# Course-Project-2

# Part 2: Basic Inferential Data Analysis

## Overview

This section contains an exploratory analysis of the ToothGrowth data in R’s datasets package, a basic summary of the data, and an analysis of the effects of the supp and dose fields.

## Loading and Exploring the Data

Make sure the datasets package is loaded, and then grab the ToothGrowth data:

```
library(datasets)
tg <- ToothGrowth
```

Next, explore some basic aspects of the data, to get a sense of what’s going on.

```
dim(tg) #dimensions of data
```

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

```
head(tg) #examine the first bit of data
```

```
##    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
```

```
apply(tg, 2, class) #classes of the fields
```

```
##         len        supp        dose 
## "character" "character" "character"
```

## Basic Summary

After very basic exploratory analysis, check the contents of the data itself

```
summary(as.numeric(tg$len))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##    4.20   13.07   19.25   18.81   25.27   33.90
```

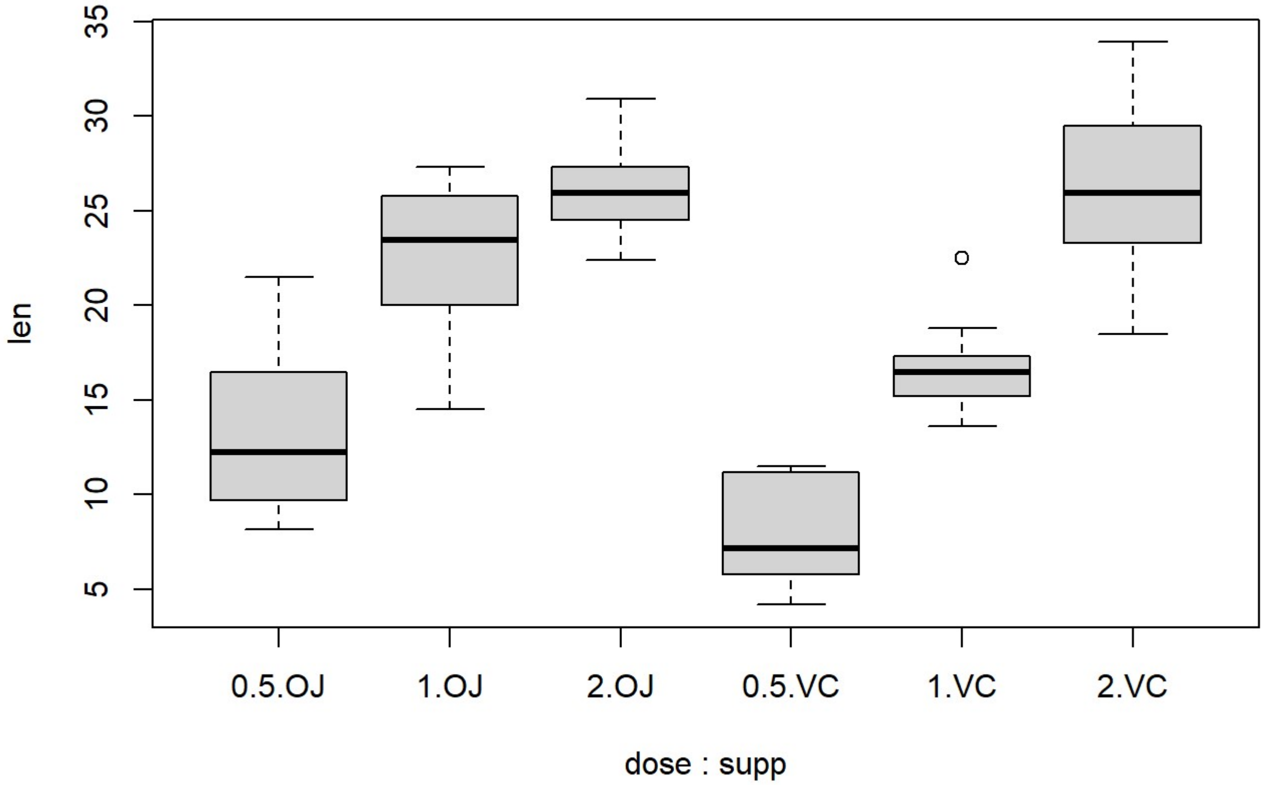
```
table(tg$supp)
```

```
## 
## OJ VC 
## 30 30
```

```
table(tg$dose)
```

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

```
boxplot(len ~ dose*supp, data=tg)
```



Since supp and dose are discrete while len is continuous, it seems likely that this is data from an experiment testing different doses of two different supplements and measuring tooth length. Visually, dose seems strongly correlated with greater len. The relationship of sup with len is harder to discern.

# Inferential Statistical Analysis

Use the statistical tools learned in class (t-test, p-values) to discern the relationships of supp and dose with len.

```
t.test(tg$len~tg$supp)[c("p.value", "estimate")] #test difference between supp="VC" and supp="OJ"
```

```
## $p.value
## [1] 0.06063451
##
## $estimate
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

```
t.test(subset(tg$len, tg$dose==1), subset(tg$len, tg$dose==0.5), alternative="greater")[c("p.value", "estimate")] #test difference between dose 0.5 and 1
```

```
## $p.value
## [1] 6.341504e-08
##
## $estimate
## mean of x mean of y
##    19.735    10.605
```

```
t.test(subset(tg$len, tg$dose==2), subset(tg$len, tg$dose==1), alternative="greater")[c("p.value", "estimate")] #test difference between dose 1 and 2
```

```
## $p.value
## [1] 9.532148e-06
##
## $estimate
## mean of x mean of y
##    26.100    19.735
```

# Conclusions and Underlying Assumptions

The results of the t-tests above are as follows: at an alpha=0.05 level, there is not sufficient evidence for a difference in population means between the OJ and VC supplements. However, there is sufficient evidence for differences between all three dose levels, pairwise.

These results rely on the following assumptions: observations are iid, groups are unpaired, variables are roughly normally distributed. There are other assumptions of course, but these three seem most pertinent.