Statistical Inference Course Project (part 2)

Overview

Here we will analyze the `ToothGrowth` data set from R. This data set evaluates the effect of vitamin C on tooth growth in guinea pigs.

The `ToothGrowth` data set consist of 3 variables with 60 observations. The variable `len` is a numeric variable that gives the tooth length in millimeters. The variable `supp` is a factor variable with two levels ("OJ" and "VC") that gives the supplement type administered to the guinea pigs. The variable `dose` is a numeric variable with values of 0.5mg, 1mg, and 2mg.

There were 60 guinea pigs total and each was assigned to one of six groups. Three groups of ten guinea pigs were given orange juice in the various dosage. Three groups of ten guinea pigs were given ascorbic acid in the various dosage.

The goal of this analysis is to determine which treatment yielded the best result.

Eploratory data analysis.

Exploratory data analysis was performed using R and ggplot2. Initial findings reveal that the mean tooth length was 18.81 with a standard deviation of 7.65.

```r
round(mean(ToothGrowth$len), 2)
```

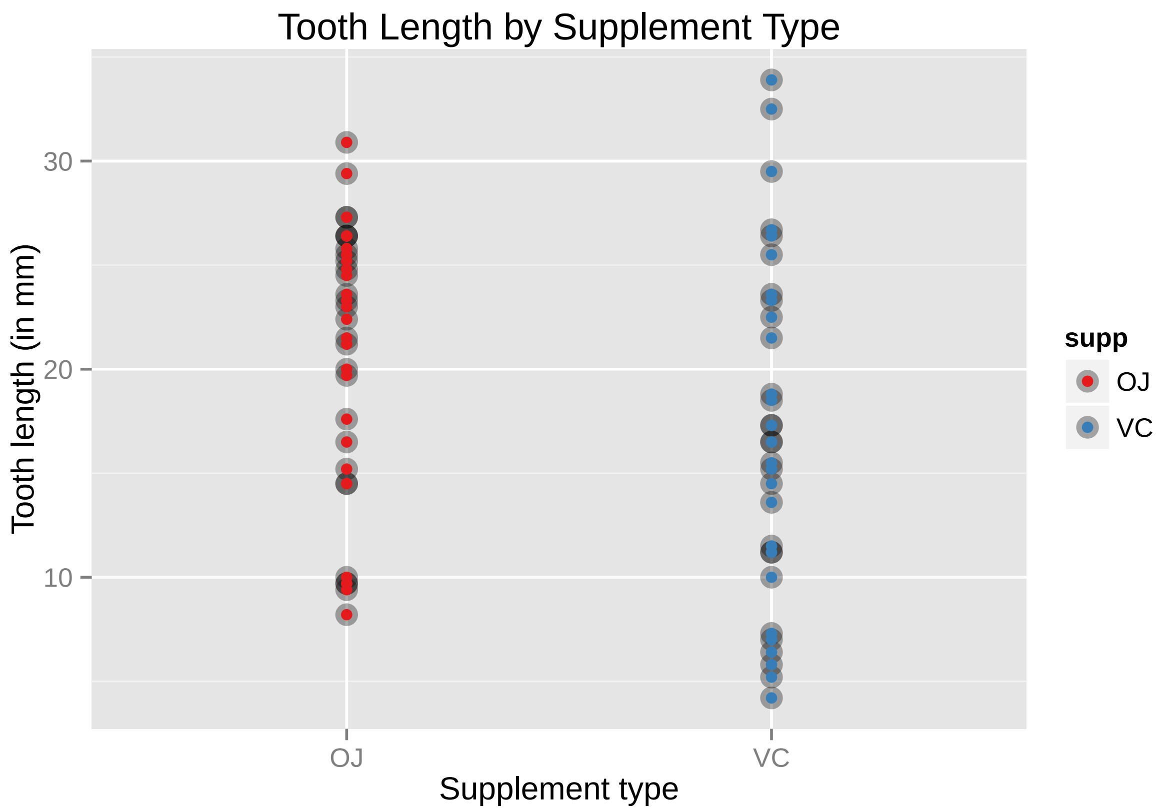
```
## [1] 18.81
```

```r
round(sd(ToothGrowth$len), 2)
```

```
## [1] 7.65
```

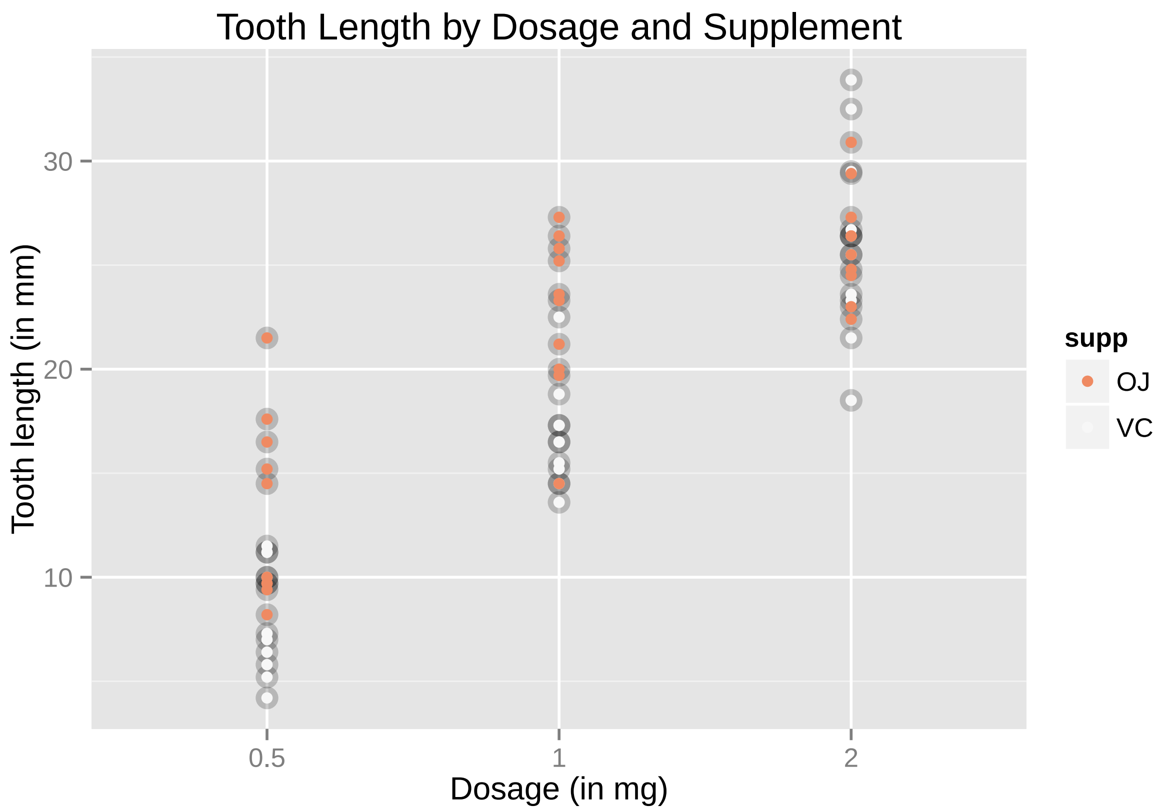
However, these results do not reveal anything relating to treatment type or dosage, so we will look at how these variables may have impacted tooth length using visualization of the results.

```r
library(RColorBrewer)
library(ggplot2)
LengthSupp <- ggplot(ToothGrowth, aes(supp, len, fill = supp)) +
  geom_point(stat = "identity", size = 4, alpha = 1/3) +
  geom_point(aes(col = supp)) +
  scale_colour_brewer(palette = "Set1") +
  xlab("Supplement type") + ylab("Tooth length (in mm)") +
  ggtitle("Tooth Length by Supplement Type")
LengthSupp
```



The plot above reveals much more variation in tooth length for the guinea pigs given ascorbic acid. However, to get a better picture of the data, we should examine the results when supplement type is combined with dosage.

```
LengthInter <- ggplot(ToothGrowth, aes(x = as.factor(dose), y = len, group = supp)) +
  geom_point(stat = "identity", size = 4, alpha = 1/5) +
  geom_point(aes(col = supp))+
  scale_fill_discrete(name = "Supplement",
                      breaks = c("OJ", "VC"),
                      labels = c("Orange Juice", "Ascorbic Acid"))+
  scale_color_brewer(palette = "RdBu") +
  xlab("Dosage (in mg)") + ylab("Tooth length (in mm)") +
  ggtitle("Tooth Length by Dosage and Supplement")
LengthInter
```



This plot reveals that dosage has a significant influence on tooth length and that supplement type is also an influence. In order to understand the data more fully, it will be necessary to run further tests of statistical inference.

Statistical testing

While we may assume that the guinea pigs were randomly assigned to the six groups, thus ensuring independence of measurement, because the number of tests subjects for each condition is less than 30, we must evaluate the data with t distributions.

A t test with unequal variance based on supplement type reveals the following:

```
t_supp <- t.test(len ~ supp, ToothGrowth, var.equal = FALSE)
round(t_supp$conf.int[1:2], 3)
```

```
## [1] -0.171  7.571
```

While, at first, there appears to be a large difference between the means of the two groups, a p-value above 0.05 indicates that this result is not significant. Thus, we fail to reject the null hypothesis. This assessment is supported by the confidence interval.

The exploratory graph suggested that dose level influenced tooth growth. To evaluate whether dose level *significantly* influenced tooth growth in the guinea pigs, we will need to run three separate pairwise t tests.

```
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##     filter, lag
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(broom)
pair_t <- t.test(ToothGrowth$len[ToothGrowth$dose == 2],
       ToothGrowth$len[ToothGrowth$dose == 1]) %>%
  tidy %>%
  mutate(null_hypothesis = "µ2mg - µ1mg = 0") %>%
  select(9, 1:8)

pair_t <- t.test(ToothGrowth$len[ToothGrowth$dose == 2],
                 ToothGrowth$len[ToothGrowth$dose == 0.5]) %>%
  tidy%>%
  mutate(null_hypothesis = "µ2mg - µ0.5mg = 0") %>%
  select(9, 1:8) %>%
  bind_rows(pair_t,.)

pair_t <- t.test(ToothGrowth$len[ToothGrowth$dose == 1],
                 ToothGrowth$len[ToothGrowth$dose == 0.5]) %>%
  tidy%>%
  mutate(null_hypothesis = "µ1mg - µ0.5mg = 0") %>%
  select(9, 1:8) %>%
  bind_rows(pair_t, .)

print.data.frame(pair_t)
```

```
##     null_hypothesis estimate estimate1 estimate2 statistic      p.value
## 1   µ2mg - µ1mg = 0    6.365    26.100    19.735  4.900484 1.906430e-05
## 2 µ2mg - µ0.5mg = 0   15.495    26.100    10.605 11.799046 4.397525e-14
## 3 µ1mg - µ0.5mg = 0    9.130    19.735    10.605  6.476648 1.268301e-07
##   parameter  conf.low conf.high
## 1  37.10109  3.733519  8.996481
## 2  36.88259 12.833833 18.156167
## 3  37.98641  6.276219 11.983781
```

The results of each paired t test reveals a p-value significance well below 0.05, suggesting that the null hypothesis be rejected. This is supported by the confidence intervals for each test. Thus, we are 95% confident that dosage levels influence tooth growth in guinea pigs.

Conclusion

The tests of significance and the confidence intervals associated with each reveal that while dosage level has a significant influence to tooth growth in guinea pigs, there appears to be little difference between orange juice and ascorbic acid supplements. That is, the effective difference in tooth length between the two supplement types was not significantly different from zero.