MTH 1210 Introduction to Statistics – Final Project Information Spring 2018

The final project is worth 15% of your course grade. So how can you try to earn a high score?

1. **3% of the project comes from submissions throughout the term.**
    a. **Project Proposal, 1% (5 pts): Tuesday 30 January**: I will be talking to everyone in class to get your project ideas. If you are ill or otherwise cannot attend class on this day, you must talk to me next class (Tuesday 6 February). *Note: You can **definitely** change your topic if you change your mind later in the term. I still encourage you to talk to me about your new topic to ensure it satisfies the requirement.*
    b. **Hypothesis Check, 1% (5 pts): Tuesday 27 February:** Again, I will be talking to everyone in class to solidify your hypothesis. If you are ill or otherwise cannot attend class on this day, you must talk to me next class (Tuesday 6 March).
    c. **Data submission, 1% (5 pts): Tuesday 3 April:** Bring a printout of your data (just the lists of numbers) to class (or you can show me your data on your laptop or tablet). For this submission, your printout should fit on one page (front and back is fine). Data may be submitted up to two weeks late by 17 April for 3 points.

2. **12% of the project is the final paper worth 60 points, due by noon on Friday 11 May.**
    a. Give yourself plenty of time to gather your data, use the software to do all computations and graphs, and write interpretations and conclusions.
    b. Follow the guidelines document on page 3 and the checklist on page 5 to make sure you have all of the correct pieces.

3. **Avoid common mistakes!**
    - Make sure you have all of the required items!
        - Don't forget the abstract! Ask Kathryn if you need help.
        - Don't forget the CIs and interpretations.
    - Stop by Kathryn's office hours or make an appointment to go over your interpretations and data analyses.
    - Do not use the empirical rule for item 3d. Help for this piece is available on page 4.
    - You must have at least 30 observations for each group.

4. Some sample papers are available on Kathryn's website, http://sites.msudenver.edu/kvernig/

Project Guidelines

Purpose

To experience the collection and analysis of real data using the tools and techniques learned in this course. To demonstrate your understanding of statistical techniques and terminology, including the use of the statistical software StatCrunch.

Project Description

Describe a problem from your field of interest in which statistical tools may be used. Define the variable and the underlying assumptions you are using. State a hypothesis regarding the problem. The variable must be a numerical variable (not categorical or ranking). Describe the sampling techniques you are using. (The sample size for each of the two groups you are studying has to be at least 30.) Use the statistical methods and tools you have learned to analyze your data. Present your results in a well-organized, written report.

** TL;DR: Find some number you can compare between two groups. (See below for examples).

Project Ideas

These are high-level problem statements. You will need to refine the statement into a project by defining the variable to be studied, the populations, hypothesis, sampling methodology and data collection strategy.

You should try to NOT choose from this list! I'm just providing these ideas to indicate the types of comparisons that can be made. Try to come up with something that you are genuinely interested in comparing.

1. Compare the average waiting time for a nurse to answer a buzzer on two different floors of a hospital.
2. Compare the average resting heart rates (beats per minute) of men and women.
3. Compare the average length of hospitalization in a teaching hospital to that in a non-teaching hospital.
4. Compare the average price of prescriptions at a supposed discount drug store (e.g. Walmart) to a retail store (e.g. King Soopers).
5. Compare average monthly rent for comparable apartments or homes in two different locations.
6. Compare the average price of gas for stations close to an Interstate with those farther away from an Interstate.
7. Did the average age of a person's first marriage change from the early 80s to current times?
8. Compare the average run time for two types of movies (e.g. comedies and science fiction)
9. Compare the average performance of low-priced golf balls to expensive golf balls.
10. Are lift-ticket prices more expensive in one ski resort versus another?
11. Compare the average file download rates (Megabytes per second) over a given network for two different times of day.
12. Compare the average cost of getting an education in two different fields (e.g. medicine versus law).
13. Compare the average of the difference between the planned arrival time and the actual arrival time for two different airlines.
14. Compare the average number of pages in history books with the average number of pages in mathematics books.
15. Compare the average points per game scored by the Denver Nuggets at home to the average points per game on the road.
16. Compare the average number of ingredients in Asian recipes with the average number of ingredients in French recipes.
17. In the last 30 years, has the AL or the NL scored more runs per season (on average)?

NOTE: The data you collect should ***NOT*** be "yes/no" responses (e.g. from a poll) or "win/lose", etc. – you need quantitative data, not qualitative data.

The Report

A written report of 600-900 words, not including visuals or bibliography. It must be word-processed using 11 or 12 point font. All graphs and charts must be generated in StatCrunch and included in the electronic document (just like we do in labs). The report may be submitted either printed to my office or online via Blackboard.

Organization of the Report

- **Title Page (1 pts):** Your report must have a single title page that presents (i) the title of your report, (ii) your name, (iii) the date of submission of the report.
- **Abstract (2 pts):** The body of the paper must be preceded by a 50-150 word abstract that includes (i) the problem statement, (ii) the hypothesis that you are testing, (iii) the population you are studying, and (iv) the main results of your study. This is meant to be a summary of the entire report in one to two paragraphs.
- **Body:** The body of the paper should be organized as follows:
  1. *Introduction* **(2 pts)** - Describe the problem. Include background information that might be necessary for the reader to understand the problem. Why did you choose it? What is your hypothesis?
  2. *Data Collection Methodology* **(2 pts)** - When, where and how you collected your data. How was randomness achieved (you MUST randomly sample *at least* 30 observations per group!)? Were any adjustments made to the raw data (e.g. did you convert percentages to decimals or perform other calculations)?
  3. *Data Analysis* (Note: do all calculations in StatCrunch except part d.)

     For *each* of your two samples find:
     - **a.** Mean, standard deviation, five number summary and range **(4 pts)**
     - **b.** Relative frequency table **(2 pts)**
     - **c.** Stem and leaf plot, histogram, boxplot **(2 pts each, 6 pts total)**
     - **d.** Percentage of observations that fall within 1, 2, and 3 standard deviations of the mean (note: this is the only thing you will not be able to generate with MINITAB/StatCrunch, see page 4 for help) **(3 pts)**
     - **e.** Draw any conclusions that seem appropriate. Include a discussion about the skewness of your data by using your graphs and statistics. Be complete and thorough. State and support your ideas. You proposed a hypothesis in the introduction – do the data appear to support it? ***This is the most important part of the Data Analysis!*** **(5 pts)**
  4. *Inferential Statistics* (Note: do ALL calculations in StatCrunch.)
     - **a.** Find **(2 pts)** and interpret **(4 pts)** the confidence intervals for each of your two groups (copy/paste MINITAB/StatCrunch output).
     - **b.** Explain if your data is nonpooled or paired and why. **(1 pt)**
     - **c.** Perform all 6 steps of the two-sample hypothesis: state the hypothesis, choose a significance level **(2 pts)**, report the *p*-value **(2 pts)**, make a decision based on the *p*-value, and interpret your result. Make sure to copy/paste your MINITAB/StatCrunch output for this problem (**2 pts)**. Also include the formula for the appropriate test statistic – either paired t or nonpooled t (can be neatly handwritten).
     - **d.** Describe your findings and the results of your hypothesis test in a few sentences. Comment on how the test supports your data analysis. **(5 pts)**
  5. *Conclusions* - **Detailed discussion that includes your overall conclusions, comments, interpretations, and suggestions. (10 pts)**
- **References:** Web sites or documents used as data sources. Books or articles other than our textbook referenced in your analysis.
- **Data:** The last page of your paper should be a table with your data. You may need to make several tables in order to fit it all onto one page. **(1 pt)**

Language and Level

Assume that your reader has a good working knowledge of the ideas covered in MTH 1210. Use correct spelling, complete sentences, and correct grammar. Language skills are not a large component of the grade, but well-written papers inevitably get better grades. (Grammar, sentence structure, etc: **(5 pts)**)

Paper item 3.d. Percentage of observations that fall within 1, 2, and 3 St. Dev. of the Mean

(This document is available as a .docx to download from Kathryn's website so you can easily include this information in your paper.)

You will need to compute 6 values for each sample: the mean plus and minus 1 St. Dev, plus and minus 2 St. Dev, and plus and minus 3 St. Dev. Then, count up how many of your observations in your sample lie between these values, divide by the total observations in that sample, and multiply by 100 to get the percentage.

Important! I am NOT looking for the empirical rule! I want to know your specific percentages for your data.

The following set up may be helpful in finding these percentages.

Sample 1:

$\overline{x_1} - s_1 =$ _____ Number of observations between these values: _____

$\overline{x_1} + s_1 =$ _____ Percentage: _____

$\overline{x_1} - 2s_1 =$ _____ Number of observations between these values: _____

$\overline{x_1} + 2s_1 =$ _____ Percentage: _____

$\overline{x_1} - 3s_1 =$ _____ Number of observations between these values: _____

$\overline{x_1} + 3s_1 =$ _____ Percentage: _____

Sample 2:

$\overline{x_2} - s_2 =$ _____ Number of observations between these values: _____

$\overline{x_2} + s_2 =$ _____ Percentage: _____

$\overline{x_2} - 2s_2 =$ _____ Number of observations between these values: _____

$\overline{x_2} + 2s_2 =$ _____ Percentage: _____

$\overline{x_2} - 3s_2 =$ _____ Number of observations between these values: _____

$\overline{x_2} + 3s_2 =$ _____ Percentage: _____

Checklist

- ☐ Title Page (1pt)
- ☐ Abstract (2 pts)
- ☐ Introduction (2 pts)
- ☐ Data Collection Methodology (2 pts)
- ☐ Mean, St. Dev, Five Number Summary, Range (4 pts)
- ☐ Relative Frequency Table (2 pts)
- ☐ Stem and Leaf Plot (2 pts)
- ☐ Histogram (2 pts)
- ☐ Boxplot (2 pts)
- ☐ Percentages within 1, 2, 3 st. dev (3 pts)
- ☐ Data Analysis Conclusions (5 pts)
- ☐ CI for each group (2 pts)
- ☐ Interpretation of CIs (4 pts)
- ☐ Nonpooled versus paired (1 pt)
- ☐ State Hypothesis and choose a significance level (2 pts)
- ☐ Report p-value (2 pts)
- ☐ Include StatCrunch results (2 pts)
- ☐ Interpret and describe result (4 pts)
- ☐ Conclusion (10 pts)
- ☐ Grammar, sentence structure, etc (5 pts)
- ☐ Include one page with your data as the last page of your paper (1 pt)

Total possible points on paper: 60 points