

Assignment #3 – Data Analysis using Regression

Introduction:

Tent caterpillars are a native insect to North America. Populations of tent caterpillar fluctuate, with outbreaks every few years. During outbreaks, the foliage of broadleaved trees, such as aspen (*Populus tremuloides*) can be almost completely consumed, resulting in dramatic declines in tree growth, and even death. Your company has conducted an experiment to evaluate the effectiveness of aerial spraying of the biological insecticide *Bacillus thuringiensis* (Bt) to kill tent caterpillars in aspen stands in Michigan.

Early in the year, during a previous outbreak, ten different rates of application of Bt, plus a control (no application), were made to aspen stands. Application rates were 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10 kg active ingredient (a.i.) per hectare. Each application rate was applied to three stands, for a total of 33 stands. A week later, eight plots in each stand were sampled and the density of live caterpillars was measured (in number per m² land area). The cost of application was approximately \$140 per hectare for helicopter time, plus \$12 per kg a.i. for Bt.

Instructions:

You've been charged with analyzing the data, which can be downloaded from the class website. It is reasonable to expect that the insecticide reduces caterpillar density, but the shape of the reduction with increasing application rate is not known. It might be linear, and it might not. Follow these steps to complete this assignment:

- Generate a scatterplot of the response against the predictor variable. What course of action is suggested for your regression analysis? Summarize your answer to this question, briefly.
- Fit a simple linear regression model to the data, and evaluate the regression assumptions. Why is this model not an appropriate solution? What would be the consequences of using it anyway?
- Find the best model for these data. You might wish to consult the Catalog of Curves for Curve Fitting (link on class website). Find at least one other possible model, and compare these two to each other and to the simple linear regression. Evaluate regression assumptions, compare descriptive fit statistics, and consider issues that might come up if someone tries to apply your best model to predict the response during future outbreaks.
- What does the final model tell you about the data? Draw as many conclusions as you can.

Product:

Prepare short written report in memo format. Present your **analysis** and include **relevant** detail.

Due Date:

Friday, March 04 2005. I am leaving town at noon on March 4th and I will not be back until after spring break.