



**ROSSICLEY RANGEL DE PAIVA**

**HEPATITE A, B e C EM LAVRAS-MG: ESTUDO DE  
CASO VIA ANÁLISE DE SOBREVIVÊNCIA  
UTILIZANDO MODELOS ENCAIXADOS**

**LAVRAS-MG**

**2013**

**ROSSICLEY RANGEL DE PAIVA**

**HEPATITE A, B e C EM LAVRAS-MG: ESTUDO DE CASO VIA ANÁLISE DE  
SOBREVIVÊNCIA UTILIZANDO MODELOS ENCAIXADOS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para obtenção do título de Mestre.

Orientador

Dr. Mário Javier Ferrua Vivanco

**LAVRAS-MG**

**2013**

**Ficha Catalográfica Elaborada pela Coordenadoria de Produtos e  
Serviços da Biblioteca Universitária da UFLA**

Paiva, Rossicley Rangel de.

Hepatite A, B e C em Lavras-MG : estudo de caso via análise de sobrevivência utilizando modelos encaixados / Rossicley Rangel de Paiva. – Lavras : UFLA, 2013.

87 p. : il.

Dissertação (mestrado) – Universidade Federal de Lavras, 2013.

Orientador: Mário Javier Ferrua Vivanco.

Bibliografia.

1. Modelos de regressão. 2. Distribuições paramétricas. 3. Seleção de covariáveis. 4. Tipos de hepatite - Análise - Modelos estatísticos. 5. Teste da razão de verossimilhança. I. Universidade Federal de Lavras. II. Título.

CDD – 519.536

**ROSSICLEY RANGEL DE PAIVA**

**HEPATITE A, B e C EM LAVRAS-MG: ESTUDO DE CASO VIA ANÁLISE DE  
SOBREVIVÊNCIA UTILIZANDO MODELOS ENCAIXADOS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para obtenção do título de Mestre.

APROVADA em 30 de agosto de 2013.

Dr. Marcelo Ângelo Cirillo

UFLA

Dr. Clailton Ataídes de Freitas

UFMS

Orientador

Dr. Mário Javier Ferrua Vivanco

**LAVRAS-MG**

**2013**

A DEUS.

A minha esposa Noêmia e filhas Thaís, Sofia e Raissa.

A minha mãe Glória.

**DEDICO**

## AGRADECIMENTOS

A DEUS, todo-poderoso que me abriu esta porta: o ingresso à Universidade Federal de Lavras para cursar um mestrado. Neste período, foi quem realmente esteve ao meu lado em todos os instantes, nas madrugadas, nos momentos de cansaço, nos muitos momentos de dificuldades de compreensão. Meu DEUS agradeço-Lhe por ter me confortado e fortalecido nos momentos de angústia.

A minha esposa Noêmia, por entender a minha ausência em alguns momentos e supri-la junto às minhas filhas, por sempre me incentivar e por acreditar que seria possível.

As minhas filhas, Thaís, Sofia e Raissa, por entenderem a minha ausência e acreditarem em mim.

A minha mãe, as palavras de conforto e fé em DEUS, que em alguns momentos de angústia me foram ditas.

Aos meus irmãos, Risia, Rondinely e Rayanne, além de todos os meus familiares e amigos que torceram por mim.

Ao Professor Dr. Mário Javier Ferrua Vivanco, pela orientação e palavras de incentivo, também pela paciência com as minhas limitações.

Aos amigos, Gilberto Rodrigues Liska, Guido Gustavo Humada Gonzalez, Juliano Bortolini e todos os colegas de curso que, de alguma forma, em algum momento, contribuíram com essa conquista.

A senhora Myrienne Antunes Camargo Amorim, enfermeira e responsável pelo Serviço de Epidemiologia da Cidade de Lavras-MG, pela grande contribuição na captação dos dados deste trabalho.

Ao Dr. Silvio Augusto Corsini Menicucci, médico epidemiologista da Santa Casa de Lavras, que muito contribuiu com este trabalho, por meio de seus conhecimentos e experiências.

Aos professores que participaram da banca de defesa, Professor Dr. Marcelo Ângelo Cirillo, Professor Dr. Marcelo Silva de Oliveira, Professor Dr. Clailton Ataídes de Freitas e Professor Dr. Fabyano Fonseca e Silva, pelo direcionamento, pelas palavras de incentivo, pelas muitas críticas que certamente contribuíram com a melhora deste trabalho.

Ao professor Renato Ribeiro de Lima, pelos ensinamentos e pela a oportunidade de aprender um pouco como interagir com os docentes, ainda pelos momentos de ajuda como coordenador do curso.

À secretária do curso Josiane Oliveira, pela sua disposição, dedicação e eficiência em tratar dos assuntos pertinentes ao curso.

À Universidade Federal de Lavras (UFLA), pela oportunidade conferida a mim, às instituições CAPES e CNPq.

Finalmente, meus sinceros agradecimentos a todos que não mencionei, mas me incentivaram e torceram por mim.

“Não ameis o mundo nem as coisas que há no mundo. Se alguém amar o mundo, o amor do Pai não está nele; porque tudo que há no mundo, a concupiscência da carne, a concupiscência dos olhos e a soberba da vida, não procede do Pai, mas procede do mundo.

Ora, o mundo passa, bem como a sua concupiscência; aquele, porém, que faz a vontade de Deus permanece eternamente.” I João 1:15-17.



## RESUMO

A partir do momento da descoberta do vírus da hepatite, o indivíduo busca a cura da doença. Quando o tratamento tem um período curto de tempo, esse indivíduo consegue afastar-se por completo ou parcialmente de seus maus hábitos cotidianos, contribuindo para a cura. O problema mais grave encontra-se com o paciente que necessita de um período maior de tratamento. A necessidade de estimar um menor tempo possível até a cura e a relação de tempo até a cura entre as hepatites (A, B e C), as fases (aguda e crônica), o sexo (masculino e feminino) e a idade dos 96 indivíduos foi o objetivo deste trabalho. O presente trabalho propôs inicialmente dois modelos: um modelo de regressão paramétrico e um modelo de regressão semi paramétrico de Cox, para que ambos pudessem retratar essas relações de tempos até a cura das hepatites. Contudo, ao utilizar a distribuição Gama-Generalizada com o propósito de obter uma distribuição paramétrica que fosse um caso particular da GG, as distribuições candidatas (Exponencial, Weibull e Lognormal) não se ajustaram aos dados segundo o teste da razão de verossimilhança. Portanto, o modelo de regressão de Cox foi ajustado, considerando as covariáveis significativas selecionadas pelo método de Collett (tipos de hepatite, fases da hepatite e sexo do indivíduo). Foram todas as covariáveis avaliadas segundo a suposição de riscos proporcionais pelo teste de Schoenfeld e todas foram favoráveis a esta condição, os resíduos de Cox-Snell confirmaram a adequação do modelo e então as estimativas dos parâmetros desse modelo indicaram, entre outras, que indivíduos do sexo masculino têm risco de cura das hepatites maior que indivíduos do sexo feminino.

Palavras-chave: Tipos de Hepatite. Distribuições Paramétricas. Teste da Razão de Verossimilhança. Seleção de Covariáveis. Modelos de Regressão.

## ABSTRACT

From the moment of discovery of hepatitis virus, the individual seeks a cure. When treatment has a short period of time, this individual can move away completely or partially from their bad habits everyday, contributing to healing in virtually all cases. The most serious problem is that the patient requires a longer period of treatment. The need to estimate a shortest possible time to cure hepatitis and the relationships between these cure times considering the types of hepatitis (A, B and C), phases (acute and chronic), sex (male and female) , and the age of the 96 individuals, was the objective of this work. This work originally proposed two models, a regression model parametric and semi-parametric regression model of Cox, so they could portray these relationships from time to cure hepatitis, however, by using the Generalized Gamma distribution for the purpose obtain a parametric distribution that is a particular case of GG, the candidate distributions (Exponential, Weibull and Lognormal) did not fit the data according to the likelihood ratio test. Therefore, the Cox regression model was adjusted, considering the significant covariates selected by the method of Collett (types of hepatitis, hepatitis stages and sex of the individual). All covariates were evaluated according to the proportional hazards assumption by Schoenfeld test and all were favorable to this condition, waste Cox-Snell confirmed the adequacy of the model and then estimates the parameters of this model indicated that males get less time to cure hepatitis.

Keywords: Types of Hepatitis, Parametric Distributions, Likelihood Ratio Test, Selection of covariates, Regression Models.

## LISTA DE FIGURAS

Figura 1	Censura à direita.....	35
Figura 2	Censura à direita do tipo aleatória.....	35
Figura 3	Curvas de sobrevivência obtidas por método não paramétrico.....	38
Figura 4	Gráfico dos resíduos de Schoenfeld.....	73
Figura 5	Gráfico dos resíduos de Cox-Snell.....	73

## LISTA DE TABELAS

Tabela 1	Estimativas através da função de sobrevivência pelo método de Kaplan-Meier.....	43
Tabela 2	Variáveis consideradas preliminarmente.....	67
Tabela 3	Fases da hepatite.....	68
Tabela 4	Sexo dos pacientes com hepatite.....	69
Tabela 5	Tipos de hepatite.....	69
Tabela 6	Censura.....	69
Tabela 7	Escolha de uma distribuição de probabilidade através da distribuição gama generalizada.....	70
Tabela 8	Seleção de covariáveis usando o modelo de regressão de Cox.....	71
Tabela 9	Teste de proporcionalidade dos riscos.....	72
Tabela 10	Estimativas dos parâmetros referentes ao modelo de Cox ajustado aos dados de hepatites.....	74
Tabela 11	Estimativas dos parâmetros referentes ao modelo de Cox ajustado aos dados de hepatites considerando a covariável idade.....	75

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	14
<b>2 REFERENCIAL TEÓRICO</b> .....	17
2.1 O fígado e suas funções.....	17
2.2 Hepatite.....	18
2.3 História das hepatites .....	19
2.4 Sintomas das hepatites .....	20
2.5 Transmissão das hepatites.....	21
2.6 Hepatite A.....	21
2.6.1 Sintomas da hepatite A.....	22
2.6.2 Transmissão da hepatite A .....	22
2.6.3 Prevenção da hepatite A .....	23
2.7 Hepatite B.....	23
2.7.1 História da hepatite B.....	24
2.7.2 Sintomas da hepatite B .....	26
2.7.3 Transmissão da hepatite B .....	26
2.7.4 Grupos de risco da hepatite B .....	27
2.8 Hepatite C.....	28
2.8.1 História da hepatite C.....	28
2.8.2 Sintomas da hepatite C.....	29
2.8.3 Transmissão da hepatite C .....	30
2.8.4 Grupos de risco da hepatite C .....	31
2.9 Análise de sobrevivência .....	32
2.9.1 Censura.....	33
2.9.2 Função de sobrevivência.....	36
2.9.3 Função taxa de risco .....	38
2.9.4 Estimação da função de sobrevivência.....	39
2.9.5 Estimador de Kaplan-Meier.....	40
2.9.6 Casos de aplicação de análise de sobrevivência.....	41
2.10 Modelos Probabilísticos.....	44
2.10.1 Modelo exponencial .....	44
2.10.2 Modelo Weibull.....	45
2.10.3 Modelo Log-normal .....	46
2.10.4 Modelo Gama-Generalizada .....	46

2.11. Métodos de estimação paramétrico.....	48
2.12 Modelo de regressão paramétrico .....	50
2.12.1 Modelo de regressão exponencial.....	51
2.12.2 Modelo de regressão Weibull .....	53
2.13 Adequação do modelo paramétrico .....	54
2.14 Modelo de regressão de Cox.....	55
2.15 Verossimilhança parcial .....	56
2.16 Teste da Razão de Verossimilhança.....	59
2.17 Razão de Risco.....	60
2.18 Adequação do modelo de Cox .....	60
2.19 Modelo de Cox Estratificado.....	61
2.20 Teste de Schoenfeld para Riscos proporcionais.....	62
2.21 Seleção de covariáveis.....	62
3 METODOLOGIA E MATERIAL.....	64
3.1 Metodologia .....	64
3.2 Material .....	66
4 RESULTADOS E DISCUSSÃO .....	68
4.1 Resultados descritivos.....	68
4.2 Ajuste do modelo paramétrico.....	70
4.3 Ajuste do modelo semi paramétrico.....	70
4.4 Discussão.....	76
5 CONCLUSÃO .....	79
REFERÊNCIAS.....	81

## 1 INTRODUÇÃO

A hepatite é uma doença viral que afeta o fígado, causa fibrose, cirrose, cancro hepático e câncer hepático e que, em muitos casos, acarreta morte dos infectados. As hepatites são doenças provocadas por diferentes agentes etiológicos e apresentam características epidemiológicas, clínicas e laboratoriais semelhantes, porém com algumas particularidades. Algumas rotinas clínicas indicam diferenças entre as hepatites virais, essas diferenças geralmente não são confirmadas se não por métodos mais eficazes. A disseminação das hepatites é um fator mundial e a infecção pelo vírus já é a maior responsável por cirrose e transplante hepático no mundo. A incidência de cada tipo varia para cada região de acordo com os vetores já conhecidos como transmissão vertical, transmissão parenteral, transmissão sexual e as condições de saneamento básico e higiene estabelecidas e desenvolvidas. Em geral, qualquer pessoa está sujeita a contrair hepatite, no entanto, alguns grupos de pessoas, devido a questões profissionais ou mesmo hábitos culturais, estão mais susceptíveis a contrair hepatite. Em termos de Brasil, o Governo Federal, através do Ministério da Saúde em parcerias com Universidades Federais e Estaduais, Secretarias Estaduais e Municipais de Saúde, faz um trabalho de captação da real situação em que se encontra o país no contexto macrorregional de incidência de casos de hepatite.

A maneira de relacionar-se sexualmente, a questão de higiene praticada, a disposição para o uso de drogas, as transfusões de sangue e o saneamento básico precário existente são alguns dos meios de contrair o vírus da hepatite. Estudos indicam que grande parte dos indivíduos que possuem o vírus da hepatite desconhece tal fato e somente terão acesso a essa informação em um primeiro estágio da doença quando através de uma consulta médica o clínico identifica algum sintoma, ou ainda através de um exame laboratorial onde,

geralmente, o interesse inicial seria observar outra enfermidade. Em um segundo estágio, o paciente terá acesso à informação de que possui o vírus, quando a doença já se encontra instalada há algum tempo, o que vai gerar um desgaste emocional intenso, devido à dificuldade acometida pelo tratamento no que se refere ao tempo necessário dispendido. Nesse estágio, os efeitos colaterais causados pela medicação e o tempo necessário para se obter a cura sobrepõem ao desgaste psicológico causado, entre outras, pela necessidade de romper com comportamentos rotineiros dos pacientes.

Os órgãos públicos competentes precisam organizar-se por meio de campanhas publicitárias eficientes e difundir informações que possam conscientizar a população do risco permanente e silencioso que são, respectivamente, os meios de contrair o vírus da hepatite e a possibilidade de tê-lo e desconhecer tal fato.

A estatística tem um papel importante nesse processo, ao modelar o tempo até a cura de pacientes que contraíram o vírus de hepatite A, B ou C, em função de covariáveis como idade, sexo, tipo de vírus de hepatite, fases da doença e outras covariáveis, a serem recomendadas para estudo por médicos especialistas nesse tipo de doença.

Deve-se ressaltar que especialistas precisam conhecer as covariáveis que são influentes no processo de tratamento de pacientes com hepatite, a fim de que se possa tratar cada paciente em um menor intervalo de tempo, com ganho de qualidade de vida, como também diminuir na evolução de câncer hepático e cirrose. Nisso, a estatística, pela análise de sobrevivência, pode agregar resultados para a tomada de decisões relativa ao estudo do tempo de diagnóstico até a cura de um tratamento dado a pacientes com hepatites do tipo A, B ou C.

Neste trabalho, será feita uma análise de regressão paramétrica e/ou semiparamétrica a fim de estabelecer estatisticamente quais são as covariáveis que mais influenciam no tratamento de pacientes com hepatites A, B ou C. Para



isso, serão aplicados métodos de análise de sobrevivência de uso comum ou não, como, por exemplo, regressão paramétrica e regressão de Cox. Como é de praxe em qualquer modelo de regressão aplicado, estudar-se-á a seleção de covariáveis, teste de proporcionalidade dos riscos, qualidade do ajuste do modelo e os pontos influentes.

Este trabalho tem como objetivo propor um modelo paramétrico e semiparamétrico que relacione o tempo até a cura de portadores do vírus de hepatite (A, B e C) da cidade de Lavras-MG, bem como verificar o grau de influência sobre a cura das hepatites ao considerar os níveis de cada covariável. Para isso, foram observadas as seguintes covariáveis: Tipo de vírus, sexo dos pacientes, idade dos pacientes e fases da doença.

O modelo proposto permitirá saber quais são as covariáveis e seus respectivos níveis que mais influenciam a cura das hepatites, informação valiosa para a tomada de decisões de médicos especialistas.

## **2 REFERENCIAL TEÓRICO**

As hepatites são um grave problema de saúde pública no Brasil e no mundo. A hepatite é a inflamação do fígado. Pode ser causada por vírus, uso de alguns remédios, álcool e outras drogas, além de doenças autoimunes e genéticas. São doenças silenciosas que nem sempre apresentam sintomas, mas quando aparecem podem ser cansaço, febre, mal-estar, tontura, enjojo, vômitos, dor abdominal, pele e olhos amarelados, urina escura e fezes claras. No Brasil, as hepatites virais mais comuns são as causadas pelos vírus A, B e C (BRASIL, s.n.a).

### **2.1 O fígado e suas funções**

O fígado é o maior órgão interno do corpo. Localiza-se na parte superior do abdômen, à direita. Ter um fígado saudável é importante para todos nós, mas para as pessoas que vivem com o HIV (vírus da imunodeficiência humana) é especialmente importante. O fígado tem um papel vital no processamento da medicação utilizada para o tratamento do HIV e outras doenças. As infecções virais como, a hepatite A, a hepatite B e a hepatite C, atacam o fígado de tal forma, tornando-o comprometido, entre outras funções, no seu desempenho de processar as medicações para cura de doenças. O fígado tem quatro funções principais:

1. O armazenamento e desintoxicação do sangue, filtrando substâncias indesejadas.

2. A produção de uma substância chamada bílis, que se liberta no intestino e ajuda a digerir a gordura.
3. O processamento dos nutrientes dos alimentos, aportando energia à circulação sanguínea e armazenando vitaminas e minerais.
4. A produção de proteínas e de certas vitaminas (SASSO et al.,2005).

Segundo Carter (2010), o consumo excessivo de bebidas alcoólicas, o uso de drogas e ter hepatite B ou hepatite C podem provocar lesões no tecido do fígado. Utilizam-se dois termos para descrever essas lesões: fibrose e cirrose. Fibrose é um endurecimento do fígado que provoca cicatrizes, quando tratada inicialmente a fibrose pode ser reversível. A cirrose é um estado de cicatrização grave do fígado, comprometendo-o de forma a não retornar às suas funções, em alguns casos. A cirrose resultante duma infecção prolongada por hepatite B e hepatite C aumenta significativamente as hipóteses de desenvolver um cancro hepático. O cancro hepático é difícil de tratar e a operação, que exige a extração de parte do fígado ou um transplante, é frequentemente a única opção. Os tumores de pequena dimensão podem ser extraídos, mas existe um elevado risco de desenvolvimento de um novo tumor nos cinco anos a seguir. A quimioterapia não tem mostrado benefícios na sobrevivência, mas pode aliviar os sintomas.

## **2.2 Hepatite**

A hepatite é uma doença do fígado altamente contagiosa e às vezes fatal, para Fuks (2012), ela pode ser decorrente de causas diversas e sua prevalência é bastante elevada, atingindo milhões de pessoas no mundo inteiro.

A hepatite alcoólica é uma lesão grave, conseqüente ao consumo excessivo de etanol. Na forma intensa da hepatite alcoólica os índices de mortalidade são muito altos (MINCIS; MINCIS, 2010).

A esteatose hepática caracteriza-se por acúmulo de gordura no fígado. Chamado também de “fígado gorduroso”, as causas da esteatose são o consumo de bebidas alcoólicas e a obesidade. Outros fatores de risco para o desenvolvimento de gordura no fígado são a diabetes tipo dois, a resistência à insulina e o colesterol elevado. A hepatite medicamentosa é causada pelo uso de alguns remédios de uso clínico em indivíduos suscetíveis, onde não se pode prever qual indivíduo terá hepatite por utilizar determinada droga. Algumas doenças fazem com que as substâncias de defesa do próprio indivíduo causem inflamação e dano ao fígado, chamadas hepatite autoimune (ABPH, 2013).

### **2.3 História das hepatites**

Em todo o mundo, as hepatites A e B continuam a ser um grande problema de saúde para as sociedades. O Brasil tem diversas características de países em rápido desenvolvimento, como a migração de comunidades rurais para as cidades, o que tem levado à formação de áreas urbanas de baixa renda e socialmente carentes (CLEMENS et al., 2000).

A hepatite viral aguda é uma doença antiga, no entanto acredita-se que o primeiro surto de hepatite B tenha ocorrido há cerca de 100 anos, na Alemanha. A hepatite é definida como uma infecção que acomete primordialmente o fígado, sendo que atualmente aceita-se como agentes etiológicos seis tipos de vírus, com características imunológicas e atributos epidemiológicos específicos. Os vírus da hepatite são: VHA (vírus da hepatite viral A), VHB (vírus da hepatite viral B),

VHC (vírus da hepatite viral C), VHD (vírus da hepatite viral D, ou Delta), VHE (vírus da hepatite viral E) e VHG (vírus da hepatite viral G). Os agentes virais possuem ação com lesão do hepatócito, seguida de reação inflamatória (CRUZ; BONETTO; CARNEIRO, 2000).

A hepatite crônica pode ter características clínicas muito semelhantes às da hepatite aguda viral, porém, do ponto de vista laboratorial, endoscópico e histológico, dados muito diferentes, com evolução na grande maioria dos casos para cirrose hepática. A hepatite crônica é uma doença inflamatória contínua do fígado, com início e duração variáveis, desencadeada por fatores desconhecidos. Constitui uma síndrome caracterizada pela presença de elementos clínicos, bioquímicos, sorológicos e histológicos que sugerem reação imunológica contra antígenos do hospedeiro, levando a danos celulares irreversíveis (PORTA, 2000).

Existem vacinas disponíveis para a hepatite A e para a hepatite B. Atualmente não existe vacina para a hepatite C. Para a hepatite A, a vacina está disponível em clínicas particulares e deve-se tomar duas doses. A vacina da hepatite B está disponível em clínicas particulares, para uma determinada faixa-etária na rede pública e são necessárias três doses (CARTER, 2010).

## **2.4 Sintomas das hepatites**

Segundo Fuks (2012), na hepatite infecciosa, os sintomas são semelhantes aos de uma gripe, com febre, dores articulares, náuseas, vômitos, falta de apetite e enfraquecimento. Pode também surgir uma cor amarelada da pele e das mucosas.

## **2.5 Transmissão das hepatites**

Embora as vias de transmissão do vírus das hepatites A (VHA) fecal-oral, B (VHB) contato com sangue, relação sexual e C (VHC) contato com sangue sejam diferentes, a disseminação de qualquer doença dentro da população é influenciada pelas condições ambientais prevalentes. Sem dúvida, melhorias adicionais na higiene, em particular, no suprimento de uma água segura seriam a longo prazo os meios de reduzir a soro prevalência de anticorpos anti-VHA no Brasil (CLEMENS et al., 2000).

## **2.6 Hepatite A**

A hepatite A é uma doença contagiosa, causada pelo vírus A, é também conhecida como “hepatite infecciosa”. Sua transmissão é fecal-oral, por contato entre indivíduos ou por meio de água ou alimentos contaminados pelo vírus. Pelo seu modo de transmissão, esse tipo de hepatite é típico de áreas menos desenvolvidas, com más condições de higiene e falta de saneamento básico. No Brasil, a hepatite A predomina em crianças de 2 a 6 anos, porém indivíduos que não tiveram a doença quando crianças podem adquiri-la em qualquer idade (BRASIL, s.d.a).

### **2.6.1 Sintomas da hepatite A**

Apresenta sinais prodrômicos frequentemente. A febre e a icterícia não são frequentes. Pode apresentar vômitos e raramente evolui para cronicidade, não deixando sequelas. A hepatite A agride com maior gravidade os adolescentes e adultos, possui mortalidade inferior a 1% e quase nunca evolui para a forma fulminante, conferindo imunidade homóloga (CRUZ; BONETTO; CARNEIRO, 2000).

Quando se apresenta clinicamente, os sintomas mais comuns são icterícia (pele e olhos amarelados), fadiga, falta de apetite, náuseas e dores articulares e musculares, ocasionalmente com febre baixa e dor no fígado (JORGE, 2005).

### **2.6.2 Transmissão da hepatite A**

A hepatite A é uma doença endêmica no Brasil e está diretamente relacionada às condições socioambientais, ou seja, sua prevalência está associada à falta de saneamento básico, uma vez que sua transmissão ocorre por via fecal-oral por meio de alimentos e água contaminados (VILLAR; DE PAULA; GASPAR, 2002).

Segundo Carter (2010), a hepatite A transmite-se através do contato com fezes humanas infectadas. A contaminação dos alimentos, da água para beber e das pedras de gelo é uma forma comum de transmissão, mas também pode ser transmitida durante o sexo.

O Brasil tem índices elevados de transmissão de hepatite A, em razão de condições deficientes ou inexistentes de saneamento básico (SANTOS et al., 2005).

### **2.6.3 Prevenção da hepatite A**

Sem dúvida, melhorias adicionais na higiene e, em particular, no suprimento de uma água segura seriam a longo prazo os meios de reduzir a soroprevalência de anticorpos anti-VHA no Brasil (CLEMENS et al., 2000).

Existem atualmente vacinas para prevenção das hepatites A e B. O Ministério da Saúde oferece vacina contra a hepatite A nos Centros de Referências Especiais (CRIEs) e vacina contra a hepatite B nos postos de saúde do Sistema Único de Saúde (SUS). Não existe vacina contra a hepatite C.

A vacina contra a hepatite A não faz parte do calendário nacional de vacinação. O encaminhamento, quando indicado, deverá ser feito pelo médico. No entanto, essa vacina está disponível no (CRIE) em algumas situações, como por exemplo, crianças menores de 13 anos com HIV e candidatos a transplante de órgão sólido cadastrados em programas de transplantes. (Ministério da Saúde).

## **2.7 Hepatite B**

Por ser uma doença sexualmente transmissível, a hepatite B desperta um pouco mais a atenção das pessoas, já que a forma de contrair o vírus pode estar



relacionada, entre outros casos, a uma prática sexual desprotegida, ou seja, sem o devido uso de preservativos, ou envolvimento com parceiros sexuais pertencentes a grupos de risco.

### **2.7.1 História da hepatite B**

As hepatites virais são uns dos maiores problemas de saúde em todo o mundo. A infecção pelo VHB atinge aproximadamente, 300 milhões de portadores crônicos, cerca de 5% da população mundial. No Brasil a endemicidade do VHB é heterogênea, sendo mais prevalente na região norte do país, incidindo nos estados do Acre, Amazonas, Rondônia e Roraima (BRASIL et al., 2003).

A infecção pelo VHB pode ocorrer em qualquer pessoa. No entanto, alguns grupos de indivíduos são particularmente expostos a esses vírus, em função de determinadas circunstâncias, pela adoção de certas atitudes comportamentais ou da atividade profissional que exercem (FERNANDES et al., 1999).

Em relação à distribuição segundo o sexo, pode-se observar que o sexo masculino foi mais predominante em todos os indivíduos infectados pelo VHB. Estas observações estão de acordo com o que é tradicionalmente descrito na literatura, em que na maioria da população a taxa de homens para mulheres portadoras é cerca de dois ou mais, e que são atribuídas a uma maior exposição do homem, aos riscos de infecção, seja pelo maior número de parceiros sexuais, homossexualismo, compartilhamento de lâmina de barbear, maiores riscos de acidente com necessidade de transfusão de sangue, hemofilia etc. (BRASIL et al., 2003).

No Brasil como um todo, avalia-se que de um a três por cento da população seja portadora crônica do VHB. Assim, a região Amazônica apresenta incidência da infecção, comparável às maiores do mundo, tendo-se demonstrado prevalência superior a 7% dos habitantes da região como portadores assintomáticos do antígeno de superfície do vírus da hepatite B (LONGHI; CORDEIRO; RAMALHAES, 2010).

Conforme Carter (2010), a maioria dos adultos com hepatite B não apresenta sintomas. Frequentemente a doença só é diagnosticada através de análises de sangue de rotina e da monitorização do funcionamento do fígado. Mesmo que não existam sintomas, pode transmitir-se o vírus a outras pessoas. No entanto, os sintomas podem surgir logo após a infecção.

Atualmente, a hepatite B é considerada uma das mais prevalentes infecções ocupacionais contraídas no ambiente hospitalar. O inter-relacionamento frequente entre profissionais de saúde e pacientes, e a manipulação de sangue e outros fluidos corporais contaminados com o vírus, representam fatores de risco de contágio. Foi demonstrado que nos indivíduos que trabalham em algumas áreas hospitalares a prevalência de infecção vigente ou pregressa pelo VHB chega a alcançar índices superiores a 30%, correspondendo a uma frequência cerca de 10 vezes maior que a encontrada na comunidade onde o hospital está localizado (FERNANDES et al., 1999).

A vacina contra a hepatite B faz parte do calendário de vacinação da criança e do adolescente e está disponível nas salas de vacinação do (SUS). É indicada para quem tem até 29 anos, 11 meses e 29 dias. Além disso, todo recém-nascido deve receber a primeira dose logo após o nascimento, preferencialmente nas primeiras 12 horas de vida. A oferta dessa vacina estende-se, também, a outros grupos em situações de maior vulnerabilidade, independentemente da faixa etária, como, por exemplo, trabalhadores da área da saúde e coletadores de lixo hospitalar e domiciliar (Ministério da Saúde).

### **2.7.2 Sintomas da hepatite B**

Na maioria dos casos, a hepatite B apresenta-se como assintomática ou sinais e sintomas inespecíficos, como anorexia, náuseas, vômitos e dores abdominais, dentre outros (ARRAES et al., 2003).

Os sintomas podem ser alguns dos seguintes: a pele e os olhos ficam amarelados (icterícia), perda de apetite, dor no estômago, náuseas e vômitos, febre alta, dores musculares nas articulações e sensação geral de mal-estar (CARTER, 2010).

### **2.7.3 Transmissão da hepatite B**

O período de incubação médio é de 120 dias a 160 dias e a transmissão se dá por sangue e secreções de pacientes infectados, principalmente através do contato sexual, acidentes com material contaminado e por meio da gestação (transmissão vertical) (CRUZ; BONETTO; CARNEIRO, 2000).

A transmissão do VHB se faz fundamentalmente através das vias parenteral e sexual. O rigoroso controle, hoje realizado, em grande parte dos bancos de sangue, praticamente eliminou de algumas áreas geográficas, a transmissão transfusional e na atualidade, a aquisição parenteral ocorre, com poucas exceções, apenas em viciados em drogas injetáveis, em inoculações acidentais com quantidades mínimas de sangue ou mais raramente, através das realizações de acupuntura e tatuagens. Em regiões de prevalência alta, a transmissão perinatal adquire grande importância (FERREIRA, 2000).

Os riscos de transmissão vertical aumentam quando se tem alta carga viral materna, Além disso, o risco de infecção do concepto após uma hepatite B aguda, autolimitada, ocorrida na gravidez, depende do período gestacional em que a infecção ocorreu. É maior o risco se a mãe se infectar no terceiro trimestre. Do ponto de vista clínico, a associação de hepatite viral aguda e gravidez é pouco frequente, porém, essa infecção é a causa mais comum de icterícia na gestação (ARRAES et al., 2003).

Os principais mecanismos envolvidos na transmissão do VHB estão relacionados à exposição percutânea de sangue e seus derivados, transmissão perinatal (vertical) e transmissão sexual. Outros mecanismos de transmissão seriam a transmissão intrafamiliar e em ambiente fechado onde ocorram respingos de sangue contaminado pelo VHB, nas paredes, como exemplo, em unidades de hemodiálise. Em áreas onde a hepatite B é altamente endêmica, o vírus é frequentemente transmitido durante a infância, seja através de mães portadoras para o recém-nascido, durante a gravidez (a chance de transmissão aumenta à medida que se aproxima o término da gravidez e é muito maior nas portadoras agudas do que nas crônicas) ou parto, ou através de crianças para crianças, provavelmente devido a lesões na pele, tais como escabiose (BRASIL et al., 2003).

#### **2.7.4 Grupos de risco da hepatite B**

Esses grupos populacionais, considerados mais expostos ao VHB, são denominados grupos de risco, nos quais estão incluídos: receptores de transfusões de sangue e/ou derivados, hemodialisados crônicos, hemofílicos, toxicômanos, filhos de mães portadoras de vírus, crianças com retardo mental,

presidiários, homossexuais masculinos, prostitutas, barbeiros, manicures e os profissionais da área da saúde (FERNANDES et al., 1999).

## **2.8 Hepatite C**

A hepatite C é a maior responsável por transplante de fígado em todo o mundo, estima-se que 90% das contaminações de vírus de hepatite devido à transfusão de sangue provêm da hepatite C. Infelizmente, é uma doença que ainda não tem vacina preventiva.

### **2.8.1 História da hepatite C**

A hepatite C é um sério problema de saúde pública mundial, não somente pelo grande número de pessoas infectadas pelo vírus da hepatite C (VHC), mas também devido ao considerável número de indivíduos que toma conhecimento de sua situação patológica somente ao doar sangue, tornando-os um importante elo na cadeia de transmissão viral, perpetuando a doença. Mais de 180 milhões de pessoas no mundo são portadores crônicos do VHC, sendo 2 milhões no Brasil (CIORLIA; ZANETTA, 2007).

Sabemos hoje que a hepatite C compete com a doença hepática alcoólica como a maior causa de doença crônica do fígado, podendo ser a vencedora em várias áreas geográficas. Estima-se que 3% da população mundial esteja contaminada, sendo relevante o número de pessoas que desconhece o fato de albergar o vírus. As altas porcentagens de cronicidade da doença, seu potencial

evolutivo para cirrose, assim como o fato de ser a mais frequente etiologia diagnosticada em casos de transplante hepático, fazem com que constitua grave problema de saúde pública (STRAUSS, 2001).

A hepatite C é a causa mais comum de doença hepática crônica, a primeira causa de transplante de fígado no mundo inteiro. O vírus da hepatite C (HCV) foi identificado em 1989 como o agente da antiga hepatite “não A, não B”. O único hospedeiro natural conhecido é o ser humano, hoje são reconhecidas diversas causas de falhas de tratamentos. Pode-se observar outras causas, como o consumo de álcool e drogas endovenosas. Entre os fatores relacionados com as características do paciente, podemos citar idade mais avançada, sexo masculino, obesidade, raça negra e presença de fibrose avançada e cirrose (ROSENTHAL, 2007).

### **2.8.2 Sintomas da hepatite C**

O surgimento de sintomas em pessoas com hepatite C aguda é muito raro. Entretanto, os que mais aparecem são cansaço, tontura, enjoos e/ ou vômitos, febre, dor abdominal, pele e olhos amarelados, urina escura e fezes claras, na fase crônica a doença pode evoluir para cirrose hepática e câncer de fígado (BRASIL, s.d.c).

### 2.8.3 Transmissão da hepatite C

Devido à via de transmissão compartilhada, cerca de um terço dos pacientes infectados com HIV (vírus da imunodeficiência humana) também estão co-infectados com o vírus da hepatite C. Portanto, é recomendável a triagem sorológica para o HIV em pacientes com hepatite B e C (ROSENTHAL, 2007).

A hepatite C transmite-se majoritariamente por contato direto de sangue com sangue. A forma mais comum de transmissão é a partilha de material utilizado para a injeção de droga principalmente pelas agulhas e seringas com sangue contaminado. O vírus pode sobreviver durante semanas nas seringas e nos tubos de lubrificantes. Está comprovado que partilhar o material para aspirar drogas (palhinhas ou notas) também implica risco de transmissão (CARTER, 2010).

Para Oliveira (2008), a influência do tipo de parto na transmissão perinatal do VHC é completamente desconhecida. O parto via vaginal já foi associado a aumento de risco da transmissão no passado.

Embora existam evidências de que a transmissão sexual do VHC ocorra, essa forma de transmissão tem importância epidemiológica secundária quando comparada às formas de transmissão percutânea (PRADO, 2008).

São justamente as exposições a materiais contaminados, que normalmente não são valorizadas, como procedimentos feitos por maus profissionais em acupuntura, piercings, tatuagens, tesouras, alicates, giletes etc. que passam despercebidas, sendo atualmente os meios de transmissão mais importantes para a disseminação silenciosa e oculta da epidemia, pois a triagem sorológica em bancos de sangue passou a ser instituída após a descoberta do vírus, a transmissão ocorre através do contato percutâneo com sangue ou seus

derivados, comportamento sexual de alto risco e, mais raramente, através da transmissão vertical e do aleitamento. (ROSENTHAL, 2007).

Segundo Jorge (2005), relatos recentes mostraram a presença do vírus em outras secreções (leite, saliva, urina e esperma), a quantidade do vírus parece ser pequena demais para causar infecção por essas vias.

#### **2.8.4 Grupos de risco da hepatite C**

- a) São considerados grupos de alto risco para aquisição do vírus da hepatite C:
- b) usuários de drogas injetáveis e inaláveis;
- c) pessoas com exposições percutâneas frequentes – por exemplo, profissionais da saúde; imigrantes de países onde há alta prevalência; (ROSENTHAL, 2007).

A ocorrência de VHC entre os profissionais da área de saúde varia de 2% a 10%, associando-se o risco de contágio com o tempo de serviço, realização de procedimentos invasivos e ocorrência de acidentes percutâneos. Suas altas taxas de prevalência estão relacionadas com os chamados grupos de risco (hemofílicos, pacientes hemolizados, aqueles que receberam múltiplas transfusões de sangue, toxicômanos) (CIORLIA; ZANETTA, 2007).



## 2.9 Análise de sobrevivência

Análise de sobrevivência é um conjunto de métodos estatísticos que estudam o tempo até a ocorrência de um evento de interesse. Em termos probabilísticos, representamos o tempo por uma variável aleatória não negativa,  $T$ , contínua, com função densidade de probabilidade dada por  $f(t), t \geq 0$ . É importante salientar uma peculiaridade nos dados de sobrevivência: a presença de censuras, termo que será abordado na seção (2.9.1). A ocorrência de censuras impossibilita o uso de procedimentos estatísticos convencionais como análise de experimentos. Um interesse em análise de sobrevivência é a estimação da probabilidade de um indivíduo sobreviver após um tempo  $t$ , conhecida como função de sobrevivência que será melhor detalhada na seção (2.9.2). Essa função pode ser estimada por uma técnica não paramétrica (as probabilidades são determinadas a partir de contagens, sem uso de uma distribuição de probabilidade conhecida), por exemplo, o produto limite de Kaplan-Meier, a ser visto em (2.9.5), porém, esse estimador não considera as covariáveis relacionadas com o tempo de sobrevivência. Para considerar as covariáveis nos tempos de sobrevivência, devem-se utilizar os modelos paramétricos que assumem uma distribuição de probabilidade conhecida para os tempos de sobrevivência.

### 2.9.1 Censura

Os estudos em análise de sobrevivência envolvem uma resposta temporal e são frequentemente de longa duração. Mesmo sendo longos, os estudos clínicos de sobrevivência podem terminar antes que ocorra o evento de interesse para todos os casos da amostra. Uma característica importante decorrente destes estudos é a presença de observações incompletas ou parciais do tempo de sobrevivência denominadas censuras. É importante notar que, mesmo censuradas, todas as observações de um estudo de sobrevivência devem ser usadas na análise estatística.

Para Allison (1995), as censuras são especificadas de formas diferentes e ocorrem por diversos motivos. As formas diferentes estão basicamente norteadas entre censura à esquerda e censura à direita.

Em um estudo em que se observa o tempo até a morte de pacientes, a censura pode ocorrer devido à morte fora do período de duração do estudo, ou quando não for possível observar o acontecimento de interesse, porque se perdeu o acompanhamento do indivíduo durante o estudo. Em ambos os casos, o indivíduo entra no estudo no instante  $t_0$  e morre no instante  $t_0 + t$  onde  $t$  é desconhecido, seja pelo indivíduo ainda se encontrar vivo no final do estudo, seja por ter sido perdido durante o acompanhamento. Se houver informação de que o indivíduo se encontra vivo no instante  $t_0 + c$  em que  $c < t$ , o instante é designado por tempo de sobrevivência com censura à direita (BASTOS; ROCHA, 2006).

Todos os pacientes que necessitaram de um tempo superior ao tempo final do acompanhamento para alcançarem a cura da hepatite são dados com censura à direita. Todos os pacientes que morreram por quaisquer motivos, abandonaram o tratamento ou mudaram de cidade são casos de censura à direita

do tipo aleatório, a censura aleatória ocorre quando as observações são encerradas por razões que não estão sob controle do investigador.

Para representar o processo de censura aleatória, é necessário o uso de duas variáveis aleatórias. Suponha que o tempo de ocorrência de um evento de interesse (falha) de um indivíduo seja representado pela variável aleatória  $T$  em que  $T$  assume valores positivos, usualmente contínua e geralmente especificada em análise de sobrevivência pela sua função de sobrevivência ou pela função de taxa de risco. Seja  $C$  uma variável aleatória independente de  $T$  representando o tempo de censura associado a esse indivíduo, então os dados observados consistem em  $t = \min(T, C)$  e o indicador de censura é dado por.

$$\delta = \begin{cases} 0, & \text{se o tempo de sobrevivência é censurado} \\ 1, & \text{para tempo não censurado.} \end{cases} \quad (2.1)$$

A censura à esquerda ocorre quando alguns dos indivíduos podem ter já experimentado o evento antes do início do estudo (ALLISON, 1995). Consideremos um estudo em que se observa a menarca (início da menstruação), por exemplo, ao considerar a idade da menarca como sendo 12 anos. No acompanhamento de cada adolescente, pode-se verificar que algumas já tiveram a menarca antes dos 12 anos, para essas adolescentes temos um tipo de censura à esquerda.

Conforme definição anterior, os indivíduos C e E representados na Figura 1 não obtiveram o evento de interesse durante todo o estudo, sendo, portanto casos de censura à direita.

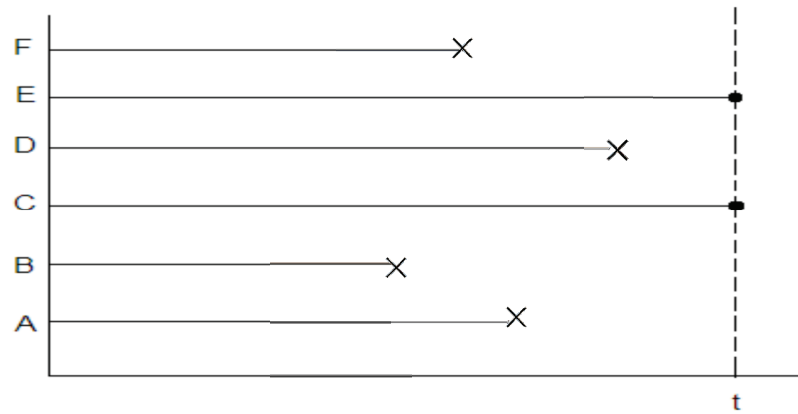


Figura 1 Censura à direita  
 Fonte: Carvalho et al.(2011)

Os indivíduos B, C e F representados na Figura 2 são casos de censura à direita do tipo aleatória.

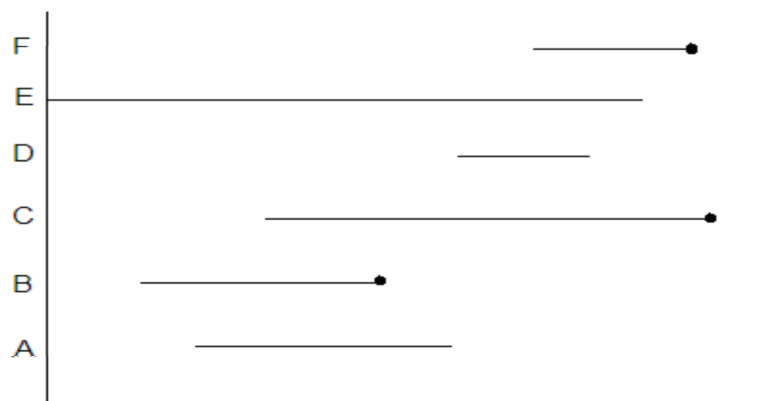


Figura 2 Censura à direita do tipo aleatória  
 Fonte: Carvalho et al.(2011)

### 2.9.2 Função de sobrevivência

Em um estudo onde se acompanha cada indivíduo até que ocorra o evento de interesse, como, por exemplo, a morte, a função de sobrevivência, denotada por  $S(t)$ , do tempo até a ocorrência do evento de interesse ( $T$ ) é dada por:

$$S(T) = P(T > t) = \int_t^{\infty} f(u) du = 1 - F(t), \quad (2.2)$$

em que  $F(T)$  é a função de distribuição acumulada da variável aleatória  $t$ .

A função de distribuição acumulada  $F(t)$ , é a probabilidade de que tal evento de interesse ocorra até o tempo  $t$ .

$$F(t) = P(T \leq t) \quad (2.3)$$

A função de sobrevivência é uma função não crescente no tempo com as propriedades de que a probabilidade de sobreviver pelo menos ao tempo zero é um e a probabilidade de sobreviver no tempo infinito é zero. Isto é,

$$S(t) = 1 \text{ para } t = 0 \quad (2.4)$$

$$S(t) = 0 \text{ para } t = \infty. \quad (2.5)$$

Para descrever a função de sobrevivência, é utilizada uma representação gráfica de  $S(t)$ , ou seja, o gráfico de  $S(t)$  versus  $t$  que é chamado de curva de sobrevivência. Uma curva íngreme representa razão de sobrevivência baixo ou curto tempo de sobrevivência, e uma curva de sobrevivência gradual ou plana representa taxa de sobrevivência alta ou sobrevivência longa.

A curva de sobrevivência pode ser usada para comparar distribuições de sobrevivência de dois ou mais grupos e também para determinar quantidades relevantes como a mediana (COLOSIMO; GIOLO; 2006).

Para Alisson (1995) e Cantor (2003), o tempo de sobrevivência mediano é a medida mais citada e interpretada em análise de sobrevivência. É mais usada do que a média, pois, como a distribuição do tempo de sobrevivência é pouco simétrica, as medidas robustas são mais indicadas. Por definição, o tempo mediano é o tempo depois do qual 50% dos indivíduos estão vivos, ou seja, o tempo no qual  $\hat{S}(t) = 0,5$ . Como o tempo de sobrevivência normalmente não é observado de forma contínua, a sobrevivência é uma função em degraus, e por isso o tempo de sobrevivência mediano é definido como o menor tempo para o qual o valor estimado de  $S(t)$  é menor ou igual a 50% :

$$\hat{t}_{mediano} = \min \{t_j \mid \hat{S}(t_j) \leq 0,5\} \quad (2.6)$$

em que  $j$  é o índice dos tempos observados de ocorrência do evento.

A Figura 3 representa as curvas de sobrevivência para três condições A, B e C estimadas por método não paramétrico, a linha vertical sobre a condição B determina o tempo mediano como sendo igual a 7 meses.

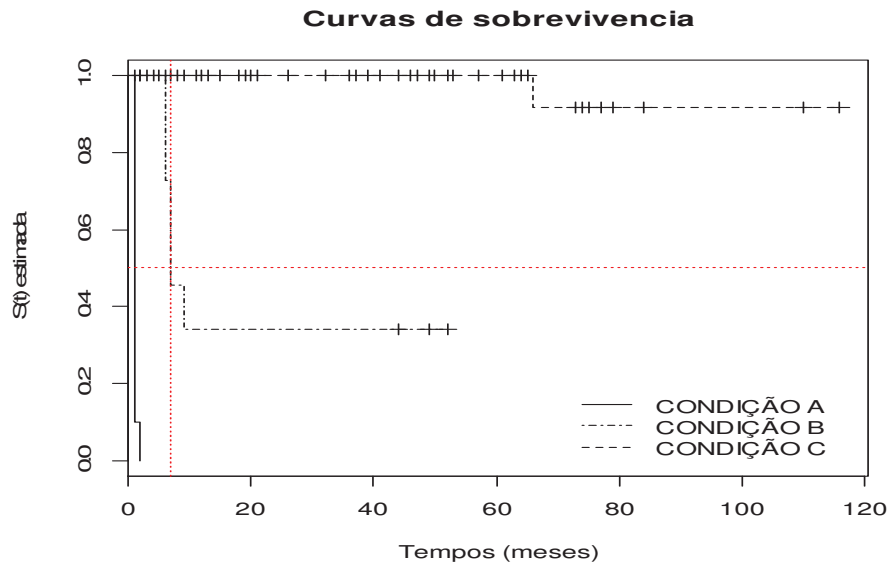


Figura 3 Curvas de sobrevivência obtidas por método não paramétrico

### 2.9.3 Função taxa de risco

A probabilidade do evento de interesse ocorrer em um intervalo de tempo  $[t, t + \Delta t]$  pode ser representada em termos da função de sobrevivência da seguinte forma:

$$S(t) - S(t + \Delta t) \quad (2.7)$$

A taxa de risco no intervalo  $[t, t + \Delta t]$  é definida como a probabilidade de que o evento ocorra neste intervalo, dado que não ocorreu antes de  $t$  dividida pelo comprimento do intervalo. Dessa forma, a função taxa de risco no intervalo  $[t, t + \Delta t]$  é representada por:

$$\frac{S(t) - S(t + \Delta t)}{\frac{t + \Delta t}{S(t)}} = \frac{S(t) - S(t + \Delta t)}{[(t + \Delta t) - t] S(t)} \quad (2.8)$$

Logo,  $h(t)$  pode ser escrita do seguinte modo:

$$h(t) = \frac{S(t) - S(t + \Delta t)}{\Delta t S(t)} \quad (2.9)$$

A função taxa de risco é o risco de o evento ocorrer em um tempo " $t$ ".

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \quad (2.10)$$

Pode-se escrever a função de risco em termos da função de distribuição acumulada  $F(t)$  e da função densidade de probabilidade  $f(t)$  da forma:

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)} \quad (2.11).$$

#### 2.9.4 Estimação da função de sobrevivência

Um passo inicial nos estudos de tempos de sobrevivência é a estimação da função de sobrevivência. Esses estudos frequentemente apresentam observações censuradas, o que requer técnicas estatísticas especializadas para acomodar a informação contida nestas observações.

O estimador não paramétrico de Kaplan-Meier pode ser utilizado para estimar a função de sobrevivência, pois usa os próprios dados para estimar as quantidades necessárias da análise, sem fazer uso de uma distribuição dos tempos de sobrevivência.



### 2.9.5 Estimador de Kaplan-Meier

No ano de 1958, Kaplan e Meier propuseram um estimador não paramétrico, que foi assim nomeado como estimador de Kaplan-Meier. O estimador de Kaplan-Meier para estimar a função de sobrevivência é também chamado de estimador limite-produto. A função de sobrevivência é uma função escada com degraus nos tempos observados de ocorrência do evento de interesse de tamanho  $1/n$ , em que  $n$  é o tamanho da amostra. Se existirem empates em um certo tempo  $t$ , o tamanho do degrau fica multiplicado pelo número de empates. São considerados como empates os dados observados que apresentam um mesmo tempo de ocorrência do evento de interesse, mesmo mês de cura da hepatite, já que, para este trabalho, o tempo até a cura da hepatite foi determinado em dias.

O estimador de Kaplan-Meier é então definido como:

$$\hat{S}(t) = \prod_{j:t_j < t} \left(1 - \frac{d_j}{n_j}\right), \quad (2.12)$$

em que  $t_1 < t_2 \dots < t_k$  correspondem aos  $k$  tempos distintos e ordenados, com  $d_j$  e  $n_j$  representando, respectivamente, o número de eventos de interesse ocorridos e número de dados sob risco de ocorrência do evento, considerando cada intervalo.

Um estimador como apresentado na equação (2.12) que não considera uma distribuição de probabilidade para modelar os tempos de sobrevivência é conhecido como um estimador não paramétrico. Esse estimador é frequentemente utilizado para estimar a função de sobrevivência e através de análises gráficas compara os tipos distintos de grupos existentes em cada

covariável, exemplo masculino e feminino, bem como pode ser utilizado para escolher entre os modelos paramétricos qual é o mais indicado.

### **2.9.6 Casos de aplicação de Análise de Sobrevivência**

Em um estudo sobre o tempo de permanência no curso de graduação de física de uma universidade brasileira, Junior, Silveira e Ostermann (2012), utilizaram o método de análise de sobrevivência para descrever quando a evasão e a diplomação ocorrem e quais fatores influenciam na permanência prolongada no curso. Para esse estudo,  $F(t)$  foi entendida como a fração dos indivíduos que foram contemplados com o evento de interesse.

Silveira e Lamounier (2007) utilizaram uma curva de sobrevivência para representar o tempo mediano de aleitamento materno em crianças, que foi de 10,85 meses, em três municípios carentes do Estado de Minas Gerais na Região do Alto Jequitinhonha. Menesia et al. (2001), em estudo sobre pacientes com aids em uma cidade do sudeste brasileiro, compararam através da curva de sobrevivência um grupo de indivíduos do sexo masculino e outro grupo de indivíduos do sexo feminino e concluíram que o grupo feminino possui sobrevivência maior.

Menesia et al. (2001) em um estudo com 2214 pacientes diagnosticados com aids em Ribeirão Preto, São Paulo, cujo objetivo foi avaliar o tempo de vida desses pacientes, concluiu que a mediana foi de 355 dias para as mulheres enquanto que para os homens a mediana foi de 296 dias, considerado o período de 1986 a 1997.

Com a finalidade de conhecer a prática do aleitamento materno e identificar fatores associados à sua duração em crianças de famílias de baixa

renda residentes na região sul da cidade de Curitiba-PA, Baptista, Andrade e Giolo (2009) apresentaram  $d_j$  como o número de crianças que deixaram de ser amamentadas, sendo este o evento de interesse e  $n_j$  representa o número de crianças observadas em amamentação que estavam sob risco de ocorrência do evento de interesse.

Na tabela 1, (NAKANO; CARRASCO, 2006), em um estudo do tempo para a gravidez para casais que desejavam ter filhos, as mulheres em estudo pararam de usar qualquer tipo de contraceptivo a partir do dia de início do experimento. Nesse estudo, a variável  $T$  representa o número de meses até que a mulher tenha engravidado. Foi considerado  $t=0$  quando a mulher engravidou no primeiro mês de estudo. O tempo  $t$  foi dado em meses,  $d_j$  representa o número de mulheres que engravidaram em cada mês considerado,  $n_j$  o número de mulheres expostas ao risco de engravidar em cada mês e  $\hat{S}(t)$  a probabilidade em cada tempo estimada por Kaplan-Meier.

Tabela-1 Estimativas através da função de sobrevivência pelo método de Kaplan-Meier

Tempos	$n_j$	$d_j$	$\hat{S}(t)$
01	611	199	0,6527
1	412	103	0,4729
2	309	64	0,3613
3	245	36	0,2984
4	197	33	0,2365
5	157	30	0,1769
6	118	18	0,1371
7	95	13	0,1058
8	77	9	0,0814
9	63	10	0,0488
10	48	3	0,0342
11	43	0	0,0342

Fonte: Dados de Tietze

## 2.10 Modelos Probabilísticos

Alguns modelos probabilísticos são muito utilizados na análise de sobrevivência. Caso seja possível associar  $T$  com uma distribuição de probabilidade, entre eles tem-se:

### 2.10.1 Modelo exponencial

Dos modelos probabilísticos usados para descrever o tempo até a ocorrência do evento de interesse, o modelo exponencial é o mais simples. Essa distribuição apresenta um único parâmetro e é a única que se caracteriza por ter uma função de taxa de risco constante. A função de densidade é dada por:

$$f(t; \alpha) = \frac{1}{\alpha} \exp\left\{-\left(\frac{t}{\alpha}\right)\right\}, t \geq 0 \quad (2.13)$$

em que o parâmetro  $\alpha > 0$  é o tempo médio de vida. As funções de sobrevivência  $S(t)$  e de taxa de falha  $h(t)$  são dadas, respectivamente, por:

$$S(t) = \exp\left\{-\left(\frac{t}{\alpha}\right)\right\}, \quad (2.14)$$

$$h(t) = \frac{1}{\alpha} \text{ para } t \geq 0 \quad (2.15)$$

### 2.10.2 Modelo Weibull

A distribuição de Weibull (1939) teve a sua aplicabilidade discutida por este mesmo autor (WEIBULL, 1951, 1954). A sua popularidade em aplicações práticas se deve ao fato de ela apresentar uma grande variedade de formas, todas com uma propriedade básica: a sua função de taxa de risco é monótona, isto é, ela é crescente, decrescente ou constante. Para uma variável aleatória  $T$  com distribuição de Weibull, tem-se a função de densidade de probabilidade dada por:

$$f(t; \alpha, \gamma) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\gamma\right\}, \text{ para } t \geq 0 \quad (2.16)$$

em que  $\gamma$  é o parâmetro de forma,  $\alpha$  o parâmetro de escala e são ambos positivos. Para esta distribuição, as funções de sobrevivência e de risco são, respectivamente,

$$S(t) = \exp\left\{-\left(\frac{t}{\alpha}\right)^\gamma\right\} \quad (2.17)$$

$$h(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1}, \quad (2.18)$$

para  $t \geq 0$ ,  $(\alpha \text{ e } \gamma) > 0$ . Para  $\gamma=1$ , tem-se a distribuição exponencial e, sendo assim, a distribuição exponencial é um caso particular da distribuição de Weibull. Ao se verificar a duração do desemprego na região metropolitana de Curitiba, Lira e Sampaio (2006) concluíram que a distribuição de Weibull foi, entre as comparadas, a que melhor moldou os dados observados. Portanto, por essa distribuição, verificaram que existe uma relação positiva entre o número de anos de estudo e a duração do desemprego.

### 2.10.3 Modelo Log-normal

A função de densidade de uma variável aleatória  $T$  com distribuição log-normal é dada por:

$$f(t; \mu, \sigma) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\log(t) - \mu}{\sigma}\right)^2\right\}, \quad t > 0, \quad (2.19)$$

em que  $\mu$  é a média do logaritmo do tempo de falha e  $\sigma$  é o desvio padrão. Como o nome sugere, o logaritmo de uma variável com distribuição log-normal de parâmetros  $\mu$  e  $\sigma$  tem uma distribuição normal com media  $\mu$  e desvio padrão  $\sigma$ . As funções de sobrevivência e de taxa de falha de uma variável log-normal não apresenta uma forma analítica explícita e são, desse modo, representadas, respectivamente, por:

$$S(t) = \Phi\left(\frac{-\log(t) + \mu}{\sigma}\right) \quad (2.20)$$

$$h(t) = \frac{f(t)}{S(t)}, \quad (2.21)$$

em que  $\Phi(\cdot)$  é a função de distribuição acumulada de uma normal padrão (COLOSIMO; GIOLO, 2006).

### 2.10.4 Modelo Gama-Generalizada

A distribuição Gama-Generalizada foi introduzida por Stacy (1962) e é caracterizada por três parâmetros,  $\gamma, k, \alpha$ , todos positivos. Sua função de densidade é dada por:

$$f(t) = \frac{\gamma}{\Gamma(k)\alpha^{\gamma k}} t^{\gamma(k-1)} \exp\left\{-\left(\frac{t}{\alpha}\right)^{\gamma}\right\}, \quad t > 0 \quad (2.22)$$

Em que  $\Gamma(k) = \int_0^{\infty} x^{k-1} \exp\{-x\} dx$  é a função gama. Utilizando integração por

partes na integral  $\int_0^{\infty} x^{k-1} \exp\{-x\} dx$ , é possível mostrar que a função

$$\Gamma(k) = (k-1) \cdot \Gamma(k-1), \text{ pois}$$

$$\begin{aligned} \Gamma(k) &= \int_0^{\infty} x^{k-1} \exp\{-x\} dx \\ &= \left(-x^{k-1} \cdot \exp\{-x\}\right)\Big|_0^{\infty} + \int_0^{\infty} (k-1)x^{k-2} \exp\{-x\} dx \\ &= 0 + (k-1)\Gamma(k-1) \\ &= (k-1)\Gamma(k-1) \end{aligned}$$

Os parâmetros  $\alpha$  de escala e os dois de forma  $\gamma$  e  $k$  tornam a distribuição Gama-Generalizada bastante flexível. Observe a partir da função de densidade da distribuição Gama-Generalizada que:

Para  $\gamma=k=1$  tem-se  $T \sim \text{Exp}(\alpha)$ ,

Para  $k=1$  tem-se  $T \sim \text{Weibull}(\gamma, \alpha)$

e para  $k \rightarrow \infty$  tem-se  $T \sim \text{Log-normal}$

Portanto, tem-se que a distribuição Gama-Generalizada inclui, como casos particulares, as distribuições: Exponencial, Weibull e Log-normal. Essa propriedade da Gama-Generalizada faz com que a mesma seja de grande utilidade, por exemplo, na discriminação entre modelos probabilísticos.



### 2.11. Métodos de estimação paramétrico

Como estimador paramétrico, o método de mínimos quadrados é inapropriado para estudos de tempo de vida. A principal razão é a sua incapacidade de incorporar censuras no seu processo de estimação. O método de máxima verossimilhança surge como uma opção apropriada para esse tipo de dado. Ele incorpora as censuras e possui propriedades ótimas para grandes amostras. Se, por exemplo, a distribuição do tempo de ocorrência do evento de interesse é a de Weibull para cada combinação diferente de  $\gamma$  e  $\alpha$ , têm-se diferentes distribuições de Weibull. O estimador de máxima verossimilhança determina aquele par de  $\gamma$  e  $\alpha$  que melhor represente a amostra observada. A tradução em termos matemáticos para a frase “a distribuição que melhor represente a amostra observada” é encontrar o valor de  $\theta$  (um ou vários parâmetros), que maximize a probabilidade da amostra observada ocorrer. Suponha que a população é caracterizada pela função de densidade  $f(t)$ . Por exemplo, na equação (2.13) a função de verossimilhança denotada por  $L$  para um parâmetro genérico  $\theta$  dessa população é então expressa por:

$$L(\theta) = \prod_{i=1}^n f(t_i; \theta) \quad (2.23)$$

A dependência de  $f$  em  $\theta$  é preciso agora ser mostrada, pois  $L$  é função de  $\theta$ . Nessa expressão,  $\theta$  pode representar um único parâmetro ou um conjunto de parâmetros. A função de verossimilhança  $L(\theta)$  mostra que a contribuição de cada observação não censurada é a sua função de densidade. A contribuição de cada observação censurada não é, contudo, a sua função de densidade. Essas observações nos informam que o tempo até a ocorrência do evento de interesse é maior que o tempo de censura observado e, portanto, a

contribuição dos tempos de censura para  $L(\theta)$  é a sua função de sobrevivência  $S(t)$ , portanto a expressão para a função de verossimilhança é dada por:

$$L(\theta) = \prod_{i=1}^n [f(t_i; \theta)]^{\delta_i} [S(t_i; \theta)]^{1-\delta_i}, \quad (2.24)$$

em que  $\delta_i$  é a variável indicadora do tempo até ocorrência do evento, como visto em (2.1).

A função de verossimilhança para a distribuição exponencial, obtida a partir das equações (2.13) e (2.14) é, portanto a que se segue.

$$\begin{aligned} L(\alpha) &= \prod_{i=1}^n \left[ \frac{1}{\alpha} \exp\left(-\left(\frac{t_i}{\alpha}\right)\right) \right]^{\delta_i} \left[ \exp\left(-\left(\frac{t_i}{\alpha}\right)\right) \right]^{1-\delta_i} \\ \Rightarrow L(\alpha) &= \prod_{i=1}^n \left[ \frac{1}{\alpha} \right]^{\delta_i} \exp\left(-\left(\frac{t_i}{\alpha}\right)\right) \end{aligned} \quad (2.25)$$

A função de verossimilhança para a distribuição Weibull, obtida a partir das equações (2.16) e (2.17) é, portanto a que se segue:

$$\begin{aligned} L(\gamma, \alpha) &= \prod_{i=1}^n \left[ \frac{\gamma}{\alpha^\gamma} t_i^{\gamma-1} \exp\left(-\left(\frac{t_i}{\alpha}\right)^\gamma\right) \right]^{\delta_i} \left[ \exp\left(-\left(\frac{t_i}{\alpha}\right)^\gamma\right) \right]^{1-\delta_i} \\ \Rightarrow L(\gamma, \alpha) &= \prod_{i=1}^n \left[ \frac{\gamma}{\alpha^\gamma} t_i^{\gamma-1} \right]^{\delta_i} \exp\left(-\left(\frac{t_i}{\alpha}\right)^\gamma\right) \end{aligned} \quad (2.26)$$

Os estimadores de máxima verossimilhança são os valores de  $\theta$  que maximizam  $L(\theta)$  ou, equivalentemente, o logaritmo de  $L(\theta)$ , isto é,  $\log(L(\theta))$ . Estes são obtidos resolvendo-se o sistema de equações:

$$U(\theta) = \frac{\partial \log L(\theta)}{\partial \theta} = 0 \quad (2.27)$$

Como as equações são não lineares em  $\beta$  e não apresentam solução analítica, devem ser resolvidas numericamente, o que, usualmente, envolve o método numérico de Newton Raphson e a utilização de um pacote estatístico.

## 2.12 Modelo de regressão paramétrico

Os modelos de regressão paramétricos, também conhecidos como modelo de tempo de vida acelerada, utilizam famílias paramétricas de distribuições para o tempo de sobrevivência  $T$ . Utilizar um modelo linear como:

$$Y = \beta_0 + \beta_1 x + \varepsilon, \quad (2.28)$$

em que  $Y$  é a variável resposta,  $x$  é a covariável,  $\beta_0$  e  $\beta_1$  são os parâmetros a serem estimados e  $\varepsilon$  é o erro aleatório com distribuição normal. Não é indicado para dados em análise de sobrevivência devido à especificidade dos dados de sobrevivência, que tem como variável resposta o tempo até ocorrência de um evento, a presença de censura e ainda que a distribuição da variável resposta tenda também a ser assimétrica na direção dos maiores tempos de sobrevivência, o que torna inadequado o uso da distribuição normal para o erro. Para solucionar esse problema deve-se utilizar o componente determinístico não linear da forma  $\exp\{\beta_0 + \beta_1 x\}$  nos parâmetros e uma distribuição assimétrica para o erro.

### 2.12.1 Modelo de regressão exponencial

O problema estatístico na modelagem em dados de análise de sobrevivência citado anteriormente deve ser solucionado ao utilizar o componente determinístico da forma  $\exp\{\beta_0 + \beta_1 x\}$  nos parâmetros e uma distribuição assimétrica para o erro, neste caso, uma distribuição exponencial. Portanto, a combinação desse componente determinístico e a distribuição exponencial para o erro gera o seguinte modelo:

$$T = \exp\{\beta_0 + \beta_1 x\} \varepsilon, \quad (2.29)$$

que é o modelo de regressão exponencial. Esse modelo admite uma relação linear entre  $T$  e  $x$  no seu componente determinístico e erro com distribuição assimétrica. Ainda pode ser linearizado se for considerado o logaritmo de  $T$ . Assim, teremos:

$$Y = \log(T) = \beta_0 + \beta_1 x + \nu \quad (2.30)$$

em que  $\nu = \log(\varepsilon)$ . O modelo (2.30) é semelhante ao modelo linear (2.28), com exceção da distribuição dos erros que não são normais. O erro  $\nu$  segue uma distribuição do valor extremo padrão ( $f(\nu) = \exp(\nu - \exp(\nu))$ ), a função de sobrevivência para o modelo de regressão exponencial fica assim definida:

$$S(t | x) = \exp(-\exp(\beta_0 + \beta_1 x)t) \quad (2.31)$$

O processo de inferir o vetor de parâmetros  $\theta = (\beta_0, \beta_1)$  do modelo (2.30) deve ser o método de estimação de máxima verossimilhança que apresenta duas partes na construção da função de verossimilhança: primeiro, a questão dos tempos até a ocorrência do evento de interesse, a contribuição dos tempos dos indivíduos que foram curados para a função de verossimilhança é a probabilidade de que o mesmo tenha sido curado de hepatite, isto é dado pela

função de densidade  $f(t_i | x_i)$ ; e, em segundo, os dados censurados, a contribuição de cada dado censurado para a função de verossimilhança é a probabilidade de sobreviver ao tempo  $t_i$  e isto é dado por sua função de sobrevivência  $S(t_i | x_i)$ . A função de verossimilhança para o modelo linear na forma (2.30) é então escrita para uma amostra de tamanho  $n$  como:

$$L(\theta) = \prod_{i=1}^n [f(y_i, |x_i)]^{\delta_i} [S(y_i, |x_i)]^{(1-\delta_i)}, \quad (2.32)$$

em que  $y_i = \log(t_i)$ , ou, ainda, para o modelo escrito na forma (2.29), tem-se

$$L(\theta) = \prod_{i=1}^n [f(t_i, |x_i)]^{\delta_i} [S(t_i, |x_i)]^{(1-\delta_i)} \quad (2.33)$$

Para obtenção dos estimadores de máxima verossimilhança, é necessário substituir as funções de densidade e sobrevivência pela distribuição do valor extremo em (2.32) ou da exponencial em (2.33). Considerando isso em (2.32) e tomando-se o logaritmo de  $L(\theta)$ , tem-se:

$$\begin{aligned} \log L(\theta) &= \sum_{i=1}^n [\delta_i ((y_i - \beta_0 - \beta_1 x_i) - \exp(y_i - \beta_0 - \beta_1 x_i)) \\ &\quad - (1 - \delta_i) \exp(y_i - \beta_0 - \beta_1 x_i)] \\ \log L(\theta) &= \sum_{i=1}^n [\delta_i (y_i - \beta_0 - \beta_1 x_i) - \exp(y_i - \beta_0 - \beta_1 x_i)] \end{aligned} \quad (2.34)$$

Os estimadores de máxima verossimilhança são os valores de  $\theta = (\beta_0, \beta_1)$  que maximizam a função  $\log L(\theta)$  em (2.34). Portanto, é necessário determinar as derivadas de (2.34) em função de  $\theta = (\beta_0, \beta_1)$ , igualar as expressões obtidas a zero e resolver o sistema de equações resultante. Como as equações são não lineares em  $\beta$  e não apresentam solução analítica, devem ser

resolvidas numericamente, o que, usualmente, envolve o método numérico de Newton-Raphson e a utilização de um pacote estatístico.

### 2.12.2 Modelo de regressão Weibull

Uma forma de generalizar o modelo (2.30) é incluir um parâmetro extra de escala em sua formulação. Isso, em modelos lineares, é equivalente a assumir, para os erros, uma distribuição normal com variância  $\sigma^2$ , em vez de uma distribuição normal padrão. O modelo linear (2.30) passa, então, a ter a forma

$$Y = \log(T) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \sigma W = \beta x + \sigma W \quad (2.35)$$

conhecido como modelo de regressão Weibull, pois  $T$  deve ter uma distribuição de Weibull para que  $\log(T)$  tenha uma distribuição ( $f(v) = \exp(v - \exp(v))$ ) do valor extremo padrão com parâmetro de escala  $\sigma$ . Também que  $\beta = (\beta_0, \beta_1, \dots, \beta_p)$ ,  $x = (1, x_1, \dots, x_p)$ ,  $\sigma$  é mais um parâmetro a ser estimado e  $W$  é um vetor de erros com distribuição especificada. Deve-se observar que  $Y$  está associado com as covariáveis  $x$  através de um modelo linear e, portanto, o papel das covariáveis é acelerar ou desacelerar o tempo de sobrevivência  $T$ . Os parâmetros  $\beta$  e  $\sigma$  no modelo acima são estimados pelo método de máxima verossimilhança (COX; HINKLEY, 1974). Logo, a função de sobrevivência fica assim definida:

$$S(t | x) = \exp(-(\exp(\beta x)t)^\gamma) \quad (2.36)$$

em que  $\gamma$  é o parâmetro de forma definido para (2.16).

### 2.13 Adequação do modelo paramétrico

Uma avaliação da adequação do modelo ajustado é parte fundamental da análise dos dados. Neste trabalho, é utilizada uma ferramenta gráfica para essa finalidade.

Os resíduos de Cox-Snell (1968) e os resíduos padronizados são úteis para examinar o ajuste global do modelo. Os resíduos de Cox-Snell, denotado por  $\hat{\epsilon}_i$ , são quantidades determinadas por:

$$\hat{\epsilon}_i = \hat{\Lambda}(t_i | x_i), \quad (2.37)$$

em que  $\Lambda(\cdot)$  é a função de risco acumulado obtida do modelo ajustado. Para os modelos de regressão exponencial, Weibull e Log Normal, os resíduos de Cox-Snell são dados, respectivamente, por:

$$\hat{\epsilon}_i = [t_i \exp(-\hat{\beta} \mathbf{X}_i)] \quad (2.38)$$

$$\hat{\epsilon}_i = [t_i \exp(-\hat{\beta} \mathbf{X}_i)]^\gamma \quad (2.39)$$

$$\hat{\epsilon}_i = -\log \left[ 1 - \Phi \left( \frac{\log(t_i) - \hat{\beta} \mathbf{X}_i}{\sigma} \right) \right] \quad (2.40)$$

Os resíduos  $\hat{\epsilon}_i$  são de uma população homogênea e devem seguir uma distribuição exponencial padrão se o modelo for adequado (Lawless, 1982). Logo, o gráfico  $\hat{\epsilon}_i$ , versus  $\hat{\Lambda}(\hat{\epsilon}_i)$  deve ser aproximadamente uma reta com inclinação igual a 1, quando o modelo proposto for adequado, uma vez que  $\hat{\Lambda}(\hat{\epsilon}_i) = -\log(\hat{S}(\hat{\epsilon}_i))$ , sendo  $\hat{S}(\hat{\epsilon}_i)$  a função de sobrevivência dos  $\hat{\epsilon}_i$ 's obtida pelo estimador de Kaplan-Meier.

## 2.14 Modelo de regressão de Cox

Os estudos em análise de sobrevivência utilizam covariáveis que estão relacionadas com o tempo de sobrevivência. Uma forma de incorporar essas covariáveis e verificar os seus efeitos nos tempos de sobrevivência é através do modelo de riscos proporcionais. Portanto, a função de risco no tempo  $t$ , dado  $x$ , pode ser escrita na forma:

$$h(t/x) = h_0(t) g(\beta, x), \quad (2.41)$$

em que  $g$  é uma função não negativa que deve ser especificada,  $\mathbf{X}$  é o vetor de covariáveis,  $\beta$  é o vetor de parâmetros relacionado com as covariáveis. Esse modelo é composto por dois componentes, um não paramétrico e outro paramétrico, sendo essa a razão do modelo ser do tipo semiparamétrico. O componente não paramétrico  $h_0(t)$  não é especificado e é uma função positiva no tempo, conhecida por função de base ou basal, já que  $h(t) = h_0(t)$  quando  $\mathbf{X} = 0$ . O componente paramétrico é usualmente escrito na forma multiplicativa:

$$g(\beta, \mathbf{X}) = \exp(\beta, \mathbf{X}) = \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) \quad (2.42)$$

Observe que a constante  $\beta_0$ , presente nos modelos paramétricos, não aparece na expressão (2.42), isso deve a presença do componente não paramétrico no modelo que absorve este termo constante. O modelo (2.42), pode então ser escrito na forma:

$$h(t/x) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) \quad (2.43)$$

O vetor de covariáveis  $\mathbf{X}$  no modelo de riscos proporcionais é dado por  $\mathbf{X} = (x_1, x_2, \dots, x_p)$ , sendo  $t_i$  o tempo de sobrevivência do  $i$ -ésimo indivíduo que depende do valor dessas covariáveis, o que se deseja é verificar de que forma e o quanto estas covariáveis influenciam  $t_i$ .



O modelo (2.43) é chamado de riscos proporcionais de Cox, pois a razão das taxas de risco de dois indivíduos diferentes é constante no tempo, isto é, a razão das funções de taxa de risco de dois indivíduos  $i$  e  $j$  é dada por:

$$\frac{h_i(t)}{h_j(t)} = \frac{h_0(t) \exp(\beta \mathbf{X}_i)}{h_0(t) \exp(\beta \mathbf{X}_j)} \quad (2.44)$$

e não depende do tempo.

A suposição básica para o uso do modelo de riscos proporcionais de Cox é que as taxas de risco de dois indivíduos sejam proporcionais ao longo de todo o acompanhamento. A presença do componente não paramétrico no modelo faz com que a estimativa dos parâmetros pelo método de máxima verossimilhança seja inadequada. Assim, Cox (1975) propôs o método da máxima verossimilhança parcial que condiciona a função de verossimilhança nos tempos de sobrevivência e elimina a função basal.

O efeito das covariáveis sobre a função de risco é caracterizado pelos coeficientes  $\beta$ , os quais devem ser estimados pelo método da máxima verossimilhança parcial.

## 2.15 Verossimilhança parcial

Considere uma amostra de  $n$  indivíduos e que nela existam  $k \leq n$  falhas distintas (momentos em que ocorreu cura da hepatite) nos tempos  $t_1 < t_2 < \dots < t_k$ . A probabilidade condicional da  $i$ -ésima observação vir a falhar no tempo  $t_i$ , conhecido o conjunto das observações sob risco no tempo  $t_i$  é:

$$\frac{h_i(t_i)}{\sum_{j \in R(t_i)} h_j(t_i)} = \frac{h_0(t_i) \exp(\boldsymbol{\beta} \mathbf{X}_i)}{\sum_{j \in R(t_i)} h_0(t_i) \exp(\boldsymbol{\beta} \mathbf{X}_j)} = \frac{\exp(\boldsymbol{\beta} \mathbf{X}_i)}{\sum_{j \in R(t_i)} \exp(\boldsymbol{\beta} \mathbf{X}_j)}, \quad (2.45)$$

em que  $R(t_i)$  é o conjunto das observações sob risco no tempo  $t_i$ . É possível verificar que, ao utilizar a probabilidade condicional, o componente não paramétrico  $h_0(t)$  desaparece da equação (2.45). A função de verossimilhança parcial  $L(\boldsymbol{\beta})$  é obtida fazendo o produto das probabilidades condicionais, ou seja:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^k \frac{\exp(\boldsymbol{\beta} \mathbf{X}_i)}{\sum_{j \in R(t_i)} \exp(\boldsymbol{\beta} \mathbf{X}_j)} = \prod_{i=1}^n \left( \frac{\exp(\boldsymbol{\beta} \mathbf{X}_i)}{\sum_{j \in R(t_i)} \exp(\boldsymbol{\beta} \mathbf{X}_j)} \right)^{\delta_i}, \quad (2.46)$$

em que

$$\delta_i = \begin{cases} 0, & \text{se o } i\text{-ésimo tempo de sobrevivência é censurado} \\ 1, & \text{caso contrário.} \end{cases}$$

Os valores de  $\boldsymbol{\beta}$  que maximizam a função de verossimilhança parcial,  $L(\boldsymbol{\beta})$ , são obtidos resolvendo-se o sistema de equações definido por  $U(\boldsymbol{\beta})=0$ , em que  $U(\boldsymbol{\beta})$  é o vetor escore de primeiras derivadas da função  $l(\boldsymbol{\beta})=\log(L(\boldsymbol{\beta}))$ . Isso equivale a:

$$U(\boldsymbol{\beta}) = \sum_{i=1}^n \delta_i [\boldsymbol{\beta} \mathbf{X}_i - \log \sum_{j \in R(t_i)} \exp(\boldsymbol{\beta} \mathbf{X}_j)] = 0 \quad (2.47)$$

A função de verossimilhança parcial (2.46) assume que os tempos de sobrevivência são contínuos e, portanto, não considera a possibilidade de empates nos valores observados. Mas, na prática, podem ocorrer empates nos tempos de falha (indivíduos que obtiveram cura da doença no mesmo mês) ou de censura devido à escala de medida do tempo. Para este estudo sobre hepatites, vários foram os indivíduos que alcançaram a cura da hepatite no primeiro mês de acompanhamento, já que o tempo foi medido em meses. Quando houver

empates entre falhas e censuras, usa-se a convenção de que a censura ocorreu após a falha, para definir quais observações serão incluídas no conjunto de risco em cada tempo de falha. A função de verossimilhança parcial deve ser modificada para incorporar as observações empatadas quando estas estão presentes. A aproximação proposta por Breslow (1972) e Peto (1972) é frequentemente usada nos softwares estatísticos. Considere  $s_i$  o vetor composto pela soma das correspondentes  $p$  covariáveis para os indivíduos que falham no mesmo tempo  $t_i (i=1, \dots, k)$  e  $d_i$  o número de falhas neste mesmo tempo. A aproximação considera a seguinte função de verossimilhança parcial:

$$L(\beta) = \prod_{i=1}^k \frac{\exp(\beta s'_i)}{\left[ \sum_{j \in R(t_i)} \exp(\beta s'_j) \right]^{d_i}} \quad (2.48)$$

Essa aproximação é adequada quando o número de observações empatadas em qualquer tempo não é grande. Caso o número de observações empatadas seja grande é aconselhável utilizar o modelo de regressão de Cox para dados agrupados (LAWLESS, 1982; PRENTICE E GLOECKLER, 1978).

As funções relacionadas a  $h_0(t)$  são importantes no modelo de Cox, considerando que, para um determinado indivíduo, todas as covariáveis têm valores iguais a zero. Sendo  $\Lambda(t)$  a função de taxa de falha acumulada, tem-se:

$$\Lambda(t) = \int_0^t h_0(u) du \quad (2.49)$$

Portanto, a função de sobrevivência de base é:

$$S_0(t) = \exp(-\Lambda(t)) = \exp\left(-\int_0^t h_0(u) du\right) \quad (2.50)$$

Uma estimativa para  $\Lambda(t)$  proposta por Breslow (1972), é uma função escada com saltos nos tempos distintos de falha e expressa por:

$$\Lambda(t) = \sum_{j:t_j < t} \frac{d_j}{\sum_{l \in R_j} \exp(\beta' x_l)}, \quad (2.51)$$

em que  $d_j$  é o número de eventos ocorridos em  $t_j$ .

### 2.16 Teste da Razão de Verossimilhança

Uma forma de discriminar modelos é através de testes de hipóteses com modelos encaixados. O teste da razão de verossimilhança em modelos encaixados é um teste de hipótese muito utilizado. Esse teste geralmente considera como  $H_0$ : O modelo de interesse é adequado; e  $H_1$ : O modelo não é adequado. Para a realização deste teste, deve ser observado um modelo generalizado de tal forma que os modelos de interesse sejam casos particulares. O teste é verificado a partir de ambos os ajustes: (a) cálculo do logaritmo da função de verossimilhança do modelo generalizado ( $\ln L(\hat{\theta}_G)$ ); (b) cálculo do logaritmo da função de verossimilhança do modelo de interesse ( $\ln L(\hat{\theta}_M)$ ). Através desses resultados, é possível calcular a estatística da razão de verossimilhança:

$$TRV = -2 \log \left[ \frac{L(\hat{\theta}_M)}{L(\hat{\theta}_G)} \right] = 2 \left[ \log L(\hat{\theta}_G) - \log L(\hat{\theta}_M) \right] = \quad (2.52)$$

A estatística do teste da razão de verossimilhança tem distribuição assintótica Qui-quadrado sob a hipótese nula, com  $(k - 2)$  graus de liberdade. A hipótese  $H_0$  será rejeitada para valores grandes de  $RV$  comparado com o valor crítico obtido da distribuição  $\chi^2_{(k-2)}$  para um nível de significância nominal  $\alpha$

fixado. Em caso de não rejeição de  $H_0$ , o modelo mais parcimonioso, ou seja, o modelo com um número menor de parâmetros deve ser escolhido.

### 2.17 Razão de Risco

Ao considerar a expressão (2.43), os coeficientes de regressão ( $\beta$ ) são estimados pelo método da máxima verossimilhança parcial. Ao se fazer a divisão dos dois lados da equação por  $h_0(t)$ , obtém-se:

$$\frac{h(t|x)}{h_0(t)} = \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) \quad (2.53)$$

O quociente  $\frac{h(t|x)}{h_0(t)}$  é chamado de função das razões de riscos (ANDERSEN, 1991), (COX; OAKES, 1984) ou de índice de prognóstico. Portanto, para estimar a razão entre as funções de riscos (HR) para cada uma das covariáveis independentes ( $x_i$ ), basta calcular o exponencial do  $\beta_i$  estimado, ou seja,  $HR(x_i) = \exp(\beta_i)$ . As suposições feitas são as de que diferentes indivíduos têm funções de riscos proporcionais entre si, e que a razão entre essas funções de risco não varia no tempo.

### 2.18 Adequação do modelo de Cox

O modelo de regressão de Cox é bastante flexível devido à presença do componente não paramétrico. Ainda assim, ele não se ajusta a qualquer situação,

sendo, portanto, necessária a suposição de riscos proporcionais. A violação dessa suposição acarreta erros na estimação dos coeficientes do modelo.

### 2.19 Modelo de Cox Estratificado

Em um modelo de Cox, onde se considera uma amostra de  $n$  indivíduos que esteja dividida em  $p$  estratos e que existe um risco de base específico para cada estrato, se considerado que os efeitos das covariáveis não variam entre os estratos, o modelo de Cox específico para o indivíduo  $i$  do estrato  $j$  é dado por:

$$h_j(t | x_{ij}) = h_{0j}(t) \exp(\beta x_{ij}) \quad (2.54)$$

para  $i = 1, 2, 3, \dots, n_j$  e  $j = 1, 2, \dots, p$ , sendo que  $n_j$  é o número de indivíduos no estrato  $j$  e  $h_{0j}(t)$  é o risco basal para o estrato  $j$ . Assume-se que  $\beta$ , é comum a todos os estratos. O modelo de Cox Estratificado é útil quando a suposição de riscos proporcionais é violada, diante desse fato, uma solução é estratificar a amostra, de maneira que a suposição seja verificada em cada estrato. A estratificação não dificulta na estimação de  $\beta$ . Uma função de verossimilhança parcial como em (2.46) é construída para cada estrato e a estimação dos parâmetros é baseada na soma dos logaritmos das funções de verossimilhança parciais, isto é, em:

$$l(\beta) = [l_1(\beta) + \dots + l_m(\beta)] \quad (2.55)$$

com  $l_j(\beta) = \log(L_j(\beta))$  obtida usando-se somente os dados dos indivíduos no  $j$ -ésimo estrato.

O modelo de Cox estratificado (2.54) assume que as covariáveis atuam de modo igual na função de risco de base de cada estrato, ou seja,  $\beta$  é assumido ser comum para todos os estratos.

## 2.20 Teste de Schoenfeld para Riscos proporcionais

Uma proposta para avaliar a suposição de riscos proporcionais no modelo de Cox é uma análise gráfica dos resíduos de Schoenfeld. Grambsch e Therneau (1994) sugerem o gráfico das covariáveis versus o tempo como um método de visualizar a suposição de riscos proporcionais. Não existe violação da proporcionalidade quando a linha horizontal se mantém dentro do intervalo de confiança em todos os instantes, assim como inclinação zero mostra evidências a favor da proporcionalidade dos riscos.

A obtenção de medidas estatísticas bem como a realização de testes de hipóteses é de grande utilidade, haja vista que a técnica gráfica envolve conclusões subjetivas. O coeficiente de correlação de Pearson ( $\rho$ ) entre os resíduos de Schoenfeld e o tempo, para cada covariável, é uma dessas medidas. O teste apresenta ainda “chisq” como valor observado da estatística qui-quadrado e o respectivo *p*-valor.

## 2.21 Seleção de covariáveis

Neste trabalho, foi utilizada uma estratégia de seleção de covariáveis derivada da proposta de Collet (1994). Os passos utilizados no processo de seleção são apresentados a seguir:

Passo 1 - Ajustar todos os modelos contendo uma única covariável. Incluir todas as covariáveis que forem significativas ao nível de 0,10. Utilizar o teste da razão de verossimilhança neste passo.



Passo 2 - As covariáveis significativas no passo 1 são, então, ajustadas conjuntamente. Na presença de certas covariáveis, outras podem deixar de ser significativas. Consequentemente, ajustam-se modelos reduzidos, excluindo uma única covariável de cada vez. Verificam-se as covariáveis que provocam um aumento estatisticamente significativo na estatística da razão de verossimilhanças. Somente aquelas que atingirem a significância permanecem no modelo.



Passo 3 - Ajusta-se um novo modelo com as covariáveis retiradas no passo 2. Neste passo, as covariáveis excluídas no passo 2 retornam ao modelo para confirmar que elas não são estatisticamente significativas.



Passo 4 - As eventuais covariáveis significativas no passo 3 são incluídas ao modelo juntamente com as que foram significativas no passo 2. Neste passo, retorna-se com as covariáveis excluídas no passo 1 para confirmar que elas não são estatisticamente significativas.



Passo 5 - Ajusta-se um modelo incluindo-se as covariáveis significativas no passo 4. Neste passo é testado se alguma delas pode ser retirada do modelo.



Passo 6 - Utilizando as covariáveis que sobreviveram ao passo 5, ajusta-se o modelo final para os efeitos principais. Para completar a modelagem, deve-se verificar a possibilidade de inclusão de termos de interação dupla entre as covariáveis incluídas no modelo. O modelo final fica determinado pelos efeitos principais identificados no passo 5 e os termos de interação significativos identificados neste passo.



### 3 METODOLOGIA E MATERIAL

A metodologia deste trabalho norteia-se nos moldes da análise de sobrevivência que é uma técnica estatística, que tem como característica a presença de censuras.

#### 3.1 Metodologia

O estudo realizado é epidemiológico do tipo descritivo (estudo com somente uma amostra, sendo então não comparativo) os dados observados são de pacientes portadores do vírus da hepatite (A, B e C) e alguns destes pacientes co-infectados pelo vírus HIV.

Neste trabalho, propõe-se um modelo de regressão paramétrico e um modelo de regressão semi-paramétrico para descrever o tempo  $T$  de cura de indivíduos que possuem hepatite viral do tipo (A, B e C), considerando as fases aguda e crônica da doença, o sexo do indivíduo, como também sua idade.

Com o intuito de selecionar o modelo paramétrico (exponencial, weibull e log-normal) mais adequado ao tempo  $T$  de sobrevivência para estes dados, foi utilizado o Teste da Razão da Verossimilhança para modelos encaixados utilizando o modelo Gama-Generalizada. O teste deverá selecionar o modelo com menor valor para a razão de verossimilhanças a um nível de significância de 5%, sendo a hipótese nula  $H_0$  de que o modelo é adequado.

Outra etapa importante será a utilização da distribuição Gama-Generalizada e a estratégia de seleção de covariáveis para modelos paramétricos proposta por Collett (1994). Em cada passo do processo de seleção de

covariáveis, visto na seção (2.20), a estatística de teste é obtida utilizando-se o teste da razão de verossimilhança com uma distribuição qui-quadrado de referência com graus de liberdade iguais à diferença entre o número de parâmetros dos dois modelos a serem comparados.

O modelo paramétrico já selecionado deve receber as covariáveis também anteriormente selecionadas pela Gama-Generalizada e o teste da razão de verossimilhança para modelos encaixados deve avaliar se a introdução de uma ou mais covariáveis no modelo aumenta de modo significativo a verossimilhança de um modelo em relação a um modelo com menos parâmetros, que, sob a hipótese nula  $H_0$  de que não há diferença entre os modelos, o teste segue uma distribuição qui quadrado com graus de liberdade igual à diferença no número de covariáveis dos modelos em questão.

A adequação do modelo deverá ser avaliada pelos resíduos de Cox-Snell. Conforme visto na seção (2.13), os resíduos de Cox-Snell auxiliam na verificação do ajuste global do modelo.

Mediante a proposta de determinar um modelo semi paramétrico, deve-se verificar a seleção das covariáveis significativas utilizando a mesma estratégia de seleção proposta por Collet (1994), neste caso, utilizando o próprio modelo de regressão de Cox.

Anteriormente ao ajuste do modelo, considerando as covariáveis já selecionadas, deve-se realizar um estudo conforme descrito na seção (2.20) para verificar se estas covariáveis atendem à pressuposição de riscos proporcionais através do teste de Schoenfeld. A confirmação pode ainda ser ratificada através de uma análise gráfica dos resíduos de Schoenfeld em que, para cada covariável apresentada, teremos uma reta que, se horizontal ou próximo disso, indica proporcionalidade dos riscos para cada covariável.

O Teste da razão de verossimilhança para modelos encaixados deverá ser utilizado para avaliar o ajuste do modelo, deverá ser verificado se a inclusão

de covariáveis aumenta de modo significativo a verossimilhança de um modelo em relação a outro com menos parâmetros, que, sob a hipótese nula  $H_0$  de que não há diferença entre os modelos, o teste segue uma distribuição qui quadrado, com graus de liberdade igual à diferença no número de covariáveis dos modelos em questão.

A adequação do modelo será feita usando os resíduos de Cox-Snell da mesma forma como mencionado para o caso paramétrico.

Por fim utilizou-se a razão de riscos para interpretar os parâmetros do modelo.

Para a realização das análises e obtenção dos resultados, serão utilizados os pacotes estatísticos Survival e Flexsurv do Sistema Computacional Estatístico R (R DEVELOPMENT CORE TEAM, 2012).

### **3.2 Material**

Os dados para formulação dos modelos foram cedidos pela Secretaria Municipal de Saúde da cidade de Lavras-MG e correspondem a 96 indivíduos infectados pelo vírus da hepatite no período de 2002 a 2011. As variáveis utilizadas no estudo são as observadas nos prontuários de cada indivíduo acompanhado. Essas variáveis são classificadas como: dependentes (tempo até a ocorrência do evento de interesse) ou independentes (sexo, idade, tipos de hepatite (A, B, C) e fases da hepatite (aguda ou crônica)), de cada paciente.

É interessante salientar que, ao catalogar os dados e entrevistar o profissional de saúde responsável, algumas informações pertinentes e curiosas foram observadas, como:

- a) Um grande número de abandonos ao tratamento da doença, devido às dificuldades encontradas por alguns pacientes no tocante aos vícios de drogas ilícitas, injetáveis, inaláveis etc. bem como o uso de bebidas alcólicas no período de tratamento em que se recomenda abolir o uso dessas substâncias.
- b) O desconforto de alguns pacientes em admitir a doença publicamente e mesmo no contexto familiar, já que hepatite B é considerada uma doença sexualmente transmissível e, portanto, para um casal, sugere que alguma das partes possa ter cometido adultério.

Diante os 96 dados observados pertinentes a indivíduos com hepatite, foram incluídas covariáveis no estudo, as quais foram admitidas preliminarmente de acordo com as opiniões dos profissionais da área da saúde que acompanham os pacientes, conforme se vê na Tabela 2.

Tabela 2 Variáveis consideradas preliminarmente

<b>Variáveis</b>	<b>Descrição</b>	<b>Classificação</b>
Tipo	Hepatite A, B, C	Categórica
Fase	0=Aguda e 1=Crônica	Categórica
Sexo	Masculino ou Feminino	Categórica
Idade	Idade com que iniciou o tratamento	Contínua
Censura	0 = Censura e 1 = Falha	Categórica
Tempo	Meses em que os pacientes foram acompanhados	Continua

## 4 RESULTADOS E DISCUSSÃO

A partir dos dados de indivíduos da cidade de Lavras-MG, infectados por vírus de hepatites, foi utilizada uma abordagem estatística nos moldes da análise de sobrevivência para aferir a importância de cada covariável envolvida no tempo de cura da hepatite.

### 4.1 Resultados descritivos

A fim de um melhor entendimento dos dados, é necessária uma análise descritiva dos mesmos. Nas tabelas abaixo, os dados serão abordados de maneira que se possa melhor compreendê-los.

Na Tabela 3, é possível observar que o número de pacientes com hepatite na fase crônica é praticamente duas vezes o número de pacientes com hepatite na fase aguda.

Tabela 3 Fases da Hepatite – Lavras-MG - 2002 a 2011

Fases	Frequência	Percentual
Aguda	35	36,46
Crônica	61	63,54
Total	96	100

Fonte: Secretária Municipal de Saúde

Na Tabela 4, pode-se verificar que o predomínio de infectados por hepatite incide sobre o grupo de indivíduos do sexo masculino.

Tabela 4 Sexo dos pacientes com hepatite – Lavras-MG - 2002 a 2011

Sexo	Frequência	Percentual
Masculino	64	66,67
Feminino	32	33,33
Total	96	100

Fonte: Secretária Municipal de Saúde

Na Tabela 5, a incidência do tipo de hepatite C supera largamente tanto o tipo de hepatite A, quanto o tipo de hepatite B.

Tabela 5 Tipos de hepatite – Lavras-MG - 2002 a 2011

Tipos	Frequência	Percentual
A	20	20,83
B	12	12,50
C	64	67,67
Total	96	100

Fonte: Secretária Municipal de Saúde

Na Tabela 6, é possível verificar que, ao longo do acompanhamento dos pacientes, apenas 28 conseguiram a cura da doença.

Tabela 6 Censura – Lavras-MG - 2002 a 2011

Censuras	Frequência	Percentual
Curados	28	29,17
Abandonos ao tratamento	18	18,75
Mortes	11	11,46
Doentes	39	40,62
Total	96	100

Fonte: Secretária Municipal de Saúde

## 4.2 Ajuste do modelo paramétrico

O Teste da razão de verossimilhança utilizado através da distribuição Gama-Generalizada, segundo a Tabela 7, rejeitou a hipótese nula de adequação de cada um dos modelos candidatos a um nível de significância de 5%.

Tabela 7 Escolha de uma distribuição de probabilidade através da distribuição gama generalizada

Distribuições	$-\log(L(\theta))$	TRV	P-valor
Gama Generalizada	216,53		
Exponencial	246,80	60,54	7,31e-14
Weibull	219,30	5,54	1,93e-2
Log-normal	220,45	7,84	5,07e-3

Haja vista que, dentre os modelos paramétricos propostos na Tabela 7, nenhum deles foi adequado ao tempo  $T$  de sobrevivência para os dados deste trabalho. A seguir propõem-se o modelo semi paramétrico de Cox e estuda-se a adequação desse modelo.

## 4.3 Ajuste do modelo semi paramétrico

Foi aplicado o método de seleção de covariáveis proposto por Collet (1994), visto na seção (2.21), utilizando o modelo de regressão de Cox, cujos resultados encontram-se descritos na Tabela 8. Nota-se que as covariáveis selecionadas foram as confirmadas no 5º Passo. Assim, observou-se que as covariáveis significativas foram fases da hepatite, tipos de hepatite e sexo dos

indivíduos. A covariável idade não apresenta significância no tempo até a cura dos indivíduos, ausentando-se, portanto, no modelo que será proposto.

Tabela 8 Seleção de covariáveis usando o modelo de regressão de Cox

Modelos	$-2\text{Log}L(\theta)$	Estatística de teste (TRV)	$p - \text{valor}$
1º Passo			
Surv~1	240,44		
Surv~Tipo	119,11	121,33	4,50e-27
Surv~Fase	172,41	68,03	1,61e-16
Surv~Sexo	238,52	1,92	0,17
Surv~Idade	199,45	40,99	1,53e-10
2º Passo			
Tipo+Fase+Idade	109,94		
Tipo+Idade	119,38	9,44	2,13e-3
Tipo+Fase	110,07	0,13	0,72
Fase+Idade	152,62	42,68	5,41e-10
3º Passo			
Fase+Tipo	110,07		
Fase+Tipo+Idade	109,94	0,13	0,72
4º Passo			
Tipo+Fase	110,07		
Tipo+Fase+Sexo	104,67	5,4	2,01e-2
5º Passo			
Fase+Tipo+Sexo	104,67		
Fase+Tipo	110,07	5,4	2,01e-20
Fase+Sexo	172,31	67,64	2,05e-15
Tipo+Sexo	113,85	9,18	2,45e-3
6º Passo			
Fase+Tipo+Sexo	104,67		
Fase+Tipo+Sexo+Fase*Tipo	104,72	0,05	0,99
Fase+Tipo+Sexo+Tipo*Sexo	103,96	0,71	0,87
Fase+Tipo+Sexo+Fase*Sexo	104,17	0,5	0,78
Etapa Final			
Modelo Fase+Tipo+Sexo			



Na Tabela 9, apresentam-se os resultados do teste de Schoenfeld para testar a suposição de proporcionalidade dos riscos. Pode-se verificar que o teste global e o teste para cada covariável apresentaram evidências favoráveis à não rejeição da hipótese nula de riscos proporcionais.

Tabela 9 Teste de proporcionalidade dos riscos

Covariável	$\rho$	chisq	$p$ -valor
TipoB	-0,187	4,64e-05	0,995
TipoC	-0,557	8,16e-04	0,977
Sexo Feminino	-0,146	5,12e-01	0,474
Fase Crônica	-0,272	5,08e-04	0,982
Global	NA	5,17e-01	0,972

Os gráficos dos resíduos de Schoenfeld são apresentados na Figura 4. Pode-se observar que, para cada uma das covariáveis, obteve-se uma reta praticamente horizontal, o que caracteriza que a covariável em questão apresenta riscos proporcionais ao longo do tempo. Mesmo a covariável sexo, onde a reta não é exatamente horizontal, a pouca variação apresentada não é suficiente para descaracterizar a condição de riscos proporcionais para essa covariável.

Observando a Figura 5, pode-se concluir que o modelo de Cox é adequado para se ajustar aos dados de tempo até a cura da hepatite, considerando as covariáveis tipos de hepatite (A, B e C), fase da hepatite (aguda e crônica) e sexo do indivíduo.

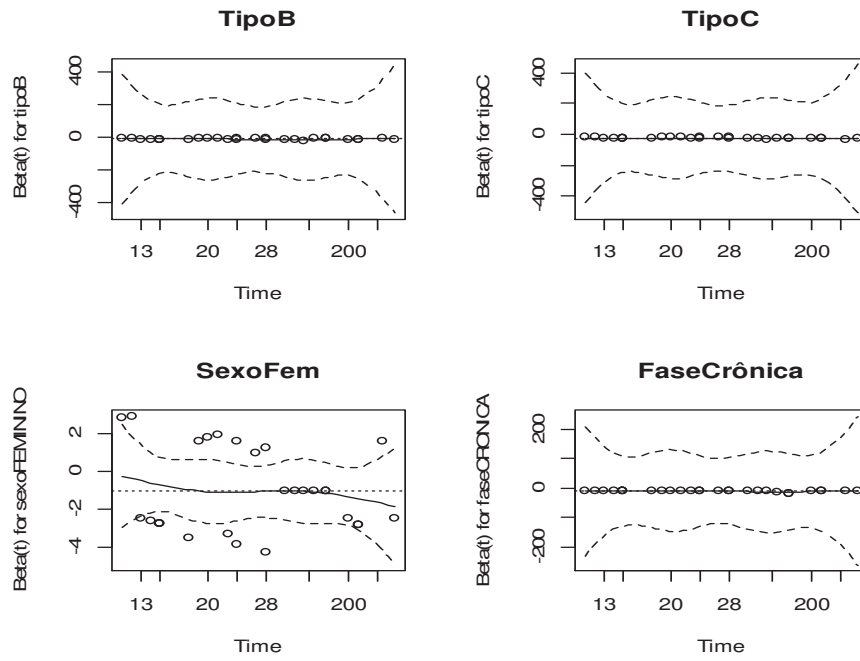


Figura 4 Gráfico dos resíduos de Schoenfeld

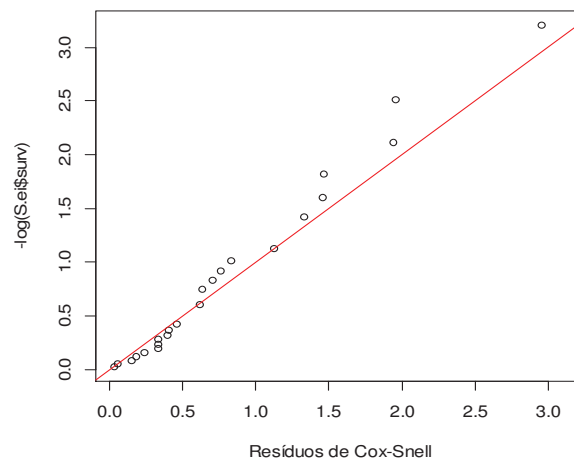


Figura 5 Gráfico dos resíduos de Cox-Snell

O modelo de Cox, conclusivo, ratificado pelos processos anteriores, é determinado pela equação (4.1) e será objeto de estudo para determinar índices de interesse relativo ao tempo até a cura de indivíduos com hepatites A, B e C na cidade de Lavras-MG.

$$M_{Cox} = h(t | \mathbf{x}) = h_0(t) \exp\{\beta_{10}x_{10} + \beta_{11}x_{11} + \beta_{12}x_{12} + \beta_{20}x_{20} + \beta_{21}x_{21} + \beta_{30}x_{30} + \beta_{31}x_{31}\} \quad (4.1)$$

A Tabela 10 apresenta as estimativas dos parâmetros à razão de riscos com os respectivos intervalos de confiança desses riscos, considerando o modelo apresentado pela equação (4.1).

Tabela 10 Estimativas dos parâmetros referentes ao modelo de Cox ajustado aos dados de hepatites

Covariável	Estimativa	Erro Padrão	Razão de Riscos (RR)	$IC_{95\%}(RR)$
TipoB	-1,95	0,80	0,14	(0,03 ; 0,68)
TipoC	-5,16	1,26	0,58e-2	(0,49e-3 ; 0,07)
SexoFeminino	-0,30	0,41	0,74	(0,33 ; 1,67)
Fase Crônica	-0,23	0,85	0,80	(0,15 ; 0,42e + 1)

Fica, portanto, o modelo assim determinado com as covariáveis tipo, sexo e fase cada covariável com seus respectivos níveis.

$$M_{Cox} = h(t | x) = h_0(t) \exp\{-1,95x_{11} - 5,16x_{12} - 0,30x_{21} - 0,23x_{31}\} \quad (4.2)$$

Um paciente com hepatite do tipo B tem risco 0,14 vezes, ou aproximadamente (86%), menor de curar-se da hepatite do que um paciente com hepatite A.

Um paciente com hepatite do tipo C tem risco 0,006 vezes, ou aproximadamente (99,4%), menor de curar-se de hepatite do que um paciente com hepatite A.

Um paciente do sexo feminino tem risco 0,74 vezes, ou aproximadamente (26%), menor de curar-se de hepatite do que um paciente do sexo masculino.

Um paciente com hepatite na fase crônica tem risco 0,80 vezes, ou aproximadamente (20%), menor de curar-se de hepatite do que um paciente com hepatite na fase aguda.

Um paciente com hepatite do tipo C tem risco 0,03 vezes, ou aproximadamente (97%), menor de curar-se de hepatite do que um paciente com hepatite do tipo B.

A Tabela 11 apresenta as estimativas dos parâmetros do modelo, considerando as quatro covariáveis iniciais. A covariável idade, excluída pelo método de Collet foi introduzida neste modelo com o objetivo de contribuir na esfera clinica.

Tabela 11 Estimativas dos parâmetros referentes ao modelo de Cox ajustado aos dados de hepatites considerando a covariável idade

Covariável	Estimativa	Erro Padrão	Razão de Riscos (RR)	$IC_{95\%}(RR)$
TipoB	-1,99	1,02	0,14	(0,02 ; 0,99)
TipoC	-5,20	1,40	0,01	(0,00 ; 0,08)
SexoFem.	-0,29	0,41	0,74	(0,33 ; 1,68)
Fase Crônica	-0,21	0,87	0,81	(0,15 ; 4,46)
Idade	0,001	0,01	1,00	(0,98 ; 1,03)

$$M_{Cox} = h(t | x) = h_0(t) \exp\{\beta_{10}x_{10} + \beta_{11}x_{11} + \beta_{12}x_{12} + \beta_{20}x_{20} + \beta_{21}x_{21} + \beta_{30}x_{30} + \beta_{31}x_{31} + \beta_{40}x_{40}\} \quad (4.3)$$

Fica então o modelo com as covariáveis tipo, sexo, fase e idade assim determinado:

$$M_{Cox} = h(t|x) = h_0(t) \exp\{-1,99x_{11} - 5,20x_{12} - 0,29x_{21} - 0,21x_{31} + 0,001x_{40}\} \quad (4.4)$$

#### 4.4 Discussão

Um modelo de regressão paramétrico, proposto, com o objetivo de aferir o tempo até a cura dos portadores das hepatites virais (A, B e C) da cidade de Lavras-MG, considerando as covariáveis tipo e fase das hepatites, sexo e idade dos indivíduos infectados, não foi adequado ao tempo  $T$  de sobrevivência para os dados deste trabalho, segundo o teste da razão de verossimilhança, a partir da distribuição gama generalizada.

Um modelo de regressão semi paramétrico, proposto, com o objetivo de verificar a influência sobre a cura das hepatites considerando os níveis de cada covariável foi ratificado e os critérios de seleção de covariáveis adotados selecionaram as covariáveis tipos de hepatites, fases da doença e sexo como as que estatisticamente são significativas (importantes) no tratamento das hepatites.

Uma forma de sistematizar as informações referentes aos vírus das hepatites considerando perfis epidemiológicos heterogêneos com representatividade é o maior desafio da vigilância epidemiológica (GAZE; CARVALHO; WERNECK, 2002).

A distribuição dos casos confirmados de hepatites virais no Brasil, no período entre 1996 e 2000, foi de 43% para VHA, 25% para VHB, 12% para VHC, 1% para hepatite não-A, não-B, não-C, 1% para outros tipos de hepatite, e 17% foi constituído de casos ignorados, provavelmente pela falta de comprovação sorológica (LADEHOF; BUENO, 2005).

Em um estudo compreendido entre 1996 a 2002, Chávez, Campana e Haas (2003) constataram que no Brasil houve um maior número de indivíduos infectados por VHA com faixa etária de 5 a 9 anos. Para o VHB e VHC a faixa etária incide para indivíduos acima de 30 anos. Devido ao comportamento e estilo de vida que oferecem maiores risco de infecção, a incidência pelo VHB está entre os indivíduos com idade acima dos 15 anos.

Ferreira e Silveira (2004) alegaram que o vírus da hepatite A é a causa mais comum de hepatite viral aguda no mundo. Em termos de Brasil, verificam-se 130 novos casos por cada 100.000 habitantes. Segundo os autores a incidência dos casos acomete na maior parte as populações carentes que não possuem saneamento básico satisfatório.

O Ministério da Saúde estima que, no Brasil, pelo menos 15% da população já esteve em contato com vírus da hepatite B. O VHB está muito presente no sangue e em fluidos orgânicos (FERREIRA; SILVEIRA, 2004).

Brasil et al. (2003), em um estudo de prevalência para hepatite B, constataram, em relação ao sexo, que o sexo masculino foi mais predominante entre todos os indivíduos infectados pelo VHB, possivelmente devido a uma maior exposição do homem aos riscos de infecção, seja pelo maior número de parceiros sexuais, homossexualismo, maiores riscos de acidentes com necessidade de transfusão de sangue etc.

Todas as pessoas estão sujeitas a contrair o VHB, no entanto, devido a atitudes comportamentais ou profissionais alguns grupos de indivíduos são particularmente expostos a esses vírus (FERNANDES et al., 1999).

As principais vias de transmissão da hepatite B são parenteral, sexual e vertical, com predomínio em indivíduos de 20 a 40 anos de idade. O Ministério da Saúde tem participado de forma atuante através de triagem sorológica em bancos de sangue e imunização passiva, que tem contribuído com uma diminuição na incidência de hepatite B (LADEHOF; BUENO, 2005).

O uso de drogas ilícitas injetáveis é o principal fator de risco da hepatite C. Assim, 50% a 80% dos usuários se infectam no primeiro ano e 100% até o oitavo ano de uso de drogas endovenosas. A frequência de transmissão vertical não supera 3% dos casos (RESENDE et al., 2010).

Cerca de 70 a 85% dos pacientes com VHC evoluem para infecção persistente e assintomática, desenvolvendo em longo prazo para hepatite crônica ativa com progressão para cirrose hepática. Um maior controle no sangue que será disponibilizado para transfusão é essencial, já que, a transmissão parenteral foi apontada como o principal meio de disseminação viral (LADEHOF; BUENO, 2005).

## 5 CONCLUSÃO

Um modelo de regressão semi paramétrico, proposto, e os critérios de seleção de covariáveis adotados selecionaram as covariáveis tipos de hepatites, fases da doença e sexo como as que estatisticamente são significativas no tratamento das hepatites. Este modelo indica que a covariável tipos de hepatite é a mais influente no tempo de cura das hepatites.

A cidade de Lavras-MG, diferentemente de índices apresentados por Ladehof e Bueno (2005), apresentou um número de indivíduos com VHC três vezes o número dos indivíduos com VHA. Informação relevante, já que, segundo esses autores, cerca de 70 a 85% dos pacientes com VHC têm caso de evolução para infecção persistente e quadro que evoluem para cirrose hepática.

Os indivíduos do sexo masculino na cidade de Lavras-MG, igualmente ao estudo de Brasil et al. (2003), apresentou número de infecção por hepatites superior aos indivíduos do sexo feminino.

Um percentual expressivo de infectados em Lavras-MG, aproximadamente 19%, abandonaram o tratamento, índice muito elevado comparado ao estudo apresentado por Garcia et al.(2012), realizado no Hospital Estadual Mário Covas (HEMC) em São Paulo, que apresentou percentual de abandono ao tratamento entre 2 a 5%.

Entre os onze casos de mortes notificados em Lavras-MG, dez foram de indivíduos com VHC e um com VHB, sendo que 70% das mortes por VHC foi por cirrose hepática, que é uma doença terminal do fígado.

Na cidade de Lavras-MG, entre os indivíduos infectados por VHB ou VHC, todos apresentaram, no início do estudo, idade superior aos vinte anos. Uma nova tendência em relação à idade de infecção por VHB é apontada, já que os jovens têm iniciação sexual antecipada ano a ano e, sendo a hepatite B uma



doença sexualmente transmissível, estudos apontam para infecção por VHB ainda na adolescência.

## REFERÊNCIAS

ABPH-Associação Brasileira dos Portadores de Hepatite; **Esteatose Hepática**. São Paulo, 2013. Disponível em: < <http://www.hepatite.org.br/hepatite/esteatose> > Acesso em: 08/06/2013.

ALLISON, P.D. **Survival Analysis Using SAS A Practical Guide**, 2ª ed.USA, SAS Institute Inc., 1995.

ANDERSEN, P.K. **Survival analysis The second decade of the proportional hazards regression model**. Statistics in Medicine, 1991.

ARRAES, L. C. et al. Prevalência de hepatite B em parturientes e perfil sorológico perinatal. **Revista brasileira Ginecologia Obstetria**, Rio de Janeiro v.25, n. 8, p.571-576, julho, 2003.

BAPTISTA, G.H. et al. Fatores associados à duração do aleitamento materno em crianças de famílias de baixa renda da região sul da cidade de Curitiba. **Revista Brasileira de Saúde Pública**, Rio de Janeiro, v.25, n.3, p.596-604, 2009.

BASTOS, J. ; ROCHA, C. Serviço de Higiene e Epidemiologia da Faculdade de Medicina do Porto. **Revista Arquimed**, Lisboa, vol.20, nº5/6, p.185-187, 2006.

BRASIL, L.M. et al. Prevalência de marcadores para o vírus da hepatite B em contatos domiciliares no Estado do Amazonas. **Revista da Sociedade Brasileira de Medicina Tropical**, Uberaba, V.36, n.5, p.565-570, setembro, 2003.

BRASIL. Ministério da Saúde. Departamento de DST, **Aids e Hepatites Virais**. **O que são hepatites?** Portal sobre aids, doenças sexualmente transmissíveis e hepatites virais. Brasília, s.d.a. Disponível em

<<http://www.aids.gov.br/pagina/o-que-sao-hepatites-virais>>, acesso em 3 out. 2013.

BRASIL. Ministério da Saúde. Departamento de DST, **Aids e Hepatites Virais. Vacina?** Portal sobre aids, doenças sexualmente transmissíveis e hepatites virais. Brasília, s.d.b. Disponível em <<http://www.aids.gov.br/pagina/vacina-hepatites>>, acesso em 2 jun 2013.

BRASIL. Ministério da Saúde. Departamento de DST, **Aids e Hepatites Virais. Hepatite C.** Portal sobre aids, doenças sexualmente transmissíveis e hepatites virais. Brasília, s.d.c. Disponível em <<http://www.aids.gov.br/pagina/hepatite-c>>, acesso em 2 jun 2013.

BRASIL. Ministério da Saúde. Programa Nacional de Hepatites Virais. **Avaliação da Assistência as Hepatites Virais no Brasil**, Brasília, v.1, p.61,2002.

BRESLOW, N.E. Contribution to the discussion of D.R.Cox. **Journal Royal Statistical Society. London.**, v.34, p.216-217, 1972.

CANTOR, A. B. **Survival Analysis Techniques for Medical Research**, 2.ed. SAS Institute Inc., USA, 2003.

CARTER, M. Grupo Português de Activistas sobre Tratamentos de **VIH/SIDA**, 2010 Disponível em: <<http://www.aidsmap.com/O-VIH-e-as-hepatites/page/1445944/>>.\_Acesso em: 02/06/2013.

CARVALHO, M. S. et al. **Teoria e aplicações em saúde**,2ª ed. Rio de Janeiro,Fiocruz,2011.

CHÁVEZ, J.H.; CAMPANA, S.G.; HAAS, P. **Revista Panamericana de Salud Pública**, São Paulo, v.14, n.2, p.91-96, 2003.

CIORLIA, L.A.S.; ZANETTA, D.M.T. Hepatite B em profissionais de saúde: prevalência, a vacinação e sua relação com fatores ocupacionais. **Revista Brasileira de Doenças Infecciosas**, Salvador, v.9, n. 5, p. 384-389, jan.2007.

CLEMENS, S.A. et al. Soro prevalência para hepatite A e B em quatro centros do Brasil. **Revista da Sociedade Brasileira de Medicina Tropical**, São Paulo, SP, v.33, n.1, p.1-10, fev. 2000.

COLLETT, D. **Modelling Survival Data in Medical Research**. Chapman and Hall, New York, 1994.

COLOSIMO, E. A; GIOLO, S. R. **Análise de Sobrevida Aplicada**, 1ª ed. São Paulo, Blucher, 2006.

COX, D.R.; HINKLEY, D.V. **Theoretical Statistics**. Chapman and Hall, London, 1975.

COX, D. R.; OAKES, D. **Analysis of Survival Data** .London, Chapman and Hall,1984.

COX, D.R.; SNELL, E . J. A General Definition of Residuals. **Journal of the Royal Statistical Society B**, 30, 248-275, 1968.

CRUZ, F. M. M.; BONETTO, D.; CARNEIRO, R. M. Hepatite viral aguda: um enfoque para uma enfermidade antiga. **Revista adolescência latino-americana**, Porto Alegre, v.2, p.16-22, 2000.

FERNANDES, J. V. et al. Prevalência de marcadores sorológicos do vírus da hepatite B em trabalhadores do serviço hospitalar. **Revista Saúde Pública**, São Paulo, v.33, n.2, p. 122-128, abril. 1999.

FERREIRA, C.T.; SILVEIRA, T.R. Hepatites virais: aspectos da epidemiologia e da prevenção, **Revista Brasileira de Epidemiologia**, v.7, n.4, p.473-487,2004.

FERREIRA, M. S. Diagnóstico e tratamento da hepatite B, **Revista sociedade Brasileira de Medicina Tropical**, Brasil, v.33, n.4, p.389-400, julho. 2000.

FUKS, A. G.; Medicina Diagnóstica Lavoisier, **Hepatite**, São Paulo, 2012. Disponível em: < <http://www.lavoisier.com.br/clientes/artigo/hepatite>>. Acesso em: 04/06/2013.

GARCIA, T. J. et al. Efeitos colaterais do tratamento da hepatite C no polo aplicador do ABC. **Revista Associação Médica Brasileira**, São Paulo, v.58, n.5, p.543-549, maio, 2012.

GAZE, R.; CARVALHO, D. M.; W, G.L. Soroprevalência das infecções pelos vírus das hepatites A e B em Macaé, Rio de Janeiro, Brasil. **Cad. Saúde Pública**, Rio de Janeiro, v.18, n.5, p. 1251-1259, setembro 2002.

GRAMBSCH, P.M.; THERNEAU, T.M. **Proportional Hazards Tests and Diagnostics based on Weighted Residuals**, v.81, p. 515-526, 1994.

JORGE, S.G. **Hepatite B**, São Paulo, 2005. Disponível em: < [http://www.hepcentro.com.br/hepatite\\_b.htm](http://www.hepcentro.com.br/hepatite_b.htm)> Acesso em: 15/09/2011.

JORGE, S.G. **Histologia hepática**, São Paulo, 2007. Disponível em: <<http://www.hepcentro.com.br/histologia.htm>> Acesso em: 15/09/2011.

JUNIOR, P.L.; SILVEIRA, F.L.; OSTERMANN, F. Análise de sobrevivência aplicada ao estudo do fluxo escolar nos cursos de graduação em física: um exemplo de uma universidade brasileira, **Revista Brasileira de Ensino de Física**, v. 34, n. 1, p.1403-1410, 2012.

KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. **Journal of the American Statistical Association**, v.53, p.457-81, 1958.

LADEHOF, M.L.; BUENO, E.C. Incidência de Hepatites Virais em Blumenau-SC, Brasil. **Acta Farmaceutica Bonaerense: Latin american journal of pharmacy**, Buenos Aires, v.24, n.2, p.436-440,2005.

LAWLESS, J. F. **Statistical Models and Methods for Lifetime Data**. John Wiley and Sons, New York, 1982.

LIRA, S. A.; SAMPAIO, A. V. Duração do desemprego na região metropolitana de Curitiba. **Revista Paranaense de Desenvolvimento**, Curitiba, v.109, p.139-160, 2005. Curitiba.

LONGHI, M.; CORDEIRO, V.; RAMALHAES, M.A. Prevalência do marcador sorológico anti-HBC da hepatite B em doadores de sangue, segundo banco de dados da homorrede de Rondônia. **Revista Arte e Ciência**, São Paulo, v.33, n. 2, p.1, 2010.

MENESIA, E.O. et al. Sobre valência de Pacientes com AIDS em uma cidade do sudeste brasileiro. **Rev. Pan. Saúde Pública**, v.10, n.1, p.29-36, 2001.

MINCIS, M.; MINCIS, R. Doença hepática alcoólica. **Revista Brasileira de Medicina**, São Paulo, v. 67, p.21-31, 2010.

NAKANO, E.Y.; CARRASCO, C.G. Uma avaliação do uso de uma modelo contínuo na análise de dados discretos de sobrevivência, **Revista Sociedade Brasileira de Matemática Aplicada e Computacional**, v.7, n.1, p.91-100, 2006.

OLIVEIRA, U.B. Transmissão Perinatal do Vírus da Hepatite C. In: **Sociedade Brasileira de Infectologia para o Manuseio e Terapia da Hepatite C**, São Paulo, p.17-19, 2008. Disponível em:<[http://www.infectologia.org.br/anexos/I%20Consenso%20para%20hepatite%20C\\_em%20portugu%C3%AAs.pdf](http://www.infectologia.org.br/anexos/I%20Consenso%20para%20hepatite%20C_em%20portugu%C3%AAs.pdf)> Acesso em: 05/06/2013.

OMS-Organização Mundial de Saúde; **Hepatite B**, Genebra, Disponível em: <  
<http://www.who.int/csr/disease/hepatitis/whocdscsrlyo20022/en/index.html>>  
Acesso em: 05/06/2013.

PETO, R. Contribuiu a Discussão do Artigo de D.R. Cox. **Journal of the Royal Statistical Society B**, v.34, p. 205-207, 1972.

PORTA, G. Hepatite auto-imune. **Jornal de Pediatria**, São Paulo, v.76, p.S181-S185 supl.2. 2000.

PRADO, K.D. Transmissão Sexual do VHC, **Sociedade Brasileira de Infectologia para o Manuseio e Terapia da Hepatite C**, São Paulo, p.14-16, (2008). Disponível em: <  
<http://www.infectologia.org.br/publicoleigo/default.asp>> Acesso em: 16/01/2013.

PRENTICE, R. L.; GLOECKLER, L.A. Regression analysis of grouped survival data with application to breast cancer data. **Biometrics**, Washington, DC, v.34, p.57-67, 1978.

R DEVELOPMENT CORE TEAM. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Viena, Áustria, 2012.

RESENDE, V.L.S.; ABREU, M.H.N.G.; TEIXEIRA, R; PORDEUS, I.A. Hepatites Virais na Prática Odontológica: Riscos e Prevenção. **Revista Pesquisa Brasileira em Odontopediatria e Clínica Integrada**, João Pessoa, v.10, n.2, p.317-323, 2010.

ROSENTHAL, C. A epidemia oculta. **Jornal do Cremesp**, Paraíba, edição 236, p. 4, 2007.

SANTOS, A.M. et al. Usando Redes Neurais Artificiais e Regressão Logística na Predição da Hepatite A. **Revista Brasileira de Epidemiologia**, Brasil, v.8, n.2 p. 117-126, 2005.

SASSO, K.D. et al. Transplante de Fígado : Resultados de aprendizagem de pacientes que aguardam a cirurgia. **Revista Latino americano de Enfermagem**, São Paulo, v.13, n.4, p.481-488, 2005.

SILVEIRA, F.J.F.; LAMOUNIER, J.A. Fatores associados à duração do aleitamento materno em três municípios na região do Alto Jequitinhonha, Minas Gerais. **Revista Saúde Pública**, Rio de Janeiro, v. 22, n.1, p.69-77, 2007.

STACY, E.W. A Generalization of the Gamma Distribution. **Ann. Math. Stat.**, 33, 1187-1192, 1962.

STRAUSS, E. Hepatite C. **Revista da Sociedade Brasileira de Medicina Tropical**, Uberaba, v.34, n.1, p. 69-82, 2000.

TIETZE, C. Fertility after discontinuation of intrauterine and oral contraception, **International Journal of Fertility**, n.31, p.385-389, 1968.

VILLAR, L.M.; DE PAULA, V.S.; GASPAR, A.M. C. Variação sazonal da infecção pelo vírus da hepatite A na cidade do Rio de Janeiro, **Revista do Instituto de Medicina Tropical de São Paulo**, v. 44, n. 5, p. 289-292, 2002.

WEIBULL, W. A Statistical Representation of Fatigue Failure in Solids. **Royal Institute of Technology**, 1954.

WEIBULL, W. A Statistical Theory of the Strength of Materials. **Ingeniors Vetenskaps Akademien Handlingar Stockholm**, n.151, p.293-297, 1939.

WEIBULL, W. A Statistical Distribution of Wide Applicability. **Journal of Applied Mechanics**, v.18, p. 293-297, 1951.