**Problem Set 2**

On Moodle, there is an excel file titled, “Problem Set 2.xlsx”. This file contains two data sets: one for each of the two problems described below.

For each problem, write a 2-3 page summary of the analysis you performed to answer the specified question(s). Make sure to include any appropriate graphs and summary statistics. Describe how you decided which statistical model to use. Be sure to include a discussion on the assumptions necessary to apply the model and whether the data met those assumptions and whether any transformations are appropriate. Make sure you clearly state your answer to the specified question(s) and include appropriate calculations that quantify the uncertainty in your conclusions. Make sure you address issues of model fitting. Address any potential shortcomings you see in the data sets. Also make sure you address any appropriate inferences and assessments of causality.

1. The data file given in the tab labeled “Problem 1” contain the construction costs of 32 light water reactor (LWR) nuclear power plants constructed in the late 1960s and early 1970s. The data also include information on the construction of the plants and specific characteristics of each power plant. The research goal is to determine which of the explanatory variables are most strongly related to the capital cost of the plant. The following table provides explanations of the variables in the data file.

|  |  |
| --- | --- |
| C | Cost in dollars X 10-6, adjusted to 1976 base |
| D | Date construction permit issued (year.proportion of year) |
| T1 | Time between application for and issue of permit (months) |
| T2 | Time between issue of operating license and construction permit (months) |
| S | Power plant net capacity (MWe) |
| PR | Prior existence of an LWR on same site (=1) |
| NE | Plant constructed in northeast region of USA (=1) |
| CT | Use of cooling tower (=1) |
| BW | Nuclear steam supply system manufactured by Babcock-Wilcox (=1) |
| N | Cumulative number of power plants constructed by each architect-engineer |
| PT | Partial turnkey plant (=1) |

1. Grubs damage lawns by feeding on grass roots. A group of researchers wish to evaluate four new treatments for lawns. They decided to evaluate the chemicals on a variety of soils and terrains. Thus, plots of land at six different sites were selected for use in the experiment.

A convenient way to conduct the experiment would be to use the same chemical at all test sites at a given location. However, this would result in the confounding of the effectiveness of the chemical with the location of the test sites. Therefore, the following experimental protocol was implemented. Within each of the six plots, there were 12 test sites randomly assigned to the control. A week after applying the treatments to the test sites, the researchers returned to the test sites and counted the number of surviving grubs on each of the 72 test sites.

The researchers were interested in determining if the average numbers of grubs on the test sites receiving the four treatments were less than the average numbers on the control sites. Furthermore, they wanted to determine if there were differences in the four treatments relative to their average counts. The data collected during the experiment are contained in the excel file titled, “Problem Set 2.xlsx”.