

MA 214: Real Data Project

Dr. Neil Martinsen-Burrell

November 29, 2006

Summary: Write a report on a statistical analysis of one of two possible data sets

Length: As long as necessary (likely 3-5 pages)

Due Date: Friday, December 8th, 5:00 pm

1 Introduction

Working with real-world data is the most fundamental thing in the practice of statistics. This includes understanding the research questions to be answered, how the data were collected, summarizing the data, estimating parameters and making inferences where appropriate. In this project, you will do all of these things for one of two data sets described below.

You will write a report that should address the following topics

- What are the research questions that this data is attempting to address?
- How was this data collected? How does this affect the rest of your analysis?
- Summarize the data using the techniques of Chapters 2 and 3. This may include graphical displays such as bar graphs and histograms, and numerical measures such as means and variances. Summarize only the data that is relevant to the research questions that you described above.
- Estimate some of the parameters of the population. These may include confidence intervals for the mean, population proportion and standard deviation.
- Formulate statistical hypotheses that are relevant to the research questions and test those hypotheses

2 Expectations

Your report should be neatly typed. It should include all of the figures. It should have all of the features of a paper including an introduction and conclusion which summarize the contents of the report. You should *not* turn in your raw data, nor should you try to overwhelm me with the size of your report. All of the tables, figures, etc. that you include should be relevant to the topics above.

Your report is due to me **Friday, December 8th** by **5:00 pm**. You may submit it to me by email (nmb@wartburg.edu), in person, by carrier pigeon or any other means, but it must be in my possession by 5:00.

3 The Data

3.1 Transportation Needs Survey

The Department of Social Work conducted a survey of local residents and businesses in an effort to determine the interest in a local transportation service. They sent out questionnaires to four groups of people: the general population, elderly residents, developmentally disabled residents and businesses. More information about the purpose of the survey and the questions asked are available on the questionnaires themselves. The returned questionnaires were processed by the Social Work Research class and made available for our use.

Due to the large number of questions and types of data, I suggest the following questions of interest.

- Using the χ^2 goodness of fit test (section 12.1), determine whether or not the age distribution of respondents matches the age distribution of people in Bremer County. (See <http://quickfacts.census.gov/> for census data on age distribution.)
- Perform linear regression for the age of respondents and the number of vehicle trips per week. Is there any correlation?
- Using the grouped data for number of trips per week, estimate the mean number of trips per week for all residents of Bremer County. Using the tests of section 11.2, determine whether the elderly population make less trips than the general population.
- Using the grouped data on how much people are willing to pay for a form of public transportation, estimate the mean amount that people are willing to pay. Using section 11.2, compare those means for the different groups surveyed.
- Using the techniques of section 11.3, compare the proportion of people who are aware of the current bus service in Bremer County between the groups surveyed.

3.2 Historical Macroeconomic Data

The economy of the United States is an interesting and well-studied object. Economic indicators such as the Gross Domestic Product, Consumer Price Index and the Unemployment Rate are collected and tabulated by various governmental and non-governmental agencies. The data of interest here are the quarterly change in GDP (economic growth rate), quarterly change in price index (inflation) and the monthly unemployment rate since 1948.

Possible questions of interest include:

- Has there been a significant change in economic growth rates in the past 20 years compared to the previous 40 years?
- Has the rate of inflation been less volatile while Alan Greenspan was chairman of the Federal Reserve than under his predecessors?
- Is the current rate of unemployment significantly lower than its historical average?

and many others that may be answered using this data.