Statistical Inference Course Project 1

Chuk Yong

6 September 2016

Coursera Statistical Inference Course Project Part 2: Basic Inferential Data Analysis Instructions

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.

Start program. Load ToothGrowth Data and investigate its structure

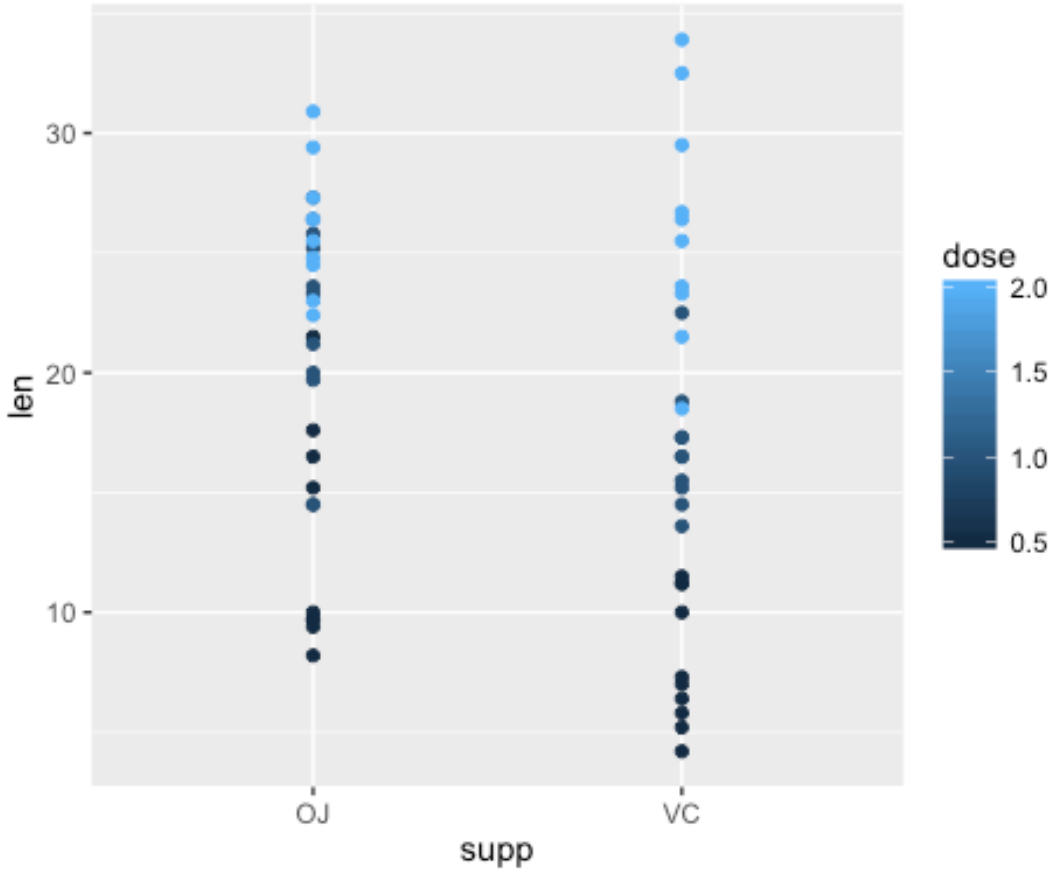
```
library(datasets)
library(dplyr)
library(ggplot2)

# Load ToothGrowth Data
data("ToothGrowth")
str(ToothGrowth)
```

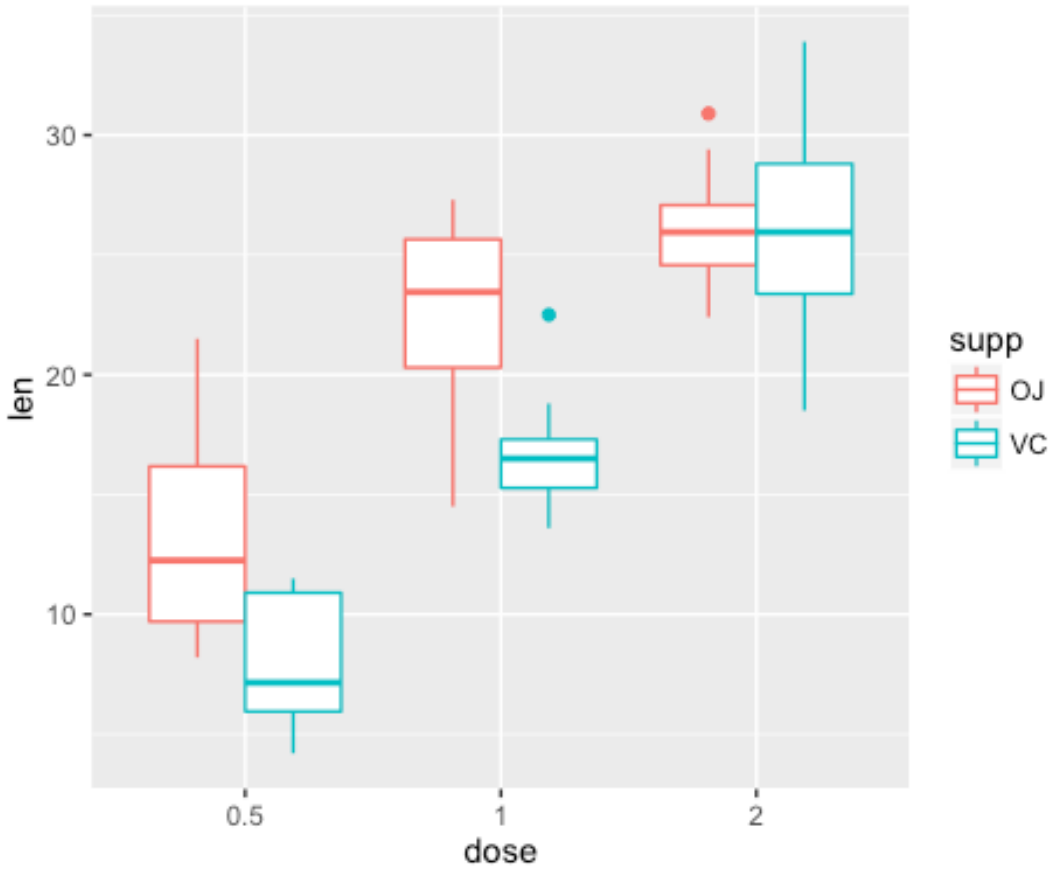
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

len : Tooth Length is numeric
supp: Supplement is Factor consisting of OJ and VC
dose: Dosage is numeric. As it is just 0.5, 1 and 2, we will have to convert it to factor so that it is easier for analysis

Doing a quick plot to see if there are any patterns



There seems to be some patterns between supp and dose. Let's use boxplot to investigate further



We noticed: 1. There is a trend between dose and len.
2. There is a trend between dose and supp.

Hypothesis Testing No.1: Overall impact of dosage on tooth growth

Null Hypothesis - There is no effect of dosage on tooth growth

Alternate Hypothesis - Higer doses produce greater tooth growth.

```
# arrange our dateset
dose05 <- filter(ToothGrowth, dose == 0.5)
dose1 <- filter(ToothGrowth, dose == 1)
dose2 <- filter(ToothGrowth, dose == 2)
```

Comparing dosage of 1mg vs 0.5mg

```
# compare between dose of 0.5 and 1
t.test(dose1$len, dose05$len, alternative = "greater")

##
##  Welch Two Sample t-test
##
## data:  dose1$len and dose05$len
## t = 6.4766, df = 37.986, p-value = 6.342e-08
```

```
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  6.753323      Inf
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

As P value is less than 0.001, it is highly significant and we can reject the null hypothesis

Comparing dosage of 2mg vs 1mg

```
t.test(dose2$len, dose1$len, alternative = "greater")

##
##  Welch Two Sample t-test
##
## data:  dose2$len and dose1$len
## t = 4.9005, df = 37.101, p-value = 9.532e-06
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  4.17387     Inf
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

As P value is less than 0.001, it is highly significant and we can reject the null hypothesis

So in both casess, we reject the null hypothesis.
We accept the althernative: higher dosages do have a positive influence on toothgrowth.

Hypothesis Testing No 2: Differences between supplements by dosage

Null Hypothesis - There is no effect between supplement OJ and VC

Alternate Hypothesis - OJ is better on toothgrowth than VC

We will do three tests and compare them based on dosages.

Dosage: 0.5mg

```
OJdose05 <- filter(dose05, supp == "OJ")
VCdose05 <- filter(dose05, supp == "VC")
t.test(OJdose05$len, VCdose05$len, alternative = "greater")

##
##  Welch Two Sample t-test
##
## data:  OJdose05$len and VCdose05$len
## t = 3.1697, df = 14.969, p-value = 0.003179
## alternative hypothesis: true difference in means is greater than 0
```

```
## 95 percent confidence interval:
##  2.34604     Inf
## sample estimates:
## mean of x mean of y
##     13.23      7.98
```

P value is lower than 0.05. We reject the null hypothesis.
We accept the alternative: OJ is a better supplement than VC at dosage of 0.5mg

Dosage: 1mg

```
OJdose1 <- filter(dose1, supp == "OJ")
VCdose1 <- filter(dose1, supp == "VC")
t.test(OJdose1$len, VCdose1$len, alternative = "greater")

##
##  Welch Two Sample t-test
##
## data:  OJdose1$len and VCdose1$len
## t = 4.0328, df = 15.358, p-value = 0.0005192
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  3.356158      Inf
## sample estimates:
## mean of x mean of y
##     22.70     16.77
```

P value is lower than 0.05. We reject the null hypothesis.
We accept the alternative: OJ is a better supplement than VC at dosage of 1mg

Dosage: 2mg

```
OJdose2 <- filter(dose2, supp == "OJ")
VCdose2 <- filter(dose2, supp == "VC")
t.test(OJdose2$len, VCdose2$len, alternative = "greater")

##
##  Welch Two Sample t-test
##
## data:  OJdose2$len and VCdose2$len
## t = -0.046136, df = 14.04, p-value = 0.5181
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -3.1335     Inf
## sample estimates:
## mean of x mean of y
##     26.06     26.14
```

P value is higher than 0.05.
We cannot reject the null hypothesis: at dosage of 2mg, there is no significant difference between the two supplement

Conclusion for Hypothesis Testing 2

For dosages of 0.5mg and 1mg, there are significant differences: OJ is better than VC in promoting tooth growth. At 2mg however, there is no significant differences.