PSM2 assignment

Applied Crime Analysis Project

The final project of this module is designed so that you can demonstrate that you have a good command of the statistical concepts covered in the lectures and tutorials and can put them to use in your own project. We will provide a dataset and your task is to address a research question stemming from crime and security science. You will also be able to demonstrate your understanding of open science practices through pre-registering your analyses, producing re-producible code, and documenting your data properly. You will have to write a report on your findings (a template will be provided) and have to submit the R code needed to reproduce your analysis in the form of an R Notebook.

Basic info

- Weight for final grade: 50%
- Learning outcomes tested:
    - understanding the logic of modelling and its value in understanding real-world phenomena
    - applying and understanding methods of statistical inference which can be applied when the assumptions of standard approaches are violated
    - understanding the null-hypothesis significance testing framework and the Bayesian approach
    - applying and understanding methods for testing and modelling datasets with particular structure
    - performing said analyses and demonstrate practical skills of data manipulation and cleaning, alongside the implementation of the statistical approaches introduced in the module
- Deadline: **29 March 2019** and **16 April 2019**.
- Feedback deadlines: **22 March 2019** for the 1-on-1 feedback (see below)
- Word count limit: 1500 words (excl. code supplement; do not exceed this word count limit!)

Project requirements

In this assessment you will work with a dataset of offenders who were charged with crimes resulting from gang activity. You can find details on that dataset here (pdf version).

Imagine you are working for a police force as an analyst. Your department is embracing an evidence-based and problem-driven approach to policing and asks you to work with the dataset to answer three questions:

1. Are there household characteristics that are related to an offender's household income 5 years after the initial measurement?
2. What (if any) is the relationship between an offender's gender and their punishment?
3. Which factors might help in predicting whether or not an offender found a job 5 years after the offense?
4. Is there are relationship between the family situation, whether or not the offender was victim of domestic abuse and whether or not the offender had previous mental health problems?
5. Does it make a difference to the household income if the mother or the father left, or is there no difference?

You are also asked to test one other research question that you can freely choose.

Your task is to answer each of these six (5 given + 1 of your own) questions. To do this, formulate hypotheses that allow you to answer these questions and preregister these hypotheses. You can formulate multiple hypotheses per question but we expect at least one.

Remember to preregister any data exclusions that you find meaningful, any data transformations, etc. and justify any deviations from your preregistered procedure.

The dataset will allow you to test all of these blocks but your task is to only use 4 of them.

Stepwise data release:

An important part of modern science is transparency and openness about the data collection and analytical process. Since the data are provided for you, you will focus exclusively on the Open Science aspect of the analysis instead. We specifically ask that you preregister your analyses using the OSF preregistration form and include **a link to the anonymised preregistration** in your assignment submission.

To help you better understand the data that you will work with, we will provide you with a pseudo-dataset that contains the same variables and structure as the full dataset but is smaller and does not necessarily have the same relationships among variables.

Your task for the preregistration is to detail all data exclusions, analytical steps, hypotheses, exploratory analyses, etc. (as detailed in the OSF preregsitration form). You will then present the preregistered analyses as "confirmatory analyses" in your report and will detail and justify any deviation from the preregistration. You can conduct non-preregistered analyses and report these as "exploratory analyses".

On the pseudo-dataset, you can also start writing your R code. The full dataset will be structured in the same way and any code that works on the pseudo-dataset will also work on the full dataset.

You can find the pseudo-dataset for the preregistration on Moodle. The individual datasets will be sent to you via email.

Deadline for preregistration: Your preregistration should be "frozen" the latest on the **29th of March**. We will release the full datset on the **30th of March** and you cannot preregister any analyses once the full dataset is released. Note that the preregistration is a requirement for this assessment.

Grading guidelines

Requirements to be graded:

- submit the 1-on-1 feedback form in time
- attend the feedback session
- freeze your preregistration publicly before the deadline (29 March)
- submit the report + code supplement before the deadline (16 April)

Grading criteria:

- Appropriateness of analytical methods – 15%
- Quality of preregistration – 15%
- Quality of analytical techniques – 20%
- Interpretation of findings – 20%
- Clarity of report – 10%
- Quality of report (layout, formatting) – 10%
- Quality of R code (clarity, documentation, reproducibility) – 10%

Feedback sessions

We will help you with an individualised feedback session.

1-on-1 feedback session: you will receive individualised feedback from both Bennett and Isabelle in a 1-on-1 session where we will help you with questions and give you final advice to fine-tune your project. These sessions will take 10 minutes per student and will be held on **26 March 2019** (timeslots to be arranged). Submit your feedback on Moodle through using this template (pdf version).

Deliverables

- Feedback report for 1-on-1 session

- Link to publicly frozen preregistration
- Report of your project: the report has a word count limit of 1,500 words and should contain the following sections
  - link to the preregistration
  - your hypotheses (you do not have to write an introduction)
  - method (i.e. what did you do with the data, which exclusions/transformations did you apply, which analytical models did you apply, . . .)
    * outline each method briefly assuming that the reader is not familiar with the analytical technique
  - data (i.e. brief details on the dataset)
  - analytical plan (i.e. clearly set out which analysis you used to test each of your hypotheses)
  - results
  - discussion (interpretation, answer to the question, limitations of your work)
- Code supplement: you also submit the code needed to fully reproduce your work in an R Notebook
  - the code should be clearly documented
  - the header should contain your student assessment ID (i.e. the code is fully anonymised)
  - the code must be reproducible in all steps