# Homework 3

Your Name: Minguk Kim

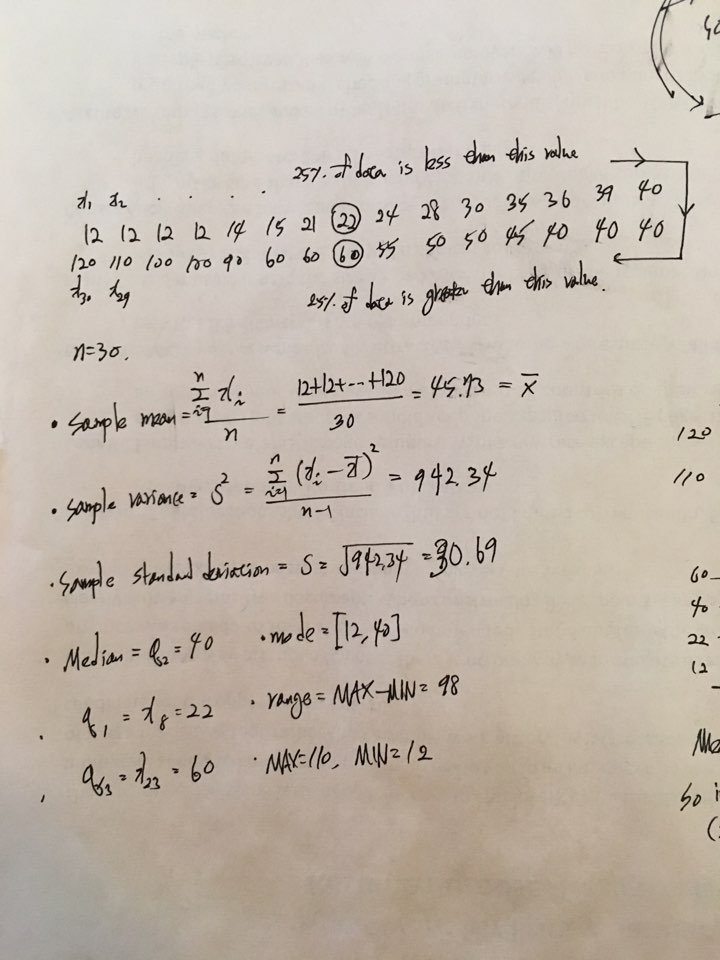
Student ID: A20437179

Manually solve the following problem

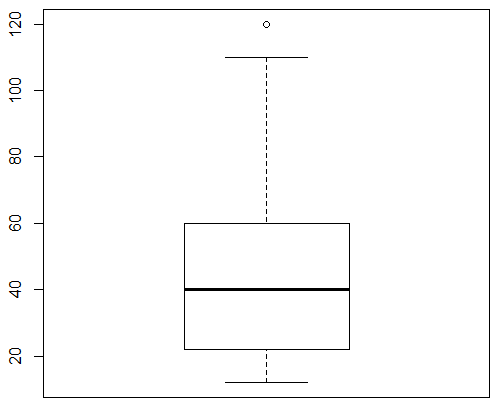
Chicago Ventra Transit Card can be used on both CTA bus, metro and Pace buses. We are going to explore a resident’s average monthly cost on CTA transportations. In this case, we performed a survey, and collect monthly cost on CTA transits from 30 people, their monthly cost can be listed as follows:

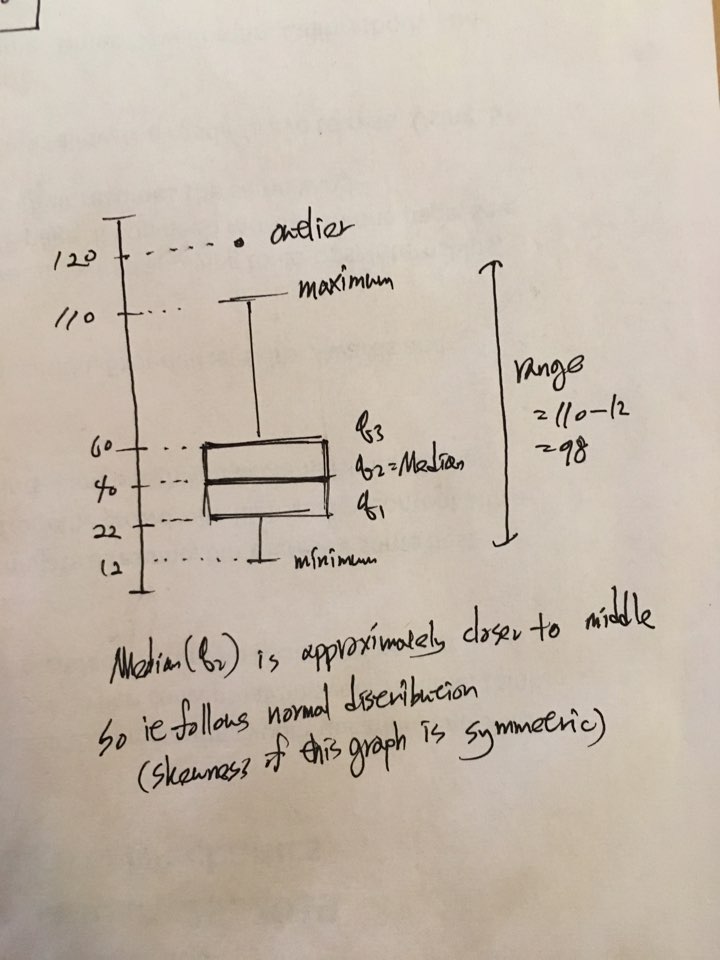
12, 12, 12, 15, 24, 35, 14, 12, 120, 55, 45, 30, 40, 40, 40, 60, 60, 40, 50, 22, 36, 28, 21, 50, 39, 60, 90, 100, 110, 100

1). [10] Calculate the mean, variance, median, mode, range, Q1 and Q3, standard deviation, min and max values

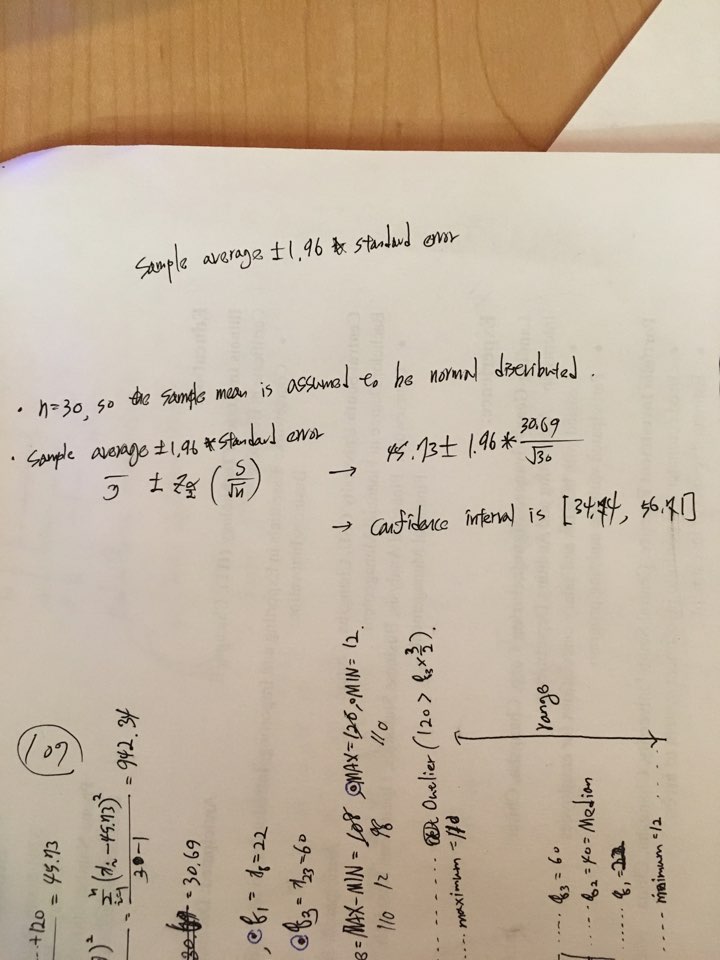


2). [10] To further understand the distribution, we draw a boxplot as follows. Interpret the box plot





3). [15] Use the sample statistics to estimate the average monthly cost on CTA transits by Chicago residents by using 95% as the confidence level



So this interval estimate procedure will catch the true population average mu in [34.74, 56.71] with a confidence level 95%

4). In addition, they can use Chicago metro trains in their daily life. We ask the same group of sample users to use Chicago metro trains only and record their monthly cost on trains. In this case, we get two groups of data as follows. We display it as two tables, since it is not able to put them on a single table. For each table, row 1 is the monthly cost by using CTA for each person, row2 is the monthly cost by using train for each person. Each column contains the costs by a same person but use different transportation (CTA vs Metro Trains)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| G1 | 12 | 12 | 12 | 15 | 24 | 35 | 14 | 12 | 120 | 55 | 45 | 30 | 40 | 40 | 40 |
| G2 | 10 | 16 | 13 | 14 | 28 | 41 | 16 | 10 | 80 | 40 | 75 | 25 | 41 | 29 | 40 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| G1 | 60 | 60 | 40 | 50 | 22 | 36 | 28 | 21 | 50 | 39 | 60 | 90 | 100 | 110 | 100 |
| G2 | 50 | 50 | 60 | 60 | 80 | 40 | 25 | 25 | 40 | 25 | 25 | 120 | 120 | 120 | 100 |

It is told that there are no differences if they are going to use the CTA or Metro trains. However, we do not believe it. We are going to use hypothesis testing to examine whether the costs by two different means should be the same or not. Assume we use 95% confidence level.

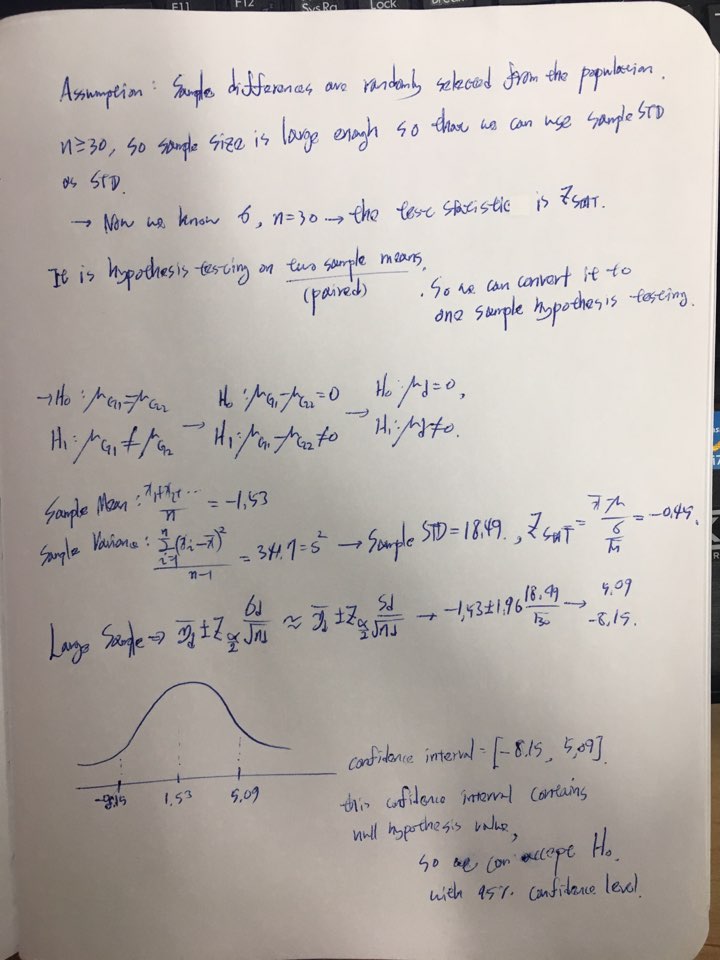
4.1), [15] write down your null and alternative hypothesis, and tell me is it a two-tailed or one-tailed test

Null hypothesis: Average monthly cost on CTA is equal to average monthly cost on Metro trains

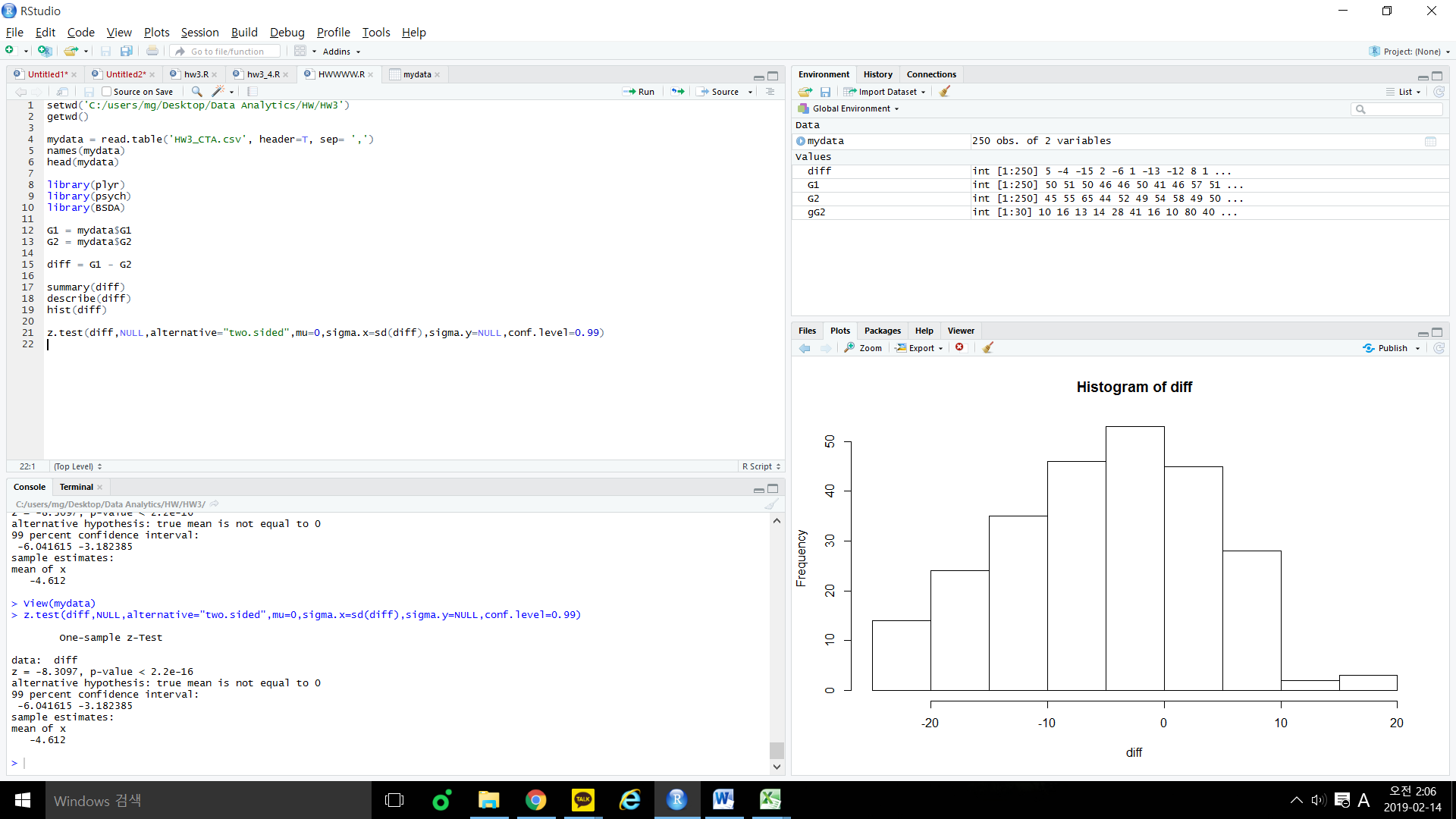
Alternative hypothesis: Average monthly cost on CTA is not equal to average monthly cost on Metro trains.

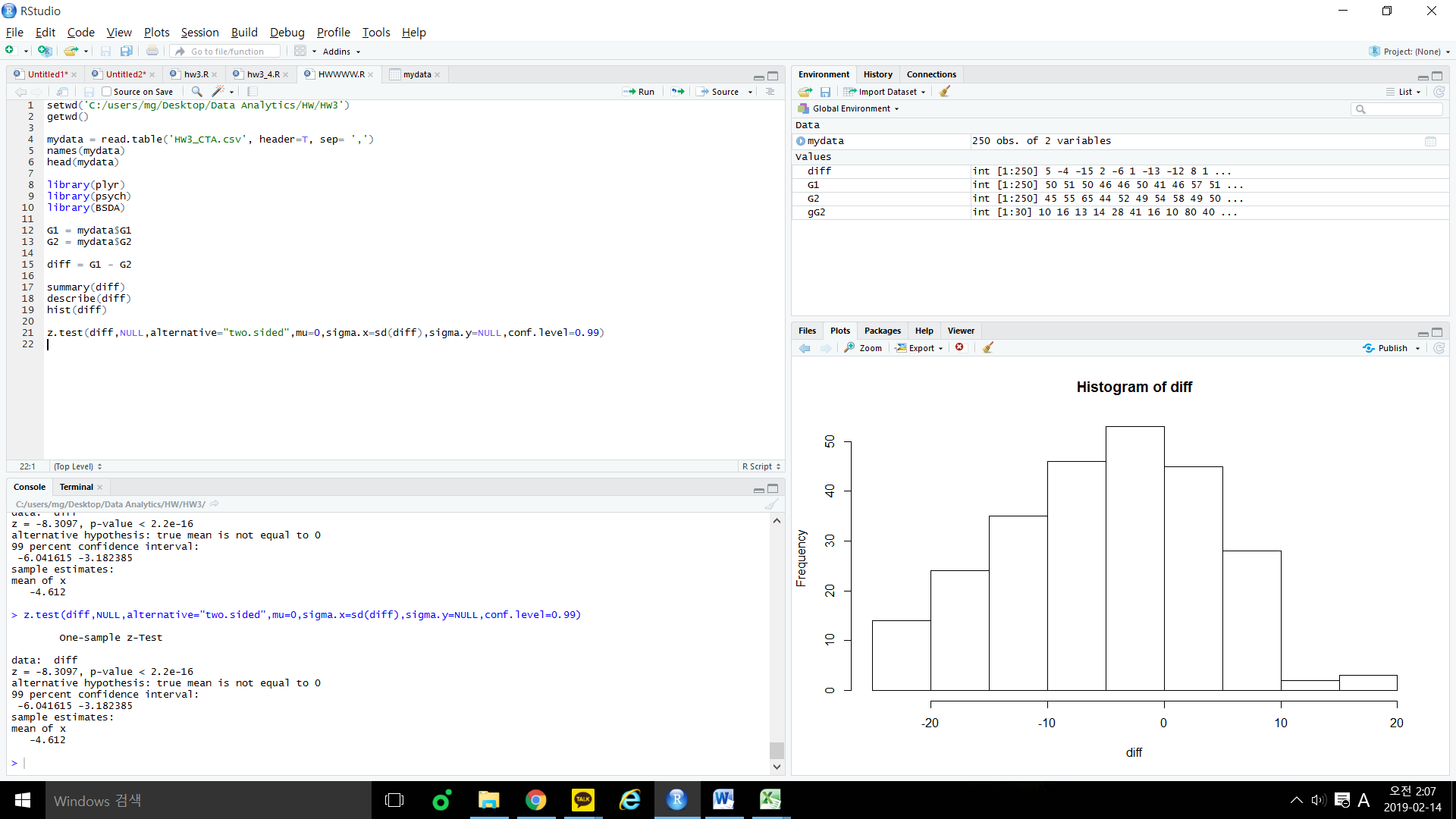
And It will be two-tailed test.

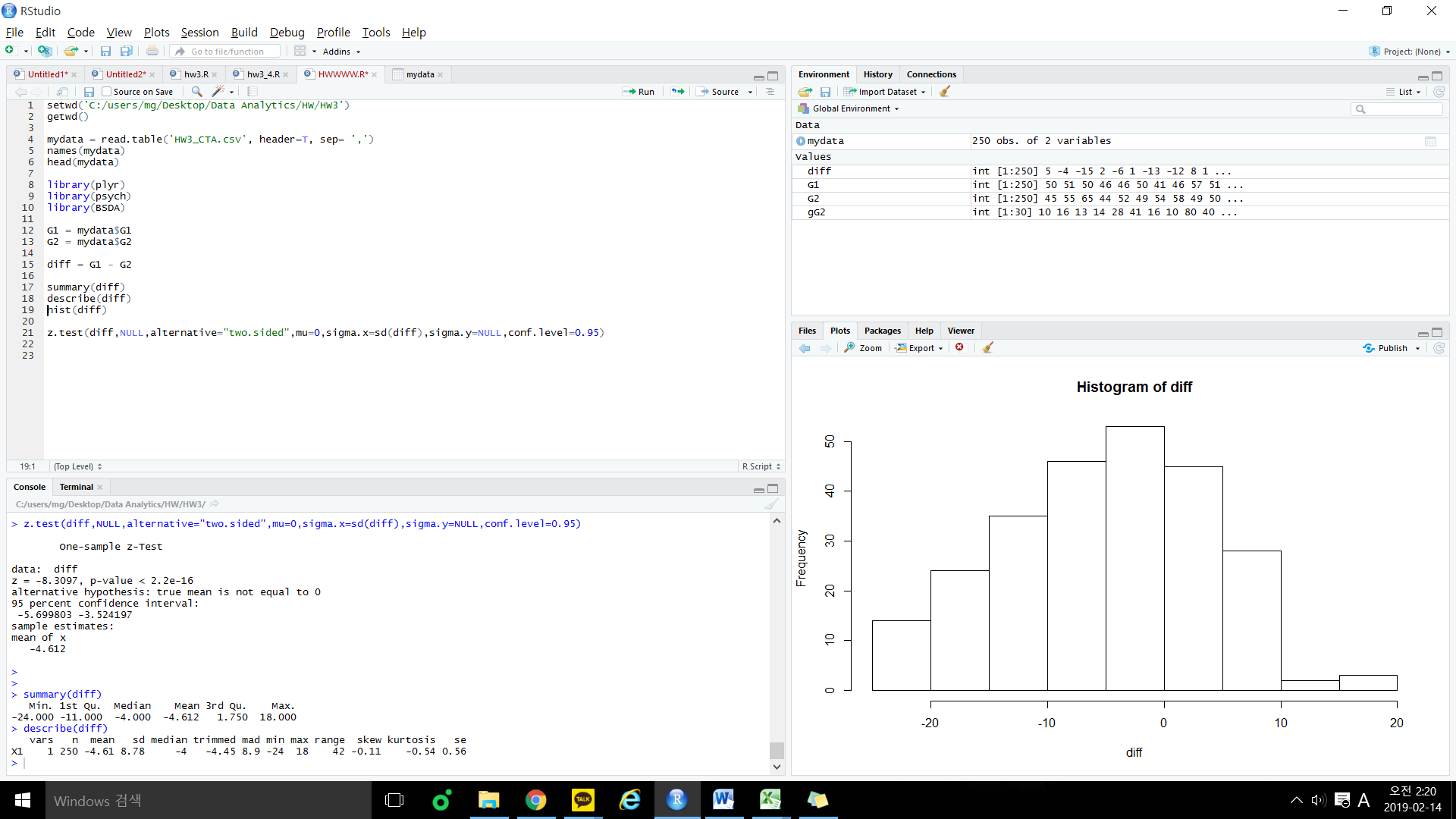
4.2), [30] Perform hypothesis testing to tell whether the costs by two different means are the same or not based on 95% confidence level

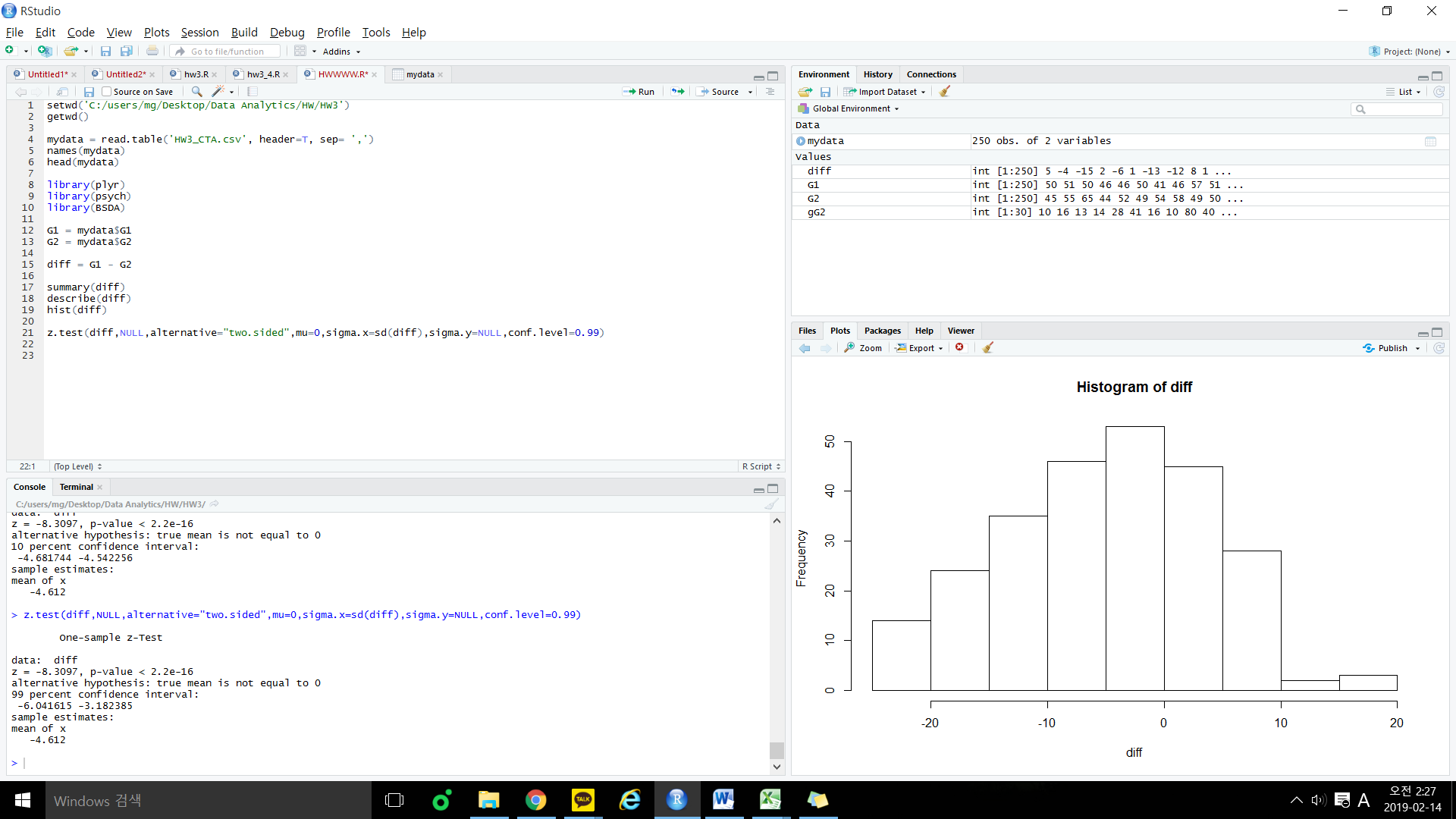


4.3), [20] Use R to solve the problem by using the data in HW3\_CTA.csv AND using 99% as the confidence level, paste your codes and the outputs in this section, and give the final conclusions. Note that the data in the file has exactly the same setting with part 4.2). The only difference is that it is a larger data with 250 rows









Let’s set the same hypothesis as 4.2).

P-value is less than 2.2e-16 and absolutely level of significance is greater than P-value so that we can reject H0.

Therefore, we know that there is difference between CTA and Metro trains monthly cost with a 99% confidence level. In addition, we noticed that importance of sample size, sample STD and level of significance.