PS6

Sam Olson

Q1

See Nettleton, HW7, Q1

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

```
library(emmeans)
```

```
## Warning: package 'emmeans' was built under R version 4.4.3
```

```
## Welcome to emmeans.
## Caution: You lose important information if you filter this package's results.
## See '? untidy'
```

```
pigs
```

```
##    source percent conc
## 1    fish       9 27.8
## 2    fish       9 23.7
## 3    fish      12 31.5
## 4    fish      12 28.5
## 5    fish      12 32.8
## 6    fish      15 34.0
## 7    fish      15 28.3
## 8    fish      18 30.6
## 9    fish      18 32.7
## 10   fish      18 33.7
## 11    soy       9 39.3
## 12    soy       9 34.8
## 13    soy       9 29.8
## 14    soy      12 39.8
## 15    soy      12 40.0
## 16    soy      12 39.1
## 17    soy      15 38.5
## 18    soy      15 39.2
## 19    soy      15 40.0
## 20    soy      18 42.9
## 21   skim       9 40.6
## 22   skim       9 31.0
## 23   skim       9 34.6
## 24   skim      12 42.9
## 25   skim      12 50.1
```

```
## 26    skim       12 37.4
## 27    skim       15 59.5
## 28    skim       15 41.4
## 29    skim       18 59.8
```

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.,

```r
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`.

Q2

See Nettleton, HW6, Q1

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 c so that a test of

$$H_0 : c^T\beta = 0$$

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 c_1, c_2, c_3, and c_4 for the linear, quadratic, cubic, and quartic tests, respectively.

b)

Are $c_1^T\beta$, $c_2^T\beta$, $c_3^T\beta$, and $c_4^T\beta$ contrasts? Explain.

c)

Are $c_1^T\beta$, $c_2^T\beta$, $c_3^T\beta$, and $c_4^T\beta$ orthogonal? Explain.

Q3

See Nettleton, HW6, Q2

Suppose H is a symmetric matrix. Prove that 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.)