

Aplicação do Modelo de Fragilidade Gama na análise da cessação do tabagismo no Brasil no ano de 2019

Cleane S. Santos[†], Cleide M. M. Lima, Valmaria R. S. Ferraz, Jackelia A. Silva

Departamento de Estatística, Universidade Federal do Piauí, Teresina, PI, Brasil.

Resumo: *É globalmente reconhecido que o consumo de tabaco é um fator de risco para várias doenças, tornando sua redução uma preocupação fundamental para a saúde pública. No Brasil, embora existam diversas literaturas sobre tabagismo e cessação, estas geralmente se limitam a estudos experimentais. Este estudo se propõe a analisar os dados da Pesquisa Nacional de Saúde de 2019, visando apresentar a Análise de Sobrevida como uma alternativa de investigação nos comportamentos de fumantes até a cessação do hábito. As variáveis consideradas incluem sexo, renda, estado civil e condições de saúde como Acidente Vascular Cerebral, hipertensão, doença pulmonar, prática de exercício físico e consumo de álcool. Aplicou-se o modelo paramétrico de Cox-Gompertz para analisar fatores que influenciam o tempo de tabagismo, assumindo tempos de consumo de tabaco independentes e riscos proporcionais. Além disso, utilizou-se o modelo de fragilidade univariado de Gama-Gompertz para lidar com fatores não observados que afetam a heterogeneidade dos dados. As análises indicaram uma maior probabilidade de cessação entre mulheres e indivíduos que praticam exercício físico, e o modelo de fragilidade indicou a presença de fatores não observados que afetam o tempo de tabagismo, uma vez que 0,17 da variabilidade dos dados não foi explicada pelas variáveis abordadas.*

Palavras-chave: *Sobrevida, Tabagismo, Heterogeneidade.*

Aplicação do Modelo de Fragilidade Gama na análise da cessação do tabagismo no Brasil no ano de 2019

Abstract: *It is globally recognized that tobacco consumption is a risk factor for various diseases, making its reduction a fundamental concern for public health. In Brazil, although there is a wealth of literature on smoking and cessation, these studies often limit themselves to experimental research. This study aims to analyze data from the 2019 National Health Survey, with the goal of presenting Survival Analysis as an alternative approach for investigating the behavior of smokers until they quit the habit. The considered variables include gender, income, marital status, and health conditions such as Stroke, hypertension, lung disease, physical exercise practice, and alcohol consumption. The Cox-Gompertz parametric model was applied to analyze factors influencing the duration of smoking, assuming independent tobacco consumption times and proportional risks. Furthermore, the univariate Gama-Gompertz frailty model was used to address unobserved factors affecting data heterogeneity. The analyses revealed a higher probability of smoking cessation among women and individuals who engage in physical exercise. The frailty model indicated the presence of unobserved factors influencing the duration of smoking, as 0.17 of the data variability was unexplained by the variables considered.*

Keywords: *Survival, Smoking, Heterogeneity.*

Introdução

O tabagismo é considerado pela Organização Mundial de Saúde (OMS) o principal fator de risco para seis das oito principais causas de morte no mundo, bem como para as quatro doenças crônicas não transmissíveis mais evitáveis e prevalentes: doenças cardiovasculares, diabetes, câncer e enfermidades respiratórias crônicas.

Em contrapartida, os principais benefícios individuais da cessação do tabagismo se traduzem em grandes reduções nas mortes por todas as suas causas atribuíveis e do câncer na população em geral. No entanto, apesar de todo o conhecimento divulgado em mídias governamentais e científicas sobre os riscos do tabaco, as tendências do seu consumo ainda são alarmantes.

[†] Autora correspondente: cleanesantos@ufpi.edu.br.

De acordo com o Relatório Global da OMS sobre Tendências na Prevalência do Uso de Tabaco 2000-2025, em 2020, 22,3% da população mundial usava tabaco, sendo a estimativa maior entre os homens (36,7%) em relação às mulheres. Já os dados nacionais divulgados pela Vigilância De Fatores De Risco e Proteção Para Doenças (VIGITEL, 2021), apontam que o percentual de fumantes com 18 anos ou mais no Brasil é de 9,1%.

A análise de sobrevivência assume um papel de significativa importância ao abordar a avaliação do período decorrido desde o início do consumo de tabaco até a sua interrupção. Este enfoque torna-se fundamental, uma vez que permite a identificação de grupos de indivíduos e momentos específicos da vida em que a cessação do tabagismo apresenta desafios particulares. A consequente compreensão destes padrões possibilita a implementação de políticas de combate ao tabagismo de maneira mais precisa, ajustada às necessidades desses grupos específicos. A análise de sobrevivência, técnica estatística aplicada na análise do tempo até a ocorrência de eventos, designados como "tempos de falha", reconhece a possibilidade de ocorrência de censura, quando a observação do evento de interesse não se dá por completo durante o período de análise. A presença de censura, inerente aos dados de sobrevivência, demanda a utilização de metodologias estatísticas específicas para sua investigação.

Em contextos de pesquisa em saúde, frequentemente se depara com a impossibilidade de inclusão de todos os fatores de risco relevantes, seja por limitações de dados individuais, seja devido a restrições financeiras ou de tempo. De acordo com Wienke, A (2011), nesses casos é útil considerar duas fontes de variabilidade nos dados de duração: Variabilidade contabilizada por fatores de riscos observáveis incluídos no modelo e heterogeneidade causada por fatores desconhecidos, e que, portanto, são teoricamente imprevisíveis, a utilização desses mecanismos dá-se por meio da aplicação do modelo de fragilidade (frailty model). Segundo Colosimo e Giolo (2006), nesse modelo, um efeito aleatório, denominado fragilidade, é introduzido na função de taxa de falha para descrever ainda possíveis associações entre os tempos de sobrevivência.

Estes modelos se desdobram em duas categorias, os univariados, que se concentram na análise de um único indivíduo em uma observação, e os multivariados, que consideram múltiplos tempos de sobrevivência agrupados. Ambos se apresentam como instrumentos úteis para a investigação de associações entre os tempos de sobrevivência em estudos de saúde.

Pesquisas recentes, exemplificadas pelo estudo de Vogt et al. (2021), têm aplicado técnicas de análise de sobrevivência, como o modelo de regressão de Cox, com o intuito de analisar o padrão de consumo de tabaco no Brasil. Neste contexto, constatou-se que fatores sociais e econômicos exercem influência significativa na cessação do tabagismo, o que enfatiza a necessidade de considerar tais variáveis no planejamento de políticas públicas direcionadas ao tratamento do tabagismo. No entanto, é importante ressaltar a escassez de estudos no Brasil que adotem modelos de fragilidade na análise de dados relacionados ao tabagismo, o que sublinha a urgência de empreender pesquisas mais abrangentes nesse domínio. Em face dessas considerações, trabalhos como o presente, que se vale de dados da Pesquisa Nacional de Saúde (PNS) de 2019 e emprega a metodologia de análise de sobrevivência conjugada com modelos de fragilidade, adquirem relevância incontestável no âmbito da compreensão dos fatores que impactam o tempo até a cessação do tabagismo, e, por conseguinte, no aprimoramento de políticas eficazes neste campo.

Diante das considerações e debates em torno da cessação do tabagismo, torna-se inquestionável a necessidade de conduzir pesquisas que possam enriquecer a compreensão dos fatores relacionados ao período necessário para alcançar a cessação desse hábito prejudicial à saúde. Nesse contexto, este estudo apresenta os resultados de uma análise aplicada aos dados da Pesquisa Nacional de Saúde (PNS) de 2019, empregando uma abordagem baseada na metodologia de análise de sobrevivência e integrando os princípios dos modelos de fragilidade. Assim, tem-se como objetivo aqui, analisar os efeitos de fatores não observados em estudo, sobre o tempo até a cessação do tabagismo, de indivíduos fumantes utilizando-se do modelo de fragilidade Gama-Gompertz.

Material e Métodos

Descrição dos Dados

O banco de dados utilizado neste estudo foi obtido por meio da base de dados da PNS (2019). De acordo com a Fundação Oswaldo Cruz (FIOCRUZ, 2023), a PNS parte de um projeto do Ministério da Saúde com foco nas condições de saúde da população brasileira e na avaliação do desempenho do sistema nacional de saúde tendo assim, o objetivo principal de produzir dados em âmbito nacional sobre a situação de saúde e os estilos de vida da população brasileira, bem como sobre a atenção à saúde, no que diz respeito ao acesso e uso dos serviços, às ações preventivas, à continuidade dos cuidados e ao financiamento da assistência.

A base de dados de estudo não considerou todas as informações presentes na Pesquisa Nacional de Saúde. Inicialmente, para analisar os possíveis fatores que podem influenciar o hábito de fumar, considerou-se estimar o tempo de tabagismo. Essa variável foi calculada, baseando-se na idade do indivíduo, em que idade começou a fumar e quanto tempo levou para parar de fumar (caso tenha parado).

Dessa forma, foram selecionados 20.498 indivíduos, sendo que estes afirmaram que iniciaram a fumar em algum momento da vida. Desconsiderou-se indivíduos com idade menor que 18 anos. Dentre as observações de estudo, tem-se que 6.589 relataram ainda serem fumantes na data da entrevista, resultando assim, em dados censurados. Denomina-se, censura à direita, visto que a falha não ocorreu durante o tempo de estudo. E entendeu-se como falha os indivíduos que no ato da pesquisa informaram que haviam parado de fumar.

No âmbito deste trabalho, foram contempladas diversas variáveis de relevância para a análise, compreensão e caracterização do grupo de estudo. Estas variáveis abrangeram a presença de hipertensão arterial, diagnosticada por um profissional médico, bem como a ocorrência de Acidente Vascular Cerebral (AVC) ou derrame, devidamente confirmados por um profissional de saúde. Além disso, foi investigado o padrão de consumo de bebidas alcoólicas, a existência de diagnóstico de doenças pulmonares crônicas, como enfisema pulmonar, bronquite crônica ou Doença Pulmonar Obstrutiva Crônica (DPOC) realizado por um médico. Também se incluiu a variável estado civil, que se relaciona com o estado civil do indivíduo, abrangendo categorias como solteiro, casado, divorciado, viúvo, entre outras. A prática de exercício físico nos últimos três meses também foi objeto de análise, bem como a renda domiciliar per capita, que se classificou em faixas de rendimento. Por fim, a variável sexo foi considerada, indicando o gênero do entrevistado, sendo categorizado como masculino ou feminino. A inclusão destas variáveis foi fundamental para a contextualização e aprofundamento da investigação realizada no presente estudo.

Análises Estatísticas

Modelo de Regressão de Cox

O modelo de riscos proporcionais de Cox (Cox, 1972) permite modelar dados de sobrevivência na presença de covariáveis, por meio da função de taxa de falha. A expressão do modelo de Cox em termos da função de taxa de falha é dada, para o i -ésimo indivíduo em um dado tempo t por:

$$h(t | x) = h_0(t)g(\mathbf{x}'\boldsymbol{\beta}) \quad (1)$$

em que $g(\mathbf{x}'\boldsymbol{\beta})$ é uma função não – negativa que deve ser especificada, tal que $g(0) = 1$.

O modelo é composto pelo produto dois componentes $h_0(t)$ que é não-paramétrico e uma função não-negativa do tempo. E o segundo componente $g(\mathbf{x}'\boldsymbol{\beta})$ que é paramétrico e frequentemente utilizado na seguinte maneira multiplicativa:

$$g(\mathbf{x}'\boldsymbol{\beta}) = \exp \{x'\boldsymbol{\beta}\} = \exp \{\beta_1x_1 + \dots + \beta_px_p\} \quad (2)$$

que $\boldsymbol{\beta}$ é o vetor de parâmetros associado às covariáveis.

Também denominado modelo de taxa de falhas proporcionais a suposição básica para o uso do modelo de regressão de Cox, é que as taxas de falhas sejam proporcionais ou, de forma equivalente para este modelo, que as taxas de falhas sejam também proporcionais (COLOSIMO e GIOLO, 2006). Então, a razão das funções de taxa de falha para os indivíduos i e j é dada por:

$$\frac{h(t | x_i)}{h(t | x_j)} = \frac{h_0(t) \exp(\beta' x_i)}{h_0(t) \exp(\beta' x_j)} = \exp(\beta'(x_i - x_j)) \quad (3)$$

e não depende do tempo. Por exemplo, se um indivíduo no início do estudo tem uma taxa de falha igual a duas vezes a de um segundo indivíduo, então esta razão será a mesma para todo o período de acompanhamento (COLOSIMO e GIOLO, 2006).

Para interpretação dos coeficientes estimados utiliza-se a propriedade de taxas de falhas proporcionais. Tomando-se a razão das taxas de falha de dois indivíduos, i e j , que possuem os mesmos valores para as covariáveis com exceção da i -ésima, tem-se:

$$\frac{\lambda(t | x_i)}{\lambda(t | x_j)} = \exp \{ \beta (x_{il} - x_{jl}) \} \quad (4)$$

que pode ser interpretado como a razão de taxas de falha instantânea no tempo t . No entanto, como esta razão é constante durante todo o período de observação, pode-se omitir a palavra instantânea da interpretação. Por exemplo, suponha que x_1 seja uma covariável dicotômica indicando pacientes hipertensos. Então, a taxa de morte entre os hipertensos é $\exp(\beta_1)$ vezes a de pacientes com pressão normal, mantidas fixas as outras covariáveis, além disso, pode-se ainda considerar a interpretação das estimativas β' s que quando apresenta-se negativa em variáveis quantitativas, implica que quanto maior o valor de x_1 , menor a probabilidade de sobrevivência estimado. E quando, $\hat{\beta}$ é negativo em uma variável categórica, entende-se que os indivíduos do suposto grupo i apresentam probabilidade de sobrevivência menor do que a dos indivíduos do possível grupo j , portanto a interpretação dos coeficientes para este caso utiliza a razão de taxas de falha como medida de associação.

No modelo de Cox a função de sobrevivência é definida por:

$$S(t_i | x_i) = \exp \left\{ - \int_0^{t_i} h_0(u) \exp \{ \beta' x_i \} du \right\} = [S_0(t_i)]^{\exp \{ \beta' x_i \}} \quad (5)$$

onde $S_0(t_i)$ é uma função de sobrevivência base, ou seja, a função de sobrevivência de um indivíduo com todas as covariáveis iguais a zero. Neste trabalho, considera-se uma abordagem paramétrica para o modelo de Cox (1) onde foi feita uma aplicação das distribuições Weibull e Gompertz para verificar a que melhor se adéqua.

A Tabela 1 mostra as funções de densidade de probabilidade ($f(t)$), funções de risco ($h_0(t)$), função sobrevivência ($S_0(t)$) e função de risco acumulado base ($H_0(t)$) para as distribuições Gompertz e Weibull que serão consideradas neste estudo.

Tabela 1: Funções de densidade de probabilidade ($f(t)$), funções de risco ($h_0(t)$), função sobrevivência ($S_0(t)$) e função de risco acumulado base ($H_0(t)$) para as distribuições Gompertz e Weibull.

	Gompertz	Weibull
$f(t)$	$\lambda \exp(\gamma t) \exp \left\{ - \left(\frac{\lambda}{\gamma} \right) (\exp(\gamma t) - 1) \right\}$	$\frac{\lambda}{\gamma} t^{\gamma-1} \exp \left\{ - \left(\frac{t}{\lambda} \right)^\gamma \right\}$
$h_0(t)$	$\lambda \exp(\gamma t)$	$\frac{\gamma}{\lambda} t^{\gamma-1}$
$H_0(t)$	$\frac{\lambda}{\gamma} (\exp(\gamma t) - 1)$	$\left(\frac{t}{\lambda} \right)^\gamma$
$S_0(t)$	$\exp \left\{ - \left(\frac{\lambda}{\gamma} \right) (\exp(\gamma t) - 1) \right\}$	$\exp \left\{ - \left(\frac{t}{\lambda} \right)^\gamma \right\}$

Fonte: Elaboração própria, 2023.

Por exemplo, considerando a função de sobrevivência e risco base da distribuição Gompertz, temos o modelo de riscos proporcionais de Cox-Gompertz, cuja função de risco fica assim definida:

$$h(t|\mathbf{x}) = \lambda \exp(\gamma t) \exp(\boldsymbol{\beta}' \mathbf{x}), \quad (6)$$

Substituindo a função ($S_0(t)$) da distribuição Gompertz em (5) a função de sobrevivência é reescrita da seguinte forma:

$$S(t|\mathbf{x}) = \left[\exp - \frac{\lambda}{\gamma} (\exp(\gamma t) - 1) \right]^{\exp\{\boldsymbol{\beta}' \mathbf{x}\}} \quad (7)$$

Inferências

Em análise de sobrevivência, o método mais utilizado para estimação dos parâmetros $\theta = (\gamma, \lambda, \boldsymbol{\beta})$, do modelo 6, é o método de máxima verossimilhança, devido ao fato de o mesmo ser capaz de incorporar dados censurados.

Considere o conjunto de dados $W = (n, \mathbf{t}, \boldsymbol{\delta}, \mathbf{X})$, em que $\mathbf{t} = (t_1, \dots, t_n)$ são os tempos de falha observados, $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)'$ são os tempos de falha censurados e \mathbf{X} é uma matriz $n \times k$ contendo as covariáveis. Supondo que os dados sejam independentes e identicamente distribuídos, a função de verossimilhança com dados censurados, segundo Lawless (2011) é definida por:

$$L(\theta) = \prod_{i=1}^n [h(t_i|\mathbf{x}_i)]^{\delta_i} S(t_i|\mathbf{x}_i), \quad (8)$$

em que δ_i é o indicador de censura. Por exemplo, substituindo as funções 6 e 7 em 8 tem-se a função de verossimilhança para o modelo Cox-Gompertz.

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n \left[\lambda \exp(\gamma t) \exp\{\boldsymbol{\beta}' \mathbf{x}_i\} \right]^{\delta_i} \left[\exp\left\{-\frac{\lambda}{\gamma} (\exp(\gamma t) - 1)\right\} \right]^{\exp\{\boldsymbol{\beta}' \mathbf{x}_i\}}. \quad (9)$$

As estimativas de máxima verossimilhança são obtidas através da maximização numérica da função log-verossimilhança (9). Os parâmetros foram estimados utilizando o pacote *optim* do *software* R e o método "BFGS" para maximização.

Os intervalos de confiança para os parâmetros do modelo são baseados nas propriedades de normalidade assintótica dos estimadores de máxima verossimilhança. Se $\hat{\boldsymbol{\theta}}$ denota os estimadores de máxima verossimilhança de $\boldsymbol{\theta}$ então a distribuição de $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ é aproximada por uma distribuição normal q-variada com média zero e matriz de covariância $I^{-1}(\hat{\boldsymbol{\theta}})$, onde $I(\boldsymbol{\theta})$ é a matriz de informação observada. Assim, um intervalo de confiança assintótico com $100(1 - \alpha)\%$ para cada parâmetro θ_i é:

$$IC(\boldsymbol{\theta}, 100(1 - \alpha)\%) = \left(\hat{\theta}_i - z_{\alpha/2} \sqrt{\hat{V}(\theta_i)}, \hat{\theta}_i + z_{\alpha/2} \sqrt{\hat{V}(\theta_i)} \right) \quad (10)$$

onde $\hat{V}(\theta_i)$ é o elemento da diagonal principal de $I^{-1}(\hat{\boldsymbol{\theta}})$ correspondente a cada parâmetro e $z_{\alpha/2}$ é o quantile $(1 - \alpha)\%$ da distribuição normal padrão.

Para testar as hipóteses relacionadas ao vetor de parâmetros $\boldsymbol{\theta}$, três testes são, em geral, utilizados para esta finalidade: Wald, Escore e Razão de Verossimilhança (COLOSIMO e GIOLO, 2006). Critérios de seleção de modelos como o critério de informação de Akaike (AIC) proposto por Akaike (1974), e o critério de informação bayesiano (BIC), proposto por Schwarz et al. (1978), são frequentemente utilizados para selecionar modelos em diversas áreas. Os melhores modelos são considerados aqueles que apresentam menores valores de AIC e BIC. Para avaliar a adequação do modelo de Cox são utilizados os resíduos de Cox-Snell (COX e SNELL, 1968); resíduos martingale; resíduos deviance.

Modelo de Fragilidade Univariada

Os modelos de fragilidade para dados de sobrevivência univariada, leva em conta que a população é não homogênea. A heterogeneidade pode ser explicada por covariáveis, mas quando importantes covariáveis não são incorporadas no modelo, isso conduz a heterogeneidade não observada (Colosimo e Giolo, 2006). O modelo de fragilidade é uma extensão do modelo de Cox (COX, 1972), onde o risco individual depende de uma variável aleatória não observável, não negativa, Z , a qual age de forma multiplicativa ou aditiva na função de risco. Neste estudo, considerando o modelo de fragilidade univariado multiplicativo, tem-se que a função de risco na presença de covariáveis no tempo t para o i -ésimo indivíduo é dado por:

$$h(t|\mathbf{x}, z_i) = z_i h_0(t) \exp \{ \mathbf{x}'\beta \} \quad (11)$$

em que z_1, \dots, z_n representam as fragilidades individuais, $h_0(t)$ é a função de risco base comum a todos os indivíduos, x é o vetor de covariáveis e β é o vetor de coeficientes. O modelo de fragilidade (11) assume estrutura de risco proporcional condicionado ao efeito aleatório. Como z_i representa um valor da variável aleatória não observável Z , o risco individual cresce quando $z_i > 1$, decresce se $z_i < 1$ e para $z_i = 1$ o modelo de fragilidade (11) reduz ao modelo de risco proporcional de Cox (1)

O fato da variável de fragilidade atuar de forma multiplicativa na função de risco implica quanto maior for o valor da variável de fragilidade, maior será a chance de ocorrer a falha. Dessa forma, quanto maior for z_i , mais “frágeis” as observações pertencentes ao indivíduo i estão para falhar. Assim, é esperado que o evento de interesse ocorra para os indivíduos mais “frágeis”.

Considerando o modelo (1) sem a presença de covariáveis, a função de risco para o i -ésimo indivíduo é dado por:

$$h(t|z_i) = z_i h_0(t) \quad (12)$$

com função de sobrevivência

$$S(t|z_i) = (S_0(t))^{z_i}, \quad (13)$$

em que $S_0(t)$ é a função de sobrevivência comum a população. A fim de obter a função de verossimilhança, é necessário encontrar a função de sobrevivência não condicional, assim é preciso integrar o termo de fragilidade, isto é:

$$S(t) = E_z (S(t|z)) = \int_0^\infty S(t|z)g(z)dz = \int_0^\infty e^{-H_0(t)z}g(z)dz = L_g [H_0(t)] \quad (14)$$

em que $g(z)$ é a função densidade de probabilidade da variável de fragilidade e $L_g [H_0(t)]$ denota a transformada de Laplace aplicada na função de risco acumulada, $H_0(t)$, o que mostra a importância desta transformação nos modelos de fragilidade (Wienke, 2011)

O modelo de fragilidade para o indivíduo i ($i = 1, \dots, n$) fica, então, expresso por:

$$\lambda_i(t) = z_i h_0(t) \exp (\mathbf{x}'_i \beta) . \quad (15)$$

Em que β é o vetor de parâmetros desconhecidos associados às covariáveis x_i , $h_0(t)$ é a função de taxa de falha de base não especificada e z_i, \dots, z_n são os valores das fragilidades assumidas serem uma amostra proveniente de variáveis aleatórias Z_i independentes e identicamente distribuídas com distribuição de probabilidade conhecida de média 1 e com variância desconhecida. Dessa forma a fragilidade introduzida neste modelo além de explicar a heterogeneidade dos indivíduos permite ainda avaliar o efeito de covariáveis que por algum motivo não foram incluídas na análise.

Modelo de Fragilidade Gama-Gompertz

Um problema importante em modelos de fragilidade é na escolha da distribuição para o efeito aleatório. Devido à forma como o termo de fragilidade atua na função de risco, as candidatas à distribuição de fragilidade são supostamente não negativas, usualmente contínuas e não dependentes do tempo como, por exemplo, as distribuições Gama, log-Normal e Gaussiana inversa (HOUGAARD, 1995). Segundo Elbers e Ridder (1982) quando se trabalha com fragilidade é necessário que a distribuição do efeito aleatório tenha média finita para o modelo ser identificável. Nesse estudo, considerando que a variável aleatória Z que representa a fragilidade segue uma distribuição Gama, $Z \sim G(a, b)$ e $a = b = \alpha$, a fim de garantir que a média seja igual a 1, para o modelo ser identificável (TOMAZELLA, 2003), a sua função densidade de probabilidade é dada por:

$$f(z) = \frac{\alpha^\alpha}{\Gamma(\alpha)} z^{\alpha-1} \exp\{-\alpha z\}, \quad z > 0, \alpha > 0 \quad (16)$$

em que, $E(Z) = 1$ e $Var(Z) = 1/\alpha$.

Assim, sem capturar o efeito de covariáveis, sua função de sobrevivência não condicional será dada por:

$$S(t) = L_z(H_0(t)) = \left(1 + \frac{H_0(t)}{\alpha}\right)^{-\alpha} \quad (17)$$

com $LZ(\cdot)$ representando a transformada de Laplace e $H_0(t)$ a função de risco acumulada. Consequentemente, a função de risco não condicional será dada por:

$$h(t) = \frac{\alpha h_0(t)}{\alpha + H_0(t)} \quad (18)$$

Considerando a distribuição Gompertz como função de risco base, tem-se o modelo de fragilidade Gama-Gompertz, onde agora a função de sobrevivência não-condicional (17) e risco não condicional (18), na presença de covariáveis são dadas respectivamente por:

$$S(t|x) = L_Z(H_0(t)) = \left(1 + \frac{H_0(t) \exp\{x'\beta\}}{\alpha}\right) \quad (19)$$

e

$$h(t|x) = \frac{\exp(x'\beta)}{1 + \frac{H_0(t) \exp(x'\beta)}{\alpha}} h_0(t) \quad (20)$$

A variância da variável de fragilidade, neste caso $1/\alpha$, quantifica a heterogeneidade não observável entre os indivíduos. Quando α é grande, implica em pouca variabilidade entre os indivíduos, ou seja, uma população mais homogênea, assim os valores das variáveis de fragilidade serão iguais a 1, consequentemente, a distribuição gama fica degenerada no ponto 1 e, com isso, tem-se o modelo de riscos proporcionais padrão para dados independentes. Por outro lado, com valor pequeno de α indica que há uma grande heterogeneidade não observável entre os indivíduos. A comparação entre os modelos regressão de Gompertz com fragilidade gama são feitas considerando o critério de informação de AIC e do BIC. O menor AIC e o BIC indica o melhor modelo ajustado aos dados

Estimação dos parâmetros do modelo

Considere um conjunto de dados $W = (n, t, \delta, \mathbf{X})$, onde $\mathbf{t} = (t_1, \dots, t_n)$ são os tempos de falhas observados e $\delta = (\delta_1, \dots, \delta_n)$ são os tempos de falhas censurados. Supondo que os dados são independentes e identicamente distribuídos, e vêm de uma distribuição com função de risco e sobrevivência especificados por $h(\cdot; \theta)$ e $S(\cdot; \theta)$, respectivamente, onde $\theta = (\alpha, \lambda, \gamma, \beta)$ denota um vetor de parâmetros, com $\beta = (\beta_1, \dots, \beta_k)$ coeficientes relativos às covariáveis que

influenciam no tempo de sobrevivência dos indivíduos. A função de verossimilhança não-condicional de θ para o modelo de fragilidade gama pode ser escrita por:

$$L(\theta) = \prod_{i=1}^n [h_i(\cdot; \theta)]^{\delta_i} S_i(\cdot; \theta), \quad (21)$$

$$L(\theta) = \prod_{i=1}^n \left(\frac{\exp\{\mathbf{x}'\beta\}}{1 + \left(\frac{H_0(t)\exp\{\mathbf{x}'\beta\}}{\alpha}\right)} h_0(t) \right)^{\delta_i} \left(1 + \frac{H_0(t)\exp\{\mathbf{x}'\beta\}}{\alpha} \right)^{-\alpha}. \quad (22)$$

onde δ_i é o indicador de censura, $h_0(t)$ e $S_0(t)$ são funções de risco e sobrevivência base de uma distribuição qualquer. Substituindo $h_0(t)$ e $S_0(t)$ pelas funções de risco e sobrevivência da distribuição Gompertz, temos a função de verossimilhança não-condicional do modelo de fragilidade Gama-Gompertz, reescrita por:

$$L(\theta) = \prod_{i=1}^n \left(\frac{\lambda \exp(\gamma t) \exp\{\mathbf{x}'\beta\}}{1 + \left(\frac{\frac{\lambda}{\gamma}(\exp(\gamma t) - 1) \exp\{\mathbf{x}'\beta\}}{\alpha}\right)} \right)^{\delta_i} \left(1 + \frac{\frac{\lambda}{\gamma}(\exp(\gamma t) - 1) \exp\{\mathbf{x}'\beta\}}{\alpha} \right)^{-\alpha}. \quad (23)$$

A função de log-verossimilhança de θ pode ser escrita por:

$$\begin{aligned} \log(L(\theta)) = \sum_{i=1}^n \left\{ \delta_i \log \left(\frac{\lambda \exp(\gamma t) \exp\{x'\beta\}}{1 + \left(\frac{\frac{\lambda}{\gamma}(\exp(\gamma t) - 1) \exp\{x'\beta\}}{\alpha}\right)} \right) \right\} \\ - \alpha \log \left(1 + \frac{\frac{\lambda}{\gamma}(\exp(\gamma t) - 1) \exp\{x'\beta\}}{\alpha} \right) \end{aligned} \quad (24)$$

A estimativas de máxima verossimilhança são obtidas através de maximização numérica da função de log-verossimilhança (24). Foi utilizado o pacote Optim do software R e o método "BFGS" para maximização.

Resultados e Discussões

O tempo até a cessação do tabagismo pode variar de pessoa para pessoa, dependendo de vários fatores, estes serão discutidos por meio dos resultados obtidos. Na Tabela 2 abaixo é possível visualizar as estimativas dos percentuais de censura e falha para as variáveis que foram consideradas no estudo. Tem-se que dos 20.498 indivíduos que iniciaram o consumo do tabaco por meio do cigarro em algum momento da vida, 67,86% deles apresentaram falha, ou seja, cessaram o consumo do cigarro e 32,14% deles apresentaram censura, permanecendo então fumando até o momento da entrevista da PNS.

Em relação ao sexo, uma maioria de 50,05% é do sexo masculino. Tem-se a maioria dos indivíduos são solteiros (41,14%) e possuem renda de até um salário-mínimo (51,18%). Em relação, as variáveis relacionadas a questões de saúde podem-se observar que 65,48% das pessoas não praticam exercício físico e 56,43% consomem bebida alcoólica. Além disso, considerando toda a população de estudo 2,45% destas apresentaram alguma doença crônica no pulmão (tais como enfisema ou bronquite) e 34,72% são hipertensos. Segundo os estudos já mencionados, tais questões de saúde mostraram-se como possíveis fatores de risco que poderiam ser ocasionadas e/ou intensificadas a partir do consumo do tabaco. Portanto, o estudo concentrou-se em avaliar a ligação destas variáveis com o tempo até a cessação do tabagismo.

Nesse estudo, os indivíduos entrevistados pela PNS apresentaram um tempo mediano de 40.5 anos praticando o ato de fumar. Além disso é possível observar na Figura (1) que existe

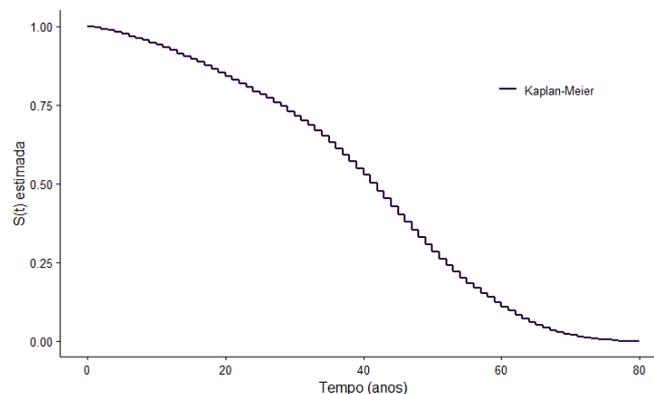
Tabela 2: Estatísticas descritivas para as variáveis referente ao consumo de tabaco, Brasil - PNS (2019)

Variáveis	Censura	Falha
Sexo		
Feminino	2.849 (18,25%)	7.411 (31,70%)
Masculino	3.740 (13,90%)	6.498 (36,15%)
Estado Civil		
Casado (a)	1.925 (9,39%)	5.673 (27,68%)
Divorciado (a)	1.322 (6,45%)	3.144 (15,34%)
Solteiro (a)	3.342 (16,30%)	5.092 (24,84%)
Renda		
Até 1 salário-mínimo	3.602 (17,57%)	6.889 (33,61%)
1-3 salários-mínimos	2.259 (11,02%)	5.113 (24,94%)
Mais de 3 salários-mínimos	728 (3,55%)	1.907 (9,30%)
Prática Exercício Físico		
Sim	1.879 (9,17%)	5.198 (25,36%)
Não	4.710 (22,98%)	8.711 (42,50%)
Consome Bebida Alcoólica		
Sim	4.034 (19,68%)	7.533 (36,75%)
Não	2.555 (12,46%)	6.376 (31,11%)
Apresentou AVC		
Sim	165 (0,80%)	496 (2,42%)
Não	6.424 (31,34%)	13.413 (65,44%)
Possui outra Doença no Pulmão		
Sim	148 (0,72%)	354 (1,73%)
Não	6.441 (31,42%)	13.555 (66,13%)
Hipertensão		
Sim	1.663 (8,11%)	5.455 (26,61%)
Não	4.926 (24,03%)	8.454 (41,24%)

Fonte: Elaboração própria, 2023.

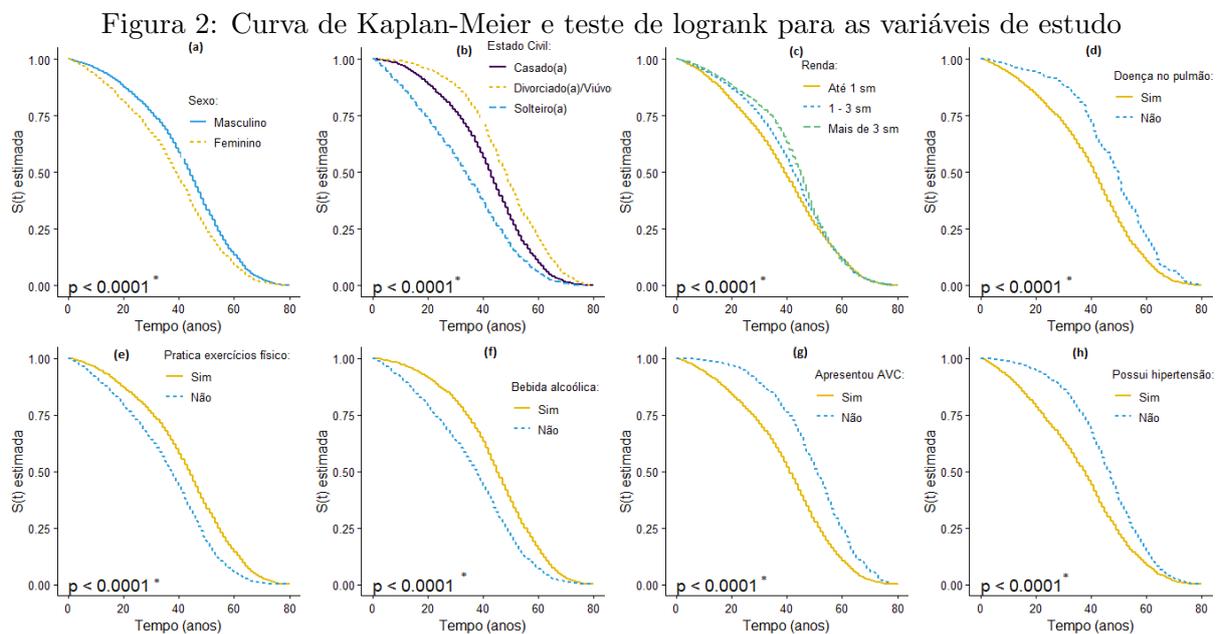
uma queda exponencial indicando, dessa forma, que quanto maior o tempo fumando menor é a probabilidade de o indivíduo continuar fumando, por exemplo, a probabilidade de uma pessoa continuar fumando 20 anos após ter iniciado o consumo é de 89,66% enquanto a de o indivíduo que fuma há 50 anos, tem probabilidade de 28,32% de permanecer fumando.

Figura 1: Curva de sobrevivência global estimadas pelo método de Kaplan-Meier para os dados de cessação de tabagismo



Fonte: Elaboração própria, 2023.

Na Figura (2), considerando os dados de tempo até a cessação do tabagismo dos 20.498 indivíduos, tem-se a apresentação da curva de sobrevivência estratificada para as categorias das covariáveis de estudo e a aplicação do teste de logrank. Através da análise gráfica, é possível notar que existem diferenças entre todas as covariáveis de estudo. Por exemplo, percebe-se que indivíduos do sexo masculino possuem uma maior probabilidade de sobrevivência do que pessoas do sexo feminino, ou seja, as mulheres passam menos tempo fumando e apresentam maior probabilidade de parar de fumar. Assim como, quem pratica exercício físico e consome bebida alcoólica, apresentam maior probabilidade de continuar fumando, já quem apresentou AVC, Hipertensão e doença no pulmão possuem menor probabilidade de continuar fumando, portanto, esses estão mais propícios a cessarem a consumo do cigarro. Dessa forma, utilizou-se o teste de logrank objetivando verificar quais grupos apresentam diferenças nas curvas de sobrevivência entre si, nota-se então que todas as covariáveis realmente apresentaram diferenças significativas entre as curvas de sobrevivência.



Fonte: Elaboração própria, 2023.

Ajuste do Modelo de Regressão Paramétrico de Cox-Gompertz

Realizou-se um estudo comparativo do modelo de riscos proporcionais de Cox-Gompertz com o modelo de Cox-Weibull ambos sem covariáveis para identificar a melhor distribuição que se adequa para o tempo de sobrevivência (tempo até a cessação do tabagismo). A Tabela 5 abaixo apresenta os resultados do modelo de estudo, assim como, os critérios de Akaike e critério de informação bayesiano (BIC) que são utilizados para a seleção de modelos. Logo após, a Figura 7 retrata o método gráfico da curva de sobrevivência versus a curva do modelo estimado, que contém a função de sobrevivência estimada pelo método de Kaplan-Meier e as curvas de sobrevivência estimadas para cada modelo. Analisando tais gráficos, nota-se que o modelo Cox-Gompertz efetivamente apresenta melhor adequação a curva de Kaplan-Meier. Tendo em vista a análise gráfica e os resultados obtidos por meio do AIC e BIC para cada um dos modelos, o modelo Cox-Gompertz foi escolhido por ser um modelo mais parcimonioso que indica uma melhor qualidade de ajuste aos tempos de sobrevivência estimado.

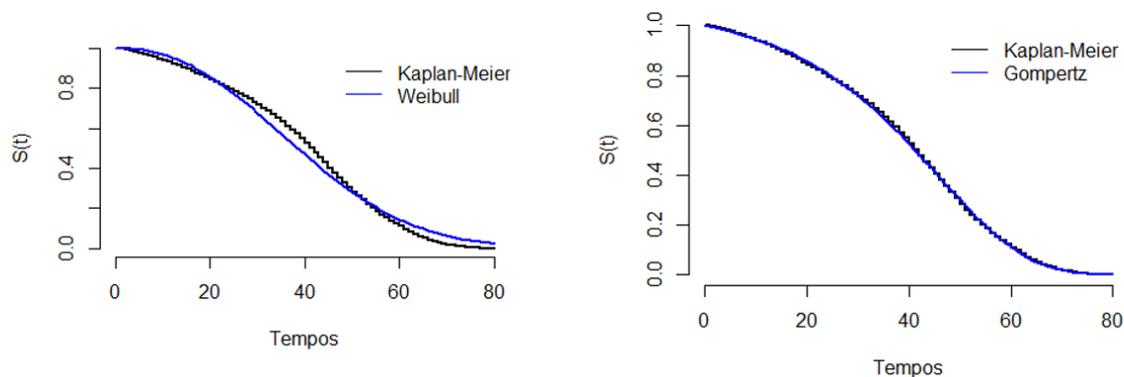
A Tabela 4 apresenta as estimativas de máxima verossimilhança, erro padrão, razão da taxa de falha, intervalo de confiança e p-valor para os parâmetros do modelo proposto. Nota-se que todas as variáveis apresentaram-se significativas, como já esperado pelos resultados do teste

Tabela 3: Estimativas de máxima verossimilhança (EMV), Erro padrão (EP) e Intervalo de confiança (IC 95%) para os parâmetros dos modelos ajustados.

Parâmetros	Gompertz			Weibull		
	EMV	EP	IC (95%)	EMV	EP	IC (95%)
λ	0,0575	0,0005	[0,0567; 0,0040]	2,3066	0,0159	[2,2798; 44,9250]
γ	0,0041	0,0001	[0,0584; 0,0043]	45,174	0,1514	[2,3321; 45,4229]
AIC	124.144,40			125.954,50		
BIC	124.120,50			125.970,40		

Fonte: Elaboração própria, 2023.

Figura 3: Função de sobrevivência estimada pelo método de Kaplan-Meier com as curvas de sobrevivência estimada para cada modelo ajustado



Fonte: Elaboração própria, 2023.

de log-rank. Para o modelo ajustado $\hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_7, \hat{\beta}_8, \hat{\beta}_7, \hat{\beta}_8$ e $\hat{\beta}_9$ apresentam parâmetros negativos, indicando um menor risco de falha, implicando que, indivíduos fumantes com renda de 1-3 e maior que 3 salários mínimos, quem apresentou Acidente Vascular Cerebral, doença no pulmão e hipertensão tendem a apresentar menor probabilidade de falhar, ou seja, parar o consumo do tabaco. Já as estimativas $\hat{\beta}_1, \hat{\beta}_5, \hat{\beta}_6$ e $\hat{\beta}_{10}$, possuem parâmetros positivos, o que implica que pessoas do sexo masculino, solteiros, pessoas que praticam exercícios físicos e quem consomem bebidas alcoólicas, apresentam maior probabilidade de falhar, ou seja, esses indivíduos apresentam maiores chances de pararem de fumar.

Analisando a Razão de Taxa de Falha do modelo tem-se que uma pessoa do sexo feminino apresenta 54,2% mais de chance de apresentar falha (parar o consumo do tabaco) em relação a indivíduos do sexo masculino. Quanto à renda, quem recebe de 1 a 3 salários-mínimos possui 17,2% de chance a menos de parar de fumar, já quem possui renda superior a três salários-mínimos apresenta 36,3% menos chance de falhar (parar de fumar) quando comparados a quem recebe até 1 salário-mínimo. Os divorciados e/ou viúvos também possuem 38,7% menos chance de parar de fumar, enquanto os solteiros 46,6% mais chance de parar de fumar do que os casados. Em relação aos praticantes de exercícios físicos, estes apresentam 53,3% a mais de chance de parar de fumar do que quem não pratica. No entanto, quem apresentou AVC possui 27,8% de chance a menos de parar de fumar do que quem não apresentou AVC. Indivíduos que têm doença no pulmão e hipertensão possuem, respectivamente, 25,1% e 36,0% menos de chance de parar de fumar do que quem não possui essas características e as pessoas que consomem bebida alcoólica possuem 58,1% mais de chance de parar de fumar do que quem não consome.

Considerando os resultados previamente apresentados, podemos concluir que as estimativas relacionadas ao sexo indicam que o tempo de tabagismo é menor para as mulheres, que também

Tabela 4: Estimativas de máxima verossimilhança para os parâmetros do modelo de riscos proporcionais de Cox com funções de risco base Gompertz

Parâmetros	Estimativas	Erro Padrão	RTF*	IC (95%)	p-valor
γ	0.071	0.001	-	-	< 0.001
λ	0.002	0.000	-	-	< 0.001
$\hat{\beta}_1$ (Sexo [Feminino])	0.433	0.018	1.542	[1.496; 1.587]	< 0.001
$\hat{\beta}_2$ (Renda [1-3 SM])	-0.188	0.019	0.828	[0.802; 0.853]	< 0.001
$\hat{\beta}_3$ (Renda [\geq 3 SM])	-0.450	0.027	0.637	[0.608; 0.665]	< 0.001
$\hat{\beta}_4$ (Estado civil [Divorciado/Viúvo])	-0.490	0.023	0.613	[0.589; 0.636]	< 0.001
$\hat{\beta}_5$ (Estado Civil [Solteiro])	0.383	0.020	1.466	[1.417; 1.514]	< 0.001
$\hat{\beta}_6$ (Exercício Físico [Sim])	0.427	0.018	1.533	[1.486; 1.579]	< 0.001
$\hat{\beta}_7$ (AVC [Sim])	-0.325	0.047	0.722	[0.667; 0.777]	< 0.001
$\hat{\beta}_8$ (Doença Pulmão [Sim])	-0.290	0.054	0.749	[0.681; 0.815]	< 0.001
$\hat{\beta}_9$ (Hipertensão [Sim])	-0.446	0.018	0.640	[0.621; 0.658]	< 0.001
$\hat{\beta}_{10}$ (Bebida Alcoólica [Sim])	0.458	0.018	1.581	[1.533; 1.628]	< 0.001

Fonte: Elaboração própria, 2023.

apresentam maiores chances de cessar o hábito de fumar. Esses resultados corroboram com as descobertas mencionadas por Vogt et al. (2021), que também apontaram uma maior adesão ao tratamento e uma maior taxa de sucesso na cessação do tabagismo entre as mulheres.

Além disso, características associadas aos hábitos de vida reforçam a importância da prática de atividades saudáveis como um elemento-chave para a cessação do tabagismo. Indivíduos que se exercitam regularmente tendem a fumar por menos tempo. No entanto, é importante notar que o consumo de álcool também foi associado a um menor tempo de tabagismo, embora seja reconhecido que existem estudos que apontam uma relação entre o consumo de álcool e o tabagismo. Esses resultados podem ser explicados pela limitação da base de dados da Pesquisa Nacional de Saúde (PNS) em identificar o nível específico de consumo de álcool que pode afetar a relação entre álcool e tabaco.

Além disso, ao analisar a Figura 10, pode-se observar que as taxas de falha proporcionais para as variáveis de estudo não indicam uma violação significativa da suposição de taxas de falha proporcionais. Embora as curvas não sejam perfeitamente paralelas ao longo do tempo, não observa-se desvios substanciais dessa característica. Isso sugere uma boa qualidade do modelo ajustado. No entanto, posteriormente, será realizado uma análise dos resíduos para fins de confirmação.

Considerando a função de sobrevivência para a distribuição Gompertz apresentada em (1) e a função de sobrevivência do modelo de Cox em (5), tem-se que a função de sobrevivência Cox-Gompertz para o modelo final é dada por:

$$\hat{S}(t_i | x_i) = \left[\exp\left(\frac{0.002}{0.071}\right) \times (\exp(0.071 - 1)) \right]^{\exp(\hat{\beta}'x_i)} \quad (25)$$

em que

$$\exp(\hat{\beta}'x_i) = 0,433 \cdot \text{Sexo}(\text{Feminino}) +, \dots, 0,458 \cdot \text{BebidaAlcoolica}(\text{Sim})$$

A Figura 6 apresenta os gráficos dos resíduos de Cox-Snell, resíduos martingale e resíduos deviance. Por meio dos gráficos (a) e (b), pode-se observar que os resíduos de Cox-Snell seguem uma distribuição exponencial padrão, indicando uma boa qualidade de ajuste global do modelo. E nos gráficos (c) e (d) nota-se que não há presença de pontos que possam ser considerados atípicos (outliers).

Após as análises, verificou-se que o modelo de regressão paramétrica de Cox-Gompertz se mostrou adequado para avaliar os fatores que influenciam a cessação do tabagismo entre os

Figura 4: Curvas de sobrevivência estimadas pelo modelo de regressão Cox-Gompertz para os indivíduos fumantes de estudo para avaliação das taxas de falhas proporcionais

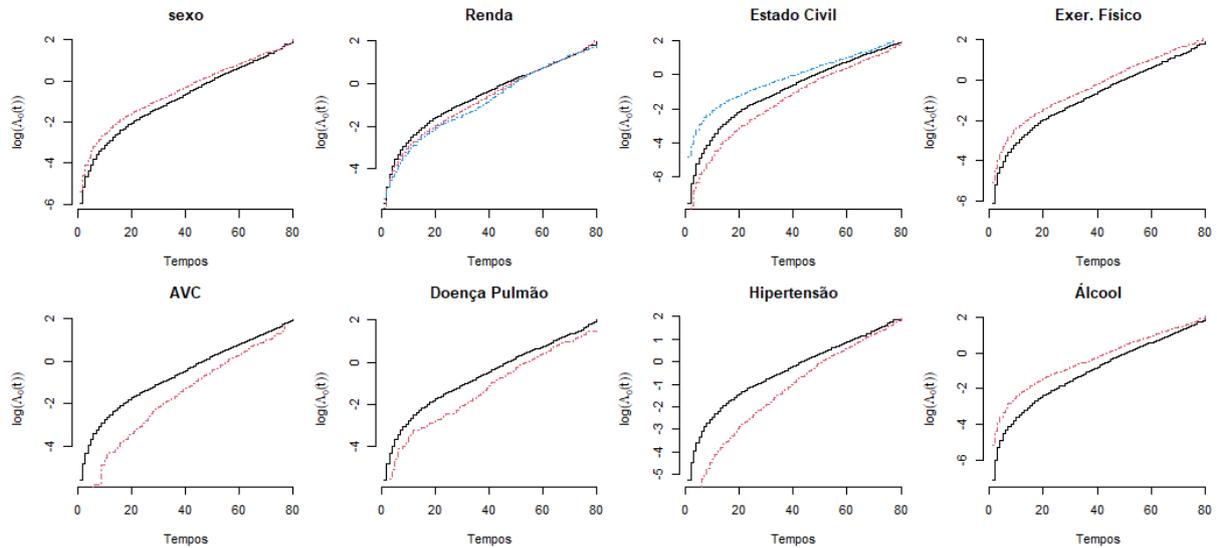
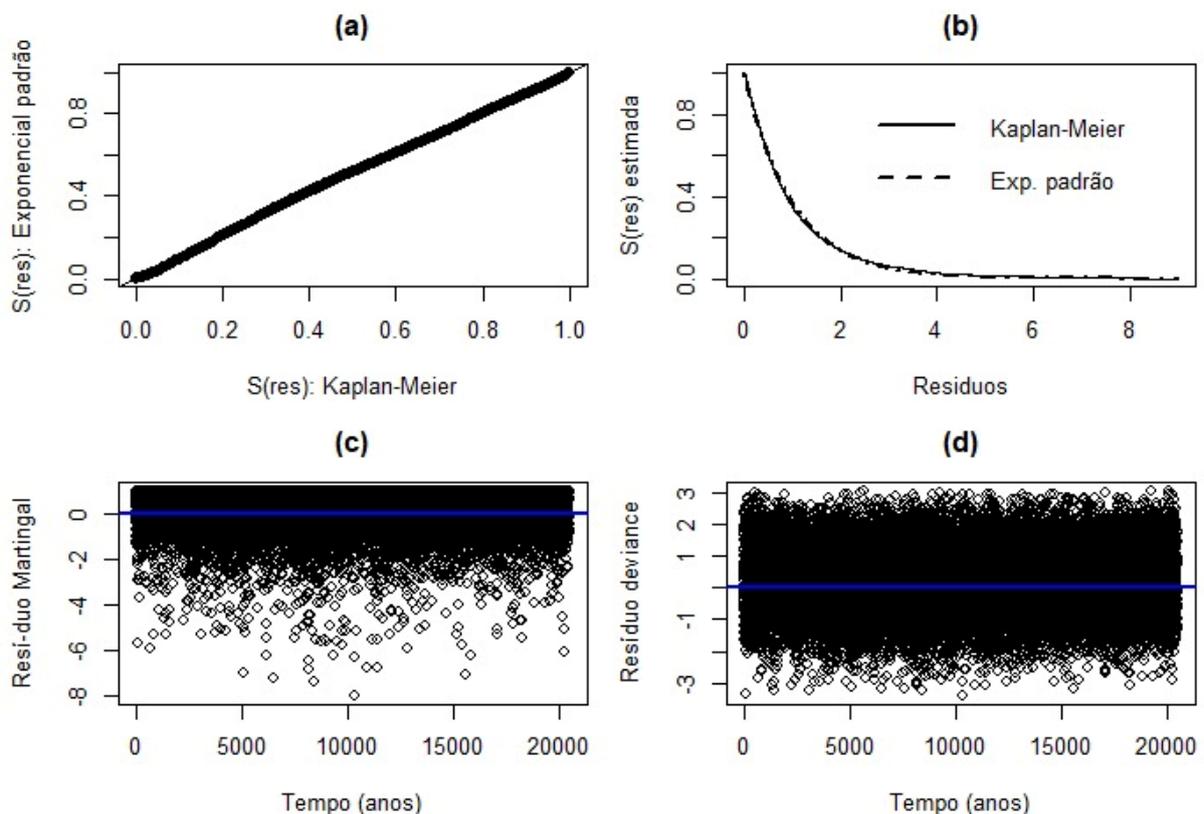


Figura 5: *

Fonte: Elaboração própria, 2023.

Figura 6: Resíduos de Cox-Snell (a e b), martingal (c) e deviance (d) do modelo de modelo de Cox-Gompertz para dados de tabagismo.



Fonte: Elaboração própria, 2023.

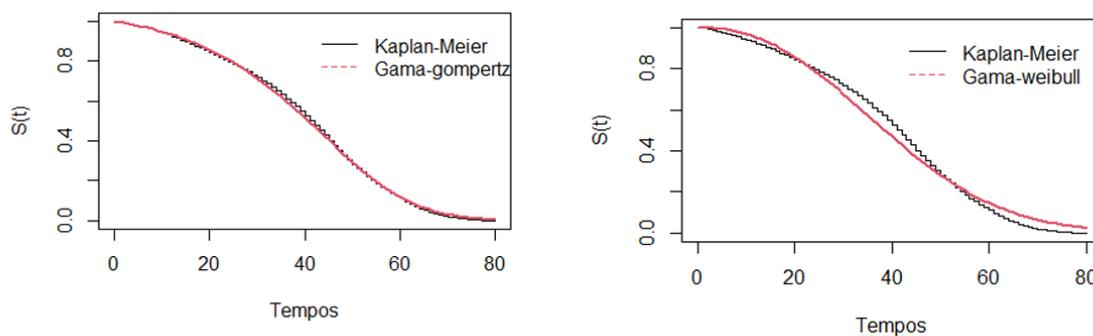
indivíduos em estudo. No entanto, é importante reconhecer que os elementos que afetam o consumo de tabaco e sua cessação podem abranger aspectos que não foram previamente consi-

derados. Essas observações escapam à capacidade de mensuração da metodologia convencional de análise de sobrevivência apresentada. Portanto, a seguir, apresentará-se a aplicação dos dados ao modelo de Fragilidade Gama, com o objetivo de quantificar os fatores não observados neste estudo.

Ajuste do Modelo de Fragilidade Gama-Gompertz

Para aplicação do modelo de fragilidade univariada, primeiramente, considerou-se o modelo sem capturar o efeito das covariáveis já apresentadas no modelo de Cox paramétrico, realizando uma comparação dos modelos Gama-Gompertz e Gama-Weibull, buscando verificar qual melhor se adequa aos tempos de sobrevivência. A Figura (7) abaixo, apresenta a comparação desses modelos, pode-se observar que o modelo de fragilidade univariado Gama-Gompertz apresenta uma melhor adequabilidade aos dados. No entanto, para efeitos de validação, realizou-se a análise por meio dos critérios de AIC e BIC e posteriormente, a análise dos resíduos.

Figura 7: Curva de Kaplan-Meier e teste de logrank para as variáveis de estudo



Fonte: Elaboração própria, 2023.

Na Tabela (5), encontram-se os resultados das estimativas e erro padrão para os parâmetros do modelo sem efeito das covariáveis. Assim como, os valores dos Critérios AIC e BIC, é possível observar que o modelo de fragilidade univariado Gama-Gompertz apresenta uma melhor adequabilidade aos dados, visto que os valores apresentados pelo AIC e BIC foram menores para este modelo.

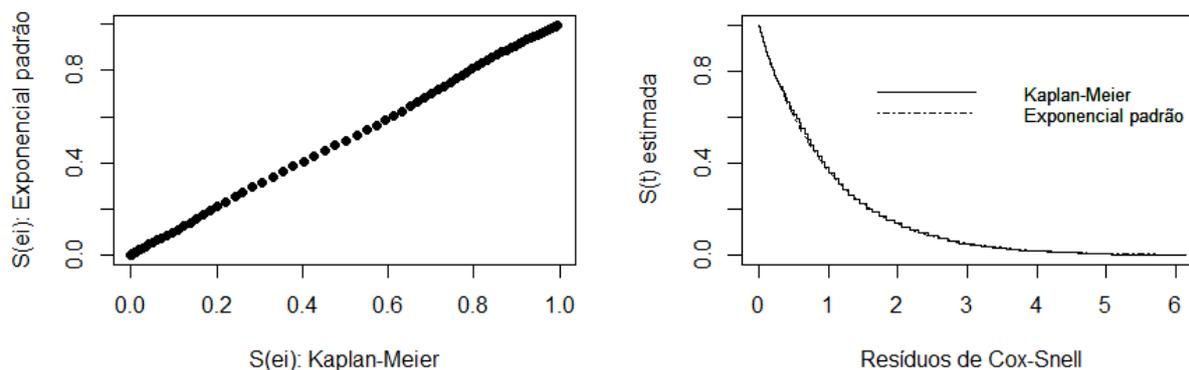
Tabela 5: Verificação de adequabilidade dos modelos de fragilidade univariada Gama-Gompertz e Gama-Weibull por meio do AIC e BIC

Parâmetros	Gompertz		Weibull	
	Estimativas	Erro Padrão	Estimativas	Erro Padrão
α	4,6498	0,2714	268,6591	87,6917
λ	0,0039	0,0001	45,1385	0,1666
γ	0,0635	0,0006	2,3087	0,016
AIC	124.251,10		125.969,70	
BIC	124.274,90		125.993,40	

Fonte: Elaboração própria, 2023.

Por meio da Figura (8), é possível observar o comportamento dos resíduos de Cox-Snell referentes ao modelo estimado Gama-Gompertz, permitindo uma análise da qualidade do ajuste. Ao examinar os gráficos, é evidente que os resíduos de Cox-Snell seguem uma distribuição exponencial padrão. Esse padrão indica que o modelo possui uma boa qualidade de ajuste global.

Figura 8: Ajuste dos resíduos para o modelo de fragilidade univariado Gama-Gompertz sem covariáveis



Fonte: Elaboração própria, 2023.

Obtendo um modelo de fragilidade adequado para os dados é possível quantificar a heterogeneidade presente não observada (fatores/variáveis não inclusas) nesse estudo. A variância estimada para a fragilidade e que é comum a todos os indivíduos foi obtida na Tabela 4, sendo que $\alpha = 4,6498$, o que transparece uma heterogeneidade significativa ($\neq 1$) entre os indivíduos mesmo que sem a presença de covariáveis e a variância estimada de $0,21 (1/\alpha)$.

Desta forma, o modelo de fragilidade pressupõe a presença de fatores não observados, comuns a todos os indivíduos, que podem influenciar na duração do tabagismo. Considerando que as análises anteriores indicaram que o modelo de fragilidade Gama-Gompertz é apropriado para os dados, procederá à aplicação desse modelo às variáveis iniciais do estudo, bem como à avaliação de sua adequação.

Por se tratar de uma extensão do modelo de Cox, o modelo de fragilidade não apresenta diferenças quanto à sua interpretação. Portanto, tem-se que, assim como no modelo paramétrico ajustado, alguns parâmetros apresentam estimativas negativas, porém houve algumas mudanças nas estimativas do modelo de fragilidade. Assim como no caso do modelo paramétrico, aqui as estimativas, $\hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_7, \hat{\beta}_8, \hat{\beta}_7, \hat{\beta}_8$, e $\hat{\beta}_9$, apresentaram-se negativas e indicam um menor risco de falha dos indivíduos. Já as demais apresentam maior risco de falha, e, portanto, cessam o tabagismo em mais tempo. Portanto, as interpretações desses parâmetros se darão conforme a seção *Ajuste do Modelo de Regressão Paramétrico de Cox-Gompertz* deste trabalho.

Após a aplicação do modelo de fragilidade com a inclusão das covariáveis de estudo, foi observado que a variância da variável de fragilidade $Var(z) = 1/\alpha$ foi de $0,17$. Isso representa um valor menor em comparação ao modelo sem covariáveis, indicando que as covariáveis tiveram impacto na redução da heterogeneidade dos dados. Contudo, essa redução parcial da variabilidade ainda evidencia a existência de fatores não observados que exercem influência sobre a duração do tabagismo entre os indivíduos estudados.

Após a aplicação do modelo de fragilidade é possível também calcular e verificar as estimativas das fragilidades para cada um dos indivíduos de estudo. Por meio da Tabela (7) é possível observar o comportamento da fragilidade para os 10 primeiros e 10 últimos indivíduos. Nota-se que as fragilidades individuais são diferentes para cada pessoa, ou seja, alguns indivíduos tendem a ser mais frágeis que outros, portanto, aqueles que apresentam fragilidades maior que 1 tendem a ser mais frágeis.

Além disso, obteve-se também que 12.112 (59,09%) dos indivíduos apresentaram estimativa de fragilidade superior a 1, já 586 (2,86%) possuem fragilidade igual a 1 e 7.800 (38,05%) são menores que 1. As visualizações gráficas dessas informações podem ser observadas na Figura (9) abaixo.

A Figura (10) apresenta os gráficos dos resíduos de Cox-Snell, resíduos martingale e resíduos deviance para o modelo de fragilidade Gama-Gompertz. Por meio dos gráficos (a) e (b), pode-se

Tabela 6: Estimativas e teste associado à fragilidade obtidos para o modelo paramétrico de fragilidade gama - Gompertz ajustado aos dados de tabagismo

Parâmetros	Estimativas	Erro Padrão	RTF	I.C (95%)	p-valor
α	5,812	0,773	-	-	< 0,001
λ	0,003	0,000	-	-	< 0,001
γ	0,073	0,001	-	-	< 0,001
$\hat{\beta}_1$ (Sexo [Feminino])	0,371	0,019	1,449	[1,40; 1,49]	< 0,001
$\hat{\beta}_2$ (Renda [1-3 SM])	-0,303	0,022	0,739	[0,702; 0,774]	< 0,001
$\hat{\beta}_3$ (Renda [> 3 SM])	-0,551	0,032	0,576	[0,516; 0,632]	< 0,001
$\hat{\beta}_4$ (Estado civil [Divorciado/Viúvo])	-0,578	0,027	0,561	[0,497; 0,624]	< 0,001
$\hat{\beta}_5$ (Estado Civil [Solteiro])	0,333	0,021	1,395	[1,369; 1,420]	< 0,001
$\hat{\beta}_6$ (Exercício Físico [Sim])	0,392	0,020	1,480	[1,460; 1,498]	< 0,001
$\hat{\beta}_7$ (AVC [Sim])	-0,411	0,052	0,663	[0,614; 0,710]	< 0,001
$\hat{\beta}_8$ (Doença Pulmão [Sim])	-0,443	0,062	0,642	[0,499; 0,784]	< 0,001
$\hat{\beta}_9$ (Hipertensão [Sim])	-0,557	0,023	0,573	[0,516; 0,628]	< 0,001
$\hat{\beta}_{10}$ (Bebida Alcoólica [Sim])	0,383	0,019	1,467	[1,446; 1,487]	< 0,001

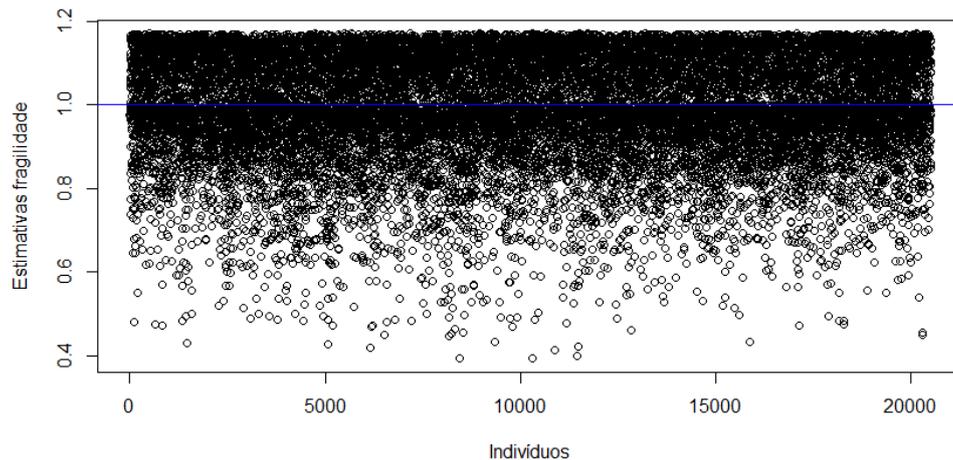
Fonte: Elaboração própria, 2023.

Tabela 7: Estimativas de algumas das fragilidades individuais

Indivíduos	1	2	3	4	5	6	7	8	9
Fragilidade	1,8006	2,9639	2,9639	0,7986	2,9639	0,2372	0,8527	0,8805	1
Indivíduos	20.489	20.490	20.491	20.492	20.493	20.494	20.495	20.496	20.497
Fragilidade	0,5608	0,4143	1,6973	1,5115	0,9238	2,2366	2,2366	1,3302	1,5699

Fonte: Elaboração própria, 2023.

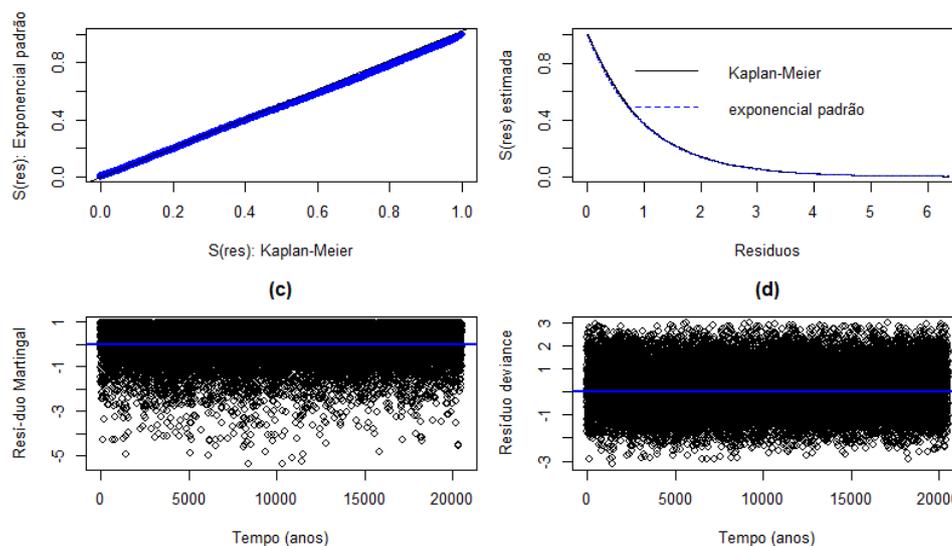
Figura 9: Visualização gráfica das estimativas de fragilidades individuais para os dados de fumantes de estudo.



Fonte: Elaboração própria, 2023.

observar que os resíduos de Cox-Snell seguem uma distribuição exponencial padrão, indicando uma boa qualidade de ajuste global do modelo. Já nos gráficos (c) e (d) nota-se que não há presença de pontos que possam ser considerados atípicos (outliers).

Figura 10: Ajuste dos resíduos para o modelo de fragilidade Gompertz com covariáveis



Fonte: Elaboração própria, 2023.

Conclusão

Este estudo apresentou uma aplicação da técnica de análise de sobrevivência, utilizando o modelo de regressão paramétrico Cox-Gompertz e o modelo de fragilidade univariada Gama-Gompertz, com o propósito de analisar o tempo de consumo de tabaco (por meio do ato de fumar) entre os indivíduos da pesquisa amostral da Pesquisa Nacional de Saúde de 2019 no Brasil. Por meio da aplicação do modelo paramétrico Cox-Gompertz, identificou-se que mulheres, indivíduos que praticam exercício físico regularmente e aqueles que são solteiros tendem a cessar o tabagismo em um período de tempo mais curto em comparação com indivíduos que possuem características opostas.

Além disso, a análise da variância da fragilidade permitiu quantificar a heterogeneidade presente nos indivíduos, a qual não pode ser explicada pelas variáveis consideradas no estudo. Isso sugere que fatores externos não abordados podem influenciar o tempo de consumo de cigarro. Portanto, o modelo de fragilidade demonstrou ser eficaz para alcançar os objetivos de identificar a existência de heterogeneidade nos dados.

Para trabalhos futuros, recomenda-se a inclusão de outras variáveis, como escolaridade, idade e outros fatores sociodemográficos dos indivíduos, a fim de ampliar a compreensão dos determinantes do tempo de consumo de tabaco. Além disso, a aplicação da modelagem de fragilidade compartilhada, que permite a análise em subgrupos específicos, pode enriquecer ainda mais a investigação nesse campo.

Referências

- AKAIKE, Hirotugu. *A New Look at the Statistical Model Identification*. *IEEE Transactions on Automatic Control*, v. 19, n. 6, p. 716-723, 1974.
- COLOSIMO, E. A e GIOLO, S. R. *Análise de Sobrevivência Aplicada*. Editora Blucher, 2006.
- COX, D. R. *Regression Models and Life Tables (with discussion)*. *J. Roy. Statist. Soc., B*, v. 34, p. 187-202, 1972.
- COX, David R.; SNELL, E. Joyce. *A General Definition of Residuals*. *Journal of the Royal*

Sigmae, Alfenas, v.12, n.3, p.140-157. 2023.

67ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria (RBras) e 20º Simpósio de Estatística Aplicada à Experimentação Agronômica (SEAGRO)

Statistical Society: Series B (Methodological), v. 30, n. 2, p. 248-265, 1968.

ELBERS, C. e RIDDER, G. *True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model*. *The Review of Economic Studies*, v. 49, n. 3, p. 403-409, 1982.

FIOCRUZ. Projeto de Pesquisa e Desenvolvimento Tecnológico - Pesquisa Nacional de Saúde (PNS), Rio de Janeiro. Acesso em 20 de janeiro de 2023, disponível em: <https://www.pns.icict.fiocruz.br/>.

GOMPERTZ, B. *On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. In a Letter to Francis Baily, Esq. FRS &c. By Benjamin Gompertz, Esq. FRS*. In: *Abstracts of the Papers Printed in the Philosophical Transactions of the Royal Society of London*. London: The Royal Society, 1833. p. 252-253.

HOUGAARD, Philip. *Frailty Models for Survival Data*. *Lifetime Data Analysis*, v. 1, p. 255-273, 1995.

SCHWARZ, Albert S. *The Partition Function of Degenerate Quadratic Functional and Ray-Singer Invariants*. *Letters in Mathematical Physics*, v. 2, n. 3, p. 247-252, 1978.

TOMAZELLA, Vera Lucia Damasceno. *Modelagem de Dados de Eventos Recorrentes via Processo de Poisson com Termo de Fragilidade*. 2003. Tese (Doutorado em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2003.

VIGITEL (Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito). *Estimativas sobre frequência e distribuição sociodemográfica de fatores de risco e proteção para doenças crônicas nas capitais dos 26 estados brasileiros e no distrito federal em 2021*. Ministério da Saúde, 2022. Acesso em 20 de Fevereiro de 2023, disponível em: [link](#).

VOGT, C. M, et al. *Smoking Cessation in Brazil: A Survival Analysis Based on Consumer Profile*. *Ciência & Saúde Coletiva*, p. 12, 2021. Acesso em 10 de Setembro de 2023, disponível em [link](#).

WEIBULL, W. *A Statistical Theory of Strength of Materials*. IVB-Handl., 1939.

WEIBULL, W. *A Statistical Distribution Function of Wide Applicability*. *Journal of Applied Mechanics*, 1951.

WEIBULL, W et al. *A Statistical Representation of Fatigue Failures in Solids*. Elander, 1949.

WIENKE, A. *Frailty Models in Survival Analysis*. Boca Raton, FL, EUA: Chapman & Hall/CRC, 2011.

WHO (World Health Organization). *Tobacco and Its Environmental Impact: An Overview*, 2017. Acesso em 20 de Janeiro de 2023, disponível em OMS: <https://apps.who.int/iris/handle/10665/255574>.

WHO (World Health Organization). *Global Report on Trends in Prevalence of Tobacco Use 2000–2025*. 2021. Acesso em 20 de Fevereiro de 2023, disponível em <https://www.who.int/publications/i/item/9789240039322>.

Sigmae, Alfenas, v.12, n.3, p.140-157. 2023.

67ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria (RBras) e 20º Simpósio de Estatística Aplicada à Experimentação Agrônômica (SEAGRO)