BIST/STAT 5505.
Applied Statistics I (Fall 2019)
Small Project

The small project consists of a data analysis report on a dataset of your choosing *or* on the "CT Crash Datasets" from 2016, available on the HuskyCT course website. The source of this data is `https://www.ctcrash.uconn.edu/`, which contains much more data and spans multiple years.

CT Crash Data Description: Crash data for rural two-lane intersections and segments in Connecticut are collected in order to study highway safety. Generally, crash severity counts may be aggregated into three categories: PDO (property damage only), or B+C (possible injury or non-incapacitating injury), or K+A (fatal or incapacitating injury) crashes. Three different data sets are available on the HuskyCT course website, corresponding to the type of intersections. Dataset 3ST Intersections_CT.csv consists of $n = 385$ cases for 3ST (three-way stop controlled) intersections. Dataset 4ST Intersections_CT.csv consists of $n = 61$ cases for 4ST (four-way stop controlled) intersections. Dataset 4SG Intersections_CT.csv consists of $n = 102$ cases for 4SG (four-way signalized) intersections. Each dataset includes the following variables at each intersection: Intersection ID, Annual Average Daily Traffic (AADT) for major roads and minor roads, three dummy variables to indicate presence of lighting, presence of a left-turn lane, and presence of a right-turn lane, followed by skew angle, and then the counts of PDO, KA, and BC crashes.

If you choose to use a dataset other than the CT Crash Datasets that are provided on the course website, your dataset *must* contain multiple variables, with at least one quantitative variable and at least one qualitative variable on which all (or the majority) of sampling units are measured.

In the report, you must posit clear research question(s) and document a *thorough* exploration of the relevant data for answering the research question(s) using the graphical and numerical summary techniques as well as perform statistical inference procedures discussed in this class.

You may work alone on the project, but it is highly recommended that you work on this project with a partner. The typed final report is to be submitted to the instructor via email by 11:59pm, Friday, **December 6, 2019**. The report needs to be in a Microsoft Word (.doc or .docx) or pdf file. Name the report as LastName_FirstName_project.docx(pdf), where LastName (one of the group members if you work with a partner) and FirstName (one of the groups members if you work with a partner) are filled in appropriately. Only one report per group should be submitted.

The report should include the following components:

- Title Page

  This includes title of the project, your name(s), and email address(es).

- Abstract/Summary Page

  This includes an abstract or summary of your work.

- Section 1. Introduction and Description of Data

  This section should include, at minimum, (i) clear statement(s) about the research question(s) that you will investigate and (ii) which variables are qualitative and quantitative.

- Section 2. Exploratory Data Analysis

  This section should include labeled figures, graphs, and tables that are referenced and discussed within the main text (in Section 2 and also possibly elsewhere).

- Section 3. Statistical Inference

  This section should include statistical inferences such as confidence intervals, hypothesis tests, nonparametric inference procedures, and/or categorical data analysis techniques. Do NOT use tools and/or techniques not covered in this course (e.g., regression).

- Section 4. Discussion of Results and Concluding Remarks

  This section should include, at minimum, a summary of your findings. Discussion about further statistical techniques that could be used to explore the data, but are beyond the scope of this course, are also recommended for the section.

- Section 5. Appendix

  This section should include, at minimum, the computer code used in this project. You may also include the dataset itself if you choose your own data, or any other important additional materials.

- References (if applicable)

  Cite all sources, if any, used in your analyses/report. Citations should be formatted as follows: Authors (Year). Title. *Journal*, **Volume**, page numbers.