Statistical Inference Course Project

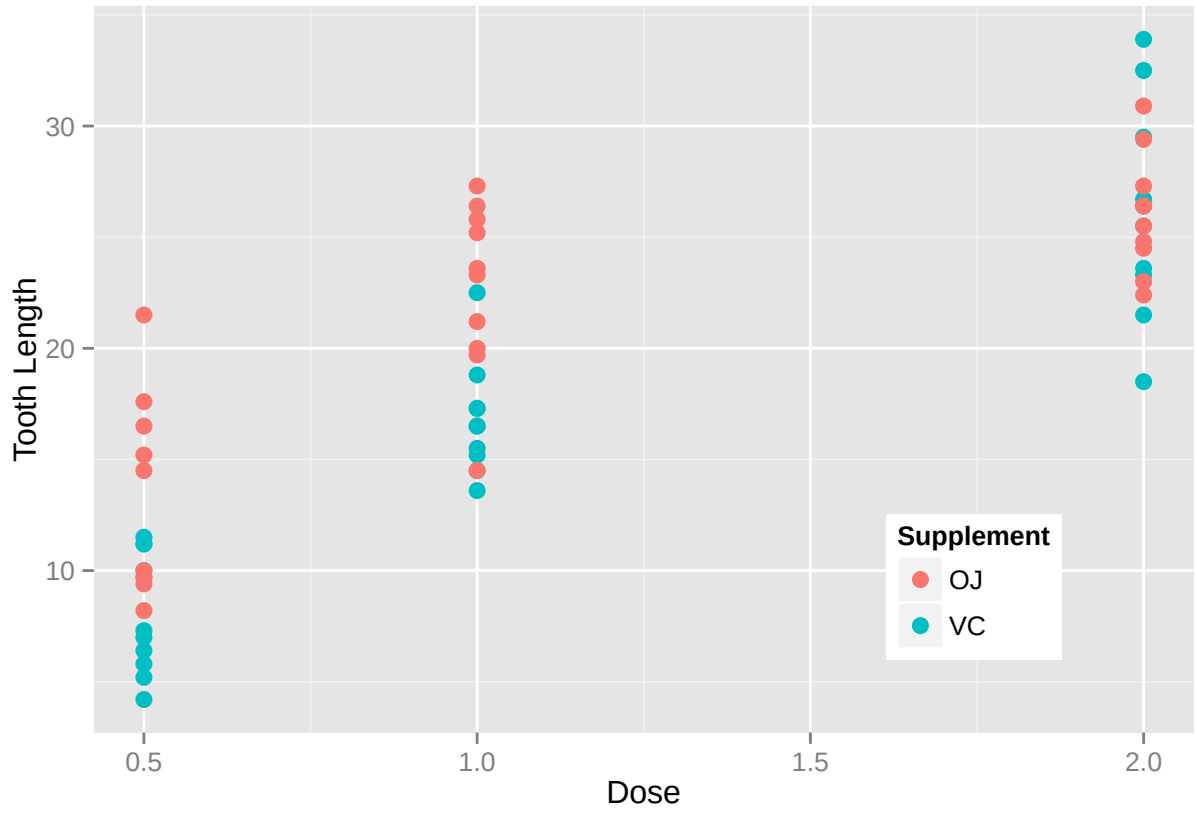
Statistical Analysis of the the Tooth Growth Database in R

Author: Gaurav Bansal

In this exercise, we'll examine the ToothGrowth database in the Datasets package in R. We'll develop a 95% confidence interval to compare tooth growth by supp and dose.

Let's begin by loading the data and plotting tooth length by dose for each supplement.

```
library(datasets)
library(ggplot2)
data(ToothGrowth)
ggplot(ToothGrowth, aes(x=dose, y=len, color=supp)) +
  geom_point(size=3) +
  labs(x="Dose", y="Tooth Length", color="Supplement") +
  theme(legend.position=c(0.8, 0.2))
```



Now let's perform some T tests on the data to see the effects of supplement and dose. To do this, we need to break up the data into groups arranged by supp and dose.

```
vc5 <- subset(ToothGrowth, supp=="VC" & dose==0.5)
vc1 <- subset(ToothGrowth, supp=="VC" & dose==1)
```

```r
vc2 <- subset(ToothGrowth, supp=="VC" & dose==2)
oj5 <- subset(ToothGrowth, supp=="OJ" & dose==0.5)
oj1 <- subset(ToothGrowth, supp=="OJ" & dose==1)
oj2 <- subset(ToothGrowth, supp=="OJ" & dose==2)
```

Now let's run T tests on the three groups to see differences in tooth length between ascorbic acid (vc) and orange juice (oj) at different doses. We will assume the data are NOT paired.

```r
t.test(vc5$len, oj5$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  vc5$len and oj5$len
## t = -3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.780943 -1.719057
## sample estimates:
## mean of x mean of y
##      7.98     13.23
```

```r
t.test(vc1$len, oj1$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  vc1$len and oj1$len
## t = -4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -9.057852 -2.802148
## sample estimates:
## mean of x mean of y
##     16.77     22.70
```

```r
t.test(vc2$len, oj2$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  vc2$len and oj2$len
## t = 0.0461, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.63807  3.79807
## sample estimates:
## mean of x mean of y
##     26.14     26.06
```

From the T tests shown above, we see that at the lower two doses–0.5 and 1.0–the supplement makes a significant impact. This is because the 95 percent confidence interval does not contain 0, which means we reject the null hypothesis in favor of the alternative hypothesis. However, at the 2.0 dose, this does not hold true. In this case we fail to reject the null hypothesis.

Let's take a look at the impact on tooth growth of dose alone, regardless of supplement.

```
half <- subset(ToothGrowth, dose==0.5)
one <- subset(ToothGrowth, dose==1)
two <- subset(ToothGrowth, dose==2)
t.test(one$len, half$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  one$len and half$len
## t = 6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   6.276219 11.983781
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

```
t.test(two$len, one$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  two$len and one$len
## t = 4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.733519 8.996481
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

Here it's clear that increasing the dose from 0.5 to 1 and then from 1 to 2 makes a significant impact on tooth growth, regardless of supplement, because the 95 percent confidence interval for both T tests does not contain 0. This means we accept the alternative hypothesis that the difference in means is not equal to zero. The very low p-value also confirms that we reject the null hypothesis in favor of the alternative hypothesis.