

# Basic Inferential

*NiaS*

*October 28, 2018*

## Exploratory Data

Load the data and required packages

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##   filter, lag
##
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
##
#inspect the structure of dataset
ToothGrowth %>% str()
## '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 ...
## $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
#summary
ToothGrowth %>% summary()
```

```
##      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
```

The dataset consist of 60 obs of 3 var: - len : tooth length, numeric - supp: supplement type; Vitamin(VC) and Orange Juice(OJ), factor -dose: in miligram, numeric

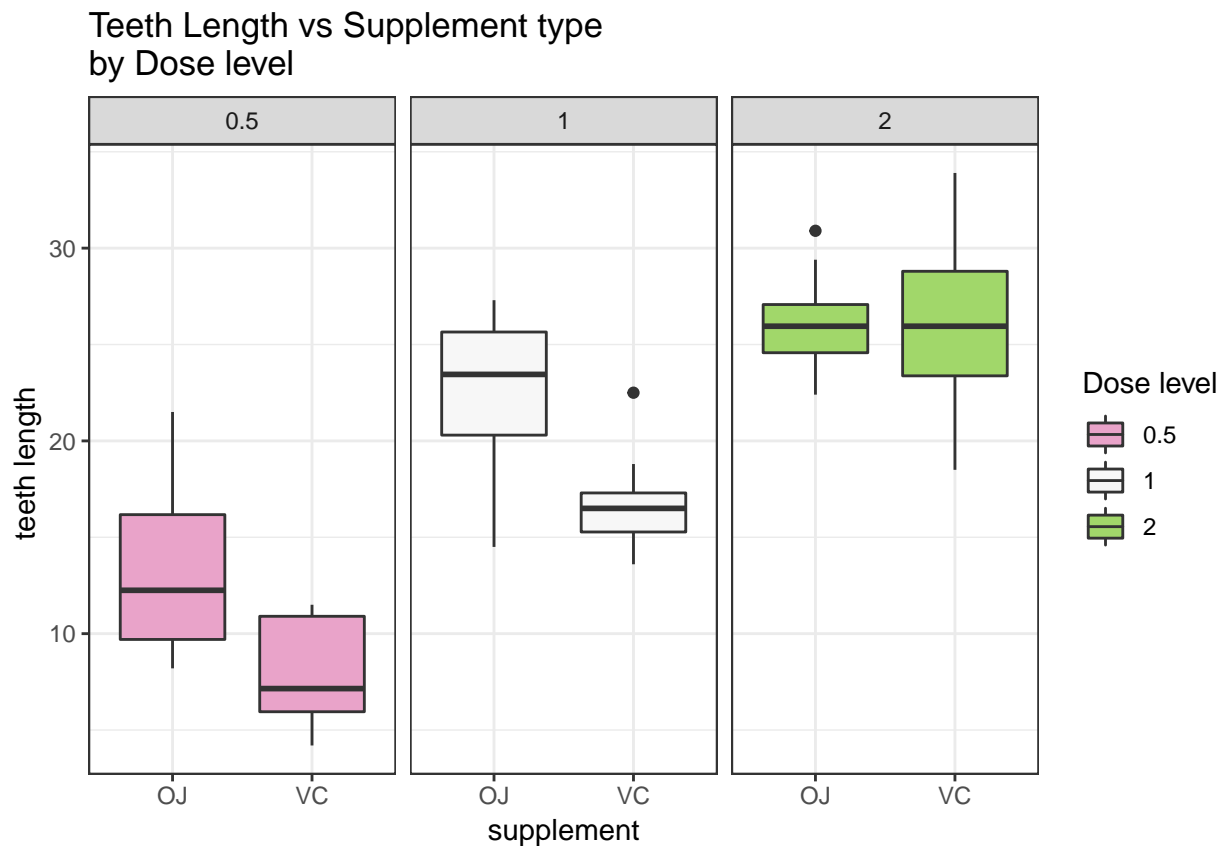
```
#find unique values in dose
table(ToothGrowth$dose)
```

```
##
## 0.5    1    2
## 20   20   20
```

```
#convert to factor var
ToothGrowth <- ToothGrowth %>% mutate(dose=as.factor(dose))
```

The goal is to find differences in tooth length across different groups of supplement types, dose levels and their respective combinations.

```
library(ggplot2)
#supplement and tooth length relationship
ToothGrowth %>% ggplot(aes(supp, len, fill=dose))+geom_boxplot()+facet_grid(. ~ dose) +
scale_fill_brewer(palette = "PiYG") +
theme_bw() +
ggtitle("Teeth Length vs Supplement type \nby Dose level") +
labs(x="supplement", y= "teeth length ") +
guides(fill=guide_legend(title="Dose level"))
```



Grouping by dose level reveals even greater differences in tooth length and their respective means. The differences between the dose level groups are so big that not even the interquartile range boundaries overlap:

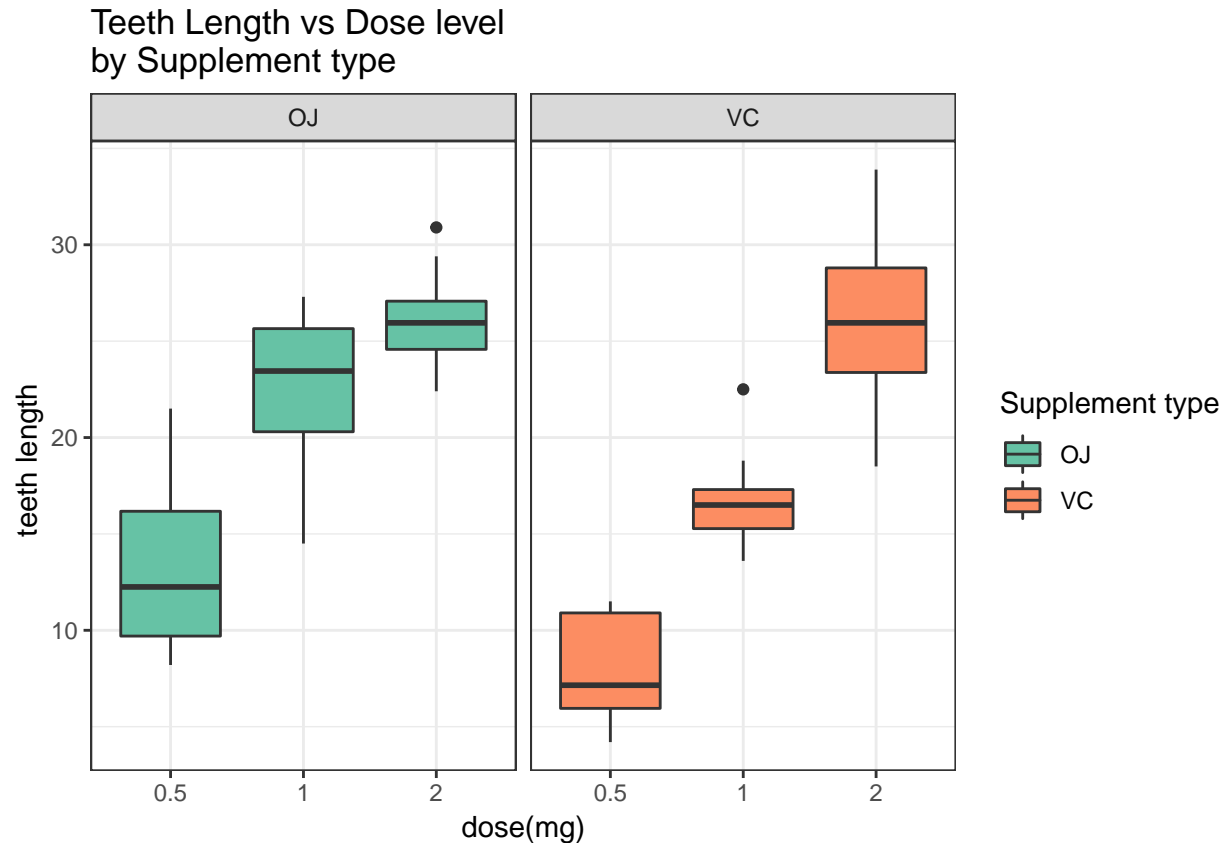
```
ToothGrowth %>% group_by(dose) %>% summarize(
  q25 = quantile(len, 0.25),
  q75 = quantile(len, 0.75)
)%>% as.data.frame
```

```
##   dose   q25   q75
## 1  0.5  7.225 12.250
## 2   1 16.250 23.375
## 3   2 23.525 27.825
```

Both trends can be further observed when examining tooth length differences across both groups at the same time.

```
# dose and tooth length relationship
ToothGrowth %>% ggplot(aes(dose, len, fill=supp))+geom_boxplot()+facet_grid(. ~ supp) +
scale_fill_brewer(palette = "Set2") +
```

```
theme_bw() +
ggtitle("Teeth Length vs Dose level \nby Supplement type") +
labs(x="dose(mg)", y= "teeth length ") +
guides(fill=guide_legend(title="Supplement type"))
```



The left box (OJ) above show that OJ appears to be more effective at dose level 0.5 and 1, while on level 2 there are not exist any significant difference compared to level 0.5 and 1. While on right box(VC) shows that VC seems to be more effective on dose level 0.5 and 2.

## Hypothesis test

### Conditions

Since  $n \leq 30$ , we will use the t distribution for our hypothesis tests and when constructing confidence intervals.

Based on the given information, it is difficult to know the following: - Independence of the observations - Observation come from a nearly normal distribution

Therefore, we assume we assume that the guinea pigs were randomly assigned to one of the groups (guarantees independence) and that they were sampled from a nearly normal population.

### Difference in dose level

We apply 3 t-tests to cover all factor level combinations since there are 3 difference dose levels.

#### Test 1, Dose 0.5 and 1

```
len1 <- ToothGrowth %>% filter(dose %in% c(0.5,1)) %>% select(len) %>% unlist()
dose1 <- ToothGrowth %>% filter(dose %in% c(0.5,1)) %>% select(dose) %>% unlist()
t.test(len1 ~ dose1, paired=FALSE)
```

```
##
## Welch Two Sample t-test
##
## data: len1 by dose1
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -11.983781 -6.276219
## sample estimates:
## mean in group 0.5 mean in group 1
## 10.605 19.735
```

### Test 2, Dose 0.5 and 2

```
len2 <- ToothGrowth %>% filter(dose %in% c(0.5,2)) %>% select(len) %>% unlist()
dose2 <- ToothGrowth %>% filter(dose %in% c(0.5, 2)) %>% select(dose) %>% unlist()
t.test(len2~dose2, paired = FALSE)
```

```
##
## Welch Two Sample t-test
##
## data: len2 by dose2
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -18.15617 -12.83383
## sample estimates:
## mean in group 0.5 mean in group 2
## 10.605 26.100
```

### Test 3, Dose 1 and 2

```
len3 <- ToothGrowth %>% filter(dose %in% c(1,2)) %>% select(len) %>% unlist()
dose3 <- ToothGrowth %>% filter(dose %in% c(1,2)) %>% select(dose) %>% unlist()
t.test(len3~dose3, paired = FALSE)
```

```
##
## Welch Two Sample t-test
##
## data: len3 by dose3
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
## 19.735 26.100
```

In all 3 cases, it appear to be a positive relationship between dose level and teeth length, indicate by the p-value is lower than the default significance level 0.05. Thus, we reject  $H_0$ .

### Difference in Supplement

```
len <- ToothGrowth %>% select(len) %>% unlist()
supp <- ToothGrowth %>% select(supp) %>% unlist()

diff_supp <- t.test(len~supp, var.equal=F)
```

Since p-value of the test is 0.061 which is larger than the significance value, 0.05, we fail to reject the null hypothesis.

There is no strong evidence that the average difference in tooth length across the two supplement types is different from 0.

The hypothesis test result corresponds to the respective confidence interval -0.171, 7.571 because it includes 0.

### Conclusion

- The observed difference of tooth length mean across supplement types is statistically not different from 0 - There is a fact where the tooth length indeed varies across group of dose level and all of those differences are statistical significant.