Statistical Package for Factorial Experiments (SPFE) 1.0

Sangeeta Ahuja, Rajender Parsad, V.K. Gupta and P.K. Malhotra
Indian Agricultural Statistics Research Institute, New Delhi – 110012

SUMMARY

The SPFE gives the designs for symmetrical and asymmetrical factorial experiments and also performs analysis of the data generated. It generates randomized layout of the designs for factorial experiments with or without confounding. The software requires user input as a list of independent interactions to be confounded. Different interaction(s) can be assigned for confounding in different replications. It also generates regular fractional factorial plans for symmetrical factorial experiments. The data are analyzed as per procedure of blocked/unblocked designs for single factor experiments. The treatment sum of squares can be partitioned into sum of squares due to main effects and interactions. A null hypothesis on any contrast of interest can also be tested. The package is also useful for illustration purposes in the classroom teaching as well as for the researchers in Statistics with interest in experimental designs particularly in factorial experiments. The package has been developed using Microsoft Visual C++ 6.0, Microsoft Visual Basic 6.0 and Microsoft Access 2000.

Key words : Factorial experiments, Contrast analysis, Confounding, Block designs, Unblocked designs, Computer software.

1. Introduction

Agricultural and industrial scientists/research workers conduct experiments in complete factorial experiment without confounding in two or three levels of factors using a randomized complete block (RCB) design. For small number of treatment combinations, an RCB design may be the most efficient design due to no loss of information in estimating factorial effects as well as block effects. However, with the increase in number of factors and/or levels of factors, the number of treatment combinations increase too rapidly. As a result the sizes of the complete blocks become large and thus fail to maintain homogeneity within blocks. This results in large intra-block variances and lower precision of the estimates of treatment contrasts. In such situations, the concept of confounding has been advocated {see e.g. Cochran and Cox (1957); Yates (1937)}. The generation of appropriate confounded factorials is one issue due to a complex process and obtaining the randomized layout of the design is the other. The

experimenters would also be interested in the analysis of the data generated from these designs. Keeping these issues in view, a software Statistical Package for Factorial Experiments (SPFE) has been developed at Indian Agricultural Statistics Research Institute that enables the users/experimenters to select a design, to generate a randomized layout and analyze data. This paper is divided in two parts. Section 2 discusses various features of the SPFE. Section 3 describes how SPFE works with some details of algorithms used. The help features are described in Section 4.

2. Design and Development of SPFE

SPFE has been written in Microsoft Visual C++ 6.0, Visual Basic 6.0 and Microsoft Access 2000. Visual C++ is an object-oriented language and suitable for software development. It utilizes various key features of object oriented technologies such as its ability to programme 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, portability and modular development. For details, a reference may be made to Kruglinski (1996) and Ritcher (1999). SPFE offers most of the usual features of any typical windows based programme, 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. A complete Context Sensitive Help with Contents, Index and Search Facilities are available. The package is very easy to use. 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. For establishing database connectivity with Microsoft Access, ActiveX Data Objects (ADO) has been used. For all mathematical calculations and functions standard library that comes with the above languages are used. SPFE requires a microcomputer with a minimum of 16 MB RAM and 5 MB free hard disk space and any 32 Bit Windows Operating System.

3. Features of SPFE

The main features of the package SPFE are (i) Generation of selected design (ii) Randomized layout of the design (iii) Analysis of the data and (iv) Probability calculation as described in the following.

3.1 Generation of the Design

This module generates designs for the following four situations *viz.*, (i) Complete factorial without confounding (ii) Complete factorial with confounding (iii) Fractional factorial plans, and (iv) Balanced confounded designs.
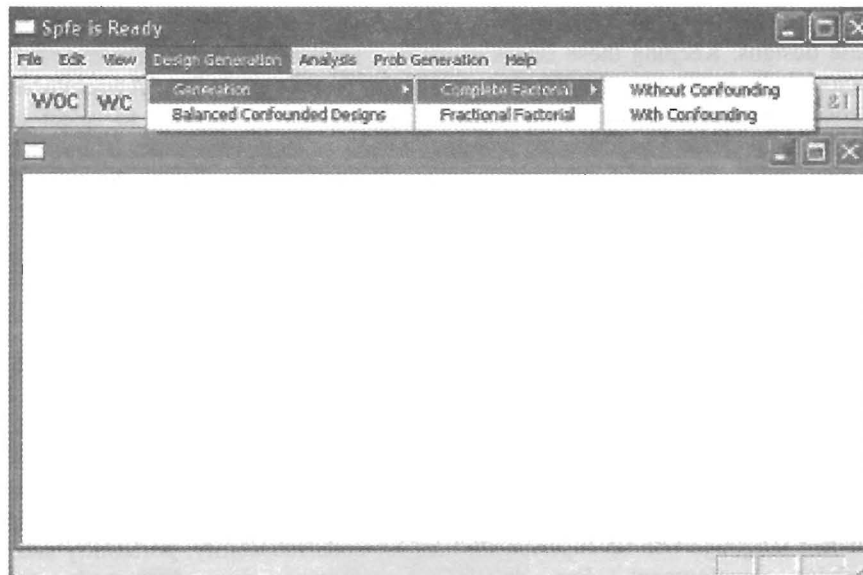
Fig. 1. First screen of SPFE after entering password

When the option *Complete factorial without confounding* is selected, the user is asked to enter the number of factors (n), number of levels of each factor (s) and the number of replications (r) required to generate a design, where the levels of the factors are denoted by 0, 1, ..., s-1. The s^n factorial treatment combinations are generated in the lexicographic order given by $a_1 \times a_2 \times ... \times a_i \times ... \times a_n$, where $\times$ denotes the symbolic direct product and $a_i = (0,1,...,s-1)'$; $i = 1,2,...,n$.. When the option *Complete factorial with confounding* is selected, the user is prompted to enter also the number of independent interactions to be confounded in individual replications. The option *Fractional factorial* gives regular fractional factorials for main effects alone, or main effects and two factor interactions. Selection of option *Balanced confounded designs* provides a list of commonly used balanced confounded designs for symmetrical and asymmetrical factorial experiments. The details of the algorithms used for generation of designs are discussed in the following.

3.1.1 Designs for complete factorial without confounding

The complete set of treatment combinations (s^n) is generated and sets of complete blocks (replications) is formed. Independent randomization is carried out in each block.
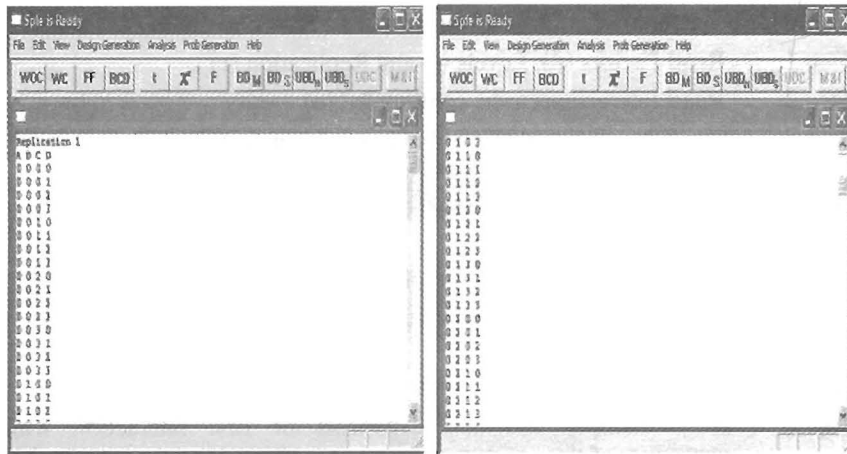
Fig. 2. Generation of design for 4^4 complete factorial experiment without confounding

3.1.2 Generation of designs for complete factorial with confounding

With a large number of factorial combinations for experimentation, the blocks are likely to be incomplete. With incomplete blocks, confounding is inevitable. So one should choose to confound less important effects, such as higher order interactions or those known to be negligible.

This module of software generates factorial design by confounding any number of interactions for the situations when the levels of the factors are prime or prime power.

Let there be n factors each at s levels where $s = p^m$ and p is a prime. The total number of treatments in a factorial is s^n. There are $(s^n - 1)$ degrees of freedom (df) due to treatments. These are divided into $(s^n - 1)/(s - 1)$ main effects and interaction components each at (s-1) df. The components are represented by equation of the form

$$p_1 x_1 + p_2 x_2 + p_3 x_3 + ... + p_n x_n = i \qquad (i = 0,1,\alpha_2,\alpha_3,...,\alpha_{s-1})$$

where $p_1, p_2, ..., p_n$ and $\alpha_2, \alpha_3, ..., \alpha_{s-1}$ are the elements of Galois field, p^m and $x_1, x_2, ..., x_n$ are the variates corresponding to the factors as defined. The package generates the complete factorial with confounding by solving the above equations. For further details on complete factorials with confounding one may refer to Yates (1937), Kempthorne (1952), Cochran and Cox (1957), Federer (1967) and Das and Giri (1986) and the references cited therein.

Steps for generation of factorial experiments with confounding

Step 1 : **Read** Factors (**n**), Levels (**s**) and Replication (**r**)

Step 2 : **Read** Interactions to be confounded (**a**)

Step 3 : Generate all possible combinations for **n** factors each at **s** levels which is equal to s^n and say it **Set A**

Step 4 : Generate total number of blocks, which should be equal to s^{n-a} and say it **Set B**

Step 5 : Take first combination from the **Set A**

//Determination of the location of the each treatment combination in different blocks

Step 6 : Take a variable named **Block** and initialize it to empty (NULL)

Step 7 : Now multiply and add exponents of first interaction with exponents of this combination and append the output number to the **Block** variable

Step 8 : Repeat Step 7 for all the interactions

Step 9 : Now **Block** variable consists of Block Number to which the selected treatment from **Set A** should go

Step 10 : **Write** this treatment to the file named with **Block**

Step 11 : **Go to** Step 5 and **Repeat** till all the combinations are exhausted

Step 12 : **Go to** Step 4 and **Repeat r** times
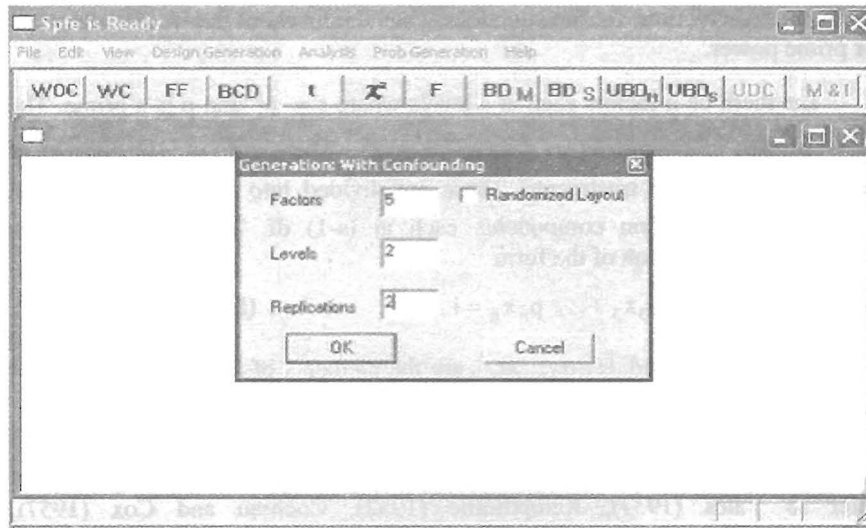
Step 13 : Combine all files and **Display**

Fig. 3. Generation of design for 2^5 complete factorial with confounding in 2 replications
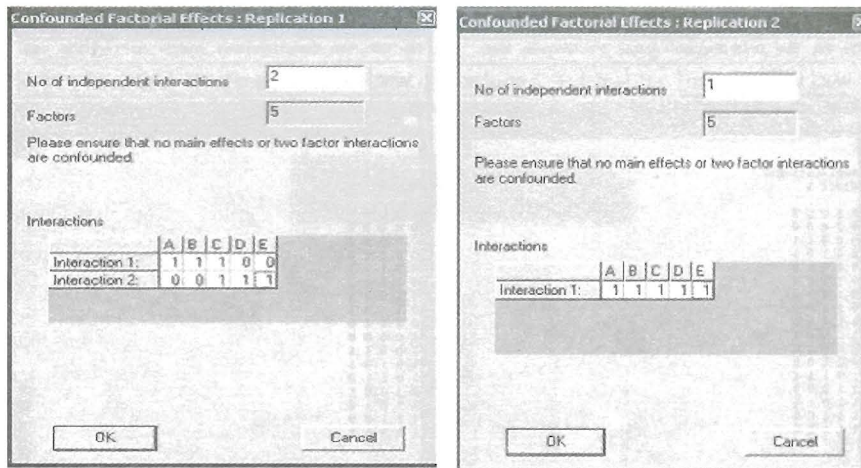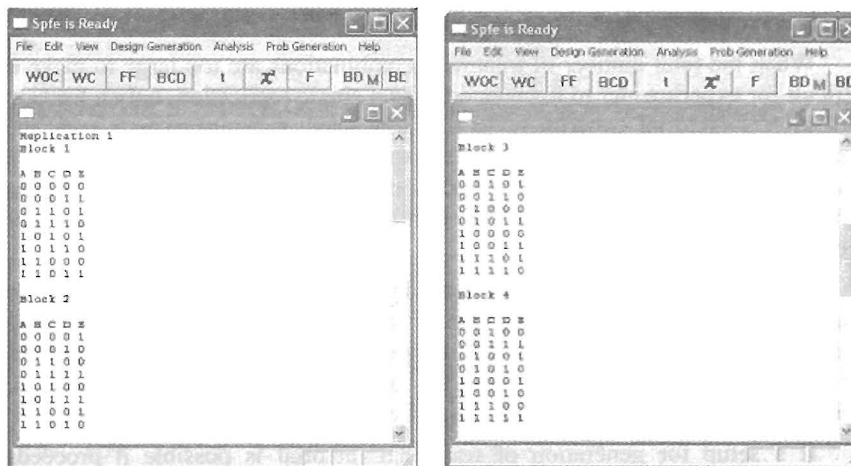
Fig. 4. Confounded factorial effects {Replication 1: 2 independent interactions (ABC and CDE); Replication 2: 1 independent interaction ABCDE}
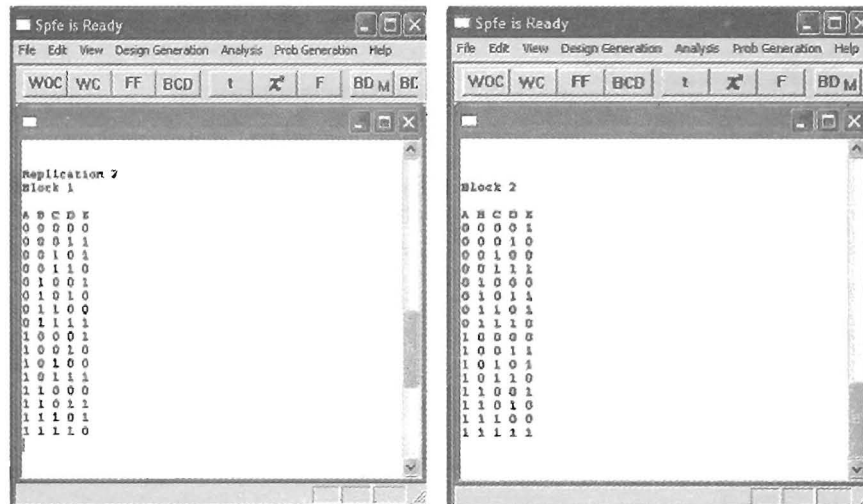
Fig. 5. Results of generation of 2^5 complete factorial with confounding in
2 replications

3.1.3 Generation of designs for fractional factorial

When the total number of treatment combinations are too large to make it difficult for experimenter to manage on even a single replication, only a fraction of complete factorial experiment is normally used. In these experiments, the higher order interactions are taken as the experimental error. For more details on fractional factorial plans, a reference may be made to Finney (1945), Cochran and Cox (1957), Dey (1985) and Dey and Mukerjee (1999) and the references cited therein.

When the option fraction factorial plans is selected, the user is asked to enter the number of factors (n), number of levels of each factor (s), and number of independent interactions to be confounded. After entering these parameters the user is asked to select one of the two available options *viz.* (1) estimating main effects only, or (2) estimating main effects and two factor interactions.

If a setup for generation of fractional factorial is possible it proceeds further; otherwise it gives an error message to the user, and prompts the user to select some other setup. SPFE makes this decision by checking that the number of runs is equal to more than the number of parameters to be estimated. In other words, the lower bound to the number of runs in the first case where only the main effects are estimable is $1 + n * (s - 1)$. Similarly for the second case where main effects and two factor interactions are to be estimated, the lower bound to the number of runs is $1 + n * (s - 1) + {}^n C_2 * (s - 1)^2$.

The confounded independent interaction components are the set of Defining Contrasts or the Identity Group of Interactions. While entering the

confounded interaction components, user should confirm that not less than three-factor interactions (including the generalized interactions) get confounded when the option Main Effects is selected, and not less than five factor interactions (including the generalized interaction components) get confounded when the option Main Effects and Two Factor Interactions is selected.

Steps for generation of fractional factorial plans

Step 1 : **Read** Factors (**n**) and Levels (**s**)

Step 2 : **Read** Interactions to be confounded (**a**)

Step 3 : Ask User to select Main Effect plan/Main effect and Two factor plan

Step 4 : Check, whether generation is possible or not (discussed above)

Step 5 : If generation is possible proceed to step 6 else **return**

Step 6 : Generate confounded factorial design in the same way as explained in **Section 3.1.2**

Step 7 : Retain and display only first block of confounded factorial

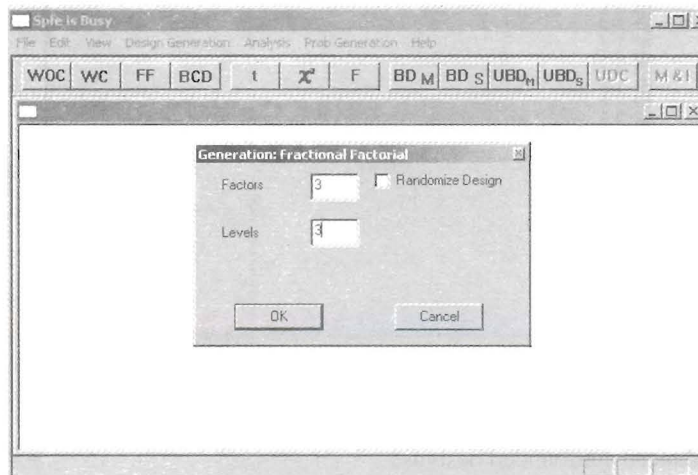This is illustrated with the help of generating $1/3$rd replicate of 3^3 factorial.

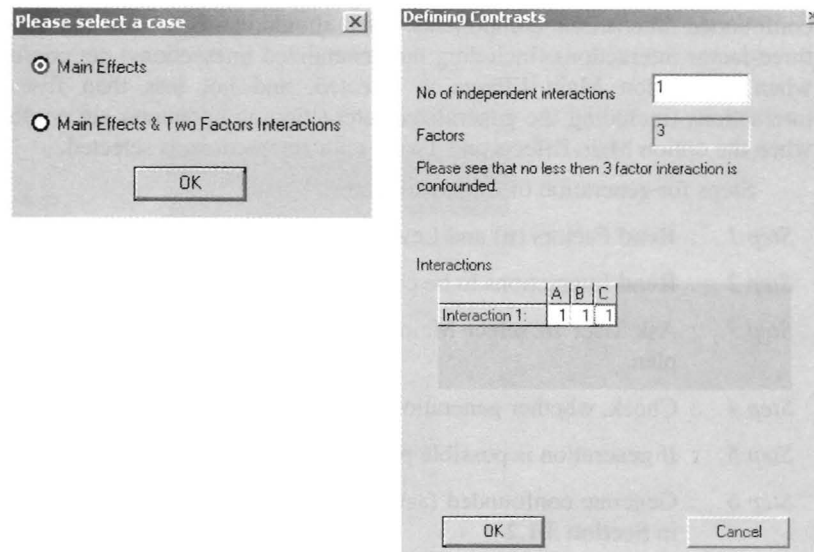**Fig. 6.** Generation of design for fractional factorial plan ($1/3^{rd}$ replicate of 3^3)

Fig. 7. Defining Contrast: ABC

Fig. 8. $1/3(3^3)$ main effect plan

3.1.4 Balanced confounded designs

In a partially confounded design if each of the interaction components of a particular order is confounded an equal number of times, which may be different for different orders of interaction, then the design is called balanced for such interactions. Such designs are useful in situations where the loss of information is to be equally distributed over the different components of the interaction of

given orders. One may refer to Cochran and Cox (1957) for balanced confounded designs and Smith *et al.* (1998) for Database programming advanced concepts and Jeff (1999) and Richter (1999) for Code optimization.

To generate balanced confounded designs from SPFE, select the option Balanced Confounded Designs from the Design Generation. After selecting the option, a design catalog dialog box will open up. Select the particular design and produce the randomized layout.

Steps for Generation of Balanced Confounded Designs

In this module of SPFE a catalogue of available designs is kept in database, and as per requirement, SPFE picks up a design from the database, randomizes it and displays it to the user.

Step 1 : Establish connection to SPFE Database and display available designs

Step 2 : Fetch selected design from database
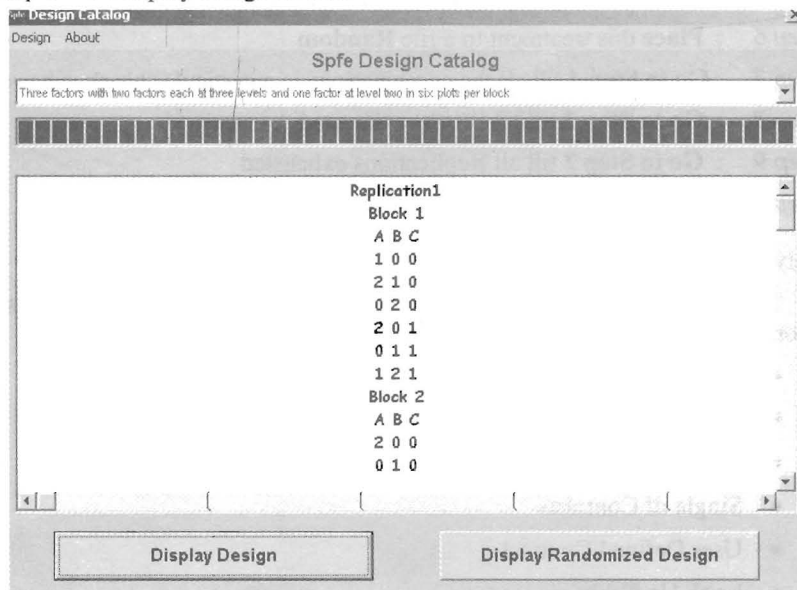
Step 3 : Display design to user

Fig. 9. Results of generation of balanced confounded designs

3.2 Randomized Layout of the Design

Randomization is defined as the random allocation of the treatments to the experimental units without any predefined sequence or procedure. The purpose of randomization is to prevent systematic and personal biases from being introduced into the experiment by the experimenter.

SPFE has a built-in facility to generate randomized layout of designs for all the options discussed in Section 3.1. This randomization includes randomization of the Replications, Blocks within Replications and Treatment combinations within each block. The randomization is achieved by using some standard library functions of Visual C++. These functions generate uniform random variates by taking seed as time the system clock.

3.2.1 Steps for randomized layout of the designs

The randomization of any generated design consists of some simple steps. All these steps are written here.

Step 1 : **Generate** the Required Design (See **Section 3.1**)

Step 2 : **Generate** a random number from 1 to number of Replications

Step 3 : **Generate** a random number from 1 to number of Blocks

Step 4 : **Generate** a random number from 1 to Total Number of Treatments in Block say it as T_Loc

Step 5 : **Read** the treatment combination from the block at location T_Loc

Step 6 : **Place** this treatment to a file **Random**

Step 7 : **Go to Step 4** till all the combinations in a particular block exhausted

Step 8 : **Go to Step 3** till all Blocks exhausted

Step 9 : **Go to Step 2** till all Replications exhausted

Step 10 : Display Design to user in file **Random**

Analysis of Data Generated

The option **Analysis** in the Menu-Bar consists of the following sub-options, which are displayed as POP-Up Menus

- Single Factor
- Multiple Factor
- Main Effects and Interactions
- Single df Contrasts
- User Defined Contrast
- Look Up Table

The options Single Factor, Multiple Factor and Look Up Table always remain activated and other options remain deactivated. Each of the options Single Factor and Multiple Factor have two sub-options viz. **Unblocked Designs (Design with no blocking)** and **Block Designs**. For the analysis of data user has to create a data file in ASCII mode. The details for creation of data file and steps to carry out the analysis are described in the sequel.

(i) Single Factor Unblocked Design : Here v treatments are applied to a total of N homogeneous experimental units. Let the treatment i be replicated r_i times such that $\sum_{i=1}^{v} r_i = N$. For the analysis, a data file with two columns and N rows is to be created. In the first column enter the treatment number and in the second column enter the corresponding value of the observed response. The two columns are separated by a Space/Tab.

(ii) Single Factor Block Design : relates to an experimental situation where v treatments are applied on N experimental units, which are grouped in b homogeneous blocks such that the jth block contains k_j experimental units (Thus, $\sum_{j=1}^{b} k_j = N$). For the analysis of the data generated from this experiment, a data file with 3 columns and N rows is to be created. The three columns correspond to the block number, treatment number and the corresponding value of the observed response, respectively. Space/Tab separates each column.

(iii) Multiple Factor Unblocked Design : In this case, an experiment with n factors, denoted as, $F_1, F_2, \cdots, F_n$ and the ith factor having s_i levels, $i = 1, 2, \ldots, n$ is considered. The treatment combinations $\prod_{i=1}^{n} s_i = v$ (say) in lexicographic order are given by $a_1 \times a_2 \times \ldots \times a_i \times \ldots \times a_n$, where $\times$ denotes the symbolic direct product and $a_i = (0, 1, \ldots, s_i - 1)'$; $i = 1, 2, \ldots, n$. Suppose that this factorial experiment is conducted using a completely randomized design in N experimental units. The replications of the treatment combinations may vary. For the analysis of the data generated from this experiment, the data file should be prepared in n + 1 columns and N rows. First n columns represent the level codes of the n-factors and the last column represents the value of the corresponding observed response. The number of rows represents the number of observations. Space/Tab separates each column.

(iv) Multiple Factor Design with Blocking relates to an experimental situation described in Multiple Factor Unblocked Design with the only change that the factorial experiment is conducted using a block design in N experimental units. The block size and replication number of the treatment combinations may vary. For the analysis of the data generated from this experiment, the data file is prepared in n + 2 columns and N rows. First column represents the block number and next n columns represent the level codes of the n-factors and the last column represents the corresponding observation. The number of rows represents the number of observations. Space/Tab separates each column.

Once the data file is created, the user can analyze the data using the appropriate option from the Menu-Bar.

The option *Single Factor* is useful for the analysis of data from single factor experiments conducted using an unblocked (completely randomized) design or a block design. If the data is generated through a completely randomized design, then select the option *Unblocked Designs* under the option Single Factor. If the data is generated from a block design (randomized complete block design or an incomplete block design), then select the option *Block Designs* under the option Single Factor. Once the option *Unblocked Designs/Block Designs* is selected, the data file name is to be provided either directly writing the complete path of the file or selecting the file through the Browse Button on the Window Dialog Box. The data is then analyzed as per usual procedure of unblocked/ block designs for single factor experiments. It provides Analysis of Variance (ANOVA) table with probability level for testing the equality of the treatment or treatment and block effects. In field experiments, blocks are fixed segment of the land and unlike treatments, blocks cannot be randomized. Therefore, one may disagree on testing the significance of block effects. For these situations, one may assess the contribution of blocks by adjusted sum of squares provided in the ANOVA table. Besides ANOVA table, it also gives the value of coefficient of determination (R^2) of the model fitted root mean square error. Further, the package is also capable of testing the significance of user-defined treatment contrasts. This is followed by

- Treatment Means and their Standard Errors

- All possible Pairwise Treatment Comparisons Prob > F ($v \times v$ matrix with diagonal elements as . and off-diagonal elements as exact level of significance of the equality of corresponding treatment effects)

- The critical differences at 1% and 5% level of significance are also displayed if the treatments are equireplicated (in case of unblocked designs) and equireplicated within each block in case of blocked designs. In other words, critical differences are displayed for balanced and orthogonal data only.

The analysis of all block designs *viz.* RCB designs, balanced incomplete block (BIB) designs, partially balanced incomplete block (PBIB) designs, variance balanced block designs, etc. follows as a particular case of this analysis and has been implemented in SPFE. SPFE also validates the *disconnected design* and prompts the message that design is disconnected and analysis cannot be carried out.

Note: Once the file is selected and the analysis is carried out, the option User Defined Contrasts gets activated. This option can be used for testing the significance of Comparisons of Interest. A Window Dialog Box opens up on selection of the option User Defined Contrasts. For details on this option see User Defined Contrasts in Section 3.3.1.

The option *Multiple Factor* is useful for the analysis of data from multiple factor experiments conducted using an unblocked (completely randomized) design or a block design. If the data is generated through a completely randomized design then select the option *Unblocked Designs* under the option Multiple Factor. If the data is generated from a block (complete or incomplete) design, then select the option *Block Designs* under the option Multiple Factor.

For multiple factor experiments, the $\prod_{i=1}^{n} s_i = v$ treatment combinations are renumbered automatically in lexicographic order from 1 to v, where s_i denotes the levels of factor i, $i = 1, 2, \ldots, n$. The identity relationship between the renumbered treatments and the treatment combinations can be seen using the Look Up Tables explained in Section 3.3.2. The data is then analyzed as per usual procedure of unblocked/block designs for single factor experiments. The contrasts for main effects and interactions are generated by the package. The sum of squares for main effects and interactions are then obtained along with probability level of significance. The package also facilitates generation of the single degree of freedom contrasts for various factorial effects using the orthogonal polynomials for equi-spaced factor-levels and testing their significance.

The means of all combinations of p $(1 \leq p \leq n)$ factors averaged over levels of all the other $(n - p)$ factors can be obtained. Further, if all the level combination of p-factors are appearing same number of times in each of the b blocks, say 'a', then we can obtain the least significant difference for comparing the effect of p-factor combinations means averaged over levels of all other factors by computing a least significant difference (lsd) given by

$$lsd = t_{\alpha, edf} \times \sqrt{\frac{2\,MSE}{ab}}$$

where edf is experimental error degrees of freedom and α is level of significance (Cochran and Cox (1957)). For a detailed discussion and further reading one may refer Gupta and Mukerjee (1989), Parsad *et al.* (2004), Nigam *et al.* (1988) and references cited therein.

3.2.2 User Defined Contrasts

Once analysis is carried out, the experimenter can test the significance of treatment contrasts of his/her interest. This option can be used for testing the significance of treatment contrasts of interest. SPFE computes and analyzes the user defined contrasts and gives results such as ANOVA along with contrasts and their estimates. Facility is there in SPFE to input + and – coefficients to represent the contrasts, in parallel to the columns of the treatments. This is shown with the help of following screen shots.
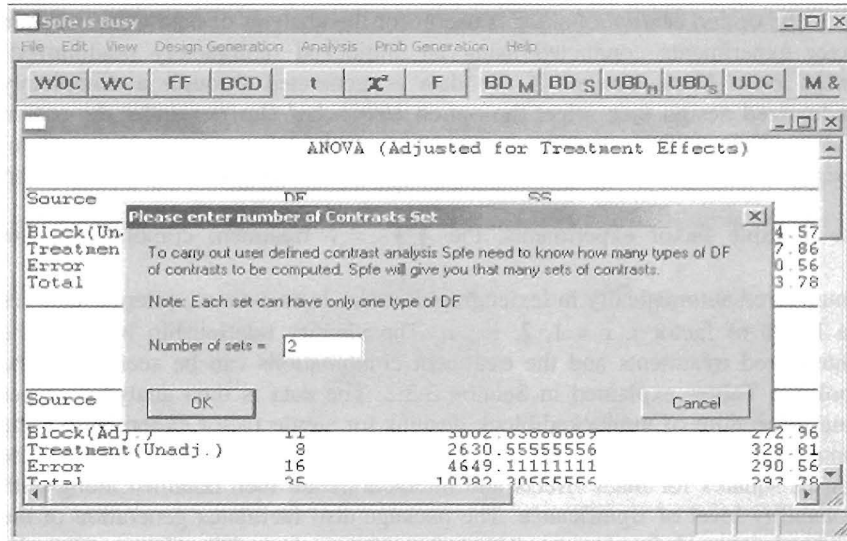
Fig. 10. User defined contrast analysis after ANOVA for the data for 9 treatment combinations (3^2) in 12 blocks. Number of the sets to be entered by the user in the above case are 2
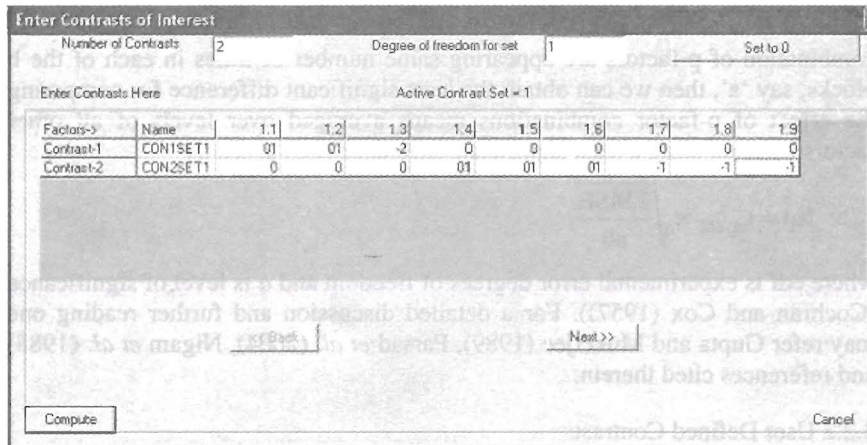

Fig. 11. Enter the number of contrasts, degree of freedom and the positive and negative coefficients representing the contrasts for first set
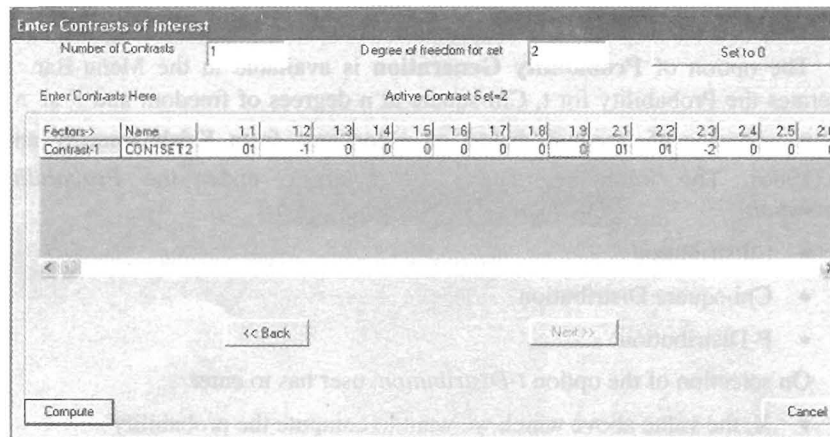
Fig. 12. Enter the number of contrasts, degree of freedom and the positive and negative coefficients representing the contrasts for second set
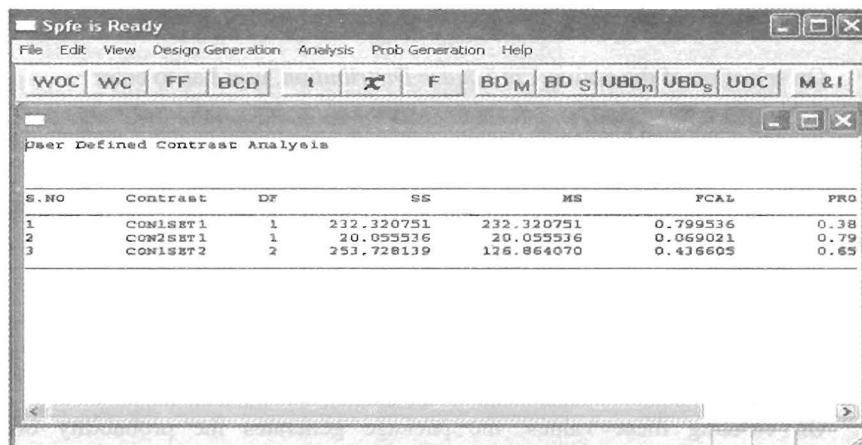

Fig. 13. Results of the User defined contrast analysis for all the sets

3.2.3 Lookup Table

The option Lookup Table essentially gives the identity relationship between the treatment combinations and the renumbered treatments. On selection of option Lookup Table from the option Analysis from the Menu-Bar, Window Dialog Box with options of Factors and Levels opens up. User is required to enter the number of factors and levels of each of the factors. On clicking OK button one gets the identity relationship between the renumbered treatments and the treatment combinations.

3.3 Probability Generation

The option of **Probability Generation** is available in the Menu-Bar. It generates the Probability for t, Chi-square at n degrees of freedom and F at n_1 and n_2 degrees of freedom, based on algorithms from Krishnamurthy and Sen (1986). The following options are available under the *Probability Generation*

- t-distribution
- Chi-square Distribution
- F-Distribution

On selection of the option *t-Distribution*, user has to enter

- X, the value above which we want to compute the probability
- n, the degrees of freedom

On entering these values, it generates the probability of t-distribution greater than X i.e. $\text{Prob}[|t| > X]$.

On selection of the option *Chi Square Distribution*, user has to enter

- X, the value above which we want to compute the probability
- n, the degrees of freedom

On entering these values, it generates the probability of Chi-square distribution greater than X i.e. $\text{Prob}[\chi^2 > X]$.

On selection of the option *F-Distribution*, user has to enter

- X, the value above which we want to compute the probability
- n_1 and n_2, the degrees of freedom

On entering these values, the package generates the probability of F-distribution greater than X i.e. $\text{Prob}[F > X]$. The procedure of entering the required values and output are displayed in the following figure.
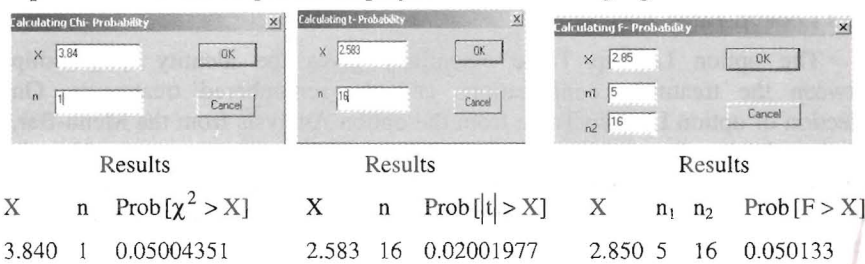

Results			Results			Results			
X	n	$\text{Prob}[\chi^2 > X]$	X	n	$\text{Prob}[|t| > X]$	X	n_1	n_2	$\text{Prob}[F > X]$
3.840	1	0.05004351	2.583	16	0.02001977	2.850	5	16	0.050133

Fig. 14. Results of the Chi-Square, t and F probability values

4. SPFE Help

A complete Context Sensitive Help with **Index**, **Contents** and **Search** facility is available. The package can be run without the aid of a manual.
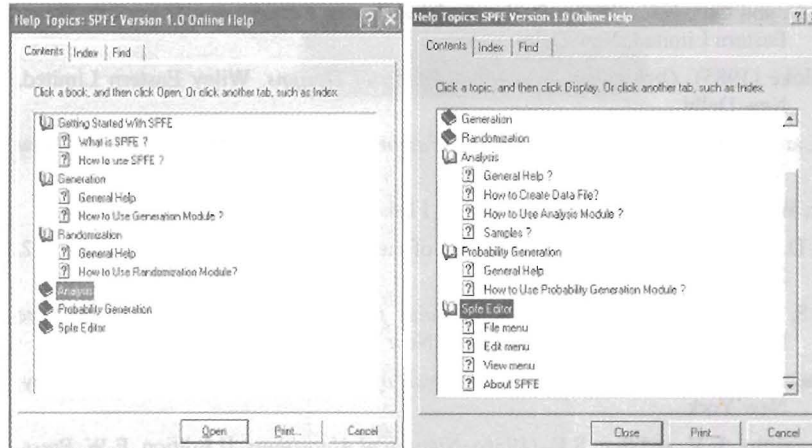

Fig 15.   SPFE Context Sensitive Help

5. Limitations and Future Requirements

At present the analysis of the data from factorial experiments with single replication, completely confounded factorial experiments and fractional factorial plans can not be performed using the software. Therefore, these features are yet to be included in the package. The generation of generalized interaction components confounded has been left to the user. The code need to be improved further, so that the set of generalized interactions is generated once the user enters the independent interactions. The package assumes that there are no missing observations in the data. However, the data with missing observations can be analyzed by deleting the row corresponding to the missing observation from the data file and assuming that the experiment was conducted with only those experimental units from which the response is available.

REFERENCES

Cochran, W.G. and Cox, G.M. (1957). *Experimental Designs*, II Edition. John Wiley and Sons, Singapore.

Das, M.N. and Giri, N.C. (1986). *Design and Analysis of Experiments*, II Edition. Wiley Eastern Limited, New Delhi.

Dey, Aloke (1985). *Orthogonal Fractional Factorial Designs.* Wiley Eastern Limited, New Delhi.

Dey, A. and Mukerjee, R. (1999). *Fractional Factorial Plans.* John Wiley and Sons, New York.

Federer, W. T. (1967). *Design of Experiments,* I Ed. John Wiley and Sons, New York.

Finney, D.J. (1945). The fractional replication of factorial arrangements. *Ann. Eugen.*, **12**, 291-301.

Gupta, S. and Mukerjee, R. (1989). *A Calculus for Factorial Arrangements.* Lecture Notes in Statistics, Springer-Verlag, New York.

Kempthorne, O. (1952). *The Design and Analysis of Experiments,* I Edition. John Wiley, New York.

Krishnamurthy, E.V. and Sen, S.K. (1986). *Numerical Algorithms*, II Edition. E.W. Press, New Delhi.

Kruglinski, David. (1996). *Inside Visual C++*, I Edition. Microsoft Press, Washington.

Nigam, A.K, Puri, P.D and Gupta, V.K. (1988). *Characterizations and Analysis of Block Designs.* Wiley Eastern, New Delhi.

Parsad, R., Gupta, V.K., Batra, P.K, Srivastava, R., Kaur, R., Kaur, A. and Arya, P. (2004). *A Diagnostic Study of Design and Analysis of Field Experiments.* Technical Report, IASRI, New Delhi.

Jeff, Prosise (1999). *Programming Windows with MFC,* II Edition. Microsoft Press, Washington.

Richter, Jeffrey (1999). *Programming Applications for Microsoft Windows*, IV Edition. Microsoft Press, Washington.

Smith Eric, A., Whisler, Valore and Marquis, Hank (1998). *Visual Basic 6 Programming Bible,* I Ed. John Wiley and Sons, New York.

Yates, F. (1937). The Design and Analysis of Factorial Experiments. *Imp. Bur. Soil. Sci.*, Tech. Comm., 35.