Statistical inference project 2

YOUNGSANG KIM

EDA

```r
library(dplyr, warn.conflicts = F)
library(ggplot2)
data("ToothGrowth")
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  
```

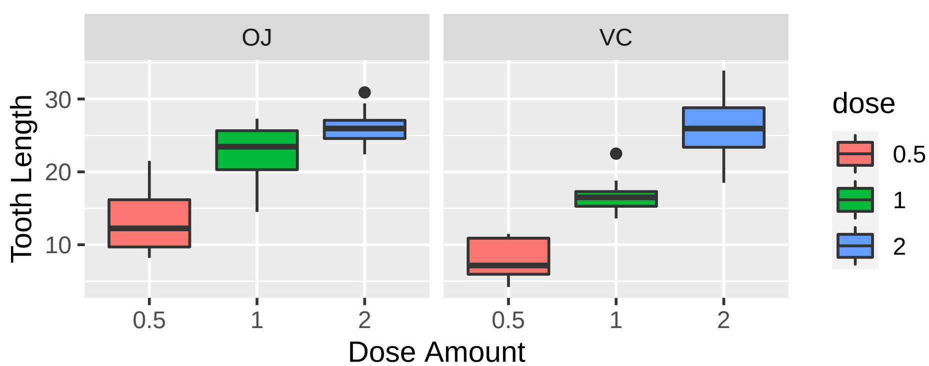
```r
str(ToothGrowth)
```

```
## '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 ...
```

```r
ToothGrowth$dose<-as.factor(ToothGrowth$dose) # Convert dose to a factor
```

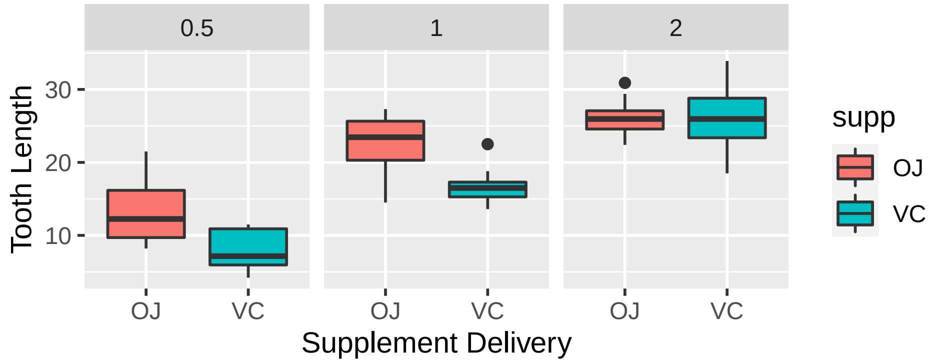
Plot tooth length ('len') vs. the dose amount ('dose'), broken out by supplement delivery method ('supp')

```r
ggplot(aes(x=dose, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=dose)) +
  xlab("Dose Amount") + ylab("Tooth Length") + facet_grid(~ supp)
```



Plot tooth length ('len') vs. supplement delivery method ('supp') broken out by the dose amount ('dose')

```r
ggplot(aes(x=supp, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=supp)) +
  xlab("Supplement Delivery") + ylab("Tooth Length") + facet_grid(~ dose)
```



Hypothesis Tests

Now we want to further compare teeth growth by supplement type and dose levels.
This time we'll use statistical tests, t-test.

```r
len_1 <- ToothGrowth %>% filter(dose %in% c(0.5,1)) %>% select(len) %>% unlist()
dose_1 <- ToothGrowth %>% filter(dose %in% c(0.5,1)) %>% select(dose) %>% unlist()
t.test(len_1 ~ dose_1, paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len_1 by dose_1
## 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
```

```r
len_2 <- ToothGrowth %>% filter(dose %in% c(0.5,2)) %>% select(len) %>% unlist()
dose_2 <- ToothGrowth %>% filter(dose %in% c(0.5, 2)) %>% select(dose) %>% unlist()
t.test(len_2 ~ dose_2, paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len_2 by dose_2
## 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
```

```r
len_3 <- ToothGrowth %>% filter(dose %in% c(1,2)) %>% select(len) %>% unlist()
dose_3 <- ToothGrowth %>% filter(dose %in% c(1,2)) %>% select(dose) %>% unlist()
t.test(len_3 ~ dose_3, paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len_3 by dose_3
## 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
```

We went through all possible combinations of levels.
Thus, we reject Ho. In other words there appears to be a positive relationship between dose level and teeth length

```r
len <- ToothGrowth %>% select(len) %>% unlist()
supp <- ToothGrowth %>% select(supp) %>% unlist()
t.test(len~supp, paired=F)
```

```
##
##  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
```

The p-value is 0.06, and the confidence interval contains zero.
This indicates that we can not reject the null hypothesis that the different supplement types have no effect on tooth length.

Conclusions

A. Increasing the dose level leads to increased tooth growth.
B. Supplement type has no effect on tooth growth.
Given the following assumption:
A. The sample is representative of the population.
B. The distribution of the sample means follows the Central Limit Theorem.