

Tooth Growth Data

Harris Wohl

Brief Overview of the Data

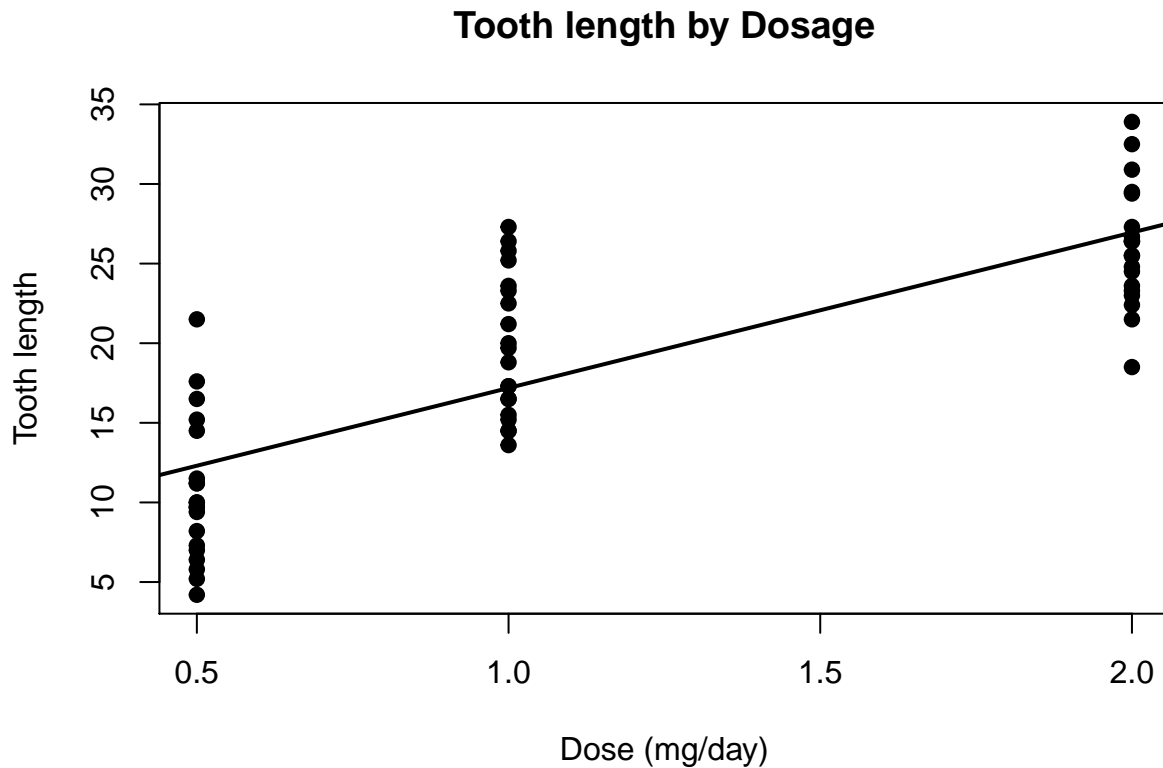
The description of the ToothGrowth dataset in R Documentation is as follows:

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).

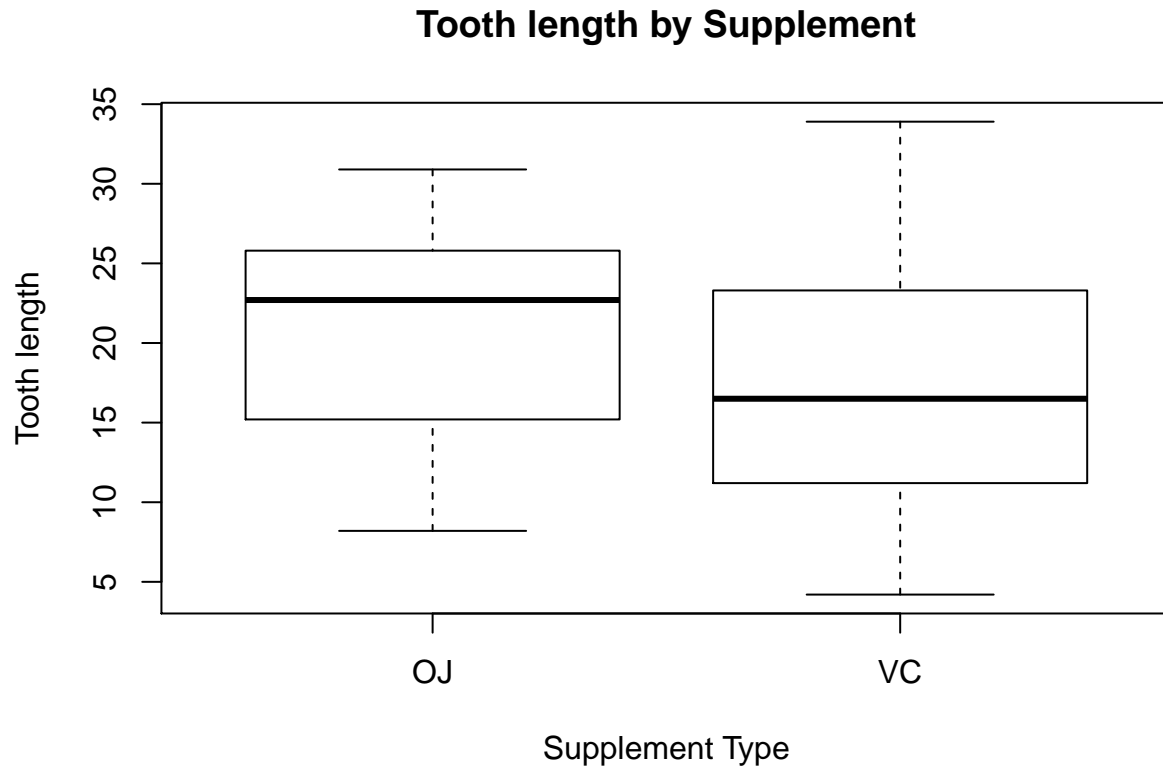
The dataframe included 60 observations among the three variables (tooth length, supplement type, dose in mg/day).

Summary of the Data

The following is a scatterplot with dose on the x axis and tooth length on the y axis. There appears to be a negative correlation between dose and tooth length, but the difference will need to be tested formally.



The following boxplot represents the difference in tooth length by supplement type.



Using the following code, the sample mean and sample standard deviation for each relevant group is produced. The first two tables show the sample mean for each supplement type and dosage, and the latter two tables show the sample standard deviations.

```
suppmean <- aggregate(tooth, by = list(tooth$supp), FUN = mean)
dosemean <- aggregate(tooth, by = list(tooth$dose), FUN = mean)
```

```
suppsd <- aggregate(tooth, by = list(tooth$supp), FUN = sd)
dosesd <- aggregate(tooth, by = list(tooth$dose), FUN = sd)
```

```
suppmean[, 1:2]
```

```
##   Group.1    len
## 1      OJ 20.66333
## 2      VC 16.96333
```

```
dosemean[, 1:2]
```

```
##   Group.1    len
## 1    0.5 10.605
## 2    1.0 19.735
## 3    2.0 26.100
```

```
suppsd[, 1:2]
```

```
##   Group.1      len
## 1      OJ 6.605561
## 2      VC 8.266029
```

```
dosesd[, 1:2]
```

```
##   Group.1      len
## 1      0.5 4.499763
## 2      1.0 4.415436
## 3      2.0 3.774150
```

Hypothesis Tests for the Difference in Tooth Length

Using the following code, the dataset was split into supplement types (creating a list of two dataframes) and dose levels (creating a list of three dataframes).

```
suppsplit <- split(tooth, tooth$supp)
dosesplit <- split(tooth, tooth$dose)
```

First, I performed a t-test to test the true difference in mean tooth length by supplement type. For the test, I chose a significance level of .05 and a two-sided alternative, as well as an assumption of unequal variances since the sample variance of VC was more than twice as large as the sample variance for OJ. Since the variances were assumed to not be equal, a Welch t-test was performed.

```
suppsplit <- split(tooth, tooth$supp)
dosesplit <- split(tooth, tooth$dose)
```

```
#t tests with assumption of equal variance
```

```
oj <- suppsplit[[1]]$len
vc <- suppsplit[[2]]$len
```

```
t.test(oj, vc, paired = FALSE, var.equal = F)
```

```
##
## Welch Two Sample t-test
##
## data:  oj and vc
## 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
```

The resulting p-value of .06063 is greater than .05, so we fail to reject the null hypothesis and conclude that there is not a significant difference in mean tooth length by supplement type.

Since there were three dosage levels, three t-tests were performed. First, to test the difference in tooth level between a dose of 2 mg and .5mg , second to test the difference in tooth level between a dose of .5 mg and 1 mg, and third to do the same but with the doses 1 mg and 2 mg. All tests assumed a two-sided alternative and equal variances, and were performed at a .05 level of significance.

```
pointfive <- dosesplit[[1]]$len
one <- dosesplit[[2]]$len
two <- dosesplit[[3]]$len

t.test(pointfive, two, paired = F, var.equal = T)

##
## Two Sample t-test
##
## data: pointfive and two
## t = -11.799, df = 38, p-value = 2.838e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -18.15352 -12.83648
## sample estimates:
## mean of x mean of y
## 10.605 26.100

t.test(pointfive, one, paired = F, var.equal = T)

##
## Two Sample t-test
##
## data: pointfive and one
## t = -6.4766, df = 38, p-value = 1.266e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -11.983748 -6.276252
## sample estimates:
## mean of x mean of y
## 10.605 19.735

t.test(one, two, paired = F, var.equal = T)

##
## Two Sample t-test
##
## data: one and two
## t = -4.9005, df = 38, p-value = 1.811e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -8.994387 -3.735613
## sample estimates:
## mean of x mean of y
## 19.735 26.100
```

The p-values for all three tests were much smaller than .05, so we reject the null for all three and conclude that there is a significant difference in tooth length between all three dosage levels.