

Introduction to Quantitative Research (GEOG/ENST 2006)

Assignment – Choose your own stats adventure!

March 22, 27, *plus* March 29, April 3, 2019

Due on the 6th day following your *second* lab session (at 23:55)

Location: Loeb A200

Introduction:

This lab exercise will serve as your first chance to examine a data set on your own, without step by step instructions. You will pose your own research questions and answer them using the quantitative methods that you have learned in this course.

You are working with EcoLiving, a not-for-profit organization that seeks to promote environmentally sustainable and healthy choices in urban centers. They have just concluded a study in summer 2018 that looks at using bikes for transportation in the city and they need a report that analyzes the data. Your employer is generally interested in any the following:

- Characteristics of the bike trips during that day and the relationships between these
- How this activity can improve personal health
- How this activity can be beneficial for the environment
- The relationship between bike trip distance/duration and any health/environmental benefits
- How interested people are in using bikes for transportation and whether this opinion is different based on the characteristics of their bike trips
- Differences in attitude and trip characteristics between men and women
- Whether there are any shortcomings in the data set

Background information:

The data set available to you (bike.Rdata on cuLearn) is the result of the study where participants without bikes were selected at random to do a day's worth of chores and other daily activities by bicycle instead of using a car. Each participant was provided with a cargo bike and instructed to complete the daily activities they had originally planned to do (school/work, shopping for groceries, shopping for clothes, going to the hardware store, piano lessons, meeting friends for coffee, etc.). The research tracked various aspects of their cumulative trips for the day, recorded information about the participants, including health indicators and determined some of the environmental benefits. At the end of the study, participants were asked whether they would use bikes as a mode of transportation in the future. Below is specific information about each variable.

ParticipantID - A unique identifier for each study participant

Age - The age in years of each participant

Gender – The gender (Female or Male) of the participant

Duration – The length of time in minutes that the participant used the bike for that day (not just biking)

Distance – The cumulative distance travelled in metres by the participants over the time elapsed

HR_before – The resting heart rate of each participant prior to getting their bike

HR_after – The heart rate of each participant when they returned their bike

Calories – The Calories (kcal) that each participant burned by using the bike

CO2 – The grams of carbon dioxide that were diverted from the atmosphere by using the bike relative to what each participant would have used if they took their car, bike or walked

Opinion – Each participant was asked the question “Now that you have had this experience, would you use a bike in the future in place of your regular mode of transportation” and their response is given here

Assignment [85]

Before you start writing the assignment and start working on the questions below, do a thorough exploratory analysis of the data set and consider the goals that EcoLiving has for this study. This will help you isolate variables of interest.

Write a professional report that analyzes this dataset for your client. Your report will need an introductory paragraph (Q1), three sections that analyze that data using descriptive, inferential and relational statistics, respectively (Q2-4) and a concluding paragraph (Q5) that summarizes your analysis. Insert tables and figures with captions that illustrate your analysis in your report and cite them as appropriate. An appendix must be included after the text and figures of your report. Copy and paste the R output into the appendix to show how you got all your results (figures, descriptive statistics, and inferential tests, etc.). This represents the 'show your work' element of this exercise. Without this, your report will not be evaluated.

Consider what you might focus on to answer questions that are related to at least some of your client's interests.

- 1) Write a paragraph (3-4 sentences) that introduces your analysis. This should include the purpose of your report and how your research will address your client's needs. [5]
- 2) Generate and report descriptive statistics (central tendency, variability and distribution shape, *as appropriate*) for all variables. Include at least one figure and one table in this section and refer to them in the text of your write up for this section. [15]
- 3) Use inferential statistics to test a hypothesis about a mean or median. For example, chose from the following according to the assumptions met: One or two sample test of difference of means (*t*-test or Wilcoxon test), paired *t*-test or Wilcoxon test, ANOVA or Kruskal-Wallis test [20]. You will receive *up to 5* bonus marks if you choose to conduct an ANOVA or Kruskal-Wallis test. Ensure that you provide the following elements in your report:
 - A) Write what you are trying to accomplish in plain language [3]
 - B) Declare the significance level you are testing at [1]
 - C) Write out your statistical hypothesis in the traditional way [2]
 - D) Report the test statistic and the *p*-value for your test [2]
 - E) Explain the result of the test (give your decision and what it means in plain language using the word significant) [4]
 - F) Make sure to select the appropriate statistical tests by *thoroughly examining and commenting on* test assumptions. Write a justification in the report and make sure to include your R output in the report appendix [8]

- 4) Use relational statistics to examine the **association** and **model the relationship** between two continuous variables. [30]
 - A) Write what you are trying to accomplish in plain language [5]
 - B) Declare the significance level you are testing at [1]
 - C) Write out your statistical hypothesis in the traditional way [4]
 - D) Report the test statistic and the p -value for your test [4]
 - E) Explain the result of the tests (give your decision and what it means in plain language using the word significant) [4]
 - F) Make sure to select the appropriate statistical tests by *thoroughly examining and commenting on* test assumptions. Write a justification in the report and make sure to include your R output in the report appendix [12]
- 5) Write a short paragraph that sums up and concludes your report. [5]

Important notes (read thoroughly):

- Throughout this lab you will be expected to *write in complete sentences* and report your data with *correct units* and a *reasonable number of digits*, and provide proper *captions and formatting* for figures and tables. In addition to the marking scheme above:
 - **5 marks** will be given for overall formatting and professionalism
 - **5 marks** will be given for spelling and grammar
- While you may have similar ideas as your fellow students, you must pursue your own work and make the task of marking your labs a varied and exciting experience for the TAs! (i.e., 'choose your *own* adventure')

How to do the various tests that we have not covered in lab 1-7:

Getting the mode of a categorical variable

- Menu/Statistics/Summaries/Frequency Counts
- Find the category with the most observations (that's the mode).

Equal variance

- Menu/Statistics/Variances/Two variances F-test (for 2 samples)
- Menu/Statistics/Variances/Levene's Test (for 2 or more samples); leave "centre=median"

Normal distribution

- Menu/Summaries/Tests of normality
- Select 'Shapiro-Wilk'
- Select the continuous variable(s) you want to test.
- If you have a continuous variable accompanied by a nominal variable (a factor) and you wish to test by each group, select the name of that variable under "Test by Groups:"

One-way ANOVA

- Menu/Statistics/Means/One-way ANOVA
- Select a categorical variable (factor)
- Select a continuous variable as the response variable
- Check pairwise comparisons of means to do a post-hoc comparison
- If the the equal variance by groups assumption is not met, you can select the Welch's test below. In that case, the last output will be the p-value associated with that test. Note the post-hoc test may not be valid in this case.

Kruskal-Wallis

- Menu/Statistics/Non-parametric test/Kruskal-Wallis
- Select a categorical variable (factor)
- Select a continuous variable as the response variable

Assignment Rubric

Your assignment will be assessed against the following criteria. These categories are meant to be illustrative, not prescriptive.

90-100% Exceptional - A superior performance with consistent strong evidence of:

- a comprehensive, incisive grasp of the subject matter;
- an ability to make insightful critical evaluation of the material given;
- an exceptional capacity for original, creative and/or logical thinking;
- an excellent ability to organize, to analyze, to synthesize, to integrate ideas, and to express thoughts fluently

80-89% Excellent - An excellent performance with strong evidence of:

- a comprehensive grasp of the subject matter;
- an ability to make sound critical evaluation of the material given;
- a good capacity for original, creative and/or logical thinking;
- an excellent ability to organize, to analyze, to synthesize, to integrate ideas, and to express thoughts fluently

70-79% Good - A good performance with evidence of:

- a substantial grasp of the subject matter;
- a good understanding of the relevant issues and a good familiarity with the relevant literature and techniques;
- some capacity for original, creative and/or logical thinking;
- a good ability to organize, to analyze, to synthesize, to examine the subject material in a critical and constructive manner

60-69% Satisfactory - A generally satisfactory performance with evidence of:

- an acceptable but basic grasp of the subject matter;
- a fair understanding of the relevant issues and techniques;
- an ability to develop solutions to moderately difficult problems related to the subject material;
- a moderate ability to examine the subject material in a critical and analytical manner

50-59% Minimal pass - A barely acceptable performance with evidence of:

- a familiarity with the subject matter;
- some evidence that analytical skills have been developed;
- some understanding of relevant issues;
- some familiarity with the relevant literature and techniques;
- attempts to solve moderately difficult problems related to the subject material and to examine the material in an analytical manner, which may only be partially successful

<50% Failure - An unacceptable performance

- In general, the above criteria are not met