Framingham Analysis Report

1 INTRODUCTION

This analysis report will explore the dataset from the Framinghan Heart Study which is a long-term, ongoing cardiovascular cohort study of residents of the city of Framingham, Massachusetts. Our dataset has 4240 observations and 16 variables. This report will utilize the person's heart rate and the person's systolic blood pressure to create a new variable named shockIndex. The purpose of this report is to analyze how the shock index differs based on whether a person has diabetes, to explore how the shock index differs for different cholesterol levels in women and to explore the shock index between the non-smoker group and smoker group.

This report used bootstrap, difference in means, t-test to build confidence interval and compute p-value to conclude that we are 95% sure that the difference of the mean shock index in nondiabetic group and the mean shock index in diabetic group is between -0.004 and 0.051. Together, this report concluded that the shock index means of those who are diabetic and of those who aren't diabetic is with no significant difference.

2 ANALYSIS

2.1 EXPLORE THE SHOCKINDEX VARIABLE AND DIABETES VARIABLE

2.1.1 Data Exploration

To analyze how the shock index differs based on whether a person has diabetes, we want to know the spread of shock index in different diabetes group. The side-by-side boxplot and histogram of shock index in different diabetes group would show the comparison and differences. (See figure 1)
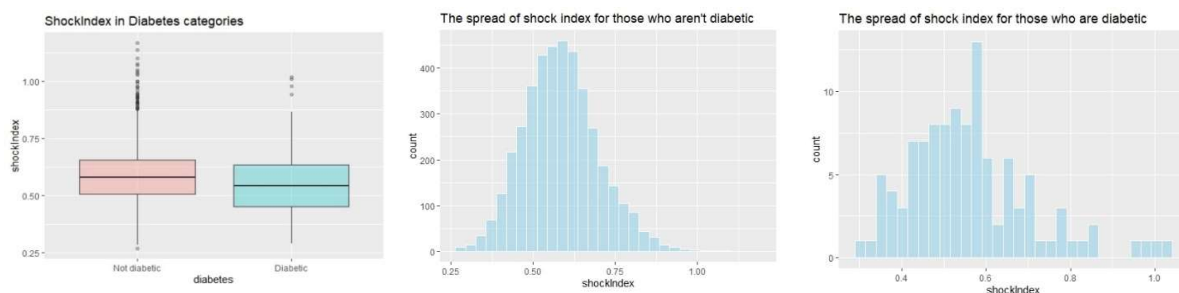

Figure 1: left: shock index of diabetes categories. Middle: the spread of shock index for those who aren't diabetic. Right: the spread of shock index for those who are diabetic

2.1.2 Inference

From the Framingham dataset, the diabetes variable is a categorical variable with two categories and the shock index variable is a quantitative variable. This report used the inference of Difference in Means to analyze these two variables. The summarize of shock

index were computed for each of two groups (One for those who aren't diabetic, one for those who are diabetic). (See table 1)

```
   diabetes avg_shockIndex sd_shockIndex      n
   <fct>             <dbl>         <dbl> <int>
1 0                  0.586         0.115  4130
2 1                  0.562         0.148   109
```
Table 1

The sample difference in means with our two groups were calculated, so the point estimate of difference in means is 0.024(Mean shock index in nondiabetic group minus mean shock index in diabetic group). Then this report created two subsets of the dataset, one for those who aren't diabetic, one for those who are diabetic. Using the shock index in these two subsets, this report simulated a 95% confidence interval with the bootstrap simulation. The 95% confidence interval computed is (-0.004, 0.051). Therefore, we are 95% sure that the difference of the mean shock index in nondiabetic group and the mean shock index in diabetic group is between -0.004 and 0.051.

Then, this report tried to answer the question: whether the shock index mean of each group is the same or not.

The null hypothesis was conducted that the population means of two groups is with no difference. The alternative hypothesis is that the population means of two group is different. Because the sample is from the Framinghan Heart Study, we can say it is random. Also, the sizes of two subsets are both larger than 30. With the conditions for t-distribution are met, we used 0.05 as significant level to do the t-test. The output is as follows:

```
        Welch Two Sample t-test

data:  fm2.nodiabete$shockIndex and fm2.diabete$shockIndex
t = 1.6901, df = 111.49, p-value = 0.0938
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.004165615  0.052489710
sample estimates:
mean of x mean of y
0.5858176 0.5616556
```

The p-value is 0.0938. It is greater than significant level (0.05). So, the null hypothesis is failed to reject. We can conclude that the shock index means of those who are diabetic and of those who aren't diabetic is with no significant difference.

2.2 EXPLORE THE SHOCKINDEX VARIABLE AND CHOLESTEROL VARIABLE IN WOMEN

The shock index variable and cholesterol variable are both quantitative variables. The report would like to use the scatterplot graph to show the correlation of these two variables.
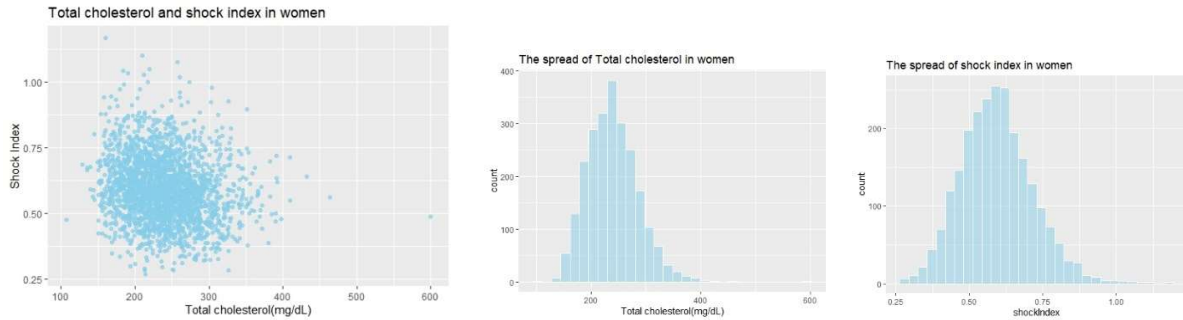
Figure 2: Left: Total cholesterol and shock index in women. Middle: The spread of Total cholesterol in women. Right: The spread of shock index in women

2.3 EXPLORE THE SHOCKINDEX VARIABLE AND SMOKERS VARIABLE

The shock index variable is quantitative variable and the smokers variable is categorical variable. The report created a boxplot graph and computed the statistics data. See figure 3 and table 2.
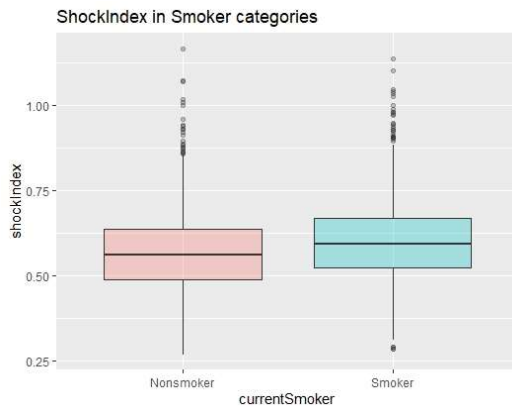

Figure 3

currentSmoker	n	avg_shockindex	median_shockindex	sd_shockindex	`min(shockIndex)`	`max(shockIndex)`
<fct>	*<int>*	*<dbl>*	*<dbl>*	*<dbl>*	*<dbl>*	*<dbl>*
0	2126	0.568	0.561	0.115	0.269	1.17
1	2063	0.603	0.594	0.116	0.285	1.14

Table 2

The spread of shock index in nonsmoker group and in smoker group is a little bit difference. So, later the analysis for the difference in means could be done to test how the shock index differs whether the person is a smoker or not.

3 CONCLUSION

This report concludes that we are 95% sure that the difference of the mean shock index in nondiabetic group and the mean shock index in diabetic group is between -0.004 and 0.051.

Also, we can conclude that the shock index means of those who are diabetic and of those who aren't diabetic is with no significant difference.

After exploring shockIndex variable and cholesterol variable in women, we found the spread of shock index in women is the bell-shape normal distribution and the spread of cholesterol in women is the bell-shape normal distribution too. And there seems to be no correlation between shock index and cholesterol in women through the scatterplot.

After exploring the shockIndex variable and smokers variable, we found the spread of shock index in nonsmoker group and in smoker group is a little bit different. So, later the analysis for the difference in means could be done to test how the shock index differs whether the person is smoker or not.