Econometrics 461: Economic Data Analysis   Spring 2024
Term Project Assignment

As noted in the course syllabus, the research paper is one of the major requirements of this writing-intensive course. To reduce the amount of frustration and uncertainty associated with the project, I have compiled a group of datasets of different types among which you can choose as the basis of your project.

- Among the datasets described below choose a specific variable as the focus of your analysis – this is your *Criterion variable.* This **MUST** be a numeric variable. Note that to conduct a meaningful statistical analysis of your criterion variable, it may be necessary to transform how the variable is measured. For example, converting a "gross" income variable to a "per capita" basis by dividing by population, or converting a "levels" variable to logarithms.
  - *If you would like to propose a topic using data that you collect/compile, you may do so. However, you need to speak with me first to get approval. This is to verify that you will be able to find data suitable for the nature of this project.*
- Next, choose a *Predictor variable* that you believe is associated (correlated) with your chosen criterion variable (this could come from a different data set than the source of your *Criterion* variable). As with the *Criterion* variable, this **MUST** be a numeric variable.
- Next, choose one or two variables – *different than your criterion or predictor variables* – to create a category variable that can be used to sub-set your overall sample into **at least three mutually exclusive sub-groups**. This could be variables related to geography, industry classifications, gender, ethnicity, etc. See the Term Project Grouping Examples workbook for examples.
- You may want to combine data from two or more of the datasets noted below.

Datasets
- County Labor Force and Unemployment Data
- County Real Wage Data
- County Real GDP Data
- County Population Characteristics
- MLB Pitching and Batting Data for 2021 and 2022
- MLB Team Level Data for 1996-2022
- Texas 8th Grade STARR Test Performance Data
- County Health Insurance Data
- County Presidential Voting Data

The dataset and data definition documents are available on Term Project page at the class website: http://people.tamu.edu/~cschulman/TermPr461.html

Given your choice of datasets, criterion variable, predictor variable, and sub-sample grouping strategy, your assignment is to provide a detailed statistical analysis, *presented in a formal written report* that should include most of the analytical tools that we cover during the semester.

Important Dates:
- **February 23**: Proposal due in **Canvas**. This should include sections 1 and 2 in the outline below.
- ~~**March 5**~~: **April 5:** Initial version of the paper due in **Canvas**. This should incorporate feedback on your proposal and include sections 1-4 in the outline below.
- **May 3**: Final version of the paper due in **Canvas**. Final submission incorporating feedback on your proposal and initial version of the paper.

Term Project Report/Analysis Outline (NOTE: While presented in outline format below, your Proposal, Initial Partial draft, and Final Report should be presented in formal report format.)

1) Introduction
   a) Description of the criterion and predictor variables to be analyzed.
      i) What is the economic significance of the criterion variable?
      ii) What is the economic basis to expect a particular type of association (positive, negative, no correlation) between the criterion and predictor variables?
   b) Sub-group identification
      i) Description of the chosen method for identifying sub-group samples and why it is economically meaningful.
      ii) What is the economic basis for whether to expect differences in the mean and/or variance of the criterion variable among sub-groups?
      iii) What is the economic basis for whether to expect differences in the degree of association (correlations) between the criterion and predictor variables among sub-groups?
   c) Statement of formal hypotheses of differences among sub-groups

2) Literature Review
   a) Discuss previous analyses of the chosen criterion variable.

3) Descriptive/Graphical Analysis
   a) Provide basic descriptive statistics (e.g., central tendency, variability, and correlation statistics) for the overall sample and for sub-group samples in both tabular and graphical formats.
   b) Provide frequency distributions of the criterion variable for the overall sample and for sub-group samples in both tabular and graphical formats.
   c) Provide Box & Whisker Plots of the overall sample and for sub-group samples.
   d) Discussion of observed differences in descriptive statistics, frequency distributions and Box & Whisker Plots.

4) Single Sample Confidence Intervals and Hypothesis Tests
   a) For the overall sample and sub-group samples, construct confidence intervals for the student's chosen criterion variable. These results should be presented in graphical and tabular form, at various levels of significance (e.g., 90%, 95%, 99%). Confidence intervals should be constructed for:
      i) the sample means
      ii) the sample variances
   b) Confidence interval results should be presented in graphical form. Discussion of confidence interval results.
   c) Taking the overall sample mean and overall sample variance of the criterion variable as "fixed" values, conduct single sample hypothesis tests comparing:
      i) The sub-group sample means to the "fixed" overall mean.
      ii) The sub-group sample variances to the "fixed" overall variance.
   d) Hypothesis test results should be presented in tabular form and discussed.

5) Two Sample Confidence Intervals and Hypothesis Tests
   a) Present pair-wise hypothesis tests of equal variances among sub-groups. Note that if the grouping method you are using results in 3 sub-groups this will involve 3 pair-wise tests, with 4 sub-groups 6 pair-wise tests, and with 5 sub-groups 10 pair-wise tests. Results should be presented in tabular form and discussed.
   b) Construct pair-wise confidence intervals and hypothesis tests of the differences between the sub-group sample means. Note that the pair-wise variance tests from above should dictate whether your means confidence intervals and tests are conducted under assumptions of equal or unequal variances. Results should be presented in tabular form and discussed.

6) ANOVA Tests
   a) Conduct a single-factor ANOVA test for joint equality among all the sub-group means.
   b) Discussion of ANOVA results.
7) Correlation Analysis
   a) Present X-Y scatter plots of the criterion ("Y") and predictor ("X") variables including "trend lines" for the overall sample and sub-group samples.
   b) Conduct tests of a significant correlation between the criterion and predictor variables for the overall and sub-group samples. Results should be presented in a tabular form and discussed.
   c) Conduct pair-wise tests for differences in the degree of correlation between the criterion and predictor variables among the sub-group samples. Results should be presented in a tabular form and discussed.
   d) Conduct a test of joint equality among all the sub-group correlations. Results should be presented in a tabular form and discussed.
8) Conclusion/Discussion of possible extensions and/or limitations

All your analyses and discussion should be presented in a formal written report complete with any citations to literature discussed in your literature review section. Any standard citation style is acceptable.

Note that this document represents the definitive statement of what is required for your term project.