Statistical Inference Course Project Part 2

Pulan Yu

2/29/2020

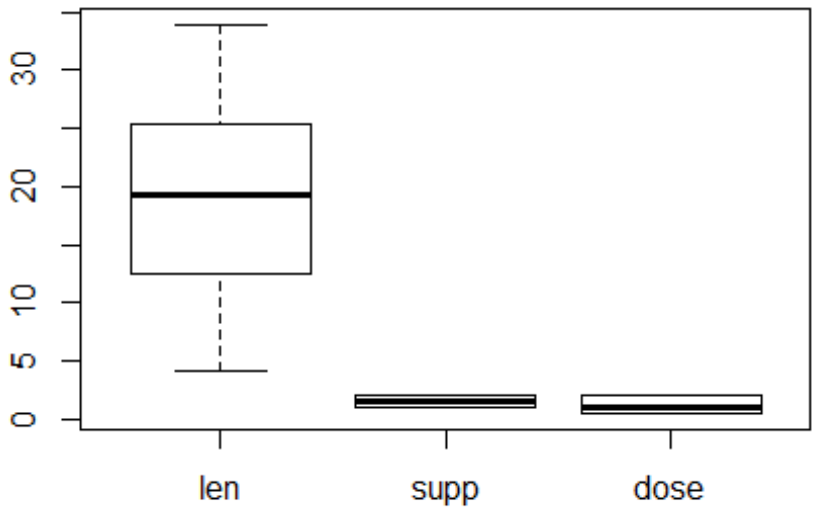
Part 2: Basic Inferential Data Analysis

ToothGrowth dataset has three columns: len, supp and dos. The following code show the information about the three columns.

```
  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

  boxplot(ToothGrowth)
```



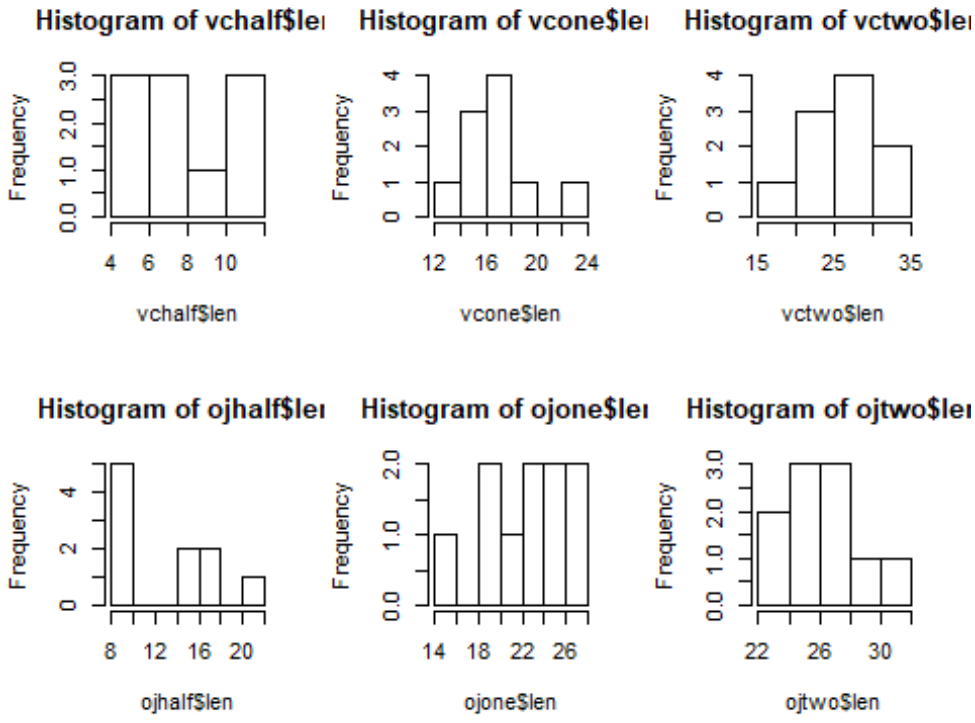
The supp is a factor with two levels: VC and OJ. The dose has three values: 0.5, 1.0 and 2.0.

The following code will do an exploeratory analysis

```
  vc <- ToothGrowth[ToothGrowth$supp =='VC',]
  vchalf <- vc[vc$dose==0.5,]
  vcone <- vc[vc$dose==1.0,]
  vctwo <- vc[vc$dose==2.0,]
  par(mfrow=c(2,3))
  hist(vchalf$len)
  hist(vcone$len)
  sd(vcone$len)
```

```
## [1] 2.515309
```

```
  hist(vctwo$len)
  oj <- ToothGrowth[ToothGrowth$supp =='OJ',]
  ojhalf <- oj[oj$dose==0.5,]
  ojone <- oj[oj$dose==1.0,]
  ojtwo <- oj[oj$dose==2.0,]
  hist(ojhalf$len)
  hist(ojone$len)
  hist(ojtwo$len)
```




```
  mean(vchalf$len)
```

```
## [1] 7.98
```

```
  sd(vchalf$len)
```

```
## [1] 2.746634
```

```
  mean(vcone$len)
```

```
## [1] 16.77
```

```
  sd(vcone$len)
```

```
## [1] 2.515309
```

```
  mean(vctwo$len)
```

```
## [1] 26.14
```

```
  sd(vctwo$len)
```

```
## [1] 4.797731
```

```
  mean(ojhalf$len)
```

```
## [1] 13.23
```

```
  sd(ojhalf$len)
```

```
## [1] 4.459709
```

```
  mean(ojone$len)
```

```
## [1] 22.7
```

```
  sd(ojone$len)
```

```
## [1] 3.910953
```

```
  mean(ojtwo$len)
```

```
## [1] 26.06
```

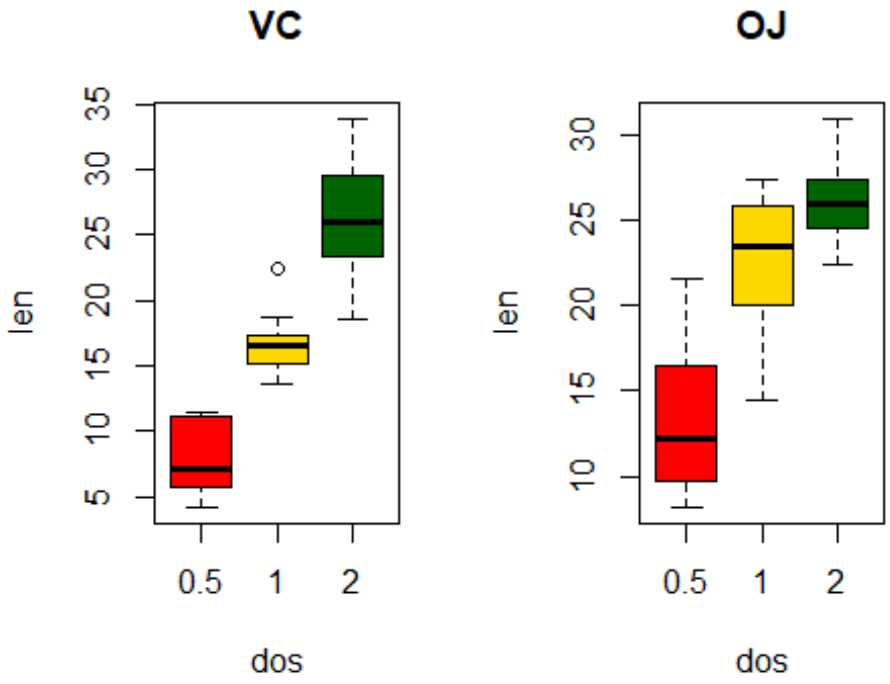
```
  sd(ojtwo$len)
```

```
## [1] 2.655058
```

Supp has two values: VC and OJ respectively. The dose has five levels from 0.5,1.0,to 2.0.

```
  par(mfrow=c(1,2))
  vc <- ToothGrowth[ToothGrowth$supp =='VC',]
  boxplot(vc$len~vc$dos,col=(c("red","gold","darkgreen")),xlab =
'dos',ylab='len',main="VC")
  oj <- ToothGrowth[ToothGrowth$supp =='OJ',]

boxplot(oj$len~oj$dos,col=(c("red","gold","darkgreen")),xlab='dos',ylab
='len',main='OJ')
```



Since there are two variables: supp and dose, firstly, the effect of dose is investigated.

1. the 0.5 level

```
  vchalf <- vc[vc$dose==0.5,]
  ojhalf <- oj[oj$dose==0.5,]
  t.test(vchalf$len,ojhalf$len,var.equal=TRUE)

##
##  Two Sample t-test
##
## data:  vchalf$len and ojhalf$len
## t = -3.1697, df = 18, p-value = 0.005304
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.729738 -1.770262
## sample estimates:
## mean of x mean of y
##      7.98     13.23
```

2. the 1.0 level

```
  vcone <- vc[vc$dose==1.0,]
  ojone <- oj[oj$dose==1.0,]
  t.test(vcone$len,ojone$len,var.equal=TRUE)

##
##  Two Sample t-test
```

```
##
## data:  vcone$len and ojone$len
## t = -4.0328, df = 18, p-value = 0.0007807
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -9.019308 -2.840692
## sample estimates:
## mean of x mean of y
##     16.77     22.70
```

3. the 2.0 level

```
  vctwo <- vc[vc$dose==2.0,]
  ojtwo <- oj[oj$dose==2.0,]
  t.test(vctwo$len,ojtwo$len,var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  vctwo$len and ojtwo$len
## t = 0.046136, df = 18, p-value = 0.9637
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.562999  3.722999
## sample estimates:
## mean of x mean of y
##     26.14     26.06
```

Then the effect of dose is investigated.

For VC supply

```
  vchalf <- vc[vc$dose==0.5,]
  vcone <- vc[vc$dose==1.0,]
  vctwo <- vc[vc$dose==2.0,]
  t.test(vchalf$len,vcone$len,var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  vchalf$len and vcone$len
## t = -7.4634, df = 18, p-value = 6.492e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.264346  -6.315654
## sample estimates:
## mean of x mean of y
##      7.98     16.77
```

```
  t.test(vchalf$len,vctwo$len,var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  vchalf$len and vctwo$len
## t = -10.388, df = 18, p-value = 4.957e-09
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -21.83284 -14.48716
## sample estimates:
## mean of x mean of y
##      7.98     26.14
```

```
  t.test(vctwo$len,vcone$len,var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  vctwo$len and vcone$len
## t = 5.4698, df = 18, p-value = 3.398e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   5.77104 12.96896
## sample estimates:
## mean of x mean of y
##     26.14     16.77
```

For OJ supply

```
  ojhalf <- oj[oj$dose==0.5,]
  ojone <- oj[oj$dose==1.0,]
  ojtwo <- oj[oj$dose==2.0,]
  t.test(ojhalf$len,ojone$len,var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  ojhalf$len and ojone$len
## t = -5.0486, df = 18, p-value = 8.358e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.410814  -5.529186
## sample estimates:
## mean of x mean of y
##     13.23     22.70
```

```
  t.test(ojhalf$len,ojtwo$len,var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  ojhalf$len and ojtwo$len
```

```
## t = -7.817, df = 18, p-value = 3.402e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -16.278223  -9.381777
## sample estimates:
## mean of x mean of y
##     13.23     26.06
```

```
  t.test(ojtwo$len,ojone$len,var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  ojtwo$len and ojone$len
## t = 2.2478, df = 18, p-value = 0.03736
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.2194983 6.5005017
## sample estimates:
## mean of x mean of y
##     26.06     22.70
```

Conclusions:

For both OJ and VC supply, the level of difference has effect on the tooth length since the p-value of all t-test are less then 0.05 at 95% confidence level. The null hypothesis is rejected, the level has effect on the length and higher does is better for tooth growth.

For the same level, at low dose level 0.5 and 1.0, the supply has effect on the tooth length and OJ is better for tooth growth. Since the p-value of 0.5 and 1.0 level are 0.005304 and 0.0007807 respectively, so the null hypothesis is rejected and supply has impact on the length. For 2.0 level, the p-value is 0.9637, the null hypothese is failed to be rejected at 95% confidence level. It suggests that at high dose, the supply may not have statistical significant impact on tooth length.