Statistical Inference Project - Part 2 - Basic Inferential Data Analysis

Tim Westran

Dec 27, 2018

Part 2: Basic Inferential Data Analysis Instructions

Now in the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package. We will analyze the `ToothGrowth` data.

Load Libraries and data

```
## Let's load some libraries we need.
library(data.table)
library(ggplot2)
library(datasets)
data(ToothGrowth)
```

Basic information about the data

```
## Dimensions of dataset
dim(ToothGrowth)
```

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

```
## Structure of dataset
str(ToothGrowth)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
## Are there any missing values?
sum(is.na(ToothGrowth))
```

```
## [1] 0
```

```
## What are the first 10 values?
head(ToothGrowth, 10)
```

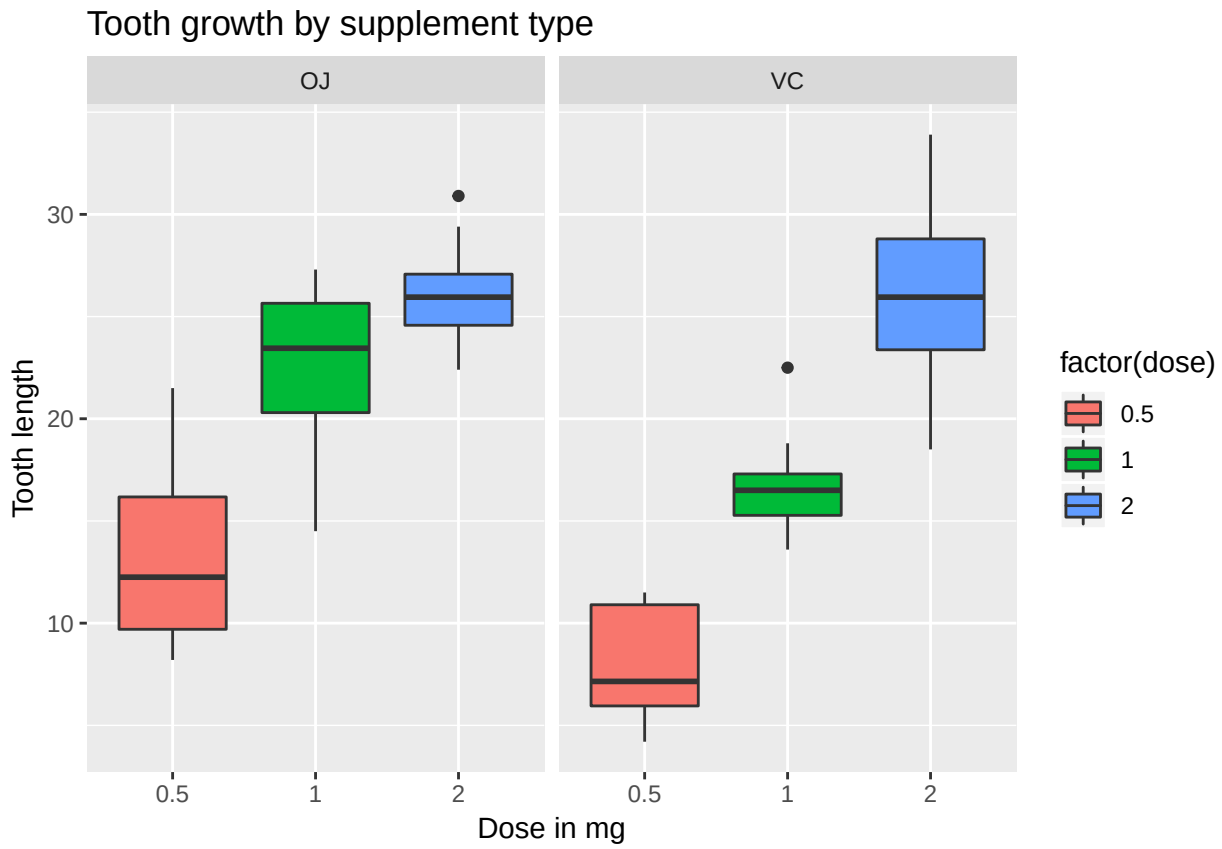
```
##     len supp dose
## 1   4.2   VC  0.5
## 2  11.5   VC  0.5
## 3   7.3   VC  0.5
## 4   5.8   VC  0.5
## 5   6.4   VC  0.5
## 6  10.0   VC  0.5
## 7  11.2   VC  0.5
## 8  11.2   VC  0.5
## 9   5.2   VC  0.5
## 10  7.0   VC  0.5
```

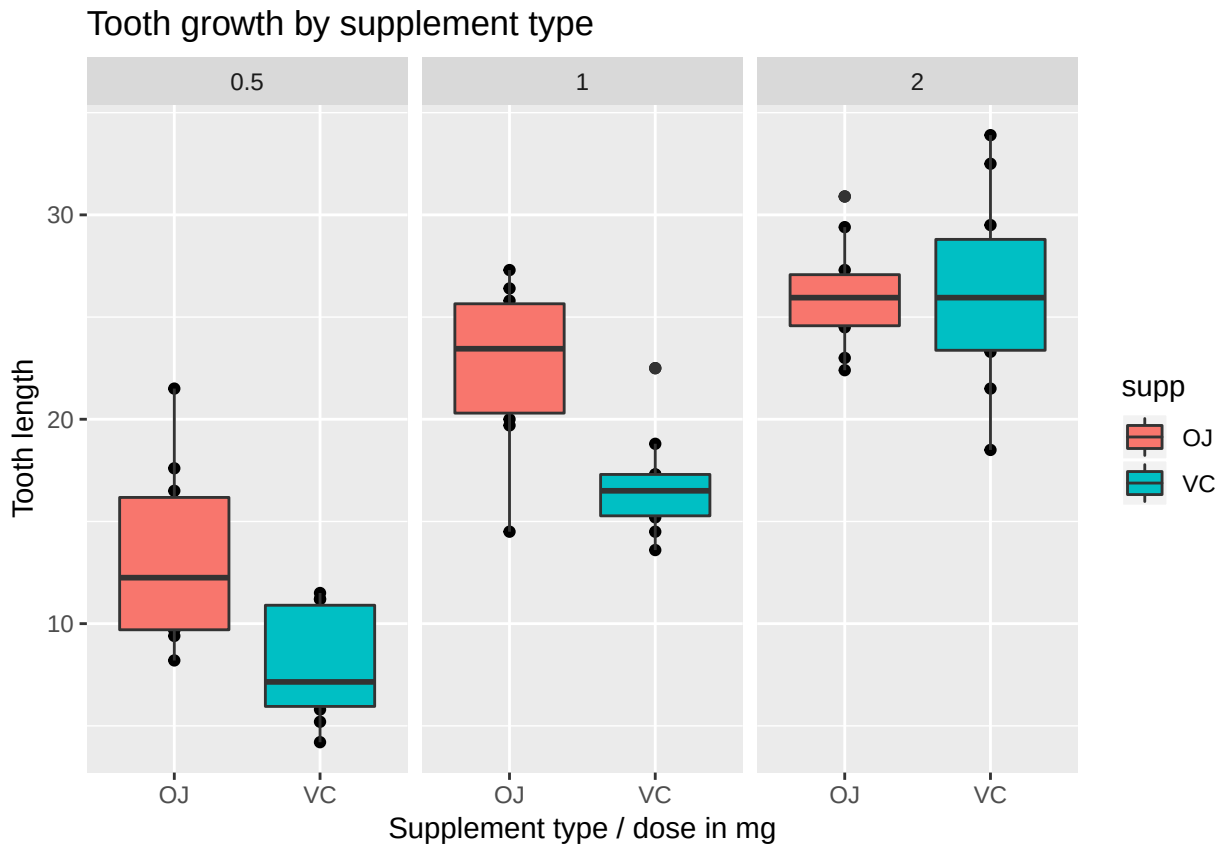
```
## What are some basic statistics of the data?
summary(ToothGrowth)
```

```
##       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
```

From the basic data analysis above, we see that the data contains 60 observations, with 3 variables. We see that there are no NAs for the data set, there are 30 samples for both VC and OJ, and the mean is around 19.

Now, let's see if there is a correlation between dosage of supplement and tooth growth.





From the above plots, we can see that the dosage does positively correlated with tooth growth. The larger the dosage, the longer the tooth.

From the two plots, we see that the OJ supplement seems to promote more tooth growth over VC, for .5 and 1mg doses.

Confidence intervals and hypothesis testing

Now, let's verify if there is a correlation between tooth size and supplement dosage. We want to see how much variance in tooth length can be attributed to supplement dosage.

Tooth growth and dosage

To begin, let's look at each group of data, by dosage, individually. We need to take three subsets of the data to calculate statistics on each individual group, then run a T-test on each group to get P vals and confidence intervals. Basically, we want to see if there is a statiatically significant difference between the different supplements at various dosages, to determine if VC or OJ is the best supplement.

T-test for dose of .5:

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
```

```
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

T-test for dose of 1:

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

T-test for dose of 2:

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

For a dose of .5, we obtain a P value of 0.006359, and a 95% confidence interval of 1.719057 to 8.780943

For a dose of 1, we obtain a P value of 0.001038, and a 95% confidence interval of 2.802148 to 9.057852

For a dose of 1, we obtain a P value of 0.9639, and a 95% confidence interval of -3.79807 to 3.63807

As such, we can reject the null hypothesis (that dosage has no impact on tooth growth) for the .05 and 1 mg dose with 95% confidence, however we cannot reject the null for dose of 2mg. We must conclude that, at the 2mg dose, there is no difference between the two supplements with respect to tooth growth.

Tooth growth and supplement type

Now, let's determine if there is a correlation between tooth growth and higher dosages for each supplement. Basically, we want to see if there is a statistically significant correlation between higher dosages of each supplement and tooth growth.

For Vitamin C:

```
##
##  Welch Two Sample t-test
##
## data:  VC.5$len and VC1$len
## t = -7.4634, df = 17.862, p-value = 6.811e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.265712  -6.314288
## sample estimates:
```

```
## mean of x mean of y
##      7.98     16.77
```

The 95% confidence interval for the T-test between the .5 and 1 mg doses is -11.265712 to -6.314288. As such, we can reject the null hypothesis that increasing the dose from .5 to 1 mg of vitamin C has no impact on tooth growth, as the confidence interval does not include 0.

```
##
##  Welch Two Sample t-test
##
## data:  VC1$len and VC2$len
## t = -5.4698, df = 13.6, p-value = 9.156e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.054267  -5.685733
## sample estimates:
## mean of x mean of y
##     16.77     26.14
```

The 95% confidence interval for the T-test between the 1 and 2 mg doses is -13.054267 to -5.685733. As such, we can reject the null hypothesis that increasing the dose from 1 to 2 mg of vitamin C has no impact on tooth growth, as the confidence interval does not include 0.

For both of the above T-tests, we can conclude that an increase in dosage of Vitamin C, up to 2mg, has a positive impact on tooth growth.

For OJ:

```
##
##  Welch Two Sample t-test
##
## data:  OJ.5$len and OJ1$len
## t = -5.0486, df = 17.698, p-value = 8.785e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.415634  -5.524366
## sample estimates:
## mean of x mean of y
##     13.23     22.70
```

The 95% confidence interval for the T-test between the .5 and 1 mg doses is -13.415634 to -5.524366. As such, we can reject the null hypothesis that increasing the dose from .5 to 1 mg of OJ has no impact on tooth growth, as the confidence interval does not include 0.

```
##
##  Welch Two Sample t-test
##
## data:  OJ1$len and OJ2$len
## t = -2.2478, df = 15.842, p-value = 0.0392
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -6.5314425 -0.1885575
## sample estimates:
## mean of x mean of y
##     22.70     26.06
```

The 95% confidence interval for the T-test between the 1 and 2 mg doses is -6.5314425 to -0.1885575. As such, we can reject the null hypothesis that increasing the dose from 1 to 2 mg of OJ has no impact on tooth

growth, as the confidence interval does not include 0.

For both of the above T-tests, we can conclude that an increase in dosage of OJ, up to 2mg, has a positive impact on tooth growth.

Thus, for both the OJ and VC supplements, we can conclude, based on both the 95% confidence interval and the P-values, that increasing supplement dosage up to 2mg has a positive impact on tooth growth.

Conclusion

OJ at .5 and 1mg appears to have a better impact on positive tooth growth over VC. However, both OJ and VC at 2mg seem to have the same impact on positive tooth growth. Increasing supplement dosage does seem to increase tooth size.

We are assuming that the sample data is not paired, and that the sample populations are different, with different variances.