Part 2: Basic Inferential Data Analysis

Muhammed Afifi Ibrahim

November 5, 2016

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.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

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

Load dataset ToothGrowth

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

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: 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
```

```
          tail(ToothGrowth)
```

```
##      len supp dose
## 55 24.8   OJ    2
## 56 30.9   OJ    2
## 57 26.4   OJ    2
## 58 27.3   OJ    2
## 59 29.4   OJ    2
## 60 23.0   OJ    2
```

Number of Rows and Columns.

```
          dim(ToothGrowth)
```

```
## [1] 60  3
```

Sample Size n

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

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

Mean group 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     1.0 22.70
## 4      VC     1.0 16.77
## 5      OJ     2.0 26.06
## 6      VC     2.0 26.14
```

Standard Deviation group 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     1.0 3.910953
## 4      VC     1.0 2.515309
## 5      OJ     2.0 2.655058
## 6      VC     2.0 4.797731
```

Boxplot graph of the tooth length vs the does

```
ggplot(ToothGrowth, aes(x = factor(dose), y = len, fill = factor(dose)))+
        geom_boxplot()+
        facet_grid(.~supp)+
        labs(title = "Tooth Length vs. Dose  by for OJ & VC",
        x = "Doses", y = "Tooth Length")
```
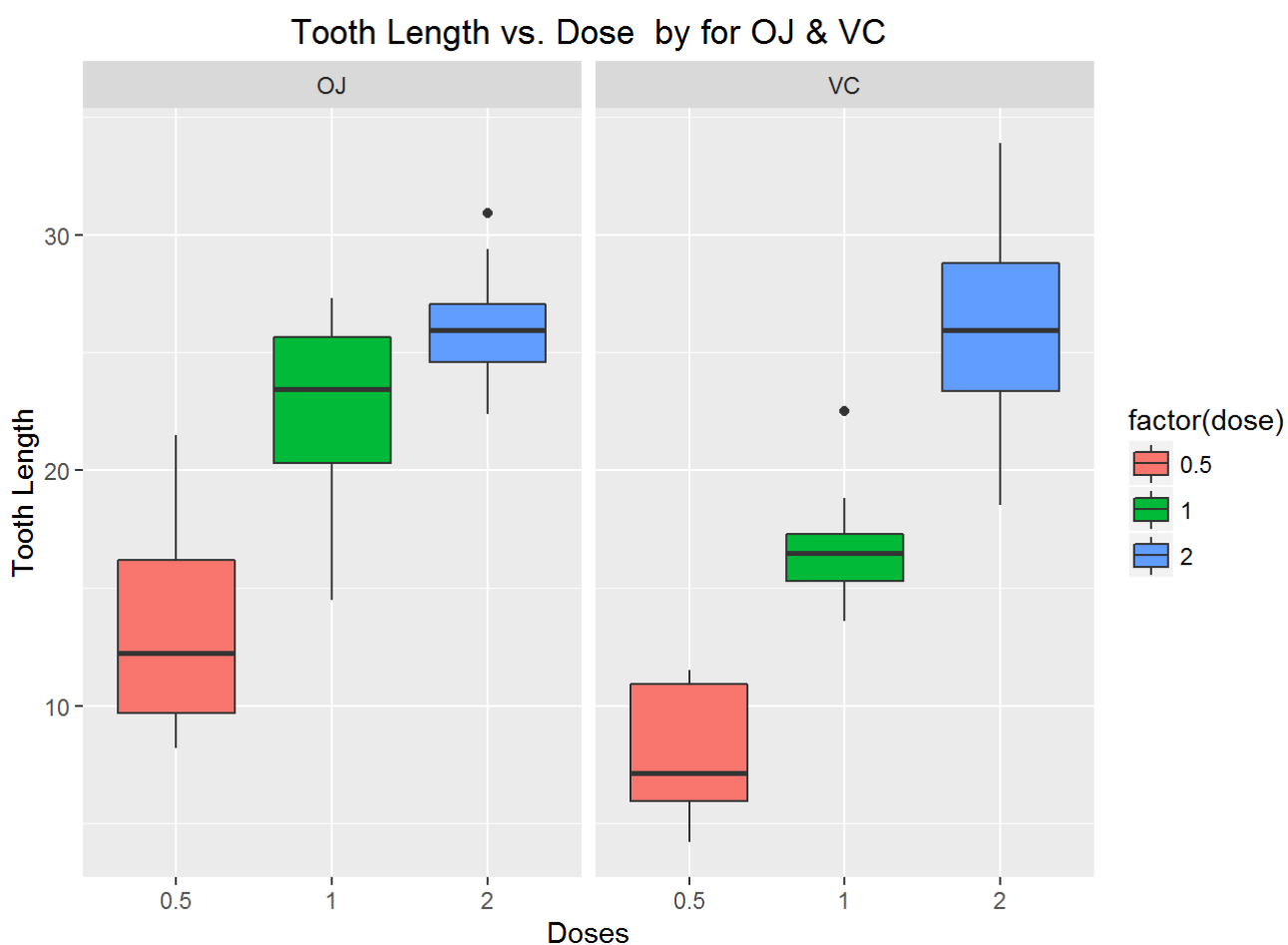

Question 2: Provide a basic summary of the data.

```
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
```

Table view of data.

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

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

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

From the box plot, OJ looks to be doing better with dose 0.5 and 1 effect on teeth growth than VC. To test this hypothesis by hold a the mean of OJ and VC does not cross zero.

Dose 0.5

We are 95% confident that the limits of 1.719057 and 8.780943 actually do contain the difference between the two population means.Because those limts do not contain zero, this confodence interval suggests that it is very possible that the two population means are not equal.

```
ojdose0.5 <- ToothGrowth %>% filter(supp=="OJ" & dose=="0.5")
vcdose0.5 <- ToothGrowth %>% filter(supp=="VC" & dose=="0.5")
t.test(ojdose0.5$len,vcdose0.5$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  ojdose0.5$len and vcdose0.5$len
## 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 of x mean of y
##     13.23      7.98
```

Dose 1

We are 95% confident that the limits of 2.802148 and 9.057852 actually do contain the difference between the two population means. Because those limts do not contain zero, this confodence interval suggests that it is very possible that the two population means are not equal.

```
ojdose1 <- ToothGrowth %>% filter(supp=="OJ" & dose=="1")
vcdose1 <- ToothGrowth %>% filter(supp=="VC" & dose=="1")
t.test(ojdose1$len,vcdose1$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  ojdose1$len and vcdose1$len
## 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 of x mean of y
##     22.70     16.77
```

Dose 2

We are 95% confident that the limits of -3.79807 and 3.63807 actually do contain the difference between the two population means. However, because those limts do contain zero, this confodence interval suggests that it is very possible that the two population means are equal.

```
ojdose2 <- ToothGrowth %>% filter(supp=="OJ" & dose=="2")
vcdose2 <- ToothGrowth %>% filter(supp=="VC" & dose=="2")
t.test(ojdose2$len,vcdose2$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  ojdose2$len and vcdose2$len
## 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 of x mean of y
##     26.06     26.14
```

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

We are 95% confident that dose 0.5 and dose 1 of OJ result in longer tooth length than dose 0.5 and dose 1 of VC. However, at the highest dose of 2, there is no statistically significant difference between the effects of OJ and VC.

Appendix

```
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
```

```
dim(ToothGrowth)
```
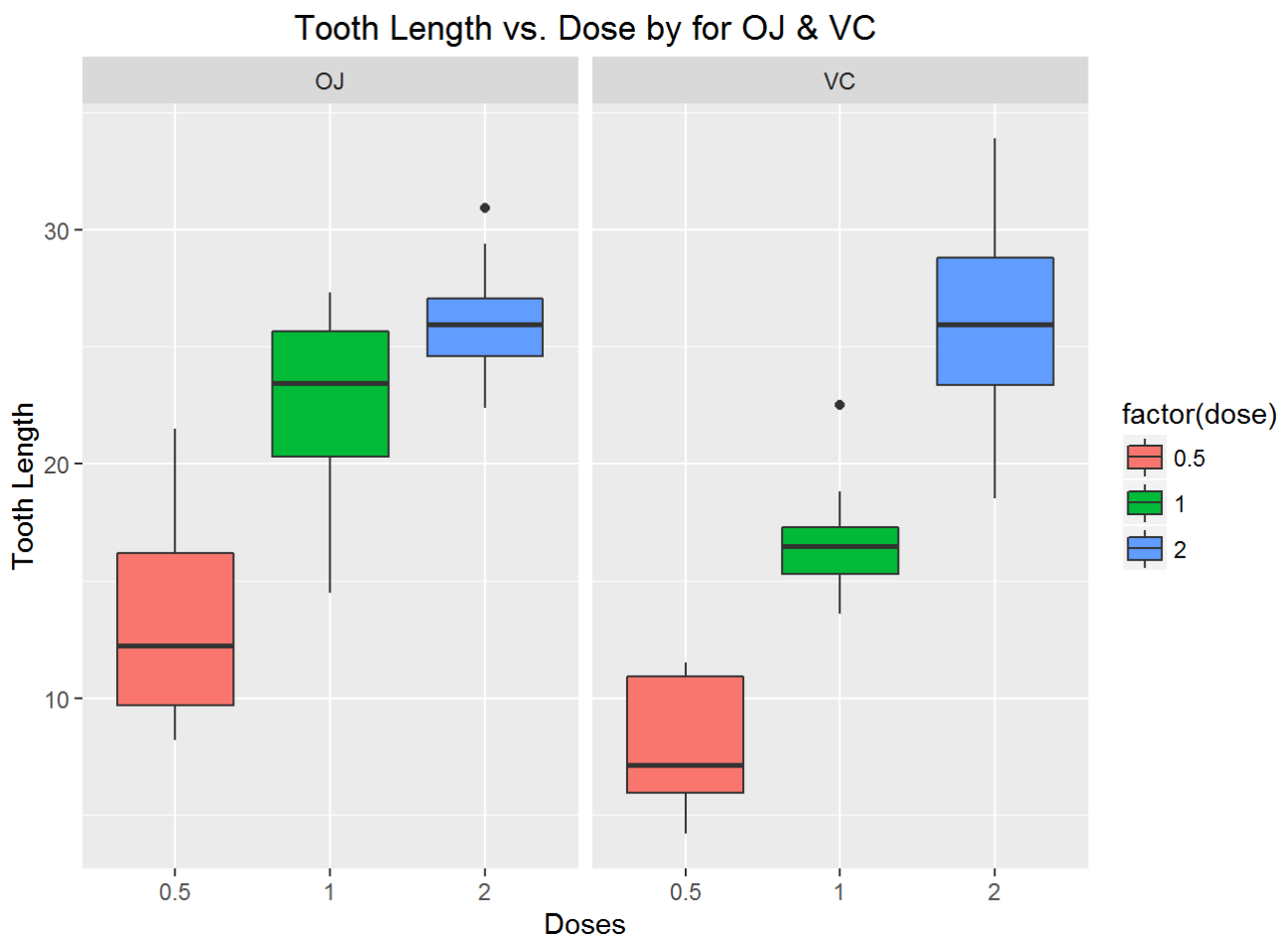
```
## [1] 60  3
```

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

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

Figure 1 Code

```
ggplot(ToothGrowth, aes(x = factor(dose), y = len, fill = factor(dose)))+
geom_boxplot()+
facet_grid(.~supp)+
labs(title = "Tooth Length vs. Dose by for OJ & VC",
x = "Doses", y = "Tooth Length")
```


Dose 0.5

```
ojdose0.5 <- ToothGrowth %>% filter(supp=="OJ" & dose=="0.5")
vcdose0.5 <- ToothGrowth %>% filter(supp=="VC" & dose=="0.5")
t.test(ojdose0.5$len,vcdose0.5$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  ojdose0.5$len and vcdose0.5$len
## 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 of x mean of y
##     13.23      7.98
```

Dose 1

```
ojdose1 <- ToothGrowth %>% filter(supp=="OJ" & dose=="1")
vcdose1 <- ToothGrowth %>% filter(supp=="VC" & dose=="1")
t.test(ojdose1$len,vcdose1$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  ojdose1$len and vcdose1$len
## 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 of x mean of y
##     22.70     16.77
```

Dose 2

```
ojdose2 <- ToothGrowth %>% filter(supp=="OJ" & dose=="2")
vcdose2 <- ToothGrowth %>% filter(supp=="VC" & dose=="2")
t.test(ojdose2$len,vcdose2$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  ojdose2$len and vcdose2$len
## 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 of x mean of y
##     26.06     26.14
```