Tooth Growth Data - Part 2

1. Synopsis

R version used for analysis is "R version 3.3.1 (2016-06-21)". OS used is "x86_64, Windows 10 64 Bit".

2. Basic settings

```
echo = TRUE  # Make code always visible
options(scipen = 1)  # Turn off scientific notations for numbers
```

3. Loading the libraries and data

```
library(datasets)
library(ggplot2)
data("ToothGrowth")
```

4. Exploratory Data Analysis and Summary of data

```
# check the dataset structure
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 ...
```

```
#check the number of rows in dataset
nrow (ToothGrowth)
```

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

```
#check the header rows
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
```

```r
#check the footer rows
tail (ToothGrowth)
```

```
##     len supp dose
## 55 24.8   OJ    2
## 56 30.9   OJ    2
## 57 26.4   OJ    2
## 58 27.3   OJ    2
## 59 29.4   OJ    2
## 60 23.0   OJ    2
```

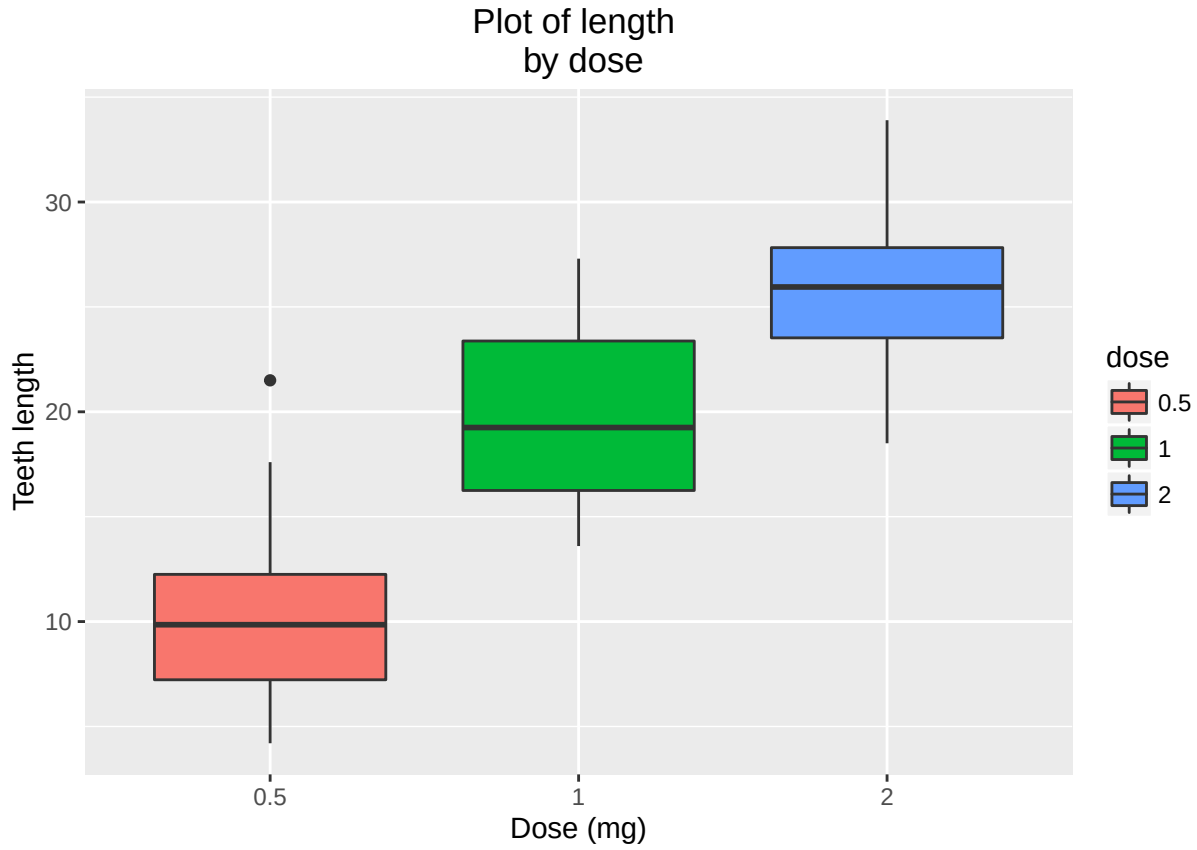
```r
#check the dataset summary
summary (ToothGrowth)
```

```
##       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
```

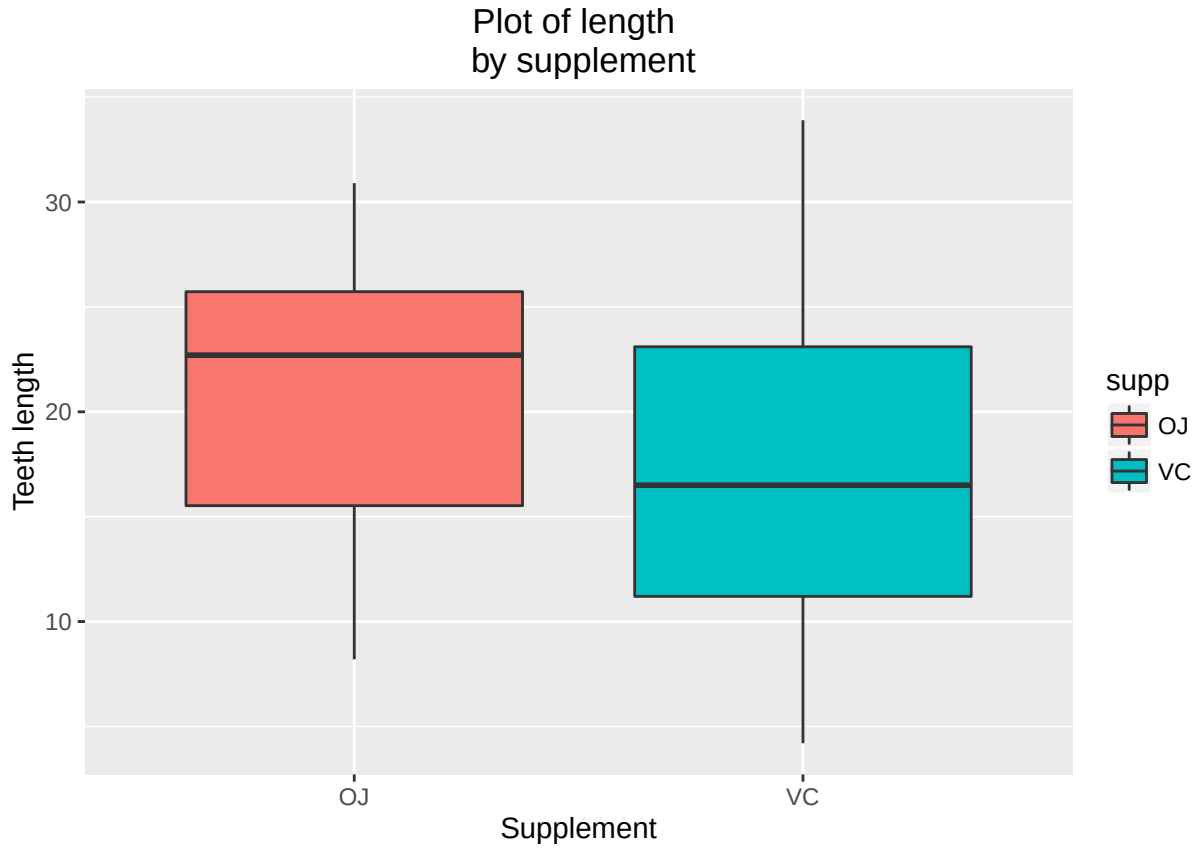
```r
ToothGrowth$dose<-as.factor(ToothGrowth$dose)

# check the summary of cases between different dose levels and delivery methods
table(ToothGrowth$dose, ToothGrowth$supp)
```

```
##
##       OJ VC
##   0.5 10 10
##   1   10 10
##   2   10 10
```

```r
# tooth length vs Dose plot
p = ggplot(aes(x=dose, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=dose))
p = p + ggtitle("Plot of length \n by dose")
p = p + xlab("Dose (mg)") + ylab("Teeth length")
p
```




```
p = ggplot(aes(x=supp, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=supp))
p = p + ggtitle("Plot of length \n by supplement")
p = p + xlab("Supplement") + ylab("Teeth length")
p
```



5. Statistical Inferential Analysis

5.1 Supplement as a factor

```
# check for group differences due to different supplement type
# assuming unequal variances between the two groups
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
```

The p-value is 0.06 and confidence interval of [-0.171, 7.571] does not allow us to reject the null hypothesis (that there is no correlation between delivery method and tooth length).

5.2 Dosage as a factor

```
# first create three sub-groups as per dose level pairs
ToothGrowth.dose1 = subset (ToothGrowth, dose %in% c(0.5, 1.0))
ToothGrowth.dose2 = subset (ToothGrowth, dose %in% c(0.5, 2.0))
ToothGrowth.dose3 = subset (ToothGrowth, dose %in% c(1.0, 2.0))

# Check for group differences due to different dose levels (0.5, 1.0)
# assuming unequal variances between the two groups
t.test(len ~ dose, data = ToothGrowth.dose1)
```

```
##
##  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
```

```
# Check for group differences due to different dose levels (0.5, 2.0)
# assuming unequal variances between the two groups
t.test(len ~ dose, data = ToothGrowth.dose2)
```

```
##
##  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
```

```
# Check for group differences due to different dose levels (1.0, 2.0)
# assuming unequal variances between the two groups
t.test(len ~ dose, data = ToothGrowth.dose3)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 37.101, p-value = 0.00001906
## 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
```

The confidence intervals ([-11.98, -6.276] for doses 0.5 and 1.0, [-18.16, -12.83] for doses 0.5 and 2.0, and [-8.996, -3.734] for doses 1.0 and 2.0) allow for the rejection of the null hypothesis and a confirmation that there is a significant correlation between tooth length and dose levels.

5.3 Considering Supplement as a Factor within the Dose levels

Analyzing the ddata for correlation between dose level and change in tooth growth within each dose level:

```
ToothGrowth.dose4 <- subset(ToothGrowth, dose == 0.5)
ToothGrowth.dose5 <- subset(ToothGrowth, dose == 1.0)
ToothGrowth.dose6 <- subset(ToothGrowth, dose == 2.0)

t.test(len ~ supp, data = ToothGrowth.dose4)
```

```
##
##  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.dose5)
```

```
##
##  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.dose6)
```

```
##
##  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
```

The confidence intervals for dose levels 0.5mg and 1.0mg([1.72, 8,78] within 0.5mg, [2.80, 9.06] within 1.0mg) allow for the rejection of the null hypothesis and a confirmation that there is a significant correlation between tooth length and dose levels. However, the confidence interval for dose level 2.0[-3.80, 3.64] is not enough to reject the null hypothesis.

6. Conclusions and Assumptions

To make conclusions with the data in this dataset, we must assume that
1. The poplulations are independent
2. The variances between populations are different
3. A random population was used
4. The population was comprised of similar guinea pigs
5. Measurement error was accounted for with significant digits
6. & double blind research methods were used. For the populations to be independent, 60 guinea pigs would have to be used so each combination of dose level and delivery method were not affected by the other methods. To ensure double blind research methods are followed, the researchers taking the measurements must have been unaware of which guinea pigs were given which dose level or delivery method. The guinea pigs must also be unaware that they are being given a specific treatment
7. For the t-tests, the variances are assumed to be different for the two groups being compared. This assumption is less stronger than the case in which the variances are assumed to be equal

If all the preceding assumptions are true, we may infer that
1. there is a significant difference between tooth length and dose levels across both delivery methods
2. A higher dose level consistently led to longer teeth
3. Initially it appeared that the delivery method had no significant impact on tooth length, but when controlling for dose level we discovered that there was a significant difference at 0.5mg and 1.0mg, but not at 2.0mg
4. Based on the above conclusion, it appears that orange juice is a better delivery method with a larger impact on tooth length for a given dose of Vitamin C, but above a maximum dose level there is no further improvement