Week 10 - Discussion

Problem 1. We want to know the mean percentage of butterfat in milk produced by a farm by sampling multiple loads of milk. Previous records indicate the average percent butterfat in milk is 3.35 and the standard deviation among loads is 0.15. After this farm importing some new cows, we hope to detect if the percentage of butterfat in milk increases (we assume that the standard deviation does not change and the percentage of butterfat in milk obeys a normal distribution).

- a. How many loads do we need to sample so that the margin of error for a 90% confidence interval of the mean percent butterfat is no more than 0.06?
- b. Suppose 100 loads of milk are sampled. What is the power of the test for detecting a change of the mean to 3.40 under $\alpha = 0.05$?
- c. How large does the sample size need to be to make the power at least be 0.8? (under the same significance level)

Problem 2. Explore $lm(\cdot)$ function with *cars* dataset. (You can get this dataset without importing any data file or package and we assume all assumptions of performing a linear regression are satisfied)

- a. Use ?cars command to learn this dataset. What relationship you can imagine for these two variables
- b. Present a scatter plot of *speed* and *dist* variables of this dataset.
- c. Substantiate your finding from parts (a) and (b) by applying $cor(\cdot)$ function of R.
- d. Try to perform a linear regression using $lm(\cdot)$ function.
- e. Add the regression line which you got from part (d) to your scatter plot.