STAT 510 **Exam 1** **Spring 2014**

Instructions: The is a closed-notes, closed-book exam. No calculator or electronic device of any kind may be used. Use nothing but a pen or pencil and blank paper on which to write answers. For questions that require extensive numerical calculations that cannot be done easily without a calculator, simply set up the calculation and leave it at that. For example, $(3.45 - 1.67)/\sqrt{2.34}$ would be an acceptable answer.

1. Researchers were interested in studying the effects of diet and voluntary exercise on the health of mice. Suppose two diets (labeled D_1 and D_2) were assigned to 8 mice using a completely randomized design with 4 mice for each diet. The 8 mice were randomly assigned to 8 individual cages. Within each diet group, 2 of the 4 mice were randomly assigned to cages without running wheels, and the other 2 mice were assigned to cages with running wheels. The running wheels permit mice to exercise, while exercise is very limited in cages without running wheels. After two months of life on the assigned diet in the assigned cage, a measure of overall health was recorded for each mouse. Let y_{ijk} denote the measure of health for the kth mouse on diet D_i and exercise treatment j, where $j = 1$ indicates no running wheel access and $j = 2$ indicates running wheel access.

   Suppose the model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$ is appropriate for these data, where

$$\boldsymbol{y} = \begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \end{bmatrix}, \ \boldsymbol{X} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix}, \ \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}, \text{ and } \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \end{bmatrix}.$$

   In this model, β_1, β_2, and β_3 are unknown real-valued parameters, and σ^2 is some unknown real-valued positive variance component.

   (a) According to the model, what is the mean response of a mouse who received diet 2 and was housed in a cage without a running wheel?

   (b) Determine a solution to the Normal Equations and express your solution in terms of the y_{ijk} data. Simplify your solution as much as possible according to the notational conventions we have used in course notes.

   (c) Find the Best Linear Unbiased Estimator (BLUE) of the mean response of a mouse who received diet 2 and was housed in a cage without a running wheel. Simplify your answer as much as possible.

   (d) In terms of the β_1, β_2, and β_3 parameters, write the null hypothesis of no diet main effect. Simplify your answer as much as possible.

   (e) Find an expression for the LSMEAN of diet D_2. Simplify your answer as much as possible.

2. Consider a two-factor experiment with a completely randomized design. Suppose factor A has 2 levels ($A1$, $A2$) and factor B has 3 levels ($B1$, $B2$, $B3$). Suppose the observed responses are provided in the table below.

Factor A	Factor B: $B1$	B_2	B_3
$A1$	7 9	8 10	12
$A2$	2	0 2	0 4

The observed responses in the table have been presented as integers to make calculations easier. However, please assume that the cell means model we have discussed in class with i.i.d. $N(0, \sigma^2)$ error terms is appropriate for these data. Use the data in the table and/or the R code and output provided below to answer the following questions.

```
> y=c(7,9,8,10,12,2,0,2,0,4)
> A=factor(c(1,1,1,1,1,2,2,2,2,2))
> B=factor(c(1,1,2,2,3,1,2,2,3,3))
> o=lm(y~A+B+A:B)
> summary(o)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    8.000      1.323   6.047  0.00377 **
A2            -6.000      2.291  -2.619  0.05888 .
B2             1.000      1.871   0.535  0.62131
B3             4.000      2.291   1.746  0.15579
A2:B2         -2.000      2.958  -0.676  0.53604
A2:B3         -4.000      3.240  -1.234  0.28461

> anova(o)
Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value   Pr(>F)
A          1 144.400 144.400 41.2571 0.00302 **
B          2   6.667   3.333  0.9524 0.45890
A:B        2   5.333   2.667  0.7619 0.52438
Residuals  4  14.000   3.500
```

(a) Compute LSMEANS for the levels of factor B.

(b) Compute an F statistic for testing the null hypothesis of no factor A main effect.

(c) Compute an F statistic for testing the null hypothesis of no interactions between the factors A and B.

3. In 1981, a group of 2,584 students took four tests covering arithmetic reasoning, mathematics knowledge, paragraph comprehension, and word knowledge, respectively. The quantitative scores on the four tests were recorded in variables x_1 for arithmetic reasoning, x_2 for mathematics knowledge, x_3 for paragraph comprehension, and x_4 for word knowledge. In 2006, these same students were asked to report their 2005 incomes. The cube roots of these incomes were recorded in the variable y. Let subscript i index individual students so that $x_{1i}, x_{2i}, x_{3i}, x_{4i}$, and y_i denote the values of the variables x_1, x_2, x_3, x_4, and y for the ith student. Suppose, for $i = 1, \ldots, n$,

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \epsilon_i, \tag{1}$$

where $\epsilon_1, \ldots, \epsilon_n \overset{i.i.d.}{\sim} N(0, \sigma^2)$. Use the code and partial output below to answer the following questions.

```
> o=lm(y~x1+x2+x3+x4)
> anova(o)
Analysis of Variance Table

Response: y
            Df Sum Sq Mean Sq
x1           1  31544   31544
x2           1   2033    2033
x3           1    172     172
x4           1     14      14
Residuals 2579 214109      83
```

(a) Explain what the sum of squares for x_1 (31544) represents.

(b) Provide an estimate of σ^2 based on the fit of model (1).

(c) The sum of squares for x_4 in the ANOVA table above has a scaled non-central chi-square distribution with 1 degree of freedom. Provide an expression for the non-centrality parameter of this non-central chi-square distribution. Simplify your answer as much as possible.

(d) Compute an F-statistic for testing the null hypothesis $H_0 : \beta_3 = \beta_4 = 0$.

(e) The `R` commands

```
> o=lm(y~x1+x2)
> anova(o)
```

will produce and ANOVA table with an F-statistic for the `x2` term. Compute the value of that F-statistic.