STAT 510 **Exam 1** **Spring 2017**

Instructions: This 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. Please write your name and answers on blank paper. Please do NOT write your answers on the pages with the questions. For questions that require extensive numerical calculations that you should not be expected to do 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. On the other hand, some quantities that are very difficult to compute one way may be relatively easy to compute another way. Part of this exam tests your ability to figure out the easiest way to compute things, based on the information provided and the relationships between various quantities. If you find yourself trying to do exceedingly complex or tedious calculations, there is probably a better way to solve the problem.

1. Although some details have been altered for the purpose of this exam, this question is based on an actual experiment described in the paper: Stone, J., Lynch, C. I., Sjomeling, M., and Darley, J. M. (1999). Stereotype threat effects on Black and White athletic performance. *Journal of Personality and Social Psychology*, **77**(6), 1213–1227.

   A total of 80 college students (40 Black and 40 White) participated in the experiment. Each student was asked to complete a golf task that involved hitting a golf ball into a series of holes with a club. Each student received a score based on how many strokes (swings of the club) it took to complete the task. Low scores indicate better performance on the task. Prior to completing the task, the 40 students in each race category were assigned to four groups using a balanced and completely randomized design. Thus, 10 Black students and 10 White students were assigned to each group. The groups for this experiment are defined as follows:

   Group 1 : Participants in group 1 were told prior to completing the golf task that the task provided a measure of *natural athletic ability*, defined as "one's natural ability to perform complex tasks that require hand-eye coordination, such as shooting, throwing, or hitting a ball or other moving objects."

   Group 2 : Participants in group 2 were told prior to completing the golf task that the task provided a measure of *sports intelligence*, defined as "personal factors correlated with the ability to think strategically during an athletic performance."

   Group 3 : Participants in group 3 were told prior to completing the golf task that the purpose of the task was to measure *general sports performance*. Participants in group 3 were also required to complete a questionnaire before completing the golf task. The first question on the questionnaire asked the participant to indicate his or her race.

   Group 4 : Participants in group 4 were treated exactly as those in group 3 except that the questionnaire was administered after the golf task was performed.

   The average score of the 10 participants for each combination of race and group is provided in the following table.

		Group			
		1	**2**	**3**	**4**
Race	**Black**	23.0	27.2	27.3	22.1
	White	27.8	23.2	22.9	24.5

The researchers fit a Gauss-Markov model with normal errors to the scores from the golf task used in this experiment. Their model allowed for eight unrestricted means, with one unrestricted mean for each of the eight combinations of the factors race and group. This is an example of the "cell means model" we discussed in class. When answering the following questions, please assume this model is appropriate as a full model for the scores obtained from the golf task.

For $j = 1, 2, 3, 4$, let $\mu_{B,j}$ be the expected value of the score response variable for the Black race category and group j. For $j = 1, 2, 3, 4$, let $\mu_{W,j}$ be the expected value of the score response variable for the White race category and group j. Let σ^2 be the error variance. Use the following information and all other information above to complete parts (a) through (e).

- The mean squared error (MSE) from the fit of the full model is 18.3.
- The F statistic for testing the null hypothesis

  $$H_0 : \mu_{B,1} = \mu_{B,2} = \mu_{B,3} = \mu_{B,4} = \mu_{W,1} = \mu_{W,2} = \mu_{W,3} = \mu_{W,4}$$

  vs. the alternative that not all eight means are equal is $F = 2.95$.

(a) Construct a 95% confidence interval for the simple effect of race in group 4. (To fully specify the interval, you need a quantile from a t-distribution. Assume the relevant quantile is 1.99.)

(b) Compute a t statistic that can be used to test whether the simple effect of race for group 1 is the same as the simple effect of race for group 2.

(c) Give the noncentrality parameter for the t distribution of the t statistic computed in part (b). Your answer should be simplified and expressed in terms of model parameters.

(d) Compute the F statistic corresponding to the test for group main effects.

(e) Provide an ANOVA table with columns labeled *Source*, *Degrees of Freedom*, and *Sum of Squares*. Under the *Source* column, include lines with the following labels: *Race*, *Group*, *Race* $\times$ *Group*, *Error*, and *C. Total*. Determine the degrees of freedom and sum of squares values for each line of the table. Use any information provided or determined above to help with calculations.

2. Researchers were studying the effect of three treatments on the level of a chemical in the blood of mice. A total of nine mice were randomly assigned to the three treatments using a completely randomized design with three mice per treatment. Researchers also measured the pre-treatment weight of each mouse in the experiment. The researchers suspected that pre-treatment weight might be associated with the level of the chemical of interest, and they noticed the mice in their experiment varied considerably in pre-treatment weight. For $i = 1, 2, 3$ and $j = 1, 2, 3$, let y_{ij} be the measurement of the chemical level in the blood of the jth mouse that received treatment i, and let x_{ij} be the pre-treatment weight of the jth mouse that received treatment i. Study the following R code and output on page 3 and use it to answer parts (a) through (e) on the top of page 4.

```
> #trt is a vector of treatment labels.
> #x is a vector of pre-treatment mouse weights in grams.
> #y is vector of measurements of the chemical level in mouse blood samples.
> #There is one row in the dataset for each mouse.
>
> data.frame(trt, x, y)
  trt  x   y
1   1 10 4.9
2   1 21 3.4
3   1 30 2.8
4   2 20 6.0
5   2 29 5.3
6   2 42 3.6
7   3 31 6.9
8   3 41 6.3
9   3 49 5.5
>
> is.factor(trt)
[1] TRUE
>
> anova(lm(y~trt))
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value Pr(>F)
trt        2   9.63    4.81  4.5314 0.0632 .
Residuals  6   6.37    1.06
>
> anova(lm(y~x+trt))
Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value    Pr(>F)
x          1    0.19    0.19   3.453    0.1223
trt        2   15.54    7.77 142.124 3.929e-05 ***
Residuals  5    0.27    0.05

> anova(lm(y~x+trt+x:trt))
Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq  F value    Pr(>F)
x          1    0.19    0.19   3.7146 0.1495590
trt        2   15.54    7.77 152.8917 0.0009576 ***
x:trt      2    0.12    0.06   1.1894 0.4165346
Residuals  3    0.15    0.05
```

(a) For $i = 1, 2, 3$ and $j = 1, 2, 3$, consider the model

$$y_{ij} = \beta_1 + \beta_2 x_{ij} + \epsilon_{ij}, \tag{1}$$

where β_1 and β_2 are unknown parameters and the ϵ_{ij} terms are independent normal random variables with mean 0 and some unknown variance σ^2. Provide an unbiased estimator of σ^2 based on the fit of this model.

(b) Compute the F statistic for testing $H_0 : \beta_2 = 0$ in model (1).

(c) For $i = 1, 2, 3$ and $j = 1, 2, 3$, consider the model

$$y_{ij} = \alpha_1 + \alpha_2 x_{ij} + \delta_i + \varepsilon_{ij}, \tag{2}$$

where α_1, α_2, δ_1, δ_2, and δ_3 are unknown parameters and the ε_{ij} terms are independent normal random variables with mean 0 and some unknown variance ν^2. Provide the F statistic for testing $H_0 : \delta_1 = \delta_2 = \delta_3$ in model (2).

(d) Compute the t statistic for testing $H_0 : \alpha_2 = 0$ in model (2). Choose the sign of the t statistic to match the sign of the ordinary least squares estimate of α_2 in model (2).

(e) Based on the code, output, and conclusions that can be drawn from the code and output, write a brief report for the researchers that explains the results of the statistical analysis of their experiment. A few well-chosen sentences should be sufficient.

3. Consider matrices defined as follows:

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 3 & 9 \end{bmatrix} \quad \text{and} \quad \boldsymbol{B} = \begin{bmatrix} 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & 2 \\ 1 & 0 & 0 & 3 \\ 1 & 0 & 0 & 4 \\ 1 & 1 & 1 & 5 \\ 1 & 1 & 1 & 6 \end{bmatrix}.$$

Find the matrix product $\boldsymbol{P_A P_B}$, where $\boldsymbol{P_A}$ and $\boldsymbol{P_B}$ are the orthogonal projection matrices for projecting onto the column spaces of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively.