Analysis of ToothGrowth dataset related to Dosage and Supplements

Jay Yanamandala

October 16, 2021

Overview

I will use 't.test' to analyze confidence intervals and hypothesis tests to compare tooth growth by supp and dose.

Load ToothGrowth data and perform some basic exploratory data analyses

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

A basic summary of the data.

Dimensions of ToothGrowth data set: **60, 3**
Columns names of ToothGrowth data set: **len, supp, dose**

```
str(tg)
```

```
## '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 ...
```

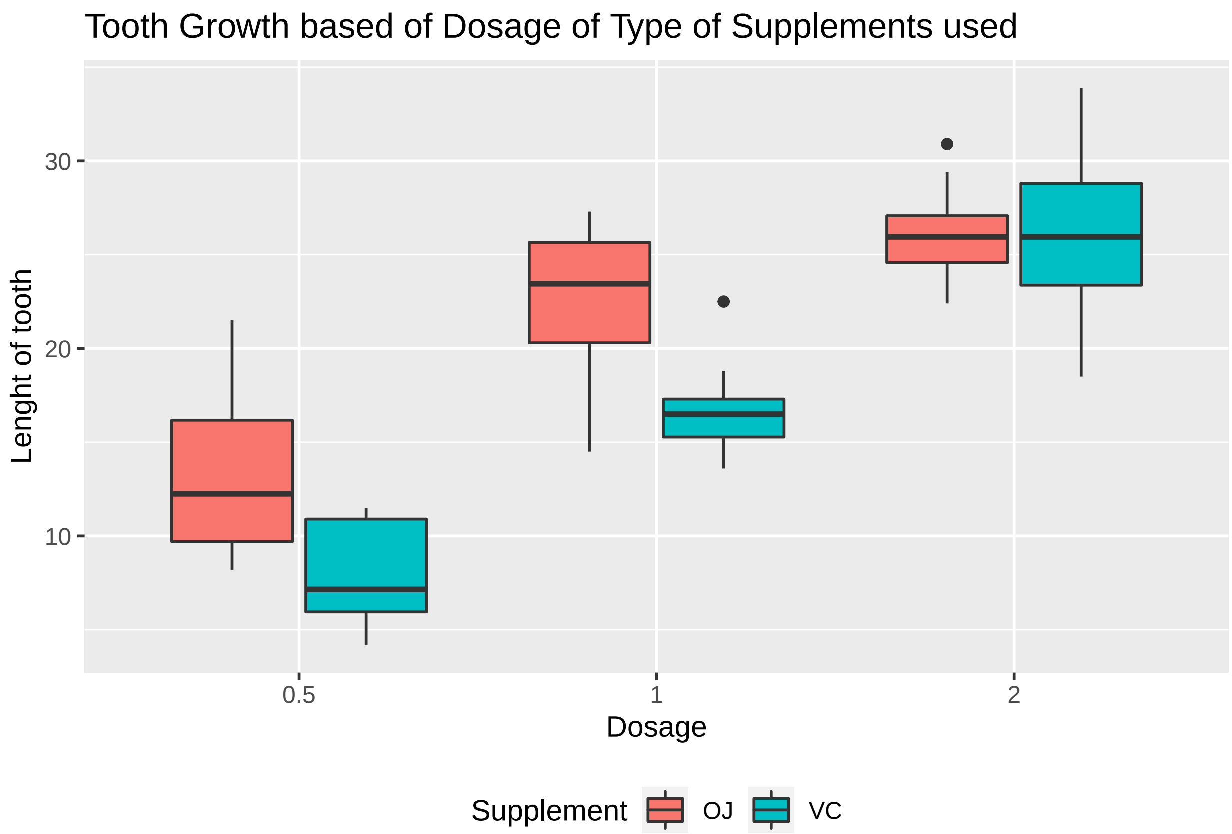
Table Summary for Tooth Growth dose and supplement:

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

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

Tooth Growth - boxplot

```
library(ggplot2)
g <- ggplot(tg, aes(x=factor(dose), y=len, fill=supp))
g + geom_boxplot() +
  ggtitle("Tooth Growth based of Dosage of Type of Supplements used") +
  ylab("Lenght of tooth") + xlab("Dosage") +
  theme(legend.position = "bottom") +
  scale_fill_discrete("Supplement")
```



From the boxplot we can observe that the dosage for 0.5 and 1.0 have bigger differences in length compared to the other dosages. As dosage increases, length of tooth increases. Of the two supplements OJ seems to have more effect on tooth growth than VC

We can look at the mean for the length of tooth growth for each dose and supp.

```
tg %>%
  group_by(dose, supp) %>%
  arrange(as.factor(dose), supp) %>%
  summarize(mean = mean(len)) %>%
  spread(supp, mean) %>%
  mutate("VC-OJ"=abs(VC - OJ))
```

```
## # A tibble: 3 x 4
## # Groups:   dose [3]
```

```
##    dose    OJ    VC `VC-OJ`
##   <dbl> <dbl> <dbl>   <dbl>
## 1   0.5  13.2  7.98  5.25
## 2   1    22.7 16.8   5.93
## 3   2    26.1 26.1   0.0800
```

Observing the differences between supplement OJ and VC, we see that 2mg/day doses has very small differences as compared to dosages; 0.5 and 1. This means it's harder to compare the effectiveness between OJ and VC for 2mg/day.

To test effectiveness between the two supplement types, we will use t.test to find the p-values and confidence intervals to perform hypothesis testing.

t.test Hypothesis Testing

- Our null hypothesis would is there would be no difference between using OJ and VC
- Our alternative hypothesis is there is a difference between sign OJ and VC
- alpha rate is set at 0.05 as standard

For hypothesis test, we will filter the dataset on dosage

```
dosage_half <- tg %>% filter(dose == 0.5)
dosage_one <- tg %>% filter(dose == 1)
dosage_two <- tg %>% filter(dose == 2)
```

sort by 'dose' and then by 'supplement'

```
dosage_half <- dosage_half[order(dosage_one$dose, dosage_one$supp), ]
# dosage_half
```

```
dosage_one <- dosage_one[order(dosage_one$dose, dosage_one$supp), ]
# dosage_one
```

```
dosage_two <- dosage_two[order(dosage_one$dose, dosage_one$supp), ]
# dosage_two
```

When dosage is '1mg/day' when using supplement 'OJ' tooth growth was biggest
When dosage is '2mg/day', the trend starts to reverse -
supplement 'VC' seem to gain on 'OJ" but result is conclusive since we don't have data past two dosages

Some information on Welchs t-test daniellakens blogspot
[From Berkley EDU(https://statistics.berkeley.edu/computing/r-t-tests)

```
### t-test for 0.5 mg/day dose
t.test(len ~ supp, dosage_half)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
```

```
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means between group OJ and group VC is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

```
###  t-test for 1 mg/day dose
t.test(len ~ supp, dosage_one)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means between group OJ and group VC is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

```
###  t-test for 2 mg/day dose
t.test(len ~ supp, dosage_two)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means between group OJ and group VC is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

Assumptions

The central assumption for the results is that the sample is representative of the population, and the variables are IID random variables.
For the t.test, two assumptions are made,
1. The data isn't paired, meaning they're independent
2. The variance are different.

Normally, whenever p-value is less than 0 .05, the null hypothesis has a
significantly low probability of being false.

Thus, for dose 0.5 and 1.0, since p-values are smaller than 0.5, we reject the null hypotheses that the supplement types don't have a difference on tooth growth. But for dose 2.0 mg/day.

With that, in reviewing the t.test, supplement type OC are more effective than VC for doses less than 1.0mg/day. But for dose at 2.0 mg/day, there is no difference between the supplement types.

Conclusion

The p-value associated with the test 'dosage_two' is 0.9638516, so we cannot reject the null hypothesis (H0) of no difference between the (true) averages of the two groups. Since the p-value is greater than the usual significance level alpha = 0.05