



Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

Modelo de Cox Bivariado através de cópulas

Erique Pereira Neto

Brasília

2014

Erique Pereira Neto

Modelo de Cox Bivariado através de cópulas

Relatório apresentado à disciplina Estágio Supervisionado II do curso de graduação em Estatística, Departamento de Estatística, Instituto de Exatas, Universidade de Brasília, como parte dos requisitos necessários para o grau de Bacharel em Estatística.

Orientador: Prof. Demerson Andre Polli

Brasília

2014

Dedicatória

*Ao meu tio **Adão**,*

que me deu a oportunidade e todo apoio para prosseguir com meus estudos.

E por ter sido exemplo de determinação e coragem.

Agradecimentos

Ao meu tio **Adão** por todo apoio e carinho.

A minha tia **Maria** por todo apoio, generosidade e carinho.

Aos meus primos **Welberth** e **Washington** pelo acolhimento fraterno e amigo.

Ao meu professor orientador **Démerson** pela paciência, confiança e amizade ao longo deste trabalho.

A todos os professores que tive na UnB, em especial agradeço a **Ana Maria Nogales** pelo carinho e incentivo.

Ao meu amigo **Felipe Veloso** que, desde o segundo dia de aula, tanto me apoiou e me alegrou nesses 4 anos de UnB.

A minha amiga **Geiziane** que sempre me apoiou e aconselhou, mas principalmente pela motivação e carinho.

Aos meus amigos **Augusto Hung** e **Felipe Nery** pelo apoio e conselhos.

Aos meus pais e irmãos.

A **Lenara Mendes** por todo amor e carinho.

A todos os meus colegas de faculdade que contribuíram de alguma forma para minha formação.

Resumo

Modelo de Cox Bivariado através de cópulas

Neste trabalho é apresentado o modelo de Cox bivariado através da modelagem de cópulas. A vantagem de usar o modelo de Cox é que não precisamos fazer suposições sobre a distribuição dos tempos de falha. Por outro lado, o modelo exige que os riscos sejam proporcionais. Foram usadas cópula de Gumbel e cópula de Clayton, para modelar a dependência das variáveis respostas. A estimação dos parâmetros do modelo foi realizada estimando as funções de sobrevivência marginais usando o método de Breslow (1972), e calculando a estimativa do coeficiente τ de Kendall para dados com censura e a partir da relação do parâmetro de dependência das cópulas com o τ de Kendall foi feita a estimação dos mesmos. Foi utilizado o método Jackknife para verificar o viés e erro padrão dos estimadores do parâmetro de dependência dos modelos. O conjunto de dados usado na aplicação dos modelos corresponde às observações de tempo até ocorrência de dois eventos distintos de infecção em pacientes com insuficiência renal.

Palavras-chave: Modelo de Cox bivariado, Riscos proporcionais, Cópulas, Dados censurados, Jackknife, Análise de sobrevivência.

Lista de Figuras

4.1	Estimador de Kaplan-Meier do tempo da primeira e segunda infecções por sexo	32
4.2	Resíduos de Cox-Snell para a variável tempo até a ocorrência da infecção 1	33
4.3	Resíduos de Cox-Snell para a variável tempo até a ocorrência da infecção 2	33
4.4	Resíduos deviance para a variável tempo até a ocorrência da infecção 1	34
4.5	Resíduos deviance para a variável tempo até a ocorrência da infecção 2	34
4.6	Gráfico das funções de sobrevivência marginais do modelo de Cox . .	36
4.7	Gráfico da função de sobrevivência do modelo de Cox bivariado através da cópula de Gumbel.	37
4.8	Gráfico da função de sobrevivência do modelo de Cox bivariado através da cópula de Gumbel por sexo.	37
4.9	Gráfico da cópula de Gumbel versus cópula Produto.	38
4.10	Gráfico da cópula de Gumbel versus cópula Produto por sexo.	39
4.11	Gráfico da função de sobrevivência do modelo de Cox bivariado através da cópula de Clayton.	40

4.12 Gráfico da função de sobrevivência do modelo de Cox bivariado através da cópula de Clayton por sexo.	40
4.13 Gráfico da cópula de Clayton versus cópula Produto por sexo.	41
4.14 Gráfico da cópula de Clayton versus cópula Produto.	42

Sumário

RESUMO	4
1 Introdução	11
1.1 Objetivos	12
2 Revisão Bibliográfica	13
2.1 Conceitos básicos	13
2.1.1 Tempo de falha	13
2.1.2 Censura	13
2.1.3 Distribuição dos tempos de sobrevivência	14
2.2 Estimação Não-Paramétrica	15
2.2.1 Estimador de Kaplan-Meier	15
2.2.2 Estimador de Nelson-Aalen	16
2.3 Estimação Semi-Paramétrica	17
2.3.1 Modelo de Cox	17
2.3.2 Avaliação do ajuste do modelo	19
2.3.2.1 Resíduos de Schoenfeld	19
2.3.2.2 Resíduos de Cox-Snell	20
2.3.2.3 Resíduo Deviance	20

2.4	Cópu­las	21
2.4.1	Cópu­las Arquimedianas	21
2.4.1.1	Definição	21
2.4.1.2	Teorema de Sklar	22
2.4.1.3	Independência de Variáveis Aleatórias	22
2.4.2	Exemplos de cópu­las	22
2.4.2.1	Cópu­la de Gumbel	22
2.4.2.2	Cópu­la de Clayton	23
2.4.2.3	Tau de Kendall	23
2.5	Modelo de Cox bivariado usando cópu­las	24
2.5.1	Introdução	24
2.5.2	Modelo de Cox através de cópu­las	26
3	Metodologia	27
3.1	Banco de dados	27
3.2	Métodos	28
3.2.1	Método Jackknife	29
4	Resultados	31
4.1	Análise descritiva	31
4.2	Modelo	36
4.2.1	Modelo com cópu­la de Gumbel	36
4.2.2	Modelo com cópu­la de Clayton	39
5	Conclusão	43

Anexo	44
Bibliografia	50

Capítulo 1

Introdução

A área da Estatística que estuda o tempo decorrido até a ocorrência de um evento de interesse é a análise de sobrevivência. As técnicas de análise de sobrevivência são usadas em várias áreas, mas, principalmente, na medicina e nas engenharia. Na área médica o interesse é estimar o tempo até o óbito de um paciente, por exemplo. Na engenharia o uso mais frequente da análise de sobrevivência é no estudo do tempo até a falha de equipamentos, daí o tempo até a ocorrência de um evento ser comumente chamado de tempo de falha. Em estudos de sobrevivência alguns indivíduos podem não sofrer o evento de interesse, neste caso são chamados de censuras.

Em alguns casos, a ocorrência de uma falha pode alterar a probabilidade de outra falha sucessiva em componentes de um mesmo indivíduo (exemplo: ao perder a visão em um dos olhos o outro tem maior risco de ser perdido). Nesta situação, a suposição de independência entre os tempos de sobrevivência (para componentes de um mesmo elemento amostral) pode não ser adequada. Então, temos dois tempos de sobrevivência podendo haver dependência entre eles. Pode acontecer também de algumas covariáveis influenciarem significativamente os tempos de sobrevivência sendo incluídas no modelo de regressão.

Este trabalho irá considerar o modelo de riscos proporcionais de Cox bivariado, utilizando cópulas para modelar a relação de dependência entre os tempos.

1.1 Objetivos

O objetivo deste trabalho é estudar a teoria de cópulas e suas aplicações aos modelos de riscos proporcionais de Cox bivariados.

Objetivos específicos:

- Modelar a dependência dos tempos de falha, em modelos de Cox bivariados, através de cópulas.
- Estudar o comportamento do modelo bivariado de Cox para dados da área da saúde.

Capítulo 2

Revisão Bibliográfica

2.1 Conceitos básicos

2.1.1 Tempo de falha

A principal variável de interesse em estudos de sobrevivência é o tempo de falha, que é o tempo decorrido até a ocorrência do evento de interesse. O evento de interesse é arbitrário sendo desde a morte ou infecção de indivíduos na medicina; falência de empresas ou inadimplência de clientes na área financeira ou a falha de equipamentos na indústria e na engenharia. Segundo Colosimo e Giolo (2006), a definição do tempo de falha é constituída por três elementos:

- fixar o tempo de início do estudo,
- a escala de medida a ser utilizada e
- estabelecer o evento de interesse, chamado de falha.

2.1.2 Censura

Censura é a não ocorrência do evento de interesse até o fim do estudo ou até o último instante de observação do elemento amostral. Pode ocorrer por diversos motivos e

em momentos diferentes do estudo. Existem diversos tipos de censuras. Quando no momento da observação do elemento amostral o evento de interesse já tenha acontecido, temos censura à esquerda. Se a observação é censurada antes de se observar a falha, a censura é chamada censura à direita, que pode ser de três tipos:

- Censura à direita tipo I: ocorre quando é fixado o tempo de estudo e após o término do estudo um ou mais indivíduos são censurados por não sofrerem a falha.
- Censura à direita tipo II: Ocorre quando fixamos a quantidade de falhas como critério de parada do estudo, e após o fim do estudo um ou mais indivíduos não sofreram o evento de interesse.
- Censura à direita aleatória: ocorre quando indivíduos não sofrem o evento de interesse por motivos aleatórios.

Além desses tipos de censura, existe ainda a censura intervalar que ocorre em estudos em que os elementos tem acompanhamento periódico e o evento de interesse ocorre num intervalo de tempo entre duas observações, entretanto não se sabe o momento exato. Neste trabalho será abordado apenas dados com censura à direita.

2.1.3 Distribuição dos tempos de sobrevivência

O tempo de sobrevivência é uma variável aleatória T que toma valores no intervalo $[0, \infty)$. A função densidade de probabilidade de T é expressa da seguinte forma:

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t} \quad (2.1)$$

em que $f(t) \geq 0$. A função de distribuição acumulada da variável aleatória T é dada por $F(t) = \int_0^t f(u)du$.

A função de sobrevivência, denotada por $S(t)$, definida como a probabilidade de um indivíduo sobreviver a um tempo t . É expressa por

$$S(t) = P(T > t) = \int_t^{\infty} f(u)du = 1 - F(t). \quad (2.2)$$

Outra forma de representar a distribuição da variável aleatória T é através da função risco de falha, definida como limite da probabilidade de um indivíduo falhar no intervalo de tempo $[t; \Delta t)$, assumindo que este mesmo indivíduo sobreviveu até o tempo t , dividida pelo comprimento do intervalo e é representada por:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T > t)}{\Delta t} \quad (2.3)$$

Da equação acima segue que $\lambda(t) = \frac{f(t)}{S(t)} = -\frac{(1-F(t))'}{S(t)}$. Então, $\lambda(t) = -(\log(S(t)))'$.

Assim, temos $S(t) = \exp\left(-\int_0^t \lambda(u)du\right) = \exp(-\Lambda(t))$, com $\Lambda(t) = \int_0^t \lambda(u)du$.

2.2 Estimação Não-Paramétrica

2.2.1 Estimador de Kaplan-Meier

Em qualquer estudo que envolva análise estatística a análise descritiva dos dados é parte essencial do estudo. Na análise de sobrevivência, devido a presença de censura nos tempos de sobrevivência devemos usar métodos descritivos diferentes dos tradicionais. Para a estimação da função de sobrevivência, um dos métodos mais utilizados é o estimador de Kaplan-Meier, que é calculado fazendo o seguinte procedimento Colosimo e Giolo (2006):

- ordenar os k tempos de falhas distintas, $t_1 < t_2 < \dots < t_k$;
- d_j o número de falhas no tempo t_j , $j = 1, 2, \dots, k$;
- n_j o número de indivíduos sob risco (indivíduos que não falharam e não censurados) no tempo t_j .

O estimador de Kaplan-Meier é dado por

$$\hat{S}(t) = \prod_{j:t_j < t} \left(\frac{n_j - d_j}{n_j} \right) = \prod_{j:t_j < t} \left(1 - \frac{d_j}{n_j} \right). \quad (2.4)$$

O estimador de Kaplan-Meier coincide com o estimador de máxima verossimilhança da função de sobrevivência Colosimo e Giolo (2006).

2.2.2 Estimador de Nelson-Aalen

O estimador de Nelson-Aalen para a função risco acumulado é definido por:

$$\hat{\Lambda}_{NA}(t) = \sum_{j:t_j < t} \left(\frac{d_j}{n_j} \right). \quad (2.5)$$

em que d_j é o número de falhas no tempo t_j , n_j é o número de indivíduos sob risco em t_j e $j = 1, \dots, k$.

2.3 Estimação Semi-Paramétrica

2.3.1 Modelo de Cox

O modelo de regressão de Cox, proposto em 1972, é um dos mais utilizados para a análise de dados de sobrevivência. Este grande interesse pelo modelo se deve ao fato deste ser bastante versátil, permitindo a análise de dados sem que seja necessário identificar a distribuição de probabilidade associada aos tempos de falha. Além disto, tal modelo incorpora a informação referente aos tempos de censura, corrigindo desta forma as estimativas. O modelo de Cox é definido a partir da função de risco, Cox (1972):

$$\lambda(t; X) = \lambda_0(t)g(X). \quad (2.6)$$

A função de sobrevivência é dada por

$$S(t; X) = \exp\left(-\int_0^t \lambda_0(u)g(X)du\right) = \exp(-\Lambda_0(t))^{g(X)} = [S_0(t)]^{g(X)} \quad (2.7)$$

em que $S_0(t) = \exp(-\Lambda_0(t))$ é a função de sobrevivência basal e $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ é a função de risco acumulada. A estimação dos parâmetros é feita através do método de Máxima Verossimilhança Parcial, que condiciona a função de risco na história dos dados. Cox, em 1975, definiu a seguinte função de Verossimilhança Parcial

$$L(\beta) = \prod_{i=1}^n \left[\frac{\exp(\beta' X)}{\sum_{j \in R(t_i)} \exp(\beta' X)} \right]^{\delta_i} \quad (2.8)$$

em que δ_i é o indicador de falha e $R(t_i)$ é o conjunto de indivíduos sob risco no tempo t_i . Breslow (1972) e Peto (1972) propõem uma modificação na função (2.8) para incluir as observações empatadas. Considere s_i o vetor formado pela soma das

correspondentes p covariáveis para os indivíduos que falharam no mesmo tempo t_i , $i = 1, \dots, k$ e d_i o número de falhas neste mesmo tempo. A modificação considera a seguinte função de verossimilhança parcial:

$$L(\beta) = \prod_{i=1}^k \frac{\exp(\beta' s_i)}{\left[\sum_{j \in R(t_i)} \exp(\beta' x_j) \right]^{d_i}} \quad (2.9)$$

Os valores dos parâmetros do vetor β são encontrados maximizando o logaritmo da função de Verossimilhança Parcial

$$l(\beta) = \log L(\beta) = \sum_{i=1}^k \beta' s_i - \sum_{i=1}^k d_i * \log \left[\sum_{j \in R(t_i)} \exp(\beta' x_j) \right]. \quad (2.10)$$

Tem-se que :

$$\frac{\partial l(\beta)}{\partial \beta_r} = \sum_{i=1}^k s_{ir} - \sum_{i=1}^k d_i \frac{\sum_{j \in R(t_i)} \exp(\beta' x_j) x_{jr}}{\sum_{j \in R(t_i)} \exp(\beta' x_j)} \quad (2.11)$$

em que $r = 1, 2, \dots, p$ e s_{ir} o r -ésimo componente em $s'_{ir} = (s_{i1}, s_{i2}, \dots, s_{ir}, \dots, s_{ip})$.

Resolvendo o sistema $U(\beta) = \frac{\partial l(\beta)}{\partial \beta_r} = 0$ encontramos os Estimadores de Máxima Verossimilhança (EMV) $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_r, \dots, \hat{\beta}_p$. Para estimação da função de sobrevivência basal utilizamos um estimador não paramétrico para $\Lambda_0(t)$ proposto por Breslow(1972);

$$\hat{\Lambda}_0(t) = \sum_{j:t_j < t} \frac{d_j}{\sum_{i \in R_j} \exp(\beta' x_i)} \quad (2.12)$$

segue que

$$\hat{S}_0(t; X) = \exp(-\hat{\Lambda}_0(t))$$

e

$$\hat{S}(t; X) = [\hat{S}_0(t; X)]^{\exp(\beta' X)}. \quad (2.13)$$

2.3.2 Avaliação do ajuste do modelo

2.3.2.1 Resíduos de Schoenfeld

Uma das formas de avaliar a suposição de riscos proporcionais no modelo de Cox é através dos resíduos de Schoenfeld. O resíduo Schoenfeld é a diferença entre os valores observados de covariáveis de um indivíduo com tempo de ocorrência do evento no tempo t_i e os valores esperados em t_i dado o grupo de risco $R(t_i)$. Considere o vetor de covariáveis $\mathbf{x}_i = (x_{1i}, x_{2i}, \dots, x_{pi})'$ para o i -ésimo indivíduo que sofreu uma falha. Tem-se para este indivíduo um vetor de resíduos de Schoenfeld $\mathbf{r}_i = (r_{i1}, r_{i2}, \dots, r_{ip})'$ em que cada componente r_{iq} , para $q = 1, 2, \dots, p$, é definido por:

$$r_{iq} = x_{iq} - \frac{\sum_{j \in R(t_i)} x_{jp} \exp(\hat{\beta}' x_j)}{\sum_{j \in R(t_i)} \exp(\hat{\beta}' x_j)}. \quad (2.14)$$

Uma forma padronizada dos resíduos de Schoenfeld é definida por:

$$s_i^* = [I(\hat{\beta})]^{-1} \times \mathbf{r}_i \quad (2.15)$$

em que $I(\hat{\beta})$ é a matriz de informação observada. Um teste para a hipótese global de proporcionalidade dos riscos sobre todas as covariáveis no modelo, pode ser realizado usando-se a estatística de teste:

$$T = \frac{(g - \bar{g})' S^* I S^* (g - \bar{g})}{d \sum_k (g_k - \bar{g})^2}, \quad (2.16)$$

em que I é a matriz de informação observada, d é o número de falhas e $S^* = dRI^{-1}$, sendo R a matriz $d \times p$ dos resíduos de Schoenfeld não padronizados e $g = g(t)$ é uma função de t , geralmente tomada como sendo $\log(t)$. Sob hipótese nula de proporcionalidade dos riscos, T tem distribuição qui-quadrado com p graus de liberdade. Valores de $T > \chi_{p,1-\alpha}^2$, sendo $\chi_{p,1-\alpha}^2$ o quantil $1 - \alpha$ de uma distribuição qui-quadrado

com p graus de liberdade e α é o nível de significância do teste, mostram evidências contra a suposição de riscos proporcionais.

2.3.2.2 Resíduos de Cox-Snell

Os resíduos de Cox-Snell são utilizados também para avaliar a qualidade geral de ajuste do modelo de Cox. Os resíduos de Cox-Snell são definidos por:

$$\hat{e}_i = \hat{\Lambda}_0(t_i) \exp \left[\sum_{k=1}^p x_{ip} \hat{\beta}_k \right] = \hat{\Lambda}_0(t_i) \exp \left[\hat{\beta}' X \right]. \quad (2.17)$$

Como $S(t_i)$ é uma uniforme no intervalo $(0, 1)$, pode-se mostrar que $-\log[S(t_i)]$ segue uma distribuição exponencial padrão. Logo, é de se esperar que

$$-\log[\hat{S}_i(t_i)] = \hat{\Lambda}(t_i) = \hat{e}_i$$

tenha a mesma distribuição exponencial padrão, ou seja, se o modelo tiver bem ajustado, o gráfico de $\hat{\Lambda}(\hat{e}_i)$ por \hat{e}_i será aproximadamente uma reta.

2.3.2.3 Resíduo Deviance

O resíduo deviance é definido por :

$$\hat{d}_i = \text{sinal}(\hat{m}_i) [-2(\hat{m}_i + \delta_i \log(\delta_i - \hat{m}_i))]^{\frac{1}{2}}, \quad (2.18)$$

em que

$$\hat{m}_i = \delta_i - \hat{\Lambda}_0(t_i) \exp \left[\sum_{k=1}^p x_{ip} \hat{\beta}_k \right] \quad (2.19)$$

é o resíduo martingal e o gráfico de \hat{d}_i versus o preditor linear $\sum_{i=1}^p x_{ip} \hat{\beta}_i$ é utilizado para verificar a presença de dados atípicos.

2.4 Cópulas

Cópula (bivariada) é uma função $C : [0, 1]^2 \longrightarrow [0, 1]$ que satisfaz as seguintes propriedades:

- Para todo u, v em $[0, 1]$, $C(0, v) = C(u, 0) = 0$;
- Para todo u, v em $[0, 1]$, $C(1, v) = v$ e $C(u, 1) = u$;
- Se $u_1 \leq u_2$ e $v_1 \leq v_2$, então $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$.

Algumas características de cópulas

- São funções que “acasalam” funções de distribuições e suas marginais.
- São funções de distribuições tais que $C(u, v) = P(U \leq u, V \leq v)$ com $u, v \in [0, 1]$.

2.4.1 Cópulas Arquimedianas

2.4.1.1 Definição

Uma cópula Arquimediana é uma função $C : [0, 1]^2 \longrightarrow [0, 1]$, dada por $C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v))$ em que φ uma função contínua convexa estritamente decrescente de $[0, 1] \longrightarrow [0, \infty]$ tal que $\varphi(1) = 0$ e, em que $\varphi^{[-1]}$ denota a “pseudo-inversa” de φ , isto é

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t) & \text{se } t \in [0, \varphi(0)) \\ 0 & \text{se } t \geq \varphi(0) \end{cases} \quad (2.20)$$

Quando $\varphi(0) = \infty$, φ é dito ser estrito e $\varphi^{[-1]} \equiv \varphi^{-1}$. φ é chamado de gerador.

2.4.1.2 Teorema de Sklar

Seja o vetor aleatório (X, Y) com função de distribuição conjunta $\mathbb{H}(x, y) = \mathbb{P}(X \leq x, Y \leq y)$ e marginais \mathbb{F}_X e \mathbb{G}_Y . Então existe uma cópula C tal que para todo x, y em $\bar{\mathbb{R}}$ ($\bar{\mathbb{R}} = [-\infty, \infty]$ é uma extensão da reta real)

$$\mathbb{H}(x, y) = C(F_X(x), G_Y(y)). \quad (2.21)$$

Se \mathbb{F}_X e \mathbb{G}_Y são contínuas, então C é única; Caso contrário, C é unicamente determinada em $Im(\mathbb{F}_X) \times Im(\mathbb{G}_Y)$. Reciprocamente, se C é uma cópula e \mathbb{F}_X e \mathbb{G}_Y são funções de distribuição acumulada univariadas então a função \mathbb{H} é uma função de distribuição conjunta com marginais \mathbb{F}_X e \mathbb{G}_Y . Veja prova desse teorema em Nelsen (2006).

2.4.1.3 Independência de Variáveis Aleatórias

Seja X e Y um vetor de variáveis aleatórias contínuas. Então, X e Y são independentes se, e somente se, $C(F(x), G(y)) = \Pi$. Em que $\Pi = \Pi(u, v) = uv$ é a cópula produto. Ou seja, X e Y são independentes se, e somente se, $\mathbb{H}(x, y) = C(F(x), G(y)) = F(x)G(y)$.

2.4.2 Exemplos de cópulas

2.4.2.1 Cópula de Gumbel

A cópula de Gumbel pertence à classe de cópulas arquimedianas com gerador $\varphi_\theta(t) = (-\log t)^\theta$ e é definida da seguinte forma

$$C_\theta(u, v) := \exp \left[-\{(-\log(u))^\theta + (-\log(v))^\theta\}^{\frac{1}{\theta}} \right] \quad (2.22)$$

com $\theta \geq 1$.

2.4.2.2 Cópula de Clayton

A cópula de Clayton é da classe de cópulas arquimedianas com gerador dado por $\varphi_\theta(t) = t^{-\theta} - 1$ e é definida da seguinte forma

$$C_\theta(u, v) := (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}} \quad (2.23)$$

com $\theta > 0$.

2.4.2.3 Tau de Kendall

Uma das medidas de dependência ou associação entre duas variáveis aleatórias X e Y mais usadas é o coeficiente τ de Kendall. A ideia por trás do cálculo do coeficiente τ de Kendall compara o número de pares concordantes e discordantes de um vetor de variáveis aleatórias contínuas (X, Y) . Sejam (x_i, y_i) e (x_j, y_j) duas observações do vetor (X, Y) . Dizemos que (x_i, y_i) e (x_j, y_j) são concordantes se $(x_i - x_j)(y_i - y_j) > 0$, discordantes se $(x_i - x_j)(y_i - y_j) < 0$ e empatados se $(x_i - x_j)(y_i - y_j) = 0$. Seja $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ uma amostra aleatória de n observações de um vetor (X, Y) de variáveis aleatórias contínuas. Então a versão amostral do τ de Kendall é dada por

$$\hat{\tau} = \frac{C - D}{C + D + E} = \frac{C - D}{\binom{n}{2}}, \quad (2.24)$$

em que C é o número de pares concordantes, D é o número de pares discordantes, E é o número de empates e existem $\binom{n}{2}$ pares de observações (x_i, y_i) e (x_j, y_j) na amostra. A versão populacional do τ de Kendall é definida de forma semelhante. Seja (X_1, Y_1) e (X_2, Y_2) vetores aleatórios contínuos independentes e identicamente distribuídos, cada um com função de distribuição conjunta H . O τ de Kendall

populacional é definido como a probabilidade de concordância menos a probabilidade de discordância:

$$\tau = \tau_{X,Y} = \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0]. \quad (2.25)$$

Se X e Y forem variáveis aleatórias contínuas com uma cópula arquimediana C gerada por φ , então o τ de Kendall populacional para X e Y é dado por

$$\tau_C = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt \quad (2.26)$$

Para a cópula de Gumbel com gerador $\varphi_\theta(t) = (-\log t)^\theta$ temos

$$\frac{\varphi_\theta(t)}{\varphi'_\theta(t)} = \frac{t \log t}{\theta}, \theta \geq 1$$

e assim

$$\tau_\theta = \frac{\theta - 1}{\theta}.$$

Já a cópula de Clayton com gerador $\varphi_\theta(t) = t^{-\theta} - 1$ temos

$$\frac{\varphi_\theta(t)}{\varphi'_\theta(t)} = \frac{t^{\theta+1} - t}{\theta}, \theta > 0$$

e assim

$$\tau_\theta = \frac{\theta}{\theta + 2}.$$

2.5 Modelo de Cox bivariado usando cópulas

2.5.1 Introdução

Em alguns estudos de análise de sobrevivência há duas variáveis respostas, ou seja, estamos investigando o tempo decorrido até dois eventos de interesse num mesmo elemento amostral e estes eventos competem entre si, podendo haver uma relação

de dependência entre os tempos de sobrevivência. Existem várias formas de se analisar e olhar para estas falhas. A falha dos motores de um avião bimotor pode ser considerada como falhas distintas no modelo estatístico (neste caso o tempo de falha de cada motor pode ser observado) ou a falha de algum dos motores interrompe a observação do avião (elemento amostral). Este último caso é um exemplo de riscos competitivos, no qual apenas a falha de um dos motores é observada e o outro motor é censurado. Nestes casos, são usados modelos bivariados para a análise dos tempos de sobrevivência. Atendida a suposição de riscos proporcionais podemos usar o modelo de Cox bivariado, que é definido da seguinte forma

$$\begin{cases} \lambda(t_1; Z) = \lambda_0(t_1)\phi(Z) \\ \lambda(t_2; Z) = \lambda_0(t_2)\psi(Z) \end{cases} \quad (2.27)$$

Frequentemente, as funções positivas $\phi(Z)$ e $\psi(Z)$ são assumidas como sendo funções paramétricas, com $\phi(Z) = \exp(\alpha'Z)$ e $\psi(Z) = \exp(\beta'Z)$, em que α e β são vetores de parâmetros e Z é o vetor de covariáveis. Raramente estamos interessados apenas nos comportamentos marginais como dados no modelo (2.29). Este modelo ignora a relação de dependência entre T_1 e T_2 . Assim, um modelo mais interessante e que será considerado neste trabalho é o que leva em consideração a dependência dos tempos de sobrevivência, definido como

$$\begin{cases} \lambda(t_1; Z) = \lambda_0(t_1)\phi(Z) \\ \lambda(t_2; Z|t_1) = \lambda_0(t_2|t_1)\psi(Z) \end{cases} \quad (2.28)$$

2.5.2 Modelo de Cox através de cópulas

A função de sobrevivência para o modelo de Cox bivariado pode ser definida, através de cópulas, como

$$S(t_1, t_2) = P(T_1 > t_1, T_2 > t_2) = C_\theta (S_1(t_1), S_2(t_2)), t_1, t_2 \geq 0 \quad (2.29)$$

com θ sendo um parâmetro de dependência. Na presença de covariáveis, o modelo assume a seguinte forma

$$S(t_1, t_2; Z) = C_\theta (S_1(t_1; Z), S_2(t_2; Z)), t_1, t_2 \geq 0. \quad (2.30)$$

Adotando a cópula de Gumbel definida em (2.24), temos

$$\begin{aligned} S(t_1, t_2; Z) &= C_\theta (S_1(t_1; Z), S_2(t_2; Z)) \\ &= \exp \left[- \left\{ (-\log(S_1(t_1; Z)))^\theta + (-\log(S_2(t_2; Z)))^\theta \right\}^{\frac{1}{\theta}} \right]. \end{aligned}$$

Usando a cópula de Clayton, a função de sobrevivência é expressa por

$$\begin{aligned} S(t_1, t_2; Z) &= C_\theta (S_1(t_1; Z), S_2(t_2; Z)) \\ S(t_1, t_2; Z) &= \left[(S_1(t_1; Z))^{-\theta} + (S_2(t_2; Z))^{-\theta} - 1 \right]^{-\frac{1}{\theta}}. \end{aligned}$$

Capítulo 3

Metodologia

3.1 Banco de dados

O conjunto de dados corresponde a observações de tempo até ocorrência de dois eventos distintos de infecção em pacientes com insuficiência renal, que estão descritos na tese de doutorado de Fachini (2011, p.144). Os pacientes utilizam máquinas de diálise portáteis e considera-se a ocorrência das infecções no ponto em que um cateter é inserido. Quando uma infecção ocorre o cateter deve ser removido até que a infecção seja curada para que o cateter seja reinsertado. Os tempos entre a inserção do cateter e a ocorrência de uma infecção são registrados. Dessa forma podem ocorrer diversas infecções nos pacientes, mas apenas duas são registradas nesse conjunto de dados. Existem situações em que o cateter é removido sem que a infecção tenha ocorrido; neste caso, ocorrem as censuras à direita. As variáveis resposta de interesse são os tempos entre as inserções do cateter e a ocorrência da infecção. Será considerada apenas a covariável sexo referente a 38 pacientes. Para cada paciente $i, i = 1, 2, \dots, 38$, as variáveis associadas estão descritas a seguir:

- t_{i1} : tempo até a ocorrência da infecção 1, em semanas;

- t_{i2} : tempo até a ocorrência da infecção 2, em semanas;
- δ_{i1} : indicador de falha do evento 1;
- δ_{i2} : indicador de falha do evento 2;
- z_i : sexo do paciente (0: masculino, 1: feminino).

3.2 Métodos

A estimação do modelo seria feita, naturalmente, usando máxima verossimilhança para encontrar estimativas para o vetor $\beta' = (\beta_1, \beta_2, \dots, \beta_k)$ e o parâmetro θ da cópula, estimando-os simultaneamente. A função de verossimilhança é expressa por Lawless (2002):

$$L = \prod_{j=1}^n c_{\theta}(u, v)^{\delta_{1j}\delta_{2j}} \frac{-\partial C_{\theta}(u, v)^{\delta_{1j}(1-\delta_{2j})}}{\partial t_{1j}} \times \frac{-\partial C_{\theta}(u, v)^{(1-\delta_{1j})\delta_{2j}}}{\partial t_{2j}} C_{\theta}(u, v)^{(1-\delta_{1j})(1-\delta_{2j})} \quad (3.1)$$

em que δ_{1j} e δ_{2j} são os indicadores de censura das variáveis T_1 e T_2 , $c_{\theta}(u, v) = \frac{\partial^2 C_{\theta}(u, v)}{\partial t_{1j} \partial t_{2j}}$ e $(u, v) = (S_1(t_1; Z), S_2(t_2; Z))$. Mas este procedimento envolve uma enorme complexidade computacional quando se considera um modelo semi-paramétrico. Assim, optou-se por estimar o valor de θ não-parametricamente usando a sua relação com τ de Kendall que para cópula de Gumbel é definido por $\hat{\theta} = \frac{1}{1-\hat{\tau}}$ e na cópula de Clayton o estimador de θ é definido por $\hat{\theta} = \frac{2\hat{\tau}}{1-\hat{\tau}}$ em que $\hat{\tau}$ é o estimador empírico do τ de Kendall, Nelsen (2006, p.158). Como os dados possuem censuras, foi utilizada uma estimação para τ levando este fato em consideração. Brown et al (1974) citado

em Wang e Wells (2000) adota o seguinte estimador para τ

$$\hat{\tau} = \frac{\sum_{i,j=1}^n a_{ij} b_{ij}}{(\sum_{i,j=1}^n a_{ij}^2 \sum_{i,j=1}^n b_{ij}^2)^{1/2}} \quad (3.2)$$

com a_{ij} calculado como na tabela 3.1 e b_{ij} calculado de forma similar.

Tabela 3.1. Valores de a_{ij} do estimador de Brown et al.

$(\delta_{1i}, \delta_{1j})$	$T_{1i} > T_{1j}$	$T_{1i} = T_{1j}$	$T_{1i} < T_{1j}$
(1,1)	1	0	-1
(0,1)	1	1	$2\{\hat{\mathbb{F}}_1(T_{1j})/\hat{\mathbb{F}}_1(T_{1i})\} - 1$
(1,0)	$1 - 2\{\hat{\mathbb{F}}_1(T_{1j})/\hat{\mathbb{F}}_1(T_{1i})\}$	-1	-1
(0,0)	$1 - \{\hat{\mathbb{F}}_1(T_{1j})/\hat{\mathbb{F}}_1(T_{1i})\}$	$1 - \{\hat{\mathbb{F}}_1(T_{1j})/\hat{\mathbb{F}}_1(T_{1i})\}$	$\{\hat{\mathbb{F}}_1(T_{1j})/\hat{\mathbb{F}}_1(T_{1i})\} - 1$

Em que $\mathbb{F}_1(\cdot)$ é o estimador marginal de Kaplan-Meier para a variável T_1 . No cálculo de τ foi usado a função “cenken” do pacote “NADA” do software R que calcula o coeficiente τ de Kendall para dados censurados.

3.2.1 Método Jackknife

O método jackknife é usado principalmente para calcular variância e viés de estimadores. Sua metodologia é bastante simples e robusta. Seja $\mathbf{x} = (x_1, x_2, \dots, x_n)$ uma amostra aleatória de tamanho n e deseja-se estimar um parâmetro θ usando um estimador $\hat{\theta} = s(\mathbf{x})$. O método jackknife subdivide a amostra em n amostras com $n - 1$ elementos cada. A i -ésima amostra jackknife consiste do conjunto de dados com a i -ésima observação removida, ou seja,

$$\mathbf{x}_{(i)} = (x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \quad (3.3)$$

Seja

$$\hat{\theta}_{(i)} = s(\mathbf{x}_{(i)}) \quad (3.4)$$

a i -ésima replicação jackknife de $\hat{\theta}$. A estimativa jackknife do viés é definido por

$$vies_{jack} = (n - 1)(\hat{\theta}_{(\cdot)} - \hat{\theta}) \quad (3.5)$$

sendo que

$$\hat{\theta}_{(\cdot)} = \sum_{i=1}^n \frac{\hat{\theta}_{(i)}}{n}. \quad (3.6)$$

O estimador jackknife do erro padrão é definido por

$$se_{jack} = \left[\frac{n-1}{n} \sum_{i=1}^n (\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)})^2 \right]^{\frac{1}{2}}. \quad (3.7)$$

Como uma regra do polegar, podemos desprezar o viés quando $\left| \frac{vies_{jack}}{se_{jack}} \right| \leq 0.25$, Efron e Tibshirani (1993, p.141). Uma outra forma de pensar o método jackknife é em termos dos “pseudo-valores” definidos por

$$\hat{\theta}_i = n\hat{\theta} - (n-1)\hat{\theta}_{(i)}. \quad (3.8)$$

Então o estimador do erro padrão de um estimador $\hat{\theta}$ é expresso por

$$se_{jack} = \left\{ \frac{\sum_{i=1}^n (\tilde{\theta}_i - \tilde{\theta})^2}{(n-1)n} \right\}^{\frac{1}{2}}. \quad (3.9)$$

com

$$\tilde{\theta} = \sum_{i=1}^n \frac{\tilde{\theta}_i}{n}. \quad (3.10)$$

Segundo Tukey (1958), $\tilde{\theta}_i$ são independentes e identicamente distribuídos. Consequentemente, temos

$$\frac{\tilde{\theta} - \theta}{se_{jack}} \sim t_{(n-1)} \quad (3.11)$$

em que $t_{(n-1)}$ é a distribuição t-Student com $n - 1$ graus de liberdade. E, portanto, um intervalo de confiança de $100(1 - \alpha)\%$ para θ é dado por

$$\tilde{\theta} \pm t_{(n-1);(1-\alpha)} se_{jack}. \quad (3.12)$$

Capítulo 4

Resultados

4.1 Análise descritiva

Definindo as seguintes variáveis, de acordo com o descrito na seção 3.1:

- T_1 : tempo até a ocorrência da infecção 1, em semanas;
- T_2 : tempo até a ocorrência da infecção 2, em semanas;

Fazendo o gráfico de Kaplan-Meier considerando a covariável sexo podemos notar que na variável T_1 há um cruzamento entre as curvas, e na variável T_2 não apresenta essa característica. O cruzamento das curvas de sobrevivência de Kaplan-Meier para a covariável sexo do paciente é um indicativo de que possivelmente a variável T_1 não segue riscos proporcionais. Isto será verificado com o teste dos resíduos de Schoenfeld.

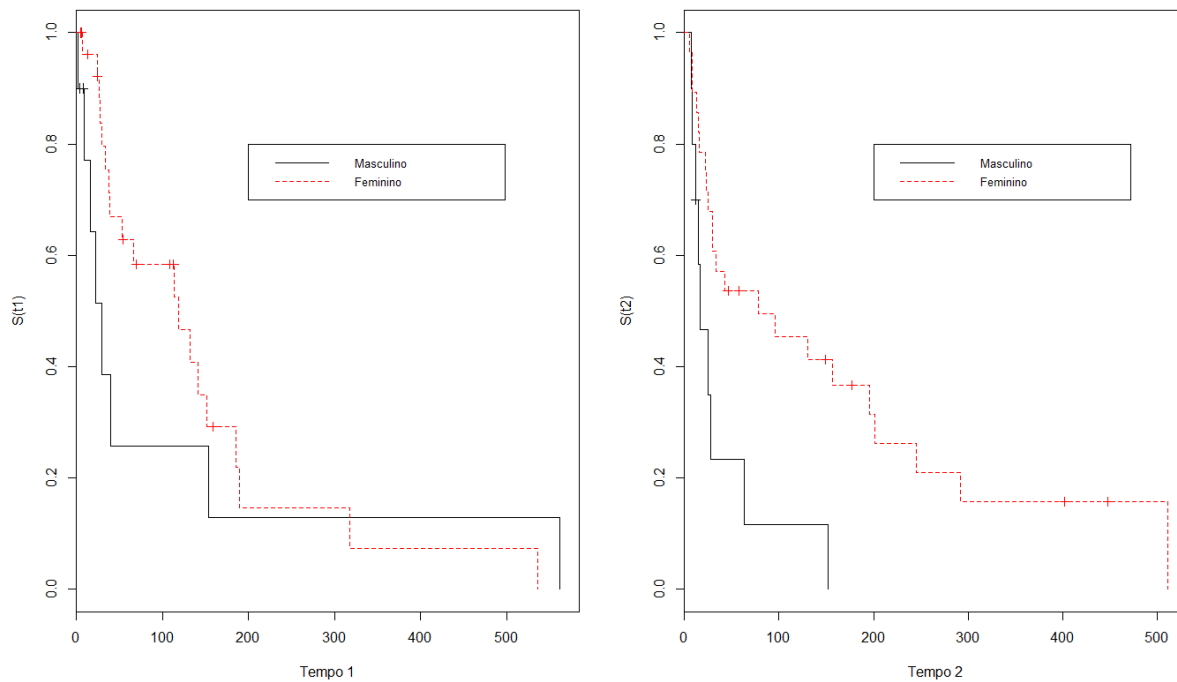


Figura 4.1: Estimador de Kaplan-Meier do tempo da primeira e segunda infecções por sexo

Os gráficos de resíduos de Cox-Snell para as variáveis tempo até a ocorrência da infecção 1 e tempo até a ocorrência da infecção 2 indicam que o ajuste do modelo está adequado.

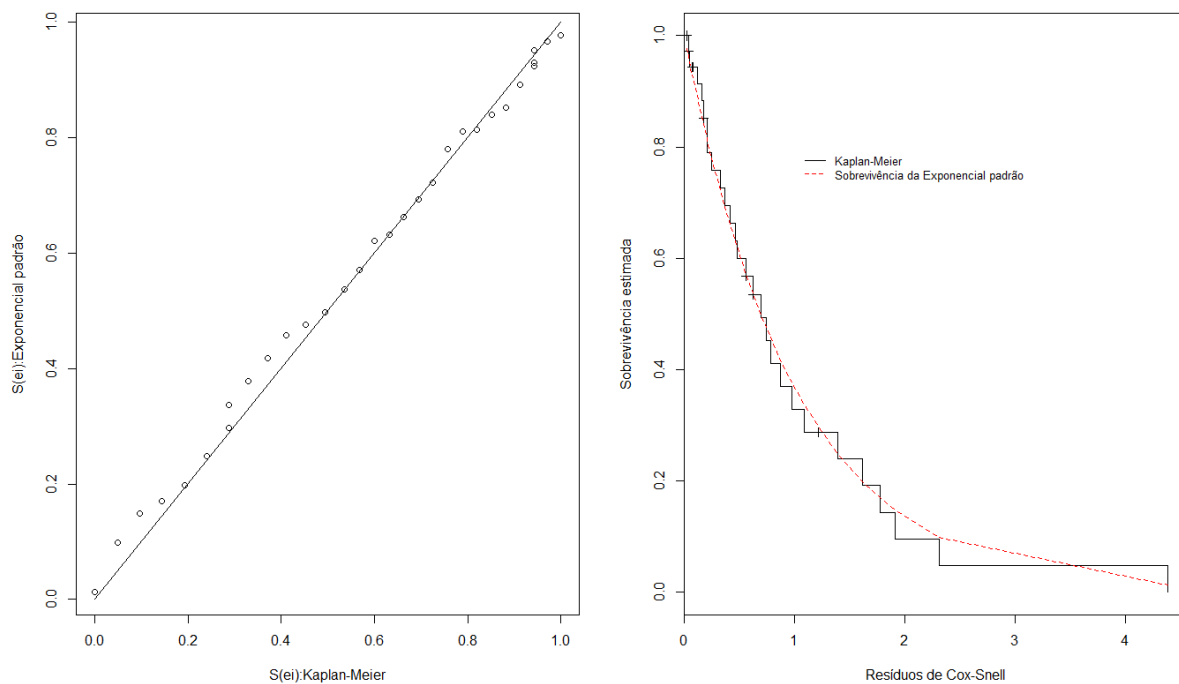


Figura 4.2: Resíduos de Cox-Snell para a variável tempo até a ocorrência da infecção 1

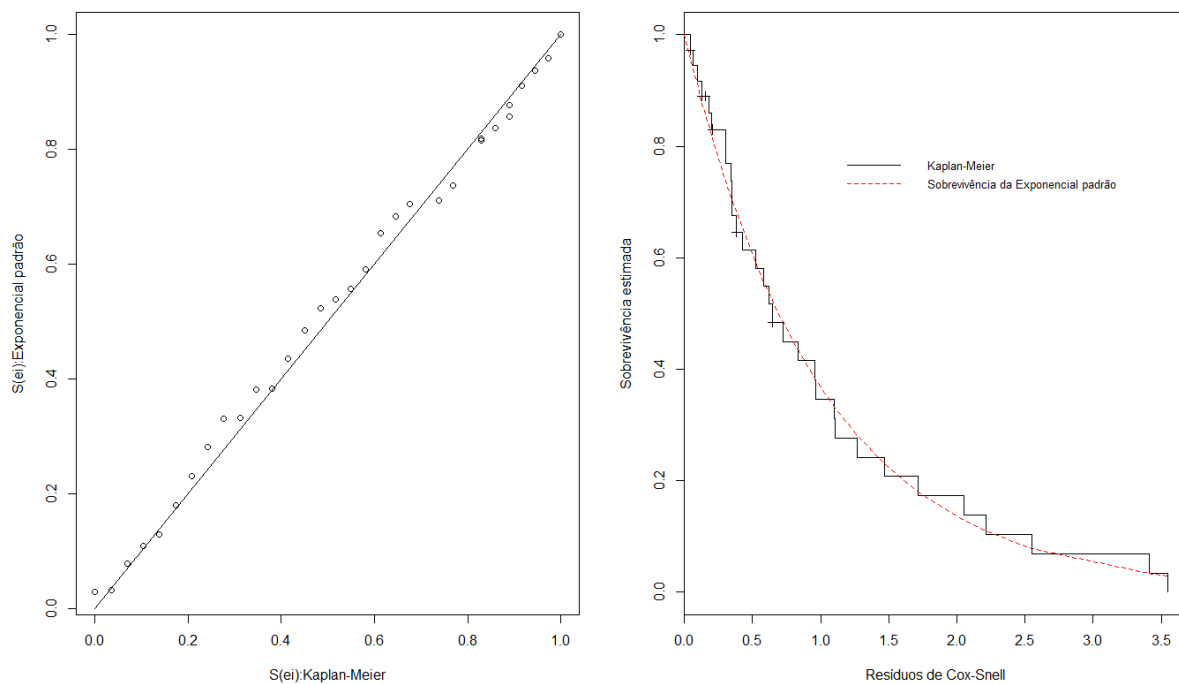


Figura 4.3: Resíduos de Cox-Snell para a variável tempo até a ocorrência da infecção 2

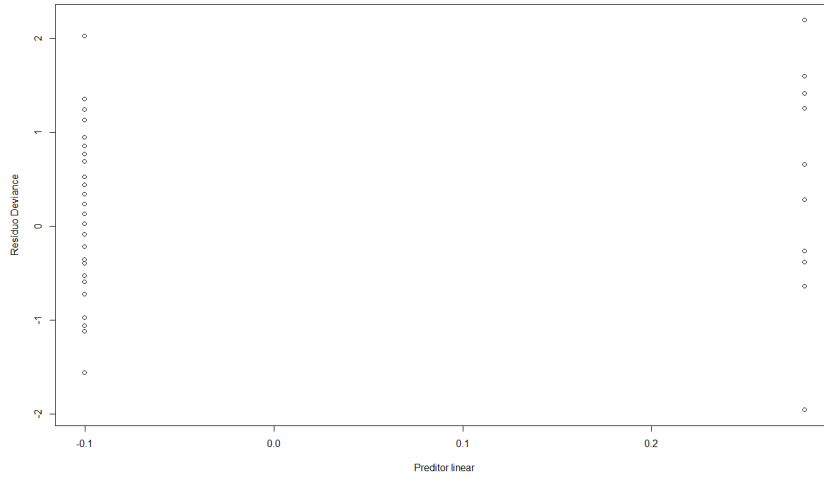


Figura 4.4: Resíduos deviance para a variável tempo até a ocorrência da infecção 1

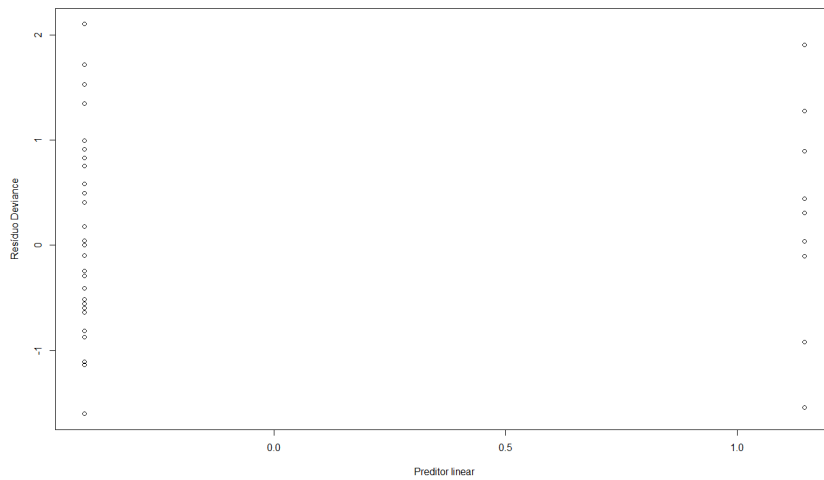


Figura 4.5: Resíduos deviance para a variável tempo até a ocorrência da infecção 2

Olhando para os resíduos deviance em ambas as variáveis, percebemos que não existem pontos atípicos, os resíduos estão entre $-2,5$ e $2,5$, mais um indicativo que o ajuste está adequado.

A estatística do teste para o resíduo de Schoenfeld para o tempo da primeira infecção foi $3,90$ ($p\text{-valor} = 0,0484$) e para a segunda infecção foi $< 0,01$ ($p\text{-valor} = 0,988$) mostrando que para o segundo tempo de infecção é possível considerar riscos proporcionais. Para o primeiro tempo de infecção podemos considerar riscos proporcionais se for considerado um alpha menor que 0.05 . O p-valor do teste dos resíduos de Schoenfeld para a variável tempo até a ocorrência da infecção 1 fica bem próximo de 0.05 , então será mantido o modelo. Portanto, o modelo de Cox pode ser utilizado. Foi calculado o coeficiente de correlação de τ de Kendall para dados com censura, resultando em $\hat{\tau} = 0,04$, mostrando ausência de correlação entre os tempos de sobrevivência T_1 e T_2 . Mesmo assim será usado o modelo com a relação de dependência para exemplo de aplicação. Podemos ajustar as cópulas. As estimativas das marginais do Modelo de Cox estão dadas na figura 4.6, que são obtidas fazendo

$$S_k(t_k; Z) = \lim_{t_i \rightarrow 0} C_\theta(S_k(t_k; Z), S_i(t_i; Z)) = C_\theta(S_k(t_k; Z), 1), k, i = 1, 2, k \neq i. \quad (4.1)$$

A figura 4.6 mostra proximidade entre as curvas de Kaplan-Meier e as marginais do modelo de Cox estimadas, indicando que é adequado usar o modelo de Cox neste caso.

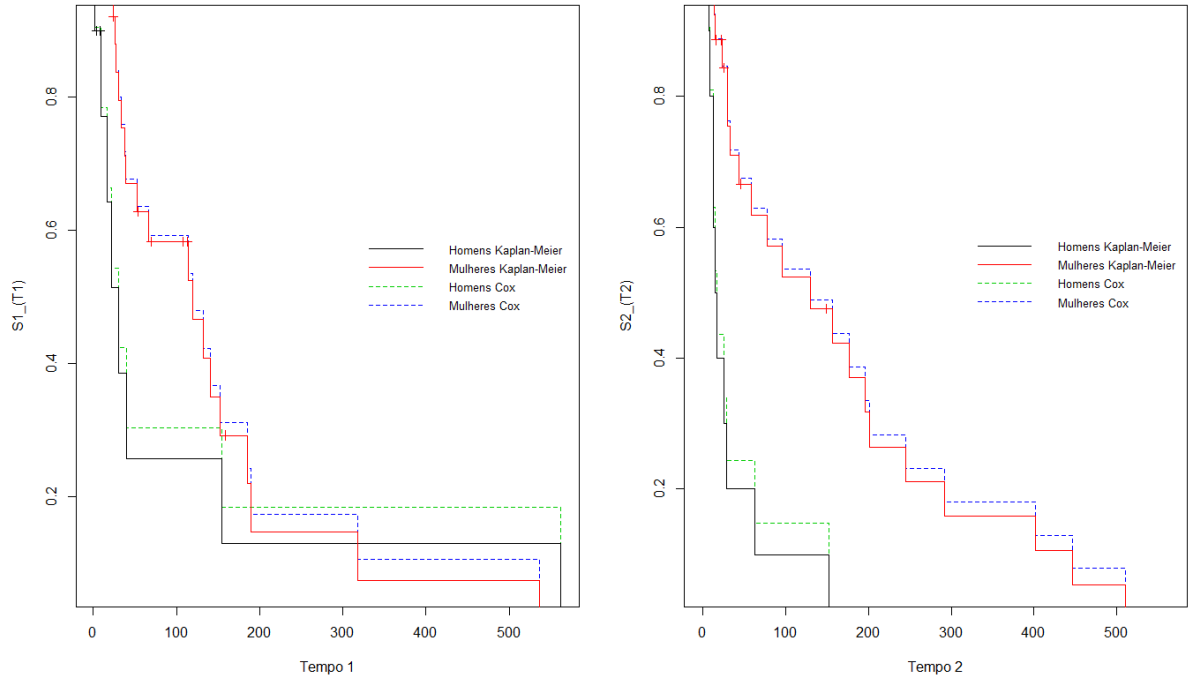


Figura 4.6: Gráfico das funções de sobrevivência marginais do modelo de Cox

4.2 Modelo

4.2.1 Modelo com cópula de Gumbel

Usando o método jackknife, discutido na seção **3.2.1**, a estimativa do viés do estimador $\hat{\theta} = \frac{1}{1-\hat{\tau}}$ foi de $vies_{jack} = 0,0009$ e o erro padrão foi de $se_{jack} = 0,03$. Fazendo a relação $\frac{vies_{jack}}{se_{jack}} = 0,0296$, temos que o viés do estimador $\hat{\theta} = \frac{1}{1-\hat{\tau}}$ pode ser desprezado. Com um coeficiente de correlação τ de Kendall entre os tempos da primeira infecção e da segunda infecção estimado em $\hat{\tau} = 0,04$ tem-se, como visto na seção 2.4.2.1, que a estimativa do parâmetro de dependência θ da cópula de Gumbel é $\hat{\theta} = 1,041$. Assim a cópula de Gumbel para o modelo é

$$C_{\hat{\theta}}(\hat{S}_1(t_1; Z), \hat{S}_2(t_2; Z)) = \exp \left[- \left\{ \left(-\log(\hat{S}_1(t_1; Z)) \right)^{1.041} + \left(-\log(\hat{S}_2(t_2; Z)) \right)^{1.041} \right\}^{\frac{1}{1.041}} \right] \quad (4.2)$$

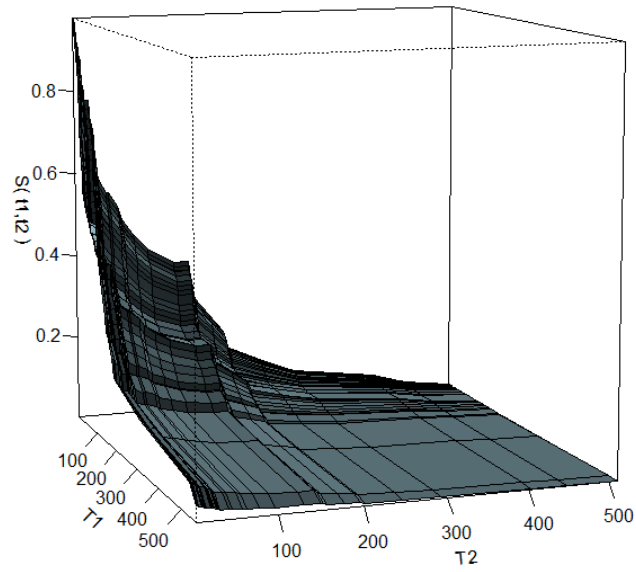


Figura 4.7: Gráfico da função de sobrevivência do modelo de Cox bivariado através da cópula de Gumbel.

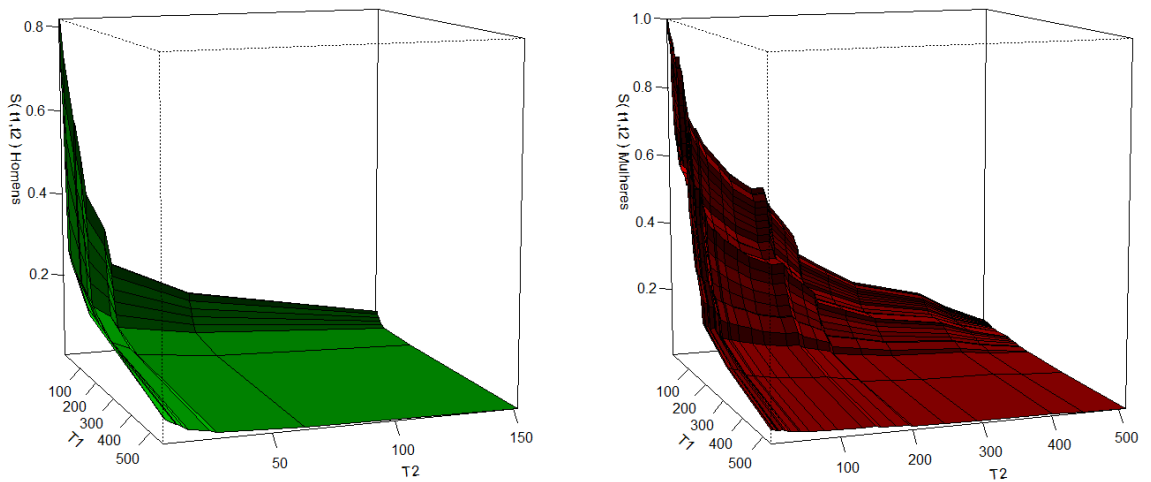


Figura 4.8: Gráfico da função de sobrevivência do modelo de Cox bivariado através da cópula de Gumbel por sexo.

Como o valor de $\hat{\theta} = 1,041$ é muito próximo de 1, temos que as variáveis T_1 e T_2 são independentes, como mostra a figura 4.9, mantendo também essa característica quando consideramos o sexo. Os gráficos mostrados nas figura 4.9 e figura 4.10 foram construídos plotando-se $C_{\hat{\theta}}(\hat{S}_1(t_1; Z), \hat{S}_2(t_2; Z))$ contra $\Pi(\hat{S}_1(t_1; Z), \hat{S}_2(t_2; Z)) = \hat{S}_1(t_1; Z)\hat{S}_2(t_2; Z)$.

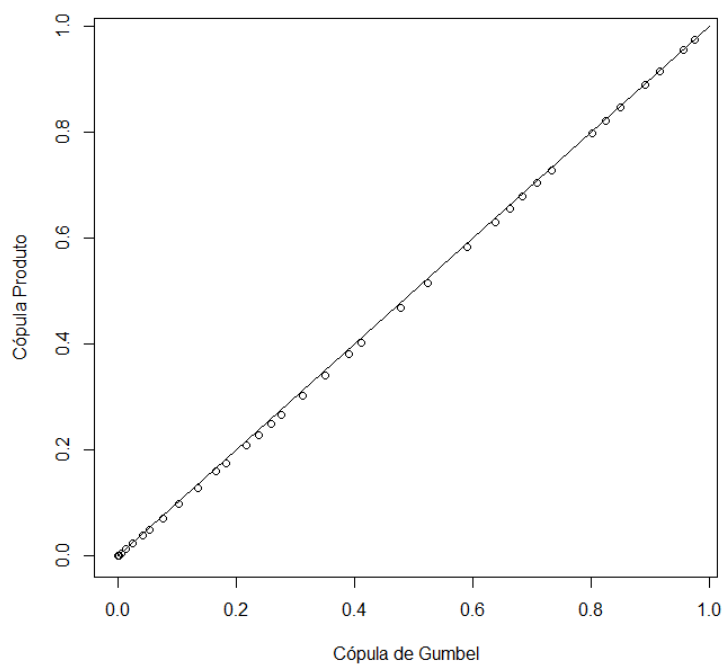


Figura 4.9: Gráfico da cópula de Gumbel versus cópula Produto.

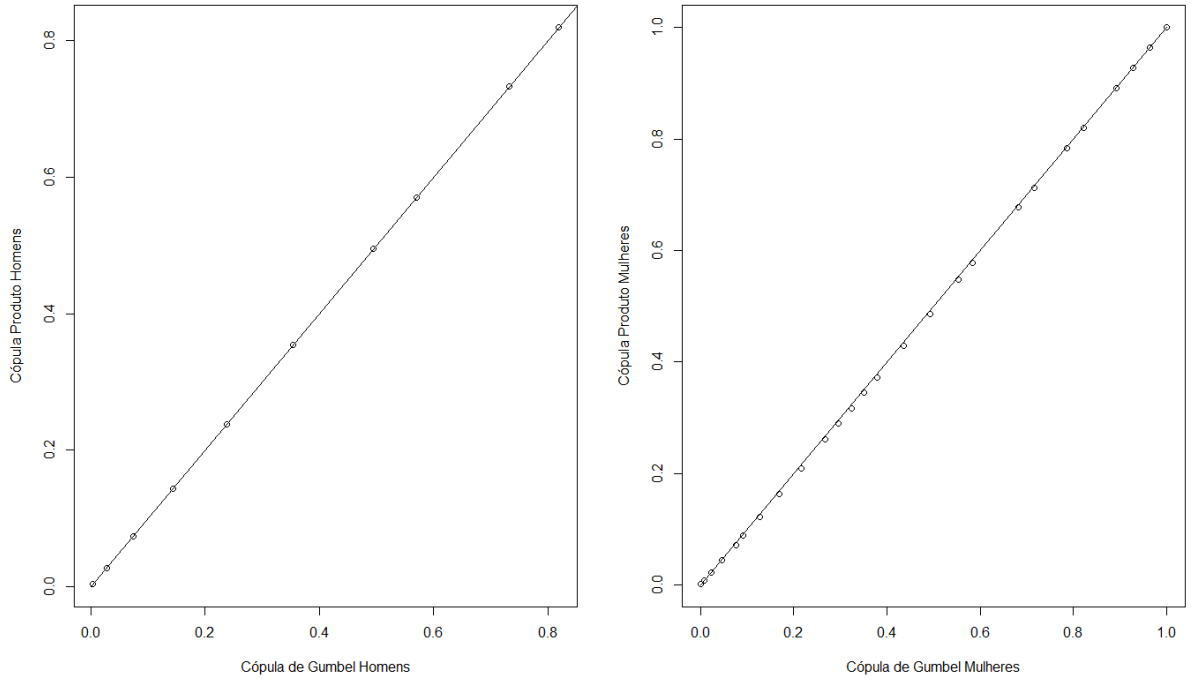


Figura 4.10: Gráfico da cópula de Gumbel versus cópula Produto por sexo.

4.2.2 Modelo com cópula de Clayton

Usando o método jackknife, discutido na seção 3.2.1, a estimativa do viés do estimador $\hat{\theta} = \frac{2\hat{\tau}}{1-\hat{\tau}}$ foi de $vies_{jack} = 0,0018$ e o erro padrão foi de $se_{jack} = 0,061$. Fazendo a relação $\frac{vies_{jack}}{se_{jack}} = 0,0297$, temos que o viés do estimador $\hat{\theta} = \frac{2\hat{\tau}}{1-\hat{\tau}}$ pode ser desprezado. O valor estimado de θ na cópula de Clayton foi $\hat{\theta} = 0,083$. Assim a cópula de Clayton para o modelo é

$$C_{\hat{\theta}}(\hat{S}_1(t_1; Z), \hat{S}_2(t_2; Z)) = \left(\left(\hat{S}_1(t_1; Z) \right)^{-0.083} + \left(\hat{S}_2(t_2; Z) \right)^{-0.083} - 1 \right)^{-\frac{1}{0.083}}$$

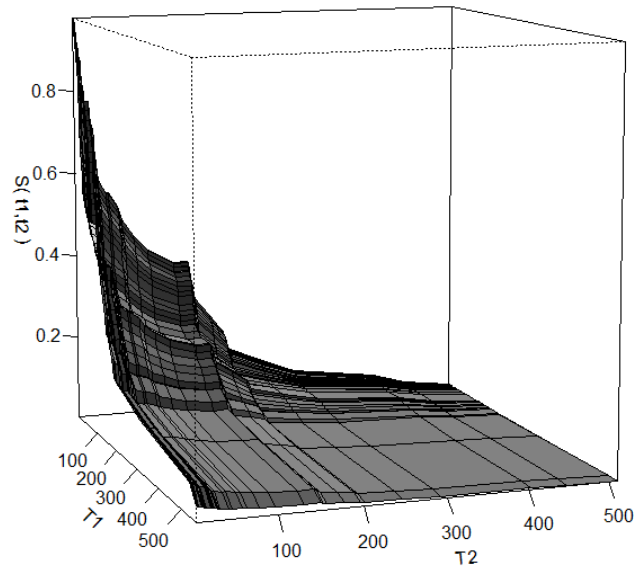


Figura 4.11: Gráfico da função de sobrevivência do modelo de Cox bivariado através da cópula de Clayton.

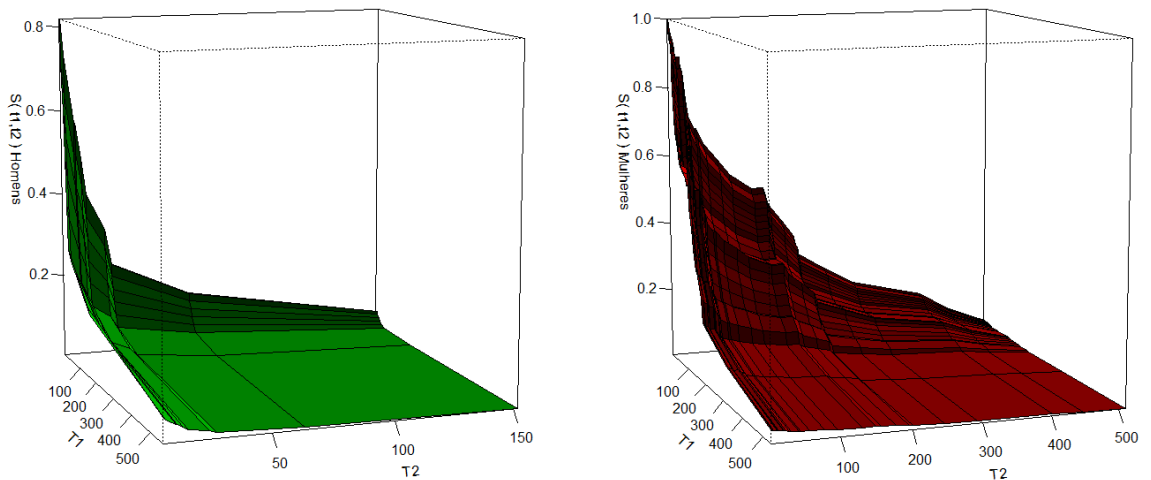


Figura 4.12: Gráfico da função de sobrevivência do modelo de Cox bivariado através da cópula de Clayton por sexo.

Como o valor de $\hat{\theta} = 0,08$ é muito próximo de 0, temos que as variáveis T_1 e T_2 são independentes, como mostra a figura 4.13, mantendo também essa característica quando consideramos o sexo. Os gráficos mostrados nas figura 4.13 e figura 4.14 foram construídos plotando-se $C_{\hat{\theta}}(\hat{S}_1(t_1; Z), \hat{S}_2(t_2; Z))$ contra $\Pi(\hat{S}_1(t_1; Z), \hat{S}_2(t_2; Z)) = \hat{S}_1(t_1; Z)\hat{S}_2(t_2; Z)$.

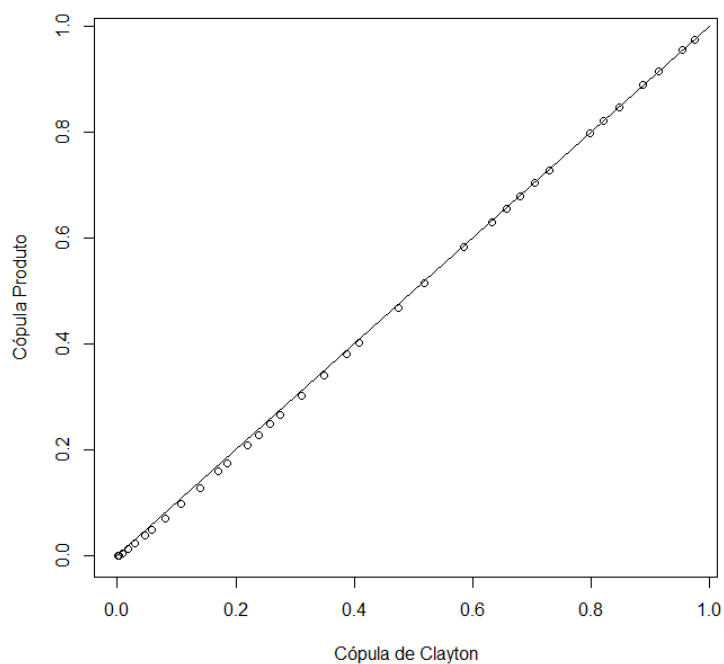


Figura 4.13: Gráfico da cópula de Clayton versus cópula Produto por sexo.

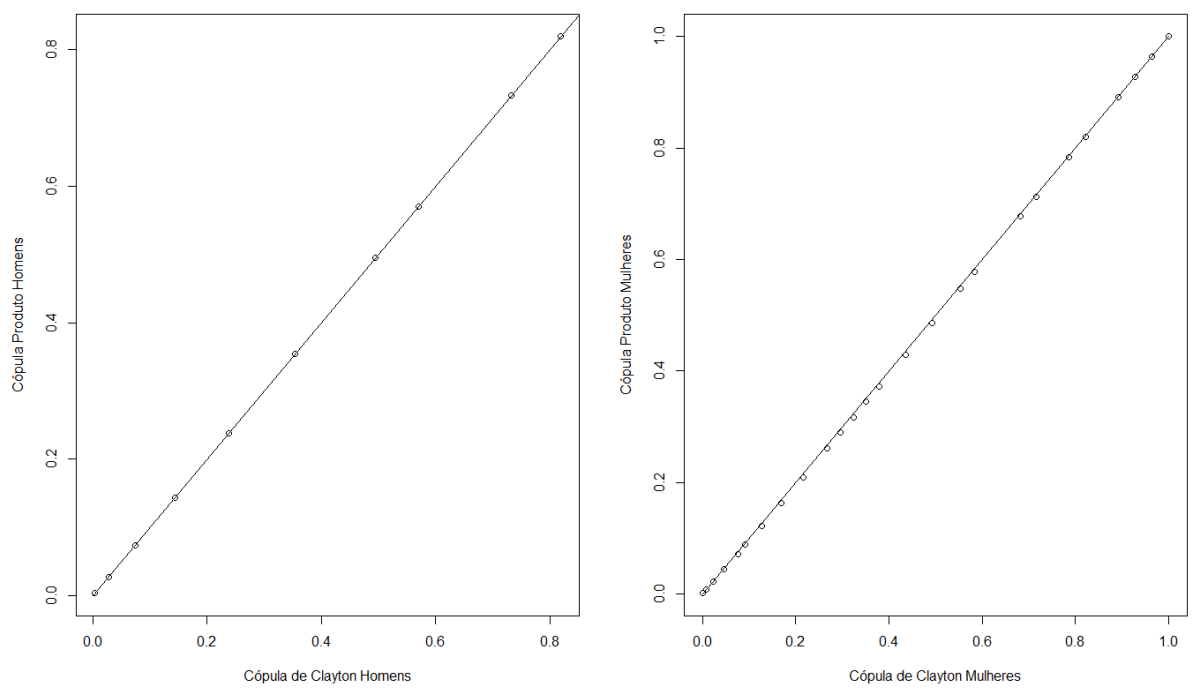


Figura 4.14: Gráfico da cópula de Clayton versus cópula Produto.

Capítulo 5

Conclusão

Este trabalho visou estudar modelos de Cox bivariados através do uso de cópulas para dados com censura. O modelo de Cox tem a vantagem de não precisarmos supor uma distribuição para os tempos de falha, mas exige que os riscos sejam proporcionais, então um dos pressupostos do modelo é a verificação da proporcionalidade dos riscos através de métodos gráficos ou testes de hipóteses nos resíduos.

Foram usadas as cópulas de Gumbel e de Clayton para modelar a dependência dos tempos de falha, em que os parâmetros de dependência foram estimados calculando o coeficiente τ de Kendall para dados censurados e dada a relação deste coeficiente e o parâmetro de dependência das cópulas. Foi verificado o viés e o erro padrão dos estimadores através do método jackknife que se mostrarão consistentes e acurados. Através de técnicas gráficas foi verificada a independência entre os dois tempos de falha, como acusava o valor dos parâmetros. Assim, uma alternativa seria estudar as duas variáveis separadamente, como se fossem de indivíduos diferentes.

Anexo

Sobre o parâmetro de dependência na cópula de Clayton

A cópula de Clayton é dada por

$$C_\theta(u, v) := (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}$$

Quando $\theta \rightarrow 0$ temos que $C_\theta(u, v) \rightarrow \Pi(u, v) = uv$

Prova

$$\lim_{\theta \rightarrow 0} (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}} = \lim_{\theta \rightarrow 0} e^{-\frac{\log(u^{-\theta} + v^{-\theta} - 1)}{\theta}}$$

Como a base está fixa, temos que o resultado do limite acima é igual à base elevada ao limite do expoente. No expoente temos um limite da forma $\frac{0}{0}$, logo, podemos usar a regra de L'hospital. Assim temos,

$$\begin{aligned} \lim_{\theta \rightarrow 0} -\frac{\log(u^{-\theta} + v^{-\theta} - 1)}{\theta} &= -\lim_{\theta \rightarrow 0} \frac{-\log(u)u^{-\theta} - \log(v)v^{-\theta}}{(u^{-\theta} + v^{-\theta} - 1)} \\ &= -\frac{-\log(u) - \log(v)}{1} = \log(u) + \log(v). \end{aligned}$$

Portanto,

$$\lim_{\theta \rightarrow 0} (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}} = e^{\log(u) + \log(v)} = e^{\log(u)} e^{\log(v)} = uv \quad \square$$

Tabelas

Modelo com cópula de Gumbel

Tabela A.1. Valores de $\hat{S}(t_1, t_2)$ usando cópula de Gumbel com $\hat{\tau} = 0.04$ e $\hat{\theta} = 1.041$.

T_1 em semanas	T_2 em semanas	$\hat{S}(t_1, t_2)$
2	5	0,9744
4	7	0,9560
5	8	0,9163
6	8	0,9163
7	8	0,8908
8	12	0,8500
9	12	0,8247
13	13	0,8022
16	15	0,7340
22	15	0,7094
24	16	0,6844
24	17	0,6627
26	22	0,6377
27	23	0,5906
30	25	0,5233
30	25	0,5233
34	28	0,4782
38	30	0,4112
39	30	0,3910
40	33	0,3510
53	43	0,3130
54	46	0,3130
66	58	0,2760
70	63	0,2586
108	78	0,2380
113	96	0,2180
114	130	0,1828
119	149	0,1671
132	152	0,1361
141	156	0,1038
152	177	0,0766
154	196	0,0542
159	201	0,0424
185	245	0,0260
190	292	0,0142
318	402	0,0064
536	447	0,0019
562	511	0,0002

Considerando a análise com separação por sexo temos que, quando sexo é feminino, o valor de $\hat{\tau} = 0,026$ que resulta em um $\hat{\theta} = 1.027$.

Tabela A.2. Valores $\hat{S}(t_1, t_2)$ usando cópula de Gumbel com $\hat{\theta} = 1.027$ e sexo feminino.

T_1 em semanas	T_2 em semanas	$\hat{S}(t_1, t_2)$
5	5	1.0000
6	8	0.9636
7	8	0.9285
13	13	0.8926
24	15	0.8226
24	16	0.8226
26	22	0,7867
27	23	0.7165
30	25	0.6821
34	30	0.5842
38	30	0.5531
39	33	0.4928
53	43	0.4359
54	46	0.4359
66	58	0.3788
70	78	0.3513
108	96	0,3236
113	130	0.2960
114	149	0.2683
119	156	0.2157
132	177	0.1688
141	196	0.1275
152	201	0.0919
159	245	0.0755
185	292	0.0462
190	402	0.0240
318	447	0.0090
536	511	0.0013

Quando sexo é masculino, o valor de $\hat{\tau} = 0$ que resulta em um $\hat{\theta} = 1$.

Tabela A.3. Valores de $\hat{S}(t_1, t_2)$ usando cópula de Gumbel com $\hat{\tau} = 0$ e $\hat{\theta} = 1$

para o sexo masculino.

T_1 em semanas	T_2 em semanas	$\hat{S}(t_1, t_2)$
2	7	0.8187
4	8	0.7326
8	12	0.5706
9	12	0.4946
16	15	0.3544
22	17	0.2376
30	25	0,1441
40	28	0.0740
154	63	0.0272
562	152	0.0037

Modelo com cópula de Clayton

Na cópula de Clayton, quando $\hat{\tau} = 0$ tem-se que $\hat{\theta} = 0$.

Tabela A.4. Valores de $\hat{S}(t_1, t_2)$ usando cópula de Clayton com $\hat{\tau} = 0$ e $\hat{\theta} = 0$ e

sexo masculino.

T_1 em semanas	T_2 em semanas	$\hat{S}(t_1, t_2)$
2	7	0.8187
4	8	0.7326
8	12	0.5706
9	12	0.4946
16	15	0.3544
22	17	0.2376
30	25	0,1441
40	28	0.0740
154	63	0.0272
562	152	0.0037

Tabela A.5. Valores de $\hat{S}(t_1, t_2)$ usando cópula de Clayton com $\hat{\tau} = 0.026$ e

$\hat{\theta} = 0.053$ e sexo feminino.

T_1 em semanas	T_2 em semanas	$\hat{S}(t_1, t_2)$
5	5	1.0000
6	8	0.9636
7	8	0.9274
13	13	0.8911
24	15	0.8201
24	16	0.8201
26	22	0,7839
27	23	0.7132
30	25	0.6787
34	30	0.5807
38	30	0.5498
39	33	0.4897
53	43	0.4333
54	46	0.4359
66	58	0.3768
70	78	0.3500
108	96	0,3224
113	130	0.2951
114	149	0.2680
119	156	0.2163
132	177	0.1703
141	196	0.1287
152	201	0.0947
159	245	0.0784
185	292	0.0492
190	402	0.0267
318	447	0.0108
536	511	0.0019

Tabela A.6. Valores de $\hat{S}(t_1, t_2)$ usando cópula de Clayton com $\hat{\tau} = 0.04$ e

$\hat{\theta} = 0.083$.

T_1 em semanas	T_2 em semanas	$\hat{S}(t_1, t_2)$
2	5	0,9744
4	7	0.9549
5	8	0.9146
6	8	0.9146
7	8	0.8883
8	12	0.8469
9	12	0,8210
13	13	0.7983
16	15	0.7294
22	15	0.7045
24	16	0.6793
24	17	0.6576
26	22	0.6325
27	23	0.5855
30	25	0.5186
30	25	0.5186
34	28	0,4740
38	30	0,4080
39	30	0.3882
40	33	0.3489
53	43	0.3118
54	46	0,3118
66	58	0.2757
70	63	0,2587
108	78	0,2387
113	96	0.2192
114	130	0.1849
119	149	0.1697
132	152	0.1395
141	156	0.1079
152	177	0.0812
154	196	0,0580
159	201	0,0469
185	245	0,0301
190	292	0,0175
318	402	0,0087
536	447	0,0032
562	511	0,0005

Bibliografia

- [1]. Cox, D. *Regression models and life-tables*. JRSS B, v. 34, n. 2, p. 187-220, 1972.
- [2]. Cox, D. *Partial likelihood*. Biometrika, v. 62, n. 2, p. 269-276, 1975.
- [3]. Efron, B., Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, London.
- [4]. Nuñez et al. *Large sample properties for a class of copulas in bivariate survival analysis*. Metrika (Heidelberg), v. 76, p. 997-1015, 2013.
- [5]. Nelsen, R. B. (2006). *An introduction to copulas* (2nd ed.). New York ; Berlin ; Heidelberg: Springer.
- [6]. Oakes, D. *Bivariate survival induced by frailties*. JASA, v. 84, n. 406, p. 487-493, 1989.
- [7]. Achibi, M., Broniatowski, M. (2010). *Bivariate Cox Model and Copulas*. arXiv, 1 a 34.
- [8]. Fachini, J. B. *Modelos de regressão com e sem fração de cura para dados bivariados em análise de sobrevivência*. 2011. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2011.

- [9]. Wang, W., Wells, M. T (2000). *Estimation of Kendall's tau under Censoring*.
Statist. Sinica 10,p. 1199 a 1215.
- [10]. Colosimo, E.A.; Giolo, S.R. *Análise de sobrevivência aplicada*. São Paulo:
Blucher, 2006. 392p.
- [11]. Lawless, J.F. *Statistical Models and Methods for Lifetime Data*. United States:
Jhon While & Sons, 2003. 350p.
- [12]. R Core Team (2014). R:A language and environment for statistical computing,
Vienna, Austria, Disponível em: <http://www.R-project.org/>.