SI-Project Assignment Part 2

Analyze the ToothGrowth data in the R datasets package.

Load and inspect the ToothGrowth data

```
library(ggplot2);library(datasets);data(ToothGrowth);head(ToothGrowth);str(ToothGrowth);names(ToothGrowth);dim(ToothGrowth);summary(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
```

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

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

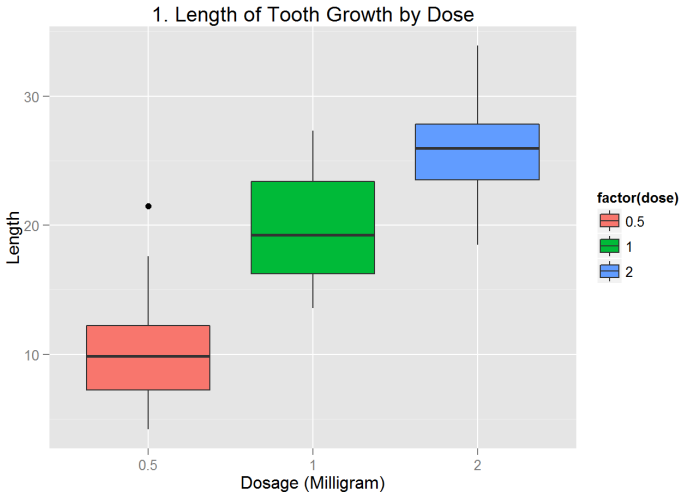
```
## [1] 60  3
```

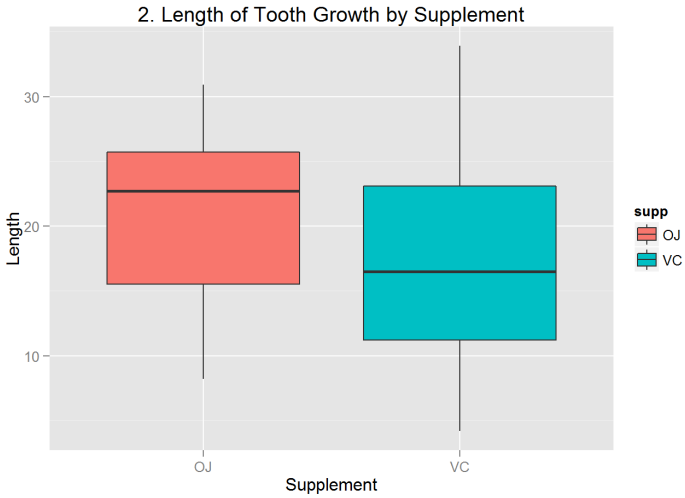
```
##       len       supp         dose
##  Min.   : 4.2   OJ:30   Min.   :0.50
##  1st Qu.:13.1   VC:30   1st Qu.:0.50
##  Median :19.2           Median :1.00
##  Mean   :18.8           Mean   :1.17
##  3rd Qu.:25.3           3rd Qu.:2.00
##  Max.   :33.9           Max.   :2.00
```

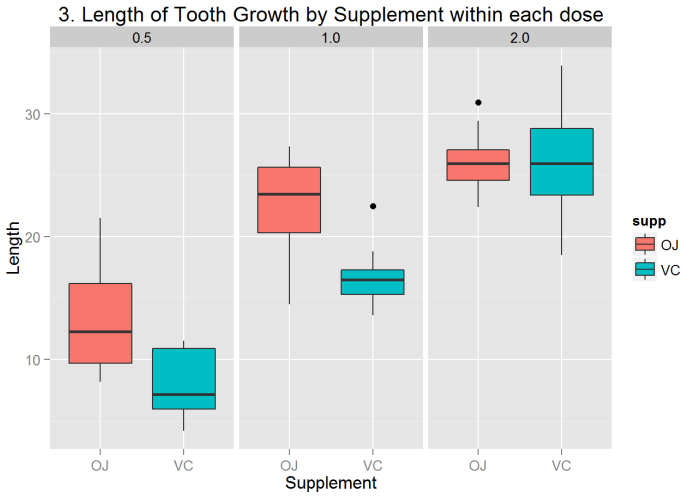
Inspecting the data, it is clear that, the data collected is to know the effect of Vitamin C on Tooth Growth in Guinea Pigs. The response is the length of teeth in 60 guinea pigs at three dose levels of Vitamin C (0.5, 1, and 2 mg) with two delivery methods (orange juice or ascorbic acid). The data frame contains 60 observations on 3 variables. The variables are:

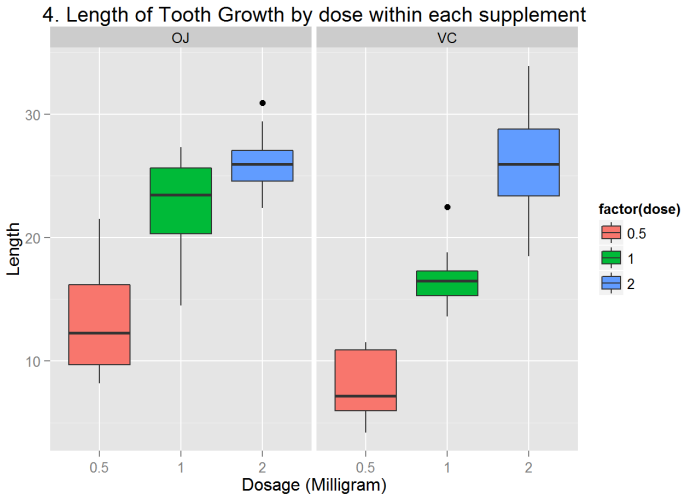
1. len - Tooth Length (a numeric variable) : This is the response variable in this case.
2. supp > Supplement Type (a factor (categorical) variable) > 2 Levels : VC (as Ascorbic Acid) & OJ (as Orange Juice)
3. dose - Dosage of the supplement in milligrams (a factor (categorical) variable) > 3 Levels :(0.5,1.0 and 2.0)

Perform some basic Exploratory Data Analyses (EDA)










```
table(ToothGrowth$supp, ToothGrowth$dose)
```

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

As observed above with simple EDA the dosage appears to affect tooth length, the higher the dosage of a supplement, the longer the tooth length. The supplement type may affect the tooth length, with OJ being higher than VC, but it is difficult to say if the differences are statistically of any significant. Finally, it seems that OJ has larger effect at lower dosages, but at a higher dosage the difference between OJ and VC is minimal, if any at all. Figure 3, and 4 Capturing in essence the same thing.

The study is done to see impact of OJ - VC with different dosage level, as you can see above that each supplement - dosage combinations has same number of set i.e 10.

Confidence Intervals and Hypothesis Testing

Test by Dosage

For these tests, we will ignore the the type of supplement, and see if there is a difference in tooth length based on dosage levels. We create three separate data frames to compare 0.5 vs 1.0, 0.5 vs 2.0, and 1.0 vs 2.0.

```
Tooth.dose51 <- subset(ToothGrowth, dose %in% c(0.5, 1.0))
Tooth.dose52 <- subset(ToothGrowth, dose %in% c(0.5, 2.0))
Tooth.dose12 <- subset(ToothGrowth, dose %in% c(1.0, 2.0))
```

```
t.test(len ~ dose, paired = F, var.equal = F, data = Tooth.dose51)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.477, df = 37.99, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.984  -6.276
## sample estimates:
## mean in group 0.5   mean in group 1
##             10.61             19.73
```

```
t.test(len ~ dose, paired = F, var.equal = F, data = Tooth.dose52)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.8, df = 36.88, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.16 -12.83
## sample estimates:
## mean in group 0.5   mean in group 2
##             10.61             26.10
```

```
t.test(len ~ dose, paired = F, var.equal = F, data = Tooth.dose12)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.901, df = 37.1, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996 -3.734
## sample estimates:
## mean in group 1 mean in group 2
##           19.73           26.10
```

- There is a statistically significant difference between the means of 0.5 and 1mg dose, and 95% confidence interval of [-11.98, -6.276] does not include 0, and pvalue is less than 0.05, hence we can reject the null hypothesis and say that there is a significant difference in tooth length between dosages of 0.5 mg and 1.0 mg (1.0 mg dosage creates longer teeth in guinea pigs).
- There is a statistically significant difference between the means of 0.5 and 2 mg dose, and 95% confidence interval of [-18.16, -12.83] does not inlcude 0 and pvalue is less than 0.05, hence we can reject the null hypothesis and say that there is a significant difference in tooth length between dosages of 0.5 mg and 2.0 mg (2.0 mg dosage creates longer teeth in guinea pigs).
- There is a statistically significant difference between the means of 1 and 2 mg dose, and 95% confidence interval of [-8.996, -3.734]does not include 0 and pvalue is less than 0.05, hence, we can reject the null hypothesis and say that there is a significant difference in tooth length between dosages of 1.0 mg and 2.0 mg (2.0 mg dosage creates longer teeth in guinea pigs).

Test by Supplement

Test by supplement only and do not consider dosages.

```
t.test(len ~ supp, paired = F, var.equal = F, data = ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.915, df = 55.31, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.171  7.571
## sample estimates:
## mean in group OJ mean in group VC
##            20.66            16.96
```

- There is a statistically significant difference between the means of OJ and VC but 95% confidence interval of [-0.171, 7.571] does contain 0 and pvalue is greater than 0.05, so we cannot reject the null hypothesis that there is not a significant difference in tooth length between the two supplement types.

Test by Supplement across Dosage Levels

Finally, for this test we will see if, given certain dosage levels, there is a significant difference in tooth growth between the two supplement types (i.e. at dose level 0.5 mg, is there a significant difference in tooth growth between VC and OJ supplement types?).

```
Tooth.dose05 <- subset(ToothGrowth, dose == 0.5)
Tooth.dose10 <- subset(ToothGrowth, dose == 1.0)
Tooth.dose20 <- subset(ToothGrowth, dose == 2.0)
```

```
t.test(len ~ supp, paired = F, var.equal = F, data = Tooth.dose05)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.17, df = 14.97, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719 8.781
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

```
t.test(len ~ supp, paired = F, var.equal = F, data = Tooth.dose10)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.033, df = 15.36, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802 9.058
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

```
t.test(len ~ supp, paired = F, var.equal = F, data = Tooth.dose20)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.0461, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.798  3.638
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

- At dose level 0.5 mg there is a statistically significant difference between the means of OJ and VC, and the 95% confidence interval of [1.72, 8,78] does not contain 0, and pvalue is less than 0.05, hence we can reject the null hypothesis and say that there is a significant difference in tooth length between the two supplement types at 0.5 mg dose level.
- At dose level 1 mg there is a statistically significant difference between the means of OJ and VC, and the 95% confidence interval of [2.80, 9.06] does not contain 0, and pvalue is less than 0.05, hence we can reject the null hypothesis and say that there is a significant difference in tooth length between the two supplement types at 1.0 mg dose level.
- At dose level 2 mg there is a statistically significant difference between the means of OJ and VC, and the 95% confidence interval of [-3.80, 3.64] does contain 0, and pvalue is substantially greater than 0.05, hence we cannot reject the null hypothesis that there is not a significant difference in tooth length between the two supplement types at 2.0 mg dose level.

State your conclusions and the assumptions needed for your conclusions.

Conclusions

1. When ignoring the supplement types, there is a significant difference in tooth length between the dose levels, with higher doses resulting in longer teeth.
2. When ignoring dosage levels, there is no significant difference in the tooth length between the supplement types.
3. When considering dosage levels and supplement types, OJ creates longer teeth than VC at dose levels of 0.5 mg and 1.0 mg, but at dose levels of 2.0 mg, there is no significant difference in teeth length between the two supplement types.

Assumptions

1. We assume that the variances between the separate populations tested are different (used var.equal = FALSE for all the t tests).
2. We assume that the populations are independent. They should be because in order to take these samples we would need at least 60 guinea pigs, and would not be able to ‘re-use’ one for a different test.(used paired=False for all t tests)
3. We assume that other statistical rules were followed, such as random populations of guineau pigs, the guinea pigs are more or less similar as a population, the researchers took accurate measurements, and the reasearchers that took the measurements were unaware of the dosages and supplement types that each individual guinea pig had been treated with.

Based on the analysis above, we can conclude that

- The 2mg dose has larger impact on tooth growth than 1mg and 0.5mg, while 1mg dose has more impact than 0.5mg dose. So there is a different in the growth of the tooth while the doses are larger.
- There is no doubt that orange juice and ascorbic acid have obvious different impact on tooth growth.