

11ª Semana da Estatística

21-24 de Outubro de 2019

UFF, Niterói, Rio de Janeiro, Brasil

Anais do Evento

Departamento de Estatística
Instituto de Matemática e Estatística
Universidade Federal Fluminense

ISBN: 978-85-94029-29-4

Sumário

Sobre	5
11ª Semana da Estatística	5
Departamento de Estatística	5
Comissão Organizadora	5
Comissão Científica	5
Programação Geral	6
Segunda-feira, 20/10/2019 - Saúde	6
Terça-feira, 21/10/2019 - Ciência de Dados	6
Quarta-feira, 22/10/2019 - Educação Estatística	7
Quinta-feira, 24/10/2019 - Finanças	7
Lista de Resumos	8
Palestras	8
Como ser um estatístico aplicado	8
Índice de fragilidade do idoso brasileiro: uma abordagem multidimensional	8
Modelo de regressão para dados longitudinais: estudo do perímetro cefálico de crianças expostas ao vírus Zika	9
Estatística: O motor do Data Science	9
Testes de hipótese em um cenário de big data	9
Genealogia da Inteligência Artificial	10
Modelagem estatística para investigação de fatores associados ao desempenho escolar	10
Educação Estatística: desafios e perspectivas	11
Microestrutura de Mercado e Finanças Comportamentais no Brasil	11
Sistema Financeiro Internacional: Mercado de Capitais, Crises Econômicas e Contágio	11
Minicursos	12
Desenhos de estudo na área da saúde: de que forma o estatístico pode atuar?!	12
Conhecendo e manipulando os Sistemas de Informação em Saúde usando o R	12
Introdução à Ciência de Dados: teoria e prática	12
Uma breve visão sobre a integração das linguagens R e Python	13
Introdução das ideias poderosas da Estatística via atividades “mão na massa”	13
Métodos Ativos de aprendizagem de Estatística utilizando o Programa R	13
Uma introdução da Teoria de Valores Extremos (TVE) com aplicações em finanças	14
Modelagem de series financeiras utilizando o R-Project	14
Sessão oral	15
Análise espaço-temporal de indicadores para a Dengue e de qualidade habitacional: Um estudo sobre o Rio de Janeiro	15
Estimação do preço hedônico de imóveis no município do Rio de Janeiro a partir de uma Análise de Regressão	15
Óleo de linhaça, ofertado no período pós-lactação, atua no fêmur e na adiposidade	16
Previsão de Churn – Estudo de Caso no QConcursos.com	17
Random Forest para Classificação de Dados Textuais	17
Testando empates não revelados entre pilotos da F1	17

Pôster	19
Aplicação do modelo de Markowitz na otimização de carteiras de investimento de risco	19
Avaliação de modelos espaciais para dados de contagem: um estudo sobre a incidência de tuberculose no Estado do Rio de Janeiro	19
Cointegração no Mercado Financeiro	20
Fatores associados à depressão em universitários de Niterói	20
Índice de segurança mineiro: indicador composto para visualização espacial da segurança no estado de Minas Gerais	21
Medidas indiretas para o número de armas no Estado do Rio de Janeiro: uma abordagem com efeitos espaciais correlacionados	21
Modelo de tomada de decisão para aumento de resiliência à desastres em comunidades da baixada Fluminense: Uma análise para orientação e direcionamento de esforços dos órgãos públicos.	22
Perfil dos municípios brasileiros que votaram nos candidatos à presidência no segundo turno das eleições de 2018	22
R Ladies Niterói: Ensino da Linguagem R	23
Uma análise da qualidade de recomendações via Filtragem Colaborativa e Regressão Logística	23
Instituições parceiras e patrocinadores	24
Instituições parceiras	24
Patrocinadores	24

11ª Semana da Estatística

A Semana da Estatística (SEMEST) é um evento que ocorre dentro da Agenda Acadêmica da Universidade Federal Fluminense (UFF) com palestras e minicursos, abordando diferentes áreas de aplicação da Estatística. Em 2019, a SEMEST foi organizada em dias temáticos. Os temas escolhidos para essa edição foram: Saúde, Ciência de Dados, Educação Estatística e Finanças. O evento tem o objetivo de criar um ambiente em que discentes e docentes, da UFF e de outras instituições, possam interagir de forma a ampliar e complementar experiências acadêmicas e profissionais na área de Estatística. Uma sessão pôster e uma sessão oral proporcionam oportunidades para os participantes apresentarem trabalhos realizados.

Departamento de Estatística

O Departamento de Estatística (GET), que está situado no Instituto de Matemática e Estatística da Universidade Federal Fluminense, é o responsável pela organização da 11ª Semana da Estatística.

Comissão Organizadora

Jessica Quintanilha Kubrusly - GET/UFF
Jony Arrais Pinto Junior - GET/UFF
Mariana Albi de Oliveira Souza - GET/UFF
Patrícia Lusié Velozo da Costa - GET/UFF

Comissão Científica

Ana Beatriz Monteiro Fonseca	Hugo Henrique Kegler dos Santos	Jessica Quintanilha Kubrusly
Keila Mara Cassiano	Ludmilla da Silva Viana Jacobson	Luis Guillermo Coca Velarde
Patrícia Lusié Velozo da Costa	Marcio Watanabe Alves de Souza	Marco Aurélio dos Santos Sanfins

Programação Geral

PL: Palestra, MI: Minicurso.

Segunda-feira, 20/10/2019 - Saúde

9:00–9:30		Credenciamento	
9:30–10:20	PL	Dani Gamerman UFRJ	Como ser um estatístico aplicado
10:20–11:10	PL	Mariá da Silva Estácio	Índice de fragilidade do idoso brasileiro: uma abordagem multidimensional
11:10–12:00	PL	Luis Guillermo Velarde UFF	Modelo de regressão para dados longitudinais: estudo do perímetro cefálico de crianças expostas ao vírus Zika
12:45–14:00		Almoço	
14:00–17:00	MI	Leda da Silva UFF	Desenhos de estudo na área da saúde: de que forma o estatístico pode atuar?!
14:00–17:00	MI	Beatriz de Oliveira FIOCRUZ	Conhecendo e manipulando os Sistemas de Informação em Saúde usando o R

Terça-feira, 21/10/2019 - Ciência de Dados

9:00-12:00	MI	Aline Paes, Daniel de Oliveira e Marcos Lage UFF	Introdução à Ciência de Dados: teoria e prática
9:00-12:00	MI	Bruno Lucian Costa Rede Globo	Uma breve visão sobre a integração das linguagens R e Python
10:45–11:10		Almoço	
9:30–10:20	PL	Thiago Marques EstaTiDados	Estatística: O motor do Data Science
10:45–11:10		Café	
10:20–11:10	PL	Dimas S. Lima Globo.com	Testes de hipótese em um cenário de big data
11:10–12:00	PL	Christian Aranha Centro de Pesquisas do Futuro	Genealogia da Inteligência Artificial
16:30–18:00		Monitoria	

Quarta-feira, 22/10/2019 - Educação Estatística

9:00-09:50	PL	Denise Britz ENCE	Modelagem estatística para investigação de fatores associados ao desempenho escolar
9:50-10:20		Café	
10:45-11:10	PL	Alexandre Silva UNIRIO	Educação Estatística: desafios e perspectivas
10:20-12:00		Sessões orais	
12:00-14:00		Almoço	
14:00-17:00	MI	Lisbeth Cordani USP	Introdução das ideias poderosas da Estatística via atividades "mão na massa"
14:00-17:00	MI	Alexandre Silva UNIRIO	Métodos Ativos de aprendizagem de Estatística utilizando o Programa R

Quinta-feira, 24/10/2019 - Finanças

9:00-12:00	MI	Marcos Sanfins/Daiane Oliveira UFF/PUC	Uma introdução da Teoria de Valores Extremos (TVE) com aplicações em finanças
9:00-12:00	MI	Daiane Santos Cândido Mendes	Modelagem de series financeiras utilizando o R-Project
12:00-14:00		Almoço	
14:00-14:50	PL	Eduardo Camilo UFF	Microestrutura de Mercado e Finanças Comportamentais no Brasil.
14:50-15:40	PL	Dimas S. Lima Globo.com	Testes de hipótese em um cenário de big data
15:40-17:00		Sessão Pôster e Coquetel de encerramento	

Lista de Resumos

Palestras

Segunda-feira, 21/10/2019

Como ser um estatístico aplicado

Dani Gamerman

Universidade Federal do Rio de Janeiro

A palestra visa ilustrar formas de atuação de um estatístico aplicado, ilustrando a indissociabilidade em Estatística entre teoria e prática. Usando técnicas modernas de Séries Temporais, Estatística Espacial, Análise Multivariada e Teoria de Valores Extremos, a palestra apresentará aplicações em diversas áreas da Ciência. Através desses exemplos, vai se procurar mostrar como um estatístico pode e deve atuar. A apresentação se concentrará nos resultados obtidos nas análises, também apresentando a teoria desenvolvida para elas.

Índice de fragilidade do idoso brasileiro: uma abordagem multidimensional

Mariá Gonçalves Pereira da Silva

Universidade Estácio de Sá

O objetivo deste trabalho foi construir um Índice de Fragilidade (IF) do idoso brasileiro usando dados de base populacional e caracterizar a fragilidade com uma abordagem multidimensional. Utilizando a Pesquisa Nacional de Saúde, foi calculado um IF de Referência (IF-R) que, a partir de sua associação com outros determinantes da fragilidade, originou o IF Brasileiro (IF-B), adaptado ao contexto brasileiro. Uma análise de múltiplos fatores foi aplicada aos determinantes do IF-B visando caracterizar a fragilidade em uma abordagem multidimensional, originando o IF Multidimensional (IF-M). Adicionalmente, os IFs foram estratificados em 4 grupos de fragilidade (não frágil, vulnerável, frágil e muito frágil) com base na literatura: e comparados quanto à concordância de classificação entre os grupos. Os três índices (IF-R, IF-B e IF-M) apresentaram melhor ajuste à distribuição gamma de probabilidades com médias de 0,13, 0,14 e 0,16, respectivamente. A concordância da classificação entre os grupos de fragilidade dada pelo IF-B com o IF-R foi de 79% e com o IF-M, de 76%. Quanto à presença de fragilidade, a concordância entre o IF-B e o IF-R foi de 94% e entre o IF-B e o IF-M, de 92%. Esses achados mostram uma boa concordância entre os três índices, no entanto, aponta a existência de outras dimensões que podem estar influenciando a caracterização da fragilidade, o que sugere a necessidade de continuidade da caracterização da fragilidade como fenômeno multidimensional.

Modelo de regressão para dados longitudinais: estudo do perímetro cefálico de crianças expostas ao vírus Zika

Luis Guillermo Coca Velarde

Universidade Federal Fluminense

Os modelos de regressão são comumente usados na área da saúde com o objetivo de quantificar o efeito de diversas variáveis consideradas causais sobre algumas variáveis de interesse, chamadas resposta. Uma das principais suposições destes modelos é a independência entre observações. Em diversas situações reais esta suposição não pode ser comprovada e, no caso de dados longitudinais, isto se deve ao fato de que as observações, ao longo do tempo, são naturalmente correlacionadas. Nestes casos é necessário considerar a estrutura de variância e covariância entre as observações de um indivíduo, o que pode ser feito usando um modelo exponencial de variograma. Este modelo está sendo usado para estudar a evolução do perímetro cefálico de crianças acompanhadas no Hospital Universitário Antônio Pedro através do projeto de pesquisa “Estudo clínico e epidemiologista de crianças expostas ao vírus Zika durante o período gestacional: estudo prospectivo de coorte”

Terça-feira, 22/10/2019

Estatística: O motor do Data Science

Thiago Marques

EstaTiDados

Os modelos de regressão são comumente usados na área da saúde com o objetivo de quantificar o efeito de diversas variáveis consideradas causais sobre algumas variáveis de interesse, chamadas resposta. Uma das principais suposições destes modelos é a independência entre observações. Em diversas situações reais esta suposição não pode ser comprovada e, no caso de dados longitudinais, isto se deve ao fato de que as observações, ao longo do tempo, são naturalmente correlacionadas. Nestes casos é necessário considerar a estrutura de variância e covariância entre as observações de um indivíduo, o que pode ser feito usando um modelo exponencial de variograma. Este modelo está sendo usado para estudar a evolução do perímetro cefálico de crianças acompanhadas no Hospital Universitário Antônio Pedro através do projeto de pesquisa “Estudo clínico e epidemiologista de crianças expostas ao vírus Zika durante o período gestacional: estudo prospectivo de coorte”

Testes de hipótese em um cenário de big data

Dimas Soares Lima

Globo.com

Uma das metodologias mais utilizadas em ciência de dados é o teste A/B. Testes A/B são experimentos randomizados, muito utilizados para auxiliar a tomada de decisão, por exemplo, quando não se sabe qual é a melhor opção de uma página de um site é melhor para se ofertar para um usuário. Em um cenário de big data, as metodologias de teste hipótese clássicos utilizadas para analisar os resultados

dos testes, acabam declarando qualquer efeito como uma diferença significativa. Ainda não existe uma resposta certa para contornar esse problema, na globo.com estamos testando algumas formas de contornar esse desafio e chegar a resultados válidos, que façam sentido para empresa e para nossos usuários.

Genealogia da Inteligência Artificial

Christian Aranha

Rede Entropia – Centro de Pesquisas do Futuro

Nesta palestra vamos abordar a história da Inteligência Artificial. Porque cresce tanto hoje. Porque causa tanto impacto. Qual a diferença entre IA, Machine Learning, Estatística e Big Data. Em que ponto estamos do seu desenvolvimento. Qual o impacto no mercado de trabalho. Como nos preparar para isso. Quais serão as habilidades que nos diferenciarão delas no futuro.

Quarta-feira, 23/10/2019

Modelagem estatística para investigação de fatores associados ao desempenho escolar

Denise Britz do Nascimento Silva

Escola Nacional de Ciências Estatísticas - ENCE

Os fatores associados ao desempenho escolar têm sido estudados há algumas décadas no cenário nacional. A disponibilidade de informações provenientes de avaliações externas em larga escala no Brasil impulsionou o desenvolvimento de estudos sobre a qualidade da educação e o desempenho escolar. Os fatores que influenciam a proficiência do aluno são múltiplos e complexos, com dinâmicas que atuam em diferentes níveis, desde características socioeconômicas e culturais dos alunos e de suas famílias, até as que ocorrem nas salas de aula, entre professores e alunos, e incluem também as características estruturais da escola. Sendo assim, a análise do fenômeno requer o emprego de modelos multiníveis adequados para a estrutura hierárquica dos dados. Esta palestra tem como foco a apresentação de trabalhos desenvolvidos na ENCE para investigação dos diferentes fatores que impactam o desempenho dos alunos da rede pública brasileira medido pelos sistemas de avaliação utilizados pelo INEP/MEC. Esses estudos identificam variáveis associadas com o desempenho escolar, seguindo a literatura especializada sobre o tema, mas também investigam elementos pouco abordados como composição da família, efeito dos pares e composição das turmas, ou incluem comparações longitudinais da proficiência.

Educação Estatística: desafios e perspectivas

Alexandre Sousa da Silva

Universidade do Estado do Rio de Janeiro - UNIRIO

A Estatística é uma ciência cada vez mais presente no nosso dia a dia. Logo, conhecimentos em Estatística têm sido uma exigência tanto na formação técnica quanto na formação de cidadãos críticos. E assim como toda a ciência, a Estatística é dinâmica e exige atualização constante. Entretanto, será que a forma de ensinar/aprender Estatística acompanhou o desenvolvimento da sociedade? Nesta palestra serão apresentados os principais resultados de pesquisas científicas em Educação Estatística, tendo como objetivo discutir ações e atividades que permitirão trabalhar os principais conceitos Estatísticos no Ensino Fundamental, Médio e Superior, utilizando metodologias ativas de ensino/aprendizagem por meio de programas computacionais. Vamos conversar sobre Educação Estatística!??

Quinta-feira, 24/10/2019

Microestrutura de Mercado e Finanças Comportamentais no Brasil

Eduardo Camilo

Universidade Federal Fluminense - UFF

O estudo da microestrutura do mercado de ações utiliza dados de alta frequência (intradiários) e permite a identificação de indicadores que não podem ser medidos de outro modo. Indicadores que medem humor do mercado, presença de insiders e ocorrência de efeito manada são muito úteis tanto para os investidores quanto para os órgãos reguladores.

Sistema Financeiro Internacional: Mercado de Capitais, Crises Econômicas e Contágio

Tiago Costa Ribeiro

FAPES-BNDES

Arquitetura do Sistema Financeiro e comunicação com o resto do mundo. Agentes econômicos formais, fluxo financeiro e comercial. Análise da disputa tarifária EUA x China, eleições argentinas e seus impactos para a economia doméstica. Mercados de capitais em países desenvolvidos, principais agentes. Análise comparativa de risco soberano e seu impacto na formação de preço dos ativos.

Minicursos

Segunda-feira, 21/10/2019

Desenhos de estudo na área da saúde: de que forma o estatístico pode atuar?!

Leda Maria Ferraz da Silva

Universidade Federal Fluminense

O conceito de desenho de estudo envolve a identificação do tipo de abordagem metodológica que se utiliza para responder a uma determinada questão, implicando, assim, a definição de certas características básicas do estudo, por exemplo, como será a população e a amostra a ser estudada, a unidade de análise, a existência ou não de intervenção direta sobre a exposição, a existência e tipo de seguimento dos indivíduos, dentre outras. Tendo como base as características básicas do estudo criaram-se uma série de padrões terminológicos que definem algumas dessas características, e que constituem aquilo que se designa como tipos ou desenhos de estudo. Exemplos de desenhos de estudo frequentemente encontrados na área da saúde são: os ensaios clínicos, os estudos de coorte, os estudos de casos e controles, os estudos transversais, observacionais, entre outros. Porém, para se ter uma correta avaliação dos resultados, os dados coletados precisam passar por uma análise estatística. Com base na utilização de testes e modelos estatísticos, possibilita-se a tomada de decisões de forma a minimizar a incerteza decorrente do uso de uma amostra. Assim, o método estatístico passa a se caracterizar por razoável grau de precisão, o que o torna bastante aceito por parte dos pesquisadores com preocupações de ordem quantitativa. Nesse sentido, os procedimentos estatísticos fornecem considerável reforço às conclusões obtidas, sobretudo, mediante a experimentação e a observação.

Conhecendo e manipulando os Sistemas de Informação em Saúde usando o R

Beatriz Fátima Alves de Oliveira

Fiocruz

Terça-feira, 22/10/2019

Introdução à Ciência de Dados: teoria e prática

Aline Paes, Daniel de Oliveira e Marcos Lage

Universidade Federal Fluminense

Ciência de Dados é uma linha de pesquisa multidisciplinar que tem ganhado relevância nos últimos anos por conta da facilidade de acesso e da enorme quantidade de dados produzidos nos mais diversos contextos como por exemplo as grandes cidades, aplicações web, estudos científicos, entre outros. Neste contexto as disciplinas de Gerência de Dados, Aprendizado de Máquina e Visualização formam o tripé fundamental para a construção de aplicações computacionais capazes de organizar, extrair e comunicar as informações contidas em grandes bases de dados. Neste curso daremos uma visão geral

destas três disciplinas e usaremos ferramentas open source desenvolvidas em Python para construir softwares que implementem de forma simples os conceitos discutidos.

Uma breve visão sobre a integração das linguagens R e Python

Bruno Lucian Costa

Rede Globo

O minicurso irá abordar a integração entre R e Python em um fluxo de análise comum no dia a dia do cientista de dados, utilizando os pacotes `reticulate` e `feather` para realizar essa interoperabilidade entre as linguagens. O fluxo de análise utilizará do `reticulate` a possibilidade de importar módulos e scripts Python numa sessão do R, de modo a prosseguir a análise na linguagem. Além disso será apresentado o pacote `feather`, que proporciona rapidez de leitura e gravação de objetos e é codificado independente da linguagem. Ao expor esses pacotes será possível utilizar as duas linguagens em prol de uma boa análise de dados, nos beneficiando dos pontos fortes do R e do Python. Vamos elaborar um processo estruturado para data cleaning, data processing e data visualization, utilizando como base um conjunto de dados anonimizado de e-commerce brasileiro, disponibilizado pelo olist que pode ser encontrado no kaggle. *mini-cv*: Graduado em Estatística pela UFF e Mestre em Matemática Aplicada pela FGV. Atualmente é Cientista de dados líder na Rede Globo trabalhando com modelos de series temporais, classificação e clusterização de usuários, com passagens pelo IBOPE DTM onde também atuou como líder dos cientistas de dados liderando a área responsável pelos sistemas de recomendação. Gerador de conteúdo para o `dadosaleatorios.com.br` por mais de 4 anos. Possui interesse em modelagem preditiva, sistemas de recomendação, machine learning, aplicações web e visualização de dados.

Quarta-feira, 23/10/2019

Introdução das ideias poderosas da Estatística via atividades “mão na massa”

Lisbeth Kaiserlian Cordani

Universidade de São Paulo

Dado o crescimento da disponibilidade de dados em todas as áreas do conhecimento, desde os últimos anos do século passado o ensino da estatística, antes só oferecido no curso universitário (com algumas exceções de escolas técnicas), vem sendo considerado como parte integrante da disciplina de matemática na escola básica. Não se pode pretender que esse conhecimento seja restrito aos seguidores das áreas de exatas, mas sim que seja parte integrante da cultura básica de todos os cidadãos. Neste minicurso será apresentada uma forma ativa do ensino da estatística, não meramente instrumental, através de atividades que desenvolvam o espírito crítico para tomar decisões na presença de variabilidade e incerteza, que favorecerá a integração dos alunos em projetos e atividades interdisciplinares.

Métodos Ativos de aprendizagem de Estatística utilizando o Programa R

Alexandre Sousa da Silva

Universidade do Estado do Rio de Janeiro

A Estatística é um componente curricular presente em quase todos os cursos de graduação. O uso crescente deste conhecimento é motivado pela necessidade de gerar informações a partir de grandes volumes de dados disponíveis em nossa sociedade cada vez mais informatizada. A Estatística, assim como toda a ciência é dinâmica e precisa ser atualizada constantemente, entretanto são tímidas as mudanças no processo de ensino/aprendizagem de Estatística em todos os níveis de ensino. Neste minicurso serão apresentadas e discutidas metodologias ativas de ensino, utilizando o programa computacional R. O programa R é um poderoso ambiente de desenvolvimento, livre e de código aberto. Possui uma comunidade de usuários extremamente ativa e colaborativa, dentre suas interfaces destaca-se o Rcmdr por ser bastante intuitivo e o Rstudio por ser muito versátil e multifuncional. As metodologias ativas proporcionam o desenvolvimento de habilidade de raciocínio e a autonomia, além do pensamento crítico. Assim os estudantes são estimulados a tornarem-se responsáveis pelo processo de aprendizagem de Estatística.

Quinta-feira, 24/10/2019

Uma introdução da Teoria de Valores Extremos (TVE) com aplicações em finanças

Marco Aurélio dos Santos Sanfins¹ e Daiane de Souza Oliveira²

¹ Universidade Federal Fluminense

² Pontifícia Universidade Católica

A Teoria de Valores Extremos é um ramo da probabilidade que estuda o comportamento estocástico dos valores extremos associados a um conjunto de variáveis aleatórias. No minicurso será apresentado uma breve introdução sobre a teoria de valores extremos, bem como algumas aplicações em finanças, realizadas através do software R-project.

Modelagem de series financeiras utilizando o R-Project

Daiane Rodrigues dos Santos

Universidade Cândido Mendes

O objetivo do minicurso é ensinar o aluno a utilizar o R-Project com a finalidade de modelar séries financeiras como preço de ações, taxa de câmbio e Juros. O curso está dividido em três partes. A primeira contém a apresentação dos modelos ARMA (Modelos Auto Regressivos de Médias Móveis), a segunda a coleta da série financeira que será modelada pelo aluno e a terceira parte é a criação de um script com os modelos ARMA.

Sessão oral

Análise espaço-temporal de indicadores para a Dengue e de qualidade habitacional: Um estudo sobre o Rio de Janeiro

Brunno Costa¹, Lucas Moura¹, Maquise Pinheiro¹, Ricardo Junqueira¹, Rodolfo Hauret¹

¹Universidade Federal Fluminense

O Estado do Rio de Janeiro sofre com epidemias recorrentes de Dengue, o desenvolvimento destas epidemias é condicionado por diversos fatores ambientais e socioeconômicos. Para verificar a relação entre os fatores condicionantes e a ocorrência de doença foram criados dois indicadores compostos, um de Morbimortalidade da Dengue e um de Qualidade de Habitação. Os indicadores foram calculados para dados dos municípios do Estado do Rio de Janeiro para o período de 2007 a 2013 e então verificou-se a existência de um padrão espacial em sua distribuição nos municípios e também a correlação entre ambos.

Palavras-chave: Epidemiologia, Dengue, Indicador, Estatística Espacial.

Email de contato: brunnocosta@id.u.br, lucasmfes@id.u.br, maquise@gmail.com, rjunqueira@id.u.br, rodolfohauret@id.u.br.

Estimação do preço hedônico de imóveis no município do Rio de Janeiro a partir de uma Análise de Regressão

Thiago Marques¹, Marcos dos Santos² e Carlos Francisco Simões Gomes³

¹Escola Nacional de Ciências Estatísticas

²Instituto Militar de Engenharia

³Universidade Federal Fluminense

O mercado imobiliário do município do Rio de Janeiro, em especial os imóveis residenciais, nestes últimos anos sofreram grandes transformações devidas às obras de acessibilidade para apoiar a realização da Copa do Mundo em 2014 e as Olimpíadas em 2016. Destarte, este trabalho teve por objetivo desenvolver um modelo de regressão, para as áreas de planejamento do Município do Rio de Janeiro, avaliando as características que mais agregaram valor aos imóveis residenciais nos anos de 2010 e 2011, empregando a metodologia de precificação hedônica (PH), que possibilita identificar quais são as características que melhor representam o preço de um imóvel e qual o tipo de relacionamento entre os preços e estas características. Para isto, foi utilizada a base de dados de uma pesquisa realizada pelas imobiliárias, de imóveis vendidos. Esta pesquisa foi realizada a pedido da Caixa Econômica Federal do Rio de Janeiro, contendo dados referentes ao período de 2004 a 2012 e envolve todos os imóveis vendidos pelas imobiliárias, com ou sem o financiamento da Caixa Econômica Federal. Desta forma, para os três níveis usuais de significância, 1%, 5% e 10%, a forma funcional que melhor se adequou ao estudo foi a semi-logarítmica, apresentando um R² ajustado de aproximadamente 0,87 e as características que se mostraram mais relevantes para a estimação do preço de um imóvel, foram o índice de desenvolvimento social, o valor unitário do imóvel em reais por metros quadrados (R\$/m²), localização da residência por área de planejamento, Quantidade de domicílios com Sanitário e Esgotamento Sanitário por área de planejamento e as variáveis características dummies,

que correspondem ao número de banheiros na residência, número de quartos da residência e número de vagas de garagem na unidade residencial.

Palavras-chave: Modelos Hedônicos, Mercado Imobiliário, Estimação de preços imobiliários.

Email de contato: t_hiagomarques@hotmail.com, marcosdossantos_doutorado_u@yahoo.com.br, cfsg1@bol.com.br

Óleo de linhaça, ofertado no período pós-lactação, atua no fêmur e na adiposidade

Aline D'Avila Pereira¹, Danielle Ribeiro¹, Letícia Cardosos¹, Eduardo Silva¹, Maurício Chagas¹, Carlos Alberto Soares da Costa², Gilson Teles Boaventura¹ e Luis Guillermo Coca Velarde¹

¹Universidade Federal Fluminense

²Universidade Federal do Recôncavo da Bahia

O óleo de linhaça é fonte de ácido -linolênico que é importante para os estágios de crescimento e desenvolvimento corporal e atua no tecido adiposo (TA) e ósseo. Sendo assim, o objetivo foi avaliar a influência do óleo de linhaça, em diferentes fases da vida, sobre a adiposidade e estrutura óssea. Ao nascimento, houve a randomização em dois grupos: grupo controle (C) e grupo óleo de linhaça (OL). Aos 21 dias, os ratos foram desmamados e houve uma nova divisão: dos 18 animais do C, 9 continuaram recebendo C e os outros 9 passaram a receber OL; dos 18 animais do OL, 9 continuaram OL e os outros 9 passaram a receber C. Foi analisado a massa da ninhada; a massa e comprimento corporal; ingestão alimentar; composição corporal. Aos 67 dias, houve a eutanásia por punção cardíaca. No soro, foram analisados o perfil lipídico e hormônios. O TA intra- abdominal, inguinal (TAI), marrom (TAM) foram dissecados e analisados quanto à massa absoluta (MA) e relativa (MR). No fêmur (F) e na tíbia (T) direitos, foi analisado o comprimento, largura, MA, MR, DMO, parâmetros biomecânicos e radiodensidade da cabeça do F. No F esquerdo, foi mensurada a espessura cortical e área medular. A 4ª vertebra lombar foi analisada quanto à: MA, MR, altura e DMO. Foi utilizado o teste ANOVA de dois fatores, sendo o nível de significância de 0,05. O consumo de OL durante a lactação promoveu ($p < 0,05$): menor massa corporal dos 60 aos 67 dias; MA do TAI; área medular do F; largura e força máxima da T; e maior MA e MR do TAM; e radiodensidade da cabeça do F. No período pós-lactação, observou-se ($p < 0,05$): maior lipoproteína de alta densidade; MA e MR do TAM; largura, MA, DMO e força máxima no F; e força de ruptura da T; e menor colesterol total e lipoproteína de baixa densidade. Assim, o estudo demonstra que o óleo de linhaça no período pós-lactação pode atuar na prevenção de doenças crônicas.

Palavras-chave: Óleo de linhaça, Ratos, Osso, Adiposidade, Ômega-3.

Email de contato: alinedavila@id.u.br, guilleco@terra.com.br, gilsontb@gmail.com.

Previsão de Churn – Estudo de Caso no QConcursos.com

Aline Pinheiro Brandao¹, Arthur Vinicius Nascimento¹ e Rodrigo Dlugokenski¹

¹QConcursos.com

O QConcursos.com é uma das maiores plataformas de estudo do Brasil. Com um modelo baseado em assinaturas, a renovação do contrato é indispensável a sustentabilidade e manutenção de seus serviços. Este estudo teve como objetivo a classificação de clientes quanto ao risco de churn. Embora apresente bons resultados com amostras, o modelo descrito não se mostrou escalável quando em produção.

Palavras-chave: Machine Learning, Classificação, XGBoost, SaaS, Previsão de churn.

Email de contato: aline.pinheiro@qcx.com.br, arthur.nascimento@qcx.com.br, rodrigo@qcx.com.br.

Random Forest para Classificação de Dados Textuais

Ana Luiza Neves¹ e Jessica Kubrusly¹

¹Universidade Federal Fluminense

Este projeto une técnicas de mineração de texto ao método de classificação Random Forest. O banco usado é composto por textos redigidos por clientes que compraram peças de roupa pela internet, junto com uma variável binária que indica se o cliente faz uma recomendação positiva ou negativa da peça comprada. O objetivo central do estudo é identificar, a partir somente dos textos, que descrevem as peças compradas, quais as peças com recomendações positivas e quais as com recomendações negativas. Para tal, foram realizados procedimentos de pré-processamento textual a fim de transformar o banco textual em uma matriz numérica, a matriz termo-documento. Em seguida o método Random Forest foi aplicado na matriz termo-documento para realizar a classificação. Os resultados foram satisfatórios, indicando que a junção dos dois métodos foi capaz de acertar a classificação com acurácia, sensibilidade e especificidade maiores que 75%.

Palavras-chave: Mineração de Texto, Random Forest, E-commerce.

Email de contato: contato: ana_neves@id.u.br, jessicakubrusly@id.u.br.

Testando empates não revelados entre pilotos da F1

Rodrigo Dias¹ e Wilson Calmon¹

¹Universidade Federal Fluminense

Em nosso trabalho, buscamos testar formalmente a existência de empates não revelados em dados de ranqueamento, com base no modelo Plackett-Luce (PL). Dentro das suposições desse modelo, podemos calcular probabilidades associadas a ranks através de escalares positivos, cada qual associado a um objeto distinto. Apesar da forma da distribuição dos ranks ser bem conhecida, a estimação dos seus parâmetros não é uma tarefa fácil. De fato, Hunter (2004) propôs um algoritmo do tipo Minorization Maximization (MM) para estimar os parâmetros considerando um espaço paramétrico irrestrito. A

hipótese (nula) de que alguns dos indivíduos estão empatados pode ser escrita como uma restrição do espaço paramétrico e, portanto, pode ser testada através do Teste da Razão de Verossimilhança (TRV). Para efetuar tal teste, propusemos uma modificação do algoritmo MM que nos permite estimar o vetor de parâmetros admitindo restrições de igualdade entre quaisquer de suas componentes. Dessa forma, o algoritmo modificado tem, como caso particular, o próprio algoritmo do Hunter. Ambos os casos (restrito e irrestrito) foram implementados no software R e aplicados em um conjunto de dados referentes a corridas de Fórmula 1. Estimamos o parâmetro do modelo considerando o caso restrito em que pilotos de uma mesma equipe estão empatados (hipótese nula) e, ainda, o caso mais geral (irrestrito), em que nenhuma restrição foi imposta sobre o modelo de PL. As estimativas irrestritas nos permitem ranquear, um a um, os pilotos. Apesar de observarmos uma hierarquia nítida entre as equipes no ranking induzido pelas estimativas irrestritas, nós rejeitamos a hipótese de que a performance dos pilotos é unicamente determinada pelas equipes.

Palavras-chave: Dados de ranqueamento, Plackett-Luce, Empates não revelados, Algoritmo MM, Produção de Ranking estatístico dos Pilotos da Fórmula 1.

Email de contato: rodrigocaland@id.u .br, calmonwilson@gmail.com.

Referências

[1] Hunter, D. R. MM algorithms for generalized bradley-terry models, *The Annals of Statistics*, 2004.

Pôster

Aplicação do modelo de Markowitz na otimização de carteiras de investimento de risco

Tuany Esthefany Barcellos de Carvalho Silva¹, Marco Aurélio dos Santos Sanfins¹ e Daiane Rodrigues dos Santos²

¹Universidade Federal Fluminense

¹Universidade Veiga de Almeida

Os estudos de Markowitz (1952) foram à base para a Moderna Teoria de Carteiras, que apresenta a diversificação como principal instrumento para a redução do risco global de uma carteira (portfolio) de investimentos. A Moderna Teoria de Carteiras teve início com a publicação do artigo Portfolio Selection por Markowitz (1952). A utilização da diversificação como forma de redução do risco de uma carteira foi amplamente discutida e comprovada por meio de estudos sobre a correlação entre os ativos. A eficiência de uma carteira é relacionada pelo binômio risco e retorno, ou seja, o investidor pode sempre que desejar reduzir o risco de seus investimentos, alterando a alocação, com o intuito de manter o retorno desejado. Assim sendo, é necessário que carteiras sejam submetidas periodicamente ao monitoramento da performance e da composição dos ativos investidos. Para resolver tal problemática é de grande utilidade a aplicação de modelos matemáticos que ofereçam suporte às escolhas dos ativos e na definição de seus percentuais em uma carteira.

Palavras-chave: Markowitz, Portfólio, Otimização, Diversificação, Ativos.

Email de contato: tuanybarcellos@id.uff.br, marcosanfins@gmail.com, daianesantoseco@gmail.com.

Referências

[1] Markowitz, H. M. Portfolio Selection. [S.l.]: Journal of finance n. 1, v. 7, p.77-91, 1952.

Avaliação de modelos espaciais para dados de contagem: um estudo sobre a incidência de tuberculose no Estado do Rio de Janeiro

Ricardo Junqueira de Souza¹ e Jony Arrais Pinto Junior¹

¹Universidade Federal Fluminense

Dados de contagem são comumente encontrados quando trabalha-se com dados georreferenciados. Usualmente, estas contagens representam o número de ocorrências de um evento de interesse agregado por uma unidade geográfica (municípios, bairros, etc). Considerar as unidades geográficas independentes não parece ser uma suposição realista, pois espera-se que regiões próximas apresentem comportamentos similares, uma das premissas de Estatística Espacial. Este trabalho propôs o uso de três modelos para quantificar o impacto de variáveis como o índice de GINI, o índice de vulnerabilidade social (IVS), entre outras, no número de notificações de casos de tuberculose nos municípios do estado do Rio de Janeiro entre os anos de 2008 e 2012. Tratam-se de dois modelos hierárquicos bayesianos com efeitos espaciais que possuem distribuições *a priori* condicionais autoregressivas (CAR) e um

modelo que considera os efeitos espaciais independentes. Dentre os três modelos avaliados, o que apresentou melhor ajuste foi o modelo Leroux, que considera uma estrutura de dependência para os efeitos espaciais, indicando que entre as variáveis explicativas testadas os municípios com maiores desigualdades de renda apontadas pelo índice de Gini e maiores vulnerabilidades sociais possuem maiores médias de notificações de tuberculose.

Palavras-chave: Tuberculose, Dados de Área, Estatística Espacial, R, Leroux, CAR.

Email de contato: rjunqueira@id.u.br, jarrais@id.u.br.

Cointegração no Mercado Financeiro

Ítalo de Moraes¹, Daiane Oliveira², Soraida Aguilar² e Marco Aurélio dos Santos Sanfins¹

¹Universidade Federal Fluminense

²Pontifícia Universidade Católica

O objetivo deste trabalho é explorar a capacidade da cointegração no mercado financeiro para análise de dados que auxiliem na tomada de decisão dos agentes econômicos, procuramos então mensurar a relação entre duas variáveis como seu risco nas operações em um mercado complexo e globalizado onde o controle de risco de mercado necessitam ser cada dia mais explorados para se obter mais certeza na alocação de recursos eficiente.

Palavras-chave: Cointegração, mercado financeiro, risco de mercado, análise de dados, alocação de recursos.

Email de contato: italodolores@gmail.com, daiane.dso@gmail.com, soraguilar@gmail.com, marco-sanfins@id.u.br

Fatores associados à depressão em universitários de Niterói

Rafael Rangel Gemesio¹ e Luis Guillermo Coca Velarde¹

¹Universidade Federal Fluminense

A incidência de depressão é cada vez mais comum ao redor do mundo. Diante deste cenário, instrumentos que possam avaliar a ocorrência deste transtorno têm sido elaborados e aplicados em muitas culturas. O Inventário de Depressão de Beck (BDI – “Beck Depression Inventory”) é uma dessas ferramentas. Este estudo avaliou a condição de depressão em 357 universitários de Niterói/RJ e verificou a relação entre algumas variáveis com a ocorrência da doença.

Palavras-chave: Depressão, Jovens, Universitários, Niterói.

Email de contato: rafael_gemesio@hotmail.com, lgcocavelarde@id.u.br.

Índice de segurança mineiro: indicador composto para visualização espacial da segurança no estado de Minas Gerais

Luiz Fernando Coelho Passos¹, Lyncoln Sousa de Oliveira¹, Beatriz da Silva Mello¹, Júlia Oliveira Dias de Souza¹ e Ludmilla da Silva Viana Jacobson¹

¹Universidade Federal Fluminense

A segurança pública é um assunto muito discutido atualmente no Brasil. O objetivo deste trabalho é elaborar um indicador de segurança pública para os municípios do Estado de Minas Gerais, no período de 2013 a 2016. Foram utilizados os dados de segurança pública do Estado, assim como informações sobre o PIB, densidade demográfica e indicadores de saúde. A relação entre os indicadores simples foi avaliada por meio da matriz de correlação de Spearman. Ao final, o indicador composto incluiu os indicadores simples de assalto, estupro, homicídios, lesão corporal, internações e pib preço corrente entre os anos de 2013 a 2016. A densidade demográfica foi utilizada como um fator de ponderação. Os resultados mostraram que a violência no Estado é maior nos municípios da região noroeste, no entanto ao longo do tempo o indicador apontou redução da violência.

Palavras-chave: Indicador, Segurança Pública, Minas Gerais, R.

Email de contato: lfcpassos@id.u.br, lyncolnsousa@id.u.br, bsmello@id.u.br, juliaods@id.u.br, ludmillajacobson@id.u.br.

Medidas indiretas para o número de armas no Estado do Rio de Janeiro: uma abordagem com efeitos espaciais correlacionados

Lucas Moura Faria e Silva¹ e Jony Arrais Pinto Junior¹

¹Universidade Federal Fluminense

Existe um consenso na literatura que a proporção de suicídios por arma de fogo é a melhor medida indireta de prevalência de arma de fogo. Entretanto, esta proxy não é uma boa medida em locais com baixa densidade populacional, visto que suicídios são eventos raros. Recentemente, Cerqueira et al. (2018) [2] propuseram medidas indiretas que levam em consideração características pessoais das vítimas como novas alternativas de medidas indiretas para a prevalência de armas de fogo. Entretanto, os efeitos das microrregiões são considerados independentes nas proxies mais recentes, o que não parece apropriado dado o contexto espacial inerente. Este trabalho propôs modificações nessas medidas indiretas, incluindo diferentes estruturas de dependência espacial e usando uma abordagem completamente bayesiana. A medida foi ajustada utilizando dados de suicídios do Rio de Janeiro disponíveis no SIM - DATASUS entre os anos de 2000 e 2017.

Palavras-chave: Medidas Indiretas, Suicídios, Dependência espacial, Inferência Bayesiana.

Email de contato: lucasmfes@id.u.br, jarraais@id.u.br.

Modelo de tomada de decisão para aumento de resiliência à desastres em comunidades da baixada Fluminense: Uma análise para orientação e direcionamento de esforços dos órgãos públicos.

Pablo Berriel

¹Universidade Federal Fluminense

O aumento de desastres se dá em sua maioria ao crescimento urbano desordenado e segundo a Organização das Nações Unidas (ONU), que através da criação de muitos programas internacionais relacionados a redução do risco, afirmam que a redução dos riscos de desastres além de primordiais para a segurança, são de suma importância para o bem estar da sociedade. Em vista dos constantes desafios e demandas apresentados pela numerosa e carente população da Baixada Fluminense do Rio de Janeiro fazemos uso do objetivo da lei 12608/12, que instituiu a Política Nacional de Proteção e Defesa Civil e tem em sua razão de ser, o papel de promover o desenvolvimento de cidades resilientes e processos sustentáveis para a urbanização. Este estudo foi baseado na pesquisa de Junior e Santos (2018) e apresenta uma metodologia baseada no Modelo Fuzzy COPPE-COSENZA para realizar hierarquizar as necessidades de investimentos em aumento de resiliência entre 4 selecionadas comunidades da Baixada Fluminense. Para a aplicação de tal modelo, utilizaremos o software estatístico R e seus respectivos pacotes de algoritmos inerentes à solução do problema.

Palavras-chave: Desastres, Resiliência, Risco, Fuzzy, R.

Email de contato: pabloberriel90@gmail.com

Referências

[1] Junior, S. Modelo para obtenção do índice de resiliência a desastres em comunidades: uma análise para orientar os investimentos dos órgãos públicos. SBPO – Simpósio Brasileiro de pesquisa Operacional, Rio de Janeiro, 2018.

Perfil dos municípios brasileiros que votaram nos candidatos à presidência no segundo turno das eleições de 2018

Thalita Costa do Nascimento¹ e Jony Arrais Pinto Junior¹

¹Universidade Federal Fluminense

Muito se especula ainda sobre qual é o alinhamento de voto do eleitorado brasileiro nos dias atuais, quais motivações e circunstâncias que levam o eleitor a votar em determinado candidato em detrimento do outro. Nas últimas eleições presidenciais brasileiras, falou-se muito sobre aspectos que poderiam ter decidido a eleição. Este trabalho tem a intenção de delinear um perfil dos municípios brasileiros que votaram nos candidatos à presidência nas eleições de 2018 no segundo turno, utilizando análise de correspondência múltipla. Neste estudo serão consideradas características dos municípios como o Índice de Desenvolvimento Humano Municipal, as proporções de homens, mulheres, escolaridade, religião, cor/raça e estado civil.

Palavras-chave: candidatos, eleições, análise de correspondência múltipla.

Email de contato: thalita_nascimento@id.u.br, jarrais@id.u.br.

R Ladies Niterói: Ensino da Linguagem R

Julia Hellen Franco Ferreira¹, Juliany Freitas Moura¹ e Luíza de Oliveira Machado Pinto¹

¹Universidade Federal Fluminense

O R Ladies é uma organização mundial cuja missão é promover a diversidade de gênero na comunidade da linguagem R [R LADIES GLOBAL, 201-]. O propósito do R Ladies Niterói é inserir mais mulheres na cidade de Niterói, com idades variadas e diversas áreas profissionais a programarem em R. Este incentivo é dado através de encontros regulares como palestras e minicursos. Os eventos organizados são gratuitos e aberto ao público.

Palavras-chave: R, programação, Rstudio, Niterói, mulheres, educação, ensino.

Email de contato: juliaferreira@id.u.br, julianymoura@id.u.br, luizaomp@id.u.br

Uma análise da qualidade de recomendações via Filtragem Colaborativa e Regressão Logística

Amanda de Araujo Gomes¹ e Jessica Quintanilha Kubrusly¹

¹Universidade Federal Fluminense

Com o avanço da tecnologia, a quantidade de dados disponíveis aumentou em proporção maior do que a capacidade de processá-los. A internet oferece um vasto número de páginas Web destinadas à oferta de produtos, tanto de compra/venda como roupas, eletrodomésticos, livros, como de produtos conceituais como artigos, filmes, músicas. O surgimento dos sistemas de recomendação facilitou o acesso e diminuiu o tempo despendido nas buscas pelas informações de interesse, auxiliando tanto aos que disponibilizam conteúdo, podendo-o fazer de maneira direcionada, quanto aos consumidores com conteúdos personalizados. O presente trabalho tem como objetivo estudar o sistema de Filtragem Colaborativa e aplicá-lo no caso da recomendação de filmes para usuários. Além disso, serão propostos modelos logísticos para avaliar se o número de usuários que avaliaram um certo filme e o número de filmes avaliados por um dado usuário influenciam na qualidade da recomendação do filme em questão para esse usuário.

Palavras-chave: Filtragem Colaborativa, Sistemas de Recomendação, Regressão Logística.

Email de contato: amandaaraujogms@gmail.com, jessicakubrusly@id.u.br.

Instituições parceiras e patrocinadores

A 11ª Semana da Estatística é um evento que faz parte da Agenda Acadêmica da Universidade Federal Fluminense de 2019.

Instituições parceiras



Patrocinadores



