Take Home Exam 2 Example

MATH 441

For Take Home Exam 2, consider two examples from lab #2, problems #2 and 3 (see the course schedule on the web site for October 16):

- *Example 1*: Classification of sex of child based on ultrasound at 20 weeks of pregnancy.
- *Example 2*: Relationship between mean global temperature, year, carbon and solar magnetic cycle length.

Note that not all parts of the exam are answered for these examples.

1. **Make a hypothesis**.
   - *Example 1*: I guessed that most people had their babies correctly identified by ultrasound at 20 weeks.
   - *Example 2*: Many climatologists and politicians today claim that increased carbon in the atmosphere results in higher mean global temperatures.

2. **Collect Data**.
   - *Example 1 is categorical*: I asked 17 mothers whether the sex of their child was correctly classified at 20 weeks, and I encoded 0's for no respones and 1's for yes responses.
   - *Example 2 is quantitative*: There is one quantitative response (**temp**) and three different explanatory quantitative variables (**time, carbon, solar magnetic cycle length**.) You can choose to have as many explanatory variables as you wish (or only one, such as when you are only fitting a line):
     - Temperature measurements from 1890-2000 are published by NASA's Goddard Institute from Space Studies in Figure 1 at http://data.giss.nasa.gov/gistemp/2005/.
     - Carbon measurements from 1950 to present from the well known "Keeling Curve" as reported in the movie *Inconvenient Truth*. For carbon measurements from 1890-1990, see Figures 1 and 12 in Robinson's "Environmental Effects of Increased Atmospheric Carbon Dioxide" paper at http://www.oism.org/pproject/review.pdf.
     - Solar magnetic cycle length measurements from 1890-1980 were published by in a 1991 *Science* paper by Friis-Christensen, E. and K. Lassen, "Length of the solar cycle: an indicator of solar activity closely associated with climate" (254, p698-700). See the first figure at http://www.tmgnow.com/repository/solar/lassen1.html.

3. **Analyze the data**.
   - *Example 1*: The model was $y = a$, we just fit a horizontal line to the 0 and 1 responses. The least squares solution is $z_{ls} = a_{ls} = .8235$. Thus about 82% of the people in the sample had correct classifications.

- *Example 2*: The model was $y = a + bx_{year} + cx_{carbon} + +dx_{solar}$, a hyper-plane. Note that $y = a + bx_{carbon}$ is a familiar line fit. The least squares solution is

$$z_{ls} = [a_{ls} \;\; b_{ls} \;\; c_{ls} \;\; d_{ls}]^T = [-2.4865 \;\; 0.0012 \;\; 0.0034 \;\; -0.0922]^T.$$

  The value $a_{ls} = -2.4865$ is not interpretable since it corresponds to all variables equal to zero, and $x_{carbon} = 0$ and $x_{solar} = 0$ are impossibilities. Other interpretations: the model predicts that global mean temperature increases by $b_{ls} = 0.0012$ degrees celsius each year. Global mean temperature increases by $c_{ls} = 0.0034$ degrees celsius for each ppm increase in carbon in the atmosphere. And $d_{ls} = -0.0922$ implies that global mean temperature decreases by .0922 degrees celsius for each year increase in solar magnetic cycle length.

4. **Test your hypotheses using the data**.

- *Example 1*: The matrix A is 17×1 column vector of ones, $n = 17$ and $p = 1$. It is easy to verify that $||r||_2 = 1.5718$, $MSE = \frac{2.4706}{16} = .1544$, and $Var_{z_{ls}} = 0.0091$ and so $SE_{z_{ls}} = 0.0953$.
- *Example 1*: The matrix A is 10×4

```
1.0e+003 *

  0.0010    1.8900    0.2910    0.0117
  0.0010    1.9000    0.2920    0.0114
  0.0010    1.9100    0.2940    0.0111
  0.0010    1.9200    0.2970    0.0107
  0.0010    1.9300    0.3010    0.0103
  0.0010    1.9400    0.3050    0.0102
  0.0010    1.9500    0.3110    0.0105
  0.0010    1.9600    0.3180    0.0107
  0.0010    1.9700    0.3270    0.0106
  0.0010    1.9800    0.3380    0.0104
```

  where the first column is all ones, the second is year, the third is composed of carbon measurements, and the last is of solar measurements. So $n = 17$ and $p = 4$. It is easy to verify that $||r||_2 = .2115$, $MSE = \frac{.0447}{6} = .0075$, and $Var_{z_{ls}}$ is the matrix

```
116.4860   -0.0689    0.1012   -1.3213
 -0.0689    0.0000   -0.0001    0.0007
  0.1012   -0.0001    0.0001   -0.0010
 -1.3213    0.0007   -0.0010    0.0185
```

  Thus,

$$SE_{z_{ls}} = [10.7929 \;\; 0.0064 \;\; 0.0099 \;\; 0.1360]^T.$$