

One day crash course for medics on Biostatistics with R

Andrew C Parnell

28 June 2016

Class 1: Basics and experimental design

9:30-10:30

Learning outcomes:

- Know how to create and interpret a two-sample t-test
- Understand what a p-value means
- Be able to perform a simple sample size calculation
- Understand the basics of experimental design

Slides:

- Intro slide
- Syllabus/timetable for the day
- General goal: be able to create a statistical model of a biomarker panel and check that it is robust
- Basics of data: continuous vs discrete, ordinal vs interval vs nominal
- Two examples: prostate cancer (regression)
- South African Heart Rate data (classification)
- Testing differences between groups; the two-sample t-test
- Sampling distributions of data
- Null and alternative hypotheses
- Drawing pictures
- Getting and understanding the p-value
- What the p-value is not
- Introduction to sample size calculations
- Type 1 and Type 2 error
- Drawing pictures
- The magic formula
- Getting the values to put in to the formula
- Possible extensions
- Design of Experiments
- The golden rule of designing an experiment
- Blocking
- Randomisation
- Replication
- More complicated experiments

Class 2: Regression and classification

10:45-11:45

Learning outcomes:

- Be able to understand the structure of regression and classification models
- Know how to read and interpret the output of a statistical model
- Be familiar with some of the extensions to basic regression and classification models

Slides:

- Regression vs Classification: what's the difference?
- Response and explanatory variables
- A basic regression model
- Example: prostate cancer
- Reading the output of the model
- Plotting the fit
- Expanding the model with two explanatory variables
- Expanding the fit even more
- Regularisation and shrinkage
- Lasso; Ridge and Elastic Net
- Dealing with interactions
- Even more advanced regression approaches
- Intro to classification models
- The logit transformation
- Example: SA Heart rate
- Extending the model
- Understanding the output
- Plotting the fitted model
- Regularisation and shrinkage for classification
- More advanced classification approaches

Class 3: Model checking and performance

12:00-13:00

Learning outcomes:

- Be able to read and understand regression diagnostic plots
- Be able to compare statistical models using information criteria and cross-validation
- Understand the different types of classification metrics
- Understand and interpret ROC curves and AUC values

Slides:

- Regression diagnostics
- Residual plots
- The ANOVA table
- Classification diagnostics
- Model comparison
- Cross-validation
- Right ways and wrong ways to do cross validation
- Sensitivity and specificity
- Choosing a probability cut-off: Youden's index

- The ROC curve
 - AUC
 - Calibration
 - Decision curve analysis
-

Practical 1: Introduction to Rstudio

14:00-15:00

The different parts of Rstudio Using the console window Using the Script window Creating a script Writing comments Running a script Reading in data Different data types Combine, data frames and lists Running functions Creating simple scatter plots

Practical 2: Running models and diagnostics

15:30-16:30

Pick one of the script files (one for Prostate data - regression, one for South African data - classification) and run the files, answering the questions as you go.