

# Statistical Inference Course Project Part 2

*Yan*

*5/8/2018*

## Part 2: Basic Inferential Data Analysis Instructionsless

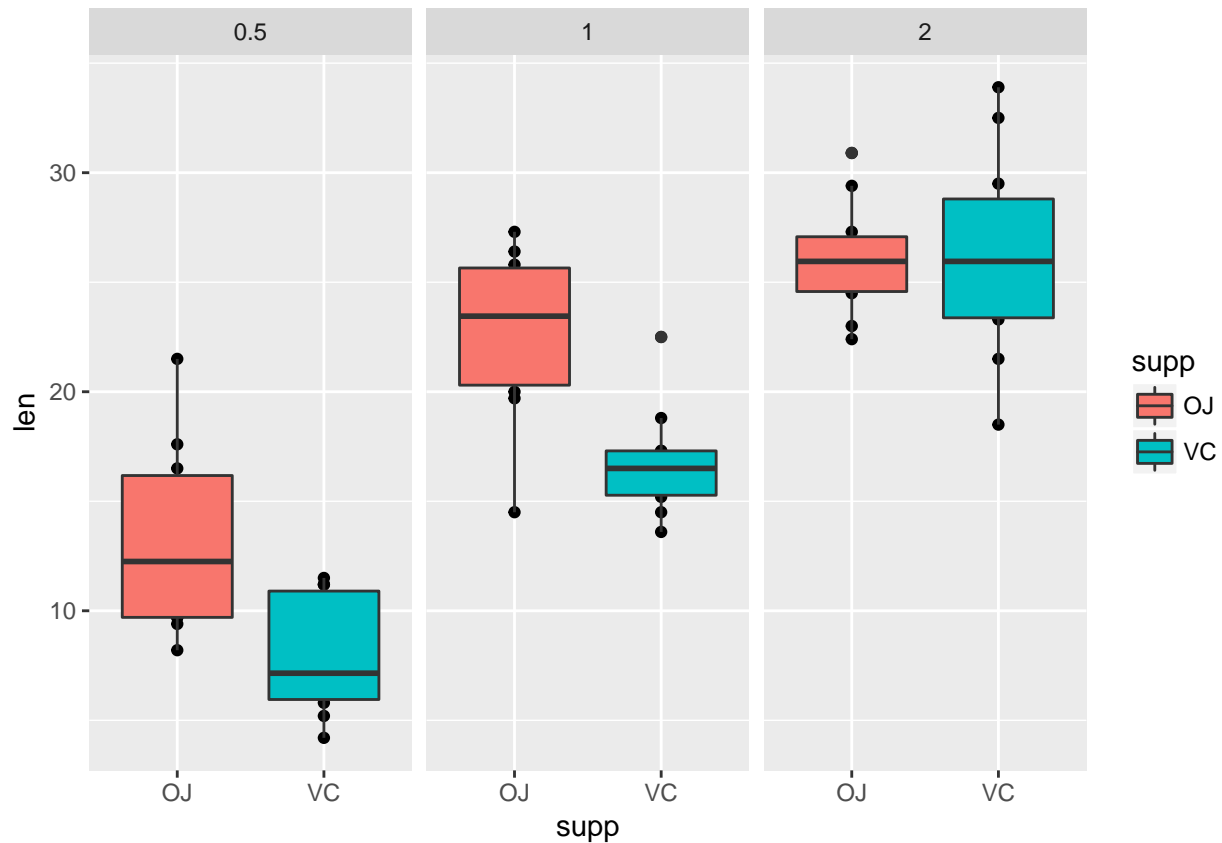
Now in the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.

**Load the ToothGrowth data and perform some basic exploratory data analyses**

```
library(datasets)
library(ggplot2)
data(ToothGrowth)
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
```

```
tg <- ToothGrowth
tg$dose <- as.factor(tg$dose)
qplot(x=supp,y=len,data = tg,facets = .~dose) + geom_boxplot(aes(fill = supp))
```



Provide a basic summary of the data.

```
dim(tg)
```

```
## [1] 60 3
```

```
summary(tg)
```

```
##      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
```

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

Note: (Only use the techniques from class, even if there's other approaches worth considering)

split the dataset according to dose and supp

```
OJdose1 <- tg[tg$supp == "OJ" & tg$dose == 0.5,]
OJdose2 <- tg[tg$supp == "OJ" & tg$dose == 1,]
OJdose3 <- tg[tg$supp == "OJ" & tg$dose == 2,]
VCdose1 <- tg[tg$supp == "VC" & tg$dose == 0.5,]
```

```
VCdose2 <- tg[tg$supp == "VC" & tg$dose == 1,]
VCdose3 <- tg[tg$supp == "VC" & tg$dose == 2,]
```

Check if tooth growth is affected by dose

**Get confidence intervals for different supp regardless different dose**

```
t.test(len ~ supp, data = tg, paired=FALSE, var.equal = FALSE)$conf
```

```
## [1] -0.1710156  7.5710156
## attr("conf.level")
## [1] 0.95
```

So when doesn't consider the dosage, doesn't show significant difference between two supplements.

**Get confidence intervals for low/median/large dose with different supp**

```
t.test(OJdose1$len, VCdose1$len, paired = FALSE, var.equal = FALSE)$conf
```

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

```
t.test(OJdose2$len, VCdose2$len, paired = FALSE, var.equal = FALSE)$conf
```

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

```
t.test(OJdose3$len, VCdose3$len, paired = FALSE, var.equal = FALSE)$conf
```

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

From the 95% confidence intervals, 0 is not included when the dosage is small or median. So there's significant difference between two supplement with OJ leads to faster tooth growth than VC.

**Get confidence intervals for same supp with different dosage**

```
t.test(OJdose1$len, OJdose2$len, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1] -13.415634 -5.524366
## attr("conf.level")
## [1] 0.95
```

```
t.test(OJdose1$len, OJdose3$len, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1] -16.335241 -9.324759
## attr("conf.level")
## [1] 0.95
```

```
t.test(OJdose2$len, OJdose3$len, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1] -6.5314425 -0.1885575
## attr("conf.level")
## [1] 0.95
```

```
t.test(VCdose1$len, VCdose2$len, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1] -11.265712 -6.314288  
## attr(,"conf.level")  
## [1] 0.95
```

```
t.test(VCdose1$len, VCdose3$len, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1] -21.90151 -14.41849  
## attr(,"conf.level")  
## [1] 0.95
```

```
t.test(VCdose2$len, VCdose3$len, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1] -13.054267 -5.685733  
## attr(,"conf.level")  
## [1] 0.95
```

From the 95% confidence intervals, 0 is not included in all cases. So there's significant difference between different dosage of same supplement, the tooth growth becomes faster with increasing dosage.

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

Assumptions:

- The variance with different supplement and different dosage is not the same.
- The data between different groups are not paired
- The sample population is not skewed

Conclusions:

- With same supplement, OJ or VC, increasing the dosage leads to faster tooth growth
- With small or median dosage, OJ leads to faster tooth growth than VC
- With large dosage, no significant difference observed using OJ or VC

### Review criteria (self check)

- [x] Did you show where the distribution is centered at and compare it to the theoretical center of the distribution?
- [x] Did you show how variable it is and compare it to the theoretical variance of the distribution?
- [x] Did you perform an exploratory data analysis of at least a single plot or table highlighting basic features of the data?
- [x] Did the student perform some relevant confidence intervals and/or tests?
- [x] Were the results of the tests and/or intervals interpreted in the context of the problem correctly?
- [x] Did the student describe the assumptions needed for their conclusions?

Links used:

link 1, link 2, link 3, link 4, link 5, link 6, link 7, link 8