Statistical Inferences Course Work - Part 2: Basic Inferential Data Analysis

Ts. Nurul Haszeli Ahmad

11/25/2020

Overview

This is a simulation exercise done to comply with the requirement to complete the Statistical Inferences course. In this simulation exercise, we will investigate and analyse the ToothGrowth data set by comparing the guinea tooth growth by supplement and dose. We will perform exploratory data analysis on the data set followed with the comparison using defined confidence intervals in order to make conclusions about the tooth growth.

Null Hypothesis
The Null Hypothesis is that **increasing** the Dose does not have any impact on the tooth growth.

Assumptions
For this analysis, we have few assumptions
1. Normal distribution of the tooth lengths.
2. No other unmeasured factors are affecting tooth length.
3. Data provided is independently distributed.
4. Data follows T distribution as the observations are limited.
5. Data is derived from samples representative of the population.

Exporation

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

# Explore the data set
# Review Data Structure
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 ...
```

```
# Review some of the actual 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
```

```r
# Review Data Statistics
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
```

```r
# Review unique values of dose control variable
unique(ToothGrowth$dose)
```

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

Based on the exploration, we can see that there are 3 unique doses given with minimum and maximum tooth length are 4.20 and 33.90. We will now perform the data analysis on the data set.

Data Analysis

We begin analysis by plotting the data into a graph.
What we know so far:
1. The 3 level of dose
2. 2 types of supplement - OJ and VC
Based on this two information, we will now plot the data to understand and analyze. For this, we will use boxplot.

```r
library(ggplot2)
t = ToothGrowth
levels(t$supp) <- c("OJ", "VC")
ggplot(t, aes(x=factor(dose), y=len)) +
  facet_grid(.~supp) +
  geom_boxplot(aes(fill = supp), show.legend = TRUE) +
  labs(title="The tooth length by dosage for each type of supplement",
    x="Dose (mg/day)",
    y="Tooth Length")
```

The tooth length by dosage for each type of supplement
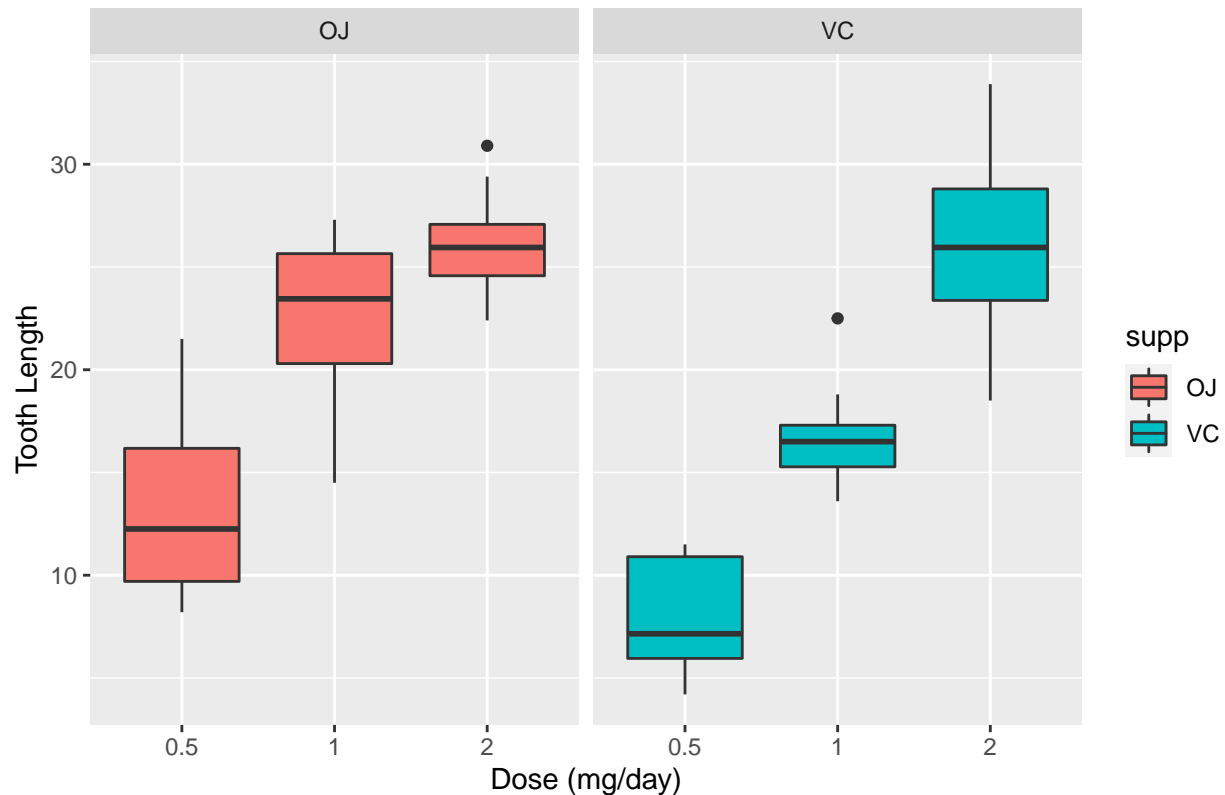
Basic summary of the data. The box plots seem to show, increasing the dosage increases the tooth growth. Between the two supplement, the **OJ seem to be have more effective affect than VC** for tooth growth when the dosage is at **0.5 to 1.0 milligrams per day**.

Both types of supplements are equally as effective when the dosage is 2.0 milligrams per day. From this initial analysis, we can pre-conclude that our hypothesis is correct.

Next, we will apply confidence interval and hypothesis evaluation to compare tooth growth by supplement and dose.

Hypothesis Testing Hypothesis #1: OJ & VC deliver the same tooth growth across the data set.

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

```
##
##  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
```

Result:
* P-Value, 0.0606 is greater than 0.05 (confidence interval of 95%)
* Confidence interval, (-0.171, 7.571) for the difference of the means of each group spans 0
* *Conclusion* - Null hypothesis is Failed to Reject, hence Hypothesis #1 is Rejected.

Hypothesis #2: For the dosage of 0.5 mg/day, the two supplements deliver the same tooth growth.

```
G = t.test(len ~ supp, data = subset(t, dose == 0.5));
print(G)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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 in group OJ mean in group VC
##            13.23             7.98
```

Result:
* P-Value, 0.006359 is below than 0.05 (confidence interval of 95%)
* Confidence interval, (1.719057, 8.780943) does not include 0
* *Conclusion* - Null hypothesis is can be Reject, hence Hypothesis #2 is Accepted - The alternative hypothesis that 0.5 mg/day dosage of OJ causing more tooth growth than VC is accepted

Hypothesis #3: For the dosage of 1.0 mg/day, the two supplements deliver the same tooth growth

```
G = t.test(len ~ supp, data = subset(t, dose == 1.0));
print(G)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

Result:
* P-Value, 0.001038 is below than 0.05 (confidence interval of 95%)
* Confidence interval, (2.802148, 9.057852) does not include 0

* *Conclusion* - Null hypothesis is can be Reject, hence Hypothesis #3 is Accepted - The alternative hypothesis that 1.0 mg/day dosage of OJ causing more tooth growth than VC is accepted

Hypothesis #4: For the dosage of 2.0 mg/day, the two supplements deliver the same tooth growth

```
G = t.test(len ~ supp, data = subset(t, dose == 2.0));
print(G)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

Result:
* P-Value, 0.9639 is more than 0.05 (confidence interval of 95%)
* Confidence interval, (-3.79807, 3.63807) for the difference of the means of each group spans 0
* *Conclusion* - Null hypothesis is Failed to Reject, hence Hypothesis #4 is Rejected.

Conclusions & Assumptions

Conclusions   Based on the above evaluation of the four hypothesis, following are the conclusions:
1. For impact of control variable "supp" only, there is no significant difference on the tooth growth for both supplement.
2. For impact of control variable "dose" only, higher the dose, higher impact to the tooth growth.
3. For combined impact of control variables (dose and supp), there is significant difference on the tooth growth between the supplement for dose of 0.5 mg/day and 1.0 mg/day. However, for does of 2.0 mg/day, there are no significant differences on the tooth growth for the two supplement.

Assumptions   Therefore, we can assume that
1. OJ is a better supplement compared to VC.
2. The right amount of dose will help in the tooth growth.