LAB 3 – MATH 463

Part I.

The purpose of this part is to fit a multiple regression model to the data in

`http://www.uoregon.edu/ dlevin/DATA/sat.csv`

The variables here are state per pupil spending on K-12, the median salary of teachers, the percent of seniors taking the SAT, and the verbal, math, and total average scores for students taking the test.

Plot total score again expenditure. What does the plot suggest? Fit a simple linear regression model, and give a 95% confidence interval for estimated slope of the simple linear regression line. What can you conclude?

Now create a multiple regression model including takers as a dependent variable. (You could include other variables as well.) Give a 95% confidence interval for the estimated coefficient of the expenditure. What can you conclude?

Do you think the model

$$Y_i = a + bx_i + cz_i + \varepsilon_i \quad [\varepsilon_i \text{ i.i.d.}]$$

where Y_i is the average total score in the state, x_i is takers, and z_i is expenditure is a reasonable model for how the data was generated?

Part II.

The ANOVA set-up is a special case of multiple linear regression. We will look at this relationship using the data `corn.txt`.

Fit a linear model as you would do for a one-way ANOVA (yield is the response, and there are 5 varieties).

For example,

```
cornlm = lm(yield~as.factor(variety, data=corn)
```

To get the ANOVA table, use `anova(cornlm)`

Note that in fact `lm` fits a multiple regression model. Use `model.matrix(cornlm)` to see the different values in the regressors, i.e. to obtain the matrix A appearing in the model $Y = AX + \varepsilon$ which `lm` is assuming.

THe columns of this matrix are the variables entering the multiple linear regression.

Note these variables take on only the values $\{0, 1\}$.

In fact, these five variables are all determined by the single variable `variety`. Give a formula defining these variables in terms of the variable `variety` in the corn data frame.

What is the meaning of the coeeficients of these variables in terms of the μ_i appearing in the analysis of variance model

$$Y_{i,j} = \mu_i + \varepsilon_i \, .$$