

MA 404 ~ PROBABILITY AND STATISTICS

STATISTICAL DATA ANALYSIS PROJECT ~ 250 POINTS

The Process of a Statistical Data Analysis:

- Get the data set and “figure out” what you have (mostly this is play around with stuff and make sure you understand what the variables you are working with mean).
- Get a firm grasp on the client’s point of view. What will they be trying to accomplish with your results? What type of results are more useful?
- Look for variables that are more closely connected with the piece you are trying to predict/understand, either because they logically have a strong connection or because there is high correlation in the data.
- Figure out how different values in the variables impact the outcome (differences in averages/proportions between groups).
- Consider a predictive analysis. Can a reasonable regression model be built?
- Write a coherent, integrated, concise report for the decision makers. Use the following general structure when writing:
 - Restate the general problem(s), and then give a very brief overview of how you address the problem (this means tell them what to expect in the paper in a sentence or two - do NOT tell them all the things you did to get your results, they don’t care).
 - Give pertinent descriptives, starting with single variable and then moving into grouped stats. (Graphs may often be important here - see Resources for Writing/Communicating Data.) Be sure that your variables are clearly defined to the reader.
 - Use APA style to present both descriptive and inferential statistical results clearly and efficiently.
 - Use the descriptives to point out the “question” that the inferential stats will answer, and use the HT/CI to answer it. (It should be a smooth ride from sample data all the way through to conclusions and how to act going forward.)
 - Include smooth transitions, recaps, and conclusions as appropriate to make the document more understandable to your audience. Assume that they have zero insight into what the data says.
 - Proof-read/revise (multiple times). You should proof-read and revise **the entire report** at least twice for every submission.

Presentation and style matter, so consider formatting (for clear communication and ease of reading), proof-read and revise (at least 2-3 revisions for each draft), avoid egocentric presentation (the report is not for you, you already know what it says), and say only exactly what needs to be said, nothing extra, but nothing less.

At no point should there be a play-by-play approach telling what you did. This is not a lab report; it is summary of key findings for decision makers to use to improve the quality of their decisions. Avoid first person (it just reads awkwardly; using phrases like “we find that ...” is generally okay) and use complete sentences/paragraphs with introductions, conclusions, and transitions (do not use an outline format, although judicious use of section headers is acceptable, as long as intros/conclusions/transitions are used).

- Review these instructions regularly so you maintain a big picture view of what you are trying to accomplish and don’t get lost in the weeds of the nitty-gritty.

Goals of this assignment:

1. to help you apply statistical analysis and understand what all the theory is designed to do (nothing builds experience and confidence like actually doing an analysis project from start to finish)
2. to help you learn to manage larger scale projects and less than detailed instructions
3. to give valuable insight to our clients in how they can improve their business

This project will be completed in several stages. You will need the data set available on Canvas to complete this assignment. You will use both descriptive (numerical and graphical) and inferential statistical techniques to analyze the data set.

BIG PICTURE STATISTICS HANDOUTS ~ DUE SEPTEMBER 1 AND 8 - 2 @ 10 POINTS EACH
SEE CLASS WEBSITE/HANDOUTS GIVEN IN CLASS

PROJECT UPDATES ~ 5 @ 5 POINTS EACH
SEE CANVAS QUIZZES

DESCRIPTIVE DRAFT WITH ≥ 3 QUESTIONS ~ DUE SEPTEMBER 15 - 15 POINTS
SUBMIT A PAPER COPY IN CLASS

ANALYSIS PLAN ~ DUE SEPTEMBER 22 - 15 POINTS
SUBMIT A PAPER COPY IN CLASS

DRAFT WITH ≥ 8 QUESTIONS ~ DUE OCTOBER 20 - 25 POINTS

FINAL DRAFT ~ DUE NOVEMBER 19 - 150 POINTS
AN EXTENSION TO NOVEMBER 29 MAY BE APPROVED UPON REQUEST
A DETAILED SCHEDULE FOR COMPLETION WILL BE REQUIRED

EXPECTATIONS:

- You are to clearly present all findings in a type-written, neatly formatted and organized report. You are to assume that you are writing for a client who has commissioned you to complete this study.
- Your report should contain compelling statistical evidence to support your claims as well as clear explanations of the question you are addressing and the answer that you are presenting.
(The average person should be able to read your report and understand what the results were, and the statistician should be able to read your report and be convinced based on the statistics that your conclusions are accurate).
- Your report will be developed in four stages.
 1. The first will contain a strictly descriptive analysis of the data (including numerical and graphical descriptions) that address at least 3 of the questions posed. They should include at least 3 different types of tests/analysis for better feedback on your work. The organization of your report should be starting to emerge in this draft (even though you do not have all the questions addressed or the inferential content yet). Develop an organizational structure that is based on the questions asked that makes it easiest for the client to understand (most of the questions fall into groups with a common theme, but they are not presented this way).

2. The second will be a list (handwritten by the questions is fine) of all the questions and which available analysis tool you will use to address the questions (include the name of the test/tool you will use and the names of all the variables that you will need to complete the test). State the assumptions for the test, and note what you will do if you don't meet them.
 3. The third will include everything from the first draft (plus revisions) as well as the inferential analysis (with conclusions) for at least 8 of the 16+ questions. The purpose of this draft is accountability for working on the project and providing feedback as to how your analysis is going.
 4. The final phase will be the completed draft of all questions (including revisions).
- Your conclusions for each question should include a discussion of what the statistic tell us is true about the data. You should also proceed one step future and suggest what your conclusions imply about Ma 105 and its students. Be sure to use a standard reporting style to present your test data (APA is good when presenting statistical results).
 - Recall that every report should include an introduction, an executive summary (different from an introduction and not an abstract) as well as a conclusion summarizing the study with recommendations.
 - As a final section to your report, you are to include a brief appendix outlining the technology used to address the various questions and create your report and include a (brief) discussion of what each tool did most effectively (or what worked poorly and needed a different tool). The purpose of this section is to demonstrate that you have learned to use technological tools to efficiently solve problems (problems of doing statistical analysis, keeping track of your work/results, presenting both written and graphical work clearly and effectively).
 - A side note on collapsing variables: sometimes it is necessary to group different categories together to get enough people in each group. For instance, you may need to group some of the high school types together or perhaps some of the grade categories, such as the Ds and the D-s. Just be careful not to collapse to insignificance.

GRADING RUBRIC FOR FINAL DRAFT:

- Application of Math
 - 45% Correctly used appropriate tests to address questions
 - 8% Constructed at least one additional (distinct, relevant, etc.) question and addressed it
 - 12% Appropriately incorporated descriptive stats
- Communication
 - 20% Clear communication of an answer to the question based on the statistics (including clear comm of statistical ideas for a non-technical audience) and appropriate interpretation results for the reader
 - 5% Clarity with conciseness (many students need to write more to make it to the clarity threshold)
 - 0% Penalty for inappropriate usage of English, poor presentation, improper type-setting, lack of professionalism, failure to follow direction, etc.
- Use of Technology
 - 5% Neat, easy to read graphics that are intentionally designed to communicate effectively
 - 5% Brief appendix outlining the technology used to address the various questions (includes appropriateness of the tool for the task and demonstrated that technology skills were developed to efficiently solve problems)

QUESTIONS TO BE ADDRESSED:

1. With 95% confidence how many students can we expect to pass Ma 105?
2. Is the student's grade in Ma 105 consistent with their Math ACT score?
3. Do students in "tech" majors do better in Ma 105 than students in "non-tech" majors?
4. Do men have a greater chance of passing Ma 105 than the women have?
5. What is the average Math ACT score of students who pass Ma 105 with 90% confidence? What is the difference between the average Math ACT scores of students who pass and who don't pass with 90% confidence?
6. Do students who take Ma 103 first do better in Ma 105 than students who did not take Ma 103? Can we assume that the variability of students who have taken Ma 103 first and those who haven't is the same?
7. Consider three possible categories for placement test results: didn't take it, placed into Ma 103, and placed into Ma 105. Is the average Ma 105 grade different for these three groups?
8. Do the "tech" majors and "non-tech" majors have equal representation in the class?
9. Does one high school type better prepare students for Ma 105 than the others? Is grade independent of high school type?
10. Do the students in the fall do better in Ma 105 than the students in the spring? Is there more variability in the grades in the fall or the spring?
11. Do students who took Ma 103 get the same (or similar) grade in Ma 105 as they did in Ma 103?
12. Do the male students do better in Ma 105 than the female students?
13. Is the grade distribution (# As, Bs, etc.) independent of gender?
14. Is the student's grade in Ma 105 consistent with their GPA?
15. Is the grade distribution (# As, Bs, etc.) different in the fall and the spring?
16. Is the student's grade in Ma 105 consistent with their Composite ACT score?
17. Are there other questions that should be asked? (In case you were wondering, the answer to this question is almost always yes) Address at least one more distinct, relevant question. Consider including more than 2 variables.