



UNIVERSIDADE FEDERAL DA PARAÍBA (UFPB)
CENTRO DE CIÊNCIAS SOCIAIS APLICADAS (CCSA)
DEPARTAMENTO DE FINANÇAS E CONTABILIDADE (DFC)
CURSO DE GRADUAÇÃO EM CIÊNCIAS ATUARIAIS (CCA)

TASSIANO VIRGULINO DE SOUZA

**PRECIFICAÇÃO DE RENDAS PARA CASAIS: UMA ABORDAGEM VIA MODELOS
DE FRAGILIDADE**

JOÃO PESSOA

2026

TASSIANO VIRGULINO DE SOUZA

**PRECIFICAÇÃO DE RENDAS PARA CASAIS: UMA ABORDAGEM VIA MODELOS
DE FRAGILIDADE**

Trabalho de Conclusão de Curso apresentado
como requisito parcial à obtenção do título de
Bacharel em Ciências Atuariais pela
Universidade Federal da Paraíba.

Área de concentração: Matemática Actuarial.

Orientador: Prof. Bel. Hugo Vieira Sá Ferreira
Gomes.

Coorientador: Prof. Dr. Luiz Carlos Santos
Júnior.

Coorientador: Prof. Dr. Rumenick Pereira da
Silva.

JOÃO PESSOA
2026

Catálogo na publicação
Seção de Catalogação e Classificação

S729p Souza, Tassiano Virgulino de.

Precificação de rendas para casais: uma abordagem
via modelos de fragilidade / Tassiano Virgulino de
Souza. - João Pessoa, 2026.

76 f. : il.

Orientação: Hugo Vieira Sá Ferreira Gomes.

Coorientação: Luiz Carlos Santos Júnior, Rumenick
Pereira da Silva.

TCC (Graduação) - UFPB/CCSA.

1. Cálculo atuarial. 2. Múltiplas vidas. 3. Modelo
de fragilidade compartilhada. 4. Anuidades. 5.
Dependência. I. Gomes, Hugo Vieira Sá Ferreira. II.
Santos Júnior, Luiz Carlos. III. Silva, Rumenick
Pereira da. IV. Título.

UFPB/CCSA

CDU 368(043)

TASSIANO VIRGULINO DE SOUZA

**PRECIFICAÇÃO DE RENDAS PARA CASAIS: UMA ABORDAGEM VIA MODELOS
DE FRAGILIDADE**

Trabalho de Conclusão de Curso como
requisito parcial à obtenção do título de
Bacharel em Ciências Atuariais pela
Universidade Federal da Paraíba.

BANCA EXAMINADORA

Documento assinado digitalmente
 **HUGO VIEIRA SA FERREIRA GOMES**
Data: 07/04/2026 07:50:11-0300
Verifique em <https://validar.iti.gov.br>

Prof. Bel. Hugo Vieira Sá Ferreira Gomes
Orientador
Universidade Federal da Paraíba (UFPB)

Documento assinado digitalmente
 **YURI MARTI SANTANA SANTOS**
Data: 07/04/2026 16:23:21-0300
Verifique em <https://validar.iti.gov.br>

Prof. Me. Yuri Martí Santana Santos
Membro Avaliador
Universidade Federal da Paraíba (UFPB)

Documento assinado digitalmente
 **VERA LUCIA DAMASCENO TOMAZELLA**
Data: 07/04/2026 14:47:15-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dra. Vera Lucia Damasceno Tomazella
Membro Avaliador
Universidade Federal de São Carlos (UFSCar)

AGRADECIMENTOS

Agradeço primeiramente a Deus por esta realização, por iluminar os meus caminhos, me cobrir de bênçãos, me dar sabedoria e por me fazer forte em momentos difíceis em que só o Senhor conhece a realidade.

À minha família, meus pais, Edimilson e Rozelândia, as pessoas mais importantes de minha vida, agradeço não somente pelos ensinamentos, mas por todo o amor por mim, pelo carinho, pelos esforços e pelos incentivos para que eu pudesse trilhar pelo caminho dos estudos. Às minhas irmãs: Tâmirys, por sempre estar ao meu lado em todos os momentos, agradáveis e difíceis, sempre me apoiando e me aconselhando; e à pequena Tauany, pelo brilho e alegria que trouxe às nossas vidas. Vocês são os pilares da minha vida.

Aos meus tios, em especial Jeremias, Rosivania, Rosimere, por todo o apoio durante várias etapas do processo para que eu conseguisse chegar até aqui, assim como aos demais familiares que sempre estiveram ao meu lado e me apoiaram nessa jornada.

A todos os amigos feitos ao longo da minha graduação, em especial a Carlos e Cléo, pelos ensinamentos e pelas contribuições para este trabalho. A Janailson e Stênio, pela parceria durante os desafios enfrentados na graduação, à Adria Cristina, pelo companheirismo, apoio e conselhos, e à Clara, que, mesmo em pouco tempo, foi especial no encerramento dessa jornada, pelo carinho, incentivos e conselhos de vida que contribuíram para a minha evolução acadêmica e profissional.

Ao meu orientador, professor Hugo, pela paciência e empenho na revisão deste trabalho, assim como pelo apoio e pela confiança nesta etapa. Aos meus coorientadores: professor Luiz, por todo o apoio durante as disciplinas do curso e nos projetos, assim como pelas contribuições para este trabalho, pelos conselhos e por sempre enxergar potencial em seus alunos; e ao professor Rumenick, pelo apoio e pelas contribuições para a conclusão deste trabalho.

A todos os professores que tive ao longo da minha graduação e que contribuíram de alguma forma para este momento, em especial aos professores Hugo e Luiz citados, e à professora Vera pelo empenho, dedicação em ensinar e pelo carinho.

Por fim, agradeço a todos que, direta e indiretamente, fizeram parte da minha jornada e que, mesmo não mencionados, foram alicerce para esta vitória. Desde já, a minha imensa gratidão.

RESUMO

Nos cálculos atuariais para mais de um indivíduo, é comum assumir a independência entre os tempos residuais de vida. Contudo, para determinados grupos de indivíduos, como casais, gêmeos e famílias, essa suposição pode não ser válida, visto que os tempos de vida podem estar interligados por diversas causas; dentre elas, destacam-se os efeitos de desastres comuns, a dependência por estilo de vida e a síndrome do coração partido, observada especialmente entre casais. Consequentemente, essa realidade, ao ser ignorada, pode afetar as premissas do cálculo atuarial, uma vez que se supõe independência. Para verificar essa questão empiricamente, este trabalho implementa uma abordagem de dependência por meio de um modelo de fragilidade compartilhada com distribuição gama para verificar se existe dependência entre os tempos residuais de vida de casais e explorar se essa relação gera impacto na precificação de anuidades para múltiplas vidas. Para aplicação do modelo, utilizou-se uma base de dados secundária coletada a partir de registros históricos do site da Familysearch. Os resultados revelaram existência significativa de dependência entre os residuais de vida e a ausência de diferença estatisticamente significativa entre a sobrevivência por sexo. Ao considerar a fragilidade na sobrevivência, observa-se que a sobrevivência com a dependência tende a ser superior em relação à prática comum de independência, e que essa diferença cresce à medida que as idades aumentam. Isso também corresponde ao efeito gerado nas rendas, o que indica que o valor da anuidade independente é subvalorizado. Assim, conclui-se que há evidências estatísticas de dependência entre os tempos residuais de vida de casais e que essa dependência gera impacto no preço das rendas, que passam a ser superiores em comparação com a prática de mercado.

Palavras-chave: Cálculo Atuarial. Múltiplas Vidas. Modelo de Fragilidade Compartilhada. Anuidades. Dependência.

ABSTRACT

In actuarial calculations for more than one individual, it is common to assume independence between residual lifetimes. However, for certain groups of individuals, such as couples, twins, and families, this assumption may not be valid, since lifetimes may be interconnected due to various causes; among them, the effects of common disasters, dependence due to lifestyle, and the broken heart syndrome, observed especially among couples. Consequently, when this reality is ignored, it may affect the assumptions of actuarial calculations, as independence is presumed. To verify this issue empirically, this study implements a dependence approach through a shared frailty model with a gamma distribution to assess whether there is dependence between the residual lifetimes of couples and to explore whether this relationship impacts the pricing of multi-life annuities. For the model application, a secondary dataset collected from historical records on the FamilySearch website was used. The results revealed a significant existence of dependence between residual lifetimes and the absence of a statistically significant difference in survival by sex. When considering frailty in survival, it is observed that survival under dependence tends to be higher compared to the common independence practice, and that this difference increases as ages rise. This also corresponds to the effect generated in annuities, indicating that the independent annuity value is undervalued. Thus, it is concluded that there is statistical evidence of dependence between the residual lifetimes of couples and that this dependence impacts annuity pricing, which becomes higher compared to standard market practice.**Keywords:** Actuarial Calculation. Multiple Lives. Shared Frailty Model. Annuities. Dependency.

LISTA DE FIGURAS

Figura 1 – Resultado da distribuição de óbitos por idade.	48
Figura 2 – Resultados das curvas de sobrevivência estimadas pelo método de Kaplan-Meier, estratificadas por sexo.	49
Figura 3 – Resultado da estimação do gráfico do perfil de verossimilhança de θ	52
Figura 4 – Resultado da distribuição empírica de θ	53
Figura 5 – Curvas de sobrevivência do modelo de Cox COM e SEM fragilidade e o estimador de Kaplan-Meier.	54

LISTA DE QUADROS

Quadro 1 – Descrição das variáveis utilizadas.....	27
Quadro 2 – Tipos de censura para dados de sobrevivência.....	30

LISTA DE TABELAS

Tabela 1 - Dados dos três primeiros registros de casais contidos na base.....	28
Tabela 2 – Resultados das principais estatísticas descritivas	47
Tabela 3 – Resultados das probabilidades de sobrevivência idades específicas.	50
Tabela 4 - Resultados estimados para o modelo de COX sem a fragilidade.....	50
Tabela 5 - Resultados do modelo de Cox com fragilidade compartilhada (distribuição Gama).	51
Tabela 6 - Resultados das anuidades, calculadas sob a hipótese de independência (modelo de Cox x KM).....	55
Tabela 7 – Resultados das anuidades, calculadas sob as hipóteses de independência e dependência.	56

SUMÁRIO

1 INTRODUÇÃO	10
1.1 CONTEXTUALIZAÇÃO E PROBLEMA	10
1.2 OBJETIVOS	12
1.2.1 Objetivo geral	12
1.2.2 Objetivos específicos	12
1.3 JUSTIFICATIVA	13
1.4 ESTRUTURA DO TRABALHO	14
2 REFERENCIAL TEÓRICO E EMPÍRICO	15
2.1 TEMPO RESIDUAL DE VIDA E DESCONTO FINANCEIRO	15
2.2 PRÊMIOS E RENDAS ATUARIAIS	18
2.3 ABORDAGEM POR DEPENDÊNCIA	21
2.4 ESTUDOS CORRELATOS	22
3 METODOLOGIA	26
3.1 TIPO DE PESQUISA	26
3.2 COLETA DE DADOS	27
3.3 NOÇÕES BÁSICAS DE ANÁLISE DE SOBREVIVÊNCIA	29
3.4 MODELOS DE FRAGILIDADES	32
3.4.1 Modelo univariado de fragilidade	34
3.4.2 Modelo compartilhado de fragilidade	39
3.5 VALIDAÇÃO DOS RESULTADOS DO MODELO	44
3.6 COMPARAÇÕES ENTRE MODELOS DE DEPENDÊNCIA E INDEPENDÊNCIA ENTRE OS TEMPOS RESIDUAIS DE VIDA	45
4. RESULTADOS E DISCUSSÕES	47
4.1 ANÁLISE DESCRITIVA DOS DADOS	47
4.2 ANÁLISE NÃO PARAMÉTRICA DA SOBREVIVÊNCIA	49
4.3 ANÁLISE DO MODELO DE FRAGILIDADE	50
4.4 ANÁLISE COMPARATIVA ENTRE A SOBREVIVÊNCIA E OS VALORES DAS ANUIDADES	54
5 CONSIDERAÇÕES FINAIS	59
REFERÊNCIAS	61
APÊNDICES	67
APÊNDICE A – Código	68

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO E PROBLEMA

A vida cotidiana envolve diferentes fenômenos de risco que podem ser caracterizados pela possibilidade de ocorrência de um evento aleatório, em que exista incerteza sobre qual resultado ocorrerá (Harrington; Niehaus, 2004; Rodrigues, 2008).

À luz dos fenômenos de risco predominantes na sociedade contemporânea, Fromm (2022) destaca como a Indústria da Proteção – a tríade seguro, segurança e seguridade – constitui um sistema complexo voltado para a mitigação de riscos e para garantir o "bem-estar" de indivíduos, famílias e organizações. Nessa tríade, o seguro constitui um instrumento de cobertura de riscos ao oferecer compensação financeira por perdas e danos.

Assim, a racionalidade do seguro envolve precificar riscos de eventos incertos para oferecer garantias mediante o acontecimento desses eventos inesperados (Fromm, 2022). Portanto, a precificação expressa diretamente a funcionalidade do seguro, já que deve refletir o risco realmente assumido em um evento para que as seguradoras consigam obter os seus objetivos e, assim, possam arcar com as garantias estipuladas nos contratos (Rodrigues, 2008). Visto que quaisquer distorções no processo de precificação podem gerar impacto sobre a solvência das seguradoras e comprometer as indenizações devidas aos segurados (Rodrigues, 2008).

No que se refere à precificação e a sua forma, os seguros se dividem em dois ramos, seguros de danos e de pessoas (Rodrigues, 2008; Fromm, 2022). O seguro de danos é uma modalidade que visa proteger o patrimônio do segurado contra prejuízos financeiros causados por eventos diversos, como incêndios, roubos, acidentes e desastres naturais (Susep, 2025). Os do ramo de pessoas são contratos que oferecem proteção financeira para o indivíduo e para a sua família em caso de eventos referentes à sobrevivência, morte, invalidez ou morbidez (Susep, 2025; Rodrigues, 2008). Por cobrirem eventos biométricos, os seguros de vida consideram, na sua precificação, os valores das probabilidades de ocorrências desses eventos, mensurados a partir das estimativas dos tempos residuais de vida dos indivíduos – também conhecidos como tempos restantes de vida, tempos futuros de vida ou tempos até a ocorrência da morte – usualmente considerados independentes (Alai; Landmann; Sherris, 2013).

Em contrapartida, a suposição de independência nem sempre é considerada válida, especialmente nos cálculos para múltiplas vidas, já que há casos em que alguns indivíduos, por

compartilharem riscos semelhantes devido a fatores genéticos, ambientais ou estilos de vida parecidos, podem apresentar tempos de sobrevivência correlacionados (Alai; Landsman; Sherris, 2013). Exemplos disso podem ser observados em casais, gêmeos ou grupos de familiares (Colosimo; Giolo, 2006; Frees; Carriere; Valdez, 1996).

Neste sentido, alguns estudos têm apresentado evidências de que, em determinados casos, os tempos residuais de vida podem ter dependências, especialmente entre casais (Giovanni; Pirra; Viviano, 2025; Gomes, 2024; Henshaw; Constantinescu; Pamen, 2020; Lee; Cha, 2018; Spreeuw; Owadally, 2013; Gourieroux; Lu, 2013; Luciano *et al.*, 2008). Os estudos sobre dependência entre casais indicam três principais causas para a existência de dependência, que são desastre comum, estilo de vida e síndrome do coração partido. A primeira causa, desastre comum, refere-se aos eventos que podem impactar ambos os integrantes de um grupo; o estilo de vida reflete o impacto comportamental dos hábitos entre indivíduos dos casais, sejam eles de saúde, de alimentação ou de diferentes fatores, como, por exemplo, socioeconômicos; e, a terceira, a síndrome do coração partido, que reflete as mudanças nas probabilidades de morte de um integrante do casal após a morte de um cônjuge (Gomes, 2024; Gourieroux; Lu, 2013).

Lee e Cha (2018) indicaram vários estudos que buscaram desconsiderar a independência presente nos tempos residuais de vida para grupos, entre eles estão: a) Frees, Carriere e Valdez (1996), Youn e Shemyakin (1999), Youn e Shemyakin (2001), Carriere (2000), Luciano, e Spreeuw e Vigna (2008), que implementaram modelos de cópulas; b) Marshal e Olkin (1967), Panjer (1994), e Denuit, Frostig e Levikson (2006), que destacaram o modelo de *common shock*; c) por fim, os estudos de Norberg (1989), Wolthuis (2003), Ji, Hardy e Lee (2011), e Spreeuw e Wang (2008), que abordaram os modelos multiestados.

A dependência do risco de morte é geralmente levada em consideração na literatura atuarial pela introdução de modelos de cópulas especiais com densidade contínua (Gourieroux; Lu, 2013). Estes se apresentam como ferramentas amplamente utilizadas, dadas as suas vantagens de flexibilização e a sua capacidade de capturar estruturas complexas. Ademais, os modelos realizam análises univariadas e as conectam para formar distribuições multivariadas. Assim, conseguem modelar diferentes estruturas de dependência entre variáveis (Joe, 1997; Nelsen, 2006).

Contudo, embora as cópulas forneçam ferramentas poderosas e flexíveis, os modelos podem, em alguns casos, apresentar complexidades e desafios, devido aos seus custos computacionais, à necessidade de grandes volumes de dados observacionais e às dificuldades da interpretação dos resultados (Frees; Carriere; Valdez, 1996).

Nesse cenário, e diante dos desafios para se modelar, ganha destaque a relevância dos modelos estatísticos capazes de lidar com a heterogeneidade não observada e com a estrutura de dependência entre indivíduos ou grupos (Onchere, 2022). A literatura recente aponta que abordagens baseadas em riscos proporcionais, modelos de sobrevivência de longa duração e distribuições flexíveis têm avançado significativamente, oferecendo alternativas robustas para representar fenômenos complexos em contextos biomédicos, atuariais e sociais. Essas contribuições reforçam a importância de metodologias que ampliem a capacidade explicativa dos modelos, ao mesmo tempo em que preservam a consistência inferencial e a aplicabilidade prática.

Wienke (2010) destaca os modelos de fragilidade que podem capturar heterogeneidade e riscos comuns de uma população via fatores latentes (ex.: genéticos e socioeconômicos), o que pode evitar vieses de estimação que ignoram as heterogeneidades, que, nesse contexto, referem-se a fatores aleatórios não medidos e não observados que afetam o risco. Logo, a utilização de modelos de fragilidade pode contribuir para cálculos atuariais mais precisos ao verificar se há impacto e, se houver, qual a magnitude do impacto da dependência entre os tempos futuros de vida de grupos.

Diante do contexto apresentado, sobre a necessidade de aprimoramento nos cálculos atuariais para múltiplas vidas, tem-se como pergunta de pesquisa: **qual é o impacto da dependência entre os tempos residuais de vida de cônjuges, incorporada por meio de fatores não observados, sobre a precificação de rendas para casais?**

1.2 OBJETIVOS

1.2.1 Objetivo geral

Analisar o impacto da dependência entre os tempos residuais de vida de cônjuges, incorporada por meio de fatores não observados, sobre a precificação de rendas para casais.

1.2.2 Objetivos específicos

- Estimar a dependência, por meio dos modelos de fragilidade, entre os tempos de residuais de vida de cônjuges;
- Calcular as anuidades para múltiplas vidas a partir das estimativas da dependência entre os tempos residuais de vida de cônjuges;

- Comparar os valores das rendas calculadas a partir das premissas de dependência e independência entre os tempos de vida citados.

1.3 JUSTIFICATIVA

Os cálculos de seguros e rendas para dois ou mais indivíduos, na maioria dos casos, são realizados de forma a considerar que os tempos residuais de vida entre indivíduos não possuem ligação ou relação (Alai, Landsman e Sherris (2013). Contudo, essa aplicação pode não refletir diretamente a realidade e precisão das estimativas sobre a mortalidade conjunta, visto que ao se considerar a dependência, os valores de prêmios de seguros e rendas podem ser super ou subestimados (Gomes, 2024; Gourieroux; Lu, 2013).

A suposição de independência, nas associações sobre tempos de vida, é frequentemente adotada na prática atuarial (Onchere, 2022). Os modelos mais utilizados e as principais premissas, as tábuas de vida, são construídos sem medir o efeito de associação. A precificação de forma mais simples pode apresentar maior agilidade e uma complexidade reduzida nos cálculos atuariais, uma vez que não necessita de modelos estatísticos complexos ou demora/obstáculos na coleta de dados bem estruturados que reflitam a realidade sobre grupos populacionais específicos.

No entanto, segundo Alai, Landsman e Sherris (2013) as aplicações dessa forma aumentam o risco sistematicamente devido à dependência entre as vidas. Visto que, esse aumento de risco muda as condições de convergência da lei dos grandes números, que é importante na gestão de riscos de carteiras de seguros. A dependência entre os eventos gera resultados correlacionados, levando a uma variabilidade maior do que a esperada. Assim, a média dos resultados pode demorar mais para convergir para o valor esperado, o que impacta a teoria da lei dos grandes números e pode prejudicar a confiabilidade das estratégias de gerenciamento de risco em seguros.

O presente trabalho busca realizar a análise de dependência mediante o uso de modelos de fragilidade, que são modelos de sobrevivência que incorporam variabilidade não capturada entre indivíduos ou grupos e que capturam a dependência não observada entre tempos de sobrevivência, o que evita, assim, o viés de estimativas que ignoram heterogeneidade latente (Colosimo; Giolo, 2006).

O estudo visa preencher lacunas e reforçar debates sobre a consideração de dependência, dado que, no campo atuarial, as aplicações dos modelos de fragilidades podem ser bastante úteis para cálculos de múltiplas vidas, pois possibilitam entender o efeito da heterogeneidade não

observada e permite a análise da dependência entre os tempos futuros de vida de indivíduos integrantes de um mesmo grupo, como casais, famílias ou populações específicas e, assim melhorar a precisão nas projeções de mortalidade conjunta (Carannante *et al.*, 2023; Carannante *et al.*, 2024).

1.4 ESTRUTURA DO TRABALHO

Para alcançar os objetivos propostos e responder à pergunta de pesquisa, o trabalho foi organizado da seguinte forma: O capítulo 1 descreve o contexto introdutório em torno do tema e os fatores que levam ao questionamento de pesquisa, com os objetivos a serem cumpridos e a justificativa que fundamenta a realização do estudo; o capítulo 2 apresenta uma revisão literária dos principais fundamentos atuariais para os cálculos de rendas, além dos principais estudos sobre dependência entre vidas e que apresentaram resultados relevantes sobre aplicações em cálculos atuariais; o capítulo 3 detalha a metodologia e os dados utilizados para conduzir o estudo; no capítulo 4 são apresentados os principais resultados da pesquisa, além das discussões; e no último, capítulo 5, apresentam-se as considerações finais.

2 REFERENCIAL TEÓRICO E EMPÍRICO

2.1 TEMPO RESIDUAL DE VIDA E DESCONTO FINANCEIRO

Em atuária, a abordagem do ramo vida, que concentra os campos de seguros e previdência, está relacionada à cobertura de sobrevivência e falha de eventos biométricos que afetam os indivíduos. A abordagem predominante de estudo sobre as chances de sobreviver ou falhar a determinadas causas é comumente tratada pela análise de sobrevivência, a qual, conforme descrita por Colosimo e Giolo (2006), é um campo de estudo que envolve diferentes contextos e pode ser naturalmente aplicada à ocorrência de diversos eventos, tais quais os ligados à vida de pessoas ou até mesmo ao funcionamento de máquinas e objetos, e envolve uma compreensão completa dos fatores que influenciam a expectativa e chances de ocorrências desses eventos.

Em campos de estudos sobre análise de sobrevivência, como estatística, medicina, engenharia e demografia, as chances de os eventos ocorrerem são medidas a partir dos tempos decorridos até o momento de permanência ou falha (evento contrário à sobrevivência) do evento (Colosimo; Giolo, 2006).

Dessa forma, nos cálculos atuariais, a avaliação estatística das funções de sobrevivência para eventos biométricos (ligados à medida de vida) é analisada por meio das implicações de variáveis demográficas (Dickson; Hardy; Waters, 2013). Para isso, considera-se a incerteza quanto ao tempo de vida futuro dos indivíduos, definindo essa grandeza como uma variável aleatória (Pires *et al.*, 2021). Logo, por meio dessa abordagem, é possível entender a dinâmica da sobrevivência e a probabilidade de eventos relacionados à vida e à morte (Dickson; Hardy; Waters, 2013).

Matematicamente, de acordo com Dickson, Hardy e Waters (2013), o tempo de sobrevivência T pode ser considerado uma variável aleatória contínua não negativa (isto é, $t \geq 0$), com função densidade $f(t)$. No contexto da atuária, o tempo futuro de vida de um indivíduo de idade x pode ser representado pela variável aleatória contínua T_x . Assim, dada uma idade x para um indivíduo, é possível estimar a probabilidade de ele vir a falecer antes de decorridos t anos por meio da Função de Distribuição Acumulada $F_x(t)$, dada por:

$$F_x(t) = P(T_x \leq t) \quad (1)$$

Logo, (1) representa a probabilidade de morte entre as idades x e $x + t$.

A partir da função de distribuição acumulada $F_x(t)$, obtêm-se duas funções fundamentais na análise de sobrevivência: a função de sobrevivência e a função de densidade de probabilidade. De forma complementar a equação (1), pode-se obter a função de sobrevivência $S_x(t)$.

$$S_x(t) = 1 - F_x(t) = P[T_x \geq t] \quad (2)$$

A equação (2) representa a probabilidade de um indivíduo de idade x sobreviver por ao menos t anos, isto é, de estar vivo na idade $x + t$. Como característica, a função de sobrevivência é monótona decrescente, e possui as seguintes propriedades: $S_x(0) = 1$ e $S_x(\infty) = 0$.

Por meio da função de distribuição acumulada, é possível descrever também a função densidade de probabilidade, que corresponde à probabilidade de um indivíduo sofrer um evento em um instante específico de tempo. Ela é denotada por $f_x(t)$ e obtida por meio da derivação da função de distribuição acumulada $F_x(t)$, conforme (3).

$$f_x(t) = \frac{d}{dt} F_x(t) = -\frac{d}{dt} S_x(t) \quad (3)$$

A equação (4) representa a função de risco, que também pode ser vista com a denominação de taxa instantânea de mortalidade ou coeficiente instantâneo de mortalidade ou, ainda, como o risco de (x) morrer no instante de tempo infinitamente pequeno (Vilanova, 1969).

$$\mu_x(t) = \frac{f_x(t)}{S_x(t)} \quad (4)$$

Assim, é possível relacionar as funções exibidas em (1), (2), (3) e (4). Elas são relevantes para o ramo atuarial pois auxiliam na análise sobre o comportamento da mortalidade dos indivíduos a cada idade. Para compreender esse comportamento, faz-se necessário tomar como base o número de vivos em cada idade, que pode ser representada pela função l_x , definida, segundo Bowers *et al.* (1997), como o número de pessoas vivas na idade exata x .

De acordo com Cordeiro Filho (2009), a partir da função número de vivos, é possível entender os riscos ligados às ocorrências da mortalidade e estimar, com base nela, os prêmios de seguros e rendas. Por meio do comportamento da mortalidade, é possível estimar a medida

de risco para o evento, ou seja, a probabilidade, definida como a razão entre o número de mortes observadas e o número total de expostos ao risco de morte (Ross, 2010). Logo, com base no número de vivos a uma idade, têm-se as probabilidades de morte e sobrevivência a certa idade.

Neste sentido, apresentam-se as expressões (5) e (6).

$${}_tq_x = \frac{l_x - l_{x+t}}{l_x} \quad (5)$$

$${}_tp_x = \frac{l_{x+t}}{l_x} \quad (6)$$

A expressão (5) é a probabilidade de morte entre as idades x e $x + t$. Já ${}_tp_x$ representa a probabilidade de uma pessoa de idade x sobreviver até a idade $x + t$. Por serem eventos complementares, a soma dessas duas probabilidades resulta em 1, conforme (7) (Cordeiro Filho, 2009):

$${}_tp_x + {}_tq_x = 1 \quad (7)$$

Além das funções biométricas, Azevedo (2005), pontua a necessidade da atualização da importância segurada, que é realizada por meio do desconto financeiro v , definido conforme (8).

$$v = \frac{1}{1 + i} \quad (8)$$

Em que i é a taxa de juros.

O fator de desconto é incorporado no cálculo atuarial para que o valor das rendas ou indenizações pagas futuramente sejam atualizadas para o tempo 0, conforme descrito em Bowers *et al.* (1997). Assim, todos os pagamentos futuros são trazidos a valor presente, considerando a taxa de juros i , e representarão o valor no instante de contratação.

A seguir, são apresentadas as metodologias e formulações para os cálculos de rendas para os casos individuais e, posteriormente, as formulações e montagem para múltiplas vidas. As abordagens apresentadas fundamentam-se em Dickson, Hardy e Waters (2013), Cordeiro Filho (2009), Bowers *et al.* (1997) e Pires *et al.* (2021).

2.2 PRÊMIOS E RENDAS ATUARIAIS

Apresentadas as principais premissas para os cálculos atuariais, que são as funções biométricas e o fator de desconto, introduzem-se os conceitos fundamentais relacionados aos cálculos atuariais, junto a suas formulações e métodos de cálculo.

Os cálculos atuariais estão relacionados à cobertura por morte, sobrevivência e demais eventos biométricos. No caso de único pagamento, tem-se o seguro, enquanto para mais de um pagamento, tem-se a denominação de rendas (Bowers *et al.*, 1997). Para o presente estudo, focaremos apenas nas rendas.

A renda por sobrevivência é um produto atuarial baseado em uma anuidade, definida como uma sequência de pagamentos periódicos que cessam com a morte do segurado (Dickson; Hardy; Waters, 2013). Esse produto é adquirido mediante o pagamento de um prêmio, que corresponde ao valor que o segurado paga à seguradora para que essa assuma a cobertura por sobrevivência do segurado (Pires *et al.*, 2021). A forma de cálculo e dos pagamentos podem ser antecipados (realizados no início do período acordado), representados por $\ddot{a}_x$, ou postecipados (realizado no fim do período acordado), representados por a_x (Pires *et al.* 2021; Dickson; Hardy; Waters, 2013).

O cálculo do preço desse produto é realizado mediante o uso do fator de desconto atuarial, que é aplicado a cada um dos pagamentos futuros a ser realizado pela seguradora ao segurado. Este fator é dado pelo produto entre o desconto financeiro v e a probabilidade de um indivíduo de idade x sobreviver a t anos (${}_t p_x$) (Bowers *et al.*, 1997; Dickson; Hardy; Waters, 2013).

$${}_t E_x = v^t \cdot {}_t p_x \quad (9)$$

A soma dos descontos atuariais é denominada de anuidade, em que se tem a atualização de uma série de pagamentos. Quanto ao tipo, as anuidades podem ser classificadas de diferentes formas, dado a forma com que o pagamento da seguradora é realizado, sendo realizado de forma antecipada ou postecipada (Pires *et al.*, 2021).

Outro aspecto significativo é que as anuidades podem ter os pagamentos imediatos (no início do período de contratação) ou diferidas (em que o pagamento se inicia após decorridos m anos, havendo assim um período de carência para o início do recebimento, para o caso diferido, estas podem representadas por ${}_m | \ddot{a}_x$. Quanto a duração, elas podem ser vitalícias

(cessamento com a morte do segurado) ou temporárias em que se define uma temporariedade de n anos, em que, no caso de sobrevivência, o segurado só recebe até no máximo n anos, e podem ser representadas por $\ddot{a}_{x:\overline{n}|}$ ou, de forma geral, apresentar os dois casos, diferida e temporária ${}_m|\ddot{a}_{x:\overline{n}|}$. Para o caso de uma anuidade antecipada, diferida de m anos e temporária por n anos, o valor presente atuarial é dado por (10):

$${}_m|\ddot{a}_{x:\overline{n}|} = \sum_{t=m}^{m+n-1} {}_t p_x \cdot v^t \quad (10)$$

A atualização depende da variável t , presente no somatório, que representa o tempo de sobrevivência a partir da idade x . Observa-se que t se inicia em m no caso antecipado e, caso fosse postecipado, em $m + 1$, totalizando n pagamentos (Bowers *et al.*, 1997).

Para as coberturas de produtos ligados a eventos biométricos de duas ou mais vidas, podem-se considerar diferentes status, assumindo inicialmente independência entre os tempos de vida. Assim, para uma cobertura que envolve duas pessoas de idades x e y em status conjunto, podem ser definidos de duas diferentes formas: “vida conjunta” e “último sobrevivente”. No primeiro caso, a sobrevivência do status ocorre apenas enquanto ambos estão vivos; no segundo, ela ocorre quando pelo menos uma pessoa estiver viva.

Assim, para analisar o tempo até a ocorrência do evento coberto (ou até o término dos pagamentos), faz-se necessário analisar a distribuição conjunta relativa a T_x e T_y , que representam os tempos futuros de vidas de dois indivíduos de idades x e y .

Para o status de vida conjunta, a variável aleatória T analisada é definida como:

$$T_{xy} = \min\{T_x, T_y\} \quad (11)$$

Para o status de último sobrevivente, tem-se que:

$$T_{\overline{xy}} = \max\{T_x, T_y\} \quad (12)$$

Assumindo independência entre as vidas, a probabilidade de sobrevivência conjunta é dada pelo produto entre as probabilidades de pessoas de idades x e y sobreviverem a t anos:

$${}_t p_{xy} = {}_t p_x \cdot {}_t p_y \quad (13)$$

No caso de último sobrevivente o pagamento só é cessado após a morte de ambos os indivíduos, têm-se:

$${}_t p_{\overline{xy}} = {}_t p_x + {}_t p_y - {}_t p_{xy} \quad (14)$$

As equações (13) e (14) representam, respectivamente, a chance de x e y sobreviverem a t anos e a chance de x ou y sobreviverem a t anos.

Apresentadas as probabilidades de sobrevivência para os dois diferentes status, para duas ou mais vidas, pode-se obter o valor da anuidade para cada status. Onde as anuidades conjuntas são representadas por $\ddot{a}_{xy}$ e as anuidades para o status do último sobrevivente por $\ddot{a}_{\overline{xy}}$ (Pires *et al.*, 2021). As formulações seguintes consideram uma anuidade conjunta com diferimento m e temporariedade n . Assim, tem-se:

$${}_m | \ddot{a}_{xy:n} | = \sum_{t=m}^{m+n-1} {}_t p_{xy} \cdot v^t \quad (15)$$

$${}_m | \ddot{a}_{\overline{xy}:n} | = \sum_{t=m}^{m+n-1} {}_t p_{\overline{xy}} \cdot v^t \quad (16)$$

As expressões (15) e (16) representam o valor presente atuarial de uma série de pagamentos unitários, pagável enquanto ambos, x e y , estiverem vivos, no caso da expressão (15) ou enquanto pelo menos um estiver vivo, x ou y , no caso da expressão (16) (Bowers *et al.*, 1997; Dickson; Hardy; Waters, 2013).

Nas expressões apresentadas, assume-se independência entre as probabilidades de sobrevivência das duas vidas, o que nem sempre pode ser válido, mesmo eliminando a necessidade de modelos complexos (Denuit *et al.*, 2005). Neste sentido, a literatura tem apresentado evidências de que pode haver associação entre os tempos restantes de vida. Dessa forma, as probabilidades de sobrevivência descritas para os status de vida conjunta e de último sobrevivente podem não ser uma boa representatividade para vidas dependentes (Walter *et al.*, 2021). Como exemplo mais comum, destacam-se os tempos de vida restantes do marido e da esposa que possuem um certo grau de exposição a riscos semelhantes (Denuit *et al.*, 2005).

A seguir, a próxima seção apresenta uma revisão de estudos na literatura que investigaram e obtiveram resultados significativos sobre a correlação entre vidas, majoritariamente sobre casais.

2.3 ABORDAGEM POR DEPENDÊNCIA

Os estudos e as práticas atuariais pressupõem a independência entre tempos residuais de vida para a aplicação da precificação de produtos de cobertura ligados a múltiplas vidas, em seguros e em anuidades. Os produtos acima mencionados são precificados de acordo com as projeções de mortalidade, que necessitam de análises apuradas, para que seja possível gerenciar os riscos e garantir as estimativas corretas sobre a definição dos prêmios e constituições de reservas matemáticas (Carannante *et al.*, 2024). Todavia, a suposição das premissas sobre a mortalidade pode nem sempre refletir a realidade, devido à possível correlação entre a sobrevivência e a mortalidade dentro de grupos de indivíduos (Onchere, 2022).

A esse respeito, a literatura tem fomentado o debate sobre a correlação entre os tempos residuais de vida, especialmente para os indivíduos pertencentes ao mesmo grupo, como em situações de casais, irmãos, gêmeos ou até populações que apresentam relações socioambientais ou socioeconômicas comuns (Giovanni; Pirra; Viviano, 2025). Dentre os indivíduos pertencentes a um determinado grupo, a sobrevivência deles pode não ser considerada como independente (Gomes, 2024).

Na década de 1980, Oakes (1982) foi um dos primeiros a trabalhar com modelos que introduzem a dependência entre os tempos de sobrevivência de pessoas. As aplicações foram realizadas por meio de modelos de cópulas para estruturar a relação entre as vidas sem alterar suas distribuições marginais. A pesquisa evidenciou que, para certos casos, a dependência apresentada pode afetar de forma significativa as inferências estatísticas. Outro estudo que se pode citar é o de Hougaard (1986), que trabalhou com modelos de fragilidade para tempos de vida correlacionados entre gêmeos e evidenciou que os tempos de vida não são independentes. Windholz, Marmar e Horowitz (1985) realizaram uma revisão de literatura sobre o impacto na mortalidade de um cônjuge dada a morte do outro, o que é denominado como a Síndrome do coração partido, e evidenciaram que pode haver aumentos na mortalidade de homens em idades mais avançadas, dado que perderam suas esposas.

Entre a década de 1990 e 2000, muitos estudos atingiram evidências empíricas em populações específicas sobre semelhanças entre tempos de vida, a título de exemplo, Jagger e Sutton (1991) que evidenciaram, empiricamente, o agravamento de morte de um cônjuge

devido à morte de outro; Hougaard, Harvald e Holm (1992), que estudaram a sobrevivência de gêmeos dinamarqueses e obtiveram como resultado que a dependência para tempos de vida é significativa e supostamente causada pelo compartilhamento de riscos não observados; Frees, Carriere e Valdez (1996), que utilizaram modelos de cópulas para estimar a dependência da mortalidade e verificar o seu impacto sobre o preço de anuidades para múltiplas vidas no status de último sobrevivente.

Nos anos 2000, outros pesquisadores continuaram na busca por uma maior precisão das estimativas dos efeitos da dependência sobre a sobrevivência e a mortalidade de indivíduos em conjunto. Nesta linha, Denuit *et al.* (2000) investigaram a dependência presente entre indivíduos e o seu impacto sobre o contexto de vidas seguradas. O estudo baseou-se em dados do National Institute of Statistics (NIS) e em observações reais da cidade de Bruxelas e, ao quantificar a dependência entre vidas para analisar o seu impacto em cálculos de seguros, obteve resultados significativos, uma vez que a dependência incorporada afetou o preço dos produtos de seguros de vida.

Outros diferentes estudos apresentaram avanços metodológicos com abordagens de diferentes modelagens para entender a dependência e a interação conjunta entre vidas, dentre eles: cópulas de alta dimensão, modelos híbridos, integração de fragilidades e/ou com modelos estocásticos. Para mais detalhes, consultar Spreeuw e Owadally (2012), Jevtić e Hurd (2017), Lu (2017), Henshaw, Constantinescu e Pamen (2020).

Diante das evidências e modelos utilizados na literatura, referentes aos modelos estudados para analisar a sobrevivência ou a mortalidade conjunta, atuários e demais pesquisadores buscaram entender se essa dependência poderia ser significativa, de forma a gerar variações sobre os prêmios e as provisões técnicas de seguros com coberturas para grupos de indivíduos. Alguns desses estudos são apresentados na seção 2.4.

2.4 ESTUDOS CORRELATOS

A dependência no contexto da precificação de rendas tem sido analisada empiricamente e fundamentada através de diversos estudos que utilizam distintas abordagens metodológicas. Desde antes dos anos 2000, estudos já analisavam a questão da dependência na mortalidade e avaliavam seus efeitos na precificação de produtos. Como destaque, tem-se o estudo de Frees, Carriere e Valdez (1996), que a partir da hipótese de que marido e mulher tendem a estar expostos a riscos semelhantes, devido a diferentes fatores de riscos e hábitos de vida que compartilham, concluíram, a partir de dados reais de uma grande seguradora canadense, que

existe uma forte dependência entre as vidas conjuntas. Os resultados de suas estimativas indicaram que há uma redução de cerca de 10% dos valores de anuidades, quando comparadas com o modelo que considera vidas independentes.

Gourieroux e Lu (2013), por sua vez, propuseram uma extensão do modelo de *Freund* com fragilidade para incrementar fatores não observados capazes de capturar a dependência da mortalidade para casais, por meio de saltos de mortalidade, o que permite separar dois tipos de dependências: a síndrome do coração partido e a dependência induzida por fatores de riscos comuns. O estudo identifica a correlação da mortalidade conjunta entre vidas e a necessidade de ela ser incorporada na gestão de riscos em seguros e anuidades para múltiplas vidas, que são precificadas de forma a não considerar a relação entre vidas.

O fenômeno da síndrome do coração partido foi também explorado empiricamente por Lu (2017), em que, por meio de dados de anuidades conjuntas, revelaram-se efeitos persistentes e assimétricos sobre a mortalidade de cônjuges devido à morte do parceiro.

Seguindo a literatura, o estudo de Walter *et al.* (2021) propôs tábuas de mortalidade que incorporam a dependência entre vidas, utilizando uma abordagem de fragilidade estável positiva. Para isso, fizeram uso dos dados da Associação de Seguradoras do Quênia (AKI) como taxas de riscos-base para homens e mulheres. O estudo empregou a metodologia da fragilidade referente a fatores não observados e, assim, capturar a dependência. Para a análise de vidas conjuntas e último sobrevivente, os pesquisadores utilizaram dados de anuidades de apólices em vigor entre 2001 e 2013, e os resultados mostraram que quando a dependência é considerada, há uma redução da responsabilidade da seguradora nas idades iniciais do beneficiário, seguida de aumentos progressivos com o avanço da idade. As aplicações utilizaram a abordagem de Fragilidade Estável Positiva para contabilizar a correlação de mortalidade entre as vidas.

Assim como observado nos estudos indicados acima, em que a dependência pode levar a distorções relevantes na precificação de produtos, Henshaw *et al.* (2023) analisaram dados reais de pensões no Egito, por meio dos quais evidenciaram variações significativas na força de associação entre os diferentes tipos de vínculos familiares, que não são capturados em modelos atuariais tradicionais ou em tábuas estrangeiras, e pode resultar em variações no preço de produtos para múltiplas vidas.

Na literatura brasileira, o caso de dependência na abordagem para múltiplas vidas foi abordado e investigado por Gomes (2024), que analisou a relação de dependência entre tempos futuros de vida de casais por meio de modelos de cópulas. A implementação do modelo é feita com dados simulados de 2000 contratos, incluindo um homem e uma mulher em cada um, e

considerando três cenários de mortalidade diferentes (cada um fundamentado em tábuas de mortalidade distintas, duas segregadas por sexo e uma para ambos os sexos). As simulações são analisadas por meio de estimativas paramétricas e não paramétricas a partir de diferentes técnicas de análise de sobrevivência. O estudo constatou que há casos em que a avaliação da dependência representa diferenças significativas na precificação de rendas para casais, pois as estimativas resultam em rendas menores em relação aos cálculos de rendas tradicionais, o que pode implicar superestimação das anuidades para os casos de últimos sobreviventes.

De volta à literatura internacional, Cabrignac, Charpentier e Gallic (2020) não apenas estudaram o caso de dependência entre cônjuges, mas também analisaram os casos de associações entre os membros familiares, isso a partir de dados históricos. O estudo encontrou evidências de associação positiva, ainda que pequena, mas significativa, entre vidas de pais e filhos. As evidências mostram que a dependência pode impactar as esperanças de vida, assim como os valores das anuidades e dos seguros, especialmente para casos reversíveis.

Outro estudo que se concentra na análise da mortalidade e na precificação de produtos de renda, especialmente anuidades, foi o de Onchere (2022), que foi desenvolvido com foco em aprimorar os modelos existentes para lidar com a heterogeneidade e a dependência presente na relação entre as vidas seguradas. O estudo utilizou um processo de fragilidade composta e abordou o caso de seguro de anuidades. Para o caso de anuidades para múltiplas vidas, foram utilizados dados reais de uma importante seguradora queniana. No estudo, o autor analisou um total de 398 contratos de anuidades, com vigência entre os anos de 2001 e 2018. Foram extraídas, a partir dos contratos de anuidades com coberturas para os status de vida conjunta e de último sobrevivente, informações como sexo, data de nascimento, início e duração da vida dos cônjuges e dos contratos, dados estes que possibilitaram uma análise da correlação entre os tempos de vida dos segurados ao longo do tempo.

A adoção do modelo com fragilidade permitiu a captura da dependência entre vidas, e os resultados mostraram que, nos contratos de anuidades, os pagamentos aos segurados diminuem nos primeiros estágios, mas aumentam posteriormente no contrato. O fato pode ser explicado pelo fato de que casais mais frágeis tendem a morrer primeiro. Estes resultados encontrados também evidenciam a necessidade de melhores ajustes nos cálculos atuariais, em produtos ligados à mortalidade, para garantia da avaliação precisa das precificações nos mercados de seguros e previdência.

Abordadas algumas das premissas do cálculo atuarial ligadas a uma ou mais vidas, bem como os estudos presentes na literatura que evidenciam a associação de dependência entre tempos de vida dentro de certos grupos de indivíduos – como casais, gêmeos e famílias – e o

impacto dessa dependência na precificação dos seguros, o próximo capítulo detalha os procedimentos metodológicos adotados neste estudo para atingir os objetivos propostos.

3 METODOLOGIA

Neste capítulo, são apresentados os principais elementos metodológicos do estudo, incluindo: i) o tipo de pesquisa; ii) a coleta e a simulação dos dados referentes à mortalidade de casais; iii) as noções básicas de análise de sobrevivência; iv) os modelos de fragilidade, assim como, a sua estimativa para vidas dependentes; v) a determinação de anuidades para o status de último sobrevivente; vi) a comparação entre os modelos com dependência e independência de sobrevivida.

Para aplicação dos pontos citados, as análises empíricas de cálculos serão realizadas por meio do *software* estatístico R (R Core Team, 2024). O uso do *software* possibilitará a implementação dos modelos estatísticos e a avaliação da mortalidade conjunta.

3.1 TIPO DE PESQUISA

A pesquisa científica é descrita como um processo de investigação para solucionar, responder ou se aprofundar sobre uma indagação no estudo de um fenômeno (Sindelar; De Conto; Ahlert, 2014). Ela é desenvolvida mediante o acúmulo de conhecimentos, assim como a sua natureza, a utilização de métodos, as técnicas e outros procedimentos científicos (Gil, 2017). Para esse estudo, tem-se por finalidade realizar uma pesquisa de abordagem quantitativa, a qual consiste em um método de se fundamentar aplicações da teoria estatística da probabilidade, de forma a generalizar dados e quantificar resultados, sendo importante no auxílio e na investigação de estudos das ciências sociais (Nascimento; Souza, 2017).

Ademais, enquadra-se como: descritiva, que envolve descrever características de uma população ou fenômenos, assim como a relação entre variáveis; inferencial, que se refere à realização de testes estatísticos que permitam a comparação dos resultados obtidos (Gil, 2017). Dadas as definições, justifica-se o enquadramento da pesquisa, porque fornecerá descrições detalhadas sobre os dados, os modelos, as teorias, os resultados e sobre a população em um estado conjunto, exposição e o teste de fatos, métodos atuariais comuns e estatísticas sobre as variáveis.

Quanto à delimitação temporal, o estudo se enquadra como uma análise longitudinal, que consiste no acompanhamento de movimentos e fenômenos ao longo do tempo, o que se enquadra com o estudo que trata de acompanhar o comportamento do fenômeno da mortalidade

para múltiplas vidas com a hipótese de dependência ao longo do tempo (Fitzmaurice; Laird; Were, 2011).

3.2 COLETA DE DADOS

O conjunto de dados utilizado para o desenvolvimento da pesquisa foi construído a partir das informações secundárias coletadas dos registros históricos disponíveis no site *FamilySearch* (<https://www.familysearch.org/pt/home/portal/>). O *FamilySearch* é uma organização internacional, sem fins lucrativos, que tem como objetivo ajudar pessoas a descobrir seu histórico familiar (FamilySearch, 2025).

No site da organização, é possível encontrar bilhões de registros históricos digitalizados de diferentes eventos, como nascimento, casamento, óbito, além de outras informações sobre os locais dos eventos e histórico familiar. Com base nos registros, foi possível coletar um conjunto de dados com informações de diferentes casais, assinalando um identificador numérico único para cada indivíduo, para anonimizar as informações, e outro identificador numérico para cada casal. Para o tratamento e fins dos ajustes das informações coletadas para a formação da base de dados, utilizou-se o *software* estatístico R (R Core Team, 2024).

Um resumo das principais variáveis é apresentado no Quadro 1, um misto de variáveis coletadas, bem como os identificadores de cada casal e indivíduos.

Quadro 1 - Descrição das variáveis utilizadas

Variáveis	Tipo de informação	Descrição
Id do grupo	Texto	Caracteres de identificação dos indivíduos pertencentes a cada casal.
Id individual	Texto	Id de identificação individual dos indivíduos.
Ano de nascimento	Numérica	Informação do período de nascimento dos indivíduos
Ano de casamento	Numérica	Informação do ano de casamento
Ano de morte ambos	Numérica	Informação sobre morte dos indivíduos
Sexo	Qualitativa	Informação do sexo dos indivíduos.

Fonte: Elaborado pelo autor (2026) com base em Gomes (2024) e Onchere (2022).

As informações de registros foram coletadas com base em um recorte temporal, entre os anos 1900 e 2020. O país selecionado para a coleta dessas informações foi o Brasil. Os registros que continham informações incompletas, aproximadas, intervalares incompatíveis

(datas de casamento após a morte de um integrante, casamento com idades muito abaixo de 15 anos, etc.) foram excluídos, a fim de minimizar erros.

Inicialmente, a base total foi formada por 902 casais (1804 observações) e a coleta foi realizada de maneira ordenada conforme os filtros aplicados (intervalo temporal e as datas de nascimento, óbito e casamento). No entanto, durante a análise exploratória e a verificação de erros nos dados, 7 casais foram removidos por apresentarem inconsistências (anos de casamento divergentes entre o casal ou morte ocorrida antes/durante o ano de casamento). Após o ajuste final dessas inconsistências, a base de dados definitiva contou com 895 casais (1790 registros individuais).

A Tabela 1 apresenta as seis primeiras linhas da base de dados coletada por meio das informações de registros.

Tabela 1 - Dados dos três primeiros registros de casais contidos na base

ID_casal	ID_ind	sexo	Ano_nasci	Ano_morte	ano_casam	tempo	status
1	ID_0001	Masculino	1900	1976	1926	76	1
1	ID_0002	Feminino	1909	1978	1926	69	1
2	ID_0003	Masculino	1905	2005	1929	100	1
2	ID_0004	Feminino	1910	2000	1929	90	1
3	ID_0005	Masculino	1915	1979	1935	64	1
3	ID_0006	Feminino	1907	1989	1935	82	1

Fonte: Elaborado pelo autor, no *software* R, a partir dos dados coletados junto ao Family Search (2025).

Como todos os registros coletados para a amostra apresentam o evento de morte, além do tempo exato, considera-se que não há presença de censura. Dessa forma, a variável status, que indica se os indivíduos apresentaram o evento, assume valor igual a 1 para todas as observações; o valor 1 indica a morte e, caso houvesse censura, o valor da variável status seria 0 para aquele registro. A utilização de dados sem censura partiu de uma escolha deliberada para simplificar a análise e evitar incertezas quanto ao tratamento de observações incompletas.

Conforme exibido no Quadro 1 e na Tabela 1, essas informações viabilizaram a análise do tempo residual de vida dos integrantes de casais e o seu efeito sobre a precificação dos contratos de seguros para vidas múltiplas.

Na seção 3.3 é apresentado as noções básicas para realização de aplicações práticas em análise de sobrevivência, assim como, a definição das principais funções dos tempos de falhas e censura, estimador de Kaplan-Meier assim como outros pontos que se fizeram importantes para desenvolvimento do estudo.

3.3 NOÇÕES BÁSICAS DE ANÁLISE DE SOBREVIVÊNCIA

Esta seção é dedicada à apresentação dos principais conceitos e descrições de análise de sobrevivência e as técnicas utilizadas para a realização da pesquisa. De modo que, por meio das abordagens apresentadas, se espera: i) apresentar os elementos básicos para a aplicação de uma análise de sobrevivência; ii) estimar a função de sobrevivência para identificar a estrutura do tempo residual de vida dos segurados; iii) a partir da sobrevivência empírica, realizar o ajuste por meio de uma distribuição paramétrica; iv) obter as funções de risco e sobrevivência condicionada à fragilidade.

A análise de sobrevivência consiste em uma metodologia utilizada na estatística para investigar o tempo até a ocorrência de um evento de interesse, chamado de tempo de falha (Colosimo; Giolo, 2006). A falha, muitas vezes, é um evento indesejado em um acompanhamento (aparição de uma doença, morte, recaída após uma resposta ao tratamento...). O tempo de falha pode ser exemplificado como o número de dias até a morte de uma pessoa após ser internada ou, no presente estudo, o tempo de idade ao óbito de um ou mais indivíduos. Nesse caso, em que se observa a sobrevivência, o tempo de falha é definido pelo tempo, em anos, entre o nascimento ou a contratação de um seguro de vida e a data de óbito (Colosimo; Giolo, 2006).

No entanto, uma característica comum quando se procura observar o tempo até a ocorrência do evento é que muitas vezes não é possível obter informações completas sobre o tempo de ocorrência ou se realmente o evento ocorreu durante o período analisado, o que pode ocasionar problemas na estimação. Esse aspecto é um dos principais fatores que diferenciam as técnicas de análise de sobrevivência dos modelos clássicos de regressão e é denominado de censura (Tomazella, 2003). A censura corresponde, segundo Colosimo e Giolo (2006), à observação parcial da resposta, ou seja, quando não há a informação completa de acompanhamento, o que dificulta estimar as curvas de sobrevivência por meio de abordagens paramétricas, sendo comumente feita por meio de estimativas não paramétricas.

Ressalta-se que os tempos de censuras, mesmo representando informações incompletas, podem ser relevantes para investigar o evento de interesse, por duas razões: i) as observações censuradas fornecem informações sobre o tempo de vida dos expostos ao evento (ainda que parciais); ii) a omissão no cálculo das estatísticas de interesse pode levar a conclusões viciadas (Colosimo; Giolo, 2006).

A censura pode ser definida em diferentes tipos, sendo: Censura à direita, Censura à esquerda e Censura intervalar. O Quadro 2 apresenta os tipos e as suas respectivas definições:

Quadro 2 – Tipos de censura para dados de sobrevivência

Tipo de Censura	Descrição
Censura a direita	Ocorre quando o evento de um indivíduo não ocorreu até o final do período de estudo ou quando ele é perdido para acompanhamento. Imagine o estudo, com acompanhamento por 20 anos; a informação de censura ocorreria caso o indivíduo sobrevivesse ao evento por mais de 20 anos.
Censura a esquerda	Isso ocorre quando um evento já aconteceu antes do início da observação, mas o tempo exato é desconhecido. Sabe-se que o evento ocorreu antes de um certo tempo.
Censura intervalar	A censura intervalar ocorre quando o tempo exato do evento de interesse (falha, morte, etc.) não é observado com precisão, mas sabe-se que ele ocorreu dentro de um intervalo de tempo.

Fonte: Elaboração própria, a partir de Carvalho *et al.*, (2011); Borges (2014).

É importante destacar que a censura pode ser informativa e não informativa, esta se refere a um cenário em que a censura dos dados não fornece nenhuma informação adicional sobre o tempo de sobrevivência dos indivíduos, enquanto aquela se refere à censura que está relacionada ao evento de interesse e pode representar relevância. Para mais detalhes, consultar Carvalho *et al.* (2011), Cap. 2, seção 2.2.

Uma outra característica comum para estudos em análise de sobrevivência é o truncamento. Esse é caracterizado pela condição de exclusão de certos indivíduos do estudo. O truncamento pode ser exemplificado para estudos em que o acompanhamento inicia após os indivíduos experimentarem um certo evento, por exemplo, indivíduos que apresentam invalidez, e o evento de interesse seja a morte (Colosimo; Giolo, 2006).

Para determinar o tempo de falha, é necessário definir o evento e, a partir disso, observar o tempo de origem (que pode ser o ano de nascimento ou o ano de início de um contrato de seguro).

No que diz respeito à representação, os dados de sobrevivência para um indivíduo i ($i = 1, 2, \dots, n$) pode ser codificado por um par de variáveis (T_i, δ_i) na qual o T_i seria o tempo observado até ocorrência da falha e δ_i é a variável responsável por indicar a censura ou o evento.

Para o estudo, os dados coletados e tratados não possuem censura, pois as informações são tidas como completas, abrangendo as informações iniciais e finais de vida dos indivíduos, isto é, todos apresentam o evento de interesse.

Em relação as funções utilizadas para representar o comportamento da variável de estudo, a função de risco e sobrevivência foram descritas na seção 2.1, assim como a função densidade e a função de distribuição acumulada.

É importante destacar que as 4 funções mencionadas são relacionadas entre si e, a partir de uma, é possível obter as demais por meio de relações matemáticas (Carvalho *et al.*, 2011). Além das 4 funções apresentadas, $F_x(t), f_x(t), S_x(t), \lambda_x(t)$, é possível destacar outra função bastante útil em análise de sobrevivência. Ela é conhecida como função de risco cumulativo e pode ser interpretada como a taxa de falha instantânea de o evento ocorrer no momento t (Calsavara, 2011), definida pela integração da função de risco $\lambda_x(t)$, conforme a equação (17):

$$\Lambda(t) = \int_0^t \lambda(u) \cdot du; t \geq 0 \quad (17)$$

A equação (17) representa acumulação do risco à medida que o tempo aumenta, e possui como propriedades: i) sempre positiva ii) o valor tende para o infinito (∞) caso $t = \infty$ (Bedia, 2022). Essa pode ser útil na avaliação da função de maior interesse, que é a função de risco (Colosimo; Giolo, 2006).

A estimativa da função de sobrevivência dos segurados foi realizada pelo estimador de Kaplan-Meier, que consiste em uma abordagem não paramétrica que é realizada a partir dos tempos de sobrevivência observados, com a presença ou não de censura (Kaplan; Meier, 1958).

O estimador de Kaplan-Meier é não paramétrico e calcula a probabilidade de sobrevivência após um tempo específico t . O método possibilita a utilização de informações completas (eventos observados) e parciais (censuras) com menor viés, ajustando dinamicamente o número de indivíduos em risco n_i a cada tempo t_i . A estimação da função de sobrevivência pelo método Kaplan-Meier, leva em consideração: $t_1 < t_2 \dots < t_k$, para os k tempos ordenados e distintos de falha; d_i , que representará o número de mortes (falhas) em t_i , para $i = 1, \dots, k$; e n_i , que é o número de indivíduos expostos ao risco em t_i , ou seja, os indivíduos sobreviventes e que não foram censurados até o instante imediatamente anterior a t_i (Colosimo; Giolo, 2006).

Logo, define-se o estimador de Kaplan-Meier como:

$$\hat{s}(t) = \prod_{i: t_i \leq t} \left(\frac{n_i - d_i}{n_i} \right) = \prod_{i: t_i \leq t} \left(1 - \frac{d_i}{n_i} \right) \quad (18)$$

A equação (18) descreve o estimador de Kaplan-Meier, que, segundo Colosimo e Giolo (2006), pode ser visto como um estimador de máxima verossimilhança não paramétrico por não requerer a especificação de um modelo paramétrico de sobrevivência. Esse estimador delinea as probabilidades de sobrevivência de forma ajustada para dados censurados (Colosimo; Giolo, 2006). Os valores podem assumir constância entre os instantes dos eventos e, portanto, a curva de sobrevivência estimada é uma função escada com degraus que muda de valor no instante de ocorrência dos eventos, conforme figura a seguir (Colosimo; Giolo, 2006).

Naturalmente, na ausência de censura, o estimador se reduz à função de sobrevivência empírica (Colosimo; Giolo, 2006). Porém, mesmo com o estudo não apresentando censuras, optou-se por manter a aplicação, por convenção metodológica e para fins de comparabilidade com a literatura.

A partir do Kaplan-Meier, estimou-se as curvas de sobrevivência para cada sexo. Ao estimá-las, muitas vezes, faz-se relevante compará-las para verificar se há diferença estatisticamente significativa, entre as curvas de sobrevivência, por meio de hipóteses (Rocha; Papoila, 2009). Para isso, utilizou-se o teste de log-rank. Este teste é comumente utilizado para verificar se as curvas de sobrevivência entre grupos são diferentes (Colosimo; Giolo, 2006). O teste se destaca por ser o mais utilizado na análise de sobrevivência (Colosimo e Giolo, 2006 pag. 55).

A hipótese testada foi:

H_0 : Não há diferença entre as curvas de sobrevivência.

Onde, para a rejeição da hipótese nula (H_0), o valor-P do teste deve ser inferior ao nível de significância considerado. Para o estudo se considerou um nível de significância de 0,05. A seção 3.4 a seguir expõe os modelos de fragilidades, que são responsáveis por tentar explicar o efeito aleatório para capturar informações que não são capturadas ou identificadas. Os modelos são fundamentados por diferentes autores, como Colosimo e Giolo (2006), Carvalho *et al.* (2011) Balan e Putter (2020) e Wienke (2010). Estes trabalhos fornecem uma visão geral da teoria e da prática no campo de modelos de fragilidade e aplicações por meio de exemplos distintos.

3.4 MODELOS DE FRAGILIDADES

A análise de sobrevivência frequentemente enfrenta desafios quanto à heterogeneidade entre indivíduos, que pode impactar significativamente a função de risco e a sobrevivência (Vaupel; Manton; Stallard, 1979).

A heterogeneidade corresponde a fatores não observados ou não identificados que não se conseguem medir e que podem alterar o comportamento da função de risco e a sobrevivência dos indivíduos (Colosimo; Giolo, 2006). Os fatores não observados, e que geram a heterogeneidade, se fazem relevantes para estudos de análise de sobrevivência, dado que a sua captura pode representar informações mais ajustadas em relação ao que é observado (Wienke, 2010).

Quando observada sob o critério, o nível hierárquico ou a unidade da análise de atribuição, a heterogeneidade pode aparecer de forma individual ou coletiva entre os indivíduos (Balan; Putter, 2020). Na individual, cada indivíduo recebe seu próprio termo de heterogeneidade, que representa as diferenças não observadas entre os indivíduos que não conseguem ser explicadas por covariáveis conhecidas. Na coletiva, é atribuída a um grupo de indivíduos, e todos dentro do grupo compartilham o mesmo termo aleatório.

Para abordar essa questão, modelos de fragilidade têm sido amplamente utilizados nas últimas décadas, pois correspondem a modelos de sobrevivência que incorporam explicitamente a heterogeneidade não observada por meio de um efeito aleatório sobre o risco (Onchere, 2022).

De acordo com Wienke (2010), a ideia geral desses modelos é que os indivíduos possuem diferentes fragilidades, de forma individual ou coletiva (dentro de determinados grupos), e que os mais frágeis morrerão mais cedo do que os menos frágeis.

Antes dos modelos de fragilidade, os estudos ligados à mortalidade, por exemplo, buscavam apenas a inclusão de covariáveis que são observadas, de forma que podem ser incluídas por meio de um modelo bastante conhecido na literatura, o modelo de riscos proporcionais (Cox, 1972). No entanto, é quase impossível incluir todos os fatores de risco importantes, talvez porque o pesquisador tenha pouca ou nenhuma informação no nível individual. Isso ocorre, por exemplo, em estudos populacionais sobre a mortalidade, em que muitas vezes as únicas variáveis conhecidas são sexo e idade (Wienke, 2010).

O surgimento da fragilidade, conforme destacado por Onchere (2022) e Wienke (2010), iniciou-se com Beard (1959) para explicar o padrão de comportamento da mortalidade, porém utilizava-se o termo “fator de longevidade” e não “fragilidade”. Posteriormente, no campo da demografia, Vaupel, Manton e Stallard (1979) introduziram aplicações usando o termo “fragilidade” para expressar um fator de risco, um efeito aleatório não observável, o que

formalizou a ideia sobre riscos que não são capturados por covariáveis. O estudo trouxe evidências matemáticas de que a inclusão da fragilidade poderia afetar as taxas de mortalidade dentro de uma população.

A inclusão da fragilidade sobre a função de risco ocorre para representar fatores não observáveis, como fatores ambientais e genéticos ou informações que, por algum motivo, não foram capturadas e consideradas no estudo ou no modelo (Carannante et al., 2024; Wienke, 2010).

Os modelos de fragilidade podem ser aplicados: a) de forma univariada, situação em que se analisa apenas uma função de risco de forma individual para contabilizar a heterogeneidade não observada por indivíduo; b) de forma multivariada, onde a fragilidade é representada por um vetor de fragilidades, usado quando existem múltiplos riscos correlacionados e de forma compartilhada, caso em que se supõe-se uma fragilidade comum a todas as observações de um mesmo *cluster* (grupo) ou a todos os eventos de um mesmo sujeito, dado que compartilham fatores de risco comum, o que pode ocasionar em tempos de falha correlacionados (Colosimo; Giolo, 2006). As seções a seguir descrevem os modelos para os casos univariado e compartilhado, sendo esse o caso a ser aplicado para identificar uma possível dependência entre os tempos residuais de vida dentro de grupos.

3.4.1 Modelo univariado de fragilidade

Como ponto principal para o estudo dessa heterogeneidade individual, Wienke (2010) enfatiza que a desaceleração da mortalidade em idades avançadas é resultado de uma heterogeneidade populacional, em que os indivíduos mais frágeis morrem mais cedo, deixando um subconjunto mais saudável da população. Esse entendimento levou à integração de variáveis latentes, muitas vezes não observadas, ao modelo de riscos proporcionais (Cox, 1972) e foi o ponto de partida para o desenvolvimento dos modelos de fragilidade (Wienke, 2010).

O modelo de fragilidade modifica o modelo de Cox (Cox, 1972), que considera apenas fatores de risco declarados que afetam a mortalidade, e passa a incorporar a fragilidade. Este modelo incorpora melhorias que não são compreendidas na sua forma padrão (Aalen, 1988).

O modelo de Cox se destacou por ser um dos mais eficientes devido ao fato de incorporar covariáveis dependentes do tempo (Colosimo; Giolo, 2006). A seguir, apresenta-se a função geral do modelo de Cox:

$$\lambda_i(t) = \lambda_0(t) \cdot \exp(X'_i \cdot \beta_i) \quad (19)$$

A equação (19) descreve a formalização do modelo de Cox, em que $\lambda(t)_i$ descreve a função de risco geral (ou taxa de risco) para o indivíduo i , enquanto $\lambda_0(t)$ representa o componente não-paramétrico e não especificado, sendo uma função não negativa e denominada como função de base. Já $\exp(X'_i \cdot \beta_i)$ é o componente paramétrico, em que x representa as covariáveis e β , os coeficientes, ambos associados a cada indivíduo i (Colosimo; Giolo, 2006). O modelo descrito é conhecido por apresentar uma razão constante na taxa de falhas para cada par de indivíduos (Balan; Putter, 2020).

Ele se apresenta eficiente para compreender as variáveis relevantes e que influenciam a função de risco. Entretanto, o modelo não leva em consideração à presença de heterogeneidades nos tempos de sobrevivência (Colosimo; Giolo, 2006).

Logo, para analisar o efeito da fragilidade, ela é incluída como uma variável de efeito aleatório, de forma a atuar de forma multiplicativa ou aditiva sobre o risco (Colosimo; Giolo, 2010). De acordo com Wienke (2010), os modelos apresentam as formas aditiva e multiplicativa, para o estudo, o foco será no modelo multiplicativo.

Dessa forma, a variável aleatória que indica essa heterogeneidade atua sobre a função de risco, podendo agravar ou suavizar o risco de morte (evento). A sua inclusão no modelo de Cox (1972) modifica o modelo, que se torna um modelo de fragilidade, pois indivíduos com maior fragilidade estarão mais suscetíveis ao evento (Calsavara, 2011). Logo, a função de risco para um indivíduo no tempo de forma condicionada à fragilidade:

$$\lambda_i(t|Z_i) = Z_i \cdot \lambda_0(t) \exp(X_i \cdot \beta) \quad (20)$$

Na equação (20), Z é representa uma variável aleatória não negativa e desconhecida que varia entre os i indivíduos de uma população (Wienke, 2010). Essa variável caracteriza o efeito da heterogeneidade sobre a função de risco base $\lambda_0(t)$ e sobre $\exp(X \cdot \beta)$, na qual X representa o vetor de covariáveis e β é o vetor de coeficientes associados a X (Wienke, 2010).

De forma a simplificar, podemos reescrever o modelo univariado em uma estrutura sem as covariáveis:

$$\lambda_i(t|Z_i) = Z_i \cdot \lambda_0(t) \quad (21)$$

Com apenas a fragilidade, a função de risco basal pode ser impactada pela fragilidade Z , de forma que quanto maior o valor da fragilidade, os indivíduos estarão mais suscetíveis ao evento, assim o risco é crescente quando se tem $Z > 1$, decrescente para $Z < 1$, e para $Z = 1$, a fragilidade não tem influência sobre o risco e, dessa forma, o modelo se reduz ao modelo de Cox (Cox, 1972; Calsavara, 2011; Wienke, 2010).

Observa-se que a função de risco de base no modelo absorve por completo o fator de fragilidade. Logo, para garantir que o modelo seja identificável, faz-se necessário que a variável fragilidade se comporte como sendo distribuída identicamente, de forma a apresentar uma distribuição de probabilidade conhecida e padronizada com $E(Z)=1$ (Calsavara, 2011). Enquanto a variância $Var[Z] = \theta$ para essa distribuição (se existir) pode ser interpretada como uma medida da heterogeneidade da população em relação ao risco basal.

A necessidade da identificação é comprovada por Elbers e Ridder (1982), na forma (20), de modo que o modelo deve atender às três seguintes suposições:

- i) A variável da fragilidade deve seguir uma distribuição qualquer degenerada e padronizada em $E[Z] = 1$, caso exista. Enquanto a variância dela é desconhecida, $Var[Z] = \theta$ (se existir), sendo essa interpretada como a medida de heterogeneidade entre a população no risco de base, que pode ser nula θ quando vai para 0. Assim, a partir da normalização, torna-se possível identificar o que realmente representa a heterogeneidade, o risco de base e o componente paramétrico;
- ii) A função de risco cumulativo $\Lambda(t)$ é obtida a partir de uma função não negativa $\lambda(t)$;
- iii) As covariáveis X devem apresentar variabilidade suficiente, enquanto a função paramétrica deve ser diferenciável e não constante. E assumir a normalização para que um dado valor de X , $\exp(X \cdot \beta) = 1$ (Honoré, 1993).

Essas condições foram inicialmente apresentadas por Elbers e Ridder (1982) e, posteriormente, Honoré (1993) ampliou esses resultados para os modelos múltiplos de duração, de modo que a identificação pode ser obtida mesmo sem restrições adicionais quando há mais de uma observação por indivíduo.

De acordo com Wienke (2010), assumem-se as hipóteses, de forma arbitrária, de que a fragilidade é independente do tempo e de que ela atua multiplicativamente sobre a função de risco basal. Essas hipóteses serviram de base para grande parte dos trabalhos sobre heterogeneidade não observada.

Com a fragilidade incluída, a função de sobrevivência condicionada a ela correspondente ao modelo univariado é dada por:

$$S(t|Z_i) = \exp(-Z_i \cdot \Lambda_0(t)) \quad (22)$$

Na equação (22), o termo $\Lambda_0(t)$ representa o risco cumulativo. A sobrevivência da população heterogênea, composta por indivíduos com diferentes níveis de fragilidade, como é desconhecida (não observável), faz-se necessário conhecer o modelo marginal, ou seja, o nível populacional em que a fragilidade é integrada (Wienke, 2010; Bedia, 2022).

Segundo Wienke (2010), a função de sobrevivência populacional corresponde a uma média ponderada das funções de sobrevivência condicionais dos indivíduos integrada com a fragilidade, dada por:

$$S(t|Z) = E_Z[\exp(-Z \cdot \Lambda_0(t))] \quad (23)$$

A equação (23) revela que a expressão da sobrevivência populacional é precisamente a transformada de Laplace de variável fragilidade Z . Desse modo, a função de sobrevivência pode ser reescrita sob a forma da Laplace:

$$S(t) = \mathcal{L}_Z(\Lambda_0(t)) \quad (24)$$

De acordo com Bedia (2022), a relação com a transformada de Laplace tem um papel importante para a estimação dos modelos de fragilidade e foi explorada por Hougaard, (1984), Hougaard (1986b). O foco especial é dado à transformada pela facilidade da estimação da função de sobrevivência e dos riscos incondicionais de forma a expressar a estrutura do modelo em termos de uma única função.

Por meio da transformada de Laplace, é possível caracterizar a função de densidade e a função de risco dos tempos até a ocorrência do evento (Wienke, 2010).

A função de densidade com a fragilidade pela transformação de Laplace assume a seguinte forma:

$$f(t) = -\lambda_0(t) \cdot \mathcal{L}'_Z(\Lambda_0(t)) \quad (25)$$

Já a função de risco pode ser expressa por meio da razão entre a função densidade e a função de sobrevivência:

$$\lambda(t) = -\lambda_0(t) \cdot \frac{\mathcal{L}'_Z(\Lambda_0(t))}{\mathcal{L}_Z(\Lambda_0(t))} \quad (26)$$

É possível caracterizar, ainda, a esperança e a variância da fragilidade Z:

$$E[Z] = -\mathcal{L}'^{(0)} \quad (27)$$

$$\text{Var}[Z] = \mathcal{L}''(0) - (\mathcal{L}'^{(0)})^2 \quad (28)$$

Na equação (28) $\mathcal{L}'$ e $\mathcal{L}''$ denotam a primeira e a segunda derivada da transformada de Laplace.

A distribuição da fragilidade descreve o comportamento da heterogeneidade da população. Onde, a função de sobrevivência condicional pode ser expressa, de forma indireta, por meio da transformada da Laplace. De modo que por meio dela, torna-se possível construir a função de verossimilhança, assim como obter os parâmetros da distribuição da fragilidade, geralmente via máxima verossimilhança ou métodos semiparamétricos (Wienke, 2010). Mais detalhes sobre a relação entre o modelo incondicional e o condicional para o caso univariado podem ser consultadas no teorema apresentado por Vaupel, Manton e Stallard (1979), Wienke (2010) e Bedia (2022).

De acordo com Colosimo e Giolo (2006), o fator da heterogeneidade se faz interessante para introduzir os efeitos não observados, assim como, a associação entre indivíduos, que é o foco do presente estudo. No entanto, para o modelo univariado, a heterogeneidade desse modelo é aplicada para descrever um único tempo até o evento por unidade, considerando que cada pessoa é independente e que cada um possui a sua própria fragilidade (Wienke, 2010). Assim, ela não é interessante para casos em que a exposição a riscos semelhantes, em que se tenha associação ou a formação de grupos (Balan; Putter, 2020).

No presente estudo, a fragilidade é considerada no contexto compartilhado, para que se possa entender a associação entre as mortalidades no processo de precificação das rendas. Dessa forma, a subseção a seguir descreve o modelo compartilhado de fragilidade, bem como a sua aplicação para casais ou grupos familiares onde as expectativas de vida podem estar associadas devido a fatores genéticos compartilhados, condições socioeconômicas ou estilos de vida similares.

3.4.2 Modelo compartilhado de fragilidade

A maioria dos modelos e aplicações estatísticas consideram que determinados eventos são estatisticamente independentes uns dos outros, como é o caso do modelo de Cox (Bedia, 2022). No entanto, segundo Colosimo e Giolo (2006), embora a suposição de independência seja válida em muitas observações, em outros ela pode ser inadequada. Por exemplo, para casos em que os tempos de sobrevivência são observados para grupos ou famílias de indivíduos que compartilham um termo de fragilidade comum.

Um exemplo frequente ocorre em casais ou grupos familiares, cujos tempos de vida tendem a ser associados devido a fatores latentes, como genética compartilhada, condições socioeconômicas ou estilos de vida semelhantes. Esses fatores podem gerar correlação entre as heterogeneidades individuais, afetando diretamente a mortalidade conjunta e introduzindo dependência entre os tempos de vida (Wienke, 2010). A identificação dessa dependência é especialmente relevante em áreas como a atuária, pois fornece subsídios mais precisos para a precificação de produtos envolvendo mais de uma vida (Onchere, 2022).

De acordo com Onchere (2022), a dependência surge principalmente pelos fatores já mencionados – os efeitos agregados compartilhados –, como o estilo de vida semelhante ou a exposição conjunta a desastres. Essa dependência pode influenciar e gerar estimativas incorretas sobre o risco real dentro de classes ou grupos populacionais (Colosimo; Giolo, 2006), pois os diferentes fatores não observados que os membros de um mesmo grupo compartilham geralmente ocasionam uma correlação significativa entre os membros (Bedia, 2022).

De forma a entender a dependência ou essa associação entre os casais, faz-se necessária a utilização de técnicas que capturem a dependência ou associação de tempos de vida dentro de grupo de indivíduos (casais) (Gourieroux; Lu, 2013). Para isso, estudos e trabalhos utilizam uma extensão natural da forma univariada do modelo de fragilidade, o chamado modelo de fragilidade compartilhado (Wienke, 2010). Nesse modelo, considera-se que compartilhar um valor de fragilidade comum pode gerar dependência entre os tempos de vida de indivíduos (Gutierrez, 2001).

Para a aplicação de dados consistindo em j grupos ($j=1,...,J$), compostos por n indivíduos ($i = 1, \dots, n_j$), sendo que cada indivíduo i dentro de j compartilham a fragilidade comum. Logo, considerando a observação para grupos de indivíduos e com a presença de covariáveis, tem-se:

$$\lambda_{ij}(t|Z_{ij}) = Z_{ij} \cdot \lambda_0(t) \cdot \exp(\beta \cdot X_i) \quad (29)$$

Na equação (29), o termo Z_{ij} representa a fragilidade compartilhada por i indivíduos dentro do grupo j , com $E[Z]=1$ e $\text{Var}[z]=\theta$. Destaca-se que as fragilidades são consideradas independentes entre as observações de diferentes grupos (Wienke, 2010). Com a fragilidade comum dentro dos j grupos torna-se possível modelar a associação entre as vidas, pois sofrem o efeito da heterogeneidade comum (Balan; Putter, 2020).

De forma similar ao apresentado para o caso univariado, a função de sobrevivência conjunta condicional para o grupo j é dada por:

$$S(t_1, \dots, t_k | Z_j) = \exp \left(-Z_j \sum_{i=1}^k \Lambda_0(t_j) \cdot \exp(\beta \cdot X_{ij}) \right) \quad (30)$$

Ao integrar em relação a Z , pode-se obter a função de sobrevivência marginal conjunta:

$$S(t_1, \dots, t_k) = \mathcal{L}_Z \left(\sum_{i=1}^k \Lambda_0(t_j) \cdot \exp(\beta \cdot X_{ij}) \right) \quad (31)$$

Logo, para isso, faz-se necessário estimar a fragilidade, e, no que se refere à estimativa, devido à aleatoriedade, faz-se necessário supor que ela segue uma distribuição de probabilidade (Gutierrez, 2001; Leão *et al.*, 2017). Estas distribuições podem se fazer relevantes para ajustar o comportamento das variáveis não identificadas, e podem alterar significativamente o resultado da análise de sobrevivência (Balan; Putter, 2020). Logo, faz-se importante entender como a fragilidade pode se comportar, mediante o uso de determinadas distribuições.

Como os modelos de fragilidade, conforme Gutierrez (2001), são utilizados para modelar tempos de sobrevivência na presença de sobredispersão ou efeitos aleatórios específicos de grupos não capturados, permite-se introduzir a heterogeneidade não observada e as associações em modelos para dados de sobrevivência.

De acordo com Calsavara (2011), as principais distribuições utilizadas para modelar o comportamento da fragilidade são a Gama e a Gaussiana Inversa, com uma prevalência crescente em modelos de confiabilidade e na análise de dados agrupados e eventos recorrentes. Destas, a Gama é a mais frequentemente empregada devido à sua eficácia e suas expressões algébricas diretas para funções de sobrevivência e de risco (Hanagal; Sharma, 2013).

Colosimo e Giolo (2006) mencionam o estudo de Vaupel e Yashin (1983), que propuseram outros diferentes tipos de distribuição para descrever o comportamento da fragilidade, como a Uniforme, a Weibull e a Log-normal, mas elas não apresentam os mesmos poderes analíticos da Gama, e geram uma maior dificuldade de serem aplicadas.

Outros estudos presentes na literatura destacam outras distribuições para a estimação paramétrica da fragilidade. Notavelmente, Hougaard (1986, 2000) explorou, junto com os atributos teóricos das distribuições Gama, o uso da Inversa-Gaussiana. Mais recentemente, Wienke (2010) introduziu a Função de variância de potência (PVF), e Leão *et al.* (2017) se concentraram na distribuição Birnbaum-Saunders para modelar o comportamento da fragilidade, que pode ser aplicado a dados com e sem censura.

A distribuição escolhida para o estudo foi a Gama. A escolha da Gama se dá devido à sua simplicidade analítica e computacional e por conta das propriedades atrativas que o modelo pode apresentar, como, por exemplo, a sua facilidade de uso e a praticidade para a estimativa de variáveis aleatórias (Hanagal; Sharma, 2013; Colosimo; Giolo, 2006). O modelo com distribuição de fragilidade gama vem sendo frequentemente o mais empregado para fragilidade em estudos de dados de sobrevivência bivariados (Wienke, 2010). Neste ínterim, destacam-se, na literatura, os trabalhos de Vaupel e Yashin (1983), Hougaard (1986, 2000) e Wienke (2010), que são particularmente notáveis.

Balan e Putter (2020) destacam que distribuição Gama integra a família de distribuições de fragilidade com transformada de Laplace tratável e infinitamente divisíveis. A densidade de uma variável aleatória com distribuição gama é dada por:

$$f(z) = \frac{1}{\Gamma(\alpha)} \cdot \delta^\alpha \cdot z^{\alpha-1} \cdot e^{\{-\delta z\}}, z \geq 0 \quad (32)$$

De modo que o parâmetro de forma α e o parâmetro de escala δ são positivos. Logo, a transformada da fragilidade gama é dada por:

$$\mathcal{L}_Z(s) = \left(1 + \frac{s}{\delta}\right)^{-\alpha} \quad (33)$$

Conforme Wienke (2010) e Bedia (2020), para o caso da fragilidade Gama, e considerando o Laplaciano (33), a esperança de Z pode ser dada por $E[Z] = \frac{\alpha}{\delta}$, $Var[Z] = \frac{\alpha}{\delta^2}$. Para a garantia de identificação do modelo, impõe-se a restrição $E[Z] = 1$. Essa restrição

implica que os parâmetros de forma e escala da distribuição sejam iguais $\alpha = \delta$. Assim, a distribuição passa a ter esperança unitária $E[Z] = 1$ (Wienke, 2010). Assumindo $\theta = \frac{1}{\delta}$ obtêm-se θ . O parâmetro θ é dito como o responsável por controlar o grau de heterogeneidade não observada e a dependência entre indivíduos dentro de um mesmo grupo (Wienke, 2010). Dessa forma, caso θ assumira valor zero, a implicação é de independência entre os tempos residuais dentro dos grupos; caso contrário, existe heterogeneidade comum entre tempos (Hanagal; Sharma, 2013).

Dada a variância representada por θ , tem-se que a variável aleatória da distribuição da fragilidade Gama pode ser representada por:

$$Z_j \sim \Gamma\left(\frac{1}{\theta}, \frac{1}{\theta}\right) \quad (34)$$

De forma que a sua principal vantagem reside na sua transformada de Laplace, a qual permite, segundo Hougaard (2000), converter funções complexas para obtenção de formas de derivação explícitas de trabalhar. De modo que a sua forma fechada é escrita por:

$$\mathcal{L}_Z(s) = (1 + \theta \cdot s)^{-\frac{1}{\theta}} \quad (35)$$

Ao substituir transformada de Laplace da distribuição Gama (Equação 35) na expressão da sobrevivência marginal conjunta (Equação 31) obtém-se a função de sobrevivência marginal para um indivíduo do grupo j , baseado no livro de Therneau e Grambsch (2000) e conforme Wienke (2010, pág. 76):

$$\hat{S}(t) = [1 + \hat{\theta} (\hat{\lambda}_0(t) \cdot \exp^{X' \hat{\beta}})]^{-\frac{1}{\hat{\theta}}} \quad (36)$$

em que:

- $\hat{\lambda}_0(t)$ geralmente o estimador de Nelson-Aalen para função de risco acumulada de base;
- $\hat{\theta}$ estimativa da variância da fragilidade/efeito aleatório;
- $\hat{\beta}$ estimativa dos efeitos/coeficientes das covariáveis; e
- X' matriz de covariáveis.

Analogamente, ao aplicar a Equação (31) ao caso conjunto, obtém-se a função de sobrevivência conjunta e a partir dessa a sobrevivência para cada casal, sendo a forma fechada, conforme Hanagal e Sharma (2013), dada por:

$$\hat{S}(t_1, t_2) = 1 + \hat{\theta} \cdot [\{\hat{\Lambda}_{01}(t_1) + \hat{\Lambda}_{02}(t_2)\} \cdot \exp^{\hat{\beta}^* X_j}]^{-\frac{1}{\hat{\theta}}} \quad (37)$$

Onde: $\hat{\Lambda}_{01}(t_1)$ e $\hat{\Lambda}_{02}(t_2)$ representam as funções de risco basal cumulativas para as variáveis de tempos de vida T1 e T2 (Indivíduos do grupo) (Hanagal: Sharma, 2013).

De forma similar à Equação (14) que representa a probabilidade de sobrevivência para o status de último sobrevivente, ao considerar o modelo com dependência, podemos escrever a sobrevivência com a dependência para o status de último sobrevivente, conforme Walter *et al* (2021):

$$S_{\overline{xy}} = S_x(t) + S_y(t_2) - \hat{S}_{xy}(t_1, t_2 | Z_j) \quad (38)$$

$S_{\overline{xy}}$ representa sobrevivência para o status de último sobrevivente com a dependência induzida pela fragilidade compartilhada. Ao substituir a sobrevivência $P_{\overline{xy}}$ pela sobrevivência dependente $S_{\overline{xy}}$, é possível obter o valor da anuidade sob a hipótese de dependência.

Para realizar a estimação do modelo e obtenção dos parâmetros, utilizou-se o *software* R (R Core Team, 2024), o qual, por meio da função *Coxph*, pacote *Survival*, estima os parâmetros do modelo de Cox estendido, com a fragilidade compartilhada. Para garantir a convergência do modelo, foi necessário aumentar o número máximo de iterações, procedimento comum em contextos com fragilidades, uma vez que a presença de efeitos aleatórios aumenta a complexidade da superfície de verossimilhança (Therneau, 2024; Wienke, 2010). Esse procedimento é comum, uma vez que, nos casos sem fragilidade, a convergência é obtida com menos iterações. Ao obter os resultados para o modelo de Cox com fragilidade, foram estimadas as probabilidades de sobrevivência conjunta para cada casal, sobre o status de último sobrevivente, a partir da expressão (38).

Mais detalhes sobre as parametrizações, suposições e detalhes sobre os modelos, assim como o processo de implementação, podem ser consultados em Hougaard (2000) e Wienke (2010).

Em relação à mensuração da dependência, destaca-se que a variância da fragilidade estimada, acima de zero, pode ser um indicativo de dependência dentro dos grupos. No entanto,

Balan e Putter (2020) apontam que, embora ela indique uma heterogeneidade não observada entre grupos, ela fornece pouca informação sobre a associação entre os tempos dos eventos. Ainda de acordo com os autores, uma medida de correlação que pode ser verificada é o coeficiente de concordância de Kendall τ . Este é um indicador não paramétrico que mede a força de associação (dependência) entre duas variáveis medidas (Hougaard, 2000).

A obtenção do parâmetro de dependência, τ (Kendall), pode ser obtida de forma simples para o caso particular do modelo compartilhado de fragilidade com distribuição gama (Wienke, 2010, p. 156-157), onde é visto que essa medida assume uma forma analítica simplificada para o caso da fragilidade gama. Ao considerar o parâmetro θ como a variância do modelo, o τ é obtido por:

$$\tau = \frac{\theta}{\theta + 2} \quad (39)$$

Como o τ depende do valor da variância do modelo, à medida que ela aumenta, o τ aumenta, indicando que o tempo de vida de um cônjuge está fortemente ligado ao do outro.

Por meio do τ , tornou-se possível avaliar uma métrica de dependência entre os tempos de vida. Ballan e Putter (2020) apontam que para modelo de fragilidade o valor do τ , dada a sua especificação, assume valor nulo ou positivo (entre 0 e 1). Um valor do $\tau = 0$ indica independência, enquanto valores maiores que 0 sinalizam uma associação positiva entre os tempos. Essa métrica é aplicada para verificar a possibilidade de mensurar o valor ou comportamento de uma influência entre os tempos de vida dos cônjuges (Hougaard, 2000).

De forma a validar os resultados obtidos no modelo de fragilidade compartilhado a partir da base de dados utilizada, realizou-se um procedimento de reamostragem e reaplicação do modelo para testar a plausibilidade das estimativas, processo descrito na próxima seção.

3.5 VALIDAÇÃO DOS RESULTADOS DO MODELO

É sabido na literatura que situações sem censura e grupos pequenos (2 elementos) podem gerar problemas na suposição da distribuição qui-quadrado com um grau de liberdade para a estatística do teste utilizado no TRV (Colosimo; Giolo, 2006 p. 302). Nesse sentido, buscando fortalecer a possível hipótese de dependência, estipulada pelo modelo, entre os tempos residuais de vida de casais, utilizou-se um teste simulado por meio de técnicas de bootstrap não paramétrico para construir uma distribuição amostral empírica para o estimador

de θ (variância da fragilidade) e, então, utilizar essa distribuição para avaliar o quão significativamente diferente de zero pode ser θ à luz da amostra analisada.

O bootstrap não paramétrico consiste em um método bem semelhante ao de Monte Carlo, onde se busca realizar um experimento de simulação (Zivot, 2021). No entanto, para o método de bootstrap, busca-se criar dados aleatórios por meio de técnicas de reamostragem com substituição a partir dos dados originais (Zivot, 2021).

Logo, por meio dessa técnica e a partir das amostras dos casais, foi realizado o procedimento de reamostragem e ajuste do modelo, gerando assim um valor de θ para cada nova observação criada. Este processo foi repetido 1000 vezes, e a cada repetição, armazenou-se um valor de θ . Dessa forma, tornou-se possível construir uma distribuição empírica para os valores de θ e, então, utilizar essa distribuição para avaliar a estabilidade e a significância do parâmetro em ser diferente de zero.

Validada essa hipótese, a próxima seção descreve como foi realizada a comparação entre as anuidades, calculadas a partir das hipóteses de dependência e independência.

3.6 COMPARAÇÕES ENTRE MODELOS DE DEPENDÊNCIA E INDEPENDÊNCIA ENTRE OS TEMPOS RESIDUAIS DE VIDA

Feita a estimativa da fragilidade, ela será implementada na função de risco para representar o cenário de dependência e o agravamento ou a suavização da mortalidade conjunta para os tempos de vida. A partir da função risco calculada de forma paramétrica e da variância da fragilidade, foi obtido as probabilidades de sobrevivência para o status de último sobrevivente utilizada nos cálculos de pagamentos das anuidades atuariais.

Posteriormente, serão calculadas e comparadas as rendas para casais levando em consideração o modelo por fragilidade e o modelo clássico com independência para identificar quaisquer disparidades significativas entre as duas técnicas de precificação.

A métrica utilizada para comparar os resultados dos modelos foi a razão entre as anuidades. As medidas tomadas por meio da razão podem indicar a posição e o quanto determinados objetos ou fatos estão distantes entre si em relação à determinada característica (Sindelar; De Conto; Ahlert, 2014). A obtenção da razão é feita pela divisão entre duas quantidades (Hogg *et al.*, 2015). Neste trabalho, o numerador representa a anuidade com dependência ($\ddot{a}_{x,y}^{dep}$) e o denominador será a anuidade com independência ($\ddot{a}_{x,y}^{ind}$) (Ross, 2014). Logo, tomando o resultado da razão como R, então pode definir-se:

$$R = \frac{\ddot{a}_{x,y}^{dep}}{\ddot{a}_{x,y}^{ind}} \quad (40)$$

A saída da equação (40) permitirá a análise das discrepâncias entre as duas abordagens de cálculo:

- Se $R < 1$, isso indica que a anuidade determinada com dependência é menor do que a anuidade calculada sob independência. Nesse cenário, a suposição de independência tende a inflar o valor da anuidade;
- Se $R = 1$, ambos os métodos produzem resultados que não são significativamente diferentes, sugerindo que o efeito da dependência é nulo;
- Se $R > 1$, a anuidade com dependência excede a independência, indicando que negligenciar a dependência pode resultar em uma subestimação do valor da anuidade.

Assim, usando essa métrica, podemos avaliar o efeito relativo da dependência de preços.

Apresentado o caminho utilizado para verificar uma resposta ao problema e atender os objetivos, parte-se ao capítulo 4. Este apresenta os resultados obtidos, assim como as discussões sobre.

4. RESULTADOS E DISCUSSÕES

Este capítulo apresenta os resultados referentes aos tempos residuais de vida dos casais analisados, bem como a investigação sobre a existência de dependência entre tais tempos e o seu impacto sobre o preço de rendas. Nesta linha, apresentam-se: i) as principais estatísticas da análise descritiva dos dados coletados; ii) a estimação da sobrevivência e os resultados dos ajustes paramétricos; iii) os resultados da fragilidade estimada, além de discussões e análises; e iv) os resultados dos cálculos das anuidades, considerando e desconsiderando a dependência, bem como os testes de diferenças entre elas.

4.1 ANÁLISE DESCRITIVA DOS DADOS

A análise descritiva foi realizada a partir da base de dados composta com 1790 observações, dividida por sexo, e totalizando 895 casais. A Tabela 2, indica as principais estatísticas de tendência central e de dispersão sobre os tempos residuais de vida.

Tabela 2 – Resultados das principais estatísticas descritivas

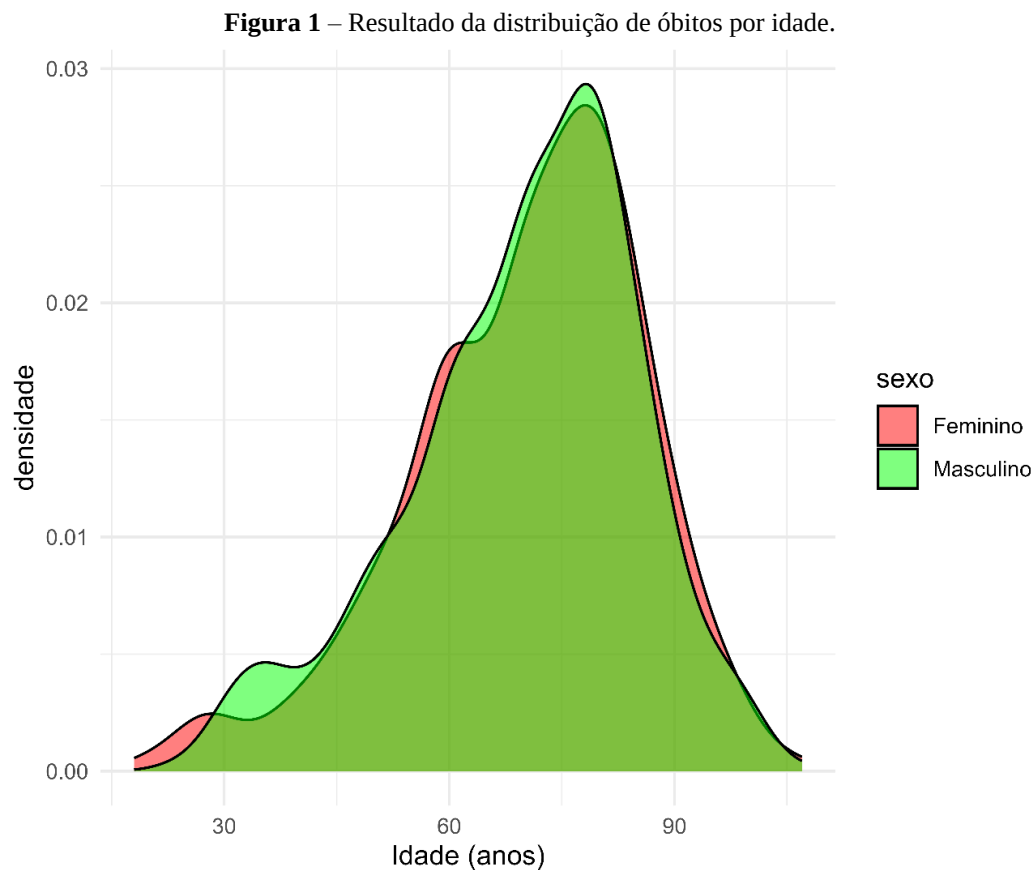
Sexo	Observações	Idade Média	Idade Mediana	DP	Mínimo	Máximo
Total	1790	70,1	72	15,5	18	107
Masculino	895	69,9	72	15,4	23	107
Feminino	895	70,4	72	15,6	18	107

Fonte: Elaborado pelo autor (2026).

A idade média ao óbito foi 70,4 anos para mulheres, uma superioridade de 0,5 anos em relação à média de vida dos homens, que foi 69,9 anos. Para ambos os sexos, a idade média de vida é de 70,1 anos. A dispersão dos tempos de vida, medida pelo desvio padrão, indicou um valor de 15,5 anos. Ao analisarmos por sexo, os resultados indicam uma proximidade entre a dispersão por sexos, onde o desvio padrão para homens é 15,4 anos e para as mulheres é de 15,6. As informações salientam uma variabilidade próxima entre a longevidade por sexo e indicam um comportamento central próximo entre os tempos residuais de vida, com disparidades semelhantes em torno da média.

Outrossim, quanto às grandezas de mínimo e máximo das idades de óbitos, observou-se que as idades máximas de óbito foram idênticas para ambos os sexo, já a idade mínima, essa foi de 18 anos para as mulheres e de 23 anos para os homens. Com o objetivo de visualizar o

comportamento dos óbitos, a Figura 1 demonstra a variação distribuição de óbitos por idade e sexo.



Fonte: Elaboração própria (2026).

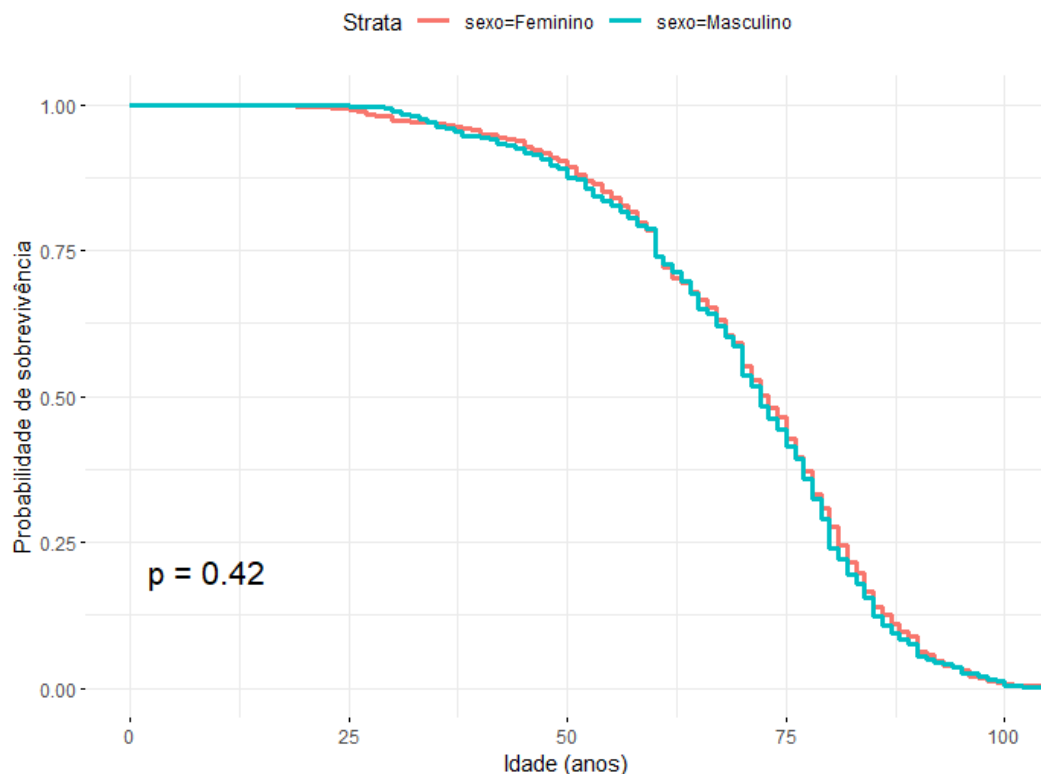
Observa-se que a distribuição de densidade da idade ao óbito apresenta assimetria à direita, com concentração mais elevada entre as idades 60 e 85, sendo o intervalo em que se localiza o pico da curva de densidade. As curvas entre as distribuições por sexo exibem similaridade (poucas variações) em alguns intervalos de idade, o que sugere um comportamento parecido das idades ao óbito por sexo. No entanto, observa-se, entre 30 e 45 anos e nas idades de pico, que a densidade masculina é mais elevada em relação à feminina. Enquanto para idades mais avançadas, após os 85 anos, percebe-se uma densidade maior para os registros do sexo feminino, se sobrepondo à curva masculina – tal fenômeno pode ser compatível com uma maior sobrevivência relativa das mulheres nas idades extremas.

Apresentados os resultados obtidos na análise descritiva, a próxima seção se destina a apresentar os resultados da estimação de sobrevivência não paramétrica e o ajuste paramétrico, detalhados a seguir.

4.2 ANÁLISE NÃO PARAMÉTRICA DA SOBREVIVÊNCIA

A seção atual apresenta os valores estimados da função de sobrevivência, de modo a entender o comportamento do tempo residual de vida dos homens e das mulheres. A função de sobrevivência foi obtida pelo estimador Kaplan–Meier. As curvas de sobrevivência podem ser visualizadas na Figura 2.

Figura 2 - Resultados das curvas de sobrevivência estimadas pelo método de Kaplan-Meier, estratificadas por sexo.



Fonte: Elaboração própria (2026).

Observa-se que, mesmo as curvas sendo bem similares em intervalos específicos, inicialmente, entre as idades de 25 a 35 anos, a sobrevivência feminina tende a ser menor. Contudo, após a idade 35, a sobrevivência se comporta de forma igual ou minimamente superior à masculina, mantendo uma superioridade marginal na maior parte do horizonte temporal. Apesar disso, não houve evidência estatística de diferença entre as curvas de sobrevivência por sexo, dado que o p-valor do teste de log-rank foi igual a 0,42, superior ao nível de significância de 0,05, não rejeitando, portanto, a equivalência entre as curvas. O teste de log-rank é amplamente utilizado para verificar se há diferença estatisticamente significativa entre as curvas de sobrevivência entre grupos (Colosimo; Giolo, 2006).

A seguir, a Tabela 3 exibe as probabilidades de sobrevivência por sexo, para quatro idades, indicando que em idades específicas, há uma vantagem da sobrevivência feminina, mas ela é fraca e insuficiente para gerar significância estatística no teste global.

Tabela 3 – Resultados das probabilidades de sobrevivência idades específicas.

Idades	Masculino	Feminino
45	0.9184	0.9274
60	0.7408	0.7408
75	0.4145	0.4268
90	0.0559	0.0626

Fonte: Elaborado pelo autor (2026).

Apresentados os dados da estimação da sobrevivência pelo estimador de Kaplan-Meier, a seguir, apresentam-se os resultados da estimação do modelo de fragilidade aplicado para capturar a dependência.

4.3 ANÁLISE DO MODELO DE FRAGILIDADE

A análise da fragilidade foi realizada por meio da extensão do modelo de Cox, que possibilita a inclusão da fragilidade para cada casal (grupo de indivíduos). Inicialmente, aplicou-se o modelo de Cox sem a fragilidade, considerando apenas a covariável sexo. Assim, o risco foi modelado em função do sexo e assumindo independência entre o tempo de vida de cônjuges.

Tabela 4 - Resultados estimados para o modelo de COX sem a fragilidade.

Variável	β	Erro padrão	P-valor	Exp(β)
Sexo masc.	0,0391	0,0473	0,4089	1,0398

Fonte: Elaborado pelo autor (2026).

Observa-se na Tabela 4 que no primeiro ajuste do modelo de COX, o coeficiente estimado para o sexo masculino foi de 0,0391, resultando em uma razão de risco de 1,04. Isso sugere que os homens apresentam risco de óbito 4% superior ao das mulheres. No entanto, não houve evidência estatística significativa para essa diferença, dado que o P-valor do teste foi de 0,409, superior ao nível significância de 0,05. Isso indica que o fato de ser homem ou mulher não altera o risco de morte. Em relação ao intervalo de confiança (com 95% de confiança), observa-se que o valor 1 é incluído, corroborando a ausência de diferença significativa entre as diferenças de óbitos por sexo.

Depois, aplicou-se o modelo com a fragilidade compartilhada para cada casal. Ao incluir a fragilidade, aumentou-se o número de interações do modelo para 1000, conforme mencionado na seção 3.4 (página 32), de forma a garantir a convergência do modelo. Considerou-se novamente o sexo como uma covariável, além da fragilidade do grupo.

Os resultados estimados podem ser visualizados na Tabela 5, onde são apresentadas as informações da covariável considerada e da fragilidade compartilhada entre os grupos.

Tabela 5 - Resultados do modelo de Cox com fragilidade compartilhada (distribuição Gama).

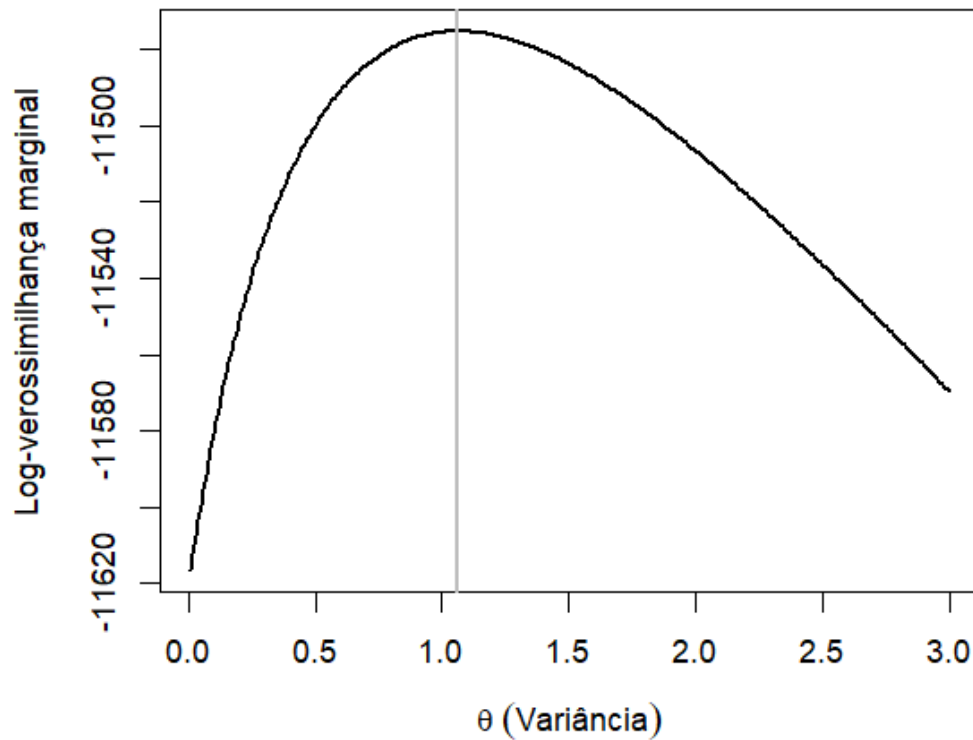
Efeitos	Parâmetros	Cox (Fragilidade (Gama))
Sexo (Masculino)	β	0,0307
	Erro P.	0,0540
	Exp (β)	1,031
	Valor-P	0,57
	θ (Variância)	1,0613
Fragilidade Compartilhada (id_casal)	Chisq (test. Wald)	3699.99
	P-valor	0,0000
	τ (Kendall's)	0,3464

Fonte: elaborado pelo autor (2026).

Os resultados para o modelo, com a inclusão da fragilidade compartilhada, indicam que a variável sexo continua sendo não significativa, dado que o Valor-P (0,57) foi superior ao nível de significância considerado (0,05). Contudo, a fragilidade compartilhada foi significativa (valor do teste de Wald = 3699.99, p-valor < 0,00001), enquanto o valor de θ (variância da fragilidade) estimada foi de 1,0613, como ela é superior a 0, o modelo sugere dependência relevante. Em outras palavras, a sobrevivência de homens e mulheres, isoladamente, não difere muito nesta base, entretanto, existe um componente comum de heterogeneidade que altera os seus riscos de morte.

Para verificar a convergência do modelo e descartar a possibilidade de um máximo local na log-verossimilhança marginal, foi construído o gráfico de perfil de verossimilhança para o parâmetro θ (variância da fragilidade). Na Figura 3, nota-se que o valor estimado para θ corresponde, de fato, ao máximo da função de log-verossimilhança marginal.

Figura 3 - Resultado da estimação do gráfico do perfil de verossimilhança de θ .



Fonte: elaborado pelo autor no *software* R (2026).

O gráfico de perfil de verossimilhança é uma ferramenta visual interessante para garantir que o algoritmo de otimização não “travou” em um pico secundário (máximo local) (Therneau; Grambsch, 2000). Em modelos de fragilidade, a função de verossimilhança pode ser malcomportada, e esse gráfico confirma a concavidade da função no ponto estimado.

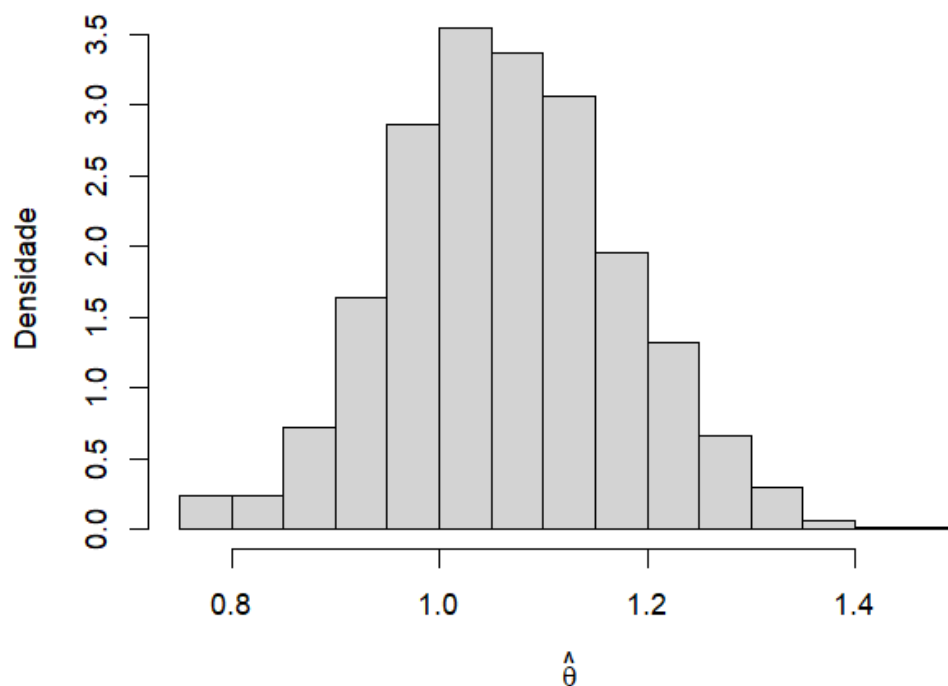
Obtidos os valores do modelo, o Teste da Razão de Verossimilhança (TRV) foi utilizado para testar a hipótese nula $H_0 : \theta = 0$ (ausência de correlação entre os tempos de vida do casal) contra a alternativa $H_1 : \theta \neq 0$ (presença de correlação). O teste resultou em um p-valor = 0 (estatística de teste = 292,1535). Portanto, há evidências estatísticas para rejeitar H_0 e admitir a existência de dependência entre os tempos de vida dos cônjuges. Essa evidência confirma os resultados de existência de dependência entre os tempos de vida obtidos por Frees, Carriere e Valdez (1996), Dufresne *et al.* (2018), Lee e Cha (2018), Henshaw, Constantinescu e Pamen (2020) e Henshaw *et al.* (2023), sendo, na maioria dos casos, para casais. Destaca-se também o estudo de Hougaard, Harvald e Holm (1992), que embora não trate especificadamente da sobrevivência de casais, evidenciou que fatores genéticos e ambientais compartilhados podem gerar associação entre os tempos de vida de gêmeos.

Dada a plausibilidade de dependência entre os tempos de vida, utilizou-se a métrica do coeficiente de Kendall para medir a associação, pois a variância da fragilidade é uma medida de heterogeneidade, e a dependência dentro dos grupos é uma consequência dessa heterogeneidade compartilhada, assim, o τ representa heterogeneidade em forma de essa dependência (Wienke, 2010). Como ele apresentou um valor superior a 0, igual a 0,3464, ou seja, existe dependência positiva entre os tempos de sobrevivência dos cônjuges. Contudo, a associação medida é induzida pela fragilidade compartilhada, e não uma dependência independente da heterogeneidade (Wienke, 2010).

Para fortalecer o resultado da hipótese de dependência entre o tempo de vida de casais, conforme mencionado na seção 3.5 (página 44), utilizou-se um teste simulado por meio de técnicas de bootstrap não paramétrico para construir uma distribuição amostral empírica para o estimador de θ e, então, utilizar essa distribuição para avaliar o quão significativamente diferente de zero pode ser θ à luz da amostra analisada.

Como se pode observar na Figura 4, valores maiores do que zero são mais prováveis de ocorrer na distribuição empírica. O intervalo de 95% de confiança também indica que valores entre 0,85 e 1,28 são plausíveis para θ . Deve-se notar, ainda que a própria estimativa obtida no modelo com fragilidade se enquadra dentro do intervalo.

Figura 4 - Resultado da distribuição empírica de θ



Fonte: Elaborado pelo autor (2026).

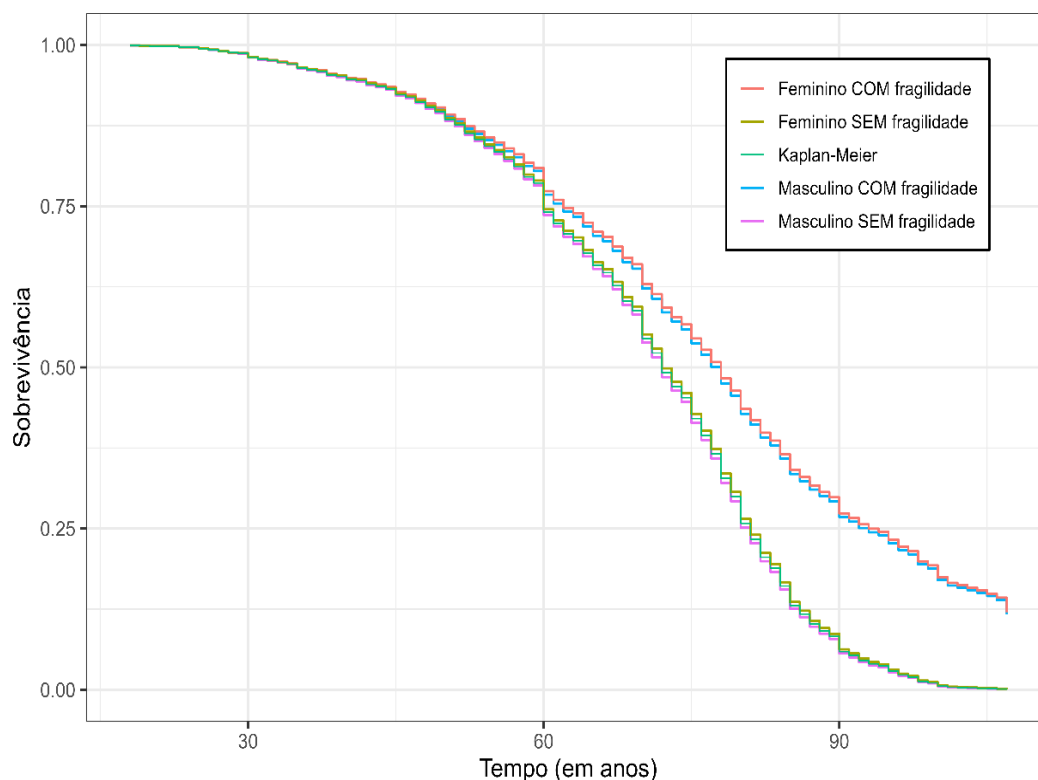
Com base nas evidências de validação do valor de θ e da hipótese de dependência entre os tempos residuais de vida, a próxima seção apresenta o impacto de se considerar a dependência no cálculo das anuidades em comparação com a hipótese comum de independência.

4.4 ANÁLISE COMPARATIVA ENTRE A SOBREVIVÊNCIA E OS VALORES DAS ANUIDADES

A partir dos métodos descritos no capítulo 3, aqui são apresentados os resultados para os cálculos das anuidades por meio do modelo com dependência. A sobrevivência com a fragilidade, com a inclusão da dependência, é obtida considerando que os indivíduos dentro de cada casal compartilham o efeito aleatório comum, o que gera a correlação entre as vidas.

Os resultados demonstraram um aumento da sobrevivência ao aplicar a fragilidade em comparação com a estimação sem a fragilidade, por meio do modelo de Cox e o estimador de Kaplan-Meier. Tal fenômeno, pode ser visualizado na Figura 5.

Figura 5 – Curvas de sobrevivência do modelo de Cox COM e SEM fragilidade e o estimador de Kaplan-Meier.



Fonte: Elaborado pelo autor

Observa-se que ao considerar a independência, o valor das probabilidades de sobrevivência é subvalorizado, logo, a fragilidade indica que a sobrevivência dependente, em idades mais avançadas, tende a ser superior. Nesse caso, casais mais velhos tendem a apresentar uma probabilidade de sobrevivência maior, quando a dependência é incluída, isso pode ser explicado pelo fato de os casais mais frágeis morrerem mais cedo e menos frágeis chegarem mais saudáveis às idades finais. Estes resultados condizem com os apresentados por Onchere *et al.* (2022) e por Walter *et al.* (2021). No entanto, esses trabalhos indicaram que a sobrevivência com a fragilidade tende a ser menor do que a probabilidade considerando a independência em idades menores. Esse fato contradiz os resultados atuais obtidos e visualizados na Tabela 6, e pode ser justificado pelo fato da distribuição escolhida, pois a fragilidade com a distribuição gama associa apenas a fragilidade de forma positiva, não permitindo incorporar um efeito negativo na sobrevivência (Onchere *et al.* 2021).

Conforme a sobrevivência, com o estimador de Kaplan-Meier e o modelo de Cox SEM e COM fragilidade, foi obtido o valor das anuidades. Como mencionado, as anuidades foram calculadas para o status de último sobrevivente, e, em relação às idades, consideraram-se as seguintes idades específicas para os segurados: 40, 50, 60, 70 e 80 anos para os homens; já para as mulheres, considerou-se uma diferença nas idades em relação aos homens, de 10, 5 e 0 anos, tanto maior quanto menor. Por exemplo, para um homem de 40 anos, calcularam-se 5 anuidades diferentes, com a idade da mulher assumindo os valores 30, 35, 40, 45 e 50 anos. Para fins de cálculo, considerou uma taxa de juros de 5% a.a. e uma idade de tempo de vida final w (idade em que todos estão mortos) de 110 anos. Em relação ao tipo de anuidade, estas foram calculadas de forma antecipada, imediata e vitalícia sob o status de último sobrevivente. Os resultados dos cálculos, inicialmente é apresentado a comparação entre o modelo de Cox e o estimador Kaplan-Meier, que são descritos na Tabela 6.

Tabela 6 - Resultados das anuidades, calculadas sob a hipótese de independência (modelo de Cox x KM).

Anuidades independentes – Prob. do modelo de Cox					
Idade da mulher	Idade dos homens (x)				
	40	50	60	70	80
-10	18,3168	16,8622	14,8546	12,3629	9,5054
-5	17,9405	16,2896	14,1075	11,3907	8,4941
0	17,5512	15,7425	13,4858	10,5516	7,9454
5	17,1766	15,2302	12,8367	9,8362	7,6448
10	16,8383	14,8210	12,3174	9,4595	7,3283
Anuidades independentes – Prob. do modelo de Cox					
Idade da mulher	Idade dos homens (x)				
	40	50	60	70	80

-10	18,3432	16,8632	14,8075	12,3854	9,5627
-5	17,9399	16,2923	14,0737	11,3808	8,5896
0	17,5495	15,7400	13,4824	10,5436	7,9366
5	17,1746	15,2441	12,8064	9,8310	7,6438
10	16,8319	14,8595	12,2823	9,3747	7,2866
Razão entre as anuidades (ind KM/ind)					
Idade da mulher	Idade dos homens (x)				
	40	50	60	70	80
-10	1,0014	1,0001	0,9968	1,0018	1,0060
-5	1,0000	1,0002	0,9976	0,9991	1,0112
0	0,9999	0,9998	0,9998	0,9992	0,9989
5	0,9999	1,0009	0,9976	0,9995	0,9999
10	0,9996	1,0026	0,9971	0,9910	0,9943

Fonte: Elaborado pelo autor com base em Gomes (2024) (2026).

Dentro das expectativas, observa-se que os resultados das anuidades com as probabilidades do modelo de Cox e do KM, com a hipótese de independência, foram praticamente idênticas, dado que os valores das razões são ligeiramente próximos de 1. Dada a comparação entre o modelo de Cox e Kaplan-Meier, parte-se ao caso principal de objetivo do estudo, a comparação entre o modelo de Cox e o modelo de fragilidade que considera a dependência. O Resultado da comparação é apresentado na Tabela 7.

Tabela 7 – Resultados das anuidades, calculadas sob as hipóteses de independência e dependência.

Anuidades independentes – Prob. do modelo de Cox					
Idade da mulher	Idade dos homens (x)				
	40	50	60	70	80
-10	18,3168	16,8622	14,8546	12,3629	9,5054
-5	17,9405	16,2896	14,1075	11,3907	8,4941
0	17,5512	15,7425	13,4858	10,5516	7,9454
5	17,1766	15,2302	12,8367	9,8362	7,6448
10	16,8383	14,8210	12,3174	9,4595	7,3283
Anuidades dependentes – Prob. do modelo de fragilidade					
Idade da mulher	Idade dos homens (x)				
	40	50	60	70	80
-10	18,6489	17,4212	15,8259	14,0805	12,3874
-5	18,3184	16,9285	15,2281	13,3872	11,7167
0	17,9736	16,4615	14,7608	12,7849	11,3175
5	17,6704	16,0789	14,3499	12,4364	11,3553
10	17,4060	15,8060	14,0564	12,3698	11,3839
Razão entre as anuidades (dep/ind)					
Idade da mulher	Idade dos homens (x)				
	40	50	60	70	80
-10	1,0181	1,0331	1,0654	1,1389	1,3032

-5	1,0211	1,0392	1,0794	1,1753	1,3794
0	1,0241	1,0457	1,0945	1,2116	1,4244
5	1,0287	1,0557	1,1179	1,2643	1,4854
10	1,0337	1,0665	1,1412	1,3077	1,5534

Fonte: Elaborado pelo autor com base em Gomes (2024) (2026).

O resultado da razão para todos os cenários de idades descritos foi superior a 1. Tal resultado indica uma subvalorização das anuidades ao assumir independência, o que sugere que, ao não considerar a dependência entre os casais, as seguradoras ou planos previdenciários cobrariam prêmios inferiores ao devido para cobrir suas obrigações futuras junto aos beneficiários, o que pode impactar a solvência das entidades. Em termos de impacto, o resultado indica que, ao considerar um homem e uma mulher, ambos aos 80 anos, a anuidade dependente é aproximadamente 42% superior a independente. Logo, assumindo a independência, o prêmio único puro cobrado seria bem menor, ao considerar um casal em idades extremas, mas resultaria em um prêmio insuficiente para fazer frente às obrigações futuras das entidades.

É importante observar que a subvalorização aumenta em idades mais avançadas, e, consequentemente, reservas futuras formadas abaixo das obrigações reais de seguradoras e de planos previdenciários. Em termos regulatórios atuariais, esses cenários podem indicar uma magnitude desse desvio poderia superar margens de solvência das instituições.

Em relação à literatura, entre os principais estudos, os estudos de Onchere (2022), Onchere *et al.* (2022), Walter *et al.* (2021) apontam que há uma subvalorização em idades elevadas. Porém, em idades iniciais, o valor das anuidades tende a ser supervalorizado sob a hipótese de independência. Dessa forma, o modelo de fragilidade implicaria em uma diminuição na responsabilidade esperada da seguradora para beneficiários em idades iniciais da anuidade, seguida por um aumento em idades mais elevadas, quando a dependência é considerada. Entretanto, o estudo aplica o modelo de fragilidade com função de risco na forma paramétrica e fragilidade seguindo uma distribuição positiva estável.

Outra observação importante é que a fragilidade tende a gerar um efeito maior à medida que a idade de um dos indivíduos aumenta, mesmo com a anuidade reduzindo. Por exemplo, ao selecionar um casal em que o homem possui 70 anos e a mulher, 60, a anuidade dependente é igual a 14,08, a independente é igual a 12,36 e o resultado da razão é 1,14. Todavia, ao aumentarmos a idade da mulher para 80, as anuidades caem, algo esperado dado o aumento do risco de mortalidade, e passam a valer, respectivamente, 12,37 (dependente) e 9,46 (independente), enquanto a razão aumenta para 1,31. Ao olhar para a diferença, em termos absolutos (dependente – independente), a diferença entre a anuidade dependente e a

independente é de aproximadamente 1,72. Quando a idade da mulher avança para 80 anos, essa diferença sobe para cerca de 2,91, mesmo com a redução do valor de ambas as anuidades. Essas variações sugerem que as implicações financeiras nos produtos atuariais, dado o envelhecimento, e pode gerar preocupação quanto à solvência das instituições

Ao olhar para a carteira de uma seguradora, tem-se que essa formada com casais, em que os homens e as mulheres têm em média 60 anos, e é pago um benefício de R\$ 1,00 por período. Ela constitui uma reserva; supondo independência, de R\$13,49 por casal. Mas o valor com a dependência, deveria ser R\$14,76. O déficit médio por casal seria de 1,27 unidade do benefício. Para uma carteira de 1.000 casais com benefício mensal de R\$ 1.000, isso representa cerca de R\$ 1,27 milhão de provisão insuficiente. Já com casais com idade média de 80/80 anos, o déficit por casal sobe para 3,37 unidades, ou seja, R\$ 3,37 milhões para a mesma carteira. Em um caso como esse, o déficit poderia impedir que a seguradora conseguisse arcar com as suas obrigações.

Em síntese, as implicações indicam que a dependência traz variações positivas em cálculos para múltiplas vidas e essas ocasionam resultados superiores para as rendas. Tais evidências reforçam a necessidade de se investigar e de se considerar estruturas de dependência na modelagem atuarial, especialmente em produtos destinados a casais.

5 CONSIDERAÇÕES FINAIS

O trabalho propôs uma análise utilizando modelos de fragilidade para analisar uma possível associação entre os tempos restantes de vida de casais e o seu impacto sobre a precificação de rendas para casais. A motivação para o estudo partiu de evidências presentes na literatura sobre a dependência entre as vidas, o que poderia gerar implicações nos preços de produtos atuariais, e a necessidade de revisão das premissas atuariais. Dessa forma, a partir de tal motivação, aplicou-se um modelo de fragilidade compartilhada com distribuição gama em dados reais coletados a partir do site do *FamilySearch*. Os resultados indicaram haver existência de dependência entre os tempos de vida dos cônjuges, sendo essa evidência significativa por meio dos testes e validações aplicadas. Ao incrementar os resultados do modelo ao cálculo atuarial, observam-se variações relevantes nas rendas, sendo superiores quando a dependência é aplicada.

Dada a evidência válida do modelo e as variações dos valores das anuidades, os resultados podem implicar uma possibilidade de revisão em premissas atuariais, visto que a desconsideração da dependência entre vidas pode levar à cobrança de prêmios e à formação de provisões matemáticas subavaliadas em relação ao risco real. Logo, a validação da dependência e de variações sobre a precificação implicam um risco à sustentabilidade do sistema de seguros e previdência ao se considerar coberturas para casais baseadas em modelos puros e independentes.

Como limitações, destaca-se a utilização de dados sem observações censuradas, o que limita uma análise a situações mais reais de mercado. Pontua-se também a utilização de um modelo não paramétrico, que pode não ser o mais adequado e confiável em comparação com um modelo paramétrico ajustado, dada a necessidade de inferências além da amostra, esperança de vida, suavização e extrapolação. Devido a esses fatores, o modelo paramétrico pode trazer maiores ganhos de precisão (Lawless, 2011; Colosimo; Giolo, 2006). Outra desvantagem a ser considerada seria a utilização da distribuição gama, que é invariante no tempo, conforme mencionado por Onchere (2022). Além disso, a estrutura resultante dessa distribuição admite basicamente associação positiva entre tempos de falha, não evidenciando, portanto, restrições formais a associações negativas.

No entanto, haja vista resultados significativos sobre a dependência entre os casais e a possibilidade de um avanço metodológico, que sugere a alteração na forma de cálculos atuariais para produtos do ramo vida, pesquisas futuras podem avaliar as diferentes distribuições de fragilidades, como em um processo composto com a Estável Positiva (*Positive Stable*), que,

conforme os resultados obtidos por Onchere *et al.* (2022), possibilitam capturar os efeitos da heterogeneidade que variam com o tempo, associados de forma positiva e negativa à dependência. A inclusão desse efeito permite capturar o aumento do risco de morte, dada a morte do cônjuge, efeito não capturado pela metodologia aqui aplicada, dado que a distribuição gama considera o impacto de associação apenas de forma positiva. Sugere-se também uma análise entre modelos paramétricos e semiparamétricos e a comparação de diferentes distribuições para fragilidade, como a log-normal e a inversa gaussiana, que poderiam ser comparadas com a gama em um modelo de fragilidade compartilhada. Em relação a outros modelos, essa metodologia poderia ser comparada também com o modelo de cópulas.

Além disso, recomenda-se analisar o impacto da dependência em outros produtos atuariais do ramo vida, tais como os seguros de vida, as anuidades para o status de vida conjunta e as anuidades reversíveis. Para além da análise dos tempos residuais de vida de casais, sugere-se a extensão da metodologia para seguros coletivos contratados por empresas para os seus funcionários. Nesses contextos, pode-se observar que dentro de determinadas empresas é possível haver diferentes polos com setores independentes, e dentro de cada um deles os funcionários podem estar expostos a riscos semelhantes e que podem ocasionar dependência entre os eventos.

Por fim, espera-se que o estudo contribua metodologicamente como base para novas pesquisas que investiguem a dependência entre vidas no meio atuarial e na literatura nacional, ainda pouco explorada no Brasil. Dessa forma, novas evidências sobre a dependência entre vidas e variações em produtos atuariais podem indicar a necessidade de revisão das hipóteses subjacentes às práticas de precificação comum de mercado e promover um maior rigor técnico e sustentabilidade nos setores de seguros e previdência.

REFERÊNCIAS

- ALAI, D. H.; LANDSMAN, Z.; SHERRIS, M. Lifetime dependence modelling using a truncated multivariate gamma distribution. **Insurance: Mathematics and Economics**, v. 52, n. 3, p. 542–549, 2013. Disponível em: <https://doi.org/10.1016/j.insmatheco.2013.03.011>. Acesso em: 15 de ago. 2025.
- AZEVEDO, Gustavo H. W. de. **Matemática financeira e atuarial**: noções aplicadas ao seguro. Rio de Janeiro, RJ: FUNENSEG, 2005.
- BALAN, Theodor A.; PUTTER, Hein. A tutorial on frailty models. **Statistical Methods in Medical Research**, v. 29, n. 11, p. 3424–3454, 2020. Disponível em: <https://journals.sagepub.com/doi/pdf/10.1177/0962280220921889>. Acesso em: 02 de ago. 2025.
- BEARD, R. E. Appendix: Note on Some Mathematical Mortality Models. **Novartis Foundation symposium**, p. 302–311, 1959. Disponível em: <https://doi.org/10.1002/9780470715253.app1>. Acesso em: 08 de ago. 2025.
- BEDIA, E. C. **Modelos de sobrevivência bivariados induzidos por fragilidade**. 2022. 119 f. Tese (Doutorado em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2022.
- BORGES, Alexandra Isabel Monteiro. **Análise de Sobrevivência com o R**. 2014. Dissertação (Mestrado) - Curso de Matemática, Universidade da Madeira, Funchal, 2014.
- BOWERS, N. L. et al. **Actuarial mathematics**. 2. ed. Schaumburg: Society of Actuaries, 1997.
- BRASIL. Superintendência de Seguros Privados (SUSEP). **Plano de Regulação para 2025**. Rio de Janeiro: SUSEP, 2025.
- CABRIGNAC, O.; CHARPENTIER, A.; GALLIC, E. **Modeling Joint Lives within Families**. arXiv (Cornell University), 1 jan. 2020. Disponível em: <https://arxiv.org/abs/2006.08446>. Acesso em: 15 ago. 2025.
- CALSAVARA, Vinícius Fernando. **Modelos de sobrevivência com frações de cura usando um termo de fragilidade e tempo de vida Weibull modificados generalizadamente**. 2011. 72 f. Dissertação (Mestrado em Estatística) – Universidade Federal de São Carlos, São Carlos, 2011.
- CARANNANTE, M. et al. Lee-Carter family of frailty-based stochastic mortality models. **Quality and Quantity**, v. 58, n. 6, p. 5081–5105, 2024. Disponível em: <https://doi.org/10.1057/s41288-024-00319-y>. Acesso em: 19 jul. 2025.
- CARANNANTE, M. et al. Frailty-based Lee–Carter family of stochastic mortality models. **Quality & Quantity**, 2023. Disponível em: <https://doi.org/10.1007/s11135-023-01786-6>. Acesso em: 10 jul. 2025.

CARVALHO, M. S. et al. **Análise de sobrevivência: teoria e aplicações em saúde**. [s.l.] Editora Fiocruz, 2011.

COLOSIMO, Enrico Antônio; GIOLO, Suely Ruiz. **Análise de sobrevivência aplicada**. São Paulo: Editora da Universidade de São Paulo, 2006. 392 p.

CORDEIRO FILHO, Antônio. **Cálculo Atuarial Aplicado: teoria e aplicações, exercícios resolvidos e propostos**. São Paulo: Atlas, 2009.

COX, D. R. Regression models and lifetables. **Journal of the Royal Statistical Society: Series B (Methodological)**, London, v. 34, n. 2, p. 187–220, 1972. Disponível em: <https://academic.oup.com/jrssb/article/34/2/187/7027194>. Acesso em: 30 jul. 2025.

DENUIT, Michel et al. Measuring the impact of dependence among insured lifelengths. **Belgian Actuarial Bulletin**, v. 1, n. 1, p. 18-39, 2000. Disponível em: <https://hdl.handle.net/11245/1.184531>. Acesso em: 30 jul. 2025.

DICKSON, D. C. M.; HARDY, M. R.; WATERS, H. R. **Actuarial Mathematics for Life Contingent Risks**. 2. ed. Cambridge: Cambridge University Press, 2013. p. 45-60.

DUFRESNE, F. et al. On age difference in joint lifetime modelling with life insurance annuity applications. *Annals of Actuarial Science*, v. 12, n. 2, p. 350–371, 3 abr. 2018. Disponível em: <https://doi.org/10.1017/S1748499518000076>. Acesso em: 08 out. 2025.

FAMILYSEARCH. **Sobre o FamilySearch**. Salt Lake City: Igreja de Jesus Cristo dos Santos dos Últimos Dias, 2025. Disponível em: <https://www.familysearch.org/pt/about>. Acesso em: 27 dez. 2025.

FREES, E. W.; CARRIERE, J.; VALDEZ, E. Annuity valuation with dependent mortality. **The Journal of Risk and Insurance**, v. 63, n. 2, p. 229-261, jun. 1996. JSTOR. Disponível em: <http://dx.doi.org/10.2307/253744>. Acesso em: 15 ago. 2025.

FROMM, D. **A indústria da proteção: sobre as interfaces entre seguro, segurança e seguridade**. 2022. 276 p. Tese (Doutorado em Antropologia Social) – Universidade Estadual de Campinas, Instituto de Filosofia e Ciências Humanas. Campinas, 2022.

FITZMAURICE, G. M.; LAIRD, N. M.; WARE, J. H. **Applied longitudinal analysis**. Hoboken, N.J.: Wiley, 2011.

GIL, Antônio Carlos. **Como elaborar projetos de pesquisa**. 6. ed. São Paulo: Atlas, 2017.

GIOVANNI, D. D.; PIRRA, M.; VIVIANO, F. Joint mortality models based on subordinated linear hypercubes. **ASTIN Bulletin**, v. 55, n. 2, p. 1-20, 2025. Disponível em: https://ideas.repec.org/a/cup/astinb/v55y2025i2p332-351_6.html. Acesso em: 15 ago. 2025.

GORFINE, Malka; ZUCKER, David. **Shared frailty methods for complex survival data: a review of recent advances**. 2022. Disponível em: <https://www.annualreviews.org/content/journals/10.1146/annurev-statistics-032921-021310>. Acesso em: 15 ago. 2025.

GOMES, H. V. S. F. **Teoria das múltiplas vidas: uma abordagem por dependência**. 2024. 78 f. Trabalho de Conclusão de Curso (Graduação em Ciências Atuariais) – Universidade Federal da Paraíba, João Pessoa, 2024.

GOURIEROUX, C.; LU, Y. Love and death: A Freund model with frailty. **Insurance: Mathematics and Economics**, v. 63, p. 191–203, jul. 2015. Disponível em: <https://doi.org/10.1016/j.insmatheco.2015.03.016>. Acesso em: 12 ago. 2025.

GUTIERREZ, Roberto G. Frailty in survival analysis models (parametric frailty, parametric shared frailty, and frailty in Cox models). **ideas.repec.org**, (2001). Disponível em: <https://ideas.repec.org/p/boc/usug01/17.html>. Acesso em: 22 ago. 2025.

HANAGAL, D. D.; SHARMA, R. Modeling heterogeneity for bivariate survival data by shared gamma frailty regression model. **Model Assisted Statistics and Applications**, v. 8, n. 2, p. 85–102, 15 Maio 2013. Disponível em: <https://doi.org/10.3233/MAS-130259>. Acesso em: 20 ago. 2025.

HENSHAW, K.; CONSTANTINESCU, C.; PAMEN, O. M. Stochastic mortality modelling for dependent coupled lives. **Risks**, v. 8, n. 1, p. 1–28, 2020. Disponível em: <https://doi.org/10.3390/risks8010017>. Acesso em: 12 jul. 2025.

HOGG, Robert V.; TANIS, Elliot A.; ZIMMERMAN, Dale L. **Probability and Statistical Inference**. 9. ed. Boston: Pearson Education, 2015.

HOUGAARD, P. **Analysis of Multivariate Survival Data**. [s.l.] Springer Nature, 2000.

HOUGAARD, P.; HARVALD, B.; HOLM, N. V. Measuring the Similarities between the Lifetimes of Adult Danish Twins Born between 1881–1930. **Journal of the American Statistical Association**, v. 87, n. 417, p. 17–24, 1 mar. 1992. Disponível em: <https://doi.org/10.1080/01621459.1992.10475170>. Acesso em: 15 ago. 2025.

HONORÉ, B. Identification results for duration models with multiple spells. **Review of Economic Studies**, v. 60, p. 241–246, 1993. Disponível em: <https://ideas.repec.org/a/oup/restud/v60y1993i1p241-246.html>. Acesso em: 01/12/2025.

HARRINGTON, S. E.; NIEHAUS, G. **Risk management and insurance**. McGraw-Hill/Irwin, 2004.

JAGGER, C.; SUTTON, C. J. Death after marital bereavement-is the risk increased? **Statistics In Medicine**, [S.L.], v. 10, n. 3, p. 395–404, mar. 1991. Wiley. Disponível em: <http://dx.doi.org/10.1002/sim.4780100311>. Acesso em: 30 jun. 2025.

JEVTIĆ, P.; HURD, T. R. The joint mortality of couples in continuous time. **Insurance: Mathematics and Economics**, v. 75, p. 90–97, jul. 2017. Disponível em: <https://ideas.repec.org/a/eee/insuma/v75y2017icp90-97.html>. Acesso em: 15 set. 2025.

KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. **Journal of the American Statistical Association**, v. 53, p. 457 – 481, 1958. Disponível em: <https://doi.org/10.1080/01621459.1958.10501452>. Acesso em: 10 ago. 2025.

LAWLESS, Jerald F. **Statistical Models and Methods for Lifetime Data**. 2ª ed. Hoboken: John Wiley & Sons, Inc., 2003.

LEÃO, J., LEIVA, V., SAULO, H., & TOMAZELLA, V. Birnbaum–Saunders frailty regression models: Diagnostics and application to medical data. **Biometrical Journal**, 59(2), 291-314. 2017.

LEE, H.; CHA, J. H. A dynamic bivariate common shock model with cumulative effect and its actuarial application. **Scandinavian Actuarial Journal**, [S.L.], v. 2018, n. 10, p. 890-906, 11 maio 2018. Disponível em: <http://dx.doi.org/10.1080/03461238.2018.1470562>. Acesso em: 23 jun. 2025.

LIYANAGE, J. C.; KARUNARATHNA, G. H. S. Shared Frailty model for joint survival data - A simulation study. **Sri Lankan journal of Applied Statistics**, v. 23, n. 2, p. 105–120, 2022.

LU, Y. Bivariate survival analysis with latent factors: theory and applications to mortality and long-term care. 2015. 150 f. Tese (Doutorado) - Université Paris Dauphine, Paris, 2015. Disponível em: <https://theses.hal.science/tel-01212470v1/document>. Acesso em: 10 ago. 2025.

LU, Y. Broken-heart, common life, heterogeneity: analyzing the spousal mortality dependence. *Astin Bulletin*, [S.L.], v. 47, n. 3, p. 837-874, 20 abr. 2017. Cambridge University Press (CUP). <http://dx.doi.org/10.1017/asb.2017.8>.

LUCIANO, Elisa; SPREEUW, Jaap; VIGNA, Elena. Modelling stochastic mortality for dependent lives. **Insurance: Mathematics and Economics**, v. 43, n. 2, p. 234–244, 2 fev. 2008. Disponível em: <http://dx.doi.org/10.1016/j.insmatheco.2008.06.005>. Acesso em: 15 ago. 2025.

NASCIMENTO, Francisco Paulo; SOUSA, Flávio Luís Leite. **Metodologia da pesquisa científica teoria e prática: como elaborar TCC**. 2. ed. Fortaleza: INESP, 2017.

NELSEN, R. B. **An introduction to copulas**. 2. ed. New York: Springer, 2006. Disponível em: <https://link.springer.com/book/10.1007/0-387-28678-0>. Acesso em: 01 ago. 2025.

ONCHERE, W. O. et al. Graduation of term assurance data using frailty approach, **Afrika Statistika**, v. 16, n. 3, p. 2805–2817, 2021.

OAKES, D. A model for association in bivariate survival data. *Journal of the Royal Statistical Society: Series B*, v. 44, n. 3, p. 414-422, 1982. Disponível em: <https://www.scirp.org/reference/referencespapers?referenceid=1787114>. Acesso em: 16 Jul. 2025.

ONCHERE, W. O. **Shared compound frailty model with application in joint life annuity insurance**. 2022. 154 p. Tese (Doutorado em Ciências Atuariais) – Department of Mathematics, University of Nairobi, Nairobi, 2022.

R CORE TEAM. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 2024. Disponível em: <https://www.R-project.org/>.

ROCHA, Cristina; PAPOILA, Ana Luísa. **Análise de sobrevivência**. Lisboa: Instituto Nacional de Estatística; Sociedade Portuguesa de Estatística, 2009.

RODRIGUES, José Ângelo. **Gestão de risco atuarial**. São Paulo: Atlas, 2008.

ROSS, Sheldon Mark. **A first course in probability**. 9. ed. Boston: Pearson, 2014.

SANTOS JÚNIOR, Luiz Carlos. **Análise de Sobrevivência Aplicada a premissas atuariais: o caso da previdência pública municipal de Cabedelo/PB**. 2018, 185 f. Tese (Doutorado em Biometria) — Instituto de Biociências, UNESP, Botucatu, SP, 2018.

SINDELAR, Fernanda Cristina Wiebusch; DE CONTO, Samuel Martim; AHLERT, Lucildo. **Teoria e prática em estatística para cursos de graduação**. 1. ed. Lajeado: Editora Univates, 2014.

SPREEUW, J.; OWADALLY, I. Investigating the broken-heart effect: a model for short-term dependence between the remaining lifetimes of joint lives. **Annals of Actuarial Science**, v. 8, n. 1, p. 48-69, 2014. Disponível em: <https://doi.org/10.1017/S1748499512000292>. Acesso em: 15 ago. 2025.

SPREEUW, Jaap; WANG, Xu. Modelling the short-term dependence between two remaining lifetimes. **Cass Business School Discussion Paper**, v. 2, n. 3, 2008.

THERNEAU, T. M.; GRAMBSCH, P. M. **Modeling Survival Data: Extending the Cox Model**. New York, NY: Springer New York, 2000.

THERNEAU, T. M. **A Package for Survival Analysis in R**. R package version 3.5-8, 2024. Disponível em: <https://CRAN.R-project.org/package=survival>. Acesso em: 2 fev. 2026.

VAUPEL, J. W.; MANTON, K. G.; STALLARD, E. The impact of heterogeneity in individual frailty on the dynamics of mortality. **Demography**, v. 16, n. 3, p. 439–54, ago. 1979.

VILANOVA, Wilson. **Matemática atuarial**: destinado aos cursos de ciências econômicas, contábeis e atuariais. Liv. Pioneira, Ed. da Universidade, 1969.

WALTER, O. et al. Positive Stable Frailty Approach in the Construction of Dependence Life-Tables. **Open Journal of Statistics**, v. 11, n. 04, p. 506–523, 2021. Disponível em: <https://www.scirp.org/journal/paperinforcitation?paperid=111296>. Acesso em: 10 set. 2025.

WIENKE, A. **Frailty models in survival analysis**. Boca Raton: Chapman & Hall/CRC, 2010.

WINDHOLZ, M. J.; MARMAR, C. R.; HOROWITZ, M. J. A review of the research on conjugal bereavement: Impact on health and efficacy of intervention. **Comprehensive Psychiatry**, v. 26, n. 5, p. 15, set. 1985. Disponível em: [https://doi.org/10.1016/0010-440X\(85\)90080-X](https://doi.org/10.1016/0010-440X(85)90080-X). Acesso em: 15 Jul. 2025.

ZIVOT, Eric. **Introduction to Computational Finance and Financial Econometrics with R**. Seattle: University of Washington, 2021. Disponível em: <https://bookdown.org/compfinezbook/introcompfinr/>. Acesso em: 20 fev. 2026.

APÊNDICES

APÊNDICE A – Código

```

rm(list=ls()) # Limpando diretório
# Instalação de pacotes
# Instalação de pacotes
# install.packages("tidyverse") # Manipulação e visualização de dados
# install.packages("readxl")    # Leitura de arquivos em Excel
# install.packages("survival")   # Análise de sobrevivência
# install.packages("flexsurv")   # Modelos paramétricos flexíveis de sobrevivência
# install.packages("survminer") # Visualização de curvas de sobrevivência
# Carregando pacotes necessarios
library(tidyverse)
library(survival)
library(flexsurv)
library(dplyr)
library(survminer)
library(readxl)
##### Tratamento dos dados #####
# importando dados coletados
dados <- read_excel("Estudos/TCC - TASSIANO/Base de dados - TCC.xlsx",
                    sheet = "Planilha1")
head(dados)

# Removendo coluna
dados$`Links de acesso` = NULL
print(dados)

# Ajustando o nome das colunas
# Renomeando colunas e criando colunas de tempos, status e tempo de casam. até a morte.
dados <-
  dados |>
  transmute(id_casal = id_grupo,
            id_ind = Id_ind,
            sexo = Sexo,
            ano_nasc = `ano de nascimento`,
            ano_morte = `ano de morte`,
            ano_casam = `data de casamento`,
            tempo = ano_morte - ano_nasc,
            status = 1, # Coluna de informação sobre o status.
            tempo_casam_ate_morte_ind = ano_morte - ano_casam
  )

# Tornando os dados anônimos.
head(dados)
dados <- dados %>%
  mutate(id_ind = sprintf("ID_%04d", row_number()))

# Visualizar
print(dados)

```

```

# Transformando as colunas para o tipo n merico.
dados <- dados %>% mutate(across(c(ano_nasc, ano_morte, ano_casam,
tempo_casam_ate_morte_ind), as.numeric))
# Verificando inconsist ncia nos dados.

# casais com ano de casamento diferente.
casais_dif <- dados |>
  group_by(id_casal) |>
  summarise(ind_divergencia_ano_casamento = if_else(length(unique(ano_casam)) > 1,
"Sim", "N o")) |>
  filter(ind_divergencia_ano_casamento == "Sim")

# retirando colunas
dados <- dados |>
  group_by(id_casal) |>
  filter(n_distinct(ano_casam) == 1) |>
  ungroup()

# Verif. casais com ano de morte inferior ao ano de casamento e
# com morte no ano do casamento.
casais_exc <- dados |>
  filter(tempo_casam_ate_morte_ind <= 0) |>
  distinct(id_casal) |>
  mutate(ind_inconsistencia = "Sim")

# Retirando os casais com em os indiv duos apresentem inconsist ncia.
dados <- dados |>
  filter(!id_casal %in% casais_exc$id_casal)

# identificando se h  valores ausentes na base
nrow(model.frame(Surv(tempo, status) ~ sexo, data = dados))
colSums(is.na(dados[, c("tempo", "status", "sexo")]))

#####  AN LISE DESCRITIVA DOS DADOS  #####

# Estat sticas gerais

cat("Total de indiv duos:", nrow(dados), "\n")
cat("Total de casais:", length(unique(dados$id_casal)), "\n")

summary_tempo <- summary(dados$tempo)
print(summary_tempo)
cat("Desvio-padr o:", round(sd(dados$tempo, na.rm = TRUE), 2), "\n")
cat("Vari ncia:", round(var(dados$tempo, na.rm = TRUE), 2), "\n")
cat("Coef. de Varia  o:", round(sd(dados$tempo, na.rm = TRUE)/mean(dados$tempo, na.rm
= TRUE)*100, 2), "%\n\n")

# An lise dos tempos de sobreviv ncia por sexo
descricao_sexo <- dados %>%

```

```

group_by(sexo) %>%
summarise(
  N = n(),
  Ida_Média = round(mean(tempo, na.rm = TRUE), 2),
  Ida_Mediana = round(median(tempo, na.rm = TRUE), 2),
  DP = round(sd(tempo, na.rm = TRUE), 2),
  Mínimo = min(tempo),
  Máximo = max(tempo, na.rm = TRUE)
)
descricao_total <- dados %>%
summarise(
  sexo = "Total",
  N = n(),
  Ida_Média = round(mean(tempo, na.rm = TRUE), 2),
  Ida_Mediana = round(median(tempo, na.rm = TRUE), 2),
  DP = round(sd(tempo, na.rm = TRUE), 2),
  Mínimo = min(tempo, na.rm = TRUE),
  Máximo = max(tempo, na.rm = TRUE)
)

descricao_final <- bind_rows(descricao_sexo, descricao_total)

print(descricao_final)

# Gráficos

# Densidade por sexo
ggplot(dados, aes(x = tempo, fill = sexo)) +
  geom_density(alpha = 0.5) +
  scale_fill_manual(values = c("red", "green")) +
  labs(x = "Idade (anos)", y = "densidade") +
  theme_minimal()
ggplot2::ggsave(
  filename = "Dens_óbito.png",
  plot = last_plot(),
  width = 16,
  height = 12,
  units = "cm",
  dpi = 1000
)
##### Estimação da sobrevivência #####

# Kaplan-Meier

obj_surv <- Surv(time = dados$tempo,
                 event = dados$status)
# Curvas de sobrevivência por sexo
km_sexo <- survfit(obj_surv ~ sexo, data = dados)
summary(km_sexo)
str(km_sexo)

```

```

# Visualizando as curvas de sobrevivência
ggsurvplot(
  km_sexo,
  data = dados,
  pval = TRUE,          # mostrar p-valor do teste log-rank
  # conf.int = TRUE,    # intervalo de confiança
  xlab = "Idade (anos)",
  ylab = "Probabilidade de sobrevivência",
  ggtheme = theme_minimal(base_size = 10)
)

# Testando se há diferença estatística entre as curvas de sobrevivência
survdifff(obj_surv ~ sexo, data = dados)

# Probabilidade por sexo para diferentes idades.
# Criando Km com tempos separados
dados_m <- subset(dados, sexo == "Masculino") # Separando dados do sexo masculino
dados_f <- subset(dados, sexo == "Feminino")  # Separando dados do sexo feminino

km_m <- survfit(Surv(tempo, status) ~ 1, data = dados_m)
km_f <- survfit(Surv(tempo, status) ~ 1, data = dados_f)

idades_ekm <- c(45, 60, 75, 90)

Sobre_m <- summary(km_m, times = idades_ekm)
Sobre_f <- summary(km_f, times = idades_ekm)

tabela <- data.frame(
  idades = idades_ekm,
  Masculino = round(Sobre_m$surv, 4),
  Feminino = round(Sobre_f$surv, 4)
)

tabela

##### Estimando o modelo de fragilidade para os grupos (casais) #####

# Modelo de Cox sem fragilidade
(fit_cox_sem_fragilidade <- coxph(Surv(tempo, status) ~ sexo, data = dados))

(fit_cox_com_fragilidade <- coxph(Surv(tempo, status) ~ sexo + frailty(id_casal, distribution
= "gamma"),
                                data = dados,
                                control = list(iter.max = 1000)))
summary(fit_cox_com_fragilidade)

# Perfil de verossimilhança de theta
# 1. Definir uma sequência de valores para theta (variância da fragilidade)

```

```

theta_vals <- seq(0.01, 3, length.out = 300)
loglik_vals <- numeric(length(theta_vals))

# 2. Loop para ajustar o modelo fixando cada valor de theta
for (i in seq_along(theta_vals)) {
  # O argumento theta= fixado dentro de frailty() desativa a estimação automática
  m <- coxph(Surv(tempo, status) ~ sexo + frailty(id_casal, distribution = "gamma", theta =
theta_vals[i]),
    data = dados,
    control = list(iter.max = 1000))

  # Extrair a log-verossimilhança parcial (segundo elemento do vetor loglik)
  loglik_vals[i] <- m$history$`frailty(id_casal, distribution = "gamma", theta =
theta_vals[i])`$c.loglik
}

# 3. Construir o gráfico de perfil
plot(theta_vals, loglik_vals, type = "l", col = "black", lwd = 2,
  xlab = expression(theta ~ (Variância)),
  ylab = "Log-verossimilhança marginal", # I-likelihood (disponível apenas para fragilidade
gama)
  main = "Perfil de Verossimilhança")

# Adicionar linha vertical no MLE (valor que o R estimaria automaticamente)
abline(v = fit_cox_com_fragilidade$history$`frailty(id_casal, distribution = "gamma"`$theta,
col = "grey", lwd = 2, pch = 3)

# Teste da razão de Verossimilhança

est_teste <- 2 * (fit_cox_com_fragilidade$history$`frailty(id_casal, distribution =
"gamma"`$c.loglik - fit_cox_sem_fragilidade$loglik[2])
p = pchisq(est_teste, df = 1, lower.tail = F)

# Teste do BooTstrap não paramétrico para theta

get_theta <- function(data, indices) {

  # Reamostrar e renomear clusters em uma única etapa
  d_boot <- do.call(rbind, lapply(seq_along(indices), function(i) {
    temp <- data[data$id_casal == unique(data$id_casal)[indices[i]], ]
    temp$id_casal <- i # usa a posição diretamente como novo id
    temp
  })))

  fit <- try(
    coxph(Surv(tempo, status) ~ sexo + frailty(id_casal, distribution = "gamma"),
      data = d_boot, control = list(iter.max = 1000)),
    silent = TRUE
  )

  if (inherits(fit, "try-error")) return(NA)

```

```

fit$history[[1]]$theta
}
# 2. Executar o Bootstrap (ex: 1000 replicações)
set.seed(123)
n_casais <- length(unique(dados$id_casal))
boot_results <- replicate(1000, get_theta(dados, sample(1:n_casais, replace = TRUE)))

# Remover NAs de falhas de convergência
boot_results <- boot_results[!is.na(boot_results)]

# Visualizar a distribuição de theta
hist(boot_results, breaks=25, main="Distribuição empírica",
     xlab = expression(hat(theta)), probability = T, ylab = "Densidade")

# Intervalo de Confiança de 95% (Percentil)
quantile(boot_results, probs = c(0.025, 0.975))

##### Análise da sobrevivência com a fragilidade #####
(fit_cox_nulo <- survfit(coxph(Surv(tempo, status) ~ sexo,
                             data = dados)))
# Extrair o beta (coeficiente do sexo) e o theta (variância da fragilidade)
beta_com_fragilidade <- fit_cox_com_fragilidade$coefficients # coef - sexo
theta <- fit_cox_com_fragilidade$history$frailty(id_casal, distribution = "gamma")$theta

df <- dplyr::tibble(risco_acumulado_base = basehaz(fit_cox_com_fragilidade)$shazard,
                   tempo = fit_cox_nulo$time,
                   s_pop_m = (1 + theta * risco_acumulado_base * exp(beta_com_fragilidade))^(
1/theta),
                   s_pop_f = (1 + theta * risco_acumulado_base)^(-1/theta),
                   s_f = summary(survfit(fit_cox_sem_fragilidade, newdata = data.frame(sexo =
"Feminino"))), times = tempo)$surv,
                   s_m = summary(survfit(fit_cox_sem_fragilidade, newdata = data.frame(sexo =
"Masculino"))), times = tempo)$surv)
df$s_pop_m
km <- survfit(obj_surv ~ 1, data = dados)
df %>%
  ggplot2::ggplot(ggplot2::aes(x = tempo, y = s_pop_m)) +
  ggplot2::geom_step(ggplot2::aes(colour = "Masculino COM fragilidade")) +
  ggplot2::geom_step(ggplot2::aes(y = s_pop_f, colour = "Feminino COM fragilidade")) +
  ggplot2::geom_step(ggplot2::aes(y = s_m, colour = "Masculino SEM fragilidade")) +
  ggplot2::geom_step(ggplot2::aes(y = s_f, colour = "Feminino SEM fragilidade")) +
  ggplot2::geom_step(
    data = data.frame(tempo = km$time, surv = km$surv),
    ggplot2::aes(x = tempo, y = surv, colour = "Kaplan-Meier"),
    linewidth = 0.3
  ) +
  ggplot2::scale_y_continuous(limits = c(0, 1)) +
  ggplot2::labs(x = "Tempo (em anos)", y = "Sobrevivência", colour = NULL) +
  ggplot2::theme_bw() +

```

```

ggplot2::theme(
  legend.position = c(0.8, 0.8),
  legend.background = ggplot2::element_rect(fill = "white", colour = "black"),
  legend.key.size = ggplot2::unit(0.5, "cm"),
  legend.text = ggplot2::element_text(size = 8)
)
ggplot2::ggsave(
  filename = "curvas_sobrevivencia.png",
  plot = last_plot(),
  width = 16,
  height = 12,
  units = "cm",
  dpi = 1000
)

```

Cálculo das anuidades

```

w <- 110 # última idade das
i <- 0.05 # Taxa de Juros
v <- 1/(1+i)
x <- c(40, 50, 60, 70, 80)
nome_linhas <- c("-10", "-5", "0", "5", "10")
nome_colunas <- c("40", "50", "60", "70", "80")

```

Auidades - Modelo de COX x KM

```

a_ind <- as.data.frame(matrix(nrow = 5, ncol = 5), row.names = nome_linhas)
names(a_ind) <- nome_colunas
a_ind_km <- a_ind
a_raz <- a_ind_km
Sobre_m <- summary(km_m)
Sobre_f <- summary(km_f)
for (j in 1:ncol(a_ind)) {
  x <- as.numeric(names(a_ind)[j])
  for (k in 1:nrow(a_ind)) {
    y <- x + as.numeric(row.names(a_ind)[k])
    t = 0:(max(w - x, w - y))
    # COX
    px <- approx(df$stempo, df$s_m, xout = x, rule = 2)$y
    py <- approx(df$stempo, df$s_f, xout = y, rule = 2)$y
    #Probabilidades condicionais
    t_px_ind <- approx(df$stempo, df$s_m, xout = x + t, rule = 2)$y / px
    t_py_ind <- approx(df$stempo, df$s_f, xout = y + t, rule = 2)$y / py
    t_pxyus_ind <- t_px_ind + t_py_ind - (t_px_ind * t_py_ind)
    ind <- sum((v^t) * t_pxyus_ind )
    # Km
    Sx <- approx(Sobre_m$time, Sobre_m$surv, x, rule = 2)$y
    Sy <- approx(Sobre_f$time, Sobre_f$surv, y, rule = 2)$y

    t_Sx <- approx(Sobre_m$time, Sobre_m$surv, x + t, rule = 2)$y
    t_Sy <- approx(Sobre_f$time, Sobre_f$surv, y + t, rule = 2)$y
  }
}

```

```

t_px_km <- t_Sx/ Sx
t_py_km <- t_Sy/ Sy
t_pxyus_km <- t_px_km + t_py_km - (t_px_km*t_py_km)

ind_km <- sum((v^t) * t_pxyus_km)
a_ind_km[k, j] <- ind_km
a_ind[k, j] <- ind
a_raz[k, j] <- ind_km / ind
}
}
round(a_ind, 4)
round(a_ind_km, 4)
round(a_raz, 4)
##### Modelo de Cox SEM x Modelo COM Fragilidade
a_ind <- as.data.frame(matrix(nrow = 5, ncol = 5), row.names = nome_linhas)
names(a_ind) <- nome_colunas
a_dep <- a_ind
a_raz <- a_dep
# Extrair o Risco Base apenas uma vez
H0_base <- basehaz(fit_cox_com_fragilidade, centered = FALSE)
# Criar a função de risco acumulado via interpolação
H0_fun <- function(tempo_alvo) {
  approx(H0_base$time,
    H0_base$hazard,
    xout = tempo_alvo,
    method = "constant", # Mantendo seu método
    rule = 2)$y
}
for (j in 1:ncol(a_ind)) {
  x <- as.numeric(names(a_ind)[j])
  for (k in 1:nrow(a_ind)) {
    y <- x + as.numeric(row.names(a_ind)[k])
    t = 0:(max(w - x, w - y))
    # SEM fragilidade - modelo de cox
    px <- approx(df$tempo, df$ss_m, xout = x, rule = 2)$y
    py <- approx(df$tempo, df$ss_f, xout = y, rule = 2)$y
    #Probabilidades condicionais t_p_x e t_p_y
    t_px_ind <- approx(df$tempo, df$ss_m, xout = x + t, rule = 2)$y / px
    t_py_ind <- approx(df$tempo, df$ss_f, xout = y + t, rule = 2)$y / py
    t_pxyus_ind <- t_px_ind + t_py_ind - (t_px_ind * t_py_ind)
    ind <- sum((v^t) * t_pxyus_ind )
    # COM fragilidade
    Sx <- approx(df$tempo, df$ss_pop_m, xout = x, rule = 2)$y
    Sy <- approx(df$tempo, df$ss_pop_f, xout = y, rule = 2)$y
    t_Sx <- approx(df$tempo, df$ss_pop_m, xout = x + t, rule = 2)$y
    t_Sy <- approx(df$tempo, df$ss_pop_f, xout = y + t, rule = 2)$y
    t_px_dep <- t_Sx/ Sx # individuais
    t_py_dep <- t_Sy/ Sy # individuais
    # conjunta marginal
    Sxy <- (1 + theta * (

```

```

    H0_fun(x) * exp(beta_com_fragilidade) + H0_fun(y)
  ))^(-1/theta)
  Sxy_t <- (1 + theta * (
    H0_fun(x + t) * exp(beta_com_fragilidade) + H0_fun(y + t)
  ))^(-1/theta)
  # Probabilidade conjunta condicional
  t_pxy_dep <- Sxy_t / Sxy
  t_pxyus_dep <- t_px_dep + t_py_dep - t_pxy_dep
  dep <- sum((v^t) * t_pxyus_dep)
  a_dep[k, j] <- dep
  a_ind[k, j] <- ind
  a_raz[k, j] <- dep / ind
}
}
round(a_ind, 4)
round(a_dep, 4)
round(a_raz, 4)

```