Tooth Growth Data

Brianne Calandra

April 26, 2015

Overview

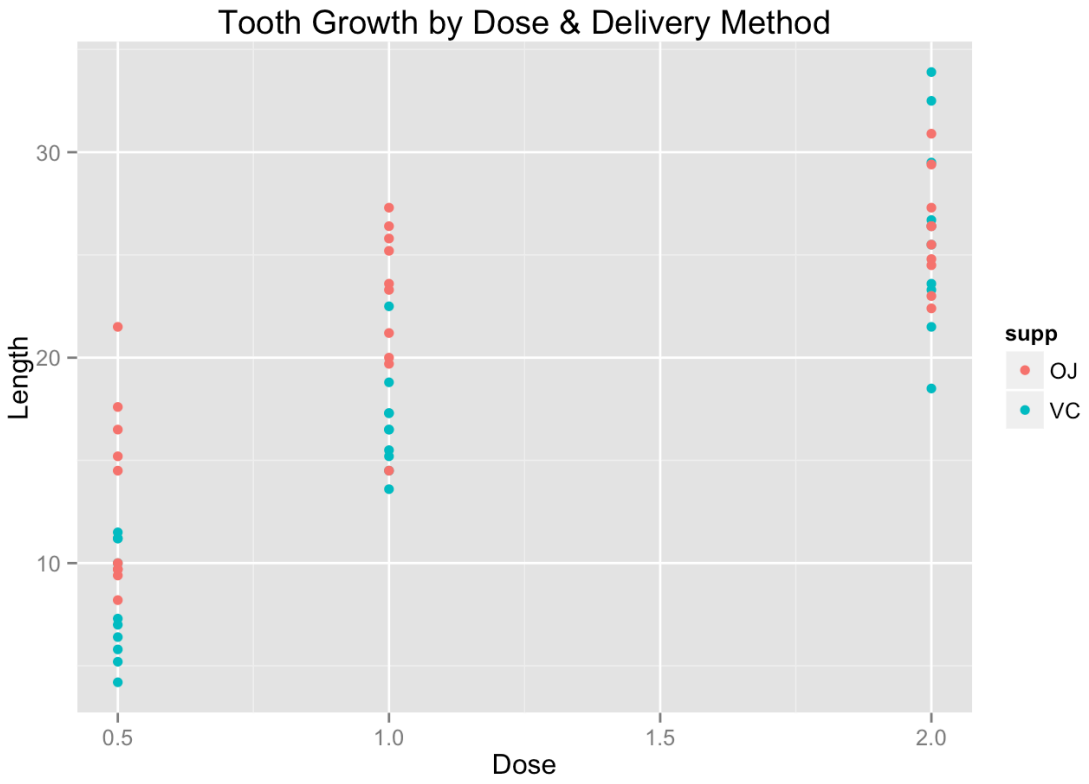
This report will investigate the ToothGrowth data set through exploratory data analysis and a hypothesis test.

Please note, instead of separating code and diagrams into the appendix, I chose to use the “knit and weave” approach of mixing my explanations and diagrams. However, in total this report should meet the page maximum (less than 6 total).

Exploratory Analysis

Here we load the data and do some exploratory plotting.

```
data(ToothGrowth)
library(ggplot2)
qplot(dose,len,data=ToothGrowth,color=supp)+labs(title = "Tooth Growth b
y Dose & Delivery Method", x="Dose",y="Length")
```



The plot above shows that there are three levels of dosage: .5, 1 and 2. There are two delivery methods: OJ and Vitamin C. The length of tooth growth seems to be related to the dosage - as the dosage goes up, so does the observed lengths.

Next, let's look at the data a little differently and calculate the mean length for each delivery method (OJ vs VC).

```
oj <- subset(ToothGrowth,supp=="OJ")
mean(oj$len)
```

```
## [1] 20.66333
```

```
vc <- subset(ToothGrowth,supp=="VC")
mean(vc$len)
```

```
## [1] 16.96333
```

There is a difference in mean, which leads us next into the hypothesis test.

Hypothesis Test

Our hypothesis test is as follows:

H0 - There is no statistically significant difference in tooth length based on delivery method.

HA - There is a statistically significant difference in tooth length based on delivery method.

We explore this using a T-test. First we check to see if the variance in these two groups is different:

```
var(oj$len)
```

```
## [1] 43.63344
```

```
var(vc$len)
```

```
## [1] 68.32723
```

Since variance is different, we use the t.test function with the following arguments:

```
t.test(len ~ supp, paired=FALSE,var.equal=FALSE,data=ToothGrowth)
```

```
##
##  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
```

The resulting T-test has a p-value of .06063, which is more than our significance level of .05, so we fail to reject the null hypothesis for this case. The confidence interval is (-0.171, 7.7571) and zero is contained within this interval. Therefore, this also agrees with our failure to reject the null hypothesis.

Conclusion and Assumptions

Based on the above analysis, we fail to reject our null hypothesis which states that there is no statistically significant difference in tooth length based on delivery method.

The T-test uses the T-interval, which assumes that the data are idd normal. The T-intervals is also used in situations where n is small, which is representative of the data we are working with here.