Stastical Inference Part 2: Tooth Growth Analysis

Rhay

3/14/2021

Overview of Requirements

Analyze the ToothGrowth data in the R datasets package. The data is set of 60 observations, length of odontoblasts (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 or ascorbic acid).

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.

Data Processing

Load data

```
data("ToothGrowth")

#look at data structure
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 ...
```

Data Investigation

```
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
```

```r
head(ToothGrowth,5)
```

```
##    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
```

```r
#display summary statistics from 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
```

```r
#display first 5 rows of data
head(ToothGrowth,5)
```

```
##    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
```

Find unique values for each column

```r
#display summary statistics from data
unique(ToothGrowth$len)
```

```
##  [1]  4.2 11.5  7.3  5.8  6.4 10.0 11.2  5.2  7.0 16.5 15.2 17.3 22.5 13.6 14.5
## [16] 18.8 15.5 23.6 18.5 33.9 25.5 26.4 32.5 26.7 21.5 23.3 29.5 17.6  9.7  8.2
## [31]  9.4 19.7 20.0 25.2 25.8 21.2 27.3 22.4 24.5 24.8 30.9 29.4 23.0
```

```r
#Find the range to values for tooth length
cat("The range to tooth values is:", range(ToothGrowth$len))
```

```
## The range to tooth values is: 4.2 33.9
```

```r
#Show unique supplements
unique(ToothGrowth$supp)
```

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

```
#Find the unique values for dose
cat("The unique dose values is:", unique(ToothGrowth$dose))
```

```
## The unique dose values is: 0.5 1 2
```

```
#Find the range to values for dose
cat("The range to dose values is:", range(ToothGrowth$dose))
```

```
## The range to dose values is: 0.5 2
```

Calculate the mean for each supplement type

```
MeanSupp = split(ToothGrowth$len, ToothGrowth$supp)
sapply(MeanSupp, mean)
```

```
##       OJ       VC
## 20.66333 16.96333
```

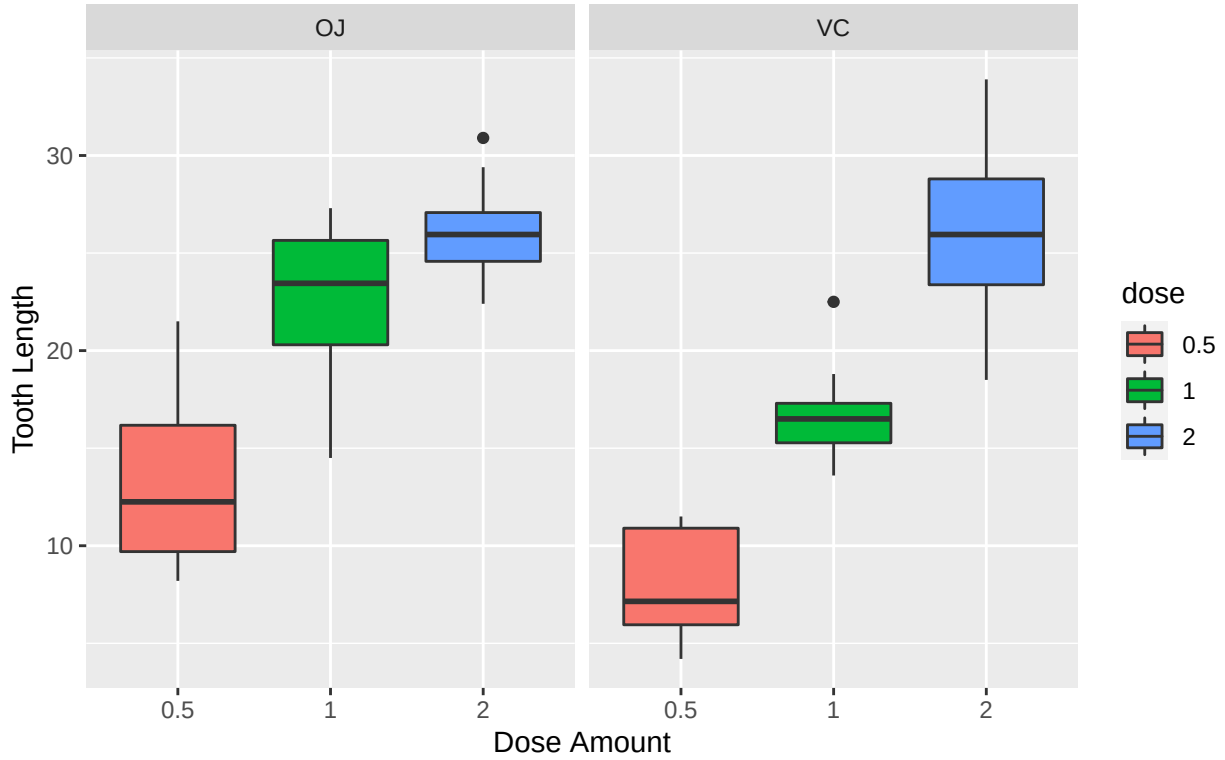
Plot Data

Create a box plot to show the tooth length by dose amount by delivery method

```
# Convert dose to a factor
ToothGrowth$dose<-as.factor(ToothGrowth$dose)

# Plot tooth length ('len') vs. the dose amount ('dose'), broken out by supplement delivery method ('su
ggplot(aes(x=dose, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=dose)) + xlab("Dose Amount") + ylab
     theme(plot.title = element_text(lineheight=.8, face="bold"))
```

Figure 1
Tooth Length vs. Dose Amount by Delivery Method

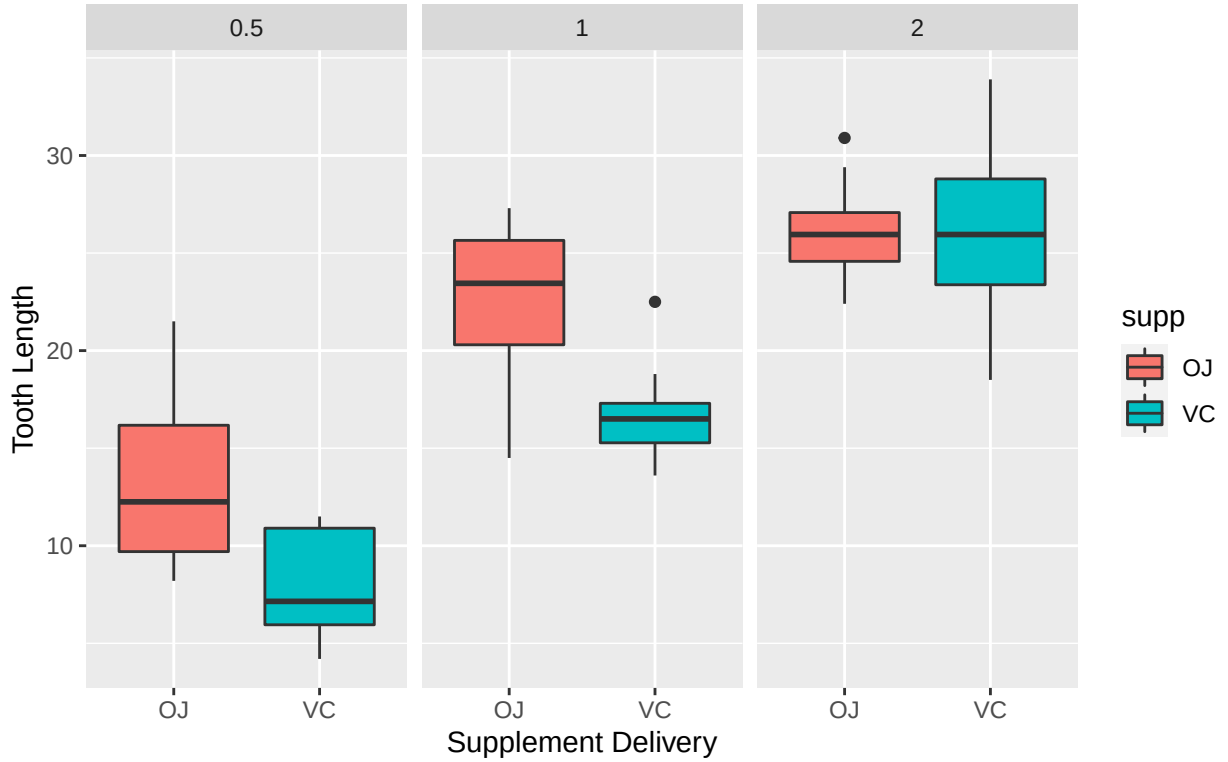


The blue and green boxes indicate that higher doses of both supplements resulted in longer tooth length (blue and green boxes). The green boxes show that for a dose=1, the mean and variance is higher for the OJ supplement. For the 0.5 dose, the mean for the OJ supplement is higher than for the OJ supplement.

Let's plot the tooth length vs. Delivery Method by Dose Amount

```
ggplot(aes(x=supp, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=supp)) + xlab("Supplement Delivery
     theme(plot.title = element_text(lineheight=.8, face="bold"))
```

Figure 2:
Tooth Length vs. Delivery Method by Dose Amount



For doses of 0.5 and 1 it appears that the OJ supplement results in larger tooth length although more so for the does equal to 1. For the dose equal to 2 it appears that the supplements result in similar tooth length. However, there is a higher variance for the VC supplement.

Inferential Statistics

Does the tooth length depend on the delivery method? Do a t-test.

```
len<-ToothGrowth$len
supp<-ToothGrowth$supp
dose<-ToothGrowth$dose

sapply(MeanSupp, var)
```

```
##       OJ       VC
## 43.63344 68.32723
```

```
t.test(len[supp=="OJ"], len[supp=="VC"], paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len[supp == "OJ"] and len[supp == "VC"]
## 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 of x mean of y
##  20.66333  16.96333
```

The pvalue is 0.06 which is higher than 0.05. From this we would conclude that supplements have no impact on tooth length.

Let's look at tooth length by dose. Have to compare:

Compare doses 0.5 and 1.0

```
ToothGrowth_sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(1.0,0.5))
t.test(len~dose,data=ToothGrowth_sub)
```

```
##
##  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
```

Compare doses 0.5 and 2.0

```
ToothGrowth_sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(0.5,2.0))
t.test(len~dose,data=ToothGrowth_sub)
```

```
##
##  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
```

Compare doses 1 and 2

```
ToothGrowth_sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(1.0,2.0))
t.test(len~dose,data=ToothGrowth_sub)
```

```
##
```

```
##  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
```

For each of these comparisons, the pvalues are close to zero. The confidence intervals is not greater than zero. Therefore we conclude that the tooth length increases with increasing does and the null hypothesis can be rejected.

Conclusions

From the t tests analysis we can conclude that the supplement had no effect on the tooth length but that increase doses of either supplement resulted in larger tooth length.