

You will undertake a personal research paper/project as a major component of our class. The topic must be chosen in consultation with me (via a proposal) and should be focused on a relatively new statistical topic to you (perhaps you heard the name in a class, but didn't study the method, or didn't study the method in depth). This paper/project is an opportunity for you to demonstrate achieving some of the learning goals for the statistics major. As such, successful completion of the paper/project is required to pass the course (which, in turn, serves as the comprehensive evaluation for the statistics major). The final deliverable for the paper/project includes a report written in R Markdown. You may include other deliverables (Shiny apps, etc.) as appropriate.

### Assignment

In the learning goals for the major, you can see that we expect graduates of the program to be able to understand statistical techniques in primary literature and to be able to communicate their understanding about statistics, as well as have the ability to identify resources to expand and fill in gaps in their knowledge. In this paper/project, you will:

1. Identify a new statistical topic of interest to you.
2. Research the topic, identifying multiple sources to use to learn about it.
3. Prepare an expository review of the topic at a level that a classmate could understand. (Imagine you are charged with teaching them the topic.)
4. Demonstrate your understanding of the topic with application(s) to a data set and/or simulations.
5. Communicate your findings/results of the application(s)/simulations.
6. Collate the exposition and findings/results of the application(s)/simulations into a final report.

This paper/project is designed to challenge you. Learning about a simple extension of a technique and applying it to a data set that you downloaded in .csv format is not acceptable. You should seek to challenge yourself and demonstrate your statistical analysis skills. The project is deliberately open-ended to allow you to fully explore your creativity - you can have more deliverables than the final report - think about Shiny apps or example simulation code that could be shared with a wider audience. Consider using data sets that require considerable wrangling skills to let you tackle a problem of interest to you with a new technique. You can easily combine the exposition of the new technique with a comparison to another technique (learned in our class or another course).

Note that “toy” examples may be used to demonstrate techniques, but will not count as an “application”. Toy examples may be included (this is encouraged!) for the exposition but the idea for the application is to try out the technique on more complicated data.

There are five main rules that must be followed:

1. The paper/project is individual.
2. The statistical topic must be new to you. That means something you did not formally study in-depth in a class, although you can study variants of such methods, or propose going into detail about a method that you studied briefly. If you learned about a topic during an internship or other experience, please consult with me to see if it is appropriate. If you have already applied the technique to a data set previously, it is likely not appropriate. What is new for one student may not be new to another - you've taken different courses and had different experiences.
3. You must prepare a paper/report as part of your final submission.
4. The paper/report must include 1) an exposition on the new topic showing that you were able to research the topic appropriately and 2) either application(s) to a data set beyond a simple “toy” example, or simulations to demonstrate your understanding of the technique.
5. The entire report/analysis, etc. must be reproducible, in line with our learning goals for the statistics major.

## 1 Components and Organization

In order to help you tackle this project, there are a number of components / deliverables designed to be turned in along the way. These are described roughly below. Details will be provided with each particular assignment / component.

### 1.1 Github

In your personal repo for the course, you should create a “Project” folder (or “Report”, etc.) to keep all relevant project materials. Remember that Github has a MB upload limit. Don't try to upload anything over 50 MB (100 MB will fail). When you get to the data for your application (if going that route), consult with me for alternatives to an upload for a large data set. You should put this assignment file and all future work on the project in this folder. You will also be asked to upload some materials to Gradescope for review. A final copy of your paper will complete your portfolio in the repo. Remember that maintaining your repo and keeping things organized, as well as doing all portfolio assignments for the semester, is a separate part of your course grade (and should be easy to earn!).

### 1.2 Brainstorm Topics

Homework 6 is an idea brainstorm for the project. This is so I can help you isolate a topic that is appropriate for the project. Although labeled homework, the points earned here go to the project components. Basically, you'll be asked to brain dump potential topics you

are interested in for the project. This will allow me to help you narrow down ideas, propose others, and make sure you are heading into the proposal phase with reasonable ideas for the project (i.e. not something too involved, or not involved enough). See the homework assignment for details.

### **1.3 Proposal**

Based on the homework 6 feedback, you will select and refine a topic and prepare a proposal for your topic/report. A proposal template will be provided that you can fill in. Revisions may be needed for clarification, and will be handled individually.

The remaining components are designed to help you with researching the topic, writing the paper/report, and staying on track to complete the final product.

### **1.4 Annotated Bibliography**

You will need to assemble several sources to learn about your new topic. (The assignment sheet for this will have more details about how many and what type, etc.) You will then construct an annotated bibliography to help in construction of the exposition portion of the report. The bibliography is submitted as a component. This will help you learn about your technique from multiple sources and begin to organize your expository review of the topic.

### **1.5 Exposition**

The portion of the paper where you describe the new technique as though you were teaching it to a classmate (with whatever name you want to give it - Literature, Background, or Exposition) will be based on your annotated bibliography. You can add sources to what you already have in the annotated bibliography (especially if my feedback says you need to!). To be sure that you are on track with the exposition, a full draft of this portion of the paper is due before Thanksgiving break. This will also help with incorporating proper citation early in the writing process, as you'll have a bibliography that we can implement for this portion of the paper at a minimum.

### **1.6 Update I**

Also due before Thanksgiving break is an update on what work you have left to complete before the paper is finished. You will need to create an issue in your course repo with answers to the following:

1. What is left to be completed before the paper/project is finished?
2. What is your proposed timeline to complete those items?
3. Do you have any questions for me?

## 1.7 Update II

After we return from break, you'll want to focus on having a complete draft ready for our peer review / exchange the Friday evening of the final weekend of classes. An update is also due that week. You'll create another issue in your repo with answers to the following:

1. What is the single biggest unresolved issue you are having? Please describe it briefly, and what your plan is for resolving this issue.
2. Do you have any questions for me?

## 1.8 Complete draft for peer review and peer review form

A complete draft of your paper/report for peer review is due by midnight on Friday, December 2nd to your repos. As best I can, you will be placed in a group of 4 (or 3, as numbers work out) for peer review. You will only review ONE other submission (pairs, or a three-way exchange) but you will have access to all submissions in your group so you can see other submissions for ideas on how your classmates have structured things, etc.

I will organize submissions and groups on Saturday the 3rd via Google Drive. You will have a form (Google doc to fill in, not officially a "form") to complete for the submission you are reviewing. Completed forms should be in the Drive before class on Monday, December 5th. That class day will be devoted to sharing feedback with each other. This is an important contribution to helping your classmates with their reports. For example, if you don't understand something in their exposition or in their application section, you have an opportunity to help them improve their paper by telling them so, and offering suggestions for improvement!

## 1.9 Report

Your final report and any other deliverables will be due by 5 pm on the last day of classes, Friday, December 9th. All materials should be in your repo folder for the project. A final copy of the report should be uploaded to the corresponding Gradescope assignment and a copy of the pdf put into your Portfolio folder.

I expect to prepare a template for the report, etc. as the semester proceeds. However, you may each need to adapt it based on your topic, plans for application(s)/simulations, and deliverables.

## 1.10 Portfolio Reflection

The only assignment due for our class after the end of classes is a final portfolio reflection after the final paper has been added to your portfolio. The deadline for this is likely to be mid-finals week (TBD).

## 2 Assessment Criteria

Your project will be evaluated based on the following criteria:

- **Understanding of Topic:** Does the paper demonstrate strong understanding of the topic? Are there any issues with the statistical content in the development of the topic? Does the write-up show appropriate synthesis of material from multiple sources?
- **Communication to Audience:** Does the paper introduce the topic well to the audience? Does the writing help communicate the author's statistical knowledge to a peer? Are there appropriate definitions, background, development, and examples to communicate understanding of the topic?
- **Application(s) / Simulations - Statistical Content:** Are the application(s) / simulations performed appropriately from a statistical standpoint? Are there any issues with the analysis or simulation setup?
- **Application(s) / Simulations - Write-up:** How well does the write-up convey the findings/results of the application(s)/simulations? Are data sets and simulations described in enough depth for a peer to follow along? Is appropriate support (figures / output, etc.) provided? Are tables / graphics / figures crafted well (captions, look nice, etc.)? Is each table / graphic / figure explained and interpreted well enough that a peer can understand the data it is presenting, as well as its purpose/relevance to the point it's meant to illustrate?
- **Presentation / Organization:** Is the paper organized well? Are ideas presented in a logical order with a guiding voice / narrative thread that makes clear to the reader what each section is doing, the relationship between the sections, and the overall logic of the paper? Has the paper been proof-read and run through spell-check?
- **Code:** Is the code reproducible? Is all code readable and formatted well? Does all code (and associated comments) fit on the pages? All code should be visible, except for the initial setup chunk where packages are loaded.
- **Citation / Use of Sources:** Are sources incorporated appropriately into the text? Are attributions for direct quotes and figures included? Are sources listed appropriately in a bibliography at the end of the report?

Some students may be interested in submitting their final projects to USPROC, a competition for undergraduate statistics projects. More information can be found here: <https://www.causeweb.org/usproc/>. If you are interested, final projects for this class would qualify for the USRESP category of the competition. Projects may need some revision (for length) and code refinement (hide code) for submission to the competition.