Project 2: Mixed Effects Regression Analysis

As with Project 01, you should do a "complete" analysis, and present your results in an IDMRAD paper. Your IDMRAD paper should have all the elements that were in your Projech 01 paper; review the guidelines and extra materials from week 02 as well as comments from me, the TA, and/or you peers on your Project 01 papers.

You should follow this general approach:

- First, write the technical appendix showing all of the work that you need to do (R code, output, graphs, tables, and comments explaining what you did and why) to answer the questions below. This should not contain false starts and side-tracks, but should contain all the work in R you need to justify your results/answers for the questions below.
  - Remember to organize the appendix into logical sections that you can refer the reader to, for detailed analyses supporting the data and results sections of the paper.
  - Remember to put some text before and after each chunk of R code and/or R output, helping the reader see why you are doing that analysis, and how the results relate to some part of the project.
- Then, write the Results section, listing each question, and the highlights from the technical appendix that are needed to answer the question.
  - Divide the Results section up into subsections that correspond to the research questions / analyses that you will list in the Introduction.
  - The Results section should be more than just text, for example it may contain displayed equations that describe models, and tables of fitted coefficients, SE's, pvalues, residual and random effect variances, etc.
  - However, it should focus on results, and not process. Graphs and tables (if any) should display models and results, not diagnostics. Process, diagnostic graphs, etc., can go in the appropriate section(s) of the appendix. Be sure to refer to specific pages or sections in the appendix at each place in the Results section where the reader may want more detail.
- Then write the rest of the IDMRAD paper (**Title, Author/Email, Abstract, Introduction, Data, Methods,** Results, **Discussion, References,** Technical Appendix). Remember that the Introduction should list all of the research questions for the paper. Write the Title and Abstract *last*.

The rubrics for grading Project 02 will be the same as for Project 01.

The Data

Dietrich College at Carnegie Mellon University is in the process of implementing a new "General Education" program for undergraduates. This program specifies a set of courses and experiences that all undergraduates must take, and in order to determine whether the new program is successful, the college hopes to rate student work performed in each of the "Gen Ed" courses each year. Recently the college has been experimenting with rating work in Freshman Statistics, using raters from across the college. In a recent experiment, 91 project papers—referred to as "artifacts"—were randomly sampled from a Fall and Spring section of Freshman Statistics. Three raters from three different departments were asked to rate these artifacts on seven rubrics, as shown in Table1. The rating scale for all rubrics is shown in Table 2. The raters did not know

Short Name	Full Name	Description
RsrchQ	Research Question	Given a scenario, the student generates, critiques or evaluates a relevant empirical research question.
CritDes	Critique Design	Given an empirical research question, the student critiques or evaluates to what extent a study design convincingly answer that question.
InitEDA	Initial EDA	Given a data set, the student appropriately describes the data and provides initial Exploratory Data Analysis.
SelMeth	Select Method(s)	Given a data set and a research question, the student selects appropriate method(s) to analyze the data.
InterpRes	Interpret Results	The student appropriately interprets the results of the selected method(s).
VisOrg	Visual Organization	The student communicates in an organized, coherent and effective fashion with visual elements (charts, graphs, tables, etc.).
TxtOrg	Text Organization	The student communicates in an organized, coherent and effective fashion with text elements (words, sentences, paragraphs, section and subsection titles, etc.).

Table 1: Rubrics for rating Freshman Statistics projects. *NOTE: These are* **not** *the rubrics used by instructors or TA's in Freshman Statistics. They are* **only** *approved to be used in this experiment.*

Rating	Meaning
1	Student does not generate any relevant evidence.
2	Student generates evidence with significant flaws.
3	Student generates competent evidence; no flaws, or only minor ones.
4	Student generates outstanding evidence; comprehensive and sophisticated.

Table 2: Rating scale used for all rubrics. *NOTE: This is* **not** *the rating scale used by instructors or TA's in Freshman Statistics. It is* **only** *approved to be used in this experiment.*

which class or which students produced the artifacts that they rated. Thirteen of the 91 artifacts were rated by all three raters; each of the remaining 78 artifacts were rated by only rater. The variables available for analysis are defined in Table 3. The file `ratings.csv` contains data organized exactly as in Table 3. The file `tall.csv` contains the same data, but organized so that each row contains just one rating, in the column labelled Rating, and the rubric for that rating is listed in the column labelled Rubric.

The Research Questions

Your IDMRAD paper will be read by the associate dean in charge of this experiment. The four key research questions are:

1. Is the distribution of ratings for each rubrics pretty much indistinguishable from the other rubrics, or are there rubrics that tend to get especially high or low ratings? Is the distribution of ratings given

Variable Name	Values	Description
(X)	1, 2, 3, ...	Row number in the data set
Rater	1, 2 or 3	Which of the three raters gave a rating
(Sample)	1, 2, 3, ...	Sample number
(Overlap)	1, 2, ..., 13	Unique identifier for artifact seen by all 3 raters
Semester	Fall or Spring	Which semester the artifact came from
Sex	M or F	Sex or gender of student who created the artifact
RsrchQ	1, 2, 3 or 4	Rating on Research Question
CritDes	1, 2, 3 or 4	Rating on Critique Design
InitEDA	1, 2, 3 or 4	Rating on Initial EDA
SelMeth	1, 2, 3 or 4	Rating on Select Method(s)
InterpRes	1, 2, 3 or 4	Rating on Interpret Results
VisOrg	1, 2, 3 or 4	Rating on Visual Organization
TxtOrg	1, 2, 3 or 4	Rating on Text Organization
Artifact	(text labels)	Unique identifier for each artifact
Repeated	0 or 1	1 = this is one of the 13 artifacts seen by all 3 raters

Table 3: Variables in the file `ratings.csv`. Variables that are **not** expected to be useful for analysis are shown in parentheses.

by each rater pretty much indistinguishable from the other raters, or are there raters that tend to give especially high or low ratings?

2. For each rubric, do the raters generally agree on their scores? If not, is there one rater who disagrees with the others? Or do they all disagree?

3. More generally, how are the various factors in this experiement (Rater, Semester, Sex, Repeated, Rubric) related to the ratings? Do the factors interact in any interesting ways?

4. Is there anything else interesting to say about this data?

Suggestions for Answering the Research Questions

Here is a (possibly incomplete) list of suggestions for exploring and analyzing the data.

1. To explore Research Question #1:
   - You can use either `ratings.csv` or `tall.csv`, but you may find `ratings.csv` easier to work with.
   - Numerical summaries (perhaps counts, percents, means, SD's, etc.) and graphs (e.g. histograms, bar plots, etc.) are a good place to start.
   - It will be useful to make a subset of the data for just the 13 artifacts seen by all three raters, and reproduce whatever you did for the full data set, just for this subset. That way you can compare and determine whether these thirteen artifacts are representative of the whole set of 91 artifacts.
2. To explore Research Question #2:

- For this question, focus on the subset of the data for just the 13 artifacts seen by all three raters.
- One measure of agreement among the raters is the intraclass correlation (ICC). You derived a formula for it in HW10, problem 2(b); it is the common correlation among the raters' ratings for each artifact. To calculate it here, we can treat each artifact as a cluster of three ratings, and fit the random-intercept model[1]. Fit seven random-intercept models, one for each rubric, and calculate the seven ICC's.
- The ICC's can help us determine whether the raters are generally in agreement (high ICC = high correlation among the raters) or not (low ICC = low correlation among the raters) on each rubric, but they cannot tell us which raters might be contributing to disagreement. One way to do this is to make a 2-way table of counts for the ratings of each pair of raters, on each rubric (since there are three pairs of raters, each rubric will get three tables). For each table, the percentage of observations on the main diagonal is the *percent exact agreement* between the two raters. The tables, and the percent exact agreement from each table, can help to determine who is agreeing with whom on each rubric.
- You can re-do the ICC calculations on the full data set (but not the percent exact agreement calculations—why not?). Do the seven ICC's for the full data set agree with the seven ICC's for the subset corresponding to the 13 artifacts that all three raters saw?

3. To explore Research Question #3:

   - One way to do this is to add fixed effects for Rater, Semester, Sex and/or Repeated to the random intercept models for the full data set, perhaps look at interactions, and perhaps do variable selection. Do the ICC's from these models agree with your earlier ICC's? Do you find that any of these fixed effects have a significant effect in predicting ratings? Are there any other random effects that you can justify adding to these models?
   - This approach doesn't let you directly examine interactions with Rubric, since each model considers only one Rubric at a time (though you may find differences between the models, or in variable selection, that do suggest interactions with Rubric). One way to explore interactions with Rubric directly would be to switch to `tall.csv`: you might begin with the model `Rating` $\sim$ `(0 + Rubric | Artifact)`, and then add fixed effects (and possibly interactions) for all of the variables Rater, Semester, Sex, Repeated and/or Rubric, and try to answer the same kinds of questions as in the previous bullet.
   - As you explore these models, `lmer()` won't always behave. Sometimes it may warn you that the fitting algorithm has failed to converge, or that the variance-covariance matrix for the random effects is singular, or something else. If you run into problems like this, please let me know and I will either suggest a fix just for your particular analysis, or for the whole class to consider.

4. To explore Research Question #4:

   - Now that you're pretty familiar with the data, what other things can you say about it, that will be of interest to the associate dean?

[1] Since the ratings can only be 1, 2, 3 or 4, the data is not really normally distributed, and we might consider fitting a multilevel multinomial logit model. However, the results for rating data are usually similar to what we would get from plain old `lmer()`, and I just expect you to use `lmer()` here. (Feel free to look at residual plots, etc., to see how well or poorly the model assumptions are satisfied, if you wish.)

You are to do this project on your own, without collaborators (except for peer reviews & office hours). If you are unsure of what something means, feel free to look it up on the web or elsewhere, but you may not post questions on discussion websites or blogs like stackexchange, etc. Questions on Piazza should be **private** *to the instructors.* ***You are welcome to discuss this project with me or the TA (office hours are also fine), but no one else.*** *Please remember to cite all the sources that you used, including webpages, in the reference list at the end of your report.*

Due Dates

Wed Nov 17 First draft of Technical Appendix work will be due with HW 10.

Wed Nov 24 (or earlier) First draft of complete IDMRAD paper due.

Wed Dec 1 Peer reviews due.

Fri Dec 10 (or earlier) Final draft of complete IDMRAD paper due.

Note: Depending on due dates for other MSP classes, we may have to adjust the due dates above a bit.

Grading

On the next page is a summary of what I will be looking for. See the materials in the week02 folder for more detail.

The percentages in the table on the next page assume that all parts of the paper are there. If one or more parts is missing, it may result in a much lower grade than the percentages suggest.

Part	Looking For...	Percent
Title	*Clear, interesting, focused.*	5%
Author/Contact Info	*Your name & email addr!*	∞!!
Abstract	*Summarize I, D, M R and D sections of the paper* (typically one sentence each).	5%
Introduction	Brief, clear, to the point; context for the problem; What is the problem/aim of the study? *Why would anyone want to read this paper? What questions will be addressed?*	10%
Data	*What data set was used in this study?* Typically, include variable definitions, sample size, quick numerical summaries of the variables and initial EDA, but *no model fitting or analysis*.	5%
Methods	*What did you do, to address these questions?* List the methods and/or analyses that will be used to answer eacb question stated in the **Introduction**. *No data analysis, graphing, model fitting, etc. appears here*; you just say what methods and analyses you will use with which variables, to answer each question.	5%
Results	*Statistical analysis & results* in order parallel to **Introduciton** and **Methods** sections. Here you *finally* get to show the data analyses (model fitting, graphics, etc.) that you did, and what the results were. Don't overload the reader: put the highlights here so the reader understands what you did and why, and refer the reader to specific pages or sections of the **Technical Appendix** for more details. It should be clear which data analyses and results go with which question from the **Introduction**. *Every analysis that is presented here should have been mentioned in the* **Methods** *section.*	10%
Discussion	*What does it all mean? Recap findings; address main problem/question; strengths & weaknesses; implications, unanswered questions, future research.* Typically you will say, for each question from the **Introduction**, how the analyses that you did the **Results** section answers that question. You might also mention EDA and so forth from the **Data** section if that makes clearer to the reader what answers you found for one (or more) of the questions. Then you will talk about the big picture, what future work or generalizations of your work might look like, and any limitations of your study. But *there should be no additional analyses or results in this section; just use the analyses you did for the* **Results** *section (and possibly the* **Data** *section).*	10%
Mechanics	Follows C-C-C[2]as much as possible (sentences, paragraphs & sections); *Grammatical; Complete sentences and paragraphs; Easy to follow.*	5%
Statistical Content	*Correctly and appropriately uses technical and non-technical material* we have learned in class. Easy to follow; Analyses makes sense/not crazy (roughly 10% per research question)	40%
References & Citations	*Follow ASA guide*[3]*, "The Reference List".* *Follow ASA guide, "Reference Citations".* Be sure to cite **all** sources!	5%
Technical Appendix	*Contains complete versions of the analyses listed in the* **Methods** *section and presented in the* **Results** *section*: R code, output, graphs, tables, and comments explaining what you did and why. There may be additional analyses here (e.g. to support the **Data** section of the paper, or to show why the methods and analyses that you chose for the paper were the right ones). Make it easy for me to follow.	0%[4]

[4]See Rule 3 in `10 rules for better organized papers.pdf`.

[4]See `ASA Style Guide.pdf`. See also the references at the end of the Sheather book for good examples of this style.

[4]You will get credit for this as part of a hw assignment instead.