

Statistical Inference:ToothGrowth exploratory data analysis

Madhavi Pallela

February 5, 2017

Objective :

The purpose of the this data analysis is to analyze the ToothGrowth data set by comparing the guinea tooth growth by supplement and dose. First, I will do exploratory data analysis on the data set. Then I will do the comparison with confidence intervals in order to make conclusions about the tooth growth.

Load the ToothGrowth data and perform some basic exploratory data analyses

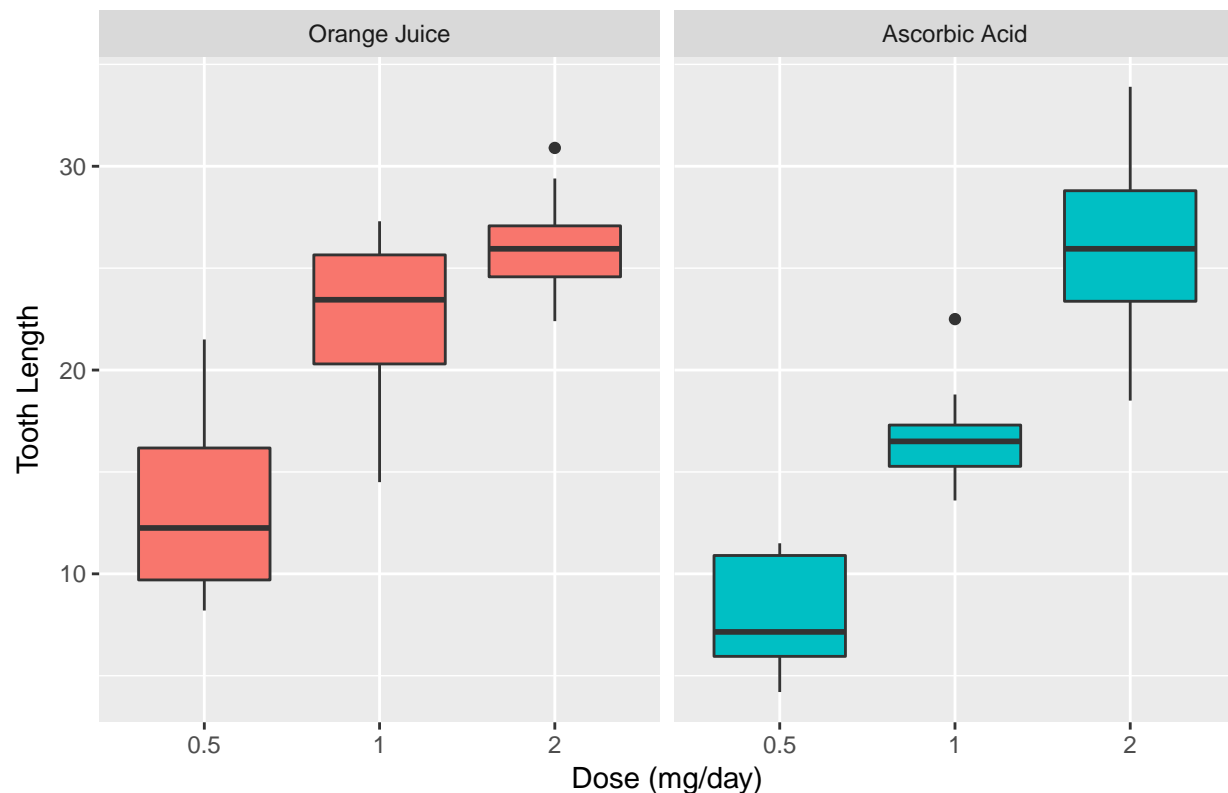
```
library(datasets)
data(ToothGrowth)
str(ToothGrowth)
head(ToothGrowth)
summary(ToothGrowth)
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.3.2
```

```
t = ToothGrowth
levels(t$supp) <- c("Orange Juice", "Ascorbic Acid")
ggplot(t, aes(x=factor(dose), y=len)) +
  facet_grid(.~supp) +
  geom_boxplot(aes(fill = supp), show.legend = FALSE) +
  labs(title="Guinea pig tooth length by dosage for each type of supplement",
       x="Dose (mg/day)",
       y="Tooth Length")
```

Guinea pig tooth length by dosage for each type of supplement



Generate Basic summary of the data

The box plots display increasing the dosage increases the tooth growth. Orange juice is more effective than ascorbic acid for tooth growth when the dosage is .5 to 1.0 milligrams per day. Both types of supplements are equally as effective when the dosage is 2.0 milligrams per day.

Use confidence intervals & hypothesis tests to compare tooth growth by supplement and dose

Hypothesis #1

Orange juice & ascorbic acid result in the same tooth growth across the data set.

```
hypothesis1<-t.test(len ~ supp, data = t)
hypothesis1$conf.int
```

```
## [1] -0.1710156 7.5710156
## attr("conf.level")
## [1] 0.95
```

```
hypothesis1$p.value
```

```
## [1] 0.06063451
```

The confidence intervals includes 0 and the p-value is greater than the threshold of 0.05. The null hypothesis cannot be rejected.

Hypothesis #2

For the dosage of 0.5 mg/day, the two supplements deliver the same tooth growth.

```
hypoth2<-t.test(len ~ supp, data = subset(t, dose == 0.5))  
hypoth2$conf.int
```

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

```
hypoth2$p.value
```

```
## [1] 0.006358607
```

The confidence interval does not include 0 and the p-value is below the 0.05 threshold. The null hypothesis can be rejected. The alternative hypothesis that 0.5 mg/day dosage of orange juice delivers more tooth growth than ascorbic acid is accepted.

Hypothesis #3

For the dosage of 1 mg/day, the two supplements deliver the same tooth growth

```
hypoth3<-t.test(len ~ supp, data = subset(t, dose == 1))  
hypoth3$conf.int
```

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

```
hypoth3$p.value
```

```
## [1] 0.001038376
```

The confidence interval does not include 0 and the p-value is smaller than the 0.05 threshold. The null hypothesis can be rejected and the alternative hypothesis that 1 mg/day dosage of orange juice delivers more tooth growth than ascorbic acid is accepted.

Hypothesis #4

For the dosage of 2 mg/day, the two supplements deliver the same tooth growth

```
hypoth4<-t.test(len ~ supp, data = subset(t, dose == 2))  
hypoth4$conf.int
```

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

```
hypoth4$p.value
```

```
## [1] 0.9638516
```

The confidence interval does include 0 and the p-value is larger than the 0.05 threshold. The null hypothesis cannot be rejected.

Conclusions & assumptions

Orange juice delivers more tooth growth than ascorbic acid for dosages 0.5 & 1.0. Orange juice and ascorbic acid deliver the same amount of tooth growth for dose amount 2.0 mg/day. For the entire data set we cannot conclude orange juice is more effective than ascorbic acid.

Assumptions

- Normal distribution of the tooth lengths
- No other unmeasured factors are affecting tooth length