Part 2 Basic Inferential Data Analysis

Prasanti

December 13, 2018

Overview:

The purpose of the this data analysis is to analyse the ToothGrowth data set by comparing the guinea tooth growth by supplement and dose.

1. Load the ToothGrowth data and perform exploratory data analysis

```
library(datasets)
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 ...
```

```
unique(ToothGrowth$dose)
```

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

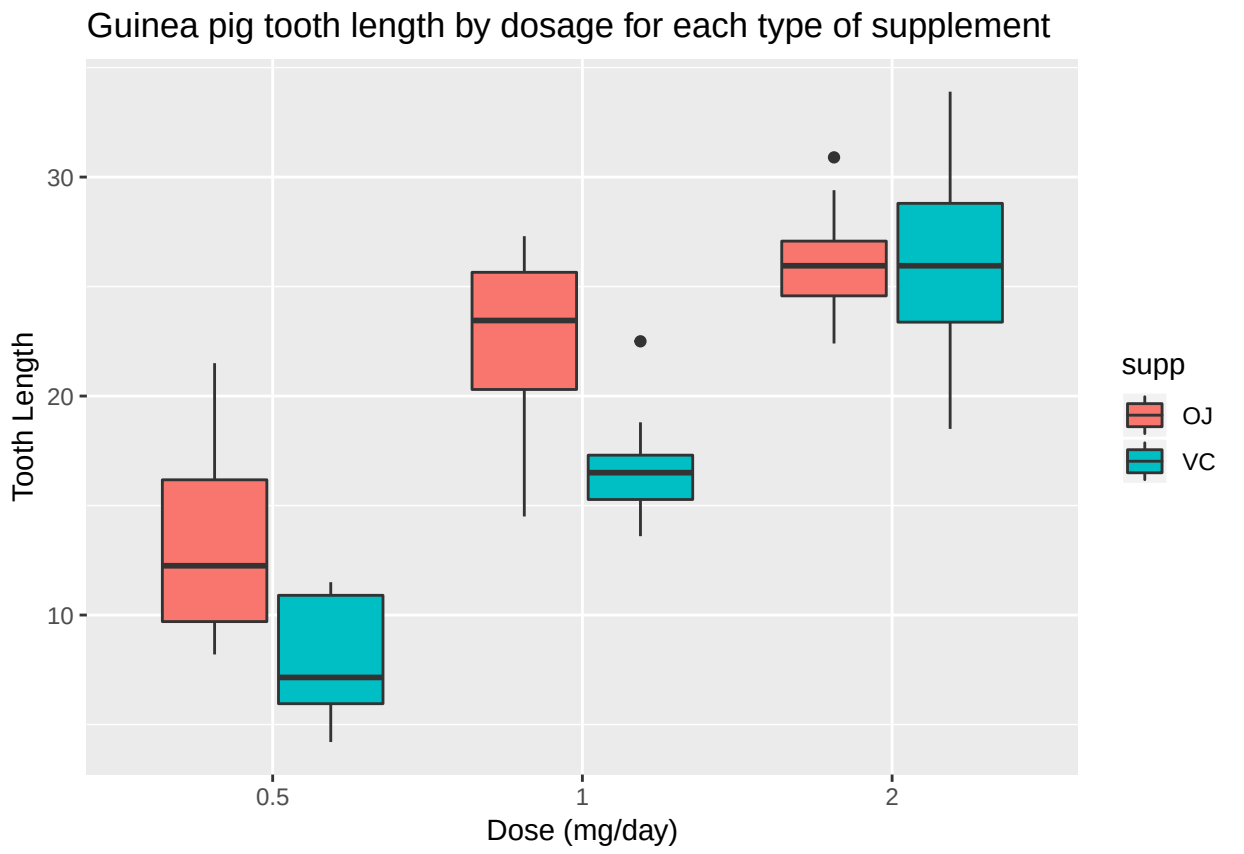
```
summary(ToothGrowth)
```

```
##       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
```

About the dataset:

The dataset has 60 observations, 3 variables. There are 2 types of supplements - OJ and VC and 3 levels of dose - 0.5, 1.0 and 2.0. The length ranges from 4.2 to 33.9 with Median length as 19.25. The mean is (18.81) slightly less than the median

```
library(ggplot2)
m <- ggplot(ToothGrowth , aes(x = factor(dose), y = len, fill = supp)) +
  geom_boxplot()+
  labs(title="Guinea pig tooth length by dosage for each type of supplement",
    x="Dose (mg/day)",
    y="Tooth Length")
m
```



2. Basic summary of the data

   . increasing the dosage increases the tooth growth.

   . OJ is more effective than VC for tooth growth when the dosage is .5 to 1.0 milligrams per day. But both types of supplements are equally effective when the dosage is 2.0 milligrams per day.

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

Hypothesis #1

Null Hypothesis : For the dosage of 0.5 mg/day, the two supplements deliver the same tooth growth. alternate hypothesis : OJ group delivers more tooth growth than VC.

```
data = subset(ToothGrowth, dose == 0.5)
g_oj = data$len[data$supp == "OJ"]
g_vc = data$len[data$supp == "VC"]
lower <- t.test(g_oj, g_vc, paired = F, var.equal = F)
lower$conf.int
```

```
## [1] 1.719057 8.780943
## attr(,"conf.level")
## [1] 0.95
```

```
lower$p.value
```

```
## [1] 0.006358607
```

The p-value is below the 0.05. So the null hypothesis can be rejected. The 95% confidence interval is entirely above zero when considered the OJ group first and then the VC group. It suggests that for 0.5 mg/day dosage, the OJ supplement delivers more tooth growth than VC. So the alternative hypothesis is accepted.

Hypothesis #2

Null Hypothesis : For the dosage of 1 mg/day, the two supplements deliver the same tooth growth. alternate hypothesis :OJ group delivers more tooth growth than VC.

```
data = subset(ToothGrowth, dose == 1)
g_oj = data$len[data$supp == "OJ"]
g_vc = data$len[data$supp == "VC"]
middle <- t.test(g_oj, g_vc, paired = F, var.equal = F)
middle$conf.int
```

```
## [1] 2.802148 9.057852
## attr(,"conf.level")
## [1] 0.95
```

```
middle$p.value
```

```
## [1] 0.001038376
```

The p-value is below the 0.05. So the null hypothesis can be rejected. The 95% confidence interval is entirely above zero when considered the OJ group first and then the VC group. It suggests that for 1 mg/day dosage, the OJ supplement delivers more tooth growth than VC. So the alternative hypothesis is accepted.

Hypothesis #3

Null Hypothesis : For the dosage of 2 mg/day, the two supplements deliver the same tooth growth.

alternate hypothesis : OJ group delivers more tooth growth than VC.

```
data = subset(ToothGrowth, dose == 2)
g_oj = data$len[data$supp == "OJ"]
g_vc = data$len[data$supp == "VC"]
upper <- t.test(g_oj, g_vc, paired = F, var.equal = F)
upper$conf.int
```

```
## [1] -3.79807  3.63807
## attr(,"conf.level")
## [1] 0.95
```

```
upper$p.value
```

```
## [1] 0.9638516
```

The p-value is larger than the 0.05 threshold. So we fail to reject the null hypothesis.The confidence interval includes 0 which suggest no difference between two groups. So the null hypothesis is accepted.

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

conclusions : OJ is more effective than VC for tooth growth when the dosage is .5 to 1.0 milligrams per day. But both types of supplements are equally effective when the dosage is 2.0 milligrams per day. So for the entire data set we cannot conclude if OJ is more effective than VC.

Assumptions: tooth lengths data follows normal distribution variances are not equal No other factors are affecting tooth length