Homework 1

STAT:3210 Experimental Design and Analysis

Yubing Li

**1. Problem 1.3, pre-experimental planning:**

Step 1: Recognition and statement of problem:

The experiment intend to test how we can get the maximum growth of garden flowers by changing different conditions.

Step 2: Response variable:

The response variables of the experiment is the growth of the garden flowers. In general, it could be the size of the flowers, the color of the flowers, the number of the flowers that are alive, etc.

Step 3: Factors, levels and range:

The four factors are sunlight, water, fertilizer and soil condition.

Under sunlight, there could be two levels: low and high. The ranger is from low to high.

Under water, there could be two levels: water with minerals and water without minerals. The range would be no minerals per liter to 50 grams minerals per liter.

Under fertilizer, there’re many levels which depend on the different brands of fertilizer we use. We can label the different fertilizer “A”, “B”, “C”, etc. The range is various.

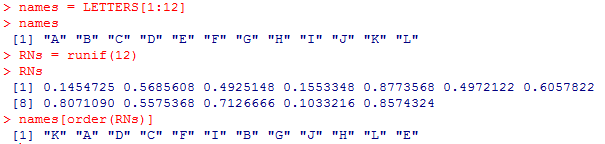
Under soil condition, there’re three levels: soil with pH less than 7, soil with pH equal to 7 and soil with pH larger than 7. The range is pH value from 0 to 14.

**2. Problem 1.9, why is randomization important:**

Randomization is one of the basic principles of experimental design. By randomization, we let the experimental material and the order of runs to be randomly determined.

Since the statistical methods require the observations to be independently distributed, using randomization could eliminate the bias in the experiment and make the assumption valid. Randomization ensures that the possibility of assign treatments to every experimental units are equal.

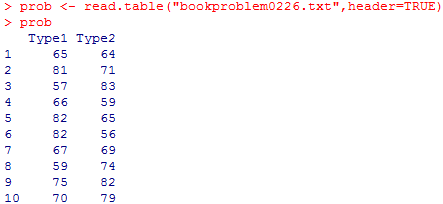
For example, people might have preference on which treatment he/she is assigned, if the experiment is not randomized and he/she could get whatever treatment to enter, he/she would probably give a bias response to that treatment.

**3. CRD:**

Assign K, A, D, C to trt1. Assign F, I, B, G to trt2. Assign to J, H, L, E to trt3.

**4. Basic principles of experimental design:**

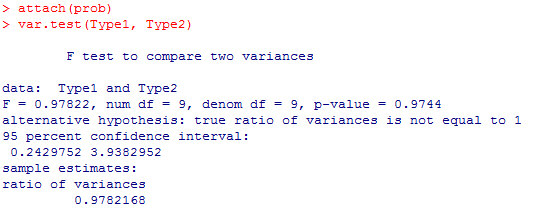
The three basic principles of experimental design are Randomization, Replication and Blocking.

**5. Problem 2.26:**

**(a) Test hypothesis of variance:**

H0: σ12 = σ22 vs. Ha: σ12 ≠ σ22

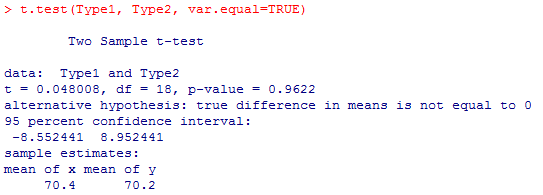
The P-value is 0.9744, which is larger than α = 0.05. Therefore, we say there’s insufficient evidence to say that the two variances are different. In a word, we fail to reject the null hypothesis that the two variances are equal.



**(b) Test hypothesis of mean (assuming variances are equal):**

Since in part (a) we can’t reject that the variances are equal, we assume equal variances and choose pooled t-test for comparing population means.

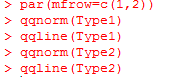
H0: µ12 = µ22 vs. Ha: µ12 ≠ µ22

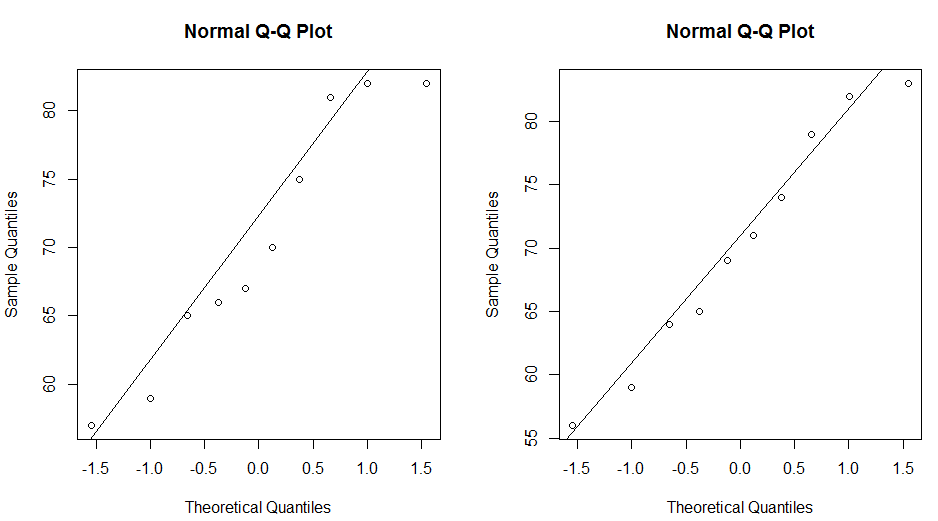


The P-value is 0.9622, which is larger than α level = 0.05. Therefore, we say there’s insufficient evidence to say that the two population mean are different. In a word, we fail to reject the null hypothesis that the two population mean are equal.

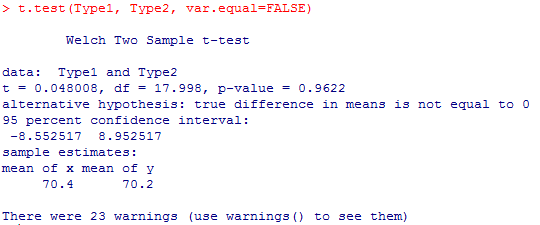
The 95% confidence interval for the population mean difference is (-8.5524, 8.9524). The confidence interval agrees to the result of the null hypothesis testing, indicates that the two population mean can be equal.

**(c) Normality assumption:**

The normality assumption is required for pooled t-test and f-test. In the normal QQ-plots, we can tell that the observations are generally close to the two diagonal lines. The two plots indicate that the burning time for the two treatments follow the normal distribution.



**(d) Test hypothesis of mean (assuming variances are unequal):**

H0: µ12 = µ22 vs. Ha: µ12 ≠ µ22

The P-value is 0.9622, which is larger than α level = 0.05. Therefore, we say there’s insufficient evidence to say that the two population mean are different. In a word, we fail to reject the null hypothesis that the two population mean are equal.

The 95% confidence interval for the population mean difference is (-8.5525, 8.9525). The confidence interval agrees to the result of the null hypothesis testing, indicates that the two population mean can be equal.

**(e) Side-by-side box plots:**

From the side-by-side boxplots, we can tell that the observations in type 1 and type 2 overlap a lot. Therefore, we can say that the two variances are probably equal. In addition, the median of the two samples are approximately the same. Since the two samples both follow the normal distribution, we can say that the mean of the two samples are also equal. The box plots give the information that are consistent to the data analysis result from this experiment.

