STA2005S – EXPERIMENTAL DESIGN PROJECT – 2025

The following project should be done in groups of up to a maximum of three students.

- Please hand in only one electronic report per group (pdf only) on Amathuba (Gradescope). Please name all group members.
- Also hand in an electronic version of your R code (either R code or a Quarto or RMarkdown file). This should be separate from your report. Again, only one file per group.

Handin Date: Monday 20th October 2025. The penalty for late handins is 10% per day. All students must sign the project and plagiarism form.

Satisfactory completion of the project is a DP requirement.

Some Guidelines

Design and Randomization	10
Analysis and presentation of results	15
Conclusions and Recommendations	5
Overall impression and clarity	5
Total	35

- Limit tables, computer output and graphs to what is informative.
- The report must be typeset.
- If you want to attach an appendix, limit this to what is needed to understand your report.
- Give the original data somewhere, either in the report or in the appendix.
- Also give the original randomisation used.
- No raw computer output in report: edit all tables, leaving in only the information necessary. Limit decimal places to three or less.

- Graphs with poor resolution make a bad impression. Find a format to save R graphics that don't look terrible in MSWord (if you are using Word).
- Every graph should have a short caption which explains just enough to understand it without having to read in the main text. So should tables.
- As a guideline, I would estimate that 10 pages should be (more than) enough space for the report.
- Use a spell check!
- Cite software used (you should cite at least R, and state the version used). See citations() in R for how to cite R. Also cite R packages used.
- Cite any literature or websites from which you have obtained information, including notes, code from tutorials, ChatGPT, etc. It is okay to use AI if you ask it to help you, but not if you ask it to do the work for you.
- Your report should have three sections, corresponding to the above rubric. Mark these clearly in your report. The design section should define the hypotheses you want to test. Remember that it is important that you think about these hypotheses at the design stage and design your experiment so that you will be able to answer these questions. Should other hypotheses be suggested by your data, distinguish between a-priori hypotheses and post-hoc tests.
- In your methods section describe the statistical methods you have used. You should give enough detail here so that anybody wanting to replicate your results will have enough information to know exactly what you did. In the methods section, define the model that you will fit, mention how you will treat multiple testing, missing values, etc. (any special methods that are required).
- Your results section should show everything that is required to understand what the data from the experiment can tell you. Plots are useful. Estimates without confidence intervals are not so useful. Anything you talk about in your conclusions needs to be backed up by what you show in your results.
- It is important to put your conclusions into context. What does everything mean? From the perspective of the farmer, what can they learn from this experiment, what should they be doing, what are the uncertainties, limitations of your study.

Two farmers, Ms Hopeful and Mr Growing, grow chillies in Ceres for the South African market. After several years, Ms Hopeful clearly gets higher yields than Mr Growing. Unfortunately, Ms Hopeful and Mr Growing have difficulty determining what causes this discrepancy because several factors (heat, light, and variety) differ in their methods of growing chillies. Mr Growing is conservative. He uses standard heating and lighting and a variety of chilli called "Redhot" while Ms Hopeful uses supplementary heating and lighting and a variety called "Furious."

Your assignment is to design and analyse an experiment to help the two chilli farmers find the best settings for growing chillies and to help to understand how the requirements for the two varieties may differ.

The response variable is a quality score, which combines taste, yield and look. These are lean times and research money is tight; therefore you only have two small greenhouses available, and will need to repeat the experiment over two growing seasons (years).

In each greenhouse, there are eight plots facing north and eight plots facing south. You can assume that each of the 16 plots inside a greenhouse are large enough for several chilli plants and that light, heat and variety can be separately controlled per plot without affecting neighbouring plots.

The farmers are interested in four light levels (1, 2, 3, 4), and four heat / temperature levels (1, 2, 3, 4). There are two varieties (Redhot and Furious).

Two following questions are of interest:

- Which light and heat settings produce the highest quality chillies?
- Does the above depend on variety?
- Does quality depend on variety?
- How big are the differences?

The procedure is as follows:

- Design your experiment and put this design (i.e. your design matrix) into a .csv file called `design.csv`. See my example, called `design.csv`, (which is deliberately a very bad design!). Each row corresponds to the settings for one plot.

Be careful to use the correct code for the factors, and the correct names for the variables (the easiest is to just change the levels from the above example).

season	1, 2
greenhouse	A, B
light	1, 2, 3, 4
heat	1, 2, 3, 4
variety	R, F
side	n, s

The `plot` column identifies the experimental unit.

- Obtain observations / scores for your experiment. These are simulated, but treat them as if they are observations from your experiment. See the file `ProjectED.R`. Read your design file `design.csv` into R, use the function `get.observations` to get simulated observations. Save your design and observations. To obtain the same observations in another simulation, you will need to set the seed in R (`set.seed(...)`).

- You might have some missing observations due to unforeseen circumstances (the heat setting malfunctioned, chillies in a plot caught a disease, an animal dug out the plants, ...). These observations were deemed as not representative and therefore removed. Adjust your analysis accordingly.
- Analyse and interpret your results.
- Make (justified) recommendations as to which combination gives the best yield. Extra heat and light are expensive, consider this when making your recommendations.