# Statistical Inference Course Project - Part 2

*Jonathan Chang*

*February 28, 2017*

## Overview

### 1. Exploratory Data Analysis

Load the ToothGrowth data and perform some basic exploratory data analyses

```
library(datasets)
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
```

```
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 ...
```

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose) # convert dose to factor
```

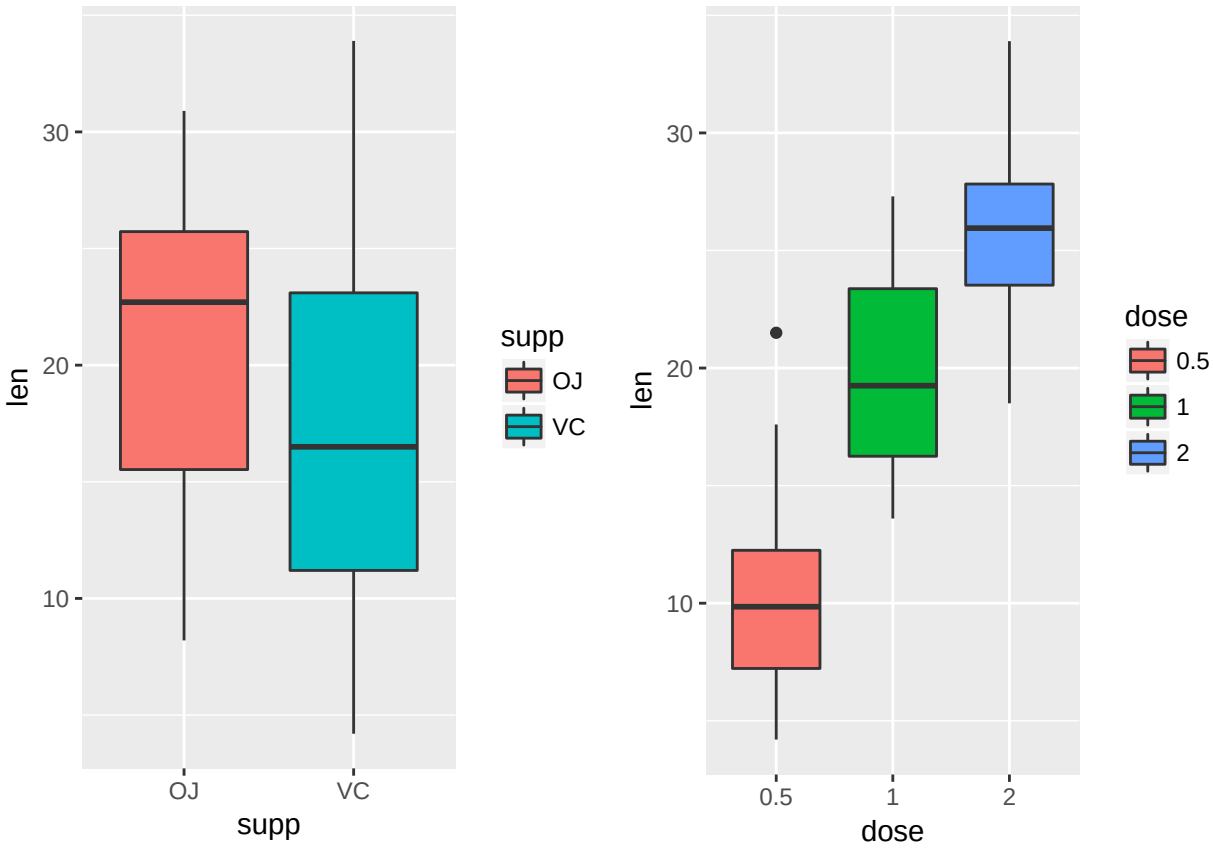
### 2. Basic Summary of the Data

Provide a basic summary of the data.

```
library(ggplot2)
library(gridExtra)
summary(ToothGrowth)
```

```
##       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
```

```
# basic boxplot
supp_plot = ggplot(ToothGrowth, aes(supp, len, fill = supp)) + geom_boxplot()
dose_plot = ggplot(ToothGrowth, aes(dose, len, fill = dose)) + geom_boxplot()
grid.arrange(supp_plot, dose_plot, ncol = 2)
```




### 3. Compare Tooth Growth by supp and dose

Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there are other approaches worth considering)

We perform the following T tests.

By supplement type:

```
t.test(len ~ supp, data=ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

By dose intervals:

```
tooth_dose_half_one <- subset(ToothGrowth, dose %in% c(0.5, 1.0))
tooth_dose_half_two  <- subset(ToothGrowth, dose %in% c(0.5, 2.0))
tooth_dose_one_two <- subset(ToothGrowth, dose %in% c(1.0, 2.0))
```

```
t.test(len ~ dose, data=tooth_dose_half_one)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## 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
```

```
t.test(len ~ dose, data=tooth_dose_half_two)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## 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
```

```
t.test(len ~ dose, data=tooth_dose_one_two)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## 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
```

And by supplement type and dose level:

```
tooth_dose_half <- subset(ToothGrowth, dose == 0.5)
tooth_dose_one <- subset(ToothGrowth, dose == 1.0)
tooth_dose_two <- subset(ToothGrowth, dose == 2.0)

t.test(len ~ supp, data=tooth_dose_half)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

```
t.test(len ~ supp, data=tooth_dose_one)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

```
t.test(len ~ supp, data=tooth_dose_two)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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 in group OJ mean in group VC
##            26.06            26.14
```

### 4. Conclusions

State your conclusions and the assumptions needed for your conclusions.

#### Assumptions

- the variables are independent and identically distributed
- tooth growth follows the normal distribution

#### Conclusions

For the tests where we obtained a p-value $> 0.05$ we could not reject the null hypothesis. Based on the cases where the reported p-values were lower than 0.05, we conclude the following:

- Dosage is positively correlated with tooth growth (independent of supplement type).
- At lower dosages (.5 - 1 Mg), orange juice causes greater tooth growth than vitamin c.