VIT®
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

Activity 4:

Name: Preyash **Date: 01/02/2022**

Registration Number: 20BPS1022

Aim: Using R Studio to study correlation and moments

Tools Used: R Studio

Syntax/ Commands used:

We are using a special library available in R Studio which is moments. It has routines for calculating moments, Pearson's kurtosis, Geary's kurtosis, and skewness, as well as tests relating to these concepts.

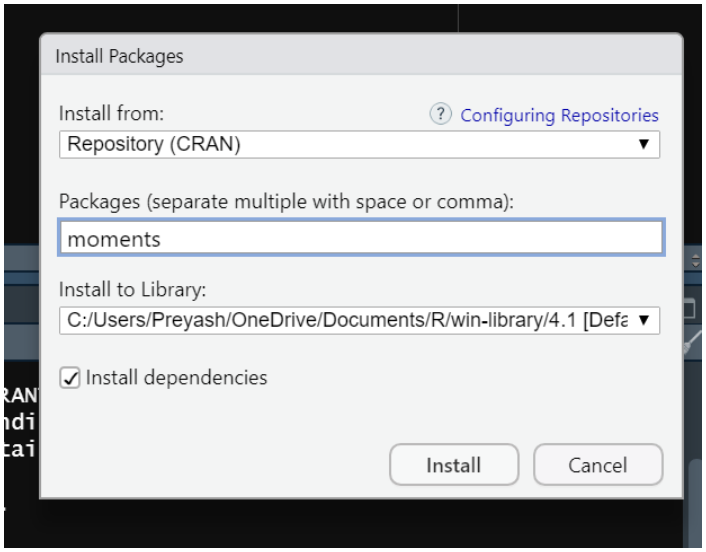
We majorly used:

all.moments(): This function computes all the sample moments of the chosen type up to a given order

kurtosis(): This function computes the estimator of Pearson's measure of kurtosis.

cor(): Cor computes the variance of x and the covariance or correlation of x and y

Installing packages:



Questions:

Task 1:

1. Calculate the (a) first (b) second (c) third and (d) fourth moments of the set 4, 8, 12, 20, 6, 7, 3 for both raw and central moments with and without package.

Code:

```
library(moments)

dataset=c(4,8,12,20,6,7,3)

datasetLen=length(dataset)

mean=sum(dataset)/datasetLen

mom1=sum(dataset-mean)/datasetLen

mom2=sum((dataset-mean)^2)/datasetLen

mom3=sum((dataset-mean)^3)/datasetLen

mom4=sum((dataset-mean)^4)/datasetLen

momr1=sum(dataset)/datasetLen

momr2=sum((dataset)^2)/datasetLen

momr3=sum((dataset)^3)/datasetLen

momr4=sum((dataset)^4)/datasetLen

print("raw")

print("package")

all.moments(dataset,order.max=4,absolute = TRUE)

print("manual")

momr1

momr2

momr3

momr4

print("central")

print("package")

all.moments(dataset,order.max=4,central = TRUE, absolute = FALSE)

print("manual")
```

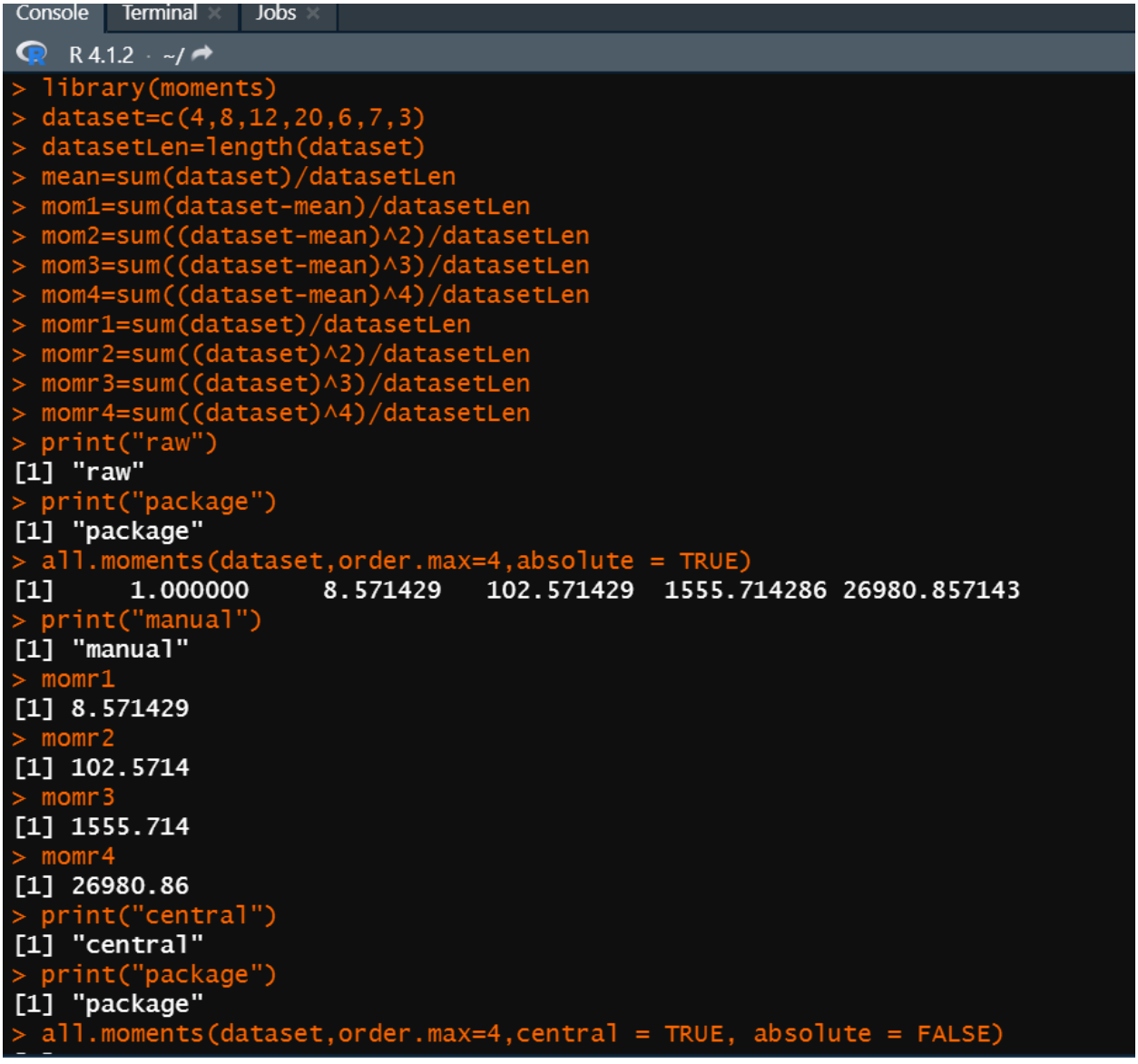
mom1

mom2

mom3

mom4

Output:



```
[1] "package"
> all.moments(dataset,order.max=4,central = TRUE, absolute = FALSE)
[1] 1.000000e+00 2.537653e-16 2.910204e+01 1.776385e+02 2.663989e+03
> print("manual")
[1] "manual"
> mom1
[1] 2.537653e-16
> mom2
[1] 29.10204
> mom3
[1] 177.6385
> mom4
[1] 2663.989
>
```

Task 2:

2. A computer, while calculating the correlation coefficient between two variables X and Y from 25 pairs of observations, obtained the following results:

$$n = 25, \quad \Sigma X = 125, \quad \Sigma X^2 = 650,$$
$$\Sigma Y = 100, \quad \Sigma Y^2 = 460, \quad \Sigma XY = 508.$$

It was, however, later discovered at the time of checking that two pairs had been copied *incorrectly as*

X	Y
6	14
8	6

while the correct values were

X	Y
8	12
6	8

Use R code to correct the correlation coefficient.

Code:

n=25

isigx=125

isigxx=650

isigy=100

isigyy=460

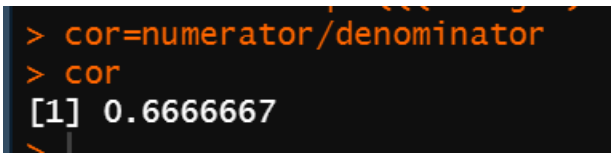
isigxy=508

ip1x=6

ip1y=14

```
ip2x=8
ip2y=6
p1x=8
p1y=12
p2x=6
p2y=8
sigx=isigx-ip1x-ip2x+p1x+p2x
sigy=isigy-ip1y-ip2y+p1y+p2y
sigxx=isigxx-(ip1x*ip1x)-(ip2x*ip2x)+(p1x*p1x)+(p2x*p2x)
sigyy=isigyy-(ip1y*ip1y)-(ip2y*ip2y)+(p1y*p1y)+(p2y*p2y)
sigxy=isigxy-(ip1x*ip1y)-(ip2x*ip2y)+(p1x*p1y)+(p2x*p2y)
numerator=(n*sigxy)-(sigx*sigy)
denominator=sqrt(((n*sigxx)-(sigx^2))*((n*sigyy)-(sigy^2)))
cor=numerator/denominator
cor
```

Output:



Task 3:

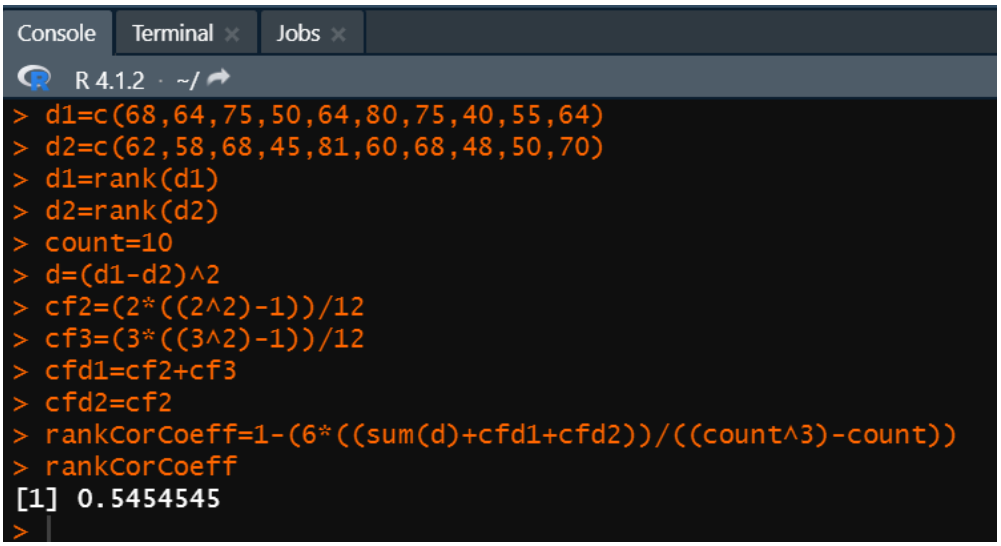
3. For the following data, write R code to obtain the rank correlation coefficient (use formula)

X	68	64	75	50	64	80	75	40	55	64
Y	62	58	68	45	81	60	68	48	50	70

Code:

```
d1=c(68,64,75,50,64,80,75,40,55,64)

d2=c(62,58,68,45,81,60,68,48,50,70)

d1=rank(d1)

d2=rank(d2)

count=10

d=(d1-d2)^2

cf2=(2*((2^2)-1))/12

cf3=(3*((3^2)-1))/12

cfd1=cf2+cf3

cfd2=cf2

rankCorCoeff=1-(6*((sum(d)+cfd1+cfd2))/((count^3)-count))

rankCorCoeff
```

Output:



Task 4 :

4. The time x in years that an employee spent at a company and the employee's hourly pay, y, for 5 employees are listed in the table below. Calculate and interpret the correlation coefficient r. Include a plot of the data in your discussion.

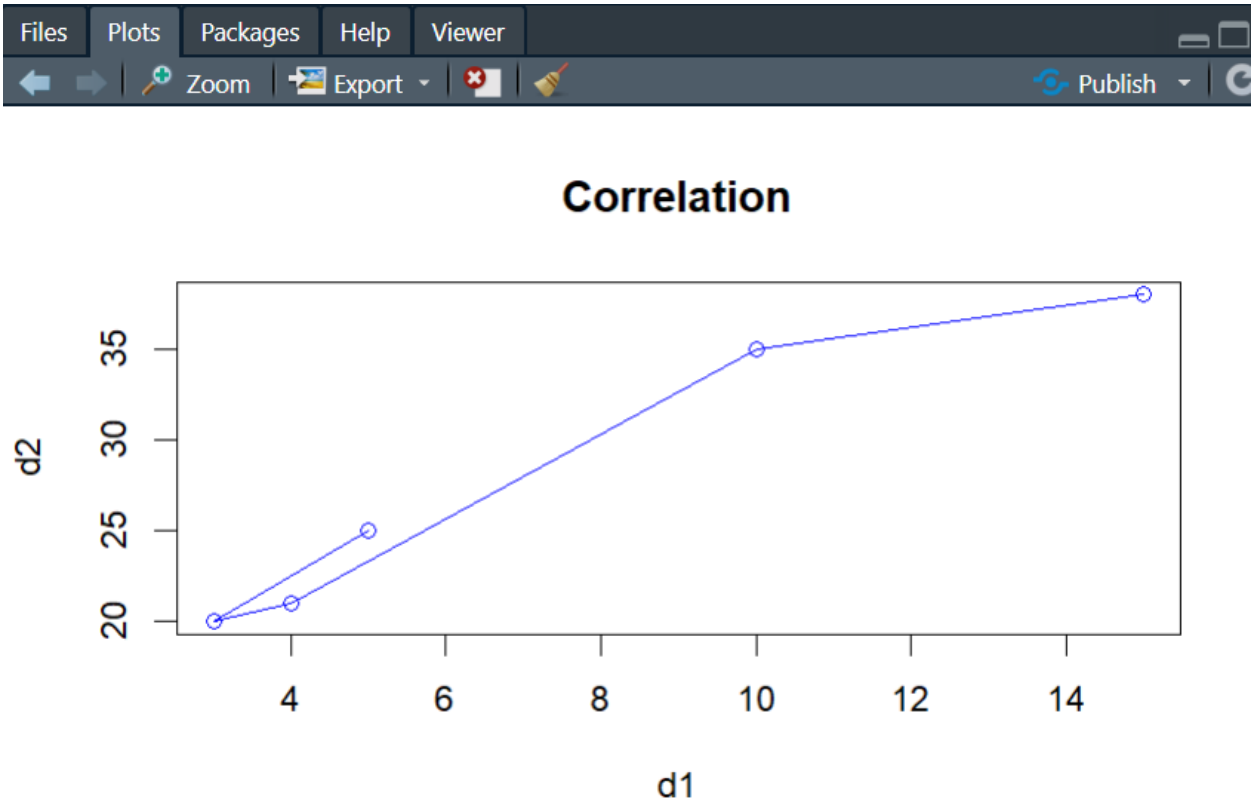
x	y
5	25
3	20
4	21
10	35
15	38

Code:

```
d1=c(5,3,4,10,15)

d2=c(25,20,21,35,38)

count=5

cor.test(d1,d2)

plot(d1,d2,type="o",d1lab ="d1",d2lab = "d2", main = "Correlation",col="blue")
```

Output:

```
> cor.test(d1,d2)

        Pearson's product-moment correlation

data:  d1 and d2
t = 6.7846, df = 3, p-value = 0.006545
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.5970187 0.9980275
sample estimates:
      cor
0.9689241

> plot(d1,d2,type="o",d1lab ="d1",d2lab = "d2", main = "Correlation",col="blue")
There were 12 warnings (use warnings() to see them)
```



Task 5:

5.

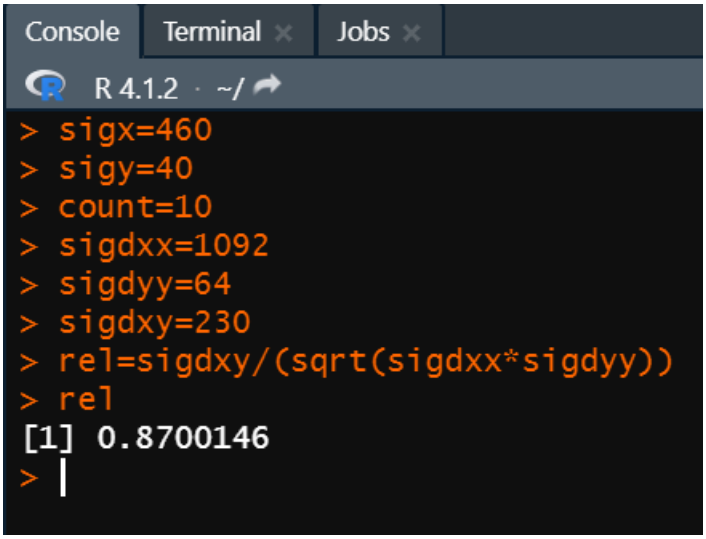
The soil temperature (X) and the germination time (Y) of 10 pairs are recorded for investigation. Calculate the Karl Pearson's correlation coefficient based on the following information.

	Soil Temperature	Germination time
Sum of the observations	460	40
Square of deviations from Mean	1092	64
Summation of the product of deviations of X and Y from their respective average is 230		

Code:

```
sigx=460
sigy=40
count=10
sigdxx=1092
sigdyy=64
sigdxy=230
rel=sigdxy/(sqrt(sigdxx*sigdyy))
rel
```

Output:



Result: We successfully solved all the questions and they were verified manually.