

**MODELOS DE FRAÇÃO DE CURA APLICADOS AOS TEMPOS DE
SOBREVIVÊNCIA DE PACIENTES SUBMETIDOS À LIGADURA
ELÁSTICA DE VARIZES NO ESÔFAGO**

AGDA JÉSSICA DE FREITAS GALLETTI

Dissertação apresentada à Universidade Estadual Paulista “Júlio de Mesquita Filho” para a obtenção do título de Mestre em Biometria.

BOTUCATU
São Paulo - Brasil
Fevereiro – 2018

**MODELOS DE FRAÇÃO DE CURA APLICADOS AOS TEMPOS DE
SOBREVIVÊNCIA DE PACIENTES SUBMETIDOS À LIGADURA
ELÁSTICA DE VARIZES NO ESÔFAGO**

AGDA JÉSSICA DE FREITAS GALLETTI

Orientador: Prof. Dr. **Rogério Antonio Oliveira**

Dissertação apresentada à Universidade Estadual Paulista "Júlio de Mesquita Filho" para a obtenção do título de Mestre em Biometria.

BOTUCATU
São Paulo - Brasil
Fevereiro – 2018

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSANGELA APARECIDA LOBO-CRB 8/7500

Galletti, Agda Jéssica Freitas.

Modelos de fração de cura aplicados aos tempos de sobrevivência de pacientes submetidos à ligadura elástica de varizes no esôfago / Agda Jéssica Freitas Galletti. - Botucatu, 2018

Dissertação (mestrado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu

Orientador: Rogério Antonio Oliveira

Capes: 33004064

1. Análise de sobrevida. 2. Varizes esofágicas e gástricas. 3. Bioestatística. 4. Fígado - Cirrose.

Palavras-chave: Análise de Sobrevivência; Bioestatística; Fração de Cura; Ligadura Elástica; Varizes no Esôfago.

ATA DA DEFESA PÚBLICA DA DISSERTAÇÃO DE MESTRADO DE AGDA JESSICA DE FREITAS GALLETTI, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA, DO INSTITUTO DE BIOCIÊNCIAS - CÂMPUS DE BOTUCATU.

Aos 23 dias do mês de fevereiro do ano de 2018, às 14:00 horas, no(a) Departamento de Bioestatística, reuniu-se a Comissão Examinadora da Defesa Pública, composta pelos seguintes membros: Prof. Dr. ROGERIO ANTONIO DE OLIVEIRA - Orientador(a) do(a) Departamento de Bioestatística / Instituto de Biociências de Botucatu - UNESP, Profa. Dra. HILDETE PRISCO PINHEIRO do(a) Depto. de Estatística / Instituto de Matemática, Estatística e Computação Científica, Profa. Dra. LICIANA VAZ DE ARRUDA SILVEIRA do(a) Departamento de Bioestatística / Instituto de Biociências de Botucatu - UNESP, sob a presidência do primeiro, a fim de proceder a arguição pública da DISSERTAÇÃO DE MESTRADO de AGDA JESSICA DE FREITAS GALLETTI, intitulada **MODELOS DE FRAÇÃO DE CURA APLICADOS AOS TEMPOS DE SOBREVIVÊNCIA DE PACIENTES SUBMETIDOS À LIGADURA ELÁSTICA DE VARIZES NO ESÔFAGO**. Após a exposição, a discente foi arguida oralmente pelos membros da Comissão Examinadora, tendo recebido o conceito final: APROVADA. Nada mais havendo, foi lavrada a presente ata, que após lida e aprovada, foi assinada pelos membros da Comissão Examinadora.

Prof. Dr. ROGERIO ANTONIO DE OLIVEIRA



Profa. Dra. HILDETE PRISCO PINHEIRO



Profa. Dra. LICIANA VAZ DE ARRUDA SILVEIRA



Dedicatória

*À minha tia Ireni (in memoriam),
por sempre cuidar de mim e me amar como
filha.*

*À minha mãe Joana,
por sempre me apoiar e ser para mim exemplo
de generosidade e honestidade.*

Agradecimentos

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, pela concessão de bolsa de estudos.

Ao Prof Dr. Rogério Antonio Oliveira pela orientação, amizade, conversas, conselhos, compreensão, incentivos e apoio incondicional.

À Profa. Dra. Liciana Vaz, pela amizade, incentivos, conversas, preocupação e pelas oportunidades de conhecimento a mim oferecida.

À todos professores e funcionários do Departamento de Bioestatística , Unesp, Botucatu.

À toda minha família que sempre esteve ao meu lado, principalmente à minha mãe por ter todo apoio e confiança.

À todos os alunos de mestrado e doutorado, que direta ou indiretamente, contribuíram positivamente para que este trabalho fosse realizado.

Aos meus amigos que fizeram dessa jornada mais agradável e feliz. Obrigada pelos abraços, sorrisos, eventuais momentos de lazer, sessões de cinema, dias e noites de estudo, alimentação saudável, besteiras também, companheirismo, confiança e inúmeros incentivos.

À todos que de alguma forma contribuíram para a realização deste trabalho.

Sumário

	Página
LISTA DE FIGURAS	vi
LISTA DE TABELAS	vii
RESUMO	viii
SUMMARY	x
1 INTRODUÇÃO	1
2 ANÁLISE DE DADOS CENSURADOS	3
2.1 Conceitos básicos	3
2.2 Funções associadas ao tempo de sobrevivência	5
2.3 Estimação não paramétrica da função de sobrevivência	7
2.4 Estimação paramétrica da função de sobrevivência	11
2.4.1 Distribuição Gama Generalizada e particularidades	11
2.4.2 Outras distribuições	14
2.5 Modelos de fração de cura ou longa duração	15
2.5.1 Teste para presença de imunes e tempo de acompanhamento suficiente	16
3 MODELO DE REGRESSÃO COM	
FRAÇÃO DE CURA	21
3.1 Formulação	21
3.2 Estimação dos parâmetros do modelo	22

	v
3.2.1 Algoritmo EM	24
3.2.2 Implementação no R (R Core Team, 2017)	26
3.3 Teste paramétrico para presença de imunes	28
3.4 Seleção de modelo	29
3.4.1 Teste de Wald	29
3.4.2 Testa da razão de verossimilhanças	30
3.4.3 AIC e BIC	31
3.5 Diagnóstico	31
3.5.1 Análise de resíduos	31
3.5.2 Teste de Kolmogorov-Smirnov	33
3.5.3 Teste para bondade do ajuste	33
4 APLICAÇÃO	35
4.1 Entendendo a doença	35
4.2 Caracterização dos pacientes	37
4.3 Seleção das covariáveis para o modelo de regressão	44
4.4 Modelagem Estatística	47
4.4.1 Definição do modelo	47
4.4.2 Diagnóstico	51
4.4.3 Interpretação do modelo	56
5 CONSIDERAÇÕES FINAIS	59
ANEXOS	61
REFERÊNCIAS BIBLIOGRÁFICAS	81
APÊNDICES	86

Lista de Figuras

		Página
1	Gráfico ilustrativo de alguns TTT plots	8
2	Histograma e Boxplot das covariáveis numéricas do estudo.	39
3	Boxplot e curva de sobrevivência KM para relacionar covariável com o tempo de acompanhamento.	41
4	Boxplot e curva de sobrevivência KM para relacionar covariável com o tempo de acompanhamento.	43
5	Boxplot e curva de sobrevivência KM para relacionar covariável com o tempo de acompanhamento.	45
6	Sobrevivência dos pacientes submetidos à LEVE estimada via KM. . . .	48
7	TTT-plot dos tempo de falha dos pacientes suscetíveis.	50
8	Resíduos Cox-Snell para os modelos ajustados.	52
9	Resíduos Cox-Snell para os modelos ajustados II.	53
10	Resíduos de Martingal e Deviance para o modelo Weibull Modificado com Fração de Cura.	55
11	Funções de sobrevivência estimadas para pacientes submetidos à LEVE .	58

Lista de Tabelas

		Página
1	Relação entre as funções associadas à função densidade de uma v.a. . . .	8
2	Percentis do coeficiente de correlação.	34
3	Medidas resumo das variáveis do estudo.	38
4	Teste para selecionar as covariáveis candidatas.	46
5	Percentil de $\hat{p}_n$ com distribuição da censura $Exp(1/\mu)$	49
6	Percentis de q_n com $1 - p = 0,2$ (proporção de suscetíveis)	49
7	Teste de Kolmogorov-Smirnov	53
8	Percentis do coeficiente de correlação r com $1 - p = 0,2$	54
9	Critérios de seleção de modelos paramétricos com fração de cura	55
10	Teste para bondade do ajuste	56
11	Estimativas do modelo de regressão Weibull Modificada com fração de cura	57

MODELOS DE FRAÇÃO DE CURA APLICADOS AOS TEMPOS DE SOBREVIVÊNCIA DE PACIENTES SUBMETIDOS À LIGADURA ELÁSTICA DE VARIZES NO ESÔFAGO

Autora: AGDA JÉSSICA DE FREITAS GALLETTI

Orientador: Prof. Dr. ROGÉRIO ANTONIO OLIVEIRA

RESUMO

A cirrose é uma doença hepática assintomática que, muitas vezes, é descoberta quando o quadro é irreversível. Por isso, o tratamento consiste em uma série de medidas para controlar o avanço da enfermidade, visto que a principal consequência da cirrose é o aumento da pressão na veia portal, que por sua vez, acarreta no surgimento de varizes e no seu respectivo rompimento, podendo ser fatal. Estudos relacionados à esta doença são muito importantes, pois a análise estatística é uma ferramenta que permite auxiliar na tomada de decisões nos procedimentos médicos e acompanhamento de pacientes. Um método estatístico bastante explorado nas ciências biomédicas é a análise de sobrevivência, que consiste em descrever o tempo de um evento inicial até a ocorrência de um outro de interesse. No entanto, existem situações em que uma proporção da amostra não vivencia o desfecho de interesse, mesmo que acompanhado por um período longo de

tempo. Nestes casos, tais observações são dita imunes ao desfecho de interesse e as metodologias tradicionais de análise de sobrevivência não são indicadas. Logo, os modelos de fração de cura ou de longa duração, desenvolvido a partir do modelo de mistura, são os utilizados nestas situações. Neste trabalho diversos modelos foram considerados para analisar os tempos de vida de pacientes submetidos à Ligadura Elástica de Varizes Esofágicas, ao qual foram anotados os tempos até o óbito durante o acompanhamento de 129 pacientes do Hospital das Clínicas da Faculdade de Medicina da UNESP, Campus Botucatu (SP), no período de 2006 a 2010. O modelo Weibull Modificado com fração de cura foi considerado adequado, indicando que quanto maior a idade e o grau da doença hepática, se os pacientes não usam o medicamento β _Bloqueador e são homens, menor é a chance deles serem sobreviventes de longa duração.

Palavras-chave: Análise de sobrevivência, Modelos de Longa duração, Fração de cura, Ligadura Elástica, Varizes no Esôfago.

Author: AGDA JÉSSICA DE FREITAS GALLETTI
Adviser: Prof. Dr. ROGÉRIO ANTONIO OLIVEIRA

SUMMARY

Cirrhosis is an asymptomatic liver disease that is often discovered when the patient's condition is irreversible. Therefore, the treatment consists of a series of measures to control the progression of the disease, since the main consequence of the cirrhosis is the increase of the portal venous pressure, which causes the appearance of varices and their respective rupture may be fatal. Studies related to that disease are very important, so the statistical analysis is a tool that helps to make decisions in medical procedures and patient follow-up. The most applied statistical method in the biomedical sciences is survival analysis, which consists of describing the time of occurrence until the event of interest. However, there are situations in which a proportion of the sample does not experience the interest outcome, even if they are accompanied by a long period of time. In such cases, such observations are said to be immune to the outcome of interest and traditional survival analysis methodologies are not appropriated. Therefore, the cure fraction or long duration models can be

used in these situations because they incorporate mixtures of models to solve the complexity inherent in the actual study. In this work, some statistical models were considered to analyze the survival times of patients, after surgery of Endoscopic Band Ligation of the Esophageal Varices, such as Exponential, Gamma and Weibull models. The data is related to survival times of 129 patients, who were treated in the Hospital das Clínicas of the Medical School of UNESP, Campus Botucatu (SP), from 2006 to 2010. The Modified Weibull distribution with cure rate was considered adequate to the data, indicating that older male patients with higher Child-Pugh score without taking beta blockers medicine are more likely not to be a long-term survivor.

Palavras-chave: Survival analysis, Long-Term Survivors, Cure Rate, Band Ligation, Esophageal Varices.

1 INTRODUÇÃO

A cirrose é uma doença hepática assintomática que, muitas vezes, é descoberta quando o quadro é irreversível. Esta doença é caracterizada como o desenvolvimento histológico de nódulos regenerativos cercados por bandas fibrosas, em resposta à lesão hepática crônica, que leva a hipertensão portal (Schuppan & Afdhal, 2009). Infelizmente, a cura somente é possível por meio do transplante de fígado. Todavia, existem tratamentos que podem auxiliar no processo da não progressão da doença.

A principal consequência da cirrose (Albillos & Garcia-Tsao, 2011) é a hipertensão portal, que é uma síndrome clínica caracterizada pelo aumento mantido na pressão venosa em níveis acima dos fisiológicos, isto é, pressão superior a 5 mmHg (Coelho et al., 2014). Sucessivamente, o aumento da pressão aumenta o risco do surgimento de varizes esofágicas e, conseqüentemente, seu rompimento, “uma das principais complicações” (Biecker, 2013).

Toshikuni et al. (2016), Coelho et al. (2014) e Biecker (2013) apontaram algumas medidas tomadas a fim de que não haja o sangramento agudo, hemorragia ativa devido à ruptura de varizes esofágicas ou sangue no esôfago/estômago sem que se tenha detectado outra fonte durante a endoscopia. Dentre as principais medidas, tem-se a profilaxia primária, que visa prevenir o primeiro sangramento de varizes e a profilaxia secundária, que tem como objetivo, minimizar o risco da recidiva hemorrágica de varizes.

A Ligadura Elástica de Varizes Esofágicas (LEVE) é frequentemente recomendada como um método eficaz para o tratamento de varizes no esôfago e apresenta efetividade no controle do sangramento em 86 a 92% dos casos (Li et al., 2017).

Estudos relacionados ao desenvolvimento desta doença em dados coletados de

pacientes são muito importantes, pois a análise estatística é uma ferramenta que permite auxiliar na tomada de decisões nos procedimentos médicos e acompanhamento de pacientes. No âmbito de pesquisas médicas, estas envolvem diagnóstico, prognóstico ou indicação de tratamento. Carvalho et al. (2005) apontam como uma das vantagens da utilização de modelos estatísticos para a descrição e representação do conhecimento científico, a possibilidade de afirmações probabilísticas.

Um método estatístico bastante explorado nas ciências biomédicas é a análise de sobrevivência, que consiste em descrever o tempo até a ocorrência de um evento de interesse, caracterizado pelo tempo de falha e pela censura (por exemplo, óbito devido a falência hepática), que juntos compõem a variável resposta a ser estudada. No entanto, existem estudos em que uma proporção da amostra não vivencia o evento de interesse, mesmo que acompanhado por um período longo de tempo. Nestes casos, tais observações são ditas imunes (curadas ou não suscetíveis) ao desfecho de interesse e as metodologias tradicionais de análise de sobrevivência não são indicadas, pois não comportam a proporção de imunes da população. Por isso, os modelos de fração de cura, desenvolvidos a partir do modelo de mistura proposto por Berkson & Gage (1952), são os utilizados nestas situações.

Diante disso, devido aos indícios de haver uma proporção de pacientes imunes na amostra, este trabalho tem como objetivo ajustar modelos estatísticos com fração de cura a fim de avaliar o tempo de sobrevivência de pacientes que fizeram a LEVE para o tratamento de varizes no esôfago. O banco de dados consiste de informações coletadas de 129 pacientes acompanhados no Hospital das Clínicas da Faculdade de Medicina da UNESP, Campus Botucatu (SP), submetidos ao procedimento no período de 2006 a 2010.

O Capítulo 2 é composto por um breve resumo de conceitos básicos sobre análise de sobrevivência. No Capítulo 3 é exposto o modelo de regressão com fração de cura geral, abordando que os tempos de falha dos indivíduos suscetíveis podem assumir qualquer distribuição. O Capítulo 4 mostra uma análise estatística do conjunto de dados em estudo e por fim, o Capítulo 5 apresenta as considerações finais.

2 ANÁLISE DE DADOS CENSURADOS

2.1 Conceitos básicos

Kleinbaum & Klein (2012) definem análise de sobrevivência como uma coleção de procedimentos estatísticos para análise de dados em que a variável resposta de interesse é o tempo até a ocorrência de um evento, denominado tempo de falha (tempo de seguimento).

O tempo de falha é o intervalo de tempo entre a exposição e a ocorrência do evento de interesse, que pode ser a morte, a incidência ou a remissão de uma doença, a inadimplência de um cliente, a falha de um equipamento, o período de desemprego ou o acionamento do seguro, entre outros.

Em geral, é necessário para o tempo de falha:

- Fixar o tempo de início do estudo;
- Determinar a escala de medida a ser utilizada;
- Estabelecer o evento de interesse.

Frequentemente ocorre que para alguns indivíduos em estudo, o desfecho de interesse nem sempre ocorre, ou a informação dele se perde durante o processo, resultando em observações parciais ou incompletas. Essa ocorrência é chamada de censura e a presença de indivíduos censurados é a característica primordial neste tipo de análise.

Alguns exemplos de censura mais encontradas são: morte por causas não relacionadas ao evento de interesse; perda de contato com o paciente; recusa em continuar

participando do estudo (desistência); mudança de procedimento, por exemplo, o paciente recebe um transplante e termina o tratamento.

Diferentemente de truncamento, que é a exclusão de certos indivíduos do estudo que apresentam alguma característica indesejada. Por exemplo, em uma pesquisa retira-se uma amostra dos dados de um hospital, no entanto, tem-se como objetivo, estudar o tempo de vida de mulheres com câncer de mama. Dessa forma, apenas mulheres que apresentam a doença fará parte da amostra a ser modelada e os outros pacientes serão truncados.

O tempo de falha t_i , $1 \leq i \leq n$ assume o menor valor entre t_i^* , tempo de sobrevivência até a falha, e u_i , tempo de falha, ou seja,

$$t_i = \min(t_i^*, u_i), \quad 1 \leq i \leq n.$$

A variável indicadora de falha (censura) é definida como,

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

Ao contrário da teoria clássica, que descarta os indivíduos censurados, na análise de sobrevivência todas as informações incompletas ou parciais são levadas em consideração. Colosimo & Giolo (2006) apresentam dois motivos para essa abordagem:

- i) mesmo sendo incompletas, as observações censuradas fornecem informações sobre o tempo de vida de pacientes;
- ii) a omissão das censuras no cálculo das estatísticas de interesse pode acarretar conclusões viciadas.

A censura pode ser à direita, à esquerda ou intervalar. A censura à direita é a mais comum na literatura, que é caracterizada pelo tempo de ocorrência do evento estar à direita do tempo registrado. Ela pode ser classificada em

- **Censura do tipo I :** O estudo termina após um período de tempo fixo t_0 e ao final desse tempo um ou mais indivíduos não falharam.
- **Censura do tipo II:** O término do estudo acontece quando ocorre o evento de interesse em um número pré-fixado $k \leq n$ de indivíduos.
- **Censura Aleatória:** é o caso em que os indivíduos não sofreram o evento de interesse por motivos não controláveis, como por exemplo, a morte por razão diferente da estudada.

2.2 Funções associadas ao tempo de sobrevivência

Seja T o tempo de sobrevivência, ou seja, o tempo decorrido até a ocorrência do evento de interesse, uma variável aleatória não-negativa contínua (em certos estudos pode ser discreta). Então, as funções que descrevem a sua distribuição de probabilidade são apresentadas a seguir.

- Função densidade

A função densidade $f(t)$ pode ser interpretada como a probabilidade de um indivíduo falhar em um intervalo de tempo $[t, t + \Delta t)$ e denotada por:

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t},$$

tal que $f(t) \geq 0$ para todo $t \geq 0$.

E a função distribuição acumulada é dada por

$$F(t) = P(T < t) = \int_0^t f(u)du, \text{ para } t > 0$$

- Função de sobrevivência

A função de sobrevivência é definida como a probabilidade de uma observação não falhar até um certo tempo t , ou seja, a probabilidade de uma

observação sobreviver ao tempo t (Colosimo & Giolo, 2006). Em outros termos,

$$S(t) = P(T \geq t) = 1 - F(t),$$

satisfazendo às seguintes propriedades:

- $S(0) = 1$;
- $\lim_{t \rightarrow \infty} S(t) = 0$; e
- $S(t)$ é decrescente.

- Função de risco

A função de risco é amplamente utilizada para expressar o risco (instantâneo) de morte em algum momento t e pode ser obtida a partir da probabilidade de um indivíduo morrer no tempo t , condicional ao fato de ele ter sobrevivido até aquele momento (Collett, 2003).

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

A função de risco acumulada fornece a taxa de falha acumulada do indivíduo até um determinado momento e pode ser útil na estimação não paramétrica de $h(t)$. Algebricamente, pode ser denotada por:

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

Também denominada taxa de falha, “a função de risco é mais informativa do que a função de sobrevivência. Diferentes funções de sobrevivência podem ter formas semelhantes, enquanto as respectivas funções taxa de falha podem diferir drasticamente” (Colosimo & Giolo, 2006).

Esta característica possibilita a indicação de um modelo probabilístico mais apropriado para o tempo de sobrevivência T por meio do gráfico do Tempo

Total em Teste (*TTT plot*). A curva TTT é obtida construindo um gráfico de

$$G(r/n) = \frac{(\sum_{i=1}^r T_{i:n}) + (n-r)T_{r:n}}{\sum_{i=1}^n T_i},$$

por r/n , sendo que $r = 1, \dots, n$, e $T_{i:n}, i = 1, \dots, n$ são as estatística de ordem da amostra.

Uma característica importante da transformação TTT é que ela fornece informações imediatas sobre a forma (*shape*) da taxa de falha (Aarset, 1987).

Barlow & Campo (1975) mostraram que a função de risco de uma variável aleatória T cresce, decresce ou se mantém constante se a nuvem de pontos obtida pela transformação TTT for côncava, convexa ou diagonal, respectivamente. Em seguida, Aarset (1987) explorou a taxa de falha da forma de U (banheira) e, anos depois, Mudholkar et al. (1996) evidenciaram a função taxa de falha unimodal, como consequência dos estudos sobre a distribuição Weibull Generalizada.

A Figura 1 ilustra as formas citadas anteriormente que a taxa de falha pode assumir. A curva côncava (A), indica que a taxa de falha é monótona crescente; a curva convexa e depois côncava (B), indica a forma de banheira; a curva côncava e depois convexa (C), indica forma unimodal; a curva diagonal (D), indica taxa constante; e a curva convexa (E), indica forma monótoma decrescente.

As funções associadas ao tempo de sobrevivência estão relacionadas entre si, conforme a Tabela 1, ou seja, “o conhecimento de uma das funções, por exemplo, $S(t)$, implica no conhecimento das demais” (Colosimo & Giolo, 2006). As demonstrações das respectivas relações, Tabela 1, podem ser facilmente encontradas na literatura.

2.3 Estimação não paramétrica da função de sobrevivência

Uma técnica utilizada para se estimar a curva de sobrevivência na presença de censura é o estimador de Kaplan-Meier (EKM). Existem outros estimadores que

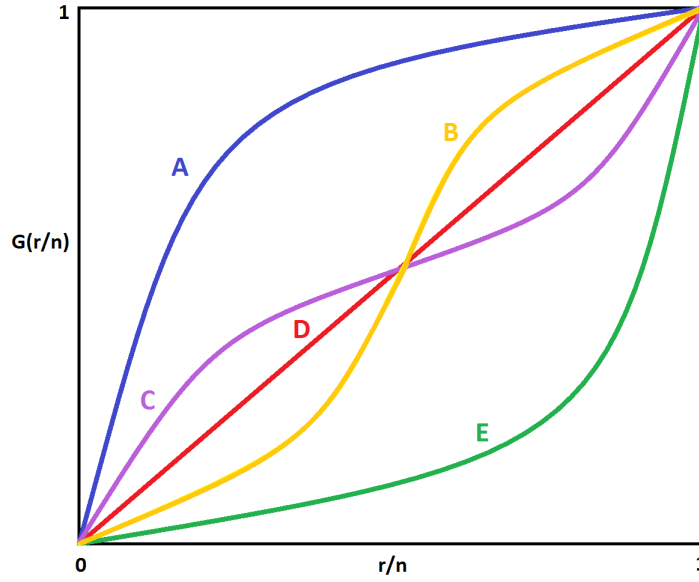


Figura 1: Gráfico ilustrativo de alguns TTT plots

Tabela 1: Relação entre as funções associadas à função densidade de uma v.a.

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

também podem ser explorados a fim de estimar a curva de sobrevivência, como o estimador de Nelson-Aalen e a aquelas baseadas na Tabela de Vida. No entanto, o estimador de Kaplan-Meier é preferível pelo fato de ser um estimador de máxima verossimilhança.

Proposto por Kaplan & Meier (1958), o EKM é também denominado de estimador produto-limite, sendo definido por:

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

tal que,

- $t_j, j = 1, \dots, k$, representa o j -ésimo tempo de falha. Os tempos devem ser ordenados e distintos;

- d_j é o número de falhas em t_j ;
- n_j é o número de indivíduos em risco até t_j .

A probabilidade de sobreviver até o período t_j , considerando as sobrevivências até certo tempo são independentes, é o produto da probabilidade de se chegar até cada um dos tempos anteriores.

Kaplan & Meier (1958) provaram que $\hat{S}(t)$ é o estimador de máxima verossimilhança de $S(t)$, consequentemente não viciado para amostras grandes. Além disso, o EKM é fracamente consistente e converge assintoticamente para um processo gaussiano (Colosimo & Giolo, 2006).

A fórmula de Greenwood fornece a variância assintótica do EKM, obtida a partir das propriedades do estimador de máxima verossimilhança, sendo dada por:

$$\widehat{Var}(\hat{S}(t)) = [\hat{S}(t)]^2 \sum_{j:t_j < t} \frac{d_j}{n_j(n_j - d_j)}. \quad (1)$$

O intervalo aproximado de $100(1 - \alpha)\%$ de confiança para $S(t)$ é expresso por:

$$\hat{S}(t) \pm z_{\alpha/2} \sqrt{\widehat{Var}(\hat{S}(t))},$$

em que, $z_{\alpha/2}$ é o percentil $(1 - \alpha/2)100\%$ da distribuição normal padrão. Entretanto, quando $S(t)$ assume valores extremos, isto é, a função de sobrevivência estimada é próxima de zero ou um, alguns intervalos estimados podem não estar contidos em $[0, 1]$, sendo assim, inapropriado. Logo, um procedimento alternativo é transformar $\hat{S}(t)$.

As principais transformações possíveis são:

- transformação logística: $\log\{S(t)/(1 - S(t))\}$; e
- transformação complemento log-log: $\log\{-\log S(t)\}$.

Considerando a transformação, $\hat{U}(t) = \log\{-\log \hat{S}(t)\}$, tem-se que a variância assintótica é expressa por:

$$\widehat{Var}(\hat{U}(t)) = \frac{\sum_{j:t_j < t} \frac{d_j}{n_j(n_j - d_j)}}{[\log \hat{S}(t)]^2}. \quad (2)$$

O intervalo aproximado de $100(1 - \alpha)\%$ de confiança para $S(t)$, pode ser obtido a partir do intervalo aproximado de $U(t)$, conforme demonstrado a seguir.

$$\begin{aligned} IC_{(1-\alpha)100\%} &= \hat{U}(t) \pm z_{\alpha/2} \sqrt{\widehat{Var}(\hat{U}(t))} \\ &= \log\{-\log \hat{S}(t)\} \pm z_{\alpha/2} \sqrt{\widehat{Var}(\hat{U}(t))} \\ &= -\exp\left\{\log\{-\log \hat{S}(t)\} \pm z_{\alpha/2} \sqrt{\widehat{Var}(\hat{U}(t))}\right\} \\ &= \log \hat{S}(t) \exp\left\{\pm z_{\alpha/2} \sqrt{\widehat{Var}(\hat{U}(t))}\right\} \\ &= \exp\left\{\log \hat{S}(t) \exp\left\{\pm z_{\alpha/2} \sqrt{\widehat{Var}(\hat{U}(t))}\right\}\right\} \\ &= [\hat{S}(t)]^{\exp\{\pm z_{\alpha/2} \sqrt{\widehat{Var}(\hat{U}(t))}\}} \end{aligned}$$

Kalbfleisch & Prentice (2002) e Collett (2003) discutem detalhadamente os passos da demonstração da Equação 1 e também da transformação de $S(t)$ que resultou na Equação 2.

Quando há variáveis explicativas categóricas no estudo, compara-se o efeito das categorias no tempo, estimando-se o EKM para $S(t)$ para cada grupo. Os gráficos das curvas de sobrevivência das categorias podem ser ilustrados para verificar se há diferença entre as categorias ou não. Entretanto, deve-se utilizar testes de hipóteses para verificar se as diferenças encontradas são estatisticamente significativas.

Dentre os testes não paramétricos, os mais utilizados são os testes logrank (Mantel, 1966), que é particularmente apropriado quando a razão das funções de taxa de falha dos grupos a serem comparados é aproximadamente constante (Colosimo & Giolo, 2006) e a generalização do teste Wilcoxon proposto por (Gehan, 1965). A principal diferença entre os testes é que o de Wilcoxon utiliza peso igual ao número

de indivíduos sob risco, consequentemente apresenta maior penalização no início do acompanhamento, diferente do logrank, que coloca o mesmo peso para qualquer que seja o tempo de seguimento.

O modelo de Cox é o mais popular dentre os utilizados em análise de sobrevivência. A proporcionalidade dos riscos é a suposição essencial para a utilização deste modelo e pode ser verificada ao construir gráficos da taxa de falha e também da função de sobrevivência empírica KM, considerando cada covariável em estudo. Também pode ser verificada pelo teste de riscos proporcionais, que utiliza a medida de correlação de Pearson entre os resíduos padronizados de Schoenfeld.

Quando não se verifica a proporcionalidade dos riscos, então o ajuste do modelo de Cox é inadequado, sendo necessário a aplicação de outros modelos de sobrevivência que incorpore esta característica.

2.4 Estimação paramétrica da função de sobrevivência

Nesse tipo de modelagem, assume-se que o tempo T até o evento segue uma distribuição conhecida de probabilidade e estimam-se os parâmetros correspondentes do modelo (Carvalho et al., 2005).

Para escolha da distribuição deve ser levado em consideração que a variável aleatória T é contínua, não negativa e assimétrica à direita. Nesse sentido, serão apresentados algumas distribuições mais utilizadas em dados de sobrevivência.

2.4.1 Distribuição Gama Generalizada e particularidades

A distribuição Gama Generalizada é caracterizada por ter um parâmetro de escala α e dois de forma, γ e k , tornando-a mais flexível.

A função densidade é dada por:

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

em que, γ , α e $k > 0$

Destaca-se como principal propriedade, a obtenção de outras distribuições dependendo dos valores dos seus respectivos parâmetros. A seguir destacam-se os principais casos na literatura:

a) Distribuição Exponencial ($\gamma = k = 1$)

A distribuição Exponencial é uma das mais simples utilizadas para descrever o tempo de falha. Este modelo é amplamente utilizado em estudos de confiabilidade (“sobrevivência para engenharias”). A função de densidade é definida como:

$$f(t) = \frac{1}{\alpha} \exp \left\{ - \left(\frac{t}{\alpha} \right) \right\}, t \geq 0 \text{ e } \alpha > 0.$$

As funções de sobrevivência $S(t)$ e função de risco $h(t)$ são expressas por:

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

e

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

Note que $h(t)$ não depende de t (taxa de falha é constante), ou seja, um indivíduo que passou mais tempo sob risco tem a mesma chance de falhar que um com tempo inferior. Tem-se também que a média é igual a α e a variância, α^2 .

b) Distribuição Weibull ($k = 1$)

A Weibull é bastante utilizada em estudos biomédicos e sua popularidade se deve ao fato de sua função taxa de falha apresentar uma grande variedade de formas, podendo ser crescente, decrescente ou constante (Colosimo & Giolo, 2006).

Dado T uma variável aleatória com distribuição Weibull, então as funções a ela associadas são respectivamente:

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

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

e

$$h(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1}, \alpha \text{ e } \gamma > 0.$$

O parâmetro de forma da Weibull é γ , que determina a forma da função de risco. Neste caso, para $\gamma < 1$, a $h(t)$ é decrescente; para $\gamma > 1$, $h(t)$ é crescente; e para $\gamma = 1$, $h(t)$ é constante e a distribuição equivale à Exponencial.

c) Distribuição Gama ($\gamma = 1$)

De acordo com Colosimo & Giolo (2006), a distribuição Gama inicialmente era aplicada a problemas de confiabilidade por se ajustar adequadamente a uma variedade de fenômenos da área médica.

A sua função densidade é denotada por:

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

em que $\Gamma(k) = \int_0^\infty x^{k-1} \exp(-x) dx$.

As demais funções não são triviais de se obter, devido à dificuldade em integrar a função gama $\Gamma(x)$. Assim, as devidas equações são expressas da seguinte forma:

$$S(t) = \int_t^\infty \frac{1}{\Gamma(k)\alpha^k} u^{k-1} \exp \left\{ - \left(\frac{u}{\alpha} \right) \right\} du$$

e a função taxa de falha pode ser obtida da relação $h(t) = f(t)/S(t)$, podendo assumir forma constante, crescente ou decrescente.

d) Distribuição Lognormal ($k \rightarrow \infty$)

Ao assumir que T segue uma distribuição Lognormal, tem-se que a transformação $\log(T)$ tem distribuição Normal (Carvalho et al., 2005).

A função densidade da variável T com distribuição lognormal é dada por:

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

em que $\sigma, t > 0$ e $-\infty < \mu < \infty$.

Entretanto, as funções de sobrevivência e taxa de falha da distribuição lognormal não apresentam uma forma analítica explícita e são representadas de acordo com as seguintes expressões:

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

e $h(t) = f(t)/S(t)$, tal que $\Phi(\cdot)$ é a função distribuição acumulada de uma normal padrão (Colosimo & Giolo, 2006).

Diferentemente das distribuições apresentadas anteriormente, a forma da sua função de risco é unimodal.

2.4.2 Outras distribuições

Distribuição Log-logística

A distribuição Log-logística, apesar de não ser uma particularidade da Gama Generalizada, tem características em comum com a distribuição Lognormal, pois a função taxa de falha também apresenta forma unimodal. Da mesma maneira em que, para Weibull e a própria Lognormal, a transformação de T ($\log(T)$) segue distribuição do Valor Extremo e Normal, respectivamente, para distribuição Log-logística, $\log(T)$ segue uma logística.

A função densidade pode ser expressa por:

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

É simples determinar as funções de sobrevivência e taxa de falha, que são apresentadas, respectivamente, por:

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

e

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

As distribuições apresentadas contemplam, com exceção da banheira, todas as formas que a função de risco pode assumir. Por este motivo, a seguir, serão apresentadas exemplos de distribuições que podem assumir forma de banheira (U), além das demais formas. Nesta etapa, apenas será evidenciado a densidade e a função de sobrevivência para cada distribuição.

Distribuição Weibull Exponencializada (Mudholkar & Srivastava, 1993)

$$f(t) = \frac{\alpha\theta}{\sigma} \left[1 - \exp \left(- \left(\frac{t}{\sigma} \right)^\alpha \right) \right]^{\theta-1} \exp \left(- \left(\frac{t}{\sigma} \right)^\alpha \right) \left(\frac{t}{\sigma} \right)^{\alpha-1}$$

e

$$S(t) = 1 - \left[1 - \exp \left(- \left(\frac{t}{\sigma} \right)^\alpha \right) \right]^\theta,$$

em que α, θ, σ e $t > 0$.

Distribuição Weibull Modificada (Lai et al., 2003)

$$f(t) = a(b + \lambda t)t^{b-1} \exp(\lambda t) \exp[-at^b \exp(\lambda t)]$$

e

$$S(t) = \exp[-at^b \exp(\lambda t)],$$

tal que, $a, \lambda, t > 0$ e $b \geq 0$.

Distribuição Weibull Estendida (Xie et al., 2002)

$$f(t) = \lambda\beta \left(\frac{t}{\alpha} \right)^{\beta-1} \exp \left[\left(\frac{t}{\alpha} \right)^\beta + \lambda\alpha \left(1 - \exp \left[\left(\frac{t}{\alpha} \right)^\beta \right] \right) \right]$$

e

$$S(t) = \exp \left[\lambda\alpha \left(1 - \exp \left[\left(\frac{t}{\alpha} \right)^\beta \right] \right) \right],$$

em que, $\lambda, \alpha, \beta > 0$ e $t \geq 0$.

2.5 Modelos de fração de cura ou longa duração

Na teoria de análise de sobrevivência usual, o pressuposto do modelo é que se o indivíduo for acompanhado por um período suficiente de tempo, o evento de interesse irá ocorrer. Contudo, na prática, pode ocorrer que a falha nunca seja observada, devido à possibilidade do indivíduo ser imune ao evento ("curado"). Modelos que tratam de dados de sobrevivência com estas características são chamados de modelos com fração de cura ou modelos de longa duração. Entretanto, Maller & Zhou (1996)

sugerem que o tempo de acompanhamento deve ser suficientemente grande para ter indícios reais de que existe uma fração de curados.

Um dos modelos de fração de cura mais utilizados é o modelo de mistura padrão discutido por Berkson & Gage (1952). Nele, assume-se que uma proporção (fração) π da população é imune (não-suscetíveis ou curados) e $1 - \pi$ é susceptível a experimentar um determinado evento de interesse, em que $0 < \pi < 1$.

Maller & Zhou (1996) definiram que quando uma função de sobrevivência não satisfaz à propriedade $\lim_{t \rightarrow \infty} S(t) = 0$ é dita imprópria ou de longa duração. Uma maneira de verificar é a partir da construção de um gráfico da função de sobrevivência empírica de Kaplan-Meier, em que poderá ser observado que a cauda da função está tendendo a um valor constante diferente de zero durante um período longo de tempo. Outra característica dos modelos de longa duração é a presença de grande quantidade de censuras ao final do estudo.

A função de sobrevivência imprópria da população é expressa por

$$S(t) = \pi + (1 - \pi)S^*(t), \quad (3)$$

em que, $S^*(t)$ denota a função de sobrevivência para a porção não curada da população. $S(t)$ possui as seguintes propriedades:

- i) Se $\pi = 0$, então $S(t) = S^*(t)$;
- ii) $S(0) = 1$;
- iii) $S(t)$ é decrescente; e
- iv) $\lim_{t \rightarrow \infty} S(t) = \pi$.

As demais funções associadas a variável aleatória T , tempo de sobrevivência, podem ser obtidas a partir das relações listadas na Tabela 1.

2.5.1 Teste para presença de imunes e tempo de acompanhamento suficiente

A presença da proporção de imunes é a principal justificativa para a utilização dos modelos de fração de cura. Portanto, verificar se realmente existem indivíduos

não suscetíveis na população ao qual a amostra estudada foi retirada, é uma questão relevante. No entanto, quando se detecta esta proporção de imunes, tem-se também que se certificar de que o tempo de acompanhamento dos indivíduos amostrados foi suficiente.

Suponha que a função de sobrevivência $S(t)$ seja imprópria, isto é, $\lim_{t \rightarrow \infty} S(t) = \pi$, em que π é a proporção de imunes, então tem-se interesse em testar

$$H_0 : S(t_{(n)}) = S(t_G) = 0, \quad t_{(n)} = \max(t_i), \quad 1 \leq i \leq n,$$

em que, t_G representa o limite da observação na amostra em que nenhum maior tempo de sobrevivência será observado ($t_{(n)}$).

Ao considerar H_0 verdadeira, neste caso, tem-se duas implicações:

- i) assume-se não haver imunes na população; ou
- ii) assume-se que o tempo de acompanhamento foi suficiente.

Se H_0 for rejeitada, então,

- i) existem imunes na população, detectada por um tempo suficiente de acompanhamento da amostra; ou
- ii) pode ou não existir imunes na população, mas o tempo de acompanhamento da amostra foi insuficiente.

Dado que tem-se indícios de que $S(t_G) > 0$, então será testado se o tempo de acompanhamento foi suficiente, isto é,

$$H_0 : t_{F_0} \leq t_G,$$

tal que, t_{F_0} , o maior tempo de sobrevivência que poderá ser observado dentre os suscetíveis.

Desta forma pode-se concluir que:

- i) se H_0 não for rejeitada, então, $S(t_{F_0}) > 0$, ou seja, tem-se evidências de haver proporção de imunes na população;
- ii) se H_0 for rejeitada, então, o tempo de acompanhamento foi insuficiente.

Logo, os testes para presença de censura e tempo de acompanhamento podem ser não paramétricos ou paramétricos como descritos a seguir.

Teste não paramétrico

Considere que $\delta_{(n)} = \max(\delta_i)$, $1 \leq i \leq n$, quando $\delta_{(n)} = 1$, implica que $S(t_{(n)}) = 0$, então, $H_0 : S(t_{(n)}) = 0$ é não rejeitado, ou seja, há indícios de não haver imunes na população.

Agora, suponha que $1 - S(t_{(n)}) = F(t_{(n)}) < 1$ ocorre na amostra, então, o maior tempo observado é censurado. Assim, o teste proposto para $H_0 : 1 - S(t_{(n)}) = F(t_{(n)}) = 1$ é:

Rejeita-se H_0 se $F(t_{(n)}) < \delta_{0,05}$, em que $\delta_{0,05}$ é o 5º percentil da distribuição $F(t_{(n)})$ calculado sob H_0 , tal que $P(F(t_{(n)}) < \delta_{0,05}) \leq 0,05$.

Para obter $\delta_{0,05}$ é necessário conhecer a distribuição de $F(t_{(n)})$, quando $F(t_{(n)}) = 1$. Como a distribuição é desconhecida, esta pode ser obtida via simulação, discutida por Maller & Zhou (1996) no Capítulo 4 e organizada em tabelas, estruturadas para critério de decisão, reproduzidas no Anexo 1.

Procedimento para testar a proporção de imunes:

1. Estimar $F(t_{(n)}) = 1 - S(t_{(n)})$ pelo método KM;
2. Estimar a proporção de observações censuradas;
3. Observar nas Tabelas A1 e A2 (Anexo 1) a porcentagem de pontos a qual corresponde o valor mais próximo de n (tamanho da amostra) e também a proporção $\hat{k}$ mais próxima de censura, a partir de

$$\hat{k} \approx \frac{\hat{\phi}}{\hat{\phi} + \hat{\mu}}$$

para obter o valor de $\hat{\mu}$.

4. Uma vez obtidas as informações, $\hat{F}(t_{(n)})$ é comparado com o percentual de pontos da tabela. Se $\hat{F}(t_{(n)})$ for menor que o valor observado na tabela, então, rejeita-se H_0 .

Ao rejeitar $H_0 : S(t_{(n)}) = 0$, o segundo passo é testar $H_0 : t_{F_0} \leq t_G$.

Sejam,

$$t_{(n)} = \max(t_i) \text{ e } t_{(n)}^* = \max(t_i : t_i = t_i^*), \quad 1 \leq i \leq n,$$

e suponha que $n \rightarrow \infty$, então,

$$t_{(n)} - t_{(n)}^* = \begin{cases} t_G - t_{F_0}, & \text{se } t_{F_0} \leq t_G \\ 0, & \text{se } t_{F_0} > t_G \end{cases}.$$

Se a diferença $t_G - t_{F_0}$ for grande, então, H_0 não é rejeitado, isto é, o tempo de acompanhamento é suficiente. Dessa forma, o teste proposto por Maller & Zhou (1996) é definido a seguir.

Seja N_n , o número de t_i 's não censurados contidos no intervalo $(2t_{(n)}^* - t_{(n)}, t_{(n)}^*]$, sendo a normalização de N_n expresso por,

$$q_n = \frac{N_n}{n},$$

em que, q_n é proporção de tempos não censurados no intervalo $(2t_{(n)}^* - t_{(n)}, t_{(n)}^*]$.

Então, não rejeita-se $H_0 : S(t_{(n)}) = 0$ se q_n exceder o valor crítico, isto é, têm-se indícios de que o tempo de acompanhamento foi suficiente.

Entretanto, para tomar uma decisão é necessário obter a distribuição de q_n , também desconhecida. Como no teste anterior, foram gerados valores para q_n por meio de simulações, melhor detalhado por Maller & Zhou (1996), e organizado em 8 tabelas disponíveis no Anexo 2.

Regra de decisão do teste Considere $c_{1-\alpha;B}$, em que $1 - \alpha$ é o quantil da distribuição da censura simulado, $U[0, B]$, e B limite superior da distribuição uniforme. Logo,

- se $q_n > c_{0,95;8}$, então, o tempo de acompanhamento foi suficiente;
- se $q_n < c_{0,05;2}$, então, o tempo de acompanhamento não foi suficiente;
- se $c_{0,10;6} < q_n < c_{0,90;4}$, então, o teste é inconclusivo.

Testado a presença de imunes na população, o próximo passo é obter os efeitos das covariáveis no tempo de sobrevivência, bem como, obter os estimadores de máxima verossimilhança dos parâmetros da distribuição associados ao tempo de falha

dos indivíduos suscetíveis. Logo, na sessão a seguir será apresentado o modelo de regressão com fração de cura.

3 MODELO DE REGRESSÃO COM FRAÇÃO DE CURA

3.1 Formulação

Os modelos de fração de cura (FC) são caracterizados pela presença de indivíduos suscetíveis e imunes ao evento de interesse prefixado. A formulação de um modelo de regressão, ao qual possa incluir um conjunto de covariáveis $\mathbf{X}$, é essencial porque permitirá traçar um perfil para os indivíduos em relação ao risco de falhar.

A fim de facilitar questões computacionais na estimação dos parâmetro do modelo ajustado, Maller & Zhou (1996) definiram uma reparametrização de π , proporção de imunes, e dos parâmetros da distribuição ajustada ao grupo de suscetíveis. Como π pertence ao intervalo $[0, 1]$, utilizando a função de ligação logito, pode-se relacionar π com as variáveis preditoras $\mathbf{X}$:

$$\pi(\boldsymbol{\beta}, \mathbf{X}) = \frac{e^{\boldsymbol{\beta}' \mathbf{X}}}{1 + e^{\boldsymbol{\beta}' \mathbf{X}}} \quad (4)$$

$$1 - \pi(\boldsymbol{\beta}, \mathbf{X}) = \frac{1}{1 + e^{\boldsymbol{\beta}' \mathbf{X}}}$$

em que, $\boldsymbol{\beta}' = (\beta_0 \ \beta_1 \ \dots \ \beta_k)_{1 \times n}$ é o vetor de parâmetros que representa os efeitos das k covariáveis e $\mathbf{X} = (x_1 \ x_2 \ \dots \ x_k)_{n \times k}$ é a matriz de covariáveis.

A partir das relações definidas na Tabela 1 e da Equação 3, pode-se definir as funções abaixo

$$f(t; \boldsymbol{\phi} | \mathbf{X}) = (1 - \pi(\boldsymbol{\beta}, \mathbf{X}))f^*(t; \theta) = \frac{1}{1 + e^{\boldsymbol{\beta}' \mathbf{X}}} f^*(t; \theta), \quad (5)$$

$$\begin{aligned}
S(t; \phi | \mathbf{X}) &= \pi(\beta, \mathbf{X}) + (1 - \pi(\beta, \mathbf{X}))S^*(t; \theta) \\
&= \frac{e^{\beta' \mathbf{X}}}{1 + e^{\beta' \mathbf{X}}} + \frac{1}{1 + e^{\beta' \mathbf{X}}} S^*(t; \theta) \\
&= \frac{e^{\beta' \mathbf{X}} + S^*(t; \theta)}{1 + e^{\beta' \mathbf{X}}}
\end{aligned} \tag{6}$$

e

$$\begin{aligned}
h(t; \phi | \mathbf{X}) &= \frac{f(t; \phi | \mathbf{X})}{S(t; \phi | \mathbf{X})} = \frac{\frac{f^*(t; \theta)}{1 + e^{\beta' \mathbf{X}}}}{\frac{e^{\beta' \mathbf{X}} + S^*(t; \theta)}{1 + e^{\beta' \mathbf{X}}}} \\
&= \frac{f^*(t; \theta)}{e^{\beta' \mathbf{X}} + S^*(t; \theta)},
\end{aligned} \tag{7}$$

em que, $\phi = (\beta, \theta)$ e θ é o vetor de parâmetros da distribuição de probabilidade dos indivíduos não imunes ao evento de interesse. $f^*(t; \theta)$, $S^*(t; \theta)$ e $h^*(t; \theta)$ são funções associadas à variável aleatória tempo de falha T dos indivíduos suscetíveis, que deve seguir uma distribuição de probabilidade à qual, seus valores sejam positivos e a distribuição assimétrica à direita.

Alguns exemplos de modelos de longa duração, utilizando algumas distribuições apresentadas na Seção 2.4 e das Equações 5, 6 e 7, pode ser encontrado no Apêndice B ?? deste trabalho.

3.2 Estimação dos parâmetros do modelo

A estimação de máxima verossimilhança é a mais popular dentre os métodos de estimação dos parâmetros desconhecidos de modelos paramétricos por ter boas propriedades. A função de verossimilhança deve ser maximizada sobre o espaço paramétrico a fim de encontrar o estimador de máxima verossimilhança dos parâmetros.

Diferentemente do método de estimação tradicional, para modelos de sobrevivência com fração de cura, deve-se ser levado em conta as variáveis indicadoras de,

- falha δ_i

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

e

- suscetibilidade c_i

$$c_i = \begin{cases} 0, & \text{se a observação é imune} \\ 1, & \text{se a observação é suscetível} \end{cases},$$

em que $i = 1, \dots, n$.

Assim, a função de verossimilhança pode ser definida como:

$$\begin{aligned} L(\phi) &= \prod_{i=1}^n f(t_i; \phi | \mathbf{X}_i)^{\delta_i} S(t_i; \phi | \mathbf{X}_i)^{1-\delta_i} \\ &= \prod_{i=1}^n [(1 - \pi(\beta, \mathbf{X}_i)) f^*(t_i; \theta)]^{\delta_i} \\ &\quad \times [\{\pi(\beta, \mathbf{X}_i)\}^{1-c_i} + \{(1 - \pi(\beta, \mathbf{X}_i)) S^*(t_i; \theta)\}^{c_i}]^{1-\delta_i}, \end{aligned}$$

em que, $\phi = (\beta, \theta)$, θ é o vetor de parâmetros da distribuição de probabilidade dos indivíduos não imunes ao evento de interesse e β é o vetor de parâmetros que representa os efeitos das k covariáveis.

A obtenção dos estimadores de máxima verossimilhança é mais simples de se obter por meio do logaritmo natural de função de verossimilhança, denotado por:

$$\begin{aligned} l(\phi) &= \sum_{i=1}^n \delta_i \ln[1 - \pi(\beta, \mathbf{X}_i)] + \sum_{i=1}^n \delta_i \ln[f^*(t_i; \theta)] \\ &\quad + \sum_{i=1}^n (1 - \delta_i) \ln [\{\pi(\beta, \mathbf{X}_i)\}^{1-c_i} + \{(1 - \pi(\beta, \mathbf{X}_i)) S^*(t_i; \theta)\}^{c_i}]. \end{aligned} \quad (8)$$

Note que, se uma observação tiver sua respectiva variável indicadora $\delta_i = 1$, então, necessariamente $c_i = 1$, ou seja, o indivíduo que falhar é suscetível ao evento de interesse, por definição. Entretanto, se $\delta_i = 0$, o indivíduo pode ser imune ou suscetível, logo, tem-se um problema de dados incompletos. Nestes casos especiais, as estimativas de máxima verossimilhança de $\phi = (\beta, \theta)$ podem ser obtidas por meio

do algoritmo EM, pois, segundo Dempster et al. (1976), o algoritmo se destaca pela simplicidade e generalidade da teoria associada.

3.2.1 Algoritmo EM

O algoritmo *Expectation-Maximization* (EM) (Dempster et al., 1976) é uma abordagem amplamente aplicada ao processo iterativo a fim de obter as estimativas de máxima verossimilhança. Este método é utilizado, principalmente, em problemas envolvendo dados incompletos, que para algoritmos como o método de Newton-Raphson podem ser mais complicados (McLachlan & Krishnan, 2008).

Considere que $t_1, t_2, \dots, t_m$ representam o tempo de sobrevivência dos indivíduos que falharam, $t_{m+1}, \dots, t_n$ os tempos censurados e $\mathbf{X}_j$, a matriz de covariáveis para o j -ésimo indivíduo. Então, a variável aleatória c é igual a 1 para os primeiros m indivíduos e, desconhecida para os restantes $n - m$ indivíduos, pois como foram censurados não se sabe o real desfecho deles. Assim, essas n observações são tratadas como perdidas.

A função de log-verossimilhança apresentada na Equação 8 passa a ser expressa por:

$$\begin{aligned} l(\phi) = & \sum_{i=1}^m \delta_i \ln[1 - \pi(\beta, \mathbf{X}_i)] + \sum_{i=m+1}^n \delta_i \ln[1 - \pi(\beta, \mathbf{X}_i)] \\ & + \sum_{i=1}^m \delta_i \ln[f^*(t_i; \theta)] + \sum_{i=m+1}^n \delta_i \ln[f^*(t_i; \theta)] \\ & + \sum_{i=1}^m (1 - \delta_i) \ln [\{\pi(\beta, \mathbf{X}_i)\}^{1-c_i} + \{(1 - \pi(\beta, \mathbf{X}_i))S^*(t_i; \theta)\}^{c_i}] \\ & + \sum_{i=m+1}^n (1 - \delta_i) \ln [\{\pi(\beta, \mathbf{X}_i)\}^{1-c_i} + \{(1 - \pi(\beta, \mathbf{X}_i))S^*(t_i; \theta)\}^{c_i}]. \end{aligned}$$

O EM consiste em dois passos. No **passo E**, calcula-se a função pseudo log-verossimilhança com base nas observações faltantes. Assim, para um tempo de censura t , é construído duas “pseudo-observações” parcialmente completas da forma

$(t; w_1(\mathbf{X}; t))$ e $(t; w_2(\mathbf{x}; t))$. Especificamente, $w_1(\mathbf{x}; t)$ e $w_2(\mathbf{x}; t)$ denotam a probabilidade condicional do indivíduo pertencer ao grupo imune ou suscetível, respectivamente, dada a sobrevivência até o tempo t (Kannan et al., 2010), ou seja,

$$\begin{aligned} w_1(\mathbf{X}, t, \phi) &= P(c_i = 0 | T > t) \\ &= \frac{P(c_i = 0) \times P(T > t | c_i = 0)}{P(T > t)} = \frac{\pi(\boldsymbol{\beta}, \mathbf{X}) \times 1}{S(t; \phi | \mathbf{X})} \\ &= \frac{\pi(\boldsymbol{\beta}, \mathbf{X})}{\pi(\boldsymbol{\beta}, \mathbf{X}) + (1 - \pi(\boldsymbol{\beta}, \mathbf{X}))S^*(t; \theta)} \end{aligned}$$

e

$$\begin{aligned} w_2(\mathbf{X}, t, \phi) &= P(c_i = 1 | T > t) \\ &= \frac{P(c_i = 1) \times P(T > t | c_i = 1)}{P(T > t)} = \frac{(1 - \pi(\boldsymbol{\beta}, \mathbf{X})) \times S^*(t; \theta)}{S(t; \phi | \mathbf{X})} \\ &= \frac{(1 - \pi(\boldsymbol{\beta}, \mathbf{X}))S^*(t; \theta)}{\pi(\boldsymbol{\beta}, \mathbf{X}) + (1 - \pi(\boldsymbol{\beta}, \mathbf{X}))S^*(t; \theta)} \end{aligned}$$

em que, $\phi = (\boldsymbol{\beta}, \theta)$.

Em Kannan et al. (2010), tem-se que a pseudo log-verossimilhança, definida como a esperança do logaritmo da verossimilhança dos dados completos, considerando as observações perdidas, é denotada por:

$$l_{pseudo}(\boldsymbol{\beta}, \theta) = g_1(\boldsymbol{\beta}) + g_2(\theta),$$

tal que,

$$g_1(\boldsymbol{\beta}) = \sum_{i=1}^m \ln(1 - \pi(\boldsymbol{\beta}, \mathbf{X}_i)) + \sum_{i=m+1}^n w_1 \ln(\pi(\boldsymbol{\beta}, \mathbf{X}_i)) + \sum_{i=m+1}^n w_2 \ln(1 - \pi(\boldsymbol{\beta}, \mathbf{X}_i))$$

e

$$g_2(\theta) = \sum_{i=1}^m \ln f^*(t_i; \theta) + \sum_{i=m+1}^n w_2 \ln S^*(t_i; \theta).$$

O **passo M** consiste em maximizar $l_{pseudo}(\boldsymbol{\beta}, \theta)$ em relação aos parâmetros desconhecidos, dado $w_1(\mathbf{X}; t)$ e $w_2(\mathbf{X}; t)$. Para isso um artifício simples é maximizar as funções $g_1(\boldsymbol{\beta})$ e $g_2(\theta)$ separadamente, sendo que a primeira obtém as estimativas de $\boldsymbol{\beta}$, vetor de efeito das covariáveis, e a segunda obtém θ , vetor de parâmetros da distribuição associada ao tempo de falha dos indivíduos suscetíveis.

Para k iterações, $k > 1$, a maximização de $g_1(\boldsymbol{\beta})$ e $g_2(\theta)$ no passo $k + 1$, requer, a partir de $\boldsymbol{\phi}^{(k)}$ e $\theta^{(k)}$, que $w_1(\mathbf{X}, t, \phi)^{(i)}$ e $w_2(\mathbf{X}, t, \phi)^{(i)}$ sejam expressos por:

$$w_1(\mathbf{X}, t, \phi)^{(k+1)} = \frac{\pi(\boldsymbol{\beta}^{(k)}, \mathbf{x})}{\pi(\boldsymbol{\beta}^{(k)}, \mathbf{X}) + (1 - \pi(\boldsymbol{\beta}^{(k)}, \mathbf{X}))S^*(t; \theta^{(k)})}$$

e

$$w_2(\mathbf{X}, t, \phi)^{(k+1)} = \frac{(1 - \pi(\boldsymbol{\beta}^{(k)}, \mathbf{X}))S^*(t; \theta^{(k)})}{\pi(\boldsymbol{\beta}^{(k)}, \mathbf{X}) + (1 - \pi(\boldsymbol{\beta}^{(k)}, \mathbf{X}))S^*(t; \theta^{(k)})}.$$

As estimativas de máxima verossimilhança de $\hat{\boldsymbol{\beta}}$ e $\hat{\theta}$ são obtidas quando um critério de parada for satisfeito, isto é, quando $|\hat{\boldsymbol{\phi}}^k - \hat{\boldsymbol{\phi}}^{k-1}| = e \approx 0$, por exemplo, $e = 10^{-100}$.

A descrição do algoritmo neste trabalho tomou como base os trabalhos de Kan-nan et al. (2010) e Silva (2015a). Existem outros diversos trabalhos que tratam deste tema, dentre eles, Balakrishnan & Pal (2016), Rodrigues et al. (2008) e Dempster et al. (1976). Aos interessados em se aprofundar, McLachlan & Krishnan (2008) expõem o tema de forma mais completa em seu livro, além de abordar outros algoritmos para maximizar uma função.

3.2.2 Implementação no R (R Core Team, 2017)

Quando a pesquisa começou a ser desenvolvida, não havia pacotes publicados no R que ajustassem modelos paramétricos com fração de cura. Entretanto, agora pode-se encontrar o pacote *flexsurvcure* (Amdahl, 2017), que não está muito bem estruturado, e o pacote não publicado, desenvolvido por Silva (2015b), que está desatualizado, não sendo possível executar. Por isso, houve a necessidade de implementação do modelo.

A implementação do algoritmo do modelo de regressão, tomando como base o *script* desenvolvido por Silva (2015a), que contribuiu para o ajuste do modelo para mais de uma variável regressora e para o desenvolvimento de uma pseudo-generalização do algoritmo, sendo possível ajustar modelos com quaisquer distribuições paramétricas, desde que tenha sido implementada no R ou que as funções de densidade e acumulada tenham forma analítica.

Função *f.cura*

A função *f.cura*, disponibilizada no Apêndice, retorna ao usuário as estimativas dos parâmetros da distribuição associada ao tempo de falha dos suscetíveis, o valor da log-verossimilhança e uma tabela com as estimativas dos efeitos das covariáveis, erro padrão das estimativas, o intervalo de confiança de 95%, significância (valor-p) e as razões de chances (para interpretação dos efeitos das covariáveis).

Os argumentos necessários para utilizar a função são:

- Os valores iniciais para β :

$$\beta_0 = (\beta_{01} \ \beta_{02} \ \dots \ \beta_{0n}) = Mdia \left[\frac{e^{\mathbf{X}}}{1 + e^{\mathbf{X}}} \right], \quad (9)$$

resultando em um vetor de tamanho igual a quantidade de covariáveis no modelo. Outros valores poderão ser obtidos a partir da soma ou multiplicação por escalar do resultado obtido pela Equação 9.

- Os valores iniciais para θ_0 : o chute inicial para todos os parâmetros da distribuição será 0.1, para quando $\theta_i > 0, i = 1, \dots, k$, em caso de outras restrições (pertencer aos reais ou a algum intervalo), sugere-se estudar a curva da função de risco ou também tenta o estimador de máxima verossimilhança para cada parâmetro.

- A matriz de covariáveis $\mathbf{X}$: necessidade de criação de variáveis “dummy” quando a variável é categórica.

- O tempo de acompanhamento.

- A variável indicadora de falha.

- A distribuição para explicar o tempo de sobrevivência dos indivíduos suscetíveis.

As distribuições disponíveis *f.cura*: Gama Generalizada, Exponencial, Weibull, Log-normal, Gama, Weibull Exponencializada, Weibull Modificada e Weibull Estendida. Na necessidade de alguma distribuição específica, a inclusão de novas distribuições é bem simples, desde que a ela esteja implementada ou tenha uma equação analítica.

Para facilitar a programação do algoritmo de otimização, Silva (2015a) enumerou os seguintes passos, que são os mesmos utilizados na função *f.cura*.

1. (Passo E): A partir de $\phi^{(k-1)} = (\beta^{(k-1)}, \theta^{(k-1)})$, obter $w_1(x)^{(k)}$ e $w_2(x)^{(k)}$ e

calcular $Q(\phi, \phi^{(k)})$

2. (Passo M): Atualizar os parâmetros β e θ . $\beta^{(k)}$ é obtido através da maximização de $g_1(\beta)$ com respeito a β , enquanto $\theta^{(k)}$ é obtido através da maximização de $g_2(\theta)$ com respeito a θ .
3. Declarar convergência se um critério de parada for atingido.

3.3 Teste paramétrico para presença de imunes

Como no teste não paramétrico, tem-se interesse em testar $H_0 : p = 0$ ou $H_0 : q = 1 - p = 1$, em que p é a proporção de imunes e q , a proporção de suscetíveis.

Procedimento para testar a proporção de imunes:

1. Obter $\hat{q}$ e as estimativas dos parâmetros $\hat{\theta}$ da distribuição associado ao tempo de sobrevivência dos suscetíveis;
2. Calcular

$$d_n = 2(\ln(\hat{\theta}_n) - \ln(\hat{\theta}_{H_0})),$$

em que $\ln(\hat{\theta}_n)$ é a log-verossimilhança resultante do modelo de fração de cura e $\ln(\hat{\theta}_{H_0})$, a log-verossimilhança sob $H_0 : q = 1$, ou seja, supondo não haver imunes na população.

A distribuição de d_n é uma mistura 50 – 50 de uma variável aleatória Qui-Quadrado com 1 grau de liberdade e massa pontual de 1. Assim, o percentil 95, $c_{0,95}$, da distribuição de d_n é dado por:

$$\frac{1}{2} + \frac{1}{2}P(X_1^2 \leq c_{0,95}) = 0,95.$$

Maller & Zhou (1996) definem que $c_{0,95} = 2,71$, bastando tomar a decisão após obter d_n

3. Tomar uma decisão baseado na seguinte regra:

- se $d_n > c_{0,95}$, então, rejeita-se H_0 , ao nível de significância de 0,05, considerando haver fortes evidências de que $q < 1$, ou seja, existem imunes na população.

3.4 Seleção de modelo

3.4.1 Teste de Wald

O teste de Wald é utilizado para averiguar quais covariáveis possuem efeitos significativos na estimação da variável resposta.

$$\text{As hipóteses a serem testadas são: } \begin{cases} H_0 : \beta_i = 0 \\ H_1 : \beta_i \neq 0 \end{cases},$$

em que $i = 1, 2, \dots, k$, representando as covariáveis a serem testadas.

A estatística do teste de Wald é dada por:

$$z_i = \frac{\hat{\beta}_i}{\sqrt{\text{var}(\hat{\beta}_i)}},$$

em que, $\sqrt{\text{var}(\hat{\beta}_i)}$ é o erro padrão da estimativa do parâmetro β_i .

A estatística z possui uma distribuição aproximadamente normal padrão quando $\beta_i = 0$. De maneira equivalente, a estatística do teste de Wald também pode ser escrita como:

$$z^2 = \frac{\hat{\beta}_i^2}{\text{var}(\hat{\beta}_i)},$$

em que, $\text{var}(\hat{\beta}_i)$ é a estimativa da variância do estimador do parâmetro β_i .

A estatística z^2 possui uma distribuição aproximadamente qui-quadrado com 1 grau de liberdade.

Para critério de decisão para a hipótese nula (H_0), calcula-se o valor-p da estatística do teste, que é equivalente a probabilidade de se obter um valor mais extremo do que a estatística do teste z , ou seja:

$$\text{valor} - p = P(|Z| > z)$$

Considerando um nível de significância α previamente definido, obtém-se que se o *valor - p* for maior que α , então, não há evidências suficientes para rejeitar a hipótese nula e a covariável x_i não deve ser considerada no modelo. Caso contrário, se o *valor - p* for menor q α , então rejeita-se a hipótese nula e a covariável x_i deve ser considerada no modelo.

3.4.2 Testa da razão de verossimilhanças

O Teste da razão de verossimilhanças (TRV) tem o mesmo objetivo que o teste de Wald, entretanto, o TRV é indicado para quando se tem modelos encaixados.

As hipóteses a serem testadas são:

$$\begin{cases} H_0 : \text{O modelo mais simples é o mais adequado} \\ H_1 : \text{O modelo completo é o mais adequado} \end{cases}$$

A estatística do teste é dada por:

$$TRV = 2[\log(L_s) - \log(L_c)], \quad (10)$$

em que L_s é a função de verossimilhança do modelo mais simples, que possui menos parâmetros, e L_c é a função de verossimilhança do modelo completo.

A estatística TRV possui uma distribuição aproximadamente qui-quadrada com $g_c - g_s$ graus de liberdade, sendo g_c é o número de covariáveis do modelo completo e g_s é o número de covariáveis do modelo simples.

Para tomar uma decisão, calcula-se o p-valor da estatística do teste, que é equivalente a probabilidade de se obter um valor mais extremo do que a estatística do teste TRV , ou seja:

$$valor - p = P(TRV > \chi^2_{(g_c - g_s)}) \quad (11)$$

Para um nível de significância α previamente definido, obtém-se que se o valor-p for maior que α , então, não há evidências suficientes para rejeitar a hipótese nula

e o modelo mais simples deve ser utilizado. Caso contrário, se o valor-p for menor que α), rejeita-se a hipótese nula e o modelo mais completo deve ser utilizado.

3.4.3 AIC e BIC

Outros métodos que auxiliam na seleção de modelos são os Critério de Informação de Akaike (AIC) e Critério de Informação Bayesiano (BIC) definidos, respectivamente, como:

$$AIC_p = -2\log(L_p) + 2[(p + 1) + 1]$$

e

$$BIC_p = -2\log(L_p) + [(p + 1) + 1]\log(n),$$

em que L_p é a função de verossimilhança do modelo e p é o número de covariáveis ajustadas no modelo.

O principal objetivo destes critérios é selecionar o modelo que mais se aproxima do verdadeiro. Logo, o modelo mais apropriado será aquele que apresentar menor AIC e BIC.

3.5 Diagnóstico

3.5.1 Análise de resíduos

Uma etapa imprescindível após o ajuste do modelo é a análise de resíduos, que verifica as suposições dos erros, a qualidade do ajuste e a presença de valores influentes.

Resíduo Cox-Snell

O resíduo de Cox-Snell tem como objetivo avaliar a qualidade do ajuste do modelo, sendo definido por:

$$e_i = \hat{H}(t_i|\mathbf{x}), i = 1, 2, \dots, n$$

em que $\hat{H}$ é a função de risco acumulada obtida do modelo ajustado. Se o modelo for adequado, então os resíduos e_i vem de uma população homogênea e devem seguir uma distribuição exponencial. Portanto o gráfico de e_i versus $\hat{H}(e_i)$ deve ser aproximadamente uma reta caso o modelo seja adequado.

Como $\hat{H}(e_i) = -\log(\hat{S}(e_i))$, o gráfico das curvas de sobrevivência desses resíduos, obtidas por Kaplan-Meier e pelo modelo exponencial padrão, também auxiliam na verificação da qualidade do modelo ajustado, o gráfico $\exp(e_i)$ versus $\hat{S}(e_i)$.

Resíduo Martingal

O resíduo martingal foi proposto inicialmente para processos de contagem e foi adaptado para modelos de sobrevivência com presença de censura e é escrito na seguinte forma:

$$r_{Mi} = \delta_i - \hat{H}(y_i|\mathbf{x}),$$

em que δ_i é a indicadora de falha e $\hat{H}(y_i|\mathbf{x})$ é a função de risco acumulada do modelo ajustado. O resíduo martingal é assimétrico com valor máximo 1 e mínimo em $-\infty$ e, portanto não é simetricamente distribuído em torno de zero.

Resíduo Deviance

O resíduo deviance foi proposto como uma tentativa de tornar os resíduos martingal mais simétricos em torno de zero e, desta maneira, facilitar a detecção de pontos discrepantes. O resíduo deviance é dado por:

$$r_{Di} = \text{sign}(r_{Mi})[-2(r_{Mi} + \delta_i \log(\delta_i - r_{Mi}))]^{1/2}.$$

Gráficos dos resíduos martingal, ou deviance, contra os valores ajustados ou o índice das observações fornecem uma maneira de verificar a adequação do modelo ajustado, bem como a presença de observações atípicas.

3.5.2 Teste de Kolmogorov-Smirnov

A qualidade do ajuste será testada, ao comparar as sobrevivências teórica e empírica, isto é, tem-se interesse em testar a hipótese $F_X = F$ contra a hipótese alternativa $F_X \neq F$, tal que, $F(x)$ representa a função distribuição acumulada proposta aos dados e $F_n(x) = 1 - \hat{S}_{KM}$, a função distribuição acumulada empírica dos dados.

A estatística do teste é definida como:

$$D_n = |\hat{S}_{KM}(t_{(n)}^*) - \hat{S}_{proposto}(t_{(n)}^*)|,$$

em que $t_{(n)}^*$ é o maior tempo de falha observado de um indivíduo suscetível.

O modelo proposto é rejeitado, se $D_n > q_{95\%}$, sendo $q_{95\%}$ o quantil do teste de Kolmogorov-Smirnov.

3.5.3 Teste para bondade do ajuste

A proposta deste teste é utilizar o coeficiente de correlação r para testar a bondade do ajuste. Como a distribuição teórica de r não é conhecida, Maller & Zhou (1996) organizaram os quantis simulados em 9 tabelas, disponível no Anexo 3, e uma tabela mais resumida reproduzida na Tabela 2.

A hipótese a ser testada é denotada por:

$$H_0 : F = \bar{F},$$

em que, $\bar{F}$ é a função distribuição acumulada do modelo proposto. Seja, $\hat{F}$ uma função distribuição empírica da amostra $t_1, t_2, \dots, t_n$, então, ao tomar $\hat{F}$ como EKM, então, o gráfico de $\bar{F}(t_i)$ contra $\hat{F}(t_i)$ é uma reta com inclinação igual a 1.

A regra de decisão, sob as condições apresentadas nas tabelas da distribuição dos quantis de r , é definida como:

- Se $r > percentil80$, então, o modelo proposto está bem ajustado;
- Se $r > percentil10$, então, o modelo tem ajuste razoável.

Tabela 2: Percentis do coeficiente de correlação.

		Distribuição da censura			
		$exp(1/\mu)$		$U[0, B]$	
	1-p	$\mu = 1$	$\mu = 2$	$B = 1$	$B = 2$
80%	0,4	0,973	0,984	0,98	0,987
	0,6	0,982	0,989	0,986	0,991
	0,8	0,986	0,992	0,989	0,993
	1	0,99	0,994	0,991	0,994
5%	0,4	0,848	0,914	0,874	0,926
	0,6	0,904	0,942	0,925	0,956
	0,8	0,933	0,966	0,948	0,969
	1	0,957	0,974	0,964	0,975

4 APLICAÇÃO

4.1 Entendendo a doença

A cirrose é uma doença hepática assintomática que, muitas vezes, é descoberta quando o quadro é irreversível. A cura somente é possível por meio de um transplante de fígado. Por isso, o tratamento consiste em uma série de medidas para controlar o avanço da enfermidade, visto que a principal consequência da cirrose é o aumento da pressão na veia portal, que por sua vez, acarreta no surgimento de varizes, resultando quase como certo no seu respectivo rompimento, podendo ser fatal. A pior complicação da ruptura de varizes é a hemorragia digestiva alta (HDA) e da HDA, o óbito.

A gravidade da doença hepática é avaliada a partir das seguintes características clínicas: nível de bilirrubina total, albumina sérica, tempo de protrombina, grau de ascite e grau de encefalopatia hepática. Tais características resultam em um escore que determinam a classe Child-Pugh do paciente, descrita inicialmente por Child (1964) e adaptada anos depois por Pugh et al. (1973).

Pugh et al. (1973) classificaram os pacientes hepáticos como: baixo risco (grau A: 5 ou 6 pontos) de óbito em intervenções cirúrgicas; moderado (grau B: com 7 a 9 pontos) e alto risco (grau C: com 10-15 pontos). O escore Child-Pugh sofreu algumas modificações. Atualmente é amplamente utilizado para avaliar a gravidade da insuficiência hepática, sendo classificado em grau A, B ou C que representam os estágios inicial, moderado ou avançado da doença, respectivamente.

Durante o curso da doença cerca de 90% dos pacientes desenvolvem varizes esofágicas (Biecker, 2013), sendo que os pacientes em estágio inicial (Child A) apre-

sentam varizes em aproximadamente 30% dos casos, enquanto os em estágio moderado ou avançado (Child B e C) da doença em até 60% (Bittencourt et al., 2010). Além disso, os indivíduos podem apresentar novas varizes ou até mesmo o aumento do calibre da veia, sendo importante o acompanhamento médico constante.

Os tratamentos podem ser classificados em duas categorias:

- 1) aqueles utilizados para diminuir a pressão do portal, como medicação (por exemplo, β -Bloqueadores não seletivos), intervenção radiológica ou abordagem cirúrgica;
- 2) os utilizados para obstruir as varizes, como a endoscopia (ligadura elástica de varizes esofágicas (LEVE), escleroterapia endoscópica de varizes esofágicas (EEVE) e injeção adesiva de tecido) ou intervenção radiológica (Toshikuni et al., 2016).

A recidiva de varizes costuma ser comum em tratamentos endoscópicos como profilaxia, principalmente em pacientes em estágio avançado de cirrose. Quando há o rompimento das varizes, além dos procedimentos listados, também são incluídos ao tratamento a utilização de antibióticos e a realização de transfusão de sangue.

A LEVE, como mencionado anteriormente, é uma medida para minimizar as adversidades consequentes da hipertensão portal.

A EEVE é também uma alternativa para o tratamento das varizes no esôfago. Entretanto, como o risco de sangramento é maior para essa técnica ela é indicada aos indivíduos em estágio avançado ou numa emergência quando a LEVE não pode ser realizada.

A cirurgia é uma outra estratégia para o tratamento de varizes e utilizada como último recurso devido o alto risco de complicações.

Laine & Cook (1995) destacaram ainda que, comparada à EEVE, a LEVE apresenta menor risco de ressangramento, menor frequência de efeitos adversos, menor número de sessões necessárias para obliteração de varizes e melhor sobrevida (Coelho et al., 2014).

As informações coletadas dos pacientes submetidos à LEVE, são utilizadas para identificar os fatores que aliados ao procedimento garantiram maior sobrevida dos pacientes.

Durante o procedimento LEVE, o paciente é sedado para realização de uma endoscopia a fim de identificar as varizes. A variz que será tratada é aspirada e em seguida, um anel de borracha é colocado, que a comprime, fazendo com que as varizes reduzam até secar. O tratamento é composto por 3 a 6 sessões, repetidas a cada 15 ou 30 dias, até o desaparecimento das varizes.

Na próxima seção, serão apresentadas as variáveis coletadas no estudo e uma análise descritiva que auxilie na modelagem e permita a caracterização dos indivíduos participantes da pesquisa.

4.2 Caracterização dos pacientes

O presente trabalho é continuidade do apresentado por Oliveira et al. (2015). Na pesquisa acompanhou-se 129 pacientes cirróticos, que necessitavam realizar um procedimento para evitar a ruptura de varizes no esôfago ou conter uma hemorragia, desde o procedimento LEVE até o óbito devido à doença. Aqueles que não experienciaram o evento de interesse (óbito pela doença) ou tiveram que sair do estudo, os tempos de seguimento foram censurados.

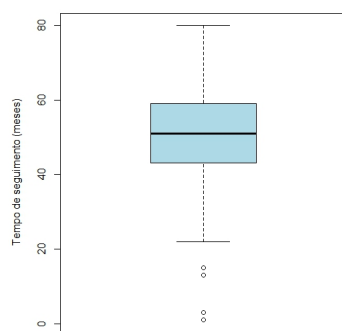
O atendimento aos pacientes foi realizado no Hospital das Clínicas da Faculdade de Medicina da UNESP, Campus Botucatu (SP), entre 2006 e 2010, totalizando em 47 meses de tempo de acompanhamento.

As variáveis observadas neste estudo estão descritas na Tabela 3, bem como, algumas medidas resumo para as variáveis quantitativas e as frequências das classes, para as qualitativas. Com isso, pode-se traçar o perfil dos indivíduos pertencentes a amostra, que será descrito a seguir.

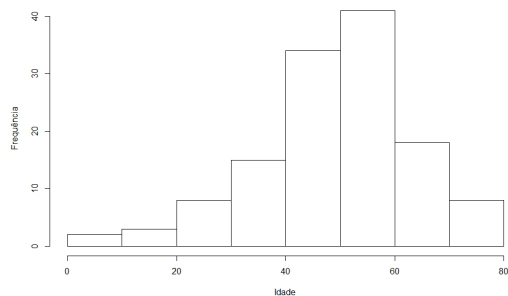
A idade dos pacientes no início do estudo varia de 1 a 80 anos, sendo que metade tem mais que 51 anos e 25% mais do que 59 anos. No histograma, Figura 2(b), pode-se observar que a distribuição da idade é levemente assimétrica à esquerda, isto é, os pacientes são na maior parte mais velhos, 50% apresentam idade entre 43 e 49 anos. Apesar disso, na Figura 2(a), pode-se observar que a idade mesmo tendo amplitude grande, a maior parte dos pacientes têm 40 a 60 anos.

Tabela 3: Medidas resumo das variáveis do estudo.

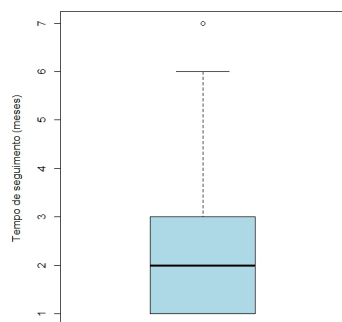
Variáveis	Descrição	Medidas resumo			
Óbito	Variável indicadora de falha	0	= 109 (84,496%)		
	0 = censurada	1	= 20 (15,504%)		
	1 = óbito				
Tseg	Tempo de acompanhamento até o óbito ou censura.	Min.	= 1	1o quartil	= 4
		Max.	= 47	Mediana	= 11
		Média	= 15,05	3o quartil	= 23
		Desvio Pad.	= 13,116		
Idade	Idade dos pacientes no início do acompanhamento.	Min.	= 1	1o quartil	= 43
		Max.	= 80	Mediana	= 51
		Média	= 49,6	3o quartil	= 59
		Desvio Pad.	= 14,4711		
Sessões	Número de sessões que foram necessárias para o tratamento.	Min.	= 1	1o quartil	= 1
		Max.	= 7	Mediana	= 2
		Média	= 2,318	3o quartil	= 3
		Desvio Pad.	= 1,431		
Anéis	Número de anéis de borracha utilizados no tratamento.	Min.	= 1	1o quartil	= 4
		Max.	= 29	Mediana	= 7
		Média	= 8,38	3o quartil	= 12
		Desvio Pad.	= 5,952		
Ressangramento	Reincidência do rompimento de alguma variz.	Não	= 110 (85,271%)		
		Sim	= 19 (14,729%)		
Complicações	Complicações devido ao procedimento.	Não	= 111 (86,046%)		
		Sim	= 18 (13,954%)		
Recidiva	Reaparecimento de varizes no esôfago.	Não	= 102 (79,070%)		
		Sim	= 27 (20,930%)		
Sexo	F = Feminino	F	= 48 (37,209%)		
	M = Masculino	M	= 81 (62,791%)		
Child-Pugh	Gravidade da doença hepática	A	= 67 (51,938%)		
	A = estágio inicial	B	= 35 (27,132%)		
	B = estágio intermediário	C	= 27 (20,930%)		
	C = estágio avançado				
HDA_PRE	Apresentaram Hemorragia	Não	= 24 (18,605%)		
	Digestiva Alta antes do tratamento.	Sim	= 105 (81,395%)		
Beta-Bloqueador	Uso do medicamento	Não	= 42 (32,558%)		
	Beta-Bloqueador	Sim	= 87 (67,442%)		
EEVE	Submissão à Escletoterapia	Não	= 104 (80,620%)		
	Endoscópica de Varizes Esofágicas.	Sim	= 25 (19,380%)		
Cirurgia	Submissão à cirurgia no tratamento de varizes.	Não	= 112 (86,822%)		
		Sim	= 17 (13,178%)		



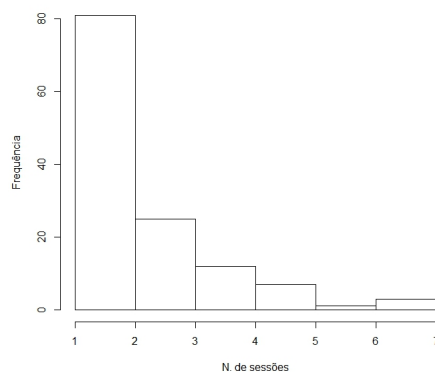
(a) Distribuição da idade



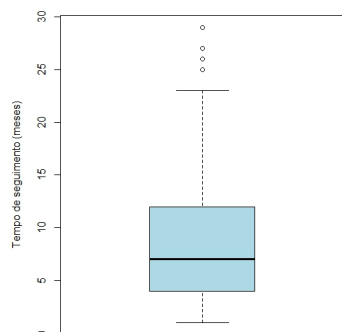
(b) Distribuição da idade



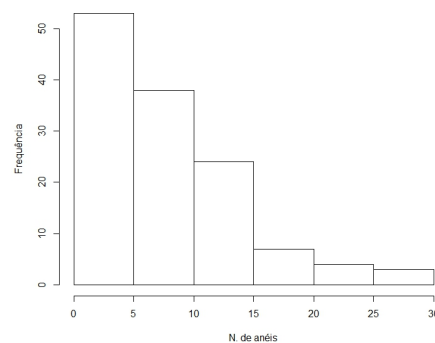
(c) Dist. do n. sessões



(d) Dist. do n. sessões



(e) Dist. do n. anéis



(f) Dist. do n. anéis

Figura 2: Histograma e Boxplot das covariáveis numéricas do estudo.

O número de sessões para o tratamento ligadura variou entre um e sete, sendo que metade dos pacientes necessitaram de mais de duas sessões e um quarto mais que três sessões. O número de sessões está distribuído assimetricamente à direita, o que pode ser observado nas Figuras 2(c) e 2(d).

Os anéis de borracha são importantes para a realização do procedimento da LEVE. O número de anéis utilizados variou de 1 a 29 anéis. Sendo que 50% dos pacientes necessitaram de mais de sete anéis em seu tratamento e 25% mais de 12. Como o número de sessões, a distribuição do número de anéis é assimétrica à direita (Figuras 2(e) e 2(f)).

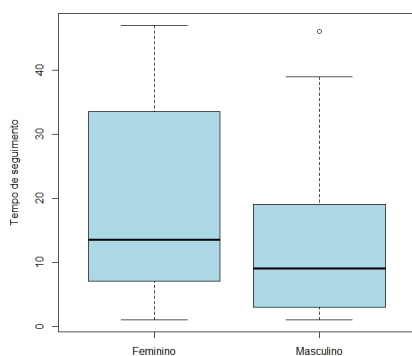
Quanto a variável sexo, 37, 21% são mulheres e 62, 79% homens. Na Figura 3(a), apresenta-se o boxplot para cada categoria para visualizar a existência de relação entre o tempo de seguimento e o sexo. Enquanto que, na Figura 3(b), as curvas de sobrevivência de Kaplan-Meier evidencia maior sobrevivência das mulheres em relação aos homens.

Aproximadamente 52% do pacientes apresentaram grau A de cirrose, enquanto que, 27% e 20% dos pacientes apresentaram grau B e C, respectivamente. Pela Figura 3(c), pode-se notar que existe uma relação entre o tempo de seguimento e o Child-Pugh, isto é, quanto maior a severidade da doença menor o tempo de seguimento. O gráfico das curvas de sobrevivência, Figura 3(d), evidenciam que os pacientes em estágio inicial de cirrose apresentam maior sobrevida que os pacientes em estágio intermediário e avançado.

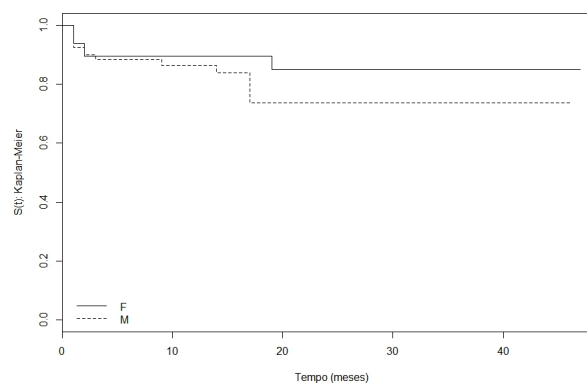
A HDA é uma séria consequência do rompimento das varizes no esôfago e, caso não seja tratado com emergência, o indivíduo pode morrer. No estudo, 81% dos pacientes apresentaram HDA antes do tratamento.

Ao observar a Figura 3(e), nota-se que não parece haver associação entre o tempo de seguimento e o fato do paciente ter tido hemorragia digestiva alta. No entanto, na Figura 3(f), tem-se que inicialmente os pacientes que não tiveram HDA tem uma sobrevida maior e depois esse quadro se inverte.

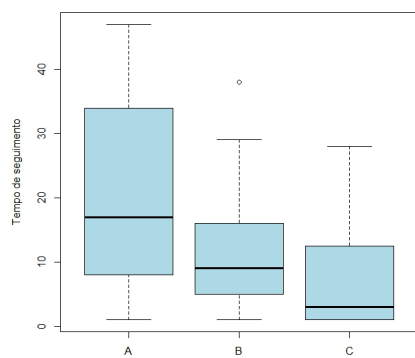
O medicamento β -bloqueador é utilizado como prevenção do sangramento das



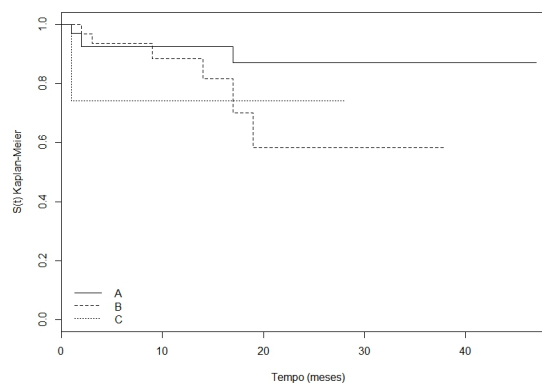
(a) Tempo de seguimento por sexo



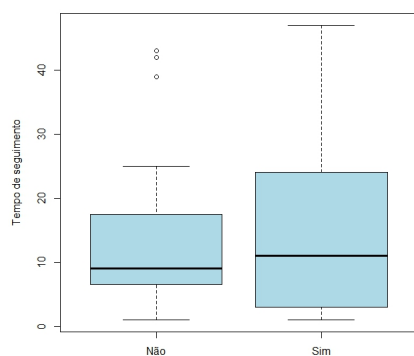
(b) Sobrevivência dos pacientes por sexo



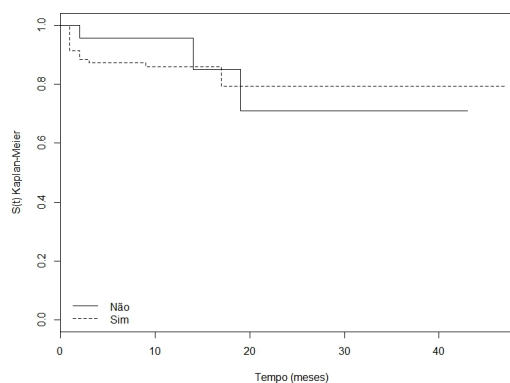
(c) Temp. de seg. X grau de cirrose



(d) Sobrevivência dos pacientes por Child



(e) Tempo de seguimento por HDA



(f) Sobrevivência dos pacientes por HDA

Figura 3: Boxplot e curva de sobrevivência KM para relacionar covariável com o tempo de acompanhamento.

varizes, reduzindo a pressão do sistema portal. Neste estudo, 67% dos pacientes utilizaram esse tipo de medicamento.

Observa-se a partir da Figura 4(a) que existe uma possível relação entre o uso de medicamento β -bloqueador e o tempo de seguimento, sendo que os que não utilizaram apresentaram tempo um pouco menor. Enquanto que, na Figura 4(b), pode-se verificar que os indivíduos que não fizeram uso de medicamentos β -bloqueadores tiveram curva de sobrevivência inferior aos que utilizaram tal medicamento.

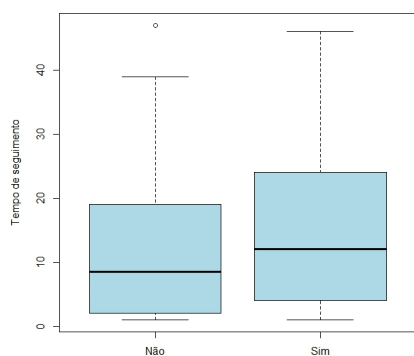
Na Figura 4(c), pode-se observar a existência de uma relação entre o tempo de seguimento e a realização do procedimento EEVE. Neste caso, àqueles que foram submetidos ao tratamento tiveram tempo de seguimento maior do que os que não foram submetidos, fato também observado na Figura 4(d), sendo que, cerca de 19% dos pacientes foram submetidos à EEVE.

No acompanhamento, 13% pacientes necessitaram de cirurgia. Note que, na Figura 4(e), aparentemente existe relação entre os indivíduos que foram submetidos à cirurgia e o tempo de seguimento. A Figura 4(f) reforça essa ideia visto que, os pacientes que foram tratados com cirurgia tiveram maior sobrevida.

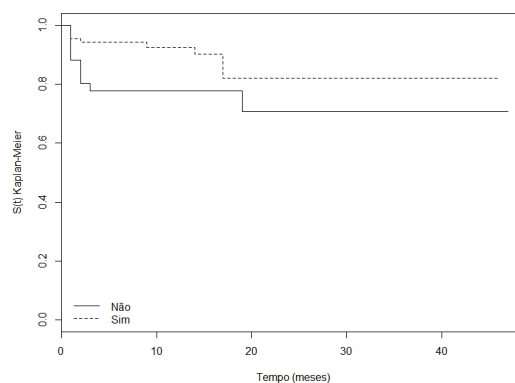
Aproximadamente 85% dos pacientes tiveram ressangramento de alguma variz, sendo o ressangramento a reincidência de rompimento de varizes. Na Figura 5(a), pode ser observada a existência de uma fraca relação entre o tempo de seguimento e o ressangramento das varizes, em que, quando ocorre o ressangramento o tempo de seguimento é inferior. A Figura 5(b) confirma também a conclusão anterior.

Durante o acompanhamento, aproximadamente 86% teve algum tipo de complicação. Estes indivíduos apresentaram tempo de seguimento inferior comparado aos que não tiveram nenhuma complicação, podendo ser observado nas Figura 5(c) e Figura 5(d).

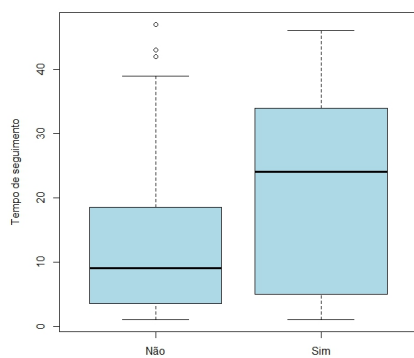
Neste estudo, 21% dos pacientes tiveram recidiva de varizes. Na Figura 5(e) pode ser observado que esses pacientes com varizes reincidentes tiveram tempo de seguimento maior. Enquanto que, nas curvas de sobrevivência KM (Figura 5(f)), observa-se que pacientes que tiveram recidiva de varizes apresentaram maior sobre-



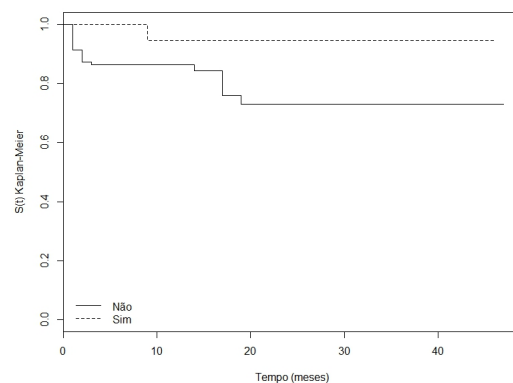
(a) T. de seg. por uso de B.Bloq



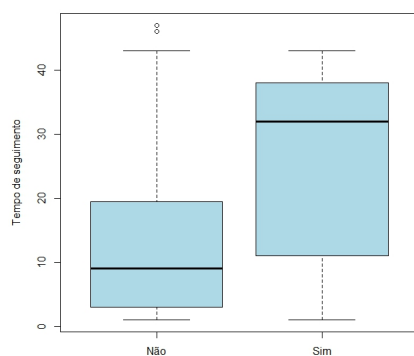
(b) Sobrevivência dos pacientes por B.Bloq



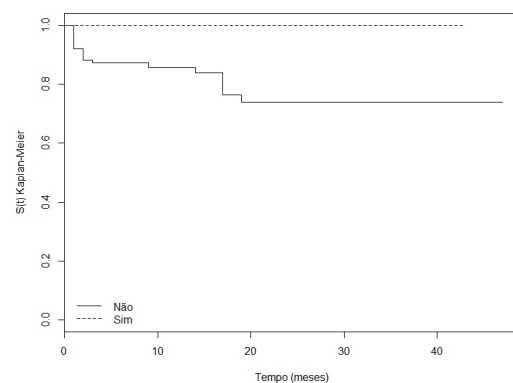
(c) T. de seg. por EEVE



(d) Sobrevivência dos pacientes por EEVE



(e) T. de seg. por cirurgia



(f) Sobrevivência dos pacientes por cirurgia

Figura 4: Boxplot e curva de sobrevivência KM para relacionar covariável com o tempo de acompanhamento.

vida.

Apresentado as características da amostra e, observado possíveis influências individuais das covariáveis no tempo de sobrevivência do paciente após serem submetidos à LEVE, em seguida, tem-se interesse de identificar as possíveis variáveis explicativas que poderão compor o modelo de regressão.

4.3 Seleção das covariáveis para o modelo de regressão

A seleção prévia de covariáveis para compor o modelo de regressão é uma estratégia para acelerar o processo de modelagem, que consiste em ajustar cada variável com a resposta. Segundo Colosimo & Giolo (2006), esta etapa tem por objetivo selecionar quais variáveis explicativas poderão seguir na análise.

Para o caso da variável não ser selecionada, ainda é interessante que ela seja testada, porque ela pode influenciar a variável resposta quando combinada com outras covariáveis.

O critério de inclusão é dado por,

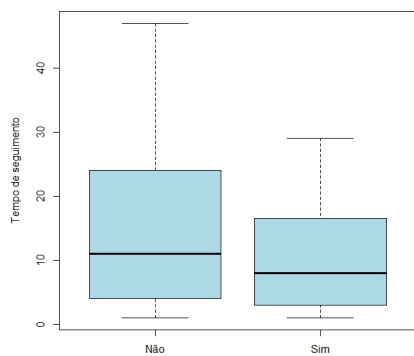
i) Covariável qualitativa:

Se o valor-p do teste logrank for menor do que 0,4, então, a consideramos possível candidata a compor o modelo.

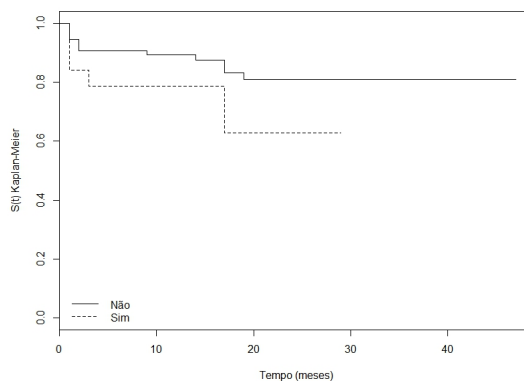
ii) Covariável quantitativa:

Se o valor-p do teste de Wald, calculado via modelo de sobrevivência paramétrico, for menor do que 0,4, então, ela é selecionada.

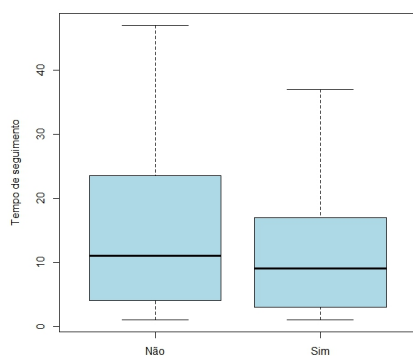
O modelo paramétrico escolhido para seleção das covariáveis numéricas foi o Weibull por se tratar de uma distribuição com função risco mais flexível e por não ter obtido grandes diferenças nas estimativas dos parâmetros quando comparado ao Modelo Cox, que não assume distribuição aos tempos de sobrevivência. Destaca-se também, que o nível de significância α utilizado nas análises de todo o trabalho será de 0,05.



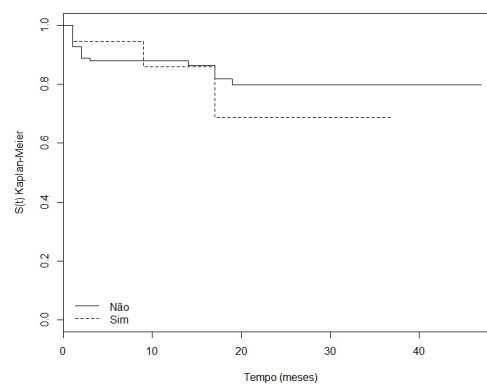
(a) T. de seg. por ressangramento



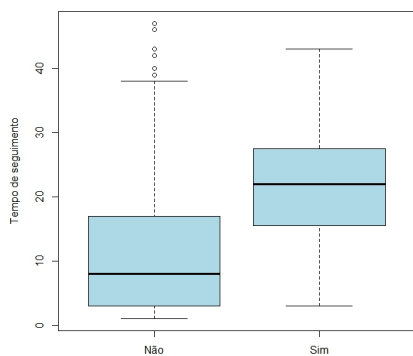
(b) Sobrevivência dos pacientes por ressangramento



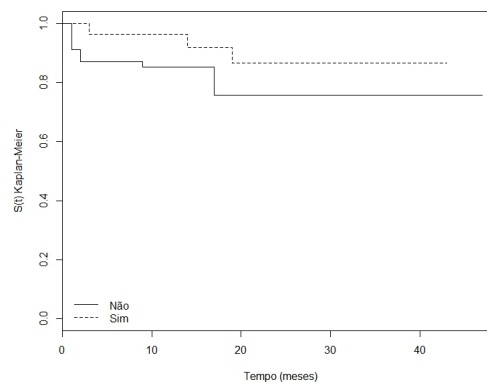
(c) T. de seg. por complicação



(d) Sobrevivência dos pacientes por complicação



(e) T. de seg. por recidiva



(f) Sobrevivência dos pacientes por recidiva

Figura 5: Boxplot e curva de sobrevivência KM para relacionar covariável com o tempo de acompanhamento.

Tabela 4: Teste para selecionar as covariáveis candidatas.

Variável	Valor-p	Selecionada
Log-rank		
Sexo	0,359	Sim
Child- Pugh	0,0307	Sim
HDA _PRE	0,728	Não
Beta-Bloqueador	0,049	Sim
EEVE	0,0486	Sim
Cirurgia	0,042	Sim
Ressangramento	0,122	Sim
Complicações	0,819	Não
Recidiva	0,205	Sim
Wald		
Idade	0,0645	Sim
Sessões	0,272	Sim
Anéis	0,351	Sim

Para as variáveis numéricas Idade, Sessões e Anéis ajustou-se o modelo Weibull, a fim de testar $H_0 : \beta_1 = 0$ e verificar a significância da covariável no modelo, bem como identificar se ela é candidata a compor o modelo de regressão segundo o critério proposto. Os resultados dos testes foram registrados na Tabela 4.

Com isso, observou-se que Idade (valor-p=0,0654), Sessões (valor-p=0,272) e Anéis (valor-p=0,351) foram selecionadas para seguir na análise, segundo o critério. Verifica-se também que todas as três variáveis tiveram valor-p maior do que o nível de significância preestabelecido, logo, há indícios de que estas covariáveis numéricas não explicam significativamente o tempo de sobrevivência do paciente.

Para as demais covariáveis, realizou-se o teste logrank, para comparar as curvas de sobrevivência estimadas por KM de cada categoria. Assim, foi testado H_0 : as curvas de sobrevivência não se diferem contra H_1 : as curvas diferem-se.

Segundo o critério de inclusão da variável, as covariáveis qualitativas selecionadas como candidatas a permanecer na análise são: Sexo (valor-p=0,359), Child-Pugh

(valor-p=0,031), β -Bloqueador (p-valor=0,049), EEVE (valor-p=0,0486), Cirurgia (valor-p=0,042), Ressangramento (valor-p=0,122) e Recidiva (valor-p=0,205). Dentre elas, Child-Pugh, β -Bloqueador, EEVE e Cirurgia foram significativas, indicando que estas covariáveis podem influenciar na sobrevida dos pacientes cirróticos do estudo.

As variáveis restantes HDA (valor-p=0,73) e Complicações (valor-p=0,82) não irão prosseguir na análise.

4.4 Modelagem Estatística

4.4.1 Definição do modelo

Um modelo estatístico tem como finalidade a modelagem de um ou mais desfecho de interesse em termos de fatores a ele(s) associado(s).

Neste estudo, o evento de interesse é o óbito do paciente, desta forma, a variável resposta é o tempo T, em meses, da realização da LEVE até o óbito ou até a censura.

Observou-se um total de 109 censuras, que corresponde a 84,5% dos pacientes. O maior tempo de óbito registrado foi igual a 19 meses e o menor foi 1 mês. O tempo máximo observado foi 47 meses, ou seja, entre 19 e 47 meses só houve a ocorrência de indivíduos censurados, que correspondem a 30,23% dos pacientes.

A Figura 6 mostra a estimativa de Kaplan-Meier da curva de sobrevivência e através dela é possível perceber a existência de uma fração de curados, visto que $\hat{S}(t)$ não tende a zero e ainda, mantém-se constante por um período de 28 meses (2 anos), caracterizando a presença de indivíduos imunes.

A confirmação de haver indícios de imunes na população pode ser obtida pelo teste não paramétrico para presença de indivíduos não suscetíveis. Neste caso, tendo como base as etapas do teste, foram obtidos os seguintes valores,

- i) $\hat{F}(t_{(n)}) = 1 - \hat{S}(t_{(n)}) = 0,216$;
- ii) $\hat{p} = 0,845 \approx \frac{4}{5}$ (proporção de observações censuradas);
- iii) Assim, $\frac{\hat{\phi}}{\hat{\phi} + \hat{\mu}} = \frac{4}{4 + 1}$, isto é, $\hat{\mu} = 1$.

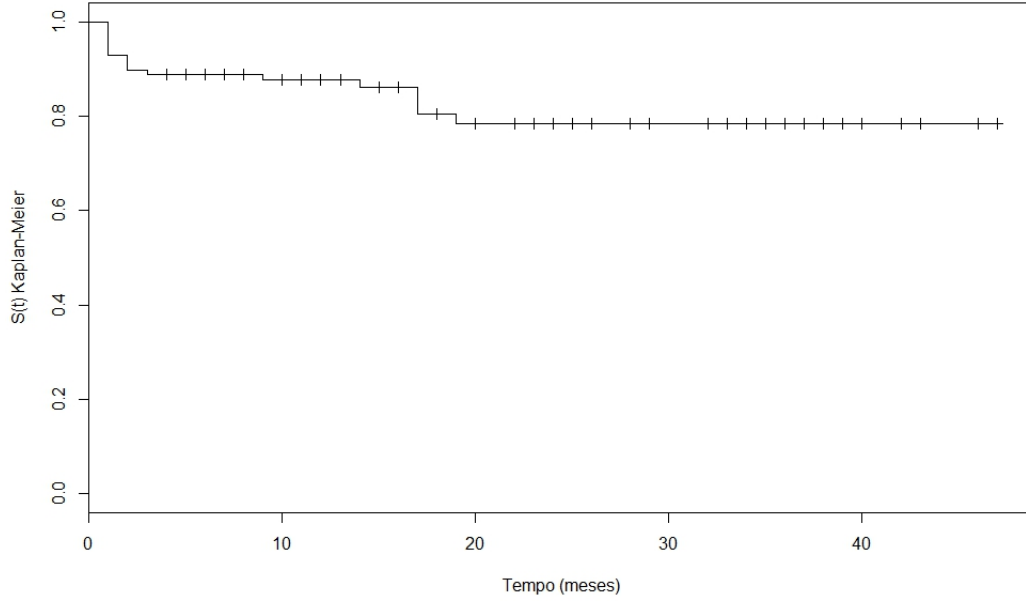


Figura 6: Sobrevivência dos pacientes submetidos à LEVE estimada via KM.

Para $n = 100$ e $\hat{\mu} = 1$, então $\hat{p}_n = 0,74$ (1º *percentil*), valor obtido na Tabela 5.

iv) Como $\hat{F}(t_{(n)}) = 0,216 < 0,74 = \hat{p}_n$, então rejeita-se H_0 , ou seja, há indícios de ter proporção de imunes na população.

O segundo passo do teste não-paramétrico para presença de imunes é testar se o tempo de acompanhamento foi suficiente para concluir que há presença de indivíduos não suscetíveis. Desta forma, tem-se $\hat{q}_n = \frac{20}{129} = 0,17$, a proporção de indivíduos que falharam no intervalo $(2t_{(n)}^* - t_{(n)}, t_{(n)}^*] = (2 * 19 - 47, 19] = (-9, 19] \approx (0, 19]$, pois $t > 0$.

Na Tabela 6, os valores obtidos via simulação por (Maller & Zhou, 1996), partiram da suposição de que a censura segue uma distribuição $U[0, B]$, logo, B é o limite superior do intervalo da distribuição uniforme. Assim, observou-se os quantis de q_n para $n = 100$; como,

$$c_{0,10;6} = 0,03 < q_n = 0,167 < c_{0,90;4} = 0,17,$$

então, o resultado do teste é inconclusivo, não sendo possível afirmar sobre a sufici-

ência do tempo de acompanhamento.

Com isso, modelos de fração de cura (longa duração) são indicados para esse estudo.

Tabela 5: Percentil de $\hat{p}_n$ com distribuição da censura $Exp(1/\mu)$

$n = 100$					
%	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 4$	$\mu = 5$
1	0,7427*	0,8856	0,9286	0,9476	0,9575
5	0,8087	0,9227	0,9560	0,9677	0,9753
10	0,8430	0,9405	0,9667	0,9768	0,9821
20	0,8841	0,9604	0,9791	1,0000	1,0000

Fonte: Apêndice A1 (Maller & Zhou, 1996).

Nota: *Valor utilizado no teste para presença de imunes.

Na Subseção 4.2, os pacientes foram caracterizados de acordo com cada variável. Além disso, na Subseção 4.3, cada variável foi avaliada a fim de selecionar quais iriam compor o modelo completo e assim, selecionaram-se aquelas que apresentaram *valor* $-p < 0,4$. Logo, as covariáveis a serem testadas são Idade, Sexo, Child-Pugh, β -bloqueador, EEVE, Cirurgia, Sessões, Anéis, Ressangramento e Recidiva.

Para dar continuidade, deve-se escolher a distribuição que será associada ao tempo de falha dos indivíduos suscetíveis. Visto que a função de risco pode assumir

Tabela 6: Percentis de q_n com $1 - p = 0,2$ (proporção de suscetíveis)

$n = 100$					
%	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 10$
5	0,0100	0,0100	0,0100	0,0400	0,1000
10	0,0100	0,0100	0,0300	0,0900	0,1200
90	0,1200	0,1700	0,2000	0,2200	0,2300
95	0,1300	0,1900	0,2200	0,2300	0,2400

Fonte: Apêndice B1 (Maller & Zhou, 1996).

várias formas, por meio do gráfico da curva TTT, que fornece indicativo da forma da função de risco da distribuição associada ao tempo de falha, pode-se pré-sugerir algumas distribuições.

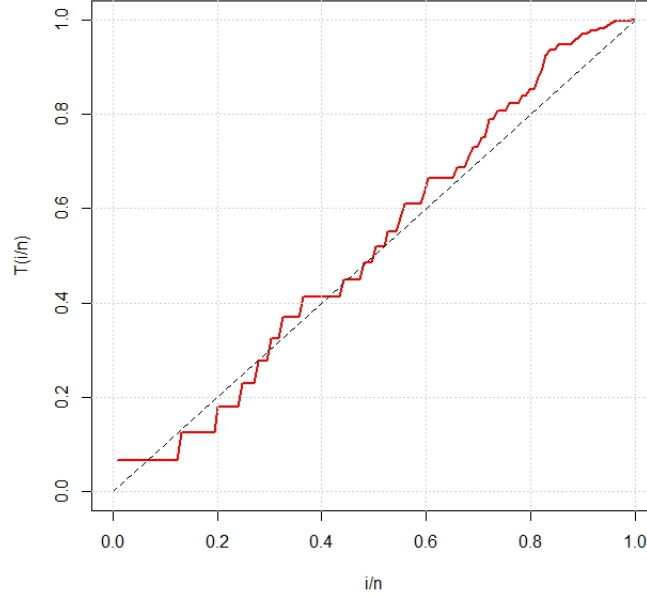


Figura 7: TTT-plot dos tempo de falha dos pacientes suscetíveis.

A Figura 7 apresenta a curva TTT, que indica que a forma da função taxa de falha é aproximadamente constante, por estar oscilando em torno da reta com inclinação igual a 1, podendo até ser em forma de banheira ou crescente. A maioria das distribuições apresentam a função de risco constante, por isso, optou-se incluir distribuições que contemplassem a forma de banheira, além da forma constante e crescente, como a Weibull Modificada e a Weibull Estendida.

As variáveis foram selecionadas até que todas fossem estatisticamente significativas, a partir de um modelo completo e utilizando a distribuição Gama Generalizada. Definido quais variáveis entrariam no modelo final, foi testado todas as demais variáveis que não foram selecionadas, inclusive as HDA e Complicações, as quais não foram testadas inicialmente devido ao critério estabelecido na Subseção 4. No en-

tanto, nenhuma variável foi acrescida ao modelo já estabelecido. Assim, o modelo até então definido foi,

$$\log \left(\frac{\pi_i(\boldsymbol{\beta}, \mathbf{X})}{1 - \pi_i(\boldsymbol{\beta}, \mathbf{X})} \right) = \hat{\beta}_0 + \hat{\beta}_1 * Idade + \hat{\beta}_2 * \beta_Bloqueador + \hat{\beta}_3 * Child.Pugh.$$

O próximo passo para definição do modelo final foi verificar a existência de interação dupla. Combinando as covariáveis e testando cada uma separadamente. Como resultado, obteve-se que a interação entre Idade e Sexo é significativa. Dessa forma, o novo modelo conjunto de variáveis a ser considerado é expresso por,

$$\begin{aligned} \log \left(\frac{\pi_i(\boldsymbol{\beta}, \mathbf{X})}{1 - \pi_i(\boldsymbol{\beta}, \mathbf{X})} \right) = & \hat{\beta}_0 + \hat{\beta}_1 * Idade + \hat{\beta}_2 * \beta_Bloqueador \\ & + \hat{\beta}_3 * Child.Pugh + \hat{\beta}_4 * Sexo + \hat{\beta}_5 * Sexo \times Idade. \end{aligned}$$

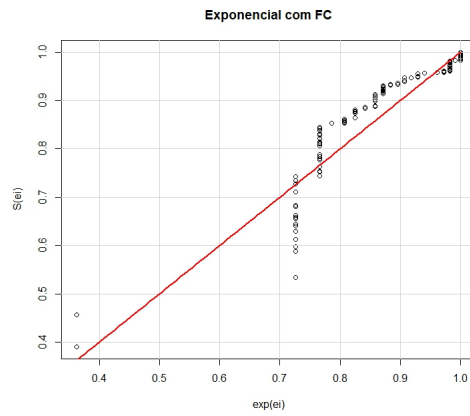
As distribuições Gama Generalizada, Weibull, Weibull Modificada, Weibull Extendida, Gama, Lognormal, Log-logística e Exponencial foram associadas ao tempo de sobrevivência dos suscetíveis, acrescidas das covariáveis significativas definidas anteriormente.

4.4.2 Diagnóstico

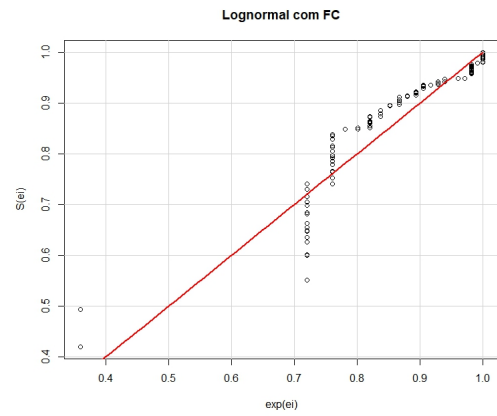
Análise dos resíduos

O modelo é adequado quando os resíduos de Cox-Snell e_i vem de uma população homogênea e devem seguir uma distribuição exponencial. Neste caso, o gráfico $exp(e_i)$ versus $\hat{S}(e_i)$ deve ser aproximadamente uma reta com inclinação de 45°.

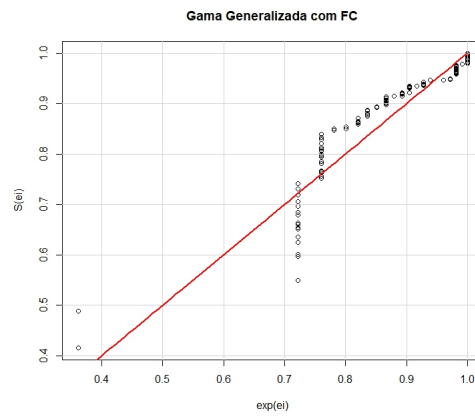
A Figura 8 e a Figura 9 apresentam os gráficos $exp(e_i)$ versus $\hat{S}(e_i)$ para cada modelo ajustado. No geral, pode-se notar que há um problema no ajuste na cauda a esquerda que provavelmente indica que a quantidade excessiva de censura, de certa forma, prejudica as estimativas dos tempos de sobrevivência dos pacientes.



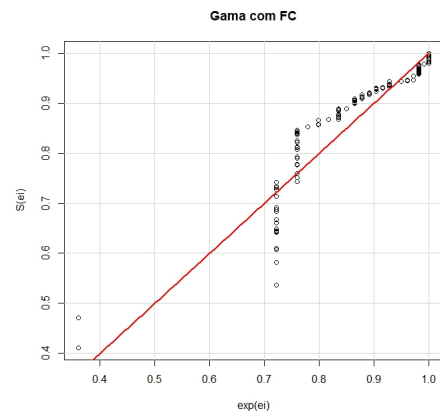
(a) Exponencial



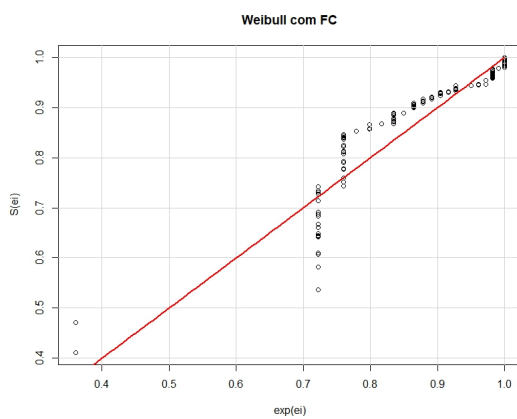
(b) Lognormal



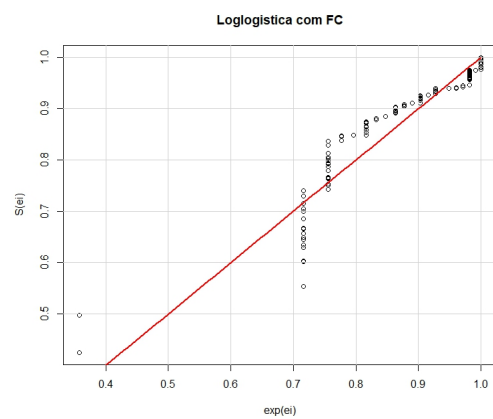
(c) Gama G.



(d) Gama



(e) Weibull



(f) Loglogística

Figura 8: Resíduos Cox-Snell para os modelos ajustados.

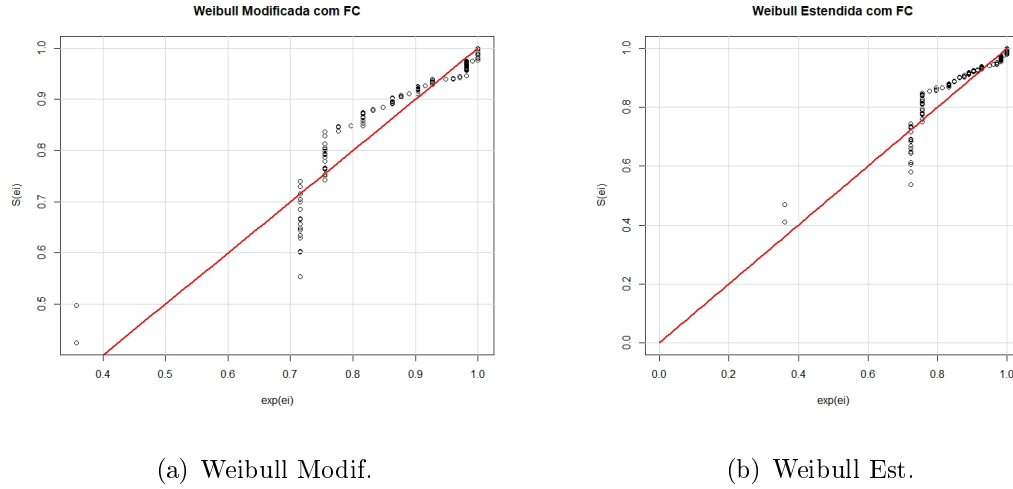


Figura 9: Resíduos Cox-Snell para os modelos ajustados II.

Bondade de Ajuste

Testes para bondade do ajuste indica a qualidade do ajuste de um modelo aos dados observados.

Tabela 7: Teste de Kolmogorov-Smirnov

Distribuição	$\hat{S}(t_{(n)}^*)$	D_n
Exponencial FC	0,8069	0,0229
Lognormal FC	0,8157	0,0317
Gama Generalizada FC	0,8150	0,0310
Gama	0,8108	0,0268
Weibull FC	0,8128	0,088
Loglogística FC	0,8168	0,0328
Weibull Modificada FC	0,8143	0,0303
Weibull Estendida FC	0,8109	0,0269

Nota: $t_{(n)}^* = 19$; $S_{KM}(t_{(n)}^*) = 0,784$; $q_{95\%} = 0,294$

O primeiro teste explorado foi o de Kolmogorov-Smirnov, que testa se um par de amostras são retirados da mesma distribuição. O resultado e o quantil de referência deste teste estão apresentados na Tabela 7.

O modelo proposto é rejeitado, se $D_n > q_{95\%}$, sendo $q_{95\%}$ o quantil do teste de Kolmogorov-Smirnov e

$$D_n = |\hat{S}_{KM}(t_{(n)}^*) - \hat{S}_{proposto}(t_{(n)}^*)|,$$

em que $t_{(n)}^*$ é o maior tempo de falha observado de um indivíduo suscetível.

Logo, tem-se indícios de que os modelos ajustados podem ser adequados aos dados. Diante disso, optou-se por realizar mais um teste de bondade apresentado por Maller & Zhou (1996), que propuseram utilizar o coeficiente de correlação r para testar a bondade do ajuste e, da mesma forma que o teste de Kolmogorov-Smirnov, tem-se interesse em verificar se um par de amostras são retirados da mesma distribuição.

Tabela 8: Percentis do coeficiente de correlação r com $1 - p = 0,2$

$n = 100$					
%	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 10$
1	0,7691	0,8517	0,8877	0,9214	0,9269
5	0,8698	0,9193	0,9353	0,9490	0,9537
10	0,9033*	0,9390	0,9514	0,9602	0,9644
20	0,9314	0,9562	0,9654	0,9707	0,9744

Fonte: Apêndice C1 (Maller & Zhou, 1996).

Nota: *Valor utilizado no teste bondade de ajuste da Tabela 10.

A regra de decisão para esse teste é definida como:

- 1) Se $r > \text{percentil } 80$, então, o modelo proposto está bem ajustado (Tabela 2);
- 2) Se $r > \text{percentil } 10$, então, o modelo tem ajuste razoável (Tabela 8).

A primeira regra não é possível comparar devido não conter na tabela ($1 - p = 0,2$) a proporção de indivíduos suscetíveis da amostra. No entanto, é possível ter uma noção de que há indícios de nenhum modelo estar bem ajustado. No entanto, para o segundo teste, tem-se indícios de que as distribuições Weibull Modificada FC e Weibull Estendida FC apresentam ajuste razoável, tendo em vista que $r = 0,9126 > 0,903 = 10\% \text{ percentil}$ e $r = 0,906 > 0,903 = 10\% \text{ percentil}$

Tabela 9: Critérios de seleção de modelos paramétricos com fração de cura

Distribuição	AIC	BIC
Exponencial	262,255	290,853
Lognormal	267,776	299,234
Gama Generalizada	269,088	303,406
Gama	269,604	301,062
Weibull	271,408	302,866
Loglogística	271,503	302,961
Weibull Modificada	271,864	306,182
Weibull Estendida	272,005	306,322

Portanto, considerando os critérios AIC e BIC, o gráfico de resíduos de Cox-Snell e os testes de bondade de ajuste, a distribuição Weibull Modificada (Lai et al., 2003), dentre os modelos ajustados, é a que pode obter melhores estimativas dos tempos de sobrevivência dos pacientes cirróticos.

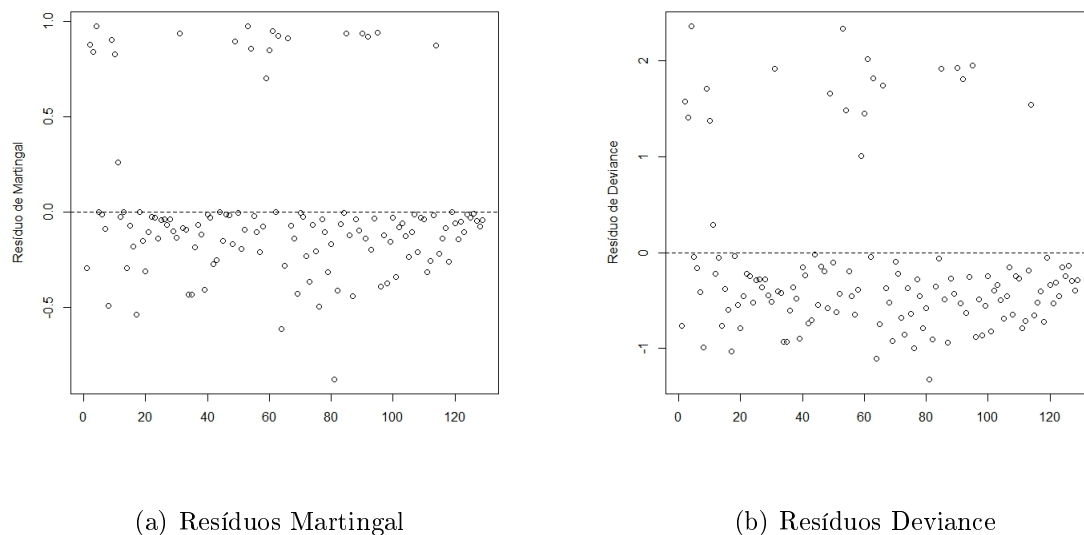


Figura 10: Resíduos de Martingal e Deviance para o modelo Weibull Modificado com Fração de Cura.

Os resíduos de Martingal e Deviance podem ser observados na Figura 10. Nota-

Tabela 10: Teste para bondade do ajuste

Distribuição	Coefficiente de correlação r
Exponencial FC	0,8838
Lognormal FC	0,8718
Gama Generalizada FC	0,8753
Gama FC	0,9016
Weibull FC	0,9020
Loglogística FC	0,8647
Weibull Modificada FC	0,9126
Weibull Estendida FC	0,9060
10% perc., q=0,2	0,903

se a existência de uma separação dos resíduos em dois conjuntos (um grupo positivo e outro negativo), caracterizando a existência de duas subpopulações no estudo, isto é, indicando a presença de pacientes imunes ao evento de interesse ou sobreviventes de longa duração. Além disso, observa-se que os resíduos apresentam comportamento aleatório em torno de zero, apontando que o modelo proposto parece ser apropriado.

4.4.3 Interpretação do modelo

Na Tabela 11 é apresentado as estimativas dos efeitos das covariáveis (β_i) do modelo de regressão ajustado, idade, Child-Pugh, $\beta_{\text{Bloqueador}}$, sexo e a interação sexo*idade, que foram significativas para explicar os tempos de sobrevivência dos pacientes.

As estimativas dos parâmetros da distribuição Weibull Modificada,

$$f(t; a, b, \lambda) = a(b + \lambda t)t^{b-1} \exp(\lambda t) \exp[-at^b \exp(\lambda t)],$$

tal que, $a, \lambda, t > 0$ e $b \geq 0$, são dadas por:

$$\hat{a} = 0,149, \quad \hat{b} = 0,812 \quad e \quad \hat{\lambda} = 6,623 \times 10^{-9}$$

A partir da Equação 4 apresentada na Seção 3.1, tem-se, que a probabilidade de um indivíduo ser sobrevivente de longa duração, isto é, pertencer ao grupo de imunes pode ser estimada por:

$$\hat{\pi}_i = \left[\frac{e^{8,87-5,997*SexoM-0,124*Idade-1,057ChildB-1,857*ChildC+0,994\beta_Bloq+0,091*SexoM*Idade}}{1 + e^{8,87-5,997*SexoM-0,124*Idade-1,057ChildB-1,857*ChildC+0,994\beta_Bloq+0,091*SexoM*Idade}} \right],$$

sendo $i = 1, 1, \dots, n$. Como na regressão logística, por meio da função de ligação logito o preditor lineariza e as razões de chances pode ser obtido pela exponenciação dos β'_i s, estimativa dos efeitos das covariáveis.

Tabela 11: Estimativas do modelo de regressão Weibull Modificada com fração de cura

Parâmetro	Estimativa	Erro Padrão	IC _{95%}	z	Valor-p	$\hat{RC} = \exp(\hat{\beta}_i)$
Intercepto	8,871	3,108	(2,778;14,963)	2,854	0,0022	-
SEXO:M	-5,997	3,308	(-12,481;0,487)	-1,813	0,0349	0,0025
IDADE	-0,124	0,048	(-0,218;-0,03)	-2,587	0,0048	0,8834
CHILD:B	-1,057	0,568	(-2,17;0,055)	-1,863	0,0312	0,3473
CHILD:C	-1,857	0,616	(-3,064;-0,65)	-3,017	0,0013	0,1561
B_BLOQ:Sim	0,944	0,486	(-0,009;1,897)	1,942	0,0261	2,5701
SEXO:M*IDADE	0,091	0,053	(-0,014;0,195)	1,695	0,0451	1,0947

Considerando as demais covariáveis fixas, quanto maior a idade menor a chance de um paciente pertencer ao grupo de imunes, isto é, com o acréscimo de um ano na idade do paciente, considerando as demais covariáveis constante, a chance de sobrevivência diminui em aproximadamente em 1,32 vezes. Em relação ao grau de cirrose (Child-Pugh), o paciente em estágio intermediário diminui 2,88 vezes, enquanto que em o em estágio avançado diminui 6,41, a chance de ser um sobrevivente de longa duração comparado aos pacientes em estágio inicial.

Entretanto, os indivíduos que fizeram o uso do medicamento β -bloqueador tem 2,57 vezes mais chance de pertencer ao grupo imune quando comparado aos que não ingeriram esse medicamento.

Devido à interação significativa, homens tem menos chance de ser sobreviventes de longa duração, quando comparado com as mulheres, no entanto, homens tem 1,09

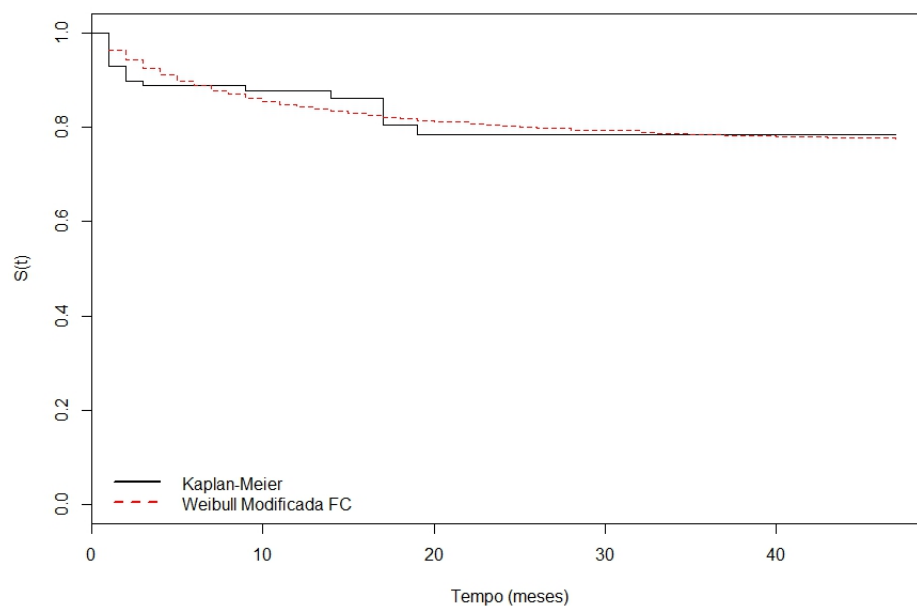


Figura 11: Funções de sobrevivência estimadas para pacientes submetidos à LEVE
vezes mais chance de pertencer ao grupo do imunes quando aumentado a idade em um ano comparado ao acréscimo de um ano na idade das mulheres.

A Figura 11 apresenta as curvas de sobrevivência empírica de KM e Weibull com fração de cura dos pacientes submetidos à LEVE. Pode ser observado que a curva paramétrica acompanha bem a curva de sobrevivência KM.

5 CONSIDERAÇÕES FINAIS

Nesse trabalho foram apresentados alguns modelos de análise de sobrevivência com fração de cura, que é uma alternativa quando se tem uma função de sobrevivência imprópria, para estudo de longa duração, não se sabe a qual grupo um indivíduo censurado pertence, suscetível ou imune. Um algoritmo de otimização EM foi apresentado para estimar os parâmetros dos modelos. Diversas distribuições foram considerados para analisar os tempos de vida de pacientes submetidos à LEVE, sendo que o modelo Weibull Modificado com fração de cura foi considerado o mais adequado dentre os analisados.

Por meios das estimativas do modelo de cura obtidos via o método EM, conclui-se que pacientes homens mais velhos e alto grau da doença hepática e sem uso de medicamento β _Bloqueador possuem menor chance de serem sobreviventes de longa duração.

Na literatura é possível encontrar trabalhos que também utilizam técnicas de análise de sobrevivência com dados relacionados a doença hepática. Por exemplo, Krige et al. (2006) observam que a escleroterapia repetida reduz o ressangramento de varizes mas que mesmo assim, a sobrevivência dos pacientes foi de 26%. Enquanto que, Ezeoke et al. (2015) verificaram que a LEVE não é mais tão eficaz na redução do ressangramento de varizes. Além disso, os pacientes que ingeriram antibiótico tem pior prognóstico (sobrevida) comparados aos que não ingeriram. Entretanto, em ambos casos, a sobrevivência apenas foi estudada utilizando o método KM, caracterizando um estudo mais exploratório da doença, diferente do executado neste trabalho.

Em relação a técnica estatística envolvendo fração de cura, observou-se que na

maioria dos trabalhos consultados, testes para presença de imunes não-paramétrico e paramétrico não são realizados, o que causa estranheza, porque pode acarretar em conclusões equivocadas. Uma justificativa para isso, é a dificuldade para encontrar as tabelas utilizadas pelos testes propostos por Maller & Zhou (1996). Infelizmente os testes não estão disponíveis nos pacotes estatísticos.

Uma limitação encontrada no trabalho foi a grande quantidade de censura (aproximadamente 84%) no conjunto de dados. Devido a isso, todas as distribuições ajustadas tiveram a cauda esquerda dos resíduos afastado da reta de referência, indicando-se como possível solução a busca por um modelo que comporte melhor o excesso de censura. Uma possível abordagem poderia ser o modelo de fração de cura inflacionado de zeros como utilizado por Oliveira-Jr et al. (2017) ou modelo de tempo de promoção (Yakovlev et al., 1996). No entanto, mais estudos são necessários.

Anexos

Anexo 1

Nos Anexos 1 a 3, serão apresentados tabelas resultantes do estudo de simulações feitos por Maller & Zhou (1996) como ferramenta que auxilie na tomada de decisões em testes estatísticos paramétricos e não-paramétricos. As tabelas foram incluídas neste trabalho com a finalidade de divulgar tais testes porque o acesso ao livro desses autores não é simples, além disso, não encontra-se facilmente os testes para presença de imunes em outras fontes.

Statistical Tables

Table A.1 Percentiles of $\hat{p}_n$ with $\text{Exp}(1/\mu)$ censoring.

%	$n = 20$					$n = 150$					%
	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 4$	$\mu = 5$	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 4$	$\mu = 5$	
1	0.4034	0.6607	0.7565	0.8075	0.8329	0.7872	0.9095	0.9476	0.9610	0.9700	1
5	0.5699	0.7734	0.8510	0.8830	0.9052	0.8424	0.9396	0.9664	0.9760	0.9824	5
10	0.6500	0.8273	0.8890	0.9167	0.9297	0.8712	0.9539	0.9753	0.9829	0.9871	10
20	0.7412	0.8871	0.9299	1.0000	1.0000	0.9048	0.9687	0.9845	0.9900	1.0000	20
$n = 30$											
1	0.5306	0.7431	0.8118	0.8616	0.8876	0.8160	0.9281	0.9588	0.9696	0.9763	1
5	0.6500	0.8297	0.8871	0.9163	0.9302	0.8652	0.9525	0.9734	0.9813	0.9860	5
10	0.7139	0.8689	0.9174	0.9377	0.9495	0.8891	0.9634	0.9800	0.9865	0.9898	10
20	0.7881	0.9123	0.9464	0.9622	1.0000	0.9180	0.9751	0.9874	0.9923	1.0000	20
$n = 40$											
1	0.6005	0.7913	0.8600	0.8926	0.9131	0.8306	0.9365	0.9621	0.9752	0.9804	1
5	0.6998	0.8569	0.9095	0.9361	0.9469	0.8776	0.9571	0.9772	0.9843	0.9882	5
10	0.7557	0.8921	0.9324	0.9532	0.9610	0.9013	0.9674	0.9832	0.9887	0.9915	10
20	0.8212	0.9292	0.9568	1.0000	1.0000	0.9281	0.9786	0.9892	1.0000	1.0000	20
$n = 50$											
1	0.6413	0.8123	0.8750	0.9087	0.9266	0.8495	0.9430	0.9680	0.9781	0.9835	1
5	0.7340	0.8746	0.9228	0.9427	0.9556	0.8895	0.9626	0.9795	0.9865	0.9902	5
10	0.7786	0.9042	0.9433	0.9594	0.9671	0.9095	0.9718	0.9852	0.9904	0.9927	10
20	0.8376	0.9370	0.9644	1.0000	1.0000	0.9333	0.9809	0.9905	0.9951	1.0000	20
$n = 60$											
1	0.6668	0.8335	0.8970	0.9206	0.9359	0.8591	0.9496	0.9716	0.9805	0.9851	1
5	0.7515	0.8925	0.9325	0.9509	0.9614	0.8958	0.9663	0.9823	0.9882	0.9911	5
10	0.7976	0.9177	0.9503	0.9645	0.9721	0.9159	0.9744	0.9869	0.9916	0.9935	10
20	0.8500	0.9449	0.9684	0.9799	1.0000	0.9384	0.9830	0.9916	1.0000	1.0000	20
$n = 70$											
1	0.6890	0.8531	0.9055	0.9315	0.9418	0.8725	0.9531	0.9750	0.9821	0.9870	1
5	0.7736	0.8999	0.9409	0.9570	0.9652	0.9039	0.9697	0.9845	0.9893	0.9923	5
10	0.8147	0.9250	0.9560	0.9695	0.9750	0.9218	0.9768	0.9884	0.9923	0.9943	10
20	0.8637	0.9494	0.9728	1.0000	1.0000	0.9429	0.9844	0.9925	0.9954	1.0000	20
$n = 80$											
1	0.7026	0.8665	0.9169	0.9382	0.9485	0.8744	0.9560	0.9757	0.9840	0.9880	1
5	0.7807	0.9102	0.9460	0.9619	0.9704	0.9074	0.9715	0.9848	0.9902	0.9926	5
10	0.8226	0.9323	0.9610	0.9720	0.9780	0.9247	0.9787	0.9891	0.9930	0.9947	10
20	0.8672	0.9541	0.9754	0.9846	1.0000	0.9447	0.9856	0.9931	1.0000	1.0000	20
$n = 90$											
1	0.7230	0.8715	0.9195	0.9419	0.9546	0.8817	0.9601	0.9781	0.9852	0.9892	1
5	0.7945	0.9152	0.9516	0.9650	0.9729	0.9137	0.9732	0.9862	0.9910	0.9935	5
10	0.8332	0.9362	0.9651	0.9746	0.9800	0.9294	0.9795	0.9901	0.9935	0.9952	10
20	0.8785	0.9577	0.9773	0.9849	1.0000	0.9479	0.9864	0.9938	1.0000	1.0000	20
$n = 100$											
1	0.7427	0.8856	0.9286	0.9476	0.9575	0.9170	0.9746	0.9866	0.9913	0.9939	1
5	0.8087	0.9227	0.9560	0.9677	0.9753	0.9386	0.9829	0.9917	0.9949	0.9965	5
10	0.8430	0.9405	0.9667	0.9768	0.9821	0.9504	0.9873	0.9940	0.9963	0.9974	10
20	0.8841	0.9604	0.9791	1.0000	1.0000	0.9635	0.9914	0.9962	1.0000	1.0000	20

STATISTICAL TABLES

Table A.2 Percentiles for $\hat{p}_n$ with $U[0, B]$ censoring.

Table A.2						B = 2						B = 4						B = 6						B = 8						B = 10							
%		B = 2		B = 4		B = 6		B = 8		B = 10		B = 2		B = 4		B = 6		B = 8		B = 10		B = 2		B = 4		B = 6		B = 8		B = 10							
n = 150																																					
n = 20																																					
1	5	10	20	0.7204	0.7763	0.8150	0.8391	0.8607	0.8966	0.9296	0.9509	0.9615	0.9680	0.9802	0.9851	0.9898	1	5	10	20	0.7204	0.7763	0.8150	0.8391	0.8607	0.8966	0.9296	0.9509	0.9615	0.9680	0.9802	0.9851	0.9898				
n = 30																																					
1	5	10	20	0.7665	0.8230	0.8541	0.8795	0.8945	0.9048	0.9399	0.9586	0.9678	0.9738	0.9837	0.9878	0.9917	1	5	10	20	0.7665	0.8230	0.8541	0.8795	0.8945	0.9048	0.9399	0.9586	0.9678	0.9738	0.9837	0.9878	0.9917				
n = 40																																					
1	5	10	20	0.8040	0.8529	0.8859	0.9001	0.9167	0.9152	0.9463	0.9634	0.9726	0.9786	0.9861	0.9896	0.9929	1	5	10	20	0.8040	0.8529	0.8859	0.9001	0.9167	0.9152	0.9463	0.9634	0.9726	0.9786	0.9861	0.9896	0.9929				
n = 50																																					
1	5	10	20	0.8222	0.8718	0.8994	0.9146	0.9288	0.9219	0.9515	0.9664	0.9750	0.9813	0.9875	0.9907	0.9936	1	5	10	20	0.8222	0.8718	0.8994	0.9146	0.9288	0.9219	0.9515	0.9664	0.9750	0.9813	0.9875	0.9907	0.9936				
n = 60																																					
1	5	10	20	0.8399	0.8875	0.9091	0.9273	0.9387	0.9276	0.9549	0.9700	0.9778	0.9829	0.9888	0.9916	0.9944	1	5	10	20	0.8399	0.8875	0.9091	0.9273	0.9387	0.9276	0.9549	0.9700	0.9778	0.9829	0.9888	0.9916	0.9944				
n = 70																																					
1	5	10	20	0.8509	0.8941	0.9181	0.9354	0.9435	0.9309	0.9573	0.9718	0.9790	0.9845	0.9900	0.9925	0.9948	1	5	10	20	0.8509	0.8941	0.9181	0.9354	0.9435	0.9309	0.9573	0.9718	0.9790	0.9845	0.9900	0.9925	0.9948				
n = 80																																					
1	5	10	20	0.8625	0.8995	0.9229	0.9408	0.9495	0.9338	0.9589	0.9731	0.9809	0.9859	0.9906	0.9928	0.9952	1	5	10	20	0.8625	0.8995	0.9229	0.9408	0.9495	0.9338	0.9589	0.9731	0.9809	0.9859	0.9906	0.9928	0.9952				
n = 90																																					
1	5	10	20	0.8695	0.9054	0.9295	0.9467	0.9545	0.9357	0.9612	0.9741	0.9818	0.9866	0.9910	0.9932	0.9954	1	5	10	20	0.8695	0.9054	0.9295	0.9467	0.9545	0.9357	0.9612	0.9741	0.9818	0.9866	0.9910	0.9932	0.9954				
n = 100																																					
1	5	10	20	0.8755	0.9160	0.9357	0.9498	0.9585	0.9501	0.9712	0.9822	0.9880	0.9912	0.9941	0.9955	0.9969	1	5	10	20	0.8755	0.9160	0.9357	0.9498	0.9585	0.9501	0.9712	0.9822	0.9880	0.9912	0.9941	0.9955	0.9969				

Anexo 2

STATISTICAL TABLES

Table B.1 Percentiles of q_n with $p = 0.2$.

255

%	$n = 20$					$n = 200$					%
	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 10$	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 10$	
5	0.0000	0.0500	0.0500	0.0500	0.0500	0.0050	0.0050	0.0050	0.0150	0.0500	5
10	0.0500	0.0500	0.0500	0.0500	0.0500	0.0050	0.0050	0.0100	0.0350	0.1250	10
90	0.2000	0.2500	0.2500	0.3000	0.3000	0.0800	0.1500	0.1850	0.2050	0.2150	90
95	0.2000	0.2500	0.3000	0.3000	0.3500	0.1000	0.1650	0.2000	0.2150	0.2250	95
$n = 30$											
5	0.0333	0.0333	0.0333	0.0667	0.0667	0.0033	0.0033	0.0033	0.0067	0.0233	5
10	0.0333	0.0333	0.0667	0.0667	0.1000	0.0033	0.0033	0.0067	0.0200	0.0800	10
90	0.1667	0.2333	0.2333	0.2667	0.2667	0.0533	0.1267	0.1767	0.1967	0.2067	90
95	0.2000	0.2333	0.2667	0.3000	0.3000	0.0700	0.1467	0.1867	0.2067	0.2133	95
$n = 40$											
5	0.0250	0.0250	0.0500	0.0750	0.0750	0.0025	0.0025	0.0025	0.0050	0.0150	5
10	0.0250	0.0250	0.0750	0.1000	0.1000	0.0025	0.0025	0.0050	0.0125	0.0500	10
90	0.1500	0.2000	0.2250	0.2500	0.2500	0.0350	0.0950	0.1700	0.1925	0.2025	90
95	0.1750	0.2250	0.2500	0.2750	0.2750	0.0500	0.1300	0.1800	0.2000	0.2100	95
$n = 50$											
5	0.0200	0.0200	0.0400	0.0600	0.0800	0.0020	0.0020	0.0020	0.0040	0.0120	5
10	0.0200	0.0400	0.0600	0.1000	0.1000	0.0020	0.0020	0.0040	0.0080	0.0340	10
90	0.1400	0.2000	0.2200	0.2400	0.2400	0.0280	0.0680	0.1640	0.1880	0.1980	90
95	0.1600	0.2200	0.2400	0.2600	0.2800	0.0380	0.1000	0.1740	0.1960	0.2060	95
$n = 60$											
5	0.0167	0.0167	0.0333	0.0667	0.0833	0.0017	0.0017	0.0017	0.0033	0.0083	5
10	0.0167	0.0333	0.0500	0.1000	0.1167	0.0017	0.0017	0.0033	0.0067	0.0250	10
90	0.1333	0.1833	0.2167	0.2333	0.2500	0.0217	0.0517	0.1567	0.1850	0.1967	90
95	0.1500	0.2167	0.2333	0.2500	0.2667	0.0300	0.0750	0.1700	0.1933	0.2033	95
$n = 70$											
5	0.0143	0.0143	0.0286	0.0571	0.1000	0.0014	0.0014	0.0014	0.0029	0.0071	5
10	0.0143	0.0143	0.0429	0.1000	0.1143	0.0014	0.0014	0.0029	0.0057	0.0200	10
90	0.1286	0.1857	0.2143	0.2286	0.2429	0.0186	0.0414	0.1457	0.1829	0.1943	90
95	0.1429	0.2000	0.2286	0.2429	0.2571	0.0257	0.0614	0.1643	0.1900	0.2014	95
$n = 80$											
5	0.0125	0.0125	0.0250	0.0500	0.1000	0.0012	0.0012	0.0012	0.0025	0.0050	5
10	0.0125	0.0125	0.0375	0.1000	0.1125	0.0012	0.0012	0.0025	0.0037	0.0150	10
90	0.1250	0.1750	0.2125	0.2250	0.2375	0.0162	0.0350	0.1225	0.1813	0.1937	90
95	0.1375	0.2000	0.2250	0.2375	0.2500	0.0213	0.0500	0.1575	0.1875	0.1988	95
$n = 90$											
5	0.0111	0.0111	0.0111	0.0444	0.1000	0.0011	0.0011	0.0011	0.0022	0.0044	5
10	0.0111	0.0111	0.0333	0.0889	0.1222	0.0011	0.0011	0.0011	0.0022	0.0044	10
90	0.1222	0.1778	0.2111	0.2222	0.2333	0.0133	0.0289	0.1033	0.1789	0.1922	90
95	0.1333	0.1889	0.2222	0.2333	0.2444	0.0189	0.0422	0.1489	0.1856	0.1978	95
$n = 100$											
5	0.0100	0.0100	0.0100	0.0400	0.1000	0.0010	0.0010	0.0010	0.0010	0.0040	5
10	0.0100	0.0100	0.0300	0.0900	0.1200	0.0010	0.0010	0.0020	0.0030	0.0110	10
90	0.1200	0.1700	0.2000	0.2200	0.2300	0.0120	0.0250	0.0870	0.1770	0.1910	90
95	0.1300	0.1900	0.2200	0.2300	0.2400	0.0170	0.0370	0.1340	0.1840	0.1960	95

STATISTICAL TABLES

256

Table B.2 Percentiles of q_n with $p = 0.3$.

%	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 10$	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 10$	%
$n = 20$											
5	0.0500	0.0500	0.0500	0.1000	0.1000	0.0050	0.0050	0.0050	0.0100	0.0350	5
10	0.0500	0.0500	0.1000	0.1000	0.1500	0.0050	0.0050	0.0100	0.0250	0.1150	10
90	0.2500	0.3000	0.3500	0.4000	0.4000	0.0750	0.1900	0.2650	0.2950	0.3050	90
95	0.2500	0.3500	0.4000	0.4000	0.4500	0.1050	0.2200	0.2800	0.3050	0.3200	95
$n = 30$											
5	0.0333	0.0333	0.0333	0.1000	0.1333	0.0033	0.0033	0.0033	0.0067	0.0200	5
10	0.0333	0.0333	0.1000	0.1333	0.1667	0.0033	0.0033	0.0067	0.0133	0.0600	10
90	0.2000	0.3000	0.3333	0.3667	0.3667	0.0467	0.1200	0.2500	0.2833	0.3000	90
95	0.2333	0.3333	0.3667	0.4000	0.4000	0.0633	0.1733	0.2663	0.2933	0.3100	95
$n = 40$											
5	0.0250	0.0250	0.0250	0.1000	0.1500	0.0025	0.0025	0.0025	0.0050	0.0125	5
10	0.0250	0.0250	0.0750	0.1500	0.1750	0.0025	0.0025	0.0050	0.0100	0.0375	10
90	0.2000	0.2750	0.3250	0.3500	0.3500	0.0325	0.0775	0.2350	0.2775	0.2925	90
95	0.2250	0.3000	0.3500	0.3750	0.3750	0.0450	0.1125	0.2525	0.2875	0.3025	95
$n = 50$											
5	0.0200	0.0200	0.0200	0.0800	0.1400	0.0020	0.0020	0.0020	0.0040	0.0080	5
10	0.0200	0.0200	0.0600	0.1400	0.1800	0.0020	0.0020	0.0040	0.0080	0.0240	10
90	0.1800	0.2800	0.3200	0.3400	0.3400	0.0260	0.0560	0.2040	0.2720	0.2900	90
95	0.2200	0.3000	0.3400	0.3600	0.3800	0.0340	0.0820	0.2420	0.2820	0.2980	95
$n = 60$											
5	0.0167	0.0167	0.0167	0.0500	0.1500	0.0017	0.0017	0.0017	0.0033	0.0067	5
10	0.0167	0.0167	0.0500	0.1333	0.1833	0.0017	0.0017	0.0033	0.0050	0.0183	10
90	0.1833	0.2667	0.3000	0.3333	0.3333	0.0200	0.0433	0.1517	0.2683	0.2867	90
95	0.2000	0.2833	0.3333	0.3500	0.3667	0.0283	0.0633	0.2233	0.2783	0.2950	95
$n = 70$											
5	0.0143	0.0143	0.0143	0.0429	0.1429	0.0014	0.0014	0.0014	0.0014	0.0043	5
10	0.0143	0.0143	0.0429	0.1286	0.1857	0.0014	0.0014	0.0014	0.0043	0.0143	10
90	0.1714	0.2571	0.3000	0.3286	0.3429	0.0171	0.0357	0.1186	0.2643	0.2843	90
95	0.1857	0.2714	0.3286	0.3429	0.3571	0.0229	0.0514	0.1814	0.2743	0.2929	95
$n = 80$											
5	0.0125	0.0125	0.0125	0.0375	0.1375	0.0012	0.0012	0.0012	0.0012	0.0037	5
10	0.0125	0.0125	0.0375	0.1125	0.1875	0.0012	0.0012	0.0012	0.0037	0.0125	10
90	0.1625	0.2500	0.3000	0.3250	0.3375	0.0150	0.0300	0.0962	0.2612	0.2825	90
95	0.1875	0.2750	0.3125	0.3375	0.3500	0.0200	0.0425	0.1462	0.2713	0.2900	95
$n = 90$											
5	0.0111	0.0111	0.0111	0.0333	0.1222	0.0011	0.0011	0.0011	0.0011	0.0033	5
10	0.0111	0.0111	0.0222	0.0889	0.1889	0.0011	0.0011	0.0011	0.0033	0.0100	10
90	0.1556	0.2444	0.2889	0.3111	0.3333	0.0133	0.0256	0.0800	0.2567	0.2811	90
95	0.1778	0.2667	0.3111	0.3333	0.3444	0.0178	0.0367	0.1211	0.2678	0.2878	95
$n = 100$											
5	0.0100	0.0100	0.0100	0.0300	0.1100	0.0010	0.0010	0.0010	0.0010	0.0030	5
10	0.0100	0.0100	0.0200	0.0800	0.1900	0.0010	0.0010	0.0010	0.0030	0.0080	10
90	0.1500	0.2400	0.2900	0.3100	0.3200	0.0110	0.0220	0.0680	0.2520	0.2800	90
95	0.1700	0.2600	0.3100	0.3300	0.3400	0.0150	0.0310	0.1040	0.2650	0.2870	95

STATISTICAL TABLES

Table B.3 Percentiles of q_n with $p = 0.4$.

257

%	$n = 20$					$n = 200$					%
	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 10$	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 10$	
5	0.0500	0.0500	0.0500	0.1000	0.1500	0.0050	0.0050	0.0050	0.0100	0.0300	5
10	0.0500	0.0500	0.1000	0.2000	0.2000	0.0050	0.0050	0.0100	0.0250	0.0950	10
90	0.3000	0.4000	0.4500	0.5000	0.5000	0.0700	0.1850	0.3400	0.3800	0.4000	90
95	0.3500	0.4500	0.5000	0.5000	0.5500	0.1000	0.2550	0.3550	0.3950	0.4100	95
$n = 30$											
5	0.0333	0.0333	0.0333	0.1000	0.2000	0.0033	0.0033	0.0033	0.0067	0.0167	5
10	0.0333	0.0333	0.1000	0.2000	0.2333	0.0033	0.0033	0.0067	0.0133	0.0467	10
90	0.2667	0.3667	0.4333	0.4667	0.4667	0.0433	0.1033	0.3133	0.3700	0.3900	90
95	0.3000	0.4000	0.4667	0.5000	0.5000	0.0600	0.1500	0.3367	0.3800	0.4000	95
$n = 40$											
5	0.0250	0.0250	0.0250	0.0750	0.2000	0.0025	0.0025	0.0025	0.0050	0.0100	5
10	0.0250	0.0250	0.0750	0.1750	0.2500	0.0025	0.0025	0.0050	0.0075	0.0300	10
90	0.2500	0.3500	0.4000	0.4500	0.4500	0.0300	0.0675	0.2425	0.3600	0.3825	90
95	0.2750	0.3750	0.4500	0.4750	0.4750	0.0425	0.1000	0.3150	0.3725	0.3925	95
$n = 50$											
5	0.0200	0.0200	0.0200	0.0600	0.2000	0.0020	0.0020	0.0020	0.0020	0.0080	5
10	0.0200	0.0200	0.0600	0.1600	0.2600	0.0020	0.0020	0.0020	0.0060	0.0220	10
90	0.2200	0.3400	0.4000	0.4200	0.4400	0.0240	0.0500	0.1700	0.3540	0.3800	90
95	0.2600	0.3600	0.4200	0.4600	0.4800	0.0320	0.0720	0.2580	0.3660	0.3880	95
$n = 60$											
5	0.0167	0.0167	0.0167	0.0500	0.1833	0.0017	0.0017	0.0017	0.0017	0.0050	5
10	0.0167	0.0167	0.0333	0.1333	0.2500	0.0017	0.0017	0.0017	0.0050	0.0167	10
90	0.2167	0.3333	0.3833	0.4167	0.4333	0.0200	0.0400	0.1283	0.3467	0.3767	90
95	0.2500	0.3500	0.4167	0.4500	0.4667	0.0267	0.0567	0.1967	0.3600	0.3850	95
$n = 70$											
5	0.0143	0.0143	0.0143	0.0429	0.1429	0.0014	0.0014	0.0014	0.0014	0.0043	5
10	0.0143	0.0143	0.0286	0.1000	0.2571	0.0014	0.0014	0.0014	0.0043	0.0129	10
90	0.2000	0.3143	0.3857	0.4143	0.4286	0.0171	0.0314	0.1000	0.3400	0.3743	90
95	0.2286	0.3429	0.4000	0.4286	0.4571	0.0214	0.0457	0.1514	0.3557	0.3814	95
$n = 80$											
5	0.0125	0.0125	0.0125	0.0375	0.1250	0.0012	0.0012	0.0012	0.0012	0.0037	5
10	0.0125	0.0125	0.0250	0.0875	0.2500	0.0012	0.0012	0.0012	0.0037	0.0100	10
90	0.1875	0.3125	0.3750	0.4125	0.4250	0.0137	0.0262	0.0812	0.3313	0.3713	90
95	0.2250	0.3375	0.4000	0.4250	0.4500	0.0188	0.0375	0.1225	0.3500	0.3800	95
$n = 90$											
5	0.0111	0.0111	0.0111	0.0333	0.1000	0.0011	0.0011	0.0011	0.0011	0.0033	5
10	0.0111	0.0111	0.0222	0.0778	0.2556	0.0011	0.0011	0.0011	0.0022	0.0089	10
90	0.1778	0.3000	0.3778	0.4000	0.4222	0.0122	0.0222	0.0678	0.3022	0.3700	90
95	0.2111	0.3333	0.3889	0.4222	0.4444	0.0167	0.0322	0.1022	0.3444	0.3778	95
$n = 100$											
5	0.0100	0.0100	0.0100	0.0200	0.0900	0.0010	0.0010	0.0010	0.0010	0.0030	5
10	0.0100	0.0100	0.0200	0.0600	0.2500	0.0010	0.0010	0.0010	0.0020	0.0070	10
90	0.1600	0.3000	0.3700	0.4000	0.4200	0.0110	0.0200	0.0570	0.2510	0.3680	90
95	0.2000	0.3200	0.3900	0.4200	0.4400	0.0150	0.0280	0.0870	0.3380	0.3750	95

STATISTICAL TABLES

Table B.4 Percentiles of q_n with $p = 0.5$.

%	$n = 20$					$n = 200$					%
	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 10$	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 10$	
5	0.0500	0.0500	0.0500	0.1000	0.2000	0.0050	0.0050	0.0050	0.0100	0.0250	5
10	0.0500	0.0500	0.1000	0.2000	0.3000	0.0050	0.0050	0.0100	0.0200	0.0850	10
90	0.3500	0.4500	0.5500	0.5500	0.6000	0.0650	0.1650	0.4050	0.4650	0.4900	90
95	0.4000	0.5000	0.5500	0.6000	0.6500	0.0900	0.2400	0.4300	0.4800	0.5000	95
$n = 300$											
5	0.0333	0.0333	0.0333	0.1000	0.2333	0.0033	0.0033	0.0033	0.0067	0.0133	5
10	0.0333	0.0333	0.0667	0.2333	0.3000	0.0033	0.0033	0.0067	0.0100	0.0400	10
90	0.3000	0.4333	0.5000	0.5333	0.5667	0.0433	0.0933	0.3333	0.4533	0.4800	90
95	0.3333	0.4667	0.5333	0.5667	0.6000	0.0567	0.1367	0.4033	0.4667	0.4900	95
$n = 400$											
5	0.0250	0.0250	0.0250	0.0750	0.2250	0.0025	0.0025	0.0025	0.0025	0.0100	5
10	0.0250	0.0250	0.0500	0.1750	0.3250	0.0025	0.0025	0.0025	0.0075	0.0250	10
90	0.2750	0.4250	0.5000	0.5250	0.5500	0.0300	0.0625	0.2125	0.4425	0.4725	90
95	0.3250	0.4500	0.5250	0.5500	0.5750	0.0400	0.0900	0.3250	0.4550	0.4825	95
$n = 500$											
5	0.0200	0.0200	0.0200	0.0600	0.1800	0.0020	0.0020	0.0020	0.0020	0.0060	5
10	0.0200	0.0200	0.0400	0.1400	0.3200	0.0020	0.0020	0.0020	0.0060	0.0180	10
90	0.2600	0.4000	0.4800	0.5200	0.5400	0.0240	0.0460	0.1480	0.4320	0.4680	90
95	0.3000	0.4400	0.5000	0.5400	0.5600	0.0320	0.0660	0.2260	0.4480	0.4780	95
$n = 600$											
5	0.0167	0.0167	0.0167	0.0500	0.1333	0.0017	0.0017	0.0017	0.0017	0.0050	5
10	0.0167	0.0167	0.0333	0.1167	0.3167	0.0017	0.0017	0.0017	0.0050	0.0133	10
90	0.2333	0.3833	0.4667	0.5000	0.5333	0.0183	0.0350	0.1117	0.4200	0.4650	90
95	0.2833	0.4167	0.5000	0.5333	0.5500	0.0250	0.0517	0.1683	0.4417	0.4733	95
$n = 700$											
5	0.0143	0.0143	0.0143	0.0286	0.1143	0.0014	0.0014	0.0014	0.0014	0.0043	5
10	0.0143	0.0143	0.0286	0.0857	0.3143	0.0014	0.0014	0.0014	0.0029	0.0100	10
90	0.2143	0.3857	0.4571	0.5000	0.5286	0.0157	0.0286	0.0886	0.3886	0.4614	90
95	0.2571	0.4143	0.4857	0.5286	0.5429	0.0214	0.0414	0.1314	0.4329	0.4700	95
$n = 800$											
5	0.0125	0.0125	0.0125	0.0250	0.1000	0.0012	0.0012	0.0012	0.0012	0.0037	5
10	0.0125	0.0125	0.0250	0.0750	0.3000	0.0012	0.0012	0.0012	0.0025	0.0088	10
90	0.2000	0.3625	0.4625	0.5000	0.5125	0.0137	0.0250	0.0700	0.3100	0.4588	90
95	0.2500	0.4000	0.4750	0.5125	0.5375	0.0188	0.0350	0.1063	0.4225	0.4675	95
$n = 900$											
5	0.0111	0.0111	0.0111	0.0222	0.0889	0.0011	0.0011	0.0011	0.0011	0.0022	5
10	0.0111	0.0111	0.0222	0.0667	0.2778	0.0011	0.0011	0.0011	0.0022	0.0078	10
90	0.1778	0.3556	0.4556	0.4889	0.5111	0.0122	0.0211	0.0589	0.2533	0.4567	90
95	0.2333	0.3889	0.4778	0.5111	0.5333	0.0156	0.0300	0.0900	0.3944	0.4656	95
$n = 1000$											
5	0.0100	0.0100	0.0100	0.0200	0.0700	0.0010	0.0010	0.0010	0.0010	0.0020	5
10	0.0100	0.0100	0.0200	0.0500	0.2300	0.0010	0.0010	0.0010	0.0020	0.0060	10
90	0.1500	0.3500	0.4500	0.4900	0.5100	0.0110	0.0180	0.0500	0.2130	0.4540	90
95	0.2100	0.3800	0.4700	0.5100	0.5300	0.0140	0.0260	0.0750	0.3360	0.4630	95

STATISTICAL TABLES

259

Table B.5 Percentiles of q_n with $p = 0.6$.

n	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 10$	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 10$	%
$n = 20$											
5	0.0500	0.0500	0.0500	0.1000	0.2000	0.0050	0.0050	0.0050	0.0100	0.0200	5
10	0.0500	0.0500	0.1000	0.2000	0.3500	0.0050	0.0050	0.0100	0.0200	0.0700	10
90	0.6500	0.5500	0.6000	0.6500	0.6500	0.0650	0.1500	0.4700	0.5500	0.5750	90
95	0.6800	0.5300	0.6500	0.7000	0.7000	0.0900	0.2200	0.5050	0.5650	0.5900	95
$n = 30$											
5	0.0333	0.0333	0.0333	0.0667	0.2000	0.0033	0.0033	0.0033	0.0033	0.0133	5
10	0.0333	0.0333	0.0667	0.1667	0.3667	0.0033	0.0033	0.0033	0.0100	0.0367	10
90	0.3333	0.5000	0.5667	0.6333	0.6333	0.0400	0.0833	0.2900	0.5333	0.5667	90
95	0.4000	0.5333	0.6333	0.6667	0.6667	0.0533	0.1233	0.4400	0.5500	0.5800	95
$n = 40$											
5	0.0250	0.0250	0.0250	0.0500	0.1750	0.0025	0.0025	0.0025	0.0025	0.0075	5
10	0.0250	0.0250	0.0500	0.1500	0.3750	0.0025	0.0025	0.0025	0.0075	0.0225	10
90	0.3250	0.4750	0.5750	0.6000	0.6250	0.0300	0.0575	0.1850	0.5200	0.5600	90
95	0.3500	0.5250	0.6000	0.6500	0.6750	0.0400	0.0825	0.2850	0.5375	0.5725	95
$n = 50$											
5	0.0200	0.0200	0.0200	0.0400	0.1400	0.0020	0.0020	0.0020	0.0020	0.0060	5
10	0.0200	0.0200	0.0400	0.1200	0.3800	0.0020	0.0020	0.0020	0.0060	0.0160	10
90	0.2800	0.4600	0.5600	0.6000	0.6200	0.0220	0.0420	0.1340	0.5060	0.5560	90
95	0.3400	0.5000	0.5800	0.6200	0.6400	0.0300	0.0600	0.2040	0.5280	0.5660	95
$n = 60$											
5	0.0167	0.0167	0.0167	0.0333	0.1167	0.0017	0.0017	0.0017	0.0017	0.0050	5
10	0.0167	0.0167	0.0333	0.0833	0.3667	0.0017	0.0017	0.0017	0.0033	0.0117	10
90	0.2500	0.4500	0.5500	0.6000	0.6167	0.0183	0.0333	0.1000	0.4400	0.5517	90
95	0.3167	0.4833	0.5833	0.6167	0.6333	0.0250	0.0467	0.1500	0.5167	0.5617	95
$n = 70$											
5	0.0143	0.0143	0.0143	0.0286	0.1000	0.0014	0.0014	0.0014	0.0014	0.0043	5
10	0.0143	0.0143	0.0286	0.0714	0.3143	0.0014	0.0014	0.0014	0.0029	0.0100	10
90	0.2143	0.4286	0.5429	0.5857	0.6143	0.0157	0.0271	0.0786	0.3386	0.5486	90
95	0.2857	0.4714	0.5714	0.6143	0.6286	0.0200	0.0386	0.1186	0.5000	0.5586	95
$n = 80$											
5	0.0125	0.0125	0.0125	0.0250	0.0750	0.0012	0.0012	0.0012	0.0012	0.0025	5
10	0.0125	0.0125	0.0250	0.0625	0.2625	0.0012	0.0012	0.0012	0.0025	0.0075	10
90	0.1875	0.4125	0.5375	0.5750	0.6000	0.0137	0.0225	0.0650	0.2688	0.5450	90
95	0.2500	0.4625	0.5625	0.6000	0.6250	0.0175	0.0325	0.0975	0.4238	0.5550	95
$n = 90$											
5	0.0111	0.0111	0.0111	0.0222	0.0667	0.0011	0.0011	0.0011	0.0011	0.0022	5
10	0.0111	0.0111	0.0222	0.0556	0.2222	0.0011	0.0011	0.0011	0.0022	0.0067	10
90	0.1667	0.4000	0.5222	0.5778	0.6000	0.0122	0.0200	0.0533	0.2233	0.5422	90
95	0.2333	0.4444	0.5556	0.6000	0.6222	0.0156	0.0278	0.0800	0.3467	0.5522	95
$n = 100$											
5	0.0100	0.0100	0.0100	0.0200	0.0600	0.0010	0.0010	0.0010	0.0010	0.0020	5
10	0.0100	0.0100	0.0200	0.0400	0.2000	0.0010	0.0010	0.0010	0.0020	0.0050	10
90	0.1500	0.3700	0.5200	0.5700	0.6000	0.0100	0.0170	0.0460	0.1860	0.5390	90
95	0.2000	0.4400	0.5500	0.5900	0.6200	0.0140	0.0240	0.0690	0.2920	0.5500	95

STATISTICAL TABLES

260

Table B.6 Percentiles of q_n with $p = 0.7$.

%	B = 2	B = 4	B = 6	B = 8	B = 10	B = 2	B = 4	B = 6	B = 8	B = 10	%
n = 20											
5	0.0500	0.0500	0.0500	0.0500	0.1000	0.0050	0.0050	0.0050	0.0050	0.0200	5
10	0.0500	0.0500	0.0500	0.1000	0.3500	0.0050	0.0050	0.0050	0.0150	0.0600	10
90	0.4000	0.6000	0.7000	0.7000	0.7500	0.0600	0.1350	0.4900	0.6300	0.6650	90
95	0.4500	0.6500	0.7000	0.7500	0.8000	0.0850	0.2000	0.5700	0.6450	0.6800	95
n = 30											
5	0.0333	0.0333	0.0333	0.0333	0.1333	0.0033	0.0033	0.0033	0.0033	0.0100	5
10	0.0333	0.0333	0.0333	0.1000	0.4000	0.0033	0.0033	0.0033	0.0100	0.0300	10
90	0.3667	0.5667	0.6667	0.7000	0.7333	0.0400	0.0767	0.2633	0.6133	0.6533	90
95	0.4333	0.6000	0.7000	0.7333	0.7667	0.0533	0.1100	0.4033	0.6333	0.6667	95
n = 40											
5	0.0250	0.0250	0.0250	0.0500	0.1250	0.0025	0.0025	0.0025	0.0025	0.0075	5
10	0.0250	0.0250	0.0500	0.1000	0.3750	0.0025	0.0025	0.0025	0.0075	0.0200	10
90	0.3250	0.5250	0.6500	0.7000	0.7250	0.0275	0.0525	0.1675	0.5975	0.6475	90
95	0.4000	0.5750	0.6750	0.7250	0.7500	0.0375	0.0775	0.2575	0.6200	0.6600	95
n = 50											
5	0.0200	0.0200	0.0200	0.0400	0.1000	0.0020	0.0020	0.0020	0.0020	0.0060	5
10	0.0200	0.0200	0.0400	0.0800	0.3400	0.0020	0.0020	0.0020	0.0040	0.0140	10
90	0.3000	0.5200	0.6400	0.6800	0.7000	0.0220	0.0400	0.1180	0.5260	0.6440	90
95	0.3600	0.5600	0.6600	0.7000	0.7400	0.0300	0.0560	0.1780	0.6060	0.6540	95
n = 60											
5	0.0167	0.0167	0.0167	0.0333	0.0833	0.0017	0.0017	0.0017	0.0017	0.0033	5
10	0.0167	0.0167	0.0333	0.0667	0.2667	0.0017	0.0017	0.0017	0.0033	0.0100	10
90	0.2500	0.5000	0.6167	0.6667	0.7000	0.0183	0.0317	0.0900	0.3967	0.6383	90
95	0.3333	0.5500	0.6500	0.7000	0.7167	0.0233	0.0433	0.1367	0.5867	0.6483	95
n = 70											
5	0.0143	0.0143	0.0143	0.0286	0.0714	0.0014	0.0014	0.0014	0.0014	0.0029	5
10	0.0143	0.0143	0.0286	0.0571	0.2286	0.0014	0.0014	0.0014	0.0029	0.0086	10
90	0.2143	0.4714	0.6143	0.6714	0.7000	0.0143	0.0257	0.0714	0.3057	0.6343	90
95	0.2857	0.5286	0.6429	0.6857	0.7143	0.0200	0.0357	0.1071	0.4757	0.6457	95
n = 80											
5	0.0125	0.0125	0.0125	0.0250	0.0625	0.0012	0.0012	0.0012	0.0012	0.0025	5
10	0.0125	0.0125	0.0250	0.0500	0.2000	0.0012	0.0012	0.0012	0.0025	0.0063	10
90	0.1750	0.4500	0.6125	0.6625	0.6875	0.0125	0.0225	0.0587	0.2412	0.6313	90
95	0.2500	0.5125	0.6375	0.6875	0.7125	0.0175	0.0312	0.0875	0.3787	0.6413	95
n = 90											
5	0.0111	0.0111	0.0111	0.0222	0.0556	0.0011	0.0011	0.0011	0.0011	0.0022	5
10	0.0111	0.0111	0.0222	0.0444	0.1778	0.0011	0.0011	0.0011	0.0022	0.0056	10
90	0.1556	0.4000	0.6000	0.6556	0.6889	0.0111	0.0189	0.0489	0.1989	0.6267	90
95	0.2222	0.5000	0.6222	0.6778	0.7111	0.0156	0.0267	0.0733	0.3133	0.6378	95
n = 100											
5	0.0100	0.0100	0.0100	0.0200	0.0500	0.0010	0.0010	0.0010	0.0010	0.0020	5
10	0.0100	0.0100	0.0200	0.0400	0.1600	0.0010	0.0010	0.0010	0.0020	0.0050	10
90	0.1400	0.3500	0.5900	0.6500	0.6800	0.0100	0.0160	0.0430	0.1680	0.6230	90
95	0.1900	0.4800	0.6200	0.6800	0.7000	0.0140	0.0230	0.0630	0.2610	0.6360	95

STATISTICAL TABLES

Table B.7 Percentiles of q_n with $p = 0.8$.

261

%	B = 2	B = 4	B = 6	B = 8	B = 10	B = 2	B = 4	B = 6	B = 8	B = 10	%
$n = 20$											
5	0.0500	0.0500	0.0500	0.0500	0.0500	0.0050	0.0050	0.0050	0.0050	0.0150	5
10	0.0500	0.0500	0.0500	0.0500	0.1000	0.0050	0.0050	0.0050	0.0150	0.0500	10
90	0.4500	0.6500	0.7500	0.8000	0.8000	0.0600	0.1200	0.4300	0.7100	0.7500	90
95	0.5000	0.7000	0.7500	0.8000	0.8500	0.0800	0.1800	0.6200	0.7300	0.7650	95
$n = 30$											
5	0.0333	0.0333	0.0333	0.0333	0.0333	0.0033	0.0033	0.0033	0.0033	0.0100	5
10	0.0333	0.0333	0.0333	0.0667	0.1333	0.0033	0.0033	0.0033	0.0100	0.0267	10
90	0.4000	0.6000	0.7333	0.7667	0.8000	0.0367	0.0733	0.2300	0.6933	0.7400	90
95	0.4667	0.6667	0.7667	0.8000	0.8333	0.0500	0.1033	0.3533	0.7133	0.7533	95
$n = 40$											
5	0.0250	0.0250	0.0250	0.0250	0.0500	0.0025	0.0025	0.0025	0.0025	0.0075	5
10	0.0250	0.0250	0.0250	0.0500	0.1750	0.0025	0.0025	0.0025	0.0050	0.0175	10
90	0.3500	0.5750	0.7000	0.7750	0.8000	0.0275	0.0500	0.1525	0.6575	0.7350	90
95	0.4250	0.6250	0.7500	0.8000	0.8250	0.0350	0.0700	0.2275	0.6975	0.7450	95
$n = 50$											
5	0.0200	0.0200	0.0200	0.0200	0.0600	0.0020	0.0020	0.0020	0.0020	0.0040	5
10	0.0200	0.0200	0.0200	0.0600	0.1800	0.0020	0.0020	0.0020	0.0040	0.0120	10
90	0.2800	0.5600	0.7000	0.7600	0.7800	0.0220	0.0360	0.1060	0.4720	0.7300	90
95	0.3800	0.6200	0.7400	0.7800	0.8200	0.0280	0.0520	0.1620	0.6780	0.7400	95
$n = 60$											
5	0.0167	0.0167	0.0167	0.0167	0.0500	0.0017	0.0017	0.0017	0.0017	0.0033	5
10	0.0167	0.0167	0.0167	0.0500	0.1667	0.0017	0.0017	0.0017	0.0033	0.0100	10
90	0.2333	0.5333	0.6833	0.7500	0.7833	0.0167	0.0300	0.0817	0.3483	0.7250	90
95	0.3167	0.6000	0.7167	0.7833	0.8000	0.0233	0.0417	0.1217	0.5483	0.7350	95
$n = 70$											
5	0.0143	0.0143	0.0143	0.0143	0.0429	0.0014	0.0014	0.0014	0.0014	0.0029	5
10	0.0143	0.0143	0.0143	0.0429	0.1571	0.0014	0.0014	0.0014	0.0029	0.0071	10
90	0.2000	0.4857	0.6857	0.7429	0.7714	0.0143	0.0243	0.0657	0.2700	0.7200	90
95	0.2714	0.5714	0.7143	0.7714	0.8000	0.0200	0.0343	0.0986	0.4229	0.7314	95
$n = 80$											
5	0.0125	0.0125	0.0125	0.0125	0.0500	0.0012	0.0012	0.0012	0.0012	0.0025	5
10	0.0125	0.0125	0.0125	0.0375	0.1375	0.0012	0.0012	0.0012	0.0025	0.0063	10
90	0.1625	0.4000	0.6750	0.7375	0.7750	0.0125	0.0213	0.0538	0.2150	0.7163	90
95	0.2375	0.5500	0.7000	0.7625	0.7875	0.0162	0.0288	0.0800	0.3313	0.7275	95
$n = 90$											
5	0.0111	0.0111	0.0111	0.0111	0.0444	0.0011	0.0011	0.0011	0.0011	0.0022	5
10	0.0111	0.0111	0.0111	0.0333	0.1222	0.0011	0.0011	0.0011	0.0022	0.0056	10
90	0.1444	0.3556	0.6667	0.7333	0.7667	0.0111	0.0178	0.0456	0.1789	0.7100	90
95	0.2000	0.5111	0.7000	0.7556	0.7889	0.0144	0.0244	0.0678	0.2788	0.7244	95
$n = 100$											
5	0.0100	0.0100	0.0100	0.0100	0.0400	0.0010	0.0010	0.0010	0.0010	0.0020	5
10	0.0100	0.0100	0.0100	0.0300	0.1100	0.0010	0.0010	0.0010	0.0020	0.0040	10
90	0.1300	0.3000	0.6600	0.7300	0.7700	0.0100	0.0160	0.0390	0.1500	0.7010	90
95	0.1800	0.4500	0.6900	0.7500	0.7800	0.0130	0.0220	0.0570	0.2360	0.7210	95

STATISTICAL TABLES

Table B.8 Percentiles of q_n with $p = 0.9$.

%	B = 2	B = 4	B = 6	B = 8	B = 10	B = 2	B = 4	B = 6	B = 8	B = 10	%
n = 20											
5	0.0500	0.0500	0.0500	0.0500	0.0500	0.0050	0.0050	0.0050	0.0050	0.0100	5
10	0.0500	0.0500	0.0500	0.0500	0.0500	0.0050	0.0050	0.0050	0.0100	0.0300	10
90	0.4500	0.6500	0.7500	0.8500	0.8500	0.0550	0.1050	0.3450	0.7850	0.8350	90
95	0.5500	0.7000	0.8000	0.8500	0.9000	0.0750	0.1550	0.5400	0.8050	0.8450	95
n = 30											
5	0.0333	0.0333	0.0333	0.0333	0.0333	0.0033	0.0033	0.0033	0.0033	0.0067	5
10	0.0333	0.0333	0.0333	0.0333	0.0333	0.0033	0.0033	0.0033	0.0067	0.0200	10
90	0.4000	0.6333	0.7667	0.8333	0.8667	0.0367	0.0633	0.1967	0.7600	0.8233	90
95	0.4667	0.7000	0.8000	0.8667	0.9000	0.0467	0.0933	0.3033	0.7900	0.8367	95
n = 40											
5	0.0250	0.0250	0.0250	0.0250	0.0250	0.0025	0.0025	0.0025	0.0025	0.0050	5
10	0.0250	0.0250	0.0250	0.0250	0.0250	0.0025	0.0025	0.0025	0.0050	0.0150	10
90	0.3250	0.6000	0.7500	0.8250	0.8500	0.0275	0.0450	0.1300	0.5600	0.8200	90
95	0.4250	0.6750	0.8000	0.8500	0.8750	0.0350	0.0650	0.2000	0.7700	0.8300	95
n = 50											
5	0.0200	0.0200	0.0200	0.0200	0.0200	0.0020	0.0020	0.0020	0.0020	0.0040	5
10	0.0200	0.0200	0.0200	0.0200	0.0400	0.0020	0.0020	0.0020	0.0040	0.0100	10
90	0.2600	0.5400	0.7600	0.8200	0.8600	0.0200	0.0340	0.0940	0.3960	0.8140	90
95	0.3600	0.6400	0.7800	0.8400	0.8800	0.0280	0.0500	0.1440	0.6280	0.8240	95
n = 60											
5	0.0167	0.0167	0.0167	0.0167	0.0167	0.0017	0.0017	0.0017	0.0017	0.0033	5
10	0.0167	0.0167	0.0167	0.0167	0.0333	0.0017	0.0017	0.0017	0.0033	0.0083	10
90	0.2167	0.4500	0.7500	0.8167	0.8500	0.0167	0.0267	0.0733	0.2950	0.8100	90
95	0.3000	0.6167	0.7833	0.8333	0.8667	0.0217	0.0383	0.1100	0.4617	0.8200	95
n = 70											
5	0.0143	0.0143	0.0143	0.0143	0.0143	0.0014	0.0014	0.0014	0.0014	0.0029	5
10	0.0143	0.0143	0.0143	0.0143	0.0429	0.0014	0.0014	0.0014	0.0029	0.0071	10
90	0.1714	0.3857	0.7429	0.8143	0.8571	0.0143	0.0229	0.0586	0.2314	0.8043	90
95	0.2571	0.5714	0.7714	0.8429	0.8714	0.0186	0.0314	0.0871	0.3643	0.8157	95
n = 80											
5	0.0125	0.0125	0.0125	0.0125	0.0125	0.0012	0.0012	0.0012	0.0012	0.0025	5
10	0.0125	0.0125	0.0125	0.0125	0.0500	0.0012	0.0012	0.0012	0.0025	0.0050	10
90	0.1500	0.3250	0.7250	0.8125	0.8500	0.0125	0.0188	0.0487	0.1875	0.7987	90
95	0.2125	0.5000	0.7625	0.8375	0.8625	0.0162	0.0262	0.0725	0.2925	0.8125	95
n = 90											
5	0.0111	0.0111	0.0111	0.0111	0.0222	0.0011	0.0011	0.0011	0.0011	0.0022	5
10	0.0111	0.0111	0.0111	0.0222	0.0444	0.0011	0.0011	0.0011	0.0022	0.0044	10
90	0.1333	0.2889	0.7222	0.8111	0.8444	0.0111	0.0167	0.0411	0.1556	0.7811	90
95	0.1889	0.4333	0.7556	0.8333	0.8667	0.0144	0.0233	0.0611	0.2456	0.8089	95
n = 100											
5	0.0100	0.0100	0.0100	0.0100	0.0200	0.0010	0.0010	0.0010	0.0010	0.0020	5
10	0.0100	0.0100	0.0100	0.0200	0.0500	0.0010	0.0010	0.0010	0.0020	0.0040	10
90	0.1200	0.2500	0.7100	0.8100	0.8500	0.0100	0.0140	0.0350	0.1320	0.6520	90
95	0.1600	0.3800	0.7500	0.8300	0.8600	0.0130	0.0200	0.0520	0.2050	0.8040	95

Anexo 3

STATISTICAL TABLES

263

Table C.1 Percentiles of correlation coefficient r with $\rho = 0.2$.

%	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 10$	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 10$	%
$n = 20$						$n = 200$					
1	0.3964	0.4598	0.5210	0.5364	0.5546	0.8866	0.9357	0.9558	0.9631	0.9668	1
5	0.5262	0.6192	0.6542	0.6837	0.6965	0.9362	0.9612	0.9709	0.9760	0.9780	5
10	0.6299	0.7065	0.7345	0.7632	0.7728	0.9515	0.9710	0.9774	0.9808	0.9827	10
20	0.7392	0.7949	0.8197	0.8368	0.8461	0.9649	0.9782	0.9831	0.9861	0.9877	20
$n = 30$						$n = 300$					
1	0.4036	0.5015	0.5870	0.6168	0.6625	0.9278	0.9523	0.9689	0.9753	0.9784	1
5	0.6050	0.6885	0.7629	0.7816	0.8058	0.9561	0.9740	0.9810	0.9844	0.9866	5
10	0.6983	0.7782	0.8223	0.8375	0.8517	0.9670	0.9798	0.9853	0.9876	0.9892	10
20	0.7942	0.8442	0.8750	0.8889	0.8963	0.9761	0.9855	0.9889	0.9908	0.9918	20
$n = 40$						$n = 400$					
1	0.4905	0.6246	0.6722	0.7230	0.7670	0.9466	0.9703	0.9787	0.9827	0.9840	1
5	0.6889	0.7767	0.8192	0.8413	0.8613	0.9677	0.9812	0.9860	0.9884	0.9896	5
10	0.7664	0.8375	0.8690	0.8854	0.8975	0.9748	0.9854	0.9890	0.9909	0.9918	10
20	0.8362	0.8852	0.9089	0.9191	0.9276	0.9813	0.9891	0.9918	0.9932	0.9939	20
$n = 50$						$n = 500$					
1	0.5515	0.6983	0.7420	0.7967	0.8188	0.9583	0.9761	0.9830	0.9858	0.9876	1
5	0.7340	0.8321	0.8649	0.8883	0.8932	0.9752	0.9849	0.9886	0.9907	0.9919	5
10	0.8022	0.8736	0.8995	0.9136	0.9201	0.9807	0.9882	0.9911	0.9926	0.9935	10
20	0.8625	0.9098	0.9262	0.9388	0.9421	0.9858	0.9912	0.9934	0.9945	0.9952	20
$n = 60$						$n = 600$					
1	0.6132	0.7510	0.8104	0.8333	0.8477	0.9659	0.9803	0.9858	0.9888	0.9893	1
5	0.7766	0.8554	0.8880	0.9054	0.9167	0.9781	0.9874	0.9908	0.9924	0.9931	5
10	0.8319	0.8932	0.9163	0.9283	0.9371	0.9833	0.9901	0.9929	0.9939	0.9946	10
20	0.8845	0.9255	0.9400	0.9492	0.9551	0.9878	0.9926	0.9947	0.9955	0.9960	20
$n = 70$						$n = 700$					
1	0.6817	0.7858	0.8373	0.8715	0.8937	0.9683	0.9832	0.9884	0.9897	0.9917	1
5	0.8081	0.8738	0.9065	0.9201	0.9305	0.9815	0.9892	0.9922	0.9935	0.9943	5
10	0.8574	0.9082	0.9288	0.9402	0.9477	0.9858	0.9914	0.9938	0.9950	0.9955	10
20	0.9025	0.9356	0.9494	0.9568	0.9619	0.9894	0.9936	0.9953	0.9961	0.9966	20
$n = 80$						$n = 800$					
1	0.7184	0.8212	0.8614	0.8977	0.8999	0.9738	0.9852	0.9898	0.9915	0.9922	1
5	0.8349	0.8941	0.9227	0.9332	0.9391	0.9839	0.9906	0.9932	0.9943	0.9950	5
10	0.8779	0.9206	0.9394	0.9491	0.9527	0.9874	0.9926	0.9946	0.9955	0.9961	10
20	0.9141	0.9454	0.9562	0.9625	0.9663	0.9907	0.9945	0.9960	0.9966	0.9970	20
$n = 90$						$n = 900$					
1	0.7309	0.8383	0.8826	0.9115	0.9161	0.9773	0.9864	0.9901	0.9922	0.9928	1
5	0.8535	0.9095	0.9313	0.9404	0.9471	0.9857	0.9916	0.9938	0.9949	0.9956	5
10	0.8909	0.9321	0.9464	0.9557	0.9606	0.9887	0.9932	0.9951	0.9960	0.9964	10
20	0.9220	0.9504	0.9615	0.9682	0.9710	0.9918	0.9951	0.9963	0.9970	0.9973	20
$n = 100$						$n = 1000$					
1	0.7691	0.8517	0.8877	0.9214	0.9269	0.9787	0.9879	0.9915	0.9934	0.9938	1
5	0.8698	0.9193	0.9353	0.9490	0.9537	0.9872	0.9923	0.9945	0.9954	0.9960	5
10	0.9033	0.9390	0.9514	0.9602	0.9644	0.9901	0.9939	0.9956	0.9965	0.9969	10
20	0.9314	0.9562	0.9654	0.9707	0.9744	0.9927	0.9955	0.9967	0.9973	0.9976	20

STATISTICAL TABLES

Table C.2 Percentiles of correlation coefficient r with $p = 0.3$.

%	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 10$	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 10$	%
$n = 200$											
$n = 20$											
1	0.4337	0.5259	0.5973	0.6471	0.6621	0.9330	0.9620	0.9738	0.9778	0.9801	1
5	0.6111	0.7110	0.7644	0.7901	0.8084	0.9582	0.9757	0.9826	0.9849	0.9872	5
10	0.7099	0.7869	0.8240	0.8453	0.8554	0.9685	0.9810	0.9861	0.9882	0.9898	10
20	0.7930	0.8509	0.8787	0.8923	0.9027	0.9770	0.9861	0.9897	0.9912	0.9923	20
$n = 300$											
$n = 30$											
1	0.5187	0.6661	0.7516	0.7633	0.8074	0.9514	0.9775	0.9826	0.9852	0.9871	1
5	0.7129	0.8099	0.8505	0.8685	0.8874	0.9720	0.9846	0.9888	0.9906	0.9911	5
10	0.7851	0.8604	0.8868	0.9022	0.9158	0.9787	0.9879	0.9910	0.9925	0.9931	10
20	0.8511	0.9024	0.9211	0.9327	0.9393	0.9843	0.9908	0.9932	0.9943	0.9949	20
$n = 400$											
$n = 40$											
1	0.6270	0.7612	0.8212	0.8441	0.8730	0.9656	0.9832	0.9866	0.9890	0.9900	1
5	0.7883	0.8674	0.8973	0.9104	0.9200	0.9801	0.9888	0.9912	0.9927	0.9936	5
10	0.8424	0.8995	0.9228	0.9320	0.9392	0.9845	0.9910	0.9932	0.9941	0.9948	10
20	0.8867	0.9288	0.9447	0.9508	0.9568	0.9884	0.9932	0.9949	0.9956	0.9961	20
$n = 500$											
$n = 50$											
1	0.7346	0.8101	0.8636	0.8919	0.9043	0.9751	0.9860	0.9890	0.9915	0.9924	1
5	0.8365	0.8958	0.9184	0.9328	0.9424	0.9839	0.9905	0.9931	0.9942	0.9951	5
10	0.8777	0.9199	0.9388	0.9490	0.9545	0.9875	0.9927	0.9945	0.9954	0.9960	10
20	0.9118	0.9423	0.9559	0.9624	0.9661	0.9908	0.9945	0.9959	0.9965	0.9970	20
$n = 600$											
$n = 60$											
1	0.7351	0.8453	0.8908	0.9148	0.9237	0.9789	0.9883	0.9915	0.9929	0.9940	1
5	0.8530	0.9129	0.9343	0.9461	0.9536	0.9868	0.9924	0.9944	0.9953	0.9958	5
10	0.8931	0.9341	0.9505	0.9584	0.9633	0.9898	0.9940	0.9955	0.9962	0.9966	10
20	0.9229	0.9524	0.9639	0.9697	0.9728	0.9924	0.9955	0.9966	0.9972	0.9975	20
$n = 700$											
$n = 70$											
1	0.7980	0.8756	0.8999	0.9291	0.9379	0.9827	0.9901	0.9928	0.9938	0.9945	1
5	0.8844	0.9280	0.9436	0.9536	0.9607	0.9887	0.9935	0.9952	0.9959	0.9964	5
10	0.9116	0.9459	0.9571	0.9644	0.9696	0.9912	0.9949	0.9961	0.9967	0.9971	10
20	0.9365	0.9601	0.9687	0.9736	0.9769	0.9934	0.9962	0.9971	0.9975	0.9979	20
$n = 800$											
$n = 80$											
1	0.8172	0.8973	0.9270	0.9365	0.9482	0.9834	0.9912	0.9939	0.9947	0.9950	1
5	0.8975	0.9375	0.9540	0.9596	0.9659	0.9898	0.9943	0.9958	0.9964	0.9968	5
10	0.9225	0.9521	0.9633	0.9686	0.9726	0.9921	0.9954	0.9966	0.9972	0.9975	10
20	0.9432	0.9651	0.9731	0.9773	0.9793	0.9941	0.9965	0.9975	0.9979	0.9981	20
$n = 900$											
$n = 90$											
1	0.8400	0.9087	0.9366	0.9460	0.9519	0.9869	0.9922	0.9941	0.9953	0.9958	1
5	0.9077	0.9451	0.9600	0.9651	0.9681	0.9915	0.9949	0.9963	0.9970	0.9972	5
10	0.9312	0.9571	0.9684	0.9727	0.9749	0.9931	0.9958	0.9971	0.9975	0.9978	10
20	0.9497	0.9692	0.9765	0.9801	0.9817	0.9949	0.9968	0.9978	0.9981	0.9984	20
$n = 1000$											
$n = 100$											
1	0.8466	0.9149	0.9363	0.9511	0.9586	0.9884	0.9932	0.9948	0.9957	0.9963	1
5	0.9165	0.9508	0.9626	0.9700	0.9723	0.9921	0.9954	0.9966	0.9971	0.9975	5
10	0.9373	0.9626	0.9712	0.9763	0.9781	0.9938	0.9964	0.9973	0.9977	0.9980	10
20	0.9552	0.9722	0.9784	0.9825	0.9841	0.9953	0.9972	0.9980	0.9983	0.9985	20

STATISTICAL TABLES

265

Table C.3 Percentiles of correlation coefficient r with $p = 0.4$.

Table 1. Critical values of T with $p = 0.4$.											
%	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 10$	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 10$	%
$n = 20$											
1	0.5168	0.6571	0.7436	0.7686	0.7833	0.9494	0.9740	0.9807	0.9834	0.9849	1
5	0.6936	0.7938	0.8448	0.8600	0.8693	0.9700	0.9830	0.9875	0.9895	0.9903	5
10	0.7644	0.8478	0.8801	0.8929	0.9012	0.9767	0.9864	0.9901	0.9916	0.9922	10
20	0.8357	0.8935	0.9124	0.9219	0.9309	0.9833	0.9901	0.9925	0.9937	0.9942	20
$n = 30$											
1	0.6468	0.7820	0.8297	0.8634	0.8738	0.9713	0.9836	0.9880	0.9898	0.9904	1
5	0.7933	0.8721	0.9009	0.9143	0.9215	0.9813	0.9890	0.9916	0.9930	0.9936	5
10	0.8450	0.9021	0.9244	0.9334	0.9390	0.9851	0.9913	0.9934	0.9945	0.9950	10
20	0.8906	0.9286	0.9451	0.9533	0.9561	0.9889	0.9935	0.9950	0.9958	0.9963	20
$n = 40$											
1	0.7596	0.8493	0.8843	0.9036	0.9158	0.9777	0.9878	0.9908	0.9919	0.9930	1
5	0.8502	0.9051	0.9306	0.9394	0.9448	0.9851	0.9920	0.9938	0.9946	0.9953	5
10	0.8843	0.9270	0.9472	0.9535	0.9585	0.9886	0.9936	0.9951	0.9959	0.9963	10
20	0.9182	0.9482	0.9604	0.9663	0.9693	0.9915	0.9952	0.9963	0.9969	0.9972	20
$n = 50$											
1	0.8040	0.8755	0.9106	0.9265	0.9348	0.9828	0.9906	0.9922	0.9935	0.9945	1
5	0.8771	0.9264	0.9459	0.9529	0.9572	0.9887	0.9934	0.9951	0.9957	0.9964	5
10	0.9065	0.9448	0.9576	0.9635	0.9665	0.9911	0.9948	0.9961	0.9967	0.9971	10
20	0.9329	0.9591	0.9687	0.9730	0.9747	0.9934	0.9960	0.9971	0.9975	0.9978	20
$n = 60$											
1	0.8162	0.9026	0.9259	0.9396	0.9501	0.9859	0.9918	0.9937	0.9949	0.9956	1
5	0.8962	0.9389	0.9534	0.9613	0.9671	0.9908	0.9946	0.9957	0.9966	0.9969	5
10	0.9224	0.9523	0.9641	0.9700	0.9736	0.9927	0.9956	0.9967	0.9972	0.9975	10
20	0.9444	0.9656	0.9739	0.9777	0.9803	0.9945	0.9966	0.9975	0.9979	0.9982	20
$n = 70$											
1	0.8500	0.9155	0.9381	0.9501	0.9560	0.9881	0.9932	0.9946	0.9954	0.9959	1
5	0.9137	0.9481	0.9607	0.9672	0.9712	0.9920	0.9954	0.9965	0.9971	0.9973	5
10	0.9349	0.9605	0.9694	0.9745	0.9776	0.9936	0.9963	0.9972	0.9977	0.9979	10
20	0.9531	0.9715	0.9779	0.9812	0.9832	0.9952	0.9972	0.9979	0.9982	0.9984	20
$n = 80$											
1	0.8593	0.9266	0.9467	0.9541	0.9616	0.9893	0.9939	0.9956	0.9961	0.9964	1
5	0.9219	0.9563	0.9669	0.9726	0.9750	0.9931	0.9959	0.9970	0.9974	0.9976	5
10	0.9415	0.9655	0.9743	0.9779	0.9803	0.9946	0.9968	0.9976	0.9979	0.9981	10
20	0.9576	0.9750	0.9807	0.9837	0.9854	0.9959	0.9976	0.9982	0.9984	0.9986	20
$n = 90$											
1	0.8919	0.9378	0.9541	0.9634	0.9652	0.9911	0.9945	0.9963	0.9966	0.9968	1
5	0.9356	0.9609	0.9710	0.9755	0.9775	0.9938	0.9964	0.9973	0.9977	0.9979	5
10	0.9512	0.9699	0.9768	0.9801	0.9825	0.9950	0.9971	0.9978	0.9981	0.9983	10
20	0.9642	0.9778	0.9829	0.9854	0.9870	0.9963	0.9978	0.9984	0.9986	0.9987	20
$n = 100$											
1	0.8940	0.9463	0.9587	0.9654	0.9692	0.9915	0.9952	0.9963	0.9971	0.9973	1
5	0.9417	0.9654	0.9732	0.9782	0.9795	0.9946	0.9968	0.9976	0.9980	0.9982	5
10	0.9554	0.9731	0.9792	0.9828	0.9840	0.9957	0.9974	0.9981	0.9984	0.9985	10
20	0.9674	0.9800	0.9844	0.9871	0.9879	0.9967	0.9981	0.9985	0.9988	0.9989	20

STATISTICAL TABLES

Table C.4 Percentiles of correlation coefficient r with $p = 0.5$.

%	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 10$	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 10$	%
$n = 20$											
1	0.5975	0.7510	0.7975	0.8410	0.8555	0.9648	0.9811	0.9855	0.9868	0.9880	1
5	0.7492	0.8427	0.8787	0.8990	0.9080	0.9780	0.9876	0.9905	0.9915	0.9922	5
10	0.8194	0.8819	0.9067	0.9220	0.9280	0.9832	0.9900	0.9924	0.9933	0.9938	10
20	0.8696	0.9143	0.9330	0.9413	0.9461	0.9875	0.9923	0.9942	0.9949	0.9954	20
$n = 30$											
1	0.7221	0.8361	0.8730	0.9002	0.9112	0.9779	0.9873	0.9904	0.9915	0.9921	1
5	0.8449	0.9003	0.9230	0.9364	0.9445	0.9858	0.9918	0.9936	0.9944	0.9950	5
10	0.8799	0.9261	0.9404	0.9484	0.9551	0.9889	0.9935	0.9948	0.9955	0.9960	10
20	0.9152	0.9466	0.9567	0.9624	0.9666	0.9917	0.9951	0.9961	0.9966	0.9970	20
$n = 40$											
1	0.7955	0.8827	0.9148	0.9286	0.9311	0.9842	0.9908	0.9928	0.9936	0.9942	1
5	0.8823	0.9285	0.9452	0.9551	0.9577	0.9896	0.9940	0.9952	0.9958	0.9962	5
10	0.9106	0.9457	0.9571	0.9634	0.9673	0.9915	0.9952	0.9962	0.9967	0.9970	10
20	0.9364	0.9605	0.9688	0.9724	0.9752	0.9935	0.9963	0.9971	0.9975	0.9977	20
$n = 50$											
1	0.8455	0.9104	0.9307	0.9465	0.9481	0.9878	0.9924	0.9945	0.9948	0.9953	1
5	0.9084	0.9454	0.9563	0.9651	0.9676	0.9917	0.9949	0.9962	0.9966	0.9970	5
10	0.9293	0.9578	0.9665	0.9715	0.9739	0.9935	0.9959	0.9970	0.9973	0.9976	10
20	0.9482	0.9692	0.9753	0.9788	0.9806	0.9950	0.9970	0.9977	0.9980	0.9982	20
$n = 60$											
1	0.8635	0.9270	0.9456	0.9538	0.9601	0.9891	0.9940	0.9954	0.9958	0.9963	1
5	0.9215	0.9534	0.9658	0.9703	0.9730	0.9931	0.9959	0.9969	0.9973	0.9975	5
10	0.9394	0.9631	0.9728	0.9766	0.9787	0.9944	0.9967	0.9975	0.9978	0.9980	10
20	0.9565	0.9732	0.9798	0.9825	0.9843	0.9957	0.9975	0.9981	0.9983	0.9985	20
$n = 70$											
1	0.8887	0.9422	0.9548	0.9637	0.9655	0.9910	0.9946	0.9961	0.9962	0.9968	1
5	0.9370	0.9622	0.9706	0.9745	0.9773	0.9940	0.9965	0.9974	0.9976	0.9978	5
10	0.9509	0.9705	0.9763	0.9798	0.9822	0.9952	0.9972	0.9979	0.9981	0.9983	10
20	0.9634	0.9784	0.9825	0.9852	0.9865	0.9963	0.9978	0.9984	0.9986	0.9987	20
$n = 80$											
1	0.9068	0.9452	0.9592	0.9669	0.9694	0.9920	0.9955	0.9963	0.9969	0.9970	1
5	0.9455	0.9660	0.9746	0.9787	0.9808	0.9948	0.9970	0.9975	0.9979	0.9981	5
10	0.9570	0.9736	0.9797	0.9830	0.9844	0.9958	0.9976	0.9981	0.9983	0.9985	10
20	0.9683	0.9802	0.9849	0.9872	0.9882	0.9968	0.9981	0.9985	0.9987	0.9989	20
$n = 90$											
1	0.9140	0.9548	0.9656	0.9704	0.9722	0.9930	0.9959	0.9967	0.9972	0.9975	1
5	0.9495	0.9709	0.9777	0.9809	0.9821	0.9953	0.9973	0.9979	0.9981	0.9983	5
10	0.9609	0.9769	0.9820	0.9852	0.9857	0.9963	0.9978	0.9983	0.9985	0.9987	10
20	0.9715	0.9825	0.9867	0.9886	0.9893	0.9972	0.9983	0.9987	0.9989	0.9990	20
$n = 100$											
1	0.9258	0.9569	0.9703	0.9735	0.9772	0.9941	0.9964	0.9972	0.9974	0.9977	1
5	0.9547	0.9729	0.9799	0.9828	0.9843	0.9960	0.9976	0.9981	0.9984	0.9985	5
10	0.9648	0.9788	0.9839	0.9863	0.9877	0.9968	0.9981	0.9985	0.9987	0.9988	10
20	0.9740	0.9845	0.9878	0.9897	0.9908	0.9975	0.9985	0.9988	0.9990	0.9991	20

STATISTICAL TABLES

267

Table C.5 Percentiles of correlation coefficient r with $p = 0.6$.

%	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 10$	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 10$	%
$n = 20$						$n = 200$					
1	0.6799	0.8050	0.8480	0.8707	0.8859	0.9735	0.9847	0.9880	0.9897	0.9904	1
5	0.8009	0.8804	0.9042	0.9162	0.9260	0.9827	0.9895	0.9919	0.9928	0.9936	5
10	0.8456	0.9087	0.9253	0.9350	0.9411	0.9864	0.9916	0.9936	0.9943	0.9948	10
20	0.8906	0.9318	0.9444	0.9517	0.9570	0.9897	0.9936	0.9953	0.9957	0.9961	20
$n = 30$						$n = 300$					
1	0.7949	0.8783	0.9075	0.9159	0.9247	0.9835	0.9904	0.9920	0.9928	0.9931	1
5	0.8738	0.9208	0.9406	0.9470	0.9509	0.9889	0.9934	0.9946	0.9953	0.9955	5
10	0.9023	0.9400	0.9522	0.9580	0.9610	0.9911	0.9947	0.9957	0.9963	0.9965	10
20	0.9295	0.9555	0.9646	0.9693	0.9708	0.9931	0.9959	0.9968	0.9972	0.9974	20
$n = 40$						$n = 400$					
1	0.8514	0.9138	0.9319	0.9419	0.9455	0.9878	0.9923	0.9941	0.9949	0.9951	1
5	0.9081	0.9431	0.9563	0.9620	0.9656	0.9918	0.9950	0.9960	0.9965	0.9967	5
10	0.9287	0.9563	0.9653	0.9705	0.9727	0.9934	0.9959	0.9967	0.9971	0.9974	10
20	0.9476	0.9675	0.9742	0.9782	0.9794	0.9950	0.9969	0.9975	0.9979	0.9981	20
$n = 50$						$n = 500$					
1	0.8755	0.9355	0.9467	0.9551	0.9589	0.9896	0.9943	0.9953	0.9956	0.9963	1
5	0.9234	0.9573	0.9664	0.9708	0.9718	0.9934	0.9960	0.9968	0.9971	0.9974	5
10	0.9418	0.9663	0.9735	0.9762	0.9777	0.9946	0.9968	0.9975	0.9977	0.9980	10
20	0.9568	0.9745	0.9796	0.9819	0.9831	0.9959	0.9975	0.9981	0.9983	0.9985	20
$n = 60$						$n = 600$					
1	0.9049	0.9457	0.9554	0.9618	0.9666	0.9918	0.9950	0.9962	0.9964	0.9968	1
5	0.9394	0.9643	0.9719	0.9755	0.9781	0.9946	0.9967	0.9974	0.9977	0.9978	5
10	0.9531	0.9719	0.9776	0.9799	0.9825	0.9957	0.9973	0.9979	0.9981	0.9983	10
20	0.9656	0.9787	0.9830	0.9848	0.9866	0.9967	0.9979	0.9984	0.9986	0.9987	20
$n = 70$						$n = 700$					
1	0.9217	0.9532	0.9632	0.9685	0.9722	0.9931	0.9958	0.9965	0.9971	0.9972	1
5	0.9497	0.9698	0.9754	0.9784	0.9807	0.9953	0.9972	0.9977	0.9980	0.9981	5
10	0.9606	0.9761	0.9808	0.9833	0.9848	0.9962	0.9977	0.9982	0.9984	0.9985	10
20	0.9706	0.9819	0.9855	0.9874	0.9885	0.9971	0.9982	0.9986	0.9988	0.9989	20
$n = 80$						$n = 800$					
1	0.9274	0.9598	0.9684	0.9729	0.9756	0.9939	0.9964	0.9971	0.9972	0.9975	1
5	0.9544	0.9735	0.9789	0.9819	0.9836	0.9958	0.9976	0.9980	0.9982	0.9984	5
10	0.9644	0.9790	0.9835	0.9857	0.9868	0.9967	0.9980	0.9984	0.9986	0.9987	10
20	0.9742	0.9842	0.9875	0.9892	0.9899	0.9975	0.9985	0.9988	0.9989	0.9990	20
$n = 90$						$n = 900$					
1	0.9362	0.9626	0.9726	0.9757	0.9774	0.9946	0.9965	0.9975	0.9977	0.9979	1
5	0.9603	0.9761	0.9812	0.9840	0.9848	0.9964	0.9978	0.9983	0.9984	0.9986	5
10	0.9692	0.9811	0.9852	0.9873	0.9881	0.9971	0.9982	0.9986	0.9987	0.9988	10
20	0.9770	0.9858	0.9887	0.9906	0.9910	0.9977	0.9986	0.9989	0.9991	0.9991	20
$n = 100$						$n = 1000$					
1	0.9452	0.9678	0.9753	0.9777	0.9798	0.9950	0.9971	0.9977	0.9979	0.9981	1
5	0.9657	0.9793	0.9829	0.9858	0.9864	0.9967	0.9981	0.9984	0.9986	0.9987	5
10	0.9726	0.9833	0.9867	0.9886	0.9893	0.9973	0.9984	0.9987	0.9989	0.9990	10
20	0.9792	0.9873	0.9899	0.9914	0.9919	0.9980	0.9988	0.9990	0.9991	0.9992	20

Table C.6 Percentiles of correlation coefficient r with $p = 0.7$.

		B = 2	B = 4	B = 6	B = 8	B = 10	B = 2	B = 4	B = 6	B = 8	B = 10	%
		n = 20					n = 200					
1	5	0.7237	0.8509	0.8772	0.8901	0.8987	0.9797	0.9880	0.9890	0.9903	0.9913	1
5	10	0.8374	0.9008	0.9198	0.9282	0.9317	0.9861	0.9916	0.9929	0.9938	0.9942	5
10	15	0.8747	0.9223	0.9352	0.9441	0.9463	0.9890	0.9933	0.9945	0.9951	0.9954	10
20	25	0.9095	0.9418	0.9518	0.9583	0.9607	0.9915	0.9949	0.9958	0.9963	0.9965	20
		n = 30					n = 300					
1	5	0.8375	0.9016	0.9220	0.9321	0.9367	0.9868	0.9917	0.9934	0.9935	0.9942	1
5	10	0.8979	0.9366	0.9496	0.9564	0.9577	0.9906	0.9943	0.9954	0.9958	0.9961	5
10	15	0.9199	0.9496	0.9598	0.9652	0.9665	0.9926	0.9955	0.9963	0.9968	0.9969	10
20	25	0.9407	0.9622	0.9699	0.9736	0.9746	0.9943	0.9966	0.9972	0.9975	0.9977	20
		n = 40					n = 400					
1	5	0.8837	0.9304	0.9432	0.9503	0.9537	0.9900	0.9940	0.9951	0.9953	0.9956	1
5	10	0.9244	0.9535	0.9636	0.9673	0.9694	0.9931	0.9958	0.9966	0.9968	0.9971	5
10	15	0.9401	0.9635	0.9707	0.9738	0.9754	0.9945	0.9965	0.9972	0.9975	0.9977	10
20	25	0.9564	0.9724	0.9778	0.9799	0.9814	0.9958	0.9974	0.9979	0.9981	0.9983	20
		n = 50					n = 500					
1	5	0.9038	0.9479	0.9560	0.9612	0.9631	0.9923	0.9947	0.9958	0.9963	0.9965	1
5	10	0.9396	0.9644	0.9706	0.9741	0.9754	0.9947	0.9966	0.9973	0.9975	0.9977	5
10	15	0.9529	0.9715	0.9765	0.9793	0.9808	0.9958	0.9973	0.9978	0.9980	0.9981	10
20	25	0.9652	0.9781	0.9822	0.9840	0.9856	0.9967	0.9979	0.9983	0.9985	0.9986	20
		n = 60					n = 600					
1	5	0.9239	0.9595	0.9641	0.9693	0.9695	0.9932	0.9959	0.9964	0.9969	0.9971	1
5	10	0.9506	0.9699	0.9757	0.9792	0.9802	0.9955	0.9973	0.9977	0.9979	0.9981	5
10	15	0.9608	0.9762	0.9808	0.9833	0.9841	0.9964	0.9978	0.9981	0.9983	0.9984	10
20	25	0.9707	0.9820	0.9854	0.9873	0.9877	0.9972	0.9983	0.9986	0.9987	0.9988	20
		n = 70					n = 700					
1	5	0.9376	0.9619	0.9707	0.9732	0.9755	0.9947	0.9966	0.9972	0.9974	0.9974	1
5	10	0.9586	0.9753	0.9800	0.9818	0.9834	0.9961	0.9976	0.9980	0.9982	0.9984	5
10	15	0.9672	0.9798	0.9840	0.9852	0.9867	0.9969	0.9981	0.9984	0.9986	0.9987	10
20	25	0.9759	0.9845	0.9879	0.9889	0.9898	0.9976	0.9985	0.9988	0.9989	0.9990	20
		n = 80					n = 800					
1	5	0.9442	0.9665	0.9740	0.9757	0.9777	0.9950	0.9969	0.9975	0.9978	0.9979	1
5	10	0.9639	0.9780	0.9822	0.9835	0.9849	0.9966	0.9979	0.9983	0.9984	0.9986	5
10	15	0.9711	0.9823	0.9858	0.9870	0.9879	0.9973	0.9983	0.9987	0.9987	0.9988	10
20	25	0.9784	0.9865	0.9891	0.9903	0.9910	0.9979	0.9987	0.9990	0.9990	0.9991	20
		n = 90					n = 900					
1	5	0.9536	0.9718	0.9772	0.9796	0.9797	0.9958	0.9974	0.9978	0.9979	0.9980	1
5	10	0.9683	0.9811	0.9842	0.9864	0.9864	0.9971	0.9982	0.9985	0.9986	0.9987	5
10	15	0.9746	0.9849	0.9872	0.9888	0.9890	0.9976	0.9986	0.9988	0.9989	0.9990	10
20	25	0.9809	0.9883	0.9904	0.9915	0.9919	0.9982	0.9989	0.9991	0.9991	0.9992	20
		n = 100					n = 1000					
1	5	0.9539	0.9732	0.9798	0.9812	0.9828	0.9961	0.9977	0.9980	0.9982	0.9983	1
5	10	0.9713	0.9829	0.9862	0.9869	0.9883	0.9974	0.9984	0.9986	0.9988	0.9989	5
10	15	0.9769	0.9862	0.9887	0.9896	0.9905	0.9979	0.9987	0.9989	0.9990	0.9991	10
20	25	0.9828	0.9894	0.9914	0.9923	0.9928	0.9983	0.9990	0.9992	0.9992	0.9993	20

STATISTICAL TABLES

269

Table C.7 Percentiles of correlation coefficient r with $p = 0.8$.

%	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 10$	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 10$	%
$n = 20$						$n = 200$					
1	0.7699	0.8675	0.8953	0.8998	0.9153	0.9827	0.9894	0.9911	0.9916	0.9920	1
5	0.8642	0.9171	0.9320	0.9353	0.9433	0.9880	0.9927	0.9940	0.9945	0.9945	5
10	0.8929	0.9347	0.9449	0.9490	0.9541	0.9904	0.9940	0.9951	0.9956	0.9956	10
20	0.9220	0.9505	0.9591	0.9626	0.9653	0.9927	0.9954	0.9962	0.9966	0.9967	20
$n = 30$						$n = 300$					
1	0.8661	0.9224	0.9365	0.9400	0.9415	0.9889	0.9929	0.9939	0.9942	0.9946	1
5	0.9111	0.9476	0.9581	0.9585	0.9608	0.9925	0.9952	0.9958	0.9961	0.9963	5
10	0.9301	0.9585	0.9659	0.9675	0.9690	0.9939	0.9961	0.9967	0.9970	0.9971	10
20	0.9484	0.9682	0.9738	0.9757	0.9767	0.9952	0.9970	0.9974	0.9977	0.9978	20
$n = 40$						$n = 400$					
1	0.8971	0.9384	0.9518	0.9553	0.9596	0.9921	0.9948	0.9953	0.9958	0.9960	1
5	0.9382	0.9606	0.9677	0.9713	0.9726	0.9943	0.9964	0.9970	0.9972	0.9972	5
10	0.9514	0.9682	0.9740	0.9768	0.9778	0.9954	0.9971	0.9976	0.9978	0.9978	10
20	0.9635	0.9760	0.9802	0.9822	0.9830	0.9964	0.9977	0.9981	0.9983	0.9984	20
$n = 50$						$n = 500$					
1	0.9236	0.9538	0.9608	0.9633	0.9662	0.9933	0.9957	0.9964	0.9966	0.9970	1
5	0.9494	0.9697	0.9751	0.9761	0.9773	0.9955	0.9972	0.9976	0.9977	0.9979	5
10	0.9603	0.9758	0.9802	0.9811	0.9821	0.9963	0.9977	0.9980	0.9982	0.9983	10
20	0.9704	0.9815	0.9846	0.9857	0.9865	0.9971	0.9982	0.9985	0.9986	0.9987	20
$n = 60$						$n = 600$					
1	0.9389	0.9617	0.9690	0.9706	0.9717	0.9945	0.9965	0.9971	0.9973	0.9974	1
5	0.9602	0.9743	0.9793	0.9801	0.9817	0.9962	0.9976	0.9980	0.9982	0.9982	5
10	0.9680	0.9793	0.9831	0.9844	0.9857	0.9969	0.9981	0.9984	0.9985	0.9985	10
20	0.9756	0.9842	0.9868	0.9880	0.9890	0.9976	0.9985	0.9987	0.9988	0.9989	20
$n = 70$						$n = 700$					
1	0.9479	0.9672	0.9720	0.9760	0.9755	0.9952	0.9971	0.9974	0.9975	0.9977	1
5	0.9655	0.9789	0.9815	0.9837	0.9839	0.9968	0.9980	0.9982	0.9984	0.9985	5
10	0.9724	0.9828	0.9854	0.9867	0.9873	0.9974	0.9984	0.9985	0.9987	0.9988	10
20	0.9790	0.9866	0.9888	0.9899	0.9904	0.9980	0.9987	0.9989	0.9990	0.9991	20
$n = 80$						$n = 800$					
1	0.9534	0.9725	0.9763	0.9791	0.9801	0.9960	0.9973	0.9978	0.9980	0.9981	1
5	0.9698	0.9809	0.9844	0.9854	0.9863	0.9972	0.9982	0.9985	0.9986	0.9987	5
10	0.9753	0.9849	0.9872	0.9884	0.9892	0.9977	0.9985	0.9988	0.9989	0.9989	10
20	0.9814	0.9883	0.9902	0.9912	0.9917	0.9982	0.9989	0.9990	0.9991	0.9992	20
$n = 90$						$n = 900$					
1	0.9604	0.9747	0.9801	0.9808	0.9818	0.9964	0.9977	0.9981	0.9981	0.9983	1
5	0.9727	0.9829	0.9861	0.9871	0.9878	0.9975	0.9984	0.9987	0.9988	0.9988	5
10	0.9784	0.9864	0.9890	0.9895	0.9901	0.9980	0.9987	0.9989	0.9990	0.9990	10
20	0.9836	0.9897	0.9915	0.9922	0.9924	0.9984	0.9990	0.9992	0.9992	0.9993	20
$n = 100$						$n = 1000$					
1	0.9641	0.9774	0.9813	0.9820	0.9835	0.9969	0.9979	0.9983	0.9983	0.9984	1
5	0.9759	0.9853	0.9877	0.9885	0.9892	0.9978	0.9986	0.9988	0.9989	0.9989	5
10	0.9808	0.9881	0.9898	0.9907	0.9914	0.9982	0.9988	0.9990	0.9991	0.9991	10
20	0.9854	0.9908	0.9921	0.9928	0.9934	0.9986	0.9991	0.9992	0.9993	0.9993	20

STATISTICAL TABLES

Table C.8 Percentiles of correlation coefficient r with $p = 0.9$.

%	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 10$	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 10$	%
$n = 20$						$n = 200$					
1	0.8246	0.8899	0.9101	0.9146	0.9216	0.9865	0.9909	0.9914	0.9921	0.9926	1
5	0.8847	0.9296	0.9400	0.9452	0.9471	0.9903	0.9936	0.9942	0.9947	0.9949	5
10	0.9093	0.9440	0.9514	0.9554	0.9575	0.9920	0.9949	0.9954	0.9958	0.9959	10
20	0.9334	0.9573	0.9634	0.9661	0.9673	0.9939	0.9959	0.9965	0.9967	0.9969	20
$n = 30$						$n = 300$					
1	0.8914	0.9341	0.9412	0.9450	0.9476	0.9905	0.9937	0.9945	0.9949	0.9953	1
5	0.9274	0.9547	0.9608	0.9648	0.9646	0.9936	0.9957	0.9963	0.9965	0.9966	5
10	0.9426	0.9635	0.9684	0.9713	0.9723	0.9948	0.9965	0.9970	0.9972	0.9973	10
20	0.9563	0.9717	0.9753	0.9780	0.9787	0.9959	0.9973	0.9977	0.9979	0.9979	20
$n = 40$						$n = 400$					
1	0.9176	0.9500	0.9567	0.9588	0.9617	0.9928	0.9954	0.9959	0.9960	0.9961	1
5	0.9481	0.9664	0.9709	0.9724	0.9737	0.9951	0.9968	0.9972	0.9973	0.9975	5
10	0.9575	0.9729	0.9768	0.9781	0.9788	0.9960	0.9975	0.9978	0.9978	0.9980	10
20	0.9678	0.9790	0.9823	0.9830	0.9838	0.9969	0.9980	0.9983	0.9984	0.9984	20
$n = 50$						$n = 500$					
1	0.9364	0.9615	0.9672	0.9700	0.9698	0.9944	0.9963	0.9967	0.9969	0.9970	1
5	0.9575	0.9740	0.9769	0.9793	0.9799	0.9963	0.9974	0.9977	0.9978	0.9980	5
10	0.9662	0.9786	0.9816	0.9829	0.9835	0.9970	0.9979	0.9982	0.9983	0.9984	10
20	0.9744	0.9834	0.9858	0.9867	0.9874	0.9976	0.9984	0.9986	0.9987	0.9987	20
$n = 60$						$n = 600$					
1	0.9489	0.9697	0.9724	0.9754	0.9756	0.9950	0.9970	0.9973	0.9973	0.9975	1
5	0.9652	0.9781	0.9813	0.9826	0.9828	0.9967	0.9978	0.9981	0.9982	0.9983	5
10	0.9720	0.9822	0.9848	0.9860	0.9861	0.9973	0.9983	0.9985	0.9986	0.9986	10
20	0.9783	0.9862	0.9882	0.9892	0.9891	0.9979	0.9987	0.9988	0.9989	0.9989	20
$n = 70$						$n = 700$					
1	0.9587	0.9733	0.9761	0.9771	0.9781	0.9961	0.9974	0.9977	0.9978	0.9979	1
5	0.9712	0.9815	0.9838	0.9850	0.9855	0.9973	0.9982	0.9984	0.9985	0.9986	5
10	0.9768	0.9849	0.9870	0.9881	0.9881	0.9978	0.9985	0.9987	0.9988	0.9988	10
20	0.9821	0.9883	0.9900	0.9907	0.9911	0.9983	0.9989	0.9990	0.9991	0.9991	20
$n = 80$						$n = 800$					
1	0.9607	0.9758	0.9792	0.9799	0.9814	0.9965	0.9977	0.9980	0.9982	0.9981	1
5	0.9744	0.9838	0.9856	0.9866	0.9877	0.9976	0.9984	0.9986	0.9987	0.9988	5
10	0.9791	0.9867	0.9886	0.9891	0.9897	0.9981	0.9987	0.9989	0.9989	0.9990	10
20	0.9840	0.9896	0.9912	0.9917	0.9920	0.9985	0.9990	0.9991	0.9992	0.9992	20
$n = 90$						$n = 900$					
1	0.9682	0.9793	0.9821	0.9831	0.9825	0.9970	0.9979	0.9981	0.9982	0.9983	1
5	0.9783	0.9857	0.9874	0.9886	0.9883	0.9979	0.9986	0.9987	0.9988	0.9989	5
10	0.9822	0.9882	0.9899	0.9907	0.9906	0.9983	0.9988	0.9990	0.9991	0.9991	10
20	0.9861	0.9910	0.9921	0.9928	0.9929	0.9986	0.9991	0.9992	0.9993	0.9993	20
$n = 100$						$n = 1000$					
1	0.9712	0.9812	0.9839	0.9843	0.9846	0.9973	0.9982	0.9984	0.9985	0.9985	1
5	0.9799	0.9872	0.9889	0.9891	0.9897	0.9981	0.9987	0.9989	0.9990	0.9990	5
10	0.9836	0.9894	0.9909	0.9913	0.9918	0.9984	0.9990	0.9991	0.9992	0.9992	10
20	0.9873	0.9919	0.9929	0.9933	0.9937	0.9988	0.9992	0.9993	0.9994	0.9994	20

STATISTICAL TABLES

271

Table C.9 Percentiles of correlation coefficient r with $p = 1$.

%	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 10$	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 10$	%
$n = 20$											
1	0.8445	0.9068	0.9185	0.9259	0.9239	0.9876	0.9918	0.9923	0.9928	0.9930	1
5	0.9041	0.9379	0.9459	0.9510	0.9514	0.9917	0.9942	0.9948	0.9952	0.9952	5
10	0.9247	0.9498	0.9574	0.9600	0.9617	0.9932	0.9952	0.9958	0.9961	0.9961	10
20	0.9418	0.9622	0.9667	0.9697	0.9706	0.9948	0.9964	0.9968	0.9969	0.9970	20
$n = 30$											
1	0.9099	0.9415	0.9458	0.9517	0.9528	0.9916	0.9943	0.9952	0.9954	0.9950	1
5	0.9387	0.9603	0.9643	0.9670	0.9688	0.9943	0.9963	0.9965	0.9967	0.9967	5
10	0.9513	0.9672	0.9716	0.9730	0.9746	0.9954	0.9970	0.9972	0.9974	0.9973	10
20	0.9618	0.9750	0.9781	0.9795	0.9804	0.9965	0.9976	0.9978	0.9979	0.9980	20
$n = 40$											
1	0.9339	0.9565	0.9612	0.9616	0.9651	0.9943	0.9960	0.9963	0.9963	0.9963	1
5	0.9552	0.9705	0.9736	0.9745	0.9766	0.9959	0.9972	0.9975	0.9975	0.9976	5
10	0.9632	0.9755	0.9785	0.9797	0.9810	0.9967	0.9977	0.9979	0.9979	0.9981	10
20	0.9717	0.9812	0.9837	0.9844	0.9854	0.9974	0.9982	0.9984	0.9984	0.9985	20
$n = 50$											
1	0.9466	0.9646	0.9683	0.9706	0.9716	0.9951	0.9966	0.9970	0.9970	0.9970	1
5	0.9643	0.9762	0.9783	0.9805	0.9808	0.9967	0.9977	0.9979	0.9980	0.9980	5
10	0.9712	0.9804	0.9826	0.9840	0.9841	0.9973	0.9981	0.9983	0.9984	0.9984	10
20	0.9776	0.9850	0.9867	0.9878	0.9879	0.9979	0.9986	0.9987	0.9988	0.9988	20
$n = 60$											
1	0.9545	0.9712	0.9755	0.9746	0.9757	0.9960	0.9972	0.9975	0.9976	0.9976	1
5	0.9700	0.9805	0.9831	0.9841	0.9834	0.9972	0.9980	0.9983	0.9984	0.9984	5
10	0.9761	0.9842	0.9859	0.9869	0.9868	0.9977	0.9984	0.9986	0.9987	0.9987	10
20	0.9814	0.9877	0.9891	0.9897	0.9900	0.9982	0.9988	0.9989	0.9990	0.9990	20
$n = 70$											
1	0.9654	0.9760	0.9782	0.9794	0.9798	0.9966	0.9976	0.9979	0.9979	0.9980	1
5	0.9746	0.9834	0.9850	0.9862	0.9864	0.9976	0.9984	0.9985	0.9986	0.9986	5
10	0.9795	0.9864	0.9879	0.9887	0.9892	0.9981	0.9987	0.9988	0.9988	0.9989	10
20	0.9844	0.9895	0.9906	0.9914	0.9917	0.9985	0.9990	0.9991	0.9991	0.9991	20
$n = 80$											
1	0.9692	0.9789	0.9807	0.9826	0.9826	0.9972	0.9980	0.9981	0.9982	0.9982	1
5	0.9783	0.9856	0.9869	0.9880	0.9879	0.9980	0.9986	0.9987	0.9988	0.9988	5
10	0.9822	0.9884	0.9894	0.9901	0.9901	0.9983	0.9989	0.9989	0.9990	0.9990	10
20	0.9864	0.9909	0.9918	0.9923	0.9924	0.9987	0.9991	0.9992	0.9992	0.9993	20
$n = 90$											
1	0.9727	0.9822	0.9837	0.9837	0.9837	0.9975	0.9982	0.9984	0.9983	0.9985	1
5	0.9818	0.9877	0.9886	0.9888	0.9891	0.9982	0.9987	0.9989	0.9989	0.9989	5
10	0.9849	0.9900	0.9909	0.9910	0.9912	0.9985	0.9990	0.9991	0.9991	0.9991	10
20	0.9881	0.9921	0.9929	0.9931	0.9933	0.9989	0.9992	0.9993	0.9993	0.9993	20
$n = 100$											
1	0.9754	0.9834	0.9837	0.9852	0.9854	0.9976	0.9984	0.9985	0.9986	0.9985	1
5	0.9832	0.9885	0.9895	0.9900	0.9904	0.9983	0.9989	0.9990	0.9990	0.9990	5
10	0.9862	0.9905	0.9916	0.9919	0.9922	0.9986	0.9991	0.9992	0.9992	0.9992	10
20	0.9891	0.9926	0.9934	0.9939	0.9940	0.9989	0.9993	0.9994	0.9994	0.9994	20

REFERÊNCIAS BIBLIOGRÁFICAS

AARSET, M. V. How to Identify a Bathtub Hazard Rate. **IEEE Transactions on Reliability**, v.R-36, n.1, p.106–108, 1987.

ALBILLOS, A.; GARCIA-TSAO, G. Classification of Cirrhosis: The Clinical Use of HVPG Measurements. **Disease Markers**, v.31, n.3, p.121—128, 2011.

AMDAHL, J. **flexsurvcure: Flexible Parametric Cure Models**, 2017. R package version 0.0.2.

BALAKRISHNAN, N.; PAL, S. Expectation maximization-based likelihood inference for flexible cure rate models with Weibull lifetimes. **Statistical Methods in Medical Research**, v.25, n.4, p.1535—1563, 2016.

BARLOW, R. E.; CAMPO, R. A. Total Time on Test Processes and Applications to Failure Data Analysis. **Reliability and fault tree analysis**, p.451–481, 1975.

BERKSON, J.; GAGE, R. P. Survival curve for cancer patients following treatment. **Journal of the American Statistical Association**, v.47, n.259, p.501–515, 1952.

BIECKER, E. Portal hypertension and gastrointestinal bleeding: Diagnosis, prevention and management. **World Journal of Gastroenterology**, v.19, n.31, p.5035–5050, 2013.

BITTENCOURT, P. L.; FARIAS, A. Q.; STRAUSS, E.; MATTOS, A. A. Variceal bleeding: consensus meeting report from the Brazilian Society of Hepatology. **Arquivos de Gastroenterologia**, v.47, p.202–216, 2010.

CARVALHO, M. S.; ANDREOZZI, V. L.; CODEÇO, C. T.; BARBOSA, M. T. S.; SHIMAKURA, S. E. **Análise de sobrevida: Teoria e aplicações em saúde**. Rio de Janeiro: Editora FIOCRUZ, 2005. 396p.

CHILD, C. G. The liver and portal hypertension. **British Journal of Surgery**, v.51, n.11, p.879–879, 1964.

COELHO, F. F.; PERINI, M. V.; KRUGER, J. A. P.; FONSECA, G. M.; ARAÚJO, R. L. C.; MAKDISSI, F. F.; LUPINACCI, R. M.; HERMAN, P. Tratamento da hemorragia digestiva alta por varizes esofágicas: conceitos atuais. **ABCD - Arquivos Brasileiros de Cirurgia Digestiva**, v.27, n.2, p.138–144, 2014.

COLLETT, D. **Modelling Survival Data in Medical Research**. 2. ed. Florida: Chapman and Hall/CRC, 2003. 391p.

COLOSIMO, E. A.; GIOLO, S. R. **Análise de sobrevivência aplicada**. São Paulo: Blucher, 2006. 370p.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum Likelihood from Incomplete Data via the EM Algorithm. **Journal of the Royal Statistical Society**, v.39, n.1, p.1–38, 1976.

EZEIKE, C. C.; WUNG, W.; MA, D.; LIN, L.; CHRISTOPHER, K. M.; NEUSCHWANDER-TETRI, B. Survival Analysis of Antibiotics in Cirrhosis and Esophageal Varices Patients Based on MELD and Childs-Pugh Score. **Gastroenterology - Click here to go back to the homepage**, v.148, n.4, 2015.

GEHAN, E. A. A Generalized Wilcoxon Test for Comparing Arbitrarily Singly-Censored Samples. **Biometrika**, v.52, n.1/2, p.203–223, 1965.

IBRAHIM, J. G.; CHEN, M. H.; SINHA, D. **Bayesian Survival Analysis**. 1. ed. New York: Springer, 2001. 479p. Springer Series in Statistics.

KALBFLEISCH, J. D.; PRENTICE, R. L. **The Statistical Analysis of Failure Time Data**. 2. ed. New Jersey: John Wiley & Sons, 2002. 439p. Wiley Series in Probability and Statistics.

KANNAN, N.; KUNDU, D.; NAIR, P.; TRIPATHI, R. C. The generalized exponential cure rate model with covariates. **Journal of Applied Statistics**, v.37, n.10, p.1625—1636, 2010.

KAPLAN, E. L.; MEIER, P. Nonparametric Estimation from Incomplete Observations. **Journal of the American Statistical Association**, v.53, n.282, p.457–481, 1958.

KLEINBAUM, D. G.; KLEIN, M. **Survival Analysis: A Self-Learning Text**. 3. ed. New York: Springer-Verlag, 2012. 700p. Statistics for Biology and Health.

KRIGE, J. E. J.; BORNMAN, P. C.; SHAW, J. M. Variceal Recurrence, Rebleeding, and Survival After Endoscopic Injection Sclerotherapy in 287 Alcoholic Cirrhotic Patients With Bleeding Esophageal Varices. **Annals of Surgery**, v.244, n.5, p.764–770, 2006.

LAI, C. D.; XIE, M.; MURTHY, D. N. P. A Modified Weibull Distribution. **IEEE Transactions on Reliability**, v.52, n.1, p.33–37, 2003.

LAINE, L.; COOK, D. Endoscopic ligation compared with sclerotherapy for treatment of esophageal variceal bleeding: a meta-analysis. **Ann Intern Med.**, v.123, p.280–287, 1995.

LI, X.; JIANG, T.; GAO, J. Endoscopic variceal ligation combined with argon plasma coagulation versus ligation alone for the secondary prophylaxis of variceal bleeding: a systematic review and meta-analysis. **European Journal of Gastroenterology & Hepatology**, v.29, n.6, p.621—628, 2017.

MALLER, R. A.; ZHOU, X. **Survival Analysis with Long-Term Survivors**. New York: Wiley, 1996. 278p.

MANTEL, N. Evaluation of survival data and two new rank order statistics arising in its consideration. **Cancer Chemotherapy Rep.**, v.50, p.163–170, 1966.

MCLACHLAN, G. J.; KRISHNAN, T. **The EM Algorithm and Extensions**. 2. ed. New York: Wiley-Interscience, 2008. 359p. Wiley Series in Probability and Statistics.

MUDHOLKAR, G. S.; SRIVASTAVA, D. K. Exponentiated Weibull Family for Analyzing Bathtub Failure-Rate Data. **IEEE Transactions on Reliability**, v.42, n.2, p.299–302, 1993.

MUDHOLKAR, G. S.; SRIVASTAVA, D. K.; KOLLI, G. D. A Generalization of the Weibull Distribution With Application to the Analysis of Survival Data. **Journal of the American Statistical Association**, v.91, n.436, p.1575–1583, 1996.

OLIVEIRA, R. A.; SILVA, G. F.; SILVEIRA, L. V. A. Aplicação do método de Buckley-James como alternativa ao modelo de Cox na violação da pressuposição de riscos proporcionais. **Rev. Bras. Biom.**, v.33, n.3, p.395–402, 2015.

OLIVEIRA-JR, M. R.; MOREIRA, F.; LOUZADA, F. The zero-inflated promotion cure rate model applied to financial data on time to default. **General & Applied Economics**, v.5, p.1–14, 2017.

PUGH, R. N. H.; MURRAY-LYON, I. M.; DAWSON, J. L.; PIETRONI, M. C.; WILLIAMS, R. Transection of the oesophagus for bleeding oesophageal varices. **British Journal of Surgery**, v.60, n.8, p.646–649, 1973.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing, Vienna, Austria, 2017.

RODRIGUES, J.; CANCHO, V. C.; CASTRO, M. Teoria Unificada de Análise de Sobrevivência. In: 18 SINAPE, 2008. **Teoria Unificada de Análise de Sobrevivência**; resumos. São Pedro: Associação Brasileira de Estatística (ABE), 2008. 94.

SCHUPPAN, D.; AFDHAL, N. H. Liver Cirrhosis. **Lancet**, v.371, p.838–851, 2009.

SILVA, C. A. Modelo de regressão Weibull discreto com fração de cura em dados de sobrevivência. Brasília - DF, 2015a. 78p. Dissertação (Mestrado) - Universidade de Brasília.

SILVA, R. P. Modelos Flexíveis de sobrevivência com Fração de Cura: Implementação Computacional. Natal - RN, 2015b. 152p. Dissertação (Mestrado) - Universidade Federal do Rio Grande do Norte.

TOSHIKUNI, N.; TAKUMA, Y.; TSUTSUMI, M. Management of gastroesophageal varices in cirrhotic patients: current status and future directions. **Official Journal of the Mexican Association of Hepatology**, v.15, n.3, p.314–325, 2016.

XIE, M.; TANG, Y.; GOH, T. N. A modified Weibull extension with bathtub-shaped failure rate function. **Reliability Engineering and System Safety**, v.76, p.279–285, 2002.

YAKOVLEV, A. Y.; TSODIKOV, A. D.; ASSELAIN, B. **Stochastic Models of Tumor Latency and Their Biostatistical Applications**. 1. ed. Singapore: World Scientific, 1996. 269p.

APÊNDICES

Apêndice A

Script R - Fração de cura

Neste apêndice apresenta-se o código desenvolvido em linguagem R de computação estatística (R Core Team, 2017).

```
f.cura<-function(beta0,p.mod0,X0,tempo,censura,dist){  
  require(flexsurv)  
  require(reliaR)  
  require(extraDistr)  
  require(stats)  
  
  #Distribuicoes  
  if(dist=="gama.gen"){  
    acumulada<-pgengamma.orig  
    densidade<-dgengamma.orig  
  }  
  
  if(dist=="exp"){
```

```

    acumulada<-pexp
    densidade<-dexp
}

if ( dist=="weibull" ){
    acumulada<-pweibull
    densidade<-dweibull
}

if ( dist=="lognormal" ){
    acumulada<-plnorm
    densidade<-dlnorm
}

if ( dist=="llogis" ){
    acumulada<-pllogis
    densidade<-dllogis
}

if ( dist=="gama" ){
    acumulada<-pgamma
    densidade<-dgamma
}

if ( dist=="weibull.exp" ){
    acumulada<-function ( t , alfa , sigma , theta ){
        (1-exp(-(t/sigma)^ alfa )) ^ theta
    }
    densidade<-function ( t , alfa , sigma , theta ){

```

```

      ( alfa*theta/sigma)*
      ((1-exp(-(t/sigma)^ alfa ))^( theta -1))*
      ( exp(-(t/sigma)^ alfa ))*( t/sigma)^( alfa -1)
    }
  }

```

```

if ( dist=="weibull.modif"){
  acumulada<-function ( t ,a ,b ,lambda){
    1-exp(-a*(t^b)*exp(lambda*t))
  }

  densidade<-function ( t ,a ,b ,lambda){
    a*(b+lambda*t)*( t^(b-1))*exp(lambda*t)*
    exp(-a*(t^b)*exp(lambda*t))
  }
}

```

```

if ( dist=="weibull.estend"){
  acumulada<-function ( t ,lambda , alfa , beta){
    1-exp(lambda*alfa*(1-exp(( t/ alfa )^ beta )))
  }

  densidade<-function ( t ,lambda , alfa , beta){
    lambda*beta*(( t/ alfa )^( beta -1))*
    exp((( t/ alfa )^ beta)+lambda*alfa*
    (1-exp(( t/ alfa )^ beta )))
  }
}

```

```

if ( dist=="weibull.bn"){
  acumulada<-function(y,eta , teta , alfa , lambda){
    1-(((1+eta*
      teta*(1-exp(-(y^alfa)*exp(lambda))))^( -1/eta))-
      ((1+eta*teta)^(-1/eta)))/
      (1-((1+eta*teta)^(-1/eta)))
  }

  densidade<-function(y,eta , teta , alfa , lambda){
    (( teta*alfa*(y^(alfa -1))*
      exp(lambda-(y^alfa)*exp(lambda)))*
      ((1+eta*teta*
        (1-exp(-(y^alfa)*exp(lambda))))^( -(1/eta)+1)))/
      (1-((1+eta*teta)^(-1/eta)))
    }
  }

#####
#####
####
E  <- 1e-100  #Erro
Av <- E + 1  # Av (diferenca das estimativas entre iteracoes)
k  <- 1

beta  <-beta0  #vetor das estimativas dos parametros

```

```

p.mod <-p.mod0 # parametros da distribuicao
X <- X0        # variaveis

while(Av > E){ #condicao de parada
  pi<-(exp(t(beta)%*%t(X)))/(1+exp(t(beta)%*%t(X)))

##### distribuicao com 1 parametro #####
  if (length(p.mod)==1){
    # PASSO E
    w1<-pi/(pi+(1-pi)*(1- acumulada(tempo,p.mod))) #S(t)=1-F(t)
    w2<-(1-w1)

    # PASSO M
    # estimacao dos parametros do modelo
    g1<-function(beta,tempo,censura,X,w1,w2){
      pi<-(exp(t(beta)%*%t(X)))/(1+exp(t(beta)%*%t(X)))

      -1*(sum(censura*log(1-pi))+sum((1-censura)*w1*log(pi))+
        sum((1-censura)*w2*log(1-pi)))
    }

    # estimacao dos parametros da distribuicao
    g2<-function(parametro,tempo,censura,X,w1,w2){
      if(parametro>0)
        return(-1*(sum(censura*
          log(densidade(tempo,parametro[1]))+
          sum((1-censura)*w2*log(1-acumulada(tempo,
            parametro[1])))))
      else return(-Inf)
    }
  }
}

```

```

    }
}

```

```

##### distribuicao com 2 parametros #####
if (length(p.mod)==2){
  # PASSO E
  w1<-pi/(pi+(1-pi)*(1-
                                     acumulada(tempo,p.mod[1],p.mod[2])))

  w2<-(1-w1)

  # PASSO M
  # estimacao dos parametros do modelo
  g1<-function(beta,tempo,censura,X,w1,w2){
    pi<-(exp(t(beta)%*%t(X)))/(1+exp(t(beta)%*%t(X)))

    -1*(sum(censura*log(1-pi))+sum((1-censura)*w1*log(pi))+
          sum((1-censura)*w2*log(1-pi)))
  }

  # estimacao dos parametros da distribuicao
  if(dist!="lognormal"){
    g2<-function(parametro,tempo,censura,X,w1,w2){
      if(parametro[1]>0 | parametro[2]>0)
        return(-1*(sum(censura*log(densidade(tempo,
                                                  parametro[1],
                                                  parametro[2]))) +
                sum((1-censura)*w2*log(1-acumulada(tempo,parametro[1],

```

```

                                parametro[2]))))
    else return(-Inf)
  }
} else {
  g2<-function(parametro,tempo,censura,X,w1,w2){
    if(sum(parametro>0)==2)
      return(-1*(sum(censura*log(densidade(tempo,
                                parametro[1],
                                parametro[2])))+
      sum((1-censura)*w2*log(1-acumulada(tempo,parametro[1],
                                parametro[2])))))
    else return(-Inf)
  }
}

}

##### distribuicao com 3 parametros #####
if (length(p.mod)==3){
  # PASSO E
  w1<-pi/(pi+(1-pi)*(1-acumulada(tempo,p.mod[1],
                                p.mod[2],
                                p.mod[3]))) #S(t)=1-F(t)

  w2<-(1-w1)

  # PASSO M
  # estimacao dos parametros do modelo
  g1<-function(beta,tempo,censura,X,w1,w2){

```

```

pi<-(exp(t(beta)%*%t(X)))/(1+exp(t(beta)%*%t(X)))

-1*(sum(censura*log(1-pi))+sum((1-censura)*w1*log(pi))+
      sum((1-censura)*w2*log(1-pi)))
}

# estimacao dos parametros da distribuicao
g2<-function(parametro,tempo,censura,X,w1,w2){
  if(sum(parametro>0)==3)
    return(-1*(sum(censura*log(densidade(tempo,
                                             parametro[1],
                                             parametro[2],
                                             parametro[3])))+
            sum((1-censura)*w2*log(1-acumulada(tempo,
                                                  parametro[1],
                                                  parametro[2],
                                                  parametro[3]))
            else return(-Inf)
  }
}

##### distribuicao com 4 parametros #####
if (length(p.mod)==4){
  # PASSO E
  w1<-pi/(pi+(1-pi)*(1-acumulada(tempo,p.mod[1],p.mod[2],
                                     p.mod[3],p.mod[4]))) #S(t)=1-F(t)
  w2<-(1-w1)

  # PASSO M

```

```

# estimacao dos parametros do modelo
g1<-function(beta,tempo,censura,X,w1,w2){
  pi<-(exp(t(beta)%*%t(X)))/(1+exp(t(beta)%*%t(X)))

  -1*(sum(censura*log(1-pi))+sum((1-censura)*w1*log(pi))+
      sum((1-censura)*w2*log(1-pi)))
}

# estimacao dos parametros da distribuicao
g2<-function(parametro,tempo,censura,X,w1,w2){
  if(sum(parametro>0)==3 | sum(parametro>0)==4)
    return(-1*
      (sum(censura*
        log(densidade(tempo,
          parametro[1],
          parametro[2],
          parametro[3],
          parametro[4])))+
        sum((1-censura)*w2*
          log(1-acumulada(tempo,
            parametro[1],
            parametro[2],
            parametro[3],
            parametro[4])))))
  else return(-Inf)
}
}

```

```

#maximizacao da g1 e g2
max.g1<-optim(beta,g1,tempo=tempo,censura=censura,X=X,
              w1=w1,w2=w2,hessian = TRUE)
max.g2<-optim(p.mod,g2,tempo=tempo,censura=censura,X=X,
              w1=w1,w2=w2)

# Atualizando os parametros
mg1<-c(max.g1$par[1:length(max.g1$par)])
mg2<-c(max.g2$par[1:length(max.g2$par)])
Av<-max(max(abs(mg1-beta)),max(abs(mg2-p.mod)))
beta<-mg1
p.mod<-mg2
k<-k+1
}

# matriz de variancias (diag principal) e covariancias
sd<-ginv(max.g1$hessian)
erro.padrao<-c() #erro padrao
z<-c() # estatistica do teste
p.valor<-c() #pvalor
Li<-c() #limite inferior
Ls<-c() #limite superior

for (i in 1:length(beta)){
  erro.padrao[i]<-sqrt(sd[i,i])
  z[i]<-beta[i]/erro.padrao[i]
  p.valor[i]<-1-pnorm(abs(z[i]))
  Li[i]<-beta[i]-1.96*erro.padrao[i]

```

```

      Ls[i]<-beta[i]+1.96*erro.padrao[i]
    }

    OR<-c() #OR
    OR[1]<-"-"
    for (i in 2:length(beta)){
      OR[i]<-round(exp(beta[i]),4)
    }

    return(list(estima=data.frame(estimativa=beta,
                                   erro.padrao=round(erro.padrao,4),
                                   Li=round(Li,4),
                                   Ls=round(Ls,4),
                                   z=round(z,4),
                                   valor.p=round(p.valor,4),
                                   RC=OR),
              e.dist=p.mod,
              lverossimilanca=(max.g1$value+max.g2$value)))
  }

```

Apêndice B

Exemplos - Modelos de Fração de Cura Paramétricos

Utilizando os modelos apresentados na Seção 2.4 e as Equações 5, 12 e 7, pode-se encontrar as funções de densidade, sobrevivência e taxa de falha com fração de cura ou de longa duração para cada distribuição exemplificada

- Exponencial com fração de cura

$$f_{EFC}(t; \phi|\mathbf{x}) = \frac{1}{1 + \exp\{\boldsymbol{\beta}'\mathbf{x}\}} \left[\frac{1}{\alpha} \exp\left\{-\frac{t}{\alpha}\right\} \right],$$

$$S_{EFC}(t; \phi|\mathbf{x}) = \frac{\exp\{\boldsymbol{\beta}'\mathbf{x}\} + \exp\left\{-\left(\frac{t}{\alpha}\right)\right\}}{1 + \exp\{\boldsymbol{\beta}'\mathbf{x}\}},$$

e

$$h_{EFC}(t; \phi|\mathbf{x}) = \frac{\frac{1}{\alpha} \exp\left\{-\left(\frac{t}{\alpha}\right)\right\}}{\exp\{\boldsymbol{\beta}'\mathbf{x}\} + \exp\left\{-\left(\frac{t}{\alpha}\right)\right\}},$$

em que $\alpha > 0$ e $t \geq 0$.

- Weibull com fração de cura

$$f_{WFC}(t; \phi|\mathbf{x}) = \frac{1}{1 + \exp\{\boldsymbol{\beta}'\mathbf{x}\}} \left[\frac{\gamma}{\alpha^\gamma} t^{\gamma-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\gamma\right\} \right],$$

$$S_{WFC}(t; \phi|\mathbf{x}) = \frac{\exp\{\boldsymbol{\beta}'\mathbf{x}\} + \exp\left\{-\left(\frac{t}{\alpha}\right)^\gamma\right\}}{1 + \exp\{\boldsymbol{\beta}'\mathbf{x}\}},$$

e

$$h_{WFC}(t; \phi|\mathbf{x}) = \frac{\frac{\gamma}{\alpha^\gamma} t^{\gamma-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\gamma\right\}}{\exp\{\boldsymbol{\beta}'\mathbf{x}\} + \exp\left\{-\left(\frac{t}{\alpha}\right)^\gamma\right\}}$$

em que, $\alpha, \gamma > 0$ e $t \geq 0$.

- Lognormal com fração de cura

$$f_{LNFC}(t; \boldsymbol{\phi}|\mathbf{x}) = \frac{1}{1 + \exp\{\boldsymbol{\beta}'\mathbf{x}\}} \left[\frac{1}{t\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\frac{\log(t) - \mu}{\sigma} \right)^2 \right\} \right],$$

$$S_{LNFC}(t; \boldsymbol{\phi}|\mathbf{x}) = \frac{\exp\{\boldsymbol{\beta}'\mathbf{x}\} + \Phi \left(\frac{-\log(t) + \mu}{\sigma} \right)}{1 + \exp\{\boldsymbol{\beta}'\mathbf{x}\}}$$

e

$$h_{LNFC}(t; \boldsymbol{\phi}|\mathbf{x}) = \frac{\frac{1}{t\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\frac{\log(t) - \mu}{\sigma} \right)^2 \right\}}{\exp\{\boldsymbol{\beta}'\mathbf{x}\} + \Phi \left(\frac{-\log(t) + \mu}{\sigma} \right)},$$

em que, $\sigma > 0$, $\mu \in \Re$ e $\Phi(\cdot)$ é a função distribuição acumulada de uma normal padrão.

- Loglogística com fração de cura

$$f_{LLFC}(t; \boldsymbol{\phi}|\mathbf{x}) = \frac{1}{1 + \exp\{\boldsymbol{\beta}'\mathbf{x}\}} \left[\frac{\gamma}{\alpha^\gamma} t^{\gamma-1} \left(1 + \left(\frac{t}{\alpha} \right)^\gamma \right)^{-2} \right],$$

$$S_{LLFC}(t; \boldsymbol{\phi}|\mathbf{x}) = \frac{\exp\{\boldsymbol{\beta}'\mathbf{x}\} + \frac{1}{1 + \left(\frac{t}{\alpha} \right)^\gamma}}{1 + \exp\{\boldsymbol{\beta}'\mathbf{x}\}}$$

e

$$h_{LLFC}(t; \boldsymbol{\phi}|\mathbf{x}) = \frac{\frac{\gamma}{\alpha^\gamma} t^{\gamma-1} \left(1 + \left(\frac{t}{\alpha} \right)^\gamma \right)^{-2}}{\exp\{\boldsymbol{\beta}'\mathbf{x}\} + \frac{1}{1 + \left(\frac{t}{\alpha} \right)^\gamma}},$$

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

- Gama com fração de cura

$$f_{GFC}(t; \boldsymbol{\phi}|\mathbf{x}) = \frac{1}{1 + \exp\{\boldsymbol{\beta}'\mathbf{x}\}} \left[\frac{1}{\Gamma(k)\alpha^k} t^{k-1} \exp \left\{ -\left(\frac{t}{\alpha} \right) \right\} \right],$$

$$S_{GFC}(t; \boldsymbol{\phi} | \boldsymbol{x}) = \frac{\exp\{\boldsymbol{\beta}' \boldsymbol{x}\} + \int_t^\infty \frac{1}{\Gamma(k)\alpha^k} u^{k-1} \exp\left\{-\left(\frac{u}{\alpha}\right)\right\} du}{1 + \exp\{\boldsymbol{\beta}' \boldsymbol{x}\}}$$

e

$$h_{GFC}(t; \boldsymbol{\phi} | \boldsymbol{x}) = \frac{\frac{1}{\Gamma(k)\alpha^k} t^{k-1} \exp\left\{-\left(\frac{t}{\alpha}\right)\right\}}{\exp\{\boldsymbol{\beta}' \boldsymbol{x}\} + \int_t^\infty \frac{1}{\Gamma(k)\alpha^k} u^{k-1} \exp\left\{-\left(\frac{u}{\alpha}\right)\right\} du},$$

em que $k, \alpha > 0$ e $t \geq 0$.