Stat-Inference Assignment-02

Vinsent P.

September 26, 2015

Objective: 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 ### techniques from class, even if there's other approaches worth considering) ### 5. State your conclusions and the assumptions needed for your conclusions.

```
library(ggplot2)
```

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

```
### Load the dataset
data(ToothGrowth)
```

Explore the Data

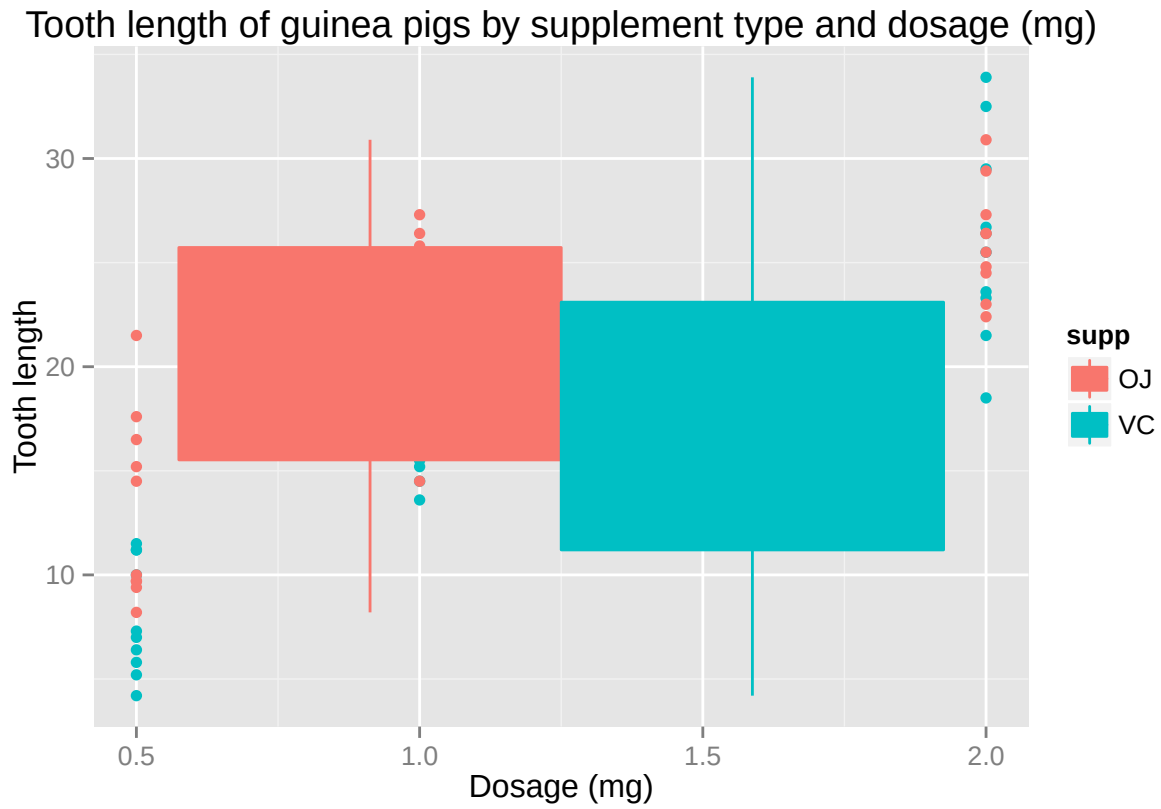
```
head(ToothGrowth)
```

```
##    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
```

```
summary(ToothGrowth)
```

```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

```
# visualize the data
qplot(dose,len,data=ToothGrowth, col=supp, main="Tooth length of guinea pigs by supplement type and dosa
```



##Hypothesis Testing

H0: There is no significant difference in treatment from both the medicine OJ and VC

H1: alternate of H0

Tooth length in VC = OJ H0

Tooth length in VC NOT equal to OJ treaments

Split the two treatment groups

```
OJ = ToothGrowth$len[ToothGrowth$supp == 'OJ']
VC = ToothGrowth$len[ToothGrowth$supp == 'VC']
```

T- Test

one tail test

```
t.test(OJ, VC, alternative = "greater", paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

##

```
##  Welch Two Sample t-test
##
## data:  OJ and VC
## t = 1.9153, df = 55.309, p-value = 0.03032
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.4682687       Inf
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

Hypothesis for dosage

HO: there is no sifnificant difference in dosage in two sizes

H1: There is difference in the two dosage

```
##Divide the dataset with dosage group

OJ2 = ToothGrowth$len[ToothGrowth$supp == 'OJ' & ToothGrowth$dose == 2]
VC2 = ToothGrowth$len[ToothGrowth$supp == 'VC' & ToothGrowth$dose == 2]

half = ToothGrowth$len[ToothGrowth$dose == 0.5]
one = ToothGrowth$len[ToothGrowth$dose == 1]
two = ToothGrowth$len[ToothGrowth$dose == 2]


# T-test

##One-tailed independent t-test with unequal variance.

t.test(half, one, alternative = "less", paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  half and one
## t = -6.4766, df = 37.986, p-value = 6.342e-08
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -6.753323
## sample estimates:
## mean of x mean of y
##    10.605    19.735
```

p-value = 6.342e-08 is less than the 5% significance level .05 Hence accepting the Alternate hypothesis as there is difference in tooth growth.

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.