Statistical Inference Course Project Part 2 - Kevin O'Leary

Now we're going to analyze the ToothGrowth data in the R datasets package which concerns the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs, each receiving one of three dose levels of Vitamin C (0.5, 1.0, and 2.0 mg) with one of two delivery methods (orange juice or an aqueous solution of ascorbic acid).

1. Load the ToothGrowth data and perform some basic exploratory data analyses

```
library(datasets)
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 ...
```

We can see that 'dose' is a numeric vector when in fact the contents can take only one of three levels. It is necessary then to convert this vector to a factor vector.

```
ToothGrowth$dose <- factor(ToothGrowth$dose)
summary(ToothGrowth)
```

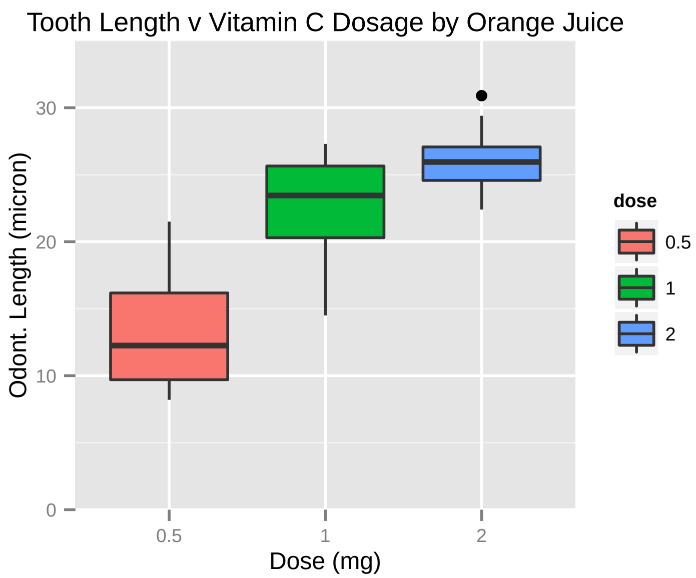
```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

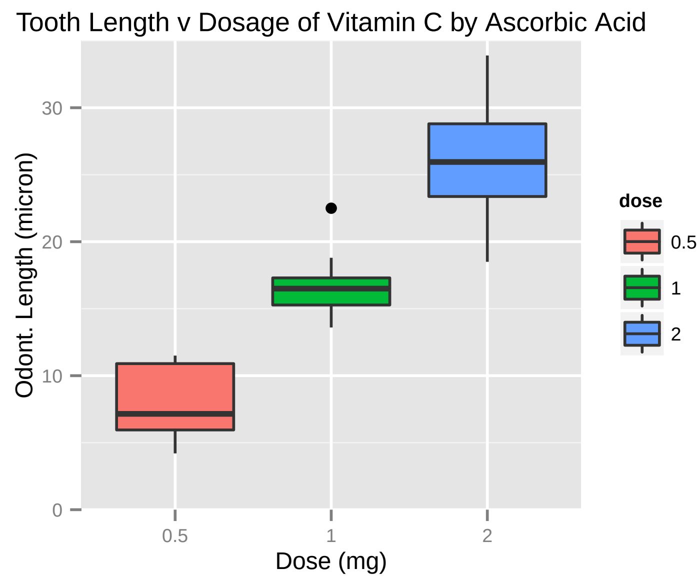
Let's take a graphical look at the data:

```
library(ggplot2)

ojsubset <- ToothGrowth[ToothGrowth$supp == "OJ",] ##subset data where supp = OJ
orangejplot <- ggplot(ojsubset, aes(x=dose,y=len, fill=dose)) +
    geom_boxplot() +
    theme(plot.title = element_text(size = 10),
         text = element_text(size = 9))+
    coord_cartesian(ylim=c(0,35))+
    labs(title='Tooth Length v Vitamin C Dosage by Orange Juice', x='Dose (mg)', y='Odont. Length (micron)')

acidsubset <- ToothGrowth[ToothGrowth$supp == "VC",] ##subset data where supp = VC
acidsubsetplot <- ggplot(acidsubset, aes(x= dose, y=len, fill= dose)) +
    geom_boxplot() +
    theme(plot.title = element_text(size = 10),
         text = element_text(size = 9))+
    coord_cartesian(ylim=c(0,35))+
    labs(title='Tooth Length v Dosage of Vitamin C by Ascorbic Acid', x='Dose (mg)', y='Odont. Length (micron)')
```





2. Provide a basic summary of the data.

The boxplots appear to show that orange juice delivery provides greater tooth length at lower dosages only (0.5 and 1) than ascorbic acid. Increases in dosage however, results in increases in tooth length for either delivery method.

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

To compare tooth growth by supplement and dose we will use the `t.test` function to determine if the two sets of data are significantly different to each other.

The null hypothesis is that the means of both data sets are equal with the alternative hypothesis being that the difference in tooth length is statistically significant. The null hypothesis will be rejected if we find a p-value of less than 0.05.

Let's first subset the data further.

```
acid05 <- acidsubset[acidsubset$dose==0.5,]
acid10 <- acidsubset[acidsubset$dose==1.0,]
acid20 <- acidsubset[acidsubset$dose==2.0,]
oj05 <- ojsubset[ojsubset$dose==0.5,]
oj10 <- ojsubset[ojsubset$dose==1.0,]
oj20 <- ojsubset[ojsubset$dose==2.0,]
```

First we t-test the effect of supplement type on tooth length while holding dosage constant.

```
# Perform t test on acidsubset vs ojsubset at 0.5mg dose
t.05.acidoj <- t.test(len ~ supp, data=rbind(acid05,oj05), var.equal=FALSE)

#1.0mg dose
t.10.acidoj <- t.test(len ~ supp, data=rbind(acid10,oj10), var.equal=FALSE)

#2.0mg dose
t.20.acidoj <- t.test(len ~ supp, data=rbind(acid20,oj20), var.equal=FALSE)
```

Table 1: Summary of Results

supplements compared	dose level	p-value	conf. int. (-)	conf. int. (+)
Ascorbic Acid and OJ	0.5mg	0.0063586	1.719057	8.780943
Ascorbic Acid and OJ	1.0mg	0.0010384	1.719057	9.057852
Ascorbic Acid and OJ	2.0mg	0.9638516	1.719057	3.638070

Now we'll test the effect of dosages on tooth length:

```
# Perform t test on 0.5mg vs 1.0mg, within each supplement
t.acid.0510 <- t.test(len ~ dose, data=rbind(acid05,acid10), var.equal=TRUE)
t.oj.0510 <- t.test(len ~ dose, data=rbind(oj05,oj10), var.equal=TRUE)

# 1.0mg vs 2.0mg
t.acid.1020 <- t.test(len ~ dose, data=rbind(acid10,acid20), var.equal=TRUE)
t.oj.1020 <- t.test(len ~ dose, data=rbind(oj10,oj20), var.equal=TRUE)

# 0.5mg vs 2.0mg
t.acid.0520 <- t.test(len ~ dose, data=rbind(acid05,acid20), var.equal=TRUE)
t.oj.0520 <- t.test(len ~ dose, data=rbind(oj05,oj20), var.equal=TRUE)
```

Table 2: Summary of Results

doses compared	supplement	p-value	conf. int. (-)	conf. int. (+)
0.5mg and 1.0mg	Acid	0.0000006	-11.264345	-6.3156545
0.5mg and 1.0mg	OJ	0.0000836	-13.410814	-5.5291857
1.0mg and 2.0mg	Acid	0.0000340	-12.968960	-5.7710402
1.0mg and 2.0mg	OJ	0.0373628	-6.500502	-0.2194983
0.5mg and 2.0mg	Acid	0.0000000	-21.832843	-14.4871567
0.5mg and 2.0mg	OJ	0.0000003	-16.278223	-9.3817774

4. State your conclusions and the assumptions needed for your conclusions.

The analysis shows a direct connection between dosage and tooth growth. p-values for all tests were below threshold and confidence intervals do not include zero, indicating that the increase in tooth growth when supplement dose is increased is statistically significant.

We see however, that the correlation between supplement type and tooth growth diminishes at the highest dose. For 0.5mg and 1.0mg, p-values are below threshold and the confidence interval does not include zero, indicating that the increase in tooth growth that's seen with orange juice vs ascorbic acid is statistically significant. At 2.0mg, however, this discrepancy vanishes and our null hypothesis at this level is accepted.