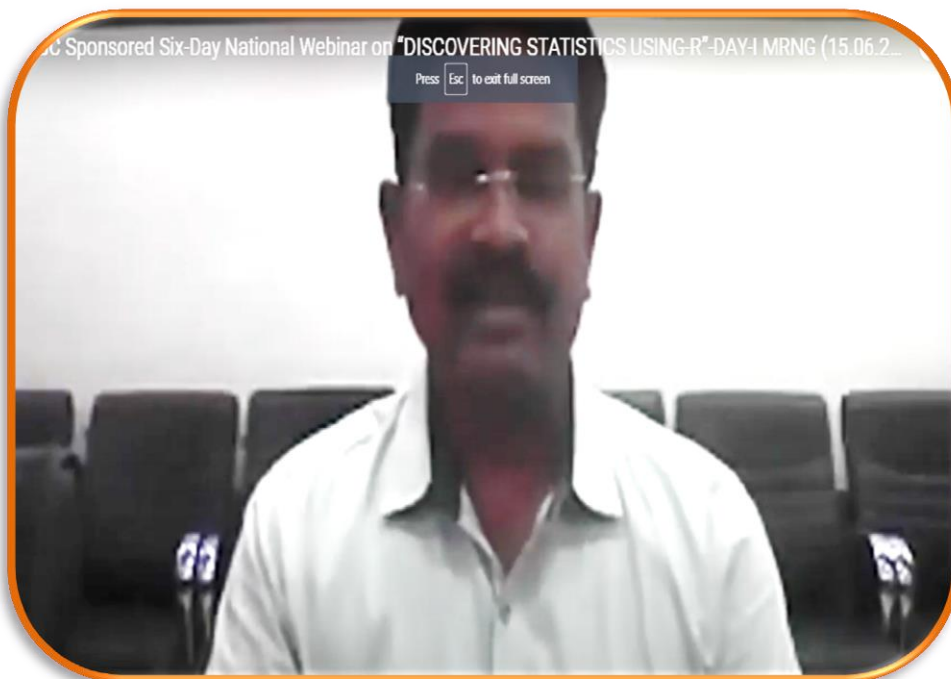


UGC Sponsored Six-Day National Webinar on
"Discovering Statistics using R"

15th-20th June, 2020

Organized by: **Department of Mathematics & Statistics**

A banner for the webinar. It features the logos of the UGC and the Department of Mathematics & Statistics at the top left. The text reads: "UGC Sponsored Six-Day National Webinar on 'Discovering Statistics using R' 15th-20th June, 2020 Organized by Department of Mathematics & Statistics". The background includes a globe, a magnifying glass over a bar chart, and various data-related icons.

President Opening Remarks by Mr. E. Vara Prasad, Vice - Principal

DAY – I – 15th June, 2020

Introduction to R & Understanding the Data



*In a National Webinar on **Discovering Statistics using R**
organized by
Dept. of Statistics, KBN College, Vijayawada, Andhra Pradesh
June 15-20, 2020*

Dr. R. Vishnu Vardhan

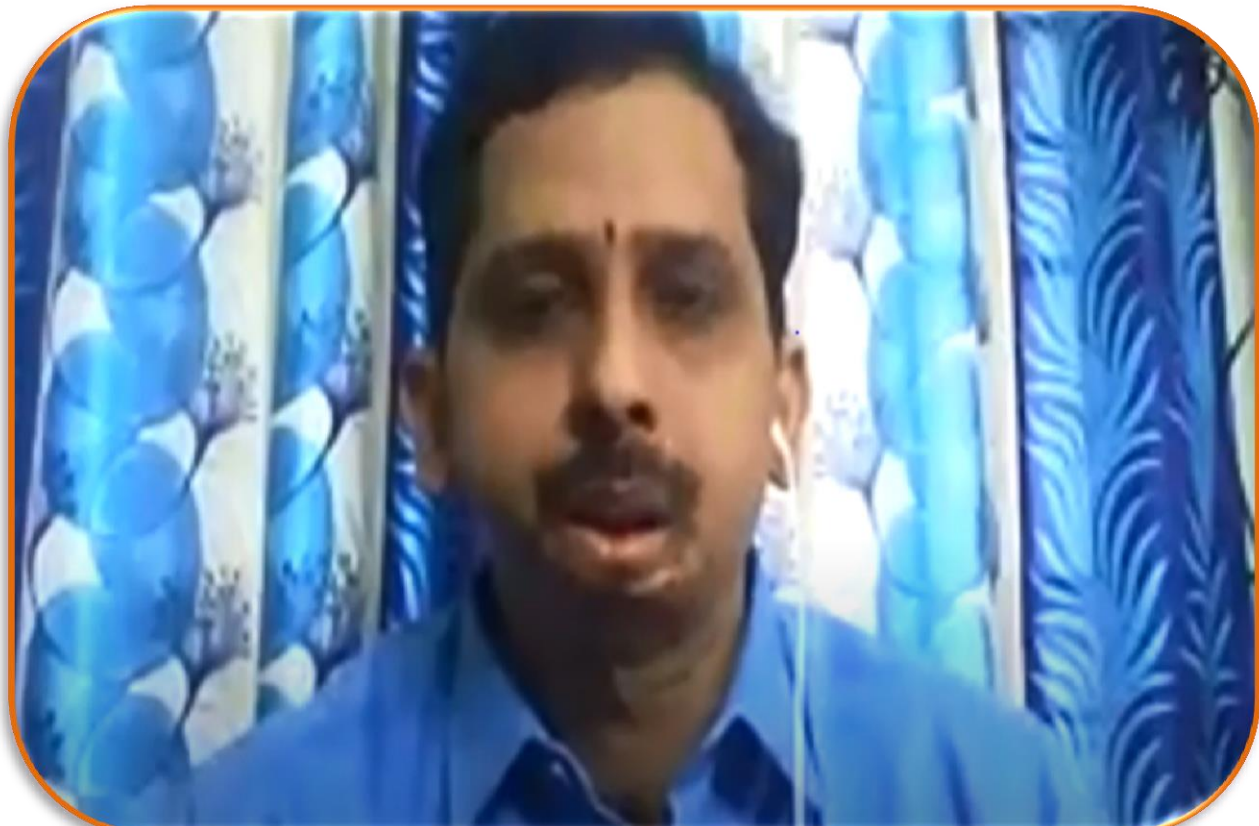
*Assistant Professor, Department of Statistics
Pondicherry University, Puducherry
Email: vrstatsguru@gmail.com; 9487609603*



8:39 / 1:19:14



*Introduction to R by Dr. R. Vishnu Vardhan, Asst. Prof. Dept. of Statistics,
Pondicherry University, Puducherry*



Installing a package and using it



```
> install.packages("tree", dependencies=TRUE)
installing package into 'C:/Users/vishnu/Documents/R/win-library/4.0'
(as 'lib' is unspecified)
--- Please select a CRAN mirror for use in this session ---
trying URL 'https://ftp.fau.de/cran/bin/windows/contrib/4.0/tree_1.0-45'
Content type 'application/zip' length 178304 bytes (174 KB)
downloaded 174 KB

package 'tree' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:/Users/vishnu/AppData/Local/Temp/library/BYNI/downloaded_packages

> library(tree)

> ##Calculations
> 2+6; 7-2; 8*3
[1] 8
[1] 5
[1] 24
> log(102/22)
[1] 1.53393
> ##Complex numbers
> c=8.2-3i
> c
[1] 8.2-3i
> Re(c)
[1] 8.2
> is.complex(c)
[1] TRUE
> as.complex(4.1)
[1] 4.1+0i
> floor(4.2)
[1] 4
> round(5.8)
[1] 6
> |
```

Operators & Logical, relational operations



Operator	Description
+ - * / %/ % % ^	arithmetic (plus, minus, times, divide, integer quotient, modulo, power)
>= < <= == !=	relational (greater than, greater than or equals, less than, less than or equals, equals, not equals)
! &	logical (not, and, or)
~	model formulae ('is modelled as a function of')
<- -> =	assignment (gets)
\$	list indexing (the 'element name' operator)
:	create a sequence

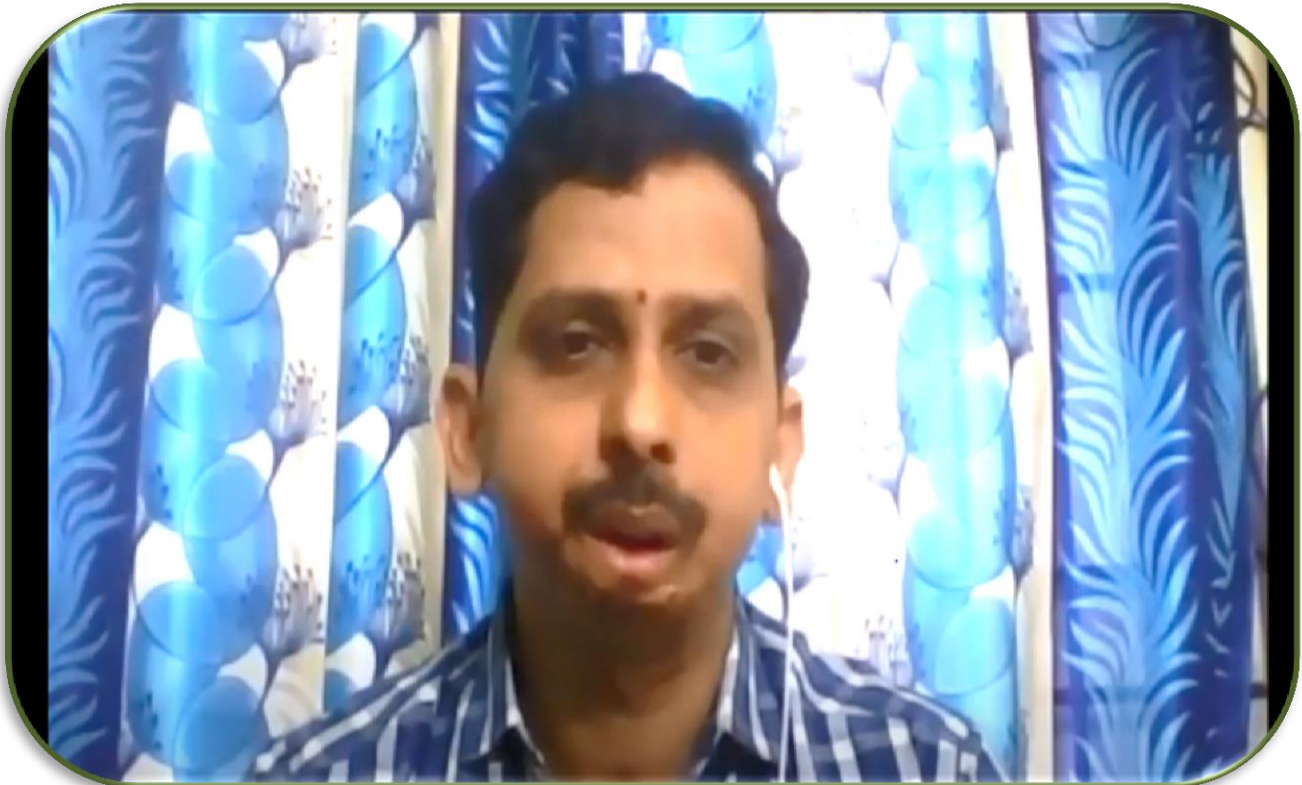
```
R Console
> ##Logical Arithmetic
> x=3:10
> x>5
[1] FALSE FALSE FALSE TRUE TRUE TRUE TRUE TRUE
> x<=7
[1] TRUE TRUE TRUE TRUE TRUE FALSE FALSE FALSE
> any(x<4)
[1] TRUE
> any(x>10)
[1] FALSE
> any(x%>10)
[1] TRUE
> ##Generating Sequences
> 2:10
[1] 2 3 4 5 6 7 8 9 10
> 10:2
[1] 10 9 8 7 6 5 4 3 2
> seq(0.1,1.0,0.2)
[1] 0.1 0.3 0.5 0.7 0.9 1.1 1.3 1.5 1.7
> seq(0.1,2.0,0.2)
[1] 0.1 0.3 0.5 0.7 0.9 1.1 1.3 1.5 1.7 1.9 2.1
> seq(10,1,-1)
[1] 10 9 8 7 6 5 4 3 2 1
> seq(from=1.2,by=0.2,length=6)
[1] 1.2 1.4 1.6 1.8 2.0 2.2
> sequence(c(2,3,5,3,4))
[1] 1 2 1 2 3 1 2 3 4 5 1 2 3 1 2 3 4
> |
```

```
> names_row=c("A1", "A2", "A3")
> rownames(A)=names_row
> A
  [,1] [,2] [,3]
A1    1    2    7
A2    4    5    9
A3    6    8    3
> solve(A)
      A1      A2      A3
B1 -1.39024390  1.21951220 -0.41463415
B2  1.02439024 -0.95121951  0.46341463
B3  0.04878049  0.09756098 -0.07317073
> A%*% solve(A)
      A1      A2      A3
A1  1.000000e+00  0.000000e+00  0.000000e+00
A2 -6.661338e-16  1.000000e+00 -2.220446e-16
A3 -2.220446e-16  7.21645e-16  1.000000e+00
> round(A%*% solve(A))
      A1 A2 A3
A1  1  0  0
A2  0  1  0
A3  0  0  1
> diag(A)
[1] 1 5 3
> diag(1,3,3)
  [,1] [,2] [,3]
[1,]  1  0  0
[2,]  0  1  0
[3,]  0  0  1
> |

> rowSums(A)
A1 A2 A3
10 18 17
> colSums(A)
B1 B2 B3
11 15 19
> rowMeans(A); colMeans(A)
      A1      A2      A3
3.333333 6.000000 5.666667
      B1      B2      B3
3.666667 5.000000 6.333333
> |
```



DAY - II - 16th June, 2020



DAY 2: 16/06/2020

DESCRIPTIVE STATISTICS AND BASIC DATA VISUALIZATION IN R

#loading/importing data in R
`lbd=read.csv("M:\\Desktop contents\\KBN Webinar\\lbw.csv", header=TRUE)`

#listing first six and last six records of the dataset
`head(lbd); tail(lbd)`

##Accessing the objects/variables in the database

DAY - III - 17th June, 2020

Sindhura Lakshmi Mathemsetty joined



Fasna K left



Radhika Chivukula

Arun Kumar Saripalli

Vishnu Vardhan R

Saroj Kumar Ghosh

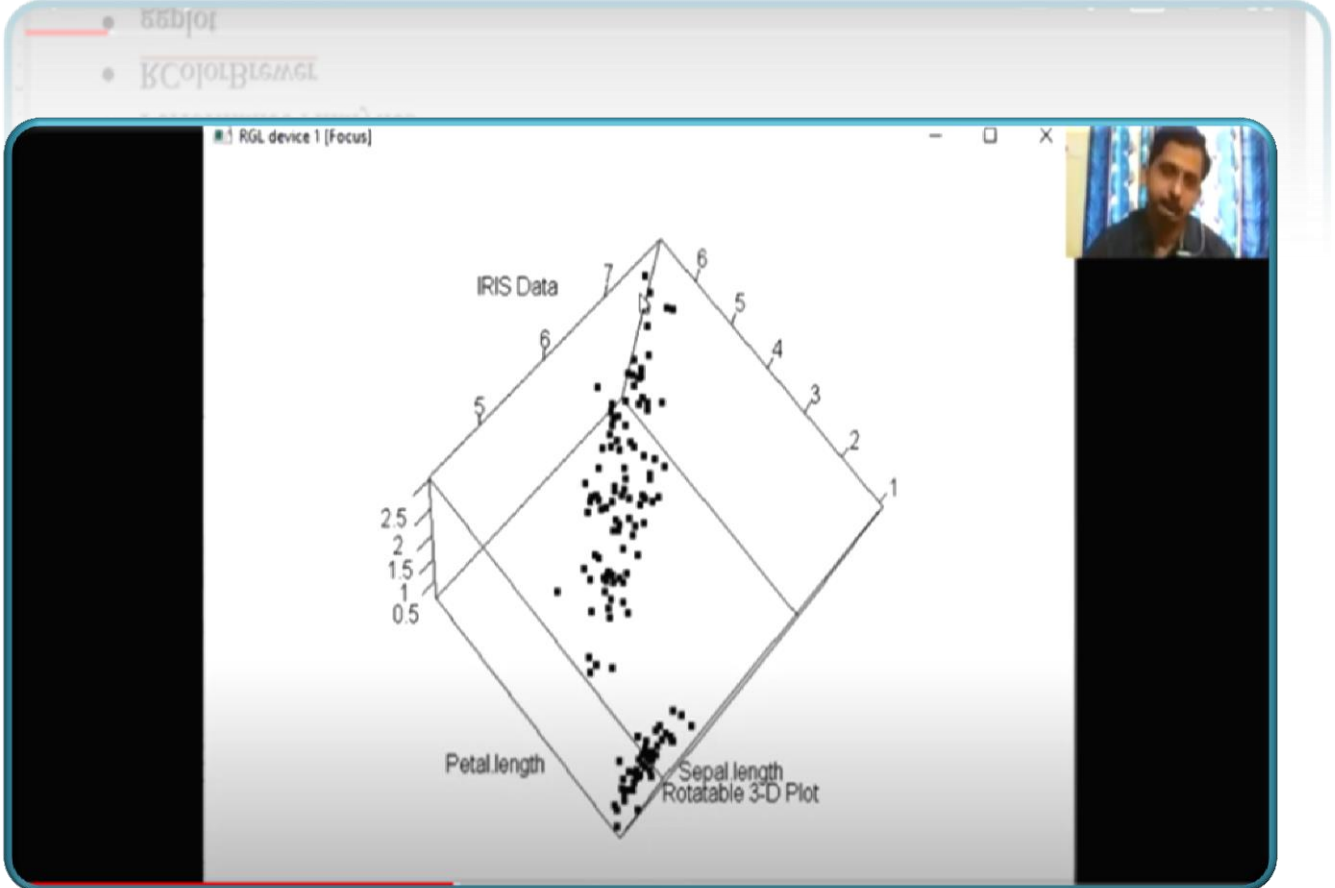


Day 3- 3D data Visualization - Word (Product Activation Failed)

File Home Insert Design Layout References Mailings Review View Font Reader PDF Tell me what you want to do...


##3D Data Visualization##

- ggplot2
- Lattice
- rgl
- plotly
- GGally
- easyGplot2
- scatterplot3D
- Rcmdr, ez
- corrplot
- Performance Analytics
- RColorBrewer
- ggplot



DAY - IV - 18th June, 2020

Kethana Nikitha left



From Peerzada Zubair to Everyone
Good morning everyone



REC

Radhika Chivukula

Vishnu Vardhan R



KBN College Webinar

PAVANI TADAVARTHI



KBN College Webinar

PAVANI TADAVARTHI

Correlation and Regression - Word (Product Activation Failed)

File Home Insert Design Layout References Mailings Review View Foxit Reader PDF Tell me what you want to do...

Clipboard Font Paragraph Styles Editing

##Day 4: Multiple Linear Regression

```

library(readxl)
mlr=read_excel("M:\\Desktop contents\\KBN Webinar\\Data sets.xlsx",sheet = 3)
#sp=read_excel("M:\\Desktop contents\\KBN Webinar\\multiple.xlsx",sheet = "Sparrow")

library(corrplot)
S <- cor(mlr, method="pearson")
round(S,3)
corrplot(S, method="circle")
corrplot(S, method="circle", type = "upper") ## method: ellipse, shade
corrplot.mixed(S, lower.col = "black", number.cex = .8)

```

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Environment History Connections

Global Environment

S num [1:5, 1:5] 1 0.701 0.85 -0.56 -0.813 ...

Files Plots Packages Help Viewer

```

5 plot(mlr) ## matrix scatterplot
6 library(corrplot)
7 S = cor(mlr, method="pearson")
8 round(S,3)
9 corrplot(S, method="circle")
10 corrplot(S, method="circle", type = "upper")
11 corrplot.mixed(S, lower.col = "black", number.cex = .8)
12 corrplot.mixed(S, lower="ellipse", upper="circle")
13 corrplot.mixed(S, lower="number", upper="circle")
14

```

Console Terminal Jobs

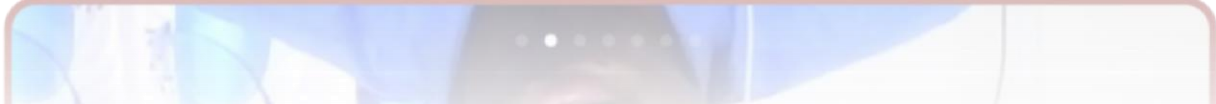
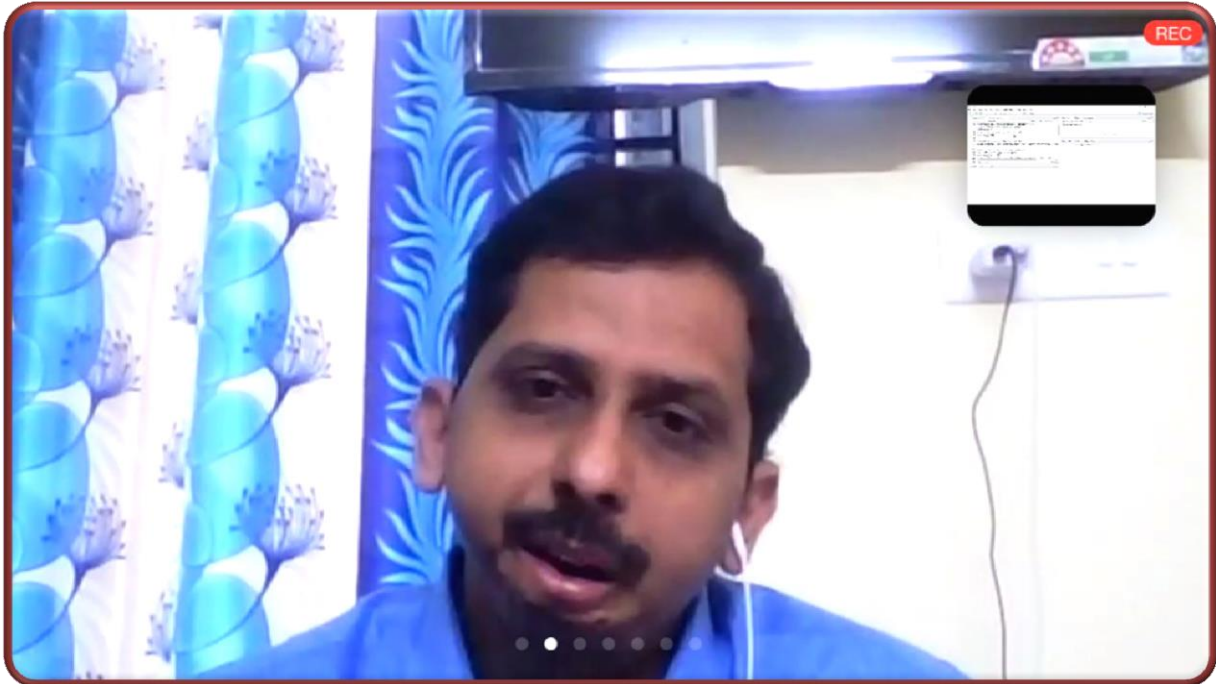
```

> cor(mlr)
X2 0.850 0.260 1.000 -0.193 -0.963
X3 -0.560 -0.835 -0.193 1.000 0.064
X4 -0.813 -0.253 -0.963 0.064 1.000
> corrplot(S, method="circle")
> corrplot(S, method="circle", type = "upper") ##
method: ellipse, shade
> corrplot.mixed(S, lower.col = "black", number.cex
x = .8)
> corrplot.mixed(S, lower="ellipse", upper="circle")
> corrplot.mixed(S, lower="number", upper="circle")

```

Y	X1	X2	X3	X4
1	0.701	0.85	-0.56	-0.813
0.7		0.26	-0.193	0.064
0.85			0.064	1.000
-0.56				-0.813
-0.813				

DAY - V - 19th June, 2020



RAJITA ARVINI joined

Radhika Chivukula

Vishnu Vardhan R

Sundaram

Siva G

A grid of four video call participants. The top-left participant is Radhika Chivukula, with a red slash icon indicating she is muted. The top-right participant is Vishnu Vardhan R, shown in a smaller video window. The bottom-left participant is Sundaram, and the bottom-right is Siva G. A notification at the top says 'RAJITA ARVINI joined'. A row of five dots is at the bottom.

RStudio interface showing R code for data reading and plotting:

```

1 nld=read.csv(file.choose(), header = T)
2 # and attach the data
3 attach(nld)
4 # ask for a summary of the data
5 summary(nld)
6
7 # make a plot of Age vs. Length
8 plot(Length, Age, pch=16,col="red",main="Polynomial Reg
9
10 # now, let's fit a linear regression
11 method1 <- lm(Age ~ Length)
12 summary(method1)
13 # and add the line to the plot make it thick and red
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

```

The Environment pane shows "Global Environment" and "Environment is empty".

RStudio interface showing R code for generating random samples and plotting histograms:

```

14 geom_histogram(bins=10,fill="steel blue")
15
16 ##random samples from Normal Distribution
17 rnorm(100)
18 nd=rnorm(5000,mean = 0, sd=1)
19 ggpplot()+aes(x=nd) +
20 geom_histogram(bins=100,fill="steel blue")
21
22
23 #binomial distribution
24 rbinom(25, size = 1, prob = 0.5)##Bernouli
25
26
27
28
29
30
31

```

The Environment pane shows "Global Environment".

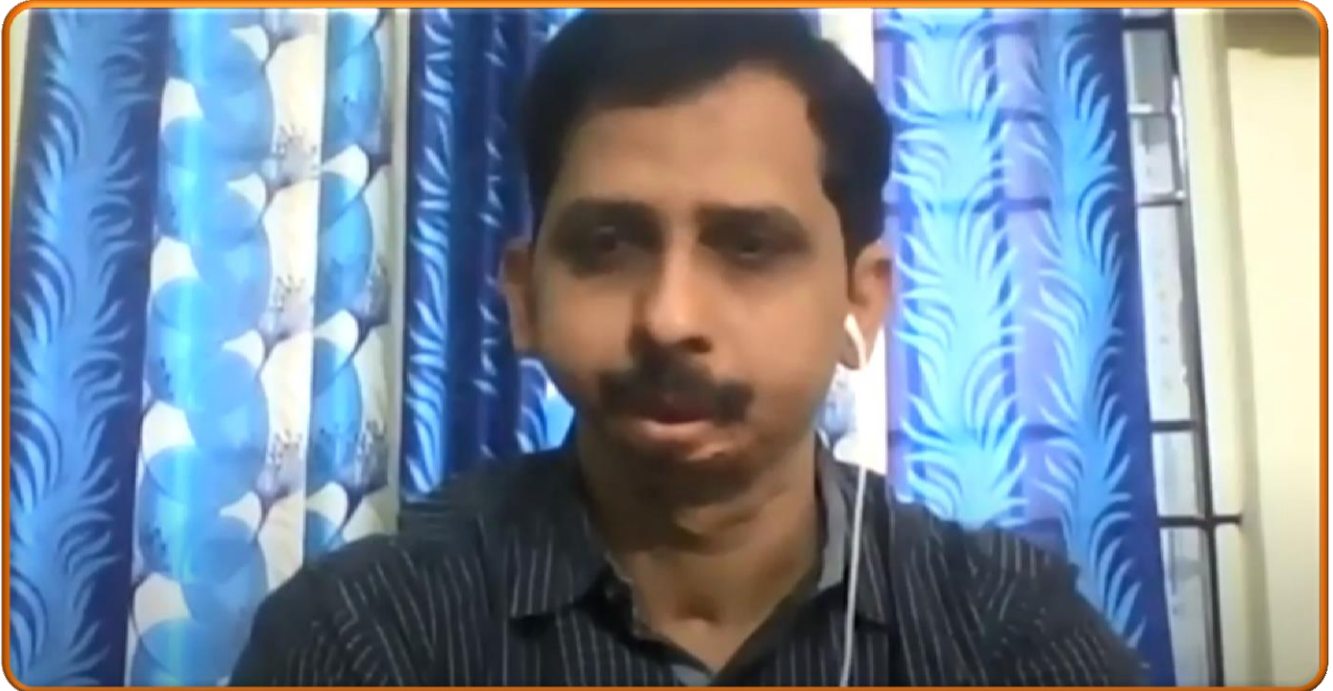
The Plots pane shows a histogram with the following data:

Index	Value 1	Value 2	Value 3	Value 4	Value 5
[166]	8.758277	7.587959	1.157891	1.584862	2.858398
[171]	1.546274	4.774019	6.207907	5.197579	1.557689
[176]	1.302178	4.036713	5.951854	7.672099	7.900190
[181]	2.592798	3.855478	2.218365	9.405516	4.489028
[186]	2.021774	8.993772	8.949835	3.725872	8.253885
[191]	5.980897	2.375709	4.523288	7.007648	5.327008
[196]	1.945211	3.722265	7.740510	3.881422	3.729611

The Plots pane also shows a histogram with the following data:

Index	Value 1	Value 2	Value 3	Value 4	Value 5
[780]	1.842577	3.155302	1.140270	3.887455	3.153877
[787]	2.880881	3.852108	4.253588	1.001848	2.351008
[788]	3.037114	8.883115	8.848832	3.152815	8.523882
[789]	5.885188	3.882418	3.578382	8.402278	4.488058
[790]	1.305718	4.036713	2.827824	1.813088	1.800780

DAY - VI - 20th June, 2020



```
14- f = function(d, i) {  
15-   d2 = d[i,]  
16-   means = tapply(sleep$Bone_Strength, sleep$AgeGrp, mean)  
17-   return(means[1]-means[2])  
18- }  
19- bootDiffMean = boot(sleep, f, R=500, strata=sleep$AgeGrp  
20- bootDiffMean  
21-  
22- sleep=read.csv("M:\\Desktop contents\\KBN Webinar\\wor  
23-  
24- library(wboot)  
25- boot.two.bca(sleep$Young, sleep$MA, mean, stacked = FAL  
26-   conf.level = 0.95)  
27-  
(Top Level) >
```

Graphs data - Excel (Product Activation Failed)

File Home Insert Page Layout Formulas Data Review View Tell me what you want to do...

Male, Female

	A	B	C	D	E	F	G	H	I
1	Age	heartrate	sex	Chestpain			Variables	Type of Variable	Labels
2	67	160	Female	Moderate			Age	Continuous	
3	64	105	Male	High			Heart Rate	Continuous	
4	74	121	Female	Low			Sex	Categorical	Male, Female
5	65	140	Male	High			Chestpain	Categorical	Low, Moderate, No Pain an 1-Low;2-Modera
6	59	161	Male	High			270		
7	53	111	Male	High					
8	44	180	Male	Moderate					
9	57	159	Female	High					
10	71	125	Female	High					
11	64	144	Male	No Pain					
12	40	178	Male	No Pain					
13	48	180	Male	Low					
14	43	181	Male	High					
15	47	143	Male	High					
16	54	159	Female	Low					
17	48	139	Female	Moderate					

17	48	139	Female	Moderate
16	54	159	Female	Low
15	47	143	Male	High



Closing Remarks by Smt. Ch. Radhika, Head, Dept. of Statistics & Convenor

UGC Sponsored Six-Day National Webinar on

“Discovering Statistics using R”

15th-20th June, 2020

Organized by: **Department of Mathematics & Statistics**

Report

UGC – Sponsored Six-Day National Webinar on

“Discovering Statistics using R”

Organized by

Department of Statistics

Dated: 15th to 20th June 2020

Opening Remarks by Vice Principal - Sri. E.Vara Prasad:

In the present situation of COVID-19 pandemic, educational institutions are one of the most affected besides the other areas. To overcome this each and every sector is trying its best to balance the loss occurred, as a result the teaching fraternity across the globe has come out with online teaching resources as a substitute to the learning system. One of such activities is conducting webinars to enhance learning horizons. The university grants Commission have been requesting the academic community to make productive use of the lockdown period. One of the attempts made by the department of Statistics of our college is UGC – Sponsored Six day National Webinar on “Discovering Statistics using R”. I appreciate Department of statistics for coming forward to organize this Six day webinar. I wish u all the learners a very happy learning.

Day-1

Morning Session:

In this session, we have discussed about the concepts like role of statistician and general introduction to R programming which includes the things like installation of R studio along with it, we also learn to work with basic functions and objects in ‘R’ and also about the installation of different packages and went through some aspects like creation and application of various matrix operations using ‘R’. Get them an idea about defining vectors and data about basic operators like logical, relational, assignment operators etc.

Afternoon Session:

In this session, we get to know about the ways of approach for better understanding of given data and reason behind deep understanding and also different statistical tools that we have to possess while tackling with types of data. The perspectives in statistics which we have to learn are analysis like correlation, Regression analysis and classification techniques.

Day-2

Morning Session:

During this session, we discussed about various formats for presenting the data and concepts like descriptive variables and its types, categorical variables etc. The way to compute data with

basic statistics like preparation of frequency tables, Summary statistics, Construction of cross tabulations and computation of different types of proportions.

Afternoon Session:

From this session, we can be able to understand the visualization techniques for better presentation of data graphically, using different tools like histogram, box plots, scatter plots etc and also about certain functions like colour, size etc of that particular mode of presentation.

Day-3

Morning Session:

In this session, we came to know about 'Inferential Statistics'. In drawing conclusions, we learned about the things like, reasons behind the calling descriptive statistics and also about the concept of estimation and its types of point and interval estimations. Have a good idea about Hypothesis testing and its components like null and alternative hypothesis

Afternoon Session:

In this session, we get idea about different types of tests to be used to tackle certain situations with data like one sample t-test, second sample t-test, one way ANOVA. Make them aware of terms like significance, p-value etc. The phases of accepting or rejecting null Hypothesis basing on the derived p-value. The basic concepts like uncertainty and confidence intervals are also covered.

Day-4

Morning Session:

During this session, the concepts like paired sample t-test and two-way ANOVA with suitable examples. Gave Introduction about the concepts of correlation and Regression. Make them better understand about their differences and also gave realistic examples to solve simple correlation and explained about the types of regression that include, simple linear regression and multiple regression, along with an example is also give to do simple linear regression.

Afternoon Session:

Continuation from morning session we came to Know about multiple linear regression with example and got to know the difference between simple and multiple linear regression and also certain assumptions to care of while dealing with those, are discussed. And make them aware of terms like modal and construction of models and also the way to represent modals graphically. Importance of model parameters and beta coefficients are well explained.

Day-5

Morning Session:

In this session, we have discussed about concept of non linear polynomial particularly of 2nd order polynomial and ogive curve polynomial. Also discussed very interesting topics like, how to fit a curvy linear approach kind of data, how to know whether we stick on the 2nd order polynomial or 3rd order polynomial or whether it is fine to use straight line to fit the date when we

have partial curvy linear approach. Generating random numbers with various probability distributions and also get to know the relations between different types of distributions like Binomial distribution, Poisson distribution and Normal distribution.

Afternoon Session:

In this session, we discussed about the areas of using multivariate analysis of variance and what exactly it mean by. The way to handle this multivariate concept using R is better explained. The construction of tables for comparison test and exact approach to understand the outcomes of multivariate analysis are well discussed during this session.

Day-6

Morning Session:

In the session, we have discussed Bootstrapping technique. In statistical knowledge it is a simple notion to understand that building constructing model, the set of inferential statistics using the available data with multiple number of times. In python we applied the technique by using library function BOOT for the data.

In the session, how do we understand the data and what sort of summary statistics we can do and learn how the inferences have to made basing on the data. We have to observe how many variables are present in the data set. Then we have to identify the type of the variable, whether it is nominal, ordinal or categorical.

Valedictory Report

The Valedictory started at 3.45pm on the second day of the Online webinar by In-Charge Principal Sri.E.Vara Prasad. He gave an overall report on the two day online webinar and conveyed heartfelt thanks to all the Resource Persons and the participants which followed with a formal Vote of Thanks proposed by Smt. Ch. Radhika, Convenor of the two day Online Seminar on UGC – Sponsored Six day National Webinar on “Discovering Statistics using R”

Total No of Participants for 6 days	:	Nearly 900
International Level	:	5
From other States	:	nearly 400

YouTube Links

Day	YouTube Link
Day – I (Morning)	https://www.youtube.com/watch?v=JFtDEAYCYbY&feature=youtu.be
Day – I (Afternoon)	https://www.youtube.com/watch?v=TDRNfXBCAUy&feature=youtu.be
Day – II (Morning)	https://www.youtube.com/watch?v=yDTC-LmV-e8&feature=youtu.be
Day – II (Afternoon)	https://www.youtube.com/watch?v=yDTC-LmV-e8&feature=youtu.be
Day – III (Morning)	https://www.youtube.com/watch?v=8r4SW48N31E
Day – III (Afternoon)	https://www.youtube.com/watch?v=u-TbP-YCOL8&t=2s
Day – IV (Morning)	https://www.youtube.com/watch?v=0LYXKHpw2N
Day – IV (Afternoon)	https://www.youtube.com/watch?v=0LYXKHpw2NM
Day – V (Morning)	https://www.youtube.com/watch?v=K6ZSr0FoJCQ
Day – V (Afternoon)	https://www.youtube.com/watch?v=Ok6mIXPfdpk
Day – VI (Morning)	https://www.youtube.com/watch?v=IgiOLVIUf7w
Day – VI (Afternoon)	https://www.youtube.com/watch?v=ob2WmkoSQhs