Experiment Report for Assignment 2 of *Using R in Financial Statistics*

Ming Chen

May 8, 2019

1 Experiment Targets

1. Learn to draw pie chart with R

2. Learn to plot scatter graph with R

3. Learn to draw density graph of distributions with R

4. Learn to conduct hypothesis test for single sample and double samples with R

2 Experiment Contents

1. Draw a pie chart to show the percentage of the employees in different provinces.

2. Draw a scatter graph to show the association relationship between age and salary, with different colors representing different groups.

3. Generate three samples of 100,000 observations from t distribution with degrees of freedom 5, 10 and 30, respectively. Compare the (estimated) density plots of these samples with the standard normal density function in one plot.

4. Use the "mtcars" data set to conduct the hypothesis testing that the mean of mpg (miles per gallon) is larger than 15.

5. Use the "mtcars" data set to test whether the means of wt (weight) of the cars are different in the automatic and manual (am, 0 = automatic, 1 = manual) cases.

6. Use hypothesis testing to show whether these two drugs have significant difference.

3 Experiment Instruments

- RStudio of version 1.1.463

- R X64 of version 3.5.2

4 Experiment Designs

Problem 1

Constructing vectors of slices and labels respectively;
Using pie function to construct pie chart for "Employees Distribution"

Problem 2

Generating random values $s.t.$ uniform distribution of ages for the three sets of people;
Generating random values $s.t.$ normal distribution of wages for the three sets of people;
Marking the three sets of people by "22-30 years old","21-45 years old","46-60 years old" respectively;

Constructing data frame consisted of "Groups","Ages", "Wages" respectively;
Factorizing "Groups" of the data frame;
Importing *ggplot* function and drawing scatter plot of the three sets of people.

Problem 3

Generating random values *s.t.* t distribution of 10000 each for sample1,2,3 respectively by using *rt* function;
Generating random value *s.t.* normal distribution of 10000 for sample 4 by using *rnorm* function;
Marking the three samples by using *rep* function;
Constructing data frame consisted of vectors of *Samples* and *Markers*;
Importing *ggplot2* and use the function to draw density graphs in a plot, in which samples are marked by four different colours.

Problem4

Making null hypothesis H_0 as "The mean of mpg is less or equal to 15" and making alternative hypothesis H_1 as "The mean of mpg is larger than 15";
Using *t.test* function to perform t test with $mu = 15$ and $alternative == greater$.

Problem 5

Making null hypothesis H_0 as "means of wt of automatic and manual are equal" and making alternative hypothesis H_1 as "means of wt of automatic and manual are unequal";
Using *subset* function to select data required and then subtracting data of *wt* from selected data;
Using *t.test* function to perform t test with $alternative == two - sided$.

Problem 6

Making null hypothesis H_0 as "the two drugs have similar effects" and making alternative hypothesis H_1 as "the two drugs have significant different effects"; Using *subset* function to select data required and then subtracting data of *wt* from selected data; Using *t.test* function to perform t test with $alternative == two-sided$.
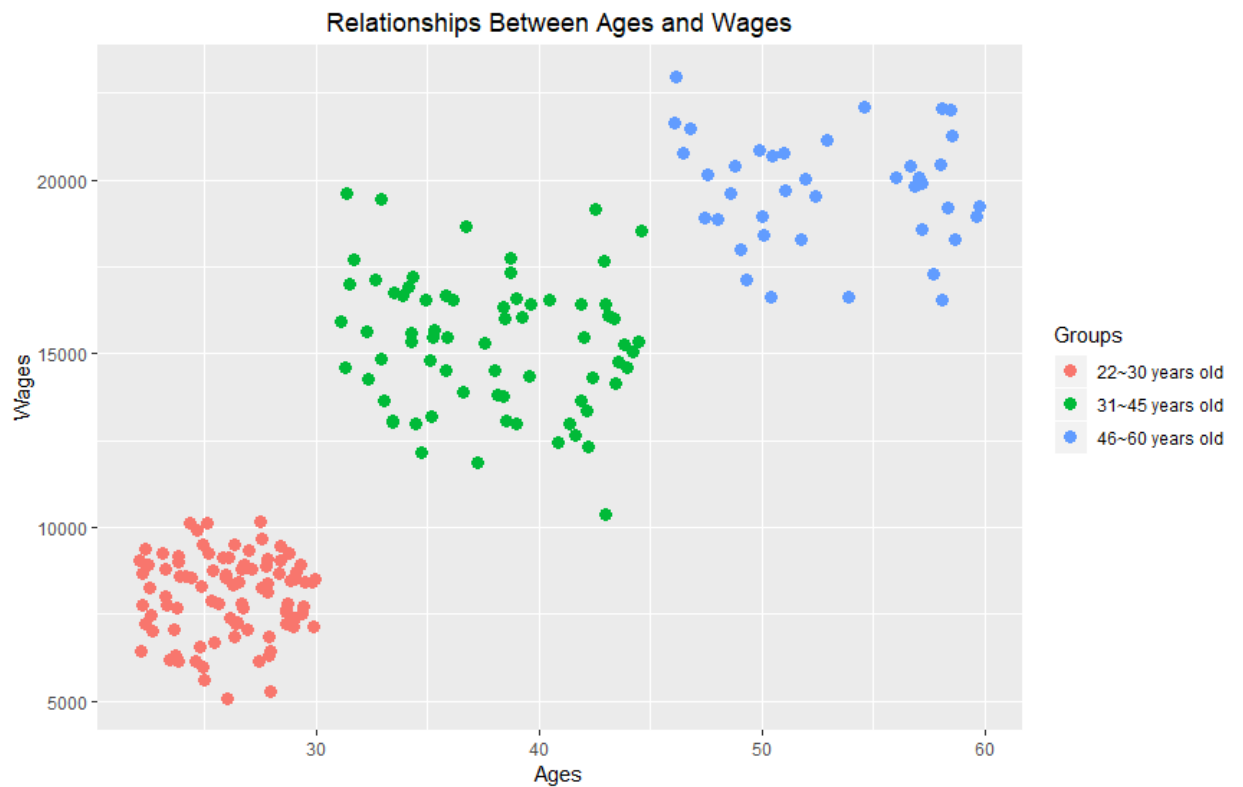
5 Experiment Results And Conclusions

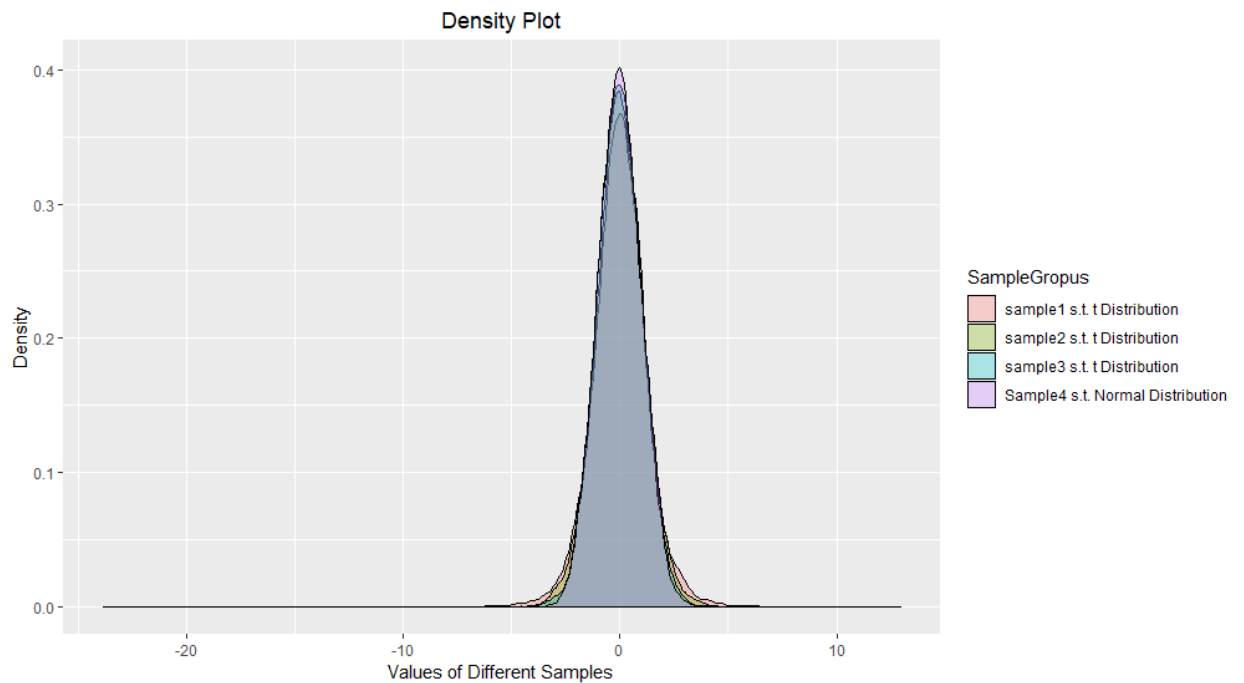## Problem 1

Employee Distributions
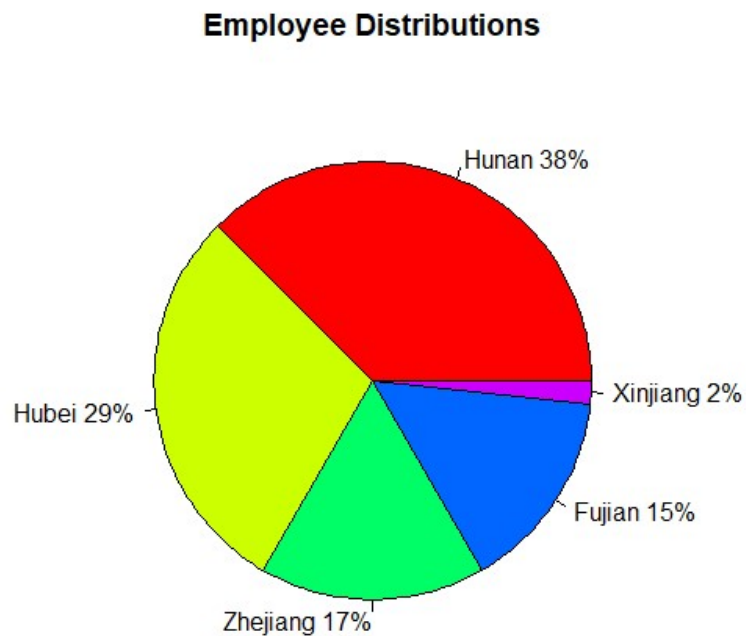
According to the experiment results, most of the employees of the company come from Hunan province (38%), and that the least employees come from Xinjiang ; The descending ranking of provinces in terms of employees distributions are Hunan, Hubei, Zhejiang, Fujian, Xinjiang.

Problem 2

According to the scatter plot, most of employees of 21-45 years old have wages of 7000-9000; most of employees of 31-45 years old have wages of 13000-17000; most of employees of 46-60 years old have wages of 18500-21500; wages of employees are positively propotional to their ages.

Problem 3

Density Plot

Problem 4

```
data:  mtcars$mpg
t = 4.778, df = 31, p-value = 2.027e-05
alternative hypothesis: true mean is greater than 15
95 percent confidence interval:
 18.28418       Inf
sample estimates:
mean of x
 20.09062
```

According to t statistics of 4.778 and p-value of 2.023e-05, null hypothesis is rejected at significance level of 0.05. we can accept the alternative hypothesis that mean of mpg (miles per gallon) is larger than 15.

Problem 5

```
data:  am_coloum_automatic and am_coloum_manual
t = 1.8772, df = 348.4, p-value = 0.06132
alternative hypothesis: true difference in means is not equ
al to 0
95 percent confidence interval:
 -0.7538182 32.3497623
sample estimates:
mean of x mean of y
 46.02645  30.22848
```

According to t statistics of 1.8772 and p-value of 0.06132, null hypothesis is rejected at significance level of 0.05. We can accept the alternative hypothesis that wt (weight) of the cars are different in the automatic and manual.

Problem 6

```
data:  subset1 and subset2
t = -1.8608, df = 17.776, p-value = 0.07939
alternative hypothesis: true difference in means is not equ
al to 0
95 percent confidence interval:
 -3.3654832  0.2054832
sample estimates:
mean of x mean of y
     0.75      2.33
```

According to t statistics of -1.8608 and p-value of 0.07939, null hypothesis is rejected at significance level of 0.05. We can accept the alternative hypothesis that the two drugs have significant difference in effects.