Statistical Inference Part2 Tooth Growth

rachelxi

27 April 2015

Loading Data and Basic Analysis

Firstly, we output the names of the data

```
data(ToothGrowth)
names(ToothGrowth)
```

```
## [1] "len"  "supp" "dose"
```

Then the first 10 lines of the data

```
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
```

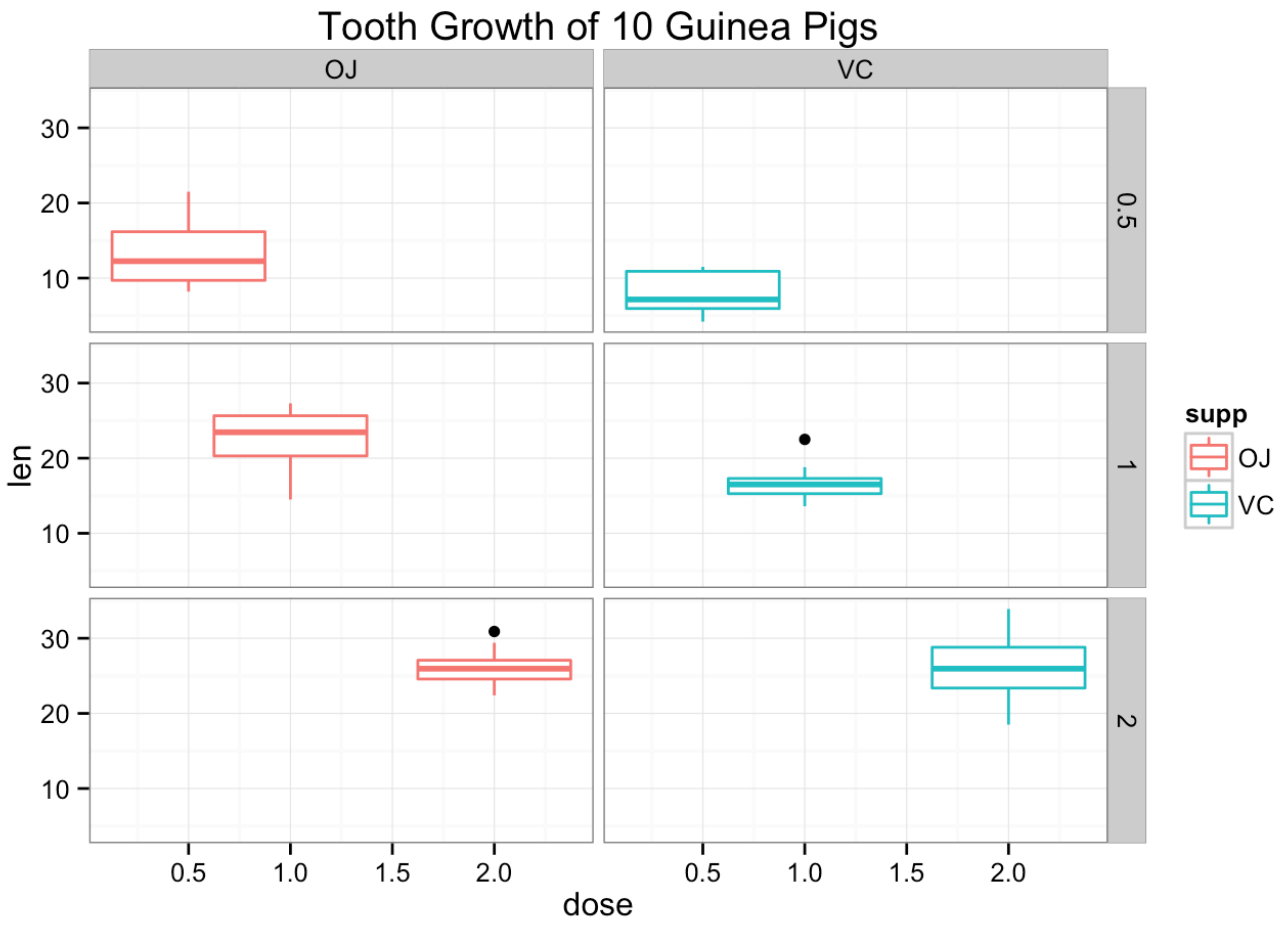
And also the size of the data (60 observations, 3 variables)

```
dim(ToothGrowth)
```

```
## [1] 60  3
```

Also, by looking at the help files, we know the data describes: 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). Then we can carry exploratory analysis to the data with plot:

```
library(ggplot2)
p <- ggplot(ToothGrowth,aes(x=dose,y=len,colour=supp))+geom_boxplot()
p+theme_bw()+facet_grid(dose~supp)+labs(title="Tooth Growth of 10 Guinea Pigs")
```



Summary of Data

We can perform the summary of data by separating supplement type and dose, calculating the mean, variance and standard deviation respectively:

```
library(plyr)
ddply(ToothGrowth,.(supp,dose),summarize,mean=mean(len),var=var(len),sd=sd(le
n))
```

```
##   supp dose  mean       var       sd
## 1   OJ  0.5 13.23 19.889000 4.459709
## 2   OJ  1.0 22.70 15.295556 3.910953
## 3   OJ  2.0 26.06  7.049333 2.655058
## 4   VC  0.5  7.98  7.544000 2.746634
## 5   VC  1.0 16.77  6.326778 2.515309
## 6   VC  2.0 26.14 23.018222 4.797731
```

Hypothesis Test and Confidence Level (95%)

Before we conduct the t test, it is essetial to process the data first

```
inter <- split(ToothGrowth,ToothGrowth$dose)
inter2 <- split(ToothGrowth,ToothGrowth$supp)
suppOJ <- inter2[[1]]
suppVC <- inter2[[2]]
```

Then first t test is about length to supplement with dose=0.5mg

```
t.test(len~supp,data=inter[[1]],paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

Then second t test is about length to supplement with dose=1.0mg

```
t.test(len~supp,data=inter[[2]],paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

Then third t test is about length to supplement with dose=2.0mg

```
t.test(len~supp,data=inter[[3]],paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

The fourth t test is about length to dose with supplement=OJ(orange juice)

```
t.test(suppOJ$dose,suppOJ$len,paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  suppOJ$dose and suppOJ$len
## t = -16.0923, df = 29.535, p-value = 3.699e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -21.97262 -17.02071
## sample estimates:
## mean of x mean of y
##  1.166667 20.663333
```

The fifth t test is about length to dose with supplement=VC(ascorbic acid)

```
t.test(suppVC$dose,suppVC$len,paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  suppVC$dose and suppVC$len
## t = -10.4365, df = 29.341, p-value = 2.168e-11
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.89076 -12.70257
## sample estimates:
## mean of x mean of y
##  1.166667 16.963333
```

Assumptions and Conclusions

The t tests are using default settings because I assume there is no paired relations between the tested data; confidence interval is calculated with “two.sided” method; the variances of two are assumed to be non-equal.

For different supplements, the pigs given orange juice seem to have a higher average in tooth length than that given ascorbic acid (under same dose quantity). Besides, with the same supplement for vitamin C, the more doses pig has taken, the longer tooth it seems to have.

The last part is about p values. Then third t test is about length to supplement with dose=2.0mg, and calculated p value is large, which indicating hypothesis would be accepted. For other four test, however, the null hypotheses are true.