ToothGrowth Data Report

Xiaowen Wang

May 23, 2015

Overview

In this report we will use a dataset about ToothGrowth in R package. WE will do the following actions with this dataset:

1. Perform basic exploratory data analyses
2. Provide a basic summary of the data. 3 Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.
3. State assumptions and conclusion.

Load the Data

```
data(ToothGrowth)
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
```

```
names(ToothGrowth)
```

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

```
dim(ToothGrowth)
```

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

We know there are three columns of data, they're "len","supp" and "dose". There are 60 observations in this dataset.

Basic Exploratory Data Analysis about this dataset

First, according to the two VC delivery method, we want to subset the dataset.

```
vc <- subset(ToothGrowth,supp=="VC")
oj <- subset(ToothGrowth,supp=="OJ")
```

we want to look at a brief summary about the data.

```
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
```

```
summary(vc)
```
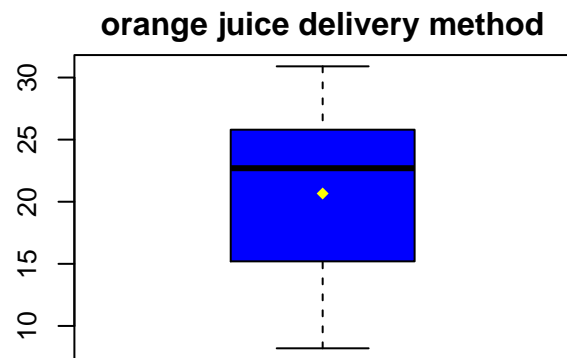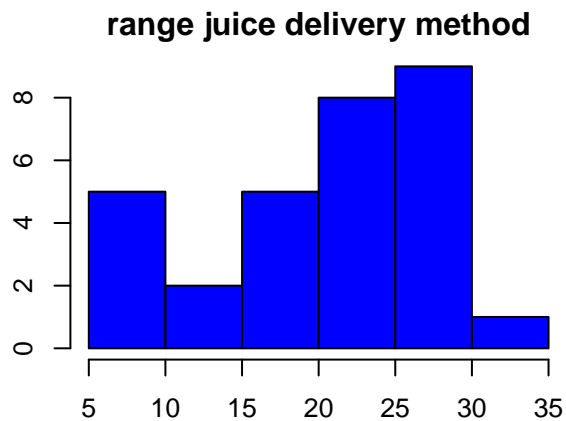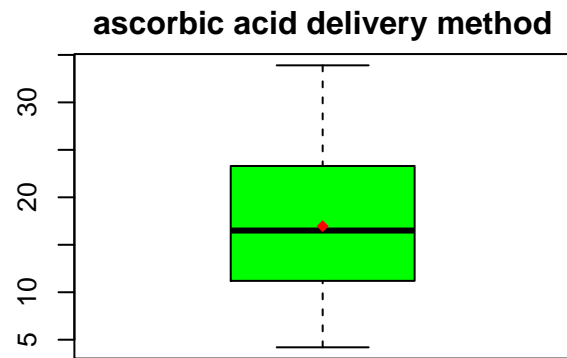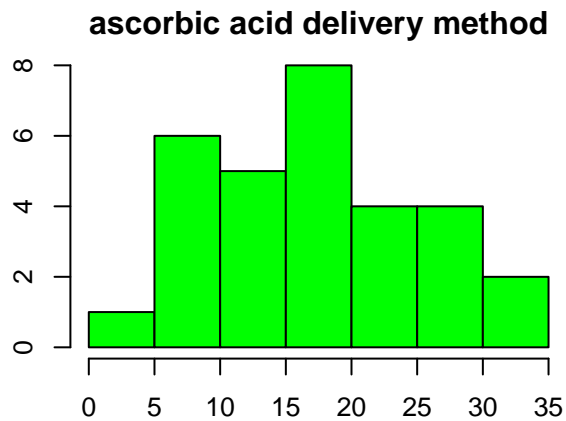
```
##       len          supp         dose
##  Min.   : 4.20   OJ: 0   Min.   :0.500
##  1st Qu.:11.20   VC:30   1st Qu.:0.500
##  Median :16.50           Median :1.000
##  Mean   :16.96           Mean   :1.167
##  3rd Qu.:23.10           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

```
summary(oj)
```

```
##       len          supp         dose
##  Min.   : 8.20   OJ:30   Min.   :0.500
##  1st Qu.:15.53   VC: 0   1st Qu.:0.500
##  Median :22.70           Median :1.000
##  Mean   :20.66           Mean   :1.167
##  3rd Qu.:25.73           3rd Qu.:2.000
##  Max.   :30.90           Max.   :2.000
```

Second,we want to know how the numeric Tooth length distributed and what's sample mean(lb) and standard deviation(s) of len. We will draw a histgram and box plot to have a basic idea.

```
par(mfrow=c(2,2),mar=c(2,2,2,2))
hist(vc$len,main="ascorbic acid delivery method",col="green")
boxplot(vc$len,main="ascorbic acid delivery method",col="green")
points(mean(vc$len),col="red",pch=18)
hist(oj$len,main="range juice delivery method",col="blue")
boxplot(oj$len,main="orange juice delivery method",col="blue")
points(mean(oj$len),col="yellow",pch=18)
```

ascorbic acid delivery method (histogram, top-left)

ascorbic acid delivery method (boxplot, top-right)

range juice delivery method (histogram, bottom-left)

orange juice delivery method (boxplot, bottom-right)

```r
lb1 = mean(vc$len)
s1 = sd(vc$len)
lb2 = mean(oj$len)
s2 = sd(oj$len)
```

According to the boxplot, we can see the length of tooth are quite balanced distributed on the two sides of mean value, but it's hard to tell if it's a normal distribution. So we try to do a two sample t test, to see if length is normal distributed with mean of 18.81,standard deviation of 7.65.

```r
dd1 <- rnorm(30,mean=lb1,sd=s1)
t.test(vc$len,dd1,paired=FALSE,var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  vc$len and dd1
## t = 0.406, df = 58, p-value = 0.6862
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.466443  5.230399
## sample estimates:
## mean of x mean of y
##  16.96333  16.08136
```

```
dd2 <- rnorm(30,mean=lb2,sd=s2)
t.test(oj$len,dd2,paired=FALSE,var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  oj$len and dd2
## t = 1.2137, df = 58, p-value = 0.2298
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.420074  5.794309
## sample estimates:
## mean of x mean of y
##  20.66333  18.47622
```

According to the result, we have 95% condifence to say length is a normal distribution in both VC delevery method experiment group.

Third, we want see if there's any relationship between length and dose. x axis is the dose, y axis is the length.

```
par(mfrow=c(2,1),mar=c(2,2,2,2))
plot(y=vc$len,x=vc$dose,col="green",pch=18,main="ascorbic acid",xlab="dose(mg)",ylab="tooth length")
plot(y=oj$len,x=oj$dose,col="blue",pch=18,main="orange juice",xlab="dose(mg)",ylab="tooth length")
```
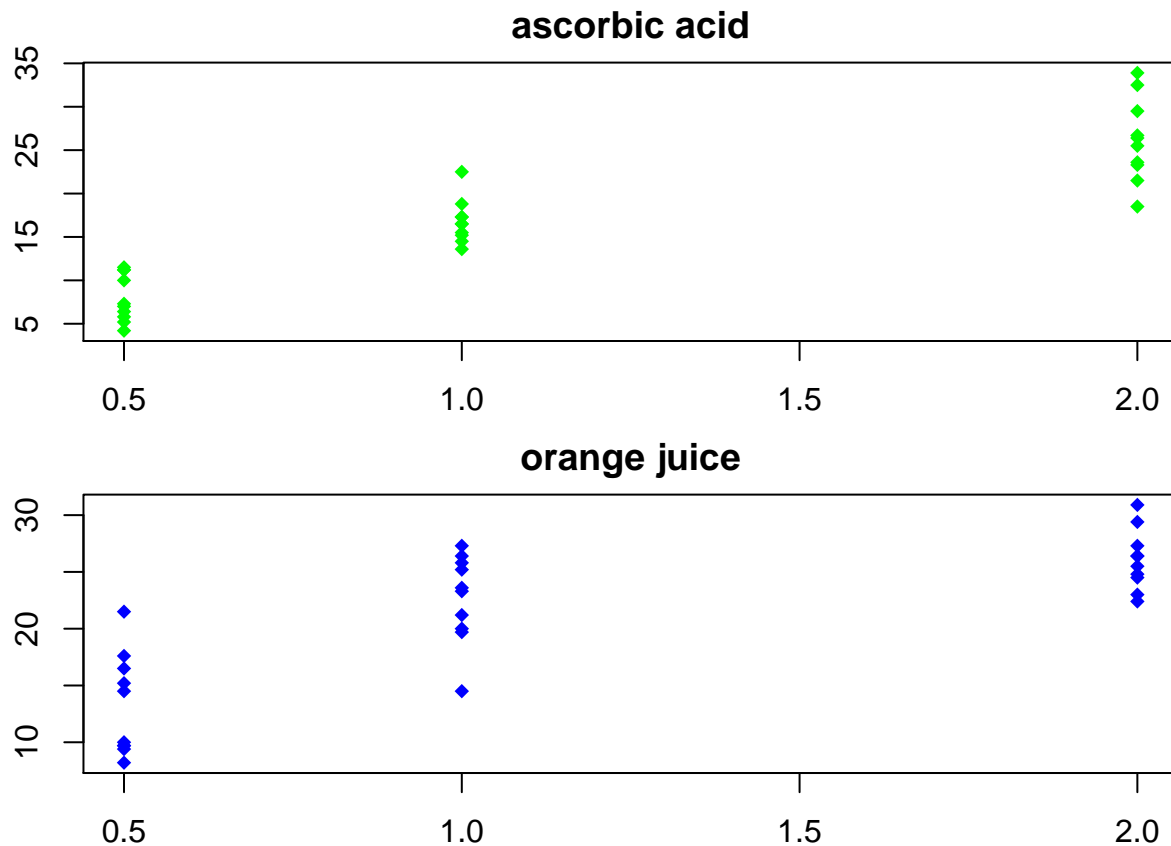

According to the graph, we can see as the dose increse, the tooth length increase.

Forth, we want to see how the two group length compare to each other.

```
par(mfrow=c(1,1))
plot(y=ToothGrowth$len,x=ToothGrowth$supp,col="yellow")
```
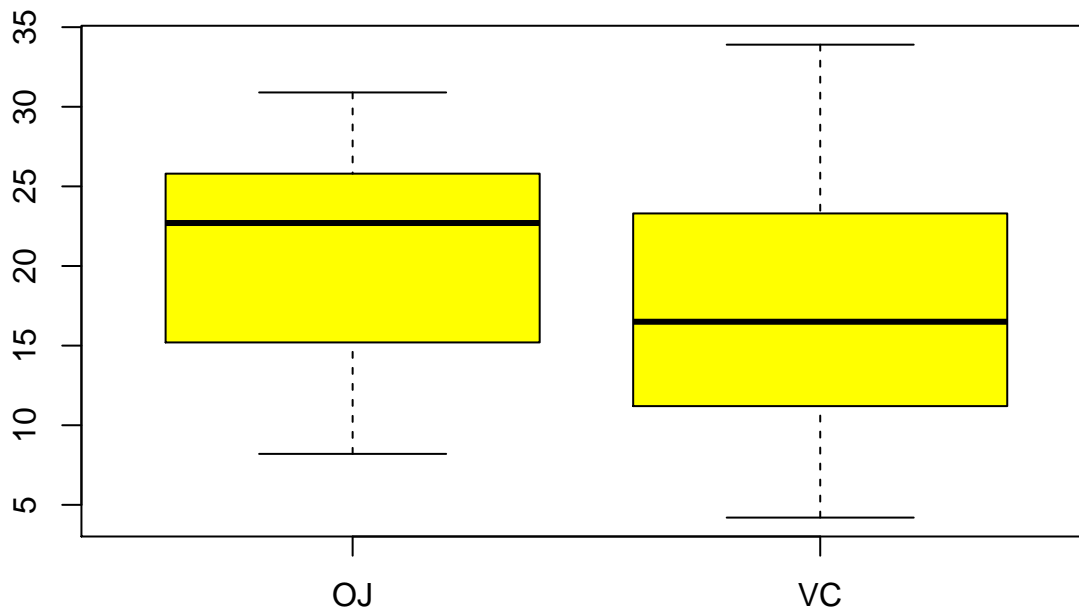

We can guess orange juice has more significant affect than ascorbic acid method to help tooth grow longer.

Summay and assumptions after exploratory analysis

1. The tooth length of either groups is a 95% confidence normal distribution.
2. Give more mg of dose help the tooth grow longer.
3. Give orange juice has more significant affects on helping the tooth grow longer.

Use Hypothesis Test

Compare Tooth Growth by Supp

Null Hypothesis is that different supply makes no difference on tooth growth. Ho: lb2 = lb1 Ha: lb2 > lb1

To prove our assumption that orange juice make tooth longer, we need to reject the null hypothesis and prove lb1 is larger than lb2.

```
t.test(oj$len,vc$len,paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  oj$len and vc$len
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
```

```
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

According to the result, we can't reject the null hypothesis, which means there's no significant increase in tooth length when we use orange juice.

Compare Tooth Growth by Dose.

Null Hypothesis is that different mg of dose makes no difference on tooth growth. We need to compare 3 times: 0.5mg to 1.0mg; 1.0mg to 2.0mg; 0.5mg to 2.0mg. We subset the dataset into 3 groups.

```
g1 <- subset(ToothGrowth,dose==0.5)
g2 <- subset(ToothGrowth,dose==1.0)
g3 <- subset(ToothGrowth,dose==2.0)
```

1. First comparison(0.5mg to 1.0mg of dose) Ho: mu1 = mu2 Ha: mu1 < mu2

```
t.test(g1$len,g2$len,paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  g1$len and g2$len
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean of x mean of y
##    10.605    19.735
```

According to the result , as $p < 0.05$, we can reject the null hypothesis, that's say there is significant different between these two groups. Compared to 0.5mg, 1.0mg of dose makes the tooth grow longer.

2. Second comparison(1.0mg to 2.0mg of dose) Ho: mu2 = mu3 Ha: mu2 < mu3

```
t.test(g2$len,g3$len,paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  g2$len and g3$len
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```

According to the result , as p < 0.05, we can reject the null hypothesis, that's say there is significant different between these two groups. Compared to 1.0mg, 2.0mg of dose makes the tooth grow longer.

2. Third comparison(0.5mg to 2.0mg of dose) Ho: mu1 = mu3 Ha: mu1 < mu3

```
t.test(g1$len,g3$len,paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  g1$len and g3$len
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean of x mean of y
##    10.605    26.100
```

According to the result , as p < 0.05, we can reject the null hypothesis, that's say there is significant different between these two groups. Compared to 0.5mg, 2.0mg of dose makes the tooth grow longer.

Conclusion

1. Giving orange juice or ascorbic acid as Vitamin C source doesn't make significant difference on the tooth growth.
2. Giving more dose of Vitamin C makes the tooth grow longer.