INDIAN NARS STATISTICAL COMPUTING PORTAL
http://stat.iasri.res.in/sscnarsportal

INDIAN AGRICULTURAL STATISTICS RESEARCH INSTITUTE
(INDIAN COUNCIL OF AGRICULTURAL RESEARCH)
LIBRARY AVENUE, PUSA, NEW DELHI – 110 012, INDIA

www.iasri.res.in

Please visit: www.iasri.res.in/sscnars; www.iasri.res.in/design

NDIAN NARS STATISTICAL COMPUTING PORTAL has been designed, developed and implemented under NAIP Consortium on Strengthening Statistical Computing for NARS with IASRI, New Delhi as lead centre and eight other consortium partners as National Academy of Agricultural Research Management, Hyderabad; National Dairy Research Institute, Karnal; Indian Veterinary Research Institute, Izatnagar; Central Institute of Fisheries Education, Mumbai; Directorate of Water Management, Bhubaneshwar; ICAR RC North-Eastern Hilly Regions, Barapani; Maharana Pratap University of Agriculture and Technology, Udaipur and GKVK University of Agricultural Sciences, Bengaluru. Using IP authentication, this portal provides service oriented computing for statistical analysis of data generated by researchers in the Indian National Agricultural Research System (NARS) users. The Portal is dynamic in nature and new additions/ updates are being made from time to time, to strengthening it further. The financial support for this portal was from National Agricultural Innovation Project(NAIP) of Indian Council of Agricultural Research (ICAR). Major contributors are Dr. Rajender Parsad and Dr. A. Dhandapani. The others involved in various stages of development, testing and implementations were Shri RS Tomar, Shri Sachin Kumar, Shri Manoj Khandelwal, Ms Sanju, Shri Subhash Chand, Shri Rakesh Saini and Shri Pramod Kumar from IASRI, New Delhi and Shri Badshah Mukerjee, Shri Jay Shah, Shri Santosh Tiwari, Shri Kushal Ramani, Shri Satyajit Dwivedi and Shri Ajit Sharma from SAS India Pvt. Ltd.

Copyright

How to Quote

Indian NARS Statistical Computing Portal (http://stat.iasri.res.in/sscnarsportal): <Name of the analysis performed. <Date of analysis performed>

INDIAN NARS STATISTICAL COMPUTING PORTAL

1. Introduction

Statistical analysis is an important tool to extract as much information as possible from given data. Statistical computing methods enable to answer quantitative biological questions from research data and help to plan new experiments in a way that the amount of information generated from each experiment is maximized. Applications of appropriate statistical techniques helps in drawing statistically valid inferences from a large amount of data generated/ collected by researchers. The complexities involved in choosing appropriate statistical analysis, computing issues and lack of knowledge of advanced analysis force many researchers to adopt simple analytical tools. The Indian NARS Statistical Computing Portal is an attempt to remove these bottlenecks and to enable researchers of NARS to perform online statistical data analysis using appropriate and advanced statistical methods.

Indian NARS Statistical Computing Portal

- hosted at http://stat.iasri.res.in/sscnarsportal

- provides service oriented computing (paradigm of computing techniques that operate on software-as-a-service) to Indian NARS researchers through IP authentication

- provides a single, easy-to-use interface for performing statistical analysis of data across the NARS organizations

- does not require installation of statistical packages for analysis of data pertaining to the available modules on the users' machine

- a rich e-learning resource consisting of reference manuals, sample data sets, worked out examples and video presentations

- complemented with web page of Strengthening Statistical Computing for NARS www.iasri.res.in/sscnars and Design Resources Server www.iasri.res.in/design

- is dynamic in nature and new additions/updates are being made from time to time

- would help scientists, research scholars and students in performing statistical analysis of their data and would save on cost and time for NARS researchers

- would enable to develop applications for customized modules for All India Co-ordinated Research Projects/ Network Projects

2. Access to the Portal

The service oriented computing feature of Indian NARS Statistical Computing Portal is restricted for use by authorized personnel from Indian NARS organizations and a two-level authorization is being followed viz. IP Authentication and User ID-Password.

2.1 IP authentication

The nodal officers of Strengthening Statistical Computing for NARS (SSCNARS) have been asked to provide the static IP addresses of their respective organizations and the same have been included in the authorized list for accessing the portal. As of now, there are 136 organizations in the authenticated list. The requests from such IP addresses to the portal will be allowed and a login screen is shown as
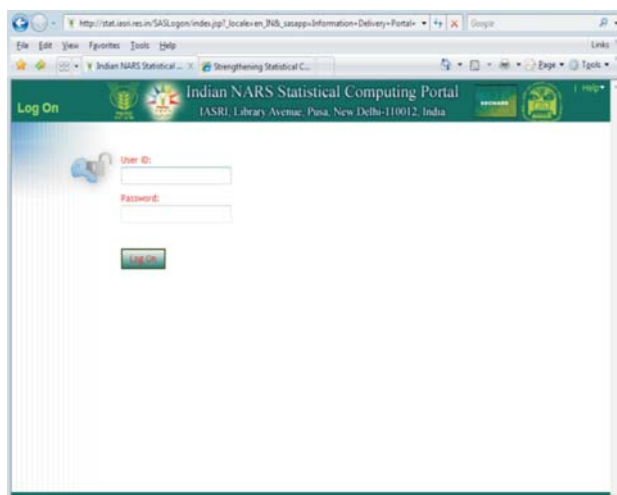

Login Screen

2.2 User-level authorization

Any user from IP authenticated machines of NARS organizations can log in to the portal either as (i) Guest user or (ii) Registered user depending upon the rights provided to the user. The data uploaded by a guest user are not saved on the server whereas the data uploaded by a registered user is automatically saved. To become a registered user of the portal, one can request through their nodal officer or by directly contacting the portal administrator depending upon the need. A common "Guest User" account has been created for those who wish to access the portal frequently for research and teaching with following User ID and Password (both are case sensitive):
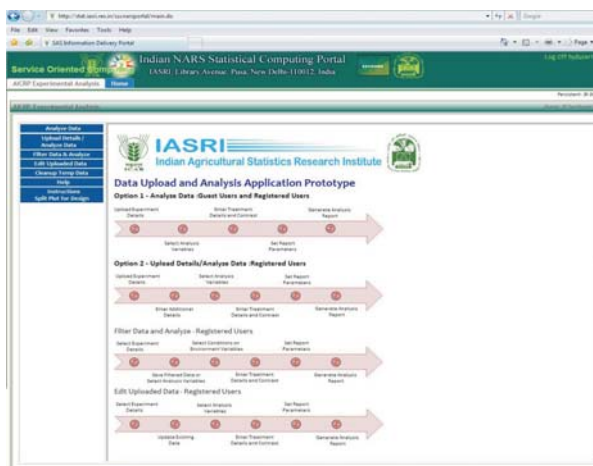
User ID: -------------

Password:------------

On successful log in, the guest user and registered users are respectively directed to the following screens
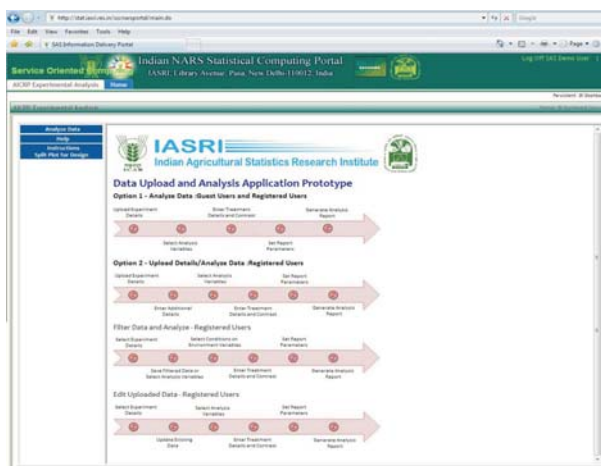
Home Page of Guest User


Home Page of Registered User

3. Statistical Analysis

The following online statistical analyses are presently available on the portal:

- Block Designs (complete and incomplete block designs)
- Resolvable Block Designs
- Augmented Block Designs
- Combined Block Designs
- Row-Column Designs
- Nested Block Designs
- Split Plot Designs
- Univariate Distribution Fitting
- Test of Significance
- Correlation
- Regression Analysis
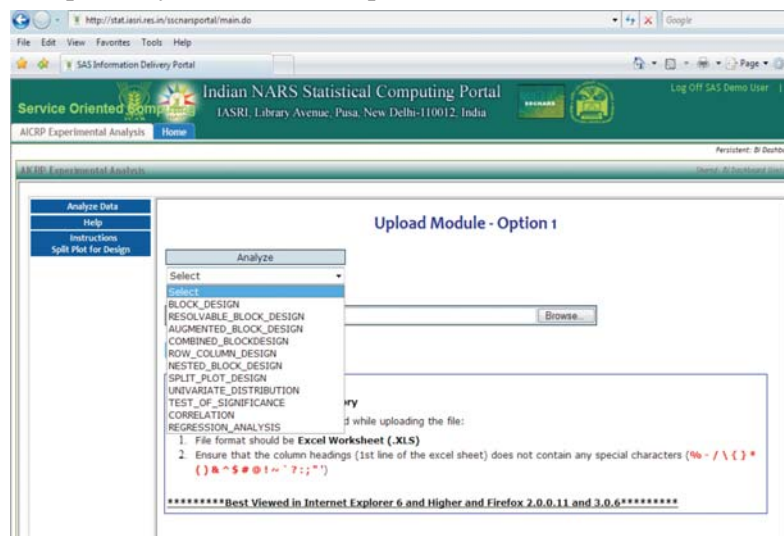
First step of Statistical Analysis

More analyses modules are being added to this list.

3.1 Uploading data into portal

Both the Registered users and Guest users must upload the data into the portal as MS-Excel file for performing any of the available analysis. Only the MS-Excel files with extension .xls (i.e. Excel 97/2003 format) are allowed and at present, MS-Excel 2007/2010 file format is not supported. However, before uploading the file with .xls extension the users must check to their satisfaction that their files satisfy the following conditions:

- As far as possible, the file being uploaded should have only one worksheet. In case of files with multiple worksheets, only the worksheet with name appearing first in lexicographic order gets uploaded.
- For missing values, leave the missing observations as blank cells. Do not put period (.) or Zero (0) to display missing values in the data.
- Do not use special characters in the variable/column names. Also variable names should not start with spaces.
- Do not use any expressions to the cell values. It should be data value only.
- The classification variables can be both numeric and character except for Augmented Designs.
- In case of errors/ wrong results, check Excel file again for blank rows/columns or extra characters. Go to first column, first cell and then press Ctrl+Shift+End. It will select all the filled rows and columns. If it selects some blank rows and columns then delete those rows, columns, also delete extra characters, if any.
- If the first row cells have been merged then it will not detect it as Column/Variable names.
- If any row(s) or column(s) are hidden then it will be displayed during the analysis.

Besides these conditions, the data file must satisfy the requirements of the statistical analysis to be performed. The guidelines for preparing data files for different statistical analysis are given in the sequel.

3.2 Steps for preparing data files for different statistical analysis

Block Designs: Prepare a data file containing two Classificatory variables to describe the Block and Treatment details. There should be at least one response/ dependent variable in the experimental data to be analyzed. Also, the block/treatment details may be coded or may have actual names (i.e. data values for variables describing block and treatment column may be in numeric or character). The maximum length of treatment value is 20 characters. The variables can be entered in any order. (These conditions are applicable to other experimental designs also)

Resolvable Block Designs: The data file should contain three Classificatory variables to describe the Replication, Block, Treatment variables and at least one Dependent/ response variable.

Augmented Block Designs: The data file should contain two Classificatory variables to describe Block & Treatment variables and at least one Dependent variable. At present, Portal supports only numeric treatment and block variables for augmented designs. An augmented block design involves two sets of treatments known as check / control and test treatments. The treatments should be numbered in such a fashion that the check / control treatments are numbered first followed by test treatments. For example, if there are 4 control treatments and 8 test treatments, then the control treatments are renumbered as 1, 2, 3, 4 and tests are renumbered as 5, 6, 7, 8, 9, 10, 11, 12.

Combined Block Designs: The data file should contain three Classificatory variables to describe Environment, Block, Treatment variables and at least one Dependent variable.

Row-Column Designs: The data file should contain three Classificatory variables to describe Row, Column, Treatment variables and at least one Dependent variable.

Nested Block Designs: The data file should contain three Classificatory variables to describe Block, SubBolck, Treatment variables and at least one Dependent variable.

Split Plot Designs: The data file should contain three Classificatory variables to describe Replication, Main Plot, Sub Plot variables and at least one Dependent variable.

Univariate Distribution Fitting: The data file should contain at least one quantitative numeric variable.

Test of Significance: The data file should contain at least one quantitative variable and one classificatory variable.

Correlation: The data file should contain at least two quantitative variables.

Regression Analysis: The data file should contain one Dependent and at least one Independent variable.

More details for data file preparation are available from Help Link Button.

3.3 Steps for performing statistical analysis

After successful log on to the portal, the user must select the statistical analysis required and upload the data file to be used for the analysis to start the process of analysing the data. Upon successful upload of the data file, the user is allowed to proceed for the subsequent steps in the selected statistical analysis. The users are advised to check the information provided about the data file by the portal in the second screen, particularly the number of rows and names of the variables. Also, they must ensure that all the variables are available for the users to select role(s) for the variable. In case of any discrepancy, the users must check the data file once again before proceeding for the analysis.

The workflow for performing a statistical analysis is Upload Experiment Details (Select analysis and Upload data file) → Select Analysis variables → Enter Treatment Details and Contrasts, if any → Set Report Parameters → Generate Analysis Report → Save Analysis Report as *.pdf or *.rtf file.

Depending upon the type of statistical analysis, users must provide/select various options to perform the analysis. Usually, the following information is required:

- Role of the variables (such as analysis variable/dependent variable, classification variables, etc.)

- Labels of the variables (to provide meaningful names to the variables) – optional

- Details of treatments, if required

- Level of significance, multiple comparisons tests, etc. to be used

- Report formats (tables to include)

- Output format (PDF or RTF or both)

Once the above information is entered, one can View the results or download the results as PDF or RTF files.

In case "SAS Stored Process Error" comes, go back and Check the EXCEL file and details provided for any discrepancies as given in section 3.1.

Analysis of data generated from Block Designs: Block designs are useful for controlling the heterogeneity in the experimental units due to one non-treatment factor. Block design may be complete, if all treatments appear in a block or incomplete when there is at least one block in which all treatments do not appear.

One can analyse any block design (complete or incomplete) by selecting the option Block Design from Analyze drop down list. The portal checks automatically whether the uploaded data is balanced or unbalanced (through incidence matrix of the block design). The output will be different depending on the design used for data generation. If the data is balanced, then it displays the standard error of difference and minimum significant difference (for Tukey/LSD options). The group letters on the treatments for significant differences are displayed if chosen by the user. The group letters on treatments are displayed only when the null hypothesis of equality of treatments effects is rejected through analysis of variance (ANOVA). If the data is unbalanced, group letters are displayed on the treatments whenever null hypothesis of equality of treatments effects is rejected through ANOVA, irrespective of whether the user has checked the option of Group Letters (treatment) or not. Some help on guided interpretation is also provided as the analysis report displays the significance at 5% and 1% level of significance. If the null hypothesis regarding equality of treatment effects is not rejected, then non-significant(NS) is displayed against the column. One can also choose ranking the treatments in descending order of the adjusted treatment means/treatment means. The steps to perform the analysis are given in the sequel.

Step 1: Select Analysis as Block Design from the drop down list, upload data file using browse button and click on Next button.

Step 2: Check the number of records and the variables names given on the screen and match it with the original data file. If there are no discrepancies, select one variable each from the variables shown for Block and Treatment under classification variables. Choose one or more variable(s), from the list of variables shown under dependent variable(s). For selecting more than one dependent variable, use Control+Select. Finally, select the multiple comparison procedure to be used (Tukey for Tukey's Honest Significant Difference test, Duncan for Duncan's Multiple Range Test, LSD for Fisher's Least Significant Difference or Critical Difference). It is also possible to enter descriptive label for different variable(s) including units if any (This, however is optional).

Click on Next and proceed to Step 3.
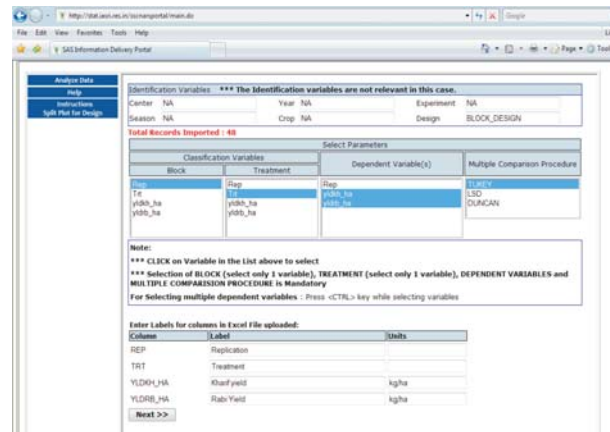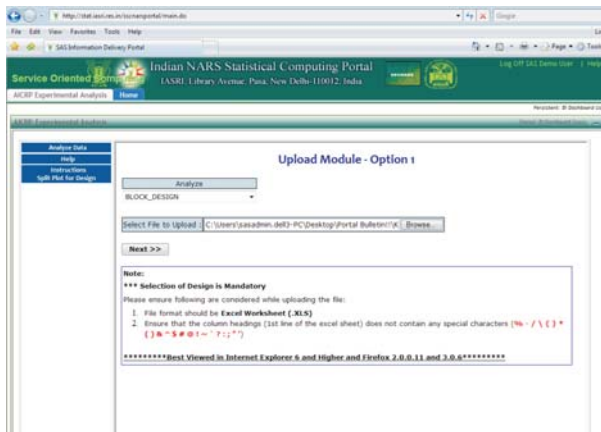
Steps 1 and 2 of Block Design Analysis

Step 3: Fill in the treatment description, if required in output. If the treatment contrast analysis is required, then Check on Enable Contrast box. If treatment contrast analysis is required, enter the label for the contrast and coefficient and click on save button. For testing the hypothesis regarding another treatment contrasts, give the contrast label and coefficients and check that the sum of coefficients is zero. Then Click on Save Button and continue in the similar fashion for testing additional treatment contrasts. Once coefficients for all required treatment contrasts are given and saved, Click on Next button to get the screen for Report Configuration. However, the treatment description and treatment contrasts are optional and to skip this step, click on Next without entering anything in the text boxes.
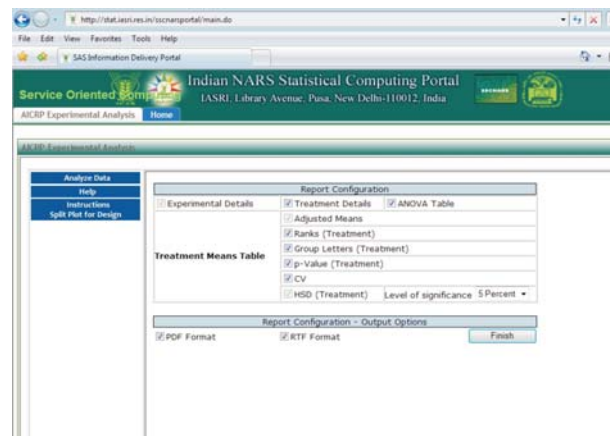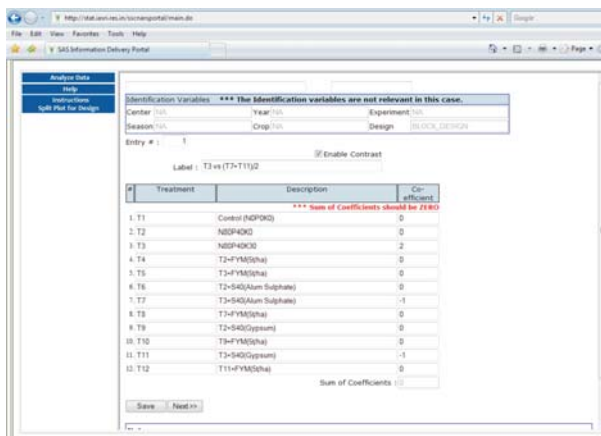
Steps 3 and 4 of Block Design Analysis

Step 4: On the report configuration screen, the adjusted treatment means and minimum significant difference test are selected by default. Users can select Treatment details, ANOVA table, Rank (Treatment) for ranking the treatments, Group Letter (Treatment) for displaying group letters depicting significant differences between treatment effects, p-value for testing the equality of treatment effects, CV (for percent coefficient of variation, level of significance (1% or 5%). Finally choose the format for analysis report file as PDF (Portable Document Format) or RTF (Rich Text Format) or both. Click on "Finish" button to display output tables on the screen which can be saved on computer of the user as *.pdf or *.rtf file.

Analysis Output Report from a Block Design

Analysis of data generated from Resolvable Block Designs: Resolvable block designs are the important class of incomplete block designs wherein the blocks can be formed together into sets with the blocks within each set constituting a complete replication. In resolvable block designs, the between blocks within replication variation helps in reducing the experimental error, increasing thereby the precision of estimation of treatment contrasts of interest. The steps for analysis of a resolvable block design are same as that of a block design except that the data file contains three Classificatory variables viz. Replication, Block nested within replication and Treatment. The classification variables and dependent variable(s) are to be typed by the user or through copy paste from name(s) of the variables displayed. More than one dependent variable should be comma separated.

Analysis of Data from a Resolvable Block Design

Analysis of data generated from Augmented Block Designs: Augmented block designs are useful for the situations in which the experimental materials of new practices/test varieties (called test treatments) are just sufficient to have one replication in the experiment and sufficient experimental materials for standard practices/ check varieties (called control treatments) are available to have replications. The selection of classification variables, dependent variable(s) and multiple comparison procedure are similar to that of block designs. Additionally, the user is required to enter the number of control treatments and number of test treatments. At present, testing the hypothesis regarding user defined treatment contrasts and entry of treatment description are not available. The analysis report includes ANOVA table, adjusted treatment means and minimum significant differences for control treatment means, test treatments in same block, test treatments in different blocks and test versus control treatment.
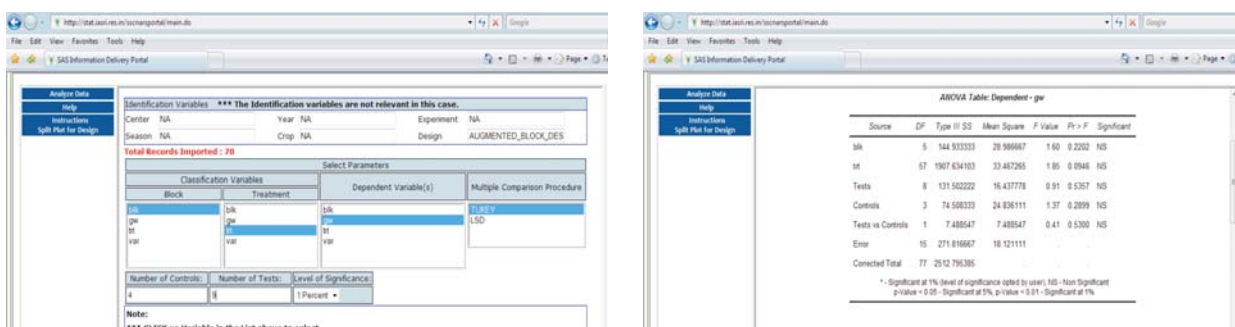
Analysis of Data from an Augmented Block Design

Analysis of data generated from block designs conducted in different environments: To study the performance of treatments across environments, the experiments are conducted in different environments. The option **Combined Block Design** performs analysis of data collected from such experiments and analysis can easily be performed along the lines similar to that of Block Designs through self explanatory screens. Depending upon the experimental situation, the environment effect may be chosen as fixed or random. There is also a provision to drop the non-significant treatment × environment interaction from the combined model, if required. The separate analysis for each of the environments is given along with the combined analysis according to the choice of nature of environment effect as random / fixed. One can also generate SREG (site regression) biplot.
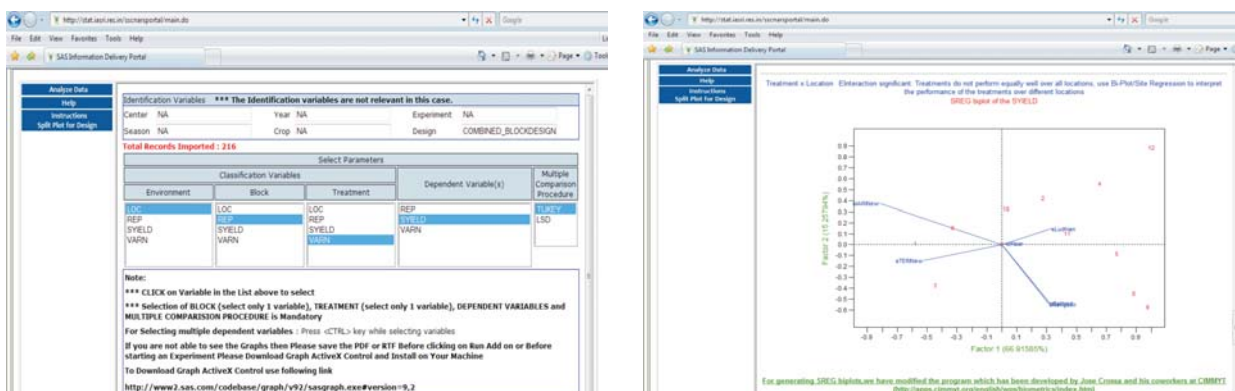
Combined Analysis of Block Designs

Analysis of data generated from Row-Column Designs: Row-Column Designs are useful for the control of non-treatment variability in experimental units arising due to two cross classified sources of heterogeneity. The steps for analysis of a row-column design are similar to that of a block design except that the data file contains three classification variables viz. Row, Column and Treatment. The classification variables and dependent variables are to be typed by the user or through copy paste from name of the variables displayed.
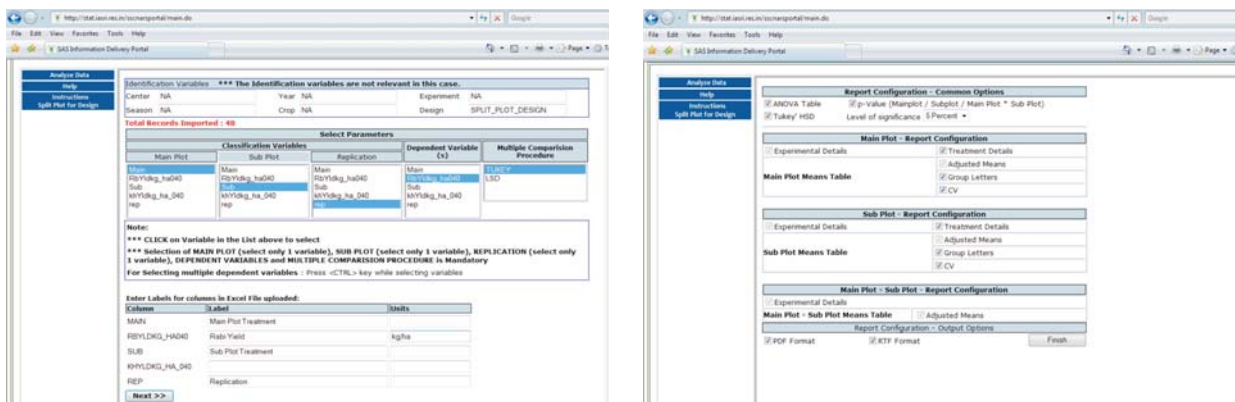

Analysis of Data from a Row-Column Design

Analysis of data generated from Nested Block Designs: A nested block design is defined as a design with two systems of blocks where the second system (called sub-blocks) is nested within the first system (called blocks). These designs have been developed to take care of the heterogeneity in the experimental units or the experimental conditions caused by two non-treatment factors (one factor is nested within the other factor). The steps of analysis are same as that of a resolvable block design except that the replication variable in resolvable block design is represented by block in nested block designs and block variable in resolvable block design is represented by SubBlock.

Analysis of Data from a Nested Block Design

Analysis of data generated from Split Plot Designs: These designs are useful for the experimental situations with two factors in which it is difficult to change the levels of one of the factors due to practical constraints. In field experiments conducted using a split plot design, one set of treatments is assigned to large plots called main plots. Each main plot is subdivided to accommodate levels of another set of treatments called sub plots. Each main plot becomes a block for the sub plot treatments. These designs should only be

used when the treatments applied to main plots cannot be allocated to smaller plots of the sizes that of sub plots. The steps of split plot design are similar to that of a block design. There is a provision of treatment contrast analysis for both main plot and sub plot treatments. The analysis report includes ANOVA table, table of means for main plot × sub plot treatment interaction, marginal means for main plot and sub plot treatments, Standard error of difference and Tukey's HSD/ LSD for main plot treatment comparisons, sub plot treatment comparisons, comparing sub plot treatment at same level of main plot treatment and for comparing main plot treatments at same or different level of sub plot treatment.

Analysis of Data from a Split Plot Design

Univariate Distribution Fitting: The option "Univariate Distribution" in Analyze drop down list can be used for testing whether the data follows Normal, Log normal, Exponential, Weibull, Beta distribution or not. For testing whether a given data follows a specified distribution, user has to enter the names of the variables in the text box for analysis of variables from the variables list shown on the screen. More than one variable should be entered as space separated. If distribution fitting is required for several groups, group variable in the Classification variable text box is required. A radio button is provided to select the distribution to be fitted. One can also specify the desired values of secondary parameters (Theta, Sigma & Alpha) or they would be estimated automatically.
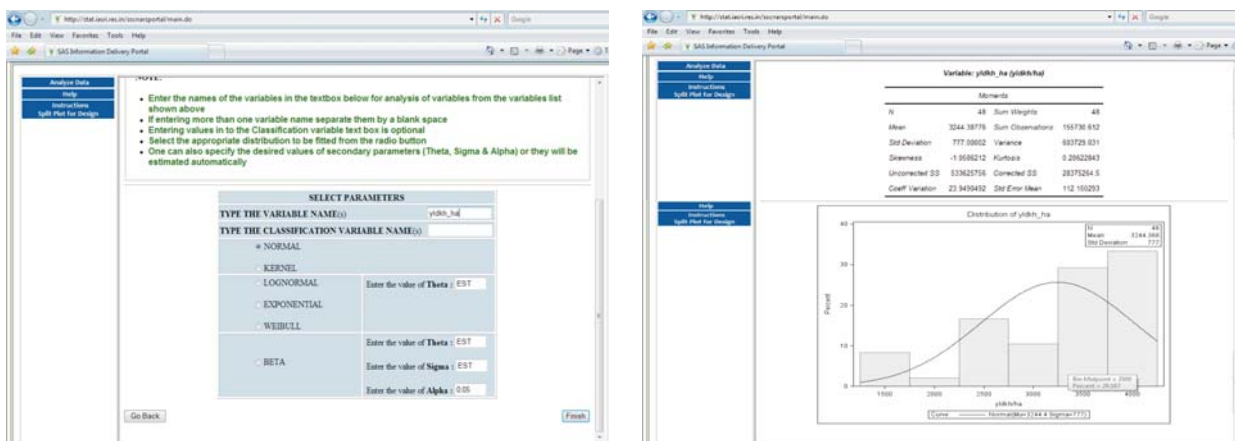

Univariate Distribution Fitting

Test of Significance: The option "TEST_OF_SIGNIFICANCE" in Analyze drop down list can be used for one sample t-test, two independent samples t-test and paired t-test. The desired test can be performed by selecting the option using the radio button. User has to enter the names of the variables in the corresponding text box. One can define left tail, right tail or two tailed test. One can also define the hypothesized value and level of significance (α). The choice is used for 100 (1-α)% confidence interval only and in the output only p-values are shown for t-test and no indication whether the test is significant at chosen level of α.
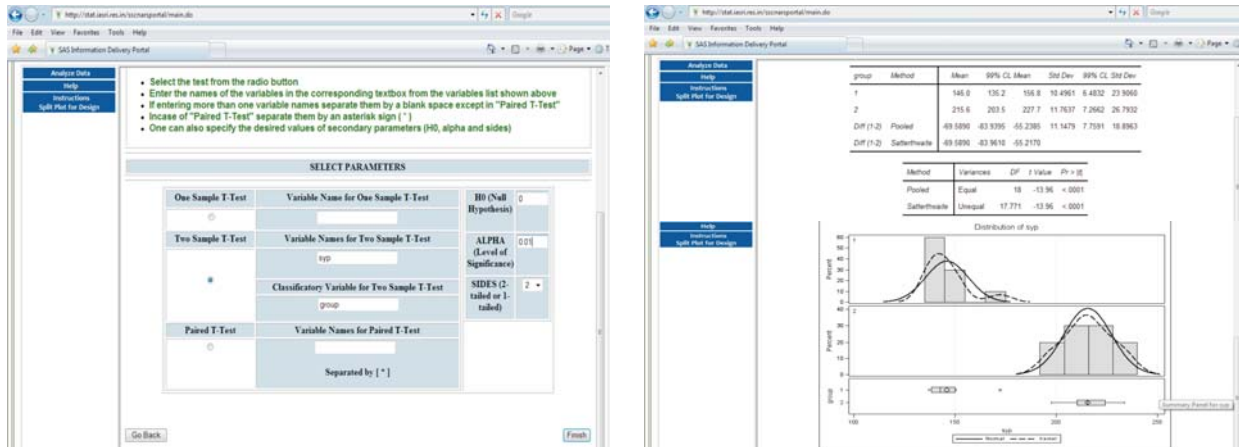
Test of Significance using t-test

Correlation: The option "CORRELATION" in Analyze drop down list can be used for computing Pearson-product moment correlation coefficient, Spearman's rank correlation coefficient, Hoeffding's measure of dependence, Kendall's tau-b and their tests of significance. It can also be used for testing the hypothesis of correlation coefficient being equal to a given value through Fisher's Z-transformation. One can also obtain partial correlation coefficients by defining the variable(s) whose linear effect is to be removed from the analysis variables.
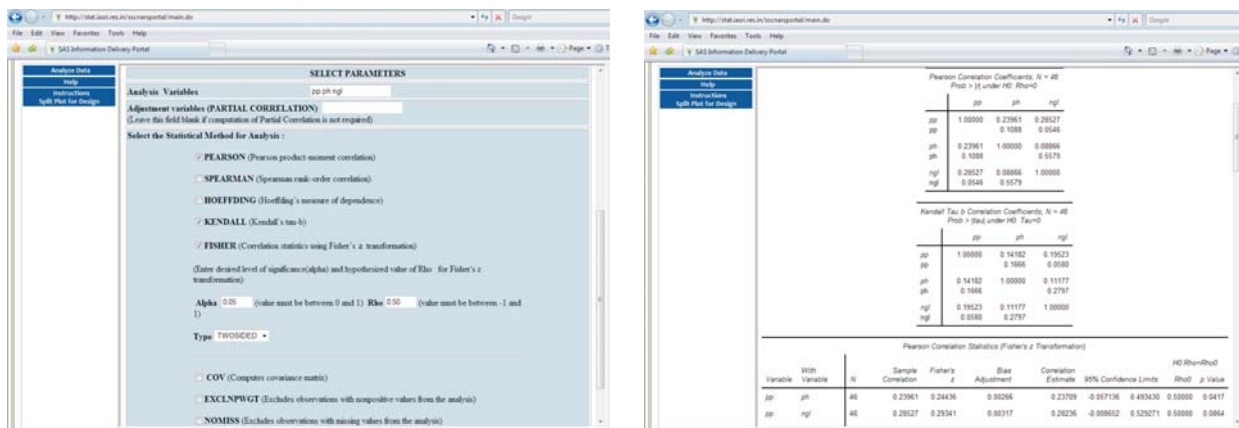
Analysis from Correlation option

Regression Analysis: The option "REGRESSION_ANALYSIS" in Analyze drop down list can be used for fitting simple/ multiple linear regression equation. Variable selection procedures such as Forward, Backward, Stepwise, Mallow's C_p statistic, etc. can also be used.
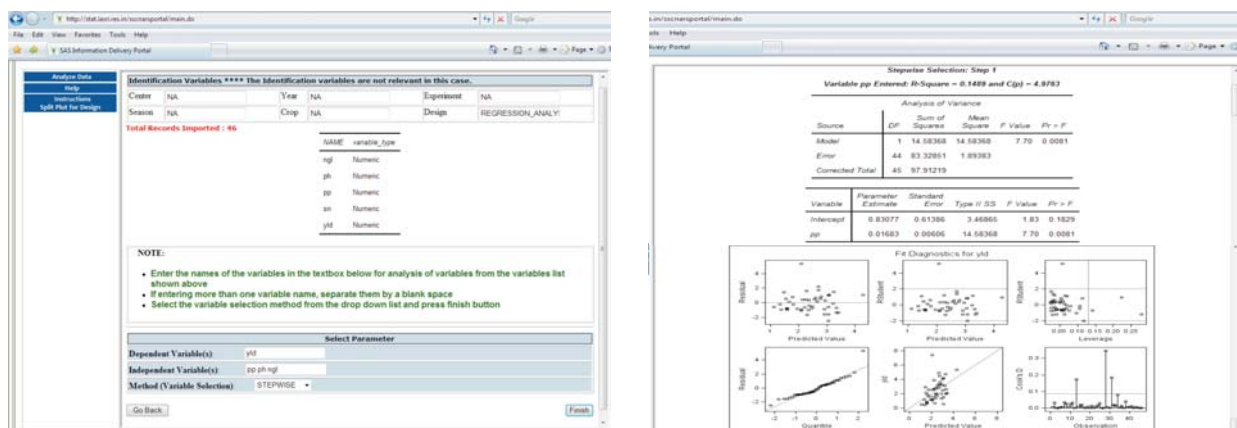
Analysis from Regression Analysis option

All the above analyses can be performed by both the users by selecting the link "Analyze Data". For registered users, links Upload Details/Analyze Data, Filter Data and Analyze and Edit Uploaded Data provide additional facilities in terms data storage, data sharing and performing analysis on the filtered data.

Upload Details / Analyze Data

The Registered users of the portal are allowed to upload the data files along with information about the experiments on selection of **Upload Details/ Analyze Data** link. The uploaded files along with the experiment details are saved in the database by default. The registered user would generally be the member of All India Coordinated Research Project (AICRP) / Network Project. The statistical analyses performed using this option are block designs and split plot designs only.

Filter Data and Analyze

At present, the registered users are divided into two different categories, namely experimenters and coordinators. The registered users with role of experimenters are allowed to access and perform the analysis of data which have been uploaded by them and cannot view/ analyze the data from other users. On the other hand, the registered users with coordinator role can access and perform the analysis of all the data those have been uploaded by different experimenters from their respective centres. Using this option, the registered user can select the experiments to create a single data file from the multiple data files uploaded earlier. In the option **Filter Data / Analyze**, the registered user can select Experiment Details (such as Center at which experiment was conducted, year of conduct of experiment, type of experiment, season, crop and design) and save filtered data as MS-Excel file or perform a combined analysis of the experiments across environments. While performing the combined analysis of filtered data, option of dropping of environments specifying cut off value based on percent coefficient of variation (CV) and / or cut off based on state means or include all environments are available.

Edit Uploaded Data

The **Edit Uploaded Data** link allows the registered user to modify the data file uploaded. This option is provided to facilitate the data entry and experiment details in multiple sessions and rectify the errors, if any in the additional / auxiliary details.

This module is being strengthened further and would be developed as a working prototype for an All India Coordinated Research project as described in the section 7.

4. Resource Page of Portal

A Resource Page http://stat.iasri.res.in/sscnarsportal/public has been created for sharing the information across all Indian NARS Users. It has selective access through IP Authentication. Here all reference manuals, presentations, sample data sets, codes and examples, video links and license files are available to NARS organizations. A screenshot of resource page is

Resource Page

5. Other IP Authenticated Services

Besides the service oriented computing services and resource page, the portal offers several other services for the users of NARS. These include access to advanced tools such as Enterprise Miner for data mining, Web Report Studio, BI Dashboards and OLAP Cube Viewer to create reports/graphs/cubes, etc. through web. The URLs for accessing these services are

Web Report Studio: http://stat.iasri.res.in/sscnarswebreportstudio

BI Dashboard: http://stat.iasri.res.in/sscnarsbidashboard

Web OLAP Viewer: http://sas.iasri.res.in:8080/sscnarswebolapviewer

E-Miner 6.1: http://sas.iasri.res.in:6401/AnalyticsPlatform

E-Miner 7.1: http://stat.iasri.res.in/SASEnterpriseMinerJWS/Status

All the links to these IP Authenticated services are also available from the project website, www.iasri.res.in/sscnars.

6. Usage of the Portal

The portal is being widely used by scientists, research scholars and students of Indian NARS. IP Addresses of 136 NARS organizations have been entered in the authentication list. From March 2011 to June 19, 2012, there are 30,479 hits from Indian NARS organizations on the portal, i.e. on an average 64 hits per day.

7. Future Directions

The Indian NARS Statistical Computing Portal created and being strengthened aims to culminate into a service oriented computing resource for standardization for statistical analysis of data generated by researchers in the Indian NARS and facilitate customized modules for AICRPs / Network Projects. To achieve this end,

the analysis modules would be strengthened further. Some of the following online analysis modules in service oriented computing mode have been planned to be included in near future are

- Block designs for factorial experiments
- Strip plot designs
- Response surface fitting
- Experiments with mixtures methodology
- Analysis of data generated through mating designs
- Categorical data analysis
- Analysis of covariance
- Multivariate analysis of variance
- Cluster analysis
- Discriminant analysis
- Estimation of genetic variance components, heritability coefficient, genetic correlation, etc.
- Path analysis
- Probit analysis
- Graphs

For facilitating customized modules for AICRPs/ Network Projects, current facility of Registered Users would be upgraded to a working prototype which would

- facilitate planning of experiments at AICRPs,
- facilitate randomized layouts / data file preparation,
- maintain information about the experiments at a centralized place,
- allow enter/upload experimental data during the course of experiment (or at the end), ability to carry out appropriate statistical analysis and automate uniform reporting process,
- provide secured access to data for authorized users and different access rights at different levels of users
- store the data and results in personal/project folder by the users
- be linked to Indian NARS Statistical Computing Portal for online design generation and online analysis of experimental data.
- help in standardization of data collection and statistical analysis across AICRPs.

The prototype would be flexible / generic so that it can be modified for use across several AICRPs.

Other Information Systems on Designed Experiments being developed at IASRI, New Delhi would also be linked to Indian NARS Statistical Computing Portal for value addition.
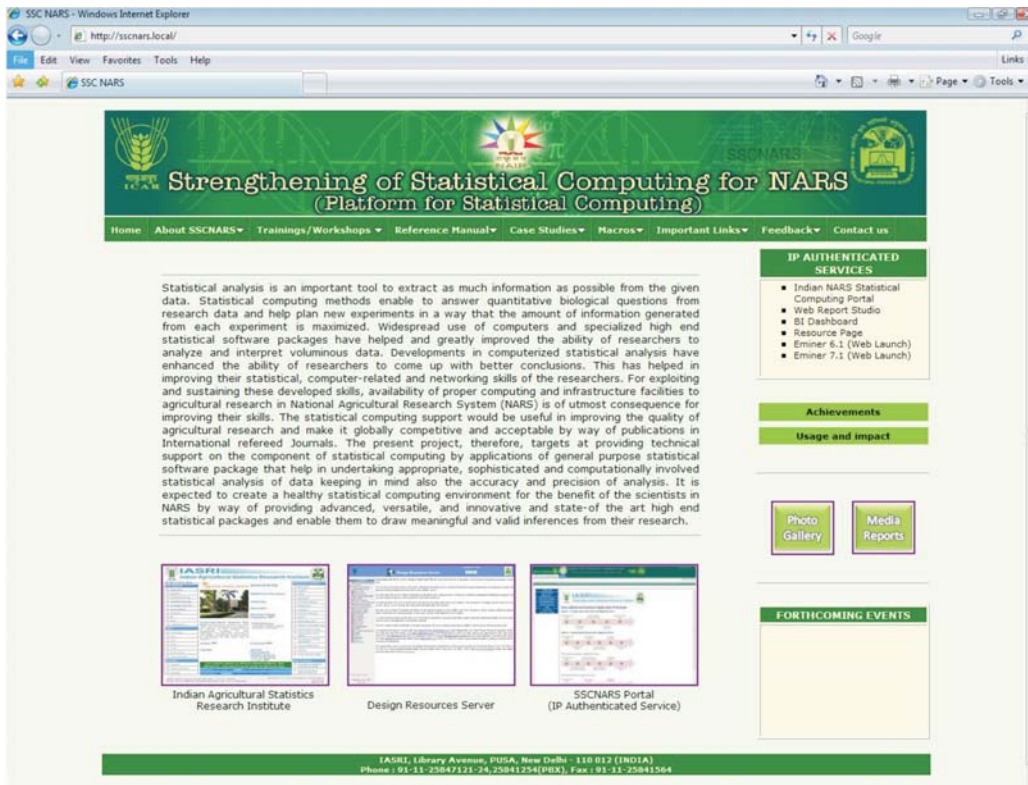
NAIP Consortium
Strengthening Statistical Computing for NARS
(Platform for Statistical Computing)
www.iasri.res.in/sscnars

The reference manuals, case studies, macros for customized analysis developed under the Consortium are made available across the globe at the website hosted at www.iasri.res.in/sscnars.

- During November 15, 2010 – June 19, 2012, Google Analytics gave **11,754 hits across 252 cities of 60 countries**
- The e-manual developed has been cited in Journal of Doctoral Research in Economics of the Bucharest Academy of Economic Studies. The macro developed for augmented designs has been cited by Jennifer Kling, Oregon State University in Introduction to Augmented Designs.

हर कदम, हर डगर
किसानों का हमसफर
भारतीय कृषि अनुसंधान परिषद

*Agri*search *with a* human *touch*

For any critical comments / suggestions that help in improving the contents and the presentation so as to make the portal more meaningful to the researchers inclusive of students, please write to

Director, IASRI, New Delhi
director@iasri.res.in

Dr. Rajender Parsad
Consortium Principal Investigator
rajender@iasri.res.in