Part2:Part 2: Basic_Inferential_Data_ Analysis

Aleksander Petrovskii

June 11, 2017

Overview

The purpose of the this data analysis is to analyze the ToothGrowth data set by comparing the guinea tooth growth by supplement and dose. First, I will do exploratory data analysis on the data set. Then I will do the comparison with confidence intervals in order to make conclusions about the tooth growth.

Exploring the Data (Question 1 and 2)

Load data

```
library(datasets)
data(ToothGrowth)
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 ...
```

Look at 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
```

Provide basic summary of dataset

```
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
```

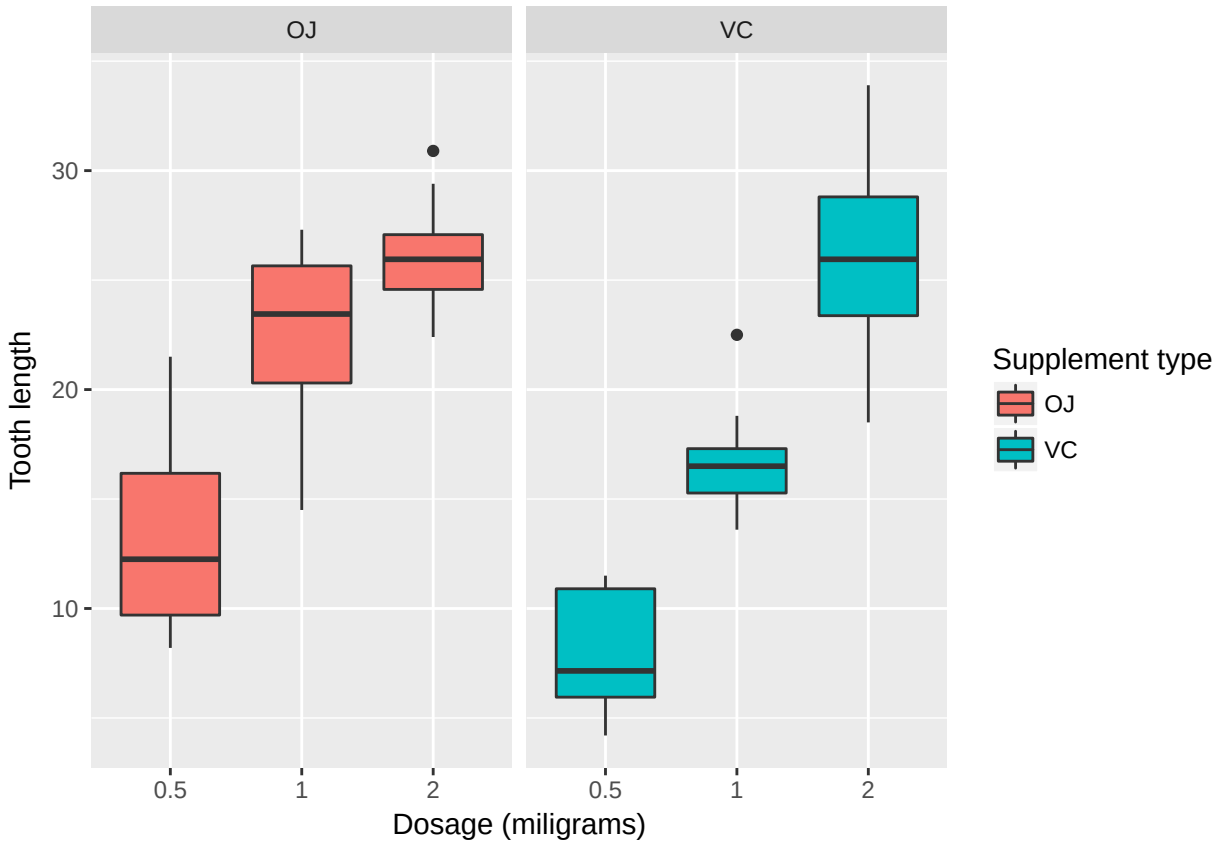
Combining the dosage and the delivery method to analyse the statistical data

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

```
## : OJ
## : 0.5
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    8.20    9.70   12.25   13.23   16.18   21.50
## ------------------------------------------------------------
## : VC
## : 0.5
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4.20    5.95    7.15    7.98   10.90   11.50
## ------------------------------------------------------------
## : OJ
## : 1
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   14.50   20.30   23.45   22.70   25.65   27.30
## ------------------------------------------------------------
## : VC
## : 1
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   13.60   15.27   16.50   16.77   17.30   22.50
## ------------------------------------------------------------
## : OJ
## : 2
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   22.40   24.58   25.95   26.06   27.08   30.90
## ------------------------------------------------------------
## : VC
## : 2
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.50   23.38   25.95   26.14   28.80   33.90
```

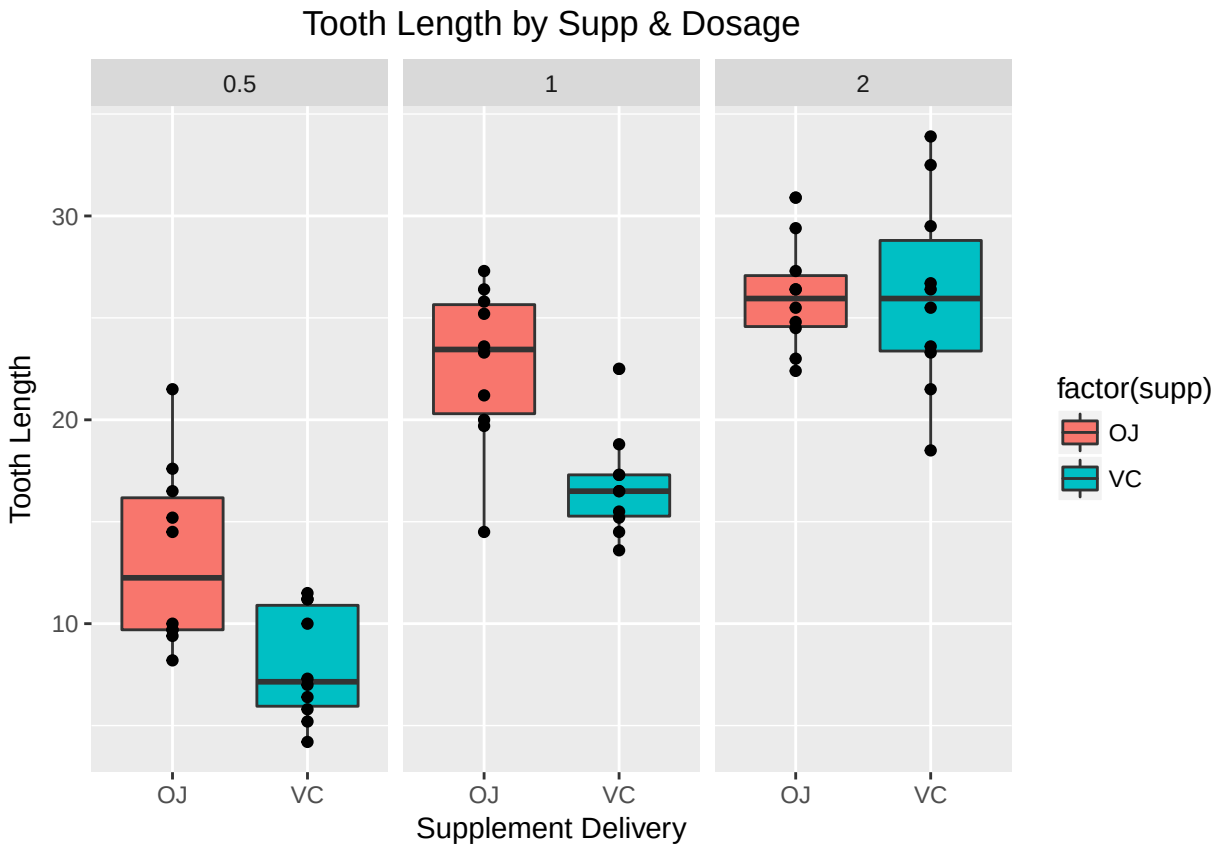
Look at possible relation between tooth length and delivery methods

```
ggplot(data=ToothGrowth, aes(x=as.factor(dose), y=len, fill=supp)) +
  geom_boxplot() +
  facet_grid(. ~ supp) +
  xlab("Dosage (miligrams)") +
  ylab("Tooth length") +
  guides(fill=guide_legend(title="Supplement type"))
```



Look at possible relation between tooth length vs the delivery methods by dose amount

```
ggplot(ToothGrowth, aes(x=supp, y=len)) +
        geom_boxplot(aes(fill=factor(supp)))+
        geom_point() +
        facet_grid(.~dose) +
        xlab("Supplement Delivery") +
        ylab("Tooth Length") +
        theme(plot.title = element_text(hjust = 0.5)) +
        ggtitle("Tooth Length by Supp & Dosage")
```



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

First, let's analyze Dosage as a factor

```
    dosage_1 <- subset(ToothGrowth, dose %in% c(0.5, 1.0))
    dosage_2 <- subset(ToothGrowth, dose %in% c(1.0, 2.0))
    dosage_3 <- subset(ToothGrowth, dose %in% c(0.5, 2.0))
```

For dosage from 0.5 to 1.0 (milligrams)

```
t.test(len ~ dose, paired = F, var.equal = F, data = dosage_1)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
```

```
##            10.605            19.735
```

For dosage from 1.0 to 2.0 (milligrams)

```
t.test(len ~ dose, paired = F, var.equal = F, data = dosage_2)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

For dosage from 0.5 to 2.0 (milligrams)

```
t.test(len ~ dose, paired = F, var.equal = F, data = dosage_3)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

Second, let's analyze Supplement as a factor

```
t.test(len ~ supp, paired = F, var.equal = F, data = ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

Conclusions and assumptions (Question 3)

Assumptions

1. For Dosage as a factor the confidence intervals ([-11.98, -6.276] for doses 0.5 and 1.0, [-18.16, -12.83], and [-8.996, -3.734] for doses 1.0 and 2.0) and for doses 0.5 and 2.0, allow for the rejection of the null hypothesis and a confirmation that there is a significant correlation between tooth length and dose levels.
2. For Supplement as a factor the confidence interval of [-0.171, 7.571] does not allow us to reject the null hypothesis (that there is no correlation between delivery method and tooth length).

Conclusions

1. Increasing the dose level leads to increased tooth growth
2. Supplement type has no effect on tooth growth.