

Syllabus Research Methodology

Dr. Marieke van Vugt

February 2, 2015

The goal of *Research Methodology* is to learn how research is being done, and to put that knowledge into practice. You learn how to apply a great number of statistical techniques, draw conclusions from those, and determine what statistical technique would be suitable for a given dataset and/or research question. In contrast to your *Statistics* course, here the emphasis lies on interpreting results and communicating those to the world at large. **In other words, you learn how to take a step back and think about what you can conclude from a certain experiment or statistical test.**

This course is meant as preparation for your Bachelor project. The course builds on knowledge and skills you have (hopefully) acquired during *Statistics* and *Basic Scientific Skills*. Lectures in this course are meant to be a complement to the knowledge you can obtain by reading the textbook *Empirical Methods for Artificial Intelligence*. You will practise your skills by making use of datasets from the research of some of the AI faculty.

Learning objectives

- Defending the use of Research Methodology
- Judging the reliability and validity of experiments
- Being able to perform exploratory data analysis
- Using parametric and non-parametric hypothesis tests (and interpreting their results). Those include t-tests, Wilcoxon rank sum, ANOVA
- Being able to draw conclusions from categorical data
- Using computer-intensive methods for data analysis
- Drawing conclusions from statistical test results
- Being able to compare statistical models
- Being able to argue when to use Bayesian vs Frequentist statistics

These objectives will be achieved by means of lectures, discussions in the lectures, assignments and blogs.

Book:

Empirical Methods for Artificial Intelligence
Paul R. Cohen

Software:

We will primarily use R (<http://www.r-project.org>) for the statistical testing. In case you are not familiar with R, I advise you to do an R tutorial. (<http://cran.r-project.org/doc/manuals/R-intro.pdf>). An easy way to use R is through the free Rstudio software: (<http://rstudio.org/>). Handy tips for R use: <http://www.r-bloggers.com/a-crash-course-in-r/>, http://www.computerworld.com/s/article/9239672/Beginner_s_guide_to_R_Easy_ways_to_do_basic_data_analysis?taxonomyId=11&pageNumber=1

Schedule:

Lecture:		
Monday 13:00-14:45		
Friday 11:00-12:45		
Practical:		
group 1	group 2	group 3
Tue 9:00-11:00 BB216	Thu 13:00-15:00 BB273	Thu 15:00-17:00 X5116.0310

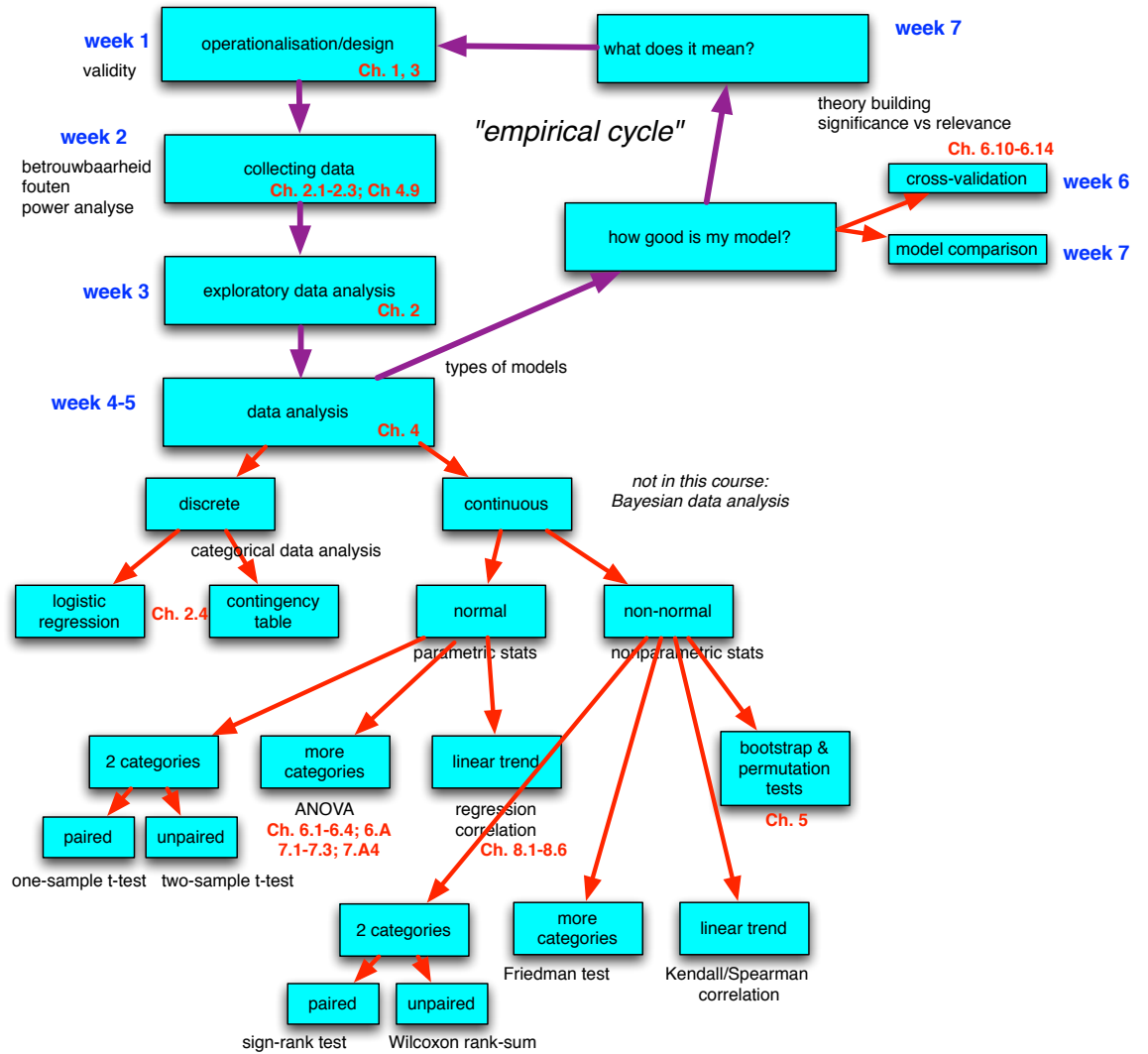
Attendance at practicals is not obligatory, but of course handing in the assignments is. At the practicals you can ask the student assistants questions, and work together with other students. Note that everyone has to hand in their own assignment, from which it should be clear that you have done it yourself and not copied from another student. The assignments are a perfect preparation for the exam, which will consist of questions that are similar to the assignments.

You are also very much encouraged to attend the lectures, where material will be discussed that is not in the book but may still occur on the exam and in the assignments. We will have a lot of discussions, and there will be presentations about statistical tests by you!

Contributing to the lectures

For every Monday class, you are requested to submit a question about one of the chapters that you were assigned for that week. This should be a question that you think could occur on an exam, and that can be discussed in class. You are required to hand in the question on paper in class. We will then choose the best exam question every Monday classes. These questions count towards your class participation grade!

Weekly overview:



Grade composition:

- Exam (50%) - this is a closed-book exam. The exam takes place on April 9th, 9:00-12:00 in rooms 5419.0013 and 5419.0113. The re-exam will happen on June 10th, 9:00-12:00 Aletta Jacobshal 01.
- Presentation about a statistical test and application to a dataset (5%)
- Active class participation, exam questions (5%)
- Assignments (together 40%)

Bloggng about an idea for a Bachelor project

During this class you are encouraged to starting thinking about a research question for your Bachelor project. To make this a bit interactive, you will post these reflections on a blog so it can be shared with other students and the rest of the world. You can find the blog at reflectiesonderzoeksmethodologie.wordpress.com. Blog assignments are part of the assignments.

Goals of blogging: reflecting on how Research Methodology is relevant for your own work. Besides, this will make you more comfortable with scientific writing and scientific discussions. I expect contributions of about 1-3 paragraphs to the blog.

Topics: It may be a challenge to come up with topics. You can find a list of thesis topics of PhD students at ALICE here: <http://research.ai.rug.nl>
Your writing will be judged based on the following *criteria*:

score	interpretation
1	<i>Assignment to performed:</i> Nothing submitted or merely a few incoherent sentences
2	<i>Too restricted:</i> Without focus, merely repeats what has been said before. There is no evidence that the student has worked with the material.
3	<i>Underdeveloped:</i> Mostly a summary or description of phenomena, no connections are made or alternative viewpoints are given. The student has spent a limited amount of time/effort on the material.
4	<i>Good:</i> There is a decent focus, with explanation or analysis of examples or data. Connections are made between ideas, or insights are being given but they haven't been fully worked out. The student clearly has spent some time with the material.
5	<i>Exceptional:</i> The blog has a clear focus in which examples are integrated with explanation or analysis. There is also an awareness of the limitations of this perspective, or alternative perspectives are given. The student clearly has intensively worked with the material.

Assignments

Goal of the assignments: becoming familiar with the scientific method by applying it to realistic datasets. In addition to correctly performing statistical tests, you will go beyond what you learned in *Statistics* by interpreting the results of these tests and thinking about those critically.

Methods: Every week you will be given datasets and assignments to do (partly in R). During the practicals you can ask questions to the student assistants. Be sure to submit your assignments as PDF!

Presentations

Every student will give a presentation about a statistical test. In the presentation you will explain the statistical test (not necessarily the equations, but at least the intuition behind it). You explain for what kind of data and research question the test can be used, what the assumptions are of the test, and what conclusions you can draw from it. It is also useful to give R code and show your classmates how the test is put into action. You get bonus points for giving an entertaining and memorable presentation. As your sources you can use: the *Empirical Methods* book, *Using R for introductory statistics* by Verzani (book from your *Statistics* course). You are also welcome to use other sources, but be aware that not all sources are equally reliable (not everything that is said on Wikipedia is true!). Make sure the presentation does not exceed 10 minutes!

I will judge the presentations on:

- Enthusiasm and originality/creativity
- Clear description of the test, including some explanation of the logic of the mathematics behind it. You can also choose to do a simulation of the test on a dummy dataset to illustrate your results
- Clear description of data requirements and assumptions of the test
- Clear description of conclusions you can draw from this test

date	test	student 1	student 2	student 3	student 4
9/2	unpaired t-test				
9/2	paired t-test				
13/2	(nonparametric) correlation				
13/2	(multiple) linear regression				
16/2	signrank test				
20/2	Wilcoxon rank sum test				
20/2	one-way ANOVA				
27/2	two-way ANOVA				
27/2	repeated measures ANOVA				
2/3	Friedman test				
6/3	chi-square test				
9/3	logistic regression				
13/3	loglinear analysis				
20/3	nonlinear regression				
20/3	clustering				

Submission deadlines

You are not allowed to submit assignments after the deadline (always on Monday at midnight), unless there is a legitimate reason. If you are sick or have another valid reason, contact the course instructor in time to request an extension.

Plagiarism

It is allowed to work together, but you are required to submit your own assignments, and it should be clear that those are your own ideas and reasonings. This means you are not allowed to submit assignments which look very much like that of a fellow student. In case of a detection of plagiarism, we will notify the exam committee.

Contact information:

I am best reachable by e-mail (m.k.van.vugt@rug.nl), but also by phone (050-363-9487), and in my room in Bernoulliborg (326). You can also ask the student assistants for help:

- Leonoor Ellen (Leonoor.Ellen@live.nl)
- Rene Mellema (rene.mellema@gmail.com)
- Xeryus Stokkel (x.l.x.stokkel@student.rug.nl)