Statistical Inference Course Project_Part 2

Shee

In the second part, we're going to analyze the ToothGrowth data in the R datasets package.

1. Load the ToothGrowth data and perform some basic exploratory data analyses

```
#Load the ToothGrowth data.
data(ToothGrowth)
ToothGrowth$dose <- as.factor(ToothGrowth$dose)

#Some basic exploratory data analyses.
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: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

```
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
```

2. Basic summary of the data

The ToothGrowth dataset explains the relation between the growth of teeth of guinea pigs at each of three dose levels of Vitamin C (0.5, 1 and 2 mg) with each of two delivery methods(orange juice and ascorbic acid).

```
summary(ToothGrowth)
```

```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

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

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

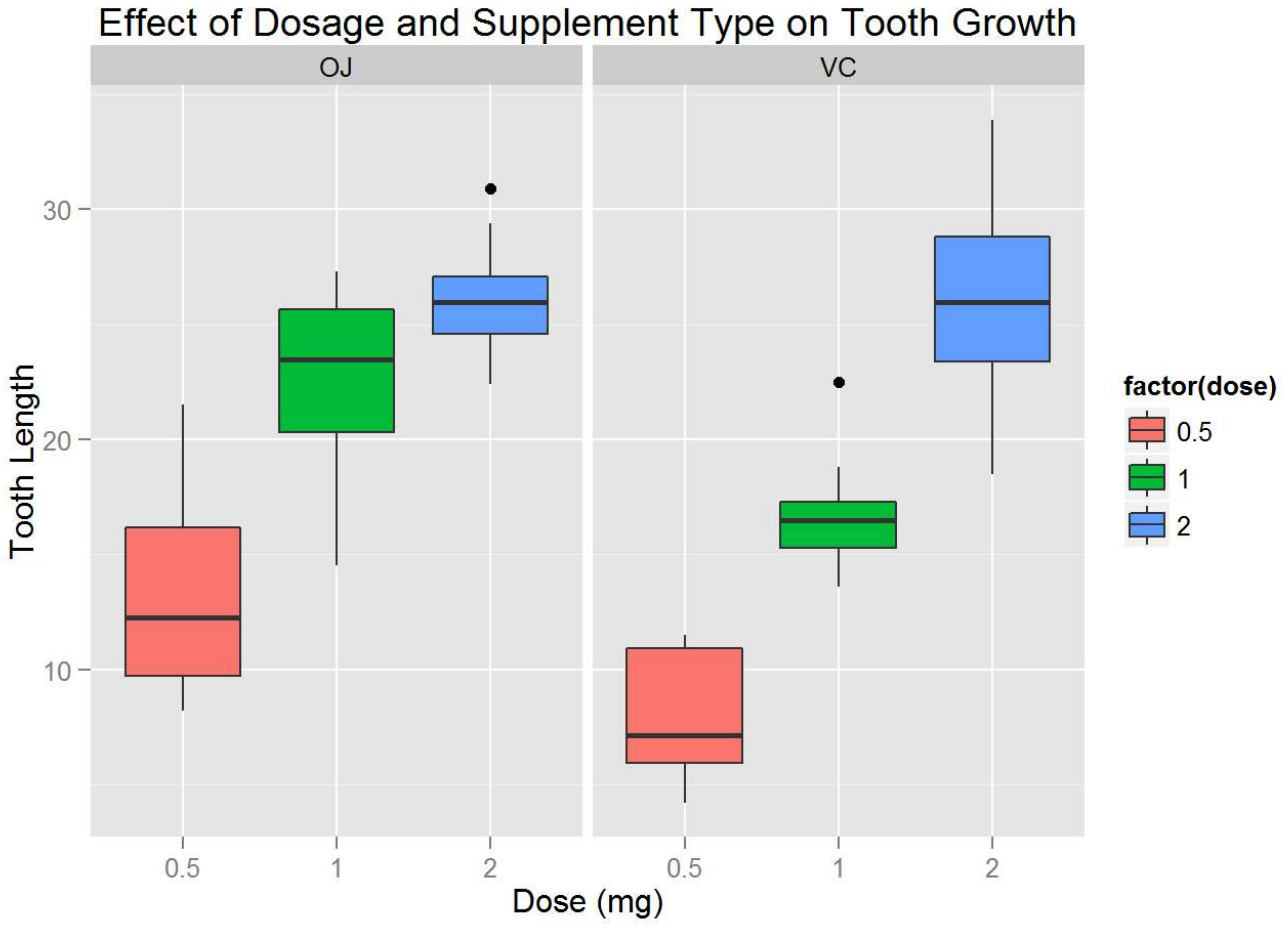
```
#Mean (by dose and by OJ & VC).
aggregate(ToothGrowth$len,list(ToothGrowth$supp,ToothGrowth$dose),mean)
```

```
##   Group.1 Group.2     x
## 1      OJ     0.5 13.23
## 2      VC     0.5  7.98
## 3      OJ       1 22.70
## 4      VC       1 16.77
## 5      OJ       2 26.06
## 6      VC       2 26.14
```

```
#Standard Deviation (by dose and by OJ & VC).
aggregate(ToothGrowth$len,list(ToothGrowth$supp,ToothGrowth$dose),sd)
```

```
##   Group.1 Group.2        x
## 1      OJ     0.5 4.459709
## 2      VC     0.5 2.746634
## 3      OJ       1 3.910953
## 4      VC       1 2.515309
## 5      OJ       2 2.655058
## 6      VC       2 4.797731
```

```
#Boxplot graph of the tooth length on the dose.
library(ggplot2)
ggplot(ToothGrowth, aes(x = factor(dose), y = len, fill = factor(dose))) +
       geom_boxplot() + facet_grid(.~supp) +
       labs(title = "Effect of Dosage and Supplement Type on Tooth Growth",
       x = "Dose (mg)", y = "Tooth Length")
```



From the boxplot, the length of the tooth increases while the dosage level increases.

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

```
# T-test by supplemant type
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
```

```
# T-test by dose level
Tooth_dose0510 <- subset(ToothGrowth, dose %in% c(0.5, 1.0))
Tooth_dose0520 <- subset(ToothGrowth, dose %in% c(0.5, 2.0))
Tooth_dose1020 <- subset(ToothGrowth, dose %in% c(1.0, 2.0))
t.test(len ~ dose, data = Tooth_dose0510)
```

```
##
##  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
```

```
t.test(len ~ dose, data = Tooth_dose0520)
```

```
##
##  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
```

```
t.test(len ~ dose, data = Tooth_dose1020)
```

```
##
##  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
```

```
# T-test for supplement by dose level
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, data = Tooth_dose05)
```

```
##
##  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 = Tooth_dose10)
```

```
##
##  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 = 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:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

For dose 0.5, the p-value of OJ compared to VC is 0.006359, which is less than 0.05, we reject null hypothesis, there is a difference between both methods.

For dose 1.0, the p-value of OJ compared to VC is 0.001038, which is less than 0.05, we reject null hypothesis, there is a difference between both methods.

For dose 2.0, the p-value of OJ compared to VC is 0.9639, which is greater than 0.05, we failed to reject null hypothesis, there is no significant difference between both methods.

4. Conclusions and Assumptions

Based of the results, we can conclude that:

1. The 2mg dose has larger impact on tooth growth than 1.0mg and 0.5mg, so there is a difference in the tooth growth while the doses are larger.
2. For doses of 0.5mg and 1.0mg, OJ has a greater effect on tooth growth than VC.
3. For dose 2.0mg there is no significant difference between the methods, thus it is uncertain whether there will be a greater effect for tooth growth from either OJ or VC.

Assumption: There is no other factor that will affect the tooth growth.