Inference Project4: Basic Inference Analysis

Shuai Wang

Dec, 11, 2016

Part 2: Basic Inferential Data Analysis

load the ToothGrowth data and perform some basic exploratory data analyses

Here I load the ToothGrowth data with basic exploratory data analyses using ggplot2(). I also use summary() function to provide some basic exploratory data analysis.

```
library(datasets)
data(ToothGrowth)
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.3.2
```

```
library(reshape2)
ToothGrowth$supp<-as.factor(ToothGrowth$supp)
ToothGrowth$dose<-as.factor(ToothGrowth$dose)
melted_data<-melt(ToothGrowth,id.vars=c("supp","dose"))
#some ggplot examples are from http://www.cookbook-r.com/Graphs/Plotting_means_and_error_bars_(ggplot2)/
library(Rmisc)#get the mean, sd information for plotting
```

```
## Loading required package: lattice
```

```
## Loading required package: plyr
```

```
melted_data.Summary <- summarySE(melted_data, measurevar="value", groupvars=c("supp",
"dose"))
  # Use 95% confidence intervals
  p.ci<-ggplot(melted_data.Summary, aes(x=dose, y=value, fill=supp)) +
    geom_bar(position=position_dodge(), stat="identity") +
    geom_errorbar(aes(ymin=value-ci, ymax=value+ci),
              width=.2,                      # Width of the error bars
              position=position_dodge(.9))+
      labs(title="Exploratory Analysis of ToothGrowth Data",x="Dose Level",y="Length")
+
      scale_fill_brewer(name="Supply Name")+
          theme_bw()
plot(p.ci)
```
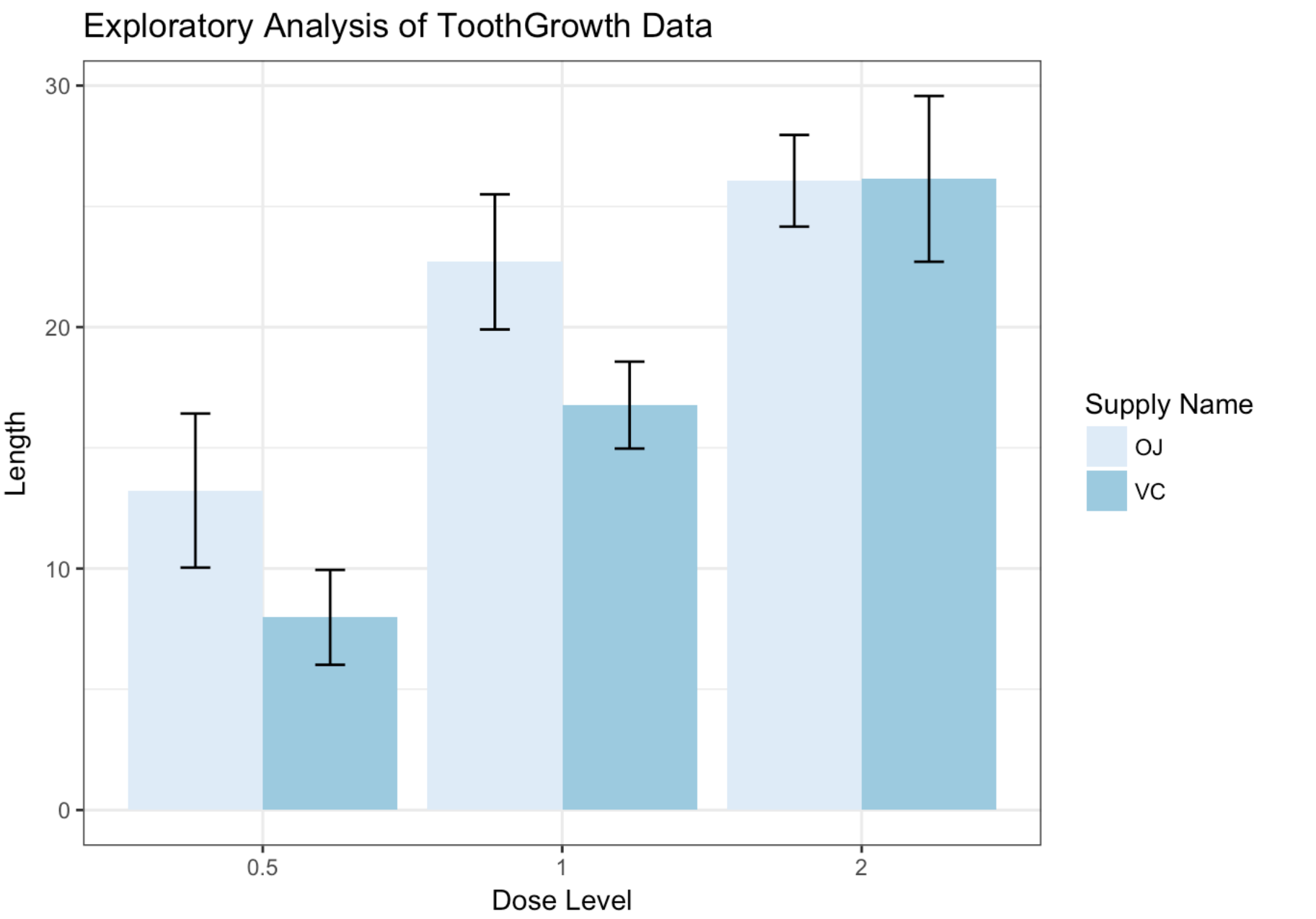

In the Figure above, The general summary of the data is presented. At x-axis, Mean lengths and 95% confidence Intervals of each supply at each dose level are plotted.

Provide a basic summary of the data.

Explanation: The ToothGrowth dataset has 60 observations with three variables. Variable "length" responses to the length of the cell that reflects to the tooth growth of guinea pigs. The "supply" variable has 2 factors. The "dose" variable, which responses to the dose level of corresponding supplies.

```
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
```

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)

Below are codes to run t tests at different dose levels. Welch t-test is used.

```
alpha<-0.05
nTestAlpha<-seq(1,3)/3*alpha
Dose.05<-subset(ToothGrowth,dose==0.5)
Dose.1<-subset(ToothGrowth,dose==1)
Dose.2<-subset(ToothGrowth,dose==2)
test.05<-t.test(len~supp,Dose.05)
test.1<-t.test(len~supp,Dose.1)
test.2<-t.test(len~supp,Dose.2)
sprintf("T test Result at Dose 0.5")
```

```
## [1] "T test Result at Dose 0.5"
```

```
test.05
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23            7.98
```

```
sprintf("T test Result at Dose 1")
```

```
## [1] "T test Result at Dose 1"
```

```
test.1
```

```
##
##   Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

```
sprintf("T test Result at Dose 2")
```

```
## [1] "T test Result at Dose 2"
```

```
test.2
```

```
##
##   Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

State your conclusions and the assumptions needed for your conclusions.

My conclusions: At Dose level 0.5,1: The supply has a significant effect on the mean length. Lengths supplied by OJ are significantly longer than those supplied by TC. At Dose level 2, the supply doesn't have a significant effect on the mean length.