Statistical Inference Project 2 - Part 2

Daniel

1/20/2020

Loading and preprocessing the data

Set Global Options

```
knitr::opts_chunk$set(warnings=FALSE, message=FALSE, echo = TRUE)
```

Question 1

Load the ToothGrowth data and perform some basic exploratory data analyses

```
# Load dataset
data("ToothGrowth")
# use str function to look at data
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 ...
```

```
# Change dose to factor
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
# Check that supp and dose are of same type
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: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

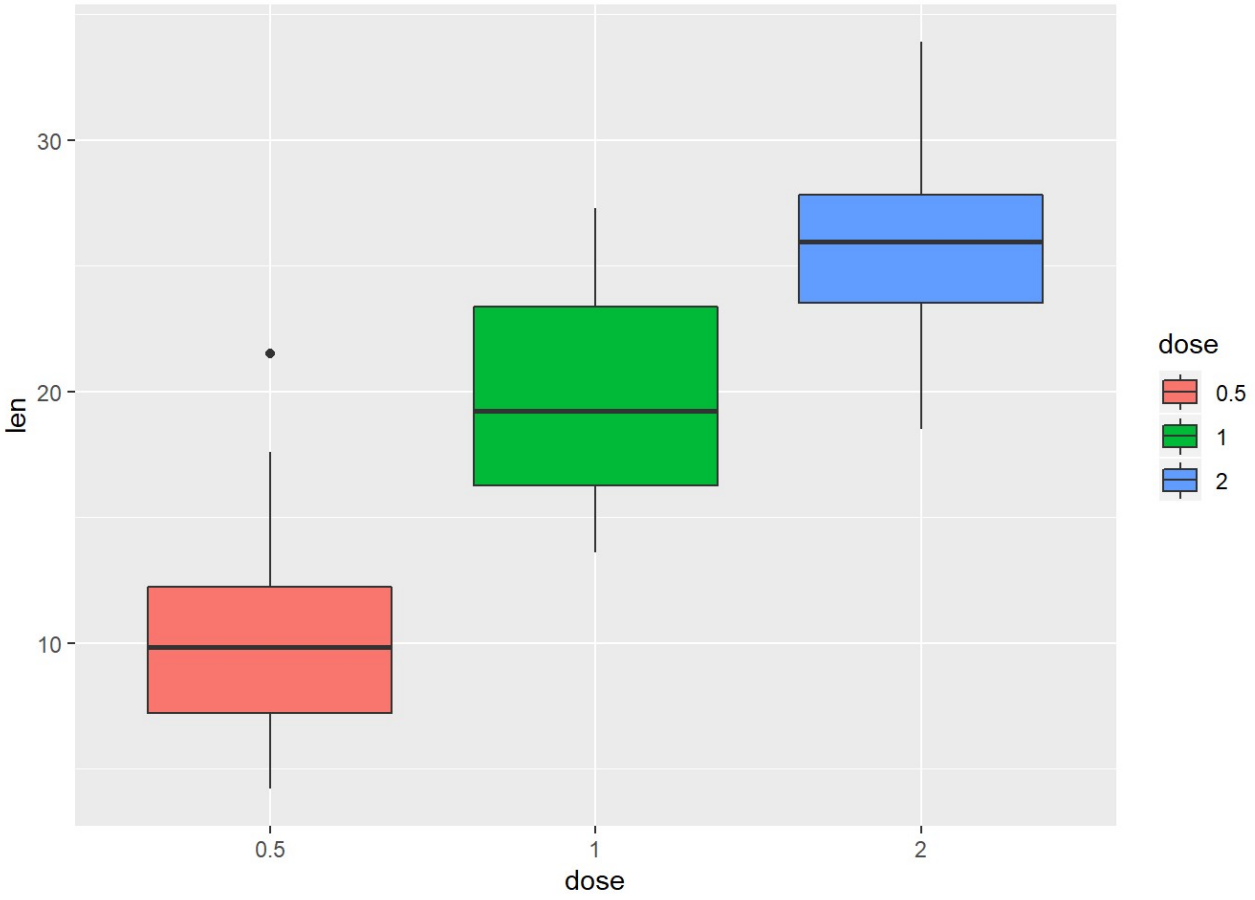
Question 2

Provid a basic summary of the data

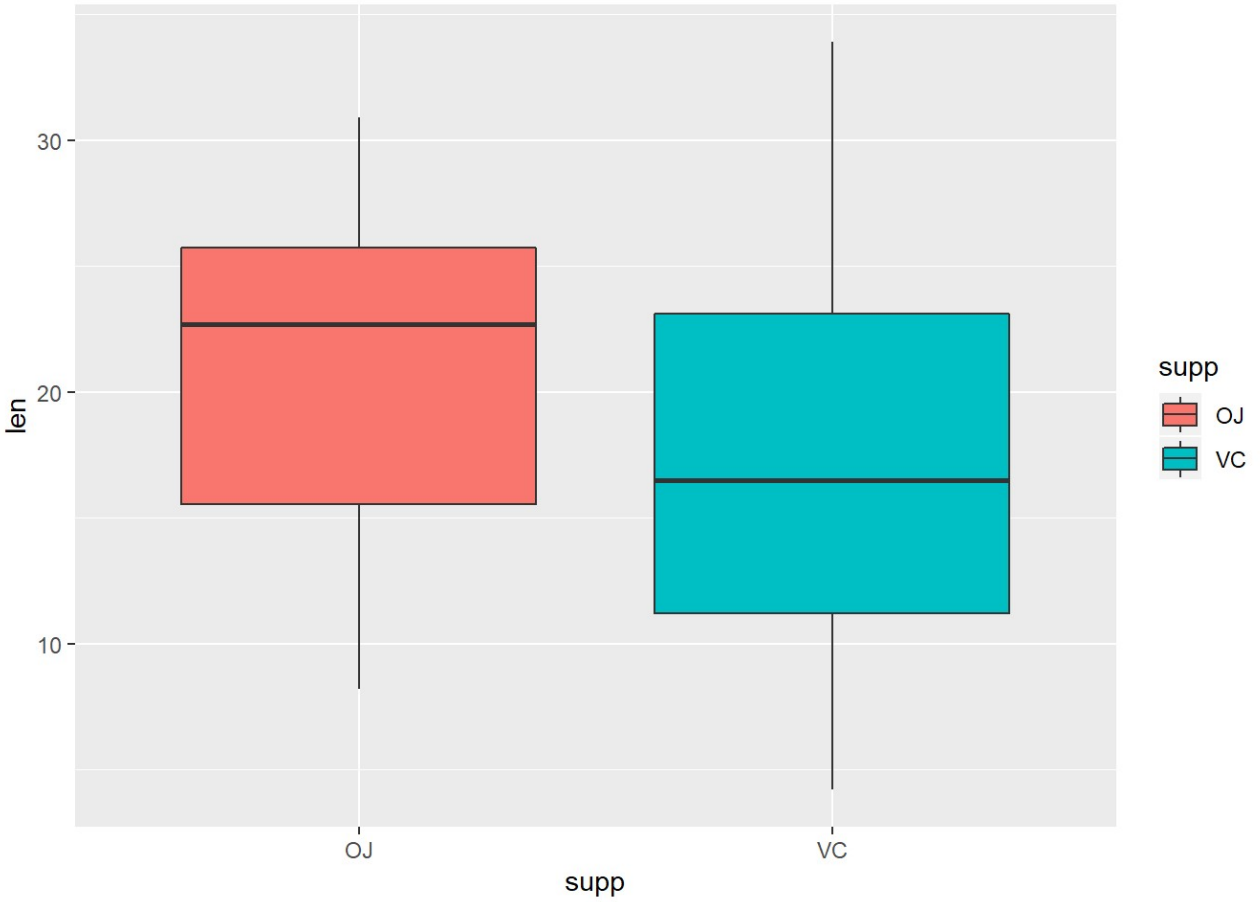
```r
# Call ggplot2
library(ggplot2)
# use summmary fct to show breakdown of dataset
summary(ToothGrowth)
```

```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

```r
# Create a boxplot graph
ggplot(data = ToothGrowth, aes(x = dose, y = len)) + geom_boxplot(aes(fill = dose))
```




```r
# Create graphics for toothgrowth from supplement type
ggplot(data = ToothGrowth, aes(x = supp, y = len)) + geom_boxplot(aes(fill = supp))
```



Question 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)

```
# Check for difference due to different supplement type. Assume unequal variances for the two groups
t.test(len~supp, 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
```

The p value is 0.06 and the range of the confidence interaval so we cannot reject the null hypothesis that different supplements have no effect on tooth length

Next, we can look at the three different doses and their effects on tooth length

```
# First we will break the ToothGrowth data into three subsets, one for each dose
dose0.5_1.0 <- subset(ToothGrowth, dose %in% c(0.5, 1.0))
dose0.5_2.0 <- subset(ToothGrowth, dose %in% c(0.5, 2.0))
dose1.0_2.0 <- subset(ToothGrowth, dose %in% c(1.0, 2.0))

# Test for doses
t.test(len ~ dose, data = dose0.5_1.0)
```

```
##
##  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
```

```
t.test(len ~ dose, data = dose0.5_2.0)
```

```
##
##  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
```

```
t.test(len ~ dose, data = dose1.0_2.0)
```

```
##
##  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
```

All three t tests produce p values under 5% and do not include 0 in their confidence intervals so we can reject the hypothesis. These test show that the length of teeth increases as you increase the dose

Question 4

State your conclusions and the assumptions needed for your conclusions

Conclusions

1. The type of the supplement does not effect tooth length
2. The dose given does effect tooth length

Assumptions

1. The sample population was representative of the total population
2. The experiment was performed in such a way that the doses and supplements were randomly administered