Statistical Inference Project - Basic Inferential Data Analysis

Lena Horsley

Overview

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.

Setup

```
library(ggplot2)
data("ToothGrowth")

#Make a copy of the data frame and rename some columns
myToothGrowthData <- data.frame(ToothGrowth)
colnames(myToothGrowthData)[2] <- c("supplement")
colnames(myToothGrowthData)[1] <- c("length")

#Replace OJ and VC with Orange Juice and Vitamin C
myToothGrowthData$supplement <-gsub("VC","Vitamin
C",myToothGrowthData$supplement)
myToothGrowthData$supplement <-gsub("OJ","Orange
Juice",myToothGrowthData$supplement)

#Convert dose to a factor (this will ensure the data is analyzed properly)
myToothGrowthData$dose <- as.factor(myToothGrowthData$dose)

#Some checks
myToothGrowthData$length
```

```
##  [1]  4.2 11.5  7.3  5.8  6.4 10.0 11.2 11.2  5.2  7.0 16.5 16.5 15.2 17.3
## [15] 22.5 17.3 13.6 14.5 18.8 15.5 23.6 18.5 33.9 25.5 26.4 32.5 26.7 21.5
## [29] 23.3 29.5 15.2 21.5 17.6  9.7 14.5 10.0  8.2  9.4 16.5  9.7 19.7 23.3
## [43] 23.6 26.4 20.0 25.2 25.8 21.2 14.5 27.3 25.5 26.4 22.4 24.5 24.8 30.9
## [57] 26.4 27.3 29.4 23.0
```

```
myToothGrowthData$supplement
```

```
##  [1] "Vitamin C"    "Vitamin C"    "Vitamin C"    "Vitamin C"
##  [5] "Vitamin C"    "Vitamin C"    "Vitamin C"    "Vitamin C"
##  [9] "Vitamin C"    "Vitamin C"    "Vitamin C"    "Vitamin C"
## [13] "Vitamin C"    "Vitamin C"    "Vitamin C"    "Vitamin C"
## [17] "Vitamin C"    "Vitamin C"    "Vitamin C"    "Vitamin C"
## [21] "Vitamin C"    "Vitamin C"    "Vitamin C"    "Vitamin C"
## [25] "Vitamin C"    "Vitamin C"    "Vitamin C"    "Vitamin C"
```

```
## [29] "Vitamin C"    "Vitamin C"    "Orange Juice" "Orange Juice"
## [33] "Orange Juice" "Orange Juice" "Orange Juice" "Orange Juice"
## [37] "Orange Juice" "Orange Juice" "Orange Juice" "Orange Juice"
## [41] "Orange Juice" "Orange Juice" "Orange Juice" "Orange Juice"
## [45] "Orange Juice" "Orange Juice" "Orange Juice" "Orange Juice"
## [49] "Orange Juice" "Orange Juice" "Orange Juice" "Orange Juice"
## [53] "Orange Juice" "Orange Juice" "Orange Juice" "Orange Juice"
## [57] "Orange Juice" "Orange Juice" "Orange Juice" "Orange Juice"
```

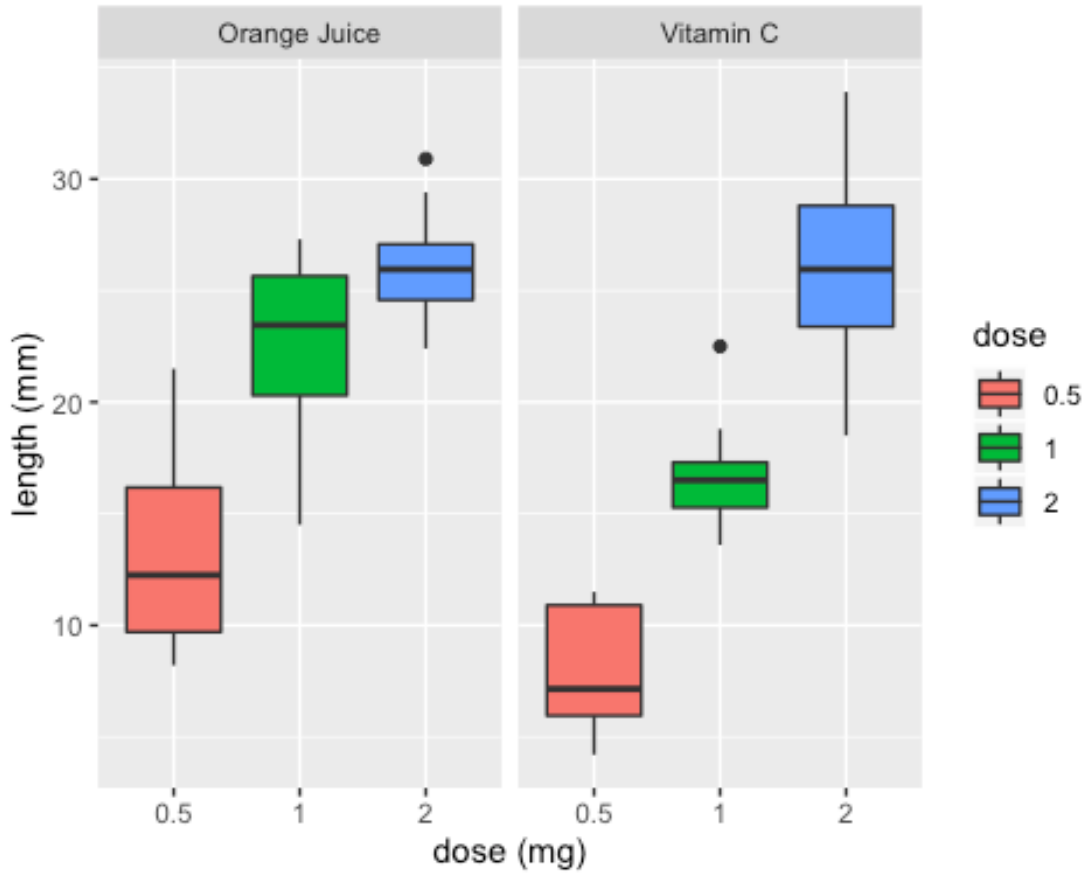
```
myToothGrowthData$dose
```

```
##  [1] 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 1   1   1   1   1   1   1
## [18] 1   1   1   2   2   2   2   2   2   2   2   2   2   0.5 0.5 0.5 0.5
## [35] 0.5 0.5 0.5 0.5 0.5 0.5 1   1   1   1   1   1   1   1   1   1   2
## [52] 2   2   2   2   2   2   2   2   2
## Levels: 0.5 1 2
```

Analysis

Compare tooth length vs dose for Vitamin C and orange juice

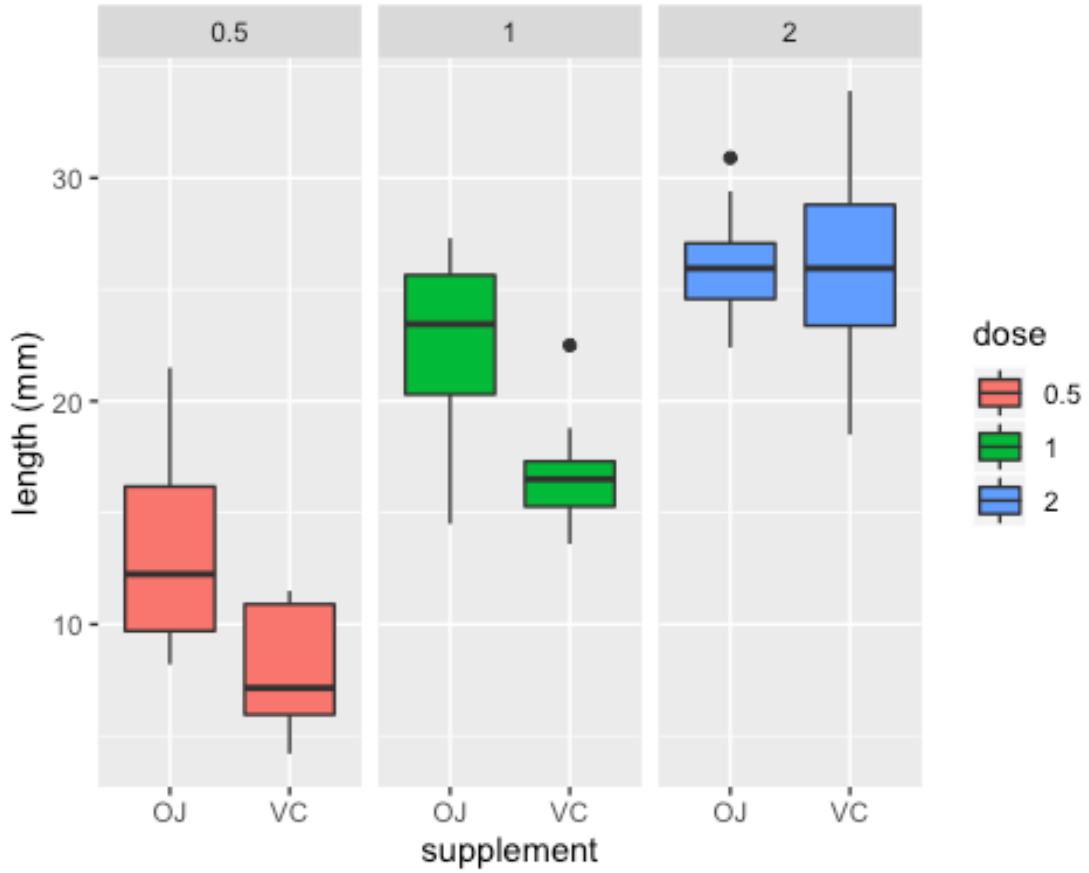
```
ggplot(myToothGrowthData, aes(x=dose, y=length, fill=dose)) +geom_boxplot() +
facet_grid(~supplement) + xlab("dose (mg)") + ylab("length (mm)")
```



Compare tooth length vs supplement type for different supplement doses

```
#Replace Orange Juice and Vitamin C with OJ and VC
myToothGrowthData$supplement <-gsub("Orange
Juice","OJ",myToothGrowthData$supplement)
myToothGrowthData$supplement <-gsub("Vitamin
C","VC",myToothGrowthData$supplement)

ggplot(myToothGrowthData, aes(x=supplement, y=length, fill=dose))
+geom_boxplot() + facet_grid(~dose) + ylab("length (mm)")
```



Based upon the information from the graphs, it appears higher supplement doses result in more tooth growth whereas supplement type does not. Let's do some t tests.

HYPOTHESIS: Supplement type does not affect tooth growth
t test for Supplement Type/Delivery Method

```
dose <- myToothGrowthData$dose
supplement <- myToothGrowthData$supplement
toothLength <- myToothGrowthData$length
t.test(toothLength~supplement, paired=FALSE, var.equal = FALSE)

##
##  Welch Two Sample t-test
##
```

```
## data:  toothLength by supplement
## 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

#the confidence interval contains 0
```

With CI = 95% and p (0.06063) greater than α (0.05), the result is determined to be ***statistically insignificant*** and we fail to reject the hypothesis "Supplement type does not affect tooth growth."

HYPOTHESIS: Supplement dose does not affect tooth growth

t test for Dosage (0.5 mg vs 1 mg)

```
t.test(toothLength[dose==0.5],toothLength[dose==1], paired = FALSE)

##
##  Welch Two Sample t-test
##
## data:  toothLength[dose == 0.5] and toothLength[dose == 1]
## 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 of x mean of y
##    10.605    19.735

#the confidence interval does not contain 0
```

t test for Dosage (1 mg vs 2 mg)

```
t.test(toothLength[dose==1],toothLength[dose==2], paired = FALSE)

##
##  Welch Two Sample t-test
##
## data:  toothLength[dose == 1] and toothLength[dose == 2]
## 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 of x mean of y
##    19.735    26.100

#the confidence interval does not contain 0
```

t test for Dosage (0.5 mg vs 2 mg)

```
t.test(toothLength[dose==0.5],toothLength[dose==2], paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  toothLength[dose == 0.5] and toothLength[dose == 2]
## 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 of x mean of y
##    10.605    26.100

#the confidence interval does not contain 0
```

For each test, CI = 95% and α (0.05) was greater than p. Each result is ***statistically significant***. The hypothesis "Supplement dose does not affect tooth growth" can be rejected.

Conclusion

Higher supplement doses result in more tooth growth. The type of supplement, whether it's Vitamin C or orange juice, does not affect tooth growth.