Problem Set 5 (two pages)

Statistics 24510-30040 (W19)

Due Tuesday, Feb. 19, at the beginning of class.

Requirements Provide detailed derivations. Select only the relevant part of the output to be inserted. Attach your code or output as an appendix if necessary. Discussions allowed, the assignment should be devised and written by yourself completely.

Problem assignments (Relevant text: Rice chapter 12 on two-way ANOVA and multiple comparison.)

1. (*Two-way ANOVA interaction effect*)

   Below is part of an ANOVA table for the Two-way ANOVA model with interaction.

```
Analysis of Variance
Response: Survival
              Df Sum Sq   Mean Sq   F value   Pr(>F)
Poison         1  93.161
Method         1   8.251
Poison:Method  1   0.012
Residuals     44 196.960
```

   (a) Complete the partial ANOVA table. additive model: means no interaction

   (b) Derive the ANOVA table for the additive Two-way ANOVA model (for the same data). Do the p-values differ from the corresponding p-values in the model in(a)? Explain why.

2. (*Two-way vs. One-way ANOVA*)

   A Baseball League changed the strike Zone each year for three years. The data below show the number of Home Runs that each of four players hit for the three years.

Y_{ij}	Player 1	Player 2	Player 3	Player 4
Zone 1	23	21	31	15
Zone 2	12	8	25	10
Zone 3	18	15	30	15

   You may input the data to R and conduct the analysis below.

   (a) Perform a one-way analysis of variance for the Player effect on Home Runs. Is there a significant Player effect by the analysis?

   (b) Perform a one-way analysis of variance for the Zone effect on Home Runs. Is there a significant Zone effect by the analysis?

   (c) Perform a two-way analysis of variance for the additive effects of Player and Zone on Home Runs. Is there a significant Player effect or Zone effect by the analysis?

   (d) Comment on the reasons of different conclusions from the above analyses.

   (e) Can we use a two-way analysis of variance to test the interaction effect of Player and Zone? Comment.

3. (*Bivariate normal properties*)

Suppose the natural logarithms of the heights of mothers and daughters (in centimeters) from a region follow a bivariate normal distribution with the mean for the mother of log(160), the mean for the daughters is log(165), the standard deviations for both distributions is 0.05 and the correlation is 0.5.

   (a) Find the probability that a mother is taller than her daughter.
   (b) Find the probability that a mother is at least 90% as tall as her daughter.
   (c) Give the joint density for the logarithms of the mother's and daughter's heights when heights is measured in inches (1 inch = 2.54 cm).
   (d) Find the density of the ratio of a mother's height to the daughter's height.

4. (*Bivariate normal probability calculation*)

Continuation of the above problem. Consider finding the probability that a daughter is at least 10 cm taller than her mother. It is not possible to give an exact answer fo this probability in terms of any standard special functions. Using R, come up with a way to approximate this probability.

5. (*Pairwise comparison, t-test vs. F-test*)

Input the data at https://www.stat.uchicago.edu/ meiwang/courses/w19stat2/multitest-Stat2.txt.
You may use the R command `mydata = read.table("multitest-Stat2.txt")`

The dataset contains (transformed) measurements from four (column) treatment groups. For this exercise, regard the treatment groups as independent, and the ith group as an i.i.d. sample from $N(\mu_i, \sigma^2)$.

   (a) Using R, perform all two-sample t-test for the hypotheses H_{ij}: $\mu_i = \mu_j, 1 \le i < j \le 4$.
   Which hypotheses are rejected at test level $\alpha = 0.1$?
   (b) Now make a Bonferroni adjustment to the test level in order to control the probability that at least one null hypothesis in (a) is falsely rejected at test level 0.1. Which of the tests in (a) would reject their null hypotheses if this Bonferoni-adjusted test level is employed?
   (c) In order to test the hypothesis H_o: $\mu_1 = \mu_2 = \mu_3 = \mu_4$, consider the test statistic

$$T = \max\{|T_{ij}|, 1 \le i < j \le 4\},$$

   where T_{ij} is the test statistic from the hypothesis test H_{ij}: $\mu_i = \mu_j$ in (a).
   The idea is to reject H_o if T is too large.
   Give a simple upper bound on the p-value of the test based on T. Show your derivations.
   (d) Does the upper bound in (c) allow you to reject H_o at test level $\alpha = 0.1$? In other words, is the upper bound smaller than 0.1?
   (e) Perform the F-test from the one-way analysis of variance to test H_o in (c).
   Do you reject H_o at test level $\alpha = 0.1$?

R remarks: The following R commands converts the data to the vector form as the example in the handout.

```
y=as.vector(as.matrix(mydata))
```

You may need to create a factor vector for analysis of variance.