Statistical Inference Course Project Part 2

Question 1

> Load the ToothGrowth data and perform some basic exploratory data analyses

```
data(ToothGrowth)
nrow(ToothGrowth)
```

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

```
ncol(ToothGrowth)
```

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

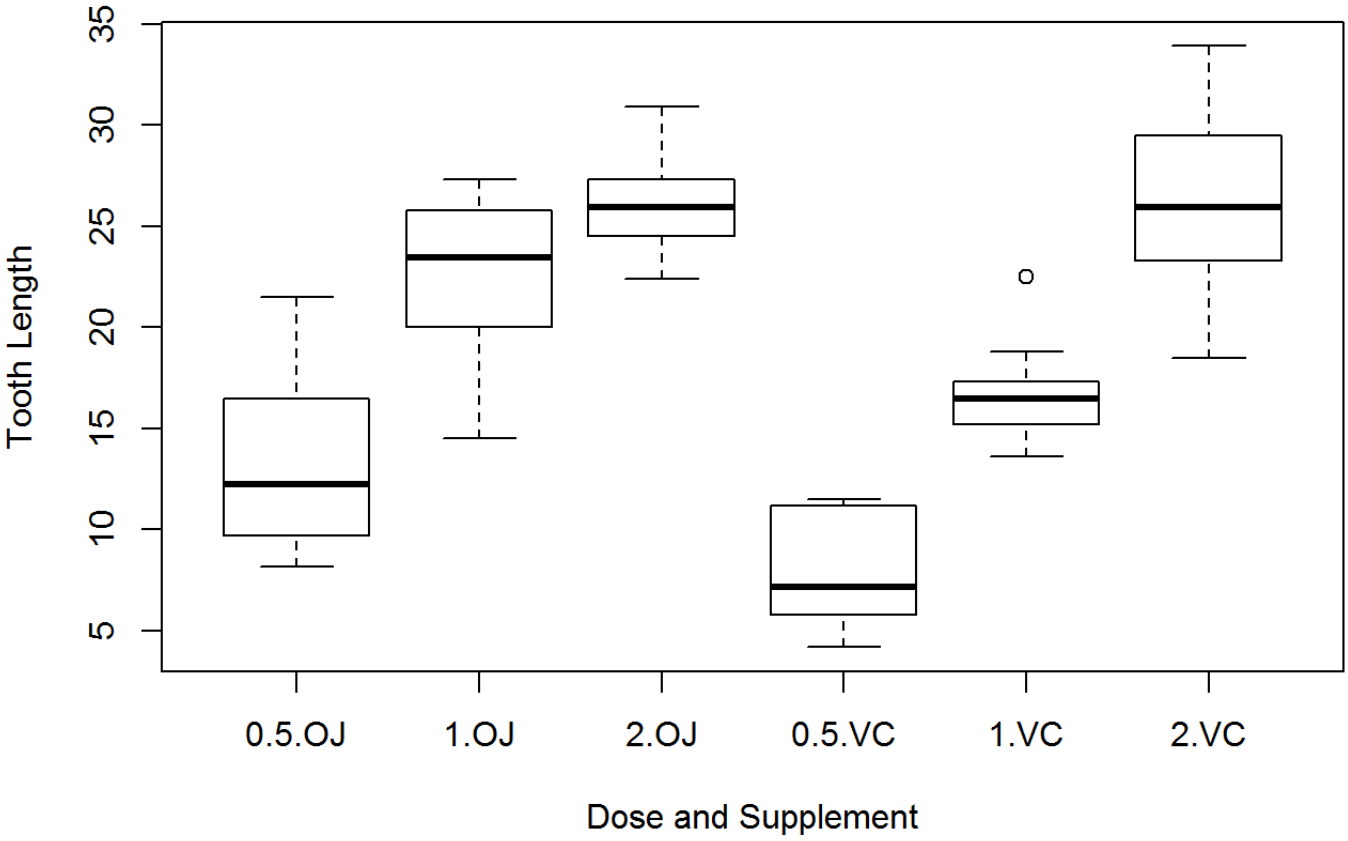
```
str(ToothGrowth)
```

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

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

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

```
toothgrowth <- transform(ToothGrowth, supp = factor(supp))
boxplot(len ~ dose * supp, toothgrowth, xlab = "Dose and Supplement", ylab = "Tooth Length")
```



Question 2

> Provide a basic summary of the data

```
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
```

The data is the length of teeth in each of 10 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.

Based on plotting the data as boxplots some correlations were observed. The length of the tooth increases as the dosage increases. The orange juice delivery method yields a greater length than the ascorbic acid for smaller dosages but the difference is very small.

Question 3

> Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose.

The tooth growth was compared by supplement for each dosage under the null hypothesis that each supplement has the same effect at a certain dosage on the tooth.

```
# We split the data up by dosages
d0.5 <- subset(ToothGrowth, dose == 0.5)
d1.0 <- subset(ToothGrowth, dose == 1.0)
d2.0 <- subset(ToothGrowth, dose == 2.0)

# We do a t-test between the supplements
test0.5 <- t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = d0.5)
test0.5$p.value; test0.5$conf[1]
```

```
## [1] 0.006358607
```

```
## [1] 1.719057
```

```
test1.0 <- t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = d1.0)
test1.0$p.value; test1.0$conf[1]
```

```
## [1] 0.001038376
```

```
## [1] 2.802148
```

Dosages 1.0 and 1.5 have significant p-values of 0.006359 and 0.001038 respectively indicating that the difference in the mean values between the supplements is significant. Dosage 1.0 has a confidence interval of 1.719-8.781 and dosage 2.0 has a confidence interval of 2.802-9.058.

```
test2.0 <- t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = d2.0)
test2.0$p.value; test2.0$conf[1]
```

```
## [1] 0.9638516
```

```
## [1] -3.79807
```

Dosage 3.0 has a very high p-value of 0.9639 and a confidence interval below zero -3.798-3.638. This indicates that there is no significant difference between the supplements at this dosage.

Question 4

> State your conclusions and the assumptions needed for your conclusions.

The supplements orange juice and ascorbic acid have different effects on tooth length for lower dosages of vitamin C according to the t-test. We can conclude that orange juice is more effective than ascorbic acid at the 0.5 and 1.0 dose levels. However, there is no difference between these two at the 3.0 dose level.

These conclusions are based on the assumption that the underlying data is normally distributed and that each data observation was an independent trial.