5P8130 Fall 2019: Biostatistical Methods I

Homework 3

Due, Nov 1st @ 5:00pm

P8130 Guidelines for Submitting Homework

Your homework should be submitted only through CourseWorks. No email submissions!

All derivations, graphs, output and interpretations to each section of the problem(s) must be included in the PDF (not the code), otherwise it will not be graded.

Only 1 PDF file should be submitted. When derivations were required and handwriting was allowed, scan the derivations and merge ALL PDF files (http://www.pdfmerge.com/) into a single one.

We are encouraged to use R for calculations, but you still have to show the mathematical formulae. Also, make sure to also submit your commented code as a separate R/RMD file.

DO NOT FORGET:

You are encouraged to collectively look for answers, explain things to each other, and use questions to test each other knowledge.

But

Do NOT hand out answers to someone who has not done any work. Everyone ought to have ideas about the possible answers or at least some thoughts about how to probe the problem further. Write your own solutions!

Problem 1 can be handwritten - legibly – scanned and incorporated into the HW PDF.

Problem 1 (15p)

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables from $N(\mu, \sigma^2)$. Show that:

a) $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is an unbiased estimator of μ.
b) $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is an unbiased estimator of σ^2.
c) In the context of ANOVA model, prove the partitioning of the total variability (sum of squares), i.e.,

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{\bar{y}})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{y}_i - \bar{\bar{y}})^2$$

For all problems below, assume a significance level of 0.05 unless stated otherwise.
You can use R to generate descriptive statistic, run models, but you still need to type the mathematical formulations of the hypotheses, test statistic, and decision rule.
NO PICS will be allowed!

Problem 2 (20p)

Studies have shown that ascorbate might have a toxic effect to a variety of cancer cell lines The purpose of the study was to determine if the survival times of patients treated with ascorbate differ with respect to the organ affected by the cancer (stomach, bronchus, colon, ovary or breast cancer).

Note: this problem assumes complete information on all surviving patients (no deaths occurred).

a) Generate and present descriptive statistics in a tabular form for each group and comment on the differences observed. Add a graphical display of the survival times by group.
b) Using a type I error of 0.01, obtain the ANOVA table. State the hypotheses, decision rule and provide interpretation in the context of the problem.
c) Based on your response in part b), perform pairwise comparisons with the appropriate adjustments (Bonferroni, Tukey, and Dunnett – 'breast cancer' as reference). Report your findings and comment on the differences/similarities between these three methods.
d) One of the ANOVA assumptions is normality of the group data. Check the assumption using a method(s) of your choice.
   i) If normality is questionable, provide at least one solution to address this issue.
   ii) Implement the solution(s) from part i) and re-run the ANOVA analysis. Comment on the new findings.

A plot

Graph

One is KW;
another is transformation

Problem 3 (15p)

You propose a new trial testing the addition of zinc to the diet of HIV-positive children in South Africa. In various other populations, zinc and other micronutrients have been found to reduce the occurrence of diarrhea, which is associated with immune system problems, as well as to slow the progress of HIV. Collaborators have shared the following articles/results with you and asked your help with sample size calculations.

Treatment	Sample size	Avg. # episodes in a year $\pm$ s.e.
placebo	56	1.1 ± 0.2
iron	54	1.4 ± 0.2
zinc	54	0.7 ± 0.1
zinc + iron	55	0.8 ± 0.1

For the new trial you will combine and consider the iron vs non-iron groups.

a) Estimate the averages and the estimated difference in the averages based of these two new groups. The estimated difference will be considered the 'true' effect size to be detected.
b) Next, calculate the standard deviations for the four initial groups and compare the values. Can you assume that the standard deviation is equal b/w 'zinc' and 'zinc-iron' and also b/w 'placebo' and 'iron' only? If so, use them as 'true' standard deviations for the new combined groups.
c) Using the values in a) + b), calculate the sample size required to attain 80% power at 0.05 alpha level for:
    i) Equal allocation b/w groups
    ii) 2:1 allocation b/w groups (larger in the non-iron groups).

Note: Sample sizes are always rounded up.