

# STAT 525 Integrative Experience (IE) Project Topics

Maryclare Griffin

Spring 2020

Here is a description of potential projects for the IE component of Stat 525.

1. Car values : How Much Is Your Car Worth? Data collected from Kelly Blue Book for several hundred 2005 used General Motors (GM) cars in order to develop a regression model to determine car values based on a variety of characteristics such as mileage, make, model, engine size, interior style, and cruise control. Please see the following files: car\_info.txt, car.xls
2. Home Prices Using Realtor Data The data file contains information on 76 single-family homes in Eugene, Oregon during 2005. At the time the data were collected, the data submitter was preparing to place his house on the market and it was important to come up with a reasonable asking price. Whereas realtors use experience and local knowledge to subjectively value a house based on its characteristics (size, amenities, location, etc.) and the prices of similar houses nearby, statistical analysis provides an alternative that more objectively models local house prices using these same data. Please see the following files: homeprice\_info.txt, homeprice.xls
3. Air pollution and traffic volume The data are a subsample of 500 observations from a data set that originate in a study where air pollution at a road is related to traffic volume and meteorological variables, collected by the Norwegian Public Roads Administration. The response variable consist of hourly values of the logarithm of the concentration of NO<sub>2</sub> (particles), measured at Alnabru in Oslo, Norway, between October 2001 and August 2003. The predictor variables are the logarithm of the number of cars per hour, temperature 2 meter above ground (degree C), wind speed (meters/second), the temperature difference between 25 and 2 meters above ground (degree C), wind direction (degrees between 0 and 360), hour of day and day number from October 1. 2001. For more details, please see the following files: NO2\_info.txt, NO2.xls
4. Wages and gender This uses data from 1985 current population survey to asses the impact of gender on wages after accounting for other variables. These data consist of a random sample of 534 persons from the CPS, with information on wages and other characteristics of the workers, including sex, number of years of education, years of work experience, occupational status, region of residence and union membership. For more details, please see the following files: cps85wages.txt, cpswages.dat
5. Pollution and mortality The data is from the U.S. Department of Labor Statistics on 60 Standard Metropolitan Statistical Areas (a standard Census Bureau designation of the region around a city) in the United States, collected from a variety of sources. The data include information on the social and economic conditions in these areas, on their climate, and some indices of air pollution potentials. The goal of this project is to examine how the age adjusted mortality rate potentially depends on the other (predictor) variables and in particular the effects of various pollutants. For more details, please see the following files: smsa\_info.txt, smsa.dat
6. Hospitals and infections. There have been increasing problems with individuals picking up infections during a hospital stay. This project will use the SENIC data from the book to examine the relationship between the probability of infection and hospital characteristics. For more details, please see the following files: senic\_info.txt, senic.dat
7. Voting patterns across counties in Florida for the 200 presidential election (and the potential outlying nature of Palm Beach county). This project will model the vote counts for candidates as a function of characteristics of the county (omitting Palm Beach) and then at the end examine whether the Palm Beach vote was outlying in some way. For more details, please see the following files: flovote\_info.txt, flovote.txt
8. Homeowner Insurance policies. This project will use data from a number of Chicago neighborhoods to examine how the voluntary issuance of insurance policies relates to race and other variables (fire rates, theft rates, income, and age). For more details, please see the following files: insurance\_info.txt, insurance.dat
9. Crime Rates Criminologists are interested in the effect of punishment regimes on crime rates. This has been studied using aggregate data on 47 states of the USA for 1960. Please see the following files: crime\_info.txt, crime.xls