# STAT 510 Homework 11

**Due Date:** 11:00 A.M., Wednesday, April 22

1. 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

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

   Let $y_{ijk}$ denote the heart rate at the $k$th time point for the $j$th woman treated with the $i$th drug. Suppose

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

   where $\mu_{ik}$ 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 10 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 10 minutes after treatment with drug A minus the mean heart rate 10 minutes after treatment with drug B.

2. Again consider the data on heart rate from problem 1. Now suppose

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

   where $\mu_{ik}$ is an unknown constant for each combination of $i = 1, 2, 3$ and $k = 1, 2, 3, 4, 5$ and $\epsilon_{ijk}$ 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 10 minutes after treatment with drug A minus the mean heart rate 10 minutes after treatment with drug B.

(i) Using the preferred structure for $\boldsymbol{W}$, compute a 95% confidence interval for the mean heart rate 10 minutes after treatment with drug A minus the mean heart rate 5 minutes after treatment with drug A.

3. 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, 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

$$y_{ij} = s_i + \mu_j + e_{ij}, \tag{1}$$

where $\mu_j$ 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, 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

```
library(nlme)
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) Use R (or SAS if you prefer) to find REML estimates of $\sigma_s^2$, $\sigma_1^2$, $\sigma_2^2$, and $\sigma_3^2$.

(b) Use R (or SAS if you prefer), to find the EBLUP of student 1's exam 3 score.

(c) Find an expression for $E(y_{13}|y_{11}, y_{12})$ in terms of model (1) parameters.

(d) Compute an estimate of $E(y_{13}|y_{11}, y_{12})$ by replacing parameters with their estimates and replacing $y_{11}$ and $y_{12}$ with their observed values.

(e) For $i = 1, \ldots, 50$, let $\boldsymbol{y}_i = [y_{i1}, y_{i2}, y_{i3}]'$, and suppose $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{50} \overset{iid}{\sim} N([\mu_1, \mu_2, \mu_3]', \boldsymbol{W})$, where

$$\boldsymbol{W} = \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{12} & w_{22} & w_{23} \\ w_{13} & w_{23} & w_{33} \end{bmatrix}$$

is an unknown, positive definite variance-covariance matrix. Repeat parts (c) and (d) for this new model.

(f) Multiple linear regression could also be used to predict the missing exam 3 score for student 1. Using the data from students 2 through 50, fit a multiple linear regression of exam 3 score on exam 1 score and exam 2 score. Use an intercept and assume additive effects for exam 1 score and exam 2 score. Provide the estimated regression equation and provide a prediction for the exam 3 score of student 1 based on the estimated regression equation.

[*Note: This problem has asked for three predictions of the exam 3 score for student 1. Two of the three predictions should be the same.*]