Intro to Quantitative methods

Chuvakin Sergey

«School of Advanced Studies»

October 11, 2020

Welcome

- **About me:** Data Scientist in Arcadia inc. (SPb), mostly worked with pharma companies. Former sociologist with focus on criminology.
- **My interests:** Business and Research Analytics, Natural Language Processing, Python, Digital Humanities, Robotics
- **Feel free to contact me:**
  - Email: sergeychuvakin1@mail.ru
  - telegram: @sergechuvakin
- **Office hours:** TBA

Quantitative methods

Quantitative methods – is a set of rules and algorithms to reach stable, sustained results of research.

Research – common term for defining procedure of finding answer on question.

Quantitative methods

After completing this course, you will be able:

- ▶ to read and understand (sic!) quantitative research papers
- ▶ to speak the language of data fluently
- ▶ move further to Data Science and Machine Learning
- ▶ able to understand and explain to others such words as «variable», «distribution», «regression», «p-value», etc.
- ▶ able to choose statistical methods appropriate to your research problem
- ▶ use R for your programming needs

Course Structure

- ▶ Week 1 - Intro ξ
- ▶ Week 2 - R Language ξ
- ▶ Week 3 - Key statistical concepts ξ
- ▶ Week 4 - Data Management ξ
- ▶ Week 5 - Basic statistical tests ϕ
- ▶ Week 6 - Regression ξ
- ▶ Week 7 - Regression advanced
- ▶ Week 8 - Methodology of Quantitative research ξ ϕ

ξ – Quiz on the week
ϕ – Home Assignment

Prerequisites

- ▶ Math (at least school level)
- ▶ Calculus (basics) (recap today)
- ▶ Linear Algebra (basics) (recap today)
- ▶ Computer literacy

Major (approximate)

Assignment or Task	Due date/s	Percent
Recap of math and Probability	22 November	8
R language practice	29 November	8
Data management practice	13 December	8
Regression practice	27 December	8
1 Home assignment	11 January	30
Design research	20 January	8
2 Home assignment	20 January	30

* it's not the end version, small changes can be (TBA)

Minor (approximate)

Assignment or Task	Due date/s	Percent
Recap of math and Probability	22 November	15
R language practice	29 November	15
Data management practice	13 December	15
Regression practice	27 December	15
1 Home assignment	11 January	25
Design research	20 January	15

[*] it's not the end version, small changes can be (TBA)

Reading

- Field A., J. Miles, and Z. Field. 2012.Discovering Statistics Using R. SAGE publications ltd
- Wickham, H., and Grolemund, G. 2016.R for data science.O'Reilly Media

Software

- ▶ R - Language itself
- ▶ Rstudio - Application for comfortable work
- ▶ GitHub - all the materials stored here

Intro - today

- ▶ Course introduction
- ▶ Calculus and linear algebra recap
- ▶ quiz

R Language

R – general purpose programming language. We will practice:

- ▶ writing a good readable code
- ▶ use basic built-in functions
- ▶ modify and create data tables
- ▶ prepare data to analysis
- ▶ use statistical functions
- ▶ run models and diagnostics
- ▶ visualize data and models
- ▶ use RMarkdown extension
- ▶ a lot of other activities

Key statistical concepts

- ▶ what is variable
- ▶ types of variables
- ▶ sample and population
- ▶ Representativeness
- ▶ central measures
- ▶ distributions
- ▶ standard deviation

Data management and visualization

- ▶ Relational data structure
- ▶ What is dplyr
- ▶ What is ggplot2
- ▶ basic data transformations
- ▶ visualization basics

Basic statistical tests

- ▶ compare two means
- ▶ compare more than two means
- ▶ t-test
- ▶ chi-square
- ▶ significance
- ▶ ANOVA
- ▶ non-parametrics tests
- ▶ correlation

Regression

- ▶ assumptions
- ▶ formula
- ▶ t-test
- ▶ interpretation
- ▶ significance

Regression advanced

- ▶ Diagnostics
- ▶ types of errors
- ▶ GLM

Methodology

- ▶ Design of research
- ▶ **A**/B tests