Homework 8

Due Thursday, 4/8/20 on Canvas.

1. Suppose $Y_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta} + \delta_i + \epsilon_{ij}$ for $j = 1, \ldots, n_i$ and $i = 1, \ldots, r$, where $\boldsymbol{x}_{ij}, \boldsymbol{\beta} \in \mathbb{R}^p$ are non-random, $\delta_i \overset{i.i.d}{\sim} N\left(0, \sigma_\delta^2\right)$, $\epsilon_{ij} \overset{i.i.d}{\sim} N\left(0, \sigma^2\right)$ and δ_i, ϵ_{ij} are independent. Note that this is not necessarily a balanced design. Let $n = \sum_{i=1}^{r} n_i$ and define $\boldsymbol{Y} \in \mathbb{R}^n$ to be such that entries are arranged by individual i, and then replicate j (i.e. the first n_1 entries of $\boldsymbol{Y}$ are $Y_{11}, \ldots, Y_{1n_1}$).

   (a) Show that for $\boldsymbol{\delta} \sim N(\boldsymbol{0}, \sigma_\delta^2 I_r)$, $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2 I_n)$ and matrices $\boldsymbol{X} \in \mathbb{R}^{n \times p}$, $\boldsymbol{Z} \in \mathbb{R}^{n \times r}$,

   $$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\delta} + \boldsymbol{\epsilon}.$$

   Find an expression for $\boldsymbol{Z}$.

   (b) Suppose $\boldsymbol{X}$ has full column rank. Show that there exists a matrix $\boldsymbol{Q} \in \mathbb{R}^{n \times (n-p)}$ with orthonormal columns such that for all $\boldsymbol{\beta} \in \mathbb{R}^p$,

   $$\tilde{\boldsymbol{Y}} = \boldsymbol{Q}^T \boldsymbol{Y} = \left(\boldsymbol{Q}^T \boldsymbol{Z}\right) \boldsymbol{\delta} + \tilde{\epsilon}, \quad \tilde{\epsilon} \sim N\left(\boldsymbol{0}, \sigma^2 I_{n-p}\right).$$

   Is matrix $\boldsymbol{Q}$ unique (i.e. can you find another matrix with orthonormal columns that satisfies the above equality for all $\boldsymbol{\beta}$)? Why or why not?

   (c) Suppose $\tilde{\boldsymbol{Z}} = \boldsymbol{Q}^T \boldsymbol{Z}$ has column rank $d \le r$ (if $\boldsymbol{X} = \boldsymbol{1}_n$, then $d = r - 1$), and let $\tilde{\boldsymbol{H}}$ be the orthogonal projection matrix for $\tilde{\boldsymbol{Z}}$ (note that $\tilde{\boldsymbol{Z}}^T \tilde{\boldsymbol{Z}}$ will not be invertible in general). Define $SSE = \tilde{\boldsymbol{Y}}^T \left(I_{n-p} - \tilde{\boldsymbol{H}}\right) \tilde{\boldsymbol{Y}}$ and $SSTR = \tilde{\boldsymbol{Y}}^T \tilde{\boldsymbol{H}} \tilde{\boldsymbol{Y}}$.

      (i) Show that SSE and $SSTR$ are independent, and do not depend on the choice of $\boldsymbol{Q}$ from part (b).

      (ii) Show that $MSE = (n - p - d)^{-1} SSE$ and $MSTR = d^{-1} SSTR$ are unbiased estimators for σ^2 and $\sigma^2 + d^{-1} \operatorname{Tr}\left(\tilde{\boldsymbol{Z}}^T \tilde{\boldsymbol{Z}}\right) \sigma_\delta^2$.

      (iii) If the null hypothesis $H_0 : \sigma_\delta^2 = 0$ is true, show that $F = \frac{MSTR}{MSE} \sim F_{d, n-p-d}$.

2. (a) Suppose $Y_1, \ldots, Y_k$ all have finite second moment with $\operatorname{Corr}(Y_r, Y_s) = \rho$ for all $r \neq s \in \{1, \ldots, k\}$ (note that $Y_1, \ldots, Y_k$ do not necessarily have the same variance). Show that $\rho \ge \frac{-1}{k-1}$.

   (b) Give an intuitive argument as to why $\rho > -1$ for $k \ge 3$.

3. Suppose $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ is full rank, $\boldsymbol{\beta} \in \mathbb{R}^p$, $\mathbb{E}(\boldsymbol{\epsilon}) = \boldsymbol{0}_n$ and $\operatorname{Var}(\boldsymbol{Y}) = \sigma^2 I_n$.

   (a) Let

   $$\left\{\hat{\sigma}_{ML}^2, \hat{\boldsymbol{\beta}}\right\} = \underset{\sigma^2 > 0,\, \boldsymbol{\beta} \in \mathbb{R}^p}{\arg\max} \left[-\frac{1}{2} \log\left\{\det\left(\sigma^2 I_n\right)\right\} - \frac{1}{2\sigma^2} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})\right].$$

   be the maximum quasi-likelihood estimates for σ^2 and $\boldsymbol{\beta}$. Show that $\mathbb{E}\left(\hat{\sigma}_{ML}^2\right) = \frac{n-p}{n} \sigma^2 < \sigma^2$.

(b) Let $\hat{\sigma}^2_{REML}$ be the REML estimate for σ^2. Show that $\hat{\sigma}^2_{REML} = \hat{\sigma}^2_{OLS}$, where $\hat{\sigma}^2_{OLS}$ is the unbiased estimate for σ^2 that we typically use in ordinary least squares (i.e. the MSE).

4. The data in the file "sleep.txt" (with 40 rows of data) give the average hours (times 10) of sleep (in the third column) of 10 insomniacs (indexed with the numbers from 1 to 10, in the first column of the file) without treatment (A) and with three different drugs (B,C and D), of which C and D are of the same general type (but are not identical) and B is a different type of drug. The averages are over a varying number of nights (from 3 through 9), but the specific number of nights for each entry is unavailable.

   (a) Fitting an additive fixed effects model (treating treatment and individual as fixed effects), estimate σ, the standard deviation of the errors. Next, using only the data for treatments C and D (i.e., leave out B and A), fit an additive fixed effects model and again estimate σ, the standard deviation of the errors. Compare the two estimates of σ and give an explanation for any difference you find.

   (b) Is there any evidence of a person × treatment interaction? If so, what is the nature of this interaction? In particular, does your answer to the previous point provide evidence of an interaction?

   (c) Suppose the 10 patients can be thought of as a random sample from the population of insomniacs. Assume for the rest of the problem that the average hours of sleep is normally distributed

      (i) Write down a model for the average hours of sleep for insomniac patients, assuming treatment is a fixed effect. Make sure to allow for the possibility of a patient × treatment interaction if you found evidence for it in part (b).

      (ii) Assuming a patient × treatment interaction, use part (i) to derive a model for the average number of hours a patient will sleep after being treated with drug B, C or D, given that they slept y_0 hours before treatment.

      (iii) Given your model from part (ii), do you think the mean effect of each drug over this population is a quantity of clinical importance? When answering this question, think about whether or not the drug you would recommend an insomniac take depends on how many hours they currently sleep.