Statistical Inference Course Project: Part 2

Emma Ideal

June 13, 2015

Overview

In this analysis, we will investigate the effect of Vitamin C supplements on tooth growth in guinea pigs. Sixty guinea pigs were randomly assigned a supplement type (VC or OJ) and given one of three doses (0.5mg, 1mg, or 2mg). Their tooth lengths were then recorded; we will analyze differences in these tooth lengths as a function of the supplement type and dosage.

Load the Data and Basic Exploration

First load the data in R:

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

We can look at the first several rows of the data frame:

```
head(ToothGrowth)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

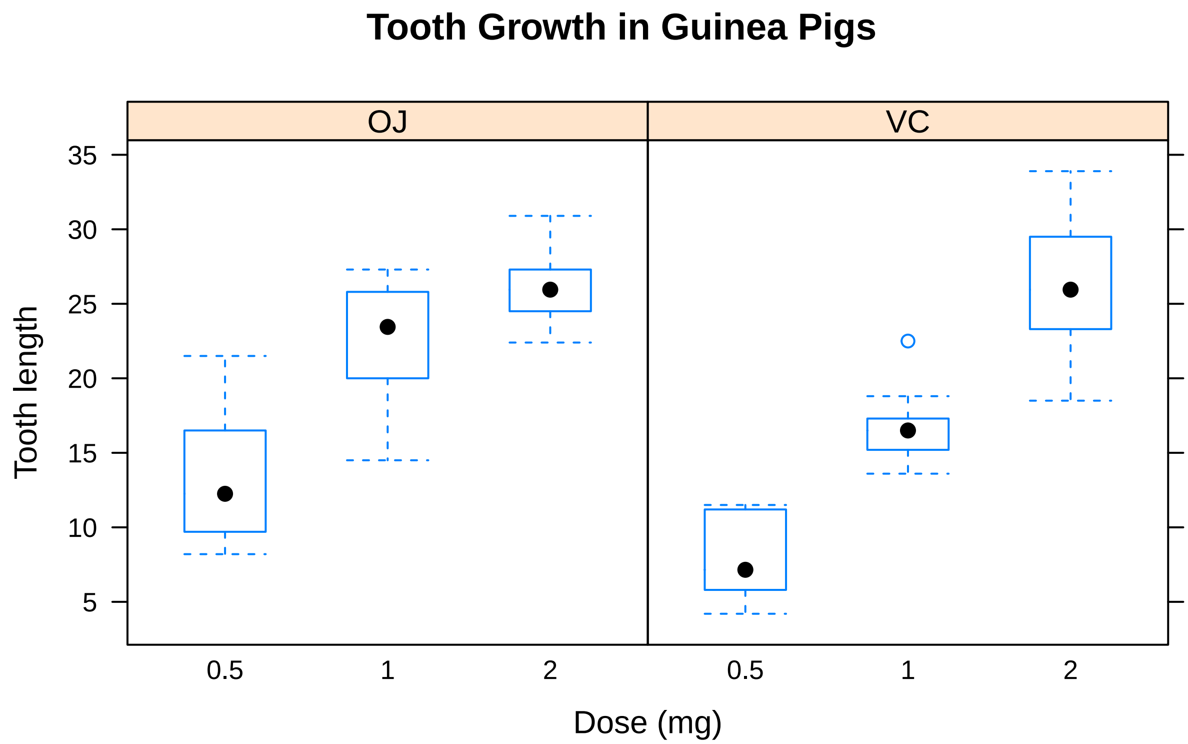
A simple summary can be found using the **summary** function:

```
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  
```

A boxplot of the data can be made:

```
library(lattice)
bwplot(len ~ as.factor(dose) | supp, data = ToothGrowth,
       main = 'Tooth Growth in Guinea Pigs', xlab = 'Dose (mg)', ylab = 'Tooth length')
```



We can see from the boxplot that larger doses lead to more tooth growth. This is true for both the VC and OJ supplements; however, for smaller doses, it seems the OJ group sees larger tooth growth than the VC group.

Summary of the Data

We can summarize this data by computing the means and standard deviations of the tooth lengths by type of supplement (VC or OJ) and by dose size (0.5, 1, or 2 mg).

```
library(plyr)
ddply(ToothGrowth, .(supp, dose), summarize, means = mean(len), sd = round(sd(len),2))
```

```
##   supp dose means   sd
## 1   OJ  0.5 13.23 4.46
## 2   OJ  1.0 22.70 3.91
## 3   OJ  2.0 26.06 2.66
## 4   VC  0.5  7.98 2.75
## 5   VC  1.0 16.77 2.52
## 6   VC  2.0 26.14 4.80
```

Confidence Intervals

(Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose)

First, we run a **T test** to test the null hypothesis H_0: the mean of the tooth growth from OJ is the same as the mean of the tooth growth from VC (two-sided test). Here, we assume the VC and OJ groups are properly randomized independent samples.

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

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1670064  7.5670064
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

The 95% confidence interval (narrowly) contains 0, and therefore, we cannot reject the null hypothesis.

We now want to test the following hypotheses H_0: the mean of the tooth growth from i dose (0.5, 1, 2 mg) is equal to the mean of the tooth growth from j dose (for all possible combinations i and j where $i \neq j$).

```r
# Subset the data by dose
subset_051 <- subset(ToothGrowth, dose %in% c(0.5, 1))
subset_052 <- subset(ToothGrowth, dose %in% c(0.5, 2))
subset_12  <- subset(ToothGrowth, dose %in% c(1,2))

# Compute the p-values
pval_051 <- t.test(len ~ dose, data=subset_051, var.equal = TRUE)$p.value
pval_052 <- t.test(len ~ dose, data=subset_052, var.equal = TRUE)$p.value
pval_12  <- t.test(len ~ dose, data=subset_12,  var.equal = TRUE)$p.value

# Compute the 95% confidence intervals for the differences in the means
int_051 <- round(t.test(len ~ dose, data=subset_051, var.equal = TRUE)$conf.int, 2)
int_052 <- round(t.test(len ~ dose, data=subset_052, var.equal = TRUE)$conf.int, 2)
int_12  <- round(t.test(len ~ dose, data=subset_12,  var.equal = TRUE)$conf.int, 2)
```

With some formatting, we get the following results:

```
##   compared doses(mg)           95% CI     p values
## 1          (0.5, 2)  [-11.98,-6.28] 1.266297e-07
## 2          (0.5, 1) [-18.15,-12.84] 2.837553e-14
## 3            (1, 2)   [-8.99,-3.74] 1.810829e-05
```

At the levels specified by the p values, we can rule out the null hypothesis for all three cases. Since all confidence intervals are entirely below 0, this suggests that higher doses lead to more tooth growth.

Conclusions and Assumptions

We can conclude from this study that higher doses of OJ or VC lead to greater tooth growth. At 95% confidence, it is inconclusive whether the OJ supplement leads to greater tooth growth compared with VC, however, it is likely (we would need to check) that at 90% confidence, we can conclude OJ is better than VC. In this analysis, we assume that the guinea pigs are sufficiently randomized in their assignment to VC or OJ. This justifies the assumption of equal population variances in the two samples. We also assume normality of the underlying population data, which allows us to use T tests to compute confidence intervals.