



Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

Análise do tempo até o alívio da dor em pacientes com dor lombar crônica via modelo de riscos proporcionais de Cox

Jéssica Ferreira Silva

Brasília - DF, 16 de dezembro de 2015

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

Análise do tempo até o alívio da dor em pacientes com dor lombar crônica não específica via modelo de riscos proporcionais de Cox

Discente: Jéssica Ferreira Silva
Orientador: Prof. Dr. Eduardo Yoshio Nakano

Relatório apresentado à disciplina Trabalho de conclusão de curso 2 do curso de graduação em estatística, Departamento de Estatística, Instituto de Exatas, Universidade de Brasília, como parte dos requisitos necessários para o grau de Bacharel em Estatística.

Brasília - DF, 16 de dezembro de 2015

1 Agradecimentos

Agradeço ao professor Eduardo Nakano por toda a paciência, dedicação e disponibilidade de sempre me ajudar para a realização deste trabalho.

Agradeço à minha família que sempre foi a minha base e me ajudou em todos os momentos, a minha mãe que sempre foi mãe e pai e me ensinou que conhecimento é algo que ninguém pode nos roubar, em especial a minha tia Lourdes que mesmo não estando mais nesse plano, lembrar de você dizendo "já deu tudo certo" me incentivou a nunca desistir, a minha avó que me ensinou o significado de garra.

A minha irmã que sempre foi a minha amiga e companheira, ao meu padrasto e meu padrinho que em vários momentos foram um pai pra mim e aos meus amigos que foram pequenos anjos ao longo desses anos.

2 Resumo

Análise do tempo até o alívio da dor em pacientes com dor lombar crônica não específica via modelo de riscos proporcionais de Cox

O modelo de Cox é um dos modelos mais citados dentro da literatura de análise de sobrevivência e este trabalho visa uma modelagem dos dados de pessoas que apresentam dor lombar crônica não específica, que foram separadas em dois grupos, denominados: Eletropilates Ativo e Eletropilates Placebo.

Os pacientes alocados no grupo Eletropilates Ativo são submetidos ao método Pilates e a corrente interferencial ativa e o grupo Eletropilates Placebo são submetidos ao método Pilates combinado com o método corrente interferencial placebo. O objetivo é constatar diferenças significativas entre os dois tratamentos, visando observar essas diferenças o alívio da dor é analisado levando em consideração uma redução de 50 e de 30% da dor em relação a dor inicial.

As informações são referentes a 137 pacientes e as covariáveis analisadas são: idade, IMC e grupo de tratamento. Dos 137 pacientes, 69 estão alocados no grupo eletropilates ativo e 68 no grupo eletropilates placebo, porém esses pacientes não sabem diferenciar o tratamento que foram submetidos.

A censura neste estudo ocorrer em duas situações, quando não foi possível acompanhar o paciente até o fim do estudo ou ao fim do estudo o paciente não observa o alívio da dor em relação as porcentagens definidas.

Palavras-chave: Análise de sobrevivência; Modelos de riscos proporcionais de Cox, Dor lombar crônica não específica; Eletropilates.

Sumário

1	Agradecimentos	3
2	Resumo	4
3	Introdução	6
4	Revisão Bibliográfica	8
4.1	Introdução	8
4.2	Variável resposta	8
4.2.1	Tempo de falha	9
4.2.2	Tempo de censura	9
4.3	Representação dos dados de sobrevivência	10
4.4	Tempo de sobrevivência	10
4.5	Função de sobrevivência	10
4.6	Função taxa de falha ou de Risco	11
4.6.1	Relações entre as funções	12
4.7	Estimação da função de sobrevivência	12
4.7.1	Estimador de Kaplan-Meier	13
4.8	Modelo de Regressão de Cox	14
4.8.1	Ajustando o modelo de COX	14
4.8.2	Estimação dos parâmetros	15
4.9	Adequação do Modelo de COX	16
5	Base de dados	17
6	Resultados	19
6.1	Análise descritiva	19
6.2	Modelo de riscos proporcionais de Cox	23
6.3	Conclusões	26
7	Referências	27

3 Introdução

A dor lombar crônica não específica é uma afecção musculoesquelética que apresenta um tempo de duração maior que 12 semanas. Segundo Franco (2014), o tratamento indicado nestes casos são os exercícios supervisionados. Há também diretrizes que direcionam a prática clínica para tratar o problema, porém ambas as diretrizes não indicam quais são os exercícios que apresentam os melhores resultados.

Dentre as diversas modalidades de exercícios existentes na Fisioterapia, aquelas que utilizam a metodologia Pilates (LATEY, 2011) apresentam comprovação científica que são eficientes para a redução da dor em pacientes com dor lombar crônica não específica. No entanto, o Pilates não é o único tratamento indicado para esse fim. Um outro método eficaz para o combate da dor lombar crônica não específica é a Corrente Interferencial.

Segundo Franco (2014), ainda não existem estudos que comparem os resultados da combinação desses dois tratamentos com o uso isolado de cada uma delas. No entanto, esse mesmo autor avaliou a efetividade da adição da Corrente Interferencial aos exercícios do método Pilates no tratamento para pacientes com dor lombar crônica.

O estudo foi conduzido em dois grupos, denominados: Eletropilates Ativo e Eletropilates Placebo. No grupo Eletropilates Ativo os pacientes eram submetidos a exercícios do método Pilates combinado com o método da corrente interferencial ativa e o grupo Eletropilates Placebo os pacientes eram submetidos a exercícios do método Pilates combinado com o método corrente interferencial placebo.

Os pacientes foram acompanhados por um período de 18 sessões de tratamento, essas 18 sessões foram realizadas em 6 semanas e o evento de interesse é se houve ou não o alívio da dor nesses pacientes. Neste trabalho, a variável resposta é o número de sessões necessárias até o alívio ou a diminuição da dor, a diminuição é caracterizada como a redução em no mínimo 50% da dor, em relação a dor no início do tratamento.

Neste estudo, os dois grupos citados apresentaram melhora significativa no alívio da dor e os pacientes não foram capazes de distinguir se estavam sendo submetidos a método corrente interferencial ativa ou corrente interferencial placebo. Nesse contexto, a proposta desse projeto é analisar o efeito do Eletropilates no alívio da dor utilizando o modelo de regressão de Cox (Cox, 1972). O "tempo" até o alívio da dor será o número de sessões que cada paciente foi submetido até ser observado o evento de interesse, que é a diminuição ou alívio da dor. Outros trabalhos que apresentam aplicações do modelo de riscos proporcionais de Cox podem ser vistos em Nakano e Cunha (2012), Santos e Nakano (2015), entre outros.

Diferentemente do trabalho de Franco (2014) (em que a diminuição da dor é caracterizada como a redução em no mínimo 50% da dor), a proposta desse projeto foi alterar a definição de diminuição da dor para uma redução de 50% e 30% da dor. Essa mudança na definição do evento de interesse visa aumentar a sensibilidade da análise e, conseqüentemente, comprovar uma possível diferença significativa entre os dois grupos estudados. Ademais o uso do modelo de riscos proporcionais de Cox permitirá utilizar na análise algumas covariáveis que não foram consideradas do trabalho de Franco (2014). A análise será baseada em uma amostra de 137 pacientes com dor lombar crônica e considerando as covariáveis idade, IMC (índice de massa corporea), sexo, tempo de dor (em semanas) e grupo de tratamento. Todas as análises realizadas foram realizadas pelo software livre R (R CORE TEAM, 2015).

4 Revisão Bibliográfica

4.1 Introdução

A palavra estatística está associada à palavra latina STATUS (Estado). As primeiras aplicações da estatística estavam voltadas para a descrição dos bens do estado, na formulação de políticas públicas, fornecendo dados demográficos e econômicos. O campo de atuação da estatística deixou de ser limitado apenas à demografia no começo do século XIX, ao ser introduzido aos métodos matemáticos da estatística a Teoria das Probabilidades.

A Estatística está presente em todas as áreas da ciência que envolvam o planejamento do experimento, a construção de modelos, a coleta, o processamento e a análise de dados, para postular, refutar ou validar hipóteses sobre um fenômeno observável. Assim, a Estatística pode ser conceituada como a ciência de aprendizagem a partir de dados. Com a alta competitividade na busca de tecnologias e de mercados tem provocado uma constante corrida pela informação e torna o aprendizado a partir de dados em alta na era da informação em que vivemos. Em linhas gerais, podemos dizer que a Estatística, com base na Teoria das Probabilidades, fornece técnicas e métodos de análise de dados que auxiliam o processo de tomada de decisão nos mais variados problemas onde existe incerteza.

As técnicas de sobrevivência é uma das áreas da estatística que mais cresceu nas últimas décadas do século passado. Uma evidência desse sucesso é o número de aplicações da Análise de Sobrevivência na medicina, o uso deste técnica cresceu de 11% em 1979, para 32% em 1989, sendo a área da estatística que mais se destacou no período avaliado (COLOSIMO; GIOLO, 2006).

Em análise de sobrevivência três fatores precisam estar bem definidos logo no início de qualquer estudo, esses itens são: o instante inicial, a escala de medida e o evento de interesse. O instante inicial é o momento marcado no tempo em que o estudo teve início, que pode não ser tão bem definido dependendo se é um estudo experimental ou um estudo de coorte. A escala de medida é a medição do tempo, podendo ser classificada em dias, meses, anos. E o evento de interesse é o que está sendo analisado no seu estudo.

4.2 Variável resposta

O tempo ou variações do tempo é a variável resposta para grande parte dos estudos em análise de sobrevivência. A principal característica que diferencia a técnica de Análise de Sobrevivência das técnicas convencionais é a presença de censura, que é basicamente a observação parcial da resposta, ou

seja, por alguma razão o relacionamento do cliente observado foi interrompido antes do final do estudo. Tempo não pode ser medido no instante exato de ocorrência do evento de interesse.

4.2.1 Tempo de falha

O tempo de falha que é definido como o tempo que decorreu desde o instante inicial do estudo até a ocorrência do evento de interesse. Esses eventos são, na maioria dos casos, indesejáveis e usualmente denominado falha.

4.2.2 Tempo de censura

Tempo de censura é a observação parcial da variável resposta. Ainda assim, os dados censurados devem ser incluídos na análise pois eles fornecem informações sobre o tempo de vida de indivíduos e a omissão deles pode fazer com que conclusões viciadas sejam feitas. A censura pode ser subdividida em censura à direita, censura à esquerda e censura aleatória.

Censura à esquerda

A censura à esquerda ocorre quando o tempo de vida é menor que o tempo observado, ou seja, o evento de interesse já ter ocorrido quando o indivíduo começou a fazer parte do estudo.

Censura à direita

Ocorre quando o tempo de ocorrência do evento de interesse está à direita do evento de interesse. Isto significa que toda a informação referente à resposta se resume ao conhecimento de que o tempo de falha é superior àquele observado. A censura à direita pode ser classificada como:

1. Censura tipo I: Consiste em preestabelecer um período de tempo para cada um dos itens em estudo, de maneira que o tempo de vida é conhecido apenas se for menor que o período. Esse tipo de censura é usada quando se pode planejar o tempo de duração de um estudo.
2. Censura tipo II: Consiste em estudos que terminam após ter ocorrido a falha de um número preestabelecido de itens sob teste. Esse tipo de censura é utilizado quando se tem pouco ou nenhuma informação do tempo de vida do seu objeto de estudo.
3. Censura aleatória: Consiste em um item ser retirado durante o estudo sem ter atingido o tempo de falha, podendo ocorrer se o item falha por uma razão diferente da estudada. É um tipo de censura do tipo I.

Censura intervalar

Ocorre em estudos em que os elementos tem acompanhamento periódico. O evento de interesse ocorreu entre esses períodos, porém não é possível determinar o instante exato.

4.3 Representação dos dados de sobrevivência

Em análise de sobrevivência a informação da variável resposta associada a cada indivíduo é representada pelo par (t_i, δ_i) , sendo que $i=1, \dots, n$ indica o número de indivíduos no estudo, t_i é o tempo de falha ou censura e δ_i é a variável de indicadora de falha ou censura, que é igual a 1 se t_i é tempo de falha e 0 se t_i é tempo de censura.

4.4 Tempo de sobrevivência

O tempo de vida do indivíduo, conhecido como tempo de sobrevivência, é representado pela variável não negativa T , geralmente contínua e de natureza longitudinal. O tempo de sobrevivência, T_s , é dado por:

$$T_s = T_f - T_i,$$

onde T_f é o instante onde fica observado o evento de interesse ou o indivíduo é censurado. T_i é o momento onde o elemento deu entrada no estudo.

4.5 Função de sobrevivência

A função de sobrevivência é uma das principais funções probabilísticas usadas para descrever estudos de sobrevivência. A função de sobrevivência é definida como a probabilidade de uma observação não falhar até um certo tempo t , ou seja, a probabilidade de uma observação sobreviver ao tempo t . Em termos probabilísticos, para T contínua, isto é escrito como:

$$S(t) = P(T > t) = \int_t^{\infty} f(x)dx \quad (1)$$

Trabalhos que consideram o tempo de sobrevivência, T , discreto podem ser vistos em Nakano e Carrasco (2006), Brunello e Nakano (2015), entre outros. Em consequência, a função de distribuição acumulada é definida como a probabilidade de uma observação não sobreviver ao tempo t , isto é escrito como:

$$F(t) = 1 - S(t).$$

A função de sobrevivência é uma função não crescente no tempo com as propriedades de que a probabilidade de sobreviver pelo menos ao tempo zero é 1 e a probabilidade de sobreviver no tempo infinito é 0.

Para descrever a função de sobrevivência é geralmente utilizada uma representação gráfica de $S(t)$, ou seja, o gráfico de $S(t)$ versus t que é chamado de curva de sobrevivência. Uma curva íngreme representa razão de sobrevivência baixo ou curto tempo de sobrevivência e uma curva de sobrevivência gradual ou plana representam taxa de sobrevivência alta ou sobrevivência longa.

4.6 Função taxa de falha ou de Risco

A probabilidade de falhar ocorre em um intervalo de tempo $[t_1, t_2)$ pode ser expressa em termos da função de sobrevivência como:

$$S(t_1) - S(t_2) \quad (2)$$

A taxa de falha ou taxa de risco no intervalo $[t_1, t_2)$ é definida como a probabilidade de que a falha ocorre neste intervalo, dado que não ocorreu antes de t_1 dividida pelo comprimento do intervalo. Assim, a taxa de falha no intervalo $[t_1, t_2)$ é expressa por:

$$\frac{S(t_1) - S(t_2)}{(t_2 - t_1)S(t_1)} \quad (3)$$

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

$$\lambda(t) = \frac{S(t) - S(t + \Delta t)}{\Delta t S(t)} \quad (4)$$

Assumindo Δt bem pequeno, $\lambda(t)$ representa a taxa de falha instantânea no tempo t condicional à sobrevivência até o tempo t . A taxa de falha é útil para descrever a distribuição do tempo de vida.

A taxa de falha é um importante método para os dados de sobrevivência, pois pode ter forma crescente, decrescente ou constante. A função crescente indica que a taxa de falha do paciente aumenta com o transcorrer do tempo, a função constante indica que a taxa de falha não se altera com o passar do tempo e a função decrescente mostra que a taxa de falha diminui a medida que o tempo passa.

A função de risco é mais informativa do que a função de sobrevivência. Diferentes funções de sobrevivência podem ter formas semelhantes, enquanto as respectivas funções de taxa de falha podem diferir drasticamente.

Outra função útil em análise de dados de sobrevivência é a função de taxa de falha acumulada. Essa função fornece a taxa de falha acumulada do indivíduo e é definida por:

$$\Lambda(t) = \int_0^t \lambda(u) du \quad (5)$$

4.6.1 Relações entre as funções

As função de sobrevivência, função de taxa de falha, função de taxa de falha acumulada e a função de densidade de probabilidade, são matematicamente relacionadas. Essa relação é útil no processo de estimação e quando somente o valor de uma das funções é conhecida e deseja-se obter as outras funções. Logo, para T uma variável aleatória contínua e não-negativa, tem-se, as seguintes relações matemáticas definidas:

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt}(\log S(t)), \quad (6)$$

$$\Lambda(t) = \int_0^t \lambda(u) du = -\log(S(t)),$$

$$\text{e } S(t) = \exp\{-\Lambda(t)\} = \exp\left\{-\int_0^t \lambda(u) du\right\}.$$

4.7 Estimação da função de sobrevivência

A primeira parte em estudos de tempo de vida é usualmente a estimação função de sobrevivência. A presença de observações censuradas, requer técnicas estatísticas especializadas para acomodar a informação contida nestas observações. Algumas técnicas estatísticas podem ser utilizadas para analisar dados de tempo de sobrevivência na presença de censura. Podem ser citadas três estimadores não-paramétricos usados para estimação da função de sobrevivência, sendo eles:

1. Estimador de Kaplan-Meier,
2. Estimador de Nelson Aalen

3. Estimador da Tabela de Vida

Estes estimadores são conhecidos como não-paramétricos, pois usam os próprios dados para estimar as quantidades necessárias da análise, sem fazer uso de suposições a respeito da forma da distribuição dos tempos de sobrevivência. Um dos objetivos desse trabalho é fazer uma comparação entre os resultados obtidos no Modelo de Regressão de COX e o número de sessões necessárias até observar o alívio da dor no Estimador de Kaplan-Meier.

4.7.1 Estimador de Kaplan-Meier

O estimador limite-produto, ou Kaplan-Meier, como é usualmente chamado, é um estimador não-paramétrico para a função de sobrevivência. Essa estimativa de $S(t)$ é uma função escada com degraus nos tempos observados de falha.

O estimador de Kaplan-Meier, na sua construção, considera tantos intervalos de tempo quanto os número distintos de falha. Os limites dos intervalos de tempo são os tempos distintos de falha. Assumindo:

1. $t_1 < t_2 < \dots < t_k$, os k tempos distintos e ordenados de falha,
2. d_j o número de falhas em t_j , $j=1, \dots, k$ e
3. n_j o número de indivíduos sob risco em t_j , ou seja, os indivíduos que não falharam e não foram censurados até o instante imediatamente anterior a t_j .

O estimador é, então definido como:

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

Ele possui as seguintes propriedades:

1. Não é viciado para amostras grandes;
2. É fracamente consistente;
3. Converge assintoticamente para um processo gaussiano; e
4. É um estimador de máxima verossimilhança de $\hat{S}(t)$

Um intervalo aproximado de $100(1-\alpha)\%$ de confiança de $\hat{S}(t)$ é dado por:

$$\hat{S}(t) \pm z_{\alpha/2} \sqrt{\text{Var}(\hat{S}(t))}$$

onde

$$\text{Var}(\hat{S}(t)) = [\hat{S}(t)]^2 \sum_{j:t_j < t} \left(\frac{d_j}{n_j(n_j - d_j)} \right).$$

Aqui $z_{\alpha/2}$ é o quantil $\alpha/2$ de uma distribuição normal padrão.

4.8 Modelo de Regressão de Cox

O modelo de regressão de COX permite a análise de dados provenientes de estudos de tempo de vida em que a resposta é o tempo até o ocorrência de um evento de interesse, ajustado por covariáveis. Considere p covariáveis, de modo que x seja um vetor com componentes $x = (x_1, x_2, \dots, x_p)$. A expressão geral do modelo de regressão de COX considera:

$$\lambda(t) = \lambda_0(t)g(x'\beta)$$

em que $g(t)$ é uma função não-negativa que deve ser especificada, tal que $g(0) = 1$. Este modelo é composto pelo produto de dois componentes, um não-paramétrico e o outro paramétrico. O componente não-paramétrico $\lambda_0(t)$, não é especificado e é uma função não negativo do tempo. Ele é usualmente chamado de função de base ou basal, pois $\lambda(t) = \lambda_0(t)$ quando $x=0$ (todas as covariáveis são iguais a zero). O componente paramétrico é frequentemente usado na seguinte forma multiplicativa:

$$g(x'\beta) = \exp(x'\beta) = \exp(\beta_1x_1 + \dots\beta_px_p) \quad (7)$$

em que β é o vetor de parâmetros associado às covariáveis. Esta forma garante que $\lambda(t)$ seja sempre não-negativa. Observe que a constante β_0 , presente nos modelos paramétricos, não aparece no componente mostrado em (7). Isto ocorre devido à presença do componente não paramétrico no modelo que absorve este termo constante. Este modelo também é denominado modelo de riscos proporcionais, pois a razão das taxas de falhas de dois indivíduos diferentes é constante no tempo. Isto é, a razão das funções de taxa de falha para o indivíduos i e j é dada por:

$$\frac{\lambda_i(t)}{\lambda_j(t)} = \frac{\lambda_0(t)\exp\{x'_i\beta\}}{\lambda_0(t)\exp\{x'_j\beta\}} = \exp\{x'_i\beta - x'_j\beta\}$$

, que não depende do tempo.

A suposição básica para o uso do modelo de regressão de COX é, portanto, que as taxas de falha sejam proporcionais ou, de forma equivalente para este modelo, que as taxas de falha acumulada sejam também proporcionais.

O modelo de regressão de COX é utilizado extensivamente em estudos médicos. A principal razão deste popularidade é a presença da componente não-paramétrico que torna o modelo bastante flexível.

4.8.1 Ajustando o modelo de COX

O modelo de regressão de COX é caracterizado pelos coeficientes β 's, que medem os efeitos das covariáveis sobre a função de taxa de falha. Estas

quantidades devm ser estimadas a partir das observações amostrais para que o modelo fique determinado.

Um método de estimação é necessário para se fazer inferência a cerca dos parâmetros do modelo. O método de máxima verossimilhança é bastante conhecido (COX e HINKLEY, 1974) e frequentemente utilizado para este propósito. A presença do componente não-paramétrico λ_0 na função de verossimilhança torna este método inapropriado, ou seja, sabe-se que:

$$L(\beta) = \prod_{i=1}^n [\lambda f(t_i|x_i)]^{\delta_i} S(t_i|x_i) \quad (8)$$

No modelo de Cox,

$$S(t_i|x_i) = \exp - \int_0^{t_i} \lambda_0(u) \exp\{x'_i \beta\} du = [S_0(t_i)]^{\exp\{x'_i \beta\}} \quad (9)$$

Assim, aplicando-se este resultado em (8), segue que:

$$L(\beta) = \prod_{i=1}^n [\lambda_0(t_i) \exp\{x'_i \beta\}]^{\delta_i} [S_0(t_i)]^{\exp\{x'_i \beta\}}$$

, que é a função do componente não-paramétrico $\lambda_0(t)$.

4.8.2 Estimação dos parâmetros

Para a estimação dos parâmetros do modelo, o método de máxima verossimilhança é inapropriado devido a presença do componente não-paramétrico na função de verossimilhança. Assim, o método de verossimilhança parcial foi proposto por COX para condicionar a construção da função de verossimilhança ao conhecimento da história passada de falhas e censuras para eliminar a função de risco base.

Dada uma amostra n indivíduos com $k \leq n$ falhas distintas nos tempos $t_1 < t_2 < \dots < t_k$, o conceito de verossimilhança considera o argumento de que a probabilidade condicional da i-ésima observação vir a falhar no tempo t_i conhecendo quais observações estão sob risco em t_i .

Assim, a função de verossimilhança parcial é dada por:

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

em que δ_i é o indicador de falha. Os valores de β que maximizam $L(\beta)$ são obtidos a partir de $U(\beta)=0$ onde $U(\beta)$ é o vetor escore de derivadas de

primeira ordem da função $l(\beta) = \log(L(\beta))$. Ou seja,

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

4.9 Adequação do Modelo de COX

O modelo de regressão de COX é bastante flexível devido a presença do componente não-paramétrico. Mesmo assim, ele não ajusta a qualquer situação clínica e, como qualquer outro modelo estatístico, requer o uso de técnicas para avaliar a sua adequação. Em particular, ele tem uma suposição básica que é a de riscos proporcionais, a violação desta suposição pode acarretar sérios vícios na estimação das coeficientes do modelo (STRUTHERS e KALBFLEISCH, 1986).

Existem diversos métodos para avaliar a adequação do modelo de COX, dentre eles podemos citar:

1. Avaliação da Qualidade Geral de Ajuste do Modelo

2. Avaliação da Proporcionalidade dos Riscos

- Método gráfico descritivo

- Método com coeficiente dependente do tempo

- Método com covariável dependente do tempo

3. Avaliação de Outros Aspectos do Modelo de COX

- Pontos atípicos e forma funcional das covariáveis

- Pontos influentes

5 Base de dados

Inicialmente o banco de dados tinha a informação referente a 148 pacientes, porém alguns tiveram que sair da análise por alguns motivos, como: não apresentavam dor inicial (alegavam dor 0 antes da primeira sessão) ou não tinha a informação de qual era a dor antes da primeira sessão, não havia a informação referente ao tempo de duração da dor. Sendo excluídos esses indivíduos, a análise foi feita levando em consideração 137 pacientes. As covariáveis que foram consideradas para a análise da nossa variável resposta (tempo até a redução da dor em 50% ou 30% da dor inicial) foram idade, IMC, sexo, tempo de dor e grupo de tratamento.

O número de pacientes agrupados em cada grupo, onde eletropilates ativo é a associação do método eletropilates e a corrente interferencial e o eletropilates placebo é apenas método eletropilates.

O IMC é o índice de massa corporal e trata-se do nível de gordura de cada pessoa. É uma medida de referência internacional reconhecida pela OMS (Organização mundial da saúde). O cálculo é realizado dividindo a massa da pessoa (em kg) pelo quadrado da altura (em metros).

$$IMC = \frac{peso}{altura^2}$$

No estudo as pessoas que apresentavam IMC até 29,9 que são as pessoas consideradas abaixo do peso, normaç e excesso de peso, foram agrupadas todas no grupo 0 e as que apresentam IMC maior que 29,9 são pessoas que já apresentam algum grau de obesidade, essas pessoas foram agrupadas no grupo 1.

Já a idade que foi contada em anos exatos, as pessoas foram divididas em dois grupos, onde o grupo 0 são as pessoas que apresentam até 50 anos completo e o grupo 1 são as pessoas que apresentam mais que 50 anos.

A dicotomização das variáveis quantitativas e as definições dos intervalos foram realizados considerando o risco relativo univariado de cada variável. As variáveis foram inicialmente através de seus quantis e aquelas categorias com riscos similares foram agrupadas.

As variáveis sexo e tempo de dor violaram a suposição de proporcionalidade dos riscos e portanto não foram consideradas na análise.

A Tabela 1 apresenta as frequências de observações de cada uma das três covariáveis consideradas no estudo.

Tabela 1: Resumo das covariáveis presentes no banco de dados

Grupo	Definição	Número de pacientes
0	Eletropilates placebo	69
1	Eletropilates ativo	68
Grupo	Intervalo do IMC	Número de pacientes
0	Até 29,9	105
1	Maior que 29,9	32
Grupo	Intervalo da idade	Número de pacientes
0	Até 50 anos	88
1	Maior que 50 anos	49

A censura neste caso ocorrerá em duas situações:

1. Por alguma situação adversa, não é possível acompanhar o paciente até o fim do estudo;
2. Após o fim do estudo, o paciente não constata alívio da dor.

Das 137 observações do tempo até o alívio da dor observadas, 8 (5,8%) observações foram censuradas quando o evento de interesse foi definido como diminuição de 50% da dor. Quando o evento de interesse (alívio da dor) foi definido como diminuição de 30% da dor, 3 (2,2%) das observações foram censuradas.

6 Resultados

6.1 Análise descritiva

Devido a presença de censuras, as análises descritivas do número de sessões até o alívio da dor foram obtidas por meio da função de sobrevivência estimada pelo método de Kaplan-Meier. Toda a análise apresentada nesse trabalho considerou o alívio da dor como diminuição de 50% na escala de dor. Resultados similares foram observados definindo o alívio da dor como a queda de 30% na escala da dor.

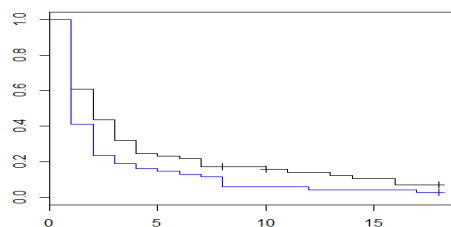
Tabela 2: Estimativa de Kaplan-Meier da função de sobrevivência, segundo grupo

Sessões	Função de sobrevivência (IC 95%)		Percentual de pacientes que apresentaram a melhora da dor	
	Eletro Ativo	Eletro Placebo	Eletro Ativo	Eletro Placebo
1	0.4118±0.0597	0.6087±0.0588	0.5882	0.3913
2	0.2353±0.0514	0.4348±0.0597	0.7647	0.5652
3	0.1912±0.0477	0.3188±0.0561	0.8088	0.6811
4	0.1618±0.0477	0.2464±0.0519	0.8302	0.7536
5	0.1471±0.0429	0.2319±0.0508	0.8529	0.7681
6	0.1324±0.0411	0.2174±0.0497	0.8676	0.7826
7	0.1176±0.0391	0.1739±0.0456	0.8823	0.8360
8	0.0588±0.0285	0.1739±0.0456	0.9411	0.8360
9	0.0588±0.0285	0.1739±0.0456	0.9411	0.8360
10	0.0588±0.0285	0.1581±0.0441	0.9411	0.8405
11	0.0588±0.0285	0.1405±0.0426	0.9411	0.8550
12	0.0441±0.0249	0.1405±0.0426	0.9558	0.8550
13	0.0441±0.0249	0.1230±0.0407	0.9558	0.8695
14	0.0441±0.0249	0.1054±0.0385	0.9558	0.8840
15	0.0441±0.0249	0.1054±0.0385	0.9558	0.8840
16	0.0441±0.0249	0.0703±0.0327	0.9558	0.9130
17	0.0294±0.0205	0.0703±0.0327	0.9705	0.9130
18	0.0294±0.0205	0.0703±0.0327	0.9705	0.9130

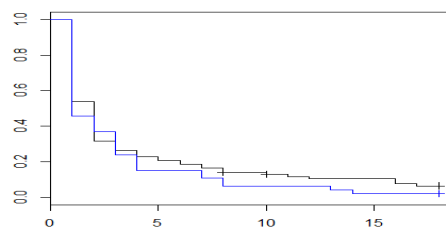
Analisando a Tabela 2, é possível notar que mais que 50% dos pacientes agrupados no grupo Eletro Ativo já apresentam a melhora significativa da dor logo na primeira sessão e os alocados no grupo Eletro Placebo apresentam essa melhora na segunda sessão. Após a oitava sessão mais de 90% dos pacientes do grupo Eletro Ativo apresentam a redução de 50% em relação a dor inicial, enquanto no grupo Eletro Placebo após as 8 sessões apenas cerca de 80% dos pacientes apresentam melhora significativa.

A Figura 1 apresenta a curva de sobrevivência estimada por Kaplan-Meier para cada uma as variáveis presentes no banco de dados.:

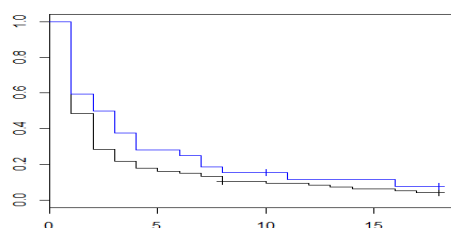
Figura 1: Estimativas da função de sobrevivência por Kaplan Meier para redução de 50% da dor



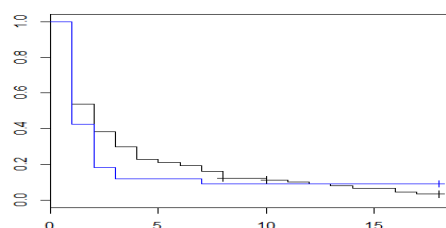
(a) Variável grupo



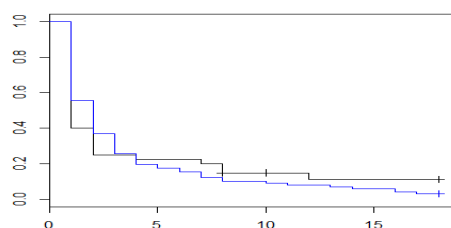
(b) Variável idade



(c) Variável imc



(d) Variável tempo de dor



(e) Variável sexo

- Grupo
 1. Preto: eletro placebo
 2. Azul: eletro ativo
- Idade
 1. Preto: até 50 anos

- 2. Azul: maior que 50 anos
- IMC
 - 1. Preto: até 29,9
 - 2. Azul: maior que 29,9
- Tempo de dor
 - 1. Preto: até 96 semanas
 - 2. Azul: maior que 96 semanas
- Sexo
 - 1. Preto: feminino
 - 2. Azul: masculino

Analisando a Figura 1 é possível notar que apenas as variáveis grupo, idade e IMC aparentam não violar a suposição de riscos proporcionais, pois as mesmas não apresenta cruzamento de sua função de sobrevivência. Note que isso não ocorre com as variáveis tempo de dor e sexo. Por esse motivo, o ajuste do modelo de riscos proporcionais de Cox apresentado na seção seguinte não considerará essas duas variáveis.

Analisando a Figura 1 é possível também observar, considerando cada variável separadamente, que as pessoas que apresentaram o menor tempo de sobrevivência (e portanto, o menor tempo até o alívio da dor) foram as pessoas alocadas no grupo eletropilates ativo, que apresentam mais de 50 anos e que o IMC é menor que 29,9.

6.2 Modelo de riscos proporcionais de Cox

O modelo de regressão de Cox foi utilizado nesse estudo com o interesse de se avaliar o poder de explicações das covariáveis. A suposição básica para o seu uso é que as taxas de falha sejam proporcionais e a avaliação dessa proporcionalidade pode ser observada analisando o gráfico de Kaplan Meier e analisando se há cruzamento entre as curvas de sobrevivência estimadas e por meio do teste de proporcionalidade. A Tabela 3 apresenta os resultados do teste de proporcionalidade considerando 4 possíveis modelos:

1. Modelo 1: considerando todas as três covariáveis (Grupo, Idade e IMC);
2. Modelo 2: somente Idade e IMC;
3. Modelo 3: somente Grupo e IMC;
4. Modelo 4: somente Grupo e idade.

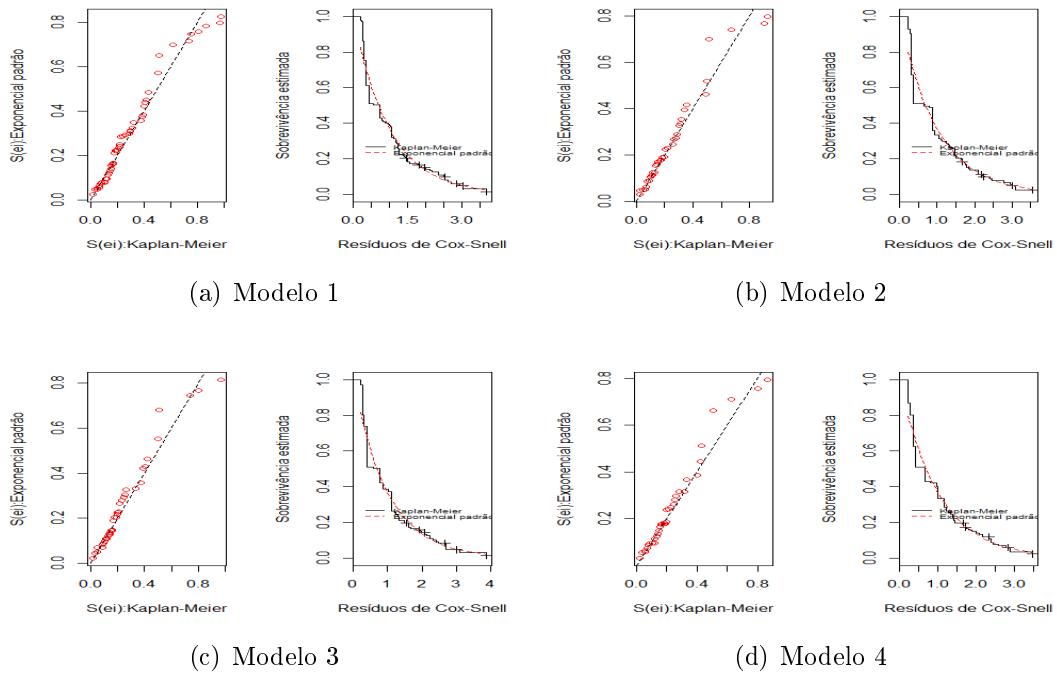
Tabela 3: Resultados para o teste de proporcionalidade dos riscos

Covariável	ρ	χ^2	P valor
Grupo	-0.0825	0.879	0.349
Idade	0.0389	0.198	0.657
IMC	0.0447	0.259	0.611
GLOBAL		1.441	0.696
Idade	0.0309	0.124	0.725
IMC	0.0488	0.305	0.581
GLOBAL		0.411	0.814
Grupo	-0.0889	1.015	0.314
IMC	0.0406	0.214	0.643
GLOBAL		1.329	0.515
Grupo	-0.0845	0.906	0.341
Idade	0.0406	0.215	0.643
GLOBAL		1.146	0.564

Veja pela Tabela 3 que em todos os modelos considerados não há evidência para rejeitar a hipótese de proporcionalidade dos riscos considerando um nível de significância de 5% , pois p-valor > 0.05 . Logo, admite-se a proporcionalidade dos riscos das três covariáveis consideradas.

A qualidade do ajuste global do modelo pode ser verificada pelos resíduos de Cox-Snell (COX e SNELL, 1968). Segundo Lawless (2003), os resíduos de Cox-Snell vêm de uma população homogênea e devem seguir uma distribuição exponencial com média 1. A Figura 2 compara graficamente a função de sobrevivência dos resíduos calculados segundo o modelo de Cox. Quanto mais próximo a função de sobrevivência dos resíduos se aproxima da função de sobrevivência da exponencial, melhor é o ajuste. Todos os modelos considerados apresentaram um ajuste adequado aos dados.

Figura 2: Resíduos de Cox-Snell



Os resultados das estimativas dos parâmetros do Modelo 1 (completo) são apresentados pela Tabela 4.

Note na Tabela 4 que a única variável significativa foi o Grupo, indicando que o risco de um paciente que está no grupo Eletro Ativo ter alívio da dor é 1,46 vezes (46% maior) o risco de um paciente do grupo Eletro Placebo. As variáveis Idade e IMC não apresentaram associação significativa com o tempo até a melhora da dor. Apesar de não significativas, a Idade e o IMC foram

Tabela 4: Estimativa dos parâmetros do modelo 1.

Covariáveis	β	EP	RR	IC 95%	P-valor
Grupo					
Placebo	0	-	1	-	-
Ativo	0.3778	0.1782	1.4591	(1.0291;2.069)	0.0339
Idade					
Até 50 anos	0	-	1	-	-
Maior que 50 anos	0.1797	0.1860	1.1969	(0.8312;1.723)	0.34
IMC					
Até 29.9	0	-	1	-	-
Maior que 29.9	-0.2513	0.2126	0.7778	(0.5127;1.180)	0.2372

mantidas no modelo por se acreditar que são variáveis que podem influenciar no tratamento.

As estimativas dos valores da Função de sobrevivência base são apresentadas na Tabela 5.

Tabela 5: Estimativa da função de sobrevivência base

t	$\hat{S}(t)$	t	$\hat{S}(t)$	t	$\hat{S}(t)$
1	0.5771	7	0.1986	13	0.1124
2	0.4040	8	0.1639	14	0.0907
3	0.3195	9	0.1541	15	0.0801
4	0.2642	10	0.1438	16	0.0801
5	0.2480	11	0.1336	17	0.0801
6	0.2317	12	0.1231	18	-

Assim, tomando como exemplo, temos que a probabilidade estimada de um paciente do grupo Ativo, com até 50 anos e IMC maior que 29.9 sobreviver (ou não apresentar alívio da dor) na décima seção ($t=10$) é igual a $(0,1438)^{\exp(0,3778+0-0,2513)} = 0,1107$.

6.3 Conclusões

O objetivo inicial nesse trabalho era analisar o tempo de recuperação de paciente de relação a uma melhora de 50% e 30% em relação a dor inicial, sendo esses pacientes divididos em um grupo denominado eletropilates ativo onde ocorre a metodologia pilates e a corrente interferencial e o grupo eletropilates placebo onde é aplicada apenas a metodologia pilates, porém os pacientes não sabiam distinguir em qual grupo estavam alocados.

O modelo de COX ajustou bem ao problema, porém a única variável que realmente foi significativa no estudo foi a variável grupo. No entanto, IMC e idade permaneceram no modelo para controle.

O resultado mostrou que o grupo eletropilates ativo melhora mais rápido que o eletropilates placebo. A chance de um paciente alocado no grupo ativo obter melhora é 46% maior ($RR=1.46$) que os pacientes alocado no grupo placebo.

A mesma análise foi realizada considerando uma redução de 30% em relação a dor inicial, porém não houve uma diferença significativa e os modelos foram bastante similares.

7 Referências

BRUNELLO, G. H. V; NAKANO, E. Y. Inferência bayesiana no modelo weibull discreto em dados com presença de censura. TEMA - Tend. Mat. Apl. Comput., v.16, n.2, p.97-110, 2015.

COLOSIMO, E. A.; GIOLO, S.R. Análise de sobrevivência aplicada. 1.ed. São Paulo: Edgard Blucher, 2006. 370p.

COX, D. R. Regression models and life tables (with discussions). J.R. Stat. Soc. B, v.34, n.2, p.187-202, 1972.

COX, D. R., & HINKLEY, D. V. Theoretical Statistics. London: Chapman and Hall, 1974.

COX, D. R.; SNELL, E. J. A general definition of residuals. J.R. Stat. Soc. B, v.30, n.2, p.248-275, 1968.

FRANCO, Y.L.S. Efetividade da adição da corrente interferencial ao método pilates na dor lombar crônica não específica: estudo controlado aleatorizado. Dissertação de Mestrado em Fisioterapia. Universidade Cidade de São Paulo, São Paulo, 2014, 47P.

KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. J. Am. Stat. Assoc., v.53, n.282, p.457-481, 1958.

LAKATOS, E. M.; MARCONI, M. A. Fundamentos de Metodologia Científica. Editora Atlas, São Paulo, 2001.

LATEY P. The Pilates method: history and philosophy. J Bodyw Mov Ther. 2001;5(4):275-82.

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

NAKANO, E. Y.; CARRASCO, C. G. Uma avaliação do uso de um modelo contínuo na análise de dados discretos de sobrevivência. TEMA - Tend. Mat. Apl. Comput., v.7, n.1, p.91-100, 2006.

NAKANO, E. Y.; CUNHA, J. F. Análise do efeito da camuflagem no tempo de segregação em regiões texturizadas utilizando o modelo de riscos

proporcionais de Cox. Semina: Ciências Exatas e Tecnológicas, v.33, n.2, p.141-148, 2012.

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

SANTOS, R. O; NAKANO, E. Y. Análise do tempo de permanência de trabalhadores no mercado de trabalho do Distrito Federal via modelo de riscos proporcionais de Cox e Log-normal. Rev. Bras. Biom., v.33, n.4, 2015.