Part2: Basic Inferential Data Analysis Instructions

Minki Jo

November, 19, 2017

Overview

In the second portion of the project, ToothGrowth data of datasets package in R is analyzed.

Load data

Load libraries that are used for this part in advance and get ToothGrowth data

```
library(datasets)
library(ggplot2)
data <- ToothGrowth
```

Summary of the data

Summary and dim function show how original data looks like. Below are how these data look like.

```
dim(data)
```

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

```
names(data)
```

```
## [1] "len"  "supp" "dose"
```

```
summary(data)
```

```
##       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
```

ToothGrowth data is a data.frame format with 60 observations on 3 variables. Three variables each are len, supp, and dosetooth. From the ?ToothGrowth, we can see what each variable mean (tooth length, supplement type, and dose in milligrams/day respectively)

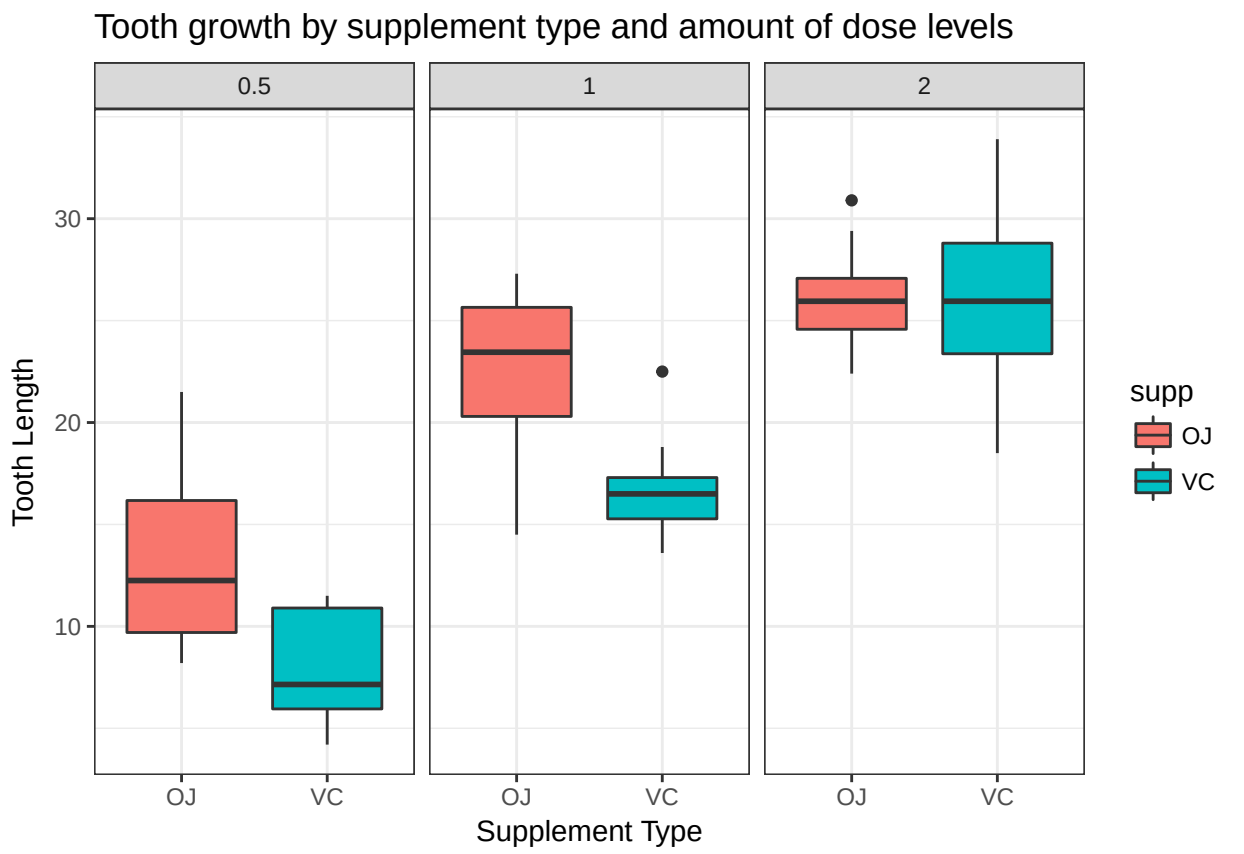
The ranges for len and dose are from 4.2 to 33.9 and from 0.5 to 2 respectively. But variable "dose" only has integer number between its range. The another variable "suppp" consists of two factors such as OJ(Orange Juice), VC(ascorbic acid).

Exploratory Data Analyses

Let's plot of len versus dose by supp factor.

```
g <- ggplot(data = data, aes(x = factor(supp), y = len))
g <- g + geom_boxplot(aes(fill = supp)) + facet_grid(~dose)
```

```
g <- g + xlab("Supplement Type") + ylab("Tooth Length") + ggtitle("Tooth growth by supplement type and a
g
```



Hypothesis test and confidence interval

As looking at boxplot above, the effectivenss of supplement type and dose levels should be looked into. All hypothesis tests in this report use 0.05 significant level.

First, let's see how supplement type affects tooth growth in general

```
t.test(len ~ supp, paired = FALSE, var.equal = TRUE, data=data)
```

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1670064  7.5670064
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

Thre result shows that it fails to reject null hypothesis p-value > 0.05), which means there is **no impact on tooth growth by supplement type.** In order to conclude such that, it is needed to take more tests that show there is still no diffrence for supplement type by dose level.

```r
t.test(len~supp,paired = FALSE, var.equal = TRUE, data=subset(data, dose == 0.5))
```

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 18, p-value = 0.005304
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.770262 8.729738
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

```r
t.test(len~supp,paired =FALSE, var.equal= TRUE, data=subset(data, dose == 1))
```

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 18, p-value = 0.0007807
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.840692 9.019308
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

```r
t.test(len~supp,paired = FALSE, var.equal = TRUE, data=subset(data, dose == 2))
```

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 18, p-value = 0.9637
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.722999  3.562999
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

From the above hypothesis test results, we can say **there is difference between OJ and VC for dose level of 0.5 and 1 milligrams/day**, which contradicts initial conclusion above. **For 2 milligrams/day dose level, there is still no difference.** (see 95 percent confience interval for diffrence from -.3.7 to 3.6)

Now, it it time to test dose level affects tooth growth. Since hypothysis test for dose level betwen 0.5 and 1, and between 1 and 2 can cover dose level of 0.5 and 2, it is not necessary to see the difference for dose level of 0.5 and 2.

```r
t.test(len ~ dose, paired = FALSE, var.equal = TRUE, data = subset(data, dose %in% c(0.5,1.0)))
```

```
##
##  Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 38, p-value = 1.266e-07
```

```
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983748  -6.276252
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

```
t.test(len ~ dose, paired = FALSE, var.equal = TRUE, data = subset(data, dose %in% c(1.0,2.0)))
```

```
##
##  Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 38, p-value = 1.811e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.994387 -3.735613
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

The results shows there is significant difference for given dose levels.

Conclusion

Hypothesis test is assumed that 60 observations are randomly sampled so that it can be a representative of the population. Also, it is assumed that distribution of sample means follows the Central Limit Theorem

Based on those assumptions, we can conclude that :

- Increased dose level results in increased tooth length.
- Supplement type has no effect when dose level is 2.0, but it has when it is 0.5 and 1.0