Statistical Inference Part 2

Nazmi Anik

Friday, October 23, 2015

Overview

We are going to analyze the ToothGrowth data in the R datasets package. A basic summary of the data will be provided. We will use confidence intervals and/or hypothesis tests to compare tooth growth by supplement (vitamin c and ascorbic acid) and dose.

Load & Explore the Data

We will load the data from the R package dataset.

```
#load the data & look at a quick summary
data(ToothGrowth); str(ToothGrowth)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

Now let's try to visualize the data with some boxplots. That will enable us to see the mean and variance in tooth growth with different dosages of each supplement.

```
g <- ggplot(data = ToothGrowth, aes(x = factor(dose), y = len)) + geom_boxplot() + facet_grid(.~supp) +
    xlab("Dose") + ylab("Tooth Length") + ggtitle("Tooth Growth with\nDifferent Supplements")
```

Here's another boxplot to see the mean and variance in tooth growth with different supplements for each dosage.

```
h <- ggplot(data = ToothGrowth, aes(x = factor(supp), y = len)) + geom_boxplot() + facet_grid(.~dose) +
    xlab("Supplement") + ylab("Tooth Length") + ggtitle("Tooth Growth with\n Different Dosages")
#Plot them together
plot_grid(g, h, align='h', labels=c('A', 'B'))
```

A
Tooth Growth with Different Supplements

B
Tooth Growth with Different Dosages
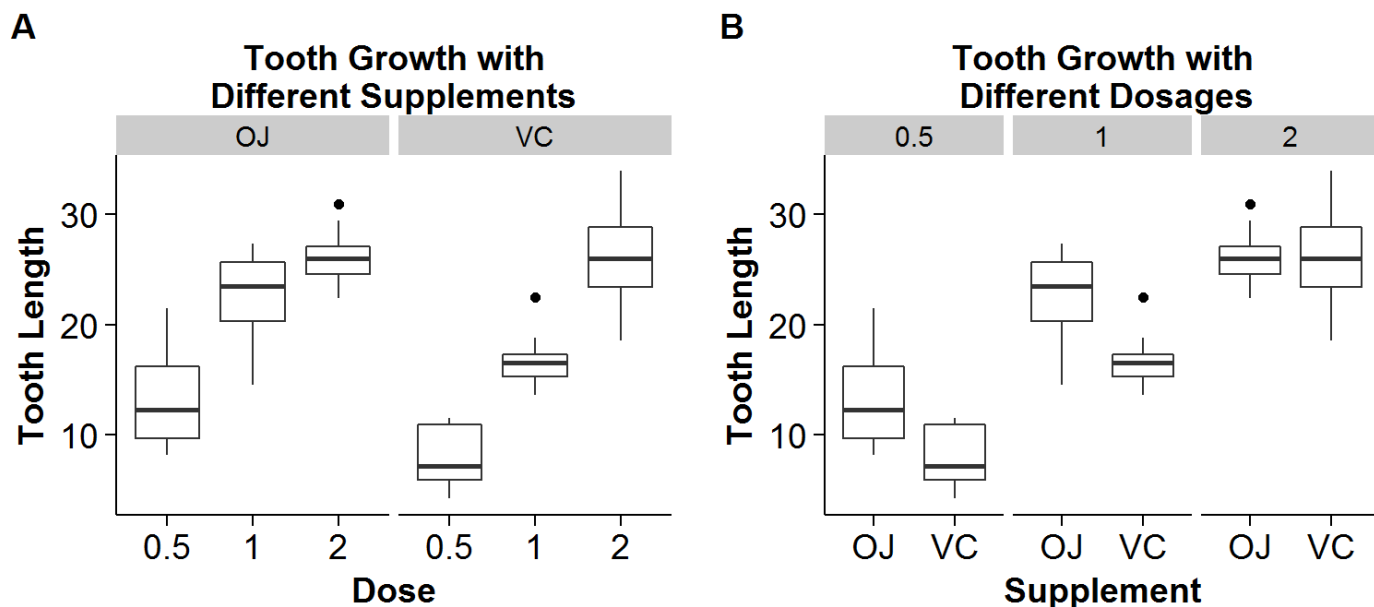
It is clear from the plots that higher dosage of the supplements cause more tooth growth. The effects of using different supplements is less clear as their effects seems to be similar.

Basic Summary of Data

Let's take a basic look at the data.

```
#group the data by dose and then by supplement
growth <- group_by(ToothGrowth, dose, supp)
#add mean and standard deviation to the summary table
growthSummary <- summarize(growth, mean = mean(len), standard.deviation = sd(len)); growt
hSummary
```

```
## Source: local data frame [6 x 4]
## Groups: dose
##
##   dose supp  mean standard.deviation
## 1  0.5   OJ 13.23           4.459709
## 2  0.5   VC  7.98           2.746634
## 3  1.0   OJ 22.70           3.910953
## 4  1.0   VC 16.77           2.515309
## 5  2.0   OJ 26.06           2.655058
## 6  2.0   VC 26.14           4.797731
```

Tooth Growth Comparison

Assumptions

We cannot use paired t tests because we are assuming that the pigs were chosen randomly and therefore are independent from different samples. Therefore, paired = FALSE for all the t tests.
We are assuming that the variances between separate groups of pigs are not equal. Therefore, var.equal = FALSE for all the t tests.

Comparison

Based on our assumptions, let's test if using ascorbic acid (VC) or vitamin c (OJ) gives us a better result in tooth length:

```
tsupp <- t.test(len~supp,paired=FALSE,var.equal=FALSE,data=growth)
tsuppDF <- data.frame("p-value"=tsupp$p.value, "Conf-Low"=tsupp$conf[1], "Conf-High"=tsup
p$conf[2], "Mean Len OJ"=tsupp$estimate[1], "Mean Len VC"=tsupp$estimate[2], row.name
s=c("OJ vs VC")); tsuppDF
```

```
##              p.value   Conf.Low Conf.High Mean.Len.OJ Mean.Len.VC
## OJ vs VC 0.06063451 -0.1710156  7.571016    20.66333    16.96333
```

The confidence interval includes zero. Therefore difference between OJ and VC being zero cannot be ruled out.
Now let's see if effectiveness of OJ and VC in tooth growth are different with different dosages:

```
#low dose
growthLow <- subset(ToothGrowth, dose == 0.5)
t1 <- t.test(len~supp,paired=FALSE,var.equal=FALSE,data=growthLow)
#medium dose
growthMed <- subset(ToothGrowth, dose == 1.0)
t2 <- t.test(len~supp,paired=FALSE,var.equal=FALSE,data=growthMed)
#high dose
growthHi<- subset(ToothGrowth, dose == 2.0)
t3 <- t.test(len~supp,paired=FALSE,var.equal=FALSE,data=growthHi)
#compare
tcompare <- data.frame("p-value"=c(t1$p.value, t2$p.value, t3$p.value), "Conf-Low"=c(t1$c
onf[1],t2$conf[1],t3$conf[1]), "Conf-High"=c(t1$conf[2],t2$conf[2],t3$conf[2]), "Mean Len
OJ"=c(t1$estimate[1],t2$estimate[1],t3$estimate[1]), "Mean Len VC"=c(t1$estimate[2],t2$es
timate[2],t3$estimate[2]), row.names=c("Low Dose","Medium Dose","High Dose" )); tcompare
```

```
##                  p.value  Conf.Low Conf.High Mean.Len.OJ Mean.Len.VC
## Low Dose    0.006358607  1.719057  8.780943       13.23        7.98
## Medium Dose 0.001038376  2.802148  9.057852       22.70       16.77
## High Dose   0.963851589 -3.798070  3.638070       26.06       26.14
```

We can see that in low and medium doses, OJ is more effective. In the high dose, both OJ and VC have similar effectiveness.

Conclusions

As we have seen from the boxplots, increasing the dosage of either ascorbic acid or vitamin c increased tooth growth in guinea pigs. Based on out t tests, we verified that in low and medium doses vitamin c is more effective, and in high dosage both vitamin c and ascorbic acid show similar effectiveness.