statistical Inference

Tanika

June 18, 2018

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.

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

```
library(datasets)
data("ToothGrowth")
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
```

```
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 ...
```

2.Provide a basic summary of the data.

```
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
```

```
mean(ToothGrowth$len)
```

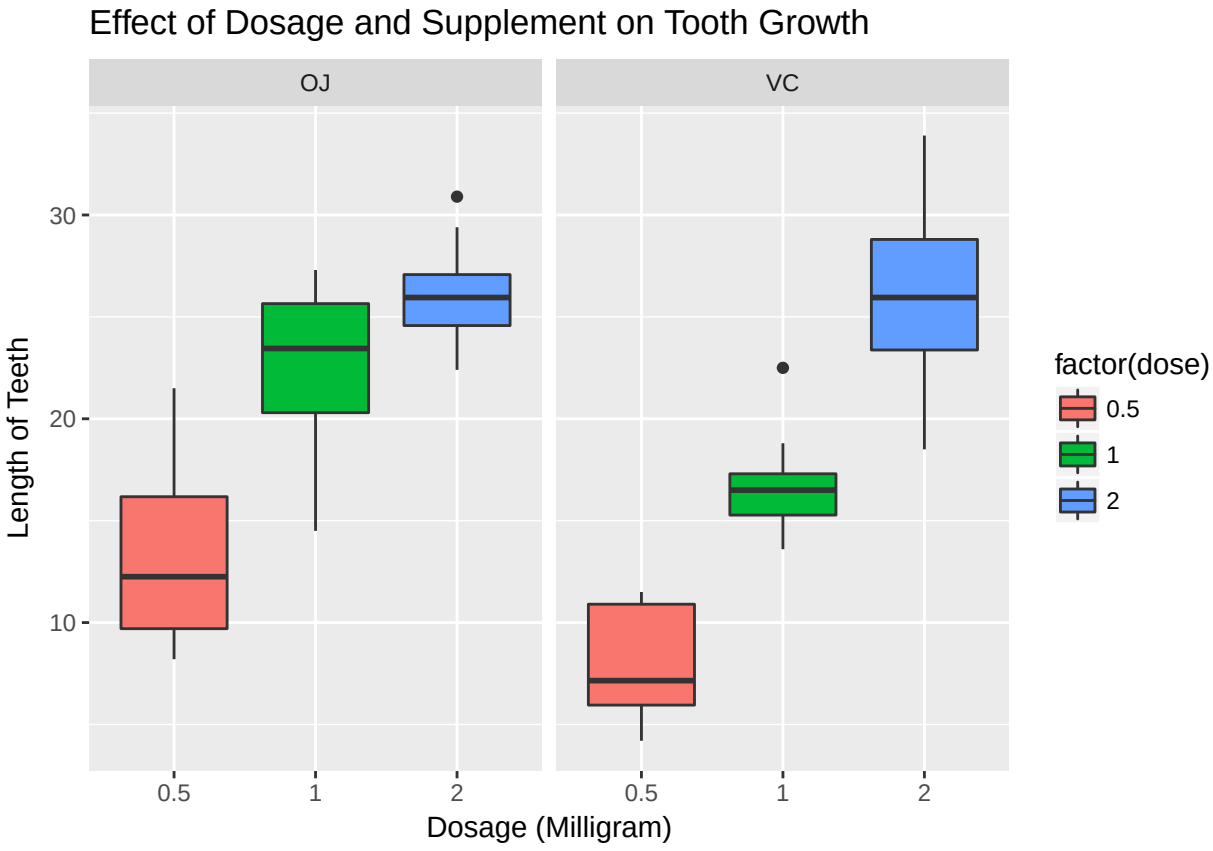
```
## [1] 18.81333
```

```
sd(ToothGrowth$len)
```

```
## [1] 7.649315
```

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)

```
library(ggplot2)
p <- ggplot(ToothGrowth,
               aes(x=factor(dose),y=len,fill=factor(dose)))
p + geom_boxplot(notch=F) + facet_grid(.~supp) +
     scale_x_discrete("Dosage (Milligram)") +
     scale_y_continuous("Length of Teeth") +
     ggtitle("Effect of Dosage and Supplement on Tooth Growth")
```



T.Test on the Toothgrowth data with supp variable

```
supp <- ToothGrowth$supp
```

```r
t.test(ToothGrowth$len[supp == "OJ"],ToothGrowth$len[supp == "VC"],paired = FALSE,var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  ToothGrowth$len[supp == "OJ"] and ToothGrowth$len[supp == "VC"]
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1670064  7.5670064
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

With the Supp T-Test, the p-value is 0.06063 which is close to 0.05, and the Confidence Interval ranges from -0.1710156 to 7.5710156 which includes ZERO.

T.Test on the Toothgrowth data with dose variable

```r
dose <- ToothGrowth$dose
t.test(ToothGrowth$len[dose == 1],ToothGrowth$len[dose == 2],paired = FALSE,var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  ToothGrowth$len[dose == 1] and ToothGrowth$len[dose == 2]
## t = -4.9005, df = 38, p-value = 1.811e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.994387 -3.735613
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```

The Dose Variable T-Test has a p-value that is 1.811e-05, which is almost 0, and the Confidence Interval ranges from -8.994387 to -3.735613.

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

With the help of t-tests to compare Tooth Growth by supp and dose, we found out that the higher the dose, the longer the tooth grows in length.

We noticw that OJ is more effective in Tooth Growth than VC in low doses such as 0.5 and 1.0 as compare to 2.0.