

Statistical Inference course project02

Fan Ouyang

Overview

This is an analysis of the ToothGrowth data in the R datasets package. ### Load the ToothGrowth data and perform some basic exploratory data analyses

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

Provide a basic summary of the data.

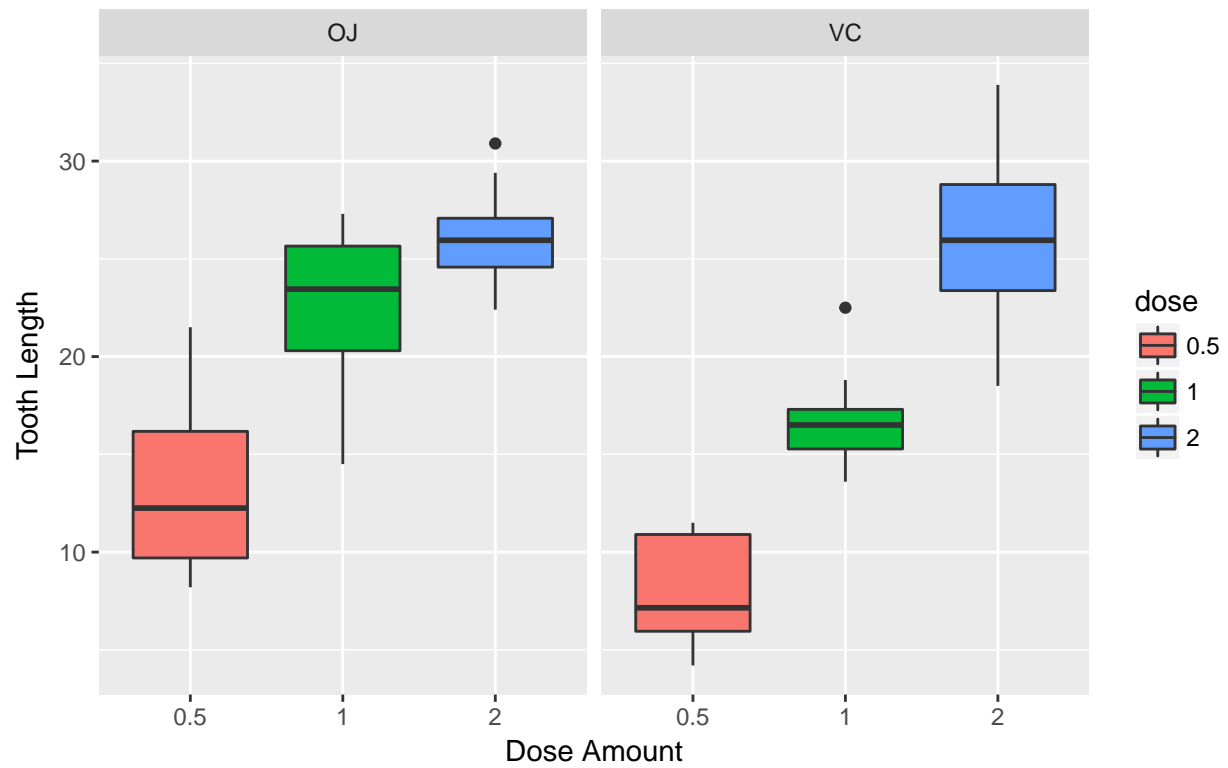
```
summary(ToothGrowth)
```

##	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

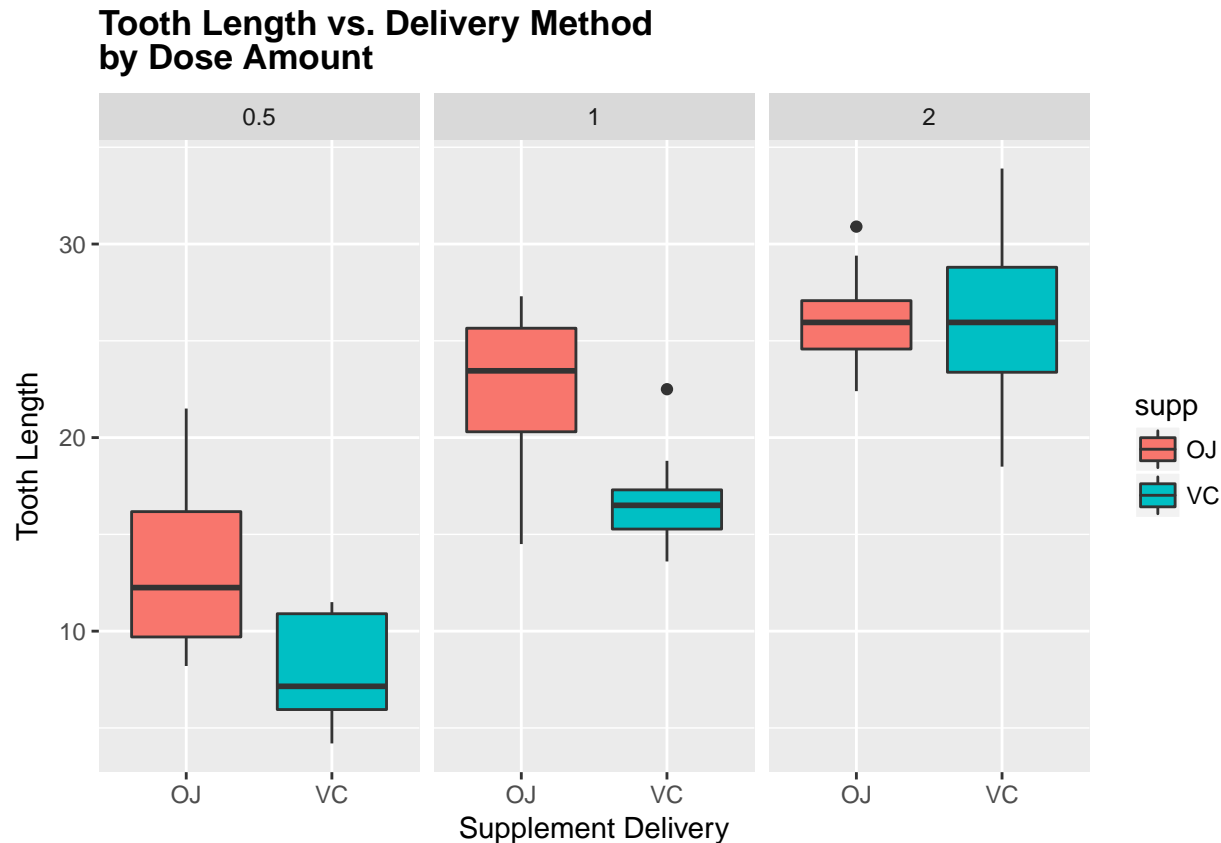
exploratory analysis of data.

```
library(ggplot2)
ToothGrowth$dose<-as.factor(ToothGrowth$dose)
par(mfrow=c(1,2))
ggplot(aes(x=dose, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=dose)) + xlab("Dose Amount") + ylab("len")
```

Tooth Length vs. Dose Amount by Delivery Method



```
ggplot(aes(x=supp, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=supp)) + xlab("Supplement Delivery")
```



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

compare tooth growth by supplement using a t-test. The p-value of this test was 0.06. Since the p-value is greater than 0.05, we could conclude that supplement types seems to have no impact on Tooth growth based on this test.

```
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
```

compare tooth growth by dose, looking at the different pairs of dose values. As can be seen, the p-value of each test was essentially zero and the confidence interval of each test does not cross over zero (0). Based on this result we can assume that the average tooth length increases with an increasing dose, and the null hypothesis can be rejected.

```
# run t-test using dose amounts 0.5 and 1.0
ToothGrowth_sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(1.0,0.5))
```

```

t.test(len~dose,data=ToothGrowth_sub)

##
##  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
# run t-test using dose amounts 0.5 and 2.0
ToothGrowth_sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(0.5,2.0))
t.test(len~dose,data=ToothGrowth_sub)

##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5    mean in group 2
##      10.605          26.100
# run t-test using dose amounts 1.0 and 2.0
ToothGrowth_sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(1.0,2.0))
t.test(len~dose,data=ToothGrowth_sub)

##
##  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

```

Conclusions and the assumptions needed for conclusions.

Conclusions 1. Supplement type has no effect on tooth growth. 2. Increasing the dose level leads to increased tooth growth.

Given the following assumptions: 1. The sample is representative of the population 2. The distribution of the sample means follows the Central Limit Theorem 3. 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.