Part 2: Tooth Growth from Vitamin C

Nikolai Alexander

Overview

For *Part 2* of the course project, we are analysing the ToothGrowth data set, which is observing the affect of Vitamin C on Guinea Pig tooh growth.. We are using confidence intervals and hypothesis tests to compare tooth growth by Delivery Method and Dose.

Data Summarization

The first thing we need to do is load the **ToothGrowth** data.

```
data(ToothGrowth)

head(ToothGrowth)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```
dim(ToothGrowth)
```

```
## [1] 60  3
```

```
class(ToothGrowth)
```

```
## [1] "data.frame"
```

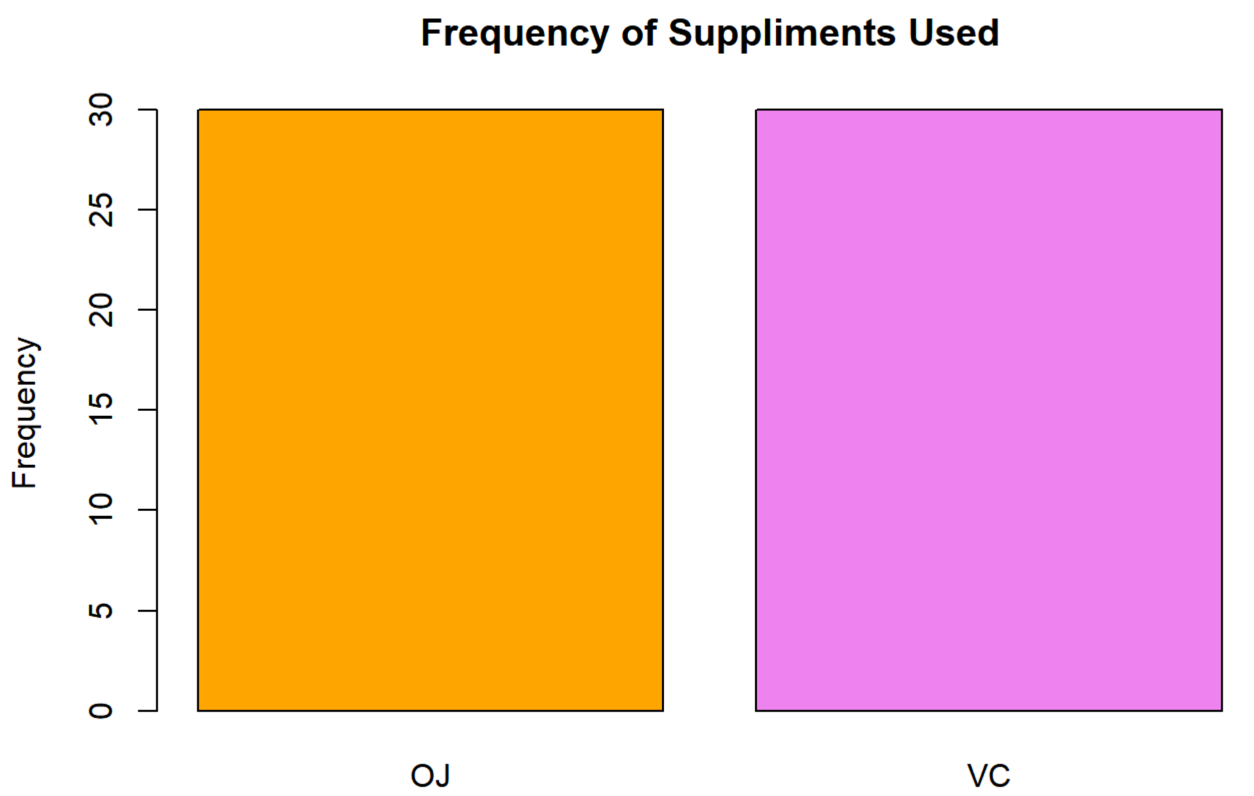
So now that we have the **ToothGrowth** data loaded, we see it is a 60x3 data frame. Now we need to observe the data in this data frame. First thing we can look at is the suppliment, particularly how many different suppliments are in the study

```
unique(ToothGrowth$supp)
```

```
## [1] VC OJ
## Levels: OJ VC
```

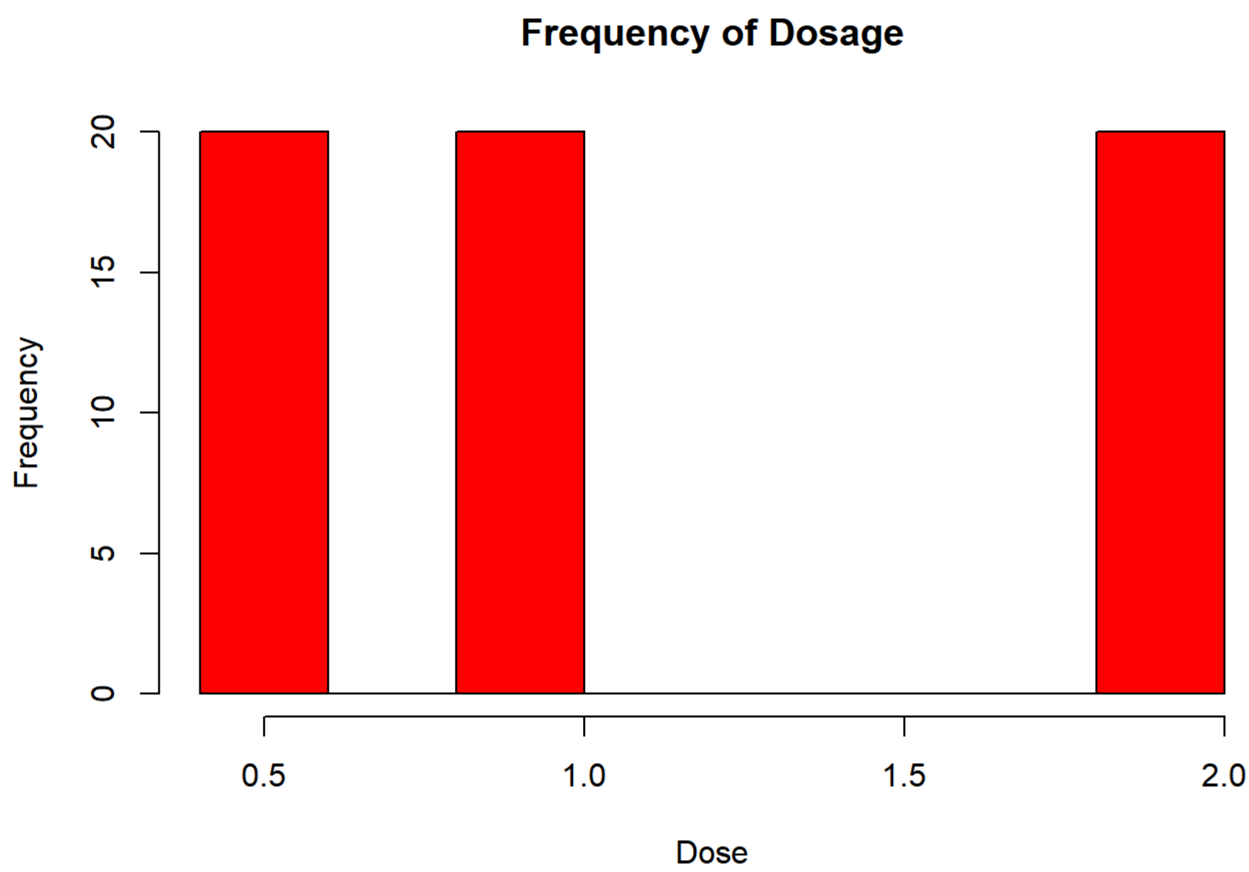
There are only two **Delivery Methods**. Now let's look at how frequency each is used…

```
barplot(
  table(ToothGrowth$supp),
  col = c("orange","violet"),
  ylab = "Frequency",
  main = "Frequency of Suppliments Used"
)
```



So we know both were used equally in the data set. Now let us look at **Dosage**…

```
hist(
  ToothGrowth$dose,
  col = "red",
  xlab = "Dose",
  main = "Frequency of Dosage"
)
```



For these two observations, we know there was an equal use for both dosages and suppliments, making the experiments pretty even in terms of data collection. Now we need to observe the most important aspect: the **Dosage/Delivery Method** comparison to Tooth Growth. The easiest way to observe all 3 variables is through a bar plot.

```
par(mfrow = c(1,3))
boxplot(
  len ~ supp,
  ToothGrowth[ToothGrowth$dose == 0.5,],
  col = c("orange","violet"),
  ylab = "Tooth Length",
  main = "0.5",
  ylim = range(0,40)
)
boxplot(
  len ~ supp,
  ToothGrowth[ToothGrowth$dose == 1,],
  col = c("orange","violet"),
  main = "1",
  ylim = range(0,40)
)
boxplot(
  len ~ supp,
  ToothGrowth[ToothGrowth$dose == 2,],
  col = c("orange","violet"),
  main = "2",
  ylim = range(0,40)
)
title("Tooth Length vs Delivery Method by Dose", outer = TRUE, line = -1)
```

As we can see from the graphs, **OJ** seems to lead to higher tooth growth at lower doses, but the rate at which tooth growth increases is faster for **VC**; However, how true is this? We will need to use confidence intervals and hypothesis tests to observe the vaility of this argument.

Confidence Intervals / Hypothesis Testing

From out observation above, we have two things to test:

1. The effectiveness of **Delivery Method** on **Tooth Growth**
2. The effectiveness of **Dosage** on **Tooth Growth**

The easiest method to test the confidence intervals is to observe the *p-value* through a *t-test*.

Delivery Method affect on Tooth Growth

```
DMvsTG <- t.test(len ~ supp, data = ToothGrowth)
DMvsTG
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

The *p-value* for the Delivery Method is **0.0606345**, which is **greater** than the accepted *p-value* of **0.05**. Therefore, the argument is too weak for the hypothesis to be valid. The next thing we can look at is the affect of dosage on tooth growth.

Dosage affect on Tooth Growth

From the 3 dosage types, we need to do 3 seperate *t-tests*: **0.5 to 1.0**, **0.5 to 2.0**, and **1.0 to 2.0**.

Dosage Comparison: 0.5 to 1.0

```
sub_5to1 <- subset(ToothGrowth, dose == 0.5 | dose == 1)
dosvstg1 <- t.test(len ~ dose, data = sub_5to1)
dosvstg1
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

In this case our *p-value* is equal to **1.268300710^{-7}**, which is much less than the accepted *p-value* of **0.5**. Therefore, this argument is strong enough to be valid.

Dosage Comparison: 0.5 to 2.0

```
sub_5to2 <- subset(ToothGrowth, dose == 0.5 | dose == 2)
dosvstg2 <- t.test(len ~ dose, data = sub_5to2)
dosvstg2
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

In this case our *p-value* is equal to **4.39752510^{-14}**, which is much less than the accepted *p-value* of **0.5**. Therefore, this argument is strong enough to be valid.

Dosage Comparison: 1.0 to 2.0

```
sub_1to2 <- subset(ToothGrowth, dose == 1 | dose == 2)
dosvstg3 <- t.test(len ~ dose, data = sub_1to2)
dosvstg3
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

In this case our *p-value* is equal to **1.906429510^{-5}**, which is much less than the accepted *p-value* of **0.5**. Therefore, this argument is strong enough to be valid.

Conclusion

The **Delivery Method** has little to no affect on Tooth Growth in Guinea Pigs. However, **Dosage** seems to play a major role in tooth growth.