Project 2 Info

Project Overview

The second course project will use a dataset of your choice to explore regression modeling. You are welcome to work individually or with one partner for this project. For the report, you should identify a research question of interest to explore with data visualization and regression modeling. I am providing a curated dataset that looks at the distance a butterfly travels before landing. I have filtered the dataset to contain a single measurement for each butterfly and removed some additional predictors. The curated version is available at https://raw.githubusercontent.com/STAT505/project2/master/butterfly.csv. Additional information about the study design and the data can be found from the references below.

- https://www.sciencedirect.com/science/article/pii/S2352340919309667
- https://data.mendeley.com/datasets/kpcgkfmpv8/1

If you choose an alternative dataset, make sure that the response variable is continuous and there are at least two other predictors related to the response variable in the dataset. If you need another location to find datasets, the data in brief journal is a good place to look: https://www.sciencedirect.com/journal/data-in-brief.

The report should contain sections such as introduction, data overview + data visualization, statistical procedures, and results and discussion. The following is a rubric that will be used to assess the report. All items are graded with the following scheme, where some have multipliers:

1. **No Credit**: Criterion was not addressed or was written in a way that was not understandable.
2. **Beginning**: Ideas are not clear and supporting ideas are not presented.
3. **Developing**: Ideas are identified but not well supported and developed or are minimally supported and developed.
4. **Advanced**: Ideas are clearly identified and are adequately supported and developed.

Report generalities	**Points**	**Multiplier**
Spelling, grammar, writing clarity	/12	3
Paragraphs, section labels	/4	1
Appendix with complete code	/4	1
Figures (reporting, labels, captions) included in text	/4	1
Citations for papers and packages used	/4	1
Acknowledgments for other resources	/4	1

Introduction	**Points**	**Multiplier**
Project motivation	/4	1
Sample size(s)	/4	1
Data source and study design	/4	1
Research question	/4	1

Data Overview + Data Visualization	**Points**	**Multiplier**
Variables with units and descriptive statistics	/4	1

Data Overview + Data Visualization	**Points**	**Multiplier**
Titles, Labels, and Captions	/4	1
Figure Clarity	/4	1
Figure Quality	/4	1

Statistical Procedures	**Points**	**Multiplier**
Define model to fit (with complete notation)	/8	2
Defense of model choice	/4	1
Discuss model fitting procedure	/4	1

Results	**Points**	**Multiplier**
State and assess model assumptions	/8	2
Concise evidence statement for estimated model	/4	1
Summarize estimates from final model (intervals)	/4	1

Discussion	**Points**	**Multiplier**
Discuss Results in the context of the research question	/4	1
Scope of Inference: how can the results be generalized?	/4	1