STAT 510 Homework 8

Due Date: 11:00 A.M., Wednesday, March 21

1. Suppose $\boldsymbol{y}$ is an $n \times 1$ random vector with distribution $N(\boldsymbol{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a positive definite variance matrix with orthonormal eigenvectors $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n$ and corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Find the distribution of the random vector $[\boldsymbol{p}_1'\boldsymbol{y}, \boldsymbol{p}_2'\boldsymbol{y}, \ldots, \boldsymbol{p}_n'\boldsymbol{y}]'$ and simplify your answer as much as possible.

2. A meat scientist is studying the effect of storage temperature on meat quality. The temperatures of interest are 34, 40, and 46 degrees Fahrenheit. Twelve coolers are available for the study. The three temperatures are randomly assigned to the twelve coolers using a balanced and completely randomized design. Two large cuts of fresh beef are stored in each cooler. After three days, each member of a team of experts independently assigns a quality score to each cut of beef. The experts are not told about the storage conditions of each cut. The scores assigned by the team to each cut of beef are averaged to produce an overall quality score for each cut.

   (a) Let y_{ijk} denote the overall quality score for the kth cut of beef stored in the jth cooler set at temperature i, where $k = 1, 2$, $j = 1, 2, 3, 4$, and $i = 1, 2, 3$ for temperatures 34, 40, and 46 degrees Fahrenheit, respectively. Specify $\boldsymbol{y}$, $\boldsymbol{X}$, $\boldsymbol{\beta}$, $\boldsymbol{Z}$, $\boldsymbol{u}$, and $\boldsymbol{\epsilon}$ for this situation, and specify a linear-mixed effects model for the data as we have done in class for other examples.

   (b) Write down an ANOVA table for the overall quality score data. Include Source, Degrees of Freedom, Sums of Squares, Mean Squares, and Expected Mean Squares columns.

   (c) Suppose the researchers wish to know if the mean overall quality score for 34° is significantly different from the mean overall quality score for 40°. Provide a formula for the test statistic you would use to address this question.

   (d) State the degrees of freedom associated with the test statistic in part (c).

   (e) State the noncentrality parameter associated with the test statistic in part (c).

3. An experiment was conducted to study mean plant height of two genotypes exposed to three watering levels. The experiment was conducted in 4 greenhouses. Each greenhouse contained three tables. On each table, were 2 pots with 1 plant in each pot. The 2 plants on any given table consisted of 1 plant of one genotype and 1 plant of the other genotype, with genotypes randomly assigned to the pots. Within each greenhouse, the three watering levels were randomly assigned to the three tables, with one table per watering level. Thus, the two plants on any given table received the same amount of water throughout the experiment. At the conclusion of the experiment, the height of each plant was recorded.

   For $i = 1, 2, 3, 4$; $j = 1, 2, 3$; and $k = 1, 2$; let y_{ijk} denote the height recorded for the plant associated with greenhouse i, watering level j, and genotype k. Consider the following model that will be referred to henceforth as MODEL 1.

   $$y_{ijk} = \mu + g_i + \omega_j + t_{ij} + \gamma_k + \phi_{jk} + e_{ijk} \quad (i = 1, 2, 3, 4; j = 1, 2, 3; k = 1, 2),$$

   where the g_i terms are $N(0, \sigma_g^2)$, the t_{ij} terms are $N(0, \sigma_t^2)$, the e_{ijk} terms are $N(0, \sigma_e^2)$, all these random terms are mutually independent, and the remaining terms in the model are unknown fixed parameters.

   (a) According to MODEL 1, what is the correlation between the heights of two plants growing together on the same table? (Note that this question is asking for the correlation rather than the covariance.)

(b) In terms of the MODEL 1 parameters, write down the null hypothesis of no watering level main effects.

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

$$\begin{aligned}\boldsymbol{y} &= (y_{111}, y_{112}, y_{121}, y_{122}, y_{131}, y_{132}, y_{211}, y_{212}, y_{221}, y_{222}, y_{231}, y_{232}, \\ &\quad y_{311}, y_{312}, y_{321}, y_{322}, y_{331}, y_{332}, y_{411}, y_{412}, y_{421}, y_{422}, y_{431}, y_{432})'.\end{aligned}$$

Provide corresponding expressions for $\boldsymbol{X}$, $\boldsymbol{\beta}$, $\boldsymbol{Z}$, and $\boldsymbol{u}$. You may wish to express parts of your answer using Kronecker product notation.

4. 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, S1, and 50% 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 i of the first factor is the OLS estimator of $\bar{\mu}_{i\cdot\cdot}$, the average of the cell means for all treatments that involve level i of the first factor. Find LSMEANS for the levels of the factor filler type.

(c) We can also compute LSMEANS for estimable marginal means like $\bar{\mu}_{\cdot jk}$, the average of the cell means for all treatments involving level j of the second factor and level k of the third factor. Find the LSMEAN for surface treatment S2 and 25% 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 first factor if $\bar{\mu}_{i\cdot\cdot} = \bar{\mu}_{i^*\cdot\cdot}$ for all levels $i \neq i^*$. Conduct a test for filler type 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.