SHORT COMMUNICATION   **OPEN ACCESS**

Application of Statistical Analysis Software and Packages for Research Data of Agriculture and Allied Sectors – a Short Communication

Pramit Pandit1, Sourav Roy2, Devayan Chatterjee3, Sk Masudul Islam2, Upasana Mohapatra4
1- Department of Agricultural Statistics,Bidhan Chandra Krishi Viswavidyalaya, Mohanpur, India
2 -Department of Vegetable Science, Bidhan Chandra Krishi Viswavidyalaya, Mohanpur, India
3 -Department of Agricultural Extension, Bidhan Chandra Krishi Viswavidyalaya, Mohanpur, India
4 -Department of Plant Biotechnology, University of Agricultural Sciences, Bengaluru, India
Email: pramitpandit@gmail.com

ABSTRACT
Numerous research works are carried out every year for the development of Agriculture and allied sectors in India. As a consequence, huge amount of data are generated and in need of proper statistical analysis for the sake of integrity of the inference drawn on the basis of experimental data. Hence, an effort has been made to discuss various statistical software used in analysis of agriculture and allied sectors' research data.
Keywords: Agriculture, Research data, Software, Statistical Analysis.

INTRODUCTION
Agriculture is considered as the backbone of India. More than 1,00,000 agricultural scientists have been working in different ICAR (Indian Council of Agricultural Research) institutes including KVKs (Krishi Vigyan Kendras), various SAUs (State Agricultural Universities), IITs (Indian Institute of Technology), state as well as central Government departments, private companies and NGOs (Non-government Organizations) for sustainable development of agriculture and allied sectors as a whole [1]. Numerous research works are carried out every year, generating a huge amount of data subjected to proper statistical analysis. As different experiments are conducted emphasizing different objectives, collected experimental data vary in nature and scale. This variation leads to the requirement of different statistical methods and tools, from univariate distribution fitting to multivariate ANOVA and so on. Under these circumstances, various statistical software and packages have been developed for proper statistical analysis for the sake of integrity of the inference drawn on the basis of experimental data. Hence, an effort has been made to discuss various statistical software and packages used in data analysis of agriculture and allied sectors.

SOFTWARE USED IN ANALYSING RESEARCH DATA OF AGRICULTURE AND ALLIED SECTORS
Out of various options available for offline and online data analysis, MS Excel, SPSS, R, SAS,Tableau, Minitab,CropStat,SPAR 2.0,SPAD,SPFE 1.0,SPBD Release 1.0,SPAB 2.0,SSDA1.0, WASP 2.0, OPSTAT and online analysis platforms available at IASRI (Indian Agricultural Statistical Research Institute) is most popular [1, 2].

MS Excel
There are plenty of statistical functions and data analysis tools available in MS Excel. Descriptive statistics, fitting of distributions and some important parametric tests (t test, F test etc.) can be done with statistical functions. However, by using 'Analysis Tool Pack' of MS Excel, ANOVA for single as well as multiple factors, regression, data summary can be computed directly for the given set of data.

SPSS

SPSS (Statistical Package for the Social Sciences), developed by IBM, offers advanced statistical analysis, a vast library of machine-learning algorithms, text analysis, open-source extensibility, integration with big data and seamless deployment into applications [3].It is one of the widely used statistical packages in the field of social science and human behaviour research. Descriptive statistics, various dimension reduction and classification techniques, parametric as well as non-parametric tests, regression and correlation analysis tests etc. can be performed with ease by dint of this user-friendly software.

R

R is a free software environment specifically meant for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS.R along with its vast library help widely to implement time-series analysis, classification techniques, clustering, classical statistical tests, linear and nonlinear modelling etc. Another notable thing about R is its active community, members of which are always engaged in building packages and improving R and the associated plug-ins incessantly. Many of R's standard functions are written in R itself, making it more user-friendly to follow the algorithmic choices made.

SAS

SAS (previously, "Statistical Analysis System") is a software suite developed by SAS Institute for advanced analytics, predictive analytics, multivariate analyses, data management and business intelligence. Using the Output Delivery System, it is possible to publish SAS data in Excel,HTML, PDF and other formats. The SAS Enterprise Guide is SAS's point-and-click interface, generating code with a view to manipulate data or to perform analysis automatically.

Tableau

Tableau is a powerful and fastest growing data visualization tool used in the Business Intelligence Industry. It substantially helps to simplify raw data into the easily understandable formats. Visualizations created by Tableau are in the form of dashboards and worksheets. Some best features of Tableau are: data blending, real time analysis and collaboration of data. It is even suitable for non-technical users to create customized dashboards. Tableau has garnered interest among the people from all sectors such as business, research, industries etc.

Minitab

Minitab is a statistical package developed by the researchers of Pennsylvania State University in the year of 1972. It began as a light version of OMNITAB 80. Commands can be executed through either GUI orscripted commands, making it friendly to both novices as well as experts. Few other software, produced by Minitab, Inc., can also be used in conjunction with Minitab.

CropStat

CropStat offers several features with a view to analyse data from agricultural field trials and several other sources. It offers descriptive statistics, scatter plot graphics, linear as well as mixed modeling, repeated measure ANOVA, combined analysis of variance, single-site analysis for variety trials, spatial analysis, pattern analysis, regression and correlation analysis, genotype × environment interaction analysis, QTL analysis and few more salient features. In some cases, data management is done after balanced analysis of variance. Any 32-bit Windows operating system is a pre-requisite to run CropStat.

SPAR 2.0

Primarily, Statistical Package for Agricultural Research Data Analysis was developed at IASRI for statistical analysis of experimental research data in Plant Breeding and Genetics using Fortran Language. Some salient features of this are: (i) Data Management,(ii) Descriptive Statistics (iii) Variance and Covariance Components Estimation (iv) Correlation and Regression Analysis (v) Estimation of Breeding values (vi) Mating Design Analysis (vii) Multivariate Analysis (viii) Stability Analysis. However, this is not suitable for large data handling.

SPAD

Statistical Package for Augmented Designs (SPAD) has been developed at IASRI. This Package is useful for generating randomized layout and performing the analysis of the generated data for the agricultural experiments conducted for the purpose of comparison between existing practices/check

varieties/controls and new practices/varieties/germplasm collections/tests, where the experimental material is scarce.

SPFE 1.0
SPFE 1.0 (Statistical Package for Factorial Experiments) essentially generates the designs and performs analysis for symmetrical and asymmetrical factorial experiments with and without confounding. User inputs as a list of independent interactions to be confounded is required. It is also suitable for partial confounding. Regular fractional factorial plans for symmetrical factorial experiments can also be generated by using this package.

SPBD Release 1.0
SPBD Release 1.0, a Statistical Package for Balanced Incomplete Block Designs, was developed at IASRI. This software is considered user-friendly and complete in every aspect, enabling experimenters to select, generate and provide randomized layout of a BIBD design. This package generates BIB designs with number of replication up to a maximum of 20 for asymmetric BIB designs and 30 for symmetric BIB designs.

SPAB 2.0
SPAB 2.0 (Statistical Package for Animal Breeding) was developed for data generated in Animal Breeding and Animal Genetics research. The package is Window based, Menu driven and user-friendly as well.

SSDA1.0
SSDA 1.0 (Software for Survey Data Analysis) is useful for the analysis of data collected by simple random sampling (SRS), stratified random sampling, systematic sampling, probability proportional to size (PPS), cluster sampling, two stage and stratified two stage sampling schemes. Estimates of population mean, variance and design efficiency of the sampling scheme can be obtained by SSDA 1.0.

WASP 2.0
WASP2.0 (Web Agri Stat Package) is the first Web Based Agricultural Statistics Software Package, designed and developed by ICAR-Central Coastal Agricultural Research Institute. This package is for the researchers who are not comfortable with standalone statistical packages [2]. Descriptive statistics, analysis for basic experimental designs, different parametric and non-parametric tests can be performed comfortably by this statistical package.

OPSTAT
OPSTAT is a free online agriculture data analysis tool developed by O. P. Sheoran, Computer Programmer at CCS HAU, Hisar, India. One factor analysis, two factor analysis, three factor analysis, correlation and regression analysis, generation means analysis, diallel analysis, path analysis, line × tester analysis, partial diallel analysis and stability analysis can be easily done with OPSTAT.

ONLINE ANALYSIS PLATFORMS AVAILABLE AT IASRI WEBSITE
Data analysis facilities available online for agricultural scientists of NARS in India are presented in table 1.

Table 1: ONLINE ANALYSIS OF DATA AT IASRI WEBSITE

Purpose	URL
Block Designs Analysis (General)	http://www.iasri.res.in/WebAnalysis/index.aspx
Augmented Block Design Analysis	http://www.iasri.res.in/spadweb/default.aspx
Survey Data Analysis	http://nabg.iasri.res.in/ssda2web/. IP
Row Column Design Analysis	http://iasri.res.in/css/Home.aspx
Web Generation of Experimental Designs Balanced for Indirect Effects of Treatments	http://iasri.res.in/webdbie
Statistical Package for Factorial Experiments	http://iasri.res.in/spfe/home.aspx
Web Based Fuzzy Clustering Software	http://proj.iasri.res.in/wfcm
Online Decision Tree Classification using C4.5	http://proj.iasri.res.in/odtc
Web Solution of Estimation of Compound Growth Rate	http://iasri.res.in/cgr
Web Generation of Polycross Designs	http://design.iasri.res.in/webpd

CONCLUSION
There are various statistical software and packages available to the researchers. However, choice of appropriate statistical software would depend on the analysis required and the technical expertise of the user.

REFERENCES
1. Jayade, K.G., Deshmukh, P.D.&Khot P.G. (2015). Statistical Analysis Software for Agricultural Research Data Analysis. International Journal of Advanced Research in Computer Science and Software Engineering, 5(3): 885-890.
2. Gupta, N.K., Isaac R.K. & Singh, R.K. (2016). Maintenance and analysis of agricultural data: a challenge. International Journal of Bioassays,5(9): 4842-4848.
3. Green, S.B.& Salkind, N.J.(2016). Using SPSS for Windows and Macintosh, Books a la Carte (8th Edition), Pearson,New York.