BIOS 830: Homework 2

Due on February, 23 2017

February 7, 2017

Instructions: Students are encouraged to work together on this problem set. However, each student is expected to *independently* write up the assigned problems. Please provide the code Questions 3 and and 4 in a clean and readable document; points will be deducted for "messy code". **Assignments are to be turned in at the beginning of lecture on the due date above. Any assignments not turned in at this time will be considered late.**

Question 1: (*The following is a fabricated example and does not reflect the true efficacy of the drugs listed*)
The data given in "Cholesterol.csv" were collected from a study comparing the effectiveness of five different treatments for treating high cholesterol and include: the standard treatment (Treatment 1) and four new or experimental treatments (Treatments 2-5). Briefly, 100 subjects with high cholesterol, ranging in age from 45 to 65 years old, were enrolled in the study and were randomly assigned to one of the five treatments. Treatment was administered for a total of 16-weeks, at which cholesterol (mg/dL) was measured for each subject.

Treatment	Generic names	Brand Names	How they work
1	rosuvastatin with amlodipine	Crestor	Raise good (HDL) cholesterol and lower bad (LDL) cholesterol
2	atorvastatin with amlodipine	Caduet	Lower how much cholesterol your body makes and lower blood pressure
3	lovastatin with niacin	Advicor	Raise good (HDL) cholesterol and lower bad (LDL) cholesterol
4	simvastatin with ezetimibe	Vytorin	Lower how much cholesterol your body makes and the amount of cholesterol your body absorbs
5	simvastatin with niacin	Simcor	Raise good (HDL) cholesterol and lower bad (LDL) cholesterol

In addition to determining whether there is a difference in cholesterol between any of the five treatment groups following the 16-week treatment period, the investigators are also interested in the following two-sided preplanned contrasts:

- Experimental drugs versus the standard treatment

- Drugs with niacin versus non-niacin containing drugs among only the experimental treatments.
- Drugs with simvastatin versus non-simvastatin containing drugs among only the experimental treatments.

(a) Is this study adequately powered ($\geq 80\%$ power) to be able to detect a difference of 10 mg/dL between at least two of the treatments at a type 1 error rate of $\alpha = 0.05$?

(b) Using a one-way ANOVA model, test the null hypothesis that $H_0 : \mu_1 = \mu_2 = \ldots = \mu_5$, where $\mu_1, \mu_2, \ldots, \mu_5$ reflect the individual treatment-specific means. Does there appear to be a difference in cholesterol between any of the five treatment groups following the 16-week treatment period?

(c) Create 95% simultaneous confidence intervals for the three preplanned contrasts using an appropriate multiple comparison method (justify your choice). What do you conclude?

(d) Suppose that after the investigators had a chance look through the data they decided that they were interested in creating 95% simultaneous confidence intervals all pairwise contrasts. What multiple comparison method would you recommend and why? What is the *critical coefficient* equal to in this scenario for the method you selected?

(e) What is the overall *experiment-wise* error rate for this study across the two sets of contrasts (c and d)? Since the investigators want to maintain an overall overall *experiment-wise* error rate of $\alpha = 0.05$, readjust the simultaneous confidence intervals computed in part (c) to satisfy this requirement.

Question 2:

Consider a CRD on the quantitative treatment factor X with levels 24, 28, 32, and 36. We are interested in the linear contrast $L = c_1 E(Y_{ij}|X_{ij} = 24) + c_2 E(Y_{ij}|X_{ij} = 28) + \ldots c_4 E(Y_{ij}|X_{ij} = 36)$ with contrast coefficients $\mathbf{c}' = (-3, -1, 1, 3)$.

Using the simple linear regression model,

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \epsilon_{ij}, \;\; i = 1, ..., 4 \;\; j = 1, ..., n_i$$

write L as a function of β_1 and show that a test of $L = 0$ is equivalent to a test of $\beta_1 = 0$.

Question 3:
Based on the iterative method described in lecture, create your own R or SAS function for calculating the sample size for a **one-way ANOVA model**. Compare the output of your function to the "built in" sample size functions in R (pwr.anova.test()) or SAS (SAS-PROC POWER) to check for consistency. Please provide the code for your function as well as the output showing consistency with one of the above "built in" functions.

Question 4:
Consider a CRD with a single treatment factor consisting of two levels, whose sample sizes

are n_1 and n_2 respectively. Thus $N = n_1 + n_2$ represents the total number of experimental units planned for this study. Assume the following model:

$$Y_{ij} = \mu + \tau_i + \epsilon, \ i = 1, 2 \ j = 1, ..., n_i,$$

where $\epsilon_{ij} \overset{IID}{\sim} N(0, \sigma^2)$. For a fixed N, empirically demonstrate that statistical power is maximized in a balanced CRD, i.e., $n_1 = n_2$ and that power decreases decreases as a function of increasing unbalancedness. *Hint, the following items may be useful for this problem*:

- `pwr.t2n.test` function in the R-package "pwr"
- `rnorm` function in R for generating random samples from a normal distribution
- Be sure that N, μ_1, μ_2, and σ^2 are fixed as you assess statistical power for varying n_1 and n_2.

In addition to the above questions, question 7 from Dean & Voss Chapter 4 and questions 2, 3, and 8 from Dean & Voss Chapter 5.