

Statistical Inference Course Project Part 2

Eric Scuccimarra

4 January 2018

Overview

Now in the second portion of the project, I will analyze the ToothGrowth data in the R datasets package. I will:

1. Provide a basic summary of the data
2. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

Loading the data

The data is part of the R datasets. We will load it and convert the dose to a factor for ease of grouping the data by dose, in the interest of more readable plots I will also change the levels of the supplement factor.

```
library(datasets)
data(ToothGrowth)
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
levels(ToothGrowth$supp) <- c("Orange Juice", "Vitamin C")
```

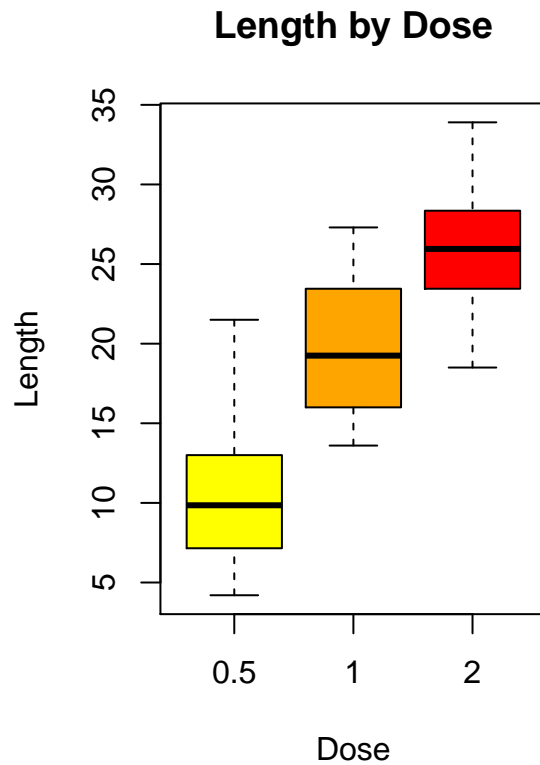
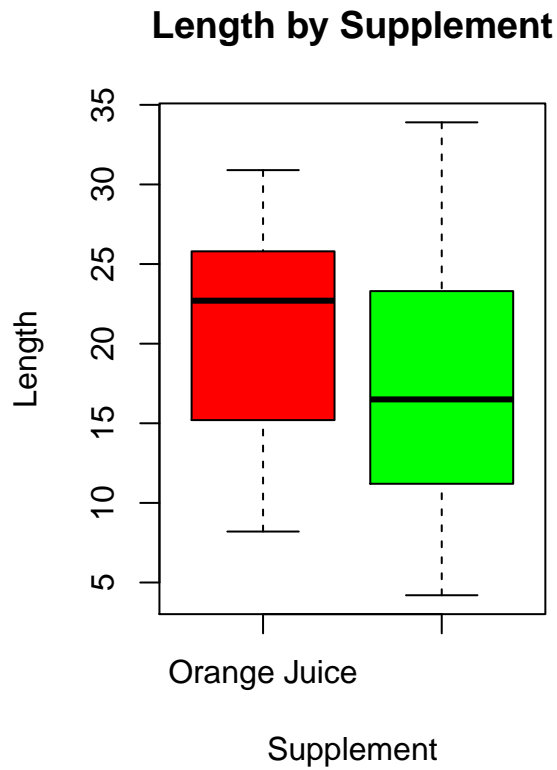
Summary of Data

```
summary(ToothGrowth)
```

```
##      len      supp      dose
## Min.   : 4.20   Orange Juice:30  0.5:20
## 1st Qu.:13.07   Vitamin C   :30  1   :20
## Median :19.25                2   :20
## Mean   :18.81
## 3rd Qu.:25.27
## Max.   :33.90
```

Let's look at some boxplots of the length by supplement and by dose.

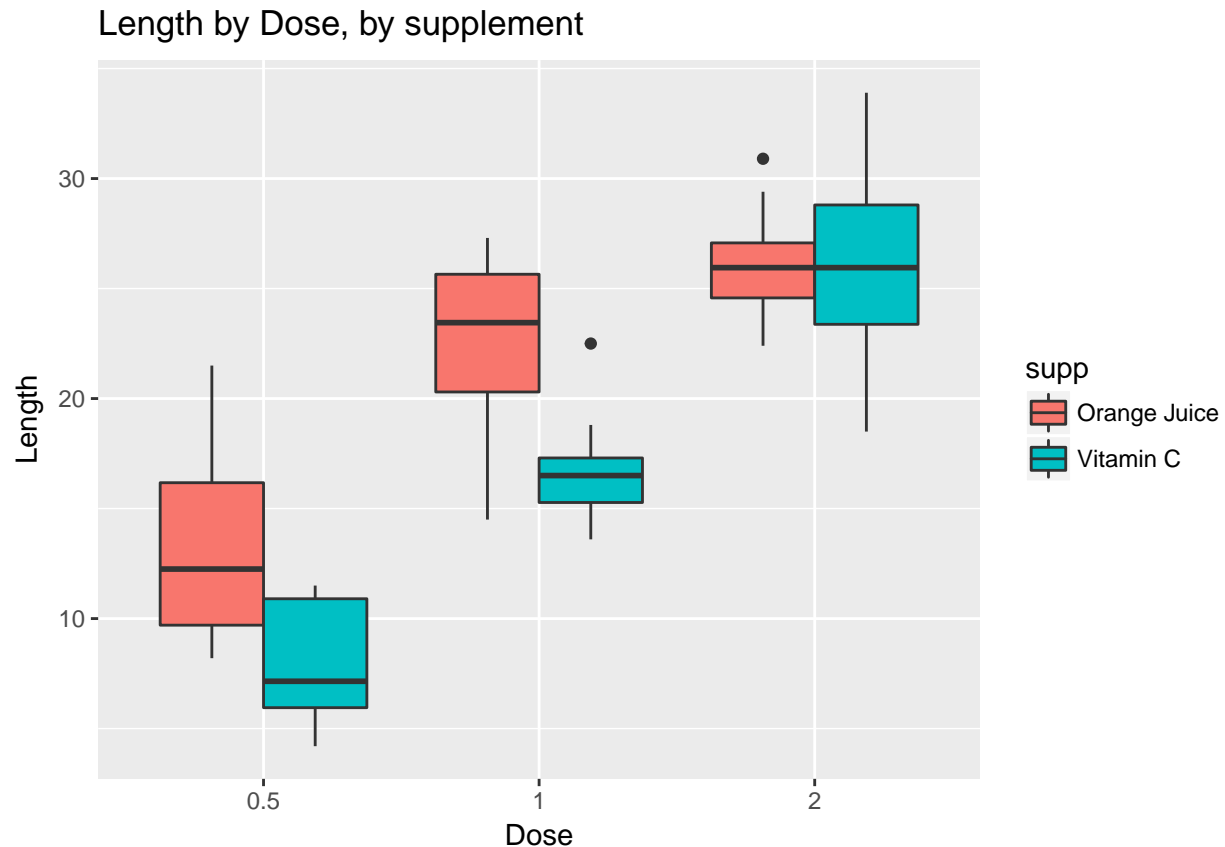
```
par(mfrow=c(1,2))
boxplot(ToothGrowth$len ~ ToothGrowth$supp, col=c("red", "green"))
title(main="Length by Supplement", xlab="Supplement", ylab="Length")
boxplot(ToothGrowth$len ~ ToothGrowth$dose, col=c("yellow", "orange", "red"), xlab="Dose", ylab="Length")
title(main="Length by Dose")
```



The Length by Dose plot above indicates that there is a clear relationship between the dose of the supplement and the length. The Length by Supplement suggests a relationship, however it seems to be much less clear as the medians of each are contained within the 1st to 3rd quartiles of the other.

Let's take a closer look at the breakdown of the length data:

```
ggplot(ToothGrowth, aes(x=dose,y=len)) + geom_boxplot(aes(fill=supp)) + labs(title="Length by Dose, by s
```



It appears that at a dose of 2 the two supplements are equally effective, although Vitamin C has a greater variance than Orange Juice. At lower doses Orange Juice appears to be more effective than Vitamin C.

Confidence Testing of Growth by Supplement and Dose

Based on the plots above I will test the null hypotheses that the two supplements provide equal growth at each of the three dosage levels.

Hypthosis 1

Null Hypothesis 1 is that the two supplements provide equivalent growth at a dose of 0.5.

```
lowdose <- subset(ToothGrowth, dose == 0.5)
conf <- t.test(len ~ supp, data=lowdose)
conf$conf.int
```

```
## [1] 1.719057 8.780943
## attr(,"conf.level")
## [1] 0.95
```

The confidence interval does not include 0, the test indicates that the difference in means is not equal to 0 with 95% confidence. The p value of 0.0063586 indicates that we can reject the null hypothesis and conclude that Orange Juice does provide greater growth at a dose of 0.5mg/day.

Hypothesis 2

Null Hypothesis 2 is that the two supplements provide equivalent growth at a dose of 1.0mg/day.

```
middose <- subset(ToothGrowth, dose == 1)
conf2 <- t.test(len ~ supp, data=middose)
conf2$conf.int
```

```
## [1] 2.802148 9.057852
## attr(,"conf.level")
## [1] 0.95
```

Again, the confidence interval does not include 0 with a confidence level of 95%. The p-value of 0.0010384 is lower than the threshold of 0.05, so we can also reject this hypothesis and conclude that Orange Juice also provides greater growth at a dose of 1mg/day.

Hypothesis 3

Null Hypothesis 3 is that the two supplements provide equivalent growth at a dose of 2.0 mg/day.

```
highdose <- subset(ToothGrowth, dose == 2)
conf3 <- t.test(len ~ supp, data=highdose)
conf3$conf.int
```

```
## [1] -3.79807 3.63807
## attr(,"conf.level")
## [1] 0.95
```

For this hypothesis the confidence level does include 0, in fact it seems to center around 0, with 95% confidence. The p value of 0.0063586 is almost double our threshold. We can not reject this hypothesis so conclude that at a dose of 2.0mg/day Orange Juice and Vitamin C provide equivalent benefits to tooth growth.

Conclusions

Based on the T tests, I must conclude that Orange Juice is more effective at improving tooth growth than Vitamin C at doses of 0.5mg/day and 1mg/day. At a dose of 2.0mg/day there does not appear to be a significant difference between the two supplements.

This analysis assumes that the tooth lengths have a normal distribution, which may be a questionable assumption given the high variability of the data.