



UNIVERSIDADE FEDERAL DO PARANÁ
SETOR DE CIÊNCIAS EXATAS
DEPARTAMENTO DE ESTATÍSTICA
CURSO DE ESTATÍSTICA

Michele Mottin

Renato de Souza Brito

ANÁLISE DE SOBREVIDA DE PACIENTES COM CÂNCER DE OVÁRIO DE UM CENTRO MÉDICO DE CURITIBA

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

Orientadora: Profa. Dra. Suely Ruiz Giolo

**CURITIBA
2015**

Sumário

1 INTRODUÇÃO	3
2 OBJETIVOS	5
2.1 Objetivo Geral	5
2.2 Objetivos Específicos	5
3 PACIENTES E MÉTODOS	6
3.1 Pacientes	6
3.1.1 Conjunto de Dados	6
3.1.2 Recursos Computacionais	6
3.2 Métodos	8
4 CRONOGRAMA DE ATIVIDADES	10
REFERÊNCIAS	11

1 INTRODUÇÃO

O câncer de ovário é considerado o câncer ginecológico mais difícil de ser diagnosticado, pois a maioria dos tumores malignos de ovário são descobertos tardiamente, quando já estão em estágio avançado. E, conseqüentemente, faz com que o câncer de ovário seja o câncer ginecológico mais letal. Este é um dos fatores que instiga estudos sobre a sobrevivência de pacientes com esse tipo de câncer.

As taxas de sobrevivência são frequentemente utilizadas pelos médicos para discussão do prognóstico de pacientes. Essas taxas são calculadas com base em resultados anteriores de um grande número de pacientes que já tiveram a doença (INSTITUTO ONCOGUIA, 2014). Muitas variáveis estão envolvidas na medição das taxas de sobrevivência. Portanto, saber o tipo do tumor, o estadiamento, o grau de diferenciação do tumor e o tratamento realizado, dentre outros, são fatores muito importantes para a estimativa mais precisa e assertiva da sobrevivência das pacientes.

Um estudo realizado pelo NCI - *National Cancer Institute*, mostrou haver grande discrepância entre as taxas de sobrevivência de pacientes diagnosticadas entre os estágios I e IV da neoplasia. De acordo com esse estudo, que se baseou em dados coletados de 1975 a 2012 (HOWLADER et al., 2015), a taxa de sobrevivência média em cinco anos de pacientes diagnosticadas com câncer de ovário no estágio I da neoplasia foi de 92,1%. Esta taxa decresceu drasticamente para 28,3% para pacientes no estágio IV da neoplasia.

Nos Estados Unidos cerca de 20.000 mulheres são diagnosticadas por ano com câncer de ovário. Entre os cânceres do sistema reprodutor feminino, ele é o maior causador de morte. Em 2012 cerca de 14.400 mulheres foram a óbito devido a esta neoplasia (U.S. CANCER STATISTICS WORKING GROUP, 2015).

O câncer de ovário pode atingir qualquer faixa etária, contudo a maioria dos casos está concentrada em mulheres acima dos 50 anos de idade. Essa neoplasia é a sexta causa de morte entre os cânceres devido à dificuldade de diagnóstico. De acordo com dados registrados pelo Hospital Erasto Gaertner, o câncer de ovário é o quinto câncer que mais atinge a população feminina de Curitiba e região metropolitana estando em oitavo lugar entre os mais incidentes em todas as regiões do Brasil (LIGA PARANAENSE DE COMBATE AO CÂNCER, 2011).

Mesmo em países desenvolvidos, a taxa de mortalidade de pacientes com câncer de ovário ainda é bem elevada, pois pouco se avançou em termos de diagnóstico precoce.

A única maneira de interferir na história natural do câncer de ovário é o estabelecimento precoce do seu diagnóstico e a correta abordagem terapêutica (LUIZ et al., 2009).

Não há nenhuma maneira simples e confiável para detectar o câncer de ovário em mulheres que não possuem quaisquer sinais ou sintomas. Testes de diagnóstico como exame pélvico rectovaginal, ultrassonografia transvaginal ou um exame de sangue específico são realizados em pacientes que já possuem algum sintoma da doença. O exame Papanicolau, que é indicado para todas as mulheres realizarem rotineiramente, não consegue verificar a existência do câncer ovariano (CDC, 2016).

Os fatores de risco que potencializam o desenvolvimento do câncer ovariano ainda estão sendo fortemente estudados. Histórico familiar é o fator de risco isolado mais importante, principalmente se o grau de parentesco for primário, ou seja, mãe, avó ou filha tiverem desenvolvido o câncer de ovário. Fatores que também aumentam o risco do desenvolvimento da neoplasia são, dentre outros: ter tido câncer de mama, de útero ou colorretal ou nunca ter engravidado.

Embora não haja maneiras de prevenir o câncer de ovário, da mesma forma que existem fatores que aumentam o risco de desenvolvimento da doença, há também fatores que diminuem a chance do seu desenvolvimento. Fazer uso de pílulas anticoncepcionais, ter retirado os dois ovários, ter dado à luz e ter amamentado por um ano ou mais, são fatores que, dentre outros, podem diminuir o risco de um possível desenvolvimento de câncer ovariano (CDC, 2016).

Dada a relevância desse tema para a saúde da população feminina, este projeto tem como foco o estudo do tempo de sobrevida de pacientes brasileiras diagnosticadas com câncer de ovário que foram tratadas em um centro médico de Curitiba, Paraná, buscando a identificação de possíveis fatores (covariáveis) que afetam a sobrevida das mesmas.

2 OBJETIVOS

2.1 Objetivo Geral

Investigar fatores que afetam a sobrevivência de pacientes diagnosticadas com câncer de ovário utilizando técnicas e ferramentas estatísticas para a predição da sobrevivência dessas pacientes.

2.2 Objetivos Específicos

1. Revisar a literatura no que diz respeito às abordagens de análise propostas para dados de sobrevivência;
2. Realizar análise exploratória dos dados de pacientes diagnosticadas com câncer de ovário descritos na Seção 3.1.1 para entendimento detalhado e consistente das informações;
3. Ajustar modelos estatísticos aos dados descritos na Seção 3.1.1 com a finalidade de investigar quais as covariáveis (fatores), dentre as disponíveis, influenciam o tempo de sobrevivência das pacientes diagnosticadas com câncer de ovário;
4. Comparar e discutir os resultados dos modelos ajustados.

3 PACIENTES E MÉTODOS

3.1 Pacientes

3.1.1 Conjunto de Dados

O conjunto de dados que será utilizado para as análises é oriundo de um centro médico de Curitiba, Paraná, que é referência na luta contra o câncer. Foram coletadas duas décadas de dados de pacientes diagnosticados com câncer e que foram tratados no próprio hospital. Dado que a neoplasia de interesse neste estudo é o câncer ovariano, foram extraídas desta base todas as informações referentes a essa doença, totalizando 390 pacientes que foram acompanhadas no período de 1990 a 2009.

A base de dados das 390 pacientes contém diversas informações (covariáveis) que serão analisadas e classificadas como relevantes ou não para o estudo em questão. No Quadro 1 tem-se a descrição de algumas dessas covariáveis. As pacientes apresentam idade média de 49,6 anos, sendo que cerca de 53% estão com 50 anos ou mais. Além disso, em torno de 55% delas são casadas e aproximadamente 49% foram diagnosticadas no estágio IV (mais avançado) da neoplasia.

A variável resposta a ser considerada será o tempo (em meses) entre o diagnóstico da doença e o evento de interesse, que no caso desse estudo será o óbito das pacientes tendo como causa o câncer ovariano. No contexto de análise de sobrevivência, é comum a inclusão de censuras, ou seja, de informações de pacientes que não apresentaram o evento (óbito). Essas censuras constituem 46,5% do conjunto de dados a ser analisado.

3.1.2 Recursos Computacionais

O *software* R, versão 3.2.2 (R CORE TEAM, 2015) será utilizado para a análise exploratória e o ajuste de modelos aos dados descritos. Alguns pacotes a serem investigados com esse propósito são: *timereg* e *survival*.

Quadro 1 – Descrição de algumas covariáveis disponíveis no banco de dados para as pacientes diagnosticadas com câncer de ovário

COVARIÁVEIS	DESCRIÇÃO
Idade	Idade das pacientes em anos
Estado Civil	Casada; Solteira; Viúva; Outros
Diagnóstico e Tratamento Prévio	1 - Sem diagnóstico e sem tratamento 2 - Com diagnóstico e sem tratamento 3 - Com diagnóstico e com tratamento 4 - Outros
Avaliação da extensão da doença	2 - Localizado 3 - Extensão Direta 4 - Envolvimento de linfonodos regionais 5 - Extensão direta com envolvimento de linfonodos regionais 6 - Metástase 7 - Não aplicável 9 - Ignorado
Base mais importante para o diagnóstico	1 - Exame clínico e/ou Patologia Clínica 2 - Diagnóstico por imagem 3 - Endoscopia 4 - Cirurgia Exploradora/Necrópsia 5 - Citologia ou hematologia 6 - Histologia metástase (AP) 7 - Histologia do TU primário (AP) 9 - Sem informação
Morfologia do tumor primário	Classificação em números de 5 dígitos do tipo histológico da neoplasia
Grau de diferenciação	1 - Bem Diferenciado 2 - Moderadamente Diferenciado 3 - Pouco Diferenciado 4 - Indiferenciado
Tratamento feito na instituição	1 - Nenhum 2 - Cirurgia 3 - Radioterapia 4 - Quimioterapia 5 - Hormonioterapia 6 - TMO > transplante de medula óssea 7 - Imunoterapia 8 - Outros 9 - Sem informação
Estádio Clínico	1 - Tumor limitado aos ovários 2 - Tumor que envolve um ou ambos os ovários e tem extensão para a pelve 3 - Invasão dos peritônio ou dos linfonodos da pelve ou do abdômen 4 - Metástases à distância ou metástases para fígado ou derrame pleural positivo para malignidade

Fonte: Relatório Epidemiológico: 1990 a 2009 – Curitiba: LPCC, 2011.

3.2 Métodos

Para a análise do banco de dados descrito na Seção 3.1.1 pretende-se utilizar metodologias estatísticas no contexto de análise de sobrevivência, na qual a variável resposta é medida pelo tempo decorrido até um determinado evento de interesse. No caso dos dados a serem analisados, a variável resposta se caracteriza pelo tempo decorrido entre o diagnóstico da neoplasia e o óbito de pacientes devido ao câncer de ovário.

Dados de sobrevivência são usualmente caracterizados pela presença de censuras. Censuras são observações parciais da resposta de interesse decorrentes de situações em que, por alguma razão, o acompanhamento foi interrompido (COLOSIMO; GIOLO, 2006). Exemplos de censuras no estudo que será analisado neste trabalho são: óbito das pacientes por outra causa que não o câncer de ovário, finalização do estudo ou interrupção do acompanhamento das pacientes por motivos pessoais.

Um dos interesses na análise de dados de sobrevivência, como os que serão analisados neste trabalho, está em avaliar a probabilidade de uma observação não falhar até certo tempo t ou, equivalentemente, a probabilidade de uma observação sobreviver ao tempo t . Tal probabilidade, caracteriza a função de sobrevivência, representada por:

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

com T a variável aleatória representando o tempo até a ocorrência do evento.

Para a análise descritiva dos dados pretende-se utilizar o estimador de Kaplan-Meier (KAPLAN; MEIER, 1958), também conhecido como estimador limite produto. Este é um método não paramétrico que permite, dentre outras coisas, avaliar a influência de covariáveis (uma a uma) sobre a variável resposta T , obter o tempo médio e mediano e percentis de interesse. O estimador de Kaplan-Meier é definido por:

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

em que t_j são os tempos distintos e ordenados de falha ($t_1 < t_2 < \dots < t_k$), n_j é o número de observações sob risco (não falhou e não foi censurado) até o tempo t_j (exclusive), e d_j é o número de falhas no tempo t_j . A partir das estimativas obtidas para

$S(t)$ serão construídos gráficos, por meio dos quais será possível responder a alguns questionamentos, bem como avaliar a influência de covariáveis sobre o tempo T .

Para testar se há diferença estatística significativa entre as curvas de sobrevivência obtidas para cada uma das categorias de uma covariável, pretende-se utilizar testes tais como: o teste *logrank* (MANTEL, 1966) e o teste de Wilcoxon (GEHAN, 1965).

Para a modelagem pretende-se utilizar, inicialmente, o modelo de Cox (COX, 1972). Este é um modelo semi-paramétrico que tem a propriedade de taxas de falha constante ao longo do tempo (suposição de riscos proporcionais). Em termos de função de sobrevivência, o modelo de Cox fica expresso por:

$$S(t | \mathbf{x}) = [S_0(t)]^{\exp(\mathbf{x}'\boldsymbol{\beta})},$$

em que \mathbf{x} é o vetor de valores observados de p covariáveis, $\boldsymbol{\beta}$ é o vetor que mede o efeito dessas covariáveis e $S_0(t)$ é denominada função de sobrevivência de base.

O modelo de Cox requer a avaliação do pressuposto de taxas de falha proporcionais. Para essa finalidade, serão utilizados gráficos baseados nos resíduos de Schoenfeld (1982). Quanto à seleção das covariáveis que farão parte do modelo, há diversos métodos propostos na literatura para essa finalidade. Aqueles os quais se pretende considerar são: (a) os métodos de seleção *forward*, *backward* e *stepwise*; (b) o método de seleção proposto por Collet (2003) e; (c) o critério de informação de Akaike (AKAIKE, 1973).

Caso o modelo de Cox não apresente um bom ajuste aos dados e/ou a suposição de taxas de falha proporcionais não seja atendida, será necessário considerar modelos alternativos como, por exemplo, extensões do modelo de Cox (que permitem que o efeito das covariáveis variem ao longo do tempo) e o modelo aditivo de Aalen (AALEN, 1989), que permite que tanto os parâmetros quanto as covariáveis variem ao longo do tempo. Ou ainda, modelos de regressão paramétricos como, por exemplo, os modelos de regressão exponencial, Weibull e log-normal (COLOSIMO; GIOLO, 2006).

Para analisar a adequação dos modelos ajustados serão utilizadas técnicas estatísticas propostas na literatura tais como: (1) análise dos resíduos de Cox-Snell e dos resíduos padronizados para examinar o ajuste global do modelo; (2) análise dos resíduos martingal para avaliar a forma funcional de cada covariável no modelo; (3) análise dos resíduos *deviance*, que auxiliam na avaliação de possíveis pontos atípicos que podem estar influenciando a qualidade de ajuste do modelo (COLOSIMO; GIOLO, 2006).

4 CRONOGRAMA DE ATIVIDADES

ATIVIDADES	OUT 2015	NOV 2015	JAN 2016	FEV 2016	MAR 2016	ABR 2016	MAI 2016	JUN 2016
1 Projeto de Pesquisa								
Definição do tema de estudo								
Definição do conjunto de dados e dos métodos estatísticos								
Elaboração e entrega do projeto de pesquisa ao orientador								
Apresentação do projeto de pesquisa								
2 Elaboração do Trabalho de Conclusão de Curso								
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								
3 Defesa do Trabalho de Conclusão de Curso								
Preparação e apresentação do TCC								
4 Elaboração da Versão Final do Trabalho de Conclusão de Curso								
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 lifetimes. **Statistics in Medicine**, v. 8, p. 907-925, 1989.
- AKAIKE, H. Information Theory as an Extension of the Maximum Likelihood Principle. In: SECOND INTERNATIONAL SYMPOSIUM ON INFORMATION THEORY, 1973, Budapest. **Proceedings...** Budapest: Akademiai Kiado, p. 267-281.
- CDC - CENTERS FOR DISEASE CONTROL AND PREVENTION. **Gynecologic cancers**. 2016. Disponível em: <<http://www.cdc.gov/cancer/ovarian/index.htm>>. Acesso em: 15/03/2016.
- COLLET, D. **Modelling survival data in medical research**, 2nd edition. London: Chapman and Hall, 2003. 410 p.
- COLOSIMO, E. A.; GIOLO, S. R. **Análise de sobrevivência aplicada**. São Paulo: Editora Blucher, 2006. 392 p.
- COX, D. R. Regression models and life tables. **Journal Royal Statistical Society, Series B**, v. 34, n. 2, p. 187-220, 1972.
- GEHAN, E. A. A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. **Biometrika**, v. 52, 203-223, 1965.
- HOWLADER, N.; NOONE A. M.; KRAPCHO, M.; GARSHELL, J.; MILLER, D.; ALTEKRUSE, S. F.; KOSARY, C. L.; YU, M.; RUHL, J.; TATALOVICH, Z., MARIOTTO, A.; LEWIS, D. R., CHEN, H. S.; FEUER, E. J.; CRONIN, K. A. (eds). **SEER Cancer Statistics Review, 1975-2012**. National Cancer Institute. Disponível em: <http://seer.cancer.gov/csr/1975_2012/>, based on November 2014 SEER data submission, posted to the SEER web site, April 2015.
- INSTITUTO ONCOGUIA. **Taxa de sobrevivência para o câncer de ovário por estágio**. Disponível em: <<http://www.oncoguia.org.br/conteudo/taxa-de-sobrevivencia-para-o-cancer-de-ovario-por-estagio/6048/229>>. Equipe Oncoguia, 2014. Acesso em: 16/11/2015.
- KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations, **Journal of the American Statistical Association**, v. 53, p. 457-481, 1958.
- LIGA PARANAENSE DE COMBATE AO CÂNCER. **Relatório Epidemiológico: 1990 a 2009**. Curitiba: LPCC, 2011. 124 p. Disponível em: <http://www.erastogaertner.com.br/arquivos/rhc/DuasDecadas_RHC_HEG_1990a2009.pdf>.
- LUIZ, B. M.; MIRANDA, P. F.; MAIA, E. M.; MACHADO, R. B.; GIATTI, M. J.; FILHO, A. A.; BORGES, J. B. Estudo epidemiológico de pacientes com tumor de ovário no município de Jundiaí no período de junho de 2001 a junho de 2006. **Revista Brasileira de Cancerologia**, v. 55, p. 247-253, 2009.

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

R CORE TEAM. **R: A language and environment for statistical computing**. Vienna, Austria, 2015. ISBN 3-900051-07-0. Disponível em: <http://www.R-project.org/>.

SCHOENFELD, D. Partial residuals for the proportional hazard regression model. **Biometrika**, v. 69, 239-241, 1982.

U.S. CANCER STATISTICS WORKING GROUP. **United States Cancer Statistics: 1999–2012 Incidence and Mortality Web-based Report**. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute; 2015. Disponível em: <www.cdc.gov/uscs>.