Spring 2018, Stat 850: Midterm exam

3/12/2018

Attention: This is an open notes exam. To get full credit, you need to show your work and do the actual calculations with the information provided in the questions. Simply writing down the formulas for various quantities will not get you any partial credits.

1. **(8 pts)** A group of researchers at UW-Madison were interested in studying the effects of diet and voluntary exercise on the health of mice. They considered two diets (labeled D_1 and D_2) and assigned these to 8 mice using a completely randomized design with 4 mice for each diet. The 8 mice were randomly assigned to 8 individual cages. Within each diet group, 2 of the 4 mice were randomly assigned to cages without running wheels, and the other 2 mice were assigned to cages with running wheels. The running wheels permit mice to exercise, while exercise is very limited in cages without running wheels. After mice spent two months on the assigned diet in the assigned cage, a measure of overall health was recorded for each mouse. Let y_{ijk} denote the measure of health for the kth mouse on diet D_i and exercise treatment j, where $j = 1$ indicates no running wheel access and $j = 2$ indicates running wheel access. The researchers assumed the model $\mathbf{y} = \mathbf{X}\beta + \epsilon$, $\epsilon \sim \mathcal{N}(0, \sigma^2\mathbf{I})$ to be appropriate for these data, where

$$\mathbf{y} = \begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \end{bmatrix}.$$

   In this model, β_1, β_2, and β_3 are unknown real-valued parameters and σ^2 is some unknown real valued parameter.

   Answer the following questions based on this model, simplify your answers as much as possible. You will not get any credits for just writing down *generic* formulas.

   (a) What is the mean response of a mouse that received diet 2 and was housed in a cage without a running wheel.

   (b) Estimate β_1, β_2, and β_3 using least squares and express your solution in terms of y_{ijk}s and their marginal means using the . notation we used in class.

   (c) Find the best linear unbiased estimator of the mean response of a mouse that received diet 2 and was housed in a cage without a running wheel.

   (d) Write the null hypothesis of no diet main effect in terms of model parameters.

   (e) Find matrix $\mathbf{A}$ and vector c such that $\mathbf{A}\beta = c$ represents the null hypothesis of no exercise main effect.

2. **(10 pts)** Consider a completely randomized experiment in which a total of 10 rats were randomly assigned to 5 treatment groups with 2 rats in each treatment group. Suppose the different treatments correspond to different doses of a drug in milliliters per gram of body weight as indicated in the following table.

Treatment	1	2	3	4	5
Dose of drug (ml/g)	0	2	4	8	16

Suppose for $i = 1, \cdots, 5$ and $j = 1, 2$, y_{ij} is the weight of the jth rat from the ith treatment group at the end of the study. Furthermore, suppose

$$y_{ij} = \mu_i + \epsilon_{ij},$$

where $\mu_1, \cdots, \mu_5$ are unknown parameters and the $\epsilon_{ij} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$ for some unknown $\sigma^2 > 0$. Use the R code and partial output provided below to answer the following questions.

(a) Provide an estimate of μ_1.

(b) Provide an estimate of μ_2.

(c) Determine the standard error of your esimate of μ_2 from part (b).

(d) Conduct a test of $H_0 : \mu_1 = \mu_2$. Provide a test statistics, the distribution of the test statistic (be precise), a p-value, and a conclusion.

(e) Provide an F-statistic for testing $H_0 : \mu_3 = \mu_4$.

(f) The researchers would like to consider a simple linear regression model with body weight as a response and dose as a quantitative explanatory variable (m2 in the output). Provide a matrix $\mathbf{A}$ and a vector c such that the null hypothesis corresponding to the model in m2 may be written as $H_0 : \mathbf{A}\beta = c$, where $\beta = (\mu_1, \cdots, \mu_5)^T$.

```
d <- rep(c(0, 2, 4, 8, 16), each = 2)
#y is the data vector with entries ordered to appropriately
#match the vector d.
dose <- factor(d)
m1 <- lm(y ~ dose)
```

```
summary(m1)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  348.007      5.945  58.535 2.75e-08 ***
dose2         -4.589      8.408  -0.546 0.608669
dose4         -6.217      8.408  -0.739 0.492901
dose8         -9.491      8.408  -1.129 0.310182
dose16       -66.791      8.408  -7.944 0.000509 ***
---

anova(m1)
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value   Pr(>F)
dose         6187.8
Residuals     353.5
---
```

```
is.numeric(d)
```

```
## [1] TRUE
```

```
m2 <- lm(y~d)
```

```
anova(m2)
Analysis of Variance Table
```

```
Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
d              5380.8
Residuals      1160.5
---
```

3 **(4 pts)** Suppose that I have planned an experiment with three treatments, 20 units per treatment, and anticipated error standard deviation $\sigma = 10$. Will my power increase more if I spend money to double my sample size (to 40 units per treatment), or spend money to halve my σ to 5? Explain your answer.

4 **(8 pts)** Researchers conducted an experiment to compare the effectiveness of several commercial weight-loss plans. Twenty five subjects were recruited for the study. They were divided at random into five groups of $n = 5$ subjects each. The first four groups were assigned to follow Plans A, B, C, and D, and the fifth group was not assigned to any plan. After three weeks, the weight loss of each subject was recorded. The results are summarized below.

	n_i	$y_{i.}$	s_i^2
Plan A	5	$y_{1.}$	2
Plan B	5	$y_{2.}$	2
Plan C	5	$y_{3.}$	1
Plan D	5	$y_{4.}$	2
No plan	5	$y_{5.}$	3

(a) Find error sum of squares and mean error sum of squares.

(b) Suppose that the largest group mean is $y_{4.} = 3.0$ and the smallest is $y_{2.} = -1$. Consider the contrast $L = \mu_4 - \mu_2$ and test $H_0 : L = 0$. Show your calculations and explicitly state your conclusion.

(c) Assess the significance of the contrast in part (b) by Scheffe's method.

(d) Assess the significance of the contrast in part (b) by Tukey's method.

(e) Assess the significance of the contrast in part (b) by the Bonferroni method, considering it to be the largest in the family of

$$\binom{5}{2} = 10$$

possible pairwise comparisons.

(f) Which of the three procedures (Scheffee, Tukey, or Bonferroni) is the most powerful one in this example? Provide a clear argument with support from the data.