# R Final Report

*Chen Ming*

*May 20th*

## Part 1

**1.1 Report descriptive statistics of the data set obtained in (a).**

The descriptive statistics of the data are summarized in the following table.

```
##        item Wr.Hnd NW.Hnd  Pulse Height   Age
## 1    Min.   13.0  12.50  35.00  152.0 16.92
## 2 1st Qu.   17.5  17.50  66.75  165.0 17.67
## 3  Median   18.5  18.50  72.00  170.6 18.58
## 4    Mean   18.8  18.73  74.02  172.5 20.43
## 5 3rd Qu.   20.0  20.00  80.00  180.0 20.17
## 6    Max.   23.2  23.50 104.00  200.0 70.42
```

**1.2 Use boxplot to show the distributions of the height of male and female students.**

The distributions of the height of male and female students are as follows.

**1.3 Which numerical variables might have an influence on the student's pulse?**

To indentify variables that have influence on students' pulse, we can conduct regression analysis. By fitting models with "Pulse" as dependent variables and other numerical variables as independent variables respectively, we can easily find that all the numerical variables might have an influence on students' pulse, with p-value of those fitted model all below 0.01.

**1.4 Is the probability of a student clapping his/her left hand on top less than 0.2?**

To make sure whether the probability of a student clapping hand on top less than 0.2, we can use **prop.test** funtion to conduct proportion test. The null hypothesis is that "The probability of a student clapping his/her left hand on top greater than or equal to 0.2", and the alternative hypothesis is that "probability of a student clapping his/her left hand on top less than 0.2". The result suggests that the p-value is 0.1626, which means that we cannot reject the null hypothesis at .1 level of significance. Therefore we can infer that the probability of a student clapping his/her left hand on top is equal to or greater than 0.2.

**1.5 Is the span of the writing hand in general larger than the span of the non-writing hand?**

To make sure whether the span of the writing hand in general larger than the span of the non-writing hand, we can make a hypothesis test. The null hypothesis is 'the span of the writing hand is in general smaller than or equal to the span of the non-writing hand', and the alternative hypothesis is 'the span of the writing hand is in general larger to the span of the non-writing hand'. With a p-value of 0.3696, The hypothesis test result suggests that we cannot reject the null hypothesis. Therefore the the span of the writing hand is in general equal to or smaller than the span of the non-writing hand.

# Part 2

**2.1 According to CLT, what is the approximated distribution of the sample means?**

According to CLT, the means of the sample tends towards a normal distribution. The approximated distribution of the sample means is normal distribution.

**2.2 Draw the density plots of the sample means and its approximated distribution on one graph.**

**2.3 Show the qq plot of the distribution of the sample means.**

The QQ plot of the distribution of the sample means is as follows.

**2.4 What conclusion can you draw from this simulation?**

From the density plot, we can know that the sample means distribution tends towards normal distribution as the sample means number get greater. From QQ plot, we can also observe that the points fall on the straight 45-degree line, which suggests that the residual values are normally distributed with a mean of 0. Hence the sample means are normally distributed.

# Part 3

**3.1 Draw 5000 random samples of X.**

The random samples generated are omitted here.

**3.2 Given $\alpha = 0.05$, find the VaR of the samples obtained in a).**

Since VaR denotes the critical value $L$ such that $P(X > L) = \alpha$ ,to find the VaR of the samples, we first need to sort the array of the change of the stock price ascendingly. Then we can use **quantile** function to find out the change value situated at given $\alpha$. The VaR is -2.130284.

**3.3 Find the CVaR of the samples obtained in a).**

# Part 4

**4.1 Report on data exploration.**

**4.2 Report on modeling (baseline model and alternatives).**

**4.3 Report on results and interpretation of the fitted model.**

**4.4 Report on model assumptions.**