# Homework 7 (60 pts)

## Problem 1 (16 pts)

(This is a conceptual question, no need to compute) Consider a model to predict the average length of a physician's stay based on the four regions. The one-way ANOVA model $\log_{10} Y \sim \text{Region}$ can be represented with different factor effects model, depending on the baseline. Hence, the mean responses, $\mu_1, \mu_2, \mu_3$ and $\mu_4$ are the means computed on $\log_{10} Y$.

Build a model summary using the first region (1) as the baseline, the R output shows table as the following. The values in the corresponding cells are denoted by $A_i, i = 1 \text{ to } 16$.

Table 1

| Coefficients: | | | | |
|---|---|---|---|---|
| | Estimate | Std. Error | t value | Pr(>\|t\|) |
| (Intercept) | $A_1$ | $A_5$ | $A_9$ | $A_{13}$ |
| Region 2 | $A_2$ | $A_6$ | $A_{10}$ | $A_{14}$ |
| Region 3 | $A_3$ | $A_7$ | $A_{11}$ | $A_{15}$ |
| Region 4 | $A_4$ | $A_8$ | $A_{12}$ | $A_{16}$ |

Next, build a model summary using the second region (2) as the baseline, the R output shows table as the following. The values in the corresponding cells are denoted by $B_i, i = 1 \text{ to } 16$. Hint: use the relevel() function.

Table 2

| Coefficients: | | | | |
|---|---|---|---|---|
| | Estimate | Std. Error | t value | Pr(>\|t\|) |
| (Intercept) | $B_1$ | $B_5$ | $B_9$ | $B_{13}$ |
| Region 1 | $B_2$ | $B_6$ | $B_{10}$ | $B_{14}$ |
| Region 3 | $B_3$ | $B_7$ | $B_{11}$ | $B_{15}$ |
| Region 4 | $B_4$ | $B_8$ | $B_{12}$ | $B_{16}$ |

Fill in the blanks with $A_i$, $B_i$, or both.

(a) [4 pts] Based on Table 1, The mean, $\mu_1, \mu_2. \mu_3, \mu_4$ can be represented as _____, _____, _____, _____

(b) [4 pts] Based on Table 2, The mean, $\mu_1, \mu_2. \mu_3, \mu_4$ can be represented as _____, _____, _____, _____

(c) [2 pts] To perform the hypothesis: $H_0$: $\mu_2 = \mu_1$ , the point estimate $\hat{Y}_2 - \hat{Y}_1 =$_____, the p.value is _____.

(d) [2 pts] To perform the hypothesis: $H_0$: $\mu_3 = \mu_1$ , the point estimate $\hat{Y}_3 - \hat{Y}_1 =$_____, the p.value is _____.

(e) [2 pts] To perform the hypothesis: $H_0: \mu_3 = \mu_2$, the point estimate $\hat{Y}_3 - \hat{Y}_2 =$_____, the p.value is _____.

## Problem 2 (10 pts)

According to the following data summary, compute the following confidence intervals.

| Group | Group Mean | N |
|---|---|---|
| 1 | 25 | 108 |
| 2 | 46 | 103 |
| 3 | 18 | 152 |
| 4 | 62 | 77 |
| MSE | 100 | |

(a) [5 pts] The difference between the average $Y$ in Groups 1 and 2 and in Groups 3 and 4, i.e.,

$$\frac{(\mu_1 + \mu_2)}{2} - \frac{(\mu_3 + \mu_4)}{2}$$

(b) [5 pts] The simultaneous confidence intervals for the differences of the following:

$$\mu_1 - \mu_2, \qquad \mu_1 - \mu_3, \qquad \mu_2 - \mu_3$$

Problems 3 and 4 are both based on the following scenario:

You are part of a team that is investigating the effectiveness of two different ingredients in a drug. For the first ingredient (factor A), you test three different amounts, 10mL, 20mL, and 30mL. For the second ingredient (factor B), you test 2 different amounts, 10mL and 20 mL. With three levels of the first ingredient and two levels of the second ingredient, you can create six different formulations for the drug. You obtain a random sample of 60 individuals. For each of the six formulations, you randomly select 10 individuals to receive that formulation of the drug and then collect a response from each individual.

## Problem 3 (14 pts)

(a) [4 pts] Write down the two-way ANOVA model that would be used for this experiment. Use the factor-effects notation for the model, not the cell-means notation. Define each component of the model in context of the problem. For example, don't just refer to factor A or factor B, but refer to drugs themselves. Also be sure to include information about the number of levels for each factor and the number of replicates.

(b) [5 pts] Pretend the table below contains information about the sample means for each of the six formulations of the drug. Draw an interaction plot for these data. For these data, does there appear to be an interaction? Why or why not?

| | | Drug B | |
|---|---|---|---|
| | | *10 mL* | *20 mL* |
| Drug A | *10 mL* | $\bar{y}_{11} = 10$ | $\bar{y}_{12} = 40$ |
| | *20 mL* | $\bar{y}_{21} = 25$ | $\bar{y}_{22} = 50$ |
| | *30 mL* | $\bar{y}_{31} = 30$ | $\bar{y}_{32} = 55$ |

(c) [5 pts] Regardless of your answer to part (b), assume that the interaction between drugs A and B is not important and can be ignored. You are interested in comparing the average of levels 2 (20 mL) and 3 (30 mL) of Drug A with level 1 of drug A (10 mL) to see if there is a significant difference between them. That is, you are interested in the following contrast:

$$L = \frac{\mu_{2\cdot} + \mu_{3\cdot}}{2} - \mu_{1\cdot}$$

(i) What is the point estimate for this contrast, $\hat{L}$?

(ii) What is the confidence interval for this contrast? Is there a significant difference? (Assume that the MSE is 60)

## Problem 4 (20 pts)

Below is an example of what the ANOVA table could look like for this experiment:

| Source | $df$ | $SS$ | $MS$ | $F$ | p − value |
|---|---|---|---|---|---|
| A | | 4290.30 | | | |
| B | | 12367.30 | | | |
| A ∗ B | | 23.20 | | | |
| Error | | 3833.70 | | | |
| Corrected Total | | 20514.5 | | | |

(a) [15 pts] Fill in the spots in the ANOVA table that are missing with the correct values.

(b) [5 pts] Based on the information in the ANOVA table, does the interaction appear to be significant?