**MA684**

**Hongbo xue**

**Feb 2, 2015**

**HW 1**

Based on a problem from Ch. 13 of the KKM and N text. The following gives weight loss (in pounds) over a three week period for people initiating two different weight-loss programs:

Diet A: 21 , 33 , 35 , 42 , 36 , 34 , 13, 39, 40

Diet B: 24 , 20 , 12 , 9 , 31 , 34, 8, 12

Our ultimate goal is to compare the average weight loss between the two groups, to see if one program leads to significantly greater weight loss than the other. But first, some preliminary analyses:

1a) Calculate the mean and standard deviation for these two samples. (It is OK to do all of Question 1 using the computer – see Question 2 – but you should know how to do these calculations by hand as well.)

**Answer:**

**Diet A:**

**Mean = (x1+x2+….+xn)/n=32.56**

**Standard Deviation = 9.5**

**N=9**

**Diet B:**

**Mean = (x1+x2+….+xn)/n=18.75**

**Standard Deviation = 10.10**

**N=8**

1b) Calculate a 95% confidence interval for the average weight loss for people following Diet A. Interpret this interval.

**Answer:**

**We can find the critical value is 2.306 from the 95% confidence interval with n=8.**

**=32.56=(25.25,39.86)**

1c) Would a 99% confidence interval for the average weight loss for people following Diet A (calculated from this same sample) be wider or narrower than the above interval?

From R, we can find the 99% confidence interval will be wider.

1d) For our primary aim, find and interpret a p-value to compare the mean weight loss for those following Diet A vs. those following Diet B. What statistical method did you use to find this p-value?

**Answer:**

**A t-test is the appropriate statistical procedure to compare two means. Here, the two means are from independent samples, and the two independent sample version of the t-test is appropriate. The null hypothesis is that there is no difference between the mean times for the two routes. The (two-tailed) alternative hypothesis is that the mean time differs for the two routes. We can follow the formula to find:**

**meanA=32.56**

**meanB=18.75**

**sdA=9.5**

**sdB=10.10**

**nA=9**

**nB=8**

**Sp=70.41**

**Tabs=(32.56-18.75)/70.41\*(1/9+1/8)^(1/2)=0.4**

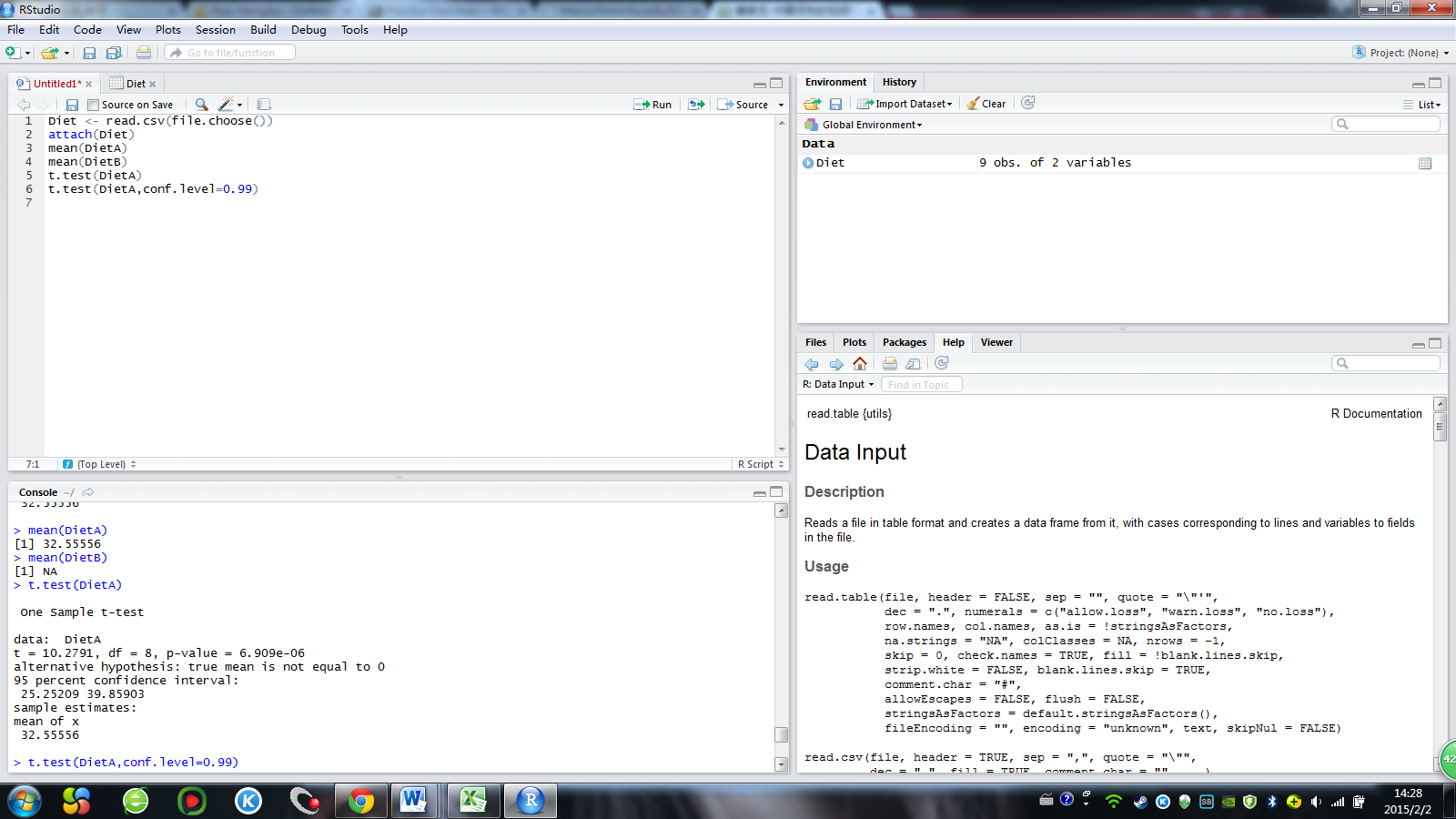
**To test at the (two-tailed) α=.05 level, the critical value with n1+n2-2 = 15 degrees of freedom is tcritical=2.131. Since tobs is not greater than tcritical, we cannot reject the null hypothesis. These data do not provide significant evidence to show a difference in the mean commuting times from the two routes.**

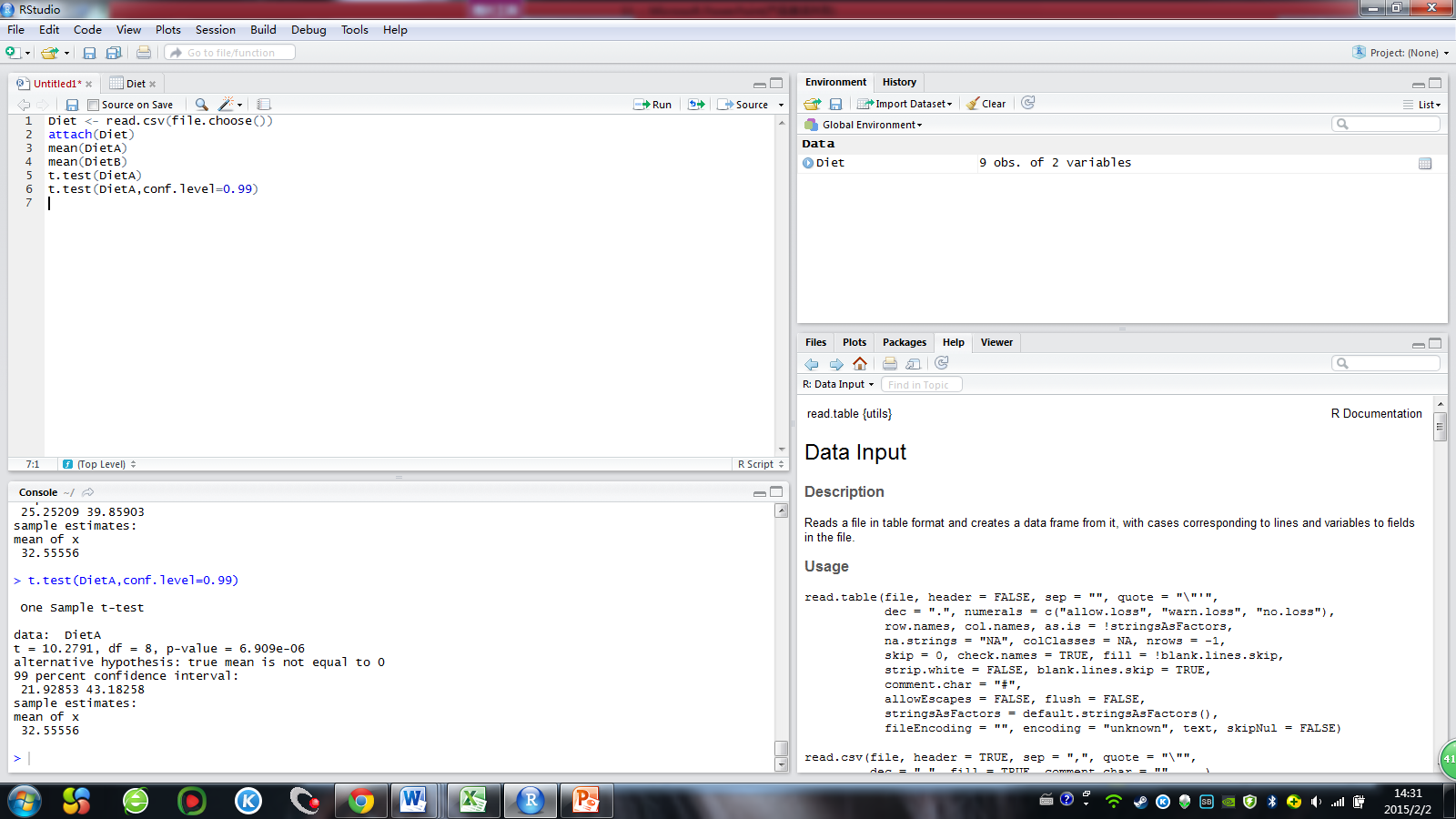
1e) Based on the above data, one researcher reported that those following Diet A lost, on average, 10.5 lbs more than those following Diet B, with a 95% confidence interval of ( -2.5 , 23.6 ) lbs. Discuss the relationship between this confidence interval and the p-value calculated in part d.

**Answer:**

If the null hypothesis of no difference in means was true, then the difference in means would be 0. So, if the confidence interval for the difference in means contains 0, there is no strong evidence against the null, and the p-value from the two-sample t-test will be greater than 0.05 (assuming a 95% CI). If the confidence interval for the difference in means does not contain 0, then we are 95% confident that the means are not equal, and the p-value from the two-sample t-test will be less than 0.05.

2. Analyze the data from problem 1 above using R, SAS or SPSS.





3. What’s the difference between a standard deviation and a standard error?

**Answer:**

**The standard deviation describes variability in a variable from subject**

**to subject. The standard error describes variability in a sample statistic (such as a**

**sample mean) from sample to sample. The standard error depends on the**

**standard deviation and sample size: SE=SD/(n)^(1/2)**

4. A t-test was performed to compare two groups of students, with 20 students in one group and 30 in the other, on the mean score of a standardized exam.

The resulting t-statistic was -2.20.

4a) What are the null and alternative hypotheses for this situation?

**Answer:**

**The null hypothesis is that there is no difference in the means of the two**

**populations represented by the two study samples. The alternative hypothesis is**

**that the means of the two populations differ, although the specifics of the**

**alternative hypothesis depend on whether we do a 1-tailed or 2-tailed test.**

4b) What are the degrees freedom for this t-statistic?

**Answer:**

**The degrees of freedom here are n1+n2-2=20+30-2=48**

4c) Find and interpret a 1-tailed p-value for this test statistic.

**Answer:**

**Since the degrees of freedom are greater than 30, we can use the z-table to approximate the**

**distribution of the t-statistic. The p-value is the tail area beyond the observed value of the test statistic. Using a z-table, the area between the mean and -2.20 0.4772. So, the area beyond -2.20 is 0.0228, and so the 1-tailed p-value is p=0.0228. Since this p-value is below the conventional cut-off value of 0.05, we can conclude that we have significant evidence to show a difference between the two means at the (one-tailed) p<0.05 level.**

4d) Find and interpret a 2-tailed p-value for this test statistic.

**Answer:**

**The 2-tailed p-value is the two-tailed area beyond the observed value of the test statistic. Here, the two-tailed area beyond ±2.20 is 2(0.0228)=0.0456. Since this p-value is still below the conventional cut-off value of 0.05, we can conclude that we have significant evidence to show a difference between the two means at the (one-tailed) p<0.05 level.**