempo de sobrevivência).

Para analisar o tempo de falha é necessário:

- fixar o tempo de início de estudo;
- definir a escala de medida a ser utilizada;
- e estabelecer o evento de interesse, que frequentemente é indesejável e conhecido como falha.

2.1.1 Censura

A censura ocorre quando há perda de informação decorrente de não se ter observado o momento exato da ocorrência do desfecho, resultando em observações parciais ou incompletas. Ou seja, o evento de interesse não pode ser observado por algum motivo. As observações censuradas devem ser incluídas na análise dos dados, pois mesmo sendo incompletas, essas observações fornecem informações sobre o tempo de vida de objetos e

indivíduos e a omissão das censuras pode acarretar conclusões viciadas na análise estatística. Dessa forma, existe a necessidade da introdução de uma variável extra na análise, que indica se o valor do tempo de sobrevivência de um determinado indivíduo foi, ou não observado. Essa variável, geralmente representada por δ , é conhecida como variável indicadora de censura e é expressa por:

$$\delta = \begin{cases} 1, & \text{se o } t \text{ indivíduo falhou} \\ 0, & \text{se o } t \text{ indivíduo foi censurado} \end{cases} \quad (2.1)$$

Em cada estudo a censura pode ocorrer por motivos diferentes. Por isso, existem formas distintas de censura. Elas podem ser censura a direita, censura a esquerda e censura intervalar.

Censura à direita

A censura à direita é caracterizada pelo tempo de ocorrência do evento estar à direita do tempo registrado, ou seja, do tempo de início do estudo. Essa, por sua vez, abrange alguns mecanismos de censura:

1. Censura tipo I: o estudo será terminado após um período de tempo fixo, t_f , e nesse ínterim de tempo uma ou mais observações em estudo não falham. O tempo de início t_0 e o tempo final do estudo t_f devem ser determinado antes do início do estudo;
2. Censura tipo II: o estudo terminará após ocorrer um número fixo k ($k \leq n$) de falhas. Esse número deve ser determinado antes do início do estudo;
3. Censura aleatória: são todos os casos em que as observações não experimentam o evento de interesse por motivos não controláveis.

A seguir estão ilustrados os mecanismos de censura à direita. A falha é representada por "●" e a censura é representada "○".

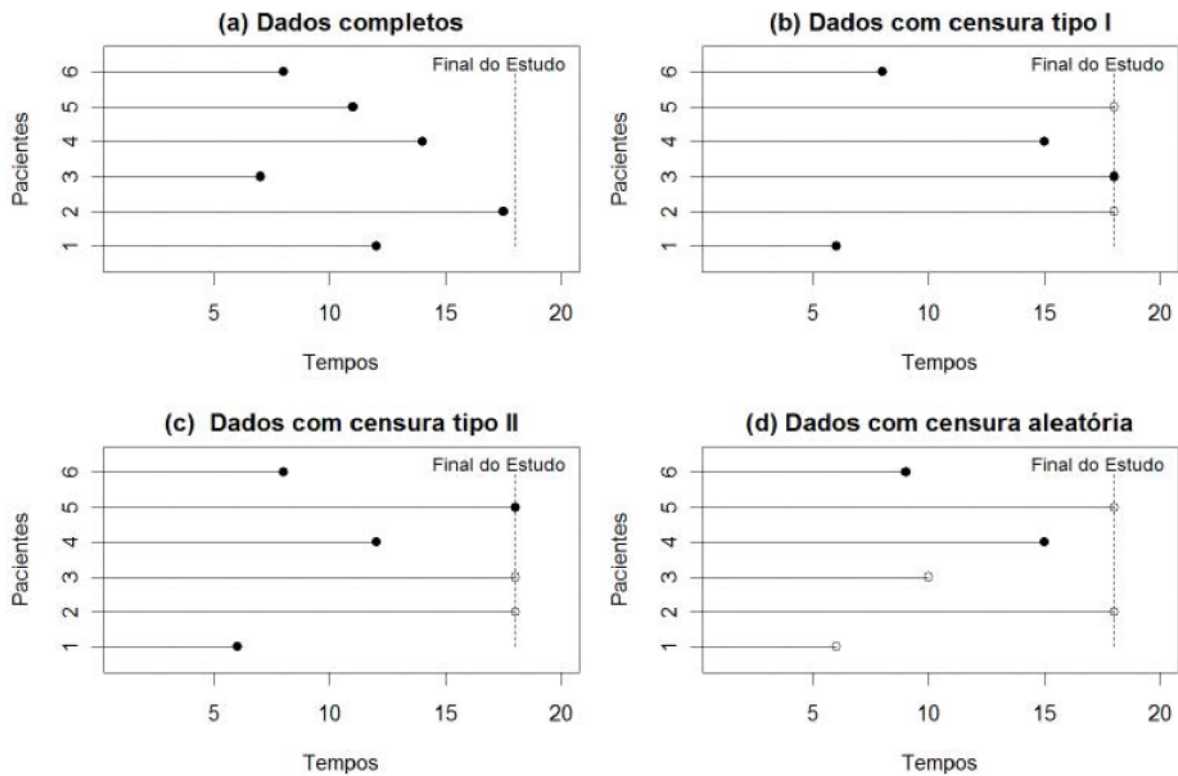


Figura 2.1: Tipos de mecanismos de censura à direita. (Colosimo;Giolo, 2006).

A análise de dados deste trabalho irá considerar a censura à direita.

Censura à esquerda

A censura à esquerda acontece quando o evento de interesse já ocorreu quando o indivíduo começou a ser observado, ou seja, o evento de interesse ocorreu antes do estudo começar.

Censura intervalar

A censura intervalar ocorre em estudos nos quais os indivíduos são acompanhados periodicamente. O evento de interesse ocorre em um intervalo de tempo. Assim, o tempo de falha não é conhecido precisamente, mas pertence a um determinado intervalo.

2.1.2 Descrição do Tempo de Sobrevivência

Os dados de sobrevivência são usualmente representados pela variável aleatória não negativa T , geralmente contínua. Essa variável representa o tempo de sobrevivência do indivíduo. Algumas maneiras de especificá-la é por meio da função densidade de probabilidade, $f(t)$, função de sobrevivência, $S(t)$; função de risco, $h(t)$; e por meio de relações matemáticas existentes entre essas três funções. Trabalhos que tratam de dados de sobrevivência discretos podem ser vistos em Nakano e Carrasco (2006), Carrasco et al. (2012) e Brunello e Nakano (2015).

A variável indicadora de falha ou censura, δ_i , e o tempo de falha, t_i , representam os dados de sobrevivência para o indivíduo i ($i = 1, \dots, n$). Se houver um vetor de covariáveis x_i , com $i = 1, \dots, n$, os dados de sobrevivência serão representados por (t_i, δ_i, x_i) .

Função Densidade de Probabilidade

A função densidade de probabilidade é definida como o limite da probabilidade de um indivíduo experimentar o evento de interesse em um intervalo de tempo $[T_i, T_f + \Delta_t]$. Essa função é expressa por:

$$f(t) = P\left(\lim_{\Delta_t \rightarrow 0} \frac{T_i \leq T \leq T_i + \Delta_t}{\Delta_t}\right), \quad (2.2)$$

$f(t) \geq 0 \forall T$, e a área abaixo de $f(t)$ é igual a 1.

Função de Distribuição

A função de distribuição também é conhecida como Função de Distribuição Acumulada, pois acumula as probabilidades dos valores inferiores ou iguais a t . Essa função comporta muitas informações sobre a variável aleatória estudada (Magalhães, 2006). Sua expressão matemática é dada por:

$$F(t) = P(T \in (-\infty, t]) = P(T \leq t), \quad (2.3)$$

no qual, $t \in \mathbb{R}$.

$F(t)$ possui as seguintes propriedades:

1. $\lim_{t \rightarrow -\infty} F(t) = 0$;
2. $\lim_{t \rightarrow \infty} F(t) = 1$;
3. $F(t)$ é contínua à direita;
4. $F(t)$ é não decrescente, ou seja, $F(t) \leq F(y)$, sempre que $t \leq y$, $\forall t, y \in \mathbb{R}$.

Função de Sobrevivência

A função de sobrevivência, $S(t)$, é definida como a probabilidade de uma observação não falhar até um certo tempo t , ou seja, é a probabilidade de uma observação sobreviver ao tempo t (Colosimo; Giolo, 2006). Em termos probabilísticos essa função é expressa por:

$$S(t) = P(T \geq t), \quad (2.4)$$

na qual, $S(t)$ é uma função monótona decrescente e contínua.

Função de Risco

A função de risco, também chamada de Função de Taxa de Falha, é caracterizada pelo limite da probabilidade da falha ocorrer no intervalo de tempo $[t, t + \Delta_t)$, condicional à sobrevivência até o tempo t , dividida pelo comprimento do intervalo de tempo Δ_t . Essa expressão assume a seguinte forma:

$$h(t) = P\left(\lim_{\Delta_t \rightarrow 0} \frac{t \leq T < t + \Delta_t | T \geq t}{\Delta_t}\right), \quad (2.5)$$

A função $h(t)$ pode assumir a forma crescente, constante ou decrescente. Além disso, pode assumir a forma unimodal ou a forma de curva da banheira.

2.1.3 Função de Risco Acumulada

A função de risco acumulada determina a taxa de falha que é acumulada até o tempo t , e pode ser usada para obter $h(t)$ na estimação não paramétrica. Sua expressão é dada por:

$$H(t) = \int_0^t h(u) du. \quad (2.6)$$

2.1.4 Relações entre as funções

Segundo Colosimo e Giolo (2006), para uma variável aleatória T contínua e não negativa, podem ser definidas relações relevantes entre elas a partir das funções definidas anteriormente. Algumas relações importantes são:

$$f(t) = \frac{d}{dt} F(t), \quad (2.7)$$

$$F(t) = 1 - S(t), \quad (2.8)$$

$$f(t) = \frac{d}{dt}(1 - S(t)) = -\frac{d}{dt} S(t), \quad (2.9)$$

$$h(t) = \frac{f(t)}{S(t)}, \quad (2.10)$$

$$h(t) = \frac{-\frac{d}{dt}S(t)}{S(t)} = -\frac{d}{dt}\log[S(t)], \quad (2.11)$$

$$\log[S(t)] = -\int_0^t h(u)du \quad e \quad (2.12)$$

$$S(t) = \exp[-H(t)]. \quad (2.13)$$

2.1.5 Estimação Não Paramétrica da Função de sobrevivência

Observações censuradas é um problema para as técnicas convencionais de análise estatística descritiva, que consiste principalmente em encontrar medidas de tendência central e variabilidade, pois a interpretação dos resultados se tornam mais difíceis. Sendo assim, o principal componente da análise descritiva, envolvendo dados de tempo de vida, é a função de sobrevivência $S(t)$ (Colosimo; Giolo, 2006). Pode-se estimá-la a partir do estimador não paramétrico de Kaplan-Meier (Kaplan e Meier, 1958).

Estimador de Kaplan-Meier

O estimador não paramétrico de Kaplan-Meier (Kaplan; Meier, 1958), também chamado de estimador limite-produto, é uma adaptação da função de sobrevivência (Colosimo; Giolo, 2006).

O estimador de Kaplan-Meier é definido como:

$$\hat{S} = \prod_{i:t_i < t} \frac{n_i - d_i}{n_i} = \prod_{i:t_i < t} \left(1 - \frac{d_i}{n_i}\right), \quad (2.14)$$

no qual, $t_1 < \dots < t_k$ são os k valores distintos dos tempos de falha; d_i é o número de falhas no instante t_i , $i = 1, \dots, k$; e n_i é o número de indivíduos sob risco (sobreviventes e não censurados) em até t_i .

As propriedades principais desse estimador são:

- não viciado para grandes amostras;
- fracamente consistente;
- converge assintoticamente para um processo gaussiano;
- estimador de máxima verossimilhança de $S(t)$.

A variância assintótica do estimador de Kaplan- Meier é dada por:

$$\widehat{Var\hat{S}(t)} = [\hat{S}(t)]^2 \sum_{i:t_i < t} \frac{d_i}{n_i(n_i - d_i)} \quad (2.15)$$

E, levando em consideração que $\hat{S}(t)$ tem distribuição assintótica Normal, para t fixo, um intervalo aproximado de $100(1 - \alpha)\%$ de confiança para $\hat{S}(t)$ é dado por:

$$\hat{S}(t) \pm Z_{\alpha/2} \sqrt{\widehat{Var(\hat{S}(t))}} \quad (2.16)$$

no qual, $Z_{\alpha/2}$ indica o $\alpha/2$ quantil da distribuição Normal padrão.

2.1.6 Comparação de curvas de sobrevivência

Quando existem covariáveis tem-se o objetivo de comparar seu efeito no tempo de sobrevivência. Quando as covariáveis são categóricas, podem ser aplicados os testes de *logrank* e/ou de Wilcoxon.

O teste *logrank* (Mantel, 1966) é apropriado quando razão das funções de risco dos grupos a serem comparados são aproximadamente constante, ou seja, quando os grupos têm riscos proporcionais.

O teste de Wilcoxon não tem exigência de proporcionalidade. Além disso, ele utiliza peso igual ao número de indivíduos sob risco, isto é, coloca mais peso na porção inicial do eixo do tempo, pois no início do estudo, todos os indivíduos estão sob risco e à medida que falham ou são censurados vão saindo do estado "sob risco".

2.1.7 Determinação Empírica da Função de Risco

A forma da função de risco será identificada por meio do método gráfico baseado no tempo total em teste (TTT) transformado (Barlow; Campo, 1975). O intuito é encontrar o modelo adequado para os tempos de vida. Contudo, esse método só será eficiente se não houver censuras, ou se existir poucas, pois a curva do gráfico não as leva em consideração. Nesse contexto, seu uso deve ser feito com cautela, principalmente quando a proporção de observações censuradas é alta.

A expressão empírica do TTT transformado (Aarset, 1985) é a seguinte:

$$G(r/n) = \frac{\sum_{i=1}^r t_i - (n - r)t_r}{\sum_{i=1}^r t_i} \quad (2.17)$$

no qual, $r = 1, \dots, n$ e t_i , com $i = 1, \dots, n$, representa as estatísticas dos tempos, $i = 1, \dots, n$.

A Figura 1.2 mostra várias formas que a curva TTT pode assumir.

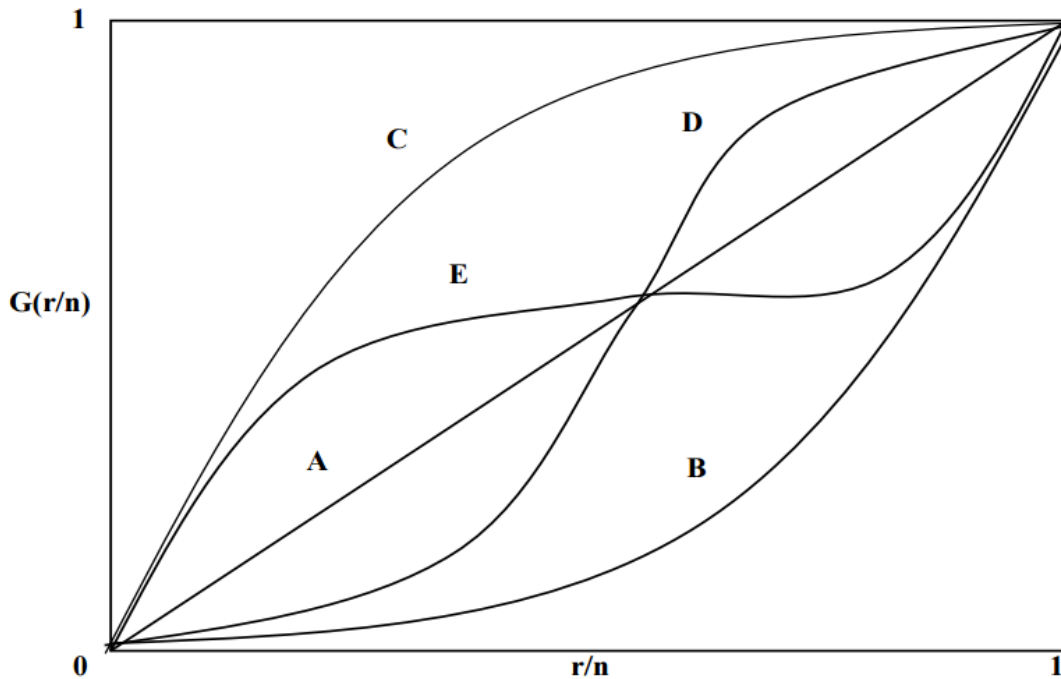


Figura 2.2: Formas da curva do Tempo Total em Teste - TTT

Se a curva TTT assume a forma de uma reta, como no modelo A, é adequado trabalhar com uma função de risco constante. Nesse caso, a distribuição exponencial é indicada, pois sua função de risco é constante para todo o tempo observado.

Quando a curva é convexa, modelo B, ou côncava, modelo C, a função de risco é monotonamente, a função de risco é monotonamente decrescente ou crescente, respectivamente. Assim, a distribuição a ser utilizada será a Weibull, pois apresenta função de risco decrescente quando o parâmetro de forma γ é menor que 1 e crescente quando γ é maior que 1.

Caso a curva apresente uma forma convexa e depois côncava, como na curva D, trata-se de uma função de risco que toma forma de banheira. As distribuições, referentes à essa função de risco, indicada para modelagem dos dados são: Weibull exponencializada, Weibull modificada, Weibull aditiva, Burr XII Aditiva, Beta Weibull Generalizada, entre outras.

Se a curva for côncava e em seguida convexa, a função de risco possui característica unimodal. Dessa forma é adequado utilizar distribuições como a Log-Normal e a Log-Logística.

2.1.8 Principais Modelos Paramétricos em Análise de Sobrevida

Os modelos paramétricos são utilizados para modelar o tempo de sobrevivência até a ocorrência do evento de interesse de forma mais fidedigna (Costa, 2013). Alguns

modelos recebem notoriedade por serem adequados à várias situações. Alguns desses modelos que serão utilizados neste trabalho são as distribuições Weibull, Exponencial, Log-Normal e Log-Logística.

Distribuição Weibull

A distribuição Weibull é popular por apresentar uma grande variedade de formas. A propriedade básica de todas essas formas é que a função de risco é monótona, ou seja, ela é crescente, decrescente ou constante.

A variável aleatória tempo de falha T , descrito pela distribuição Weibull, tem a seguinte função de densidade de probabilidade:

$$f(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1} \exp \left\{ - \left(\frac{t}{\alpha} \right)^\gamma \right\}, \quad (2.18)$$

no qual, $t \geq 0$ e o parâmetro de escala, $\alpha \geq 0$, tem a mesma unidade de medida de t e o parâmetro de forma, $\gamma \geq 0$, não tem unidade

A função de sobrevivência da distribuição Weibull é dada por:

$$S(t) = \exp \left\{ - \left(\frac{t}{\alpha} \right)^\gamma \right\}, t \geq 0. \quad (2.19)$$

E a função de risco é dada por:

$$h(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1}, t \geq 0. \quad (2.20)$$

A função de risco, $h(t)$, é estritamente crescente quando $\gamma > 1$, estritamente decrescente quando $\gamma < 1$ e constante quando $\gamma = 1$

Distribuição Exponencial

A distribuição exponencial é um caso particular da distribuição Weibull. Esse modelo é um dos mais simples usados para descrever o tempo de falha. Apresenta apenas um parâmetro e o único que tem função de risco constante (propriedade de falta de memória). Por esse motivo, é indicado que essa distribuição seja usada quando o tempo é curto.

A variável aleatória tempo de falha T , descrito pela distribuição exponencial, tem a seguinte função de densidade de probabilidade:

$$f(t) = \frac{1}{\alpha} \exp \left\{ - \left(\frac{t}{\alpha} \right) \right\}, \quad (2.21)$$

no qual, $t \geq 0$ e o parâmetro $\alpha > 0$ é o tempo médio de vida e tem a mesma unidade do tempo de falha T .

Além disso, a função de sobrevivência $S(t)$ do modelo em questão é dada por:

$$S(t) = \exp \left\{ - \left(\frac{t}{\alpha} \right) \right\}, t \geq 0. \quad (2.22)$$

E a função de risco é dada por:

$$h(t) = \frac{1}{\alpha}, t \geq 0. \quad (2.23)$$

Distribuição Log-Normal

A distribuição log-normal, assim como a distribuição weibull, é útil para caracterizar tempos de vida de indivíduos e produtos. A variável aleatória tempo de falha T , com distribuição log-normal, tem a seguinte função densidade:

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp \left\{ - \frac{1}{2} \left(\frac{\log(t) - \mu}{\sigma} \right)^2 \right\}, t \geq 0. \quad (2.24)$$

Aqui, $-\infty \leq \mu \leq \infty$ é a média do logaritmo do tempo de falha T , e $\sigma > 0$ é o desvio padrão.

Os dados provenientes de uma distribuição log-normal podem ser analisados segundo uma distribuição normal, considerando o logaritmo dos dados em vez dos valores originais, isto é, $\log(t) \sim \text{Normal}(\mu, \sigma^2)$.

Uma variável log-normal não apresenta uma forma analítica explícita para as funções de sobrevivência e de função de risco, sendo representadas, respectivamente, da seguinte maneira:

$$S(t) = \Phi \left(\frac{-\log(t) + \mu}{\sigma} \right) \quad (2.25)$$

e

$$h(t) = \frac{f(t)}{S(t)}, \quad (2.26)$$

no qual, $\Phi(\cdot)$ é a função de distribuição acumulada de uma normal padrão.

Distribuição Log-Logística

A variável aleatória tempo de falha T cuja distribuição é dada pelo modelo log-logístico, com parâmetros α e γ , tem a seguinte função de densidade:

$$f(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1} (1 + (t/\alpha))^{-2}, \quad t > 0. \quad (2.27)$$

Aqui, $\alpha > 0$ e $\gamma > 0$ são os parâmetros de escala e forma da distribuição, respectivamente.

A função de sobrevivência $S(t)$, a função de risco $h(t)$ são expressos da seguinte maneira:

$$S(t) = \frac{1}{1 + (t/\alpha)^\gamma}, \quad t > 0 \quad (2.28)$$

e

$$h(t) = \frac{\gamma(t/\alpha)^{\gamma-1}}{\alpha[1 + (t/\alpha)^\gamma]}, \quad t > 0. \quad (2.29)$$

2.1.9 Estimação dos Parâmetros dos Modelos

Os estudos que envolvem tempos de falha são modelados por distribuições. Os parâmetros dessas distribuições devem ser estimados por meio das observações amostrais. Assim, é possível responder às perguntas de interesse.

Um dos métodos mais conhecidos de estimação, dentro do contexto de regressão linear, é o método de mínimos quadrados. Contudo, usá-lo na análise de estudos de tempo de vida é inapropriado, pois não é possível incorporar censura em seu processo.

Outro método relevante de estimação na literatura é o de máxima verossimilhança. Esse, por sua vez, é mais apropriado para estudos em análise de sobrevivência, pois é possível incorporar censuras a ele. Além disso, é relativamente trivial de ser compreendido e possui ótimas propriedades para grandes amostras.

Método de Máxima Verossimilhança

O estimador de máxima verossimilhança, dado uma distribuição, escolhe valores para os parâmetros que melhor explique a amostra observada. Dessa forma, o objetivo é encontrar o valor do vetor de parâmetros θ que maximiza a função de verossimilhança, $L(\theta)$, a qual é expressa por:

$$L(\theta) = \prod_{i=1}^n f(t_i; \theta), \quad (2.30)$$

no qual t_1, \dots, t_n representam os tempos observados de uma amostra de tamanho n de uma população.

A função de verossimilhança mostra que cada observação não censurada contribui com sua função densidade $f(t)$. E, cada informação censurada à direita contribui informando que o tempo de falha é maior que o tempo de censura observado, ou seja, contribuem com sua função de sobrevivência $S(t)$.

Dessa forma, a função de verossimilhança levará em consideração a censura à direita, que é a censura considerada neste trabalho. Assim, tem-se a seguinte expressão para a função de verossimilhança:

$$L(\theta) \propto \prod_{i=1}^n [f(t_i; \theta)]^{\delta_i} [S(t_i; \theta)]^{1-\delta_i}, \quad (2.31)$$

no qual, θ é o vetor de parâmetros a ser estimado e δ_i é a variável indicadora de falha ou censura.

É sempre conveniente, contudo, trabalhar com o logaritmo da função de verossimilhança. Segundo Colosimo e Giolo (2006), os estimadores de máxima verossimilhança são os valores de θ que maximizam $L(\theta)$ ou equivalentemente o logaritmo de $L(\theta)$, isto é, $\log(L(\theta))$. Esses estimadores são encontrados resolvendo-se o sistema de equações:

$$U(\theta) = \frac{\partial \log L(\theta)}{\partial \theta} = 0, \quad (2.32)$$

no qual, $l(\theta) = \log(L(\theta))$ é dado por:

$$l(\theta) = \sum_{i=1}^n \delta_i \log[f(t_i; \theta)] + (1 - \delta_i) \log[S(t_i; \theta)]. \quad (2.33)$$

Teste da Razão de Verossimilhanças

Geralmente, há interesse em testar hipóteses relacionadas ao vetor de parâmetros $\theta = (\theta_1, \dots, \theta_p)'$, ou relacionada a um subconjunto desse vetor. Um teste bastante utilizado para isso é o teste da razão de verossimilhanças. A estatística do teste é dada pela razão dos valores do logaritmo da função de verossimilhança maximizada restrigido aos valores definido na hipótese nula, H_0 e dos valores do logaritmo da função de verossimilhança maximizada sem restrição. De uma maneira mais matemática, temos a representação das hipóteses em teste e a expressão da estatística do teste, respectivamente:

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta \neq \theta_0, \quad (2.34)$$

$$TRV = -2 \log \left[\frac{L(\theta_0)}{L(\theta)} \right]. \quad (2.35)$$

TRV segue aproximadamente uma distribuição qui-quadrado com p graus de liberdade, χ_p^2 . H_0 é rejeitada, ao nível de $100\alpha\%$ de significância, se $TRV > \chi_{p,1-\alpha}^2$.

2.1.10 Adequação do Modelo Ajustado

É importante avaliar se o modelo proposto está bem ajustado aos dados. Isso pode ser feito por meio de técnicas gráficas. Essas técnicas avaliam a distribuição dos erros e, sua principal utilidade é rejeitar modelos inapropriados, ou seja, o interesse não é aprovar um modelo particular, até porque, em muitos casos, existem mais de um modelo que podem ser utilizados.

Em análise de sobrevivência, os resíduos de Cox-Snell, martingal e deviance são os mais utilizados.

Resíduos de Cox-Snell

Segundo Colosimo e Giolo (2006), os resíduos de Cox-Snell (1968), são úteis para examinar o ajuste global do modelo. Esses resíduos são determinados por:

$$\hat{e}_i = \hat{H}(t_i), \quad (2.36)$$

no qual, $\hat{H}(t_i)$ é a função de risco acumulado obtido do modelo ajustado.

Os resíduos \hat{e}_i vêm de uma população homogênea e, se o modelo for adequado, devem seguir uma distribuição exponencial padrão (Lawless,2003). Para que o modelo exponencial seja adequado, o gráfico de \hat{e}_i versus $\hat{H}(\hat{e}_i)$ deve ser aproximadamente uma reta. O gráfico das curvas de sobrevivência dos resíduos, $\hat{S}(\hat{e}_i)$, e pelo modelo exponencial padrão, $\exp(-\hat{e}_i)$, também auxiliam na verificação da qualidade do modelo ajustado, quanto mais próximas, o ajuste do modelo aos dados é melhor.

Resíduos de Martingal

Segundo Colosimo e Giolo (2006), os resíduos *martingal* são úteis para determinar a forma funcional (linear, quadrática, etc.). Para os modelos de regressão paramétricos, os resíduos *martingal* são definidos por:

$$\hat{m}_i = \delta_i - \hat{e}_i, \quad (2.37)$$

no qual δ_i é a variável indicadora de falha e \hat{e}_i são os resíduos de Cox-Snell.

Os resíduos de *martingal*, inicialmente, foram introduzidos nos processos de contagem, posteriormente, foram reescritos para modelos de regressão paramétricos.

Resíduos Deviance

Os resíduos deviance auxiliam a examinar a acurácia do modelo para cada indivíduo sob estudo. Os resíduos deviance, para os modelos de regressão paramétrico, são definidos por:

$$\hat{d}_i = \text{sign}(\hat{m}_i)[-2(\hat{m}_i + \delta_i \log(\delta_i - \hat{m}_i))]^{1/2}. \quad (2.38)$$

Esses resíduos são uma tentativa de tornar os resíduos *martingal* mais simétricos em torno de zero, pois facilitam, em geral, a detecção de pontos atípicos. Se o modelo for apropriado, esses resíduos devem apresentar um comportamento aleatório em torno de zero (Colosimo e Giolo, 2006).

2.2 Modelo de Regressão de Cox

O modelo de regressão de Cox (Cox, 1972), proposto por Cox em 1972, possibilita que as covariáveis de interesse sejam consideradas na análise dos tempos de falha.

O modelo de Cox é formulado pela seguinte função de risco:

$$h_i(t|x_i) = h_0(t) \exp(\beta'x_i), \quad (2.39)$$

no qual, h_0 é conhecida como a função de risco base, ou seja, é a função de risco de um indivíduo com vetor de covariáveis nulo (iguais a zero) $x=0$, β é o vetor de dimensão p dos coeficientes de regressão desconhecidos e x_i é o vetor de dimensão p de covariáveis observadas para o indivíduo i , $i = 1, \dots, n$.

O modelo de Cox é chamado semi-paramétrico, pois é composto por h_0 , que não assume nenhuma forma paramétrica, e pela parte paramétrica que são as covariáveis que atuam de forma multiplicativa por meio de uma função de ligação $g(x, \beta) = \exp(\beta x_i)$.

A razão das funções de risco para dois indivíduos, i e j , é dada por:

$$\frac{h(t|x_i)}{h(t|x_j)} = \frac{h_0(t) \exp(\beta'x_i)}{h_0(t) \exp(\beta'x_j)} = \frac{\exp(\beta'x_i)}{\exp(\beta'x_j)}, \quad (2.40)$$

que não depende do tempo t . Portanto, suposição básica para o uso do modelo de Cox é que as taxas de falha sejam proporcionais.

A função de sobrevivência para esse modelo, considerando um indivíduo com vetor de covariável x é dada por:

$$S(t|x) = (S_0(t))^{\exp(\beta'x)}, \quad (2.41)$$

no qual, $S_0(t) = \exp[-\int_0^t h_0(u)du]$ é a função de sobrevivência base e β é o vetor de covariáveis do modelo de Cox.

2.2.1 Estimação dos parâmetros no modelo de Cox

A presença do componente não-paramétrico no modelo de Cox torna inviável o uso do método de máxima verossimilhança para estimar o vetor de coeficientes β . Então, Cox(1975) propôs a função de verossimilhança parcial, $L(\beta)$, que é dada pela seguinte expressão:

$$L(\beta) = \prod_1^n \left[\frac{\exp(\beta' X_i)}{\sum_{j \in R_i} \exp(\beta' X_j)} \right]^{\delta_i}, \quad (2.42)$$

no qual R_i é o conjunto de todos os indivíduos ainda em risco no tempo t_i .

Os indivíduos censurados entram na função de verossimilhança parcial $L(\beta)$ por meio do conjunto de risco R_i , contribuindo somente enquanto permanecem em risco. Em geral, o vetor de parâmetros β é estimado maximizando o logaritmo da função $L(\beta)$. Em seguida deriva-se em relação a cada parâmetro do vetor e iguala o resultado a zero. Isso é dado pela seguinte expressão matemática:

$$\frac{\partial l(\beta)}{\partial \beta_k} = \sum_1^n \delta_i \left[x_{ik} - \frac{\sum_{j \in R_i} x_{jk} \exp(\beta' X_j)}{\sum_{j \in R_i} \exp(\beta' X_j)} \right] = 0, k = 1, \dots, p, \quad (2.43)$$

no qual $l(\beta) = \log(L(\beta))$ e x_{jk} são os indivíduos que permanecem em risco, ou seja que ainda não falharam ou não foram censurados.

2.2.2 Estimando funções relacionadas a $h_0(t)$

Uma estimativa para $H_0(t)$, não paramétrica, proposta por Breslow (1972), é uma função escada com saltos nos diferentes tempos de falha e, é expresso por:

$$\hat{H}_0(t) = \sum_{j:t_j < t} \frac{d_j}{\sum_{l \in R_j} \exp(x_l' \hat{\beta})}, \quad (2.44)$$

no qual d_j é o número de falhas em t_j .

A partir da equação (2.44), podem ser estimadas as funções de sobrevivência $S_0(\hat{t})$ e $\hat{S}(t)$, dadas respectivamente por:

$$\hat{S}_0(t) = \exp(-\hat{H}_0(t)), \quad (2.45)$$

e

$$\hat{S}(t|x) = [\hat{S}_0(t)]^{\exp(x' \hat{\beta})}. \quad (2.46)$$

Na ausência de covariáveis, $\hat{H}_0(t)$ se reduz a :

$$\hat{H}_0(t) = \sum_{j:t_j < t} \frac{d_j}{n_j}, \quad (2.47)$$

que é chamado de Estimador Nelson-Aalen. Por isso, o estimador proposto por Breslow(1972) na equação (2.44) também é conhecido como estimador de Nelson-Aalen-Breslow.

2.2.3 Adequação do modelo de Cox

O modelo de regressão de Cox, apesar de ser bastante flexível, não se ajusta a qualquer situação. Para investigar a adequação do modelo aos dados é necessário verificar a suposição de riscos proporcionais, fazer uma avaliação geral do ajuste do modelo de Cox e analisar outros aspectos do modelo de Cox.

Avaliação da suposição de riscos proporcionais

Existem algumas técnicas gráficas e testes estatísticos para avaliar a suposição de riscos proporcionais no modelo de Cox. Algumas dessas técnicas serão descritas a seguir:

- **Método gráfico descritivo**

Essa técnica consiste em dividir os dados em m estratos, usualmente de acordo com alguma covariável. Em seguida, deve-se estimar $\hat{H}_0(t)$ usando a equação (2.44) para cada estrato da covariável qualitativa. Se a suposição de riscos proporcionais for pertinente, as curvas do logaritmo de $\hat{H}_0(t)$ versus t , ou $\log(t)$, apresentarão diferenças aproximadamente constantes ao longo do tempo.

• **Método com coeficiente dependente do tempo (Resíduos de Schoenfeld (1982))**

Segundo Colosimo e Giolo (2006), para definir os resíduos de Schoenfeld no modelo de Cox, considere que se o i -ésimo indivíduo com o vetor de covariáveis $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})'$ é observado falhar, tem-se para esse indivíduo um vetor de resíduos de Schoenfeld $\mathbf{r}_{iq} = (r_{i1}, r_{i2}, \dots, r_{ip})$, em que cada componente r_{iq} , para $q = 1, 2, \dots, p$, é definido por:

$$r_{iq} = x_{iq} - \frac{\sum_{j \in R(t_i)} x_{jq} \exp(x'_j \hat{\beta})}{\sum_{j \in R(t_i)} \exp(x'_j \hat{\beta})}, \quad (2.48)$$

Uma forma padronizada dos resíduos de Schoenfeld geralmente é usada, pois possibilita que a estrutura de correlação dos resíduos seja considerada. Sua expressão matemática é dada por:

$$\mathbf{s}_{iq}^* = [I(\hat{\beta})]^{-1} r_{iq}, \quad (2.49)$$

no qual, $[I(\hat{\beta})]^{-1}$ é a matriz de informação observada.

A suposição de riscos proporcionais avaliada pelos resíduos de Schoenfeld é baseada em um resultado em Grambsch e Therneau (1994), expresso por:

$$h(t) = h_0(t) \exp(x' \beta(t)), \quad (2.50)$$

Tem-se a restrição de que $\beta(t) = \beta$, como uma forma alternativa de representar o modelo de Cox. A restrição $\beta(t) = \beta$ resulta na proporcionalidade dos riscos. Se $\beta(t)$ não é constante, o impacto de uma ou mais covariáveis no risco pode variar com o tempo. Assim, o gráfico de $\beta_q(t)$ versus t deve ser uma linha horizontal. Grambsch e Therneau (1994), sugerem o gráfico de $\mathbf{s}_{iq}^* + \hat{\beta}_q$ versus t ou $g(t)$. Inclinação zero mostra evidências a favor da suposição de riscos proporcionais (Colosimo e Giolo, 2006).

• **Teste de hipótese e coeficiente de correlação**

O coeficiente de correlação de Pearson (ρ), entre os resíduos padronizados e $g(t)$, para cada covariável, é uma medida que permite avaliar a suposição de riscos proporcionais. Valores de ρ próximos de zero mostram não haver evidências para rejeitar a suposição de riscos proporcionais.

Um teste para a hipótese global de proporcionalidade dos riscos sobre todas as covariáveis no modelo, levando em consideração $g_q(t) = g_t$, tem como estatística do teste:

$$T = \frac{(g - \bar{g})' S^* I S^* (g - \bar{g})}{d \sum_k (g_k - \bar{g})^2}, \quad (2.51)$$

no qual, I é a matriz de informação observada, d é o número de falhas e $S^* = dR\Gamma^{-1}$, sendo R a matriz dXp dos resíduos de Schoenfeld não padronizados.

As hipóteses testadas são:

H_0 : Os riscos são proporcionais. H_1 : Os riscos não são proporcionais.

T tem distribuição qui-quadrado com p graus de liberdade. Valores de $T > \chi_{p,1-\alpha}^2$ mostram que há evidências estatística para rejeitar a suposição de riscos proporcionais.

Avaliação geral do ajuste do modelo de Cox

Os resíduos de Cox-Snell (1968), martingal e deviance, descritos na Seção 2.1.11, também são utilizados para avaliar a qualidade geral de ajuste do modelo de Cox.

Capítulo 3

Resultados e Discussões

Os dados utilizados neste trabalho foram obtidos em um estudo realizado no Instituto de Psiquiatria (IPq) do Hospital das Clínicas (HC) da Universidade de São Paulo (USP), no período de 1 de dezembro de 1997 a 31 de dezembro de 2004, que teve como intento avaliar o tempo (em dias) até a re-hospitalização de pacientes com esquizofrenia tratados por diferentes tipos de antipsicóticos. Resultados desse estudo, sem a consideração de covariáveis na análise, foram apresentados em Werneck et al (2011).

Inicialmente, faremos uma análise exploratória do conjunto de dados que será utilizado na modelagem do tempo até a re-hospitalização de pacientes com esquizofrenia.

Tabela 3.1: Descrição das covariáveis do estudo

Idade (X_1)	$X_1 = 0$	< 50 anos
	$X_1 = 1$	≥ 50 anos
Sexo (X_2)	$X_2 = 0$	masculino
	$X_2 = 1$	feminino
Idade no início da doença (X_3)	$X_3 = 0$	< 20 anos
	$X_3 = 1$	≥ 20 anos
Tempo de duração da doença (X_4)	$X_4 = 0$	< 20 anos
	$X_4 = 1$	≥ 20 anos
Número de internações (X_5)	$X_5 = 0$	< 10
	$X_5 = 1$	≥ 10
Quantidade de medicamentos (X_6)	$X_6 = 0$	< 3
	$X_6 = 1$	≥ 3
Tempo de internação (X_7)	$X_7 = 0$	< 80 dias
	$X_7 = 1$	≥ 80 dias
Grupo de medicação (X_8)	$X_8 = 0$	Clozapina
	$X_8 = 1$	PGA*
	$X_8 = 2$	SGA**

* Primeiro Grupo de Antipsicóticos

**Segundo Grupo de Antipsicóticos

Tabela 3.2: Frequências absolutas e relativas das covariáveis

Variável	Frequência Absoluta	Frequência Relativa
Idade		
< 50 anos	167	87,9%
≥ 50 anos	23	12,1%
Sexo		
masculino	105	44,74%
feminino	85	55,26%
Idade no início da doença		
< 20 anos	92	48,42 %
≥ 20 anos	98	51,58%
Tempo de duração da doença		
< 20 anos	131	68,95%
≥ 20 anos	59	31,05%
Número de internações		
< 10	177	93,16%
≥ 10	13	6,84%
Quantidade de medicamentos		
< 3	139	73,16%
≥ 3	51	26,84%
Tempo de internação (x_7)		
< 80 dias	162	14,74%
≥ 80 dias	28	85,26%
Grupo de medicação		
Clozapina	42	22,1%
PGA*	58	30,53%
SGA**	90	47,37%

* Primeiro Grupo de Antipsicóticos

**Segundo Grupo de Antipsicóticos

A Tabela 3.1 e a Tabela 3.2 descrevem as covariáveis do nosso estudo e as frequência absolutas e relativas de cada uma delas, respectivamente. A análise descritiva possibilita conhecer melhor os dados. Porém, essa não será feita de maneira tradicional, pois os dados observados possuem muita censura, cerca de 81,6%. O banco de dados é composto por 190 observações cujas 155 são censuras e 35 são falhas.

A função de sobrevivência dos dados, que é a probabilidade acumulada de permanecer em alta, ou seja, sem retornar ao hospital, se comporta como na Figura 3.1.

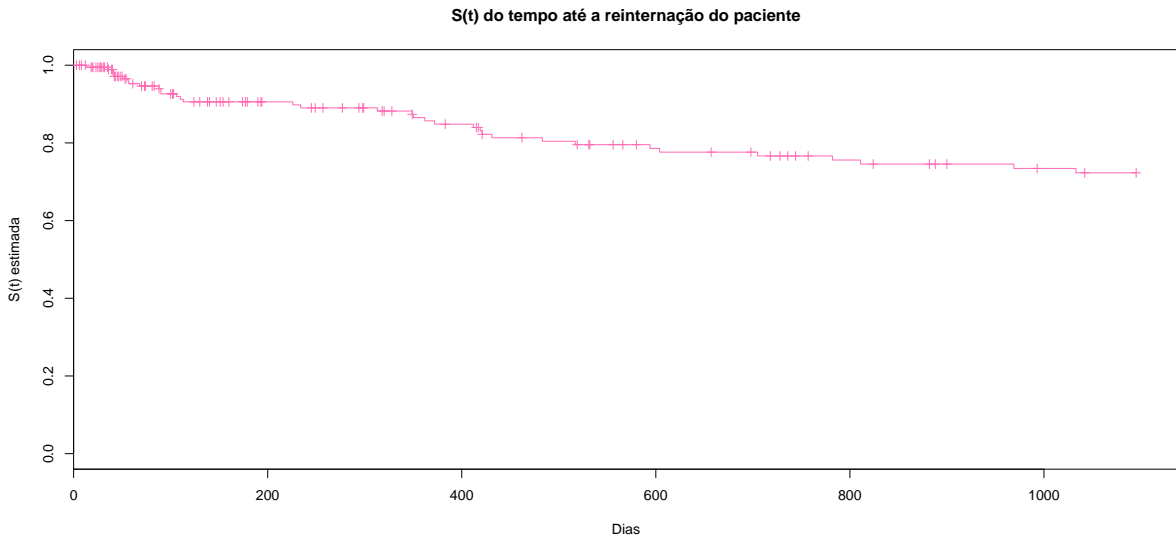


Figura 3.1: Função de Sobrevivência do tempo até a reinternação do paciente

O gráfico da Figura 3.1 mostra que a $S(t)$ do tempo até a reinternação do paciente está decrescendo. O tempo de duração do estudo não foi suficiente para que a função de sobrevivência chegasse a zero. Contudo, a curva de $\hat{S}(t)$ não fica estável em nenhum momento, ela está sempre decrescendo, ainda que lentamente. Se o estudo tivesse um tempo de duração maior, cerca de dois mil dias a mais, provavelmente seria observado que à medida que o tempo tende ao infinito, a função de sobrevivência tende a zero. Dessa forma, não há evidências estatística de que há fração de cura no banco de dados em estudo.

A principal covariável estudada neste trabalho é o Grupo de medicação. Portanto, é interessante fazer uma análise exploratória. Será observado como a função de sobrevivência dos dados se comporta em relação a essa covariável. O gráfico ajudará a mostrar se as curvas são diferentes. Esse gráfico é apresentado na Figura 3.2. Além disso, será verificado se realmente existe diferença ou não entre os grupos por meio dos testes de log-rank e wilcoxon. As hipóteses que serão testadas são as seguintes:

$$H_0: \text{Não há diferença entre os grupos} \quad \text{vs} \quad H_1: \text{Há diferença entre os grupos.}$$

O teste verifica se há evidências ou não para rejeitar a hipótese nula, H_0 . Para um nível de significância de 10%, rejeitamos a hipótese H_0 se o p-valor for menor que o nível de significância, caso contrário, não existe evidências pra rejeitar a hipótese nula. Os resultados serão apresentados na Tabela 3.3.

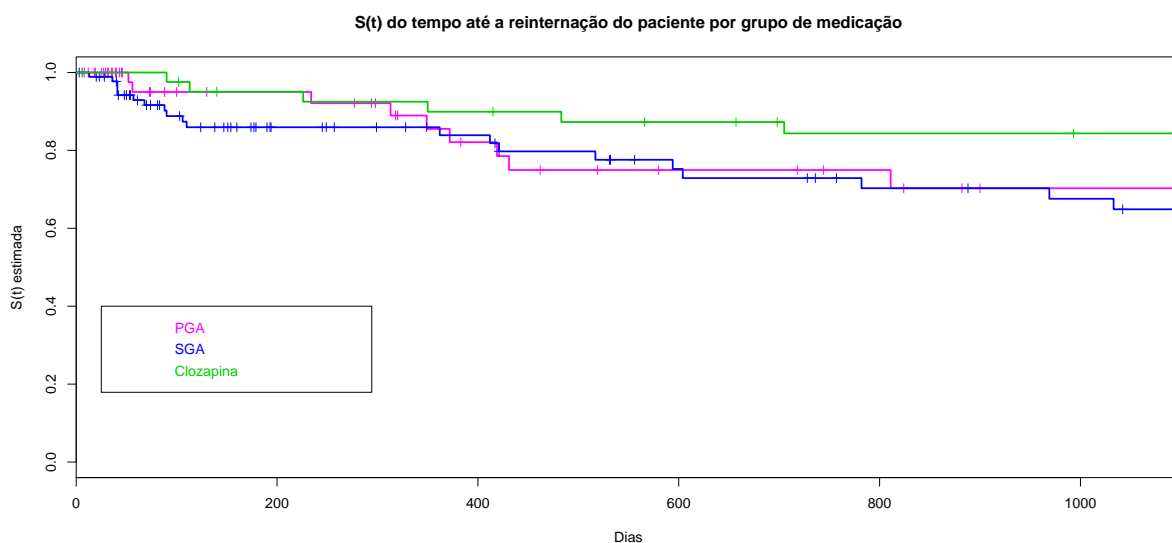


Figura 3.2: Função de Sobrevivência do tempo até a reinternação do paciente por grupo de medicação

Tabela 3.3: Testes log-rank e de Wilcoxon utilizados para testar a igualdade das curvas de sobrevivência obtidas para as covariáveis consideradas neste estudo

Variáveis	p-valor (<i>log-rank</i>)	p-valor (Willcoxon)
PGA vs SGA	0,565	0,585
PGA vs Clozapina	0,191	0,181
SGA vs Clozapina	0,0479	0,0462

Nota-se, no gráfico da Figura 3.2, que aparentemente o grupo SGA tem uma reação diferente da reação do grupo controle Clozapina. Porém, as curvas da função de sobrevivência dos grupos PGA e Clozapina tem um comportamento parecido. E, segundo Werneck et al (2011), os resultados demonstraram que o risco de re-hospitalização para o grupo PGA foi similar para o grupo SGA. Além disso, segundo o teste de *log-rank* e Wilcoxon, só existe diferença entre o segundo grupo de antipsicóticos e a clozapina. Assim, a partir de agora, serão considerados apenas as observações dos dados relativos ao segundo grupo de antipsicóticos (SGA) e à Clozapina, serão apresentadas as covariáveis e suas frequências absolutas e relativas na Tabela 3.4 e Tabela 3.5, respectivamente.

Tabela 3.4: Descrição das covariáveis do estudo com observações referentes ao segundo grupo de antipsicóticos e à Clozapina.

Idade (X_1)	
$X_1 = 0$	< 50 anos
$X_1 = 1$	\geq 50 anos
Sexo (X_2)	
$X_2 = 0$	masculino
$X_2 = 1$	feminino
Idade no início da doença (X_3)	
$X_3 = 0$	< 20 anos
$X_3 = 1$	\geq 20 anos
Tempo de duração da doença (X_4)	
$X_4 = 0$	< 20 anos
$X_4 = 1$	\geq 20 anos
Número de internações (X_5)	
$X_5 = 0$	< 10
$X_5 = 1$	\geq 10
Quantidade de medicamentos (X_6)	
$X_6 = 0$	< 3
$X_6 = 1$	\geq 3
Tempo de internação (X_7)	
$x_7 = 0$	< 80 dias
$x_7 = 1$	\geq 80 dias
Grupo de medicação (X_8)	
$X_8 = 0$	Clozapina
$X_8 = 1$	Segundo Grupo de Antipsicóticos (SGA)

Tabela 3.5: Frequências absolutas e relativas das covariáveis com observações referentes ao segundo grupo de antipsicóticos e à Clozapina.

Variável	Frequência Absoluta	Frequência Relativa
Idade		
< 50 anos	118	89,4%
≥ 50 anos	14	10,6%
Sexo		
masculino	81	61,36%
feminino	51	38,64%
Idade no início da doença		
< 20 anos	66	50,0 %
≥ 20 anos	66	50,0%
Tempo de duração da doença		
< 20 anos	94	71,21%
≥ 20 anos	38	28,79%
Número de internações		
< 10	124	93,94%
≥ 10	8	6,06%
Quantidade de medicamentos		
< 3	88	66,67%
≥ 3	44	33,33%
Tempo de internação (x_7)		
< 80 dias	20	15,15%
≥ 80 dias	112	84,85%
Grupo de medicação		
Clozapina	42	31,82%
SGA	90	68,18%

A função de sobrevivência para o tempo até a reinternação do paciente com relação à todas as covariáveis consideradas neste estudo foi obtida por meio do estimador Kaplan-Meier. Os gráficos são apresentados na Figura 3.3 e permitem avaliar se existe diferença, ou não, entre os grupos principais e os grupos controle. A partir dessa análise será atestado se é vantajoso manter a separação dos grupos em cada covariável.

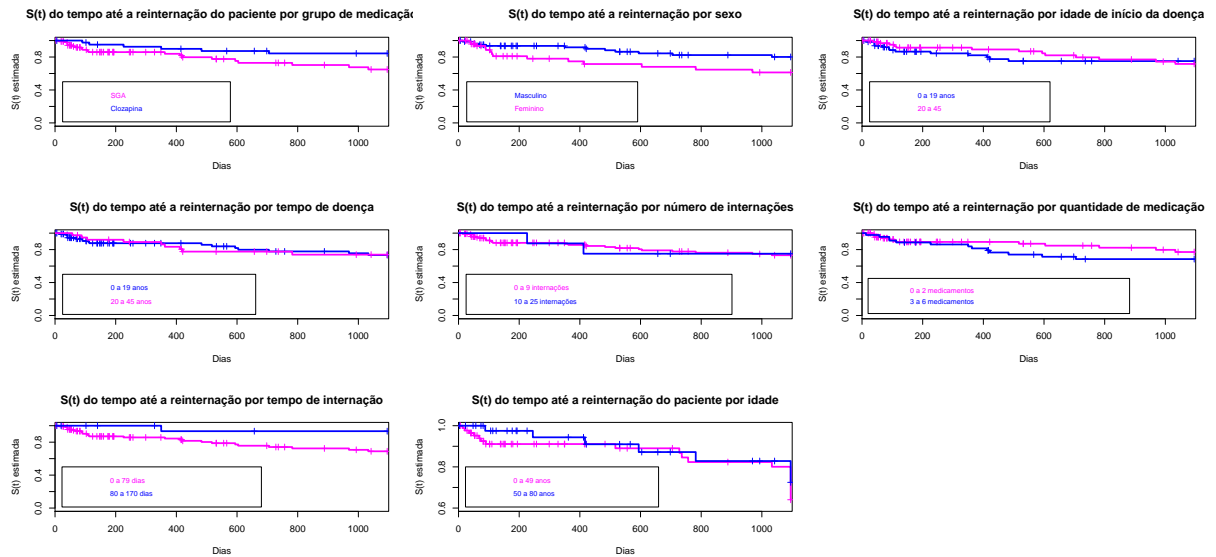


Figura 3.3: Gráficos Kaplan-Meier para as covariáveis medidas no estudo do tempo até a re-hospitalização de pacientes com esquizofrenia

A Figura 3.3 mostra que o grupo SGA tem uma reação diferente do grupo controle Clozapina, como já visto em tela. Os sexos masculino e feminino, aparentemente, também reagem diferente aos medicamentos. Observa-se também que diferentes quantidades de medicamentos e o tempo de internação provocam uma reação diferente nos pacientes. As outras covariáveis parecem não apresentar diferença em relação aos grupos controle.

A verificação, se realmente existe diferença ou não entre os grupos das covariáveis, foi realizado utilizando o teste de log-rank e wilcoxon. As hipóteses que serão testadas são as seguintes:

$$H_0: \text{Não há diferença entre os grupos} \quad \text{vs} \quad H_1: \text{Há diferença entre os grupos.}$$

O teste verifica se há evidências ou não para rejeitar a hipótese nula, H_0 . Para um nível de significância de 10%, rejeitamos a hipótese H_0 se o p-valor for menor que o nível de significância, caso contrário, não existe evidências pra rejeitar a hipótese nula.

Tabela 3.6: Testes log-rank e de Wilcoxon utilizados para testar a igualdade das curvas de sobrevivência obtidas para as covariáveis consideradas neste estudo

Variáveis	valor-p (<i>log-rank</i>)	valor-p (Willcoxon)
Grupo de antipsicóticos (SGA vs Clozapina)	0,0479	0,0462
Idade	0,405	0,418
Sexo	0,0308	0,0309
Idade no início da doença	0,658	0,804
Tempo de duração da doença	0,985	0,987
Número de internações	0,979	0,991
Número de medicamentos ingeridos	0,301	0,304
Tempo de internação em dias	0,0578	0,0558

Com base nos resultados apresentados na Tabela 3.6, conclui-se que a covariável grupo de medicação apresenta diferença entre o grupo SGA e o grupo de controle Clozapina. A covariável sexo também apresenta diferença entre as curvas, ou seja, os medicamentos respondem de maneira diferente para os sexos feminino e masculino. Além disso, os resultados para quem fica internado menos de 80 dias é diferente para quem fica internado 80 dias ou mais. As covariáveis idade, idade no início da doença, tempo de duração da doença, número de internações, número de medicamentos ingeridos não apresentam diferença entre seus grupos.

Ainda explorando os gráficos da Figura 3.3, podemos supor proporcionalidade entre os grupos das covariáveis, apesar do leve cruzamento entre os grupos de algumas covariáveis.

3.1 Ajuste do modelo

3.1.1 Ajuste do modelo paramétrico

A primeira etapa no ajuste do modelo do tempo até a re-hospitalização dos pacientes com esquizofrenia é buscar modelar os dados com base em algum modelo paramétrico, desconsiderando as covariáveis.

Visto que existem várias formas que podem ser assumidas pelo gráfico da função de risco da variável Tempo até a re-hospitalização, foi plotado o gráfico do tempo total em teste (curva TTT) para os tempo, como pode ser observado na Figura 3.4.

A curva não assume uma forma explícita. O gráfico indica que o modelo que se ajusta bem aos dados pode ter função de risco em forma de U, porém a curva está próxima de uma função crescente ou contante. Com base nesses fatos, os modelos Weibull e exponencial foram testados. Além disso pode-se perceber que os modelos log-normal e log-logística têm um bom ajuste.

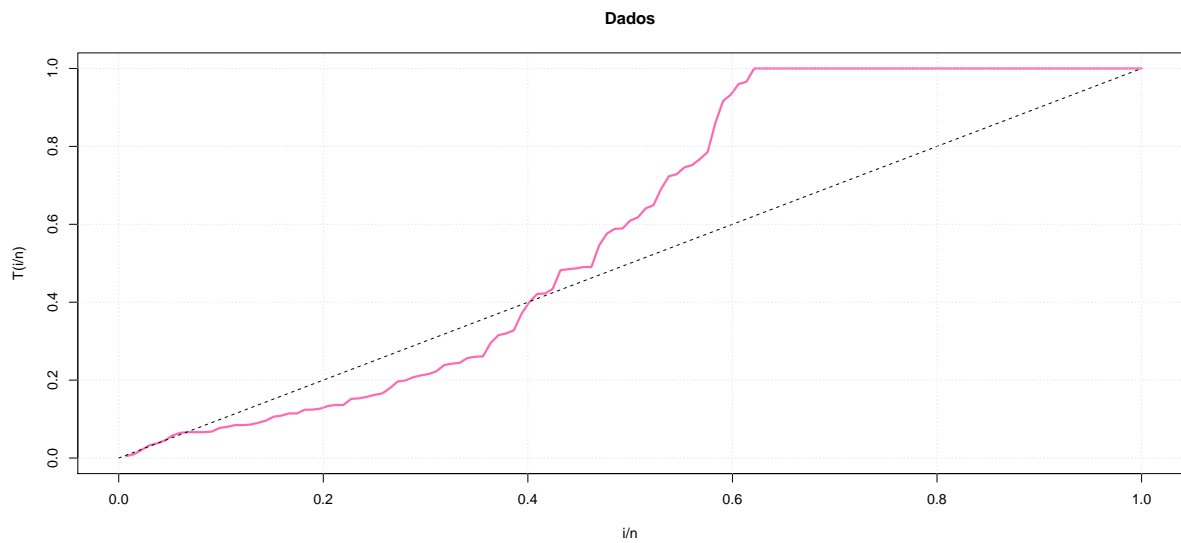


Figura 3.4: Curvas TTT dos dados do tempo até a re-hospitalização de pacientes com esquizofrenia

Para comparar os quatro modelos propostos e escolher o mais adequado foi construído o gráfico com a função de sobrevivência estimada por Kaplan-Meier e a função de sobrevivência dos modelos Weibull, Exponencial, Log-Normal e Log-Logística, o resultado encontra-se apresentado na Figura 3.5 .

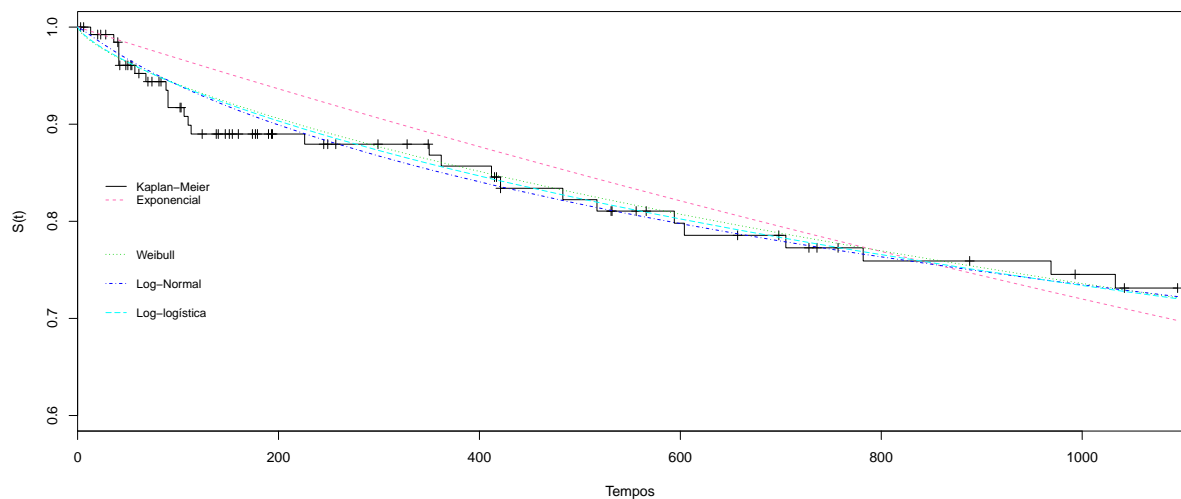


Figura 3.5: Função de Sobrevivência estimada por Kaplan-Meier e pelos modelos paramétricos Weibull, Exponencial, Log-Normal e Log-Logística

É possível observar que os modelos log-normal e log-logístico estão mais próximos da curva de Kaplan-Meier, portanto, se ajustam melhor aos dados. No entanto, a diferença observada entre as curvas mostra-se pequena.

A análise gráfica, contudo, é uma técnica subjetiva. Assim, na tentativa de confirmar os resultados obtidos na análise em tela, foram utilizadas as técnicas de seleção de variáveis AIC, AICc e BIC. Os resultados são apresentados na Tabela 3.7.

Tabela 3.7: Critérios de seleção de variáveis AIC, AICc e BIC paos modelos Exponencial, de Weibull, Log-Normal e Log-Logística

Modelo	Parâmetro	AIC	AICc	BIC
Exponencial	$\alpha = 3043,5$	471,0797	471,1105	473,9625
Weibull	$\alpha = 5377,75$ $\gamma = 0,703$	468,2877	468,3808	474,0533
Log-Normal	$\mu = 8,46$ $\sigma = 2,47$	465,8023	465,8953	471,5679
Log-Logística	$\alpha = 3816,14$ $\gamma = 0,75$	467,7195	467,8126	473,4851

Considerando os três critérios de seleção o modelo que apresentou menor valor em todos os casos foi o modelo log-normal.

Análise de Resíduos

Os resíduos de Cox-Snell serão apresentados com finalidade de avaliar o ajuste do modelo proposto.

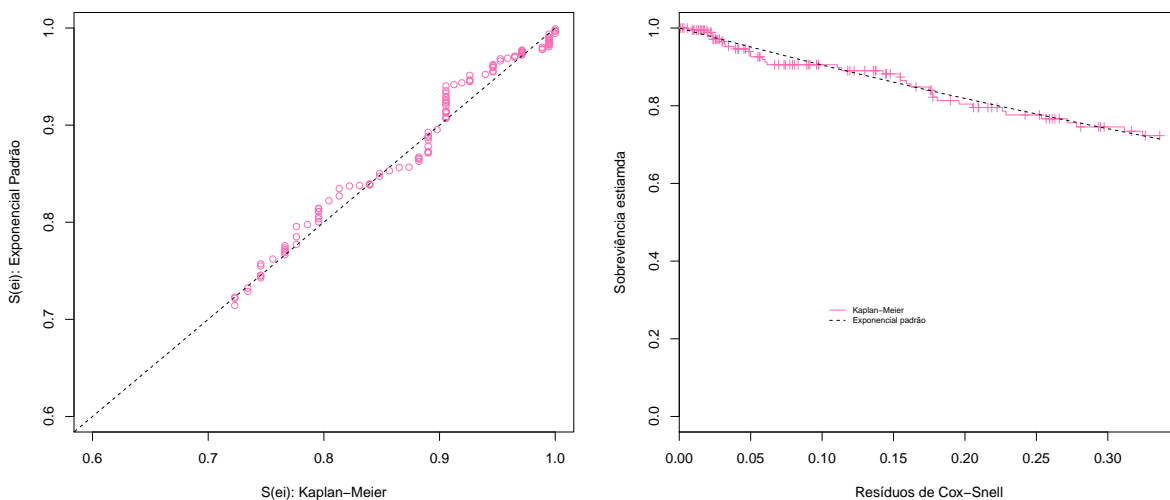


Figura 3.6: $\hat{S}(t)$ dos resíduos de Cox-Snell estimada pelo método de Kaplan-Meier e pelo modelo exponencial padrão e curva de $\hat{S}(t)$ estimada

Nesta etapa, será analisado se é possível que o modelo log-normal não seja adequado aos dados, pois a análise de resíduos apenas mostra se o modelo claramente não

é adequado aos dados, pois pode ser que exista mais de um modelo que se ajuste bem aos dados.

Os pontos, no gráfico $S(e_i)$ da Figura 3.6, estão próximos de uma reta diagonal. E, com relação ao gráfico de resíduos de Cox-Snell, a curva da distribuição exponencial padrão e a de Kaplan-Maier apresentam estar bastante próximas. A partir dos gráficos apresentados na Figura 3.5 e da análise dos resíduos, observa-se que o modelo log-normal encontra-se bem ajustado.

Interpretação do modelo

Uma estimativa dos pacientes não voltarem a ser rehospitalizados num determinado tempo, com base no modelo log-normal, é dado por:

$$\hat{S}(t) = \phi\left[\frac{-\log(t) + \mu}{\sigma}\right], \quad (3.1)$$

no qual, $\phi(\cdot)$ é a função de distribuição acumulada de uma normal padrão.

Um exemplo dessa situação é a estimativa dos pacientes não voltarem a ser rehospitalizados antes de 782 dias, que é dada por:

$$\hat{S}(t) = \phi\left[\frac{-\log(782) + 8,46}{2,47}\right] = 0,7667. \quad (3.2)$$

Ou seja, 76,67% dos pacientes voltam a ser rehospitalizados depois de 782 dias. A mesma estimativa obtida pelo estimador de Kaplan-Meier fornece o valor de 75,9%. Dessa forma, um paciente com esquizofrenia apresenta uma probabilidade de cerca de 77% de não voltar a ser reinternado em menos de 782 dias após ter recebido alta.

A diferença entre o AIC dos modelos propostos inicialmente não foi muito discrepante. Como foi visto em todos os passos para seleção do modelo paramétrico, não está claro qual modelo melhor se ajustaria aos dados. Visto que houve suposição de riscos proporcionais em todas as covariáveis, podemos utilizar o modelo de Cox no estudo do tempo até a reinternação dos pacientes com esquizofrenia. Apesar dos modelos paramétricos serem mais precisos, o modelo de Cox é mais flexível por possuir uma parte não-paramétrica.

3.1.2 Ajuste do modelo de Cox

O modelo de regressão de Cox permite a análise de dados provenientes de estudos de tempo de vida em que a resposta é o tempo até a ocorrência de um evento de interesse, ajustando por covariáveis (Colosimo e Giolo, 2006).

O modelo proposto por Cox modela diretamente a função de risco. Inicialmente, na construção do modelo de Cox foi feita a seleção das covariáveis. No decorrer do processo foram usados os códigos identificadores das covariáveis da Tabela 2.4. Em cada passo do processo de seleção das covariáveis foi verificado o nível de significância das covariáveis separadamente ou em conjunto. As hipóteses testadas são:

H_0 : O modelo de interesse é adequado. vs H_1 : O modelo de interesse não é adequado.

O primeiro passo foi ajustar o modelos com apenas uma covariável, e, separadamente, as únicas que foram significativas, ou seja, com valor-p menor que 0,1, foram X_2 (sexo) com valor-p 0,036, X_7 (tempo de internação) com valor-p 0,0902 e X_8 (grupo de medicação) com valor-p 0,0537. Todas as outras covariáveis tiveram valor-p acima de 0,1.

Foi construído, no segundo passo, o modelo com todas as covariáveis, e como o modelo não ficou bem ajustado, foram retiradas as covariáveis com o maior valor-p, uma de cada vez, e assim, o ajuste do modelo foi verificado novamente.

Inicialmente, retiramos a covariável X_5 (número de internações), no modelo com todas as covariáveis, seu valor-p foi 0,9952. Em seguida, retiramos a covariável X_7 (tempo de internação), que em conjunto com as outras covariáveis, deixou de ser significativa, seu valor-p no modelo, com as outras covariáveis que permaneceram, foi de 0,4029. Depois, retiramos a covariável X_3 (idade no início da doença), com o valor-p igual a 0,32201, mantendo as outras covariáveis que não foram eliminadas. Posteriormente, a covariável X_1 (idade), com valor-p igual a 0,25482, foi excluída, e foram mantidas as outras covariáveis restantes. Logo a seguir, eliminamos a covariável X_4 (tempo de duração da doença), seu valor-p foi de 0,24624.

O modelo de Cox, dessa forma, foi ajustado com as variáveis X_2 (sexo), X_6 (quantidade de medicamento) e X_8 (grupo de medicação).

As covariáveis selecionadas para o modelo e o modelo global, ademais, precisam ter a suposição de riscos proporcionais satisfeita. A seguir, são apresentados na Tabela 3.8 os resultados do coeficiente de correlação ρ ; a estatística do teste χ^2 ; e o valor-p. Além disso, serão mostrados os gráficos dos resíduos padronizados de Schoenfeld nas Figuras 3.7, 3.8 e 3.9. As hipóteses que serão testadas são:

H_0 : As funções de riscos são proporcionais.

H_1 : As funções de riscos não são proporcionais.

Tabela 3.8: Teste de proporcionalidade para as funções de risco do modelo final de Cox

Covariável	rho (ρ)	χ^2	valor-p
Sexo	-0,0150	0,00575	0.940
Quantidade de medicamentos	-0,0210	0,00923	0,923
Grupo de medicação	-0.0233	0,01204	0,913
Global	-	0.02214	0.999

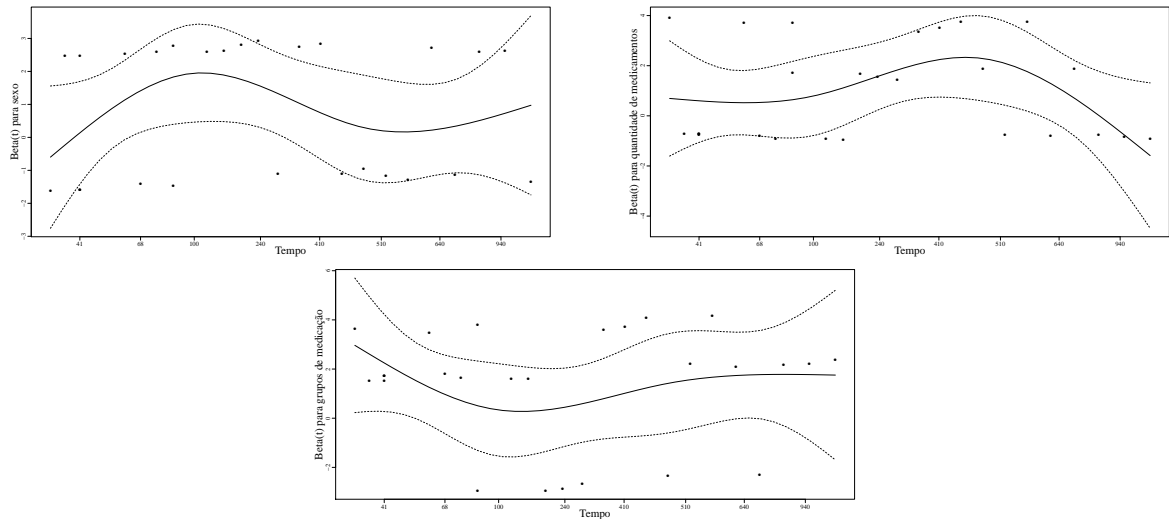


Figura 3.7: Resíduos padronizados de Schoenfeld para as covariáveis do modelo de Cox final

Observando-se os gráficos apresentados na Figura 3.7, não são evidentes tendências ao longo do tempo. Isso permite afirmar que as covariáveis em questão possuem riscos proporcionais. Ainda assim, é necessário analisar o teste de proporcionalidade para as funções de risco do modelo final de Cox, pois a análise gráfica é subjetiva. Os valores dos coeficientes de correlação de Pearson, ρ , são todos próximos de zero e todos os valor-p são superiores a 0.10, tanto para o teste global quanto para os testes para cada covariável separadamente. Portanto, não há evidências estatística para rejeitar a hipótese nula. Assim, pode-se confirmar que as funções de risco das covariáveis sexo, quantidade de medicamentos e grupo de medicação são proporcionais.

Análise de resíduos

Vimos que o melhor ajuste encontrado foi referente ao modelo de Cox que relaciona as covariáveis sexo, quantidade de medicamentos e grupo de medicação, com o tempo de sobrevivência. Dessa forma, o próximo passo é analisar globalmente se o modelo encontrado é adequado para os dados sobre esquizofrenia. Portanto, pode-se fazer uso da análise de resíduos considerando o método de Cox-Snell.

Segundo Lawless (2003), os resíduos de Cox-Snell vêm de uma população homogênea e devem seguir uma distribuição exponencial com média 1. Quanto mais a função de sobrevivência dos resíduos do modelo de Cox se aproxima da função de sobrevivência da exponencial, melhor é o ajuste do modelo de Cox. Segundo Klein e Moeschberger (2003), essa interpretação deve ser feita com cautela, visto que a distribuição exponencial dos resíduos somente é válida quando os verdadeiros valores dos parâmetros são considerados. Desvios da distribuição exponencial podem ocorrer, em algumas situações, devido

à incerteza no processo de estimação dos parâmetros. E, esses desvios podem ser maiores, principalmente, na cauda direita da distribuição e em casos de amostras pequenas.

Foi observado, a partir da Figura 3.8, que a função de sobrevivência dos resíduos se ajustou bem à função de sobrevivência da exponencial padrão, sobretudo para os valores pequenos dos resíduos. Esse resultado é um indicativo do bom ajuste do modelo de Cox, que pode ser confirmado pelo teste de Kolmogorov-Smirnov (CONOVER, 1999). O teste de Kolmogorov-Smirnov para verificação do ajuste da distribuição Exponencial padrão dos resíduos resultou na estatística do teste $D=0,215$. Considerando um nível de significância de 5% e $n=32$ (32 valores não censurados), o valor crítico do teste de Kolmogorov-Smirnov é igual a $D_{crítico}=0,221$, não rejeitando a hipótese do resíduo de Cox-Snell seguir uma distribuição exponencial padrão.

Note que o alto valor $D=0,215$ da diferença entre as funções de sobrevivência observado na Figura 3.8, percebido na cauda direita da distribuição dos resíduos é um fenômeno já previsto por Moeschberger. Visto que as observações (neste caso os resíduos de Cox-Snell) vão experimentando o evento de interesse ou sendo censurados ao longo do tempo, o número de observações na cauda da função de sobrevivência é pequeno. Neste contexto, os desvios presentes na cauda da função de sobrevivência não devem ser tratados da mesma forma que os desvios das funções nos valores iniciais, corroborando ainda mais a suposição de bom ajuste do modelo de riscos proporcionais de Cox para os dados dessa aplicação.

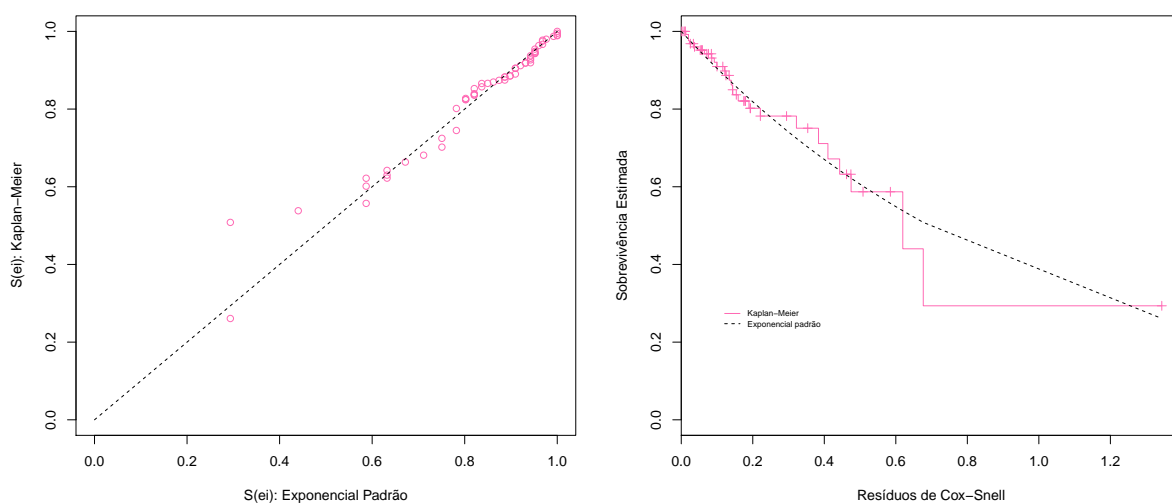


Figura 3.8: Gráficos da função de sobrevivência estimada dos resíduos de Cox-Snell e Exponencial Padrão.

3.2 Interpretando as estimativas dos parâmetros

As estimativas dos parâmetros do modelo de regressão final de Cox encontram-se na Tabela 3.9. Ao nível de 10 %, todas as covariáveis são significativas.

Tabela 3.9: Resultado do ajuste do modelo de Cox para os dados do tempo até a re-hospitalização de pacientes com esquizofrenia e correspondentes razões de riscos (RR).

Covariável	Coefficiente	Erro Padrão	Exponencial (coeficiente)	$IC_{95\%}(RR) = e^{coeficiente}$	valor-p
Sexo	$\beta_{masculino} = 0$				
	$\beta_{feminino} = 0,83$	0,39	2,3	(1,06 ; 4,98)	0,03
Quantidade de medicamento	$\beta_{<3} = 0$				
	$\beta_{\geq 3} = 0,97$	0,42	2,64	(1,16 ; 6,03)	0,02
Grupo de medicação	$\beta_{Clozapina} = 0$				
	$\beta_{SGA} = 1,33$	0,5	3,8	(1,42; 10,14)	0,007

Foi calculado, também, o fator inflacionário da variância (FIV) para cada covariável selecionada. Assim, podemos avaliar se há multicolinearidade entre as covariáveis. Se o valor do FIV das covariáveis for baixo não há multicolinearidade. Os resultados são apresentados na Tabela 3.10.

Tabela 3.10: Valor do Fator Inflacionário da Variância (FIV) para cada covariável selecionada para o modelo.

Covariável	Sexo	Quantidade de medicamento	Grupo de medicação
FIV	1.001669	1.144589	1.146374

Visto que todos os FIVs têm valores próximos a 1, não há indício de multicolinearidade entre as covariáveis.

O modelo final com as covariáveis sexo, quantidade de medicamento e grupo de medicação, é apresentado, em termos da função de risco, na equação a seguir :

$$h(t|x) = h_0(t) * exp(\beta_{sexo} + \beta_{quantidadedemedicamentos} + \beta_{grupodemedicao}). \quad (3.3)$$

O componente não paramétrico, $h_0(t)$, não é especificado e é uma função não-negativa. Os coeficientes da regressão do componente paramétrico medem os efeitos das covariáveis sob a função de risco. Esse efeito é o de acelerar ou desacelerar a função de risco. As seguintes interpretações podem ser obtidas a partir dos resultados da Tabela 3.9.

- O risco de pacientes do sexo feminino voltarem a ser rehospitalizados é estimado

aproximadamente 2,3 vezes o risco dos pacientes do sexo masculino (130 % maior), com todas as outras covariáveis mantidas constantes.

- E, com o aumento de uma unidade na covariável quantidade de medicamentos, o risco de pacientes que tomam de mais de 3 medicamentos voltarem a ser rehospitalizados é estimado aproximadamente 2,64 vezes o risco dos pacientes que tomam menos de 3 medicamentos (164 % maior), com todas as outras covariáveis mantidas fixas.
- Ademais, com o aumento de uma unidade na covariável grupo de medicação, o risco de pacientes que fazem parte do segundo grupo de antipsicóticos voltarem a ser rehospitalizados é estimado aproximadamente 3,8 vezes o risco dos pacientes que fazem parte do grupo Clozapina (280 % maior), fixando todas as outras covariáveis.

É possível, também, calcular a função de sobrevivência de base.

Temos à disposição, também, as estimativas da função de sobrevivência de base, $\hat{S}_0(t)$, apresentadas na Tabela 3.11.

$$\hat{S}_0(t) = \exp[-\hat{H}_0(t)]. \quad (3.4)$$

Tabela 3.11: Estimativas da função de sobrevivência de base $\hat{S}_0(t)$.

t	$\hat{S}_0(t)$	t	$\hat{S}_0(t)$	t	$\hat{S}_0(t)$
1	1	106	0,948	483	0,8475
13	0,9939	110	0,9172	517	0,8360
36	0,9877	113	0,9095	594	0,8241
41	0,9687	226	0,9003	604	0,8120
57	0,9619	350	0,8904	705	0,7995
68	0,9548	362	0,8806	782	0,7863
88	0,9474	412	0,8703	969	0,7728
90	0,9325	421	0,8590	1033	0,7589

A função de sobrevivência, dessa forma, para um indivíduo com vetor de covariáveis $X = (x_2, x_6, x_8)'$ é estimada por:

$$\hat{S}(t) = [\hat{S}_0(t)]^{\exp(X'\hat{\beta})}, \quad (3.5)$$

no qual, $x'\hat{\beta} = \hat{\beta}_{sexo} + \hat{\beta}_{quantidade\ de\ medicamentos} + \hat{\beta}_{grupo\ de\ medicação}$.

A probabilidade, por conseguinte, de um indivíduo do sexo feminino que toma de 3 a 6 medicamentos por dia e faz parte do segundo grupo de antipsicóticos não voltar a ser reinternado em menos de 782 dias após ter recebido alta é:

$$\hat{S}(782) = [\hat{S}_0(782)]^{\exp(x'\hat{\beta})} = 0,7863^{\exp(3,13)} = 0,004, \quad (3.6)$$

no qual $X'\hat{\beta} = 0,83 + 0,97 + 1,33$ e $\hat{S}_0(782)$ é dada pela Tabela 3.11.

A probabilidade, visto isso, de um indivíduo cujo sexo é feminino, toma de 3 a 6 medicamentos por dia e faz parte do segundo grupo de antipsicóticos, não voltar a ser reinternado em menos de 782 dias após ter recebido alta é de 0,004.

Capítulo 4

Conclusões

Os dados analisados neste trabalho possui algumas covariáveis, uma delas é o grupo de medicação que possui três classificações: primeiro grupo de antipsicóticos, segundo grupo de antipsicóticos e clozapina. E, no artigo de Werneck et al(2012), foi observado que o primeiro e o segundo grupo de antipsicóticos apresentaram comportamentos semelhantes. Além disso, o primeiro grupo de antipsicóticos não apresentou diferença em relação ao grupo clozapina. Portanto, foram considerados somente os dados relativos ao segundo grupo de antipsicóticos e à clozapina.

Considerando os resultados obtidos a partir da análise dos ajustes dos modelos por meio de gráficos e de testes, observamos que o modelo paramétrico que melhor se ajustou aos dados foi o log-normal. Porém, os outros modelos (exponencial, weibull e log-logística) tiveram um ajuste bastante próximo do modelo escolhido. Ademais, o modelo de regressão de Cox também mostrou-se adequado para os dados.

Aqui, não há competição entre os modelos. Contudo, segundo Santos (2015) a escolha do melhor modelo a ser utilizado depende do objetivo do pesquisador. Assim, o modelo lognormal é preferível quando há interesse em realizar previsões em relação ao tempo de sobrevivência de um paciente com esquizofrenia e quando esse tempo de sobrevivência for maior que os observados na amostra.

O modelo de Cox, por sua vez, será útil quando desejarmos fazer previsão para um tempo dentro da amplitude de tempos observados (interpolação). Por ser um modelo não paramétrico, o modelo de Cox falha em prever tempos superiores àqueles observados na amostra.

Visto que os resultados desse trabalho indicaram que a suposição de riscos proporcionais do modelo de Cox não foi violada, a preferência do uso do modelo de Cox ao invés de um modelo paramétrico foi devido às vantagens na hora de interpretar os coeficientes estimados. Certamente a proporcionalidade dos riscos é uma característica muito conveniente, principalmente quando o interesse é interpretar os resultados. Os resultados obtidos nesse estudo sugerem, por exemplo, que o risco de re-hospitalização de pacientes que fazem uso de antipsicóticos de segunda geração (ASG) apresenta um risco de ser re-hospitalizado 280% maior do que pacientes que fazem uso de Clozapina, isso

para qualquer tempo t . Já o modelo de regressão log-normal não permite essa mesma interpretação, pois não impõe que os riscos sejam proporcionais.

Referências Bibliográficas

- [1] Berkson, J.; Gage, R. P. (1952). **Survival curve for cancer patients following treatment**. Journal of the American Statistical Association, 47, 501-515.
- [2] Brunello, G. H. V.; Nakano, E. Y. (2015) Inferência bayesiana no modelo Weibull discreto em dados com presença de censuras. TEMA - TEMA - Tend. Mat. Apl. Comput., Vol.16, n.2, p. 97-110.
- [3] Carrasco, C. G.; Tutia, M. H.; Nakano, E. Y. (2012). Intervalos de confiança para os parâmetros do modelo geométrico com inflação de zeros. TEMA - Tend. Mat. Apl. Comput., v.13, n.3, p.247-255.
- [4] Conover, W. J. (1999). Practical nonparametric statistics. 3. ed. New York: J Wiley, 584p.
- [5] Colosimo, E. A.; Giolo, S. R. (2006). **Análise de Sobrevida Aplicada**, (1 ed.). EDGARD BLUCHER.
- [6] Costa, N.S.S. (2013). **Modelo de riscos múltiplos com fração de cura**. Monografia(Graduação em Estatística) - Instituto de ciências Exatas, Universidade de Brasília, Brasília, p. 11.
- [7] Cox, D. R.; Snell, E. J. (1968). A general definition of residuals. J.R. Stat. Soc. B, v.30, n.2, p.248-275.
- [8] Fachini-Gomes, J. B. (2015) **Análise de Sobrevida**; Notas de Aula. Departamento de Estatística, Universidade de Brasília, Brasília.
- [9] Fachini-Gomes, J. B. (2011). **Modelos de regressão com e sem fração de cura para dados bivariados em análise de sobrevivência**. Tese (Doutorado em Ciências).Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba.
- [10] Garcia, P.N.A. (2013). **Aplicação de técnicas de Análise de Sobrevida em pacientes submetidos à intervenção coronária curânea**. Monografia(Graduação em Estatística) - Instituto de ciências Exatas, Universidade de Brasília, Brasília.
- [11] Kaplan, E. L.; Meier, P. (1958). Nonparametric estimation from incomplete observations. J. Am. Stat. Assoc., v.53, n.282, p.457-481.

- [12] Klein, J. P; Moeschberger, M. L. (2003). Survival analysis: Techniques for censored and truncated data. 2.ed. New York: Springer-Verlag, 536p.
- [13] Lawless, J. F. (2003). Statistical models and methods for lifetime data. 2.ed. New Jersey: John Wiley and Sons, 664p.
- [14] Magalhães, M. N. (2006). **Probabilidade e Variáveis Aleatórias**, (2 ed.). EDUSP.
- [15] Nakano, E. Y.; Carrasco, C. G. (2006). Uma avaliação do uso de um modelo contínuo na análise de dados discretos de sobrevivência. TEMA - Tend. Mat. Apl. Comput., v.7, n.1, p.91-100.
- [16] Nakano, E. Y.; Cunha, J. F. (2012). Análise do efeito da camuflagem no tempo de segregação em regiões texturizadas utilizando o modelo de riscos proporcionais de Cox. Semina: Ciências Exatas e Tecnológicas, v.33, n.2, p.141-148.
- [17] Oliveira, M.L. (2014). **Análise de dados de transplante de medula óssea: Proposta do modelo de regressão Kumaraswamy-Weibull com fração de cura**. Monografia(Graduação em Estatística) - Instituto de ciências Exatas, Universidade de Brasília, Brasília.
- [18] R Core Team (2013). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.Rproject.org/>.
- [19] Santana, T.V.F. (2010). **As distribuições Kumaraswamy-log-logística e Kumaraswamy-logística**. Dissertação (Mestrado em Ciências).Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba.
- [20] Santos, R. O; Nakano, E. Y. (2015). Análise do tempo de permanência de trabalhadores no mercado de trabalho do Distrito Federal via modelo de riscos proporcionais de Cox e Log-normal. Rev. Bras. Biom., v.33, n.4.
- [21] Santos, R.O. (2014). **Análise do tempo de permanência do trabalhador formal no mercado de trabalho no Distrito Federal**. Monografia(Graduação em Estatística) - Instituto de ciências Exatas, Universidade de Brasília, Brasília.
- [22] Silva, R. C. B. (2006) **Esquizofrenia: uma revisão**. Psicologia USP, São Paulo, v.17, n° 04. p. 1.
- [23] Werneck, A.; Hallak, J.; Nakano, E. Y; Elkis, H. (2011). Time to rehospitalization in patients with discharged on first generation antipsychotics, non-clozapine second generation antipsychotics, or clozapine. Psychiatry Research, v.188, n.3, p.315-319.

Apêndice A

Modelos propostos para o modelo final de Cox

Neste apêndice apresentamos tabelas com o nível de significância de cada variável dos modelos testados para ser o modelo final de Cox. A descrição desses modelos está no capítulo 2, em Ajuste do modelo de Cox.

Tabela A.1: Modelo com apenas uma covariável

Covariável	valor-p
X_1	0,43
X_2	0,036
X_3	0,801
X_4	0,984
X_5	0,989
X_6	0,308
X_7	0,0902
X_8	0,0537

Tabela A.2: Modelo: $X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8$

Covariável	valor-p
X_1	0,2967
X_2	0,0304
X_3	0,3017
X_4	0,1518
X_5	0,9952
X_6	0,0322
X_7	0,4038
X_8	0,0200

Tabela A.3: Modelo: $X_1 + X_2 + X_3 + X_4 + X_6 + X_7 + X_8$

Covariável	valor-p
X_1	0,2316
X_2	0,0295
X_3	0,3010
X_4	0,1513
X_6	0,0315
X_7	0,4029
X_8	0,0197

Tabela A.4: $X_1 + X_2 + X_3 + X_4 + X_6 + X_8$

Covariável	valor-p
X_1	0,1969
X_2	0,0270
X_3	0,3220
X_4	0,0943
X_6	0,0113
X_8	0,0028

Tabela A.5: Modelo: $X_1 + X_2 + X_4 + X_6 + X_8$

Covariável	valor-p
X_1	0,2548
X_2	0,0300
X_4	0,1452
X_6	0,0112
X_8	0,0038

Tabela A.6: Modelo: $X_2 + X_4 + X_6 + X_8$

Covariável	valor-p
X_2	0,0237
X_4	0,2462
X_6	0,0121
X_8	0,0050

Tabela A.7: Modelo: $X_2 + X_6 + X_8$

Covariável	valor-p
X_2	0,0351
X_6	0,0209
X_8	0,0077