Basic inferential data analysis

Amy Lee

11/9/2019

Basic Inferential Data Analysis

Now in the second portion of the project, we're going to 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.

Overview

This project was completed if the tooth length could differ by the supplement type. The test concluded that the supplement type does not affect the tooth length.

Necessary Packages

```r
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
```

```r
library(ggplot2)
```

1) Load the ToothGrowth data and perform some basic exploratory data analyses

```r
data<-ToothGrowth
str(data)
```

```
## '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 ...
```

```r
head(data)
```

```
##    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
```

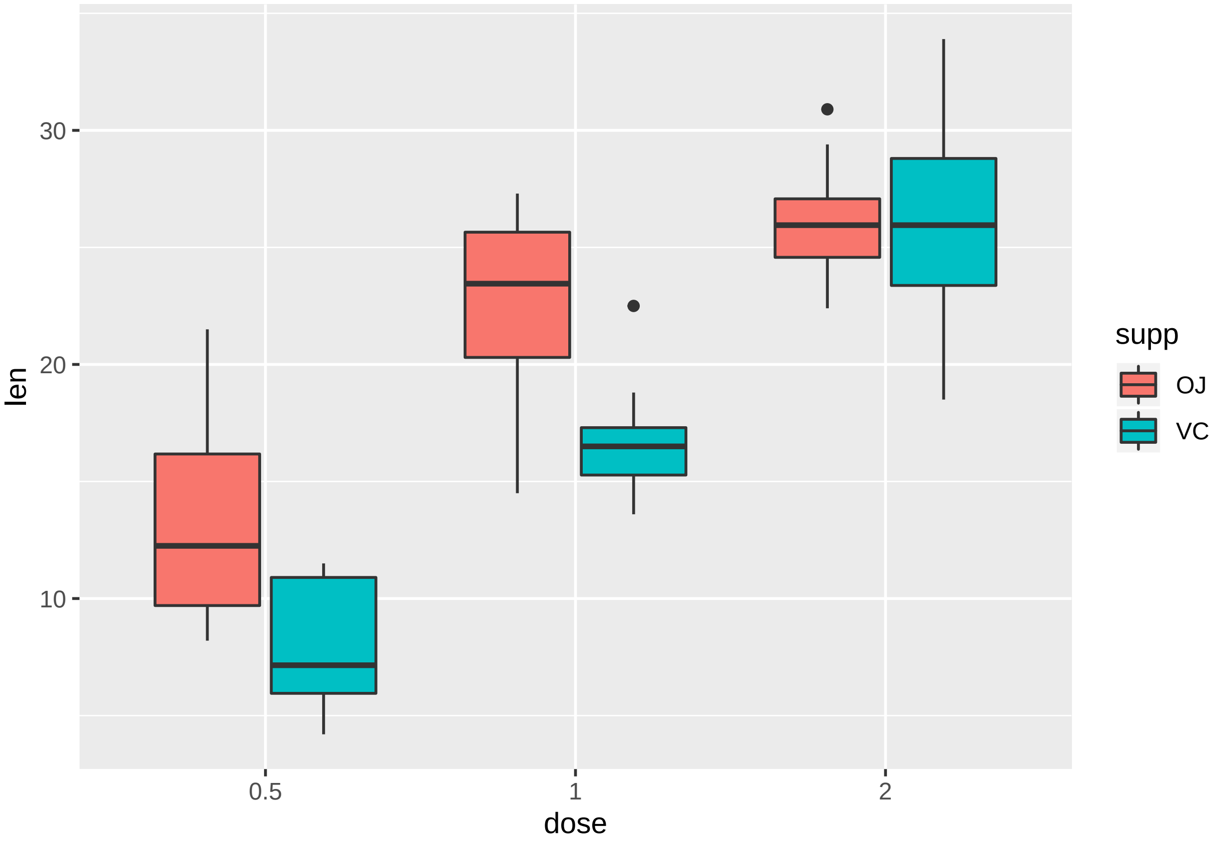
```
data$dose<-as.factor(data$dose)
```

According to R Documentation, data ToothGrowth is based on an experiment on 60 guinea pigs. These guinea pigs received different dosages (0.5,1,2) of vitamine C with two supplementary method (OJ, VC) and see if the response of tooth lengths differ significantly.

According to the outcome of str function, it is better to have dose variable as a factor instead of numeric. Thus, the conversion was made.

2) Provide a basic summary of the data.

```
ggplot(data, aes(x=dose, y=len, fill=supp))+geom_boxplot()
```



```
g_data<-data%>%group_by(dose, supp)
summarize(g_data,mean=mean(len))
```

```
## # A tibble: 6 x 3
## # Groups:   dose [3]
##   dose  supp   mean
##   <fct> <fct> <dbl>
## 1 0.5   OJ    13.2
## 2 0.5   VC     7.98
## 3 1     OJ    22.7
## 4 1     VC    16.8
## 5 2     OJ    26.1
## 6 2     VC    26.1
```

```r
summarize(g_data,n())
```

```
## # A tibble: 6 x 3
## # Groups:   dose [3]
##   dose  supp  `n()`
##   <fct> <fct> <int>
## 1 0.5   OJ       10
## 2 0.5   VC       10
## 3 1     OJ       10
## 4 1     VC       10
## 5 2     OJ       10
## 6 2     VC       10
```

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)

```r
OJ<-data%>%filter(supp=="OJ")%>%select(len,dose)
VC<-data%>%filter(supp=="VC")%>%select(len,dose)
t.test(OJ$len, VC$len, paired=FALSE)
```

```
## 
##  Welch Two Sample t-test
## 
## data:  OJ$len and VC$len
## 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
```

According to the t test, the p value is equal to 0.06, which is greater than 0.05. Which means, the null hypothesis is rejected.

4) State your conclusions and the assumptions needed for your conclusions.

With the p value greater than 0.05, the two sample t test fails to reject the hypothesis of mu1-mu2=0, which means the data is not sufficient to prove that one supplementary type is better than the other.

To come to this conclusion, some assumption has to be made. These are: 1) normality 2) randomness 3) equal variance