Statistical Inference Part 2

Kris Jacomet

8/14/2020

Load Libraries

```
        library(ggplot2)
        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(datasets)
```

1. Loading Data & Exploratory Analyses

```
        data(ToothGrowth)
        dim(ToothGrowth)
```

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

```
        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
```

```
        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
```

2. Summary 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
```

3. Graphs & Hypothesis Testing

Boxplot Tooth Length Vs. Dose

```
           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")
```

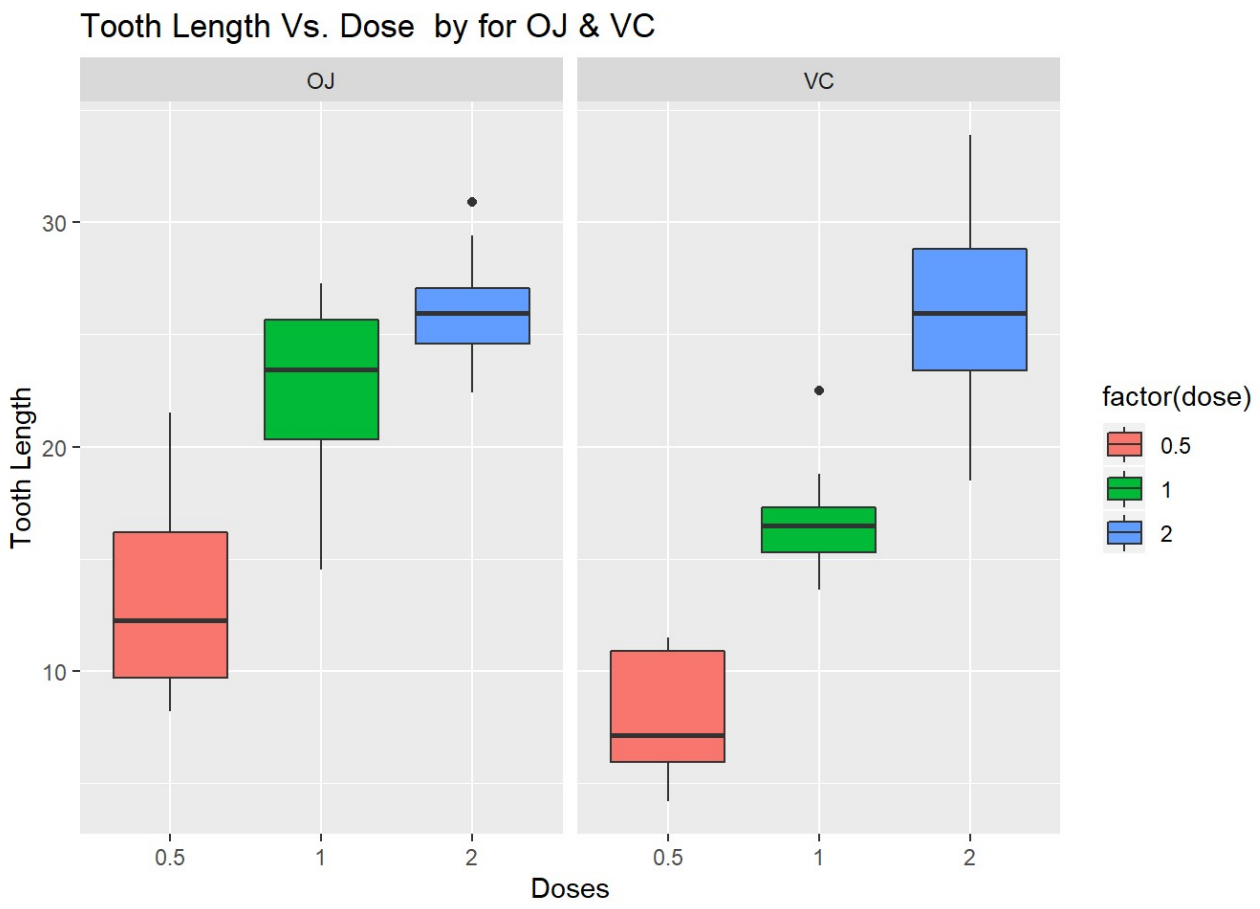

Table View

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

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

Confidence Intervals / Hypothesis Tests - Comparison of Tooth Growth by Supp & Dose

In accordance with the box plot, OJ appears better with a dose 0.5 and 1.0 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 the limits of 1.719057 and 8.780943 contain the difference between the two population means. The confidence interval suggests that it is possible that the two population means are not equal since the limits do not contain zero.

```
          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 1.0

We are 95% confident the limits of 2.802148 and 9.057852 contain the difference between the two population means. The confidence interval suggests that it is possible that the two population means are not equal since the limits do not contain zero.

```
            ojdose1.0 <- ToothGrowth %>% filter(supp=="OJ" & dose=="1")
            vcdose1.0 <- ToothGrowth %>% filter(supp=="VC" & dose=="1")
            t.test(ojdose1.0$len,vcdose1.0$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  ojdose1.0$len and vcdose1.0$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 2.0

We are 95% confident the limits of -3.79807 and 3.63807 contain the difference between the two population means. The confodence interval suggests that it is possible that the two population means are equal since the limits contain zero.

```
            ojdose2.0 <- ToothGrowth %>% filter(supp=="OJ" & dose=="2")
            vcdose2.0 <- ToothGrowth %>% filter(supp=="VC" & dose=="2")
            t.test(ojdose2.0$len,vcdose2.0$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  ojdose2.0$len and vcdose2.0$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
```

4. Conclusions & Assumptions

We are 95% confident the OJ doses of 0.5 and 1 result in longer tooth length than VC doses of 0.5 and 1. However, at the highest dose of 2.0, there is no statistically significant difference between the effects of OJ and VC. The null hypotheses that there is a relationship between the supplements and the length of the tooth has a p value of less then 5% (3%). The 5% interval is a rule of thumb used. This means we have to reject the null hypotheses and assume that there is a correlation.