

Introdução ao LATEX - 3h

Daniel Furtado Ferreira - UFLA

Este minicurso tem pretende fazer uma pequena introdução ao LaTeX e aos seus recursos. A estrutura básica do curso está dividida em várias etapas. A primeira delas fará uma explicação básica ao programa e aos recursos que são necessários para sua instalação, que será focada essencialmente na plataforma Windows. Na segunda etapa serão abordadas a estrutura básica de um arquivo LaTeX, destacando alguns comandos básicos do preâmbulo e do uso de pacotes e de formatação do documento criado. Nesta etapa serão apresentados caracteres especiais, comandos e detalhes de espaçamento e de criação de parágrafos e inserção de comentários. Concluimos esta etapa com o objetivo deste minicurso. Na terceira etapa se apresentam várias características do LaTeX quanto a digitação de textos, desde a digitação de parágrafos até a apresentação dos corpos flutuantes. Posteriormente, na quarta etapa, apresentam-se as fórmulas matemáticas e os pacotes básicos para lidar com elas de forma otimizada. Apresentam-se comandos para matrizes, espaçamento no modo matemático, entre outros recursos. Na quinta etapa, apresentam-se conceitos básicos da bibliografia, índices remissivos, entre outros recursos adicionais. Na sexta etapa, pretende-se fazer uma breve introdução ao pacote tikz. Este pacote permite lidar com gráficos com um potencial extremamente elevado. Finalmente, se encerrará o minicurso com as conclusões principais da exposição dando-se os créditos a Donald E. Knuth e lembrando que o minicurso de baseou nas notas de Tobias Oetiker.

Python para usuários de R – 4 h

Lincoln Frias e Patrícia de Siqueira Ramos - UNIFAL

A linguagem Python tem sido cada vez mais utilizada para a análise de dados e *machine learning* (www.kaggle.com/surveys/2017). Por isso, o objetivo deste minicurso é apresentar um panorama da análise de dados em Python, indicando os pacotes essenciais e apontando semelhanças e diferenças em relação ao R. As principais funcionalidades que um estatístico espera estão nos seguintes pacotes: pandas (leitura de dados e operações com dataframes), numpy (operações vetoriais), scipy.stats (distribuições estatísticas), statsmodels (modelos de regressão), pymc3 e pyStan (inferência bayesiana) e scikit-learn (*machine learning*). Duas das bibliotecas mais utilizadas para a elaboração de redes neurais são escritas em Python: tensorflow/keras e pytorch. Há diversas opções para a elaboração de gráficos. O pacote matplotlib é o mais utilizado, maduro e versátil e serve de base para os pacotes seaborn e plotnine (uma versão do ggplot2), que permitem gerar gráficos mais elaborados e maneira mais simples. O pacote altair é uma excelente novidade, pela elegância de sua sintaxe, grande capacidade de gerar gráficos interativos e facilidade de integração com páginas da web. Outra opção, um equivalente mais próximo do pacote shiny do R, é o pacote dash. Porém, para além da análise de dados, a principal vantagem da Python é o fato de ser uma linguagem de uso geral. Embora ainda não possua certos pacotes estatísticos especializados disponíveis em R, há outros 130 mil pacotes no PyPI (o equivalente do CRAN), em áreas tão diversas quanto edição de áudio e vídeo, construção de sites, astronomia, processamento de texto e manipulações algébricas. A

página analisededados.org/pythonr apresenta uma comparação entre Python e R, incluindo uma videoaula. No restante do site há diversos materiais explicativos sobre Python.

Uma introdução ao *Machine Learning* com o R – 4 h

Paulo Henrique Sales Guimarães - UFLA

Machine Learning (Aprendizagem de Máquina) é um conjunto de técnicas e procedimentos que permite que os computadores (máquinas) possam agir e tomar decisões baseados em dados, isto é, aprendam por meio dos dados, constituindo atualmente, um dos campos de trabalho mais importantes dentro da Inteligência Artificial. Assim, possibilita a predição e o aprendizado de certos padrões e comportamentos automaticamente, sem a intervenção humana. Este minicurso tem por objetivo introduzir o conceito de *Machine Learning* utilizando o *software* R, tratando conceitos básicos na área como aprendizagem supervisionada e não supervisionada, além da classificação, árvores de decisão, aprendizagem baseada em instância (kNN), Naive Bayes, e conceitos básicos sobre máquinas de vetores de suporte (SVM).

Expressões Regulares para Preparação de Base de Dados – 6 h

Crysttian Arantes Paixão - UFSC

Serão apresentados os conceitos básicos sobre o uso de expressões regulares para a manipulação de base de dados. O curso abordará as principais aplicações para o processamento de dados para análise utilizando o *software* R, destacando a coleta de informações na internet (*web scraping*).

Documentos dinâmicos usando o R Markdown – 3h

Kelly Pereira de Lima - UFLA

A pesquisa reproduzível é um tema da ciência que tem despertado muito a atenção de pesquisadores nos últimos anos, pois sua ideia geral é a realização de uma análise de dados para que a publicação contenha os dados e código para que outras pessoas consigam reproduzir e verificar os resultados. Existem diversos relatos de que um estudo foi repetido e gerou resultado que diferente do original ou mesmo discordante de um estudo anterior. Desta forma, a reprodutibilidade na ciência é altamente reconhecida nos dias atuais, mas ainda não é largamente praticada. Pois, os cientistas ainda não adotaram uma ferramenta para as necessidades da pesquisa reproduzível. Desta forma, esta apresentação mostrará uma ferramenta que é capaz de criar documentos dinâmicos utilizando uma linguagem simples, o R Markdown que possibilita a inclusão de códigos e suas saídas diretamente do R, os documentos em Markdown. Nesta palestra, apresentaremos esta ferramenta que tanto facilita a criação de relatórios e apresentações de análise de dados.

**An Intuitive Approach to Machine Learning: Boosting, Nearest Neighbors,
Random Forests and Support Vector Machines – 6h**

**Andreas Ziegler - Institute of Medical Biometry and Statistics at the University of Lübeck,
Germany**

Modelagem de dados complexos: uma introdução aos GAMLSS – 8h

Luiz Ricardo Nakamura - UFSC