Final Assignment – 51735 Statistics for Graduate Students

Submission by: 28/02/2023 [*Part 0 submission: by 04/01/2023, see below; in-class brief presentations on two additional dates – details to appear separately*]

Please submit *in pairs* via Moodle.

General explanatory note: The goal of this assignment is to guide you through applying what you've learned in class *on your design and in the context of your research question*. We hope that the assignment will be informative as you plan your design, prepare for (or progress with) data collection, and think about the models and methods you plan to apply on your data. However, to make sure you are working on data that fits the requirement for this assignment, you will first need to send it to us for approval– **see details under Part 0**, below. Once approved, you can start working on Parts 1-3 and on your presentation. We are available if you have any questions or need help identifying a dataset.

Technical notes:

(1) *Word limits serve as general guidelines*. We will not deduct scores for being *slightly* over the limit (rest assured, we will not run word counts!). But, excessively and needlessly long responses will be penalized. Try to be succinct and avoid including unnecessary information.

(2) While you can contact us via email, as always, we encourage you to use the *Q&A forum on Moodle*. However, we will not be able to answer questions outside the limits of clarifications (e.g., asking which test is better for your data).

(3) *Separate instructions will be sent regarding the in-class presentation*, which will provide a floor to discuss progress and questions.

Part 0: Choosing a design and dataset to work with.

NOTE: This part must be completed before you start working on Parts 1 onwards.

Identify **a research question, a task/design, and a dataset** that is similar to or relevant for your research. It can be from your own study (even a pilot), from that of a lab member, or from a publicly available dataset from a paper in your research domain[1]. The only requirements are that:

(1) The data was previously analyzed using a "classic" NHST test (i.e., t-tests, ANOVAs of any kind, contrasts, correlations, simple regression, or multiple [not mixed-effect] regression [linear or logistic]).
(2) Raw data at the smallest observation unit is available for you to work with. For example, in the case of an in-lab task, you need to have access to the trial-level data (e.g., in a RT-based task, you need to have access to the RT in each trial, not just means for each condition/participant; in the case of clinical longitudinal data, you should have access to

[1] There are multiple ways to find published papers with publicly available data. E.g., you can search for papers in relevant journals that include an "Open Data Badge". See here.

the data at each time-point, etc.). If data is only available at an aggregated form [e.g., by-condition means], you will not be able to work with it in Part 3, below.

(3) In addition to a dependent variable(s), there are at least two potential predictors in the dataset (either independent variables or controls), even if they were not previously analyzed by the original authors/team. The dataset should also include unique identifiers for participants.

0.0 Before you start working on Part 1 and beyond, **please share with us your dataset for approval and complete a short survey**, via the relevant assignment on Moodle. **This part should be completed by 04/01/2023**, so we can inspect your dataset and leave you enough time to work on your in-class presentation and Parts 1-3. Earlier submissions will receive earlier responses.

Part 1: Describing the design and dataset [7 points overall]

1.1 Describe the research question, design, and dataset, following the sub-questions below. You can also include a figure to describe the design/task, if helpful. [300 words overall; 7 points]

1.1.1 What was the study's main research question?

1.1.2 What was the study type (i.e., correlational, or experimental[2])? Was it a between-subject or within-subject design?

1.1.3 What were the task(s) used, and the experimental conditions in the task (if any)? How many stimuli and trials were included? Was the data longitudinal (i.e., collected at multiple time-points)?

1.1.4 What were the study's variables? Refer to independent variable(s), control variable(s), and dependent variable(s).

1.1.5 Who were the participants in the sample and what was the sample size? How was sample size originally determined?

1.1.6 Are there any other methodological details you think we should know about for us to understand the rest of this assignment?

Part 2: What was done already? [45 points overall]

2.1 How was the data previously cleaned and pre-processed? Were any observations (at the participant- or trial-level) removed? Based on what criteria? Were any transformations applied to variables? Briefly describe the processes applied, pertaining to these questions and others you deem relevant [80 words; 4 points].

[2] Recall your basic research methods classes: An experimental design is one where there's a manipulation on the independent variable(s). A correlational design is one where there is no such manipulation.

2.2 Repeat the exact data cleaning procedure using your own code. Submit the code with detailed comments. [4 points]

2.3 Inspect and describe the dataset after the data cleaning and pre-processing procedure. Answer the following questions, and submit the lines of code that helped you answer them [80 words; 5 points]

> **2.3.1** How many participants are left in the sample overall?
>
> **2.3.2** *If trial-level task:* How many trials are included per participant (and per condition, if applicable) on average?
>
> **2.3.3** *If longitudinal data:* How many time-points are included per participant (and per condition, if applicable) on average?
>
> **2.3.4** What is the distribution of number of trials/time-points included by participant? Draw a histogram.

2.4 What was the statistical procedure previously employed on the dataset to answer the main theoretical question? Briefly describe the statistical test(s) in the context of the study. [100 words; 4 points].

2.5 Given the sample size originally used, and the statistical procedure utilized, what was the study's a priori statistical power[3]? Compute power for different population effect sizes that you deem as "small", "medium", and "large" in the context of your research domain. Given the power levels computed and the alpha used, and given what you see as a reasonable value for p(H1)/p(H0) odds (i.e., R), what is your estimate of the study's PPV? For this section, you can assume that all test's assumptions were met. You can use an off-the-shelf library for power calculations (e.g., library(pwr)) or run a simulation. Submit the code and briefly describe your findings and their implications [150 words; 10 points].

2.6 Repeat the NHST statistical procedure originally conducted. Submit the code used. Summarize your findings (in your own words). You can (and are encouraged) to use a Table and a Figure (max 1 of each) to present the findings. Were you able to replicate the original findings? If not, discuss what might be the source of the differences. [250 words; 10 points].

2.7 Were effect size estimates and/or confidence intervals originally computed/reported? **If yes**, re-estimate them using your own code. Plus, discuss whether additional estimates are required in your opinion, and if so, compute them. **If not**, choose and justify what estimates to include, and write a script that computes them. **In both cases:** Submit the code used, discuss the computed estimates, the justification behind their use, and discuss the implications of the estimates. [250 words; 8 points].

---

[3] To clarify – since the question here is about *a priori* statistical power, it wouldn't be impacted by the actual data observed (only the sample size and the procedure used).

Part 3: What could be done differently? [48 points overall]

3.1 Discuss – is the procedure previously utilized proper for the dataset at hand? Were any assumptions potentially violated by using this procedure? Are there any additional sources of variance not being accounted for by the method used (e.g., ignored random effects or control variables?)? Any other methodological issues related to the employed test? Critically discuss the identified issues. In parallel, inspect the data and try to demonstrate evidence for the potential issue(s) you've identified. Submit this code. If relevant, accompany your description with a figure. [200 words; 10 points]

3.2 Identify (some type of) a mixed-effect model analysis that can be used to analyze the data and address at least one of the issues discussed in **3.1**. Fully describe the planned analysis and all pre-processing steps it requires (e.g., transformations, codings, random effect determination method, etc.). Write a brief 'analysis' section (as if you are reporting your analysis plans in a paper/pre-registration), containing all these details. Also submit the code that instantiates these preparatory steps on your data. Your mixed-effect model should contain at least two predictors, and in any case should address the work's theoretical question and the issues you've identified in **3.1** above. [200 words; 8 points]

3.3 Run the mixed-effect model. Submit the code used to run it. Describe your findings, as if you are reporting it in a paper. You can add a Table and/or a Figure (max. 1 of each). [200 words; 10 points]

3.4 Discuss how the results of the mixed-effect analysis compare to those originally observed/reported (using the "classic" tests). Do these differences (or lack thereof) fit your expectations? Are they in line with general differences between the two analytical approaches? Discuss what may have contributed to differences/similarities you observe. [150 words; 10 points].

3.5 Assuming that the population effect size is the one observed in the mixed-effect analysis of the current dataset, conduct a prospective power analysis for a replication study. We encourage you to use the *simR* package to circumvent needing complex simulations for this section[4]. What is the power of a replication with an identical sample size (and all other design properties)? If you think this power is too low or too high: What is the sample size needed to result in what you see as reasonable power? Submit the code used to run the power analysis, and describe what you found. You can accompany your description with a Figure. Do these results change your data collection plans as you prepare for your next study? [150 words; 10 points].

[4] Check out the original paper describing the package for helpful examples and explanations: https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.12504