Statistical Inference Course Project - Part 2

Amit Kohli

February 12, 2016

Overview:

The reports investigates the ToothGrowth data in the R datasets package by loading the ToothGrowth data and performing some basic exploratory data analyses. The report provides a basic summary of the data and uses confidence intervals to compare tooth growth by supplement and dose.

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

Number of Rows

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

Number of Columns

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

Display Structure of the ToothGrowth Dataset

```
## '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 ...
```

Display Header of the dataset

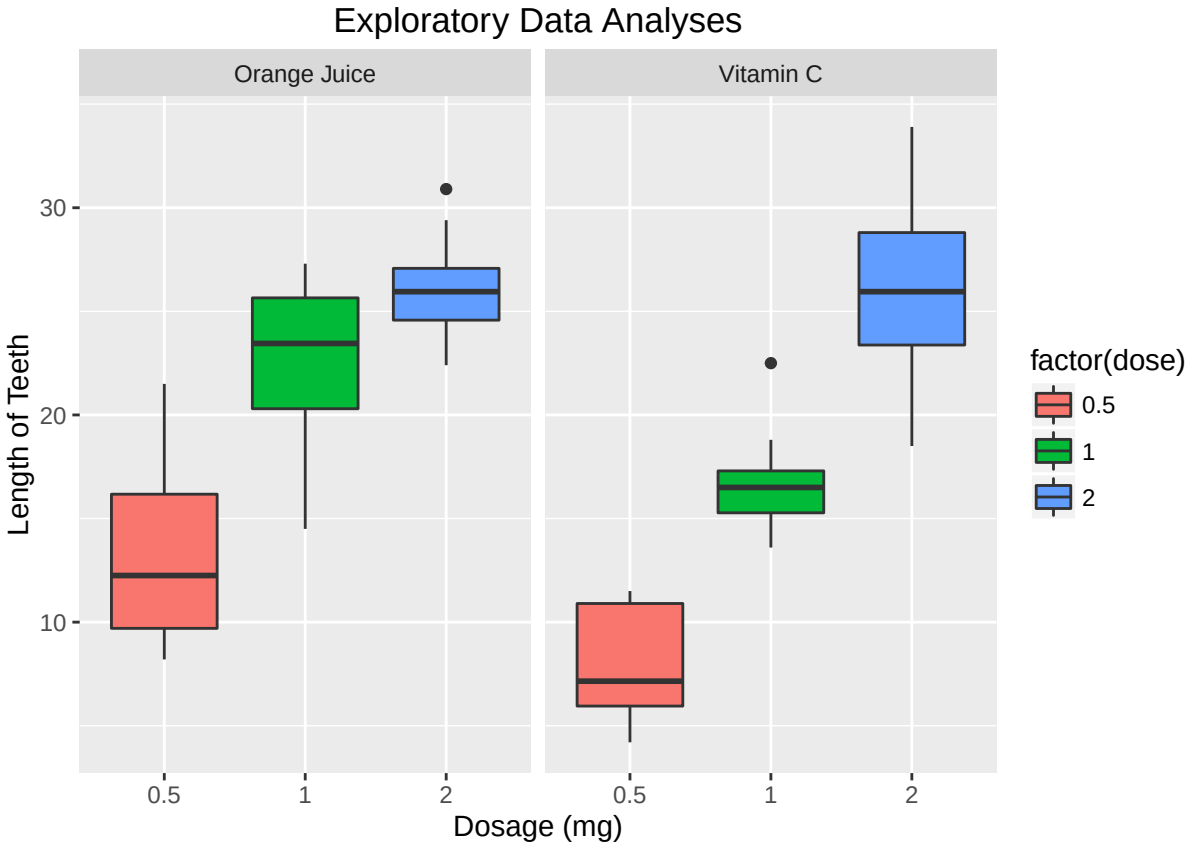
```
##
##
##   len  supp    dose
## -----  -----  -----
##   4.2  VC       0.5
##  11.5  VC       0.5
##   7.3  VC       0.5
##   5.8  VC       0.5
##   6.4  VC       0.5
##  10.0  VC       0.5
```

Disaplay Unique Dose values

```
## [1] 0.5 1.0 2.0
```

The ToothGrowth dataset has 3 columns and 60 rows. The response is the length of tooth growth in 60 guinea pigs. Each animal received one of three dose levels of Vitamin C and Orange Juice (0.5, 1, and 2 mg/day). There are two factors for supplement: "OJ" and "VC" – and there are three unique dosage values: "0.5", "1.0", and "2.0" based on the exploratory analysis above.

The plot below visualizes the ToothGrowth data – side-by-side comparison of both supplements.



2. Provide a basic summary of the data

```
##       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
```

Summary of data by Supplement

```
## ToothGrowth$supp: OJ
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    8.20   15.52   22.70   20.66   25.72   30.90
## ------------------------------------------------------------
## ToothGrowth$supp: VC
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4.20   11.20   16.50   16.96   23.10   33.90
```

Summary of data by Dosage

```
## ToothGrowth$dose: 0.5
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   4.200   7.225   9.850  10.600  12.250  21.500
## ------------------------------------------------------------
## ToothGrowth$dose: 1
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   13.60   16.25   19.25   19.74   23.38   27.30
## ------------------------------------------------------------
## ToothGrowth$dose: 2
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.50   23.52   25.95   26.10   27.83   33.90
```

3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

```
##                p.value   Conf.Low Conf.High
## Equal Var   0.06039337 -0.1670064  7.567006
## Unequal Var 0.06063451 -0.1710156  7.571016
```

The confidence results above allow for the rejection of the null hypothesis and a confirmation that there is a significant correlation between tooth length and supplement dose levels.

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

Based on the analysis above, there is definitive evidence that small dosages (0.5 and 1.0) of orange juice is more effective than Vitamin C in tooth growth. Although the results of higher dosage (2.0) is overall better than small dosage, the results are inconclusive between which supplement is better.

Assumptions: 1) The population are independent, the variances between populations are different and a random population was used; 2) The population was comprised of similar guinea pigs and measurement error was accounted for with significant digits; 3) For the populations to be independent, 60 guinea pigs were used so each combination of dose level and delivery method were not affected by the other methods.

Code for loading the ToothGrowth data and performing basic exploratory data analyses

```
data(ToothGrowth)

nrow(ToothGrowth)
ncol(ToothGrowth)
str(ToothGrowth)
knitr::kable(head(ToothGrowth))
unique(ToothGrowth$dose)

grid_names <- as_labeller(c('OJ' = "Orange Juice",'VC' = "Vitamin C"))
plot <- ggplot(ToothGrowth,
               aes(x=factor(dose),y=len,fill=factor(dose)))
plot + geom_boxplot(notch=F) + facet_grid(.~supp, labeller = grid_names) +
  scale_x_discrete("Dosage (mg)") +
  scale_y_continuous("Length of Teeth") +
  ggtitle("Exploratory Data Analyses")
```

Code for providing a basic summary of the data

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
summary(ToothGrowth)

 by(ToothGrowth$len, INDICES = ToothGrowth$supp, summary)
 by(ToothGrowth$len, INDICES = ToothGrowth$dose, summary)
```

Code for # using confidence intervals to compare tooth growth by supplement and dose

```
supplement1 <- t.test(len~supp, paired=F, var.equal=T, data=ToothGrowth)
supplement2 <- t.test(len~supp, paired=F, var.equal=F, data=ToothGrowth)
confResult <- data.frame("p-value"=c(supplement1$p.value, supplement2$p.value),
                         "Conf-Low"=c(supplement1$conf[1],supplement2$conf[1]),
                         "Conf-High"=c(supplement1$conf[2],supplement2$conf[2]),
                         row.names=c("Equal Var","Unequal Var"))
confResult
```