Research Master Methodology and Statistics in the Behavioural and Social Sciences
Department: Methodology and Statistics

MSBBSS010 - Markup languages and reproducible programming in statistics (2020)

Coordinator: Dr. Gerko Vink

Course description
This course gives an overview of the state-of-the-art in statistical markup, reproducible programming and scientific digital representation. Students will get to know the professional field of statistical markup and its innovations and challenges. It consists of 6 meetings in which students will learn about markup languages (LaTeX and Markdown), learn efficient and reproducible programming with rMarkdown, experience developing Shiny web apps, get to know version control with Git and will create and maintain their own data archive repository and personal (business card) page through GitHub. Combining these lectures, the students get acquainted with different viewpoints on marking up statistical manuscripts, areas of innovation, and challenges that people face when working with, analysing and reporting (simulated) data. Knowledge obtained from this course will help students face multidimensional problems during their professional career.

Assignment
Students will individually choose one statistical topic and work on a manuscript about this topic. Students will need to perform calculations and program code for this manuscript. All work for the student needs to be combined in an easy understandable and insightful data archive and will need be posted on a personal GitHub repository. This end result will be graded on 1) Quality of the markup language skills, 2) Quality of the data archive and 3) Quality of the online repository.

Grading
Students will be evaluated on the following aspects:

1. Developing and publishing a research archive that contains code, data and a typeset manuscript following a markup language;
2. Developing and publishing a research archive that demonstrate a reproducible workflow.
3. Developing and publishing a personal repository page;

a. Students develop fundamental knowledge and understanding in the state of the art in statistical markup languages and reproducible programming (Knowledge and Understanding)
b. They apply their knowledge in a multi-disciplinary context to contemporary problems (Applying)
c. They can determine the most effective markup strategies to address a typesetting problem (Applying)
d. They can efficiently organise a reproducible programming process (Applying)
e. They can advise researchers in applying the current state of the art in markup and programming (Judgment)
f. They can produce reproducible repositories up to the standards of international programming and coding conventions and initiatives (Communication)
g. They can produce publications up to the typesetting standards of international peer-reviewed journals (Communication)
h. They are capable of autonomous scholarly self-development (Learning skills)

i. They give proof of being a responsible and scholarly professional (Learning skills)

After taking this course students can understand innovations in statistical markup, statistical simulation and reproducible research. Students are also able to approach challenges from different professional viewpoints. They have gained experience in marking up a professional manuscript and designing a state-of-the-art statistical archive in an open source repository.

Grading
To develop the necessary skills for completing the assignment and the presentation, 6 exercises must be made and submitted. These exercises are not graded, but students must fulfil them to pass the course.
The final grade is computed as follows

1. Markup manuscript (50%)
2. Research repository (40%)
3. Personal repository (10%)

In order to pass the course, the final grade must be 5.5 or higher, your contribution to the course should be sufficient and all assignments and practical assignments should be handed in and/or passed. Otherwise, additional work is required concerning the assignments and/or exercises you have failed.

Literature
Relevant documents that are used in the course and/or may provide useful additional reading will be placed on or referred to on Blackboard.

Instructions for preparing the repositories
The research repository has to be prepared as a supplementary archive that can serve as an extensive documentation of the research (e.g. as a supplement to be submitted to a journal). The archive has to be published in a public or private GitHub repository.

Time schedule
This course takes place in the first semester. The course will run over 6 meetings (Wednesdays 9-12am) during the semester, with the first meeting on September 14, 2020.

Prerequisites
Students will need their own laptop computer. Students should have experience in programming with R and should be familiar with the IDE RStudio.

Course schedule

Wk	Date	Topic	Exercise (Y/N)
38	14-Sep	Monte Carlo simulation and replication	Yes
40	28-Sep	Reproducible workflows	Yes
44	26-Oct	Typesetting equations with LaTeX	Yes
45	02-Nov	Version control and Github in depth	Yes
48	23-Nov	Presentations with RMarkdown	Yes
50	07-Dec	Github pages and Shiny	Yes

Course materials

All course materials can be found on https://www.gerkovink.com/markup/. A detailed overview of the course, topics, required readings, code and other materials can be obtained there.