Instructor: Bernhard Angele, PhD
Email: bangele@bournemouth.ac.uk

1 Course overview and goals

This course is designed to give postgraduate students in Psychology a solid foundation in the inferential statistics typically used in Psychology. In typical undergraduate statistics courses in Psychology, the focus is on teaching students the very basics: What are the required tests, and how does one apply them (in statistical software) without making errors? This course aims to go beyond this. It assumes that students have learned about the basic statistical methods before when obtaining their undergraduate degree and that they know how to apply them in SPSS. This course aims to supplement these basic skills with more in-depth knowledge: Why do we do hypothesis testing? What is at stake? Why do the statistical tests work? What happens if assumptions are violated? There are two possible ways to address these questions. The first approach is analytic: it involves deriving the different test statistics and mathematically proving that they are related to the properties of the population being studies. This approach is usually very challenging to beginning postgraduate students.

This course favours a second approach: studying the mathematical properties of the different statistical tests not by derivations and proofs, but by simulation. It tries to provide students with the tools needed to build intuitions about these tests and aims to instill a "What if?" approach to statistics. These intuitions can later be used to gain an understanding of the underlying mathematics, if the student so desires. Even so, having an intuitive understanding of how hypothesis tests work and what influences their outcome (rather than blindly applying step-by-step procedures)is invaluable in recognising and dealing with the statistical issues that will inevitably come up in any serious research endeavor.

Of course, gaining a deeper understanding of how data and statistics work comes at a price: in this case, students will have to learn how to use a new tool, namely the R statistical software system. While it is not impossible to manipulate data and do simulations in the SPSS language, it is much more difficult to use than the R language. For those students who have only used the SPSS menu interface and have never worked with command-line or script-based programming languages before, this will be a challenge.

However, the skills gained with R will likely be highly useful to students in many areas: basic programming skills are extremely helpful both in research and in everyday situations. Combined with basic data manipulation skills, they can save one many hours of time in preparing experiments, collecting data, reading the data into statistical software, and cleaning and preparing it for analysis.

Furthermore, R is on its way to becoming the predominant statistical software in many research fields and even private companies. It is quite likely that students will encounter R in some form or shape in the future, and being familiar with it can increase employability quite a bit.

2 Statistical software

In this course, we will mainly use the R statistical software. It will be very useful for students to follow along with the analyses as I perform them in class. Also, all homework and class assignments will be done in R. Because of this, it will be highly recommended for students to install R on their laptop computers and bring them to class. R is freely available at *http://cran.r-project.org/*. Additionally, students should install *RStudio* (desktop version), which is also available free of charge. RStudio is a major help in interacting with R and writing script files. Detailed installation instructions are available in Navarro's Textbook (see below), Chapter 3.1. Students should spend some time before the first class session to familiarise themselves with the R software and the RStudio environment. A great place to start is Try R by Codeschool. This tutorial doesn't even require students to have R installed yet, as it uses an online R interpreter. Students are strongly encouraged to work through the 7 short chapters before the first class session.

3 Textbooks (required)

1. Shravan Vasishth's statistics lecture notes.

This is an updated version of

Vasishth, S., & Broe, M. (2010). *The Foundations of Statistics: A Simulation-based Approach: A Simulation-Based Approach.* Springer.

2. Navarro, D. *Learning statistics with R: A tutorial for psychology students and other beginners (Version 0.4).*

Both of these books are available on the authors' websites free of charge (see links).

Optional reading:

3. Field, Andy, Miles, Jeremy and Field, Zoë (2012) *Discovering Statistics Using R.* Sage Publications, London. ISBN 978-1446200469

4. Field, Andy (2013) *Discovering statistics using IBM SPSS Statistics: and sex and drugs and rock n roll (4th edition).* Sage Publications, London. ISBN 9781446249178

5. Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22(11)*, 1359-1366.

4   Lectures

Lectures will take place every Thursday from October 2nd to November 13. Note that, on some Thursdays, there will be two lectures, one from 10 am to 12 pm and one from 1 pm to 4 pm. The last hour of every class day (3–4 pm) will be in independent study period: students will be asked to work on a homework assignment independently. I will be available for questions and help during these independent study periods.

5   Course website and homework

The easiest way to get the newest versions of the course materials is at the following URL (they will also be available on muBU):

https://github.com/bernhardangele/BU-adv-stats.

Homework will be assigned at the end of each class and will be due at the beginning of the next class. Homework and coursework assignments will be submitted via Turnitin on MyBU.

6   Marking policy

Your mark will consist of three components: One short coursework assignment (40%), one longer coursework assignment (50%), and the homework (10%).

6.1   Homework

Homework assignments will not be marked in detail; all you have to do is turn something in by the beginning of the next class in order to get a point. If you abuse this policy (e.g. by turning in an empty file or turning in the same file twice), you will lose all homework points. Homework points translate into marks as follows:

Mark	Number of homework assignments turned in
100	5+
70	4
60	3
50	2
40	1
0	0

6.2   Coursework

Coursework assignments will involve analysing one or two data sets. Each data set will require data screening, data aggregation, performing hypothesis tests, and reporting the results in essay form. The assignments will be marked as follows: Each assignment will receive three marks:

- **Statistical methods (70%)**. Does the assignment use the correct statistical methods? Are appropriate statistics reported? Is the conclusion warranted given the hypothesis test results? Are potential issues discussed properly?

- **Writing (20%)**. Are the hypothesis tests and statistics presented in an understandable way? Are the data in question introduced and described properly? Are conclusions and interpretations described clearly and concisely (so that a non-psychologist could understand them)?

- **Formatting and style (10%)**. Is the document formatted properly? Are conventions such as APA citation style, number and table formatting followed? (If students use Markdown and Knitr as suggested, there should be no major problems here.)

7   Course Outline:

Date and Topic

Date and Topic

Oct 2nd   **1**

Double class session: 10–12 am and 1–3 pm
Introduction to Advanced Statistics
Introduction to the R software
Introduction to the simulation approach
Descriptive statistics Probability and inferential statistics
Why is the normal distribution so important?
The central limit theorem
Estimating population parameters from a sample
R basics: Plotting, reading data files, basic programming
Readings:
Vasishth, Chapter 1–2
Navarro, Chapters 1, 3 – 11

9th   **2**

Double class session: 10–12 am and 1–3 pm
Hypothesis testing:
Confidence intervals
z-test
t-test: one-sample, two sample, unequal variances, repeated measures
Power
Readings:
Navarro, Chapters 12–13
Vasishth, Chapter 3

16th   **3**

Analysis of Variance
One-way ANOVA
Planned comparisons
Assumption tests
Writing reports using R Markdown and Knitr
Navarro, Chapters 14–16

DATE AND TOPIC
23rd   **4** Multiway ANOVA Repeated measures ANOVA Power in ANOVAs Reporting ANOVA-type experiments *Readings:* *Navarro, Chapters 14–16*
30th   **5** Linear regression Generalised linear model Contrast coding ANCOVA *Readings: Vasishth, Chapter 4.1*
Nov 6th *Reading Week, no class*
13th   **6** *Double class session: 10–12 am and 1–3 pm* Transformations Logistic regression Linear mixed models Nonparametric tests Power "Researcher degrees of freedom" *Readings:* *Vasishth, Chapter 4* *Simmons et al. (2011)*

Date and Topic
20th *Coursework Week, no class*

The **first coursework assignment** will be set on **Thurday, October 23rd** and will be due on **Thursday, November 13th at 12 pm**.

The **second coursework assignment** will be set on **Thurday, November 13th** and will be due on **Thursday, January 29th at 12 pm**.

Important: This document (including the calendar) is subject to change. Please always get the newest version on myBU.