

## STAT2003 Analytics for Experimental and Simulated Data

### Assessment Task 1: Case study (25%)

2% informal class presentation for feedback before submission + 23% poster

**Aim:**

The purpose of this task is to allow you explore applied experimental design, analysis and reporting.

*A published scientific paper will be different in style to a purely statistical/data-analysis report. A goal of this task is to give you some experience with reading and writing in a scientific format for communication to a wider audience.*

**Learning Outcomes:**

- 1 Explain and recognise the difference between a controlled experiment and an observational study
- 2 Apply the concepts of sources of variability and ANOVA
- 3 Analyse data from controlled experiments
- 4 Identify potential weaknesses in the design of an experiment or the conduct of data analysis

**Task:**

Your task is to present a case study for an experimental design based on an existing study & dataset.

- Your goal is to find, understand and interpret an original study.
- You will attempt to replicate the analysis (or a simplified analysis).
- You will present your understanding and analysis as a 'case study' for other students.

**Mode:**

Individual, however, as each person's assignment will be different you can discuss and help each other (collaborate) with the process.

**Format:**

Powerpoint poster

- Some default poster designs are provided on BB.

Supporting files:

Appendix: can be in smaller font as additional page on the poster; or you can submit a separate word/html file (if you create an r script or rmd file make sure you include a compiled version in word or html).

Datafile (.csv or .xlsx)

**Length:**

Negotiable: I suggest 750-1500 words (plus tables/graphs/diagrams) with an appendix of output or code.

**Submission:**

Online submission on Blackboard, and a brief (5 min), informal, pre-submission presentation to the rest of the class of what you found in week 5 for feedback *before* you submit. (This may be extended to week 6 if needed).

**Structure:**

Title: Include the design, general topic area, and if it's a 'real' or 'simulated' experiment. You can use the original title or reword it for a general audience.

Citation: Include the original citation with links to the paper and data.

Introduction

- Context (give a enough background information that we can understand the application).
- Experimental Design (brief explanation of the design & its key features; for example, does it include randomisation, replication, blocking? Is it a 'screening' experiment or a 'full' design intended to describe effects in detail?)
- Conceptual diagram (see chapter 1 of the text for examples).
- Make a clear statement of the research question/goal/aim of the original study. You can quote this from the original study if it was clearly written or paraphrase it for non-experts. Please use quote marks or italics if quoting the original study for clarity.

Methods

- Describe the design and analysis plan of the original study, make the explanation suitable for other undergraduate students (it can include an experimental diagram of the setup either from the original study or your interpretation; this would represent the actual experiment rather than a conceptual diagram of the whole context).
- Describe the data which is provided for the study (eg is it a full raw data set, has it been reduced or redacted) and make a table of variables.
  - Describe if you had to reformat or handle the data in some way to analyse it.
  - Describe any challenges(!) For example some papers say that they 'blocked' over a variable but then this is not included in the data or analysis.

Results

- Present the results of *your* analysis.
  - Your analysis can be simplified if the original one was very complex.

- Summarise your findings, and if they match the original findings.
  - You can present your findings and the original findings side by side if that is of interest.

### Discussion

- Reflect on the success/appropriateness of the original study.
  - You can include strengths and limitations of their reporting – was it clear what they did? Could you reproduce the analysis and findings?
- Discuss how accessible or inaccessible you found the original paper and dataset.
- If you were going to design a follow-up study would you make any recommendations?

### **HOW TO FIND A CASE STUDY:**

#### Process:

- Go to Dryad data repository: <https://datadryad.org/stash>
  - *There are other data repositories, we are looking specifically for data which will have a scientific publication attached which most of the Dryad ones do. Another source is <https://data.mendeley.com/>, this is a much large repository so you will need to have more key words to narrow down your search.*
- Use the search function to search for ‘experiment’, it can be any topic (humans, animals, agriculture, ecology etc). The only requirement is that it must be an experiment, not an observational study.
  - You can narrow this down by using a topic area, or journal, or adding other key words like ‘factorial experiment’.
- Find a title that sounds interesting to you, and, check the abstract & data to see if its suitable.
  - You may want to add in other search terms to narrow down a field of interest.
  - Find the original publication if there is one.
    - Some datasets have not been published as full journal articles, this might be ok if they give a structured report on the repository.
  - Sometimes the data files do not actually contain raw data (which is frustrating!) before you spend too much time on a possible topic check what is contained files.
- I suggest you find 2 or 3 possible studies to choose from rather than stopping at the first one you look at. And you can ask for help/feedback.
  - We will begin this process in a workshop as it can be a little intimidating when there is a lot of specialist terminology to understand.

#### Requirements:

1. It must be an experiment.
2. One of the outcome variables should be continuous (these is so you can use ANOVA rather than, say, logistic regression which we aren’t covering in this subject).

Recommendation:

It's more important that you find a paper & dataset that you can understand then finding something on a very specific topic. For example, bumble bee dancing behaviour is very interesting, but if the experimental design & data is very complicated then you are potentially making your task more difficult than it needs to be.

Having a lot of variables is not necessarily a problem, often we can trim the dataset to the core components and simplify the analysis. (It is always interesting if you get approximately the same results with a simplified analysis as they get in the paper with a more sophisticated approach.)

Show me what you are doing in workshops, I am here to help.

Do not use either of these studies as we will look at them during the unit:

- Reinhardt, Jan D. (2021), *Telerehabilitation program for COVID-19 survivors (TERECO) - Randomized controlled trial*, Dryad, Dataset, <https://doi.org/10.5061/dryad.59zw3r27n>
- Baines, Celina B.; McCauley, Shannon J.; Rowe, Locke (2014), *Data from: The interactive effects of competition and predation risk on dispersal in an insect.*, Dryad, Dataset, <https://doi.org/10.5061/dryad.j8b02>