

# Regresní analýza dat

## 01RAD - Cvičení 00

Jiri Franc

Czech Technical University  
Faculty of Nuclear Sciences and Physical Engineering  
Department of Mathematics

- ▶ **Regresní analýza dat - 01RAD** 2+2 z,zk.
- ▶ **Podmínky pro získání zápočtu.**
  - ▶ Povinná docházka:  
**max. 3 absence**  
při distanční výuce bude docházka brána z účasti na meetingu na MS Teams další absence lze odčinit vypracováním úkolů z absentujícího cvičení a jejich odevzdáním.
  - ▶ **aktivní** účast na hodinách - vypracování zadaných úloh.
  - ▶ vypracování a odevzdání 3 úkolů (reportů) během semestru (zadání cca na 4, 8, a 12 hodině)
- ▶ **Konzultace**  
po předchozí domluvě na MS Teams (prosím nepoužívejte e-mail ani pro domluvu) nebo v T-40.

### Literatura k přednášce

- ▶ M. Golberg, H A. Cho, Introduction to Regression Analysis  
Link to Amazon  
Link to Google books

### Volně dostupné tutoriály k R na r-project.org

- ▶ Contributed Documentation

### Online kurz

- ▶ Codecademy.com: learn-r

### Literatura v češtině

- ▶ K. Zvára, Regrese, MatfyzPress, 2008

### **Volně dostupná literatura k datové analýze v R**

- ▶ Garrett Golemund, Hadley Wickham, R for Data Science
- ▶ Max Kuhn and Kjell Johnson, Feature Engineering and Selection: A Practical Approach for Predictive Models
- ▶ Rafael A. Irizarry, Introduction to Data Science Data Analysis and Prediction Algorithms with R
- ▶ Hadley Wickham , Advanced R
- ▶ Malte Grosser, Henning Bumann, and Hadley Wickham, Advanced R Solutions
- ▶ Claudia A Engel, Data Wrangling with R
- ▶ Colin Gillespie, Robin Lovelace, Efficient R programming

### **Volně dostupná literatura kombinující R a regresí.**

- ▶ G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning with Applications in R.
- ▶ J. J. Faraway, Practical Regression and Anova using R.
- ▶ J. J. Faraway, Extending the Linear Model with R (Generalized Linear, Mixed Effects and Nonparametric Regression Models).
- ▶ R Core team, An Introduction to R (Notes on R: A Programming Environment for Data Analysis and Graphics).
- ▶ J. Verzani, simpleR - Using R for Introductory Statistics.

## Základní informace

### Code style for R

- ▶ tidyverse style
- ▶ Google style

### Git and R

- ▶ Happy Git and GitHub for the useR

### Různé cheatsheets

- ▶ R cheatsheet
- ▶ data wrangling cheatsheet
- ▶ lubridate cheatsheet

### R for MATLAB users

- ▶ Translate MATLAB to R

### Blogs and Webs

- ▶ r-bloggers.com
- ▶ Kaggle
- ▶ Julia Silge and her TidyTuesday

## Obsah cvičení (Předběžný - bude upřesněno dle přednášek):

1. Úvod do R
2. Práce s daty v R
3. Jednoduchá Regrese I
4. Jednoduchá Regrese II
5. Vícerozměrná Regrese I
6. Vícerozměrná Regrese II
7. Post-hoc analýza: Rezidua, Diagnostika, Homoscedasticita, Normalita, ...
8. Transformace: polynom, log, Box-Cox, GLM
9. Podmodel - Kritéria výběru modelu
10. Multikolinearita
11. Vlivná pozorování, influenční míry
12. Úvod do robustních metod v regresní analýze
13. Práce na zápočtové úloze