STAT 510 Homework 7

Due Date: 11:00 A.M., Wednesday, March 7

1. 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
⋮	⋮	⋮
20	29	46
21	35	-
22	81	-
⋮	⋮	⋮
40	55	-
41	-	17
42	-	54
⋮	⋮	⋮
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, μ_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 σ_u^2 and σ_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 σ_e^2 and σ_u^2 are known. Provide a simplified expression for the best linear unbiased estimator of $\mu_1 - \mu_2$ in terms of $\bar{d}.$, $\bar{a}.$,$\bar{b}.$, σ_u^2, and σ_e^2.

2. Suppose the responses in problem 1 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 $\boldsymbol{u} = [u_1, u_2, \ldots, u_{60}]'$. Use Kronecker product notation to simplify your answer.

3. If $\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{\epsilon}$, where $\text{Var}(\boldsymbol{\epsilon}) = \sigma_1^2\boldsymbol{I} + \sigma_2^2\mathbf{1}\mathbf{1}'$ for some $\sigma_1^2, \sigma_2^2 > 0$, then the BLUE of μ is $\bar{y}.$, the average of the entries in the vector $\boldsymbol{y}$. Use this fact, along with whatever else you may know about best linear unbiased estimation, to find a simplified expression (in terms of the elements of $\boldsymbol{y}$) for the BLUE of μ in the following situation.

Suppose

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu \\ \mu \\ \mu \\ \mu \\ \mu \end{bmatrix}, \begin{bmatrix} 5 & 1 & 1 & 1 & 0 \\ 1 & 5 & 1 & 1 & 0 \\ 1 & 1 & 5 & 1 & 0 \\ 1 & 1 & 1 & 5 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix} \right).$$

Note that finding the BLUE of μ in this situation does not require inverting any matrices. All the computations needed to obtain the BLUE can be carried out by hand in a minute or two.