ToothGrowth Project

We analyze the ToothGrowth data in the R datasets package.

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

From the R datasets package …

"The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, (orange juice or ascorbic acid (a form of vitamin C and coded as VC))."

We load the libraries that we'll use and the ToothGrowth data set.

```
library(datasets)
library(ggplot2)

data(ToothGrowth)
```

Let's look at what's inside.

```
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 ...
```

So we have 60 observations (i.e. rows) of 3 variables (i.e. columns).

Provide a basic summary of the data.

Let's take a quick summary of the data.

```
summary(ToothGrowth)
```

```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

Build a contingency table of the counts at each combination of factor levels.

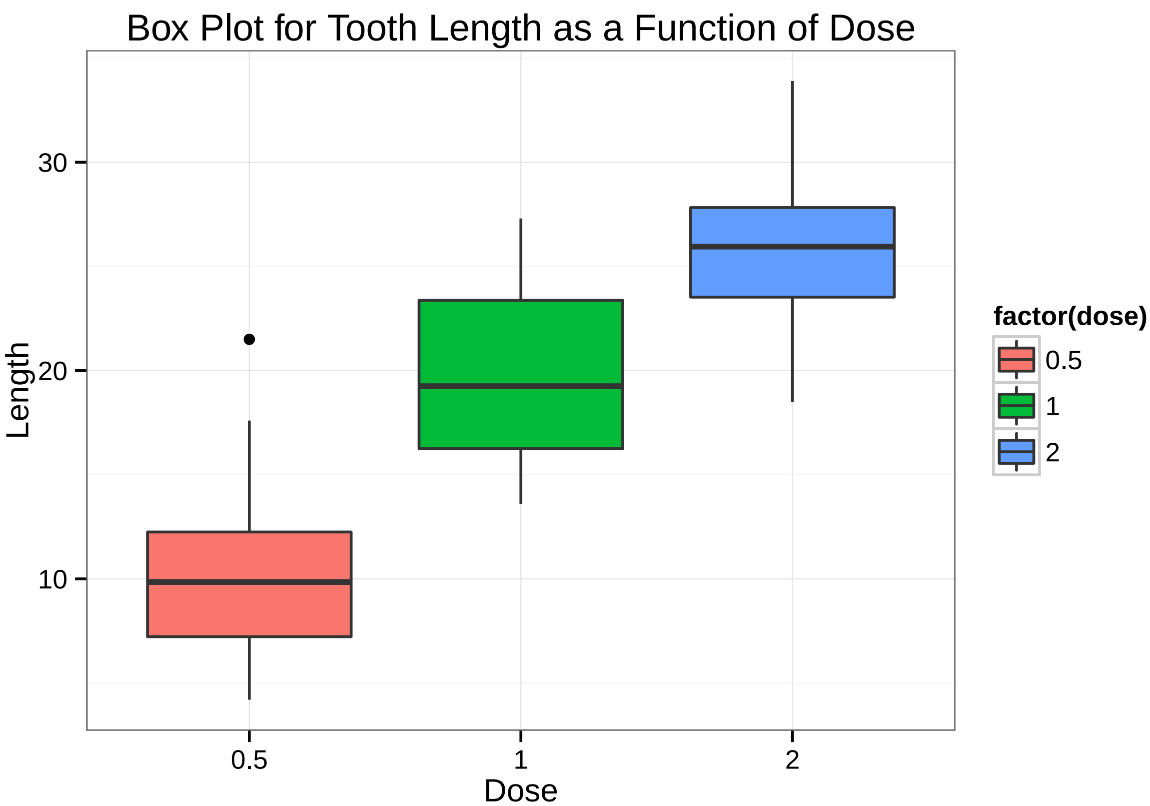
```
table(ToothGrowth$dose, ToothGrowth$supp)
```

```
##
##        OJ VC
##    0.5 10 10
##    1   10 10
##    2   10 10
```

Let's look at box plots for tooth growth as a function of dose and as a function of delivery method.

```
dose_box <- ggplot(data=ToothGrowth, aes(x=as.factor(dose), y=len)) +
            xlab("Dose") + ylab("Length") + theme(legend.title=element_blank()) +
            ggtitle("Box Plot for Tooth Length as a Function of Dose") +
            geom_boxplot(aes(fill=factor(dose))) + theme_bw()


dose_box
```



From the box plot above, it appears that a greater dose a greater effect on tooth length.

```
supp_box <- ggplot(data=ToothGrowth, aes(x=as.factor(supp), y=len)) +
            xlab("Delivery Method") + ylab("Length") + theme(legend.title=element_blank()) +
            ggtitle("Box Plot for Tooth Length as a Function of Supplement Type") +
            geom_boxplot(aes(fill=factor(supp))) + theme_bw()


supp_box
```

From the box plot above, it appears that orange juice *might* have a slightly greater effect on tooth length, than vitamin C.

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)

We use t-tests for doses and for supplements.

First, we look at doses.

```
# t-test for different dosages
dose_0.5 <- subset(ToothGrowth, dose == 0.5)
dose_1 <- subset(ToothGrowth, dose == 1)
dose_2 <- subset(ToothGrowth, dose == 2)
```

We consider three cases:

```
t.test(dose_0.5$len,dose_2$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose_0.5$len and dose_2$len
## 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 of x mean of y
##    10.605    26.100
```

```r
t.test(dose_1$len,dose_2$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose_1$len and dose_2$len
## 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 of x mean of y
##    19.735    26.100
```

```r
t.test(dose_0.5$len,dose_1$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose_0.5$len and dose_1$len
## 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 of x mean of y
##    10.605    19.735
```

Looking at the t-test results for the differing doses, we see small p-values. For the difference between a dosage of 0.5 and 2, we have a p-value < 0.00000000001. The ecidence is strong, that increased doses results in increased tooth length.

Next, we look at supplement type.

```r
# t-test for supplement types
supp_OJ <- subset(ToothGrowth, supp == "OJ")
supp_VC <- subset(ToothGrowth, supp == "VC")
```

There is only one case to consider:

```r
t.test(supp_OJ$len,supp_VC$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  supp_OJ$len and supp_VC$len
```

```
## 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 of x mean of y
##  20.66333  16.96333
```

We have a p-value equal to 0.06063 which is greater than the 0.05 significance level. This leads us to conclude that the effect of the supplements is similar.

State your conclusions and the assumptions needed for your conclusions.

We conclude that supplement types have similar effects on tooth growth. On the other hand, increasing doses leads to a significant increase in tooth length.

The assumptions underlying this analysis are that the guinea pigs were sampled from a normally distributed population. The samples were drawn randomly.