Statistical Inference Course Project pt.2

Junaid Khan

March 25, 2018

In second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.

Loading the dataset

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

Summary 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
```

```
str(ToothGrowth)
```

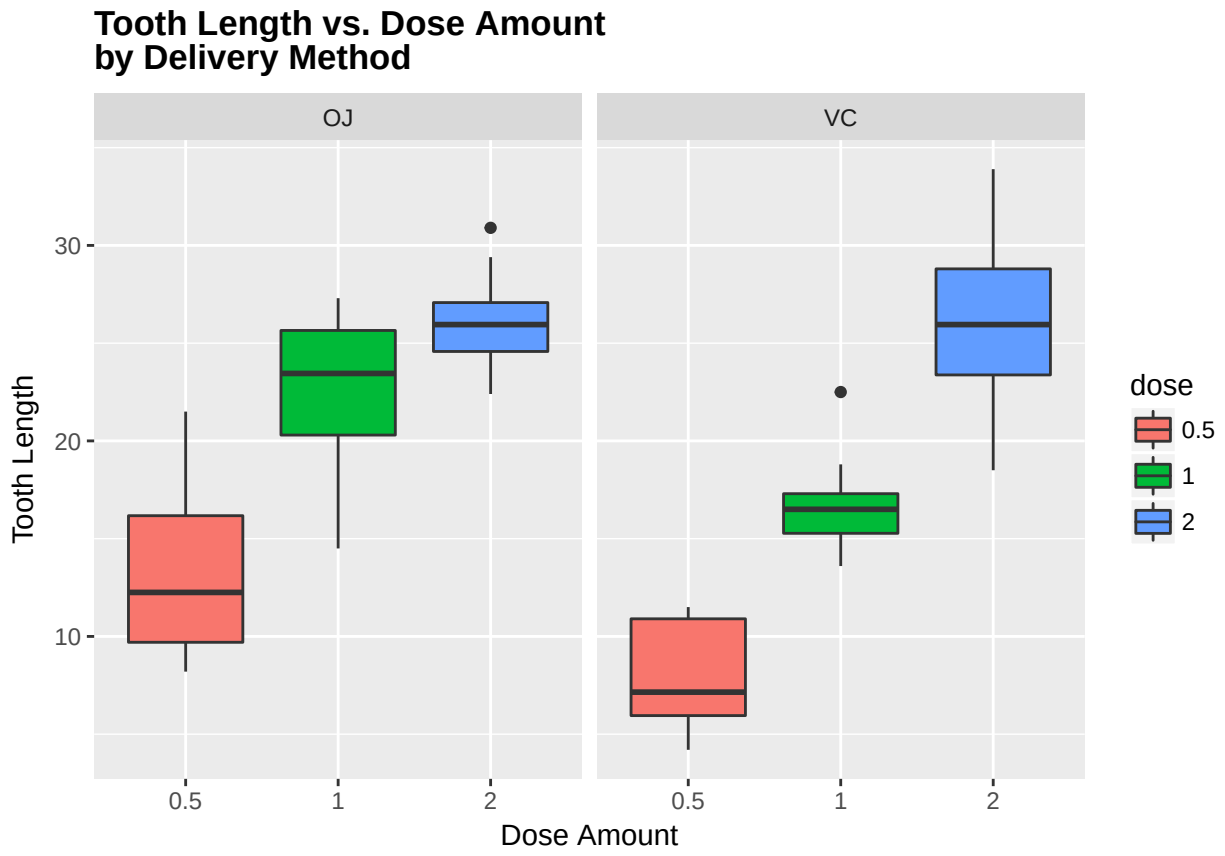
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

Thus we have the basic summary of the ToothGrowth dataset, i.e The data types and the quantiles of the different columns.

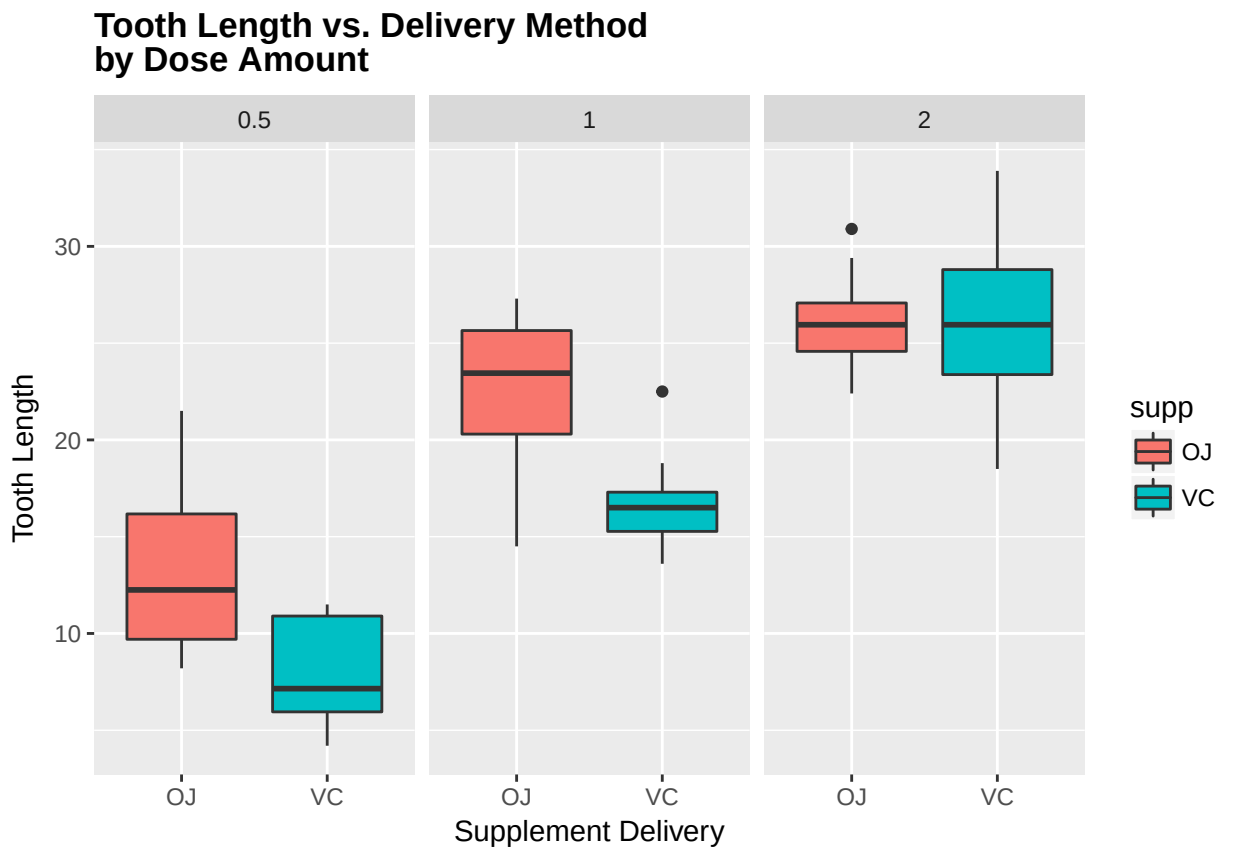
some basic exploratory data analyses

```
library(ggplot2)
```

```
ToothGrowth$dose<-as.factor(ToothGrowth$dose)
ggplot(aes(x=dose, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=dose)) + xlab("Dose Amount") + ylab
     theme(plot.title = element_text(lineheight=.8, face="bold"))
```




```
ggplot(aes(x=supp, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=supp)) + xlab("Supplement Delivery
     theme(plot.title = element_text(lineheight=.8, face="bold"))
```



Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose

T-tests :

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

```
##
##  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
```

Since the p-value of this test is greater than 0.05 and the confidence intervals contains zero, we can say that supplement types do not have an impact on the lenght of the teeth

Now, we will run t-tests on len and subsets of dose.

```
ToothGrowth_1 <- subset(ToothGrowth, ToothGrowth$dose %in% c(0.5,1.0))
t.test(len~dose,data = ToothGrowth_1)
```

```
##
##  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
```

```
ToothGrowth_2 <- subset(ToothGrowth, ToothGrowth$dose %in% c(1.0,2.0))
t.test(len~dose,data = ToothGrowth_2)
```

```
##
##  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
```

As we can see from the above tests, the p-value is essentially zero

and the confidence interval never crosses zero

Conclusion

Given the following assumptions: 1.The sample is representative of the population 2.The distribution of the sample means follows the Central Limit Theorem

In reviewing our t-test analysis from above, we can conclude that supplement delivery method has no effect on tooth growth/length, however increased dosages do result in increased tooth length.