

## STAT 191 PROBLEM SET 2

Due date: Monday, Feb. 12

- (1) In one-way ANOVA, the model is:

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad \epsilon_{i,j} \sim N(0, \sigma^2).$$

The mean sum-of-treatment-squares term is

$$MSTR = \sum_{i=1}^r n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2 / (r - 1).$$

Assume that  $n_i = n$ , i.e. each level has  $n$  observations. Show the following:

- (a) For each  $i$ ,

$$\bar{Y}_{i.} - \bar{Y}_{..} = \alpha_i + (\bar{\epsilon}_{i.} - \bar{\epsilon}_{..}),$$

where  $\bar{\epsilon}_{i.} = \frac{1}{n} \sum_{j=1}^n \epsilon_{ij}$  and  $\bar{\epsilon}_{..} = \frac{1}{nr} \sum_{i=1}^r \sum_{j=1}^n \epsilon_{ij}$

- (b)  $\sum_{i=1}^r (\bar{Y}_{i.} - \bar{Y}_{..})^2 = \sum \alpha_i^2 + \sum (\bar{\epsilon}_{i.} - \bar{\epsilon}_{..})^2 + 2 \sum \alpha_i (\bar{\epsilon}_{i.} - \bar{\epsilon}_{..}),$   
 (c)  $E[2 \sum \alpha_i (\bar{\epsilon}_{i.} - \bar{\epsilon}_{..})] = 0,$   
 (d) Use (a-c), and the fact

$$E \left[ \sum (\bar{\epsilon}_{i.} - \bar{\epsilon}_{..})^2 \right] = \frac{(r-1)\sigma^2}{n},$$

to show that

$$E(MSTR) = \sigma^2 + \frac{n \sum_{i=1}^r \alpha_i^2}{r-1}.$$

- (2) In the surgery rehab data of lecture 8 (data file Rehab.txt), design and test the hypothesis that the recovery times of patients in the below and above average fitness group have the same absolute deviation from the average group. Write out explicitly the null hypothesis and the way you conducted the test.
- (3) A research studied the sodium content in lager beer by selecting at random six brands from the large number of brands of U.S. and Canadian beers sold in a metropolitan area. The researcher then chose eight 12-ounce cans or bottles of each selected brand at random from retail outlets in the area and measured the sodium content (in milligrams) of each can or bottle. Assume a random effects ANOVA model.

- (a) Test whether or not the mean sodium content is the same in all brands sold in the metropolitan area; use  $\alpha = 0.1$ . State the null and alternative hypothesis, and show the ANOVA table. What is the P-value of the test?
  - (b) Estimate the mean sodium content for all brands; give a 99 percent confidence interval.  
(data file Beer.txt)
- (4) RABE Exercise 5.7 (Data file: Election.txt)
  - (5) RABE Exercise 6.3 (Data file: Election.txt)
  - (6) RABE Exercise 6.5 (Data file: Megabyte.txt)

*RABE: Regression Analysis by Example by Chatterjee and Hadi, Ed. 4.*