

Statistical Inference Part2

Vamshi Krishna

12/11/2019

Synopsis: Now in the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package. The response is the length of tooth when they receive one of the 2 supplements namely OJ (Orange Juice) and VC (Vitamin C) and one of the 3 doses (0.5, 1, 2) on total 60 animals.

Instructions:

1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)
4. State your conclusions and the assumptions needed for your conclusions.

1. Loading necessary packages and data, summarizing the ToothGrowth data

```
library(ggplot2)

library(datasets)
data("ToothGrowth")
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 ...
```

There are two types of supplements and 3 types of doses.

Lets check for the null values present in the data

```
table(is.na(ToothGrowth$len))
```

```
##
## FALSE
##      60
```

There are no null values in the length column

Lets calculate the mean of each supplements

```
vc = subset(ToothGrowth, supp %in% "VC")
vc_mean = mean(vc$len)
```

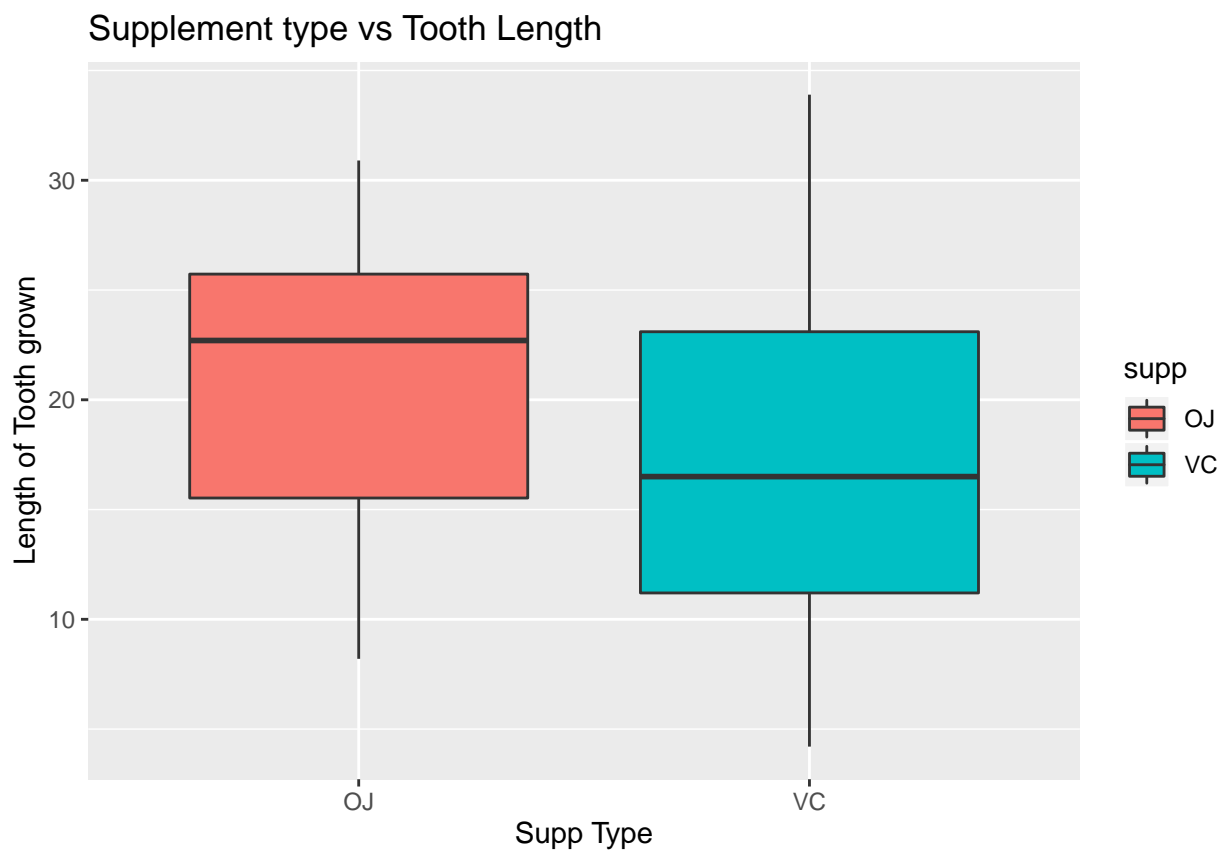
Mean of VC is 16.9633333

```
oj = subset(ToothGrowth, supp %in% "OJ")
oj_mean = mean(oj$len)
```

Mean of OJ is 20.6633333. oj_mean is greater than vc_mean.

Lets draw a boxplot to show each supplement

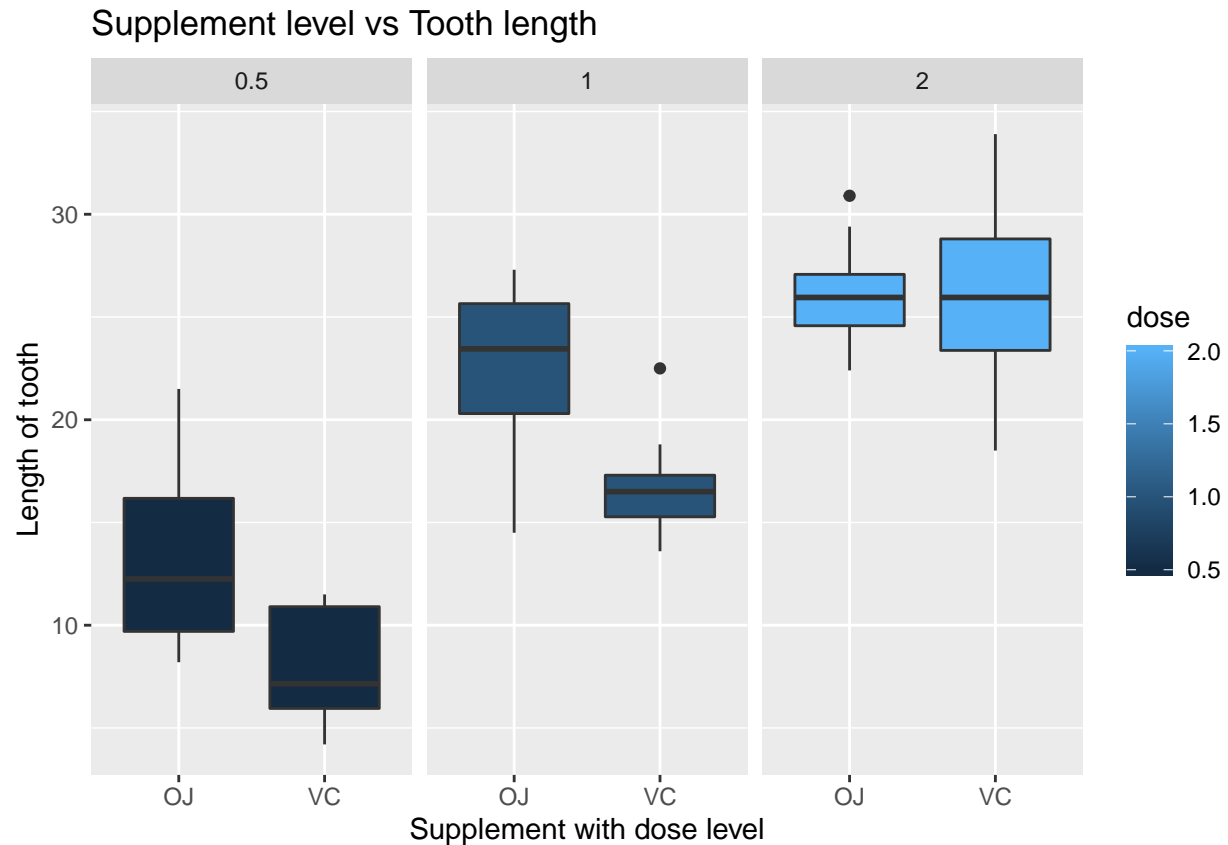
```
g1 = ggplot(data=ToothGrowth, mapping = aes(supp,len)) + geom_boxplot(aes(fill = supp)) + xlab("Supp Ty")
print(g1)
```



This graph shows the growth of tooth depending on the given supplement to subject. we can conclude that the overall growth is higher with OJ when compared to VC.

Lets plot one more plot to show the difference in growth of the tooth with respect to dose:

```
g2 = ggplot(data = ToothGrowth, aes(supp,len)) + geom_boxplot(aes(fill = dose)) + xlab("Supplement with")
print(g2)
```



From the above plot we can conclude that mean growth of tooth length is high with supplement OJ and dose 0.5 and 1 but in case of 2 as dose level the mean growth is high with supplement VC with a little difference.

2. Lets see the basic summary of data:

Lets see how the data is with the help of head function

```
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
```

There are 3 columns and length depends on the supplement and dosage level.

Lets also see the summary of the dataframe

```
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
```

3. Lets use t.test to find confidence level and p value to compare tooth growth

Lets do a normal t-test initially for VC and OJ lengths irrespective of dose:

```
t.test(vc$len,oj$len,paired = FALSE)
```

```
##
## Welch Two Sample t-test
##
## data: vc$len and oj$len
## 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:
## -7.5710156 0.1710156
## sample estimates:
## mean of x mean of y
## 16.96333 20.66333
```

We can see the p-value is near to 0.05 and confidence interval contains 0 in it and mean length is greater when supplement is OJ overall.

Now, lets do t-test with respect to dosage level:

```
t.test(len~supp, data = ToothGrowth[ToothGrowth$dose==0.5,], paired = FALSE)
```

```
##
## 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(len~supp, data = ToothGrowth[ToothGrowth$dose==1,], paired = FALSE)
```

```
##
## 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(len~supp, data = ToothGrowth[ToothGrowth$dose==2,], paired = FALSE)
```

```
##
## 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
```