Case Study 3

Effect of sewage disposal in river biodiversity

The Experiment[1]

A field ecologist is interested in examining the effect of sewage disposal in the diversity of invertebrates normally found in rivers. She dispatches a team of graduate students to count how many different species can be found in 100-Liter samples of water from seven rivers that have untreated sewage disposal at some point in their course, from both 100m before and after the point where the sewage is disposed. For each river, 10 samples are collected (one each month, from february to november). The simulated data is available in the file *riverbio.csv*, available from GitHub.

Activities

Based on the available sample, your task is to answer the following question:

Does the disposal of raw sewage in rivers affect the mean number of species?

Your analysis should follow a simple procedure:

1. Describe the experimental design required to answer the technical question of interest. Detail the hypotheses being tested and the relevant design for testing those hypotheses.
2. Perform the statistical analysis using the observations contained in the data file provided. This includes:
    a. Perform the actual test of statistical significance;
    b. Estimate the effect size (including the confidence interval);
    c. Check the assumptions of your test;
    d. Discuss the power of your test, **if relevant**
    e. Describe your conclusions and recommendations.

Remember that your conclusions should always be placed in the context of the original technical/scientific question.

Report

You must deliver a short report detailing your analysis and the results obtained. Instructions for writing your report depend on the course you're taking, and are given below:

1) For graduate (PPGEE) Students

Your report will be evaluated according to:

- Compliance with the required format (see below);
- Reproducibility of results;
- Technical correctness;
- Structure of argumentation;

[1] Based on M.J. Crawley (2007), "The R Book", Ch. 8.

- Correct use of language (grammar, orthography, etc.);

The report **must** be produced using R Markdown, and should contain the reproducible analysis code embedded as code blocks within the document. Please send me both the **.Rmd** file and the **.csv** data file generated by the simulation app. The analysis should assume that the data file is in the same directory as the report file.

A template for the case study reports is available in our GitHub repo.

Reports written in either Portuguese, English, or Spanish will be accepted.

For undergrad (Systems Engineering) Students

Your report will be evaluated according to:

- Technical correctness;
- Correct use of language (grammar, orthography, etc.);

The report has no particular template. R Markdown is suggested, but not mandatory. Please send me the **.pdf** file (**NO** .doc, .odt, or other exotic formats please), the **.R** analysis file, and the **.csv** data file generated by the simulation app