HW: Introduction to R and Statistical Thinking

Jack Ruder

Sep 06, 2022

1

```
> head(trees,6)
  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
```

3

The sample mean of timber yielded from the trees is approximately 30.17 cubic feet.

a

The response variable is the blood pressure of the participants (systolic and diastolic). The explanatory variable is whether the participants were given white chocolate or dark chocolate.

b

This study is an experiment, researchers split a group into two and controlled (to some extent) what participants ate in order to observe an effect.

c

The high-blood-pressure participants are the sample, the population is people whose blood pressure is around (loosely, no data is provided here) 153/84.

d

The aim of the model was to evaluate a treatment.

e

Let S, D be systolic and diastolic pressures respectively. Let $C = 0$ represent eating white chocolate, and let $C = 1$ represent eating dark chocolate. The form of the model is thus

$$S_{end} = S_{initial} - C \cdot (\mu_S + \epsilon) \sim f_S(\mu_S)$$
$$D_{end} = D_{initial} - C \cdot (\mu_D + \epsilon) \sim f_D(\mu_D)$$

where f_S and f_D are unknown distributions of the decreases in blood pressure, distributed with mean μ_i.

f

The given data is $S_{initial} = 153$, $\mu_S = 5$, $D_{initial} = 84$, $\mu_S = 2$. Without any more datapoints it would be incorrect to pick a particular distribution of data,

there is not enough information here. For this particular experimint, we have

$$S_{end} = 153 - C \cdot f_S(5),$$

$$D_{end} = 84 - C \cdot f_D(2).$$

g

Again, there is too little information about the statistical significance of the results. But, if the result was statistically significant then causation could be esdablished because this was a controlled experiment.