Statistical Inference Project - Inferential Data Analysis

William Matthews

January 19, 2018

Load packages

```
library(ggplot2)
```

Summary:

We will study the The effect of vitamin c on tooth growth in Guinea Pigs using two supplements (orange juice and ascorbic acid (vc)) administered at three different dosage levels (0.5, 1, 2).

Exploratory data analysis of the ToothGrowth dataset

Load the ToothGrowth dataset

```
data("ToothGrowth")
```

View dimensions of the dataset

```
dim(ToothGrowth)
```

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

View variable names in the datasets

```
names(ToothGrowth)
```

```
## [1] "len"  "supp" "dose"
```

View structure of the dataset.

```
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 ...
```

Since there are only three unique values in the dose variable convert from numeric to factor

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
```

View structure of data

```
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: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```
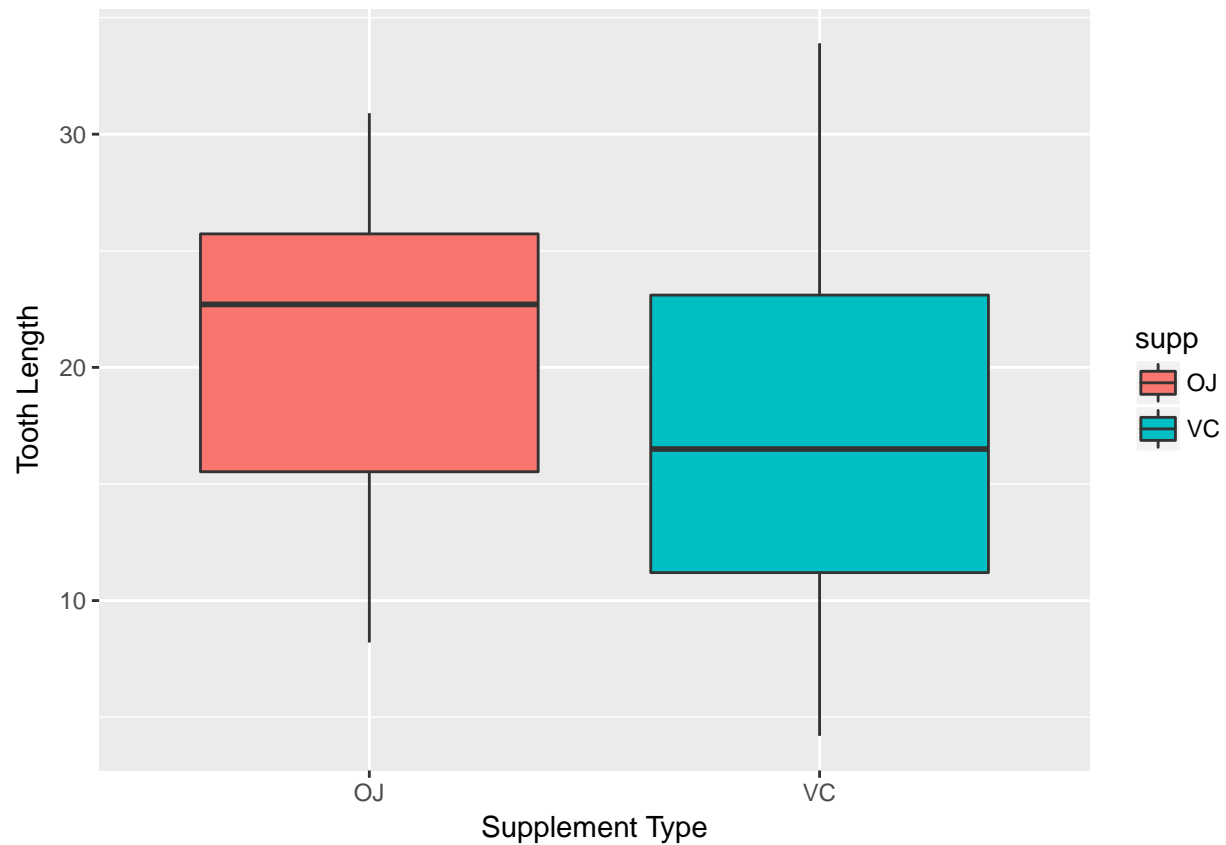
View summary of the dataset

```
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
```

Create a boxplot graph to show the tooth lenght by supplement

```
ggplot(aes(x = supp, y = len), data = ToothGrowth) + geom_boxplot(aes(fill = supp)) +
  xlab("Supplement Type") + ylab("Tooth Length")
```
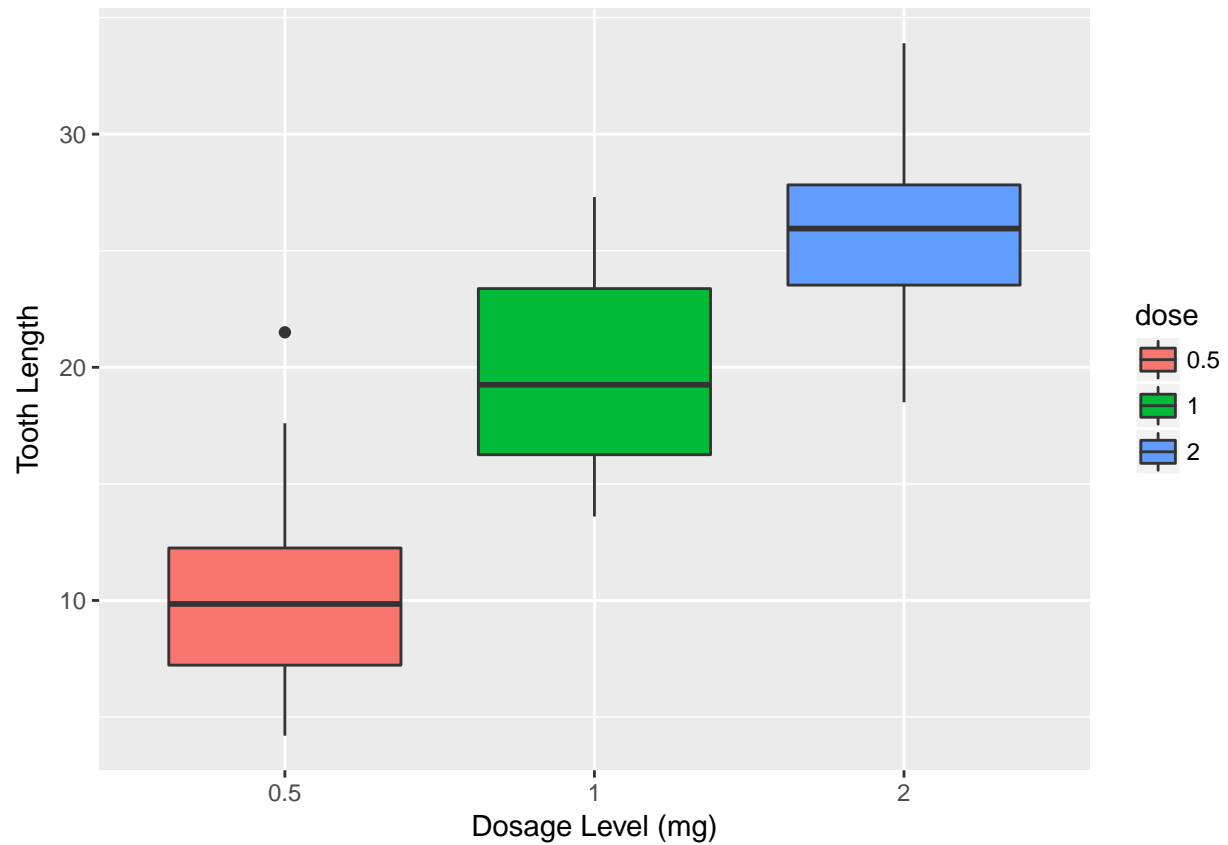
Calculate the mean tooth growth by dosage level.

```
meanDosage <- split(ToothGrowth$len, ToothGrowth$dose)
sapply(meanDosage, mean)
```

```
##    0.5      1      2
## 10.605 19.735 26.100
```

Create a boxplot graph to show the tooth length by dosage level

```
ggplot(aes(x = dose, y = len), data = ToothGrowth) + geom_boxplot(aes(fill = dose)) +
  xlab("Dosage Level (mg)") + ylab("Tooth Length")
```

Create a boxplot graph to compare the supplements at each dosage level

```
ggplot(aes(x = supp, y = len), data = ToothGrowth) + geom_boxplot(aes(fill = supp)) +
  facet_wrap(~dose) + xlab("Dosage Level (mg)") + ylab("Tooth Length")
```
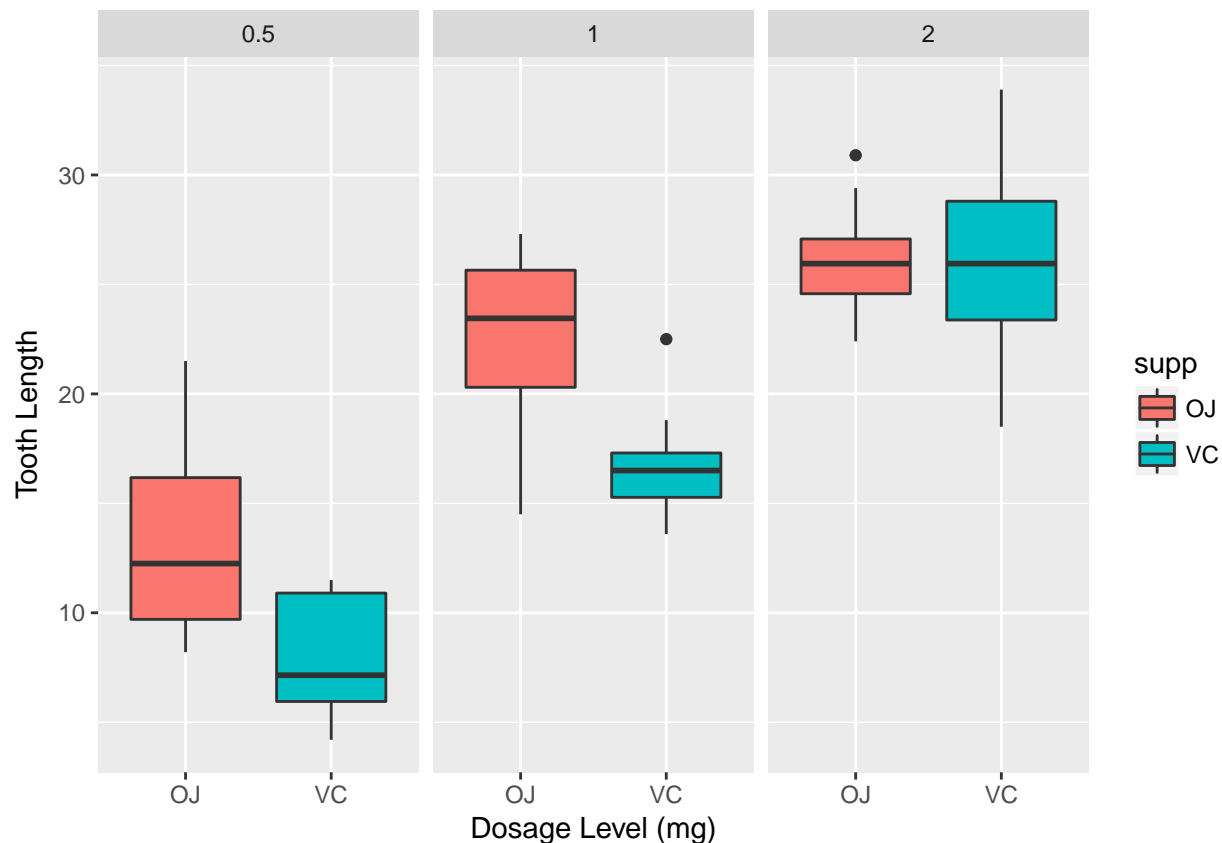
Analysis & Hypothesis testing OJ appears to be more effective for tooth growth at lower dosage levels. At 2mg both supplements are relatively equal.

The graphs seem to indicate there is an increase in tooth growth with the increase in supplement dose and supplement type. Let the null hypothesis be that supp type does not affect tooth growth.

Use t-test to test for differences in tooth growth due to 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
```

With a p-value at 0.06 and the confidence interval containing 0 we cannot reject the null hypothesis that supp type has no affect on tooth growth.

Use t-test to test for differences in tooth growth due to dosage level.

First create three groups to compare the dosage levels

```
dose_a <- subset(ToothGrowth, dose %in% c(0.5,1.0))
dose_b <- subset(ToothGrowth, dose %in% c(0.5,2.0))
dose_c <- subset(ToothGrowth, dose %in% c(1.0,2.0))
```

Now test each level

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

```
##
##   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 = dose_b)
```

```
##
##   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 = dose_c)
```

```
##
##   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
```

For each of these tests the p-value was less than 0.05 and the confidence interval does not contain 0. Therefore, we can reject the null hypothesis that changes in dosage levels do not affect tooth growth.

Conclusion:

We cannot reject the null hypothesis that supplement type does not affect tooth growth. We can rejet the null hypothesis that dosage levels do not affect tooth growth.

Assumptions:

1. The samples are independent and the variances are assumed to be different.
2. The sample represents the true population.