

Análise de sobrevivência - Aplicação do modelo de regressão de Cox

Survival analysis - The Cox regression model application

Maria Helena Rigão¹, Anaelena Bragança Moraes², Roselaine Ruviaro Zanini³

Resumo

O principal propósito, desta pesquisa, foi realizar um estudo sobre análise de sobrevivência utilizando o modelo semi-paramétrico de riscos proporcionais de Cox. Para essa análise utilizaram-se dados obtidos no Hospital Universitário de Santa Maria, referentes a 69 pacientes com diagnóstico de câncer no aparelho digestivo. Estes pacientes foram acompanhados até a morte, ou o fim do estudo, ou a perda do seguimento. As covariáveis avaliadas foram: idade, tipo de cirurgia, estágio da doença, tipo de tratamento cirúrgico e data de óbito ou de censura. A média de idade dos pacientes foi de 60 anos, sendo que 71% deles tiveram como desfecho o óbito. A análise mostrou pelo modelo de riscos proporcionais de Cox ajustado para os dados, que apenas a covariável estágio da doença foi significativa ($p=0,001$), verificado pelo teste da razão de verossimilhança parcial e teste de Wald. Com o modelo encontrado, concluiu-se que o risco relativo de óbito aumenta com o avanço do estágio da doença.

Palavras-chave: Análise de sobrevivência, Kaplan-Meier, Regressão de Cox.

Abstract

The main purpose of this research was to conduct a study on survival analysis using the semi-parametric Cox proportional hazards model. For this analysis we used data obtained from the University Hospital of Santa Maria, referring to 69 patients diagnosed with cancer. in the digestive tract. These patients were followed until death, or end of study, or loss of follow-up. The covariates evaluated were: age, type of surgery, disease stage, type of surgical treatment and date of death or censorship. The average age of the patients was 60 years, and 71% of them had death as their outcome. The analysis showed by the data adjusted Cox proportional hazards model that only the covariate stage of the disease was significant ($p = 0.001$), verified by the partial likelihood ratio test and the Wald test. With the model found, it was concluded that the relative risk of death increases with advancing disease stage.

Keywords: Survival analyses, Kaplan-Meier, Cox proportional hazards regression.

¹ Universidade Federal de Santa Maria, Brasil; e-mail: hrigao@yahoo.com.br

² Universidade Federal de Santa Maria. Brasil. e-mail: anaelena@smail.ufsm.br;

³ Universidade Federal de Santa Maria, Brasil; e-mail: rrrzanini63@gmail.com



1 Introdução

A análise de sobrevivência é o conjunto de técnicas e modelos estatísticos, utilizados na análise do comportamento de variáveis aleatórias positivas. Nesse tipo de análise, a variável resposta é, geralmente, o tempo até a ocorrência de um evento de interesse. Esse tempo é denominado tempo de falha, podendo ser o tempo até a falha de um equipamento, a morte de um paciente, bem como até a cura ou recidiva de uma doença. A principal característica de dados de sobrevivência é a presença de censura, que é a observação parcial da resposta (COLOSIMO, 2001; BUSTAMANTE-TEIXEIRA, FAERSTEIN, LATORRE, 2002).

A área de sobrevivência é uma das que mais cresceu nas últimas duas décadas do século passado. Isso se evidencia pelo número de aplicações de análise de sobrevivência em medicina (COLOSIMO; GIOLO, 2006). Na área da saúde a análise de sobrevivência é utilizada quando se deseja identificar fatores de prognóstico, ou de risco, de certa doença, ou comparar tratamentos de interesse em um estudo clínico controlado (COLOSIMO, 2001). Técnicas estatísticas não-paramétricas são, freqüentemente, utilizadas para esse fim. Entre elas, o estimador de Kaplan-Meier para a função de sobrevivência e o teste log-rank para comparar funções de sobrevivência. Quando se deseja encontrar um modelo que explique a sobrevivência dos pacientes, considerando todas as covariáveis, ao mesmo tempo, pode-se utilizar o modelo semi-paramétrico de riscos proporcionais de Cox (CARVALHO, ANDREOZZI, CODEÇO et al., 2005; COLOSIMO, GIOLO, 2006). Com isso, a análise estatística final de dados de sobrevivência é, geralmente, realizada utilizando-se modelos de regressão.

Os registros de câncer e institutos como o INCA (Instituto Nacional do Câncer) reúnem as informações gerando relatórios relacionados à sobrevivência de pessoas com os vários tipos de câncer, os quais são divulgados por meio impresso e eletrônico. Esses relatórios são fontes importantes que subsidiam de informações os órgãos de saúde nas suas tomadas de decisões. As análises de sobrevivência, em geral, são feitas com base nas técnicas clássicas de sobrevivência considerando uma causa para o óbito (remissão, recidiva, etc.) como desfecho de interesse (FERRAZ; MOREIRA-FILHO, 2017).

Nesse trabalho, o objetivo foi estudar as técnicas de análise de sobrevivência com foco principal no modelo de regressão de Cox, bem como realizar a aplicação das técnicas de sobrevivência utilizando dados da área médica.

2 Materiais e métodos

Para realizar a análise de sobrevivência serão definidos alguns conceitos importantes, nesta área, para posteriormente, desenvolver-se a teoria do modelo de regressão de Cox, que é o objeto deste trabalho.

Tempo de sobrevivência (t): é o tempo observado desde a entrada do indivíduo no estudo, até a ocorrência de um evento de interesse; tempo inicial: é a data de início do estudo; evento de interesse ou falha: refere-se a eventos geralmente indesejáveis, como a morte de um paciente, a recidiva de uma doença, etc.; censura: são as informações incompletas sobre o tempo de sobrevivência, caracterizando assim a presença de dados censurados. A não consideração de informações incompletas sobre esses tempos de vida pode levar a inferências viciadas ou menos eficientes.

Razões pelas quais a censura pode ocorrer: não ocorre o evento de interesse antes do fim do estudo; o indivíduo é "perdido" durante o período de estudo; o indivíduo retirou-se do estudo porque morreu (se a morte não é o evento de interesse) ou alguma outra razão. Também, a função de sobrevivência $S(t)$ é a probabilidade de um indivíduo sobreviver durante algum tempo especificado t:

$$S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t)$$

Principais características da função de sobrevivência:

$S(t)$ é uma função monótona decrescente; no tempo $t = 0$, $S(t) = 1$ ou 100% e $\lim_{t \rightarrow 0} S(t) = 0$.

A estimação de $S(t)$ e da função acumulada da taxa de falha $H(t)$, realizada a partir dos tempos de sobrevivência observados, pode ser feita utilizando-se procedimentos paramétricos ou procedimentos não-paramétricos.

Se o procedimento for paramétrico a distribuição de probabilidade de T é conhecida, sendo as mais utilizadas: a distribuição exponencial, a Weibull, a Gamma e a lognormal. No caso não-paramétrico, a distribuição de probabilidade de T não precisa ser especificada. Destacam-se a função de sobrevivência empírica, as Tabelas de Vida, o estimador Kaplan-Meier e o modelo de regressão de Cox (BUSTAMANTE-TEIXEIRA, FAERSTEIN, LATORRE, 2002).

A função de sobrevivência empírica, para dados sem censura é definida como:

$$\hat{S}(t) = \frac{\text{número de observações que não falharam até o tempo } t}{\text{número de observações no estudo}}$$

$\hat{S}(t)$ é uma função escada com degraus nos tempos observados de falha de tamanho $1/n$, onde n é o tamanho da amostra. Se existirem empates em certo tempo t , o tamanho do degrau fica multiplicado pelo número de empates. O estimador de Kaplan-Meier (1958), na sua construção, considera o número de intervalos de tempo equivalentes ao número de falhas distintas. As probabilidades de falha são calculadas somente no instante em que elas ocorrerem.

Suponha que n indivíduos sejam observados e, além disso, que o tempo de falha é observado para um dos indivíduos em k tempos distintos $t_1 < t_2 < \dots < t_k$. Se uma censura e uma falha ocorrem ao mesmo tempo, considera-se a falha antes da censura.

A estimativa da curva de sobrevida, usando o estimador de Kaplan-Meier é:

$$\hat{S}(t) = \prod_{j: t_j < t} \left(1 - \frac{d_j}{n_j}\right), \quad t_k \leq t \leq \text{a maior observação censurada}; \text{ onde:}$$

$$n_j = n - d_{j-1} - w_{j-1};$$

n_j = números de indivíduos sob risco, isto é, que não falharam e não foram censurados até o tempo t_j , exclusive;

$$d_{j-1} = \text{número de falhas no tempo } t_{j-1};$$

$$w_{j-1} = \text{número de observações censuradas no tempo } t_{j-1}.$$

Para o modelo de regressão de Cox, que é um modelo de regressão semi-paramétrico, a função de risco, no caso geral, assume a forma:

$$h(t) = h_0(t) g(X'\beta) = h_0(t) \cdot \exp(\beta_1 X_1 + \dots + \beta_p X_p)$$

O componente não-paramétrico, $h_0(t)$, não é especificado e é uma função não-negativa do tempo, chamada de função de risco base. O componente paramétrico $g(X'\beta)$, é uma combinação linear das covariáveis, e assumirá um valor diferente para cada uma delas, sendo β o seu vetor de parâmetros associado Cox (1972). O modelo de Cox é também chamado modelo de riscos proporcionais, pois a razão das taxas de falhas entre dois indivíduos distintos, i e j , é constante no tempo:

$$\frac{h_i(t)}{h_j(t)} = \frac{h_0(t)\exp(X'_i\beta)}{h_0(t)\exp(X'_j\beta)} = \exp(X'_i\beta - X'_j\beta)$$

$$HR_{(t,i,j)} = \exp\beta(X'_i - X'_j)$$

Para estimar os coeficientes β do modelo, Cox (1972, 1975) sugere o método de máxima verossimilhança parcial ou condicional, para estimar os coeficientes.

$$\frac{h_i(t_i)}{\sum_{j \in R(t_i)} h_j(t_i)} = \frac{h_0(t)\exp(X'_i\beta)}{\sum_{j \in R(t_i)} h_0(t)\exp(X'_j\beta)} = \frac{\exp(X'_i\beta)}{\sum_{j \in R(t_i)} \exp(X'_j\beta)}$$

onde:

$R(t_i)$ = conjunto dos índices das observações sob risco no tempo t_i .

Ao condicionar falhas e censuras até o tempo t_i , o componente não-paramétrico $h_0(t)$ desaparece e a função de verossimilhança a ser usada é:

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

δ_i = indicador de falha.

Os valores de β que maximizam $L(\beta)$, são obtidos resolvendo o sistema de equações definido por $U(\beta)=0$, onde $U(\beta)$ é o vetor escore de primeiras derivadas da função $\ln(L(\beta))$. Isto é:

$$U(\beta) = \sum_{i=1}^n \left[\delta_i X_i - \frac{\sum_{j \in R(t_i)} X_j \exp(X'_j\beta)}{\sum_{j \in R(t_i)} \exp(X'_j\beta)} \right] = 0$$

Quando existem observações empatadas, a verossimilhança parcial deve ser modificada (BRESLOW E PETO *apud* COLOSIMO (2001)). O teste de Wald para a significância dos coeficientes do modelo de Cox compara as estimativas de verossimilhança dos parâmetros β com a estimativa do seu erro padrão. A razão resultante, sob a hipótese de que $\beta=0$, terá uma distribuição normal padrão (HOSMER; LEMESHOW, 1999).

Considerando o quadrado desta razão:

$$W^2 = \left(\frac{\hat{\beta}_i}{EP(\hat{\beta}_i)} \right)^2 \sim \chi_1^2$$

onde:

$\hat{\beta}_i$ = coeficiente;

$EP(\hat{\beta}_i)$ = erro padrão do coeficiente;

e W^2 segue uma distribuição do Qui-quadrado com 1 grau de liberdade.

Também pode ser realizado o teste da razão da verossimilhanças parcial para a significância dos coeficientes do modelo de Cox. Sendo a estatística do teste dada por:

$$\chi^2 = 2 [\ln L(0) - \ln L(\beta)]$$

onde:

$\ln L(0)$ = logaritmo da verossimilhança do modelo básico (sem covariáveis);

$\ln L(\beta)$ = logaritmo da verossimilhança do modelo a ser testado (com covariável).

Sob a hipótese nula, de que os coeficientes são iguais a zero, a estatística do teste segue uma distribuição do Qui-quadrado, com um grau de liberdade, para uma covariável contínua, e n-1 graus de liberdade, se a covariável for categórica, sendo n a quantidade de categorias ou faixas em que essa covariável é dividida (HOSMER; LEMESHOW, 1999). Para a comparação dos modelos utiliza-se o teste da razão da verossimilhanças parcial.

A estatística do teste é:

$$\chi^2 = 2 [\ln L(\text{modelo maior}) - \ln L(\text{modelo menor})]$$

que segue uma distribuição de χ^2 com o número de graus de liberdade igual a diferença do número de covariáveis entre os dois modelos, onde $\ln L$ é o valor do logaritmo da verossimilhança de cada modelo ajustado (FISHER; VAN BELLE, 1995).

Para verificar a proporcionalidade de riscos pode-se utilizar a técnica gráfica, que consiste em dividir os dados em “m” faixas de acordo com a covariável. Após estima-se $\hat{H}_0(t)$ para cada faixa e constrói-se o gráfico $\hat{H}_0(t)$ versus $t, j = 1, \dots, m$. Se a suposição for válida, estas curvas devem ser aproximadamente paralelas no tempo. Curvas não paralelas significam desvios da suposição de riscos proporcionais. Ou, pode-se testar se a covariável é dependente do tempo.

Para examinar a suposição de riscos proporcionais, Cox, *apud* Colosimo (2001) propôs o teste que consiste em acrescentar ao modelo uma covariável dependente do tempo. O interesse é verificar se a razão das taxas de falhas é a mesma em qualquer tempo t .

O modelo de Cox para esta situação é

$$h(t) = h_0(t)\exp(X_1\beta_1)$$

O risco relativo de falha em qualquer tempo t da covariável X_1 é $\exp(\beta_1)$, se o modelo for adequado para os dados.

Uma outra covariável $X_2 = t$ pode ser adicionada ao modelo

$$h(t) = h_0(t)\exp(X_1\beta_1 + t\beta_2)$$

sendo que o risco relativo é agora $\exp(\beta_1 + t\beta_2)$ e, portanto, não é mais constante no tempo e nem o modelo é mais de riscos proporcionais. Em particular, se $\beta_2 < 0$, o risco relativo decresce com o tempo. Se $\beta_2 > 0$, o risco relativo de falha aumenta com o tempo. No caso em que $\beta_2 = 0$, o risco relativo é constante e igual a $\exp(\beta_1)$, a regressão não é significativa, mostrando que esta hipótese corresponde à suposição de riscos

proporcionais (COLOSIMO, 2001). Para comparar duas ou mais curvas de sobrevivência, pode-se utilizar o teste log-rank ou Qui-quadrado.

Um teste aproximado para a igualdade das duas funções de sobrevivência, pode ser baseado na estatística:

$$T^2 = \frac{\left(\sum_{j=1}^k d_{1j} - \sum_{j=1}^k e_{1j} \right)^2}{\sum_{j=1}^k V_j} \sim \chi^2_{(1)}$$

que é a estatística do teste log-rank, tendo distribuição Qui-quadrado, com 1 grau de liberdade.

O teste log-rank também pode ser utilizado para comparar três ou mais curvas de sobrevivência (ARMITAGE; BERRY, 1994). Para realizar a análise de sobrevivência em dados reais foram utilizados dados de prontuários, obtidos do Hospital Universitário de Santa Maria - HUSM, referentes a pacientes atendidos e acompanhados no período de outubro de 1991 a abril de 1999. Estes pacientes eram portadores de diagnóstico de câncer no aparelho digestivo e foram submetidos a procedimento cirúrgico (data de entrada do paciente no estudo), conforme a avaliação do caso. Com isso, a amostra foi de conveniência (não-probabilística).

O evento de interesse do estudo foi a morte desses pacientes por complicações relacionadas à doença. A morte não causada pela doença foi considerada censura. As covariáveis avaliadas na internação e logo após a cirurgia foram: idade, tipo de cirurgia, estágio da doença, tipo de tratamento cirúrgico e data de óbito ou de censura. Os tipos de cirurgias foram: gastrectomia subtotal, gastrectomia total e irressecável. Quanto ao estadiamento, os pacientes foram classificados conforme os estágios I, II, III ou IV, dependendo do estágio da doença. O tipo de tratamento cirúrgico foi classificado em: curativo ou paliativo.

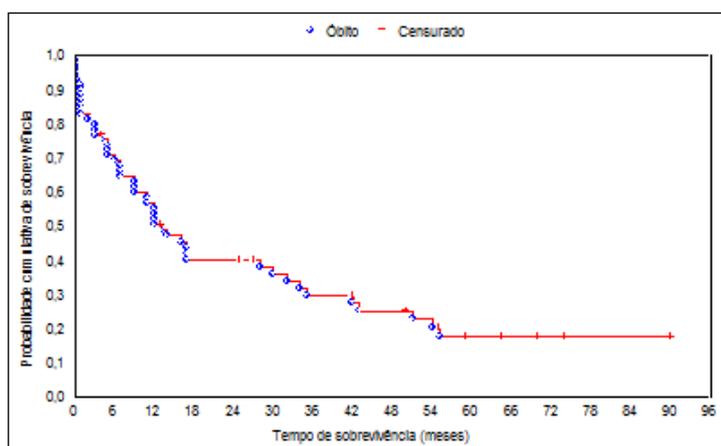
Para a análise dos dados utilizou-se o estimador de Kaplan-Meier e o modelo de regressão de Cox. Para a análise estatística foram usados os aplicativos computacionais SPSS 10.0 e Statistica 5.0. Não foram considerados os aspectos éticos, pois na época (ano de 2004) esses aspectos não eram observados.

3 Resultados

A amostra dos 69 pacientes foi observada por, aproximadamente, 7,5 anos (tempo máximo do estudo), sendo que o tempo mediano de sobrevida desses pacientes foi de um ano. O momento de entrada do paciente no grupo de estudo, foi a data da cirurgia e como tempo final, foi considerada a data do óbito do paciente ou a data do término do estudo (paciente censurado). A média de idade dos pacientes foi de 60 anos, sendo a mínima de 23 e a máxima de 80 anos e o desvio padrão de 12,3 anos.

A função de sobrevida acumulada foi estimada pelo método de Kaplan-Meier e a curva de sobrevivência dos 69 pacientes pode ser observada na Figura 1.

Figura 1 - Curva de sobrevivência global



Pode-se observar na Figura 1 que, para os 69 pacientes estudados, a curva de sobrevivência estimada cai para 17,6% a partir do tempo de sobrevivência de 55 meses, permanecendo este mesmo valor até 90 meses (7,5 anos), pois neste período não ocorreu mais nenhum óbito e o tempo mediano de sobrevivência foi de 12 meses.

A probabilidade de sobreviver mais de 12 meses, dos pacientes submetidos à gastrectomia subtotal foi de 57,1%, com tempo mediano de sobrevivência de 17 meses. Os pacientes submetidos à gastrectomia total tiveram 58,2% de probabilidade de sobreviver até 12 meses e tempo mediano de sobrevida de aproximadamente 13 meses, enquanto que os submetidos à cirurgia irresssecável tiveram 17,6% de probabilidade de sobrevida em 12 meses e apresentaram um tempo mediano de sobrevida inferior aos outros dois grupos. O resultado do teste do Qui-quadrado para

a comparação das curvas, não foi significativo, pois $\chi_c^2=4,74$ (2 gl; $p=0,094$), concluindo-se que não existe diferença significativa entre as curvas de sobrevivência, por tipo de cirurgia.

Para os pacientes classificados no estágio I a probabilidade de sobrevivência em um ano foi 77,8%. No estágio II, a probabilidade de sobrevivência em 16 meses foi de 71,4%. Já, no estágio III, a probabilidade de sobrevida em 13 meses foi 44,9%. Os pacientes classificados no estágio IV apresentaram probabilidade de sobrevivência em 12 meses de 7,7%, sendo este resultado já esperado, pois este é o estágio mais avançado da doença. O resultado do teste do Qui-quadrado foi $\chi_c^2=19,03$ (3 gl; $p=0,000$), sendo significativa a diferença entre as curvas de sobrevivência por estágio da doença. Para identificar quais as curvas que diferiam entre si, foi utilizado o teste log-rank, que mostrou que a curva do estágio I diferia da curva do estágio III ($p=0,003$) e da curva do estágio IV ($p=0,001$) e, a curva do estágio II, diferia da curva do estágio IV ($p=0,004$), sendo as demais comparações não-significativas, ao nível de 5%.

Para o grupo de pacientes que realizaram tratamento cirúrgico curativo, a probabilidade de sobrevida em 12 meses foi de 62,6%. Os pacientes que realizaram tratamento cirúrgico paliativo tiveram probabilidade de sobrevivência de 7,7%, em 12 meses. O teste log-rank mostrou resultado significativo ($p=0,003$), isto é, existe diferença significativa entre as curvas de sobrevivência por tipo de tratamento. Para mais detalhes da análise descritiva e da análise de sobrevivência não-paramétrica ver Chiesa (2002). Para o ajuste do modelo de regressão de Cox, inicialmente realizou-se uma análise simples das covariáveis. Como resultado apenas a covariável idade não foi significativa. Assim, considerou-se o modelo múltiplo somente com as covariáveis significativas no modelo univariado. Como a covariável tratamento possui uma relação linear com a covariável estágio ela foi retirada do modelo, sendo mantida a covariável estágio pelo seu maior grau de importância no modelo.

Na Tabela 2 são mostrados os resultados obtidos no modelo de regressão de Cox, para a análise múltipla das covariáveis, cirurgia e estágio.

Tabela 2 – Modelo de regressão de Cox - covariáveis cirurgia e estágio

Covariável	β	SE	W ²	gl	p	exp β
Gastrectomia subtotal			0,507	2	0,776	
Gastrectomia total	0,303	0,434	0,487	1	0,485	1,354
Irressecável	0,036	0,485	0,006	1	0,941	1,037
Estágio I			14,332	3	0,002	
Estágio II	1,165	0,896	1,689	1	0,194	3,205
Estágio III	1,763	0,746	5,589	1	0,018	5,829
Estágio IV	2,695	0,795	11,502	1	0,001	14,804

Observa-se pelo teste de Wald que a covariável cirurgia não é significativa ($p > 0,05$) no modelo. Pelo teste da razão de verossimilhanças, com as covariáveis cirurgia e estágio no modelo, obteve-se $\chi^2 = 20,151$ ($p = 0,001$), resultado significativo. Porém, pelo teste de Wald, a covariável cirurgia não é significativa, indicando que esta não contribui para o modelo.

A análise múltipla aponta a covariável estágio como sendo a que oferece a melhor contribuição para a modelagem dos dados em estudo, concluindo-se, assim, que este é o modelo que melhor se ajusta aos dados. Na Tabela 3 é apresentado o modelo final, a um nível de significância de 5%.

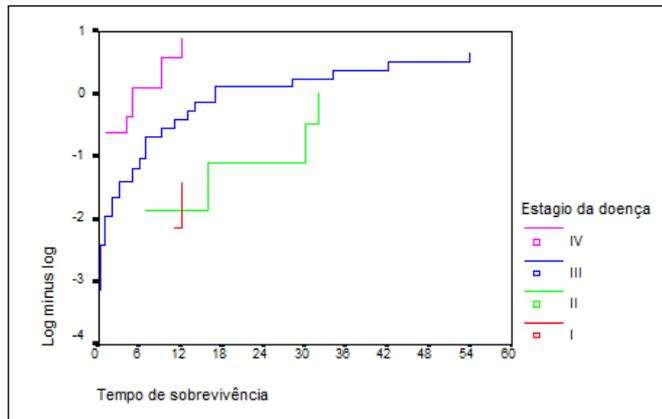
Tabela 3 – Modelo final

Estágio	β	SE	W	gl	p	exp β	IC 95%
I (X ₁)			15,46	3	0,001		
II (X ₂)	1,04	0,87	1,43	1	0,232	2,82	0,52; 15,42
III (X ₃)	1,79	0,74	5,78	1	0,016	5,99	1,39; 25,73
IV (X ₄)	2,60	0,77	11,37	1	0,001	13,46	2,97; 61,04

Nesta fase da análise é necessário verificar a proporcionalidade dos riscos no tempo (teste gráfico).

Na Figura 2 são apresentadas as curvas, por categoria, da covariável estágio. Devido ao pequeno tamanho da amostra a conclusão gráfica sobre a proporcionalidade dos riscos relativos fica prejudicada, mas pode-se dizer que os riscos são, aproximadamente, proporcionais.

Figura 2 - Curvas comparativas por estágio da doença



Pode-se também testar a proporcionalidade dos riscos pelo teste de dependência no tempo, para verificar a pressuposição de riscos proporcionais com maior segurança.

Para realizar esse teste foi criada uma nova covariável, dependente do tempo definida como:

$$\text{deptemp} = X (\ln t - 2,17)$$

sendo:

X = covariável estágio;

$\ln t$ = logaritmo neperiano do tempo de sobrevivência;

2,17 = média do $\ln t$.

Os resultados são apresentados na Tabela 4.

Tabela 4 - Modelo com a covariável dependente do tempo

Covariável	β	SE	W^2	gl	p	$\exp\beta$
Estágio	0,787	0,196	16,10	1	0,000	2,198
deptemp	-0,075	0,124	0,366	1	0,545	0,928

Na Tabela 4 são apresentados os resultados do teste, sendo que a variável dependente do tempo não é significativa ($p=0,545$), não contribuindo para o modelo. Os riscos se mantêm constantes no tempo, atendendo a pressuposição da metodologia da regressão de Cox, com $\beta_2 = -0,075$, ou seja, $\beta_2 \cong 0$, indicando que o modelo é um modelo de riscos proporcionais.

Assim, pode-se afirmar que entre os modelos avaliados, o modelo que inclui apenas a covariável estágio é o de melhor ajuste dos dados.

A função de sobrevivência do modelo é:

$$\hat{S}(t) = \hat{S}_0(t)^{\exp(1,036X_2+1,790X_3+2,600X_4)}$$

A função razão de riscos relativos do modelo é:

$$HR = \exp(1,036X_2+1,790X_3+2,600X_4)$$

O valor numérico da razão de riscos relativos de óbito mostra o estágio I (X_1) como a faixa de referência. Os pacientes classificados no segundo estágio (X_2), têm um risco de óbito 2,8 vezes maior do que os pacientes do estágio I. Os pacientes do estágio III (X_3) têm 6,0 vezes mais risco de óbito que pacientes do estágio I. E, finalmente, os pacientes do estágio IV (X_4), tem 13,5 vezes mais risco de óbito que os pacientes do estágio I da doença, conforme os valores de $\exp(\beta)$ da Tabela 3.

À época, o objetivo deste trabalho, que resultou em uma monografia de especialização, era realizar um estudo sobre as técnicas de análise de sobrevivência, principalmente, utilizando o modelo de regressão semi-paramétrico de Cox, com uma aplicação a dados da área médica. Com isso, não foi objetivo discutir os resultados com outras publicações da área médica sobre câncer do aparelho digestivo, e sim utilizar a técnica em dados reais utilizando um aplicativo computacional, sem foco na discussão.

4 Conclusões

Utilizando-se o modelo de regressão de Cox, a covariável estágio foi a única a apresentar resultado significativo no modelo ($p=0,001$), sendo esta uma covariável categórica, ou seja, avaliada em faixas de risco. Pode-se concluir que quanto maior o estágio da doença maior o risco de óbito, sendo este resultado bem claro, tanto na modelagem de Cox, quanto na estimação por Kaplan-Meier, ou seja, há uma clara concordância entre as duas análises, o que era esperado.

Embora algumas covariáveis apresentem diferenças significativas pelo teste log-rank na comparação das curvas de sobrevivência estimadas por Kaplan-Meier, como é

o caso do tipo de tratamento, quando estas variáveis são colocadas juntas, no modelo múltiplo de Cox, algumas perdem a significância estatística, ou poder de estimação devido a contribuições competitivas, neste caso a relação entre tratamento e estágio era de perfeita dependência linear, não havendo sentido algum em manter as duas covariáveis no modelo.

A covariável idade não se mostrou significativa em relação ao risco de óbito, possivelmente por tratar-se de um conjunto de pacientes com idade em geral, mais avançada, cuja média foi de 60 anos.

Em função do pequeno tamanho da amostra disponível para este estudo sugere-se que estes resultados sejam revistos com um número maior de pacientes, utilizando-se esta metodologia e que este texto possa ser útil como referência para futuros estudos.

Referências

ARMITAGE, P & BERRY, G. **Statistical Methods in Medical Research**. 3 ed. Blackwell Science Ltda. USA, 1994.

BUSTAMANTE-TEIXEIRA MT, FAERSTEIN E, LATORRE MR. Técnicas de análise de sobrevivência. **Cadernos de Saúde Pública**, Rio de Janeiro, 18(3):579-594, mai-jun, 2002.

CARVALHO MS, ANDREOZZI VL, CODECO CL, et al. **Análise de Sobrevivência: Teoria e Aplicações em Saúde**. Rio de Janeiro: Editora Fiocruz, 2005. 295p.

CHIESA, EMF. **Análise de Sobrevivência através de um método não paramétrico** [monography]. Santa Maria: Curso de Especialização em Estatística e Modelagem Quantitativa /UFSM; 2002.

COLOSIMO, EA. Análise de Sobrevivência Aplicada. In: **Anais 9º, SEAGRO, REUNIÃO ANUAL DA RBRAS**, 46ª, Piracicaba, SP. Brasil. 2001.

COLOSIMO EA, GIOLO SR. **Análise de sobrevivência aplicada**. São Paulo: Edgard Blücher, 2006. 370p.

COX, DR. Regression Models and Life Tables. **Journal of the Royal Statistical Society B**, 1972; 34, 187-220.

COX, DR. Partial likelihood. **Biometrika**, 1975; 62, 269-276.

FERRAZ RO, MOREIRA-FILHO DC. Análise de sobrevivência de mulheres com câncer de mama: modelos de riscos competitivos. **Ciência & Saúde Coletiva**, 22 (11) 3743-3753, 2017.

FISCHER, LD & BELLE, GV. Biostatistics – A Methodology for the Health Sciences. **New York: John Wiley & Sons, Inc**, 1995.

HOSMER, DW, & LEMESHOW, S. Applied Survival Analysis - Regression Modeling of Time to Event Data. **New York: John Wiley & Sons, Inc**, 1999.

KAPLAN, EL & MEIER, P. Nonparametric estimation for incomplete observations. **Journal of the American Statistical Association**, 1958; v. 53, n.282, p.457-481.