Course Group Project Specifications

Neil J. Hatfield

Last Updated: February 06, 2024

The purpose of the Course Group Project is multifaceted. On the one hand, you get to put into practice (and show off) your understandings of the course material in an authentic way. That is to say, there is little room to hide what you do and do not understand. On the other, you will further practice your communication and teamwork skills that are vital to you getting a job and becoming a top-notch statistician, data scientist, or other wielder of statistical tools.

The course group project is fairly straightforward with two options:

- **Option 1**: Design and carryout a study (experiment, quasi-experiment, or observational) to collect data and then analyze that data to answer your research question.
- **Option 2**: Find and adopt a real data set that works for your research question (you must describe the way the data was collected and for what original purpose) and then analyze the data to answer your research question.

In either option, each group will submit a single report (including an Executive Summary) at the end of the semester.

Since this is a project for our ANOVA class, your central analysis **must** come from the ANOVA Toolkit. You may not propose a regression project. If you are proposing multiple analyses, the primary one should be ANOVA. Any secondary analyses are up to you.

Groups

With relatively few exceptions each person *must* complete this project in a group with at least one other person. Only in special circumstances (e.g., using this project as an Honor's project or part of the Master's thesis) will I consider allowing a person to work alone. If you feel that you have a special situation, you need to tell me as soon as possible.

While I will take input from each student, *I* will make the groups for the project.

Details

Let's take a look at the details a bit more closely. Short assignment names appear in parentheses.

1) Begin thinking about ideas and teams (GP #1)
2) Teams get formed
    A) Neil will form teams of roughly 2-4 people. The creation of teams is a joint enterprise between the class and Neil. Be sure to check out the important notice at the end of this document.
    B) At the end of the semester, each member of a team will be responsible for grading each team member as well as themselves in terms of group work and contributions to the project.
3) Develop a Statistical Research Question (GP #2)
    A) As a team, you will
        i) state what you're trying to learn,
        ii) come up with one or more SRQs
        iii) identify potential key design elements (e.g., measurement units, response, factor)

iv) explain how you are going to get your data
v) generate a Hasse diagram for your proposed studies

B) Your group will need to schedule a "pitch meeting" with Neil to go through your proposed study for feedback and approval.

4) Develop your Study/Experiment

A) As a team, use feedback to finish designing your study
B) Special Note–Involvement of Human and Animal Subjects
i) In the event you need to use either Humans or Animals subjects, you will need to have your study scrutinized carefully in coordination with Neil.
ii) If you need to go down this route, your timetable is now ASAP.
iii) Data may NOT be collected until after IRB approval is secured, if necessary.

5) Collect your Data

A) As a team, carry out study to collect your data.
B) If you are adopting data, you need to identify whether any Data Sharing Agreements need to be signed. If so, you need to work with Neil so that we can get you access to the data. DO NOT FILL OUT ANY DATA AGREEMENT ON YOUR OWN.
C) I have limited funds which may help to defray costs.

6) Register Your Study (GP #3)

A) Your group will practice some of the principles of Open Science by registering your study.

7) Analyze your Data

A) Clean your Data
B) Write your Data Narrative
C) Carry out the appropriate method(s) of analysis from the **ANOVA toolkit**. (Note: you can't use a regression model as the primary analysis for this project)

8) Write your report.

A) Structure (You're structure should look similar to this)
1) Executive Summary (2 page max)
2) Introduction
a) Literature Review & Background, as appropriate
b) Statement of Research Questions and Hypotheses
3) Methods
a) Describe the study
b) Describe the analytical/statistical methods you used
c) Describe the sample (i.e., Exploratory Data Analysis)
4) Results
a) Assumption Checking
b) Omnibus Results
c) Post Hoc Results (as applicable)
5) Discussion
a) Limitations
b) Future Work
6) References (include where others can get access to your data)
7) Author Contributions
8) Code Appendix

9) Submit your Project Report (GP #4).

10) Complete Peer Evaluations (GP #5)

Potential Topics

You are only bound by four things:

1) ethics,
2) what you can complete (i.e., collect and analyze) in the remaining time this semester,
3) the interests of your team, and
4) what you can afford (I can help out with some limited costs).

I will not approve any SRQ and subsequent study design that I believe to be ethically questionable. The same is true for using data obtained by someone else. I will do my best to help your team refine your SRQs into something that you can explore this semester and I will give you pointers on your study design. Try to draw from your experiences for inspiration.

Word about Kaggle

I must issue a few important cautions about trying to adopt data from Kaggle. The vast majority of data sets found on Kaggle are *not* designed for ANOVA. This means that students who are insistent on using data from Kaggle are either having to manufacture a way to get a Kaggle data to fit within the parameters of the project or they try to substitute a regression analysis. Neither option will result in a successful course project.

Second, I have noticed two disturbing trends the last several years with Kaggle data sets: more and more of the data are fake/synthetic (i.e., not real but potentially simulated from real) and many of the people posting the data are not the originators of the data. You need to make use of real data for this project. Simulated (fake/synthetic) data often is generated from a particular model–you finding that model is neither interesting nor the point of this project. People posting the data sets created by others often with little to no attribution and/or no clear permission is an ethically ambiguous space. How would you feel if someone posted your work product, under their name, without your knowledge?

Important Notice–Working in Groups

A potentially useful way to think about this group project is to think of it as a team project at a job. To this end, I will play several roles throughout the project:

- **Technical Supervisor:** I will act as your immediate supervisor on the job, periodically checking in with the team, asking questions, answering questions/offering advice, etc.
- **Human Resources Office:** I will also check in to see how well the team is working together. If there are interpersonal conflicts, I need to know so that I can help resolve them.
- **Course Instructor:** Ultimately, I will be assessing the final product your group produces.

Given these many hats, I end up in the position where I might need to "fire" and/or otherwise remove a member from a team. Thus, I reserve the right to remove individuals from any team for the course group project. Depending on when that removal happens, I might reassign them to another group or seek an alternative.

Such removals can happen *after* the submission of the final report. These cases are generally the result of a team member ghosting their team and not contributing to the report. In such cases, it is at my discretion about whether the removed student will have an opportunity to submit an alternative assignment.

Additional Project Ideas

Here are some ideas that have been done by students in the past:

Factors	Response
seat height, generator, tire pressure	bike course completion time/pulse rate
popcorn brand, batch size, popcorn to oil ratio	yield of popcorn
amount of yeast, amount of sugar, liquid type, rise temp, rise time	quality of bread
hours of illumination, water temp, specific gravity of water	growth rate of algae
blending speed, amount of water, water temp, soaking time	blending time for soy beans
width/height ratio of balsa wood, slant angle, dihedral angle, weight added, wood thickness	flight length for model airplane
type of drink, number of drinks, rate of drinking, hours after last meal	time to get ball through maze
stamp type, zip code, time of day when mailed	days required for delivery of letter
distance to target, type of gun, type of powder	number of shot penetrating 1ft diameter circle on target
amounts of cooking wine, oyster sauce, sesame oil	taste of stewed chicken
ambient temp, choke setting, number of charges	number of kicks to start motorcycle
amounts of flour, eggs, milk	taste of pancakes (consensus of housemates)
brand of tape deck, bass level, treble level, synthesizer	clearness and quality of sound
child's weight, spring tension, swing orientation	number of swings and duration of an infant swing
orientation of football, kick, steps taken prior, shoe type	distance football is kicked
amount of detergent, bleach, fabric softener	ability to remove oil and grape juice stains
weight of bowling ball, spin, bowling line	bowling pins knocked down
freq. of watering, use of plant food, temp of water	plant growth rate
temp of gas chromatograph column, tube type, voltage	size of unwanted droplet
concentration of lactose crystal, crystal size, rate of agitation	spreadability of caramel candy
proportional band, manual reset, regulator pressure	sensitivity of pneumatic valve control system
temp, nitrate concentration, amount of added preservatives	nitrate concentration in sewage
pH, dissolved oxygen content of water, temp	extent of iron corrosion
amperage, contact tube height, travel speed, edge preparation	quality of weld
brand, mess type, liquid amount, amount of towels used	quality of paper towel clean up
type of OS, amount of RAM available, language used, parallel or serial processing	Computation Speed
oven temp, baking time	Quality of cake
brand of pop/soda, type of candy	volume of displaced foam
species of Wisconsin Fast Plant, fertilizer type, light conditions	height of plant
Minecraft Character Handedness, Shooting Hand, Type of Bow	Shot power