# Overview

This course covers the basics of statistics and data analysis. The ability to extract insights from data is an essential skill for both academic and non-academic work, and “data literacy” is increasingly important in a world where data are collected about every aspect of our lives. After completing this course, you will be able to independently analyze data, interpret and report your findings, and assess the results of analyses performed by others, such as you might find in scientific articles.

This GitBook contains all relevant information about this course. It is assumed that every student reads it carefully. If you have any questions, first consult this GitBook, then ask a fellow student, and only if your question is still not answered, then contact the course coordinator.

Communication about the course occurs through [Canvas](https://tilburguniversity.instructure.com/courses/12004) (Login with your student ID and password).

## Course Description

In the course, the following techniques will be discussed:

**All Majors**

* Information on the use of SPSS and interpretation of the output.
* Descriptive statistics;
* Normal distribution; standard scores;
* Sampling distributions; Z and t distributions;
* Hypothesis tests and confidence intervals for the mean.
* The power of a statistical test.
* One way Between Subjects Analysis of Variance.
* Linear regression analysis

**Major Business and Economics**

* Moderation
* Logistic regression

**Major Cognitive Neuroscience**

* Post hoc tests
* Factorial Analysis of Variance.
* ANCOVA
* Analysis of Repeated Measures.

**Major Social Sciences**

* Moderation
* Reliability analysis
* PCA and Factor Analysis.

## Learning goals

After taking this course, students will be able to…

### All majors

1. compute and interpret commonly used descriptive statistics such as the sample mean, the median, the mode, variance and standard deviation, the standard error, and the correlation coefficient.
2. recognize different probability distributions such as the normal distribution, and make computations for these probability distributions.
3. explain the essential aspects of null-hypothesis significance testing, including sampling distributions, Type I and Type II errors, one-tailed versus two-tailed testing, and statistical power.
4. apply different statistical tests such as the Z-test, the one sample t-test, the one way Between Subjects Analysis of Variance test, and statistical tests related to (multiple) linear regression analysis with continuous and categorical predictors; and clarify the statistical and/or methodological assumptions that apply to the techniques that are discussed in this course.
5. explain basic concepts in regression analysis, including: linear association, least-squares estimation, explained variance, Multiple R, multiple correlation, adjusted R-square, raw and standardized regression coefficients, model-comparison tests, predicted scores, residuals and the assumptions;
6. choose the appropriate analysis technique for answering a specific research problem from the range of techniques that are covered in the course.
7. use the software package SPSS to perform several statistical data analyses and be able to correctly interpret and report the output to an informed audience (e.g., Liberal arts students, researchers from the social sciences/business and economics/cognitive neuroscience).
8. draw valid conclusions from the results of empirical data analyses given specific research questions envisaged.

### Major Business and Economics

1. apply statistical tests in the context of multiple linear regression models with interaction terms and logistic regression models; interpret the corresponding output.
2. describe the concepts of probabilities, odds and logits; describe the relationship between the three scales; transform one into another (formulae are provided).

### Major Cognitive Neuroscience

1. apply statistical tests in the context of factorial ANOVA, ANCOVA and Analysis of Repeated measures; interpret the corresponding output; and calculate and interpret effect size estimates relevant for these statistical techniques (e.g., (partial) eta squared)

### Major Social Sciences

1. apply statistical tests in the context of multiple linear regression models with interaction terms and interpret the corresponding output.
2. gauge the reliability of measurements from questionnaires and identify problematic items.
3. explore the dimensionality of questionnaire data.

The **course schedule** is available at [My Timetable](https://rooster.uvt.nl/schedule). The information below might go out of date. For a general overview of the content, see below:

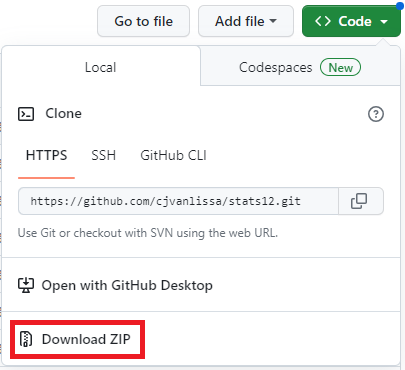
## Using this GitBook

**You do not need a book for this course!**

All essential information is contained within this GitBook. To ensure that you always have access to the GitBook, it is recommended that you download it to your local computer as follows:

1. Go to <https://github.com/cjvanlissa/stats12>
2. Click the green Code button
3. Select “Download ZIP” (see figure below)
4. Save the ZIP archive to your drive
5. Right-click the downloaded file, and select “Extract here” (or similar option, depending on your operating system)
6. You should now have a folder with the contents of the book.
7. To launch the book in your browser, open the file docs\index.html.
8. The “data” folder contains all SPSS datasets you need for the course.
9. The “pdfs” folder contains all PDF files you need for the course.

**It is possible that the book will be updated during the course. If this happens, I will notify you via Canvas to re-download the book.**



## Software

During lab sessions, you work on the exercises and your portfolio using the commercial SPSS software installed on university computers.

If you want to use your own computer instead, you might consider trying some free alternatives to SPSS:

* PSPP, which is designed to be nearly identical to SPSS with all the same basic functionality: <https://www.gnu.org/software/pspp/pspp.html>
* JASP, which is more modern, looks nicer and is very easy to use – but looks less similar to SPSS: <https://jasp-stats.org/>

## Attendance

Attendance is mandatory based on our experience that students who actively participate tend to pass the course, whereas those who do not tend to drop out or fail. All lectures and practicals ‘build’ on each other, so if you have to miss either one, absolutely make sure you have caught up with the materials before the next session.

## Study Load

Below is a breakdown of the expected study load:

| Activity | Duration | Times | Total |
| --- | --- | --- | --- |
| Synchronous |  |  |  |
| Lectures | 2 | 14 | 28 |
| Tutorials | 2 | 14 | 28 |
| Asynchronous |  |  |  |
| Knowledge clips | 1 | 14 | 14 |
| Formative tests | 1 | 14 | 14 |
| Reading time | 1 | 14 | 14 |
| Studying | 1 | 1 | 37 |
| Assessment type(s) |  |  |  |
| Portfolio | 10 | 3 | 30 |
| Exam | 3 | 1 | 3 |
| Total |  |  | 168 |
| ECTS |  |  | 6 |

## Staff

**Coordinator:**

[dr. Caspar J. van Lissa](https://www.tilburguniversity.edu/nl/medewerkers/c-j-vanlissa)

**Lab sessions**

(Thu) Tra Lê

### Why group assignments?

Contact with fellow students is a key aspect of the university experience. We want to stimulate you to engage with the material and with one another. Therefore, the portfolio assignments are made in groups. There are also aspects of learning in groups that can really improve your knowledge, like peer feedback. To ensure that every group member pulls their weight, the final exam tests each student’s individual comprehension of all material covered in the portfolios.

Groups comprise 3-5 members and are assigned randomly when the course starts. However, it is allowed to switch with a consenting member of another group, or to join/merge with another small group if your group has become smaller than 3 members. There are three portfolio registration deadlines. At this point, one group member submits the definitive group composition via a Google form.

### Why use portfolio assessment?

Portfolio assignments are well-suited for a skills-based course like Statistics 1 & 2. They also take a lot of the pressure off because you can work at your own pace, and keep improving the work until it is good enough. We entrust you with the responsibility of making these portfolio assignments in good faith, without instrumental assistance from outside your group or plagiarism, so I kindly ask you to make good on this trust, and hand in original work to show what you’ve learned.

## Grading

Your grade is based on three portfolio assignments made in groups, and one individual exam to test comprehension of the material covered in the portfolios.

A grade of 5.5 or higher is required for each component to pass the course.

### Portfolios 40% (3 x 13.3%)

You work on the portfolio assignments with your group, both during the lab sessions and outside of class. For each assignment, register your group membership before the set deadline at <http://tiny.cc/stats12_portfolio>. You hand in your group’s portfolio assignment before the set deadline, at which point it is graded. If your grade is below the passing level of 5.5, your group will have the opportunity to revise the portfolio based on teacher feedback to receive a maximum grade of 6.

Groups should equally distribute the work load for the portfolio assignments. In case doubts are raised about the equal distribution of labor in a particular group, the portfolio assignment in question will be supplemented with individual oral examination and an individual grade, which can not exceed the original grade for the group assignment. In other words, failing to distribute the work properly can not have positive effects, but it can have negative effects on your grade. To prevent this, make clear agreements about the distribution of work with your group mates.

### Exam 60%

To make sure that all students are equally involved in the making of the portfolio assignments, an individual exam assesses comprehension of the material covered therein. It is a digital multiple choice exam. Note that some of the questions are developed by you, the students (see portfolio)! You thus have an advantage if you develop questions that are good enough to be on the exam.

You may bring all course materials to the exam, including the portfolio. The exam consists of a common part and a major-specific part. Note: As per university policy, a guessing correction is applied to your grade.

**Note: I recommend that you download the GitBook onto a USB stick so you have access during your exam!**

## Assignments

Below is a description of the assignments. For each assignment, every element labeled with a lower case letter is graded fail (0 points), pass (1 point), or excellent (1.5 points). Grades are summed for each assignment, and rescaled from 1-10. The final grade is the average across assignments of the rescaled grades. Note the stated word limit for each section. If you can write a good report with fewer words, that’s fine. If you exceed the word limit however, your grade for that section cannot exceed a pass (1 point).

The focus of the assignments should me on motivating, reporting, interpreting, and discussing your analyses. You will get a good grade for well-reasoned and discussed analyses.

See the Appendices section to access data sources for the assignment.

### Assignment 1

Descriptive statistics and statistical inference

1. Select at least three variables for further analysis, and motivate your selection based on theory, using at least one reference to explain why are you interested in the properties of the selected variables (150 words)
   1. Include one continuous variable
   2. Include one nominal variable
   3. Include one ordinal variable
2. Describe the dataset (200 words + tables/figures)
   1. Use appropriate univariate descriptive statistics for all variables
   2. Plot data using appropriate plots
   3. Include at least one frequency- or crosstable
3. For a continuous variable:
   1. Select one or more values with clinical/societal/statistical relevance (i.e., provide some justification for the choice of value)
   2. Using probability calculus, calculate and report the probability of observing values that fall below/between/exceed the chosen value(s)
4. For a continuous variable:
   1. Formulate a specific null- and alternative hypothesis
   2. Report a one-sample t-test or Z-test for the specific null-hypothesis
   3. Calculate the probability of comitting a Type II error
5. Discuss your analyses (300 words)
   1. Explain your rationale for important modeling decisions
   2. Motivate your choice for the type of statistics and analyses
   3. Discuss assumptions
   4. Discuss what you have learned from it and how you might improve it
6. Use APA style throughout your report
7. Create four multiple choice questions about the first 4 weeks. You can use some of your analyses as inspiration, or make the questions purely about understanding. The best questions will be part of the real exam, so you have an advantage if you make questions that are good, but that you can still answer correctly!
   1. Introduction to statistics
   2. Probability
   3. Sampling distribution
   4. Testing

The multiple choice questions must be in plain text (no images or special characters, in the format below. You can just copy-paste this into the (Word?) document of your assignment:

question = "This is the question",  
answer = "This is the correct answer",  
distractor1 = "This is an incorrect distractor answer",  
distractor2 = "This is an incorrect distractor answer",  
distractor3 = "This is an incorrect distractor answer",  
expanation = "Why is the correct answer correct?"

### Assignment 2

General linear model

1. Select at least three variables for further analysis, and using at least one reference, explain what research questions you will investigate and what hypotheses you will test (150 words)
   1. Include one continuous outcome variable
   2. Include one continuous predictor
   3. Include one nominal or ordinal predictor
2. Construct a model with only the continuous predictor (200 words)
   1. Report and interpret the different sums of squares
   2. Report and interpret the explained variance
   3. Conduct a separate correlation analysis. Compare the results with the regression analysis.
3. Construct a model with only the categorical predictor (200 words)
   1. Report and interpret the model results
   2. Conduct a separate ANOVA or t-test with the same variables, whichever one is suitable. Compare the results with the regression analysis.
4. Construct a model with both the continuous and categorical predictor (200 words)
   1. Report and interpret the model results
   2. Conduct and report a nested model test
5. Discuss your analyses (300 words)
   1. Explain your rationale for important modeling decisions
   2. Motivate your choice for the type of statistics and analyses
   3. Discuss assumptions
   4. Discuss what you have learned from it and how you might improve it
6. Use APA style throughout your report
7. Create four multiple choice questions about the first 4 weeks. You can use some of your analyses as inspiration, or make the questions purely about understanding. The best questions will be part of the real exam, so you have an advantage if you make questions that are good, but that you can still answer correctly!
   1. Linear regression, correlation, or sums of squares
   2. Categorical predictors in linear regression (e.g., t-test or ANOVA)
   3. Multiple regression
   4. Nested models

The multiple choice questions must be in plain text (no images or special characters, in the format below. You can just copy-paste this into the (Word?) document of your assignment:

question = "This is the question",  
answer = "This is the correct answer",  
distractor1 = "This is an incorrect distractor answer",  
distractor2 = "This is an incorrect distractor answer",  
distractor3 = "This is an incorrect distractor answer",  
expanation = "Why is the correct answer correct?"

### Assignment 3 (BE)

Logistic regression

1. Select at least three variables for further analysis, and using at least one reference, explain what research questions you will investigate and what hypotheses you will test (150 words)
   1. Choose a binary outcome variable
   2. Include at least two predictors
2. Construct a model with only main effects (200 words)
   1. Report and interpret the results
3. Construct a model with the interaction effect (200 words)
   1. Report and interpret the model results
   2. Conduct and report a nested model test
4. Throughout your report, report confidence intervals. For at least one hypothesis, use interval testing (optionally, alongside p-value based testing).
5. Discuss your analyses (300 words)
   1. Explain your rationale for important modeling decisions
   2. Motivate your choice for the type of statistics and analyses
   3. Discuss assumptions
   4. Discuss what you have learned from it and how you might improve it
6. Use APA style throughout your report
7. Create four multiple choice questions about the major-specific material. The best questions will be part of the real exam, so you have an advantage if you make questions that are good, but that you can still answer correctly!

The multiple choice questions must be in plain text (no images or special characters, in the format below. You can just copy-paste this into the (Word?) document of your assignment:

question = "This is the question",  
answer = "This is the correct answer",  
distractor1 = "This is an incorrect distractor answer",  
distractor2 = "This is an incorrect distractor answer",  
distractor3 = "This is an incorrect distractor answer",  
expanation = "Why is the correct answer correct?"

### Assignment 3 (SS)

1. Select at least three constructs for further analysis, and using at least one reference, explain what research questions you will investigate and what hypotheses you will test (150 words)
   1. At least one of these constructs must be a scale with 3+ items
   2. Include at least two predictors
2. Perform some kind of data reduction analysis (200 words)
   1. Report and interpret the results
3. Perform reliability analysis (200 words)
   1. Report and interpret the results
4. Construct a model with only main effects (200 words)
   1. Report and interpret the results
5. Construct a model with the interaction effect (200 words)
   1. Report and interpret the model results
   2. Conduct and report a nested model test
6. Discuss your analyses (300 words)
   1. Explain your rationale for important modeling decisions
   2. Motivate your choice for the type of statistics and analyses
   3. Discuss assumptions
   4. Discuss what you have learned from it and how you might improve it
7. Use APA style throughout your report
8. Create four multiple choice questions about the major-specific material. The best questions will be part of the real exam, so you have an advantage if you make questions that are good, but that you can still answer correctly!

The multiple choice questions must be in plain text (no images or special characters, in the format below. You can just copy-paste this into the (Word?) document of your assignment:

question = "This is the question",  
answer = "This is the correct answer",  
distractor1 = "This is an incorrect distractor answer",  
distractor2 = "This is an incorrect distractor answer",  
distractor3 = "This is an incorrect distractor answer",  
expanation = "Why is the correct answer correct?"

### Assignment 3 (CN)

1. Select at least three variables for further analysis, and using at least one reference, explain what research questions you will investigate and what hypotheses you will test (150 words)
   1. At least one of these variables must be categorical with 3+ levels
   2. Include at least two predictors
2. Perform an ANOVA (200 words)
   1. Report and interpret the results
   2. Use either a different kind of coding, or post-hoc tests, or planned comparisons
3. Perform a factorial ANOVA (200 words)
   1. Report and interpret the results
   2. Report on the interaction effect
4. Include a continuous control variable in one of the preceding models (200 words)
   1. Report and interpret the results
5. Discuss your analyses (300 words)
   1. Explain your rationale for important modeling decisions
   2. Motivate your choice for the type of statistics and analyses
   3. Discuss assumptions
   4. Discuss what you have learned from it and how you might improve it
6. Use APA style throughout your report
7. Create four multiple choice questions about the major-specific material. The best questions will be part of the real exam, so you have an advantage if you make questions that are good, but that you can still answer correctly!

The multiple choice questions must be in plain text (no images or special characters, in the format below. You can just copy-paste this into the (Word?) document of your assignment:

question = "This is the question",  
answer = "This is the correct answer",  
distractor1 = "This is an incorrect distractor answer",  
distractor2 = "This is an incorrect distractor answer",  
distractor3 = "This is an incorrect distractor answer",  
expanation = "Why is the correct answer correct?"

## Use of Large Language Models (LLMs)

Be warned: LLMs learn from all text on the internet, which includes a *lot* of incorrect information about statistics. As a result, my experience is that e.g. ChatGPT produces a lot of incorrect output for statistics assignments.

If you decide to use LLMs, it is your responsibility to thoroughly check its output for logical consistency and correctness. You may not yet have the level of expertise required to know when ChatGPT generates irrelevant nonsense - but the teacher who grades your work does. Consider this carefully when deciding what makes more sense: doing your work manually, making sure each step is correct - or outsourcing it to AI, and then checking its work before submitting.

## Credit

This book was authored by Caspar J. Van Lissa. Its code and layout are derived from Lisa DeBruine’s “booktem”:

DeBruine L, Lakens D (2023). *booktem: Methods Book Template*. https://github.com/debruine/booktem, https://debruine.github.io/booktem/.

Also see: https://psyteachr.github.io/

library(officer)  
library(revise)  
format\_text <- function(doc, x, sections){  
 # x = "coursedesc"  
 txt <- sections$sections[[x]]  
 txt <- gsub("\\\*\\\*(?![a-zA-Z])", "\\\*\\\*\n", txt, perl = T)  
 txt <- strsplit(txt, split = "\n")[[1]]  
 for(i in txt){  
 if(grepl("^(\\\*\\\*|##)", i)){  
 i <- gsub("\*", "", i, fixed = T)  
 i <- gsub("#", "", i, fixed = T)  
 i <- trimws(i)  
   
 doc <- body\_add(doc, fpar(ftext(i, fp\_text(bold = TRUE))))  
 } else {  
 doc <- officer::body\_add\_par(doc, i)  
 }  
 }  
 doc  
}  
sections <- revise::read\_manuscript("index.qmd", to\_envir = FALSE)  
file.remove("Syllabus\_Statistis\_1\_2.docx")  
doc <- read\_docx("c:/gitbook\_stats12/LAS Syllabus Template.docx")  
doc <- body\_replace\_all\_text(doc, "XXcoursename", "Statistics 1 and 2")  
doc <- body\_replace\_all\_text(doc, "XXcoursecode", "840099-B-6")  
doc <- body\_replace\_all\_text(doc, "XXcurrentyear", paste0(format(Sys.Date(), format = "%Y"), "-", (as.integer(format(Sys.Date(), format = "%Y"))+1)))  
  
doc <- officer::cursor\_reach(doc, keyword = "XXcoursedesc")  
doc <- format\_text(doc, "coursedesc", sections)  
doc <- body\_replace\_all\_text(doc, "XXcoursedesc", "")  
  
  
doc <- officer::cursor\_reach(doc, keyword = "XXlearninggoals")  
doc <- format\_text(doc, "learninggoals", sections)  
doc <- body\_replace\_all\_text(doc, "XXlearninggoals", "")  
  
doc <- officer::cursor\_reach(doc, keyword = "XXstudyload")  
tab\_load[] <- lapply(tab\_load, as.character)  
tab\_load[is.na(tab\_load)] <- ""  
doc <- officer::body\_add\_table(doc, tab\_load)  
doc <- body\_replace\_all\_text(doc, "XXstudyload", "")  
  
  
  
doc <- officer::cursor\_reach(doc, keyword = "XXschedule")  
schedule <- schedule[ordr, ]  
#tab\_load[is.na(tab\_load)] <- ""  
doc <- officer::body\_add\_table(doc, schedule)  
doc <- body\_replace\_all\_text(doc, "XXschedule", "")  
  
doc <- officer::cursor\_reach(doc, keyword = "XXgrading")  
doc <- format\_text(doc, "grading", sections)  
doc <- body\_replace\_all\_text(doc, "XXgrading", "")  
  
print(doc, target="Syllabus\_Statistis\_1\_2.docx")