Statistical Inference: Analyzing ToothGrowth Data

Carl Fredriksson

Overview

In this part of the project we will analyze the ToothGrowth data from the R datasets package.

Load the ToothGrowth data

```
library(datasets)
df <- datasets::ToothGrowth
nrow(df)
```

```
## [1] 60
```

```
names(df)
```

```
## [1] "len"  "supp" "dose"
```

```
sapply(df, class)
```

```
##       len      supp      dose
## "numeric"  "factor" "numeric"
```

```
levels(df$supp)
```

```
## [1] "OJ" "VC"
```

```
unique(df$dose)
```

```
## [1] 0.5 1.0 2.0
```

The dataset is about the effect of vitamin C on tooth growth in guinea pigs. There are 60 observations of three variables: len, supp, dose.

Basic summary and data exploration

"len" is a numeric variable that describes tooth length. "len" takes values between 4.2 and 33.9. "supp" is a factor that describes what supplement the guinea pig received. "supp" has two levels: OJ, VC. "OJ" means orange juice, "VC" means ascorbic acid. "dose" is a numeric variable that describes the dose of the supplement in milligrams/day. In the dataset "dose" takes the values: 0.5, 1, 2. Each animal received on of the doses of one of the supplements.

```
summary(df)
```
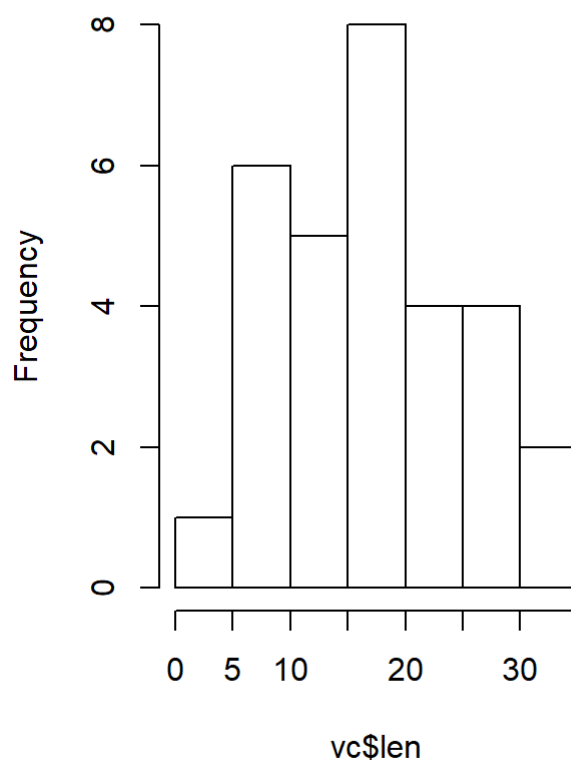
```
##      len          supp       dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

```
oj <- subset(df, supp=="OJ")
vc <- subset(df, supp=="VC")
par(mfrow=c(1, 2))
hist(oj$len)
hist(vc$len)
```
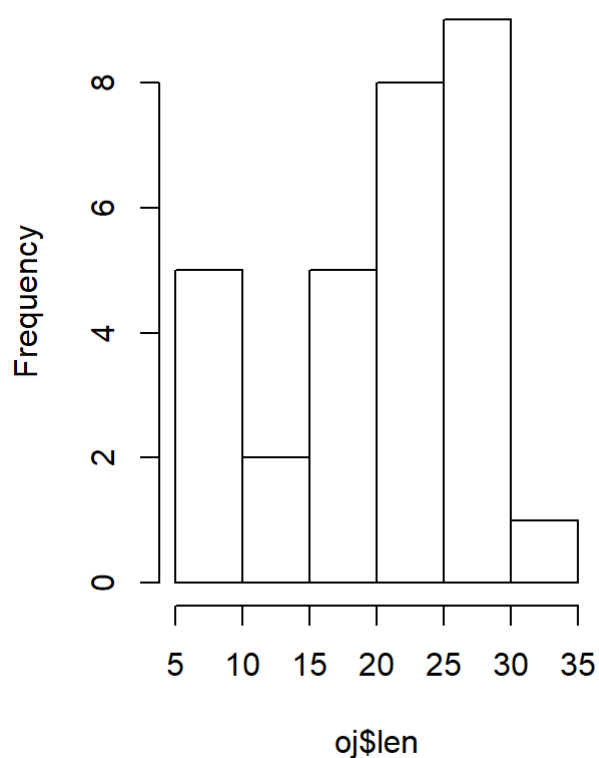

Hypothesis testing

I think that "OJ" is a more effective supplement than "VC". To verify this idea I am going to do a hypothesis test where the null hypothesis is that there is no difference between the supplements and the alternative hypothesis is that tooth lengths from "OJ" are longer on average. I am going to use a significance level of $\alpha = 0.05$.

$$H_0 : \mu_{OJ} = \mu_{VC}$$
$$H_a : \mu_{OJ} > \mu_{VC}$$

```
t.test(oj$len, vc$len, alternative="greater")$p.value
```

```
## [1] 0.03031725
```

Conclusion

Since the p-value is smaller than alpha ($p = 0.03031725 < \alpha = 0.05$) I conclude that "OJ" is a more effective supplement for guinea pig tooth growth. I assume that the observations are independent and identically distributed (iid). There are 30 observations for each supplement so that should be enough for the t-test.