

R-Assignment-7

In the following assignment, do the following for each problem:

- (i) **Set up the null hypothesis and the alternative hypothesis.**
- (ii) **Use R to test the hypothesis. Include the R-code.**
- (iii) **Give your conclusion.**

Read pages 59-71 in "Using R for Introductory Statistics" by John Verzani and look at my lecture notes.)

1. The data set given in table 7-1 represents the treadwear measures of tires. The data is also provided in an excel-file, called Data7-1. Each tire was subject to measurement by two methods, the first based on weight loss and the second based on groove wear.

Reference: R.D. Stichler, G.G. Richey, and J. Mandel, "Measurement of Treadware of Commercial Tires, Rubber Age, 73:2 (May 1953) and

DASL at <http://lib.stat.cmu.edu/DASL/DataArchive.html>

(A) Use the **read.delim** command in R, to read in your data set given in Data7-1. (See examples in lecture notes from 4.4 or 4.5).

(B) Create two separate vectors from the table with measurements from the two methods. Call the vector with the data based on weight loss for X, and the vector with the data based on the groove wear method for Y. Do this by letting $X = Z[,1]$ and $Y = Z[,2]$, where Z is the data set that you read in. (See examples in lecture notes from 4.4 or 4.5).

(C) Draw a horizontal boxplot of each data set on the same axis.

(D) Is there a significant difference of the mean treadwear measure of tires subject to these two methods at the 5% significance level? Briefly explain why you can use the test of hypothesis that you are using.

(E) Describe how you would design such an experiment.

2. The data given in Data7-2 represents the waiting time between eruptions for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA for two different periods, one measured 8/1 to 8/1, 1985 and the other measured 8/6 to 8/10, 1985.

WT1 = Minutes between eruptions 8/1 to 8/1, 1985

WT2 = Minutes between eruptions, 8/6 to 8/10, 1985

Reference: A. Azzalini and A.W. Bowman, "A Look at Some Data on the Old Faithful Geyser," Applied Statistics, 1990, pp 57-365. and

DASL at <http://lib.stat.cmu.edu/DASL/DataArchive.html>

(A) Use the **read.delim** command in R, to read in your data set given in Data7-2.

(B) Create two separate vectors from the table with measurements from the two methods. Call the vector with the data measured, 8/1 to 8/1, 1985 for X, and the vector with the data measured 8/6 to 8/10, 1985 for Y.

(C) Are there significant difference in the variance at the 5% significance level for these two measurements? (You can do the test-statistics by hand, but use R to find the critical value.)

(D) Use R to determine a 90% confidence interval for the mean.

(E) Is there a difference in the mean of those two measurement at the 10% significance level?

3. The data in Data6-3 contain the outcomes for two items in the Montana Economic Outlook Poll conducted in May 1992, with accompanying demographics for 179 out of 418 poll respondents. The items are whether the respondents view the state economic outlook as

better over the next year. Respondents are classified by age.

AGE = 1 under 35,

Age= 2 35-54,

Age= 3 55 and over.

STAT = State economic outlook 1 better, 0 not better than a year ago

Source: Bureau of Business and Economic Research, University of Montana, May 1992 and DASL at <http://lib.stat.cmu.edu/DASL/DataArchive.html>

(A) Use the **read.delim** command in R, to read in your data set given in Data7-3.

Create three separate vectors from the table with the outlook from the three age group. Let X be the vector with the outlook from people in age group 1. Let Y be the vector with the outlook from people in age group 2 and let Z be the vector with the outlook from people in age group 3. (C) Use the **sum** command in R to count the number of people in each age group that view the state economic outlook as better over the next year.

(D) Is there a difference at the 5% significance level in the proportions of people in the age-group under 35 and age-group, 35-54, that view the state economic outlook as better over the next year?

(E) Is there a difference at the 5% significance level in the proportions of people in the age-group, 35-54, and age group 55 and over, that view the state economic outlook as better over the next year?

4. We are interested in testing whether the mean SAT scores from a high school, called high school A, located in the Midwest, and a high school, called high school B, located in the North East, differs. Suppose both distributions of scores follows a normal distribution. The SAT score of 40 randomly selected students from each high school is given in Data7-4. Is there a difference in the mean score at the 5% significance level ?