Assignment 6 STP 526, Fall 2009, S. Lohr
Due on Monday, 9 November 2009

1. Read Sections 7.1, 7.2, and 7.3.1 of the text.

2. Use the famous stackloss data for this exercise. You may access the data by typing `data(stackloss)` in R and find a description by typing `help(stackloss)`. The variable `stack.x` contains the **X** matrix, and `stack.loss` contains the **y** vector. Answer the questions in this problem by finding the appropriate matrices and using the theoretical results we derived in class. You may want to check your answers by using the function lm, but I want you to do the calculations using matrices.

   (a) Using R, fit the model $E[Y] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$ where x_1 = 'Air Flow', x_2 = 'Water Temp', and x_3 = 'Acid Conc.'

   (b) Using the formulation in Result 7.2.1, perform a test of the hypothesis

   $$H_0 : \beta_1 = 1, \beta_3 = 0,$$

   Your work for this problem should include the matrices $\mathbf{X}$, $\mathbf{X}'\mathbf{X}$, $\mathbf{C}$, and other matrices relevant for calculating the F statistic. Also give your F statistic and p-value. Show your work in R, and highlight the relevant output so that IT IS EASY FOR YOUR INSTRUCTOR TO READ. Do NOT just hand in pages of unannotated output.

   (c) Find a joint 95% confidence region for β_1 and β_2.

3. Two-way unbalanced ANOVA. Use the data below, and do all your work in R.

Column	name	description
1	amino	type of amino acid
2	diet	protein (diets)
3	gain	average daily weight gain

```
1 1 0.83
1 1 1.02
1 1 0.81
2 1 1.12
1 2 0.82
1 2 1.22
2 2 1.29
2 2 1.38
2 2 1.32
1 3 1.45
1 3 1.30
2 3 1.33
```

Consider the model

$$Y_{ijk} = \mu + \tau_i + \beta_j + \epsilon_{ijk}, \tag{1}$$

where τ refers to amino and β refers to diet. The response is gain.

(a) Write the model in the form $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon$. Write down $\mathbf{X}$ and $\boldsymbol{\beta}$ explicitly. Find $\mathbf{P}_X$.

(b) Now consider the reduced model

$$Y_{ijk} = \mu + \tau_i + \epsilon_{ijk}. \tag{2}$$

What is $\mathbf{X}_1$ for this model, if written as $\mathbf{Y} = \mathbf{X}_1\gamma + \varepsilon$.

(c) Find $\mathbf{P}_1$, the projection matrix onto $C(\mathbf{X}_1)$.

(d) State the null hypothesis corresponding to the reduced model. Form the F statistic using (SSE_0 - SSE), obtaining the SSE's by using the projection matrices. What is your p-value?

(e) Write the null hypothesis from question 3d in terms of a linear parametric function

$$H_0 : \mathbf{C}^T\boldsymbol{\beta} = 0.$$

Give $\mathbf{C}$ explicitly. Use Result 7.2.1 to carry out the test and find the p-value.

(f) Suppose that $\sigma = 0.2$ and we wish to use a balanced design (the same number of observations in each cell). How many observations should be taken in each cell if it is desired to have power 0.85 for the test in question 3e against the alternative hypothesis with $\beta_1 = 0.1$, $\beta_2 = -0.2$, and $\beta_3 = 0.1$?

(g) Suppose diet 2 is a blend of equal parts diet 1 and diet 3. The investigator would like to know whether the blend does better or worse than the individual diets 1 and 3. Test

$$H_0 : \beta_2 = \frac{\beta_1 + \beta_3}{2}$$

versus a two-sided alternative. As above, give $\mathbf{C}$ explicitly and give the p-value.

(h) What is the power of your test in question 3g against the alternative that $\beta_2 - \frac{\beta_1 + \beta_3}{2} = 0.2$?

(i) Fill in the values in the following ANOVA table, using your matrices. Verify that the first four SS's are independent.

Source	df	SS
Mean		$\mathbf{Y}'(\frac{1}{n}\mathbf{J}_n)\mathbf{Y}$
Amino		$\mathbf{Y}'(\mathbf{P}_1 - \frac{1}{n}\mathbf{J}_n)\mathbf{Y}$
Diet|Amino		$\mathbf{Y}'(\mathbf{P}_X - \mathbf{P}_1)\mathbf{Y}$
Residual		$\mathbf{Y}'(\mathbf{I} - \mathbf{P}_X)\mathbf{Y}$
Total		$\mathbf{Y}'\mathbf{Y}$

(j) Consider an alternative reduced model

$$Y_{ijk} = \mu + \beta_j + e_{ijk}. \tag{3}$$

Let $\mathbf{X}_2$ be the reduced design matrix, and let $\mathbf{P}_2$ be the projection matrix for model (3). Find $\mathbf{X}_2$ and $\mathbf{P}_2$, and write the ANOVA table (theoretically, in terms of the matrices $\mathbf{J}_n, \mathbf{P}_X$, and $\mathbf{P}_2$) with Mean, Diet, Amino|Diet, and Residual.

(k) Fill in the numbers for the ANOVA table in question 3j. Are the sums of squares the same as in question 3i?