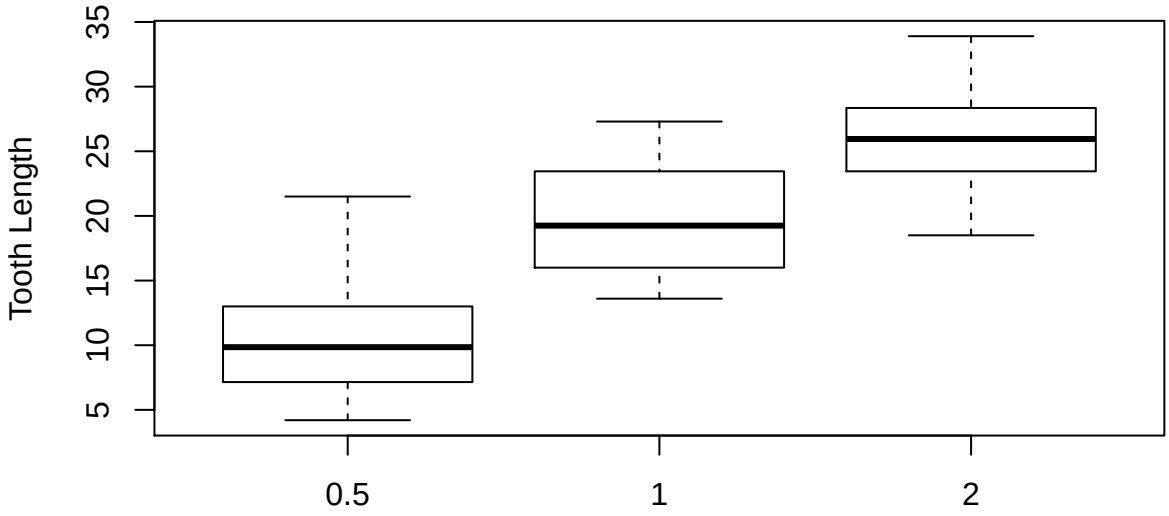
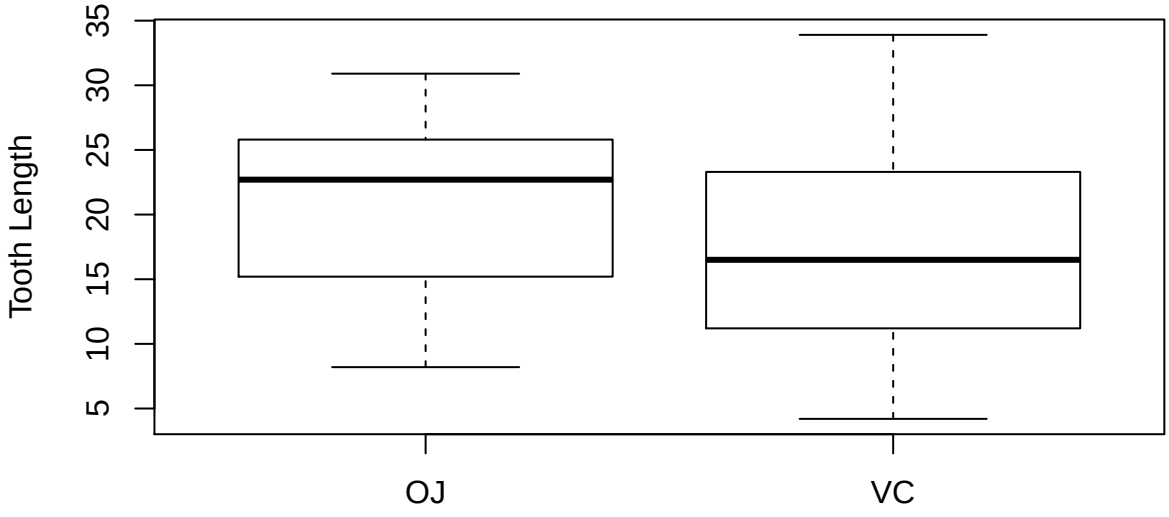
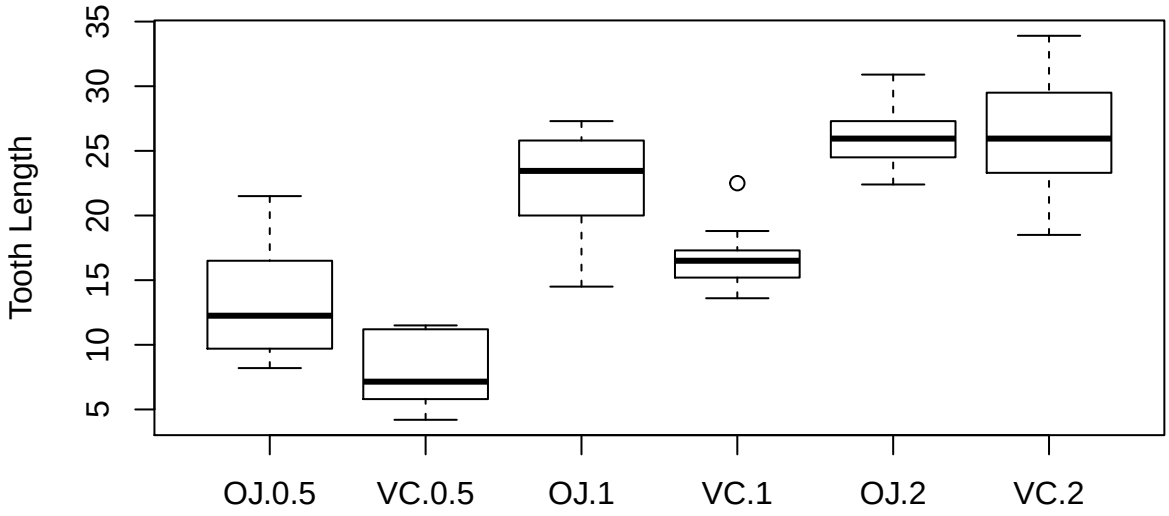
Course Project Report: Part 2

Section 1: Load the ToothGrowth data and perform some basic exploratory data analyses







As we can see on the first boxplot, on average the length of tooth increases as the dose increases. From the second boxplot, the length of tooth associated with OJ method seems to be greater than that associated with the VC method, but only marginally. The third boxplot shows possible interaction between supplement types and dose levels.

Section 2: Provide a basic summary of the data.

The summary of the data is given below:

```
##       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
```

We can group the sample by supplement types, and calculate the mean and standard deviation for each subsample.

```
sapply( split(ToothGrowth$len,ToothGrowth$supp),mean)
```

```
##       OJ       VC
## 20.66333 16.96333
```

```
sapply( split(ToothGrowth$len,ToothGrowth$supp),sd)
```

```
##       OJ       VC
## 6.605561 8.266029
```

Similarly, we can split the sample by dose levels, and calculate the mean and standard deviation for each subsample.

```
sapply( split(ToothGrowth$len,ToothGrowth$dose),mean)
```

```
##    0.5      1      2
## 10.605 19.735 26.100
```

```
sapply( split(ToothGrowth$len,ToothGrowth$dose),sd)
```

```
##      0.5        1        2
## 4.499763 4.415436 3.774150
```

We can also divide the sample by both supplement types and dose levels, and calculate the mean and standard deviation for each subsample.

```
sapply(split(ToothGrowth$len,list(ToothGrowth$supp,ToothGrowth$dose)),mean)
```

```
## OJ.0.5 VC.0.5   OJ.1   VC.1   OJ.2   VC.2
##  13.23   7.98  22.70  16.77  26.06  26.14
```

```
sapply(split(ToothGrowth$len,list(ToothGrowth$supp,ToothGrowth$dose)),sd)
```

```
##   OJ.0.5   VC.0.5     OJ.1     VC.1     OJ.2     VC.2
## 4.459709 2.746634 3.910953 2.515309 2.655058 4.797731
```

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

We first conduct t test for the split based on supplment types with two different assumptions on sample variance.

```
supp.t1<-t.test(len~supp, data=ToothGrowth)
supp.t2<-t.test(len~supp, data=ToothGrowth,var.equal=TRUE)
supp.result <- data.frame("p-value"=c(supp.t1$p.value, supp.t2$p.value),
                          "Conf-Low"=c(supp.t1$conf[1],supp.t2$conf[1]),
                          "Conf-High"=c(supp.t1$conf[2],supp.t2$conf[2]),
                          row.names=c("Equal Variance","Unequal Variance"))
supp.result
```

```
##                     p.value   Conf.Low Conf.High
## Equal Variance   0.06063451 -0.1710156  7.571016
## Unequal Variance 0.06039337 -0.1670064  7.567006
```

The p values in both cases are larger than 5%. Given the significant level of 5%, we would fail to reject the null hypothesis, which suggests that there is no difference in tooth growth between the two methods. Next, we conduct three t tests for the split based on dose levels. As the procedure is quite similar to the previous one, we drop the R code from the main text. The result is as follows:

```
##                p.value   Conf.Low  Conf.High
## 0.5 vs 1 1.266297e-07 -11.983748  -6.276252
## 0.5 vs 2 2.837553e-14 -18.153519 -12.836481
## 1 vs 2   1.810829e-05  -8.994387  -3.735613
```

The p values for all three comparisons are much smaller than 5%. Given the significant level of 5%, we reject the null, which suggests that there are indeed differences in tooth growth among these three subgroups. Lastly, we compare tooth growth by both supplement types and dose levels.

```
dose_5 <- subset(ToothGrowth, dose == 0.5)
dose_1 <- subset(ToothGrowth, dose == 1)
dose_2 <- subset(ToothGrowth, dose == 2)
dose.supp.t1<-t.test(len ~ supp, data = dose_5)
dose.supp.t2<-t.test(len ~ supp, data = dose_1)
dose.supp.t3<-t.test(len ~ supp, data = dose_2)
dose.supp.result <- data.frame("p-value"=
             c(dose.supp.t1$p.value, dose.supp.t2$p.value,dose.supp.t3$p.value),
            "Conf-Low"=c(dose.supp.t1$conf[1],dose.supp.t2$conf[1],dose.supp.t3$conf[1]),
            "Conf-High"=c(dose.supp.t1$conf[2],dose.supp.t2$conf[2],dose.supp.t3$conf[2]),
            row.names=c("0.5","1","2"))
dose.supp.result
```

```
##         p.value  Conf.Low Conf.High
## 0.5 0.006358607  1.719057  8.780943
## 1   0.001038376  2.802148  9.057852
## 2   0.963851589 -3.798070  3.638070
```

For the dose levels of 0.5 and 1, there is a difference in the length of tooth between the OJ method and the VC method. For the dose level of 2, however, we fail to reject the null hopothesis, which suggests that there is no difference in tooth growth between the two methods.

Section 4: State your conclusions and the assumptions needed for your conclusions.

Collecting findings in section 3, we can conclude that the deliveray methods would affect tooth growth when dose stays at relatively low levels, while they would have no difference when the dose is at high levels. Since we rely heavily on t statistics to do hypothesis testing, all our observations of tooth growth are assumed to be independently drawn from some normal distribution.