STAT 510 Homework 5

Due Date: 11:00 A.M., Wednesday, February 27

1. Suppose $\boldsymbol{X}$ is an $n \times p$ matrix and $\boldsymbol{B}$ is a $p \times p$ non-singular matrix. Prove that

$$\mathcal{C}(\boldsymbol{X}) = \mathcal{C}\left(\boldsymbol{X}\boldsymbol{B}^{-1}\right).$$

2. Consider the Gauss-Markov model with normal errors $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$. For any nonsingular $p \times p$ matrix $\boldsymbol{B}$, the model can be reparameterized by

$$\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}\boldsymbol{B}^{-1}\boldsymbol{B}\boldsymbol{\beta} = \boldsymbol{W}\boldsymbol{\alpha}, \text{ where } \boldsymbol{W} = \boldsymbol{X}\boldsymbol{B}^{-1} \text{ and } \boldsymbol{\alpha} = \boldsymbol{B}\boldsymbol{\beta}.$$

From problem 1, we know the column spaces of $\boldsymbol{X}$ and $\boldsymbol{W}$ are identical so that $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ and $\boldsymbol{y} = \boldsymbol{W}\boldsymbol{\alpha} + \boldsymbol{\epsilon}$ are the same models. Suppose $\boldsymbol{C}$ is a $q \times p$ matrix of rank $q < p$. Then there exists a $(p-q) \times p$ matrix $\boldsymbol{A}$ such that

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{C} \end{bmatrix} \text{ has rank } p \text{ and is therefore nonsingular.}$$

Then we can write

$$\boldsymbol{B}\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{C} \end{bmatrix}\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{A}\boldsymbol{\beta} \\ \boldsymbol{C}\boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \end{bmatrix} = \boldsymbol{\alpha},$$

where $\boldsymbol{\alpha}_1 = \boldsymbol{A}\boldsymbol{\beta}$ and $\boldsymbol{\alpha}_2 = \boldsymbol{C}\boldsymbol{\beta}$. If we let $\boldsymbol{W}_1$ be the matrix consisting of the first $p-q$ columns of $\boldsymbol{W} = \boldsymbol{X}\boldsymbol{B}^{-1}$ and $\boldsymbol{W}_2$ be the matrix consisting of the last q columns of $\boldsymbol{W} = \boldsymbol{X}\boldsymbol{B}^{-1}$, then

$$\boldsymbol{W}\boldsymbol{\alpha} = [\boldsymbol{W}_1, \boldsymbol{W}_2]\begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \end{bmatrix} = \boldsymbol{W}_1\boldsymbol{\alpha}_1 + \boldsymbol{W}_2\boldsymbol{\alpha}_2.$$

Now consider testing

$$H_0 : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{0} \text{ vs. } H_A : \boldsymbol{C}\boldsymbol{\beta} \neq \boldsymbol{0}.$$

(a) Rewrite these hypotheses in terms of the $\boldsymbol{\alpha}$ parameter vector.

(b) If you wanted to fit a reduced model corresponding to the null hypothesis, what model matrix would you use?

(c) Consider the unbalanced experiment described in slide set 8. Assume the full model given on slide 6. Provide a matrix $\boldsymbol{C}$ for testing the main effect of temperature.

(d) Provide a matrix $\boldsymbol{A}$ so that

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{C} \end{bmatrix}$$

is a 4×4 matrix of rank 4.

(e) Provide a model matrix for a reduced model that corresponds to the null hypothesis of no temperature main effect.

(f) Find the error sum of squares for the reduced and full models.

(g) Find the degrees of freedom associated with the sums of squares in part (f).

(h) Compute the F-statistic for testing the null hypothesis of no temperature main effect using the sums of squares and degrees of freedom computed in parts (f) and (g).

3. Consider the dataset `pigs` provided in the `R` package `emmeans`. The data can be accessed in `R` with the following commands.

```
install.packages("emmeans")
library(emmeans)
pigs
```

To learn a more about the data, type `?pigs` at the `R` prompt.

For the purposes of this problem, use the natural logarithm of the variable `conc` as the response. Consider both `source` and `percent` as categorical factors. Assume the cell-means model with one unrestricted treatment mean for each combination of `source` and `percent`.

(a) Generate an ANOVA table with Type I (sequential) sums of squares for `source`, `percent`, `source` × `percent`, `error`, and `corrected total`. In addition to sums of squares, your ANOVA table should include degrees of freedom, mean squares, F statistics, and p-values where appropriate.

(b) Generate an ANOVA table with Type II sums of squares for `source`, `percent`, `source` × `percent`, `error`, and `corrected total`. In addition to sums of squares, your ANOVA table should include degrees of freedom, mean squares, F statistics, and p-values where appropriate.

(c) Generate an ANOVA table with Type III sums of squares for `source`, `percent`, `source` × `percent`, `error`, and `corrected total`. In addition to sums of squares, your ANOVA table should include degrees of freedom, mean squares, F statistics, and p-values where appropriate.

(d) Find LSMeans for `source` and `percent`.

(e) Consider simplifying the model so that `percent` is treated like a quantitative variable with linear effects on log `conc` and linear interactions; i.e.,

```
lm(y ~ source + percent + source:percent),
```

where `y=log(conc)` and `percent` is numeric. Does such a model fit adequately relative to the cell-means model? Conduct a lack of fit test and report the results.

(f) The reduced model fit in part (e) implies that, for each source, there is a linear relationship between the expected log concentration and percentage. Based on the fit of the reduced model in part (e), provide the estimated linear relationship for each source.