STA 2101/442 Assignment Eleven[1]

The non-computer questions are just practice for the quiz, and are not to be handed in. Use R for Question 4, and bring your printout to the quiz. **Your printout should show *all* R input and output, and *only* R input and output**. Do not write anything on your printouts except your name and student number.

1. The general mixed linear model is $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zb} + \boldsymbol{\epsilon}$, where

   - $\mathbf{X}$ is an $n \times p$ matrix of known constants.
   - $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown constants.
   - $\mathbf{Z}$ is an $n \times q$ matrix of known constants.
   - $\mathbf{b} \sim N_q(\mathbf{0}, \boldsymbol{\Sigma}_b)$ with $\boldsymbol{\Sigma}_b$ unknown but often diagonal.
   - $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n)$, where $\sigma^2 > 0$ is an unknown constant.

   (a) What is the distribution of $\mathbf{y}$? Include expressions for the expected value and covariance matrix.

   (b) Suppose you use ordinary least squares to estimate $\boldsymbol{\beta}$. What is the distribution of $\widehat{\boldsymbol{\beta}}$? Include expressions for the expected value and covariance matrix.

   (c) Is $\widehat{\boldsymbol{\beta}}$ still an unbiased estimator of $\boldsymbol{\beta}$? Answer Yes or No.

   (d) As preparation for the next question, let $\mathbf{w}$ be a random vector with expected value $\boldsymbol{\mu}_w$ and covariance matrix $\boldsymbol{\Sigma}_w$. Find a convenient expression for $cov(\mathbf{Aw}, \mathbf{Bw})$, where $\mathbf{A}$ and $\mathbf{B}$ are matrices of the right size.

   (e) All the standard F-tests and t-tests for the normal linear model rely on the independence of $\widehat{\boldsymbol{\beta}}$ and the vector of residuals $\mathbf{e}$. Are $\widehat{\boldsymbol{\beta}}$ and $\mathbf{e}$ still independent under this model? Carry out the calculation and answer Yes or No.

2. In lecture, the following model was suggested for paired normal data. In practice, if you believed this model you'd calculate differences and do a matched t-test. Here is the model. Independently for $i = 1, \ldots, n$,

$$\begin{aligned} y_{i,1} &= \mu_1 + \tau_i + \epsilon_{i,1} \\ y_{i,2} &= \mu_2 + \tau_i + \epsilon_{i,2}, \end{aligned}$$

   where $\tau_i \sim N(0, \sigma^2_\tau)$, $\tau_i \sim N(0, \sigma^2_1)$ and $\tau_i \sim N(0, \sigma^2_2)$ are all independent. The task is to fit this into the matrix format of Question 1, sticking to the specific case of $n = 5$ to keep the matrices small. Just put symbols from the model above into the matrices. Don't re-parameterize. This may take more than one sheet of paper but I think it's worth it.

   (a) What is $\mathbf{y}$? Give all 10 elements.

   (b) What is $\mathbf{X}$?

   (c) What is $\boldsymbol{\beta}$? It has 2 elements.

[1]Copyright information is at the end of the last page.

(d) What is $\mathbf{Z}$?

(e) What is $\mathbf{b}$? It has 5 elements.

(f) What is $\boldsymbol{\epsilon}$? Give all 10 elements.

(g) Finally, the two observations coming from the same individual are definitely not independent. What is $Cov(y_{i,1}, y_{i,2})$?

3. Here is a model for a single random factor in which there are q randomly selected factor level and k observations are collected at each factor level; say k fish are caught at each of q randomly selected lakes. Let $y_{ij} = \mu_{.} + \tau_i + \epsilon_{ij}$, where

   $\mu_{.}$ is an unknown constant parameter.

   $\tau_i \sim N(0, \sigma^2_\tau)$

   $\epsilon_{ij} \sim N(0, \sigma^2)$

   τ_i and ϵ_{ij} are all independent.

   $\sigma^2_\tau \geq 0$ and $\sigma^2 > 0$ are unknown parameters.

   $i = 1, \ldots q$ and $j = 1, \ldots, k$

   (a) What is $Var(y_{ij})$?

   (b) What is $Cov(y_{ij}, y_{ij'})$, where $j \neq j'$. This is the covariance of the weights of two fish taken from the same lake. Show your work.

   (c) Suppose $k = 4$ fish are caught at each lake. Give the covariance matrix of the vector of observations from lake i.

4. In a taste test of wine, 6 professional judges judged 4 specific wines, tasted in a different random order for each judge. The numbers they gave do not exactly represent quality. Instead, they are maximum prices in dollars per bottle that the judge thinks the company can charge and still sell most of the wine. I suppose we are assuming that the 6 judges are some kind of random sample, even though they probably are not.

   The data are available in the file `Wine.data.txt`. The question is whether these wines differ in mean potential price. Go ahead and assume normality; use a mixed model. If the overall test is significant, don't get fancy; follow up with all pariwise matched t-tests, using a Bonferroni correction. This way you will be able to draw directional conclusions if any are justified.

Please bring your printout for Question 4 to the quiz. **Your printout should show *all* R input and output, and *only* R input and output**. Do not write anything on your printouts except your name and student number.

This assignment was prepared by Jerry Brunner, Department of Statistics, University of Toronto. It is licensed under a Creative Commons Attribution - ShareAlike 3.0 Unported License. Use any part of it as you like and share the result freely. The LaTeX source code is available from the course website: `http://www.utstat.toronto.edu/~brunner/oldclass/appliedf17`