

Análise de Sobrevivência aplicada à dados de morte por acidente de trabalho: um estudo de caso no município de Lavras-MG

Isabela S. Lima^{1†}, Carla Regina G. Brighenti²

¹Universidade Federal de Lavras, UFLA.

²Universidade Federal de São João Del-Rei, UFSJ.

Resumo: Neste estudo foi analisado o tempo até a morte por acidente de trabalho de indivíduos no município de Lavras-MG, entre os anos de 2016 à 2019. Para isso, utilizou-se análise de sobrevivência, uma técnica estatística que estuda o tempo até a ocorrência de um determinado evento de interesse, e é caracterizada pela presença de censura nos dados. Foram avaliados os efeitos de algumas covariáveis na variável resposta, como as atividades econômicas que referem às ocupações que os indivíduos exerciam, e o estado civil. Para este propósito, foram utilizados ajustes simples com estimador não-paramétrico (Kaplan-Meier), com modelo semi-paramétrico (regressão de Cox) e modelos paramétricos (Exponencial, Log-normal e de Weibull). Em cada caso apresentou-se a respectiva curva de sobrevivência. Os ajustes foram testados quanto a significância, optando-se pelo modelo de regressão de Weibull que atendeu as pressuposições, além de que neste modelo foi possível incluir o efeito da covariável significativa referente às atividades econômicas.

Palavras-chave: Estimador de Kaplan-Meier; Modelos de Regressão; Modelo de Weibull; Modelo de Cox.

Abstract: In this study, the time until death due to an accident at work of individuals in the city of Lavras-MG was analyzed, between the years 2016 to 2019. For this, survival analysis was used, a statistical technique that studies the time until the occurrence of a particular event of interest, and is characterized by the presence of censorship in the data. The effects of some covariates on the response variable were evaluated, such as economic activities that refer to the occupations that individuals exercised, and marital status. For this purpose, simple adjustments were used with a non-parametric estimator (Kaplan-Meier), with a semi-parametric model (Cox regression) and parametric models (Exponential, Log-normal and Weibull). In each case, the respective survival curve was presented. The adjustments were tested for significance, opting for the Weibull regression model, which met the assumptions, and in this model it was possible to include the effect of the significant covariate referring to economic activities.

Keywords: Kaplan-Meier estimator; Regression Models; Weibull Model; Cox Model.

[†]Autor correspondente: isabela_lima30@hotmail.com.

Introdução

Os acidentes de trabalho ocorrem pela combinação de inúmeros fatores, de maneira que destacam-se as falhas humanas e materiais, e em sua maioria pelas condições de trabalho inadequadas, máquinas inseguras, ou ainda, falta de preparo dos profissionais para enfrentar certos riscos. Tal problemática trazem tristeza para os parentes, além de prejuízos para as famílias, que perdem seus provedores e também para o mercado (MIRANDA et al., 2012).

A mortalidade por acidentes de trabalho resulta em um problema de saúde pública de grande relevância, que expressa a degradação social a que estão expostos os trabalhadores e coloca em discussão as condições em que se realiza o trabalho (LACERDA; FERNANDES; NOBRE, 2014). Além disso, os acidentes de trabalho impactam a produtividade e a economia, pois são uma das principais causas de afastamento de trabalhadores dos seus respectivos serviços (SANTANA et al., 2006).

No Brasil, segundo dados do Observatório Digital de Segurança e Saúde do Trabalho (2021), entre 2012 e 2018 foram contabilizadas 17.200 mortes em razão de algum acidente ou doença relacionados à atividade laboral. As atividades econômicas com maior número de mortes por acidentes de trabalho foram: atendimento hospitalar, comércio varejista, especialmente supermercados, administração pública, construção de edifícios, transporte de cargas e correio.

Para avaliar as condições de riscos existentes e conseqüentemente auxiliar nas medidas que podem posteriormente serem adotadas, exige-se a necessidade do conhecimento de informações estatísticas relativas aos acidentes do trabalho (CORREA; ASSUNÇÃO, 2003). Várias são as técnicas estatísticas que utilizam dados de acidentes de trabalho. Como por exemplo, Hennington et al. (2004), usaram estatística descritiva para analisar o número de mortes decorrentes por acidente de trabalho. Amorim et al. (2012), utilizaram regressão logística para realizar a análise simultânea de algumas variáveis estudadas.

Da Silva e Júnior (2020), analisaram o risco de acidente no trabalho em uma companhia elétrica por meio de modelos de sobrevivência na ausência e presença de riscos competitivos, isto é, quando a ocorrência de um evento altera a probabilidade de ocorrência de um outro evento de interesse, onde consideram os efeitos de algumas covariáveis, utilizando regressão de Cox. Ainda neste trabalho, os autores alegam que encontraram limitações para direcionar suas discussões nos resultados, visto que não foram encontrados trabalhos que tratam do tempo até a ocorrência de acidente, utilizando-se os modelos de sobrevivência.

Desse modo, utilizar técnicas de análise de sobrevivência para estimar o tempo até a ocorrência de mortes por acidente no trabalho pode ser eficaz, já que a análise de sobrevivência é uma área da estatística que estuda o tempo até a ocorrência de um determinado evento de interesse, podendo ser o tempo até a morte por determinada causa, bem como até a cura ou recidiva de uma doença, ou ainda, o tempo até a falha de um equipamento, entre outros. Entretanto, é muito comum que o evento de interesse não ocorra em alguns elementos amostrais, assim constituindo em dados censurados, pois há a ausência da ocorrência do evento no tempo de análise. Desse modo, utilizar a análise de sobrevivência é a forma mais adequada para tratar dessas censuras, que são observações incompletas da resposta, já que a presença de censura caracteriza os modelos de análise de sobrevivência. Para isso, pode-se utilizar o estimador não paramétrico de Kaplan-Meier, modelos paramétricos que fazem uso de distribuições de probabilidade e modelos semiparamétricos, como o modelo de regressão de Cox (COLOSIMO; GIOLO, 2006).

Neste sentido, o presente trabalho objetivou em utilizar as técnicas de análise de sobrevivência para modelar o tempo até a morte ocasionada por acidente de trabalho, e assim, escolher o melhor modelo que descreve esses dados e incorpora os efeitos das covariáveis que forem significativas para o ajuste.

Metodologia

Para avaliar as mortes por acidente de trabalho, utilizou-se um conjunto de dados que foi extraído do Sistema de Informações de Mortalidade - SIM, desenvolvido pelo Departamento de Informática do Sistema Único de Saúde - DATASUS (MINISTÉRIO DA SAÚDE, 2016, 2017, 2018, 2019). Este sistema consolida todas as declarações de óbito (DO) emitidas pelos profissionais de serviços de saúde e por cartórios de registro civil. Além de seu caráter jurídico-civil, a DO é um instrumento oficial importante para a quantificação dos agravos fatais na população, que, recentemente, vem-se tornando uma fonte ágil para enumerar e quantificar os acidentes relacionados ao trabalho, identificar riscos e subsidiar os serviços que trabalham com vigilância em saúde do trabalhador (CORREA; ASSUNÇÃO, 2003).

Em virtude disso, foram selecionadas as declarações de óbito por acidente de trabalho registradas no referido banco ocorridas no município de Lavras, entre os anos de 2016 a 2019, que totalizou em 28 observações. Para cada indivíduo, foram registrados: a data de nascimento, a data de óbito, a ocupação de trabalho e o estado civil. Assim, para este estudo foi considerado como o evento de interesse a morte por acidente de trabalho, em anos, que foi contabilizado através da diferença entre a data de óbito e a data de nascimento, e foi considerado como censura onde o tempo de trabalho foi interrompido por alguma razão desconhecida. Pois, apesar da planilha ser de acidente de trabalho o tempo registrado foi interrompido por motivo desconhecido, caracterizando uma censura aleatória.

As covariáveis ocupação e estado civil são categóricas, sendo assim, foram codificadas para serem incluídas nas análises. Para ocupação, que se refere às atividades econômicas em questão, foram atribuídos os seguintes códigos: 1 - Outros, 2 - Motorista, 3 - Construção civil e 4 - Comercial. Para estado civil: 1 - Outros, 2 - Solteiro e 3 - Casado.

A função extensivamente usada na análise de dados de sobrevivência, é a função de sobrevivência, que segundo Colosimo e Giolo (2006), 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, isto é escrito como:

$$S(t) = P(T \geq t) \quad (1)$$

em que T é uma variável aleatória não-negativa, usualmente contínua, que representa o tempo de falha. A função de distribuição acumulada é definida como a probabilidade de uma observação não sobreviver ao tempo t , isto é, $F(t) = 1 - S(t)$.

Para estimar a função de sobrevivência, primeiramente, utilizou-se o método não paramétrico, o estimador de Kaplan-Meier, que conforme Colosimo e Giolo (2006), é definido por:

$$\hat{S}(t) = \prod_{j:t_j < t} \left(\frac{n_j - d_j}{n_j} \right) = \prod_{j:t_j < t} \left(1 - \frac{d_j}{n_j} \right) \quad (2)$$

onde, $t_j < t$ são os tempos ordenados de falha e são distintos entre si, onde $j = 1, 2, \dots, k$, d_j representa o número de falhas nos tempos de falha e n_j o número de indivíduos sob risco nos tempos de falha.

Assim sendo, foi construído o gráfico para as estimativas da função de sobrevivência $S(t)$ e seus respectivos intervalos a 95% de confiança utilizando-se o estimador de Kaplan-Meier, sem considerar determinado estrato da amostra inicialmente. Além disso, foi estimado o tempo médio de vida dos indivíduos, que de acordo com Colosimo e Giolo (2006) é obtido calculando-se a área sob a curva de Kaplan-Meier estimada. Como esta curva é uma função escada, esta integral é simplesmente a soma de áreas de retângulos, isto é:

$$\hat{t}_m = t_1 + \sum_{j=1}^{k-1} \hat{S}(t_j)(t_{j+1} - t_j) \quad (3)$$

Posteriormente, para comparar as curvas de sobrevivência considerando as covariáveis ocupação e estado civil, ainda utilizando o estimador de Kaplan-Meier, foram construídas as curvas para as estimativas da função de sobrevivência separadamente, para investigar o efeito dessas covariáveis no tempo até a ocorrência de morte por acidente de trabalho. Para isso, utilizou-se o teste logrank para testar a hipótese de igualdade entre as curvas, este é um teste muito utilizado em análise de sobrevivência, e sua estatística é dada pela diferença entre o número observado de falhas em cada ocupação ou estado civil e uma quantidade que, para muitos propósitos, pode ser pensada como o correspondente número esperado de falhas sob a hipótese nula (CARVALHO et al., 2005).

Para a abordagem paramétrica, foram utilizados modelos probabilísticos com as distribuições de probabilidade mais usuais em análise de sobrevivência, que são, exponencial, de Weibull e log-normal. Tais distribuições têm se mostrado bastante adequadas para descrever dados de sobrevivência (ISIDRO et al., 2020). Assim, foram estimados os parâmetros pelo método da máxima verossimilhança para os três modelos sem considerar covariáveis, para escolher dentre eles, o que melhor se ajusta aos dados. A função de densidade de probabilidade, a função de sobrevivência e de taxa de falha para esses três modelos estão apresentadas na Tabela 1.

Tabela 1: Funções importantes para os modelos

Modelo	$f(t)$	$S(t)$	$\lambda(t)$
Exponencial	$\frac{1}{\alpha} \exp \left\{ - \left(\frac{t}{\alpha} \right) \right\}$	$\exp \left\{ - \left(\frac{t}{\alpha} \right) \right\}$	$\frac{1}{\alpha}$
de Weibull	$\frac{\gamma}{\alpha^\gamma} t^{\gamma-1} \exp \left\{ - \left(\frac{t}{\alpha} \right)^\gamma \right\}$	$\exp \left\{ - \left(\frac{t}{\alpha} \right)^\gamma \right\}$	$\frac{\gamma}{\alpha^\gamma} t^{\gamma-1}$
Log-normal	$\frac{1}{\sqrt{2\pi}t\sigma} \exp \left\{ - \frac{1}{2} \left(\frac{\log(t) - \mu}{\sigma} \right)^2 \right\}$	$\Phi \left(\frac{-\log(t) + \mu}{\sigma} \right)$	$\frac{f(t)}{S(t)}$

De acordo com Colosimo e Giolo (2006), a escolha de um modelo probabilístico adequado para descrever o tempo de falha deve ser feita com bastante cuidado, pois a utilização de um modelo inadequado acarreta erros grosseiros nas estimativas dessas quantidades. Desse modo, para a escolha do modelo foram utilizados dois métodos gráficos, um que consiste em construir os gráficos das estimativas das sobrevivências obtidas pelo método Kaplan-Meier versus as estimativas das sobrevivências obtidas a partir dos modelos exponencial, de Weibull e log-normal e, o outro que consiste na linearização da função de sobrevivência.

Além disso, foi realizado o teste da razão de verossimilhanças para confirmar os resultados, com as hipóteses de que: i) H_0 : o modelo exponencial é adequado, ii) H_0 : o modelo de Weibull é adequado e iii) H_0 : o modelo log-normal é adequado. Utilizou-se a estatística da razão de verossimilhanças em modelos encaixados, assim, partiu-se da distribuição gama generalizada, pois as distribuições exponencial, de Weibull e log-normal são casos particulares dela, o que a torna útil na discriminação dos modelos. Após escolhido o melhor ajuste, foram construídas as curvas de sobrevivência e estimado o tempo médio e mediano.

Ainda utilizando modelos paramétricos, mas agora considerando as covariáveis que foram significativas pelo teste logrank, na etapa do estimador de Kaplan-Meier, foram utilizados modelos de regressão paramétricos para acomodar o efeito dessas covariáveis. Pois, as técnicas vistas até o momento não permitem a inclusão conjunta das covariáveis na análise, desse modo, a forma mais eficiente de acomodar o efeito de covariáveis é utilizar um modelo de regressão apropriado para dados censurados (COLOSIMO; GIOLO, 2006).

Assim, requer a especificação de uma distribuição de probabilidade adequada aos dados. Foi realizado o teste da razão de verossimilhanças como anteriormente, entretanto, considerando as covariáveis no modelo. Após a escolha do modelo e antes de proceder a interpretação das esti-

mativas dos parâmetros do modelo ajustado, utilizou-se os resíduos de Cox-Snell para confirmar a adequação do modelo. E então, construiu-se as curvas de sobrevivência estimadas.

Por fim, utilizou-se uma abordagem semiparamétrica, considerando-se o modelo de regressão de Cox, que é amplamente utilizado por conter a presença de um componente não-paramétrico, que torna o modelo bastante flexível. A suposição básica para o uso deste modelo, é que as taxas de falha sejam proporcionais (KLEINBAUM; KLEIN, 1995). A expressão geral do modelo de regressão de Cox, considerando p covariáveis, é dada por:

$$\lambda(t|\mathbf{x}) = \lambda_0(t)g(\mathbf{x}'\boldsymbol{\beta}) \quad (4)$$

em que \mathbf{x} é um vetor com os componentes $\mathbf{x} = (x_1, \dots, x_p)'$; $\boldsymbol{\beta}$ é o vetor de parâmetros associados às covariáveis; $g(\mathbf{x}'\boldsymbol{\beta})$ é uma função não-negativa que deve ser especificada e λ_0 o componente não-paramétrico, usualmente denominado função de taxa de falha de base.

Logo, foram estimados os parâmetros para quatro modelos possíveis utilizando o modelo de regressão de Cox, em seguida, foi realizado o teste da razão de verossimilhanças para escolher qual modelo se ajusta melhor aos dados. Para avaliar esse ajuste, ou seja, para avaliar a suposição de taxas de falhas proporcionais, realizou-se os resíduos padronizados de Schoenfeld (COLOSIMO; GIOLO, 2006).

Para obter as estimativas da sobrevivência, e plotar as curvas, primeiramente foi necessário estimar os valores para a função taxa de falha acumulada de base dada por:

$$\Lambda_0(t) = \int_0^t \lambda_0(u)du \quad (5)$$

e à correspondente função de sobrevivência de base:

$$S_0(t) = \exp\{-\Lambda_0(t)\} \quad (6)$$

Já que a função de sobrevivência é dada por:

$$S(t|\mathbf{x}) = [S_0(t)]^{\exp\{\mathbf{x}'\boldsymbol{\beta}\}} \quad (7)$$

Assim, construiu-se as curvas de sobrevivência estimadas e as correspondentes curvas das taxas de falha acumulada estimadas para o melhor ajuste aos dados. Utilizou-se o nível de significância de 5% para avaliação da aceitação dos modelos considerados.

Todas as análises estatísticas foram realizadas no programa R (R CORE TEAM, 2020), um software estatístico gratuito, por meio do pacote survival (THERNEAU, 2015).

Resultados e Discussão

Os resultados são apresentados aqui de acordo com a metodologia citada.

Utilizando o estimador de Kaplan-Meier tem-se as estimativas da função de sobrevivência $S(t)$ e seus respectivos intervalos a 95% de confiança, como pode ser visto na Figura 1.

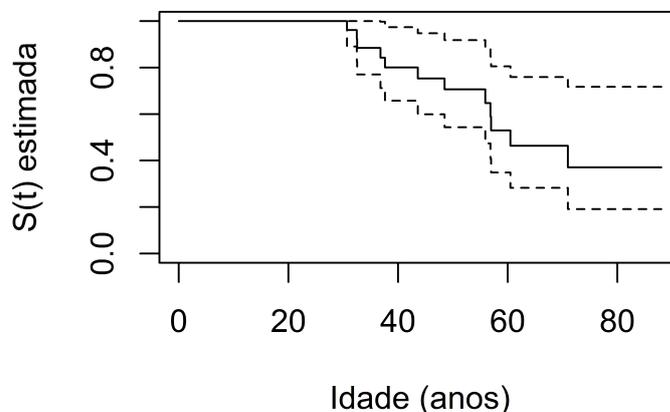


Figura 1: Sobrevivência estimada por Kaplan-Meier para os dados

Pode-se observar que a função de sobrevivência estimada é uma função escada com degraus nos tempos observados que houve morte (falha), de tamanho $\frac{1}{28}$. Nos lugares em que o topo do degrau (parte horizontal) é maior, indica que se passaram mais tempo até acontecer uma morte. O tempo médio estimado pela Equação (3), foi igual à 60,01 anos. Alguns trabalhos obtiveram resultados próximos, como em Santana et al. (2005) observou uma elevação da mortalidade com a idade chegando ao patamar de 23,0 mortes/100.000 acidentes na faixa de 65 anos, o que está próximo do tempo médio estimado. Além disso, Santana et al. (2009) ao avaliar atendimento de acidentes, observaram que a maior parte dos casos graves ocorreu entre os homens que tinham pelo menos 37 anos de idade.

O gráfico considerando estado civil a partir do estimador de Kaplan-Meier está na Figura 2.

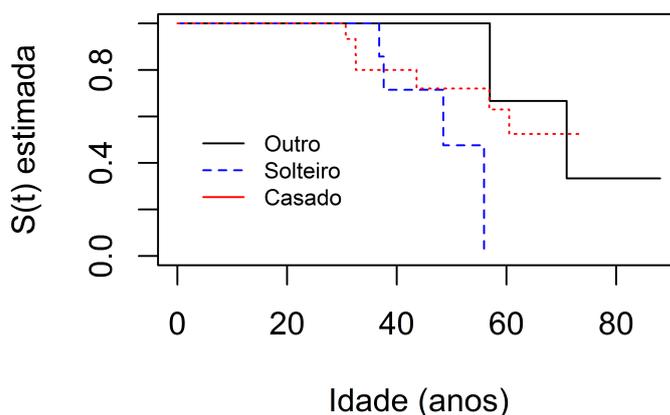


Figura 2: Sobrevivência estimada por Kaplan-Meier para os dados considerando o estado civil

O teste logrank resultou em valor p maior que 0,05, portanto não é possível rejeitar a hipótese nula, ao nível de 5% de probabilidade. Indicando, assim, que não há diferença entre as curvas de sobrevivência de estado civil, ou seja, o efeito do estado civil dos indivíduos não influencia na

morte por acidente de trabalho. Ruiz et al. (2004), realizaram um trabalho em que analisaram descritivamente também o estado civil e encontraram um índice de 51% dos acidentes com vítimas casadas, enquanto 42,7% eram com vítimas solteiras, ainda neste trabalho, os autores alegam que a respeito da variável estado civil não se encontrou outras pesquisas semelhantes para comparação dos resultados, apesar de ser um dado referenciado como relevante na quantificação da exposição ao risco de acidentarse. Assim, na literatura não encontra-se muitas referências para discussão dos resultados.

O gráfico com a covariável ocupação utilizando o estimador de Kaplan-Meier é apresentado na Figura 3.

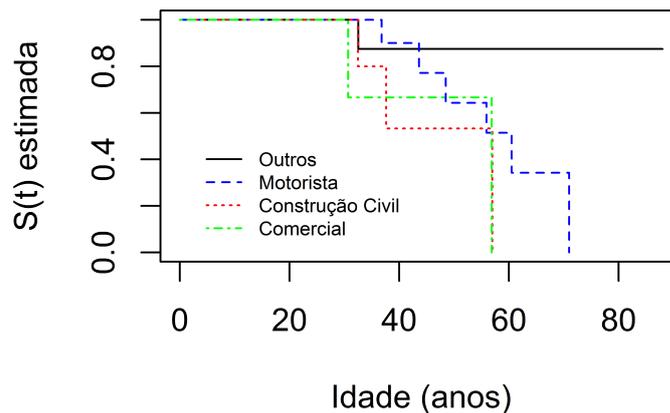


Figura 3: Sobrevivência estimada por Kaplan-Meier para os dados de mortes por acidentes no trabalho considerando a ocupação da vítima

O teste logrank, para comparação entre as ocupações, resultou em um valor p menor que 0,05, logo, rejeita-se a hipótese nula, ao nível de 5% de probabilidade. Indicando, assim, que há diferença entre as curvas de sobrevivência de ocupação, ou seja, a ocupação dos indivíduos influencia na morte por acidente de trabalho. Pela Figura 3, pode-se observar que para motorista o tempo até a morte é maior que nas demais ocupações, o comportamento de construção civil e comercial apresentam semelhanças em alguns tempos. Ao comparar este resultado na literatura, encontrou-se que o ramo de atividade econômica de maior risco de acidentes fatais foi a construção civil, em Campinas (Salerno, 1998), mas isso diferiu dos dados de Waldvogel (2002), que estimou como o mais perigoso o ramo dos transportes e comunicações, ficando a indústria manufatureira e a construção civil em segundo lugar, seguido pelo serviço e comércio.

Para o modelo paramétrico, sem considerar covariáveis, foram estimados os parâmetros pelo método da máxima verossimilhança para os modelos: exponencial, de Weibull e log-normal. Dessa forma, substituindo essas estimativas nas expressões das funções de sobrevivência, tem-se para os modelos exponencial, de Weibull e log-normal, respectivamente:

$$\hat{S}(t) = \exp\left\{\frac{-t}{118,67}\right\}, \quad \hat{S}(t) = \exp\left\{-\left(\frac{t}{73,66}\right)^{3,04}\right\}, \quad \hat{S}(t) = \Phi\left[\frac{(-\log(t) + 4,15)}{0,43}\right]$$

Para escolher um dos modelos, utilizou-se inicialmente o primeiro método gráfico, que indica que o modelo adequado será aquele em que a sua curva de sobrevivência mais se aproxima da curva do estimador de Kaplan-Meier. Como conclusão, obteve-se que o modelo de Weibull e o log-normal parecem os mais adequados para estes dados, já que acompanham um pouco mais de perto uma reta. Construiu-se também os gráficos linearizados (método 2), e concluiu-se que o modelo de Weibull e log-normal continuam sendo mais adequados que o exponencial, visto que se assemelham mais a uma reta.

Para confirmar qual modelo é mais adequado, foi realizado o teste da razão de verossimilhança, cujos valores do logaritmo da função de verossimilhança resultaram nos valores apresentados na Tabela 2:

Tabela 2: Logaritmo da função de verossimilhança e resultados dos TRV

Modelo	Logaritmo	TRV	Valor p
Gama Generalizado	-58,77	-	-
Exponencial	-69,32	$2(69,32 - 58,77) = 21,1$	0,0000262
Weibull	-62,18	$2(62,18 - 58,77) = 36,8$	0,0089613
Log-normal	-61,05	$2(61,05 - 58,77) = 4,6$	0,0325250

O valor p dos modelos exponencial e de Weibull são menores que 0,01, assim rejeita-se a hipótese de que o modelo exponencial é adequado, assim como a hipótese que o modelo de Weibull é adequado. O modelo log-normal apresentou adequação para análise destes dados, já que o seu valor p foi maior que 0,01, não rejeitando a hipótese de que o modelo log-normal é adequado, ao nível de 1% de probabilidade, confirmando assim, as conclusões obtidas nas técnicas gráficas.

Na Figura 4 tem-se a curva de sobrevivência estimada por meio do ajuste do modelo log-normal versus a curva de sobrevivência estimada por Kaplan-Meier:

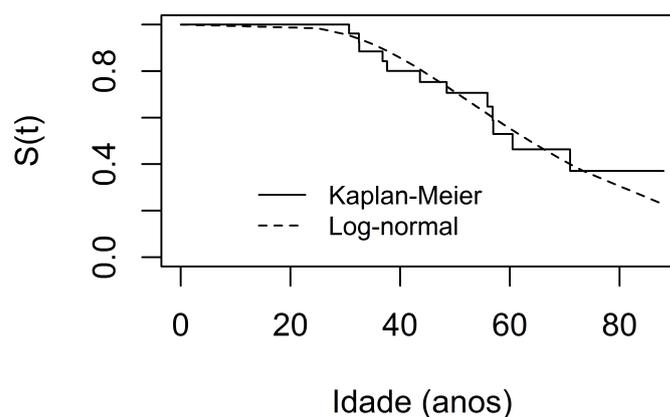


Figura 4: Curva de sobrevivência estimada pelo modelo log-normal versus a curva de sobrevivência estimada por Kaplan-Meier

Pode-observar que o modelo ajustado Log-normal acompanha adequadamente o que foi estimado por Kaplan-Meier. O tempo médio com base na distribuição log-normal é dado por:

$$\hat{E}(t) = \exp \left[4,15 + \frac{0,43^2}{2} \right] = 69,87$$

E o tempo mediano, é calculado através da expressão dos percentis:

$$\hat{t}_{0,5} = \exp [0 \cdot 0,43 + 4,15] = 63,58$$

Isto indica que 50% dos indivíduos permanecem vivos em 63,58 anos. Vale salientar que não foi encontrado na literatura para comparação trabalhos que utilizassem uma abordagem paramétrica para modelar o tempo até a morte por acidente de trabalho.

Inicialmente, utilizou-se métodos não paramétricos para construir as estimativas de Kaplan-Meier para cada covariável separadamente e compará-las, e pelo teste logrank obteve-se que apenas a covariável ocupação é significativa, diferindo assim suas curvas de sobrevivência. Desse modo, apenas a covariável ocupação deve ser incluída na modelagem estatística, e para isso, a forma mais eficiente de acomodar o efeito dessa covariável é usar um modelo de regressão apropriado.

Os resultados do teste da razão de verossimilhança para determinar qual distribuição de probabilidade é mais adequada na presença da covariável ocupação, estão na Tabela 3:

Tabela 3: Resultados do teste da razão de verossimilhanças

Modelo	-log	TRV	Valor p
Gama Generalizado	53,965	-	-
Exponencial	66,601	$2(66,601 - 53,965) = 25,272$	0,000003
Weibull	55,885	$2(55,885 - 53,965) = 3,840$	0,050043
Log-normal	57,318	$2(57,318 - 53,965) = 6,706$	0,009608

Apesar do valor p encontrado ser muito próximo da região de decisão, este pode indicar o modelo de regressão de Weibull como o apropriado. Acredita-se que pelo perfil do modelo de Weibull, este poderia ser mais expressivo se adotado um tamanho de amostra maior. No entanto, para o objetivo deste trabalho, apenas os dados para este período estavam disponíveis, e optou-se, pelo menos de forma didática em prosseguir a análise considerando o modelo de regressão de Weibull. De acordo com Coster (2009), sabe-se que estimadores possuem boas propriedades se o tamanho de amostra for grande. Entretanto, nem sempre o tamanho amostral utilizado nos estudos é o ideal.

Desse modo, todas as análises seguintes são baseadas no modelo de regressão de Weibull. Vale salientar que ocupação é uma covariável categórica, logo ela é tratada como fator no modelo, assim, para cada categoria de ocupação tem-se uma estimativa diferente para o parâmetro, que pode ser visto na Tabela 4:

Tabela 4: Estimativas dos parâmetros

Covariável	Estimativa	Erro padrão	Valor p
Intercepto	4,724	0,257	0,0000
Motorista	-0,566	0,272	0,0372
Construção Civil	-0,767	0,289	0,0080
Comercial	-0,792	0,307	0,0098

O parâmetro de forma será único, cujo valor estimado é 4,189.

Antes de proceder a interpretação das estimativas dos parâmetros do modelo ajustado, examinou-se os resíduos de Cox-Snell, para confirmar a adequabilidade do modelo, em que concluiu-se que o modelo de Weibull é satisfatório, pois as estimativas das sobrevivências dos resíduos obtidas por Kaplan-Meier e pelo modelo de Weibull, estão bem próximas, assim como o gráfico dos pares de pontos é aproximadamente uma reta, como pode ser visto na Figura 5:

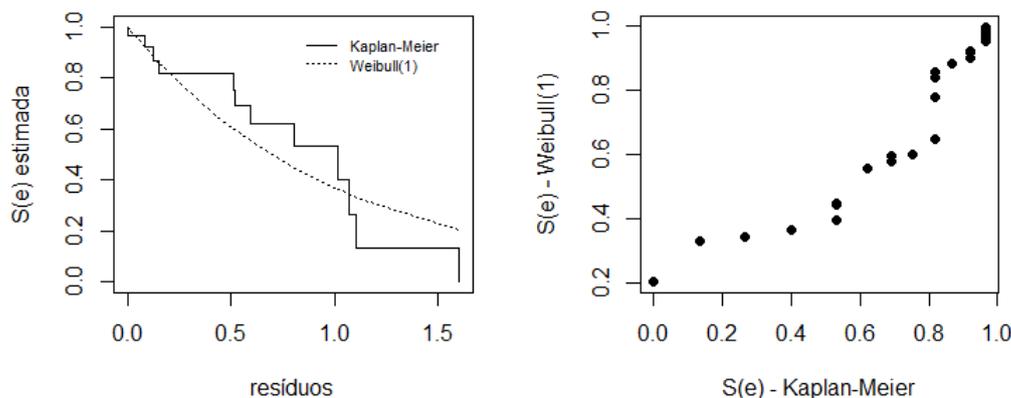


Figura 5: Análise dos resíduos de Cox-Snell do modelo de Weibull para os dados

Estimou-se as curvas de sobrevivências (Figura 6), cujas expressões das funções são, para $t \geq 0$:

$$\hat{S}(t|x) = \exp \left\{ - \left(\frac{t}{\exp \{4,724\}} \right)^{4,189} \right\}, \text{ para outros}$$

$$\hat{S}(t|x) = \exp \left\{ - \left(\frac{t}{\exp \{4,724 - 0,566\}} \right)^{4,189} \right\}, \text{ para motorista}$$

$$\hat{S}(t|x) = \exp \left\{ - \left(\frac{t}{\exp \{4,724 - 0,767\}} \right)^{4,189} \right\}, \text{ para construção civil}$$

$$\hat{S}(t|x) = \exp \left\{ - \left(\frac{t}{\exp \{4,724 - 0,792\}} \right)^{4,189} \right\}, \text{ para comercial}$$

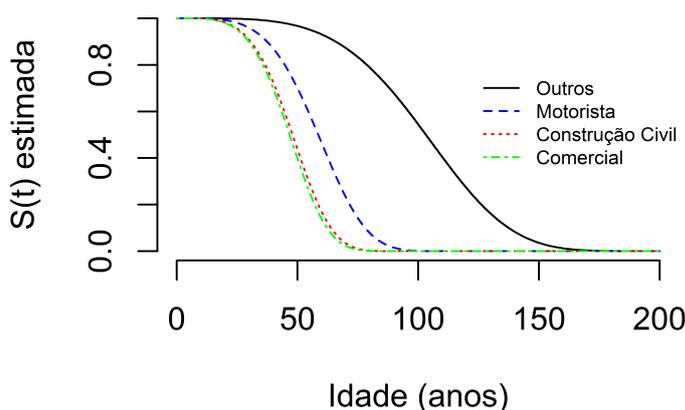


Figura 6: Curvas de sobrevivência estimadas

Tomando-se o exponencial dos coeficientes estimados apresentados na Tabela 4, obtém-se a razão dos tempos medianos de sobrevivência. Assim, para a covariável ocupação codificada em (1, 2, 3 e 4), esta razão compara o tempo mediano de sobrevivência das ocupações 2, 3 e 4 em relação à ocupação 1. Desse modo, as interpretações dos resultados obtidos são as seguintes:

- o tempo mediano até a morte de motoristas é 0,57, aproximadamente a metade daqueles indivíduos que possuem outra ocupação.
- o tempo mediano até a morte de indivíduos que possuem ocupações relacionadas à construção civil, como pedreiro, servente, entre outros é aproximadamente 0,46 daqueles indivíduos que possuem outra ocupação.
- o tempo mediano até a morte de indivíduos que possuem ocupações relacionadas ao comércio é aproximadamente 0,45 daqueles indivíduos que possuem outra ocupação.

Pode-se observar que com relação à construção civil e comercial, as conclusões são quase as mesmas. Santana et al. (2009) chegaram a conclusão que foi elevada a gravidade de acidentes de trabalho, especialmente os ocorridos com trabalhadores do ramo de transporte, que teve o maior percentual de casos críticos 7%, e comércio 6,9%, seguido pelo setor de construção civil 4,7%.

Para a modelagem semiparamétrica dos dados, assumiu-se que o modelo de Cox é adequado para esses dados e foram ajustados quatro modelos possíveis, cujos resultados estão na Tabela 5:

Tabela 5: Estimativas obtidas para os modelos de Cox ajustados aos dados

Modelos	Covariáveis	Estimativas	Log verossimilhança parcial
1	Nenhuma	-	- 32,17535
2	X_1 : ocupação	2 $\hat{\beta}_1 = 2,133$	-27,18271
		3 $\hat{\beta}_2 = 2,859$	
		4 $\hat{\beta}_3 = 3,187$	
3	X_2 : estado civil	2 $\hat{\beta}_1 = 1,3476$	- 30,83431
		3 $\hat{\beta}_2 = 0,1843$	
4	X_1 : ocupação	2 $\hat{\beta}_1 = 2,313$	- 25,27126
		3 $\hat{\beta}_2 = 3,3498$	
		4 $\hat{\beta}_3 = 3,7280$	
	X_2 : estado civil	2 $\hat{\beta}_4 = 2,0417$	
		3 $\hat{\beta}_5 = 0,6687$	

Os resultados do teste da razão de verossimilhança para escolher qual modelo se ajusta melhor aos dados podem ser vistos na Tabela 6:

Tabela 6: Resultados do teste da razão de verossimilhanças

Efeito	Estatística do teste	Valor p
$X_1 nenhuma$	$2(32,17 - 27,18) = 9,98$	0,019
$X_2 nenhuma$	$2(32,17 - 30,83) = 2,68$	0,261
$X_2 X_1$	$2(27,18 - 25,27) = 3,82$	0,148

A partir do teste da razão de verossimilhanças conclui-se que o modelo que inclui somente o efeito da covariável ocupação foi o melhor, pois o valor p foi menor que 0,05, rejeitando a hipótese nula de que não existe efeito da covariável ocupação na variável resposta, ao nível de 5% de probabilidade. A Tabela 7 mostra as estimativas obtidas para este modelo:

Tabela 7: Resultados do ajuste do modelo de Cox e as correspondentes razões de taxas de falha

Covariável	Estimativa	Erro padrão	Valor p	RTF
Motorista	2,133	1,107	0,0540	8,442
Construção Civil	2,859	1,238	0,0209	17,442
Comercial	3,187	1,313	0,0152	24,205

em que RTF indica razão de taxa de falha.

Para avaliar esse ajuste realizou-se os resíduos padronizados de Schoenfeld, os resultados se encontram na Figura 7 e Tabela 8:

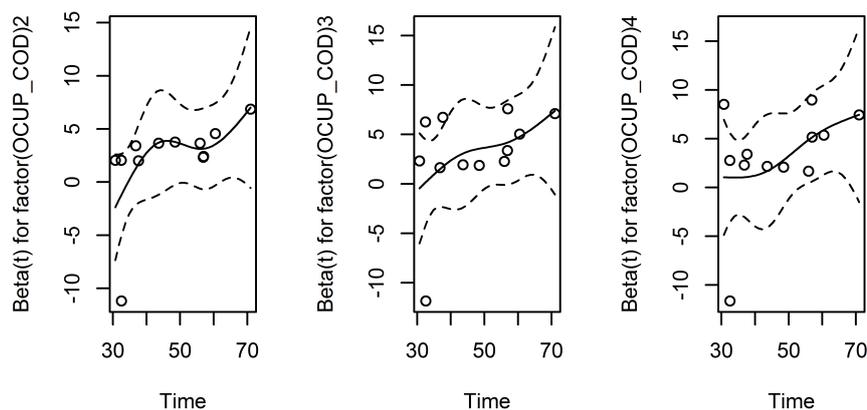


Figura 7: Resíduos padronizados de Schoenfeld do modelo de Cox com a covariável ocupação

Pela Figura 7 é possível observar que as curvas indicam tendências crescentes que, embora não sejam muito acentuadas, não sugerem riscos proporcionais. Mas a partir da Tabela 8, pode-se observar que tanto no teste global quanto os testes para cada covariável apresentaram evidências que não permitem a rejeição da hipótese nula de taxas de falha proporcionais, já que todos os valores de p são superiores a 0,05, indicando assim riscos proporcionais.

Tabela 8: Testes da proporcionalidade das taxas de falha no modelo de Cox com a covariável ocupação

Covariável	χ^2	Valor p
Motorista	1,0963	0,30
Construção Civil	0,0163	0,90
Comercial	0,0405	0,84
Global	3,8983	0,27

Logo, o modelo de Cox com a presença da covariável ocupação, é uma opção satisfatória para a análise de dados de mortes por acidentes no trabalho. Assim, as funções de sobrevivência e de taxas de falha estimadas para o modelo ajustado são expressas, respectivamente, por:

$$\hat{S}(t|\mathbf{x}) = \begin{cases} [\hat{S}_0(t)]^{exp(0)} & \text{para outros} \\ [\hat{S}_0(t)]^{exp(2,133)} & \text{para motorista} \\ [\hat{S}_0(t)]^{exp(2,859)} & \text{para construção civil} \\ [\hat{S}_0(t)]^{exp(3,187)} & \text{para comercial} \end{cases}$$

e

$$\hat{\lambda}(t|\mathbf{x}) = \begin{cases} \hat{\lambda}_0(t)exp(0) & \text{para outros} \\ \hat{\lambda}_0(t)exp(2,133) & \text{para motorista} \\ \hat{\lambda}_0(t)exp(2,859) & \text{para construção civil} \\ \hat{\lambda}_0(t)exp(3,187) & \text{para comercial} \end{cases}$$

As estimativas de $\hat{S}_0(t)$ são necessárias para obtenção de $\hat{S}(t|\mathbf{x})$. Elas foram obtidas de acordo com a Equação (6), e $\hat{\lambda}_0(t)$ de acordo com a Equação (5). Assim, tem-se as curvas de sobrevivência estimadas apresentadas na Figura 8 e as correspondentes curvas das taxas de falha acumulada estimadas na Figura 9, para cada ocupação:

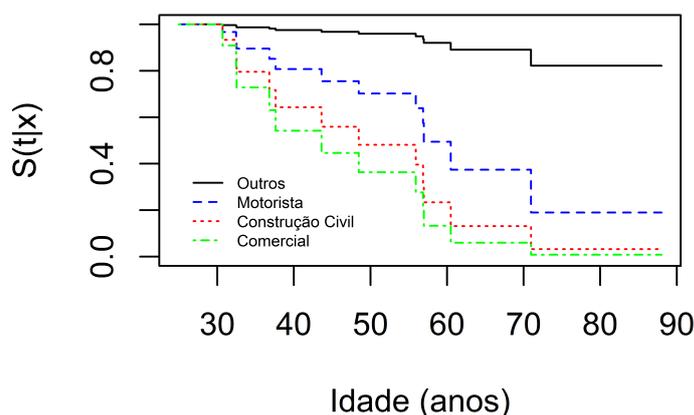


Figura 8: Curvas de sobrevivência estimadas pelo modelo de Cox para os dados

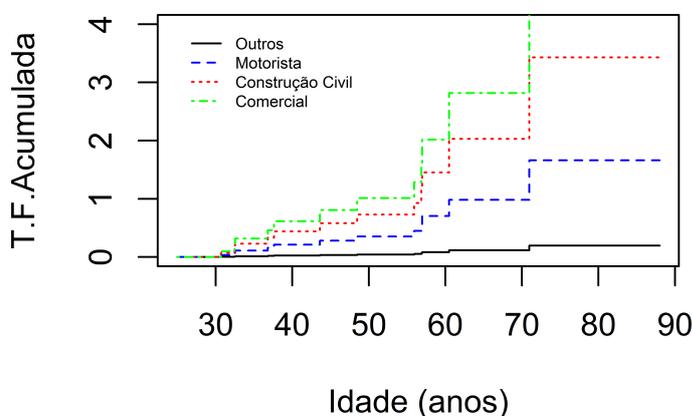


Figura 9: Taxas de falha acumulada estimadas para os dados

As interpretações para as taxas de falhas foram feitas tomando-se o exponencial dos coeficientes estimados apresentados na Tabela 7 e dividindo pela exponencial de 0, que refere-se a ocupação 1, que foi tida como referência para a comparação. Desse modo, tem-se:

- a taxa de morte de motoristas é aproximadamente 8,44 vezes a taxa de morte daqueles indivíduos que possuem outra ocupação.
- a taxa de morte de indivíduos que possuem ocupações relacionadas à construção civil, como pedreiro, servente, entre outros é aproximadamente 17,44 vezes a taxa de morte daqueles indivíduos que possuem outra ocupação.
- a taxa de morte de indivíduos que possuem ocupações relacionadas ao comércio é aproximadamente 24,20 vezes a taxa de morte daqueles indivíduos que possuem outra ocupação.

Da Silva e Júnior (2020) utilizaram o modelo de Cox para estimar o tempo até o acidente de trabalho em uma companhia elétrica, neste modelo também foram consideradas algumas covariáveis qualitativas, como sexo, tipo de acidente e afastamento do trabalho. Estes concluíram que o ajuste de Cox se mostrou adequado para os dados em questão, mas não encontraram na literatura outros trabalhos que utilizassem essa abordagem para discutir melhor seus resultados.

Conclusão

Foi possível aplicar as técnicas de sobrevivência ao conjunto de dados de morte por acidente de trabalho, como proposto. Entretanto, a baixa significância dos modelos mostra o problema do tamanho pequeno de amostra, mas isto nos proporciona caminhos para mostrar que trabalhos futuros precisam considerar tamanhos amostrais maiores para evitar problemas deste tipo.

Agradecimentos

Os autores agradecem ao CNPq pelo apoio financeiro concedido.

Referências

- AMORIM, C. R.; ARAÚJO, E. M. D., ARAÚJO, T. M. D.; OLIVEIRA, N. F. D. Acidentes de trabalho com mototaxistas. *Revista Brasileira de Epidemiologia*, v.15. n.1, p.25-37. 2012.
- CARVALHO, M. S.; ANDREOZI, V. L.; CODEÇO, C. T.; BARBOSA, M. T.; SHIMAKURA, S. E. *Análise de sobrevivência: teoria e aplicações em saúde*. Rio de Janeiro: Editora Fiocruz, 2005.
- COLOSIMO, E. A.; GIOLO, S. R. *Análise de Sobrevivência Aplicada*. São Paulo: Edgard Blucher. 2006. 369p.
- CORREA, P. R. L.; ASSUNÇÃO, A. A. A subnotificação de mortes por acidentes de trabalho: estudo de três bancos de dados. *Epidemiologia e Serviços de Saúde*, v.12, n.4, p.203-212, 2003.
- COSTER, R. Um alerta sobre o uso de amostras pequenas na regressão logística. Trabalho de conclusão de curso. Universidade Federal do Rio Grande do Sul, Rio Grande do Sul, 2009.
- DA SILVA, B. G.; JÚNIOR, L. C. S. Análise do risco de acidentes no trabalho de uma companhia elétrica paraibana. *Refas-Revista Fatec Zona Sul*, v.7, n.2, p.1-18, 2020.

HENNINGTON, E. A.; CORDEIRO, R.; MOREIRA FILHO, D. C. Trabalho, violência e morte em Campinas. *Cadernos de Saúde Pública*, São Paulo, Brasil. v. 20, p. 610-617, 2004.

ISIDRO, M. J.; ISIDRO, R. A.; BRITO, A. L.; OLIVEIRA, T. A. Modelos de fragilidade aplicados a análise de fatores contribuintes na morte de pacientes portadores de leucemia. *Brazilian Journal of Development*, v. 6, n.8, p. 54802-54820, 2020.

KLEINBAUM, D. G.; KLEIN, M. *Survival Analysis: A Self-Learning Text*. New York: Springer. 1995.

LACERDA, K. M.; FERNANDES, R. C. P.; NOBRE, L. C. C. Acidentes de trabalho fatais em Salvador, BA: descrevendo o evento subnotificado e sua relação com a violência urbana. *Revista Brasileira de Saúde Ocupacional*, v.39, n.129, p.63-74, 2014.

MINISTÉRIO DA SAÚDE, DATASUS. *Sistema de Informações de Mortalidade - SIM*, 2016.

MINISTÉRIO DA SAÚDE, DATASUS. *Sistema de Informações de Mortalidade - SIM*, 2017.

MINISTÉRIO DA SAÚDE, DATASUS. *Sistema de Informações de Mortalidade - SIM*, 2018.

MINISTÉRIO DA SAÚDE, DATASUS. *Sistema de Informações de Mortalidade - SIM*, 2019.

MIRANDA, F. M. D.; SCUSSIATO, L. A.; KIRCHHOF, A. L. C.; CRUZ, E. D. A.; SARQUIS, L. M. M. Caracterização das vítimas e dos acidentes de trabalho fatais. *Revista Gaúcha de Enfermagem*, v.33, n.2, p.45-51, 2012.

OBSERVATÓRIO DIGITAL DA SAÚDE E SEGURANÇA DO TRABALHO.

Disponível em: <https://observatoriosst.mpt.mp.br>. Acessado em: 16 mar. 2021.

R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2020. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

RUIZ, M. T.; BARBOZA, D. B.; SOLER, Z. A. S. G. Acidentes de trabalho: um estudo sobre esta ocorrência em um hospital geral. *Arq ciênc saúde*, v. 11, n. 4, p. 219-24, 2004.

SALERNO V. L. Considerações sobre a terceirização e as possíveis implicações nos riscos de acidentes de trabalho fatais: estudo da região de Campinas no período de 1990 a 1995. Dissertação de mestrado. Faculdade de Saúde Pública, São Paulo, 1998.

SANTANA, V. S.; FILHO, J. B. A.; OLIVEIRA, P. R. A.; BRANCO, A. B. Acidentes de trabalho: custos previdenciários e dias de trabalho perdidos. *Revista de Saúde Pública*, São Paulo, v.40, n.6, p.1004-1012, 2006.

SANTANA, V.; NOBRE, L.; WALDVOGEL, B. C. Acidentes de trabalho no Brasil entre 1994 e 2004: uma revisão. *Ciência & Saúde Coletiva*, São Paulo, v. 10, n. 4, p. 841-855, 2005.

SANTANA, V. S.; XAVIER, C.; MOURA, M. C. P.; OLIVEIRA, R.; ESPÍRITO-SANTO, J. S.; ARAÚJO, G. Gravidade dos acidentes de trabalho atendidos em serviços de emergência. *Revista de Saúde Pública*, São Paulo, v.43, n.5, p.750-760, 2009.

THERNEAU, T. *A package for survival analysis in s.* n. 38, p.1-5, 2015.

WALDVOGEL, B. qC. Acidentes do trabalho: os casos fatais a questão da identificação e da mensuração. *Segrac, Coleção Prodat Estudos e Análises*, Belo Horizonte, v.1, n.1, 2002.