

## Ajuste de modelo de análise de sobrevivência para dados de tuberculose do município de Patos de Minas - Minas Gerais

Bruno Henrique Gomes<sup>1†</sup>, Fabrícia M. Oliveira<sup>2</sup>

<sup>1</sup>Graduando em Biotecnologia, Instituto de Genética e Bioquímica-INGEB-UFU.

<sup>2</sup>Professora da Faculdade de Matemática-FAMAT-UFU.

**Resumo:** *A tuberculose tem sido uma das doenças mais estudadas nos seus aspectos biológico, epidemiológico, diagnóstico, terapêutico e profilático. No entanto essa doença teima em persistir, em nível mundial, causando sofrimento humano e morte em pleno século XXI. Com isso há a necessidade de estudos epidemiológicos associados a técnicas estatísticas, como, a Análise de Sobrevivência, que têm como objetivo estudar dados de experimentos em que a variável resposta é, geralmente, o tempo até a ocorrência de um evento de interesse, sendo muito aplicada na área médica. Neste trabalho foi utilizado o modelo paramétrico AFT para modelar o tempo até a cura em função das covariáveis consideradas, determinando a distribuição que melhor se ajusta aos dados, para analisar o efeito de covariáveis no tempo até a ocorrência de um dos eventos, morte por tuberculose ou cura em 437 pacientes detectados com tuberculose atendidos pela Secretaria Regional de Saúde, da cidade de Patos de Minas - MG, no período de 2003 a 2012. Concluiu-se, em relação à covariável sexo, que um paciente do sexo feminino tem 109,38% a mais na probabilidade de cura que pacientes do sexo masculino. Os resultados sugerem ainda pacientes do município de Patos de Minas tem 259,18% a mais na probabilidade de cura que pacientes de outras regiões.*

**Palavras-chave:** Modelo paramétrico, Covariáveis, Probabilidade.

**Abstract:** *Tuberculosis has been one of the most studied their biological epidemiology, diagnosis, prophylactic and therapeutic aspects diseases. However this disease stubbornly persists worldwide, causing human suffering and death in the 21st century. Thus there is a need for epidemiological studies associated with statistical techniques such as Analysis of Survival have to study data from experiments in which the response variable is usually the time to occurrence of an event of interest being applied extensively in the medical field . In this paper the AFT parametric model was used to model the time until healing function of the covariates considered in determining the distribution that best fits the data , to analyze the effect of covariates on time to occurrence of an event , death from tuberculosis or healing in 437 patients detected with tuberculosis treated by the Regional Department of Health - Patos de Minas - MG , in the period 2003 - 2012. We conclude that in relation to the covariate sex, a female patient is 109.38 % more likely to cure that in masculine patients . The results also suggest patients from Patos de Minas city is 259.18 % more in the probability of cure than patients from other regions.*

**Keywords:** Parametric model, Covariates, Probability.

---

†Autor correspondente: [b.hgomes@hotmail.com](mailto:b.hgomes@hotmail.com).

## Introdução

A tuberculose é uma doença contagiosa relacionados com a pobreza, a desnutrição e má função imunológica. A Morbidade e mortalidade da tuberculose são mais altas nos países de renda baixa e média (WHO, 2013) e tem sido uma das doenças mais estudadas nos seus aspectos biológico, epidemiológico, diagnóstico, terapêutico e profilático. No entanto essa doença teima em persistir, em nível mundial, causando sofrimento humano e morte em pleno século XXI (HIJJAR; OLIVEIRA; TEIXEIRA, 2001).

Causada pelo *Mycobacterium tuberculosis*, pode acometer uma serie de órgãos e/ou sistemas, é transmitida por via aérea em praticamente todos os casos (WHO, 2012), no entanto seu alvo principal é o pulmão, uma vez que o agente etiológico se reproduz e desenvolve rapidamente em áreas do corpo muito oxigenadas (SOUZA, 2009). A infecção ocorre a partir da inalação de núcleos secos de partículas contendo bacilos expelidos pela tosse, fala ou espirro do doente com tuberculose ativa de vias respiratórias (pulmonar ou laríngea). Os doentes bacilíferos, isto é, aqueles cuja baciloscopia de escarro é positiva, são a principal fonte de infecção (WHO, 2012).

Segundo Souza (2009) acreditava-se que os medicamentos utilizados para o tratamento da tuberculose erradicariam a doença entre as décadas de 70 e 80. No entanto, o que se observou foi o reaparecimento no início da década de 80 como um sério problema de saúde mundial. Vários foram os fatores responsáveis pelo ressurgimento da tuberculose no mundo, dentre eles podemos destacar a não adesão dos pacientes ao tratamento, que contribui para o aparecimento de bactérias resistentes aos fármacos utilizados. No Brasil a taxa de abandono é alta e de acordo com Chirinos e Meirelles (2011) situa-se em 17%, porém, em muitas regiões, atinge níveis mais elevados, chegando a 20 % na grande São Paulo. Isso leva ao não rompimento da cadeia de transmissão, pois as pessoas com tuberculose que não aderem à terapêutica continuam doentes e permanecem como fonte de contágio. Além disso, a pobreza, os aglomerados populacionais, a falta de informação, e os movimentos migratórios contribuem para seu reaparecimento (SOUZA, 2009).

Segundo a World Health Organization - WHO, a taxa de incidência da tuberculose vem diminuindo desde 2002 em 1,3% ao ano. A estimativa de novos casos de tuberculose no mundo é de 8,8 milhões em 2010, equivalente a 128/100.000 habitantes. Desse total, 59%, 26%, 7%, 5% e 3%, estão na Ásia, África, região do Mediterrâneo, Europa e Américas, respectivamente. Entre 1 e 1,2 milhões de casos são de pessoas vivendo com HIV/AIDS. A África é responsável por 82% dos casos de coinfeção tuberculose/HIV no mundo. 22 países são responsáveis por 81% de todos os casos de tuberculose no mundo, sendo os cinco primeiros a Índia (um quarto dos casos), China, África do Sul, Indonésia e Paquistão, enquanto que o Brasil ocupa o 19º lugar. A World Health Organization (2013) declarou a tuberculose como emergência sanitária mundial em 1993 e em 2000, a Organização das Nações Unidas estabeleceu as Metas do Milênio, que incluíram o controle da tuberculose até 2015.

No Brasil em 2009, foram notificados 72 mil novos casos, correspondendo a um coeficiente de incidência de 38/100.000 habitantes. Esses indicadores colocaram o país na 19ª posição em relação ao numero de casos e na 104ª posição em relação ao coeficiente de incidência (MS, 2011). A tuberculose continua sendo um grande problema de saúde global. Isso faz com que problemas de saúde entre os milhões de pessoas a cada ano e classifica como a segunda principal causa de morte de doenças infecciosas em todo o mundo, após o vírus da imunodeficiência humana (WHO, 2012).

Devido ao aumento de casos de tuberculose no Brasil nos últimos anos, faz-se necessário o estudo de ferramentas estatísticas apropriadas, aqui às técnicas de análise de sobrevivência são consideradas, pois se ajustam cada vez mais aos dados, que, frequentemente, são encontrados em vários tipos de estudos, especialmente, clínicos e observacionais. O objetivo deste estudo é aplicar análise de sobrevivência aos fatores de risco associados ao tempo até ocorrência de cura da tuberculose em pacientes atendidos pela Superintendência Regional de Saúde da cidade de Patos de Minas - MG.

## Fundamentação teórica

A análise de sobrevivência refere-se ao estudo de dados relacionados ao tempo até a ocorrência de um determinado evento de interesse, a partir de um tempo inicial até o tempo final de estudos predefinidos. Em geral, o tempo inicial coincide com o final do tempo de recrutamento de indivíduos para um determinado estudo experimental, ou com a data de aleatorização, em um estudo clínico aleatório, ou a data do diagnóstico ou do início do tratamento de uma determinada doença. Em estudos médicos o evento em questão, geralmente relaciona-se a morte do paciente, ao tempo até a ocorrência de uma determinada doença ou a complicação da mesma, ao tempo de duração da doença, ao tempo de cura (MEDRONHO et al., 2006).

### Caracterizando dados de sobrevivência

Os conjuntos de dados de sobrevivência são caracterizados pelos tempos de falha e, muito frequentemente, pelas censuras. Estes dois componentes contituem a resposta. Em estudos clínicos, um conjunto de covariáveis é também medido em cada paciente (COLOSIMO e GIOLO, 2006).

### Tempo de falha

O tempo de início de estudo deve ser precisamente definido. Os indivíduos devem ser comparáveis na origem do estudo, com exceção de diferenças medidas pelas covariáveis. Em um estudo clínico aleatorizado, a data da aleatorização é a escolha natural para a origem do estudo. A falha pode ainda ocorrer devido a uma única causa ou mais. Situações em que causas de falha competem ente si são denominadas de risco competitivo (COLOSIMO e GIOLO, 2006).

### Censura e dados truncados

Estudos que envolvem resposta temporal são frequentemente prospectivos e de longa duração. Uma característica decorrente destes estudos é, a presença de observações incompletas ou parciais, denominadas censuras (COLOSIMO e GIOLO, 2006).

Segundo Colosimo e Giolo (2006) alguns mecanismos de censuras são diferenciados em estudos clínicos. Censura tipo I é aquela em que o estudo será terminado após um período pré estabelecido de tempo. Censura do tipo II é aquela em que o estudo será terminado a pós ter ocorrido o evento de interesse em um número pré estabelecidos de indivíduos. Um terceiro mecanismo de censura, o do tipo aleatório é o que mais ocorre é estudos clínicos. Isto acontece quando um paciente é retirado no decorrer do estudo sem ter ocorrido uma falha.

### Variáveis explicativas

Valores grandes da variável tempo necessitam de mais tempo e persistência para serem observadas. Em situações extremas este fato pode comprometer a observação do valor da variável para alguns indivíduos, uma vez que o evento de interesse pode não ocorrer até o tempo final de estudo. Também o paciente pode abandonar o estudo antes da observação do evento devido a uma possível mudança de localidade (lost follow-up) ou devido a efeitos colaterais relacionados à terapia (drop out) ou falecer devido a outras causas de morte, mas não a em estudo. Entretanto, apesar de não termos o tempo até a ocorrência de um evento de interesse para um determinado indivíduo, temos o tempo até, por exemplo, o abandono do estudo. Dessa forma existe a necessidade de introdução de uma variável dicotômica na análise que indique se o valor de tempo de sobrevida para um determinado indivíduo foi ou não observado. Essa variável é conhecida como variável indicadora de censura, ou simplesmente censura, e é definida com sendo igual a um, se o tempo de sobrevida é observado, e igual à zero, caso o tempo de sobrevida seja censurado (MEDRONHO et al., 2006).

## Especificando o tempo de sobrevivência

A variável aleatória não negativa  $T$ , usualmente contínua, que representa o tempo de falha, é geralmente especificada em análise de sobrevivência pela sua função de sobrevivência. Essa função é definida como a probabilidade de uma observação falhar até um certo tempo  $t$ , ou seja, a probabilidade de uma observação sobreviver ao tempo  $t$ . Em termos probabilísticos, isto é escrito como,

$$S(t) = P(T \geq t)$$

### Função de taxa de falha ou de risco

Segundo Colosimo e Giolo (2006) a probabilidade de falha de ocorrer em um intervalo de tempo  $[t_1, t_2)$  pode ser expressa em termos da função de sobrevivência como:

$$S(t_1) - S(t_2) \quad (1)$$

a taxa de falha no intervalo  $[t_1, t_2)$  é dada como a probabilidade de que a falha ocorra neste intervalo, dado que não ocorreu antes de  $t_1$ , dividida pelo comprimento do intervalo. Assim, a taxa de falha no intervalo  $[t_1, t_2)$  é expressa por,

$$\frac{S(t_1) - S(t_2)}{(t_2 - t_1)S(t_1)} \quad (2)$$

De forma geral, redefinindo o intervalo  $[t, t + \Delta t)$ , a expressão (2) assume a seguinte forma:

$$\lambda(t) = \frac{S(t) - S(t + \Delta t)}{\Delta t S(t)}$$

assumindo  $\Delta t$  bem pequeno,  $\lambda(t)$  representa a taxa de falha instantânea no tempo  $t$  condicional a sobrevivência até o tempo  $t$ . A função de taxa de falha  $\lambda(t)$  é bastante útil para descrever a distribuição do tempo de vida de pacientes. Ela descreve a forma em que a taxa instantânea de falha muda com o tempo. A função de taxa de falha de  $T$  é definida como:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \quad (3)$$

## Técnicas não paramétricas

Quando se faz uma análise estatística, o primeiro passo é fazer uma análise descritiva, mas a presença de censura nos dados faz com que a aplicação destas técnicas descritivas não possam ser utilizadas como habitualmente, uma vez que não se conhece a frequência exata associada a cada intervalo. Pode-se colocar como exemplo, o gráfico de dispersão de cada covariável contínua versus a variável resposta, que pode ser utilizado para verificar se há relação linear entre as variáveis ou ainda identificar se há adequação de um modelo proposto.

A existência de dados censurados gera dificuldades na interpretação deste gráfico, mas com alguns critérios tal gráfico gera informações descritivas sobre a relação entre as variáveis. Na estatística básica são utilizados principalmente medidas de tendência central e variabilidade. Mas quando são encontrados dados censurados não se faz uso destas medidas. Na presença de tais dados a medida mais utilizada é a função de sobrevivência, que descreve a probabilidade de uma observação não falhar antes de determinado tempo. Pode-se estimar a função de sobrevivência de várias formas, uma delas é utilizando técnicas não paramétricas como o estimador de Kaplan-Meier.

### Estimador de Kaplan-Meier

O estimador não paramétrico de Kaplan-Meier, proposto por Kaplan e Meier em 1958 para estimar a função de sobrevivência, é também chamado de estimador limite-produto. Ele é uma adaptação da função de sobrevivência empírica que, na ausência de censuras é definido como

$$S(t) = \frac{n_t}{n_e} \quad (4)$$

em que  $S(t)$  é uma função escada com degraus nos tempos observados de falha de tamanho  $\frac{1}{n}$ , em que  $n$  é o tamanho da amostra,  $n_t$  é o número de observações que falharam até o tempo  $t$ ,  $n_e$  é o total de observações em estudo. O estimador de Kaplan-Meier considera tantos intervalos de tempo quantos forem o número de falhas distintas. Os limites dos intervalos de tempo são os tempos de falha da amostra. Este estimador é definido como:

$$\hat{S}(t) = \prod_{j:t_j < t} \left( \frac{n_j - d_j}{n_j} \right) = \prod_{j:t_j < t} \left( 1 - \frac{d_j}{n_j} \right) \quad (5)$$

em que  $t_1 < t_2 \dots < t_k$ , os  $k$  tempos distintos e ordenados de falha,  $d_j$  o número de falhas em  $t_j$ ,  $j = 1, \dots, k$ , e  $n_j$  o número de indivíduos sob risco em  $t_j$ , ou seja, os indivíduos que não falharam e não foram censurados até  $t$  instante imediatamente anterior a  $t_j$ . As principais propriedades do estimador de Kaplan-Meier são basicamente as seguintes:

- i. é não viciado para amostras grandes,
- ii. é fracamente consistente,
- iii. converge assintoticamente para um processo gaussiano e
- iv. é estimador de máxima verossimilhança de  $S(t)$ .

### Modelos probabilísticos

A existência de técnicas não paramétricas para estimar os parâmetros em estudos de análise de sobrevivência, pode-se ajustar os dados utilizando modelos paramétricos de probabilidade para o tempo de falha. Existe uma série de modelos probabilísticos que podem ser utilizados em análise de sobrevivência, sendo que alguns deles ocupam uma posição de destaque por sua comprovada adequação a várias situações práticas como modelos exponencial, Weibull, Log-normal e Log-logística.

### Distribuição exponencial

A distribuição exponencial é um dos modelos probabilísticos mais simples usados para descrever o tempo de falha. Esta distribuição apresenta um único parâmetro e é a única que se caracteriza por ter uma função de taxa de falha constante. A função densidade de probabilidade para a variável aleatória tempo de falha  $t$  com distribuição exponencial é dada por:

$$f(t) = \frac{1}{\alpha} \exp \left\{ - \left( \frac{t}{\alpha} \right) \right\}, t \geq 0 \quad (6)$$

em que o parâmetro  $\alpha > 0$  é tempo médio de vida, o parâmetro  $\alpha$  tem a mesma unidade do tempo de falha  $t$ . Isto é, se  $t$  é medido em horas,  $\alpha$  também será fornecido em horas (COLOSIMO e GIOLO, 2006).

### Distribuição de Weibull

A distribuição de Weibull é frequentemente utilizada em estudos biomédicos e industriais. A sua popularidade em aplicações práticas se deve ao fato dela apresentar uma grande variedade de formas, todas com as propriedades básicas: a função de taxa de falha 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

$$f(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1} \exp \left\{ - \left( \frac{t}{\alpha} \right)^\gamma \right\}, t \geq 0 \quad (7)$$

em que  $\gamma$ , o parâmetro de forma, e  $\alpha$ , o de escala, são ambos positivos. O parâmetro  $\alpha$  tem a mesma unidade de medida de  $t$  e  $\gamma$  não tem unidade. Para esta distribuição, as funções de sobrevivência e de risco são respectivamente,

$$\hat{S}(t) = \exp \left\{ - \left( \frac{t}{\alpha} \right)^\gamma \right\} \quad (8)$$

e

$$\lambda(t) = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1} \quad (9)$$

para  $t \geq 0$ ,  $\alpha$  e  $\gamma > 0$ . Observa que  $\gamma=1$ , tem-se distribuição exponencial e, sendo assim, a distribuição exponencial é um caso particular da distribuição de Weibull (COLOSIMO e GIOLO, 2006).

### Distribuição Log-normal

A distribuição Log-normal é muito utilizada para caracterizar tempos de vida de produtos e indivíduos. Ela também é utilizada para descrever situações clínicas, como o tempo de vida de pacientes. A função de densidade de uma variável aleatória  $T$  com distribuição Log-normal é dada por,

$$f(t) = \frac{1}{\sqrt{2\pi t\alpha}} \exp \left\{ - \frac{1}{2} \left( \frac{\log(t) - \mu}{\sigma} \right)^2 \right\}, t > 0, \quad (10)$$

em que  $\mu$  é a média do logaritmo do tempo de falha, assim como  $\sigma$  é o desvio padrão (COLOSIMO e GIOLO, 2006).

### Distribuição Log-logística

Segundo Colosimo e Giolo (2006) uma variável aleatória  $T$  com distribuição log-normal, a função de densidade é expressa por:

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

sendo  $\alpha > 0$  o parâmetro de forma e  $\gamma > 0$  o de escala. As funções de sobrevivência e de taxa de falha são expressas, respectivamente, por:

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

e

$$\lambda(t) = \frac{\gamma(t/\alpha)^{\gamma-1}}{\alpha[1 + (t/\alpha)^\gamma]}$$

## Distribuição gama e gama generalizada

A distribuição gama foi inicialmente usada descrever o tempo de vida de copos de vidro. Em problemas da área médica, sua utilização na descrição de tempos de vida de pacientes é mais recente. A função de densidade da distribuição gama, que é caracterizada por dois parâmetros,  $K$  e  $\alpha$ , em que  $k > 0$  é chamado de parâmetro de forma e  $\alpha > 0$  de escala, é expressa por:

$$f(t) = \frac{1}{\Gamma(k)\alpha^k} t^{k-1} \exp \left\{ - \left( \frac{t}{\alpha} \right) \right\}, t > 0. \quad (11)$$

com  $\Gamma(k)$  a função gama definida por  $\Gamma(k) = \int_0^{\infty} x^{k-1} \exp \{-x\} dx$ . Para  $k > 1$ , esta função de densidade apresenta um único pico em  $t = (k - 1)/\alpha$ . A respectiva função de sobrevivência desta distribuição é dada por:

$$S(t) = \int_t^{\infty} \frac{1}{\Gamma(k)\alpha^k} u^{k-1} \exp \left\{ - \left( \frac{u}{\alpha} \right) \right\} du \quad (12)$$

A função de taxa de falha, obtida da relação  $\lambda(t) = f(t)/S(t)$ , apresenta um padrão crescente ou decrescente convergindo, no entanto, para um valor constante  $t$  cresce de 0 a infinito (COLOSIMO e GIOLO, 2006). Sendo a variável resposta o tempo, em meses, até a cura do paciente. O tempo referente aos pacientes que morrem neste período será considerado como falha, e aquele referente aos pacientes que foram curados será considerado como censura.

Os modelos de tempo de falha acelerado (AFT), são modelos que assumem a relação entre  $S(\cdot)$  de 2 indivíduos,  $i$  e  $j$ , seguem a seguinte relação:

$$S_i t = S_j(\phi_{ij} t) \quad (13)$$

em que,  $S_i t$  é função do tempo  $t$ ;  $\phi$  constante específica para o par  $(i, j)$ , caso  $\phi_{ij}$  for igual a 1, as curvas de sobrevivência se sobrepõem. A expressão seguinte é uma sub-classe de modelos AFT,

$$T_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \sigma \varepsilon \quad (14)$$

Em que,  $T_i$  tempo até a ocorrência do evento;  $\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$  os parâmetros a serem estimados;  $x_i$  são as covariáveis;  $\sigma \varepsilon$  erro aleatório.

No modelo tempo de falha acelerado o erro não necessariamente assume uma distribuição normal, são permitidas outras distribuições, como por exemplo, a distribuição de Weibull, Log-normal, Exponencial, Log-logística e a Gama. Enquanto que no modelo regressão clássica o erro só assume a distribuição normal.

Neste trabalho será utilizado o estimador de Kaplan-Meier para modelar o tempo até a cura em função das covariáveis medidas, determinando a distribuição que mais se ajusta aos dados, partindo da distribuição gama generalizada, pois é um caso particular das distribuições de Weibull, Exponencial e Log-normal, que serão testadas. No processo de seleção de covariáveis, a estatística de teste será obtida usando o teste da razão de verossimilhanças com uma distribuição de qui-quadrado, onde os graus de liberdade são os parâmetros do modelo a ser testado, dado a seguir;

$$TRV = -2[\log L(\hat{\theta}) - \log L(\hat{\theta}_0)] \quad (15)$$

em que,  $L(\hat{\theta})$  e  $L(\hat{\theta}_0)$  são as funções de verossimilhanças maximizadas do modelo gama generalizado sem restrição e sob a hipótese nula, respectivamente. A estatística do teste, TRV, segue aproximadamente uma distribuição qui-quadrado, com número de graus de liberdade igual à diferença entre o número de parâmetros dos dois modelos. A escolha do modelo mais adequado também será feita utilizando o TRV entre o modelo gama generalizado e os modelos Weibull, Log-normal e Exponencial.

## Metodologia

Os dados para o desenvolvimento deste estudo foram fornecidos pela Superintendência Regional de Saúde (SRS) - Patos de Minas. Foram obtidos registros de 437 pacientes diagnosticados com tuberculose em toda a região abrangida pela SRS no período de janeiro de 2003 a dezembro 2012. A variável resposta foi o tempo de ocorrência, em dias, de um dos possíveis eventos.

Os eventos considerados neste estudo são: óbito e cura, no sentido de que a ocorrência de um evento exclui a possibilidade de ocorrência do outro. É importante destacar que pacientes que mudaram de região foram excluídos do estudo. Para cada paciente o tempo é computado no momento em que ele inicia tratamento e, termina no momento em que um dos eventos venha a ocorrer. De cada paciente foram obtidas as informações descritas na tabela 1, que constituíram as covariáveis em estudo.

Tabela 1: Informações sobre os pacientes

Dados	Legenda	Situação
Tempo de tratamento em dias	Tempo	
Censura	Cens	0(Censurado),1(Falha)
Sexo do paciente	Sexo	0(Masculino),1(Feminino)
Idade do paciente	Idade	0( $\leq 40$ ), 1( $> 40$ )
Raça do paciente	Raça	0 (Branca),1( Outras)
Região de origem do paciente	Região	0 (Patos de Minas),1(outras)

Os modelos de tempo de falha acelerado (AFT), são modelos que assumem a relação entre  $S(.)$  de 2 indivíduos,  $i$  e  $j$ , seguem a seguinte relação:

$$S_i t = S_j(\phi_{ijt}) \quad (16)$$

em que,  $S_i t$  é função do tempo  $t$ ;  $\phi$  constante específica para o par  $(i, j)$ , caso  $\phi_{ij}$  for igual a 1, as curvas de sobrevivência se sobrepõem. A expressão seguinte é uma sub classe de modelos AFT,

$$T_i = \beta_o + \beta_1 x_{i1} + \dots + \beta_k x_{1k} + \sigma \varepsilon \quad (17)$$

Em que,  $T_i$  tempo até a ocorrência do evento;  $\beta_o + \beta_1 x_{i1} + \dots + \beta_k x_{1k}$  os parâmetros a serem estimados;  $x_i$  são as covariáveis;  $\sigma \varepsilon$  erro aleatório.

Neste trabalho foi utilizado o modelo paramétrico AFT para modelar o tempo até a cura em função das covariáveis consideradas, determinando a distribuição que melhor se ajusta aos dados, partindo da distribuição gama generalizada. No processo de seleção de covariáveis, utilizou-se o teste de razão de verossimilhanças, cuja estatística é dada por:

$$TRV = -2[\log L(\hat{\theta}) - \log L(\hat{\theta}_0)] \quad (18)$$

em que,  $L(\hat{\theta})$  e  $L(\hat{\theta}_0)$  são as funções de verossimilhanças maximizadas do modelo gama generalizado sem restrição e sob a hipótese nula, respectivamente. A estatística do teste, TRV, segue aproximadamente uma distribuição qui-quadrado, com número de graus de liberdade igual à diferença entre número de parâmetros dos dois modelos. A escolha do modelo mais adequado também foi feita utilizando o TRV entre o modelo gama generalizado e os modelos Weibull, Log-normal e Exponencial.

As análises foram feitas utilizando o software estatístico R (R CORE TEAM, 2012), versão 2.13.3.

## Resultados e discussão

No processo de seleção das covariáveis utilizando o modelo gama generalizado, foram significativas a idade, a raça, região de origem e sexo, a um nível de significância de 0,05. Foram testados os modelos de Weibull, Log-normal e Exponencial, entre estes três modelos, o modelo Log-normal foi considerado o mais adequado, de acordo com o teste de razão de verossimilhanças.

O gráfico das probabilidades de sobrevivência dos resíduos estimadas por Kaplan-Meier e pelo modelo Log-normal, bem como de suas respectivas curvas de sobrevivência estimadas, estão na Figura 1. A partir desta pode-se acreditar que o modelo de regressão Log-normal se encontra bem ajustado aos dados sob análise.

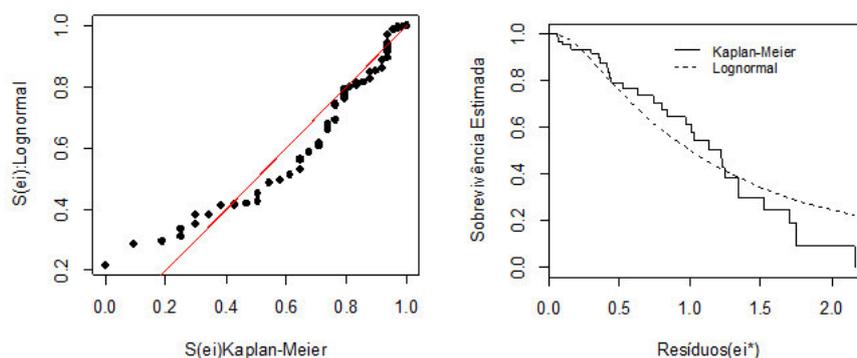


Figura 1: Sobrevivências dos resíduos  $e_i$  estimadas pelo método de Kaplan-Meier e pelo modelo Log-normal (gráfico à esquerda) e respectivas curvas de sobrevivência estimadas (gráfico à direita)

Tabela 2: Modelo de regressão Log-normal.

Parâmetro (covariável)	Estimativa	Erro padrão	$z$	$Pr > z$
(Intercepto)	10.840	1.691	6.411	1.44e-10
$\beta_1$ (Idade)	-0.755	0.823	-0.918	3.59e-01
$\beta_2$ (Sexo)	0.739	0.859	0.860	3.90e-01
$\beta_3$ (Município)	-0.300	0.835	-0.360	7.19e-01
$\beta_4$ (Raça)	0.747	0.757	0.986	3.24e-01

Tomando-se o exponencial dos coeficientes estimados, apresentados na tabela 2, obtém-se a razão dos tempos medianos de sobrevivência. Assim, tem-se que para a covariável Idade, por exemplo, o tempo mediano até a cura de pacientes com mais de 40 anos, é de  $((1 - \exp - 0,755) = 52,99)$  52,99% a mais que pacientes com menos de 40 anos. Os demais coeficientes são interpretados usando o mesmo procedimento.

## Conclusões

A interpretação dos coeficientes estimados, que para uma covariável codificada (0 e 1), compara a probabilidade de sobrevivência do grupo 1 em relação ao grupo 0, permitiu obter as seguintes conclusões: Pacientes com mais de 40 anos tem 52,99% mais probabilidade de cura que pacientes com menos de 40 anos; pacientes do sexo feminino tem 109,38% a mais na probabilidade de cura que pacientes do sexo masculino; pacientes do município de Patos de Minas

tem 259,18% a mais na probabilidade de cura que pacientes de outras regiões; pacientes da raça branca tem 111,06% a mais na probabilidade de cura que pacientes de outras raças.

## Agradecimento

À Superintendência Regional de Saúde - Patos de Minas pelo apoio e empenho em fornecimento de todos os dados utilizados neste estudo.

## Referências

CHIRINOS, N. E. C.; MEIRELLES, B. H. S. *Fatores Associados ao Abandono do Tratamento da Tuberculose: uma revisão integrativa*. Texto & Contexto - Enfermagem. Florianópolis, v. 20, n. 3, p.599-606, Jul. 2011. Disponível em: <http://www.scielo.br/pdf/tce/v20n3/23.pdf>. Acesso em: 05 mar. 2014.

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

HIJJAR, M. A.; OLIVEIRA, M. J. P. R.; TEIXEIRA, G. *A tuberculose no Brasil e no mundo*. Boletim de Pneumologia Sanitaria, Rio de Janeiro, v. 9, n. 2, p.9-16, dez. 2001.

MEDRONHO, R. A; BLOCH, K. V; LUIZ, R.R; WERNECK, G. *Epidemiologia*. São Paulo: Atheneu, 2006.

MINISTÉRIO DA SAÚDE (MS). *Manual de recomendações para o controle da tuberculose no Brasil*. Brasília, 2011.

R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2012. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

SOUZA, M. V. N. *A tuberculose e o desenvolvimento de novos tuberculostáticos em Far-Manguinhos (FIOCRUZ) no Rio de Janeiro*. Revista Virtual de Química: Rio de Janeiro, v. 1, n. 1, p.9-16, fev, 2009. Disponível em: <http://www.uff.br/RVQ/index.php/rvq/article/viewFile/11/1j>. Acesso em: 03 mar. 2014.

WORLD HEALTH ORGANIZATION (WHO). *Global Tuberculosis Report*. Geneva, 2013. Disponível em: <http://www.who.int/tb/publications/global/report/en/>. Acesso em: 03 mar. 2014.

WORLD HEALTH ORGANIZATION (WHO). *Global Tuberculosis Report*. Geneva, 2012. Disponível em: <http://www.who.int/tb/publications/global/report/en/>. Acesso em: 28 jan. 2014.