

Universidade Federal do Paraná  
Setor de Ciências Exatas  
Departamento de Estatística

Ana Flávia do Carmo Santos

Felipe Werner

**Análise Estatística de Sobrevivência: um estudo  
de pacientes com câncer de mama tratados no  
município de Curitiba**

**Curitiba**

**2015**

Ana Flávia do Carmo Santos  
Felipe Werner

**Análise Estatística de Sobrevivência: um estudo de  
pacientes com câncer de mama tratados no município de  
Curitiba**

Projeto de Pesquisa apresentado à disciplina  
Laboratório A do Curso de Graduação em Es-  
tatística da Universidade Federal do Paraná,  
como requisito para elaboração do Trabalho  
de Conclusão de Curso.

Orientadora: Profa. Dra. Suely Ruiz Giolo

Curitiba  
2015

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b> . . . . .	<b>3</b>
<b>2</b>	<b>OBJETIVOS</b> . . . . .	<b>4</b>
<b>2.1</b>	<b>Objetivos Gerais</b> . . . . .	<b>4</b>
<b>2.2</b>	<b>Objetivos Específicos</b> . . . . .	<b>4</b>
<b>3</b>	<b>MATERIAL E MÉTODOS</b> . . . . .	<b>5</b>
<b>3.1</b>	<b>Material</b> . . . . .	<b>5</b>
3.1.1	Banco de Dados . . . . .	5
3.1.2	Software . . . . .	5
<b>3.2</b>	<b>Métodos</b> . . . . .	<b>5</b>
<b>4</b>	<b>CRONOGRAMA DE ATIVIDADES</b> . . . . .	<b>9</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>10</b>

# 1 Introdução

O câncer de mama é um tumor maligno resultante da multiplicação de células anormais da mama (INCA, 2015). Segundo o Instituto Nacional de Câncer - INCA, este é o tipo de câncer mais comum entre as mulheres no Brasil e no mundo, tanto em países em desenvolvimento quanto em países desenvolvidos. Devido à dimensão da doença, ela possui uma das mais populares campanhas de prevenção, conhecida como "Outubro Rosa", que visa chamar atenção para a realidade atual do câncer de mama.

O câncer é uma doença que pode se espalhar para outros órgãos do corpo, fenômeno chamado de metástase. No caso do câncer de mama, o mais comum é que a doença afete os ossos, os pulmões, o fígado ou o cérebro (IMAMA - INSTITUTO DA MAMA RS, 2014). Mesmo que o câncer de mama metastático tenha se espalhado para outra parte do corpo, ele é considerado e tratado como câncer de mama. Por exemplo, se o câncer de mama se espalhou para os pulmões, ele ainda é considerado câncer de mama e é tratado com medicamentos para câncer de mama.

Importantes avanços na abordagem do câncer de mama aconteceram nos últimos anos, principalmente no que diz respeito à cirurgias menos mutilantes, assim como a busca da individualização do tratamento. O tratamento varia de acordo com o estadiamento da doença, suas características biológicas, bem como das condições da paciente (idade, status menopausal e comorbidades).

O prognóstico do câncer de mama depende da extensão da doença. Quando a doença é diagnosticada no início, o tratamento tem maior potencial curativo. Entretanto, quando há evidências de metástases (doença se espalhou), o tratamento tem por objetivos principais prolongar a sobrevida e melhorar a qualidade de vida. As modalidades de tratamento do câncer de mama podem ser divididas em duas, sendo a primeira o tratamento local, onde é feito um tratamento mais simples com cirurgia e radioterapia, e a segunda o tratamento sistêmico, que é um tratamento mais árduo para o paciente e é feito quando a doença já se espalhou pelo corpo. O tratamento consiste em sessões de quimioterapia, hormonioterapia e terapia biológica (INCA, 2014a).

Cerca de 1,67 milhões de casos novos dessa neoplasia foram esperados para o ano de 2012 em todo o mundo, o que representa 25% de todos os tipos de câncer diagnosticados nas mulheres (INCA, 2014b). Pesquisa divulgada pela Revista Veja aponta que, no Brasil, pessoas com tumores de mama e próstata estão vivendo mais (ALLEMANI et al., 2015). Entretanto, foi estimado para 2014 e 2015, o diagnóstico de 57.120 novos casos de câncer de mama no Brasil. Desses 57 mil novos casos, 10.370 seriam na região Sul do Brasil, sendo 3.490 no estado do Paraná (INCA, 2014b).

Considerando as informações mencionadas, o objetivo deste trabalho é o de analisar fatores associados à sobrevida de pacientes com câncer de mama que foram tratados no município de Curitiba, entre os anos de 1990 a 2009.

## 2 Objetivos

### 2.1 Objetivos Gerais

Utilizando técnicas de análise de sobrevivência, identificar fatores associados à sobrevida de pacientes com câncer de mama, que foram tratados em um centro médico de Curitiba.

### 2.2 Objetivos Específicos

- a) Revisar a literatura no que diz respeito a fatores associados à sobrevida de pacientes com câncer de mama e seus índices no Brasil e no mundo;
- b) Realizar uma análise descritiva dos dados para entendimento detalhado e consistente das informações;
- c) Utilizar o estimador de Kaplan-Meier a fim de investigar a influência de covariáveis na variável resposta (tempo desde o diagnóstico até o óbito de pacientes com câncer de mama);
- d) Ajustar modelos de regressão aos dados descritos na Seção 3.1.1 com a finalidade de verificar as covariáveis que influenciam no tempo desde o diagnóstico até o óbito de pacientes com câncer de mama;
- e) Comparar os resultados dos modelos de regressão a fim de obter um melhor entendimento do conjunto de dados;
- f) Apresentar conclusões a partir das análises realizadas a fim de obter informações sobre os fatores associados à sobrevida de pacientes com câncer de mama.

## 3 Material e Métodos

A seguir, são descritos brevemente os métodos estatísticos que se pretende utilizar neste trabalho para a análise dos dados descritos na Seção 3.1.1.

### 3.1 Material

#### 3.1.1 Banco de Dados

Para execução do projeto será utilizado um banco de dados fornecido por um centro médico de Curitiba o qual contém informações de pacientes com câncer de mama. As informações foram coletadas em um período de 20 anos de série histórica, compreendendo os anos de 1990 a 2009 (LIGA PARANAENSE DE COMBATE AO CÂNCER, 2011). Foram observados 3.569 pacientes, dos quais 1.549 foram a óbito e os demais, denominados censuras, não foram a óbito ou deixaram o estudo por outras razões. O Quadro 1 apresenta uma descrição das covariáveis disponíveis no conjunto de dados.

#### 3.1.2 Software

Para as análises e tratamento do banco de dados será utilizado o *software* livre R, versão 3.2.2 (R CORE TEAM, 2015). Tal *software* permite a utilização de códigos abertos, propiciando um amplo espaço para análises estatísticas, gerenciamento de banco de dados e também análises gráficas.

### 3.2 Métodos

A análise de sobrevivência, técnica muito utilizada atualmente em estudos na área da saúde, tem interesse em avaliar o tempo até a ocorrência de um evento. Neste projeto, o evento de interesse refere-se ao tempo, em meses, a partir da data de diagnóstico da doença até o óbito de pacientes com câncer de mama. A presença de censuras, informação parcial da resposta, é comum em diversos estudos em que se tem interesse em avaliar o tempo de sobrevivência (COLOSIMO; GIOLO, 2006). No estudo descrito na Seção 3.1.1, tem-se um percentual de 56,6% de censuras.

Para a análise dos dados mencionados na Seção 3.1.1 será realizado, inicialmente, um estudo descritivo, buscando observar o perfil dos pacientes diagnosticados com câncer de mama. Para isso, será utilizado o estimador de Kaplan-Meier (KAPLAN; MEIER, 1958), o qual é uma técnica não-paramétrica utilizada para estimar a função de sobrevivência  $S(t) = P(T \geq t)$  quando o conjunto de dados apresenta censuras.

Quadro 1 – Descrição de covariáveis disponíveis nos dados de câncer de mama

COVARIÁVEL	DESCRIÇÃO	CATEGORIA
Tumor	Paciente possui mais do que um tumor	S - Sim N - Não
Sexo	Sexo do paciente	F - Feminino M - Masculino C - Casado
Civil	Estado civil do paciente	O - Outros S - Solteiro V - Viúvo
Idade	Idade do paciente	-
AED	Avaliação e extensão da doença	Localizado Metástase Extensão direta Envio por linfonodos regionais In situ (no local) Sem Informação / Não aplicável
Topografia	Localização do tumor	-
Morfologia	Tipo de tumor	Carcinoma Adenocarcinoma Comedocarcinoma Neoplasma Sarcoma Outros
GD	Grau de diferenciação	Pouco diferenciado Moderadamente diferenciado Bem diferenciado Indiferenciado Sem Informação
Estadiamento	Quanto o câncer se espalhou pelo corpo	Tumor: 1 - Limitado a mama, sem metástase 2 - Pode estar estendido além da mama 3 - Envolve regiões vizinhas 4 - Com metástase a distância
TFI	Tratamento feito na instituição	2 - Cirurgia 3 - Radioterapia 4 - Quimioterapia 5 - Hormonioterapia 8 - Outros

A função de sobrevivência é uma função escada, com degraus nos tempos em que foram observadas as falhas. A partir desse estimador pode-se obter estatísticas de interesse como o tempo médio, mediano e alguns percentis.

Também será utilizado o teste *logrank* (MANTEL, 1966) a fim de comparar as curvas de sobrevivência associadas às categorias de cada covariável analisada. Posteriormente, modelos de regressão serão considerados com o intuito de modelar a relação entre as covariáveis e o tempo de sobrevivência (COLOSIMO; GIOLO, 2006).

Inicialmente, será considerado o modelo de regressão de Cox, que é amplamente utilizado em estudos clínicos devido à sua versatilidade (COLOSIMO; GIOLO, 2006). Tendo em vista que o modelo de regressão de Cox assume o pressuposto de taxas de falha proporcionais, ou seja, que a razão das taxas de falhas de dois indivíduos quaisquer é

constante no tempo, serão considerados outros modelos, tal como o modelo aditivo de Aalen, caso a suposição de taxas de falha proporcionais não seja atendida. A função taxa de falha  $\lambda(t)$  representa a taxa de falha instantânea no tempo  $t$  condicional à sobrevivência até o tempo  $t$  (COLOSIMO; GIOLO, 2006).

A expressão do modelo de regressão de Cox em termos da função taxa de falha é dada, para o  $i$ -ésimo indivíduo em um dado tempo  $t$ , por:

$$\lambda(t | \mathbf{x}_i) = \lambda_0(t) \exp \left( \sum_{k=1}^p \beta_k x_{ik} \right), \quad (3.1)$$

em que  $p$  é o número de covariáveis,  $x_{ik}$  é o valor da  $k$ -ésima covariável observada para o  $i$ -ésimo indivíduo,  $\beta_k$  é o parâmetro que descreve o efeito da  $k$ -ésima covariável e  $\lambda_0(t)$  é uma função do tempo não-negativa e não especificada, usualmente denominada função taxa de falha de base ou basal, que corresponde a taxa comum a todos os indivíduos.

Para o modelo aditivo de Aalen, a função taxa de falha em um dado tempo  $t$  é dada por:

$$\lambda(t | \mathbf{x}_i(t)) = \beta_0(t) + \sum_{k=1}^p \beta_k(t) x_{ik}(t), \quad (3.2)$$

em que  $p$ ,  $x_{ik}$  e  $\beta_k$  são especificados como na equação (3.1), porém variam com o tempo  $t$ , e  $\beta_0(t)$  corresponde à função taxa de falha de base.

Os modelos de Cox e Aalen diferem em alguns aspectos. O de Cox tem uma função taxa de falha de base não paramétrica (nenhuma forma paramétrica é assumida para esta função), mas o efeito das covariáveis é modelado de forma paramétrica. Já o modelo de Aalen é completamente não paramétrico (COLOSIMO; GIOLO, 2006).

O modelo de regressão de Cox apresenta vantagens, entre elas apresenta interpretação simples dos resultados, é facilmente estendido para agregar covariáveis dependentes do tempo, além de estar disponível em vários pacotes estatísticos. No entanto, segundo Aalen (1989), o modelo possui limitações, entre elas, a suposição de taxas de falha proporcionais pode não ser válida. Ademais, o modelo de Cox não é adequado para detectar mudanças de efeitos de covariáveis ao longo do tempo. Aalen (1989) também cita que a suposição de proporcionalidade das taxas de falha é vulnerável à mudanças no número de covariáveis no modelo. Dadas tais limitações, o modelo aditivo de Aalen se constitui em uma alternativa ao modelo de Cox. Este modelo permite que os parâmetros e os vetores de covariáveis variem com o tempo, sendo capaz de fornecer informações detalhadas sobre a influência temporal de cada covariável. No modelo de Aalen a função de sobrevivência não é monótona (dado que o modelo não é paramétrico), ou seja, pode haver intervalos nos quais ela é crescente e intervalos nos quais é decrescente. Desta forma, a função taxa de falha pode assumir valores negativos e isto é visto como uma desvantagem deste modelo (AALEN, 1989).

De acordo com Abadi et al. (2011), ainda não há métodos estatísticos que possibilitem a comparação dos dois modelos citados (Cox e Aalen), pois o fato de não serem

modelos encaixados dificulta a comparação dos mesmos. Abadi et al. (2011) também comentam que os modelos de Cox e Aalen não devem ser vistos como modelos alternativos, mas sim como métodos complementares que juntos oferecem uma melhor compreensão dos dados sob estudo.

Devido ao número elevado de covariáveis disponíveis no banco de dados, métodos de seleção de covariáveis, tais como o *backward*, *forward* e *stepwise*, podem ser necessários. A seleção *forward* parte da suposição de que não há covariáveis no modelo, apenas o intercepto. O intuito do método é adicionar uma covariável de cada vez, iniciando-se com a inclusão da covariável com maior correlação com a variável resposta. Considerando que a primeira covariável foi selecionada, o passo seguinte é encontrar a próxima covariável com a maior correlação com a variável resposta na presença da primeira no modelo e, assim por diante.

Enquanto o método *forward* começa sem nenhuma covariável no modelo e vai adicionando-as a cada passo, o método *backward* faz o caminho oposto. Inclui, inicialmente, no modelo todas as covariáveis e depois, por etapas, vai eliminando as não correlacionadas com a resposta. Por sua vez, o método *stepwise*, assim como o método *forward*, tem início com a inclusão da covariável que tiver a maior correlação com a variável resposta. Contudo, nos passos subsequentes uma covariável que tenha sido incluída em uma passo anterior pode vir a ser excluída em uma passo seguinte. O algoritmo é realizado até que não possa ser incluída ou excluída nenhuma covariável (FERREIRA, 2012).

Quanto à verificação da adequação dos modelos ajustados, pretende-se realizá-la por meio da análise de resíduos propostos na literatura para os modelos de Cox e de Aalen.

## 4 Cronograma de Atividades

ATIVIDADES	NOV 2015	DEZ 2015	JAN 2016	FEV 2016	MAR 2016	ABR 2016	MAI 2016	JUN 2016
<b>1 Projeto de Pesquisa</b>								
Elaboração do Projeto de Pesquisa								
Entrega da versão final do Projeto de Pesquisa ao orientador								
<b>2 Elaboração do Trabalho de Conclusão de Curso</b>								
Revisão de literatura sobre o tema								
Análise dos dados e discussão dos resultados obtidos								
Redação do trabalho de conclusão de curso								
Leitura do trabalho pelo orientador e correções								
Entrega do trabalho redigido aos membros da banca								
<b>3 Defesa do Trabalho de Conclusão de Curso</b>								
Preparação e apresentação do TCC								
<b>4 Elaboração da Versão Final do TCC</b>								
Elaboração da versão final do TCC								
Entrega da versão final do TCC ao orientador								

## Referências

- AALEN, O. O. A linear regression model for the analysis of life times. *Statistics in Medicine*, Wiley Online Library, v. 8, n. 8, p. 907–925, 1989.
- ABADI, A. et al. Comparison of aalen’s additive and cox proportional hazards models for breast cancer survival: analysis of population-based data from british columbia, canada. *Asian Pacific Journal of Cancer Prevention*, v. 12, n. 11, p. 3113–3116, 2011.
- ALLEMANI, C. et al. Global surveillance of cancer survival 1995–2009: analysis of individual data for 25 676 887 patients from 279 population-based registries in 67 countries (concord-2). *The Lancet*, Elsevier, v. 385, n. 9972, p. 977–1010, 2015.
- COLOSIMO, E. A.; GIOLO, S. R. *Análise de sobrevivência aplicada*. São Paulo, Brasil: Edgard Blücher, 2006. 392 p. (ABE - Projeto Fisher). ISBN 9788521203841.
- FERREIRA, A. *Modelos Lineares: seleção de variáveis*. Rio de Janeiro, UERJ, 2012. 9 p. Disponível em: <[http://wiki.nosdigitais.teia.org.br/images/f/fa/Apostila\\_selecao\\_variaveis.pdf](http://wiki.nosdigitais.teia.org.br/images/f/fa/Apostila_selecao_variaveis.pdf)>.
- IMAMA - INSTITUTO DA MAMA RS. *Câncer de Mama Metastático*. Porto Alegre, 2014. Acesso em 08 de Dez de 2015. Disponível em: <<http://www.imama.org.br/index.php/cancer-de-mama>>.
- INCA. *Controle do câncer de mama*. Rio de Janeiro, 2014. Acesso em 08 de Dez de 2015. Disponível em: <[http://www2.inca.gov.br/wps/wcm/connect/acoes\\_programas/site/home/nobrasil/programa\\_controle\\_cancer\\_mama/tratamento](http://www2.inca.gov.br/wps/wcm/connect/acoes_programas/site/home/nobrasil/programa_controle_cancer_mama/tratamento)>.
- INCA. *Incidência do Câncer no Brasil*. Rio de Janeiro, 2014. 124 p. Acesso em 15 de Nov de 2015. Disponível em: <<http://www.inca.gov.br/estimativa/2014/>>.
- INCA. *O câncer de mama*. Rio de Janeiro, 2015. Acesso em 15 de Nov de 2015. Disponível em: <<http://www.inca.gov.br/wcm/outubro-rosa/2015/cancer-de-mama.asp>>.
- KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, Taylor & Francis, v. 53, n. 282, p. 457–481, 1958.
- LIGA PARANAENSE DE COMBATE AO CÂNCER. *Relatório Epidemiológico: 1990 a 2009*. Curitiba: LPCC, 2011. 124 p. Acesso em 15 de Nov de 2015. Disponível em: <[http://www.erastogaertner.com.br/arquivos/rhc/DuasDecadas\\_RHC\\_HEG\\_1990a2009.pdf](http://www.erastogaertner.com.br/arquivos/rhc/DuasDecadas_RHC_HEG_1990a2009.pdf)>.
- MANTEL, N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports*, v. 50, n. 3, p. 163–170, 1966.
- R CORE TEAM. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2015. Disponível em: <<https://www.R-project.org/>>.