UNIVERSITY OF TEXAS AT AUSTIN

The Final Project

Congratulations! You are almost at the end of your *Applied Statistics* course. Many of you already independently chose a topic on which to work for their final project. Those of you who have not done so already **can still** decide to pursue a topic we talked about in class (or you worked on in a previous project). However, if nothing piqued your interest yet, fear not: here are some choices of topics to choose from.

All written work handed in by the student is considered to be **their own work, prepared without unauthorized assistance.** The University of Texas provides detailed information on prohibited conduct at https://catalog.utexas.edu/general-information/appendices/appendix-c/student-discipline-and-conduct/. In particular, "*Plagiarism* includes, but is not limited to, the appropriation of, buying, receiving as a gift, or obtaining by any means material that is attributable in whole or in part to another source without any indication of the original source, including words, ideas, illustrations, structure, computer code, and other expression or media, and presenting that material as one's own academic work being offered for credit or in conjunction with a program course or degree requirements."

Please, include the following statements along with your electronic signature in your final-project submission.

The University Code of Conduct

"The core values of The University of Texas at Austin are learning, discovery, freedom, leadership, individual opportunity, and responsibility. Each member of the university is expected to uphold these values through integrity, honesty, trust, fairness, and respect toward peers and community. As a student of The University of Texas at Austin, I shall abide by the core values of the University and uphold academic integrity. I have complied with the UT Honor Code during my completion of this project."

Signature:

1. PROJECT IDEA #1.

Walkability and rent. Go to https://www.zillow.com/. Then, either look at Austin as a whole, or at the neighborhood in which you currently live, or at a neighborhood in which you would like to live but do not (yet!). For the latter two options you will need to scroll a bit down the webpage. Concentrate on *one bedroom apartments*. This is essential! Do not mix-and-match different properties!!!

(i) Pick out 20 different apartments in the above category. For each of the chosen apartments, record the following in a table: the monthly rent, the Walk Score, and the Transit Score. Make your table neat and display it.

(ii) Make a scatterplot of the Walk Score and monthly rent. Be careful about your choice of the explanatory and response variables. Perform a simple linear regression analysis and report your findings. The best way to do this is to output the summary from R and accompany it with your own conclusions/comments about the obtained result.

(iii) How much in rent do you predict to have to pay if your required Walk Score is 70?

(iv) Imagine that your friend wants an apartment with a WalkScore of 90. What would you report to them based on your statistical analysis?

(v) Make a scatterplot of the Transit Score and monthly rent. Be careful about your choice of the explanatory and response variables. Perform a simple linear regression analysis and report your findings. Again, the best way to do this is to output the summary from R and accompany it with your own conclusions/comments about the obtained result.

(vi) How much in rent do you predict to have to pay if your required Transit Score is 60?

(vii) Think of an improvement to the statistical analysis you used above. I am not referring to data presentation or choice of software, but to the model and the statistical tools you used. Describe your proposed improvement, and explain why you think it would be an improvement.

---

2. PROJECT IDEA #2.

Recreational activity vs. major. Conduct a survey of sufficiently many of your friends, relations, and colleagues asking them about their major in college and their favorite recreational activity (from the offered choices; say, running, swimming, tennis, other). So, your first task would be to create a survey (say, using a Google Form as I did for your project #5). Then you would collect the data and visually display them. How would you organize your survey results in a table? What is your research question? Which statistical procedure is appropriate? Conduct the test and provide your findings!

---

3. PROJECT IDEA #3.

Education (or age or major) vs. most credible medium The problem I have given you in the exam was a bit dated since it only allowed for "radio", "newspaper," and "television" as categories of media people rely on. Here, you would first figure out what the possible values of the two variables would be. For the age, you would have to create age bins to be able to treat age as categorical (say, under 25, 25-35, older than 35). Then, you would create a survey (say, using a Google Form as I did for your project #5). Then, you would collect the data and visually display them. How would you organize your survey results in a table? What is your research question? Which statistical procedure is appropriate? Conduct the test and provide your findings!

---

4. PROJECT IDEA #4.

Case study: Using stents to prevent strokes The very first section in our textbook contains a case study. Now that you know more statistics, you can actually complete the necessary tests. Your first step would be to give some context to the research problem. Then, consider the original data sets available at

https://www.openintro.org/data/

Then, you would want to visualize the data. Propose the appropriate type of statistical inference and procedures (say, a confidence interval and a hypothesis test in this case). Write down your conclusions.

---

5. PROJECT IDEA #5.

Smallpox in Boston, 1721. **Subsection 3.2.4** in your textbook contains a contingency table for the data set available here

https://www.openintro.org/data/index.php?data=smallpox

For this project, you should start with some historical background. Then, you should visualize the data. What is your research question? Which statistical procedure is appropriate? What are your findings? Your conclusions?

---

6. PROJECT IDEA #6.

Heart transplants Consider **Exercise 2.26** from the textbook. Use the original paper cited on the bottom of that page and the exercise statement as your guide for what you need to do. If needed, consult the solution to the similar **Exercise 2.25** available in the back of the textbook. Conduct the appropriate test and write down your conclusions.

---

7. PROJECT IDEA #7.

Global warming **Exercise 3.15** in your textbook is interesting. The original data are available at

https://www.openintro.org/data/index.php?data=global_warming_pew

Start with some background on global warming as scientific fact. Then, consider the data. Do you know of any background information that might explain what's going on? How would you visually display the data? What would the two-way table of counts look like? What is our research question? Which statistical procedure is appropriate? Conduct the test and provide your findings?

---

8. PROJECT IDEA #8.

Health coverage and health status **Exercise 3.16** in your textbook is also rather interesting. The original data are available at

https://www.openintro.org/data/index.php?data=health_coverage

Start with some information about the structure of health insurance in the US. Why is the system organized the way it is? What are its origins? What would your ideas be about the potential relationship between the health coverage and the health status had you not seen the data? If this is impossible to imagine, survey your friends and relations.

Now, focus on the data. How would you visually display the data? What would the two-way table of counts look like (for the sample size, go into the original data set, please)? What is your research question? Which statistical procedure is appropriate? Conduct the test and provide your findings!

---

9. PROJECT IDEA #9.

Burger preferences. As opposed to the above follow-ups on textbook problems, you could conduct your own survey of your friends, relations, and colleagues on their burger preferences as they relate to another one of their traits (say, gender or major in college). Some of the chains mentioned in **Exercise 3.17** exist in Austin, but there are others. Your first task would be to create a survey (say, using a Google Form as I did for your fifth project). Then you would collect the data and display them graphically. How would you organize your survey results in a table? What is your research question? Which statistical procedure is appropriate? Conduct the test and provide your findings!

---

10. PROJECT IDEA #10.

Full body scan. Consider the data available in **Exercise 6.36**. Your task would be to conduct a similar survey by asking the people about their views on full body scans on airports as they relate to their major in college. Your first task would be to create a survey (say, using a Google Form as I did for your fifth project). Then, you would collect the data and display them graphically. How would you organize your survey results in a table? What is your research question? Which statistical procedure is appropriate? Conduct the test and provide your findings!

---

11. PROJECT IDEA #11.

Lobbying. You might consider updating or redoing the sample paper available here:
https://www.openintro.org/go/?id=analysis-sopa_pipa&referrer=/book/os/index.php.
Is there a different issue you are passionate about and would like to find out the legislative leanings on? Are there other interesting data available on: https://www.propublica.org/?

---

12. PROJECT IDEA #12.

Population growth. You might consider updating or redoing the sample paper available here:
https://www.openintro.org/go/?id=analysis-county_census&referrer=/book/os/index.php.
Are there other interesting data available on: https://data.census.gov/cedsci/?

---

13. PROJECT IDEA #13.

Hans Rosling. Learn more about Hans Rosling. One starting point is his obituary you can search for at:
https://www.newsecuritybeat.org/
Explain at least five of his data visualization ideas and how they might change our understanding of the world and our policy.

---

14. PROJECT IDEA #14.

The salary gap. Learn more about the salary gap both with respect to gender and with respect to other demographics. One starting point is the *Bureau of Labor Statistics*:
https://www.bls.gov/cps/earnings.htm
For the data analysis: look at this excel spreadsheet: https://www.bls.gov/cps/womens-earnings-tables-2020.xlsx. It contains the statistics in terms of medians (which is appropriate for salary/income data). There is a test of equality of medians which is a special implementation of the chi-squared test. Here is the Wikipedia entry on it: https://en.wikipedia.org/wiki/Median_test.
You would include the following in your final project:

- Some background information (with references). Examples: https://www.pewresearch.org/fact-tank/2021/05/25/gender-pay-gap-facts/, https://en.wikipedia.org/wiki/Racial_pay_gap_in_the_United_States, https://www.pewresearch.org/fact-tank/2016/07/01/racial-gender-wage-gaps-persist-i https://www.census.gov/library/stories/2021/03/unequally-essential-women-and-gender-pay-gap-duri html
- The appropriate data visualization.
- A description of the median test.
- The implementation of the median test and your conclusion.

---

15. PROJECT IDEA #15.

When data meets [sic] diversity. Provide an overview of the content and statistics available in
https://sinews.siam.org/Details-Page/when-data-meets-diversity
and in the citations on the bottom of the linked article.

---

16. PROJECT IDEA #16.

Fuel efficiency. Pick a data set from https://www.fueleconomy.gov/feg/download.shtml. Think about a question about fuel efficiency you would like to answer. One option is to look at whether there is a significant difference between the fuel efficiency of cars with automatic transmission and those with manual transmission, e.g. Visualize the data. Clean up your data set so that you retain just the information that you need. Conduct a hypothesis test choosing a significance level you find appropriate. What are your findings?

---

17. PROJECT IDEA #17.

Gaming and food. Consider this article: https://academic.oup.com/ajcn/article/93/2/308/4597636
Please, try to ignore the horrible British food choices in the article. Provide an overview of the content and statistics available in it. Which graphs do you think the article would have benefited from? Provide at least five different ways in which you would improve or re-do an observation study and/or experiment on the same (or similar) topic.