Introduction to R and Statistical Thinking
Xinyi (Anny) Cui

1. We can firstly use > treesData< − trees to assign the values and then use the function >head(treesData) to display the first six rows of the build-in data frame, the picture is shown below:

```
> treesData<-trees
> head(treesData)
  Girth Height Volume
1   8.3     70   10.3
2   8.6     65   10.3
3   8.8     63   10.2
4  10.5     72   16.4
5  10.7     81   18.8
6  10.8     83   19.7
>
```

2. The average volume of the given timber sample is 30.2 cubic feet. (I did not attach treesData, so I directly use >round(mean(treesData$Volume), 1) to get the mean of sample volumes with 1 digit in decimal places.)

3. Article questions:

   a. The response variable is blood pressure and this is a quantitative variable. The explanatory variable is the type of chocolates. The researchers want to see if the dark chocolate contained plant substances called polyphenols can lower blood pressure.

   b. This study is an experiment and researchers select adults with high blood pressure and assign them into two groups. One group is given dark chocolate and the other one is given white chocolate. The researchers control other probable influenced factors and make sure the two groups obtain the same amount of different types of chocolate every day for two weeks.

   c. The sample is the selected thirteen adults with untreated mild hypertension, and the population is those people who have hypertension or face the problem of high blood pressure.

   d. The scientists are trying to evaluate a treatment to see whether eating dark chocolate can help people who have hypertension lower their blood pressure or not.

   e. The intuited model is $BloodPressure = f(type\ of\ chocolate) + \epsilon$. We know the study divides selected thirteen adults into two groups and the only difference between these groups is that half of them get dark chocolate and the other half get white chocolate. Therefore, we can assume the model can be chosen as the response variable, blood pressure, influenced by the explanatory variable, dark chocolate, and some unavoidable errors (maybe weights of people or their ages). We can denote Y_s as the systolic blood pressure change and Y_d as the diastolic

blood pressure change. We also can let $\mu_{s,1}$ be the mean of systolic blood pressure dropped and $\mu_{d,1}$ be the mean of diastolic blood pressure dropped in the dark chocolate group. Similarly, let $\mu_{s,2}$ and $\mu_{d,2}$ be variables in the white chocolate group. Thus, the model is shown below:

$$Y_s = \mu_{s,1} + \epsilon \text{ and } Y_d = \mu_{d,1} + \epsilon \qquad \text{for people who eat dark chocolate}$$
$$Y_s = \mu_{s,2} + \epsilon \text{ and } Y_d = \mu_{d,2} + \epsilon \qquad \text{for people who eat white chocolate}$$

f. To fit the model, we can use some given data from the experiment, the sample mean. The average blood pressure of those participants is about 153 over 84. After two weeks, in the dark chocolate group, systolic blood pressure drops an average of five points and the diastolic one drops an average of almost two points. While, blood pressure almost does not change in the white chocolate group. Then, we can see a table with the given information below:

Variable	Group	N	SystolicMean	DiastolicMean
BloodPressureDrop	DarkChocolate	half of 13	5	≈ 2
	WhiteChocolate	half of 13	≈ 0	≈ 0

We can assume that $\bar{Y}_{s,1} = 5$ and $\bar{Y}_{d,1} = 2$ for average blood pressure drops in the dark chocolate group, while, $\bar{Y}_{s,2} = 0$ and $\bar{Y}_{d,2} = 0$ for average blood pressure drops in the white chocolate group. The fitted model then expresses as $\hat{Y} = 5$ points of systolic blood pressure change, and $\hat{Y} = 2$ points of diastolic blood pressure change, for people who eat dark chocolate. Then $\hat{Y} = 0$ points of both systolic and diastolic blood pressure change for people who eat white chocolate. Actually, we can not really figure out the sample number for each group in the study as it only says that each group has half of the selected thirteen adults without specific number. It can be either 6 versus 7 or 7 versus 6. Also, we do not know the number of average blood pressure of each group. We only be told "The participants had an average blood pressure reading of about 153 over 84." and "While their blood pressure did not fall enough to be considered in the desirable range, below 120 over 80." These unknown data might influence our models and the outcome of this experiment. In addition, the lack of standard deviation might be helpless for us to check the rationality of our models.

g. The results can help us establish causation to some extent. The researchers indeed make control groups and make sure the only difference between their two groups is the chocolate. The experiment provides us with a clue that eating dark chocolate will cause both systolic and diastolic blood pressure lower. Even though the size of sample is a little small (only 13 patients in total), the blood pressure does not fall enough to be considered in the desirable range, and some details like number of people in each group are unclear, the results show that the group of people who eat dark chocolate drops their blood pressure compared with those who eat white chocolate. Dark chocolate "might serve as a promising approach to reduce systolic blood pressure," said lead author Dr. Dirk Taubert of the University of Cologne. At least, we take a glance of the conclusion that dark chocolate can affect the blood pressure to drop.