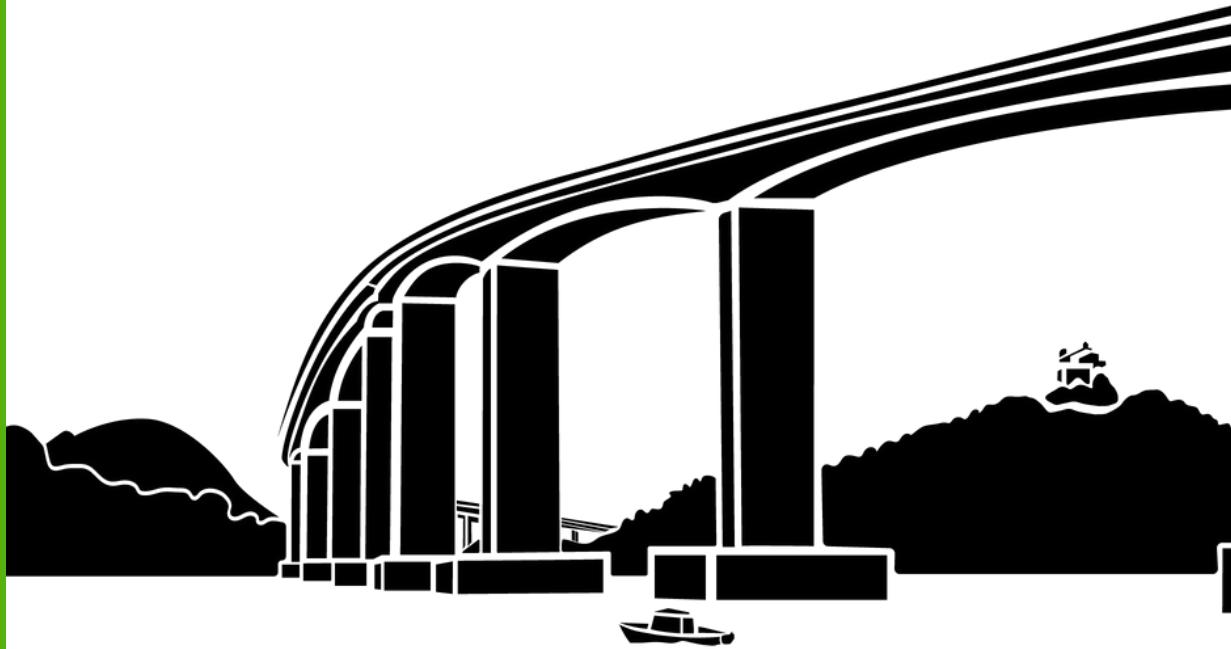


**69^a Reunião Anual da
RBras**
21º SEAGRO

Livro dos Resumos



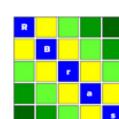
Vitória - ES
4 a 8 de agosto de 2025

VITÓRIA-ES

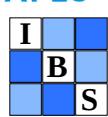
**69^a
2025**

**Reunião Anual da
RBras
21º SEAGRO**

Realização:



Apoio:



Boas Vindas

Seja bem-vindo(a) à 69^a RBras e ao 21º SEAGRO

Caros participantes,

É com grande satisfação que recebemos todos vocês em Vitória-ES para a 69^a Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria (RBras) e o 21º Simpósio de Estatística Aplicada à Experimentação Agronômica (SEAGRO). Este evento, organizado pela RBras em parceria com o Departamento de Estatística da Universidade Federal do Espírito Santo (UFES), reafirma seu papel como uma das principais iniciativas científicas do país nas áreas de Estatística, Biometria e Ciência de Dados.

Ao longo dos últimos meses, a comissão organizadora local, sob coordenação da Profa. Dra. Agatha Sacramento Rodrigues, e a comissão científica, coordenada pela Profa. Dra. Vera Tomazella, trabalharam de forma incansável para planejar e executar cada detalhe deste encontro. O esforço coletivo teve como objetivo proporcionar uma experiência acadêmica rica e acolhedora, que valorize a produção científica, o diálogo interdisciplinar e a troca de conhecimentos entre os participantes.

Neste ano, o tema central do evento — “Estatística e Ciência de Dados a serviço do Desenvolvimento Sustentável” — reflete nosso compromisso com a ciência orientada por evidências e com os desafios globais delineados pela Agenda 2030 da ONU. Acreditamos no papel fundamental da Estatística, da Biometria e da Ciência de Dados na promoção de soluções para problemas complexos como as mudanças climáticas, a segurança alimentar, a equidade de gênero, a erradicação da pobreza e o acesso à educação e saúde de qualidade.

A programação foi cuidadosamente elaborada para incluir conferências, miniconferências, sessões temáticas, apresentações orais e de pôsteres, minicursos e tutoriais, reunindo uma ampla gama de especialistas nacionais e internacionais. Esse ambiente plural favorece não apenas a formação de redes e colaborações, mas também o fortalecimento da nossa comunidade científica. Em nome da diretoria da RBras, expresso minha profunda gratidão a todos os participantes, palestrantes, membros das comissões e avaliadores. Cada contribuição foi fundamental para que este evento se concretizasse com qualidade e entusiasmo. Estendo também nossa gratidão aos patrocinadores e instituições apoiadoras, cujo compromisso com a ciência, a pesquisa e a educação têm sido essenciais para a realização desta edição. Com o apoio de vocês, temos um espaço fértil para a disseminação do conhecimento e o fortalecimento da comunidade científica brasileira e internacional. Desta forma, desejo a todos uma semana produtiva, repleta de reflexões, descobertas e conexões duradouras. Que as interações aqui estabelecidas inspirem novas ideias e impulsionem avanços científicos com impacto social.

Sejam muito bem-vindos à RBras 2025!

Sejam muito bem-vindos a Vitória!

Atenciosamente,

Diogo Francisco Rossoni

Presidente da RBras

Sumário

Minicursos	20
• M1 - Aprendizagem e Inferência Causal: Um Guia Prático	21
• M2 - Obtenção de dados geoespaciais e integração com modelos de machine learning	
22	
• M3 - Machine-Learning for Handling Imbalanced Data: Multi-Class Classification, Anomaly Detection, and Accuracy Metrics in R	23
Conferências	24
• C1 - Are we already FAIR? – The future of data sharing	25
• C2 - Bayesian learning for infectious disease surveillance	26
• C3 - Continuous monitoring of delayed outcomes in basket trials	27
• C4 - Modelagem Estatística de Redes Complexas	28
• C5 - Statistical Pitfalls in Measuring Biological Aging with Epigenetic Clocks: Insights from a Chronic Disease Setting	29
• C6 - Árvore de decisão para identificação dos indicadores determinantes da desnutrição no Brasil	30
• C7 - Influência de Ancestralidade Local na Expressão Gênica de Tipos Celulares da População Brasileira	31
• C8 - Recurrent Event Process Models: change point models and clustering of events	
32	
• C9 - Causal Inference on Flexible Non-mixture Cure Rate Modeling with Piecewise Hazard and Gaussian Process	33
Sessões Temáticas	34
• ST1 - Sessão Fiocruz: Saúde com Dados	35
• ST2 - A Network dos ODS na Embrapa	36
• ST3 - Da Imagem à Decisão: Fenotipagem de alto rendimento e Ciência de Dados no melhoramento de plantas	37
• ST4 - Sessão Butantan: Aplicação de métodos estatísticos para o desenvolvimento de vacinas	38
• ST5 - Riscos climáticos na agricultura: contribuições da Ciência de Dados	39
• ST6 - Dados que Cuidam: Estatística e Inteligência Artificial para a Saúde de Amanhã	40

• ST7 - Sessão Brazilian Journal of Biometrics: Melhores Artigos	41
• ST8 - Aprendizado Estatístico de Máquina e Aplicações	43
• ST9 - Jovens Pesquisadores RBras - Políticas públicas: A formação do profissional em estatística e ciência de dados está de fato o preparando para a tomada de decisões?	44
Mesas Redondas	45
• MR1 - Estatística e Ciência de Dados como Ferramentas para a Cidadania	46
• MR2 - Sessão ABE - Uso de IA em Estatística	47
• MR3 - Publicações Científicas na Era da Inteligência Artificial e Pré-prints: Desafios e Oportunidades	48
• MR4 - Desafios e avanços na fiscalização e regulamentação da profissão de estatístico	49
Minicoferências	50
• Y1 - Processos de contagem e derivações: Uma aplicação para dados ambientais e epidemiológicos	51
• Y2 - Abordagem Bayesiana para a Predição do Estoque de Carbono em Florestas Plantadas	52
• Y3 - Um novo modelo de fração cura com causas concorrentes Geométrica Zero-Inflacionada	53
• Y4 - A Bayesian joint model of multiple longitudinal and categorical outcomes with application to multiple myeloma using permutation-based variable importance .	54
Tutoriais	55
• T1 - Análise Estatística de Desigualdades no Acesso a Água Potável e Saneamento: Explorando os ODS 6 e 10 com Dados do Censo Brasileiro	56
• T2 - Similaridade de Imagens usando deep learning com Python	57
• T3 - Obtendo dados do SIDRA/IBGE com R utilizando o pacote SIDRAR	58
• T4 - Testes de hipóteses em Ensaios Clínicos	59
Comunicações Orais	60
• Análise de Sobrevida	61
– Um novo modelo de tempo de promoção considerando a distribuição de série de potências modificada com uma rede neural para variáveis não estruturadas	62

– Modelo de regressão de fragilidade discreta Waring sob esquemas de ativação aleatória	63
– Modelagem espacial do desmatamento no sul do Amazonas: uma abordagem utilizando análise de sobrevivência	64
– Modelo de regressão odd log-logística generalizada na presença de censura informativa	65
– Modelo de Regressão Weibull Generalizada Restrita para Eventos Competitivos	
66	
– Análise Bayesiana de um modelo de fragilidade discreta baseado na distribuição Hurdle de série de potência zero-modificada: Aplicação em câncer de pulmão	
67	
● Aprendizado de máquina	68
– Active learning in process mining: an active sampling using dimensionality reduction	69
– randomMachines: um modelo de ensemble com máquinas de vetores de suporte	
70	
● Estatística Aplicada à Saúde e ao Meio Ambiente	71
– Modelagem espaço-temporal e previsão hierárquica de séries temporais dos incêndios florestais no Brasil	72
– Desafios na preparação e análise de dados geoespaciais com R: Lições do projeto ICOLMA	73
– Modelação conjunta de respostas binárias longitudinais: aplicação a dados de malnutrição aguda e crônica sob uma abordagem bayesiana não paramétrica	
74	
– Análise de série temporal e previsão para dados de sífilis congênita no estado de Minas Gerais: 2007 a 2024	75
– Comprehensive statistical analysis reveals significant benefits of COVID-19 vaccination in hospitalized patients: propensity score, covariate adjustment, and feature importance by permutation	76
– Selection of objective priors for Gumbel distribution parameters with application to maximum rainfall data	77
● Estatística aplicada em Agronomia e Biologia	78
– Avaliação dos métodos de estimativa de peso e digestibilidade em éguas no terço final da gestação	79
– Estratégias de Amostragem para uma Pesquisa Nacional Agropecuária Brasileira - PNAgro	80
– Índice de seleção MGIDI na avaliação de genótipos superiores de linhaça ..	81
– Dissimilaridade genética e componentes de variância em genótipos de linhaça dourada e marrom	82
– Seleção de Índices de vegetação para predição da produtividade com VSURF	
83	

– Árvores filogenéticas de <i>Bacillus nitratireducens</i> obtidas por modelos probabilísticos evolutivos usando inferência clássica e bayesiana, com marcadores multi-locus (16S rRNA e gyrB)	84
● Métodos estatísticos em Agronomia e Biologia	85
– Parametrizações do modelo de von Bertalanffy para descrição de curvas de crescimento	86
– Determinação da Origem Geográfica de Amostra de Maconha Utilizando SVM e Análise de Isótopos Estáveis: Uma Abordagem de Aprendizado de Máquina 87	
– Análise de sobrevivência com riscos competitivos aplicada à duração da proteção de cultivares	88
– Utilização de modelos não lineares na avaliação do acúmulo de nitrogênio em plantas de chicória sob diferentes formas de cultivo	89
– Aplicação de Técnicas de Tempo até o Evento com Censura Intervalar na Germinação de Sementes	90
– Modelos multi-estados para transições duplas associadas ao parasitismo no controle biológico	91
● Modelos de Regressão, MLG e modelos não lineares	92
– Identificação de Ineficiências de Licitações Públicas por Meio do Modelo de Mistura de Regressões Binomiais Negativas	93
– Overdispersion models for clustered toxicological data in a bioassay of entomopathogenic fungus	94
– Bessel regression and bbreg package for bounded data	95
– Regression models for binary data with scale mixtures of centered skew-normal link functions	96
– Sandwich variance estimation in scale mixture of skew-normal linear mixed models	97
– A robust quantile regression for bounded variables based on the Kumaraswamy Rectangular distribution	98
● Séries Temporais, Análise espacial e Análise espaço-temporal	99
– Bayesian Hierarchical Modeling for Predicting Spatially Correlated Curves in Irregular Domains with Application to PM10 Air Pollution Data	100
– Spatial distribution of multifractal parameters to characterize the formations of <i>Sargassum</i> spp	101
– Terrorism in Cabo Delgado, Mozambique: A Spatio-Temporal Analysis of Fatalities and Their Determinants (2017–2024) with Forecasts to 2030	102
– Zero-Inflated Generalized Poisson GARCH regression models for describing <i>Pseudo-nitzschia</i> in Lisbon Bay	103
– A Parsimonious GAS Model for NDVI Curve Estimation in Soybean Crop Monitoring	104
– Análise de séries temporais para dados de desembarques do <i>Katsuwonus</i> pelamis via SARIMA e modelos híbridos	105

Concursos	106
● Melhor IC	107
– Estimação por máxima verossimilhança em modelos espaciais lineares mistos generalizados baseada na aproximação de Laplace	108
– D-Valor: comparação e aplicação para dados massivos	109
– Avaliação aprimorada da capacidade do processo por meio de modelagem semiparamétrica por partes	110
● Melhor Dissertação de Mestrado	111
– New Generalized Autoregressive Score and Regression Models	112
– Identificação do efeito causal no modelo de mediação com variáveis latentes	113
– Delineamentos ótimos para modelos polinomiais fracionários	114
● Melhor Tese de Doutorado	115
– Functional Data Analysis: Spatial Association of Curves and Irregular Spacing	116
– Multivariate modeling and linear and partially linear mixed models using scale mixtures of centered skew-normal distributions	117
– Essays on Cure Rate Models	118
– Inferência para modelos com respostas censurados usando a distribuição Student-t e skew-t	119
 Pôsters	120
● Séries Temporais	121
– Análise de precipitação máxima anual na região do Matopiba	122
– Modelo Hierárquico de Séries Temporais da Inflação nos Preços de Alimentos no Brasil	123
– Series Predictor: Aplicativo para análise e previsão de séries temporais em shiny	124
– Mortalidade por acidentes de transporte em Minas Gerais: um estudo de séries temporais	125
– Estudo comparativo da qualidade das previsões em séries temporais utilizando as metodologias SSA e GRNN	126
– Imputação múltipla de dados faltantes em séries temporais de poluição atmosférica	127
– Análise da comercialização do etanol hidratado no estado de São Paulo por meio do preço de revenda	128
– Educação de Qualidade: O Papel das Boas Práticas no Ensino Básico	129
– Modelos de séries temporais para dados de contagem: Um estudo comparativo	130

– Influência da Precipitação, Temperatura e os Fenômenos El Niño e La Niña no número de casos de dengue na Cidade de São Paulo	131
– Análise de Ciclos da Exportação de Café no Brasil via Modelos de Decomposição de Séries Temporais (2015-2022)	132
– Modelagem Preditiva de Temperatura: Uma Análise Comparativa para o Município de Juiz de Fora – MG	133
– Previsão do número de suicídios no Brasil a partir de modelos de séries temporais e machine learning	134
– Modelo de Previsão do Preço do Etanol: Influência de Fatores Agrícolas, Energeticos e Macroeconômicos	135
– Padrões de Similaridade da Base do Milho no Brasil: Uma Análise de Agrupamento de Séries Temporais	136
– Evasão feminina no curso de Agronomia da Universidade Federal de Lavras: uma análise temporal	137
– Avaliações Preditivas de Modelos Temporais com Integração de Informações Espaciais: Um Estudo de Incidência da Covid-19 no Estado de São Paulo em 2020	138
– Previsão do PIB do Agronegócio Brasileiro para 2025: Análise Temporal e Modelagem Estatística	139
• Probabilidade e Processos Estocásticos	140
– Comparação do desempenho de métodos de seleção de árvores de contexto	141
– Processos de Poisson não homogêneos de n tipos - Uma possível aplicação no poder Judiciário	142
– Tempo de espera para a ocorrência de palavras em um alfabeto finito	143
• Planejamento do Experimento e Controle Estatístico de Qualidade	144
– Aplicações de Planejamento de Experimentos (DOE) na Engenharia: Uma Revisão de Estudos de Caso e Práticas	145
– Tamanho ótimo de parcela para avaliar a massa de parte aérea de grão-de-bico com base no modelo linear de resposta com platô	146
– Tamanho de amostra para estimação da média das massas fresca e seca da cultivar Trifecta de alfafa	147
– Tamanho de amostra para estimação da média de caracteres produtivos da cultivar Crioula de alfafa	148
– Galpões de avícolas comerciais são parcelas apropriadas para experimentos de nutrição?	149
– Desempenho do algoritmo de troca por coordenada em modelos polinomiais fracionários	150
• Modelos lineares generalizados e modelos não lineares	151
– Estimando incertezas na predição do excesso de mortalidade por COVID-19, a partir do MLG Misto, via simulação	152

– Modelagem e Análise do Processo de Secagem de Grãos de Soja em Diferentes Temperaturas	153
– Comparação de modelos não lineares na descrição da mineralização do nitrogênio no solo com resíduos da produção animal	154
– Análise da influência de características estruturais e variabilidade no mercado imobiliário: uma abordagem GAMLSS	155
– Inferência bayesiana em modelos não lineares: Ajuste do modelo logístico em dose-resposta de herbicidas	156
– Modelagem não linear para a atividade de água no armazenamento de açúcar mascavo	157
– Índice para diagnóstico de superdispersão em dados longitudinais nominais politônicos com estrutura agrupada	158
– Modelagem não linear da variação do parâmetro a^* na coloração do açúcar mascavo durante seu armazenamento	159
– Modelagem não linear do Crescimento de Novilhas alimentadas com 10% de Proteína Bruta	160
– Comparação de Modelos Não Lineares na Análise de Crescimento na Cunicultura	161
– Influência da Temperatura no Desenvolvimento de Pupas de Crisopídeos: Uma Comparação entre Machos e Fêmeas	162
– Temperaturas Extremas e Mortalidade no Rio de Janeiro: Uma Análise com Modelos de Regressão Condisional	163
– Contribuições à Modelagem de Dados Duplamente Limitados: Uma Nova Proposta com Ênfase em Aplicações Hidrológicas	164
– Métodos de Aprendizado de Máquina para Detecção de Quasares de Alto Redshift com Mitigação de Viés de Seleção	165
– Modelagem da taxa de mortalidade por intoxicações e lesões autoprovocadas intencionalmente por indivíduos de 15 a 29 anos no Brasil, 2014 a 2022 ..	166
– Funções e Afilamentos não-lineares com GAMLSS	167
– Previsão de Preços de Carros Usados com Kolmogorov–Arnold Networks: Uma Abordagem para Captura de Relações Não Lineares	168
– Classe Complementar Generalizada Seno Bibaseline	169
● Modelos de Regressão	170
– Preditores da taxa de incidência de Malária em Tete, Moçambique: uma comparação de modelos para dados de contagem com superdispersão	171
– GAMLSS na predição volumétrica de plantações de eucalipto	172
– Uncovering the key factors behind South American football team performance	173
– Análise do peso vivo de alpacas utilizando uma abordagem GAMLSS ..	174
– Análise de influência local via busca progressiva no modelo de regressão heteroscedástico linear simétrico	175
– Custos de Transação Econômico nos Processos de Contratações Públicas: Uma Análise Comparativa de Modelos de Regressões	176

– Estudos das relações entre produto interno bruto per capita e os setores do valor adicionado bruto do estado de mato grosso	177
– Modelagem da inserção informal no mercado de trabalho: uma análise socioeconômica via regressão logística para dados de pesquisas complexas	178
– Antedependence models for nonstationary longitudinal data with centered skew-t distribution in a Bayesian approach	179
– Correlação e regressão linear entre caracteres da cultivar Trifecta de alfafa	180
– Correlação e regressão linear entre caracteres da cultivar Crioula de alfafa	181
– GAMLSS na modelagem de decisões de crédito: uso de dados sintéticos para previsão de aprovações	182
– Avaliação de modelos não lineares na descrição do acúmulo de nitrogênio em plantas de alface	183
– Modelagem estatística do estoque de carbono em florestas plantadas de eucaliptos	184
– Comparação de modelos não lineares com erro assimétrico para o acúmulo de matéria seca do feijoeiro cultivar Bolinha	185
– Ajuste do modelo logístico no acúmulo de matéria seca em frutos de café conilon: identificação de pontos críticos com intervalos de confiança via bootstrap	186
– Modelagem gammss para a descrição de peso vivo de alpacas de raça hucaya	
– Análise de regressão para identificar fatores altistas e baixistas no preço do suíno em Mato Grosso	188
– Incorporating gmrf-based distributional models for analyzing violence against women in Minas Gerais	189
– Impact of social indicators on infant mortality: An approach using distributional models	190
– Universalidades em Sistemas Urbanos: Uma Análise Física com GAMLSS	191
– Análise de tendência temporal via Regressão Segmentada Joinpoint da taxa de incidência da malária nos estados da Amazônia Legal	192
– Outlier Detection in the Circular Regression Model	193
– Unified Beta Regression Model with Random Effects for the Analysis of Sensory Attributes	194
– Modelo de Regressão Quantílica Marshall-Olkin Cosseno Fréchet	195
– Análise de regressão: uma previsão do ENADE 2025 para os cursos de licenciatura em matemática de Pernambuco, Brasil	196
● Inferência Estatística	197
– Estimadores em Forma-Fechada para a Distribuição de Probabilidade Nakagami Bivariada	198
– Análise da Taxa do Medo do Fracasso no Empreendedorismo na América Latina e Caribe: Tendências Regionais e Ajuste de Modelos de Distribuições Probabilísticas (2001–2021)	199

– Estimação via estatísticas do tipo Mellin para distribuição na classe slash modificada	200
– Modelagem de dados de suporte limitado: a distribuição power logit inflacionada em zero e um	201
– Seleção de modelos de captura-recaptura com heterogeneidade temporal e efeito comportamental à marcação	202
● Inferência Bayesiana	203
– Spatial deformation in a Bayesian spatiotemporal model for matrix-variate responses	204
– Bayesian inference for zero-and/or-one augmented unit-gamma	205
– Análise de dados para um experimento com variável resposta ordinal	206
– Bayesian Inference in Stochastic Processes to Identify Mortality Attributed to Sepsis	207
– Modelagem bayesiana de proporções de germinação em Fitopatologia	208
– Determinant point process to clustering poverty data	209
– Generalized Bayesian parametric quantile regression modelling	210
● Geoestatística	211
– Análise espacial para verificar índices socioeconômicos dos municípios de Minas Gerais	212
– Análise de experimentos em blocos casualizados com várias plantas por parcela considerando a dependência espacial	213
– Previsão espacial de eventos de precipitação extrema: uma abordagem bayesiana e geoestatística	214
– Análise Geoestatística do Risco de Fogo em Minas Gerais	215
– Comparação de funções de base radial para interpolação espacial de variáveis categóricas	216
– Agrupamentos espaciais da precipitação pluviométrica no estado do Rio Grande do Sul	217
– Geostatistical estimation and prediction for censored responses using STAN Bayesian approach	218
– Ajuste de distribuições de probabilidade para algumas métricas obtidas do semivariograma	219
– Rodovias Paranaenses: Um estudo exploratório sobre a ocorrência de acidentes	220
– Visualização e análise de dados georreferenciados de soros para acidentes com animais peçonhentos	221
● Estatística Multivariada	222
– Relações lineares entre caracteres de milho	223
– Análise de Agrupamento das mesorregiões do Espírito Santo (2021–2023) baseado nas produções de diferentes culturas	224

– Correlação dos teores de antocianinas e compostos fenólicos totais de alfaves com parâmetros colorimétricos	225
– Seleção de variáveis para modelos de regressão multivariada para dados composicionais caso pessoas com deficiência	226
– Estimativa de Densidade por Programação Quadrática com Restrições de CDF: Uma Abordagem Baseada em SV-Density	227
– Modelagem Estatística da Distribuição de Espécies Florestais na Amazônia com Análise de Componentes Principais	228
– Associações entre Características Morfológicas e Produtivas em Açaizeiros via Análise de Correlação Canônica	229
– Análise Estatística da Composição Química de extrato de café verde	230
● Estatística Computacional	231
– Utilização de redes neurais convolucionais para classificação de imagens SAR	232
– Desenvolvimeto de um Aplicativo Shiny para Mineração de Texto	233
– skewlmm: An R Package for Fitting Skewed and Heavy-Tailed Linear Mixed Models	234
– ExtremeTS: um pacote em linguagem de programação R para determinação do tamanho do bloco de extremos	235
– Automatizando a análise de regressão linear no RStudio: um aplicativo interativo em shiny	236
– Rede bayesiana baseada na distribuição G0I com parâmetros FUZZY para classificação de imagens SAR	237
– Análise de taxas máximas de câmbio do dólar por meio do aplicativo UniMaC no R	238
– GenomicLand: uma ferramenta gratuita e intuitiva para predição e associação genômica baseada em R	239
– Análise Comparativa entre Métodos de Estimação para a Distribuição Maxwell Reparametrizada	240
– Modelos Relacionais e Grafos: Um Estudo Comparativo Sobre Eficiência Computacional	241
– Processamento de Dados de Reabilitação Física e Neurofuncional	242
● Estatística Aplicada em Ciências Sociais Aplicadas, Demografia e Estatísticas Públicas	243
– Cooperative Statistics in Brazil: A Classification Proposal using Machine Learning models Aligned with International Standards	244
– Caracterização da dependência espacial da dengue em Minas Gerais: possíveis causas associadas aos padrões espaciais	245
– INFOVIS: um portal de visualização de dados para a análise da segurança pública na Bahia	246
– Construção de pirâmides etárias para a região de Ouro Preto - um estudo demográfico	247

– Classificação e agrupamento de crimes relacionados à violência contra mulheres: um estudo baseado em Doc2vec	248
– A trajetória dos menores de idade nos Jogos Olímpicos	249
– Modelos de Mistura para Dados Longitudinais Aplicados ao Estudo de Bullying Escolar	250
– Análise estatística do perfil dos estudantes egressos da UFMT após 10 anos da lei de cotas	251
– Perfil sociodemográfico das vítimas fatais de trânsito no Brasil (1997–2023)	252
– Impacto das Mortes por Psicoativos no Espírito Santo: Uma Análise Sociodemográfica (2013–2022)	253
– Apreensões de Drogas e Mortalidade no Espírito Santo: Uma Abordagem Socioespacial	254
– Desigualdades Regionais na Fecundidade Adolescente: um Estudo Longitudinal do Semiárido Brasileiro (2019–2023)	255
– Análise de aspectos qualitativos e sociodemográficos relacionados ao incômodo causado por gases odorantes em ambiente urbano industrializado	256
– Modelagem Espacial das Mortes Violentas Intencionais (MVI) nos Municípios de Pernambuco	257
– Óbitos Classificados com Causas Garbage no Brasil	258
– Aplicação do Modelo Nominal da Teoria da Resposta ao Item (TRI) na Avaliação do SAEB 2023 – 5º Ano	259
– Estudo do Padrão Espacial dos Crimes de Roubo e Furto nas Áreas Urbanas de Cuiabá e Várzea Grande – MT	260
● Estatística Aplicada em Ciências Médicas, Saúde e Meio Ambiente	261
– Detecção de padrões de incidência da malária na província de Tete, Moçambique de 2016 a 2022	262
– Análise temporal entre grupos etários dos atendimentos ambulatoriais do SUS referentes à intoxicação por drogas, medicamentos e substâncias biológicas no Estado do Paraná (2018–2023)	263
– Mudanças Climáticas no Rio Grande do Sul e seus Impactos Econômicos	264
– Confiabilidade de Equipamentos Médicos: Métodos e Aplicações na Segurança Hospitalar	265
– Previsão da ocupação hospitalar usando análise multivariada de séries temporais	266
– Caracterização dos casos de morte materna tardia no Brasil	267
– Construção de Redes Complexas de Interações entre SNPs: Uma Abordagem Computacional e Estatística	268
– Simulações sobre o Impacto de Parâmetros Genéticos e Ambientais na Variabilidade do Risco de Malária	269
– Clusterização para identificação da similaridade temporal e espacial da situação endêmica e epidêmica da dengue nos municípios do Paraná	270
– Análise Estatística dos desastres causados por chuvas intensas na região metropolitana de Pernambuco	271

– Estimação de valores faltantes em dados de poluição do ar utilizando imputação única	272
– Violência e Lesões Autoprovocadas em Adolescentes do Espírito Santo: Uma Análise dos Fatores Associados	273
– Análise de sobrevivência para avaliação dos processos demográficos de árvores da Mata Atlântica Brasileira	274
– Using correspondence analysis to characterize the dynamics between depth and chlorophyll index in a semi-arid reservoir	275
– Análise do comportamento da série de COVID-19 via Autorregressão Quantílica	276
– Características de seca do município de Recife-PE utilizando o Standardised Precipitation-Evapotranspiration Index	277
– Metodologias para Análise de Trajetórias Latentes: Implementação em Software Estatístico	278
– Comparison of spatial statistical models for areal data with outliers	279
– Tendências Climáticas no Rio Grande do Norte: Análise de Dias Secos e Chuvas Intensas entre 1991 e 2020	280
– Estimativa ultrassonográfica do peso fetal em gestantes diabéticas ou com diabetes gestacional	281
– Modelagem de risco climático-financeiro para o setor agrícola via cópulas bivariadas	282
– Irradiância solar extrema e sua relação com variáveis atmosféricas: uma abordagem com teoria dos valores extremos e cópulas	283
– Distribuição Espacial de Focos de Queimadas no Estado do Ceará com Base em Processos Pontuais (2022–2024)	284
– Análise Espacial da Incidência de Dengue em Minas Gerais no Ano de 2024 com Ênfase na Autocorrelação Espacial	285
– Painel de Vigilância da Saúde Materna e Perinatal	286
– Grafos e Fluxos: Redes Fúngicas sob um olhar da Modelagem e Estatística	287
– Avaliação da qualidade da água subterrânea do estado da Paraíba: uma abordagem por modelos hierárquicos	288
– Investigando o comportamento endêmico e epidêmico da dengue no Paraná e sua associação com variáveis climáticas	289
– Séries Temporais Aplicadas no Estudo da Variação Temporal da Temperatura e Precipitação em Estações Climatológicas do Estado de São Paulo, Período: 1970–2023	290
– Análise bivariada da Temperatura máxima e Umidade relativa do ar mínima em Brasília (DF)	291
– Diagnóstico de Pacientes com Alto Risco de Doença Mental por Meio da Análise do Discurso Verbal e Não Verbal	292
– Evaluation of Arterial Blood Pressure Indices in Young Adults with Sickle Cell Anaemia and Haemoglobin AA in Nigeria	293
– Análise da saúde mental na população brasileira: uma abordagem com amostragem complexa	294

– Análise Abrangente do Covid-19 no Município de Teodoro Sampaio/SP: Um estudo retrospectivo e mapeamento epidemiológico setorial	295
– Distribuição Seno Burr XII Unitária: Uma Aplicação à Vulnerabilidade Climática Mundial	296
– Robust spectral discriminant analysis: an M-quantile periodogram approach	
297	
– Estudo Preditivo da Dengue com Base em Séries Temporais Climáticas e Epidemiológicas	298
● Estatística Aplicada em Arquitetura e Urbanismo, Engenharias e Ciências Exatas	
299	
– Análise multivariada dos determinantes da produtividade na produção de ferro manganês alto carbono em uma indústria siderúrgica	300
– Por que tão poucas? Análise do Efeito Tesoura e Leaky Pipeline na carreira de mulheres pesquisadoras na Engenharia	301
● Educação Estatística	302
– Desafios da estatística aplicada na análise dos efeitos da temperatura na toxicidade de microplásticos e cobre em organismos aquáticos	303
– Tradução do livro “R for Data Science” (2 ^a Edição) feita pela comunidade brasileira	304
– Material Didático Interativo para o Ensino e Aprendizado de Distribuições de Probabilidade	305
– Análise do Desempenho Acadêmico dos Estudantes do Curso de Estatística/UFSCar a partir de 2018	306
– Uso de Monte Carlo para a Previsão de Resultados do Campeonato Brasileiro	
307	
– Ciência de dados e Estatística: Uma comparação descritiva dos cursos de graduação em Universidades Brasileiras	308
– Fatores sociais que influenciam as notas de Matemática e suas Tecnologias no ENEM 2023 em Recife, uma modelagem GAMLSS	309
– Visualização de distribuições conjugadas e interpretação da posteriori em termos de área	310
– Paradoxo de Simpson: Aplicação em Dados do Campeonato Brasileiro de Futebol	311
– Análise de fatores que afetam a conclusão dos estudantes do curso de bacharelado em Estatística da Ufes	312
– Grupo StatUFSM: Iniciativas de Ensino e Extensão para Divulgar a Ciência Estatística nas Mídias Sociais	313
– Assessoria Estatística da UFPel: Qualificação de Pesquisas Científicas ..	314
– Um jogo de memória que explora a variabilidade	315
– Leitura e interpretação de gráficos a partir dos níveis de Curcio	316
● Dados Funcionais, Dados em Alta Dimensão e Aprendizado Estatístico de Máquinas	
317	

– Similaridade de Processos Judiciários Utilizando Processamento de Linguagem Natural	318
– Uso de Redes Neurais na Modelagem de Credit Scoring	319
– Comparação de Medidas de Capacidade Preditiva em um Conjunto de Dados com Classes Desbalanceadas	320
– Classificação Textual com Uso de Florestas Aleatórias em Processos Jurídicos	
321	
– Análise Preditiva da Satisfação do Usuário em Telessaúde de um Hospital Universitário	322
– Avaliação da qualidade de sementes por imagens de raio X utilizando uma abordagem comparativa entre classificadores de aprendizado de máquina	
323	
– Multifractal analysis of orbital products classified as <i>Sargassum spp.</i>	324
– Performance of Regularization Techniques in High-Dimensional Linear Mixed Models	325
– Optimizing Robot Soccer Image Classification Using Transfer Learning and Data Augmentation for Deep Learning Experiments	326
– Arquitetura da Luta: O Algoritmo por trás da Era do Vale-tudo (1994–1997)	
327	
– Utilização de Algoritmos de Machine Learning baseado em estrutura de Árvores aplicados a dados SAR	328
– Comparação dos Modelos de Regressão Logística, Boosting e Floresta Aleatória no Problema de Credit Scoring utilizando Dados Simulados	329
– Previsão de preços de insumos utilizados na cadeia do leite com modelo de aprendizado de máquina	330
– Comparação de Modelos de Classificação para Detecção de Fraudes Financeiras	
331	
– TRI multidimensional como decoder explícito em VAEs: retropropagação direta via diferenciação automática	332
– Avaliação de Modelos de Machine Learning para Previsão da Velocidade do Vento na Cidade da Beira – Moçambique	333
– Bayesian Active Learning Powered by Local Entropy and MCMC-Based Predictive Uncertainty	334
– Comparação de Métodos em Aprendizado Semi-Supervisionado para a Classificação de Notícias Falsas	335
– Avaliação de Respostas Dissertativas em Testes Adaptativos Informatizados baseada em Large Language Models	336
– Por que ETFs podem ser mais eficientes que carteiras de ações individuais? Uma análise com aprendizado de máquina	337
– Qualidade Estatística de Dados Sintéticos Gerados por LLMs: Estudo com Base Clínica de Reabilitação Pós-AVC	338
– O Supremo Tribunal Federal (STF) e os Objetivos de Desenvolvimento Sustentável (ODS)	339
• Análise de Sobrevida	340

– Machine learning em Análise de Sobrevivência – aplicações em dados oncológicos	341
– Modelagem Semiparamétrica de Dados de Confiabilidade Usando Aproximações Polinomiais de Funções de Risco	342
– Extended Semiparametric Cox Model: A Cure Fraction Approach for Long-Term Survival	343
– Bayesian Regression Models for Survival Data with Cure Fraction Using Marshall-Olkin Defective Distributions	344
– Reliability analysis of repairable systems considering unobserved heterogeneity and competing risks	345
– Log-normal cure rate model in GAMLSS	346
– Zero-Adjusted Defective Gompertz Regression Model with Gamma Frailty: Applications in Medical Data	347
– Análise de Sobrevivência Aplicada à Longevidade Reprodutiva de Vacas Nelore: Influência de Fatores Ambientais, Produtivos e Genéticos no Terceiro Parto ..	348
– Análise dos fatores associados à demissão de empregados em empresas não públicas no Brasil	349
– Análise de sobrevivência aplicada no processo de melhoramento genético do cafeeiro	350
– Enraizamento de miniestacas de <i>*Eucalyptus*</i> spp: uma aplicação da análise de sobrevivência semiparamétrica	351
– Sobrevida de Pacientes Acometidos Pela COVID-19 na Paraíba e Fatores de Risco	352
– Regressão Inversa Gaussiana Parametrizada na Moda e Dispersão e Aplicações em Análise de Sobrevivência	353
– Aplicação de modelos de análise de sobrevida via aprendizado de máquina em dados de câncer de mama	354
– Parametric Model for Correlated Survival Data in the Presence of Competing Risks and Interval Censored	355
– Análise de sobrevivência aplicada à germinação de sementes expostas à água ativada por plasma	356
– Globally orthogonal parametrisations for log-location-scale models	357
– Modelo de cura com tempo de promoção integrado com algoritmos de classificação de aprendizado de máquina	358
– Survival Prediction in Colorectal Cancer: A Random Forest Approach ..	359
– The Defective ZTP-Gompertz regression models in the presence of competing and complementary risks for modeling long-term survivors	360
– Modelos de Sobrevivência com Fração de Cura e Fragilidade Espacial usando Distribuições Defeituosas: Uma Aplicação em Obstetrícia	361
– Uso de Redes Neurais Recorrentes na Previsão de Precipitação em Séries Temporais para Campina Grande-PB	362
– Adesão de políticas públicas da área da saúde por municípios brasileiros via modelo de fragilidade compartilhada	363

• Estatística Aplicada em Agronomia e Biologia 1	364
– Frequência de mensurações de variáveis meteorológicas para o cálculo da evapotranspiração de referência	365
– Abordagem Bayesiana utilizando Modelos Não Lineares na Predição do Crescimento de Coelhos	366
– Influência de variáveis meteorológicas nas produtividades de caroço, pluma e semente de algodão no estado do Mato Grosso	367
– Aplicação de redes neurais e inteligência artificial para a predição da qualidade do farelo de soja	368
– Impact on genetic gain from assuming different statistical models in generating designs for early generation plant-breeding experiments	369
– Aprimoramento do Mapeamento de Herdabilidades Regionais para Aumento do Poder de Detecção Genômica	370
– Ordem de adição em experimentos de mistura aplicados na composição de silagens para ruminantes	371
– Blocking Factorial Experiment with Unequal Number of Replications for Factors Levels	372
– From Models to Fields: Simulating Beet Productivity Using AI and Statistical Tools	373
– Classificação de produtividade em soja utilizando imagens RGB de VANTs e Redes Neurais	374
– Curva de lactação de ovelhas da raça Santa Inês	375
– Protocolo para contagem automatizada de cápsulas de linhaça utilizando o pacote R <i>pliman</i>	376
– Avanços no melhoramento do linho no Sul do Brasil: Resultados do Programa de Melhoramento NEPEM/UFSC	377
– Modeling <i>Chrysodeixis includens</i> behavior using generalized linear mixed-effects fractional polynomial models	378
– Insights para a delimitação de blocos experimentais com base em atributos do solo em diferentes profundidades	379
– Estudo Temporal da Herdabilidade e Correlação de Índices de Vegetação com Características de Interesse Agronômico em Cana-de-Açúcar	380
– Dimensionamento de parcela ótima na cultura de arroz considerando dependência espacial usando Geoestatística	381
– Fenotipagem digital da relação entre atributos espectrais e compostos bioativos em sementes de linhaça dourada e marrom	382
– Análise de componentes principais aplicada na caracterização agronômica de pimentas do gênero <i>Capsicum</i> para fins ornamentais	383
– Eficácia de derivados salicil-N-hidrazônicos na inibição de <i>Fusarium</i> sp. e <i>Corynespora cassiicola</i>	384
– Ambientônica aplicada à caracterização climática em ensaios de melhoramento de plantas	385

– Análise da mortalidade de <i>Diaphorina citri</i> induzida por <i>Cordyceps fumosorosea</i> : um estudo de funções de ligação para modelos de dose-resposta dicotômica	386
– Análise de Correlação e PCA em Genótipos de Linhaça (<i>Linum usitatissimum</i> L.) Dourada e Marrom	387
– Análise de componentes principais aplicada ao estudo de atributos químicos do solo de diferentes sistemas de produção agropecuária	388
– Adaptabilidade e estabilidade de genótipos de linhaça utilizando análise AMMI e WAAS	389
– Análise da degradabilidade ruminal <i>in vitro</i> por meio do ajuste de modelos não lineares mistos	390
– Comparative analysis of thermal efficiency in solar collectors with spiral polyethylene tubes	391
– Produção de batata-doce de polpa alaranjada em resposta à adubação com esterco de frangos de corte em Moçambique	392
– Efeitos de diferentes concentrações de Poliacrilamida Aniônica sobre propriedades físicas do solo	393
– O Uso da Modelagem para Interpretar o Efeito do Genoma e de Mudanças Climáticas sobre a Dispersão de <i>Coccidioides</i> spp.	394
– Classificação de Doenças em Folhas de Cafeiro com Redes Neurais Convolucionais	395
– Modelagem linear e não linear na estimativa volumétrica de clones de eucaliptos no Semiárido brasileiro	396
● Estatística Aplicada em Agronomia e Biologia 2	397
– Da Agricultura 1.0 à 5.0: A Transformação Estatística no Setor Agrícola	398
– Protocolo de fenotipagem de alto rendimento no melhoramento de batata-doce	399
– Testando se uma folha pertence as espécies amostradas	400
– Teoria de Valores Extremos na análise da taxa de câmbio no Brasil: uma aplicação com o pacote ExtremeTS	401
– Precipitação Extrema e Riscos Climáticos na Agricultura: Um Estudo em Cascavel (PR)	402
– Potencial da Análise Discriminante de Padrões Espectrais na Detecção da Murcha-de-Ceratocystis em Eucaliptos	403
– Comparação de Índices para Seleção de Regiões Genômicas em Estudos de Associação Genômica Amplia	404
– Variações na fonte de Fósforo (P) no desenvolvimento inicial da Altura e do Diâmetro do caule do cafeiro	405
– Visualização Gráfica da Adaptabilidade, Estabilidade e Produtividade de Genótipos de Cana-de-Açúcar com Índices Baseados em Modelos Mistos	406
– Distribuição Espacial de Bovinos no Brasil	407
– Anomalias na produtividade municipal de feijão, milho e soja, em Goiás, de 1992 a 2021, definidas por regressão linear	408

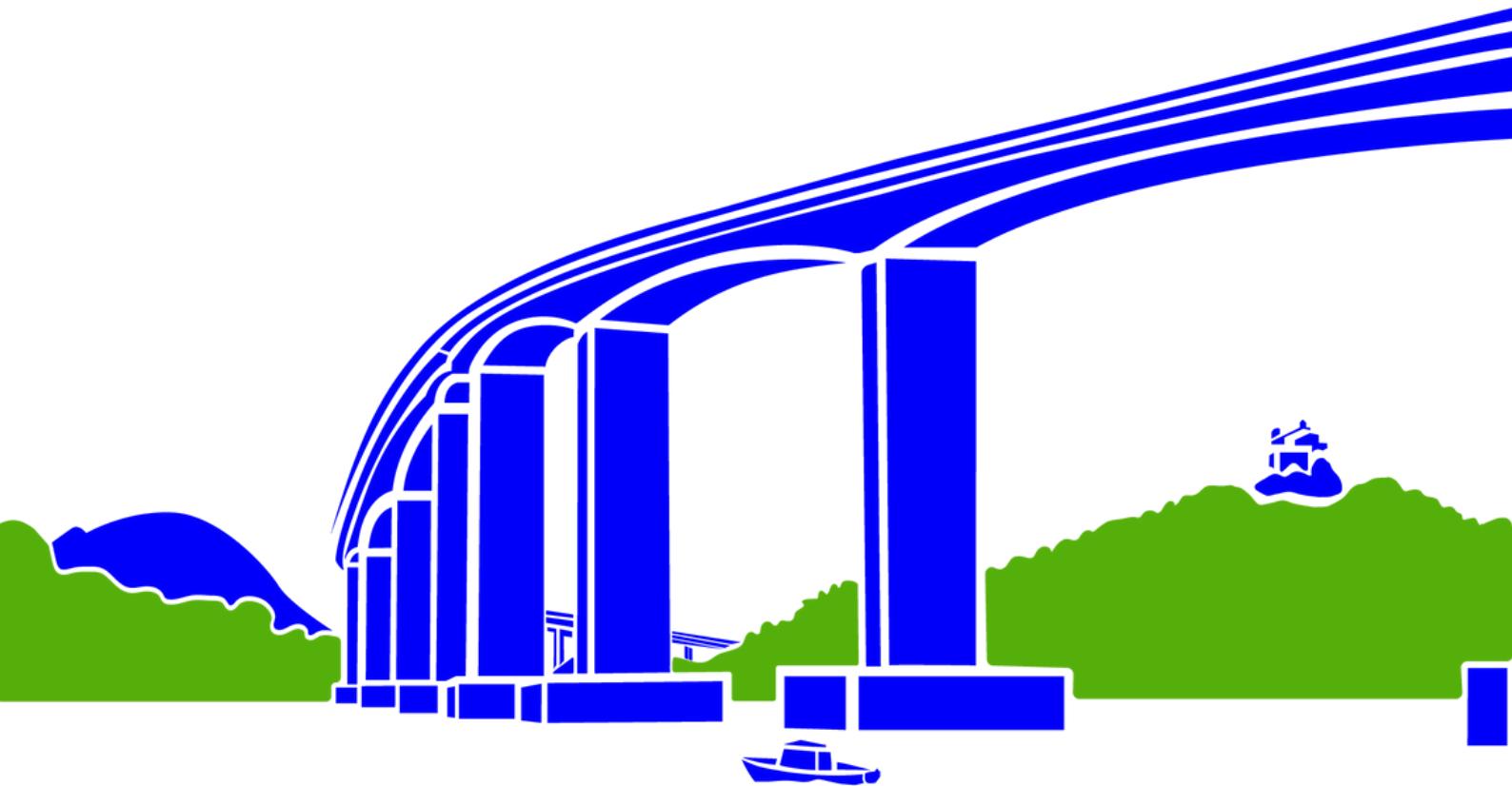
– Modelos de Apenas-Presença: Uma Análise Comparativa de Métodos Estatísticos Computacionais	409
– Comparação de índices de seleção multivariados em ensaios multiambientais	
410	
– Avaliação da podridão branca do capítulo do girassol	411
– Avaliação da maturação de peixes com regressões binárias: explorando novas funções de ligações	412
– Famílias ruins podem originar clones superiores? Análise comparativa da seleção na fase inicial (T1) em cana-de-açúcar	413
– Influência da agregação dos dados na análise de trilha de genótipos de linhaça	
414	
– Comportamento temporal da herdabilidade via imagens multiespectrais de cana-de-açúcar	415
– Democratizando a modelagem de curvas de crescimento com o aplicativo plimanshiny	416
– Dinâmica temporal e diversidade varietal da cana-de-açúcar em Pernambuco/Paraíba/Rio Grande do Norte	417
– Informações espaciais incorporadas à análise de vizinhança para avaliação de famílias de cana-de-açúcar	418
– Métodos de validação de testes para análise de sementes – um aplicativo Shiny	
419	
– Comparação entre CART e RF como métodos de imputação em dados de pecuária de precisão	420
– Modelagem de Extremos Meteorológicos via GEV e GPD: Uma Análise Comparativa em Areado-MG	421
– Análise temporal da área plantada em região semiárida do Nordeste Brasileiro, Brasil	422
– Influência da Precipitação Pluviométrica no Rendimento da Produção de Mandioca no Agreste Pernambucano: Uma Análise Correlacional	423

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**

2025

Minicursos





M1 - Aprendizagem e Inferência Causal: Um Guia Prático

Data: Segunda e terça-feira, 04/08 e 05/08 - 08:00 - 10:00

Palestrantes: Adèle Helena Ribeiro (Universidade de Münster, Alemanha) e Júlia Maria Pavan Soler (IME-USP)

Resumo

Neste minicurso, adotaremos uma abordagem prática para apresentar os conceitos fundamentais para inferência causal a partir de dados observacionais, demonstrando como diferentes técnicas nessa área podem expandir as possibilidades de inferência, indo além da identificação de associações estatísticas (inferências do nível 1), avançando para explicações dos processos de decisão, incluindo a identificação tanto efeitos de intervenções quanto contrafactual (inferências dos níveis 2 e 3, respectivamente). Entre as técnicas que exploraremos estão os algoritmos de aprendizado de estrutura causal, capazes de identificar, ao menos parcialmente, relações de causa efeito entre variáveis. Discutiremos, em particular, o algoritmo Fast Causal Inference (FCI), que se destaca pelo seu rigor metodológico e pela menor dependência de suposições não realistas em comparação com outros métodos de descoberta causal. Notavelmente, o FCI é capaz de corretamente identificar relações causais, mesmo na presença de confundimento latente e viés de seleção. No entanto, assume-se que há um teste estatístico adequado disponível para a identificação das relações de independência condicional entre as variáveis aleatórias envolvidas. Neste contexto, discutiremos os desafios e possíveis soluções para aplicações que envolvem variáveis heterogêneas (quantitativas, categóricas, de contagem, etc.), avaliações longitudinais e a presença de dados faltantes. Além de algoritmos de aprendizado causal, também discutiremos técnicas para identificabilidade e estimativa dos efeitos, apresentando um pipeline de inferência causal inteiramente baseado em dados observacionais.

Apoio Recebido: O presente trabalho foi realizado com apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brasil. Processo nº 2022/11963-3. As opiniões, hipóteses e conclusões ou recomendações expressas nesse material são de responsabilidade dos autores e não necessariamente refletem a visão da FAPESP.



M2 - Obtenção de dados geoespaciais e integração com modelos de machine learning

Data: Segunda e terça-feira, 04/08 e 05/08 - 08:00 - 10:00

Palestrantes: Deoclecio Jardim Amorim (USP) e Maria Gabriella da Silva Araújo (USP)

Resumo

O curso oferece uma introdução aos princípios fundamentais da teoria do geoprocessamento e dos Sistemas de Informações Geográficas (SIG). Serão abordadas as principais bases de dados geoespaciais ambientais, além da aplicabilidade, importância e desafios na utilização desses dados em diferentes contextos. Os alunos aprenderão a obter e manipular dados geoespaciais de maneira eficaz, combinando habilidades de manipulação de dados e programação para realizar análises robustas. As técnicas apresentadas serão aplicáveis em áreas como monitoramento ambiental, planejamento territorial, gestão de recursos naturais, estudos de mudanças climáticas, análise de padrões espaciais e modelagem de cenários futuros, promovendo uma visão integrada para tomada de decisões em contextos científicos e de gestão ambiental. Portanto, objetiva-se introduzir os conceitos iniciais referente a coleta de dados geoespaciais das principais bases de dados, envolvendo a aquisição de dados, organização e integração de dados e visualização utilizando linguagem de programação R. A abordagem adotada busca equilibrar teoria e prática, proporcionando uma compreensão abrangente e aplicada dos métodos de coleta e análise de dados.



M3 - Machine-Learning for Handling Imbalanced Data: Multi-Class Classification, Anomaly Detection, and Accuracy Metrics in R

Data: Segunda e terça-feira, 04/08 e 05/08 - 08:00 - 10:00

Palestrantes: Niamh Mimnagh (Maynooth University, Ireland) e Darshana Jayajumari (Maynooth University, Ireland)

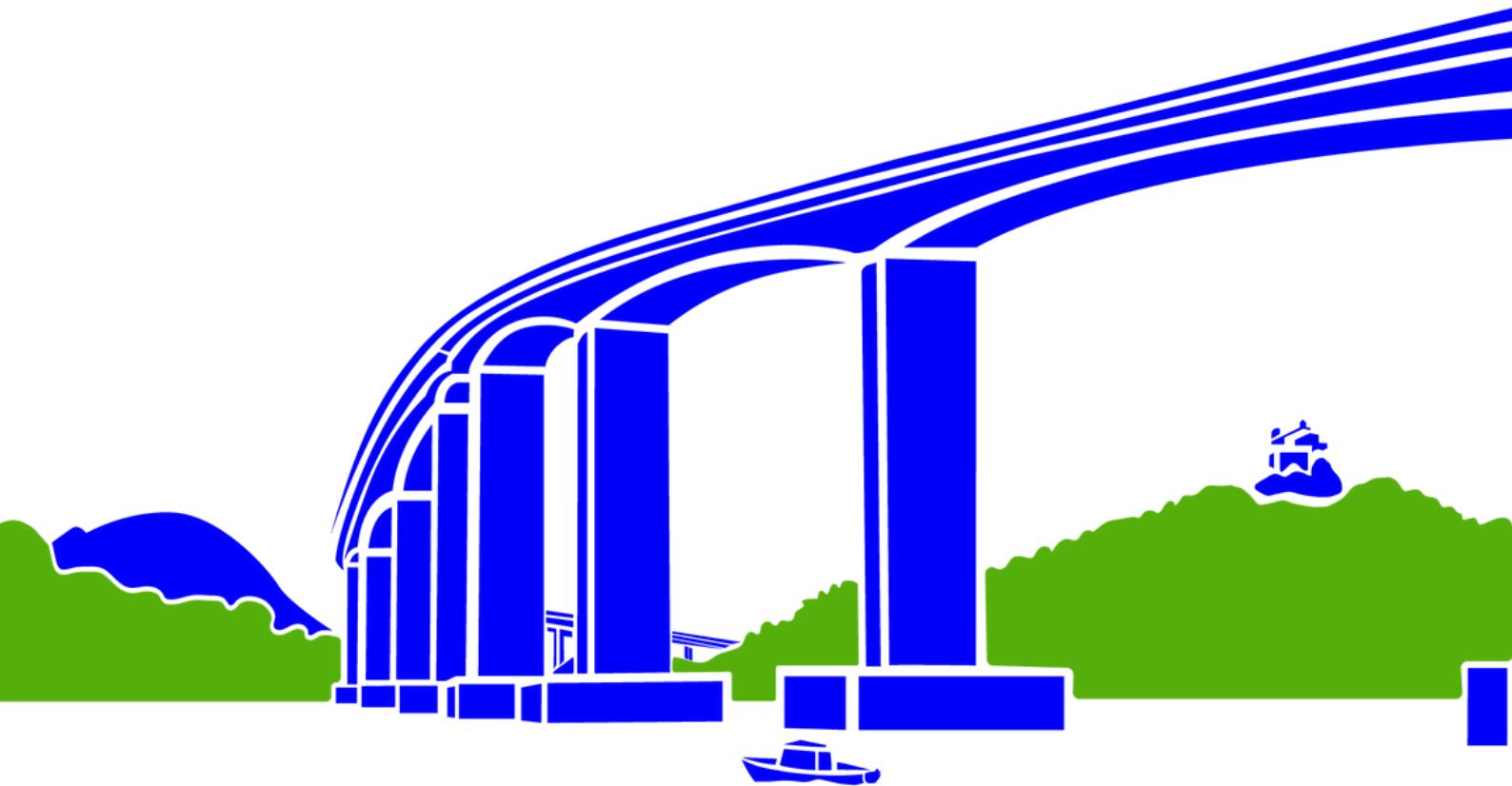
Resumo

This short course provides a comprehensive introduction to handling imbalanced data in machine learning, a critical challenge in fields like fraud detection, medical diagnosis, and anomaly detection. Participants will explore effective techniques to address class imbalance, focusing on resampling methods (both oversampling and undersampling), weighted classification, and multi-class classification strategies. The course will also cover anomaly detection approaches for rare event prediction, such as Local Outlier Factors, Isolation Forests, Angle-Based Outlier Detectors and Minimum Covariance Determinants, as well as performance metrics tailored for imbalanced datasets, such as Precision-Recall curves, F1 score, and AUC. Through hands-on case studies and examples in R, attendees will gain practical experience in implementing these techniques and evaluating model performance in real-world scenarios. The course is designed for data scientists and analysts seeking to build robust predictive models when faced with skewed data distributions, offering insights into both theory and practical applications.

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Conferências





C1 - Are we already FAIR? – The future of data sharing

Data: Segunda-feira, 04/08 - 11:10 - 12:00

Apresentadores: Iris Pigeot (Leibniz Institute, Germany)

Resumo

'Data sharing' is becoming increasingly important in terms of efficient use of resources. In 2007, the OECD called for easy access to research data for the scientific community. In 2016, the FAIR principles (findable, accessible, interoperable, reusable) for research data were published. In 2018, the German government decided to establish a National Research Data Infrastructure (NFDI) where NFDI4Health takes responsibility for personal health data. This talk will present the infrastructures that have been realized so far and discuss potential (statistical) hurdles by giving illustrative examples. Further European developments such as the European Health Data Space will be addressed.



C2 - Bayesian learning for infectious disease surveillance

Data: Segunda-feira, 04/08 - 15:35 - 16:30

Apresentadores: Leonardo Soares Bastos (FIOCRUZ)

Resumo

Public health decision-makers need to decide when to act based on expected or predicted case numbers and to act as quickly as possible based on the current situation of a particular infectious disease to minimize worst outcomes, such as deaths. As part of the epidemiological surveillance of infectious diseases, it is necessary to understand disease dynamics, often with limited data. Bayesian learning is a powerful ally in epidemiological surveillance due to its flexibility in combining expert knowledge and data in complex statistical models. In this talk, I will present problems and proposed solutions related to real-time infectious disease surveillance of arboviruses (e.g. dengue and chikungunya) in Brazil, severe cases of acute respiratory viruses (e.g. Influenza, COVID-19, RSV) in Brazil, and measles reported cases in the US.



C3 - Continuous monitoring of delayed outcomes in basket trials

Data: Terça-feira, 05/08 - 14:00 - 15:00

Apresentadores: Márcio Diniz (Mount Sinai, NY)

Resumo

Precision medicine has led to a paradigm shift, enabling the development of targeted drugs that are agnostic to the tumor location. In this context, basket trials aim to identify which tumor types – or baskets – benefit from the targeted therapy among patients with the same molecular marker or mutation. We propose a continuous monitoring framework for basket trials to increase the likelihood of early identification of non-promising baskets. Although existent Bayesian trial designs allow multiple interim analyses, most are computationally intensive and do not address delayed outcomes that are typical for targeted treatments such as immunotherapies. To address these limitations, we leverage the Bayesian empirical approach proposed by Fujiwara et al. that has low computational cost, and built upon the ideas of Cai et al to handle multiple interim analyses via multiple imputation. We compare operating characteristics of five different strategies for handling delayed outcomes in an extensive simulation study, using the strategy where trial accrual is paused until complete data is observed as a benchmark. Practical recommendations for designing basket trials with continuous monitoring using a delayed outcome are provided.



C4 - Modelagem Estatística de Redes Complexas

Data: Terça-feira, 05/08 - 15:35 - 16:30

Apresentadores: Andressa Cerqueira (UFSCar) **Resumo**

Redes complexas têm recebido uma grande atenção da comunidade estatística, especialmente no contexto de analisar e descrever as interações de sistemas complexos aleatórios. Redes estão por toda parte, desde redes sociais até redes biológicas e sistemas de transporte. No entanto, compreender e analisar essas redes complexas é um desafio científico. Nesta palestra discutiremos como técnicas estatísticas nos permitem desvendar os padrões subjacentes, identificar comunidades, medir a robustez e prever o comportamento futuro das redes complexas.



C5 - Statistical Pitfalls in Measuring Biological Aging with Epigenetic Clocks: Insights from a Chronic Disease Setting

Data: Quarta-feira, 06/08 - 14:00 - 15:00

Apresentadores: Fernanda Schumacher (Ohio State University, US)

Resumo

The third goal of the United Nations 2030 Agenda for Sustainable Development is to “Ensure healthy lives and promote well-being for all at all ages”. Its targets are a good representation of how vast health care is, covering maternal mortality, epidemics of infectious diseases, premature mortality from chronic diseases, prevention and treatment of substance abuse, access to sexual and reproductive health care, and affordable health care, just to name a few. In addition to the essential and undeniable role of statisticians in measuring all the diverse indicators related to this goal, how can statisticians contribute to reaching the targets? In this talk, I will focus on statistical challenges faced when working with chronic diseases, with particular emphasis on people with Multiple Sclerosis (MS). Beyond chronological age, biological age reflects the cumulative damage to a person’s cells over time and is increasingly recognized as a critical factor in understanding health disparities and the progression of chronic conditions. Several biomarkers have recently been proposed to measure biological age and assess the cumulative burden of aging, which directly affect not only life expectancy but also quality of life, especially for people with debilitating diseases. As one of the popular markers, epigenetic “clocks” are based on statistical and machine learning tools to detect DNA methylation (DNAm) patterns. Epigenetic modifications represent a reversible mechanism in regulating the function of the genome without altering the underlying DNA sequence and have been linked to aging through several factors, allowing DNAm to be affected by environmental exposures and lifestyle habits. Since the first proposed epigenetic clock model in 2011, multiple epigenetic clocks have been reported with increasing accuracy, precision, and broader application prospects in aging research. Still, they are based on regression coefficients determined on general training populations, which are then used for out-of-sample prediction. Commonly, the predictions are then regressed on chronological age and technical variables, and the corresponding residuals are called Epigenetic Age Acceleration (EAA), which are then often used for hypothesis testing. The lack of interval estimates for individual predictions, the availability of several algorithms, the lack of a gold standard measure of biological age, and the use of prediction-based inference add to the statistical challenges of these markers. This talk will discuss such challenges, illustrated using data from a clinical study on biological aging in people with MS. I will also discuss the role of statisticians in ensuring that such issues are properly considered, especially when such measures could evolve to be outcomes in clinical trials for anti-aging treatment and the prevention of chronic diseases progression.



C6 - Árvore de decisão para identificação dos indicadores determinantes da desnutrição no Brasil

Data: Quarta-feira, 06/08 - 15:35 - 16:30

Apresentadores: Alcione Miranda dos Santos (UFMA)

Resumo

A desnutrição infantil (DI) é caracterizada pela deficiência de nutrientes básicos que culminam em déficits de peso e estatura, sendo considerado um problema de saúde pública nos países de média e baixa renda. A partir do sucesso no alcance de muitas metas dos Objetivos de Desenvolvimento do Milênio (ODM), foi instituída a Agenda 2030, com 17 Objetivos do Desenvolvimento Sustentável (ODS). Uma das metas internacionais é, até 2030, acabar com todas as formas de desnutrição, e até 2025 atingir as metas internacionais sobre atraso no crescimento (reduzir 40%) e baixo peso (reduzir e manter <5%) em crianças menores de 5 anos. Dessa forma, torna-se relevante o desenvolvimento de estudos que busquem identificar e monitorar os fatores determinantes da desnutrição, a fim de avaliar as intervenções necessárias que possam contribuir para o alcance das metas. Nesta palestra, apresentaremos os principais indicadores que determinam a ocorrência da desnutrição no país, utilizando a técnica de árvore de decisão. Todos os indicadores determinantes da DI estão relacionados à pobreza e às desigualdades sociais, dificultando o alcance da meta para a DI no país.



C7 - Influência de Ancestralidade Local na Expressão Gênica de Tipos Celulares da População Brasileira

Data: Quinta-feira, 07/08 - 14:00 - 15:00

Apresentadores: Benilton de Sá Carvalho (UNICAMP)

Resumo

A diversidade genética da população brasileira, resultante de um complexo histórico de migração e miscigenação, oferece uma oportunidade única para estudar como a ancestralidade genética influencia a expressão gênica. Este estudo propõe uma análise integrada de dados de RNA-seq de célula única (scRNA-seq) e de sequenciamento do genoma completo (WGS) para investigar a relação entre ancestralidade local e expressão diferencial de genes em diferentes tipos celulares. Neste trabalho, serão discutidos métodos de pré-processamento de dados de scRNA-seq, incluindo a filtragem de células de baixa qualidade, normalização, remoção de efeitos de lote e anotação de tipos celulares. Paralelamente, serão também apresentados métodos de pré-processamento de dados de sequenciamento de genoma completo, genotipagem e inferência de ancestralidade local. Um modelo de regressão para respostas do tipo binomial negativa será discutido e utilizado na integração de ambas modalidades de dados, permitindo a comparação da expressão gênica entre os grupos de diferentes ancestrais locais. Extensões que permitam o uso de variáveis medidas com erros e também o uso de observações correlacionadas serão discutidas. Esta estratégia permitirá a identificação de genes cuja expressão é significativamente influenciada pela ancestralidade local em diferentes tipos celulares. Esses achados podem fornecer novos insights sobre as bases genéticas das diferenças fenotípicas observadas na população brasileira e contribuir para a compreensão das interações entre genética e ambiente. Este estudo demonstra uma abordagem inovadora e integrada para explorar a influência da ancestralidade genética na expressão gênica em nível celular, utilizando dados de scRNA-seq e WGS. A aplicação dessas metodologias pode abrir novas perspectivas para pesquisas em genética populacional e medicina personalizada, especialmente em populações geneticamente diversas como a brasileira.

Financiamento: BRAINN/FAPESP 2013/07559-3



C8 - Recurrent Event Process Models: change point models and clustering of events

Data: Quinta-feira, 07/08 - 15:35 - 16:30

Apresentadores: Elizabeth Juarez-Colunga (University of Colorado, US)

Resumo

Recurrent event data arise when an event may occur repeatedly over time. Examples include recurrence of bladder cancer tumors, epileptic seizures, or pulmonary exacerbations. This talk will main discuss two projects. The first focuses on modeling pulmonary exacerbations and their relationship to a longitudinal binary outcome, and the second aims to understand the clustering of events within individuals. The first project was motivated by a study of cystic fibrosis, a hereditary lung disease characterized by progressive loss of lung function. Chronic Pseudomonas aeruginosa (PA) infection is associated with worse clinical outcomes, including more frequent pulmonary exacerbations (PE). The longitudinal progression of PA infection and recurrent PE events are likely intrinsically linked, but their temporal interrelationship has not been fully characterized. It is known that the rate of PA progression increases as individuals age, with potential sharp changes in its trajectory. Using data from the Early Pseudomonas Infection Control Observational Study, we propose a joint model to examine longitudinal PA and recurrent PE events. This model incorporates individual-specific random effects in the longitudinal sub-model, linked to those in the recurrent event sub-model. The longitudinal sub-model includes two change points to represent sharp changes in the trajectory, while the recurrent event sub-model employs a counting process for recurrent events and accommodates delayed entry. The results indicate that children experience a modest increase of 5.13% per year in the odds of PA starting at age 6.9, followed by a more pronounced rise of 27.12% around age 14.5. Additionally, an increased probability of PA is associated with a higher risk of experiencing subsequent PE events. The second project focuses on epileptic seizures, with the primary goal of understanding the clustering of seizures within individuals. We model clustering using a self-exciting stochastic process.



C9 - Causal Inference on Flexible Non-mixture Cure Rate Modeling with Piecewise Hazard and Gaussian Process

Data: Quinta-feira, 07/08 - 17:05 - 18:00

Apresentadores: Dipak Dey (University of Connecticut, US)

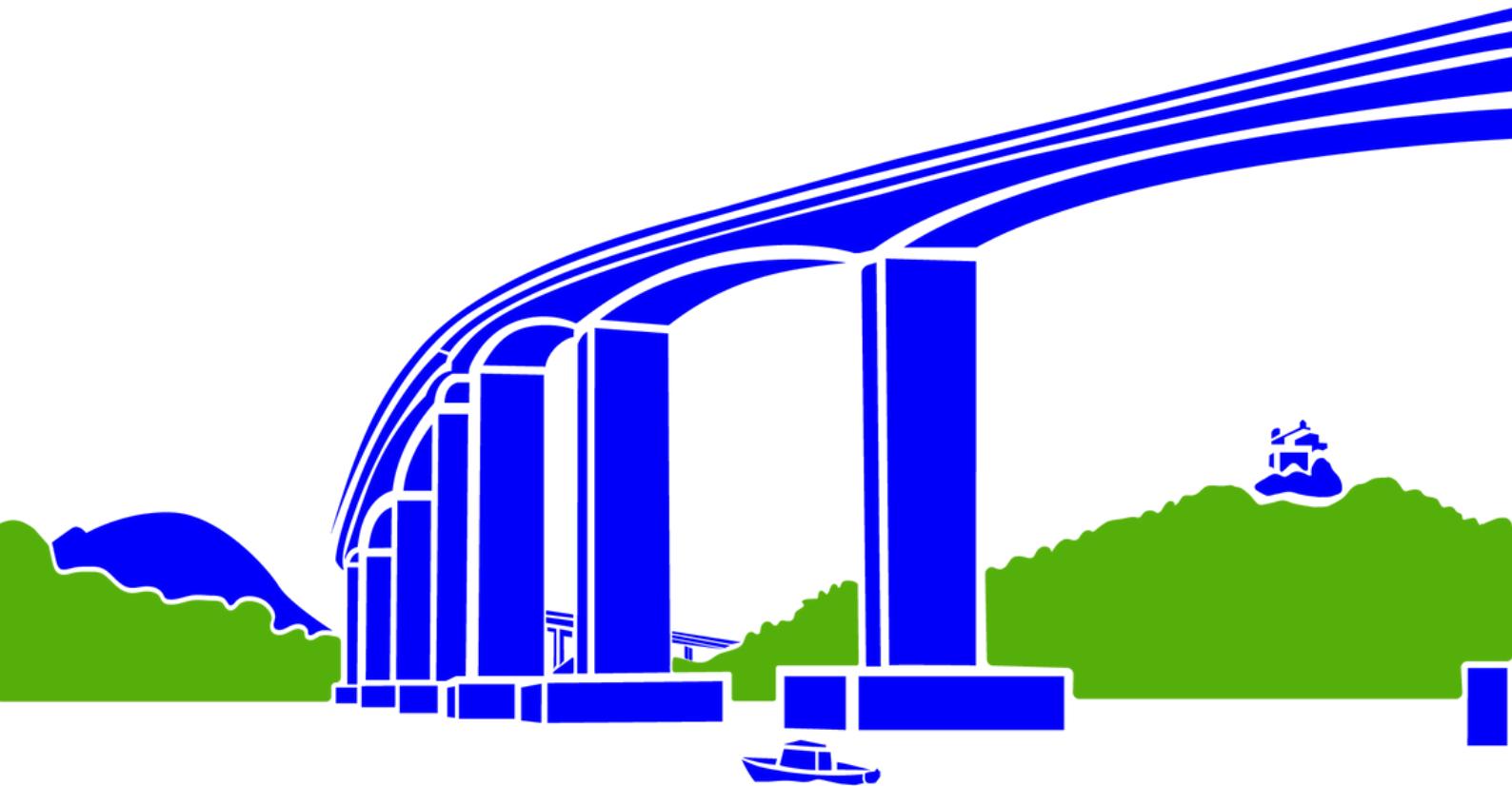
Resumo

In the field of oncology, survival analysis often requires the inclusion of a cure fraction to account for individuals who are effectively cured of their disease. The concept of a cure fraction in survival analysis was introduced in a study examining long term survival following cancer therapy. The early work laid the foundation for defining cured individuals who were not at risk of experiencing the cancer recurrence after a certain period. Recently, there has been increasing interest in semiparametric mixture cure models, which relax some parametric assumptions. The non-mixture cure rate model, which is another branch of the cure rate modeling, represents a significant advance in cure rate modeling. The non-mixture cure rate model, unlike the traditional mixture cure rate model, addresses this need by introducing a latent variable, often interpreted as the unobserved count of cancer cells, to indirectly estimate an individual's cure status. This latent factor approach offers several benefits, including the flexibility to integrate a proportional hazards structure and enhanced computational efficiency. Over time, the non-mixture cure rate model has been extensively extended to handle complex data, leveraging semiparametric methods for modeling survival function. The non-mixture cure rate model typically incorporates covariates into the cure rate parameter through a log-linear form, assuming a Poisson distribution for the unobserved cancer cell count. Additionally, the common use of a linear functional form for covariate effects can be restrictive, particularly for continuous covariates that often display nonlinear relationships. In modeling cure rates, it is often assumed that the effects of continuous covariates vary smoothly over their domain. However, the exact relationship between these covariates and the event of interest is not typically known a priori and may exhibit complex, nonlinear patterns. To flexibly capture these nonlinear covariate effects, we impose a Gaussian Process prior over the effects of the continuous covariates. In this presentation, we consider non-mixture cure rate models in presence of Gaussian Process and further develop causality approach to decide the order of treatment procedures. The methodology is exemplified on a breast cancer study.

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Sessões Temáticas





ST1 - Sessão Fiocruz: Saúde com Dados

Data/hora: Segunda-feira, 04/08 - 17:05 - 18:35

Apresentadores: Elisângela Rodrigues (Chair, UFC/Cidacs-Fiocruz-BA), Maria Yury Travassos Ichihara (Cidacs/Fiocruz-BA), Laís Picinini Freitas (PROCC - Fiocruz-RJ) e Raphael de Freitas Saldanha (ICICT - Fiocruz-RJ)

Resumo

Na sessão temática “Fiocruz: Saúde com Dados”, destacamos avanços e aplicações na análise de dados voltados para a saúde pública, apresentando abordagens inovadoras em modelagem estatística e estudos sobre desigualdades sociais. Na palestra “Modelos Bayesianos espaço-temporais para dados de vigilância: distribuição, fatores associados, e predição de epidemias de arboviroses”, foi abordada a utilização de modelos estatísticos bayesianos para analisar padrões espaço-temporais de epidemias, como dengue, zika e chikungunya. Os pesquisadores exploraram a capacidade desses modelos de identificar fatores de risco, prever surtos e auxiliar na alocação de recursos para vigilância epidemiológica.

Já a palestra “Desigualdades sociais em saúde e impacto de políticas sociais no Brasil: evidências produzidas no Cidacs” trouxe à tona os resultados de estudos realizados pelo Centro de Integração de Dados e Conhecimentos para Saúde (Cidacs). As análises revelaram que esses estudos vêm produzindo evidências robustas sobre desigualdades sociais e avaliação de políticas sociais a partir do uso de grandes bases de dados integradas e o desenvolvimento de tecnologias de Big Data. Serão apresentadas as evidências que foram produzidas pelos pesquisadores do Cidacs em relação à saúde materno-infantil, doenças infecciosas, nutrição e doenças crônicas não transmissíveis como câncer, por exemplo.

Na terceira palestra, “Democratização de acesso a dados de saúde, ambiente e clima com o R: a experiência do LIS/ICICT/Fiocruz”, destacamos o uso da linguagem de programação R como ferramenta para facilitar o acesso a dados de saúde e clima, promovendo a transparência e o uso de informações em pesquisas e políticas públicas. A combinação dessas temáticas ilustra como a estatística e a ciência de dados podem ser uma ferramenta poderosa para enfrentar desafios complexos de saúde pública, desde a previsão de epidemias até a promoção de equidade no acesso à saúde, além de reforçar a importância de dados e tecnologias como bases para ações mais eficazes e equitativas em saúde pública.

Palestras:

1. **Maria Yury Travassos Ichihara (Cidacs/Fiocruz-BA)** - Desigualdades sociais em saúde e impacto de políticas sociais no Brasil: evidências produzidas no Cidacs
2. **Laís Picinini Freitas (PROCC – Fiocruz-RJ)** – Modelos Bayesianos espaço-temporais para dados de vigilância: distribuição, fatores associados, e predição de epidemias de arboviroses
3. **Raphael de Freitas Saldanha (ICICT – Fiocruz-RJ)** – Democratização de acesso a dados de saúde, ambiente e clima com o R: a experiência do LIS/ICICT/Fiocruz



ST2 - A Network dos ODS na Embrapa

Data/hora: Segunda-feira, 04/08 - 17:05 - 18:35

Apresentadores: Waldomiro Barioni Júnior (Chair, Embrapa Pecuária Sudeste), Betúlia de Morais Souto (Embrapa Sede), Wilson Tadeu Lopes da Silva (Embrapa Instrumentação) e Daniel de Castro Victoria (Embrapa Agricultura Digital)

Resumo

A Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA, vinculada ao Ministério da Agricultura, Pecuária e Abastecimento – MAPA, tem como missão “Viabilizar soluções de pesquisa, desenvolvimento e inovação para a sustentabilidade da agricultura, em benefício da sociedade brasileira”. A EMBRAPA implantou na empresa a política vinculada a Agenda 2030 da ONU que procura alinhar as suas ações de pesquisa a cada um dos 17 ODS de forma a promover a agropecuária brasileira como um importante setor para o protagonismo do desenvolvimento humano, econômico e ambiental, com foco na segurança alimentar e em alimentos seguros, produzidos com responsabilidade social e saúde única. A Embrapa instituiu a sua Rede ODS com olhar corporativo, tanto estratégico como operacional. Durante a sessão será feito uma abordagem geral sobre os ODS, suas metas e indicadores, com destaque de alguns casos de sucesso, de tecnologias e ações, desenvolvidas pela Embrapa, associadas a um ou mais ODS. Serão discutidas, também, algumas estatísticas dos ODS a nível mundial. Bem como, de que maneira a ciência de dados pode auxiliar no cumprimento das metas.

Palestras:

1. **Betúlia de Morais Souto (Embrapa Sede)** – Implantação da rede ODS na Embrapa, metas, indicadores e casos de sucesso
2. **Wilson Tadeu Lopes da Silva (Embrapa Instrumentação)** – Dados e informações atuais do Saneamento rural e ações para a universalização
3. **Daniel de Castro Victoria (Embrapa Agricultura Digital)** – Gestão e processamento dos dados para geração do Zoneamento Agrícola de Risco Climático (ZARC)



ST3 - Da Imagem à Decisão: Fenotipagem de alto rendimento e Ciência de Dados no melhoramento de plantas

Data/hora: Terça-feira, 05/08 - 10:30 - 12:00

Apresentadores: Luiz Alexandre Peternelli (Chair, UFV), Maicon Nardino (UFV) e Tiago Olivoto (UFSC)

Resumo

O aumento populacional gera a necessidade de aumento da oferta de alimentos dos mais diversos. Em especial, em se tratando de alimentos de origem vegetal, destacam-se diversas culturas como o milho, a soja, o trigo e a cana-de-açúcar. Especificamente, o trigo é a base para a produção de pães e massas alimentícias, enquanto que a cana-de-açúcar contribui tanto para a produção de açúcar quanto para etanol. O aumento da produção alimentícia sem maiores danos ou invasões de áreas é possível quando se pensa no aumento da produtividade, de modo mais eficiente e econômico. Nesse sentido podemos falar sobre o melhoramento como forma de melhorar a oferta de alimentos sem ter que aumentar a área cultivada, mas sim melhorando a produtividade por meio de variedades geneticamente superiores. Atualmente tem-se buscado otimizar os programas de melhoramento genético por meio de técnicas de fenotipagem de alto rendimento e que, por meio de algoritmos computacionais, possam auxiliar o melhorista na tomada de decisão quanto à seleção dos genótipos mais adequados. Nesta sessão temática, serão abordados exemplos de aplicação, na cultura do trigo e da cana-de-açúcar, de uso de tecnologias de imagens multiespeciais e de espectros no infravermelho próximo como meio de otimizar as coletas de dados “indiretos” para seleção de genótipos nas etapas iniciais dos programas de melhoramento. Fechando a sessão, será apresentado um pacote computacional que permite, de maneira fácil, rápida e eficiente, as análises dos dados obtidos de imagens digitais.

Palestras:

1. **Luiz Alexandre Peternelli (UFV)** – Otimização do programa de melhoramento da cana com dados espectrais
2. **Maicon Nardino (UFV)** – Como posso otimizar meu programa de melhoramento: uma visão de HTP para trigo
3. **Tiago Olivoto (UFSC)** – Recursos computacionais para as análises de imagens



ST4 - Sessão Butantan: Aplicação de métodos estatísticos para o desenvolvimento de vacinas

Data/hora: Terça-feira, 05/08 - 17:05 - 18:35

Apresentadores: Patricia Emilia Braga (Chair, Butantan), Elizabeth Gonzalez Patiño (Butantan), João Italo Dias França (Butantan), Marcos Alves De Lima (Butantan) e Marco Antonio De Oliveira (Butantan)

Resumo

A vacina é uma estratégia eficiente de prevenção primária contra diversas doenças. No processo de desenvolvimento de uma nova vacina, são empregados métodos estatísticos para avaliar a relação risco/benefício.

O Instituto Butantan (IB) é o principal produtor de imunobiológicos do Brasil, trata-se de um grande centro de referência em: pesquisas biológicas, produção de soros e vacinas. O IB possui como objetivo pesquisar, desenvolver, fabricar e fornecer produtos e serviços para a saúde da população, além de ter como política: melhorar a eficiência na produção de biológicos, intensificar a produção de biológicos buscando parcerias, buscar a melhoria contínua do sistema de gestão, minimizar os desvios dos processos produtivos e divulgar e popularizar o ensino de ciências. Assim o IB visa proteger a saúde pública e proporcionar o acesso a vacinas para todos.

O IB, dentro do processo de desenvolvimento de imunobiológicos, realiza ensaios clínicos de Fase I, II, III e IV. Nesses estudos são avaliados desfechos tais como: segurança, eficácia/efetividade e imunogenicidade. Para a avaliação destes desfechos são aplicados métodos estatísticos que subsidiam as tomadas de decisão para aprovação da nova vacina junto à ANVISA.

Nesta sessão temática, iremos apresentar alguns estudos desenvolvidos, pelo IB, contra COVID-19, infecção pelo vírus da dengue e influenza, com ênfase nos delineamentos dos estudos e nas metodologias empregadas.

Ao final, espera-se que os participantes desta sessão temática tenham acesso a aplicações da ciência estatística em estudos reais de relevância à saúde pública.

Palestras:

1. **Elizabeth Gonzalez Patiño (Butantan)** – Estudo de eficácia e segurança de uma vacina: pontos de atenção
2. **Marcos Alves Lima (Butantan)** – Planejamento e desenho do estudo da vacina CoronaVac em Serrana
3. **Patrícia Emilia Braga (Butantan)** – Papel do Butantan na Pandemia
4. **João Ítalo Dias França (Butantan)** – Estatística com foco em Ensaios Clínicos de vacinas



ST5 - Riscos climáticos na agricultura: contribuições da Ciência de Dados

Data/hora: Quarta-feira, 06/08 - 10:30 - 12:00

Apresentadores: Alfredo José Barreto Luiz (Chair, Embrapa Meio Ambiente), José Eduardo Macedo Pezzopane (UFES) e Fernando Antônio Macena da Silva (Embrapa Cerrados)

Resumo

Com influências de diferentes comunidades, a Ciência de Dados evoluiu para fornecer percepções em muitos ambientes diferentes orientados por dados, incluindo as ciências agrária e climática. Nesta sessão, será apresentada uma breve revisão da Ciência de Dados e sua conexão com a agrometeorologia. Além disso, serão discutidos os rumos da área para quantificar os riscos das mudanças climáticas e discutidas algumas áreas chave de pesquisa em Estatística e métodos quantitativos na análise de riscos climáticos na agricultura. Especialistas exibirão sobre as técnicas usadas em zoneamentos agrícolas de risco climático, em análise de risco de doenças e de incêndios florestais e suas relações com as variáveis agrometeorológicas e as mudanças climáticas. O objetivo será mostrar como a Ciência de Dados é vital na nossa luta contra as alterações climáticas, através da criação de modelos e da análise de dados que permitem a avaliação de riscos, redução de perigos, adaptação às mudanças e orientação de políticas, na busca de um futuro sustentável.

Palestras:

1. Prof. Dr. José Eduardo Macedo Pezzopane (UFES) – As mudanças climáticas e os riscos florestais
2. Dr. Fernando Antônio Macena da Silva (Embrapa Cerrados) – Avaliação de Riscos e Resiliência Agroclimática



ST6 - Dados que Cuidam: Estatística e Inteligência Artificial para a Saúde de Amanhã

Data/hora: Quarta-feira, 06/08 - 10:30 - 12:00

Apresentadores: Marcos Oliveira Prates (Chair, UFMG), Alexandre Loureiros Rodrigues (UFES) e Deisy Morselli Gysi (UFPR)

Resumo

A integração entre a Estatística e a área da Saúde possui, há muitos anos, um papel crucial no desenvolvimento de métodos, técnicas e modelos estatísticos. Tais pesquisas, também impulsionam avanços significativos em pesquisa e prática clínica. Em um cenário de crescente complexidade dos dados de saúde e a facilidade da coleta dos mesmos provenientes de fontes como registros hospitalares, ensaios clínicos e dados de sensores, a Estatística fornece as ferramentas adequadas para transformar esses dados em informações que auxiliem à tomada de decisões e na formulação de políticas de saúde mais eficazes.

A necessidade de avanços em modelagem estatística para dados médicos torna-se evidente frente aos desafios contemporâneos, como a análise de dados de larga escala, dados espaciais e informações de alta dimensionalidade. Essas novas demandas exigem modelos estatísticos inovadores, capazes de lidar com a estrutura do sistema de saúde, heterogeneidade dos dados e com incertezas inerentes à fatores biológicos.

Além disso, a aplicação de métodos estatísticos na área da saúde é fundamental para alcançar os Objetivos de Desenvolvimento Sustentável (ODS) da ONU, fornecendo evidências e análises necessárias para enfrentar os desafios de saúde, reduzir as desigualdades e promover o bem-estar em uma escala global. Especificamente nossa proposta de Seção Temática apresenta técnicas estatísticas que atuam diretamente nas ODS 3, 10 e 17.

Palestras:

1. **Alexandre Loureiros Rodrigues (DEST/UFES)** – Matching de Procedimentos Médicos: Soluções Inteligentes com Modelos de Linguagem de Grande Porte
2. **Deisy Morselli Gysi (DEST/UFPR)** – Integrando dados para diagnóstico e tratamento de transtornos mentais
3. **Marcos Oliveira Prates (DEST/UFMG)** – Latent Archetypes of the Spatial Patterns of Cancer



ST7 - Sessão Brazilian Journal of Biometrics: Melhores Artigos

Data/hora: Quinta-feira, 07/08 - 08:30 - 10:00

Apresentadores: Tales Jesus Fernandes (Chair, UFLA), Jesús Alberto Peña Guillén (Universidad de Los Andes, Venezuela) e Eder Silva de Brito (Instituto Federal de Educação, Ciência e Tecnologia de Goiás)

Resumo

Será feita uma breve apresentação da história e contexto atual da BJB, enquanto revista oficial da RBras. Além disso, serão apresentados os 2 melhores artigos publicados na BJB durante o ano de 2024.

Palestras:

1. **Jesús Alberto Peña Guillén (Universidad de Los Andes, Venezuela)** – A bivariate survival model for events with dependent failure times based on Archimedean copula functions. Application case: A sample of HIV patients

Resumo: This paper proposes a bivariate survival model for dependent failure times based on copula functions of the Archimedean family and the mean cumulative function for non-recurrent events of different types (MCFRE) and uses it to estimate the probability of survival from the occurrence of events of different types on the same HIV/AIDS patient. The copula functions evaluate the dependence structure between the failure times of the events experienced by the same patient throughout their follow-up period, and the MCFRE generates the marginal survival function for each event. The marginal function is a nonparametric estimator that gives the same estimated survival probability as the Kaplan-Meier estimator if the failure times of the different types of events are independent. If each patient experiences at least one event, a subset of them generates a compound event that affects the estimated probability of survival. The results show that the traditionally estimated survival probabilities are biased if dependent failure times are treated as independent.

2. **Eder Silva de Brito (Instituto Federal de Educação, Ciência e Tecnologia de Goiás)** – Reliability of repairable systems with Non-Central Gamma frailty

Resumo: Maintenance actions on industrial equipment are essential to reduce expenses associated with equipment failures. Based on a well-fitted model, it is possible, through the estimated parameters, to predict several functions of interest, such as the cumulative average and reliability functions. In this paper, a new frailty model is proposed to analyze failure times of repairable systems subject to unobserved heterogeneity actions. The Non-Central Gamma distribution is assumed to the frailty random variable effect. The class of minimal repair models for repairable systems is explored considering an approach that includes the frailty term to estimate the unobserved heterogeneity over the systems' failure process. Classical inferential methods



69^a Reunião Anual da RBras 21º SEAGRO

Estatística e Ciência de Dados
a serviço do Desenvolvimento Sustentável

Vitória - ES

4 a 8 de agosto de 2025

were used to parameter estimation and define the reliability prediction functions. A simulation study was conducted to confirm the properties expected in the estimators. Two real-world data known in literature were used to illustrate the estimation procedures and validate the proposed model as a viable alternative to those already established in the literature. The results obtained highlight the potential of our proposed approach, particularly for industries dealing with such systems, where unquantifiable factors may impact equipment failure times.



ST8 - Aprendizado Estatístico de Máquina e Aplicações

Data/hora: Quinta-feira, 07/08 - 10:30 - 12:00

Apresentadores: Anderson Ara (Chair, UFPR), Gabriela Maria Rodrigues (ESALQ/USP) e Eufrásio de Lima Neto (UFPB)

Resumo

As técnicas de Aprendizado Estatístico de Máquina, do inglês Statistical Machine Learning, têm recebido muita atenção nos últimos anos devido ao fato de automatizarem modelos analíticos com um bom poder preditivo e que podem ser aplicados para resolução de diversos problemas. Nesta sessão temática serão apresentados algumas metodologias de Aprendizado Estatístico e suas aplicações.

Palestras:

1. **Gabriela Maria Rodrigues (ESALQ/USP)** – New regression model and machine learning with application to human development index
2. **Anderson Ara (DEST-UFPR)** – Identifying 124 new anti-HIV drug candidates in a 37 billion compound database: An integrated approach of machine learning (QSAR), molecular docking, and molecular dynamics simulation
3. **Eufrásio de Lima Neto (DEST-UFPB)** – Identifying early help referrals for local authorities with machine learning and bias analysis



ST9 - Jovens Pesquisadores RBras - Políticas públicas: A formação do profissional em estatística e ciência de dados está de fato o preparando para a tomada de decisões?

Data/hora: Quinta-feira, 07/08 - 10:30 - 12:00

Apresentadores: Allison Queiroz de Oliveira (Chair, ESALQ/USP), Elisangela Aparecida da Silva Lizzi (Universidade Tecnológica Federal do Paraná) e Doris Satie Maruyama Fontes (CONRE-3)

Resumo

A tomada de decisões pelo estado, governo ou qualquer órgão competente impacta diretamente o caminho trilhado pela comunidade. Contudo tais decisões devem ser feitas de forma clara, inteligente, precisa e que considere as consequências (sejam elas positivas ou negativas) geradas. Com isso o papel de um profissional que alinhe as demandas sociais e suas complexidades à técnicas que auxiliem nas melhores escolhas possui um impacto crucial neste processo. No atual contexto, com a alta geração de dados informativos acerca da sociedade, o uso de métricas e técnicas matemáticas e estatísticas são indispensáveis. Diante disso, durante esta mesa redonda traremos à luz como está a formação atual dos profissionais de estatística e ciência de dados e se de fato esta está colaborando para que o mercado receba profissionais qualificados para a tomada de decisões em políticas públicas e, caso contrário, o que deve ser proposto em termos de ensino para que o perfil do novo profissional destas áreas venha a ser condizente com tal demanda. Além deste eixo, também iremos abordar quais os impactos na qualidade da formação de tais profissionais quando se refere à políticas relacionadas principalmente ao crescimento econômico, da garantia de alimento no prato de todos, qualidade de vida e redução da pobreza em nosso país.

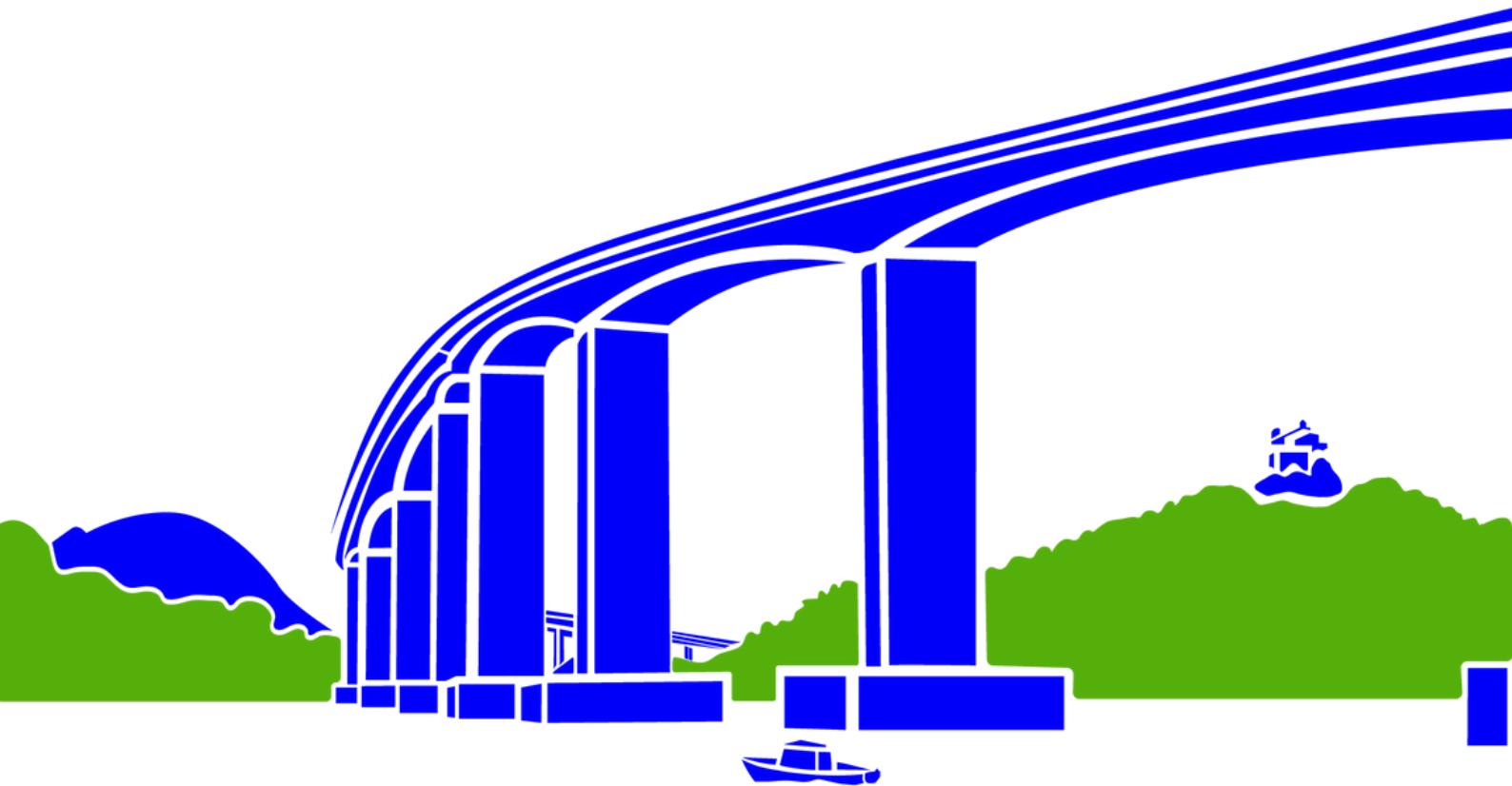
Palestras:

- Elisangela Aparecida da Silva Lizzi (Universidade Tecnológica Federal do Paraná)** - Ensino de estatística para profissionais de tomada de decisões em saúde pública e vigilância em saúde. Qual nossa missão nesse cenário?
- Doris Satie Maruyama Fontes (Conre-3)** - Há Descompasso entre Formação e Necessidades do Mercado de Trabalho?

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Mesas Redondas





MR1 - Estatística e Ciência de Dados como Ferramentas para a Cidadania

Data/hora: Segunda-feira, 04/08 - 17:05 - 18:35

Apresentadores: Doris Satie Maruyama Fontes (Chair, CONRE-3), Geovane Carlos Barbosa (IFES) e Mário Hissamitsu Tarumoto (Unesp)

Resumo

Estamos vivendo uma era onde, de forma muito rápida, a tecnologia vem ganhando espaço no cotidiano do cidadão, seja dentro das escolas, do mercado de trabalho, no lazer, no convívio familiar. No entanto, grande parte dessas tecnologias demanda um mínimo de conhecimento técnico e/ou cultural do usuário para melhor aproveitamento da “inteligência artificial”. Vamos dizer que as novas tecnologias demandam um Pensamento Crítico ou Pensamento Estatístico minimamente desenvolvido do usuário. Um ChatGPT só produzirá uma resposta aproveitável se o prompt tiver sido bem elaborado. No entanto, para elaborar um prompt aproveitável, é necessário que o usuário tenha um mínimo de conhecimento do assunto.

Uma parte significante dos problemas hoje gira em torno da dificuldade de se desenvolver pensamentos lógicos, principalmente pela falta de compreensão numérica (numeracia). As falhas no aprendizado iniciam-se já no EF1 por um ensino fraco em matemática e, muito pior, senão nulo, em estatística e probabilidade, que vai seguir se agravando ao longo de todo o EF2 e EM.

Quando o jovem ingressa num curso superior, na grande maioria dos cursos, a estatística e probabilidade serão lecionadas como curso introdutório. No entanto, o formato frequentemente utilizado, baseado em modelos dos anos 70 ou 80, além de não atrair a atenção dos alunos, exige um conhecimento mínimo de matemática que os alunos nunca tiveram. Mais uma etapa onde há desperdício de oportunidade para aprender a pensar criticamente.

Assim, uma EDUCAÇÃO DE QUALIDADE, sobretudo em Matemática, Estatística e Probabilidade, é fundamental para que cada pessoa possa exercer plenamente a sua cidadania, seja nas decisões pessoais para sua sobrevivência, seja na leitura de notícias e ter capacidade/discriminamento para decidir se é legítimo ou fake-news.



69^a Reunião Anual da RBras 21º SEAGRO

Estatística e Ciência de Dados
a serviço do Desenvolvimento Sustentável

Vitória - ES

4 a 8 de agosto de 2025

MR2 - Sessão ABE - Uso de IA em Estatística

Data/hora: Terça-feira, 05/08 - 10:30 - 12:00

Apresentadores: Clécio da Silva Ferreira (Chair, UFJF e Diretoria da ABE), Fernanda Schumacher (Ohio State University, US), Márcio Diniz (Mount Sinai, NY) e Victor Hugo Lachos (University of Connecticut, US)

Resumo

O objetivo é discutir o uso de novas tecnologias baseadas em inteligência artificial no ambiente de trabalho, tanto em termos de pesquisa como em aplicações.



MR3 - Publicações Científicas na Era da Inteligência Artificial e Pré-prints: Desafios e Oportunidades

Data/hora: Terça-feira, 05/08 - 17:05 - 18:35

Apresentadores: Rafaela Figueiredo Rodrigues (Chair, Unifal-MG), Eric Batista Ferreira (Unifal-MG) e Tales Jesus Fernandes (UFLA).

Resumo

A revista científica Sigmae, que desde 2012 contribui para o avanço da Matemática e Estatística por meio da divulgação de pesquisas de qualidade, propõe a realização de uma mesa redonda para explorar o impacto das tecnologias emergentes e novas práticas na comunicação acadêmica. Com o tema Publicações Científicas, Inteligência Artificial e Pré-prints, a ideia da mesa é buscar reunir editores experientes para debater os desdobramentos dessa interseção tecnológica e metodológica.

A mesa redonda contará com a participação de três editores de revistas científicas e um moderador, de diferentes áreas, que discutirão os seguintes aspectos: 1) a influência da inteligência artificial nas publicações científicas; 2) o papel dos pré-prints no ciclo de publicação; e 3) como as revistas podem se posicionar.

A discussão será moderada por uma pesquisadora, garantindo que os debates abordem perspectivas variadas e atendam às demandas do público presente. A revista Sigmae propõe essa atividade como uma forma de fomentar reflexões relevantes para a comunidade acadêmica, promovendo o diálogo sobre como as publicações científicas podem evoluir em tempos de mudanças tecnológicas aceleradas.

Espera-se que o evento ofereça insights práticos e estratégicos tanto para pesquisadores quanto para outros atores envolvidos no ecossistema da comunicação científica, contribuindo para o aperfeiçoamento das práticas editoriais e a adoção de soluções inovadoras.



MR4 - Desafios e avanços na fiscalização e regulamentação da profissão de estatístico

Data/hora: Quinta-feira, 07/08 - 08:00 - 08:30

Apresentadores: Stéfane Rossoni (Chair, CONRE-4) e Doris Satie Maruyama Fontes (CONRE-3).

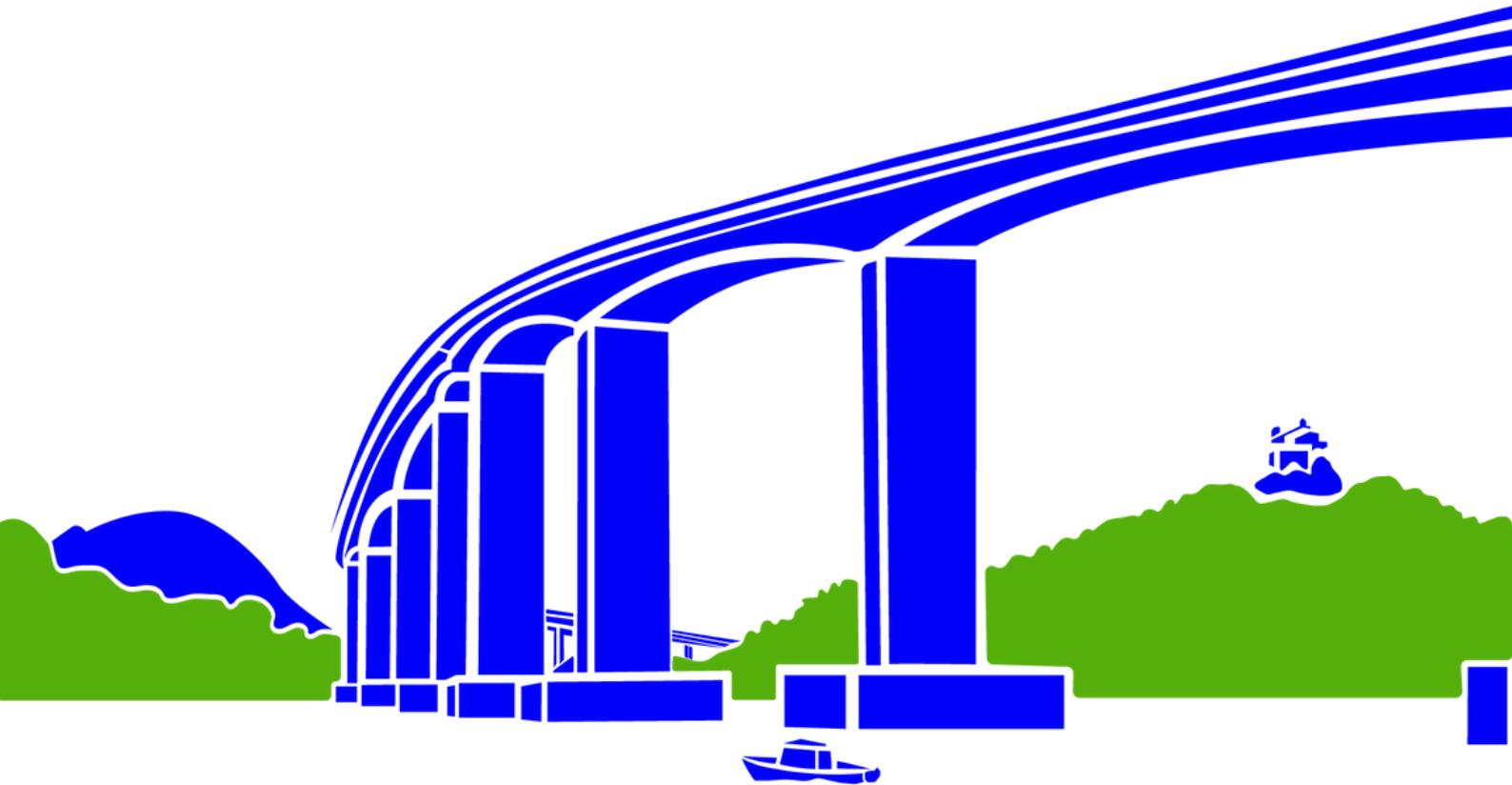
Resumo

A mesa-redonda “Desafios e Avanços na Fiscalização e Regulamentação da Profissão de Estatístico” propõe um debate sobre os principais entraves enfrentados na consolidação e valorização da profissão no Brasil. Serão abordados temas como a baixa adesão ao registro profissional, a contratação de estatísticos sob outras denominações para burlar direitos trabalhistas, a escassez de fiscais nos Conselhos Regionais e a importância do engajamento coletivo da categoria para garantir o exercício ético e legal da Estatística. A discussão também destacará iniciativas em andamento, propostas de modernização e o papel dos órgãos de classe na defesa da profissão.

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Miniconferências





Y1 - Processos de contagem e derivações: Uma aplicação para dados ambientais e epidemiológicos

Data: Segunda-feira, 04/08 - 15:05 - 15:35

Palestrantes: Ana Julia Alves Câmara (UFES)

Resumo

Nos últimos anos, o crescimento econômico e o desenvolvimento urbano decorrentes da expansão das cidades e comunidades trouxeram consigo tanto benefícios quanto desafios. Enquanto impulsionaram a economia e o progresso urbano, também resultaram em problemas ambientais e de saúde, devido à geração de resíduos que impactam a qualidade de vida da população. Modelos para dados de contagem têm sido amplamente empregados em estudos epidemiológicos para avaliar o impacto da poluição do ar na saúde humana. A presença de observações atípicas nesse tipo de dado pode prejudicar significativamente a estimativa e inferência dos parâmetros de interesse, impactando os resultados e as conclusões de estudos científicos na área. Devido à natureza do dado, um modelo GLARMA robusto é proposto baseado na robustificação da função de quase-verossimilhança. As observações atípicas são tratadas separadamente, usando funções de peso para as variáveis explicativas e a função de perda Huber para a variável resposta. Estudos de simulação demonstraram que essa abordagem robusta é mais confiável do que o modelo GLARMA clássico quando os dados estão contaminados. Para ilustrar a aplicabilidade do modelo, um estudo utilizando dados reais investiga o impacto do material particulado (PM10) no número de mortes por doenças respiratórias na cidade de Vitória/ES, Brasil.



Y2 - Abordagem Bayesiana para a Predição do Estoque de Carbono em Florestas Plantadas

Data: Terça-feira, 05/08 - 15:05 - 15:35

Palestrantes: Fábio Prataviera (ESALQ/USP)

Resumo

O estoque de carbono em florestas plantadas desempenha um papel fundamental no sequestro de carbono e na mitigação das mudanças climáticas (ODS 13). Modelar e prever essa variável de forma precisa é importante para o planejamento sustentável e a tomada de decisões estratégicas no manejo florestal. Neste estudo, propomos uma abordagem bayesiana para a predição do estoque de carbono, explorando a flexibilidade dos Modelos Aditivos Bayesianos para Localização, Escala e Forma (BAMLSS). Essa metodologia permite capturar relações complexas entre variáveis ambientais, estruturando a incerteza de forma mais robusta e incorporando efeitos suavizados sem pressuposições rígidas sobre a forma das relações. Além disso, analisamos a influência e o comportamento de diferentes fatores para o carbono sequestrado na biomassa vegetal, como a idade das árvores, características do solo e condições climáticas. A modelagem bayesiana possibilita uma inferência robusta e interpretável, oferecendo estimativas para a predição e monitoramento do carbono em plantações florestais. Dessa forma, o estudo busca quantificar o carbono armazenado em florestas plantadas, ressaltando seu papel essencial na mitigação dos gases de efeito estufa.



Y3 - Um novo modelo de fração cura com causas concorrentes Geométrica Zero-Inflacionada

Data: Quarta-feira, 06/08 - 15:05 - 15:35

Palestrantes: Marcia Brandão (UFAM)

Resumo

Uma nova modelagem para análise de sobrevivência é proposta, ao considerar que o número de causas concorrentes relacionadas à ocorrência do evento de interesse pode ser correlacionado. Diferentemente da suposição tradicional de independência entre essas causas, o modelo reconhece a complexidade biológica observada em pacientes com câncer testicular, onde diferentes fatores interagem de forma não observável. Especificamente, considera-se que o número de causas competitivas segue uma distribuição geométrica inflacionada em zero (ZIG) com dependência estrutural, permitindo capturar aspectos latentes específicos deste tipo de câncer que não podem ser diretamente quantificados. São apresentadas propriedades estatísticas do modelo, e a estimação dos parâmetros é realizada por intermédio do método da máxima verossimilhança. Adicionalmente, um estudo de simulação de Monte Carlo é conduzido para avaliar o comportamento dos estimadores e a precisão dos intervalos de confiança obtidos. Por fim, o modelo é aplicado a um conjunto de dados reais, referentes a casos de câncer testicular registrados em São Paulo, Brasil, demonstrando seu desempenho competitivo em relação a modelos clássicos e reforçando sua utilidade em aplicações práticas.



Y4 - A Bayesian joint model of multiple longitudinal and categorical outcomes with application to multiple myeloma using permutation-based variable importance

Data: Quinta-feira, 07/08 - 15:05 - 15:35

Palestrantes: Danilo Alvares (University of Cambridge, UK)

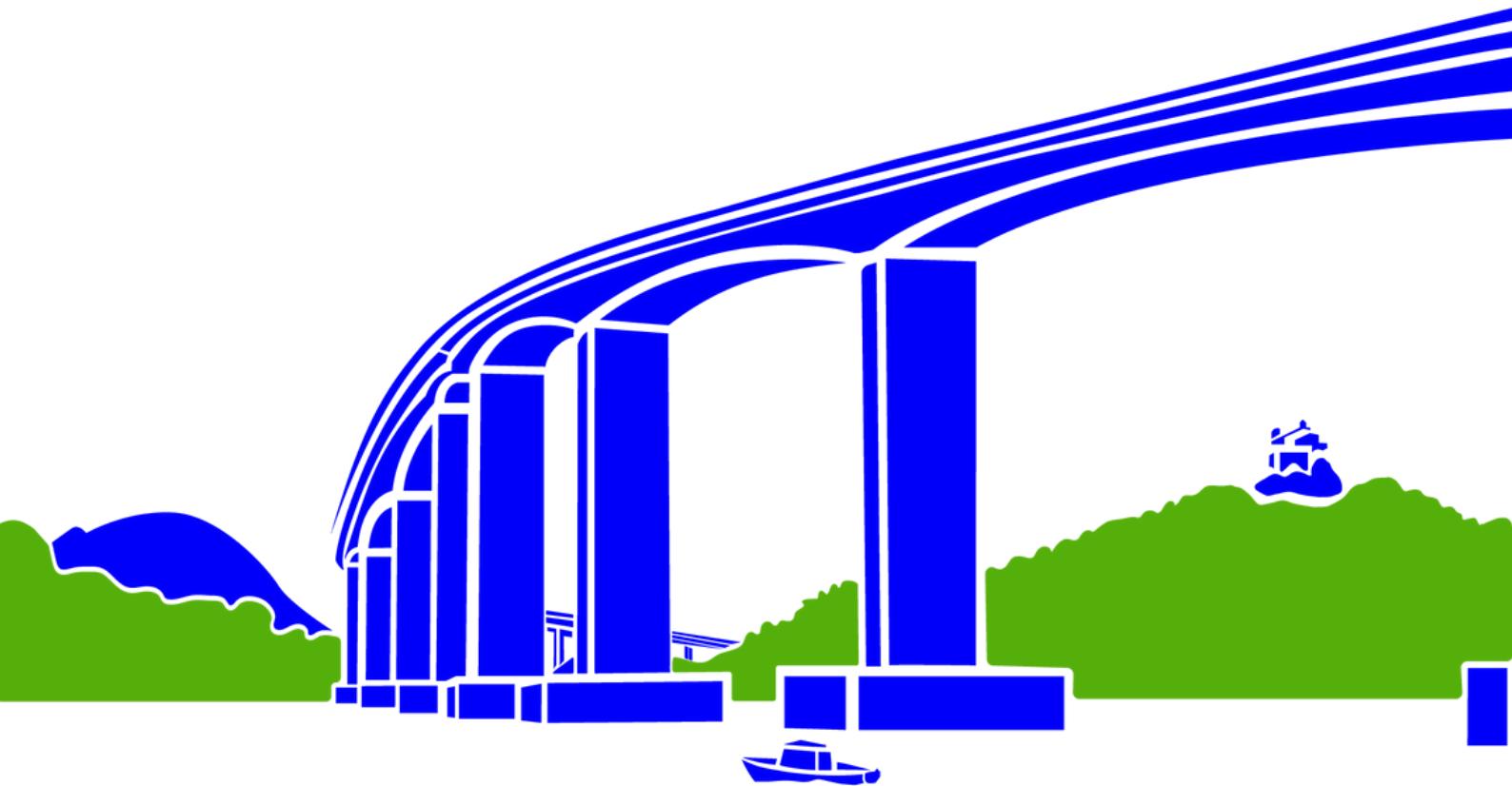
Resumo

Joint models have proven to be an effective approach for uncovering potentially hidden connections between various types of outcomes, mainly continuous, time-to-event, and binary. Typically, longitudinal continuous outcomes are characterized by linear mixed-effects models, survival outcomes are described by proportional hazards models, and the link between outcomes are captured by shared random effects. Other modeling variations include generalized linear mixed-effects models for longitudinal data and logistic regression when a binary outcome is present, rather than time until an event of interest. However, in a clinical research setting, one might be interested in modeling the physician's chosen treatment based on the patient's medical history to identify prognostic factors. In this situation, there are often multiple treatment options, requiring the use of a multiclass classification approach. Inspired by this context, we develop a Bayesian joint model for longitudinal and categorical data. In particular, our motivation comes from a multiple myeloma study, in which biomarkers display nonlinear trajectories that are well captured through bi-exponential submodels, where patient-level information is shared with the categorical submodel. We also present a variable importance strategy to rank prognostic factors. We apply our proposal and a competing model to the multiple myeloma data, compare the variable importance and inferential results for both models, and illustrate patient-level interpretations using our joint model.

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Tutoriais





T1 - Análise Estatística de Desigualdades no Acesso a Água Potável e Saneamento: Explorando os ODS 6 e 10 com Dados do Censo Brasileiro

Data/hora: Terça-feira, 05/08 - 10:30 - 12:00

Palestrante: Fernando Almeida Barbalho (Ministério da Fazenda)

Resumo

Com base nos dados do Censo 2022 divulgados pelo IBGE, é possível conduzir análises detalhadas considerando características dos domicílios, acesso a redes de água, presença de banheiros, tipos de esgotamento sanitário, bem como informações sobre cor e raça dos moradores e a localização regional das residências, incluindo favelas. Esses dados permitem um panorama abrangente das desigualdades no acesso a recursos básicos em diferentes cidades brasileiras.

O tutorial proposto seguirá um fluxo estruturado, começando com a coleta automática de dados usando o pacote R {sidrar}, avançando para uma análise exploratória de dados para compreender padrões e tendências, e culminando em análises avançadas, como a clusterização de cidades com base em critérios de acesso a água e saneamento. O objetivo é capacitar os participantes a realizar análises robustas e aprofundadas, que possam embasar políticas públicas e intervenções direcionadas para melhorar o acesso equitativo a esses recursos essenciais.



T2 - Similaridade de Imagens usando deep learning com Python

Data/hora: Terça-feira, 05/08 - 17:05 - 18:35

Palestrante: Jodavid de Araújo Ferreira (UFPE)

Resumo

Neste tutorial, será utilizado uma abordagem para fazer correspondência de imagens utilizando a similaridade do cosseno com o algoritmo Vision Transformer (ViT). O objetivo é demonstrar uma técnica de deep learning para realizar correspondência de imagens, que se refere ao processo de identificar e emparelhar partes semelhantes ou correspondentes em duas ou mais imagens. Esta tarefa pode ser considerada complexa devido a variações de iluminação, ângulo de captura e qualidade da imagem, especialmente em cenários do mundo real, como o reconhecimento de objetos, animais ou pessoas em diferentes imagens. Assim, este tutorial tem como objetivo abordar as etapas de pré-processamento de imagens, geração de embeddings usando Vision Transformers e cálculo de similaridade através da similaridade do cosseno com a linguagem de programação Python. Espera-se que, ao final, os participantes entendam como aplicar esses métodos tanto em contextos acadêmicos quanto práticos no Mercado de Trabalho.



T3 - Obtendo dados do SIDRA/IBGE com R utilizando o pacote SIDRAR

Data/hora: Quarta-feira, 06/08 - 10:30 - 12:00

Palestrante: Beatriz Milz (UFABC)

Resumo

Este tutorial apresenta uma abordagem prática para acessar dados do SIDRA/IBGE diretamente no R, utilizando o pacote SIDRAR para interagir com a base de dados oficial do IBGE. Focaremos na obtenção da série histórica de Produção Agrícola Municipal (PAM) relacionadas à agricultura como exemplo prático, demonstrando como essas informações podem ser utilizadas para análises que contribuem com os Objetivos de Desenvolvimento Sustentável (ODS). Além disso, utilizaremos pacotes do tidyverse para organizar e transformar os dados, facilitando a preparação para análises mais complexas. O objetivo é capacitar os participantes a extrair dados por meio da API do SIDRA e prepará-los de maneira eficiente e estruturada.



T4 - Testes de hipóteses em Ensaios Clínicos

Data/hora: Quinta-feira, 07/08 - 10:30 - 12:00

Palestrantes: Elizabeth Gonzalez Patiño (Butantan) e Marcos Alves de Lima (Butantan)

Resumo

Este tutorial possui a intenção de demonstrar o papel da Estatística nas atividades científicas pela busca de tratamentos mais eficazes e seguros para algumas doenças. No desenvolvimento clínico de novos produtos, tais como vacinas e medicamentos, o uso de métodos estatísticos específicos, surgem como ferramentas fundamentais para proporcionar e apoiar tomadas de decisões. Para que um novo medicamento ou vacina sejam aprovados para uso, é necessário comprovar que este novo produto traz benefícios, que seja seguro e no caso de comparações com outros produtos, que seja similar ou até mesmo superior. Desta forma, a comparação entre dois ou mais medicamentos ou vacinas podem possuir objetivos distintos, tais como: determinar se há evidência de diferença estatística na comparação dos tratamentos (um produto é superior a outro), demonstrar que há uma padronização e/ou uma consistência na fabricação de um medicamento (equivalência) ou até mesmo mostrar que um novo produto não é inferior clinicamente em relação a outro tratamento (não-inferioridade).

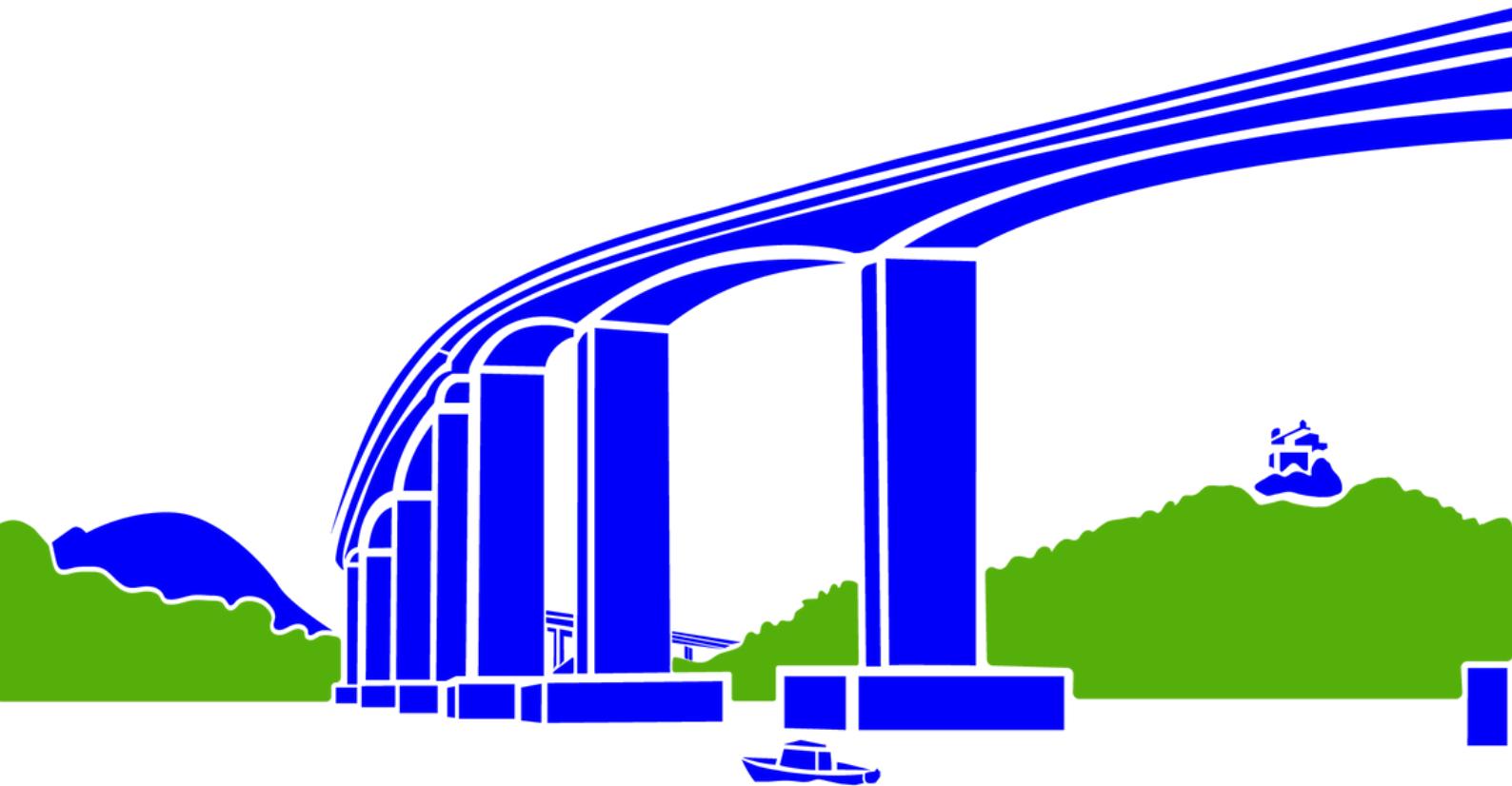
Neste tutorial, iremos nos aprofundar nos conceitos e nas formulações de hipóteses para testes de superioridade, não-inferioridade e equivalência, explorando suas aplicações e limitações. Além disso, traremos exemplos práticos de estudos do Centro de Ensaios Clínicos e Farmacovigilância do Instituto Butantan que utilizaram tais técnicas estatísticas.

Ao final deste tutorial, espera-se que os participantes, compreendam os diferentes tipos de testes de hipóteses utilizados em ensaios clínicos para o desenvolvimento de vacinas e consigam avaliar criticamente os resultados de estudos clínicos já publicados.

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

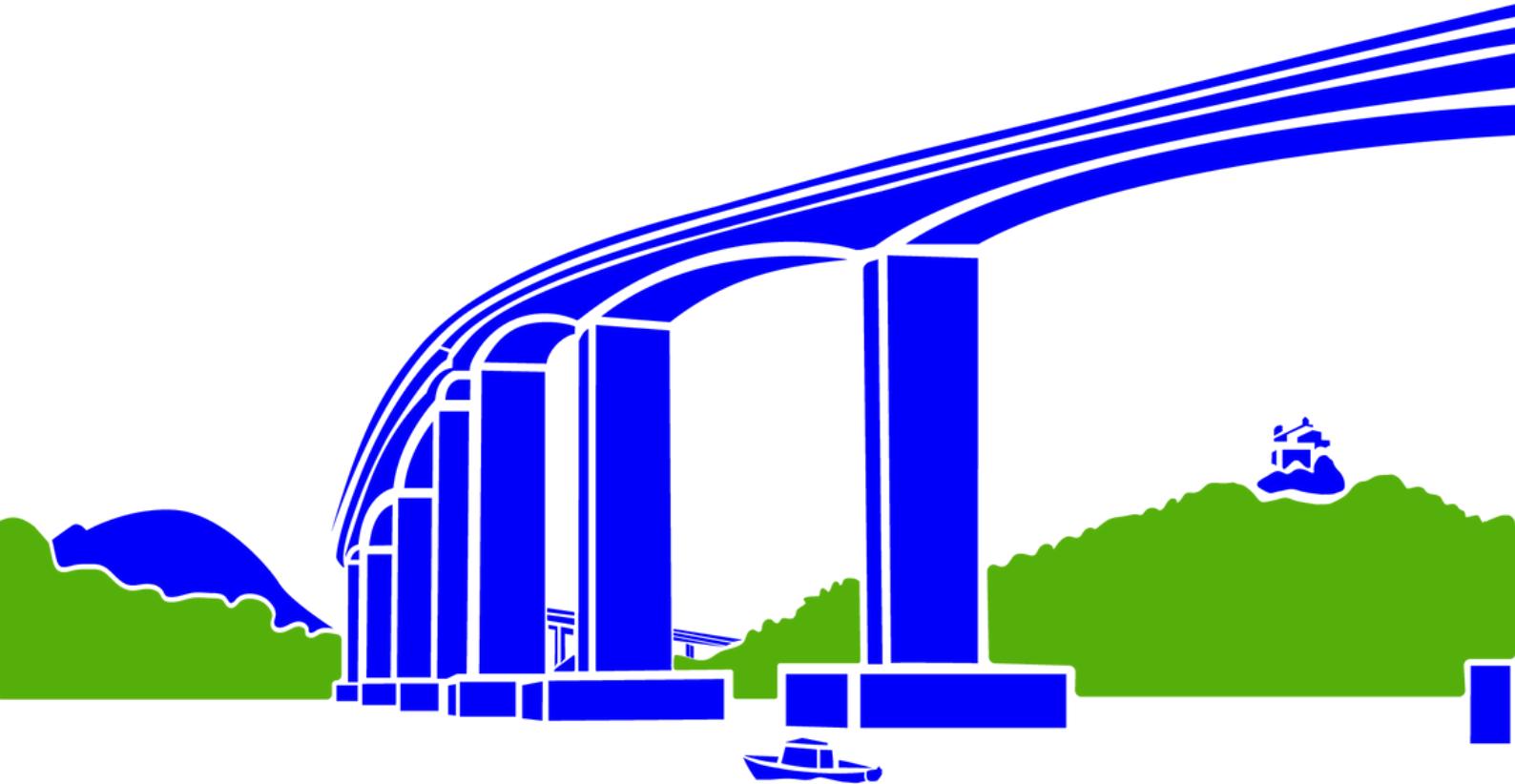
Comunicações Orais



VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Comunicações Orais
Análise de
Sobrevivência





Um novo modelo de tempo de promoção considerando a distribuição de série de potências modificada com uma rede neural para variáveis não estruturadas

Natan Hilario da Silva ¹ Adriano Kamimura Suzuki ² Vicente Garibay Cancho ² Josemar Rodrigues ² Narayanaswamy Balakrishnan ³

Resumo

Neste trabalho, generalizamos a aplicação dos modelos de cura por tempo de promoção para a inclusão de variáveis não estruturadas por meio de uma rede neural. Consideraremos que o número de riscos competitivos segue uma distribuição Série de Potências Modificada, que generaliza distribuições como Poisson, Binomial Negativa, Poisson generalizada restrita, entre outras. A estimativa é feita por meio do algoritmo EM. Estudos de simulação avaliam a convergência do modelo e a qualidade das métricas consideradas. Propomos uma nova métrica para a seleção de modelos de cura. Uma aplicação é feita utilizando tempos até o diagnóstico de pacientes de Alzheimer, flexibilizando modelos pré-existentes.

Palavras-chave: análise de sobrevivência; modelos de cura; redes neurais; variáveis não-estruturadas; algoritmo EM

¹Programa Interinstitucional de Pós-Graduação em Estatística, UFSCar / USP - natan.hilario@usp.br

²Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo (USP)

³Department of Mathematics and Statistics, McMaster University



Modelo de regressão de fragilidade discreta Waring sob esquemas de ativação aleatória

Jonathan K. J. Vasquez;¹ Vera Tomazella² Pedro Rafael D. Marinho³

Resumo

Este estudo propõe um modelo de regressão para análise de sobrevivência incorporando fragilidade discreta com base na distribuição Waring, combinada com um mecanismo de ativação aleatória. Identificar as causas primárias de eventos críticos continua sendo um desafio significativo, particularmente em cenários com etiologia multifatorial. A distribuição Waring fornece uma estrutura robusta para capturar a heterogeneidade dos fatores de risco e modelar a fração de cura de forma mais realista. No modelo proposto, o número de causas latentes segue uma distribuição Waring, enquanto os tempos de ativação dessas causas são modelados usando uma distribuição Weibull. A influência da ativação aleatória é examinada em termos de como ela afeta a dispersão distribucional e captura fontes internas e externas de variabilidade, frequentemente negligenciadas em modelos tradicionais. A abordagem metodológica se baseia em estudos anteriores enfatizando a importância da ativação na modelagem da taxa de cura. O desempenho do modelo será avaliado por meio de um estudo de simulação de Monte Carlo para explorar suas propriedades inferenciais em vários cenários. Além disso, a aplicabilidade prática da metodologia proposta será ilustrada por meio de uma aplicação a dados reais sobre pacientes com câncer de melanoma, destacando sua flexibilidade e capacidade de capturar estruturas de risco complexas. As descobertas ressaltam o potencial da distribuição de Waring em contextos biomédicos, contribuindo para o avanço de técnicas modernas em análise de sobrevivência.

Palavras-chave: Modelo de Fragilidade, Distribuição de Waring, Análise de Sobrevivência, Causas de Risco, Esquemas de Ativação Aleatória, Fração de Cura.

¹Institute of Mathematics and Computer Sciences, University of São Paulo, São Carlos, SP, Brazil - kevin.jv@usp.br

²Department of Statistics, Federal University of São Carlos, São Carlos, SP, Brazil - vera.tomazella@gmail.com

³Departamento de Estatística, Universidade Federal de Pernambuco, Recife, Brazil - pedro.rafael.marinho@gmail.com



Modelagem espacial do desmatamento no sul do Amazonas: uma abordagem utilizando análise de sobrevivência

Maurício dos Anjos da Silva ¹

Marciel Lelis Duarte ²

Nerilson Terra Santos ³

Sebastião Martins Filho ⁴

Resumo

O desmatamento é um problema mundial relacionado a diversos fatores ecológicos, climáticos e socioeconômicos. O presente trabalho objetivou analisar o efeito de variáveis socioeconômicas e ambientais no desmatamento no sul do Amazonas, por meio da aplicação de métodos de análise de sobrevivência e análise espacial. Além disso, é igualmente relevante inserir a dependência espacial dos dados no modelo e avaliar o efeito aleatório espacial associado a cada grupo de área analisado. Em primeira análise, o tempo até o desmatamento em cada unidade de área foi estimado por meio do método não paramétrico de Kaplan Meier, considerando o nível de cada covariável para análise separadamente. Posteriormente, todas as covariáveis foram inseridas na análise, para uma avaliação em conjunto. A dependência espacial observada nos dados de desmatamento foi modelada por meio do modelo ICAR, considerando uma matriz de vizinhanças que captura a dependência espacial dos dados. Os efeitos espaciais associados a cada grupo foram inseridos no modelo de fragilidade, com o intuito de capturar a dependência espacial. As variáveis unidades de conservação, terras indígenas e pastagem foram significativas e demonstraram uma importância maior para explicar o desmatamento na região. A magnitude dos efeitos aleatórios mostraram que regiões que fazem fronteira ao arco do desmatamento possuem um maior tempo esperado associado ao desmatamento e ligados a fatores não observados, não mensurados nos dados. Dessa maneira, os modelos de fragilidade são adequados para a modelagem do desmatamento.

Palavras-chave: Modelos de fragilidade compartilhada; Modelagem Bayesiana; Modelo ICAR; Análise de vizinhança; Floresta Amazônica.

¹ Departamento de Estatística, Universidade Federal de Minas Gerais – mauricioanjos@ufmg.br

² Departamento de Estatística, Universidade Federal de Viçosa – marciel.duarte@ufv.br

³ Departamento de Estatística, Universidade Federal de Viçosa – nsantos@ufv.br

⁴ Departamento de Estatística, Universidade Federal de Viçosa – martinsfilho@ufv.br



Modelo de regressão odd log-logística generalizada na presença de censura informativa

Valdemiro Piedade Vigas;¹ Edwin M. M. Ortega;² Gauss M. Cordeiro;³ Giovani L. Silva;⁴ Paulo C. dos Santos Junior;⁵

Resumo

A escolha do mecanismo de censura é muito importante quando se analisa dados de sobrevivência, pois, em algumas situações práticas, existe dependência entre os tempos de falha e de censura, o que contraria a suposição de censura não-informativa. O presente trabalho propõe uma regressão para dados censurados na presença de censura informativa, com base na família de distribuições odd log-logística generalizada. O modelo proposto se baseia na suposição de que os tempos de falha e de censura são condicionalmente independentes, dado um termo de fragilidade. A escolha família odd log-logística generalizada para os tempos de falha e de censura é justificada pelo fato de estender algumas distribuições já consolidadas na área de sobrevivência e, simultaneamente, por oferecer grande flexibilidade na modelagem de dados práticos, além da capacidade de representar diversas formas da função de risco, incluindo a bimodal. O método de máxima verossimilhança é utilizado para estimar os parâmetros do modelo. Para diferentes configurações de parâmetros e tamanhos amostrais, são realizadas simulações com o objetivo de analisar o comportamento das estimativas de máxima verossimilhança. Aplicação a um conjunto de dados reais ilustra a utilidade do modelo proposto.

Palavras-chave: Censura informativa; Distribuição odd log-logística generalizada; Fragilidade; Modelo de regressão.

Abstract

The choice of the censoring mechanism is very important when we are analyzing survival data, because in some practical situations there is a dependence between the failure and censoring times, contradicting the assumption of non-informative censoring. The present work proposes a censored-data regression in the presence of informative censoring based on the generalized odd log-logistic family of distributions. The proposed model is based on the assumption that the failure and censoring times are conditionally independent given a frailty. The choice of the widespread odd log-logistics family for failure and censoring times is justified by the fact that it generalizes some distributions already consolidated in the area of survival and concomitantly with the fact that it offers great flexibility in data modeling in practice, in addition to the capacity to present various forms of the risk function, among them, bimodal. The maximum likelihood method is used to estimate the model parameters. For different parameter configurations and sample sizes, different simulations are performed to analyze the behavior of the maximum likelihood estimates. An application to a real dataset illustrates the usefulness of the proposed model.

Keywords: Frailty; generalized odd log-logistic distribution; informative censoring; regression model.

¹INMA, UFMS - valdemiro.vigas@ufms.br

²ESALQ, USP - edwnin@usp.br

³DEINFO, UFPE - gausscordeiro@gmail.com

⁴IST, ULisboa - giovani.silva@tecnico.ulisboa.pt

⁵ICEN, UFPA - pauloest16@gmail.com



Modelo de regressão Weibull generalizada restrita em eventos competitivos: aplicação no tempo de emprego dos funcionários do Município de Simões Filho - Bahia - 2021

Rafael Toledo Costa de Almeida;¹ Giovana Oliveira Silva;² Edleide de Brito;³

Resumo

Em análise de sobrevivência, geralmente o interesse está no tempo até a ocorrência de um evento, porém existem situações que o interesse está em mais de um evento e, além disso, a ocorrência de um dos eventos impede que os demais eventos aconteçam. Nesta situação, usa-se modelos que consideram a presença de eventos competitivos.

Inicialmente, considerou-se eventos competitivos, dados censurados e modelo de regressão a partir da distribuição Weibull Generalizada. No estudo de simulação conduzido para diferentes combinações de valores dos parâmetros, tamanhos amostrais e percentuais de censura, observou-se que as estimativas obtidas por máxima verossimilhança apresentavam variâncias elevadas. Diante desse resultado, optou-se pelo modelo com restrição na distribuição de probabilidade.

Adicionalmente, o bom desempenho deste modelo restrito foi verificado por meio de um novo estudo de simulação nos diversos cenários citados. Os resultados indicam que o modelo restrito apresentou estimativas precisas e com menor variância, com melhora significativa à medida que o tamanho da amostra aumenta. Testes de hipóteses para verificar a significância dos coeficientes de regressão foram aplicados utilizando a distribuição assintótica dos estimadores de máxima verossimilhança.

Com o propósito de analisar a adequabilidade do modelo restrito, foi usado o resíduo *quantile*. Por fim, a metodologia foi aplicada ao conjunto de dados que trata da duração do vínculo empregatício de empresas privadas de Simões Filho (Bahia), em 2021, sendo tomado como eventos competitivos demissão e desligamento sem justa causa.

Palavras-chave: Análise de Sobrevivência; Distribuição Weibull Generalizada; Evento competitivos; Resíduos; Simulação.

¹Programa de Pós-Graduação em Matemática, UFBA - rafael.tole.doc@hotmail.com

²Departamento de Estatística, UFBA - giovana@ufba.br

³Departamento de Estatística, UFBA - edbrito@ufba.br



Análise Bayesiana de um modelo de fragilidade discreta baseado na distribuição Hurdle de série de potência zero-modificada: Aplicação em câncer de pulmão

Katy Rocio Cruz Molina¹; Joaquín Martínez-Minaya²; Danilo Alvares³; Vera Tomazella⁴

Resumo

Os modelos de sobrevivência com a presença do termo de fragilidade são ferramentas essenciais para capturar a heterogeneidade não observada entre indivíduos em estudos clínicos e epidemiológicos. Propõe-se um modelo de sobrevivência Bayesiano em que a fragilidade discreta é modelada utilizando a versão Hurdle da distribuição de série de potência zero-modificada (HSPZM). A distribuição HSPMZ é estruturada de forma a permitir que um dos parâmetros modele a proporção de indivíduos imunes ao risco, frequentemente interpretada como a fração de cura, enquanto outro parâmetro captura a heterogeneidade na suscetibilidade entre os indivíduos. Essa capacidade combinada confere ao modelo a flexibilidade necessária para representar realisticamente cenários com ou sem fração de cura. Para a estimativa dos parâmetros, adotou-se uma metodologia Bayesiana, que possibilita a incorporação de conhecimento prévio e fornece inferências robustas, especialmente útil em amostras de tamanho limitado. Como aplicação prática, o modelo proposto foi ajustado a dados de sobrevivência de pacientes com câncer de pulmão. Este é um cenário clinicamente relevante, onde a variabilidade individual influencia significativamente a progressão da doença e a resposta ao tratamento. Os resultados indicam que a abordagem proposta oferece um ajuste superior aos dados em comparação aos modelos de fragilidade convencionais, além de permitir uma interpretação mais refinada do impacto da heterogeneidade na trajetória de sobrevivência dos pacientes.

Palavras-chave: Câncer de pulmão; Distribuição HSPZM; Inferência Bayesiana; Modelo de fragilidade discreta.

¹Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, Brasil - rocio.cm@usp.br

²Departamento de Estadística e Investigación Operativa Aplicadas y Calidad, Universitat Politècnica de València, Valencia, España - jmarmin@eio.upv.es

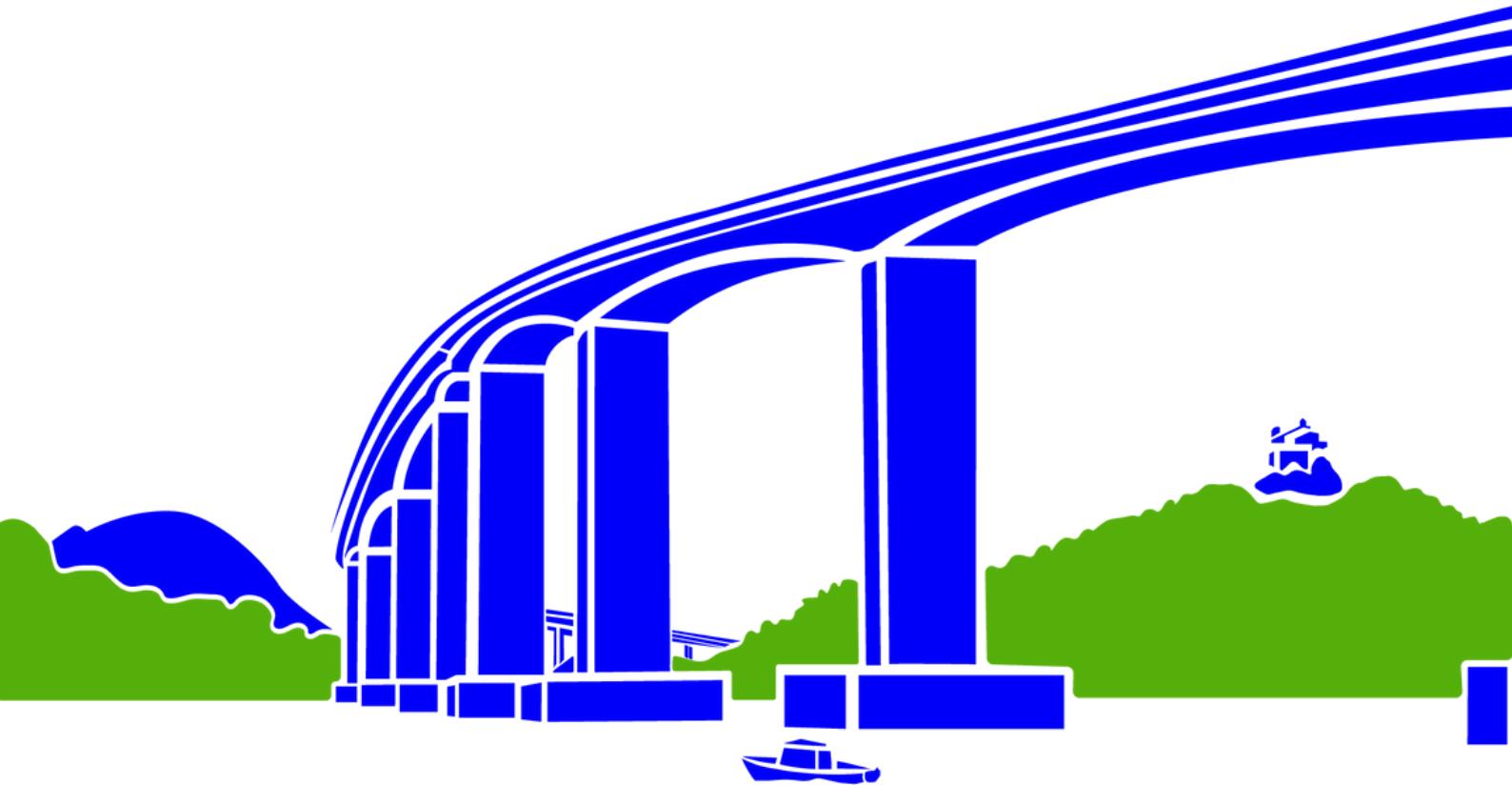
³MRC Biostatistics Unit, University of Cambridge, Cambridge, Kingdom United - danilo.alvares@mrc-bsu.cam.ac.uk

⁴Departamento de Estatística, Universidade Federal de São Carlos, São Carlos, Brasil - vera@ufscar.br

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Comunicações Orais
Aprendizado de
Máquina





Active learning in process mining: an active sampling using dimensionality reduction

Rafael Aguilar Magalhães ¹ Anderson Andolfato Filho ²
Eduardo Alves Portela Santos ³ Anderson Ara ⁴

Resumo

This study presents an integrated methodology that combines Active Learning and Process Mining to enhance the efficiency of CRM systems, focusing on reducing unnecessary calls and improving dialer performance. Unlike traditional model-centric approaches, the method proposes the use of t-SNE for dimensionality reduction and entropy as a criterion for informative sample selection (GSx), reducing dependence on large labeled datasets. The methodology was applied during the transition of a Brazilian company from a telephony-based system to an omnichannel CRM platform. Event logs, call history, and dialer activity were analyzed to identify patterns and estimate success rates. Four predictive models were evaluated — Linear Regression (LR), Support Vector Regression (SVR), Random Forest (RF), and Multilayer Perceptron (MLP) — with LR and SVM using a linear kernel demonstrating superior performance in minimizing inefficient calls. Process mining complemented the analysis by identifying critical variables impacting success rates. The results indicate that integrating entropy and t-SNE with active learning allows for the selection of more representative samples, improves predictive model performance, and significantly reduces computational cost. This approach stands out for its practical applicability in real-world CRM environments and promotes a more intelligent use of data in decision-making and process optimization.

Palavras-chave: Active Learning, Process Mining, Software CRM

¹Federal University of Paraná, Curitiba, PR, Brazil - rafael.aguilar@ufpr.br

²Federal University of Paraná, Curitiba, PR, Brazil - anderson.andolfato@ufpr.br

³Federal University of Paraná, Curitiba, PR, Brazil - portela@ufpr.br

⁴Federal University of Paraná, Curitiba, PR, Brazil - ara@ufpr.br



randomMachines: um modelo de ensemble com máquinas de vetores de suporte

Cristian Pessatti dos Anjos;¹ Anderson Ara²

Resumo

No cenário do aprendizado de máquina, os modelos de classificação assumem um papel fundamental, apresentando grande relevância na Estatística e na Ciência de dados. Suas aplicações na extração de padrões a partir de dados complexos têm sido de grande importância para classificação tanto de variáveis binárias como multiclasse. A classificação multiclasse é crucial no aprendizado de máquina, permitindo a categorização de dados em várias classes distintas. Isso é essencial em muitas aplicações do mundo real, como diagnóstico médico, classificação de espécies e análise de sentimentos em texto. Dentro desse contexto, o modelo de *ensemble* Máquinas Aleatórias se apresenta entre os modelos de aprendizado de máquina com uma notável capacidade preditiva em comparação com abordagens tradicionais. O pacote randomMachines é uma implementação em R do algoritmo Máquinas Aleatórias, disponível no CRAN desde 14 de dezembro de 2023. Em suma, o algoritmo utiliza da técnica de *bagging* para modelos base de *Support Vector Machines* (SVM), utilizando uma amostragem aleatória adicional de funções *kernel* para sua construção fazendo uso da biblioteca *kernlab*. Em sua forma mais básica, os modelos de SVM não modelam classificação multiclasse nativamente, assim executam classificação binária somente em duas classes, e para a classificação multiclasse, o mesmo princípio é utilizado, decompondo o problema de multiclassificação em vários problemas de classificação binária. Atualmente, o pacote randomMachines está preparado exclusivamente para tarefas de classificação em classes binárias. Neste trabalho, propomos uma extensão da metodologia bem como a atualização no pacote randomMachines para lidar com tarefas de classificação envolvendo duas ou mais classes.

Palavras-chave: Machine Learning; SVM; Classificação;

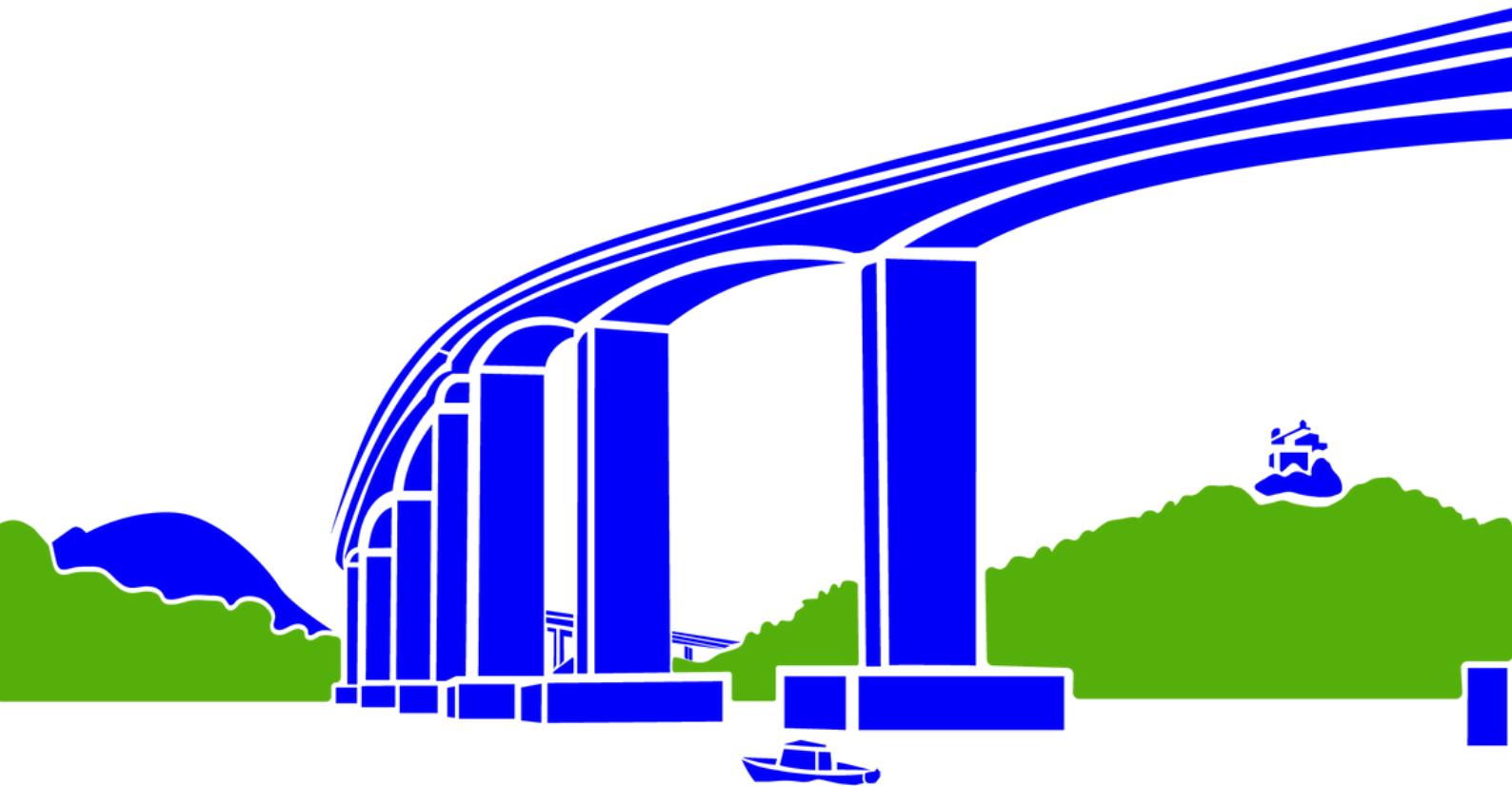
¹Departamento de Estatística, UFPR - cristian.anjos@ufpr.br

²Departamento de Estatística, UFPR - ara@ufpr.br

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Comunicações Orais
Estatística aplicada à
Saúde e Meio Ambiente





Modelagem espaço-temporal e previsão hierárquica de séries temporais dos incêndios florestais no Brasil

Paulo Canas Rodrigues¹

Resumo

Incêndios florestais estão entre os desastres naturais mais comuns em muitas regiões do mundo e têm impacto ativo na qualidade de vida. Esses eventos tornaram-se mais frequentes devido às mudanças climáticas, outras políticas locais e ao comportamento humano. Nesta palestra, considero os dados históricos com as localizações geográficas de todos os "focos de incêndio" detectados pelos satélites de referência que cobrem o território brasileiro entre janeiro de 2011 e dezembro de 2022, totalizando mais de 2,2 milhões de focos. Inicialmente, apresentarei os resultados de um modelo linear generalizado espaço-temporal para dados de unidades espaciais, cujas inferências sobre seus parâmetros são realizadas por meio de uma abordagem Bayesiana, utilizando variáveis meteorológicas (precipitação, temperatura do ar, umidade e velocidade do vento) e uma variável humana (transição e ocupação do uso da terra) como covariáveis. Em seguida, apresentarei os resultados para a previsão hierárquica de séries temporais, onde os seis biomas brasileiros e os 5570 municípios formam a hierarquia. (Trabalho conjunto com Jonatha Pimentel, Rodrigo Bulhões e Ana Pinheiro).

Palavras-chave: modelagem espaço-temporal, previsão de séries temporais hierárquicas, incêndios florestais

¹ Statistical Learning Laboratory (SaLLy) e Departamento de Estatística, Universidade Federal da Bahia, Salvador, BA – paulocanas@gmail.com



Desafios na preparação e análise de dados geoespaciais com R: Lições do projeto ICOLMA

Beatriz Milz ¹; Sandra Momm ²; Gabriel Machado Araujo ³

Resumo

A pandemia de COVID-19 evidenciou desigualdades socioespaciais profundas, sobretudo em grandes centros urbanos como São Paulo. O projeto "Impacto da COVID-19 no modo de vida, mobilidade e acessibilidade dos grupos marginalizados" (ICOLMA) investiga os impactos da COVID-19 na mobilidade de grupos marginalizados em três cidades: São Paulo (Brasil), Cidade do Cabo (África do Sul) e Dortmund (Alemanha). Neste projeto, utilizamos o software Maptionnaire para aplicar um questionário baseado em mapas (*map-based survey*). Esta comunicação oral tem como objetivo apresentar os desafios enfrentados na preparação e análise dos dados geográficos, enfatizando a importância da linguagem R para garantir a reproduzibilidade da análise dos dados. Serão apresentados desafios nos processos de validação, limpeza e integração dos dados, com destaque para as estratégias de correção de inconsistências nas coordenadas geográficas e a unificação de informações oriundas de múltiplos arquivos. Em seguida, detalharemos a aplicação de técnicas de análise geoespacial, como o cálculo de distâncias entre residências e pontos de atividade coletados. Também apresentaremos o dashboard interno em Shiny, desenvolvido com o objetivo de facilitar a exploração dos dados e a visualização dos resultados, além de facilitar a identificação de possíveis erros de coleta. As lições aprendidas durante este processo oferecem importantes contribuições metodológicas, servindo de referencial para futuras pesquisas que utilizam questionários baseados em mapas.

Palavras-chave: map-based survey, mobilidade urbana, COVID-19, grupos marginalizados.

¹ Laboratório de Planejamento Territorial (LaPlan), Universidade Federal do ABC, São Bernardo do Campo – milz.bea@gmail.com

² Laboratório de Planejamento Territorial (LaPlan), Universidade Federal do ABC, São Bernardo do Campo – sandra.momm@ufabc.edu.br

³ Laboratório de Planejamento Territorial (LaPlan), Universidade Federal do ABC, São Bernardo do Campo – g.machado@ufabc.edu.br



Modelação conjunta de respostas binárias longitudinais: aplicação a dados de malnutrição aguda e crónica sob uma abordagem bayesiana não paramétrica

Giovani L. Silva;¹ André Nunes;² Luzia Gonçalves³

Resumo

Na análise de dados longitudinais, são geralmente adotados efeitos aleatórios gaussianos para controlar a heterogeneidade não observada dos indivíduos ao longo do tempo. Neste trabalho, questiona-se essa suposição gaussiana numa análise de variáveis respostas binárias que são conjuntamente distribuídas e com estrutura de dependência via efeitos aleatórios sob uma abordagem bayesiana não paramétrica (BNP). Estatística bayesiana não paramétrica não impõe uma dada distribuição de probabilidade para as variáveis respostas ou efeitos aleatórios ou ambas. Considera-se modelos BNP para os efeitos aleatórios e modelos paramétricos (distribuição de Bernoulli) para as variáveis respostas binárias. Relativamente ao modelo BNP, há uma variedade de processos tais como o processo de Dirichlet (DP), o processo de quebra-vara (*stick-breaking process*), o processo de Pitman-Yor e a árvore de Polya que estão implementados no pacote R NIMBLE. São também empregues métodos de Monte Carlo via cadeias de Markov (MCMC) para fazer inferência sobre os parâmetros do modelo. Por fim, a motivação deste trabalho deve-se a um estudo de dados binários de um ensaio longitudinal aleatorizado em paralelo com quatro braços e conduzido na província de Bengo - Angola, envolvendo um total de 121 crianças com infecções parasitárias intestinais que receberam tratamento inicial. Um dos objetivos do estudo é investigar os efeitos de quatro intervenções nas variáveis respostas de malnutrição aguda e crónica.

Palavras-chave: Estatística não paramétrica bayesiana; modelo misto; malnutrição.

Referências

- 1 Gasparinho, C., Gonçalves, M.H., Chissaque, A., Silva, G.L., Fortes, F., Gonçalves, L. (2022). Wasting, stunting, and anemia in Angolan children after deworming with Albendazole or a test-and-treat approach for intestinal parasites: binary longitudinal models with temporal structure in a four-arm randomized trial. *Nutrients*, 14, 2185.
- 2 Nunes, A., Silva, G.L., Gonçalves, L. (2025). Joint Models of Longitudinal Binary Responses: A Bayesian Nonparametric Approach. In: Henriques-Rodrigues, L., Meñezes, R., Machado, L.M., Faria, S., de Carvalho, M. (Editors), *New Frontiers in Statistics and Data Science*. SPE 2023. Springer Proceedings in Mathematics & Statistics, vol 469, pg.437–454. Springer.

¹CEAUL & Dep. Matemática - IST, Universidade de Lisboa, Portugal - giovani.silva@tecnico.ulisboa.pt

²Dep. Matemática - IST, Universidade de Lisboa, Portugal - andre.b.nunes@tecnico.ulisboa.pt

³CEAUL, Universidade de Lisboa & IHMT, Universidade Nova de Lisboa, Portugal - luziag@ihmt.unl.pt



Análise de série temporal e previsão para dados de sífilis congênita no estado de Minas Gerais: 2007 a 2024

Louiziane Ribeiro Carvalho¹

Valdeline de Paula Mequelinha Ferreira²

Maria Vitória Neves³

Thelma Sáfadi⁴

Joel Augusto Muniz⁵

Resumo

A sífilis é uma infecção crônica e sistêmica, de distribuição global, transmitida entre humanos. Apesar da redução na sua prevalência com o uso da penicilina, observa-se um aumento nos casos, especialmente de sífilis congênita. A sífilis congênita é provocada pela bactéria espiroqueta *Treponema pallidum* ocorrendo por via transplacentária e também durante o parto e a amamentação se houver contato com lesões maternas. O objetivo do estudo foi analisar e fazer previsão da série dos casos de sífilis congênita em Minas Gerais. Trata-se de um estudo de séries temporais considerando para análise o estado de Minas Gerais, a partir do número de casos de sífilis congênita e de nascidos vivos (NV), registrados entre 2007 e 2024. Calculou-se as taxas de incidência da doença expressa como o número de casos por 1000 NV por ano. Ajustou-se o modelo para a série e foi feita a previsão para o período de julho de 2024 a dezembro de 2025. No período de janeiro de 2007 a junho de 2024, foram notificados 22.464 casos de sífilis congênita. A taxa de incidência variou entre 0,69/1000 NV em 2007 a 9,71/1000 NV em 2023. Observou-se uma tendência temporal na série, com um aumento no número de casos até o ano de 2019, decréscimo nos anos de 2020 e 2021, seguido um novo aumento. Os resultados indicam que a sífilis congênita permanece um problema de saúde pública, visto que o número de casos foi crescente para o período analisado e com previsão de aumento para 2025.

Palavras-chave: Séries temporais; sífilis congênita; previsão; modelo ARMA; tendência.

¹ Programa de Pós-Graduação em Estatística, UFLA – louiziane.carvalho2@estudante.ufla.br

² Programa de Pós-Graduação em Estatística, UFLA – valdeline.ferreira1@estudante.ufla.br

³ Programa de Pós-Graduação em Estatística, UFLA – maria.neves5@estudante.ufla.br

⁴ Departamento de Estatística, UFLA – safadi@ufla.br

⁵ Departamento de Estatística, UFLA – joamuniz@ufla.br



Comprehensive statistical analysis reveals significant benefits of COVID-19 vaccination in hospitalized patients: propensity score, covariate adjustment, and feature importance by permutation

Eduardo V. Moraes¹, Magda C. Pires², Guilherme F. Nascimento³, Leonardo C. D. Rocha⁴, Marcos A. Gonçalves⁵, Unaí Tupinambas¹, Milena S. Marcolino⁶

Abstract

COVID-19 vaccines effectively prevent infection and hospitalization. This study aimed to compare the clinical characteristics and outcomes of vaccinated and unvaccinated COVID-19 in hospital patients using advanced statistical methods: propensity score analyses, covariate adjustment and feature importance by permutation. The retrospective cohort included adult COVID-19 patients admitted from March 2021 to August 2022 from 27 hospitals across five Brazilian states during the Delta and Omicron waves.

From the 3,188 patients, 1,963 (61.6%) were unvaccinated and 1,225 (38.4%) were fully vaccinated. Among these, 558 vaccinated individuals were matched with 558 unvaccinated ones. Vaccinated patients had lower rates of mortality (19.4% vs. 33.3%), invasive mechanical ventilation (IMV-18.3% vs. 34.6%), noninvasive mechanical ventilation (NIMV-10.6% vs. 22.0%), intensive care unit admission (ICU-32.0% vs. 44.1%) vasoactive drug use (21.1% vs. 32.6%), dialysis (8.2% vs. 14.7%) hospital length of stay (7.0 vs. 9.0 days), and thromboembolic events (3.9% vs. 7.7%), $p < 0.05$ for all. Risk-adjusted multivariate analysis demonstrated a significant inverse association between vaccination and in-hospital mortality (adjusted odds ratio [aOR] = 0.42, 95% confidence interval [CI]: 0.31-0.56; $p < 0.001$) as well as IMV (aOR = 0.40, 95% CI: 0.30-0.53; $p < 0.001$). These results were consistent in all analyses, including feature importance by permutation.

In conclusion, vaccinated patients admitted to hospital with COVID-19 had significantly lower mortality and other severe outcomes than unvaccinated ones during the Delta and Omicron waves. These findings have important implications for public health strategies and support the critical importance of vaccination efforts, particularly in low-income countries, where vaccination coverage remains suboptimal.

Keywords: COVID-19; vaccine; severe illness; propensity score; machine learning

¹ Medical School and University Hospital, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

² Department of Statistics, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

³ Universidade Federal de São João del-Rei, São João del-Rei, Brazil

⁴ Computer Science Department, Universidade Federal de São João del-Rei, São João del-Rei, Brazil

⁵ Computer Science Department, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

⁶ Department of Internal Medicine, Medical School & Telehealth Center, University Hospital, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil



Selection of objective priors for Gumbel distribution parameters with application to maximum rainfall data

Fernando Antonio Moala;¹ Adriano Buran Moala² Nixon Jerez-Lillo² Pedro Luiz Ramos²

Abstract

The selection of priors is a critical aspect of Bayesian analysis, although the literature lacks studies concerning the application of the Gumbel distribution using different objective priors. We derive objective priors for the two-parameter Gumbel distribution and present a fully Bayesian analysis. Our primary goal is to choose a prior that represents a state of “little knowledge” a priori for both parameters. To yield this, we implement Markov Chain Monte Carlo algorithms to sample from the posterior distribution and to calculate the Bayes estimators. This investigation is made in the context of extreme weather events, using maximum rainfall data.

Keywords: Gumbel distribution; Bayesian inference; objective priors; reference prior; Jeffreys.

¹Department of Statistics, State University of São Paulo, Brazil, f.moala@unesp.br

²Department of Statistics, State University of São Paulo, Brazil, adriano.moala@portoseguro.com.br

²Facultad de Matemáticas, Pontificia Universidad Católica de Chile, Chile, njerez1@mat.uc.cl

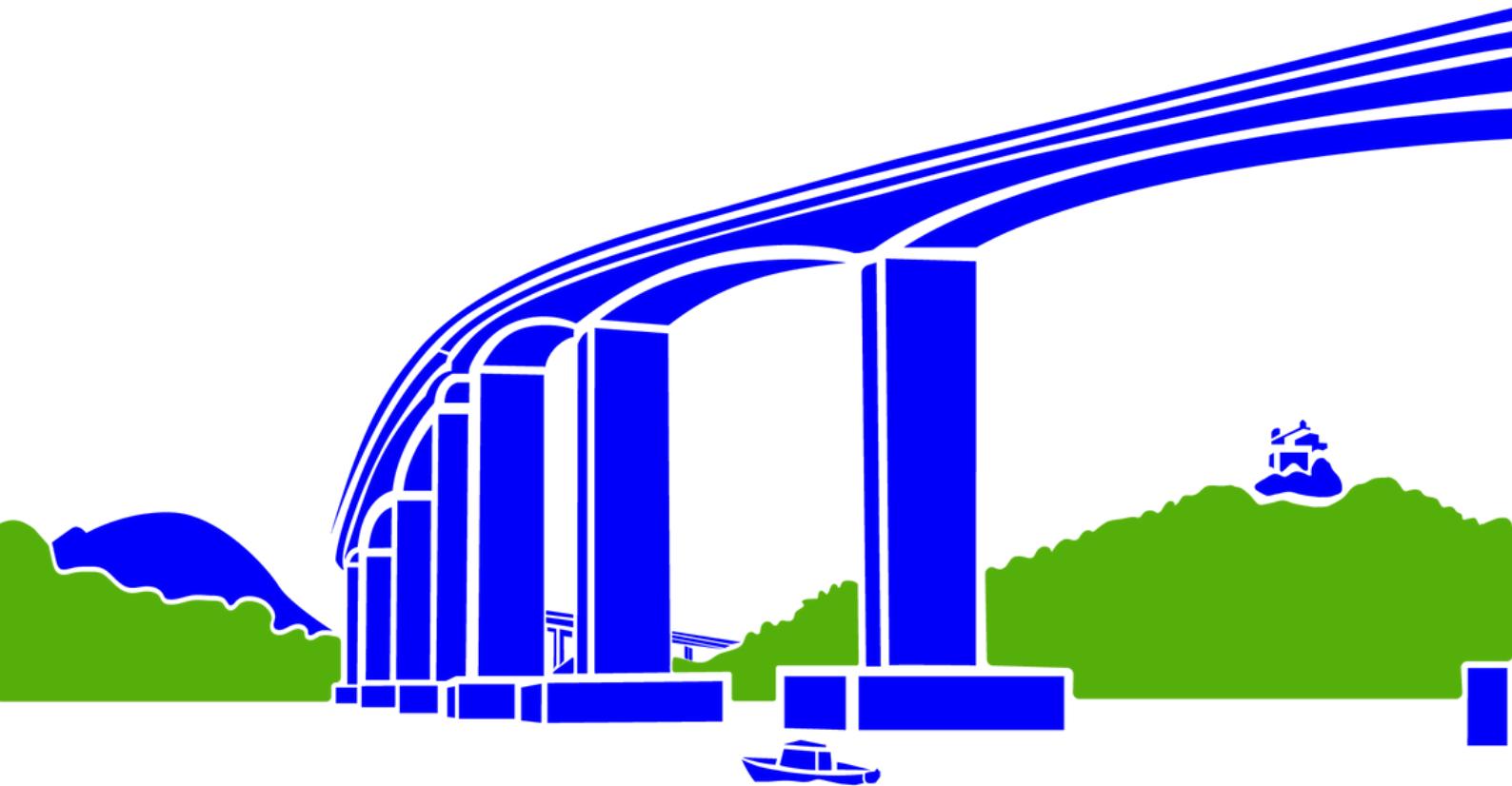
²Facultad de Matemáticas, Pontificia Universidad Católica de Chile, Chile, pedro.ramos@mat.uc.cl

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Comunicações Orais

**Estatística aplicada em
Agronomia e Biologia**





Avaliação dos métodos de estimativa de peso e digestibilidade em éguas no terço final da gestação

Marckis Lyandro Farias de Lima;¹

Resumo

O objetivo deste estudo foi avaliar a digestibilidade da dieta de éguas no terço final da gestação, alimentadas exclusivamente com forragem, e a utilização de fita de pesagem e equações da literatura para estimar o peso corporal. O experimento foi realizado na Universidade Federal da Paraíba, utilizando quatro éguas alimentadas com feno de capim-Tifton 85. O ensaio de digestibilidade durou cinco dias, e as fezes foram coletadas em dois turnos diários. O peso das éguas foi aferido semanalmente com balança comercial, fita métrica e equações, utilizando medidas biométricas (circunferência torácica, comprimento corporal, altura na cernelha, entre outras). A análise estatística incluiu testes de normalidade, homogeneidade de variâncias, e comparação dos dados com o teste de Kruskal-Wallis. O consumo de matéria seca foi, em média, de 1,57% do peso vivo, e a digestibilidade da matéria seca do feno foi de 56,18%. Os resultados mostraram que os métodos alternativos de pesagem diferem estatisticamente do peso real, mas ainda são úteis considerando sua praticidade e custo. A digestibilidade do feno foi considerada satisfatória e não impactou negativamente o desempenho das éguas. Conclui-se que, embora os métodos alternativos não sejam tão precisos quanto a balança, a fita métrica e as equações podem ser opções viáveis para estimar o peso de éguas gestantes, desde que aplicados com correções apropriadas.

Palavras-chave: equinos; gestante; morfometria; ganho de peso; condição corporal.

¹Departamento de Estatística e Informática da Universidade Federal Rural de Pernambuco (UFRPE), Recife - PE - marckislima19@gmail.com



Estratégias de Amostragem para uma Pesquisa Nacional Agropecuária Brasileira - PNAgro

Cristiano Ferraz ¹

João Eudes M. M. Torres ²

Resumo

Brasil é um país com tradição centenária em Censos Agropecuários, realizados usualmente a cada dez anos, desde 1920. No entanto, é fato que entre censos, não existe ainda uma Pesquisa Nacional Agropecuária (PNAgro) que forneça informações sustentadas por um método probabilístico de amostragem. Estudos metodológicos de viabilidade e eficiência para uma PNAgro podem tender a adotar uma estratégia herdada da tradição exitosa brasileira de pesquisas domiciliares, à exemplo da PNADC, cujo foco está na identificação de estabelecimentos agropecuários a partir de endereços. Neste trabalho, uma estratégia distinta é apresentada, recomendada pela FAO – Organização das Nações Unidas para a Agricultura e Combate à Fome, com foco na identificação pelo monitoramento de uso do solo. As principais vantagens desta estratégia são discutidas, e uma proposta de planejamento amostral que a utiliza, baseada no uso de um cadastro de área de segmento quadrado, é investigada. Os resultados de um estudo de simulação computacional preliminar, com base em informações do Censo Agropecuário de 2017 para o município de Sorriso -MT, são apresentados. Os números observados, em conjunto com a experiência prática de outros países que já adotam a estratégia, dão suporte à viabilidade da proposta para o Brasil.

Palavras-chave: Cadastro de área; Segmentos quadrados ponderados; Cadastro duplo.

¹ Departamento de Estatística, CASTLab – Laboratório de Estatística Agropecuária Computacional, Universidade Federal de Pernambuco – cferraz@castlab.org

² Coordenação de Informações Gerenciais - CIG, Universidade Federal de Pernambuco.



Índice de seleção MGIDI na avaliação de genótipos superiores de linhaça

Carlos Eduardo Forcelini Assoni^{1*}

Kamilly Siqueira Tonet¹

Lalesca Dadam Gomes¹

Luiz Paulo da Silva¹

Matheus Lopes Machado¹

Leonardo Talavera Campos²

Analu Mantovani³

Tiago Olivoto⁴

Resumo

O melhoramento genético da linhaça (*Linum usitatissimum* L.) é essencial para aumentar a resiliência das culturas frente à instabilidade climática e impulsionar a produção sustentável de grãos com elevado valor nutricional e comercial. Este estudo teve como objetivo aplicar o índice MGIDI (Multi-trait Genotype-Ideotype Distance Index) para a seleção de genótipos superiores com base em múltiplas características agronômicas. O experimento foi conduzido em Campos Novos (SC), em 2024, utilizando delineamento de blocos aumentados, com 41 genótipos (40 linhagens e uma testemunha comercial) de linhaça marrom e dourada. Foram mensuradas nove características agronômicas, incluindo rendimento de grãos por planta e componentes de rendimento como número de grãos e cápsulas por planta. A seleção foi realizada com o índice MGIDI em dois cenários: com e sem atribuição de pesos (peso 5 para massa de grãos e 1 para as demais), adotando-se uma pressão de seleção de 25%. As análises foram realizadas no software R. Os resultados demonstraram que o MGIDI foi eficiente na identificação de genótipos superiores, sendo G13 (marrom) e G69D (dourada) os genótipos que se destacaram. O G29 foi selecionado apenas com pesos, indicando a importância da ponderação. O G69D obteve o maior rendimento de grãos (5,69 g), superando a testemunha ST_PIONEIRA (4,16 g). Os maiores diferenciais de seleção foram observados para massa de grãos e massa de mil grãos, especialmente com pesos. Conclui-se que o uso do MGIDI, com ou sem ponderações, é uma ferramenta robusta e eficaz para apoiar decisões em programas de melhoramento da linhaça.

Palavras-chave: MGIDI; *Linum usitatissimum* L.; melhoramento genético; seleção.

¹ Discente do curso de Agronomia, Universidade Federal de Santa Catarina.

² Pós-doutorando, PPG/RGV, Universidade Federal de Santa Catarina.

³ Docente da Universidade do Oeste de Santa Catarina, Câmpus Campos Novos-SC.

⁴ Docente do Departamento de Fitotecnia, Universidade Federal de Santa Catarina.

* Autor correspondente: caduforcelini@gmail.com



Dissimilaridade genética e componentes de variância em genótipos de linhaça dourada e marrom

Tiago Olivoto^{1*}, Carlos Eduardo Forcelini Assoni², Lalesca Dadam Gomes², Luiz Paulo da Silva², Matheus Lopes Machado², Kenji da Cruz Konno², Isabela Martins Ghizoni², Maria Laura Faustino Monteiro², Kamilly Vitória Siqueira Tonet², Leonardo Talavera Campos³, Analu Mantovani⁴

Resumo

A linhaça (*Linum usitatissimum* L.) é uma cultura de crescente interesse agronômico e funcional, com aplicações na alimentação humana, animal e na indústria. Apesar do seu potencial, a produção de linhaça no Brasil ainda é limitada, reflexo da escassez de cultivares adaptadas e do baixo investimento em pesquisa e desenvolvimento. Para contribuir com o avanço dessa cultura, o Núcleo de Estudos e Pesquisas em Experimentação e Melhoramento Vegetal (NEPEM) iniciou um programa de melhoramento genético de linhaça no litoral sul do Brasil. Neste trabalho, avaliou-se a diversidade fenotípica e os componentes de variância de 40 genótipos oriundos de duas populações contrastantes (Dourada e Marrom), conduzidos em delineamento de blocos aumentados, utilizando a cultivar ST PIONEIRA como testemunha. Foram mensurados dez caracteres agronômicos, e as análises estatísticas foram realizadas por meio de modelos lineares mistos. As herdabilidades em sentido amplo (H^2) variaram de 0,485 a 0,837, com valores elevados para altura de planta ($H^2 = 0,837$) e número de cápsulas ($H^2 = 0,701$), indicando bom potencial de resposta à seleção. A variância intra-genotípica foi próxima de zero para a maioria das características, refletindo boa uniformidade fenotípica. A análise de agrupamento evidenciou variabilidade genética, mesmo se tratando de um painel com genótipos avançados, subsidiando a escolha estratégica de genótipos para cruzamentos dirigidos, como G69, G93 e G145 (Dourada), e G72, G13, G29 e G4 (Marrom). Os resultados demonstram o potencial do programa em gerar cultivares superiores e adaptadas às condições brasileiras, contribuindo para o fortalecimento da cadeia produtiva da cultura no país.

Palavras-chave: *Linum usitatissimum* L; alimento funcional; análise multivariada; melhoramento vegetal.

¹ Docente do Departamento de Fitotecnia, Universidade Federal de Santa Catarina.

² Discente do curso de Agronomia, Universidade Federal de Santa Catarina.

³ Pós-doutorando, PPG/RGV, Universidade Federal de Santa Catarina.

⁴ Docente da Universidade do Oeste de Santa Catarina, Câmpus Campos Novos-SC.

* Autor correspondente: tiago.olivoto@ufsc.br



Seleção de Índices de vegetação para predição da produtividade com VSURF

Lucas Coelho da Silva ¹ Luiz Alexandre Peternelly ²

Resumo

Os índices de vegetação (IVs) são combinações matemáticas de duas ou mais bandas espectrais, RGB, RedEDGE e NIR. Dada a grande variedade de possíveis IVs, surge a necessidade de selecionar aqueles que estão mais associados à produtividade. O *Variable Selection Using Random forest* (VSURF) utiliza uma estratégia de seleção baseada no algoritmo Random Forest (RF), empregando o método stepwise com base na importância das variáveis. Este estudo utiliza o algoritmo VSURF para selecionar os IVs em cada estádio fenológico da soja mais associados a produtividade e comparar o desempenho preditivo das variáveis selecionadas com todas as variáveis, pelo algoritmo da RF. O experimento foi conduzido em delineamento de blocos casualizados, com 297 parcelas, distribuídas em três blocos. Foram capturadas 11 imagens, aproximadamente uma vez por semana, ao longo do ciclo da soja, a partir das quais foram calculados 36 IVs. A validação dos modelos foi realizada por meio de validação cruzada em 10 etapas, utilizando acurácia e área sob a curva ROC (AUC) como métricas de avaliação. Entre as variáveis mais selecionadas, destacam-se o NIR, NGBDI, o SIPI e o TGI. O modelo com variáveis selecionadas obteve acurácia de 0,63 e AUC de 0,67, comparado a 0,62 e 0,65 do modelo completo. A fase do início do enchimento dos grãos (R_5) apresentou o melhor desempenho preditivo, com acurácia de 0,72 e AUC de 0,79. O VSURF demonstrou-se eficaz na seleção de variáveis, uma vez que manteve o desempenho preditivo semelhante ao modelo completo.

Palavras-chave: classificação; curva ROC; multiespectral; random forest; sensoriamento remoto.

¹ Departamento de Estatística, Universidade Federal de Viçosa – lucas.silva2@ufv.br

² Departamento de Estatística, Universidade Federal de Viçosa – peternelly@ufv.br



Árvores filogenéticas de *Bacillus nitratireducens* obtidas por modelos probabilísticos evolutivos usando inferência clássica e bayesiana, com marcadores multi-locus (16S rRNA e gyrB).

Elisângela Aparecida da Silva Lizzi¹; Marcus Vinicius da Silva Rodrigues²³; Sandriele Aparecida Noriler²; Glaucia Maria Bressan¹; Admilton Gonçalves de Oliveira Junior²

Resumo

Este estudo investigou as relações filogenéticas de linhagens de *Bacillus nitratireducens* com abordagens estatísticas computacionais avançadas de modelos evolutivos usando inferência clássica e bayesiana. Foram analisados 19 genomas (14 do GenBank e 5 inéditos) e sequências dos genes 16S rRNA (conservado) e gyrB (variável), com *B. proteolyticus* como grupo externo. Os modelos probabilísticos evolutivos de substituição como TN+F para gyrB; HKY para 16S rRNA e revelaram alta concordância entre as topologias geradas por genomas completos e genes concatenados, com apoio computacional do IQTree e MrBayes, com uso de estimadores de máxima verossimilhança e simulação de Monte Carlo em Cadeias de Markov. A análise demonstrou que todas as linhagens pertencem à mesma espécie, com valores de suporte robustos. O gene gyrB forneceu maior resolução filogenética intraespécie em comparação ao 16S rRNA, cuja baixa variabilidade limitou a discriminação de linhagens próximas. Os resultados destacam a eficácia de abordagens multi-locus combinando marcadores com distintas taxas evolutivas, com aplicações em taxonomia microbiana e biotecnologia. A robustez e concordância das análises filogenéticas usando inferência bayesiana baseada em simulações de MCMC mostrou uma análise conservadora para reconstruir as relações evolutivas entre as linhagens estudadas, em comparação e complemento com métodos evolutivos de inferência clássica. Conclui-se que esses resultados destacam como a integração de métodos estatísticos avançados, inferência clássica e bayesiana dentro do contexto da microbiologia, com marcadores genéticos complementares mostra a elucidação de relações evolutivas em *Bacillus nitratireducens* usando modelos probabilísticos evolutivos.

Palavras-chave: Modelos probabilísticos evolutivos; Árvores Filogenéticas; Inferência Clássica e Bayesiana; *Bacillus nitratireducens*

¹ Departamento Acadêmico de Matemática e Programa de Pós Graduação em Bioinformática, UTFPR- Cornélio Procópio – elisangelalizzi@gmail.com; galbressan@gmail.com

² Programa de Pós Graduação em Bioinformática, UTFPR- Cornélio Procópio – marcusr@alunos.utfpr.edu.br

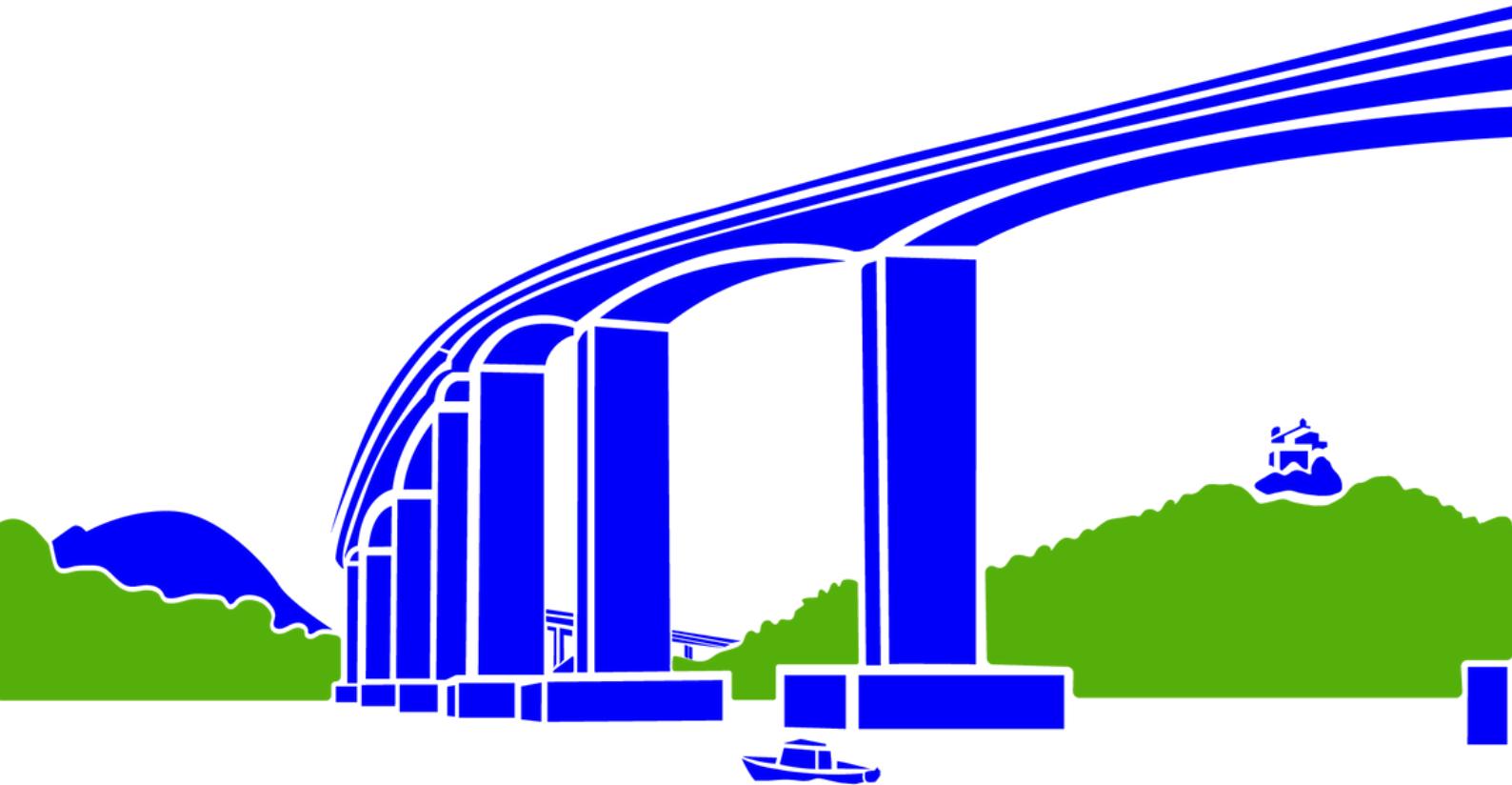
³ Laboratório de Biotecnologia Microbiana (LABIM), Universidade Estadual de Londrina (UEL), Londrina-PR – norilersnd@gmail.com; admilton@uel.br

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Comunicações Orais

**Métodos estatísticos em
Agronomia e Biologia**





Parametrizações do modelo de von Bertalanffy para descrição de curvas de crescimento

Felipe Augusto Fernandes ¹

Resumo

As curvas de crescimento de animais, em geral, possui formato de “S”, conhecidas também como curvas sigmoidais. Este tipo de curva é bem ajustada por modelos de regressão não linear, dentre eles o de von Bertalanffy que tem sido muito aplicado em diversas áreas, sendo apresentado na literatura por meio de diferentes parametrizações, que na prática, pode além de complicar seu entendimento, afetar as medidas de não linearidade e as inferências sobre os parâmetros. Para quantificar a não linearidade presente em um modelo Bates e Watts utilizaram um conceito geométrico de curvatura. O objetivo deste trabalho foi desenvolver analiticamente três parametrizações do modelo não linear de von Bertalanffy, referente à sua não linearidade, as implicações nas inferências e estabelecer relações entre os parâmetros nas diferentes formas de expressar os modelos. Estas as parametrizações foram ajustadas à dados de crescimento de ovinos. Para cada parametrização foram calculadas as medidas de curvatura intrínseca e paramétrica descritas por Bates e Watts. A escolha da parametrização afeta as medidas de não linearidade, consequentemente, influencia na confiabilidade e nas inferências sobre os parâmetros estimados. As formas mais utilizadas na literatura apresentaram os maiores afastamentos da linearidade, evidenciando a importância de se analisar estas medidas em qualquer estudo sobre curva de crescimento. Deve ser utilizada a parametrização na qual a estimativa de b representa a abscissa do ponto de inflexão por apresentar menores desvios de linearidade e interpretação biológica direta para todos os parâmetros.

Palavras-chave: Interpretação biológica; medidas de curvatura; regressão não linear; reparametrizações

¹Departamento de Estatística, Universidade Estadual de Campinas - fernandesfelipe@gmail.com



Determinação da Origem Geográfica de Amostra de Maconha Utilizando SVM e Análise de Isótopos Estáveis: Uma Abordagem de Aprendizado de Máquina

Hozana Francielle do Nascimento Borges;¹ Luiz Antonio Martinelli;² Paulo José Duarte Neto;³ Antonio Samuel Alves da Silva;⁴ Francisco Gustavo da Silva;⁵

Resumo

Este estudo investigou a aplicação do Support Vector Machine (SVM) como classificador para a determinação da origem geográfica de amostras de maconha (*Cannabis sativa*), combinando técnicas de aprendizado de máquina com análise de isótopos estáveis. Foram testados quatro kernels do SVM: linear, sigmoide, polinomial e radial, com tamanhos de amostra variando de 50 a 500. O kernel sigmoide, com tamanho de amostra 50, apresentou o melhor desempenho, alcançando uma acurácia de 0,79. Além disso, foram calculadas métricas como Precisão Média, Recall Médio, F1-Score Médio, Desvio Padrão (SD) da Acurácia, e intervalos de confiança (inferior e superior) para avaliar a robustez do modelo. O kernel sigmoide destacou-se não apenas pela acurácia, mas também por valores elevados de Precisão Média, Recall Médio e F1-Score Médio, com baixo desvio padrão e intervalos de confiança consistentes. Conclui-se que o SVM com kernel sigmoide é uma abordagem eficaz e confiável para a classificação da origem geográfica de amostras de maconha, sendo uma ferramenta promissora para aplicações forenses e científicas.

Palavras-chave: Análise isotópica; Aprendizado de Máquina; Cannabis; Biomass.

¹Programa de Pós Graduação em Biometria e Estatística Aplicada, UFRPE - hozana_francielle@hotmail.com

²Centro de Energia Nuclear na Agricultura, USP - martinelli@cena.com.br

³Programa de Pós Graduação em Biometria e Estatística Aplicada, UFRPE - pjduarteneto@gmail.com

⁴Programa de Pós Graduação em Biometria e Estatística Aplicada, UFRPE - antonio.sasilva@ufrpe.br

⁵Programa de Pós Graduação em Biometria e Estatística Aplicada, UFRPE - fr.gustavosilva.098@gmail.com



Análise de sobrevivência com riscos competitivos aplicada à duração da proteção de cultivares

Verônica Manhães Saint'Clair ¹

Marciel Lelis Duarte ²

Sebastião Martins Filho ³

Resumo

A proteção de cultivares assegura direitos de propriedade intelectual sobre novas variedades vegetais por meio do Certificado de Proteção de Cultivar. No entanto, eventos como renúncia, cancelamento e anulação podem encerrar essa proteção antes do prazo legal, configurando um cenário de riscos competitivos. Esta pesquisa investiga como as características do titular (instituição pública, setor privado e parcerias) e do cultivar (presença de transgenia e tipo de cultura), influenciam a duração do certificado. Foram analisadas 3421 proteções obtidas da plataforma *CultivarWeb* do Ministério da Agricultura, Pecuária e Abastecimento, entre 1997 e 2024. O evento de interesse (expiração por prazo) e os eventos competitivos (renúncia, cancelamento e anulação da proteção) foram analisados por meio do modelo de Fine-Gray. Os resultados indicaram que cultivares transgênicas apresentam menor incidência tanto de expiração quanto dos eventos competitivos. Proteções do setor privado têm menor incidência de expiração, mas com maior propensão à renúncia, cancelamento e anulação. Cultivares ornamentais registraram as maiores incidências de renúncia, independentemente do ciclo da cultura. Dessa forma, conclui-se que características do titular e da cultivar impactam a duração da proteção e que o modelo de Fine-Gray se mostra promissor para modelar a duração do certificado de proteção de cultivares.

Palavras-chave: dados censurados; modelo de Fine-Gray; propriedade intelectual.

¹ Departamento de Estatística, Universidade Federal de Viçosa – veronica.saintclair@ufv.br

² Departamento de Estatística, Universidade Federal de Viçosa – marciel.duarte@ufv.br

³ Departamento de Estatística, Universidade Federal de Viçosa – martinsfilho@ufv.br



Utilização de modelos não lineares na avaliação do acúmulo de nitrogênio em plantas de chicória sob diferentes formas de cultivo

Gabriel Edson Sousa da Silva ¹; Louziane Ribeiro Carvalho ²; Joel Augusto Muniz ³;
Edilson Marcelino Silva ⁴

Resumo

Este estudo avaliou o acúmulo de nitrogênio em plantas de chicória sob diferentes formas de cultivo, utilizando modelos não lineares. A chicória é uma hortaliça importante, rica em inulina, com potencial para a saúde humana. O nitrogênio é um nutriente essencial para o crescimento das plantas, influenciando diretamente a fotossíntese e o desenvolvimento. Modelos não lineares, como Logístico, Gompertz e Von Bertalanffy, foram utilizados para analisar o acúmulo de nitrogênio ao longo do tempo. O experimento foi conduzido com duas cultivares de chicória, com e sem cobertura de tecido de polipropileno. Os resultados indicaram que o modelo Logístico foi o mais adequado para descrever o acúmulo de nitrogênio na maioria das cultivares. Não houve diferença significativa no acúmulo máximo de nitrogênio entre as cultivares, mas houve diferença no ponto de inflexão entre a cultivar AF-218 coberta e a crespa descoberta.

Palavras-chave: Chicória; Macronutrientes; Modelos não lineares; Acúmulo de nitrogênio; Polipropileno.

¹Departamento de Estatística, UFLA - gabriel.silva66@estudante.ufla.br

²Departamento de Estatística, UFLA - louziane.carvalho1@estudante.ufla.br

³Departamento de Estatística, UFLA - joamuniz@ufla.br

⁴Departamento de Matemática, UFRRJ - edilsonest@ufrrj.br



Aplicação de Técnicas de Tempo até o Evento com Censura

Intervalar na Germinação de Sementes

Sara Silvério ¹

Marciel Lelis Duarte ¹

Carla Regina Guimarães Brighenti ²

Lausanne Soraya de Almeida ³

Sebastião Martins Filho ¹

Resumo

Devido à presença de dados censurados nos ensaios de germinação, métodos convencionais de análise de dados podem não ser a escolha mais apropriada. Uma alternativa é a aplicação da análise de sobrevivência. Entretanto, em alguns casos, o momento exato da germinação não é conhecido, pois as observações ocorrem em intervalos de tempo, caracterizando censura intervalar. Neste trabalho, foram utilizados dois conjuntos de dados distintos sobre a germinação de sementes de pitaia (*Hylocereus* spp.) com o objetivo de exemplificar e avaliar a aplicação de diferentes abordagens na análise de sobrevivência intervalar paramétrica, não paramétrica e semiparamétrica. O primeiro conjunto refere-se a um experimento que analisou o efeito de diferentes temperaturas sobre sementes armazenadas por 12, 13 e 14 meses. O segundo envolveu a germinação de sementes armazenadas por diferentes períodos, em dois ambientes: câmara fria e condições ambiente. A partir dos resultados obtidos com a abordagem não paramétrica, a 15 °C a germinação foi significativamente inferior e diminuiu com o aumento do tempo de armazenamento. Na análise paramétrica a interação entre tempo e local de armazenamento foi significativa (-0,1593), indicando que a câmara fria reduz o impacto do tempo de armazenamento nas sementes, em comparação com as condições ambiente. Na análise semiparamétrica a razão de taxas de falha de aproximadamente 1,174 ($\exp(0,16)$) indica que, a cada mês adicional de armazenamento, o risco de germinação das sementes aumenta aproximadamente 17,4% quando as sementes são armazenadas em câmara fria. Os métodos utilizados mostraram-se valiosos para analisar a germinação das sementes com censura intervalar.

Palavras-chave: *Hylocereus* spp; armazenamento; temperatura; análise de sobrevivência.

¹ Departamento de estatística, UFV – sara.silverio@ufv.br



Modelos multi-estados para transições duplas associadas ao parasitismo no controle biológico

Idemauro Antonio Rodrigues de Lara;¹ Gabriel Rodrigues Palma;² Victor José Bon;³
Carolina Reigada;⁴ Rafael de Andrade Moral.⁵

Resumo

O percevejo marrom *Euschistus heros* é uma das pragas potenciais da cultura da soja. Neste trabalho, apresenta-se um estudo experimental desenvolvido para avaliar as mudanças comportamentais dos parasitoídeos de acordo com a qualidade dos hospedeiros, ou seja, quando previamente parasitados ou não, bem como os efeitos do parasitismo prévio dos ovos do *Euschistus heros* na taxa de parasitismo das espécies *Trissolcus basalis* e *Telenomus podisi*. Assim, foram implementados modelos markovianos multi-estados para modelar sucessivamente a escolha dos ovos (não parasitados, parasitados por *T. podisi* e parasitados por *T. basalis*) e o comportamento condicional dada a escolha (caminhar, tamborilar, ovipositar ou marcar). Neste trabalho, enfatizamos a contribuição metodológica de transições duplas devido a dois estágios sucessivos de respostas ao longo do tempo. A metodologia estatística é baseada em processos estocásticos de tempo contínuo e no procedimento de máxima verossimilhança. Usando o modelo de Cox e assumindo um processo estacionário, verificou-se que o efeito do tratamento foi significativo para a escolha, indicando que as duas espécies de parasitoídeos têm padrões de escolha diferentes. Numa segunda etapa, os resultados também mostraram a influência da espécie no comportamento condicional, especialmente, que o agente *T. podisi* evita a competição intraespecífica. O método estatístico usado contribuiu para seleção do melhor agente de controle biológico da praga da soja, que é uma prática da agricultura sustentável e ecologicamente correta.

Palavras-chave: *Telenomus podisi*; *Trissolcus basalis*; comportamento de forrageamento; controle do percevejo marrom, processos estocásticos.

¹Escola Superior de Agricultura Luiz de Queiroz, USP, Piracicaba SP, idemauro@usp.br

²Department of Mathematics and Statistics, Maynooth Ireland, Gabriel.Palma.2022@mumail.ie

³Escola Superior de Agricultura Luiz de Queiroz, USP, Piracicaba SP,- victorbon@usp.br

⁴Departamento de Biologia, Universidade Federal de São Carlos, creigada@ufscar.br

⁵Department of Mathematics and Statistics, Maynooth Ireland, rafael.deandrademoral@mu.ie

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Comunicações Orais

**Modelos de Regressão, MLG
e modelos não lineares**





Identificação de Ineficiências de Licitações Públicas por Meio do Modelo de Mistura de Regressões Binomiais Negativas.

Emerson Antonio Rocha Pazeto
Poliano Bastos da Cruz ¹
Alessandro José Queiroz Sarnaglia
Emerson Antonio Rocha Pazeto
Fabio Alexander Fajardo Molinares ²

Resumo

Nos últimos anos, o setor público brasileiro tem passado por transformações relevantes sob o ponto de vista da transparéncia e *compliance*. Por isso, os gastos públicos cada vez mais têm sido objeto de interesse da sociedade. Este trabalho focou na análise estatística de processos licitatórios, mais especificamente na discriminação não supervisionada entre processos “eficientes” e “ineficientes”. Foi utilizado como indicador de ineficiência o número de aditivos contratuais. Após a coleta e preparação dos dados, foram identificadas evidências de sobredispersão e excesso de zeros. A fim de resolver esse problema, propôs-se um modelo de Misturas de Regressões Binomiais Negativas, o qual foi ajustado com a aplicação do algoritmo EM. Esse algoritmo tornou o modelo capaz de classificar as observações em dois grupos com as características desejadas. Com o fito de mostrar a efetividade do modelo, foram realizadas comparações com modelos de regressões Poisson, Binomial Negativa, Poisson Inflada de Zeros e Binomial Negativa Inflada de Zeros. Os resultados mostraram que o modelo proposto se ajustou satisfatoriamente aos dados e apresentou o melhor desempenho associado à parcimônia em termos do Critério de Informação Bayesiano. Esse resultado torna-se interessante na medida que o modelo proposto tem potencial de se transformar em uma ferramenta de gestão pública capaz de identificar probabilidades de um processo ser “ineficiente” antes mesmo de ser contratado.

Palavras-chave: Aditivos, Ineficiência, Regressões, Classificação não Supervisionada, processos de contratações ineficientes, Infladas de Zeros, Binomial Negativa, Poisson, Algoritmo Expectation-Maximization.

¹PPG Acadêmico - Fucape Business School, sediada em Vitória-ES - emebompaz@gmail.com

²Departamento de Estatística, Universidade Federal do Espírito Santo (UFES), sediado em Vitória-ES - alessandro.sarnaglia@ufes.br



Overdispersion models for clustered toxicological data in a bioassay of entomopathogenic fungus

Silvia M. Freitas;¹ Lida Fallah;² Clarice G.B. Demétrio;³ John Hinde⁴

Resumo

We consider discrete mortality data for groups of individuals observed over time. The fitting cumulative mortality curves as a function of time involves the longitudinal modelling of the multinomial response. Typically such data exhibit greater variation than predicted by the multinomial distribution (overdispersion). To model the extra-multinomial variation we consider a Dirichlet-multinomial model, a random intercept model and a random intercept and slope model. We construct asymptotic and robust covariance matrix estimators for the regression parameter standard errors. Applying these models to a specific insect bioassay of the fungus *Beauveria bassiana*, we note some simple relationships in the results and explore why these are simply a consequence of the data structure. Fitted models are used to make inferences on the effectiveness and consistency of different isolates of the fungus to provide recommendations for its use as a biological control in the field.

Palavras-chave: Dirichlet-multinomial; Extra-multinomial variation; Generalized estimating equations; Generalized linear models; Random effects models..

¹Department of Statistics and Applied Mathematics, Federal University of Ceará, Fortaleza, CE, Brazil - silvia@dema.ufc.br

²Trinity Translational Medicine Institute and School of Computer Science and Statistics, Trinity College Dublin, Ireland - l.fallah22@gmail.com

³Department of Exact Sciences, São Paulo University, ESALQ, Piracicaba, SP, Brasil - clarice.demetrio@usp.br

⁴School of Mathematical and Statistical Sciences, University of Galway, Ireland - john.hinde@universityofgalway.ie



Bessel regression and **bbreg** package for bounded data.

Vinícius D. Mayrink ¹
Wagner Barreto-Souza ²
Alexandre B. Simas ³

Resumo

Beta regression is widely used for modeling continuous bounded data, with few competitors offering similar flexibility. The normalized inverse-Gaussian (N-IG) process, a Bayesian alternative to the Dirichlet process, has attracted growing interest; however, its univariate distribution remains largely unexplored in the context of classical inference. This study introduces bessel regression, based on the univariate N-IG, as a novel alternative to the beta regression model. Parameter estimation is performed via an EM algorithm, and inference procedures are outlined. We propose a practical model selection criterion to distinguish between bessel and beta regressions. Additionally, we present the R package **bbreg** for fitting both models, which includes tools for assessing model adequacy and selection. A simulation study evaluates performance under model misspecification, and an empirical example illustrates the comparison between bessel and beta regressions using **bbreg**. The first author acknowledges research support from CNPq, CAPES, and FAPEMIG.

Palavras-chave: beta regression; EM algorithm; normalised inverse-Gaussian distribution; misspecification; model selection.

¹Departamento de Estatística, ICEX, Universidade Federal de Minas Gerais, Brasil - vdm@est.ufmg.br

²University College Dublin, Irlanda - wagner.barreto-souza@ucd.ie

³King Abdullah University of Science and Technology, Arábia Saudita - alexandre.simas@kaust.edu.sa



Regression models for binary data with scale mixtures of centered skew-normal link functions

João Victor Bastos de Freitas ¹

Caio Lucidius Naberezny Azevedo ²

Resumo

For binary regression, the use of symmetric link functions is not appropriate when we have evidence that the probability of success increases at a different rate than decreases. In these cases, the use of link functions based on the cumulative distribution function of a skewed and heavy-tailed distribution can be useful. The most popular choice is some scale mixtures of skew-normal distributions. This family of distributions can have some identifiability problems caused by the so-called direct parameterization. To circumvent this problem, in this paper we proposed link functions based on the scale mixtures of skew-normal distributions under the centered parameterization. Also, in binary modelling with skewed link functions, we may have another identifiability problem caused by the presence of the intercept and skewness parameters. Furthermore, we proposed to fix the sign of the skewness parameter, which is a new perspective in the literature to deal with the identifiability problem in skewed link functions. Bayesian inference using MCMC algorithms and residual analysis are developed. Simulation studies are carried out to evaluate the performance of the model. The methodology is applied to a heart disease data.

Palavras-chave: Binary regression; Scale mixtures of skew-normal distributions; Centered parameterization; Bayesian inference; Skewed link functions.

¹ Departamento de Estatística, Instituto de Matemática, Estatística e Computação Científica, Universidade Estadual de Campinas, Campinas, São Paulo, Brasil - joao.victor.ufc13@gmail.com

² Departamento de Estatística, Instituto de Matemática, Estatística e Computação Científica, Universidade Estadual de Campinas, Campinas, São Paulo, Brasil - cnaber@unicamp.br



Sandwich variance estimation in scale mixture of skew-normal linear mixed models

Keyliane Travassos ¹; Larissa A. Matos ¹; Fernanda L. Schumacher ²

Abstract

Linear mixed models are often used to analyze studies with repeated measurements over time, where the data present a correlation between observations of the same individual. However, one of the main challenges in using these models is correctly defining the covariance structure associated with error and random effects. In this context, we obtain the sandwich variance estimator as a robust alternative to estimate the standard error in the scale mixture of skew-normal linear mixed models. We present the definition of the estimator, its derivation for this model, and its practical implementation. In addition, through simulation studies, we evaluate its performance in different scenarios, demonstrating its ability to provide reliable inferences even when the model or the covariance structure is misspecified.

Keywords: sandwich variance estimator; linear mixed-effects model; scale mixture of skew-normal distributions.

¹Department of Statistics, Universidade Estadual de Campinas, Campinas, São Paulo, Brazil

¹Department of Statistics, Universidade Estadual de Campinas, Campinas, São Paulo, Brazil

²Division of Biostatistics, College of Public Health, The Ohio State University, Columbus, OH, U.S.A



A robust quantile regression for bounded variables based on the Kumaraswamy Rectangular distribution

Matheus O. Castro;¹ Caio L. N. Azevedo² Juvêncio S. Nobre³

Abstract

Quantile regression (QR) models offer an interesting alternative compared with ordinary regression models for the response mean. Besides allowing a more appropriate characterization of the response distribution, the former is less sensitive to outlying observations than the latter. Indeed, the QR models allow modeling other characteristics of the response distribution, such as the lower and/or upper tails. However, in the presence of outlying observations, the estimates can still be affected. In this context, a robust quantile regression model for bounded responses is developed, considering the Kumaraswamy Rectangular (KR) distribution. The KR model corresponds to a finite mixture structure similar to the Beta Rectangular distribution. That is, the KR distribution has heavier tails compared to the Kumaraswamy model. Indeed, we show that the correspondent KR quantile regression model is more robust and flexible than the usual Kumaraswamy one. Bayesian inference, which includes parameter estimation, model fit assessment, model comparison, and influence analysis, is developed through a hybrid-based MCMC approach. To link both conditional and marginal quantiles in terms of a regression structure, a two-step estimation algorithm under a Bayesian approach is proposed to obtain samples from the posterior distributions. Our proposal showed to be robust against outlying observations related to the response without adding too much complexity to the estimation process. Its performance is validated through three simulation studies, highlighting its accuracy and flexibility. Finally, we apply the proposed model to two real datasets involving socio-economic indicators in Brazil, demonstrating its practical utility and comparative advantages over existing alternatives.

Keywords: Kumaraswamy Rectangular; Double bounded response; Bayesian inference; MCMC algorithms; Latent variables.

¹Instituto de Matemática, Estatística e Computação Científica, Universidade Estadual de Campinas - m234976@dac.unicamp.br

²Instituto de Matemática, Estatística e Computação Científica, Universidade Estadual de Campinas - cnaber@unicamp.br

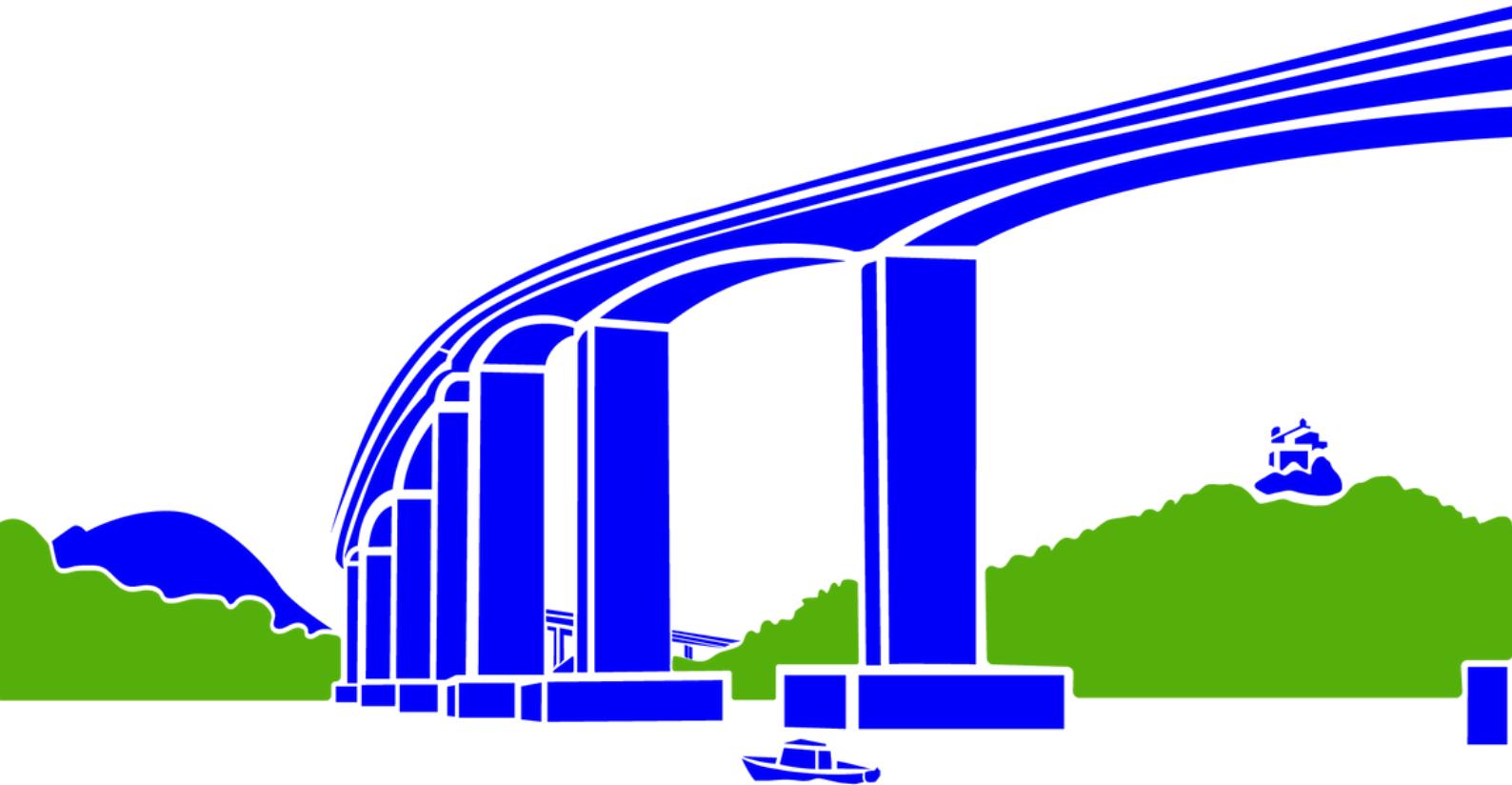
³Departamento de Estatística e Matemática Aplicada, Universidade Federal do Ceará - juvencio@ufc.br

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Comunicações Orais

**Séries Temporais, Análise espacial
e Análise Espaço-Temporal**





Bayesian Hierarchical Modeling for Predicting Spatially Correlated Curves in Irregular Domains with Application to PM₁₀ Air Pollution Data.

Alvaro Alexander Burbano Moreno ¹

Abstract

This study presents a Bayesian hierarchical model for analyzing spatially correlated functional data and handling irregularly spaced observations. The model uses Bernstein polynomial (BP) bases combined with autoregressive random effects, allowing for nuanced modeling of spatial correlations between sites and dependencies of observations within curves. Moreover, the proposed procedure introduces a distinct structure for the random effect component compared to previous works. Simulation studies conducted under various challenging scenarios verify the model's robustness, demonstrating its capacity to accurately recover spatially dependent curves and predict observations at unmonitored locations. The model's performance is further supported by its application to real-world data, specifically PM₁₀ particulate matter measurements from a monitoring network in Mexico City. This application is of practical importance, as particles can penetrate the respiratory system and aggravate various health conditions. The model effectively predicts concentrations at unmonitored sites, with uncertainty estimates that reflect spatial variability across the domain. This new methodology provides a flexible framework for the FDA in spatial contexts and addresses challenges in analyzing irregular domains with potential applications in environmental monitoring.

Palavras-chave: Bayesian Hierarchical Modeling; Spatial Functional Data; Irregular Sampling; Bernstein Polynomials.

¹Departamento de Estatística, IMECC, Universidade Estadual de Campinas - aamoreno@unicamp.br



Spatial distribution of multifractal parameters to characterize the formations of *Sargassum* spp.

Francisco Gustavo-Silva¹, Cristele Chevalier², Marine Laval³, Luc Courtrai⁴, Alex Costa de Silva⁵, Jacques Descloitres⁶, Yamina Aimene⁷, Adan Salazar⁸, Audrey Minghelli⁹, Vincent Vantrepotte¹⁰, Paulo Duarte-Neto¹¹

Abstract

This study proposes an innovative approach to analyzing aggregations of *Sargassum* spp. in the Western Mid-Atlantic Region (WCWA) using multifractal analysis of segmented images from the OLCI sensor of the Sentinel-3 satellite. Using products from the SargAlert project, classified into presence or absence of Sargassum, the methodology applied multifractal analysis to 289 image partitions to extract parameters such as dominant singularity (α_0), spectrum asymmetry ($f(\Delta\alpha)$), spectrum width ($\Delta\alpha$) and lacunarity (Λ). The spatial distribution of these parameters made it possible to identify distinct morphological patterns associated with the density and organization of the aggregations. The results showed that the parameter α_0 correlates with the amount of cover, while negative values of $f(\Delta\alpha)$ were associated with dense, spiral structures, possibly ocean vortices. The analysis also revealed morphologically coherent groupings such as: absence of algae (SN), low density (QSN), dispersed distribution (SD), intense agglomeration (PREV), and spiral structures (VRT). The detection of VRT groups highlighted the multiscale complexity of these formations and their association with multiple oceanographic processes, such as turbulence and detachment of algal masses. This work contributes to advancing methods for the detection and spatial interpretation of Sargassum aggregations, offering new metrics for understanding ocean dynamics and mitigation strategies.

keywords: Dominant singularity; Vortices; Ocean dynamics; Central-West Atlantic

¹Departamento de Estatística e Informática - DEINFO/UFRPE, Recife - Pernambuco, Brasil - fr.gustavosilva.098@gmail.com

²Mediterranean Institute of Oceanography (MIO), IRD, Marseille, France - cristele.chevalier@mio.osupytheas.fr

³Laboratoire des Matériaux et Molécules en Milieu Agressif (L3MA), Schoelcher, France - marine.laval@mio.osupytheas.fr

⁴IRISA, Université de Bretagne Sud, Vannes, France - luc.courtrai@univ-ubs.fr

⁵Departamento de Oceanografia da Universidade Federal de Pernambuco – DOCEAN/UFPE, Recife - Pernambuco, Brasil - alex.csilva@ufpe.br

⁶IRISA, Université de Bretagne Sud, Vannes, France - Jacques.Descloitres@univ-lille.fr

⁷Laboratoire des Matériaux et Molécules en Milieu Agressif (L3MA), Schoelcher, France - yamina.aimene@univ-antilles.fr

⁸Mexican Space Agency (AEM), Ciudad de Mexico, Mexico - asalazar.cdmx.2023@gmail.com

⁹Laboratoire d’Informatique et Système (LIS), Toulon, France - audrey.minghelli@univ-tln.fr

¹⁰Laboratoire d’Océanologie et de Géosciences - LOG, Wimereux, France - vincent.vantrepotte@univ-littoral.fr

¹¹Departamento de Estatística e Informática - DEINFO/UFRPE, Recife - Pernambuco, Brasil - pjduarteneto@gmail.com



Terrorism in Cabo Delgado, Mozambique: A Spatio-Temporal Analysis of Fatalities and Their Determinants (2017–2024) with Forecasts to 2030

Alex Nhancololo ¹; Airlane Pereira ²; Giovani Festa Paludo ³; Katerine Zuniga ⁴; André Cuinica ⁵; Welson Oliveira ⁶; Gean Damaceno ⁷; Elias Sabe ⁸; Valderine Ferreira ⁹; Minelda Arao Lhone ¹⁰; João Scalon ¹¹

Abstract

Since 2017, Mozambique's Cabo Delgado province has experienced sustained insurgent violence. While the exploitation of mineral resources is frequently cited as a primary driver, this study challenges such oversimplified narratives through a spatio-temporal analysis spanning 2017–2024. By integrating Armed Conflict Location and Event Data with demographic variables from the U.S. Census International Database, the study employs Spatial Autoregressive with Autoregressive Disturbances and Geographically Weighted Regression models. These are combined with Generalized Autoregressive Moving Average forecasts to identify associated factors and project fatality trends through 2030. Results reveal that 76.9% of fatalities cluster near strategic resource hubs such as Palma's gas projects, Balama's graphite mines, and Mocímboa da Praia's deep-water port, as well as in artisanal gold mining areas (Muidumbe and Nangade), coastal zones, and international borders. The Spatial Lag of X model (AIC: 8750.19) highlights border dynamics: Nampula border districts exhibit 36.2% lower fatality rates, while spillover effects amplify risks near internal borders and the Tanzanian border. Non-state armed groups including the Dyck Advisory Group, Islamic State Mozambique, and Southern African Development Community forces—demonstrate significantly higher fatality rates compared to Rwandan forces. Population growth correlates with increased fatalities, though socio-demographic factors, such as a higher female-to-male ratio, marginally attenuate this effect. Forecasts suggest declining post-2024 fatalities, signaling potential stabilization; however, biases from political underreporting necessitate caution.

Keywords: Armed conflict in northern Mozambique; fatality rate modeling; SARAR; GARMA; Impacts of state and non-state armed forces in Cabo Delgado.

¹Alex Monito Nhancololo, São Paulo - amnhancololo@gmail.com

²Airlane Alencar Pereira, São Paulo - lane@ime.usp.br

³Giovani Festa Paludo, Lavras - gfpaludo@gmail.com

⁴Katerine Zuniga, São Paulo - zunigalastrakaterine@gmail.com

⁵André Silvestre Cuinica, Mozambique - andre.cuinica@gmail.com

⁶Welson Antônio de Oliveira, Lavras - welson.oliveira2@estudante.ufla.br

⁷Gean Damaceno, Lavras - gean.damaceno1@estudante.ufla.br

⁸Elias Sabe, Lavras - sabehelio@gmail.com

⁹Valdeline Ferreira, Lavras - valdeline.ferreira1@estudante.ufla.br

¹⁰Minelda Arao Lhone, Mozambique - mineldamimi@gmail.com

¹¹João Domingos Scalon, Lavras - scalon@ufla.br



Zero-Inflated Generalized Poisson GARCH regression models for describing Pseudo-nitzschia in Lisbon Bay

Rafaela Rodrigues;¹ Helena Mourão² Valderio Anselmo Reisen³

Resumo

Phytoplankton play a vital role in aquatic ecosystems, serving as a primary food source for a wide range of marine organisms. However, excessive nutrient input can lead to harmful algal blooms that produce toxic compounds, posing risks to marine birds, mammals, and humans. Among these, diatoms of the genus Pseudo-nitzschia are particularly concerning due to their ability to produce domoic acid—a potent neurotoxin responsible for outbreaks of Amnesic Shellfish Poisoning worldwide. Understanding the seasonal and spatial dynamics of Pseudo-nitzschia blooms is therefore essential for evaluating their ecological and public health impacts. This study analyzes weekly water samples collected in Cascais (Lisbon Bay) from June 2001 to May 2005 to determine Pseudo-nitzschia spp. concentrations. While previous research employed a Zero-Inflated Generalized Poisson Regression Model to characterize Pseudo-nitzschia variability and assess the influence of environmental factors such as sea surface temperature and upwelling indices, the present study extends the analysis by incorporating Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models to capture time-series volatility.

Palavras-chave: Pseudo-nitzschia; Time series of counts; Overdispersion; Zero-Inflated Generalized Poisson distribution; GARCH models.

¹Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal - fc62715@alunos.ciencias.ulisboa.pt

²CEAUL, Faculdade de Ciências, Universidade de Lisboa, Lisboa - mhnunes@fc.ul.pt

³Universidade Federal do Espírito Santo, Espírito Santo - valderioanselmoreisen@gmail.com



A Parsimonious GAS Model for NDVI Curve Estimation in Soybean Crop Monitoring

Renata Rojas Guerra;¹ María Agustina Gonzalez ²

Abstract

This work proposes a new generalized autoregressive score (GAS) model based on the Unit-Lindley distribution for modeling the Normalized Difference Vegetation Index (NDVI) curve throughout the soybean growth cycle. The NDVI is a widely used remote sensing indicator of vegetation health, with values naturally bounded between -1 and 1 , making double-bounded continuous distributions particularly suitable for modeling its dynamics. The Unit-Lindley distribution is a one-parameter model for double-bounded random variables. One of its main advantages is the possibility of reparameterization in terms of the mean, enabling a flexible yet interpretable way to model the conditional expectation of NDVI over time. Another advantage of considering this distribution under the GAS framework is its parsimony when compared with the beta distribution, which is the most commonly used double-bounded GAS model in the literature. Parameters are estimated via conditional maximum likelihood, and a Monte Carlo simulation study is conducted to evaluate estimator performance in finite samples. The model's predictive ability is assessed using MODIS satellite-derived real NDVI data from soybean fields in southern Córdoba, Argentina. This approach offers a novel contribution to the analysis of double-bounded time series, with practical applications in environmental monitoring and precision agriculture.

Key-words: double-bounded random variables; Generalized Autoregressive Score models; remote sensing; time series analysis; unit distributions; unit-Lindley distribution.

¹Departamento de Estatística, Universidade Federal de Santa Maria (UFSM), Santa Maria, Rio Grande do Sul, Brasil - renata.r.guerra@ufts.br

²Instituto Gulich, Universidad Nacional de Córdoba, Córdoba, Argentina, - agustina.gonzalez@ig.edu.ar



Análise de séries temporais para dados de desembarques do *Katsuwonus pelamis* via SARIMA e modelos híbridos

Nicoly Rodrigues da Costa ¹

Paulo Emilio Costa Santos ²

Resumo

Este estudo visa modelar e prever os desembarques mensais do *Katsuwonus pelamis* (bonito-listrado) no Oceano Atlântico, utilizando modelos de séries temporais do tipo SARIMA e abordagens híbridas com redes neurais artificiais (RNA). Foram utilizados os dados da Tarefa II da ICCAT (*International Commission for the Conservation of Atlantic Tunas*), que organiza informações pesqueiras por espécie, mês, país, tipo de isca e local de pesca, segmentadas em quatro quadrantes espaciais (Q1 a Q4). Inicialmente foi ajustados modelos SARIMA para captura de componentes lineares da série temporal. Após avaliar os resíduos, e com o objetivo de captar padrões não lineares residuais, ajustou-se modelos híbridos SARIMA + RNA. A previsão final do modelo híbrido resulta da soma das previsões de ambas as etapas. Foram ajustados modelos individualmente para cada um dos quadrantes do Atlântico. A seguir, os desempenhos dos modelos foram avaliados por meio das métricas MAPE, MSE e MAE: Q1 e Q4 tiveram melhor desempenho com o SARIMA, Q2 e Q3 tiveram melhor desempenho com o modelo híbrido. Com base nos melhores modelos de cada quadrante, foi realizada a previsão dos desembarques para os 12 meses subsequentes. As previsões demonstraram boa aderência à tendência recente dos dados históricos, com variações notáveis entre quadrantes Q2 e Q4, que apresentaram maior volume e variabilidade. A abordagem híbrida mostrou-se promissora, especialmente quando havia sinais de não linearidade residual, o que é um diferencial em relação ao SARIMA. Essa flexibilidade evidencia a importância de avaliar múltiplas abordagens em problemas com diferentes padrões temporais e estruturais.

Palavras-chave: Séries temporais; atuns; SARIMA; modelos híbridos; seleção de modelo.

¹ Laboratório Computacional de Estatística, UFPA – Nicolyrodrigues1111@gmail.com

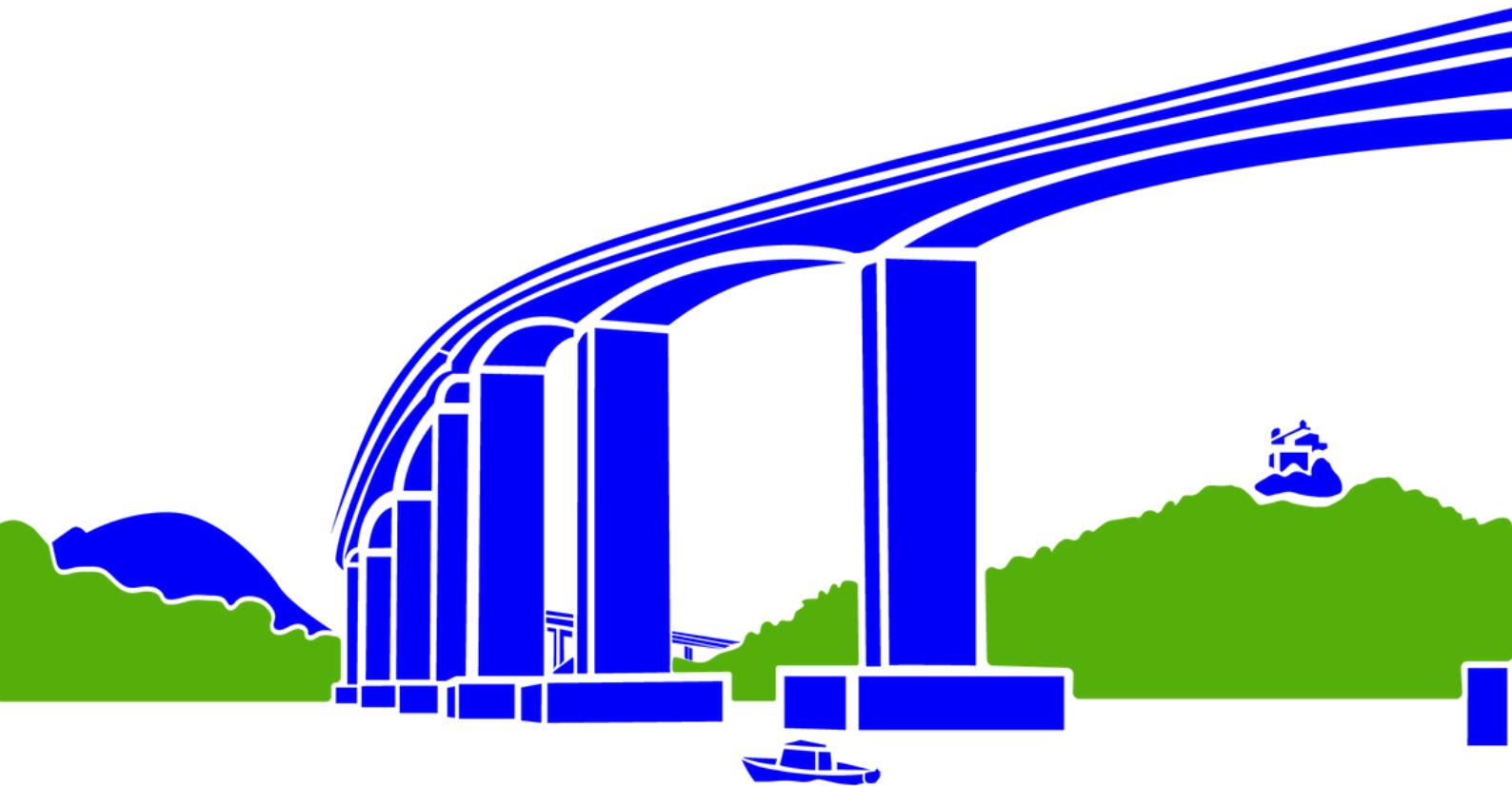
² Núcleo de Ecologia Aquática e Pesca da Amazônia, UFPA – paulo.e.c.s.mat@gmail.com

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**

2025

Concursos

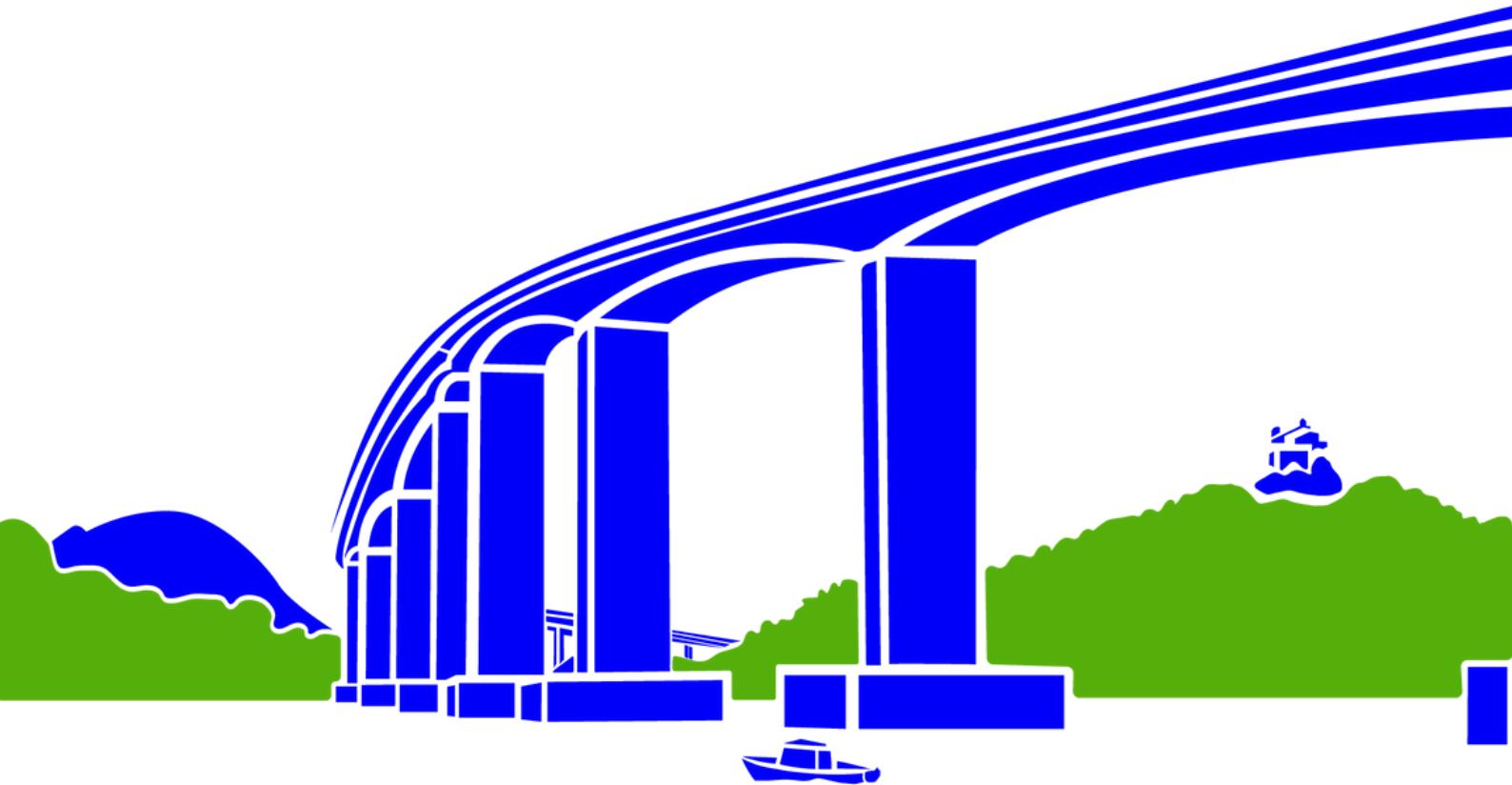


VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Concursos

Melhor IC





Estimação por máxima verossimilhança em modelos espaciais lineares mistos generalizados baseada na aproximação de Laplace

Caio Gomes Alves ¹

Resumo

O trabalho apresenta uma abordagem computacional para estimação de parâmetros em Modelos Espaciais Lineares Mistos Generalizados (SGLMMs) utilizando a aproximação de Laplace, como alternativa a métodos tradicionais baseados em MCMC, que são computacionalmente intensivos e possuem desafios de convergência. O estudo foi desenvolvido em R, com funções adaptadas para simplificar a sintaxe e otimizar o ajuste de modelos para dados não-gaussianos, como contagens e proporções georreferenciadas.

Foram realizados estudos de simulação para avaliar propriedades dos estimadores, como não-viés e consistência, sob diferentes condições (tamanho amostral, funções de correlação espacial e regiões amostrais). Os resultados mostraram que a aproximação de Laplace é computacionalmente mais eficiente que MCMC (sob condições de regularidade), com estimativas comparáveis e menor tempo de execução.

A aplicação a dados reais (contagem de ervas daninhas e sondagens geotécnicas) demonstrou a eficácia do método, permitindo predições espaciais precisas e seleção de modelos via verossimilhança.

Palavras-chave: Geoestatística; Modelos Mistos; Aproximação de Laplace; Verossimilhança; Estatística Espacial.

¹Instituto de Matemática, Estatística e Computação Científica, UNICAMP - c267583@dac.unicamp.br



D-Valor: comparação e aplicação para dados massivos

Jefferson Caponero;¹

Resumo

Este trabalho traz uma discussão sobre o uso do p-valor nos testes de significância da hipótese nula e as questões relacionadas à sua interpretação. Foram avaliadas críticas à abordagem dos p-valores e os desafios específicos ao usar o p-valor para dados massivos. Com alternativa, se avaliou aplicação do d-valor em testes de hipóteses. O d-valor é uma métrica estatística proposta Demidenko focando na probabilidade de que uma observação de um grupo seja maior que uma de outro grupo (controle). Essa métrica é especialmente útil em áreas onde a interpretação prática da magnitude do efeito é mais relevante do que a significância estatística. Simulações realizadas indicaram que, em populações com distribuição normal e gama, o tamanho da amostra tem um impacto pouco significante no d-valor. Foram ainda abordados os desafios associados ao uso em análises estatísticas de dados massivos reais. Assim, aplicou-se o d-valor a amostras aleatórias do conjunto de dados formado pelo Censo da Educação Superior Brasileira do ano de 2019, avaliando categorias como: gênero, turno, auxílio financeiro, modalidade de ensino, grau acadêmico e desempenho territorial, destacando diferenças estatisticamente significativas entre essas categorias. O d-valor demonstrou estabilidade nos resultados, com uma distribuição normal e pouco sensível ao tamanho da amostra. Esse estudo revelou informações significativas sobre o ensino superior nacional. Os dados demonstram que em todas as categorias estudadas a diferença entre o grupo de referência e demais alunos foi ínfima e que a diferença de desempenho entre os estados brasileiros é no máximo pequena.

Palavras-chave: D-Valor; P-Valor; Dados Massivos; Teste de Hipóteses; Censo da Educação Superior Brasileira.

¹Universidade Federal da Bahia, Salvador/BA - caponero@gmail.com



Avaliação aprimorada da capacidade do processo por meio de modelagem semiparamétrica por partes

Vinícius da Costa Soares ¹

Paulo Henrique Ferreira da Silva ²

Resumo

Os modelos por partes (piecewise models) têm ganhado popularidade como uma ferramenta útil em confiabilidade e monitoramento/controle da qualidade, especialmente quando os dados do processo não seguem uma distribuição normal. Neste estudo são desenvolvidos estimadores de máxima verossimilhança (EMVs) para os índices de capacidade do processo Cpk, Cpm, C* pm e Cpmk, usando um modelo semiparamétrico. Para remover o viés nos EMVs com tamanhos de amostra pequenos, é proposta uma abordagem de correção de viés (bias-correction approach). Além disso, o método proposto é estendido para situações em que os pontos de quebra (changepoints) na função de densidade são desconhecidos. Para estimar os parâmetros do modelo de forma eficiente, é utilizado o método de máxima verossimilhança perfilada. O estudo de simulação revelou que o método sugerido produz estimativas com baixo viés e erro quadrático médio. Por fim, são fornecidas três aplicações a dados reais, a fim de demonstrar a superioridade do procedimento proposto em relação aos existentes.

Palavras-chave: Índice de capacidade do processo; método de correção de viés; modelo de ponto de quebra; modelo exponencial por partes.

¹ Departamento de Estatística, Instituto de Matemática e Estatística, Universidade Federal da Bahia, Salvador-BA, Brasil – vinicius4burame@gmail.com

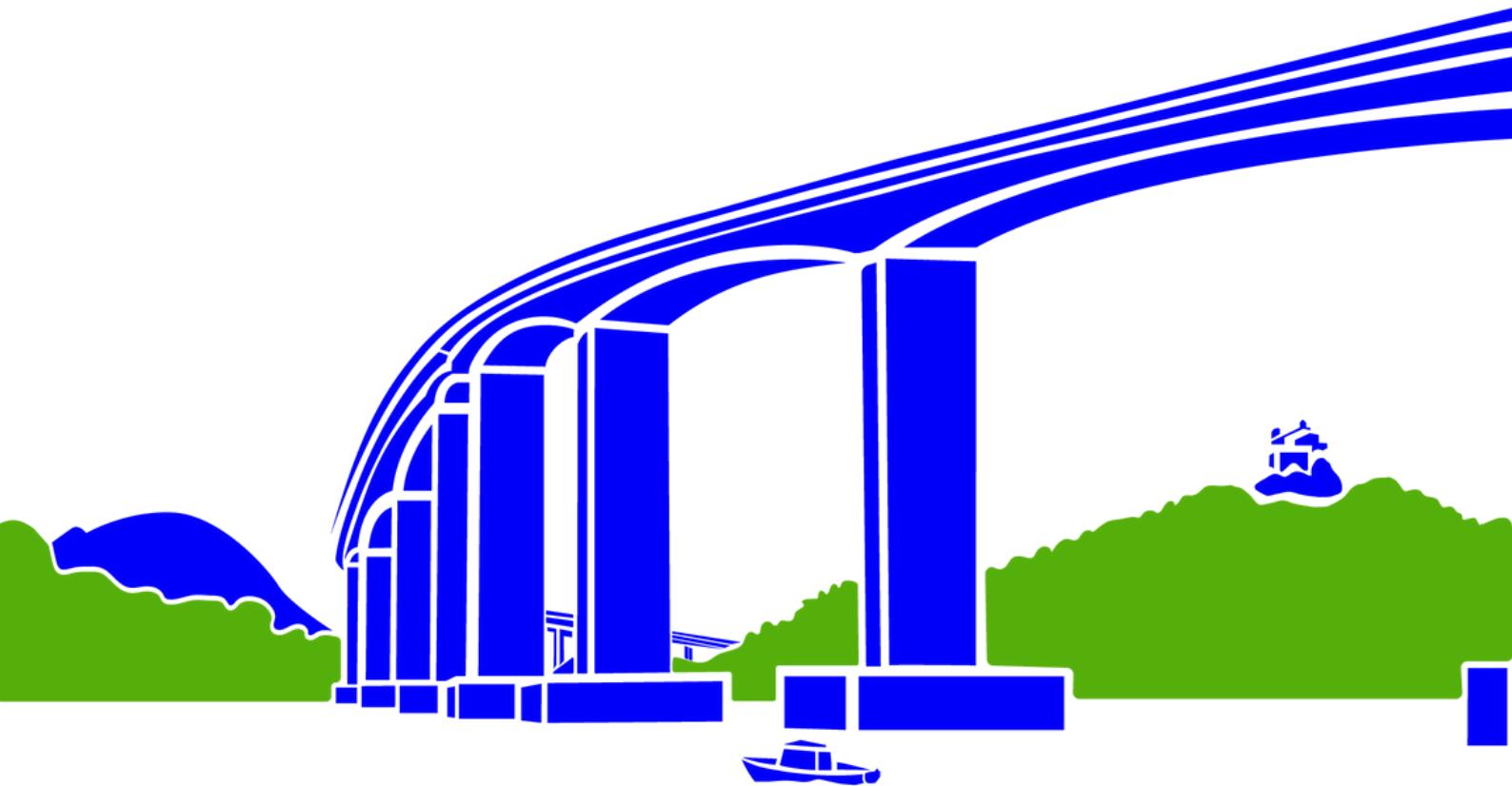
² Departamento de Estatística, Instituto de Matemática e Estatística, Universidade Federal da Bahia, Salvador-BA, Brasil – paulohenri@ufba.br

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Concursos

**Melhor Dissertação de
Mestrado**





New Generalized Autoregressive Score and Regression Models

Fernando José Monteiro de Araújo;¹ Renata Rojas Guerra;² Fernando Arturo Peña-Ramírez³

Resumo

This work proposes new dynamic models that combine the unit gamma (UG) and Burr XII (BXII) conditional distributions with the generalized autoregressive score (GAS) framework, referred to as UG-GAS and BXII-GAS. These proposals aim to enhance the understanding of stochastic behavior and improve the forecasting of hydro-environmental indicators. To model time-dependent effects, dynamic structures such as GAS models provide a useful framework, allowing for the update of time-varying parameters based on past observations and the weighted score function. The estimation of GAS models is performed via conditional maximum likelihood, and a Monte Carlo simulation study is conducted to assess the estimators' performance. The proposed UG-GAS and BXII-GAS models are applied to real data on capacity factors in wind power plants and river flow rates in hydroelectric plants, respectively. In addition, we also propose a new quantile regression model based on a reparameterization of the unit ratio-Weibull (URW) distribution. We define a systematic structure for two parameters of the distribution: one representing the quantiles of the URW and the other corresponding to the shape parameter. The estimation of the regression model parameters is carried out using the maximum likelihood method, and its performance is evaluated through Monte Carlo simulations and an application to mortality rates. The results illustrate the model's usefulness in understanding and quantifying the effects of economic, social, demographic, and public health variables on the quantiles of the COVID-19 mortality rate in Latin American countries.

Key words: Generalized autoregressive score models; Hydro-environmental indicators; Monte Carlo simulation; Mortality rates; Quantile regression.

¹Graduate Program of Statistics, Universidade Federal do Rio Grande do Sul, Avenida Bento Gonçalves, Porto Alegre, 91509-900 Rio Grande do Sul, Brazil. Email: fernando.monteiro@ufrgs.br.

²Department of Statistics, Universidade Federal de Santa Maria, Avenida Roraima, Santa Maria, 97105-900, Rio Grande do Sul, Brazil. Email: renata.r.guerra@ufrsm.br.

³Department of Statistics, Universidade Federal de Santa Maria, Avenida Roraima, Santa Maria, 97105-900, Rio Grande do Sul, Brazil. Email: fernando.p.ramirez@ufrsm.br.



IDENTIFICAÇÃO DO EFEITO CAUSAL NO MODELO DE MEDIAÇÃO COM VARIÁVEIS LATENTES

Michelle P.V. Passos*¹

Marcelo M. Taddeo¹

Leila D.A.F. Amorim¹

Resumo

A análise de mediação causal baseada em respostas potenciais (contrafactual) tem sido amplamente utilizada para decompor o efeito causal de uma intervenção sobre desfechos em diversas áreas, especialmente na epidemiologia e nas ciências sociais. Os métodos mais conhecidos se baseiam em variáveis contínuas, geralmente por meio de modelos lineares e mensuração sem erro. No entanto, em certos contextos, o mediador e/ou o desfecho podem ser variáveis latentes, caracterizadas via modelos de classes latentes. Assim, com a crescente disseminação de modelos de mediação com variáveis latentes, torna-se necessária a formalização das condições de identificação causal dos efeitos naturais direto e indireto, para garantir interpretação e estimação sem viés. Neste contexto, esta dissertação avalia o comportamento dos estimadores dos efeitos direto e indireto em modelos com variáveis latentes categóricas, via análise de classes latentes (LCA), em cenários com mediadores e/ou desfechos latentes. As metodologias para estimação dos efeitos natural indireto (NIE) e direto (NDE) são estendidas para variáveis latentes com mais de duas classes. Propõe-se também a inclusão de escores de propensão em modelos marginais estruturais com variáveis latentes. Estudos de simulação de Monte Carlo avaliaram as propriedades dos métodos propostos sob diferentes cenários de violação das suposições. As metodologias são ilustradas com dados reais de duas aplicações na área da saúde. Os resultados destacam a importância dos critérios de identificação causal e oferecem *insights* relevantes para avanços metodológicos e futuros estudos.

Palavras-chave: Inferência Causal, Análise de Classes Latentes, Efeito Natural Indireto, Mediação Causal, Modelos Marginais Estruturais.

¹Instituto de Matemática e Estatística, Universidade Federal da Bahia, Salvador, Bahia, Brasil.



Delineamentos ótimos para modelos polinomiais fracionários

Samantha Navarro Janine ¹; Marcelo Andrade da Silva ²

Resumo

A vasta aplicabilidade de modelos não lineares em diferentes contextos tem motivado o desenvolvimento de estudos para a definição de algoritmos envolvendo estas funções. Contudo, ainda existem poucos trabalhos realizados em cenários de delineamentos ótimos para modelos não lineares. A teoria de delineamentos de experimentos consiste em planejar o experimento e determinar a estimativa dos parâmetros de acordo com a função dada entre as variáveis de interesse. Para obter a máxima informação possível sobre os parâmetros, utilizam-se os delineamentos ótimos, com base em um determinado critério. Em funções com covariáveis contínuas, cujos comportamentos não são capturados por modelos polinomiais, o uso de polinômios fracionários, dado por um conjunto de potências usualmente utilizado, torna o ajuste mais flexível e de simples interpretação. Logo, o objetivo deste trabalho é estudar a teoria de delineamentos ótimos em modelos polinomiais fracionários a partir da construção de um algoritmo de troca das coordenadas da matriz de planejamento. Através do estudo das abordagens de delineamentos localmente ótimos e pseudo-Bayesianos considerando o critério D, o interesse é avaliar como a atribuição dos valores iniciais dos parâmetros e suas respectivas prioris podem influenciar no resultado do delineamento ótimo obtido a fim de garantir delineamentos mais eficientes. Além disso, deseja-se motivar a programação de algoritmos flexíveis a diferentes cenários que não estão implementados em pacotes usuais disponíveis em softwares para análises estatísticas.

Palavras-chave: Delineamentos ótimos; Polinômios fracionários; Algoritmo de troca por coordenada; Planejamento de experimentos.

¹Departamento de Ciências Exatas - ESALQ/USP, Piracicaba-SP - samantha.nj@usp.br

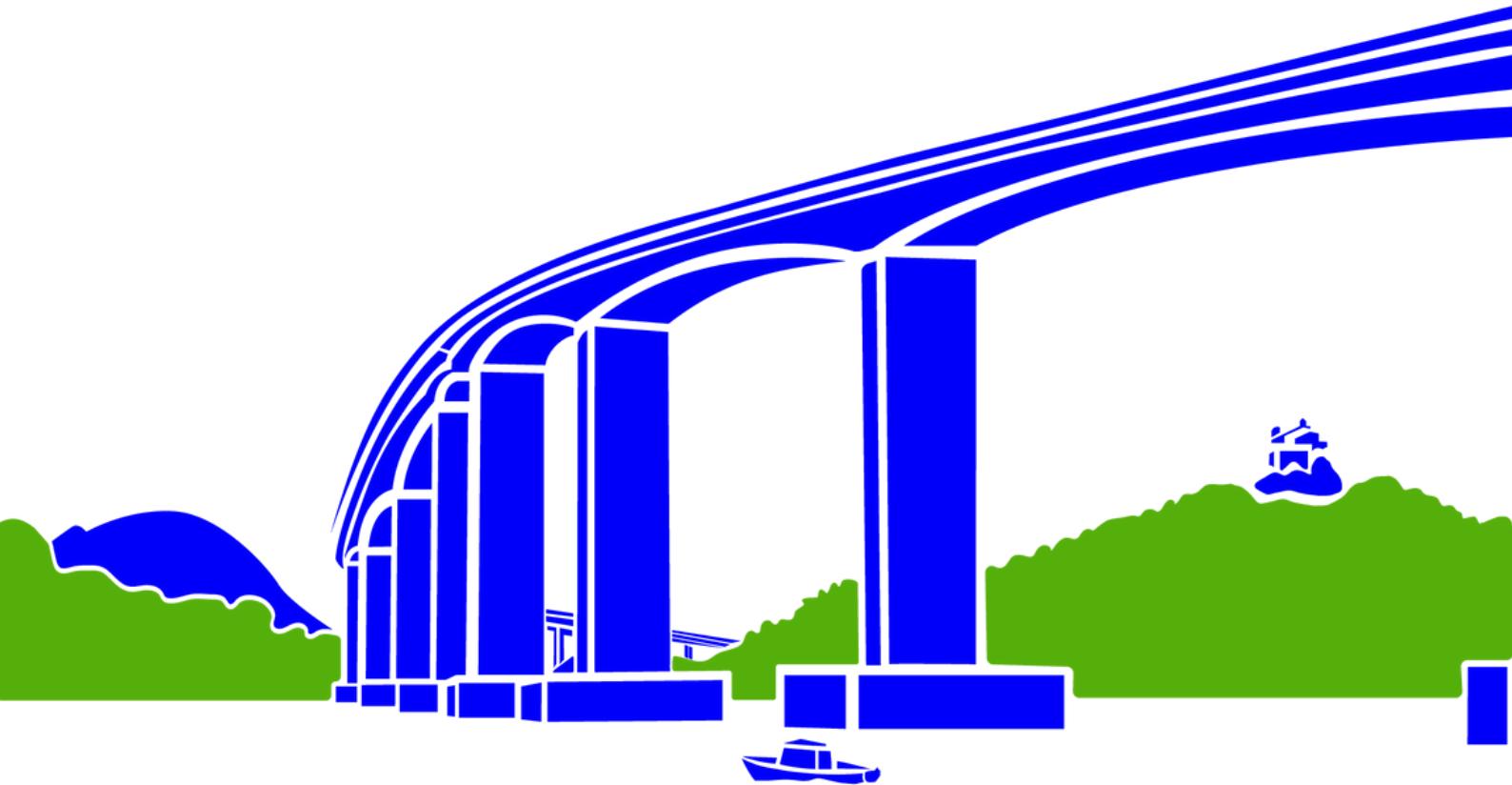
²Departamento de Ciências Exatas - ESALQ/USP, Piracicaba-SP - silva.marcelo@usp.br

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Concursos

**Melhor Tese de
Doutorado**





Functional Data Analysis: Spatial Association of Curves and Irregular Spacing

Alvaro Alexander Burbano Moreno ¹ Vinícius Diniz Mayrink ²

Abstract

Spatial Functional Data (SFD) analysis is an emerging statistical framework that combines Functional Data Analysis (FDA) and spatial dependency modeling. Unlike traditional statistical methods, which treat data as scalar values or vectors, SFD considers data as continuous functions, allowing for a more comprehensive understanding of their behavior and variability. This approach is well-suited for analyzing data collected over time, space, or any other continuous domain. SFD has found applications in various fields, including economics, finance, medicine, environmental science, and engineering. This thesis proposes new functional Gaussian models incorporating spatial dependence structures, focusing on irregularly spaced data and reflecting spatially correlated curves. The models are based on B-spline basis expansions and Bernstein Polynomials (BP) and utilize a Bayesian approach for estimating unknown quantities and parameters. The thesis explores the advantages and limitations of B-spline-based and BP-based models in capturing complex shapes and patterns while ensuring numerical stability. The main contributions of this work include the development of an innovative model designed for SFD using B-spline or BP structures, including a random effect to address associations between irregularly spaced observations, and a comprehensive simulation study to evaluate models' performance under various scenarios. The thesis also presents two real applications related to levels of PM10 and Temperature in Mexico City, showcasing practical illustrations of the proposed models.

Palavras-chave: Functional Data Analysis; Spatial Dependency; Bayesian Statistics; Bernstein Polynomial; B-spline.

¹Departamento de Estatística, IMECC, Universidade Estadual de Campinas - aamoreno@unicamp.br

²Departamento de Estatística -ICEX, Universidade Federal de Minas Gerais - vdm@est.ufmg.br



Multivariate modeling and linear and partially linear mixed models using scale mixtures of centered skew-normal distributions

Candidato: João Victor Bastos de Freitas ¹

Orientador: Prof. Dr. Caio Lucidius Naberezny Azevedo ²

Co-orientador: Prof. Dr. Valdério Anselmo Reisen ³

Resumo

Em muitos casos, em Estatística, as distribuições associadas às variáveis de interesse podem apresentar assimetria e/ou caudas mais pesadas do que as da normal, fazendo com que esta distribuição seja inapropriada. Assim, introduzimos e exploramos as distribuições de mistura de escala da skew-normal centrada (MESNC) multivariada para contornar alguns problemas inferenciais e de interpretação apresentados por sua versão sob a parametrização usual. Além disso, motivados pelo Millennium Cohort Study, propomos um modelo linear misto com efeitos aleatórios seguindo uma MESNC multivariada e erros seguindo uma mistura de escala da normal multivariada. Por fim, motivados pelo estudo da qualidade do ar na cidade de Vitória, Espírito Santo, Brasil, propomos um modelo misto aditivo parcialmente linear sob independência condicional com erros seguindo uma MESNC univariada, considerando também os parâmetros de escala e assimetria variando no tempo. Para todas as metodologias propostas foram feitos estudos de simulação para avaliar características de interesse. Quando pertinente, análise residual, critérios de informação e análise de influência foram propostas. As análises de conjuntos de dados reais constataram que os novos modelos (distribuições) apresentaram melhores resultados, em termos de qualidade de ajuste do que abordagens usuais da literatura.

Palavras-chave: Modelos mistos; Modelagem multivariada; Misturas de escala; Skew-normal centrada; Modelos aditivos parcialmente lineares.

¹ Departamento de Estatística, Instituto de Matemática, Estatística e Computação Científica, Universidade Estadual de Campinas, Campinas, São Paulo, Brasil - joao.victor.ufc13@gmail.com

² Departamento de Estatística, Instituto de Matemática, Estatística e Computação Científica, Universidade Estadual de Campinas, Campinas, São Paulo, Brasil - cnaber@unicamp.br

³ Programa de Pós-graduação em Engenharia Ambiental (PPGEA), Universidade Federal do Espírito Santo, Vitória, Espírito Santo, Brasil - valderioanselmoreisen@gmail.com



Essays on Cure Rate Models

Marcia Brandao de Oliveira Martins;¹ Jeremias da Silva Leão;² Marcelo Bourguignon Pereira;³ Diego Ignacio Gallardo Mateluna;⁴

Resumo

Este trabalho foca em modelos análise de sobrevivência no contexto de fração de cura, fundamentais em aplicações onde uma parcela significativa dos indivíduos nunca experimenta o evento de interesse, mesmo quando observados por longos períodos. Esses modelos assumem uma população homogênea e incorporam uma variável aleatória não observada para representar informações ausentes. O estudo é dividido em três capítulos. O primeiro capítulo apresenta uma introdução aos modelos de fração de cura. O segundo capítulo estende o modelo de fração de cura ao considerar causas competitivas dentro da classe de Séries de Potências, assumindo uma mistura de duas causas concorrentes. Essa generalização inclui vários modelos conhecidos como casos especiais. A estimação dos parâmetros é realizada pelo método da máxima verossimilhança, e um algoritmo do tipo EM (Expectation-Maximization) é proposto. Simulações de Monte Carlo são conduzidas para avaliar propriedades assintóticas. A metodologia é ilustrada com dados reais de um estudo populacional de casos de melanoma cutâneo diagnosticados em São Paulo, Brasil. O terceiro capítulo introduz uma nova abordagem de modelagem onde o número de causas competitivas segue uma mistura das distribuições Poisson e Birnbaum-Saunders. Essa abordagem generaliza modelos existentes, sendo o modelo do tempo de promoção um caso limite. Propriedades estatísticas são analisadas, e a estimação dos parâmetros é novamente realizada pelo método da máxima verossimilhança com um algoritmo EM. Experimentos de Monte Carlo avaliam propriedades assintóticas e o poder do teste da razão de verossimilhança. Por fim, a metodologia é aplicada a dados reais de um estudo populacional de casos de câncer de mama em São Paulo, Brasil.

Palavras-chave: Misturas; Distribuição em série de potências, Poisson, Birnbaum-Saunders, causas concorrentes; Algoritmo EM; Melanoma; Câncer de Mama.

¹Universidade Federal do Amazonas - mbrandao07@gmail.com

²Universidade Federal do Amazonas - leaojeremiass@gmail.com

³Universidade Federal do Rio Grande do Norte - m.p.bourguignon@gmail.com

⁴Universidad del Bío-Bío - diego.gallardo.mateluna@gmail.com



Inferência para modelos com respostas censurados usando a distribuição Student-t e skew-t

Katherine Loor Valeriano;¹ Larissa Avila Matos;² Christian Galarza Morales³

Resumo

O estudo de modelos nos quais a variável de resposta está sujeita a limites de detecção, tem sido de interesse em muitas áreas da estatística. Este tipo de dados surgem frequentemente em monitoramento ambiental, medicina, economia, agronomia e biologia. A maioria dos modelos existentes na literatura para lidar com dados censurados assume uma distribuição normal para a variável de resposta, e essa suposição pode ser irrealista na presença de desvios da normalidade ou de outliers. Neste trabalho, propomos uma série de modelos que consideram distribuições assimétricas e de caudas pesadas, como as distribuições Student-t e skew-t, para lidar com observações censuradas e/ou faltantes na variável de resposta.

Os parâmetros dos modelos são estimados utilizando o algoritmo Expectation-Maximization (EM) (Dempster et al., 1977), um método amplamente utilizado para aproximar iterativamente as estimativas de máxima verossimilhança (ML). Este algoritmo exige o cálculo de algumas esperanças condicionais. Em nossos modelos, isso inclui os dois primeiros momentos das distribuições Student-t, skew-t e extended skew-t. Para calcular os momentos da distribuição Student-t, desenvolvemos um método baseado na integração de Monte Carlo, complementado por resultados derivados da esperança condicional (veja, por exemplo, Galarza et al., 2021). Além disso, quando as esperanças condicionais não podem ser derivadas de forma explícita, empregamos uma versão de aproximação estocástica do algoritmo EM, conhecido como algoritmo SAEM (Delyon et al., 1999), para a estimativa dos parâmetros. Para cada modelo, também fornecemos procedimentos para aproximar o erro padrão das estimativas e expressões para prever observações futuras. As propriedades assintóticas e a robustez das estimativas são demonstradas através de estudos de simulação, e aplicações em conjuntos de dados reais são apresentadas para esses modelos.

Palavras-chave: algoritmo EM; família de distribuições elípticas; distribuições de caudas pesadas; distribuições assimétricas; distribuições truncadas; observações censuradas.

Referências

Delyon, B., M. Lavielle, E. Moulines, et al. (1999). Convergence of a stochastic approximation version of the EM algorithm. *The Annals of Statistics* 27(1), 94–128.

¹Escuela Superior Politécnica del Litoral (ESPOL), Facultad de Ciencias Naturales y Matemáticas, Ecuador - kataloor@espol.edu.ec

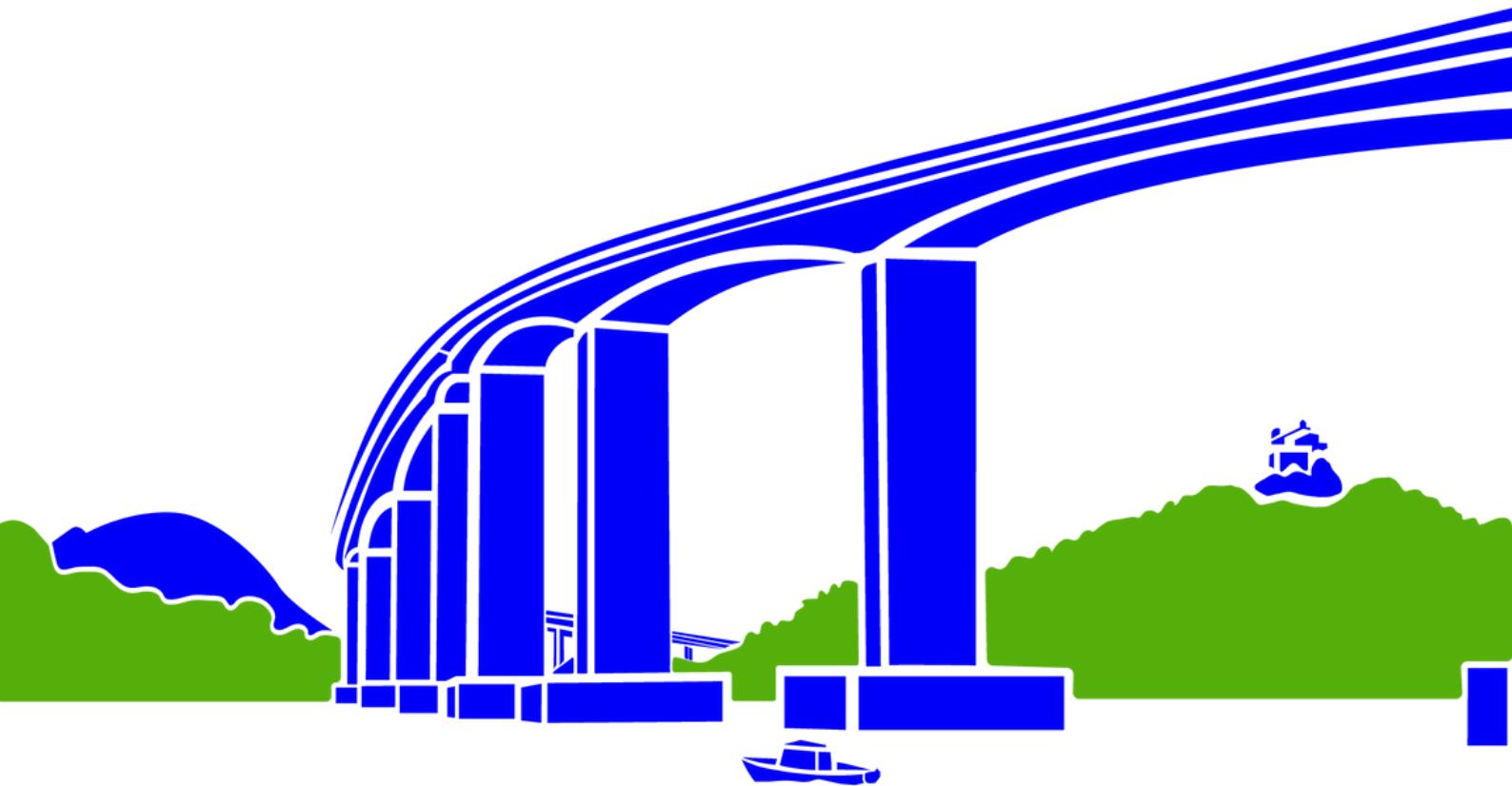
²Universidade Estadual de Campinas (UNICAMP), Instituto de Matemática, Estatística e Computação Científica, Brasil - larissam@unicamp.br

³Escuela Superior Politécnica del Litoral (ESPOL), Facultad de Ciencias Naturales y Matemáticas, Ecuador - chedgala@espol.edu.ec

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Pôsters

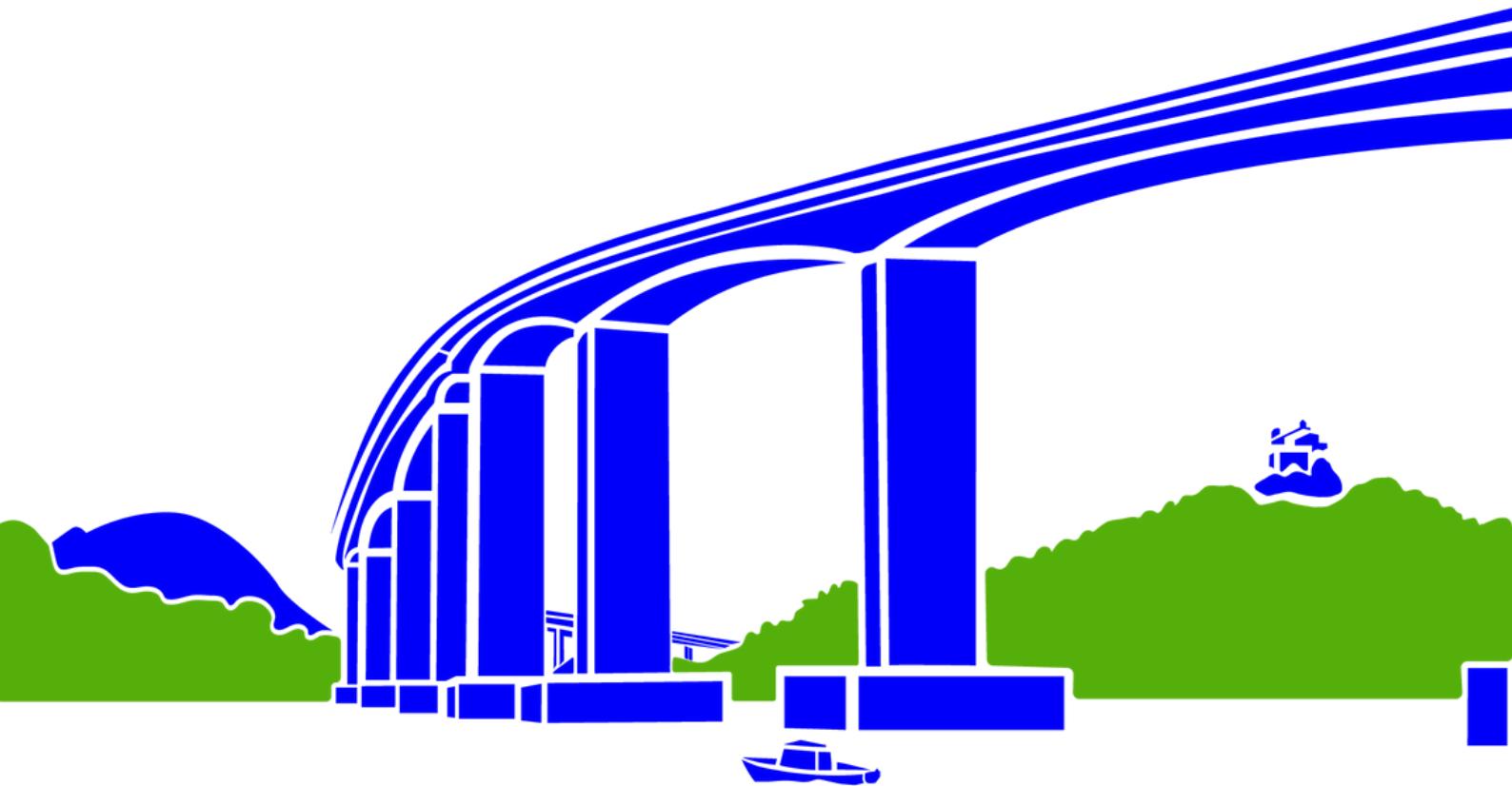


VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Pôsters

Séries Temporais





Análise de precipitação máxima anual na região do Matopiba

Vinicius Silva Begnami¹; Samantha Gouvêa Oliveira²; Paulo César Emiliano³; Luciano Gonçalves Batista⁴; Kamila Andrade de Oliveira⁵

Resumo

O avanço da tecnologia revolucionou a mecanização agrícola e possibilitou a adaptação de cultivares para as mais diversas regiões do Brasil. O Matopiba é uma região compreendida entre os estados do Maranhão, Tocantins, Piauí e Bahia, conhecida como a nova fronteira agrícola. Essa região se destaca devido à sua alta produção de grãos, principalmente a soja. Na safra 2022/2023, foi responsável por 18,5 milhões de toneladas de soja, o que corresponde a 12,3% da produção total do país. Devido à relevância da localidade, pesquisas sobre eventos climáticos vêm se intensificando na região com o intuito de auxiliar no planejamento e manejo das culturas. Diante do exposto, este trabalho tem como objetivo modelar dados de precipitações máximas de quatro cidades do Matopiba: Carolina (MA), Porto Nacional (TO), Cristino Castro (PI) e Barreiras (BA). Os dados, referentes ao período de 1961 a 2024, constituem-se de valores de precipitação máxima anual. Foram obtidos no Banco de Dados Meteorológicos para Ensino e Pesquisa (BDMEP) e modelados por meio da Distribuição Generalizada de Valores Extremos (GEV). Os parâmetros da distribuição foram estimados pelo método da máxima verossimilhança, entretanto, devido à não linearidade das equações, o método de otimização Broyden-Fletcher-Goldfarb-Shanno (BFGS) foi empregado. Os níveis de retorno foram estimados para 10, 20, 30, 50 e 100 anos em cada cidade. Os resultados indicam que é esperado um volume máximo anual de no mínimo 109,31 mm e no máximo 174,31 mm de precipitação entre os anos estipulados. Dessa forma, políticas públicas e planejamento serão cruciais para manter a produção agrícola na região.

Palavras-chave: Tendências climáticas; Modelagem; Chuva extrema; Valores extremos.

¹Departamento de estatística, Viçosa-MG - vinicius.begnami@ufv.br

²Departamento de estatística, Viçosa-MG - samantha.gouvea@ufv.br

³Departamento de estatística, Viçosa-MG - paulo.emiliano@ufv.br

⁴Departamento de estatística, Viçosa-MG - luciano.batista@ufv.br

⁵Departamento de engenharia agrícola, Chapadinha-MA - kamilla.andrade@ufma.br



Modelo Hierárquico de Séries Temporais da Inflação nos Preços de Alimentos no Brasil

Gustavo Jun Yakushiji ¹ Fábio Prataviera ²

Resumo

A inflação é caracterizada pelo aumento persistente dos preços de produtos e serviços em uma economia. No Brasil, o principal indicador utilizado para medir a inflação ao consumidor é o Índice Nacional de Preços ao Consumidor Amplo (IPCA), estruturado em quatro níveis hierárquicos: grupos, subgrupos, itens e subitens. A partir de 2007, o grupo Alimentação e Bebidas — um dos nove que compõem o IPCA — apresentou crescimento acumulado superior ao índice geral, indicando uma inflação mais elevada nos alimentos. Esse fenômeno afeta, de forma mais acentuada, a população de menor renda, cujo dispêndio orçamentário com alimentação é proporcionalmente maior. Neste contexto, o presente estudo tem como objetivo aplicar modelos hierárquicos de séries temporais para a inflação de alimentos ao consumidor no Brasil, com base em três abordagens metodológicas: *bottom-up*, *top-down* e reconciliação ótima. Foram utilizadas séries mensais do grupo Alimentação e Bebidas e de seus subgrupos e itens, no período de janeiro de 2000 a dezembro de 2024. Utilizou-se o teste de raiz unitária Dickey-Fuller Aumentado para verificar a estacionariedade das séries, e as funções de autocorrelação e autocorrelação parcial para identificar as possíveis ordens dos modelos ARIMA e ETS. A seleção dos modelos foi realizada com base na raiz do erro quadrático médio, e a adequação dos resíduos, por meio do teste Ljung-Box. As previsões foram realizadas para os anos de 2023 e 2024. Os resultados visam contribuir para a compreensão do comportamento da inflação de alimentos e auxiliar no planejamento de políticas públicas e em futuras investigações na área.

Palavras-chave: IPCA; Reconciliação ótima; *Bottom-up*; *Top-down*.

¹ Departamento de Ciências Exatas - LCE, ESALQ/USP – gustavojun@usp.br

² Departamento de Ciências Exatas - LCE, ESALQ/USP – fabio_prataviera@usp.br



Series Predictor: Aplicativo para análise e previsão de séries temporais em shiny

Gabriel Cardoso ¹ Anderson Castro Soares de Oliveira ² Elianara Martins de Almeida ³
Juliano Bortolini ⁴

Resumo

A análise de séries temporais é fundamental para compreender e prever padrões ao longo do tempo em diferentes áreas do conhecimento. O software R destaca-se como uma plataforma robusta para esse tipo de análise, oferecendo amplos recursos estatísticos e de modelagem. No entanto, sua natureza programática pode dificultar o acesso para usuários com pouca experiência em programação. Para superar essa barreira, este trabalho apresenta o desenvolvimento de um aplicativo interativo baseado no pacote Shiny, com o objetivo de tornar a análise de séries temporais mais acessível e intuitiva. A ferramenta permite ao usuário importar conjuntos de dados, convertê-los em séries temporais, realizar testes de tendência e sazonalidade, ajustar modelos ARIMA/SARIMA e gerar previsões futuras, tudo por meio de uma interface gráfica amigável. Essa abordagem visa democratizar o uso de técnicas estatísticas avançadas, promovendo uma maior compreensão dos dados temporais e apoiando a tomada de decisões em áreas como economia, saúde, meio ambiente e gestão pública. Os resultados obtidos demonstram que o aplicativo é eficaz tanto na análise quanto na visualização dos dados, oferecendo insights relevantes de forma prática. Assim, o estudo contribui para a ampliação do uso de análises temporais em contextos aplicados, consolidando o Shiny como uma solução viável para integrar análise estatística e acessibilidade.

Palavras-chave: Análise Preditiva, Modelagem Temporal, Aplicativo Interativo, Modelos ARIMA, Modelos SARIMA

¹Departamento de Estatística/ICET/UFMT, Cuiabá - gabriel.cardoso@sou.ufmt.br

²Departamento de Estatística/ICET/UFMT, Cuiabá - anderson.oliveira@ufmt.br

³Departamento de Estatística/ICET/UFMT, Cuiabá - elianara.almeida@ufmt.br

⁴Departamento de Estatística/ICET/UFMT, Cuiabá - juliano.bortolini@ufmt.br



Mortalidade por acidentes de transporte em Minas Gerais: um estudo de séries temporais

Lucas Pereira Barbosa ¹; Luiz Otávio Pala ²; Thelma Sáfadi ³

Resumo

Os acidentes de trânsito representam uma importante causa de morbimortalidade, impactando a sociedade em diferentes aspectos, como nos sistemas de saúde. Este estudo tem como objetivo analisar a taxa de mortalidade por acidentes de transporte (Grupo CID V01-V99) em Minas Gerais, com foco na identificação de medidas e políticas públicas que possam contribuir para a desaceleração do número de mortes. A justificativa deste estudo baseia-se no terceiro Objetivo de Desenvolvimento Sustentável (ODS), que visa assegurar uma vida saudável e promover o bem-estar para todas as pessoas, em todas as idades, incluindo a redução das mortes e dos ferimentos causados por acidentes de trânsito. Foram utilizadas as séries temporais das taxas de mortalidade de homens e mulheres, ajustadas a cada cem mil habitantes, em decorrência de acidentes de transporte no período de janeiro de 2000 a dezembro de 2023, cujos dados foram extraídos do Departamento de Informática do Sistema Único de Saúde. A análise inicial indicou uma tendência cúbica em ambas as séries, caracterizada por duas inflexões: a primeira após 2011, marcando uma redução no número de mortes, e a segunda em 2020, com o menor número de óbitos, seguida por um aumento até 2023. Após subtraídas as tendências das séries masculinas e femininas, a correlação significativa nas séries livres da tendência evidenciou a presença de sazonalidade, resultado que foi corroborado pelo teste de Friedman. Para a taxa de mortalidade masculina, identificou-se uma dependência sazonal de 1, 2 e 3 meses, além de 1 e 2 anos. Já para a taxa feminina, a dependência sazonal encontrada foi de 4 meses e 1 ano. Para o ajuste das séries temporais, foram utilizados modelos da classe SARIMA $(p, d, q)(P, D, Q)_{12}$, seguidos de previsões para os meses de 2024. Os modelos ajustados para cada série foram, respectivamente, SARIMA $(3, 0, 0)(2, 1, 0)_{12}$ e SARIMA $(4, 0, 0)(1, 0, 0)_{12}$. As previsões para os 12 meses de 2024 indicam um movimento decrescente para a série masculina e estabilidade com pequenas variações para a série feminina. Esses resultados oferecem informações para a formulação de políticas públicas mais eficazes, que possam estar alinhadas e cumprir os objetivos estabelecidos nos ODS.

Palavras-chave: Acidentes de trânsito; previsão de óbitos; modelos SARIMA.

¹Departamento de Estatística e Experimentação Agropecuária, Universidade Federal de Lavras - lucas.barbosa6@estudante.ufla.br

²Departamento de Estatística, Universidade Federal de Lavras - luizpala@ufla.br

³Departamento de Estatística, Universidade Federal de Lavras - safadi@ufla.br



Estudo comparativo da qualidade das previsões em séries temporais utilizando as metodologias SSA e GRNN

Vitor Yuji Kiemo;¹ Fernando Lucambio Perez ²

Resumo

A transformação log-retorno é amplamente utilizada no mercado financeiro para modelagem e análise de séries temporais, especialmente em índices de ativos voláteis, como criptomoedas. A compreensão dos padrões dessas séries é interessante para a identificação de tendências e sazonalidades, e também na estruturação de estratégias de investimento. Dentro deste contexto, neste estudo comparamos a performance preditiva de dois métodos: Singular Spectrum Analysis (SSA) e Generalized Regression Neural Networks (GRNN), aplicados a múltiplos índices de criptomoedas. Nossa objetivo é avaliar qual abordagem captura melhor as dinâmicas do mercado, considerando a não estacionariedade das séries e possíveis padrões sazonais, contribuindo assim para a escolha de estratégias mais eficazes na modelagem de séries temporais econômico-financeiras.

Palavras-chave: SSA; GRNN; séries temporais; mercado financeiro; criptomoeda

¹Discente de Estatística e Ciência de Dados, UFPR - vitor.kiemo@ufpr.br

²Departamento de Estatística, UFPR - lucambio@ufpr.br



Imputação múltipla de dados faltantes em séries temporais de poluição atmosférica

Wendy Nascimento Silva;¹
Ana Júlia Alves Câmara²

Resumo

A poluição atmosférica é um fator crítico à saúde humana, estando associada a doenças cardiovasculares, respiratórias e neurológicas. A análise de séries temporais é uma ferramenta essencial para identificar padrões entre os níveis de poluição e seus impactos na saúde da população. No entanto, a presença de valores ausentes, comum em dados ambientais, pode comprometer a qualidade das análises.

Este trabalho teve como objetivo aplicar o método K-Nearest Neighbors (KNN) para imputação de dados faltantes em séries temporais, avaliando sua eficiência por meio das métricas RMSE e MAE. Foram utilizados dois conjuntos de dados: um real, com 84 observações de concentrações de poluentes atmosféricos e casos de bronquite aguda na Grande Vitória (sendo analisados especificamente o CO, PM10 e NO), e outro simulado, com 200 observações geradas por um processo AR(1) $\phi = 0,1$.

As ausências foram simuladas com o mecanismo MAR, em proporções de 5%, 10% e 20%. A imputação foi realizada em linguagem R, utilizando os pacotes Caret (imputação), miss-Methods (remoção de dados) e Metrics (avaliação de desempenho). Diferentes valores de k (número de vizinhos mais próximos considerados na imputação) foram testados para definição do melhor desempenho. Os resultados demonstraram que o KNN é uma alternativa eficiente para imputação de dados faltantes em séries temporais, com baixos erros médios e boa preservação do padrão da série original.

Palavras-chave: valores faltantes; séries temporais; imputação múltipla; KNN; poluição atmosférica.

¹Departamento de Estatística, Universidade Federal do Espírito Santo, Vitória, ES - wendy.silva@edu.ufes.br

²Departamento de Estatística, Universidade Federal do Espírito Santo, Vitória, ES - ana.j.camara@ufes.br



Análise da comercialização do etanol hidratado no estado de São Paulo por meio do preço de revenda

Mariana Gabriela Pantoja¹
Gilberto Rodrigues Liska²
Marta Cristina Marjotta-Maistro³

Resumo

A importância da cana-de-açúcar no Brasil se encontra desde a época da colonização, a partir do século XVI. O crescimento do setor teve destaque a partir da década de 1990 com o aumento da produção de um de seus subprodutos, o etanol hidratado, que ganhou um cenário cada vez mais enriquecido, principalmente no Estado de São Paulo, o qual foi o maior produtor no ano de 2021, correspondendo a 36,2% da produção nacional. Este estudo teve por objetivo estudar os preços de revenda na cadeia de comercialização do etanol hidratado no Estado de São Paulo entre o período de 2003 a 2022. Para tal, os dados foram obtidos da Agência Nacional do Petróleo (ANP) e a União da Indústria de Cana-de-Açúcar e Bioenergia (UNICA) e estes foram submetidos à metodologia Box & Jenkins de séries temporais. Foi considerado o ano de 2022 para divisão da série original em treinamento e teste e foram considerados o critério de informação de Akaike, erro quadrático médio de previsão e erro percentual absoluto médio para decidir qual modelo estatístico é mais apropriado. Todas as análises foram realizadas no R. Os resultados mostram evidência de presença de tendência e sazonalidade na série de preços de revenda de etanol hidratado e dentre os 15 modelos considerados, o modelo SARIMA(3,1,2)(1,0,1)₁₂ foi o que apresentou os melhores indicadores de qualidade de ajuste e pode ser recomendado para efetuar previsões. Em continuação desse estudo, pretende-se estudar o comportamento dos preços de etanol hidratado ao produtor e distribuidor.

Palavras-chave: Cadeia de produção; preços; mercado interno; séries temporais; Programa R

¹ Discente da Engenharia Agronômica, UFSCar, CCA - mariana.pantoja@estudante.ufscar.br

² Docente do Departamento de Tecnologia Agroindustrial e Socioeconomia Rural do Centro de Ciências Agrárias, Universidade Federal de São Carlos, Araras - SP, Brasil - gilbertoliska@ufscar.br

³ Docente do Departamento de Tecnologia Agroindustrial e Socioeconomia Rural do Centro de Ciências Agrárias, Universidade Federal de São Carlos, Araras - SP, Brasil - marjotta@ufscar.br



Educação de Qualidade: O Papel das Boas Práticas no Ensino Básico

João Otávio do Lago ¹
Anderson Ara ²

Resumo

A qualidade da educação básica no Brasil tem sido foco de atenção diante dos desafios enfrentados pelas redes públicas de ensino. Nesse contexto, a adoção de boas práticas educacionais tem se mostrado uma estratégia promissora para melhorar os indicadores de desempenho escolar. Este trabalho tem como objetivo analisar a influência de boas práticas na qualidade da educação no Ensino Fundamental – anos iniciais (1º ao 5º ano) – em escolas municipais do estado do Paraná. Foram utilizados dados do IDEB, Censo Escolar e coletas do Tribunal de Contas do Estado do Paraná, permitindo uma análise temporal e descritiva. As principais perguntas investigadas incluem: Quais são os impactos das boas práticas na qualidade do ensino? Esses impactos são significativos? O efeito é maior conforme o número de boas práticas implementadas? Foram utilizados como referência os municípios de Sobral (CE), Coruripe (AL), Apucarana (PR) e Paranavaí (PR), reconhecidos por bons resultados educacionais. A análise evidenciou que práticas como gestão escolar eficiente, formação continuada de professores, uso de dados na tomada de decisões e acompanhamento pedagógico têm forte associação com melhorias nos índices de qualidade. Os resultados apontam que os impactos tendem a ser mais expressivos quando as práticas são aplicadas de forma sistemática e contínua, reforçando a importância de políticas públicas baseadas em evidências. Este estudo busca contribuir com diagnósticos e reflexões que apoiam a formulação de ações para o aprimoramento da educação básica brasileira.

Palavras-chave: Ensino Básico; Boas Práticas; Educação de Qualidade; IDEB; Escolas Municipais.

¹Departamento de Estatística, UFPR - joao.lago@ufpr.br

²Departamento de Estatística, UFPR - ara@ufpr.br



Modelos de séries temporais para dados de contagem: Um estudo comparativo

Nathália Dantas Handam Nunes;¹ Ana Julia Alves Câmara.²

Resumo

Modelos de séries temporais para dados de contagem foram analisados com foco na comparação entre os modelos GLARMA (Generalized Linear Autoregressive Moving Average) e GARMA (Generalized Autoregressive Moving Average), aplicados à investigação da relação entre poluentes atmosféricos e a incidência de doenças respiratórias na Região da Grande Vitória, Espírito Santo. Inicialmente, modelos lineares generalizados (GLM) com distribuições Poisson e Binomial Negativa indicaram a presença de sobredispersão e dependência temporal nos resíduos. Para mitigar essas limitações, ajustou-se um modelo GLARMA com distribuição Binomial Negativa, incorporando componentes autorregressivos e de média móvel. A multicolinearidade entre variáveis explicativas foi tratada por meio da Análise de Componentes Principais (PCA). A avaliação do ajuste, com base nas funções de autocorrelação (ACF), autocorrelação parcial (PACF) e gráficos Q-Q plot indicou melhor desempenho do modelo GLARMA Binomial Negativo com PCA, embora limitações permaneçam na modelagem da autocorrelação residual. O prosseguimento do estudo prevê a aplicação e comparação com modelos GARMA, visando aprimorar a modelagem da estrutura temporal e contribuir para o entendimento dos impactos da poluição atmosférica na saúde pública. Tais evidências podem subsidiar políticas públicas voltadas à melhoria da qualidade do ar e à prevenção de doenças respiratórias na população.

Palavras-chave: séries temporais; dados de contagem; GLARMA; poluição atmosférica; saúde pública.

¹Departamento de Estatística, Universidade Federal do Espírito Santo, Vitória, ES - nathalia.d.nunes@edu.ufes.br

²Departamento de Estatística, Universidade Federal do Espírito Santo, Vitória, ES - ana.j.camara@ufes.br



Influência da Precipitação, Temperatura e os Fenômenos El Niño e La Niña no número de casos de dengue na Cidade de São Paulo

Giovanna Martins Verasto ¹

Isabella Viana Bambirra ²

Luan Tomimoto ³

Maria Júlia Gregório Romero ⁴

Murilo Cassiavilani ⁵

Márcio Luis Lanfredi Viola ⁶

Resumo

A dengue é uma doença causada por um arbovírus transmitida pelo mosquito Aedes aegypti, também responsável pelas transmissões de chikungunya e zika. No Brasil, esta doença tem sido um problema há muitos anos e tem sido um desafio, pois nos últimos anos o número de casos tem aumentado significativamente. Desta forma, é importante o estudo dos aumentos e quedas de número de casos de dengue, além de entender se fatores como a precipitação e a temperatura podem influenciá-lo. O objetivo deste trabalho é analisar como influenciam o número de casos de dengue no município de São Paulo utilizando modelos multivariados de séries temporais e causalidade de Granger. As séries temporais utilizadas correspondem de janeiro de 2009 até setembro de 2024. Além disso, é de interesse investigar como a ocorrência dos El Niño e La Niña podem influenciar o número de casos de dengue.

Palavras-chave: Aedes Aegypti; Causalidade de Granger; Dengue; Série Temporal Multivariada.

¹ Departamento de Estatística, Universidade Federal de São Carlos - gmverasto@estudante.ufscar.br

² Departamento de Estatística, Universidade Federal de São Carlos - isabellavb@estudante.ufscar.br

³ Departamento de Estatística, Universidade Federal de São Carlos - luantomimoto@estudante.ufscar.br

⁴ Departamento de Estatística, Universidade Federal de São Carlos - mariaromero@estudante.ufscar.br

⁵ Departamento de Estatística, Universidade Federal de São Carlos - muriloc@estudante.ufscar.br

⁶ Departamento de Estatística, Universidade Federal de São Carlos - lanfredi@ufscar.br



ANÁLISE DE CICLOS DA EXPORTAÇÃO DE CAFÉ NO BRASIL VIA MODELOS DE DECOMPOSIÇÃO DE SÉRIES TEMPORAIS (2015-2022)

Thaíi Céu;¹ Manoel Ivanildo Silvestre Bezerra;²

Resumo

O café, como *commodity* de grande relevância econômica global, é cultivado em mais de 60 países em desenvolvimento e exportado principalmente para mercados desenvolvidos, desempenhando papel crucial nas trocas comerciais internacionais. A competitividade no mercado de café intensificou-se desde os anos 80, demandando maior gestão do conhecimento em sua cadeia produtiva. Dados recentes do CECAFÉ mostram variações nas exportações brasileiras, destacando sua importância econômica. No contexto da análise de séries temporais, o estudo utiliza ferramentas com as séries harmônicas e as variáveis *dummies*, para decompor e compreender padrões nas exportações dos quatro tipos de café brasileiros: Conilon, Arábica, Torrado e Solúvel. Realizamos uma análise exploratória das variáveis volume, receita e preço médio, para os quatro tipos de cafés, calculando as principais métricas estatísticas, fazendo uma análise gráfica das séries, e aplicando o teste de estacionariedade do Dickey-Fuller. A seguir, ajustamos os modelos com variáveis *dummies*, e com variáveis harmônicas para as séries citadas, comparando-os por meio de algumas métricas usuais, e uma análise residual. Identificou-se padrões anuais e mensais, variações específicas por tipo de café e comportamentos sazonais. Por fim, foram comparados os modelos quanto à precisão e capacidade explicativa para entender as tendências e variações nos dados, e concluiu-se que os modelos com variáveis *dummies* e harmônicas têm vantagens distintas, com as *dummies* sendo mais preciso e com as harmônicas mais explicativo. Ambas as abordagens são úteis dependendo do objetivo da análise, seja para fazer previsões ou averiguar se existem tendências e/ou sazonalidade na série.

Palavras-chave: cafés: Conilon, Arábica, Torrado e Solúvel; modelos de decomposição temporal; exportações.

¹Departamento de Estatística/Unesp, Presidente Prudente/SP - thaíi.ceu@unesp.br

²Departamento de Estatística/Unesp, Presidente Prudente/SP - manoel.bezerra@unesp.br



Modelagem Preditiva de Temperatura: Uma Análise Comparativa para o Município de Juiz de Fora – MG

Geovane Fernandes da Silva ¹ Geraldo Magela da Cruz Pereira ² Gean Pereira Damaceno ³ João Paulo Assis Bonifácio ⁴ Pedro Mambelli Fernandes ⁵ Lara Reis Gualberto ⁶

Resumo

A previsão precisa de temperatura é essencial para áreas como agricultura e gestão energética. Este estudo utilizou dados meteorológicos diários do Instituto Nacional de Meteorologia (INMET) para Juiz de Fora - MG, coletados entre 2020 e 2025, totalizando mais de 1.800 observações. Implementamos quatro modelos de machine learning - Random Forest, XGBoost, Prophet e SVR (Support Vector Regression) utilizando validação cruzada temporal com janelas deslizantes na validação, respeitando a ordem temporal dos dados e evitando vieses. A otimização de hiperparâmetros foi realizada através de grid search para maximizar a precisão das previsões.

Os resultados demonstraram a superioridade do SVR (SMAPE = 1.00%) em relação aos demais modelos, seguido pelo Prophet (SMAPE = 1.14%). XGBoost e Random Forest apresentaram desempenho ligeiramente inferior, com SMAPE de 1.45% e 1.52%, respectivamente. O SVR demonstrou excelente capacidade de generalização e consistência nas previsões, com erro médio absoluto percentual simétrico (SMAPE) de apenas 1.00%, indicando alta precisão na captura de padrões temporais da temperatura. A diferença de 0.14% entre SVR e Prophet sugere que ambos os modelos são altamente adequados para esta aplicação, oferecendo excelente equilíbrio entre precisão e interpretabilidade. Este estudo contribui para a literatura sobre previsão meteorológica, fornecendo evidências empíricas sobre a eficácia relativa de diferentes abordagens de machine learning.

Palavras-chave: Machine Learning; Séries Temporais; Modelagem Preditiva; Meteorologia

¹ Departamento de Estatística, UFLA, Lavras/MG – geovanefs90@gmail.com

² Departamento de Estatística, UFLA, Lavras/MG – geraldo.pereira@ufla.br

³ Departamento de Estatística, UFLA, Lavras/MG – geandamaceno030@gmail.com

⁴ Departamento de Estatística, UFLA, Lavras/MG – jpab.27@hotmail.com

⁵ Departamento de Estatística, UFLA, Lavras/MG – pedromambelli@gmail.com

⁶ Departamento de Estatística, UFLA, Lavras/MG – larargualberto@hotmail.com



PREVISÃO DO NÚMERO DE SUICÍDIOS NO BRASIL A PARTIR DE MODELOS DE SÉRIES TEMPORAIS E MACHINE LEARNING

Kelly Pereira de Lima;¹ Raquel Araújo de Almeida;² Ricardo de Andrade Lira Rabelo.³

Resumo

O suicídio é um problema relevante de saúde pública, tanto no Brasil quanto em nível mundial, exigindo ações preventivas baseadas em dados consistentes e atualizados. Este trabalho tem como objetivo prever o número de suicídios no Brasil a partir da aplicação de modelos de séries temporais e técnicas de *machine learning*, utilizando dados oficiais do Sistema de Informações sobre Mortalidade (DATASUS) e do Instituto Brasileiro de Geografia e Estatística (IBGE), no período de 1996 a 2019. A metodologia envolveu a coleta, limpeza e integração das bases de dados, seguida de análises descritivas por região, estado, mês e ano. Posteriormente, foram aplicados os modelos de Suavização Exponencial Tripla (SET), ARIMA e Long Short-Term Memory (LSTM), com o objetivo de comparar o desempenho preditivo. Além disso, foi realizado o teste de Mann-Whitney para verificar se o crescimento do número de suicídios ao longo dos anos foi estatisticamente significativo. Os resultados revelaram um aumento expressivo e significativo no número de suicídios no período analisado, com destaque para as regiões Sul e Sudeste. Os modelos preditivos apresentaram bom desempenho, sendo o modelo LSTM o que melhor capturou os padrões temporais dos dados. Conclui-se que o uso de modelos de previsão pode contribuir de forma eficaz para a formulação de políticas públicas voltadas à prevenção do suicídio, servindo como ferramenta de apoio à tomada de decisões no âmbito da saúde pública.

Palavras-chave: Suicídio; Saúde Pública, Estatística.

¹Departamento de Computação (DC), UFPI - kelly.lima.88@gmail.com.br

²Pontifícia Universidade Católica de Minas Gerais - PUC Minas. - raquel.a.d.a@hotmail.com

³Departamento de Computação (DC), UFPI - ricardoalr@ufpi.edu.br



Modelo de Previsão do Preço do Etanol: Influência de Fatores Agrícolas, Energéticos e Macroeconômicos.

Isabele Alves Pereira;¹ Marilaine Colnago²

Resumo

Este estudo tem como objetivo construir uma análise comparativa de modelos estatísticos e de aprendizado de máquina com dados de variação do preço do etanol no estado de São Paulo. A modelagem incorpora variáveis agrícolas e macroeconômicas relevantes para o setor sucroenergético. Entre os fatores agrícolas analisados estão o preço do açúcar, o preço do ATR (Açúcar Total Recuperável) — indicador da qualidade da cana-de-açúcar — e o período de safra, que influencia diretamente a oferta do produto. Do ponto de vista macroeconômico, o estudo considera a inflação, a taxa de câmbio (dólar) e a cotação internacional do açúcar na bolsa de Nova York, variáveis que impactam commodities agrícolas em mercados globalizados. Além disso, são analisados os preços do barril de petróleo e da gasolina, uma vez que aumentos nesses combustíveis tendem a elevar a competitividade do etanol como alternativa energética. A metodologia envolve a aplicação de modelos de aprendizado de máquina como Catboost, XGBoost, Random Forest (Floresta Aleatória), Regressão Linear, LSTM (Long Short-Term Memory) e análise de séries temporais, com dados semanais de maio de 2004 a novembro de 2023. A performance dos modelos testados foram avaliados pelas métricas R^2 , MAPE, RMSE e MAE, buscando identificar os preditores mais relevantes e a acurácia das estimativas. Assim, dentre todos os modelos testados, o LSTM (Long Short-Term Memory) destacou-se como o modelo de melhor performance. Os resultados visam apoiar a tomada de decisão por produtores, usinas, agentes de mercado e formuladores de políticas públicas, promovendo previsibilidade e eficiência no setor de biocombustíveis.

Palavras-chave: Análise Preditiva; Biocombustíveis; Aprendizado de Máquina; Ciência de Dados; Cana-de-açúcar.

¹Faculdade de Ciências e Tecnologia, Presidente Prudente, Estatística - isabele.alves@unesp.br

²Faculdade de Ciências e Tecnologia, Presidente Prudente, Estatística - marilaine.colnago@unesp.br



Padrões de Similaridade da Base do Milho no Brasil: Uma Análise de Agrupamento de Séries Temporais

João Paulo Assis Bonifácio¹ Sérgio Domingos Simão² Pedro Mambelli Fernandes³

Izabela Regina Cardoso de Oliveira⁴

Resumo

A base (do inglês, basis) do milho, calculada como a diferença entre o preço físico local e o preço na B3 (Bolsa de Valores Brasileira, principal plataforma de negociação de contratos futuros), é fundamental para estratégias de negociação, logística e gestão de risco na comercialização dessa commodity. Este trabalho aplica a clusterização de séries temporais para agrupar 118 séries históricas da base de milho, cobrindo diversas localidades em todas as regiões do Brasil, com o objetivo de identificar regiões com dinâmicas semelhantes ao longo do tempo. Os dados foram extraídos da plataforma Broadcast, cujo acesso foi viabilizado pela JPA Agro, empresa de agronegócio sediada em Lavras, MG, e parceira neste trabalho. As séries temporais analisadas compreendem o período diário de 1º de janeiro de 2020 a 31 de dezembro de 2024. Para a análise, utilizamos a linguagem R e pacotes como tidyverse e dtwclust, empregando a distância Dynamic Time Warping (DTW) com avaliação de diferentes centróides. A seleção do número ótimo de clusters é um passo crucial, guiado pela avaliação de Índices de Validação de Cluster (CVIs), como o índice Silhouette, sendo que este processo apresenta considerável custo computacional. Os agrupamentos resultantes são visualizados em mapas utilizando o pacote leaflet, permitindo a exploração de possíveis correlações espaciais e a investigação da influência da proximidade regional e outros fatores geográficos nos padrões da base identificados. Os agrupamentos permitem analisar padrões de convergência ou divergência de preços, otimizar estratégias de hedge e entender as relações de mercado entre diferentes praças de negociação.

Palavras-chave: Dynamic Time Warping; Distâncias Estatísticas; Dinâmica de Preços.

Agradecemos à FAPEMIG pelo apoio financeiro e à empresa JPA Agro pela parceria.

¹Graduando em Estatística, Universidade Federal de Lavras – Lavras, MG – Brasil
joao.bonifacio2@estudante.ufla.br

²Pós-doutorando do Programa de Estatística e Experimentação Agropecuária, DES/ICET, Universidade Federal de Lavras – Lavras, MG – Brasil sergiodomigos@ufla.br

³Graduando em Estatística, Universidade Federal de Lavras – Lavras, MG – Brasil
pedromambelli@gmail.com

⁴DES/ICET Universidade Federal de Lavras – Lavras, MG – Brasil Izabela.oliveira@ufla.br



Evasão feminina no curso de Agronomia da Universidade Federal de Lavras: uma análise temporal

Melissa Wendy Ribeiro Vicente;¹ Valdeline de Paula Mequelinho Ferreira ² Marcela Silva de Araújo ³ Marluce Rodrigues Pereira ⁴ Izabela Regina Cardoso de Oliveira ⁵

Resumo

A presença feminina no ensino superior tem crescido nas últimas décadas. No entanto, ainda persistem desafios significativos para mulheres em cursos tradicionalmente masculinos, como os das ciências agrárias. A evasão universitária feminina pode ser motivada por uma variedade de fatores, incluindo questões socioeconômicas, culturais, familiares e institucionais. Nesse contexto, este estudo teve como objetivo modelar a evasão de mulheres no curso de Agronomia da Universidade Federal de Lavras, considerando o período de 1948 a 2024, em relação aos ingressos. Os dados foram cedidos pela Universidade. Para a modelagem, foi utilizado o modelo ARMA(1,1), cujos resíduos apresentaram comportamento estacionário, evidenciando a adequação do modelo aos dados. A partir desse ajuste, foram realizadas previsões para os anos de 2025 e 2026. Os resultados indicam um aumento na evasão nesse intervalo, em consonância com a tendência observada em anos anteriores. Estudos destacam a importância da formulação e implementação de políticas públicas que incentivem a permanência de mulheres no ensino superior, especialmente em áreas com histórico de desigualdade de gênero, como as ciências agrárias. Políticas que promovam inclusão, apoio financeiro, psicológico e acadêmico, bem como medidas efetivas contra a discriminação e o preconceito de gênero, são essenciais para garantir maior equidade. Essas ações contribuem para fortalecer a presença feminina tanto no meio acadêmico quanto no mercado de trabalho. .

Palavras-chave: evasão 1; universidade 2; mulheres 3; série temporal 4.

¹Departamento de Estatística, UFLA - melissa.vicente2@estudante.ufla.br

²Departamento de estatística, UFLA - valdeline.ferreira1@estudante.ufla.br

³Departamento de estatística, UFLA - marcela.araujo1@estudante.ufla.br

⁴Departamento de Computação Aplicada, UFLA - marluce@ufla.br

⁵Departamento de Estatística, UFLA - izabela.oliveira@ufla.br



Avaliações Preditivas de Modelos Temporais com Integração de Informações Espaciais: Um Estudo de Incidência da Covid-19 no Estado de São Paulo em 2020

Thiago Moraes Rizzieri;¹ José Silvio Govone²

Resumo

A disseminação global do COVID-19 representou um desafio significativo, dada sua natureza como uma infecção respiratória aguda altamente contagiosa e potencialmente grave. Em resposta a essa emergência, o Comitê de Emergência da Organização Mundial da Saúde (OMS) declarou estado de emergência global em 30 de janeiro de 2020. A complexidade da situação e o surgimento contínuo de novos modelos estatísticos destacaram a importância de abordagens preditivas para avaliar a evolução da pandemia. Este estudo apresenta uma análise espaço-temporal dos dados de COVID-19 no estado de São Paulo em 2020, fornecidos pelo SEADE. Para a análise temporal, foram utilizados os modelos ARIMA, Floresta Aleatória e Redes Neurais LSTM. Esses modelos foram aplicados para prever a média móvel semanal da incidência de casos de COVID-19. Além disso, este estudo compara dois modelos espaciais: o modelo de autoregressão espacial simultânea (SAR) e o modelo de redes neurais de grafos (GNN), para modelar a defasagem espacial. A análise espacial foi realizada para entender melhor a distribuição geográfica dos casos de COVID-19 e para verificar se a integração das informações espaciais nos modelos temporais pode melhorar o desempenho das previsões. Os resultados indicam que a consideração da dependência espacial pode aprimorar significativamente as previsões temporais, oferecendo uma compreensão mais completa da dinâmica de transmissão do vírus e contribuindo para a tomada de decisões mais informadas no combate à pandemia.

Palavras-chave: COVID-19; Série Temporal; Estatística Espacial; ARIMA; Redes Neurais.

¹Departamento de Biodiversidade e Bioestatística IBB-UNESP, Botucatu - thiago.m.rizzieri@unesp.br

²Departamento de Estatística, Matemática Aplicada e Computação IGCE-UNESP, Rio Claro - js.govone@unesp.br



Previsão do PIB do Agronegócio Brasileiro para 2025: Análise Temporal e Modelagem Estatística

Yasmin Beatriz Pereira Santana;¹ João Vitor Andrade Alves de Souza ² Melissa Wendy Ribeiro Vicente ³ Valdeline de Paula Mequelinho Ferreira ⁴ Luiz Otávio De Oliveira Pala ⁵ Renato Ribeiro de Lima ⁶

Resumo

O Produto Interno Bruto (PIB) do agronegócio brasileiro desempenha um papel fundamental na economia nacional, abrangendo os segmentos de agricultura, pecuária, insumos, agroindústria e serviços. Este trabalho tem como objetivo realizar previsões trimestrais para o ano de 2025, analisando a influência do agronegócio no PIB por meio de séries temporais. A base de dados utilizada foi fornecida pelo IBGE, contendo valores encadeados a preços de 1995, registrados trimestralmente no período de 1996 a 2024. Para a análise, utilizou-se o software R e foram aplicadas diversas técnicas, incluindo a decomposição da série para verificação dos pressupostos de tendência e sazonalidade, bem como a divisão dos dados em conjuntos de treino e teste. Foi necessário aplicar uma transformação logarítmica (\ln). Em seguida, foi ajustado um modelo SARIMA. A análise dos resíduos incluiu a verificação dos gráficos da Função de Autocorrelação (ACF) e da Função de Autocorrelação Parcial (PACF), permitindo o refinamento da estrutura do modelo. No contexto das previsões, observou-se uma leve tendência positiva no PIB do agronegócio para 2025, sugerindo um possível crescimento do setor nesse ano.

Palavras-chave: Séries Temporais; IBGE; Agropecuária.

¹Programa de Pós Graduação em Estatística e Experimentação Agrapecuária- UFLA, Lavras - yasmin.santana@estudante.ufla.br

²Programa de Pós Graduação em Estatística e Experimentação Agrapecuária- UFLA, Lavras - joao.souza20@estudante.ufla.br

³Programa de Pós Graduação em Estatística e Experimentação Agrapecuária- UFLA, Lavras - melissa.vicente2@estudante.ufla.br

⁴Programa de Pós Graduação em Estatística e Experimentação Agrapecuária- UFLA, Lavras - valde-line.ferreira1@estudante.ufla.br

⁵Departamento de Estatística- UFLA, Lavras - luizpala@ufla.br

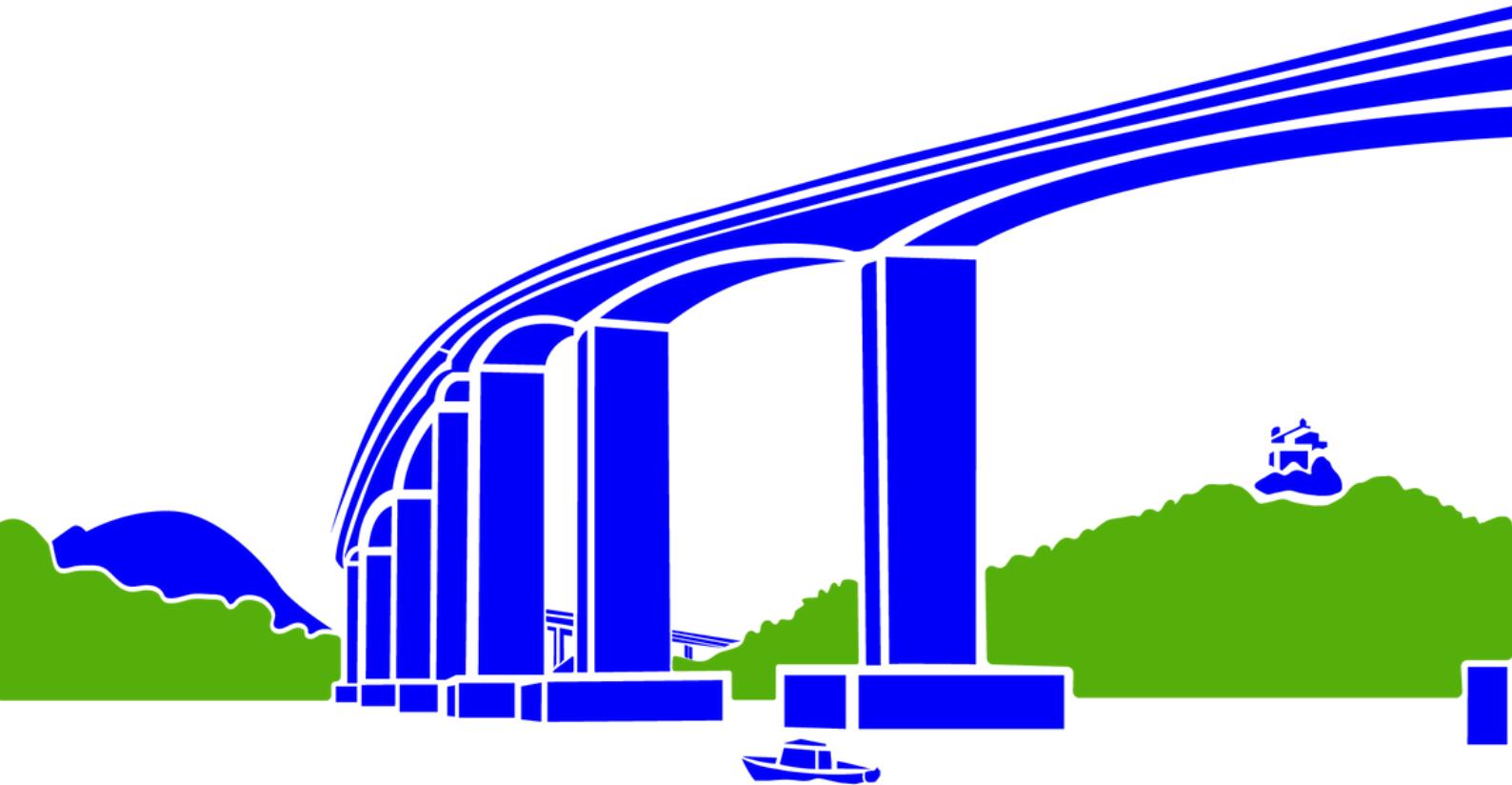
⁶Departamento de Estatística- UFLA, Lavras - rrlima@ufla.br

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Pôsters

**Probabilidade e Processos
Estocásticos**





Comparação do desempenho de métodos de seleção de árvores de contexto

Riquelme Nascimento dos Santos;¹ Ricardo Felipe Ferreira²

Resumo

Os modelos de árvore de contexto probabilísticas, propostos por Jorma Rissanen, descrevem cadeias estocásticas que incorporam grande flexibilidade em relação ao número de parâmetros dos modelos Markovianos tradicionais. Essa abordagem permite uma representação mais parcimoniosa e um método de detectar padrões complexos com maior eficiência. Essa classe de modelos tem sido amplamente aplicada em diversos campos da ciência, incluindo finanças, genética, neurociência e linguística. Neste estudo, exploramos metodologias de inferência para a estimativa de árvores de contexto, analisando o algoritmo Contexto de Rissanen, algumas as modificações propostas na literatura e um método de penalização da verossimilhança via critério de informação Bayesiano (BIC). Também investigamos uma abordagem Bayesiana recentemente desenvolvida, denominada árvores de contexto Bayesianas. Avaliamos essas metodologias por meio de cadeias estocásticas simuladas computacionalmente e em bases de dados reais, comparando o desempenho dos métodos em termos de acurácia, consistência dos estimadores e eficiência computacional dos algoritmos.

Palavras-chave: Algoritmo Contexto; Seleção de modelos; Cadeias estocásticas com memória de alcance variável; Inferência em processos estocásticos.

¹Programa Interinstitucional de Pós Graduação em Estatística - UFSCar/USP, São Carlos, SP - riquelme@usp.br

²DEs - Departamento de Estatística - UFSCar, São Carlos, SP - rferreira@ufscar.br



Processos de Poisson não homogêneos de n tipos - Uma possível aplicação no poder Judiciário

Bruno Gondim Toledo ¹

Resumo

O presente trabalho tem por objetivo responder a seguinte pergunta: "Quantos processos serão julgados em um determinado tribunal em um ano?". Modelar quantidade de processos julgados pode ser relevante para, por exemplo, estimar custos, número de funcionários necessários, possível expansão ou diminuição do tribunal em função desta produtividade. Para responder esta pergunta, modelou-se uma soma de processos de Poisson não homogêneos.

Palavras-chave: Modelagem; Variável aleatória.

¹Departamento de Estatística, Universidade de Brasília (UnB) - brunogtoledo96@gmail.com



Tempo de espera para a ocorrência de palavras em um alfabeto finito

Eduardo Rio Oliveira ¹

Renato Jacob Gava ²

Resumo

Considere uma moeda justa e denote cara (C) e coroa (K). Considere uma palavra formada apenas por C's e K's. Arremesse essa moeda repetidamente e anote o resultado de cada lançamento. Com isso, algumas perguntas relevantes poderiam ser feitas, tais como "Qual o tempo de espera até que a palavra CCCCC(podemos provar que é 62) ocorra sem interrupção?" ou "Qual a probabilidade de a palavra CCCCK ocorrer pela primeira vez antes da palavra CCCKC?(apesar de parecer contra intuitivo, o resultado é 2/3)" "O tempo de espera é finito para qualquer palavra?". Este trabalho estuda essa e outras questões mais gerais dessa ideia, introduzindo a construção de um martingal a partir do qual essas perguntas são respondidas tanto para o caso em que o alfabeto é {C,K} quanto para alfabetos finitos quaisquer. Além disso, exibe-se um algoritmo em R que, a partir da entrada de um alfabeto e sequencia de palavras, simula tal experimento e nos fornece estimativas para o tempo de espera e a probabilidade de uma palavra dada ser a primeira a ocorrer.

Palavras-chave: Tempo de espera; probabilidade; martingal; algoritmo.

¹ Departamento de Estatística, Universidade Federal de São Carlos – eduardorio@estudante.ufscar.br

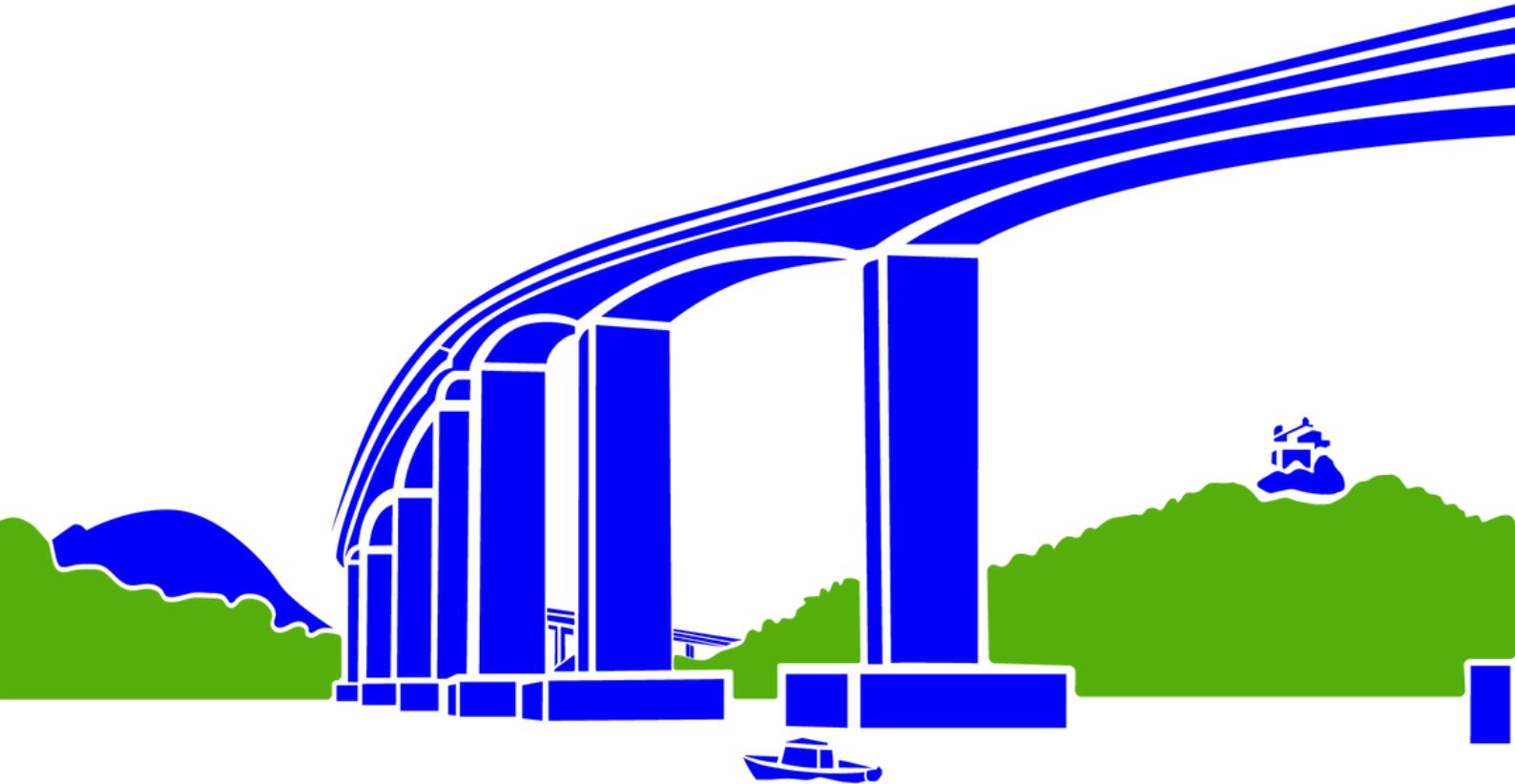
² Departamento de Estatística, Universidade Federal de São Carlos – gava@ufscar.br

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Pôsters

**Planejamento do Experimento
e Controle Estatístico de
Qualidade**





Aplicações de Planejamento de Experimentos (DOE) na Engenharia: Uma Revisão de Estudos de Caso e Práticas

Leandro Santos Ribeiro¹

Paulo César de Resende Andrade²

Resumo

O Planejamento de Experimentos (DOE) é uma metodologia estatística fundamental na engenharia, que possibilita a análise sistemática de processos e a otimização de produtos por meio da manipulação controlada de variáveis. Sua aplicação tem crescido em diversas áreas, como engenharia química, mecânica e de materiais, contribuindo significativamente para o aumento da eficiência, a melhoria da qualidade e a redução de custos operacionais. Exemplos notáveis de sua utilização incluem a otimização de processos de usinagem, o aprimoramento da eficiência térmica de trocadores de calor e o desenvolvimento de ligas metálicas de alta resistência. Apesar de sua eficácia, a implementação do DOE ainda enfrenta desafios, principalmente devido à complexidade metodológica e à necessidade de capacitação especializada. Este estudo apresenta uma revisão de casos práticos publicados entre 2009 e 2024, extraídos da base de dados acadêmica ScienceDirect, com foco nas aplicações do DOE em contextos industriais e científicos. Destacam-se os métodos fatorial e de Taguchi, bem como metodologias complementares, como a análise de variância (ANOVA) e as superfícies de resposta, que possibilitam a avaliação simultânea de múltiplos fatores. Esses métodos facilitam a tomada de decisões e contribuem para a redução de custos experimentais. Os resultados obtidos indicam que o DOE tem se consolidado como uma ferramenta crucial na otimização de processos produtivos e a melhoria da qualidade de produtos e sistemas. No entanto, ainda há oportunidades para aprimorar sua aplicação, especialmente por meio da simplificação metodológica e do fortalecimento da formação profissional, o que pode ampliar sua adoção na engenharia e impulsionar inovações tecnológicas em diversas áreas.

Palavras-chave: DOE; Design of Experiments; Process Optimization; Statistic; Taguchi.

¹ Instituto de Ciéncia e Tecnologia, Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM), Diamantina – MG - leandro.ribeiro@ufvjm.edu.br

² Instituto de Ciéncia e Tecnologia, Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM), Diamantina – MG - paulo.andrade@ict.ufvjm.edu.br



Tamanho ótimo de parcela para avaliar a massa de parte área de grão-de-bico com base no modelo linear de resposta com platô

Alberto Cargnelutti Filho ¹

Murilo Vieira Loro ²

Vithória Morena Ortiz ³

João Augusto Andretta ⁴

Resumo

O objetivo deste trabalho foi determinar o tamanho ótimo de parcela para avaliar a massa de parte área da planta de grão-de-bico (*Cicer arietinum* L.), cultivar BRS Cristalino. Foram conduzidos três ensaios de uniformidade com a cultura de grão-de-bico, cultivar BRS Cristalino, em área do Departamento de Fitotecnia da Universidade Federal de Santa Maria, estado do Rio Grande do Sul. Os ensaios de uniformidade foram divididos em 36 unidades experimentais básicas (UEB) de 1 m × 1 m. A UEB foi formada por duas fileiras de 1,0 m de comprimento, espaçadas 0,50 m entre fileiras, totalizando 1,0 m². Foi avaliada a massa de parte área da planta de grão-de-bico de cada UEB. Foi determinado o tamanho ótimo de parcela por meio do método do modelo linear de resposta com platô. O tamanho ótimo de parcela para avaliar massa de parte área da planta de grão-de-bico (*Cicer arietinum* L.), cultivar BRS Cristalino, é 7,19 m² e a precisão experimental estabiliza a partir desse tamanho.

Palavras-chave: *Cicer arietinum* L.; ensaio de uniformidade; planejamento experimental.

¹ Departamento de Fitotecnia - Universidade Federal de Santa Maria - alberto.cargnelutti.filho@gmail.com

² Departamento de Fitotecnia - Universidade Federal de Santa Maria - muriloloro@gmail.com

³ Programa de Pós-Graduação em Agronomia - Universidade Federal de Santa Maria - vithoria.ortiz159@gmail.com

⁴ Curso de Graduação em Agronomia - Universidade Federal de Santa Maria - joao.andretta@acad.ufsm.br

Agradecimentos ao CNPq (Processos 304652/2017-2 e 304878/2022-7), PIBIC/CNPq/UFSM, PROBIC/FAPERGS/UFSM e Capes (Finance Code 001), pelo auxílio financeiro e bolsas concedidas aos autores. Aos estudantes bolsistas e voluntários pelo auxílio na coleta de dados.



Tamanho de amostra para estimação da média das massas fresca e seca da cultivar Trifecta de alfafa

Manoela Conterato Della Pace ¹

Alberto Cargnelutti Filho ²

Lais Cherobini ³

Leonardo Cesar Pradebon ⁴

João Augusto Andretta ⁵

Debora Luiza Pozzebon ⁶

Mikael Brum dos Reis ⁷

Resumo

O objetivo deste trabalho foi determinar o tamanho de amostra (número de plantas) necessário para avaliar a massa fresca da parte aérea (MFPA), massa fresca da raiz (MFR), massa seca da parte aérea (MSPA) e massa seca da raiz (MSR) da cultivar Trifecta de alfafa. Foram realizados dois ensaios de uniformidade (semeaduras em 27/11/2023 e 25/01/2024), em Santa Maria - RS. Em 110 plantas de cada ensaio foram avaliadas a MFPA, MFR, MSPA e MSR. Foi determinado o tamanho de amostra assumindo erro de estimativa (semi-amplitude do intervalo de confiança) igual a 10% da estimativa da média, com grau de confiança ($1-\alpha$) de 95%. O tamanho de amostra para a estimativa da média variou entre 132 (MFR) a 283 plantas (MSPA) (semeadura em 27/11/2023) e 163 (MFR) a 284 plantas (MSPA) (semeadura em 25/01/2024). Na média das duas datas de semeadura o tamanho de amostra foi de 229, 148, 284 e 177 plantas para estimativa da média dos caracteres MFPA, MFR, MSPA e MSR, respectivamente.

Palavras-chave: *Medicago sativa* L.; ensaio de uniformidade; estimativa da média.

¹ Curso de Graduação em Agronomia - Universidade Federal de Santa Maria (UFSM) - manoelaconterato@gmail.com

² Departamento de Fitotecnia - UFSM - alberto.cargnelutti.filho@gmail.com

³ Programa de Pós-Graduação em Agronomia - UFSM - laischerobini@gmail.com

⁴ Programa de Pós-Graduação em Agronomia - UFSM - leonardopradebon@gmail.com

⁵ Curso de Graduação em Agronomia - UFSM - aundregusto@gmail.com

⁶ Curso de Graduação em Agronomia - UFSM - deborapozzebon11@gmail.com

⁷ Programa de Pós-Graduação em Agronomia - UFSM - mikareis1999@gmail.com

Agradecimentos ao CNPq (Processo 304878/2022-7), PIBIC/CNPq/UFSM,

PROBIC/FAPERGS/UFSM e Capes (Finance Code 001), pelo auxílio financeiro e bolsas concedidas aos autores. Aos estudantes bolsistas e voluntários pelo auxílio na coleta de dados.



Tamanho de amostra para estimação da média de caracteres produtivos da cultivar Crioula de alfafa

Lais Cherobini ¹

Alberto Cargnelutti Filho ²

Leonardo Cesar Pradebon ³

Mikael Brum dos Reis ⁴

João Augusto Andretta ⁵

Debora Luiza Pozzebon ⁶

Manoela Conterato Della Pace ⁷

Resumo

O objetivo deste trabalho foi determinar o tamanho de amostra (número de plantas) necessário para avaliar caracteres produtivos de alfafa (*Medicago sativa* L.), cultivar Crioula. Foram conduzidos dois ensaios de uniformidade (semeaduras em 27/11/2023 e 25/01/2024), em Santa Maria - RS. Em 110 plantas, de cada ensaio, foram avaliadas a massa fresca da parte aérea (MFPA), massa fresca da raiz (MFR), massa seca da parte aérea (MSPA) e massa seca da raiz (MSR). Foi calculado o tamanho de amostra assumindo erro de estimativa (semi-amplitude do intervalo de confiança) igual a 10% da estimativa da média, com grau de confiança ($1-\alpha$) de 95%. O tamanho de amostra variou de 119 (MFR) a 176 plantas (MSPA), na semeadura de 27/11/2023, e de 111 (MSR) a 222 plantas (MFPA), na semeadura de 25/01/2024. Na média das duas datas de semeadura o tamanho de amostra foi de 183, 116, 197 e 120 plantas para estimativa da média da MFPA, MFR, MSPA e MSR, respectivamente.

Palavras-chave: *Medicago sativa* L.; intervalo de confiança; estimativa da média.

¹ Programa de Pós-Graduação em Agronomia - Universidade Federal de Santa Maria (UFSM) - laischerobini@gmail.com

² Departamento de Fitotecnia - UFSM - alberto.cargnelutti.filho@gmail.com

³ Programa de Pós-Graduação em Agronomia - UFSM - leonardopradebon@gmail.com

⁴ Programa de Pós-Graduação em Agronomia - UFSM - mikareis1999@gmail.com

⁵ Curso de Graduação em Agronomia - UFSM - aundregusto@gmail.com

⁶ Curso de Graduação em Agronomia - UFSM - deborapozzebon11@gmail.com

⁷ Curso de Graduação em Agronomia - UFSM - manoelaconterato@gmail.com

Agradecimentos ao CNPq (Processo 304878/2022-7), PIBIC/CNPq/UFSM,

PROBIC/FAPERGS/UFSM e Capes (Finance Code 001), pelo auxílio financeiro e bolsas concedidas aos autores. Aos estudantes bolsistas e voluntários pelo auxílio na coleta de dados.



Galpões de avícolas comerciais são parcelas apropriadas para experimentos de nutrição?

Iago Marçal Costa dos Santos¹, Davi Vitório Vicentin Vitti¹, Renata Alcarde Sermarini¹, Taciana Vilella Savian¹, José Henrique Tostes Barbi², Naiara Simarro Fagundes²

Resumo

Na avaliação da nutrição de aves há muito questiona-se a possibilidade de utilização de galpões de alojamento para produção comercial para a condução de ensaios com diferentes formulações de ração. Para investigar a similaridade entre os galpões e o poder do teste na identificação de efeitos significativos da ração, foram analisados dados de uma empresa com informações de lotes de frango de corte ao longo de um ano, totalizando 52 galpões com cinco ou mais ocorrências. Utilizou-se a técnica Fuzzy C-Means para análise de agrupamento e foram definidos 21060 cenários para simulação do ganho de peso médio diário (GPD). Os galpões foram aleatoriamente designados a dois tratamentos: um grupo recebeu um incremento no GPD (tratado) e o outro não (controle). Modelos foram ajustados para avaliar os efeitos de blocos e tratamentos, registrando valores de probabilidade. Com um nível de significância de 5%, foram calculadas as proporções de significância do tratamento. Para o delineamento em blocos casualizados, um efeito de aproximadamente 2,5 g foi necessário para rejeitar a hipótese nula do teste para os efeitos de tratamentos em 95% dos casos, resultando em um ganho de 115 g por ave. No delineamento inteiramente casualizado, o efeito foi maior, cerca de 3,3 g, refletindo em 150 g por ave. Assim, um ganho de peso final substancial é necessário para identificar um efeito significativo da ração, sendo mais desafiador quando os galpões são selecionados aleatoriamente.

Palavras-chave: Avicultura, Experimentação a campo, Poder do teste, Simulação, Similaridade.

¹ Luiz de Queiroz College of Agriculture (ESALQ), University of São Paulo, Piracicaba, São Paulo, ralcarde@usp.br. The authors are grateful to CAPES.

² Adisseo Brasil Nutrição Animal.



Desempenho do algoritmo de troca por coordenada em modelos polinomiais fracionários

Samantha Navarro Janine ¹

Marcelo Andrade da Silva ²

Resumo

Quando a relação entre uma variável resposta e suas covariáveis é não linear, polinômios são frequentemente usados para representar suas curvas. Uma alternativa promissora é o uso de polinômios fracionários, cujas potências pertencem a um conjunto predefinido de valores. Para extrair o máximo de informação acerca dos parâmetros, recorre-se aos delineamentos ótimos, que consistem em escolher um determinado número de pontos na região experimental dada pela combinação dos fatores considerados viáveis para o experimento de forma que atendam a um determinado critério de otimalidade. Para obter os delineamentos ótimos, pode-se utilizar um algoritmo de troca e/ou suas extensões. Dentre elas, está o algoritmo de troca por coordenada, o qual consiste em trocar as coordenadas da matriz de delineamento a partir de um processo iterativo e avaliar se tal troca contribui para maximização do valor do critério considerado. O objetivo deste trabalho é avaliar o desempenho do algoritmo de troca por coordenada aplicado em modelos polinomiais fracionários. Para isso, diferentes cenários são considerados combinando tamanho amostral, número de fatores, grau do polinômio e abordagens (localmente ótimo ou pseudo-Bayesiano). O estudo está em andamento e espera-se comparar o tempo computacional e a eficiência da otimalidade entre as abordagens de troca por linha e por coluna nos diferentes cenários.

Palavras-chave: algoritmo de Fedorov; delineamentos ótimos; eficiência computacional; modelos não-lineares; mfp.

¹ Departamento de Ciências Exatas, ESALQ/USP, Piracicaba-SP – samantha.nj@usp.br

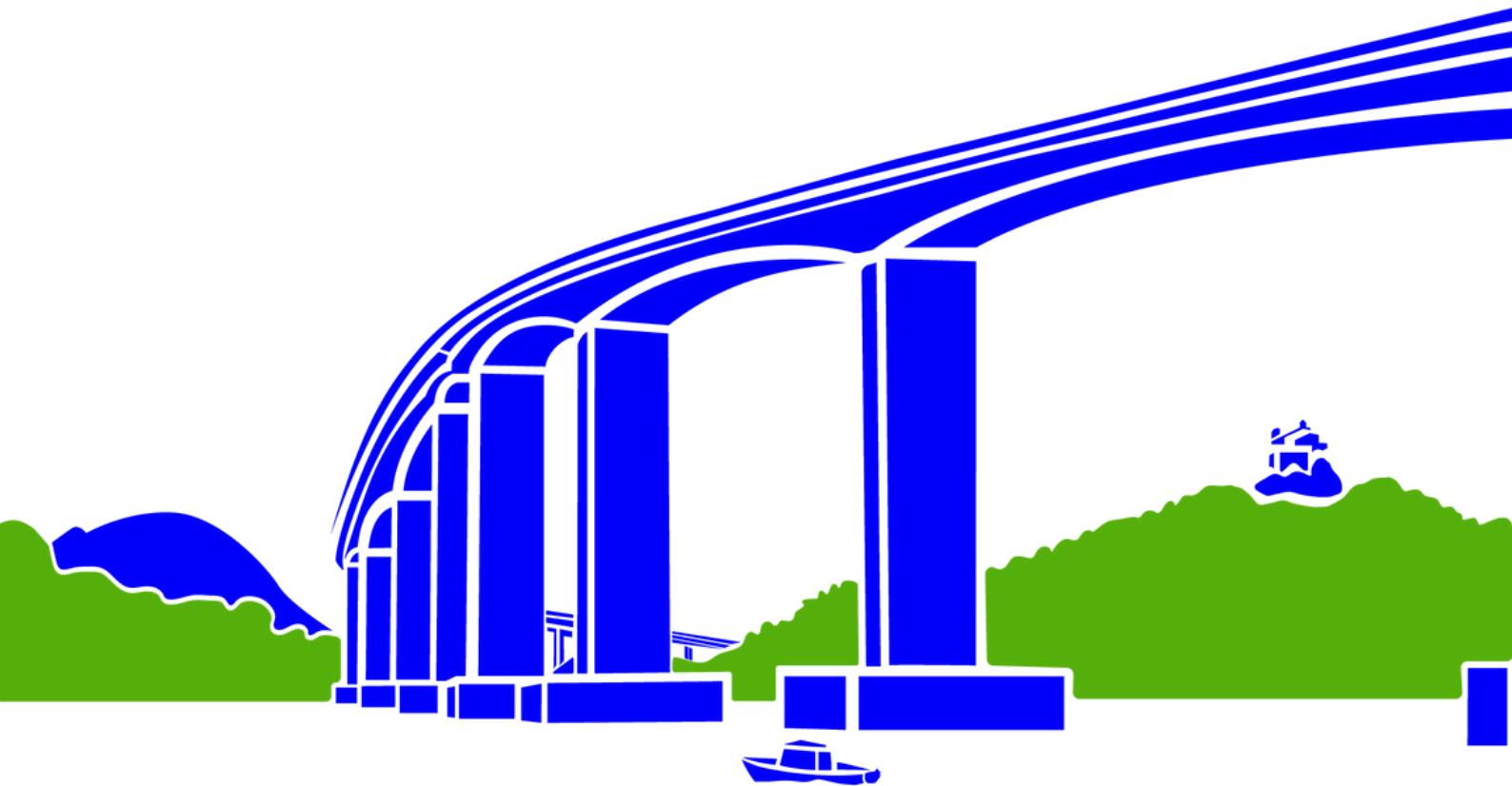
² Departamento de Ciências Exatas, ESALQ/USP, Piracicaba-SP – silva.marcelo@usp.br

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Pôsters

**Modelos lineares generalizados
e modelos não lineares**





Estimando incertezas na predição do excesso de mortalidade por COVID-19, a partir do MLG Misto, via simulação

Fernando Andrade ¹
Luzia Aparecida Trinca ²

Resumo

Resumo

A análise da mortalidade em eventos pandêmicos é fundamental para avaliar impactos na saúde populacional e orientar políticas públicas. No Brasil, a COVID-19 destacou desafios críticos: além das mais de 700 mil mortes diretas, a gestão precária da saúde em 2020-2021 agravou efeitos indiretos, tais como barreiras ao acesso a tratamentos, deterioração socioeconômica e impactos na saúde mental. Para quantificar esse impacto, este trabalho propõe técnicas de simulação computacional para estimar incertezas da predição do excesso de mortalidade – diferença entre mortes observadas e esperadas sob a hipótese da não ocorrência da COVID-19 – utilizando Modelos Lineares Generalizados Mistos (MLGM). A escolha dos MLGMs justifica-se por sua capacidade de lidar com dados de contagem, correlacionados e heterogêneos. Foram testadas as distribuições Gaussiana, justificada pela boa aproximação aos dados, e Binomial Negativa, ambas adequadas ao contexto, com destaque para a eficácia da Binomial Negativa em capturar sobredispersão nos dados de mortalidade. Os dados modelados estão disponíveis no Sistema de Informação sobre Mortalidade (SIM) do Ministério da Saúde, Brasil. A contribuição central deste estudo reside no método simulatório para construção de intervalos de predição, que consiste em gerar observações sintéticas, usando as estimativas dos parâmetros obtidas pelo ajuste do modelo, para grupos existentes e novos, permitindo quantificar incertezas mesmo em cenários sem estrutura analítica fechada. Os resultados demonstram que as simulações reproduziram adequadamente a variabilidade dos dados, com intervalos de predição coerentes para ambas as distribuições. No caso Gaussiano, as estimativas analíticas e simulacionais mostraram-se equivalentes, validando a robustez do método. Já para a Binomial Negativa, as simulações superaram as limitações teóricas. Essa abordagem inovadora em modelos de efeitos mistos reforça sua importância em estudos de mortalidade com cenários complexos, fornecendo ferramentas replicáveis para futuras crises sanitárias.

Palavras-chave: Modelo Binomial Negativa Misto;Intervalo de predição; Dados correlacionados.

¹Doutorando em Biometria, Departamento de Biodiversidade e Bioestatística, Unesp, Botucatu - f.andrade@unesp.br

²Departamento de Biodiversidade e Bioestatística, Unesp, Botucatu - luzia.trinca@unesp.br



Modelagem e Análise do Processo de Secagem de Grãos de Soja em Diferentes Temperaturas

Natiele de Almeida Gonzaga ¹; Rafaela de Carvalho Salvador ²; Edilene Cristina Pedroso Azarias ³; João Vitor Andrade Alves de Souza ⁴; Edilson Marcelino Silva ⁵; Joel Augusto Muniz ⁶; Alessandra Querino da Silva ⁷

Resumo

A colheita antecipada com grãos sadios é essencial para minimizar perdas na produção de soja. Devido ao alto teor de umidade no momento da colheita, a secagem é uma etapa crucial para a preservação da qualidade dos grãos. Os modelos não lineares descrevem a redução da quantidade de água em um alimento ao longo do tempo, geralmente representada por uma curva sigmoidal decrescente. Este estudo investigou a secagem de grãos de soja em temperaturas de 45°C, 60°C e 80°C, ajustando o modelo de Page aos dados experimentais. O ajuste foi realizado por regressão não linear, utilizando o Método de Mínimos Quadrados Generalizados e o algoritmo de Gauss-Newton no software R. O teste de Durbin-Watson indicou a presença de autocorrelação residual de primeira ordem, corrigida com a inclusão de um parâmetro autoregressivo AR(1). As taxas de secagem foram de 0,5694 g/h a 45°C, 0,6589 g/h a 60°C e 1,0191 g/h a 80°C. Além disso, o coeficiente de difusão efetiva aumentou com a temperatura, alcançando $1,23 \times 10^{-6}$ m²/s a 45°C, $1,26 \times 10^{-6}$ m²/s a 60°C e $2,21 \times 10^{-6}$ m²/s a 80°C. Os resultados evidenciam a influência da temperatura no processo de secagem, indicando que temperaturas mais elevadas aceleram a remoção de umidade dos grãos e aumentam o coeficiente de difusão efetiva. A modelagem utilizando o modelo de Page mostrou-se adequada para descrever as curvas de secagem, contribuindo para o aprimoramento de estratégias na conservação da qualidade dos grãos de soja.

Palavras-chave: Secagem de grãos; modelo de Page; difusão efetiva; regressão não linear; temperatura do ar.

¹ Departamento de Estatística, Universidade Federal de Lavras - natiele.gonzaga1@estudante.ufla.br

² Departamento de Estatística, Universidade Federal de Lavras - rafaela.salvador3@estudante.ufla.br

³ Departamento de Estatística, Universidade Federal de Lavras - edilene.azarias1@estudante.ufla.br

⁴ Departamento de Estatística, Universidade Federal de Lavras - joao.souza20@estudante.ufla.br

⁵ Departamento de Matemática, Universidade Federal Rural do Rio de Janeiro - edilsonest@ufrj.br

⁶ Departamento de Estatística, Universidade Federal de Lavras - joamuniz@ufla.br

⁷ Faculdade de Ciências Exatas e Tecnologia, Universidade Federal da Grande Dourados - alessandrasilva@ufgd.edu.br



Comparação de modelos não lineares na descrição da mineralização do nitrogênio no solo com resíduos da produção animal

Vitória Cristina Mendes da Silva ¹; Edilson Marcelino Silva ²; Joel Augusto Muniz ³

Resumo

A mineralização do nitrogênio em solos tratados com compostos orgânicos resultantes de resíduos da produção animal, é um processo essencial para disponibilização de nutrientes para as plantas. Esse processo traz melhoria na fertilidade do solo, reduz a necessidade de fertilizantes sintéticos, e contribui para a sustentabilidade agrícola, além disso, minimiza impactos ambientais, como a lixiviação de nitrato e a emissão de gases de efeito estufa. Os resíduos são de caprinos e ovinos, denominados como pequenos ruminantes, animais importantes na agricultura. Objetivou-se com este trabalho comparar o ajuste dos modelos não lineares de Stanford e Smith, Michaelis-Menten e Cabrera, ao descrever a mineralização do nitrogênio quando utilizou doses desse composto de pequenos ruminantes ao longo de 126 dias. Os dados utilizados foram extraídos de Araújo et al. (2020) e avaliados por meio do Software *R*, foram analisadas as doses de $0 \text{ M}_g \text{ha}^{-1}$; $3,75 \text{ M}_g \text{ha}^{-1}$ e $30 \text{ M}_g \text{ha}^{-1}$. A estimativa dos parâmetros dos modelos foi realizada utilizando o método dos mínimos quadrados com o algoritmo de convergência de Gauss-Newton. Para verificar os pressupostos de normalidade, homocedasticidade e independência utilizou-se os testes estatísticos de Shapiro-Wilk, Breusch-Pagan e Durbin-Watson, respectivamente. A qualidade do ajuste dos modelos foi verificada por meio do Coeficiente de Determinação (R^2), Desvio Padrão Residual (DPR) e do Critério de Informação de Akaike (AIC). Ao analisar os resultados obtidos, conclui-se que, para as doses $0 \text{ M}_g \text{ha}^{-1}$ e $3,75 \text{ M}_g \text{ha}^{-1}$ o modelo Cabrera foi o mais adequado para descrever os dados, já para dose $30 \text{ M}_g \text{ha}^{-1}$ foi o de Michaelis-Menten.

Palavras-chave: Stanford e Smith; Michaelis-Menten; Cabrera; Nutrientes; Pequenos Ruminantes.

¹ Departamento de Estatística, UFLA – vitoria.silva22@estudante.ufla.br

² Departamento de Matemática, UFRRJ – edilsonest@ufrj.br

³ Departamento de Estatística, UFLA – joamuniz@ufla.br

³



Análise da influência de características estruturais e variabilidade no mercado imobiliário: uma abordagem GAMLSS

Noé Osório Macário^{1*}; Elias M. Sabe¹; Paulo H.S. Guimarães²; Luiz R. Nakamura²;
Andréa C. Konrath³

Resumo

O mercado imobiliário tem atraído crescente interesse tanto de pesquisadores quanto de investidores, devido à sua complexidade e importância econômica. Normalmente, os preços das propriedades são influenciados por uma ampla gama de variáveis. Além disso, esses preços frequentemente apresentam comportamentos dinâmicos e não lineares, tornando a modelagem e a previsão de valores imobiliários um desafio. Este estudo analisa a influência de características estruturais e a variabilidade no mercado imobiliário de Portugal. Foram utilizadas 22.774 observações, tendo como variável resposta o preço da propriedade (em euros) e quatro covariáveis candidatas: presença de estacionamento (sim ou não), ano de construção, área total da propriedade (m^2) e o número de cômodos. Para a modelagem, utilizou-se os modelos aditivos generalizados para locação, escala e forma (GAMLSS), que permitem modelar não apenas a média da variável resposta, mas também sua variabilidade, assimetria e curtose. Dada a assimetria observada no preço da propriedade, testaram-se quatro distribuições: gama, Box-Cox Cole e Green, Box-Cox t (BCT) e Box-Cox exponencial potência. Após um processo de seleção de covariáveis baseado no *stepwise*, o modelo baseado na distribuição BCT foi considerado o melhor ajustado com base no critério de informação de Akaike, indicando que a mediana e a dispersão são influenciadas por todas as covariáveis, ao passo que a assimetria é influenciada pelo tamanho da área, e a curtose explicada por todas as covariáveis, com exceção do tamanho da área. A avaliação de resíduos indicou que o modelo é adequado para explicar os preços dos imóveis em Portugal.

Palavras-chave: regressão distribucional; mercado imobiliário; *bucket plot*; variabilidade.

¹Programa de Pós-graduação em Estatística e Experimentação Agropecuária, Lavras - {noe.macario1, elias.sabe1}@estudante.ufla.br

²Departamento de Estatística, Universidade Federal de Lavras, Lavras - {paulo.guimaraes, luiz.nakamura}@ufla.br

³Departamento de Informática e Estatística, Universidade Federal de Santa Catarina, Florianópolis - andreack@gmail.com



Inferência bayesiana em modelos não lineares: Ajuste do modelo logístico em dose-resposta de herbicidas

Edilene Cristina Pedroso Azarias¹

Rafaela de Carvalho Salvador²

Natiele Almeida Gonzaga³

Edilson Marcelino Silva⁴

Joel Augusto Muniz⁵

Resumo

O uso de herbicidas é uma das estratégias mais adotadas no controle de plantas daninhas, pelo menor custo e pela possibilidade de aplicação em grandes áreas. A relação entre a dose do herbicida e a resposta das plantas segue um padrão sigmoidal, podendo ser modelada por modelos não lineares. Este estudo teve como objetivo ajustar um modelo não linear logístico a dados de dose-resposta de herbicidas por meio da abordagem bayesiana, comparando verossimilhanças com distribuições normal, Skew Student-t e Student-t. Foram utilizados dados do herbicida S-metolachlor em duas espécies de plantas daninhas do gênero *Amaranthus*, com doses de 90, 180, 360 e 720 $g\ ha^{-1}$. O controle foi avaliado 28 dias após a aplicação do herbicida por meio de uma escala visual, onde 0 representa ausência de injúria visível e 100 indica controle total das plantas. A análise foi realizada no software R. Foram utilizados o critério de informação da deviance (DIC) e o critério da densidade preditiva ordenada (CPO). Não houve evidências de não convergência das cadeias. As estimativas para os parâmetros do modelo, considerando todas as distribuições, foram próximas. A distribuição Normal apresentou o melhor desempenho preditivo, com maior CPO, sugerindo uma boa capacidade de previsão. Já a distribuição Student-t apresentou menor DIC. No geral, a distribuição Normal foi considerada a mais adequada, seguida pela distribuição Student-t. Esta última, por ser mais flexível, é capaz de lidar melhor com caudas mais pesadas ou outliers, tornando-se uma alternativa, aos dados que apresentam características de assimetria ou valores extremos.

Palavras-chave: regressão; Skew Student-t; Student-t; *Amaranthus*; controle de planta daninha.

¹ Departamento de Estatística, UFLA – edilene.azarias1@estudante.ufla.br

² Departamento de Estatística, UFLA – rafaela.salvador3@estudante.ufla.br

³ Departamento de Estatística, UFLA – natiele.gonzaga1@estudante.ufla.br

⁴ Departamento de Matemática, UFRRJ - edilsonest@ufrj.br

⁵ Departamento de Estatística, UFLA – joamuniz@ufla.br



Modelagem não linear para a atividade de água no armazenamento de açúcar mascavo

Giulia Eduarda Bento ¹; Gilberto Rodrigues Liska ²; Yara Karolina Moura de Souza³

Carolina da Silva Perez ⁴; Gustavo Verruma Bernardi ⁵; Laura Daniele Sanches Leal ⁶

Paula Porrelli Moreira da Silva ⁷; Josiane Rodrigues ⁸; Marta Regina Verruma-Bernardi ⁹

Resumo

A atividade de água (Aw) é um parâmetro importante para a qualidade e estabilidade do açúcar mascavo, um produto de alto valor agregado no setor sucroalcooleiro e que vem se popularizando, devido ao seu menor refinamento e processamento, pois preserva nutrientes no produto, em relação aos outros tipos de açúcar existente no mercado. Durante o armazenamento em diferentes umidades e tempos, o açúcar mascavo sofre alterações em variáveis físicas e químicas, como a atividade de água, que com valores acima de 0,60, podem indicar reações enzimáticas, microbiologias e de deterioração do produto, que alteram a qualidade, estabilidade e o sensorial do produto. Dessa forma, os modelos não lineares foram ajustados aos dados, para encontrar qual melhor descreve o comportamento dessa variável, determinando pontos importantes, como de inflexão, aceleração máxima e desaceleração máxima, que fornecem interpretações práticas que são úteis para a logística desse produto. Todas as análises foram realizadas utilizando a linguagem de programação R, e o modelo que melhor se ajusta aos dados foi determinado por critérios de qualidade de ajuste, como Erro Médio Estimado, Erro Médio Relativo, Critério de informação de Akaike, Critério de Informação Bayesiano, Coeficiente de determinação (R^2) e Coeficiente de determinação R^2 ajustado. Assim, a estatística foi utilizada para transformar dados para fornecer interpretações práticas para melhorar a qualidade da logística do açúcar mascavo.

Palavras-chave: Modelos não lineares; Açúcar mascavo; Atividade de água; Programa R; Alimentos.

¹ Estudante de Biotecnologia da UFSCar - Araras - giuliabento@estudante.ufscar.br

² Docente do DTaSeR, CCA, UFSCar - gilbertoliska@ufscar.br

³ Mestre em Agroecologia e Desenvolvimento Rural, UFSCar, Araras - yara.souza@estudante.ufscar.br

⁴ Mestre em Agroecologia e Desenvolvimento Rural, UFSCar, Araras - clsilva@estudante.ufscar.br

⁵ Pesquisadora do LAN, Esalq/USP, Piracicaba - pporrelli@alumni.usp.br

⁶ Estudante de Engenharia Agronômica da Esalq/USP, Piracicaba - gustavobernardi@usp.br

⁷ Estudante de Engenharia Agronômica da Esalq/USP, Piracicaba - laurasanches@usp.br

⁸ Docente DTaSeR, CCA, UFSCar - josirodrigues@ufscar.br

⁹ Docente DTaSeR, CCA, UFSCar - verruma@ufscar.br



Índice para diagnóstico de superdispersão em dados longitudinais nominais politômicos com estrutura agrupada

Maria Letícia Salvador;¹ Gabriel Rodrigues Palma;² Rafael de Andrade Moral;³
Idemauro Antonio Rodrigues de Lara;⁴

Resumo

Experimentos em Ciências Agrícolas frequentemente envolvem análise de variáveis politômicas nominais em contextos longitudinais, tanto em estruturas individuais quanto agrupadas. Modelos de efeitos marginais e mistos são duas abordagens comuns. Uma premissa da análise de dados politômicos é que a variância observada está próxima da variância esperada pelo modelo. No entanto, pode haver casos em que a variância observada claramente excede a esperada, caracterizando a superdispersão, cuja identificação é crucial para escolher o modelo apropriado. Uma análise exploratória inicial se faz importante e pode ajudar tanto na seleção do modelo quanto na identificação de fontes de variabilidade. Neste contexto, um índice de dispersão multinomial longitudinal é proposto como uma ferramenta descritiva e diagnóstica. Este índice é calculado como a razão entre a variância observada e a esperada pelo modelo. O desempenho deste índice foi avaliado por meio de um estudo de simulação, empregando técnicas estatísticas para avaliar seu desempenho em diferentes cenários. Por meio do estudo de simulação, identificou-se que quanto mais o índice se aproxima de um, mais provável é a ocorrência de um alto grau de superdispersão. Por outro lado, valores mais próximos de zero indicam um baixo grau de superdispersão. Como estudo de caso, foi apresentada uma aplicação em ciência animal, na qual avaliou o comportamento de suínos agrupados em baías e submetidos a duas diferentes condições de enriquecimento ambiental.

Palavras-chave: distribuição multinomial; modelo dos logitos generalizados mistos; máxima verossimilhança penalizada.

¹Departamento de Ciências Exatas (LCE) - ESALQ (USP), Piracicaba/SP - marialesalvador@gmail.com

²Instituto Hamilton, Universidade de Maynooth, Irlanda - Gabriel.Palma.2022@mumail.ie

³Instituto Hamilton, Universidade de Maynooth, Irlanda - rafael.deandrademoral@mu.ie

⁴Departamento de Ciências Exatas (LCE) - ESALQ (USP), Piracicaba/SP - idemauro@usp.br



Modelagem não linear da variação do parâmetro a^* na coloração do açúcar mascavo durante seu armazenamento.

Giulia Eduarda Bento ¹; Gilberto Rodrigues Liska ²; Yara Karolina Moura de Souza³; Carolina da Silva Perez ⁴; Paula Porrelli Moreira da Silva ⁵; Gustavo Verruma Bernardi⁶ Laura Daniele Sanches Leal ⁷; Josiane Rodrigues ⁸; Marta Regina Verruma-Bernardi ⁹

Resumo

A coloração do açúcar mascavo é descrita pelo sistema de cores CIELAB, com parâmetros como luminosidade (L), eixo verde-vermelho (a^*), eixo azul-amarelo (b^*), croma (c^*) e ângulo (hue). O parâmetro a^* indica variações na tonalidade verde-vermelho, onde valores negativos correspondem a uma coloração esverdeada, enquanto valores positivos tons avermelhados. Esse último está associado à aceitação do produto pelos consumidores e fornece informações relevantes sobre sua qualidade, principalmente quando analisado em conjunto com variáveis físicas-químicas, que influenciam aspectos como sabor, umidade e processos de degradação. Para aprimorar a interpretação dos dados dessa variável, modelos de regressão não linear foram ajustados aos dados, para encontrar qual melhor descreve o comportamento de a^* . Por exemplo, para a umidade 2,44% o modelo que melhor se ajustou aos dados foi o Logístico, segundo os critérios de qualidade de ajuste Erro Médio Estimado, Erro Médio Relativo, Critério de informação de Akaike, Critério de Informação Bayesiano, Coeficiente de determinação (R^2) e Coeficiente de determinação R^2 ajustado. Além disso, também foi possível identificar pontos relevantes, como de inflexão (106,30; 5,62), aceleração máxima (-32,48; 2,37) e desaceleração máxima (305,09; 3,36), que fornecem interpretações práticas que são úteis para a logística desse produto. Todas as análises foram realizadas utilizando a linguagem de programação R. Dessa forma, a estatística foi utilizada como ferramenta para otimizar a logística e qualidade desse produto.

Palavras-chave: Modelos não lineares; Açúcar mascavo; Coloração a^* ; Programa R; Alimentos.

¹ Estudante de Biotecnologia da UFSCar - Araras - giuliabento@estudante.ufscar.br

² Docente do DTaSeR, CCA, UFSCar - gilbertoliska@ufscar.br

³ Mestre em Agroecologia e Desenvolvimento Rural, UFSCar, Araras - yara.souza@estudante.ufscar.br

⁴ Mestre em Agroecologia e Desenvolvimento Rural, UFSCar, Araras - clsilva@estudante.ufscar.br

⁵ Pesquisadora do LAN, Esalq/USP, Piracicaba - pporrelli@alumni.usp.br

⁶ Estudante de Engenharia Agronômica da Esalq/USP, Piracicaba - gustavobernardi@usp.br

⁷ Estudante de Engenharia Agronômica da Esalq/USP, Piracicaba - laurasanches@usp.br

⁸ Docente DTaSeR, CCA, UFSCar - josirodrigues@ufscar.br

⁹ Docente DTaSeR, CCA, UFSCar - verruma@ufscar.br



Modelagem não linear do Crescimento de Novilhas alimentadas com 10% de Proteína Bruta

Hellen Sonaly Silva Alves;¹ Guilherme Rocha Moreira² Felipe Antunes Magalhães³
Cristiane Rocha Albuberque⁴ Patricia Arruda de Moura⁵ Gabriela Cristina Gomes Silva
⁶ José Reinaldo Santana Junior⁷

Resumo

Modelos não lineares se destacam por representarem com precisão curvas de crescimento, sendo a escolha do modelo mais adequado essencial para descrever corretamente o desenvolvimento dos animais ao longo do tempo. Desse modo, este estudo teve como objetivo comparar modelos não lineares aplicados às curvas de crescimento de novilhas mestiças das raças Europeu-Zebu. O experimento foi realizado na Universidade Federal de Uberlândia, utilizando seis novilhas confinadas e alimentadas com dieta contendo 10% de proteína bruta. Os pesos foram registrados semanalmente por 98 dias. Os modelos testados foram: Santos (2019), Logístico Bicompartimental, Santos (2023), Santos (2024) e Brody. Para verificar a adequação dos modelos, foram aplicados testes de normalidade, independência e homogeneidade das variâncias. Todos os modelos atenderam aos pressupostos ($p > 0,05$), com exceção do modelo de Brody, que apresentou violação na homogeneidade das variâncias ($p < 0,05$). A qualidade de ajuste dos modelos foi avaliada por meio do Critério de Informação de Akaike (AIC), Critério de Informação Bayesiano (BIC) e Coeficiente de Determinação Ajustado ($R^2_{ajustado}$). O modelo Santos (2019) apresentou o melhor desempenho, com os menores valores de AIC (63,94) e BIC (68,41), além do maior $R^2_{ajustado}$ (0,99). Os parâmetros estimados indicaram duas fases de crescimento, com peso final teórico de 605,91 kg, resultado da soma das assíntotas (491,15 kg e 114,76 kg). Essa abordagem auxilia no aprimoramento do manejo nutricional e do planejamento produtivo, além de melhorar a compreensão da resposta dos animais às dietas.

Palavras-chave: Desempenho animal; Modelagem não linear; Dietas de confinamento; Ruminantes.

¹Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, Recife - hellen.silva@ufrpe.br

²Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, Recife - guilherme.rochamoreira@ufrpe.br

³Universidade Federal de Uberlândia, Departamento de Medicina Veterinária e Zootecnia, Uberlândia - felipeam.vet@gmail.com

⁴Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, Recife - cristiane.alima@ufrpe.br

⁵Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, Recife - patricia.amoura@ufrpe.br

⁶Universidade Federal de Uberlândia, Departamento de Medicina Veterinária e Zootecnia, Uberlândia - gabriela.silva6@ufu.br

⁷Universidade Federal de Uberlândia, Departamento de Medicina Veterinária e Zootecnia, Uberlândia - jose.santana@ufu.br



Comparação de Modelos Não Lineares na Análise de Crescimento na Cunicultura

Rafael Genaro;^{*1} Carlos A. F. Silva;¹ Daniel F. Ferreira;¹ Tales J. Fernandes;¹ Sarah L. C. Meirelles

Resumo

Este trabalho teve como objetivo ajustar e comparar sete modelos não lineares clássicos para descrever o crescimento em peso corporal de coelhos da raça Nova Zelândia Branco. Foram considerados dados de 1.793 animais avaliados em dez idades distintas, do nascimento aos 90 dias de vida. Os modelos analisados foram: Logístico, Gompertz, Brody, Weibull, Richards, Dos Santos e Von Bertalanffy. Os ajustes foram realizados por mínimos quadrados não lineares, e a seleção do melhor modelo baseou-se no Critério de Informação de Akaike (AIC) e no coeficiente de determinação ajustado (R^2_{aj}). O modelo de Von Bertalanffy apresentou o menor AIC (96,22) e elevado $R^2_{aj} = 0,9374$, sendo considerado o mais adequado. A validade do ajuste foi confirmada por meio de testes aplicados aos resíduos: Shapiro-Wilk para normalidade ($p = 0,9246$), Breusch-Pagan para homocedasticidade ($p = 0,2937$) e Durbin-Watson para independência dos erros ($p = 0,6363$). Os resultados indicaram que o modelo selecionado atende às suposições inferenciais clássicas, com estimativas biologicamente interpretáveis. Dessa forma, o modelo de Von Bertalanffy mostrou-se especialmente adequado para representar o crescimento dos coelhos avaliados, contribuindo de maneira prática para decisões mais precisas em programas de manejo e melhoramento genético.

Palavras-chave: crescimento animal; cunicultura; modelagem estatística; ajuste de modelos; seleção de curvas

¹Programa de Pós-graduação em Estatística e Agropecuária, Universidade Federal de Lavras (UFLA), Lavras - {rafael.genaro, carlos.silva29 }@estudante.ufla.br,{ danielff, tales.jfernandes, sarah }@ufla.br



Influência da Temperatura no Desenvolvimento de Pupas de Crisopídeos: Uma Comparação entre Machos e Fêmeas

Isabely Salles Rodrigues;¹ Thales Rangel Ferreira² Osvaldo Moisés Assinde³ Taciana Villela Savian⁴

Resumo

Os crisopídeos são insetos pertencentes à família *Chrysopidae* e, devido seu potencial como agente de controle biológico, têm sido objeto de estudos, especialmente quanto à sua aplicabilidade em sistemas agrícolas. Neste estudo, avaliou-se a duração média da fase de pupa de quatro espécies: *Ceraeochrysa cincta*, *Ceraeochrysa cubana*, *Ceraeochrysa paraguaria* e *Chrysoperla externa*, sob sete temperaturas constantes, com a diferenciação entre machos e fêmeas. Para a análise, foi utilizado o modelo não-linear baseado na equação equação de Réamur, o qual utiliza como parâmetros a temperatura base (T_b) e a constante térmica (K), representando, respectivamente, o limiar térmico inferior de desenvolvimento ($^{\circ}C$) e a constante térmica em graus-dias. Os resultados das análises evidenciaram que a duração média da fase de pupa diminui com o aumento da temperatura, sendo este padrão consistente entre espécies e sexos. Verificou-se, por meio do teste da razão de verossimilhança entre o modelo ajustado com espécie e sexo como fatores e o modelo apenas com espécie, que não houve redução significativa na soma de quadrados de resíduos ao incluir o fator sexo no modelo. O resultado do coeficiente de determinação do modelo ($R^2 = 0,9856$), indica que aproximadamente 98,56% da variação observada na duração média da fase de pupa nas diferentes temperaturas é explicada pelo modelo não-linear de Réamur.

Palavras-chave: controle biológico; equação de réamur; regressão não-linear.

¹Estatística e Experimentação Agronômica USP-ESALQ, Piracicaba - isabely.rodrigues@usp.br

²Estatística e Experimentação Agronômica USP-ESALQ, Piracicaba - - thales.rangel@usp.br

³Estatística e Experimentação Agronômica USP-ESALQ, Piracicaba - - osvaldoassinde@usp.br

⁴Estatística e Experimentação Agronômica USP-ESALQ, Piracicaba - - tvsavian@usp.br



Temperaturas Extremas e Mortalidade no Rio de Janeiro: Uma Análise com Modelos de Regressão Condisional

Ingrid Thaynara de Oliveira Brito ¹

Ludmilla da Silva Viana Jacobson ²

Jony Arrais Pinto Junior ³

Resumo

Desde a Revolução Industrial, mudanças nos meios de produção elevaram a temperatura global, impactando o meio ambiente e a saúde humana (?). Diante disso, diversos estudos investigam a associação entre temperaturas extremas e mortalidade. ? analisaram esse efeito em mortes por causas específicas na Inglaterra e País de Gales de 1993 a 2006. ? investigaram como a variação diária da temperatura influencia a mortalidade em quatro cidades do estado da Virgínia. No Brasil, ? estudaram o impacto da temperatura sobre mortes por doenças respiratórias em idosos nas capitais de 2000 a 2017. Este trabalho visa comparar dois modelos amplamente utilizados em estudos epidemiológicos — a regressão logística condicional e a regressão de Poisson condicional — na análise do impacto das temperaturas extremas na mortalidade no estado do Rio de Janeiro, entre 2000 e 2019. Utiliza-se o delineamento *case-crossover*, que controla fatores sazonais e de tendência ao comparar períodos de risco e controle para os mesmos indivíduos.

Palavras-chave: Temperatura extrema; mortalidade; modelos de regressão condicional.

References

¹UFF - ingridthaynara@id.uff.br

²UFF - ludmillajacobson@id.uff.br

³UFF - jarrais@id.uff.br



Contribuições à Modelagem de Dados Duplamente Limitados: Uma Nova Proposta com Ênfase em Aplicações Hidrológicas

Luis Gonzaga Pinheiro Felix ¹

Frank Gomes-Silva ²

Josimar Mendes de Vasconcelos ³

Thiago A. N. De Andrade ⁴

Resumo

Distribuições de probabilidade são ferramentas valiosas para estatísticos e pesquisadores aplicados, dada sua robustez na modelagem de dados sob incerteza. O avanço na geração de dados em diferentes áreas tem estimulado o desenvolvimento contínuo de novas distribuições, mesmo com a existência de modelos já consolidados. Um foco recente tem sido a criação de distribuições para dados duplamente limitados, cuja modelagem é desafiadora devido às suas características específicas. Neste trabalho, exploramos essas fronteiras do conhecimento ao apresentar um novo modelo com quatro parâmetros, o qual demonstra elevada flexibilidade para dados restritos ao intervalo unitário. Nossa abordagem é baseada na beta-generalizada proposta no artigo seminal *Beta-normal distribution and its applications*, por Eugene, Lee e Famoye (2002). Especificamente, emprega-se o gerador beta-G utilizando como distribuição base a versão modificada da distribuição Kumaraswamy para dados unitários, proposta por Sagrillo, Guerra e Bayer (2021), o que dá origem a um novo modelo denominado beta-Kumaraswamy modificado (β -MK). O modelo proposto preserva o suporte no intervalo unitário e apresenta elevada flexibilidade, sendo capaz de acomodar dados assimétricos à direita, à esquerda, unimodais, entre outras configurações. Para avaliar o desempenho do modelo na modelagem de dados reais, foram considerados oito conjuntos de dados referentes ao volume de água útil em reservatórios de usinas hidrelétricas brasileiras. Para fins comparativos, foram incluídas distribuições concorrentes consagradas na literatura, com número de parâmetros variando entre dois e quatro. Os resultados obtidos por meio de medidas de bondade de ajuste, histogramas e gráficos Q-Q favorecem o modelo proposto nesta pesquisa.

Palavras-chave: Modelagem; Beta-generalizada; Distribuição; Beta-G; Kumaraswamy.

¹Departamento de Estatística e Informática, Universidade Federal Rural de Pernambuco - UFRPE, Recife, Brasil - luis.felix@ufrpe.br

²Departamento de Estatística e Informática, Universidade Federal Rural de Pernambuco - UFRPE, Recife, Brasil - franksinatrags@gmail.com

³Departamento de Estatística e Informática, Universidade Federal Rural de Pernambuco - UFRPE, Recife, Brasil - josimar.vasconcelos@ufrpe.br

⁴Departamento de Estatística, Universidade Federal de Santa Maria - UFSM, Santa Maria, Brasil - thiagoan.andrade@gmail.com



Métodos de Aprendizado de Máquina para Detecção de Quasares de Alto Redshift com Mitigação de Viés de Seleção

Bruno Marcondes e Resende;¹ Rafael Izicki²

Resumo

Redshift é uma medida astronômica que indica o deslocamento de uma onda eletromagnética para o espectro vermelho, sendo utilizado para medir a distância de objetos astronômicos, como galáxias e quasares. Quanto maior o redshift, mais distante e antigo é o objeto. Detectar quasares com alto redshift é crucial para entender a formação de buracos negros supermassivos no início do universo. Tradicionalmente, o redshift é medido por espectroscopia, um processo demorado e caro. A fotometria, que mede a intensidade da luz em diferentes faixas de comprimentos de onda, juntamente com o aprendizado de máquina (ML), é usada para estimar o redshift de maneira mais rápida. No entanto, há o desafio do viés de seleção, que resulta em bancos de treinamento que sub-representam quasares que são geralmente mais captados pela espectroscopia.

Este projeto tem como objetivo desenvolver métodos de aprendizado de máquina que mitigam esse viés, permitindo a detecção mais eficiente de quasares de alto redshift. Utilizando dados do levantamento fotométrico S-PLUS (Southern Photometric Local Universe Survey), que cobre 9300 graus quadrados do hemisfério sul celeste e utiliza um sistema fotométrico com 12 filtros passa-banda, o projeto busca não apenas melhorar a precisão das estimativas de redshift, mas também identificar candidatos a quasares com redshift muito alto. Isso aprofundará a compreensão do universo primordial e dos mecanismos que levaram à formação dos buracos negros supermassivos.

Palavras-chave: Machine Learning; Quasar; Redshift; Estatística; Viés Seleção; Probabilidade.

¹Departamento de Estatística, UFSCar (São Carlos) - bruno.resende@estudante.ufscar.br

²Departamento de Estatística, UFSCar (São Carlos) - rizicki@ufscar.br



Modelagem da taxa de mortalidade por intoxicações e lesões autoprovocadas intencionalmente por indivíduos de 15 a 29 anos no Brasil, 2014 a 2022

Nayara Maria Barbosa de Sousa¹

Guilherme Lopes de Oliveira²

Geraldo Magela da Cruz Pereira³

Resumo

O suicídio tem sido uma das principais causas de morte entre os jovens de 15 a 29 anos, segundo a Organização Mundial da Saúde. Estudos apontam que no Brasil ocorreram 112.230 mortes por suicídio de 2010 a 2019, com aumento de 43% dos casos em 2019 em relação a 2010. Devido à relevância deste problema, neste trabalho são apresentadas propostas de modelagem estatística através de modelos de regressão Poisson, quasi-Poisson e Binomial Negativo. Os ajustes foram feitos considerando como variável resposta o total de mortes por intoxicações e lesões autoprovocadas intencionalmente por indivíduos de 15 a 29 anos, no período de 2014 até 2022. Os fatores explicativos incluíram sexo, raça/cor da pele, incidência de transtornos mentais e AIDS, taxa de analfabetismo e de desocupação, além da região macroeconômica e o ano de ocorrência. Os modelos quasi-Poisson e Binomial Negativo apresentaram bons ajustes, com coeficientes estimados muitos próximos. As regiões Nordeste, Sudeste e Norte apresentam taxas de suicídio consideravelmente menores que a região Centro- Oeste e os homens têm três vezes mais chances de cometerem suicídio do que as mulheres.

Palavras-chave: modelo Binomial Negativo; regressão Poisson; suicídio.

¹ Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, PPGEE, UFLA – nayara.sousa@estudante.ufla.br

² Departamento de Computação, CEFET - MG – guilhermeoliveira@cefetmg.br

³ Departamento de Estatística, UFLA – geraldo.pereira@ufla.br



Funções de Afilamentos não-lineares por GAMLSS

Mauro Citro Lalucci;¹ Juliano Bortolini² João Paulo Sardo Mad³ Cyro Matheus Conetti Favalessa⁴ José Nilton da Cruz⁵

Resumo

O presente estudo teve como objetivo avaliar a eficácia do Modelo Aditivo Generalizado para Locação, Escala e Forma (GAMLSS) na modelagem do perfil do fuste de árvores de *Eucalyptus sp.*. Para descrever o perfil do fuste das árvores, foram testadas quatro funções não-lineares de afilamento: Kozak (2004), Max & Burkhart (1976), Clark III, Souter & Schlaegel (1991) e Demaerschalk (1972). Essas funções foram combinadas com cinco distribuições estatísticas (NO, LO, TF2, JSU e SST). Os preditores lineares foram definidos de modo que a média (μ) fosse a função de afilamento, enquanto a escala (σ) foi modelada como uma função linear simples, e os parâmetros de forma (ν e τ) foram mantidos constantes. A análise incluiu avaliação gráfica do ajuste dos modelos e comparação entre as combinações de funções de afilamento e distribuições, utilizando pesos para cada métrica de acurácia. Os resultados indicaram que a função de afilamento segmentada de Clark, Souter e Schlaegel (1991), combinada com as distribuições JSU e SST, apresentou os melhores desempenhos na estimativa dos diâmetros ao longo do fuste das árvores. Além disso, todas as combinações testadas demonstraram bom ajuste, validando a eficácia do GAMLSS não-linear para modelagem florestal.

Palavras-chave: DAP; Modelagem Florestal; GAIC

¹Departamento de Estatística da Universidade Federal de Lavras (UFLA), Lavras - mauro.citro@gmail.com

²Departamento de Estatística da Universidade Federal de Mato Grosso (UFMT), Cuiabá - juliano.bortolini@ufmt.br

³Departamento de Engenharia Florestal da Universidade Federal do Paraná (UFPR), Paraná - joaosardomadi@gmail.com

⁴Departamento de Engenharia Florestal da Universidade Federal de Mato Grosso (UFMT), Cuiabá - cyro.favalessa@ufmt.br

⁵Departamento de Estatística da Universidade Federal de Mato Grosso (UFMT), Cuiabá - niltonn.cruz@gmail.com



Previsão de Preços de Carros Usados com Kolmogorov–Arnold Networks: Uma Abordagem para Captura de Relações Não Lineares

Igor De Mello Sandini;¹ Arthur Henrique da Rocha Hintz;² Thiago A. N. De Andrade³

Resumo

Este trabalho explora a aplicação do modelo Kolmogorov–Arnold Network (KAN) para prever o preço de carros usados, destacando sua eficácia em contextos nos quais as relações entre variáveis são marcadamente não lineares. Diferente de modelos tradicionais, o KAN mostra-se eficiente para lidar com dados complexos, nos quais as interações entre variáveis não são facilmente capturadas por abordagens lineares convencionais.

Os dados utilizados neste estudo foram extraídos de anúncios de veículos e incluem variáveis categóricas e numéricas. A relação entre o preço e atributos como ano de fabricação, quilometragem e marca não segue um padrão linear simples, o que torna desafiadora a aplicação de métodos de regressão tradicionais. O modelo KAN, com sua arquitetura adaptável, conseguiu capturar essas interações de forma eficaz, resultando em previsões mais precisas em comparação com modelos concorrentes.

O processo envolveu normalização dos dados, codificação de variáveis categóricas e a criação de atributos derivados, como interações entre quilometragem e ano de fabricação. O desempenho dos modelos foi avaliado com métricas como o Erro Absoluto Médio (MAE) e a Raiz do Erro Quadrático Médio (RMSE), que evidenciaram a superioridade do KAN em relação ao XGBoost e ao Multilayer Perceptron (MLP).

Este estudo reforça o potencial das redes KAN na ciência de dados aplicada, especialmente em contextos voltados à sustentabilidade, nos quais previsões precisas de preços podem otimizar processos de comercialização e apoiar estratégias de consumo mais consciente.

Palavras-chave: Kolmogorov–Arnold Networks; previsão de preços; relações não lineares; ciência de dados aplicada; aprendizado de máquina

¹Departamento de Engenharia de Automação, Universidade Federal de Santa Maria, Santa Maria, Rio Grande do Sul, Brasil - igor.sandini@acad.ufsm.br

²Departamento de Estatística, Universidade Federal de Santa Maria, Santa Maria, Rio Grande do Sul, Brasil - arthur_hintz@acad.ufsm.br

³Departamento de Estatística, Universidade Federal de Santa Maria, Santa Maria, Rio Grande do Sul, Brasil - thiagoan.andrade@gmail.com



Classe Complementar Generalizada Seno Bibaseline

Natália Moraes Cordeiro ¹

Cícero Carlos Ramos de Brito ²

Kleber Napoleão Nunes de Oliveira Barros ³

João Silva Rocha ⁴

Wilson Rosa de Oliveira Júnior ⁵

Resumo

O presente trabalho tem como objetivo principal apresentar a Classe Complementar Generalizada Seno Bibaseline. Essa classe constitui uma generalização construída a partir da composição de funções, permitindo a geração de novas distribuições de probabilidade, bem como a formulação de classes e famílias de distribuições probabilísticas. A partir dessa proposta, foram desenvolvidas as expressões analíticas para a função de distribuição acumulada (FDA) e para a função densidade de probabilidade (FDP) associadas à classe. Além disso, são discutidas diversas propriedades de caracterização da classe, incluindo a função de risco, os momentos e momentos centrais de ordem m, a função geradora de momentos, a função característica e o coeficiente geral. Também são apresentadas as derivadas da função de log-verossimilhança, bem como a análise do suporte da distribuição. O estudo inclui ainda uma aplicação teórica da classe proposta, complementada por aplicações a dados simulados e a um conjunto de dados reais. As distribuições obtidas foram comparadas com modelos previamente existentes na literatura, evidenciando o potencial da classe proposta em termos de ajuste e flexibilidade.

Palavras-chave: Bibaseline; Complementar; Generalizada.

¹ Departamento de Estatística e Informática, UFRPE – Recife/PE – nataliamcorddeiro@gmail.com

² Instituto Federal de Pernambuco IFPE – Recife/PE – cicerocarlosbrito@yahoo.com.br

³ Departamento de Estatística e Informática, UFRPE – Recife/PE – kleber.barros@ufrpe.br

⁴ Departamento de Estatística e Informática, UFRPE – Recife/PE – joaosilvarocha@hotmail.com

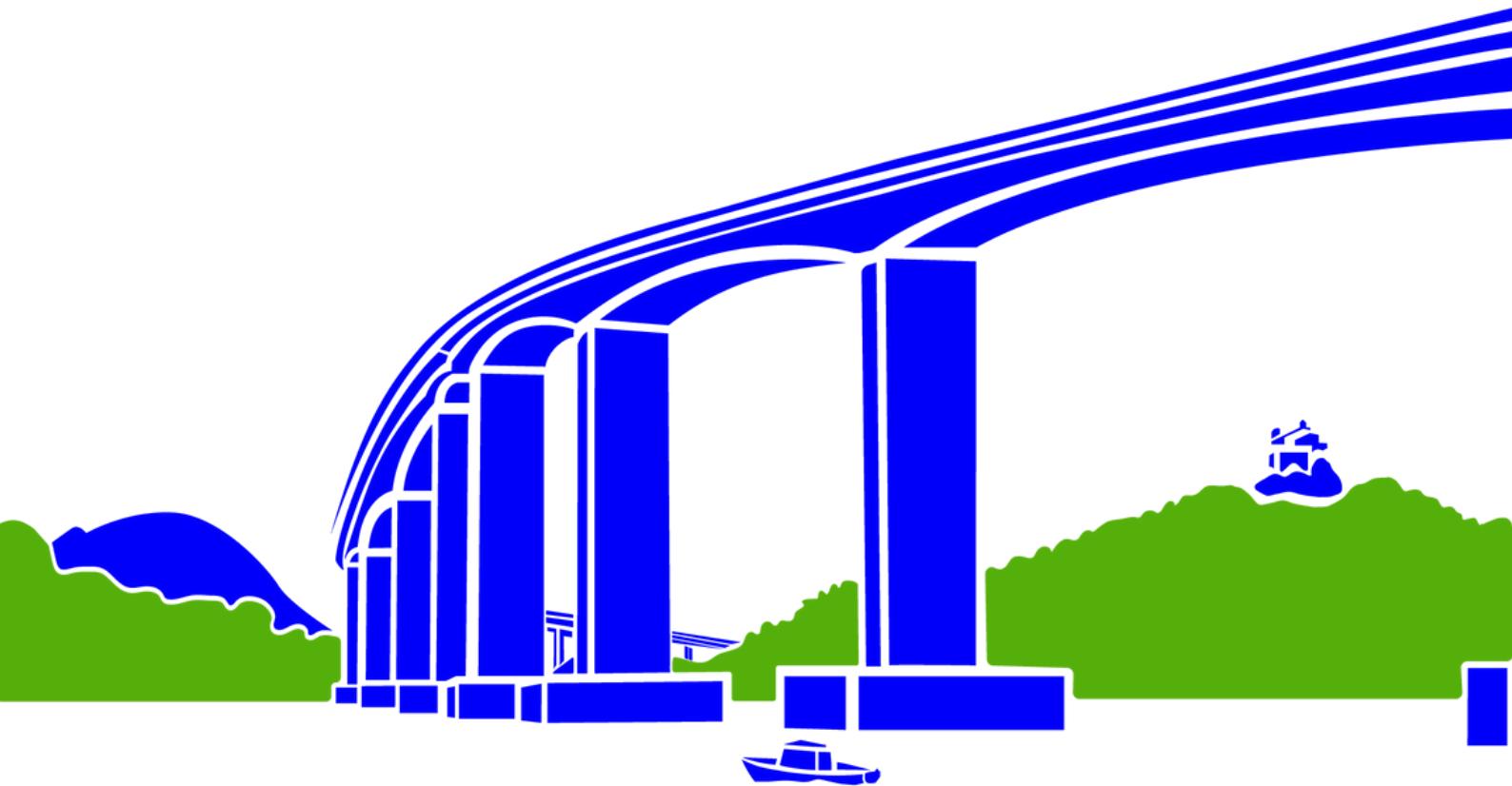
⁵ Departamento de Estatística e Informática, UFRPE – Recife/PE – wilson.oliveirajr@ufrpe.br

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Pôsters

Modelos de Regressão





Preditores da Taxa de Incidência de Malária em Tete, Moçambique: Uma Comparação de Modelos para Dados de Contagem com Superdispersão

Abrantes João Afonso Mussafo;¹ Thiago Moraes Rizzieri;² Fernando Andrade³ Victória Marchetti⁴ José Silvio Govone⁵ Liciana Vaz de Arruda Silveira⁶

Resumo

O estudo visa comparar os modelos de Poisson e binomial negativa nas incidências de malária na província de Tete, em função de variáveis climáticas e sociais nos 15 distritos da província de Tete, Moçambique, no período de 2016 a 2022. Os dados são de rotina, mensais, e foram coletados nesta província em 15 distritos e 155 unidades sanitárias, no período de 2016 a 2022. A comparação foi realizada por meio do ajuste de modelos lineares generalizados, utilizando o Critério de Informação de Akaike (AIC), o parâmetro de dispersão (razão entre a deviance residual e os graus de liberdade), bem como o diagnóstico do ajuste. Em relação aos resultados, o modelo de Poisson apresentou superdispersão igual a $\phi = 8.175219$ e uma razão entre a deviance residual e o grau de liberdade de 7.0603, ambos superiores a 1. Além disso, o AIC foi de 14060. Por outro lado, o modelo Binomial Negativo teve um ajuste satisfatório, com parâmetro de dispersão de $\phi = 1.15001$, uma razão entre a deviance residual e o grau de liberdade de 1.0539, AIC de 8766,9 e o diagnóstico adequado. Conclui-se que o modelo Binomial Negativo é o mais adequado em situações de superdispersão nos dados. Além disso, pode-se afirmar que a taxa de analfabetismo e a temperatura mínima média são preditores da taxa de incidência da malária.

Palavras-chave: Malária, incidência, Taxa de analfabetismo, Condições climáticas, Modelos de Regressão.

¹Departamento de Estatística e Biodiversidade, UNESP-Botucatu: a.mussafo@unesp.br

²Departamento de Estatística e Biodiversidade, UNESP-Botucatu: thiago.m.rizzieri@unesp.br

³Departamento de Estatística e Biodiversidade, UNESP-Botucatu: f.andrade@unesp.br

⁴Departamento de Estatística e Biodiversidade, UNESP-Botucatu: victoria.marchetti@unesp.br

⁵Departamento de Estatística, Matemática Aplicada e Computação, IGCE, e Centro de Estudos Ambientais /UNESP - Rio Claro js.govone@unesp.br

⁶Departamento de Estatística e Biodiversidade, UNESP-Botucatu :liciana.silveira@unesp.br



GAMLSS na predição volumétrica de plantações de eucalipto

Rafaela de Carvalho Salvador¹

Patrick Gustavo Machado de Souza²

Natiele Almeida Gonzaga

Edilene Cristina Pedroso Azarias

Tales Jesus Fernandes

Luiz Ricardo Nakamura

Resumo

A predição do crescimento de árvores de eucaliptos é fundamental para a gestão ambiental florestal. Diversos agentes podem interferir no crescimento das árvores como a temperatura, a precipitação e o local de plantio. Os modelos aditivos generalizados para locação, escala e forma (GAMLSS) são modelos de regressão semi-paramétricos que podem lidar com curvas de crescimento complexas, permitindo a utilização de distribuições realistas para os dados, modelando diretamente parâmetros como, por exemplo, média, variância, assimetria e curtose. O objetivo deste trabalho foi propor a utilização dos GAMLSS, recorrendo a procedimentos de validação cruzada, para realizar a predição do crescimento em volume de árvores de eucalipto ($m^3\text{árvore}^{-1}$), considerando como covariáveis: temperatura ($^{\circ}\text{C}$), precipitação (mm) e diâmetro à altura do peito (cm) em três locais distintos de plantio. Para a modelagem, foram testadas as seguintes distribuições: Box-Cox Cole e Green (BCCG), Box-Cox exponencial potência (BCPE) e Box Cox t (BCT); comparadas por meio das métricas MAE, RMSE, MAPE, MASE e sMAPE. As métricas indicaram a distribuição BCCG permite a predição e projeção precisa do crescimento volumétrico e da produtividade da floresta de eucalipto.

Palavras-chave: regressão distribucional; regressão semi-paramétrica; plantações; sivilcultura; validação cruzada.

¹ Departamento de Estatística, Universidade Federal de Lavras – rafaela.salvador3@estudante.ufla.br

² Departamento de Recursos Hídricos, Universidade Federal de Lavras.



Uncovering the key factors behind South American football team performance¹

Lucas F. Rosa; ² Pedro H.R. Cerqueira; ³ Luiz R. Nakamura; ⁴ D. Mikis Stasinopoulos ⁵

Abstract

Football is one of the most popular sports in the world and holds particular prominence in South America. In this study, we use generalised additive models for location, scale, and shape (GAMLSS) with the aim of identifying factors that directly affect the points ratio of teams at the end of the season in three South American leagues: the Argentine, Brazilian, and Chilean, over the years 2022 to 2024, comprising a total of 191 observations. Because of the nature of the response, beta and logit-normal distributions were considered. The GAMLSS based on the logit-normal distribution was chosen as the most appropriate model through a stepwise-based procedure. This process identified, among several match statistics considered as candidate variables, the following covariates for the two different regression structures: goals scored, goals conceded, clean sheets, accurate passes, tackles, saves, fouls, red cards, the league in question, number of assists, dribbles, yellow cards, and shots against. Residual analysis indicated that the fitted model provided a reasonable fit to the data, and the results can be used by coaches and analysts in their decision-making processes to improve team performance.

Keywords: games, GAMLSS, soccer, sport statistics, statistical modelling.

¹This study was funded by FAPEMIG

²Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, Universidade Federal de Lavras, Lavras - lucasferreira_77@hotmail.com.br

³Departamento de Estatística, Universidade Estadual de Londrina, Londrina - pcerqueira@uel.br

⁴Departamento de Estatística, Universidade Federal de Lavras, Lavras - luiz.nakamura@ufla.br

⁵School of Computing and Mathematical Sciences, University of Greenwich, London - dmh.stasinopoulos@gmail.com



Análise do peso vivo de alpacas utilizando uma abordagem GAMLSS

Momate Emate Ossifo ¹
Ali William Canaza Cayo ²
Viviane Costa Silva ³
Elias Manensa Sabe ⁴
Rafaela de Carvalho Salvador ⁵
Luiz Ricardo Nakamura ⁶
Daniel Furtado Ferreira ⁷

Resumo

Alpaca (*Vicugna pacos*) é um camelídeo Sul-Americano que se adaptou ao clima e à flora da região elevada dos Andes. A maioria delas é encontrada acima de 3.800 m acima do nível do mar, em países como Peru, Chile, Bolívia, Equador e Argentina. Elas adaptam-se bem às condições climáticas severas, têm baixo impacto ambiental e a sua produção de carne é uma das atividades econômicas e sociais mais importantes nas zonas altas Andinas. Ciente disso, neste trabalho objetivou-se analisar o peso vivo de alpacas, relacionando-o com algumas variáveis explicativas candidatas, a saber: perímetro torácico (cm), perímetro abdominal (cm), altura na cernelha (cm), altura da garupa (cm), comprimento do membro anterior (cm) e comprimento corpo (cm). Foram utilizadas informações de 117 alpacas machos da raça Huacaya, que fazem parte da coleção permanente da Estação Experimental de Quimsachata, Puno, Peru. Para a modelagem, foram utilizados os modelos aditivos generalizados para locação, escala e forma (GAMLSS), por conta de sua flexibilidade em explicar o comportamento da variável resposta. Três distribuições foram inicialmente ajustadas para a variável resposta: gama (GA), inversa gama generalizada (GIG) e normal inversa (IG), e para a seleção das covariáveis em todos os parâmetros de cada distribuição foi utilizada uma estratégia baseada no procedimento *stepwise*. O GAMLSS baseado na distribuição GA foi o mais adequado, pois apresentou o menor critério de informação de Akaike. Após uma análise dos resíduos quantílicos normalizados, verificou-se que o modelo GAMLSS, baseado na distribuição GA, é adequado para descrever o conjunto de dados.

Palavras-chave: Distribuição gamma; Regressão distribucional; Variabilidade.

¹ Departamento de Estatística, UFLA, Lavras, MG, Brazil – momateemate@gmail.com

² Departamento de Estatística, UFLA, Lavras, MG, Brazil – awcanaza@gmail.com

³ Departamento de Estatística, UFLA, Lavras, MG, Brazil – vivicosta19.vc@gmail.com

⁴ Departamento de Estatística, UFLA, Lavras, MG, Brazil – elias.sabe1@estudante.ufla.br

⁵ Departamento de Estatística, UFLA, Lavras, MG, Brazil – rafaela.salvador3@estudante.ufla.br

⁶ Departamento de Estatística, UFLA, Lavras, MG, Brazil – luiz.nakamura@ufla.br

⁷ Departamento de Estatística, UFLA, Lavras, MG, Brazil – danielff@ufla.br



Análise de influência local via busca progressiva no modelo de regressão heteroscedástico linear simétrico

Victor Meneses Navarro Fernandes ¹

Juvêncio Santos Nobre ²

Resumo

A adequabilidade dos modelos de regressão depende de determinadas suposições, as quais são raramente atendidas de forma exata e completa em situações práticas. Desse modo, é importante avaliar como pequenas perturbações do modelo ou das suposições podem influenciar nos seus resultados. Se essas perturbações causam mudanças substanciais nos resultados, isto pode ser indicativo da sensibilidade do modelo e pode evidenciar a sua inadequabilidade. Caso contrário, tem-se que o modelo é robusto consoante às perturbações introduzidas. Por exemplo, na presença de heteroscedasticidade e quando o modelo é incorretamente especificado, as inferências baseadas no modelo homoscedástico considerando a suposição de normalidade podem ser extremamente imprecisas. Uma forma comumente usada para identificar a violação das suposições supracitadas é através da análise de diagnóstico. Complementar a isto, diferentes procedimentos estatísticos robustos às suposições idealizadas foram propostos na literatura para limitar a influência destas observações e garantir a validade da inferência considerando um modelo parcialmente bem especificado. Nos casos heteroscedásticos, diferentes métodos robustos têm sido propostos, incluindo estimadores baseados em reponderação e aparagem. Em particular, a busca progressiva (forward search) é um procedimento iterativo baseado na aparagem adaptativa dos dados que permite detectar múltiplas observações influentes e identificar inadequabilidade do modelo resultante. Diante disto, o presente trabalho tem por objetivo conduzir uma revisão acerca de métodos de diagnóstico robustos na modelagem estatística, em particular, a análise de influência local em modelos de regressão lineares heteroscedásticos simétricos. Ademais, realiza-se uma aplicação em um conjunto de dados real.

Palavras-chave: Influência local. Modelos de regressão linear. Heteroscedasticidade. Robustez. Busca progressiva.

¹ Departamento de Estatística, Instituto de Matemática, Estatística e Computação Científica, Universidade Estadual de Campinas, Campinas, São Paulo, Brazil — victormnavarrof2@gmail.com

² Departamento de Estatística e Matemática Aplicada, Universidade Federal do Ceará, Fortaleza, Ceará, Brazil — juvencio@ufc.br



CUSTOS DE TRANSAÇÃO ECONÔMICO NOS PROCESSOS DE CONTRATAÇÕES PÚBLICAS INEFICIENTES: Uma Análise Comparativa de Modelos de Regressões.

Emerson Antonio Rocha Pazeto
Poliano Bastos da Cruz ¹

Resumo

Nos últimos anos, o setor público brasileiro tem passado por transformações relevantes sob o ponto de vista da transparência e compliance. Por isso, os gastos públicos têm desperado cada vez mais o interesse da sociedade. Com o fito de se compreender adequadamente como variáveis internas e externas aos contratos públicos afetam o resultado da ineficiência dos gastos públicos no âmbito do governo federal brasileiro, foram observados 06 (seis) modelos de regressões com diversos indicadores de ajuste e de performance. A partir da comparação, foram observadas as covariáveis que, concomitantemente, afetam o número de aditivos dos contratos públicos federais para os modelos: MQO Truncado, TOBIT, Poisson, Poisson Inflada de Zeros, Binomial Negativo e Binomial Negativo Inflado de Zeros. Uma Análise Fatorial por Componentes Principais foi realizada, a fim de compreender as correlações entre fatores socioeconômicos relevantes da sociedade brasileira e quais os possíveis impactos destes fatores na performance do poder público. Os fatores, as covariáveis e as interações entre a transformação logarítmica dos Valores Iniciais dos Contratos e as covariáveis (fatores + variáveis) do estudo trazem uma melhor interpretação dos resultados. Os resultados mostraram que os modelos propostos se ajustaram satisfatoriamente aos dados. Apesar de serem classes distintas de modelos, a combinação das covariáveis significantes em todos os modelos propostos, a um nível de significância de 5%, trazem uma interpretabilidade aderente à literatura sobre os impactos das variáveis independentes na variável resposta (Aditivos).

Palavras-chave: Aditivos; Fatores; Ineficiência; Modelos de Regressões; Processos de Contratações Públicas.

¹PPG Acadêmico - Fucape Business School, sediada em Vitória-ES - emebompaz@gmail.com



ESTUDOS DAS RELAÇÕES ENTRE PRODUTO INTERNO BRUTO *PER CAPITA* E OS SETORES DO VALOR ADICIONADO BRUTO DO ESTADO DE MATO GROSSO

Wender Gleysom Schlosser da Rosa¹ Eveliny Barroso da Silva²

Resumo

Este trabalho tem como objetivo examinar a dinâmica econômica do estado de Mato Grosso ao longo de duas décadas, explorando a relação entre o Produto Interno Bruto (PIB) *per capita* e os principais setores econômicos representados pelo Valor Adicionado Bruto (VAB). Utilizando a técnica de regressão linear múltipla, investiga-se como as variações nos diferentes setores impactam o crescimento econômico durante o período de 2002 a 2021. Os resultados mostram que os setores de agropecuária, indústria e serviços impactam positivamente o PIB *per capita*, com destaque para a agropecuária como setor mais influente. No entanto, o setor de administração apresenta efeito negativo, sugerindo que seus recursos possam ser utilizados de forma mais eficiente. O estudo ainda reforça a importância de investimentos diversificados aos demais setores para um desenvolvimento sustentável em Mato Grosso.

Palavras-chave: Economia Regional, Influência Econômica, Economia Sustentável, Setores Produtivos.

¹ Departamento de Estatística, ICET, UFMT/Campus Cuiabá – wender.rosa@sou.ufmt.br

² Departamento de Estatística, ICET, UFMT/Campus Cuiabá – eveliny.silva@ufmt.br



**69^a Reunião Anual da
RBras
21º SEAGRO**

Estatística e Ciência de Dados
a serviço do Desenvolvimento Sustentável

Vitória - ES

4 a 8 de agosto de 2025

MODELAGEM DA INSERÇÃO INFORMAL NO MERCADO DE TRABALHO: UMA ANÁLISE SOCIOECONÔMICA VIA REGRESSÃO LOGÍSTICA PARA DADOS DE PESQUISAS COMPLEXAS.

João Vitor Alemida Toste¹ Eveliny Barroso da Silva²

Resumo

Este estudo investiga os determinantes da informalidade no mercado de trabalho, focando na análise das probabilidades associadas a características individuais, ocupacionais e regionais. Utilizando uma abordagem quantitativa baseada em modelos de regressão logística, foram analisados dados de uma pesquisa amostral complexa representativa da força de trabalho. O estudo contribui para a literatura ao fornecer estimativas detalhadas e *insights* sobre a dinâmica da informalidade no mercado de trabalho brasileiro. Tais resultados são cruciais para orientar políticas públicas voltadas à promoção da formalização, redução da desigualdade e aumento da proteção social para os trabalhadores. Por fim, são analisados os principais determinantes que contribuem para a informalidade no território brasileiro.

Palavras-chave: Informalidade; mercado de trabalho; regressão logística.

Departamento de Estatística, ICET, UFMT/Campus Cuiabá – vitor.almeida.33.va@gmail.com

² Departamento de Estatística, ICET, UFMT/Campus Cuiabá – eveliny.silva@ufmt.br



Antedependence models for nonstationary longitudinal data with centered skew-t distribution in a Bayesian approach.

Carina Brunehilde P. da Silva;¹ José Roberto Silva dos Santos;² Larissa Ávila Matos¹

Resumo

Longitudinal data and repeated measures are widely observed in different and important fields of science, such as biological, medicine, or social data. However, the dependence that these observations exhibit hinders the utilization of classical analysis procedures. Extra challenges are added when data are nonstationary, skewed, or heavy-tailed. In this work, we propose to use a parsimonious and flexible class of models capable of addressing the mentioned issues: the antedependence models. Regarding nonstationarity, no restrictions in variances or same-lag correlations behavior along the time points are possible due to using the modified Cholesky decomposition for the covariance matrix, which allows a joint mean-covariance estimation based on an unconstrained parameterization. It also allows the researcher to account for different marginal skewness or kurtosis for different instants, if required by the data. Moreover, to handle departures from normality, we propose that the model's innovations follow the centered skew-t distribution, which presents some advantages, such as simplifying the interpretation of the parameters since we estimate the mean and variance directly, providing robust accommodation for non-normal data and enabling its representation in a stochastic form that facilitates computational implementation. Concerning the estimation procedures, a Bayesian approach is considered using the Hamiltonian Monte Carlo algorithm via Stan. Simulation studies were driven to analyze parameter recovery, convergence, and sensitivity with satisfactory results. Furthermore, the proposed model was employed to investigate a classical database in antedependence models literature that analyzes cattle weights, but now, we present more tools to accommodate outlier observations.

Palavras-chave: Antedependence; skewed and heavy-tailed data; Cholesky decomposition; HMC.

¹Universidade Estadual de Campinas, SP - c223716@dac.unicamp.br

²Universidade Federal do Ceará, CE - silvadosjr@dema.ufc.br

¹Universidade Estadual de Campinas, SP - larissam@unicamp.br



Correlação e regressão linear entre caracteres da cultivar

Trifecta de alfafa

Alissa Frigotto ¹

Alberto Cargnelutti Filho ²

Lais Cherobini ³

Mikael Brum dos Reis ⁴

Leonardo Cesar Pradebon ⁵

João Augusto Andretta ⁶

Debora Luiza Pozzebon ⁷

Manoela Conterato Della Pace ⁸

Resumo

A alfafa (*Medicago sativa* L.) é uma leguminosa que pode ser utilizada como pastagem, feno e silagem. Os objetivos desse trabalho foram verificar se há relações lineares entre a altura de planta (AP), número de nós (NN), massa de matéria fresca (MF) e massa de matéria seca (MS) de alfafa, cultivar Trifecta, e se há caracteres que possam ser utilizados para predição da MS. Em um ensaio de uniformidade, em Santa Maria-RS, os caracteres AP, NN, MF e MS, foram avaliados em 110 plantas no dia 23/01/2024 (1º corte) e em outras 110 plantas no dia 06/03/2024 (2º corte), realizados aos 57 e 100 dias após a semeadura, respectivamente. Foram realizadas análises de correlação e regressão linear individual (por corte) e conjunta. Constatou-se que há relações lineares positivas entre os caracteres AP, NN, MF e MS. Também, conclui-se que é possível predizer a MS em função da MF por meio da equação: $MS = 0,1168 + 0,2454 MF; R^2 = 0,8987$.

Palavras-chave: *Medicago sativa* L.; ensaio de uniformidade; relações lineares.

¹ Programa de Pós-Graduação em Agronomia - Universidade Federal de Santa Maria (UFSM) - alissafrigotto@gmail.com

² Departamento de Fitotecnia - UFSM - alberto.cargnelutti.filho@gmail.com

³ Programa de Pós-Graduação em Agronomia - UFSM - laischerobini@gmail.com

⁴ Programa de Pós-Graduação em Agronomia - UFSM - mikareis1999@gmail.com

⁵ Programa de Pós-Graduação em Agronomia - UFSM - leonardopradebon@gmail.com

⁶ Curso de Graduação em Agronomia - UFSM - aundregusto@gmail.com

⁷ Curso de Graduação em Agronomia - UFSM - deborapozzebon11@gmail.com

⁸ Curso de Graduação em Agronomia - UFSM - manoelaconterato@gmail.com

Agradecimentos ao CNPq (Processo 304878/2022-7), PIBIC/CNPq/UFSM, PROBIC/FAPERGS/UFSM e Capes (Finance Code 001), pelo auxílio financeiro e bolsas concedidas aos autores. Aos estudantes bolsistas e voluntários pelo auxílio na coleta de dados.



Correlação e regressão linear entre caracteres da cultivar

Crioula de alfafa

Debora Luiza Pozzebon ¹

Alberto Cargnelutti Filho ²

Lais Cherobini ³

Alissa Frigotto ⁴

Mikael Brum dos Reis ⁵

Leonardo Cesar Pradebon ⁶

João Augusto Andretta ⁷

Manoela Conterato Della Pace ⁸

Resumo

Os objetivos desse trabalho foram verificar se há relações lineares entre a altura de planta (AP), número de nós (NN), massa de matéria fresca (MF) e massa de matéria seca (MS) de alfafa (*Medicago sativa* L.), cultivar Crioula, e se há caracteres que possam ser utilizados para predição da MS. Em um ensaio de uniformidade, conduzido em Santa Maria - Rio Grande do Sul, foram avaliados os caracteres AP, NN, MF e MS, em 110 plantas no dia 23/01/2024 (1º corte) e em outras 110 plantas no dia 06/03/2024 (2º corte), realizados aos 57 e 100 dias após a semeadura, respectivamente. Foram realizadas análises de correlação e regressão linear individual (por corte) e conjunta. Há relações lineares positivas entre os caracteres AP, NN, MF e MS e é possível predizer a MS em função da MF por meio da equação: $MS = 0,1472 + 0,2545 MF; R^2 = 0,8857$.

Palavras-chave: *Medicago sativa* L.; ensaio de uniformidade; relações lineares.

¹ Curso de Graduação em Agronomia - Universidade Federal de Santa Maria (UFSM) - deborapozzebon11@gmail.com

² Departamento de Fitotecnia - UFSM - alberto.cargnelutti.filho@gmail.com

³ Programa de Pós-Graduação em Agronomia - UFSM - laischerobini@gmail.com

⁴ Programa de Pós-Graduação em Agronomia - UFSM - alissafrigotto@gmail.com

⁵ Programa de Pós-Graduação em Agronomia - UFSM - mikareis1999@gmail.com

⁶ Programa de Pós-Graduação em Agronomia - UFSM - leonardopradebon@gmail.com

⁷ Curso de Graduação em Agronomia - UFSM - aundregusto@gmail.com

⁸ Curso de Graduação em Agronomia - UFSM - manoelaconterato@gmail.com

Agradecimentos ao CNPq (Processo 304878/2022-7), PIBIC/CNPq/UFSM, PROBIC/FAPERGS/UFSM e Capes (Finance Code 001), pelo auxílio financeiro e bolsas concedidas aos autores. Aos estudantes bolsistas e voluntários pelo auxílio na coleta de dados.



GAMLSS na modelagem de decisões de crédito: uso de dados sintéticos para previsão de aprovações

Elias M. Sabe;^{*1} Rafael Genaro;¹ Gean Damaceno;¹ Momate Ossifo;¹ Noé Macário;¹
Luiz R. Nakamura;² Andréa C. Konrath ³

Resumo

A concessão de crédito é um desafio no sistema financeiro moderno, sendo essencial para o crescimento econômico e a inclusão financeira. A decisão de aprovar ou rejeitar um empréstimo envolve múltiplos fatores analisados para minimizar o risco de inadimplência. No entanto, devido a restrições de privacidade e sigilo bancário e regulamentações de proteção de dados, instituições financeiras raramente disponibilizam dados reais de solicitações de crédito. Este estudo utiliza dados sintéticos construídos com base em perfis reais de risco de crédito, balanceando a proporção entre casos aprovados e rejeitados com *Synthetic Minority Over-sampling Technique for Nominal and Continuous features* (SMOTENC). O conjunto possui 45 mil observações e 14 variáveis (8 numéricas e 6 categóricas), permitindo avaliar o desempenho de modelos na predição de concessão de empréstimos. A variável resposta é dicotômica, representando o estado de aprovação do empréstimo (aprovado; rejeitado). Para a modelagem, foram empregados os modelos aditivos generalizados para locação, escala e forma (GAMLSS) e testaram-se três distribuições (binomial, binomial-dupla e beta-binomial - BB), e a seleção do modelo final foi baseada no Critério de Informação de Akaike (AIC). O modelo baseado na distribuição BB apresentou o melhor ajuste, onde as seguintes variáveis foram consideradas nas estruturas de regressão: amortização anual (USD), juros (%), renda anual (USD), situação da casa, finalidade, histórico financeiro, valor de empréstimo (USD), experiência e a idade do cliente. A análise de resíduos confirmou a adequação do modelo.

Palavras-chave: dados sintéticos; empréstimo bancário; inadimplência; instituições financeiras.

¹Programa de Pós-graduação em Estatística e Agropecuária, Universidade Federal de Lavras, Lavras - {elias.sabe1, rafael.genaro, gean.damaceno, momate.ossifo, noe.macario1}@estudante.ufla.br

²Departamento de Estatística, Universidade Federal de Lavras, Lavras - luiz.nakamura@ufla.br

³Departamento de Informática e Estatística, Universidade Federal de Santa Catarina, Florianópolis - andreack@gmail.com



AVALIAÇÃO DE MODELOS NÃO LINEARES NA DESCRIÇÃO DO ACÚMULO DE NITROGÊNIO EM PLANTAS DE ALFACE

Osvaldo Moisés Assinde [\[1\]](#)

Thales Rangel Ferreira [\[2\]](#)

Isabelly Salles Rodrigues [\[2\]](#)

Taciana Villela Savian [\[3\]](#)

Resumo

O alface está entre as principais verduras folhosas e é consumida na forma de salada ou então refogada, poucas vezes, constituindo-se num alimento indicado nas refeições por fornecer nutrientes e vitaminas. O acúmulo de nutrientes, como Nitrogênio, nas plantas, ao longo do tempo, em geral, apresenta caráter sigmoidal, que pode ser descrito por modelos não lineares. O objetivo deste trabalho foi avaliar a utilização dos modelos Logístico e Gompertz no estudo do acúmulo de Nitrogênio, em mg/planta, em função da idade da planta, em dias, das cultivares de alface “Americanas” e “Butterhead”, cultivadas com e sem cobertura com tecido de polipropileno. Os dados foram obtidos de um experimento onde o conteúdo acúmulado de Nitrogênio foi anotado nas idades de 7, 14, 21, 28, 35 e 42 dias após os transplantes. Os modelos foram ajustados pelo método de mínimos quadrados utilizando o algoritmo de Gauss-Newton por meio do software R. A avaliação da qualidade do ajuste foi feita com base nos valores do coeficiente de determinação, do desvio padrão residual e do critério de informação de Akaike. O modelo Logístico apresentou menor AIC, maior R² e menor DPR, sendo, portanto, mais adequado para descrever o acúmulo de Nitrogênio em plantas de alface.

Palavras-chave: Modelo Logístico. Modelo Gompertz. Hortaliças.

[\[1\]](#) Estatística Experiência Agrônoma USP-ESALQ.Piracicaba – osvaldoassinde@usp.br

[\[2\]](#) Estatística Experiência Agrônoma USP-ESALQ.Piracicaba – thales.rangel@usp.br

[\[2\]](#) Estatística Experiência Agrônoma USP-ESALQ.Piracicaba – isabelly.rodrigues@usp.br

[\[3\]](#) Estatística Experiência Agrônoma USP-ESALQ.Piracicaba – tvsavian@usp.br



Modelagem estatística do estoque de carbono em florestas plantadas de eucaliptos

Bianca Tomazela;¹ Humberto Eufrade Junior;² Fábio Prataviera³

Resumo

A modelagem estatística do estoque de carbono florestal desempenha um papel fundamental no monitoramento e no manejo adequado dos recursos florestais, além de contribuir diretamente para a mitigação das mudanças climáticas, conforme proposto pelo Objetivo de Desenvolvimento Sustentável (ODS) 13 Ação Contra a Mudança Global do Clima. A estimativa precisa do estoque de carbono é essencial para monitorar as mudanças na quantidade de carbono armazenada em ecossistemas florestais e para avaliar os impactos de diferentes práticas silviculturais. Além disso, pode ajudar na identificação de áreas prioritárias para conservação e restauração florestal. Técnicas estatísticas para modelagem do estoque de carbono florestal fazem parte de uma abordagem amplamente utilizada na literatura científica e em estudos relacionados à floresta e mudanças climáticas. Os dados utilizados foram obtidos via análise bibliométrica. O período de busca foi de 2021 a 2023 em base de dados científica em repositório de acesso aberto, considerando as seguintes palavras-chave, “sequestro de carbono”, “plantação de eucalipto” e “Brasil”, artigos em inglês e português. Foram obtidos dados dos seguintes estados: Bahia, Distrito Federal, Espírito Santo, Minas Gerais, Mato Grosso do Sul, Paraná, Rio Grande do Sul e São Paulo, totalizando 32 publicações e 164 experimentos com 12 espécies de eucalipto em 7 diferentes tipos de solo. Neste estudo, empregamos o algoritmo de *Random Forest*, uma técnica de aprendizado de máquina baseada em árvores de decisão, com o objetivo de identificar as variáveis ambientais e silviculturais mais relevantes na predição do estoque de carbono em florestas de eucalipto. Essa abordagem é especialmente eficaz para lidar com dados não lineares, permitindo a avaliação da importância relativa de cada variável no modelo. Os resultados preliminares deste estudo revelam que as variáveis “latitude”, “longitude”, “precipitação”, “temperatura média anual”, “espécie” e “idade” emergem como os principais fatores de relevância na explicação do processo de sequestro de carbono.

Palavras-chave: Carbono; Desenvolvimento sustentável; Florestas de eucaliptos; Mudanças climáticas; Predição.

¹Departamento de Ciências Florestais, Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo ESALQ/USP, Piracicaba, São Paulo, Brasil - bitomazela@usp.br

²Departamento de Ciências Florestais, Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo ESALQ/USP, Piracicaba, São Paulo, Brasil h.eufrade@usp.br

³Departamento de Ciências Exatas, Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo ESALQ/USP, Piracicaba, São Paulo, Brasil - fabio_prataviera@usp.br



Comparação de modelos não lineares com erro assimétrico para o acúmulo de matéria seca do feijoeiro cultivar *Bolinha*

Kelly Pereira de Lima ¹ Felipe Augusto Fernandes ; ² Ricardo de Andrade Lira Rabelo; ³ Augusto Ramalho de Moraes. ⁴

Resumo

O estudo do crescimento do feijoeiro por meio do acúmulo de matéria seca ao longo do tempo é fundamental para o manejo adequado da cultura e para a identificação de fatores que afetam o desenvolvimento das plantas. A modelagem desse crescimento pode subsidiar a aplicação mais eficiente das práticas agrícolas. Entretanto, o feijão tem um crescimento bem atípico e assim pode gerar valores que algumas distribuições estatísticas simétricas não conseguem captar, como por exemplo a distribuição normal. Nesse contexto, distribuições assimétricas, como a normal assimétrica e a t assimétrica, podem ser alternativas viáveis para incorporar tais particularidades. Este trabalho teve como objetivo comparar os modelos não lineares Logístico e Gompertz utilizando diferentes distribuições de probabilidade para os erros — normal, normal assimétrica e t assimétrica — na descrição do acúmulo de matéria seca do feijoeiro cultivar “Bolinha”. O experimento foi conduzido na Universidade Federal de Lavras, durante a safra das águas de 2006/2007, em delineamento em blocos casualizados com três repetições. Utilizou-se um esquema fatorial 5 x 8, com cinco densidades de semeadura (75, 145, 215, 285 e 355 mil plantas por hectare) e oito épocas de avaliação (13 a 83 dias após a emergência). Os modelos Logístico e Gompertz mostraram-se adequados. O modelo Logístico, com erros normal assimétrico e t assimétrico, apresentou melhor desempenho, destacando-se, respectivamente, nas densidades de 145, 285 e 355 mil plantas por hectare e nas de 75 e 215 mil plantas por hectare.

Palavras-chave: Assimetria; Modelo de crescimento; Feijão.

¹Departamento de Computação (DC), UFPI - kelly.lima.88@gmail.com.br

²Departamento de Estatística, Unicamp - fernandesfelipe@ gmail.com

³Departamento de Computação (DC), UFPI - ricardoalr@ufpi.edu.br

⁴Departamento de Estatística, UFLA - armorais@dex.ufla.br



Ajuste do modelo logístico no acúmulo de matéria seca em frutos de café conilon: identificação de pontos críticos com intervalos de confiança via *bootstrap*

Sarah Martins Rezende;¹ Tales Jesus Fernandes²

Resumo

O café conilon (*Coffea canephora*) é uma cultura de grande importância agrícola, especialmente em regiões tropicais. Compreender o crescimento e o desenvolvimento dos frutos é essencial para otimizar práticas de manejo. Este estudo teve como objetivo descrever o acúmulo de matéria seca em frutos do clone P2 (Clone 11 da variedade ES8143 “Marilândia”), com base na identificação dos pontos críticos da curva de crescimento e da taxa de acúmulo, utilizando o modelo logístico em função dos dias após a florada. Os dados foram extraídos de Senra et al. (2022), e a estimativa dos parâmetros foi realizada via mínimos quadrados, utilizando o algoritmo de Gauss-Newton. A qualidade do ajuste foi avaliada por meio do coeficiente de determinação ($R^2 = 0,9871$), apresentando os resíduos com distribuição normal ($p = 0,454$), independentes (Durbin-Watson $p = 0,466$) e homocedásticos (Breusch-Pagan $p = 0,072$). Os parâmetros estimados foram $a = 25,2313$; $b = 110,4553$; e $k = 0,0166$. A partir da modelagem, identificaram-se os seguintes pontos críticos: Ponto de Aceleração Máxima (PAM) aos 31 dias após a florada, Ponto de Inflexão (PI) aos 110 dias, Ponto de Desaceleração Máxima (PDM) aos 189 dias e Ponto de Desaceleração Assintótica (PDA) aos 248 dias. Para cada um desses pontos, foram obtidos intervalos de confiança por meio de reamostragem *bootstrap*, permitindo avaliar a precisão das estimativas. O modelo logístico mostrou-se adequado para descrever o crescimento da massa seca dos frutos, contribuindo com informações relevantes para decisões agronômicas em sistemas produtivos de café conilon.

Palavras-chave: *Coffea canephora*; Desenvolvimento fenológico; Regressão não linear; Pontos críticos; *Bootstrap*.

¹Doutoranda em Estatística, UFLA - sarah.rezende5@estudante.ufla.br

²Departamento de Estatística, UFLA - tales.jfernandes@ufla.br



69º RBras
21º SEAGRO

Estatística e Ciência de Dados
a serviço do Desenvolvimento Sustentável

Vitória - ES

4 a 8 de agosto de 2025

MODELAGEM GAMLSS PARA A DESCRIÇÃO DE PESO VIVO DE ALPACAS DE RAÇA HUCAYA

Eliseu Castanheiro Alberto¹

Osvaldo Moisés Assinde²

Tales Jesus Fernandes¹

Adriele Aparecida Pereira³

Resumo

O presente estudo pressupõe a utilização de Modelos Aditivos Generalizados para Locação, Escala e Forma (GAMLSS) que são uma classe recente de modelos que auxiliam o pesquisador a entender o comportamento de fenômenos biológicos, tornando a variável resposta mais flexível. Neste estudo, os GAMLSS foram aplicados em um banco de dados referente a medidas biométricas de alpacas de raça huacaya onde a variável resposta corresponde ao Peso Vivo (PV) do animal. Neste âmbito, foi necessário estudo de algumas distribuições que se comportam melhor com esta variável resposta, onde se destacaram as distribuições tais como Gama (GA), Inversa Gaussiana (IG) e Lognormal Tipo 2 (LOGNO2). A seleção das variáveis para os modelos foi por stepwise. Os modelos GAMLSS resultantes modelaram, a média (μ) e a variância (σ). e posteriormente foi feita análise de resíduos. O objetivo de estudo foi modelar o peso vivo de alpacas de raça huacaya por meio de ajuste de GAMLSS, afim de conhecer as medidas que mais influenciam no peso vivo do animal. Atendendo todos os parâmetros com efeitos das variáveis independentes (variáveis biométricas) e sob a modelagem da classe GAMLSS, o modelo da distribuição da família IG foi ao que melhor se ajustou aos dados por ter apresentado menor valor de AIC.

Palavras-chave: Distribuição, Dados biométricos, Gaussiana Inversa

Agradecimento: Os autores agradecem à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES pelo apoio financeiro com a bolsa de estudos.

¹ Departamento de Estatística – Universidade Federal de Lavras (UFLA) Caixa Postal 3037 – 37200-000 – Lavras, MG – Brasil – eliseu.alberto1@estudante.ufla.br; tales.jfernandes@ufla.br.

² Escola Superior de Agricultura 'Luiz de Queiroz'-Universidade de São Paulo – assindenano@gmail.com

³ Instituto de Ciências Sociais Aplicada, Universidade Federal de Alfenas; adrieleapvga@gmail.com



Análise de regressão para identificar fatores altistas e baixistas no preço do suíno em Mato Grosso

Thiago Duarte da Cruz ¹

Mariano Martínez Espinosa ²

Resumo

A suinocultura é uma atividade de grande relevância para a economia de Mato Grosso, com crescimento contínuo nos últimos anos. Este estudo teve como objetivo analisar, por meio de regressão linear múltipla, os fatores que influenciam a volatilidade do preço do quilo do suíno vivo na região. Foram consideradas variáveis como preços bovinos, insumos alimentares e fatores de oferta e demanda. A pesquisa é de natureza quantitativa e utilizou dados do Instituto Mato-grossense de Economia Agropecuária (Imea), Instituto de Defesa Agropecuário de Mato Grosso (Indea), do Instituto Brasileiro de Geografia e Estatística (IBGE) e da Secretaria de Comércio Exterior (Secex), abrangendo o período de janeiro de 2015 a dezembro de 2023. O modelo foi ajustado pelo método dos mínimos quadrados e submetido à análise residual para validação dos pressupostos, incluindo gráficos dos resíduos, testes de Cook-Weisberg e Shapiro-Wilk, além da verificação de colinearidade entre as variáveis independentes. Os resultados indicaram que o aumento no número de abates de fêmeas (x_2) contribui para a redução no preço do suíno, enquanto as variáveis exportação de carne suína (x_3), preço do farelo de soja (x_7) e produção de carne suína (x_8) apresentaram efeitos altistas. Conclui-se que a modelagem proposta fornece informações relevantes para produtores, gestores e agentes do setor, oferecendo uma base estatística robusta para decisões relacionadas ao mercado suinícola mato-grossense.

Palavras-chave: Regressão linear múltipla; mínimos quadrados; preço do suíno.

¹ Universidade Federal de Lavras, UFLA, Brasil – thiagoduarte20151@outlook.com

² Universidade Federal de Mato Grosso, UFMT, Brasil – marianomphd@gmail.com



Incorporating gmrf-based distributional models for analyzing violence against women in Minas Gerais

Wélon Antônio de Oliveira ¹ Márcia Faria Barros ² Yohana Taise Hoffmann ³ Lucas Ferreira Rosa ¹ Alex Monito Nhancololo ⁴ Luiz Ricardo Nakamura ⁵
João Domingos Scaloni ⁵

Resumo

A violência contra a mulher (VAW) é uma grave violação dos direitos humanos e representa um dos principais desafios de saúde pública. As consequências dessa violência são diversas, afetando condições físicas, saúde mental, sexual e emocional das vítimas, além de comprometer sua vida profissional e social. Entre os efeitos estão transtornos psicológicos, estímulos ao uso de substâncias, gravidez indesejada e maior vulnerabilidade a infecções sexualmente transmissíveis. A VAW é definida pela OMS como qualquer ato de violência de gênero que cause ou possa causar danos físicos, sexuais ou mentais às mulheres. No Brasil, a Lei Maria da Penha e a Política Nacional de Enfrentamento à Violência contra as Mulheres são os principais instrumentos legais para combater esse fenômeno. Apesar dos avanços legislativos, persistem desafios na implementação das políticas, especialmente em áreas com poucos recursos. Fatores como desigualdade econômica, escolaridade, status familiar e idade estão entre os principais determinantes da violência doméstica. Em Minas Gerais, a diversidade dos municípios demanda análises detalhadas para compreender os fatores espaciais e sociais associados à violência. Este estudo utiliza o modelo GAMLLSS com distribuição BCT e estrutura espacial GMRF para modelar a taxa de VAW nos municípios mineiros, considerando variáveis socioeconômicas e geográficas. As funções de suavização revelam relações não lineares entre VAW e variáveis como densidade populacional, taxa de matrículas no ensino médio e proporção da população economicamente ativa. Além disso, o efeito espacial indica maior incidência de violência nas regiões norte e noroeste do estado. Parâmetros como IDEB e média salarial estão associados à redução da violência. O modelo apresentou bom ajuste e resíduos adequados, validando sua eficácia na análise da VAW. A abordagem permite identificar áreas prioritárias e fatores determinantes, auxiliando no direcionamento de políticas públicas mais eficientes e regionais de enfrentamento à violência contra a mulher em Minas Gerais.

Palavras-chave: Violência contra a mulher, Modelagem espacial, GMRF, GAMLLS.

¹Programa de Pós-Graduação em Estatística e Experimentação Agropecuária – Instituto de Ciências Exatas e Tecnológicas (ICET) - Universidade Federal de Lavras (UFLA) – Lavras, MG - Brasil - welson.oliveira.estatistica@gmail.com.br

²Philips Medical Systems Ltda., Varginha, Minas Gerais, Brasil

³Prefeitura municipal de Lavras, Minas Gerais, Brasil

⁴Instituto de Matemática e Estatística da Universidade de São Paulo (IME-USP), São Paulo, Brasil

⁵Departamento de Estatística – Instituto de Ciências Exatas e Tecnológicas (ICET) - Universidade Federal de Lavras (UFLA) – Lavras, MG – Brasil - scalon@ufla.br



Impact of social indicators on infant mortality: An approach using distributional models

Welson Antônio de Oliveira ¹ Renato Santos Elias ² Márcia Faria Barros ³ Lucas Ferreira Rosa ¹ Alex Monito Nhancololo ⁴ Érica Suélen do Nascimento ⁵ Luiz Ricardo Nakamura ⁶
João Domingos Scaloni ⁶

Resumo

A taxa de mortalidade infantil (TMI), medida pela razão entre o número de óbitos de crianças menores de um ano e o número de nascidos vivos, permanece como um importante indicador da qualidade de vida e das condições socioeconômicas de uma população. No Brasil, fatores como acesso ao saneamento básico, renda média e escolaridade têm sido apontados como determinantes para a variação desse indicador entre os municípios. Nesse contexto, o presente estudo teve como objetivo analisar a influência de indicadores sociais e econômicos sobre a TMI nos municípios do estado de Minas Gerais, utilizando de uma abordagem estatística robusta baseada em modelos distribucionais. Os dados utilizados foram extraídos das plataformas *Cidades@* do (IBGE) e do Observatório do Trabalho de Minas Gerais, referentes ao ano de 2022. Foram considerados indicadores como proporção da população com ensino superior, taxa de urbanização, acesso à coleta de lixo, cobertura de esgotamento sanitário adequado, salário médio municipal, indicador de vulnerabilidade à pobreza e produto interno bruto do município. Para a modelagem estatística, empregou-se a classe dos Modelos Aditivos Generalizados para Locação, Escala e Forma, sendo selecionada a distribuição Box-Cox t (BCPE) com base no critério de Akaike (AIC). Os resultados indicaram que o aumento de uma unidade no salário médio está associado a uma redução mediana de 2% na TMI. De modo semelhante, um aumento de 1% na proporção da população com acesso a esgotamento sanitário implica em uma redução de 0,04% na mediana da TMI. Por outro lado, a vulnerabilidade à pobreza mostrou efeito contrário: a cada aumento de 1% na proporção de pessoas em situação de pobreza, observa-se um acréscimo mediano de aproximadamente 0,71% na TMI. A taxa de urbanização também exerceu influência positiva sobre a TMI, com tendência de aumento em municípios mais urbanizados. O modelo ajustado demonstrou boa adequação aos dados, conforme indicado pelos testes de normalidade, homoscedasticidade e dependência espacial.

Palavras-chave: Mortalidade infantil; Indicadores sociais; GAMLSS; Minas Gerais.

¹Programa de Pós-Graduação em Estatística e Experimentação Agropecuária – Instituto de Ciências Exatas e Tecnológicas (ICET) - Universidade Federal de Lavras (UFLA) – Lavras, MG – Brasil - welson.oliveira.estatistica@gmail.com.br

²Graduando em Administração Pública pelo Departamento de Administração Pública – Faculdade de Ciências Sociais Aplicadas- Universidade Federal de Lavras (UFLA) – Lavras, MG – Brasil

³Philips Medical Systems Ltda., Varginha, Minas Gerais, Brasil

⁴Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, Brasil.

⁵Departamento de Administração Pública – Faculdade de Ciências Sociais Aplicadas - Universidade Federal de Lavras (UFLA) – Lavras, MG – Brasil

⁶Departamento de Estatística – Instituto de Ciências Exatas e Tecnológicas (ICET) - Universidade Federal de Lavras (UFLA) – Lavras, MG – Brasil - scalon@ufla.br



Universalidades em Sistemas Urbanos: Uma Análise Física com GAMLSS

Hudson Vinicios Tavares Mineiro ¹ Mauro Citro Lalucci; ² Fabiano Lemes Ribeiro ³ Luiz Ricardo Nakamura ⁴

Resumo

A crescente necessidade de uma abordagem quantitativa e sistemática para a gestão urbana impulsiona a busca por teorias que expliquem os fenômenos das cidades. Estudos preliminares indicam que os sistemas urbanos exibem padrões de escalonamento universal em relação a necessidades individuais, variáveis socioeconômicas e de infraestrutura. Este estudo, portanto, buscou identificar essas universalidades, testando-as com diversas métricas urbanas de cidades ao redor do mundo. Para garantir a precisão da análise, foi empregado o modelo GAMLSS (Modelo Aditivo Generalizado para Locação, Escala e Forma), que permitiu ajustar a heterogeneidade dos dados e modelar a distribuição da variável dependente de forma adequada. A análise dos coeficientes (betas) do modelo, com interpretação física baseada na literatura, revelou padrões consistentes em cidades de países com mais de 200 cidades, corroborando a existência de universalidades nos sistemas urbanos.

Palavras-chave: Urbanização; Modelagem estatística; Infraestrutura urbana; *splines*

¹Departamento de Física da Universidade Federal de Lavras (UFLA), Lavras - hudson.mineiro@estudante.ufla.br

²Departamento de Estatística da Universidade Federal de Lavras (UFLA), Lavras - mauro.citro@gmail.com

³Departamento de Física da Universidade Federal de Lavras (UFLA), Lavras - fribreiro@ufla.br

⁴Departamento de Estatística da Universidade Federal de Lavras (UFLA), Lavras - luiz.nakamura@ufla.br



Análise de tendência temporal via Regressão Segmentada Joinpoint da taxa de incidência da malária nos estados da Amazônia Legal

Carla Dominique Silva Vasconcelos ¹, Karina Silva Homobono ², Pedro dos Santos Costa Junior ³, Vitor Barbosa Louzada ⁴, Camila Rodrigues Barbosa Nemer ⁵, Inana Fauro de Araujo ⁶, Neylan Leal Dias ⁷, Simone de Almeida Delphim Leal ⁸
Edcarlos Vasconcelos da Silva ⁹

Resumo

A malária é uma doença infecciosa febril aguda, causada por protozoários do gênero *Plasmodium*, transmitidos pela picada de fêmeas de mosquitos *Anopheles*. É uma das doenças parasitárias mais importantes do mundo, com grande impacto em regiões tropicais e subtropicais, especialmente na África Subsaariana, Amazônia e Sudeste Asiático. Os casos graves ocorrem principalmente nas infecções por *P. falciparum* e podem evoluir rapidamente para óbito se não tratados. Nos últimos 5 anos, a região da Amazônia Legal acumulou 690 mil casos, dos quais 15,5% foram relacionados a espécie de parasita *P. falciparum*, forma grave da doença. Nesta direção, o objetivo do estudo foi o de analisar a evolução temporal da taxa de incidência da malária nos estados da Amazônia Legal. A metodologia empregada consistiu no uso de regressão segmentada Joinpoint para avaliar a tendência da taxa de incidência podendo ser de crescimento, decrescimento ou constante. O banco de dados é de natureza secundária, disponibilizado gratuitamente pelo departamento de informática do Ministério da Saúde, TABNET DATASUS. O estudo está em fase de desenvolvimento e é conduzido pelo Grupo de Pesquisa em Educação, Saúde e Ciências da Vida – MQBio, da Universidade Federal do Amapá, primeiros resultados serão apresentados no 69 RBrás. Como conclusão, verifica-se a relevante contribuição do estudo para suporte a políticas de mitigação dos casos de malária na região amazônica, priorizando os estados com casos mais crônico e epidêmicos desta região.

Palavras-chave: Amazônia 1; Tendência Temporal 2; Malária 3; Joinpoint 4.

¹ Departamento de Pós-Graduação, UNESP – carladominique4@gmail.com

² Departamento de Pós-Graduação, UNESP – karinahomobono@gmail.com

³ Departamento de Ciências Exatas , UNIFATECIE – pedrooss028@gmail.com

⁴ Departamento de Ciências Biológicas e da Saúde, UNIFAP – vitoinho.unifap@gmail.com

⁵ Departamento de Ciências Biológicas e da Saúde, UNIFAP – camila.barbosa@unifap.br

⁶ Departamento de Ciências Biológicas e da Saúde, UNIFAP – inana@unifap.br

⁷ Departamento de Ciências Exatas e Tecnológicas, UNIFAP – neylanmatematico@gmail.com

⁸ Departamento de Ciências Exatas e Tecnológicas, UNIFAP – leal@unifap.br

⁹ Departamento de Ciências Exatas e Tecnológicas, UNIFAP – edcarlos.vasconcellos@unifap.br



Outlier Detection in the Circular Regression Model

Carlos André Pereira de Jesus Silva;¹ Anselmo Valdério Reisen;² Paulo Canas Rodrigues³

Abstract

In recent decades, several studies have been developed in the field of Statistics that focus on modeling the relationship between a circular covariate and a circular response variable, known as circular regression. Among these studies, methods for outlier detection in circular regression models have been proposed in the literature, given the impact of such observations on the estimation of model parameters. In this work, the circular mean error statistic is considered for outlier detection on the circular regression model. Through simulation studies, the cutoff points for the procedure are obtained, and its performance is investigated. An example of the method's application is presented using a real dataset of wind direction.

Keywords: Circular regression; outlier detection; wind direction.

¹Departamento de Estatística UFBA, Salvador -BA - cpibic@gmail.com

²Departamento de Estatística - UFES, Vitória - ES - valderioanselmoreisen@gmail.com

³Departamento de Estatística UFBA, Salvador -BA - paulocanas@gmail.com



Unified Beta Regression Model with Random Effects for the Analysis of Sensory Attributes

João César Reis Alves;¹
Gabriel Rodrigues Palma;² Idemauro Antonio Rodrigues de Lara¹

Resumo

Studies involving sensory analysis are essential for evaluating and measuring the characteristics of food and beverages, including consumer acceptance of samples. For various products, the experimental designs are generally incomplete block designs, with sensory attributes assessed using hedonic scales, ratings, or scores. Statistical methods such as generalized logits are commonly used to analyze these data but face limitations, including convergence issues due to superparameterization. Furthermore, sensory attributes are traditionally analyzed separately, increasing the complexity of the process and complicating the interpretation of results. This study proposes a unified beta regression model with random effects for simultaneously analyzing multiple sensory attributes, whose scores were converted to the (0,1) interval. Simulation studies demonstrated overall agreement rates greater than 82% for the unified model compared to models fitted separately for each attribute. As a motivational example, the unified model was applied to a real dataset in which eight grape juice formulations were evaluated by 98 potential consumers for each sensory attribute: color, flavor, aroma, acidity, and sweetness. The unified model identified the same top-rated formulations as the separately fitted models, characterized by a higher proportion of juice relative to sugar. The results underscore the ability of the unified model to simplify the analytical process without compromising accuracy, offering an efficient and insightful approach to sensory studies.

Palavras-chave: Simultaneous analysis; Product selection; Maximum likelihood; Agreement rates; Simulation.

¹Department of Exact Sciences, "Luiz de Queiroz" College of Agriculture, University of São Paulo - joaocesar@usp.br

²Hamilton Institute, Maynooth University- gabriel.palma.2022@mumail.ie.

¹Department of Exact Sciences, "Luiz de Queiroz" College of Agriculture, University of São Paulo - idemauro@usp.br.



Modelo de Regressão Quantílica Marshall-Olkin Cosseno Fréchet

Leonardo Prior Migliorini;¹ Fábio Souza;² Cleber Bisognin³

Resumo

A distribuição Fréchet foi introduzida no contexto da análise de valores extremos. Tal distribuição possui um único parâmetro e suporte nos reais positivos. É amplamente aplicada em dados de desastres naturais, como precipitações máximas anuais e poluição do ar, e na análise de risco na área de finanças. O presente trabalho propõe uma nova reparametrização da distribuição Fréchet por meio do gerador Marshall-Olkin Cosseno, denotada por $MOCF(\alpha, \beta)$, para $\alpha, \beta > 0$. Além disso, define-se um novo modelo de regressão quantílica baseado na distribuição MOCF. Para definir o novo modelo de regressão, reparametriza-se a função densidade de probabilidade da distribuição MOCF em termos do seu quantil μ . A estrutura de regressão é incluída através da relação $g(\mu_t) = \mathbf{x}_t^\top \boldsymbol{\beta}$, para $t = 1, \dots, n$, em que $g(\cdot) = \log(\cdot)$ é a função de ligação. Para estimação do vetor de parâmetros dos modelos propostos, utilizou-se o método de máxima verossimilhança (EMV). O desempenho do EMV foi avaliado via simulações de Monte Carlo, considerando múltiplos cenários, com 10000 réplicas de Monte Carlo e tamanhos amostrais $n \in \{50, 100, 200, 300\}$. As medidas utilizadas na avaliação do EMV foram a média, viés, erro padrão, erro quadrático médio (EQM), assimetria e curtose. Através da análise dos resultados das simulações, verificou-se que o estimador apresenta indícios de que são assintoticamente não viesados, consistentes, além de apresentar distribuição assintoticamente normal. Em estudos futuros, pretende-se definir medidas de diagnóstico para o modelo e aplicá-lo a dados reais, comparando os ajustes a outros modelos já consolidados na literatura.

Palavras-chave: Distribuição Fréchet; Regressão Quantílica; Máxima Verossimilhança; Quantil; Simulação de Monte Carlo.

Trabalho apoiado pelo programa PIBIC-CNPq.

¹Curso de Bacharelado em Estatística, UFSM - leonardo.migliorini@acad.ufsm.br

²Curso de Bacharelado em Estatística, UFSM - fabiosouza.608@gmail.com

³Departamento de Estatística, UFSM - cleber.bisognin@ufsm.br



Análise de regressão: uma previsão do ENADE 2025 para os cursos de licenciatura em matemática de Pernambuco, Brasil

Maria Marciele de Lima Silva ¹

Natália Moraes Cordeiro ²

Moacyr Cunha Filho ³

Resumo

O ensino superior tem como objetivo fundamental o desenvolvimento do potencial dos discentes, com ênfase nas competências e habilidades adquiridas ao longo de sua formação. Desde 2004, o Exame Nacional de Avaliação de Desempenho dos Estudantes (ENADE) tem sido utilizado como indicador da qualidade dos cursos superiores no Brasil. Para os cursos de licenciatura, a aplicação do exame, que antes ocorria a cada três anos, passará a ser anual a partir de 2024. Assim, utilizando os dados disponíveis no portal do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), aplicamos Regressão Linear Simples para analisar os resultados obtidos nos anos de 2008, 2011, 2014, 2017, 2021 e 2024. O objetivo foi prever os conceitos ENADE e o desenvolvimento dos discentes nas áreas de Formação Geral (FG) e Conhecimento Específico (CE) para a edição de 2025, nos cursos de Licenciatura em Matemática de Pernambuco que participaram de todas as edições anteriores do exame. Para cada Instituição de Ensino Superior (IES), foi ajustado um modelo de regressão linear, com a verificação da qualidade do ajuste. Complementarmente, os resultados foram padronizados em variáveis discretas, com variação de 0 a 5, conforme os conceitos ENADE. Com base nos resultados obtidos, espera-se que as instituições UFPE – Campus Recife, UFRPE – Campus Recife e UPE – Campus Garanhuns apresentem os melhores desempenhos nas três variáveis analisadas. Além disso, o ranking gerado com os conceitos ENADE previstos indica que IES públicas apresentam desempenho superior em comparação às instituições privadas no que tange ao aproveitamento educacional dos discentes.

Palavras-chave: ENADE; Regressão linear; Estimativas.

¹ Departamento de Estatística e Informática, UFRPE – Recife/PE – marciele.liima@gmail.com

² Departamento de Estatística e Informática, UFRPE – Recife/PE – nataliamcorddeiro@gmail.com

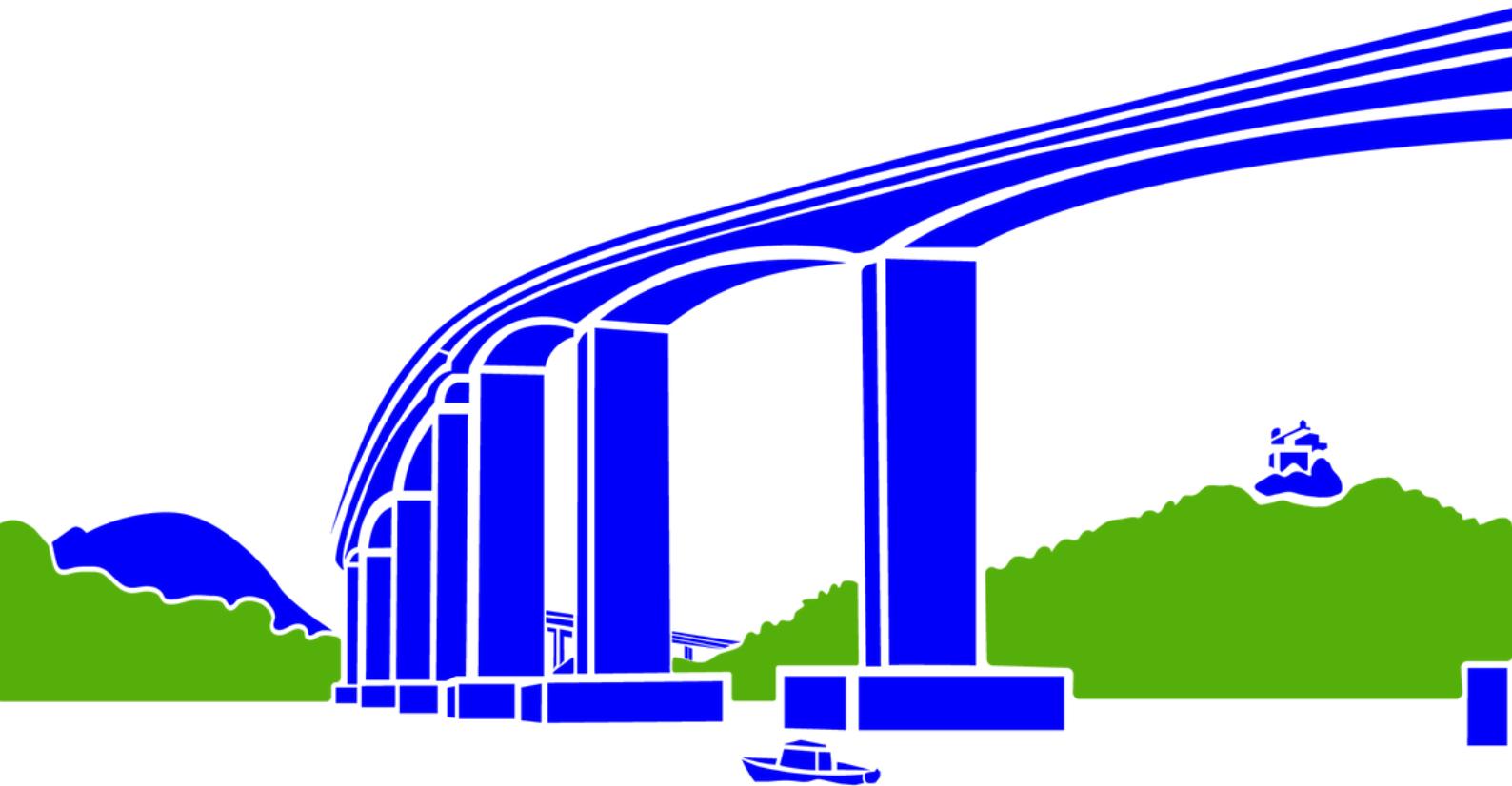
³ Departamento de Estatística e Informática, UFRPE – Recife/PE – moacyr2006@gmail.com

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Pôsters

Inferência Estatística





Estimadores em Forma-Fechada para a Distribuição de Probabilidade Nakagami-m Bivariada

Lucas P. Rezende ¹; Francisco Louzada Neto ²; Pedro Luiz Ramos ³.

Resumo

A estimação de parâmetros desempenha um papel fundamental na inferência estatística, e o Estimador de Máxima Verossimilhança (EMV) se destaca como uma das abordagens mais adotadas devido às suas propriedades desejáveis, como eficiência assintótica, consistência e invariância sob transformações. No entanto, a obtenção de EMVs em forma-fechada nem sempre é viável, especialmente quando se lida com distribuições mais complexas. Nessas situações, métodos numéricos como Newton-Raphson e Expectation-Maximization (EM), por exemplo, são frequentemente empregados, o que, por sua vez, pode resultar em altos custos computacionais, limitando sua aplicação em cenários que demandam processamento rápido e eficiente. Este estudo propõe uma abordagem generalizada para a obtenção de estimadores em forma-fechada, sob certas condições específicas. Em particular, desenvolvemos uma nova classe de estimadores para a distribuição Nakagami-Bivariada, baseada em um modelo Gama-Generalizado e sua conexão com o modelo Gama-Bivariado. Essa abordagem oferece uma alternativa eficiente e computacionalmente viável, ideal para problemas estatísticos que requerem estimativas rápidas e precisas.

Palavras-chave: Estimador; Forma-Fechada; EMV; Nakagami-Bivariada; Inferência Estatística.

¹Instituto De Ciências Matemáticas e de Computação, USP - lucasrezende@usp.br

²Instituto De Ciências Matemáticas e de Computação, USP - louzada@icmc.usp.br

³Departamento de Estadística, Pontificia Universidad Católica de Chile - pedro.ramos@mat.uc.cl



Análise da Taxa do Medo do Fracasso no Empreendedorismo na América Latina e Caribe: Tendências Regionais e Ajuste de Modelos de Distribuições Probabilísticas (2001-2021)

Carla Dominique Silva Vasconcelos ¹

Karina da Silva Homobono ²

Ricardo Puziol de Oliveira ³

Resumo

Este estudo analisa a Taxa do Medo do Fracasso no Empreendedorismo em países da América Latina e Caribe entre 2001 e 2021, utilizando a base de dados pública disponibilizada pela Global Entrepreneurship Monitor (GEM). O medo do fracasso é um fator determinante na decisão de empreender e reflete barreiras psicológicas e econômicas que influenciam o ambiente de negócios. O presente estudo adota uma abordagem estatística para modelar a variação dessa taxa ao longo do tempo, aplicando cinco distribuições probabilísticas: Weibull-Poisson, Weibull Modificada, Exponencial Weibull-Poisson, Gamma-Lomax e Weibull-Lomax. Os modelos foram avaliados por meio da estimativa de máxima verossimilhança e da razão de verossimilhança para determinar a melhor adequação aos dados. Os resultados indicam que as distribuições Weibull-Poisson e Exponencial Weibull-Poisson apresentam os melhores ajustes, demonstrando padrões de variação do medo do fracasso na região. A pesquisa destaca a importância da estabilidade econômica e políticas de incentivo ao empreendedorismo na redução dessas barreiras.

Palavras-chave: Distribuições Probabilísticas; Empreendedorismo; América Latina e Caribe; Medo do Fracasso.

¹ Departamento de Pós-Graduação – Matemática Aplicada e Computacional, Universidade Estadual Paulista – Presidente Prudente/SP – carladominique4@gmail.com

² Departamento de Pós-Graduação – Matemática Aplicada e Computacional, Universidade Estadual Paulista – Presidente Prudente/SP – karinahomobono@gmail.com

³ Departamento de Estatística – Faculdade de Ciência e Tecnologia – Universidade Estadual – Presidente Prudente/SP – rp.oliveira@unesp.br



Estimação via estatísticas do tipo Mellin para distribuição na classe slash modificada

Elisa Lopes do Nascimento;¹ Jaine Moura de Carvalho ² Josimar Mendes de Vasconcelos; ¹ Frank Gomes-Silva; ¹ **Resumo**

Este trabalho investiga a aplicação da Transformada de Mellin à distribuição Slashed Modificada Rayleigh Generalizada (SMRG). Proposta inicialmente por Barranco-Chamorro et al. (2021), essa distribuição possui representação estocástica e é capaz de modelar dados positivos com alta curtose, devido ao parâmetro adicional em sua composição. A análise estatística foca na avaliação da bondade de ajuste da distribuição. Inicialmente, são construídos os log-cumulantes do segundo tipo e elaborado o diagrama de log-cumulantes como ferramenta qualitativa, permitindo uma visualização gráfica do comportamento da distribuição em relação aos dados observados (Nicolas, 2002; Anfinsen et al., 2011). Além disso, propomos uma estatística baseada na estatística T^2 de Hotelling e nos log-cumulantes para avaliação quantitativa do ajuste, oferecendo um critério estatístico para verificar a adequação da distribuição aos dados empíricos (Anderson, 2003). Os resultados obtidos oferecem uma visão abrangente sobre a eficácia da Transformada de Mellin na modelagem estatística da distribuição SMRG. Por fim, as ferramentas desenvolvidas são aplicadas a dados reais, ilustrando sua utilidade prática. Os resultados destacam a eficiência da Transformada de Mellin na análise desta classe de distribuições, contribuindo para o avanço de métodos estatísticos na inferência de distribuições generalizadas.

Palavras-chave: Distribuição SMRG; Transformada de Mellin; log-cumulantes; Medidas de Bondade de Ajuste; Modelagem.

References

- Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis* (3 ed.). Wiley-Interscience.
- Anfinsen, S. N., A. P. Doulgeris, and T. Eltoft (2011). Goodness-of-fit tests for multilook polarimetric radar data based on the Mellin transform. *IEEE Transactions on Geoscience and Remote Sensing* 49(7), 2764–2781.
- Barranco-Chamorro, I., Y. A. Iriarte, Y. M. Gómez, J. M. Astorga, and H. W. Gómez (2021). A generalized rayleigh family of distributions based on the modified slash model. *Symmetry* 13(7), 1226.
- Nicolas, J.-M. (2002). Introduction aux statistiques de deuxième espèce: Applications des logs-moments et des logs-cumulants à l'analyse des lois d'images radar. *TS. Traitement du signal* 19(3), 139–167.

¹Departamento de Estatística e Informática, Recife, PE - Elisa.lopes@ufrpe.br

²Departamento de Estatística e Informática, Recife, PE - jaine.moura@ufrpe.br

¹Departamento de Estatística e Informática, Recife, PE - josimar.mendes@ufrpe.br

¹Departamento de Estatística e Informática, Recife, PE - franksinatrags@gmail.com



Modelagem de dados de suporte limitado: a distribuição *power logit* inflacionada em zero e um

Antonio Gabriel Freitas da Silva;¹ Francisco Felipe de Queiroz ²

Resumo

Dados contínuos limitados ao intervalo unitário são comuns em diversas áreas do conhecimento, como ecologia, biologia, economia e saúde pública, e frequentemente apresentam assimetria, variância dependendo da média e observações nos extremos do suporte. A distribuição beta é amplamente utilizada para modelar esse tipo de dado, mas possui limitações em cenários com valores cuja a variável apresenta valores nos extremos dos interava-los e observações atípicas. Neste trabalho, são propostos avanços na modelagem estatística de dados de suporte limitado por meio do desenvolvimento da classe de distribuições *power logit* inflacionada em zero e um (PLIZU), uma família flexível de distribuições que permite a modelagem de dados de suporte limitado que acomoda observações nos dois extremos do suporte. São apresentadas as propriedades das distribuições PLIZU, inferência baseadas em máxima verossimilhança e resultados de simulações para avaliar as propriedades dos estimadores em amostras de tamanho finito. A seguir, apresenta-se uma aplicação com dados reais de pesca de atuns tropicais.

Palavras-chave: dados de suporte limitado; proporções; distribuição *power logit* infla-
cionada; distribuição beta.

¹Departamento de Matemática Aplicada, IME-USP - antoniogsilva03@usp.br

²Departamento de Estatística, IME-USP - felipeq@ime.usp.br



SELEÇÃO DE MODELOS DE CAPTURA-RECAPTURA COM HETEROGENEIDADE TEMPORAL E EFEITO COMPORTAMENTAL À MARCAÇÃO

Márcio Roger Piagio;¹ Luis Ernesto Bueno Salasar² George Lucas Moraes Pezzott³

Resumo

O método de captura-marcção-recaptura é amplamente utilizado para estimar o tamanho de uma população, especialmente em estudos ecológicos nos quais é difícil observar diretamente todos os animais. O processo de seleção dos animais é realizado em etapas sucessivas (épocas) de amostragem, e comumente assume-se que a probabilidade de captura varia entre as épocas. Além disso, a marcação pode influenciar o comportamento dos animais, afetando as chances de recapturá-los. Por exemplo, a colocação de tags pode trazer desconforto ao animal, enquanto o uso de petiscos pode tornar o processo agradável. Nestes casos, a probabilidade de capturar um animal previamente marcado pode ser diferente daquela dos animais não marcados e, consequentemente, torna-se necessária a correta seleção do modelo de captura-recaptura. Neste sentido, o presente trabalho tem como objetivo o estudo e comparação entre o modelo de captura-recaptura com efeito comportamental da marcação e o modelo sem efeito, ambos considerando a heterogeneidade temporal. A estimativa dos parâmetros é feita via método de máxima verossimilhança, e a seleção entre os dois modelos é realizada utilizando os critérios AIC, BIC, AICc, HQIC e o Teste da Razão de Verossimilhança (TRV). Propõe-se um estudo de simulação com diferentes cenários para analisar o poder dos testes destes critérios. Os resultados mostraram que os critérios AIC e AICc possuem maior sensibilidade ao efeito comportamental, enquanto o BIC foi mais conservador. O TRV apresentou desempenho intermediário, alinhando-se com o HQIC. Por fim, a metodologia é aplicada em um conjunto de dados reais de captura-recaptura de ratos.

Palavras-chave: Método de captura-marcção-recaptura, Teste de hipóteses, Poder do teste, Seleção de modelos.

¹Programa de Pós-Graduação em Bioestatística - UEM - pg405670@uem.br

²Departamento de Estatística - UFSCar - luis@ufscar.br

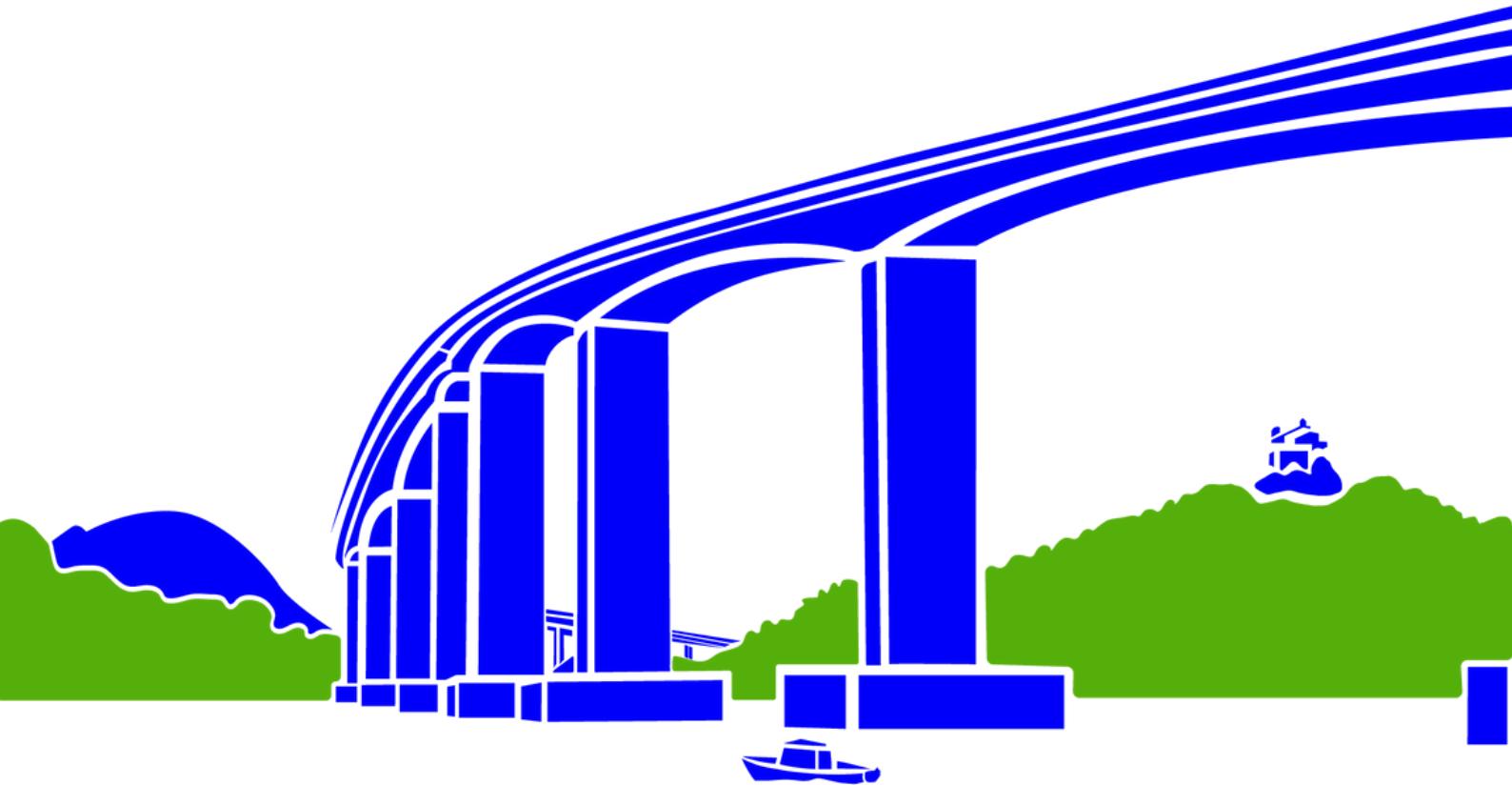
³Departamento de Estatística - UEM - glmpezott@uem.br

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Pôsters

Inferência Bayesiana





Spatial deformation in a Bayesian spatiotemporal model for matrix-variate responses

Rodrigo de Souza Bulhões ^{1,2}

Marina Silva Paez ³

Dani Gamerman ⁴

Resumo

In this work we propose a matrix-variate spatiotemporal model to fit two or more response variables, measured at different points in time and distinct locations in space. Here we relax the hypothesis of isotropy in space, since this assumption is not reasonable in various real situations. We treat the anisotropy of the spatial process using the concept of spatial deformation and we deal with temporal structure via Bayesian dynamic models. Through simulation studies and also with two examples, we show that the incorporation of spatial deformation provides better interpolations in some anisotropic scenarios. We also present a solution to deal with missing values in different response variables.

Palavras-chave: Matrix-variate spatiotemporal models; anisotropy; spatial deformation; geostatistics; Bayesian dynamic models.

¹ Programa de Pós-Graduação em Estatística, Instituto de Matemática, Universidade Federal do Rio de Janeiro, Rio de Janeiro/RJ, Brasil - <rodrigo.bulhoes@ufrj.br>.

² Departamento de Estatística, Instituto de Matemática e Estatística, Universidade Federal da Bahia, Salvador/BA, Brasil - <rbulhoes@ufba.br>.

³ Departamento de Métodos Estatísticos, Instituto de Matemática, Universidade Federal do Rio de Janeiro, Rio de Janeiro/RJ, Brasil - <marina@im.ufrj.br>.

⁴ Departamento de Métodos Estatísticos, Instituto de Matemática, Universidade Federal do Rio de Janeiro, Rio de Janeiro/RJ, Brasil - <dani@im.ufrj.br>.



Bayesian inference for zero-and/or-one augmented unit-gamma

Éric Oliveira Rocha;¹ Juvêncio Santos Nobre;² Caio Lucidius Naberezny Azevedo³

Abstract

In this work, we introduce a new distribution constructed from a mixture of the unit gamma (UG) distribution with either a degenerate distribution at c (where $c = 0$ or $c = 1$) or a Bernoulli distribution. This new distribution is based on the unit gamma distribution, reparameterized in terms of the mean, as originally proposed by Mousa et al. (2016), following the approach of Ospina and Ferrari (2010). As such, the augmented UG distribution emerges as a viable alternative to beta-based models commonly used to account for zero and/or one inflation. We detail the entire estimation procedure within a Bayesian framework, employing Markov Chain Monte Carlo (MCMC) methods implemented in Just Another Gibbs Sampler (JAGS) (Plummer, 2003), via the `jagsUI` package (Kellner, 2024) in the R software environment (R Core Team, 2024). Finally, we illustrate the theory by applying it to a real-world problem.

Keywords: Augmented unit-gamma; Bayesian inference; Mixture; Proportions; MCMC.

¹Instituto de Matemática, Estatística e Computação Científica, UNICAMP - e256291@dac.unicamp.br

²Departamento de Estatística e Matemática Aplicada, UFC - juvencio@ufc.br

³Instituto de Matemática, Estatística e Computação Científica, UNICAMP - cnaber@ime.unicamp.br



Análise de dados para um experimento com variável resposta ordinal

Fabíola Adriane Cardoso Santos ¹

Júlio Sílvio de Sousa Bueno Filho ²

Resumo

A análise de dados categorizados é uma tarefa importante em várias áreas de pesquisa uma vez que em grande parte dos estudos a variável de interesse é categorizada ordinal. Dessa maneira a modelagem e a previsão de resultados ordinais apresentam um especial interesse para muitos pesquisadores. Por esse motivo, diferentes modelos têm sido propostos, quando há alguma ordenação natural entre os grupos de dados e existe alguma correlação nas categorias ordenadas. Vários autores propuseram métodos para modelagem de dados ordinais, sejam em abordagens clássicas (McCullagh, 1980) ou sob a perspectiva bayesiana (Albert; Chib, 1993), (Cowles; Carlin; Connell, 1996).

Este trabalho tem por objetivo verificar a possibilidade de utilizar *prioris* Dirichlet para ajustar modelos de componentes da variância em genética e melhoramento de plantas quando se avalia caracteres qualitativos ordinais. Utilizando dados simulados verificamos que, por mais sofisticada que seja a simulação, os modelos resultantes são relativamente simplificados e que os parâmetros das normais subjacentes na ligação com as concentrações da Dirichlet não permitem estimar com precisão as componentes de covariância.

Palavras-chave: dados ordinais, simulação, Dirichilet, bayesiana.

¹ Departamento Estatística, UFLA – fabiola.santos1@estudante.ufla.br

² Departamento Estatística, UFLA – jssbueno@ufla.br



Bayesian Inference in Stochastic Processes to Identify Mortality Attributed to Sepsis

Victor Fossaluza ¹ Nicholas Wagner Eugenio ²

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.

Palavras-chave: Bayesian inference; stochastic processes; non-homogeneous Markov chains; attributable population fractions; adapted attributable hazard fraction.

¹Departamento de Estatística, IME-USP - victorf@ime.usp.br

²Departamento de Estatística, IME-USP - nicholaswe@gmail.com



Modelagem bayesiana de proporções de germinação em Fitopatologia

Antonio Leopoldo Cardoso Sabino;¹ Silvio Sandoval Zocchi;² Renan Fernandes Alves.³

Resumo

Os fungos são organismos heterotróficos, unicelulares ou pluricelulares, que liberam esporos durante sua fase reprodutiva. Esses esporos são responsáveis pela disseminação da espécie e podem ser transportados por diferentes meios, ajudando na sua propagação para ambientes adequados à germinação. A germinação dos esporos depende de fatores ambientais, principalmente temperatura e umidade. Entender como esses fatores afetam a proporção de germinação é fundamental para identificar os períodos climáticos mais favoráveis e aprimorar modelos preditivos. Diversos modelos são utilizados para descrever a germinação em função da temperatura (como Briére, Ratkowsky, Lactin) e do molhamento (como Gompertz, Weibull, logístico, entre outros). No entanto, muitos desses estudos apresentam dois problemas principais: a estimativa da proporção máxima pode ultrapassar o valor 1, e os intervalos de confiança podem extrapolar os limites do intervalo [0, 1]. Além disso, a maioria dos estudos assume uma distribuição normal para a variável resposta, recorrendo a transformações como raiz quadrada do arco-seno ou probit, em vez de considerar distribuições mais adequadas, como a binomial, que possui o mesmo suporte da proporção de germinação (entre 0 e 1). Apenas alguns trabalhos utilizam essa abordagem. Diante disso, apresentamos aqui uma função conjunta que é uma modificação de uma reparametrização da função beta generalizada vezes o modelo monomolecular para a análise de dados de proporção de germinação em função da temperatura e do período de molhamento de tal forma que sua imagem seja o intervalo [0, 1], com uma abordagem bayesiana. A metodologia foi implementada utilizando a linguagem de programação R, e é ilustrada por meio de um conjunto de dados reais. Os resultados revelam que o modelo pode obter as temperaturas cardinais e não extrapolar o intervalo [0, 1] para proporção de germinação.

Palavras-chave: Proporção; Temperaturas cardinais; MCMC.

¹Departamento de Ciências Exatas, ESALQ/USP - leopoldosabino@usp.br

²Departamento de Ciências Exatas, ESALQ/USP - sszocchi@usp.br

³Departamento de Fitopatologia e Nematologia, ESALQ/USP - renan.alves@usp.br



Determinant point process to clustering poverty data

Patricia Viana da Silva;¹ Rosangela Helena Loschi;² Cristiano de Carvalho Santos²

Resumo

Data-driven insights can hold the key to tackling global challenges like poverty, robust statistical methods are more critical than ever. Cluster analysis stands out as a powerful tool for processing vast amounts of information and uncovering valuable patterns, particularly with the advancements in machine learning. This work presents a flexible unsupervised clustering method that integrates repulsion among location parameters.

The proposed clusters are modeled via a Finite Mixture Model of Normal/Independent Distributions, where the behavior of location parameters follows a Determinantal Point Process (DPP) probability distribution. This approach inherently avoids redundant group formation, enhancing interpretability and precision in data segmentation. The proposal provides an a priori uncertainty structure to the DPP parameters and suggests an approach for its estimation. The model was used to address latent groups particularly for categorical variables with many levels offering an alternative to penalties.

We evaluate the specification of the model parameter effects for analyzing poverty spatial data at Minas Gerais State. The model allows robust grouping and parsimonious estimation of the number of groups, outperforming other models. Furthermore, we develop a complete Markov Chain Monte Carlo (MCMC) algorithm to estimate the model parameters, following the Bayesian paradigm with R and C++ implementation.

Palavras-chave: Clustering; Determinant Point Process; Finite Mixture Model; Normal/Independent Distribution; poverty.

¹Instituto de Matemática e Estatística, Universidade Federal de Uberlândia - patricia.viana@ufu.br

²Departamento de Esatística, Universidade Federal de Minas Gerais - loschi@est.ufmg.br

²Departamento de Esatística, Universidade Federal de Minas Gerais - ccsgaus@ufmg.br



Generalized Bayesian parametric quantile regression modelling

Amanda Merian Freitas Mendes;¹ Caio Lucidius Naberezny Azevedo ²

Resumo

Quantile regression is a technique for estimating conditional quantile functions on covariates of interest. Quantile regression models allow the effects of these covariates to be assessed along any quantile of the response variable, leading to a more complete picture of the relationship between the covariates and the response distribution. In this work, three new Bayesian quantile parametric regression models are proposed. The first model is developed for the analysis of positive data, it is defined by considering the generalized gamma distribution reparameterized by its quantiles. Secondly, we developed a class of Bayesian augmented (data) scale-mixture Birnbaum-Saunders (SMBS) quantile regression models for analyzing positive data. In this setup, the quantile of interest is modeled indirectly by using an appropriate conditional quantile based on the conditional distribution of the response given the augmented data. Finally, we developed a class of Bayesian quantile regression models based on the asymmetric exponential power (AEP) distribution as an alternative to traditional Bayesian quantile regression which is based on the asymmetric Laplace (AL) distribution. Inference tools are developed using Markov Chain Monte Carlo (MCMC) methods. Residual analysis, model comparison criteria, and influence analysis are also proposed. The developed methods have been implemented in Jags and Nimble systems, respectively, through the interface provided by the **R2jags** and **nimble** packages available in the R program. Simulation studies were performed to evaluate the performance of the estimators, the model fit assessment and the model comparison tools. Some real data sets were analyzed to illustrate the proposed models.

Palavras-chave: Bayesian quantile regression; parametric approach; generalized gamma distribution; scale-mixture Birnbaum-Saunders distribution; asymmetric exponential power family.

¹Departamento de Estatística, Universidade Federal do Paraná - amandamerianfm@gmail.com

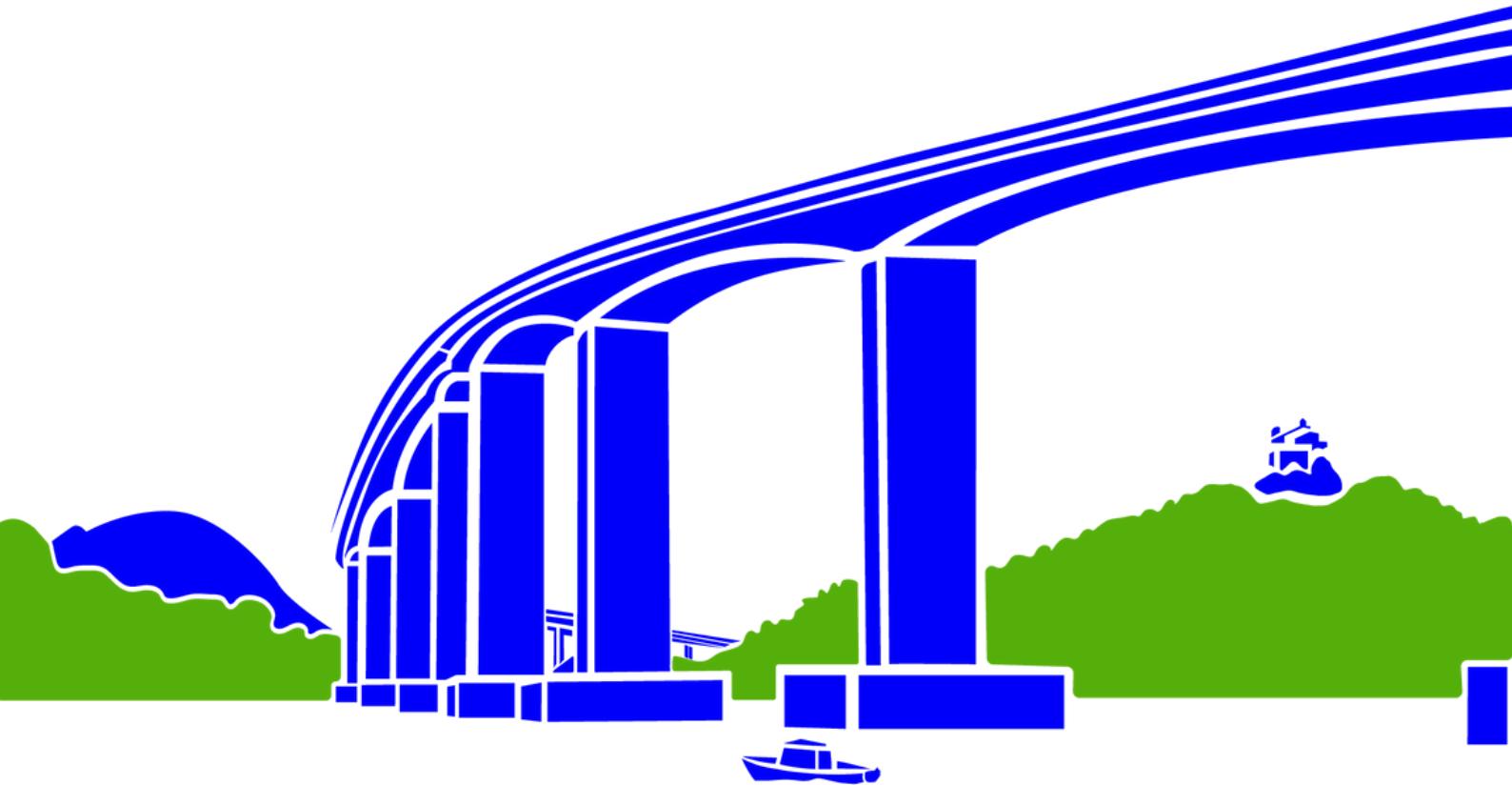
²Departamento de Estatística, Universidade Estadual de Campinas - cnaber@unicamp.br

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Pôsters

Geoestatística





Análise espacial para verificar índices socioeconômicos dos municípios de Minas Gerais

João Vitor Andrade Alves de Souza ¹ Yasmin Beatriz Pereira Santana; ² Melissa Wendy Ribeiro Vicente ³ Geraldo Magela da Cruz Pereira ⁴ Natiele de Almeida Gonzaga ⁵
Rayza Camila dos Santos Silva ⁶

Resumo

O estado de Minas Gerais, localizado na região Sudeste do Brasil, faz fronteira com São Paulo, Bahia, Espírito Santo, Rio de Janeiro, Mato Grosso do Sul, Goiás e Distrito Federal. Com Belo Horizonte como capital, é o segundo estado mais populoso, o quarto em extensão territorial e possui o terceiro maior PIB. Conta com 853 municípios, o maior número entre os estados brasileiros, e possui um Índice de Desenvolvimento Humano (IDH) de 0,774, considerado alto. Este estudo tem como objetivo realizar uma análise espacial, uma técnica estatística aplicada à modelagem de variáveis aleatórias no espaço-tempo. Os dados, extraídos do IBGE, incluem as variáveis PIB, população, densidade demográfica e IDHM, sendo analisados no software R. Inicialmente, foi aplicada estatística descritiva para compreender o comportamento dos dados. Em seguida, a análise espacial englobou a construção da Matriz de Vizinhança, o cálculo do Índice Global de Moran e do Índice Local de Moran, além da aplicação de modelos de regressão para identificar os fatores que influenciam o desenvolvimento humano nos municípios mineiros. A pesquisa permitiu identificar as variáveis mais significativas para o IDHM e ajustar o modelo estatístico mais adequado. Os resultados foram apresentados em mapas temáticos, evidenciando a influência das cidades vizinhas no IDHM de determinados municípios. Além disso, a análise contribuiu para a compreensão do desenvolvimento socioeconômico do estado, fornecendo subsídios para a formulação de políticas públicas voltadas à melhoria da qualidade de vida e à distribuição mais justa de renda.

Palavras-chave: IDHM; Minas Gerais; Estatística Espacial.

¹Programa de Pós Graduação em Estatística e Experimentação Agrícola- UFLA, Lavras - joao.souza20@estudante.ufla.br

²Programa de Pós Graduação em Estatística e Experimentação Agrícola- UFLA, Lavras - yasmin.santana@estudante.ufla.br

³Programa de Pós Graduação em Estatística e Experimentação Agrícola- UFLA, Lavras - melissa.vicente2@estudante.ufla.br

⁴Departamento de Estatística- UFLA, Lavras - geraldo.pereira@ufla.br

⁵Programa de Pós Graduação em Estatística e Experimentação Agrícola- UFLA, Lavras - natiele.gonzaga1@estudante.ufla.br

⁶Programa de Pós Graduação Interinstitucional em Estatística UFSCar-USP, São Carlos - rayzasantos@usp.br



Análise de experimentos em blocos casualizados com várias plantas por parcela considerando a dependência espacial

Luan Renner Ribeiro ¹

Renato Ribeiro de Lima ²

Resumo

Em experimentos de campo conduzidos em blocos casualizados com parcelas compostas por várias plantas, a análise estatística pode ser realizada utilizando as médias das parcelas ou considerando cada planta individualmente, semelhante a um delineamento com repetições dentro dos blocos. Contudo, a dependência espacial entre as plantas pode introduzir correlação entre os erros, violando o pressuposto de independência, fundamental para a validade da análise. Neste estudo realizou-se análise de variância e aplicou-se o teste de Tukey, considerando um experimento com a espécie *Eremanthus erythropappus* (DC.) McLeish, levando-se em conta a dependência espacial entre plantas, com o objetivo de aumentar a precisão das estimativas dos efeitos dos tratamentos. Foram conduzidas análises por quatro metodologias distintas: (1) utilizando as médias das parcelas; (2) considerando cada planta individualmente; (3) utilizando a abordagem geoestatística; e (4) utilizando a abordagem autorregressiva. O ajuste para dependência espacial foi considerado apenas em (3) e (4). As análises foram realizadas utilizando os pacotes spANOVA e ExpDes, no ambiente R. Na primeira abordagem, não houve diferença significativa entre os tratamentos. Na segunda, ocorreram diferenças significativas, porém com interação significativa entre blocos e tratamentos e discriminação desigual das médias nos blocos. Nas abordagens geoestatística e autorregressiva, ao considerar a dependência espacial, foram detectadas diferenças significativas entre as médias dos tratamentos. Entretanto, na abordagem autorregressiva, observou-se diferença significativa entre os blocos. Conclui-se que entre as metodologias avaliadas, a abordagem geoestatística se destacou por não apresentar diferença significativa entre os blocos, evidenciando seu papel no controle da heterogeneidade entre os blocos.

Palavras-chave: Experimentação; Geoestatística; Modelos autorregressivos; spANOVA; Software R.

¹ Departamento de Estatística, DES/UFLA – luan.ribeiro3@estudante.ufla.br

² Departamento de Estatística, DES/UFLA – rrlima@ufla.br



Previsão espacial de eventos de precipitação extrema: uma abordagem bayesiana e geoestatística

Thales Rangel Ferreira ^{1,7} Gilberto Rodrigues Liska ^{2,7} Carlos José dos Reis ³ Arthur Fellipy Pereira Silva ⁴ Luiz Alberto Beijo ^{5,7} Taciana Villela Savian ⁶

Resumo

A ocorrência de eventos extremos de precipitação pode causar danos significativos à infraestrutura urbana, ao meio ambiente e às atividades humanas em geral. Assim, compreender o comportamento desse fenômeno pode auxiliar no planejamento de atividades suscetíveis a tais danos. Nesse sentido, este capítulo tem como objetivo fornecer informações sobre modelagem estatística realizada por duas abordagens distintas e bastante promissoras: Bayesiana e Geoestatística. Ambas as metodologias foram aplicadas na modelagem espacial da precipitação máxima nas regiões sul e sudoeste do estado de Minas Gerais, Brasil, utilizando Inferência Bayesiana combinada com Krigagem e Ponderação Inversa da Distância (IDW). O estudo utilizou dados diários de precipitação de 29 cidades da região. Para a análise utilizando IDW, Krigagem Ordinária (OK) e Krigagem Log-Normal (LNK), as previsões de precipitação foram obtidas por Inferência Bayesiana para cada localidade e período de retorno (RP) analisados. As previsões foram geradas utilizando a melhor estrutura a priori para cada município. Para os métodos de Krigagem, serão considerados os modelos de semivariograma Gaussiano, Esférico, Exponencial e de Onda. Para decidir sobre o melhor modelo de semivariograma, foram realizadas validação cruzada e erro médio de predição (MPE). Os resultados da avaliação indicaram que, para predição espacial no período de maior retorno, o modelo mais adequado foi compatível com o semivariograma de Onda. Portanto, este modelo foi utilizado para gerar mapas de predição para valores de RP de 50 e 100 anos. Os resultados fornecem evidências de que a combinação das abordagens Bayesiana e geoestatística é uma alternativa viável e promissora.

Palavras-chave: Distribuição Generalizada de Valores Extremos; krigagem; precipitação; nível de retorno; semivariograma

¹PPGEEA - ESALQ, Piracicaba, SP - thales.rangel8@gmail.com

²DTAiSeR - CCA - UFSCar, Araras, SP - gilbertoliska@ufscar.br

³UNIFAL-MG, Alfenas, MG - carlos.mat.unifal@gmail.com

⁴CCA - UFSCar - Araras, SP - arthurfps@estudante.ufscar.br

⁵DE - UNIFAL-MG, Alfenas, MG - prof.beijo@gmail.com

⁶DCE - USP/Esalq, Piracicaba, SP - tvsavian@usp.br

⁷O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001



Análise Geoestatística do Risco de Fogo em Minas Gerais

Daví Barbosa Pereira de Sousa;¹ Marcela Silva de Araujo;² Viviane Costa Silva;³
Gabriel Graciano de Mendonça;⁴ João Domingos Scaloni⁵

Resumo

O risco de fogo é um índice que mensura a probabilidade de ocorrência de queimadas, considerando aspectos como temperatura do ar, umidade relativa, precipitação, velocidade do vento e tipo de cobertura vegetal. Neste estudo, visa-se identificar locais com maior risco de fogo em Minas Gerais, a partir de dados de focos de queimadas disponibilizados pelo Instituto Nacional de Pesquisas Espaciais (INPE) para o período de 01 de janeiro a 31 de março de 2025. Com a aplicação do semivariograma, foi possível identificar evidências de associação espacial. Como visivelmente os dados não seguem o pressuposto de normalidade, para a estimação de parâmetros foi utilizado o método dos mínimos quadrados, no qual foi identificado que a distribuição Wave é a mais indicada para a krigagem. Por meio da krigagem ordinária com essa distribuição, foram identificados maiores riscos de fogo nas áreas norte, leste e sul do estado, com baixa oscilação espacial. Já áreas próximas às fronteiras estaduais, especialmente nas porções sul, norte e oeste, apresentaram maior incerteza nas estimativas. Os resultados contribuem para a compreensão e prevenção de queimadas, auxiliando no planejamento de ações preventivas e na formulação de políticas públicas.

Palavras-chave: Queimadas; Semivariograma; Interpolação espacial; Krigagem ordinária; Fatores climáticos.

¹Universidade Federal de Lavras, Departamento de Estatística, Lavras, MG - davi.sousa@estudante.uflla.br

²Universidade Federal de Lavras, Departamento de Estatística, Lavras, MG - marcela.araujo1@estudante.uflla.br

³Universidade Federal de Lavras, Departamento de Estatística, Lavras, MG - viviane.silva6@estudante.uflla.br

⁴Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, Recife, PE - gabriel.gracianom@ufrpe.br

⁵Universidade Federal de Lavras, Departamento de Estatística, Lavras, MG - scalon@ufla.br



Comparação de funções de base radial para interpolação espacial de variáveis categóricas

Rogério Moraes de Lima ¹

Enio Júnior Seidel ²

Resumo

A interpolação espacial de variáveis categóricas é relevante em aplicações na agricultura e no meio ambiente. Este trabalho tem como objetivo comparar o desempenho de diferentes funções de base radial (FBRs) na interpolação espacial de variáveis categóricas. Foram simulados três cenários com 2, 4 e 8 categorias em um espaço bidimensional com 200 observações. As variáveis categóricas foram transformadas em variáveis indicadoras e interpoladas por meio de equações multiquádricas. Doze variações das funções multiquádricas generalizadas foram testadas com diferentes valores para as constantes a e b [$a=(0, 5, 50, 500)$; $b=(-1, 0, 1)$]. Os dados, em cada cenário, foram divididos em dados de treinamento (80%) e de validação (20%). Foi utilizada matriz de confusão entre os dados de validação observados e preditos pela interpolação especial para obter a medida de acurácia e o índice Kappa para comparação das FBRs. A FBR multiquádrica generalizada com constantes $a=50$ e $b=0$ apresentou os melhores resultados em todos os cenários avaliados, com altos valores de acurácia (0,875, 0,900 e 0,950 nos cenários com 2, 4 e 8 categorias, respectivamente) e índice Kappa (0,853, 0,863 e 0,890 nos cenários com 2, 4 e 8 categorias, respectivamente), superando as outras versões da FBR mesmo com o aumento do número de categorias. Os resultados obtidos fornecem subsídios para a escolha da FBR mais adequada em aplicações reais.

Palavras-chave: Estatística espacial; Equações multiquádricas; Mapeamento; Autocorrelação espacial.

¹ Curso de Matemática, UNIFAP, Macapá - AP – roger201415@gmail.com

² Departamento de Ciências Exatas e Tecnológicas, UNIFAP, Macapá - AP – seidel_enio@unifap.br



Agrupamentos espaciais da precipitação pluviométrica no estado do Rio Grande do Sul

Fernando Henrique Martins Fernandes ¹

Alana Pisoni ²

Enio Júnior Seidel ³

Resumo

A identificação de regiões homogêneas de precipitação é essencial para a gestão dos recursos hídricos. Este trabalho teve como objetivo identificar agrupamentos de estações meteorológicas com altos e baixos valores de precipitação anual e mensal no estado do Rio Grande do Sul (RS). Foram utilizados dados mensais de 277 estações localizadas no estado. Para a identificação dos grupos, aplicaram-se os métodos de Moran Local, Geary Local, Getis-Ord Local e Getis-Ord Local tipo 2, que permitem detectar agrupamentos do tipo Alto-Alto e Baixo-Baixo. A precipitação média anual mostrou predominância de grupos Alto-Alto no norte do RS, enquanto os grupos Baixo-Baixo concentraram-se no sul. Em relação à precipitação média mensal, nos meses de janeiro, fevereiro, março, abril, maio, junho, setembro, outubro, novembro e dezembro ocorrem grupos Alto-Alto mais ao norte do RS e grupos Baixo-Baixo mais ao sul, sudoeste e sudeste. Já nos meses de julho e agosto ocorrem grupos Alto-Alto mais a leste e grupos Baixo-Baixo mais a oeste do RS. De modo geral, os diferentes métodos apresentaram resultados semelhantes. O método de Geary Local tendeu a formar agrupamentos maiores, com mais estações. Observou-se também que os grupos Alto-Alto, no geral, abrangem mais estações do que os grupos Baixo-Baixo.

Palavras-chave: Estatística espacial; Autocorrelação local; Brasil subtropical; Estações meteorológicas.

¹ Curso de Matemática, DCET, UNIFAP, Macapá - AP – fernando.martins1001.com.br

² Itaipu Parquetec, Foz do Iguaçu – PR – alanapisoni@hotmail.com

³ Departamento de Ciências Exatas e Tecnológicas, UNIFAP, Macapá - AP – seidel_enio@unifap.br



Geostatistical estimation and prediction for censored responses using STAN Bayesian approach

Victor Eduardo Lachos Olivares ¹

Resumo

Geostatistics is a branch of spatial statistics that studies how continuous phenomena like mineral deposits, pollution, or weather patterns vary across space. Its core idea is simple: measurements taken closer together tend to be more similar than those farther apart. Traditionally, geostatistical models assume that all data are fully observed. However, real-world data often face censoring, where some values are only partially known, for instance, pollutant levels falling below detection limits. This poses challenges for classical estimation methods like the EM algorithm, which struggles when key calculations lack closed-form solutions. To address this, simulation-based methods such as Monte Carlo EM and Stochastic Approximation EM (SAEM) have been developed to better handle censored data. Building on this, the present work proposes a Bayesian approach, offering a flexible probabilistic framework that naturally incorporates uncertainty and prior knowledge. By leveraging STAN and the NUTS-HMC algorithm, this approach aims to improve parameter estimation in spatial models, particularly when using complex covariance structures. The methodology is rigorously tested through simulations under both censored and uncensored scenarios and validated with real-world data. The study not only highlights the strengths of Bayesian methods over traditional approaches but also points toward future advancements in modeling spatial data with censoring.

Palavras-chave: Gaussian process; SAEM estimation; STAN;NUTS-HMC algorithm; Censored data.

¹ Department of Statistics, Inter-institutional Graduate Program in Statistics (PIPGES), São Carlos, São Paulo, elachoso27@gmail.com



Ajuste de distribuições de probabilidade para algumas métricas obtidas do semivariograma

Samuel Braga Cavalcante ¹

Diogo Francisco Rossoni ²

Enio Júnior Seidel ³

Resumo

A avaliação da dependência espacial é fundamental no contexto da Geoestatística. Tal avaliação é realizada com base em métricas obtidas do semivariograma. O objetivo deste estudo foi realizar uma primeira aproximação das distribuições de probabilidade das principais métricas de dependência especial. São considerados o Efeito Pepita Relativo (EPR), o Grau de Dependência Espacial (GDE), o Índice de Dependência Espacial (IDE) e a Medida de Dependência Espacial (MDE). Foram geradas simulações de processos espaciais sob modelo esférico com diferentes estruturas de dependência espacial variando os valores de contribuição, efeito pepita e alcance. Foi obtido o gráfico de Cullen e Frey para avaliar as distribuições de probabilidade candidatas. Por fim, o teste de Kolmogorov-Smirnov (KS) foi aplicado para verificar o ajustamento das distribuições aos dados. O EPR e o GDE, por serem complementares ($EPR=1-GDE$) tiveram melhor ajuste pela mesma distribuição de probabilidade, que foi a distribuição beta. Já o IDE teve melhor ajuste pela distribuição normal. O MDE teve melhor ajuste pela distribuição uniforme continua. A identificação das distribuições candidatas permite melhor representar o comportamento das métricas, aprimorando as inferências sobre a dependência espacial.

Palavras-chave: Estatística espacial; Geoestatística; Dependência espacial; Variáveis aleatórias.

¹ Curso de Matemática, UNIFAP, Macapá - AP – samuelunifap@gmail.com

² Departamento de Estatística, UEM, Maringá - PR – dfrossoni@uem.br

³ Departamento de Ciências Exatas e Tecnológicas, UNIFAP, Macapá - AP – seidel_enio@unifap.br



Rodovias Paranaenses: Um estudo exploratório sobre a ocorrência de acidentes.

Vitor Gabriel Pignatari¹

Paulo Justiniano Ribeiro Junior²

Resumo

As rodovias paranaenses desempenham um papel fundamental no dia a dia tanto do estado quanto do país, sendo palco do transporte de mercadorias e pessoas. Essas vias são uma infraestrutura essencial para o desenvolvimento regional. Diante disso, há diversos fatores que influenciam suas dinâmicas de tráfego, sob a influência destes aspectos essas rodovias têm sido cenário de diversos acidentes ao longo dos anos, e disso surge uma pergunta fundamental: Seriam elas de fato seguras? Este estudo propõe uma análise descritiva e tem como objetivo analisar as condições associadas aos acidentes registrados nas rodovias federais que cruzam o estado do Paraná entre os anos de 2021 e 2024, utilizando dados públicos disponibilizados pela PRF (<https://www.gov.br/prf/pt-br/acesso-a-informacao/dados-abertos/dados-abertos-da-prf>). A partir dessa base, busca-se responder questões como: quais trechos apresentam maior incidência de acidentes? Quais fatores podem influenciar a ocorrência de acidentes? Os acidentes mantêm um padrão ao longo do tempo ou apresentam variações significativas? Há predominância de certos tipos de acidentes com condições semelhantes? Existem fatores comuns que contribuem para que os acidentes resultem em vítimas ou feridos? Além de responder estas perguntas, este estudo busca propor ideias para futuros estudos, que possam contribuir para a melhoria da segurança viária no estado do Paraná.

Palavras-chave: Rodovias paranaenses; Acidentes; Segurança Viária.

¹ Departamento de Estatística, UFPR – vitor.pignatari@ufpr.br

² Departamento de Estatística, UFPR – paulojus@ufpr.br



Visualização e análise de dados georreferenciados de soros para acidentes com animais peçonhentos

Emanuel Victor da Silva Favorato ¹; Cibele Maria Russo ²; Hellen Geremias Gatica Santos ³; Maristela Oliveira Santos ⁴; Oilson Alberto Gonzatto Junior ⁵

Resumo

Com o aumento observado na incidência de casos de ataques de animais peçonhentos a humanos, a utilização de métodos estatísticos para a determinação de novas ocorrências torna-se fundamental para que os gestores de saúde possam garantir que os estoques de soros antiveneno estejam regularmente abastecidos, de modo a otimizar o armazenamento e salvar vidas. Neste projeto, utilizamos dados provenientes de fontes públicas, como SINAN e IBGE, para a implementação dos modelos estatístico-espaciais Geographically Weighted Regression (GWR) e Geographically and Temporally Weighted Regression (GTWR) — considerando os modelos Tweedie, Poisson e Binomial Negativo —, a fim de compreender a distribuição dos ataques e a tendência de longo prazo, respectivamente, em municípios do estado de São Paulo. A ponderação geográfica foi feita utilizando kernel bi-square para os pesos e validação cruzada (Cross-Validation) para a largura das bandas. Também fez-se uso dos índices I de Moran e Moran Local, a fim de observar a autocorrelação espacial, e do Lisa Map, para a análise das regiões onde há maior suscetibilidade a ataques, implementados no software Python. Resultados preliminares apontam para o melhor desempenho da regressão de Tweedie, a existência de significativa autocorrelação espacial, bem como possível heterocedasticidade dos resíduos, o que representa oportunidades de aprimoramento na modelagem de dados.

Palavras-chave: Estatística; Saúde; Modelos estatístico-espaciais; GWR; autocorrelação espacial; Cross-Validation; SINAN; IBGE.

¹Instituto de Ciências Matemáticas e de Computação (ICMC - USP), São Carlos/SP - emanuelvictor-dasilva856@usp.br

²Instituto de Ciências Matemáticas e de Computação (ICMC - USP), departamento de Matemática Aplicada e Estatística, São Carlos/SP - cibele@icmc.usp.br

³Instituto Carlos Chagas/Fundação Oswaldo Cruz, Curitiba/PR - hellengeremias@gmail.com

⁴Instituto de Ciências Matemáticas e de Computação (ICMC - USP), departamento de Matemática Aplicada e Estatística, São Carlos/SP - mari@icmc.usp.br

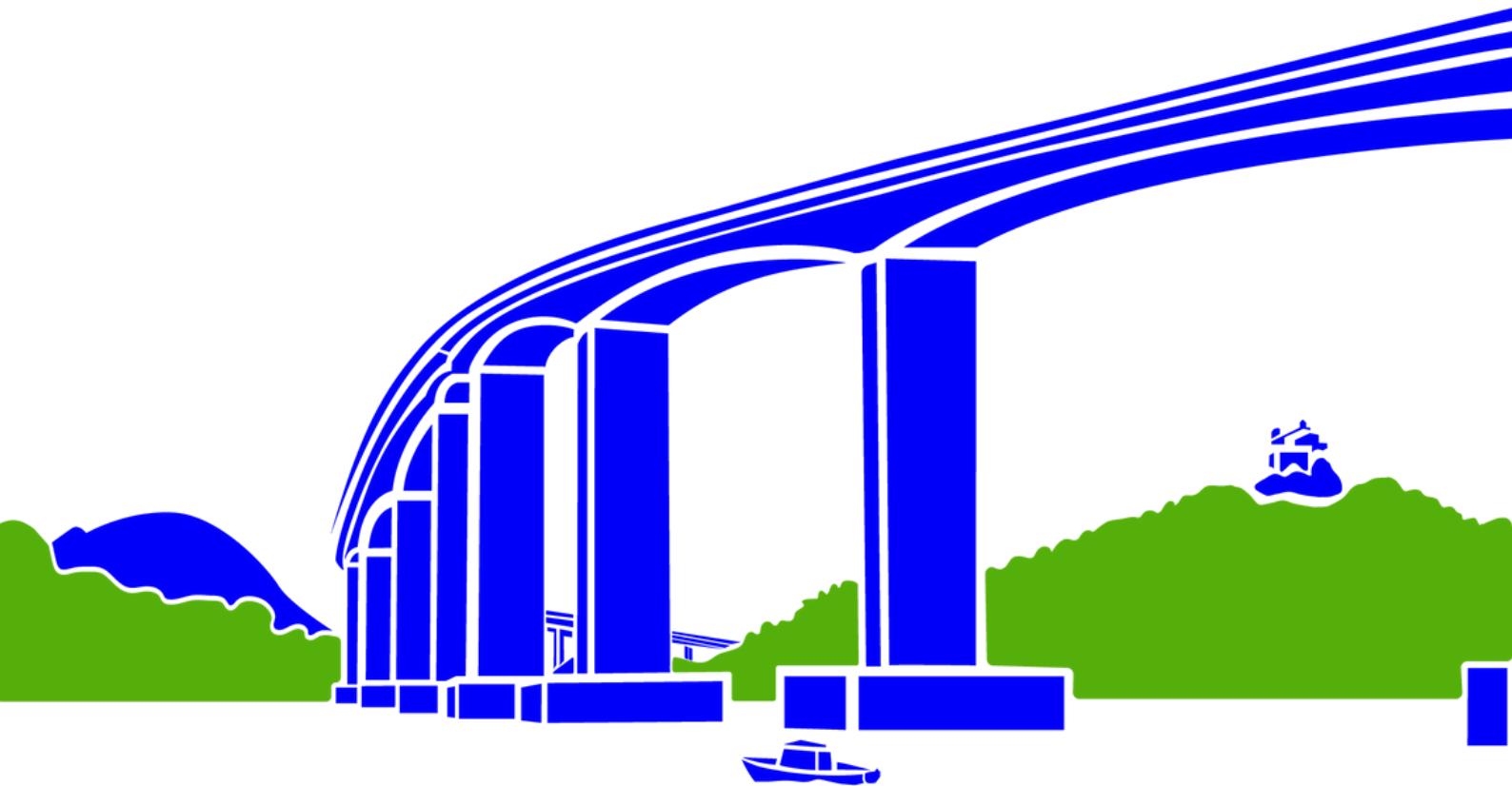
⁵Instituto de Ciências Matemáticas e de Computação (ICMC - USP), departamento de Matemática Aplicada e Estatística, São Carlos/SP - oilson.agjr@icmc.usp.br

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Pôsters

Estatística Multivariada





Relações lineares entre caracteres de milho

Daniela Lixinski Silveira ¹

Alberto Cargnelutti Filho ²

Resumo

O estudo de relações lineares entre os caracteres de milho pode possibilitar a identificação de caracteres adequados para seleção indireta. Assim, os objetivos deste trabalho foram verificar se há relações lineares entre caracteres de milho e se há caracteres que possam ser indicados para a seleção indireta. Foi conduzido um experimento no ano agrícola 2020/2021 com 16 cultivares de milho, no delineamento experimental blocos casualizados, com três repetições. Foram avaliados 21 caracteres de milho, sendo três fenológicos, três morfológicos, quatro produtivos e 11 caracteres de pendão. Foi realizada a análise de variância e estimada a matriz de coeficientes de correlação fenotípica entre os caracteres. A multicolinearidade foi diagnosticada, com base no número de condição. Com os caracteres que permaneceram, após o diagnóstico de multicolinearidade, foi realizada a análise de trilha. Para a análise de trilha, a produtividade de grãos foi considerada o caractere principal e os demais caracteres foram considerados explicativos. A altura de planta e o número de ramificações secundárias do pendão apresentam relação linear positiva com a produtividade de grãos e podem ser utilizados para seleção indireta de plantas de milho, em programas de melhoramento genético.

Palavras-chave: *Zea mays* L.; multicolinearidade; análise de trilha; seleção indireta; produtividade de grãos.

¹ Bolsista do Programa Institucional de Pós-Graduação - PIPD da Capes - junto ao Programa de Pós-Graduação em Agronomia, Universidade Federal de Santa Maria - danielisil@gmail.com

² Departamento de Fitotecnia - Universidade Federal de Santa Maria - alberto.cargnelutti.filho@gmail.com

Agradecimentos ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq - Processos 304652/2017-2 e 304878/2022-7) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes, Código de Financiamento 001) pela concessão de bolsas aos autores. Aos estudantes bolsistas e voluntários pelo auxílio na coleta de dados.



Análise de Agrupamento das mesorregiões do Espírito Santo (2021-2023) baseado nas produções de diferentes culturas

Maria Vitória Neves;¹ Louziane Ribeiro Carvalho;² Valdeline de Paula Mequelinho Ferreira;³ Daniel Furtado Ferreira⁴

Resumo

O setor agropecuário continua sendo um dos principais geradores de fonte de renda no país, inclusive no Espírito Santo. Entre 2021 e 2023, a produção agrícola capixaba apresentou variações significativas entre as diferentes culturas produzidas nas mesorregiões do estado, evidenciando a análise de padrões produtivos. Neste estudo, utilizamos técnicas de análise de agrupamento, como o método de agrupamento hierárquico de Ward, para classificar os principais produtos agrícolas produzidos no estado com base em sua similaridade. Os dados foram coletados da produção agrícola municipal (PAM) do Instituto Brasileiro de Geografia e Estatística (IBGE). Foram coletados dados referentes à área plantada(ha), produção (t) e rendimento médio da produção (kg por ha), das mesorregiões do Espírito Santo. Posteriormente, os dados foram padronizados e organizados em clusters segundo a metodologia adotada permitindo identificar grupos de culturas com comportamentos semelhantes ao longo do período analisado, utilizando o software R. Os resultados mostraram que existem padrões diferentes entre as culturas, o que pode ajudar a criar políticas agrícolas mais direcionadas e com isso, melhorar a gestão dos recursos.

Palavras-chave: Produção Agrícola; cluster; ward.

¹Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, Universidade Federal de Lavras - maria.neves5@estudante.ufla.br

²Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, Universidade Federal de Lavras -louiziane.carvalho2@estudante.ufla.br

³Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, Universidade Federal de Lavras -valdeline.ferreira1@estudante.ufla.br

⁴Departamento de estatística, Universidade Federal de Lavras - danielff@ufla.br



Correlação dos teores de antocianina e compostos fenólicos totais de alfaces com parâmetros colorimétricos

Camily Gomes da Silva¹
Gilberto Rodrigues Liska²
Maria Helena Serrano Carvalho³
Christiane de Fátima Martins França⁴
Fernando Cesar Sala⁵

Resumo

Tradicionalmente as quantificações dos teores de clorofilas, antocianinas e compostos fenólicos totais, que têm benefícios à saúde, são realizadas pela extração dos solutos foliares em extrato cetônico e posterior determinação espectrofotométrica na região do visível. Entretanto essas análises são realizadas em laboratório, de forma destrutiva e demanda muitas horas na sua determinação. A análise de cor instrumental é uma técnica rápida e simples, pois não utiliza reagentes e pode ser realizada diretamente, utilizando espectrofotômetros portáteis e os valores medidos podem ser traduzidos em alterações físicas-químicas nas hortaliças. Nesse contexto, o objetivo deste trabalho foi avaliar a correlação entre os parâmetros de cor instrumental com os parâmetros químicos. O estudo foi desenvolvido no Centro de Ciências Agrárias da UFSCar e foi planejado um experimento em DIC envolvendo 10 espécies de alface roxa. Foram quantificados a cor instrumental (a^* , b^* , C , hue , L), compostos fenólicos totais e teores de antocianina. Os dados foram submetidos e análise de componentes principais (CP) com o auxílio do software R. Os resultados mostram que 85,76% da variação total dos dados é explicada pelas duas primeiras CP's. A CP1 tem alta correlação com os parâmetros de cor instrumental, antocianina e compostos fenólicos totais. O material Betânia destacou-se pelos altos valores dos parâmetros L , b^* , C , hue e menores valores dos teores de antocianina, compostos fenólicos totais e parâmetro colorimétrico a^* . Com interpretação contrária, tem-se a espécie Rouxai. Estes resultados podem contribuir na seleção de espécies de alface roxa de interesse comercial.

Palavras-chave: Antocianina, Cor instrumental, Programa R, Análise de componentes principais

¹ Discente da Engenharia Agronômica, UFSCar, CCA - camilygomessilva@estudante.ufscar.br

² Docente do Departamento de Tecnologia Agroindustrial e Socioeconomia Rural do Centro de Ciências Agrárias, UFSCar, Araras - SP, Brasil - gilbertoliska@ufscar.br

³ Docente do Departamento de Tecnologia Agroindustrial e Socioeconomia Rural do Centro de Ciências Agrárias, UFSCar, Araras - SP, Brasil - marjotta@ufscar.br

⁴ Docente do Departamento de Tecnologia Agroindustrial e Socioeconomia Rural do Centro de Ciências Agrárias, UFSCar, Araras - SP, Brasil – arthurfps@estudante.ufscar.br

⁵ Docente do Biotecnologia e Produção Vegetal e Animal do Centro de Ciências Agrárias, UFSCar, Araras - SP, Brasil – fcsala@ufscar.br



Seleção de variáveis para modelos de regressão multivariada para dados compositionais caso pessoas com deficiência

Paulo Tadeu Meira e Silva de Oliveira¹

Resumo

Pessoas com deficiência representa parcela significativa da população e são considerados um dos grupos mais vulneráveis e excluídos. De acordo com o IBGE elas podem ser divididas em visual, auditiva, física, intelectual e múltipla; e, apresentam desvantagens no mercado de trabalho. Dados compositionais são descrições quantitativas das partes do todo. As medições que envolvem probabilidades ou proporções podem ser pensadas como dados compositionais. A regressão com resposta composicional é frequentemente empregada nas ciências sociais. Como avaliar proporcionalmente os tipos de deficiência nos diferentes municípios, estado ou país. Modelos de regressão multivariada considera p variáveis respostas Y_1, \dots, Y_p . A cada uma dessas p variáveis está associado a um grupo de diferentes variáveis explicativas tem por objetivo estimar conjuntamente os parâmetros das p regressões obtendo ganho de eficiência dos estimadores. Seleção de variáveis significa escolher um subconjunto que retenha as variáveis preditoras mais importantes excluindo as demais. Neste trabalho pretendemos estimar as proporções das diferentes deficiências nos 5565 municípios brasileiros. Naturalmente, é de interesse analisar como variam essas proporções em função de certas mudanças contextuais, por exemplo, localização geográfica, tempo ou disponibilidade de recursos disponíveis nos diferentes municípios para o atendimento desta população. Neste trabalho ajustaremos modelos variáveis resposta tipos de deficiência em função de explicativas relacionadas a nível de instrução, renda outros obtidos no banco de dados do Censo Demográfico de 2010 após aplicar transformação \ln juntamente com estudo comparativo de seleção de variáveis e de modelos utilizando para os diferentes ajustes as técnicas de Regressão Multivariada em dados compositionais.

Palavras-chave: Tipos de deficiência; Regressão multivariada; Dados compositionais; PLS; Seleção de variáveis.

¹ IME-USP, poliver@usp.br



Estimativa de Densidade por Programação Quadrática com Restrições de CDF: Uma Abordagem Baseada em SV-Density

Anderson Andolfato Filho¹ Anderson Ara²

Resumo

Neste trabalho, investiga-se uma metodologia alternativa para estimação de densidades unidimensionais, denominada **SV-Density**, baseada em formulações de programação quadrática inspiradas nos princípios da teoria de Vapnik. A proposta busca minimizar uma função de custo quadrática associada a um kernel Gaussiano normalizado, sujeita a restrições de densidade válidas (soma dos pesos unitária e não-negatividade) e, principalmente, restrições na aproximação da função de distribuição acumulada (CDF), com base na CDF empírica. Utilizando diferentes valores de largura de banda no kernel (gamma) e na função acumulada (sigma), analisamos o comportamento do estimador frente ao método clássico de Parzen (KDE). As métricas avaliadas incluíram **correlação com a densidade real** e **erro quadrático médio (MSE)** em simulações com dados de distribuição normal.

Palavras-chave: Densidade; KDE; SVDE

¹ Departamento de estatística da UFPR, Curitiba- Paraná – anderson.andolfato@ufpr.br

² Departamento de estatística da UFPR, Curitiba- Paraná – ara@ufpr.br



Modelagem Estatística da Distribuição de Espécies Florestais na Amazônia com Análise de Componentes Principais

Werlleson Nascimento ¹; Lucietta Guerreiro Martorano ²; Jose Reinaldo da Silva Cabral de Moraes ³; Leila Sheila Silva Lisboa ⁴; Carlos Tadeu dos Santos Dias ⁵

Resumo

Este estudo utilizou a Análise de Componentes Principais (PCA), para investigar a distribuição geográfica de duas espécies nativas da Amazônia: Castanha-da-Amazônia (*Bertholletia excelsa*) e Cumaru (*Dipteryx odorata*). As ocorrências das espécies foram obtidas da Base de Dados de Informações Ambientais (BDIA/IBGE), complementadas por registros do GBIF, herbários e literatura científica. Foram consideradas variáveis geográficas e climáticas extraídas de uma série histórica (1961–2022), incluindo precipitação, temperaturas (mínima, média e máxima), umidade relativa, déficit e excedente hídrico (DEF e EXC), déficit de pressão de vapor (DPV) e o Índice de Precipitação Padronizado (SPI). O SPI foi utilizado por sua capacidade de normalizar a variabilidade da precipitação, enquanto DEF e EXC foram calculados com uma capacidade de água disponível de 300 mm. A PCA revelou que os três primeiros componentes explicaram 83,5% da variação total. O PC1 destacou a umidade relativa (13,16%), DEF (12,75%) e a precipitação do trimestre mais seco (11,95%); o PC2 foi dominado por temperatura máxima (44,69%), DPV (16,44%) e temperatura média (11,70%); e o PC3, por longitude (37,70%), DPV (15,06%) e temperatura mínima (13,93%). O biplot indicou que a Castanheira tende a ocorrer em ambientes mais úmidos, enquanto o Cumaru mostrou maior plasticidade ambiental, ocorrendo em condições climáticas e geográficas mais diversas.

Palavras-chave: Amazônia Legal; análise multivariada; variáveis climáticas; distribuição geográfica; agrometeorologia.

¹ Departamento de Ciências Exatas, Universidade de São Paulo (USP/ESALQ) – werlleson.nascimento@usp.br

² NAPT Médio Amazonas, Embrapa Amazônia Oriental – lucieta.martorano@embrapa.br

³ Sombrero insurance – reinaldojmoraes@gmail.com

⁴ Secretaria municipal de educação, SEMEC – lisboa.leilass@gmail.com

⁵ Departamento de Ciências Exatas, Universidade de São Paulo (USP/ESALQ) – werlleson.nascimento@usp.br



Associações entre Características Morfológicas e Produtivas em Açaizeiros via Análise de Correlação Canônica

Werlleson Nascimento ¹; João Thiago Rodrigues de Souza ²; Clarice Garcia Borges
Demétrio ³; Carlos Tadeu dos Santos Dias ⁴

Resumo

Em pesquisas agronômicas, é comum a presença de variáveis agrupadas em diferentes conjuntos. A Análise de Correlação Canônica (ACC) é uma técnica multivariada que permite investigar associações lineares entre dois grupos distintos de variáveis. Este estudo utilizou a ACC para analisar a relação entre características vegetativas e produtivas de plantas de açaí (*Euterpe oleracea*) oriundas de duas localidades paraenses: Abaetetuba e Santa Izabel. Foram analisadas oito variáveis referentes às plantas (como número de folhas e altura) e 14 características relacionadas aos frutos (como massa de frutos maduros e número de frutos por cacho). A análise revelou uma forte correlação entre os dois grupos na primeira função canônica (0,9467), que explicou 59% da variabilidade total e foi estatisticamente significativa ($p = 0,048$). As demais funções apresentaram menor relevância e não foram significativas. No grupo das plantas, atributos como circunferência do estipe (CAP), número de folhas em emissão (NFE) e comprimento da folha (LFL) foram os mais influentes. Entre as variáveis dos frutos, destacaram-se número total de frutos e número médio de frutos por racemo. Esses resultados indicam que certas características morfológicas das plantas estão fortemente associadas ao rendimento e qualidade dos frutos. Assim, a ACC demonstrou ser uma ferramenta útil para compreender como diferentes aspectos das plantas se relacionam com a produtividade do açaizeiro em distintos contextos ambientais.

Palavras-chave: Amazônia Legal; análise multivariada; análise de correlação; fitotecnia; fruticultura.

¹ Departamento de Ciências Exatas, Universidade de São Paulo (USP/ESALQ) – werlleson.nascimento@usp.br

² Departamento de Ciências Exatas, Universidade de São Paulo (USP/ESALQ) – thiagronomo@usp.br

³ Departamento de Ciências Exatas, Universidade de São Paulo (USP/ESALQ) – clarice.demetrio@usp.br

⁴ Departamento de Ciências Exatas, Universidade de São Paulo (USP/ESALQ) – ctsdias@usp.br



Análise Estatística da Composição Química de extrato de café verde

Hudson Alfredo Cardoso Silva ¹

Diana Cardoso Silva ²

Ana Carolina Campana Nascimento ¹

Moysés Nascimento ¹

Fábio Júnior Moreira Novaes ²

Resumo

Este trabalho investigou a composição química e as propriedades bioativas de extratos produzidos a partir de grãos de café Arábica em diferentes estágios de maturação (verde, cereja e boia) e submetidos a variados níveis de torra (média, clara, escura e muito escura), buscando identificar componentes que contribuem para essas atividades. Por meio de análises por cromatografia a gás acoplado a espectrometria de massas (CG-EM) e avaliações em modelo animal, quantificou-se e correlacionou-se os componentes à atividade bioativa através do método de análise multivariada de Componentes Principais (PCA), evidenciando o potencial dos extratos. Os resultados permitiram identificar quais componentes químicos dos extratos a base de café que estavam associados a maior eficácia dos diferentes tratamentos, sendo promissores para aplicações terapêuticas.

Palavras-chave: café verde; propriedades bioativas; componentes principais; análise multivariada

¹ Laboratório de Inteligência Computacional e Aprendizados Estatístico (LICAE), Departamento de Estatística, Universidade Federal de Viçosa (UFV), Viçosa/MG – hudson.cardoso@ufv.br; ana.campana@ufv.br; moysesnascim@ufv.br

² Departamento de Química, Universidade Federal de Viçosa (UFV), Viçosa/MG – diana.cardoso@ufv.br; fabio.novaes@ufv.br

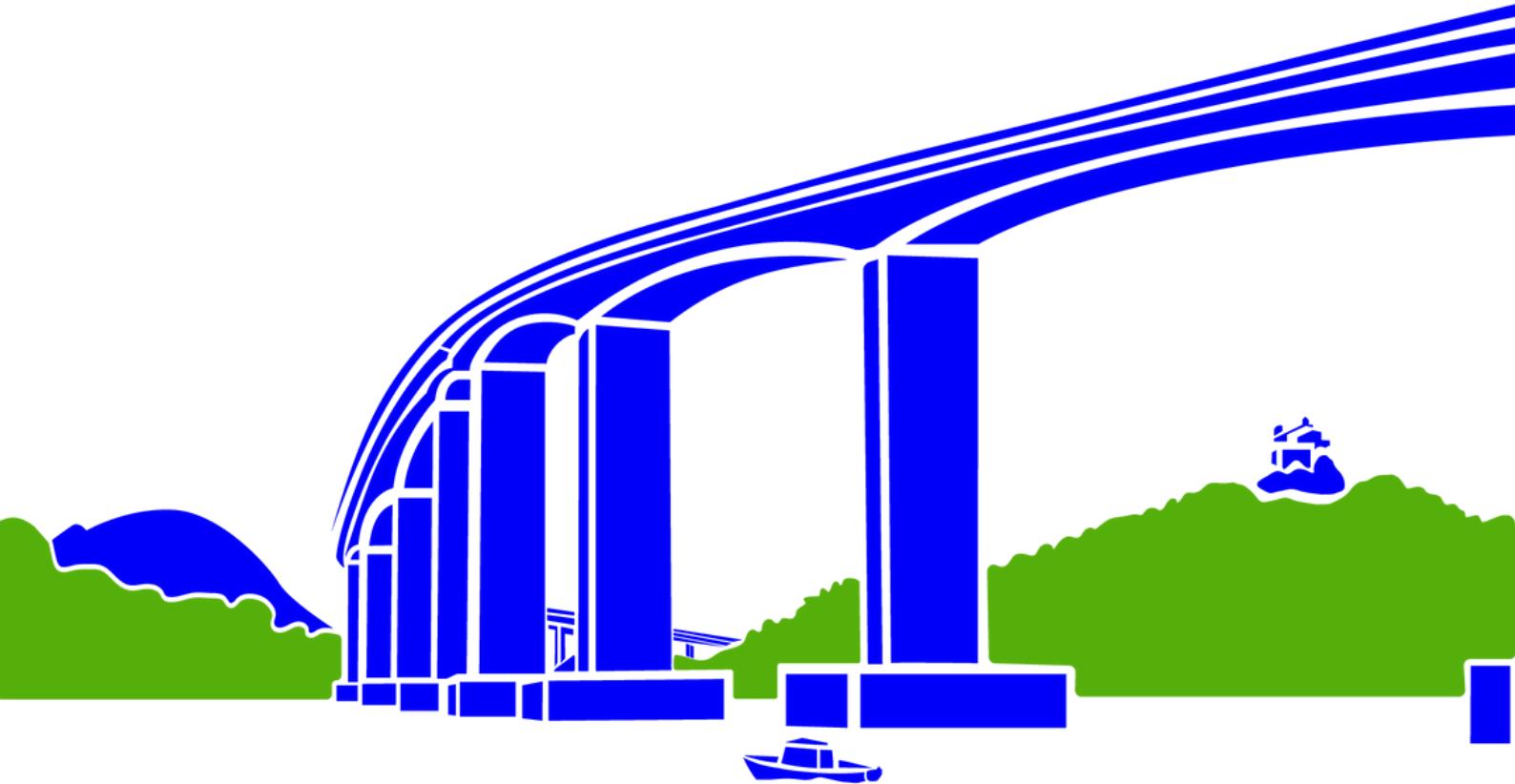
VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Pôsters

Estatística

Computacional





UTILIZAÇÃO DE REDES NEURAIS CONVOLUCIONAIS PARA CLASSIFICAÇÃO DE IMAGENS SAR

Jorge Antonio Silva Santos;¹ Jodavid de Araújo Ferreira ²

Resumo

Radares de abertura sintética (SAR) são conhecidos como instrumentos ativos de sensoriamento remoto utilizados para obtenção de imagens de uma determinada área geográfica e uma das vantagens é que esses sinais podem penetrar nuvens, chuvas e coberturas vegetais; ou seja, eles podem operar em quase todas as condições climáticas (van Zyl, 2011). As imagens geradas por sistemas SAR, conhecidas como imagens SAR, são obtidas através emissão e recepção dos sinais, estes que podem ser no formato linear e em diferentes direções, ou seja, os sinais podem ser emitidos com as seguintes combinações: HH, HV, VH e VV (H e V são representações para as polarizações horizontais e verticais, respectivamente). Apesar das vantagens discutidas previamente, as imagens SAR são afetadas por um ruído chamado speckle (Goodman, 1975). Este sinal ruído gera uma aparência granular nas imagens, criando uma dificuldade na interpretação e análise. Com o objetivo de extração de informações a respeito da cobertura do solo, classificadores para dados SAR têm sido propostos. Neste estudo, foram utilizados para classificação de dados SAR, duas arquiteturas de redes neurais: (i) a arquitetura totalmente conectada (fully connected) e (ii) a rede neural convolucional (Convolutional Neural Network- CNN) Yang et al. (2019). A imagem SAR utilizada nesse estudo é uma região de São Francisco (EUA) e representa 3 áreas bem definidas, floresta, oceano e urbano. Foram criados modelos de redes neurais fully connected e CNN utilizando como dados de entrada os 3 canais de polarização HH, HV e VV, e modelos utilizando um único canal de polarização como informação da imagem para os modelos. Métricas de avaliação de desempenho de classificação como acurácia, Matthews correlation coefficient- MCC, coeficiente foram calculadas e realizada uma discussão comparativa entre os modelos.

Palavras-chave: SAR ; Redes Neurais; CNN; Sensoriamento remoto; Classificação.

¹Departamento de Estatística, UFPE - jorge.silvasantos@ufpe.br

²Departamento de Estatística, UFPE - jodavid.ferreira@ufpe.br



Desenvolvimento de uma Aplicativo Shiny para Mineração de Texto

Wellington de Lara Martins ¹ Anderson Soares de Oliveira ² Lia Hanna Martins Morita
³

Resumo

O crescente volume de dados textuais disponíveis em diversas fontes, tem gerado uma demanda significativa por ferramentas que facilitem a análise e a extração de informações a partir desses textos. A mineração de texto, um campo do processamento de linguagem natural (PLN), oferece um conjunto de técnicas para a análise de dados textuais. No entanto, a complexidade envolvida na implementação dessas técnicas pode ser um desafio para muitos pesquisadores, cuja formação não inclua aspectos mais aprofundados de programação ou ciência de dados. O desenvolvimento de um aplicativo interativo em Shiny apresenta-se como uma solução relevante e oportuna. O Shiny é uma ferramenta de fácil uso que permite a criação de aplicações web diretamente a partir do R. A proposta desse aplicativo é preencher a lacuna existente entre a necessidade de análises textuais avançadas e a acessibilidade dessas técnicas. O aplicativo desenvolvido permite que os usuários realizem a leitura, o pré-processamento, a visualização (nuvem de palavras) e análise de sentimentos em textos. A escolha por focar nessas funcionalidades é estratégica, pois essas etapas constituem o núcleo de muitas análises textuais. Com este aplicativo, espera-se democratizar o acesso às técnicas de mineração de texto, facilitando a pesquisa e promovendo a inovação em diversas áreas.

Palavras-chave: Mineração de Texto; Processamento de Linguagem Natural; Shiny; Análise de Sentimentos; Nuvem de Palavras.

¹ Departamento de Estatística, ICET, UFMT/Campus Cuiabá-MT - wellington.martins@sou.ufmt.br

² Departamento de Estatística, ICET, UFMT/Campus Cuiabá-MT - andersoncso@ufmt.br

³ Departamento de Estatística, ICET, UFMT/Campus Cuiabá-MT - profaliaufmt@gmail.com



skewlmm: An R Package for Fitting Skewed and Heavy-Tailed Linear Mixed Models

Fernanda L. Schumacher;¹ Larissa A. Matos² Victor H. Lachos³

Resumo

Longitudinal data are commonly analyzed using linear mixed models, which, for mathematical convenience, usually assume that both random effect and error follow normal distributions. However, these restrictive assumptions may result in a lack of robustness against departures from the normal distribution and invalid statistical inferences. Schumacher et al. (2021) developed a flexible extension of linear mixed models considering the scale mixture of skew-normal class of distributions from a frequentist point of view, accommodating skewness and heavy tails, and the robust model formulation accounts for a possible within-subject serial dependence by considering some useful dependence structures. This work presents the R package **skewlmm**, which implements the method proposed by Schumacher et al. (2021) and provides a user-friendly tool to fit robust linear mixed models to longitudinal data, including model-fit tests, residual analyzes, and plot functions to support model selection and evaluation. Two data sets and a synthetic example are analyzed to illustrate the methodology and software implementation.

Palavras-chave: longitudinal data analysis; outliers; robust models; skewness.

Agradecimentos: O presente trabalho foi realizado com apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brasil. Processo nº 2020/16713-0.

Referências

Schumacher, F. L., V. H. Lachos, and L. A. Matos (2021). Scale mixture of skew-normal linear mixed models with within-subject serial dependence. *Statistics in Medicine* 40(7), 1790–1810.

¹The Ohio State University

²Universidade Estadual de Campinas - larissam@unicamp.br

³University of Connecticut



ExtremeTS: um pacote em linguagem de programação R para determinação do tamanho do bloco de extremos

Arthur Fellipy Pereira Silva ¹

Gilberto Rodrigues Liska ²

Marta Cristina Marjotta-Maistro ³

Resumo

Com o objetivo de facilitar a análise dos dados coletados, foi desenvolvido um pacote na linguagem R, preliminarmente chamado de *ExtremeTS*, para analisar valores extremos da taxa de câmbio brasileira. Durante o período analisado, foi elaborado um código de uma função em R capaz de organizar séries temporais, extrair valores máximos em períodos específicos e gerar gráficos para análise visual. O código também implementa testes estatísticos, como o Mann-Kendall, para detecção de tendência, e o Ljung-Box, para verificar a independência dos dados. Além disso, o pacote acompanha funções específicas para fornecer a possibilidade de salvar a série de dados organizados conforme customizado pelo usuário, bem como os resultados dos testes estatísticos. Para ilustração do pacote criado, bem como suas funções, foi utilizada uma série temporal de dados econômicos referentes à taxa de câmbio em dólar disponibilizado pelo Banco Central compreendendo o período de 1985 a 2025. Ressalta-se que o pacote desenvolvido poderá ser utilizado em outras séries temporais desde que contenha a informação de data (com formato dd/mm/aaaa) para o respectivo valor da variável a ser analisada. Os resultados apontam o sucesso no uso das funções do pacote proposto para a análise de eventos extremos, ampliando seu impacto na pesquisa acadêmica e no setor financeiro. Atualizações futuras incluem a otimização do pacote, a aplicação da metodologia a novos dados econômicos e a avaliação do desempenho do modelo em previsões futuras.

Palavras-chave: Linguagem R; Modelagem estatística; Teoria de Valores Extremos.

¹ Discente da graduação em Engenharia Agronômica, CCA-UFSCar – arthurfps@estudante.ufscar.br

² DTAiSER, CCA-UFSCar – gilbertoliska@ufscar.br

³ DTAiSER, CCA-UFSCar – marjotta@ufscar.br



AUTOMATIZANDO A ANÁLISE DE REGRESSÃO LINEAR NO RSTUDIO: UM APLICATIVO INTERATIVO EM SHINY

Matheus Nunes Paiva;¹ Eveliny Barroso da Silva²

Resumo

Este trabalho propõe o desenvolvimento de um aplicativo interativo em Shiny via linguagem R para automatizar a análise de regressão linear, técnica amplamente utilizada em diversas áreas. A análise de regressão linear pode ser complexa e requer conhecimento em estatística e programação, o que pode ser um obstáculo para muitos usuários. Além disso, lidar com grandes conjuntos de dados e múltiplas variáveis pode ser demorado. O aplicativo proposto visa resolver esses desafios, permitindo análises intuitivas e sem codificação. Com este aplicativo, os usuários podem inserir seu banco de dados, selecionar as variáveis independentes e a dependente, e realizar análises de regressão linear. Isso torna a técnica mais acessível, economizando tempo e esforço. O foco passa a ser a interpretação dos resultados, não a implementação do modelo. Para testes, será usado um banco de dados de consumo de energia da empresa *DAEWOO Steel Co., Ltd.* A variável resposta é o consumo de energia e as covariáveis incluem dióxido de carbono e potência relativa da corrente. O aplicativo permite a inserção de dados em Excel, apresenta os dados em tabelas, realiza análises descritivas, regressões simples e múltiplas, seleção de modelos e diagnóstico.

Palavras-chave: análise de regressão linear; seleção de modelos; linguagem R; RStudio; pacote *Shiny*.

¹Departamento de Estatística, UFMT - matheuspaiva2019@hotmail.com

²Departamento de Estatística, UFMT - eveliny.silva@ufmt.br



REDE BAYESIANA BASEADA NA DISTRIBUIÇÃO \mathcal{G}_I^0 COM PARÂMETROS FUZZY PARA CLASSIFICAÇÃO DE IMAGENS SAR

LEANDRO LUZ;¹ JODAVID FERREIRA²

Resumo

A classificação é uma das técnicas de modelagem no conjunto de ferramentas para a área de dados (?). Existem várias abordagens para realizar uma classificação, e cada algoritmo possui uma abordagem diferente em sua metodologia. No entanto, todos compartilham um objetivo final comum: classificar os dados em categorias ou classes distintas. Com o objetivo de extração de informações a respeito da cobertura do solo, classificadores para dados SAR têm sido propostos. As imagens geradas por sistemas SAR (*Synthetic Aperture Radar* - Radar de Abertura Sintética), conhecidas como imagens SAR, são imagens resultantes entre uma interação da energia emitida por um radar com o objeto em estudo e a aparência da imagem é influenciada pela forma e textura do alvo (?). Os sistemas SAR utilizam o processamento de iluminação coerente na formação da imagem e, dessa forma, as imagens SAR são afetadas por um ruído chamado *speckle*. Este sinal ruído gera uma aparência granular nas imagens, criando uma dificuldade na interpretação e análise. O *speckle* impõe aos dados SAR uma natureza multiplicativa tal que as intensidades SAR não se comportem como a lei gaussiana. Assim, estudos com dados SAR requerem métodos de processamentos de imagens específicos, baseados nas propriedades físicas de formação das imagens SAR. Neste trabalho é proposto a distribuição \mathcal{G}_I^0 (?) com parâmetros fuzzy (?), e uma rede Bayesiana baseada na distribuição proposta para classificação de imagens SAR.

Palavras-chave: Probabilidade Fuzzy; Imagens SAR; Classificação; Naive Bayes; Distribuição \mathcal{G}_I^0 .

References

¹Departamento de Estatística, UFPE - leandro.albuquerque@ufpe.br

²Departamento de Estatística, UFPE - jodavid.ferreira@ufpe.br



Análise de taxas máximas de câmbio do dólar por meio do aplicativo *UniMaC* no R

Laura de Aro Rocha ¹ Gilberto Rodrigues Liska ² Arthur Fellipy Pereira Silva³
Marta Cristina Marjotta-Mastro ⁴

Resumo

A taxa de câmbio é uma das variáveis macroeconômicas mais importantes e influencia fortemente as relações comerciais. Variações bruscas dessa taxa podem acarretar em abruptas mudanças no cenário econômico, como escassez de produtos e serviços, desvalorização da moeda, aumento da inflação, entre outros. O objetivo foi investigar o comportamento estocástico das taxas de câmbio de venda do dólar no Brasil, utilizando a Teoria de Valores Extremos (TVE). Para tal, foi utilizada uma série temporal de dados econômicos referentes à taxa de câmbio em dólar disponibilizado pelo Banco Central compreendendo o período de 1985 a 2025. Os dados foram analisados por meio do aplicativo em *Shiny R* nomeado *UniMaC*. Os resultados indicaram que a distribuição Generalizada de Valores Extremos (GVE) apresentou melhor ajuste em todos os meses. Em termos probabilísticos, os meses de setembro a novembro apresentaram as maiores probabilidades de ocorrerem taxas de câmbio superiores a 6% e espera-se que num período médio de 50 anos ocorra uma taxa superior a 6 dólares nesses meses. Por outro lado, as menores probabilidades correm nos meses de janeiro, fevereiro e junho. A abordagem estatística utilizada pode ser útil para investidores, economistas e formuladores de políticas públicas. Ainda, o aplicativo UniMaC tornou a análise pela TVE fácil, intuitiva e informativa.

Palavras-chave: Probabilidade de ocorrência; pacote Shiny; Aplicativo Web, Mercado cambial; Linguagem R

¹ Discente do Bacharelado em Biotecnologia do Centro de Ciências Agrárias, Universidade Federal de São Carlos, Araras - SP, Brasil – laura.rocha@estudante.ufscar.br

² Docente do Departamento de Tecnologia Agroindustrial e Socioeconomia Rural do Centro de Ciências Agrárias, Universidade Federal de São Carlos, Araras - SP, Brasil - gilbertoliska@ufscar.br

³Discente da Engenharia Agronômica do Centro de Ciências Agrárias, Universidade Federal de São Carlos, Araras - SP, Brasil – arthurfps@estudante.ufscar.br

⁴ Docente do Departamento de Tecnologia Agroindustrial e Socioeconomia Rural do Centro de Ciências Agrárias, Universidade Federal de São Carlos, Araras - SP, Brasil - marjotta@ufscar.br



GenomicLand: uma ferramenta gratuita e intuitiva para predição e associação genômica baseada em R

Camila Ferreira Azevedo¹

Moysés Nascimento²

Vitor Cunha Fontes³

Marcos Deon Vilela de Resende⁴

Cosme Damião e Cruz⁵

Resumo

O GenomicLand é um software gratuito desenvolvido para estudos de predição e associação genômica, baseado na linguagem R por meio do pacote Rshiny. Ele oferece uma interface gráfica amigável, permitindo o uso por pesquisadores que não possuem familiaridade com comandos de programação. O software é compatível com os sistemas operacionais Windows e Linux, está disponível em inglês e pode ser baixado gratuitamente pelo site <https://www.licae.ufv.br/genomicland/>. Além disso, o GenomicLand é capaz de processar bases de dados fenotípicos e genômicos, o que o torna apropriado para diferentes aplicações em programas de melhoramento genético. O GenomicLand implementa metodologias estatísticas baseadas em modelos mistos, inferência bayesiana, métodos de redução de dimensionalidade e algoritmos de inteligência artificial, que são aplicadas principalmente para resolver problemas de multicolinearidade, alta dimensionalidade e distribuição não normal de características fenotípicas. Estudos de predição e associação genômica ampla tem desempenhado um papel essencial nos programas de melhoramento, contribuindo para o aumento da acurácia na predição de valores genéticos, a aceleração na identificação de indivíduos superiores e a redução do intervalo de gerações. Nesse contexto, o GenomicLand surge como uma ferramenta gratuita e acessível, que disponibiliza tecnologias de ponta para pesquisadores que atuam na área de genética e melhoramento de plantas e animais.

Palavras-chave: estatística; inteligência artificial, marcadores moleculares; biometria.

¹ Departamento de Estatística, Universidade Federal de Viçosa, Viçosa – MG, camila.azevedo@ufv.br

² Departamento de Estatística, Universidade Federal de Viçosa, Viçosa – MG, moysesnascim@ufv.br

³ Pesquisador independente – vitorenfontes@gmail.com

⁴ Embrapa Café, Departamento de Engenharia Florestal, Universidade Federal de Viçosa, Viçosa – MG – marcos.deon@gmail.com

⁵ Departamento de Biologia Geral, Universidade Federal de Viçosa, Viçosa – MG, cdcrus@ufv.br



Análise Comparativa entre Métodos de Estimação para a Distribuição Maxwell Reparametrizada

Arthur Henrique da Rocha Hintz;¹ Leonardo Prior Migliorini;² Lucas de Bona Sartor;³
Thiago Alexandre Nascimento de Andrade;⁴

Resumo

Na estatística aplicada, quando se objetiva produzir inferências a partir de um modelo, a precisão assume papel fundamental. De fato, estimadores não viesados, testes de hipótese mais poderosos e intervalos de confiança mais estreitos conduzem a resultados mais precisos, os quais podem traduzir-se em maior segurança e economia de recursos, entre outros benefícios, nas mais diversas áreas de aplicação. Este trabalho tem como foco o estudo comparativo de métodos de estimação pontual em modelos probabilísticos. Especificamente, comparam-se diferentes técnicas de estimação aplicadas à distribuição de Maxwell reparametrizada, na qual o parâmetro original é expresso em função da média. Avaliamos o desempenho de quatro métodos: Máxima Verossimilhança (MLE), Produto Máximo de Distâncias (MPS), Momentos (MM) e Anderson-Darling (AD). Adicionalmente, aplicamos a correção de viés via reamostragem bootstrap a todos os estimadores considerados. O desempenho dos estimadores foi avaliado por meio de simulações de Monte Carlo, com observações geradas de forma pseudoaleatória utilizando o método da inversão, amplamente reconhecido na literatura especializada. O processo de simulação foi conduzido sob diferentes configurações experimentais, considerando-se distintos tamanhos amostrais e valores do parâmetro. A avaliação baseou-se em métricas consagradas para esse fim, a saber: viés relativo, erro quadrático médio, erro absoluto médio e a análise da consistência dos estimadores à medida que o tamanho amostral aumentava. Os resultados indicam que o MM apresentou, em geral, a melhor precisão, enquanto o MPS se mostrou o menos preciso nas configurações avaliadas. Não obstante, todos os estimadores demonstraram-se assintoticamente não viesados e consistentes, o que reforça sua adequação para a estimação do parâmetro reparametrizado.

Palavras-chave: Bootstrap; Distribuição Maxwell; Métodos de Estimação; Monte Carlo; Reparametrização.

¹Curso de Bacharelado em Estatística, UFSM - arthur_hintz@acad.ufsm.br

²Curso de Bacharelado em Estatística, UFSM - leonardo.migliorini@acad.ufsm.br

³Curso de Bacharelado em Estatística, UFSM - lucassartor100@hotmail.com

⁴Departamento de Estatística, UFSM - thiago_andrade@ufsm.br



Modelos Relacionais e Grafos: Um Estudo Comparativo Sobre Eficiência Computacional

Gabriel Müller Bittencourt;¹ Cristian Pessatti dos Anjos;¹ Anderson Luiz Ara Souza;¹

Resumo

Com o avanço do volume, variedade e velocidade dos dados, característicos da era do *Big Data*, tem exposto limitações dos métodos tradicionais de armazenamento baseados em estruturas tabulares e bancos relacionais. Embora amplamente utilizados, esses modelos enfrentam desafios crescentes à medida que os dados se tornam mais conectados, heterogêneos e de rápida expansão. Este trabalho propõe uma análise comparativa entre métodos tradicionais de armazenagem de dados — centrados em tabelas e relacionamentos — e abordagens baseadas em grafos, que oferecem uma estrutura nativa para representar relações complexas e conectadas, permitindo consultas otimizadas por meio de travessias entre nós. Essa estrutura também facilita a incorporação de hierarquias irregulares, relações de muitos-para-muitos e redes com graus de conexão variáveis. Além disso, a ciência de dados orientada a grafos evoluiu para incorporar técnicas como engenharia de atributos baseada na topologia da rede, embeddings estruturais, detecção de comunidades, predição de links e aprendizado nativo em grafos. O objetivo é avaliar aspectos como custo computacional, expressividade de modelos, facilidade de iterações, escalabilidade e acurácia em tarefas típicas da Ciência de Dados. Ao fazer essa comparação, pretende-se evidenciar as vantagens e limitações de cada paradigma, contribuindo para a escolha adequada de estratégias conforme a natureza dos dados e dos objetivos analíticos.

Palavras-chave: Grafos; Big Data; Custo computacional; Armazenamento de dados; Métodos tradicionais.

¹UFPR, Curitiba - mullerbittencourt@ufpr.br

¹UFPR, Curitiba - cristian.anjos@ufpr.br

¹UFPR, Curitiba - ara@ufpr.br



Processamento de Dados de Reabilitação Física e Neurofuncional

Victor Young Keun Kwon;¹ Diego Roberto Colombo Dias ²

Resumo

Este estudo tem como objetivo analisar dados de reabilitação física e neurofuncional provenientes do banco de dados ReBase, utilizando técnicas de análise descritiva e exploratória para otimizar sua infraestrutura e funcionalidades. Os dados, capturados por sensores de movimento como o Kinect, incluem variáveis numéricas, categóricas e temporais relacionadas ao movimento corporal de pacientes pós-AVC. Foram realizadas etapas de limpeza, identificação de outliers e tratamento de valores ausentes, seguidas de análises univariadas e bivariadas, incluindo histogramas, boxplots e correlações de Pearson. Os resultados evidenciaram diferenças significativas na amplitude de movimento dos joelhos, sugerindo assimetrias motoras compatíveis com sequelas de AVC. As análises indicaram uma forte correlação linear (0,92) entre as angulações dos joelhos, sugerindo movimentos sincronizados durante a atividade sentada. Além disso, a aplicação de técnicas de interpolação permitiu corrigir descontinuidades temporais causadas por períodos de descanso, melhorando a representação dos dados. A pesquisa contribui para a aplicação de métodos estatísticos na reabilitação neuromotora, destacando o potencial do banco de dados para suportar análises avançadas e intervenções personalizadas.

Palavras-chave: Análise de dados, Reabilitação neuromotora, Estatística descritiva, Interpolação, Movimento corporal.

¹Departamento de Estatística, UFES - victor.kwon@edu.ufes.br

²Departamento de Estatística/Programa de Pós-graduação em Informática, UFES - diego.dias@ufes.br

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Pôsters

**Estatística Aplicada em Ciências
Sociais Aplicadas, Demografia e
Estatísticas Públicas**





Cooperative Statistics in Brazil: A Classification Proposal using Machine Learning models Aligned with International Standards

Ricardo Theodoro ¹ André Gustavo Pereira ² Davi Rogério de Moura Costa ³

Elisangela Vaz ⁴

Abstract

Statistics pertaining to cooperatives in any nation play a pivotal role in comprehending the dynamics of this sector and formulating public policies aimed at ameliorating labor conditions and addressing wage disparities in specific regions. These statistics encompass vital information such as the number of active cooperatives, economic activities, geographical distribution, and typologies. The standardization of such data across nations facilitates identification of potential challenges, and exchange of comparative insights. Presently, Brazil lacks a classification system for its cooperatives that adheres to international standards. To address classification issues prevalent across various domains, the field of statistics has been pioneering numerous techniques tailored for application in machine learning algorithms, including KNN and Random Forest. By integrating these techniques with the EGA, this study aims to classify Brazilian cooperatives into the four international typologies: user, producer, worker, and multi-stakeholder. This classification hinges on the meticulous description of all economic activities undertaken by cooperatives, leveraging data sourced from the Brazilian Federal Revenue alongside their Economic Activities Code. The efficacy of this approach was yielding an impressive accuracy rate of 92.7%. These findings hold promise for advancing international cooperative studies, facilitating cross-country comparisons, and fortifying the cooperative ecosystem in Brazil and beyond.

¹ Universidade de São Paulo, São Paulo – rtheodoro@usp.br

² Universidade Federal Rio Grande do Norte , Natal – andregustavocampospereira@gmail.com

³ Universidade de São Paulo, São Paulo – drmouracosta@usp.br

⁴ Universidade de Campinas, Campinas – elisangela.vaz@unesp.br



Caracterização da dependência espacial da dengue em Minas Gerais: possíveis causas associadas aos padrões espaciais

Gean Pereira Damaceno ¹; Giovani Festa Paludo ²; João Paulo Assis Bonifácio ³; Júlio Sílvio de Sousa Bueno Filho ⁴; Douglas Mateus da Silva ⁵; João Domingos Scaloni ⁶

Resumo

A dengue é uma arbovirose que causa sérios problemas de saúde pública no Brasil, contudo, ainda não foram completamente elucidadas todas as causas das recorrentes epidemias. O estudo de sua estrutura espacial pode ajudar a linkar os possíveis processos causadores dos padrões no espaço. Com base nisso, o objetivo do presente estudo foi: (i) descrever a estrutura espacial da ocorrência da dengue em Minas Gerais e; (ii) apontar as possíveis causas associadas aos padrões encontrados. Foram obtidos os números absolutos de ocorrência da dengue em 2023 na Secretaria de Saúde (MG) e calculada a incidência (casos/população*100000). Foi calculado o índice de Moran global e local (LISA) para verificar, respectivamente, a existência da autocorrelação espacial global e agrupamentos de autocorrelação espacial local. As estimativas do índice global para incidência e casos absolutos foram, respectivamente, 0,276 e 0,104, ambos estatisticamente significativos ao nível de 5% de confiança, evidenciando a presença de autocorrelação positiva. Tal resultado indica que municípios vizinhos tendem a apresentar valores similares de incidência. Embora a maioria dos municípios não tenha apresentado autocorrelação espacial local, foram identificados agrupamentos com autocorrelação espacial tanto para o número de casos absolutos quanto para a incidência, porém em regiões distintas. Além disso, foram descritos os principais padrões observados e discutidas algumas possíveis causas para esses padrões encontrados, como exemplo o efeito transbordamento e a estrutura das rodovias. Um próximo passo é obter medições das variáveis das hipóteses sugeridas e testar quais têm efeito significativo na distribuição espacial da dengue.

Palavras-chave: padrões de distribuição; epidemia; dependência espacial; descrição de padrões; *Aedes aegypti*.

¹Graudando do curso de graduação em Estatística com ênfase em Ciência de Dados, UFLA, MG - geandamaceno030@gmail.com

²Doutorando do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, UFLA, MG - gfpaludo@gmail.com

³Graudando do curso de graduação em Estatística com ênfase em Ciência de Dados, UFLA, MG - jpad.27@hotmail.com

⁴Departamento de Estatística da Universidade Federal de Lavras, Lavras, MG - jssbueno@ufla.br

⁵Departamento de Estatística da Universidade Federal de Lavras, Lavras, MG - douglas.mateus@ufla.br

⁶Departamento de Estatística da Universidade Federal de Lavras, Lavras, MG - scalon@ufla.br



INFOVIS: UM PORTAL DE VISUALIZAÇÃO DE DADOS PARA A ANÁLISE DA SEGURANÇA PÚBLICA NA BAHIA

Italo Estrela de Souza Sá;¹ Larissa Reis Sousa Santos²

Resumo

O InfoVis é uma plataforma de visualização de dados da Bahia, criada pela Superintendência de Estudos Econômicos e Sociais da Bahia (SEI), projetada para oferecer estatísticas e indicadores de diversas áreas, inclusive a da segurança pública. O InfoVis de Segurança Pública foi desenvolvido por meio de uma parceria entre a SEI e a Secretaria da Segurança Pública da Bahia (SSP), envolvendo a Avaliação Operacional da Superintendência de Gestão Integrada da Ação Policial (SIAP) e a Coordenação de Documentação e Estatística Policial (CDEP), da Polícia Civil do Estado da Bahia. Esse painel fornece estatísticas detalhadas sobre crimes no estado, como homicídios, latrocínios e feminicídios, além de crimes contra o patrimônio, incluindo informações referentes ao número e ao perfil das vítimas, à distribuição temporal dos crimes e ao mapeamento geográfico das ocorrências. Com isso, o presente trabalho tem como objetivo principal apresentar o InfoVis de Segurança Pública, por meio da metodologia tecnológica, destacando suas excepcionalidades e funcionalidades, mostrando como essa ferramenta pode ser importante para auxiliar a implementação de projetos e políticas públicas baseadas em evidências, bem como possibilitar o monitoramento e a avaliação de seus resultados pelos gestores e a sociedade em geral. O painel InfoVis de Segurança Pública consolidou-se como ferramenta estratégica, capaz de fornecer dados robustos e detalhados que auxiliam tanto na formulação quanto na avaliação de políticas públicas. Além de oferecer informações específicas para cidadãos, pesquisadores e autoridades, esse recurso mostra-se fundamental para a tomada de decisões governamentais baseadas em dados, promovendo uma governança digital eficiente.

Palavras-chave: informação visual; segurança pública; Bahia; estatística; SEI.

¹Graduado em Estatística e mestrando em Matemática com concentração em Estatística pela Universidade Federal da Bahia (UFBA). Técnico da Superintendência de Estudos Econômicos e Sociais da Bahia (SEI). <https://orcid.org/0009-0000-6269-0279>. italosa@sei.ba.gov.br

²Graduada em Estatística e mestrandona Matemática com concentração em Estatística pela Universidade Federal da Bahia (UFBA). Técnica da Superintendência de Estudos Econômicos e Sociais da Bahia (SEI). <https://orcid.org/0009-0007-4676-6158>. larissasantos@sei.ba.gov.br



Construção de pirâmides etárias para a região de Ouro Preto

- um estudo demográfico

Michel Lima de Oliveira ¹

Fernando Luiz Pereira de Oliveira ²

Resumo

Acerca das intensas transformações demográficas, ora em curso, o município de Ouro Preto - MG encontra-se diante de fenômenos desafiadores, como o encolhimento e envelhecimento populacional. Espera-se que esse cenário, causado pela baixa natalidade e maior expectativa de vida do brasileiro, impacte negativamente o ritmo de consumo na economia municipal, uma vez que a parcela economicamente ativa tende a diminuir. Com base nessas conclusões, este artigo analisou diversos indicadores de caráter multidimensional, abrangendo áreas demográficas, sociais, econômicas e de saúde para Ouro Preto - MG e regiões comparáveis, objetivando apontar soluções políticas e econômicas às lacunas demográficas identificadas. Para o processamento e interpretação das extensas bases de dados, foi empregada a linguagem R, garantindo precisão e eficiência nas análises.

Palavras-chave: transformação demográfica; dinâmica populacional; indicadores sociais; indicadores econômicos; Ouro Preto.

¹ Departamento de Estatística, Universidade Federal de Ouro Preto - michel.lima@aluno.ufop.edu.br

² Departamento de Estatística, Universidade Federal de Ouro Preto - fernandoluiz@ufop.edu.br



Classificação e agrupamento de crimes relacionados à violência contra mulheres: um estudo baseado em *Doc2vec*

Maria de Fátima Magalhães de Souza Andrade;¹ Lia Hanna Martins Morita;² Anderson Castro Soares de Oliveira³

Resumo

Word embedding é uma técnica que transforma palavras em representações vetoriais contínuas, permitindo que algoritmos de aprendizado de máquina processem informações linguísticas com mais eficiência. O *doc2vec* é um modelo de PLN (Processamento de Linguagem Natural) desenvolvido como extensão do *word2vec*, que cria representações vetoriais contínuas de palavras usando duas arquiteturas: *continuous bag of words* (CBOW) e *continuous skip-gram*. O CBOW estima uma palavra central a partir de palavras do contexto, enquanto o *skip-gram* prevê palavras do contexto a partir de uma palavra central. Diferente do *word2vec*, que gera vetores apenas de palavras, o *doc2vec* utiliza as arquiteturas PV-DM (*Distributed Memory Model of Paragraph Vectors*) e PV-DBOW (*Distributed Bag of Words of Paragraph Vectors*). A primeira cria vetores de documentos inteiros, capturando relações sintáticas, semânticas e contextuais. A segunda, semelhante ao *skip-gram*, ignora palavras de contexto e prevê palavras do documento amostradas aleatoriamente. Utilizando a arquitetura PV-DM, treinou-se um modelo *doc2vec* para criar vetores contínuos de narrativas de boletins de ocorrência de crimes relacionados à violência contra mulheres para classificação e agrupamento de documentos, detecção de sentenças que revelem os tipos de violência mais prevalentes e as associações entre palavras, sentenças e documentos que permitam identificar melhor os crimes.

Palavras-chave: *Doc2vec*; PV-DM; PLN; classificação; crimes.

¹Departamento de Estatística, ICET, UFMT/Campus Cuiabá – Mato Grosso - maria.andrade1@sou.ufmt.br

²Departamento de Estatística, ICET, UFMT/Campus Cuiabá – Mato Grosso - profaliaufmt@gmail.com

³Departamento de Estatística, ICET, UFMT/Campus Cuiabá – Mato Grosso - anderson.oliveira@ufmt.br



A trajetória dos menores de idade nos Jogos Olímpicos

Júlia Zorzo Ferreira;¹ Alexandre Schleder Gonçalves Kobus;² Amanda Merian Freitas Mendes;³

Resumo

Os Jogos Olímpicos são o mais renomado evento multiesportivo mundial desde sua estreia em Atenas, no ano de 1896. Embora a maior parte dos atletas seja composta por adultos, historicamente, crianças e adolescentes também participam da competição. O destaque midiático com relação à quantidade de jovens olímpicos a cada edição, no entanto, pode induzir à falsa impressão de que esse número vem aumentando, quando, na verdade, ao analisar essa quantidade ao longo dos anos, observou-se uma tendência decrescente. Pensando nisso, por meio de um estudo descritivo dos dados dos Jogos Olímpicos, este trabalho tem como objetivo principal analisar a participação de atletas menores de idade, ainda em fase de desenvolvimento físico e mental, nos Jogos de Inverno e Verão, entre 1896 e 2022. Além disso, procurou-se identificar em quais modalidades há maior representatividade de atletas menores de idade, verificando seus desempenhos e países de origem. Por fim, buscou-se explorar os possíveis motivos históricos, burocráticos e políticos por trás do aumento e subsequente declínio, ao longo do tempo, da quantidade de jovens atletas no evento.

Palavras-chave: Jogos Olímpicos; Análise Exploratória; Dados Esportivos.

¹Departamento de Estatística, UFPR - zorzojulia@ufpr.br

²Departamento de Estatística, UFPR - alexsgkobus@gmail.com

³Departamento de Estatística, UFPR - amandamerian@ufpr.br



Modelos de Mistura para Dados Longitudinais Aplicados ao Estudo de Bullying Escolar

Eduardo Figueiredo Batista ^{1,3}; Leila Denise Alves Ferreira Amorim ^{2,3}

Resumo

Modelos de Mistura são técnicas estatísticas voltadas para a identificação e caracterização de subpopulações latentes de uma população através de uma amostra. Exemplos de aplicação desses modelos realizadas na literatura incluem estudos sobre comportamento ético, vulnerabilidade social, e traços psicológicos de equipes de funcionários num contexto empresarial, dentre outros. Nesse trabalho, o enfoque é dado a dois modelos de mistura para análise de dados de natureza longitudinal: a análise de classes latentes com medidas repetidas(RMLCA, em inglês) e a análise de transição latente(LTA, em inglês). A RMLCA consiste do modelo tradicional de análise de classes latentes aplicado a medidas repetidas e tem como objetivo diferenciar as classes latentes de acordo com seus comportamentos nos pontos do tempo. Já a LTA, que também é adequada para a análise de dados longitudinais, tem como objetivo principal estudar as transições que podem ocorrer entre os estados latentes ao longo dos pontos do tempo, de modo que pode ser utilizada para estudar a mudança de comportamento de indivíduos com relação a uma variável que não pode ser diretamente observada. Esse trabalho objetiva sistematizar a literatura relacionada à análise de classes latentes com medidas repetidas e à análise de transição latente, ilustrando estas metodologias com uso de dados reais de estudo conduzido com escolares do ensino médio dos Estados Unidos, que objetivou examinar mudanças de estados latentes relacionadas ao comportamento de bullying escolar. Para as análises são utilizados dois pacotes do *software* estatístico **R**: poLCA (para RMLCA) e LMest (para LTA).

Palavras-chave: análise de transições latentes; análise de classes latentes; dados longitudinais; modelos de mistura.

¹Bolsista de Iniciação Científica (PIBIC-UFBA/CNPq), Bacharelado em Estatística, Instituto de Matemática e Estatística, Universidade Federal da Bahia, Salvador - edubatis@gmail.com

²Departamento de Estatística, Instituto de Matemática e Estatística, Universidade Federal da Bahia, Salvador - leiladen@ufba.br [Apoio recebido da Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB, termo de outorga PP0021/2023) e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, termo de outorga 403465/2023-0)]

³Laboratório de Inferência Causal e Aplicações - LInCA, <https://linca.ufba.br/>



ANÁLISE ESTATÍSTICA DO PERFIL DOS ESTUDANTES EGRESSOS DA UFMT APÓS 10 ANOS DA LEI DE COTAS

Aline Martins Soares ¹

Eveliny Barroso da Silva ²

Anderson Castro Soares de Oliveira ³

Resumo

A Lei das Cotas nº 12.711, sancionada em 2012 no Brasil, foi um avanço significativo para a democratização do acesso ao ensino superior, garantindo uma reserva de vagas em universidades públicas para estudantes de escolas públicas, negros, pardos e indígenas. Sua implementação gerou debates sobre seus efeitos no ingresso e desempenho acadêmico dos beneficiados. Após uma década, os impactos da lei podem ser avaliados sob várias perspectivas, como inserção no mercado de trabalho, formação acadêmica e condições de permanência e conclusão dos cursos. Essas análises são essenciais para determinar se a política tem cumprido seu propósito de promover igualdade de oportunidades. Este estudo propõe uma análise estatística do perfil dos estudantes egressos da Universidade Federal de Mato Grosso, comparando cotistas e não cotistas. O objetivo é examinar as condições desses ex-alunos no mercado de trabalho e na vida acadêmica, permitindo uma visão mais clara sobre os resultados práticos da Lei das Cotas ao longo dos anos. Resultados parciais mostram que 42,6% dos egressos eram cotistas com idade média de 33,89 anos e Renda per capita média de R\$ 979,18. **Palavras-chave:** Lei de Cotas, Egressos UFMT, Cotistas, Mercado de Trabalho.

¹ Departamento de Estatística, ICET, UFMT/Campus Cuiabá – alinesmartines@gmail.com

² Departamento de Estatística, ICET, UFMT/Campus Cuiabá – eveliny.silva@ufmt.br

³ Departamento de Estatística, ICET, UFMT/Campus Cuiabá – anderson.oliveira@ufmt.br



Perfil sociodemográfico das vítimas fatais de trânsito no Brasil (1997–2023)

Ana Beatriz da Silva Marques;¹ Bruno Wavrzenczak;² Anderson Ara.³

Resumo

A mortalidade no trânsito representa um grave problema de saúde pública no Brasil, com impactos sociais e econômicos relevantes. Diversos estudos apontam que as vítimas de acidentes de trânsito não se distribuem de maneira homogênea entre os diferentes grupos populacionais, refletindo desigualdades sociais e regionais. Este estudo tem como objetivo descrever e analisar o perfil sociodemográfico das vítimas fatais de acidentes de trânsito ocorridos no país no período de 1997 a 2023. Para isso, foi utilizada uma base de dados composta por registros individuais de óbitos relacionados ao trânsito, contendo informações sobre idade, sexo, etnia, escolaridade, local de ocorrência e data do evento. A metodologia proposta envolve técnicas de estatística descritiva, visualizações gráficas, análise temporal e testes de associação entre variáveis. A pesquisa busca identificar padrões e desigualdades associadas à mortalidade no trânsito, bem como possíveis mudanças no perfil das vítimas ao longo do tempo. Acredita-se que os resultados obtidos podem contribuir para o entendimento das vulnerabilidades sociodemográficas relacionadas aos acidentes de trânsito e subsidiar o desenvolvimento de políticas públicas mais eficazes na prevenção de mortes e na promoção da segurança viária.

Palavras-chave: Acidentes de Trânsito; Mortalidade; Perfil Sociodemográfico; Estatística Descritiva; Segurança Viária.

¹Departamento de Estatística, UFPR - beatrizda@ufpr.br

²Departamento de Estatística, UFPR - brunowaurzenczak@ufpr.br

³Departamento de Estatística, UFPR - ara@ufpr.br



Impacto das Mortes por Psicoativos no Espírito Santo: Uma Análise Sociodemográfica (2013–2022)

Caroline de Oliveira Costa ¹; Franco Lovatti Souza Pinto ²; Iane Maria Santos Martins ³;

Matheus Monteiro Alves Melo ⁴; Samuel Martins de Medeiros ⁵; Agatha Sacramento Rodrigues⁶

Resumo

O uso de substâncias psicoativas é um problema de saúde pública que impacta a mortalidade e qualidade de vida da população. Compreender o perfil sociodemográfico das vítimas e os impactos regionais é essencial para formular políticas públicas eficazes. Este estudo analisou bases de dados governamentais de 2013 a 2022 para investigar os efeitos das drogas no Espírito Santo. Utilizando o Sistema de Informação sobre Mortalidade (SIM) e a linguagem R, foram examinadas variáveis como idade, gênero, raça, escolaridade e estado civil, a partir dos códigos da Classificação Internacional de Doenças (CID) relacionados a psicoativos. Os dados foram analisados sob quatro perspectivas: mortes totais e por psicoativos no Brasil e no Espírito Santo. Indicadores compararam os óbitos por drogas no estado com dados nacionais ajustados à população local. Os resultados indicaram que os idosos (60+) lideravam as mortes gerais, enquanto os adultos predominavam nas mortes relacionadas aos psicoativos. Homens apresentaram taxas de mortalidade significativamente superiores às mulheres. O estudo destaca a necessidade de melhorar a qualidade dos dados estaduais, visto que algumas categorias apresentaram um percentual de registros faltantes próximo a 50%. Além disso, foi constatado que o Espírito Santo teve óbitos por psicoativos acima da média ajustada nacionalmente, evidenciando um impacto desproporcional. Esses resultados reforçam a importância de investigações aprofundadas e da criação de um perfil sociodemográfico detalhado para aprimorar políticas públicas voltadas ao combate dos efeitos do uso de substâncias psicoativas.

Palavras-chave: Mortalidade 1; Substâncias psicoativas 2; Saúde Pública 3; Perfil sociodemográfico 4; Espírito Santo 5.

¹ Departamento de Estatística, Universidade Federal do Espírito Santo, Vitória, ES – Caroline.deoli.2004@gmail.com

² Departamento de Estatística, Universidade Federal do Espírito Santo, Vitória, ES - franco.lovatti@gmail.com

³ Departamento de Estatística, Universidade Federal do Espírito Santo, Vitória, ES - ianemartins99@gmail.com

⁴ Departamento de Estatística, Universidade Federal do Espírito Santo, Vitória, ES - matheus.melo@edu.ufes.br

⁵ Departamento de Informática, Universidade Federal do Espírito Santo, Vitória, ES - Samuel.martins7@hotmail.com

⁶ Departamento de Estatística, Universidade Federal do Espírito Santo, Vitória, ES - agatha.rodrigues@ufes.br



Apreensões de Drogas e Mortalidade no Espírito Santo: Uma Abordagem Socioespacial

Samuel Martins de Medeiros¹; Caroline de Oliveira Costa²; Franco Lovatti Souza Pinto³; Iane Maria Santos Martins⁴; Matheus Monteiro Alves Melo⁵; Agatha Sacramento Rodrigues⁶

Resumo

O uso de psicoativos no Espírito Santo é resultado de uma complexa interação de fatores sociais, econômicos e geográficos, que influenciam tanto os índices de apreensões quanto a ocorrência de óbitos. Este projeto de Iniciação Científica investiga essas inter-relações ao analisar dados oficiais de apreensões e registros de mortalidade, considerando as peculiaridades dos diferentes municípios do estado. A pesquisa utiliza metodologias estatísticas e modelos de análise espacial para identificar correlações entre as variáveis estudadas e mapear áreas de maior vulnerabilidade, com atenção especial à capital, Vitória, e sua comparação com os indicadores nacionais. Ao integrar informações de fontes diversas, o estudo busca compreender como determinantes socioeconômicos e estruturais afetam a dinâmica do consumo de psicoativos, contribuindo para o agravamento dos impactos na saúde pública e na segurança. Os resultados preliminares apontam que regiões com altos índices de apreensões também podem apresentar elevadas taxas de mortalidade, evidenciando a necessidade de estratégias integradas e ações preventivas. Assim, a pesquisa enfatiza o papel das políticas públicas intersetoriais na mitigação desses problemas, fornecendo subsídios para intervenções mais eficazes e o aprimoramento dos mecanismos de vigilância e apoio à população vulnerável.

Palavras-chave: Mortalidade 1; Substâncias psicoativas 2; Análise espacial 3; Apreensões de drogas 4; Espírito Santo 5.

¹ Departamento de Informática, Universidade Federal do Espírito Santo, Vitória, ES - Samuel.martins7@hotmail.com

² Departamento de Estatística, Universidade Federal do Espírito Santo, Vitória, ES - Caroline.deoli.2004@gmail.com

³ Departamento de Estatística, Universidade Federal do Espírito Santo, Vitória, ES - franco.lovatti@gmail.com

⁴ Departamento de Estatística, Universidade Federal do Espírito Santo, Vitória, ES - ianemartins99@gmail.com

⁵ Departamento de Estatística, Universidade Federal do Espírito Santo, Vitória, ES - matheus.melo@edu.ufes.br

⁶ Departamento de Estatística, Universidade Federal do Espírito Santo, Vitória, ES - agatha.rodrigues@ufes.br



Desigualdades Regionais na Fecundidade Adolescentes: um Estudo Longitudinal do Semiárido Brasileiro (2019–2023)

Tania Barbosa Tomaz¹

Everlane Suane de Araújo da Silva²

Maria Andrelly Matos de Lima³

Resumo

A confiança nos dados dos sistemas de registros de estatísticas vitais e o acompanhamento da fecundidade das adolescentes podem ser citados como metas dos Objetivos de Desenvolvimento Sustentável para 2030. Nesse sentido, teve-se como objetivo analisar os níveis da fecundidade entre adolescentes de 15 a 19 anos no Semiárido brasileiro, no período de 2019 a 2023, com base nos registros do Sistema de Informações sobre Nascidos Vivos (SINASC). Trata-se de uma pesquisa ecológica e longitudinal que abrangeu os 1.473 municípios definidos pela SUDENE. A metodologia incluiu o pré-processamento dos microdados, análise da qualidade dos registros, cálculo e correção das Taxas Específicas de Fecundidade (TEF) e a avaliação da distribuição espacial dos indicadores TEF. Os resultados demonstraram que os dados do SINASC apresentaram elevado grau de completude, com preenchimento superior a 90%, mesmo nos anos críticos da pandemia de COVID-19. Observou-se tendência de queda nas TEF adolescentes em todas as regiões do país, embora com ritmos diferenciados: a Região Norte manteve os maiores níveis, enquanto Sudeste e Sul apresentaram os menores. No Semiárido brasileiro, os valores médios foram inicialmente inferiores à média nordestina, mas, após correção, superaram-na, alcançando 0,0449 em 2023, acima da média nacional de 0,0391. A distribuição espacial revelou elevada heterogeneidade intraestadual, sem padrões homogêneos, com destaque para áreas do sul da Bahia e parte do Sudeste com menores taxas. Os achados reforçaram a importância de análises desagregadas e da aplicação de métodos de correção para estimativas precisas, com o intuito de subsidiar políticas públicas adaptadas às realidades locais.

Palavras-chave: Fecundidade; Adolescente; Semiárido brasileiro; SINASC; ODS.

¹ Departamento de Estatística, Universidade Federal da Paraíba, Campus I, João Pessoa-PB – taniabtomaz@yahoo.com.br

² Programa de Pós-Graduação em Modelos de Decisão e Saúde, Universidade Federal da Paraíba, Campus I, João Pessoa-PB – everlane.suane@academico.ufpb.br

³ Programa de Pós-Graduação em Modelos de Decisão e Saúde, Universidade Federal da Paraíba, Campus I, João Pessoa-PB – andrellymatos@gmail.com



Análise de aspectos qualitativos e sociodemográficos relacionados ao incômodo causado por gases odorantes em ambiente urbano industrializado

Franciele Ribeiro Cavalcante ¹

Valdério Anselmo Reisen ²

Milena Machado de Melo ³

Jane Meri Santos ⁴

Resumo

Este estudo investiga os fatores qualitativos que influenciam a percepção e o incômodo causados por gases odorantes em uma área urbana industrializada na Grande Vitória-ES, Brasil. Foram aplicadas pesquisas de opinião em dois períodos distintos, representando variações sazonais das condições meteorológicas. As análises envolveram testes qui-quadrado e regressão logística binária multivariada. Os resultados revelam que a percepção de odores está fortemente associada a experiências negativas no cotidiano, como a necessidade de manter janelas fechadas. A maioria dos respondentes relatou incômodo elevado (“muito” ou “extremamente”) frente aos odores, especialmente os provenientes da indústria siderúrgica e de estações de tratamento de esgoto. Diferenças significativas entre os períodos de coleta foram observadas quanto ao perfil dos participantes e à percepção da poluição do ar. Variáveis como gênero, percepção de poeira e perda de visibilidade, frequência de odores e histórico de problemas respiratórios mostraram-se estatisticamente associadas ao aumento da chance de incômodo intenso. Ao abordar as respostas subjetivas à poluição odorífera — um tema ainda pouco explorado nas avaliações ambientais —, este estudo contribui para aproximar dados técnicos das percepções sociais. Seus achados oferecem subsídios valiosos para o planejamento de políticas públicas e ações de mitigação por parte de empresas e órgãos ambientais, com foco na saúde e qualidade de vida da população exposta.

Palavras-chave: Poluição odorífera; Percepção ambiental; Incômodo por odores; Regressão logística; Qualidade do ar.

¹ Departamento de Engenharia Ambiental, Universidade Federal do Espírito Santo (UFES), Vitória – franciele97rc@gmail.com

² Departamento de Engenharia Ambiental, Universidade Federal do Espírito Santo (UFES), Vitória – valderioanselmoreisen@gmail.com

³ Instituto Federal do Espírito Santo (IFES), Campus Guarapari – milas2@gmail.com

⁴ Departamento de Engenharia Ambiental, Universidade Federal do Espírito Santo (UFES), Vitória – jmerisantos@yahoo.com.br



Modelagem Espacial das Mortes Violentas Intencionais (MVI) nos Municípios de Pernambuco

Marcela Silva de Araujo;¹ Melissa Wendy Ribeiro Vicente;³ Valdeline de Paula Mequelino Ferreira⁶ Marcelo Ângelo Cirillo;⁶

Resumo

As Mortes Violentas Intencionais (MVI), que incluem homicídios dolosos, feminicídios, latrocínios e mortes decorrentes de intervenção policial, constituem um importante indicador da violência letal no Brasil. Este estudo teve como objetivo analisar a distribuição espacial das MVI nos municípios do estado de Pernambuco entre os anos de 2016 a 2024. Os dados utilizados foram obtidos junto à Secretaria de Defesa Social de Pernambuco (SDS-PE). As análises foram conduzidas por meio de técnicas de estatística espacial, com ênfase na autocorrelação espacial global e local. O Índice de Moran Global indicou autocorrelação espacial positiva ($I = 0,10; p - valor = 0,01$), sugerindo a presença de dependência espacial entre os municípios, ainda que de forma moderada. A análise local, por meio dos Indicadores Locais de Associação Espacial (LISA), identificou clusters estatisticamente significativos do tipo "alto-alto" nos municípios da Região Metropolitana do Recife, como Recife, Olinda, Jaboatão dos Guararapes e Cabo de Santo Agostinho. Esses agrupamentos indicam áreas com altas taxas de MVI vizinhas a municípios com padrão semelhante. Por outro lado, agrupamentos "baixo-baixo" foram observados em regiões do interior do estado, caracterizadas por menor densidade populacional e menores índices de violência. Os resultados evidenciam a heterogeneidade espacial da violência letal em Pernambuco e apontam para a importância de políticas públicas direcionadas territorialmente. Estratégias de prevenção da violência devem considerar os padrões de distribuição espacial, priorizando intervenções focalizadas nas áreas críticas identificadas, com ações integradas entre os setores de segurança, saúde e desenvolvimento social.

Palavras-chave: Estatística Espacial; Índice de Moran; Homicídio; Feminicídio; Segurança Pública.

¹Universidade Federal de Lavras, Departamento de Estatística, Lavras, MG - marcela.araujo1@estudante.ufla.br

³Universidade Federal de Lavras, Departamento de Estatística, Lavras, MG - melisawendyrv@outlook.com

⁶Universidade Federal de Lavras, Departamento de Estatística, Lavras, MG - valdeline.ferreira1@estudante.ufla.br

⁶Universidade Federal de Lavras, Departamento de Estatística, Lavras, MG - macufla@ufla.br



Óbitos Classificados com Causas *Garbage* no Brasil

Everlane Suane de Araújo da Silva ¹

José Roberto Ramos de Assis ²

Resumo

Mundialmente, existe um esforço para aumentar a cobertura e melhorar a qualidade dos dados sobre mortalidade. A existência de sistemas de registros civil e de estatísticas vitais confiáveis pode ser citada como meta dos Objetivos de Desenvolvimento Sustentável para 2030, da ONU. É imprescindível considerar os Códigos *Garbage* (códigos mal definidos ou pouco específicos) quando se trata da qualidade da informação sobre as causas de morte nos estados do Brasil. O aumento do número de óbitos no período pandêmico no Brasil chamou bastante atenção, sendo o pico atingido em 2021, 1.015.350 óbitos para o sexo masculino e 816.616 para o sexo feminino. Assim, teve-se como objetivo identificar e mensurar os óbitos classificados com Códigos *Garbage* nos espaços geográficos dos estados brasileiros, no período de 2019 a 2023. No Painel de Monitoramento da Mortalidade por Causas Básicas Inespecíficas ou Incompletas (*Garbage Codes*) foram obtidas as séries temporais, permitindo calcular os percentuais de óbitos por Códigos *Garbage* no Brasil e verificar o alarmante resultado que 49% dos óbitos ocorridos no ano pandêmico de 2021 foram classificados com Códigos *Garbage*. Transcorrida a pandemia de COVID-19, em 2023, valores inferiores foram observados no país e o comportamento do indicador ficou muito similar ao que antecedeu a pandemia, aproximadamente 32%. O estudo contribuiu para avanços no conhecimento da realidade da qualidade da informação da notificação dos óbitos nas unidades federativas do Brasil e trouxe à tona a necessidade de cautela ao utilizar os dados de óbitos para a construção de importantes indicadores.

Palavras-chave: Causas de morte; Códigos *Garbage*; Qualidade dos dados; Brasil; ODS.

¹ Programa de Pós-Graduação em Modelos de Decisão e Saúde, Universidade Federal da Paraíba, Campus I, João Pessoa-PB – everlane.suane@academico.ufpb.br

² Departamento de Estatística, Universidade Federal da Paraíba, Campus I, João Pessoa-PB – joserrdeassis@gmail.com



Aplicação do Modelo Nominal da Teoria da Resposta ao Item (TRI) na Avaliação do SAEB 2023 – 5º Ano

André Felipe Zilio da Silva¹

Resumo

Este trabalho tem como objetivo analisar os itens da avaliação em larga escala do INEP – o Sistema de Avaliação da Educação Básica (SAEB) – aplicada a estudantes do 5º ano do Ensino Fundamental, por meio do Modelo de Resposta Nominal (MRN) da Teoria da Resposta ao Item (TRI). Diferentemente do modelo tradicional de 3 parâmetros (ML3), que classifica as respostas como corretas ou incorretas, o MRN permite a análise de todas as alternativas de resposta, incluindo os distratores (respostas incorretas). Isso possibilita uma compreensão mais aprofundada dos padrões de escolha dos estudantes. Além disso, o modelo é capaz de estimar o possível gabarito do item com base nos padrões de resposta observados, o que pode revelar discrepâncias em relação ao gabarito oficial. Em 2023, eram esperados mais de 2,6 milhões de estudantes do 5º ano para a aplicação da avaliação, que abrangeu as áreas de Língua Portuguesa e Matemática, com 22 itens em cada prova, totalizando 44 itens por caderno. Para este estudo, optou-se por analisar 2 cadernos dos 21 diferentes. A aplicação do MRN revelou algumas particularidades, como a identificação de possíveis gabaritos divergentes do oficial e níveis de dificuldade distintos dos esperados, destacando o potencial do modelo para aprimorar a avaliação e a interpretação dos resultados educacionais.

Palavras-chave: Avaliação de larga-escala; SAEB; TRI; Modelo de Resposta Nominal; INEP.

¹ Secretaria Municipal da Educação – Curitiba – andrefzs@hotmail.com.br



Estudo do Padrão Espacial dos Crimes de Roubo e Furto nas Áreas Urbanas de Cuiabá e Várzea Grande – MT

Elizabeth Cristina Röpck;¹ Lia Hanna Martins Morita;² Elianara Martins de Almeida;³
Anderson Castro Soares de Oliveira⁴

Resumo

Os crimes de roubo e furto têm sido uma preocupação constante em Cuiabá e Várzea Grande, demandando esforços contínuos das autoridades de segurança pública. Apesar das medidas adotadas para otimizar o policiamento e implementar ações preventivas, esses índices criminais persistem, afetando não apenas o patrimônio das vítimas, mas também o seu bem-estar emocional e psicológico. Nesse contexto, a análise de *hotspots* emerge como ferramenta essencial para direcionar os recursos disponíveis de forma mais eficaz, identificando áreas com maior incidência criminal e possibilitando a implementação de medidas específicas para cada região. Essas estratégias não só ajudam a dissuadir a atividade criminosa, mas também fortalecem a confiança da comunidade nas forças de segurança. Este trabalho consistiu em analisar o padrão espacial da distribuição dos crimes de roubo e furto consumados nas áreas urbanas de Cuiabá e Várzea Grande, no estado de Mato Grosso, visualizar a distribuição dos crimes em um mapa temático para identificar padrões e agrupamentos, delimitar áreas com características homogêneas utilizando a técnica de Kernel e investigar a interação entre os pontos das ocorrências de furto e roubo por meio das funções J e K. Os dados usados foram de ocorrências registradas no período de 1º de agosto de 2022 a 31 de julho de 2023. Foi realizada a análise com o estimador Kernel e as funções K e J para examinar a distribuição espacial dos crimes. A análise foi realizada no software R, utilizando o pacote estatístico *spatstat*. Os resultados incluíram a identificação de *hotspots* e a verificação do padrão da configuração espacial dos pontos de ocorrência dos crimes em Cuiabá e Várzea Grande. Foram estimados a densidade de Kernel e as funções K e J para verificar o padrão espacial dos dados e identificar áreas com maior densidade de crimes. Os resultados revelaram padrões de agrupamento dos crimes, com concentração nas regiões centrais de Cuiabá e Várzea Grande. A visualização dos mapas de Kernel confirmou esses agrupamentos, destacando as áreas de maior densidade de crimes.

Palavras-chave: Padrões Pontuais; Crimes; *Hotspots*; Mapa de calor; Funções de 1^a e 2^a ordem.

¹Policia Militar do Estado de Mato Grosso, Cuiabá MT - elizabete@pm.mt.gov.br

²Universidade Federal de Mato Grosso, Cuiabá MT - lia.morita@ufmt.br

³Universidade Federal de Mato Grosso, Cuiabá MT - elianara.almeida@ufmt.br

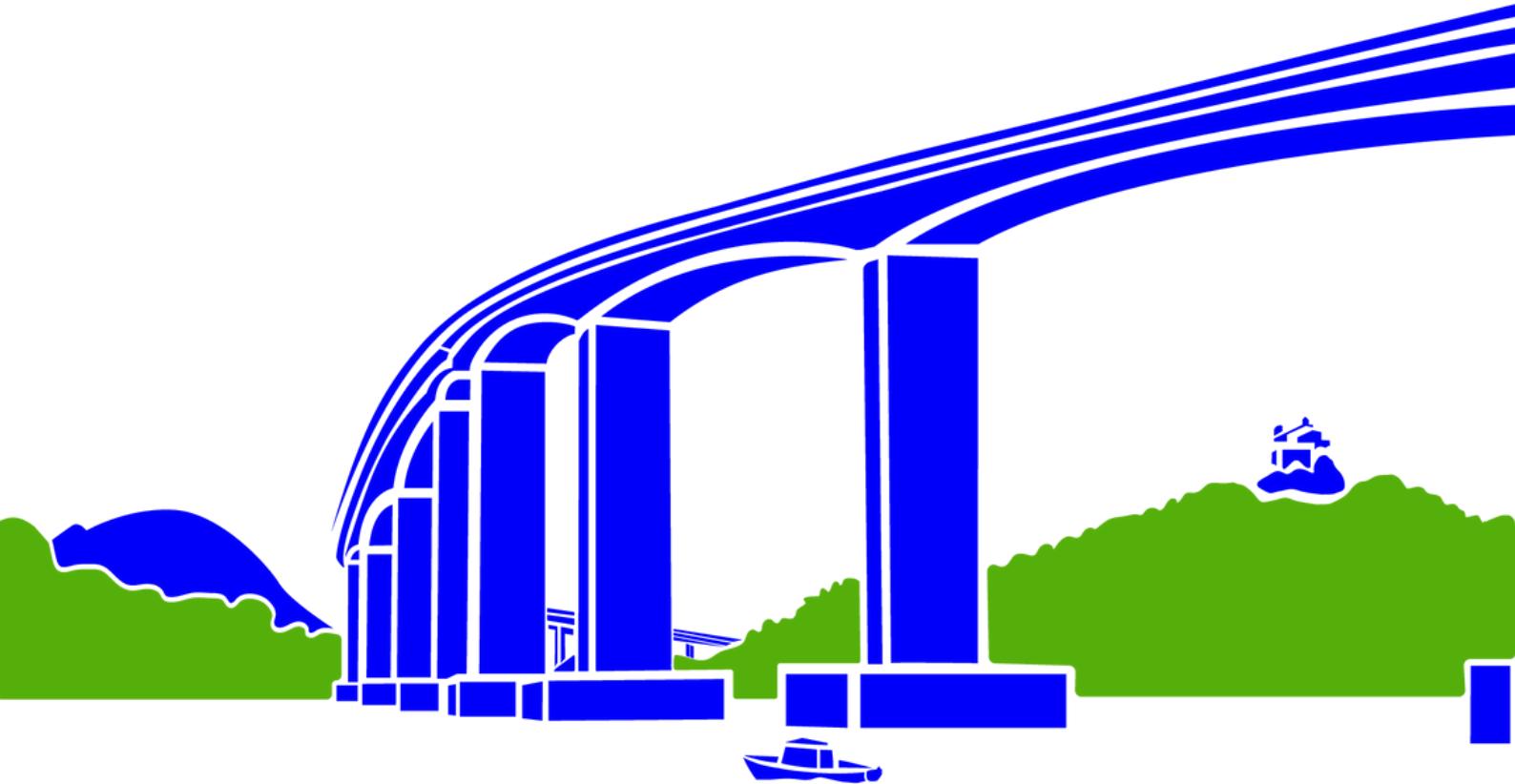
⁴Universidade Federal de Mato Grosso, Cuiabá MT - andersoncso@ufmt.br

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Pôsters

**Estatística Aplicada em
Ciências Médicas, Saúde e
Meio Ambiente**





Detecção de padrões de incidência da malária na província de Tete, Moçambique de 2016 a 2022

Abrantes João Afonso Mussafo;¹ Liciana Vaz de Arruda Silveira² José Silvio Govone³

Resumo

O estudo busca compreender a evolução anual da incidência da malária e identificar padrões locais de autocorrelação espacial. Os dados analisados são de rotina, coletados anualmente em 15 distritos e 155 unidades sanitárias da província de Tete. A metodologia empregada consistiu na análise exploratória de dados, utilizando estatística espacial para cada ano, com enfoque no Índice de Moran global e local (mapa LISA). Os resultados indicam a presença de autocorrelação espacial positiva no ano de 2021, com $I = 0,259$ e $p\text{-valor} = 0,0338$, evidenciando uma similaridade entre a incidência da malária e a localização espacial dos casos. Nos demais anos, os valores de $p\text{-valor} > 0,05$ indicam que a autocorrelação espacial não é estatisticamente significativa a 95% de significância. Além disso, observa-se que regiões com alta incidência de malária entre 2016 e 2018 e baixa incidência em 2022 estão circundadas por áreas com características semelhantes. Também foram identificadas regiões com padrões instáveis entre 2019 e 2021, caracterizadas pela presença de clusters do tipo alto-alto e baixo-baixo. Em conclusão, ao longo dos sete anos analisados, foram identificados agrupamentos espaciais de regiões com altas e baixas incidências de malária, sugerindo a necessidade de intervenções mais eficazes para mitigar a transmissão da doença.

Palavras-chave: Malária; incidência; análise espacial; I de Moran global; mapa LISA.

¹Departamento de Estatística e Biodiversidade,UNESP-Botucatu: a.mussafo@unesp.br

²Departamento de Estatística e Biodiversidade,UNESP-Botucatu:liciana.silveira@unesp.br

³Departamento de Estatística, Matemática Aplicada e Computação,IGCE, e Centro de Estudos Ambientais /UNESP - Rio Claro js.govone@unesp.br



Análise temporal entre grupos etários dos atendimentos ambulatoriais do SUS referentes à intoxicação por drogas, medicamentos e substâncias biológicas no Estado do Paraná (2018-2023)

Stéfane Lele Rossoni ¹

Diogo Francisco Rossoni ²

Resumo

Os atendimentos ambulatoriais do SUS referentes à intoxicação por drogas, medicamentos e substâncias biológicas (CID-10: T36 a T50) entre os anos de 2018 a 2023 no estado do Paraná, apresentaram uma tendência de crescimento apenas para a faixa etária de adultos (18 a 59 anos), apontado pelo teste de Mann-Kendall ($\tau = 0,55$; valor-p < 0,001). No entanto, quando as quantidades de atendimentos são ponderadas pela população das respectivas faixas etárias, o teste de ANOVA identificou que a média mensal de atendimentos de crianças/adolescentes e idosos por intoxicação é significativamente maior do que a dos adultos ($F = 10,61$; valor-p < 0,001). As linhas temporais dos registros de atendimentos indicam que os grupos de crianças/adolescentes e adultos apresentaram uma queda brusca nos atendimentos logo após o início da pandemia da Covid-19, já o grupo de idosos não acompanhou o mesmo comportamento, o qual começou a apresentar uma crescente no segundo semestre de 2019, atingindo um pico no mês de dezembro do mesmo ano e voltando a se estabilizar já a partir de janeiro de 2020, não sofrendo quedas ou picos com o início da pandemia.

Palavras-chave: Intoxicação; Séries Temporais; COVID-19; Análise de Variância; SUS.

¹ Programa de Pós-graduação em Ciências da Saúde, Universidade Estadual de Maringá – ste.tistics@gmail.com

² Programa de Pós-graduação em Bioestatística, Universidade Estadual de Maringá – dfrossoni@uem.br



Mudanças climáticas no Rio Grande do Sul e seus impactos econômicos

Lucas Oliveira da Silva ¹

Resumo

Este trabalho analisa os impactos econômicos das mudanças climáticas no estado do Rio Grande do Sul, com foco nas enchentes históricas de 2024. Através de modelos de regressão linear aplicados em séries temporais climáticas e econômicas desde 2004, busca-se identificar correlações entre variações pluviométricas e térmicas sobre o PIB estadual. A pesquisa utiliza dados do INMET e do DEE-RS, e aplica técnicas estatísticas no R e Python, incluindo análise de resíduos e testes estatísticos adequados para evidenciar tais impactos. Os resultados pretendem embasar políticas de mitigação e adaptação climática, evidenciando como eventos extremos/atípicos podem afetar a economia regional do estado, e, similarmente, outras economias.

Palavras-chave: séries temporais; correlações; mudanças climáticas.

¹ Departamento de Estatística (DEs) ..., UFSCar – lucasos@estudante.ufscar.br



Confiabilidade de Equipamentos Médicos: Métodos e Aplicações na Segurança Hospitalar

Bernardo Miguel Moreira Ramiro ¹

Paulo César de Resende Andrade ²

Resumo

A confiabilidade de equipamentos médicos é crucial para a segurança dos pacientes e a eficácia dos tratamentos, pois falhas podem resultar em consequências graves. A análise de confiabilidade envolve testes rigorosos, validação técnica e identificação de falhas, garantindo conformidade com normas regulatórias e prolongando a vida útil dos dispositivos. A Manutenção Centrada em Confiabilidade (MCC) destaca-se como uma abordagem estratégica ao priorizar ações preditivas e preventivas, reduzindo custos e aumentando a segurança dos equipamentos. Sua eficácia é ampliada pela integração com o Processo de Projeto de Equipamentos Médicos (PPEM), promovendo maior eficiência operacional e segurança hospitalar. Diversas metodologias são aplicadas para a identificação e correção de falhas, como a Análise dos Modos de Falha e seus Efeitos (FMEA), a Análise da Árvore de Falhas (FTA) e a Análise de Modos de Falha e seus Efeitos na Saúde (HFMEA), esta última utilizada na Maternidade Escola Januário Cicco (RN) para avaliação de um ventilador pulmonar. Além disso, ferramentas da qualidade, como a Folha de Verificação, o Diagrama de Pareto e o Diagrama de Ishikawa, auxiliam no monitoramento e aprimoramento da manutenção. Dessa forma, a análise de confiabilidade torna-se essencial na gestão de equipamentos médicos, assegurando conformidade com normas técnicas, segurança dos pacientes, aumento da durabilidade dos dispositivos e redução de custos operacionais no setor hospitalar.

Palavras-chave: Análise de falhas; Confiabilidade; Equipamentos médicos; Manutenção Centrada em Confiabilidade; Segurança do paciente.

¹ Engenharia Mecânica, Instituto de Ciência e Tecnologia, Universidade Federal dos Vales do Jequitinhonha e Mucuri, Diamantina/MG – bernardo.miguel@ufvjm.edu.br

² Instituto de Ciência e Tecnologia, Universidade Federal dos Vales do Jequitinhonha e Mucuri, Diamantina/MG – paulo.andrade@ict.ufvjm.edu.br



Previsão da ocupação hospitalar usando análise multivariada de séries temporais

Ediney José da Silva Magalhães Júnior ¹ Anderson Castro Soares de Oliveira ² Neuber José Segri ³ Lia Hanna Martins Morita ⁴

Resumo

Este estudo analisou as inter-relações entre variáveis fundamentais para a gestão hospitalar — Taxa de Ocupação, Tempo Médio de Permanência e Giro de Leitos — por meio da aplicação de um modelo de Vetores Autoregressivos (VAR) de ordem 3. O objetivo foi construir um modelo preditivo capaz de apoiar o planejamento estratégico da ocupação hospitalar. Os resultados revelaram que, embora o Tempo Médio de Permanência (Y_2) e o Giro de Leitos (Y_3) apresentem influência significativa nas projeções da Taxa de Ocupação (Y_1), os valores defasados da própria taxa de ocupação foram os principais determinantes nas previsões. Isso evidencia a importância das autocorrelações na dinâmica da série temporal em análise.

Além das relações estatísticas entre as variáveis, observou-se que fatores externos, não incluídos no modelo, também impactam de forma relevante o comportamento da taxa de ocupação. A análise reforça a necessidade de considerar o contexto operacional mais amplo dos hospitais, que envolve não apenas os serviços médicos, mas também setores como hotelaria, nutrição e logística. A compreensão aprofundada desses elementos permite uma melhor previsão da demanda por recursos físicos e humanos, promovendo maior eficiência e qualidade no atendimento. Conclui-se que modelos preditivos como o VAR oferecem um suporte valioso à tomada de decisão em ambientes hospitalares, sendo recomendada a ampliação futura do modelo com a inclusão de outras variáveis operacionais. Dessa forma, este trabalho contribui para o avanço das práticas de gestão baseadas em dados no setor da saúde.

Palavras-chave: Análise Preditiva, Modelagem Temporal, Aplicativo Interativo, Modelos ARIMA, Modelos SARIMA

¹Departamento de Estatística/ICET/UFMT, Cuiabá - ediney.junior@sou.ufmt.br

²Departamento de Estatística/ICET/UFMT, Cuiabá - anderson.oliveira@ufmt.br

³Departamento de Estatística/ICET/UFMT, Cuiabá - neuber.segri@ufmt.br

⁴Departamento de Estatística/ICET/UFMT, Cuiabá - lia.morita@ufmt.br



Caracterização dos casos de morte materna tardia no Brasil

Mariana Modenesi de Andrade

Nátnaly Adriana Jiménez Monroy

Resumo

A morte materna tardia define-se como o óbito obstétrico, direto ou indireto, que ocorre no puerpério tardio, ou seja, de 43 dias a um ano após o parto. Sendo um tema pouco explorado no Brasil, a escassez de referências evidenciou a necessidade de aprofundar o estudo sobre o impacto desses óbitos no contexto da saúde materna. A análise das mortes maternas tardias, não consideradas no cálculo da razão de mortalidade materna (RMM), evidenciou a grande quantidade de casos ignorados na análise dessa taxa. Notou-se que a maioria das mortes maternas tardias foi classificada como óbito materno desconsiderado, ou seja, casos com características de óbito materno que não são reconhecidos como tal pelo Ministério da Saúde, de forma que esses casos deixaram de ser considerados ao analisar a situação da saúde materna no Brasil. Observou-se aumento expressivo de mortes maternas indiretas na última década, enquanto os óbitos maternos diretos se estabilizaram nos últimos anos. A RMM tardia foi calculada para cada Unidade Federativa (UF) e o teste de Mann Kendall foi usado para verificar a tendência temporal. Além disso, foram produzidos mapas da RMM tardia no Brasil nos quais observou-se uma forte influência da pandemia, com alta variabilidade entre as UF.

Palavras-chave: Morte materna tardia. Saúde materna. Pandemia. Razão de Mortalidade Materna. Análise exploratória.



Construção de Redes Complexas de Interações entre SNPs

Uma Abordagem Computacional e Estatística

Rafaela Miglinski Lucca ¹

Andressa Cerqueira ²

Resumo

Desde o início dos anos 2000, as interações entre polimorfismos de nucleotídeo único (SNPs) têm sido investigadas em relação a condições de saúde complexas, visando elucidar mecanismos genéticos e preditores de risco. Neste estudo, empregamos técnicas de aprendizado de máquina para identificar interações entre SNPs associados a diabetes, utilizando dados do projeto ISA-Capital ("Estilo de vida, marcadores bioquímicos e genéticos como fatores de risco cardiometabólico: inquérito de saúde na cidade de São Paulo"). SNPs, variações de base única no genoma, representam a forma mais prevalente de diversidade genética, influenciando fenótipos e susceptibilidade a doenças. Foram aplicadas regressões lineares e logísticas para associações diretas entre SNPs e diabetes, complementadas por florestas de interação (interaction forests) para capturar efeitos epistáticos quantitativos e qualitativos. Dados genômicos ausentes foram imputados via cadeias de Markov e estimativas de fração de recombinação. Os resultados preliminares indicam a identificação de SNPs candidatos e redes de interação genética significativas, contribuindo para a compreensão dos mecanismos moleculares subjacentes ao diabetes.

Palavras-chave: Polimorfismos de Nucleotídeo Único, Redes de Interação Genética, Diabetes.

¹ Departamento de Estatística -Des/ ICMC, Usp- Ufscar- São Carlos – luccarafaelam@gmail.com
² Departamento de Estatística -Des , Ufscar- São Carlos – acerqueira@ufscar.br



Simulações sobre o Impacto de Parâmetros Genéticos e Ambientais na Variabilidade do Risco de Malária

Thais Sena de Paula Domingues ¹, Adèle Helena Ribeiro ², Júlia M Pavan Soler ¹,
Rodrigo Malavazi Corder ³, Marcelo Urbano Ferreira ³, Dominik Heider ²

Resumo

Como explicar que em regiões endêmicas de malária, como a Amazônia brasileira, algumas pessoas sofrem ataques recorrentes da infecção enquanto outras permanecem livres da doença durante a maior parte do tempo, é ainda um ponto em aberto na literatura. Acredita-se que esse padrão seja modulado por fatores genéticos de suscetibilidade, ambientais, socioeconômicos e de estilo de vida, que afetam a exposição e o acesso a recursos de prevenção e tratamento, além da virulência das linhagens parasitárias, configurando um fenômeno multifatorial. Neste trabalho, inspirados nos dados do projeto “Variação individual no risco de malária: causas e consequências em populações amazônicas” (Processo FAPESP nº 2022/11963-3), avaliamos o impacto de parâmetros genéticos e ambientais na heterogeneidade do risco individual de malária. Diferentes cenários de simulações de dados são propostos, variando a herdabilidade de fenótipos de interesse, padrões de agregação entre indivíduos devido a ambientes compartilhados, bem como características do tipo de moradia e da exposição dos indivíduos a ambientes de mata. Na análise dos dados gerados, são aplicados algoritmos de aprendizado de estrutura causal, como o Fast Causal Inference (FCI), visando identificar relações de causa e efeito entre variáveis.

Palavras-chave: Dados genômicos, estudos observacionais, populações heterogêneas, algoritmo de aprendizado de estrutura causal, recorrência de malária.

Apoio Recebido: O presente trabalho foi realizado com apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brasil. Processo nº 2022/11963-3.

¹ IME-USP, São Paulo, Brasil – pavan@ime.usp.br

² Universidade de Münster, Alemanha, – adele.helena@gmail.com

³ ICB-USP, São Paulo, Brasil – rodrigo.corder@usp.br



Clusterização para identificação da similaridade temporal e espacial da situação endêmica e epidêmica da dengue nos municípios do Paraná

Juliana Nascimento de Paula;¹ Dayana Telles Bertuzzi;² Eniuce Menezes³

Resumo

Dengue é uma arbovirose que impacta a saúde pública, especialmente em países tropicais e temperados. No ano epidemiológico de 2024, o Brasil registrou o maior número de casos notificados no mundo e o Paraná o terceiro estado brasileiro com maior incidência da doença. A identificação de padrões temporais e espaciais auxilia na prevenção e controle da doença. Neste estudo, foi realizada a clusterização de séries temporais para analisar a ocorrência de casos confirmados de dengue no Paraná com base tanto na situação endêmica histórica quanto epidêmica dos municípios. O algoritmo k-means para séries temporais foi empregado, e o número de clusters determinado pelo método Elbow. Foram analisados 218 municípios onde a dengue é endêmica, resultando em oito clusters para ambas as situações, endêmica e epidêmica. A clusterização na situação endêmica permitiu identificar grupos de municípios com picos sazonais na mesma semana epidemiológica. Historicamente, essa informação auxilia na previsão dos períodos de maior risco epidêmico. Abril foi identificado como o mês mais crítico, com 129 municípios (59,17%) atingindo o pico de casos entre as semanas 14 e 18 (31/03 a 04/05 aproximadamente). Os municípios com picos mais precoces (semanas 49, 52 e 5) concentraram-se predominantemente em uma faixa diagonal do mapa, do Noroeste ao Centro-Oeste do estado. Em 58 municípios, os picos endêmico e epidêmico coincidiram, enquanto nos demais houve um deslocamento temporal. Em 37 municípios, a epidemia ocorreu três semanas antes do esperado, indicando padrões distintos de disseminação e reforçando a importância do monitoramento contínuo para ações preventivas eficazes.

Palavras-chave: dengue; *clustering*; *k-means*; endemia; epidemia.

¹Programa de Pós-graduação em Bioestatística, UEM, Maringá - juliananp.es@gmail.com

²Programa de Pós-graduação em Bioestatística, UEM, Maringá - dayanatbertuzzi@gmail.com

³Departamento de Estatística, Programa de Pós-graduação em Bioestatística, UEM, Maringá - emenezes@uem.br



Análise Estatística dos Desastres Causados por Chuvas Intensas na Região Metropolitana de Pernambuco

Kaylane Sabino da Silva¹

Mylleni Silvia Chaves Durval¹

Fidel Ernesto Castro Morales²

Antonio Marcos Batista do Nascimento²

Daniele Tôrres Rodrigues³

Resumo

O estado de Pernambuco, situado na região nordeste do Brasil, é frequentemente afetado por chuvas intensas que desencadeiam desastres como alagamentos e deslizamentos de terra. Esses eventos têm grandes impactos socioambientais, levando à perda de vidas, destruição de moradias, interrupção de serviços essenciais e sérios problemas de saúde pública. Este estudo tem como objetivo principal analisar a frequência desses eventos extremos e seus impactos. Para isso, utilizamos uma base de dados jornalística, abrangendo os desastres ocorridos entre 2005 e 2021 na região metropolitana de Pernambuco, que inclui informações sobre localização, data de ocorrência, tipo de desastre e os principais impactos provocados. Por meio de uma metodologia quantitativa, realizamos uma análise descritiva utilizando a linguagem R. Foram calculadas medidas estatísticas como média, mediana e desvio padrão, essenciais para caracterizar a distribuição desses eventos. Além disso, geramos gráficos que ilustram a frequência dos desastres por estação do ano. Os resultados indicam que a maioria, 46,88% dos desastres corresponde a deslizamentos de terra, ocorrendo principalmente entre março e maio, período chuvoso na costa leste da região nordeste. Para análises futuras, pretendemos empregar metodologias mais complexas, com o objetivo de modelar esses eventos. Isso permitirá uma compreensão mais profunda dos padrões dos desastres e fornecerá um entendimento mais abrangente dos impactos causados pelas chuvas intensas no estado de Pernambuco.

Palavras-chave: Desastres naturais ; chuvas; deslizamento de terra; Estatística.

¹Departamento de Matemática, UFRN - kaylanesabino12@icloud.com

²Departamento de Estatística, UFRN

³Departamento de Estatística, UFPI



Estimação de valores faltantes em dados de poluição do ar utilizando imputação única

Rafael Jordane de Suza Oliveira;¹ Ana Júlia Alves Câmara ²

Resumo

A poluição do ar representa um grave problema de saúde pública e ambiental, particularmente em regiões urbanas e industriais. A ocorrência de dados faltantes nas séries temporais de monitoramento compromete a qualidade das análises e a eficácia de políticas de controle. Este estudo avalia métodos de imputação única (média, mediana, LOCF, NOCB e splines) para lidar com esse problema, comparando sua eficácia em diferentes cenários.

Foram simuladas séries temporais usando modelos ARMA (AR(1), AR(2) e ARMA(1,1)) com dados faltantes do tipo MAR (5% a 40%). A avaliação baseou-se no RMSE e em análises gráficas. Os resultados mostraram que métodos simples (média e mediana) tiveram desempenho inferior, especialmente em séries com alta autocorrelação, por não capturarem adequadamente os padrões temporais. A interpolação por splines destacou-se em séries de baixa a média complexidade, enquanto o NOCB foi mais eficaz em cenários complexos. A seleção do método adequado permite reduzir erros e melhorar a confiabilidade das análises. Esta pesquisa visa contribuir para o aprimoramento das análises ambientais, apoiando estratégias mais precisas de controle da poluição e proteção à saúde pública.

Palavras-chave: dados faltantes; poluição atmosférica; séries temporais; imputação única; modelos ARMA.

¹Departamento de Estatística, Universidade Federal do Espírito Santo, Vitória, ES - rafael.js.oliveira@edu.ufes.br

²Departamento de Estatística, Universidade Federal do Espírito Santo, Vitória, ES - ana.j.camara@ufes.br



Violência e Lesões Autoprovocadas em Adolescentes do Espírito Santo: Uma Análise dos Fatores Associados

Pedro de Brito Neto; Nátnaly Adriana Jiménez Monroy

Resumo

Este estudo investigou os fatores que influenciam a violência contra adolescentes no Espírito Santo, utilizando dados do Sistema de Informação de Agravos de Notificação (SINAN) registrados entre 2009 e 2021. A pesquisa adotou uma abordagem transversal e analítica, combinando análises exploratórias, testes estatísticos (como o qui-quadrado) e modelos logísticos para estimar as chances associadas aos diferentes tipos de violência, incluindo lesão autoprovocada, bem como agressões física, psicológica e sexual.

Os resultados indicaram que as adolescentes têm cerca de 5,8 vezes mais probabilidade de sofrer violência sexual em comparação aos meninos, especialmente na faixa etária de 10 a 13 anos. Em contraste, a violência física foi predominantemente observada entre os adolescentes do sexo masculino, que apresentaram aproximadamente 3 vezes mais chances de serem vítimas, com incidência maior nas idades mais avançadas. Embora os casos de violência psicológica tenham sido menos frequentes, os dados não apontaram diferenças significativas quanto ao perfil das vítimas.

No contexto da violência sexual, apesar de a maioria dos agressores ser desconhecida, uma parcela relevante dos registros identificou pais e padrastos como principais suspeitos. Além disso, a análise dos casos de lesão autoprovocada demonstrou uma prevalência maior entre as meninas, e a idade revelou-se um fator crucial, visto que adolescentes de 14 a 20 anos apresentam mais chances de se auto lesionar em comparação aos mais jovens.

Essas evidências reforçam a importância de considerar o sexo e a idade das vítimas ao identificar riscos e planejar ações de prevenção e intervenção, contribuindo para o desenvolvimento de políticas públicas mais efetivas na proteção dos adolescentes.

Palavras-chave: Violência contra adolescentes; Lesão autoprovocada; Modelo logístico; Modelo log-binomial; Epidemiologia; Sistemas de informação em saúde.



Análise de sobrevivência para avaliação dos processos demográficos de árvores da Mata Atlântica Brasileira

Marciel Lelis Duarte¹

Sebastião Martins Filho²

Carlos Moreira Miquelino Eleto Torres³

Resumo

As florestas fornecem múltiplas funções e serviços ecossistêmicos, incluindo o fornecimento de bens por meio da produção primária, regulação do clima por meio do sequestro de carbono e serviços hidrológicos. Dentre as florestas tropicais brasileiras, a Mata Atlântica é considerada uma das mais importantes em termos de biodiversidade. No entanto, sofreu uma drástica redução da sua área devida a atividades antrópicas. O objetivo geral deste trabalho foi avaliar diferentes técnicas de análise de sobrevivência para modelar o efeito de covariáveis ambientais e antrópicas no desenvolvimento de fragmentos de Mata Atlântica. No primeiro capítulo o objetivo foi avaliar a técnica de análise de sobrevivência intervalar para modelar o efeito de covariáveis ambientais no recrutamento e na mortalidade de árvores em fragmentos de Mata Atlântica. A metodologia avaliou com eficácia o efeito de covariáveis ambientais no recrutamento e na mortalidade de árvores em fragmentos de Mata Atlântica. No segundo capítulo avaliou-se a mortalidade individual de árvores de um fragmento de Mata Atlântica, por meio da técnica de análise de sobrevivência com dados discretos. Esta metodologia apresentou uma acurácia superior a 80% na avaliação da mortalidade de árvores na Mata Atlântica. No terceiro capítulo foi utilizada a abordagem de análise de sobrevivência, por meio de métodos de aprendizado de máquina, para estudar a mortalidade de árvores da Mata Atlântica, em função de covariáveis ambientais e antrópicas. A metodologia random survival forests foi o método mais eficiente em avaliar a mortalidade individual de árvores nos fragmentos de Mata Atlântica. Os métodos aplicados neste estudo podem servir como ferramentas estratégicas para a tomada de decisões por gestores florestais, contribuindo para a conservação e manejo sustentável da Mata Atlântica.

Palavras-chave: Dados Censurados; Modelos paramétricos; Modelos de regressão discretos; Aprendizado de máquina; Florestas tropicais.

¹ Departamento de Estatística, Universidade Federal de Viçosa – marciel.duarte@ufv.br

² Departamento de Estatística, Universidade Federal de Viçosa – martinsfilho@ufv.br

³ Departamento de Engenharia Florestal, Universidade Federal de Viçosa – carlos.eleto@ufv.br



Using correspondence analysis to characterize the dynamics between depth and chlorophyll index in a semi-arid reservoir

Francisco Gustavo-Silva¹, Amanda Menezes Ferreira², Alan Cezar Bezerra³, Antônio Henrique Cardoso do Nascimento⁴, Paulo Duarte-Neto⁵

Abstract

This study analyzed the relationship between depth and the chlorophyll index of a reservoir, seeking to understand ecological dynamics and provide information on primary activity and siltation. The study was conducted at the Jazigo Dam, in Serra Talhada (PE), in the Pajeú river basin. Depth was measured using an echo-bathymetric probe and an RTK GPS. The chlorophyll index was estimated by remote sensing, using the NDCI (*Normalized Difference Chlorophyll Index*), calculated from bands B5 and B4 of the MSI sensor on the Sentinel-2 satellite. The depths were then categorized as Low (0.2–3.0 m), Medium (3.0–6.0 m) and High (>6.0 m). As for the chlorophyll index, we classified it as Mesotrophic (-0.1-0.0), Eutrophic (0.0-0.2), Supereutrophic (0.2-0.5) and Hypereutrophic (>0.5), as well as assessing the quality of the water, which is considered good for $NDCI < 0,1$ and bad for $NDCI > 0,1$, according to the work of Lamparelli and Mishra & Mishira. We then applied a Correspondence Analysis, in which the first two axes explained 54.2% of the total variability. The results indicated a strong association between shallow areas and high chlorophyll levels (eutrophic and supereutrophic), reflecting high biological productivity and poor water quality. In contrast, deeper regions were associated with lower chlorophyll concentrations and better environmental conditions. These findings highlight the impact of siltation and eutrophication in semi-arid aquatic environments, reinforcing the importance of integrated monitoring of these systems.

keywords: Depth; Geographic Positioning System; Remote Sensing

¹ Laboratório de modelagem estatística e CompUtacional de fenômenos NAturais complexos - LA-CUNA/UFRPE/DEINFO, Recife - Pernambuco, Brasil - fr.gustavosilva.098@gmail.com

² Laboratório de Dinâmicas Populacionais Marinhas - DIMAR/UFRPE/DEPAq, Recife, Pernambuco, Brasil - engpescmenezes@gmail.com

³ Laboratório de Geotecnologias Aplicada ao Semiárido - LAGAS/UFRPE/UAST, Serra Talhada - Pernambuco, Brasil - alan.bezerra@ufrpe.br

⁴ Laboratório de Geotecnologias Aplicada ao Semiárido - LAGAS/UFRPE/UAST - Serra Talhada - Pernambuco, Brasil - antonio.nascimento@ufrpe.br

⁵ Laboratório de modelagem estatística e CompUtacional de fenômenos NAturais complexos - LA-CUNA/UFRPE/DEINFO, Recife - Pernambuco, Brasil - pjduarteneto@gmail.com



Análise do comportamento da série de COVID-19 via Autorregressão Quantílica

Camilly Oliveira Brito da Silva¹

Gecynalda Soares da Silva Gomes²

Resumo

Este trabalho consiste em analisar o comportamento da série temporal dos dados do COVID-19 via autorregressão quantílica. O objetivo é prever e analisar a distribuição dos dados do COVID-19 em cada quantil, e não somente na média condicional. A análise completa foi realizada utilizando três variáveis: número de casos, número de mortes e cobertura vacinal. Os dados foram acessados pelo pacote COVID-19 Data Hub e inicialmente feita uma análise exploratória, após isso foi estimada uma equação para cada quantil de 0,05 até 0,95, com incrementos de 0,05. A metodologia utilizada foi a Autorregressão Quantílica (ARQ), introduzida por Koenker e Xiao (2004), que se trata da associação entre a regressão quantílica e o modelo autorregressivo, o que permite captar assimetrias e heterogeneidades na dinâmica temporal dos dados. Este trabalho ainda mostra que os dados analisados não apresentam comportamento uniforme ao longo dos diferentes quantis condicionais. No que tange aos principais resultados, a estatística utilizada foi o Qui-Quadrado mediante o Teste de Ljung-Box, e através do teste de hipótese foi obtido p -valor menor que o nível de significância (0,05) em todos os quantis, exceto em 0,70, 0,75, 0,80, 0,85 para mortalidade por COVID-19, exceto 0,30, 0,35, e 0,95 para casos confirmados, e exceto no quantil 0,35 para cobertura vacinal, ou seja, nos quantis citados das respectivas variáveis analisadas não rejeitamos a hipótese de não-correlação da série residual, uma vez que o p -valor é maior que 5%, assim, o teste não apresenta adequabilidade apenas para os modelos para esses quantis específicos.

Palavras-chave: Séries Temporais; Autorregressão Quantílica; Distribuição Condicional; COVID-19.

¹ Departamento de Estatística - UFBA, Salvador-BA – camillysilva@ufba.br

² Departamento de Estatística - UFBA, Salvador-BA – gecynaldassg@ufba.br



Características de seca do município de Recife-PE utilizando o Standardised Precipitation-Evapotranspiration Index

Gabriel Graciano de Mendonça;¹ Hellen Sonaly Silva Alves;² Antonio Samuel Alves da Silva.³

Resumo

A seca é um dos fenômenos naturais mais frequentes no mundo e, ao mesmo tempo, um dos desastres naturais mais complexos e menos compreendidos. Essa complexidade está relacionada à dificuldade de quantificá-la, pois não há uma variável física capaz de medila ou quantificá-la. Para isso, foram desenvolvidos diversos índices, a fim de quantificar e analisar as secas, entre eles o Standardised Precipitation-Evapotranspiration Index (SPEI), que combina a sensibilidade do Palmer Drought Severity Index com a simplicidade do Standardised Precipitation Index. Este trabalho tem como objetivo calcular o SPEI para o município de Recife-PE e, a partir disso, analisar as características das secas, como duração, severidade, intensidade, frequência e picos. Foram utilizados os dados de precipitação e evapotranspiração no período de janeiro de 1961 a dezembro de 2023, retirados do Brazilian Daily Weather Gridded Data (BR-DWGD). Os resultados mostraram uma duração média de seca de 1,72 meses, com 257 meses apresentando SPEI inferior a -0,5, o que representa uma frequência de 33,99% no período analisado. A severidade total, definida pelo somatório dos valores de SPEI abaixo do limiar, foi de 273,95, com média de 1,84. A intensidade total, definida pela razão entre severidade e duração, foi de 150,1, com uma intensidade média de 1,01. O pico da série, menor valor registrado de SPEI, foi de -2,70. Tais resultados podem contribuir para o planejamento e formulação de estratégias mais eficazes para enfrentar os impactos da seca no município.

Palavras-chave: Standardised Precipitation-Evapotranspiration Index; Secas; Características.

¹Departamento de Estatística e Informática, Recife - gabriel.gracianom@ufrpe.br

²Departamento de Estatística e Informática, Recife - hellen.ssilva@ufrpe.br

³Departamento de Estatística e Informática, Recife - antonio.sasilva@ufrpe.br



Metodologias para Análise de Trajetórias Latentes: Implementação em Software Estatístico

Ruan Gabriel da Cruz Cerqueira ^{1,3}

Marcelo Magalhães Taddeo ^{2,3}

Leila Denise Alves Ferreira Amorim ^{1,2,3}

Resumo

A disponibilidade de dados individuais longitudinais tem aumentado em várias áreas do conhecimento, auxiliando na descrição de fenômenos biológicos, comportamentais e sociais. Os modelos mistos têm sido usados rotineiramente para análise destes dados. Mais recentemente, tem crescido o interesse em métodos capazes de identificar e caracterizar subpopulações homogêneas (classes latentes) a partir das trajetórias longitudinais de uma população heterogênea. Neste contexto, as principais abordagens utilizadas são os Modelos de Mistura de Crescimento (*GMM=Growth Mixture Models*) e Análise de Crescimento com Classes Latentes (*LCGA=Latent Class Growth Analysis*), que podem ser utilizadas em desfechos contínuos ou discretos. A principal diferença entre essas abordagens está na inclusão ou não de efeitos aleatórios que permitem variação entre os indivíduos de uma mesma classe latente. Sua implementação pode ser realizada em diferentes softwares, como o R, STATA e MPLUS, que adotam estratégias distintas de estimação. No STATA, o plugin **TRAJ** estima os parâmetros através da maximização da verossimilhança através de métodos Quasi-Newton. O MPLUS utiliza o algoritmo Expectation-Maximization (EM) para a estimação. No R, o ajuste dos modelos pode ser feito com os pacotes **lcmm** e **flexMix**. O pacote **flexMix** utiliza o algoritmo EM, enquanto o pacote **lcmm** realiza a maximização direta da verossimilhança através do algoritmo iterativo de Marquardt. Este trabalho tem como objetivo descrever estas metodologias, comparando as ferramentas disponíveis nos diferentes softwares. Para ilustração de métodos alternativos para análise de trajetórias latentes são considerados dados de mortalidade infantil de cidades brasileiras entre os anos de 2005 e 2012.

Palavras-chave: Trajetórias Latentes; Modelos de Mistura; Classes Latentes; Dados Longitudinais; Modelos Mistos.

¹ Programa de Pós-Graduação em Estatística e Ciência de Dados (PGECD), UFBA, Salvador-Bahia – ruangabrielc@gmail.com

² Departamento de Estatística, Instituto de Matemática e Estatística, UFBA, Salvador-Bahia – marcelo.magalhaes@ufba.br, leiladen@ufba.br [Apoio recebido da Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB, termo de outorga PP0021/2023) e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, termo de outorga 403465/2023-0)]

³ Laboratório de Inferência Causal e Aplicações - LInCA, <https://linca.ufba.br/>



Comparison of spatial statistical models for areal data with outliers

Gean Pereira Damaceno;^{*1} Giovani Festa Paludo;² Alex Nhancololo;³ Elias M. Sabe;² Geovane Fernandes da Silva;¹ Júlio Sílvio de Sousa Bueno Filho;⁴ Douglas Mateus da Silva; ⁴ João Domingos Scalon ⁴

Abstract

Tuberculosis is an infectious and transmissible disease that represent a public health challenge, especially in regions with marked social inequalities. This study aimed to compare different spatial modeling approaches applied to the incidence of tuberculosis in Maranhão, highlighting the impact of atypical observations on the results. The global and local Moran's index (LISA) were calculated to verify, respectively, the existence of global spatial autocorrelation and clusters of local spatial autocorrelation. Associations with some demographic covariates were studied and three models were adjusted: classical linear regression, spatial simultaneous autoregressive model (SAR) and spatial error model (CAR). The linear regression presented an adjusted R^2 of 0.3168 and residuals with non-random patterns, indicating violation of the assumptions of the classical model. The overall Moran's Index was 0.095, statistically significant at the 5% confidence level, evidencing the presence of spatial autocorrelation in the data and clusters of correlation of the local index in municipalities, justifying the use of spatial models. The CAR model showed the best performance (AIC = 829.49), followed by SAR (AIC = 833.76), both impacted by outliers. Since the residuals of the models violate the assumptions of normality, their use is not recommended. Despite the improvement in the fit with spatial models, the results highlight the need for techniques that avoid the problems associated with the presence of outliers. It is concluded that spatial modeling is essential for understanding the distribution of tuberculosis, but should be complemented by methods that consider the structure of the data.

Keywords: Outliers; Robustness; Spatial statistics; Tuberculosis; Public health.

¹Programa de Graduação em Estatística com ênfase em Ciência de Dados, Universidade Federal de Lavras, Lavras - {geandamaceno030@gmail.com, geovanefs90@gmail.com}

²Programa de Pós-graduação em Estatística e Agropecuária, Universidade Federal de Lavras, Lavras - {elias.sabe1@estudante.ufla.br, gfpaludo@gmail.com}

³Programa de Pós-graduação em Estatística, Universidade de São Paulo, São Paulo - amnhancololo@gmail.com

⁴Departamento de Estatística, Universidade Federal de Lavras, Lavras - {jssbueno, douglas.mateus, scalon}@ufla.br



Tendências Climáticas no Rio Grande do Norte: Análise de Dias Secos e Chuvas Intensas entre 1991 e 2020

Mylleni Silvia Chaves Durval¹

Kaylane Sabino da Silva²

Fidel Ernesto Castro Morales³

Antonio Marcos Batista do Nascimento²

Daniele Tôrres Rodrigues³

Resumo

Estudos regionais sobre mudanças climáticas indicam que o número de dias consecutivos secos pode estar aumentando em razão de fatores como o aquecimento global e a variabilidade climática. Esse fenômeno acentua os impactos negativos das secas, incluindo a degradação dos recursos naturais, a perda de biodiversidade e a redução da recarga de reservatórios hídricos. O objetivo principal deste estudo é analisar a ocorrência de tendências no número de dias secos consecutivos por ano, bem como no número de dias com precipitação superior a 20 mm por ano, em sete municípios do estado do Rio Grande do Norte, durante o período de 01/01/1991 a 31/12/2020. Para a análise, foram aplicados o teste não paramétrico de Mann-Kendall, utilizado para detectar tendências monotônicas em séries temporais, e o estimador de Sen-Slope, empregado para quantificar a magnitude das tendências observadas. Os resultados apontam para tendências estatisticamente significativas de aumento no número de dias secos consecutivos no município de Apodi. Por outro lado, verificou-se uma tendência de redução no número de dias com precipitação superior a 20 mm nos municípios de Cruzeta e Macau. Essas evidências reforçam a importância do monitoramento contínuo da variabilidade climática regional, fornecendo subsídios relevantes para a formulação de estratégias de adaptação e mitigação, especialmente no contexto da segurança hídrica, da agricultura e da sustentabilidade das comunidades locais.

Palavras-chave: Mudanças climáticas; Dias secos consecutivos; Aquecimento global; Análise de tendência.

¹Departamento de Matemática, UFRN - MMylenichaves@gmail.com

²Departamento de Estatística, UFRN

³Departamento de Estatística, UFPI



Estimativa ultrassonográfica do peso fetal em gestantes diabéticas ou com diabetes gestacional

Matheus Monteiro Alves Melo ¹; Luciana Graziela de Godoi ²

Resumo

O peso fetal (PF) está diretamente ligado à saúde gestacional, sendo atualmente as medidas observadas no exame ultrassonográfico o recurso mais utilizado para a quantificação indireta dessa medida. Hadlock et al. (1984) propôs uma das fórmulas mais utilizadas atualmente para a estimativa do PF na população de mulheres em geral, porém é sabido que a saúde da gestante interfere no PF. Por exemplo, a presença de diabetes durante a gestação é um importante fator prognóstico para a macrossomia fetal por representar risco elevado de morbimortalidade materna e perinatal. Dessa forma, este estudo investigou se fórmulas atuais de estimativa subestimam o peso ao nascer em gestantes diabéticas (gestacional ou não). Para isso, foram analisados 1140 registros de partos e exames ultrassonográficos da Clínica Obstétrica do HC-FMUSP.

Durante as análises, foram detectados indícios de heteroscedasticidade no modelo de Hadlock, indicando a necessidade de ajustes. Assim, propomos diferentes fórmulas para estimativa do PF considerando distribuições alternativas para modelar a variável resposta, tais como a normal heteroscedástica, Gama e Box-Cox Cole and Green, no contexto dos modelos aditivos generalizados para locação, escala e forma (GAMLSS).

Após ajuste, os PF estimados e reais foram classificados em adequado (AIG), grande (GIG) ou pequeno (PIG) para a idade gestacional e comparados via índices tipo Kappa. O modelo gama foi pontualmente melhor do que o de Hadlock, mas estatisticamente equivalente. Demais medidas de desempenho de classificação consideradas mostraram um desempenho superior do modelo Gama, principalmente com respeito a categoria GIG, alvo inicial do estudo.

Palavras-chave: peso fetal; obstetrícia; regressão; heteroscedasticidade; GAMLSS.

¹Departamento de Estatística, UFES - matheus.melo@edu.ufes.br

²Departamento de Estatística, UFES - luciana.godoi@ufes.br



Modelagem de risco climático-financeiro para o setor agrícola via cônjuges bivariadas

Carolyne Soares de Brito;¹ Cira Etheowalda Guevara Otiniano² Enzo Porto Brasil;³

Resumo

O aquecimento global e a crescente frequência de eventos climáticos extremos têm impactado diretamente a produção agrícola e a estabilidade financeira de empresas do agronegócio, cuja produtividade e preços dos seus produtos estão fortemente associados às condições ambientais. Este trabalho avalia o impacto de duas variáveis climáticas no valor das ações da empresa BrasilAgro, utilizando técnicas estatísticas relacionadas a distribuições de cauda pesada, Teoria dos Valores Extremos e cônjuges. Considerando os dados de março de 2021 a dezembro de 2024 e assumindo a hipótese de independência temporal, analisamos as distribuições marginais das variáveis climáticas (temperatura de orvalho e umidade relativa do ar) de Correntina-Bahia, cidade onde a empresa possui sua maior área agricultável, e dos log-retornos das ações da BrasilAgro. Para contornar a hipótese de independência temporal, extraímos subamostras de máximos dos retornos da empresa, denotados por r_t , e subamostras de mínimos das variáveis climáticas, denotadas por x_t (temperatura de orvalho) e y_t (umidade relativa do ar). Essas subamostras foram ajustadas pela distribuição GEV (Generalized Extreme Value), com destaque para uma inovação: o uso da distribuição GEV bimodal para ajustar y_t . Modelamos os pares (r_t, x_t) e (r_t, y_t) utilizando cônjuges para capturar a dependência entre marginais. Com base nos critérios AIC e BIC, as cônjuges de Tawn II, com rotações de 90° e 270°, apresentaram o melhor desempenho. O ajuste bivariado permitiu calcular a medida de risco bivariada VaR (*Value at Risk*), oferecendo uma compreensão mais aprofundada do impacto das variáveis climáticas nas ações da empresa em estudo.

Palavras-chave: Cônjuges; Distribuições GEV e BGEV; Modelagem de valores extremos; Setor agrícola; Mudanças climáticas.

¹Departamento de Estatística, Universidade de Brasília - carolynesb.01@gmail.com

²Departamento de Estatística, Universidade de Brasília - cira@unb.br

³Departamento de Estatística, Universidade de Brasília - enzopporto2000@gmail.com



Irradiância solar extrema e sua relação com variáveis atmosféricas: uma abordagem com teoria dos valores extremos e cópulas

Enzo Porto Brasil ¹; Cira Etheowalda Guevara Otiniano ²;
Carolyne Soares de Brito ³; Beatriz Gomes da Cruz Albernaz ⁴

Resumo

O estudo da irradiância solar (IS) é essencial para compreender seu impacto nos ciclos climáticos, nas dinâmicas atmosféricas e, por consequência, no contexto social. Eventos solares extremos, como as Tempestades de Halloween (2003) e a Tempestade Geomagnética de 2024, ressaltam a importância do monitoramento contínuo da IS, diante dos riscos que representam para tecnologias modernas. Esta pesquisa combina a Teoria dos Valores Extremos (TVE) com a teoria de cópulas, buscando avaliar os impactos e as relações entre variáveis atmosféricas de sistemas complexos. A análise inicial utilizou 15 anos de dados da missão espacial *Solar Radiation and Climate Experiment* (SORCE), da NASA, a partir dos quais foi modelada a IS pelas distribuições α -estável e GEV (*Generalized Extreme Value*), que capturam padrões raros nas caudas. A partir disso foram calculados valores em risco (VaRs) e níveis de retorno. Para investigar a possível influência da IS sobre outras variáveis atmosféricas, também foram analisados e modelados dados do Instituto Nacional de Meteorologia (INMET), com foco especial no Distrito Federal, Brasil, região frequentemente afetada por secas severas e queimadas acima da média nacional. O estudo aborda ainda a modelagem da densidade do ar pela distribuição GEV bimodal e a aplicação de cópulas bivariadas para avaliar as dependências. Os resultados mostram que as variáveis correspondem a um sistema complexo com duas subpopulações, sendo bem modelado pela combinação de cópulas apropriadas com as distribuições marginais das variáveis extremas. Por fim, os níveis de retorno bivariados permitem interpretar com precisão o impacto da irradiância solar em variáveis atmosféricas.

Palavras-chave: irradiância solar; densidade do ar; Teoria dos Valores Extremos (TVE); cópulas; distribuições GEV e BGEV.

¹Departamento de Estatística, Universidade de Brasília - enzoporto2000@gmail.com

²Departamento de Estatística, Universidade de Brasília - ciragotiniano@gmail.com

³Departamento de Estatística, Universidade de Brasília - carolynesb.01@gmail.com

⁴Departamento de Estatística, Universidade de Brasília - beatrizgomesalbernaz@gmail.com



Distribuição Espacial de Focos de Queimadas no Estado do Ceará com Base em Processos Pontuais (2022–2024)

Vitoria Soares de Souza ¹ Hellen Sonaly Silva Alves ² Daví Barbosa Pereira de Sousa ³
Jéssica Cristina Pereira Batista ⁴ Marcela Silva de Araujo ⁵ Guilherme Rocha Moreira ⁶
João Domingos Scalon ⁷

Resumo

O bioma Caatinga é uma região semiárida com grande vulnerabilidade a incêndios, especialmente em períodos de seca, sendo registrados grandes volumes de focos de queimadas entre o período de 2022 a 2024. Diante disso, visa-se observar o comportamento espacial dos focos de queimadas no estado do Ceará, no período supracitado, com dados disponibilizados pelo Instituto Nacional de Pesquisas Espaciais e análises realizadas no software RStudio. Observou-se uma concentração de queimadas nas regiões norte e centro-oeste do estado em 2022, com padrão espacial claramente agrupado. Em 2023, houve um aumento no número de focos, indicando o agravamento da situação. Já em 2024, o número de focos foi ainda maior, evidenciando a intensificação do problema ao longo do período analisado. O teste de Kolmogorov-Smirnov foi aplicado para os três anos e identificou-se não se tratar de um processo homogêneo. Assim, aplicou-se o estimador de densidade Kernel, que revelou baixa densidade de focos em relação à área total do estado. Utilizando da função K, notou-se que para os três anos há padrão de agrupamento, levando em consideração distâncias de até 150km. Diante do exposto, a intensificação e a concentração dos focos em áreas específicas reforçam a necessidade de ações preventivas e de monitoramento contínuo, especialmente nas regiões mais afetadas do bioma Caatinga.

Palavras-chave: Distribuição geográfica; Padrões espaciais; Queimadas.

¹Universidade Estadual da Paraíba, Departamento de Estatística, Campina Grande- vitoria-soares0104@gmail.com

²Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, Recife-hellen.silva@ufrpe.br

³Universidade Federal de Lavras, Departamento de Estatística, Lavras- davi.barbosa.5454@gmail.com

⁴Universidade Estadual da Paraíba, Departamento de Estatística, Campina Grande- jessicapstatistics@gmail.com

⁵Universidade Federal de Lavras, Departamento de Estatística, Lavras - marcela.araujo1@estudante.ufla.br

⁶Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, Recife - guilherme.rochamoreira@ufrpe.br

⁷Universidade Federal de Lavras, Departamento de Estatística, Lavras - scalon@ufla.br



Análise Espacial da Incidência de Dengue em Minas Gerais no Ano de 2024 com Ênfase na Autocorrelação Espacial

Marcela Silva de Araujo;¹ Daví Barbosa Pereira de Sousa;² Viviane Costa Silva;³ Hellen Sonaly Silva Alves;⁴ Vitoria Soares de Souza;⁵ Marcelo Ângelo Cirillo;⁶ Guilherme Rocha Moreira⁷

Resumo

A dengue é uma doença infecciosa transmitida pelo mosquito Aedes aegypti, prevalente em regiões tropicais e subtropicais. A prevenção baseia-se na eliminação de criadouros, campanhas educativas e, em alguns casos, vacinação. Este estudo analisa a taxa de infecção por dengue em Minas Gerais no ano de 2024, considerando a razão de 1:1000 habitantes. Com base nos dados do TABNET/SINAN, investigou-se a distribuição espacial dos casos e a influência das localidades nas taxas de infecção. A análise revelou autocorrelação espacial positiva, segundo o Índice de Moran Global ($I = 0,02$, valor - $p = 0,02$), indicando que municípios vizinhos tendem a apresentar taxas semelhantes. A análise local (LISA) identificou agrupamentos significativos: Ituiutaba, Frutal, Campina Verde, Monte Carmelo, Patrocínio, Patos de Minas e Araxá foram classificados como “alto-alto”, destacando áreas críticas. Por outro lado, Mariana, Ouro Preto, São João del Rei, Barbacena e Congonhas apresentaram padrões “baixo-baixo”, com menor incidência. As diferenças nas taxas podem estar associadas ao tipo de vegetação e às condições ambientais. Municípios com vegetação densa, como Ituiutaba e Patos de Minas, oferecem mais locais propícios à reprodução do vetor. Em contraste, áreas urbanizadas, como Mariana e Congonhas, com melhor infraestrutura de saneamento, tendem a registrar menor incidência. Esses resultados reforçam a importância da localização geográfica na distribuição da dengue e a necessidade de medidas focadas nas regiões mais afetadas.

Palavras-chave: Dengue; Distribuição espacial; Autocorrelação; Moran Global; Epidemiológico.

¹Universidade Federal de Lavras, Departamento de Estatística, Lavras, MG - marcela.araujo1@estudante.ufla.br

²Universidade Federal de Lavras, Departamento de Estatística, Lavras, MG - davi.sousa@estudante.ufla.br

³Universidade Federal de Lavras, Departamento de Estatística, Lavras, MG - viviane.silva6@estudante.ufla.br

⁴Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, Recife, PE - hellen.ssilva@ufrpe.br

⁵Universidade Estadual da Paraíba, Departamento de Estatística, Campina Grande, PB - vitoria-soares0104@gmail.com

⁶Universidade Federal de Lavras, Departamento de Estatística, Lavras, MG - macufla@ufla.br

⁷Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, Recife, PE - guilherme.rochamoreira@ufrpe.br



Painel de Vigilância da Saúde Materna e Perinatal

Beatriz Lima Silveira ¹; Esley Caminhos Ferreira ²; Rafael Herzog Sant'Ana ³; Mariana Costas Freitas ⁴; Agatha Sacramento Rodrigues ⁵

Resumo

Tragédias como a morte precoce do neonatal e a morte da mulher durante a gestação, parto ou até 42 dias após o término da gestação - não são eventos isolados. São os resultados de uma série de determinantes sociais, econômicos, e situacionais que atuam durante a vida daqueles afetados, diante desse contexto, o Painel de Vigilância da Saúde Materna e Perinatal foi desenvolvido como uma ferramenta para apresentar em uma plataforma, indicadores que expressam a vulnerabilidade das mulheres e dos perinatais. Os indicadores são calculados a partir de bases de dados nacionais (SINASC, SIM, SINAN, SIH, SIAB, ANS e estimativas populacionais) e abrangem municípios, estados, regiões de saúde e o país, com dados disponíveis desde 2012 com atualização anual. Para o desenvolvimento da aplicação web foi selecionado o software livre R, que possui pacotes avançados para visualização de dados como shiny, flexydashboard, golem e outros, os microdados das bases disponíveis no Opendatasus foram coletados utilizando o pacote microdatasus. Para facilitar sua utilização por gestores de saúde, os indicadores possuem visualização intuitiva e incluem valores de referência, como metas nacionais e internacionais, recomendações da Organização Mundial da Saúde, valores observados em países desenvolvidos ou a média nacional. Por fim, o painel é uma ferramenta para auxiliar ações e políticas voltadas à redução da mortalidade materna e melhoria dos desfechos perinatais no Brasil.

Palavras-chave: morte materna; indicadores obstétricos; políticas de saúde pública; painel de visualização; indicadores perinatais.

¹Departamento de Estatística, Universidade Federal do Espírito Santo, Vitória, ES - beatrizlimasilveira@gmail.com

²Departamento de Estatística, Universidade Federal do Espírito Santo, Vitória, ES - esleycaminhasferreira@gmail.com

³Departamento de Estatística, Universidade Federal do Espírito Santo, Vitória, ES - rafasanatanah21@gmail.com

⁴Departamento de Estatística, Universidade Federal do Espírito Santo, Vitória, ES - marianacfistas2004@gmail.com

⁵Departamento de Estatística, Universidade Federal do Espírito Santo, Vitória, ES - agatha.srodrigues@gmail.com



Grafos e Fluxos: Redes Fúngicas sob um olhar da Modelagem e Estatística

Cezar Augusto Fontana Teixeira¹; Alice Pitanga Rocha¹; Daniel Moreira Melo¹
Daniel Oliveira Sabbagh¹; Lorena Souza Castro Altoé¹; Matheus Correia Casotti¹
Victor Alves Lopes¹; Yasmin Moreto Guaitolini¹; Flavia de Paula¹
Iúri Drumond Louro¹; Débora Dummer Meira¹

Resumo

A compreensão das redes fúngicas tem avançado graças à união entre modelagem matemática, estatística aplicada e genômica, que fornece bases quantitativas para descrever desde o crescimento de colônias até a arquitetura do “wood-wide web”. Logo, buscou-se descrever a importância da modelagem e técnicas estatísticas para elucidar a resiliência fúngica. Representar o micélio como um grafo — em que nós são pontos de ramificação ou anastomose e ligações são hifas ou cordões — permite extrair, manualmente ou via segmentação assistida, métricas de comprimento, área e resistência ao fluxo. A partir daí, medidas como coeficiente de agrupamento e índice alfa capturam a transição de árvores radiais a redes reticuladas, enquanto o cálculo de caminhos mínimos e eficiência global quantificam a rapidez e robustez do transporte de nutrientes. Modelos baseados em equações diferenciais parciais integram produção de inibidores, autofagia e difusão de biomassa para simular substituição em ambientes bidimensionais ou tridimensionais porosos, validando-se experimentalmente com testes não-paramétricos e transformações potência-logarítmicas que corrigem assimetrias de variância. Para prever fluxos, a analogia elétrica converte a matriz de adjacência em circuito de resistores, revelando como diâmetros de hifas e redundâncias paralelas ampliam a resiliência a ataques e facilitam a reconexão pós-dano. Em escalas celulares e de colônia, o transporte combina motores moleculares ao longo de microtúbulos e fluxos de massa em cordões, mapeados por isótopos, exibindo comportamento pulsátil e bi-direcional. Assim, o uso integrado de grafos, estatística e modelagem multiescalar está redesenhandando nossa visão de fungos e aplicações em biocontrole e design de infraestruturas inspiradas em redes vivas.

Palavras-chave: Redes fúngicas; Grafos; Modelagem matemática; Estatística aplicada; Transporte de nutrientes.



Avaliação da qualidade da água subterrânea do estado da Paraíba: uma abordagem por modelos hierárquicos

Laudeline Dantas Santana¹

Renata Alcarde Sermarini²

Resumo

O estado da Paraíba, no Nordeste do Brasil, destaca-se por sua diversidade edafoclimática, com variações em vegetação, clima, relevo e uso do solo. Essa heterogeneidade ambiental, associada à presença de afloramentos rochosos, caatinga arbórea e exploração mineral, influencia diretamente a qualidade da água subterrânea, que apresenta, em grande parte do semiárido paraibano, altos teores de salinidade, comprometendo seu uso, impactando comunidades rurais e atividades produtivas. As atividades antrópicas têm se tornado um agente cada vez mais relevante no processo de salinização dos recursos hídricos. Esta problemática está diretamente relacionada ao Objetivo de Desenvolvimento Sustentável 6 (ODS 6), em especial a meta 6.3, que visa melhorar a qualidade da água, reduzindo a poluição e minimizando a liberação de produtos químicos e despejo em cursos de água. Diante disso, buscou-se aplicar modelos mistos hierárquicos para identificar padrões espaciais da condutividade elétrica (CE) da água subterrânea e, captar informações sobre a qualidade da água a nível de mesorregião, microrregião e município, integrando precipitação e ano de coleta. Os parâmetros foram estimados por máxima verossimilhança restrita, os modelos foram comparados por meio do teste da razão de verossimilhanças e adotou-se o nível de 5% de significância. Não foi evidenciado efeito significativo da mesorregião nem da precipitação sobre a CE, entretanto, os anos de 2016, 2017 e 2019 apresentaram variação estatisticamente significativa, sugerindo mudanças temporais nos níveis de salinidade e, que a CE da água subterrânea varia de acordo com o município avaliado.

Palavras-chave: salinidade; qualidade da água; condutividade elétrica; classificação hierárquica.

¹ Departamento de Ciências Exatas - Escola Superior de Agricultura "Luiz de Queiroz" - USP - laudantas@usp.br

² Departamento de Ciências Exatas - Escola Superior de Agricultura "Luiz de Queiroz" - USP - ralcarde@usp.br



Investigando o comportamento endêmico e epidêmico da dengue no Paraná e sua associação com variáveis climáticas

Dayana Telles Bertuzzi;¹ Juliana Nascimento de Paula;² Jociene Santana Pimentel;³ Eniuce Menezes;⁴

Resumo

A dengue tem se intensificado nas Américas, especialmente no Brasil. Fatores climáticos como temperatura, umidade e precipitação influenciam a dinâmica do *Aedes aegypti*. Este estudo investigou a associação entre variáveis climáticas e a incidência de dengue nos municípios do Paraná (2016–2024). Os casos confirmados foram agrupados por padrão temporal usando k-means (método Elbow), com análises separadas para contextos endêmicos e epidêmicos. As variáveis climáticas — extraídas do ERA5-Land — foram agregadas semanalmente por município e associadas aos agrupamentos por meio de correlação cruzada (CCF) para identificar defasagens temporais entre clima e casos. Os resultados revelaram padrões temporais endêmicos e epidêmicos distintos entre os municípios. Por outro lado, a identificação dos municípios com características similares permitiu verificar que a temperatura foi o fator climático com maior correlação com a incidência de dengue, especialmente em contextos epidêmicos, nos quais os efeitos foram mais imediatos. A precipitação e a umidade mostraram influências variáveis, com lags prolongados em anos endêmicos e correlações mais discretas durante surtos.

Palavras-chave: dengue; *clustering*; variáveis climatológicas; endemia; epidemia.

¹Programa de Pós-graduação em Bioestatística, UEM, Maringá – dayanatbertuzzi@gmail.com

²Programa de Pós-graduação em Bioestatística, UEM, Maringá – juliananp.es@gmail.com

³Secretaria de Estado da Saúde do Paraná – jociene@sesa.pr.gov.br

⁴Departamento de Estatística, Programa de Pós-graduação em Bioestatística, UEM, Maringá – emeneszes@uem.br



Séries Temporais Aplicadas no Estudo da Variação Temporal da Temperatura e Precipitação em Estações Climatológicas do Estado de São Paulo, Período: 1970 - 2023

Victória Marchetti ¹
José Silvio Govone ²

Resumo

As mudanças climáticas representam uma questão de extrema relevância e urgência no equilíbrio ambiental, econômico e social do planeta, caracterizada por mudanças bruscas e eventos extremos como inundações e ondas de calor que, conforme os anos passam, estão se tornando mais frequentes e preocupantes, impactando diretamente a saúde pública, a agricultura e os recursos hídricos. Vários estudos mostram a urgência de minimizar os impactos das mudanças climáticas, como o Sexto Relatório de Avaliação (AR6, 2023) do Painel Intergovernamental sobre Mudanças Climáticas (IPCC) que, de maneira geral, alerta sobre as graves consequências dessas mudanças. Dentre os impactos, se destacam as mudanças de temperatura e precipitação. A compreensão dos padrões climáticos passados e futuros é essencial para a formulação de políticas públicas sustentáveis e eficazes. Este estudo visa analisar séries temporais das variáveis temperatura e precipitação ao longo de mais de 50 anos, em estações climatológicas do Estado de São Paulo. Para isso, foram usados os dados disponibilizados pelo DAEE (Departamento de Águas e Energia Elétrica) e Esalq - USP (Escola Superior de Agricultura "Luiz de Queiroz" Universidade de São Paulo), utilizando modelos estatísticos, como o ARIMA (*Autoregressive Integrated Moving Average*), para identificar padrões, tendências e anomalias no período de 1970 a 2023, o que pode fornecer informações valiosas sobre o clima da região. Resultados preliminares indicam uma tendência de aumento na temperatura e na precipitação, além de alguns *outliers* pontuais. Tal análise é importante para compreender o fenômeno das mudanças climáticas e suas implicações na natureza e sociedade, facilitando o desenvolvimento de estratégias de adaptações na região estudada.

Palavras-chave: Mudanças Climáticas; Séries Temporais.

¹Discente do Programa de PG em Biometria, IBB, UNESP, Botucatu - victoria.marchetti@unesp.br

²Departamento de Estatística, Matemática Aplicada e Computação, IGCE, UNESP, Rio Claro - js.govone@unesp.br



Análise bivariada da Temperatura máxima e Umidade relativa do ar mínima em Brasília (DF)

Beatriz Gomes da Cruz Albernaz;¹ Cira Etheowalda Guevara Otiniano;²
Carolyne Soares de Brito;³ Enzo Porto Brasil⁴

Resumo

As transformações climáticas têm sido uma preocupação crescente, especialmente em regiões que enfrentam eventos extremos cada vez mais frequentes. No Distrito Federal, mudanças significativas nos padrões climáticos têm chamado atenção, com episódios de calor intenso, frio atípico e longos períodos de seca. Neste trabalho, utilizou-se a metodologia de valores extremos e cópulas para analisar os padrões de temperatura máxima e umidade relativa do ar mínima em Brasília, bem como a dependência entre essas variáveis. Uma análise preliminar mostrou que, ao considerar independência temporal dos dados, a distribuição α -estável ajustou-se bem à variação logarítmica de cada conjunto de dados. Sem violar a hipótese de independência, os dados foram organizados em blocos de máximos e mínimos e ajustados pela distribuição GEV. Para capturar a dependência entre temperatura e umidade, foram ajustadas cópulas arquimedianas, sendo que a cópula de Clayton apresentou os menores valores de AIC e BIC. Por fim, foi calculada a medida de risco hidrológico, determinada pelo nível de retorno (x_T, y_T) associado a um tempo de retorno T . Para períodos de 116 dias, espera-se que a temperatura ultrapasse 36°C; para a umidade relativa do ar mínima, deve-se esperar, em média, 128 dias até superar 35,99%. Em períodos maiores (5800/6400 dias), as temperaturas mais extremas esperadas são 38,14°C e a umidade mínima esperada é superior a 41,51%.

Palavras-chave: α -estável; Cópulas; GEV; Temperatura; Umidade relativa do ar.

¹Departamento de Estatística, Universidade de Brasília - beatrizgomesalbernaz@gmail.com

²Departamento de Estatística, Universidade de Brasília - cira@unb.br

³Departamento de Estatística, Universidade de Brasília - carolynesb.01@gmail.com

⁴Departamento de Estatística, Universidade de Brasília - enzoporto2000@gmail.com



Diagnóstico de Pacientes com Alto Risco de Doença Mental por Meio da Análise do Discurso Verbal e Não Verbal

Bruno Wavrzenczak;¹ Felipe Queluz;² Alexandre Loch;³ Anderson Ara⁴

Resumo

O ARMS (*At-Risk Mental States*) ou "Estados Mentais de Alto Risco" é um conceito na psiquiatria que se refere a um conjunto de critérios clínicos e sintomas psicopatológicos que indicam uma probabilidade aumentada de desenvolver um transtorno psicótico, como esquizofrenia, em um futuro próximo. Esses critérios foram desenvolvidos para identificar indivíduos que estão em um estágio prévio ao desenvolvimento completo da psicose, permitindo intervenções precoces e preventivas para evitar ou mitigar o surgimento do transtorno. Com o intuito de contribuir para a detecção precoce desses estados, este trabalho propõe uma abordagem baseada em um ensemble de modelos, combinando diferentes fontes de dados. Um dos modelos realiza a análise de landmarks faciais, extraíndo informações não verbais a partir de expressões e microexpressões faciais. O outro modelo foca na análise da fala, utilizando features extraídas a partir de grafos de fala (*speech graphs*), que capturam a estrutura e a coerência do discurso verbal. A integração dessas duas modalidades permite uma avaliação mais abrangente dos sinais verbais e não verbais associados ao risco de transtornos mentais, com o objetivo de aprimorar o diagnóstico precoce e a eficácia das intervenções.

Palavras-chave: ARMS; Ensemble; Landmarks; Speech Graphs.

¹Departamento de Estatística, UFPR - brunowaurzenczak@ufpr.br

²Departamento de Estatística, UFPR - felipeheron@ufpr.br

³Universidade de São Paulo, USP - alexandre.loch@usp.br

⁴Departamento de Estatística, UFPR - ara@ufpr.br



Evaluation of Arterial Blood Pressure Indices in Young Adults with Sickle Cell Anaemia and Haemoglobin AA in Nigeria

Adisa Jamiu saka;¹ Muhammed. A. Akintunde²

Resumo

Sickle Cell Anaemia (SCA) as a genetic blood disorder caused by a mutation in the haemoglobin gene, leading to the production of abnormal haemoglobin known as haemoglobin S (Kumar, et al. 2021). This abnormal haemoglobin causes red blood cells to become rigid, sticky, and shaped like a crescent or sickle, which can obstruct blood flow and lead to various complications such as pain, infections, and could go to the extent of damaging nerves and organs (kidneys, liver and spleen) by (Rees, et al. 2020). This research utilizes a two-level factorial experiment to evaluate the impact of four major factors (Age, Sex, Genotype, and Rhythm) on six distinct blood pressure (BP) indices: Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), Pulse Rate (PR), Pulse Pressure (PP), Mean Arterial Pressure (MAP), and Rate Pressure Product (RPP). The experimental units consist of young adults with Sickle Cell Anaemia (SCA) and Haemoglobin AA (HbAA). The results of analysis indicate that Age and Genotype are the major significant factors affecting blood pressure (BP) indices. Meanwhile, Pulse Pressure (PP) appears to be more sensitive to the aforementioned factors when compares to SBP or DBP. Also, the interaction effects between Age and Genotype, and between Age and Sex demonstrates clinical relevance. Importantly, these results highlight the importance of early detection of abnormal cardiovascular symptoms and opens way for further heart disease diagnostic test and treatments in the young adults. It is also worthwhile to note that Pulse Pressure (PP) provides a more comprehensive measure for the abnormal cardiovascular detection within the young adults.

Palavras-chave: Sickle Cell Anaemia 1; Blood Pressure Indices 2; Factorial Experiment 3; Treatment Effects 4; Young Adults 5.

¹Department of Statistics, Federal University of Lavras, Brazil - adisa.saka1@estudante.ufla.br, saka-jamiu@gmail.com

²Department of Mathematics, Obafemi Awolowo University, Ile Ife 220005, Nigeria. - maakin-tunde@student.oauife.edu.ng



Análise da saúde mental na população brasileira: uma abordagem com amostragem complexa

Daisy Santana Ferreira¹, Liciana Vaz de Arruda Silveira²

Este estudo teve como objetivo descrever e analisar a presença de sintomas depressivos na população brasileira com 15 anos ou mais, com base nas respostas ao instrumento PHQ-9, presente no banco de dados da Pesquisa Nacional de Saúde (PNS-2019). Para a análise, as respostas foram recategorizadas em duas categorias (baixa e alta frequência), a fim de garantir maior estabilidade estatística, considerando as frequências reduzidas nas categorias mais severas. As análises foram conduzidas com base no delineamento amostral complexo da PNS, que envolve estratificação, conglomerados e pesos amostrais, o que torna inadequada a utilização de métodos estatísticos clássicos que assumem amostragem aleatória simples. Na etapa descritiva, foram calculadas proporções ajustadas das respostas ao PHQ 9. Para investigar associações entre sintomas depressivos e variáveis sociodemográficas (sexo, idade, região), utilizou-se o teste qui-quadrado com correção de Rao-Scott, implementado por meio da função svychisq() do pacote survey no R (Lumley, 2004). Os resultados dos testes mostraram associação estatisticamente significativa entre sexo e alguns sintomas específicos: mulheres apresentaram maiores proporções de respostas positivas para os itens relacionados a lentidão/agitação ($p = 0,03$), sentimento de fracasso ($p = 0,0007$) e pensamento suicida ($p < 0,0001$).

Palavras-chave: Saúde Mental; PHQ-9; Amostragem Complexa; PNS-2019.

Lumley, T. (2004). Analysis of complex survey samples. Journal of Statistical Software, 9(1), 1–19. <https://doi.org/10.18637/jss.v009.i01>

¹PPG em Biometria, Unesp, Campus de Botucatu – daisy.santana@unesp.br ²Departamento de Biodiversidade e Bioestatística, IBB, Unesp, Campus de Botucatu – liciana.silveira@unesp.br 294



Análise Abrangente do Covid-19 no Município de Teodoro Sampaio/SP: Um estudo retrospectivo e mapeamento epidemiológico setorial

Ilda Barbosa Garbim ¹

Raul Borges Guimarães ²

Resumo

A pandemia de COVID-19, causada pelo vírus SARS-CoV-2, tem sido um desafio global sem precedentes, afetando todos os aspectos da vida humana. No Brasil, a situação não é diferente, com o país enfrentando enormes desafios para controlar a propagação do vírus e minimizar seu impacto na saúde pública. Este estudo se concentra no município de Teodoro Sampaio, localizado no estado de São Paulo. A escolha do município de Teodoro Sampaio/SP se deve à necessidade de compreender melhor a dinâmica da COVID-19 em diferentes contextos geográficos e demográficos dentro do Brasil. O objetivo deste projeto é realizar uma análise abrangente da COVID-19 em Teodoro Sampaio, utilizando uma abordagem retrospectiva entre o período de janeiro de 2020 a julho de 2023 e um mapeamento epidemiológico setorial. Pretendemos explorar a progressão da doença, identificar padrões e tendências, e entender os fatores que podem ter contribuído para a disseminação do vírus no município. Esperamos que os resultados deste estudo possam contribuir para a compreensão da pandemia em nível local e fornecer insights valiosos para informar estratégias futuras de prevenção e controle.

Palavras-chave: COVID-19; Mapeamento Epidemiológico;Teodoro Sampaio/SP 1.

¹ Geografia, Unesp-Presidente Prudente – ilda.garbim@unesp.br

² Geografia, Unesp-Presidente Prudente – raul.guimaraes@unesp.br



Distribuição Seno Burr XII Unitária: Uma Aplicação à Vulnerabilidade Climática Mundial

Leonardo Prior Migliorini;¹ Caroline Lopes Gonçalves² Cleber Bisognin³

Resumo

Este trabalho propõe a distribuição Seno Burr XII Unitária (SUBXII), denotada por $\text{SUBXII}(\alpha, \beta)$, para $\alpha, \beta > 0$, como alternativa para modelagem de variáveis contínuas restritas ao intervalo (0,1). A distribuição baseia-se no gerador Seno-G tendo como distribuição base a distribuição Burr XII unitária, permitindo maior flexibilidade na forma das funções de densidade de probabilidade e distribuição acumulada. Por meio da estrutura GAMLSS (Generalized Additive Models for Location, Scale and Shape), incorporam-se estruturas de regressão nos parâmetros, através de funções de ligação, isto é, $g(\theta_{it}) = \mathbf{x}_{it}^\top \boldsymbol{\beta}_i$, em que $t = 1, \dots, n$, em que $g_i(\cdot) = \log(\cdot)$ é a função de ligação, com $i \in \{1, 2\}$, $\boldsymbol{\theta} = (\alpha, \beta)$. Para ambos os modelos, foram propostos métodos de estimação dos parâmetros, mais especificamente, o estimador pela máxima verossimilhança (EMV). A performance do EMV foi avaliada realizando-se simulações de Monte Carlo, considerando métricas como viés, erro quadrático médio, coeficientes de assimetria e curtose. Analisando os resultados das simulações de Monte Carlo, percebe-se que, à medida que o tamanho amostral aumenta, a média converge para o verdadeiro valor do parâmetro, enquanto o desvio padrão, o viés, o erro quadrático médio e o coeficiente de assimetria tendem a zero, e a curtose tende a três. Com base nesses resultados, conclui-se que o EMV apresenta indícios de ser assintoticamente não viésado, consistente e de seguir uma distribuição assintoticamente normal. A aplicabilidade da SUBXII foi testada comparando-a com distribuições consagradas, utilizando dados do FMI sobre vulnerabilidade climática em 192 países. A nova distribuição mostrou ajuste superior aos modelos tradicionais, destacando-se como uma alternativa robusta para dados unitários.

Palavras-chave: Distribuição SUBXII; Variáveis no intervalo unitário; GAMLSS; Máxima verossimilhança; Aplicação.

Trabalho apoiado pelo programa PIBIC-CNPq.

¹Curso de Bacharelado em Estatística, UFSM - leonardo.migliorini@acad.ufsm.br

²Pós-Graduação em Estatística e Modelagem Quantitativa – Especialização, UFSM - caroline.goncalves@acad.ufsm.br

³Departamento de Estatística, UFSM - cleber.bisognin@ufsm.br



Robust spectral discriminant analysis: an M-quantile periodogram approach

Patrick Ferreira Patrocínio;¹ Valderio Anselmo Reisen;² Wei Biao Wu;³ Pascal Bondon

4

Resumo

This paper presents a robust spectral discriminant analysis using M-quantile periodogram function. It combines M- and quantile regression to obtain robust cepstral coefficients. We analyse scenarios in which the data contains outliers and conditional heteroscedasticity. Empirical properties of the M-quantile periodogram are discussed to support the suggested methodology. Monte Carlo tests demonstrate that the M-quantile cepstral approach, as opposed to the multitaper cepstral method, is accurate for processes with non-constant conditional variance and more resistant to additive outliers. The real scenario considers PM₁₀ air pollution data sets to demonstrate the usefulness of the proposed methodology.

Palavras-chave: M-quantile, cepstral, discriminant, robust, heteroscedastic processes.

¹Federal University of Espírito Santo, Vitoria/ES - patrick.ferreira-patrocínio@centralesupelec.fr

²Federal University of Espírito Santo, Vitoria/ES - valderioanselmoreisen@gmail.com

³University of Chicago, Illinois - wbwu@uchicago.edu

⁴Université Paris-Saclay, CentraleSupélec, Gif-sur-Yvette - pascal.bondon@centralesupelec.fr



Estudo Preditivo da Dengue com Base em Séries Temporais Climáticas e Epidemiológicas

Alexandre Schleder Gonçalves Kobus ¹

Amanda Merian Freitas Mendes ²

Kelly Durski ³

Luciana Schleder Gonçalves ⁴

Resumo

A dengue é uma doença endêmica no Brasil, cuja incidência é influenciada por variáveis climáticas. A análise de dados epidemiológicos e climáticos é essencial para prever surtos e embasar medidas preventivas. Este estudo propõe um sistema automatizado para geração de relatórios epidemiológicos no R, incluindo informações sobre casos prováveis, confirmados, descartados e graves, utilizando dados de notificação da dengue. Foi desenvolvido um modelo preditivo para investigar a relação entre variáveis climáticas (temperatura, precipitação e umidade) e o aumento dos casos da doença. A metodologia baseou-se na coleta de dados epidemiológicos do DATASUS e dados climáticos de fontes públicas, restritos ao estado do Paraná, seguida de uma análise estatística exploratória para identificar padrões sazonais e correlações. Utilizou-se modelagem estatística para estimar a defasagem temporal entre condições climáticas críticas e o aumento expressivo de casos. Os resultados indicaram correlações significativas entre fatores climáticos e a incidência de dengue, permitindo antever períodos de risco. Foi desenvolvido um dashboard interativo que automatiza a visualização e atualização dos dados, fornecendo informações acessíveis a gestores e população. O sistema aprimora a gestão da saúde pública ao possibilitar tomadas de decisão baseadas em evidências, reforçando estratégias preventivas nos períodos mais críticos.

Palavras-chave: Dengue; Modelagem Preditiva; Séries Temporais; Previsão de Surtos; Análise Climática

¹ Departamento de Estatística, UFPR – alexandre.schleder@ufpr.br

² Departamento de Estatística, UFPR – amandamerian@ufpr.br

³ Secretaria de Estado de Saúde, 3^a Região - Ponta Grossa - PR – kellydurski@gmail.com

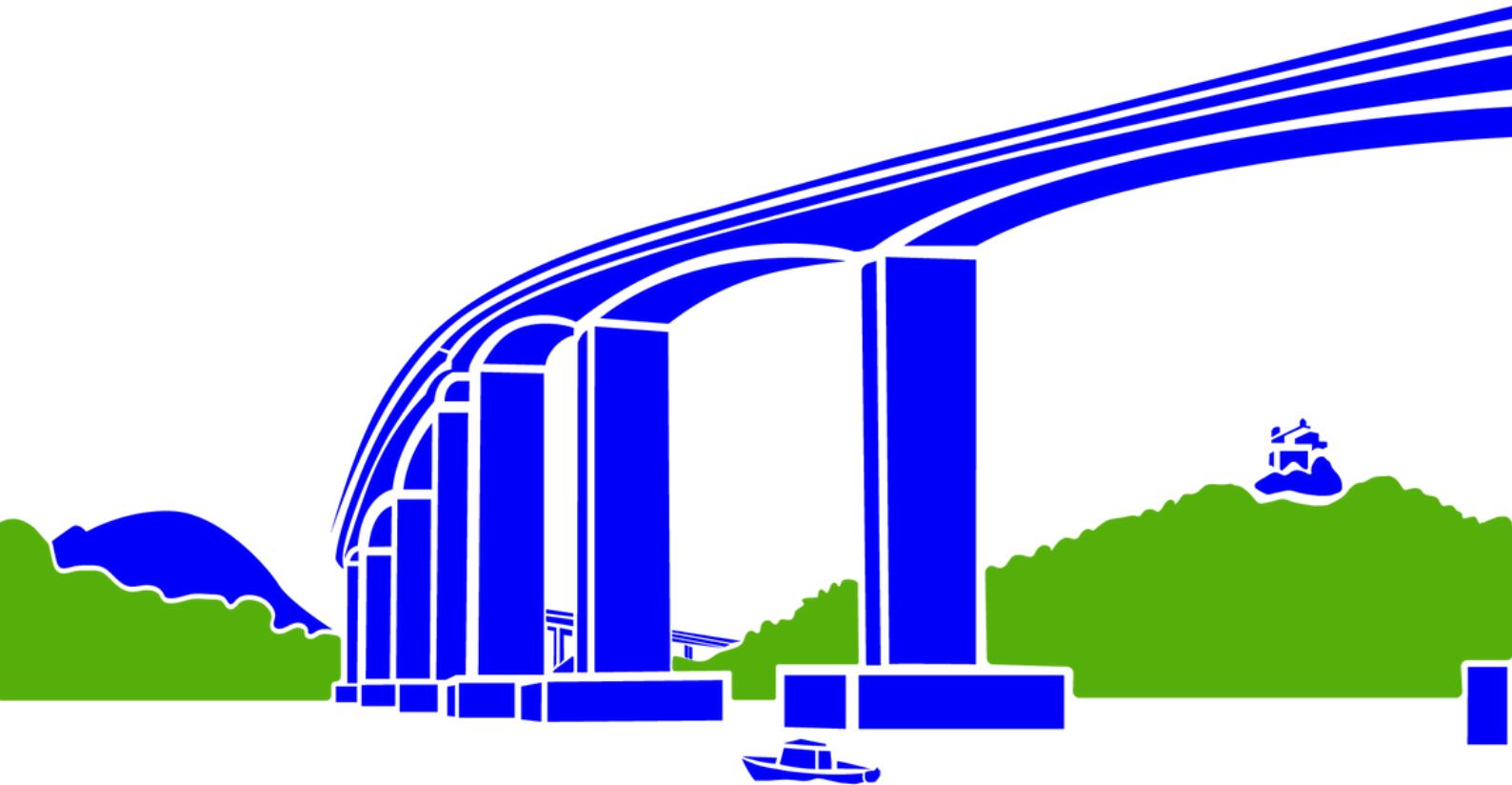
⁴ Departamento de Enfermagem, UFPR – lsgk@ufpr.br

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Pôsters

**Estatística Aplicada em
Arquitetura e Urbanismo,
Engenharias e Ciências Exatas**





Análise multivariada dos determinantes da produtividade na produção de ferro manganês alto carbono em uma indústria siderúrgica

Gislaine Cristina Batistela ¹ Fernando Hiromitus ² Pedro Augusto Rocha Reis ³

Resumo

A indústria siderúrgica enfrenta o desafio constante de otimizar a produtividade e utilizar recursos de forma eficiente, sendo fundamental compreender como matérias-primas, custos e consumo energético influenciam o desempenho industrial. Neste contexto, este estudo teve como objetivo analisar as variáveis que influenciam o desempenho do processo de produção de ferroliga, com foco na gestão e na produtividade. A base de dados utilizada refere-se ao processo de produção de ferro manganês alto carbono em uma indústria siderúrgica do Sudoeste Paulista, composta por registros diários de um forno do tipo elétrico de redução a arco entre junho de 2023 e maio de 2024. Foram consideradas informações sobre consumo de insumos, variáveis operacionais, composição das matérias-primas e indicadores produtivos, como: consumo específico de energia (*CEE*), consumo específico de carbono (*CEC*), consumo específico de manganês (*CEMn*) e produção líquida. Os resultados, por meio de análise de agrupamentos e regressão múltipla, indicam que a elevação sustentável da produção líquida está associada ao aprimoramento da composição química da carga, como óxido de cálcio, óxido de magnésio, dióxido de silício, dióxido de titânio, óxido de alumínio, ferro metálico e cinzas. Os indicadores *CEE*, *CEC* e *CEMn* mostraram-se fortemente influenciados pela escolha das matérias-primas, equilíbrio dos óxidos nas escórias e controle dos insumos redutores, impactando diretamente na eficiência e produtividade do processo industrial. Além disso, valorizar práticas como o consumo de finos de liga excedente ao lingotamento (refusão), controlar insumos energéticos e redutores e equilibrar variáveis operacionais contribuem para maximizar a produtividade.

Palavras-chave: Análise de Agrupamentos; Regressão Múltipla; Ferroliga; Escória.

¹Departamento de Ciências e Tecnologia, Instituto de Ciências e Engenharia, UNESP, Câmpus de Itapeva - gislaine.batistela@unesp.br

²Departamento de Produção, Faculdade de Engenharia e Ciências, UNESP, Câmpus de Guaratinguetá - fernando.hiromitus@unesp.br

³Instituto de Ciências e Engenharia, UNESP, Câmpus de Itapeva - pedro.r.reis@unesp.br



Por que tão poucas? Análise do Efeito Tesoura e *Leaky Pipeline* na carreira de mulheres pesquisadoras na Engenharia.

Iasmin Ferreira Nery ¹ Julia Castro Mendes ² Camila Borelli Zeller ³ Gabriela Paschoal da Silva ⁴

Resumo

A desigualdade de gênero na academia, especialmente nas áreas de Ciências, Tecnologias, Engenharias e Matemática (STEM), é uma realidade persistente, marcada por vieses e barreiras estruturais que dificultam a ascensão de mulheres. Esta pesquisa, parte do Programa PDPG – Alteridade na Pós-Graduação, promovido pela CAPES, foca na presença feminina nos programas de pós-graduação (PPGs) de engenharia no Brasil, com o objetivo de quantificar a participação das mulheres e analisar fenômenos como o “efeito tesoura” e o “Leaky pipeline” — que representam a redução progressiva da presença feminina ao longo da carreira acadêmica. O estudo abrange 13 cursos de engenharia, classificados nas subáreas Engenharias I, II, III e IV pela CAPES, além de um estudo de caso sobre a produção científica de bolsistas de produtividade (PQs) na Engenharia Civil. As perguntas centrais investigam se há parâmetros que favorecem classificações desiguais e se homens são mais frequentemente sobreclassificados em relação às mulheres. Utilizando dados da Plataforma Sucupira (CAPES) desde 2013 e dos editais PQ do CNPq a partir de 2017, a análise envolveu a identificação de docentes e discentes por meio de ferramentas como GenderAPI e ChatGPT, já que os relatórios oficiais não informam o gênero dos pesquisadores. O propósito é diagnosticar o cenário atual da desigualdade de gênero na pós-graduação em engenharia, identificar pontos críticos na carreira, mapear diferenças regionais e subáreas, além de entender o impacto da produtividade acadêmica na ocupação de cargos. O estudo visa contribuir para ambientes mais equitativos e inclusivos na ciência brasileira.

Palavras-chave: Análise Estatística; Engenharia; Gênero; Mulheres na STEM; Pós-Graduação.

¹Departamento de Engenharia Civil, UFJF – iasmin.nery@estudante.ufjf.br

² Departamento de Engenharia Civil, UFJF – juliastro_mendes@ufjf.br

³ Departamento de Estatística, UFJF – camila.zeller@ufjf.br

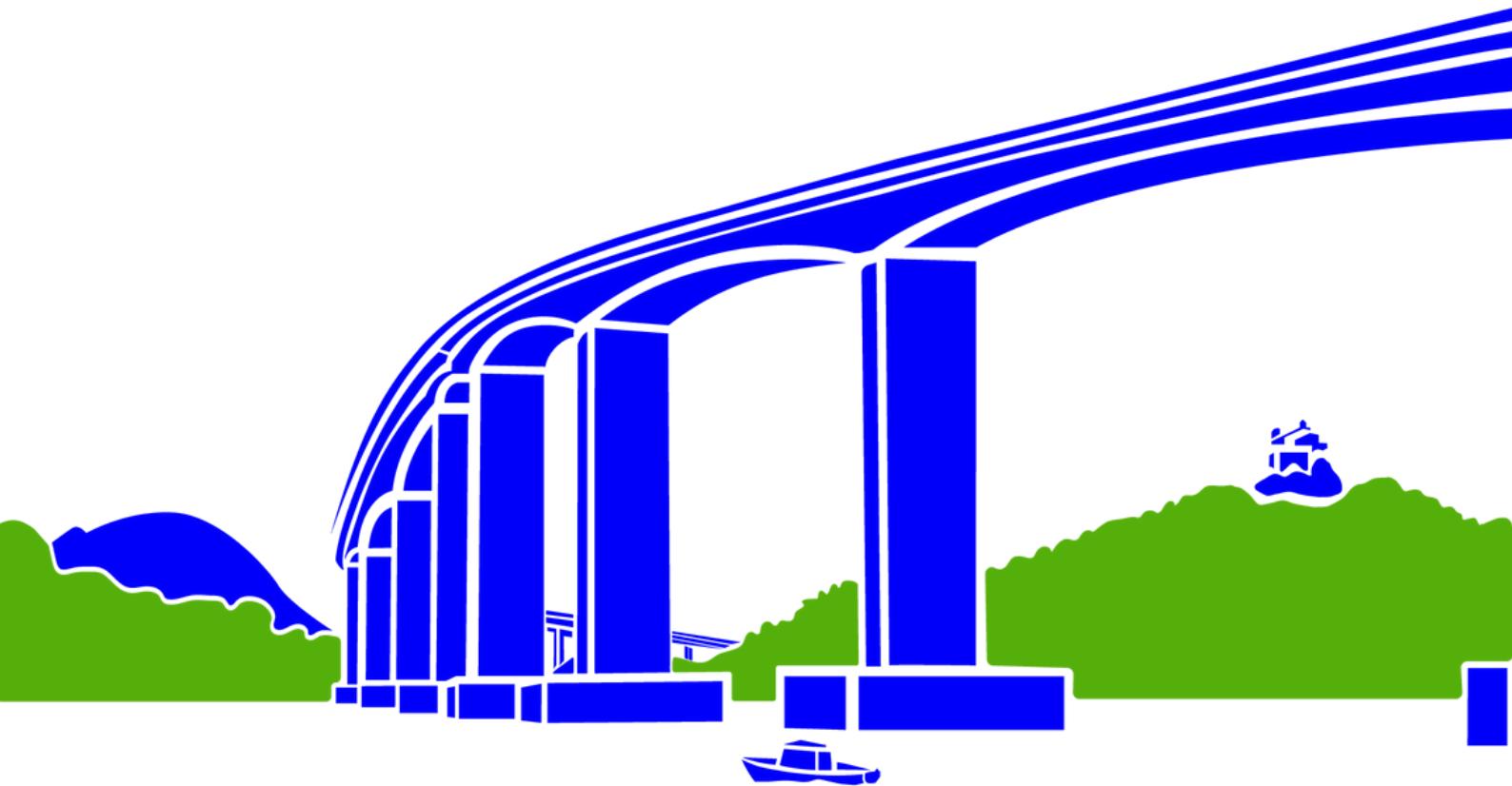
⁴ Departamento de Estatística, UFJF – gabriela.paschoal@estudante.ufjf.br

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Pôsters

Educação Estatística





Desafios da estatística aplicada na análise dos efeitos da temperatura na toxicidade de microplásticos e cobre em organismos aquáticos

Alfredo José Barreto Luiz ¹

Claudio Martín Jonsson; Júlio Ferraz de Queiroz; Márcia Mayumi Ishikawa; Vera Lúcia Scherholz Salgado de Castro ²

Resumo

O trabalho apresenta situações em que a estatística aplicada enfrenta desafios para obedecer aos princípios básicos da experimentação diante dos obstáculos do ambiente experimental. Apresentamos dois experimentos realizados como parte do projeto de pesquisa apoiado pela Fapesp (2023/14504-2) que visa avaliar os efeitos da mudança climática na toxicidade de microplásticos em peixes. Ecossistemas aquáticos sofrem influência de fatores ambientais, como temperatura e exposição aos raios UV, e das atividades antrópicas, como compostos orgânicos e inorgânicos, cuja interação pode produzir efeitos aditivos, sinérgicos ou antagônicos. Executar experimentos controlados que reproduzam essas interações, possivelmente presentes no ambiente natural, encontra inúmeras dificuldades. No primeiro experimento foi utilizado o peixe zebra (*Danio rerio*) na fase embrião-larval, mantido sob três temperaturas em microplacas de 24 poços, submetidos a quatro tratamentos: controle; microplástico; cobre; e microplástico mais cobre. Como não é possível manter poços de uma microplaca em temperaturas diferentes, foi necessário repetir o experimento para cada uma das três e, depois, verificar a possibilidade da análise conjunta. Outro experimento, com tilápia (*Oreochromis niloticus*), realizado em aquários com 120 litros de água, testou o efeito de quatro doses de cobre mais o controle. A dificuldade foi a impossibilidade de recirculação da água e sua filtragem, pois interferiria na concentração do metal (causa de variação sob análise). Então, o experimento foi conduzido em sistema estático, com renovação frequente e parcial de água, o que exigiu análises recorrentes para medir as concentrações reais de cobre em cada aquário. São apresentados detalhes dos problemas e das abordagens adotadas.

Palavras-chave: Ensino de estatística; aleatoriedade; repetição; controle local; análise conjunta de experimentos.

¹ Embrapa Meio Ambiente, Jaguariúna, SP – alfredo.luiz@embrapa.br

² Embrapa Meio Ambiente, Jaguariúna, SP



Tradução do livro “R for Data Science” (2^a Edição) feita pela comunidade brasileira

Beatriz Milz ¹

Resumo

Apresentamos a iniciativa da comunidade de R de traduzir para o português brasileiro a segunda edição do livro “R for Data Science”, originalmente de Hadley Wickham, Mine Çetinkaya-Rundel e Garrett Grolemund. O objetivo do projeto é disponibilizar o livro de forma gratuita e em português, reduzindo a barreira linguística que dificulta o acesso aos recursos em inglês para o aprendizado de programação. Mais de 20 pessoas voluntárias da comunidade de R no Brasil e na América Latina se mobilizaram nessa tradução comunitária. Cada capítulo contou com a contribuição de uma pessoa tradutora e duas revisoras. Ferramentas do GitHub, como Issues, Pull Requests e Boards, foram essenciais para gerenciar o processo, enquanto um guia de tradução facilitou a adaptação dos conteúdos mesmo para as pessoas que não tinham experiência prévia com a ferramenta. A experiência da comunidade latino-americana, que já havia traduzido a primeira edição para o espanhol, também foi fundamental para o projeto – exemplificada pelo pacote `{dados}` (disponível no CRAN), “irmão” do `{dados}`, que disponibiliza as bases de dados utilizadas no livro. O livro está disponível gratuitamente online em: <https://pt.r4ds.hadley.nz/>

Palavras-chave: programação em R; tidyverse; ensino; CRAN; análise de dados.

¹ Universidade Federal do ABC (UFABC), São Bernardo do Campo – milz.bea@gmail.com



Material Didático Interativo para o Ensino e Aprendizado de Distribuições de Probabilidade

MOREIRA, D. A. P.¹

BASTOS, F. S.²

Resumo

A Estatística é uma ciência de grande relevância por possibilitar a análise de dados e a identificação de padrões, elementos fundamentais para a tomada de decisões e a resolução de problemas do cotidiano. No entanto, seu ensino ainda enfrenta desafios, especialmente devido à escassez ou à pouca divulgação de materiais didáticos que adotem estratégias eficazes de aprendizagem. Em cursos introdutórios de Estatística, por exemplo, o estudo das distribuições de probabilidade costuma se restringir a poucas delas, sem a devida exploração de suas representações gráficas, o que pode dificultar a compreensão por parte dos alunos. Nesse contexto, este trabalho apresenta o desenvolvimento de um material didático interativo voltado para o ensino de Distribuições de Probabilidade, fundamentado na premissa de que a interatividade favorece o aprendizado e contribui para a assimilação dos conceitos estatísticos. O material contempla 14 distribuições de probabilidade, incluindo distribuições discretas e contínuas, e apresenta definições teóricas, os principais parâmetros de cada distribuição e algumas medidas estatísticas associadas. O objetivo é disponibilizar um recurso pedagógico inovador, que auxilie professores e estudantes na visualização dinâmica e na compreensão do comportamento gráfico das distribuições à medida que seus parâmetros são modificados. Além disso, foi realizada uma revisão da literatura com o intuito de identificar lacunas no ensino do tema e embasar o desenvolvimento da ferramenta proposta. Como parte dos resultados, também foram fornecidos códigos detalhados que permitem a construção de novos materiais semelhantes, tornando o processo de aprendizagem mais acessível e eficaz.

Palavras-chave: R; Ensino de Estatística; Gráficos Interativos; Ensino Interativo; Pacote learnr.

¹ Discente do Programa de Pós Graduação em Estatística Aplicada e Biometria , Viçosa – darah.moreira@ufv.br

² Professor do Programa de Pós Graduação em Estatística Aplicada e Biometria, Viçosa – fernando.bastos@ufv.br



Análise do Desempenho Acadêmico dos Estudantes do Curso de Estatística/UFSCar a partir de 2018

Emily Alves Franco ¹

Isabela Borges Farina ²

Isabella Viana Bambirra ³

Luan Tomimoto ⁴

Lucas de Andrade Marin ⁵

Murilo Cassiavilani ⁶

Rafael Coelho Soares Reis ⁷

Márcio Luis Lanfredi Viola ⁸

Resumo

O Projeto Pedagógico de um curso de graduação é o documento que estabelece as diretrizes para o seu funcionamento, em particular, é nela que são definidas a matriz curricular de um curso. Diante disso, é necessário que um Projeto Pedagógico seja constantemente atualizado devido, principalmente, a mudanças no perfil dos estudantes ingressantes e para atender às necessidades da sociedade com relação à formação do profissional. Especificamente, a atualização do Projeto Pedagógico de um Curso de Bacharelado em Estatística deve ser um processo contínuo, considerando o uso de conjuntos de dados grandes e complexos, além de métodos de aprendizado de máquina. A última atualização do Projeto Pedagógico do Curso de Bacharelado em Estatística da UFSCar começou a vigorar em 2018 e está em discussão a sua atualização. A partir de 2013, o grupo do Programa de Educação Tutorial (PET) Estatística da UFSCar passou a acompanhar os dados acadêmicos de todos os discentes ingressantes no curso a partir da implantação do PPC em 2006. O objetivo deste trabalho é avaliar o desempenho acadêmico dos estudantes egressos e dos evadidos a partir de 2018.

Palavras-chave: Análise descritiva; Análise de correspondência; Curso de Bacharelado em Estatística; Estatística; Projeto Pedagógico.

¹ Departamento de Estatística, Universidade Federal de São Carlos – emily.franco@estudante.ufscar.br

² Departamento de Estatística, Universidade Federal de São Carlos – isabelafarina@estudante.ufscar.br

³ Departamento de Estatística, Universidade Federal de São Carlos – isabellavb@estudante.ufscar.br

⁴ Departamento de Estatística, Universidade Federal de São Carlos – luantomimoto@estudante.ufscar.br

⁵ Departamento de Estatística, Universidade Federal de São Carlos – lucasmarin@estudante.ufscar.br

⁶ Departamento de Estatística, Universidade Federal de São Carlos – muriloc@estudante.ufscar.br

⁷ Departamento de Estatística, Universidade Federal de São Carlos – rafael.reis@estudante.ufscar.br

⁸ Departamento de Estatística, Universidade Federal de São Carlos – lanfredi@ufscar.br



Uso de Monte Carlo para a Previsão de Resultados do Campeonato Brasileiro

Emily Alves Franco ¹

Lucas de Andrade Marin ²

Rafael Coelho Soares Reis ³

Thiago Rafael de Souza ⁴

Luis Ernesto Bueno Salasar ⁵

Márcio Luis Lanfredi Viola ⁶

Resumo

Dados têm sido cada vez mais importantes e usados para a tomada de decisão em times e ligas esportivas. Um caso clássico é o do dirigente Billy Beane, que conseguiu levar o Oakland Athletics às fases finais da MLB, visando encontrar atletas baratos e fundamentais para o time. Atualmente, modelos de aprendizado de máquinas também estão sendo usados pelos times na tomada de decisão. O uso de dados e análises estatísticas têm sido usados em diversas modalidades, principalmente nos Estados Unidos e Europa. Os clubes europeus utilizam, com muito frequência, ferramentas estatísticas no futebol. Especificamente, no futebol, é de interesse prever quantos gols serão marcados em uma partida ou ser capaz de prever se um determinado time ganhará a partida. Neste sentido, modelos baseados na distribuição Poisson são explorados na modelagem de dados esportivos como, por exemplo, o uso do modelo Poisson bivariado e suas extensões. Este trabalho tem por objetivo o uso de simulações de Monte Carlo para prever a probabilidade de vitória, empate ou derrota de um time do Campeonato Brasileiro no ano de 2024 usando as informações dos times no campeonato do ano de 2023, comparando duas estratégias para simular os possíveis placares de uma partida: (i) distribuições independentes de Poisson e; (ii) distribuição bivariada de Holgate. Utilizamos o método dos momentos para a estimativa dos parâmetros da distribuição Poisson e, para avaliar a qualidade preditiva dos modelos, o escore de Brier, que é uma métrica comumente usada na literatura.

Palavras-chave: Distribuição de Poisson; Distribuição de Poisson bivariada; Método de Monte Carlo; Previsão esportiva; Futebol.

¹ Departamento de Estatística, Universidade Federal de São Carlos - emily.franco@estudante.ufscar.br

² Departamento de Estatística, Universidade Federal de São Carlos - lucasmarin@estudante.ufscar.br

³ Departamento de Estatística, Universidade Federal de São Carlos - rafael.reis@estudante.ufscar.br

⁴ Departamento de Estatística, Universidade Federal de São Carlos - rthiago102@gmail.com

⁵ Departamento de Estatística, Universidade Federal de São Carlos - luis@ufscar.br

⁶ Departamento de Estatística, Universidade Federal de São Carlos - lanfredi@ufscar.br



Ciência de dados e Estatística: Uma comparação descritiva dos cursos de graduação em Universidades Brasileiras

Anderson José de Souza Inácio ¹

Marco Antonio Silva Guibor ²

Anderson Ara ³

Resumo

Com o advento da popularização do uso de sistemas de inteligência artificial generativo, a busca por profissionais qualificados do ramo da ciência de dados aumentou exponencialmente. A fim de atender essa necessidade, instituições Brasileiras de Ensino Superior públicas e privadas, têm buscado melhorar seu portfólio com novos cursos e disciplinas que envolvam técnicas para manipulação de dados. A área de Ciência de Dados é recente no Brasil, em 2021 ela foi adicionada no Ofício Circular 0714111/2021/CGCES/DEED-INEP como um curso da área de computação, já em 2023 a CBO (Classificação Brasileira de Ocupações) inseriu a ocupação de Cientista de Dados, como pertencente a profissionais da área de Estatística. Com dados coletadas do portal do e-MEC (<https://emecc.mec.gov.br/>), junto com os do Censo da Educação Superior de 2023 feita pela INEP, foi identificado um aumento em cerca de 97 cursos com o nome Ciência de Dados e similares, entre o período de 2018 e 2024, sejam em formato presencial ou a distância (EAD). Com base nisso, o objetivo deste trabalho é realizar uma comparação descritiva dos cursos de Ciência de Dados e Estatística, fornecendo um diagnóstico sobre a formação destes cursos de graduação no Brasil. As principais questões abordadas são: quais as diferenças e semelhanças entre os cursos de Estatística e Ciência de Dados, como ambos cursos estão organizados geograficamente e qual é a modalidade (Presencial ou EAD) mais ofertada.

Palavras-chave: Estatística; Ciência de Dados; Universidades Brasileiras; Educação Estatística; Análise Descritiva

¹Departamento Estatística, UFPR - anderson.inacio@ufpr.br

²Departamento Estatística, UFPR - marcoguibor@ufpr.br

³Departamento Estatística, UFPR - ara@ufpr.br



Fatores sociais que influenciam as notas de Matemática e suas Tecnologias no ENEM 2023 em Recife, uma modelagem GAMLSS

Elisa Lopes do Nascimento;¹ Fernanda de Oliveira Lima² Josimar Mendes de Vasconcelos;¹ **Resumo**

Este estudo investigou os fatores que influenciam o desempenho de estudantes de Recife na prova de Matemática e suas Tecnologias do ENEM em 2023, com ênfase em variáveis demográficas e escolares, como sexo, raça, faixa etária e tipo de escola (pública ou particular). O ENEM, criado em 1998, passou a ser utilizado como instrumento de acesso ao ensino superior a partir de 2009 (INEP, 2024). A partir dos microdados disponibilizados pelo INEP, foram aplicadas técnicas de Regressão Linear Múltipla, Modelos Lineares Generalizados (Cordeiro et al., 2024) e Modelos Aditivos Generalizados para Localização, Escala e Forma (GAMLSS), conhecidos por sua flexibilidade na modelagem de dados (Ribeiro et al., s.d.; Repositório Institucional UFC, 2021). O modelo GAMLSS, com distribuição Skew Exponential Power Type 3, apresentou o melhor ajuste, evidenciado por resíduos aderentes e bons indicadores de normalidade. Os resultados refletem desigualdades estruturais no acesso à educação, como apontado por estudos que revelam defasagens de aprendizado e disparidades entre as redes pública e privada (IEDE, 2023; G1, 2011). Essas desigualdades foram ainda mais acentuadas pelos impactos da pandemia da COVID-19.

Palavras-chave: GAMLSS; ENEM 2023; Matemática e suas Tecnologias; Modelagem.

References

- INEP (2024). Microdados enem. Acesso em: 05 dez. 2024.
- Cordeiro, G. M., C. G. B. Demétrio, and R. A. Moral (2024). *Modelos Lineares Generalizados e Aplicações*. São Paulo: Blucher.
- G1 (2011). 25% das crianças do 3º ano da rede privada sabem matemática do 5º ano. Acesso em: 05 dez. 2024.
- IEDE (2023). Relatório de indicadores de aprendizagem. Acesso em: 05 dez. 2024.
- Repositório Institucional UFC (2021). Modelos gammss: fundamentos e aplicações. Universidade Federal do Ceará. Acesso em: 05 dez. 2024.
- Ribeiro, M. L. F., L. S. d. Oliveira, and J. M. d. Vasconcelos (s.d.). Modelos gammss: Aplicação em casos de síndrome respiratória aguda grave com ênfase em influenza e outras causas. Manuscrito não publicado ou sem data.

¹Departamento de Estatística e Informática, Recife, PE - Elisa.lopes@ufrpe.br

²Departamento de Estatística e Informática, Recife, PE - fernanda.lima@ufrpe.br

¹Departamento de Estatística e Informática, Recife, PE - josimar.mendes@ufrpe.br



Visualização de distribuições conjugadas e interpretação da posteriori em termos de área

Thiago Rafael de Souza ¹

Giovanna Martins Verasto ²

Maria Julia Gregório Romero ³

Pedro Henrique Duarte ⁴

Márcio Alves Diniz ⁵

Márcio Luis Lanfredi Viola ⁶

Resumo

Em geral, estudantes de Estatística possuem dificuldades no aprendizado de conceitos envolvendo Estatística Bayesiana. Uma maneira para a abordagem Bayesiana se tornar mais intuitiva aos estudantes é utilizar ferramentas simples que permitem a visualização e interpretação de tais conceitos. Por exemplo, a distribuição a posteriori pode ser interpretada usando a noção de área no caso em que os estados da natureza e ações são finitos. O objetivo deste trabalho é apresentar duas ferramentas didáticas desenvolvidas pelo grupo PET (Programa de Educação Tutorial) Estatística/UFSCar que podem ser usadas no ensino de conceitos Bayesianos: (i) aplicativo que permite a interpretação do Teorema de Bayes e, consequentemente, da posteriori usando razão de áreas em um quadrado unitário, cujos lados são associados às priori e verossimilhança; (ii) página interativa que permite a visualização de distribuições conjugadas.

Palavras-chave: Distribuição a posteriori; Distribuição a priori; Distribuições conjugadas; Estatística Bayesiana; Educação.

¹ Departamento de Estatística, Universidade Federal de São Carlos – thiagorafael@estudante.ufscar.br

² Departamento de Estatística, Universidade Federal de São Carlos – gmverasto@estudante.ufscar.br

³ Departamento de Estatística, Universidade Federal de São Carlos – mariaromero@estudante.ufscar.br

⁴ Departamento de Estatística, Universidade Federal de São Carlos – pedrohd@estudante.ufscar.br

⁵ Departamento de Estatística, Universidade Federal de São Carlos – marciodiniz@ufscar.br

⁶ Departamento de Estatística, Universidade Federal de São Carlos – lanfredi@ufscar.br



Paradoxo de Simpson: Aplicação em Dados do Campeonato Brasileiro de Futebol

Andressa Nascimento Santos ¹

Isabela Lika Morioka ²

Vinicius Cotrim Cruz ³

Márcio Luis Lanfredi Viola ⁴

Resumo

Em vários estudos, o estatístico deve preocupar-se com fatores cujo efeito pode se confundir com o de variáveis de interesse. Especificamente, em pesquisas que buscam verificar a associação entre duas variáveis, X e Y, devemos ter atenção às variáveis secundárias, que podem interferir nessa associação, alterando apenas a magnitude dos efeitos de X e Y ou tanto a magnitude quanto o sentido da relação entre elas. Nesse caso, a variável secundária é uma variável de confundimento (Z), e podemos obter conclusões diferentes a respeito da associação entre X e Y, dependendo se a variável de confundimento Z é considerada ou não na análise estatística. Isto é, o sentido da relação entre as duas variáveis pode ser influenciado pela presença de um terceiro fator, pois os efeitos de X e Y podem estar confundidos com o da variável de confundimento (fator de confusão), afetando a conclusão final. Esse fenômeno é conhecido como Paradoxo de Simpson, o qual ocorre quando a natureza de uma associação pode mudar (ou até mesmo se inverter) ou desaparecer ao se combinar dados de diferentes grupos em um único conjunto, ou ao se separá-los. Este trabalho tem por objetivo ilustrar o Paradoxo de Simpson em dados relacionados ao futebol, utilizando informações do Campeonato Brasileiro de 2023 e 2024.

Palavras-chave: Campeonato Brasileiro; Confundidores; Futebol; Paradoxo de Simpson.

¹Departamento de Estatística, Universidade Federal de São Carlos – andressans@estudante.ufscar.br

²Departamento de Estatística, Universidade Federal de São Carlos – isabela.lika@estudante.ufscar.br

³Departamento de Estatística, Universidade Federal de São Carlos – viniciuscotrim@estudante.ufscar.br

⁴ Departamento de Estatística, Universidade Federal de São Carlos – lanfredi@ufscar.br



Análise de fatores que afetam a conclusão dos estudantes do curso de bacharelado em Estatística da Ufes

Fabio Alexander Fajardo Molinares ¹

Profa. Dra. Nátaly A. Jiménez Monroy ²

Luciana G. de Godoi ³

Resumo

Este trabalho tem como objetivo identificar os principais fatores associados à evasão dos estudantes do Curso de Bacharelado em Estatística da Universidade Federal do Espírito Santo, por meio de uma abordagem quantitativa que utiliza ferramentas estatísticas e computacionais avançadas. As informações, como desempenho acadêmico e histórico escolar, serão extraídas de bases institucionais da Ufes. A análise será conduzida com modelos estatísticos tradicionais e algoritmos de aprendizado de máquina, permitindo a identificação de padrões relevantes e a antecipação de situações com maior risco de evasão. Essa estratégia permitirá a investigação de relações complexas entre variáveis e a construção de perfis detalhados dos estudantes, contribuindo para o desenvolvimento de políticas de permanência mais eficazes. O trabalho busca avaliar dimensões como autoeficácia acadêmica, adaptabilidade de carreira e impacto de aspectos financeiros, entre outros. Os resultados esperados envolvem a identificação de fatores críticos associados à evasão e a formulação de estratégias institucionais mais direcionadas. A originalidade da proposta está na combinação de métodos estatísticos robustos com técnicas de inteligência artificial para análise preditiva, oferecendo subsídios para políticas educacionais voltadas à permanência estudantil e à promoção do sucesso acadêmico.

Palavras-chave: Evasão estudantil; Modelagem estatística; Permanência acadêmica.

¹ Departamento de Estatística, Ufes – fabio.molinares@ufes.br

² Departamento de Estatística, Ufes – nataly.monroy@ufes.br

³ Departamento de Estatística, Ufes – luciana.godoi@ufes.br



Grupo StatUFSM: Iniciativas de Ensino e Extensão para Divulgar a Ciência Estatística nas Mídias Sociais

Renata Rojas Guerra;¹ Beatriz Woos Buffon;² Laís Helen Loose;³ Moizes da Silva Melo;⁴

Abstract

O projeto "StatUFSM: Estatística para Todos" tem como objetivo divulgar conceitos de estatística e ciência de dados de forma acessível, destacando a importância dessas áreas e a natureza multidisciplinar de suas aplicações. Alinhado à proposta de popularização da estatística, o projeto concentra-se em ações que envolvem indicadores relacionados aos Objetivos de Desenvolvimento Sustentável (ODS). Entre as atividades desenvolvidas pelo StatUFSM, destacam-se a produção e publicação de vídeos educativos no canal do projeto no YouTube, além de postagens relacionadas à estatística e à ciência de dados no Instagram e no LinkedIn, oferecendo conteúdo informativo ao público-alvo. As ações do projeto também incluem apresentações em eventos acadêmicos na Universidade Federal de Santa Maria, criando um ambiente propício para a troca de conhecimentos e experiências. Adicionalmente, a participação em eventos nacionais e internacionais evidencia a projeção global e a relevância do StatUFSM. A atuação em escolas, no âmbito das iniciativas de extensão universitária, demonstra o compromisso do projeto com a educação e a conscientização em diferentes níveis. Essa presença em distintos contextos contribui não apenas para a disseminação do conhecimento, mas também para a construção de colaborações e parcerias acadêmicas em âmbito nacional e internacional. Dessa forma, a divulgação de conteúdos nas mídias sociais colabora para a formação de uma comunidade engajada com a aprendizagem estatística, tornando a estatística mais acessível à sociedade e, assim, alcançando um dos principais objetivos do projeto.

Palavras-chave: Educação científica; Extensão universitária; Instagram; Objetivos de Desenvolvimento Sustentável (ODS).

¹Departamento de Estatística, Universidade Federal de Santa Maria (UFSM), Santa Maria, Rio Grande do Sul, Brasil - renata.r.guerra@ufsm.br

²Departamento de Estatística, Universidade Federal de Santa Maria (UFSM), Santa Maria, Rio Grande do Sul, Brasil - beatriz.buffon@acad.ufsm.br

³Departamento de Estatística, Universidade Federal de Santa Maria (UFSM), Santa Maria, Rio Grande do Sul, Brasil - lais.loose@ufsm.br

⁴Departamento de Estatística, Universidade Federal de Santa Maria (UFSM), Santa Maria, Rio Grande do Sul, Brasil - moizes.melo@ufsm.br



Assessoria Estatística da UFPel: Qualificação de Pesquisas Científicas

Rodrigo de Laforet Padilha Raupp¹; Clause Fátima de Brum Piana²; João Gilberto Corrêa da Silva³; Ana Rita Assumpção Mazzini⁴.

Resumo

A Estatística é um componente fundamental do método científico, oferecendo ferramentas que permitem o planejamento racional da pesquisa, a análise de dados complexos e a obtenção de conclusões confiáveis. A Estatística Experimental, em especial, fornece métodos rigorosos para o planejamento, condução e análise de experimentos, possibilitando a generalização dos resultados e a avaliação da significância dos achados. Apesar de sua importância, o planejamento estatístico é muitas vezes negligenciado, o que pode comprometer os resultados e desperdiçar recursos. A Estatística vai além de uma simples ferramenta de análise, sendo vital para o avanço da ciência e da tecnologia. Seu desenvolvimento contínuo se dá em resposta às demandas da pesquisa científica, tornando imprescindível a atuação de especialistas em todas as fases do processo investigativo. Nesse contexto, destaca-se a atividade de consultoria estatística, essencial em instituições de ensino e pesquisa. Na Universidade Federal de Pelotas (UFPel), essa atuação foi institucionalizada em 1995 com o projeto "Assessoramento Técnico em Estatística Aplicada", oferecendo apoio desde o planejamento até a interpretação dos resultados das pesquisas. O projeto, inicialmente no Instituto de Física e Matemática, foi transferido em 2012 para o Centro de Desenvolvimento Tecnológico (CDTec), onde permanece ativo. Coordenado por professores da área de Estatística e com participação de estudantes, o projeto já auxiliou diversas pesquisas. Este trabalho tem como objetivo descrever as assessorias prestadas pelo grupo, analisando perfis dos consultentes, áreas de atuação e objetivos das pesquisas, além de divulgar as atividades do projeto à comunidade científica.

Palavras-chave: Estatística Experimental; Pesquisa Científica; Assessoramento; Planejamento Experimental; Consultoria.

¹Centro de Desenvolvimento Tecnológico, UFPel - rdlpraupp@inf.ufpel.edu.br

²Centro de Desenvolvimento Tecnológico, UFPel - clause.piana@inf.ufpel.edu.br

³Departamento de Matemática e Estatística, UFPel - jgcs1804@yahoo.com.br

⁴Departamento de Matemática e Estatística, UFPel - amazzini@inf.ufpel.edu.br



Um jogo de memória que explora a variabilidade

Giovani Festa Paludo ¹; Amanda Merian Freitas Mendes ²; Gean Pereira Damaceno ³; Ali William Canaza Cayo ⁴; Alex Monito Nhancololo ⁵; Júlio Sílvio de Sousa Bueno Filho ⁶

Resumo

Durante todo o desenvolvimento das crianças, tanto dentro das escolas quanto fora, existe uma tendência a explorar áreas de raciocínio em que há uma única resposta certa para um problema. Nestes ambientes, as crianças se deparam com várias situações de resolução de problemas, especialmente aqueles relacionados à área de Matemática em que só existem soluções exatas. Por outro lado, a Estatística é a ciência que se preocupa em observar e mensurar a variabilidade existente no mundo. Baseando-se nisso, no presente trabalho nós desenvolvemos um jogo de memória para ser jogado nos anos iniciais em que o foco não está na exatidão, mas na variabilidade. Ao invés de pares de cartas idênticas, foram construídos grupos de cartas parecidas. Os jogadores deverão avaliar se pares de cartas são suficientemente parecidas para serem consideradas do mesmo grupo, sendo que também é apresentado um gabarito para conferência. Os objetos utilizados nas cartas foram as folhas de plantas que são parte de um conjunto de dados utilizado em um projeto de pós-graduação. O material biológico foi digitalizado e foram obtidos os contornos no pacote do R. Em seguida, foram escolhidos contornos de folhas de várias espécies de plantas, buscando selecionar folhas diferentes dentro da mesma espécie. O presente trabalho apresentou uma proposta de atividade que pode ajudar no desenvolvimento da observação da variabilidade presente no mundo. E ainda, serve para apresentar espécies de plantas para as crianças. O jogo apresentado permite trabalhar um dos conceitos fundamentais da estatística: a variabilidade.

Palavras-chave: atividades lúdicas; pensamento probabilístico; ensino de estatística; anos iniciais; ensino de botânica.

¹Doutorando do Programa de Pós Graduação em Estatística e Experimentação Agropecuária, UFLA, MG - gfpaludo@gmail.com

²Professora do Departamento de Estatística - Universidade Federal do Paraná

³Mestrando do Programa de Pós Graduação em Estatística e Experimentação Agropecuária, UFLA, MG

⁴Doutorando do Programa de Pós Graduação em Estatística e Experimentação Agropecuária, UFLA, MG

⁵Doutorando do Programa de Pós Graduação em Estatística da Universidade de São Paulo

⁶Professor do Departamento de Estatística - Universidade Federal de Lavras



Leitura e interpretação de gráficos a partir dos níveis de Curcio

José Victor Neves Gomes¹

Geovane Carlos Barbosa²

Resumo

Este estudo tem por objetivo analisar o desempenho de alunos de um curso técnico na leitura e interpretação de gráficos estatísticos, a partir dos níveis propostos por Curcio (1989). Trata-se de uma pesquisa de natureza qualitativa, na perspectiva diagnóstica e descritiva, que teve como participantes doze alunos do curso técnico noturno de uma instituição pública. O processo de análise dos dados tomou como referência os níveis de compreensão da linguagem gráfica propostos por Curcio (1989), que considera três níveis de compreensão. Os resultados indicaram que os alunos apresentam bom desempenho no nível I (leitura elementar), mas enfrentaram dificuldades nos níveis II (leitura entre os dados) e III (leitura além dos dados), especialmente quando é necessário realizar cálculos matemáticos e uma análise crítica sobre os dados. Por fim, o estudo destaca a importância de uma leitura gráfica de qualidade, com ênfase nas habilidades necessárias para que os alunos possam não apenas compreender os dados, mas também interpretar, refletir e criticar as informações, de modo a exercer sua cidadania em uma sociedade onde a incerteza e a variabilidade estão presentes.

Palavras-chave: Níveis de Curcio; Educação Estatística; Ensino de Estatística; Leitura gráfica

¹ Licenciado em Matemática, Instituto Federal do Espírito Santo, Campus Cachoeiro de Itapemirim – neves7838@gmail.com

² Coordenadoria da Licenciatura em Matemática, Instituto Federal do Espírito Santo, Campus Cachoeiro de Itapemirim – geovane.barbosa@ifes.edu.br

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Pôsters

**Dados Funcionais, Dados em
Alta Dimensão e Aprendizado
Estatístico de Máquinas**





Similaridade de Processos Judiciários Utilizando Processamento de Linguagem Natural

Bruno Gondim Toledo;¹ Thais Carvalho Valadares Rodrigues¹

Resumo

Com objetivo de contribuir com uma das metas de gestão da presidência do Ministro Luís Roberto Barroso de diminuição do acervo de processos de controle concentrado de constitucionalidade do Supremo Tribunal Federal (STF), nesse trabalho foram aplicadas técnicas de processamento de linguagem natural para encontrar processos semelhantes no acervo. Foram utilizadas técnicas de vetorização, como *tokenização* e *Bag of Words*, bem como métricas de similaridade, como distância do cosseno e divergência de Jensen-Shannon, a fim de encontrar semelhanças entre uma petição de entrada e os processos em tramitação no STF. Nesse trabalho também foi desenvolvido um aplicativo *Shiny* para retornar os resultados da busca por processos similares, utilizando técnicas amplamente aceitas na bibliografia e em conformidade com demais tecnologias do Tribunal sempre que possível, a fim de legar uma ferramenta útil e prática aos analistas do Tribunal. Este estudo foi capaz de produzir uma aplicação com bom desempenho e baixo custo computacional utilizando um modelo de *Bag of Words* metrificados sobre a distância do cosseno entre os dados, e, portanto, pode ser uma ferramenta útil em auxiliar no cumprimento desta meta de gestão do Ministro Presidente.

Palavras-chave: Agrupamento, Controle Concentrado de Constitucionalidade, Distância do Cosseno, Supremo Tribunal Federal, Vetorização de Texto.

¹Departamento de Estatística, Universidade de Brasília (UnB) - brunogtoledo96@gmail.com

¹Departamento de Estatística, Universidade de Brasília (UnB) - thaisrodrigues@unb.br



Uso de Redes Neurais na Modelagem de *Credit Scoring*

Nicole Maria Onofre;¹ Márcio Luis Lanfredi Viola;² Estela Maris Pereira Bereta.³

Resumo

Redes Neurais Artificiais são modelos computacionais inspirados no funcionamento e na estrutura do sistema neural biológico, que se ajusta e aprende por meio de erros e acertos. Nesses modelos, as informações são processadas em unidades chamadas neurônios artificiais, onde ocorre o aprendizado. Nos últimos anos, esse método tem sido amplamente utilizado, especialmente em problemas mais complexos de classificação, devido à sua alta capacidade de identificar padrões. No mercado de crédito, um dos principais desafios é a classificação de indivíduos como bons ou maus pagadores, o que auxilia as instituições financeiras na concessão de crédito e diminuição dos riscos. Nesse contexto, os métodos de classificação são extremamente importantes e surgem como ferramenta para ajudar nessa análise. Dessa forma, nosso objetivo foi estudar o comportamento das redes neurais em um problema de classificação binária, analisando como seu desempenho variava conforme o grau de desbalanceamento dos dados. Para isso, simulamos conjuntos de dados financeiros com proporções entre bons e maus pagadores variando de 50/50 a 90/10. Em cada cenário, treinamos e compararamos oito arquiteturas de redes neurais, variando a função de ativação, número de neurônios e camadas. Nossos resultados indicaram que as redes parecem manter um desempenho consistente em todos os cenários, conforme avaliamos por métricas como erro de predição, sensibilidade e especificidade. Além disso, redes mais complexas não apresentaram necessariamente um desempenho superior às mais simples, indicando que arquiteturas menos sofisticadas podem ser eficientes, mesmo em cenários desbalanceados.

Palavras-chave: Redes Neurais; Classificação Binária; Modelagem Estatística; Dados Desbalanceados.

¹Departamento de Estatística, Universidade Federal de São Carlos - nicoleonofre@estudante.ufscar.br

²Departamento de Estatística, Universidade Federal de São Carlos - lanfredi@ufscar.br

³Departamento de Estatística, Universidade Federal de São Carlos - estela@ufscar.br



Comparação de Medidas de Capacidade Preditiva em um Conjunto de Dados com Classes Desbalanceadas

Paulo Justiniano Ribeiro Neto ¹
Anderson Ara ²

Resumo

A escolha apropriada de modelos de aprendizado supervisionado, bem como das métricas utilizadas para avaliar seu desempenho, é fundamental em aplicações com foco preditivo. Este trabalho tem como objetivo comparar diferentes métricas de capacidade preditiva em cenários de extremo desbalanceamento entre as classes da variável resposta. Para esse fim, utilizou-se um conjunto de dados sobre acidente vascular cerebral (AVC), disponibilizado na plataforma Kaggle. O foco principal está na comparação das métricas de desempenho em um cenário severamente desbalanceado, sem a aplicação de técnicas de balanceamento das classes. Os métodos testados incluíram random forest, máquinas de vetores de suporte (SVM), regressão penalizada (GLMnet), k-vizinhos mais próximos (KNN), classificadores bayesianos e árvores de decisão, todos implementados com o pacote caret na linguagem R. A avaliação dos modelos foi realizada por meio de validação cruzada com holdout repetido, visando aumentar a robustez das estimativas de desempenho. Métricas como acurácia, acurácia balanceada, sensibilidade, especificidade e F1-score foram utilizadas para comparar os modelos. Por fim, discute-se a relevância da escolha adequada das métricas de desempenho, destacando como diferentes medidas podem levar a conclusões distintas sobre a qualidade dos modelos avaliados.

Palavras-chave: Aprendizado de Máquina; Aprendizado Supervisionado; Medidas de capacidade preditiva; Classificação; caret.

¹Departamento de Estatística, UFPR - paulo.neto@ufpr.br

²Departamento de Estatística, UFPR - ara@ufpr.br



Classificação Textual com Uso de Florestas Aleatórias em Processos Jurídicos

Rafael Luís da Silva Maciel;¹ Lia Hanna Martins Morita;² Elianara Martins de Almeida
³ Elen da Silva Moraes Carvalho⁴

Resumo

Este trabalho investiga a aplicação de técnicas de mineração de texto e aprendizado de máquina, especificamente o classificador Floresta Aleatória, na classificação de documentos judiciais de petições de Recursos Especial e Extraordinário. A digitalização dos documentos judiciais no Brasil, impulsionada por resoluções do Conselho Nacional de Justiça, viabilizou o uso dessas técnicas para facilitar a análise e extração de informações. O estudo utilizou dados de processos remetidos à Vice-Presidência do Tribunal de Justiça do Estado de Mato Grosso em 2022, selecionando os quatro assuntos mais frequentes: Indenização por Dano Moral, Contrato Bancário, Obrigaçāo de Fazer/Não Fazer e Tráfico de Drogas e Condutas Afins. O classificador Floresta Aleatória foi avaliado por meio de métricas como acurácia, sensibilidade, especificidade, precisão e F1-score. Os resultados indicaram um desempenho satisfatório do classificador, com uma acurácia média de 75,9%, sensibilidade de 75,4% para Contrato Bancário, 79 % para Indenização por Dano Moral, 52,4% para Obrigaçāo de Fazer ou Não Fazer e 100% para Tráfico de Drogas. As especificidades foram de 88,8%, 81,1%, 96,7% e 99,4%, respectivamente. As precisões foram de 67,1%, 67,3%, 83,7% e 97,9%, respectivamente. Os F1-score foram de 71,1%, 72,7%, 64,5% e 98,9%, respectivamente. As nuvens de palavras revelaram termos distintos entre os assuntos, contribuindo para a eficácia do modelo. Conclui-se que as técnicas de mineração de texto podem auxiliar na classificação de processos judiciais, contribuindo para a eficiência e celeridade na prestação jurisdicional.

Palavras-chave: Mineração de Texto; Floresta Aleatória; Recurso Especial; Recurso Extraordinário.

¹Tribunal de Justiça do Estado de Mato Grosso, Cuiabá MT - rafael.macie@tjmt.jus.br

²Universidade Federal de Mato Grosso, Cuiabá MT - lia.morita@ufmt.br

³Universidade Federal de Mato Grosso, Cuiabá MT - elianara.almeida@ufmt.br

⁴Instituto Federal de Mato Grosso, Cuiabá MT - elen.carvalho@ifmt.edu.br



Análise Preditiva da Satisfação do Usuário em Telessaúde de um Hospital Universitário

Karen Conceição Dias ¹

Marcello Dala Bernadina Dalla ²

Juliana Dantas de Araujo Santos Camargo ³

Camiliane Azevedo Ferreira ⁴

Scheyla Fraga Ferreira Rauta ⁵

Resumo

A telessaúde tem se consolidado como uma ferramenta essencial para ampliar o acesso aos serviços de saúde. Os Hospitais Universitários (HUs) no Brasil têm papel estratégico. Contudo, a experiência e satisfação do usuário ainda precisam de maior compreensão. Este estudo tem como objetivo analisar a percepção dos pacientes atendidos por teleconsulta no HU do Espírito Santo, Hospital Universitário Antonio Cassiano Antonio Moraes (Hucam), utilizando técnicas de aprendizado de máquina para identificar padrões e prever níveis de satisfação. Trata-se de um estudo transversal, realizado com pacientes atendidos no período de maio/2020 a maio/2025. Foi utilizado um questionário estruturado com variáveis relacionadas às características sociodemográficas, acesso à tecnologia e satisfação com a teleconsulta. Para a análise, que está em andamento, foram aplicadas abordagens supervisionadas de aprendizado de máquina, visando modelar e prever a satisfação dos usuários. A performance dos modelos foi avaliada por meio de Acurácia, Precisão, Revocação e F1-score, além da Área sob a Curva ROC (AUC-ROC) para medir a discriminação entre níveis de satisfação. Os resultados esperados incluem a identificação dos principais determinantes da satisfação e a construção de modelos preditivos que permitam estimar a probabilidade de alta satisfação.

Palavras-chave: Serviço de Telessaúde; Satisfação do Usuário; Machine Learning.

¹ Hucam/Ufes, Vitória/ ES – karen.estatistica@gmail.com.br

² Hucam/Ufes, Vitória/ ES – marcello.dalla@ebserh.gov.br

³ Maternidade Escola Januário Cicco (MEJC)/ UFRN, Natal/ RN – juliana.camargo@ebserh.gov.br

⁴ MEJC/UFRN, Natal/ RN – Camiliane.Ferreira@ebserh.gov.br

⁵ Hucam/Ufes, Vitória/ ES – sheyla.rauta@ebserh.gov.br



Avaliação da qualidade de sementes por imagens de raio X utilizando uma abordagem comparativa entre classificadores de aprendizado de máquina

Maria de Fátima Magalhães de Souza Andrade;¹ Thelma Safadi;²

Resumo

Este trabalho propõe uma abordagem automatizada para a classificação de sementes a partir de imagens de raio X. Foram analisadas 446 imagens, das quais foram extraídos descritores de textura (Haralick), forma, momentos e características morfométricas. As sementes foram agrupadas em três classes: cheias, deformadas e levemente danificadas. Para a tarefa de classificação, foram testados os algoritmos Regressão Logística Multinomial, Árvore de Decisão, Floresta Aleatória, XGBoost, SVM e Redes Neurais. Os modelos foram avaliados utilizando métricas como acurácia (com intervalo de confiança de 95%), sensibilidade, especificidade, valor preditivo positivo, acurácia balanceada e teste qui-quadrado. Os resultados demonstraram que o classificador XGBoost obteve acurácia de 76,4% (IC 95%: 66,2-84,8%); seguido pela Floresta Aleatória (75,3%)(IC 95%: 65,0-83,3%) e Regressão Logística (75,3%)(IC 95%: 65,0-83,8%). Embora a análise por classe tenha revelado que a sensibilidade variou entre os modelos, com melhor desempenho para as classes "cheias" e maior variabilidade para "leves", todos os modelos apresentaram resultados estatisticamente significantes ($p < 0,001$) no teste qui-quadrado, demonstrando sua capacidade de diferenciar entre as classes de sementes. A pesar do desempenho promissor, a ampliação do banco de imagens poderia melhorar a generalização dos modelos. A seleção do melhor classificador possibilita a implementação de sistemas computacionais precisos em laboratórios de análise e cadeias produtivas agrícolas, promovendo agilidade e padronização na avaliação da qualidade de sementes.

Palavras-chave: Aprendizado de máquina; classificadores; imagens de raio X; sementes; agricultura de precisão.

¹Departamento de Estatística, ICET, UFLA/Campus Lavras – Minas Gerais - maria.andrade13@estudante.ufla.br

²Departamento de Estatística, ICET, UFLA/Campus Lavras – Minas Gerais - safadi@ufla.br



Multifractal analysis of orbital products classified as *Sargassum* spp.

Francisco Gustavo¹, Cristele Chevalier², Marine Laval³, Luc Courtrai⁴, Alex Costa de Silva⁵, Jacques Descloitres⁶, Yamina Aimene⁷, Adan Salazar⁸, Audrey Minghelli⁹, Paulo Duarte Neto¹⁰

Abstract

This study applied multifractal analysis to characterize aggregations of *Sargassum* spp. in the Western Mid-Atlantic (WCWA) region, using segmented images obtained by the OLCI sensor (Sentinel-3) and made available by the SargAlert project. The methodology included the analysis of 166 images from different periods between 2020 and 2024, with the extraction of metrics such as dominant singularity (α_0), spectrum asymmetry ($f(\Delta\alpha)$), spectrum width ($\Delta\alpha$) and lacunarity (Λ), using the FracLac plug-in in ImageJ. In addition, the total coverage area and the ratio between the number of particles (QP) and the number of classified pixels (NPX) were calculated. The results showed that the images exhibit multifractal behavior, with decreasing D_q spectra ($D_0 > D_1 > D_2$) and $f(\alpha)$ singularity spectra with parabolic curvatures. A direct correlation was observed between multifractal patterns and oceanographic conditions, with images from 2020 showing less complexity, low coverage and high lacunarity - suggesting environments less favorable to the formation of dense aggregations, possibly due to the lower availability of nutrients. In contrast, the years 2021 and 2023 showed greater area, density and organization, indicating the influence of natural phenomena that intensified mass transport and agglomeration. The Λ metric stood out as an indicator of spatial heterogeneity. Multifractal analysis, therefore, proved to be effective in representing the morphology and dynamics of *Sargassum* spp. masses, contributing innovative metrics that can improve models for monitoring, forecasting and mitigating these events in the Tropical Atlantic.

keywords: Singularity spectrum; Dominant singularity; Tropical Atlantic; Ocean dynamics.

¹Departamento de Estatística e Informática - DEINFO/UFRPE, Recife - Pernambuco, Brasil - fr.gustavosilva.098@gmail.com

²Mediterranean Institute of Oceanography (MIO), IRD, Marseille, France - cristele.chevalier@mio.osupytheas.fr

³Laboratoire des Matériaux et Molécules en Milieu Agressif (L3MA), Schoelcher, France - marine.laval@mio.osupytheas.fr

⁴IRISA, Université de Bretagne Sud, Vannes, France - luc.courtrai@univ-ubs.fr

⁵Departamento de Oceanografia da Universidade Federal de Pernambuco – DOCEAN/UFPE, Recife - Pernambuco, Brasil - alex.csilva@ufpe.br

⁶IRISA, Université de Bretagne Sud, Vannes, France - Jacques.Descloitres@univ-lille.fr

⁷Laboratoire des Matériaux et Molécules en Milieu Agressif (L3MA), Schoelcher, France - yamina.aimene@univ-antilles.fr

⁸Mexican Space Agency (AEM), Ciudad de Mexico, Mexico - asalazar.cdmx.2023@gmail.com

⁹Laboratoire d’Informatique et Système (LIS), Toulon, France - audrey.minghelli@univ-th.fr

¹⁰Departamento de Estatística e Informática - DEINFO/UFRPE, Recife - Pernambuco, Brasil - pjduarteneto@gmail.com



Performance of Regularization Techniques in High-Dimensional Linear Mixed Models

Daniela C. R. Oliveira ¹ Marcos S. Oliveira ² Victor H. Lachos ³

Abstract

In high-dimensional contexts, frequently observed in the field of Genetics, where the number of explanatory variables exceeds the number of observations ($p > n$), variable selection is a significant challenge, particularly when employing linear mixed models that incorporate hierarchical structures or correlated data. This study evaluates, through simulations, the performance of the Lasso and Elastic Net regularization methods in selecting fixed effects in mixed models while keeping random effects unpenalized. Lasso, based on ℓ_1 penalization, is widely recognized for its ability to shrink irrelevant coefficients to zero, effectively performing variable selection. However, in the presence of multicollinearity among predictors, it tends to arbitrarily select only one variable from a group of correlated features, which may compromise model interpretation. Elastic Net, in contrast, combines ℓ_1 and ℓ_2 penalties, allowing for the joint selection of groups of correlated variables while preserving the sparsity-inducing properties of Lasso. The results, assessed using metrics such as specificity, sensitivity, accuracy, and mean squared error, indicate that Elastic Net delivers more robust and effective performance in variable selection for mixed models applied to high-dimensional data. These findings reinforce the applicability of the method in studies aimed at identifying variables associated with phenotypes of interest, contributing to health and well-being solutions, and thereby aligning with the Sustainable Development Goals of the United Nations agenda.

Keywords: Elastic Net; Genetic Analysis; High-Dimensional Data; Lasso; Linear Mixed Models.

¹Department of Mathematics and Statistics, Federal University of São João del-Rei, MG, Brazil - daniela@ufs.edu.br

²Department of Mathematics and Statistics, Federal University of São João del-Rei, MG, Brazil - mso@ufs.edu.br

³Department of Statistics, University of Connecticut, Storrs, CT, USA - hlachos@uconn.edu



Optimizing Robot Soccer Image Classification Using Transfer Learning and Data Augmentation for Deep Learning Experiments

Marcos S. Oliveira ¹ Franklin A. Brito; ² André L. C. Ottoni ³

Abstract

Robot soccer requires autonomous agents to navigate and interact efficiently with their environment. Computer vision plays a crucial role by extracting visual information to enhance artificial intelligence systems. This study evaluates the impact of Transfer Learning (TL) and Data Augmentation (DA) on the classification of robot soccer images using the Edge Impulse platform for Deep Learning (DL) experiments. The dataset consists of humanoid robot soccer images categorized into five classes: Ball, Ball on the line, Robot, Robot with ball, and Goal in the image. Experiments were conducted under three conditions: without TL and DA, using only TL, and applying both techniques. The chosen DL model architecture was MobileNetV2 96x96 0.35. Results showed that TL significantly improved model accuracy, reaching 100% in some cases. The addition of DA further enhanced classification performance in most experiments, except in the five-class classification, where a slight accuracy decrease (-0.6%) was observed. These findings confirm TL and DA as effective techniques for optimizing robot soccer image classification. Future work may explore hyperparameter tuning to further refine model performance.

Keywords: Computer Vision; Data Augmentation; Deep Learning; Robot Soccer; Transfer Learning.

¹Department of Mathematics and Statistics, Federal University of São João del-Rei, MG, Brazil - mso@ufsj.edu.br

²Technologic and Exact Center, Federal University of Recôncavo da Bahia, BA, Brazil - frank.engcet@gmail.com

³Department of Computing, Federal University of Ouro Preto, MG, Brazil - andre.ottoni@ufop.edu.br



Arquitetura da Luta: O Algoritmo por trás da Era do Vale-tudo (1994-1997)

Ícaro da Costa Francisco ¹ Stéfane Lele Rossoni ² Diogo Francisco Rossoni ³

Resumo

Este estudo analisa os métodos de vitória na era Vale-Tudo do UFC, entre 1993 e 1997, com base em dados históricos e modelagem preditiva. Durante esse período, o evento era marcado por regras mínimas e grande diversidade de estilos marciais, como *Wrestling*, *Kickboxing* e *Brazilian Jiu-Jitsu (BJJ)*. A partir da edição 14, mudanças nas regras, como a obrigatoriedade do uso de luvas e restrições a certos golpes, marcaram a transição para o formato moderno do MMA. Utilizando o algoritmo Random Forest, foram classificadas as formas de vitória (*submission*, *KO/TKO* e *decision*) com 91,3% de acurácia. As variáveis mais relevantes para predição foram tentativas de finalização (*R_SUB_ATT*), duração da luta e eficiência em golpes significativos. Simulações indicaram que lutas curtas favorecem vitórias por submissão, enquanto lutas mais longas tenderam a terminar em nocaute técnico. A análise descritiva também apontou a predominância de estilos como *Wrestling* e *Mixed* em vitórias, reforçando a relevância do domínio do jogo de chão e da versatilidade técnica. Os resultados demonstram que, mesmo em um ambiente de poucas regras como o Vale-Tudo, há padrões mensuráveis que podem ser utilizados para prever desfechos com base em dados objetivos. Este trabalho destaca o potencial de modelos estatísticos na compreensão estratégica do MMA e abre caminhos para estudos comparativos com edições mais recentes do UFC.

Palavras-chave: MMA; Vale-tudo; Random Forest; UFC; modelagem preditiva.

¹ Programa de Pós Graduação em Bioestatística - UEM, Maringá/PR – icarodacostafran@hotmail.com

² Programa de Pós-Graduação em Ciências da Saúde - UEM, Maringá/PR – ste.tistics@gmail.com

³ Programa de Pós-Graduação em Bioestatística - UEM, Maringá/PR – dfrossoni@uem.br



Utilização de Algoritmos de Machine Learning baseado em estrutura de Árvores aplicados a dados SAR

Raffael Marinho de Arruda Feitosa ¹

Jodavid de Araújo Ferreira ²

Resumo

Radar de Abertura Sintética (SAR) é uma técnica ativa de sensoriamento remoto que funciona pela emissão de pulsos de radar em direção à superfície terrestre que, ao retornarem após interagir com alvos no solo, possibilitam a geração de imagens de alta resolução. Uma das vantagens está na capacidade de operar independentemente das condições de iluminação solar e meteorológicas, como a presença de nuvens ou chuva. Devido às características, os dados SAR têm sido amplamente utilizados em diversas aplicações, como o monitoramento da profundidade do lençol freático, a detecção de derramamento de óleo nos oceanos e a classificação de diferentes regiões e coberturas do solo. A presença do ruído speckle representa um dos desafios na sua interpretação e análise pois, comprometem a aplicação eficaz de técnicas de processamento de imagens, especialmente baseadas na suposição de distribuições normais dos dados. Neste trabalho, busca-se avaliar o desempenho dos algoritmos de aprendizado de máquina baseados em árvores: Random Forest (RF) e Gradient Boosting (GB), aplicados à classificação de imagens SAR, que serão utilizados em uma imagem SAR da região da Baía de São Francisco, nos Estados Unidos, com três classes bem definidas: oceano, floresta e área urbana. A eficácia será avaliada por métricas de desempenho.

Palavras-chave: sensoriamento remoto; dados SAR; ruído speckle; Random Forest; Gradient Boosting.

¹ Departamento de Matemática UFPE, Pernambuco – raffael.arruda@ufpe.br

² Departamento de Estatística UFPE, Pernambuco – jodavid.ferreira@ufpe.br



Comparação dos Modelos de Regressão Logística, Boosting e Floresta Aleatória no Problema de Credit Scoring utilizando Dados Simulados

Eduarda Godoy Afonso¹

Isabela Borges Farina²

Pedro Henrique Duarte³

Márcio Luis Lanfredi Viola⁴

Resumo

No contexto da avaliação do risco de crédito, as instituições financeiras utilizam modelos de classificação para o cálculo do escore de crédito dos proponentes a fim de auxiliar na tomada de decisão, ou seja, se o proponente terá crédito aprovado. Tradicionalmente, o modelo de regressão logística foi utilizado para a obtenção do escore de crédito, mas, atualmente, métodos de aprendizado de máquina estão sendo testados e utilizados, como, por exemplo, métodos baseados em Boosting e Floresta Aleatória. As instituições financeiras utilizam a estatística do teste de Kolmogorov-Smirnov (KS) como uma métrica tanto para a avaliação do desempenho de modelos preditivos como para avaliar se o modelo ajustado continua sendo eficiente para a discriminação de inadimplentes em um público cujo perfil não é o mesmo em relação ao público no qual o modelo foi ajustado. Este trabalho tem por objetivo comparar o poder preditivo dos modelos de regressão logística, Boosting e Floresta Aleatória, assim como avaliar como estes métodos se comportam quando ocorrem mudanças de perfis dos proponentes. Para isto, dados foram simulados considerando diferentes tamanhos amostrais (500, 10000 e 100000) combinados com diferentes proporções de indivíduos inadimplentes (60%, 70% e 85%).

Palavras-chave: Métrica KS; Modelos preditivos; Mudança de perfil; Risco de crédito.

¹Departamento de Estatística, Universidade Federal de São Carlos – eduardaafonso@estudante.ufscar.br

²Departamento de Estatística, Universidade Federal de São Carlos – isabelafarina@estudante.ufscar.br

³ Departamento de Estatística, Universidade Federal de São Carlos – pedrohd@estudante.ufscar.br

⁴ Departamento de Estatística, Universidade Federal de São Carlos – lanfredi@ufscar.br



Previsão de preços de insumos utilizados na cadeia do leite com modelo de aprendizado de máquina

Sérgio Domingos Simão¹

João Paulo Assis Bonifácio²

Izabela Regina Cardoso de Oliveira³

Resumo

O milho é tradicionalmente o principal insumo energético na Nutrição de Ruminantes. No entanto, a substituição por subprodutos como o caroço de algodão e a polpa cítrica pode oferecer vantagens nutricionais e econômicas, especialmente quando se considera a relação de troca entre esses insumos. Por essa razão, o estudo da dinâmica de preços desses produtos é de grande importância para agentes do setor da cadeia produtiva do leite. O objetivo deste trabalho foi utilizar um modelo de aprendizado de máquina, especificamente Máquinas de Vetores de Suporte (*Support Vector Machine – SVM*) com kernel linear, para - predição dos preços diários por tonelada de polpa cítrica e de caroço de algodão, e obter previsões com horizontes de 7, 15 e 30 dias. Foram utilizados dados fornecidos pela empresa parceira JPA Agro, uma *trader* localizada em Lavras, MG, abrangendo o período de 2018 a 2025. O ajuste dos modelos aos dados de polpa cítrica foi realizado para nove mesorregiões nos estados da Bahia, Minas Gerais e São Paulo, enquanto para o caroço de algodão, os ajustes foram feitos para os estados de Minas Gerais e Mato Grosso. Foi utilizada a linguagem R, e os pacotes *tidyverse* e *tidymodels*. As covariáveis consideradas foram a dimensão temporal e o componente autorregressivo de primeira ordem (AR-1), visando capturar a dependência temporal. Os dados foram divididos em conjuntos de treino (80%) e teste (20%), respeitando a sequência temporal. Além da previsão de preços, o estudo também avaliou a relação de troca com o milho, oferecendo *insights* sobre o poder de compra do produtor. A aplicação do SVM permitiu gerar estimativas confiáveis para o curto prazo, oferecendo uma ferramenta eficaz para o planejamento da compra de insumos na cadeia leiteira.

Palavras-chave: Polpa cítrica; Caroço de algodão; Relação de troca; Machine Learning; *Support Vector Machine*.

Agradecemos à FAPEMIG pelo apoio financeiro e à empresa JPA Agro pela parceria.

¹ Pós-doutorando do Programa de Estatística e Experimentação Agropecuária, DES/ICET, Universidade Federal de Lavras e Professor na Escola Estadual João Melo Gomide – Perdões/MG – sergiodomilson@ufla.br

² Graduando em Estatística, Universidade Federal de Lavras – Lavras, MG – Brasil – joao.bonifacio2@estudante.ufla.br

³ Departamento de Estatística – DES/ICET, Universidade Federal de Lavras – Lavras, MG – Brasil – izabela.oliveira@ufla.br



Comparação de Modelos de Classificação para Detecção de Fraudes Financeiras

Heitor Rodrigues Sabino ¹

Bryan de Lima Naneti Barbosa ²

Renato Ribeiro de Lima ³

Resumo

O avanço das tecnologias financeiras e das transações digitais tem ampliado também as formas de atuação de agentes fraudadores, exigindo sistemas mais eficientes para identificar fraudes. Este trabalho compara o desempenho de cinco modelos de machine learning na detecção de fraudes financeiras, em dados balanceados e desbalanceados. Foi utilizada a base pública do Kaggle sobre fraudes com cartões de crédito, contendo cerca de 284 mil transações realizadas por clientes europeus, das quais apenas 0,17% são fraudes. Após o pré-processamento, aplicou-se a Análise de Componentes Principais (PCA) para redução de dimensionalidade. Os modelos avaliados foram: Random Forest, XGBoost, Redes Neurais, Regressão Logística e SVM. As métricas utilizadas incluíram acurácia, precisão, recall, f1-score, falsos positivos/negativos e AUC da curva ROC. Os testes foram conduzidos tanto com os dados originais quanto com versões balanceadas por reamostragem. Também foi aplicada validação cruzada para avaliação da robustez. Em dados平衡ados, Random Forest e XGBoost apresentaram acurácia superior a 99,9% e nenhum falso negativo. As Redes Neurais tiveram desempenho intermediário, e SVM e Regressão Logística apresentaram mais erros. Em dados desbalanceados, os modelos de árvore mantiveram bons resultados, enquanto os demais tiveram maior dificuldade. A aplicação do PCA contribuiu para reduzir a complexidade computacional sem comprometer o desempenho. Além disso, para uma melhor avaliação dos modelos, estão sendo criados cenários que imitam fraudes reais, como card testing, transações de alto valor durante a madrugada e uso sequencial de cartões em localidades distintas. Dessa forma, obtendo uma melhor avaliação dos modelos.

Palavras-chave: Fraudes Financeiras; Machine Learning; Classificação; PCA.

¹ Departamento de Ciência da Computação, UFLA – heitor.sabino@estudante.ufla.br

² Departamento de Ciência da Computação, UFLA – bryan.barbosa@estudante.ufla.br

³ Departamento de Estatística, UFLA – rrliima@ufla.br



TRI multidimensional como decoder explícito em VAEs: retropropagação direta via diferenciação automática

Flavio Margarito Martins de Barros ¹

Mariana Cúri ²

Resumo

A estimação de parâmetros em modelos multidimensionais da Teoria de Resposta ao Item (MIRT) enfrenta desafios computacionais, especialmente em modelos complexos ou grandes volumes de dados. Variational Autoencoders (VAEs) oferecem uma alternativa escalável para inferência variacional de habilidades latentes e parâmetros dos itens. Diferentemente de implementações que utilizam decoders genéricos baseados em redes neurais densas, propomos um método inovador que implementa o modelo MIRT, incluindo o logístico de três parâmetros (3PL), como um decoder explícito em um VAE. Usando diferenciação automática (AD) no framework Flux.jl (Julia), calculamos gradientes diretamente da função MIRT, garantindo retropropagação eficiente sem redes neurais intermediárias. Estendendo abordagens anteriores (Curi, 2020), nosso método incorpora o parâmetro de pseudo-chute, aumentando a flexibilidade. Simulações evidenciam maior precisão na estimação de parâmetros em cenários de alta dimensionalidade. Essa abordagem une a interpretabilidade psicométrica à escalabilidade do aprendizado profundo, com aplicações em avaliações educacionais, superando limitações de métodos clássicos.

Palavras-chave: psicometria; VAE; inferência variacional; MIRT; diferenciação automática.

¹ Programa Interinstitucional de Pós-Graduação em Estatística (PIPGES) - UFSCar/USP, São Carlos – flavio.barros@usp.br

² ICMC – Universidade de São Paulo (USP), São Carlos – mcuri@icmc.usp.br



Avaliação de Modelos de Machine Learning para Previsão da Velocidade do Vento na Cidade da Beira - Moçambique

Viviane Costa Silva;¹ Orlando Eduardo Chipura;² Elias Manensa Sabe;¹ Momate Emate Ossifo;¹ Daví Barbosa Pereira de Sousa;¹ Geraldo Magela da Cruz Pereira;³ Marcelo de Carvalho Alves⁴

Resumo

A velocidade do vento é um fator crucial em diversas áreas, como a meteorologia, a agricultura, a geração de energia e a previsão de desastres naturais. Estudar a velocidade do vento é fundamental para entender padrões climáticos, otimizar a produção de energia e proteger comunidades vulneráveis. A cidade da Beira, localizada na costa central de Moçambique, é uma das principais cidades portuárias do país. Devido à sua localização geográfica, a cidade é frequentemente afetada por fenômenos climáticos como ciclones tropicais e tempestades. A Beira está situada numa região onde os ventos podem ser intensos, o que torna importante monitorar e prever a velocidade do vento para mitigar os impactos de eventos climáticos adversos. Este trabalho visa comparar três modelos de previsão — Prophet, ARIMA e NNAR — aplicados aos dados de velocidade do vento na Beira. A avaliação dos modelos foi realizada por meio de métricas como erro absoluto médio (MAE), erro percentual absoluto médio (MAPE), erro absoluto médio escalado (MASE), erro percentual absoluto simétrico (SMAPE), erro quadrático médio (RMSE) e o coeficiente de determinação (R^2). Os resultados indicaram que o modelo Prophet apresentou o melhor desempenho entre os três, com MAE de 0,354, MAPE de 7,13%, MASE de 0,667, SMAPE de 7,25%, RMSE de 0,471 e R^2 de 0,422. O modelo NNAR(1,1,20)[7] apresentou desempenho intermediário, com MAE de 0,465, MAPE de 9,90%, MASE de 0,874, SMAPE de 9,40%, RMSE de 0,586 e R^2 de 0,0565. Já o modelo ARIMA(2,0,2)(1,0,0)[7] obteve os piores resultados, com MAE de 0,502, MAPE de 10,3%, MASE de 0,944, SMAPE de 10,2%, RMSE de 0,620 e R^2 de apenas 0,00986. Além da análise do desempenho dos modelos, o Prophet também foi utilizado para prever a velocidade do vento nos dois meses seguintes ao período avaliado. Para 30 de dezembro de 2024, a previsão foi de 4,83 m/s, com intervalo entre 3,90 e 5,76 m/s, enquanto o valor real observado em 31 de dezembro foi de 5,16 m/s. Já para 30 de janeiro de 2025, a previsão foi de 4,92 m/s, com intervalo entre 3,99 e 5,84 m/s, sendo que o valor real em 31 de janeiro foi de 5,31 m/s. Esses resultados mostram que o modelo foi capaz de gerar previsões próximas aos valores reais, reforçando sua confiabilidade para aplicações práticas.

Palavras-chave: Meteorologia; Previsão ; Séries Temporais.

¹Programa de Pós-Graduação em Estatística e Experimentação Agropecuária. Universidade Federal de Lavras, Lavras - MG - vivicosta19.vc@gmail.com; sabehelio@gmail.com; momate.ossifo@estudante.ufla.br; davi.barbosa.5454@gmail.com

²Programa de Pós-Graduação em Engenharia Agrícola - Universidade Federal de Lavras, Lavras - MG - o.chipura@gmail.com

³Departamento de Estatística - Universidade Federal de Lavras, Lavras - MG - geraldo.pereira@ufla.br

⁴Departamento de Engenharia Agrícola - Universidade Federal de Lavras, Lavras - MG - marcelo.alves@ufla.br



Bayesian Active Learning Powered by Local Entropy and MCMC-Based Predictive Uncertainty

Rafael Aguilar Magalhães ¹ Paulo Justiniano ² Anderson Ara ³

Resumo

Active Learning (AL) is a class of machine learning techniques designed to reduce the labeling effort by selecting the most informative samples for training predictive models. Traditional AL methods, such as GSx and GSy, rely on evaluating uncertainty or variance through known outputs, requiring access to the true response variable y for all candidate observations. This dependency limits applicability in scenarios where label acquisition is costly or time-consuming. In this study, we propose a novel active learning approach that combines t-SNE for nonlinear dimensionality reduction, local entropy estimation to assess structural complexity, and Bayesian Linear Regression with MCMC to quantify predictive uncertainty. A key innovation of our method is the ability to estimate the variance of predictions before querying the true label, using samples of the posterior distribution of model coefficients obtained via MCMC. The selection of new samples is guided by a composite criterion, which integrates entropy-based structural information and predictive variance. Observations with the highest d_{xy} values are iteratively added to the active set. Experimental results show that the proposed strategy achieves predictive performance comparable to models trained on the full dataset, while requiring less than half the labeled data. This methodology provides a scalable and label-efficient framework for supervised learning tasks, particularly relevant in domains with limited annotated data and high modeling costs.

Palavras-chave: Active Learning, Regressão Linear Bayesiana, Entropy

¹Federal University of Paraná, Curitiba, PR, Brazil - rafael.aguilar@ufpr.br

²Federal University of Paraná, Curitiba, PR, Brazil - paulojus@ufpr.br

³Federal University of Paraná, Curitiba, PR, Brazil - ara@ufpr.br



Comparação de Métodos em Aprendizado Semi-Supervisionado para a Classificação de Notícias Falsas

Carlos Henrique de Meo ¹

Márcio Luis Lanfredi Viola ²

Resumo

A propagação de notícias falsas representa um desafio crescente, exigindo métodos eficazes para sua detecção automatizada. Este trabalho investiga a pseudo-rotulagem como estratégia para mitigar a escassez de rótulos na classificação de *fake news*. Foram analisados três modelos: Regressão Logística, *XGBoost* e *ChatGPT-4*, avaliando seu desempenho sob diferentes níveis de remoção de rótulos, com e sem pseudo-rotulagem. O conjunto de dados utilizado foi o *Fake.br-Corpus* e os modelos foram avaliados por acurácia, precisão, recall e medida F1. Os resultados indicaram que a pseudo-rotulagem impactou negativamente a eficácia dos modelos, pois, com a remoção progressiva dos rótulos, todas as métricas de desempenho se deterioraram. Além disso, dentre os modelos estudados, o *ChatGPT-4* obteve boas métricas iniciais, mas apresentou alta variabilidade nos resultados, falhando sob critérios de reproduzibilidade. Por outro lado, os modelos de Regressão Logística e *XGBoost* mostraram maior estabilidade e adequação ao processamento de texto, mantendo desempenho mais consistente e menos custoso diante da remoção de rótulos.

Palavras-chave: Aprendizado semi-supervisionado; *ChatGPT-4*; *Fake news*; Regressão logística; *XGBoost*.

¹ Departamento de Estatística, Universidade Federal de São Carlos – carlosmeo@estudante.ufscar.br

² Departamento de Estatística, Universidade Federal de São Carlos – lanfredi@ufscar.br



Avaliação de Respostas Dissertativas em Testes Adaptativos Informatizados baseada em Large Language Models

João Vítor Monteiro de Souza ¹

Fernando de Jesus Moreira Júnior ²

Resumo

Os Testes Adaptativos Informatizados (TAI) personalizam avaliações ao ajustar dinamicamente a dificuldade das questões, porém tradicionalmente limitam-se a itens objetivos (ex.: múltipla escolha) devido à complexidade de corrigir respostas dissertativas. Este estudo propõe um sistema automatizado para superar essa limitação, utilizando Large Language Models (LLMs) na correção de respostas discursivas. Por meio de APIs (ex.: OpenAI, DeepSeek), o sistema realiza análise semântica e sintática das respostas, comparando-as a um banco de referência para atribuir notas (0-1) com base em critérios como coerência, clareza e relevância. Respostas com nota ≥ 0.7 (critério ajustável) são classificadas como adequadas, permitindo integração imediata ao TAI para otimizar o número de questões e calibrar a dificuldade em tempo real. A abordagem reduz drasticamente tempo e custos de correção manual, além de oferecer feedbacks instantâneos e detalhados. Testes preliminares com 50 participantes revelaram 85% de concordância entre a avaliação automática e humana, validando a precisão do modelo. A integração de LLMs aos TAIs representa um avanço metodológico ao expandir sua aplicação para avaliações qualitativas, mantendo a adaptabilidade e eficiência características desses testes. Os resultados sugerem que a automação via LLMs não só replica critérios humanos de correção, mas também introduz escalabilidade e rigor analítico, potencializando a personalização do ensino. Este trabalho contribui para a evolução dos TAIs, propondo uma solução inovadora e economicamente viável para avaliações educacionais complexas.

Palavras-chave: Testes Adaptativos Informatizados; Inteligência Artificial; Large Language Models (LLMs); Tecnologia na Educação; Processamento de Linguagem Natural (PLN).

¹ Curso de Estatística, Universidade Federal de Santa Maria – vitor.souza@acad.ufsm.br

² Departamento de Estatística, Universidade Federal de Santa Maria – fernando.junior@ufsm.br



Por que ETFs podem ser mais eficientes que carteiras de ações individuais? Uma análise com aprendizado de máquina

Pietros André Balbino dos Santos¹ Paulo Henrique Sales Guimaraes²

Resumo

A integração entre gestão financeira e crescimento econômico sustentável passa pela otimização de investimentos. Ao comparar carteiras de ETFs globalmente diversificados com carteiras de ações brasileiras, este trabalho demonstra como instrumentos financeiros simples podem democratizar o acesso a retornos robustos e fomentar economias resilientes. O estudo comparou o desempenho de carteiras de ações brasileiras otimizadas por modelos de aprendizado de máquina (ML) com uma carteira balanceada de ETFs. A carteira de ETFs incluiu SPY (S&P 500), VT (mercado global), AGG (renda fixa) e BOVA11 (Ibovespa), com pesos iguais. Foram construídas 10.000 carteiras aleatórias compostas por ações do IBOVESPA para treinamento dos modelos de ML. As carteiras aleatórias tiveram retorno médio diário de 0,01656%, volatilidade de 0,03751% e Sharpe de 0,4789, indicando baixo desempenho. Em contraste, a carteira de ETFs apresentou retorno quase três vezes maior (0,0482% ao dia), menor volatilidade (0,01848%) e Sharpe superior (2,6082), demonstrando melhor eficiência risco-retorno. Os modelos de ML foram treinados para prever retornos e otimizar as carteiras. O XGBoost superou o Random Forest, com RMSE de 0,0003085 e R² de 0,2881, ainda que o baixo R² sugira limitações na explicação dos retornos, possivelmente devido à ausência de variáveis adicionais e curto período analisado (dois anos). A carteira de ETFs ofereceu desempenho superior em comparação com carteiras aleatórias de ações. Embora XGBoost tenha mostrado melhores resultados, sua capacidade limitada indica a necessidade de incorporar mais fatores ou testar outras abordagens, que está sendo feito. A simplicidade e eficiência dos ETFs reforçam sua vantagem para pequenos investidores.

Palavras-chave: Exchange Traded Fund; Otimização de carteira; Mercados emergentes; risco-retorno.

¹ Departamento de estatística, Universidade Federal de Lavras - UFLA – pietrosbalbino@ufla.br

² Departamento de estatística, Universidade Federal de Lavras - UFLA – paulo.guimaraes@ufla.br



Qualidade Estatística de Dados Sintéticos Gerados por LLMs: Estudo com Base Clínica de Reabilitação Pós-AVC

Aline Pires de Angeli Ferreira;¹ Lucas Oliveira Alves² Diego Roberto Colombo Dias²

Resumo

O uso de dados sintéticos tem ganhado destaque para ampliar bases de dados reais, reduzir custos com coleta e preservar a privacidade de indivíduos, especialmente em áreas sensíveis como a saúde. Com o avanço dos Modelos de Linguagem de Larga Escala (LLMs), surge a oportunidade de explorar seu potencial na geração de dados sintéticos que são significativos estatisticamente. Este trabalho investiga o uso de LLMs para gerar dados sintéticos a partir de uma base real composta por informações clínicas, demográficas e funcionais de pacientes em reabilitação pós-acidente vascular cerebral (AVC). Inicialmente, é realizada uma análise descritiva da base original, buscando compreender suas principais características e distribuições. Em seguida, os dados reais são utilizados para condicionar um LLM na tarefa de geração sintética. A qualidade dos dados gerados é avaliada por meio de métricas estatísticas, como comparação de distribuições marginais, análise de correlação e componentes principais (PCA). Além disso, modelos estatísticos clássicos, como regressão logística e árvores de decisão, são aplicados sobre os dados sintéticos e reais para comparar a capacidade preditiva em relação a variáveis de interesse. O objetivo principal é avaliar a fidelidade estatística dos dados gerados pelo LLM e sua viabilidade como insumo para análises estatísticas. A proposta busca integrar técnicas modernas de inteligência artificial com a estatística aplicada, contribuindo para o desenvolvimento de métodos seguros e eficientes de geração de dados sintéticos em contextos clínicos.

Palavras-chave: LLM; Dados Sintéticos; Análise de Dados Clínicos; Estatística Aplicada; Modelagem Estatística.

¹Departamento de Estatística, UFES - apiresangeli@gmail.com

²Programa de Pós-graduação em Informática, UFES - lucas.o.alves@edu.ufes.br

²Departamento de Estatística/Programa de Pós-graduação em Informática, UFES - diego.dias@ufes.br



O Supremo Tribunal Federal (STF) e os Objetivos de Desenvolvimento Sustentável (ODS)

Carlos Alberto Oliveira de Matos (Prof. Aposentado Voluntário)¹

Resumo

A atuação do Supremo Tribunal Federal (STF) no contexto dos Objetivos de Desenvolvimento Sustentável (ODS) da Agenda 2030 representa uma evolução na integração da sustentabilidade ao sistema jurídico brasileiro, evidenciando a responsabilidade intergeracional prevista na Constituição de 1988. Objetivamos analisar a correlação entre resumos de decisões de 102 julgamentos do STF e os indicadores dos ODS, utilizando métodos de análise de redes complexas e processamento de linguagem natural para classificar processos segundo a Agenda 2030 (RAFA 2030). A metodologia empregou a construção de grafos a partir de dados de discordâncias entre ministros do STF, utilizando matrizes de adjacência, técnicas de otimização como spin-glass e reconhecimento simulado para identificar comunidades num grafo, além de medidas ponderadas de centralidade de rede: Grau, rank de Centralidade e de Autoridade e Centralidade de Intermediação e Proximidade com o uso de ferramentas em R. Os resultados indicam que ministros com alta centralidade nas redes de discordância são figuras-chave na formação das decisões relacionadas aos ODS. A classificação automatizada dos processos evidenciou forte aderência aos seguintes temas: Paz, Justiça e Instituições Eficazes (ODS 16), Redução das Desigualdades (ODS 10), Trabalho Decente (ODS 8), Parcerias e Meios de Implementação (ODS 17). As médias ponderadas de discordâncias em relação a eles foram: 6,83; 8,11; 6,18; 12,06 e 6,29 respectivamente. Conclui-se que a adoção de tecnologias de IA, como a RAFA 2030, potencializa a implementação dos ODS no judiciário. A correlação entre indicadores e decisões judiciais sinaliza uma governança mais alinhada aos princípios de sustentabilidade e direitos humanos.

Palavras-chave: Supremo Tribunal Federal; Objetivos de Desenvolvimento Sustentável; Inteligência Artificial; Análise de Redes Complexas.

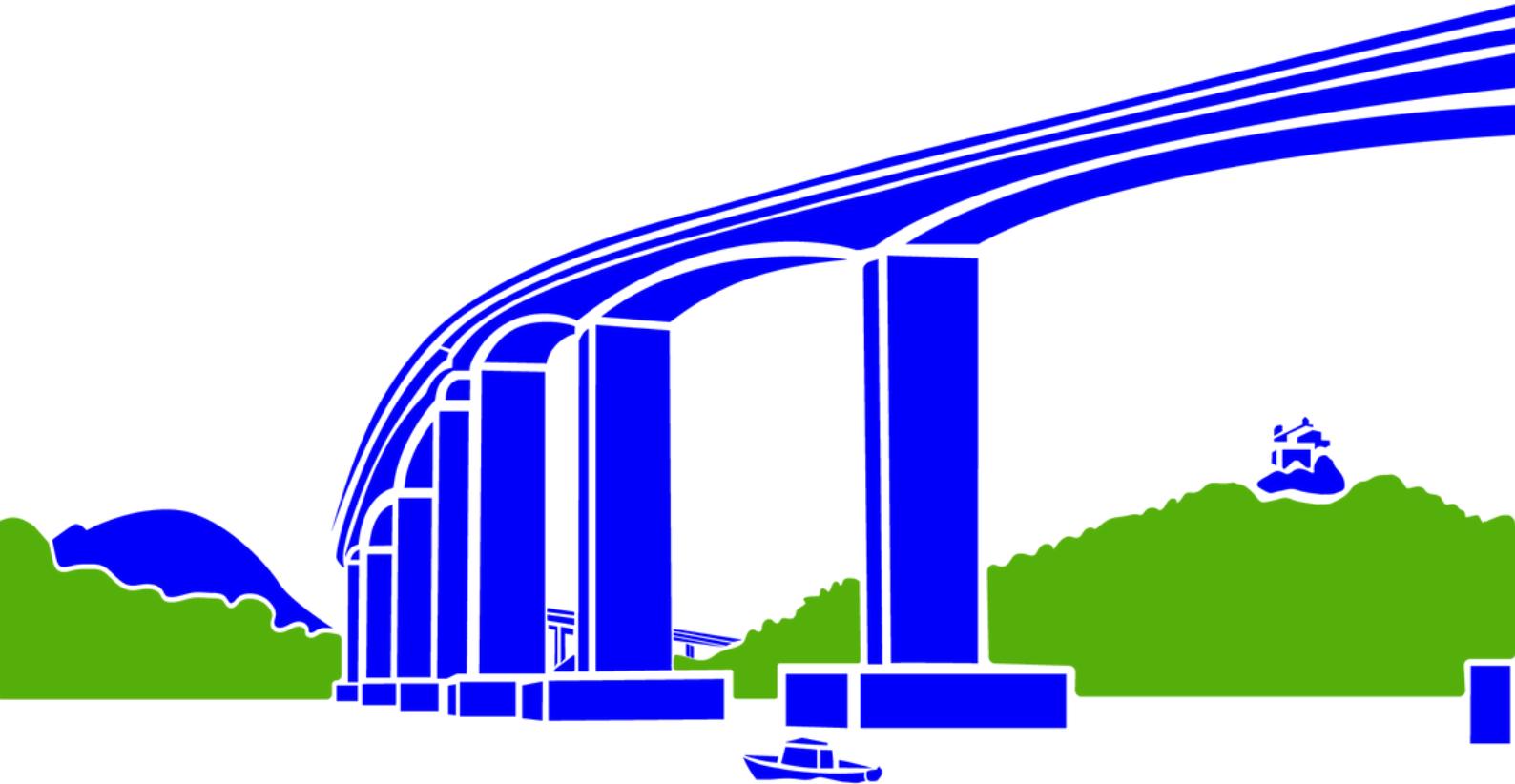
¹Instituto de Ciências e Engenharia - UNESP, Itapeva/SP - carlos.matos@unesp.br

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Pôsters

**Análise de
Sobrevivência**





Machine learning em Análise de Sobrevivência - aplicações em dados oncológicos

Mariana Freitas;¹ Profa. Dra. Agatha Rodrigues²

Resumo

A análise de sobrevivência tem como objetivo modelar o tempo até a ocorrência de um evento de interesse, sendo amplamente utilizada na área da saúde para identificar fatores de risco e prever desfechos clínicos. Com o aumento da demanda por modelos preditivos robustos, que ofereçam boa acurácia e auxiliem médicos e pesquisadores na tomada de decisões e no planejamento de políticas públicas, torna-se relevante explorar abordagens alternativas aos métodos tradicionais. Este trabalho investiga o uso da Máquina de Vetores de Suporte (SVM) na análise de sobrevivência, comparando seu desempenho com o Modelo de Regressão de Cox e a Floresta Aleatória de Sobrevivência. A avaliação é realizada por meio de estudos de simulação e de uma aplicação em dados reais de pacientes diagnosticados com câncer colorretal no estado de São Paulo, entre 2000 e 2021, obtidos junto à Fundação Oncocentro de São Paulo (FOSP). O conjunto de dados inclui informações demográficas, clínicas e o desfecho de óbito pela doença. O objetivo principal é avaliar a capacidade preditiva da SVM em diferentes cenários e contextos de censura. A comparação entre os métodos permite identificar suas principais vantagens e limitações, fornecendo subsídios importantes para a escolha de modelos em aplicações reais de sobrevivência.

Palavras-chave: Análise de Sobrevivência; Aprendizado de Máquina; Óbito por Câncer; Oncologia; Máquina de Vetores de Suporte.

¹Departamento de Estatística da Universidade Federal do Espírito Santo, Vitória (ES) - mariana.c.freitas@edu.ufes.br

²Departamento de Estatística da Universidade Federal do Espírito Santo, Vitória (ES) - agatha.rodrigues@ufes.br



Modelagem Semiparamétrica de Dados de Confiabilidade Usando Aproximações Polinomiais de Funções de Risco

Ricardo Puziol de Oliveira^{1,2}, Jorge Alberto Achcar²,
João Debastiani Neto^{2,3}, Maria Graziela da Silva Fernandes⁴

Resumo

Este trabalho apresenta um novo modelo semi-paramétrico baseado em polinômios para capturar efetivamente padrões complexos de taxa de risco, como forma de banheira, forma crescente, entre outras variações, por meio de aproximações polinomiais da função de risco. Os parâmetros foram estimados utilizando o método dos mínimos quadrados, com a regularização LASSO empregada para selecionar os termos polinomiais no ajuste do modelo, evitando assim o sobreajuste. O modelo proposto foi aplicado à análise de conjuntos de dados reais de engenharia, utilizando o estimador de Kaplan-Meier para estimar as funções de risco cumulativas empíricas de forma não-paramétrica. Nossos achados destacam a adaptabilidade e eficácia da metodologia apresentada, capturando as três fases tipicamente observadas na análise de confiabilidade de produtos: falha inicial, vida útil e desgaste eventual; além de modelar com precisão dinâmicas complexas da taxa de risco.

Palavras-chave: função de risco, modelos polinomiais, modelagem semi-paramétrica, estimação por mínimos quadrados, estimação de Kaplan-Meier, regressão LASSO.

¹ Departamento de Estatística – Faculdade de Ciência e Tecnologia – Universidade Estadual Paulista – Presidente Prudente/SP – rp.oliveira@unesp.br

² Departamento de Medicina Social – Faculdade de Medicina de Ribeirão Preto – Universidade de São Paulo – Ribeirão Preto/SP – achcar@fmrp.usp.br

³ Departamento de Ciências – Universidade Estadual de Maringá – Goioerê/PR – jdneto@uem.br

⁴ Pós-Graduação em Matemática Aplicada e Computacional - Universidade Estadual de Londrina – Londrina/PR - grazifernandes25@gmail.com



Extended Semiparametric Cox Model: A Cure Fraction Approach for Long-Term Survival

João Debastiani Neto^{1,2}, Jorge Alberto Achcar², Ricardo Puziol de Oliveira^{2,3}, Maria Graziela da Silva Fernandes⁴

Abstract

In this work, we introduce a semi-parametric mixture cure model based on transformation models to incorporate long-term survival outcomes, providing the estimation of the plateau on the survival curves. Also, our approach is capable to distinguishing between subjects who remain at risk and those who achieve long-term remission or cure addressing the limitation of the Cox proportional hazard model, which assumes that all individuals are perpetually at risk. Specifically, in our framework, the unknown hazard function is treated as a latent variable and estimated using a Bayesian approach, where prior information from the Cox model estimates. To illustrate the model usefulness, we employed it into a retrospective cohort of 295 primary invasive breast carcinoma cases from the Netherlands Cancer Institute (NKI). The results showed that the proposed semi-parametric mixture cure model provides a more accurate representation of the survival probabilities by appropriately capturing the heterogeneity between cured and at-risk subjects.

Palavras-chave: Bayesian analysis; Breast cancer; Cox PH model; Cure rate estimation; Semiparametric transformation models.

¹Departamento de Medicina Social - Faculdade de Medicina de Ribeirão Preto - Universidade de São Paulo - Ribeirão Preto/SP - email: achcar@fmrp.usp.br.

²Departamento de Ciências - Universidade Estadual de Maringá - Goioerê/PR - e-mail: jdneto@uem.br.

³Departamento de Estatística - Faculdade de Ciência e Tecnologia - Universidade Estadual Paulista - Presidente Prudente/SP - e-mail: rp.oliveira@unesp.br.

⁴Programa de Pós Graduação em Matemática Aplicada e Computacional, Universidade Estadual de Londrina - Londrina/PR - e-mail: grazifernandes25@gmail.com.



Bayesian Regression Models for Survival Data with Cure Fraction Using Marshall-Olkin Defective Distributions

Dionisio Alves da Silva Neto ^{1,2}; Vera Lucia Tomazella ²; Adriano Kamimura Suzuki¹

Abstract

Regression models play an essential role in interpreting external factors and the effects of treatment on patient survival in clinical trials. A key challenge in developing such models is the appropriate specification of the failure time distribution, especially in the presence of censoring and a subset of individuals who exhibit a cure fraction. Defective models offer a flexible approach by adjusting the parametric space to incorporate cured individuals in the estimation process. The widely used Gompertz and inverse Gaussian distributions have demonstrated enhanced flexibility when incorporated into the Marshall-Olkin family. In this study, we introduce two novel regression models based on the Marshall-Olkin Gompertz and Marshall-Olkin inverse Gaussian distributions within a Bayesian framework. Through simulation studies, we assessed their frequentist properties under vague prior specifications. The results indicated low bias, minimal variability in point estimates, and coverage probabilities close to nominal levels. The proposed models were applied to a dataset of colon cancer patients, where the Marshall-Olkin Gompertz regression model was indicated as the best approach, outperforming even some parametric variations of the standard mixture model. The inferential summary and cure fraction estimates are presented for the real data analysis.

Keywords: Long-term survivors; MCMC; Lymph node counts; Fully parametric.

¹Department of Applied Mathematics and Statistics, ICMC, University of São Paulo, São Carlos, São Paulo - dionisioneto@usp.br, suzuki@icmc.usp.br

²Department of Statistics, CCET, Federal University of São Carlos, São Carlos, São Paulo - vera@ufscar.br



Reliability analysis of repairable systems considering unobserved heterogeneity and competing risks

Éder Silva de Brito ¹

Resumo

This work investigates failure processes in repairable systems by proposing new parametric frailty models that incorporate three key aspects: the type of repair performed (perfect or imperfect), the presence of competing risks, and unobserved heterogeneity. Although most studies assume only minimal repairs, this restrictive assumption limits their applicability in real-world scenarios. To address this limitation, we develop univariate and shared frailty models for multiple repairable systems, extending the existing literature by incorporating perfect repairs and various memory structures for imperfect repairs (ARA_m and ARI_m classes) within a competing risks framework. Unobserved heterogeneity is modeled using Gamma-distributed frailty terms, allowing the capture of random effects on failure times or across systems. A Power Law Process is assumed for the baseline failure intensity, and both frequentist and Bayesian approaches are adopted. Maximum likelihood estimation and asymptotic confidence intervals are obtained using numerical methods, while Bayesian inference is carried out via Markov Chain Monte Carlo algorithms. Simulation studies validate the proposed methodologies, which are also applied to real datasets, demonstrating their practical relevance. The developed models contribute to a deeper understanding of failure behavior, which is essential for improving preventive maintenance policies and reducing operational costs.

Palavras-chave: Repairable systems; Power law process; Frailty models; Unobserved heterogeneity; Competing risks.

¹Programa Interinstitucional de Pós-Graduação em Estatística - USP/UFSCar, São Carlos (SP) - edersilvabrito@gmail.com



Log-normal cure rate model in GAMLSS¹

Vanessa M. Silvério ²; Luiz R. Nakamura ³; Thiago G. Ramires ⁴; Rodrigo R. Pescim ⁵

Abstract

Cure rate regression models play an important role in survival analysis, modelling lifetime data for different forms of cancer. In recent years, significant efforts have been made to incorporate these models into the generalised additive models for location, scale, and shape (GAMLSS) framework, as this approach explicitly captures non-linear effects (through smoothing functions) of covariates on the failure time and/or cure fraction, while also allowing for the explicit modelling of other characteristics, such as data variability. In this study, we use the log-normal cure rate model within the GAMLSS framework to analyse melanoma data. The dataset includes 205 patients monitored after surgery for the removal of malignant melanoma between 1962 and 1977, with 72% of them being right-censored. The candidate explanatory variables for explaining the time until death, as well as the dispersion and cure fraction, are: ulceration (absent or present); tumour thickness (in mm); and sex (female or male). After using a stepwise-based method, we observed that all covariates were selected to explain both the median and variability of the time to death, whereas only tumour thickness was included in the cure fraction structure. A residual analysis indicated that the model provides a reasonable fit to the data.

Keywords: distributional regression, melanoma, survival analysis.

¹ This study was funded by FAPEMIG.

² Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, Universidade Federal de Lavras, Lavras – vanessa.silverio2@estudante.ufla.br

³ Departamento de Estatística, Universidade Federal de Lavras, Lavras – lui.z.nakamura@ufla.br

⁴ Departamento de Matemática, Universidade Tecnológica Federal do Paraná, Apucarana – thiagogentil@gmail.com

⁵ Departamento de Estatística, Universidade Estadual de Londrina, Londrina – rrpescim@uel.br



Zero-Adjusted Defective Gompertz Regression Model with Gamma Frailty: Applications in Medical Data

Pedro Rafael Diniz Marinho;¹ Cleide Mayra M. Lima;² Vera Lucia D. Tomazella;³
Agatha Sacramento Rodrigues;⁴

Resumo

In this article, introduced zero-adjusted defective Gompertz model, incorporating gamma frailty, designed for the analysis of health data. The proposed approach extends existing methodologies by addressing the combined challenges of excessive zeros, structural defects in the data, and unobserved heterogeneity through a unified framework. Specifically, the model leverages the flexibility of the Gompertz distribution for survival analysis while introducing adjustments to account for zero-inflation and structural anomalies commonly encountered in health datasets. The incorporation of gamma frailty adds a new dimension, allowing for the modeling of unobserved heterogeneity and capturing individual or unit-level variations more effectively. This innovation not only broadens the applicability of Gompertz-based models but also provides a robust tool for analyzing complex datasets. Applications demonstrate the model's superior performance in public health and epidemiology, offering new possibilities for studying intricate survival patterns and disease dynamics.

Palavras-chave: zero-adjusted, cure fraction, frailty term, defective distribution, gamma distribution.

¹Universidade Federal da Paraíba, Department of Statistics, João Pessoa - PB, Brazil - pedro.rafael.marinho@gmail.com

²Universidade Federal do Piauí, Department of Statistics, Teresina - PI, Brazil - cleide.mayra@eufpi.edu.br

³Universidade Federal de São Carlos, Department of Statistics, São Carlos - SP, Brazil - veratomazella@gmail.com

⁴Universidade Federal do Espírito Santo, Department of Statistics, Vitória - ES, Brazil - agatha.srodrigues@gmail.com



Análise de Sobrevivência Aplicada à Longevidade Reprodutiva de Vacas Nelore: Influência de Fatores Ambientais, Produtivos e Genéticos no Terceiro Parto

Sabrina Luzia Caetano ¹

Danísio Prado Munari ²

Salvador Boccaletti Ramos ³

Marcos Eli Buzanskas ⁴

Fernando Sebastián Baldi Rey ⁵

Resumo

Este estudo avaliou os fatores que influenciam a idade ao terceiro parto em vacas Nelore, utilizando análise de sobrevivência em um banco de dados com aproximadamente 50 mil vacas nascidas entre 2010 e 2017. Foram consideradas variáveis como idade ao primeiro parto (IPP), produtividade acumulada da vaca (PAC), estação de parto, ano de nascimento e efeito genético do pai (frailty). Os testes de Log-Rank e Wilcoxon mostraram diferenças significativas entre grupos de PAC, IPP, ano de nascimento e entre estações de parto ($p < 0,0001$). Vacas mais produtivas (PAC elevado) e com parição na estação da seca (abril a setembro) apresentaram maior frequência de terceiro parto em idades mais jovens, indicando maior eficiência reprodutiva. O modelo de Cox com frailty confirmou essas tendências: PAC e IPP foram fortemente associados ao risco de alcançar o terceiro parto ($p < 0,001$), com maiores PACs e menores IPPs elevando a probabilidade do evento. Além disso, vacas nascidas em anos mais recentes tiveram maior risco de atingir o terceiro parto precocemente. O efeito genético do pai foi significativo, com herdabilidade aditiva estimada em 0,227, sugerindo moderada influência genética sobre a característica. A estratégia reprodutiva que concentra acasalamentos nas águas, resultando em partos na seca, mostrou-se eficiente. Os resultados reforçam o valor do uso de PAC, IPP e dados de sobrevivência como critérios de seleção funcional em programas de melhoramento genético da raça Nelore.

Palavras-chave: Longevidade; Raça Nelore; Stayability; Análise de Sobrevivência; Modelo de Cox.

¹ Departamento de Ciências Exatas, Unesp-FCAV- sabrina.caetano@unesp.br

² Departamento de Ciências Exatas, Unesp-FCAV- danisio.munari@unesp.br

³ Departamento de Ciências Exatas, Unesp-FCAV- sb.ramos@unesp.br

⁴ Departamento de Melhoramento e Nutrição Animal, UNESP/FMVZ- m.buzanskas@unesp.br

⁵ Departamento de Zootecnia, Unesp-FCAV- fernando.baldi-rey@unesp.br



Análise dos fatores associados a demissão de empregados em empresas não públicas no Brasil

Fernando Humberto de Almeida Moraes Neto ¹ Edleide de Brito ² Maristela Dias Oliveira ³ Giovana Oliveira Silva ⁴

Resumo

Este estudo investiga os fatores associados à demissão de trabalhadores em empresas não públicas no Brasil, utilizando os dados da Relação Anual de Informações Sociais (RAIS) de 2023. Por meio de modelos de Análise de Sobrevivência, buscou-se compreender os determinantes do tempo de permanência no emprego em nível nacional. Foram avaliados quatro modelos paramétricos (exponencial, Weibull, gama generalizado e lognormal), sendo este último o que apresentou melhor desempenho, segundo o Critério de Informação de Akaike (AIC). Os dados passaram por um processo de filtragem para remover inconsistências, como idade ou tempo de emprego igual a zero, além do agrupamento de categorias com baixa frequência. Devido ao grande volume de dados (aproximadamente 50 milhões de vínculos), foi aplicada uma amostragem estratificada proporcional por região geográfica brasileira, o que possibilitou uma análise robusta dos fatores de risco. Os resultados mostram que os trabalhadores mais suscetíveis à demissão são, em sua maioria, homens jovens, pretos/pardos, com baixa escolaridade, salário de até três salários-mínimos e jornada semanal entre 41-44 horas. A média de tempo de permanência no emprego foi estimada em 34,5 meses, sendo a demissão sem justa causa o principal motivo de desligamento. As covariáveis sexo, idade, raça/cor e faixa salarial influenciaram significativamente o tempo até a demissão. O modelo lognormal mostrou bom ajuste, confirmado por meio da análise de resíduos. Os achados reforçam a necessidade de políticas públicas voltadas à qualificação e valorização profissional como forma de mitigar a rotatividade, especialmente entre os grupos mais expostos à demissão.

Palavras-chave: Análise de Sobrevivência; Demissão; Vínculo não público.

¹ Departamento de Estatística - Universidade Federal da Bahia - fernando.humberto@ufba.br

² Departamento de Estatística - Universidade Federal da Bahia - edbrito@ufba.br

³ Departamento de Estatística - Universidade Federal da Bahia - maridias@ufba.br

⁴ Departamento de Estatística - Universidade Federal da Bahia - giovana@ufba.br



Análise de sobrevivência aplicada no processo de melhoramento genético do cafeeiro

Lucas Pereira Belo ¹

Paulo César Emiliano ²

Resumo

O estudo utilizou técnicas de análise de sobrevivência para avaliar o desempenho produtivo de uma cultivar de café resultante do cruzamento entre Híbrido de Timor MG0357 e Tupi Amarelo IAC5162, com dados coletados anualmente entre 2016 e 2022. O evento de interesse foi definido como o momento em que a produção média por planta (em litros) foi igual ou superior as produções médias anuais das testemunhas Paraíso MG H419-1 e Catuaí Vermelho IAC 144. Foram analisadas características fenotípicas relacionadas à sanidade (ferrugem – *Hemileia vastatrix*, cercosporiose – *Cercospora coffeicola*, bicho-mineiro – *Leucoptera coffeella*) e à qualidade de bebida (ciclo de maturação, tamanho e uniformidade dos frutos). A probabilidade de produção foi estimada por curvas de Kaplan-Meier, com comparações realizadas por meio dos testes de log-rank e Wilcoxon, corrigidos por Bonferroni. Diferenças significativas ($p < 0,05$) foram observadas para as curvas de produção das covariáveis de sanidade e qualidade de bebida. Como exemplo, a uniformidade dos frutos apresentou grupos de maturação precoce e tardia que não diferiram entre si, enquanto níveis intermediários de maturação mostraram padrões distintos dos demais, mas foram equivalentes entre si. A abordagem permitiu identificar padrões temporais de produtividade associados às características fenotípicas, indicando que plantas com baixa ou moderada incidência de doenças apresentaram produtividade esperada em comparação com as testemunhas. Além disso, a cultivar demonstrou potencial para boa qualidade de bebida, e os testes empregados foram eficazes na identificação de diferenças precoces e tardias nas curvas de produção.

Palavras-chave: produção; log-rank; Kaplan-Meier; comparações; café arábica;

¹ Departamento de estatística - UFV – lucas.p.belo@ufv.br

² Departamento de estatística - UFV – paulo.emiliano@ufv.br



Enraizamento de miniestacas de *Eucalyptus* spp: uma aplicação da análise de sobrevivência semiparamétrica

Marciel Lelis Duarte ¹

Sebastião Martins Filho ²

A propagação vegetativa é a principal técnica utilizada na produção de mudas de eucalipto, sendo a indução do enraizamento adventício em casa de vegetação climatizada, uma das etapas mais importantes. O objetivo deste trabalho foi avaliar o uso do modelo de Cox no enraizamento de miniestacas de *Eucalyptus* spp e determinar o tempo ótimo de permanência de casa de vegetação. O experimento foi conduzido no Viveiro de pesquisas do Departamento de Engenharia Florestal da Universidade Federal de Viçosa por um período de 36 dias. Foi avaliado o enraizamento de três clones em função de quatro tipos de miniestacas (T1 - estaca apical com 2 pares de folhas com redução de 50% da área foliar; T2 - estaca apical com 2 pares de folhas inteiras; T3 - estaca apical com 3 pares de folhas inteiras; e T4 - estaca basal com 1 par de folhas inteiras). Os parâmetros estimados do modelo foram utilizados para determinar as funções de sobrevivência e a função de distribuição acumulada. Os resíduos de Cox-Snell foram utilizados para verificar a qualidade do ajuste. A covariável clone violou a pressuposição de taxas de falha, desta forma o modelo foi estratificado para atender essa pressuposição. O tipo de minestaca T1 (estaca apical com 2 pares de folhas com redução de 50% da área foliar), apresentou uma taxa de enraizamento de 1,3 vezes maior que os demais tipos de miniestacas. O tempo ótimo de permanência na casa de vegetação foi de 30 dias.

Palavras-chave: Dados censurados; Modelo de Cox; Propagação vegetativa.

¹ Departamento Estatística, Universidade Federal de Viçosa – marciel.duarte@ufv.br

² Departamento Estatística, Universidade Federal de Viçosa – martinsfilho@ufv.br



Sobrevida de Pacientes Acometidos Pela COVID-19 na Paraíba e Fatores de Risco

Daví Barbosa Pereira de Sousa;¹ Hellen Sonaly Silva Alves;² Vitoria Soares de Souza;³ Guilherme Rocha Moreira;⁴ João Domingos Scalon⁵

Resumo

A pandemia mais recente enfrentada pela humanidade foi a da COVID-19, causando inúmeras perdas aos cidadãos e sequelas aos pacientes sobreviventes. Na Paraíba, a situação não foi diferente, assim, utilizando dados do Sistema de Informações Hospitalares do SUS no período de março de 2020 a dezembro de 2024, visa-se entender a sobrevida e fatores de risco de pacientes paraibanos acometidos pelo coronavírus. Por meio da estimativa da curva de sobrevida via Kaplan-Meier, percebeu-se que, a partir do décimo dia de internação a probabilidade de sobrevida do paciente era inferior a 0,77 e no 44º dia inferior a 0,20. Realizando o ajuste da regressão de Cox com as covariáveis Faixa, Sexo e Ano, notou-se que, comparado a faixa etária de 0 a 19 anos, a de 20 a 39 possui risco de morte 3,94 vezes maior, 40 a 59 anos é 5,28 vezes maior e ≥ 60 anos é 8,09 vezes maior. Em comparação ao sexo masculino, o feminino apresentava risco 5,26% menor. Quanto aos anos de internação, o risco foi reduzido em comparação a 2020, sendo 2021 12,09% menor, 2022 26,17% inferior, 2023 foi 28,01% menor e 2024 21,14% menor (não significativo, valor-p > 0,05). Os resultados obtidos mostram que, fatores como idade e sexo influenciam significativamente a sobrevida dos pacientes, evidenciando a importância de estratégias direcionadas de prevenção e cuidado com grupos mais vulneráveis. Fatores externos, não abordados, também influenciaram o ano de internação.

Palavras-chave: Internação; Regressão de Cox; Kaplan-Meier; SUS; Mortalidade.

¹Universidade Federal de Lavras, Departamento de Estatística, Lavras, MG - davi.sousa@estudante.ufla.br

²Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, Recife, PE - hellen.ssilva@ufrpe.br

³Universidade Estadual da Paraíba, Departamento de Estatística, Campina Grande - vitoria-soares0104@gmail.com

⁴Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, Recife, PE - guilherme.rochamoreira@ufrpe.br

⁵Universidade Federal de Lavras, Departamento de Estatística, Lavras, MG - scalon@ufla.br



Regressão Inversa Gaussiana Parametrizada na Moda e Dispersão e Aplicações em Análise de Sobrevida

Jessica Hemily de Sá Gonçalves;¹ Josmar Mazucheli²

Resumo

A distribuição inversa gaussiana é uma distribuição de probabilidade unimodal, contínua, com suporte nos números reais positivos e pertencente à família exponencial. Devido ao seu suporte positivo, ela é frequentemente usada na análise de dados de sobrevida.

No contexto de regressão, dependendo dos parâmetros da função densidade de probabilidade (fdp) da inversa gaussiana — e, consequentemente, da assimetria dos dados — modelar a média pode não ser a melhor escolha, uma vez que ela é sensível a valores aberrantes. Nesse sentido, uma alternativa intuitiva é modelar a moda, por sua facilidade de interpretação e maior estabilidade em relação aos valores extremos. Além disso, uma alternativa interessante para o segundo parâmetro é o parâmetro de dispersão, devido à sua interpretabilidade direta. Tal parametrização, proposta em Punzo (2019), é o objeto principal deste estudo, no qual foi desenvolvida uma rotina de modelagem considerando a distribuição inversa gaussiana parametrizada pela moda e pela dispersão.

Adicionalmente, foi empregada a função flexsurvreg, do pacote Flexsurv (disponível no software R, versão 2024), para a análise de sobrevida, mediante a especificação da nova distribuição. Por fim, realizou-se um estudo de simulação, variando a fração de observações censuradas, para demonstrar a aplicabilidade e avaliar o desempenho da modelagem modal em comparação com a abordagem tradicional, que utiliza a parametrização pela média.

Palavras-chave: Moda; Inversa gaussiana; Inversa gaussiana reparametrizada; Máxima verossimilhança; Sobrevida; Análise de sobrevida paramétrica.

Referências

PUNZO, A. A new look at the inverse gaussian distribution with applications to insurance and economic data. *Journal of Applied Statistics*, Taylor Francis, v. 46, n. 7, p. 1260–1287, 2019.

¹Programa de Pós-Graduação em Bioestatística, Universidade Estadual de Maringá, Paraná, Brasil - pg405656@uem.br

²Departamento de Estatística, Universidade Estadual de Maringá, Paraná, Brasil - jmazucheli@uem.br



Aplicação de modelos de análise de sobrevida via aprendizado de máquina em dados de câncer de mama

Nayara Maria Barbosa de Sousa¹

Geraldo Magela da Cruz Pereira²

Resumo

Este estudo tem por objetivo aplicar algoritmos de aprendizado de máquina para a previsão da sobrevida de mulheres diagnosticadas com câncer de mama. Os dados são provenientes de um estudo de coorte de casos de câncer de mama em mulheres tratadas entre 2005 e 2015 no Hospital da Fundação Assistencial da Paraíba, em Campina Grande, no estado da Paraíba, Brasil. O conjunto de dados inclui informações clínicas e terapêuticas, como idade, tipo histológico do tumor, número de doses de tratamentos (hormonioterapia, quimioterapia e radioterapia), além de marcadores moleculares (receptores de estrogênio e de progesterona, proteínas Ki67, P53 e HER2), e as datas da primeira consulta, da última nomeação e do óbito (quando aplicável). Dada a natureza censurada dos dados, foram aplicados os algoritmos: Random Survival Forest (RSF), Gradient Boosting Machine (GBM) e Support Vector Machine para sobrevida (SVMc). O desempenho dos algoritmos foi avaliado por meio do C-index. Espera-se que os resultados deste estudo contribuam para uma melhor estratificação de risco e forneça suporte à tomada de decisão clínica, considerando o poder preditivo desses algoritmos para a estimativa do tempo até o evento estudado.

Palavras-chave: sobrevida; câncer de mama; modelos de aprendizado de máquina.

¹ Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, PPGEE, UFLA – nayara.sousa@estudante.ufla.br

² Departamento de Estatística, UFLA – geraldo.pereira@ufla.br



Parametric Model for Correlated Survival Data in the Presence of Competing Risks and Interval Censored

Márcio Augusto F. Rodrigues ¹

Enrico A. Colosimo ²

Juliana V. Bastos ³

Sylvia C. Coste ⁴

Resumo

Interval censored in survival analysis occurs when the survival time of the individuals under study is known to only belong to a time interval. In classic survival analysis, a single cause for the occurrence of an event is considered. However, some studies may be interested in the occurrence of more than one event, which is called competing risks. Hudgens et al. (2014) considered the parametric modeling of the cumulative incidence function for competing risk data subject to interval censored. In this study, we extend the work of Hudgens et al. (2014) to accommodate the presence of clusters. The proposed methodology considers a Generalized Estimating Equations (GEE)-type model using an independent work matrix. The model performance was assessed using a simulated dataset. We demonstrated that the proposed methodology has good properties for small samples. The proposed method was applied to a real dental trauma dataset.

Palavras-chave: Clusters; GEE; Sandwich matrix; Weibull model; Work matrix.

¹ Instituto de Matemática e Estatística, Universidade Federal de Goiás, Goiânia - marcioaugusto@ufg.br

² Departamento de Estatística, Universidade Federal de Minas Gerais, Belo Horizonte - enricoc57@gmail.com

³ Departamento de Odontologia Restauradora, Universidade Federal de Minas Gerais, Belo Horizonte - julianavbtrauma@gmail.com

⁴ Departamento de Odontologia Restauradora, Universidade Federal de Minas Gerais, Belo Horizonte - sylviacury@hotmail.com



Análise de sobrevivência aplicada à germinação de sementes expostas à água ativada por plasma

Matheus Custódio da Silva ¹

Sara Silvério ¹

Verônica Manhães Saint'Clair ¹

Marciel Lelis Duarte ¹

Sebastião Martins Filho ¹

Resumo

Embora existam outras abordagens estatísticas para o estudo da germinação de sementes, a presença de censuras nesse tipo de investigação torna a análise de sobrevivência um método mais apropriado. O objetivo deste trabalho foi avaliar a influência da exposição à água ativada por plasma na germinação de sementes de alface e tomate, por meio da técnica de análise de sobrevivência paramétrica. Para a realização dos experimentos, as sementes foram submetidas a dois tratamentos: com e sem plasma. Após a instalação do experimento, as sementes foram transferidas para o germinador a 23 °C sob um fotoperíodo de 16 h. A germinação foi avaliada a cada 24 horas durante 12 dias. Após este tempo as sementes que não germinaram foram consideradas censura, e as germinadas as falhas. Foram ajustados os modelos de sobrevivência Exponencial, Weibull, Log-normal e Log-logístico para estimar o tempo de germinação de sementes. Os parâmetros estimados dos modelos estudados foram utilizados para determinar as funções de sobrevivência e a função de distribuição acumulada; o Critério de Akaike (AIC) para realizar a seleção do melhor modelo e os resíduos de Cox-Snell para verificar o ajuste. O modelo Log-normal apresentou o menor valor AIC entre os modelos avaliados e um ajuste satisfatório, para ambas as espécies. O parâmetro associado ao tratamento com água ativada por plasma não se mostrou estatisticamente significativo ($p > 0,05$) para a alface e o tomate. Dessa forma, conclui-se que o tratamento com plasma não tem efeito na germinação das sementes dessas espécies.

Palavras-chave: Dados censurados; Modelos paramétricos; Testes de germinação.

¹ Departamento de Estatística, Universidade Federal de Viçosa – matheus.c.silva@ufv.br; sara.silverio@ufv.br; veronica.saintclair@ufv.br; marciel.duarte@ufv.br; martinsfilho@ufv.br



Globally orthogonal parametrizations for log-location-scale models

Karim Anaya-Izquierdo;¹ Daniele Cristina Tita Granzotto;² Brian A. R. de Melo³

Resumo

This paper introduces a framework for constructing globally Fisher orthogonal parameterization for extended log-location-scale distributions. Using classical orthogonal polynomials, we extend standard parametric families by introducing a new parameter that adjusts the distribution's shape while preserving orthogonality with the scale parameter. This orthogonality simplifies maximum likelihood estimation by ensuring asymptotic independence of estimators and avoiding confounding. The method relies on projecting score functions and constructing new distributions via exponential tilting of a baseline density. Each extension uses a specific orthogonal polynomial (e.g., Hermite, Laguerre, Legendre) and a corresponding transformation function to ensure orthogonality in a baseline distribution (e.g., Lognormal, Exponential, Log-logistic). The approach is illustrated with well-known distributions demonstrating how the hazard and cumulative hazard functions are affected. We briefly explore a three-parameter extension, which includes a power parameter and shows that exact global orthogonality can still be achieved in certain cases. A simulation study is carried out and validates the theoretical developments, showing the consistency and independence of the maximum likelihood estimators.

Palavras-chave: Fisher orthogonality; log-location-scale models; orthogonal polynomials; survival analysis.

¹Department of Mathematical Sciences, University of Bath, Bath, Somerset, UK

²Departamento de Estatística, Universidade Estadual de Maringá, Maringá, PR

³Programa de Pós-graduação em Bioestatística, Universidade Estadual de Maringá, Maringá, PR - barmelo@uem.br



Modelo de cura com tempo de promoção integrado com algoritmos de classificação de aprendizado de máquina.

Daiane de Souza Santos;¹ Vicente Garibay Cancho ²

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. Existem 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. Neste 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 de cura com tempo de promoção que utiliza os algoritmos de classificação *Extreme Gradient Boosting*, florestas aleatórias, redes neurais, *support vector machine*, redes neurais e árvores de decisão para modelar a parte de incidência. O novo modelo herda as características dos algoritmos para fornecer um modelo flexível para avaliar os efeitos das covariáveis na incidência. Desenvolvemos um método de estimação para o modelo de cura e realizamos um estudo de simulação.

Palavras-chave: Análise de sobrevivência, Aprendizado de máquina; Modelo de cura com tempo de promoção; Algoritmo EM; Algoritmos de classificação.

¹Instituto de Ciências Matemáticas e de Computação-USP, São Carlos - daiane@icmc.usp.br

²Instituto de Ciências Matemáticas e de Computação-USP, São Carlos - garibay@icmc.usp.br



Survival Prediction in Colorectal Cancer: A Random Forest Approach

Rafael Sant'Ana Herzog ¹
Vinicius Fernando Calsavara ²
Agatha Sacramento Rodrigues ³

Resumo

With technological advancements and increased computational power, machine learning methods have gained significant popularity. However, as the range of available models expands, it becomes increasingly evident that understanding the structure of a dataset under study is crucial, and determining whether the use of a certain model makes sense within this structure is essential. In this context, we revisit a study on the survival of colorectal cancer patients treated in the state of São Paulo between 2000 and 2021, which, among other issues, ignored the existence of censored observations, using standard classification methods to predict the probability of patients surviving for at least a certain period of time. We discuss how ignoring the existence of censored data can bias the results of a study and present Random Survival Forests (RSFs), a model adapted from regression random forests that incorporate censoring into its formulation. We fitted several RSFs to the original dataset and obtained, for the final model, a C-index of 76% and an Integrated Brier Score of 0.149. Furthermore, using SHAP values, we determined that being in clinical stage IV, not having undergone surgery and having undergone chemotherapy during treatment were the most impactful factors in the model's predictions. Finally, we conducted a simulation study to assess the impact that different levels of censoring have on the predictions of the RSFs and the usual classification models, obtaining results that demonstrate the inadequacy of the latter models for dealing with datasets containing censored information.

Palavras-chave: Colorectal Cancer; Machine Learning; Random Survival Forests; C-index; SHAP values; Simulation Studies.

¹Department of Statistics at the Universidade Federal do Espírito Santo - rafael.herzog@edu.ufes.br

²Cedars-Sinai, Los Angeles, US - vinicius.calsavara@cshs.org

³Department of Statistics at the Universidade Federal do Espírito Santo - agatha.rodrigues@ufes.br



The Defective ZTP-Gompertz regression models in the presence of competing and complementary risks for modeling long-term survivors

Cynthia A. V. Tojeiro;¹ Pedro Luiz Ramos² Vera Tomazella³

Abstract

This article introduces a novel method for modeling cure fractions, focusing on scenarios where some observations are not affected by the primary event of interest. A new distribution model is presented, based on the Poisson distribution and the defective model known as Gompertz distribution. The proposed model present an innovative approach in the relevant scenario of competitive and complementary risk scenarios. Here, the latent variables follow a zero-truncated Poisson distribution (ZTP). The distribution's shape parameter holds crucial physical significance in competing and complementary risks. The mathematical properties and inferential procedures are discussed in detail. Maximum likelihood estimation is used for parameter estimation, and simulation studies assess its accuracy. This approach considers censored data and covariates, demonstrating the resulting distribution's applicability for modeling cure fractions. The method is illustrated with a dataset from cancer patients.

Keywords: Cure Fraction; Defective Distribution; Zero-Truncated Poisson Distribution; Gompertz Distribution; Regression Model.

¹PPGMDS-UFPB (e IME/UFG) , João Pessoa-Paraíba (Goiânia-GO) - cynthiatojairo@ufg.br

²Faculdade de Matemática- PUC, Santiago-Chile -pedro.ramos@mat.uc.cl

³PPGMDS-UFPB, João Pessoa-Paraíba - vera@ufscar.br



Modelos de Sobrevivência com Fração de Cura e Fragilidade Espacial usando Distribuições Defeituosas: Uma Aplicação em Obstetrícia

José Carlos Soares Junior;¹ Agatha Sacramento Rodrigues²

Resumo

As bases nacionais de dados obstétricos SIM-DOFET e SINASC reúnem informações detalhadas sobre óbitos fetais e nascidos vivos em todo o território brasileiro, permitindo investigar o tempo até o evento em nível individual e espacial. Entretanto, os modelos de mistura padrão convencionais frequentemente exigem múltiplos componentes para ajustar simultaneamente a fração de cura e a correlação geográfica do risco, o que pode aumentar o custo computacional devido ao aumento de parâmetros decorrentes do método de construção desse tipo de modelo. O objetivo deste trabalho é propor um modelo de análise de sobrevivência que incorpore de forma integrada a fração de cura — representando as gestações que evoluem sem óbito fetal — e um termo de fragilidade aleatório espacial, capturando a dependência espacial (por exemplo, estabelecimentos de saúde) do risco fetal e oferecendo uma solução mais parcimoniosa que os métodos de mistura padrão. Assim, adotamos a distribuição Gompertz defeituosa, capaz de descrever diretamente a dinâmica de risco ao longo da idade gestacional e a proporção de cura. A componente espacial é representada por um termo de fragilidade aleatória com estrutura SAR (Spatial Autoregressive Model), integrando informações de vizinhança geográfica. A inferência será conduzida via INLA (Integrated Nested Laplace Approximation), viabilizando estimativa bayesiana eficiente em cenários de alta dimensionalidade. O modelo será aplicado aos dados do SIM-DOFET e do SINASC para avaliar desempenho preditivo, interpretabilidade dos parâmetros espaciais e identificar padrões regionais de risco fetal.

Palavras-chave: Mortalidade fetal; Análise de sobrevivência; Fração de cura; Fragilidade espacial; INLA.

¹OONB/Departamento de Obstetrícia e Ginecologia, FMUSP - josecarlok905@usp.br

²OONB/Departamento de Estatística, UFES - agatha.rodrigues@ufes.br



Uso de Redes Neurais Recorrentes na Previsão de Precipitação em Séries Temporais para Campina Grande-PB

Gabriel Messias Santana Peixoto;¹ Mateus Santos Peixoto;² Giullber Valentim da Silva;³ Mateus Silva Rocha;⁴ Prof. Dr. Ricardo Alves de Olinda;⁵ Prof. Dr. João Domingos Scalon;⁶

Resumo

Este estudo visa prever a precipitação acumulada em Campina Grande (PB) utilizando redes neurais recorrentes do tipo Long Short-Term Memory (LSTM). Os dados foram obtidos na plataforma Google Earth Engine, especificamente do produto ERA5-Land, que oferece reanálises climáticas com alta resolução temporal e espacial. O período analisado abrange de 2010 a 2023, totalizando 168 meses de registros mensais.

Inicialmente, a série temporal foi decomposta para identificar padrões como tendências e sazonalidades, sendo identificada apenas a sazonalidade. Em seguida, aplicou-se o teste de estacionariedade Dickey-Fuller Aumentado (ADF), confirmando a estacionariedade ao nível de 5% de significância.

Utilizou-se o método de janela deslizante para transformar o problema sequencial em uma tarefa de aprendizado supervisionado. Três arquiteturas foram desenvolvidas com a biblioteca TensorFlow: (i) um modelo de referência simples com uma camada de achatamento e uma densa; (ii) um modelo com cinco camadas LSTM de 50 neurônios cada; e (iii) um modelo otimizado via Random Search, contendo três camadas LSTM (64, 64 e 50 neurônios) e uma camada densa final.

A métrica de avaliação adotada foi o Erro Quadrático Médio (EQM), sendo o modelo otimizado o que apresentou melhor desempenho. Recomenda-se, para estudos futuros, a integração de abordagens estatísticas clássicas com redes neurais, a fim de aprimorar ainda mais a previsão.

Palavras-chave: Deep Learning; long short-term memory; precipitação.

¹Estatística e Experimentação Agropecuária UFLA, Lavras - gabrielpeixotostats@gmail.com

²Estatística e Experimentação Agropecuária UFLA, Lavras - mateus_peixoto12@hotmail.com

³Estatística Estatística e Experimentação Agropecuária UFLA, Lavras - giullberguitar@gmail.com

⁴Departamento de Estatística - UFPE, Recife - mattrocha57@gmail.com

⁵Departamento de Estatística - UEPB, Campina Grande - prof_ricardo@servidor.uepb.edu.br

⁶Lavras - scalon@ufla.br



Adesão de políticas públicas da área da saúde por municípios brasileiros via modelo de fragilidade compartilhada

Milena Souza Gonçalves¹ Juliana Betini Fachini Gomes² Mateus Felipe Santos Araújo³

Celina Pereira⁴ Denilson Bandeira Coêlho⁵

Resumo

O que causa a adoção de um novo programa por outros governos? Por que apenas alguns governos adotam determinadas políticas? O que faz com que uma política se dissemine? Quais fatores explicam a velocidade dessa disseminação? Estas são algumas das questões relevantes em estudos de difusão de políticas. Embora ainda incipientes, os estudos sobre difusão de políticas públicas no Brasil são promissores, dada a estrutura federativa altamente descentralizada do país. Os entes sub-nacionais — estados e municípios — possuem significativa autonomia para implementar políticas públicas, o que cria um ambiente fértil para a difusão horizontal (entre municípios ou estados) (COÊLHO; CAVALCANTE; TURGEON, 2016). Esse cenário realça a importância de compreender como, por que e em que ritmo as políticas públicas são adotadas, especialmente em áreas críticas como a saúde pública. Desta forma, o objetivo deste trabalho é identificar fatores que influenciam o tempo até a adoção de nove políticas da área da saúde por municípios brasileiros. O modelo de fragilidade compartilhada foi utilizado por assumir que os tempos de um mesmo município são independentes condicional às variáveis não mensuradas. Variáveis explicativas das dimensões de complexidade das políticas, governabilidade, capacidades institucionais, inovação ministerial, eleitoral e de controle foram consideradas na análise. Os resultados revelaram que diversas variáveis influenciam significativamente a adesão, e a fragilidade compartilhada entre os municípios demonstrou ser relevante. O modelo mostrou-se robusto para compreender a complexa dinâmica da adesão de políticas de saúde em nível municipal no Brasil.

Palavras-chave: políticas públicas; difusão; múltiplos eventos; municípios brasileiros; modelo de fragilidade compartilhada.

¹ Departamento de Estatística, Universidade de Brasília – miawnsouza@gmail.com

² Departamento de Estatística, Universidade de Brasília – jfachini@unb.br

³ Departamento de Estatística, Universidade de Brasília – mateusfelipesa@hotmail.com

⁴ Ministério da Gestão e da Inovação em Serviços Públicos , Brasília – celina.pereira@gestao.gov.br

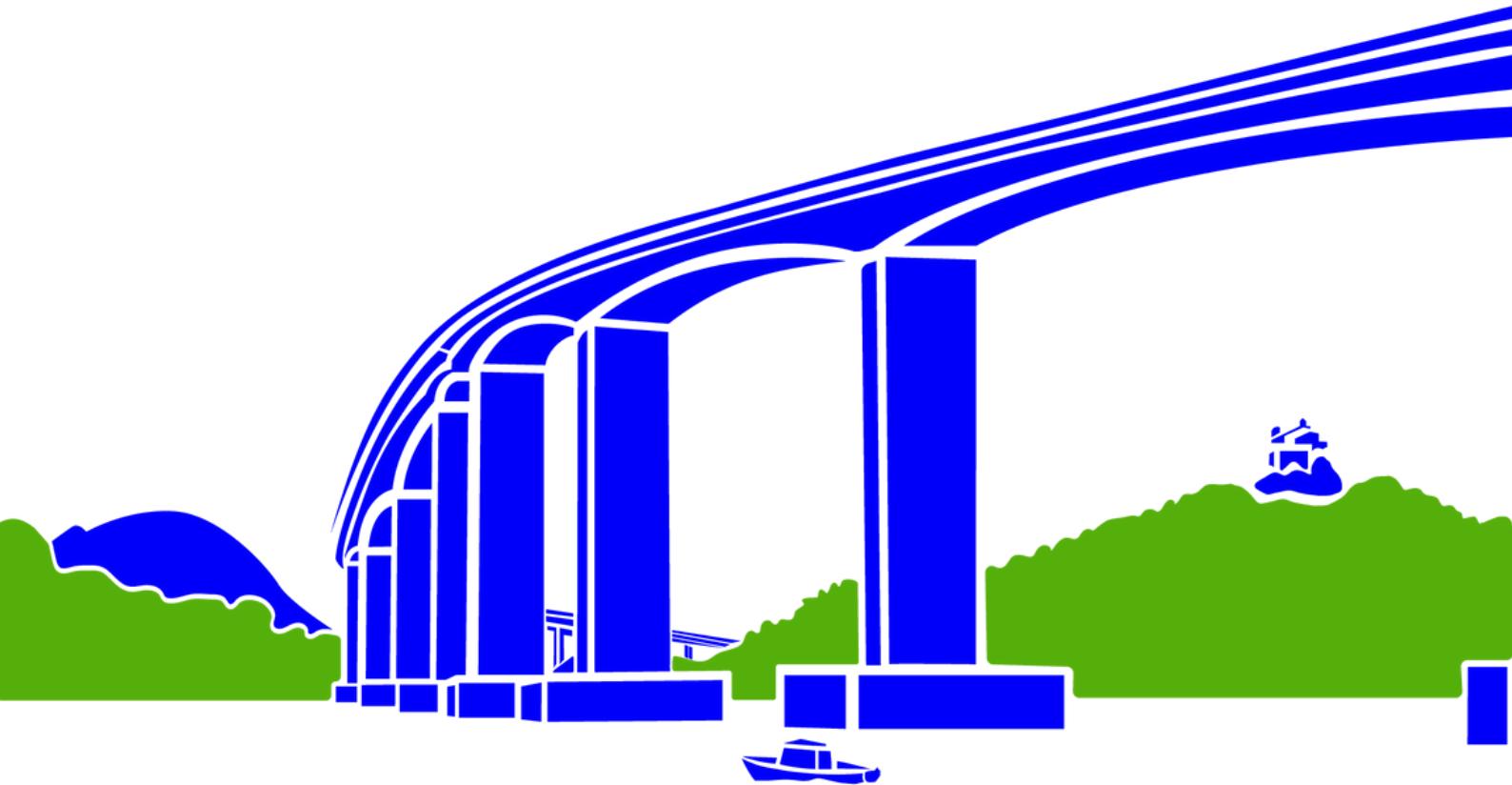
⁵ Departamento de Ciências Política, Universidade de Brasília – denilsonbc@unb.br

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Pôsters

**Estatística Aplicada em
Agronomia e Biologia 1**





Frequência de mensurações de variáveis meteorológicas para o cálculo da evapotranspiração de referência

Alessandro Dal'Col Lúcio¹ Aline Ulzefer Henck¹ Sidinei Zwick Radons²

Lana Bruna de Oliveira Engers¹

Resumo

A utilização de sistemas de irrigação via estimativa da evapotranspiração de referência com um conjunto mínimo e preciso de dados meteorológicos é fundamental nos sistemas agrícolas. Assim este trabalho objetivou caracterizar o banco de dados das variáveis meteorológicas sob distintas frequências de observações e classificações climáticas, bem como determinar a frequência mínima de observação das variáveis meteorológicas mensuradas. Os dados de temperatura, umidade relativa do ar, radiação solar global incidente e velocidade do vento foram obtidos diariamente, em intervalos de 15 em 15 minutos, por meio de três estações meteorológicas situadas em distintas regiões climáticas brasileiras (Aw, Cfa, Cwa). A partir do banco de dados original, simularam-se outras nove frequências de mensurações de cada variável meteorológica e foi estimada a evapotranspiração de referência pelo método de Penman-Monteith. Foram feitas análises descritivas e verificado o atendimento aos pressupostos de normalidade dos dados e homogeneidade das variâncias. A análise de agrupamento foi realizada pelo Método Hierárquico Aglomerativo do vizinho mais distante adotando 70% de semelhança para a formação dos grupos. O clima Cwa apresentou as maiores médias de temperatura do ar, enquanto o Cfa as menores com maior variabilidade nos dados. Para a umidade relativa do ar, o Cfa registrou as maiores médias com menor variabilidade. A radiação solar global incidente foi maior no Aw, e a velocidade do vento, no Cfa e Aw, sendo a evapotranspiração de referência também mais significativa no clima Aw. A frequência mínima de mensurações de variáveis meteorológicas por meio de estações meteorológicas visando o cálculo da evapotranspiração de referência, foi de observações em intervalos a cada 2 horas nas distintas classificações climáticas. Exceções foram verificadas no caso da temperatura do ar no clima Cfa (frequências de até 6 horas), para a velocidade do vento no clima Cfa e para a evapotranspiração de referência no clima Cwa (frequências de até 4 horas).

Palavras-chave: Dados meteorológicos, Analise de agrupamento, Dendograma.

¹ Departamento de Fitotecnia, Centro de Ciências Rurais, Universidade Federal de Santa Maria, Santa Maria - RS – adlucio@uol.com.br, alineulzeferhenck@gmail.com, engers.lana@gmail.com

² Universidade Federal da Fronteira Sul, Cerro Largo – RS – sidineiradons@gmail.com



Abordagem Bayesiana utilizando Modelos Não Lineares na Predição do Crescimento de Coelhos

Laura Lucia Dominguez Barrios ¹

Felipe Augusto Fernandes ²

Mariana Rodrigues Motta ³

Resumo

A criação de coelhos apresenta alto potencial de crescimento devido ao fácil manejo, alta proliferação, baixo impacto ambiental e adaptabilidade aos sistemas de agricultura familiar, além de produzir carne de alto valor biológico. No Brasil, a aceitação do consumo de carne ainda é baixa, o que limita a produção e favorece a criação para fins domésticos, impulsionada pela demanda crescente por animais de companhia. Estudos sobre essa atividade podem fornecer informações relevantes e incentivar sua expansão. O manejo tem papel fundamental no crescimento dos coelhos, especialmente após o desmame, período em que fatores como genética, nutrição, condições sanitárias e ambiente térmico influenciam diretamente o desenvolvimento dos animais. Uma criação bem conduzida pode otimizar o ganho de peso e garantir melhores índices produtivos. A modelagem do crescimento por meio de modelos não lineares permite identificar padrões morfológicos e compreender melhor esse processo. Assim, este estudo teve como objetivo comparar os modelos Logístico, Gompertz e von Bertalanffy na descrição das curvas de crescimento de coelhos, utilizando a abordagem Bayesiana. Os dados foram retirados de Fernandes et al. (2022) e referem-se a 10 coelhos mestiços de Gigante de Flandres e Nova Zelândia Branco, cuja pesagem ocorreu de 0 a 150 dias de idade. Os resultados indicaram o modelo Gompertz como o mais adequado para essa descrição, fornecendo médias assintóticas de 3552 kg.

Palavras-chave: curvas de crescimento; inferência bayesiana; modelo Gompertz; ; cunicultura; modelo logístico; von Bertalanffy; rstan

¹Departamento de Estatística, Universidade Estadual de Campinas, Brasil - l175089@dac.unicamp.br

²Departamento de Estatística, Universidade Estadual de Campinas, Brasil - felipeaf@unicamp.br

³Departamento de Estatística, Universidade Estadual de Campinas, Brasil - marirm@unicamp.br



Influência de variáveis meteorológicas nas produtividades de caroço, pluma e semente de algodão no estado do Mato Grosso

Leonardo Cesar Pradebon ¹

Alberto Cargnelutti Filho ²

Murilo Vieira Loro ³

Lais Cherobini ⁴

João Augusto Andretta ⁵

Resumo

O objetivo deste trabalho foi identificar se há influência de variáveis meteorológicas nas produtividades de caroço, pluma e semente de algodão (*Gossypium hirsutum* L.) no estado do Mato Grosso. Foram utilizadas as produtividades de caroço (PC), pluma (PP) e semente (PS) de algodão, do estado do Mato Grosso, dos anos agrícolas 1978/79 a 2022/23, obtidos no site da Conab. Os dados de temperaturas máxima (Tmax), média (Tmed) e mínima (Tmin) do ar, umidade relativa do ar (UR) e precipitação (Prec) foram obtidos no site do INMET. Foi realizada a análise de correlação linear de Pearson. Estimaram-se os parâmetros da árvore de regressão para a predição da PC, PP e PS em função das variáveis meteorológicas. Fez-se a análise de componentes principais para avaliar a relação multivariada entre as variáveis meteorológicas e PC, PP e PS. Conclui-se que as maiores temperaturas do ar associadas às baixas UR e Prec tendem a aumentar a PC, PP e PS de algodão.

Palavras-chave: correlação; árvore de regressão; análise de componentes principais.

¹ Programa de Pós-Graduação em Agronomia, Universidade Federal de Santa Maria (UFSM) - leonardopradebon@gmail.com

² Departamento de Fitotecnia, UFSM - alberto.cargnelutti.filho@gmail.com

³ Departamento de Fitotecnia, UFSM - muriloloro@gmail.com

⁴ Programa de Pós-Graduação em Agronomia, Universidade Federal de Santa Maria (UFSM) - laischerobini@gmail.com

⁵ Curso de Agronomia, Universidade Federal de Santa Maria (UFSM) - aundregusto@gmail.com

Agradecimentos ao CNPq (Processo 304878/2022-7), PIBIC/CNPq/UFSM, PROBIC/FAPERGS/UFSM e Capes (Finance Code 001), pelo auxílio financeiro e bolsas concedidas aos autores.



APLICAÇÃO DE REDES NEURAIS E INTELIGÊNCIA ARTIFICIAL PARA A PREDIÇÃO DA QUALIDADE DO FARELO DE SOJA.

Lucas Port Bellen Leite.¹ Renan de Siqueira Cecchin.² Denilson dos Santos Ebling.²

Lucio Strazzabosco Dorneles.³

Resumo

O avanço do conhecimento sobre redes neurais apresentaram novas possibilidades de aplicação no agronegócio. Este trabalho explora o emprego de redes neurais para fazer previsões precisas de parâmetros do farelo de soja. A obtenção dos dados se deu por amostras do grão de soja moído em espectroscopia no infravermelho próximo. O modelo foi estabelecido de acordo com a restrição de tamanho de 10 megabytes, sendo executado em aplicativo mobile, com brevidade de análise a campo. Para cada amostra analisada, há um conjunto de valores de absorbância relacionados. Em anos anteriores, foram empregadas alternativas como o *SNV (Standard Normal Variate)*, *PLSR (Partial Least-Squares Regression)*, dentre outros. Em 2024, com a implementação de redes neurais, visando reduzir o erro de predição, e acessando as métricas do *RMSE* e do *R²*, atingiram-se resultados de 0,27 e 0,29. Esses valores representaram melhoria em relação aos algoritmos anteriores. Seguindo a bibliografia mais recente na área, o modelo foi desenvolvido em rede convolucional unidimensional, passando por camadas convolucionais, função de ativação ELU, camadas Max Pool e camadas lineares. Contudo, nota-se que o modelo de rede neural é de inferência paramétrica imprevisível, exigindo constante reajuste dos hiperparâmetros intrínsecos a cada camada. No presente ano, obteve-se subsequente melhoria nos resultados. O *RMSE* e o *R²* atuais foram de 0,49 e 0,44. Assim sendo, as redes neurais são algoritmos com potencial para a predição de parâmetros, e sua escalabilidade favorece o produtor rural em sua tomada de decisão.

Palavras-chave: Agronegócio; Soja; Estatística; Redes Neurais.

¹ Departamento de Estatística, Centro de Ciências Naturais e Exatas, Universidade Federal de Santa Maria. – lucasp.pbl@gmail.com

² Zeit Soluções em Inteligência Artificial Ltda., Incubadora Pulsar da Universidade Federal de Santa Maria – denilson@zeit.com.br, renanceccchin07@gmail.com

³ Departamento de Física, Centro de Ciências Naturais e Exatas, Universidade Federal de Santa Maria. – lucio.dorneles@ufrgs.br



Impact on genetic gain from assuming different statistical models in generating designs for early generation plant-breeding experiments

Renata Alcarde Sermarini¹, Christopher James Brien^{2,3}, Clarice Garcia Borges Demétrio⁴, Alessandra dos Santos^{4,5}, Alessandra de Lima Goes⁴, Denize Palmito dos Santos⁴

Resumo

This work combines the results of three papers that investigated assumptions about the model used to create optimal designs for early generation plant breeding experiments. These experiments aim to maximize realized genetic gain (RGG) using limited resources, which is challenging due to the need for understanding variation sources and their magnitudes, often unknown. Simulations were conducted to identify robust designs for various experimental conditions. The first paper focused on partially replicated (p -rep) designs, where $p\%$ of the test lines are replicated. It explored different conditions for two experimental area sizes. It was concluded that for large experiments, the RGG from various optimal designs remains stable across different variation scenarios, suggesting resolved designs for fixed genetic effects should be used for robustness. Design assumptions impact RGG in smaller p -rep designs. The second paper evaluated p -rep and grid-plot designs, assuming fixed test line effects. It showed that the number of check varieties has minimal impact on RGG and model variance estimates, as well as the optimality criteria. The third paper combined the findings of the first two, focusing on multisite resolved experiments. It found that spatially optimized grid-plot and p -rep designs, for both individual and joint analyses, showed p -rep designs were slightly better in selecting genetic material. Despite small differences across scenarios, the p -rep design without checks achieved the highest genetic gain. It is recommended to use p -rep designs with small numbers of replicates for check varieties rather than grid-plot designs when it is desired to compare test lines to check varieties.

¹ Luiz de Queiroz College of Agriculture (ESALQ), University of São Paulo, Piracicaba, São Paulo, ralcarde@usp.br. The authors are grateful to FAPESP (#2024/17910-4 and #2016/26003-4), CAPES and The University of Adelaide.

² UniSA STEM, University of South Australia, Adelaide, South Australia.

³ School of Agriculture Food and Wine, University of Adelaide, Adelaide, South Australia.

⁴ Luiz de Queiroz College of Agriculture (ESALQ), University of São Paulo, Piracicaba, São Paulo

⁵ Center of Exact and Technological Sciences, Western Paraná State University (UNIOESTE), Cascavel, Paraná.



Aprimoramento do Mapeamento de Herdabilidades Regionais para Aumento do Poder de Detecção Genômica

OLIVEIRA, B. V.¹; AZEVEDO, C. F.²;

ROQUE, R. D.¹; NASCIMENTO, M.²; NASCIMENTO, A. C. C.²

Resumo

Na busca por compreender a base genética da variação fenotípica, a Associação Genômica Ampla visa identificar associações entre marcadores moleculares e os fenótipos. Como alternativa aos métodos estatísticos que analisam um marcador por vez, o Mapeamento de Herdabilidades Regionais (MHR) propõe modelar segmentos genômicos, permitindo detectar associações mais complexas com os fenótipos. O modelo MHR inclui efeitos poligênicos por meio da inclusão da matriz de parentesco genômico, além do efeito da região genômica. Os marcadores que compõem a região são excluídos do cálculo do efeito poligênico para evitar a dupla contagem. Estudos, baseados em um único marcador, mostram que incluir marcadores do mesmo cromossomo no efeito poligênico pode reduzir o poder de detecção. Para contornar esse problema, utiliza-se a abordagem *Leave-One-Chromosome-Out* (LOCO), que exclui o cromossomo do marcador testado ao construir a matriz de parentesco. Neste contexto, avaliou-se a integração RHM+LOCO como estratégia para aumentar o poder de detecção. Foram simulados 1.000 indivíduos, genotipados para 6.000 marcadores e fenotipados para três características com baixa, moderada e alta herdabilidade. A simulação foi repetida 30 vezes, e os resultados finais foram as médias e desvios-padrão. As regiões genômicas foram definidas considerando desequilíbrio de ligação de aproximadamente 0,20. MHR e MHR+LOCO foram comparados quanto ao poder de detecção, taxa de falso positivos (TFP) e área sob a curva ROC (AUC). A abordagem RHM + LOCO apresentou desempenho superior quanto ao poder de detecção e à AUC, indicando maior sensibilidade na identificação de regiões associadas, embora com um aumento na TFP.

Palavras-chave: *Leave-One-Chromosome-Out*; Simulação; Desequilíbrio de ligação; Características quantitativas; associação genômica.

¹ Discente do Programa de Pós Graduação em Estatística Aplicada e Biometria, Viçosa – brenda.v.oliveira@ufv.br

² Professor(a) do Programa de Pós Graduação em Estatística Aplicada e Biometria, Viçosa – camila.azevedo@ufv.br



Ordem de adição em experimentos de mistura aplicados na composição de silagens para ruminantes

Melissa Wendy Ribeiro Vicente;¹ Marcelo Ribeiro Malta² Marcelo Ângelo Cirillo³

Resumo

Em sistemas intensivos de criação de ruminantes, a alimentação representa uma das principais parcelas dos custos de produção. Uma alternativa viável para reduzir esses custos é o aproveitamento de resíduos agrícolas, como subprodutos da agroindústria que muitas vezes são descartados ou subutilizados. Entre esses, os resíduos do café apresentam grande potencial para uso na alimentação de ruminantes. Para avaliar formulações alimentares com esses subprodutos, modelos estatísticos de experimentos de mistura são ferramentas valiosas para identificar a proporção ideal entre os componentes. Além disso, esses modelos podem incorporar o conceito de ordem de adição, que se refere à sequência em que os ingredientes são combinados durante a formulação. Essa ordem pode influenciar propriedades físico-químicas e nutricionais da mistura, afetando diretamente a qualidade final do produto. Este trabalho tem como objetivo avaliar a viabilidade de um modelo de mistura com a ordem de adição incorporada, utilizando dados da EPAMIG (Empresa de Pesquisa Agropecuária de Minas Gerais). A instituição avaliou a composição bromatológica de silagens produzidas a partir de subprodutos do processamento do café, formado pela casca de café seca (CCS), a casca de café úmida (CCU), e o melaço (MEL). As variáveis-resposta analisadas foram: matéria seca (MS), proteína bruta (PB), e extrato etéreo (EE). Os resultados indicaram bom ajuste do modelo de mistura com ordem de adição, resultando em valores de R^2 ajustado variando entre 0,9384 e 0,9943. A partir do modelo ajustado, foi possível determinar proporções ótimas da mistura para maximizar cada uma das variáveis de interesse.

Palavras-chave: mistura 1; ordem de adição 2; silagem 3; ruminantes 4.

¹Departamento de Estatística, UFLA - melissa.vicente2@estudante.ufla.br

²Empresa de Pesquisa Agropecuária de Minas Gerais, EPAMIG - marcelomalta@epamig.br

³Departamento de estatística, UFLA - macufla@gmail.com



Blocking Factorial Experiment with Unequal Number of Replications for Factors Levels

Julio Silvio de Sousa Bueno Filho;¹ Adisa Jamiu Saka²

Resumo

The blocking factorial experiments are a complicated endeavour with many factors of multiple levels but with blocks of relatively small sizes, as happens to be the case in agronomical experiments. This paper examines a situation in which complete (orthogonal) block designs are impossible for the factorial design and also there are no evident of combinatory to aid the design finding of such. One of the interesting problems we deal with is unequal number of replicates of factor levels; this is mainly due to convenience of practical experimentation. The example presented here deals with an agronomy experiment of garlic vernalization with design structure of $2 \times 2 \times 3 \times 4$. An optimal (or near optimal) design is obtained for the problem using an interchange strategy for the irregular factorial structure. The resulting design has nice combinatorial properties that could not be envisaged in advance. The methods used in this paper are not new to design theory but are considerably underused in agronomical trials. The design itself is unread of in horticultural practice.

Palavras-chave: Factorial Experiment 1; Replication 2; Design Structure 3; Optimal 4; Combinatorial 5.

¹Department of Statistics, Federal University of Lavras, Brazil - jssbueno@ufla.br

²Department of Statistics, Federal University of Lavras, Brazil - adisa.saka@ufla.br



From Models to Fields: Simulating Beet Productivity Using AI and Statistical Tools

Julio Cesar Souza Vasconcelos¹ Denize Palmito dos Santos²

Abstract

Field experiments play a highly important role in scientific studies by testing hypotheses under real-world conditions. Despite their relevance, these experiments often suffer from limited sample sizes, which can impact the reliability and depth of the analysis. One alternative to address this issue is the use of synthetic data—artificially generated data that maintains the same mathematical and statistical properties as real data. This study compares synthetic data generation using Gaussian Copulas and Generative Adversarial Networks (GANs). The original dataset consists of measurements of sugar beet root dry matter under five nitrogen fertilization levels. The experimental design followed a randomized block design with three blocks and five plots per block. The main objective is to augment the dataset to enhance the application of various statistical and machine learning models. The synthetic datasets generated by each method was evaluated by comparing them to real data in terms of visual similarity and model performance. The results showed that Gaussian Copulas outperformed GANs. This finding highlight the usefulness of synthetic data in enriching agronomic research, particularly in scenarios with limited availability of field data.

Keywords: Gaussian Copulas; Generative Artificial Intelligence; Generative Adversarial Networks (GANs); Synthetic data.

¹UNIFESP – Instituto de Ciéncia e Tecnologia – ICT, SJC–juliocezarvasconcelos@hotmail.com

²INMA UFMS - Instituto de Matemática – denizepalmito@gmail.com



**Classificação de produtividade em soja utilizando imagens RGB de VANTs e
Redes Neurais**

Eleniz Aparecida Dias ¹, Luiz Alexandre Peternelly ²

Resumo

Com a crescente demanda por métodos eficientes de seleção de cultivares, drones com sensores multiespectrais têm sido amplamente utilizados no melhoramento genético. No entanto, o alto custo desses sensores limita sua aplicação. Nesse contexto, sensores RGB emergem como uma alternativa mais acessível. Este estudo comparou a eficácia de imagens RGB e multiespectrais na classificação da produtividade da soja. Foram avaliadas 99 populações e 21 genitores, totalizando 120 parcelas arranjadas em delineamento em blocos casualizados com três repetições, resultando em 360 parcelas. Ao longo do ciclo da cultura, foram conduzidos 11 voos com drone a 40 m de altura, obtendo-se imagens para cálculo dos índices de vegetação e valores das bandas red, blue, green e NIR. Os dados foram analisados por uma Rede Neural Feed Forward (FFNN) no R. Dois modelos foram desenvolvidos: um baseado exclusivamente em índices RGB (SIPI, NGRDI, PRI, RI) e nas bandas blue, green e red; e outro utilizando índices multiespectrais (DVI, EVIS, MSR, MTVI, RDVI, SAVI, SIPI) combinados com a banda NIR. Ambos classificaram as parcelas em dois grupos de produtividade: grupo 1 (acima ou igual à média) e grupo 2 (abaixo da média). Os resultados revelaram acurárias comparáveis entre as abordagens (0.643 para RGB e 0.648 para multiespectral), demonstrando que os sensores RGB apresentam potencial para substituir os sensores multiespectrais, com a vantagem adicional do menor custo.

Palavras-chave: Sensores RGB; Índices de Vegetação; Glycine max; Seleção de plantas.

¹ Departamento de Estatística, Universidade Federal de Viçosa – Eleniz.dias@ufv.br

² Departamento de Estatística, Universidade Federal de Viçosa – peternelly@ufv.br



Curva de lactação de ovelhas da raça Santa Inês

Louziane Ribeiro Carvalho¹

Maria Vitória Neves²

Valdeline de Paula Mequelinha Ferreira³

Joel Augusto Muniz⁴

Resumo

O leite de ovelha possui maior concentração de proteína e gordura quando comparado ao leite de vaca e de cabra. Os continentes Asiático e Europeu se destacam na produção de leite ovino e na fabricação de queijos com sabores característicos e alto valor commercial. A ovinocultura leiteira é recente no Brasil e a principal base genética do rebanho nacional é a raça Lacaune de origem francesa. Os ovinos Santa Inês, presentes em várias regiões do Brasil, são explorados principalmente para produção de carne, sendo limitado o conhecimento sobre o potencial leiteiro da raça. O objetivo deste trabalho foi comparar diferentes modelos não lineares para descrever a curva de lactação de ovelhas da raça Santa Inês. Foram utilizados: o modelo Cobb e Le Du, o modelo Dhanoa e o modelo Papajcsik. Para estimar os parâmetros dos modelos, foi utilizado o método de mínimos quadrados, usando o algoritmo de convergência de Gauss-Newton, no software R. A análise de resíduos incluiu testes de normalidade (Shapiro-Wilk), homocedasticidade (Breusch-Pagan) e independência (Durbin-Watson). A qualidade de ajuste dos modelos foi avaliada por meio do coeficiente de determinação ajustado (R_{aj}^2) e do critério de informação de Akaike (AIC). Com base nos critérios de avaliação e qualidade de ajuste, e também pela vantagem de estimar o tempo para atingir o pico de produção, o modelo Dhanoa foi considerado o melhor modelo para descrever a curva de lactação de ovelhas Santa Inês. O pico de lactação das ovelhas Santa Inês ocorreu entre a quarta e quinta semana de lactação.

Palavras-chave: leite de ovelha; modelos não lineares; ovinocultura leiteira, regressão.

¹ Programa de Pós-Graduação em Estatística, UFLA – louziane.carvalho2@estudante.ufla.br

² Programa de Pós-Graduação em Estatística, UFLA – maria.neves5@estudante.ufla.br

³ Programa de Pós-Graduação em Estatística, UFLA – valdeline.ferreira1@estudante.ufla.br

⁴ Departamento de Estatística, UFLA – joamuniz@ufla.br



Protocolo para contagem automatizada de cápsulas de linhaça utilizando o pacote R pliman

Lalesca Dadam Gomes^{1*}

Carlos Eduardo Forcelini Assoni¹

Kenji da Cruz Konno¹

Luiz Paulo da Silva¹

Matheus Lopes Machado¹

Leonardo Talavera Campos²

Tiago Olivoto³

Resumo

O linho (*Linum usitatissimum* L.) destaca-se por sua importância econômica e nutricional, sendo a linhaça a principal fonte vegetal de ômega-3. A contagem de cápsulas por planta é um componente essencial para o melhoramento genético da cultura, porém, a avaliação manual é trabalhosa e suscetível a erros. Este trabalho avaliou o desempenho do pacote pliman, desenvolvido no R, na contagem automatizada de cápsulas de linhaça em diferentes combinações de resoluções de imagem (100%, 50% e 25%) e uso de filtros morfológicos (opening e filtro de mediana), resultando em 12 cenários testados. A acurácia foi estimada pelo coeficiente de concordância de Lin (ρ_c) e o erro pela raiz do erro quadrático médio (RMSE). O melhor desempenho foi observado com resolução de 50% (1224 x 1224 pixels), e uso dos argumentos opening = 15 e filter = 3 na função analyze_objects() do pliman, resultando em um $\rho_c = 0,997$ e RMSE = 4,51. Nesse cenário, o tempo médio de processamento foi de ~1 s por imagem (33 vezes mais rápido comparado à resolução de 100% com o uso dos mesmos argumentos). Os piores resultados ocorreram sem aplicação de filtros, onde todos os valores foram superestimados devido a presença de ruídos. Imagens com altas resoluções não resultam em maior acurácia e tem a desvantagem de aumentar o tempo de processamento. Recomenda-se o uso de uma resolução moderada (1200 x 1200 pixels) com opening = 15 e filter = 3 para obtenção de alta acurácia na contagem automatizada de cápsulas de linhaça utilizando imagens digitais.

Palavras-chave: *Linum usitatissimum* L.; análise de imagem; filtros morfológicos; eficiência computacional.

¹ Discente do curso de Agronomia, Universidade Federal de Santa Catarina.

² Pós-doutorando, PPG/RGV, Universidade Federal de Santa Catarina.

³ Docente do Departamento de Fitotecnia, Universidade Federal de Santa Catarina.

*Autor correspondente: glalescadadam@gmail.com



Avanços no melhoramento do linho no Sul do Brasil: Resultados do Programa de Melhoramento NEPEM/UFSC

Kenji da Cruz Konno ^{1*}
Carlos Eduardo Forcelini Assoni¹
Matheus Lopes Machado¹
Lalesca Dadam Gomes¹
Luiz Paulo da Silva¹
Leonardo Talavera Campos²
Tiago Olivoto³

Resumo

Este estudo avaliou o progresso genético no rendimento do *Linum usitatissimum L.* do Programa de Melhoramento do NEPEM/UFSC, durante três anos (2022-2024), visando desenvolver cultivares com alto rendimento de grãos por planta (RGPLA). Em 2022, duas populações (Dourada e Marrom) foram semeadas para avaliar caracteres agronômicos. Nesse experimento, uma avaliação multivariada realizada em 229 plantas permitiu selecionar 101 genótipos superiores (ex., altura, maior RGPLA). Em 2023, esses genótipos foram reavaliados em linhas de 2 metros de comprimento, analisando-se os mesmos caracteres e utilizando o índice de seleção MGIDI, resultando na seleção de 32 genótipos elite (13,97% dos genótipos iniciais). Em 2024, o mesmo protocolo foi aplicado aos 32 genótipos. Os dados dos três anos foram analisados no Software-R. O diferencial de seleção observado nos dois ciclos de seleção foi calculado como $D(\%) = (Y_s - Y_o) / Y_o \times 100$. Um teste *t*-Student ($p < 0,05$) rejeitou a hipótese nula de igualdade das médias, confirmando aumento no RGPLA (0,698g - 2022 para 3,28g - 2024). Por outro lado, observou-se uma redução do coeficiente de variação (106% - 2022 para 32,5% - 2024), sugerindo uma redução da variabilidade genética ao longo do estudo. Apesar do sucesso no RGPLA, a redução na variabilidade genética põe em risco a sustentabilidade do programa, reduzindo a chance de observar ganhos futuros significativos. Pretende-se então iniciar cruzamentos entre genótipos elite e materiais comerciais para reintroduzir variabilidade sem comprometer ganhos. Este trabalho demonstra os avanços do melhoramento do linho no sul Brasil, contribuindo para o desenvolvimento de cultivares mais produtivas e resilientes.

Palavras-chave: Linho 1; Alto Rendimento 2; Progresso 3; Melhoramento Genético 4; Adaptabilidade 5

¹ Discente do curso de Agronomia, Universidade Federal de Santa Catarina.

² Pós-doutorando, PPG/RGV, Universidade Federal de Santa Catarina.

³ Docente do Departamento de Fitotecnia, Universidade Federal de Santa Catarina.

*Autor correspondente: kenji.c.konno@gmail.com



Modeling *Chrysodeixis includens* behavior using generalized linear mixed-effects fractional polynomial models

Allison Queiroz de Oliveira ¹ César Gonçalves de Lima ² Lucas Silva Barros ³ Pedro Takao Yamamoto ⁴ Kevin Burke ⁵

Abstract

One of the main features in data sets from the counting process is the great quantity of zeros that exceeds the modeling ability of classical distributions as such Poisson and negative binomial, setting scenarios of zero-inflated data. To overcome this problem extensions of these distributions are cited in the literature, well-known as zero-inflated models as: zero-inflated Poisson (ZIP), zero-inflated negative binomial (ZINB) as well as the hurdle models. The situation gets more complex when the longitudinal effect is present in the observation vector and, despite the linear analysis approach being very employed, for the counting process the response behavior has in general a nonlinear trend. A way to handle that nonlinearity is the use of fractional polynomials that presents greater flexibility in the modeling than the usual nonlinear techniques. In view of that, considering an experiment which the effect of five treatments was assessed in relation to their control over *Chrysodeixis includens* in soybean crops along thirteen weeks under a randomized completely block design, a structure of selection and fitting of generalized linear mixed-effects fractional polynomial models has been developed in R programming language, allowing the selection of most adequate powers to the average response behavior and zero probability. Besides, a simulation study was carried out to validate the quality of the selection structure as well as the best selected model.

Keywords: experimental design; soybean; hierarchical models; longitudinal data; zero inflation.

¹Graduate Program in Statistics and Agricultural Experimentation, “Luiz de Queiroz” College of Agriculture, University of São Paulo Piracicaba/SP - allisonqueiroz@usp.br

²Department of Basic Sciences, Faculty of Animal Science and Food Engineering, University of São Paulo, Pirassununga/SP - cegdlima@usp.br

³Graduate Program in Entomology, “Luiz de Queiroz” College of Agriculture, University of São Paulo, Piracicaba/SP - lucasbarros.lsb@gmail.com

⁴Department of Entomology and Acarology, “Luiz de Queiroz” College of Agriculture, University of São Paulo, Piracicaba/SP - pedro.yamamoto@usp.br

⁵Department of Mathematics and Statistics, University of Limerick, Limerick/Ireland - Kevin.Burke@ul.ie



Insights para a delimitação de blocos experimentais com base em atributos do solo em diferentes profundidades

Samantha Gouvêa Oliveira ¹, Matheus de Paula Ferreira ², Luiz Alexandre Peternelly ³

Resumo

Sabe-se que o desenvolvimento de culturas agrícolas é influenciado não somente pelas condições climáticas, mas também pela disponibilidade de nutrientes no solo. Dessa forma, o manejo adequado do solo é essencial para garantir produtividade e sustentabilidade no sistema agrícola. Este trabalho tem como objetivo estudar a distribuição de atributos gerais solo (pH, fósforo, potássio, cálcio, magnésio, alumínio, H+Al e fósforo remanescente) e atributos específicos (zinc, ferro, cobre e manganês) em uma área com diferentes clones de cana-de-açúcar, para a determinação de blocos experimentais. Foram coletadas 60 amostras de solo nas profundidades de 0 – 20 e 20 – 40 cm. Em cada profundidade, cada atributo foi mapeado pelo inverso ponderado da distância (IDW). Posteriormente, aplicou-se a técnica de clusterização (K-means) considerando as concentrações de atributos gerais e específicos, com a formação de dois e três grupos para ambas as profundidades. O índice Kappa de Cohen (κ) foi utilizado para avaliar a concordância entre os agrupamentos. Observou-se que, para os atributos gerais, os mapas apresentaram concordância substancial ($\kappa_2 = 0,6977$) para a formação de dois grupos e concordância moderada ($\kappa_3 = 0,4422$) para três. Este resultado indica que a profundidade de coleta pode não impactar a distribuição espacial destas características. Já para os específicos, a profundidade parece impactar a distribuição, visto que os mapas apresentaram concordância leve ($\kappa_2 = 0,0219$ e $\kappa_3 = 0,0524$) para dois e três grupos, respectivamente. Considerando cenários de manejo sob restrição de recursos, a clusterização com base em atributos específicos pode auxiliar na tomada de decisão quanto à formação de blocos experimentais.

Palavras-chave: clusterização; K-means; IDW, Kappa de Cohen.

¹ Departamento de Estatística, Universidade Federal de Viçosa – samantha.gouvea@ufv.br

² Departamento de Estatística, Universidade Federal de Viçosa – peternelly@ufv.br

³ Departamento de Estatística, Universidade Federal de Viçosa – matheus.paula@ufv.br



Estudo Temporal da Herdabilidade e Correlação de Índices de Vegetação com Características de Interesse Agronômico em Cana-de-Açúcar

Aloísio Fernando Silva Ribeiro¹ Matheus de Paula Ferreira² Luiz Alexandre Peternelly³

Resumo

O uso de imagens aéreas tem grande relevância no melhoramento genético vegetal, sobretudo na seleção indireta, ampliando e acelerando a coleta de dados fenotípicos. No entanto, a eficácia dessa abordagem depende da herdabilidade dos índices extraídos das imagens e de sua correlação com características primárias de interesse. Este estudo avaliou a herdabilidade e a correlação dos índices de vegetação NDVI, EVI, SAVI, NDMI, PRI, NDRE, GNDVI, OSAVI, MSAVI, PSRI e ARI com número de colmos (NC), toneladas de cana por hectare (TCH) e Brix em famílias de cana-de-açúcar ao longo de três datas entre outubro de 2023 e fevereiro de 2024. O experimento foi conduzido em delineamento em blocos aumentados, com oito blocos, duas testemunhas e 416 tratamentos. As herdabilidades foram estimadas por modelos mistos e as correlações pelo coeficiente de Pearson. As maiores correlações ocorreram entre índices e NC, variando de 0,61 a 0,67 na primeira data. TCH e Brix apresentaram correlações inferiores a 0,2 em todas as avaliações. Em geral, observou-se queda nas correlações absolutas de NC e Brix ao longo do tempo, enquanto TCH apresentou pequeno aumento. Os índices PRI e ARI mostraram maior variação temporal nas correlações, especialmente o ARI, que teve redução de cerca de 0,60 entre a primeira e as demais datas. Todos os índices apresentaram herdabilidades elevadas (0,50 a 0,70), com pouca variação ao longo do tempo. Conclui-se que a associação entre índices de vegetação e características agronômicas é variável temporalmente, indicando a necessidade de estudos adicionais para otimizar sua aplicação preditiva.

Palavras-chave: Melhoramento genético; Análise de imagens; Fenotipagem de auto rendimento; Seleção indireta; Correlação fenotípica.

¹ Departamento de estatística, UFV – aloisio.fernando@ufv.br

² Departamento de estatística, UFV – matheus.paula@gmail.com

³ Departamento de estatística, UFV – peternelly@ufv.br



Dimensionamento de parcela ótima na cultura de arroz considerando dependência espacial usando Geoestatística

Edmundo do Rosário Rodrigues Caetano ¹

Renata Alcarde Sermarini ²

Resumo

A determinação do tamanho ótimo de parcelas experimentais é fundamental no planejamento de experimentos agrícolas, para melhorar a precisão, bem como maximizar a obtenção de informações e minimizar custos operacionais do experimento. Existem poucos estudos realizados para determinação do tamanho e forma ótimos da parcela considerando a existência de dependência espacial e utilizando métodos geoestatísticos. Com o objetivo de avaliar metodologias que consideram tal característica em um experimento com a cultura de arroz, foram gerados dados a partir de um ensaio de uniformidade com a variedade IR8, em uma área plantada de 760 m², sendo 1m² a área da unidade básica. Foram efetuados agrupamentos de parcelas com 16 tamanhos e formas diferentes, nomeadamente: 1x1, 1x2, 1x3, 1x4, 1x5, 1x6, 2x2, 2x3, 2x4, 2x5, 2x6, 3x4, 3x5, 3x6, 4x5, 4x6, calculadas as variâncias e os coeficientes de variação, para uso do método da máxima curvatura modificado (MMCD) e, adicionalmente, foram utilizados modelos geoestatísticos ajustados aos dados experimentais considerando o semivariograma e uma função de curvatura esférica para encontrar o tamanho ótimo de parcela, através da análise do raio de curvatura da função. Os resultados preliminares encontrados mostraram que utilizando o método MMCD o tamanho ótimo de parcela é de 6 m², e considerando a dependência espacial usando um semivariograma e uma função de curvatura esféricos foram identificados três potenciais candidatos ao tamanho de parcela ótimos nomeadamente 4 m², 5 m² e 6 m², tendo se destacado a parcela de 6 m² que apresentou a curvatura máxima ajustada de 0,0559.

Palavras-chave: *Parcela Ótima; Dependência Espacial; Curvatura; Semivariograma.*

¹ Estudante do Doutorado do Departamento de Ciências Exatas, USP-ESALQ, Piracicaba, SP – eddycaetano@usp.br

² Professora do Departamento de Ciências Exatas, USP-ESALQ, Piracicaba, SP– ralcarde@usp.br



Fenotipagem digital da relação entre atributos espectrais e compostos bioativos em sementes de linhaça dourada e marrom

Isabella Ramos Pedroni Paulo^{1*}

Luiz Paulo da Silva¹

Kenji da Cruz Konno¹

Carlos Eduardo Forcelini Assoni¹

Tiago Olivoto²

Resumo

A linhaça (*Linum usitatissimum* L.) destaca-se como alimento funcional, com potencial na prevenção de doenças crônicas devido à presença de compostos bioativos como fenóis, flavonoides e antioxidantes. As variedades dourada e marrom diferem em aspectos da composição química, mas há poucos estudos que comparam seus atributos funcionais de forma integrada. Este trabalho investigou a relação entre a coloração das sementes e a presença de compostos bioativos por meio da fenotipagem digital e análise de correlação. Foram analisadas imagens digitais de diferentes genótipos das variedades dourada e marrom, das quais se extraíram variáveis de cor nos espaços RGB e HSB, além de índices derivados (HUE, SAT, BI, SAVI, VARI e IPCA). As correlações de Pearson revelaram padrões distintos entre as variedades. Na linhaça dourada, observaram-se menores teores de flavonoides, com correlações negativas para SAVI ($r = -0,49$) e HUE ($r = -0,42$). Fenóis totais correlacionaram-se moderadamente com o potencial antioxidante ($r = 0,53$), evidenciando seu papel funcional. Na linhaça marrom, tonalidades mais escuras (menor HUE) foram associadas a maiores concentrações de flavonoides ($r = -0,05$), e os fenóis totais apresentaram correlações com flavonoides ($r = 0,61$) e com o potencial antioxidante ($r = 0,45$), reforçando sua importância na atividade antioxidante. Os resultados indicam que a fenotipagem digital é uma ferramenta promissora e não destrutiva na triagem de genótipos com maior valor funcional, destacando o potencial da coloração como marcador visual indireto da composição bioativa em sementes de linhaça.

Palavras-chave: *Linum usitatissimum*; Índices de cor; Flavonoides; Fenotipagem digital.

¹Discente do curso de Agronomia, Departamento de Fitotecnia, Universidade Federal de Santa Catarina.

²Docente do Departamento de Fitotecnia, Universidade Federal de Santa Catarina.

*Autor correspondente: isabellapedroni2409@gmail.com



Análise de componentes principais aplicada na caracterização agronômica de pimentas do gênero *Capsicum* para fins ornamentais

Beatriz Bandini de Carvalho¹, Beatriz Romanini Morales¹,
Christiane de Fátima Martins França², Fernando César Sala³,
Gilberto Rodrigues Liska², Josiane Rodrigues²

Resumo

As pimentas são parte da riqueza cultural brasileira e um valioso patrimônio da biodiversidade do Brasil. Elas são consumidas de diversas formas, seja *in natura*, processadas em geleias e compotas, ou de forma medicinal como fonte de antioxidantes. Adicionalmente, existe ainda um mercado promissor de pimentas ornamentais, recomendadas tanto para cultivo em jardins como para espaços internos. No Brasil, dentre as espécies mais cultivadas estão a *C. annuum* e a *C. chinense*, sendo que, no que diz respeito ao comércio de pimentas ornamentais, há apenas uma pequena quantidade de variedades presentes no mercado, o que abre caminho para o melhoramento genético das variedades que já estão disponíveis. Assim, o presente trabalho tem por objetivo auxiliar os programas de melhoramento genético, caracterizando e comparando as pimentas biquinho vermelha e bode amarela, ambas da espécie *C. chinense*, e as linhagens F810070, F810082 e F810047, da espécie *C. annuum* pertencentes ao Banco de Germoplasma da Universidade Federal de São Carlos, com relação a variáveis agronômicas consideradas favoráveis para seu uso em ornamentação, utilizando-se da técnica multivariada da análise de componentes principais, cujo intuito é a obtenção de um pequeno número de combinações lineares das variáveis originais que retenham o máximo possível da informação contida no conjunto de dados. Na análise, apenas os três primeiros componentes principais foram considerados, os quais captaram cerca de 85% da variabilidade total do conjunto. Em suma, pode-se concluir que a pimenta de maior destaque foi a F810070, sobressaindo-se na maioria das variáveis agronômicas estudadas.

Palavras-chave: Análise multivariada; Melhoramento genético; Estatística aplicada em agronomia; *Capsicum chinense*; *Capsicum annuum*.

¹ Estudante de Graduação em Engenharia Agronômica, Centro de Ciências Agrárias (CCA), Universidade Federal de São Carlos (UFSCar), Araras/SP, Brasil - bbscarvalho@estudante.ufscar.br, beatrizromanini@estudante.ufscar.br

² Docente do Departamento de Tecnologia Agroindustrial e Socioeconomia Rural (DTAiSeR), Centro de Ciências Agrárias (CCA), Universidade Federal de São Carlos (UFSCar), Araras/SP, Brasil - christiane@ufscar.br, gilbertoliska@ufscar.br, josirodrigues@ufscar.br

³ Docente do Departamento de Biotecnologia e Produção Vegetal e Animal (DBPVA), Centro de Ciências Agrárias (CCA), Universidade Federal de São Carlos (UFSCar), Araras/SP, Brasil - fcsala@ufscar.br



Eficácia de derivados salicil-N-hidrazônicos na inibição de *Fusarium sp.* e *Corynespora cassiicola*

Maria Júlia R. de Grandi¹, Bárbara T. Shiga², Vitória M. Shiga³, Juliana F. S. Daniel⁴

Elizabeth M. Hashimoto⁵, Fábio Vandresen⁶, Marcelo C. Filho⁷

Resumo

As doenças fúngicas causadas por fitopatógenos, como *Fusarium sp.* e *Corynespora cassiicola*, representam um desafio significativo para o controle agrícola. A resistência desses microrganismos a condições adversas compromete a eficácia dos antifúngicos convencionais, impactando negativamente a saúde humana, animal e ambiental. Diante disso, este estudo avaliou a eficácia de derivados salicil-N-hidrazônicos na inibição do crescimento micelial desses fungos em meio BDA (batata, dextrose e ágar). Uma análise de regressão, com apoio computacional do software Jamovi versão 2.6, foi conduzida para avaliar a relação entre a concentração dos compostos (50 mg/L e 100 mg/L) e a espécie fúngica (*Fusarium sp.* e *Corynespora cassiicola*) sobre o crescimento dos microrganismos. Os resultados indicaram que os pressupostos de normalidade dos resíduos (Shapiro-Wilk, p -valor $> 0,05$), bem como homogeneidade de variâncias (Breusch-Pagan, p -valor $> 0,05$) foram atendidos. Por meio da análise, observou-se um impacto significativo da concentração dos compostos na inibição do crescimento fúngico (p -valor $< 0,05$), indicando que concentrações mais altas resultaram em maior eficácia antifúngica. Por outro lado, não foi observada diferença significativa na resposta das espécies testadas (p -valor $> 0,05$). Conclui-se que os compostos derivados salicil-N-hidrazônicos possuem atividade antifúngica dependente da concentração, sendo promissores para o controle de fitopatógenos.

Palavras-chave: compostos químicos; fitopatógenos fúngicos; crescimento micelial; análise de regressão.

¹ Programa de Pós-graduação em Engenharia Ambiental - UTFPR, Londrina - mariajuliagrandi@alunos.utfpr.edu.br

² Graduação em Engenharia Ambiental e Sanitária - UTFPR, Londrina - bshiga@alunos.utfpr.edu.br

³ Graduação em Engenharia Ambiental e Sanitária - UTFPR, Londrina - vitoriaashiga@alunos.utfpr.edu.br

⁴ Programa de Pós-graduação em Engenharia Ambiental - UTFPR, Londrina - julianasouza@professores.utfpr.edu.br

⁵ Departamento Acadêmico de Matemática - UTFPR, Londrina - ehashimoto@professores.utfpr.edu.br

⁶ Departamento Acadêmico de Química - UTFPR, Londrina - fabiovandresen@professores.utfpr.edu.br

⁷ Graduação em Engenharia Química - UTFPR, Londrina - marcelocosta.filho@alunos.utfpr.edu.br



Ambientômica aplicada à caracterização climática em ensaios de melhoramento de plantas

Maria Laura Faustino Monteiro^{1*} Isabela Martins Ghizoni¹ Kamily Siqueira Tonet¹
Kenji da Cruz Konno¹ Carlos Eduardo Forcelini Assoni¹ Tiago Olivoto²

Resumo

O estudo avaliou a precisão dos dados climáticos da plataforma NASAPOWER para temperatura máxima (Tmax), temperatura mínima (Tmin) e umidade relativa (UR%) na região litorânea de Santa Catarina, comparando-os com registros *in situ* da Fazenda Experimental da Ressacada (UFSC) em 2022 e 2023, e caracterizou ambientes de cultivo para a cultura da linhaça (*Linum usitatissimum* L.). Utilizando o coeficiente de concordância de Lin (c) e análises de distribuição de frequência, os resultados indicaram alta acurácia do NASAPOWER para Tmax ($c=0,85-0,87$) e Tmin ($c=0,79$), embora a UR% tenha apresentado superestimação ($c=0,42$), limitando sua confiabilidade para análises quantitativas. A tipagem ambiental revelou contrastes climáticos entre Florianópolis (litoral) e Campos Novos (interior), com maior amplitude térmica e noites frias no interior, além de variações interanuais que impactam fases críticas da cultura, como floração e estádio reprodutivo. Conclui-se que o NASAPOWER é uma ferramenta válida para obtenção de dados térmicos em regiões sem estações meteorológicas locais, mas sua aplicação para UR% exige validação complementar. A integração de ensaios multiambientais e tipagem ambiental é essencial para identificar genótipos adaptados a condições específicas ou estáveis frente à variabilidade climática.

Palavras-chave: Interação genótipo × ambiente (G×E); NASAPOWER; Melhoramento genético; Variáveis climáticas; Estabilidade fenotípica.

¹ Discente do curso de Agronomia, Universidade Federal de Santa Catarina.

² Docente do Departamento de Fitotecnia, Universidade Federal de Santa Catarina.

* Autor correspondente: marialaurafrm.31@gmail.com



Análise da mortalidade de *Diaphorina citri* induzida por *Cordyceps fumosorosea*: um estudo de funções de ligação para modelos de dose-resposta dicotômica

Carla Valéria dos Santos Pacheco;¹ Robson Marcelo Rossi;¹
Diego Corrêa Alves;¹ Rodolfo de Souza Nascimento.¹

Resumo

Este estudo investiga a relação entre a mortalidade de *Diaphorina citri*, uma praga significativa da citricultura, e diferentes concentrações do fungo entomopatogênico *Cordyceps fumosorosea*, utilizado no controle biológico. A modelagem foi por meio de regressão binomial, via abordagem bayesiana. Foram consideradas funções de ligação, incluindo as funções usuais (*logit*, *probit*, *cauchit*, *cloglog* e *loglog*), as de potência e as reversas de potência. O modelo Cauchy apresentou o melhor ajuste aos dados, com o menor valor de DIC (Critério de Informação de *Deviance*). A análise de resíduos quantílicos normalizados, por meio de envelope simulado, garantiu uma avaliação robusta da qualidade do ajuste. Os parâmetros do modelo foram estatisticamente significativos para o preditor linear, com as seguintes estimativas e Intervalos de Credibilidade ICr (95%): O intercepto (β_0) foi $-11,95$, $[-19,19; -7,16]$; o coeficiente angular (β_1) foi $2,37$, $[1,44; 3,81]$ e a dose letal de 50% (DL_{50}) foi $5,03$ ($\log(\text{conídios}/ml)$), $[4,84; 5,21]$, ou seja, ao aplicar essa dose, espera-se que metade dos insetos expostos morra. A ampliação das funções avaliadas contribui para um melhor entendimento da modelagem de bioensaios binários e reforça a importância da escolha da função de ligação especialmente em relação ao meio ambiente e aos custos relativos aos níveis de dose do produto.

Palavras-chave: *greening*; *HLB*; *huanglongbing*; *psilídeo asiático*; *regressão binomial*.

¹Programa de Pós-graduação em Bioestatística, Universidade Estadual de Maringá - carla.pacheco@ifsc.edu.br



Análise de Correlação e PCA em Genótipos de Linhaça

(*Linum usitatissimum* L.) Dourada e Marrom

Luiz Paulo da Silva^{1*} Lalesca Dadam Gomes¹ Carlos Eduardo Forcelini Assoni¹
Matheus Lopes Machado¹ Leonardo Talavera Campos² Tiago Olivoto³

Resumo

O presente estudo teve como objetivo investigar as relações entre características agronômicas em 41 genótipos de linhaça (*Linum usitatissimum* L.), cultivados em Campos Novos-SC durante o ano de 2024, por meio de análises estatísticas multivariadas. Foram avaliadas nove variáveis agronômicas: altura de planta (AP), número de cápsulas (NC), número de grãos por planta (NG), massa de cápsulas (MC), massa de grãos (MG), massa de mil grãos (MMG), comprimento produtivo (CP), número de ramificações (NR) e índice de colheita de cápsulas (ICC). A análise de correlação e a análise de componentes principais (PCA) permitiram identificar padrões distintos entre genótipos com sementes douradas e marrons. Os resultados revelaram que as associações entre características variam conforme o grupo, como observado entre MMG e AP, que apresentaram correlação negativa na dourada ($r = -0,30$) e positiva na marrom ($r = 0,15$). A PCA destacou diferenças estruturais no agrupamento de genótipos e na disposição das variáveis, refletindo distintas relações fenotípicas em cada grupo. Conclui-se que há variação nas inter-relações agronômicas entre os tipos de linhaça, e que o uso de abordagens multivariadas contribui para evidenciar tais diferenças, auxiliando na identificação de genótipos que apresentam caracteres desejáveis. No geral, este estudo contribui positivamente para o avanço do melhoramento da cultura do linho no Brasil, na busca de uma agricultura mais diversificada, produtiva e sustentável.

Palavras-chave: análise de correlação; PCA; *Linum usitatissimum*; melhoramento genético.

¹ Discente do curso de Agronomia, Universidade Federal de Santa Catarina.

² Pós-doutorando, PPG/RGV, Universidade Federal de Santa Catarina.

³ Docente do Departamento de Fitotecnia, Universidade Federal de Santa Catarina.

* Autor correspondente: glalescadadam@gmail.com



Análise de componentes principais aplicada ao estudo de atributos químicos do solo de diferentes sistemas de produção agropecuária

Josiane Rodrigues¹ Daniel Cheng Nascimento¹ Alberto Carlos de Campos Bernardi²

Resumo

Os sistemas integrados de produção agropecuária, podem agregar numa mesma área atividades agrícolas, pecuárias e florestais. No Brasil, a utilização desses sistemas tem aumentado de forma a minimizar os riscos das atividades praticadas de forma isolada e diminuir as áreas de pastagens degradadas. Uma vez que os atributos do solo podem variar em função de diferentes práticas de manejo, o objetivo deste estudo é comparar cinco diferentes sistemas (Integração Lavoura-Pecuária-Floresta, Integração Pecuária-Floresta, Integração Lavoura-Pecuária, Pecuária Extensiva e Pastagem Intensiva) de uma área experimental da Embrapa Pecuária Sudeste, com relação a diferentes atributos químicos do solo, utilizando a técnica da análise de componentes principais, cujo intuito é a obtenção de um pequeno número de combinações lineares das variáveis originais, que retenham o máximo possível da informação contida nos dados. Com isso, pretende-se descobrir as melhores estratégias de manejo para a área de estudo, no tocante à composição química do solo, bem como enfatizar os benefícios dos sistemas integrados.

Palavras-chave: Agricultura Sustentável; Análise Multivariada; Sistemas ILPF.

¹ Departamento de Tecnologia Agroindustrial e Socioeconomia Rural, Centro de Ciências Agrárias, Universidade Federal de São Carlos, Araras-SP, Brasil – danielcheng@estudante.ufscar.br, josirodrigues@ufscar.br

² Embrapa Pecuária Sudeste, Fazenda Canchim, São Carlos-SP, Brasil - alberto.bernardi@embrapa.br



Adaptabilidade e estabilidade de genótipos de linhaça utilizando análise AMMI e WAAS

Isabela Martins Ghizoni^{1*}
Kamilly Siqueira Tonet¹
Maria Laura Faustino Monteiro¹
Lalesca Dadam Gomes¹
Carlos Eduardo Forcelini Assoni¹
Luiz Paulo da Silva¹
Matheus Lopes Machado¹
Kenji da Cruz Konno¹
Tiago Olivoto²

Resumo

A linhaça (*Linum usitatissimum* L.) é uma cultura promissora para a diversificação agrícola devido à sua adaptabilidade, mas a variabilidade genética e a interação genótipo x ambiente (G x E) podem comprometer a seleção de genótipos superiores. Este estudo teve como objetivo avaliar a estabilidade genotípica e o rendimento de 35 genótipos de linhaça cultivados em ambientes contrastantes de Santa Catarina entre 2022 e 2024, utilizando modelos AMMI (Additive Main Effect and Multiplicative Interaction) e WAAS (Weighted Average of Absolute Scores), para quantificar a estabilidade dos genótipos avaliados. Os resultados demonstraram padrões distintos de adaptação, com destaque para G4, G29 e G228, que apresentaram alta produtividade e estabilidade, sendo recomendados para ambientes com maior imprevisibilidade climática. Em contrapartida, genótipos como G2 e G69D apresentaram adaptações específicas no ano de 2024, mas com baixa estabilidade. A integração das análises AMMI e WAAS, mostrou-se eficiente no auxílio da seleção de genótipos superiores, com potencial para apoiar futuras decisões em melhoramento voltadas à adaptabilidade e estabilidade em ambientes distintos.

Palavras-chave: interação; ambiente; estabilidade; linhaça; WAAS.

¹ Discente do curso de Agronomia, Universidade Federal de Santa Catarina.

² Docente do Departamento de Fitotecnia, Universidade Federal de Santa Catarina.

* Autor correspondente: isabelaghizoni@gmail.com



Análise da degradabilidade ruminal *in vitro* por meio do ajuste de modelos não lineares mistos

Haward Antunny da Silva Américo;¹ Vanderly Janeiro²

Resumo

A análise estatística da produção de gás acumulado tem se mostrado uma ferramenta importante para compreender processos fermentativos e avaliar a qualidade de diferentes substratos. Neste estudo, curvas de produção de gás obtidas pela incubação *in vitro* de distintos tipos de forragem foram utilizadas para o ajuste de modelos não lineares que descrevem a dinâmica da fermentação ao longo do tempo. Foram considerados o modelo de Ørskov & McDonald (1979) e o modelo de Gompertz, com a estimativa dos parâmetros realizada no software R, por meio do pacote *nlme*. Utilizou-se a abordagem de modelos mistos não lineares, incorporando efeitos fixos e aleatórios, de forma a representar a variabilidade entre repetições experimentais. Essa estrutura permitiu ajustes mais robustos, adequados à natureza dos dados. Os modelos foram comparados quanto à sua capacidade de descrever o comportamento fermentativo das amostras, observando-se diferenças na qualidade do ajuste conforme a estrutura adotada. Os resultados obtidos contribuem para o avanço na aplicação de métodos estatísticos em estudos de digestibilidade e fermentação ruminal, oferecendo suporte na escolha de modelos que melhor representem a produção de gás em experimentos *in vitro*.

Palavras-chave: Modelos mistos não lineares; Produção de gás acumulado; Ørskov & McDonald (1979); Modelo de Gompertz; Fermentação *in vitro*.

¹Programa de Bioestatística, UEM - antunnyamerico@gmail.com

²Departamento de estatística, UEM - vjaneiro@uem.br



Comparative analysis of thermal efficiency in solar collectors with spiral polyethylene tubes

Julio Cezar Souza Vasconcelos ¹ Denize Palmito dos Santos ²

Abstract

The drying stage represents one of the highest energy consumption points in grain production and remains particularly costly for small-scale farmers. Aiming to reduce expenses and encourage the use of renewable energy sources, this study evaluated the thermal performance of solar air heaters built with spiral polyethylene tubes for use in agro-industrial activities. Three different collector configurations were tested, including versions with and without glass covers, and their efficiencies were analyzed using a partially linear regression model based on an extended normal distribution. The results revealed statistically significant differences between the tested conditions, with the glass-covered collector achieving the highest average outlet temperatures. Further analysis indicated relevant variations in mean temperature readings at distances of 5, 10, 20, and 25 meters along the tube. The configuration with a 20-meter-long tube showed the best thermal performance, suggesting an optimal balance between tube length and heat retention. The new statistical modeling approach proved effective in interpreting the data, demonstrating its potential as an analytical tool for sustainable energy studies. These findings support the technical feasibility of spiral polyethylene tube solar collectors as an accessible and efficient solution for air heating in small rural properties.

Keywords: Advanced regression modelling; Spiral solar collector; Sustainable energy; Thermal efficiency.

¹ Departamento de Matemática, São José dos Campos – juliocezarvasconcelos@hotmail.com

² INMA UFMS - Instituto de Matemática – denizepalmito@gmail.com



Produção de batata-doce de polpa alaranjada em resposta à adubação com esterco de frangos de corte em Moçambique

Elpídio Ana Alberto¹, Orlando Gonçalves Brito², Renato Ribeiro de Lima³, Nelson Hilário Mubai⁴

Resumo

A batata-doce é uma cultura de duplo propósito e de maior valor nutricional usada na alimentação e as suas raízes tuberosas são importantes na segurança alimentar da população rural Moçambicana. No país, o seu rendimento é baixo e o manejo da fertilidade do solo é um dos fatores limitantes. O objetivo desta pesquisa foi avaliar a produção de cultivares de batata-doce de polpa alaranjada em resposta a adubação com esterco de frangos de corte. O experimento foi conduzido no município de maxixe, província de Inhambane, Moçambique. Os tratamentos foram implantados em esquema fatorial 2 x 4, constituídos por duas cultivares de batata-doce (Glória e Sumaia) e quatro doses de adubação com esterco de frangos de corte (0, 10, 20 e 30 ton ha⁻¹). O delineamento adotado foi em blocos casualizados com três repetições. Foram avaliados a Produtividade de raízes, o Número de raízes totais, o Número de raízes comerciais, Peso médio de raiz e a Percentagem do rendimento comercial. Em relação à produtividade de raízes, apenas a cultivar Glória apresentou resposta ao aumento das doses de adubação, destacando-se com maior número de raízes totais e comerciais. A maior produtividade de raízes para a cultivar Glória foi obtida com a aplicação de 11,85 ton ha⁻¹ de esterco de frangos de corte, atingindo uma produtividade máxima de 17,01 ton ha⁻¹. As maiores produções estimadas de raízes comerciais para as duas cultivares, foram alcançadas com a dose de 30 t ha⁻¹ de esterco de frangos, representando 48% do total de raízes produzidas.

Palavras-chave: batata-doce, doses, esterco, produtividade, raízes

¹ Departamento de Estatística, UFLA – elpidio9332@gmail.com

² Departamento de Agricultura, UFLA – orlandocefet@yahoo.com.br

³ Departamento de Estatística, UFLA – rrlima@ufla.br

⁴ Departamento de Ciência dos Alimentos, UFPel – nhmubay@live.com



Efeitos de diferentes concentrações de Poliacrilamida

Aniônica sobre propriedades físicas do solo

Douglas Henrique Bandeira ¹
Bruno Rodriguez Leite Chaves ²
Murilo Augusto Destefani ³
Julia de Paula Cau ⁴
Salatiel Bezerra Alves ⁵

Resumo

A degradação dos solos agrícolas compromete sua produtividade e sustentabilidade, tornando necessária a busca por práticas que promovam a manutenção da qualidade do solo. Nesse cenário, surgem diversas técnicas de manejo para mitigar esses problemas, entre elas o uso de condicionantes químicos, como a poliacrilamida aniônica (PAM). Embora existam resultados promissores de estudos realizados em outras regiões e condições, há uma lacuna de informações sobre os efeitos da PAM em solos tropicais. Diante disto, o objetivo desse estudo foi avaliar o efeito de doses crescentes de PAM aplicadas à lanço no solo sobre alguns atributos físicos do solo e na produção de biomassa deixada pela cultura do milho. O experimento foi conduzido em um Latossolo Vermelho distroférico, em parcelas com dimensões de 75 m². Os tratamentos foram cinco, com três repetições, e as doses estudadas foram 0, 25, 50, 75 e 100 kg ha⁻¹, aplicados à lanço imediatamente após a semeadura do milho em sistema convencional de preparo do solo. A caracterização do solo foi realizada antes e após o ciclo do milho, em três profundidades (0-5, 5-10 e 10-20 cm). Observou-se que os tratamentos com doses de 75 e 100 kg ha⁻¹ apresentaram menores valores de densidade do solo, maiores valores de porosidade total e melhor estabilidade dos agregados em água, diferindo-se significativamente dos demais tratamentos, na camada superficial. Para produção de biomassa vegetal, o tratamento que recebeu dose de 100 kg ha⁻¹, apresentou maiores valores, indicando ser a dose que melhor promoveu a preservação da qualidade do solo.

Palavras-chave: Condicionante do solo; manejo do solo; práticas conservacionistas.

¹ Departamento de Ciência do Solo, Unesp Campus Jaboticabal-SP – douglas.bandeira@unesp.br

² Departamento de Ciência do Solo, Unesp Campus Jaboticabal-SP – bruno.chaves@unesp.br

³ Departamento de Ciência do Solo, Unesp Campus Jaboticabal-SP – murilo.destefani@unesp.br

⁴ Departamento de Ciência do Solo, Unesp Campus Jaboticabal-SP – julia.cau@unesp.br

⁵ Departamento de Ciência do Solo, Unesp Campus Jaboticabal-SP – salatiel.bezerra@unesp.br



O Uso da Modelagem para Interpretar o Efeito do Genoma e de Mudanças Climáticas sobre a Dispersão de *Coccidioides spp.*

Cezar Augusto Fontana Teixeira¹ Maria Judith Serrano Rodrigues de Oliveira¹ Mariana Tozzi Rodrigues¹ Alice Pitanga Rocha¹ Lorena Souza Castro Altoé¹ Matheus Correia Casotti¹ Flavia de Paula¹ Iúri Drumond Louro¹ Débora Dummer Meira¹

Resumo

Coccidioides é um gênero de fungos dimórficos e patogênicos, adaptados a climas desérticos e solos alcalinos, predominantes nas Américas. Transmitidos pelo vento, seus esporos causam a "febre do vale", que pode evoluir para lesões pulmonares, disseminação hematogênica ou meningite, exigindo tratamento vitalício em 10% dos casos. Espera-se um aumento da dispersão até o final do séc. XXI por causa das mudanças climáticas, o que torna a análise de modelos meteorológicos e genômicos para prever esse cenário e seus efeitos. Demonstrar a importância da modelagem estatística na previsão da dispersão de *Coccidioides*, correlacionando-a com fatores genômicos e seus impactos. Revisão bibliográfica com estratégia de busca: ("fungi" OR "fungus") AND "pathogenic" AND ("onygenales" OR "coccidioides") AND ("models" OR "genome") AND "distribution". Modelos como PRISM, RCP, CMIP5 e RF preveem aumento de 6°C nos EUA até o fim do século 21, expandindo sua área endêmica até a fronteira do Canadá. No genoma, genes como DRK1 (transição sapróbia-parasítica), adesinas (BAD1, SOWgp), CBP1 e enzimas como catalase peroxidase são essenciais à virulência desse gênero. Ademais, a reprodução sexuada gera variabilidade genética, potencializando adaptações a mudanças climáticas, resistência a hospedeiros e antifúngicos. O aquecimento global poderá aumentar a dispersão da "febre do vale" na América do Norte, com riscos similares na América Central e do Sul, incluindo o Brasil. A modelagem estatística e a genômica são ferramentas vitais para prevenir e conter o contágio, especialmente diante de possíveis adaptações que aumentem a virulência e resistência.

Palavras-chave: Fungo patogênico; Coccidiomicose; Modelagem estatística; Genômica; Aquecimento global.



Classificação de Doenças em Folhas de Cafeiro com Redes Neurais Convolucionais

Karina Vieira dos Santos Fonseca;¹ Renísio Braulio Baldini² Eric Batista Ferreira²

Resumo

A cafeicultura é uma das principais atividades agrícolas do Brasil, com papel relevante na economia nacional e internacional Lopes (2018). Contudo, a produtividade dessa cultura é frequentemente comprometida por doenças foliares causadas por fungos e bactérias, que impactam negativamente a qualidade e o rendimento dos frutos. Além disso, diagnósticos imprecisos favorecem o uso incorreto de defensivos agrícolas, contribuindo para o surgimento de cepas resistentes e acarretando maiores danos ambientais.

Visando enfrentar esses desafios, este estudo propõe uma abordagem baseada em aprendizado profundo para a detecção e classificação automática de doenças foliares em folhas de cafeiro. Foi empregada a arquitetura EfficientNetB0, previamente treinada no conjunto de dados ImageNet, sendo adaptada à tarefa por meio de *fine-tuning*, conforme metodologia descrita por Tajbakhsh (2016). As imagens utilizadas pertencem ao *JMUBen Coffee Dataset*, composto por amostras reais capturadas em diferentes condições de campo, com e sem sintomas de doenças. Durante o treinamento, a rede neural alcançou acurácia de 99,45% e obteve 100% de acurácia de validação na quarta época. Na etapa de testes, o modelo demonstrou excelente desempenho, com acurácia de 99,99%, além de *precision*, *recall* e *f1-score* perfeitos em todas as cinco classes de doenças (Cercospora leaf spot, Coffee leaf rust, Brown eye spot, Anthracnose e Bacterial blight). A matriz de confusão indicou apenas um erro de classificação em mais de 11 mil amostras.

Embora tenham ocorrido pequenos *warnings* técnicos durante o treinamento, como mensagens sobre serialização de tensores e ausência do parâmetro *val_acc*, esses não afetaram a eficácia do modelo. Os resultados evidenciam o potencial da metodologia para aplicação prática na agricultura, contribuindo para diagnósticos fitossanitários mais precisos e sustentáveis. O estudo reforça a importância da integração entre redes neurais convolucionais e agricultura de precisão Mohanty et al. (2016).

Palavras-chave: Cafeicultura, Doenças foliares, Aprendizado profundo, Redes neurais convolucionais.

References

- Lopes, L. P. (2018). Predicting the price of brazilian natural coffee using statistical machine learning models. *Sigmae* 7(1), 1–16.
- Mohanty, S. P., D. P. Hughes, and M. Salathé (2016). Using deep learning for image-based plant disease detection. *Frontiers in plant science* 7, 1419.
- Tajbakhsh, N. e. a. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE transactions on medical imaging* 35(5), 1299–1312.

¹Mestrando, PPGEAB, Universidade Federal de Alfenas, Brasil - karina.santos@sou.unifal-mg.edu.br

²Mestrando, PPGEAB, Universidade Federal de Alfenas, Brasil - renisio.braulio@sou.unifal-mg.edu.br

²Professor, PPGEAB, Universidade Federal de Alfenas, Brasil - eric.ferreira@unifal-mg.edu.br



Modelagem linear e não linear na estimativa volumétrica de clones de eucaliptos no Semiárido brasileiro

Maria Marciele de Lima Silva ¹

Natália Moraes Cordeiro ²

Moacyr Cunha Filho ³

José Antonio Aleixo da Silva ⁴

Resumo

A modelagem volumétrica permite a inclusão de novas variáveis durante o ajuste de modelos. Esta técnica tem sido considerada efetiva para obter estimativas de maior precisão em relação ao volume real da árvore, no entanto, a eficiência dos modelos depende da espécie analisada, das condições em que ela se encontra e da precisão dos dados coletados. Assim, essa pesquisa tem como objetivo ajustar modelos lineares usuais para estimativa de altura de árvores, e, a partir da incorporação destes, realizar ajustes sob os modelos de Spurr, Spurr – não linear, Schumacher – Hall, e Silva – Bailey, com o intuito de identificar a equação que forneça estimativas volumétricas mais consistentes com base no Índice de Schlaegel, no erro padrão da estimativa e na análise da dispersão gráfica dos resíduos. Foram utilizados dados de um experimento composto por três clones de *Eucalyptus* plantados em densidades diferentes em Araripina – PE. A melhor estimativa para a altura foi obtida por meio da equação gerada em função de He e dos diâmetros medidos a 50 cm, 150 cm e 170 cm. Além disso, visando a incorporação da altura ajustada no ajuste dos modelos volumétricos, foram realizados dois ajustes, os quais indicaram a equação gerada pelos ajustes realizados sob o modelo de Schumacher - Hall como a mais eficiente. Embora os dois ajustes tenham apresentado estatísticas descritivas muito próximas, nota-se que as estimativas geradas por Schumacher – Hall ajustado II estão mais próximas do volume real.

Palavras-chave: Ajuste; Experimentação florestal; Estimativas; Modelos estatísticos; Volume.

¹ Departamento de Estatística e Informática, UFRPE – Recife/PE – marciele.liima@gmail.com

² Departamento de Estatística e Informática, UFRPE – Recife/PE – nataliamcorddeiro@gmail.com

³ Departamento de Estatística e Informática, UFRPE – Recife/PE – moacyr2006@gmail.com

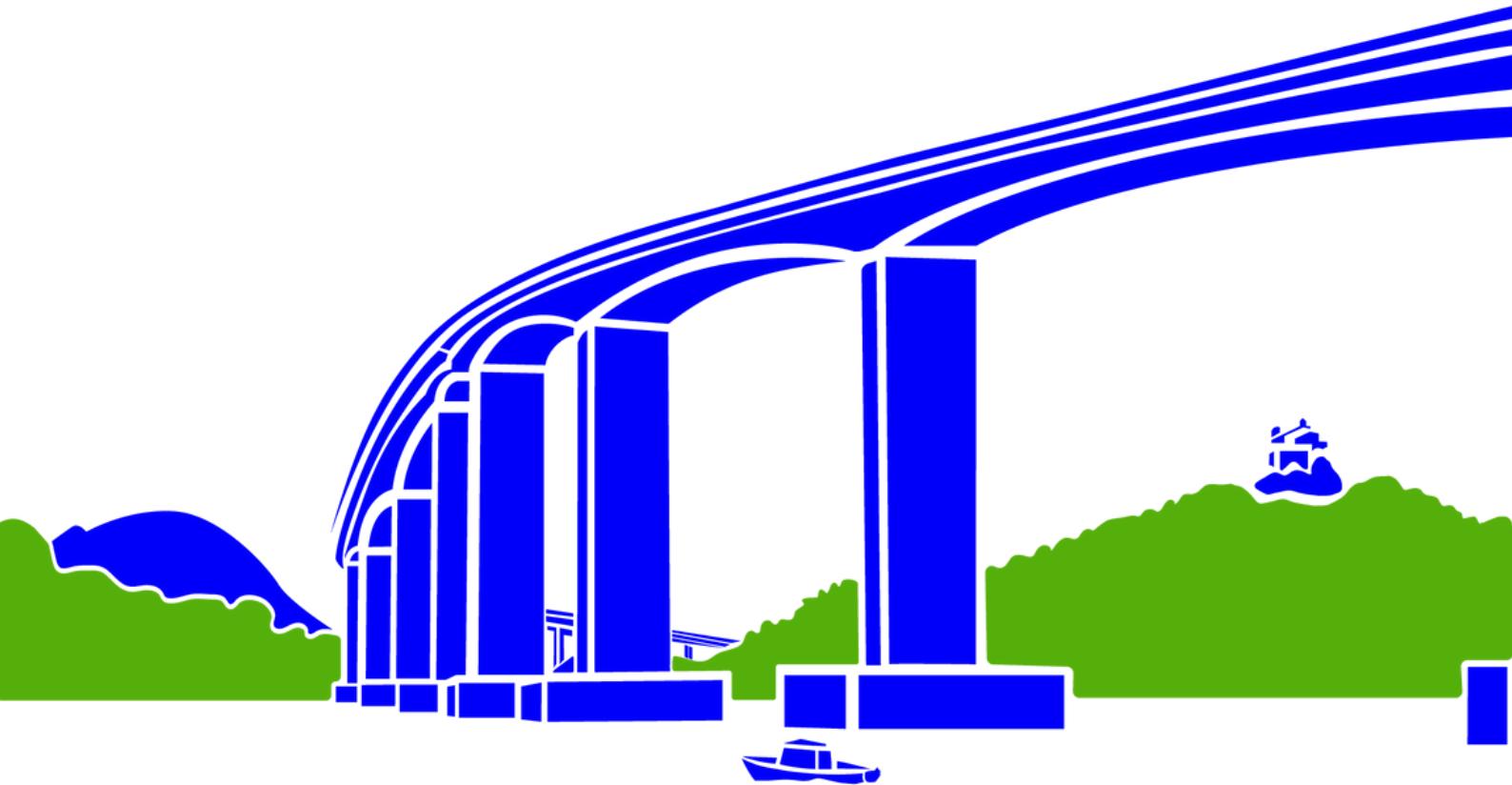
⁴ Departamento de Estatística e Informática, UFRPE – Recife/PE – jaaleixo@gmail.com

VITÓRIA-ES

**69^a Reunião Anual da
RBras
21º SEAGRO**
2025

Pôsters

**Estatística Aplicada em
Agronomia e Biologia 2**





Da Agricultura 1.0 à 5.0: A Transformação Estatística no Setor Agrícola

Leandra Cardoso Moreira ¹

Paulo César de Resende Andrade ²

Resumo

A agricultura tem evoluído continuamente para atender à crescente demanda global por alimentos e matérias-primas, impulsionada por sucessivos avanços tecnológicos. Desde a Agricultura 1.0, baseada em práticas manuais, até a Agricultura 5.0, marcada pelo uso de inteligência artificial e aprendizado de máquina, os métodos estatísticos têm desempenhado um papel fundamental nessa transformação. Com a mecanização agrícola (Agricultura 2.0), surgiram os primeiros métodos estatísticos estruturados, inicialmente aplicados manualmente ou por softwares básicos. A automação e informatização da Agricultura 3.0 permitiram o uso de ferramentas como SAS e Sisvar, possibilitando análises mais rápidas e precisas, incluindo testes de médias e análise de variância. Já na Agricultura 4.0, a integração de big data, sensores digitais e plataformas inteligentes ampliou o uso de softwares avançados e de código aberto, como R e Python, viabilizando análises multivariadas e modelagem preditiva. Atualmente, a Agricultura 5.0 potencializa essa abordagem ao incorporar algoritmos sofisticados para previsão de cenários e apoio à tomada de decisões estratégicas. A estatística tornou-se indispensável para a agricultura moderna, permitindo análises objetivas e decisões baseadas em dados, resultando em maior eficiência no manejo agrícola. Os softwares estatísticos são essenciais nesse processo, possibilitando análises complexas de forma ágil e precisa, contribuindo para o planejamento agrícola, a escolha de cultivares, a otimização de insumos e a implementação de práticas sustentáveis, elevando a produtividade e reduzindo desperdícios.

Palavras-chave: Agricultura digital; Big data; Inteligência artificial; Modelagem preditiva; Softwares estatísticos.

¹ Programa de Pós Graduação em Produção Vegetal, Universidade Federal dos Vales do Jequitinhonha e Mucuri, Diamantina/MG – leandra.cardoso@ufvjm.edu.br

² Instituto de Ciência e Tecnologia, Universidade Federal dos Vales do Jequitinhonha e Mucuri, Diamantina/MG – paulo.andrade@ict.ufvjm.edu.br



Protocolo de fenotipagem de alto rendimento no melhoramento de batata-doce

Matheus Lopes Machado¹

Resumo

A obtenção eficiente e precisa de dados morfológicos é essencial para o sucesso de programas de melhoramento genético vegetal, permitindo a identificação de atributos genéticos associados a características desejáveis. Na batata-doce, a classificação do formato das folhas é fundamental para a análise da diversidade morfológica e identificação de genótipos com potencial ornamental. Apesar da existência de classes bem definidas na literatura, a atribuição manual de genótipos a essas categorias ainda é subjetiva e ineficiente. Além disso, a classificação de características comerciais, como tamanho, sanidade, coloração e conformação, é comum em programas de melhoramento, mas depende de mão de obra intensiva e tem baixa escalabilidade. O uso de aprendizado de máquina e automação oferece uma solução eficiente, permitindo execuções rápidas por meio de paralelização e otimização de recursos. Para fortalecer o cultivo da batata-doce, é necessário aprimorar a coleta de dados fenotípicos, e uma abordagem promissora é o uso de imagens digitais para a obtenção de caracteres produtivos e qualitativos. Assim, este projeto propõe um protocolo de fenotipagem de alto rendimento para quantificação da diversidade genética, caracterização e classificação de genótipos de batata-doce. Utilizando fotometria de folhas e raízes com base em imagens digitais. A ferramenta digital desenvolvida permitirá análises rápidas, acessíveis e fácil criação de bancos de dados estruturados, facilitando sua aplicação em diferentes centros de pesquisa sem exigir conhecimentos técnicos aprofundados, nem prévia experiência.

¹ Curso de Agronomia, Centro de Ciências Agrárias, Universidade Federal de Santa Catarina, campus Florianópolis – matheuslmachado03@gmail.com



Testando se uma folha pertence as espécies amostradas

Giovani Festa Paludo ¹; Amanda Merian Freitas Mendes ²; Gean Pereira Damaceno ³;
Ali William Canaza Cayo ⁴; Júlio Sílvio de Sousa Bueno Filho ⁵

Resumo

Caracteres morfológicos de plantas historicamente tem sido estudados para o problema de identificação de espécies. Contudo, com as novas ferramentas de análise de imagens e a consequente obtenção de conjuntos maiores de variáveis com maiores níveis de correlação, trouxeram novas possibilidades de aplicações ao problema. Nesse sentido, o objetivo foi estabelecer uma estratégia para testar se um indivíduo (representado por um vetor p -variado de características da folha) veio de uma das populações já caracterizadas (isto é, uma espécie botânica). Foi construído um banco de dados com 2009 folhas de 10 espécies. As folhas foram escaneadas, processadas e obtidas as variáveis morfológicas. Foram estabelecidos e comparados vários testes, sendo destacados os 2 principais: um baseado na verossimilhança da normal multivariada e; outro na verossomilhança da log-normal multivariada, em que diferentes quantis da normal foram utilizados para decisão. Os testes foram comparados quanto à sensibilidade e especificidade, e procurou-se estabelecer o quantil da distribuição que maximiza a especificidade, e após isso, escolheu-se o teste com maior sensibilidade. Como resultado, o teste baseado na verossimilhança da log-normal foi o que obteve a maior taxa de sensibilidade (52,7%) mantendo a especificidade em 99,8% e, portanto, é o teste recomendado entre os avaliados. Tal resultado também indica a presença de assimetrias nas distribuições das variáveis. Sendo que o estabelecimento de um teste apropriado para a presente estrutura de dados é um passo que poderá contribuir na construção ou melhoria de sistemas probabilísticos para identificação de espécies vegetais a partir de dados morfológicos.

Palavras-chave: morfometria; testes de hipóteses; problema taxonômico; inferência multivariada; poder do teste.

¹Doutorando do Programa de Pós Graduação em Estatística e Experimentação Agropecuária, UFLA, MG - gfpaludo@gmail.com

²Professora do Departamento de Estatística - Universidade Federal do Paraná

³Mestrando do Programa de Pós Graduação em Estatística e Experimentação Agropecuária, UFLA, MG

⁴Doutorando do Programa de Pós Graduação em Estatística e Experimentação Agropecuária, UFLA, MG

⁵Professor do Departamento de Estatística - Universidade Federal de Lavras



Teoria de Valores Extremos na análise da taxa de câmbio no Brasil: uma aplicação com o pacote *ExtremeTS*

Arthur Fellipy Pereira Silva ¹

Gilberto Rodrigues Liska ²

Marta Cristina Marjotta-Maistro ³

Resumo

O objetivo foi investigar o comportamento estocástico das taxas de câmbio de venda do dólar no Brasil, utilizando a Teoria de Valores Extremos (TVE). Para tal, foi utilizada uma série temporal de dados econômicos referentes à taxa de câmbio em dólar disponibilizado pelo Banco Central compreendendo o período de 1985 a 2025. Foi utilizado o pacote *ExtremeTS* (autoria própria) para obter as séries máximas mensais da taxa de câmbio, o pacote *evd* (CRAN do R) para a análise de extremos e o pacote *hydroGOF* (CRAN do R) para as métricas de qualidade de ajuste. Para fins de análise, foram consideradas a série original, o logaritmo da série e a diferença dos logaritmos das taxas originais (série *log* de retorno das taxas). Os resultados indicaram que a metodologia baseada na diferença dos logaritmos das taxas originais apresentou maior estabilidade, tornando-se a abordagem indicada para modelagem da série. A comparação entre as distribuições probabilísticas mostrou que a Generalizada de Valores Extremos (GVE) apresentou melhor ajuste para 9 dos 12 dos meses. Os cálculos de *Value-at-Risk* (VaR) demonstraram probabilidades associadas a diferentes níveis de retorno, fornecendo estimativas para eventos extremos na taxa de câmbio. Em termos probabilísticos, os meses de janeiro a março apresentaram as maiores probabilidades de ocorrerem taxas de câmbio superiores a 2%. A abordagem estatística utilizada permite uma modelagem mais precisa da volatilidade cambial, útil para investidores, economistas e formuladores de políticas públicas.

Palavras-chave: Mercado Internacional; Mercado cambial; Linguagem R

¹ Discente da graduação em Engenharia Agronômica, CCA-UFSCar – arthurfps@estudante.ufscar.br

² DTAiSER, CCA-UFSCar – gilbertoliska@ufscar.br

³ DTAiSER, CCA-UFSCar – marjotta@ufscar.br



Precipitação Extrema e Riscos Climáticos na Agricultura: Um Estudo em Cascavel (PR)

Valdeline de Paula Mequelinho Ferreira;¹ Luiz Otávio de Oliveira Pala² Thelma Safadi³
Resumo

Este estudo analisou dados de temperatura máxima mensal de janeiro de 1995 a janeiro de 2025 para a cidade de Cascavel (PR), obtidos da Administração Nacional da Aeronáutica e Espaço (NASA), com o objetivo de avaliar a ocorrência de eventos extremos de precipitação e seus possíveis impactos na agricultura. Utilizou-se a Teoria de Valores Extremos, por meio da Distribuição Generalizada de Pareto (GPD), para modelar os valores acima de um limiar e calcular os tempos de retorno para diferentes intervalos. O teste de Mann-Kendall revelou uma tendência não significativa de alteração nas séries temporais (valor-p = 0,106), enquanto o teste de Kolmogorov-Smirnov indicou que os dados seguem uma distribuição adequada para a análise (valor-p = 0,2244). Os resultados mostraram que, em um período de 3 anos, espera-se que a precipitação atinja ou ultrapasse 199,92 mm com uma probabilidade de 100%, e, em um intervalo de 5 anos, a precipitação pode atingir 239,36 mm com uma probabilidade de 91,80%. Chuvas com essa intensidade podem resultar em encharcamento do solo, dificultando práticas agrícolas como o manejo fitossanitário e a aplicação de fertilizantes, além de aumentar o risco de erosão. Esses resultados destacam a importância de monitorar e planejar estratégias para mitigar os impactos de eventos climáticos extremos, especialmente para o cultivo da soja, uma cultura sensível ao excesso de umidade em fases críticas do ciclo produtivo.

Palavras-chave: Teoria dos valores extremos; eventos climáticos extremos; precipitação intensa; tempos de retorno; produtividade da soja; mudanças climáticas.

¹Departamento de Estatística e Experimentação Agrícola, Universidade Federal de Lavras (UFLA), Lavras-MG - valdeline.ferreira1@estudante.ufla.br

²Departamento de Estatística e Experimentação Agrícola, Universidade Federal de Lavras (UFLA), Lavras-MG - luizpala@ufla.br

³Departamento de Estatística e Experimentação Agrícola, Universidade Federal de Lavras (UFLA), Lavras-MG - safadi@ufla.br



Potencial da Análise Discriminante de Padrões Espectrais na Detecção da Murcha-de-Ceratocystis em Eucaliptos

Maycon Junio Ivo Vieira¹; Nerilson Terra Santos²; Luiz Alexandre Peternelli³

Resumo

A Murcha-de-Ceratocystis, causada pelo fungo Ceratocystis Fimbriata, prejudica o eucalipto e sua disseminação é rápida, manifestando-se em diferentes estágios. O sensoriamento remoto é uma técnica hábil de identificar doenças, monitorando alterações nas variações da radiação emitida pela planta infectada. Este estudo avaliou o potencial discriminatório dos padrões espectrais de clones de eucalipto para diferenciar grupos suscetíveis e resistentes, utilizando a câmera multiespectral MicaSense RedEdge-MX, que captura imagens em cinco sensores eletromagnéticos (Azul, Verde, Vermelho, RedEdge e infravermelho próximo) em condições controladas, a nível de bancada. As imagens foram processadas para extração das bandas Red, Green, Blue, RedEdge e Nir, e alguns índices de vegetação. Após reduzir a multicolinearidade, aplicou-se Análise Discriminante Linear para classificar os grupos, avaliando-se o desempenho por meio da acurácia e dos coeficientes de discriminação linear. Resultados preliminares indicam que as variáveis NIR, Red Edge e NDVI são as variáveis melhor discriminadoras para esse conjunto. Em especial o NDVI foi observado com uma média maior no grupo dos resistentes. Portanto, a análise de dados espectrais, combinada com técnicas estatísticas, apresenta potencial para uso na detecção de doenças em plantas.

Palavras-chave: Sensoriamento Remoto; NDVI; Imagens Multiespectral; Eucalipto-Doença.

¹ Universidade Federal de Viçosa, Departamento de Estatística - maycon.ivo@ufv.br

² Universidade Federal de Viçosa, Departamento de Estatística - nsantos@ufv.br

³ Universidade Federal de Viçosa, Departamento de Estatística - peternelli@ufv.br



Comparação de Índices para Seleção de Regiões Genômicas em Estudos de Associação Genômica Amplia

ROQUE, R. D¹, AZEVEDO, C. F²

NASCIMENTO, M.³, NASCIMENTO, A. C. C.⁴

Resumo

A associação genômica visa identificar regiões do genoma associadas a características fenotípicas, por meio da detecção de loci de efeitos genéticos (QTLs). Diversos métodos estatísticos têm sido empregados nesse contexto, como o *Mixed Linear Model* (MLM), *Compressed MLM* (CMLM), *General Linear Model* (GLM), *Settlement of MLM Under Progressively Exclusive Relationship* (SUPER), *Marker-Assisted Best Linear Unbiased Prediction* (MABLUP) e *FarmCPU*, cada um com suas particularidades metodológicas. O objetivo deste trabalho foi comparar cinco índices para auxiliar na seleção de regiões genômicas relevantes para essas características. Um índice foi calculado sem ponderação, enquanto os demais foram ponderados com base nos efeitos dos marcadores da região, valores p e q associados à significância desses efeitos, e na proporção da variância genética explicada, visando reduzir a detecção de falsos positivos. Foram simulados mil indivíduos genotipados com seis mil marcadores moleculares, considerando três características fenotípicas controladas por cem, cinquenta e dez QTLs. A comparação entre os índices considerou a correlação entre os efeitos dos marcadores e os índices, além da área sob a curva (AUC) entre verdadeiros e falsos positivos. O índice sem ponderação apresentou menor correlação com os efeitos, indicando baixa capacidade de capturar a importância das regiões genômicas. A ponderação, especialmente pelo valor- q , aprimorou a identificação das regiões associadas, com destaque para a característica com dez QTLs, podendo beneficiar programas de melhoramento na seleção de indivíduos geneticamente superiores e no entendimento da arquitetura genética. No entanto, os índices não se diferenciam quanto à AUC, que se manteve constante entre as abordagens testadas.

Palavras-chave: marcadores moleculares; melhoramento genético; arquitetura genética; correlação; simulação de dados.

¹ Programa de Pós Graduação em Estatística Aplicada e Biometria, UFV – renata.roque@ufv.br

² Departamento de Estatística, Universidade Federal de Viçosa – camila.azevedo@ufv.br

³ Departamento de Estatística, Universidade Federal de Viçosa – moysesnascim@ufv.br

⁴ Departamento de Estatística, Universidade Federal de Viçosa – ana.campana@ufv.br

Agradecimento à CAPES pelo apoio financeiro.



Variações na fonte de Fósforo (P) No Desenvolvimento Inicial Da Altura e do Diâmetro do Caule Do Café

Matheus da Costa Lima ¹

Joel Augusto Muniz ²

Resumo

O fósforo é essencial para o desenvolvimento inicial do cafeeiro, atuando no crescimento vegetativo e radicular. Este estudo avaliou o efeito de diferentes fontes de fósforo no crescimento inicial de mudas de Coffea arabica, com enfoque nas variáveis altura de plantas e diâmetro do caule. O experimento foi conduzido na fazenda Experimental da PROCAFÉ em Varginha-MG, utilizando delineamento inteiramente casualizado com seis tratamentos e cinco repetições. Os tratamentos incluíram: testemunha (sem fertilizante), 2 tratamentos utilizando o fertilizante biofosfato (na dose de 330g e 230g/vaso), superfosfato simples (200g/vaso), mistura biofosfato+superfosfato (165g+100g/vaso) e organomineral vegetal (270g/vaso). As variáveis altura e diâmetro do caule foram avaliadas mensalmente durante seis meses. Os dados foram submetidos à análise de variância e teste de Scott-Knott ($p<0,05$). O organomineral vegetal e a mistura de fertilizantes proporcionaram as maiores alturas finais (25,6cm e 26,3cm, respectivamente). Para o diâmetro do caule, a mistura biofosfato+superfosfato apresentou o maior valor (8,3cm), seguido pelo organomineral vegetal (7,0cm). A testemunha apresentou os menores valores para ambas as variáveis (21,0cm de altura e 5,8cm de diâmetro). Esses resultados demonstram que fontes mais solúveis de fósforo, como o organomineral vegetal, e o biofosfato promovem melhor desenvolvimento inicial do cafeeiro. O uso de fontes solúveis de fósforo, em especial o organomineral vegetal, mostrou-se eficaz para promover o crescimento inicial do cafeeiro, com destaque para o aumento do diâmetro do caule, indicador de vigor vegetativo. Esses resultados sugerem que a escolha adequada da fonte fosfatada pode contribuir para a formação de mudas mais robustas e produtivas.

Palavras-chave: coffeea arabica 1; nutrição mineral 2; crescimento vegetativo3; fertilizantes fosfatados 4.

¹ Departamento de Estatística, Lavras – matheus.lima8@estudante.ufla.br

² Departamento de Estatística, Lavras – joamuniz@ufla.br



Visualização Gráfica da Adaptabilidade, Estabilidade e Produtividade de Genótipos de Cana-de-Açúcar com Índices Baseados em Modelos Mistos

Tays Silva Batista ¹ Luiz Alexandre Peternelli ¹

Resumo

A cana-de-açúcar é uma cultura tropical de elevada importância econômica. Em programas de melhoramento, a avaliação de genótipos em ensaios multiambientais é essencial para selecionar clones com alta produtividade, adaptabilidade e estabilidade frente à interação genótipo × ambiente ($G \times A$). Este estudo teve como objetivo avaliar e visualizar o comportamento genotípico de clones de cana-de-açúcar ao longo de três cortes, utilizando duas abordagens, o índice HMRPGV (sintetiza essas características em uma única métrica), e o biplot WAASB × RPGV (visualização conjunta da estabilidade, adaptabilidade e da produtividade). Foram avaliados 30 genótipos em 36 ambientes (local × ano), em Minas Gerais. Devido ao desbalanceamento dos dados, utilizou-se modelo misto linear com a inclusão de covariáveis ambientais. A produtividade (ATR) foi ajustada por BLUP, e os índices HMRPGV, RPGV e WAASB foram estimados. A aplicação do HMRPGV permitiu ranquear genótipos com alto desempenho e estabilidade, como RB966928 e CTC4. Os biplots WAASB × RPGV revelaram padrões distintos entre os cortes, identificando genótipos com alta adaptabilidade e baixa estabilidade e outros estáveis, porém menos produtivos. A comparação com os biplots WAASB × RPGV revela nuances adicionais. Enquanto o HMRPGV sintetiza em um único índice a estabilidade e a adaptabilidade, os biplots permitem uma visualização mais detalhada das trocas de posição entre genótipos (interação $G \times A$ do tipo crossover) ao longo dos cortes. Além disso, a separação dos dados por cortes evidencia variações no comportamento genotípico ao longo do ciclo da cultura, refletindo a influência de fatores como rebrota e adaptação ao ambiente de soca.

Palavras-chave: HMRPGV; RPGV; WAASB; ensaio multiambiental; interação $G \times A$.

¹Departamento de Estatística, Universidade Federal de Viçosa, Viçosa, MG, 36570-900, Brazil – tays.batista@ufv.br



Distribuição Espacial de Bovinos no Brasil

Hellen Sonaly Silva Alves;¹ Daví Barbosa Pereira de Sousa² Gabriel Graciano De Mendonça³ Marcela Silva de Araujo⁴ Vitoria Soares de Souza⁵ Guilherme Rocha Moreira⁶ João Domingos Scaloni⁷

Resumo

O Brasil se destaca como um dos maiores produtores e exportadores mundiais de carne bovina, com isso a pecuária bovina tem um papel central na economia brasileira, sendo responsável por uma parcela significativa do Produto Interno Bruto (PIB). O objeto deste estudo foi analisar a distribuição geográfica de bovinos nos estados brasileiros de 2021 a 2023. Os dados foram obtidos da Pesquisa da Pecuária Municipal (PPM), e as análises foram realizadas no IDE RStudio. Foram aplicadas técnicas de análise espacial, incluindo os índices de Moran Global que avaliaram se a presença de bovinos em uma área influencia áreas vizinhas e Moran Local onde identificou-se agrupamentos com características semelhantes. Houve ausência de dependência espacial na quantidade de bovinos, com índice de Moran Global não significativo ($valor - p > 0,05$). No entanto, o mapa de quartis indicaram maior concentração de bovinos nas regiões Norte, Centro-Oeste e Sudeste. No período analisado, a Região Centro-Oeste, exceto o Distrito Federal, apresentou influência significativa sobre seus vizinhos, com altos números de bovinos e 1% de significância segundo o Moran Local (LISA). A análise espacial pode ser utilizada como ferramenta para compreensão da distribuição da produção de bovinos no Brasil, com isso será possível avaliar fatores como disponibilidade de pastagem, infraestrutura e demanda de mercado, contribuindo para planejamento mais eficiente e sustentável.

Palavras-chave: Análise espacial; Distribuição geográfica; Produção pecuária.

¹Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, Recife - hellen.silva@ufrpe.br

²Universidade Federal de Lavras, Departamento de Estatística, Lavras - davi.barbosa.5454@gmail.com

³Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, Recife - gabriel.gracianom@ufrpe.br

⁴Universidade Federal de Lavras, Departamento de Estatística, Lavras - marcela.araujo1@estudante.ufla.br

⁵Universidade Estadual da Paraíba, Departamento de Estatística, Campina Grande- vitoria-soares0104@gmail.com

⁶Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, Recife - guilherme.rochamoreira@ufrpe.br

⁷Universidade Federal de Lavras, Departamento de Estatística, Lavras - scalon@ufla.br



Anomalias na produtividade municipal de feijão, milho e soja, em Goiás, de 1992 a 2021, definidas por regressão linear

Alfredo José Barreto Luiz ¹

Nathália Faria de Sousa ²

Fernando Antônio Macena da Silva ³

Resumo

Identificar as consequências das mudanças climáticas na produção agrícola é fundamental para a segurança alimentar. Para localizar esses possíveis efeitos no tempo e no espaço, verifica-se a ocorrência de anomalias na produtividade local a cada safra. Embora o cálculo de anomalias geralmente se baseie na média, em séries temporais da produtividade agrícola pode haver uma tendência de mudança que precisa ser considerada. Neste trabalho, ajustaram-se equações de regressão linear simples entre o passar dos anos e a produtividade de feijão, milho e soja em 34, 61 e 33 municípios de Goiás, respectivamente. Foram utilizados dados municipais de produtividade da Pesquisa Produção Agrícola Municipal (PAM) do IBGE, de 1992 a 2021. Para selecionar municípios, o critério foi possuírem pelo menos 20 anos com a produtividade de cada cultura. Em todas as regressões ajustadas a inclinação foi positiva, o que indicou aumento da produtividade a cada ano. O coeficiente de determinação (r^2) assumiu valores médios de 0,47 para o feijão, 0,50 para o milho e 0,62 para a soja, e máximos de 0,88, 0,88 e 0,87, respectivamente. Foram consideradas anomalias positivas as diferenças entre produtividades observada e estimada maiores que erro padrão da estimativa média (STDP); e negativas as menores que menos um STDP. Isso resultou em 359 (40,2%) positivas e 346 (38,8%) negativas para o feijão, 734 (40,2%) positivas e 701 (38,4%) negativas para o milho, e 442 (44,7%) positivas e 342 (34,6%) negativas para a soja. Os anos com mais anomalias negativas foram 1999 (feijão), 2016 (milho) e 2015 (soja).

Palavras-chave: agrometeorologia; erro padrão da previsão; mudanças climáticas; precipitação pluviométrica; El Niño.

¹ Embrapa Meio Ambiente, Jaguariúna, SP – alfredo.luiz@embrapa.br

² Bolsista PIBIC, graduanda, Engenharia Florestal, UFSCar, Sorocaba – nathalia01faria@gmail.com

³ Embrapa Cerrados, Planaltina, DF – fernando.macena@embrapa.br



Modelos de Apenas-Presença: Uma Análise Comparativa de Métodos Estatísticos Computacionais

Maria Luísa da Silva Amancio ¹

Rosangela Helena Loschi ²

Marcos Oliveira Prates ³

Resumo

Este estudo investiga abordagens de modelagem para dados de apenas-presença, comuns em ecologia, epidemiologia e geociências. Esses dados registram apenas ocorrências, sem confirmação de ausências, o que impõe desafios à inferência da distribuição espacial de espécies. O objetivo é comparar métodos estatísticos aplicáveis a processos pontuais, como IWLR, DWPR, Spatstat e abordagens bayesianas (bayesPO, Inlabru). A análise, baseada em dados reais de ocorrência, avalia implementação, desempenho e aderência dos modelos à literatura. Modelos GLM, como IWLR e DWPR, são eficientes, mas exigem pseudo-ausências. O Spatstat requer pré-processamento ambiental. Modelos bayesianos, embora mais exigentes computacionalmente, dispensam pseudo-ausências e oferecem maior flexibilidade. A escolha metodológica depende do contexto, estrutura dos dados e recursos disponíveis. Como próxima etapa, dados simulados serão utilizados para avaliar a robustez dos modelos em diferentes cenários. Este trabalho contribui com uma análise comparativa sistemática, auxiliando na seleção de métodos adequados para dados de apenas-presença.

Palavras-chave: Apenas-presença; Processos Pontuais; Comparaçao de Modelos; Estatística Espacial; IWLR; DWPR; Spatstat; bayesPO; inlabru.

¹ Departamento de Estatística, UFMG – marialuisamancio@gmail.com

² Departamento de Estatística, UFMG – loschi@est.ufmg.br

³ Departamento de Estatística, UFMG – marcosop@gmail.com



COMPARAÇÃO DE ÍNDICES DE SELEÇÃO MULTIVARIADOS EM ENSAIOS MULTIAMBIENTAIS

Wanessa Alves Lima ¹

Moysés Nascimento ¹

Resumo

A existência da interação genótipo x ambiente (G x A) desafia o melhoramento genético, dificultando a identificação de genótipos superiores em diversos ambientes. Para avaliar múltiplas características simultaneamente, empregam-se índices de seleção multivariados como FAI-BLUP, MGIDI e MTSI, que superam limitações estatísticas da abordagem tradicional de Smith e Hazel. Embora baseados em análise de fatores, diferem na construção. FAI-BLUP define um genótipo ideal e calcula a probabilidade de aproximação do mesmo. Já o MGIDI usa um ideal implícito de máximo desempenho, medindo a distância euclidiana e, o MTSI incorpora o conceito de estabilidade na análise, definindo o ideal por desempenho e estabilidade, também usando como métrica a distância euclidiana. Este estudo avaliou o desempenho desses índices simulando nove experimentos em blocos casualizados, com 100 genótipos em três ambientes para quatro características com diferentes herdabilidades e correlações, utilizando os pacotes AlphaSimR e FieldSimR no R. A concordância da seleção de genótipos pelos índices foi comparada por diagramas de Venn. Os resultados indicaram alta concordância entre FAI-BLUP e MGIDI em cenários com herdabilidades homogêneas e correlações moderadas a altas. Contudo, a concordância diminuiu com herdabilidades distintas. O MTSI mostrou menor concordância geral, mas maior flexibilidade em cenários heterogêneos. Conclui-se que FAI-BLUP e MGIDI são mais adequados para herdabilidades homogêneas e correlações moderadas a altas, enquanto MTSI se destaca em cenários mais variáveis.

Palavras-chave: Interação GxA; Seleção de genótipos; Simulação experimental;

¹ Departamento de Estatística – UFV, CEP: 36570-900 – Viçosa/MG – wanessa.a.lima@ufv.br



Avaliação da podridão branca do capítulo do girassol

Santiago Germán Delgado ^{1,2}

Renata Alcarde Sermarini ²

Resumo

A podridão branca do capítulo do girassol (PBC), causada pelo fungo *Sclerotinia sclerotiorum*, é uma doença devastadora em todas as regiões temperadas do mundo onde esta espécie é cultivada. É de interesse dos programas de melhoramento genético desenvolver híbridos com bom desempenho contra o aparecimento da PBC. A severidade da doença, entendida como a proporção do capítulo afetado, não tem sido amplamente estudada devido à dificuldade de mensuração e à grande quantidade de recursos que demanda. Também pode acontecer que em ensaios de campo as condições ambientais sejam favoráveis à doença e o conjunto de dados obtido apresente um grande número de observações em seu valor máximo. O objetivo deste trabalho é propor um modelo estatístico adequado para modelagem da severidade com excesso de valores máximos e um método para seleção das cultivares resistentes. Um conjunto de medidas de severidade foi obtido de 32 híbridos comerciais de girassol, cultivados por seis anos em um delineamento em blocos completos com três repetições. Foi ajustado um modelo de regressão beta inflacionado em uns. Verificou-se que o modelo que considera a estimativa da probabilidade da resposta ser um, dependendo da interação entre o ano e o cultivar e, considerando variâncias heterogêneas devido ao ano no componente sistemático que modela a resposta (0; 1) apresentou melhor adequação aos dados, de acordo com o critério de informação de Akaike e o teste da razão de verossimilhanças para modelos encaixados. Uma abordagem de análise de componentes principais permitiu destacar cultivares resistentes.

Palavras-chave: severidade; excesso de valores máximos; regressão beta inflacionada em uns, seleção genotípica, análise de componentes principais.

¹ Universidad Nacional de Mar del Plata, Facultad de Ciencias Agrarias, Departamento de Introducción a las Ciencias Agrarias – sdelgado@usp.br

² Universidade de São Paulo, Escola Superior de Agricultura Luiz de Queiroz, Departamento de Ciências Exatas – ralcarde@usp.br



AVALIAÇÃO DA MATURAÇÃO DE PEIXES COM REGRESSÕES BINÁRIAS: EXPLORANDO NOVAS FUNÇÕES DE LIGAÇÕES.

Rodolfo de Souza Nascimento¹

Diego Corrêa Alves²

Robson Marcelo Rossi³

Carla Valéria dos Santos Pacheco⁴

Resumo

O tamanho de primeira maturação dos peixes é fundamental para o manejo pesqueiro e a definição de tamanhos mínimos de captura. Este estudo propõe novas abordagens estatísticas para estimar o comprimento no qual 50% (L50) e 99% (L99) dos indivíduos atingem a maturidade sexual, com ênfase em funções de ligação alternativas para modelos binários. Além das funções clássicas *logit* e *probit*, foram investigadas as funções potência e reversa potência, sob uma abordagem bayesiana, por oferecerem maior flexibilidade, especialmente em dados desbalanceados. Os modelos foram aplicados a dados de *Prochilodus lineatus*, provenientes da planície de inundação do Alto Rio Paraná. A comparação entre os modelos considerou critérios como o *Deviance Information Criterion* (DIC), análise gráfica, envelope simulado de resíduos e tempo de processamento. A função de ligação *Potência Cauchit* (PC) apresentou o melhor ajuste segundo o DIC e o envelope simulado, mas superestimou o L99. Já a função *Reversa de Potência Logit* (RPL) apresentou resultados mais equilibrados, com bom desempenho em todos os critérios avaliados. Assim, a RPL foi considerada a opção mais adequada para estimar a maturação da espécie estudada.

Palavras-chave: Curimba; Sábalo; Planície de inundação; regressão logística; biologia pesqueira.

¹ PBE, UEM – rodolfosouza12979@gmail.com

² PBE, UEM – dcalves@uem.br

³ PBE, UEM – rrossi@uem.br

⁴ PBE, UEM – carla.pacheco@ifsc.edu.br



Famílias ruínas podem originar clones superiores? Análise comparativa da seleção na fase inicial (T1) em cana-de-açúcar

Mayara Rodrigues e Silva ¹

Luiz Alexandre Peternelli ²

Resumo

No melhoramento da cana-de-açúcar uma das fases mais importantes corresponde à inicial T1, no qual são selecionados os primeiros indivíduos que serão clonados e cultivados nas fases subsequentes por meio da propagação vegetativa. Acredita-se que a seleção dentro das melhores famílias aumentará a probabilidade de encontrar os melhores indivíduos. Nesse trabalho, objetivou-se comparar famílias selecionadas na fase T1, com o desempenho dos genótipos selecionados dessas famílias em um experimento subsequente. Dois experimentos foram desenvolvidos no Centro de Pesquisa de cana da UFV. No Experimento I, 98 famílias foram analisadas, selecionando-se as 10 melhores e piores para TCH (Tonelada de cana por hectare) e Brix. No Experimento II, avaliaram-se oito colmos (indivíduos) previamente selecionados dessas famílias. Foi aplicado o teste de Wilcoxon para comparar os grupos das melhores e piores famílias. Além disso, foram elaborados gráficos de dispersão com as famílias do experimento I e seus indivíduos avaliados no experimento II. Há variabilidade genética entre os grupos das melhores e piores famílias para os caracteres Brix ($p < 0,01$) e TCH ($p < 0,05$). A análise gráfica confirma que famílias com alto Brix ou TCH produzem com maior probabilidade indivíduos superiores. Embora alguns indivíduos das piores famílias possam se destacar, a maioria dos indivíduos de melhor performance está diretamente associada às melhores famílias. A seleção de famílias com alto valor genotípico maximiza a probabilidade da obtenção de progêneres elite, aspecto importante na fase T1 do melhoramento. CAPES, CNPq, FAPEMIG, FINEP, RIDESA.

Palavras-chave: melhoramento genético; produtividade; *Saccharum* spp.

¹ Departamento de Estatística - UFV, Viçosa - MG – mayara.r.silva@ufv.br

² Departamento de Estatística - UFV, Viçosa - MG – peternelli@ufv.br



Influência da agregação dos dados na análise de trilha de genótipos de linhaça.

Matheus Lopes Machado^{1*} Carlos Eduardo Forcelini Assoni¹ Leonardo Talavera Campos² Tiago Olivoto³

Resumo

A precisão da análise de trilha no melhoramento vegetal é influenciada pela estruturação dos dados. Este estudo avaliou o impacto da amostragem por planta individual em comparação ao uso de médias genotípicas nos coeficientes de trilha em 40 genótipos de *Linum usitatissimum* L. A partir de dez caracteres mensurados, realizou-se o diagnóstico de multicolinearidade e a seleção de variáveis. A análise baseada em médias genotípicas resultou em maior colinearidade e perda de fidelidade causal. Em contrapartida, a abordagem por planta demonstrou maior sensibilidade estatística, revelando estruturas causais consistentes com a fisiologia da espécie e um modelo com multicolinearidade fraca ($NC \approx 40$; $VIF < 1,8$). O número de cápsulas e o índice de capacidade de colheita apresentaram os principais efeitos sobre a massa de grãos por planta, enquanto a altura da planta e a elongação dos grãos mostraram associações secundárias relevantes.

Palavras-chave: Análise de trilha, linhaça, melhoramento genético vegetal, multicolinearidade

¹ Discente do Curso de Agronomia, Universidade Federal de Santa Catarina, Florianópolis, SC

² Pós-doutorando, PPG/RGV, Universidade Federal de Santa Catarina, Florianópolis, SC

³ Docente do Departamento de Fitotecnia, Universidade Federal de Santa Catarina, Florianópolis, SC

* - Contato: matheuslmachado03@gmail.com



Comportamento temporal da herdabilidade via imagens multiespectrais de cana-de-açúcar

Matheus de Paula Ferreira ¹, Aloísio Fernando Silva Ribeiro ²,

Samantha Gouvêa Oliveira ³, Luiz Alexandre Peternelly ⁴

Resumo

A herdabilidade é uma medida genética que expressa a proporção da variação fenotípica (observada) de uma característica que pode ser atribuída à variação genética entre os indivíduos de uma população. Índices de Vegetação (IV) com alta herdabilidade e correlação com características genéticas podem ser usados para selecionar genótipos superiores. O presente trabalho teve por objetivo o mapeamento temporal da herdabilidade das bandas R, G, B, NIR e RE, além de cinco índices de vegetação (NDVI, EVI, NDRE, GNDVI, PSRI) obtidos via imagens multiespectrais. Foram realizados cinco voos sobre um experimento no delineamento em blocos aumentados, no Centro de Experimentação em Cana-de-Açúcar em Oratórios-MG. Para cada parcela do experimento foram extraídas a média e mediana de cada banda e IV, e calculadas as herdabilidades. Os valores de herdabilidade foram analisados graficamente ao longo do tempo. Nota-se uma tendência de crescimento da herdabilidade para todas as bandas e índices estudados. Consequentemente, a idade da planta está diretamente relacionada ao valor da herdabilidade. Além disso, notou-se que a utilização de valores médios ou medianos não influenciou a herdabilidade. Para as bandas o valor máximo da herdabilidade foi de aproximadamente 0,70, exceto para R (0,45). Para os IVs os valores máximos de herdabilidade variaram de 0,45 a 0,7 ($NDVI = 0,45$; $EVI = 0,65$; $NDRE = 0,5$; $GNDVI = 0,55$; $PSRI = 0,7$). Esses resultados fornecem *insights* para otimização do uso de imagens multiespectrais no programa de melhoramento de cana-de-açúcar, auxiliando na determinação de épocas mais propícias para coleta de dados aéreos.

Palavras-chave: índice de vegetação; fenotipagem; seleção indireta; melhoramento genético.

¹ Departamento de Estatística, Universidade Federal de Viçosa – matheus.paula@ufv.br

² Departamento de Agronomia, Universidade Federal de Viçosa – aloisio.fernando@ufv.br

³ Departamento de Estatística, Universidade Federal de Viçosa – samantha.gouvea@ufv.br

⁴ Departamento de Estatística, Universidade Federal de Viçosa – peternelly@ufv.br



Democratizando a modelagem de curvas de crescimento com o aplicativo plimanshiny

Tiago Olivoto¹

Resumo

Modelos de crescimento são amplamente utilizados na agricultura para descrever o desenvolvimento temporal das culturas, permitindo a estimativa de parâmetros biologicamente relevantes, como pontos de inflexão, pontos críticos e integrais definidas. No entanto, sua aplicação em larga escala ainda enfrenta limitações devido a desafios computacionais, como a necessidade de ajustar múltiplos modelos não lineares, definir valores iniciais apropriados e garantir a convergência dos ajustes. Este trabalho apresenta o módulo *Growth models*, integrado ao aplicativo plimanshiny, desenvolvido para tornar esse processo acessível, eficiente e escalável. O módulo realiza a modelagem automatizada de dados de séries temporais (ex., altura de planta) com o uso de funções *self-start* e suporte a processamento paralelo. Ele inclui uma ampla gama de modelos organizados em famílias funcionais: modelos sigmoides (Logistic 3P, Logistic 4P, Gompertz, Trans-Gompertz, Weibull, Beta growth, Hill), modelos exponenciais (Von Bertalanffy, Exponential, Janoschek, Asymptotic, Exponential-Plateau, Expolinear), modelos de pico (Asymmetric Gaussian) e modelos cíclicos (Sinusoidal). Adicionalmente, permite a coleta automatizada de dados climáticos via NASA POWER e o cálculo de variáveis derivadas, como graus-dia (GDD), ampliando a base fisiológica dos modelos ajustados. Com poucos cliques, os usuários podem ajustar múltiplos modelos de crescimento, comparar métricas de ajuste (AIC, RMSE, MAE), analisar derivadas de primeira e segunda ordem, extrair estimativas fenológicas e gerar visualizações interativas. O módulo representa uma solução robusta e versátil que democratiza o uso de modelos de crescimento, permitindo o cálculo automatizado e o pronto acesso a dados que podem ser utilizados em análises fenotípicas, estudos de modelagem preditiva e aplicações em inteligência artificial.

Palavras-chave: Fenotipagem de alto rendimento; Modelos de crescimento; modelos não lineares; shiny.

¹Docente do Departamento de Fitotecnia, Universidade Federal de Santa Catarina – tiago.olivoto@ufsc.br



Dinâmica temporal e diversidade varietal da cana-de-açúcar em Pernambuco/Paraíba/Rio Grande do Norte

Patricia Arruda de Moura,¹ Marilia Gabriela Medeiros Ordomo Lages,² Hellen Sonaly Silva Alves,³ Milene Vieira Figueira,⁴ Amaro Epifânio Pereira Silva,⁵ Victor Casimiro Piscoya,⁶ Renisson Neponuceno de Araújo Filho,⁷ Guilherme Rocha Moreira,⁸ Moacyr Cunha Filho.⁹

Resumo

Este trabalho objetivou avaliar as práticas de adoção e substituição de clones, bem como a diversidade genética da cana-de-açúcar cultivada em Pernambuco, Paraíba e Rio Grande do Norte, no período de 2016 a 2023, em alinhamento ao Objetivo de Desenvolvimento Sustentável (ODS) 2 – Fome Zero e Agricultura Sustentável. Para isso, aplicaram-se modelos de regressão e o teste de Mann–Kendall às séries de área plantada e colhida, analisando tanto variações sazonais quanto cumulativas. A monotonicidade das tendências foi quantificada pelo coeficiente τ , estimado com $\alpha = 0,05$, enquanto a diversidade varietal anual foi medida pelo Índice de Shannon (H). Os resultados indicaram que as cinco variedades mais recorrentes foram RB92579, RB867515, SP791011, RB041443 e SP813250, cada uma apresentando perfis distintos de expansão e declínio. A variedade RB041443, lançada em 2021, exibiu o maior crescimento de área ($p < 0,05$; $\tau \approx 0,71$), enquanto a RB92579, apesar de continuar dominante em extensão cultivada, apresentou declínio gradual. Além disso, a diversidade manteve-se moderada ($H \approx 2,2$) até 2019, atingiu pico em 2020 ($H \approx 2,45$), recuou em 2021 ($H \approx 2,17$) e alcançou níveis altos ($H > 2,5$) a partir de 2022. Esses achados reforçam a importância de ações práticas, tais como a diversificação por grupos de variedades e adoção de esquemas rotacionais para manutenção da diversidade, fortalecendo a resiliência varietal e o manejo sustentável. Além disso, contribuem diretamente para o ODS 2 ao promover sistemas agrícolas mais produtivos, equitativos e perenes, garantindo maior segurança alimentar e otimização dos recursos naturais.

Palavras-chave: cana-de-açúcar; RB041443; índice de Shannon; dinâmica varietal; tendências temporais.

¹Departamento de Estatística e Informática/UFRPE, Recife-PE – patricia.amoura@ufrpe.br

²Departamento de Estatística e Informática/UFRPE, Recife-PE – marilia.lages@ufrpe.br

³Departamento de Estatística e Informática/UFRPE, Recife-PE – hellen.ssilva@ufrpe.br

⁴Departamento de Estatística e Informática/UFRPE, Recife-PE – milene.figueira@ufrpe.br

⁵Estação Experimental de Cana-de-açúcar do Carpina/UFRPE, Carpina-PE, – epifaniao@hotmail.com

⁶Departamento de Tecnologia Rural/UFRPE, Recife-PE – hellen.ssilva@ufrpe.br

⁷Departamento de Tecnologia Rural/UFRPE, Recife-PE – renisson.neponuceno@ufrpe.br

⁸Departamento de Estatística e Informática/UFRPE, Recife-PE – guilherme.rochamoreira@ufrpe.br

⁹Departamento de Estatística e Informática/UFRPE, Recife-PE – moacyr2006@gmail.com



Informações espaciais incorporadas à análise de vizinhança para avaliação de famílias de cana-de-açúcar

Rodrigo da Cruz Nunes ¹

Antônio Policarpo Souza Carneiro ²

Resumo

No delineamento em blocos casualizados (DBC) os princípios básicos da experimentação, repetição e casualização, são contemplados, tendo o segundo, um controle. Os tratamentos são distribuídos em blocos, contendo cada bloco uma única repetição de cada tratamento. Esse controle tem por objetivo eliminar da estimativa do erro experimental as diferenças existentes entre blocos distintos. Entretanto, ao avaliar ensaios em DBC na avaliação genética de cana-de-açúcar, surgem fatores que podem prejudicar a precisão experimental. Por conta do pouco material de genótipos disponível, o experimento apresentará parcelas pequenas e um número reduzido de repetições nos tratamentos e, devido às grandes quantidades de cultivares a serem avaliados, os blocos gerados serão extensos, dificultando a formação de blocos uniformes. Uma das alternativas para melhorar a análise é a utilização de métodos de análise de vizinhança, médias móveis e Papadakis, que capturam a dependência espacial na forma de análise de covariância. Entretanto, essas metodologias possuem divergências quanto ao número e posição das parcelas usadas no método de vizinhança. Este estudo, propôs uma metodologia padrão para a aplicação de médias móveis e Papadakis com base na incorporação de informações espaciais, ao considerar a distância e a covariância espacial entre parcelas como ponderadores. Os resultados indicam que a metodologia proposta foi superior às metodologias tradicionais em termos de precisão experimental e eliminação ou redução da dependência entre os resíduos.

Palavras-chave: melhoramento genético; dependência espacial; precisão.

¹ Discente do Programa de Pós Graduação em Estatística Aplicada e Biometria, Viçosa – rodrigo.c.nunes@ufv.br

² Professor(a) do Programa de Pós Graduação em Estatística Aplicada e Biometria, Viçosa – policarpo@ufv.br



Métodos de validação de testes para análise de sementes – um aplicativo Shiny

Gisela Hélvia Nunes Chipenete ¹

Renato Ribeiro de Lima ²

Resumo

A qualidade da semente é muito importante para a agricultura, interferindo na produtividade, sustentabilidade, resistência a estresses e eficiência no uso de insumos. Além dos avanços científicos, melhoramento genético e biotecnologia e boas práticas agrícolas, o rigoroso processo de certificação está entre os principais fatores que asseguram tal qualidade. O Sistema de Certificação estabelece padrões de qualidade para germinação, pureza, identidade genética, sendo que a fiscalização, que utiliza testes específicos, garante que as sementes comercializadas atendam a padrões mínimos de qualidade. Uma falha no teste pode causar prejuízo ilegítimo como perda financeira para o produtor, impedimento de comercialização, risco ambiental e fitossanitário, multas e penalizações. Por isso, é de suma importância determinar o resultado correto ao aplicar esses testes. Assim, é necessário que se faça a validação do método de teste, garantindo a precisão, reproduzibilidade e repetibilidade deste. A automatização de procedimentos rotineiros resulta em maior eficiência e economia de tempo, bem como reduz riscos associados a erros humanos. Nesse intuito, foi desenvolvido o ValidMeth, um aplicativo web que realiza a análise estatística do Método ISTA de Validação de teste para análise de sementes. Este permite o uso de estatísticas automatizadas, gráficos interativos e relatórios personalizáveis, reduzindo erros e garantindo conformidade com os padrões ISTA. Os analistas e pesquisadores envolvidos com a validação de metodologias para análise de sementes podem se beneficiar desse aplicativo, promovendo maior padronização dos procedimentos, agilidade na obtenção e interpretação dos resultados, além de assegurar maior confiabilidade estatística no processo de validação.

Palavras-chave: estatística de validação; germinação; reproduzibilidade; repetibilidade; testes colaborativos.

¹ Departamento de Estatística, DES/UFLA – gisela.chipenete@estudante.ufla.br

² Departamento de Estatística, DES/UFLA – rrlima@ufla.br



Comparação entre CART e RF como métodos de imputação em dados de pecuária de precisão

Vivian Aparecida Brancaglioni ¹; Welinton Yoshio Hirai ²; Danilo Sarti ³; Adriele Giaretta Biase ⁴; Carlos Tadeu dos Santos Dias ⁵

Resumo

Estudos com a presença de observações ausentes são muito comuns em diversas áreas do conhecimento, e com dados obtidos a partir da pecuária da precisão não seria diferente. Dados de pecuária de precisão auxiliam o setor agropecuário a acompanhar, mapear e identificar problemas e buscar soluções. O conjunto de dados utilizado neste trabalho provém da pecuária de precisão, no qual pode-se acompanhar a oscilação de peso de 38 animais, das raças Nelore e Cruzado Britânico, divididos entre macho inteiro e macho castrado. Esses dados foram coletados a partir de um sistema de plataforma de pesagem automática. No entanto, durante as pesagens algumas informações de peso foram perdidas e o objetivo deste trabalho foi comparar o desempenho de dois métodos de imputação de dados da classe MICE, implementados no software R por meio do pacote *mice*: árvore de classificação e regressão (CART) e floresta aleatória (RF). Esses métodos foram comparados por meio de quatro critérios, raiz do erro quadrático médio (RMSE), pelo coeficiente de correlação de Pearson, índice de acurácia de Willmott e índice de desempenho. Os valores faltantes foram removidos do conjunto de dados original, obtendo-se um conjunto completo; e a partir dele foram criados bancos com diferentes porcentagens de dados faltantes, 5% e 15%, removidos aleatoriamente. Quando comparados os métodos, fixando-se os cenários, foi possível observar que o método de árvore de classificação e regressão teve melhor desempenho.

Palavras-chave: Imputação Múltipla; *Machine learning*; Reamostragem; Pecuária; MICE.

¹Departamento de Ciências Exatas, ESALQ/USP - vivian.brancaglioni@gmail.com

²Núcleo de Epidemiologia e Bioestatística, Hospital de Amor - welinton.hirai@hospitaldeamor.com.br

³Hamilton Institute Math and Stats - University of Ireland Maynooth - daniloasarti@gmail.com

⁴Centro de Ciências Exatas e Tecnológicas, UFAC - adriele.biase@ufac.br

⁵Departamento de Ciências Exatas, ESALQ/USP - ctsdias@usp.br



Modelagem de Extremos Meteorológicos via GEV e GPD: Uma Análise Comparativa em Areado-MG

Valdeline de Paula Mequelino Ferreira;¹ Maria Vitória Neves;² Louiziane Ribeiro Carvalho;³ Luiz Otávio de Oliveira Pala;⁴ Thelma Safadi⁵

Resumo

A modelagem dos extremos de temperatura máxima em Areado-MG foi baseada em séries mensais dos máximos anuais, extraídas dos dados diários da Estação Meteorológica de Alfenas-MG, disponibilizados pelo Instituto Nacional de Meteorologia (INMET), no período de 1980 a 2024. Utilizou-se a abordagem clássica de máxima verossimilhança para ajustar dois modelos da Teoria de Valores Extremos: o método de *block maxima*, com a distribuição Generalizada de Valores Extremos (GEV), e o método *peaks-over-threshold* (POT), com a distribuição Generalizada de Pareto (GPD). O limiar u do modelo POT foi definido com base no gráfico da média da função de excesso, que indicou linearidade a partir de aproximadamente 36,8°C, validando o uso da GPD para os excessos. As estimativas dos parâmetros foram obtidas via software R. A avaliação preditiva foi feita por meio do Desvio Absoluto Médio de Predição (MADP) e da Raiz do Erro Quadrático Médio (RMSE). Estimaram-se níveis de retorno para períodos de 2, 5 e 10 anos. Para retorno de 2 anos, obteve-se uma temperatura de 37,5°C, indicando que esse valor tende a ser superado, em média, a cada dois anos — um limiar relevante para o setor agrícola. Os maiores valores previstos ocorreram entre dezembro e fevereiro, meses de maior temperatura na região, o que demanda atenção especial. Conclui-se que o modelo POT com distribuição GPD é o mais adequado para prever extremos de temperatura máxima em Areado-MG.

Palavras-chave: Temperatura máxima; Areado-MG; Teoria dos Valores Extremos; Distribuição GPD; Tempo de retorno; Modelagem estatística clássica.

¹Departamento de Estatística e Experimentação Agrícola, Universidade Federal de Lavras (UFLA), Lavras-MG - valdeline.ferreira1@estudante.ufla.br

²Departamento de Estatística e Experimentação Agrícola, Universidade Federal de Lavras (UFLA), Lavras-MG - Maria.neves5@estudante.ufla.br

³Departamento de Estatística e Experimentação Agrícola, Universidade Federal de Lavras (UFLA), Lavras-MG - louiziane.carvalho2@estudante.ufla.br

⁴Departamento de Estatística e Experimentação Agrícola, Universidade Federal de Lavras (UFLA), Lavras-MG - luizpala@ufla.br

⁵Departamento de Estatística e Experimentação Agrícola, Universidade Federal de Lavras (UFLA), Lavras-MG - safadi@ufla.br



Análise temporal da área plantada em região semiárida do Nordeste Brasileiro, Brasil

Marilia Gabriela Medeiros Ordomo Lages¹ Patricia Arruda de Moura² Arundo Nunes da Silva Júnior³ Elisa Lopes do Nascimento⁴ Tiago Alessandro Espinola Ferreira⁵ Lucian Bogdan Bejan⁶ Tatijana Stosic⁷ Maria de Fátima Neves Cabral⁸ Victor Casimiro Piscoya⁹ Renisson Neponuceno de Araújo Filho¹⁰ Moacyr Cunha Filho¹¹

Resumo

Este trabalho teve como objetivo analisar a evolução da área plantada de milho e feijão no Nordeste entre 2016 e 2023. Essas culturas têm papel fundamental na segurança alimentar regional e na geração de renda para a agricultura familiar, sendo o feijão cultivado majoritariamente em pequenas propriedades no Nordeste e o milho representando parcela expressiva da produção agrícola familiar, conforme dados da Produção Agrícola Municipal (PAM) e do Censo Agropecuário do IBGE. A partir desses dados, aplicaram-se o teste de Mann–Whitney para comparação das áreas plantadas e a regressão para avaliação das tendências temporais. Adicionalmente, realizou-se o teste de Shapiro–Wilk para verificar a normalidade dos resíduos dos modelos. O teste de Mann–Whitney indicou diferença significativa entre as áreas plantadas de milho e feijão ($p = 0,0002$). A regressão revelou tendência de crescimento para o milho, com incremento médio anual de 69.259 hectares ($p = 0,0071$; $R^2 = 72,8\%$), e de redução para o feijão, com decréscimo médio anual de 34.979 hectares ($p = 0,0033$; $R^2 = 78,6\%$). Os resíduos dos modelos de regressão apresentaram distribuição normal para ambas as culturas ($p > 0,05$), validando a adequação dos modelos aplicados. Esses resultados indicam alteração progressiva na dinâmica agrícola do Nordeste, caracterizada pela expansão da cultura do milho e retração do cultivo de feijão. Dessa forma, este estudo destaca a necessidade de ações à valorização de culturas tradicionais e fortalecimento da agricultura familiar, contribuindo para formulação de políticas públicas que assegurem a produção diversificada e sustentável no semiárido nordestino.

Palavras-chave: dinâmica agrícola; Nordeste brasileiro; regressão linear; teste de Mann–Whitney; tendências temporais.

¹ Departamento de Estatística e Informática, UFRPE – marilia.lages@ufrpe.br

² Departamento de Estatística e Informática, UFRPE – patricia.amoura@ufrpe.br

³ Departamento de Estatística e Informática, UFRPE – arundo.nunes@ufrpe.br

⁴ Departamento de Estatística e Informática, UFRPE – elisa.lopes@ufrpe.br

⁵ Departamento de Estatística e Informática, UFRPE – tiago.espinola@ufrpe.br

⁶ Departamento de Estatística e Informática, UFRPE – lucianbb@gmail.com

⁷ Departamento de Estatística e Informática, UFRPE – tastosic@gmail.com

⁸ Campus Abreu e Lima, IFPE – fatimanevesc@bol.com

⁹ Departamento de Tecnologia Rural, UFRPE – victor.piscoya@ufrpe.br

¹⁰ Departamento de Tecnologia Rural, UFRPE – renisson.neponuceno@ufrpe.br

¹¹ Departamento de Estatística e Informática, UFRPE – moacyr2006@gmail.com



Influência da Precipitação Pluviométrica no Rendimento da Produção de Mandioca no Agreste Pernambucano: Uma Análise Correlacional

Natália Moraes Cordeiro ¹

Maria Marciele de Lima Silva ²

Moacyr Cunha Filho ³

Resumo

O semiárido brasileiro caracteriza-se por uma configuração climática particular, marcada pela escassez de água, resultante da irregularidade na distribuição das precipitações pluviométricas, além de longos períodos de estiagem e secas severas. Tal condição impõe desafios significativos à produção agrícola na região, a qual está altamente dependente das condições climáticas e, por conseguinte, sujeita às variações na precipitação. Este é o caso específico da produção de mandioca no Nordeste do Brasil. Entre os diversos cultivos agrícolas no país, a mandioca destaca-se como uma das culturas mais cultivadas no Agreste pernambucano, representando uma parcela significativa da economia nacional. Nesse contexto, o presente estudo teve como objetivo avaliar a influência da precipitação pluviométrica no rendimento médio da produção de mandioca na região do agreste de Pernambuco. Para tanto, foram realizadas análises de correlação linear entre os dados de precipitação e o rendimento médio anual da cultura de mandioca, utilizando o coeficiente de correlação de Pearson. As informações necessárias para a análise foram obtidas a partir dos bancos de dados da Embrapa e da Agência Pernambucana de Águas e Clima (APAC), no período compreendido entre 2014 e 2024. Os resultados obtidos indicaram uma correlação moderada entre a precipitação e o rendimento da mandioca, sugerindo que a variável pluviométrica exerce uma influência substancial sobre a produtividade da cultura.

Palavras-chave: Correlação de Pearson; Precipitação pluviométrica; Produção de mandioca; Rendimento médio.

¹ Departamento de Estatística e Informática, UFRPE – Recife/PE – nataliamcorddeiro@gmail.com

² Departamento de Estatística e Informática, UFRPE – marciele.liima@gmail.com

³ Departamento de Estatística e Informática, UFRPE – moacyr.cunhafo@ufrpe.br