



2017 Spring
EDMS 646: General Linear Models I
1121 Benjamin Building, Thursdays: 4:15-7:00pm
Assignment 1
(Due: Feb 9, beginning of the class)

Answer each question synthesizing the information from the lecture slides, textbook, and other resources. When appropriate, insert statistical output (and input if need be) to justify each of your answers. Answers to the questions must be word-processed using Word and Microsoft Equation Editor (an object to be inserted within Word) for statistical/mathematical notation. Answers to the homework questions should appear on 8.5" x 11" paper (not computer output) and must be legible. Students with clarifying questions about the homework should contact Dr. Yang directly. The homework grading scheme was explained during the first class and also written in the course syllabus.

PART I: Complete the commute time data analysis

For this portion of assignment, you use the provided data that were collected in EDMS 646 class. Answer the following questions and do not forget to show your work.

1. Using the population data, get descriptive statistics and report them. Use appropriate symbols when you report mean and variance. For this question, you may use any statistical software.
2. Using the sample data, answer the questions below. The proposed research question and hypothesis are as follows.

How long is the average commuting time for the students who are in EDMS 646?
It will be about 30 minutes.

a) State the null and alternative hypotheses for a two-sided test.

First, assume that you do not know the population variance and mean. Answer each question below.

b) Find the test statistic's sampling distribution under H_0 .

Hint: You might want to use either an online resource or R to draw the distribution.

c) What are the critical statistics when alpha level is 0.05? (Provide your rationale, i.e., using a proper distribution table, using R or online resources and so on)

d) What is the observed statistic? Show your calculation.

e) What is the p-value? (Provide your rationale, i.e., a proper distribution table or using R and so on)

f) Calculate 95% confidence interval for the mean and interpret (explain) what it means. Show your calculation.

g) What is your statistical decision about the null hypothesis?

h) Confirm your answers for questions d), e), and f) using a statistical software. Copy and paste relevant output to show your work.

3. Now, assume that you do know the population variance (You can obtain the number from answers to Question 1) but do not know the population mean. Conduct an appropriate statistical test to make a decision on the set of hypotheses in question a). In other words, answer questions from b) to g) under this new scenario using the values you obtain from sample data. You do not need to confirm the answers using a statistical software for this part.

PART II: Two sample t-test review

The data set that you can find under HW1 package folder, *Hypertension*, consists of hypertension data for 72 subjects who were randomly assigned to one of two treatment groups (diet or no diet). The research question revolves around knowing if diet influences hypertension. The response variable (outcome) is a hypertension score where a higher score indicates more hypertension. For the group of people who are not on a diet, the variable *Diet* is coded as 1. For the group of people who are on a diet, the variable *Diet* is coded as 2. To answer the following questions, you should use a statistical software of your choice.

1. Do an exploratory data analysis. Comment on interesting facets of the data that you uncover paying particular attention to the shape of the distributions, an appropriate measure of central tendency for each group, the variability in each group, and if outliers are present. Implicit in this discussion is an acknowledgement of how well the response variable corresponds to the assumptions of the independent samples t-test.
2. Now, perform the appropriate t-test using hypertension as the dependent variable and diet as the grouping variable (independent). Be sure to write down H_0 , your decision rule, your conclusion, and to comment on internal validity. Use $\alpha = .05$. All of your results write up should adhere to the APA-Style Write-Up (Refer to the textbook, supplemental material, or other sources).

3. Check the assumptions underlying the t-test (model that you imposed on the data). Provide both graphical as well as numerical evidence to justify your position.