**STAT 510 Homework 7**
**Due Date:** 11:00 A.M., Wednesday, March 13

1. An experiment was conducted to study the durability of coated fabric subjected to abrasive tests. Three factors were considered. One factor was filler type with two levels (F1 and F2). Another was surface treatment with two levels (S1 and S2). The third factor was proportion of filler with three levels (25%, 50%, and 75%). Using a completely randomized design with two fabric samples per treatment, the amount of fabric lost in milligrams for each fabric sample was recorded following testing. Data are available in a tab delimited text file at http://dnett.github.io/S510/FabricLoss.txt. *(Note: You may use R to complete this problem, but you may find it easier to complete with the help of SAS.)*

   (a) Consider a cell means model for these data. Estimate the mean and standard error for the treatment corresponding to F2, S2, and 25% filler.

   (b) The concept of LSMEANS has been explained carefully in lecture and course notes for the special case of a two-factor study. The concept generalizes easily to multi-factor studies. For example, in a three-factor study, the LSMEAN for level $j$ of the second factor is the OLS estimator of $\bar{\mu}_{\cdot j \cdot}$, the average of the cell means for all treatments that involve level $j$ of the second factor. Find LSMEANS for the levels of the factor surface treatment.

   (c) We can also compute LSMEANS for estimable marginal means like $\bar{\mu}_{i \cdot k}$, the average of the cell means for all treatments involving level $i$ of the first factor and level $k$ of the third factor. Find the LSMEAN for filler type F1 and 50% filler.

   (d) Provide a standard error for the estimate computed in part (c).

   (e) In a three-factor study we would say there are no main effects for the second factor if $\bar{\mu}_{\cdot j \cdot} = \bar{\mu}_{\cdot j^* \cdot}$ for all levels $j \neq j^*$. Conduct a test for surface treatment main effects. Provide an $F$-statistic, a $p$-value, and a conclusion.

   (f) In a three-factor study in which the third factor has $K$ levels, we would say there are no three-way interactions if, for all $i \neq i^*$ and $j \neq j^*$,

   $$\mu_{ij1}-\mu_{ij^*1}-\mu_{i^*j1}+\mu_{i^*j^*1} = \mu_{ij2}-\mu_{ij^*2}-\mu_{i^*j2}+\mu_{i^*j^*2} = \cdots = \mu_{ijK}-\mu_{ij^*K}-\mu_{i^*jK}+\mu_{i^*j^*K}.$$

   Note that each linear combination above can be viewed as a two-way interaction effect involving the first two factors while holding the level of the third factor fixed. If these interaction effects are all the same regardless of which level of the third factor is selected, we say there are no three way interactions. Put another equivalent way, there are no three-factor interactions if

   $$\mu_{ijk} - \mu_{ij^*k} - \mu_{i^*jk} + \mu_{i^*j^*k} - \mu_{ijk^*} + \mu_{ij^*k^*} + \mu_{i^*jk^*} - \mu_{i^*j^*k^*} = 0$$

   for all $i \neq i^*$, $j \neq j^*$, and $k \neq k^*$. Conduct a test for three-way interactions among the factors filler type, surface treatment, and filler proportion. Provide an $F$-statistic, a $p$-value, and a conclusion.

   (g) In a three-factor study, we would say there are no two-way interactions between the first and third factors if

   $$\bar{\mu}_{i \cdot k} - \bar{\mu}_{i \cdot k^*} - \bar{\mu}_{i^* \cdot k} + \bar{\mu}_{i^* \cdot k^*} = 0$$

   for all $i \neq i^*$ and $k \neq k^*$. Conduct a test for two-way interactions between the factors filler type and filler proportion. Provide an $F$-statistic, a $p$-value, and a conclusion.

2. An experiment was conducted to compare the effectiveness of two sports drinks (denoted 1 and 2). The subjects included 60 males between the ages of 18 and 31. Each subject rode a stationary bicycle until his muscles were depleted of energy, rested for two hours, and biked again until exhaustion. During the rest period, each subject drank one of the two sports drinks as assigned by the researchers. Each subject's performance on the second round of biking following the rest period was assigned a score between 0 and 100 based on the energy expended prior to exhaustion. Higher scores were indicative of better performance.

   20 of the 60 subjects repeated the bike-rest-bike trial on a second occasion separated from the first by approximately three weeks. These subjects drank one sports drink during the first trial and the other during the second trial. The drink order was randomized for each subject by the researchers, even though previous research suggested no performance difference in repeated trials when three weeks passed between trials. The other 40 subjects performed the trial only a single time, drinking a randomly assigned sports drink during the rest period. 20 of these subjects received sports drink 1, and the other 20 received sports drink 2. A portion of the entire data set is provided in the following table.

| Subject | Drink 1 | Drink 2 |
|---:|---:|---:|
| 1 | 45 | 52 |
| 2 | 69 | 73 |
| $\vdots$ | $\vdots$ | $\vdots$ |
| 20 | 29 | 46 |
| 21 | 35 | - |
| 22 | 81 | - |
| $\vdots$ | $\vdots$ | $\vdots$ |
| 40 | 55 | - |
| 41 | - | 17 |
| 42 | - | 54 |
| $\vdots$ | $\vdots$ | $\vdots$ |
| 60 | - | 61 |

   Subjects 1 through 20 in the table above represent the 20 subjects who performed the trial separately for each of the sports drinks. Note that the data set contains no information about which drink was received in the first trial and which drink was received in the second trial. Throughout the remainder of this problem, please assume that this information is not important. In other words, you may assume that the subjects would have scored the same for drinks 1 and 2 regardless of the order the trials were performed.

   Suppose the following model is appropriate for the data.

$$y_{ij} = \mu_i + u_j + e_{ij}, \tag{1}$$

   where $y_{ij}$ is the score for drink $i$ and subject $j$, $\mu_i$ is the unknown mean score for drink $i$, $u_j$ is a random effect corresponding to subject $j$, and $e_{ij}$ is a random error corresponding to the score for drink $i$ and subject $j$ ($i = 1, 2$ and $j = 1, \ldots, 60$). Here $u_1, \ldots, u_{60}$ are assumed to be independent and identically distributed as $N(0, \sigma_u^2)$ and independent of the $e_{ij}$'s, which are assumed to be independent and identically distributed as $N(0, \sigma_e^2)$.

   (a) For each of the subjects who received both drinks, the difference between the scores (drink 1 score $-$ drink 2 score) was computed. This yielded 20 score differences denoted $d_1, \ldots, d_{20}$.

Describe the distribution of these differences considering the assumptions about the distribution of the original scores in model (1).

(b) Suppose you were given only the differences $d_1, \ldots, d_{20}$ from part (a). Provide a formula for a test statistic (as a function of $d_1, \ldots, d_{20}$) that could be used to test $H_0 : \mu_1 = \mu_2$.

(c) Fully state the exact distribution of the test statistic provided in part (b).

(d) Let $a_1, \ldots, a_{20}$ be the scores of the subjects who received only drink 1. Let $b_1, \ldots, b_{20}$ be the scores of the subjects who received only drink 2. Suppose you were given only these 40 scores. Provide a formula for a 95% confidence interval for $\mu_1 - \mu_2$ (as a function of $a_1, \ldots, a_{20}$ and $b_1, \ldots, b_{20}$).

(e) Suppose you were given $d_1, \ldots, d_{20}$ from part (a) and $a_1, \ldots, a_{20}$ and $b_1, \ldots, b_{20}$ from part (d). Provide formulas for unbiased estimators of $\sigma_u^2$ and $\sigma_e^2$ as a function of these observations.

(f) Suppose you were given $\bar{d}. = \sum_{i=1}^{20} d_i/20$, $\bar{a}. = \sum_{i=1}^{20} a_i/20$, and $\bar{b}. = \sum_{i=1}^{20} b_i/20$; where $d_1, \ldots, d_{20}$ are from part (a) and $a_1, \ldots, a_{20}$ and $b_1, \ldots, b_{20}$ are from part (d). Furthermore, suppose $\sigma_e^2$ and $\sigma_u^2$ are known. Provide a simplified expression for the estimator of $\mu_1 - \mu_2$ that you would use. Your answer should be a function of $\bar{d}.$, $\bar{a}.$,$\bar{b}.$, $\sigma_u^2$, and $\sigma_e^2$.

3. Suppose the responses in problem 2 were sorted first by subject and then by drink into a response vector $\boldsymbol{y}$; i.e.,

$$\boldsymbol{y} = [45, 52, 69, 73, \ldots, 29, 46, 35, 81, \ldots, 55, 17, 54, \ldots, 61]'.$$

Provide $\boldsymbol{X}$ and $\boldsymbol{Z}$ matrices so that the model in equation (1) may be written as $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}$, where $\boldsymbol{\beta} = [\mu_1, \mu_2]'$ and $\boldsymbol{u} = [u_1, u_2, \ldots, u_{60}]'$. If possible, use Kronecker product notation to simplify your answer.

4. The following questions refer to the slide set 12 entitled *The ANOVA Approach to the Analysis of Linear Mixed-Effects Models.*

(a) Derive the expected mean square for $xu(trt)$ for the ANOVA table on slide 9 using the technique illustrated on slides 15 through 17.

(b) For the special case of $t = 2$, $n = 2$, and $m = 2$, repeat part (a) using the technique described on slide 19.