

Statistics 207

PROJECT

Winter Quarter 2016

Please choose one of the projects that are listed. This project is due at the time of the final exam. You are encouraged to form a group of 3 registered students in the class. Only one report per group needs to be submitted and names of all the three group members must appear on the front page. It is important that the name of the project also appears on the front page of the report. Every student must be in only one group so that only one project per student is submitted.

The project should be typed and the main body of the project should be no longer than 4-5 pages. You may use an appendix for the graphs and charts as needed.

Guidelines:

The project should include the following along with justifications and explanations in each step:

1. A description of the data and the goal of the analysis.
2. Graphical methods employed to obtain initial understanding of the data.
3. The type of statistical methods and models used in the analysis.
4. Transformations of variables (if needed), interactions (if needed), nonlinear terms (such as quadratic, if needed) in the analysis.
5. Initial analysis that includes all the variables and all the summary statistics such as anova table (if appropriate), parameter estimates, their standard errors, p-values etc.
6. All the relevant diagnostics needed for the initial analysis (step 5).
7. Model selection.
8. The recommended final model along with the summary statistics for this model (as in part 5), and the relevant plots (as needed).
9. Summary of findings, conclusion and recommendations for further analysis (if any).

Notes on the projects:

Project 1: Requires ANOVA.

Project 2: Requires stepwise regression, ridge, lasso and partial least squares.

Project 1. Photosynthesis rate.

Description: Photosynthetic rates (CO₂ Uptakes) for 2 plant populations (Quebec and Mississippi), under 2 treatments (control and chilled), at 7 CO₂ concentrations (95, 175, 250, 350, 500, 675, 1000 $\mu\text{L/L}$.)

In the experiment, the plants have been chosen at random. The experimenters' goal was to investigate how photosynthetic rates (CO₂ uptake) depend on the population, treatment and CO₂ concentration.

Column 1: population (Quebec1=1, Mississippi=2),

Column 2: treatment (Control=1, Chilled=2),

Column 3: plant id.,

Column 4: CO₂ concentration,

Column 5: CO₂ uptake.

Project 2: Mortality.

Mean annual precipitation (in inches) [PRECIP],

Median number of school years completed by persons of age 25 or over [EDUC],

Percentage of population in 1960 that is nonwhite [NONWHITE],

Percentage of households with annual income under \$3000 in 1960 [POOR],

Relative pollution potential of oxides of nitrogen (NO_x) [NOX],

Relative pollution potential of sulphur dioxide (SO₂) [SO2],

Total age-adjusted mortality [MORT] from all causes, in deaths per 100,000 population.

[“Relative pollution potential” is the product of the tons emitted per day per square kilometer and a factor correcting for SMSA dimension and exposure.]

The goal of the analysis would be to relate mortality (the response) to all the variables. Thus mortality is the response variable. It would be also important to see if pollution is related to mortality. Transformations may be needed for some of the variables including the response.