

Statistical Inference Course Project Part 2

Smriti Gambhir

February 28, 2016

1. Load the ToothGrowth data and perform some basic exploratory data analyses.

```
data("ToothGrowth")
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 ...
```

```
head(ToothGrowth)
```

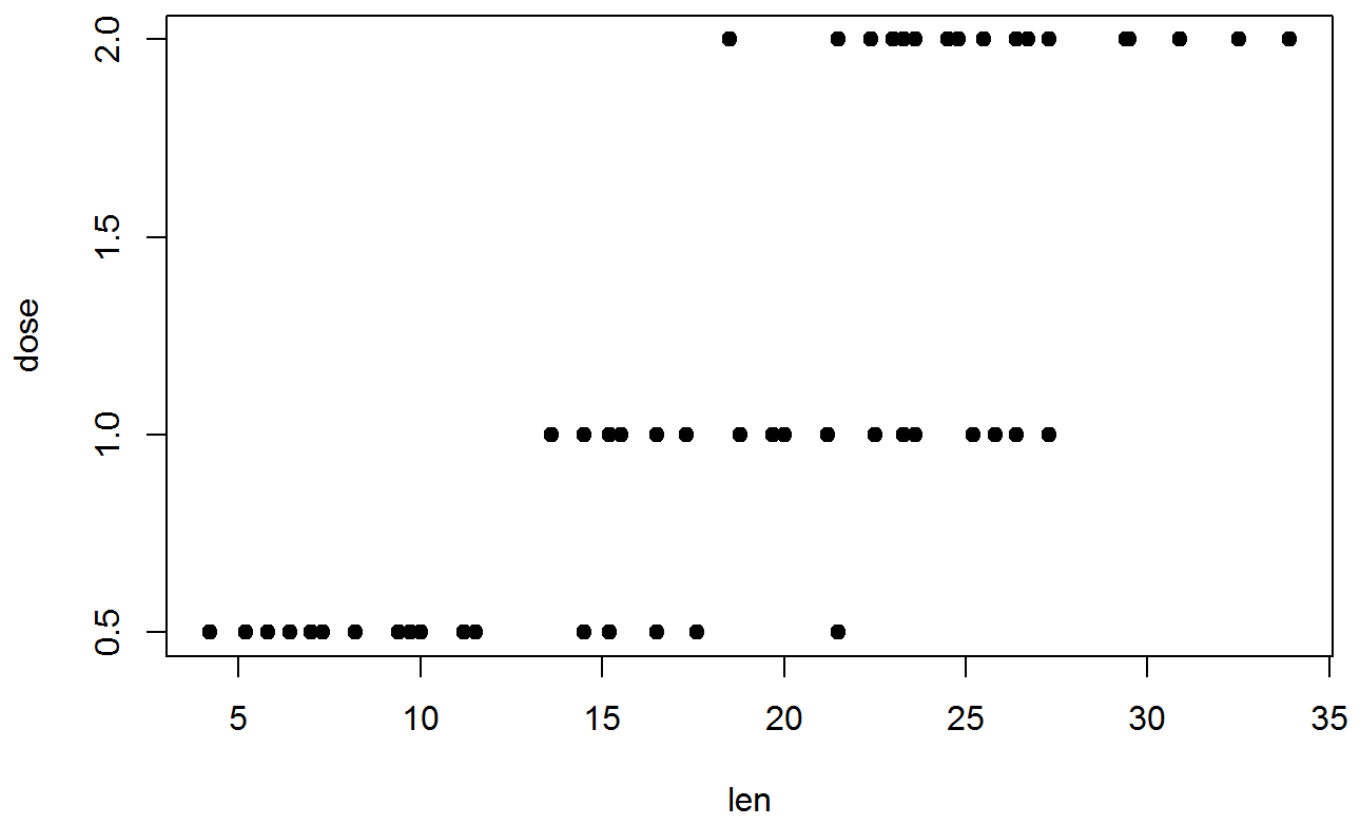
```
##      len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```
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
```

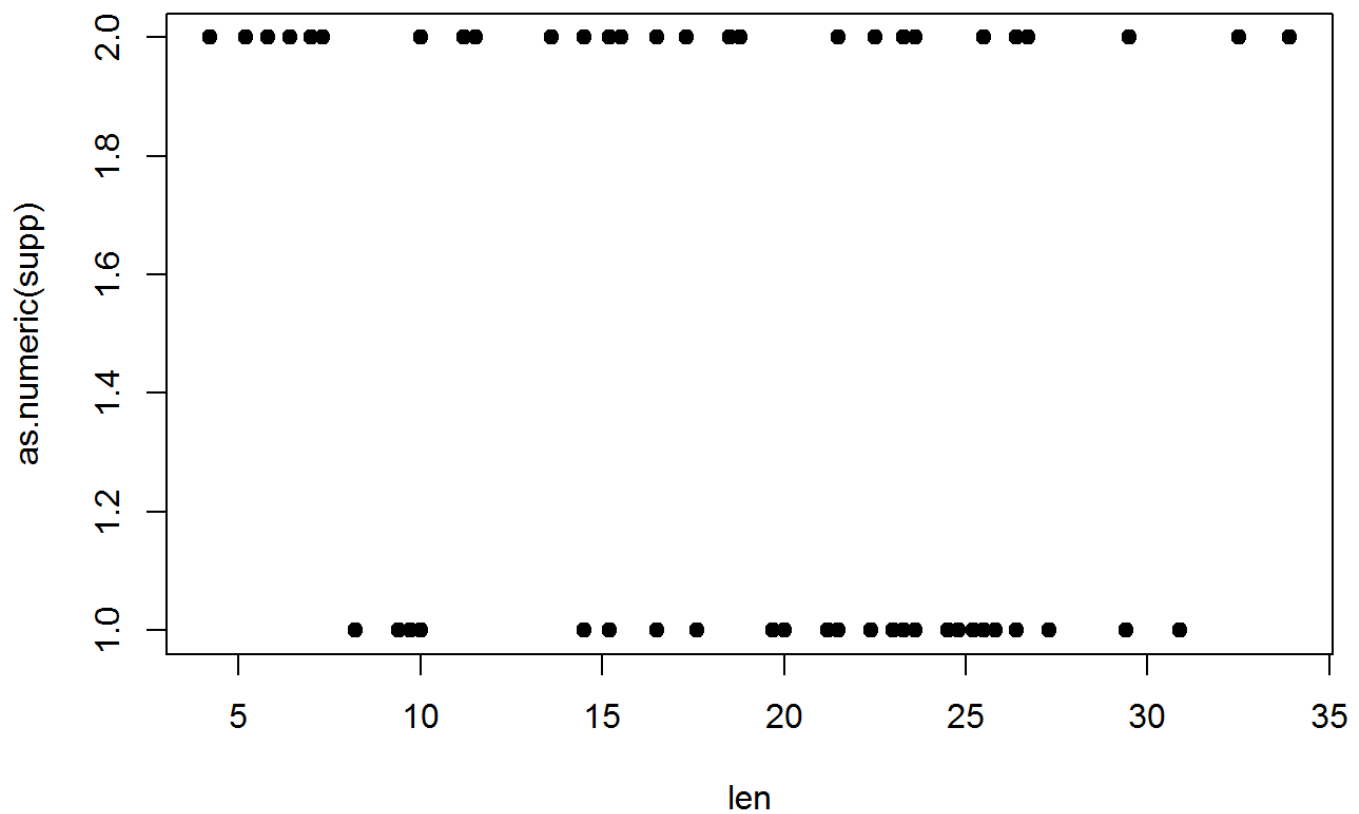
dose versus len plot

```
plot(dose~len,data = ToothGrowth,pch=19)
```



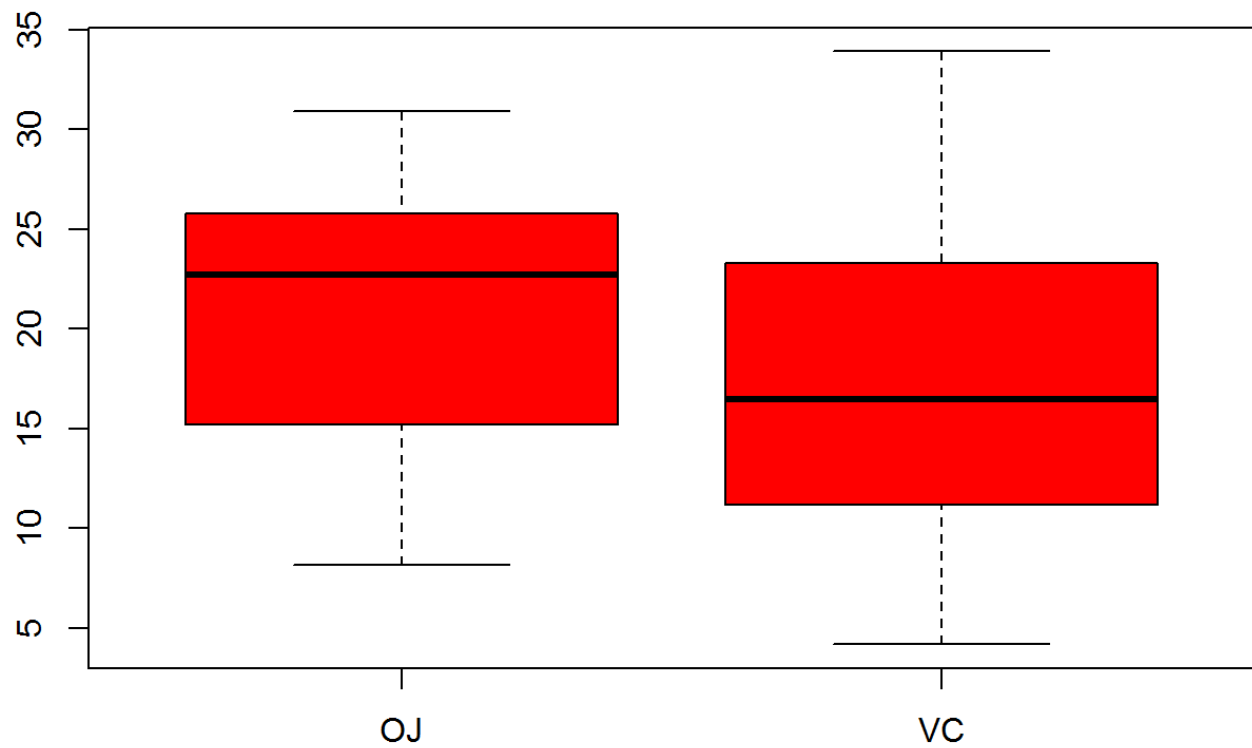
supp versus len plot

```
plot(as.numeric(supp)~len,data = ToothGrowth,pch=19)
```



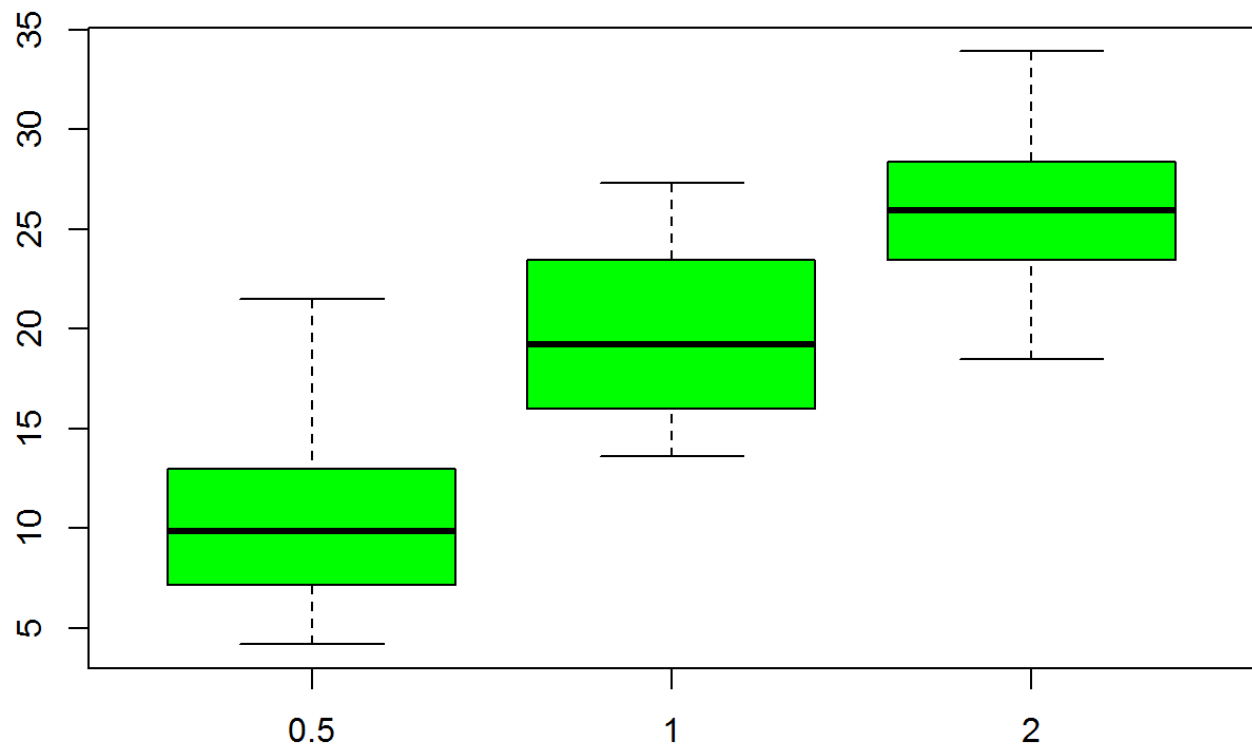
Boxplot of len based on supp

```
boxplot(len~supp,data = ToothGrowth,col="red")
```



Boxplot of len based on dose

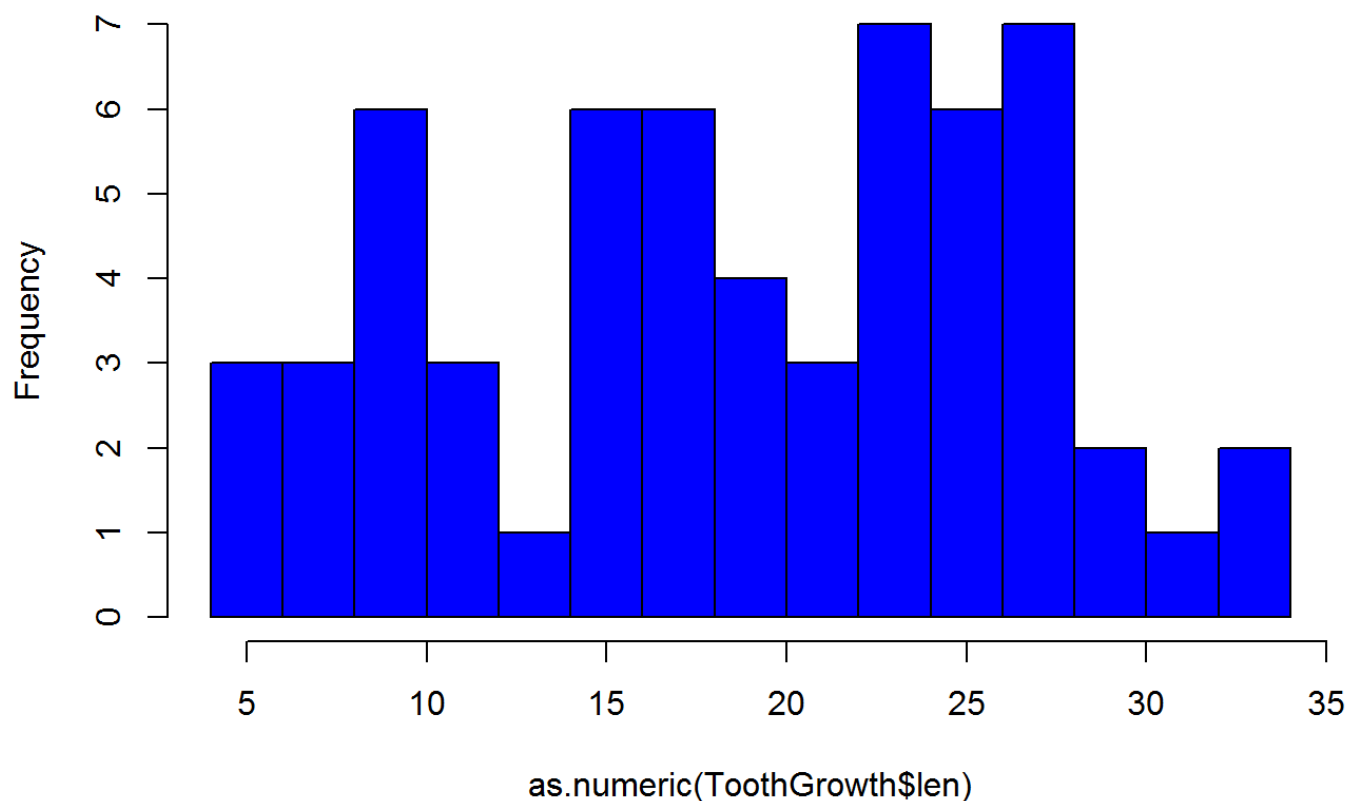
```
boxplot(len~dose,data = ToothGrowth,col="green")
```



Histogram of len

```
hist(as.numeric(ToothGrowth$len),col = "blue",breaks=15)
```

Histogram of as.numeric(ToothGrowth\$len)



2.Perform some relevant confidence intervals and/or tests

```
t.test(len ~ supp, ToothGrowth[ToothGrowth$dose == .5, ])
```

```
##
## 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
```

```
t.test(len ~ supp, ToothGrowth[ToothGrowth$dose == 1, ])
```

```
##
## 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
```

```
t.test(len ~ supp, ToothGrowth[ToothGrowth$dose == 2, ])
```

```
##
## Welch Two Sample t-test
##
## data: len by supp
## t = -0.046136, 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
```

Conclusions

Lower p-value for dose=0.5 and dose=1 indicates significant difference in means as can be seen [13.23-7.98=5.25 and 22.70-16.77=5.93] Whereas higher p-value for dose=2 indicates a very small or insignificant difference in mean.