Statistical Inference Course Project - Part 2

Bernardo Najlis

October 24, 2015

Introduction

In the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.

- Load the ToothGrowth data and perform some basic exploratory data analyses.
- Provide a basic summary of the data.
- 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)
- State your conclusions and the assumptions needed for your conclusions.

Load the ToothGrowth data and basic exploratory analysis

```
library(datasets)
data("ToothGrowth")
ToothGrowth$dose <- as.factor(ToothGrowth$dose)

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 ...
```

```
nrow(ToothGrowth)
```

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

Basic summary of the data

```
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
```

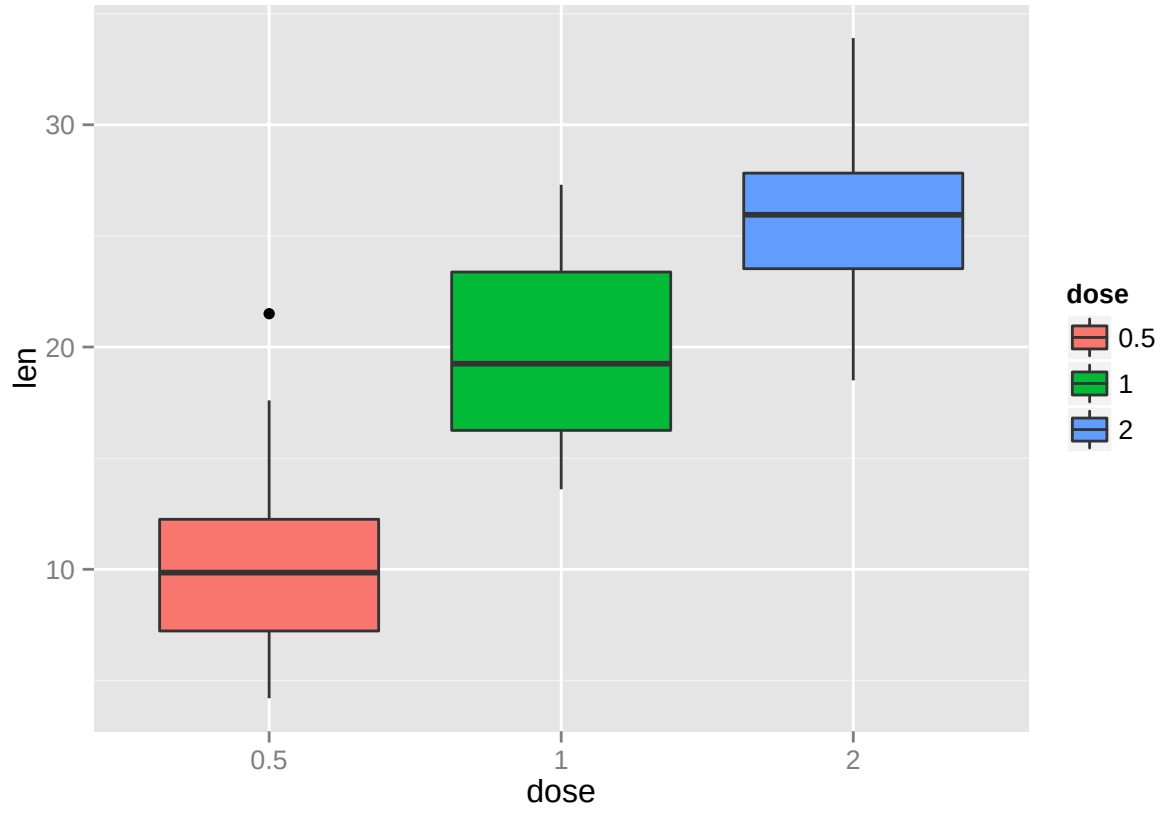
```
table(ToothGrowth$dose, ToothGrowth$supp)
```

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

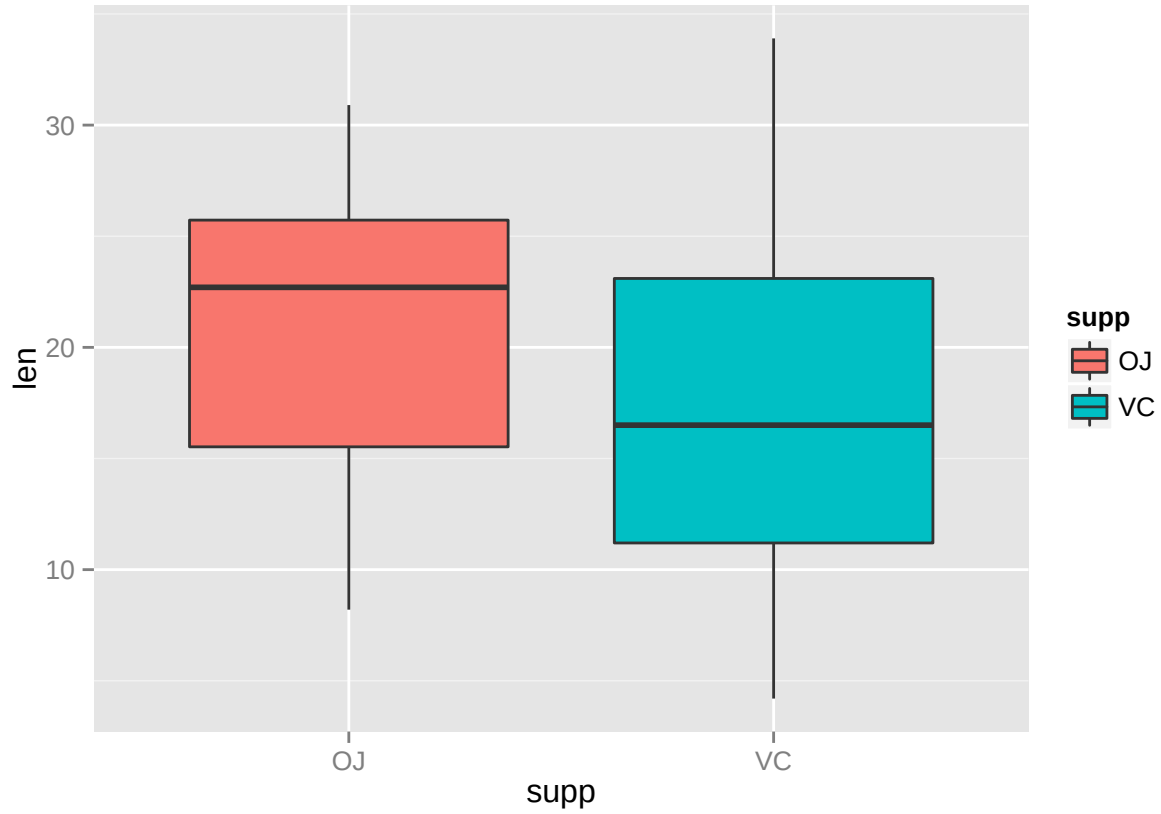
```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.2.2
```

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




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



Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose

```
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
```

P-value is 0.06 but as the confidence interval contains zero, we cannot reject the null hypothesis (different supplement types have no effect on tooth length). We will create three sub groups based on dose level pairs and repeat the test on each one.

```r
ToothGrowth_sub1 <- subset(ToothGrowth, dose %in% c(0.5, 1.0))
ToothGrowth_sub2 <- subset(ToothGrowth, dose %in% c(0.5, 2.0))
ToothGrowth_sub3 <- subset(ToothGrowth, dose %in% c(1.0, 2.0))

t.test(len ~ dose, data = ToothGrowth_sub1)
```

```
##
##  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
```

```r
t.test(len ~ dose, data = ToothGrowth_sub2)
```

```
##
##  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
```

```r
t.test(len ~ dose, data = ToothGrowth_sub3)
```

```
##
##  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 the new three subsets, the p-value is now less than 0.05 and the condifence interval does not contain 0. Also, the mean tooth length increases when raising the dose level, meaning that we can reject the null hypothesis and establish that *increasing the dose level leads to an increase in tooth length.*

Conclusions and assumptions

- The supplement type has no effect on tooth growth
- Increasing the dose level leads to an increase in tooth length

For these conclusions to be valid, the following assumptions would have to be valid:

- the guinea pigs are repesentative for the population of guinea pigs,
- dosage and supplement were randomly assigned and
- the distribution of the means is normal.