

# Final Project Guidelines

## MTH 1210: Introduction to Statistics

### 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 (data collection, data analysis, and inference) and terminology, including the use of Minitab or similar statistical software.

### Project Description

Describe a problem from your field of interest in which statistical tools may be used to compare a numerical variable in two different groups or under two different conditions. Define the variables and describe the data collection (sampling) techniques you use. The samples size for each group you are studying must be at least 30 (at least 60 observations total). State a hypothesis regarding the problem, state the statistical methods to be used, and state the underlying assumptions associated with those methods. Use the statistical methods learned in class to perform exploratory data analysis and to test the hypotheses.

### Form of the Report

**Title Page:** Your report must have a title page that presents (i) the title of your report, (ii) your name, (iii) the date of submission of the report.

**Body:** The body of the paper should be organized in sections as follows:

1. **Introduction** – Describe the problem (including the **problem statement**). Why did you choose this problem? What is your **hypothesis**?
2. **Data Collection Methodology** – What variable(s) were recorded? When, where, and how did you collect your samples? Were any adjustments or conversions made to the raw data? How big are your sample sizes? Can the samples be expected to be representative of the populations? If so, why? If you found the data online, give the url for the website.
3. **Exploratory Data Analysis and Descriptive Statistics** – **For each sample:**
  1. Find the **mean** and **standard deviation**. Discuss what these values tell you about each sample.
  2. Find the **five number summary** (minimum, first quartile, median, third quartile, and maximum).
  3. Make a **histogram** and discuss the shape of the histogram (Is it skewed? If so, in which direction? Is it unimodal or multimodal? Are there visible outliers? etc.).
  4. Also, make **side-by-side box plots** of the samples, and comment on any differences between the two sets of data that are made apparent by the box plots.
  5. Include in your report the **Minitab output** showing results of the above calculations and graphs.
  6. State any **conclusions** you can from the exploratory data analysis and support those conclusions. For example, are there any clear differences between the two groups? If so, what are they? Are there unusual patterns or outliers in the data? Is there any evidence in support of your original hypothesis? etc.

#### 4. *Inferential Statistics* -

1. For each sample, find the **confidence interval** for the population mean. In a sentence or two, discuss the information that the confidence intervals convey.
2. **Test the hypothesis** (this is the *main purpose* of this project):
  - State the null and alternative hypotheses.
  - Give the formula for the test statistic.
  - Give the computed value of the test statistic. This value should be stated explicitly in the text as well as shown in the Minitab output.
  - Give the p-value. This value should be stated explicitly in the text as well as shown in the Minitab output.
  - State the conclusion using level of significance  $\alpha = 0.05$ .
  - Describe your findings and the results of your test in a few sentences.
3. Include in your report the **Minitab output** showing the results of the confidence interval and hypothesis test calculations.
5. **Conclusions** - Discuss in detail your conclusions, comments, interpretations, and suggestions, paying particular attention to the results of the inferential statistical procedures.

#### **Language and Level**

You should use correct spelling, complete sentences, and correct grammar. You may expect that well written papers inevitably get better grades. Write your paper assuming that your reader has a good, working knowledge of the ideas covered in this class.

#### **Use of Computer**

Most of your calculations should be done with **Minitab**. You may use another statistical software programs only by permission of the instructor. You can copy and paste your output and graphs into your report which should be typed using a word processor. But you must **explicitly state the values of the descriptive statistics, the confidence intervals, the test statistic and the p-value** in your write-up.

#### **Project Proposal**

You should submit a project proposal by **Wednesday, Sept. 11**. Use the **project proposal form** from the course website. The proposal should include your problem statement, the variable(s) to be investigated, and your proposed data collection methods. I'll read it and respond with suggestions as a basis for discussion and possible improvement.

**Groups of two students** are welcome to join forces and work together on one topic. They will submit one proposal and one report. Both team members should meet together with me for consultations and help.

**The final projects are due by the last regular day of class.**

---

## Problem Statement and Data Collection Ideas

The usual way to carry out this project is to first determine the problem statement, and then decide what variables should be observed and how the data should be collected. The problem statement should involve comparing the means of a **quantitative** (numerical) **variable** in **two** different **groups** (or under two different experimental conditions). See the examples below.

Since the statistical method (hypothesis testing) that will be used requires that the data be random samples (or be from a randomized experiment), your data collection method should adhere to this assumption as closely as possible.

It may be possible to find existing data, for example on the internet or in the library. If you do not collect the data yourself, you must state the source of the data in your report, with enough detail so that I can find the data myself. Data found in textbooks of any kind will not be accepted.

The **sample size of each group** you are studying has to be **at least 30** (so at least 60 *total* observations).

## Samples of Problem Statements

1. To determine if there is a significant difference in the average waiting time for a nurse to answer a buzzer on two different floors of a hospital.
2. To determine if women, on average, have a faster resting heart rate than men.
3. To determine if the average length of hospitalization is significantly longer in a teaching hospital than in a non-teaching hospital.
4. To determine if there is a significant difference in the temperature forecast between channels 4 and 9.
5. To determine if the average waiting time at the highway on-ramp is shorter at 7:00 am or at 7:45 am.
6. To determine if country music songs are longer, on average, than rock songs.
7. To determine if the average number of pages in math books in the library is greater than that of art books.
8. To determine if the Nuggets tend to score more points in home games than in away games.
9. To compare the average price of prescriptions at a mail order drug store to a retail store.
10. Compare the average monthly rent for comparable apartments in two different areas.
11. Compare gas prices for stations close to an Interstate with those farther away from an Interstate.
12. To decide if the average age of a person's first marriage changed from the early 80's to current times.
13. Compare the average run time for two different genres of movies (i.e. comedy and science fiction).
14. Compare the average travel time for a bus from Stop A to Stop B in the morning to the time it takes to travel from Stop B back to Stop A in the evening.
15. Compare tipping percentages at a restaurant for men versus women.
16. Compare the number of pictures in two different newspapers or magazines.

17. Compare the commute time from home to work if you leave just-in-time versus leaving 10 minutes early.
18. Compare file download rates (megabytes per second) over a given network for two different times of the day or for two different service providers.
19. Compare the average number of ingredients in Asian recipes with the average number of ingredients in French recipes.
20. Compare the reading rates (in words per minute) of two different populations (e.g. men versus women, college graduates versus non-graduates, etc.).
21. Compare the average GPA for Metro students in two different age groups.