

## Biostatistics for Researchers

Programme is subject to alteration

Course Director: Rebecca Stellato

### Week 1

Weekend before the start of the course		
Time	Activity	Description
	Key pick up	You will find the exact key pick up location in the pre-departure information, which becomes available after you have paid the course fee.

Monday, July 07 2024		
Time	Activity	Description
09:30 – 09:45	Introduction	Welcome and Introduction to the Course
09:45 – 12:30	Lecture 1 & labs*: Introductory Case Studies	We explain the typical steps in a statistical analysis: from a first inspection of the data, to a description of the data in figures and plots, to some first preliminary conclusions with respect to the research question. We will do so using introductory case studies, in which real-life data sets are (partly) analyzed. These case studies will be used throughout the course to illustrate different statistical concepts. There is also ample time to practice with the software (R and/or SPSS).
	Lunch	
13:30 – 17:00	Lecture 1 & labs (continued)	

Tuesday, July 08 2024		
Time	Activity	Description
09:30 – 12:30	Lecture 2 & labs*: Estimation and Testing	The concepts of estimation (confidence intervals) and statistical testing are introduced, using examples with continuous outcomes. We use the t-tests for one and two (paired and independent) samples and their corresponding confidence intervals to introduce the idea of testing and estimation.
	Lunch	
13:30 – 17:00	Lecture 2 & labs (continued)	

Wednesday, July 09 2024		
Time	Activity	Description
09:30 – 12:30	Lecture 3 & labs*: Comparison of more than two means	In practice, the design of experiments is quite often more complicated than just comparing two groups, as we do on day 2. If the outcome or dependent variable is continuous (as with t-tests), and the dependent variable is a "factor" (categorical variable) with more than 2 levels, we can use one-way analysis of variance (ANOVA) to analyze the data. We will examine the overall ANOVA test for any difference among the groups, and also introduce a few methods of post-hoc testing as well as the concept of correcting for multiple testing. Finally, since every statistical test has its own assumptions. The assumptions of the tests used so far will also be covered, as well as several possible solutions for when these assumptions are violated.
	Lunch	
13:30 – 17:00	Lecture 3 & labs (continued)	

For the most current information about the course and Social Programme, please visit the [Utrecht Summer School website!](#)

**Thursday, July 10 2024**

Time	Activity	Description
09:30 – 12:30	Lecture 4 & labs*: Association in continuous data	Days 2 and 3 were limited to research questions with respect to the comparison of means, i.e. different types of t-tests and ANOVA. If we have two continuous variables and we want to study whether these variables are related, correlation and regression analysis are two statistical techniques that can be used. Both correlation and regression analysis look at the linear association between 2 continuous variables but there are substantial differences. We look for a linear association and express the direction and the strength of this linear association in a correlation coefficient. If on the other hand we want to know predict one variable with the other, we might use linear regression.
	Lunch	
13:30 – 17:00	Lecture 4 & labs (continued)	

**Friday, July 11 2024**

Time	Activity	Description
09:30 – 12:30	Lecture 5 & labs*: Comparison and association in categorical data	So far, we have limited ourselves to statistical techniques that are especially applicable with a continuous outcome. However, outcome variables are not always continuous, they might also be categorical. In the case of categorical outcome variables, estimation is related to sample proportions. With these sample proportions we estimate population probabilities of the several specific outcomes. When we are dealing with two groups, the parameter of interest is focused on the difference between proportions. Both should be presented with a confidence interval for the proportion or difference in proportions. We also cover hypothesis testing for the mentioned research designs. We introduce a binomial test for one sample, McNemar's test for two paired proportions, and the chi-square and Fisher's exact tests for contingency tables.
	Lunch	
13:30 – 17:00	Lecture 5 & labs (continued)	

**Week 2**

**Monday, July 14 2024**

Time	Activity	Description
09:45 – 12:30	Lecture 6 & labs*: Modeling continuous data	When we have one continuous outcome and more than one explanatory variable, we use multiple regression. We start with two continuous explanatory variables, then expand to more than two. Model building techniques are introduced, along with a brief discussion of when (not) to use model building. The assumptions of multiple regression are addressed, along with how to check those assumptions.
	Lunch	
13:30 – 17:00	Lecture 6 & labs (continued)	

## Tuesday, July 15 2024

Time	Activity	Description
09:30 – 12:30	Lecture 7 & labs*: Model building Issues	What if you want to include categorical explanatory variables in your multiple (linear) regression? What if you think there might be effect modification? Today, several issues related to model building (applicable to multiple linear regression, but also to logistic and Cox models in the coming days) will be addressed. Attention is paid to the use and interpretation of “dummy” variables and statistical interaction terms.
	Lunch	
13:30 – 17:00	Lecture 7 & labs (continued)	

## Wednesday, July 16 2024

Time	Activity	Description
09:30 – 12:30	Lecture 8 & labs*: Modeling binary data	When we have a dichotomous outcome variable and one or more explanatory variables, we use logistic regression. The concepts behind logistic regression are addressed, with emphasis on interpretation of results. Likelihood-based methods (likelihood ratio test and AIC) are introduced.
	Lunch	
13:30 – 17:00	Lecture 8 & labs (continued)	

## Thursday, July 17 2024

Time	Activity	Description
09:30 – 12:30	Lecture 9 & labs*: Survival Analysis	You will learn about the special features of “survival” or “time-to-event” data, and some of the commonly used methods for analyzing this special type of data. An example (survival times of women with advanced ovarian cancer) will be used in the lectures, and you will have a chance to practice the techniques on several other time-to-event datasets. At the end of the last lecture, a summary will be given of the course.
	Lunch	
13:30 – 17:00	Lecture 9 & labs (continued)	

## Friday, July 18 2024

Time	Activity	Description
09:30 – 13:00	Group assignment	On the last day of the course, you show us what you have learned! In small groups ( $\pm 3$ students) you will analyze a case study in the morning session. You will also be asked to examine a second case study and be the first to ask questions to the presenting group in the afternoon.
	Lunch	
14:00 – 17:00	Group presentations	In the afternoon your group presents their case to the class. There will be time for discussion and feedback for each presentation.

\* Days 1-9 are a mix of lectures and computer labs/practice in SPSS and/or R. In the afternoons, time permitting, you can also ask the lecturers – and your fellow students – questions about your own data analysis.