

Statistical Package for Agricultural Research (SPAR 2.0)

Sangeeta Ahuja, P.K. Malhotra, V.K. Bhatia and Rajender Parsad
Indian Agricultural Statistics Research Institute, New Delhi
(Received: November 2006)

SUMMARY

Statistical Package for Agricultural Research (SPAR 2.0) has been developed under the Windows platform for handling statistical analysis of agricultural experimental research data. The package has eight modules to carry out different aspects of analysis such as diallel analysis, path analysis, discriminant analysis, cluster analysis, line \times tester analysis, stability analysis etc. useful for plant breeding and Genetics research studies. This package is also useful for teaching the subject of Genetical Statistics to the Post Graduate students and researchers in statistics with special interest in Plant and Animal Sciences. There are no high end requirements for the successful execution of the package, a normal system with minimum configuration is enough to run this package.

Key words: Statistical package, Plant breeding and genetics analytical tools, Microsoft Visual C++.

1. INTRODUCTION

SPAR 2.0 has been primarily developed for the statistical analysis of experimental research data in Plant Breeding and Genetics. However, many of the modules can be used for the analysis of data coming from other disciplines of research, provided the data conform to the method and design used in the package. Earlier, Statistical Package for Agricultural Research Data Analysis (SPAR 1.0) was developed at IASRI for statistical analysis of experimental research data in Plant Breeding and Genetics using FORTRAN language, which was DOS based and had some limitations like Dynamic Memory Allocation, Graphic User Interface, etc. The package was not able to handle large data sets. So, there was a need to develop the Windows Version of SPAR 1.0. SPAR 2.0 is a Windows Version of SPAR 1.0 with some additional modules. This package has been developed in Microsoft Visual C++ 6.0 language. It is User-Friendly, Interactive, Password Protected, Menu-Driven Package and can also be operated using the Toolbars. A Context-Sensitive Web Help with Index, Contents, Search and Favorites facility is a special feature of the package. This package consists of all the necessary details to use the software and also the basic definitions including mathematical formulae related to the methods used in the package. The sample inputs and outputs of

different modules are helpful tool for quick understanding of the working of the package. This package is also useful for teaching the subject of Genetical Statistics to the Post-Graduate students and researchers in Statistics with special interest in Plant and Animal sciences. The package consists of eight modules viz. (i) Data Management, (ii) Descriptive Statistics, (iii) Estimation of Breeding Values, (iv) Correlation and Regression Analysis, (v) Variance and Covariance Components Estimation, (vi) Stability Analysis, (vii) Multivariate Analysis and (viii) Mating Design Analysis. A system with minimum of 64 MB RAM and 2 GB Hard Disk storage capacity is required to run this package and it will be compatible with Microsoft Windows 98, ME, 2000 and XP.

Section 2 discusses design and development of SPAR 2.0. Section 3 describes the features and Section 4 deals with help modules.

2. DESIGN AND DEVELOPMENT OF SPAR 2.0

SPAR 2.0 has been written in Microsoft Visual C++ 6.0. Visual C++ is an object-oriented language and is suitable for software development. It utilizes various key features of object oriented technologies such as its ability to program in an event driven operating system with great ease, write code for events automatically, optimize code

capability for native platform, etc. The basic reason behind selecting an object-oriented language was its superior abilities for code reusability, inheritability, encapsulation, portability and modular development as reported in Kruglinski (1996), and Ritcher (1999). The Software Process Model used in SPAR 2.0 is Spiral. SPAR 2.0 offers most of the usual features of any typical Windows based program, such as menu driven functions, push buttons for common operations, keyboard accelerators as hot keys, rich edit control for text editor, informative status bar and two useful toolbars for carrying out common tasks. An online SPAR 2.0 Web HTML Help with Contents, Index, Search and Favorites Facilities is available and is developed using Microsoft HTML Web Help Workshop 4.0, JavaScript and a graphics edit tool Microsoft Help Image Editor. The package is very easy to use and the password control regulates its access to valid users. The package also works as a text editor with usual facilities for preparing and editing data file. The concepts of Dynamic Linked Library (DLL) and Component Object Model (COM)/Dynamic Component Object Model (DCOM) in Microsoft Visual C++ 6.0 have been utilized in SPAR 2.0 for complex mathematical calculations.

3. FEATURES OF SPAR 2.0

Flash screen of SPAR 2.0 is given in Fig. 1. As SPAR 2.0 is password protected so after entering the password, the next screen output which is main screen, is given in Fig. 2.

3.1 Data Management

The Data Management Module consists of a sub-module for Transformation of Data. One can generate new variables from the existing variables using Numeric, Relational, Arithmetic and Logical operators. One can transform the data using various Logarithmic, Trigonometric, Inverse and other mathematical functions. User can either enter the data through SPAR 2.0 Editor or can browse the file from any destination.

3.1.1 Transformation of data

On selection of Option “Data Management”, a sub-menu “Transformation of Data” will be opened. After entering the number of Characters and Rows information a dialog box will be opened. Through this box, one can select the type of file ASCII or Excel from the Combo

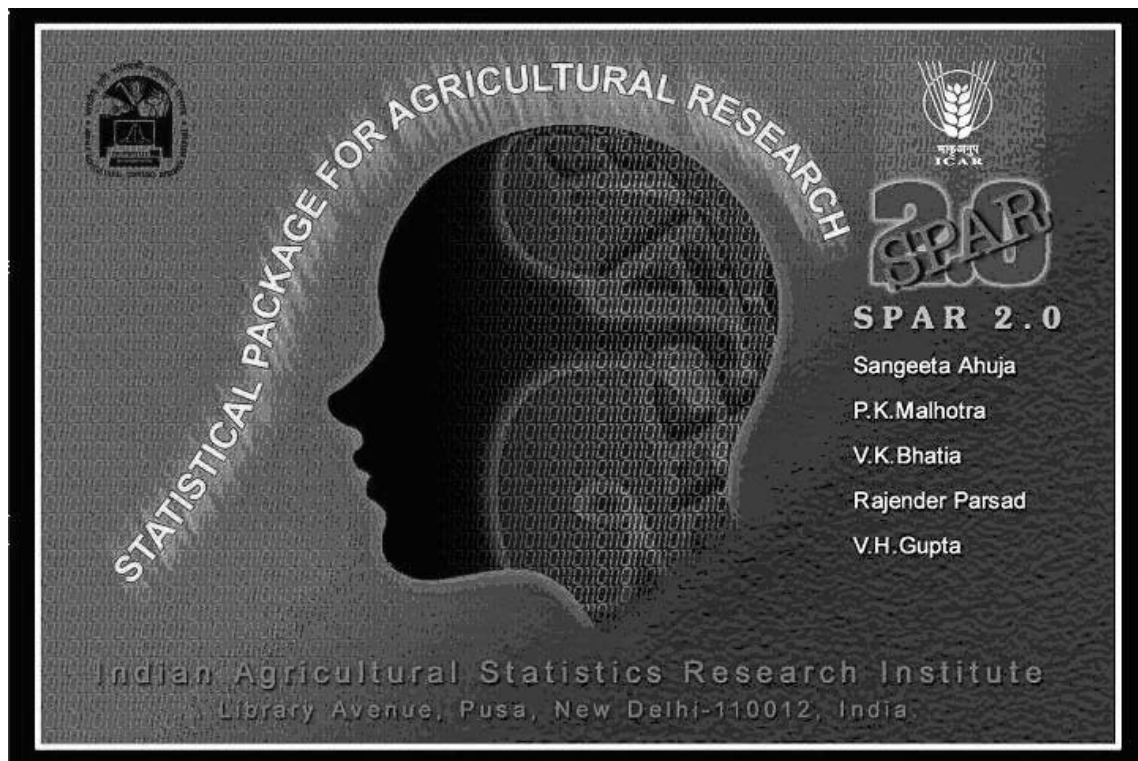


Fig. 1. Flash Screen of SPAR 2.0

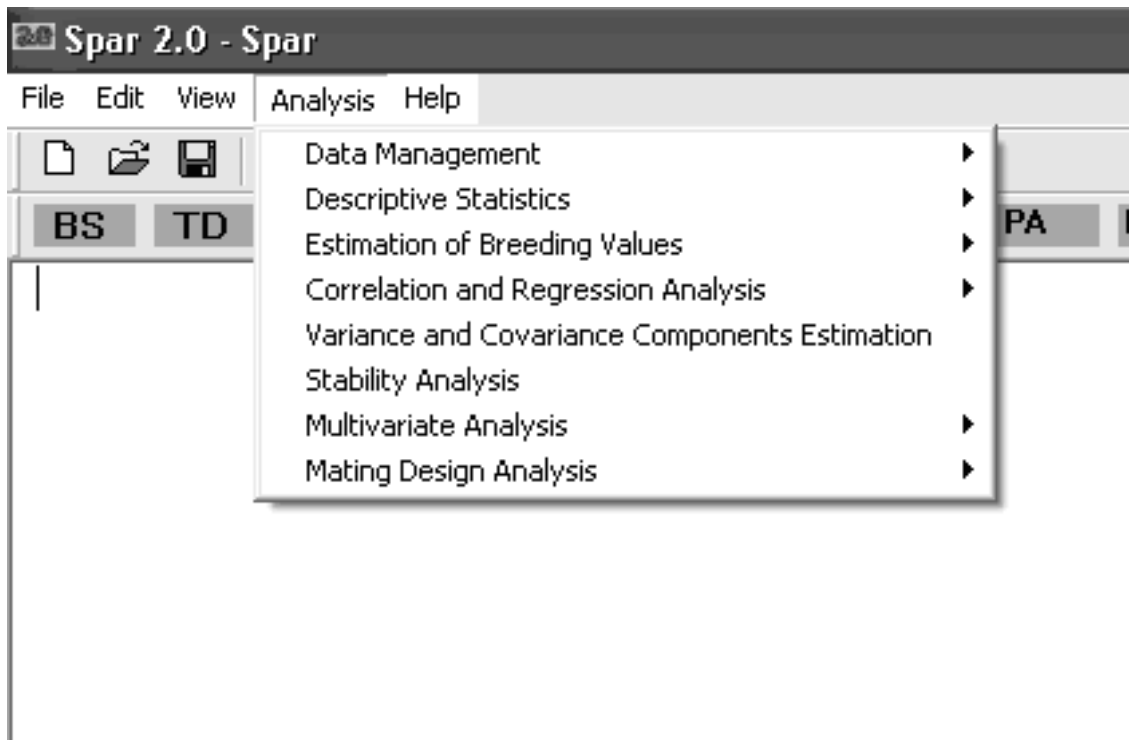


Fig. 2. First screen of SPAR 2.0 after entering password and clicking the Analysis link

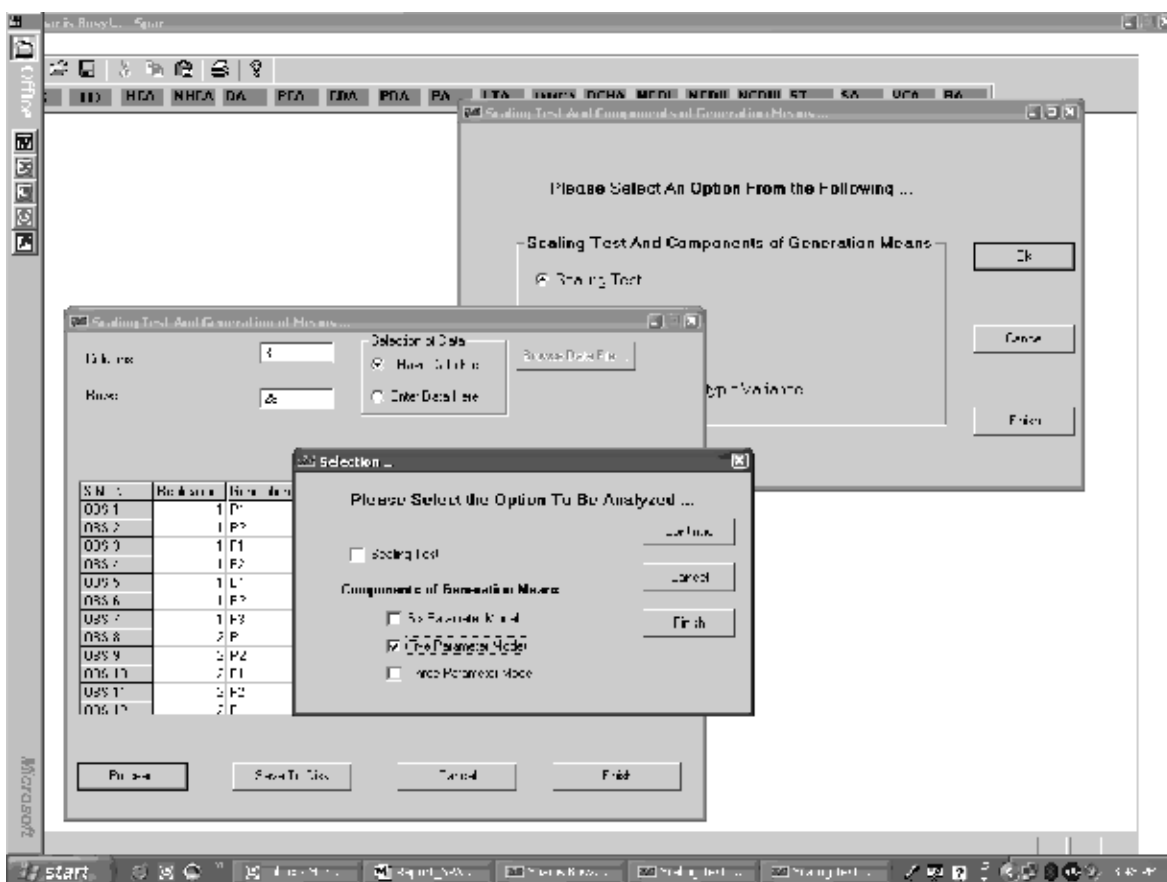


Fig. 3. Scaling Test : Selection of scaling test, and selection of five parameter model

Box. This procedure of selection of data to be processed either through an existing file or entering the data through a worksheet is followed in all the modules. After fetching the data one can click the “Proceed” command button to enter information on parameters like target variable (output variable name), type label, numerical expression, relational operators, numerals, functions. Clicking the “Finish” button will lead to termination of the transformation of data.

3.2 Descriptive Statistics

The Descriptive Statistics Module consists of various sub-modules, such as Measures of Central Tendency; Measures and Coefficients of Dispersion; Measures of Skewness and Kurtosis; Measures of Partition Values and Generation of Moments. In addition to various descriptive statistics there is also a provision to compute the Test of Independence of Attributes using Chi Square Test.

3.3 Estimation of Breeding Values

The Estimation of Breeding Values Module predicts breeding values by fitting of generations means models with three, five and six parameters along with performing Scaling and Joint Scaling Tests (Hayman and Mather 1955). Adequacy of scale must satisfy two conditions namely, additivity of gene effects and independence of heritable components from non-heritable ones. The test of first condition provides information regarding absence or presence of gene interactions. The test of adequacy of scales is important because in most of the cases the estimation of additive and dominance effects and their components of variances are important for studying different gene interactions. Sample screen for some of the options are given in Fig. 3.

3.4 Correlation and Regression

This module consists of sub-modules for correlation, regression including path analysis. The correlation analysis gives Simple Correlation, Partial and Multiple Correlations whereas the regression analysis computes the Regression Coefficient, Analysis of Variance of Regression and Regression Equation. The Path Analysis module gives Direct Effects, Indirect Effects and Residual Effects.

3.5 Variance and Covariance Components Estimation

The Variance and Covariance Component Estimation Module consists of computation of ANOVA components of variances and covariances. Both phenotypic and genotypic component of variations are obtained under RBD layout. From these estimated variances and covariance components the estimates of heritability and genetic correlation along with their standard errors and critical differences are worked out (Singh and Choudhary 1977).

3.6 Stability Analysis

Stability analysis can be performed using three models: (i) Eberhart and Russell (1966); (ii) Perkins and Jinks (1968) and (iii) Freeman and Perkins (1971).

On selection of Stability Analysis module, a Window Dialog Box will open up and demand the information on parameters on Locations, Varieties, Replications, and Characters (Fig. 4). Sample results are given in Fig. 5.

S.NO.	Locations	Replication	Varieties	Char1	Char2
OBS 1	1	1	1	36.4	36.4
OBS 2	1	1	2	40	40
OBS 3	1	1	3	32.4	32.4
OBS 4	1	1	4	33.5	33.5
OBS 5	1	1	5	41.3	41.3
OBS 6	1	1	6	27.9	27.9
OBS 7	1	1	7	38.5	38.5
OBS 8	1	1	8	38.6	38.6
OBS 9	1	1	9	41.6	41.6
OBS 10	1	1	10	22.6	22.6
OBS 11	1	2	1	41.3	41.3
OBS 12	1	2	2	38.9	38.9

Fig. 4. Information about parameters for stability analysis

3.7 Multivariate Analysis

This module has three sub-modules, viz. (i) Cluster Analysis; (ii) Discriminant Analysis and (iii) Principal Component Analysis.

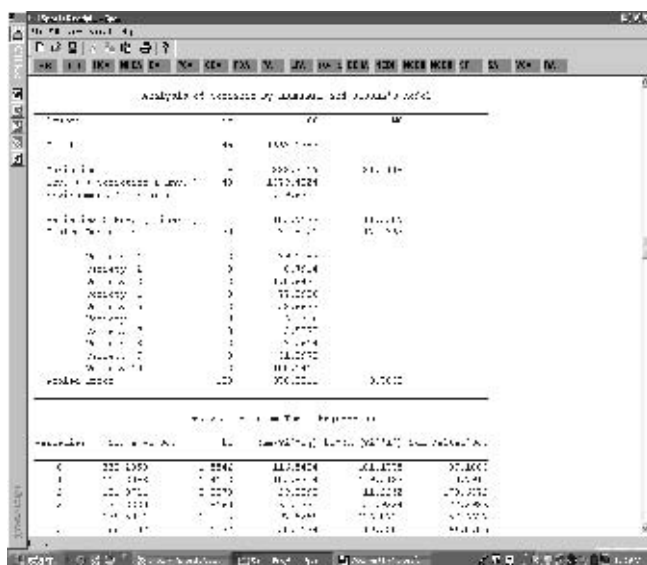
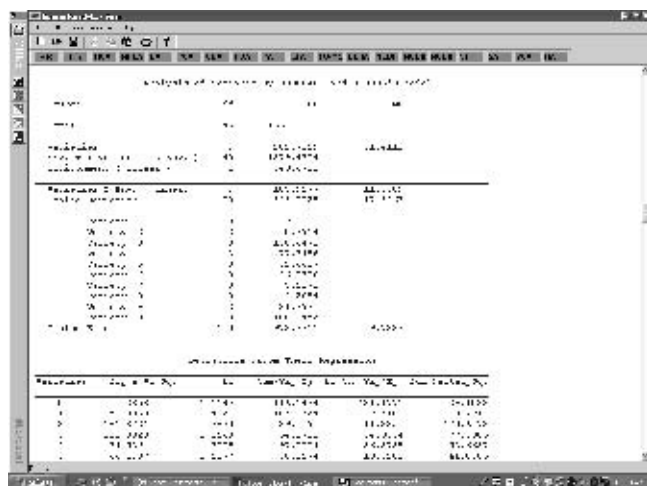
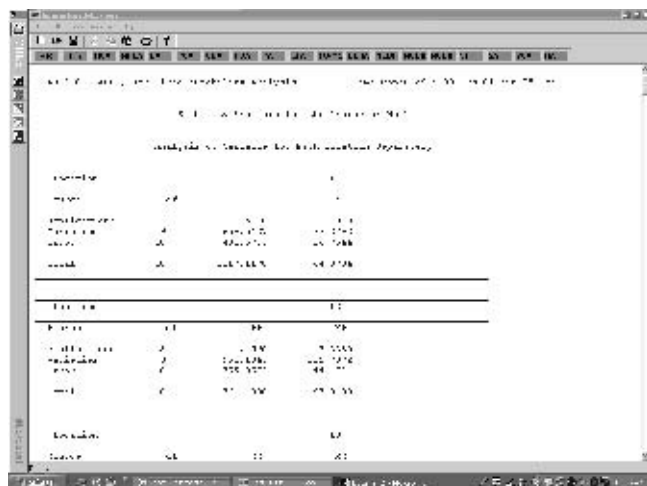


Fig. 5. Sample results generated on execution of Sub-module Stability Analysis

3.7.1 Cluster Analysis

Cluster analysis sub-module is further partitioned into Hierarchical Cluster Analysis and Non-Hierarchical Cluster Analysis. The Hierarchical Cluster Analysis by using various techniques such as Single Linkage (Minimum Distance or Nearest Neighbor), Complete Linkage (Maximum Distance or Farthest Neighbor), Average Linkage (Average Distance) and Distance Matrix Methods included in the computation are the Coefficient of Shape Difference, the Cosine Co-efficient, the Correlation Co-efficient, the Canberra Metric Co-efficient, the Bray–Curtis Coefficient, the Euclidean Distance. Sample screen outputs for Hierarchical Cluster Analysis are given in Fig. 6.

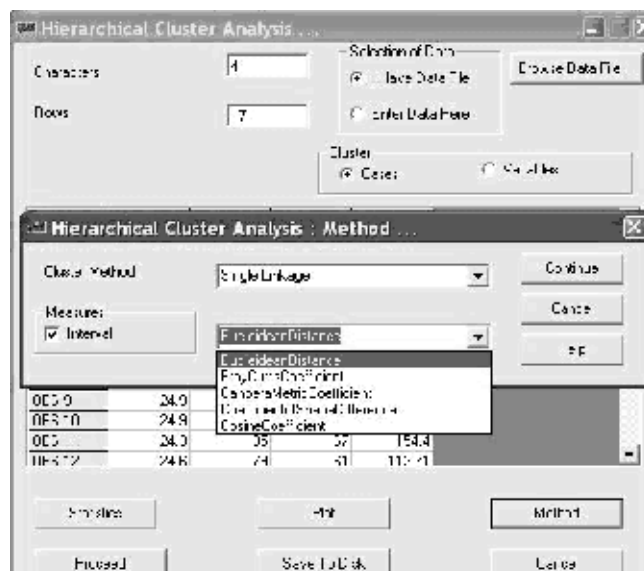
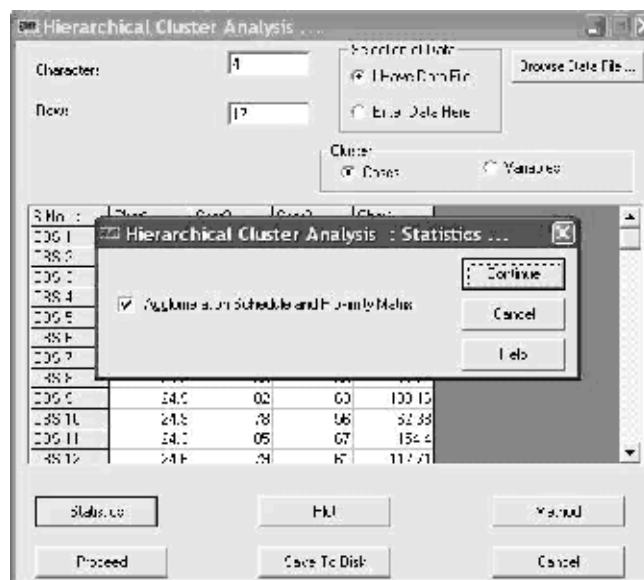


Fig. 6. Continuing with Hierarchical Cluster Analysis option

Non-Hierarchical Clustering can be done by four methods, viz. (i) K-Means (ii) K-Medians (iii) K-Modes and (iv) K-Mediod. Provision is there to analyze the data both Cases and Variables wise. On selection of Non-Hierarchical Cluster Analysis from SPAR 2.0, a Window Dialog Box will open up that demands information about the parameters on characters and rows. From the Frame “Cluster”, one can select one of the options from the two Radio Buttons, Cases and Variables. If one selects the Radio Button “Cases”, it will compute the Distance Matrix Row wise and if one selects the Radio Button “Variables”, it will compute the Distance Matrix Column wise.

After selection of input file, if one presses the Method Command Button, a Window Dialog Box will open up which demands parameter information on Cluster Method option where one can choose from the four options, K-Means, K-Median, K-Mode and K-Mediod (Fig. 7).

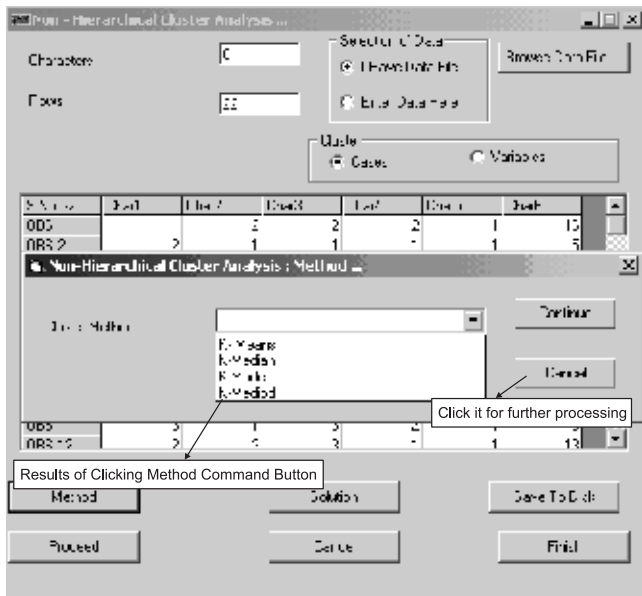


Fig. 7. Selection of the Cluster Method option

On clicking the Command Button “Solution”, a Window Dialog Box will open up which asks the following parameter information from the user on the two options, Single Solution and Multiple Solution (Fig. 8).

If one clicks the “Single Solution” Radio Button, a Window Dialog Box will open up which will ask the parameter information, Range (number of clusters). After entering the Range in the Text Box, user can click the

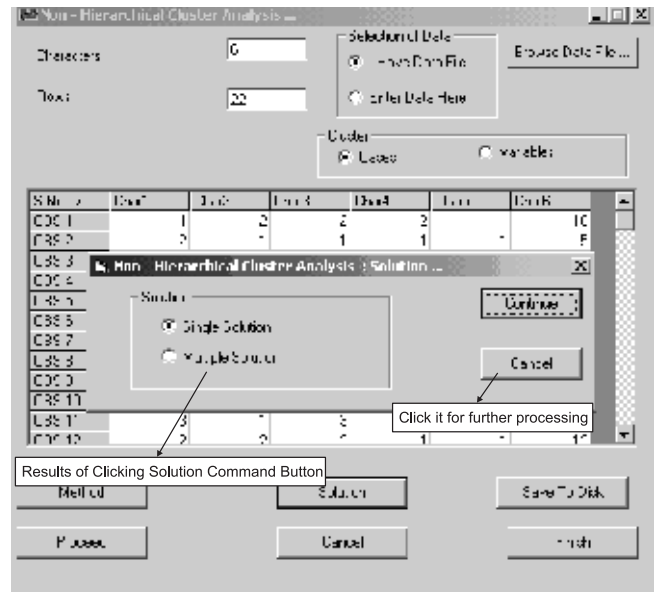


Fig. 8. Choice of the type of solution - single or multiple

“Continue” Command Button. If one clicks the Multiple Solution Radio Button Option, a Window Dialog Box will come up which asks the following parameter information, Lower Range and Upper Range. User can specify the number of cluster range using Lower Range for the minimum number of clusters to Upper Range for the maximum number of clusters (Fig. 9).

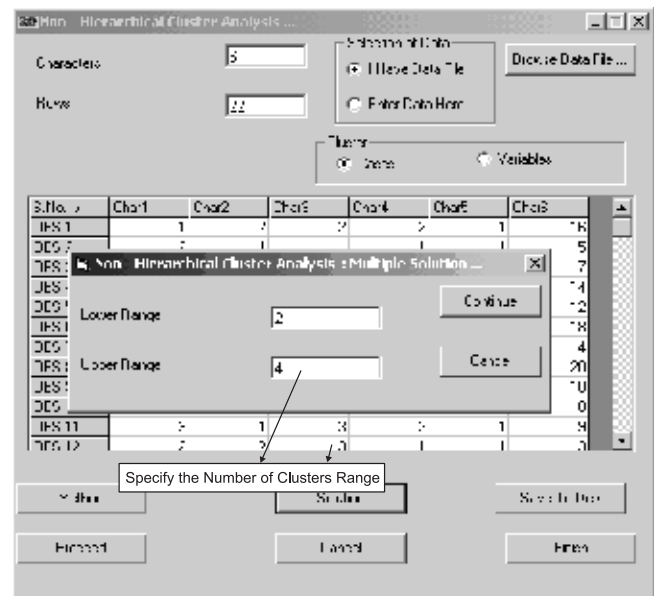


Fig. 9. Specifying the number of clusters in multiple solution

After entering all the information in the various Window Dialog Boxes, one can click the “Proceed” Command Button for Final Execution of the Non-Hierarchical Cluster Analysis.

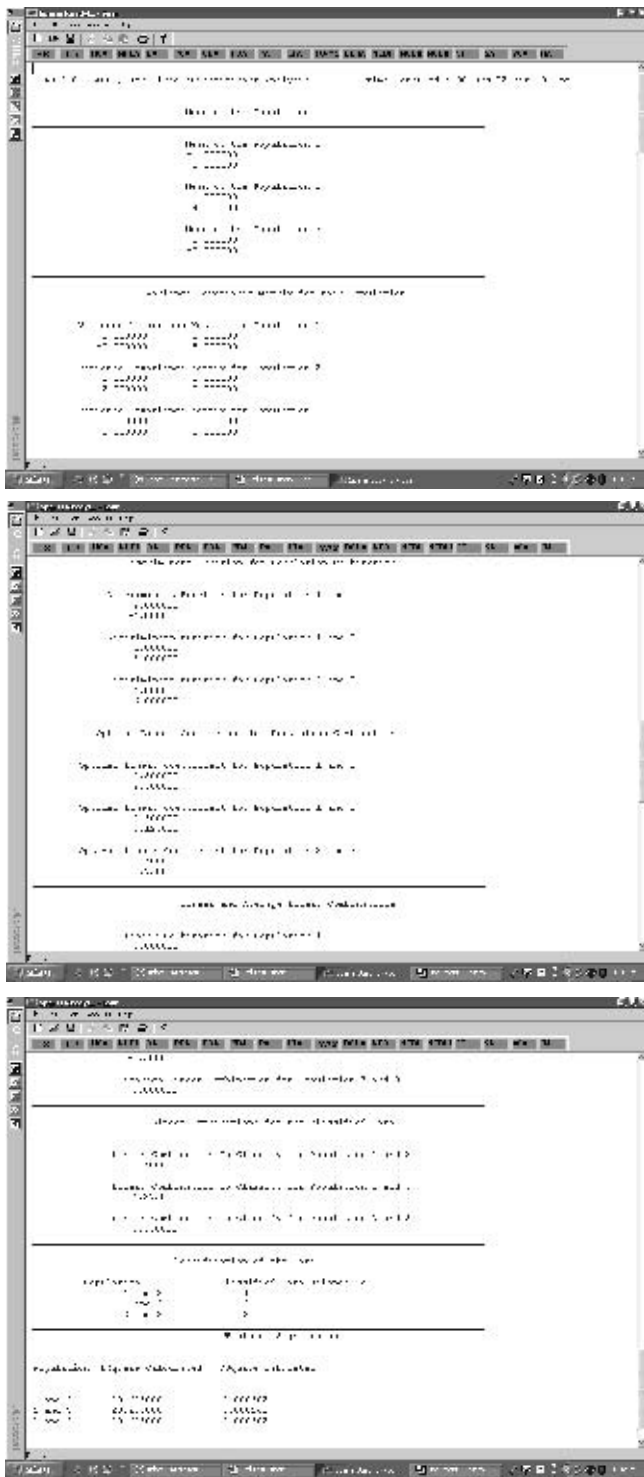


Fig. 10. Sample output of Discriminant Function Analysis

3.7.2 Discriminant Analysis

Discriminant Analysis can be used for estimation of Variance-Covariance matrix, Discriminant function, Optimallinear combination and Testing the validity of significance. Few screen shots are given in Fig. 10 which

describes the functioning of the Discriminant Procedure where the data set has 2 characters and 9 rows.

3.7.3 Principal Component Analysis

Principal Component Analysis can be employed for computation of Variance-Covariance matrix, Correlation matrix, Eigenvalues, Eigenvectors, Principal Components, Variance of Principal Components, Total variation by original variables, Total variation by Principal Components and proportion of total variation accounted for by each of the Principal Components and Principal Component scores.

On selection of Principal Component Analysis, a Window Dialog Box will open up that demands the information about the parameters on characters, rows and the computational process (Fig. 11).

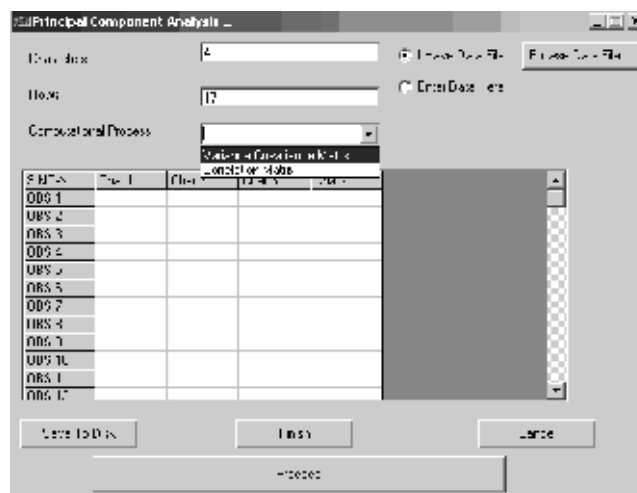


Fig. 11. Sample screen shots for Principal Component Analysis

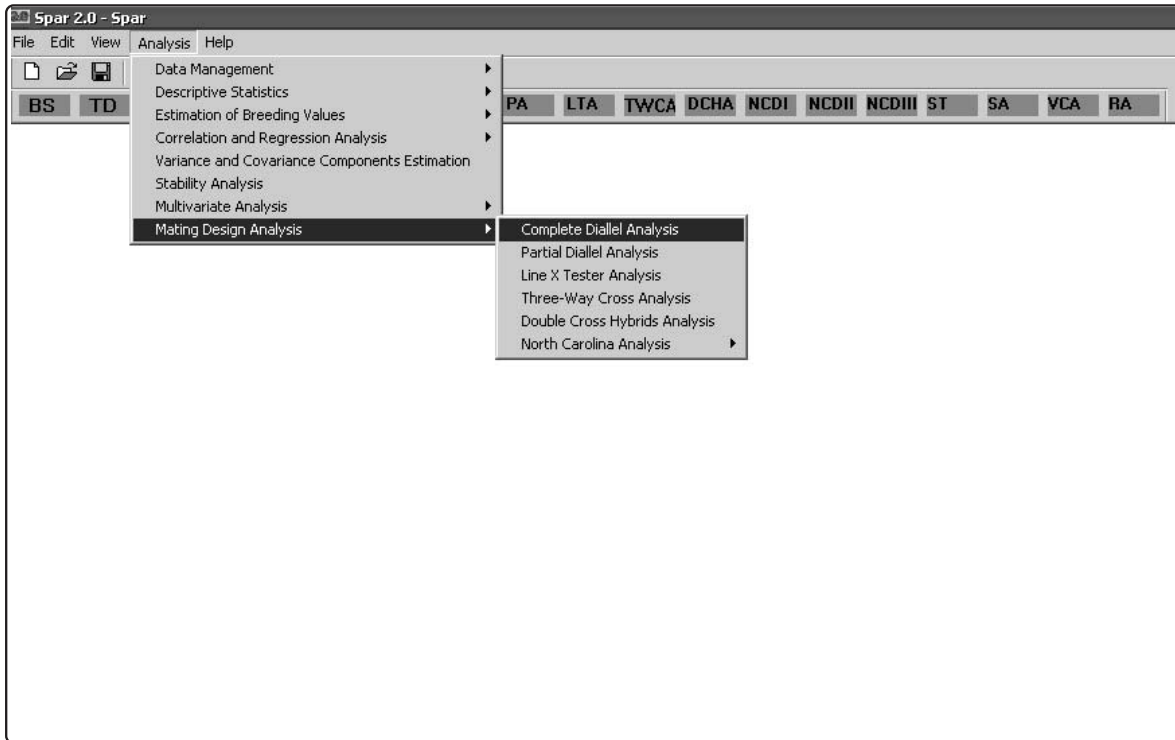


Fig. 12. Complete Diallel Analysis by Hayman's and Griffing's Approaches

3.8 Mating Design Analysis

This module consists of seven sub-modules viz. (i) Complete Diallel; (ii) Partial Diallel; (iii) Line \times Tester (with parents) and Line \times Tester (without parents); (iv) Three-way cross; (v) Double cross and (vi) North Carolina Designs.

3.8.1 Complete Diallel Analysis

Complete Diallel Analysis can be performed by using two approaches: (i) Hayman's approach (Hayman 1954 a, b; 1957, 1958) and (ii) Griffing's approach (Griffing 1956).

The Hayman's approach includes the computations, viz. ANOVA, Estimation of components of variation such as Testing the Significance of Genotypic Differences, Estimation of Variances and Covariance, Testing the Validity of the Hypothesis, W_r , V_r -Graph, Estimation of Components of Variation, Other Parameters, Estimation of Most Dominant and Recessive Parent, Estimation of Components of Variation in F_2 , Computation of Proportion of the Genetic Components.

The Griffing's approach includes the Analysis using Parents (n), $(n-1)/2 F_1$'s and reciprocals, Analysis using Parents and F_1 's only, Analysis using F_1 's and reciprocals and Analysis using F_1 's only. All the above analysis computes Significance of Genotypic Difference, Combining Ability Analysis, gca, sca, reciprocal effects, Standard Error and Critical Differences. Sample screen shots are given in Fig. 12.

3.8.2 Partial Diallel Analysis

The Partial Diallel Analysis in SPAR 2.0 computes the Sampling Procedure, Testing the Significance of genotypic differences, Combining Ability Analysis, Component of Variances and Standard Error (Narain 1990, Kempthorne and Curnow 1961).

3.8.3 Line \times Tester Analysis

Line \times Tester Analysis computes ANOVA with Parents and Crosses, ANOVA for Line \times Tester Analysis, ANOVA for Line \times Tester Analysis including Parents, gca effects, sca effects, Standard Error for Combining Ability Effects, Genetic Components and Proportional

Contribution of Lines, Testers and their interactions to Total Variance (Kempthorne 1957).

3.8.4 Three-Way Cross Analysis

Three-Way Cross Analysis computes ANOVA for Triallel, Testing the Significance of Genotypic (Crosses) Differences, Estimation of Effects and Variances, Estimation of the various effects such as Estimation of General and Two-line Specific Effects, Three-line Effect, Analysis of Variance for Specific and General Line Effects, Standard Errors, Estimation of Parameters Useful for the Evaluation of Lines, Estimation of the Variance Components (Rawling and Cockerham 1962, Hinkelmann 1965, Ponnuswamy 1972).

3.8.5 Double Cross Hybrids Analysis

Double Cross Hybrids Analysis includes the computation of Statistical Model based ANOVA, Components of Variances, Combining Ability Effects (Rawling and Cockerham 1962).

3.8.6 North Carolina Designs Analysis

North Carolina Designs Analysis computes the ANOVA and Component of Variances for NCD I, II and III Designs (Comstock and Robinson 1948, 1952).

4. HELP

A complete Context Sensitive Web HTML Help with Index, Contents, Search and Favorites facility is available. The package can be run without the aid of a manual. This package consists of all the necessary details to use the software and also the basic definitions including mathematical formulae related to the methods used in the package. It also includes sample inputs and outputs of all the modules. Fig. 13 gives how to use help for a specified analysis. Fig. 14 gives the SPAR 2.0 Help Samples Inputs and Outputs.

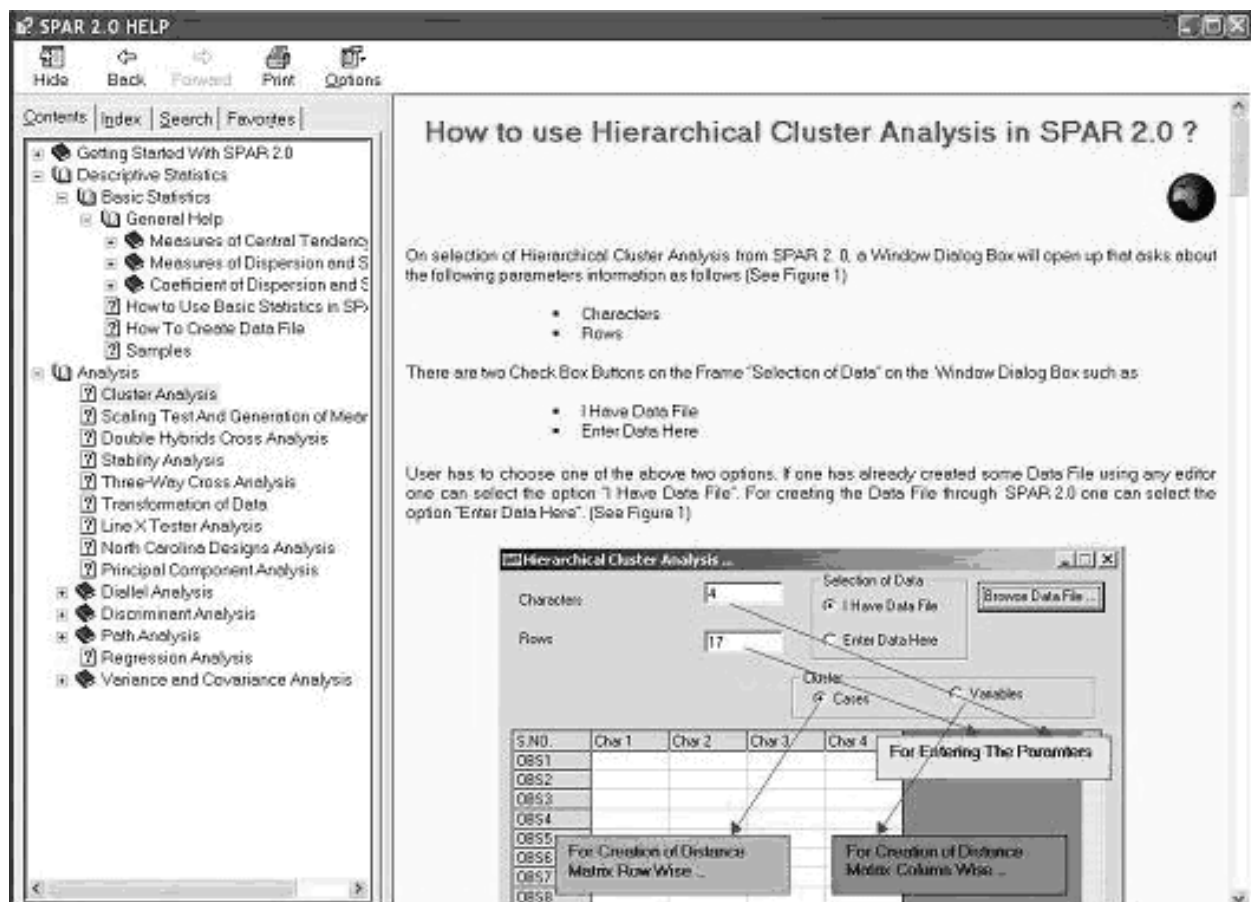


Fig. 13. SPAR 2.0 Help

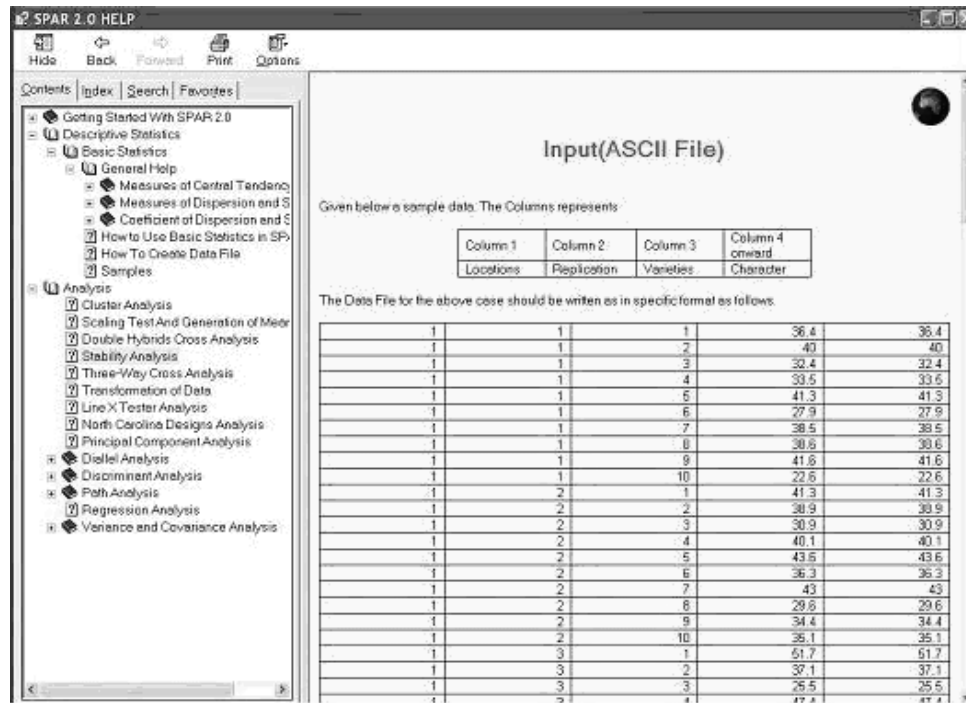


Fig. 14. SPAR 2.0 Help Samples Inputs and Outputs

REFERENCES

- Comstock, R.E. and Robinson, H.F. (1948). The components of genetic variance in populations of biparental progenies and their use in estimating the average degree of dominance. *Biometrics*, **4**, 254-266.
- Comstock, R.E. and Robinson, H.F. (1952). Estimation of average dominance of genes. In: *Heterosis*, Iowa State College Press, Ames, 494-516.
- Eberhart, S.A. and Russell, W.A. (1966). Stability parameters for comparing varieties. *Crop Sci.*, **6**, 36-40.
- Freeman, G.H. and Perkins, J.M. (1971). Environmental and genotype-environmental components of variability, VIII. Relations between genotypes grown in different environments and measure of these environments. *Heredity*, **26**, 15-23.
- Griffing, B. (1956). A concept of general and specific combining ability in relation to diallel system. *Austr. Biol. Sci.*, **9**, 463-493.
- Hayman, B.I. (1954 a). The theory and analysis of diallel crosses. *Genetics*, **39**, 789-809.
- Hayman, B.I. (1954 b). The analysis of variance of diallel tables. *Biometrics*, **10**, 235-244.
- Hayman, B.I. (1957). Interaction, heterosis and diallel crosses. *Genetics*, **42**, 336-355.
- Hayman, B.I. (1958). The separation of epistatic from additive and dominance variation in generation means. *Heredity*, **12**, 371-390.
- Hayman, B.I. and Mather, K. (1955). The description of genic interaction in continuous variation. *Biometrics*, **11**, 69-82.
- Hinkelmann, K. (1965). Partial triallel cross. *Sankhya* **A27**, 173-196.
- Kempthorne, O. (1957). *An Introduction to Genetic Statistics*. John Wiley & Sons, New York.
- Kempthorne, O. and Curnow, R.N. (1961). The partial diallel crosses. *Biometrics*, **17**, 229-250.
- Kruglinski, D. (1996). *Inside Visual C++, I Edition*. Microsoft Press, Washington.
- Narain, P. (1990). *Statistical Genetics*. Wiley, New York.
- Perkins, J.M. and Jinks, J.L. (1968). Environmental and genotype-environmental components of variability. III. Multiple lines and crosses. *Heredity*, **23**, 339-356.
- Ponnuswamy, K.N. (1972). Some contributions to design and analysis for diallel and triallel cross. Unpublished Ph.D. thesis. IARI, New Delhi.
- Rawling, J.O. and Cockerham, C.C. (1962). Triallel analysis. *Crop Sci.*, **2**, 228-231.
- Rawling, J.O. and Cockerham, C.C. (1962). Analysis of double cross hybrid populations. *Biometrics*, **18**, 229-244.
- Richter, J. (1999). *Programming Applications for Microsoft Windows*, IV Edition. Microsoft Press, Washington.
- Singh, R.K. and Choudhry, B.D. (1977). *Biometrical Methods in Quantitative Genetic Analysis*. Kalyani Publishers, New Delhi.