Statistical Inference Week 4 Project

Grejell Segura

8/13/2017

Statistical Inference - Week 4 Assignment: Course Project Part 2

Overview

This report aims to show the difference of tooth growth among various treatments. There are 2 factors considered in this analysis namely supplement and dose. The observed effect of these factors to the length of tooth growth is analysed using hypothesis testing.

Load the necessary library and ToothGrowth data from the datasets package

```r
library(ggplot2)
data(ToothGrowth)
```

Summary of the Data

```r
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 ...
```

The data has 60 observations with 3 variables (length, supplement, and dose). Supplement is of categorical type with 2 levels - OJ and VC. Meanwhile, Dose is numerical but it has to be converted to categorical with 3 levels - 0.5, 1, and 2. Length on the other hand is numerical.

```r
summary(ToothGrowth$len)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4.20   13.08   19.25   18.81   25.28   33.90
```

Length has mean 18.81 with 4.20 and 33.90 as the respective minimum and maximum values.

```r
summary(ToothGrowth$supp)
```

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

There are 30 observations both for OJ and VC supplements

Convert dose to categorical.

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

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

There are also 20 observations per levels of dose.

Visualization

To further understand the data, we need to create a chart to show the difference among groups/levels betweens the factors.

First, let us show the plot between supplements

```
plot2 <- ggplot(ToothGrowth, aes(x = as.factor(supp), y = len, color = supp)) + geom_boxplot()
plot2 <- plot2 + geom_point(size = 5, alpha = 0.8)
plot2 <- plot2 + labs(x = "Supplement", y = "Length") + theme_bw()
plot2
```
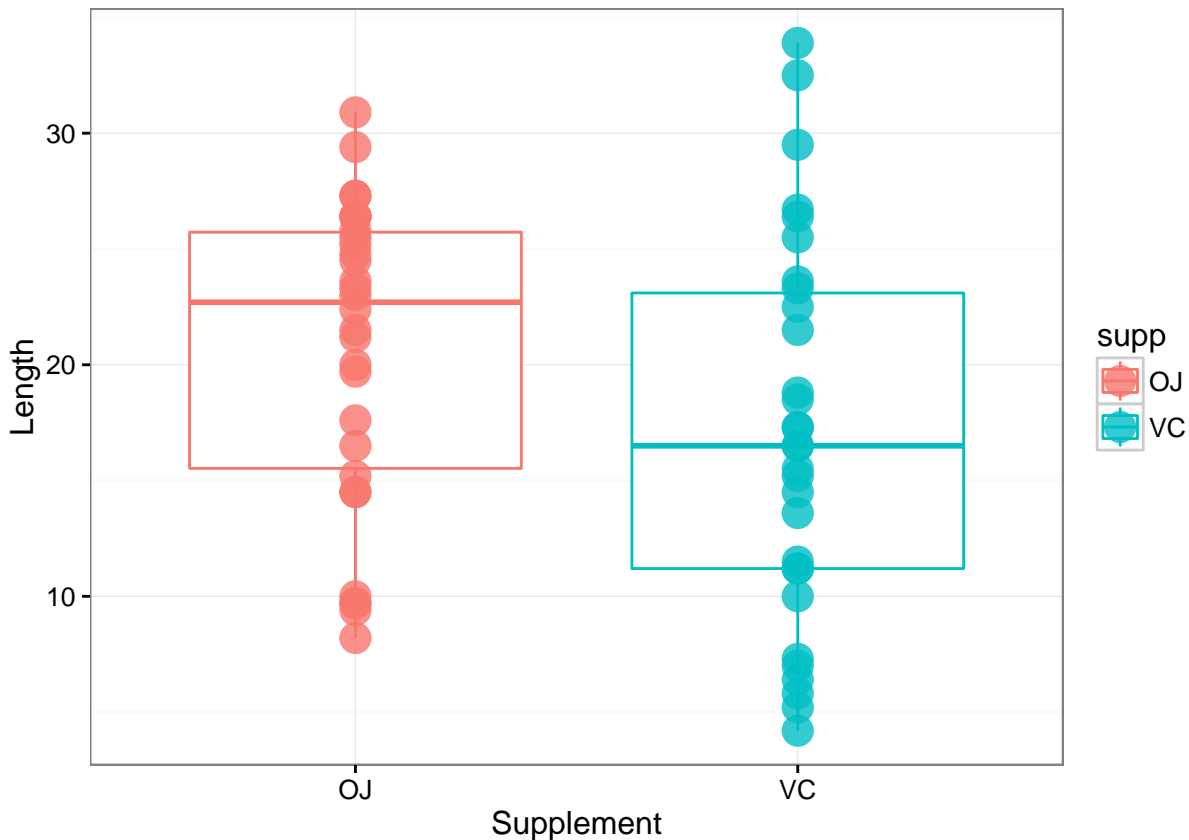

Here is the plot for the dose.

```
plot3 <- ggplot(ToothGrowth, aes(x = as.factor(dose), y = len, color = as.factor(dose))) + geom_boxplot
plot3 <- plot3 + geom_point(size = 5, alpha = 0.8)
plot3 <- plot3 + labs(x = "Dose", y = "Length") + theme_bw()
plot3
```
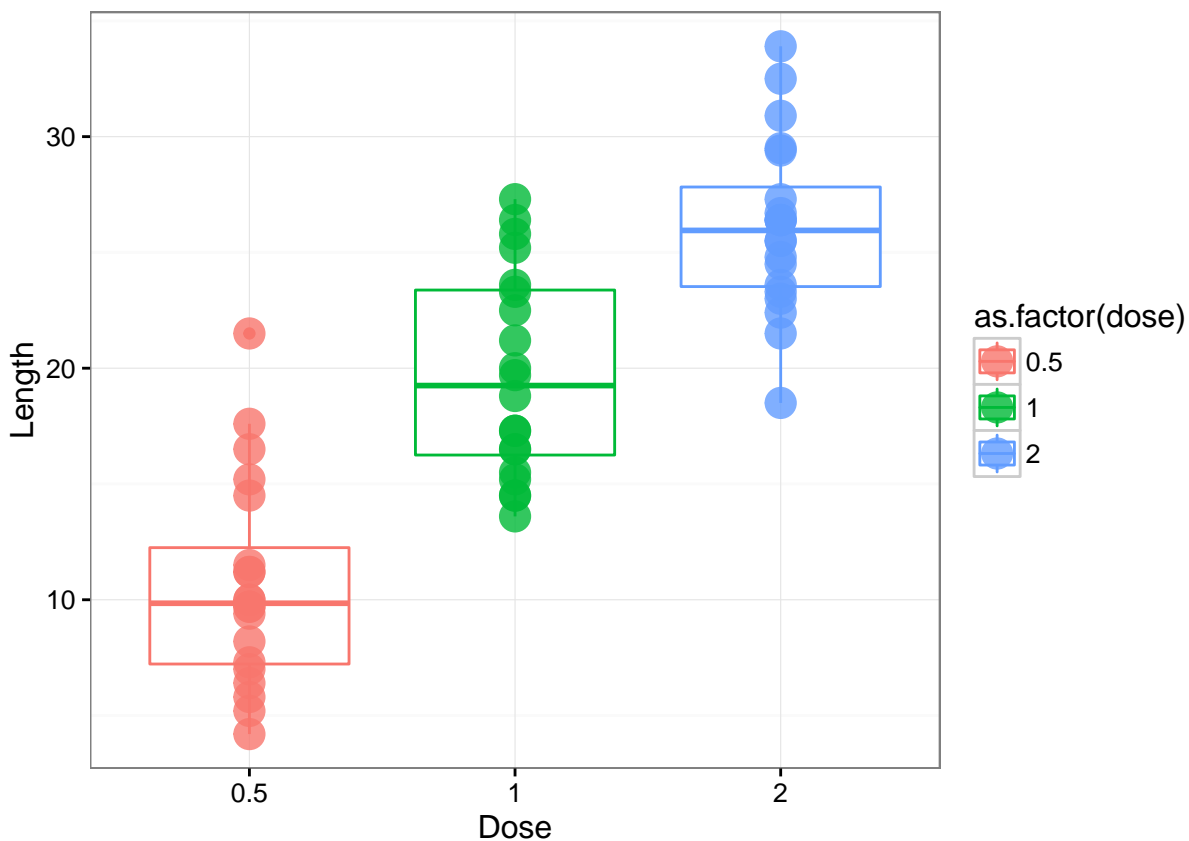
Lastly, here is the plot for the dose broken down between supplements.

```r
plot4 <- ggplot(ToothGrowth, aes(x = as.factor(dose), y = len, color = as.factor(dose))) + geom_boxplot
plot4 <- plot4 + facet_grid(~supp) + geom_point(size = 5, alpha = 0.8)
plot4 <- plot4 + labs(x = "Dose", y = "Length") + theme_bw()
plot4
```
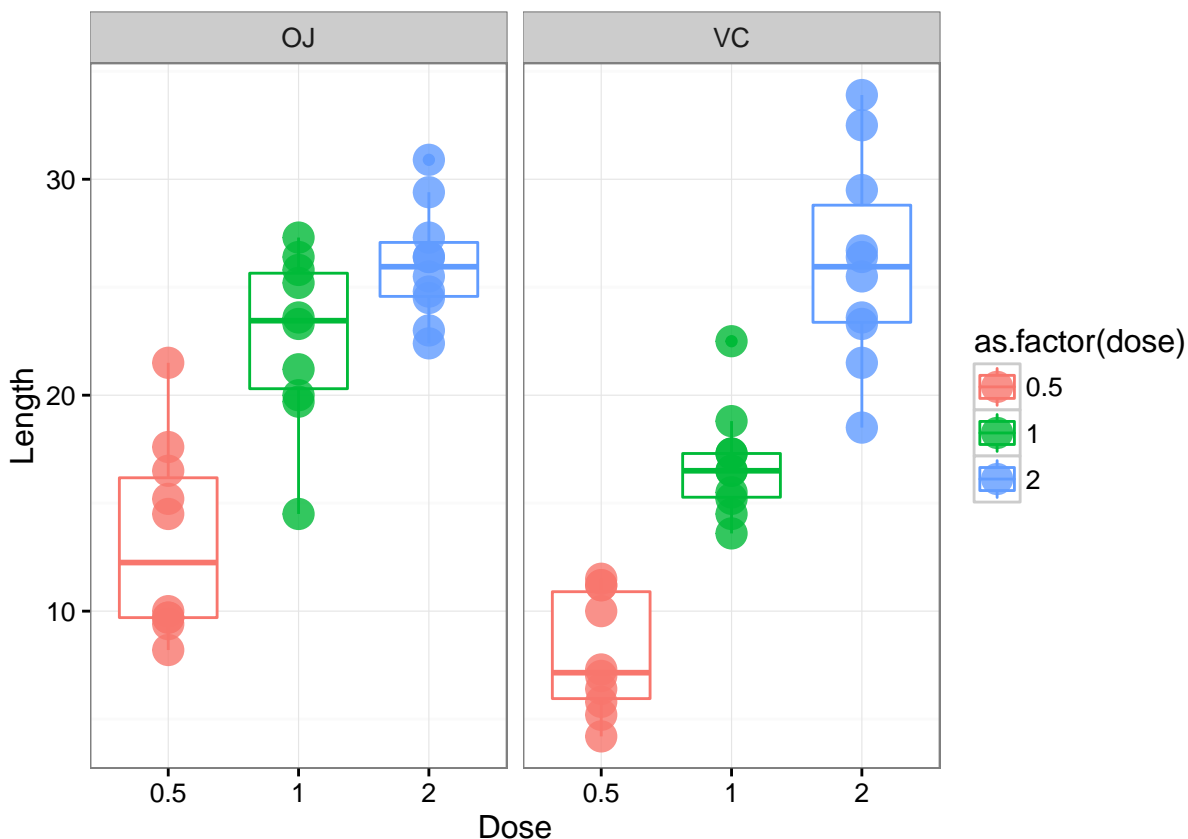
As shown in the plots, we might have a hypothesis that there are differences of tooth growth between the groups in dosage and supplements. To furthere investigate this, we will analyse using hypothesis testing.

1. Supplements

   One hypothesis we come up is there is difference between the 2 groups of supplements OJ and VC. Hence we proceed with the following experiment:

   Ho : there is no difference in tooth growth between OJ and VC.

   Ha : there is difference in tooth growth between OJ and VC.

   We wish to be 95% confident of the result.

```
### difference between supplements ###
model1 <- t.test(len ~ supp, data = ToothGrowth)
model1
```

```
##
##  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
```

As a result above, @ p = 0.06063 and alpha = 0.05 we can not reject the null hypothesis. This means we don't find any statistical evidence that OJ and VC has different tooth growth means. Furthermore, 0 also lies

between the calculated 95% confidence interval (-0.1710156, 7.5710156). This show that the mean difference may be 0 in some occassions.

2. Dose
   Another hypothesis we come up is there is difference between the 3 groups of dosage. Hence we proceed with the following experiment. This time, since we wish to compare 3 groups, we will use ANOVA.
   Ho : there is no difference in tooth growth between 3 types of dosage.
   Ha : there is difference in tooth growth between 3 types of dosage.
   We wish to be 95% confident of the result.

```
model2 <- aov(len ~ dose, data = ToothGrowth)
summary(model2)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## dose         2   2426    1213   67.42 9.53e-16 ***
## Residuals   57   1026      18
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The F value is 67.42 which converts to a p-value = 9.53e-16 which is very small. Therefore, we are 95% confident (alpha = 0.05) that there is difference of tooth growth among the 3 types of dosages.

3. Interaction between Supplement and Dose
   Lastly, we wish to see if there is an observed interaction between supplement and dosage to the tooth growth. We found the result below by using a 2-way ANOVA.

```
model3 <- aov(len ~ dose * supp, data = ToothGrowth)
summary(model3)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## dose         2 2426.4  1213.2  92.000  < 2e-16 ***
## supp         1  205.4   205.4  15.572 0.000231 ***
## dose:supp    2  108.3    54.2   4.107 0.021860 *
## Residuals   54  712.1    13.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

@ p-value = 0.02, the model show that there is a very significant interaction between supplement and dosage. This means that the effect of supplements to the tooth growth depends on the effect of dosage and vice versa.

Conclusion

The study has investigated the effect of Supplement and Dosage to the tooth growth of guinea pigs. To generate results, we conducted hypothesis testing and made the following assumptions below:
1. The sample generated was randomly picked and is representative of the population.
2. The tooth growth length follow a normal distribution.
3. The tooth growth length variance is equal among all groups of supplement and dosage.

The following results were observed:
1. At alpha = 0.05, there is no significant difference to the tooth growth between supplements OJ and VC
2. At alpha = 0.05, there is significant difference to the tooth growth between 3 types of dosages.
3. At alpha = 0.05, there is an observed interaction of effects between supplement and dosage to the tooth growth.