

Authentic analysis assessment

Find an empirical paper for which the authors have freely-shared the dataset but not their analysis scripts. Ideally this will link to your research interests, or be similar to your project analysis, but this is not required.

- A good way of finding a suitable paper would be to look through a journal known for publishing papers with open-source datasets, for example PlosOne. Another would be to search Zenodo or a similar repository for recently-published datasets, and look for where the articles reporting these data were published.
- To keep the assignment simple, choose a paper which uses statistical methods taught as part of the module, or which can be approximated by them. For example: a multiple regression, between-subjects Anova, or repeated measures Anova or mixed model.
- Some papers report many studies, or many different analyses, and may have technical details with which you are not familiar. If this is the case then you do NOT need to deal with the whole paper: Simply **focus on one of the primary analyses**. For example, if one of the main study hypotheses can be answered by a 2x2 Anova or mixed model, but there are lots of other analyses included in the paper, then focus on just this one analysis.

If you are unsure about the scope of this assignment (e.g. don't know how much of a paper to try and replicate) then please check with the module leader early in the process.

To check if a paper will be suitable, you can answer these questions:

1. Are the data available?
2. Is the design between-groups? (answer must be yes)
3. Are measurements at more than 2 timepoints per-participant? (answer should be no; if the paper reports "mixed effects" models or repeat-measures Anova it will not be suitable)
4. Is the outcome a continuous variable? (i.e. must not be a binary outcome)

Your task is to:

1. Imagine you were the author of the paper, just prior to collecting data for this study. Complete the form on 'aspredicted.org'. You don't actually need to use the aspredicted website for this — simply copy the heading structure from the form, and include your answers in your submitted manuscript. Comment on whether it is possible and/or straightforward to reconstruct what would have been in the pre-registration from what is written in the journal article.
2. Using `ggplot`, make at least one plot of the data which illustrate the main findings. Explain your design decisions for these plots: what features of the data were you trying to highlight, and how?
3. Replicate the statistical analysis using methods which are appropriate to the data. You do not need to use the *exact* method used in the original paper: For example, it's fine to use a mixed model in place of a repeated-measures Anova. Report your findings in APA format. Compare/contrast your results with the published manuscript.
4. Document your responses and version of the analysis in an Rmd file with accompanying dataset, suitable for upload to Zenodo or OSF.io.
5. Reflect on the extent to which the concept of 'researcher degrees of freedom' should influence our understanding of your chosen study.

Important notes:

- Don't assume that a published dataset and manuscript can be easily replicated simply because they have been shared: ***It may well be the case that it is hard to replicate the findings and this is OK (and interesting!) You will not lose marks simply because you cannot replicate the published findings.***
- You do not need to use the *exact* method used in the original paper; the point of the exercise is to apply methods we have learned in an appropriate way.

- Your mark will be based on how appropriate the analysis you propose is, and the range of skill and understanding you demonstrate in answering the questions.

Submitting your answers

You should submit exactly 5 files:

1. An Rmd file and
2. datafile which when ‘knitted’ produce
3. An HTML or PDF document.
4. As a separate document, upload a copy of the standard CW coversheet (and complete the feedback section).
5. Also include a pdf copy of the study you chose to replicate.

Within the Rmd file you should:

- Label each question clearly
- Where necessary, include comments explaining what specific lines of code do.
- Intersperse explanatory text with code chunks (don’t put everything in the same code chunk).

Your Rmd file should “Just Work” when the marker opens and runs it, and produce the same output as the knitted HTML or PDF file you submitted (i.e. there won’t be any errors or missing datafiles).

If you work on your own computer at home, you should check your Rmd file ‘knits’ correctly on the online Rstudio server.

See common problems when trying to knit Rmd files