Stat 5100 **Assignment 6**

Due: Friday, March 14th 11:59PM in gradescope.

Problem 1 *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` $\times$ `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` $\times$ `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` $\times$ `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.*

Problem 2 *Consider the plant density example discussed in slide set 6.*

a) *For each of the tests in the ANOVA table on slide 38, provide a vector* $\mathbf{c}$ *so that a test of* $H_0 : \boldsymbol{c}^\top \boldsymbol{\beta} = 0$ *would yield the same statistic and p-value as the ANOVA test. (You can use R to help you with the computations like we did on slides 45 and 46 of slide set 6.) Label these vectors* $\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3$, *and* $\boldsymbol{c}_4$ *for the linear, quadratic, cubic, and quartic tests, respectively.*

b) Are $\mathbf{c}_1^\top \boldsymbol{\beta}$, $\mathbf{c}_2^\top \boldsymbol{\beta}$, $\mathbf{c}_3^\top \boldsymbol{\beta}$, and $\mathbf{c}_4^\top \boldsymbol{\beta}$ contrasts? Explain.

c) Are $\mathbf{c}_1^\top \boldsymbol{\beta}$, $\mathbf{c}_2^\top \boldsymbol{\beta}$, $\mathbf{c}_3^\top \boldsymbol{\beta}$, and $\mathbf{c}_4^\top \boldsymbol{\beta}$ orthogonal? Explain.

Problem 3 *Suppose $\boldsymbol{H}$ is a symmetric matrix. Prove that $\boldsymbol{H}$ is nonnegative definite if and only if all its eigenvalues are nonnegative. (If you wish, you may use the Spectral Decomposition Theorem in your proof.)*