

Statistical Inference Course Project

S Sagara

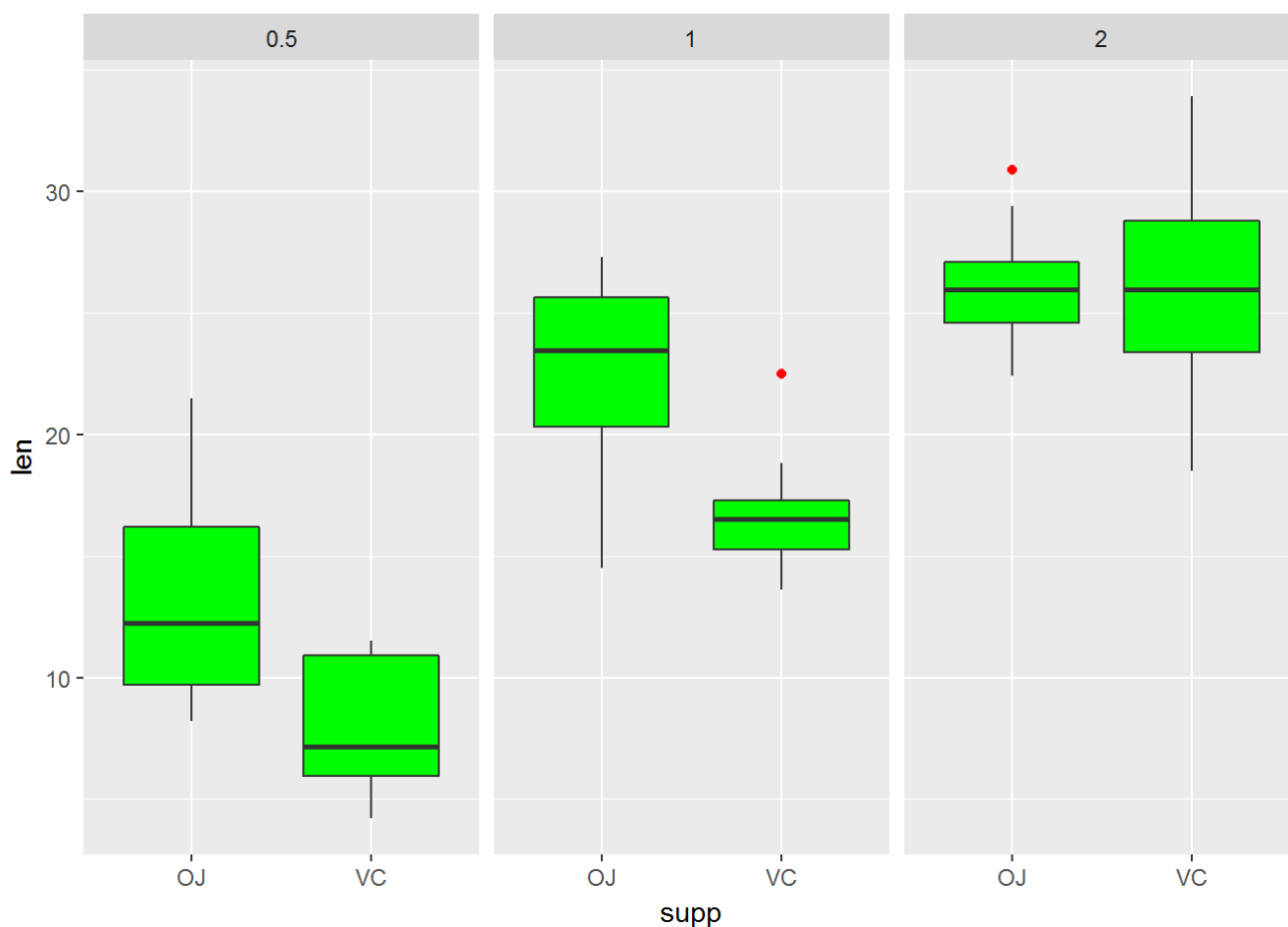
This is 2 Basic inferential data analysis.

Load the ToothGrowth data

```
library(datasets)
df <- ToothGrowth
#Summary statistics
summary(df)
```

```
##      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
```

```
#plot for len by supp in dose facets
#It appears that there is a positive correlation between len and dose.
library(ggplot2)
ggplot(df, aes(supp, len))+
  geom_boxplot(outlier.color = "red", fill="green") +
  facet_grid(.~dose)
```



```
#confidence intervals and/or hypothesis tests
```

```
df_oj <- df[df["supp"]=="OJ",]
df_oj_0.5 <- df_oj[df_oj["dose"]==0.5,]
df_oj_1.0 <- df_oj[df_oj["dose"]==1,]
df_oj_2.0 <- df_oj[df_oj["dose"]==2,]
```

```
df_vc <- df[df["supp"]=="VC",]
df_vc_0.5 <- df_vc[df_vc["dose"]==0.5,]
df_vc_1.0 <- df_vc[df_vc["dose"]==1,]
df_vc_2.0 <- df_vc[df_vc["dose"]==2,]
```

H0: There is no difference in len among doses. Ha: There is difference in len among doses. alpha = .05 in two sided test, thus .025 in one side OJ

Supp OJ for dose between 0.5 and 1.0 H0 is rejected with p-value of 8.785e-05 and the test indicates there is difference. 95% Confidence interval is (-13.415634 -5.524366).

```
##
## Welch Two Sample t-test
##
## data: df_oj_0.5$len and df_oj_1.0$len
## t = -5.0486, df = 17.698, p-value = 8.785e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -13.415634 -5.524366
## sample estimates:
## mean of x mean of y
## 13.23 22.70
```

Supp OJ for dose between 0.5 and 2.0 H0 is rejected with p-value of 1.324e-06 and the test indicates there is difference. 95% Confidence interval is (-16.335241 -9.324759)

```
##
## Welch Two Sample t-test
##
## data: df_oj_0.5$len and df_oj_2.0$len
## t = -7.817, df = 14.668, p-value = 1.324e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -16.335241 -9.324759
## sample estimates:
## mean of x mean of y
## 13.23 26.06
```

Supp OJ for dose between 1.0 and 2.0 H0 cannot be rejected with p-value of 0.0392, which is higher than .025. 95% Confidence interval is (-6.5314425 -0.1885575).

```
##
## Welch Two Sample t-test
##
## data: df_oj_1.0$len and df_oj_2.0$len
## t = -2.2478, df = 15.842, p-value = 0.0392
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -6.5314425 -0.1885575
## sample estimates:
## mean of x mean of y
## 22.70 26.06
```

VC Supp VC for dose between 0.5 and 1.0 H_0 is rejected with p-value of $6.811e-07$ and the test indicates there is difference. 95% Confidence interval is (-11.265712 -6.314288).

```
##
## Welch Two Sample t-test
##
## data: df_vc_0.5$len and df_vc_1.0$len
## t = -7.4634, df = 17.862, p-value = 6.811e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -11.265712 -6.314288
## sample estimates:
## mean of x mean of y
## 7.98 16.77
```

Supp VC for dose between 0.5 and 2.0 H_0 is rejected with p-value of $4.682e-08$ and the test indicates there is difference. 95% Confidence interval is (-21.90151 -14.41849).

```
##
## Welch Two Sample t-test
##
## data: df_vc_0.5$len and df_vc_2.0$len
## t = -10.388, df = 14.327, p-value = 4.682e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -21.90151 -14.41849
## sample estimates:
## mean of x mean of y
## 7.98 26.14
```

Supp VC for dose between 1.0 and 2.0 H_0 is rejected with p-value of $9.156e-05$ and the test indicates there is difference. 95% Confidence interval is (-13.054267 -5.685733).

```
##  
## Welch Two Sample t-test  
##  
## data: df_vc_1.0$len and df_vc_2.0$len  
## t = -5.4698, df = 13.6, p-value = 9.156e-05  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -13.054267 -5.685733  
## sample estimates:  
## mean of x mean of y  
## 16.77 26.14
```

In conclusion, the null hypothesis of no difference in len by supp among doses is rejected, except OJ len between dose 1 and 2, and as a result, it would be reasonable to assume there is indeed difference in len among doses statistically.