# Homework 2

Your Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Student ID:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. **(35 points) Manually solve the problem below (do not use R):**

Note, if you need either z value or t value, you can find them by using this tool:   
<http://www.mathcracker.com/z_critical_values.php>

<http://www.mathcracker.com/t_critical_values.php>

<https://www.socscistatistics.com/pvalues/>

A bank branch located in a commercial district of a city has the business objective of improving the process for serving customer during the noon to 1 PM (lunch period). The waiting time (defined as the time the customer enters the line until he or she reaches the teller window) of a random sample of 15 customers is collected, and the results are organized and stored as below:

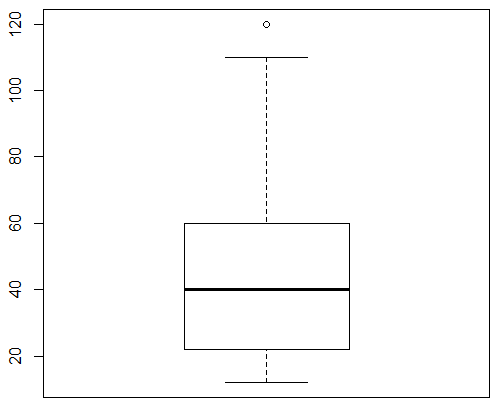
4.21, 5.55, 3.02, 5.13, 4.77, 2.34, 3.54, 3.20  
4.50, 6.10, 0.38, 5.12, 6.46, 6.19, 3.79

1. Calculate the mean and standard deviation, and find q1, q3 from the values above. Is the distribution symmetric? Why? [10]
2. As a customer walks into the branch office during the lunch period. She asks the branch manager how long she can expect to wait. If you are the manager, answer this question by using 95% confidence. [10]
3. We were told the average waiting minute will be 5 minutes. But we think it could be more than 5 minutes. By using 90% as confidence level, validate the hypothesis. Show your steps and calculations [15]

**2. (55 points)** 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). [5] To further understand the distribution, we draw a boxplot as follows. Interpret the box plot



2). [15] Use the sample statistics to estimate the average monthly cost on CTA transits by Chicago residents by using 95% as the confidence level. Assume that we know the population variance is 4. Use R to solve the problem. paste your snapshot of R coding and outputs, also deliver your conclusions

4). In addition, they can use Chicago metro trains in their daily life. We ask the same group of 30 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) . Assume that we know they have the same population variance of 4.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| user | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| row1 | 12 | 12 | 12 | 15 | 24 | 35 | 14 | 12 | 120 | 55 | 45 | 30 | 40 | 40 | 40 |
| row2 | 10 | 16 | 13 | 14 | 28 | 41 | 16 | 10 | 80 | 40 | 75 | 25 | 41 | 29 | 40 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| user | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
| row1 | 60 | 60 | 40 | 50 | 22 | 36 | 28 | 21 | 50 | 39 | 60 | 90 | 100 | 110 | 100 |
| row2 | 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 believe using CTA is more expensive than using Metro trains. 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, and it is two independent or paired samples, why?

4.2), [20] Perform hypothesis testing to tell whether the costs by two different means are the same or not based on 95% confidence level, by using R. Again, give the R coding, snapshot, outputs, deliver your conclusions by referring to/explaining the outputs

**3. (10 points)** The z-test and t-test we used for hypothesis testing (including both one-sample and two-sample hypothesis testing) are also known as "Parametric Test". Alternatively, there are other statistical tests which can be used as alternatives, and they are called "Non-Parametric Test". Search online or find other learning materials to answer the questions below.

1). (5 points) what are the differences between Parametric Test and Non-Parametric Test?

2). (5 points) what are the requirements/conditions to use z-test or t-test? If our data does not meet these requirements, which non-parametric test is the alternative for hypothesis testing?