Statistical Inference - Week 4 Course Project - Part 2

Sanjay Lonkar

14 Jun 2018

Basic Inferential Data Analysis

Overview

This assignment is Part 2 of Statistical Inference - Week 4 course project. Goal of this assignment is to perform basic inferencetial anlysis and draw fair conclustions:

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

```
# Load necessary libraries
library (ggplot2)

## Warning: package 'ggplot2' was built under R version 3.4.2

library (dplyr)

## Warning: package 'dplyr' was built under R version 3.4.2

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

# Load ToothGrowth data set
library (datasets)
data (ToothGrowth)
```

Some basic exploratory data analysis

```
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
```

```
tail (ToothGrowth)
```

```
##     len supp dose
## 55 24.8   OJ    2
## 56 30.9   OJ    2
## 57 26.4   OJ    2
## 58 27.3   OJ    2
## 59 29.4   OJ    2
## 60 23.0   OJ    2
```

```
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 ...
```

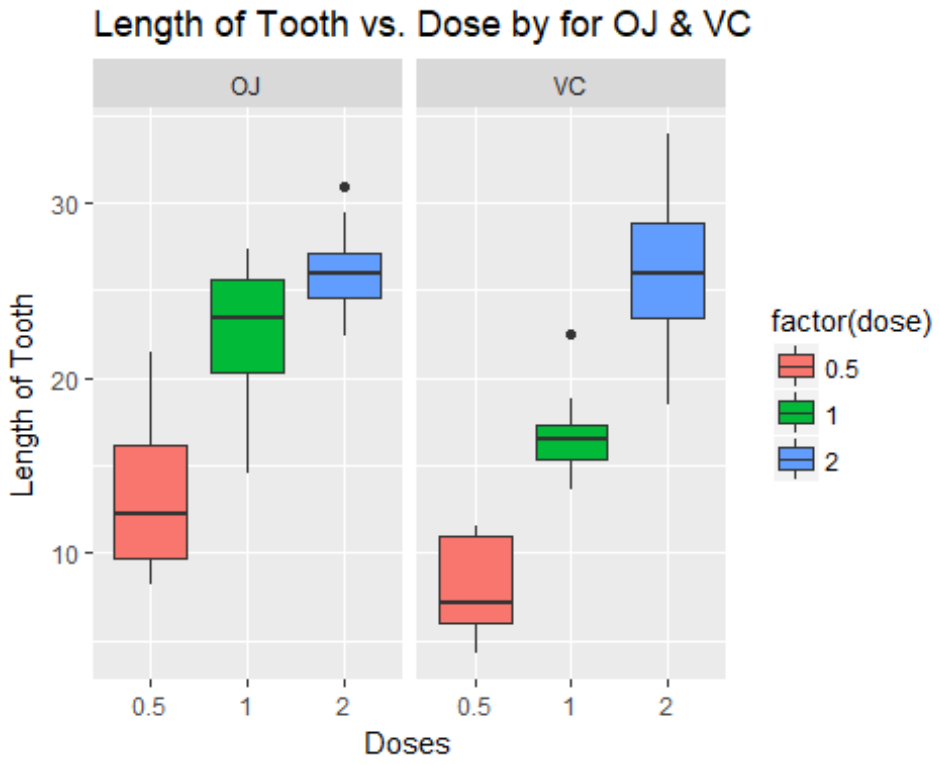
```
# Sample Size & Number of Rows and Columns of data frame
length (ToothGrowth)
```

```
## [1] 3
```

```
# Number of Rows and Columns
dim (ToothGrowth)
```

```
## [1] 60  3
```

```
# Boxplot graph of the tooth length vs the does
p <- ggplot (ToothGrowth, aes (x = factor(dose), y = len, fill =
factor(dose))) +
  geom_boxplot () +
  facet_grid (.~supp) +
  labs (title = "Length of Tooth vs. Dose by for OJ & VC",
        x = "Doses", y = "Length of Tooth")
print (p)
```



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
```

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

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

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)

```
# Dose 0.5
doseOJ0.5 <- ToothGrowth %>% filter (supp == "OJ" & dose == "0.5")
```

```
## Warning: package 'bindrcpp' was built under R version 3.4.2
```

```r
doseVC0.5 <- ToothGrowth %>% filter (supp == "VC" & dose == "0.5")
t.test(doseOJ0.5$len,doseVC0.5$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  doseOJ0.5$len and doseVC0.5$len
## 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 of x mean of y
##     13.23      7.98
```

```r
# Dose 1
doseOJ1 <- ToothGrowth %>% filter (supp == "OJ" & dose == "0.5")
doseVC1 <- ToothGrowth %>% filter (supp == "VC" & dose == "0.5")
t.test(doseOJ1$len,doseVC1$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  doseOJ1$len and doseVC1$len
## 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 of x mean of y
##     13.23      7.98
```

```r
# Dose 2
doseOJ2 <- ToothGrowth %>% filter (supp == "OJ" & dose == "0.5")
doseVC2 <- ToothGrowth %>% filter (supp == "VC" & dose == "0.5")
t.test(doseOJ2$len,doseVC2$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  doseOJ2$len and doseVC2$len
## 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 of x mean of y
##     13.23      7.98
```

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

At 95% confidence assumption, we can have draw following conclusions:

1. Dose 0.5 of OJ results in longer tooth than dose 0.5 of VC"
2. Dose 1 of OJ results in longer tooth than dose 0.5 of VC"
3. However, Dose 2 of OJ and VC result in almost similar tooth lenght"