# ToothGrowth Hypothesis Testing

*Di Zien LOW, dizien@gmail.com*

*Wednesday, December 17, 2014*

## Introduction

This is Part 2 of the Statistical Inference Coursework. Key objectives of this report is to perform exploratory data analysis on ToothGrowth dataset using R, and perform some hypothesis test to compare tooth growth by supplements types and dosage. This report is structured into 3 sections:

Section 1. Data understanding

Section 2. Exploratory data analysis

Section 3. Hypothesis testing & conclusion

## Section 1. Data understanding

### 1a. Load ToothGrowth 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 ...
```

```
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
```

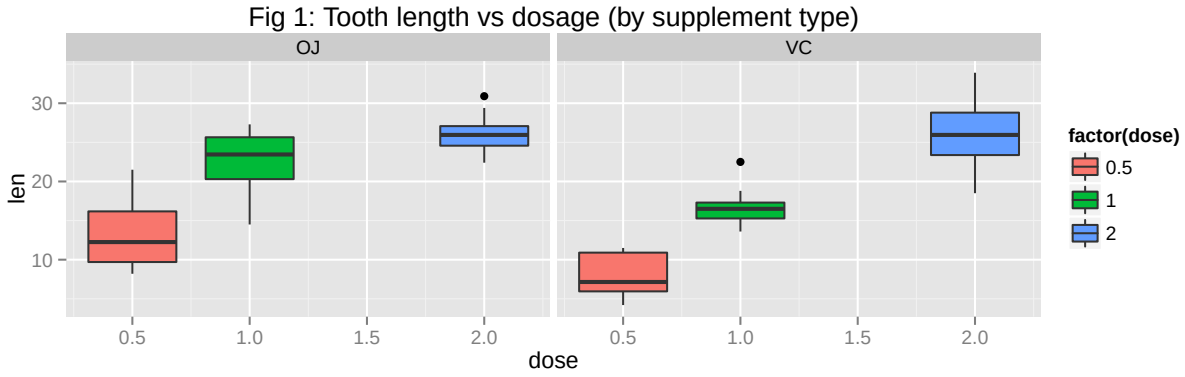
### 1b. Understand and summarize data

- The ToothGrowth dataset exploains the effect of vitamin C on tooth growth in guinea pigs
- It documents the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid)
- The dataset is in a dataframe has 60 observations and 3 variables. The 3 variables are: (i) **len** - Tooth length, (ii) **supp** - Supplement type (OJ or VC), (iii) **dose** - Dose in milligrams

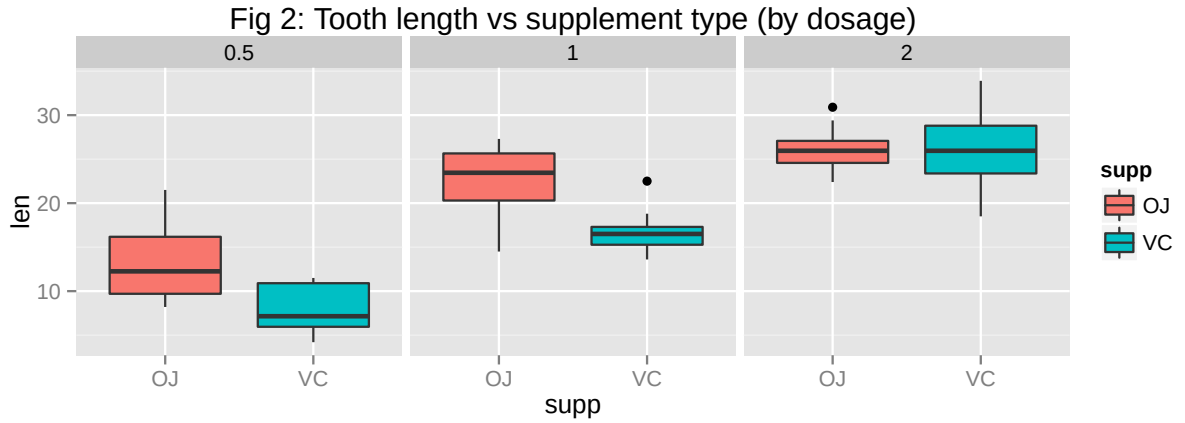
## Section 2. Exploratory Data Analysis

### 2a. Plot exploratory graph

```
library(ggplot2)
g <- ggplot(ToothGrowth, aes(x=dose, y=len, fill=factor(dose))) + geom_boxplot()
g + facet_grid(.~supp) + ggtitle("Fig 1: Tooth length vs dosage (by supplement type)")
```




```
library(ggplot2)
g <- ggplot(ToothGrowth, aes(x=supp, y=len, fill=supp)) + geom_boxplot()
g + facet_grid(.~dose) + ggtitle("Fig 2: Tooth length vs supplement type (by dosage)")
```




### 2b. Form initial hypothesis

Assuming that (i) $\mu$ represents the average length of tooth growth, (ii) all observed data points are independent and unpaired, (iii) at 95% confidence level, hypothesis to be examined are as follow:

- **Test 1**: Regardless of supplement, high dosage promotes greater tooth growth, $\mu_{dose=2.0} > \mu_{dose=0.5}$
- **Test 2**: For dosage = 0.5 mg, Orange Juice promotes greater tooth growth, $\mu_{dose=0.5,OJ} > \mu_{dose=0.5,VC}$
- **Test 3**: For dosage = 1.0 mg, Orange Juice promotes greater tooth growth, $\mu_{dose=1.0,OJ} > \mu_{dose=1.0,VC}$
- **Test 4**: For dosage = 2.0 mg, Orange Juice promotes greater tooth growth, $\mu_{dose=2.0,OJ} > \mu_{dose=2.0,VC}$

## Section 3. Hypothesis testing

### Hypothesis Test 1: Regardless of supplement, high dosage promotes greater tooth growth

#### 1A. Perform statistics calculations

$H_o$: $\mu_{dose=2.0}$ - $\mu_{dose=0.5} = 0$

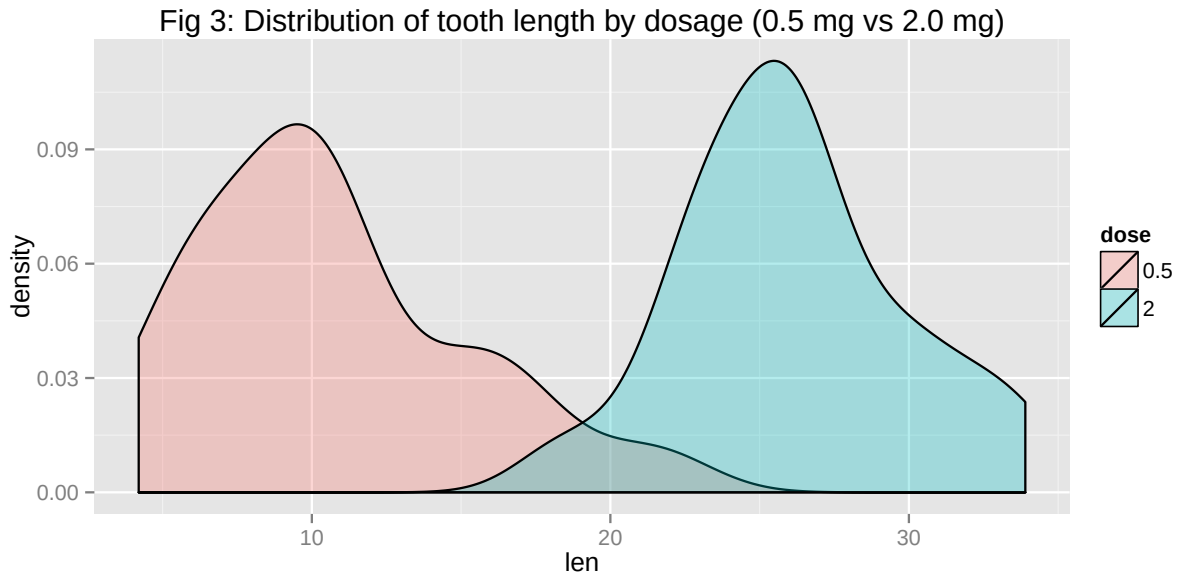
$H_1$: $\mu_{dose=2.0}$ - $\mu_{dose=0.5} > 0$

```
data1 <- subset(ToothGrowth, dose %in% 2.0)
data2 <- subset(ToothGrowth, dose %in% 0.5)
result <-t.test(data1$len-data2$len, paired = FALSE, conf.level=0.95)
c(statistic=round(result$statistic,4), p.value=round(result$p.value,4),
  CI.low=round(result$conf.int[1],4), CI.high=round(result$conf.int[2],4))
```

```
## statistic.t     p.value      CI.low     CI.high
##     11.2915      0.0000     12.6228     18.3672
```

#### 1B. Plot the two distributions for visual inspection

```
combo <- rbind(data1, data2)
combo$dose <- as.factor(combo$dose)
library(ggplot2)
g <- ggplot(combo, aes(x=len, fill=dose)) + geom_density(alpha=.3)
g + ggtitle("Fig 3: Distribution of tooth length by dosage (0.5 mg vs 2.0 mg)")
```




#### 1C.Findings and conclusion

- **Findings**: As p-value is less than than 0.05, we have enough evidence to **reject** $H_o$
- **Conclusion**: Regardless of supplement type, high dosage of Vitamin C **does** promote greater tooth growth in guinea pigs

**Hypothesis Test 2: For dosage = 0.5 mg, Orange Juice promotes greater tooth growth**

**2A. Perform statistics calculations**

$H_o$: $\mu_{dose=0.5,OJ}$ - $\mu_{dose=0.5,VC} = 0$

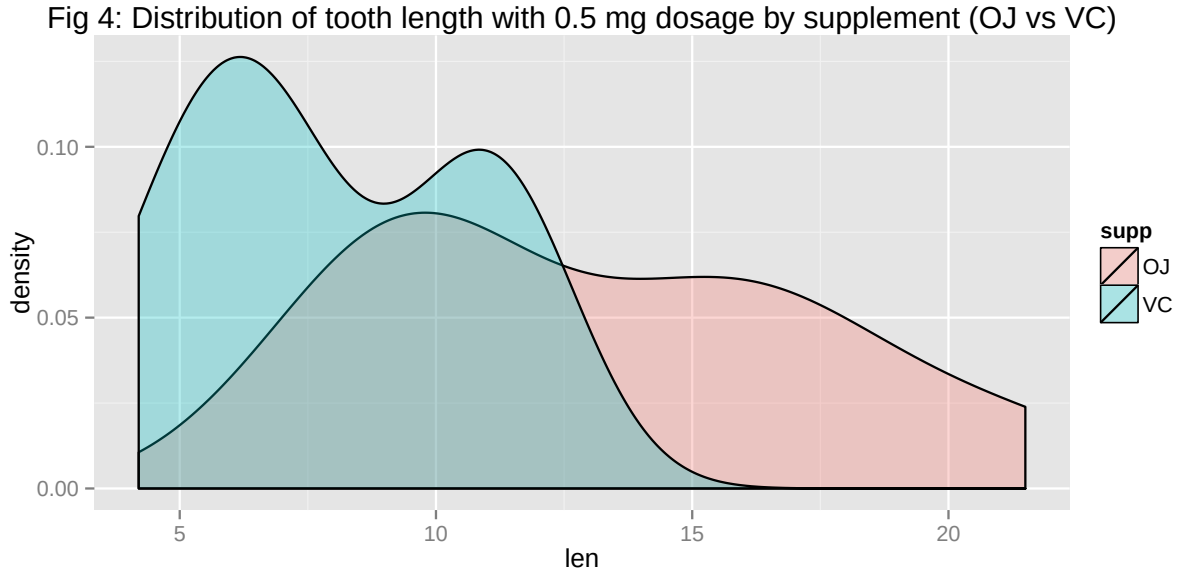
$H_1$: $\mu_{dose=0.5,OJ}$ - $\mu_{dose=0.5,VC} > 0$

```
data1 <- subset(ToothGrowth, dose %in% 0.5 & supp %in% "OJ")
data2 <- subset(ToothGrowth, dose %in% 0.5 & supp %in% "VC")
result <-t.test(data1$len-data2$len, paired = FALSE, conf.level=0.95)
c(statistic=round(result$statistic,4), p.value=round(result$p.value,4),
  CI.low=round(result$conf.int[1],4), CI.high=round(result$conf.int[2],4))
```

```
## statistic.t     p.value      CI.low     CI.high
##      2.9791      0.0155      1.2635      9.2365
```

**2B. Plot the two distributions for visual inspection**

```
combo <- rbind(data1, data2)
library(ggplot2)
g <- ggplot(combo, aes(x=len, fill=supp)) + geom_density(alpha=.3)
g + ggtitle("Fig 4: Distribution of tooth length with 0.5 mg dosage by supplement (OJ vs VC)")
```




**2C.Findings and conclusion**

- **Findings**: As p-value less than 0.05, we have enough evidence to **reject** $H_o$
- **Conclusion**: For Vitamin C dosage of 0.5mg, Orange Juice (OJ) supplement type **does** promote greater tooth growth in guinea pigs than ascorbic acid (VC)

**Hypothesis Test 3: For dosage = 1.0 mg, Orange Juice promotes greater tooth growth**

**3A. Perform statistics calculations**

$H_o$: $\mu_{dose=1.0,OJ}$ - $\mu_{dose=1.0,VC} = 0$

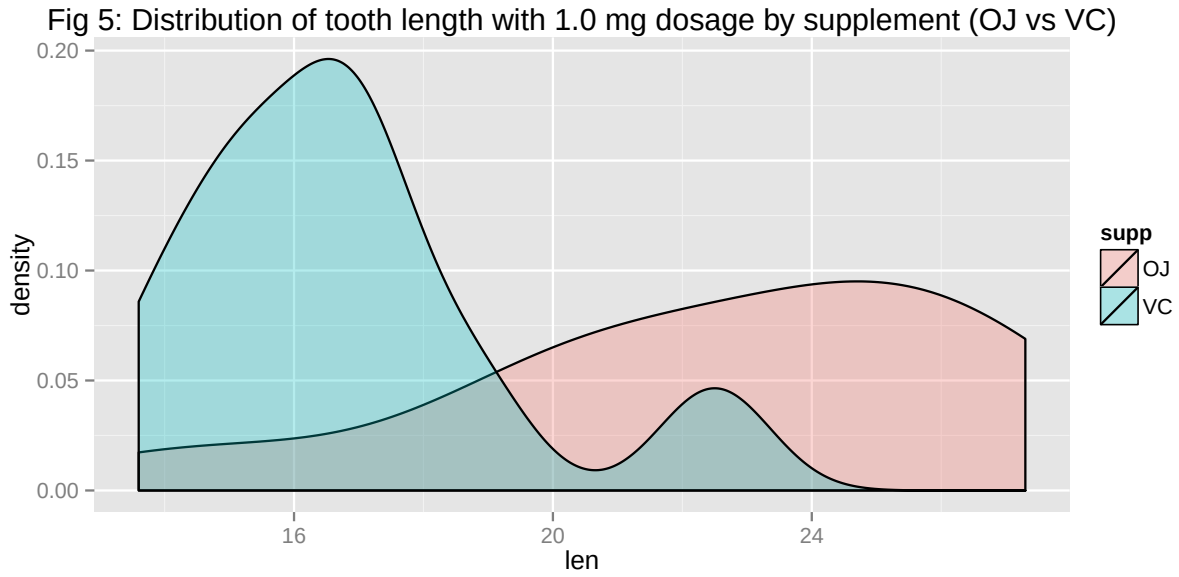
$H_1$: $\mu_{dose=1.0,OJ}$ - $\mu_{dose=1.0,VC} > 0$

```
data1 <- subset(ToothGrowth, dose %in% 1.0 & supp %in% "OJ")
data2 <- subset(ToothGrowth, dose %in% 1.0 & supp %in% "VC")
result <-t.test(data1$len-data2$len, paired = FALSE, conf.level=0.95)
c(statistic=round(result$statistic,4), p.value=round(result$p.value,4),
  CI.low=round(result$conf.int[1],4), CI.high=round(result$conf.int[2],4))
```

```
## statistic.t     p.value      CI.low     CI.high
##      3.3721      0.0082      1.9519      9.9081
```

**3B. Plot the two distributions for visual inspection**

```
combo <- rbind(data1, data2)
library(ggplot2)
g <- ggplot(combo, aes(x=len, fill=supp)) + geom_density(alpha=.3)
g + ggtitle("Fig 5: Distribution of tooth length with 1.0 mg dosage by supplement (OJ vs VC)")
```




**3C.Findings and conclusion**

- **Findings**: As p-value less than 0.05, we have enough evidence to **reject** $H_o$
- **Conclusion**: For Vitamin C dosage of 1.0mg, Orange Juice (OJ) supplement type **does** promote greater tooth growth in guinea pigs than ascorbic acid (VC)

**Hypothesis Test 4: For dosage = 2.0 mg, Orange Juice promotes greater tooth growth**

**4A. Perform statistics calculations**

$H_o$: $\mu_{dose=2.0,OJ}$ - $\mu_{dose=2.0,VC} = 0$

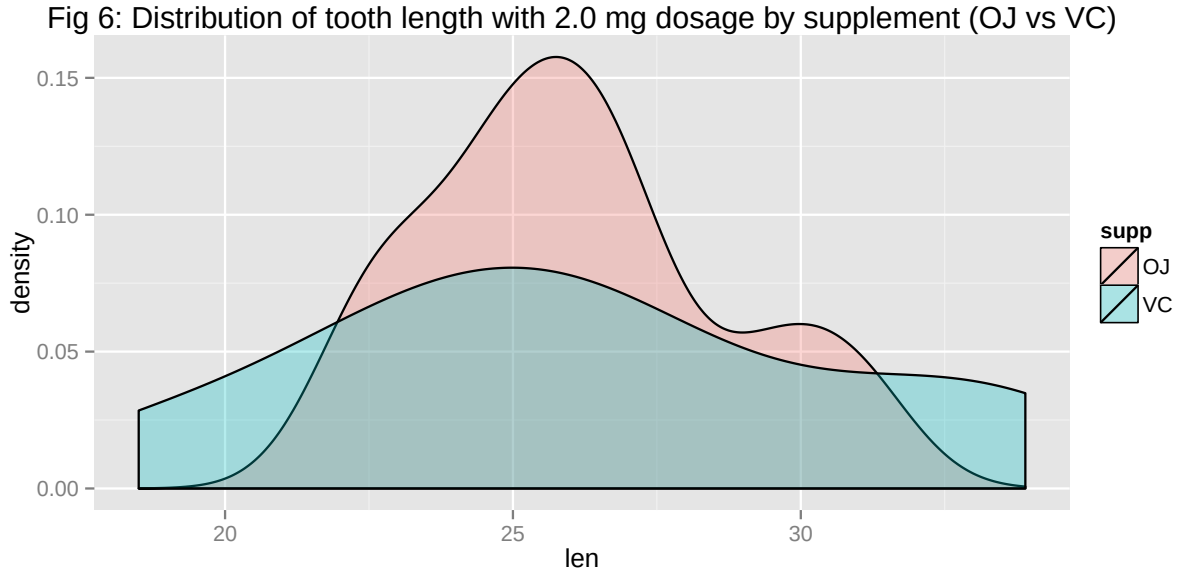
$H_1$: $\mu_{dose=2.0,OJ}$ - $\mu_{dose=2.0,VC} > 0$

```
data1 <- subset(ToothGrowth, dose %in% 2.0 & supp %in% "OJ")
data2 <- subset(ToothGrowth, dose %in% 2.0 & supp %in% "VC")
result <-t.test(data1$len-data2$len, paired = FALSE, conf.level=0.95)
c(statistic=round(result$statistic,4), p.value=round(result$p.value,4),
  CI.low=round(result$conf.int[1],4), CI.high=round(result$conf.int[2],4))
```

```
## statistic.t     p.value      CI.low     CI.high
##     -0.0426      0.9670     -4.3290      4.1690
```

**4B. Plot the two distributions for visual inspection**

```
combo <- rbind(data1, data2)
library(ggplot2)
g <- ggplot(combo, aes(x=len, fill=supp)) + geom_density(alpha=.3)
g + ggtitle("Fig 6: Distribution of tooth length with 2.0 mg dosage by supplement (OJ vs VC)")
```




**4C.Findings and conclusion**

- **Findings**: As p-value is greater than 0.05, we **do not have enough evidence** to reject $H_o$
- **Conclusion**: For Vitamin C dosage of 2.0mg, Orange Juice (OJ) supplement type **does not** promote greater tooth growth in guinea pigs than ascorbic acid (VC)