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

CURITIBA
2016

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

Trabalho de Conclusão de Curso apresentado à disciplina Laboratório B do Curso de Estatística do Setor de Ciências Exatas da Universidade Federal do Paraná, como exigência parcial para obtenção do grau de Bacharel em Estatística.

Orientadora: Profa. Dra. Suely Ruiz Giolo

CURITIBA
2016

AGRADECIMENTOS

Primeiramente a Deus, por nos ter permitido alcançar este objetivo enfrentando todos os obstáculos que surgiram em nossa vida universitária nos dando força, saúde e confiança.

A nossa orientadora Professora Doutora Suely Ruiz Giolo, pela paciência e dedicação que demonstrou em todos os momentos que precisamos e, principalmente, pela forma amiga, doce e gentil com que nos passou seus ensinamentos em todas as nossas reuniões fazendo com que acreditássemos em nossa capacidade.

Ao centro médico que nos forneceu os dados de câncer de ovário possibilitando todas as análises desse estudo.

Aos nossos pais, pelo amor incondicional, apoio e incentivo, nos fortalecendo nos momentos de desânimo e cansaço.

Aos nossos amigos e colegas que fizeram parte de nossa vida acadêmica nos acompanhando nos bons e maus momentos.

Aos professores do Departamento de Estatística pelas aulas e ensinamentos compartilhados durante nosso período acadêmico.

Aos funcionários da Universidade Federal do Paraná, em especial à Sra. Arielza Cruz dos Santos e ao Sr. Alcides Nepomuceno do Laboratório de Estatística.

À Professora Nivea da Silva Matuda pela disponibilidade em participar da banca deste trabalho.

"*Por vezes sentimos que aquilo que fazemos não é senão uma gota de água no mar.*

Mas o mar seria menor se lhe faltasse uma gota ".

(*Madre Tereza de Calcutá*)

RESUMO

O câncer de ovário é uma das doenças que mais levam as mulheres a óbito, sendo um dos motivos a descoberta tardia da neoplasia, devido aos métodos de diagnóstico apresentarem resultados pouco eficazes na descoberta da doença na fase inicial. Em decorrência disso, a doença é detectada, geralmente, quando já está em estágio avançado de malignidade, o que diminui drasticamente as chances de cura. Esse trabalho apresenta um estudo de sobrevida de 383 pacientes diagnosticadas com câncer ovariano, acompanhadas no período de 1994 a 2009, em um centro médico de Curitiba, Paraná. Para análise dos dados foram utilizadas técnicas no contexto de análise de sobrevivência, visando estudar o tempo de sobrevida das pacientes após a constatação do câncer, levando em consideração algumas covariáveis. Para isso, utilizou-se, inicialmente, o estimador de Kaplan-Meier e o teste *logrank* para a análise descritiva das covariáveis. Na sequência foram ajustados os seguintes modelos de regressão: (a) modelo de riscos proporcionais de Cox, (b) modelo de riscos aditivos de Aalen e o de riscos aditivos semiparamétrico proposto por McKeague e Sasieni e, (c) modelo de regressão log-logístico. De modo geral, os modelos apresentaram resultados bastante semelhantes. Foi constatado que: (a) pacientes com idades mais avançadas tiveram uma probabilidade de sobrevida menor quando comparadas a pacientes mais novas; (b) pacientes que deram entrada no hospital entre os anos de 2000 e 2004, apresentaram uma sobrevida maior em relação as que entraram nos demais períodos, contudo não se sabe o real motivo desse resultado, havendo necessidade de uma maior investigação e; (c) pacientes que foram avaliadas com metástase apresentaram, em geral, sobrevida menor.

Palavras-chave: Análise de Sobrevida. Modelo Aditivo de Aalen. Modelo de Cox. Modelo Log-Logístico. Ovário. Riscos Proporcionais.

LISTA DE FIGURAS

LISTA DE TABELAS

Sumário

1 INTRODUÇÃO

Câncer é definido como uma doença degenerativa resultante do acúmulo de lesões no material genético celular, que induz o processo de crescimento, reprodução e dispersão anormal das células (ALBERTS et al., 2004).

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, consequentemente, faz com que o câncer de ovário seja o câncer ginecológico mais letal. Este é um dos fatores que instiga estudos de sobrevida de pacientes com esse tipo de câncer.

As taxas de sobrevida 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 sobrevida. 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 estimação mais precisa e assertiva da sobrevida das pacientes.

Um estudo realizado pelo NCI - *National Cancer Institute*, mostrou haver grande discrepância entre as taxas de sobrevida de pacientes diagnosticadas entre os estádios 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 sobrevida média em cinco anos de pacientes diagnosticadas com câncer de ovário no estádio I da neoplasia foi de 92,1%. Esta taxa decresceu drasticamente para 28,3% para pacientes no estádio 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 mulheres de qualquer faixa etária, contudo a maioria dos casos está concentrada nas mulheres com idade acima dos 50 anos. 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 (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, o foco deste trabalho foi 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.

No geral, o trabalho está estruturado da seguinte forma. No Capítulo 2 é apresentada uma breve revisão sobre estudos estatísticos referentes ao câncer ovariano, encontrados na literatura. No Capítulo 3 faz-se uma breve descrição dos

dados utilizados no estudo, bem como uma apresentação dos métodos estatísticos mais utilizados em análises de sobrevida e que foram utilizados ao longo do estudo. Os resultados obtidos com os ajustes dos modelos aos dados de câncer de ovário estão compreendidos no Capítulo 4, seguido das considerações finais apresentadas no Capítulo 5. Na sequência estão as Referências utilizadas para a realização deste estudo e os Apêndices que complementam algumas das análises realizadas.

2 REVISÃO DE LITERATURA

O câncer de ovário é conhecido como “a doença que sussurra”, pois poucos sintomas específicos estão associados a fase inicial da doença. A detecção precoce da doença ainda é a chave para salvar a vida das mulheres, dado que cerca de 90% delas são curadas apenas pela cirurgia se o câncer for descoberto ainda no estágio I (VANDERHYDEN et al., 2003).

A etiologia do câncer de ovário ainda não é claramente conhecida. Muitos estudos foram realizados para a descoberta de fatores que aumentam ou diminuam as chances de uma mulher contrair o câncer ovariano. Estudos epidemiológicos mostraram que o fator de risco mais importante (depois da idade) é o histórico familiar da doença, representando cerca de 5 a 10% de chances de desenvolvimento da neoplasia (SALEHI et al., 2008).

Apesar dos avanços da terapia oncológica, a taxa de mortalidade de mulheres com câncer ovariano tem se mantido estável nos últimos 30 anos e pouco se avançou em termos de diagnóstico precoce. Cerca de 70% das pacientes são identificadas com a neoplasia quando estão no estágio avançado, devido a sutileza dos sintomas que surgem no início (SALEHI et al., 2008). Estudos apontaram que a taxa de mortalidade para mulheres com a neoplasia é de 70% dentro de dois anos e de 90% dentro de cinco anos (TORRES et al., 2002).

Vários métodos para diagnóstico do câncer de ovário foram relatados, tais como ultrassonografia abdominal e transvaginal, ultrassom tridimensional, ultrassonografia com Doppler colorido e marcadores tumorais. Contudo, nenhum desses métodos mostrou, individualmente, uma performance significativa na detecção da malignidade do tumor.

Torres et al. (2002) mostrou em seu artigo algumas fórmulas matemáticas que foram desenvolvidas para a obtenção do diagnóstico da malignidade do tumor. Usando um modelo logístico e incorporando as variáveis status da menopausa, marcador tumoral CA 125 e constatação da ultrassonografia, tem-se a obtenção de um sistema de escore, que está descrito na literatura e indica os índices de malignidade do tumor.

Um estudo caso-controle realizado no Canadá avaliou o impacto de atividades físicas no risco do câncer de ovário. A base continha dados coletados entre 1994 e

1997 de 442 mulheres com a etiologia e 2.135 mulheres sem, com idade variando de 20 a 76 anos. O estudo mostrou que níveis moderados de atividades físicas reduzem o risco do câncer de ovário (PAN et al., 2005).

Estudos que correlacionam atividade física com o câncer ovariano também foram desenvolvidos na China e na Itália. Em ambos, os resultados foram muito parecidos, concluindo-se que o risco diminui significativamente à medida que as pacientes desenvolvem atividades físicas (PAN et al., 2005).

Ristow et al. (2006) apontaram que a gestação está associada com a diminuição do risco do carcinoma epitelial de ovário. De acordo com os autores, a primeira gestação reduz em aproximadamente 40% o risco desse tipo de câncer e cada gestação adicional à primeira confere uma diminuição de cerca de 14% no risco. Gestações que não são levadas ao termo (por aborto espontâneo ou induzido) não protegem tanto quanto as que realmente são concluídas. Mulheres nulíparas, voluntariamente, constituem um grupo de alto risco para o desenvolvimento do carcinoma epitelial de ovário.

No estudo caso-controle de Riman et al. (2002) que envolveu 1.205 pacientes suecas com câncer de ovário, o histórico familiar mostrou-se um importante fator de risco. O fato de a mãe ou a irmã terem tido câncer ovariano aumentou em aproximadamente 3 vezes a chance do desenvolvimento da doença. E naquelas cujas mãe ou irmã tiveram câncer de mama houve um risco de desenvolvimento de câncer de ovário elevado em 1,35 vezes.

Segundo o artigo publicado por Silva-Filho et al. (2004), um dos princípios do tratamento cirúrgico do câncer ovariano é a citorredução. Pacientes submetidas a uma citorredução satisfatória (quando os tumores residuais não ultrapassam 1 centímetro) têm sobrevida global média de 35 meses. Enquanto que pacientes com doença residual maior que 1 centímetro, sobrevivem, em média, 18 meses. Uma metanálise mostrou que a citorredução satisfatória é determinante para a sobrevida de pacientes que estão nos estágios III e IV da doença.

A disseminação do câncer ovariano não se limita aos órgãos e vísceras pélvicas, portanto os procedimentos cirúrgicos como a citorredução, tornam-se bastante complicados. Dessa forma, é fundamental que o cirurgião tenha domínio técnico que extrapole o conhecimento da cirurgia ginecológica propriamente dita. Ao se tratar de câncer de ovário, a cirurgia é indicada estritamente a pacientes que

tenham uma grande chance de resultado satisfatório, pois há associação desses procedimentos com um aumento da morbidade (SILVA-FILHO et al., 2004).

Pacientes com câncer de ovário geralmente passam por uma progressão da doença ao longo do tempo. Mesmo após a remissão completa do tumor com cirurgia e quimioterapia, cerca de 20-30% das pacientes que estão nos estágios I e II têm recaída, e este número cresce para 75% quando as pacientes estão nos estágios III e IV (CHANG et al., 2015).

No artigo de Salani et al. (2007), foi apresentado um estudo de sobrevida de 55 pacientes que mesmo após a cirurgia citorredutora, tiveram recorrência do câncer ovariano. O estudo mostrou que o tempo mediano de sobrevida após a segunda cirurgia, para todo o grupo estudado, foi de 48 meses, enquanto que a sobrevida mediana após a primeira cirurgia foi de 30 meses. Adicionalmente, pacientes que tiveram a citorredução completa após a segunda cirurgia, tiveram sobrevida mediana de 50 meses, já as pacientes que não apresentaram a remissão completa com a cirurgia, tiveram sobrevida média de 7,2 meses.

Em análise de sobrevida, muitas vezes há interesse em saber se o efeito das covariáveis varia ao longo do tempo, como foi feito neste trabalho. No artigo de Chang et al. (2015), foi estudado se o efeito de algumas covariáveis relacionadas ao prognóstico do câncer de ovário apresentavam efeito tempo-dependente. Os autores chegaram à conclusão que apenas o efeito da covariável “CA125 nadir” não era constante ao longo do tempo.

No Brasil, sem considerar os tumores de pele, o câncer de ovário é o quinto mais incidente na região centro-oeste, o sétimo nas regiões sul, sudeste e nordeste, e o oitavo mais frequente na região norte (INSTITUTO NACIONAL DE CÂNCER JOSÉ DE ALENCAR GOMES DA SILVA, 2014).

3 PACIENTES E MÉTODOS

3.1 Pacientes

3.1.1 Conjunto de Dados

O conjunto de dados que foi utilizado para as análises é oriundo de um centro médico de Curitiba, Paraná, o qual 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 dessa base todas as informações referentes a essa doença, totalizando 383 mulheres que foram acompanhadas durante o período de 1990 a 2009.

A base contém diversas informações (covariáveis) que foram analisadas e classificadas como relevantes ou não para o estudo em questão, tais como: idade, ano de entrada no estudo, estado civil, tratamento feito na instituição, dentre outras. Na Tabela 1 (Capítulo 4) estão apresentadas as covariáveis que foram utilizadas nas análises descritivas com seus respectivos percentuais de falhas e censuras e suas frequências absolutas. As pacientes apresentaram idade média de 50 anos (desvio-padrão=15,2), sendo que aproximadamente 54% delas apresentaram idade igual ou superior a 50 anos e aproximadamente 55% delas eram casadas. Além disso, 49% das pacientes foram diagnosticadas no estádio IV (mais avançado) da neoplasia.

Quanto à variável resposta foi utilizado o tempo (em meses) entre o diagnóstico da doença e o evento de interesse, que no caso desse estudo é o óbito das pacientes tendo como causa o câncer ovariano. No contexto de análise de sobrevivência é comum a ocorrência de censuras, que é a não observação do tempo de falha, ou seja, tem-se apenas tempos parciais da resposta de interesse decorrente de situações em que, por alguma razão, o acompanhamento foi interrompido (COLOSIMO; GIOLO, 2006). Essas censuras constituem 46,74% do conjunto de dados.

3.1.2 Recursos Computacionais

O *software* R, versão 3.2.2 (R CORE TEAM, 2015) foi utilizado para a análise exploratória e o ajuste dos modelos aos dados descritos, por meio de pacotes como *timereg* e *survival* que estão disponíveis no *software*.

3.2 Métodos

Em análise de sobrevivência, a variável resposta é, geralmente, medida pelo tempo decorrido até um determinado evento de interesse. Esse tempo é denominado tempo de falha (COLOSIMO; GIOLO, 2006). No caso dos dados desse estudo, 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 decorrente de situações em que, por alguma razão, o acompanhamento foi interrompido (COLOSIMO; GIOLO, 2006). Isto significa que o tempo de falha, ou óbito, nesses tipos de dados é superior àquele observado. Exemplos de censuras nesse estudo são: óbito das pacientes por outra causa que não seja propriamente o câncer de ovário, finalização do estudo ou interrupção do acompanhamento das pacientes por motivos pessoais.

Um dos principais interesses na análise de dados de sobrevida, como é o caso dos dados do estudo descrito, 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:

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

sendo T a variável aleatória não-negativa representando o tempo até a ocorrência do evento de interesse.

Uma função que costuma ser mais informativa que a função de sobrevivência, é a função taxa de falha, pois diferentes funções de sobrevivência podem apresentar formas semelhantes, enquanto suas respectivas funções taxa de falha podem diferir drasticamente (COLOSIMO; GIOLO, 2006). A função taxa de falha $\lambda(t)$ é bastante útil

para descrever a distribuição do tempo de vida de pacientes. A taxa de falha no intervalo $[t, t + \Delta t)$ é definida como a probabilidade de que a falha ocorra neste intervalo, condicional à sobrevivência até o tempo t, dividida pelo comprimento do intervalo. A função $\lambda(t)$ é descrita por:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \,|\, T \geq t)}{\Delta t} .$$

Para a estimação da função de sobrevivência $S(t)$, acomodando as censuras presentes no conjunto de dados, foi utilizado neste trabalho o estimador não-paramétrico de Kaplan-Meier, apresentado a seguir.

3.2.1 Estimador de Kaplan-Meier

O estimador não-paramétrico proposto por Kaplan e Meier em 1958 é também conhecido como estimador limite-produto e é usado para estimar a função de sobrevivência. Ele é, indubitavelmente, o mais utilizado em estudos clínicos e ganha cada vez mais espaço nos estudos de confiabilidade. É um estimador não paramétrico justamente por não assumir nenhuma distribuição probabilística para a variável resposta.

O estimador de Kaplan-Meier consiste em dividir o tempo de seguimento em tantos intervalos quantos forem o número de falhas distintas. Os limites dos intervalos de tempo são os tempos de falha. A expressão do estimador é definida como:

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

em que t_j, com j = 1,...,k, são os *k* tempos distintos e ordenados de falha ($t_1 < t_2 < \cdots < t_k$), 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 a partir do estimador de Kaplan-Meier pode-se construir gráficos das curvas de sobrevivência, por meio dos quais é possível responder a alguns questionamentos, bem como avaliar a influência das covariáveis ao longo do tempo.

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 $S(t)$.

Neste trabalho, o estimador de Kaplan-Meier foi utilizado para estimar a probabilidade de sobrevida das pacientes em função das categorias de cada uma das covariáveis apresentadas na Tabela 1 (Seção 4).

3.2.2 Teste *logrank*

Proposto por Mantel (1966), o teste *logrank* é o mais utilizado em análise de sobrevivência para a comparação das curvas de sobrevida de grupos de indivíduos a fim de investigar a existência de diferenças significativas entre elas. O teste é particularmente apropriado quando a razão de taxas de falha dos grupos a serem comparados é aproximadamente constante ou, equivalentemente, as populações têm a propriedade de taxas de falha proporcionais.

O teste consiste na diferença entre o número observado de falhas em cada grupo e o número esperado de falhas sob a hipótese nula, a qual é a hipótese de igualdade das curvas de sobrevivência entre os grupos a serem analisados, isto é, $H_0: S_1(t) = S_2(t)$.

Para o teste de igualdade entre duas funções de sobrevivência $S_1(t)$ e $S_2(t)$, considere $t_1 < t_2 < \cdots < t_k$ os tempos distintos de falhas da amostra formada pela combinação das duas amostras individuais. Suponha que no tempo t_j aconteçam d_j falhas e que n_j indivíduos estejam sob risco em um tempo imediatamente inferior a t_j na amostra combinada e, respectivamente d_{ij} e n_{ij} na amostra i, com $i = 1, 2$ e $j = 1, \ldots, k$. Sabendo que d_{2j} segue uma distribuição hipergeométrica, sua média é dada por $w_{2j} = n_{2j} d_j n_j^{-1}$ e sua variância por $(V_j)_2 = n_{2j}(n_j - n_{2j}) d_j (n_j - d_j) n_j^{-2} (n_j - 1)^{-1}$. Sendo k a quantidade de tempos de falha, a estatística proposta segue uma distribuição qui-quadrado com 1 grau de liberdade e é dada por:

$$T = \frac{\left[\sum_{j=1}^{k} (d_{2j} - w_{2j})\right]^2}{\sum_{j=1}^{k} (V_j)_2} .$$

3.2.3 Modelo de Regressão de Cox

Devido a sua versatilidade, o modelo de Cox (COX, 1972) é o mais utilizado em estudos clínicos para a análise de dados de sobrevida. Foi proposto por David Cox em 1972 abrindo uma nova fase na modelagem de dados clínicos (COLOSIMO; GIOLO, 2006).

Considerando $\boldsymbol{x}$ um vetor com os valores de p covariáveis, $\boldsymbol{x} = (x_1, \dots, x_p)'$, tem-se a expressão geral do modelo de Cox como:

$$\lambda(t|\boldsymbol{x}) = \lambda_0(t) g(\boldsymbol{x}'\boldsymbol{\beta}),$$

em que o componente paramétrico $g(\boldsymbol{x}'\boldsymbol{\beta})$ é uma função não-negativa que deve ser especificada, de modo que $g(\mathbf{0}) = 1$, e o componente não-paramétrico $\lambda_0(t)$ é uma função não-negativa do tempo, o qual é usualmente denominado função taxa de falha de base. O componente paramétrico é frequentemente utilizado da forma:

$$g(\boldsymbol{x}'\boldsymbol{\beta}) = \exp\{\boldsymbol{x}'\boldsymbol{\beta}\} = \exp\{\beta_1 x_1 + \cdots + \beta_p x_p\},$$

em que $\boldsymbol{\beta}$ é o vetor de parâmetros associados às p covariáveis. Por ser composto de um componente não-paramétrico e outro paramétrico, o modelo de Cox é denominado modelo semiparamétrico.

A suposição básica para o devido uso do modelo de Cox é que as taxas de falhas sejam proporcionais, o que faz o modelo ser comumente denominado modelo de riscos proporcionais. Isto significa que a razão das funções de taxa de falha de dois indivíduos observados no início do estudo deve ser a mesma durante todo o período de acompanhamento. Desse modo, essa razão para dois indivíduos i e j é dada por:

$$\frac{\lambda(t|\boldsymbol{x}_i)}{\lambda(t|\boldsymbol{x}_j)} = \frac{\lambda_0(t)\exp(\boldsymbol{x}_i{}'\boldsymbol{\beta})}{\lambda_0(t)\exp(\boldsymbol{x}_j{}'\boldsymbol{\beta})} = \exp\{\boldsymbol{x}_i{}'\boldsymbol{\beta} - \boldsymbol{x}_j{}'\boldsymbol{\beta}\}.$$

Os métodos de estimação são necessários para a inferência acerca dos parâmetros dos modelos. No caso do modelo de Cox, os coeficientes $\beta's$ medem os efeitos das covariáveis sobre a função taxa de falha. Por conter um componente não-paramétrico no modelo, Cox propôs um método de estimação (COX, 1975)

denominado de máxima verossimilhança parcial, de maneira que, para uma amostra de n indivíduos, com $k \leq n$ falhas distintas nos tempos $\mathrm{t}_1 < \mathrm{t}_2 < \cdots < \mathrm{t}_k$, a respectiva função de verossimilhança parcial fica expressa por:

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

em que δ_i é o indicador de falha, sendo 1 quando o tempo do indivíduo i corresponde a uma falha e 0 quando censura, e $R(t_i)$ é o conjunto dos índices das observações sob risco no tempo t_i. A maximização dessa função fornece os estimadores do vetor $\boldsymbol{\beta}$, ou seja, $\widehat{\boldsymbol{\beta}}$.

A fim de estimar a função taxa de falha acumulada de base, Breslow (1972) propôs um estimador não paramétrico dado por:

$$\widehat{\Lambda}_0(t) = \sum_{j:\, t_j < t} \frac{d_j}{\sum_{l \in R(t_j)} \exp\{\boldsymbol{x}'_l \widehat{\boldsymbol{\beta}}\}} , \qquad (1)$$

com d_j o número de falhas em t_j e $R(t_j)$ o conjunto de indivíduos sob risco em t_j.

Em consequência do estimador da taxa de falha acumulada de base proposto por Breslow (1972), apresentado na equação (1), segue que as funções de sobrevivência $S_0(t)$ e $S(t|\boldsymbol{x})$ podem ser estimadas a partir dos estimadores:

$$\hat{S}_0(t) = \exp\left\{-\widehat{\Lambda}_0(t)\right\}$$

e

$$\hat{S}(t|\boldsymbol{x}) = \left[\hat{S}_0(t)\right]^{\exp\{\boldsymbol{x}'\widehat{\boldsymbol{\beta}}\}} .$$

Tanto $\hat{S}_0(t)$ quanto $\hat{S}(t|\boldsymbol{x})$ são funções escada decrescentes com o tempo.

Com o propósito de verificar o pressuposto de taxas de falha proporcionais no modelo de Cox, foram utilizados três métodos propostos na literatura. O primeiro é um método gráfico descritivo que consiste em dividir os dados em m estratos, que correspondem às categorias de cada covariável (sendo necessário categorizar as variáveis contínuas), e estimar a função taxa de falha acumulada de base $\widehat{\Lambda}_{0j}(t)$ para

cada estrato j (j = 1, ..., m). O estimador proposto por Breslow (1972) para estimar tal função para cada estrato j é dado por:

$$\widehat{\Lambda}_{0j}(t) = \sum_{j:\, t_j<t} \frac{d_j}{\sum_{l\in R_j} \exp\{\boldsymbol{x}'_l \widehat{\boldsymbol{\beta}}\}} \ .$$

Em seguida, são obtidos para cada covariável os gráficos de $\log(\widehat{\Lambda}_{0j}(t))$ *versus* t. Se a suposição for válida, as curvas devem apresentar diferenças aproximadamente constantes ao longo do tempo (COLOSIMO; GIOLO, 2006).

O segundo método também consiste em um método gráfico o qual faz uso dos resíduos padronizados de Schoenfeld, definidos por Schoenfeld (1982). Denotando esses resíduos por s_{iq}, para $i = 1, \dots, d$ e $q = 1, \dots, p$, com d sendo o número de falhas, Grambsch e Therneau (1994) sugeriram o gráfico de $s_{iq} + \hat{\beta}_q$versus *t* que deve apresentar uma linha horizontal ou, em outras palavras, apresentar inclinação nula caso a suposição de taxas de falha proporcionais seja válida.

Adicional aos dois métodos gráficos mencionados, os quais muitas vezes são subjetivos, um terceiro método utilizado foi a obtenção do coeficiente de correlação de Pearson (ρ) entre os resíduos padronizados de Schoenfeld e t, para cada covariável. Valores de ρ próximos de zero mostram não haver evidências para a rejeição da suposição de riscos proporcionais (COLOSIMO; GIOLO, 2006).

Uma limitação do modelo de Cox é a de que ele não é adequado para verificar mudanças dos efeitos das covariáveis ao longo do tempo. Uma alternativa para essa situação foi a utilização dos modelos de riscos aditivos de Aalen e de riscos aditivos semiparamétrico proposto por McKeague e Sasieni, apresentados a seguir.

3.2.4 Modelo de Riscos Aditivos

Um modelo alternativo ao de Cox que tem por finalidade estimar a função taxa de falha na presença de covariáveis e dados censurados foi proposto por Aalen (1980). Esse modelo permite que tanto os parâmetros quanto as covariáveis variem ao longo do tempo, o que o torna uma forte alternativa caso a suposição de riscos proporcionais assumida para o modelo de Cox seja violada.

Adotando as notações apresentadas anteriormente e considerando um vetor de covariáveis $\boldsymbol{x}_i(t) = (1, x_{i1}(t), x_{i2}(t), \dots, x_{ip}(t))'$, possivelmente dependentes do tempo, para $i = 1, \dots, n$, com n o número de indivíduos e p o número de covariáveis,

segue que a função taxa de falha $\lambda\big(t|\boldsymbol{x}_i(t)\big)$ para o indivíduo i no tempo t fica expressa de acordo com o modelo aditivo de Aalen por:

$$\lambda\big(t|\boldsymbol{x}_i(t)\big) = \beta_0(t) + \sum_{j=1}^{p} \beta_j(t)\, x_{ij}(t),$$

ou, na forma matricial, por:

$$\lambda\big(t|\boldsymbol{x}(t)\big) = \boldsymbol{X}(t)\boldsymbol{\beta}(t),$$

com $\boldsymbol{\beta}(t) = (\beta_0(t), \beta_1(t), \dots, \beta_p(t))'$ um vetor de funções do tempo desconhecido, e a matriz $\boldsymbol{X}(t)$ de ordem $n \times (p+1)$ é definida da seguinte maneira: se o evento considerado ainda não ocorreu para o *i*-ésimo indivíduo e ele não é censurado, então, a *i*-ésima linha de $\boldsymbol{X}(t)$ é o vetor $\boldsymbol{x}_i(t) = (1, x_{i1}(t), x_{i2}(t), \dots, x_{ip}(t))'$. Caso contrário, ou seja, se o indivíduo não está sob risco no tempo *t*, então, a *i*-ésima linha de $\boldsymbol{X}(t)$ contém apenas zeros. O primeiro elemento desse modelo, $\beta_0(t)$, é interpretado como a taxa de falha de base, enquanto os $\beta_j(t), j = 1, \dots, p$, são funções de regressão que medem a influência das respectivas covariáveis, as quais atuam de maneira aditiva na função taxa de falha.

O modelo aditivo de Aalen é considerado não-paramétrico pelo fato de nenhuma forma paramétrica ser assumida para as funções $\beta_j(t)$, nenhuma distribuição ser especificada para a variável resposta T e a estimação ser realizada utilizando apenas informação local (COLOSIMO; GIOLO, 2006).

A estimação direta dos $\beta_j(t), j = 0, 1, \dots, p,$ é bastante complicada, pois nenhuma forma paramétrica é assumida para os betas. Aalen propôs, então, um estimador baseado na técnica de mínimos quadrados para o vetor $\mathbf{B}(t)$ com os elementos $\mathrm{B}_j(t), j = 0, 1, \dots, p$, correspondendo às funções de regressão acumuladas. A estimação não é, em geral, possível para todo o tempo de acompanhamento, ficando restrita a um tempo máximo denotado por τ. Detalhes sobre o estimador de mínimos quadrados de Aalen podem ser encontrados em Aalen (1989).

Nota-se que quando todas as covariáveis forem fixadas em $t = 0$, os estimadores para a taxa de falha acumulada e para a função de sobrevivência ficam expressos, respectivamente, por:

$$\widehat{\Lambda}(t|\boldsymbol{x}) = \boldsymbol{x}'\widehat{\boldsymbol{B}}(t) = \widehat{\mathrm{B}}_0(t) + \sum_{j=1}^{p} \widehat{\mathrm{B}}_j(t)\, x_j$$

e

$$\hat{S}(t|\boldsymbol{x}) = \exp\{-\widehat{\Lambda}(t|\boldsymbol{x})\} \ .$$

Para analisar o comportamento das covariáveis ao longo do tempo foram construídos gráficos dos coeficientes de regressão acumulados $\widehat{\mathrm{B}}_j(t)$, $j = 0, 1, \ldots, p$, versus o tempo. Observando a inclinação desses coeficientes é possível obter informações sobre a influência de cada covariável, bem como verificar se uma particular covariável tem efeito constante ou variando com o tempo ao longo do estudo (COLOSIMO; GIOLO, 2006). Inclinações aproximadamente iguais a zero serão observadas em períodos em que a covariável não tem efeito sobre a taxa de falha. Inclinações positivas indicam efeito crescente sobre a taxa de falha, ao passo que inclinações negativas exercem efeito decrescente.

Para testar (a) se o efeito da covariável é significativo, isto é, H_{01}: $\beta_q(t) = 0$, para $q = 1, .., p,\ t \leq \tau$, e (b) se o efeito da covariável muda ao longo do tempo, isto é, H_{02}: $\beta_q(t) = \beta_q,\ q = 1, .., p,\ t \leq \tau$, Martinussen e Scheike (2006) propuseram estatísticas de teste cuja distribuição assintótica são obtidas via procedimentos de reamostragem. Detalhes podem ser encontrados na referência mencionada. Essas estatísticas foram utilizadas neste trabalho para testar as hipóteses citadas. Ainda, gráficos propostos por Martinussen e Scheike (2006) que mostram 50 processos escores simulados sob a hipótese nula H_2: $\beta_q(t) = \beta_q$, para $q = 1, .., p$ e $t \leq \tau$, juntamente com o processo escore observado versus o tempo, também foram investigados com o objetivo de avaliar se o efeito de cada covariável muda ou não ao longo do tempo.

Também foi utilizado neste estudo uma extensão do modelo de riscos aditivos de Aalen denominado modelo de riscos aditivos semiparamétrico, o qual permite que parte dos coeficientes variem com o tempo e parte não. Tal modelo foi proposto por McKeague e Sasieni (1994) e tem sua função taxa de falha escrita como:

$$\lambda\left(t\middle|\boldsymbol{x}_i(t),\boldsymbol{z}_i(t)\right)=\beta_0(t)+\boldsymbol{x}'_i(t)\boldsymbol{\beta}(t)+\boldsymbol{z}'_i(t)\boldsymbol{\gamma}\ ,$$

em que $\boldsymbol{x}_i(t)=(x_{i1}(t),\ldots,x_{ir}(t))$ e $\boldsymbol{z}_i(t)=(z_{i1}(t),\ldots,z_{il}(t))$ são os vetores de covariáveis de dimensão *r* e *l*, respectivamente, observados no tempo $t\,\epsilon\,[0,\tau]$, $\boldsymbol{\beta}(t)=(\beta_1(t),\ldots,\beta_r(t))$ e $\boldsymbol{\gamma}=(\gamma_1,\ldots,\gamma_r)$ são, respectivamente, os vetores dos coeficientes tempo-dependentes e tempo-independentes.

Adicional à utilização do modelo semiparamétrico de Cox e de ambos os modelos de riscos aditivos para a análise dos dados descritos na Seção 3.1.1, também foi utilizado o modelo de regressão paramétrico log-logístico, apresentado a seguir.

3.2.5 Modelo de regressão log-logístico

Modelos de regressão paramétricos são mais amplamente utilizados para descrever os tempos de vida de produtos industriais do que os de estudos clínicos. Isso se deve a uma maior facilidade em encontrar uma distribuição para os tempos de vida de produtos industriais devido aos fatores que afetam seus tempos serem mais facilmente controlados do que os que afetam indivíduos que vivem livremente e não sob o controle do pesquisador.

Os modelos de regressão que ocupam uma posição de destaque em dados de sobrevida são: o exponencial, o de Weibull, o log-normal e o log-logístico. Neste trabalho, foi utilizado o modelo de regressão log-logístico por ter sido, dentre os citados, o que melhor se adequou aos dados.

A distribuição log-logística foi proposta por Tadikamalla e Johnson (1982), sendo que, para uma variável aleatória *T*, sua respectiva função densidade de probabilidade, na presença de um vetor de covariáveis $\boldsymbol{x}=\left(x_1,\ldots,x_p\right)'$, é dada por:

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

sendo $\alpha(\boldsymbol{x})=\exp\{\boldsymbol{x}'\boldsymbol{\beta}\}$ e $\gamma>0$ os parâmetros de escala e forma, respectivamente. As funções de sobrevivência e de taxa de falha são expressas, respectivamente, por:

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

e

$$\lambda(t \mid \boldsymbol{x}) = \frac{\gamma(t/\alpha(\boldsymbol{x}))^{\gamma-1}}{\alpha(\boldsymbol{x})[1+(t/\alpha(\boldsymbol{x}))^{\gamma}]} \quad .$$

A estimação dos parâmetros do modelo log-logístico se deu pelo método de máxima verossimilhança no contexto de censuras à direita e aleatórias, ou seja, as censuram ocorreram por várias razões durante o estudo (por isso aleatórias) e sabe-se que o tempo de ocorrência do evento, neste caso o óbito, é maior que o tempo indicado pela censura (por isso censura à direita). Este método de estimação é amplamente utilizado com modelos paramétricos, podendo ser facilmente encontrado na literatura, bem como em Colosimo e Giolo (2006).

A escolha do modelo log-logístico se deu por meio de visualização gráfica das curvas de sobrevivência estimadas para cada um dos modelos (exponencial, Weibull, log-normal e log-logístico), na ausência de covariáveis, com a curva de sobrevivência estimada pelo método de Kaplan-Meier, *versus* o tempo.

Na seção a seguir são apresentados os resíduos de Cox-Snell, utilizados para a avaliação da qualidade de ajuste dos modelos de Cox, de riscos aditivos e log-logístico.

3.2.6 Métodos de adequação global e diagnóstico de observações atípicas

A avaliação da adequação dos modelos ajustados aos dados foi realizada por meio dos resíduos de Cox-Snell (1968), martingal e *deviance*. O primeiro foi utilizado para a verificação da qualidade do ajuste e os outros dois para a verificação de potenciais pontos atípicos ou influentes.

Os resíduos de Cox-Snell são quantidades definidas, para $i = 1, \ldots, n$, por $\hat{e}_i = \widehat{\Lambda}(t_i|\boldsymbol{x}_i)$, de modo que para os modelos considerados tem-se:

$$\hat{e}_i = \begin{cases} \widehat{\Lambda}_0(t_i)\exp\left(\boldsymbol{x}'_i\widehat{\boldsymbol{\beta}}\right) & \text{Cox} \\ \widehat{\mathrm{B}}_0(t_i) + \boldsymbol{x}'_i\widehat{\mathbf{B}}(t_i) & \text{Aalen} \\ \widehat{\mathrm{B}}_0(t_i) + \boldsymbol{x}'_i\widehat{\mathbf{B}}(t_i) + \boldsymbol{z}'_i\widehat{\boldsymbol{\gamma}}t_i & \text{Mackeague e Sasieni} \\ -\ln\left(\dfrac{1}{1+\left(\frac{t}{\alpha(\boldsymbol{x})}\right)^{\gamma}}\right) & \log-\text{logístico.} \end{cases}$$

Para que o ajuste dos modelos seja considerado satisfatório, tais resíduos devem seguir uma distribuição exponencial padrão, como explica Lawless (2003).

Para a análise de adequação dos modelos foram utilizadas técnicas gráficas baseadas nos resíduos de Cox-Snell. Uma delas, o da função taxa de falha acumulada estimada $\widehat{\Lambda}(\hat{e}_i)$ *versus* os resíduos de Cox-Snell $\hat{e}_i$ sugere um bom ajuste caso a curva apresentada siga aproximadamente uma reta a partir da origem com inclinação 1. Outra alternativa foi a representação da curva de sobrevivência dos resíduos estimada por Kaplan-Meier $\hat{S}_{KM}(\hat{e}_i)$ *versus* $\hat{e}_i$ e a curva de sobrevivência dos resíduos sob a Exponencial padrão $\hat{S}_{Exp}(\hat{e}_i)$ *versus* $\hat{e}_i$ em um mesmo gráfico. Quanto mais próximas estiverem estas curvas, mais evidências a favor do modelo ajustado. Finalmente, foi avaliado um terceiro gráfico com os pares de pontos ($\hat{S}_{KM}(\hat{e}_i)$, $\hat{S}_{Exp}(\hat{e}_i)$) que, similar ao primeiro, mostra um ajuste adequado se os pontos seguirem aproximadamente uma reta a partir da origem com inclinação 1.

Os resíduos martingal e *deviance* foram utilizados para verificar a presença de potenciais pontos atípicos (*outliers*) que podem estar influenciando na qualidade do ajuste do modelo. São definidos, para $i = 1, \dots, n,$ respectivamente, por:

$$\widehat{m}_i = \delta_i - \hat{e}_i$$

e

$$\hat{d}_i = sinal(\widehat{m}_i)[-2(\widehat{m}_i + \delta_i \log(\delta_i - \widehat{m}_i)]^{1/2} \ .$$

Foram construídos gráficos dos resíduos $\widehat{m}_i$ e $\hat{d}_i$ *versus* o preditor linear sabendo que a ausência de pontos atípicos é evidenciada se os resíduos estiverem distribuídos aleatoriamente em torno de zero.

3.2.7 Qualidade de predição dos modelos

Para avaliar a qualidade de predição dos modelos apresentados nas seções anteriores, bem como compará-los entre si, mesmo apresentando estruturas distintas (multiplicativa e aditiva), foi utilizada uma versão tempo-dependente da área sob a curva ROC proposta por Heagerty e Zheng (2005) para o modelo de Cox, a qual foi estendida para os demais modelos por Raminelli (2015). De acordo com os autores mencionados, a qualidade de predição é avaliada em tempos específicos por meio da área sob a curva ROC em cada tempo t, denotada por $AUC(t)$.

A curva ROC é representada pelos pares de pontos $(1 - \text{esp}(c,t),\ \text{sens}(c,t))$ sendo que $\text{esp}(c,t)$ corresponde à especificidade e $\text{sens}(c,t)$ à sensibilidade, ambas tempo-dependentes, e definidas por:

$$\text{sens}(c,t) = P(M_i(t) > c \mid T_i = t) = P(M_i(t) > c \mid \delta_i(t) = 1)$$

$$\text{esp}(c,t) = P(M_i(t) \leq c \mid T_i > t) = P(M_i(t) \leq c \mid \delta_i(t) = 0),$$

com $M_i(t),\ i = 1, \ldots, n$, um marcador tempo-dependente utilizado para a previsão de falha no tempo t e $c \in \mathbb{R}$ um ponto de corte utilizado como critério para classificar a previsão como falha ou censura no tempo *t* (RAMINELLI, 2015). Foram utilizados dois métodos para a estimação da curva ROC(*t*). O primeiro se baseia no teorema de Bayes e no estimador de Kaplan-Meier (KAPLAN; MEIER, 1958) e o segundo no estimador do vizinho mais próximo (AKRITAS, 1994), denotado por NNE. Quanto mais próximo do valor 1 estiver a AUC(*t*), melhor é a capacidade de discriminação do modelo. Para a obtenção das estimativas dos erros padrão associados às AUC(*t*) foram realizadas 600 simulações utilizando o método de reamostragem *bootstrap* (EFRON, 1982). Mais detalhes podem ser encontrados em Raminelli (2015).

4 RESULTADOS E DISCUSSÃO

4.1 Análise Exploratória

Inicialmente foi realizada uma análise exploratória do conjunto de dados apresentado na Seção 3.1.1. Tem-se, assim, na Tabela 1 as frequências absolutas e as porcentagens de pacientes em cada categoria das covariáveis, bem como os percentuais de falhas e censuras.

Tabela 1 – Frequências absolutas e respectivos percentuais de pacientes, censuras e falhas referentes aos dados com 383 pacientes com câncer de ovário

Covariável	Categoria	N	N (%)	Censura	Falha	Censura (%)	Falha (%)
Idade da Paciente	< 40	86	22%	61	25	71%	29%
	40-49	90	23%	44	46	49%	51%
	50-59	85	22%	34	51	40%	60%
	60-88	122	32%	40	82	33%	67%
Estado Civil	Casada	212	55%	99	113	47%	53%
	Solteira	87	23%	46	41	53%	47%
	Viúva	62	16%	23	39	37%	63%
	Outros	22	6%	11	11	50%	50%
Ano de entrada no Hospital	1990-1994	127	33%	57	70	45%	55%
	1995-1999	162	42%	65	97	40%	60%
	2000-2004	94	25%	57	37	61%	39%
Tratamento Prévio	0 : Sem tratamento	207	54%	84	123	41%	59%
	1 : Com tratamento	176	46%	95	81	54%	46%
Avaliação e Extensão da Doença	1 : Localizado	85	22%	68	17	80%	20%
	2 : Extensão direta	75	20%	46	29	61%	39%
	3 : Linfonodos regionais	30	8%	8	22	27%	73%
	4 : Metástase	186	49%	53	133	28%	72%
	99 : Sem informação	7	2%	4	3	57%	43%
Diferenciação do Tumor	0 : Não diferenciado	357	93%	159	198	45%	55%
	1 : Bem diferenciado	26	7%	20	6	77%	23%
Estádio da doença	1 : Limitado aos ovários	83	22%	68	15	82%	18%
	2 : Extensão para pelve	19	5%	15	4	79%	21%
	3 : Metástase na superfície	116	30%	37	79	32%	68%
	4 : Metástase à distância	62	16%	16	46	26%	74%
	99 : Sem informação	103	27%	43	60	42%	58%
Tratamento Feito na Instituição	1 : Cirurgia	102	27%	59	43	58%	42%
	2 : Radioterapia	12	3%	3	9	25%	75%
	3 : Quimioterapia	85	22%	35	50	41%	59%
	4 : Cirurgia+Quimioterapia	165	43%	72	93	44%	56%
	5 : Cirurgia+Outro	19	5%	10	9	53%	47%

Fonte: Os autores (2016).

Ao observar a covariável idade, que foi categorizada em quatro classes de idade, nota-se que o percentual de falhas (óbito da paciente devido ao câncer de ovário) aumentou progressivamente com a idade das pacientes e que a última classe com as idades mais avançadas apresentou o maior número de pacientes. A paciente mais nova do estudo estava com 16 anos, enquanto a mais velha tinha 88 anos.

Ao analisar o estado civil das pacientes, observou-se que 212 mulheres eram casadas, representando 55% das pacientes. Notou-se, também, que os percentuais de falhas e censuras não diferiram muito entre as classes dessa covariável.

O ano de entrada no centro médico foi categorizado em três segmentos. Para as que ingressaram entre 1990 e 1994 (42% da base) e entre 1995 e 1999 (33% da base), as falhas foram superiores às censuras. Contudo, para as que deram entrada entre 2000 e 2004, o percentual de censuras foi mais elevado. Talvez tenha ocorrido alguma melhora nos medicamentos que explique o menor número de falhas durante o último período de entrada, ou pode ser que este seja tão somente consequência do menor tempo de seguimento das pacientes que entraram neste período.

A covariável tratamento prévio indica se a paciente realizou algum tratamento para combater o câncer de ovário antes de dar entrada no centro médico. Observou-se que 59% das 207 pacientes que não obtiveram tratamento prévio foram a óbito, e este número se reduz a 46% quando algum tratamento prévio foi realizado.

Avaliação e extensão da doença e estádio da doença são duas covariáveis que relatam informações extremamente parecidas, dado que avaliar o estadiamento de uma neoplasia é praticamente o mesmo que avaliar o seu grau de disseminação. A partir da Tabela 1, nota-se a ausência de informação sobre a extensão da doença para apenas 2% das pacientes, enquanto para o estadiamento para 27% delas. Para as duas covariáveis citadas, suas quatro categorias estão relacionadas com a classificação do câncer de ovário na paciente, sendo que a primeira categoria expressa uma condição melhor da paciente do que a última. Desta forma, é compreensível que o número de falhas (óbitos) seja menor para a primeira categoria, aumentando gradativamente para as categorias subsequentes. Pode-se notar, para a covariável avaliação e extensão da doença, que o percentual de falhas registrado na primeira categoria foi de 20% contra 72% na última categoria. Para a covariável estádio esses percentuais foram muito semelhantes, 18% contra 74%. Como essas

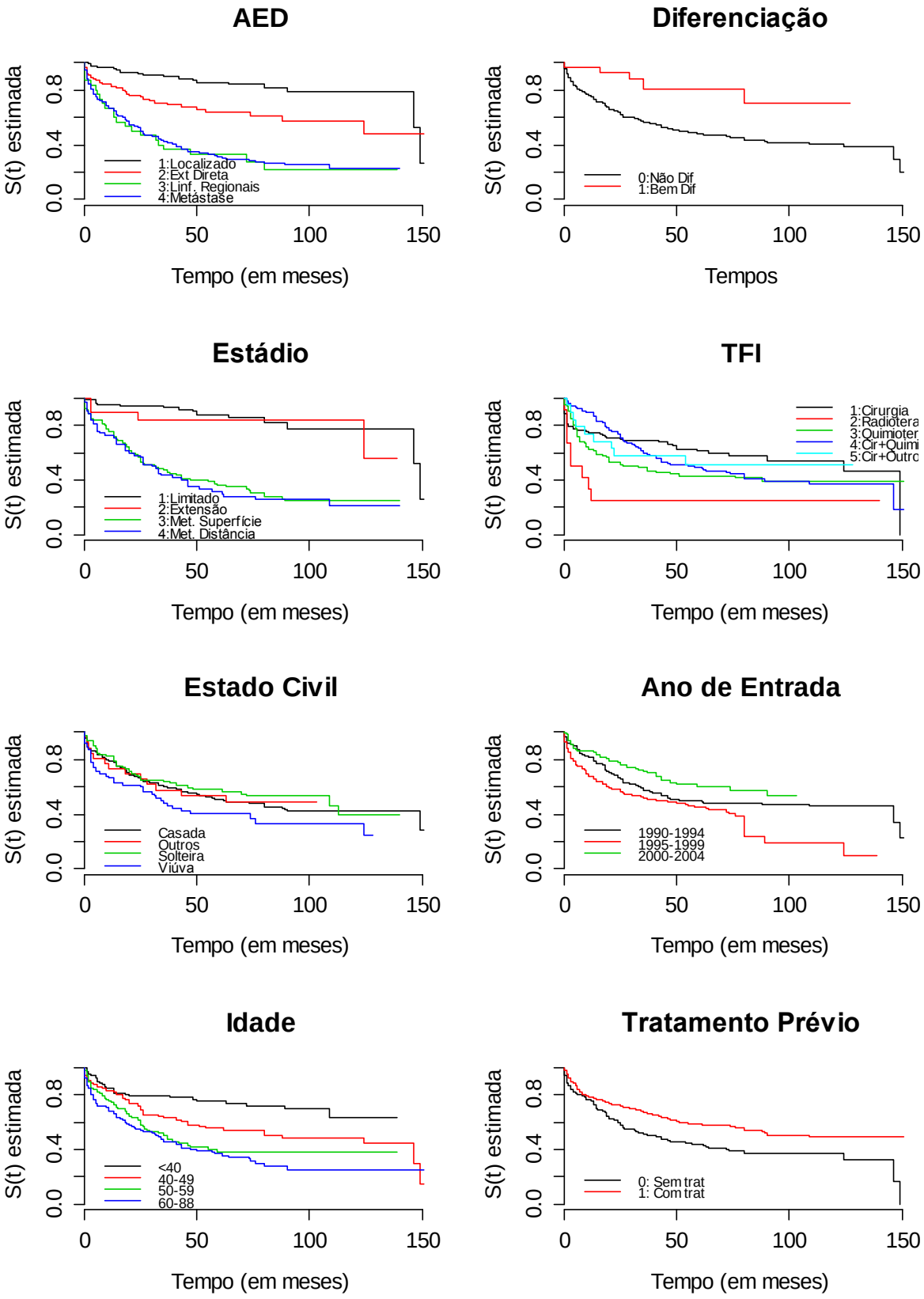
duas covariáveis expressam informações muito similares, foi avaliada a correlação entre elas, apresentada mais adiante.

O grau de diferenciação do tumor é uma covariável que mede o quanto as células neoplásicas (células atingidas pelo tumor) assemelham-se às células normais. As neoplasias benignas geralmente apresentam células bem diferenciadas, ou seja, células que são bastante semelhantes às células de origem. Na base de dados desse estudo, 93% das pacientes apresentaram grau não diferenciado, sendo que 55% delas foram a óbito. Dado que foi observado um percentual muito elevado em apenas uma das categorias desta covariável, optou-se por não considerá-la nos modelos.

A última covariável apresentada na Tabela 1 refere-se ao tratamento feito na instituição, que corresponde ao tipo de tratamento que a paciente foi submetida de acordo com a extensão e estadiamento da doença. Dentre os tipos de tratamentos registrados no estudo, cirurgia fez parte de três deles, sendo que 75% das pacientes necessitaram passar pela cirurgia. O tratamento que consistiu da realização de cirurgia seguida de quimioterapia compreendeu 43% das pacientes, sendo que 44% delas não vieram à óbito (censuras). O tratamento que apresentou o menor número de pacientes foi a radioterapia, com apenas 3%, sendo que 75% delas foram a óbito.

Devido à presença de censuras, as técnicas descritivas mais usuais em estatística, tais como gráficos de barras, de setores e histogramas, não são viáveis para dados de sobrevida. Por esse fato, foi utilizado o estimador de Kaplan-Meier para estimar as curvas de sobrevida das pacientes em cada uma das categorias das covariáveis apresentadas na Tabela 1. A finalidade em se estimar essas curvas foi a de verificar a existência de associação de cada covariável com o tempo de sobrevida, bem como analisar quais covariáveis seriam candidatas a entrar no modelo de Cox, de riscos aditivos de Aalen, de riscos aditivos semiparamétrico e de regressão log-logístico. Essas curvas, obtidas com o auxílio do *software* R, podem ser visualizadas nos gráficos da Figura 1, havendo indícios de diferenças não significativas entre as curvas apenas para a covariável estado civil.

Figura 1 – Curvas de sobrevida obtidas pelo estimador de Kaplan-Meier para as covariáveis da Tabela 1



Fonte: Os autores (2016).

Adicionalmente às curvas de sobrevida estimadas por Kaplan-Meier, foi realizado o teste *logrank* para a comparação das curvas de sobrevida entre as categorias de cada covariável. Os resultados estão na Tabela 2, sendo possível observar, ao nível de significância de 5%, que apenas a covariável estado civil não apresentou evidências de diferença significativa entre as curvas de sobrevida.

Tabela 2 – Resultados do teste *logrank* para as covariáveis do conjunto de dados das pacientes com câncer de ovário de um centro médico de Curitiba

Covariável	**Estatística do teste**	**Valor p**
Idade	32,3	<0,0001
Estado Civil	5,9	0,1164
Ano de entrada no Hospital	11,9	0,0025
Tratamento Prévio	9,98	0,0016
Avaliação e Extensão da Doença	76,6	<0,0001
Diferenciação do Tumor	7,9	0,0049
Estádio da doença	63	<0,0001
Tratamento Feito na Instituição	11,2	0,0243

Fonte: Os autores (2016).

Desta forma, seis covariáveis foram consideradas como potenciais candidatas para os modelos de regressão, sendo elas: idade, ano de entrada no estudo, tratamento prévio, avaliação e extensão da doença, estádio da doença e tratamento feito na instituição. Pelo fato de o modelo de Cox ser o mais utilizado em estudos clínicos, optou-se por começar os ajustes por ele, seguido dos modelos de riscos aditivos (de Aalen e semiparamétrico) e, por fim, o modelo de regressão log-logístico.

4.2 Resultados do Modelo de Cox

Seis covariáveis potencialmente importantes para descrever os tempos de sobrevida das pacientes com câncer de ovário foram indicadas a partir da análise exploratória dos dados. Exceção à covariável idade, as demais, que são categóricas, foram introduzidas nos modelos por meio de variáveis *dummy* com a primeira categoria de cada uma delas considerada como a categoria de referência. Para seleção das covariáveis, foi utilizado um método manual que implicou em, inicialmente, ajustar seis modelos, cada um com uma única covariável e verificar quais são significativas ao nível de significância de 0,10, bem como aquela que apresenta

maior contribuição de acordo com o teste da razão de verossimilhanças. A covariável avaliação e extensão da doença (AED) foi a que mostrou a maior contribuição.

O passo seguinte foi ajustar modelos com duas covariáveis, com AED presente em todos eles, com a finalidade de investigar possíveis correlações entre elas. Para tanto, foram analisadas as mudanças ocorridas tanto na magnitude quanto nos sinais das estimativas dos parâmetros. Correlação ficou evidenciada entre as covariáveis AED e estadiamento da doença (EST), bem como entre AED e tratamento feito na instituição (TFI). Dado que EST apresentou correlação com AED, optou-se pela covariável AED, já que para 27% das pacientes não se tem informação sobre o estadiamento (EST) contra apenas 2% para AED. Quanto à escolha entre AED e TFI, também correlacionadas, foram avaliados dois possíveis modelos: (a) modelo 1, iniciando com a covariável AED e; (b) modelo 2, iniciando com a covariável TFI. Também por método manual, foram acrescentadas a cada um dos dois modelos as demais covariáveis, uma a uma, a fim de verificar quais seriam significativas. No modelo 1, apenas a covariável tratamento prévio (TP) não mostrou significância. Já no modelo 2, todas foram significativas. As Tabelas 3 e 4 mostram, respectivamente, as estimativas dos parâmetros dos modelos com AED e TFI, bem como do coeficiente de correlação de Pearson (ρ) entre os resíduos de Schoenfeld padronizados e os tempos (utilizado para avaliar a suposição de riscos proporcionais).

A partir dos resultados apresentados nas Tabelas 3 e 4 pode-se observar que os valores do coeficiente de correlação de Pearson (ρ) estão próximos de zero, o que indica não haver evidências para a rejeição da suposição de riscos proporcionais.

Tabela 3 – Estimativas dos parâmetros e do coeficiente de correlação de Pearson relativos ao modelo 1 de Cox ajustado aos dados de câncer de ovário

	Covariável	**Coeficiente**	**Erro padrão**	**valor *p***	**rho (ρ)**
	1:Localizado*				
	2:Ext. Direta	0,8703	0,3082	0,0047	-0,1435
AED:	3:Linf. Regionais	1,9071	0,3281	<0,0001	-0,0821
	4:Metástase	1,8524	0,2641	<0,0001	-0,1214
	99: Sem inform.	1,3048	0,6303	0,0385	-0,0573
	1990-1994*				
Ano	1995-1999	0,4422	0,1633	0,0068	0,1178
	2000-2004	-0,3883	0,2083	0,0624	0,0823
Idade	(em anos)	0,0272	0,0052	<0,0001	0,0576

*Nota: *categoria de referência para cada covariável.*
Fonte: Os autores (2016).

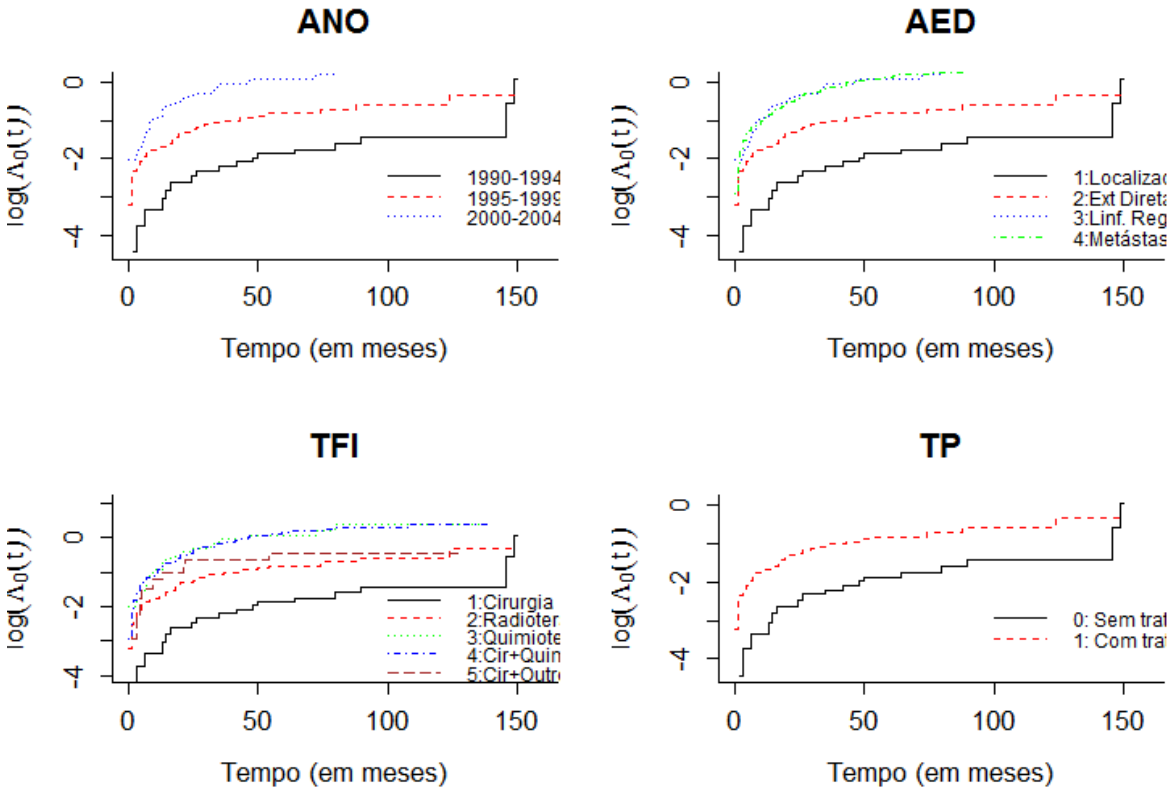
Tabela 4 – Estimativas dos parâmetros e do coeficiente de correlação de Pearson relativas ao modelo 2 de Cox ajustado aos dados de câncer de ovário

Covariável		Coeficiente	Erro padrão	valor *p*	rho (ρ)
TFI:	1:Cirurgia*				
	2:Radioterapia	1,2675	0,3835	0,0009	-0,1036
	3:Quimioterapia	0,8914	0,2222	<0,0001	-0,0291
	4:Cir+Quimio	0,3676	0,1870	0,0493	0,1676
	5:Cir+Outros	0,1432	0,3690	0,6978	-0,1460
Ano:	1990-1994*				
	1995-1999	0,3498	0,1677	0,0370	0,1079
	2000-2004	-0,3143	0,2138	0,1415	0,0483
TP:	0: Sem trat*				
	1: Com trat	-0,5226	0,1610	0,0011	0,0110
Idade (em anos)		0,0283	0,0050	<0,0001	0,0419

*Nota: *categoria de referência para cada covariável.*
Fonte: Os autores (2016).

Adicional ao coeficiente de correlação de Pearson, tem-se na Figura 2 os gráficos das curvas $\log(\widehat{\Lambda}_{0j}(t))$ *versus* os tempos para cada covariável.

Figura 2 – $\mathrm{Log}(\widehat{\Lambda}_{0j}(t))$ vesus *t* para as covariáveis consideradas nos modelos e 1 e 2



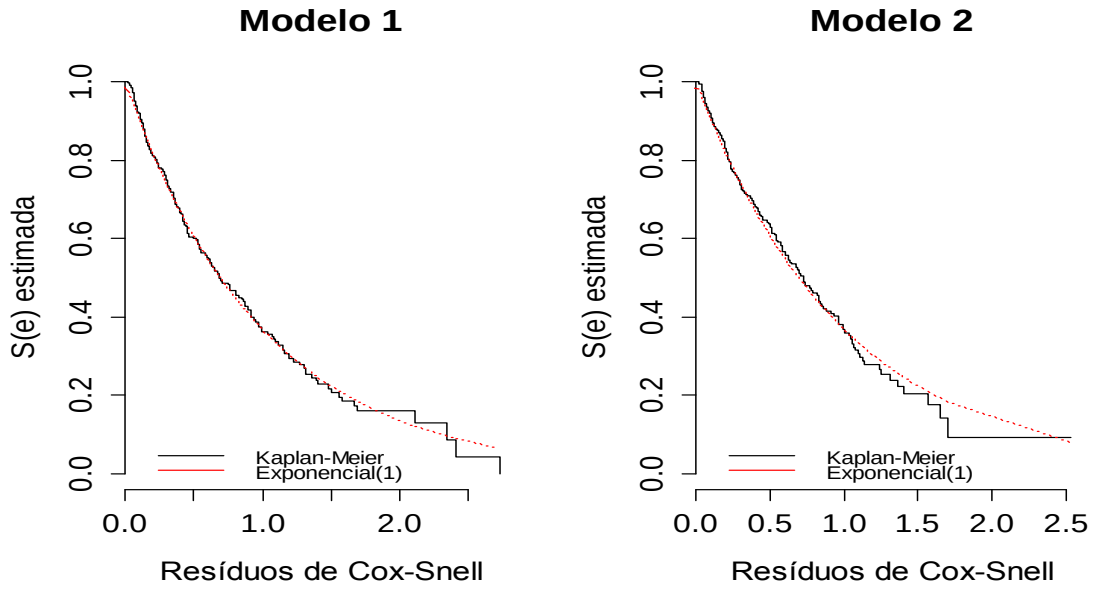
Fonte: Os autores (2016).

A partir da Figura 2, nota-se que as curvas não exibem cruzamentos entre elas, o que evidencia a não rejeição da suposição de taxas de falha proporcionais. Para complementar as análises referente à suposição de taxas de falha proporcionais, foram também obtidos os gráficos dos resíduos de Schoenfeld que, assim como os outros métodos, não indicaram violação da suposição. Os gráficos mencionados para os modelos de Cox 1 e 2 podem ser visualizados no Apêndice A.

Para analisar a qualidade de ajuste dos modelos de Cox 1 e 2, foram utilizados os resíduos de Cox-Snell. A Figura 3 exibe os gráficos das curvas estimadas por Kaplan-Meier e sob a exponencial padrão *versus* os resíduos de Cox-Snell. Como pode ser observado, as curvas estão bem próximas, o que indica boa qualidade de ajuste. Os demais gráficos utilizados na avaliação da adequação dos modelos estão apresentados no Apêndice B, bem como os gráficos referentes aos resíduos martingal e *deviance*, que não mostraram nenhum ponto atípico, estão no Apêndice C.

Do que foi apresentado, notou-se que tanto o modelo com AED quanto o com TFI apresentaram bom ajuste aos dados. Contudo, ao analisar os gráficos dos resíduos de Cox-Snell (Figura 3), optou-se pelo modelo 1 (com as covariáveis: AED, idade e ano de entrada no estudo) por este apresentar ajuste pouco melhor e também por ser mais parcimonioso.

Figura 3 – Curvas S(t) estimadas por Kaplan-Meier e sob a Exponencial padrão para os resíduos de Cox-Snell dos modelos de Cox 1 e 2

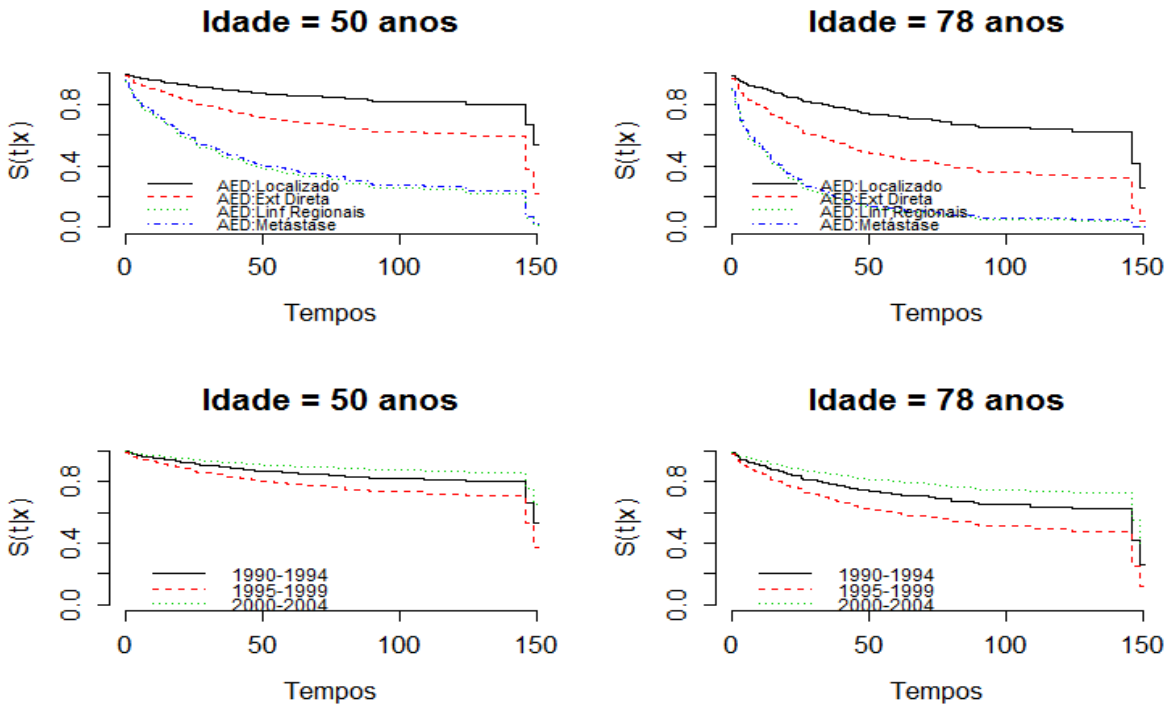


Fonte: Os autores (2016).

De acordo com o modelo 1, a idade da paciente é um fator importante ao analisar a sobrevida de pacientes com câncer de ovário. Quanto mais avançada a idade da paciente, menor a sua probabilidade de sobrevida, como pode ser observado na Figura 4 em que foi feita uma comparação entre mulheres com 50 e 78 anos e observadas as curvas de sobrevida em relação ao ano de entrada no estudo e como a doença foi avaliada. É notável que as pacientes que estão nas categorias 3 e 4 de AED são as que têm a menor probabilidade de sobrevivência, independente da idade. Também é observado que a sobrevida das pacientes que ingressaram no estudo entre os anos de 1995 e 1999 é menor comparado às outras duas categorias.

Observou-se, também, que pacientes com 78 anos apresentaram taxa de óbito de aproximadamente 2,14 vezes a das pacientes com 50 anos de idade. Além disso, se fixarmos as covariáveis idade e ano de entrada no centro médico, tem-se que a taxa de óbito de pacientes que estão com metástase (categoria 4 de AED) é aproximadamente 2,7 vezes a taxa de pacientes avaliadas na categoria 2 de AED. Considerando, ainda, as covariáveis idade e AED fixas, tem-se que a taxa de óbito para pacientes que entraram no estudo entre 1995 e 1999 é aproximadamente 2,3 vezes a taxa de pacientes que ingressaram no estudo entre 2000 e 2004.

Figura 4 – Curvas de sobrevida estimadas pelo modelo de Cox 1 ajustado aos dados



Fonte: Os autores (2016).

4.3 Resultados dos modelos de riscos aditivos

Para a análise dos dados de câncer de ovário foi também considerado dois modelos de riscos aditivos com as três covariáveis selecionadas para o modelo de Cox (AED, idade e ano de entrada no estudo). Nota-se que a covariável contínua idade foi centrada em sua respectiva média. Inicialmente, com o auxílio do *software* R, foi ajustado o modelo de Aalen, em que se considera todas as covariáveis com efeito tempo-dependente. Na Tabela 5 estão dispostos os resultados dos testes aplicados para analisar se o efeito das covariáveis são significativos e se há evidências de efeito tempo-dependente.

Tabela 5 – Resultados dos testes que avaliam o efeito das covariáveis e o efeito tempo-dependente das covariáveis para o modelo aditivo de Aalen

Covariável		$H_{01}: \beta_q(t) = 0$		$H_{02}: \beta_q(t) = \beta_q$	
		Estatística	Valor *p*	Estatística	Valor *p*
Constante		2,69	0,102	0,077	0,441
AED:	1:Localizado*				
	2:Ext. Direta	2,49	0,163	0,130	0,325
	3:Linf. Regionais	4,06	<0,001	0,286	0,269
	4:Metástase	8,04	<0,001	0,312	<0,001
	99: Sem Inform.	1,99	0,385	0,344	0,068
Ano:	1990-1994*				
	1995-1999	3,05	0,048	0,529	0,049
	2000-2004	2,69	0,108	0,167	0,198
Idade - $\bar{x}$		5,77	<0,001	0,002	0,841

*Nota: *categoria de referência para cada covariável.*
Fonte: Os Autores (2016).

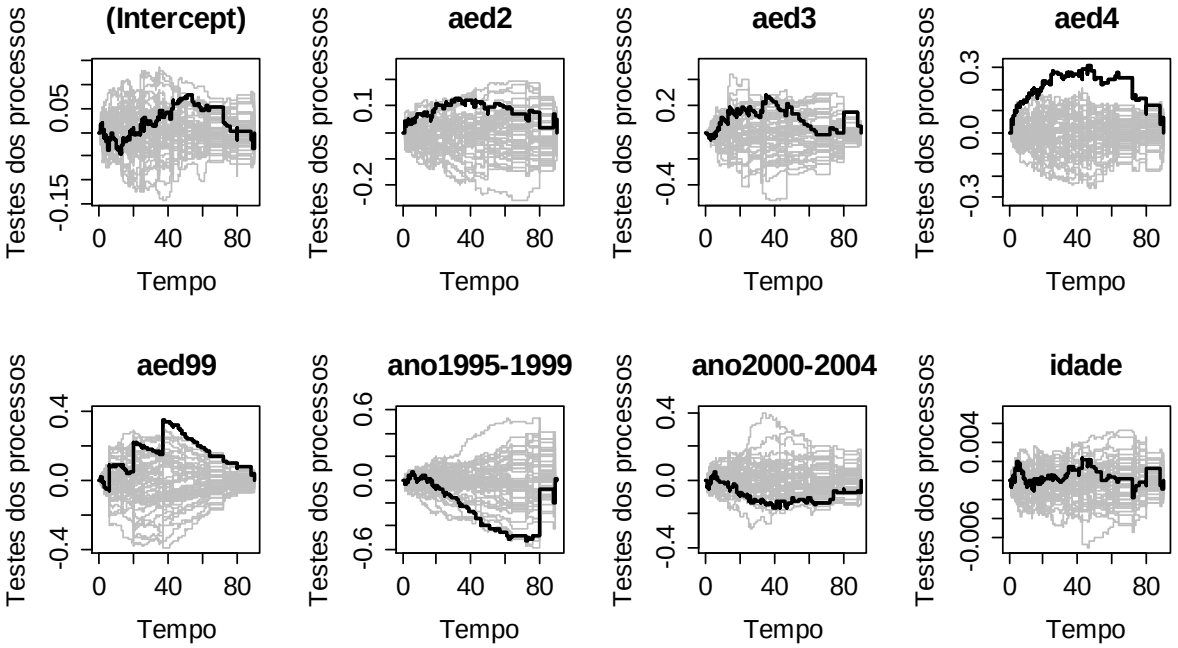
A partir dos resultados dos testes mostrados na Tabela 5, nota-se que todas as covariáveis apresentaram efeito significativo a um nível de significância de 0,10, sendo que apenas a covariável idade não mostrou efeito tempo-dependente. Evidências de que AED e ano apresentaram efeito tempo-dependente podem ser observadas também nos gráficos da Figura 5.

Desse modo, ajustou-se o modelo de riscos aditivos semiparamétrico com apenas a covariável idade sem efeito variando no tempo. Para este modelo, a estimativa do coeficiente tempo-invariante (Tabela 6) reflete que quanto maior for a idade da paciente, maior será o risco de óbito.

Tabela 6 – Estimativa do coeficiente com efeito tempo-invariante no modelo de riscos aditivos semiparamétrico ajustado aos dados de câncer de ovário

Covariável	**Coeficiente**	**Erro padrão**	**Z**	**valor *p***
IDADE - $\bar{x}$	0,0002	0,000	4,99	0,000

Figura 5 – Processo escore observado (linha preta) e os processos escores simulados (linhas cinzas) para as covariáveis do modelo aditivo de Aalen

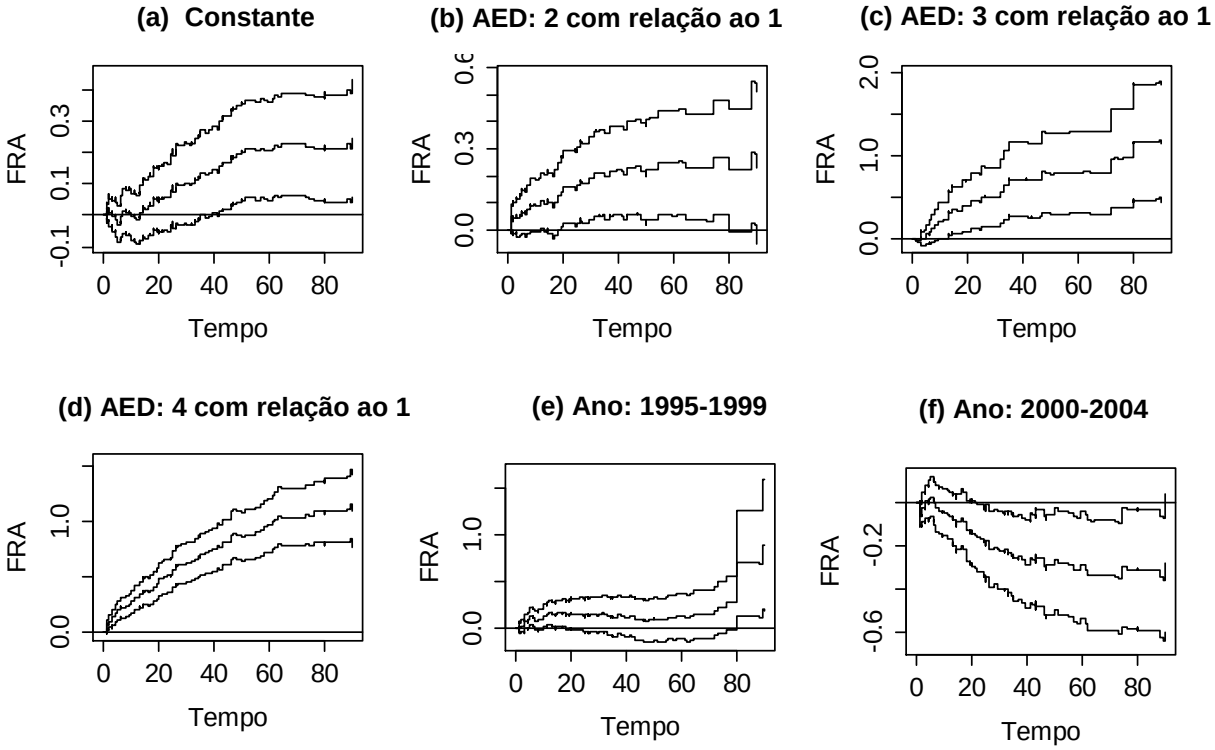


Fonte: Os autores (2016).

Quanto aos efeitos tempo-dependentes, os coeficientes de regressão acumulados que estão representados nos gráficos da Figura 6 evidenciaram que o efeito da covariável AED varia com o tempo. Na Figura 6, o gráfico (a) mostra a taxa de falha acumulada de base estimada para uma paciente de idade média (50 anos), que está na categoria 1 de AED (câncer localizado) e que deu entrada no hospital entre os anos de 1990 e 1994. A taxa de falha desta paciente é crescente com o passar do tempo. Ao analisar a covariável AED, é notável que pacientes que estão nas categorias 3 e 4 de AED (gráficos c e d) apresentam taxa de falha superior e crescente à das pacientes na categoria 1 de AED. Com relação ao ano de entrada no centro médico, nota-se que pacientes que entraram entre os anos de 1995 e 1999 apresentam taxa de falha superior e crescente às pacientes que entraram entre 1990 e 1994 (gráfico e), e o oposto ocorre com as pacientes que entram entre os de 2000 e 2004, apresentando taxa de falha menor e decrescente com relação ao tempo (gráfico f).

As estimativas dos coeficientes do modelo aditivo semiparamétrico foram estimadas até o tempo 103, que corresponde ao maior tempo em que estimação foi possível para os dados desse estudo. Essa é uma desvantagem dos modelos aditivos, que nem sempre possibilitam estimativas para todo o tempo de seguimento.

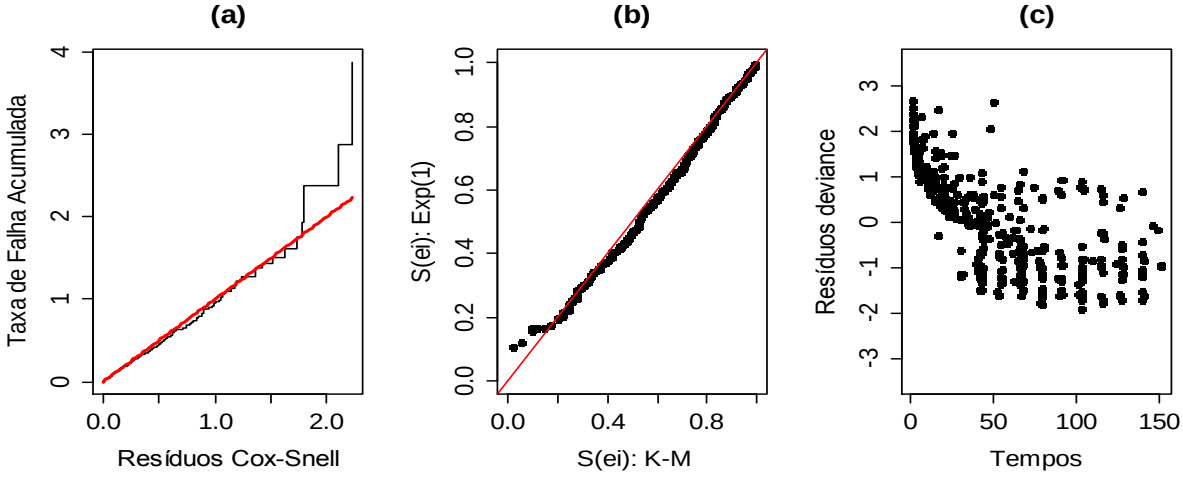
Figura 6 – Coeficientes de regressão acumulados para as covariáveis AED e ano no modelo de riscos aditivos semiparamétrico (bandas de confiança de 95%)



Fonte: Os autores (2016).

Em seguida, foram obtidos os resíduos de Cox-Snell que estão dispostos nos gráficos (a) e (b) da Figura 7 e que mostram evidências favoráveis ao modelo de riscos aditivos semiparamétrico. No gráfico (c) estão representados os resíduos *deviance* que sugerem ausência de observações atípicas.

Figura 7 – Resíduos de Cox-Snell e resíduos *deviance* para o modelo de riscos aditivos semiparamétrico para os dados de câncer de ovário

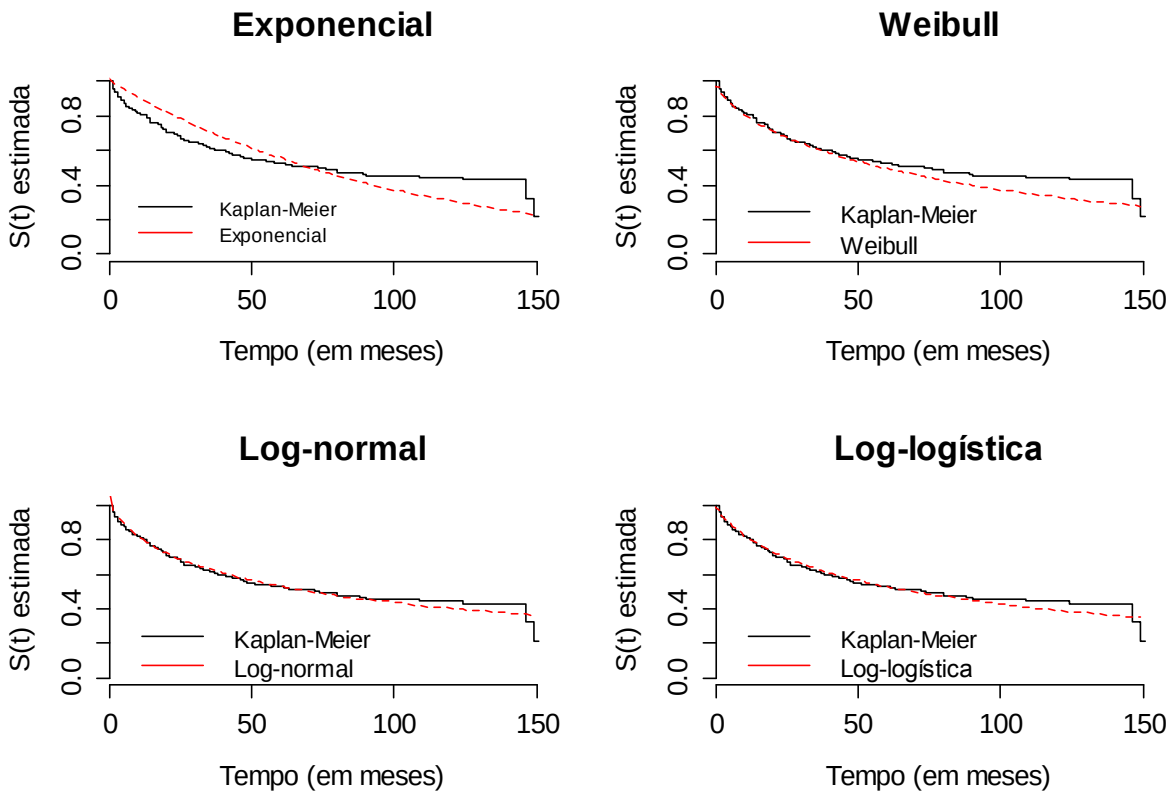


Fonte: Os autores (2016).

4.4 Resultados do modelo de regressão log-logístico

Inicialmente foram ajustados os quatro modelos de regressão paramétricos, citados na Seção 3.2.5, todos na ausência de covariáveis, e estimadas suas respectivas curvas de sobrevivência. Como pode ser visto na Figura 8, os modelos de regressão log-normal e log-logístico foram os que apresentaram, visualmente, uma melhor aproximação com a curva de Kaplan-Meier e são, consequentemente, os melhores candidatos para a análise dos dados do estudo.

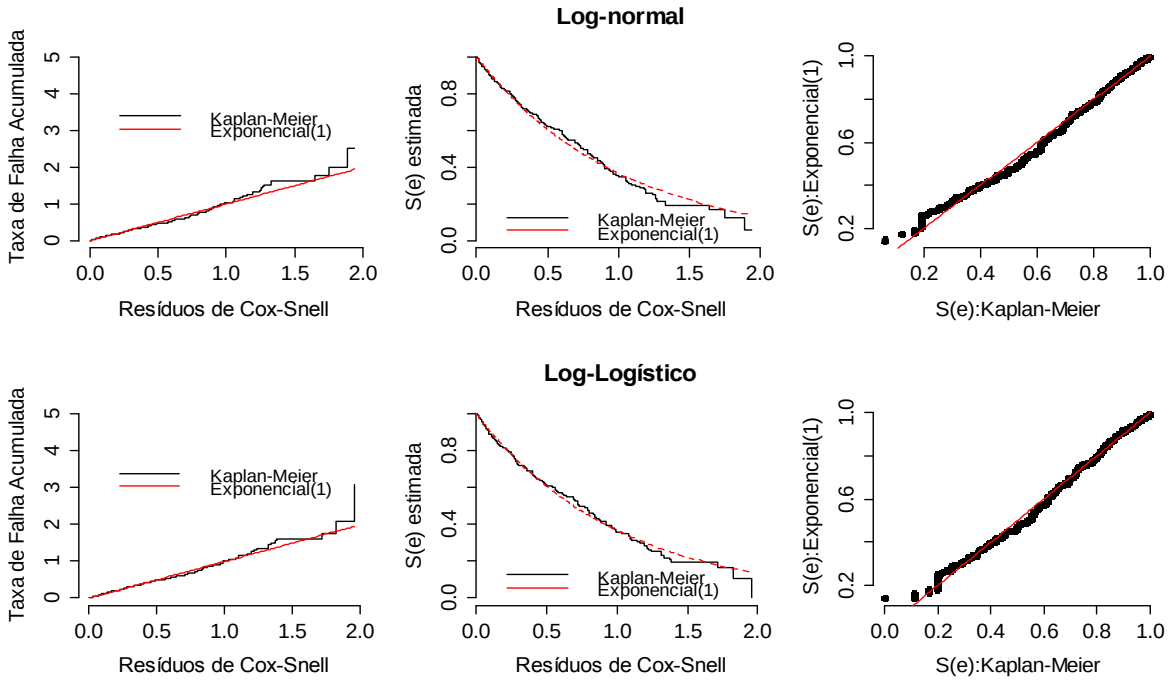
Figura 8 – Curvas de sobrevida estimada via Kaplan-Meier e modelos de regressão paramétricos para os dados do câncer de ovário



Fonte: Os autores (2016).

Dado que os modelos log-normal e log-logístico foram os melhores candidatos sem a presença das covariáveis, foram ajustados os dois modelos na presença de AED e ano de entrada no estudo como covariáveis categóricas e idade como contínua. Com a inclusão das covariáveis, foi analisado os resíduos de Cox-Snell de ambos os modelos (Figura 9), observando-se ajuste pouco superior do modelo log-logístico.

Figura 9 – Análise dos resíduos de Cox-Snell para os modelos de regressão log-normal e log-logístico ajustados aos dados de câncer de ovário



Fonte: Os autores (2016).

Selecionado o modelo de regressão log-logístico, tem-se na Tabela 7 os coeficientes estimados para esse modelo. Tomando-se o exponencial dos coeficientes apresentados na Tabela 7, obtém-se a razão dos tempos medianos de sobrevida. Deste modo, o tempo mediano de mulheres com 50 anos de idade é aproximadamente 2,6 vezes comparado ao das mulheres com 78 anos. Ainda, comparando a idade das pacientes, notou-se que 85% das pacientes com 50 anos sobreviveram até o tempo 110 meses, enquanto que o mesmo percentual de pacientes com 78 anos sobreviveu somente até o tempo 42 meses.

Com relação ao ano de entrada no estudo, verificou-se que o tempo mediano de sobrevida de mulheres que deram entrada no hospital entre os anos de 2000 e 2004 foi aproximadamente 2,72 vezes o das pacientes que deram entrada entre 1995 e 1999.

Com relação à avaliação e extensão da doença, as pacientes que estavam na categoria 2 de AED (metástase), tiverem um tempo mediano de aproximadamente 3,93 vezes ao das que estavam na categoria 4 dessa mesma covariável.

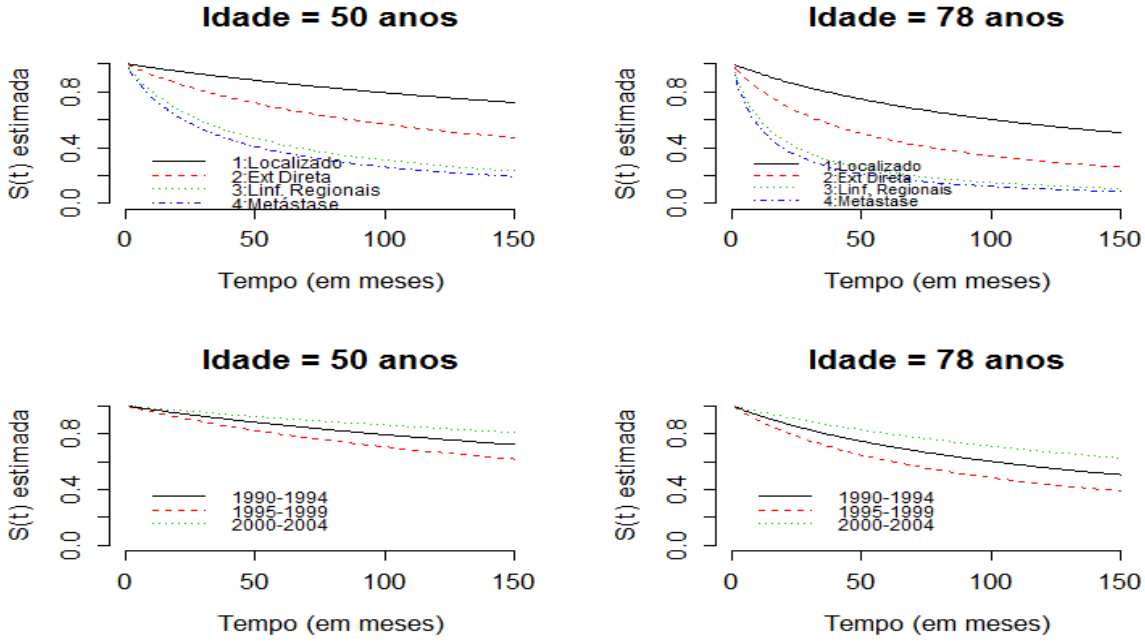
Tabela 7 – Estimativas dos coeficientes de regressão estimados pelo modelo de regressão log-logístico para os dados do câncer de ovário

Covariável		Coeficiente	Erro padrão	Estatística Z	valor *p*
Constante		7,7454	0,5168	14,98	<0,0001
AED:	1:Localizado*				
	2:Ext. Direta	-1,1224	0,3938	-2,85	0,0043
	3:Linf. Regionais	-2,2420	0,4599	-4,87	<0,0001
	4:Metástase	-2,4911	0,3348	-7,43	<0,0001
	99: Sem inform.	-1,7278	0,8244	-2,09	0,0361
Ano	1990-1994*				
	1995-1999	-0,4907	0,2462	-1,92	0,0463
	2000-2004	0,5124	0,3014	1,70	0,0892
Idade (em anos)		-0,0348	0,0076	-4,52	<0,0001

*Nota: *categorias de referência para cada covariável.*

As curvas de sobrevivência estimadas para mulheres com idade de 50 e 78 anos estão dispostas na Figura 10. Observa-se que as curvas têm um decaimento mais acentuado nos gráficos em que se tem idade de 78 anos, indicando que mulheres com idade mais avançada apresentaram uma menor probabilidade de sobreviver comparado àquelas com idade de 50 anos. Assim como nos resultados dos modelos de Cox e de riscos aditivos semiparamétrico tem-se, a partir do modelo de regressão log-logístico, que as pacientes classificadas nas classes 3 e 4 da covariável avaliação de extensão da doença apresentaram probabilidades menores de sobreviver, assim como as pacientes que deram entrada no estudo entre os anos de 1995 e 1999.

Figura 10 – Curvas de sobrevida de pacientes com 50 e 78 anos de idade ajustadas pelo modelo de regressão log-logístico

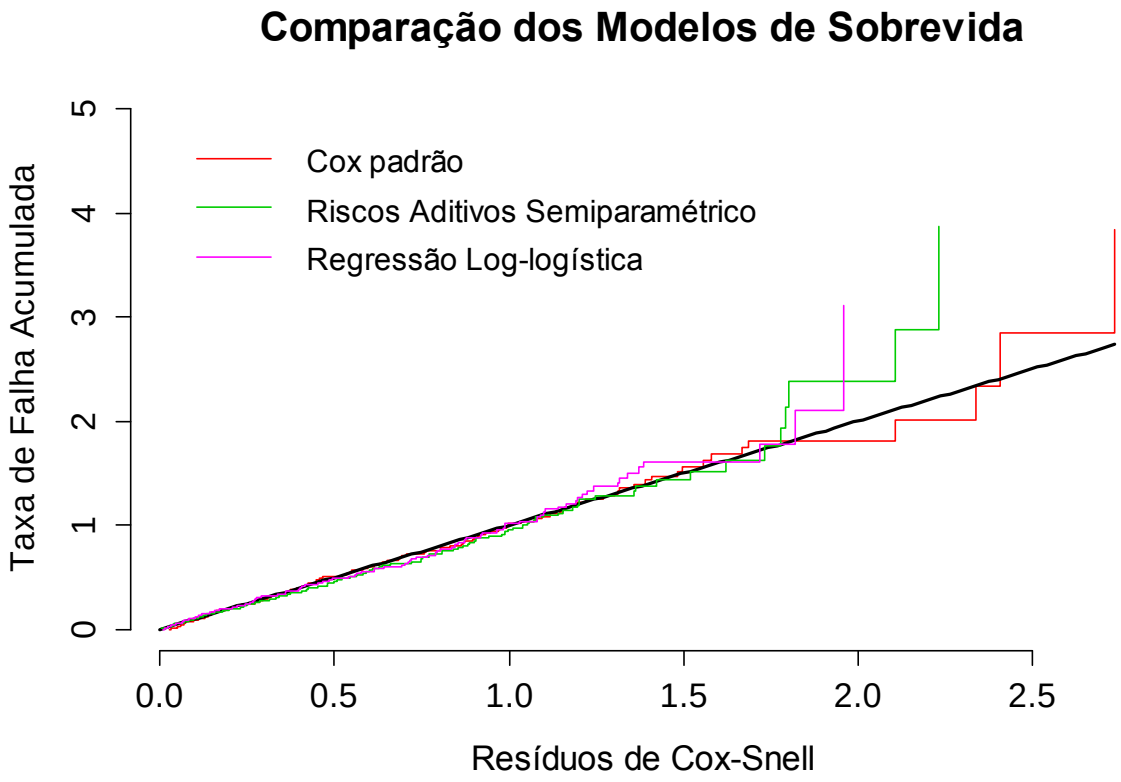


Fonte: Os autores (2016).

4.5 Análise comparativa dos modelos ajustados

Nesta seção, foi realizada uma comparação entre os três modelos de sobrevivência ajustados aos dados de câncer de ovário. Quanto à qualidade de ajuste global dos modelos, foi analisado o gráfico exibido na Figura 11, o qual compreende as taxas de falha acumulada *versus* os resíduos de Cox-Snell para os modelos de Cox, de riscos aditivos semiparamétrico e de regressão log-logístico. Pode-se perceber evidências favoráveis aos três modelos, com uma leve superioridade do modelo de Cox, que apresenta uma proximidade mais acentuada com a reta.

Figura 11 – Resíduos $\hat{e}_i$ *versus* $\Lambda(\hat{e}_i)$ para os três modelos ajustados aos dados de câncer de ovário



Fonte: Os autores (2016).

Quanto à qualidade de predição dos modelos, esta foi avaliada e comparada com base nas AUC(*t*), método apresentado na Seção 3.2.7. Os erros padrão foram estimados via reamostragem *bootstrap*. Os resultados estão dispostos na Tabela 8 e mostram que o modelo de regressão log-logístico apresentou qualidade de predição levemente superior aos demais, embora a diferença entre os valores da AUC(t) para os três modelos não seja tão distinta.

Tabela 8 – Estimativas das AUC(t) em $t\epsilon[10{,}120]$ e respectivos erros padrão (e.p.) para os três modelos ajustados aos dados do câncer de ovário

	AUC(t)					
Modelo	t = 10	t = 30	t = 50	t = 70	t = 90	t = 120
Cox	0.718 (0.029)	0.740 (0.026)	0.765 (0.024)	0.786 (0.024)	0.799 (0.028)	0.803 (0.028)
Aditivo*	0.694 (0.027)	0.748 (0.025)	0.774 (0.024)	0.794 (0.023)	0.804 (0.024)	----
Log-logístico	0.717 (0.037)	0.753 (0.029)	0.783 (0.027)	0.803 (0.028)	0.811 (0.031)	0.817 (0.032)

*Nota: *modelo de riscos aditivos semiparamétrico*

5 CONSIDERAÇÕES FINAIS

O câncer de ovário ainda é uma neoplasia bastante difícil de ser diagnosticada quando ainda está no início. Por esta razão, a taxa de óbito é grande e, portanto, há necessidade do desenvolvimento de técnicas médicas que previnam esse câncer.

As análises realizadas neste trabalho tiveram como objetivo estudar o tempo de sobrevida de pacientes diagnosticadas com câncer de ovário e que foram tratadas em um centro médico de Curitiba, Paraná. Com este objetivo, foram utilizadas técnicas estatísticas de análise de sobrevivência que possibilitaram um estudo com o ajuste de três modelos de regressão estatísticos.

Com o auxílio do estimador de Kaplan-Meier, foi realizada uma análise descritiva dos dados, com o qual foi possível compreender o comportamento das covariáveis mediante o tempo de sobrevida das pacientes, e selecionar as covariáveis candidatas a fazerem parte dos modelos de regressão.

Os três modelos ajustados, sendo eles: (a) modelo de regressão de Cox, (b) modelo de riscos aditivos semiparamétrico e, (c) modelo de regressão log-logístico, apresentaram ajustes bastante satisfatórios e, portanto, semelhantes. Dessa maneira, a proposta inicial do trabalho foi atendida.

Optou-se pelo modelo de Cox para um ajuste inicial pelo histórico de casos bem-sucedidos desse modelo, que é o mais utilizado na área médica. A seleção das covariáveis apontou para três modelos distintos, dado a forte correlação entre três delas, sendo o modelo final composto pelas seguintes covariáveis: avaliação e extensão da doença, idade da paciente e ano de entrada no centro médico.

Ao utilizar o modelo de riscos aditivos de Aalen, percebeu-se que duas covariáveis (avaliação e extensão da doença e ano de entrada no centro médico) apresentaram efeito tempo-dependente, o que nos levou a utilizar uma extensão do modelo de Aalen, denominado modelo de riscos aditivos semiparamétrico. Uma restrição deste modelo foi a estimação limitada a um tempo máximo, deixando-o em desvantagem em relação aos demais.

O modelo de regressão paramétrico escolhido foi o log-logístico, que dentre os modelos de Weibull, exponencial e log-normal, foi o que melhor se ajustou aos dados de câncer de ovário, apresentando, também, um ajuste bastante satisfatório.

Os três modelos apresentaram tempo de sobrevida maior para pacientes que deram entrada no centro médico entre os anos de 2000 e 2004, que eram mais jovens

e que foram avaliadas quando estavam ainda no início da neoplasia. Contudo, pacientes com idades mais avançadas, que foram avaliadas com metástase e que deram entrada no hospital entre os anos de 1995 e 1999, tiveram uma probabilidade de sobrevida bastante reduzida.

Vale ressaltar, que por mais que o ajuste dos modelos tenha sido bastante satisfatório, sobretudo para o modelo de Cox, não significa que os resultados obtidos a partir desses modelos possam ser estendidos para outras populações devido a vários fatores, dentre eles, às divergências entre regiões, por exemplo.

REFERÊNCIAS

AALEN, O.O. A model for non-parametric regression analysis of counting processes. **Mathematical Statistics and Probability. Lecture Notes in Statistics**, Springer, New York, v. 2, p. 1-25, 1980.

AALEN, O.O. A linear regression model for the analysis of lifetimes. **Statistics in Medicine**, v. 8, p. 907-925, 1989.

AKRITAS, M. G. Nearest neighbor estimaton of a bivariate distribution under random censoring. **The Annals of Statistics**, Beachwood, v. 22, p. 1299-1327, 1994.

ALBERTS, B.; JOHNSON, A.; LEWIS, J.; RAFF, M.; ROBERTS, K.; WALTER, P. **Biologia Molecular da Célula**. 4 ed. Porto Alegre: Artmed, p. 1463, 2004.

BRESLOW, N.E. Discussion of Professor Cox's Paper. **Journal of the Royal Statistical Society B**, v. 34, p. 216-217, 1972.

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.

CHANG, C.; CHIANG, A. J.; WANG, H.; CHEN, W.; CHEN, J. Evaluation of the time-varying effect of prognostic factors on survival in ovarian câncer. **Annals of Surgical Oncology**, v.22, p. 3976-3980, 2015.

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.

COX, D.R. Patial likelihood. **Biometrika**, v. 65, p. 269-276, 1975.

COX, D.R.; SNELL, E.J. A general definition of residuals. **Journal Royal Statistical Society, Series B**, v. 30, p. 248-275, 1968.

EFRON, B. The efficiency of Cox´s likelihood function for censored data. **Journal of the American Statistical Association**, Alexandria, v. 72, p. 557-565, 1977.

GRAMBSCH, P.M; THERNEAU, T.M. Proportional hazards tests and diagnostics based on weighted residuals. **Biometrika**, v. 81, n. 3, p. 515-526, 1994.

HEAGERTY, P. J.; ZHENG, Y. Survival model predictive accuracy and ROC curves. **Biometrics**, Washington, v. 61, p. 92-105, 2005.

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 NACIONAL DE CÂNCER JOSÉ DE ALENCAR GOMES DA SILVA. **Estimativa 2014: Incidência de cancer no Brasil**.Rio de Janeiro: INCA, 2014. 124p.

INSTITUTO ONCOGUIA. **Taxa de sobrevida para o câncer de ovário por estágio**. Disponível em: <http://www.oncoguia.org.br/conteudo/taxa-de-sobrevida-para-o-cancer-de-ovario-por-estagio/6048/229>. Equipe Oncoguia, 2014. Acesso em: 16/11/2015.

KAPLAN, E.L.; MEIER, P. Nonprametric estimation from incomplete obervations. **Journal of the American Statistical Association**, v. 53, p. 457-481, 1958.

LAWLESS, J. F. **Statistical models and methods for lifetime data**. 2nd ed. New York: John Wiley and Sons, 2003. 664 p.

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 dat and two new rank order statistics arising in its consideration. **Cancer Chemotherapy Reports**, v. 50, p. 163-170, 1966.

MARTINUSSEN, T.; SCHEIKE, T. H. **Dynamic regression models for survival data**. New York: Springer Verlag, 470 p., 2006.

MCKEAGUE, I. W.; SASIENI, P. D. A partly parametric additive risk model. **Biometrika**, Oxford, v. 81, p. 501-514, 1994.

PAN, S.Y.; UGNAT, A.M.; MAO, Y.; THE CANADIAN CANCER REGISTRIES EPIDEMIOLOGY RESEARCH GROUP. Physical activity and the risk of ovarian cancer: A case-control study in Canada. **International Journal of Cancer**, v. 117, p. 300-307, 2005.

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/.

RAMINELLI, J.A. **Métodos de adequação e diagnóstico em modelos de sobrevivência dinâmicos**. Tese (Doutorado) – Escola Superior de Agricultura “Luiz de Queiroz”, São Paulo: Piracicaba, 113 p., 2015.

RIMAN, T.; DICKMAN, P. W.; NILSSON, S.; CORREIA, N.; NORDLINDER, H.; MAGNUSSON, C. M.; PERSSON, I.R. Risk Factors for Invasive Epithelial Ovarian Cancer: Results from a Swedish Case-Control Study. **American Journal of Epidemiology**, v. 156, n. 4, 2002.

RISTOW, C. M.; YAMAMOTO, C. T.; FÁVARO, M. Fatores de risco e patogênese das neoplasias malignas epiteliais de ovário: revisão de literature. **Revista Brasileira de Cancerologia**, v. 52, p. 185-195, 2006.

SALANI, R.; SANTILLAN, A.; ZAHURAK, M.; GIUNTOLI, R.; GARDNER, G. J.; ARMSTRONG, D. K.; BRISTOW, R. E. Secondary cytoreductive surgery for localized recurrent epithelial ovarian cancer: Analysis of prognostic factors and survival outcome. **Cancer**, v. 109, n.4, p. 685-691, 2007.

SALEHI, F., DUNFIELD, L., PHILLIPS, K.P., KREWSKI, D., VANDERHYDEN, B. C. Risk Factors for Ovarian Cancer: Na Overview with Emphasis on Hormonal Factors. **Journal of Toxicology and Environmental Health**. p. 301-321. 2008.

SILVA-FILHO, A. L.; CÂNDIDO, E. B.; NOVIELLO, M. B.; SANTOS-FILHO, A. S.; TRAIMAN, P.; TRIGINELLI, S. A.; CUNHA-MELO, J. R. Cirurgia não ginecológica em pacientes com câncer de ovário. **Revista Brasileira de Ginecologia e Obstetrícia**, v. 26, n. 5, p. 411-416, 2004.

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

TADIKAMALLA, P.R.; JOHNSON, N.L. Systems of frequency curves generated by transformations of logistic variables. **Biometrika**, London, v. 69, n. 2, p. 461-465, 1982.

THERNEAU, T.M.; GRAMBSCH, P.M. **Modeling survival data: extending the Cox model**. Springer-Verlag, New York, 2000.

TORRES, J. C. C.; DERCHAIN, S. F. M.; FAÚNDES, A.; GONTIJO, R. C.; MARTINEZ, E. Z.; ANDRADE, L. A. L. A. Risk-of-Malignancy Index in preoperative evaluation of clinically restricted ovarian cancer. **Sao Paulo Medical Journal**, v. 120, n. 3, p. 72-76, 2002.

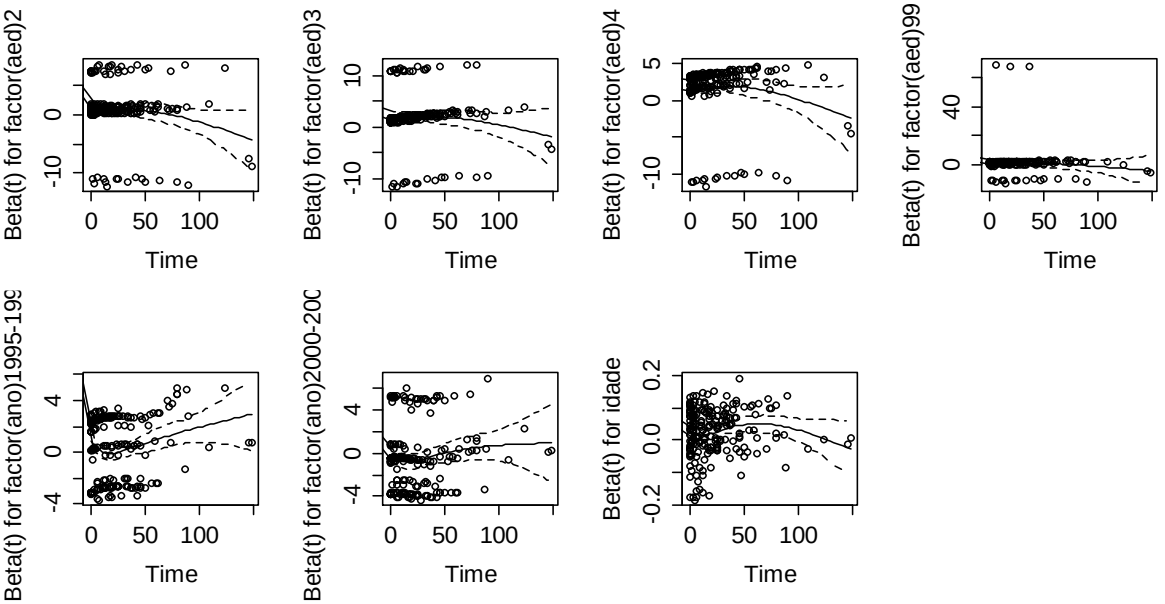
U.S. CANCER STATISTICS WORKING GROUP. **United States Cancer Statistics: 1999–2012 Incidence and Mortality Web-based Report**. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute; 2015. Disponível em: <www.cdc.gov/uscs>.

VANDERHYDEN, B. C., SHAW, T. J., GARSON, K., TONARY, A.M. **Ovarian carcinogenesis**. The Ovary, eds. P. C. K. Leung and E. Y. Adashi, p. 591-612. San Diego, CA: Elsevier Academic Press. 2003.

APÊNDICES

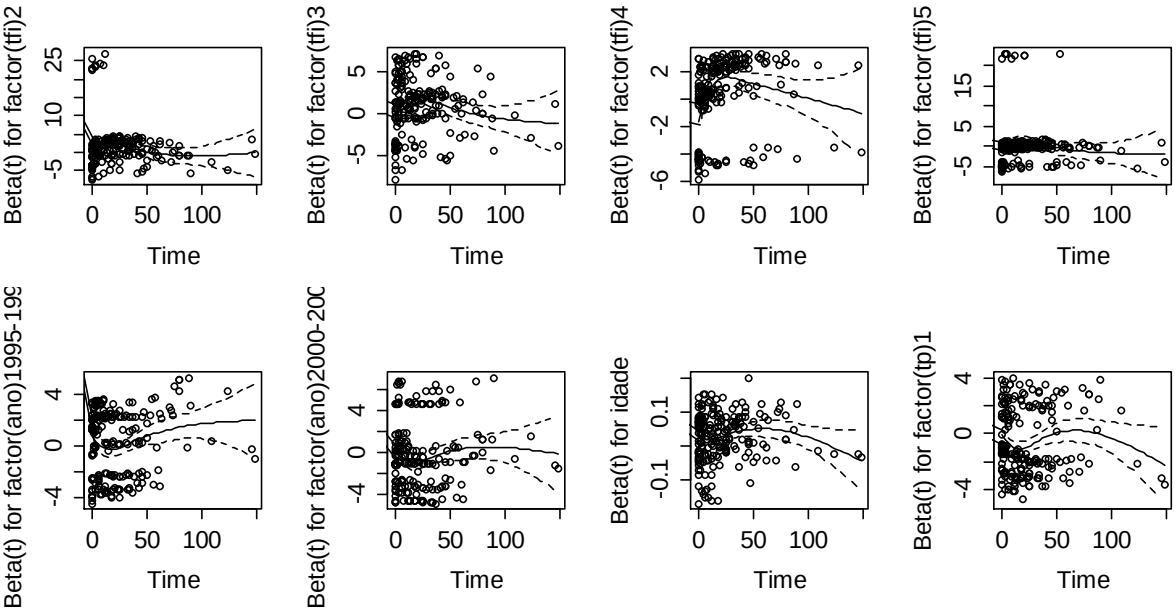
APÊNDICE A – Resíduos de Schoenfeld para os modelos de Cox 1 e 2

Figura A1 – Resíduos padronizados de Schoenfeld associados às covariáveis do modelo de Cox 1 para averiguar a suposição de proporcionalidade



Fonte: Os autores (2016).

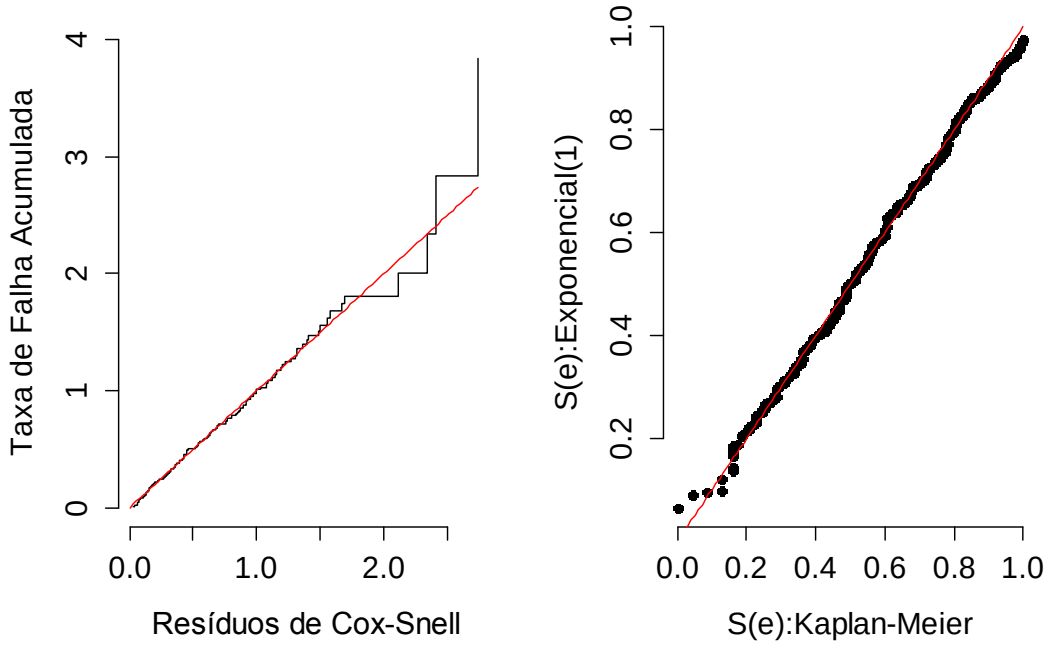
Figura A2 – Resíduos padronizados de Schoenfeld associados às covariáveis do Modelo de Cox 2 para averiguar a suposição de proporcionalidade



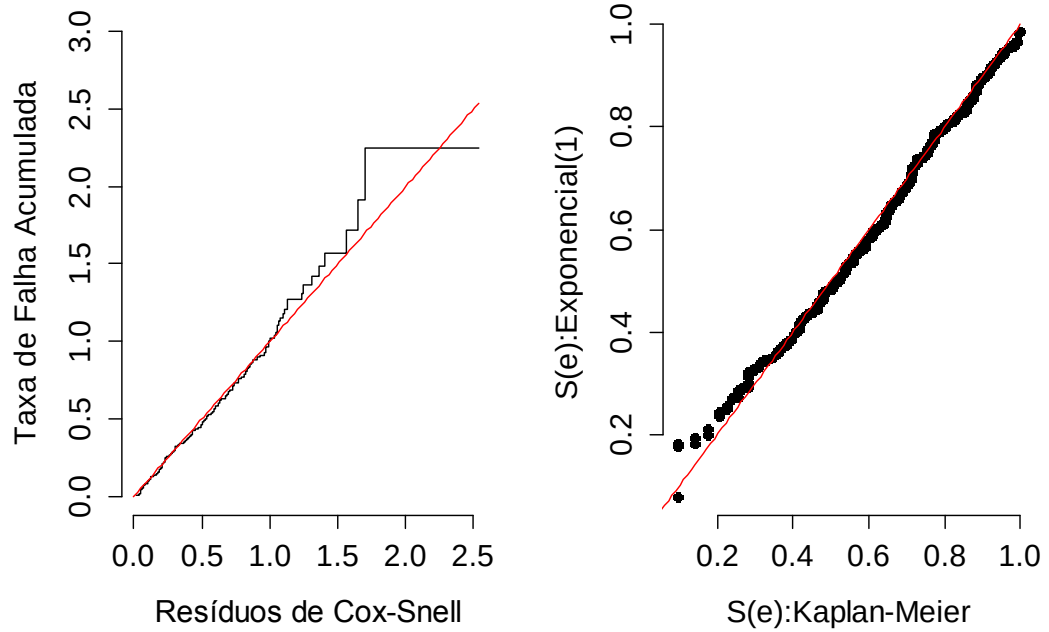
Fonte: Os autores (2016).

APÊNDICE B – Análise dos resíduos de Cox-Snell para os modelos 1 e 2

Modelo 1

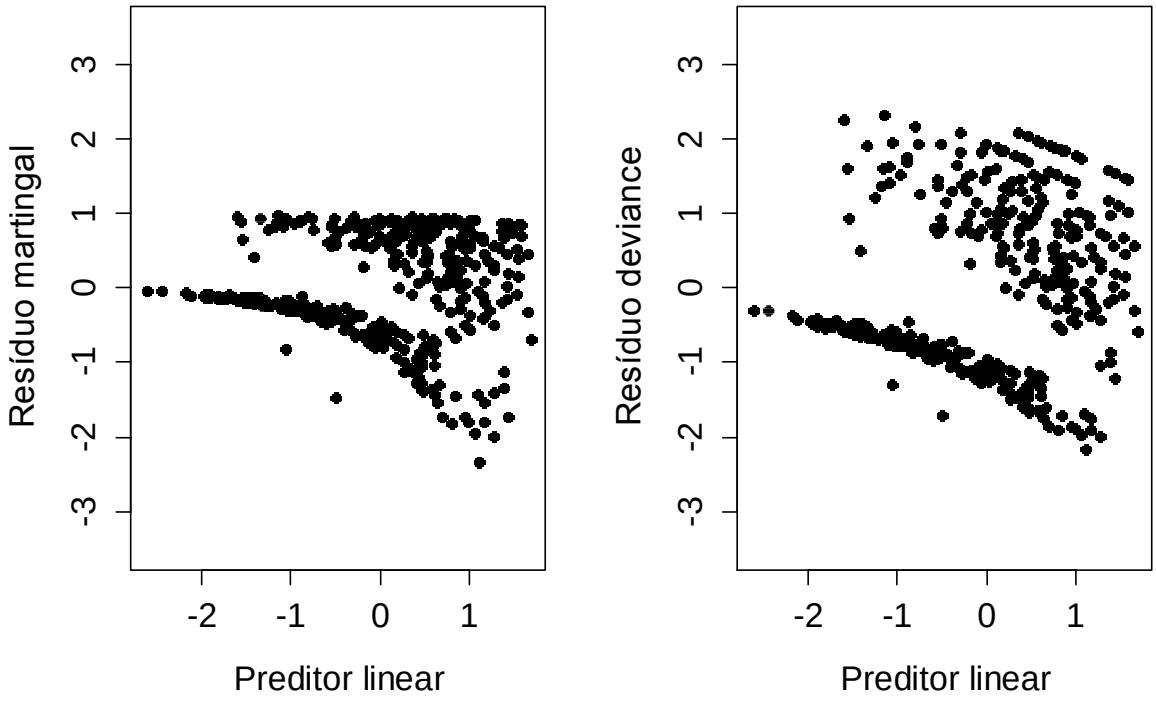


Modelo 2



APÊNDICE C – Análise dos resíduos martingal e *deviance* para os modelos 1 e 2

Modelo 1



Modelo 2