

# Statistic Inference Week4 Part 2

Jie Yang

## Part 2:

### Analysis

Analyzing the ToothGrowth data set. Tasks includes:

1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)
4. State your conclusions and the assumptions needed for your conclusions.

### Data set Description:

The response is 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).

### Exploratory analysis

```
#load dataset of ToothGrowth - The Effect of Vitamin C on Tooth Growth in Guinea Pigs
data(ToothGrowth)
#Basic exploratory analysis:
#Get names:
names(ToothGrowth)
#Take a peek of the dataset
head(ToothGrowth)
summary(ToothGrowth)
str(ToothGrowth)
```

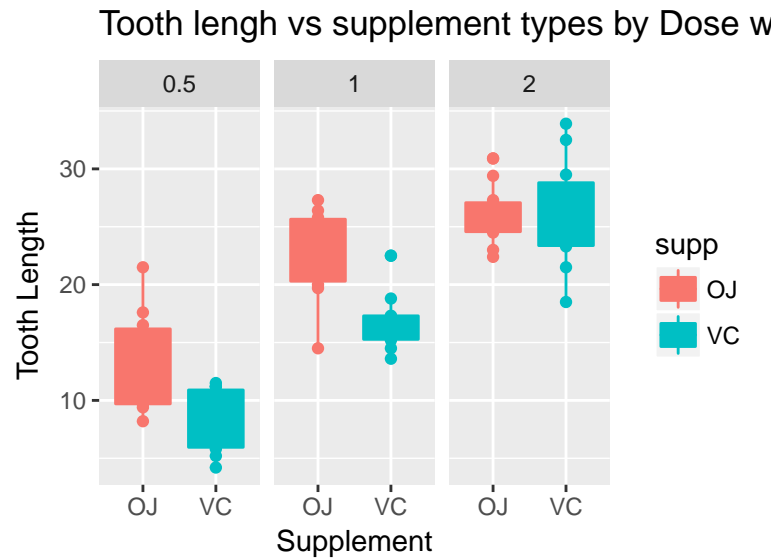
The data has 60 observations and 3 variables (from the str() we get the type of variables): 1. len (numeric) - Tooth length 2. supp (factor) - Supplement type (VC or OJ) 3. dose (numeric) - Dose in milligrams

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

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

### Visualization before generating the hypothesis:

```
library(ggplot2)
#qplot(dose, len, data=ToothGrowth, color=supp)
qplot(supp, len, data=ToothGrowth, facets=~dose, color=supp,
      xlab="Supplement", ylab="Tooth Length", main="Tooth length vs supplement types by Dose w qplot")+
  geom_boxplot(aes(fill = supp))
```



```
#g<-ggplot(ToothGrowth,aes(len,fill=supp))
```

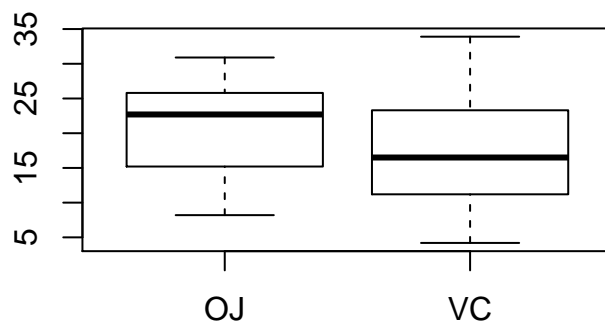
## Hypothesis Test

### Hypothesis Test I: VC vs OJ

- $H_0$ : Tooth Growth w VC and OJ doesn't show difference.
- $H_a$ : Tooth Growth w VC is lesser than OJ.

alternative hypothesis is generated from Visual plot

```
plot(ToothGrowth$supp,ToothGrowth$len)
```



## The t-test analysis

```

#subset the dataframe to get the OJ and VC
dt_OJ<-ToothGrowth[ToothGrowth$supp=='OJ',]
dt_VC<-ToothGrowth[ToothGrowth$supp=='VC',]
#Make a t-test since the n is small in this data set
#Confidence level is set to be 95%
t.test(dt_OJ$len, dt_VC$len, alternative = "greater", paired = F, var.equal = F, conf.level = 0.95)

##
## Welch Two Sample t-test
##
## data: dt_OJ$len and dt_VC$len
## t = 1.9153, df = 55.309, p-value = 0.03032
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.4682687      Inf
## sample estimates:
## mean of x mean of y
## 20.66333 16.96333

```

### Conclusion

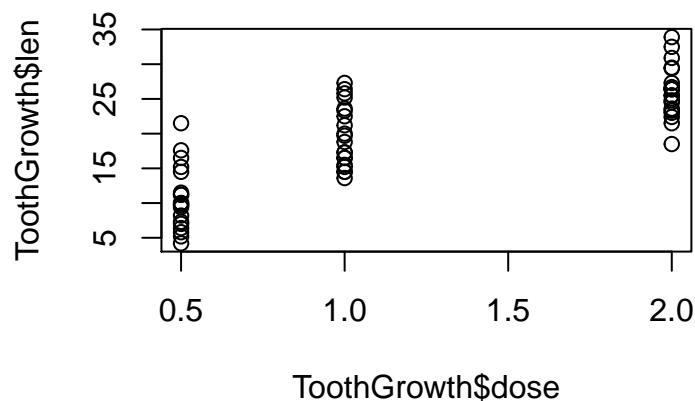
As the two group unpaired t-test show, p value is  $0.03 < 0.05$ , we can reject the Null hypothesis, and each to  $H_a$ : OJ is greater than VC for Tooth Growth.

## Hypothesis Test II: Dose

Since we have three doses, 0.5, 1.0 and 2, we will do the t-test with two doses at the same time. Given the plot below, we will generate following hypothesis in general:

- $H_0$ : Tooth Growth is independent of dose.
- $H_a$ : Tooth Growth increases as dose increases.

```
plot(ToothGrowth$dose, ToothGrowth$len)
```



Dose 0.5 vs Dose 1.0

- H0: Tooth Growth is independent of dose. 0.5 dose growth is equal to 1.0 dose.
- Ha: Tooth Growth for 0.5 dose is lesser than 1.0 dose.

```
dt_0p5<-ToothGrowth[ToothGrowth$dose==0.5,]
dt_1p0<-ToothGrowth[ToothGrowth$dose==1,]
t.test(dt_0p5$len,dt_1p0$len,alternative = "less", paired = F, var.equal = F, conf.level = 0.95)

##
## Welch Two Sample t-test
##
## data: dt_0p5$len and dt_1p0$len
## t = -6.4766, df = 37.986, p-value = 6.342e-08
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -6.753323
## sample estimates:
## mean of x mean of y
##      10.605      19.735
```

Given the p-value is very low, we'll reject the Null Hypothesis, and support Ha: ToothGrowth for 0.5 dose is less than 1.0 dose.

Similarly,

#### Dose 1.0 vs Dose 2.0

- H0: Tooth Growth is independent of dose. 1.0 dose growth is equal to 2.0 dose.
- Ha: Tooth Growth for 1.0 dose is lesser than 2.0 dose.

```
dt_2p0<-ToothGrowth[ToothGrowth$dose==2,]
t.test(dt_1p0$len,dt_2p0$len, alternative="less", paired=F, var.equal=F, conf.level=0.95)
```

Given p value is very small, we can reject the null hypothesis, and support Ha : ToothGrowth for 1.0 dose is less than 2.0 dose.

#### Conclusion

Put these two t-test conclusion together, we can conclude that in term of the ToothGrowth, 0.5 dose < 1.0 dose < 2.0 dose.

## Hypothesis Test III

Finally, as seen from the data exploratory session, @dose=2.0, we don't see much difference in lens. We want to test the null hypothesis there: whether the OJ and VC are different. (Notice our Hypothesis I only test the overall dataset VC vs OJ, here we test at dose=2.0)

Consistent Hypthesis would be:

- H0: at dose 2.0, VC and OJ doesn't show any difference in lens
- Ha: at dose 2.0, VC and OJ shows difference in lens.

```
dt_OJ2<-ToothGrowth[ToothGrowth$supp=='OJ' & ToothGrowth$dose==2.0,]
dt_VC2<-ToothGrowth[ToothGrowth$supp=='VC' & ToothGrowth$dose==2.0,]
t.test(dt_OJ2$len,dt_VC2$len, alternative = "two.sided", paired = F, var.equal = F, conf.level = 0.95)

##
## Welch Two Sample t-test
##
```

```
## data: dt_OJ2$len and dt_VC2$len
## 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 of x mean of y
## 26.06 26.14

# we also double check on the equal variance case
t.test(dt_OJ2$len,dt_VC2$len, alternative = "two.sided", paired = F, var.equal = T, conf.level = 0.95)

##
## Two Sample t-test
##
## data: dt_OJ2$len and dt_VC2$len
## t = -0.046136, df = 18, p-value = 0.9637
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -3.722999 3.562999
## sample estimates:
## mean of x mean of y
## 26.06 26.14
```

This is the “two-sided” t-test, as alternative is “unequal”

### Conclusion

As one can see, both equal and unequal variance case show very high p value, we fail to reject the NULL hypothesis, and we will conclude at dose=2.0, OJ and VC are equal in lens.

### Final Conclusions

- Overall, OJ makes ToothGrowth less than VC.  
But at dose=2.0, VC and OJ are equal.
- From dose 0.5 to 2.0, the ToothGrowth is increases.

### reference work

[https://rstudio-pubs-static.s3.amazonaws.com/67663\\_669e39c2153e4f8295f9d130abf07b7e.html](https://rstudio-pubs-static.s3.amazonaws.com/67663_669e39c2153e4f8295f9d130abf07b7e.html)

[https://github.com/UtkarshPathrabe/Statistical-Inference-Johns-Hopkins-Bloomberg-School-of-Public-Health-Coursera/blob/master/Course%20Project/Part\\_02\\_Basic\\_Inferential\\_Data\\_Analysis.md](https://github.com/UtkarshPathrabe/Statistical-Inference-Johns-Hopkins-Bloomberg-School-of-Public-Health-Coursera/blob/master/Course%20Project/Part_02_Basic_Inferential_Data_Analysis.md)

[https://github.com/alex23leem/Statistical-inference-project/blob/master/tooth\\_growth\\_analysis.Rmd](https://github.com/alex23leem/Statistical-inference-project/blob/master/tooth_growth_analysis.Rmd)