Statistical Inference Project - Part 2

Boris Badmaev

August 21, 2015

Introduction

This report uses the `ToothGrowth` data in the R datasets package. The goal of this analysis is to compare tooth growth by supp and dose. We will perform basic exploratory data analysis, provide summary of the data and use confidence intervals and/or hypothesis tests in order to come to our conclusions.

Data

The `Toothdata` dataset shows the effect of Vitamin C on tooth growth in guinea pigs.The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

The `Toothdata` data frame contains 60 observations on 3 variables:
`len` - Tooth length (numeric)
`supp` - Supplement type (VC or OJ)
`dose` - Dose in milligrams

Data Source: C. I. Bliss (1952) The Statistics of Bioassay. Academic Press
References: McNeil, D. R. (1977) Interactive Data Analysis. New York: Wiley.

Data Processing and Analysis

```
library(datasets)
data(ToothGrowth)
attach(ToothGrowth)
head(ToothGrowth, 15)
```

```
##     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
## 7  11.2   VC  0.5
## 8  11.2   VC  0.5
## 9   5.2   VC  0.5
## 10  7.0   VC  0.5
## 11 16.5   VC  1.0
## 12 16.5   VC  1.0
## 13 15.2   VC  1.0
## 14 17.3   VC  1.0
## 15 22.5   VC  1.0
```

Let's explore the data frame

```
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 ...
```

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
table(ToothGrowth$supp, ToothGrowth$dose)
```

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

```
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
```

```
mean(ToothGrowth$len)
```

```
## [1] 18.81333
```

```
sd(ToothGrowth$len)
```

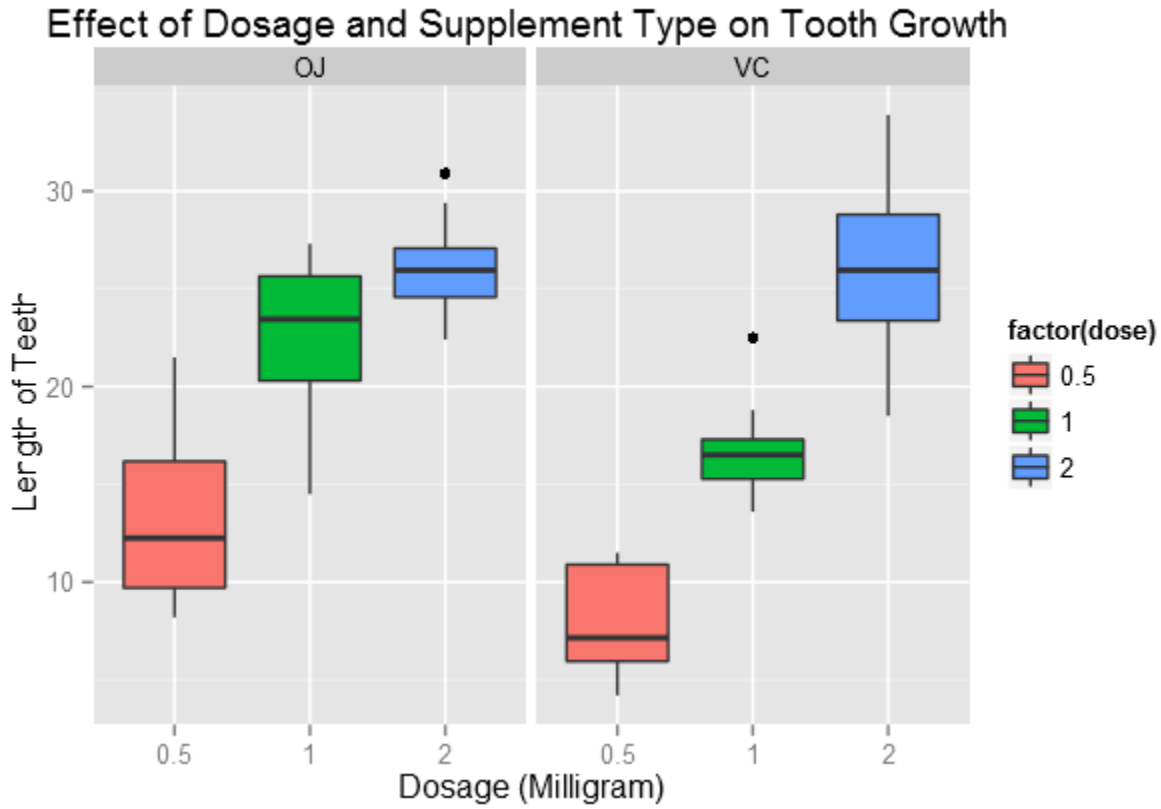
```
## [1] 7.649315
```

Let's do some plotting

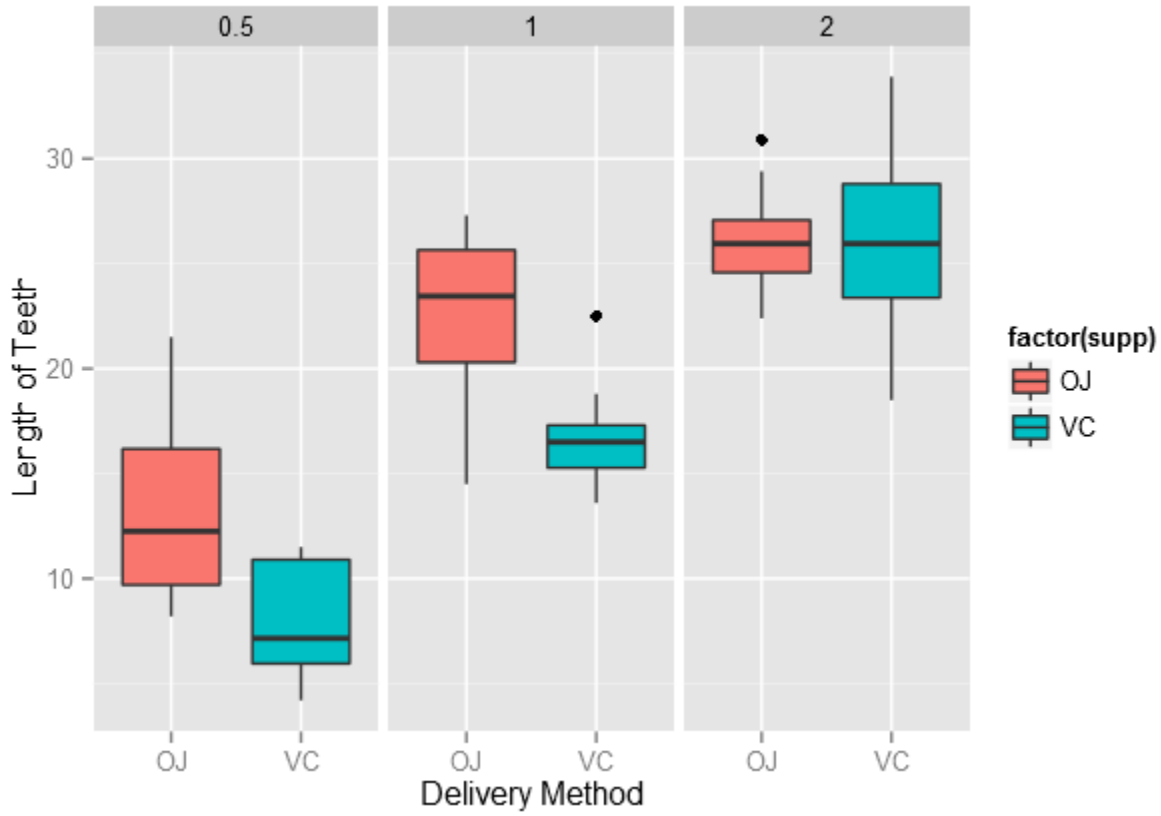
```
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
g <- ggplot(ToothGrowth, aes(x=factor(dose),y=len,fill=factor(dose)))+
            geom_boxplot(notch=FALSE) + facet_grid(.~supp) +
            scale_x_discrete("Dosage (Milligram)") +
            scale_y_continuous("Length of Teeth") +
            ggtitle("Effect of Dosage and Supplement Type on Tooth Growth ")
print(g)
```




```
q <- ggplot(ToothGrowth, aes(x=factor(supp),y=len,fill=factor(supp)))+
            geom_boxplot(notch=FALSE) + facet_grid(.~dose) +
            scale_x_discrete("Delivery Method") +
            scale_y_continuous("Length of Teeth")
print(q)
```



As we can see from these plots it seems like dosage has greater impact on tooth growth than the delivery method. Let's verify our hypotheses.

We will use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose

```
# T test for supplement by dose level
dose0.5 <- subset(ToothGrowth, dose == 0.5)
dose1.0 <- subset(ToothGrowth, dose == 1.0)
dose2.0 <- subset(ToothGrowth, dose == 2.0)

supp0.5<-t.test(len ~ supp, paired=FALSE, var.equal=TRUE, data = dose0.5)
supp1.0<-t.test(len ~ supp, paired=FALSE, var.equal=TRUE, data = dose1.0)
supp2.0<-t.test(len ~ supp, paired=FALSE, var.equal=TRUE, data = dose2.0)

supp.result <- data.frame("p-value"=c(supp0.5$p.value, supp1.0$p.value, supp2.0$p.value),

"Conf-Low"=c(supp0.5[["conf.int"]][1],supp1.0[["conf.int"]][1],supp2.0[["conf.int"]][1]),
"Conf-High"=c(supp0.5[["conf.int"]][2],supp1.0[["conf.int"]][2], supp2.0[["conf.int"]][2]),
row.names=c("Dose 0.5","Dose 1.0", "Dose 2.0"))

supp.result
```

```
##               p.value  Conf.Low Conf.High
## Dose 0.5 0.0053036613  1.770262  8.729738
## Dose 1.0 0.0007807262  2.840692  9.019308
## Dose 2.0 0.9637097790 -3.722999  3.562999
```

With assumptions that observations are not paired and different delivery methods have equal variance the results show small P-values (<0.01) and entirely positive confidence intervals for small dosages (0.5 and 1), that means we can reject Ho hypothesis. In other words changing of delivery method from ascorbic acid (VC) to orange juice (OJ) does change (increases) the average tooth length if we use small dosages (0.5 or 1). For dosage = 2.0 we observe large P-value and the confidence interval containing zero. That means we can't reject Ho hypothesis which means that for dosage = 2.o there is no change in tooth length when moving from VC to OJ delivery methods.

Now let's test the Ho hypothesis to see if changing the dosage will impact the tooth growth within each group(VC and OJ)

```
# subsetting based on supp and dosage change
VC.dose0.5_1.0 <- subset(ToothGrowth, dose %in% c(0.5, 1.0) & supp=="VC")
VC.dose0.5_2.0 <- subset(ToothGrowth, dose %in% c(0.5, 2.0) & supp=="VC")
VC.dose1.0_2.0 <- subset(ToothGrowth, dose %in% c(1.0, 2.0) & supp=="VC")

OJ.dose0.5_1.0 <- subset(ToothGrowth, dose %in% c(0.5, 1.0) & supp=="OJ")
OJ.dose0.5_2.0 <- subset(ToothGrowth, dose %in% c(0.5, 2.0) & supp=="OJ")
OJ.dose1.0_2.0 <- subset(ToothGrowth, dose %in% c(1.0, 2.0) & supp=="OJ")

VCtest0.5_1.0<-t.test(len ~ dose, paired=FALSE, var.equal=TRUE, data = VC.dose0.5_1.0)
VCtest0.5_2.0<-t.test(len ~ dose, paired=FALSE, var.equal=TRUE, data = VC.dose0.5_2.0)
VCtest1.0_2.0<-t.test(len ~ dose, paired=FALSE, var.equal=TRUE, data = VC.dose1.0_2.0)

OJtest0.5_1.0<-t.test(len ~ dose, paired=FALSE, var.equal=TRUE, data = OJ.dose0.5_1.0)
OJtest0.5_2.0<-t.test(len ~ dose, paired=FALSE, var.equal=TRUE, data = OJ.dose0.5_2.0)
OJtest1.0_2.0<-t.test(len ~ dose, paired=FALSE, var.equal=TRUE, data = OJ.dose1.0_2.0)

VCdose.result <- data.frame("p-value"=c(VCtest0.5_1.0$p.value,
                                        VCtest0.5_2.0$p.value,

                         "Conf-Low"=c(VCtest0.5_1.0[["conf.int"]][1],
                                      VCtest0.5_1.0[["conf.int"]][1],
                                      VCtest1.0_2.0[["conf.int"]][1]),

                         "Conf-High"=c(VCtest0.5_1.0[["conf.int"]][2],
                                       VCtest0.5_1.0[["conf.int"]][2],
                                       VCtest1.0_2.0[["conf.int"]][2]),

                        row.names=c("VC 0.5-1.0","VC 0.5-2.0", "VC 1.0-2.0"))

OJdose.result <- data.frame("p-value"=
                                    c(OJtest0.5_1.0$p.value,
                                      OJtest0.5_2.0$p.value,
                                      OJtest1.0_2.0$p.value),

                         "Conf-Low"=c(OJtest0.5_1.0[["conf.int"]][1],
                                      OJtest0.5_1.0[["conf.int"]][1],
                                      OJtest1.0_2.0[["conf.int"]][1]),

                         "Conf-High"=c(OJtest0.5_1.0[["conf.int"]][2],
                                       OJtest0.5_1.0[["conf.int"]][2],
                                       OJtest1.0_2.0[["conf.int"]][2]),

                        row.names=c("OJ 0.5-1.0","OJ 0.5-2.0", "OJ 1.0-2.0") )
```

```
VCdose.result
```

```
##                  p.value  Conf.Low Conf.High
## VC 0.5-1.0 6.492265e-07 -11.26435 -6.315654
## VC 0.5-2.0 4.957286e-09 -11.26435 -6.315654
## VC 1.0-2.0 3.397578e-05 -12.96896 -5.771040
```

```
OJdose.result
```

```
##                  p.value   Conf.Low  Conf.High
## OJ 0.5-1.0 8.357559e-05 -13.410814 -5.5291857
## OJ 0.5-2.0 3.401859e-07 -13.410814 -5.5291857
## OJ 1.0-2.0 3.736280e-02  -6.500502 -0.2194983
```

With the same assumptions that observations are not paired and equal variances we see consistently very low p-values and confidence intervals not containing zero across both groups (OJ and VC) and across all possible dosage changes (0.5-1.0, 0.5-2.0 and 1.0-2.0). That allow us to reject Ho hypothesis and confirm the fact that changing of dosage (using VC or OJ delivery method) will likely lead to change in tooth length (increased dosage leads to increased tooth length).