Assignment #2 — Due Friday, February 8, 2007, by 4:00 P.M.

Turn in homework in lecture, discussion, or your TA’s mailbox (just inside the main entrance to MSC). Please circle the discussion section you expect to attend to pick up this assignment.

311: Tues. 1:00–2:15  312: Wed. 2:30–3:45  313: Tue. 4:00–5:15

1. The growth of the mold Aspergillus niger is inhibited by the toxicant methyl 4-hydroxybenzoate. Various concentrations of the toxicant are available. The inhibitions are measured as growth rate changes (in mm) per 24 hours.

<table>
<thead>
<tr>
<th>Concentration (mg/L)</th>
<th>0.32</th>
<th>0.40</th>
<th>0.51</th>
<th>0.64</th>
<th>0.80</th>
<th>1.0</th>
<th>1.3</th>
<th>1.6</th>
<th>2.0</th>
<th>2.5</th>
<th>3.2</th>
<th>4.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inhibition (mm)</td>
<td>2.0</td>
<td>2.3</td>
<td>2.4</td>
<td>2.8</td>
<td>3.0</td>
<td>3.3</td>
<td>3.6</td>
<td>4.3</td>
<td>4.8</td>
<td>5.1</td>
<td>5.6</td>
<td>6.6</td>
</tr>
</tbody>
</table>

(a) Create a file with this data an read it into R. Use xyplot() from the lattice package to plot inhibition versus concentration, including a regression line with the plot.

Recall that you need to enter the command `library(lattice)` once per R session to load lattice before using xyplot().

Recall that the syntax for using `xyplot()` is as follows.

```r
> xyplot( y-variable ~ x-variable, data=name-of-data.frame)
```

To plot points and a regression line, add the argument `type=c("p","r")` to the command.

Does it appear to be reasonable to use a simple linear regression model to predict inhibition from concentration? Explain why or why not.

(b) Use `lm()` to fit the regression model. What are the slope and intercept of the fitted line?

(c) Use the model you fit in the last part to predict inhibition under similar experimental conditions when the concentration is 1.9 mg/L.

(d) Based on the scatterplot you made in part (a), do you expect the prediction to be pretty accurate or biased either high or low? Explain.

2. Read in the cow data set from the first assignment and answer these questions. Information about the variables is in the file `cow.info`. Milk yields are in pounds per day and initial weights are measured in pounds.

(a) Fit a simple linear model to predict milk yield using the initial weight of the cow as a predictor. Write the fitted model in the form

(Yield in ??) = ?? + ?? (Weight in ??)

In what units are the slope and intercept?

(b) What are the estimated standard errors for the slope and intercept?

(c) Is it reasonable to interpret the numerical value of the intercept? Explain.

(d) What is a typical distance for the actual yield to differ from the predicted yield?

*Note that there is an additional problem on the next page.*
3. Here is a small data set from an elementary student science project in which the student grew *Wisconsin Fast Plants* from seed for 21 days under several experimental conditions. Heights of plants are measured in cm, fertilizer amount is measured in ml, and sand is the percentage of sand in the soil mixture (the remaining component is commercial potting soil).

<table>
<thead>
<tr>
<th>height</th>
<th>fertilizer</th>
<th>sand</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10.0</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>18.5</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>12.5</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>19.5</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>12.0</td>
<td>10</td>
<td>20</td>
</tr>
</tbody>
</table>

Following the presentation in the lecture on the mathematical background for multiple regression, construct the model matrix $X$ for a multiple regression model to predict height from fertilizer and sand amounts if:

(a) both predictors are treated as quantitative;
(b) fertilizer is treated as quantitative, but sand is treated as a factor with levels *control* and *sandy*;
(c) fertilizer is treated as a factor with levels *control*, *low*, and *high* and sand is treated as a factor with levels *control* and *sandy*.

Work to do, but not turn in.

- Read Chapters 3–4 of the textbook.