ENGR 371 project

Winter 2024

As engineers, it is important that you be able to conduct investigation of complex problems. A typical method for this is to form a hypothesis, design an experiment to test the hypothesis, conduct the experiment and then analyze the data and draw conclusions.

In this project you will do all these four things. The overall problem here is to learn something about a large pool of objects/numbers (the "population") by looking at only a few of them (the "sample").

Hypothesis

A hypothesis is something that you think is true and want to validate. It should also be something that is important and useful: there should be a purpose as to why you have chosen your hypothesis to test. An example hypothesis might be: "2 cars pass through an intersection after 5 seconds after their traffic signal has turned yellow". This would be useful to know when deciding on signals to give pedestrians or cyclists who use the intersection.

The population that you will study might be cars going through yellow lights in all intersections with traffic lights in Montreal. Another population might be all intersections with traffic lights, in some smaller geographical entity, like Hampstead, or lights along a certain road.

Whatever you choose as a population, this may lead you to modify your hypothesis to be "2 cars pass through an intersection 5 seconds after the traffic signal turned yellow, in Hampstead."

In preparing your hypothesis you should do some research to try to understand how reasonable your hypothesis is, and ultimately how your results fit within a wider context. You will be asked to write about what research you did when forming your hypothesis.

Experiment Design

In designing your experiment, you should endeavor to get a random sample (in the meaning of what we learned in the course) from the population you have chosen. Given your resources this may not be easy. For example, you may not be prepared to stand at traffic lights in the middle of the night to gather data. You might not be drawing samples from the entire population. Also if your sampling takes place in a short window, or always at the same time every day, then the samples may not be independent. For this you may want to adjust your hypothesis. But be careful about defining a hypothesis so narrow that it has no use.

Part of your experimental design is how many samples you should take. How to determine this will not have been covered when you write your proposal, but include some discussion of the number of samples you plan to take. Aim to take at least 50 samples and more if possible.

Conduct your experiment

Conduct the experiment you designed. It may be that as you conduct the experiment, you notice things that may bias your results that you had not thought of: for example you plan to go out on a certain

Saturday afternoon to conduct your experiment, but when the day arrives it is very cold and windy so you decide to conduct the experiment later. Afterward you wonder if the weather has an impact on the population you are studying. Pay attention to these details.

If your project involves human subjects, you must fill out the Abbreviated Summary Protocol Form (SPF). You submit this with your proposal. You must have received approval of your form before doing any interviews. Note that part of this form is that you must received informed consent before conducting interviews. Before filling out this form read the document "Submission on the Ethical Dimensions of your Statistical Study". Note that your submission on ethical dimensions of your project is only due in your final submission, but reading this document may help you prepare the Abbreviated SPF.

Analyze your data and Draw Conclusions

To begin with calculate the sample mean, the sample variance and the sample median. Describe the distribution of the sample mean. Generate a confidence interval on the sample mean (decide on an appropriate confidence bound). Also set up a hypothesis test for your hypothesis and test its validity.

Software use in your project

Using software for your project will make the analysis faster and allow you to do many calculations that would otherwise be extremely time-consuming. Your group can select whatever software you want to use for your project. Concordia holds a certain number of licenses for Minitab and MATLAB. The languages R and Python are open-source and are thus also possible options without license limitations and have many available libraries to help you, and good documentation.

You will hand in all files and output that you generate as part of your project submission.

Getting started with Minitab

You should enter the data you gathered into the Minitab software including items like dates and times of when samples were taken. With the minitab software you can calculate simple statistics like sample mean etc using the stat->basic stats->display descriptive statistics and then in the panel that comes up select the panel with your data (double click) so that it appears in the variables window. Push the statistics button to get a list of statistics that you can select. When done selecting push ok to get your statistics.

You can also use Minitab to calculate confidence intervals and do Hypothesis testing (stat->basic stats->one sample z or one sample t.

There are many other features to explore in Minitab that you can use to understand your data better.

Getting started with MATLAB

There are many software packages that you can use to calculate confidence intervals and do hypothesis testing. You should use one of these for your project.

As an example, the matlab commands ztest and ttest can be used to give both confidence intervals and do hypothesis testing. For ztest,

[h,p,ci,zval] = ztest(x,m,sigma,varargin)

The above is for calculating an hypothesis test on the mean where the variance of the population is known or where we can use the sample variance because we have taken a large sample. Here x is a vector containing the data collected. m is the value for your null hypothesis (H0: μ=m). sigma is the population standard deviation or the sample standard deviation for a large sample. varargin is a spot to put in various parameters or arguments of the hypothesis test. For example, to change the significance level (α) (the default is 0.05), or to make it a one-sided test. For more details type help ztest in matlab.

The above command would return something like

h = 0

p = 0.9927

ci = *2×1*

73.2191

76.7975

zval = 0.0091

The h=0 indicates that the test fails to reject the null hypothesis. The p-value is then given, followed by the confidence interval on the mean. Finally, the test statistic is given.

ttest is a similar function for use when your test statistic follows a t distribution.

In what you hand in provide the MATLAB code and output you used.

Project Procedures

1. The first step will be team formation. This will be done by either the instructor or the project TA. There will be teams of 4. Each student will receive an email with the names and emails of their team mates.
2. The expectation is that each student will spend 12 hours working on this project. For a team of 4 that means 48 hours in total.
3. Each team will keep a blog, of every hour that is spent on the project by any team member. A sample blog excel is found on moodle.
4. Each team should prepare a proposal (see below) and submit it to through moodle. There is no hard deadline on this, but you should aim to get the proposal done by **February 15**. The project

TA may approve the proposal with minor changes or ask for major changes, in which case the changes should be made and the proposal should be submitted again. Thus, it is good to submit earlier rather than later.

5. Each team will then conduct their experiment, and then analyze the data, come to conclusions, and submit the final submission.

6. The final submission (see below) is due on **April 12, 2024**

Proposal

Your proposal should contain

1) The population you will take your sample from.

2) The hypothesis that you will test in your project.

3) A summary of the research you did when preparing your hypothesis and what you learned. Keep this section to no more than ¾ of a page. Provide references.

4) A justification on why this hypothesis is interesting or important.

5) The proposal for the experiment you have designed to test your hypothesis including

    a. How you will select certain members of your population to measure. i.e. How you will do your sampling.

    b. The quantity you will measure from each member of the sample.

    c. The size of your sample.

6) The Abbreviated Summary Protocol Form if your experiment involves human subjects.

The proposal at this stage will not be marked for grades, only critiqued. However, the items in the proposal will be submitted for the final submission where they will be graded. Note that you can correct or update what you wrote at the proposal stage, for the final submission.

Final submission

In the final submission submit

1. An expectation of originality form signed by all group members

2. The Team Blog signed by all group members

3. Your proposal, (all 4 parts). They may have been updated/corrected from what you originally submitted.

4. All of your sampling data

5. Two page submission on the Ethical Dimensions of your project. See the document on this submission.

6. Notes and details of your experiment. Dates that data was collected, information not clear from experimental design (for example weather that day if it was relevant to your experiment.)

7. The analysis of your data. See the section earlier in this document for ideas on what analysis you should consider doing. Incorporate plots from Minitab in your results.

8. A conclusion on whether your hypothesis was reasonable, and justification for your conclusion.

9. Comments

    a. After the analysis, you are required to comment on the results. In particular you should comment

        i. Did your method of sampling result in a random sample?

        ii. If your sample was not a random sample what sorts of measures could you take if you were to do this project again, to get a random sample?

        iii. Based on the experiment, would it be appropriate to write a revised hypothesis ("in Hampsteand, 5 cars go through each yellow light.")

        iv. Comment on whether you think your results can be extrapolated to draw more general conclusions, perhaps on wider populations. State your opinion **and then back it up with well-reasoned arguments.**

This is a vital part of this project and you should spend some time doing a good job on this part.

Your final report should consist of one file, either a single pdf file or a word document. You are also required to upload any supporting calculations files (MATLAB, Minitab, R, etc.).

Grading

Overall Organization and completeness 10%

Hypothesis 5%

A summary of your research and what you learned. 5%

Why is this Hypothesis interesting or important 10%

Experiment Design 15%

Ethical Dimensions on your project 5%

Notes and details on your experiment and submission 5%

Analysis 15%

Conclusion on hypothesis and justification 10%

Comments 20%

Individual +/- 10% (for students in groups where either one or a few teammates did a large proportion of the work, or where one or more did not contribute much. Most students will get 0 for this. The blog will be used for this)