

# Minicurso 1: Aprendizado de Máquina com o Pacote tidymodels

Marcus Nunes<sup>1</sup>

<sup>1</sup>UFRN, [marcus@marcusnunes.me](mailto:marcus@marcusnunes.me)

## 1 Resumo

Este minicurso será uma introdução ao aprendizado de máquina com R, mostrando como utilizar o pacote tidymodels para realizar tarefas de regressão e classificação. O pacote tidymodels simplifica o fluxo de trabalho em aplicações de aprendizado de máquina, pois suas características são análogas às do tidyverse. Conceitos como sobreajuste, validação cruzada e engenharia de atributos serão trabalhados de maneira que os alunos compreendam a sua importância para as tarefas de aprendizagem de máquina. Será um minicurso aplicado, com ajustes de modelos em dados reais. Os ajustes serão criados, refinados e analisados, de modo a definir critérios que possam ajudar o aluno a decidir qual o melhor modelo para as suas necessidades.

## 2 Sobre o instrutor

Expert em *data Science*. Ver site: <https://marcusnunes.me/>.

## 3 Data

Dias: 29/11/2021 (Segunda-feira) e 30/11/2021 (Terça-feira)

Horário: 14:00 às 15:45

## **Minicurso 2: UMA CONVERSA ENTRE O R E O POWER BI: “A VIDA COMO ELA É NAS EMPRESAS”**

Adriana Letícia dos Reis<sup>1</sup>, Eliana Rigoni<sup>1</sup>, Ianní Muliterno<sup>1</sup>, Regina  
Albanese Pose<sup>1</sup>

<sup>1</sup> *R-Ladies São Paulo.*

### **1 Resumo**

**Objetivo:** Apresentar o cotidiano do mundo corporativo e como o DS transforma as empresas

**Público Alvo:** Todes que tenham interesse e curiosidade para conhecer um pouco da “vida como ela é nas empresas”

As R-Ladies apresentam a comunidade e seu propósito, um pouco da história de todo esse movimento e contam um pouco quem são e o que fazem. A partir de alguns cenários da “vida como ela é nas empresas” serão discutidos alguns fundamentos estatísticos dentro de aplicações de R e de Power BI. Conceitos de estatística básica que poderiam ser explorados no mundo corporativo com o objetivo de orientar os tomadores de decisão

### **2 Sobre as instrutoras**

Adriana Letícia dos Reis - BI Consultant at Vivo

- <https://www.linkedin.com/in/reis-al/>
- Bacharel em Sistema de Informação - Mackenzie
- MBA Economia e Gestão Empresarial - FGV

Eliana Rigoni - Data Analyst at Ambev Tech

- <https://www.linkedin.com/in/elianarigoni/?originalSubdomain=br>
- Bacharel em ADM - FAPCAMP
- Analista de Dados – IGTI
- MBA ESALQ USP - cursando

Ianni Muliterno – DS/Estatística at Unilever

- <https://www.linkedin.com/in/iannimuliterno/>
- Bacharel em Estatística - UFPE
- Técnico em automação industrial SENAI

Regina Albanese Pose – avó da Jade e professora USCS

- <https://www.linkedin.com/in/regina-albanese-pose-2300b4110/>
- Licenciada em Matemática - FSA
- Psicopedagoga clínica e institucional/lato sensu – São Marcos
- Mestre em Ciências FMUSP
- Especialista em Poluição atmosférica e saúde humana/lato sensu – FMUSP
- Bacharel em Estatística – UNICAPITAL

### **3 Data**

Dias: 29/11/2021 (Segunda-feira) e 30/11/2021 (Terça-feira)

Horário: 14:00 às 15:45

# Minicurso 3: Análise de dados de amostras complexas usando R

Marcel Vieira<sup>1</sup> e Pedro Silva<sup>2</sup>

<sup>1</sup> UFJF, [mdtvieira@gmail.com](mailto:mdtvieira@gmail.com)

<sup>2</sup> ENCE, [pedronsilva@gmail.com](mailto:pedronsilva@gmail.com)

## 1 Resumo

Muitas pesquisas amostrais realizadas para produção de estatísticas públicas empregam planos amostrais complexos, envolvendo estratificação, probabilidades desiguais de seleção, amostragem conglomerada, calibração de pesos etc. Os dados provenientes dessas pesquisas precisam ser analisados levando em conta tais aspectos do planejamento amostral, sob pena de obtenção de inferências incorretas. O curso apresentará alguns modelos e métodos requeridos para seu ajuste quando os dados são provenientes de amostras complexas. Também apresentará as ferramentas disponíveis para ajuste dos modelos no ambiente do sistema R, com diversos exemplos baseados em pesquisas amostrais.

Conteúdo: Modelos de superpopulação; método da máxima pseudo-verossimilhança; efeitos do plano amostral; pacotes `survey` e `srvyr` do R; métodos para estimação de variâncias.

## 2 Observações

Modelagem estatística de dados amostrais - tema de trabalho de Djalma Pessoa.

Visite: <https://www.isi-web.org/events/courses/analysis-of-survey-data-using-r>

## 3 Data

Dias: 01/12/2021 (Quarta-feira) e 02/12/2021 (Quinta-feira)

Horário: 10:45 às 12:30

**Referências bibliográficas**

Pessoa, D.G.C. e Silva, P. L. d. N. (2018). Análise de dados amostrais complexos.  
<https://djalmapessoa.github.io/adac/>

Silva, P. L. N., Bianchini, Z. M., Dias, A. J. R. (2021) Amostra: teoria e prática usando R. Rio de Janeiro: [s.n.]. Available at:  
<https://amostragemcomr.github.io/livro/index.html>.

# Minicurso 4: Modelos de regressão linear uni e multivariados para dados com caudas pesadas, simétricos ou assimétricos.

Clécio Ferreira<sup>1</sup>, Camila Zeller<sup>2</sup> e Diego Gallardo<sup>3</sup>

<sup>1</sup>UFJF, [clecio.ferreira@ufff.edu.br](mailto:clecio.ferreira@ufff.edu.br)

<sup>2</sup>UFJF, [camila.zeller@ice.ufff.br](mailto:camila.zeller@ice.ufff.br)

<sup>3</sup>Universidad de Atacama-Chile., [diego.gallardo@uda.cl](mailto:diego.gallardo@uda.cl)

## 1 Resumo

Este minicurso propõe utilizar famílias de distribuições de probabilidade em modelos de regressão linear uni e multivariados de dados com caudas pesadas, simétricos ou assimétricos. As famílias de distribuições utilizadas são: Normal (N); Misturas de Escala Normal (SMN); Skew-Normal (SN); Skew-Normal-Cauchy (SNC); Misturas de Escala da Skew-Normal (SMSN), Misturas de Escala Normal Assimétricas (SSMN) e Misturas de Escala da Skew-Normal-Cauchy (SMSNC). Um pacote denominado 'skewMLRM' foi desenvolvido e publicado no CRAN do R com as seguintes funcionalidades: geração de amostras aleatórias das famílias de distribuições, ajuste de modelos de regressão linear uni e multivariados, seleção de variáveis via AIC/BIC e também via significância dos coeficientes de regressão. Por fim, o pacote permite fazer seleção de modelos, dentre e entre famílias. Com exceção do modelo Normal, nos demais os estimadores de máxima verossimilhança são baseados no algoritmo EM e suas extensões (ECM/ECME).

## 2 Observações

Utiliza pacote "skewMLRM" no R-CRAN.

## 3 Data

Dias: 01/12/2021 (Quarta-feira) e 02/12/2021 (Quinta-feira)

Horário: 10:45 às 12:30