STAT 510 **Exam 2** **Spring 2018**

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. A total of 18 men who reported difficulties with sleeping participated in an experiment to assess the effects of three factors on duration of sleep. The 18 men were randomly divided into three groups of 6 men each. Men in one group were each given a mattress of type M_1. Men in the second group were each given a mattress of type M_2. Men in the third group were given a mattress of type M_3. Each man was asked to sleep each night on his assigned mattress. The total time asleep was recorded for each man for each of four weeks (sleep was totaled from Sunday night through Thursday night each week, and weekend sleep – defined as sleep initiating on Friday and Saturday nights – was ignored). In a addition to a mattress, each man was given a machine capable of emitting light and sound. For each week that total time asleep was recorded, one of four light and sound treatment combinations was randomly assigned in such a way that all four combinations were assigned to each man over the course of the four weeks of sleep data collection. Combination L_1S_1 involved no light and no sound. Combination L_1S_2 involved no light and exposure to a sound similar to the sound of falling rain. Combination L_2S_1 involved exposure to a red light but no sound, and combination L_2S_2 involved exposure to both the red light and the rainfall sound.

   For $i = 1, 2, 3$, $j = 1, 2$, $k = 1, 2$, and $l = 1, \ldots, 6$, let y_{ijkl} be the total time asleep for the lth man sleeping on mattress type M_i during the week he was treated with light and sound combination L_jS_k. Suppose

   $$y_{ijkl} = \mu_{ijk} + m_{il} + e_{ijkl}, \tag{1}$$

   where μ_{ijk} is an unknown real-valued parameter for each i, j, and k, $m_{il} \sim N(0, \sigma_m^2)$ for each i and l, $e_{ijkl} \sim N(0, \sigma_e^2)$ for each i, j, k, and l, and all m_{il} and e_{ijkl} terms are independent.

   (a) According to model (1) described above, what is the correlation between y_{1111} and y_{1221}?

   (b) Provide the source and the degrees of freedom columns for an ANOVA table that could be used to analyze the data under the assumptions of model (1).

   (c) Model (1) can be written as $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}$, where

   $$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \boldsymbol{y}_3 \end{bmatrix}, \text{ with}$$

   $$\boldsymbol{y}_i = [y_{i111}, y_{i121}, y_{i211}, y_{i221}, y_{i112}, y_{i122}, y_{i212}, y_{i222}, \ldots, y_{i116}, y_{i126}, y_{i216}, y_{i226}]' \text{ for } i = 1, 2, 3,$$

   $$\boldsymbol{\beta} = [\mu_{111}, \mu_{112}, \mu_{121}, \mu_{122}, \mu_{211}, \mu_{212}, \mu_{221}, \mu_{222}, \mu_{311}, \mu_{312}, \mu_{321}, \mu_{322},]', \text{ and}$$

   $$\boldsymbol{u} = [m_{11}, m_{12}, m_{13}, m_{14}, m_{15}, m_{16}, m_{21}, m_{22}, m_{23}, m_{24}, m_{25}, m_{26}, m_{31}, m_{32}, m_{33}, m_{34}, m_{35}, m_{36}]'.$$

   Specify the corresponding $\boldsymbol{X}$ and $\boldsymbol{Z}$ matrices using Kronecker product notation.

2. Consider an experiment with two factors: A and B. Suppose the levels of factor A are indexed by $i = 1, 2$. Suppose the levels of factor B are indexed by $j = 1, 2$. For $i = 1, 2$ and $j = 1, 2$, let n_{ij} be the number of observations for the treatment combination of level i of factor A and level j of factor B. For $i = 1, 2$ and $j = 1, 2$ and $k = 1, \ldots, n_{ij}$, suppose

$$y_{ijk} = \mu_{ij} + e_{ijk},$$

where the μ_{ij} terms are unknown parameters and the e_{ijk} terms are independent and identically distributed as $N(0, \sigma^2)$. The following table contains response averages and the number of observations for each treatment group.

Level of Factor A	Level of Factor B	Average Response ($\bar{y}_{ij\cdot}$)	Number of Observations (n_{ij})
1	1	3.0	2
1	2	5.0	8
2	1	7.0	6
2	2	3.0	4

(a) Find the Type I sum of squares for factor A, assuming the ANOVA table follows the usual order: A, B, $A \times B$, *error*, *corrected total*.

(b) Find the LSMEAN for level 2 of factor A.

(c) Consider the contrast that could be used to test for a factor A main effect. Consider the contrast that could be used to test for a factor B main effect. Determine whether these two contrasts are orthogonal according to our definition of orthogonal contrasts in slide set 9.

3. Consider the Aitken model $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{X}$ is a known $n \times p$ matrix, $\boldsymbol{\beta}$ is an unknown parameter vector in $\mathbb{R}^p$, and $\boldsymbol{\epsilon}$ is a random vector satisfying $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $\text{Var}(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{V}$ for some known positive definite matrix $\boldsymbol{V}$ and some (perhaps unknown) variance component $\sigma^2 > 0$. For this linear model, there is a theorem that says the ordinary least squares (OLS) estimator of an estimable $\boldsymbol{C\beta}$ is the BLUE of $\boldsymbol{C\beta}$ if and only if there exists a matrix $\boldsymbol{Q}$ such that $\boldsymbol{VX} = \boldsymbol{XQ}$.

(a) Suppose $\boldsymbol{V} = \boldsymbol{I}_{n \times n} + a\mathbf{1}_n\mathbf{1}_n'$ for some $a > 0$, where $\mathbf{1}_n$ denotes a $n \times 1$ vector of ones. Show that $\boldsymbol{V}$ is positive definite.

(b) Consider a special case of the Aitken model described at the beginning of this problem, where the model matrix $\boldsymbol{X}$ has the property that $\boldsymbol{X}\mathbf{1}_p = \mathbf{1}_n$, and $\boldsymbol{V} = \boldsymbol{I}_{n \times n} + a\mathbf{1}_n\mathbf{1}_n'$ for some $a > 0$. Use the theorem stated at the beginning of this problem to show that the OLS estimator of an estimable $\boldsymbol{C\beta}$ is the BLUE of $\boldsymbol{C\beta}$ for this special case.

4. Suppose five subjects are asked to complete a task multiple times. Each time a subject completes the task, the subject is a assigned a score that indicates how well the task was completed. Each time the subject is asked to complete the task, he or she is randomly assigned one of two treatments that is in effect while the subject completes the task. Let y_{ijk} be the kth score received by subject i while completing the task under treatment j. For $i = 1, \ldots, 5$, $j = 1, 2$, and $k = 1, 2$, suppose

$$y_{ijk} = s_i + \mu_j + e_{ijk}, \tag{2}$$

where μ_1 and μ_2 are unknown real parameters, $s_1, s_2, s_3, s_4, s_5 \overset{iid}{\sim} N(0, \sigma_s^2)$, each $e_{ijk} \sim N(0, \sigma_e^2)$, and all s_i and e_{ijk} terms are independent. All the data and some summary statistics are provided in the table below. Note that not all subjects completed the task the same number of times for each treatment. Thus, y_{ijk} is not observed for all $i = 1, \ldots, 5$, $j = 1, 2$, and $k = 1, 2$; i.e., y_{ijk} is only observed for those combinations of i, j, and k that are represented in the table below.

Subject ID	Scores for Treatment 1	Scores for Treatment 2	Treatment 1 Average	Treatment 2 Average	Difference of Treatment Averages
1	$y_{111} = 56$, $y_{112} = 50$	$y_{121} = 42$, $y_{122} = 48$	53	45	8
2	$y_{211} = 86$, $y_{212} = 90$	$y_{221} = 75$, $y_{222} = 71$	88	73	15
3	$y_{311} = 72$	$y_{321} = 65$	72	65	7
4	$y_{411} = 80$, $y_{412} = 88$	$y_{421} = 83$	84	83	1
5	$y_{511} = 46$	$y_{521} = 52$, $y_{522} = 44$	46	48	-2
Average	71	60	68.6	62.8	5.8

(a) One linear unbiased estimator of $\mu_1 - \mu_2$ is given by the average of all 8 treatment 1 response values minus the average of all 8 treatment 2 response values. By looking at the very last row of the table, we can see this estimator computes to $71 - 60 = 11$ in this case. Assuming model (2) described above is correct, find an expression for the variance of this estimator in terms of model (2) variance components.

(b) Another linear unbiased estimator of $\mu_1 - \mu_2$ is given by the average of the difference between treatment averages for each subject. The difference between treatment averages is computed in the rightmost column of the data table, separately for each subject. The average of these subject-specific differences is computed as 5.8 in the bottom row and rightmost column of the table. Assuming model (2) holds, find an expression for the variance of this estimator in terms of model (2) variance components.

(c) Define and compute another linear unbiased estimator of $\mu_1 - \mu_2$ that is guaranteed to have lower variance than the linear unbiased estimator described in part (b).