STAT 510 **Exam 1** **Spring 2020**

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 the answer sheets provided. 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. Show that any two matrices $\boldsymbol{W}$ and $\boldsymbol{X}$ have the same column space if there exist matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ such that $\boldsymbol{W}\boldsymbol{A} = \boldsymbol{X}$ and $\boldsymbol{X}\boldsymbol{B} = \boldsymbol{W}$.

2. Consider a field partitioned into 10 plots of land. Two varieties of corn (labeled 1 and 2) were randomly assigned to the 10 plots using a balanced and completely randomized design. Seed of the assigned variety was planted in each plot. Fertilizer amounts of $x_1 = 0$, $x_2 = 30$, $x_3 = 60$, $x_4 = 90$, and $x_5 = 120$ pounds of nitrogen per acre were randomly assigned to the 5 plots planted with variety 1. Likewise, fertilizer amounts of $x_1 = 0$, $x_2 = 30$, $x_3 = 60$, $x_4 = 90$, and $x_5 = 120$ pounds of nitrogen per acre were randomly assigned to the 5 plots planted with variety 2. Let y_{ij} be the amount (in bushels per acre) of corn harvested at the end of the growing season from the plot planted with variety i and treated with x_j units of fertilizer ($i = 1, 2; j = 1, 2, 3, 4, 5$). Consider the following model:
$$y_{ij} = \mu + \alpha_i + \beta x_j + \varepsilon_{ij}, \tag{1}$$
where μ, α_1, α_2, and β are unknown parameters and the ε_{ij} terms are independent and identically distributed as $N(0, \sigma^2)$ for some unknown variance parameter $\sigma^2 > 0$.

   (a) Let $\boldsymbol{\beta} = [\mu, \alpha_1, \alpha_2, \beta]'$, $\boldsymbol{y} = [y_{11}, y_{12}, y_{13}, y_{14}, y_{15}, y_{21}, y_{22}, y_{23}, y_{24}, y_{25}]'$, and $\boldsymbol{\varepsilon} = [\varepsilon_{11}, \varepsilon_{12}, \varepsilon_{13}, \varepsilon_{14}, \varepsilon_{15}, \varepsilon_{21}, \varepsilon_{22}, \varepsilon_{23}, \varepsilon_{24}, \varepsilon_{25}]'$. Provide the matrix $\boldsymbol{X}$ so that model (1) can be written as $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.

   (b) Show that $\mu + \alpha_1 + \beta x$ is estimable for any $x \in \mathbb{R}$.

   (c) Show that $\boldsymbol{X}$ in part (a) has the same column space as
$$\boldsymbol{W} = \begin{bmatrix} 1 & -1 & -2 \\ 1 & -1 & -1 \\ 1 & -1 & 0 \\ 1 & -1 & 1 \\ 1 & -1 & 2 \\ 1 & 1 & -2 \\ 1 & 1 & -1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix}.$$

   (d) Suppose $\boldsymbol{y} = [3, 6, 7, 9, 10, 4, 7, 10, 12, 13]'$. Determine the ordinary least squares estimator of $\mu + \alpha_1 + \beta x$ for any fertilizer amount x in pounds of nitrogen per acre.

(e) Now consider the new model

$$y_{ij} = \theta + \gamma_i + \delta_j + \epsilon_{ij}, \tag{2}$$

where θ, γ_1, γ_2, δ_1, δ_2, δ_3, δ_4, and δ_5 are unknown parameters, and the ϵ_{ij} terms are independent and identically distributed as $N(0, \eta^2)$ for some unknown variance parameter $\eta^2 > 0$. SSE for this new model (2) is 2.4. SSE for the original model (1) is 4.8. Compute the F statistic you would use to test whether the new model (2) fits significantly better than the original model (1).

(f) State the degrees of freedom associated with the F statistic in part (e).

3. Suppose $\boldsymbol{M}$ is an $m \times n$ matrix of rank r. Suppose $\boldsymbol{z}$ is an m-dimensional standard multivariate normal random vector.

(a) Provide an expression involving $\boldsymbol{M}$ and $\boldsymbol{z}$ that has a central chi-squared distribution with r degrees of freedom.

(b) Show that the expression involving $\boldsymbol{M}$ and $\boldsymbol{z}$ that you provided in part (a) has a central chi-squared distribution with r degrees of freedom.

(c) Let a be the random variable involving $\boldsymbol{M}$ and $\boldsymbol{z}$ that you defined in part (a). Let $b = \boldsymbol{z}'\boldsymbol{z}$. Provide an expression involving a and b that has a central F distribution with r numerator degrees of freedom and $m - r$ denominator degrees of freedom.

(d) Show that the expression you provided in part (c) has a central F distribution with r numerator degrees of freedom and $m - r$ denominator degrees of freedom.

4. An experiment was conducted to study the effect of two diets (1 and 2) and two drugs (1 and 2) on blood pressure in rats. A total of 40 rats were randomly assigned to the 4 combinations of diet and drug, with 10 rats per combination. Let y_{ijk} be the decrease in blood pressure from the beginning of the study to the end of the study for diet i, drug j, and rat k ($i = 1, 2$; $j = 1, 2$; $k = 1, \ldots, 10$). Suppose

$$y_{ijk} = \mu_{ij} + \epsilon_{ijk}, \tag{3}$$

where the μ_{ij} terms are unknown parameters and the ϵ_{ijk} terms are independent and identically distributed as $N(0, \sigma^2)$ for some unknown variance parameter $\sigma^2 > 0$.

Model (3) was fit to the response values stored in a vector `y` in R. A part of the code and output are provided below.

```
> diet
 [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[21] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
Levels: 1 2


> drug
 [1] 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2
[21] 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2
Levels: 1 2

> o = lm(y ~ diet + drug + diet:drug)
```

```
> coef(o)
(Intercept)        diet2        drug2     diet2:drug2
        4.9         -5.1         -4.4             4.3

> sum(resid(o)^2)
[1] 454.0
```

(a) Determine LSMEANS for the diet factor.

(b) Compute the value of the t-statistic you would use to test for interaction between the diet and drug factors.