Effect of Vitamin C on the Tooth Length of Guinea Pigs

Toni Rib

Overview

In this paper, we will explore the ToothGrowth dataset which is provided with R. This dataset shows the effect of Vitamin C on tooth growth in Guinea Pigs when given various doses of Vitamin C (0.5mg, 1mg, 2mg) with two delivery methods or supplements (orange juice or ascorbic acid). Our analysis shows that higher doses result in more tooth growth, and delivery method is only relevant at the two lower doses.

Load The Data

First, we will load the data from R and look at what kind of variables are contained in the dataset.

```
library(datasets)
data(ToothGrowth)
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 ...
```

```
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
```

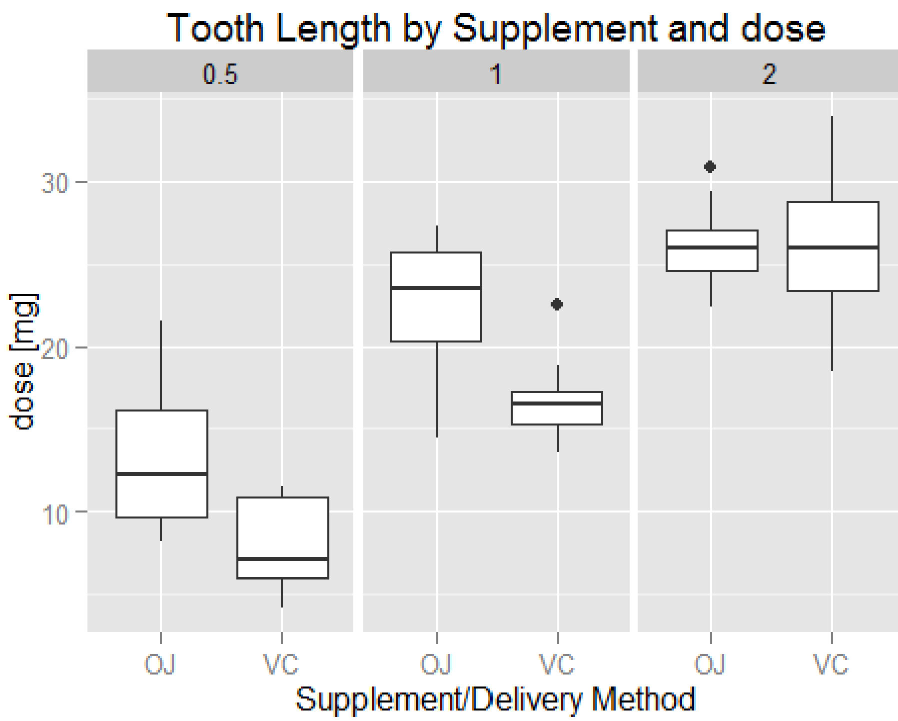
We can see that we have loaded a data frame with 60 observations of 3 variables, two of which are numbers (dose & tooth length) and one of which is a factor (supplement type). The tooth length varies from 4.2 to 33.9, with a mean of 18.81.

Exploratory Plot

It is worthwhile to create a simple plot in order to get an idea of how the data looks and how it might be correlated. To do that, we will use a boxplot showing tooth length by supplement type for each dose.

```
library(ggplot2)
qplot(supp, len, data=ToothGrowth, geom = "boxplot", facets = . ~ dose,
```

```
      main = "Tooth Length by Supplement and dose", xlab =
"Supplement/Delivery Method",
      ylab = "dose [mg]")
```



It appears that for the 0.5mg and 1mg doses, tooth length is longer when orange juice (OJ) is used to deliver the Vitamin C. However, at the 2mg dose, the tooth length average is approximately the same for both supplements, though asorbic acid (VC) does have more variability. Overall, a higher dose results in more tooth growth.

Confidence Intervals

To more accurately compare tooth length by delivery method (supplement) and dose, we will use the t.test function to perform two-sided Welch tests using a confidence level of 95%. Since our data is already split into two levels in the supplement variable, we will use this as the explanatory variable.

First we must split the data up by dose level. We can then perform a t-test on each level. Based on our barplot earlier, we assume that the variances of the two groups are not equal. In order to perform these tests, we assume that the underlying data is both iid and Gaussian.

Because we are performing a two-sample test, our null hypothesis will be that the difference in means [mean(OJ) - mean(VC)] is 0, which equates to the two delivery methods resulting in the same amount of tooth growth. If our results return a very small p-value, we know there is truly a difference between the two delivery methods and we can reject the null hypothesis. If the p-value is close to one, we can assume that the difference in delivery methods is not statistically significant. In this case, we will reject the null hypothesis unless the p-value shows a probability of at least 95% or more for obtaining the test statistic.

```
t0.5 <- t.test(len ~ supp, data = ToothGrowth, subset = ToothGrowth$dose ==
0.5, var.equal = FALSE)
t1.0 <- t.test(len ~ supp, data = ToothGrowth, subset = ToothGrowth$dose ==
1.0, var.equal = FALSE)
t2.0 <- t.test(len ~ supp, data = ToothGrowth, subset = ToothGrowth$dose ==
2.0, var.equal = FALSE)
r1 <- c(t0.5$statistic, t0.5$p.value, t0.5$conf, t0.5$estimate)
r2 <- c(t1.0$statistic, t1.0$p.value, t1.0$conf, t1.0$estimate)
r3 <- c(t2.0$statistic, t2.0$p.value, t2.0$conf, t2.0$estimate)
dat <- data.frame(round(r1, 3), round(r2, 3), round(r3, 3),
                  row.names = c("t-statistic", "p-value", "conf int (low)",
"conf int (high)", "mean (OJ)", "mean (VC)"))
colnames(dat) <- c("dose = 0.5mg", "dose = 1.0mg", "dose = 2.0mg")
print(dat)
```

```
##                 dose = 0.5mg dose = 1.0mg dose = 2.0mg
## t-statistic            3.170        4.033       -0.046
## p-value                0.006        0.001        0.964
## conf int (low)         1.719        2.802       -3.798
## conf int (high)        8.781        9.058        3.638
## mean (OJ)             13.230       22.700       26.060
## mean (VC)              7.980       16.770       26.140
```

Using the table above, we can see the p-values are very low for both the 0.5mg and 1.0mg doses, but almost equal to 1 for the 2.0mg dose.

Conclusions

Based on our testing, we can conclude that the delivery method affects tooth length when using doses of 0.5mg and 1.0mg, but does not affect tooth length using a dose of 2.0mg. At the lower dose levels, the p-values are 0.006 and 0.001, respectively. These values are very low, so we reject the null hypothesis. At the dose level of 2.0mg, the p-value is 0.964, which is above the 95% level, and thus we accept the null hypothesis.