Project: Statistical Inference - Part II

Part 2: Basic Inferential Data Analysis Instructions

Now in the second portion of the project, 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 2. Provide a basic summary of the data. 3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering) 4. State your conclusions and the assumptions needed for your conclusions.

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

```
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 ...
```

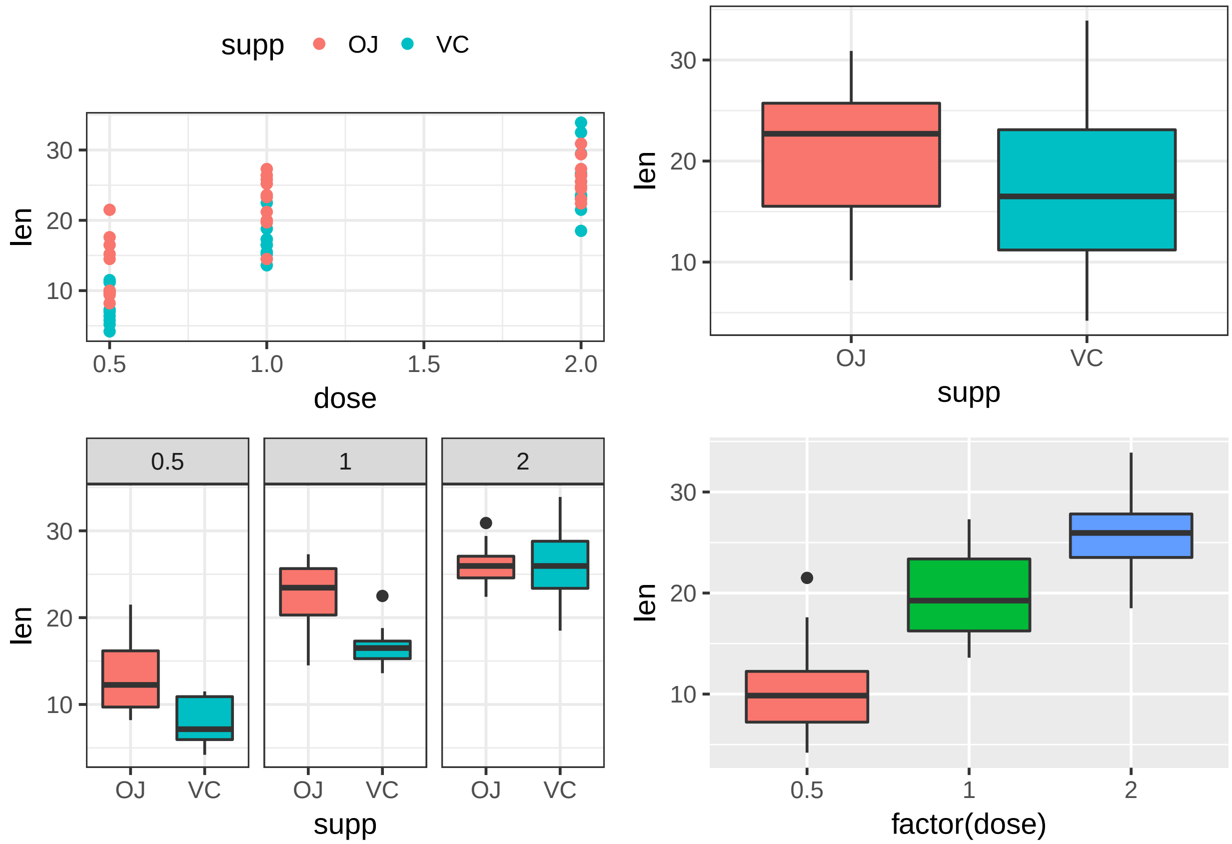
```
# To find all unique values for each causal factor
unique(ToothGrowth$supp)
```

```
## [1] VC OJ
## Levels: OJ VC
```

```
unique(ToothGrowth$dose)
```

```
## [1] 0.5 1.0 2.0
```

```
# Comparision between Tooth length and Vitamin C dose
g1 <- ggplot(aes(x=dose, y = len), data = ToothGrowth) + geom_point(aes(color = supp)) +
    theme_bw() + theme(legend.position = "top")
# Comparison between delivery methods and tooth length
g2 <- ggplot(aes(x = supp, y = len), data = ToothGrowth) +
    geom_boxplot(aes(fill = supp))+
    theme_bw() + theme(legend.position = "none")
# Comparison between delivery methods at each dose level
g3 <- ggplot(aes(x = supp, y = len), data = ToothGrowth) +
    geom_boxplot(aes(fill = supp)) + facet_wrap(~ dose)+
    theme_bw() + theme(legend.position = "none")
# Comparison between Vitamin C dose levels and tooth length
g4 <- ggplot(aes(x = factor(dose), y = len), data = ToothGrowth) +
    geom_boxplot(aes(fill = factor(dose))) + theme(legend.position = "none")
# You might have produced myPlotList using instead lapply, mc.lapply, plyr::dlply, etc
myPlotList = list(g1, g2, g3, g4)
library("gridExtra")
do.call(grid.arrange,  myPlotList)
```



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
```

```
by(ToothGrowth$len, INDICES = list(ToothGrowth$supp, ToothGrowth$dose), summary)
```

```
## : OJ
## : 0.5
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##    8.20    9.70   12.25   13.23   16.18   21.50 
## -------------------------------------------------------- 
## : VC
## : 0.5
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##    4.20    5.95    7.15    7.98   10.90   11.50 
## -------------------------------------------------------- 
## : OJ
```

```
## : 1
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   14.50   20.30   23.45   22.70   25.65   27.30
## --------------------------------------------------------
## : VC
## : 1
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   13.60   15.28   16.50   16.77   17.30   22.50
## --------------------------------------------------------
## : OJ
## : 2
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   22.40   24.57   25.95   26.06   27.07   30.90
## --------------------------------------------------------
## : VC
## : 2
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.50   23.38   25.95   26.14   28.80   33.90
```

```r
length(ToothGrowth$len)
```

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

```r
by(ToothGrowth$len, INDICES = list(ToothGrowth$supp, ToothGrowth$dose), length)
```

```
## : OJ
## : 0.5
## [1] 10
## --------------------------------------------------------
## : VC
## : 0.5
## [1] 10
## --------------------------------------------------------
## : OJ
## : 1
## [1] 10
## --------------------------------------------------------
## : VC
## : 1
## [1] 10
## --------------------------------------------------------
## : OJ
## : 2
## [1] 10
## --------------------------------------------------------
## : VC
## : 2
## [1] 10
```

3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)

```r
t.test(len ~ supp, paired = F, var.equal = F, 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
```

A confidence interval of [-0.171, 7.571] does not allow us to reject the null hypothesis (that there is no correlation between delivery method and tooth length).

```r
# Analysis of the data for correlation between the dose level and change in tooth growth
dose1 <- subset(ToothGrowth, dose %in% c(0.5, 1.0))
dose2 <- subset(ToothGrowth, dose %in% c(0.5, 2.0))
dose3 <- subset(ToothGrowth, dose %in% c(1.0, 2.0))
t.test(len ~ dose, paired = F, var.equal = F, data = dose1)
```

```
##
##  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
```

```r
t.test(len ~ dose, paired = F, var.equal = F, data = dose2)
```

```
##
##  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
```

```r
t.test(len ~ dose, paired = F, var.equal = F, data = dose3)
```

```
##
##  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
```

Confidence interval is -11.98 for Doses 0.5 Confidence interval is -6.276 for Doses 1.0 Confidence interval is -18.16 for Doses 0.5 Confidence interval is -12.83 for Doses 2.0 Confidence interval is -8.996 for Doses 1.0 Confidence interval is -3.734 for Doses 2.0

Above data indicates the rejection of the null hypothesis and a confirm that there is a significant correlation between tooth length and dose levels.

```r
# Analysis of the correlation between dose level and impact in tooth growth
Tooth.dose5 <- subset(ToothGrowth, dose == 0.5)
Tooth.dose1 <- subset(ToothGrowth, dose == 1.0)
Tooth.dose2 <- subset(ToothGrowth, dose == 2.0)
t.test(len ~ supp, paired = F, var.equal = F, data = Tooth.dose5)
```

```
##
##  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
```

```r
t.test(len ~ supp, paired = F, var.equal = F, data = Tooth.dose1)
```

```
##
##  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
```

```r
t.test(len ~ supp, paired = F, var.equal = F, data = Tooth.dose2)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, 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
```

Summary

The confidence intervals for dose levels 0.5mg and 1.0mg([1.72, 8,78] within 0.5mg, [2.80, 9.06] within 1.0mg) allow for the rejection of the null hypothesis and a confirmation that there is a significant correlation between tooth length and dose levels. However, the confidence interval for dose level 2.0[-3.80, 3.64] is not enough to reject the null hypothesis.

Conclusions and the Assumptions

With the values obtained it can be assumed that there is a different in the growth of the tooth while the doses are larger. By looking at the boxplot and the results, it can also be said that there is no other factor that will affect the growing process (it will depend greatly on the dose), in other words, the delivery methods are independent of the dose size.