STAT 510 **Exam 2** **Spring 2019**

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 answer sheets. 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 discover 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. Prove that $\boldsymbol{Hx} \neq \boldsymbol{0}$ whenever $\boldsymbol{H}$ is an $n \times n$ orthogonal matrix and $\boldsymbol{x} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$.

2. For $i = 1, 2$, suppose $\boldsymbol{y}_i$ is an $n_i \times 1$ multivariate normal random vector with mean $\mu \boldsymbol{1}_{n_i \times 1}$ and variance $\sigma^2 \boldsymbol{V}_i$, where n_i is a positive integer, $\mu \in \mathbb{R}$ is unknown, $\sigma^2 > 0$ is unknown, and $\boldsymbol{V}_i$ is a known positive definite matrix. Furthermore, suppose $\boldsymbol{y}_1$ is independent of $\boldsymbol{y}_2$. So, we have

$$\boldsymbol{y}_1 \sim N\left(\begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix}, \sigma^2 \boldsymbol{V}_1 \right) \quad \text{independent of} \quad \boldsymbol{y}_2 \sim N\left(\begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix}, \sigma^2 \boldsymbol{V}_2 \right).$$

Determine a fully simplified expression for the best linear unbiased estimator of μ.

3. Researchers were interested in studying the effect of two drugs on weight gain in pigs. Researchers set up a full-factorial design with two factors: *drug* (drug 1 vs. drug 2) and *dose* (0 vs. 10 milligrams of drug per kilogram of pig body weight at the time of treatment). Originally, eight pigs were used in the experiment with two pigs assigned to each of the four combinations of *drug* and *dose* using a completely randomized design. However, one pig had to be removed from the experiment for reasons unrelated to the experiment. The data are provided in the following `R` code and output. There is one row for each pig in the `R` `data.frame` `d`. The variables `drug` and `dose` in `d` are factors, and `y` in `d` is weight gain in kilograms over the period of study. These weight gains are presented as integers to make calculations easier. As indicated in the last line of the `R` code, the researchers fit a cell-means model to this dataset with one unrestricted mean for each of the four combinations of the levels of the factors *drug* and *dose*.

```
> d
  drug dose  y
1    1    0  6
2    1    0  2
3    1   10 12
4    1   10  6
5    2    0  4
6    2   10 16
7    2   10 10
> o = lm(y ~  drug + dose + drug:dose, data = d)
```

   (a) Following the `R` code above, the `R` command `anova(o)` would produce an ANOVA table with sequential (i.e., Type I) sums of squares. Determine these sequential sums of squares for `drug`, `dose`, `drug:dose`, `error`, and `corrected total` for this data set.

   (b) Ignoring the data and considering only the treatments used in this experiment, describe a model that should be more appropriate for this dataset than the cell-means model fit by the researchers.

4. A researcher studied factors that affect the acidity of coffee. Each morning for 14 consecutive mornings, the researcher made a pot of coffee. On 7 of the mornings, selected in a completely random way from the 14 mornings, the researcher brewed the coffee for one minute. The other 7 mornings, the researcher brewed the coffee for two minutes. On each of the 14 mornings, the researcher poured the brewed coffee from the pot into two cups, added one tablespoon of cream to one cup selected at random from the two cups, and then measured the acidity of the contents of each cup. This process provided 28 measurements of acidity, i.e., one measurement for each of $14 \times 2 = 28$ cups of coffee. Consider the model you would fit to these data to learn about the effects of the factors *brew time* (one minute vs. two minutes) and *additive* (cream vs. nothing) on acidity. Provide the *Source* and *Degrees of Freedom* columns for the ANOVA table corresponding to your model. [*Note: You are not required to write your model on the answer sheet. Only the Source and Degrees of Freedom columns of the ANOVA table will be graded.*]

5. A plant scientist was interested in comparing two plant genotypes (1 vs. 2) grown in soil of varying moisture levels (1 = low vs. 2 = medium vs. 3 = high). An experiment was conducted in a greenhouse with one table, eight trays, and 24 pots of soil. The table in the greenhouse held the eight trays with three pots on each tray. In each pot, one seed of genotype 1 and one seed of genotype 2 were planted. The three soil moisture levels were assigned, completely at random, to the three pots in each tray. The response of interest is a quantitative measurement of overall plant health that was calculated for each plant 42 days after planting. For $i = 1, 2, 3$, $j = 1, 2$, and $k = 1, \ldots, 8$, let y_{ijk} be the response for soil moisture level i, genotype j, and tray k. Suppose

$$y_{ijk} = \mu_{ij} + t_k + p_{ik} + e_{ijk}, \tag{1}$$

where μ_{ij} is an unknown real-valued parameter, $t_k \sim N(0, \sigma_t^2)$, $p_{ik} \sim N(0, \sigma_p^2)$, $e_{ijk} \sim N(0, \sigma_e^2)$, and all the random effects and errors are independent. Table 1 contains the values of $\bar{y}_{ij\cdot}$ for $i = 1, 2, 3$ and $j = 1, 2$.

Table 1. $\bar{y}_{ij\cdot}$ for $i = 1, 2, 3$ and $j = 1, 2$.

	Genotype 1	Genotype 2
Soil Moisture Level 1	4.2	5.6
Soil Moisture Level 2	6.3	6.1
Soil Moisture Level 3	9.4	8.8

The researchers obtained an ANOVA table that included one line for each of the factors *soil moisture*, *genotype*, and *tray* and all possible interaction among these factors. Mean squares and expected mean squares are provided in Table 2 for a subset of the lines in the ANOVA table obtained by the researchers.

Table 2. Mean Squares and Expected Mean Square for Selected Lines of the ANOVA Table

Source	Mean Square	Expected Mean Square
Tray	17.9	$6\sigma_t^2 + 2\sigma_p^2 + \sigma_e^2$
Tray $\times$ Soil Moisture	5.3	$2\sigma_p^2 + \sigma_e^2$
Tray $\times$ Genotype	3.3	σ_e^2
Tray $\times$ Soil Moisture $\times$ Genotype	3.9	σ_e^2

(a) Determine the value of an unbiased estimator for σ_t^2. [*Note: Multiple answers are possible. Partial credit will be given for the value of any unbiased estimator. Full credit will be given for the value of the lowest-variance unbiased estimator. There is no need to prove that the estimator you choose has lowest variance among available unbiased estimators.*]

(b) Determine the value of an unbiased estimator for σ_e^2. [*Note: Multiple answers are possible. Partial credit will be given for the value of any unbiased estimator. Full credit will be given for the value of the lowest-variance unbiased estimator. There is no need to prove that the estimator you choose has lowest variance among available unbiased estimators.*]

(c) The low, medium, and high soil moisture levels were maintained by adding either 2, 10, or 18 ounces of water per week to the soil of each pot (i.e., 2 ounces for pots with low soil moisture level, 10 ounces for pots with medium soil moisture level, and 18 ounces for pots with high soil moisture level). Determine the value of the test statistic you would use to test whether the genotype 1 means (μ_{11}, μ_{21}, and μ_{31}) are linear in the amount of water added to the soil. In other words, determine the value of the test statistic for testing the null hypothesis that says the points $(2, \mu_{11})$, $(10, \mu_{21})$, and $(18, \mu_{31})$ all fall on a single line.