Assignment 4

Perform measure of dispersion, standard deviation, quartile deviation etc

Use built in functions and write user defined functions for all. Compare the results of user defined function with built in functions.

Standard Deviation

sd() is to calculate the standard deviation

```
> sd(PlantGrowth$weight)
[1] 0.7011918
```

User defined function

```
> stdf=function(x){
+   s=sum((mean(x)-x)^2)
+   st=s/(length(x)-1)
+   print(st^0.5)
+ };
> stdf(PlantGrowth$weight)
[1] 0.7011918
```

Variance

var() is used to calculate the variance

```
> var(PlantGrowth$weight)
[1] 0.49167
```

User defined function

```
> variancef=function(x){
+   s=sum((mean(x)-x)^2)
+   st=s/(length(x)-1)
+   print(st)
```

```
+ };
> variancef(PlantGrowth$weight)
[1] 0.49167
```

Range

Range provides the highest and the lowest value.

```
> range(PlantGrowth$weight)
[1] 3.59 6.31
```

User defined function

```
> rangeFunction=function(x){
+ ma=max(x)
+ mi=min(x)
+ sprintf("%f %f",mi,ma)
+ };
> rangeFunction(PlantGrowth$weight)
[1] "3.590000 6.310000"
```

Quartile

This provides 0, 25, 50, 75 and 100 quartiles.

In this: • 0% = Lowest value,

- 25% = First quartile (Q1)
- 50% = Median (Q2)
- 75% = Third quartile (Q3)
- 100% = Highest value

Example of using quantile with a vector

```
> quantile(PlantGrowth$weight)
  0%  25%  50%  75% 100%
```

3.590 4.550 5.155 5.530 6.310

This is the default version of this function, and it produces the 0th percentile, 25th percentile, 50th percentile, 75th percentile, and 100th percentile

User defined function

```
> quartileF=function(x){
+   q=sort(x)
+   zeroeth=q[1]
+   fq=q[length(x)/4]
+   sq=q[length(x)*0.5]
+   third=length(x)*3/4
+   tq=q[third]
+   lq=q[length(x)]
+   sprintf("%f %f %f %f %f",zeroeth,fq,sq,tq,lq)
+   };
> quartileF(PlantGrowth$weight)
[1] "3.590000 4.500000 5.140000 5.500000 6.310000"
```

Finding percentile in R or quantile in R

The n^{th} **percentile** of an observation variable is the value that cuts off the first *n* percent of the data values when it is sorted in ascending order.

Using a vector

```
> x = c(5,10,12,15,20,24,27,30,35)
> quantile(x, probs = c(0.125,0.375,0.625,0.875))
12.5% 37.5% 62.5% 87.5%
   10    15    24    30
```

Using the PlantGrowth dataset

```
> quantile(PlantGrowth$weight,probs = c(0.125,0.375,0.625,0.875))
  12.5%   37.5%   62.5%   87.5%
4.26375 4.88000 5.29500 5.93000
```

The user defined function for quantile was not giving accurate results so has not been included.

InterQuartile Range: The **interquartile range** of an observation variable is the difference of its upper and lower quartiles. It is a measure of how far apart the middle portion of data spreads in value.

```
> IQR(PlantGrowth$weight)
[1] 0.9
```

User defined function

```
> interquartileF=function(x){
+ q=sort(x)
+ fq=q[length(x)/4]
+ third=length(x)*3/4
+ tq=q[third]
+ sprintf("%f",tq-fq)
+ };
> interquartileF(PlantGrowth$weight)
[1] "1.000000"
```

Coefficient Of Variation- The ratio of standard deviation to mean

```
> cov=sd(PlantGrowth$weight)/mean(PlantGrowth$weight)
> print(cov)
[1] 0.1382204
```

User defined function

```
> covariance=function(x){
+ s=sum((mean(x)-x)^2)
```

```
+ st=s/(length(x)-1)
+ sdev=st^0.5
+ umean=sum(x)/length(x)
+ cov=sdev/umean
+ print(cov)
+ };
> covariance(PlantGrowth$weight)
[1] 0.1382204
```

Summary()

```
> summary (PlantGrowth$weight)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  3.590   4.550   5.155   5.073   5.530   6.310
```

User defined function

```
userSummaryF=function(x){
mi=min(x)
me=sum(x)/length(x)
ma=max(x)
med=median(x)
fq=sort(PlantGrowth$weight)
firstq=fq[length(x)/4]
third=length(x)*3/4
tq=fq[third]
print("Min.   1st Qu.   Median.   Mean.   3rd Qu.   Max.")
sprintf("%f %f %f %f %f %f",mi,firstq,med,me,tq,ma)
};
```

```
userSummaryF(PlantGrowth$weight)
[1] "Min.   1st Qu.   Median.   Mean.   3rd Qu.   Max."
[1] "3.590000 4.500000 5.155000 5.073000 5.500000 6.310000"
```