Summary of 2nd National Workshop on R for Basic Biostatistics at AIIMS, Bhubaneswar

Vikram Kumar Gupta[1], Priya Bansal[2]

Author's Affiliation: [1]Associate Professor; [2]Assistant Professor, Dept of Community Medicine, Dayanand Medical College & Hospital, Ludhiana, Punjab
Correspondence Dr. Vikram Kumar Gupta Email:drvikramgupta1983@gmail.com

Sir,

We had the opportunity to attend the '2nd National Workshop on R for Basic Biostatistics' organized by department of pharmacology, AIIMS, Bhubaneswar. The workshop was of 03 working days from 25th -27th July, 2018[1]. A total of 32 participants attended it. Course fee was Rs 6000/- excluding the staying facility. Faculty members and residents of medical departments of multidiscipline came to attend it both locally and from places over India.

The whole day activity was divided into four sessions including after breakfast, after tea, after post lunch and again continuing after a tea break. . The topics included are listed as follows:

List of Topics

1. Types of data, data presentation and descriptive statistics
2. Inferential statistics
3. GUI intro, navigation, basic functions and data types
4. List, vectors, matrix, data frame and arrays
5. Operations – mean, sum, summary and logical
6. Subsetting, combining and other dataframe operations
7. Apply, lapply, sapply
8. Reading & writing to external data
9. Basic plots, installation of packages
10. Parametric tests for hypothesis testing (normality testing, t-test, ANOVA)
11. Non parametric tests and categorical data analysis (Wilcoxon test, Kruskal Wallis Test, Friedman Test, Chi-square test)
12. Sample size calculation
13. Correlation and linear regression
14. Logistic regression
15. ROC Curve
16. Survival analysis
17. Randomization in R
18. R Markdown
19. RStudio practical with prior installed packages like epiR, dunntest, randomizeR, pROC, OIsurv, ggfortify etc., (Practical exercise for serial no. 3 to 17)

The hallmark of the workshop was hands on training on universally freely available software by CRAN (Comprehensive R Archive Network) i.e., R & free open source license version of RStudio. Presently software version 3.5.1 & 1.1.453 is available for R & RStudio respectively[2,3] and works well on microsoft windows version on and above 7 along with Linux and (Mac) OS X.

R is a programming language used for statistical computing while RStudio uses the R language to develop statistical programs. RStudio however, must be used alongside R in order to properly function and is like an IDE (Integrated Development Environment), it allows users to develop and edit programs in R by supporting a large number of statistical packages, higher quality graphics, and the ability to manage your workspace. R and RStudio are not separate versions of the same program, and cannot be substituted for one another. R may be used without RStudio, but RStudio may not be used without R.[4]

The practical on RStudio coding and command exercises related to it started on first day of workshop. R studio interface have four size adjustable windows

a. Upper Left : 'Source' with 'run' function where one can write and save an R code

b. Lower Left: 'Console' for evaluation which can be emptied by ctrl+L command

c. Upper Right: 'Environment and History' to see objects and tables

d. Lower Right: 'Files/Plot/Packages/Help' for navigating and finding files and shows the graphics made after running various plotting commands

The computer command consisting of *"ctrl+shift+alt+0"* adjusts the four windows. One

can save files in Microsoft excel in '.csv' for loading it in RStudio. It is advised to use only first row in excel sheet for data labelling. RStudio codes data as either numeric with decimal options, character in "" commas, integers meaning the numbers without decimal and logical with either TRUE/FALSE option.

Data structure in RStudio can be homogenous, *vector* as one dimension, *matrix* as two dimension and *array* as n dimensions. The command to read a file is *read.csv ("file_nameame.csv")*. Packages are installed with command *install.packages ("name_of_package")*. The commands to plot data for already available excel sheet may resemble to *(barplot(height=tablesexgroup,width=2,xlim=c(0,6)* which gives beautiful colored graphics with seamless and easy data handling in right lower windows. Various types of statistical tests, regressions, survival analysis can be applied in it. Randomization sequence generation actually generates a particular code as 'seed' which is a number and can be quoted in needful studies as the source of randomization. To export the data, *'write.csv("file_name.csv")* command saves the file in '.csv' of the desired R folder in any of the computer drives.

There are a plethora of packages available on R and guidance files in '.pdf' format are there online.[5,6,7] There are community groups related to it on social media sites like facebook and many videos available to learn on websites like youtube.

Various avenues to use R include biostatistics, simulation, machine learning, bioinformatics, graphics, image analysis augmented by superior graphics capabilities of R. Few disadvantages of R include challenge of learning, building over the code writing slowly over time, being open source with no technical support, need to load package every time.

But in reality, it is like magic software and problems can be searched online on various search portals and one finds links available for the solutions easily.

Medical field personnel need to stand to challenge of R and its embracement.

REFERENCES

1. 2nd National Workshop on R for Biostatistics – NAWORB. Available at: https://sites.google.com/aiimsbhubaneswar.edu.in/aiims-bbsr-pharm-r-wksp/national-workshop-on-r-for-basic-biostatistics. Accessed on July 12th , 2018.

2. The Comprehensive R Archive Network. Available at: https://cran.r-project.org/ Accessed on July 15th , 2018.

3. R Studio. 250 Northern Ave, Boston, MA 02210. Available at: https://www.rstudio.com/products/rstudio/download/#download. Accessed on July 15th, 2018.

4. ST. Olaf Collage. R Vs RStudio. Available at: https://wp.stolaf.edu/it/r-vs-rstudio/ Accessed on July 29th, 2018.

5. CRAN Packages by Date of Publication. https://cran.r-project.org/web/packages/available_packages_by_date.html Accessed on July 25th, 2018.

6. The R Manuals. Available at: https://cran.r-project.org/manuals.html Accesses on: July 25th, 2018.

7. Contributed Documentation for Learning about R. Available at: https://cran.r-project.org/other-docs.html Accesses on: July 25th, 2018.