STAT 510 Homework 11
Due Date: 11:00 A.M., Wednesday, April 18

1. Suppose $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r]$ is an $n \times r$ matrix of rank r. A technique known as *Gram-Schmidt Orthogonalization* can be used to obtain an $n \times r$ matrix $\boldsymbol{W} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r]$ that has orthogonal columns and the same column space as $\boldsymbol{X}$. This can be useful because $\boldsymbol{P_W}$ is relatively easy to compute when the columns of $\boldsymbol{W}$ are orthogonal, and $\mathcal{C}(\boldsymbol{X}) = \mathcal{C}(\boldsymbol{W})$ implies $\boldsymbol{P_X} = \boldsymbol{P_W}$. An algorithm for carrying out Gram-Schmidt Orthogonalization is as follows:

   - Let $\boldsymbol{w}_1 = \boldsymbol{x}_1$.
   - Let $\boldsymbol{w}_2 = (\boldsymbol{I} - \boldsymbol{P}_{[\boldsymbol{w}_1]})\boldsymbol{x}_2$.
   - Let $\boldsymbol{w}_3 = (\boldsymbol{I} - \boldsymbol{P}_{[\boldsymbol{w}_1, \boldsymbol{w}_2]})\boldsymbol{x}_3$.

     $\vdots$
   - Let $\boldsymbol{w}_r = (\boldsymbol{I} - \boldsymbol{P}_{[\boldsymbol{w}_1, \ldots, \boldsymbol{w}_{r-1}]})\boldsymbol{x}_r$.

   Now consider an experiment with two factors: A and B. Suppose that 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) Provide a model matrix $\boldsymbol{W}$ with orthogonal columns that corresponds to the additive model where $\mu_{ij} = \mu + \alpha_i + \beta_j$ for all $i = 1, 2$, $j = 1, 2$, and some unknown parameters μ, α_1, α_2, β_1, and β_2.

   (b) Without the aid of a computer, find the value of $\boldsymbol{P_W y}$. Note that it is usually easiest to compute this by finding $(\boldsymbol{W'W})^{-1}$, $\boldsymbol{W'y}$, $(\boldsymbol{W'W})^{-1}\boldsymbol{W'y}$, and then finally $\boldsymbol{W}(\boldsymbol{W'W})^{-1}\boldsymbol{W'y}$. Even though you don't have all the elements of $\boldsymbol{y}$ it is still possible to compute $\boldsymbol{P_W y}$ from the averages in the table above.

   (c) Without the aid of a computer, find the Type II sum of squares for factor B. This question can be answered by making use of part (b).

2. An experiment was designed to compare the effect of three drugs (A, B, and C) on the heart rate of women. Fifteen women were randomly assigned to the drugs using a completely randomized design with five women for each drug. The heart rate (in beats per minute) of each woman was measured at 0, 5, 10, and 15 minutes after the drug was administered. The data are provided in the file

Let y_{ijk} denote the heart rate at the kth time point for the jth woman treated with the ith drug. Suppose

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

where μ_{ik} is an unknown constant for each combination of $i = 1, 2, 3$ and $j = 1, 2, 3, 4, 5$; $w_{ij} \sim N(0, \sigma_w^2)$ for all i and j; $e_{ijk} \sim N(0, \sigma_e^2)$ for all i, j, and k; and all random effects are independent.

(a) Using Kronecker product notation, provide an expression for the variance of

$$\boldsymbol{y} = [y_{111}, y_{112}, y_{113}, y_{114}, y_{121}, y_{122}, y_{123}, y_{124}, \ldots, y_{351}, y_{352}, y_{353}, y_{354}]'.$$

(b) Test the null hypothesis of no drug-by-time interactions. Compute a test statistic, state its degrees of freedom, find a p-value, and provide a brief conclusion.

(c) Is the mean heart rate 15 minutes after treatment the same for all three drugs? State a formal null hypothesis corresponding to this question. Compute a test statistic, state its degrees of freedom, find a p-value, and provide a brief conclusion.

(d) Compute a 95% confidence interval for the mean heart rate 15 minutes after treatment with drug A minus the mean heart rate 15 minutes after treatment with drug B.

3. Again consider the data on heart rate from problem 2. Now suppose

$$y_{ijk} = \mu_{ik} + \epsilon_{ijk},$$

where μ_{ik} is an unknown constant for each combination of $i = 1, 2, 3$ and $k = 1, 2, 3, 4, 5$ and ϵ_{ijk} is a normally distributed error term with mean 0 for all $i = 1, 2, 3$, $j = 1, 2, 3, 4, 5$, and $k = 1, 2, 3, 4$. For all $i = 1, 2, 3$ and $j = 1, 2, 3, 4, 5$, let

$$\boldsymbol{\epsilon}_{ij} = (\epsilon_{ij1}, \epsilon_{ij2}, \epsilon_{ij3}, \epsilon_{ij4})'.$$

Suppose all the $\boldsymbol{\epsilon}_{ij}$ vectors are mutually independent, and let $\boldsymbol{W}$ be the variance-covariance matrix of $\boldsymbol{\epsilon}_{ij}$, which is assumed to be the same for all $i = 1, 2, 3$ and $j = 1, 2, 3, 4, 5$.

(a) Find the REML estimate of $\boldsymbol{W}$ under the assumption that $\boldsymbol{W}$ is a positive definite, compound symmetric matrix.

(b) Find AIC and BIC for the case where $\boldsymbol{W}$ is a positive definite, compound symmetric matrix.

(c) Find the REML estimate of $\boldsymbol{W}$ under the assumption that $\boldsymbol{W}$ is a positive definite matrix with constant variance and an AR(1) correlation structure.

(d) Find AIC and BIC for the case where $\boldsymbol{W}$ is a a positive definite matrix with constant variance and an AR(1) correlation structure.

(e) Find the REML estimate of $\boldsymbol{W}$ under the assumption that $\boldsymbol{W}$ is a positive definite, symmetric matrix.

(f) Find AIC and BIC for the case where $\boldsymbol{W}$ is a positive definite, symmetric matrix.

(g) Which of the three structures for $\boldsymbol{W}$ is preferred for this dataset?

(h) Using the preferred structure for $\boldsymbol{W}$, compute a 95% confidence interval for the mean heart rate 15 minutes after treatment with drug A minus the mean heart rate 15 minutes after treatment with drug B.

4. Consider a class of 50 students. Suppose each student is required to take two midterm exams and one final exam. The midterms are coded as 1 and 2, and the final exam is coded as 3 in the dataset available at

http://dnett.github.io/S510/ExamScores.txt

In the dataset, note that student 1 has scores for only exams 1 and 2. Suppose student 1 was not able to take the final exam due to a medical emergency. For $i = 1, \ldots, 50$ and $j = 1, 2, 3$, let y_{ij} be the score for student i on exam j. For $i = 1, \ldots, 50$ and $j = 1, 2, 3$, suppose

$$y_{ij} = s_i + \mu_j + e_{ij},$$

where μ_i is an unknown parameter, $s_i \sim N(0, \sigma_s^2)$, $e_{ij} \sim N(0, \sigma_j^2)$ for some unknown variance parameter $\sigma_j^2 > 0$, and all s_i and e_{ij} terms are independent. For this model, note that the variance of the error terms is not constant and instead depends on the exam. This model may be fit to the data using the R code

```
lme(score ~ 0 + exam, random = ~ 1 | student,
                          weights = varIdent(form = ~ 1 | exam), data = d)
```

(This code assumes the data are in a data.frame `d`, where `exam` and `student` are factors. Be careful with cutting and pasting from this pdf to R as some characters (e.g., $\sim$) may not translate properly.)

(a) Find REML estimates of σ_s^2, σ_1^2, σ_2^2, and σ_3^2.

(b) Find the eBLUP of student 1's exam 3 score.