Medical Statistics

2022-06-14

Contents

About this course

In this course an introduction to basic statistical methods useful for biomedical data analysis will be given. During the course we will alternate between lectures and practicals, allowing for plenty of interaction and illustration with examples of practical interest.

The course requires little prior statistics knowledge, but assumes participants are able to work with R, R packages and RMarkdown for the practicals. Participants with no or little experience in R are strongly advised to follow an introductory R course prior to following this course, such as the one we offer.

The course program includes: exploratory data analysis, basic statistical tests, methods for count data tables, linear and logistic regression, as well as basic methods for survival data analysis. For each method, power analysis and sample size determination will be handled.

Teachers

- Renee Menezes, Biostatistics Center, NKI (**coordinator**)
- Renaud Tissier, Biostatistics Center, NKI
- Leyla Azarang, Biostatistics Center, NKI

Goals & Topics

After the course you will be able to:

- Perform data exploration analysis
- Perform comparison between two different groups using statistical tests
- Perform linear and logistic regression
- Perform survival data analysis
- Use the R software to perform these tasks

We will cover the following topics:

- Probabilistic distributions
- Hypothesis testing

- Linear regression
- Logistic regression
- Survival data analysis
- Power analysis

Prerequisites

The course assumes no prior programming knowledge. Elementary statistics knowledge is necessary.

Participants must **bring own laptops** capable of running RStudio.

Before the course **please prepare your laptop**:

1. install R, an open-source, free environment for statistical computing and graphics. You can find instructions for downloading and installing it from one of the CRAN mirrors, for example from the Univ. of Gent or from the Imperial College. A full list of mirrors can be found here.
2. install RStudio. Go to the RStudio download page, select a version of RStudio appropriate for your laptop, download it and then install. Please check whether you can start RStudio.
3. install RMarkdown, a very nice and easy tool to produce reports using RStudio. It is made available as an R package for Rstudio. One easy way to install it is as follows:

i) open RStudio

ii) click on the "File" menu on the top left, and choose "New file">"R Markdown". If RMarkdown is not yet installed on your machine, this will prompt you to install it and any packages required. Just follow the instructions that appear on the screen.

4. Several packages are required to perform the course. These packages can be installed before the course to gain some time. The required packages are:

- factoextra
- gplots
- pwr
- statmod

Materials

After the course material .zip file is downloaded, the course material can be assessed:

- as HTML pages by opening `index.html` in any browser

- by clicking on the `RcourseNKI.Rproj` file, which will open the entire course as an R project
- via the course source git repository

The materials contain a `data` directory with the **data files** used in the presentations/tasks. The directory can be also accessed at https://github.com/rxmenezes/RcourseNKI/tree/master/data

Programme

Third NKI edition, Novermber 22nd, 23rd, 24th, 25th and 26th 2021

This course will be given online via Zoom. All course days are in the period 9:00-16:00, with the last hour reserved for general Q&A.

Chapter 1

Introduction and exploratory data analysis

Placeholder

1.1 Introduction to probabilistic distributions

1.1.1 Motivation

1.1.2 Discrete random variables

1.1.3 Continuous random variable

1.2 Data representation

1.2.1 Motivating example

1.2.2 Working example

1.3 Mean-Variance

1.3.1 Working example

1.4 Correlation and heatmap

1.4.1 Motivation

1.4.2 Working example

Chapter 2

Statistical tests - Part 1

Placeholder

2.1 Comparing 2 groups

2.1.1 Motivation

2.1.2 Working Example

2.2 Power of a statistical test

2.2.1 Motivation

2.2.2 Working example

Chapter 3

Common regression models

Placeholder

3.1 Linear regression

3.1.1 Motivation

3.1.2 Working Example

3.2 Logistic regression

3.2.1 Motivation

3.2.2 Link function

3.2.3 The `glm` function

3.2.4 Working example

3.2.5 Fitted values

Chapter 4

Statistical tests - Part 2

Placeholder

Chapter 5

Survival analysis

Placeholder

5.1 Survival data - Introduction

5.1.1 Motivation

5.1.2 Working example

5.1.3 Kaplan-Meier curve

5.1.4 Working example (cont)

5.1.5 Survival data analysis in R

5.1.6 Survival data for groups: the log-rank test

5.1.7 Working example (cont)

5.2 Survival data - regression models

5.2.1 Motivation

5.2.2 Cox proportional-hazards model

5.2.3 Working example

5.2.4 Group-specific baseline hazards

5.2.5 The proportional hazards assumption

5.2.6 Others

5.3 Survival data - power analysis

5.3.1 Motivation

5.3.2 Power and sample size for Cox regression

5.3.3 Working example

5.3.4 Minimum sample size

5.3.5 Working example (cont)

5.3.6 Reference

5.3.7 Notes

5.4 Survival analysis - Notes

5.4.1 Other types of censoring

5.4.2 Right-, left- and interval censoring

5.4.3 Type I and type II censoring

5.4.4 Competing risks

Chapter 6

Exercises

Placeholder

6.1 Chapter 1

6.1.1 Exercise 1

6.1.2 Exercise 2:

6.1.3 Exercise 3(*)

6.2 Chapter 2

6.2.1 Exercise 1

6.2.2 Exercise2

6.3 Chapter 3

6.3.1 Exercise 1

6.3.2 Exercise 2

6.3.3 Exercise 3

6.4 Chapter 4

6.4.1 Exercise 1

6.4.2 Exercise 2

6.4.3 Exercise 3

6.4.4 Exercise 4

6.5 Chapter 5

6.5.1 Exercise 1

6.5.2 Exercise 2

6.5.3 Exercise 3

Chapter 7

Appendix

Placeholder

7.1 PCA Analysis

7.1.1 Motivation

7.1.2 Working example

7.1.3 Quick tasks