

Course Manual

Basics of Biostatistics

2026-2027

Contents

Introduction to Research and (Bio) Statistics.....	2
About the course.....	2
Learning objectives.....	2
Course topics.....	3
Course schedule.....	3
Course activities.....	3
Course literature.....	4
Prerequisite knowledge.....	4
Statistical software.....	4
Examination.....	5
Evaluation.....	6

Introduction to Research and (Bio) Statistics

Scientific research is defined as the application of scientific methods to the investigation of relationships among natural phenomena, or to solve a medical or technical problem. According to Wikipedia "The scientific method is a body of techniques for investigating phenomena, acquiring new knowledge, or correcting and integrating previous knowledge" (Goldhaber & Nieto, 2010). To be termed scientific, a method of inquiry is commonly based on empirical or measurable evidence subject to specific principles of reasoning (Newton, 1999). The Oxford Dictionaries Online (2016) defines the scientific method as "a method or procedure that has characterized natural science since the 17th century, consisting in systematic observation, measurement, and experiment, and the formulation, testing, and modification of hypotheses".

The scientific method is an ongoing process, which usually begins with observations about the natural world. Human beings are naturally inquisitive, so they often come up with questions about things they see or hear and often develop ideas (hypotheses) about why things are the way they are. The best hypotheses lead to predictions that can be tested in various ways, including making further observations about nature. In general, the strongest tests of hypotheses come from carefully controlled and replicated experiments that gather empirical data. Depending on how well the tests match the predictions, the original hypothesis may require refinement, alteration, expansion or even rejection.

The role of (bio) statistics within (scientific) research is that each decision in the scientific method will have statistical consequences for the next steps.

Statistics can be divided in two phases: descriptive statistics and inferential statistics. Descriptive statistics aims to describe the basic features of the data in a study, to provide simple summaries of the variable(s) in a sample, and to form, together with simple graphics analysis, the basis of virtually every quantitative analysis of data. On the other hand, inferential statistics aims to make inferences concerning some unknown aspect of a population.

About the course

This nine week course (**4.5 ECTS, 14 hours per week**) provides an introduction to statistical methodology. A number of statistical techniques are covered that are relevant for practical biomedical data analysis.

This course covers the concepts of statistical estimation (point and interval estimation) and testing. It focuses on methods developed for categorical data, in particular binary data, and quantitative data, in particular normally distributed data. The course covers, simple linear regression, correlation, one way analysis of variance, analysis of contingency tables and non-parametric statistics. Furthermore, it introduces (multiple) linear regression and (multiple) logistic regression.

The theory will be presented during web lectures and you will have the opportunity to practice your skills through exercises, discussions, quizzes, assignments and case studies of realistic and real data.

Learning objectives

By the end of this course, you will understand the principles and results of statistical analysis techniques including student t-tests, analysis of variance, simple and multiple linear regression analysis, and 1-sample, 2-sample and paired proportion tests.

In particular, you should be able to:

- Understand and reflect on the function of statistics within academic research.
- Recognize and explain basic statistical techniques used for data analysis, including the conditions under which each technique is appropriate.
- Identify which statistical methods are suitable for a given research question.
- Apply these techniques manually and/or using statistical software such as SPSS or R.
- Interpret the outcomes of statistical analyses accurately.
- Communicate and report statistical findings effectively in relation to a research question.

Course topics

This course consists of 9 Learning Units (LU's):

- LU 1 - Introductory Statistical Testing
- LU 2 - Comparisons of Two Means
- LU 3 - Comparison of More than Two Means
- LU 4 - Association in Continuous Data
- LU 5 - Comparison and Association for Categorical Data
- LU 6 - Multiple Linear Regression
- LU 7 - Multiple Logistic Regression
- LU 8 - Preparation Case Study
- LU 9 - Exam and Course Closure

Course schedule

Introduction and discussion statistical theory: week 47-51, week 1-2 (fully online; weekly deadlines)

Christmas holiday: week 52 - 53

Case study (group assignment): week 2-3 (online)

Exam (individual): week 4 (location UMC Utrecht)

Re-exam (individual): week 8 (location UMC Utrecht)

Course activities

During the course there are several activities in which you can participate.

- Weblectures
The statistical topics are explained in weblectures. Some of the lectures are a bit too long to watch in one go. It can be good to pause the lectures so you can digest the information presented. In the activities with lectures, we give an advise on what moment you can best pause. Some of the lectures contain quizzes in which you can test whether you have understood the theory. These tests are easy tools to test your comprehension of the theory that was discussed in the lecture.
- Apply your knowledge
In every Learning Unit you will find multiple activities named "Apply your knowledge". Usually, these activities follow after a web lecture. There is a document with exercises and cases related to the topics previously explained, accompanied with an answer document. If SPSS or R

is used to answer the questions, you can find the syntax and the output of these programs also in this answer document. There are also "How to in SPSS" and "How to in R" documents available. You do not have to hand in your solutions to the exercises and cases, but you can use them to practice with the statistical programs and the topics which are being discussed.

- Test your knowledge
At the end of every Learning Unit you will find a quiz to test your knowledge of that particular learning units including previous learning units. Some quiz questions are similar to what you can expect on the exam.
- Discussions
Sometimes we will add statements about a statistical matter or a news item to the forums so you can discuss it with your fellow participants. You can choose to respond with a completely new argument or to comment on one of your fellow participants. Make sure you respond and contribute to the discussion by elaborating or asking questions.
You can also start your own discussion for example if you have questions about or problems with your own research.
- Case study
You will do the case study in groups of maximum three participants. As a group you have to analyze a case study and prepare a (detailed) report of the statistical analyses of the case study including research question, research design, justification of statistical choices made, interpretation of results and discussion. The case study is mandatory and will be reviewed by the lecturers.
- Reflection
The course also contains a reflection activities in which you will reflect on your learning experience. Reflection is of vital importance to improve learning outcomes. In these activities you will be asked to discuss how the course (so far) meets your personal learning objectives.

Course literature

The literature for this course has not yet been finalized. Details about required and recommended reading materials will be communicated to students at a later stage.

Prerequisite knowledge

Although active statistical knowledge is not a prerequisite, we expect some basic knowledge of statistics and mathematics acquired through, for example, courses in biostatistics in the bachelor program or self-study.

It is assumed that you

- are familiar with the concepts of a "population" versus a "sample"
- can interpret data presented in the form of a histogram, boxplot, frequency table, scatterplot or contingency table
- can calculate and interpret from a sequence of numbers: the mean, median, variance, standard deviation, range, interquartile range, standard error of the mean
- are familiar with the concepts probability and probability distributions (in particular the standard normal distribution).

Statistical software

In this course, both SPSS and R/RStudio can be used for statistical analyses. You are free to choose the software you prefer. You should, however, be able to interpret both SPSS and R output, as examples and web lectures may use either for illustration. This is not expected to pose difficulties for most students.

Those who have no experience with SPSS can get an introduction by watching the demo's of SPSS in the learning environment and/or by making the extra SPSS exercises.

How to get SPSS

- SPSS is almost everywhere available on the UU or UMC.
- Buy SPSS through the site Surfspot, which provides a great discount for being a student. Just login to their website with your student solis username and password. The complete program of SPSS is available for around 15 euro. Don't forget to buy a new license number every year.
- You may use SPSS through SolisWorkplace. It's free that way: <http://students.uu.nl/en/free-software> and <http://students.uu.nl/en/myworkplace>

If you would like to practice with R/RStudio or learn more about it, a list of helpful resources is provided below.

- R tutorial, developed by a UMC Utrecht team. For future R-users who have no or very limited prior experience with programming: <https://datasci-biostat.juliuscentrum.nl/knowledge-base/>. Instructions for installing R or RStudio can be found in this tutorial as well.
- R manuals: <http://cran.r-project.org/manuals.html>
- Another R tutorial: <http://tryr.codeschool.com/levels/1/challenges/1>. This tutorial would take a beginner 2-3 hours to complete. You get immediate feedback on the commands you type, and some explanation of why R code works the way it works.
- The R package (or "library") *swirl*. *swirl* is a software package for the R statistical programming language. Its purpose is to teach users statistics and R simultaneously and interactively: <http://swirlstats.com/students.html>
- R search engine: <http://rseek.org/>
- R examples with code and output: <http://www.statmethods.net/index.html>

Examination

The examination consists of two parts, namely:

1. a case study (25% of final grade)
2. a final exam (75% of final grade)

The grade for the final test and the case study must both be at least 5 and the final grade at least a 5.5 to pass for the course. Furthermore, active attendance is mandatory.

Attendance

You are expected to actively participate in all learning units, the case study and the exam.

Active participation means that you have 24/7 access to the course, but we expect that you have completed the associated learning unit within the given time period. To do this, schedule enough time in your agenda for this course. Only in this way we can ensure that no-one will run behind. We have to warn you that when you are lagging behind structurally and/or still behind at the start of the

case study, that you can lose your right to obtain a grade. It also means: finalize all learning units, actively contribute to the forums, assignments and discussions.

Case study

You will do the case study in groups of maximum three students. These groups are formed based on the students who have chosen for the same case study. The deadline for the case study is mentioned on the online environment of the course. All group members are responsible for the contents and quality of the report of the case study.

As a group you have to analyze a case study and write a (detailed) report of the statistical analyses of the case study including research question, research design, justification of statistical choices made, interpretation of results and discussion. Start your report with a short summary of the most important results, including effect sizes.

Exam

The exam consists of a combination of closed (multiple-choice or fill-in) and open questions. Note that mastery of the SPSS program or R is *not* tested during the exam. You are expected to be able to interpret *output* from SPSS or R.

The exam is 'open-book', which means that you are allowed to bring any notes and books. A (graphical) pocket calculator is allowed. Phones, tablets and laptops are **not** allowed.

Practice exams will be made available on the learning platform.

Evaluation

At the end of the course you will be asked to fill in a course evaluation. We kindly request that you take the time to fill in and hand in the evaluation form. Comments and questions during the course are also encouraged!