

Homework 4

Due, Nov 6 @ 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's 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 (5p):

In the context of ANOVA model, prove the partitioning of the total variability (sum of squares), i.e.,

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

For the problems below, you can use R to generate summary statistics and perform statistical tests; however, the hypotheses and critical values need to be written using mathematical formulations.

Problem 2 (15p)

A rehabilitation center is interested in examining the relationship between physical status before therapy ('below average', 'average' and 'above average') and the time (days) required in physical therapy until successful rehabilitation. Records from patients 18-30 years old were collected and provided to you for statistical analysis (dataset "Knee.csv").

Assuming that data are normally distributed, answer the questions below:

- Generate descriptive statistics for each group and comment on the differences observed. (4p)
- Using a type I error of 0.01, obtain the ANOVA table. State the hypotheses, test statistic, critical value, and decision interpreted in the context of the problem. (5p)
- Based on your response in part b), perform pairwise comparisons with the appropriate adjustments (Bonferroni, Tukey, and Dunnett – 'below average' as reference). Report your findings and comment on the differences/similarities between these three methods. (5p)
- Write a short paragraph summarizing your overall results as if you were presenting to the rehabilitation center director. (1p)

Problem 3 (10p)

A research article was published with the following headline “For adults, chicken pox vaccine may stop shingles”. The findings were based on a randomized clinical trial with a total of 420 adults being randomized to receive either chicken pox vaccine or placebo. While the results were intriguing, some side effects emerged and required further investigation. The table below summarizes the frequencies of one of the most frequent and concerning side effect - swelling around the injection site.

	Major Swelling	Minor Swelling	No swelling
Vaccine	54	42	134
Placebo	16	32	142

Use a significance level of 0.05 to assess if the distribution of swelling status is the same for the two treatment populations.

- Justify the appropriate test to be used for addressing the question of interest. (2p)
- Provide the table with all values necessary for calculating the test statistic. (4p)
- State the hypotheses, test statistic, critical value, p value and decision rule interpreted in the context of the problem. (4p)