Statistical Inference Course Project (Part 2)

Tong Tat

December 23, 2017

Rpub Link: Click Here

Part 2

For the second part of the project, we are going to analyze the ToothGrowth data in the R datasets package.

2.1. Loading Dataset

```
# Check for missing dependencies and load necessary R packages
if(!require(ggplot2)){install.packages('ggplot2')}; library(ggplot2)

# Load dataset
data("ToothGrowth")

# Quick explanation on ToothGrowth dataset
?ToothGrowth

# Check summary
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
```

```
# Display different summary
str(ToothGrowth)
```

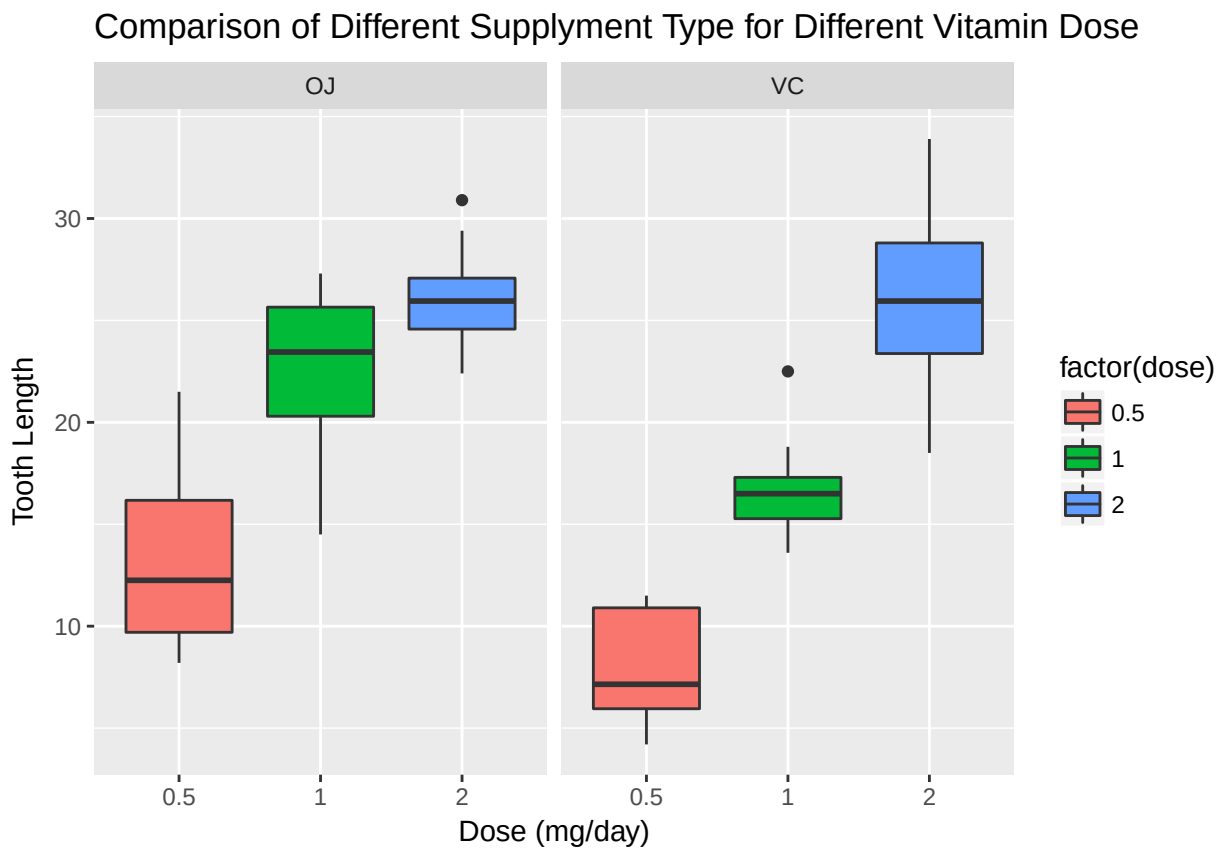
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

2.2. Data Visualization

Now since "supp" is the Supplyment type (namely **VC** represents **Ascorbic Acid** and **OJ** represents **Orange Juice**), so it is easier to look at Tooth Length vs Dose breakdown by Supplyment type.

```
gg2 <- ggplot(ToothGrowth, aes(x=factor(dose), y=len, fill=factor(dose))) +
  geom_boxplot() + facet_grid(.~supp) +
```

```
  labs(title="Comparison of Different Supplyment Type for Different Vitamin Dose", y="Tooth Length", x=
gg2
```



2.3. Calculating Confidence Interval

```
# Calculate Confidence Interval for Dose=0.5 mgrams/day.
dose.05 <- subset(ToothGrowth, dose==0.5)
test.05 <- t.test(len ~ supp, paired=F, var.equal=F, data=dose.05)


# Calculate Confidence Interval for Dose=0.5 mgrams/day.
dose.1 <- subset(ToothGrowth, dose==1)
test.1 <- t.test(len ~ supp, paired=F, var.equal=F, data=dose.1)


# Calculate Confidence Interval for Dose=0.5 mgrams/day.
dose.2 <- subset(ToothGrowth, dose==2)
test.2 <- t.test(len ~ supp, paired=F, var.equal=F, data=dose.2)


# Summarizing p-value and Confidence Interval in a table
table1 <- data.frame(
  "p.value"=c(test.05$p.value, test.1$p.value, test.2$p.value),
```

```
  "Conf.Low"=c(test.05$conf.int[1], test.1$conf.int[1], test.2$conf.int[1]),
  "Conf.High"=c(test.05$conf.int[2], test.1$conf.int[2], test.2$conf.int[2]),
  row.names=c("Dose.05", "Dose.1", "Dose.2")
)
table1
```

```
##             p.value  Conf.Low Conf.High
## Dose.05 0.006358607  1.719057  8.780943
## Dose.1  0.001038376  2.802148  9.057852
## Dose.2  0.963851589 -3.798070  3.638070
```

2.4. Assumptions and Conclusion

For this ToothGrowth dataset, I assume the data is collected from random Guinea pigs given either Orange Juice or Asorbic Acid. Hence, when doing the t-test, we have to use paired=FALSE. I also assume the experiment is conducted on the same species of Guinea pigs, hence it is assumed to be no error term.

Based on the p-value, for Dose=0.05mg/day & 1mg/day, since the p-value is less than 5%, we reject the null hypothesis and conclude there is significant difference between Orange Juice and Ascorbic Acid at low dose (less than or equal 1mg/day). For Dose=2mg/day, since the p-value is greater than 5%, we failed to reject the null hypothesis and conclude there is no significant difference between Orange Juice and Ascorbic Acid at high dose (2mg/day).