

## ENV 710: Applied Statistics Group Project

The learning goal of the group project is to use your statistical knowledge and skills to analyze a “real” dataset. Students will work in small groups of 2-3 people assigned by the instructor. Group members should work together to define a research question and obtain a dataset that applies to the question. The group will determine the appropriate statistical analyses and conduct and interpret the analyses.

The subject of the group project can be any research question related to the environmental sciences. The project data can come from online databases, faculty- or student-collected data, or your own data collection. The **project must include multivariate analyses** (i.e., multiple independent variables) such as multiple regression, generalized linear models, or mixed (multilevel) models. The analysis and all figures and graphs must be implemented in R.

This is the culminating project of the class, and therefore, it is expected that the final paper will be well written, polished, and demonstrate your knowledge of statistics and R programming. All statistical results must be presented appropriately, and all figures and graphs should be attractive, labeled correctly and include descriptive figure legends. The paper must be written in R Markdown.

### Calendar of Assignments:

- Mar. 6, 2023 - Abstract & Dataset
- Mar. 31, 2023 - Group Paper Introduction & Outline
- Apr. 21, 2023 - Final Group Paper
- Apr. 21, 2023 - Project Reflection

**Abstract (40 points - graded as a lab):** Each group will compose and turn in an abstract of the group project. **The objective of the abstract is for you to work with your group to define a research question(s) and identify an appropriate dataset.** For the final project, you will use more advanced statistics and models to analyze the dataset that those that we have studied to date; **therefore, you are not expected to know which models to use to analyze the data at this point.**

Try to identify a dataset that includes a minimum of 80-100 observations (replicates or rows of data). The dataset should also include multiple independent variables (database columns) that potentially explain the dependent variable(s).

The abstract should be approximately 1-page in length and consist of the following:

1. State your research question(s) and provide 1-2 paragraphs of background information needed to understand the question or issue.
2. Explain why this question is relevant and interesting to environmental science, management or policy in 1 paragraph.
3. Describe your data, including the dependent variable(s) and independent variables in the dataset. What is your unit of analysis (e.g., cities, nations, corporations, individuals, plots of land, rivers, etc.)? What type of data are your dependent variables (categorical, nominal, continuous)? List the data source (website, faculty project, etc.).
4. **(Optional)** With the understanding that we haven't yet discussed the multivariate statistics and models that you will use to analyze your data, if you do have an idea of the statistical tools/tests you might use (e.g., ANOVA, multiple linear regression, and generalized linear models) briefly discuss them. If not, skip this part.
5. State any concerns you have about completing the project or specific information or skills you might need that are not listed on the syllabus.

The abstract should be clearly and concisely written and demonstrate that you have a research question and dataset. Please turn in a *single abstract for the entire group*, uploading a *Word document to Sakai that is entitled with the last names and first initial of each member of the group* (e.g., “SmithTJonesCWuZAbstract”).

**Group Project Introduction & Outline (40 points - graded as a lab):** Please turn in a well-written, final version of the Introduction of your group project. See the guidelines below for what should be included in the Introduction (it should essentially be written as the introduction to a scientific paper). This assignment should also include an outline of the rest of the paper and should also include references for papers cited in your Introduction. This assignment will be graded on writing, grammar, appropriate citation of references, logical flow of the introduction to the topic, and clearness of research questions.

**Final Paper (100 points):** The final group project paper is due the last day of graduate classes (see syllabus). Please turn in a *single paper for the entire group, uploading it to Sakai entitled with the last names of each member of the group and the word FinalPaper* (e.g., “SmithTJonesCWuZFinalPaper”). *In addition, upload a copy of your data and your complete (and annotated) R script, entitling the folder with the last names of each member of the group and the word Script* (e.g., “SmithTJonesCWuZScript”).

*The final paper can be submitted as a Word document, although a pdf report created in R Markdown is preferred.*

The expectations for the paper include:

- The paper *should not exceed 4000 words (or about 5 pages single-spaced)* in length, including the text, tables, figures and bibliography.
- Appendices with R code or additional figures can be included in addition to the 4000-word limit.
- The paper must be concise and well written. Papers that are poorly written, even if the statistics are done correctly, will be significantly marked down. Writing is an important skill to be able to communicate results to your audience. See “*Picky Notes on Scientific Writing*” for guidelines for writing.
- Papers should be written in *first person* and *active* verb tense.
- Statistical analyses must be pertinent, correctly done, and correctly reported.
- Tables and graphs must be attractive, accurate, and informative. Both should include descriptive captions that explain the information the figure/table is meant to convey.
- Papers should include *a minimum of three figures/graphs*, created in R.
- The graphs included in the paper should depict the effects of the interest covariates (independent variables). *Graphs from your exploratory analysis or model checking can be included in the Appendix*, but not in the main text.
- Guidelines for outlining the paper and the grading scheme are below.
- *The paper must include at least as many multivariate models and group members (i.e. if your group has three people, you must include at least three multivariate models). Each group member must conduct one of the models. All group members must participate in data manipulation and analysis, tasks should not be distributed among group members such that one member does all the analysis and other members do the writing, etc.*

**Project Reflection (20 points):** Each group member should independently submit an approximately one-page project reflection that explains what part of the project s/he

accomplished. For example, which model specifically did you conduct? Why did you choose that model? What sections of the paper did you write? What were your responsibilities?

Briefly describe the process of conducting the entire project. How did working in a group improve or detract from this project? What did you like or dislike about the process? What did you learn about yourself or your group members?

If you could do the project over, what would you do differently (if anything).

## Guidelines for the Final Paper

The paper should be written in the format of a peer-review scientific article. The subject of the paper should be the environmental question or problem that you are investigating, not the statistical analysis. You will likely include a more depth description of your analysis than in many scientific papers, but focus should still be the environmental question.

### 1. INTRODUCTION

- a. Clearly define your research question. Include a literature review of similar studies that describe why your study is useful, interesting and situates it in the current knowledge of the subject. (~ 2 paragraphs)
- b. Clearly articulate your research question.
- c. Articulate the hypotheses to be tested.

### 2. METHODOLOGY

- a. Describe how the data were collected (this will vary with the type of project).
- b. Describe your data. What are the dependent and independent variables? What are the sources of data? Do you have missing or truncated data and how are you dealing with them?
- c. Provide an overview of your data analysis. What type of model will you use? Do you need to transform your data? How will you check the assumptions of the model?

### 3. RESULTS

- a. Exploratory Data Analysis/Descriptive Statistics
  - i. Summarize your data/results (e.g., number observations, measures of central tendency and variation).
- b. Statistical Analysis
  - i. Discuss the findings of your statistics and models. Make sure to interpret the strength of the effects of your variables, not just whether they were significantly important or not.
  - ii. Discuss the assumptions of your underlying the statistical tool and whether the assumptions were met?

### 4. DISCUSSION/CONCLUSION

- a. Discuss overall findings, statistical vs. practical significance.
- b. What is the scope of inference? To what population can you infer?
- c. How do your results compare to previous studies, your expectations or hypotheses?
- d. What were the limitations of your study? What would you do to improve the study?

### 5. BIBLIOGRAPHY

- a. Use a consistent, recognized format for putting together the bibliography, such as ESA (Ecological Society of America).

## **Grading Scheme for Final Group Projects**

| <b>Writing [25 points]</b>  | <b>Points</b> |
|---|---------------|
| 1. Easy to read, grammatically correct, concise, appropriately formal, references appropriately cited | 10            |
| 2. Use of 1 <sup>st</sup> person, particularly in Methods, Results                                    | 5             |
| 3. Introduction provided adequate background information  | 5             |
| 4. Study questions and hypotheses were clearly stated in the Introduction                             | 5             |
| <b>Analysis [25 points]</b>   |               |
| 1. Analysis was appropriately done and thorough given the dataset                                     | 5             |
| 2. Complexity of statistical models   | 5             |
| 3. Statistics were correctly cited in the text  | 5             |
| 4. Statistics were correctly explained in the text  | 5             |
| 5. Paper included interpretation of effect sizes, not just based on p-values                          | 5             |
| <b>Inference [25 points]</b>  |               |
| 1. The stated results logically followed the results of the statistical tests                         | 10            |
| 2. The conclusions logically followed the results   | 10            |
| 3. The limitations of the study were clearly articulated  | 5             |
| <b>Graphs/Tables/Presentation [25 points]</b>   |               |
| 1. The graphs were attractive, correctly labeled with axes, etc.                                      | 5             |
| 2. The tables were attractive and included the appropriate information                                | 5             |
| 3. Figures and tables were adequately described by descriptive legends                                | 5             |
| 4. The graphs and tables sufficiently represented the study   | 10            |