



# VII Workshop em Análise de Sobrevivência e Aplicações

Em Homenagem à Professora Vera Tomazella

23 a 25 de Outubro de 2024

Brasília, DF - Brasil



---

## Boas-vindas!

Prezados e Prezadas, bem-vindos!

É com grande alegria que os recebemos na cidade de Brasília para o VII Workshop em Análise de Sobrevida e Aplicações (WASA). Este evento é uma das reuniões científicas da comunidade estatística brasileira, e esta edição está sendo organizada pelo Departamento de Estatística da Universidade de Brasília (EST-UnB) em parceria com a Região Brasileira da Sociedade Internacional de Biometria (RBras). Foram meses de trabalho, reuniões, planejamento e investimentos por parte do EST-UnB, do Programa de Pós-Graduação em Estatística da UnB (PPGEST-UnB), da RBras, da Comissão Organizadora Local e da Comissão Científica com o objetivo de tornar o VII WASA um excelente evento.

A programação científica do VII WASA é composta por Conferências, Miniconferências, Sessão Temática, Tutoriais, Sessões de Comunicações Orais e Sessões de Pôsteres. Em especial, nesta edição, a Professora Vera Tomazella será homenageada. Essa homenagem é um reconhecimento público por sua notável contribuição à Estatística, especialmente à área de Análise de Sobrevida.

Agradecemos imensamente aos patrocinadores e colaboradores que ajudaram a tornar o VII WASA possível. Em especial agradecemos a você! Pela sua presença e interesse em desenvolver e aprofundar seus conhecimentos em Análise de Sobrevida, compartilhando suas experiências com os especialistas e se inteirando de novos conhecimentos na área.

Sejam bem-vindos ao VII Workshop em Análise de Sobrevida e Aplicações e tenham um excelente evento!

Profa. Dra. Juliana Betini Fachini Gomes  
Coordenadora da Comissão Organizadora Local do VII WASA



---

## Conteúdo

|          |  |          |
|----------|--|----------|
| <b>1</b> | <b>Comissões</b>                                     | <b>3</b> |
| <b>2</b> | <b>Informações Importantes</b>                       | <b>5</b> |
| <b>3</b> | <b>Grade de Programação</b>                          | <b>6</b> |
| <b>4</b> | <b>Programação Detalhada por Dia</b>                 | <b>7</b> |
| <b>5</b> | <b>Programação Científica: Atividades Detalhadas</b> | <b>8</b> |
| 5.1      | Conferências . . . . .                               | 8        |
| 5.2      | Miniconferências: Jovens Doutores . . . . .          | 10       |
| 5.3      | Sessão Temática: Mulheres na Ciência . . . . .       | 12       |
| 5.4      | Tutoriais . . . . .                                  | 13       |
| 5.5      | Sessões de Comunicações Orais . . . . .              | 14       |
| 5.6      | Sessões de Apresentação de Pôster . . . . .          | 18       |

---

# 1 Comissões

## **Comissão Organizadora Local**

- Profa. Dra. Juliana Betini Fachini Gomes (UnB) - Presidente
- Profa. Dra. Agatha Sacramento Rodrigues (UFES)
- Prof. Dr. Alan Ricardo da Silva (UnB)
- Profa. Dra. Cira Etheowalda Guevara Otiniano (UnB)
- Prof. Dr. Eduardo Yoshio Nakano (UnB)
- Prof. Dr. Frederico Machado Almeida (UnB)
- Profa. Dra. Giovana Oliveira Silva (UFBA)
- Prof. Dr. Helton Saulo Bezerra dos Santos (UnB)
- Profa. Dra. Terezinha Késsia de Assis Ribeiro (UnB)
- Profa. Dra. Thais Carvalho Valadares Rodrigues (UnB)

## **Comissão Científica**

- Profa. Dra. Vera Tomazella (UFSCar-UFPB) – Presidente
- Prof. Dr. Antônio Eduardo Gomes (UnB)
- Prof. Dr. Edwin Moises Marco (ESALQ-USP)
- Prof. Dr. Enrico Colosimo (UFMG)
- Profa. Dra. Gleici Perdoná (FMRP-USP)
- Prof. Dr. Jeremias Leão (UFAM)
- Prof. Dr. Paulo Henrique Ferreira (UFBA)
- Prof. Dr. Pedro Luiz Ramos (PUC-Chile)
- Profa. Dra. Rosemeire Leovigildo Fiaccone (UFBA)
- Prof. Dr. Victor Hugo Lachos (University of Connecticut)

## **Secretarias**

- Ednilson Lopes Carvalho (UnB)
- Karen Luíza Silva de Sousa (UnB)
- Lucas Fernandes de Albuquerque Lira (UnB)



- 
- Luis Filipe Silva de Andrade (UnB)
  - Tathyanna Martins Cordeiro (UnB)
  - Vivian Claudia Pereira Oliveira (UnB)



---

## 2 Informações Importantes

- **Certificados:** Os certificados de participação e apresentação de trabalho serão disponibilizados online no site do evento: <https://viiwasa.softaliza.com.br/>.
- **Recibo:** O recibo de pagamento está disponibilizado online no site do evento: <https://viiwasa.softaliza.com.br/>.
- **Pôster:** Os autores apresentadores deverão fixar seus trabalhos nos painéis disponíveis no Espaço Chiarini que fica localizado no 1<sup>o</sup> andar do Prédio CIC/EST até 30 minutos antes do horário de sua sessão. É obrigatória a presença do(a) autor(a) apresentador(a) durante toda a sessão de pôster para validar a apresentação do trabalho. Trabalhos com pôster fixado mas sem autor(a) apresentador(a) serão desconsiderados da certificação.
- **Festa de Abertura:** A festa de abertura do VII WASA será realizada no Salão Pérgola do Clube AABB, localizado na Asa Sul Trecho 2 Conjunto 17 20, St. de Clubes Esportivos Sul, Brasília - DF, no dia 23/10 a partir das 19h30min. Para acesso a festa será necessário apresentação do convite da festa que é recebido junto com o material do evento.

### 3 Grade de Programação

Tabela 1: Grade de Programação do VII Workshop em Análise de Sobrevivência e Aplicações.

| Início                           | Final | Atividade                             |
|----------------------------------|-------|---------------------------------------|
| <b>Quarta-feira - 23/10/2024</b> |       |                                       |
| 08:00                            | 08:20 | Inscrições e entrega de material      |
| 08:20                            | 08:30 | Abertura                              |
| 08:30                            | 09:00 | Homenagem à Professora Vera Tomazella |
| 09:00                            | 10:00 | Conferência de Abertura (C1)          |
| 10:00                            | 10:30 | Café                                  |
| 10:30                            | 11:30 | Conferência 2 (C2)                    |
| 11:30                            | 12:00 | Miniconferência 1 (MiniConf1)         |
| 12:00                            | 14:00 | Almoço                                |
| 14:00                            | 15:00 | Conferência 3 (C3)                    |
| 15:00                            | 15:30 | Miniconferência 2 (MiniConf2)         |
| 15:30                            | 16:30 | Sessão Oral 1 (SO1)                   |
| 16:30                            | 17:30 | Sessão de Pôsters 1 e Café            |
| 19:30                            | 23:30 | Festa de Abertura                     |
| <b>Quinta-feira - 24/10/2024</b> |       |                                       |
| 08:00                            | 10:00 | Tutorial 1                            |
| 10:00                            | 10:30 | Café                                  |
| 10:30                            | 11:30 | Conferência 4 (C4)                    |
| 11:30                            | 12:00 | Miniconferência 3 (MiniConf3)         |
| 12:00                            | 14:00 | Almoço                                |
| 14:00                            | 15:00 | Sessão Oral 2 (SO2)                   |
| 15:00                            | 16:00 | Conferência 5 (C5)                    |
| 16:00                            | 17:00 | Sessão de Pôsters 2 e Café            |
| 17:00                            | 18:30 | Sessão Temática “Mulheres na Ciência” |
| <b>Sexta-feira - 25/10/2024</b>  |       |                                       |
| 08:00                            | 10:00 | Tutorial 2                            |
| 10:00                            | 10:30 | Café                                  |
| 10:30                            | 11:30 | Sessão Oral 3 (SO3)                   |
| 11:30                            | 12:30 | Conferência de Encerramento (C6)      |
| 12:30                            | 13:00 | Cerimônia de Encerramento             |

## 4 Programação Detalhada por Dia

Tabela 2: Programação Detalhada - Quarta-feira, 23/10/2024.

| Horário       | Evento                                |
|---------------|---------------------------------------|
| 08:00 - 08:20 | Inscrições e entrega de material      |
| 08:20 - 08:30 | Abertura                              |
| 08:30 - 09:00 | Homenagem à Professora Vera Tomazella |
| 09:00 - 10:00 | C1 - Janet van Nierkerk (KAUST)       |
| 10:00 - 10:30 | Café                                  |
| 10:30 - 11:30 | C2 - Francisco Louzada (ICMC-USP)     |
| 11:30 - 12:00 | MiniConf1 - Márcio Rodrigues (UFG)    |
| 12:00 - 14:00 | Almoço                                |
| 14:00 - 15:00 | C3 - Francisco Medeiros (UFRN)        |
| 15:00 - 15:30 | MiniConf2 - Éder Silva de Brito (IFG) |
| 15:30 - 16:30 | Sessão Oral 1                         |
| 16:30 - 17:30 | Sessão de Pôsters 1 e Café            |
| 19:30 - 23:30 | Festa de Abertura                     |

Tabela 3: Programação Detalhada - Quinta-feira, 24/10/2024.

| Horário       | Evento   |
|---------------|--|
| 08:00 - 10:00 | Tutorial 1 - Anderson Ara (UFPR) e Marcelo R. P. Ferreira (UFPB) |
| 10:00 - 10:30 | Café   |
| 10:30 - 11:30 | C4 - Fábio Nogueira Demarqui (UFMG)                              |
| 11:30 - 12:00 | MiniConf3 - Eduardo Ramos (ICMC-USP)                             |
| 12:00 - 14:00 | Almoço   |
| 14:00 - 15:00 | Sessão Oral 2  |
| 15:00 - 16:00 | C5 - Vinícius D. Mayrink (UFMG)                                  |
| 16:00 - 17:00 | Sessão de Pôsters 2 e Café                                       |
| 17:00 - 18:30 | Sessão Temática “Mulheres na Ciência”                            |

Tabela 4: Programação Detalhada - Sexta-feira, 25/10/2024.

| Horário       | Evento  |
|---------------|---|
| 08:00 - 10:00 | Tutorial 2 - Janet van Nierkerk (KAUST) e Denis Rustand (KAUST) |
| 10:00 - 10:30 | Café  |
| 10:30 - 11:30 | Sessão Oral 3   |
| 11:30 - 12:30 | C6 - Gisela Tunes (IME-USP)                                     |
| 12:30 - 13:00 | Cerimônia de Encerramento                                       |

## 5 Programação Científica: Atividades Detalhadas

### 5.1 Conferências

#### Conferência 1 (C1) – Conferência de Abertura – 23/10 (09:00 – 10:00)

**Título:** Complex survival analysis with INLA

**Conferencista:** Janet van Nierkerk – King Abdullah University of Science and Technology (KAUST)

**Resumo:** INLA is an approximate Bayesian inference method that can fit many different statistical models including survival models. In this talk I will present two projects in which INLA was used for survival analysis. The first shows how to fit a spatial survival model based on areal data or point-indexed data. The second project employs a quantile joint model to study adherence in the midst of outliers.

#### Conferência 2 (C2) – 23/10 (10:30 – 11:30)

**Título:** Bridging the Gap: Reliability Innovations in Brazil Connecting Academia and Industry

**Conferencista:** Francisco Louzada – Universidade de São Paulo (ICMC-USP)

**Resumo:** In today's world, our reliance on mechanical and electronic devices is growing exponentially. From intelligent sensors and artificial intelligence to agricultural, financial, and medical robots, these technologies have become integral to our daily lives. However, despite their efficiency, these devices are not immune to failure. As a result, statistical reliability analysis has emerged as a crucial tool in the innovation process, ensuring the dependability and longevity of these technologies. This talk presents several reliability innovation projects that showcase the efforts being made in Brazil to bridge the gap between academia and various industrial sectors, including oil and gas, agriculture, medicine, and finance. By establishing a connection between theoretical research and practical applications, these projects aim to enhance the reliability and performance of critical equipment and systems. The focus of this talk is on the development of reliability models for a range of applications. These include oil well construction equipment, which plays a vital role in the energy sector; bucket tracking equipment, essential for efficient material handling; and agricultural machinery, crucial for maintaining food production. Additionally, the talk explores communication modeling for mobile phones, a technology that has

---

revolutionized personal and business communication. Through these projects, we demonstrate how statistical reliability analysis can be leveraged to improve the design, manufacturing, and maintenance of complex systems. By collaborating with industry partners, academia can contribute to the development of more reliable and resilient technologies, ultimately benefiting society as a whole. The talk concludes by emphasizing the importance of continued collaboration between academia and industry in the field of reliability engineering. By working together, we can drive innovation, improve safety, and ensure the long-term performance of the devices and systems upon which we increasingly depend.

### **Conferência 3 (C3) – 23/10 (14:00 – 15:00)**

**Título:** Log-symmetric models with cure fraction with application to leprosy reactions data

**Conferencista:** Francisco Medeiros – Universidade Federal do Rio Grande do Norte (UFRN)

**Resumo:** In this paper, we propose a log-symmetric survival model with cure fraction, considering that the distributions of lifetimes for susceptible individuals belong to the log-symmetric class of distributions. This class has continuous, strictly positive, and asymmetric distributions, including the log-normal, log- $t$ -Student, Birnbaum-Saunders, log-logistic I, log-logistic II, log-normal-contaminated, log-exponential-power, and log-slash distributions. The log-symmetric class is quite flexible and allows for including bimodal distributions and outliers. This includes explanatory variables through the parameter associated with the cure fraction. We evaluate the performance of the proposed model through extensive simulation studies and consider a real data application to evaluate the effect of factors on the immunity to leprosy reactions in patients with Hansen's disease.

### **Conferência 4 (C4) – 24/10 (10:30 – 11:30)**

**Título:** Survival Data Simulation With the R Package rsurv

**Conferencista:** Fábio Nogueira Demarqui – Universidade Federal de Minas Gerais (UFMG)

**Resumo:** In this talk we introduce the R package rsurv, aimed for general survival data simulation purposes. The package, which is available on CRAN at <https://CRAN.R-project.org/package=rsurv>, is built under a new approach to simulate survival data that depends deeply on the use of dplyr verbs. The proposed package allows the simulation of survival data from a wide range of regression models, including accelerated failure time (AFT), proportional hazards (PH), proportional odds (PO), accelerated hazard (AH), Yang and Prentice (YP), and extended hazard (EH) models. The package rsurv also stands out by its ability to generate survival data from an unlimited number of baseline distributions provided that an implementation of the quantile function of the chosen baseline distribution is available in R. Another nice feature of the package rsurv lies in the fact that linear predictors are specified via a formula-based approach, facilitating the inclusion of categorical variables and interaction terms. The functions implemented in the package rsurv can also be employed to simulate survival data with more complex structures, such as survival data with different types of censoring mechanisms, survival data with cure fraction, survival data with random effects (frailties), multivariate survival data, and competing risks survival data.

---

**Conferência 5 (C5) – 24/10 (15:00 – 16:00)**

**Título:** Bayesian solution to the monotone likelihood in the standard mixture cure model

**Conferencista:** Vinícius D. Mayrink – Universidade Federal de Minas Gerais (UFMG)

**Resumo:** An advantage of the standard mixture cure model over a usual survival model is how it accounts for population heterogeneity. It allows a joint estimation for the distribution related to the susceptible and non-susceptible subjects. The estimation algorithm may provide coefficients when the likelihood cannot be maximized. This phenomenon is known as Monotone Likelihood (ML), common in survival and logistic regressions. The ML appears in situations with a small sample size, many censored times, and many binary or unbalanced covariates. Particularly, it occurs when all uncensored cases correspond to one level of a binary covariate. The existing frequentist solution is an adaptation of the Firth correction, originally proposed to reduce the bias of maximum likelihood estimates. It prevents estimates by penalizing the likelihood, with the penalty interpreted as the Bayesian Jeffreys prior. This study considers the penalized likelihood of the standard mixture cure model with different penalties (Bayesian priors). A Monte Carlo simulation study indicates good inference results, especially for balanced data sets. Finally, a real application involving melanoma data illustrates the approach.

**Conferência 6 (C6) - Conferência de Encerramento – 25/10 (11:30 – 12:30)**

**Título:** Semiparametric joint modeling of competing risks survival and longitudinal data

**Conferencista:** Gisela Tunes – Universidade de São Paulo (IME-USP)

**Resumo:** In many observational studies, patients are followed until death or another outcome of interest is observed and several clinical variables are also measured during follow up. When longitudinal data is available as well as survival, it may be of interest to understand the effect of prognostic variables in both longitudinal and survival outcomes. In this work, a joint model for competing risks survival data and longitudinal data is discussed. A semiparametric approach is presented, along with simulation results and real data application.

## 5.2 Miniconferências: Jovens Doutores

**Miniconferência 1 (MiniConf1) – 23/10 (11:30 – 12:00)**

**Título:** Cause-Specific Regression Model for correlated data in the presence of competing risks and interval censoring

**Conferencista:** Márcio Rodrigues – Universidade Federal de Goiás (UFG)

**Resumo:** Interval censoring might happen in survival analysis when it is only known that the survival response belongs to a time interval. Usually in survival analysis, a single cause for the occurrence of the event is considered, however, some studies may be interested in the occurrence of more than one event, called competing risks. The Cox proportional hazards model is model might be used to the cause-specific failure rate function. We propose a cause-specific regression model for correlated data and interval censoring. The cause-specific regression model

---

is an extension of the traditional Cox regression model for each event type, in which failures of competing events are treated as censored observations. The present study was motivated by a longitudinal study conducted at the Dental Trauma Clinic at the Federal University of Minas Gerais, Brazil aiming to evaluate luxation of permanent teeth for their pulp prognosis as well as its prognostic factors. In this paper, the proposed methodology considers a GEE-type model using an independent work matrix to accommodate the presence of clusters. A Taylor series is used to approximate the log baseline hazard function in Cox proportional hazards regression. With this formulation, the likelihood ratio test can be used to select an appropriate order for the Taylor series approximation and maximum likelihood is used to estimate model parameters and provide statistical inference. The model performance was evaluated using a simulated data set and showing that the proposed methodology has good small sample properties. The proposed methodology is applied to a real dental trauma dataset.

### **Miniconferência 2 (MiniConf2) – 23/10 (15:00 – 15:30)**

**Título:** Unobserved heterogeneity for multiple repairable systems subject to competitive risks under imperfect repair

**Conferencista:** Éder Silva de Brito – Instituto Federal de Goiás (IFG)

**Resumo:** In this study, we introduce models for the failure times of repairable systems, which experience failures due to different and independent causes while being influenced by unobservable effects acting on their failure processes. Furthermore, we assume that imperfect repairs are performed after each failure event to restore the system to its standard operational condition. In this sense, the proposed model combines the concept of imperfect repairs with independent competing risks linked by unobserved heterogeneity, which is shared by the failure times of each system. In addition to presenting these novel models, the objective of this work is to develop classic inferential methodologies for estimating maximum likelihood parameters and obtaining reliability prediction functions for each system based on its failure history. Real world applications of these models are conducted, demonstrating their capacity to identify the effect of repairs related to the causes of failure and the presence or absence of unobserved heterogeneity. Therefore, this research addresses a relevant issue in the field of reliability because, in addition to presenting models that extend and generalize ones in the literature, it has potential practical applicability in diverse scenarios involving repairable systems.

### **Miniconferência 3 (MiniConf3) – 24/10 (11:30 – 12:00)**

**Título:** Propriedades da distribuição a posteriori para dados censurados usando a distribuição Gama

**Conferencista:** Eduardo Ramos – Universidade de São Paulo (ICMC-USP)

**Resumo:** Nesta apresentação, explorarei as propriedades da distribuição a posteriori da distribuição Gama, com um foco particular em dados censurados. Abordaremos as condições necessárias e suficientes para que priors impróprias resultem em posteriors próprias e examinaremos a finitude dos momentos a posteriori. Discutiremos os desafios da censura de dados e a aplicação de diversos priors objetivos. Apresentarei um novo estimador para dados censurados,

---

que melhora a eficiência do algoritmo de Monte Carlo via Cadeias de Markov (MCMC). Por meio de um estudo de simulação, avaliaremos o desempenho de estimadores Bayesianos com diferentes priors. Aplicaremos nossa metodologia a um conjunto de dados do Atlas do Genoma do Câncer, com foco em adenocarcinoma de pulmão em pacientes acima de 70 anos, oferecendo insights valiosos sobre a progressão da doença e padrões de mortalidade.

### 5.3 Sessão Temática: Mulheres na Ciência

#### Palestra 1 – 24/10 (17:00 – 17:30)

**Título:** Modelagem de classes latentes com desfechos distais em sobrevivência

**Palestrante:** Leila Denise Alves Ferreira Amorim – Universidade Federal da Bahia (UFBA)

**Resumo:** Um dos desafios na análise de dados é a necessidade de incorporação de múltiplos componentes interrelacionados em modelos estatísticos, de modo a permitir a captura de efeitos que apresentam natureza multifacetada como, por exemplo, hábitos e comportamentos de saúde que podem ser mensurados através de um conjunto de indicadores. A análise de classes latentes (LCA) é um caso especial de modelos de mistura finita, que têm sido utilizada para identificar e tipificar subgrupos (classes) latentes baseados em padrões de resposta observados. Extensões de LCA incluem a presença de covariáveis, que são utilizadas para entender que características podem prever o pertencimento a uma classe latente, e a modelagem de respostas distais, quando os preditores são latentes e a variável resposta é observada. Neste trabalho discutimos desenvolvimentos relacionados à modelagem com variáveis latentes categóricas e desfechos distais definidos pelo tempo até o evento para dados censurados. Em particular, as análises de interesse objetivam descrever o efeito de uma variável latente categórica no tempo até ocorrência do evento, em que o modelo estrutural é definido pelo modelo de Cox. Adicionalmente, propomos um procedimento de estimação bayesiano nos modelos com respostas distais em análise de sobrevivência, no qual a estimação é feita conjuntamente para os parâmetros da LCA e do modelo de Cox com abordagens em 1 e 3-passos. A metodologia proposta é ilustrada com análise de dados do projeto PrEP1519, que é um estudo de coorte multicêntrico que objetivou avaliar a adesão, segurança e viabilidade da profilaxia pré-exposição (PrEP) entre adolescentes MSM (homens que fazem sexo com homens) e mulheres transgêneros (TGW) no Brasil. Nossas análises, em particular, objetivam avaliar o efeito de padrões de comportamentos de risco no tempo até a descontinuação do uso da PrEP. Estudos de simulação são conduzidos para avaliação da metodologia proposta.

#### Palestra 2 – 24/10 (17:30 – 18:00)

**Título:** Integração do Modelo de Sobrevivência com fração de cura com o Algoritmo de Suporte Vetorial

**Palestrante:** Daiane de Souza Santos – Universidade de São Paulo (ICMC-USP)

**Resumo:** Os modelos de sobrevivência com fração de cura são extensões dos modelos de sobrevivência padrão para analisar dados de sobrevivência com uma fração de indivíduos curados ou não suscetíveis ao evento de interesse, frequentemente provenientes de estudos clínicos. Exis-

tem várias estratégias propostas na literatura para modelar a parte de latência desses modelos. No entanto, não há muitas metodologias propostas para investigar os efeitos de covariáveis na parte de incidência. Nesse sentido, a maioria dos estudos existentes assume que a fronteira que separa os sujeitos curados e não curados em relação às covariáveis é linear. Como tal, eles só podem capturar efeitos simples das covariáveis na probabilidade de curado ou não curado. Neste trabalho, estudamos um modelo com fração de cura que utiliza o algoritmo de suporte vetorial (ASV) para modelar a parte de incidência. O novo modelo herda as características do ASV para fornecer um modelo flexível para avaliar os efeitos das covariáveis na incidência. Desenvolvemos um método de estimação para estimar o modelo de cura e realizamos um estudo de simulação para mostrar que o modelo proposto supera os modelos de cura existentes, especialmente na estimação da incidência.

### **Palestra 3 – 24/10 (18:00 – 18:30)**

**Título:** Análise de sobrevivência com fração de cura utilizando florestas aleatórias de sobrevivência

**Palestrante:** Gabriela Rodrigues – Universidade de São Paulo (ESALQ-USP)

**Resumo:** As técnicas de aprendizado de máquinas, no contexto de análise de sobrevivência, têm como um dos principais objetivos, fornecer previsões precisas das probabilidades de sobrevivência. Neste estudo, são analisados dados de câncer de próstata utilizando florestas aleatórias de sobrevivência (FAS). A FAS é um método estatístico não linear e não paramétrico, baseado em um conjunto de árvores de decisão. Com o ajuste de várias árvores, as estimativas do modelo são estabilizadas, otimizando a acurácia de predição. Adicionalmente, os efeitos não lineares ou interações de ordem superior para os preditores não precisam ser definidos previamente, como em regressões usuais. O conjunto de dados analisado consiste em pacientes com câncer de próstata submetidos à prostatectomia. A variável resposta é dada pelo tempo sem a detecção da doença após o tratamento. As observações não censuradas correspondem aos pacientes que tiveram a recorrência da doença e as observações censuradas correspondem aos pacientes que não observaram a recorrência do câncer no momento da coleta dos dados. A porcentagem de indivíduos considerados curados é de aproximadamente 83%. O desempenho preditivo da FAS é comparado com modelos de regressão paramétricos com fração de cura existentes na literatura.

## **5.4 Tutoriais**

### **Tutorial 1 – 24/10 (08:00 – 10:00)**

**Título:** Introdução ao Aprendizado de Máquina para Análise de Sobrevivência

**Ministrantes:** Anderson Ara – Universidade Federal do Paraná (UFPR) e Marcelo R. P. Ferreira – Universidade Federal da Paraíba (UFPB)

**Resumo:** Devido aos recentes avanços em aprendizado de máquina, surgem novas oportunidades em como estimar a estrutura de associação entre os tempos de sobrevida e covariáveis por meio de técnicas mais flexíveis. Este tutorial visa abordar os conceitos teóricos dos métodos de

---

aprendizado de máquina para análise de sobrevivência, em específico aos métodos de Random Survival Forest e Support Vector Censored Regression, bem como explorar exemplos reais utilizando a Linguagem R. Material confeccionado em trabalho conjunto com Agatha Rodrigues (UFES) e Raydonal Ospina (UFBA).

## **Tutorial 2 – 25/10 (08:00 – 10:00)**

**Título:** Joint modeling with INLAjoint

**Ministrantes:** Janet van Nierkerk e Denis Rustand – King Abdullah University of Science and Technology (KAUST)

**Resumo:** Modeling longitudinal and survival data jointly offers many advantages such as addressing measurement error and missing data in the longitudinal processes, understanding and quantifying the association between the longitudinal markers and the survival events, and predicting the risk of events based on the longitudinal markers. A joint model involves multiple submodels (one for each longitudinal/survival outcome) usually linked together through correlated or shared random effects. Their estimation is computationally expensive (particularly due to a multidimensional integration of the likelihood over the random effects distribution) so that inference methods become rapidly intractable, and restricts applications of joint models to a small number of longitudinal markers and/or random effects. R-INLA provides a fast and reliable inference technique for applying joint models to the complex multivariate data encountered in health research. In this tutorial we will show how complex joint models can be fit using the INLA methodology with the R library INLAjoint.

## **5.5 Sessões de Comunicações Orais**

### **SO1: Sessão de Comunicações Orais 1 (23/10/2024, 15:30 - 16:30)**

- **15:30** - *Modelo de Sobrevivência Bayesiano para Dados de Câncer Colorretal – Camille Menezes Pereira dos Santos (UFBA), Maristela Dias de Oliveira (UFBA)*

**Resumo:** O câncer colorretal é uma doença grave que pode ser fatal. Muitos estudos na área biomédica tratam de análise de dados desse tipo de câncer considerando estudos longitudinais, pois possibilitam a identificação fatores prognósticos para o evento de interesse. Este trabalho analisa o tempo até sucessivas reinternações de pacientes diagnosticados com câncer colorretal e submetidos à cirurgia de remoção do tumor. Foi adotado um modelo de regressão com fragilidade compartilhada gama para incorporar possíveis influências de covariáveis na presença de eventos recorrentes, pois não é razoável supor que os tempos de reinternação de um mesmo paciente sejam independentes. Um efeito aleatório (fragilidade) é introduzido multiplicativamente na função taxa de falha Weibull para descrever a possível associação entre os tempos de um mesmo paciente. Os parâmetros do modelo foram estimados utilizando as abordagens frequentista e Bayesiana. Os resultados encontrados são condizentes com estudos semelhantes que ajustam os mesmos dados: o risco de reinternação é maior para pacientes do sexo masculino, com estágios tumorais mais avançados e índices de comorbidade mais elevados. O modelo também indicou que os tempos de um mesmo indivíduo não podem ser considerados independentes e que há heterogeneidade entre os tempos de diferentes pacientes. Por fim,

os resíduos quantílicos demonstraram a adequação do ajuste sob as duas abordagens, evidenciando superioridade das estimativas produzidas pela abordagem Bayesiana.

- **15:45** - *Acesso ao emprego formal, informal e inatividade no Brasil segundo um modelo de riscos competitivos* – Laura Cristina Melo Teixeira (UFMG), Leandro Tavares Correia (UnB)

**Resumo:** O presente trabalho analisou a recente duração do desemprego (até a transição para o emprego formal, informal ou inatividade) no Brasil, de 2021 até 2023. Utilizando dados da PNAD Contínua, investigou-se a relação entre a duração do desemprego e um conjunto de covariáveis. Foi ajustado um modelo de regressão log-normal em análise de sobrevivência na presença de riscos competitivos. Observou-se que o tempo mediano até o ingresso no emprego formal ou informal para as mulheres foi mais que o dobro do tempo dos homens. Pessoas negras apresentaram maior exposição ao emprego informal, com tempo até ingresso nessa ocupação 10% menor que o de pessoas brancas. Os grupos com maior propensão a ingressar na inatividade foram: indivíduos de 14 a 20 anos, com poucos anos de estudo, residentes na região Nordeste e em áreas rurais.

## SO2: Sessão de Comunicações Orais 2 (24/10/2024, 14:00 - 15:00)

- **14:00** - *A new class of semiparametric models for bivariate survival data* – Walmir dos Reis Miranda Filho (UFOP), Fábio Nogueira Demarqui (UFMG)

**Resumo:** We propose a new class of bivariate survival models based on the family of Archimedean copulas with margins modeled by the Yang-Prentice (YP) model. The Ali-Mikhail-Haq (AMH), Clayton, Frank, Gumbel-Hougaard (GH), and Joe copulas are employed to accommodate the dependency between marginal distributions. Baseline distributions are modeled semiparametrically by Piecewise Exponential (PE) distribution and the Bernstein polynomials (BP). Inference procedures for the proposed class of models are based on the maximum likelihood (ML) approach. The new class of models possesses some attractive features: i) the ability to take into account survival data with crossing survival curves; ii) the inclusion of the well-known proportional hazards (PH) and proportional odds (PO) models as particular cases; iii) greater flexibility provided by the semiparametric modeling of the marginal baseline distributions; iv) the availability of closed-form expressions for the likelihood functions, leading to more straightforward inferential procedures. The properties of the proposed class are numerically investigated through an extensive simulation study. Finally, we demonstrate the versatility of our new class of models through the analysis of survival data involving patients diagnosed with ovarian cancer.

- **14:15** - *Proposta de nova taxa de inovação de adoção de políticas públicas por meio de modelo de regressão paramétrico para dados censurados* – Matheus Braga Milhomem (UnB), Carolina Gomes Casado De Carvalho (UnB), Juliana Betini Fachini Gomes (UnB), Denilson Bandeira Coelho (UnB), Terezinha Késsia de Assis Ribeiro (UnB)

**Resumo:** O presente trabalho propõe uma nova medida de inovação política a partir da função de risco do modelo de regressão Weibull para dados censurados. A nova taxa de inovação é uma evolução metodológica em relação aos escores tradicionais da literatura, uma vez que incorpora características dos estados americanos por meio de inclusão das variáveis explicativas na formulação da medida. O conjunto de dados utilizado é composto

por três políticas do artigo de Boehmke e Skinner (2012). As variáveis explicativas foram coletadas de diversas fontes oficiais do governo americano. Após o ajuste dos modelos de regressão para cada política, foi calculada a taxa de inovação proposta e a mesma comparada com as medidas existentes na literatura. A taxa proposta apresentou maior nível de interpretação e a mesma pode ser calculada em diferentes momentos do tempo de estudo.

- **14:30** - *Modelagem do Tempo de Internação até a Morte de Pacientes com COVID-19 no Sistema Público de Saúde do Distrito Federal – Carolyne Soares de Brito (UnB), Juliana Betini Fachini Gomes (UnB)*

**Resumo:** O vírus da COVID-19 teve um impacto transformador no Brasil e no mundo, caracterizando-se pela alta velocidade de transmissão e gravidade dos casos. Embora as medidas de contenção, intervenções medicamentosas e vacinas tenham mostrado eficácia, a infecção ainda afetou muitos brasileiros, resultando em uma alta demanda por assistência hospitalar e, conseqüentemente, na superlotação dos hospitais e em diversos óbitos. O presente estudo tem como objetivo identificar os fatores que influenciaram o tempo de hospitalização até a morte de indivíduos que foram admitidos no sistema público de saúde com COVID-19, no Distrito Federal, entre 2020 e 2023, utilizando técnicas de análise de sobrevivência. Para isso, investigou-se a adequação de diversas distribuições de probabilidade, com foco em distribuições unimodais devido ao comportamento observado nos dados, para determinar a que melhor se ajusta à variável tempo. As distribuições analisadas incluíram Log-Logística, Log-Normal, Inversa Gaussiana Reparametrizada, Burr-XII, Kumaraswamy-Log-Normal, Kumaraswamy-Log-Logística, Kumaraswamy-Inversa-Gaussiana-Reparametrizada e Kumaraswamy-Burr-XII. Além disso, foram consideradas variáveis explicativas relacionadas aos atributos dos pacientes para propor modelos de regressão baseados na distribuição selecionada. Os modelos finais, ajustados com as distribuições Log-Normal e Burr-XII, identificaram como variáveis significativas "Idade" do paciente, "Sexo" do paciente, "Período" de internação, "Valor Total" gasto e a interação entre "Período" e "Valor Total". Os resultados indicam que pacientes mais velhos têm menor probabilidade de sobrevivência e que homens apresentam uma probabilidade de sobrevivência inferior em comparação às mulheres no contexto da hospitalização por COVID-19. Além disso, a interação entre "Período" e o logaritmo de "Valor Total" sugere que quanto maior o gasto com o paciente no período "pós-pandemia", maior a probabilidade de sobreviver do que pacientes no período de pandemia. Após o ajuste dos modelos, cujos parâmetros foram estimados pelo método de máxima verossimilhança, os mesmos foram avaliados por meio dos resíduos de Cox-Snell. Um estudo de simulação também foi realizado, mostrando que o método de estimação apresentou consistência, independentemente do tamanho da amostra e da porcentagem de censura.

- **14:45** - *A Análise de Reincidências Criminais: Uma Abordagem Baseada em Modelos Hierárquicos para a Análise de Dados de Eventos Recorrentes Inflacionados de Zeros e Correlacionados Espacialmente – Alisson Carlos da Costa Silva (IPEDF), Fábio Nogueira Demarqui (UFMG), Marcos Oliveira Prates (UFMG)*

**Resumo:** O estudo das trajetórias criminais na criminologia busca entender como as carreiras no crime começam, progridem e são interrompidas. Este trabalho apresenta novos modelos que superam as limitações dos métodos tradicionais na análise de reincidência criminal, especialmente em dados com excesso de zeros e correlação espacial.

Utilizando métodos de contagem, esses modelos estendem o Processo de Poisson Não Homogêneo, incorporando efeitos aleatórios para analisar associações espaciais. Além dos modelos paramétricos baseados na Lei de Potência, versões semiparamétricas flexíveis foram propostas, utilizando Polinômios de Bernstein. A abordagem Bayesiana facilita a incorporação de evidências externas e a modelagem de correlações. Testados em simulações e aplicados à Região Metropolitana de Belo Horizonte, os modelos ofereceram uma análise detalhada das áreas de maior risco e da taxa de reincidência ao longo do tempo, contribuindo para o entendimento das trajetórias criminais e o desenvolvimento de estratégias mais eficazes no combate à criminalidade.

### SO3: Sessão de Comunicações Orais 3 (25/10/2024, 10:30 - 11:30)

- **10:30** - *Estudo da Evasão Feminina nos Bacharelados de Estatística, Matemática e Ciência da Computação da Universidade de Brasília – Beatriz Gomes da Cruz Albernaz (UnB), Juliana Betini Fachini Gomes (UnB), Maristela Terto de Holanda (UnB), Maria Teresa Leão Costa (UnB)*

**Resumo:** A pesquisa em questão visa aprofundar o estudo sobre a evasão nos cursos do Bacharelado em Estatística, Matemática e Ciência da Computação da Universidade de Brasília (UnB), com um foco específico na evasão feminina. Ao definir a evasão como a saída da aluna do curso antes do término por diferentes razões - troca de curso, mudança de instituição, etc - ou quando ela simplesmente abandona a graduação, este trabalho busca obter resultados para subsidiar discussões sobre a evasão e estratégias para combatê-la. O estudo destaca a importância de compreender as nuances da evasão feminina contribuindo para a formulação de políticas e práticas que promovam a retenção de estudantes e a equidade de gênero nessas áreas acadêmicas. A metodologia de análise de sobrevivência foi a técnica utilizada, mais especificamente o modelo de regressão Log-normal para a Estatística, o modelo de regressão Weibull para a Matemática e o modelo de riscos proporcionais de Cox para Ciência da Computação.

- **10:45** - *Additive frailty model for recurrent events data with application to cancer data – Vera Tomazella (UFSCar), Felipe Rodrigues da Silva (UFPI), Cleide Mayra Menezes Lima (UFPI), Saralees Nadarajah (University of Manchester, UK)*

**Resumo:** Both the additive and proportional intensity models provide two principal frameworks for studying the association between risk factors and disease recurrences. When the events of interest are not terminal and can occur more than once for the same individual, we have the so-called recurrent events, these types of data appear in areas such as biomedicine, criminology and industrial reliability. In this paper, we study an additive intensity model with gamma frailty and propose an estimator for the individual frailties of patients. An advantage of the studied model is the possibility to jointly consider the heterogeneity among patients and to evaluate the dependence within recurrent events captured by the frailty variable. Such a distribution has theoretical arguments to model medical data and has been shown empirically to be a good option. We consider likelihood-based methods to estimate the model parameters, and also investigate large-sample properties of the estimators. In order to illustrate our methodology, we consider two data sets including one from an experimental animal carcinogenesis study.

- **11:00** - *A Bayesian Bivariate Regression Model for Long-Duration Survival Data: Study*

---

*on Customer Churn – Vicente Garibay Cancho (ICMC-USP), Michelle Maciel Sacramento (ICMC-USP)*

**Resumo:** O churn de clientes, ou a taxa na qual os clientes deixam uma empresa, pode impactar significativamente a receita e a estabilidade financeira. Taxas altas de churn indicam insatisfação dos clientes ou a atração por concorrentes, enquanto taxas baixas indicam satisfação e lealdade. Gerenciar o churn de forma eficaz é crucial para a retenção e o crescimento. Nosso estudo foca no churn de dois produtos de cartão de crédito no Brasil, analisando o tempo até o churn ao longo de sete anos. Introduzimos um modelo de sobrevivência bivariado com frailty, utilizando as distribuições de mistura de Poisson e Power Variance Function (PVF). O modelo é estendido para uma análise de regressão, permitindo avaliar a influência de covariáveis na lealdade dos clientes e no risco de abandono do produto. A análise bayesiana é realizada usando métodos de Hamiltonian Monte Carlo (HMC), e também discutimos a comparação de modelos. Esta pesquisa oferece insights importantes sobre o comportamento dos clientes, fornecendo uma base para estratégias de redução de churn e aumento da lealdade dos clientes.

- **11:15** - *Estratégia para Avaliação Preliminar da Confiabilidade Estatística de Sistemas em Desenvolvimento – Oilson Alberto Gonzatto Junior (ICMC-USP), Francisco Louzada Neto (ICMC-USP), Vera Tomazella (UFSCar)*

**Resumo:** Durante o desenvolvimento de um novo sistema, a falta de dados experimentais consolidados exige a busca por novas fontes de informação. A Análise de Modos e Efeitos de Falha (FMEA) é uma das poucas disponíveis para avaliar a confiabilidade inicial, mas a presença de índices RPN idênticos, mesmo com valores diferentes para S, O e D, limita sua eficácia. O método sugerido propõe entrelaçar três tipos de informações para o monitoramento preliminar da confiabilidade: (1) Associar os requisitos de confiabilidade com modelos teóricos estatísticos; (2) Associar esses requisitos com os índices da FMEA; e (3) Associar os índices da FMEA com quantidades práticas baseadas em modelos teóricos. Isso cria uma relação entre as fontes preliminares e os modelos teóricos, resultando em uma avaliação quantitativa preliminar que sistematiza a crença subjetiva dos desenvolvedores e fornece uma base probabilística antes da coleta formal de dados.

## 5.6 Sessões de Apresentação de Pôster

### Sessão de Pôsters 1 (23/10/2024, 16:30 - 17:30)

- **PO01** - *Aplicação de modelo de eventos competitivos ao tempo de permanência de estudantes universitários – Edleide de Brito (UFBA), Giovana Oliveira Silva (UFBA), Maristela Dias de Oliveira (UFBA)*

**Resumo:** Este trabalho estuda o tempo de permanência de estudantes em cursos de Graduação na Universidade Federal da Bahia - UFBA. Dois eventos mutuamente exclusivos podem ser observados ao se analisar a saída do estudante da vida acadêmica: formatura (quando o aluno obtém a graduação) ou evasão (quando o aluno deixa a vida universitária sem se graduar). Para identificar fatores que podem influenciar nesses dois resultados, um modelo de regressão Weibull Generalizado considerando eventos competitivos é adotado para avaliar o efeito de covariáveis como sexo, idade, atividades extracurriculares e quota sobre o tempo até que o estudante deixe a vida universitária. Dados oriundos do CEN-SUP (Censo da Educação Superior) de 2019 referentes a 429 estudantes dos cursos de

Matemática e Estatística tratam de tempo (em semestres) desde a sua entrada na UFBA, via ENEM ou vagas residuais. Alunos com matrícula ativa no período do censo foram considerados censurados para ambos os eventos. Visando acomodar a natureza discreta dos registros, um modelo de probabilidade Weibull Discreto também deve ser considerado. Resíduos quantílicos são utilizados para avaliar a adequabilidade do modelo, cujos parâmetros são estimados via abordagem frequentista e Bayesiana. Os achados mostram que a participação de alunos em atividades extracurriculares contribui para reduzir o tempo de graduação bem como a probabilidade de evasão estudantil.

- **PO02** - *Câncer de pâncreas: Fatores associados a longevidade e cura* – Vanessa Maculan Silvério (UFLA), Luiz Ricardo Nakamura (UFLA), Thiago Gentil Ramires (UTFPR)

**Resumo:** O câncer é uma doença caracterizada pelo crescimento descontrolado de células malignas em diferentes partes do corpo e afeta milhões de pessoas globalmente, sendo uma das principais causas de mortalidade. Alguns tipos, como o câncer de pâncreas, apresentam prognósticos mais desafiadores devido à detecção tardia e à rápida progressão da doença. Embora muitos pacientes venham a falecer devido à doença, é importante ressaltar que nem todos são fatalmente afetados por ela, evidenciando a importância da análise de sobrevivência, que desempenha um papel crucial ao estudar o tempo de vida após o diagnóstico, permitindo uma compreensão mais completa dos desfechos dos pacientes. Diante disso, este estudo investigou o tempo de sobrevivência, do diagnóstico até o fim do acompanhamento no estudo, de pacientes diagnosticados com câncer de pâncreas, a fim de explorar os fatores prognósticos. Os dados foram obtidos na Fundação Oncocentro de São Paulo, os quais incluem variáveis com informações sobre: idade, sexo, estágio do câncer, tipos de tratamento, dentre outras. Como nem todos os pacientes apresentaram óbito ao final da pesquisa, os mesmos apresentam casos de censura. Como metodologia, utilizou-se o teste de Kaplan-Meier, que permitiu estimar e comparar as curvas de sobrevivência em função das diferentes variáveis. Para comparar tais curvas, utilizou-se o teste de log-rank, cujos resultados evidenciaram diferenças significativas entre os níveis de todas as variáveis categóricas, sendo o grupo de maior risco caracterizado por: pacientes do sexo masculino, com ensino fundamental incompleto, que estão no estágio mais avançado do câncer de pâncreas, com o tumor crescendo e invadindo estruturas importantes ao redor do pâncreas, como grandes vasos sanguíneos, e se espalhando para linfonodos próximos e órgãos distantes, como o fígado e os pulmões. Em relação ao tratamento recebido no hospital, os maiores riscos são observados em pacientes que não passaram por cirurgia, terapia hormonal e imunoterapia, assim como naqueles que receberam radioterapia e quimioterapia.

- **PO03** - *A Discrete Weibull Proportional Odds Survival Model* – Marcílio Ramos Pereira Cardial (ICMC-USP), Juliana Cobre (ICMC-USP), Eduardo Yoshio Nakano (UnB)

**Resumo:** The purpose of this paper is to present the development of the Proportional Odds Model for discrete time-to-event data in a survival analysis context. In this work, inferences about the model parameters considering the discrete Weibull as baseline distribution are presented, obtaining the point and interval estimates of the model parameters. In addition, procedures for checking the proportional odds assumption were proposed. The proposed model is illustrated using a dataset on the survival time of patients with Leukemia. The proposed model showed a good fit to the real data set, proving to be a good alternative for modeling discrete time-to-event survival data with covariates.

- **PO04** - *Modelagem de Dados de Sobrevivência Grupados: Modelagem da mortalidade de árvores no bioma Mata Atlântica no Brasil* – Marciel Lelis Duarte (UFV), Sebastião Martins Filho (UFV), Carlos Moreira Miquelino Eleto Torres (UFV)

**Resumo:** Modelos para prever a mortalidade de árvores podem ajudar a entender a dinâmica da vegetação e a prever efeitos das mudanças climáticas em florestas nativas. No entanto, os padrões de mortalidade, são altamente variáveis e difíceis de descrever. Apesar de inúmeras investigações visando desenvolver modelos de sobrevivência para árvores, ainda há lacunas importantes que precisam ser preenchidas. O objetivo do presente estudo foi utilizar modelos de sobrevivência grupados para modelar o efeito de variáveis de clima e solo na mortalidade de árvores em Floresta Estacional Semidecidual no bioma Mata Atlântica. Foram utilizados dados de 20 parcelas permanentes de um fragmento de Mata Atlântica em Minas Gerais, Brasil. Como variáveis explicativas utilizaram-se as características químicas do solo (pH do solo, cátions trocáveis, teor de P e matéria orgânica) e clima (temperatura média, precipitação  $\geq$  100 mm, precipitação nos três meses mais secos, precipitação média anual total e déficit hídrico - CWD). Foram ajustados os modelos de riscos proporcionais de Cox e o modelo de chances proporcionais para estudar a influência destas variáveis na mortalidade das árvores. A escolha do melhor modelo foi baseada no AIC, AUC e capacidade preditiva. A mortalidade das árvores é influenciada pelo pH, matéria orgânica do solo, temperatura, precipitação média e pelo déficit hídrico de dois anos. Ambos os modelos apresentaram resultados semelhantes, com uma AUC de 0,976 e capacidade preditiva de 66,67%. O modelo de Cox apresentou menor AIC. A taxa de precisão de classificação do número de árvores mortas é considerada alta, uma vez que eventos de mortalidade são considerados raros em florestas nativas. A metodologia de análise de sobrevivência com dados grupados foi eficiente em avaliar o efeito de variáveis de clima e solo na mortalidade de árvores na Mata Atlântica. Esta técnica pode ser uma ferramenta útil na tomada de decisões dos gestores florestais.

- **PO05** - *Planos de saúde privados: como sobreviver com eles se eles não sobrevivem?* – Carlos Alberto Oliveira de Matos (UNESP)

**Resumo:** Operadoras de Planos de Saúde são empresas responsáveis por administrar e comercializar os planos de saúde que são produtos ou serviços oferecidos que incluem coberturas de assistência médica, odontológica, etc. São supervisionadas e regulamentadas pela ANS (Agência Nacional de Saúde Suplementar) que, para a proteção aos direitos dos consumidores, define normas para assegurar a legalidade, a transparência e a qualidade dos serviços prestados aos beneficiários. As primeiras operadoras de planos de saúde no Brasil, como uma alternativa ao sistema público surgiram nos anos 60 e até os anos 90, sem normatização ou supervisão estatal, cresceram desreguladamente com base em práticas abusivas, cobertura limitada e falta de transparência. A Lei nº 9.656, de 3 de junho de 1998, estabeleceu regras básicas para o setor, foi complementada com a criação da ANS em 2000 e objetivavam a proteção dos consumidores e a organização do setor. 323660 dados do conjunto de dados “Histórico da situação dos planos de saúde” (DIPRO - ANS) referentes a 153141 planos foram analisados. O histórico de cada plano é constituído por um conjunto que varia de 1 a 44 das quatro seguintes situações de comercialização: Atividade, Atividade com Comercialização Suspensa, Cancelamento e Transferência. Neste trabalho foram selecionados os planos de saúde com duas situações em sequência: *atividade* e *cancelamento*. Foram extraídas as datas de início dos sta-

tus dessa sequência que aparece **37846** vezes na base de dados. Das 325 comparações múltiplas de curvas de sobrevivência dos anos entre 1998 e 2023 efetuadas pelo teste de Log-Rank utilizando o método de Benjamini, Hochberg e Yekutieli para o controle das taxas de falsa descoberta do valor P apenas 8 ( $\approx 2,5$  %) rejeitaram a  $H_0$  de ausência de diferença entre as curvas. Essa ausência de homogeneidade pode ser um indicativo da baixa qualidade da população de planos de saúde no período ou da baixa eficiência da ANS como órgão regulador responsável pela supervisão e regulamentação do setor de saúde suplementar. O fato de outra quantidade frequentemente de interesse em uma análise de sobrevivência que é o tempo médio de sobrevivência dos planos de saúde quantificado neste trabalho através da mediana ponderada ter resultado em aproximadamente **2,45 anos** agrava ainda mais este panorama.

- **PO06** - *Modelos de longa duração para avaliar a germinação de sementes de espécies florestais* – Marciel Lelis Duarte (UFV), Sebastião Martins Filho (UFV)

**Resumo:** Durante os testes de germinação, as sementes frequentemente apodrecem ou não germinam. Essas informações, que são frequentemente descartadas, são essenciais e devem ser incluídas na análise estatística para se obter conclusões precisas. Os métodos tradicionais de análise de dados podem não ser ideais para lidar com esses problemas e uma alternativa apropriada é a análise de sobrevivência. Dentre estas técnicas, os modelos de longa duração paramétricos têm sido tradicionalmente considerados quando se espera que uma proporção dos indivíduos analisados não vivencie o evento de interesse. Este é, por exemplo, o caso de estudos de germinação, em que, sementes consideradas duras, vazias ou que sofreram algum dano físico não irão germinar. O objetivo deste trabalho foi avaliar os modelos de mistura paramétricos de longa duração: Exponencial, Weibull, Log-normal, Log-logístico, para modelar o tempo de germinação de sementes de *Jacaranda caroba* e *Cecropia pachystachya*. Os experimentos foram conduzidos durante 30 dias e avaliou-se a germinação em função de duas covariáveis: dois fotoperíodos (fotoperíodo diário de 12/12h e sob escuro contínuo) e duas temperaturas alternadas (30/20 e 25/15 °C). Os parâmetros estimados dos modelos estudados foram utilizados para determinar as funções de sobrevivência e a função de distribuição acumulada. O Critério de Informação de Akaike (AIC) foi utilizado para realizar a seleção do melhor modelo e a qualidade do ajuste foi verificada usando gráficos, comparando a curva do modelo ajustado com o método não-paramétrico. O modelo Log-logístico foi o mais adequado para descrever as curvas de germinação das sementes ao longo do tempo, em ambas as espécies. As duas covariáveis apresentaram efeito significativo sobre a germinação das sementes nas espécies estudadas. Não houve efeito significativo da interação. Sementes de *J. caroba* e *C. pachystachya* no fotoperíodo de 12 horas, tem 1,16 e 1,44 vezes maior tempo de germinação do que no escuro contínuo, respectivamente. Na temperatura 30/20 °C as sementes de *J. caroba* e *C. pachystachya* tem 2,16 e 1,31 vezes menor tempo de germinação do que as que estão na temperatura 25/15 °C.

- **PO07** - *Evaluating imputation strategies for missing data through statistical modeling and simulations* – Enzo Porto Brasil (UnB), Eduardo Yoshio Nakano (UnB)

**Resumo:** The presence of missing data in databases is a common issue that can hinder statistical analyses, which are typically performed on complete datasets. One potential solution is data imputation, which involves estimating and filling in missing values. This study aims to evaluate the performance of imputation techniques that utilize statistical

regression and to propose a practical guide for researchers based on both literature and simulation results. Our methodology involved creating artificial databases to simulate various missing data scenarios, allowing for a comprehensive analysis of different imputation techniques. We examined the impact of the proportion of missing data, the nature of the incomplete variable, and the number of incomplete variables. By adjusting linear regression models, we compared single and multiple imputation methods, along with techniques based on measures of central tendency and predictive modeling. Additionally, we assessed the impact on analysis when observations with missing data were either ignored or deleted, contrasting these outcomes with those obtained from applying imputation methods. The findings indicate that regression models are effective for data imputation, highlighting the importance of selecting appropriate methods based on the nature and extent of missing data.

- **PO08** - *Bayesian Inference in Stochastic Processes to Identify Mortality Attributed to Sepsis* – Nicholas W. Eugenio (IME-USP), Victor Fossaluza (IME-USP)

**Resumo:** We introduce a novel method for calculating attributable population fractions (PAF) and attributable hazard functions (AHF) within the framework of stochastic processes and non-homogeneous Markov chains. This approach is designed to align with existing literature while offering enhanced flexibility for diverse study designs. Motivated by a Brazilian study of over 3800 hospitalized patients across 38 medical centers, which explored the relationship between sepsis exposure and patient outcomes (death and discharge), our method provides a dynamic measure that accounts for time-dependent variations and the impact of covariates. Our Adapted Attributable Hazard Fraction (AAHF) incorporates elements from competing risks analysis and survival analysis, allowing for the calculation of metrics both for specific subpopulations and general measures. We apply this new approach to filtered data from the motivating study, analyzing transitions in patient outcomes and risk factors over time. Our findings highlight the delayed impact of sepsis on mortality. Early in hospitalization (days 1-13), no significant difference in mortality due to sepsis is observed, possibly due to effective interventions or undetected high-risk patients. However, from day 14 onwards, sepsis-related mortality begins to increase, peaking at around 2% by day 18, underscoring the importance of continuous monitoring and aggressive sepsis management in long-term hospitalized patients.

- **PO09** - *Análise de sobrevivência em dados de fluido de corte* – Elizabeth Mie Hashimoto (UTFPR), Fábio Prata Vieira (ESALQ-USP), Janaína Fracaro de Souza Gonçalves (UTFPR)

**Resumo:** Explorar os fatores que afetam o processo de usinagem é fundamental para prolongar a vida útil das ferramentas e assegurar a produção de peças de alta qualidade, promovendo assim maior sustentabilidade e economia. Neste contexto, o objetivo deste trabalho foi avaliar diferentes tipos de fluido e velocidades de corte sobre a vida útil de uma ferramenta. Os dados utilizados neste estudo são provenientes de um ensaio de bancada, cujas covariáveis observadas foram o tipo de fluido (mineral, semi-sintético vegetal e vegetal) e a velocidade de corte (95m/min e 120m/min). A variável resposta foi o número de furos até o desgaste de flanco máximo (0,30mm). No entanto, o número exato de furos em que ocorreu o desgaste máximo não foi definido com precisão, sendo conhecido apenas um intervalo de valores, o que indica a presença de censura intervalar. Dessa forma, uma análise de sobrevivência foi utilizada, considerando três modelos (Weibull,

log-normal e log-logística). O critério de informação de Akaike foi utilizada para escolha do modelo mais adequado e fixado um nível de significância de 5%. Com base no critério de informação de Akaike, a distribuição log-normal foi identificada como o modelo mais apropriado para representar os dados. Além disso, a análise com o modelo de regressão log-normal revelou uma interação significativa entre o tipo de fluido e a velocidade de corte sobre a vida útil da ferramenta no processo de usinagem. Portanto, para uma velocidade de corte de 95m/min, não existe diferença significativa entre o fluido mineral e o fluido vegetal. Entretanto, observou-se uma diferença significativa entre o fluido semi-sintético vegetal e o fluido vegetal em relação a vida útil da ferramenta. Por outro lado, para um velocidade de corte de 120m/min, não houve diferença significativa entre os fluidos do tipo mineral e do tipo semi-sintético vegetal em comparação com o fluido vegetal.

- **PO10** - *Preditores do Tempo de Recuperação Nutricional em Crianças Menores de 59 Meses com Desnutrição Aguda Grave de 2018 a 2022 em Moçambique: Abordagem de Análise de Sobrevivência – Audêncio Victor (USP), Bélio Castro Antônio (Universidade Lúrio), Ana Raquel Manuel Gotine (Universidade Lúrio), Manuel Mahoche (USP), Sancho Pedro Xavier (Universidade Zambeze), Osiyallê Akanni Silva Rodrigues (UFBA), Andreia Ferreira (Fundação Oswaldo Cruz), Patrícia H. Rondó (USP)*

**Resumo: Introdução:** A desnutrição é um problema de saúde pública com efeitos adversos no bem-estar físico e psicossocial, caracterizado por uma rápida deterioração do estado nutricional em um curto período de tempo e/ou edema bilateral causado por ingestão alimentar deficiente e/ou presença de doenças no passado recente. **Métodos:** O estudo é uma coorte retrospectiva envolvendo um total de 1.208 crianças com desnutrição aguda grave acompanhadas em internação em seis distritos da província de Sofala entre 2018 e 2022. Foram considerados os registros de todas as crianças de 0 a 59 meses admitidas nos hospitais dos distritos estudados por desnutrição aguda grave (DAG) e complicações associadas. A variável dependente foi a recuperação, e as variáveis independentes foram idade, sexo da criança, vômitos, desidratação e hipoglicemia, presença de edema nutricional, comorbidades associadas (pneumonia/broncopneumonia, HIV/AIDS materno e infantil, diarreia, anemia, malária e tuberculose) e antropometria na admissão e cura da DAG. O método Kaplan-Meier foi utilizado para traçar curvas de sobrevivência, e o teste de resíduos de Schoenfeld para estudar a proporcionalidade de cada covariável. Análises de regressão de riscos proporcionais de Cox bivariáveis e multivariáveis foram realizadas com razão de risco bruta e ajustada, com intervalo de confiança de 95%. **Resultados:** Na análise bruta, idade, peso, altura, malária, diarreia e desidratação foram considerados significativos. Crianças com menos de 24 meses tinham 28 % menos probabilidade de recuperação (HR = 0,72, IC 95 %: 0,57 - 0,93), peso abaixo de 6,16 kg tinha 2 % menos probabilidade de recuperação (HR = 0,77, IC 95 %: 0,61 - 0,96) e altura acima de 71,1 cm tinha 20 % menos probabilidade de recuperação (HR = 0,80, IC 95 %: 0,64 - 0,99). Por outro lado, crianças com malária, diarreia e desidratação tinham mais probabilidade de recuperação (HR = 1,48, IC 95 %: 1,13 - 1,92), (HR = 1,96, IC 95 %: 1,14 - 2,71) e (HR = 1,66, IC 95 %: 1,20 - 2,30), respectivamente. Após ajuste, a diarreia foi um preditor significativo da taxa de recuperação nutricional. **Conclusão:** O estudo mostrou uma alta taxa de recuperação em crianças com desnutrição aguda grave. Os preditores significativos da taxa de recuperação nutricional foram idade, peso, altura, malária, diarreia e desidratação, embora apenas a diarreia tenha sido um preditor significativo da taxa de recuperação nutricional após o ajuste.

- **PO11** - *Preditores de Tempo até a Morte de Crianças Menores de Cinco Anos em Moçambique: Um Modelo de Sobrevida Usando Dados das Pesquisas Demográficas e de Saúde de 2011 e 2022/23* – Sancho Pedro Xavier (UFMT), Audêncio Victor (USP), Ana Raquel Manuel Gotine (USP), Patricia Hellen Rondó (USP), Ageo Mário da Silva (UFMT)

**Resumo: Introdução:** A mortalidade infantil é um indicador vital para avaliar o estado geral de saúde de uma população. Moçambique enfrenta desafios nessa área, e compreender os fatores de risco é essencial para desenvolver intervenções eficazes. O estudo visou analisar a sobrevida global da mortalidade infantil e seus preditores em Moçambique. **Métodos:** O estudo utilizou dados das Pesquisas Demográficas e de Saúde em Moçambique (MDHS) de 2011 e 2015. Modelos de regressão de Cox e de Kaplan-Meier foram aplicados para determinar o tempo de sobrevida e identificar os preditores de morte. A Razão de Risco ajustada (AHR) com um Intervalo de Confiança (IC) de 95 % permitiram medir a associação entre os fatores analisados. **Resultados:** Em 2011, a taxa de mortalidade antes dos cinco anos foi de 7,3 % , e 3,5 % antes do primeiro ano de vida. Em 2015, essas taxas diminuíram para 5,9 % e 2,9 %, respectivamente. A idade mediana da mortalidade foi de 29,5 meses. Houve uma redução no risco de morte em crianças do sexo feminino (AHR = 0,85; p = 0,024) e de mães com educação secundária ou superior (AHR = 0,77; p = 0,031). Além disso, o risco foi menor em crianças nascidas entre o segundo e o quarto filho, e que viviam em domicílios com 5 ou mais membros (AHR = 0,80; p = 0,045). Por outro lado, o risco de mortalidade foi maior entre mães que tinham ocupação informal (AHR = 1,12; p = 0,033), idade materna no primeiro parto inferior a 20 anos (AHR = 1,25; p = 0,017), mães anêmicas (AHR = 1,18; p = 0,029), crianças com baixo peso ao nascer (AHR = 1,30; p = 0,015) e aquelas nascidas em casa, no setor privado ou em outros locais que não unidades de saúde públicas (AHR = 1,22; p = 0,027). **Conclusão:** Os achados destacam a importância de promover a educação materna e melhorar os serviços de saúde em Moçambique para reduzir as taxas de mortalidade infantil. Investir em educação materna e cuidados pré-natais e obstétricos é essencial para melhorar a sobrevida infantil.

- **PO12** - *Influência de Variáveis Demográficas e Clínicas na Sobrevida de Pacientes com Câncer de Rim* – Daví Barbosa Pereira de Sousa (UFLA), Vitoria Soares de Souza (UEPB), Hellen Sonaly Silva Alves (UFRPE), Elyda Camyla de Araújo Alves (UEPB), Jéssica Cristina Pereira Batista (UEPB), João Domingos Scalon (UFLA), Guilherme Rocha Moreira (UFRPE)

**Resumo:** Os rins são órgãos vitais para a vida e têm uma função fundamental na regulação do equilíbrio químico e de líquidos no corpo. O câncer de rim geralmente surge a partir de mutações genéticas nas células renais, levando a um crescimento celular descontrolado e à formação de tumores. Correspondendo a 3% das doenças malignas que afetam adultos globalmente e é o segundo tipo de câncer mais comum do sistema urinário, sendo superado apenas pelo câncer de bexiga. Entre os sintomas mais frequentes estão sangue na urina, tossir sangue, dor abdominal e perda de apetite. Com isso, o objetivo deste estudo é estimar o tempo, em dias, de sobrevida dos pacientes com câncer de rim em Brasília, desde o diagnóstico até o óbito, com base na faixa etária, sexo e extensão do tumor, e avaliar se essas variáveis influenciam significativamente as curvas de sobrevivência. Os dados foram obtidos do Instituto Nacional de Câncer (INCA) para

o Distrito Federal, e para a organização dos dados, foi considerado como falha o paciente que morreu de câncer e como censura o paciente que não morreu de câncer, não faleceu ou saiu do estudo por outro motivo. As análises foram realizadas no RStudio, utilizando técnicas de Kaplan-Meier e o teste de log-rank. Foram definidos três grupos etários: até 30 anos, de 31 a 60 anos e acima de 60 anos. Para o sexo, os grupos foram divididos entre "feminino" e "masculino". Em relação à extensão do tumor, foram categorizados como "localizado" ou "metástase", desconsiderando os "sem informação". Os resultados mostraram diferenças significativas no tempo de sobrevivência entre os grupos etários, os sexos e as categorias de extensão do tumor. Pacientes mais jovens, do sexo feminino ou com tumor localizado apresentaram maior probabilidade de sobrevivência. Em conclusão, é crucial considerar essas variáveis e suas interações ao planejar estratégias de tratamento e acompanhamento para melhorar os desfechos clínicos e o cuidado dos pacientes com câncer de rim.

- **PO13** - *Sobrevida de Músicas no Top 200 Global do Spotify: Taylor Swift – Davi Barbosa Pereira de Sousa (UFLA), Marcela Silva de Araujo (UFLA), Rayza Camila dos Santos Silva (USP), João Domingos Scalon (UFLA), Marcelo Angelo Cirillo (UFLA), Cibele Maria Russo Novelli (ICMC-USP)*

**Resumo:** O Spotify é hoje uma das maiores plataformas de utilização para consumir músicas globalmente. Atualmente, muitos artistas chegam ao estrelato e possuem músicas conhecidas em todo o planeta. Portanto, o objetivo deste estudo é realizar uma análise de sobrevida das músicas de uma das cantoras mais populares da atualidade, a Taylor Swift. As músicas analisadas estiveram no ranking das 200 músicas mais escutadas na plataforma entre os dias 01 de janeiro de 2024 ao dia 30 de junho do mesmo ano, obtidos no site do Spotify Charts. Todas as análises foram realizadas utilizando o software RStudio. Considerou-se como falha músicas que no período estudado saíram do *chart* e não retornaram mais e como censura a finalização do estudo sem que a música tenha saído. Para obtenção da função de sobrevivência foi utilizada a técnica de estimação de Kaplan-Meier e utilizado o teste de verossimilhança para verificar qual distribuição melhor se assemelha a função estimada. O resultado da análise de sobrevivência para as músicas de Taylor Swift mostra como a popularidade ou relevância das músicas diminui ao longo do tempo. No início, 94,44% das músicas ainda estão "ativas" após o primeiro dia, no entanto, essa porcentagem cai ao longo do tempo, chegando a apenas 44,44% após 10 dias e 8,33% após 36 dias. A pesquisa contribui para o entendimento de como as músicas atuais possuem consumação mais rápida e a constante necessidade de lançamentos para que artistas mantenham seu nome em alta.

- **PO14** - *Sobrevida de Pacientes com Câncer de Pulmão no Estado do Mato Grosso – Davi Barbosa Pereira de Sousa (UFLA), Hellen Sonaly Silva Alves (UFRPE), Marcela Silva de Araujo (UFLA), Gabriel Graciano De Mendonça (UFRPE), João Domingos Scalon (UFLA), Marcelo Angelo Cirillo (UFLA), Guilherme Rocha Moreira (UFRPE)*

**Resumo:** O câncer é uma das principais causas de morte no Brasil, e entre os diversos tipos, o câncer de pulmão se destaca como um dos mais incidentes e letais. O tabagismo, assim como a exposição passiva ao fumo, são fatores de risco significativos para o desenvolvimento desse tipo de câncer, que continua sendo uma das principais causas de mortalidade por câncer no país. Portanto, o objetivo deste estudo é realizar uma análise de sobrevida dos pacientes diagnosticados com câncer de pulmão no estado do Mato

Grosso, explorando fatores como sexo e idade do paciente. Para realização das análises, foram utilizados dados do site do Instituto Nacional de Câncer (INCA) com informações de 2005 a 2020 para o referido estado. Todas as análises foram realizadas utilizando o software RStudio. Considerou-se como falha pacientes que no período estudado faleceram por câncer de pulmão e como censura situações adversas. Para obtenção da função de sobrevivência foi utilizada a técnica de estimação de Kaplan-Meier e realizado o teste de log-rank para comparação entre os grupos. Para separação por faixa etária, foram considerados adultos pessoas de até 65 e idosos os demais, já para sexo foram consideradas duas categorias: "masculino" e "feminino". De modo geral, os resultados iniciais indicaram uma queda rápida nas probabilidades de sobrevivência, com taxa de 88,4% após 1 mês do diagnóstico e 78,8% após 2 meses, refletindo a agressividade da doença. Embora as análises tenham mostrado variações nas curvas de sobrevivência entre diferentes grupos ("Mulher Adulta", "Mulher Idosa", "Homem Adulto" e "Homem Idoso"), não foram encontradas evidências significativas de diferenças nas taxas de sobrevivência baseadas em sexo ou idade. A pesquisa contribui para o entendimento de sobrevivência em pacientes com câncer de pulmão e fornece uma base para futuras investigações clínicas, destacando a necessidade de estratégias de intervenção que possam melhorar os desfechos de saúde nessa população.

- **PO15** - *Ajustes para a Verossimilhança Perfilada no Modelo de Regressão Log-Birnbaum-Saunders sob Censura tipo II* – Thalytta Evilly Cavalcante Silva (UFPE), Audrey Helen Mariz de Aquino Cysneiros (UFPE), Aline Barbosa Tsuyuguchi (UFPE)

**Resumo:** Neste trabalho, desenvolvemos ajustes para a função de verossimilhança perfilada no modelo de regressão log-Birnbaum-Saunders na presença, ou não, de censura tipo II, com base nos ajustes de Cox e Reid (1987), Fraser e Reid (1995), e Severini (1999). Consideramos o parâmetro de forma  $\alpha$ , como o parâmetro de interesse e os coeficientes de regressão  $\gamma_s$  ( $s = 1, \dots, p$ ), os parâmetros de perturbação. Obtivemos os estimadores de máxima verossimilhança perfilado e perfilados ajustados, os intervalos de confiança assintótico, *bootstrap* percentil e *bootstrap* BCa, e os testes baseados nas estatísticas da razão de verossimilhanças perfilada e perfiladas ajustadas. Por meio de simulações de Monte Carlo, avaliamos os comportamentos dos estimadores pontuais e intervalares, assim como os dos testes em termos de tamanho e poder. Os resultados mostraram que os estimadores e testes ajustados superaram os resultados baseados na perfilada, apresentando menores vieses e taxas de rejeição mais próximas aos níveis nominais. Por fim, uma aplicação a dados reais é apresentada.

- **PO16** - *Modelo exponencial por partes potência de longa duração para o contexto de crédito* – Silvana Schneider (UFRGS), Leonardo de Miranda Pinheiro (UFRGS)

**Resumo:** Nesse trabalho propomos um modelo para dados de sobrevivência com fração de cura, em que a distribuição dos tempos é ajustada pela distribuição exponencial por partes potência. Na análise de sobrevivência com fração de cura, existem duas classes principais: os modelos de mistura, proposto por Boag (1949) e Berkson and Gage (1952), e os modelos de tempo de promoção, proposto inicialmente por Yakovlev and Tsodikov (1996). Com o intuito de ajustar conjuntos de dados com diferentes formas de taxas de falha, o uso de modelos semiparamétricos, no caso, o modelo exponencial por partes (MEP) foi considerado, dado a sua aceitação e popularidade decorrente de sua flexibilidade. Optou-se por utilizar diferentes ajustes para modelar os tempos de falhas, sendo

eles o MEP usual e o modelo exponencial por partes potência (MEPP), proposto por Gómez et al. (2017). Assim, o objetivo do presente trabalho consiste na proposta de uma combinação ainda não explorada, de forma conjunta na literatura, através da construção de um modelo de sobrevivência de longa duração, considerando a abordagem dos modelos tempos de promoção e distribuição exponencial por partes potência: sob perspectiva da inferência frequentista e aplicação no contexto de crédito. Para avaliar o desempenho do modelo proposto, realizou-se um estudo de simulação comparando o modelo proposto com os modelos disponíveis na literatura (Weibull e exponencial por partes), utilizando a linguagem de programação R. Assim como, foi conduzido uma aplicação do modelo proposto em um conjunto de dados reais relacionado a empréstimo pessoal, para avaliar a aplicabilidade do modelo em um contexto de crédito.

- **PO17** - *Modelo de Regressão Weibull ajustado ao Tempo de Sobrevivência de Empresas Mercantis Brasileiras – Jonatas Silva do Espirito Santo (Sebrae Nacional), Felipe Marcel Neves (Sebrae Nacional), Lucas de Melo Alves (Sebrae Nacional), Juliana Borges Vaz (Sebrae Nacional), Tomaz Back Carrijo (Sebrae Nacional)*

**Resumo:** Para o SEBRAE, é importante conhecer a dinâmica de aberturas e fechamentos das empresas. Para tanto, foi realizado um estudo com mais de 15 milhões de Pequenos Negócios (Microempreendedores individuais - MEI, Microempresas - ME e Empresas de Pequeno Porte -EPP) abertos entre o ano de 2019 e 2023. Foram realizadas análises das curvas de sobrevivência Kaplan-Meier dos Pequenos Negócios, estratificadas por porte, setor de atividade e a Unidade da Federação de registro da empresa. Foram ajustados modelos de regressão Weibull para o tempo de sobrevivência dos Pequenos Negócios, com as covariáveis porte, setor e unidade da federação. Dentre os resultados importantes do estudo, destaca-se que a chance de um MEI fechar é 3,12 vezes maior que a de uma empresa EPP (referência) e uma ME tem 35% mais chances de fechar que uma EPP.

- **PO18** - *Sobrevivência de Empresas Mercantis Brasileiras: Efeitos do Porte, Setor, UF e da COVID-19 – Felipe Marcel Neves (Sebrae Nacional), Jonatas Silva do Espirito Santo (Sebrae Nacional), Lucas de Melo Alves (Sebrae Nacional), Juliana Borges Vaz (Sebrae Nacional), Tomaz Back Carrijo (Sebrae Nacional)*

**Resumo:** Este estudo analisa a sobrevivência de mais de 15 milhões de empresas mercantis brasileiras criadas entre 2019 e 2023, utilizando dados do Cadastro Nacional de Pessoa Jurídica (CNPJ) da Receita Federal. O SEBRAE tem especial interesse nesse fenômeno para os pequenos negócios (MEI e MPE - Microempresas e Empresas de Pequeno Porte). MEI apresentam taxas de sobrevivência (Kaplan-Meier) significativamente menores que as MPE (71,6% vs 86,9% para 2 anos,  $p < 0,05$ ). A mediana de sobrevivência dos MEI é de aproximadamente 5 anos, evidenciando sua maior vulnerabilidade. O modelo de regressão de Cox apontou que porte, setor de atuação e unidade federativa são fatores críticos para o risco de fechamento. O impacto da pandemia de COVID-19 foi estimado usando covariáveis relacionadas à abertura e fechamento de empresas no período, revelando um aumento significativo no risco de encerramento.

- **PO19** - *Discrete Weibull distribution applied to the probability of recovery in COVID-19 patients – Fábio Prata Vieira (ESALQ-USP), Elizabeth Mie Hashimoto (UTFPR)*

**Resumo:** Survival analysis investigates the duration until the occurrence of specific events, ranging from the death of an individual to failures in mechanical or electrical

devices. Regression models play a crucial role in this field by enabling the examination of the impact of different variables on the probability of events occurring within specific periods. This study proposes a cure fraction model using the discrete Weibull distribution to analyze the effect of sex and age group on the duration and probability of recovery in Chinese patients with COVID-19. The regression coefficients and the fraction of recovered individuals are estimated using maximum likelihood and Bayesian methods. Survival functions, stratified by age group and sex, are presented. The results indicate that age group has a significant impact on patient survival time, while women exhibit a higher probability of recovery compared to men. Individuals aged  $\geq 60$  years show a lower probability of recovery compared to those aged  $< 60$  years. This model may serve as a valuable tool for analyzing the survival time of COVID-19 patients, particularly when considering data censoring and discrete time.

- **PO20** - *Análise dos fatores associados à desligamento de funcionários em empresas privadas do Nordeste – Fernando Humberto de Almeida Moraes Neto (UFBA), Mariana Thais Almeida, Edleide de Brito, Lilia Carolina Carneiro da Costa, Rafael Toledo Costa de Almeida (UFBA), Adriano Kamimura Suzuki (USP)*

**Resumo:** Este trabalho compara modelos de Análise de Sobrevivência para avaliar os fatores de risco de demissão de funcionários que trabalham em empresas privadas do Nordeste. Os dados utilizados foram da Relação Anual de Informações Socioeconômicas de todos os estados da Região Nordeste em 2015. Verificou-se que os trabalhadores que trabalham em empresas privadas do Nordeste são em sua maioria homens, pardos, brasileiros e com idade entre 18 e 39 anos. Os trabalhadores das capitais e regiões metropolitanas têm faixa salarial de até três períodos mínimos e trabalham de 41 a 44 horas semanais. Eles são contratados por meio de reemprego, permanecem nas empresas por aproximadamente três meses e são desligados por rescisão sem justa causa por iniciativa do empregador. Ao comparar os modelos de análise de sobrevivência com o estimador de Kaplan-Meier e utilizando o Critério de Informação de Akaike, observou-se que, dentre os modelos ajustados, o modelo gama generalizado foi o mais adequado aos dados. No modelo final, as covariáveis sexo do trabalhador, idade, raça/cor e faixa salarial média foram definidas em intervalos mínimos.

- **PO21** - *Análise da sobrevida de pacientes com câncer de pele através da regressão GAMLSS com fração de cura – Celso Menoti da Silva (UFRGS), Silvana Schneider (UFRGS)*

**Resumo:** Na análise de sobrevivência, o objetivo é avaliar o risco de um evento específico, bem como a probabilidade de sobrevivência. Assume-se que todos os indivíduos são suscetíveis a esse evento. Estudos de sobrevivência são amplamente aplicados na engenharia de produção, análise de crédito, saúde e muitas outras áreas. Modelos estatísticos de sobrevida tradicionais geralmente não são adequados para analisar conjuntos de dados em que uma parcela dos indivíduos não terá o evento de interesse. Em estudos da área da saúde, a proporção desses indivíduos é denominada fração de cura. A modelagem por regressão é uma ferramenta estatística muito utilizada em análise de sobrevivência. Os modelos mais tradicionais como a regressão linear simples ou os modelos lineares generalizados exigem suposições quanto ao tipo de distribuição da variável resposta e não são indicados para modelar relações não-lineares. Para superar essas limitações, surgiram os modelos de regressão GAMLSS (Generalized additive models for location, scale and shape), que permite modelar os parâmetros de localização ( $\mu$ ), escala ( $\sigma$ ) e forma ( $\nu$  e  $\tau$ ) em

função de covariáveis. O presente trabalho apresenta uma modelagem da sobrevivência de pacientes com câncer de pele do estado de São Paulo, dados fornecidos pela FOSP, com a finalidade de estimar a fração de cura. O modelo de regressão adotado foi uma extensão do GAMLSS, seguindo a metodologia proposta por Ramires et al. (2019). Por essa sistematização, a fração de cura é incorporada como um parâmetro na extensão da distribuição Weibull, para ajustar o tempo de sobrevivência. Dessa forma, foi possível avaliar a influência das seguintes covariáveis: Tipo de Tratamento; Estágio da doença; Gênero; Tipo de atendimento. Bem como, avaliar a fração de cura e a sobrevivência populacional.

- **PO22** - *A Bayesian approach to model the dependence of observed and censored lifetimes in medical applications* – Ricardo Puziol de Oliveira (UNESP), Suellen Isabela Carlette Ferreira (UEM), João Debastiani Neto (UEM), Edson Zangiacomí Martinez (USP), Jorge Alberto Achcar (USP)

**Resumo:** In medical research, specially in oncology applications, it is frequently observed that there exists a dependence structure between survival times and dropout times. This phenomenon arises because terminally ill patients, who are participating in clinical trials, may be withdrawn from the study and return to their homes, where they typically experience death shortly thereafter. Traditional methodologies often assume that survival times and censoring times are independent. However, this assumption may not be appropriate in many medical contexts and could potentially distort the analysis of lifetime data, particularly when important covariates such as disease stage or patient age are not accounted for. In this study, we address this issue by employing the Basu-Dhar discrete geometric bivariate lifetime distribution alongside frailty models to capture and model the dependence between survival and dropout times. To obtain the inferences of interest, we used Markov Chain Monte Carlo (MCMC) methods. This approach allows for a more accurate representation of the data structure and improves the robustness of the lifetime data analysis by incorporating potential dependencies and relevant covariates.

- **PO23** - *A Weibull-Based Non-Mixture Cure Rate Model Applied to Chinese Lung Cancer Data* – Ricardo Puziol de Oliveira (UNESP), Suellen Isabela Carlette Ferreira (UEM), João Debastiani Neto (UEM), Edson Zangiacomí Martinez (USP), Jorge Alberto Achcar (USP)

**Resumo:** In medical research, particularly in oncology, it is essential to analyze how subject-specific factors influence the likelihood of experiencing critical events such as death and to explore the conditions leading to right-censored observations, often indicative of "cured" patients. Traditional survival analysis techniques, such as the Cox proportional hazards model, do not provide direct estimates of cure rates, highlighting the need for alternative approaches, including mixture and non-mixture cure fraction models. This study introduces a novel non-mixture cure rate model based on the Weibull distribution, applying it to analyze survival data from a cohort of lung cancer patients. The dataset, provided by Ding et al. (2017), includes lifetimes of Chinese patients with pathologically confirmed lung cancer who underwent mutation testing for EGFR, KRAS, and BARF at the Thoracic Cancer Institute, Tongji University, between January 2012 and April 2016. Smoking status and cancer stage were also assumed as covariates potentially influencing survival outcomes. This study aims to refine the understanding of cure rates and survival

---

patterns in oncology through the application of an innovative Weibull-based non-mixture cure rate model.

**PO2: Sessão de Pôsters 2 (24/10/2024, 16:00 - 17:00)**

- **PO24** - *Modelo de regressão heterocedástico OLL-Exponencial com fração de cura e fragilidade – Cleanderson Romualdo Fidelis (UFAC), Edwin Moises Marcos Ortega (ESALQ-USP)*

**Resumo:** Os modelos de sobrevivência desempenham um papel central na literatura estatística, sendo amplamente utilizados para modelar diversos problemas que envolvem o tempo até a ocorrência de um evento. Esse tipo de modelagem frequentemente lida com dados censurados, o que implica desafios adicionais na escolha correta do modelo. Em particular, os modelos de sobrevivência com fração de cura são relevantes em situações em que parte da população estudada é imune ao evento, exigindo abordagens que considerem tanto indivíduos suscetíveis quanto não suscetíveis ao evento. Este trabalho busca apresentar o modelo de regressão odd log-logistic exponencial com fração de cura e fragilidade. O modelo proposto foi formulado a partir da flexibilização da distribuição de Poisson pela inclusão de um parâmetro adicional, que funciona como um parâmetro de dispersão para mensurar possível heterogeneidade, dispersão ou correlação entre o número de células carcinogênicas. Utilizando essa estrutura, foi estudada a incorporação de covariáveis que influenciem tanto a fração de curados quanto a distribuição associada aos tempos de vida dos indivíduos suscetíveis. Além disso, utilizamos os resíduos quantílicos aleatorizados, calculados com base na função quantílica da distribuição normal padrão aplicada à função de distribuição populacional ajustada do modelo. Essa abordagem permite identificar possíveis desvios e garantir a adequação do modelo aos dados observados, assegurando a robustez das inferências realizadas. O trabalho inclui simulações para ilustrar o desempenho do modelo proposto em cenários envolvendo ausência de covariáveis, considerando os tempos até a ocorrência do evento vindos de uma mesma distribuição, e em cenários envolvendo a presença de covariáveis nos parâmetros relacionados à fração de curados e à distribuição dos tempos dos indivíduos suscetíveis modelados pela distribuição exponencial. A modelagem foi aplicada aos tempos até a morte por câncer de cólon, incluindo diferentes tipos de tratamentos com quimioterapia, estágios do câncer e informação sobre a realização de cirurgia, destacando a importância de métodos rigorosos para a análise de modelos com fração de cura e fragilidade.

- **PO25** - *Abordagem semi-paramétrica com fração de cura e zero inflacionado para avaliar o tempo de inadimplência – Rafaela Vidal Galetto (UFRGS), Silvana Schneider (UFRGS)*

**Resumo:** Técnicas de análise de sobrevivência são utilizadas para avaliar o tempo até a ocorrência de eventos no setor financeiro. Com a utilização da análise de sobrevivência nas instituições financeiras, além de estudar o tempo até o evento, foi necessário avaliar os casos em que uma fração dos indivíduos do estudo nunca vivenciará o evento de interesse e uma fração dos indivíduos que sofrerá o evento no tempo zero. Ribeiro de Oliveira Jr et al. (2017) propuseram um modelo de mistura padrão com zero inflado e distribuição weibull, no qual foi adicionada uma proporção ( $p_0$ ) correspondente indivíduos que sofreram o evento no tempo zero. A justificativa baseia-se no cenário de crédito, em que, em diferentes contextos, é preciso lidar com clientes que sofrem o evento no início

do estudo, enquanto há clientes que nunca sofrerão o evento e, também, clientes que sofrerão o evento ao longo do tempo. O objetivo deste trabalho é estender a abordagem de Ribeiro de Oliveira Jr et al. (2017), utilizando a distribuição exponencial por partes de potência (EPP). O modelo exponencial por partes é uma abordagem semiparamétricos, comumente utilizada para ajustar tempos com diferentes formas de taxas de falha, em virtude da sua flexibilidade. Para avaliar o desempenho do modelo proposto foi realizado um estudo de simulação, comparando as distribuições Weibull, exponencial por partes e exponencial por partes potência. O modelo proposto apresentou resultados satisfatórios, com vício relativo em geral abaixo de 10%, valores de desvio padrão próximos ao erro padrão médio e a probabilidade de cobertura próxima do valor nominal de 95%, para todos os parâmetros. Por fim, foi realizada uma aplicação para estimar a probabilidade de *default*, de uma carteira de empréstimos em uma instituição financeira. Para tanto, foi estudada uma amostra de 15.000 clientes e o *default* foi definido como atingir a marca de 35 ou mais dias de atraso. O interesse foi modelar o tempo até que esse *default* aconteça. O tempo zero, nesse caso, é o cliente não cumprir com o vencimento logo no primeiro pagamento após a concessão, que em geral se dá no mês seguinte. Através do gráfico do estimador Kaplan-Meier foi possível avaliar a presença de uma proporção de clientes inadimplentes no tempo zero. Além disso, é possível avaliar que a função de sobrevivência não tende a zero, demonstrando haver um platô próximo do valor 0,41, indicando haver uma proporção de clientes que possivelmente nunca irão atrasar seus pagamentos. Portanto, o modelo proposto se mostrou uma ótima alternativa para o ajuste de dados com fração de cura e zero inflacionado.

- **PO26** - *Aplicação do Modelo de Regressão de Cox para estimar o tempo entre ultrapasses de concentração de poluentes na Cidade do México – Maria Isabela de Souza Pompei (UNESP), Mário Hissamitsu Tarumoto (UNESP), Eliane Regina Rodrigues (UNAM México)*

**Resumo:** A poluição do ar tem um impacto significativo na vida das pessoas em todo o mundo, contribuindo para problemas de saúde, como doenças respiratórias e cardiovasculares, além de reduzir a qualidade de vida e a expectativa de vida em muitas regiões. Na cidade do México, a situação não é diferente; a poluição atmosférica afeta gravemente a saúde da população, especialmente na região central da cidade. Neste estudo, adotamos um modelo de regressão de Cox para analisar o tempo entre os momentos em que a concentração de um poluente excede um limite específico de interesse, em particular, a concentração de oxigênio na atmosfera. As covariáveis consideradas foram: dióxido de nitrogênio (NO<sub>2</sub>), dióxido de carbono (CO<sub>2</sub>), temperatura (TMP), umidade relativa do ar (HR) e partículas inaláveis (PM<sub>10</sub>). As covariáveis foram utilizadas com uma hora de atraso. Esse procedimento de predição permite estimar o comportamento da função média ao longo do tempo e principalmente identificar as principais covariáveis que influenciam o tempo entre ultrapassa de um limiar (0.11 p.p.m.) de concentração de ozônio.

- **PO27** - *Gamma-Dagum Defective Frailty Regression Model – Cynthia Arantes Vieira Tojeiro (UFG), Vera Lúcia Damasceno Tomazella (UFSCar)*

**Resumo:** Cure fraction modeling, or long-duration modeling, addresses scenarios where non-susceptible observations to the studied event exist. Cure rates can be modeled using defective distributions, which adjust their usual parameter domain, enabling survival functions to converge to a value  $p \in (0; 1)$ . In such studies, there is also a need to estimate

the unobservable heterogeneity caused by prognostic factors that cannot be observed. In this context, we proposed a new defective distribution derived from the Gamma-Dagum defective distribution and use the gamma function for the frailty term introduced in the hazard function to control for unobservable heterogeneity in patient populations. Considering censored data and covariates, we show that the new distribution Gamma-Dagum with a frailty term possesses defective characteristics, making it suitable for data involving cure fractions. We utilize maximum likelihood estimation for parameter inference, assessing their asymptotic properties through simulations. We illustrate the methodology using a real dataset from São Paulo Oncocenter Foundation, São Paulo, Brazil, concerning survival time of melanoma cancer patients.

- **PO28** - *Análise de Sobrevida em Câncer Colorretal: Aplicação de Florestas Aleatórias – Rafael Sant Ana Herzog (UFES), Vinicius Fernando Calsavara (Cedars-Sinai), Agatha Sacramento Rodrigues (UFES)*

**Resumo:** Com os avanços tecnológicos e o aumento do poder computacional, os métodos de *machine learning* que permitem combinar modelos estatísticos e algoritmos computacionais para extrair mais informações de grandes conjuntos de dados têm ganhado cada vez mais popularidade. No entanto, com o aumento da variedade de modelos disponíveis, torna-se cada vez mais evidente a necessidade de se entender adequadamente a estrutura do conjunto de dados em estudo e se a utilização de um certo modelo faz sentido dentro dessa estrutura – afinal, são os modelos que devem se adequar aos dados, e não o contrário. Nesse contexto, revisitamos, neste trabalho, um estudo a respeito da sobrevivência de pacientes com câncer colorretal tratados no estado de São Paulo entre 2000 a 2021, o qual, além de outros problemas, ignorou por completo a existência de observações censuradas no conjunto de dados, utilizando métodos de classificação usuais para prever a probabilidade de um paciente sobreviver por pelo menos um certo período de tempo. Discutimos sobre como ignorar a existência de dados censurados pode enviesar o resultado de um estudo, e apresentamos modelos de aprendizado de máquina específicos para dados de sobrevivência que podem - e devem - ser utilizados quando tratamos de dados desse tipo. Mais especificamente, focamos nas Florestas Aleatórias para Análise de Sobrevida (do inglês *Survival Random Forests*, SRFs), modelos adaptados das florestas aleatórias de regressão que incorporam a censura em sua formulação. Ajustamos SRFs ao conjunto de dados utilizado no estudo original, discutindo sobre a inclusão de certas variáveis, a forma correta de se utilizar variáveis categóricas e as diferentes regras de decisão que podem ser adotadas nesse tipo de modelo, e obtivemos, para o modelo final, C-index de 76% e Integrated Brier Score de 0,149. Para o modelo final, calculamos, ainda, os valores SHAP, que revelaram que estar no estadiamento clínico IV, não ter realizado cirurgia e ter realizado quimioterapia ao longo do tratamento foram as informações que mais impactaram as previsões do modelo.

- **PO29** - *Fatores Associados ao Insucesso do Tratamento da Tuberculose na População em Situação de Rua: Uma Coorte Retrospectiva – Osiyallê Akanni Silva Rodrigues (UFBA), Hamed Oladeji Mogaji (UFBA), Layana Costa Alves (UFBA), Renzo Flores-Ortiz (Fiocruz), Cleber Cremonese (UFBA), Joilda Silva Nery (UFBA)*

**Resumo: Introdução:** A tuberculose (TB) é uma doença prevenível, curável, mas que ainda é responsável por 4.000 mortes diariamente. Pessoas em situações de pobreza, atingidas pela insegurança alimentar e nutricional, com redução no acesso aos serviços

básicos de saúde, como a população em situação de rua (PSR), são mais afetadas pela enfermidade e possuem dificuldades no tratamento. O estudo visou analisar os fatores associados ao insucesso do tratamento da TB, ao longo do tempo, entre a PSR e pessoas abrigadas. **Métodos:** O estudo foi desenvolvido com dados secundários de novos casos de TB, entre a PSR e pessoas abrigadas, registrados no Sistema de Informação de Agravos de Notificação (Sinan), para realizar uma coorte retrospectiva no período de 2015 a 2020. O modelo de regressão de Cox foi utilizado para identificar os fatores associados à não conclusão do tratamento no período de 270 dias de terapia. Análises de sensibilidade foram feitas para comparar as tendências dos resultados com 180 e 360 dias de tratamento. A razão de risco (HR) com Intervalo de Confiança (IC) de 95% foi utilizada para medir a associação entre covariáveis e desfechos. **Resultados:** Os dados apresentaram que das pessoas diagnosticadas com TB, 83,21% são homens, 59,14% são pretos ou pardos, com idade de 30 a 39 anos (31,50%). Em 270 dias, 75,47% da população abrigada conclui o tratamento, enquanto apenas 39,84% da PSR atinge esse objetivo. Entre os abrigados, o risco de não finalizar o tratamento é maior entre homens (HR 1,13), pretos e pardos (HR 1,17), indígenas (HR 1,90) e pessoas com 70 anos ou mais (HR 2,29). Ao comparar as duas populações, a PSR tem maior risco de não concluir o tratamento, seja pela análise bruta (HR: 3,54, 95% IC 3,18–3,94,  $p < 0.001$ ) ou ajustada por covariáveis (HR 2,04, 95% IC 1,82–2,28,  $p < 0.001$ ). Na PSR, os fatores associados ao insucesso do tratamento da TB foram o uso de drogas ilícitas (HR 1,38, 95% IC 1,09–1,74,  $p = 0.001$ ), doenças mentais (HR 2,12, 95% IC 1,08–4,15,  $p = 0.01$ ) e não receber o Tratamento Diretamente Observado (TDO) (HR 18,37, 95% IC 12,23–27,58,  $p < 0.001$ ). Entre as categorias que compõem o insucesso, a perda por seguimento apresentou associação com esses três fatores na PSR. **Conclusão:** Os achados apontam que a PSR é duas vezes mais suscetível a não concluir o tratamento em comparação às pessoas que não estão nessa condição social. Ademais, os resultados reforçam a relevância de abordagens intersetoriais e integrais para identificar e atuar sobre os determinantes de saúde que afetam a sobrevivência dessa população vulnerável no Brasil.

- **PO30** - *Sobrevivência de pacientes com AIDS: influência do sexo na probabilidade de sobrevivência* – Marcela Silva de Araujo (UFLA), Viviane Costa Silva (UFLA), Davi Barbosa Pereira de Sousa (UFLA), Hellen Sonaly Silva Alves (UFRPE), Marcelo Ângelo Cirillo (UFLA), João Domingos Scalon (UFLA), Guilherme Rocha Moreira (UFRPE)

**Resumo:** A AIDS (Síndrome da Imunodeficiência Adquirida) é uma doença causada pelo vírus da imunodeficiência humana (HIV) caracterizada pelo enfraquecimento do sistema imunológico, aumentando o risco do surgimento de outras doenças oportunistas. Entre os sintomas mais comuns estão febre, fraqueza e diarreia prolongada. Este estudo investiga o tempo de sobrevivência de pacientes com AIDS atendidos no Instituto de Pesquisa Clínica Evandro Chagas (Ipec/Fiocruz) entre 1986 e 2000. Utilizando uma amostra de 193 pacientes, a análise foi conduzida com o objetivo de identificar fatores que influenciam a sobrevivência e avaliar o desempenho de diferentes modelos de sobrevivência. A distribuição de Weibull foi adotada como modelo principal, comparada com outras distribuições como a log-normal e a exponencial. A função de sobrevivência foi estimada por meio do estimador de Kaplan-Meier e a comparação entre grupos foi realizada utilizando o teste de Log-Rank. O modelo de Cox foi empregado para investigar a influência de variáveis explicativas no tempo de sobrevivência. Os resultados revelaram que pacientes do sexo feminino apresentam maior sobrevivência em comparação aos masculinos e que regimes terapêuticos

combinados e potentes são eficazes em aumentar o tempo de vida. Apesar da distribuição log-normal ter demonstrado um melhor ajuste, a Weibull foi mantida como parte da metodologia para explorar diferentes aspectos dos dados. Os resultados obtidos nesse estudo são de alta relevância e de interesse social, destacando a necessidade de políticas públicas que ampliem e reforcem estratégias de prevenção contra o HIV e possível evolução à AIDS.

- **PO31** - *Uma aplicação da família Neyman tipo A odd log-logística generalizada com fração de cura – Valdemiro Piedade Vigas (UFMS), Edwin M. M. Ortega (ESALQ-USP), Rick A. F. Mangueira (UFPI)*

**Resumo:** O presente trabalho propõe uma nova família de modelos de sobrevivência denominada Neyman Tipo A odd log-logística generalizada com fração de cura. Esse modelo se baseia em diferentes esquemas de ativação, nos quais o número de fatores  $M$  segue a distribuição discreta Neyman tipo A, e o tempo até a ocorrência do evento segue uma família de distribuições odd log-logística generalizada. A análise clássica via máxima verossimilhança é utilizada para estimar os parâmetros dos modelos. Diferentes estudos de simulação são adotados para avaliar as médias, os vieses e a raízes dos erros quadráticos médios das estimativas dos modelos nos diferentes esquemas de ativação. Os critérios de seleção do modelo e o teste da razão de verossimilhanças também são apresentados, além de técnicas gráficas, como análise de TTT-Plot e Kaplan-Meier. A aplicação para um conjunto de dados reais ilustra a utilidade dessa família.

- **PO32** - *Modelo de chances proporcionais para dados com censura intervalar – Paulo Cerqueira dos Santos Junior (UFPA), Rogervan de Sousa Soares (UFPA), Vinicius Duarte Lima (UFPA), Valdemiro Piedade Vigas (UFMS)*

**Resumo:** Este trabalho investiga a aplicabilidade e eficácia dos novos modelos de chances proporcionais para dados de sobrevivência com censura intervalar. A problemática abordada reside na necessidade de modelos estatísticos flexíveis para a análise de tempos de sobrevivência, capazes de lidar adequadamente com a censura intervalar, um desafio comum em diversas áreas como medicina, biologia e engenharia. O objetivo deste trabalho consiste em avaliar as vantagens, limitações e potenciais referente aos modelos propostos. A pesquisa inclui uma revisão literária fundamentada, com base principalmente nas obras de Bennett (1983), Gómez, Gallardo e Arnaldo (2018), e Dos Santos Junior e Schneider (2022). Como também, são realizados um estudo de simulação e uma aplicação a dados reais. Os resultados da simulação demonstraram que os modelos propostos proporcionaram estimativas de alta precisão, com viés reduzido e probabilidade de cobertura próxima ao valor nominal (95%) à medida que o tamanho da amostra aumenta. Na aplicação prática, os modelos ajustados aos dados de pacientes hemofílicos com AIDS revelou que o MEPP-CP com cinco intervalos obteve os melhores valores de AICC e HQIC em comparação com outros modelos. De forma geral, portanto, os modelos mostraram um bom ajuste, com boas propriedades e consistências dos estimadores.

- **PO33** - *Power Piecewise Exponential regression model for Dependent Censoring under a Copula-based – Paulo Cerqueira dos Santos Junior (UFPA), José Ailton Nunes de Lima (UFMG), Silvana Schneider (UFRGS)*

**Resumo:** Dependent censoring occur in studies when some censoring mechanism is related to the occurrence of the event of interest, in these cases the traditional assumption

of independent censoring may not be satisfied and can lead to biased inferences. The aim of this paper is to propose the Power Piecewise Exponential (PPE) regression model for survival data with dependent censoring, considering a copula-based. In addition, the consequent purpose is present the estimation of all parameters, including the dependency parameter, and show the identifiability of this proposed method. Accordingly, the dependency structure is adjusted through the Clayton copula function, the marginal distributions are modeled via the Power Piecewise Exponential distribution and the covariates are included like in the Cox proportional hazard model. A simulation study is carried out to verify the proposed method under different association scenarios between lifetimes and dependent censoring times. Furthermore, the simulation results show that the parameters estimates is close to the real values, the Mean Squared Error decrease and coverage probability is close to the nominal level (95%), as the sample size increases. Lastly, the proposed model is considered to analyze a dataset from a study about patients with Human Immunodeficiency Virus type 1 (HIV-1), named ACTG 175, resulting in significative ( $p$ -value  $< 0.05$ ) effects of the covariates age, treatment, CD4 cell count and symptoms.

- **PO34** - *Asymmetric distributions on the unit hypercube: applications in survival analysis* – Leonardo Santos da Cruz (UFV), Roberto Vila Gabriel (UnB), Helton Saulo Bezerra dos Santos (UnB), Felipe Sousa Quintino (UnB), João Victor Monteiro de Andrade (UnB)

**Resumo:** In this work, the main functions used to describe survival time will be addressed, focusing on a new probability distribution derived from a multivariate model based on asymmetric elliptical distributions, called EUGSELL<sub>1</sub>. As an example, we will present the survival function EUGSEN<sub>1</sub> and the accumulated risk function EUGSEN<sub>1</sub>, applied to univariate continuous random variables:

$$\begin{aligned}
 - S_{y_G}(t) &= \frac{2}{\sqrt{2\pi}\sigma} \int_t^\infty e^{-\frac{1}{2} \frac{(y_G - \mu)^2}{\sigma^2}} \cdot \Phi(\lambda(y_G - \mu)) \cdot G'(y_G) dy_G; \\
 - H_{y_G}(t) &= -\ln \left[ \frac{2}{\sqrt{2\pi}\sigma} \int_t^\infty e^{-\frac{1}{2} \frac{(y_G - \mu)^2}{\sigma^2}} \cdot \Phi(\lambda(y_G - \mu)) \cdot G'(y_G) dy_G \right].
 \end{aligned}$$

where,  $\lambda, \mu, \sigma$  are univariate parameters and  $G : (0, 1) \rightarrow \mathbb{R}$  is an invertible, differentiable monotone increasing function.

- **PO35** - *Extended G-Skew-Normal: A New Approach for Asymmetric Modeling in Survival Analysis* – João Victor Monteiro de Andrade (UnB), Roberto Vila Gabriel (UnB), Helton Saulo Bezerra (UnB), Leonardo Santos (UFV), Felipe Sousa Quintino (UnB)

**Resumo:** This work presents the Extended G-skew-normal (EGSN), a new class of asymmetric Gaussian distributions, initially developed in the multivariate context and adapted for univariate Survival Analysis. The adaptation allows a detailed analysis of its properties, followed by a practical application comparing the EGSN with other asymmetric distributions in survival analysis. The objective is to demonstrate how EGSN can offer more accurate and flexible modeling, bringing new solutions to complex problems. The survival and risk functions associated with EGSN are defined by:

$$- S_{y_G}(t) = \frac{2}{\sqrt{2\pi}\sigma} \int_t^\infty e^{-\frac{1}{2} \frac{(y_G - \mu)^2}{\sigma^2}} \cdot \Phi(\lambda(y_G - \mu)) \cdot G'(y_G) dy_G;$$

$$- H_{y_G}(t) = - \ln \left[ \frac{2}{\sqrt{2\pi}\sigma} \int_t^\infty e^{-\frac{1}{2} \frac{(y_G - \mu)^2}{\sigma^2}} \cdot \Phi(\lambda(y_G - \mu)) \cdot G'(y_G) dy_G \right].$$

where  $\lambda$ ,  $\mu$  and  $\sigma$  are parameters that correspond, respectively, to asymmetry, location and scale, and  $G : (0, \infty) \rightarrow \mathbb{R}$  is a monotone increasing, invertible and differentiable function.

- **PO36** - *Análise do tempo até a aquisição de um vínculo empregatício dos doutores titulados no Brasil via regressão de Cox com covariáveis dependentes no tempo* – Stephany Lima de Oliveira (UnB)

**Resumo:** Este trabalho objetiva analisar o tempo até a aquisição de um vínculo empregatício formal de caráter acadêmico de doutores que titularam no Brasil por meio do modelo de regressão de Cox com covariáveis dependentes no tempo. Para a obtenção das informações de titulação e emprego foram utilizadas as bases de dados da Relação Anual de Informações Sociais (RAIS) dos anos 2012 a 2021, Plataforma Sucupira dos anos 2013 a 2021, Cadastro Geral de Empregados e Desempregados (CAGED) dos anos 2013 a 2021. Os resultados obtidos mostraram que não ter vínculo ativo fora da área acadêmica, ter obtido o título de doutor até os 32 anos, ter titulado em programa nota 5 da Capes e ter titulado na grande área de Linguística foram fatores que indicaram um menor tempo até a aquisição de um vínculo formal na área acadêmica.

- **PO37** - *Use of a non-mixture cure fraction gamma model to analyze long-term survival curves in breast cancer patients* – Ryan Novaes Pereira (UNESP), Ricardo Puziol de Oliveira (UNESP)

**Resumo:** Breast cancer is the second most common type of cancer globally, with approximately 2.3 million new cases in 2022, and ranks fourth in mortality, with over 666,000 recorded deaths (Global Cancer Observatory: Cancer Today). Although it has a lower mortality rate compared to other types of cancer, breast cancer remains a significant public health concern, particularly in regions like Asia and Europe, which account for most cases and deaths. Survival analysis models have been employed to understand and predict survival patterns in patients diagnosed with breast cancer. In this context, this study aims to introduce a new approach using a non-mixture cure fraction model based on the Gamma distribution for the lifetimes of 272 patients diagnosed with breast cancer. The data were obtained from the Netherlands Kanker Instituut (NKI) and include, in addition to lifetimes, some prognostic factors (chemotherapy, amputation, stage, etc.), which were used in the regression approach as covariates that may potentially influence the survival outcomes of the patients. This study highlights the importance of improving the precision in monitoring and survival patterns in oncology studies.

- **PO38** - *Análise de Fatores de Risco para Mortalidade em Mulheres Diagnosticadas com Câncer de Mama no Estado de Pernambuco* – Viviane Costa Silva (UFLA), Maria Karolina de Farias Ramos (UEPB), Ana Patricia Bastos Peixoto (UEPB), Luiz Ricardo Nakamura (UFLA), Tiago Almeida de Oliveira (UEPB)

**Resumo:** A análise de sobrevivência é uma área da estatística dedicada a avaliar o tempo até a ocorrência de um evento de interesse, sendo amplamente aplicada em estudos clínicos nos últimos anos. Este estudo tem como objetivo identificar fatores de risco que influenciam o tempo até o óbito em mulheres diagnosticadas com câncer de mama no

estado de Pernambuco, entre 1996 e 2017. A amostra é composta por 2.337 mulheres, das quais 7,06% das observações foram censuradas e 92,94% resultaram em óbito devido ao câncer de mama. Inicialmente, foi aplicada a técnica não-paramétrica de Kaplan-Meier para a análise descritiva dos dados, revelando que o risco de óbito aumenta ao longo do tempo, sendo mais elevado em mulheres diagnosticadas com tumor metastático e em faixa etária de 81 anos ou mais. O teste de logrank indicou diferenças significativas nas curvas de sobrevivência para a maioria dos grupos analisados, exceto entre as faixas etárias “21 a 40 × 41 a 60” e entre os meios de diagnóstico “histologia do tumor primário × citologia” e “histologia da metástase × clínico”. O modelo de regressão paramétrica ajustados pelas covariáveis extensão, faixa etária e meio de diagnóstico, via distribuição Weibull proporcionou o melhor ajuste indicando que a extensão de metástase aumentou o risco de morte em 57,52% em comparação com a extensão localizada, enquanto a falta de informação sobre a extensão aumentou a taxa de falha em 73,6%. Mulheres na faixa etária de 81 anos ou mais apresentaram um risco 71,2% maior de morte em comparação àquelas de 21 a 40 anos. O diagnóstico clínico e a histologia da metástase elevaram o risco de falha em 56,72% e 38,40%, respectivamente, em comparação ao diagnóstico por citologia.

- **PO39** - *Modelando episódios de absenteísmo por DME: uma aplicação em análise de sobrevivência* – Larissa Reis Sousa Santos (UFBA), Verônica Maria Cadena Lima (UFBA), Rosemeire Leovigildo Fiaccone (UFBA)

**Resumo:** O presente trabalho explora a abordagem de dados que lidam com eventos recorrentes sob a perspectiva da área de análise de sobrevivência. Os dados utilizados são provenientes de um estudo longitudinal de coorte fixa, com seguimento de 10 anos, realizado com 2661 servidores públicos com informações sobre múltiplos episódios de afastamento de licença médica relacionadas ao trabalho. Os afastamentos de interesse foram aqueles relacionados aos distúrbios musculoesqueléticos (DME), que compõem um grupo de doenças do sistema osteomuscular que são caracterizados pela inflamação e a degeneração em músculos, nervos, tendões, articulações e cartilagens, acarretando dor e limitação funcional, que podem ter longa duração e ocorrer de forma recorrente. O objetivo é identificar os fatores de riscos associados à incidência dos múltiplos episódios de afastamentos através do modelo Andersen-Gill (AG). Por meio do modelo é possível notar que a taxa de incidência do sexo feminino tem um aumento de aproximadamente 44% em relação ao sexo masculino, ajustado pelas demais covariáveis. Ter nível de escolaridade superior, reduz a taxa em cerca de 28% em relação ao nível inferior, controlando pelas outras variáveis. Considerando as faixas de idades, é possível observar aumento da taxa a medida que a faixa de idade aumenta, utilizando como base trabalhadores de até 30 anos: taxa de 1,51 para a faixa de idade 30 a 40 anos; 3,02 para 40 a 50 anos e 4,08 para maiores de 50 anos. Esses resultados indicam a relação que existe entre a taxa de incidência de episódios de afastamento por DME com as características dos trabalhadores e a influência de cada uma nos episódios.

- **PO40** - *Análise do tempo de recuperação da bolsa de valores da crise de 2020: Modelagem via Exponencial por Partes* – Luciano Santana dos Santos (UFBA), Maristela Dias de Oliveira (UFBA)

**Resumo:** Em 2020, a bolsa de valores do Brasil atingiu sua menor pontuação desde 2017 devido à instabilidade global causada pela pandemia de COVID-19. Nesse período,

as bolsas de valores ao redor do mundo sofreram quedas expressivas, com a bolsa brasileira registrando sua maior queda diária de 14,78%. Em março de 2020, o mecanismo de defesa *circuit breaker* foi acionado para proteger os investidores contra oscilações bruscas, algo que não ocorria desde 2017. O estudo avaliou o comportamento das ações do Índice BOVESPA após o acionamento do mecanismo de defesa, empregando Análise de Sobrevida para analisar o tempo de recuperação das ações após as quedas. Foi constatado que o tempo de recuperação variou ao longo do ano, com o Modelo Exponencial por Partes (MEP) sendo eficaz na modelagem dessa variação. Além disso, a inclusão de covariáveis no modelo destacou que o tempo de recuperação difere entre os setores, evidenciando que determinados setores foram mais impactados pela crise de 2020.

- **PO41** - *On the analysis of labor time and healthcare management via Generalized Gamma zero-inflated cure-rate regression model* – Gleici da Silva Castro Perdoná (USP), Hayala C. C. de Souza (USP), Francisco Louzada (ICMC-USP), Adesina Akintan, Hadiza Abdulaziz Idris, Josaphat Byamugisha, Lawal O. Oyeneying

**Resumo:** Standard survival models assume that time is greater than zero. However, there are several practical examples where a proportion of the data presents times equal to zero. Based on this issue, survival models including this proportion have been proposed and are called zero-inflated survival models. It is important to consider flexible distributions in this context in order to model more complex patterns. In this paper, we propose a Generalized Gamma zero-inflated cure-rate survival model, motivated by the pattern observed in the labor progression times. Based on simulation study, we show that estimation methods (maximum likelihood estimates and asymptotic confidence intervals) present a good performance even for small sample sizes. Concerning the model selection, we verified that Likelihood Ratio Test showed the best results. The proposed model was considered to analyze the labor time of pregnant women from sub-Saharan Africa. For diagnostic analysis, we used Cox-Snell residuals and Local Influence methods. In general, the model showed as an adequate tool to describe the labor and we conclude that this model can be a tool in the study of childbirth times supporting management in obstetrical healthcare. We acknowledge the World Health Organization for granting us permission to use the data set.

- **PO42** - *Modelo de Regressão Log de Sobrevida Proporcional para Dados Discretos na Presença de Censura* – Tiago Chandiona Ernesto Franque (UnB), Eduardo Yoshio Nakano (UnB)

**Resumo:** Um dos modelos de regressão mais populares na análise de dados de sobrevivência é o modelo de riscos proporcionais de Cox, cuja principal característica é considerar que as covariáveis atuam multiplicativamente na função de risco. No entanto, essa característica não pode ser satisfeita quando os tempos de sobrevivência são discretos, devido ao fato da função de risco ser limitada no intervalo (0,1). Neste contexto, este trabalho apresenta o desenvolvimento de um modelo de log de sobrevivências proporcionais como alternativa discreta do modelo de riscos proporcionais. Procedimentos para a verificação da suposição de proporcionalidade do logaritmo da função de sobrevivência foram propostos e inferências dos parâmetros do modelo foram formuladas considerando a distribuição Weibull discreta como distribuição basal dos dados. Estudos de simulação foram realizados para verificar as propriedades assintóticas dos estimadores e o modelo

proposto foi ilustrado por meio de um conjunto de dados sobre o tempo de sobrevivência de pacientes com leucemia.

- **PO43** - *Análise de Confiabilidade em Compressores de Ar Industriais para Manutenção Preditiva* – Marcos Jardel Henriques (UFSCar-USP), Oilson Alberto Gonzatto Junior (ICMC-USP), Éder Silva de Brito (IFG), Edilenia Queiroz Pereira (UFSCar-USP), Francisco Louzada Neto (ICMC-USP)

**Resumo:** Este projeto explora a aplicação de tecnologias IoT e conceitos de confiabilidade estatística para otimizar a manutenção preditiva de compressores de ar industriais. Utilizando sensores IoT, monitoramos variáveis críticas como temperatura do óleo, temperatura do motor e pressão do ar. A modelagem do comportamento do sistema é realizada através de processos de Poisson e análise da intensidade de falhas. A função de intensidade de falhas, descrita pela Lei de Potência, e a Função Cumulativa Média (MCF) são empregadas para avaliar a evolução do sistema e a eficácia dos reparos. A manutenção preventiva e os eventos de atenção são analisados para prever falhas e otimizar o desempenho. Os dados são integrados em um dashboard automatizado, que facilita a visualização e a previsão de falhas, aprimorando a tomada de decisões e permitindo uma manutenção mais eficiente baseada em dados estatísticos.

- **PO44** - *Defective Dagum Distributions Induced by Gamma Frailty for Analyzing Cancer Data with a Cure Fraction* – Hemílio Coelho (UFPB), Vera Tomazella (UFSCar), Cleide Mayra M. Lima (UFPI), Pedro Rafael Marinho (UFPB), Claudio Javier Tablada (UFPB)

**Resumo:** This study presents a model based on defective Dagum distributions combined with Gamma frailty to analyze cancer data that includes a cure fraction. Defective distributions are used to address scenarios where there is a non-zero probability that the event of interest, such as cancer recurrence, may never occur, which is crucial in modeling the cure fraction. These distributions allow the model to appropriately capture the presence of cured patients who are no longer at risk, differentiating them from those who may still experience the event. Gamma frailty, in turn, is introduced to model unobserved heterogeneity among patients, adjusting for individual variation in failure risks. By combining these approaches, the model provides a powerful tool for survival data analysis in populations with a cure fraction, offering a deeper understanding of the factors influencing both cure and disease recurrence. The model is validated with real cancer data, highlighting its applicability and efficiency in clinical settings.

- **PO45** - *Modelo de regressão bivariado via cópulas para dados discretos e censurados* – José Paulo Camolez (Branding Consultoria LTDA), Juliana Betini Fachini-Gomes (UnB), Cira Etheowalda Guevara Otiniano (UnB)

**Resumo:** Ao considerar presença de covariáveis, o presente estudo propôs modelos de regressão bivariados de sobrevivência baseados nas cópulas arquimedianas de Gumbel, Frank e Clayton com marginais Log-Logística discretas para modelar a dependência de dados de sobrevivência bivariados discretos e censurados. Os parâmetros dos modelos foram estimados pelo método de máxima de verossimilhança paramétrica em um único passo. Para estudar o desempenho dos estimadores de máxima verossimilhança, realizou-se uma aplicação a dados de sobrevivência com dois eventos de interesse, o tempo até a adoção da Lei de Zoneamento para Uso e Ocupação de Solo (ZUOS) e o tempo até a

---

adoção da Lei de Zona Especial de Interesse (ZEIS), por meio do *software* R. O modelo final foi definido pelo método *Forward* com covariáveis significativas a nível de 10%. Os resultados confirmam que os modelos são adequados para analisar os dados relacionados à adesão de Políticas Públicas, em que o modelo de regressão bivariado via cópula de Clayton obteve o melhor ajuste no ponto de vista estatístico. Já no ponto de vista político, os modelos de regressão bivariados via cópula de Gumbel e Frank se tornam preferíveis por conter mais informações sobre a adoção da Política ZUOS. Portanto, os modelos de regressão bivariados via cópulas propostos neste trabalho, apresentam-se como uma nova metodologia para analisar dados de sobrevivência com respostas bivariadas discretas, censuradas e na presença de variáveis explicativas.