Coursera Statistical Inference Course Project: Part 2

Brad Tweardy

Saturday, March 21, 2015

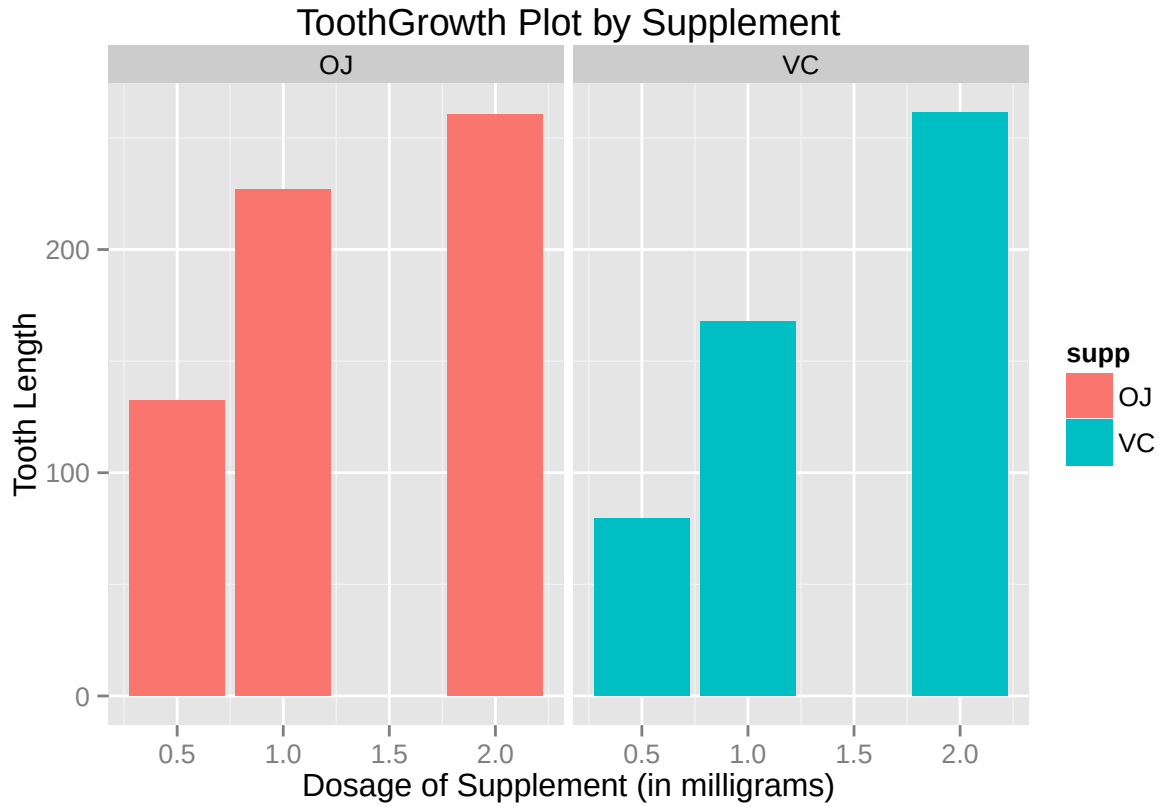
Introduction

In this project we will analyze the ToothGrowth data in the R datasets package.

Basic Analysis

Below we'll load the dataset and provide a basic analyis of it.

```
data(ToothGrowth)
ggplot(data = ToothGrowth, aes(x = dose, y = len, fill = supp))+
        geom_bar(stat = "identity") +
        facet_grid(.~supp) +
        xlab("Dosage of Supplement (in milligrams)") +
        ylab("Tooth Length") +
        ggtitle("ToothGrowth Plot by Supplement")
```



we can deduct from this data that a higher dosage of either supplement correlates to longer teeth.

Basic Summary

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
summary(ToothGrowth)
```

```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

```
table(ToothGrowth$supp, ToothGrowth$dose)
```

```
##
##      0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

We see that there are three dosage amounts and two supplement types. We're also able to verify that the tests are distributed equally by supplement and dosage.

Comparison of Tooth Growth by Supplement and Dosage

```
VC_t <- t.test(len~supp, paired = FALSE, var.equal = TRUE, data = ToothGrowth)
OJ_t <- t.test(len~supp, paired = FALSE, Var.equal = FALSE, data = ToothGrowth)
comp_supp <- data.frame("p-value" = c(VC_t$p.value, OJ_t$p.value),
                        "Conf-Low"  = c(VC_t$conf[1], OJ_t$conf[1]),
                        "Conf-High" = c(VC_t$conf[2], OJ_t$conf[2]),
                        row.names = c("Equal Variance","Unequal Variance"))
comp_supp
```

```
##                     p.value   Conf.Low Conf.High
## Equal Variance   0.06039337 -0.1670064  7.567006
## Unequal Variance 0.06063451 -0.1710156  7.571016
```

We see that the p value of Equal and Unequal variables are higher than five percent. We also see that Confidence Intervals in both t tests contain zero. With this data we cannot ignore the null hypothesis tests.

```
dose_.5 <- ToothGrowth[which(ToothGrowth$dose ==.5), 1]
dose_1.0 <- ToothGrowth[which(ToothGrowth$dose ==1), 1]
dose_2.0 <- ToothGrowth[which(ToothGrowth$dose ==2), 1]
Tcomp_.5_1.0 <- t.test(dose_.5, dose_1.0, paired = F, var.equal = T)
Fcomp_.5_1.0 <- t.test(dose_.5, dose_1.0, paired = F, var.equal = F)
Result_comp_.5_1.0 <- data.frame("p-value" = c(Tcomp_.5_1.0$p.value,
                                               Fcomp_.5_1.0$p.value),
```

```
                                  "Conf-Low" = c(Tcomp_.5_1.0$conf[1],
                                                 Fcomp_.5_1.0$conf[1]),
                                  "Conf-High"=c(Tcomp_.5_1.0$conf[2],
                                                Fcomp_.5_1.0$conf[2]),
                                  row.names=c("Equal Variance",
                                              "Unequal Variance"),
                                  "Dose" = "0.5 to 1")
Tdose_.5_2.0 <- t.test(dose_.5, dose_2.0, paired = F, var.equal = T)
Fdose_.5_2.0 <- t.test(dose_.5, dose_2.0, paired = F, var.equal = F)
Result_comp_.5_2.0 <- data.frame("p-value" = c(Tdose_.5_2.0$p.value,
                                               Fdose_.5_2.0$p.value),
                                  "Conf-Low"=c(Tdose_.5_2.0$conf[1],
                                               Fdose_.5_2.0$conf[1]),
                                  "Conf-High"=c(Tdose_.5_2.0$conf[2],
                                                Fdose_.5_2.0$conf[2]),
                                  row.names=c("Equal Var","Unequal Var"),
                                  "Dose"="0.5 to 2")
Tdose_1.0_2.0 <- t.test(dose_1.0, dose_2.0, paired=F, var.equal=T)
Fdose_1.0_2.0 <- t.test(dose_1.0, dose_2.0, paired=F, var.equal=F)
Result_comp_1.0_2.0 <- data.frame("p-value" = c(Tdose_1.0_2.0$p.value,
                                                Fdose_1.0_2.0$p.value),
                                   "Conf-Low"=c(Tdose_1.0_2.0$conf[1],
                                                Fdose_1.0_2.0$conf[1]),
                                   "Conf-High"=c(Tdose_1.0_2.0$conf[2],
                                                 Fdose_1.0_2.0$conf[2]),
                                   row.names=c("Equal Var","Unequal Var"),
                                   "Dose"="1 to 2")
Result_Dose <- rbind(Result_comp_.5_1.0,Result_comp_.5_2.0,Result_comp_1.0_2.0)
Result_Dose
```

```
##                       p.value   Conf.Low  Conf.High     Dose
## Equal Variance   1.266297e-07 -11.983748  -6.276252 0.5 to 1
## Unequal Variance 1.268301e-07 -11.983781  -6.276219 0.5 to 1
## Equal Var        2.837553e-14 -18.153519 -12.836481 0.5 to 2
## Unequal Var      4.397525e-14 -18.156167 -12.833833 0.5 to 2
## Equal Var1       1.810829e-05  -8.994387  -3.735613   1 to 2
## Unequal Var1     1.906430e-05  -8.996481  -3.733519   1 to 2
```

As we can see from the data above, we can confidently conclude that dosage levels have a impact on tooth growth. We can say that a higher dosage of either supplement will result in an increase in tooth growth. In supporting these findings, we can see in all performed tests that the p values are relatively low and the confidence intervals do not contain zeros.

Final Analysis

First, in order for this analysis to be accurate we must assume a few pieces of information that was not given. In general, we have to assume that the overall health, diet, age and weight is relatively equal in all test subjects.

In conclusion, we can say for certain that the dosage amount effects the amount of tooth growth. Higher dosage normally results in higher length of tooth growth. What we cannot say is that one supplement over the other results in more tooth growth.