

STAT 696, Spring 2011  
Homework 1 Problems  
due Thursday Feb. 3

2 Problems. Please follow the Lab report directions off the homework web page.

1. **Soil pH data.** Data on soil pH comes from an observational study in which samples of soil were collected at the nodes of a grid with 11 rows  $\times$  11 columns. The pH of the samples were measured in a laboratory.

The data is available off the class web page:

<http://www.rohan.sdsu.edu/~babailey/stat700/soilph.dat>

Use the R `read.table` command with the `header=T` option. (You do not need to make your own labels!)

We will reproduce some of the EDA plots from Lab1.

(a) Use the code in `intro1.r` to produce a plot and summary statistics for the data. In addition, include a histogram of the pH values. What is range of the data? Does the soil pH data appear to be normally distributed? Explain.

(b) Use the code in `intro2.r` to produce a contour and image plot of the data. Do the soil pH values appear to be constant over the region? Explain.

(c) Use the code in `intro4.r` to produce the mean and median of the soil pH over the rows and columns. Describe any patterns or trends that you observe.

2. **Soil pH data (cont.)** We will use the R `lm` function to fit a linear trend surface to the data.

(a) Fit the following linear trend to the data:  $m(x, y) = \beta_0 + \beta_1 x + \beta_2 y$ .

Include summary and diagnostic plots of the model fit from `lm`. Do you detect any pattern in the residuals?

(b) After examining the plots from Problem 1, suggest an alternative model for the trend. Write down the form of the model. Use the R `lm` function to fit the model. Include summary and diagnostic plots of the model fit from `lm`. Do you detect any pattern in the residuals?

(c) Which of the models in (a) and (b) best describes the trend? To answer this question use a model selection criterion such as AIC. There is a R function `AIC`. The model with the smallest AIC is “best” based on AIC.