STAT 510 Homework 12

Due Date: 11:00 A.M., Wednesday, April 26

1. An experiment was conducted at 15 research stations around the country to determine how dose of a chemical mixture affects the leaf area of a certain type of plant. Instructions for creating the chemical mixture and for carrying out the experiment were sent to the managers at each of the 15 research stations. At each station, a completely randomized design was used to assign 5 doses of the chemical (0, 25, 50, 75, and 100 mL/day) to 20 plants with 4 plants per dose. Each plant grew in its own pot and received its assigned chemical dose each day of the experiment. At three weeks of age, the leaf area of each plant was recorded. The data for this experiment are available in the file

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

   For $i = 1, \ldots, 15$, $j = 1, \ldots, 5$, and $k = 1, \ldots, 4$, let y_{ijk} be the leaf area for the kth plant that received dose j in research station i, and suppose

   $$y_{ijk} = (\beta_1 + b_{1i}) + (\beta_2 + b_{2i})x_j + e_{ijk} \tag{1}$$

   In this model (1), β_1 and β_2 are unknown parameters, $x_1 = 0$, $x_2 = 25$, $x_3 = 50$, $x_4 = 75$, $x_5 = 100$, the e_{ijk} terms are iid $N(0, \sigma_e^2)$, and the b_{1i} and b_{2i} terms are normal random effects independent of the e_{ijk} terms. More specifically, let

   $$\boldsymbol{b}_i = \begin{bmatrix} b_{1i} \\ b_{2i} \end{bmatrix} \text{ for all } i = 1, \ldots, 15.$$

   We assume $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{15} \overset{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma_b})$ for some positive definite 2×2 variance matrix $\boldsymbol{\Sigma_b}$.

   Model (1) is a special case of what is sometimes referred to as a *random coefficient model* because the regression coefficients are assumed to be random variables rather than fixed parameters. It is straightforward to fit such a model in `R` using code like the following.

```
d = read.delim("http://dnett.github.io/S510/LeafArea.txt")
library(lme4)
o = lmer(LeafArea ~ Dose + (1 + Dose | ResearchStation), data = d)
```

   The approximate BLUEs of β_1 and β_2 can be obtained with code like

```
fixef(o)
```

   As usual, the estimated variance of the estimator of the fixed effects parameters is given by `vcov(o)`.

   The empirical BLUPs of b_{1i} and b_{2i} can be obtained with code like

```
ranef(o)
```

   Typing `summary(o)` provides you with enough information to determine the REML estimates of σ_e^2 and $\boldsymbol{\Sigma_b}$. The estimate of the matrix $\boldsymbol{\Sigma_b}$ is not provided directly, but you can compute it from the given estimates of the variances and the provided estimate of the correlation between b_{1i} and b_{2i} labeled `Corr` in the `Random effects` portion of the output.

(a) Provide the REML estimate of σ_e^2.

(b) Provide the REML estimate of $\boldsymbol{\Sigma}_{\boldsymbol{b}}$.

(c) Make a scatterplot of leaf area vs. dose for the data from the 7th research station. Add a black line to the plot that shows the estimate of the regression function $\beta_1 + \beta_2 x$ for $x \in (0, 100)$.

(d) Find the prediction of the regression function for the 7th research station; i.e., predict $(\beta_1 + b_{17}) + (\beta_2 + b_{27})x$ for $x \in (0, 100)$.

(e) Using only the data from the 7th research station, find the ordinary least squares estimate of the regression function for the simple linear regression of leaf area on dose of the chemical.

(f) To the plot in part (c), add a red line that shows the regression function predicted in part (d) and a blue line that shows the regression function estimated in part (e).

(g) Compute the likelihood ratio statistic for testing $H_0 : \beta_2 = 0$.

(h) Find AIC for the fit of model (1) to the data from all 15 research stations.

(i) Find AIC for a simplified version of model (1) that assumes there is one slope coefficient common to all research stations.

(j) Find AIC for a simplified version of model (1) that assumes there is one intercept coefficient common to all research stations and one slope coefficient common to all research stations.

(k) According to AIC, which model is preferred among model (1), the model considered in part (i), and the model considered in part (j).

2. Model (1) from problem 1 can be written in linear mixed-effects model form as $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e}$. Define $\boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{Z}, \boldsymbol{u}, \boldsymbol{G} = \text{Var}(\boldsymbol{u})$, and $\boldsymbol{R} = \text{Var}(\boldsymbol{e})$ using terms from model (1). Assume the response vector $\boldsymbol{y}$ is ordered as in the dataset LeafArea.txt.

3. Consider a generic repeated measures experiment like the experiment on strength training programs that we considered in class. Suppose there are three treatments indexed by $i = 1, 2, 3$ with n_i subjects indexed by $j = 1, \dots, n_i$ for the ith treatment group. Suppose the response of interest is measured at t time points for each subject. Let y_{ijk} be the response for treatment i, subject j, and time point k $(i = 1, 2, 3; j = 1, \dots, n_i; k = 1, \dots, t)$. For all i and j, let

$$\boldsymbol{y}_{ij} = (y_{ij1}, \dots, y_{ijt})'.$$

Suppose all $\boldsymbol{y}_{ij}$ are mutually independent of one another and that, for all i and j,

$$\boldsymbol{y}_{ij} \sim N(\boldsymbol{\mu}_i, \boldsymbol{W}),$$

where $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{it})'$ and $\boldsymbol{W}$ is some unknown $t \times t$ positive definite and symmetric matrix. Let

$$\boldsymbol{y} = (\boldsymbol{y}'_{11}, \dots, \boldsymbol{y}'_{1n_1}, \boldsymbol{y}'_{21}, \dots, \boldsymbol{y}'_{2n_2}, \boldsymbol{y}'_{31}, \dots, \boldsymbol{y}'_{3n_3})', \text{ and let } \boldsymbol{\beta} = (\boldsymbol{\mu}'_1, \boldsymbol{\mu}'_2, \boldsymbol{\mu}'_3)'.$$

(a) Use Kronecker product notation to specify a matrix $\boldsymbol{X}$ so that $E(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{\beta}$.

(b) Use Kronecker product notation to specify $\text{Var}(\boldsymbol{y}) = \boldsymbol{\Sigma}$ in terms of $\boldsymbol{W}$.

(c) Find a simplified expression for $(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}$.

(d) Find a simplified expression for $(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}$.

(e) Find a simplified expression for $(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$.

(f) Give a simplified expressions for the BLUEs of $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, and $\boldsymbol{\mu}_3$.