

UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA  
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA UFSCar-USP

MURILO CANTONI

**FULL BAYESIAN SIGNIFICANCE TEST PARA DADOS DE SOBREVIVÊNCIA BIVARIADOS:  
SELEÇÃO DE MODELOS ENCAIXADOS DA CÓPULA PVF**

Tese apresentada ao Departamento de Estatística – Des/UFSCar e ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre ou Doutor em Estatística - Programa Interinstitucional de Pós-Graduação em Estatística UFSCar-USP.

Orientador: Prof. Dr. Adriano Polpo de Campos

**São Carlos  
Agosto de 2021**



UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA  
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA UFSCar-USP

MURILO CANTONI

**FULL BAYESIAN SIGNIFICANCE TEST FOR BIVARIATE SURVIVAL DATA: PVF COPULA'S  
NESTED MODEL SELECTION**

Doctoral dissertation submitted to the Departamento de Estatística – Des/UFSCar and to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP, in partial fulfillment of the requirements for the degree of the Doctorate Joint Graduate Program in Statistics DES-UFSCar/ICMC-USP. FINAL VERSION

Advisor: Prof. Dr. Adriano Polpo de Campos

**São Carlos  
August 2021**



# UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia  
Programa Interinstitucional de Pós-Graduação em Estatística

---

## Folha de Aprovação

---

Defesa de Tese de Doutorado do candidato Murilo Cantoni, realizada em 28/06/2021.

### Comissão Julgadora:

Prof. Dr. Adriano Polpo de Campos (UFSCar)

Prof. Dr. José Santos Romeo Núñez (Massey)

Profa. Dra. Verónica Andrea González López (UNICAMP)

Prof. Dr. Victor Fossaluza (USP)

Profa. Dra. Teresa Cristina Martins Dias (UFSCar)

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa Interinstitucional de Pós-Graduação em Estatística.



# AGRADECIMENTOS

---

---

Em primeiro lugar, agradeço a Deus pelo dom da vida e pela motivação ao longo destes anos de estudo.

Agradeço aos meus pais, Vera e Mauro, por todo carinho e dedicação, pelos ensinamentos e por terem batalhado muito para que eu pudesse ter a oportunidade de estudar e me formar. Agradeço também à minha irmã, Laís, por compartilhar momentos de descontração e alegria, assim como sempre me apoiar e me aconselhar em momentos delicados. Em especial, gostaria de agradecer aos meus avós, Lázara e Osvaldo, por todos os cafés aos domingos, por sempre me receberem com um abraço apertado e transmitirem tanta alegria. À Silvana, minha noiva, pelo amor, companheirismo e por cada momento que compartilhamos desde o momento em que nos conhecemos.

Ao meu orientador, Adriano Polpo de Campos, que acompanhou minha trajetória desde a graduação. Agradeço não somente por compartilhar conhecimento, incentivar o pensamento crítico e ser atencioso nas discussões de ideias, mas também pela paciência, confiança e pelos conselhos. À Teresa Cristina Martins Dias, pelas sugestões e correções que foram valiosas no processo de construção deste trabalho. Ao meu tutor no período em que fiz parte do Programa de Educação Tutorial, Pedro Ferreira Filho, pelos ensinamentos, pela amizade e por ter contribuído no meu desenvolvimento como pessoa.

Aos professores do Departamento de Estatística da UFSCar e do Programa Interinstitucional de Pós-Graduação em Estatística UFSCar-USP, que de forma direta ou indireta contribuíram para minha formação. Aos funcionários do Departamento de Estatística da UFSCar, em especial à Maria Isabel Rinaldo Pessoa de Araujo, pela atenção, solicitude e por sempre tornar as coisas mais simples e objetivas.

Aos amigos que fizeram parte da minha trajetória, por compartilhar bons momentos e contribuir para que o caminho até aqui fosse menos árduo. Em especial, gostaria de agradecer ao Eduardo, pela inestimável amizade, ao Gabriel, Sérgio e senhor Carlos, por me fazerem sentir em casa.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.



# RESUMO

CANTONI, M. **Full Bayesian Significance Test para dados de sobrevivência bivariados: seleção de modelos encaixados da Cópula PVF**. 2021. 70 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2021.

Investigar e modelar a dependência existente em um conjunto de variáveis aleatórias é um tema amplamente discutido em estatística. Neste contexto, o uso de cópulas torna-se interessante por tratar-se de uma abordagem flexível que permite estudar, em um primeiro momento, as distribuições univariadas e, posteriormente, a estrutura de dependência. Em problemas práticos, conhecer a cópula que melhor conecta as distribuições marginais à função de distribuição conjunta não é uma tarefa simples. Em geral, vários modelos são ajustados de acordo com o tipo de dependência existente no conjunto de dados e algum critério de seleção é aplicado com o intuito de escolher o “melhor modelo”. Neste trabalho, utilizamos a família Arquimediana de dois parâmetros Power Variance Function (PVF), que inclui as cópulas de Clayton, Gumbel e Gaussiana Inversa (IG) como casos particulares ou casos limites, pois oferece uma abordagem unificada e flexível para ajustar modelos de cópulas amplamente utilizadas e propomos a utilização do Full Bayesian Significance Test (FBST) como critério de seleção de modelos encaixados. Validamos os resultados através de um estudo de simulação e ilustramos a utilidade da metodologia usando dados sobre os tempos de apendicectomia para gêmeos adultos.

**Palavras-chave:** Análise de Sobrevivência Bivariada, Cópulas Arquimedianas, Dependência, Full Bayesian Significance Test, Seleção de Modelos.





# ABSTRACT

CANTONI, M. **Full Bayesian Significance Test for bivariate survival data: PVF Copula's nested model selection**. 2021. 70 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2021.

The investigation and modeling of the existing dependence in a set of random variables is a widely discussed topic in statistics. In this context, the use of copulas becomes interesting because it is a flexible approach that allows to study, in a first moment, the univariate distributions and, later, the dependency structure. In practical problems, knowing the copula that best connects marginal distributions to the joint distribution function is not a simple task. In general, several models are adjusted according to the type of dependence existing in the data set and some selection criteria are applied in order to choose the "best model". In this work, we use the two-parameter Archimedean family of Power Variance Function (PVF), which includes the Clayton, Gumbel and Inverse Gaussian (IG) copulas as special or limiting cases, once it offers a unified and flexible approach to adjust widely used copula models and we propose the use of the Full Bayesian Significance Test (FBST) as a model selection criterion. We validated the results through a simulation study and illustrated the usefulness of the methodology using data on appendectomy times for adult twins.

**Keywords:** Bivariate Survival Analysis, Archimedean Copulas, Dependence, Full Bayesian Significance Test, Model Selection.



# LISTA DE ILUSTRAÇÕES

---

---

Figura 1 – Estimativa de Kaplan-Meier das sobrevivências marginais do vetor $(T_1, T_2)$ para gêmeos MZ e DZ. . . . .	19
Figura 2 – Esboço das cópulas produto ( $\Pi$ ), limite inferior ( $W$ ) e superior ( $M$ ) de Fréchet-Hoeffding para $d=2$ . . . . .	24
Figura 3 – Esboço da cópula construída a partir da distribuição exponencial bivariada de Gumbel. . . . .	26
Figura 4 – Esboço da cópula PVF para diferentes valores de $\alpha$ e $\eta$ . . . . .	36
Figura 5 – Cópulas escolhidas para o estudo de simulação. . . . .	41
Figura 6 – Distribuições marginais escolhidas para o estudo de simulação. . . . .	41
Figura 7 – EAIC, EBIC, DIC e $e$ -value obtidos nas 100 réplicas do estudo de simulação para o caso de marginais i.d., número de amostral igual a 1000, 75% de censura em cada marginal e modelo verdadeiro Gumbel. . . . .	44
Figura 8 – EAIC, EBIC, DIC e $e$ -value obtidos nas 100 réplicas do estudo de simulação para o caso de marginais não i.d., número de amostral igual a 1000, 75% de censura em cada marginal e modelo verdadeiro Clayton. . . . .	60
Figura 9 – EAIC, EBIC, DIC e $e$ -value obtidos nas 100 réplicas do estudo de simulação para o caso de marginais não i.d., número de amostral igual a 1000, 75% de censura em cada marginal e modelo verdadeiro Gumbel. . . . .	61
Figura 10 – EAIC, EBIC, DIC e $e$ -value obtidos nas 100 réplicas do estudo de simulação para o caso de marginais não i.d., número de amostral igual a 1000, 75% de censura em cada marginal e modelo verdadeiro IG. . . . .	62
Figura 11 – EAIC, EBIC, DIC e $e$ -value obtidos nas 100 réplicas do estudo de simulação para o caso de marginais não i.d., número de amostral igual a 1000, 20% de censura em cada marginal e modelo verdadeiro Clayton. . . . .	63
Figura 12 – AIC, EBIC, DIC e $e$ -value obtidos nas 100 réplicas do estudo de simulação para o caso de marginais não i.d., número de amostral igual a 1000, 20% de censura em cada marginal e modelo verdadeiro Gumbel. . . . .	64
Figura 13 – AIC, EBIC, DIC e $e$ -value obtidos nas 100 réplicas do estudo de simulação para o caso de marginais não i.d., número de amostral igual a 1000, 20% de censura em cada marginal e modelo verdadeiro IG. . . . .	65
Figura 14 – EAIC, EBIC, DIC e $e$ -value obtidos nas 100 réplicas do estudo de simulação para o caso de marginais i.d., número de amostral igual a 1000, 75% de censura em cada marginal e modelo verdadeiro Clayton. . . . .	66

Figura 15 – EAIC, EBIC, DIC e <i>e-value</i> obtidos nas 100 réplicas do estudo de simulação para o caso de marginais i.d., número de amostral igual a 1000, 75% de censura em cada marginal e modelo verdadeiro IG. . . . .	67
Figura 16 – EAIC, EBIC, DIC e <i>e-value</i> obtidos nas 100 réplicas do estudo de simulação para o caso de marginais i.d., número de amostral igual a 1000, 20% de censura em cada marginal e modelo verdadeiro Clayton. . . . .	68
Figura 17 – EAIC, EBIC, DIC e <i>e-value</i> obtidos nas 100 réplicas do estudo de simulação para o caso de marginais i.d., número de amostral igual a 1000, 20% de censura em cada marginal e modelo verdadeiro Gumbel. . . . .	69
Figura 18 – EEAIC, EBIC, DIC e <i>e-value</i> obtidos nas 100 réplicas do estudo de simulação para o caso de marginais i.d., número de amostral igual a 1000, 20% de censura em cada marginal e modelo verdadeiro IG. . . . .	70

# LISTA DE TABELAS

---

---

Tabela 1 – Percentual de acerto da escolha do modelo mais adequado aos dados por cada critério de seleção de modelos. . . . .	42
Tabela 2 – Distribuição de pares de gêmeos australianos por apendicectomia. . . . .	44
Tabela 3 – Média, mediana, desvio padrão, limite inferior e superior do intervalo HPD de 95% considerando distribuição Weibull para as sobrevivências marginais para gêmeos MZ. . . . .	45
Tabela 4 – Média, mediana, desvio padrão, limite inferior e superior do intervalo HPD de 95% considerando distribuição Exponencial por Partes para as sobrevivências marginais para gêmeos MZ. . . . .	46
Tabela 5 – Tau de Kendall estimado e critérios de seleção de modelos para gêmeos MZ.	48
Tabela 6 – Média, mediana, desvio padrão, limite inferior e superior do intervalo HPD de 95% considerando distribuição Weibull para as sobrevivências marginais para gêmeos DZ. . . . .	49
Tabela 7 – Média, mediana, desvio padrão, limite inferior e superior do intervalo HPD de 95% considerando distribuição Exponencial por Partes para as sobrevivências marginais para gêmeos DZ. . . . .	50
Tabela 8 – Tau de Kendall estimado e critérios de seleção de modelos para gêmeos DZ.	52



# SUMÁRIO

---

---

1	INTRODUÇÃO	17
2	MÉTODOS	21
2.1	Cópulas	21
2.1.1	<i>Exemplos</i>	23
2.1.2	<i>Um Método para Obtenção de Cópulas</i>	25
2.2	Medidas de Dependência	26
2.2.1	<i>Concordância</i>	26
2.2.2	<i>Tau de Kendall</i>	27
2.2.3	<i>Rho de Spearman</i>	29
2.3	Análise de Sobrevivência	30
2.3.1	<i>Censura</i>	30
2.3.2	<i>Verossimilhança</i>	31
2.4	Seleção de Modelos	32
2.4.1	<i>EAIC, EBIC e DIC</i>	32
2.4.2	<i>FBST</i>	33
2.5	FBST para Cópula Power Variance Function (PVF)	34
3	APLICAÇÕES	39
3.1	Definições Preliminares	39
3.2	Estudo de Simulação	40
3.3	Dados de Apendicectomia	43
3.3.1	<i>Gêmeos Monozigóticos (MZ)</i>	45
3.3.2	<i>Gêmeos Dizigóticos (DZ)</i>	49
4	CONSIDERAÇÕES FINAIS	53
4.1	Sugestões para Pesquisas Futuras	54
	REFERÊNCIAS	57
APÊNDICE A	COMPORTAMENTO DOS CRITÉRIOS DE SELEÇÃO NO ESTUDO DE SIMULAÇÃO	59





---

## INTRODUÇÃO

---

Em linhas gerais, a análise de sobrevivência é uma área da estatística que se dedica ao estudo do tempo até a ocorrência de determinado evento. A caracterização de um problema de sobrevivência se dá basicamente por dois fatores: o tempo de ocorrência e a censura. O tempo de ocorrência geralmente é composto por três principais elementos, sendo o primeiro chamado de tempo inicial do estudo (ou referência), no qual todos os indivíduos devem ser comparáveis, o segundo elemento está associado à escala de medida, que usualmente é o tempo (dias, meses, anos, etc.), mas também pode ser a quantidade de ciclos, quilometragem, entre outros, e, por fim, o terceiro elemento relacionado ao tempo de ocorrência é o evento de interesse, podendo ser a falha de algum componente em um circuito, a recidiva de alguma doença, entre outros. A censura, por sua vez, é caracterizada por observações parciais ou incompletas e existem diversas razões para ocorrer na prática, como a perda de contato de um médico com seu paciente ou em função de estudos de longa duração, por exemplo. Mais detalhes sobre caracterização de um problema de sobrevivência podem ser consultados em (HOSMER; LEMESHOW, 1999), (LAWLESS, 2011) e (COLOSIMO; GIOLO, 2006).

Existem três principais tipos de censura: à direita, à esquerda e intervalar. A censura à direita acontece em situações nas quais o tempo de falha é superior ao tempo observado, a censura à esquerda, por sua vez, ocorre em situações nas quais o tempo de falha é inferior ao tempo observado e, por fim, a censura intervalar acontece em casos nos quais a informação disponível é que a falha ocorreu em um intervalo mas não se sabe o momento exato. Nos casos de censura à direita ou à esquerda, a representação dos dados de sobrevivência para um indivíduo  $i$  ( $i = 1, \dots, n$ , em que  $n$  é o número de indivíduos no estudo) é usualmente dada por  $(t_i, \delta_i, \mathbf{x}_i)$ , sendo  $t_i$  o tempo de falha,  $\delta_i$  o indicador de falha ou censura, isto é,

$$\delta_i = \begin{cases} 0 & \text{se } t_i \text{ é um tempo censurado,} \\ 1 & \text{se } t_i \text{ é um tempo de falha,} \end{cases}$$

e  $\mathbf{x}_i$  representa as covariáveis (idade, sexo, etc.). Em casos de censura intervalar, a representação é dada por  $(l_i, u_i, \delta_i, \mathbf{x}_i)$ , em que  $l_i$  e  $u_i$  são os limites inferior e superior do intervalo que foi observado para o  $i$ -ésimo indivíduo, respectivamente.

As aplicações de análise de sobrevivência ocorrem em diversas áreas do conhecimento. Um exemplo interessante em medicina pode ser visto em (TURNBULL; JR; HU, 1974), que investiga se o Programa de Transplante de Coração de Stanford prolongou o tempo de vida de pacientes que receberam novo coração. Em engenharia, uma aplicação de análise de sobrevivência pode ser vista em (MEEKER, 1987), que estuda a ocorrência de falha em circuitos integrados. Em geral, uma pequena parcela de componentes fabricados possuem defeitos que não são perceptíveis em uma simples inspeção e posteriormente podem causar algum tipo de falha. Desta forma, foram conduzidos testes de duração limitada para estimar a proporção de itens defeituosos.

Em algumas situações, no entanto, pode existir o interesse em entender o comportamento de dois ou mais tempos de sobrevivência de forma conjunta. Neste cenário, a notação utilizada geralmente é semelhante ao caso univariado. Para censura à direita ou à esquerda, denota-se, no caso bivariado,  $(\mathbf{t}_i, \boldsymbol{\delta}_i, \mathbf{x}_i)$ , em que  $\mathbf{t}_i = (t_{i1}, t_{i2})$ ,  $\boldsymbol{\delta}_i = (\delta_{i1}, \delta_{i2})$  e  $\mathbf{x}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2})$ . Para censura intervalar, denota-se  $(\mathbf{l}_i, \mathbf{u}_i, \boldsymbol{\delta}_i, \mathbf{x}_i)$ , em que  $\mathbf{l}_i = (l_{i1}, l_{i2})$  e  $\mathbf{u}_i = (u_{i1}, u_{i2})$ . Um exemplo de problema bidimensional pode ser encontrado em (GROUP *et al.*, 1976). Pacientes com retinopatia proliferativa recebem um tratamento de fotocoagulação em um dos olhos (escolhido aleatoriamente) e o outro olho é considerado como controle. O evento principal é a perda de acuidade visual. Outro exemplo pode ser visto em (DUFFY; MARTIN; MATHEWS, 1990), que investiga o risco de aparecimento de várias doenças, incluindo apendicite aguda, em gêmeos adultos registrados no Australian NH&MRC Twin Registry (ATR). O objetivo é avaliar se este risco é diferente para gêmeos monozigóticos (MZ) e dizigóticos (DZ). A Figura 1 mostra a estimativa de Kaplan–Meier (KAPLAN; MEIER, 1958) para este estudo. Se o comportamento dos dois ou mais tempos de sobrevivência fossem independentes, análises univariadas poderiam ser conduzidas sem causar impacto (ou viés) nos resultados. No entanto, na maior parte dos casos existe uma ligação entre os tempos de sobrevivência, que usualmente é chamada dependência, e somente análises univariadas não são mais capazes de produzir resultados plausíveis.

Dentre as alternativas disponíveis para tratar dados dessa natureza, a metodologia de cópulas, que inicialmente foi discutida em (SKLAR, 1959), pode ser útil. O ganho em utilizar esta metodologia está no fato de que se pode, em um primeiro momento, buscar informações das funções de distribuição (ou sobrevivência) marginais e, posteriormente, da estrutura de dependência inerente ao problema. Neste sentido, as funções cópulas podem ser vistas como um elo de ligação entre as funções de distribuição marginais (ou sobrevivência) e a função de distribuição conjunta (ou sobrevivência). Em termos matemáticos, o Teorema de Sklar garante que

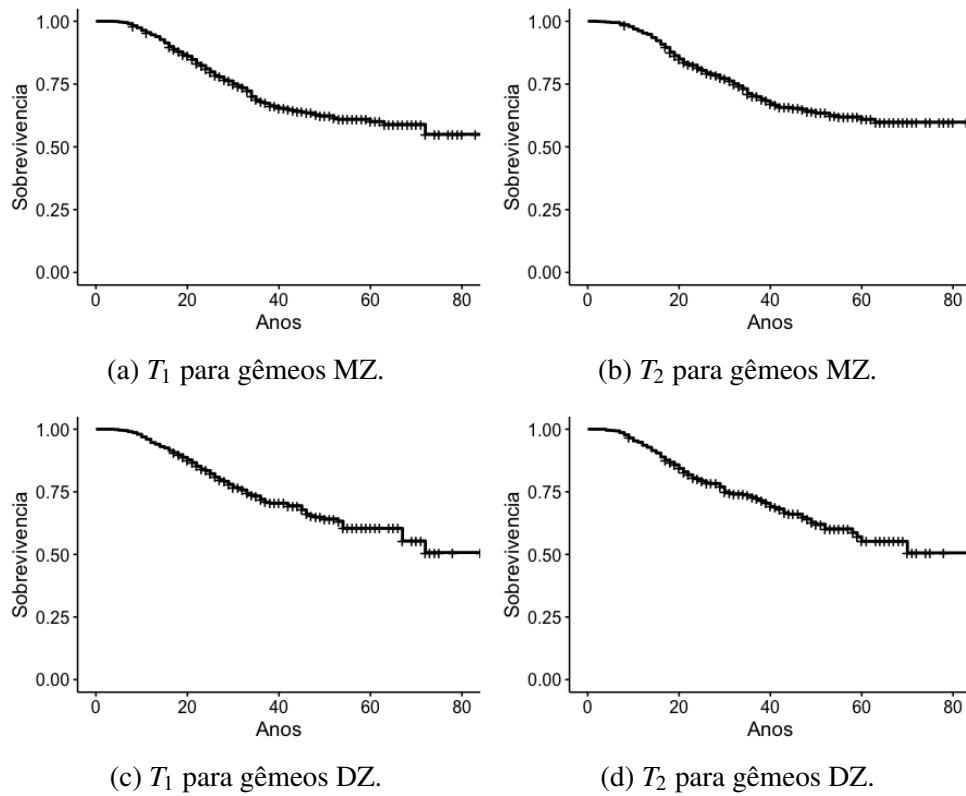


Figura 1 – Estimativa de Kaplan-Meier das sobrevivências marginais do vetor  $(T_1, T_2)$  para gêmeos MZ e DZ.

$$F(t_1, \dots, t_d) = C(F_1(t_1), \dots, F_d(t_d)),$$

em que  $F$  representa a função de distribuição conjunta,  $C$  representa a função cópula,  $F_1$  representa a função de distribuição marginal de  $T_1$ ,  $F_2$  representa a função de distribuição marginal de  $T_2$  e assim sucessivamente, e  $d$  representa a dimensão do problema. Para o caso bivariado,  $H(t_1, t_2) = C(F_1(t_1), F_2(t_2))$  e, a partir de algumas manipulações algébricas, a seguinte relação é obtida:

$$S(t_1, t_2) = C(1 - S_1(t_1), 1 - S_2(t_2)) + S_1(t_1) + S_2(t_2) - 1,$$

em que  $S$  representa a sobrevivência conjunta,  $C$  representa a função cópula,  $S_i$  representa a função de sobrevivência marginal de  $T_i$ ,  $i = 1, 2$ . Este resultado permite que a sobrevivência conjunta possa ser construída a partir das sobrevivências marginais e da estrutura de dependência entre  $T_1$  e  $T_2$ .

Embora o Teorema de Sklar seja útil no sentido de diminuir a complexidade do problema e, para o caso em que as variáveis aleatórias são contínuas, garantir, sob certas condições, que a função cópula  $C$  seja única, não é explicitada a forma para se construir diretamente  $C$ . Desta forma, em problemas práticos é comum o ajuste de diversos modelos de cópulas que acomodam

o tipo de dependência existente nos dados e a aplicação de um ou mais critérios de seleção de modelos. Neste trabalho, utilizamos a cópula PVF, que é construída a partir da distribuição Power Variance Function (PVF) de três parâmetros, sendo esta uma extensão da distribuição estável positiva (HOUGAARD, 1986). A vantagem em utilizar essa cópula está no fato de que, além da flexibilidade, algumas cópulas amplamente utilizadas na literatura, como Gumbel, Clayton e Gaussiana Inversa (IG), podem ser obtidas como casos particulares da cópula PVF.

Em termos de seleção de modelos, existe, na literatura, uma quantidade significativa de critérios no contexto bayesiano, como Expected Akaike Information Criterion (EAIC) (BROOKS *et al.*, 2002), Expected Bayesian Information Criterion (EBIC) (CARLIN; LOUIS, 2001) e Deviance Information Criterion (DIC) (SPIEGELHALTER *et al.*, 2002). Por outro lado, propomos a utilização de um teste de hipótese bayesiano como critério de seleção de modelos, o Full Bayesian Significance Test (FBST) (PEREIRA *et al.*, 2008), cuja ideia é avaliar o volume de um subespaço do espaço paramétrico, que possui apenas pontos com menor densidade a posteriori quando comparado à máxima densidade a posteriori no espaço da hipótese. A utilização deste critério permite o cálculo da evidência dos modelos menos complexos, ou casos particulares, contra o modelo mais abrangente, ou PVF, fornecendo, portanto, uma forma rigorosa da avaliação do modelo que acomoda melhor a variabilidade existente no conjunto de dados, ponderando aspectos como complexidade, por exemplo.

As principais contribuições deste trabalho são:

- Introduzir o FBST como critério de seleção para modelos encaixados da cópula PVF e explorar a utilização desta metodologia no contexto de análise de sobrevivência;
- Comparar a metodologia proposta com critérios de seleção de modelos amplamente utilizados.

Os capítulos seguintes da tese estão organizados do seguinte modo:

- Capítulo 2: base teórica e conceitos utilizados para o desenvolvimento do trabalho;
- Capítulo 3: aplicação da metodologia proposta em dados simulados e dados reais;
- Capítulo 4: conclusão, discussões gerais sobre a metodologia proposta e sugestões para pesquisas futuras.

## MÉTODOS

A relação de dependência entre duas (ou mais) variáveis aleatórias está completamente descrita na função de distribuição conjunta, que em muitos casos é desconhecida. Neste sentido, a teoria de cópulas pode ser útil pois permite estudar, em um primeiro momento, informação das distribuições marginais e, posteriormente, da estrutura de dependência.

Neste capítulo, definimos formalmente cópulas, relacionamos a teoria de cópulas com medidas de dependência e análise de sobrevivência, descrevemos alguns dos principais métodos de seleção de modelos no contexto bayesiano e propomos uma aplicação do teste de hipótese FBST no contexto de modelos encaixados.

### 2.1 Cópulas

**Definição 2.1.** Sejam  $J_1, \dots, J_d$ ,  $d \in \{2, 3, \dots\}$ , subconjuntos não vazios de  $\bar{\mathbb{R}}$ ,  $H : J = J_1 \times \dots \times J_d \rightarrow \bar{\mathbb{R}}$  uma função real d-variada e  $B = [\mathbf{a}, \mathbf{b}]$  uma d-caixa cujos vértices pertencem a  $J$ . O H-volume de  $B$ , denotado por  $V_H(B)$ , é definido como  $\Delta_{\mathbf{a}}^{\mathbf{b}} H(\mathbf{t}) = \Delta_{a_d}^{b_d} \dots \Delta_{a_1}^{b_1} H(\mathbf{t})$ , em que  $\Delta_{a_k}^{b_k} H(\mathbf{t}) = H(t_1, \dots, t_{k-1}, b_k, t_{k+1}, \dots, t_d) - H(t_1, \dots, t_{k-1}, a_k, t_{k+1}, \dots, t_d)$  é chamada de diferença de  $k$ -ésima ordem de  $H$  em  $B$ .

**Definição 2.2.** Uma função real d-variada  $H$  é d-crescente se  $V_H(B) \geq 0$  para todas as d-caixas vértices cujos pertencentes ao domínio de  $H$ .

Seja  $H : [0, 1]^2 \rightarrow [0, 1]$  uma função real bivariada definida como  $H(t_1, t_2) = \max(t_1, t_2)$  e considere  $B = [0, 1]^2$ . Pela Definição 2.1,  $V_H(B) = -1$ , ou seja,  $H$  não é bicrescente mesmo sendo não decrescente marginalmente. Por outro lado, considere  $H(t_1, t_2) = (2t_1 - 1)(2t_2 - 1)$ . Para todo  $B = [t_{11}, t_{12}] \times [t_{21}, t_{22}] \in [0, 1]^2$ , com  $t_{11} \leq t_{12}$  e  $t_{21} \leq t_{22}$ ,  $V_H(B) = 4(t_{12} - t_{11})(t_{22} - t_{21}) \geq 0$ , ou seja, pela Definição 2.2,  $H$  é bicrescente. Contudo,  $H$  é decrescente em  $t_1$  para todo  $t_2 \in (0, 1/2)$  e decrescente em  $t_2$  para todo  $t_1 \in (0, 1/2)$ . Este dois exemplos mostram que o fato de uma

função ser não decrescente marginalmente não implica que ela seja bicrescente, assim como o fato de uma função ser bicrescente não implica que marginalmente ela seja não decrescente.

**Definição 2.3.** Seja  $H : J = J_1 \times \dots \times J_d \rightarrow \bar{\mathbb{R}}$  uma função real d-variada. Considere  $a_i = \min\{J_i\}$  e  $b_i = \max\{J_i\}$ ,  $i = 1, \dots, d$ . Desta forma, H é dita aplanada se  $H(\mathbf{t}) = 0, \forall \mathbf{t} \in J$  tal que  $t_k = a_k$  para pelo menos um  $k$ . Além disso, H possui marginais univariadas, denotadas por  $H_k$ , definidas como  $H_k(t) = H(b_1, \dots, b_{k-1}, t, b_{k+1}, \dots, b_d), \forall t \in J_k$ . As marginais de ordem superiores são definidas de maneira semelhante.

**Lema 1.** Seja  $H : J = J_1 \times \dots \times J_d \rightarrow \bar{\mathbb{R}}$  uma função real d-variada aplanada e d-crescente em  $J$ . Então, H é não decrescente em cada argumento, isto é, se  $(t_1, \dots, t_{k-1}, x, t_{k+1}, \dots, t_d)$  e  $(t_1, \dots, t_{k-1}, y, t_{k+1}, \dots, t_d)$  estão em  $J$  e  $x < y$ , então

$$H(t_1, \dots, t_{k-1}, x, t_{k+1}, \dots, t_d) \leq H(t_1, \dots, t_{k-1}, y, t_{k+1}, \dots, t_d).$$

**Definição 2.4.** Uma cópula d-dimensional é uma função  $C : [0, 1]^d \rightarrow [0, 1]$ , aplanada, d-crescente e com marginais (unidimensionais), denotadas por  $C_k, k = 1, \dots, d$ , que satisfazem  $C_k(u) = u, \forall u \in [0, 1]$ .

Alternativamente, uma cópula pode ser definida a partir de um contexto de probabilidade.

**Definição 2.5.** Uma cópula d-dimensional é uma função de distribuição  $C : [0, 1]^d \rightarrow [0, 1]$  de um vetor aleatório  $\mathbf{U} = (U_1, \dots, U_d)$  tal que  $\mathbb{P}(U_i \leq u) = u, \forall u \in [0, 1]$ , ou seja,  $U_i \sim U(0, 1), i = 1, \dots, d$ .

**Teorema 2.6** (Teorema de Sklar). Seja  $F$  uma função de distribuição d-variada com marginais unidimensionais  $F_1, \dots, F_d$ . Então existe uma cópula d-dimensional  $C$  tal que para todo  $\mathbf{t} = (t_1, \dots, t_d) \in \bar{\mathbb{R}}^d$

$$F(\mathbf{t}) = C(F_1(t_1), \dots, F_d(t_d)).$$

Se  $F_1, \dots, F_d$  são contínuas, então  $C$  é única; caso contrário,  $C$  é unicamente determinada na  $ImF_1 \times \dots \times ImF_d$ . Reciprocamente, se  $C$  é uma cópula d-dimensional e  $F_1, \dots, F_d$  são funções de distribuição, então a função  $F$  é uma função de distribuição d-dimensional com marginais  $F_1, \dots, F_d$ .

A partir do Teorema 2.6, uma cópula pode ser vista como um elo de ligação entre as funções de distribuição marginais e a função de distribuição conjunta. Em um problema de estimação em que a dimensão é maior ou igual a dois ( $d \geq 2$ ), encontrar diretamente a função de distribuição conjunta adequada não é, em muitos casos, uma tarefa simples. Neste sentido, o Teorema de Sklar é muito útil no sentido de diminuir a complexidade do problema, pois permite

que se busque conhecimento primeiro das distribuições marginais e posteriormente da estrutura de dependência para se chegar à função de distribuição conjunta.

Algumas cópulas merecem destaque. Seja  $\mathbf{u} = (u_1, \dots, u_d) \in [0, 1]^d$ . Os limitantes inferior e superior de Fréchet-Hoeffding são definidos, respectivamente, como  $W^d(\mathbf{u}) = \max\{\sum_{i=1}^d u_i - d + 1, 0\}$  e  $M^d(\mathbf{u}) = \min\{u_1, \dots, u_d\}$ . Além disso, para toda cópula  $d$ -dimensional  $C$  a desigualdade  $W^d(\mathbf{u}) \leq C(\mathbf{u}) \leq M^d(\mathbf{u})$  é válida.  $M^d$  é uma cópula para todo  $d$ , no entanto  $W^d$  só é uma cópula se  $d = 2$  pois  $V_{W^d}([\mathbf{1}/2, \mathbf{1}]) = 1 - (n/2) < 0$ ,  $\forall d \geq 3$  ( $\mathbf{1}/2 = (1/2, \dots, 1/2)$  e  $\mathbf{1} = (1, \dots, 1)$ ), ou seja,  $W^d$  não satisfaz a Definição 2.2 e, portanto, não é  $d$ -crescente). Outra cópula que merece destaque é a cópula produto, definida como  $\Pi^d(\mathbf{u}) = \prod_{i=1}^d u_i$ , ou seja, um conjunto de variáveis aleatórias  $\mathbf{T} = (T_1, \dots, T_d)$  é independente se, e somente se, a cópula associada a  $\mathbf{T}$  é a cópula produto (observe que  $u_i = F_i(t_i)$ ,  $i = 1, \dots, d$ ). A Figura 2 fornece um esboço das cópulas  $W^d(\mathbf{u})$ ,  $M^d(\mathbf{u})$  e  $\Pi^d(\mathbf{u})$  para  $d = 2$ .

### 2.1.1 Exemplos

**Exemplo 2.7** (Cópulas de Sobrevivência). Seja  $\mathbf{T} = (T_1, T_2)$  um vetor aleatório que representa tempos até a ocorrência de determinado evento,  $F$  e  $S$ , a função de distribuição e a função de sobrevivência conjunta,  $F_i$  e  $S_i$ ,  $i = 1, 2$ , a função de distribuição e a função de sobrevivência marginais, respectivamente. Note que

$$\begin{aligned} S(t_1, t_2) &= \mathbb{P}(T_1 > t_1, T_2 > t_2) \\ &= F(\infty, \infty) - F(t_1, \infty) - F(\infty, t_2) + F(t_1, t_2) \\ &= 1 - F_1(t_1) - F_2(t_2) + C(F_1(t_1), F_2(t_2)) \\ &= S_1(t_1) + S_2(t_2) - 1 + C(1 - S_1(t_1), 1 - S_2(t_2)) \\ &= \bar{u}_1 + \bar{u}_2 - 1 + C(1 - \bar{u}_1, 1 - \bar{u}_2), \end{aligned}$$

em que  $\bar{u}_1 = S_1(t_1)$  e  $\bar{u}_2 = S_2(t_2)$ . A função  $\bar{C} : [0, 1]^2 \rightarrow [0, 1]$  definida como  $\bar{C}(\bar{u}_1, \bar{u}_2) = \bar{u}_1 + \bar{u}_2 - 1 + C(1 - \bar{u}_1, 1 - \bar{u}_2)$  é aplanada (Definição 2.3) pois

$$\begin{aligned} \bar{C}(0, \bar{u}_2) &= 0 + \bar{u}_2 - 1 + C(1 - 0, 1 - \bar{u}_2) \\ &= \bar{u}_2 - 1 + C(1, 1 - \bar{u}_2) \\ &= \bar{u}_2 - 1 + 1 - \bar{u}_2 \\ &= 0, \end{aligned}$$

e, analogamente,  $\bar{C}(\bar{u}_1, 0) = 0$ .  $\bar{C}$  também é bicrescente (Definição 2.2) pois para todo  $B = [\bar{u}_{11}, \bar{u}_{12}] \times [\bar{u}_{21}, \bar{u}_{22}] \in [0, 1]^2$ , com  $\bar{u}_{11} \leq \bar{u}_{12}$  e  $\bar{u}_{21} \leq \bar{u}_{22}$ ,



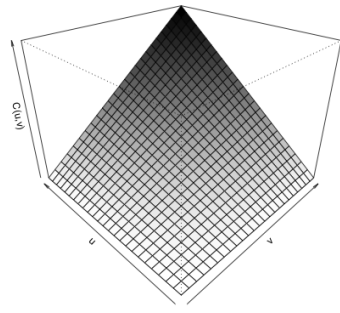
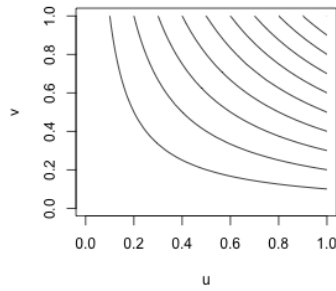
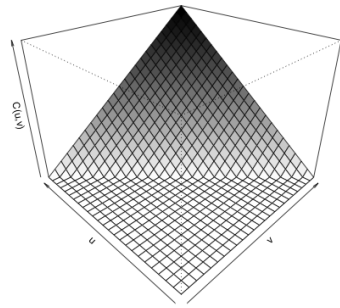
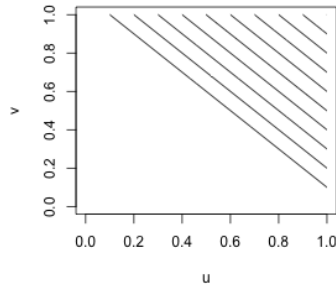
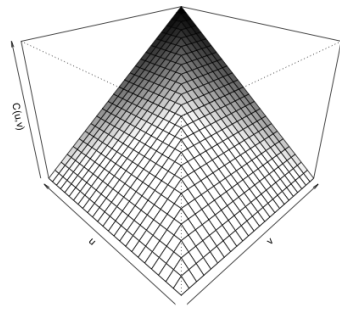
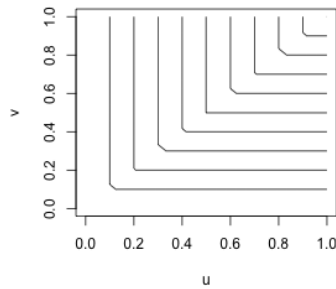
(a)  $\Pi(u, v)$ .(b)  $\Pi(u, v)$  - Curva de nível.(c)  $W(u, v)$ .(d)  $W(u, v)$  - Curva de nível.(e)  $M(u, v)$ .(f)  $M(u, v)$  - Curva de nível.

Figura 2 – Esboço das cópulas produto ( $\Pi$ ), limite inferior ( $W$ ) e superior ( $M$ ) de Fréchet-Hoeffding para  $d=2$ .

$$\begin{aligned}
 V_{\bar{C}}(\mathbf{B}) &= \bar{C}(\bar{u}_{12}, \bar{u}_{22}) - \bar{C}(\bar{u}_{12}, \bar{u}_{21}) - \bar{C}(\bar{u}_{11}, \bar{u}_{22}) + \bar{C}(\bar{u}_{11}, \bar{u}_{21}) \\
 &= [\bar{u}_{12} + \bar{u}_{22} - 1 + C(1 - \bar{u}_{12}, 1 - \bar{u}_{22})] - [\bar{u}_{12} + \bar{u}_{21} - 1 + C(1 - \bar{u}_{12}, 1 - \bar{u}_{21})] - \\
 &\quad [\bar{u}_{11} + \bar{u}_{22} - 1 + C(1 - \bar{u}_{11}, 1 - \bar{u}_{22})] + [\bar{u}_{11} + \bar{u}_{21} - 1 + C(1 - \bar{u}_{11}, 1 - \bar{u}_{21})] \\
 &= C(1 - \bar{u}_{12}, 1 - \bar{u}_{22}) - C(1 - \bar{u}_{12}, 1 - \bar{u}_{21}) - C(1 - \bar{u}_{11}, 1 - \bar{u}_{22}) + C(1 - \bar{u}_{11}, 1 - \bar{u}_{21}) \\
 &= V_C(\bar{\mathbf{B}}),
 \end{aligned}$$

em que  $\bar{B} = [1 - \bar{u}_{12}, 1 - \bar{u}_{11}] \times [1 - \bar{u}_{22}, 1 - \bar{u}_{21}]$  e, dado que  $C$  é cópula,  $V_C(\bar{B}) \geq 0$ , o que implica em  $V_{\bar{C}}(B) \geq 0$ . Por fim, as marginais de  $\bar{C}$  são dadas por

$$\begin{aligned}\bar{C}(1, \bar{u}_2) &= 1 + \bar{u}_2 - 1 + C(0, 1 - \bar{u}_2) \\ &= \bar{u}_2,\end{aligned}$$

já que, pelo fato de  $C$  ser cópula,  $C(0, 1 - \bar{u}_2) = 0$ . Analogamente,  $\bar{C}(\bar{u}_1, 1) = \bar{u}_1$ . Desta forma,  $\bar{C}$  também é uma cópula (Definição 2.4) e é chamada de cópula de sobrevivência de  $\mathbf{T}$ .

**Exemplo 2.8** (Probabilidade Condicional). Seja  $C : [0, 1]^2 \rightarrow [0, 1]$  uma cópula. Então a função de distribuição condicional  $U_2$  dado  $U_1 = u_1$ , denotada por  $c_{u_1}$ , é dada por  $c_{u_1}(u_2) = \mathbb{P}(U_2 \leq u_2 | U_1 = u_1) = \lim_{\Delta u_1 \rightarrow 0} \{C(u_1 + \Delta u_1, u_2) - C(u_1, u_2)\} / \Delta u_1 = \partial C(u_1, u_2) / \partial u_1$ . Vale ressaltar que esta quantidade pode ser usada para simular pares do tipo  $(u_1, u_2)$  (NELSEN, 2006). A densidade da cópula  $C$  é obtida a partir de  $\partial c_{u_1}(u_2) / \partial u_2 = \partial C(u_1, u_2) / \partial u_1 \partial u_2$  e pode ser utilizada, por exemplo, na estimação de máxima verossimilhança ou no processo de estimação bayesiano.

**Exemplo 2.9** (Cópulas Arquimedianas). Seja  $\varphi : [0, 1] \rightarrow [0, +\infty]$  uma função contínua, convexa e estritamente decrescente tal que  $\varphi(1) = 0$ . Considere a pseudo-inversa de  $\varphi$ , denotada por  $\varphi^{[-1]}$ , definida por  $\varphi^{[-1]}(t) = \varphi^{-1}(t) \mathbb{I}_{[0, \varphi(0)]}(t)$ , em que  $\varphi^{-1}$  é a função inversa de  $\varphi$  e  $\mathbb{I}$  é chamada função indicadora ( $\mathbb{I}_A(t) = 1$ , se  $t \in A$  e 0 caso contrário). A função  $C : [0, 1]^2 \rightarrow [0, 1]$  definida como  $C(u_1, u_2) = \varphi^{[-1]}(\varphi(u_1) + \varphi(u_2))$  é chamada cópula arquimediana e  $\varphi$  é chamado gerador da cópula arquimediana  $C$ . A extensão para o caso multivariado pode ser vista em (NELSEN, 2006).

### 2.1.2 Um Método para Obtenção de Cópulas

**Definição 2.10.** Seja  $F : \bar{\mathbb{R}} \rightarrow [0, 1]$  uma função de distribuição. A inversa generalizada de  $F$ , denotada por  $F^{(-1)}$ , é qualquer função de  $[0, 1]$  em  $\bar{\mathbb{R}}$  tal que  $F(F^{(-1)}(t)) = t, \forall t \in \text{Im}F = [0, 1]$  e  $F^{(-1)}(t) = \inf\{x | F(x) \geq t\} = \sup\{x | F(x) \leq t\}, \forall t \notin \text{Im}F = [0, 1]$ .

**Corolário 2.11** (Cópulas via Função de Distribuição). Seja  $F$  uma função de distribuição d-variada com funções de distribuição marginais  $F_1, \dots, F_d$ , cujas inversas generalizadas são, respectivamente,  $F_1^{(-1)}, \dots, F_d^{(-1)}$ . Então,  $\forall \mathbf{u} = (u_1, \dots, u_d) \in [0, 1]^d, C : [0, 1]^d \rightarrow [0, 1]$  dada por

$$C(\mathbf{u}) = F(F_1^{(-1)}(u_1), \dots, F_d^{(-1)}(u_d))$$

é uma cópula.

A partir do Corolário 2.11, uma cópula  $C : [0, 1]^d \rightarrow [0, 1]$  pode ser construída a partir de uma função de distribuição d-variada  $F : \bar{\mathbb{R}}^d \rightarrow [0, 1]$ . Considere  $F_\theta, \theta \in [0, 1]$ , a função correspondente a distribuição exponencial bivariada de Gumbel (GUMBEL, 1960) dada por

$$F_{\theta}(t_1, t_2) = \begin{cases} 1 - e^{-t_1} - e^{-t_2} + e^{-(t_1+t_2+\theta t_1 t_2)} & (t_1, t_2) \in [0, \infty] \times [0, \infty] \\ 0 & c.c. \end{cases}$$

As funções de distribuição marginais de  $F_{\theta}$  são dadas por  $F_1(t_1) = F_{\theta}(t_1, +\infty) = 1 - e^{-t_1}$ , ou seja,  $T_1 \sim Exp(1)$  e  $F_2(t_2) = F_{\theta}(+\infty, t_2) = 1 - e^{-t_2}$ , ou seja,  $T_2 \sim Exp(1)$ . A partir da Definição 2.10,  $F_i^{(-1)}(u_i) = -\log(1 - u_i)$ ,  $u_i \in [0, 1]$ ,  $i = 1, 2$ . Tendo em vista o Corolário 2.11, basta aplicar  $F_1^{(-1)}(u_1)$  e  $F_2^{(-1)}(u_2)$  em  $F_{\theta}$  para obter a cópula, isto é,  $C(u_1, u_2) = F_{\theta}(F_1^{(-1)}(u_1), F_2^{(-1)}(u_2)) = u_1 + u_2 - 1 + (1 - u_1)(1 - u_2)e^{-\theta \log(1 - u_1) \log(1 - u_2)}$ . A Figura 3 fornece um esboço desta cópula para  $\theta = 0,5$ .

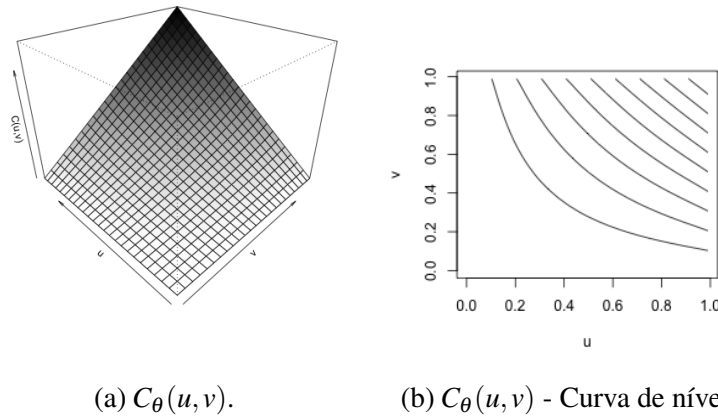


Figura 3 – Esboço da cópula construída a partir da distribuição exponencial bivariada de Gumbel.

## 2.2 Medidas de Dependência

A dependência entre duas variáveis aleatórias está contida na função de distribuição conjunta destas variáveis, que por sua vez está intimamente ligada ao contexto de cópulas pelo Teorema 2.6. Desta forma, é natural procurar meios para quantificar esta associação através de cópulas, sendo esta quantificação baseada no intervalo  $[-1, +1]$ , em que  $-1$  indica dependência perfeita negativa e  $+1$  indica dependência perfeita positiva. Diversas medidas foram propostas na literatura e cada medida se propõe a captar um tipo específico de associação. Nesta seção, duas medidas são apresentadas e a relação destas medidas com cópulas são discutidas.

### 2.2.1 Concordância

Um conceito fundamental para construção de algumas medidas de dependência, como o tau de Kendall e o rho de Spearman, é o conceito de concordância. Informalmente, um par de variáveis aleatórias é concordante se “grandes valores” de uma variável estão associados a

“grandes valores” da outra variável e, de maneira similar, “pequenos valores” de uma variável estão associados a “pequenos valores” da outra variável. Formalmente, considere  $(x_i, y_i)$  e  $(x_j, y_j)$  duas observações do vetor de variáveis aleatórias contínuas  $(X, Y)$ . Diz-se que  $(x_i, y_i)$  e  $(x_j, y_j)$  são concordantes se  $(x_i - x_j)(y_i - y_j) > 0$ , isto é,  $x_i < x_j$  e  $y_i < y_j$  ou  $x_i > x_j$  e  $y_i > y_j$ . Por outro lado, diz-se que  $(x_i, y_i)$  e  $(x_j, y_j)$  são discordantes se  $(x_i - x_j)(y_i - y_j) < 0$ , isto é,  $x_i < x_j$  e  $y_i > y_j$  ou  $x_i > x_j$  e  $y_i < y_j$ . Mais detalhes podem ser encontrados em (KRUSKAL, 1958).

### 2.2.2 Tau de Kendall

A versão amostral do tau de Kendall é definida em termos de concordância. Considere uma amostra aleatória  $\{(x_i, y_i)\}_{i=1}^n$  de tamanho  $n$  do vetor de variáveis aleatórias contínuas  $(X, Y)$ . Seja  $c$  o número de pares concordantes e  $d = n - c$  o número de pares discordantes. A versão amostral do tau de Kendall, denotada por  $t$ , é definida como

$$t = \frac{c - d}{c + d}. \quad (2.12)$$

Note que a Equação (2.12) pode ser vista como a probabilidade de concordância menos a probabilidade de discordância para os pares  $(x_i, y_i)$  e  $(x_j, y_j)$ , escolhidos aleatoriamente de  $\{(x_i, y_i)\}_{i=1}^n$ . A versão populacional do tau de Kendall para um vetor de variáveis aleatórias contínuas  $(X, Y)$  é definida de forma análoga. Suponha  $(X_1, Y_1)$  e  $(X_2, Y_2)$  vetores aleatórios independentes e com a mesma distribuição de  $(X, Y)$ . A versão populacional do tau de Kendall, denotado por  $\tau_{X,Y}$ , é definida como

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

Cópuas podem ser utilizadas no cálculo do tau de Kendall. No entanto, antes de mostrar a relação entre tau de Kendall e cópuas é necessário definir a função de concordância para dois vetores de variáveis aleatórias contínuas  $(X_1, Y_1)$  e  $(X_2, Y_2)$ , cujas funções de distribuição serão denotadas por  $H_1$  e  $H_2$ , em que  $H_1$  não é necessariamente igual a  $H_2$ . Apesar disto,  $X_1$  e  $X_2$  possuem a mesma função de distribuição, denotada por  $F$ , assim como  $Y_1$  e  $Y_2$ , cuja função de distribuição será denotada por  $G$ .

**Teorema 2.13.** Sejam  $(X_1, Y_1)$  e  $(X_2, Y_2)$  vetores de variáveis aleatórias contínuas com funções de distribuição  $H_1$  e  $H_2$  e marginais comuns  $F$ , de  $X_1$  e  $X_2$ , e  $G$ , de  $Y_1$  e  $Y_2$ . Pelo Teorema 2.6,  $H_i(x, y) = C_i(F(x), G(y))$ ,  $i = 1, 2$ . A função de concordância, denotada por  $Q$ , é definida pela diferença entre a probabilidade de concordância e a probabilidade de discordância, isto é,  $Q = \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) > 0) - \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) < 0)$ , pode ser escrita como

$$Q = Q(C_1, C_2) = \iint_{I^2} C_2(u, v) dC_1(u, v) - 1,$$

em que  $u = F(x)$  e  $v = G(y)$ .

*Demonstração.*  $(X_i, Y_i), i = 1, 2$ , são vetores de variáveis aleatórias contínuas, portanto,  $\mathbb{P}((X_1 - X_2)(Y_1 - Y_2) < 0)$  pode ser escrita como  $1 - \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) > 0)$ . Desta forma,  $Q = 2\mathbb{P}((X_1 - X_2)(Y_1 - Y_2) > 0) - 1$ . Além disso, os eventos  $(X_1 - X_2)(Y_1 - Y_2) > 0$  e  $[(X_1 > X_2) \cap (Y_1 > Y_2)] \cup [(X_1 < X_2) \cap (Y_1 < Y_2)]$  são equivalentes e, conseqüentemente,  $\mathbb{P}((X_1 - X_2)(Y_1 - Y_2) > 0) = \mathbb{P}(X_1 > X_2, Y_1 > Y_2) + \mathbb{P}(X_1 < X_2, Y_1 < Y_2)$ , uma vez que  $(X_1 > X_2) \cap (Y_1 > Y_2)$  e  $(X_1 < X_2) \cap (Y_1 < Y_2)$  são eventos disjuntos. Portanto,

$$\begin{aligned} \mathbb{P}(X_1 > X_2, Y_1 > Y_2) &= \mathbb{P}(X_2 < X_1, Y_2 < Y_1) \\ &= \iint_{\mathbb{R}^2} \mathbb{P}(X_2 < x, Y_2 < y) dC_1(F(x), G(y)) \\ &= \iint_{\mathbb{R}^2} C_2(F(x), G(y)) dC_1(F(x), G(y)) \\ &= \iint_{I^2} C_2(u, v) dC_1(u, v). \end{aligned}$$

Por outro lado,

$$\begin{aligned} \mathbb{P}(X_1 < X_2, Y_1 < Y_2) &= \iint_{\mathbb{R}^2} \mathbb{P}(X_2 > x, Y_2 > y) dC_1(F(x), G(y)) \\ &= \iint_{\mathbb{R}^2} (1 - F(x) - G(y) + C_2(F(x), G(y))) dC_1(F(x), G(y)) \\ &= \iint_{I^2} (1 - u - v + C_2(u, v)) dC_1(u, v) \\ &= \iint_{I^2} C_2(u, v) dC_1(u, v), \end{aligned}$$

pois  $U, V \sim U(0, 1)$  ( $\mathbb{E}(U) = 1/2 = \mathbb{E}(V)$ ). Ou seja,  $\mathbb{P}(X_1 > X_2, Y_1 > Y_2) = \mathbb{P}(X_1 < X_2, Y_1 < Y_2)$ . E, por fim,

$$\begin{aligned} Q &= \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) > 0) - \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) < 0) \\ &= 2\mathbb{P}((X_1 - X_2)(Y_1 - Y_2) > 0) - 1 \\ &= 2(\mathbb{P}(X_1 > X_2, Y_1 > Y_2) + \mathbb{P}(X_1 < X_2, Y_1 < Y_2)) - 1 \\ &= 4\mathbb{P}(X_1 > X_2, Y_1 > Y_2) - 1 \\ &= 4 \iint_{I^2} C_2(u, v) dC_1(u, v) - 1. \end{aligned}$$

□

A função de concordância possui propriedades interessantes. O próximo corolário descreve algumas das principais propriedades.

**Corolário 2.14.** Sejam  $C_1, C_2$  e  $Q$  como no Teorema 2.13. Então,

- $Q$  é simétrica nos seus argumentos, isto é,  $Q(C_1, C_2) = Q(C_2, C_1)$ .
- $Q$  é não decrescente em cada argumento, isto é, se  $C_1 \prec C'_1$  e  $C_2 \prec C'_2$  para todo  $(u, v) \in [0, 1]^2$ , então  $Q(C_1, C_2) \leq Q(C'_1, C'_2)$ .
- Cópias de sobrevivência podem substituir cópias em  $Q$ , isto é,  $Q(C_1, C_2) = Q(\bar{C}_1, \bar{C}_2)$ .

A versão populacional do tau de Kendall pode ser vista como um caso particular do Teorema 2.13, já que os vetores de variáveis aleatórias contínuas  $(X_i, Y_i)$ ,  $i = 1, 2$ , possuem a mesma função de distribuição, isto é,  $H_1 = H_2$ . Denotando, respectivamente, por  $H$  e  $C$  esta função de distribuição e cópula associada tal que  $H(x, y) = C(F(x), G(y))$ ,  $\forall (x, y) \in \bar{\mathbb{R}}^2$ , a versão populacional do tau de Kendall pode se escrita como

$$\tau_{X,Y} = Q(C, C) = 4 \iint_{\mathcal{I}^2} C(u, v) dC(u, v) - 1, \quad (2.15)$$

em que  $u = F(x)$  e  $v = G(y)$ . Note que a Equação (2.15) pode ser vista como quatro vezes esperança da função  $C(U, V)$  menos um, ou seja,  $\tau_{X,Y} = 4\mathbb{E}(C(U, V)) - 1$ , em que  $U, V \sim U(0, 1)$ . Se  $X$  e  $Y$  são independentes, então  $C(u, v) = uv$ , o que implica em  $\mathbb{E}(C(U, V)) = \mathbb{E}(UV) = \mathbb{E}(U)\mathbb{E}(V) = (1/2)^2 = 1/4$ , ou seja,  $\tau_{X,Y} = 0$ .

### 2.2.3 Rho de Spearman

O rho de Spearman, assim como o tau de Kendall, é definido em termos dos conceitos de concordância e discordância. A versão populacional do rho de Spearman para um vetor de variáveis aleatórias contínuas  $(X, Y)$  é definida a partir de três vetores aleatórios  $(X_i, Y_i)$ ,  $i = 1, 2, 3$ , que são independentes e possuem a mesma distribuição de  $(X, Y)$ . A versão populacional do rho de Spearman, denotado por  $\rho_{X,Y}$ , é definido como

$$\rho_{X,Y} = 3(\mathbb{P}((X_1 - X_2)(Y_1 - Y_3) > 0) - \mathbb{P}((X_1 - X_2)(Y_1 - Y_3) < 0)). \quad (2.16)$$

Note que a Equação (2.16) é a função de concordância, definida no Teorema 2.13, para os vetores  $(X_1, Y_1)$  e  $(X_2, Y_3)$ , multiplicada por três. Vale ressaltar que o vetor  $(X_3, Y_2)$  poderia ser utilizado na definição ao invés do vetor  $(X_2, Y_3)$ . Suponha que  $(X, Y)$  tenha função de distribuição conjunta  $H$  com marginais  $F$  (de  $X$ ),  $G$  (de  $Y$ ) e cópula  $C$  associada. Desta forma, a função de distribuição conjunta do vetor  $(X_3, Y_2)$  é  $F(x_3)G(y_2)$ ,  $\forall (x_3, y_2) \in \bar{\mathbb{R}}^2$ , uma vez que  $X_3$  e  $Y_2$  são independentes. Em termos de cópias, a versão populacional do rho de Spearman pode ser escrita como

$$\rho_{X,Y} = 3Q(C, \Pi) = 12 \iint_{\mathcal{I}^2} uv dC(u, v) - 3. \quad (2.17)$$

Não é difícil mostrar que  $Q(W, \Pi) = -1/3$  e  $Q(M, \Pi) = +1/3$ , em que  $W(u, v) = \max\{u + v - 1, 0\}$  e  $M(u, v) = \min\{u, v\}$  são, respectivamente, o limitante inferior e superior de Fréchet-Hoeffding. Além disto, para todo par  $(u, v)$  em  $[0, 1]^2$  e toda cópula  $C$ , a desigualdade  $W(u, v) \leq C(u, v) \leq M(u, v)$  é válida e, aplicando o segundo item do Corolário 2.14,  $Q(W, \Pi) \leq Q(C, \Pi) \leq Q(M, \Pi)$ , ou seja,  $Q(C, \Pi) \in [-1/3, +1/3]$ . Por esta razão, multiplica-se  $Q(C, \Pi)$  pela constante três, que é chamada constante normalizadora. Por fim,

$$\rho_{X,Y} = 12\mathbb{E}(UV) - 3 = \frac{\mathbb{E}(UV) - 1/4}{1/12} = \frac{\mathbb{E}(UV) - \mathbb{E}(U)\mathbb{E}(V)}{\sqrt{\text{Var}(U)\text{Var}(V)}}. \quad (2.18)$$

Isto é, a Equação (2.18) mostra que o rho de Spearman pode ser visto como o coeficiente de correlação linear de Pearson para as variáveis  $U = F(X)$  e  $V = G(Y)$ , com  $U, V \sim U(0, 1)$ . Se  $X$  e  $Y$  são independentes, então  $\mathbb{E}(UV) = \mathbb{E}(U)\mathbb{E}(V)$ , ou seja,  $\rho_{X,Y} = 0$ .

## 2.3 Análise de Sobrevivência

Em algumas situações, o interesse não está diretamente na função de distribuição conjunta das variáveis aleatórias. Exemplos de situações como esta surgem naturalmente quando se pensa no contexto de análise de sobrevivência, em que o tempo até a ocorrência de um evento de interesse é, em geral, a variável resposta. Este tempo, denominado muitas vezes por tempo de falha, pode ser o tempo até a recidiva ou cura de uma determinada doença, o tempo até a falha de um componente eletrônico, entre outros.

### 2.3.1 Censura

Uma característica particular em análise de sobrevivência é a existência de observações incompletas ou parciais. Estas observações, que geralmente são chamadas de observações censuradas, existem por uma série de fatores como a perda do acompanhamento de um paciente ao longo de um tratamento, o evento de interesse pode não ocorrer até o fim de um experimento, entre outros fatores. De acordo com (COLOSIMO; GIOLO, 2006), embora não se tenha a informação completa, é de suma importância que as observações censuradas façam parte das análises estatísticas pois, além de trazer informação, mesmo que parcial, sobre o tempo até a falha, a retirada dessas observações podem induzir a conclusões viesadas.

Existem três principais tipos de censura:

- **Tipo I:** o estudo terminará após um período de tempo estabelecido previamente e, desta maneira, os indivíduos em que o evento de interesse não ocorreu são classificados como observações censuradas;
- **Tipo II:** o estudo terminará após um número de ocorrências do evento de interesse estabelecido previamente;

- **Aleatória:** este tipo de censura é caracterizada por indivíduos que deixaram de fazer parte do estudo mesmo sem a ocorrência do evento de interesse.

Quanto aos mecanismos de censura:

- **À direita:** o tempo entre o início do estudo e a ocorrência do evento é maior que o tempo entre o início do estudo e o tempo registrado, isto é, o tempo de ocorrência do evento está à direita do tempo registrado;
- **À esquerda:** o tempo entre o início do estudo e a ocorrência do evento é menor que o tempo entre o início do estudo e o tempo registrado, isto é, o tempo de ocorrência do evento está à esquerda do tempo registrado;
- **Intervalar:** a ocorrência do evento de interesse acontece em um intervalo de tempo conhecido, porém o tempo exato em que o evento ocorreu é desconhecido.

Além de saber se a observação é ou não censurada, é importante conhecer o tipo e o mecanismo de censura pois estes dois aspectos influenciam, por exemplo, na forma como a função de verossimilhança é construída e, conseqüentemente, afeta diretamente o resultado das análises.

### 2.3.2 Verossimilhança

Seja  $(T_1, T_2)$  um vetor bivariado de tempos de falha com função de sobrevivência conjunta contínua  $S$  (e marginais  $S_1$  e  $S_2$ ) e  $(T_1^*, T_2^*)$  um vetor bivariado de tempos de censura. A princípio, vamos assumir que  $(T_1, T_2)$  é independente de  $(T_1^*, T_2^*)$ . Na prática, temos disponíveis amostras  $(z_1, z_2)$  do vetor  $(Z_1, Z_2)$ , em que, para o caso de censura à direita,  $Z_k = \min\{T_k, T_k^*\}$ ,  $k = 1, 2$ , que podem ser classificadas em quatro grandes grupos:

- $A_1: (z_1, z_2) = (t_1, t_2)$ , isto é,  $z_1$  e  $z_2$  são tempos de sobrevivência observados;
- $A_2: (z_1, z_2) = (t_1, t_2^*)$ , isto é,  $z_1$  é o tempo de sobrevivência observado e  $z_2$  é o tempo de censura;
- $A_3: (z_1, z_2) = (t_1^*, t_2)$ , isto é,  $z_1$  é o tempo de censura e  $z_2$  é o tempo de sobrevivência observado;
- $A_4: (z_1, z_2) = (t_1^*, t_2^*)$ , isto é,  $z_1$  e  $z_2$  são tempos de censura.

Os indivíduos em cada grupo contribuem na verossimilhança de maneira diferente, a saber:

- Indivíduos em  $A_1: \frac{d^2 S(z_1, z_2)}{dz_1 dz_2}$ ;



- Indivíduos em  $A_2$ :  $-\frac{dS(z_1, z_2)}{dz_1}$ ;
- Indivíduos em  $A_3$ :  $-\frac{dS(z_1, z_2)}{dz_2}$ ;
- Indivíduos em  $A_4$ :  $S(z_1, z_2)$ .

A partir de um conjunto de dados  $\mathcal{D} = (z_{11}, z_{12}), \dots, (z_{n1}, z_{n2})$  e considerando  $\boldsymbol{\theta}$  como o vetor de parâmetros de interesse, a função de verossimilhança pode ser escrita como (LAWLESS, 2011):

$$L(\boldsymbol{\theta}|\mathcal{D}) = \prod_{i \in A_1} \frac{d^2 S(z_{i1}, z_{i2})}{dz_{i1} dz_{i2}} \times \prod_{i \in A_2} \left( -\frac{dS(z_{i1}, z_{i2})}{dz_{i1}} \right) \times \prod_{i \in A_3} \left( -\frac{dS(z_{i1}, z_{i2})}{dz_{i2}} \right) \times \prod_{i \in A_4} S(z_{i1}, z_{i2}). \quad (2.19)$$

Pelo Exemplo 2.7,  $S(z_{i1}, z_{i2}) = \bar{C}(S_1(z_{i1}), S_2(z_{i2}))$ , então a Equação (2.19) pode ser reescrita como:

$$L(\boldsymbol{\theta}|\mathcal{D}) = \prod_{i \in A_1} \frac{d^2 \bar{C}(S_1(z_{i1}), S_2(z_{i2}))}{dz_{i1} dz_{i2}} \times \prod_{i \in A_2} \left( -\frac{d\bar{C}(S_1(z_{i1}), S_2(z_{i2}))}{dz_{i1}} \right) \times \prod_{i \in A_3} \left( -\frac{d\bar{C}(S_1(z_{i1}), S_2(z_{i2}))}{dz_{i2}} \right) \prod_{i \in A_4} \bar{C}(S_1(z_{i1}), S_2(z_{i2})). \quad (2.20)$$

## 2.4 Seleção de Modelos

O Teorema de Sklar garante a existência de uma função cópula que conecta as distribuições marginais à função de distribuição conjunta, o que é útil no sentido de reduzir a complexidade do problema. No entanto, a forma para se construir diretamente a cópula não é explicitada. Desta forma, na prática é usual que se ajuste vários modelos de cópulas que acomodam o tipo de dependência existente nos dados e o melhor, de acordo com algum critério de seleção de modelos, é escolhido. Nesta seção são discutidos alguns destes critérios.

### 2.4.1 EAIC, EBIC e DIC

O Expected Akaike Information Criterion (EAIC) (BROOKS *et al.*, 2002), Expected Bayesian Information Criterion (EBIC) (CARLIN; LOUIS, 2001) e Deviance Information Criterion (DIC) (SPIEGELHALTER *et al.*, 2002) são critérios de seleção de modelos no contexto bayesiano que são construídos a partir do conceito de Deviance. Para o modelo  $M$ , a Deviance é definida como

$$D_M(\boldsymbol{\theta}) = -2 \log L_M(\boldsymbol{\theta}), \quad (2.21)$$

em que  $L_M(\boldsymbol{\theta})$  é a função de verossimilhança do modelo M. Desta forma,

$$\begin{aligned} EAIC &= \mathbb{E}(D_M(\boldsymbol{\theta})) + 2p_M \\ EBIC &= \mathbb{E}(D_M(\boldsymbol{\theta})) + p_M \log(n) \\ DIC &= 2\mathbb{E}(D_M(\boldsymbol{\theta})) - D_M(\mathbb{E}(\boldsymbol{\theta})), \end{aligned} \quad (2.22)$$

em que  $p_M$  é o número de parâmetros para o modelo M e  $n$  é o tamanho da amostra. Se  $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K$  é uma amostra da distribuição a posteriori de  $\boldsymbol{\theta}$  (obtida, por exemplo, via MCMC), então as noções em (2.22) podem ser calculadas da seguinte maneira

$$\begin{aligned} \widehat{EAIC} &= K^{-1} \sum_{k=1}^K D_M(\boldsymbol{\theta}_k) + 2p_M \\ \widehat{EBIC} &= K^{-1} \sum_{k=1}^K D_M(\boldsymbol{\theta}_k) + p_M \log(n) \\ \widehat{DIC} &= 2K^{-1} \sum_{k=1}^K D_M(\boldsymbol{\theta}_k) - D_M\left(K^{-1} \sum_{k=1}^K \boldsymbol{\theta}_k\right). \end{aligned} \quad (2.23)$$

De todos os modelos considerados, escolhe-se, em geral, aquele que tem o menor valor de EAIC, EBIC e DIC e diz-se que este modelo se ajusta melhor ao conjunto de dados.

### 2.4.2 FBST

O *e-value* é a contrapartida bayesiana ao *p-value* e é baseado no *Full Bayesian Significance Test* (FBST) (PEREIRA *et al.*, 2008). A principal ideia do FBST é avaliar o volume de um subespaço do espaço paramétrico, que possui apenas pontos com menor densidade a posteriori quando comparado à máxima densidade a posteriori no espaço da hipótese. Considerando que nossa hipótese é H, uma definição formal do FBST é dada pela Definição 2.24.

**Definição 2.24.** Seja  $\pi(\boldsymbol{\theta}|\mathbf{t})$  a densidade a posteriori de  $\boldsymbol{\theta}$  dado a amostra  $\mathbf{x}$ ,  $\Theta$  o espaço paramétrico,  $\Theta_H$  o espaço paramétrico sob a restrição da hipótese H e  $T(\mathbf{t}) = \{\boldsymbol{\theta} \in \Theta : \pi(\boldsymbol{\theta} | \mathbf{t}) \geq \sup_{\boldsymbol{\theta} \in \Theta_H} \pi(\boldsymbol{\theta} | \mathbf{t})\}$ . A medida de evidência de suporte para a hipótese  $\boldsymbol{\theta} \in \Theta_H$  é definida como  $E_V(\Theta_H, \mathbf{x}) = 1 - \mathbb{P}(\boldsymbol{\theta} \in T(\mathbf{t}) | \mathbf{t})$ .

Note que  $T(\mathbf{t})$  é um subespaço de  $\Theta$ , chamado conjunto tangente.  $T(\mathbf{t})$  é tangente à hipótese H. Se  $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K$  é uma amostra da distribuição a posteriori de  $\boldsymbol{\theta}$  (obtida, por exemplo, via MCMC), então o *e-value* é dado por

$$1 - \frac{1}{K} \sum_{i=1}^K \mathbb{I} \left( \pi(\boldsymbol{\theta}_i | \mathbf{t}) \geq \sup_{\boldsymbol{\theta} \in \Theta_H} \pi(\boldsymbol{\theta} | \mathbf{t}) \right), \quad (2.25)$$

em que  $\mathbb{I}(A)$  é a função indicadora do conjunto  $A$ .

No caso de modelos encaixados, isto é, em que uma classe modelos é derivada de um modelo generalista, a metodologia do FBST pode ser útil.

No caso univariado, por exemplo, suponha que em um problema de análise de sobrevivência estamos interessados em comparar o modelo Weibull e Exponencial. Assim, se  $T \sim Weibull(\lambda, k)$ , então

$$f(t) = \frac{k}{\lambda} \left(\frac{t}{\lambda}\right)^{k-1} e^{-(t/\lambda)^k} \mathbb{I}_{R>0}(t). \quad (2.26)$$

Por outro lado, quando  $k = 1$  a Equação 2.26 pode ser escrita como

$$f(t) = \frac{1}{\lambda} e^{-(t/\lambda)} \mathbb{I}_{R>0}(t). \quad (2.27)$$

Neste caso,  $T \sim Exp(\lambda)$ . A seleção de modelos, considerando o FBST, pode ser feita a partir do teste da hipótese  $H: k = 1$ , sendo uma alternativa interessante aos critérios usuais de seleção como EAIC, EBIC, DIC, entre outros, por se tratar de um teste de hipótese.

## 2.5 FBST para Cópula Power Variance Function (PVF)

A distribuição Power Variance Function (PVF) foi derivada como uma extensão da distribuição estável positiva por (HOUGAARD, 1986). Seguindo a parametrização dada por (MALLICK; RAVISHANKER, 2006), uma variável aleatória  $W$  possui distribuição PVF( $\alpha, \delta, \theta$ ) se sua função densidade de probabilidade pode ser escrita como

$$f(w) = -\frac{1}{\pi w} \exp(-\theta w + \delta \theta^\alpha / \alpha) \sum_{k=1}^{\infty} \frac{\Gamma(k\alpha + 1)}{\Gamma(k + 1)} \left(\frac{-w^{-\alpha} \delta}{\alpha}\right)^k \sin(\alpha k \pi),$$

com  $\alpha \in (0, 1)$ ,  $\delta > 0$  e  $\theta > 0$ . A transformada de Laplace  $L_W(s) = \mathbb{E}[\exp(-sW)]$  de  $W$  é dada por

$$L_W(s) = \exp\left(-\frac{\delta}{\alpha} [(\theta + s)^\alpha - \theta^\alpha]\right).$$

(OAKES, 1989) mostra que qualquer modelo de fragilidade para dados de sobrevivência multivariados possui uma representação via cópula Arquimediana, introduzida no Exemplo 2.9. No caso  $d$ -variado, sejam  $T_1, \dots, T_d$  variáveis aleatórias contínuas não negativas com sobrevivências marginais  $S_1, \dots, S_d$ . Se  $T_1, \dots, T_d$  são condicionalmente independentes dado uma variável aleatória não negativa  $W$  em um modelo de fragilidade, isto é,  $S(t_1, \dots, t_d | w) = \prod_{i=1}^d S_i(t_i | w) = \prod_{i=1}^d \{B_i(t_i)\}^w$ , em que  $B_i$  são funções de sobrevivência basais, a função de sobrevivência conjunta possui a seguinte representação via cópulas Arquimedianas:

$$S(t_1, \dots, t_d) = \bar{C}_\phi(S_1(t_1), \dots, S_d(t_d)) = \phi^{-1} \left[ \sum_{j=1}^d \phi(S_j(t_j)) \right],$$

com gerador da cópula  $\bar{C}_\phi$  dado por  $\phi(\cdot) = L_W^{-1}(\cdot)$ , a inversa da transformada de Laplace de  $W$ . Isto é visto a partir da seguinte representação da cópula de sobrevivência

$$\bar{C}(u_1, \dots, u_d) = S(S_1^{-1}(u_1), \dots, S_d^{-1}(u_d)),$$

$$\begin{aligned} S(t_1, \dots, t_d) &= \int S(t_1, \dots, t_d | w) dF(w) \\ &= \int \prod_{i=1}^d \{B_i(t_i)\}^w dF(w) \\ &= \int \exp \left( w \sum_{i=1}^d \log(B_i(t_i)) \right) dF(w) \\ &= \mathbb{E} \left[ \exp \left( W \sum_{i=1}^d \log(B_i(t_i)) \right) \right] \\ &= L_W \left( - \sum_{i=1}^d \log(B_i(t_i)) \right), \end{aligned}$$

$S_j(t_j) = L_W(-\log B_j(t_j))$ , para  $j = 1, \dots, d$  e  $S_j^{-1}(u_j) = B_j^{-1}(\exp(-L_W^{-1}(u_j)))$ , para  $u_j \in (0, 1)$  e  $j = 1, \dots, d$ .

Assim como em (ROMEO; MEYER; GALLARDO, 2018), a transformação  $\delta = \eta^{1-\alpha}$  e  $\theta = \eta$ , com  $\alpha \in (0, 1)$  e  $\eta > 0$ , é aplicada para tornar os parâmetros identificáveis. Desta forma, a transformada de Laplace de  $W \sim PVF(\alpha, \eta)$  é dada por

$$\phi^{-1}(s) = \exp \left( -\frac{1}{\alpha} [\eta^{1-\alpha}(\eta + s)^\alpha - \eta] \right).$$

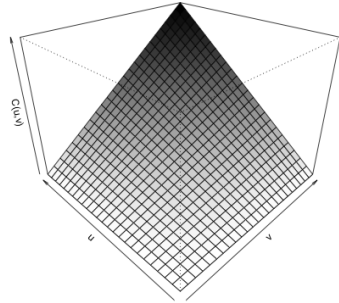
Definindo  $g(s) = \eta^\alpha - \alpha \eta^{\alpha-1} \log(s)$ , a inversa da transformada de Laplace de  $W$ , o gerador da cópula PVF é dado por  $\phi(s) = g(s)^{1/\alpha} - \eta$ . A versão multivariada da cópula PVF, com espaço paramétrico  $\Omega = \{(\alpha, \eta) \in \mathbb{R}^2 : \alpha \in (0, 1), \eta > 0\}$  pode ser escrita como

$$\bar{C}_{\alpha, \eta}(u_1, \dots, u_d) = \exp \left( -\frac{1}{\alpha} \left[ \eta^{1-\alpha} \left( \sum_{i=1}^d g(u_i)^{1/\alpha} - (p-1)\eta \right)^\alpha - \eta \right] \right),$$

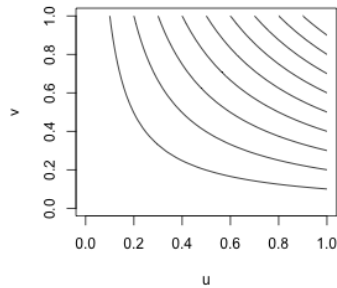
em que  $u_i = S_i(t_i)$ . Se  $d = 2$ , então

$$\bar{C}_{\alpha, \eta}(u_1, u_2) = \exp \left( -\frac{1}{\alpha} \left[ \eta^{1-\alpha} \left( g(u_1)^{\frac{1}{\alpha}} + g(u_2)^{\frac{1}{\alpha}} - \eta \right)^\alpha - \eta \right] \right).$$

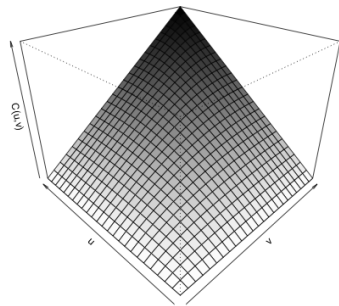
A Figura 4 traz o esboço desta cópula para alguns valores dos parâmetros  $\alpha$  e  $\eta$ .



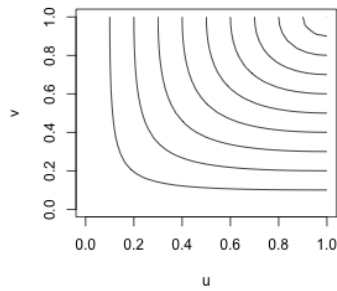
(a) PVF(0,9;10).



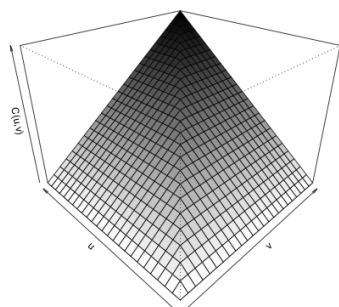
(b) PVF(0,9;10) - Curva de nível.



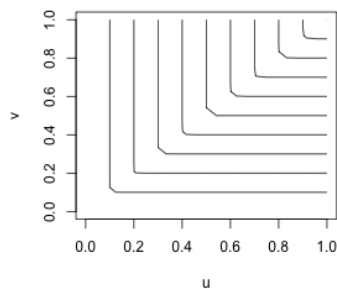
(c) PVF(0,5;0,001).



(d) PVF(0,5;0,001) - Curva de nível.



(e) PVF(0,001;0,01).



(f) PVF(0,001;0,01) - Curva de nível.

Figura 4 – Esboço da cópula PVF para diferentes valores de  $\alpha$  e  $\eta$ .

A vantagem em utilizar a cópula PVF está apoiada no fato de que esse modelo é generalista e tem como casos particulares modelos de cópulas amplamente utilizados na literatura, a saber:

1. Se  $\eta \rightarrow 0$ , a cópula bivariada de Gumbel é obtida

$$\bar{C}_\alpha(u_1, u_2) = \exp\left(-\left[(-\log u_1)^{1/\alpha} + (-\log u_2)^{1/\alpha}\right]^\alpha\right), \text{ com } \alpha \in (0, 1).$$

2. Se  $\alpha = 0,5$ , a cópula bivariada Gaussiana Inversa é obtida

$$\bar{C}_\eta(u_1, u_2) = \exp\left(-2\left[\eta^{1/2}(b(u_1)^2 + b(u_2)^2 - \eta)^{1/2} - \eta\right]\right), \text{ com } \eta > 0,$$

em que  $b(s) = \eta^{1/2} - (1/2)\eta^{-1/2} \log s$ .

3. Se  $\alpha \rightarrow 0$ , a cópula bivariada de Clayton é obtida

$$\bar{C}_\eta(u_1, u_2) = (u_1^{-\eta} + u_2^{-\eta} - 1)^{-1/\eta} \text{ com } \eta > 0.$$

Note que as cópulas de Gumbel e Clayton são obtidas quando se está na borda do espaço paramétrico. Neste trabalho, nosso interesse está em estimar, a partir de um conjunto de dados  $\mathcal{D}$  com censura à direita, os parâmetros da cópula de sobrevivência PVF e seus casos particulares por um processo bayesiano. Desta forma, se  $\boldsymbol{\theta}$  representa o vetor de parâmetros que engloba os parâmetros da cópula e os parâmetros das distribuições marginais, então a posteriori, denotada por  $\pi(\boldsymbol{\theta}|\mathcal{D})$ , pode ser escrita como

$$\pi(\boldsymbol{\theta}|\mathcal{D}) \propto L(\mathcal{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}),$$

em que  $L$  é a função de verossimilhança como na Equação (2.20) e  $\pi(\boldsymbol{\theta})$  é a priori conjunta para o vetor de parâmetros  $\boldsymbol{\theta}$ .

A partir de uma amostra da distribuição a posteriori  $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_k$  obtida, por exemplo, via MCMC, calculamos o *e-value* a partir da Equação (2.25) para testar hipóteses sobre os modelos encaixados, além de utilizar os demais métodos de seleção de modelos no contexto bayesiano descritos na Seção 2.4. Para quantificar a dependência entre as variáveis que representam os tempos de sobrevivência, utilizamos como base as medidas descritas na Seção 2.2, como Tau de Kendall e Rho de Spearman.



## APLICAÇÕES

Neste capítulo são apresentados um estudo de simulação e uma análise de dados de apendicectomia para gêmeos adultos como ilustração e validação da proposta do uso da cópula PVF e seleção de modelos via FBST apresentada no capítulo anterior. Para efeito de comparação, foram considerados critérios de seleção de modelos amplamente utilizados no contexto bayesiano como EAIC, EBIC e DIC. Os cálculos foram feitos utilizando o software R ([R Core Team, 2017](#)).

### 3.1 Definições Preliminares

Nesta seção, introduzimos a parametrização das distribuições que são utilizadas no estudo de simulação e na análise dos dados de apendicectomia para gêmeos adultos.

**Definição 3.1.** A variável aleatória  $X$  tem distribuição Beta com parâmetros  $a \in \mathbb{R}_{>0}$ ,  $b \in \mathbb{R}_{>0}$ , se  $\forall x \in \mathbb{R}$  sua função densidade de probabilidade pode ser escrita como

$$f(x) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} x^{a-1} (1-x)^{b-1} \mathbb{1}_{(0,1)}(x),$$

em que  $\Gamma$  denota a função gama.

**Definição 3.2.** Dada uma partição finita de  $\mathbb{R}_{\geq 0}$ ,  $0 = a_1 < a_2 < \dots < a_{L+1} = \infty$ , a variável aleatória  $X$  tem distribuição Exponencial por Partes (PWE) com parâmetros  $\lambda_1 \in \mathbb{R}_{>0}, \dots, \lambda_L \in \mathbb{R}_{>0}$  se sua função de risco, definida como o quociente entre a função densidade de probabilidade e a função de sobrevivência, pode ser escrita como

$$h(x) = \sum_{j=1}^L \lambda_j \mathbb{1}_{[a_j, a_{j+1})}(x).$$

**Definição 3.3.** A variável aleatória  $X$  tem distribuição Gama com parâmetros  $\alpha \in \mathbb{R}_{>0}$ ,  $\beta \in \mathbb{R}_{>0}$ , se  $\forall x \in \mathbb{R}$  sua função densidade de probabilidade pode ser escrita como



$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbb{1}_{\mathbb{R}_{>0}}(x).$$

**Definição 3.4.** A variável aleatória  $X$  tem distribuição Normal com parâmetros  $\mu \in \mathbb{R}$ ,  $\tau \in \mathbb{R}_{>0}$ , se  $\forall x \in \mathbb{R}$  sua função densidade de probabilidade pode ser escrita como

$$f(x) = \sqrt{\frac{\tau}{2\pi}} e^{-\frac{\tau(x-\mu)^2}{2}} \mathbb{1}_{\mathbb{R}}(x),$$

em que  $\tau$  é chamado de parâmetro de precisão.

**Definição 3.5.** A variável aleatória  $X$  tem distribuição Weibull com parâmetros  $\rho \in \mathbb{R}_{>0}$ ,  $\beta_0 \in \mathbb{R}$ , se  $\forall x \in \mathbb{R}$  sua função densidade de probabilidade pode ser escrita como

$$f(x) = \frac{\rho}{e^{-\beta_0}} \left( \frac{t}{e^{-\beta_0}} \right)^{\rho-1} e^{-\left( \frac{t}{e^{-\beta_0}} \right)^\rho} \mathbb{1}_{\mathbb{R}_{>0}}(x).$$

## 3.2 Estudo de Simulação

Para construção do estudo de simulação, quatro principais aspectos foram considerados: o modelo de cópula, as distribuições marginais, o número de amostras e o percentual de censura. As cópulas escolhidas foram Clayton(5), Gumbel(0,85) e IG(0,1), cuja definição está na Seção 2.5 e a ilustração está na Figura 5. Estes modelos específicos de cópulas incluem tanto casos de baixa dependência, por exemplo o modelo Gumbel(0,85), cujo tau de Kendall é 0,15, quanto casos de alta dependência, por exemplo o modelo Clayton(5), cujo tau de Kendall é aproximadamente 0,71. Para as distribuições marginais, foram considerados os casos com  $T_1$  e  $T_2$  i.d. (identicamente distribuídas), em que  $T_1 \sim \text{Weibull}(5,25; -2,75)$  e  $T_2 \sim \text{Weibull}(5,25; -2,75)$ , e o caso com  $T_1$  e  $T_2$  não i.d., em que  $T_1 \sim \text{Weibull}(2,75; -2,25)$  e  $T_2 \sim \text{Weibull}(7,5; -3)$ , cuja ilustração está na Figura 6. Para o número de amostras, foram considerados os tamanhos 100 e 1000. Por fim, foram considerados dois casos para o percentual de censura: no primeiro, tanto  $T_1$  quanto  $T_2$  tiveram 75% e, no segundo, tanto  $T_1$  quanto  $T_2$  tiveram 20% de censura.

Cada cenário foi construído a partir da combinação da escolha de um modelo de cópula (Clayton, Gumbel ou IG), da estrutura das distribuições marginais (i.d. ou não i.d.), do número de amostras (100 e 1000) e do percentual de censura (com 75% ou com 20% de censura), resultando em 24 cenários. Em todos estes cenários, foi escolhido o número de réplicas igual a 100. A estrutura do estudo de simulação consiste em obter uma amostra a partir de um cenário escolhido, ajustar todos os modelos de cópulas e avaliar a capacidade do FBST para indicar o melhor modelo frente aos demais critérios de seleção utilizados na literatura como EAIC, EBIC e DIC, apresentados na Seção 2.4.

O ajuste de cada modelo foi feito considerando distribuição Weibull( $\rho, \beta_0$ ) para as marginais, com as seguintes distribuições a priori:  $\rho \sim \text{Gama}(0,001; 0,001)$  e  $\beta_0 \sim \text{Normal}(0;$

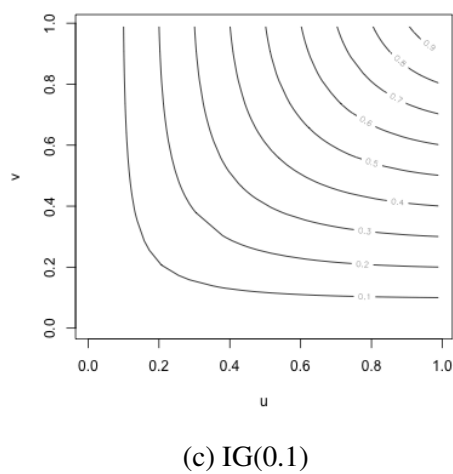
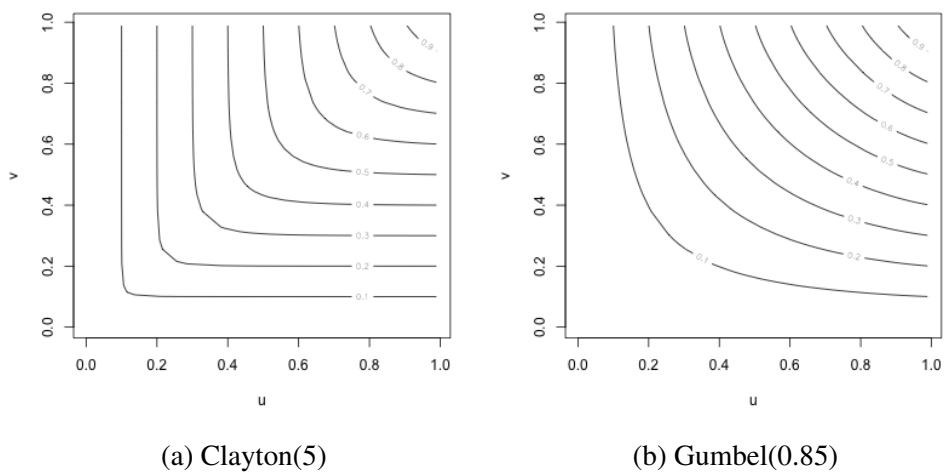


Figura 5 – Cópulas escolhidas para o estudo de simulação.

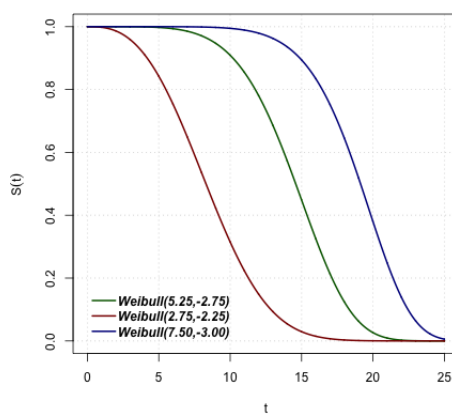


Figura 6 – Distribuições marginais escolhidas para o estudo de simulação.

0,0001), em que as distribuições Gama e Normal seguem a definição apresentada na Seção 3.1. A distribuição a priori para os parâmetros da cópula, por sua vez, foi escolhida da seguinte maneira:  $\alpha \sim \text{Beta}(1;1)$  e  $\eta \sim \text{Gama}(0,1;0,1)$ .

Duas perspectivas para o FBST foram consideradas, sendo que a primeira, indicada pelo FBST1, representa o percentual de vezes em que a hipótese referente ao modelo do qual os dados foram simulados não foi rejeitada ao nível de 5% e segunda e mais rigorosa, indicada pelo FBST2, representa o percentual de vezes em que não somente a hipótese não foi rejeitada como também é a hipótese com maior evidência dentre todas.

A Tabela 1 mostra os principais resultados do estudo de simulação. Em relação ao número de amostras, pode-se notar que os critérios de seleção de um modo geral tendem a identificar e escolher o modelo do qual os dados foram simulados com maior precisão na medida em que o tamanho amostral aumenta, assim como tendem a escolher de maneira correta na medida em que o percentual de censura diminui. O fato das marginais serem ou não identicamente distribuídas não parece afetar de maneira significativa a capacidade de escolha do modelo do qual os dados foram simulados.

Tabela 1 – Percentual de acerto da escolha do modelo mais adequado aos dados por cada critério de seleção de modelos.

Marginais i.d.	Amostras	Censura	Cópula	EAIC	EBIC	DIC	FBST1	FBST2
Não	1000	75%	Clayton	96%	<b>99%</b>	87%	<b>99%</b>	84%
Não	1000	75%	Gumbel	95%	99%	67%	<b>100%</b>	<b>100%</b>
Não	1000	75%	IG	81%	83%	76%	<b>100%</b>	99%
Não	1000	20%	Clayton	95%	<b>100%</b>	91%	99%	99%
Não	1000	20%	Gumbel	96%	99%	73%	<b>100%</b>	<b>100%</b>
Não	1000	20%	IG	94%	99%	88%	<b>100%</b>	<b>100%</b>
Não	100	75%	Clayton	72%	72%	69%	<b>88%</b>	1%
Não	100	75%	Gumbel	63%	63%	10%	<b>100%</b>	16%
Não	100	75%	IG	38%	38%	39%	<b>100%</b>	81%
Não	100	20%	Clayton	97%	<b>99%</b>	89%	<b>99%</b>	77%
Não	100	20%	Gumbel	76%	76%	9%	<b>100%</b>	33%
Não	100	20%	IG	78%	78%	67%	<b>100%</b>	87%
Sim	1000	75%	Clayton	97%	<b>99%</b>	91%	97%	89%
Sim	1000	75%	Gumbel	93%	96%	69%	<b>100%</b>	98%
Sim	1000	75%	IG	80%	81%	76%	<b>100%</b>	<b>100%</b>
Sim	1000	20%	Clayton	96%	<b>100%</b>	87%	98%	98%
Sim	1000	20%	Gumbel	96%	<b>100%</b>	79%	<b>100%</b>	<b>100%</b>
Sim	1000	20%	IG	89%	<b>99%</b>	84%	<b>99%</b>	<b>99%</b>
Sim	100	75%	Clayton	80%	80%	74%	<b>88%</b>	0%
Sim	100	75%	Gumbel	71%	71%	7%	<b>100%</b>	26%
Sim	100	75%	IG	45%	45%	41%	<b>100%</b>	74%
Sim	100	20%	Clayton	98%	<b>100%</b>	90%	99%	75%
Sim	100	20%	Gumbel	77%	77%	7%	<b>100%</b>	24%
Sim	100	20%	IG	74%	74%	71%	<b>100%</b>	87%

A Figura 7, por sua vez, mostra o comportamento dos resultados obtidos nos critérios de seleção de modelos sob as condições de marginais identicamente distribuídas, número de amostras igual a 1000, percentual de censura de 75% nas marginais e dados simulados a partir do modelo de cópula Gumbel. É possível notar pelas Figuras 7a, 7b e 7c que a distribuição das medidas EAIC, EBIC e DIC se comportam de maneira semelhante, no entanto as cópulas Gumbel e PVF apresentam valor mediano relativamente menor se comparado aos modelos de Clayton e IG. Uma vez que menores valores destas medidas indicam melhor adequação do modelo aos dados, pode-se dizer que neste caso estes três critérios de seleção tendem a escolher o modelo do qual os dados realmente foram simulados, embora essas medidas se comportem de maneira similar para todos os modelos de cópula.

O *e-value* para as réplicas consideradas no estudo, mostrados na Figura 7d, apresenta um resultado interessante. Para os modelos de Clayton e IG, os valores obtidos, com exceção dos *outliers*, são relativamente baixos, o que implica em baixa evidência desses modelos para os dados. Por outro lado, os valores obtidos para a cópula de Gumbel são significativamente maiores e implicam em maior evidência em favor deste modelo. Desta forma, mesmo que todos os critérios de seleção de modelos tenham obtido um percentual alto de acerto, o processo de decisão via FBST se mostra mais concreto e preciso quando comparado aos demais.

### 3.3 Dados de Apendicectomia

Nesta seção, dados sobre o tempo de apendicectomia para gêmeos adultos registrados no Australian NH&MRC Twin Registry (ATR) (DUFFY; MARTIN; MATHEWS, 1990) são analisados. Este estudo foi conduzido para investigar se a força de dependência entre pares de gêmeos quanto ao risco de aparecimento de várias doenças, incluindo apendicite aguda, é diferente para gêmeos monozigóticos (MZ) e dizigóticos (DZ). Uma vez que qualquer efeito potencial de um ambiente compartilhado seria semelhante para gêmeos MZ e DZ, uma dependência mais forte nos riscos de apendicite entre os membros dos pares de gêmeos MZ seria indicativo de um efeito genético sobre o risco de apendicite aguda e evidência do papel da hereditariedade no início da apendicite.

Para detalhes sobre a forma como as informações foram coletadas na pesquisa, consulte (DUFFY; MARTIN; MATHEWS, 1990) e (FAN; HSU; PRENTICE, 2000). Como em (FAN; HSU; PRENTICE, 2000) e (HU *et al.*, 2011), são apresentados os resultados da análise de pares de gêmeos do sexo feminino que realizaram apendicectomia, cuja base de dados compreende 1231 gêmeos MZ e 748 gêmeos DZ. Como mostra a Tabela 2, havia 248 pares nos quais ambos os membros foram submetidos à apendicectomia, 544 pares nos quais apenas um membro foi submetido à apendicectomia e 1187 pares nos quais nenhum membro foi submetido à apendicectomia. Indivíduos que não foram submetidos à apendicectomia antes da pesquisa ou suspeitos de serem submetidos à apendicectomia profilática foram considerados tempos de falha

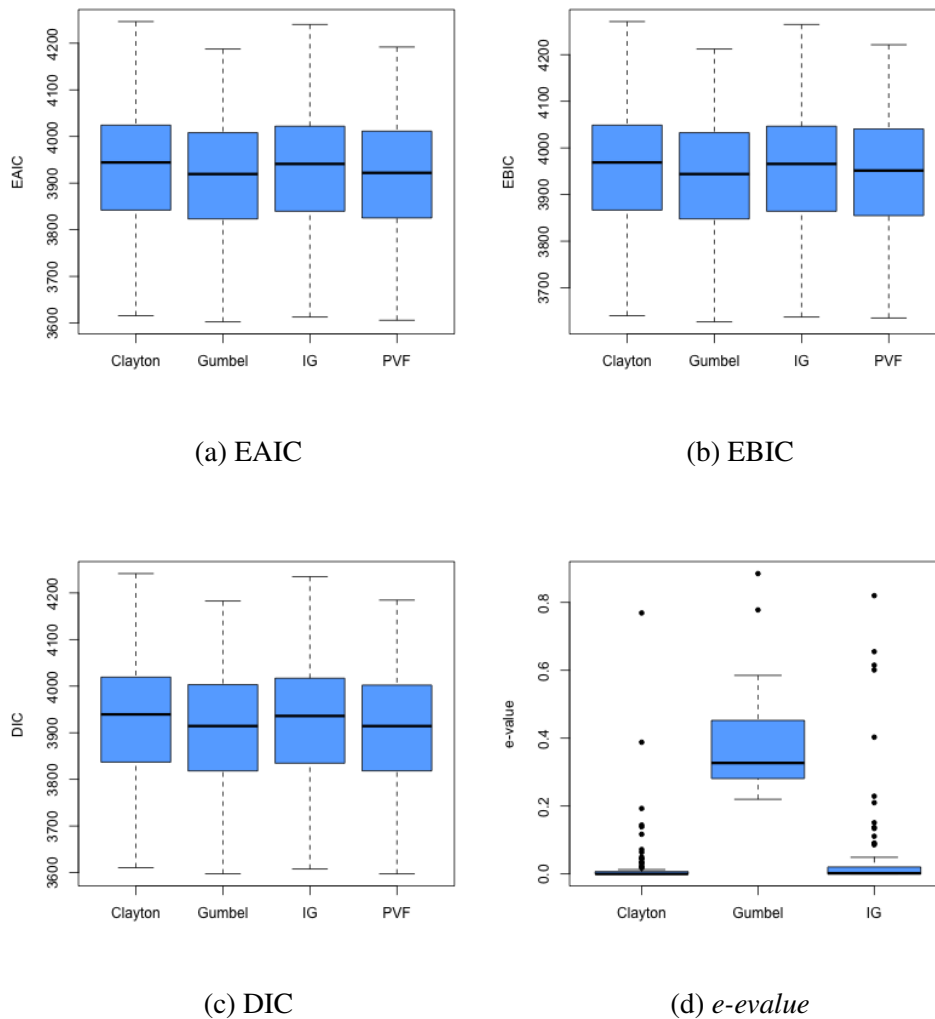


Figura 7 – EAIC, EBIC, DIC e *e-value* obtidos nas 100 réplicas do estudo de simulação para o caso de marginais i.d., número de amostral igual a 1000, 75% de censura em cada marginal e modelo verdadeiro Gumbel.

censurados (aproximadamente 73% para cada membro gêmeo em qualquer tipo de zigoto).

Tabela 2 – Distribuição de pares de gêmeos australianos por apendicectomia.

Gêmeos	Apendicectomia			Total
	Ambos	Um	Nenhum	
MZ	166	322	743	1231
DZ	82	222	444	748
Total	248	544	1187	1979

Os conjuntos de dados para gêmeos MZ e DZ são analisados separadamente. O objetivo em ambas as análises está em ajustar os modelos de cópulas PVF, Clayton, Gumbel e IG considerando, em um primeiro momento, distribuição Weibull para as marginais e, em um segundo momento, distribuição Exponencial por Partes para as marginais, que é uma alternativa mais flexível quando comparada ao modelo Weibull, e fazer a escolha do modelo que melhor

se adequa aos dados via FBST, além da comparação deste critério com os demais critérios de seleção descritos no capítulo anterior.

Para as marginais, as seguintes prioris foram escolhidas:  $\rho \sim \text{Gama}(0,001;0,001)$  e  $\beta_0 \sim \text{Normal}(0; 0,0001)$  no caso de marginais Weibull( $\rho, \beta_0$ ) e  $\lambda_i \sim \text{Gama}(0,001;0,001)$ ,  $i = 1, \dots, 7$  no caso de marginais PWE( $\lambda_1, \dots, \lambda_7$ ), com partição  $(0,9,15,30,38,44,54,\infty)$  para  $T_1$  e  $(0,8,12,27,35,44,53,\infty)$  para  $T_2$ . Para os parâmetros da cópula, as seguintes prioris foram escolhidas:  $\alpha \sim \text{Beta}(1;1)$  e  $\eta \sim \text{Gama}(0,1;0,1)$ .

### 3.3.1 Gêmeos Monozigóticos (MZ)

A Tabela 3 mostra a média, mediana, desvio padrão, limite inferior e superior do intervalo HPD de 95% a posteriori considerando distribuição Weibull para as sobrevivências marginais para gêmeos MZ. Nota-se que estas medidas para os parâmetros das marginais são muito semelhantes independente do modelo de cópula escolhido, com valores médio e mediano próximos a 1,5 para  $\rho_1$  e  $\rho_2$  e valores médio e medianos próximos a  $-4,4$  para  $\beta_{01}$  e  $\beta_{02}$ . O limite inferior do intervalo de maior densidade a posteriori (HPD) para o parâmetro  $\alpha$  da cópula PVF e da cópula de Gumbel está bem distante de 0, enquanto que o limite inferior do intervalo HPD para o parâmetro  $\eta$  está próximo de 0, indicando, desta maneira, que o modelo Gumbel possa ser plausível para este conjunto de dados, considerando distribuição Weibull para as marginais. Esta conclusão se apoia no fato de que se  $\eta \rightarrow 0$  na cópula PVF, o modelo de Gumbel é obtido.

Tabela 3 – Média, mediana, desvio padrão, limite inferior e superior do intervalo HPD de 95% considerando distribuição Weibull para as sobrevivências marginais para gêmeos MZ.

Cópula	Parâmetro	Média	Mediana	Desvio Padrão	LI HPD	LS HPD
PVF	$\alpha$	0,7069	0,7066	0,0253	0,6633	0,7596
	$\eta$	0,0009	0,0004	0,0022	0,0001	0,0037
	$\rho_1$	1,4742	1,4726	0,0666	1,3475	1,6113
	$\beta_{01}$	-4,3439	-4,3415	0,0486	-4,4411	-4,2550
	$\rho_2$	1,4593	1,4585	0,0705	1,3246	1,6056
	$\beta_{02}$	-4,3943	-4,3930	0,0535	-4,4964	-4,2935
Clayton	$\eta$	1,3843	1,3870	0,1889	1,0099	1,7212
	$\rho_1$	1,5266	1,5228	0,0671	1,4058	1,6660
	$\beta_{01}$	-4,3192	-4,3171	0,0473	-4,4042	-4,2260
	$\rho_2$	1,5080	1,5060	0,0670	1,3799	1,6420
	$\beta_{02}$	-4,3661	-4,3632	0,0485	-4,4629	-4,2763
Gumbel	$\alpha$	0,7082	0,7080	0,0244	0,6613	0,7582
	$\rho_1$	1,4688	1,4667	0,0668	1,3277	1,5961
	$\beta_{01}$	-4,3433	-4,3418	0,0492	-4,4394	-4,2486

Continua na próxima página

Tabela 3 – continuação da página anterior

Cópula	Parâmetro	Média	Mediana	Desvio Padrão	LI HPD	LS HPD
	$\rho_2$	1,4512	1,4516	0,0690	1,3057	1,5752
	$\beta_{02}$	-4,3978	-4,3960	0,0518	-4,5008	-4,3054
	$\eta$	0,1566	0,1533	0,0383	0,0904	0,2320
	$\rho_1$	1,5119	1,5098	0,0689	1,3886	1,6505
IG	$\beta_{01}$	-4,3296	-4,3249	0,0496	-4,4186	-4,2302
	$\rho_2$	1,4989	1,4985	0,0693	1,3719	1,6304
	$\beta_{02}$	-4,3762	-4,3742	0,0496	-4,4850	-4,2919

De maneira semelhante, a Tabela 4 mostra a média, mediana, desvio padrão, limite inferior e superior do intervalo HPD de 95% considerando distribuição Exponencial por Partes para as sobrevivências marginais para gêmeos MZ. Para o caso de marginais Exponencial por Partes, observa-se que a média, mediana e desvio padrão a posteriori para os parâmetros  $\lambda_{ij}$ ,  $i, j = 1, \dots, 7$  possuem semelhança considerável. Assim como no caso de marginais Weibull, o limite inferior do intervalo HPD para o parâmetro  $\alpha$  da cópula PVF e da cópula de Gumbel está bem distante de 0, indicando que este modelo possa ser plausível. Além disso, o limite inferior do intervalo HPD para o parâmetro  $\eta$  da cópula PVF está próximo de 0, o que implica que os modelos de Clayton e IG possam não ser totalmente adequados ao problema.

Tabela 4 – Média, mediana, desvio padrão, limite inferior e superior do intervalo HPD de 95% considerando distribuição Exponencial por Partes para as sobrevivências marginais para gêmeos MZ.

Cópula	Parâmetro	Média	Mediana	Desvio Padrão	LI HPD	LS HPD
	$\alpha$	0,7153	0,7193	0,0341	0,6412	0,7743
	$\eta$	0,0060	0,0013	0,0121	0,0001	0,0316
	$\lambda_{11}$	0,0025	0,0025	0,0005	0,0016	0,0034
	$\lambda_{21}$	0,0111	0,0111	0,0011	0,0090	0,0133
	$\lambda_{31}$	0,0129	0,0128	0,0012	0,0104	0,0152
	$\lambda_{41}$	0,0139	0,0139	0,0017	0,0109	0,0175
	$\lambda_{51}$	0,0118	0,0116	0,0028	0,0069	0,0172
PVF	$\lambda_{61}$	0,0060	0,0057	0,0020	0,0028	0,0100
	$\lambda_{71}$	0,0042	0,0040	0,0014	0,0017	0,0068
	$\lambda_{12}$	0,0017	0,0017	0,0005	0,0008	0,0026
	$\lambda_{22}$	0,0090	0,0089	0,0010	0,0071	0,0109
	$\lambda_{32}$	0,0142	0,0142	0,0013	0,0118	0,0169

Continua na próxima página

Tabela 4 – continuação da página anterior

Cópula	Parâmetro	Média	Mediana	Desvio Padrão	LI HPD	LS HPD
	$\lambda_{42}$	0,0097	0,0097	0,0014	0,0069	0,0124
	$\lambda_{52}$	0,0139	0,0138	0,0022	0,0098	0,0183
	$\lambda_{62}$	0,0107	0,0105	0,0026	0,0061	0,0160
	$\lambda_{72}$	0,0039	0,0038	0,0012	0,0018	0,0062
	$\eta$	1,6607	1,6581	0,2319	1,1925	2,0856
Clayton	$\lambda_{11}$	0,0023	0,0022	0,0005	0,0012	0,0032
	$\lambda_{21}$	0,0111	0,0111	0,0011	0,0089	0,0130
	$\lambda_{31}$	0,0131	0,0131	0,0012	0,0107	0,0153
	$\lambda_{41}$	0,0141	0,0140	0,0018	0,0106	0,0174
	$\lambda_{51}$	0,0115	0,0112	0,0026	0,0069	0,0167
	$\lambda_{61}$	0,0056	0,0054	0,0019	0,0025	0,0097
	$\lambda_{71}$	0,0042	0,0040	0,0014	0,0018	0,0068
	$\lambda_{12}$	0,0014	0,0014	0,0004	0,0007	0,0021
	$\lambda_{22}$	0,0089	0,0088	0,0010	0,0069	0,0108
	$\lambda_{32}$	0,0145	0,0144	0,0014	0,0119	0,0174
	$\lambda_{42}$	0,0097	0,0097	0,0014	0,0066	0,0123
	$\lambda_{52}$	0,0136	0,0135	0,0021	0,0100	0,0183
	$\lambda_{62}$	0,0101	0,0100	0,0024	0,0054	0,0145
	$\lambda_{72}$	0,0038	0,0038	0,0011	0,0019	0,0063
Gumbel	$\alpha$	0,7265	0,7272	0,0261	0,6744	0,7750
	$\lambda_{11}$	0,0026	0,0026	0,0005	0,0016	0,0035
	$\lambda_{21}$	0,0110	0,0110	0,0010	0,0091	0,0132
	$\lambda_{31}$	0,0128	0,0127	0,0012	0,0105	0,0152
	$\lambda_{41}$	0,0140	0,0139	0,0018	0,0109	0,0179
	$\lambda_{51}$	0,0117	0,0114	0,0028	0,0066	0,0173
	$\lambda_{61}$	0,0061	0,0058	0,0021	0,0028	0,0104
	$\lambda_{71}$	0,0042	0,0040	0,0015	0,0018	0,0073
	$\lambda_{12}$	0,0018	0,0018	0,0005	0,0010	0,0027
	$\lambda_{22}$	0,0089	0,0088	0,0009	0,0073	0,0110
	$\lambda_{32}$	0,0142	0,0142	0,0014	0,0118	0,0170
	$\lambda_{42}$	0,0096	0,0095	0,0016	0,0067	0,0127
	$\lambda_{52}$	0,0140	0,0139	0,0021	0,0099	0,0182
	$\lambda_{62}$	0,0105	0,0103	0,0025	0,0058	0,0158
$\lambda_{72}$	0,0040	0,0039	0,0012	0,0018	0,0063	
	$\eta$	0,1306	0,1273	0,0308	0,0738	0,1886
	$\lambda_{11}$	0,0023	0,0023	0,0005	0,0014	0,0033

Continua na próxima página



Tabela 4 – continuação da página anterior

Cópula	Parâmetro	Média	Mediana	Desvio Padrão	LI HPD	LS HPD
	$\lambda_{21}$	0,0113	0,0112	0,0011	0,0093	0,0136
	$\lambda_{31}$	0,0131	0,0130	0,0013	0,0109	0,0159
	$\lambda_{41}$	0,0142	0,0141	0,0017	0,0114	0,0180
	$\lambda_{51}$	0,0115	0,0113	0,0027	0,0066	0,0166
	$\lambda_{61}$	0,0059	0,0057	0,0019	0,0024	0,0099
	$\lambda_{71}$	0,0042	0,0041	0,0014	0,0019	0,0072
	$\lambda_{12}$	0,0014	0,0014	0,0004	0,0006	0,0022
	$\lambda_{22}$	0,0090	0,0090	0,0010	0,0072	0,0112
	$\lambda_{32}$	0,0146	0,0146	0,0013	0,0120	0,0172
	$\lambda_{42}$	0,0097	0,0096	0,0015	0,0071	0,0127
	$\lambda_{52}$	0,0138	0,0137	0,0022	0,0094	0,0178
	$\lambda_{62}$	0,0102	0,0100	0,0024	0,0058	0,0152
	$\lambda_{72}$	0,0038	0,0037	0,0011	0,0018	0,0061

A Tabela 5 traz os critérios de seleção de modelos para os dados de gêmeos monozigóticos. Conforme a análise das Tabelas 3 e 4, de fato o modelo de Gumbel possui maior evidência se comparado ao modelo de Clayton e IG, quando o *e-value* é considerado. Inclusive, segundo esta medida e através do FBST a hipótese de que estes dois modelos são plausíveis seria rejeitada ao nível de 5%, tanto para o caso de marginais Weibull quanto Exponencial por Partes. Os demais critérios de seleção de modelos também apontam a cópula de Gumbel como mais adequada ao conjunto de dados em questão. Em relação ao processo de escolha das marginais, estes critérios apontam para a distribuição Exponencial por Partes como mais adequada.

Tabela 5 – Tau de Kendall estimado e critérios de seleção de modelos para gêmeos MZ.

Modelo	$\tau$ mediano (sd)	EAIC	EBIC	DIC	<i>e-value</i>
PVF-Weibull	0,291 (0,024)	7255,809	7286,502	7248,631	–
Clayton-Weibull	0,409 (0,033)	7292,410	7317,988	7287,375	<0,001
Gumbel-Weibull	0,291 (0,024)	7252,416	7277,994	7247,342	<b>0,253</b>
IG-Weibull	0,346 (0,022)	7279,053	7304,631	7274,037	<0,001
PVF-PWE	0,275 (0,027)	7192,654	7274,504	7175,032	–
Clayton-PWE	0,453 (0,034)	7216,625	7293,359	7201,791	0,004
Gumbel-PWE	0,272 (0,026)	<b>7189,477</b>	<b>7266,211</b>	<b>7174,947</b>	<b>0,950</b>
IG-PWE	0,363 (0,020)	7201,822	7278,556	7186,914	0,303

### 3.3.2 Gêmeos Dizigóticos (DZ)

A média, mediana, desvio padrão, limite inferior e superior do intervalo HPD de 95% a posteriori considerando distribuição Weibull para as sobrevivências marginais para gêmeos DZ são apresentadas na Tabela 6. Assim como no caso dos dados de gêmeos monozigóticos, nota-se que estas medidas para os parâmetros das marginais são muito similares para todos os modelos de cópula, com valores médio e mediano próximos a 1,55 para  $\rho_1$  e 1,45  $\rho_2$  e valores médio e medianos próximos a  $-4,4$  para  $\beta_{01}$  e  $\beta_{02}$ . Em se tratando do intervalo HPD, o limite inferior para o parâmetro  $\alpha$  da cópula PVF e da cópula de Gumbel não é próximo de 0 e, por outro lado, o limite inferior para o parâmetro  $\eta$  está próximo de 0, indicando, desta maneira, que o modelo Gumbel possa ser plausível para este conjunto de dados, considerando marginais Weibull. Esta conclusão se apoia no fato de que se  $\eta \rightarrow 0$  na cópula PVF, o modelo de Gumbel é obtido.

Tabela 6 – Média, mediana, desvio padrão, limite inferior e superior do intervalo HPD de 95% considerando distribuição Weibull para as sobrevivências marginais para gêmeos DZ.

Cópula	Parâmetro	Média	Mediana	Desvio Padrão	LI HPD	LS HPD
PVF	$\alpha$	0,8301	0,8341	0,0500	0,7595	0,9100
	$\eta$	0,0179	0,0012	0,0750	0,0001	0,0779
	$\rho_1$	1,5333	1,5348	0,0907	1,3559	1,7066
	$\beta_{01}$	-4,3875	-4,3835	0,0649	-4,5088	-4,2606
	$\rho_2$	1,4446	1,4410	0,0874	1,2788	1,6156
	$\beta_{02}$	-4,3664	-4,3629	0,0680	-4,4960	-4,2382
Clayton	$\eta$	0,5970	0,5906	0,1803	0,2434	0,9542
	$\rho_1$	1,5576	1,5600	0,0915	1,3815	1,7389
	$\beta_{01}$	-4,3806	-4,3775	0,0664	-4,5299	-4,2711
	$\rho_2$	1,4663	1,4663	0,0854	1,2956	1,6186
	$\beta_{02}$	-4,3560	-4,3526	0,0636	-4,4757	-4,2353
Gumbel	$\alpha$	0,8392	0,8394	0,0311	0,7760	0,8985
	$\rho_1$	1,5360	1,5337	0,0909	1,3557	1,7035
	$\beta_{01}$	-4,3860	-4,3768	0,0665	-4,5169	-4,2597
	$\rho_2$	1,4391	1,4342	0,0877	1,2776	1,6107
	$\beta_{02}$	-4,3682	-4,3666	0,0686	-4,4947	-4,2310
IG	$\eta$	0,6512	0,5851	0,2921	0,2245	1,2116
	$\rho_1$	1,5579	1,5528	0,0952	1,3776	1,7329
	$\beta_{01}$	-4,3788	-4,3723	0,0652	-4,5079	-4,2544
	$\rho_2$	1,4577	1,4597	0,0878	1,3043	1,6505

Continua na próxima página

Tabela 6 – continuação da página anterior

Cópula	Parâmetro	Média	Mediana	Desvio Padrão	LI HPD	LS HPD
	$\beta_{02}$	-4,3602	-4,3579	0,0675	-4,4928	-4,2354

Analogamente, a Tabela 4 mostra a média, mediana, desvio padrão, limite inferior e superior do intervalo HPD de 95% considerando distribuição Exponencial por Partes para as sobrevivências marginais para gêmeos dizigóticos. Observa-se, para marginais Exponencial por Partes, que a média, mediana e desvio padrão a posteriori para os parâmetros  $\lambda_{ij}$ ,  $i, j = 1, \dots, 7$  possuem semelhança considerável. Em relação ao intervalo HPD, o limite inferior do para o parâmetro  $\alpha$  da cópula PVF e da cópula de Gumbel não está próximo de 0, o que indica que este modelo possa ser plausível para este conjunto de dados. Entretanto, o limite inferior do intervalo HPD para o parâmetro  $\eta$  da cópula PVF está próximo de 0, dando indícios de que os modelos de Clayton e IG possam não ser adequados ao conjunto de dados.

Tabela 7 – Média, mediana, desvio padrão, limite inferior e superior do intervalo HPD de 95% considerando distribuição Exponencial por Partes para as sobrevivências marginais para gêmeos DZ.

Cópula	Parâmetro	Média	Mediana	Desvio Padrão	LI HPD	LS HPD
PVF	$\alpha$	0,8017	0,8318	0,1274	0,5650	0,9127
	$\eta$	0,0662	0,0029	0,2253	0,0001	0,4177
	$\lambda_{11}$	0,0024	0,0024	0,0006	0,0013	0,0036
	$\lambda_{21}$	0,0097	0,0096	0,0015	0,0068	0,0123
	$\lambda_{31}$	0,0120	0,0120	0,0012	0,0098	0,0147
	$\lambda_{41}$	0,0104	0,0102	0,0024	0,0062	0,0151
	$\lambda_{51}$	0,0021	0,0018	0,0015	0,0001	0,0050
	$\lambda_{61}$	0,0135	0,0132	0,0038	0,0071	0,0215
	$\lambda_{71}$	0,0048	0,0042	0,0027	0,0007	0,0101
	$\lambda_{12}$	0,0030	0,0029	0,0007	0,0017	0,0045
	$\lambda_{22}$	0,0108	0,0107	0,0019	0,0073	0,0147
	$\lambda_{32}$	0,0121	0,0121	0,0012	0,0098	0,0143
	$\lambda_{42}$	0,0081	0,0080	0,0019	0,0048	0,0120
	$\lambda_{52}$	0,0115	0,0113	0,0028	0,0061	0,0169
	$\lambda_{62}$	0,0093	0,0089	0,0034	0,0034	0,0163
	$\lambda_{72}$	0,0075	0,0070	0,0037	0,0016	0,0148
		$\eta$	0,7248	0,7171	0,2093	0,3357
	$\lambda_{11}$	0,0022	0,0022	0,0006	0,0011	0,0033

Continua na próxima página

Tabela 7 – continuação da página anterior

Cópula	Parâmetro	Média	Mediana	Desvio Padrão	LI HPD	LS HPD
	$\lambda_{21}$	0,0097	0,0096	0,0015	0,0069	0,0126
	$\lambda_{31}$	0,0120	0,0120	0,0013	0,0094	0,0143
	$\lambda_{41}$	0,0103	0,0101	0,0023	0,0059	0,0148
	$\lambda_{51}$	0,0022	0,0018	0,0016	0,0001	0,0053
	$\lambda_{61}$	0,0130	0,0128	0,0036	0,0057	0,0197
	$\lambda_{71}$	0,0046	0,0041	0,0027	0,0005	0,0095
	$\lambda_{12}$	0,0028	0,0028	0,0007	0,0017	0,0043
	$\lambda_{22}$	0,0109	0,0108	0,0020	0,0071	0,0149
	$\lambda_{32}$	0,0122	0,0122	0,0012	0,0101	0,0144
	$\lambda_{42}$	0,0081	0,0080	0,0019	0,0043	0,0115
	$\lambda_{52}$	0,0113	0,0110	0,0027	0,0067	0,0171
	$\lambda_{62}$	0,0096	0,0092	0,0035	0,0036	0,0163
	$\lambda_{72}$	0,0078	0,0072	0,0039	0,0012	0,0151
	$\alpha$	0,8459	0,8471	0,0324	0,7892	0,9154
	$\lambda_{11}$	0,0024	0,0024	0,0006	0,0013	0,0035
	$\lambda_{21}$	0,0097	0,0097	0,0016	0,0064	0,0125
	$\lambda_{31}$	0,0120	0,0120	0,0012	0,0095	0,0144
	$\lambda_{41}$	0,0105	0,0103	0,0024	0,0062	0,0156
	$\lambda_{51}$	0,0022	0,0018	0,0015	0,0001	0,0052
	$\lambda_{61}$	0,0135	0,0132	0,0038	0,0068	0,0208
Gumbel	$\lambda_{71}$	0,0047	0,0041	0,0027	0,0004	0,0100
	$\lambda_{12}$	0,0031	0,0030	0,0007	0,0018	0,0045
	$\lambda_{22}$	0,0109	0,0109	0,0019	0,0075	0,0151
	$\lambda_{32}$	0,0122	0,0121	0,0012	0,0100	0,0146
	$\lambda_{42}$	0,0082	0,0081	0,0018	0,0050	0,0119
	$\lambda_{52}$	0,0114	0,0112	0,0027	0,0065	0,0168
	$\lambda_{62}$	0,0097	0,0091	0,0034	0,0038	0,0165
	$\lambda_{72}$	0,0074	0,0069	0,0037	0,0017	0,0146
	$\eta$	0,5245	0,4708	0,2489	0,1973	0,9544
	$\lambda_{11}$	0,0022	0,0021	0,0006	0,0010	0,0032
	$\lambda_{21}$	0,0096	0,0095	0,0015	0,0068	0,0125
	$\lambda_{31}$	0,0120	0,0120	0,0013	0,0096	0,0145
	$\lambda_{41}$	0,0103	0,0102	0,0023	0,0063	0,0149
	$\lambda_{51}$	0,0023	0,0019	0,0016	0,0002	0,0055
	$\lambda_{61}$	0,0132	0,0128	0,0037	0,0059	0,0199
IG	$\lambda_{71}$	0,0047	0,0042	0,0028	0,0002	0,0099

Continua na próxima página

Tabela 7 – continuação da página anterior

Cópula	Parâmetro	Média	Mediana	Desvio Padrão	LI HPD	LS HPD
	$\lambda_{12}$	0,0029	0,0029	0,0007	0,0015	0,0043
	$\lambda_{22}$	0,0109	0,0109	0,0019	0,0074	0,0147
	$\lambda_{32}$	0,0122	0,0121	0,0012	0,0097	0,0145
	$\lambda_{42}$	0,0080	0,0078	0,0018	0,0045	0,0117
	$\lambda_{52}$	0,0112	0,0110	0,0027	0,0059	0,0164
	$\lambda_{62}$	0,0092	0,0088	0,0032	0,0035	0,0153
	$\lambda_{72}$	0,0078	0,0072	0,0039	0,0017	0,0160

Por fim, os critérios de seleção de modelos para os dados de gêmeos DZ são apresentados na Tabela 8. Em concordância com as Tabelas 6 e 7, de fato o modelo de Gumbel possui maior evidência se comparado aos demais modelos, quando o *e-value* é considerado. Segundo esta medida e através do FBST a hipótese de que estes dois modelos são plausíveis seria rejeitada ao nível de 5% para o caso de marginais Weibull, no entanto o mesmo não acontece para marginais Exponencial por Partes. Os demais critérios de seleção vão na mesma direção e apontam a cópula de Gumbel como mais adequada aos dados. Em relação ao processo de escolha das marginais, estes três critérios possuem valores semelhantes para o modelo Gumbel, fazendo com tanto Weibull quanto Exponencial por Partes sejam consideradas como marginais adequadas.

Tabela 8 – Tau de Kendall estimado e critérios de seleção de modelos para gêmeos DZ.

Modelo	$\tau$ mediano (sd)	EAIC	EBIC	DIC	<i>e-value</i>
PVF-Weibull	0,159 (0,031)	4363,380	4391,084	4354,066	–
Clayton-Weibull	0,227 (0,053)	4371,709	4394,796	4366,718	0,025
Gumbel-Weibull	0,160 (0,031)	4359,937	4383,024	4354,837	<b>0,414</b>
IG-Weibull	0,205 (0,039)	4370,329	4393,416	4365,145	0,034
PVF-PWE	0,159 (0,037)	4363,144	4437,023	<b>4344,159</b>	–
Clayton-PWE	0,263 (0,056)	4369,227	4438,488	4354,683	0,345
Gumbel-PWE	0,152 (0,032)	<b>4359,356</b>	<b>4428,617</b>	4344,610	<b>0,960</b>
IG-PWE	0,229 (0,040)	4367,552	4436,813	4352,849	0,507

---

## CONSIDERAÇÕES FINAIS

---

Dados de sobrevivência podem ser caracterizados por dois principais fatores, que são o tempo de falha e a censura. No contexto bivariado, isto é, quando se tem interesse no vetor aleatório  $(T_1, T_2)$ , geralmente com  $T_1$  e  $T_2$  não independentes, a censura pode ocorrer tanto em  $T_1$  ou  $T_2$  quanto em  $T_1$  e  $T_2$ . Embora alguns métodos estejam disponíveis na literatura para lidar com dados dessa natureza, conduzir um processo de seleção de modelos para dados bivariados e com censura não é uma tarefa simples.

Neste trabalho, utilizamos a metodologia de cópulas, que inicialmente foi discutida em (SKLAR, 1959), para lidar com cenários em que  $T_1$  e  $T_2$  não são independentes, pois, a partir do Teorema de Sklar, cópulas podem ser vistas como um elo de ligação entre as sobrevivências marginais e a sobrevivência conjunta. Para as sobrevivências marginais, duas abordagens foram consideradas: Weibull e Exponencial por Partes. Em se tratando de cópulas, foi considerada a cópula Power Variance Function (PVF), que pertence à família de cópulas Arquimedianas de dois parâmetros, com casos particulares amplamente utilizados na literatura: Clayton, Gumbel e IG. Trabalhar com casos particulares induz uma situação de modelos encaixados, o que possibilitou a utilização do Full Bayesian Significance Teste (FBST) como critério rigoroso de seleção de modelos, por se tratar de um teste de hipóteses. Comparamos esta abordagem com algumas das principais abordagens disponíveis na literatura no contexto de seleção de modelos: EAIC, EBIC e DIC.

Com o intuito de comparar o FBST com os demais critérios para avaliar a performance da metodologia proposta, construímos um estudo de simulação levando em conta diferentes tamanhos amostrais, percentual de censura, características das marginais (i.d. ou não) e modelos de cópulas. Os resultados do FBST foram avaliados segundo duas perspectivas diferentes: a primeira se concentrou em avaliar o percentual de vezes em que a hipótese referente ao modelo qual os dados vieram não era rejeitada e a segunda, mais rigorosa, avaliou o percentual de vezes em que não somente a hipótese não era rejeitada como também era a hipótese com maior

evidência dentre todas. A primeira perspectiva obteve resultados expressivos em todos os cenários frente aos demais critérios enquanto que a segunda, mais rigorosa, obteve resultados expressivos para tamanhos de amostra maiores. Em relação ao percentual de censura, os critérios de seleção de um modo geral tendem a identificar o modelo do qual os dados foram simulados com maior acurácia em cenários com menor percentual de censura. O fato das marginais serem identicamente distribuídas ou não, por sua vez, não impactou significativamente os resultados.

Após o estudo de simulação, utilizamos a metodologia proposta em um conjunto de dados real, que investiga se a força de dependência entre pares de gêmeos adultos quanto ao risco de aparecimento de diversas doenças é diferente para gêmeos monozigóticos e dizigóticos. Neste contexto, é analisado o tempo até a apendicectomia para gêmeos adultos registrados no Australian NH&MRC Twin Registry (ATR) (DUFFY; MARTIN; MATHEWS, 1990). Tanto para o conjunto de dados de gêmeos monozigóticos quanto para o conjunto de dados de gêmeos dizigóticos, as medidas calculadas a posteriori para os parâmetros das marginais, tanto no caso de marginais Weibull quanto Exponencial por Partes, mostraram-se bastante similares, o que indica que a estimação das marginais praticamente não mudou independente do modelo de cópula adotado. Analisando os critérios de seleção de modelos, a cópula de Gumbel e marginais Exponencial por Partes se mostraram mais aderentes aos dados tanto para gêmeos monozigóticos quanto para gêmeos dizigóticos em comparação com os demais modelos. Este resultado está na mesma direção que as estatísticas calculadas a posteriori para os parâmetros da cópula PVF, uma vez que o limite inferior do intervalo HPD para o parâmetro  $\alpha$  ficou distante de 0 e, por outro lado, o limite inferior do intervalo HPD para o parâmetro  $\eta$  ficou próximo de 0, que configura, no limite, o caso particular da cópula de Gumbel.

## 4.1 Sugestões para Pesquisas Futuras

Neste trabalho consideramos a seleção de modelos utilizando o Full Bayesian Significance Test (FBST) no contexto bivariado a partir da cópula PVF. Este modelo possui versão multivariada e uma possibilidade seria investigar o desempenho da metodologia proposta em dimensões de ordem superior, isto é,  $d > 2$ , em que a complexidade do problema de escolha do modelo adequado é maior.

Outro ponto relevante a ser investigado é a censura pois o estudo de simulação mostrou que o percentual de dados censurados influencia na escolha do modelo. Esses resultados foram obtidos com base em censura à direita mas seria interessante analisar o comportamento dos critérios de seleção de modelos para outros tipos de censura, como censura à esquerda e censura intervalar.

Em termos de modelo de cópula, foi visto que para o caso da cópula PVF utilizar o FBST como critério de seleção produziu bons resultados frente aos demais critérios. No entanto, outro ponto de investigação seria o comportamento destes critérios para outros modelos de cópulas no

contexto de modelos encaixados.





## REFERÊNCIAS

---

---

BROOKS, S.; SMITH, J.; VEHTARI, A.; PLUMMER, M.; STONE, M.; ROBERT, C. P.; TITTERINGTON, D.; NELDER, J.; ATKINSON, A.; DAWID, A. *et al.* Discussion on the paper by Spiegelhalter, Best, Carlin and van der Linde. **Journal of the Royal Statistical Society. Series B: Statistical Methodology**, Wiley-Blackwell, v. 64, n. 4, p. 616–639, 2002. Citado nas páginas 20 e 32.

CARLIN, B. P.; LOUIS, T. A. Bayes and empirical bayes methods for data analysis. Chapman & Hall, 2001. Citado nas páginas 20 e 32.

COLOSIMO, E.; GIOLO, S. **Análise de sobrevivência aplicada**. Edgard Blücher, 2006. (ABE - Projeto Fisher). ISBN 9788521203841. Disponível em: <<https://books.google.com.br/books?id=g0-uOgAACAAJ>>. Citado nas páginas 17 e 30.

DUFFY, D.; MARTIN, N.; MATHEWS, J. Appendectomy in Australian twins. **American Journal of Human Genetics**, Elsevier, v. 47, n. 3, p. 590–592, 1990. Citado nas páginas 18, 43 e 54.

FAN, J.; HSU, L.; PRENTICE, R. L. Dependence estimation over a finite bivariate failure time region. **Lifetime Data Analysis**, Springer, v. 6, n. 4, p. 343–355, 2000. Citado na página 43.

GROUP, D. R. S. R. *et al.* Preliminary report on effects of photocoagulation therapy. **American journal of ophthalmology**, Elsevier, v. 81, n. 4, p. 383–396, 1976. Citado na página 18.

GUMBEL, E. J. Bivariate exponential distributions. **Journal of the American Statistical Association**, Taylor & Francis, v. 55, n. 292, p. 698–707, 1960. Citado na página 25.

HOSMER, D. W.; LEMESHOW, S. **Applied Survival Analysis: Regression Modeling of Time to Event Data**. 1st. ed. USA: John Wiley & Sons, Inc., 1999. ISBN 0471154105. Citado na página 17.

HOUGAARD, P. Survival models for heterogeneous populations derived from stable distributions. **Biometrika**, Oxford University Press, v. 73, n. 2, p. 387–396, 1986. Citado nas páginas 20 e 34.

HU, T.; NAN, B.; LIN, X.; ROBINS, J. M. Time-dependent cross ratio estimation for bivariate failure times. **Biometrika**, Oxford University Press, v. 98, n. 2, p. 341–354, 2011. Citado na página 43.

KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. **Journal of the American Statistical Association**, Taylor & Francis, v. 53, n. 282, p. 457–481, 1958. Citado na página 18.

KRUSKAL, W. H. Ordinal measures of association. **Journal of the American Statistical Association**, v. 53, n. 284, p. 814–861, 1958. Citado na página 27.

LAWLESS, J. F. **Statistical models and methods for lifetime data**. [S.l.]: John Wiley & Sons, 2011. v. 362. Citado nas páginas 17 e 32.

- MALLICK, M.; RAVISHANKER, N. Pvf frailty models with a flexible baseline hazard. **International Journal of Statistics and Systems**, v. 1, p. 57–80, 2006. Citado na página 34.
- MEEKER, W. Q. Limited failure population life tests: application to integrated circuit reliability. **Technometrics**, Taylor & Francis, v. 29, n. 1, p. 51–65, 1987. Citado na página 18.
- NELSEN, R. B. **An Introduction to Copulas**. [S.l.]: Springer-Verlag, 2006. Citado na página 25.
- OAKES, D. Bivariate survival models induced by frailties. **Journal of the American Statistical Association**, Taylor & Francis, v. 84, n. 406, p. 487–493, 1989. Citado na página 34.
- PEREIRA, C. A. d. B.; STERN, J. M.; WECHSLER, S. *et al.* Can a significance test be genuinely bayesian? **Bayesian Analysis**, International Society for Bayesian Analysis, v. 3, n. 1, p. 15–36, 2008. Citado nas páginas 20 e 33.
- R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2017. Disponível em: <<https://www.R-project.org/>>. Citado na página 39.
- ROMEO, J. S.; MEYER, R.; GALLARDO, D. I. Bayesian bivariate survival analysis using the power variance function copula. **Lifetime data analysis**, Springer, v. 24, n. 2, p. 355–383, 2018. Citado na página 35.
- SKLAR, A. Fonctions de répartition à  $n$  dimensions et leurs marges. **Publications de l'Institut de Statistique de l'Université de Paris**, v. 8, n. 229-231, 1959. Citado nas páginas 18 e 53.
- SPIEGELHALTER, D. J.; BEST, N. G.; CARLIN, B. P.; LINDE, A. V. D. Bayesian measures of model complexity and fit. **Journal of the royal statistical society: Series b (statistical methodology)**, Wiley Online Library, v. 64, n. 4, p. 583–639, 2002. Citado nas páginas 20 e 32.
- TURNBULL, B. W.; JR, B. W. B.; HU, M. Survivorship analysis of heart transplant data. **Journal of the American Statistical Association**, Taylor & Francis Group, v. 69, n. 345, p. 74–80, 1974. Citado na página 18.

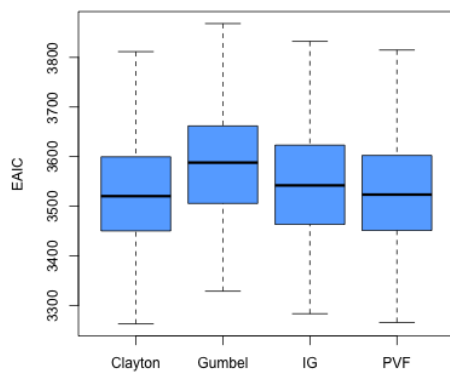
---

## COMPORTAMENTO DOS CRITÉRIOS DE SELEÇÃO NO ESTUDO DE SIMULAÇÃO

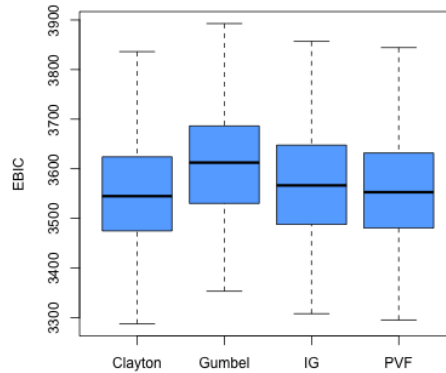
---

---

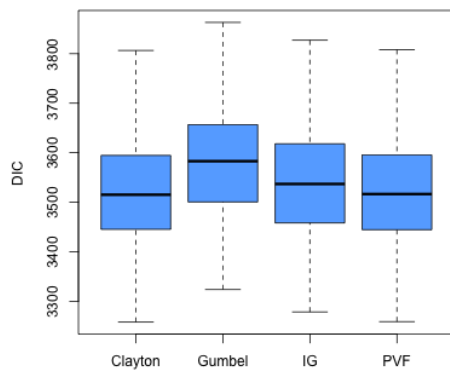
Neste apêndice é apresentado o comportamento dos critérios de seleção de modelos nas réplicas do estudo de simulação com tamanho de amostra igual a 1000, com exceção do cenário de marginais i.d., 75% de censura em cada marginal e modelo verdadeiro Gumbel, apresentado na Figura 7.



(a) EAIC



(b) EBIC



(c) DIC

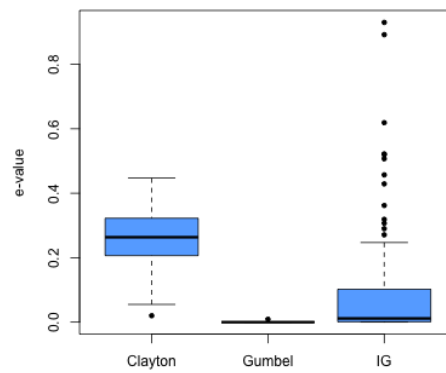
(d) *e-value*

Figura 8 – EAIC, EBIC, DIC e *e-value* obtidos nas 100 réplicas do estudo de simulação para o caso de marginais não i.d., número de amostral igual a 1000, 75% de censura em cada marginal e modelo verdadeiro Clayton.

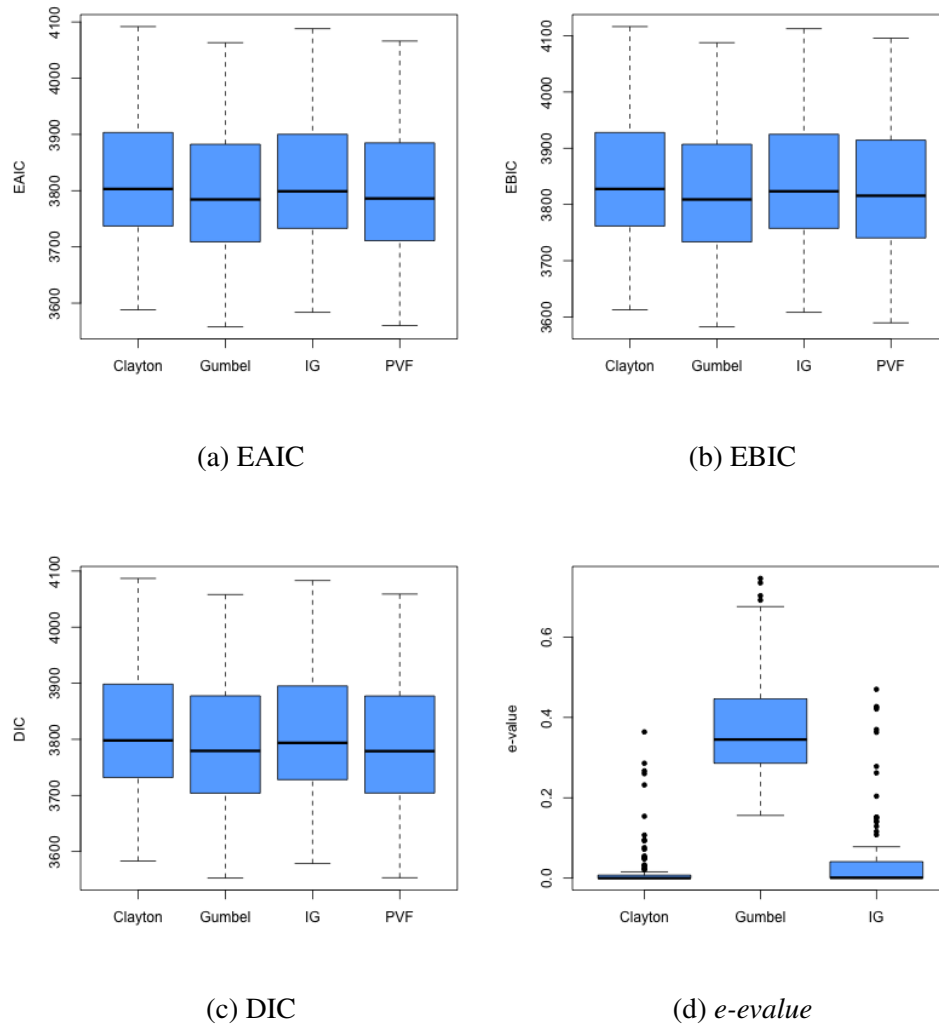


Figura 9 – EAIC, EBIC, DIC e *e-value* obtidos nas 100 réplicas do estudo de simulação para o caso de marginais não i.d., número de amostral igual a 1000, 75% de censura em cada marginal e modelo verdadeiro Gumbel.

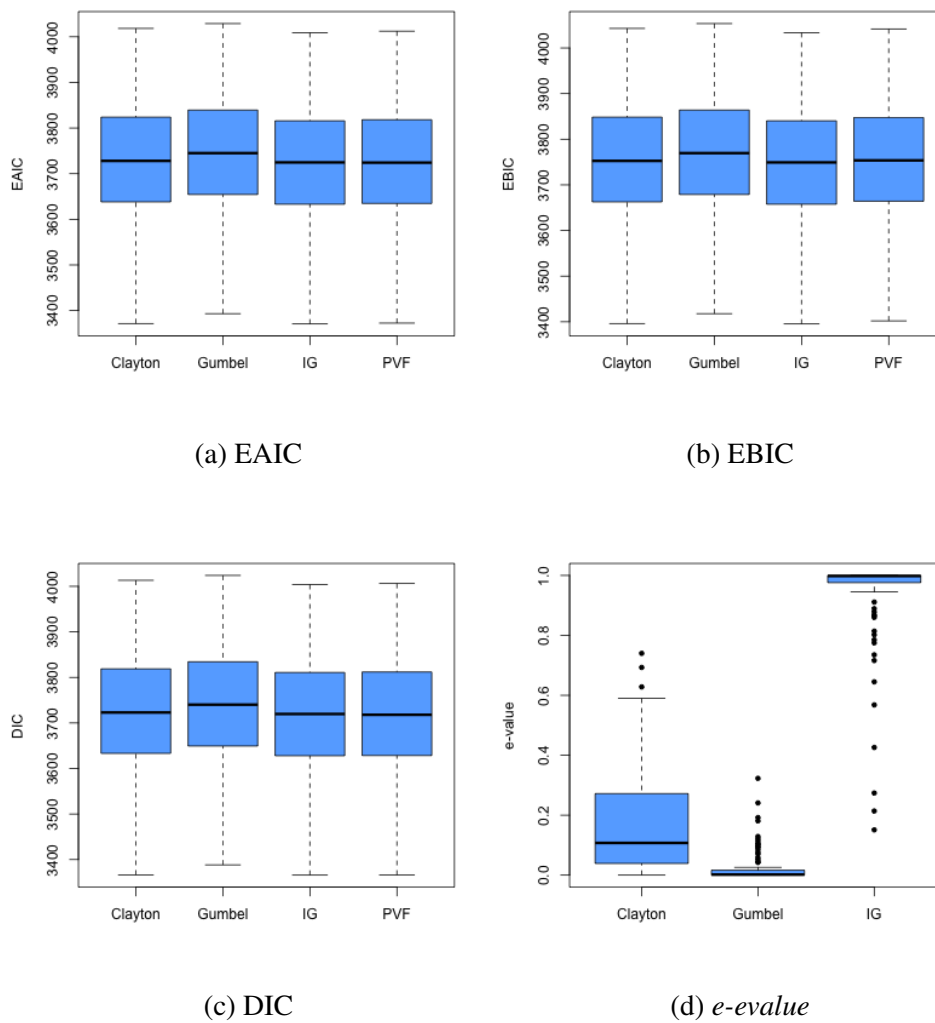


Figura 10 – EAIC, EBIC, DIC e  $e$ -value obtidos nas 100 réplicas do estudo de simulação para o caso de marginais não i.d., número de amostral igual a 1000, 75% de censura em cada marginal e modelo verdadeiro IG.

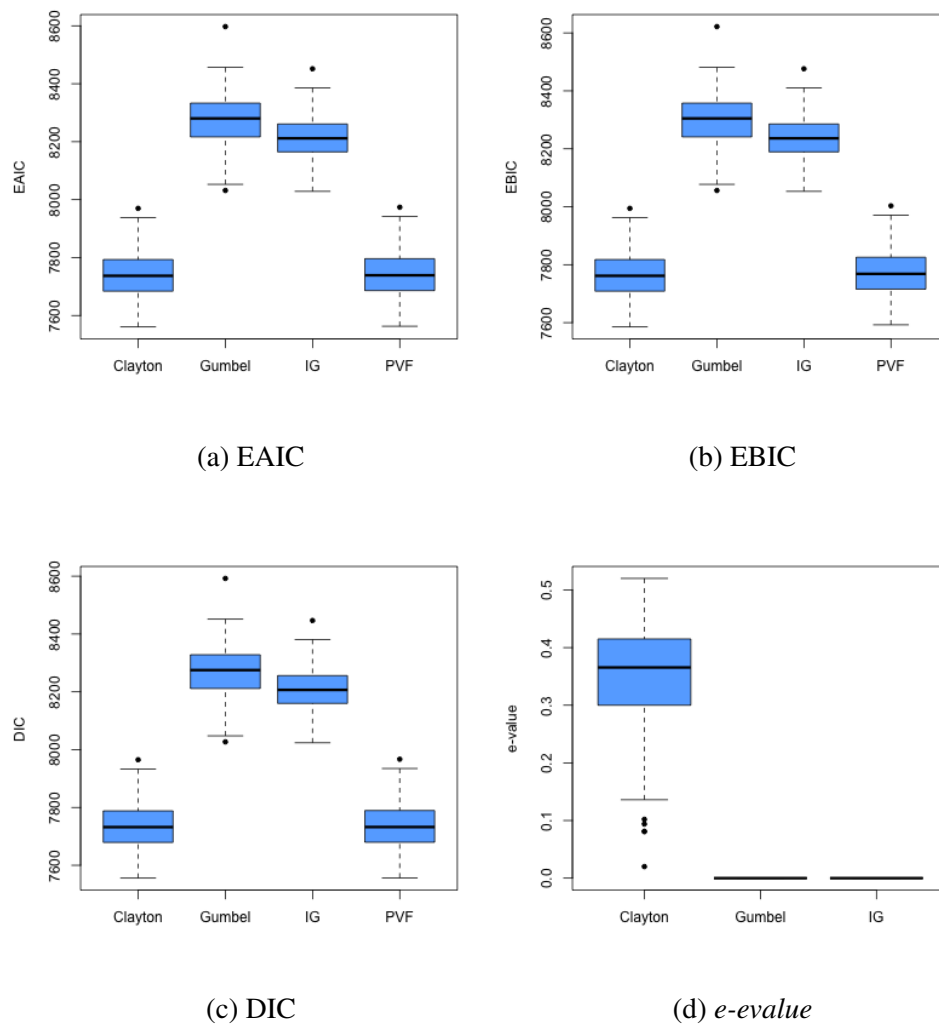
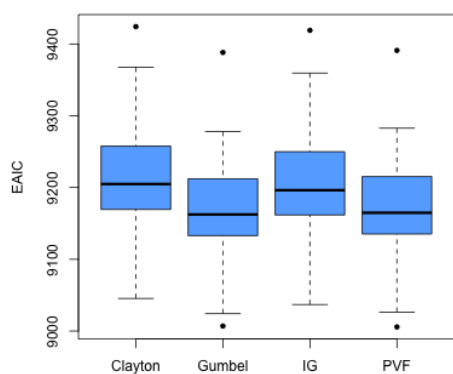
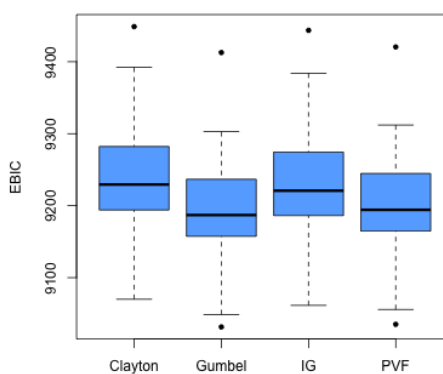


Figura 11 – EAIC, EBIC, DIC e  $e$ -value obtidos nas 100 réplicas do estudo de simulação para o caso de marginais não i.d., número de amostral igual a 1000, 20% de censura em cada marginal e modelo verdadeiro Clayton.

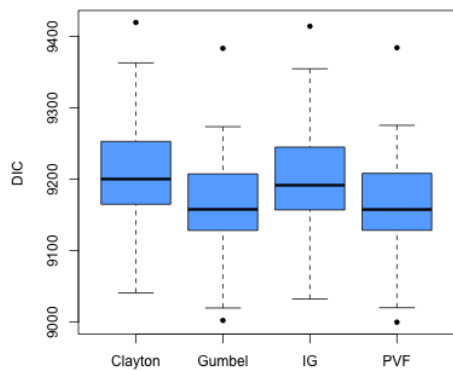




(a) EAIC



(b) EBIC



(c) DIC

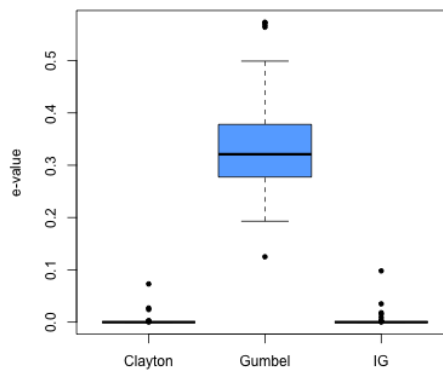
(d) *e-value*

Figura 12 – AIC, EBIC, DIC e *e-value* obtidos nas 100 réplicas do estudo de simulação para o caso de marginais não i.d., número de amostral igual a 1000, 20% de censura em cada marginal e modelo verdadeiro Gumbel.

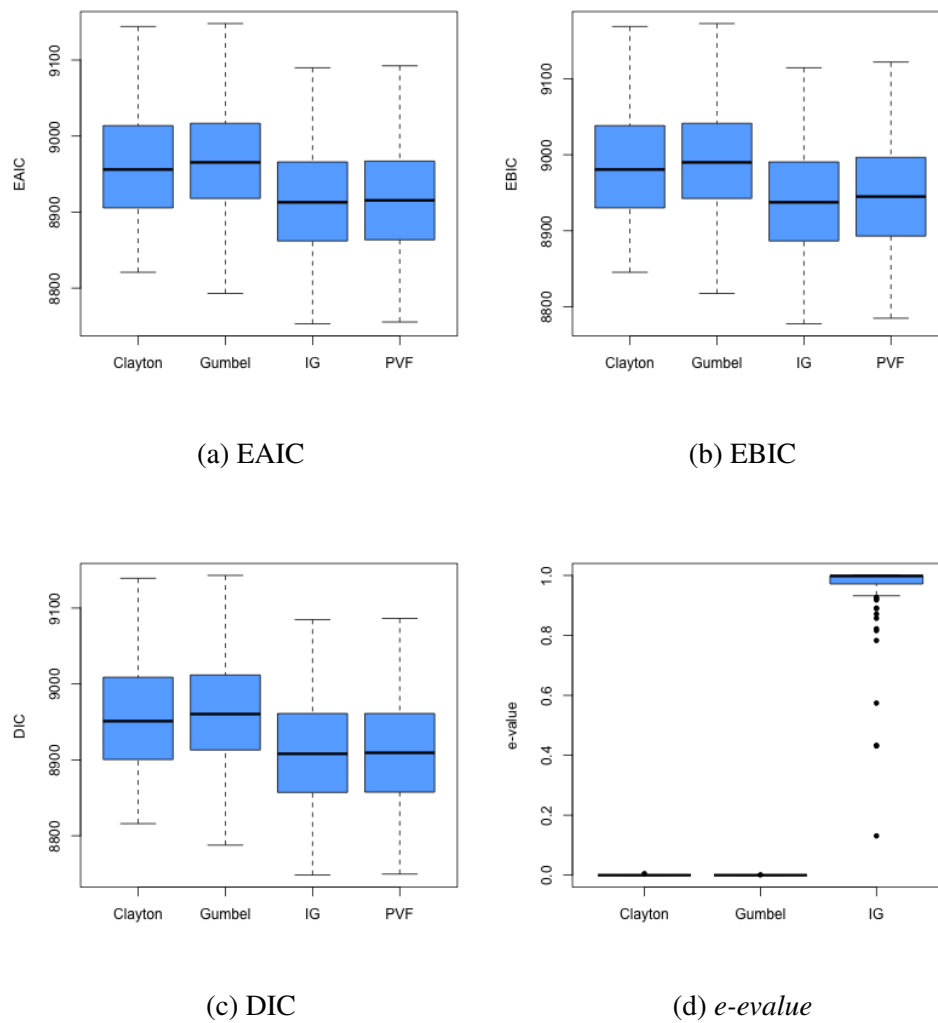
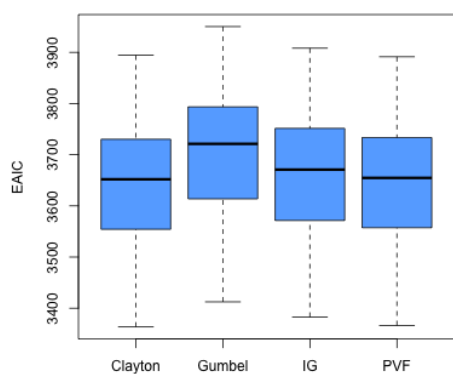
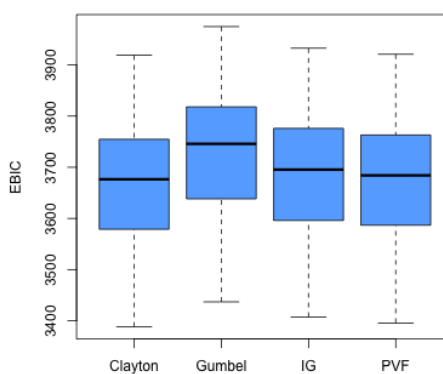


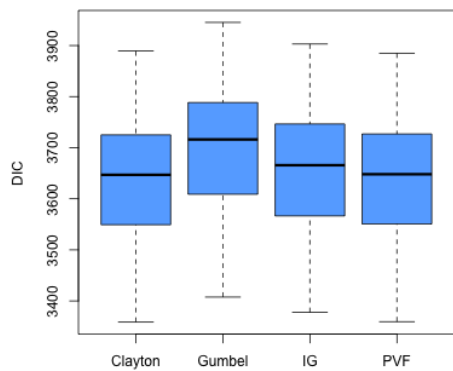
Figura 13 – AIC, EBIC, DIC e *e-value* obtidos nas 100 réplicas do estudo de simulação para o caso de marginais não i.d., número de amostral igual a 1000, 20% de censura em cada marginal e modelo verdadeiro IG.



(a) EAIC



(b) EBIC



(c) DIC

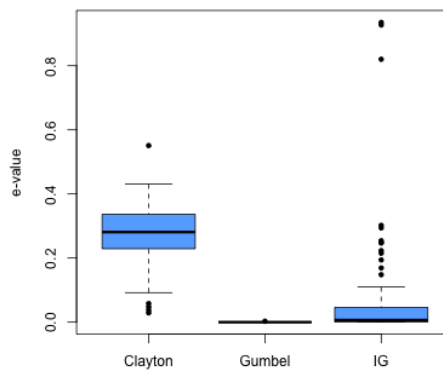
(d) *e-value*

Figura 14 – EAIC, EBIC, DIC e *e-value* obtidos nas 100 réplicas do estudo de simulação para o caso de marginais i.d., número de amostral igual a 1000, 75% de censura em cada marginal e modelo verdadeiro Clayton.

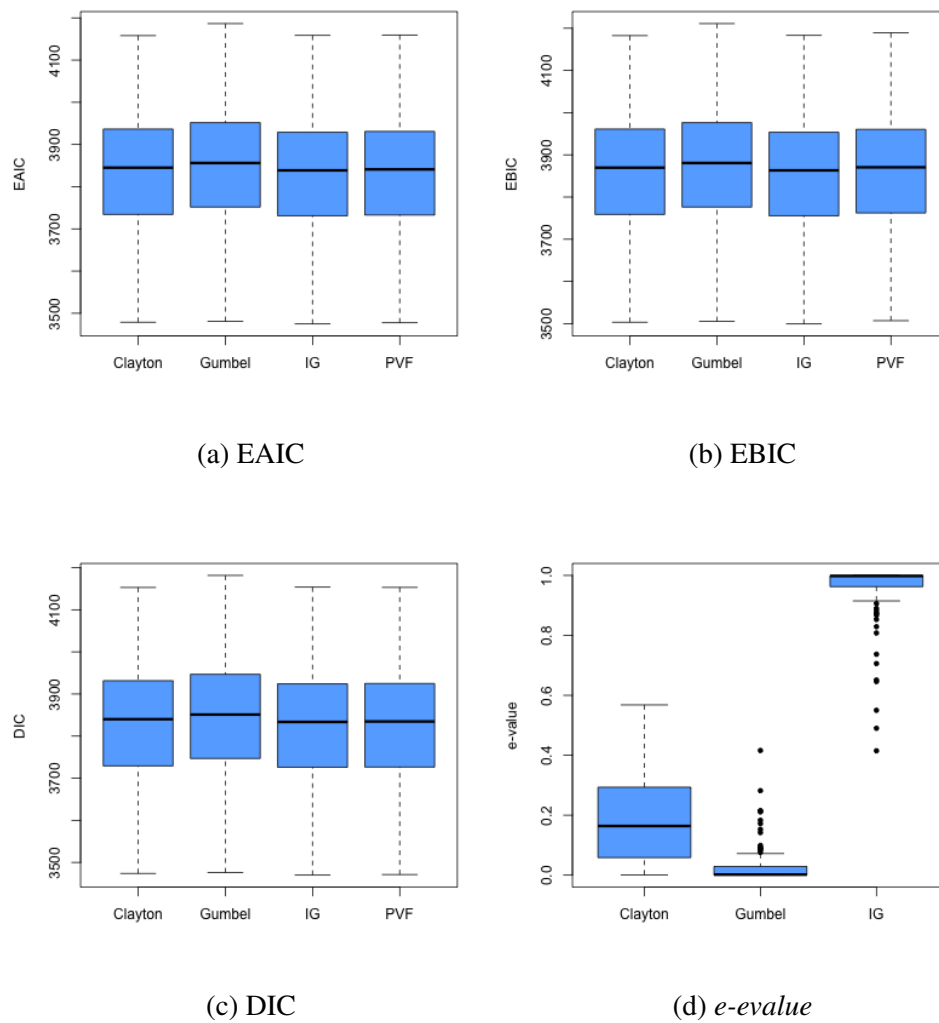


Figura 15 – EAIC, EBIC, DIC e *e-value* obtidos nas 100 réplicas do estudo de simulação para o caso de marginais i.d., número de amostral igual a 1000, 75% de censura em cada marginal e modelo verdadeiro IG.

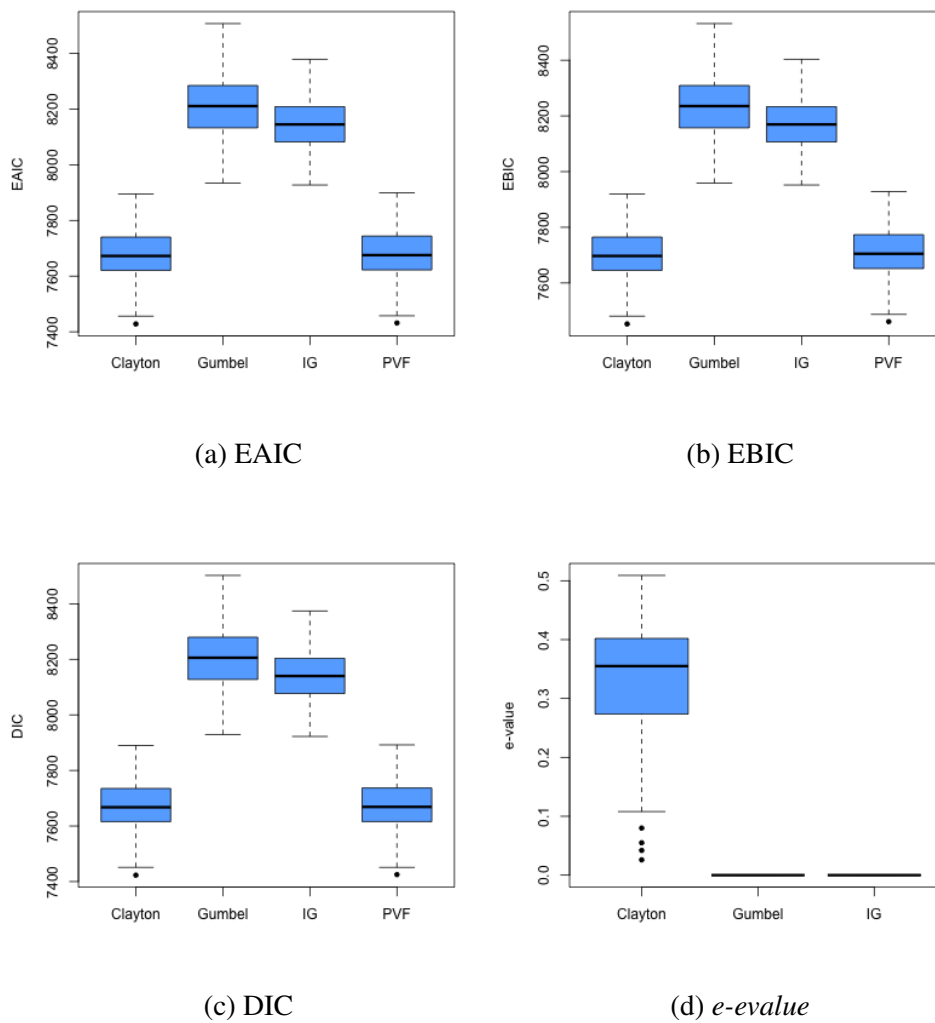


Figura 16 – EAIC, EBIC, DIC e *e-value* obtidos nas 100 réplicas do estudo de simulação para o caso de marginais i.d., número de amostral igual a 1000, 20% de censura em cada marginal e modelo verdadeiro Clayton.

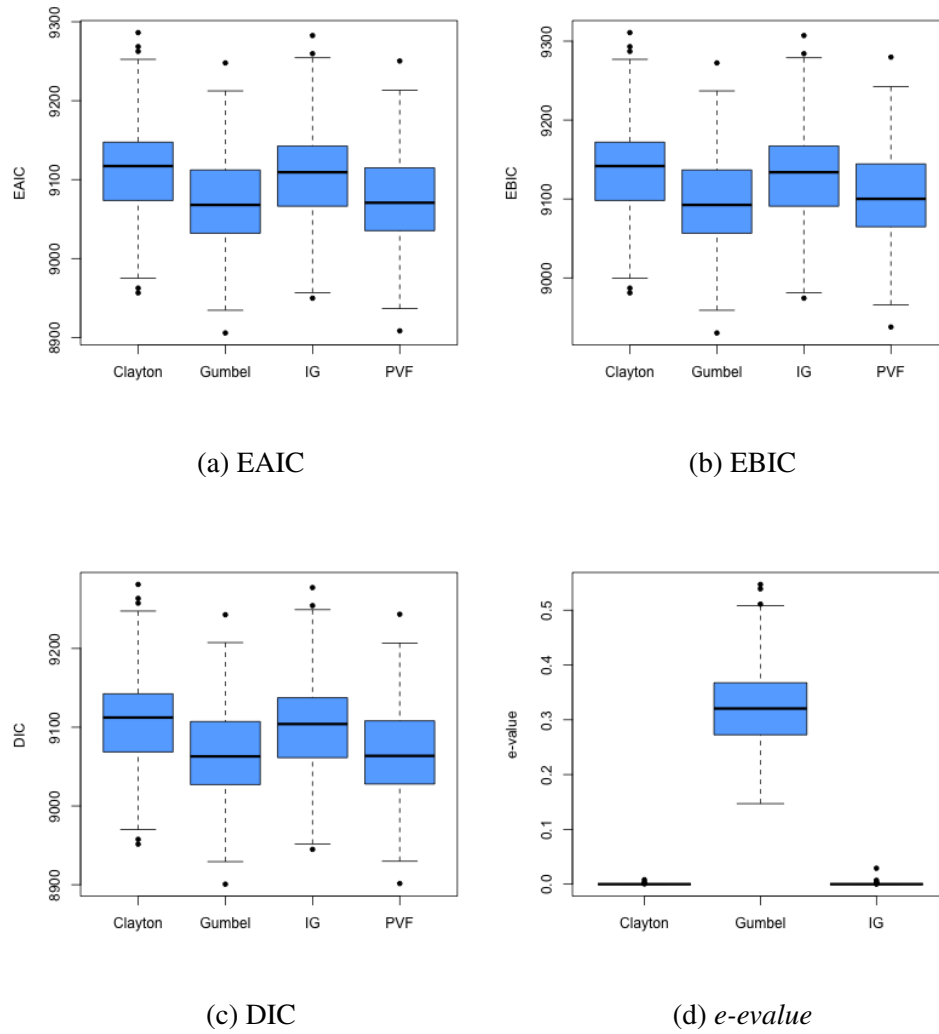


Figura 17 – EAIC, EBIC, DIC e *e-value* obtidos nas 100 réplicas do estudo de simulação para o caso de marginais i.d., número de amostral igual a 1000, 20% de censura em cada marginal e modelo verdadeiro Gumbel.

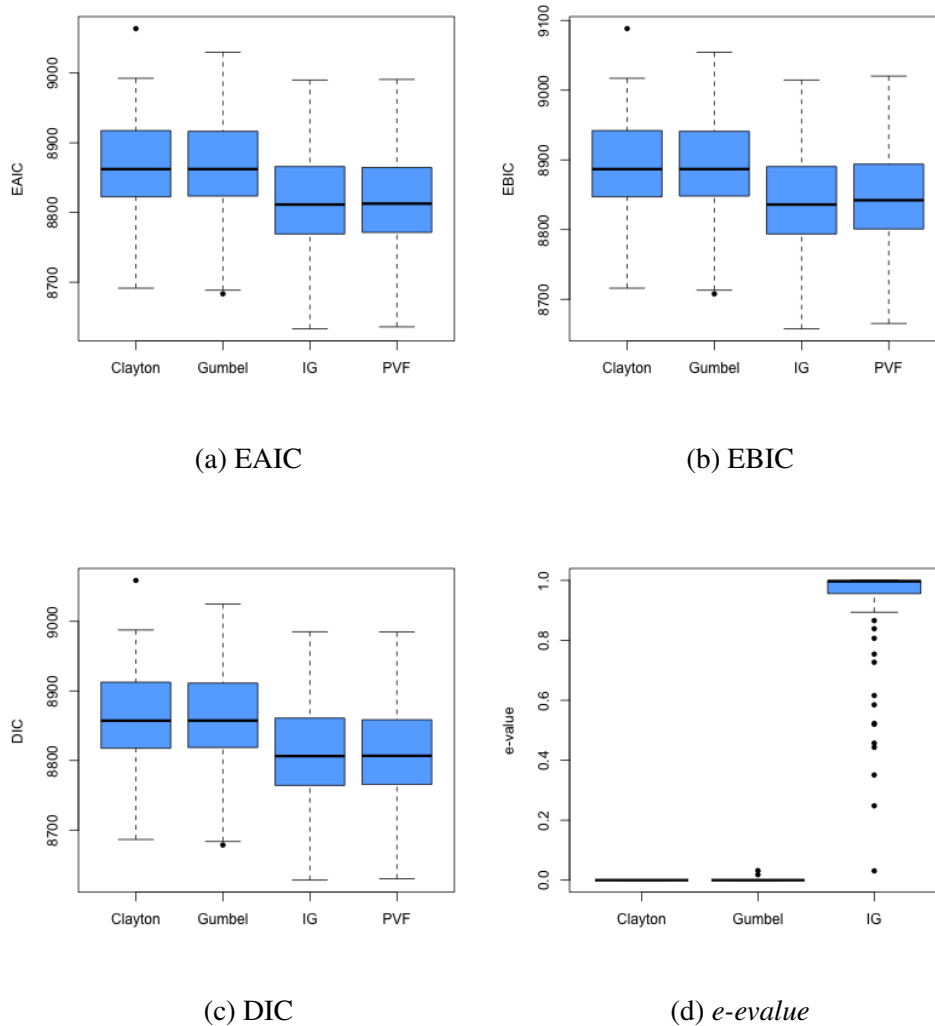


Figura 18 – EEAIC, EBIC, DIC e *e-value* obtidos nas 100 réplicas do estudo de simulação para o caso de marginais i.d., número de amostral igual a 1000, 20% de censura em cada marginal e modelo verdadeiro IG.