Statistical_Inference Course Project Part 2: ToothGrowth Dataset Analysis

Author: Michael Nichols

Basic Summary of the Dataset:

ToothGrowth - data frame with 60 observations on 3 variables - len numeric Tooth length - supp factor Supplement type (VC or OJ). —VC: ascorbic acid tablet —OJ: orange juice - dose numeric Dose in milligrams/day

```
#load libraries
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.3
```

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

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 ...
```

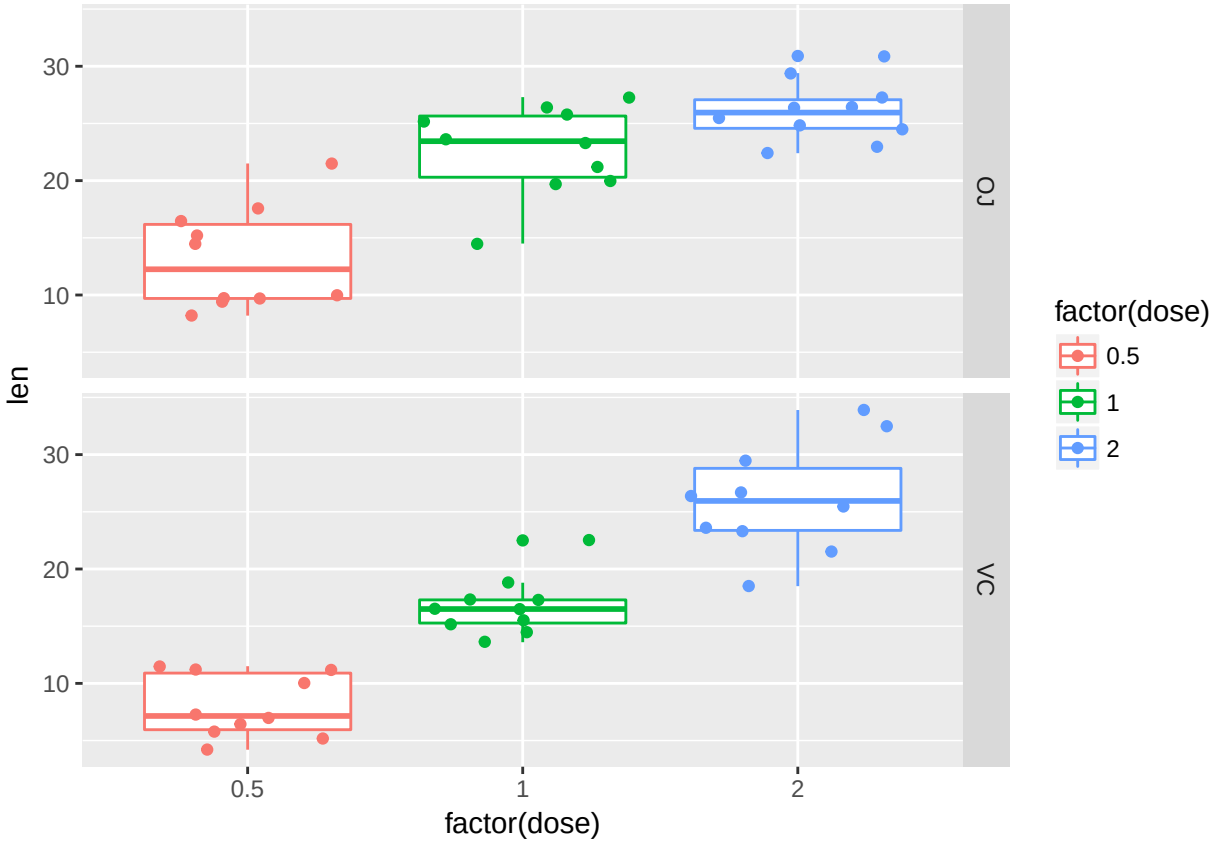
```
head(ToothGrowth)
```

```
##    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
```

```
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
```

```
#basic plot of the summary statistics
qplot(factor(dose), len, data = ToothGrowth, color = factor(dose), facets = supp ~ ., geom = c("boxplot
```



2. T-Test Evaluation

```
# Explore effects of supplement type on tooth growth by performing a two-sample t-test for the differen

ToothTTest <- t.test(len ~ supp, data = ToothGrowth, var.equal = FALSE, paired = FALSE)
```

3. Conclusions:

```
#show table of two sample t-test results for tooth growth by supplement
tTestResults <- matrix(data.frame(ToothTTest$p.value, ToothTTest$conf.int[1], ToothTTest$conf.int[2], T
rownames(tTestResults) <- c("p-value", "Lower Conf Int", "Upper Conf Int", "OJ Mean", "VC Mean")
colnames(tTestResults) <- "Results"

print(tTestResults)
```

```
##                Results
## p-value        0.06063451
## Lower Conf Int -0.1710156
## Upper Conf Int 7.571016
## OJ Mean        20.66333
## VC Mean        16.96333
```

We do not reject the Null hypothesis, as the p.value is 0.061 which is greater than the 0.05 threshold and the confidence interval contains 0.