Course Assignment – Basic inferential data analysis

Author: Sahil Grover

Analyzing the ToothGrowth 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
```

```
table(ToothGrowth$len)
```
Some random values

```
table(ToothGrowth$supp)
OJ VC
30 30
```

```
table(ToothGrowth$dose)
0.5   1   2
 20  20  20
```

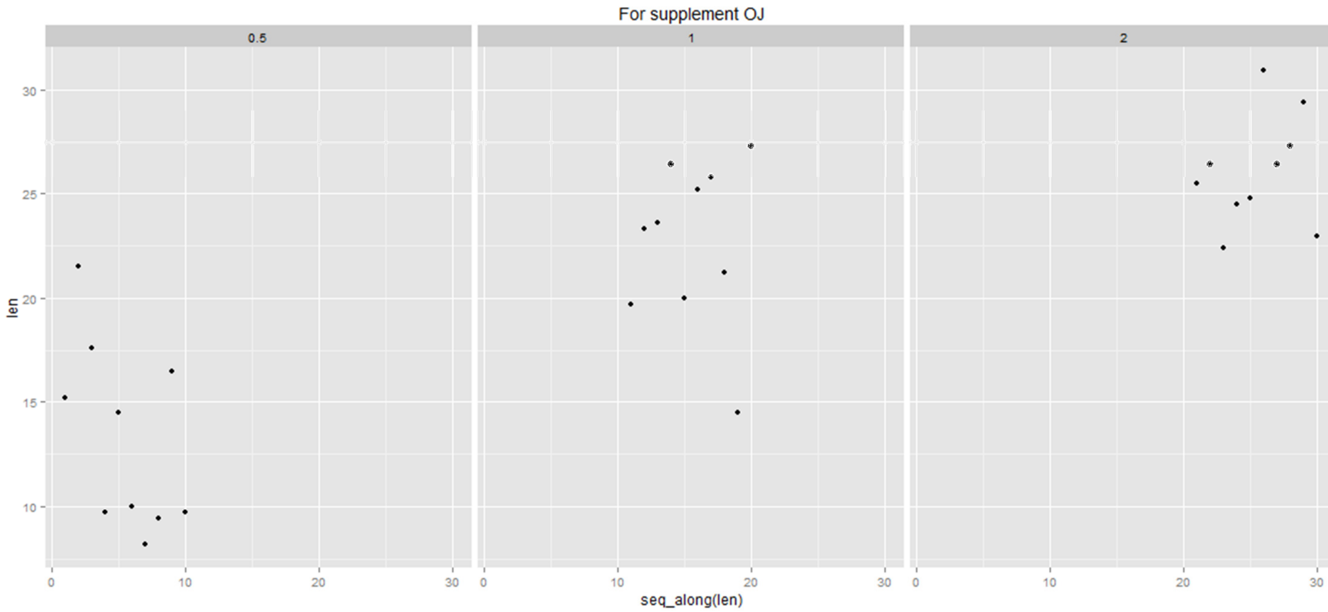
Hence, we could see that this data provides values named "len" for different amounts of doses (0.5,1,2) for 2 types of supplements (VC,OJ).

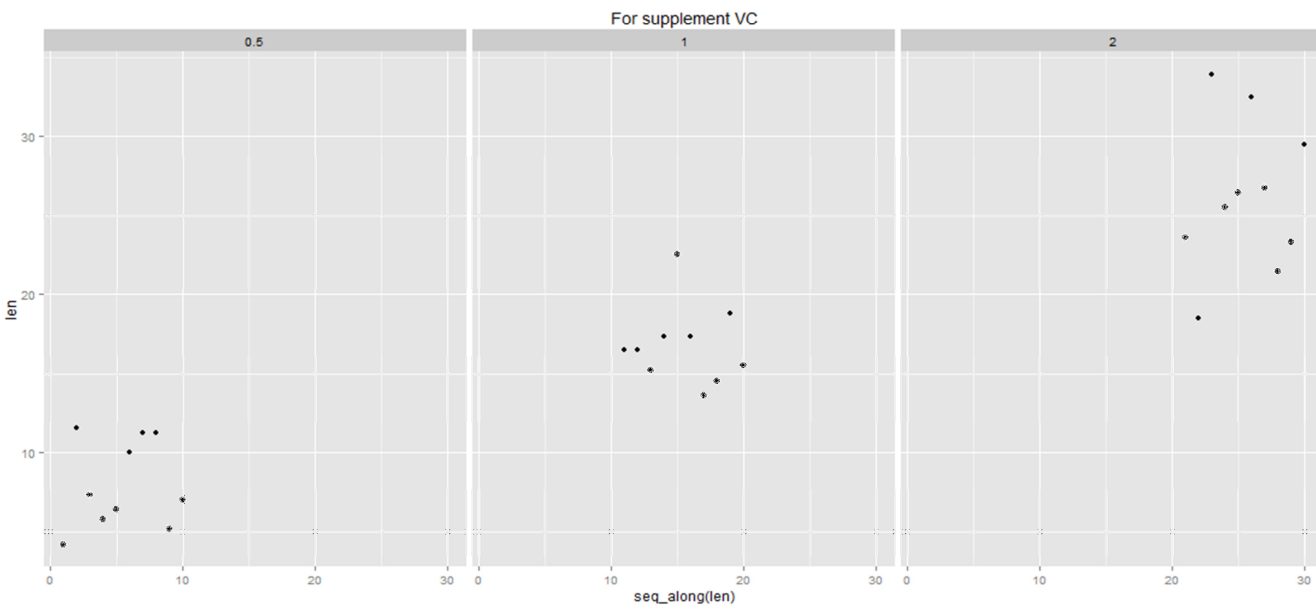
```
# Subsetting the data
      ToothGrowth_VC<-ToothGrowth[ToothGrowth$supp=="VC",]
      ToothGrowth_OJ<-ToothGrowth[ToothGrowth$supp=="OJ",]
# Plotting the both subsets for different doses
      qplot(y=len,data=ToothGrowth_OJ,facets = .~dose) + ggtitle("For supplement OJ")
      qplot(y=len,data=ToothGrowth_VC,facets = .~dose) + ggtitle("For supplement VC")
```

Please see the next page to see some exploratory analysis of the data.

Means , variances and 90% confidence intervals of the data

```
    supp  dose average  variance confidenceleft confidenceright
  (fctr) (dbl)   (dbl)     (dbl)          (dbl)           (dbl)
1     OJ   0.5   13.23 19.889000      10.644791       15.815209
2     OJ   1.0   22.70 15.295556      20.432894       24.967106
3     OJ   2.0   26.06  7.049333      24.520913       27.599087
4     VC   0.5    7.98  7.544000       6.387828        9.572172
5     VC   1.0   16.77  6.326778      15.311923       18.228077
6     VC   2.0   26.14 23.018222      23.358846       28.921154
```





Hypothesis 1

Let's make a null hypothesis that the tooth growth from both supplements is same.

To check it, we will perform a hypothesis test on the data.

Assumption

Here we are **assuming that the tooth growth is supplement dependent only**

Computation:

Let's compute the statistics of data based only on "supp"

```
     supp  average variance confidenceleft confidenceright
   (fctr)    (dbl)    (dbl)          (dbl)           (dbl)
1      OJ 20.66333 43.63344       18.61418        22.71249
2      VC 16.96333 68.32723       14.39907        19.52759
```

Hence we compute mean(OJ)-mean(VC)+c(-1,1)*qt(0.95,58)*sqrt(var(OJ)/30+var(VC)/30)*sqrt(1/30+1/30).

```
[1] 2.866229 4.533771
```

Conclusions

We can conclude that since the confidence intervals of the difference does not contain zero. Therefore, we can ***reject*** the null hypothesis.

Hypothesis 2

Let's make a null hypothesis that there is no difference in tooth growth based on the dose.

Assumption

Here we are **assuming that the tooth growth is dose dependent only**

Computation:

Let's compute the statistics of data based only on "dose"

```
   dose average variance confidenceleft confidenceright
  (dbl)   (dbl)    (dbl)          (dbl)           (dbl)
1   0.5  10.605 20.24787       8.865185        12.34481
2   1.0  19.735 19.49608      18.027790        21.44221
3   2.0  26.100 14.24421      24.640740        27.55926
```

Hence we compute relation between "0.5" and "1.0"

mean(dose_0.5)-mean(dose_1.0)+c(-1,1) * qt(0.95,38) * sqrt(var(0.5)/20+var(1.0)/20) * sqrt(1/20+1/20)

```
[1] -9.881565 -8.378435
```

Now, let's compute relation between "1.0" and "2.0"

mean(dose_1.0)-mean(dose_2.0)+c(-1,1) * qt(0.95,38) * sqrt(var(1.0)/20+var(2.0)/20) * sqrt(1/20+1/20)

```
[1] -7.143726 -5.586274
```

Even if it is evident from the respective averages that "0.5" and "2.0" have very different tooth growth, lets double check it by computing

mean(dose_0.5)-mean(dose_2.0)+c(-1,1) * qt(0.95,38) * sqrt(var(0.5)/20+var(2.0)/20) * sqrt(1/20+1/20)

```
[1] -16.29373 -14.69627
```

Conclusions:

We can conclude that since the confidence intervals of the difference in all three cases does not contain zero. Therefore, we can ***reject*** the null hypothesis.