SPSS: AN OVERVIEW

SEEMA JAGGI
Indian Agricultural Statistics Research Institute
Library Avenue, New Delhi-110 012
seema@iasri.res.in

1. Introduction

The abbreviation SPSS stands for **Statistical Package for the Social Sciences** and is a comprehensive system for analysing data. This package of programs is available for both personal and mainframe (or multi-user) computers. SPSS package consists of a set of software tools for data entry, data management, statistical analysis and presentation. SPSS integrates complex data and file management, statistical analysis and reporting functions. SPSS can take data from almost any type of file and use them to generate tabulated reports, charts, and plots of distributions and trends, descriptive statistics, and complex statistical analyses.

Features of SPSS

(i) It is easy to learn and use
(ii) It includes a full range of data management system and editing tools
(iii) It provides in-depth statistical capabilities
(iv) It offers complete plotting, reporting and presentation features.

SPSS makes statistical analysis accessible for the casual user and convenient for the experienced user. The data editor offers a simple and efficient spreadsheet-like facility for entering data and browsing the working data file. To invoke SPSS in the windows environment, select the appropriate **SPSS** icon. There are a number of different types of windows in SPSS.

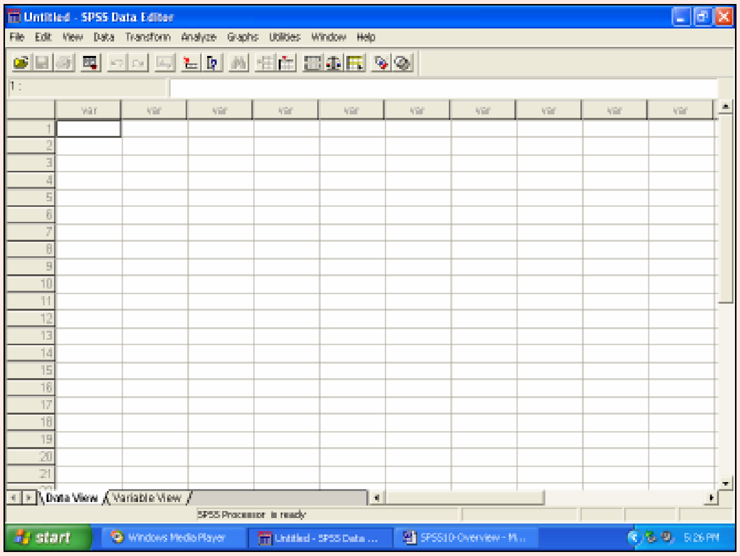
Data Editor. This window displays the contents of the data file. One can create new data files or modify existing ones. The Data Editor window opens automatically when one starts an SPSS session. One can have only one data file open at a time. This editor provides two views of the data.

- **Data view**. Displays the actual data values or defined value labels.
- **Variable view**. Displays variable definition information, including defined variable and value labels, data type etc.

With the Data Editor, one can modify data values in the Data view in many ways like change data values; cut, copy and paste data values; add and delete cases; add and delete variables, change the order of variables.

Viewer. All statistical results, tables, and charts are displayed in the Viewer. The output can be edited and saved for later use. A Viewer window opens automatically the first time you run a procedure that generates output.

Draft Viewer. The output can be displayed as a simple text in this window.



Syntax Editor. One can paste the dialog box choices into a syntax window, where the selections appear in the form of command syntax. One can then edit the command syntax to utilize special features of SPSS not available through dialog boxes. These commands can be saved in a file for use in subsequent SPSS sessions.

Pivot Table Editor. Output is displayed in pivot tables that can be modified in many ways with this editor. One can edit text, swap data in rows and columns, create multidimensional tables, and selectively hide and show results.

Text Output Editor. Text output not displayed in pivot tables can be modified with the Text Output Editor. One can edit the output and change font characteristics (type, style, colour, size).

Chart Editor. High-resolution charts and plots can be modified in chart windows. One can change the colours, select different type of fonts and sizes etc.

Many of the tasks that are to be performed with SPSS start with **menu** selections. Each window has its own menu bar with menu selections appropriate for that window type. The various procedures under SPSS are

File Edit View Data Transform Analyze Graphs Utilities Windows Help

Analyze and Graphs menus are available on all windows, making it easy to generate new output without having to switch windows. Most menu selections open dialog boxes. One can use dialog boxes to select variables and options for analysis. Since most procedures provide a great deal of flexibility, not all of the possible choices can be contained in a single dialog box. The main dialog box usually contains the minimum information required to run a procedure. Additional specifications are made in subdialog boxes. All these above mentioned options have further suboptions. To see what applications there are, we simply move the cursor to a particular option and press, when a drop-down menu will appear. To cancel a drop-down menu, place the cursor anywhere outside the option and press the left button.

The three dots after an option term (...) on a drop-down menu, such as **Define Variable**...option in Data option, signifies that a dialog box will appear when this option is chosen. To cancel a dialog box, select the **Cancel** button in the dialog box. A right-facing arrowhead after an option term indicates that a further submenu will appear to the right of the drop-down menu. An option with neither of these signs means that there are no further drop-down menus to select. There are five standard command pushbuttons in most dialog boxes.

OK. Runs the procedure. After the variables and additional specifications are selected, click OK to run the procedure.

Paste. Generates command syntax from the dialog box selections and pastes the syntax into a syntax window.

Reset. Deselects any variables in the selected variable list and resets all specifications in the dialog box.

Cancel. Cancels any changes in the dialog box settings since the last time it was opened and closes the dialog box.

Help. Contains information about the current dialog box.

Entering and Editing data

The easiest way of entering data in SPSS is to type it directly into the matrix of columns and numbered rows in the **Data Editor** window. The columns represent variables and the rows represent cases. The variables can be defined in the variable view. Variable name must be no longer than eight characters and the name must begin with a letter.

Saving data

To be able to retrieve a file, we need to save it and give it a name. The default extension name for saving files is **sav**. Thus, we could call our data file **see.sav**. To save this file on a floppy disk, we carry out the following sequence:
→**File** →**Save As...** [opens **Save Data As** dialog box]→box under **Drives:** →drive [e.g. **a**] from options listed→box under **File Name:**, delete the asterisk and type file stem name [e.g. see] →**OK**

The output file can also be printed and saved. The extension name for output file is **spo**.

Retrieving a saved file

To retrieve this file at a later stage when it is no longer the current file, use the following procedure:
→**File**→**Open**→**Data...**[opens the **Open Data File** dialog box]
→box under **Drives:** →drive [e.g. **a**]from options listed
→box under **File Name:** →file name [e.g. **see.sav**] → **OK**

Basic Steps in Data Analysis

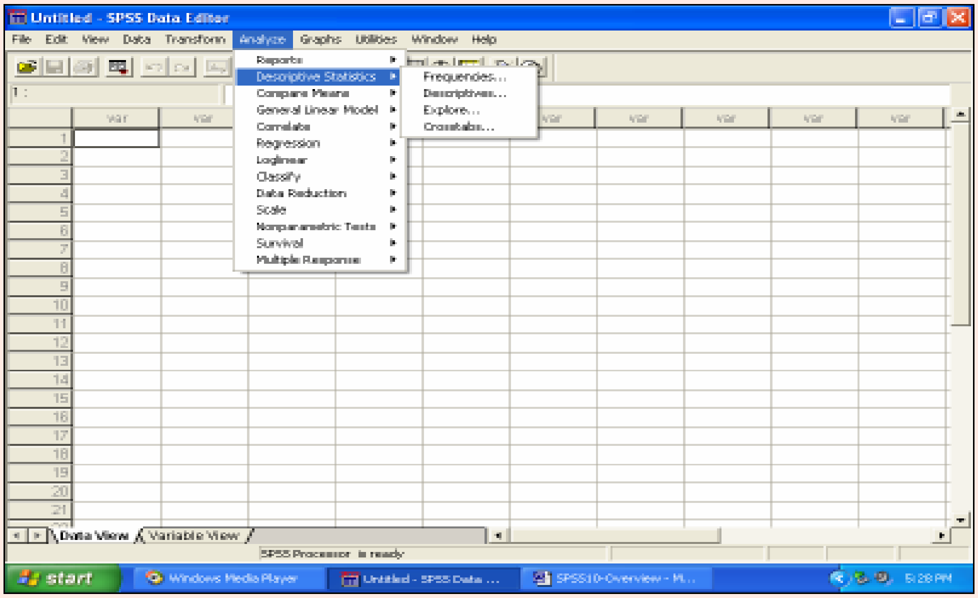
- **Get your data into SPSS**. You can open a previously saved SPSS data file, read a spreadsheet, database, or text data file, or enter your data directly in the Data Editor.

- **Select a procedure**. Select a procedure from the menus to calculate statistics or to create a chart.
- **Select the variables for the analysis**. The variables in the data file are displayed in a dialog box for the procedure.
- **Run the procedure**. Results are displayed in the Viewer.

2. Statistical Procedures

After entering the data set in **Data Editor** or reading an ASCII data file, we are now ready to analyse it. The **Analyse** option has the following sub options:

Reports, Descriptive Statistics, Compare means, General Linear model, Correlate, Regression, Loglinear, Classify, Data Reduction, Scale, Non parametric tests, Time Series, Survival, Multiple response.



2.1 Descriptive Statistics

This submenu provides techniques for summarising data with statistics, charts, and reports. The various sub-sub menus under this are as follows:

Frequencies provide information about the relative frequency of the occurrence of each category of a variable. This can be used it to obtain summary statistics that describe the typical value and the spread of the observations. To compute summary statistics for each of several groups of cases, Means procedure or the Explore procedure can be used.

Descriptives is used to calculate statistics that summarize the values of a variable like the measures of central tendency, measures of dispersion, skewness, kurtosis etc.

Explore produces and displays summary statistics for all cases or separately for groups of cases. Boxplots, stem-and leaf plots, histograms, tests of normality, robust estimates of location, frequency tables and other descriptive statistics and plots can also be obtained.

Crosstabs is used to count the number of cases that have different combinations of values of two or more variables, and to calculate summary statistics and tests. The variables you use to form the categories within which the counts are obtained should have a limited number of distinct values.

List Cases displays the values of variables for cases in the data file.

Report Summaries in Rows produces reports in which different summary statistics are laid out in rows. Case listings are also available from this command, with or without summary statistics.

Report Summaries in Columns produces reports in which different summary statistics are laid out in separate columns.

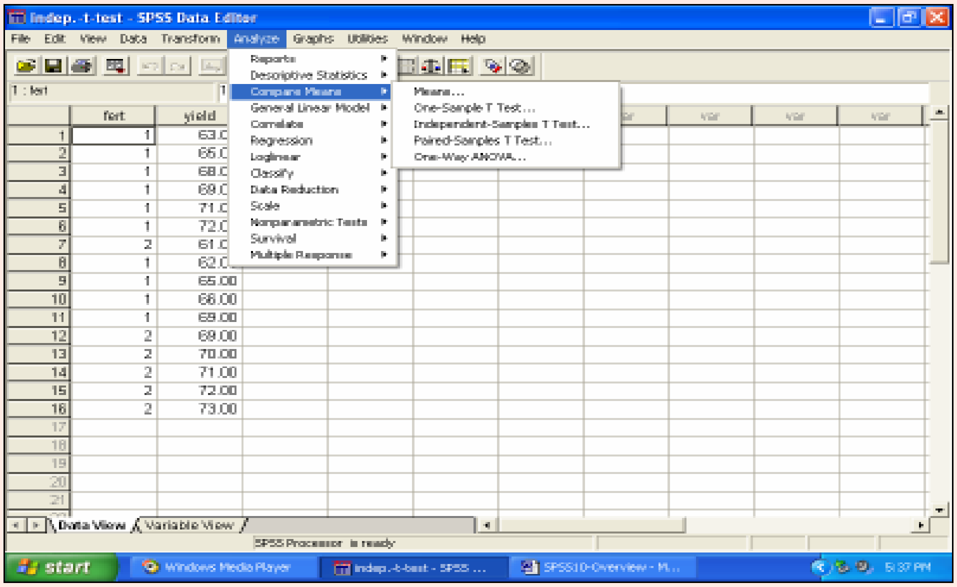
Custom Tables submenu provides attractive, flexible displays of frequency counts, percentages and other statistics.

2.2 Compare Means

This submenu provides techniques for testing differences among two or more means for both independent and related samples.

Means computes summary statistics for a variable when the cases are subdivided into groups based on their values for other variables.

Independent Sample t test is used if two unrelated samples come from populations with the same mean. The observations should be from two unrelated groups, and for testing, the mean must be an appropriate summary measure for the variable to be compared in the two groups. For more than two independent groups, the **One-way ANOVA** option could be used.



Paired Sample t test is used to compare the means of the same subjects in two conditions or at two points in time i.e. to compare subjects who had been matched to be similar in certain respects and then to test if two related samples come from populations with the same mean. The related, or paired, samples often result from an experiment in which the same person is observed before and after an intervention. If the distribution of the differences of the values between the members of a pair is markedly nonnormal you should consider one of the nonparametric tests.

One-Way ANOVA is used to test that several independent groups come from populations with the same mean. To see which groups are significantly different from each other, multiple comparison procedures can be used through **Post Hoc Multiple Comparison option** which consist of the options like **Least-significant difference, Duncan's multiple range test, Scheffe** etc.. The contrast analysis can also be performed in order to compare the different groups or treatments by using the **Contrast** option. The data obtained using completely randomised design can be analysed through this option.

2.3 General Linear Model

This submenu provides techniques for testing univariate and multivariate Analysis of Variance models, including repeated measures. The **Univariate** suboption could be used to analyse the experimental designs like Completely randomised design, Randomised block design, Latin square design, Designs for factorial experiments etc.

The covariace analysis can also be performed and alternate methods for partitioning sums of squares can be selected.

If only some of the interactions of a particular order are to be included, the **Custom** procedure should be used. If there is only one factor then One-Way ANOVA procedure should be used..

Multivariate analyses analysis-of-variance and analysis-of-covariance designs when you have two or more correlated dependent variables.

Multivariate analysis of variance is used to test hypotheses about the relationship between a set of interrelated dependent variables and one or more factor or grouping variables. For example, you can test whether verbal and mathematical test scores are related to instructional method used, sex of the subject, and the interaction of method and sex.

This procedure should be used only if there are several dependent variables which are related to each other. For a single dependent variable or unrelated dependent variables, the Univariate ANOVA procedures can be adopted. If the same dependent variable is measured on several occasions for each subject, the Repeated Measures procedure is to be used.

Repeated Measures is used to test hypotheses about the means of a dependent variable when the same dependent variable is measured on more than one occasion for each subject.

Subjects can also be classified into mutually exclusive groups, such as males or females, or type of job held. Then you can test hypotheses about the effects of the between-subject variables and the within-subject variables, as well as their interactions.

2.4 Correlate

This submenu provides measures of association for two or more variables measured at the interval level.

Bivariate calculates matrices of Pearson product-moment correlations, and of Kendall and Spearman nonparametric correlations, with significance levels and optional univariate statistics.

The **correlation coefficient** is used to quantify the strength of the linear relationship between two variables.

The **Pearson correlation coefficient** should be used only for data measured at the interval or ratio level. Spearman and Kendall correlation coefficients are nonparametric measures which are particularly useful when the data contain outliers or when the distribution of the variables is markedly nonnormal. Both the Spearman and Kendall coefficients are based on assigning ranks to the variables.

Partial calculates **partial correlation coefficients** that describe the relationship between two variables, while adjusting for the effects of one or more additional variables.

If the values of a dependent variable from a set of independent variables is to be predicted then the Linear Regression procedure may be used. If there are no control variables then the Bivariate Correlations procedure can be adopted. Nominal variables should not be used in the partial correlation procedure.

2.5 Regression

This submenu provides a variety of regression techniques, including linear, logistic, nonlinear, weighted, and two-stage least-squares regression.

Linear is used to examine the relationship between a dependent variable and a set of independent variables. If the dependent variable is dichotomous, then the logistic regression procedure should be used. If the dependent variable is censored, such as survival time after surgery, use the Life Tables, Kaplan-Meier, or proportional hazards procedure.

Logistic estimates regression models in which the dependent variable is dichotomous.

If the dependent variable has more than two categories, use the Discriminant procedure to identify variables which are useful for assigning the cases to the various groups. If the dependent variable is continuous, use the Linear Regression procedure to predict the values of the dependent variable from a set of independent variables.

Probit performs probit analysis which is used to measure the relationship between a response proportion and the strength of a stimulus.

For example, the probit procedure can be used to examine the relationship between the proportion of plants dying and the strength of the pesticide applied or to examine the relationship between the proportion of people buying a product and the magnitude of the incentive offered. The probit procedure should be used only if the response is dichotomous-buy/not buy, alive/dead and several groups of subjects are exposed to different levels of some stimulus. For each stimulus level, the data must contain counts of the totals exposed and the totals responding.

If the response variable is dichotomous but you do not have groups of subjects with the same values for the independent variables, one should use the Logistic Regression procedure.

Nonlinear estimates nonlinear regression models, including models in which parameters are constrained.

The nonlinear regression procedure can be used if one knows the equation whose parameters are to be estimated, and the equation cannot be written as the sum of parameters times some function of the independent variables. In nonlinear regression the parameter estimates are obtained iteratively.

If the function is linear, or can be transformed to a linear function, then the Linear Regression procedure should be used.

The **Loglinear** submenu provides general and hierarchical log-linear analysis and logit analysis.

2.6 Classify

This submenu provides cluster and discriminant analysis.

K-means Cluster performs cluster analysis using an algorithm that can handle large numbers of cases, but that requires you to specify the number of clusters.

The goal of cluster analysis is to identify relatively homogeneous groups of cases based on selected characteristics.

If the number of clusters to be formed is not known, then Hierarchical Cluster procedure can be used. If the observations are in known groups and one wants to predict group membership based on a set of independent variables, then the Discriminant procedure can be used.

Hierarchical Cluster combines cases into clusters hierarchically, using a memory-intensive algorithm that allows you to examine many different solutions easily.

Discriminant is used to classify cases into one of several known groups on the basis of various characteristics. To use the Discriminant procedure the dependent variable must have a limited number of distinct categories. Independent variables that are nominal must be recoded to dummy or contrast variables.

If the dependent variable has two categories, Logistic Regression can be used. If the dependent variable is continuous one may use Linear Regression.

2.7 Data Reduction

This submenu provides factor analysis, correspondence analysis, and optimal scaling.

Factor is used to identify factors that explain the correlations among a set of variables. Factor analysis is often used to summarize a large number of variables with a smaller number of derived variables, called factors.

Distances computes many different measures of similarity, dissimilarity or distance. Many different measures can be used to quantify how much alike or how different two cases or variables are. Similarity measures are constructed so that large values indicate much similarity and small values indicate little similarity. Dissimilarity measures estimate the distance or unlikeness of two cases. A large dissimilarity value tells that two cases or variables are far apart. In order to decide which similarity or dissimilarity measure to use, one must consider the characteristics of the data. Special measures are available for interval data, frequency counts, and binary data. If the cases are to be classified into groups based on similarity or dissimilarity measures, one of the Cluster procedures should be used.

The **Conjoint** submenu provides for the generation and analysis of conjoint designs.

2.8 Scale

This submenu provides reliability analysis and multidimensional scaling.

2.9 Nonparametric Tests

This submenu provides nonparametric tests for one sample, or for two and more paired or independent samples.

Chi-Square is used to test hypotheses about the relative proportion of cases falling into several mutually exclusive groups. For example, if one wants to test the hypotheses that people are equally likely to buy six different brands of cereals, one can count the number buying each of the six brands. Based on the six observed counts Chi-Square procedure could be used to test the hypothesis that all six cereals are equally likely to be bought. The expected proportions in each of the categories don't have to be equal. The hypothetical proportions to be tested should be specified.

Binomial is used to test the hypothesis that a variable comes from a binomial population with a specified probability of an event occurring. The variable can have only two values. For example, to test that the probability of an item on the assembly line is defective is one out of ten (p=0.1), take a sample of 300 items and record whether each is defective or not. Then use the binomial procedure to test the hypothesis of interest.

Runs is used to test whether the two values of a dichotomous variable occur in a random sequence. The runs test is appropriate only when the order of cases in the data file is meaningful.

1-Sample Kolmogorov-Smirnov is used to compare the observed frequencies of the values of an ordinal variable, such as rated quality of work, against some specified theoretical distribution. It determines the statistical significance of the largest difference between them. In SPSS, the theoretical distribution can be **Normal, Uniform or Poisson**.

Alternative tests for normality are available in the Explore procedure, in the Summarize submenu. The P-P and Q-Q plots in the Graphs menu can also be used to examine the assumption of normality.

2-Independent Samples is used to compare the distribution of a variable between two nonrelated groups. Only limited assumptions are needed about the distributions from which the sample are selected. The Mann-Whitney U test is an alternative to the two sample t-test. The actual values of the data are replaced by ranks. The Kolmogorov-Smirnov test is based on the differences between the observed cumulative distributions of the two groups. The Wald-Woflowitz runs tests sorts the data values from smallest to largest and then performs a runs test on the groups numbers. The Moses Test of Extreme Reaction is used to test for differences in range between two groups.

K-Independent Samples is used to compare the distribution of a variable between two or more groups. Only limited assumptions are needed about the distributions from which the samples are selected. The Kruskal-Wallis test is an alternative to one-way analysis of variance, with the actual values of the data replaced by ranks. The Median test counts the number of cases in each group that are above and below the combined median, and then performs a chi-square test.

2 Related Samples is used to compare the distribution of two related variables. Only limited assumptions are needed about the distributions from which the samples are selected. The Wilcoxon and Sign tests are nonparametric alternative to the paired samples t-test. The Wilcoxon test is more powerful than the Sign test.

McNemar's test is used to determine changes in proportions for related samples. It is often used for "before and after" experimental designs when the dependent variable is dichotomous. For example, the effect of a campaign speech can be tested by analyzing the number of people whose preference for a candidate changed based on the speech. Using McNemar's test you analyze the changes to see if change in both directions is equally likely.

K Related Samples is used to compare the distribution of two or more related variables. Only limited assumptions are needed about the distributions from which the samples are selected. The Friedman test is a nonparametric alternative to a single-factor repeated measures analysis of variance. You can use it when the same measurement is obtained on several occasions for a subject. For example, the Friedman test can be used to compare consumer satisfaction of 5 products when each person is asked to rate each of the products on a scale.

Cochran's Q test can be used to test whether several dichotomous variables have the same mean. For example, if instead of asking each subject to rate their satisfaction with five products, you asked them for a yes/no response about each, you could use Cochran's test to test the hypothesis that all five products have the same proportion of satisfied users.

Kendall's W measures the agreement among raters. Each of your cases corresponds to a rater, each of the selected variables is an item being rated. For example, if you ask a sample of customers to rank 7 ice-cream flavors from least to most liked, you can use Kendall's W to see how closely the customers agree in their ratings.

2.10 Time Series

This submenu provides exponential smoothing, autocorrelated regression, ARIMA, X11 ARIMA, seasonal decomposition, spectral analysis, and related techniques.

2.11 Survival

This submenu provides techniques for analyzing the time for some terminal event to occur, including Kaplan-Meier analysis and Cox regression.

2.12 Multiple Response

This submenu provides facilities to define and analyze multiple-response or multiple-dichotomy sets.

Weight Estimation estimates a linear regression model with differential weights representing the precision of observations. This command is in the Professional Statistics option.

If the variance of the dependent variable is not constant for all of the values of the independent variable, weights which are inversely proportional to the variance of the dependent variable can be incorporated into the analysis. This results in a better solution.

The Weight Estimation procedure can also be used to estimate the weights when the variance of the dependent variable is related to the values of an independent variable. If you know the weights for each case you can use the linear regression procedure to obtain a weighted least squares solution. The linear regression procedure provides a large number of diagnostic statistics which help you evaluate how well the model fits your data.

2-Stage Least Squares performs two-stage least squares regression for models in which the error term is related to the predictors. This command is in the Professional Statistics option.

For example, if you want to model the demand for a product as a function of price, advertising expenses, cost of the materials, and some economic indicators, you may find that the error term of the model is correlated with one or more of the independent variables. Two-stage least squares allows you to estimate such a model.

Correspondence Analysis analyzes correspondence tables (such as crosstabulations) to best measure the distances between categories or between variables. This command is in the Categories option.

Homogeneity Analysis is an optimal scaling procedure analogous in some ways to factor analysis, but capable of analyzing categorical or ordinal variables. The technique is also known as multiple correspondence analysis. This command is in the Categories option.

Nonlinear Components performs nonlinear principal-components analysis to try to reduce the dimensionality of a set of variables. This command is in the Categories option.

Overals performs nonlinear canonical correlation analysis to determine how similar sets of variables are to one another. This command is in the Categories option.

3. Other Options

3.1 Transform

Compute calculates the values for either a new or an existing variable, for all cases or for cases satisfying a logical criterion.

Random Number Seed sets the seed used by the pseudo-random number generator to a specific value, so that you can reproduce a sequence of pseudo-random numbers.

Count creates a variable that counts the occurrences of the same value(s) in a list of variables for each case.

Recode into Same Variables reassigns the values of existing variables or collapses ranges of existing values into new values.

Recode into Different Variables reassigns the values of existing variables to new variables or collapses ranges of existing values into new variables.

Rank Cases creates new variables containing ranks, normal scores, or similar ranking scores for numeric variables.

Automatic Recode reassigns the values of existing variables to consecutive integers in new variables.

Create Time Series creates a time-series variable as a function of an existing series, for example, lagged or leading values, differences, cumulative sums. This command is in the Trends option.

Replace Missing Values substitutes non-missing values for missing values, using the series mean or one of several time-series functions. This command is in the Trends option.

Run Pending Transforms executes transformation commands that are pending due to the Transformation Options setting in the Preferences dialog.

3.2 Utilities

Command Index take you to the dialog box for a command if you know its name in the SPSS command language.

Fonts lets you choose a font, style, and size for SPSS Data Editor, output, and syntax windows.

Variable Information displays the Variables window, which shows information about the variables in your working data file, and allows you to scroll the data editor to a specific variable, or copy variable names to the designated syntax window.

File Information displays information about the working data file in the output window.

Output Page Titles lets you specify a title and subtitle for output from SPSS. They appear in the page header, if it is displayed. (Preferences in the Edit menu controls the page header.)

Define Sets defines sets of variables for use in other dialog boxes.

Use Sets lets you select which defined sets of variables should appear in the source-variable lists of other dialog boxes.

Grid Lines turns grid lines on and off in the Data Editor window. This command is available when the Data Editor is active.

Value Labels turns on and off the display of Value Labels (instead of actual values) in the Data Editor window. When Value Labels are displayed you can edit data with a pop-up menu of labels. This command is available when the Data Editor is active.

Auto New Case turns on and off the automatic creation of new cases by cursor movement below the last case in the Data Editor window. This command is available when the Data Editor is active.

Designate Window designates the active window to receive output from SPSS commands (if it is an output window); or to receive commands pasted from dialog boxes (if it is a syntax window). You can also designate a window by clicking the ! button on its icon bar. This command is available when an output or syntax window is active.

3.3 Graphs

Bar generates a simple, clustered, or stacked bar chart of the data.

Line generates a simple or multiple line chart of the data.

Area generate a simple or stacked area chart of the data.

Pie generates a simple pie chart or a composite bar chart from the data.

High-Low plots pairs or triples of values, for example high, low, and closing prices.

Pareto creates Pareto charts, bar charts with a line superimposed showing the cumulative sum.

Control produces the most commonly-used process-control charts.

Boxplot generates boxplots showing the median, interquartile range, outliers, and extreme cases of individual variables.

Scatter generates a simple or overlay scatterplot, a scatterplot matrix, or a 3-D scatterplot from the data.

Histogram generates a histogram showing the distribution of an individual variable.

Normal P-P plots the cumulative proportions of a variable's distribution against the cumulative proportions of the normal distribution.

Normal Q-Q plots the quantiles of a variable's distribution against the quantiles of the normal distribution.

Sequence produces a plot of one or more variables by order in the file, suitable for examining time-series data.

Time Series: Autocorrelations calculates and plots the autocorrelation function (ACF) and partial autocorrelation function of one or more series to any specified number of lags, displaying the Box-Ljung statistic at each lag to test the overall hypothesis that the ACF is zero at all lags.

Time Series: Cross-correlations calculates and plots the cross-correlation function of two or more series for positive, negative, and zero lags.

Time Series: Spectral calculates and plots univariate or bivariate periodograms and spectral density functions, which express variation in a time series (or covariation in two time series) as the sum of a series of sinusoidal components. It can optionally save various components of the frequency analysis as new series.

EXERCISES

Exercise 1: A Completely Randomised Design was conducted with three treatments A, B, C where treatment A is replicated 6 times and B and C are replicated 4 times. Analyse the data.

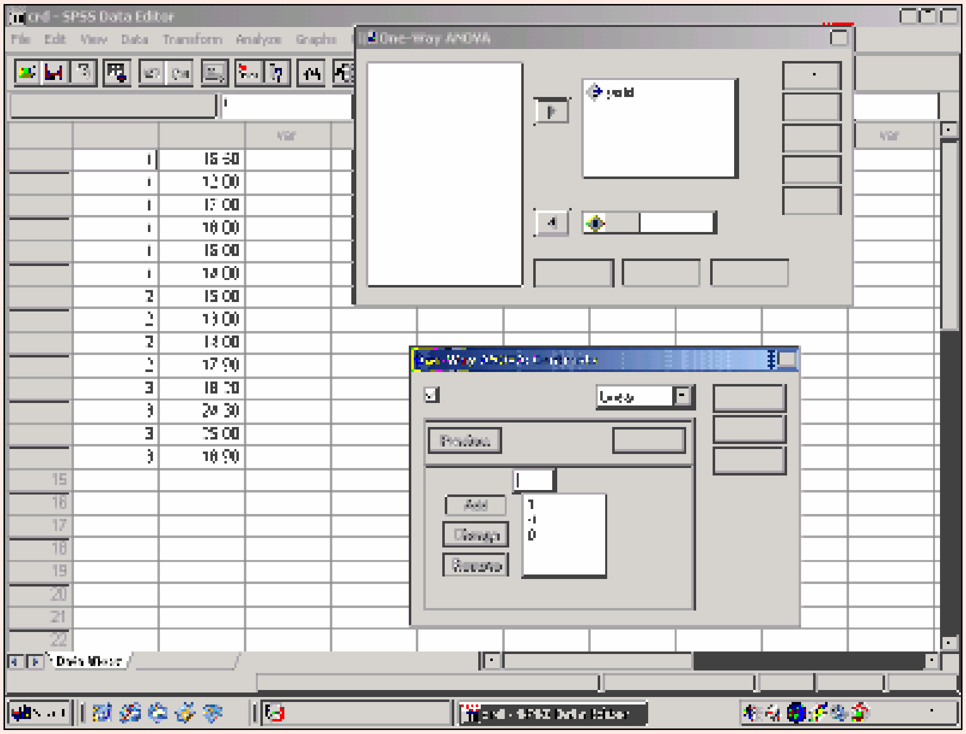
A	B	C
16.50	15.00	18.20
17.00	13.80	24.30
16.00	14.00	25.00
12.00	17.90	18.90
18.00		
14.00		

The above data should be entered as given below in the **Data Editor**:

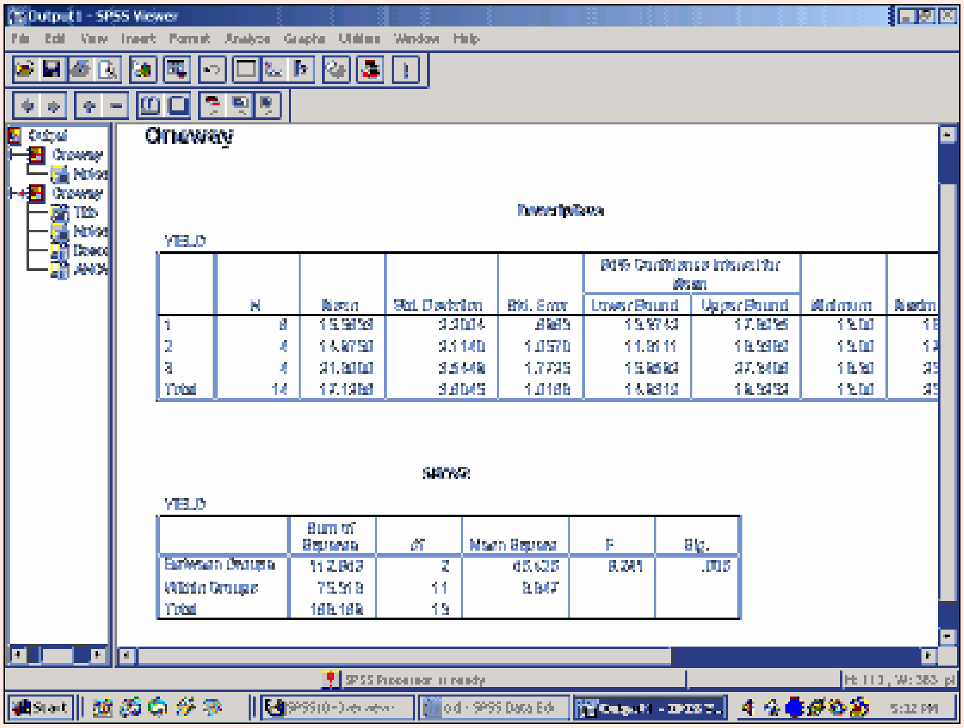
Treat	Yield
1	16.50
1	12.00
1	17.00
1	18.00
1	16.00
1	14.00
2	15.00
2	13.80
2	14.00
2	17.90
3	18.20
3	24.30
3	25.00
3	18.90

SPSS Commands

Analyze→ Compare means → One-way ANOVA → Yield → button [puts yield under **Dependent List**:] → **Treat** [puts treat under **Factor**:] → **Continue → Contrasts... → Coefficients:** -1 → **Add → Coefficients:** 1 → **Add → Continue** [This compares treatment 1 with treatment 2] → **OK**.



Output



Exercise 2: Analyse the data of a 2^3 Factorial Experiment conducted using a randomized complete block design with three replications. The three factors were the fertilizers viz. Nitrogen (N), Phosphorus (P) and Potassium (K). The purpose of the experiment is to determine the effect of different kinds of fertilizers on potato crop yield. The yields under 8 treatment combinations for each of the three randomized blocks are given below:

Block-I

npk	(1)	k	np	p	n	nk	pk
450	101	265	373	312	106	291	391

Block-II

P	nk	k	np	(1)	npk	pk	n
324	306	272	338	106	449	407	89

Block-III

P	npk	nk	(1)	n	k	pk	np
323	471	334	87	128	279	423	324

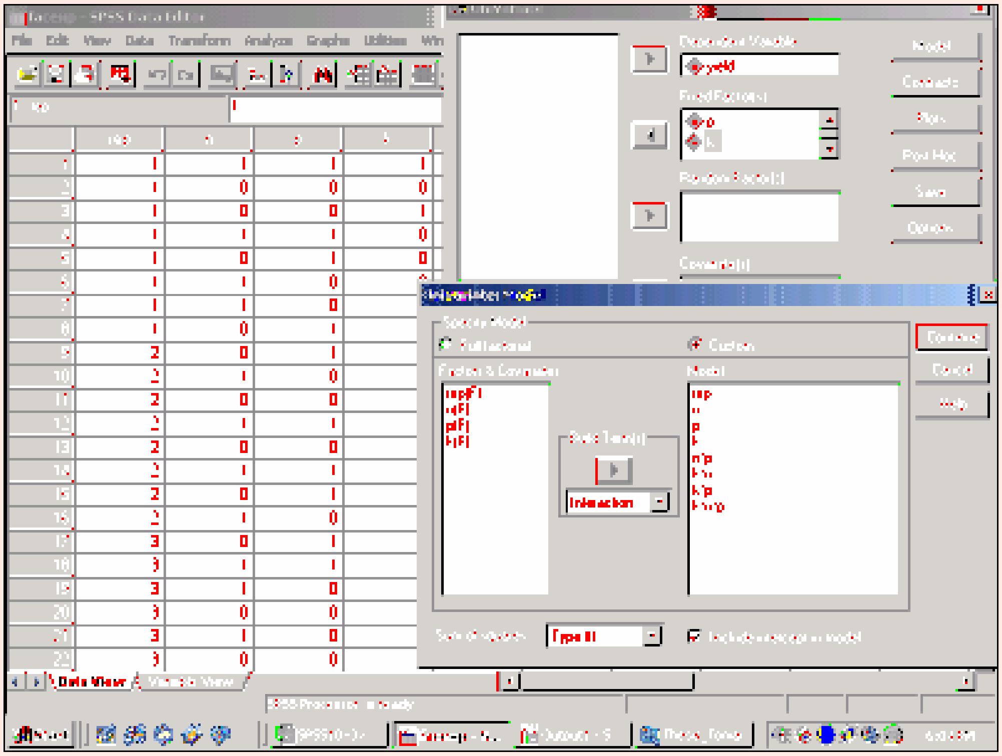
The data for the above layout should be entered in the following manner:

Rep	**N**	**P**	**K**	**Yield**
1	0	0	0	101
1	1	0	0	106
1	0	1	0	312
1	1	1	0	373
1	0	0	1	265
1	1	0	1	291
1	0	1	1	391
1	1	1	1	450
2	0	0	0	106
2	1	0	0	89
2	0	1	0	324
2	1	1	0	338
2	0	0	1	272
2	1	0	1	306
2	0	1	1	407
2	1	1	1	449
3	0	0	0	87
3	1	0	0	128
3	0	1	0	323
3	1	1	0	324
3	0	0	1	279
3	1	0	1	334
3	0	1	1	423
3	1	1	1	471

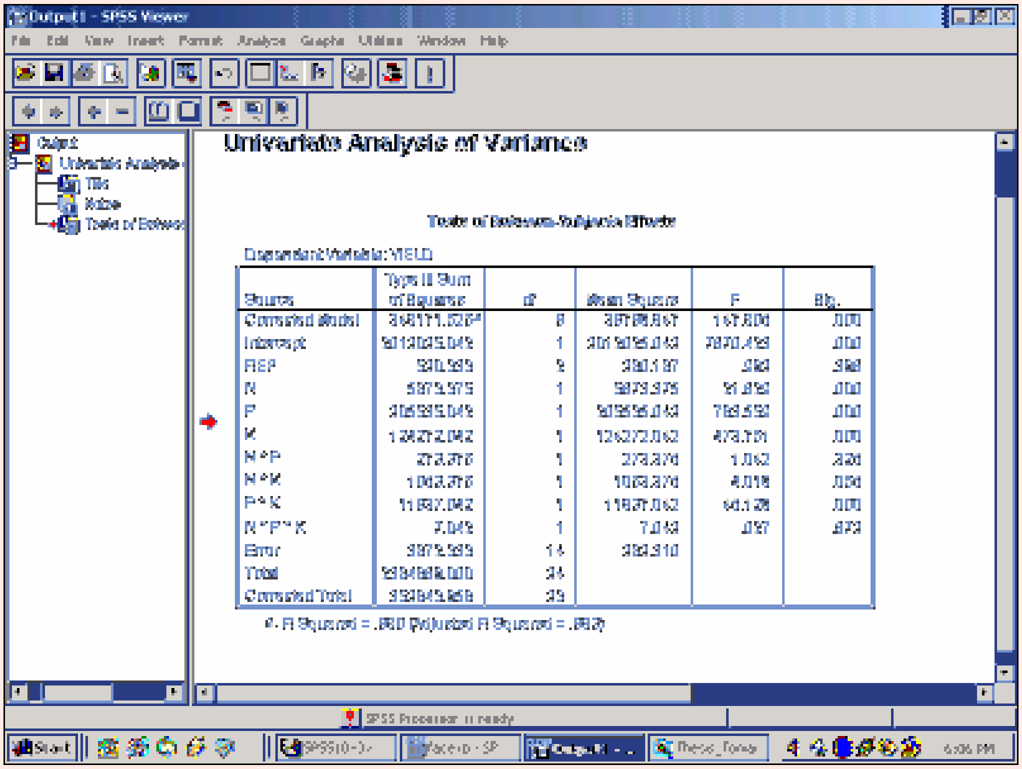
SPSS Commands

Analyze → GLM → Yield → button [puts yield under **Dependent list**:] **→ N →** [puts N under **Factor**:] **→ P → K → Rep → Continue → Model...** [opens Model dialogue box] **→ Custom → Rep. →** [puts Rep under **Model**:] **→ N → P → K →Interaction →**

N → **P** → [puts N*P under **Model**:] → [All the interactions can be entered this way] → **Continue** → **OK**.



Output



Exercise 3: Analyse the following 2^3 Factorial-experiment in blocks of 4 plots, involving three fertilizers N, P, K, each at two levels.

Replication I		Replication II		Replication III	
Block 1	**Block 2**	**Block 3**	**Block 4**	**Block 5**	**Block 6**
np	p	(1)	np	pk	n
101	88	125	115	75	53
npk	n	npk	k	nk	npk
111	90	95	95	100	76
(1)	pk	nk	pk	(1)	p
75	115	80	90	55	65
k	nk	p	n	np	k
55	75	100	80	92	82

The data for the above should be entered in the following manner in **Data Editor**:

Rep	Block	N	P	K	Yield
1	1	1	1	0	101
1	1	1	1	1	111
1	1	0	0	0	75
1	1	0	0	1	55
1	2	0	1	0	88
1	2	1	0	0	90
1	2	0	1	1	115
1	2	1	0	1	75
2	1	0	0	0	125
2	1	1	1	1	95
2	1	1	0	1	80
2	1	0	1	0	100
2	2	1	1	0	115
2	2	0	0	1	95
2	2	0	1	1	90
2	2	1	0	0	80
3	1	0	1	1	75
3	1	1	0	1	100
3	1	0	0	0	55
3	1	1	1	0	92
3	2	1	0	0	53
3	2	1	1	1	76
3	2	0	1	0	65
3	2	0	0	1	82

SPSS Commands

Analyze → GLM → Yield → button [puts yield under **Dependent list**:] **→ N →** [puts N under **Factor**:] **→ P → K → Rep → Block → Continue → Model...** [opens Model dialogue box] **→ Custom → Rep. →** [puts Rep under **Model**:] **→Block → N → P →**

K → → **Interaction** → **N** → **P** → [puts N*P under **Model**:] → [All the interactions can be entered this way] → **Continue** → **OK**.

Exercise 4: An experiment on cotton was conducted to study the effect of folliar application of urea in combinations with insecticidal sprays on the cotton yield. Six treatments were tried in a 6x6 Latin Square Design. The layout plan and yield is given below:

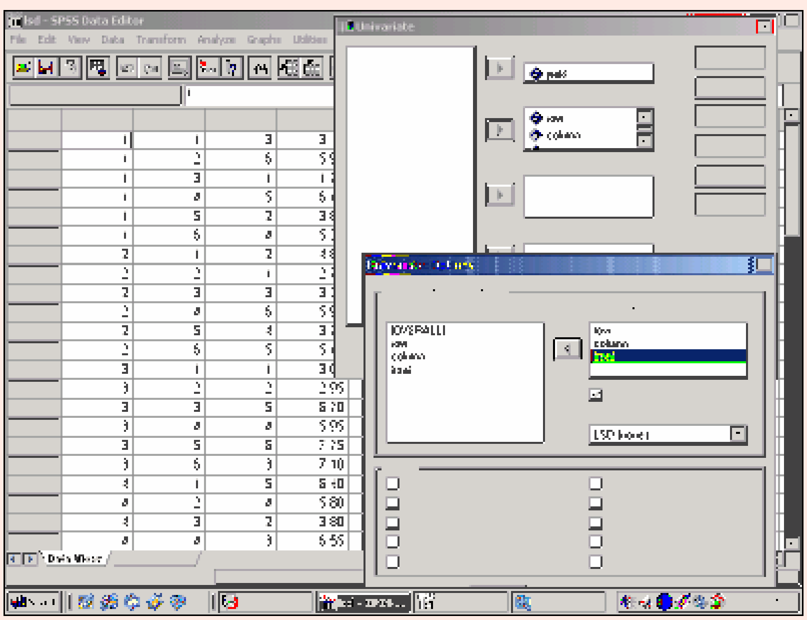
T_3 3.10	T_6 5.95	T_1 1.75	T_5 6.40	T_2 3.85	T_4 5.30
T_2 4.80	T_1 2.70	T_3 3.30	T_6 5.95	T_4 3.70	T_5 5.40
T_1 3.00	T_2 2.95	T_5 6.70	T_4 5.45	T_6 7.75	T_3 7.10
T_5 6.40	T_4 5.80	T_2 3.80	T_3 6.55	T_1 4.80	T_6 9.40
T_6 5.20	T_3 4.85	T_4 6.60	T_2 4.60	T_5 7.00	T_1 5.00
T_4 4.25	T_5 6.65	T_6 9.30	T_1 4.95	T_3 9.30	T_2 8.40

Analyse the data.
Enter the data under the following factors:

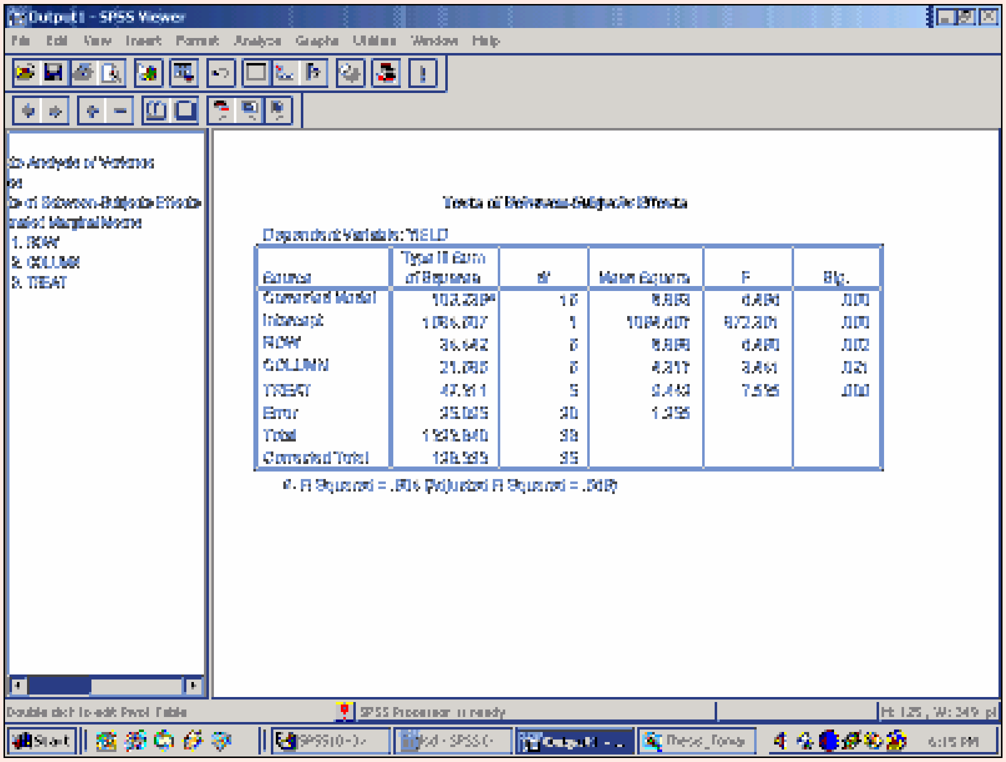
Row	**Column**	**Treat.**	**Yield**
1	1	3	3.10
1	2	6	5.95
1	3	1	1.75

and so on ...

SPSS Commands
Analyze → **GLM** → **Yield** → **button** [puts yield under **Dependent list**:] → **Row** → [puts row under **Factor**:] → **Column** → **Treat.** → **Continue** →**Model...** → **Custom** → **Row** → [puts row under **Model**:] → **Column** → **Treat.** → [puts column and treat. under **Model**:] → **Continue** → **OK**.



Output



Exercise 5: An experiment on rice crop was conducted in split plot design with three replications. Factors tried in the experiment are

1. 4 variety of rice crop: V_1 (IR8), V_2 (IR5), V_3 (C4-C6), V_4 (Peta) - Main plot treatments
2. 6 levels of N: N_0 =0, N_1=60, N_2=90, N_3=120, N_4=150 and N_5=180 kg N/ha- Sub plot treatments

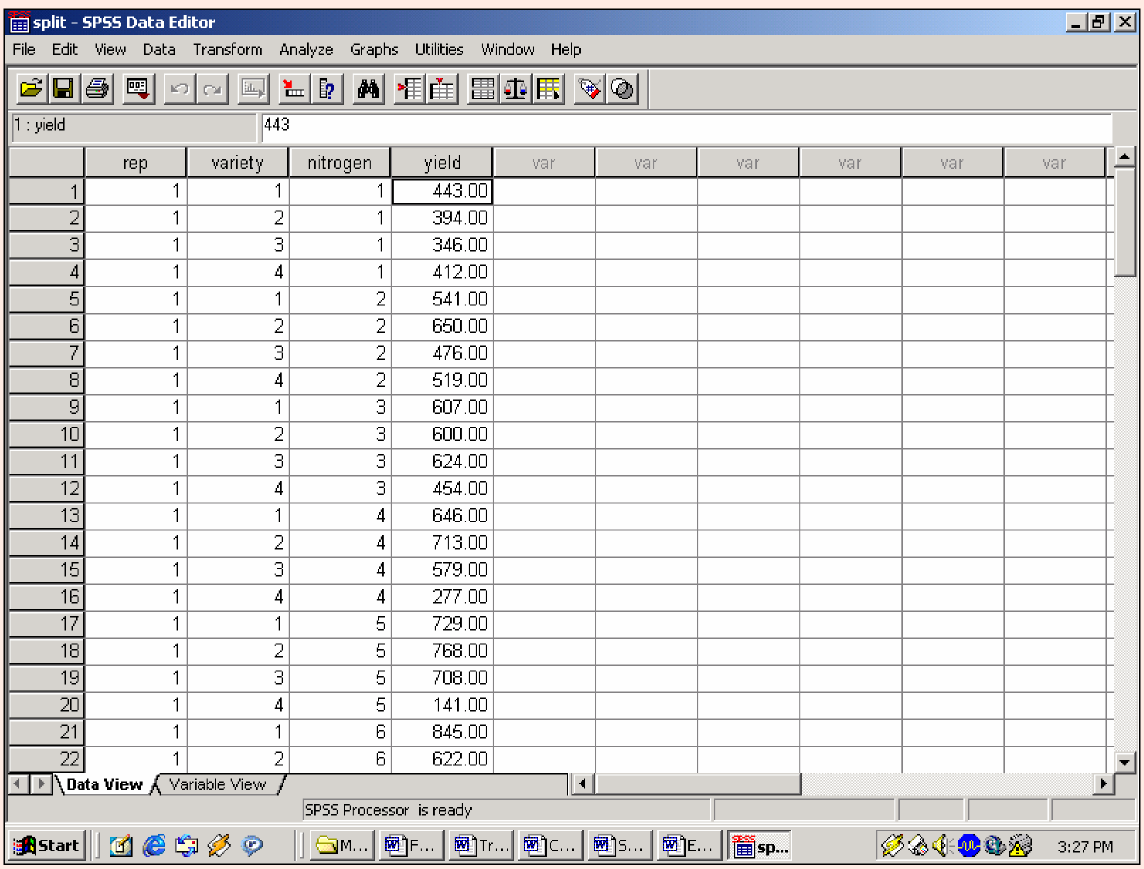
Grain yield data in kg/ha is as given below:

GRAIN YIELD Kg/ha			
VARIETY	**REP-I**	**REP-II**	**REP-III**
	N_0 (0 Kg N/ha)		
V_1	443	447	385
V_2	394	531	366
V_3	346	294	314
V_4	412	448	483
	N_1(60 Kg N/ha)		
V_1	541	516	643
V_2	650	585	558
V_3	476	600	555
V_4	519	460	465
	N_2 (90 Kg N/ha)		
V_1	607	642	670

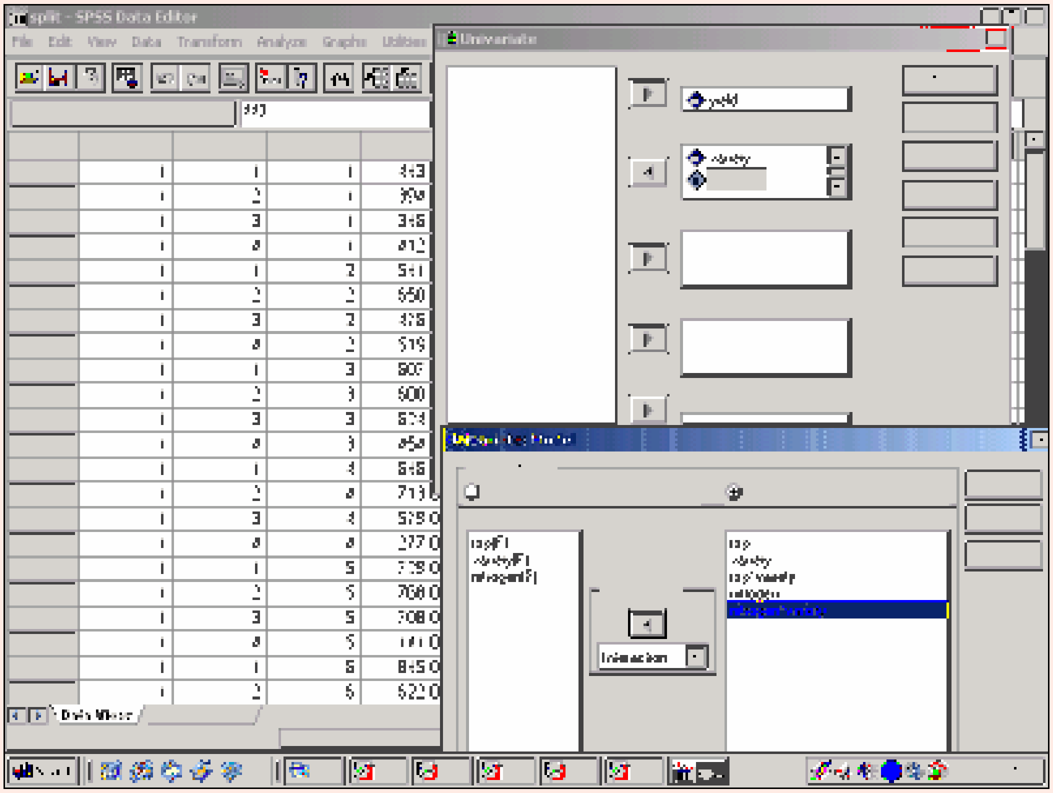
V_2	600	612	664
V_3	624	572	601
V_4	454	574	414
	N_3 (120 Kg N/ha)		
V_1	646	705	668
V_2	713	698	656
V_3	579	588	637
V_4	277	503	363
	N_4 (150 Kg N/ha)		
V_1	729	784	755
V_2	768	659	657
V_3	708	666	632
V_4	141	196	276
	N_5 (180 Kg N/ha)		
V_1	845	883	881
V_2	622	738	600
V_3	559	712	548
V_4	224	138	201

Analyse the data and draw conclusions.

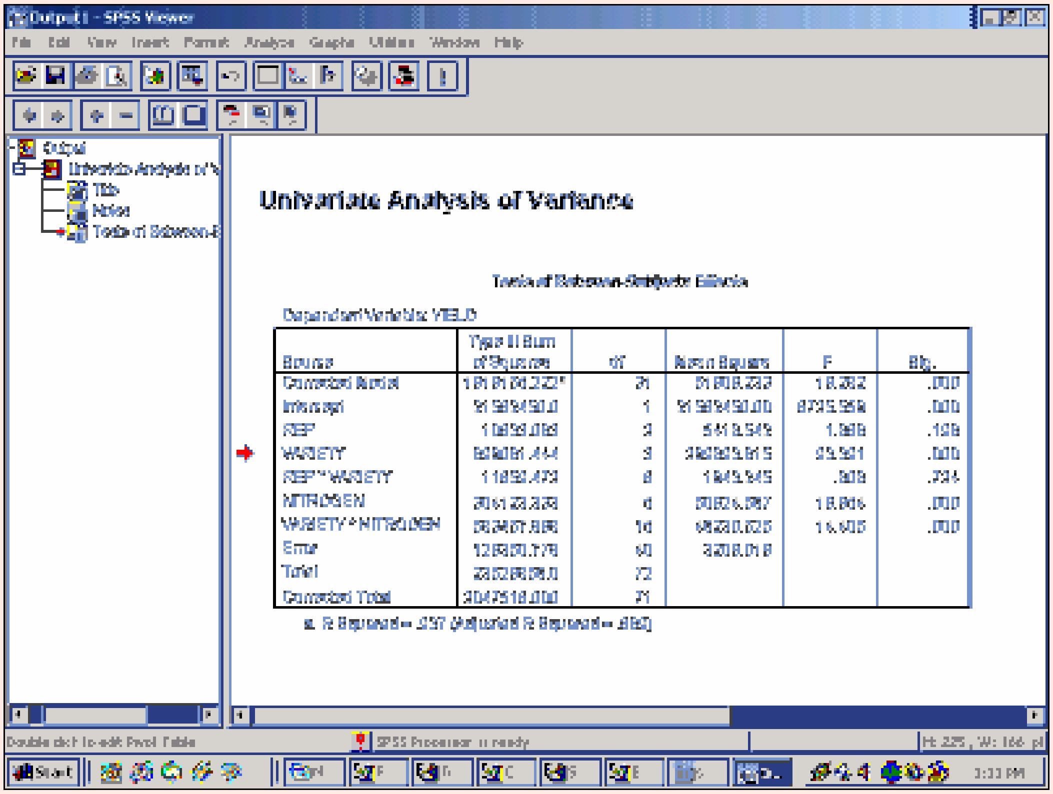
Data Entry in SPSS



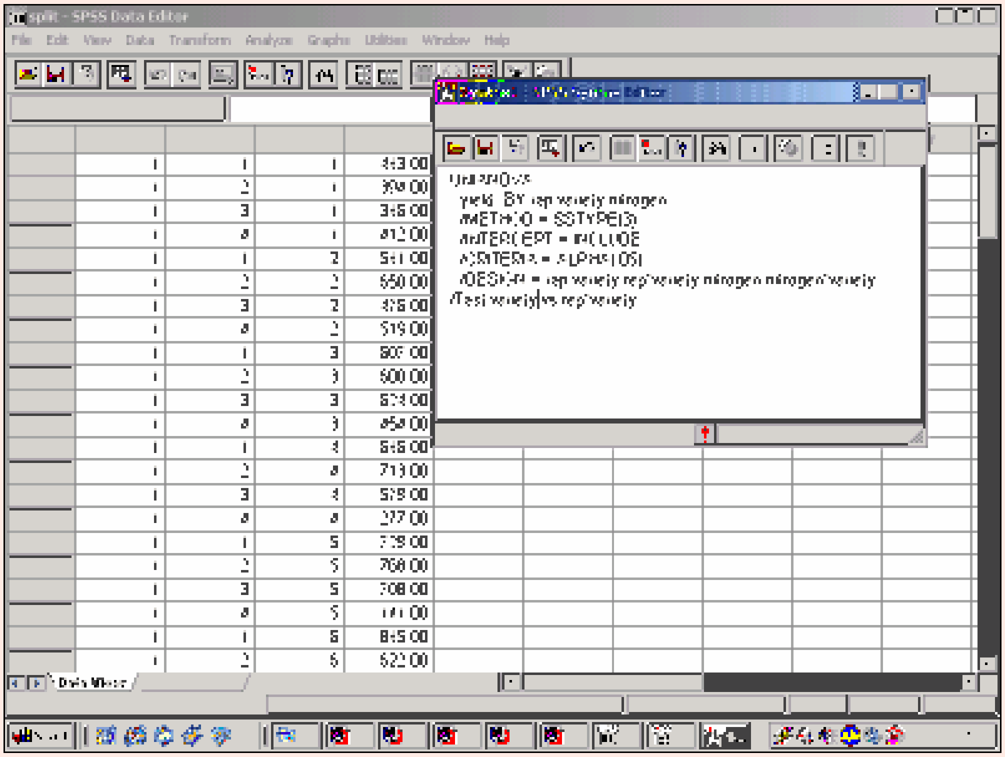
Selection of Variables and Model



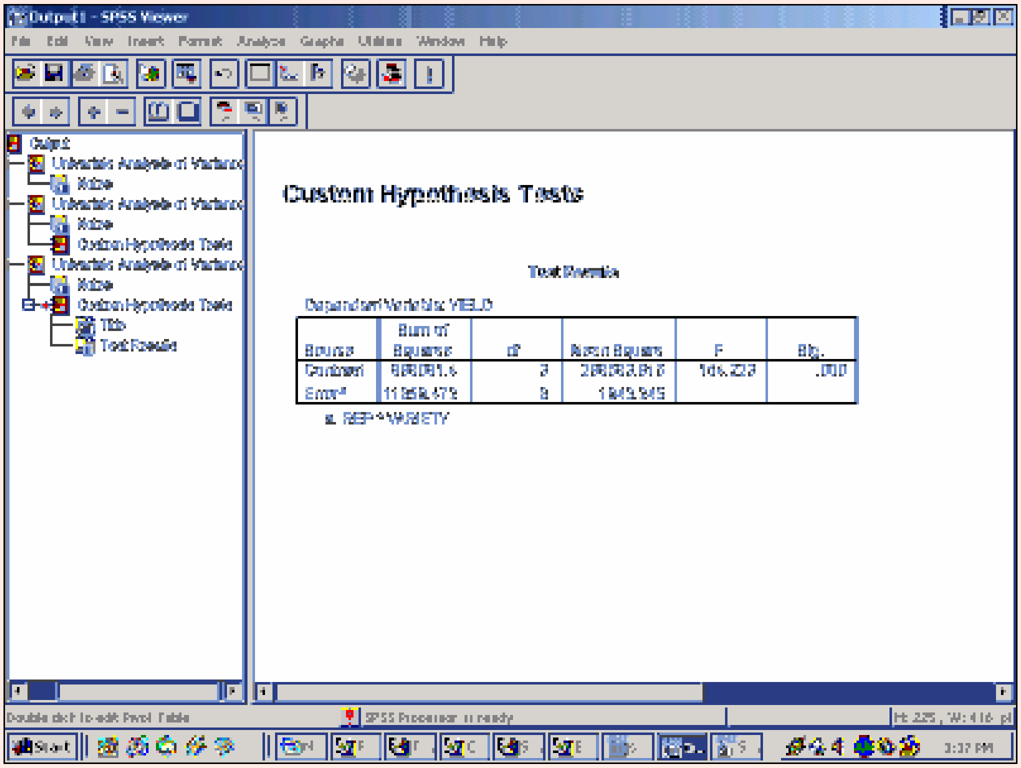
Output



Syntax for testing mainplot with Error(a)



Output of Syntax

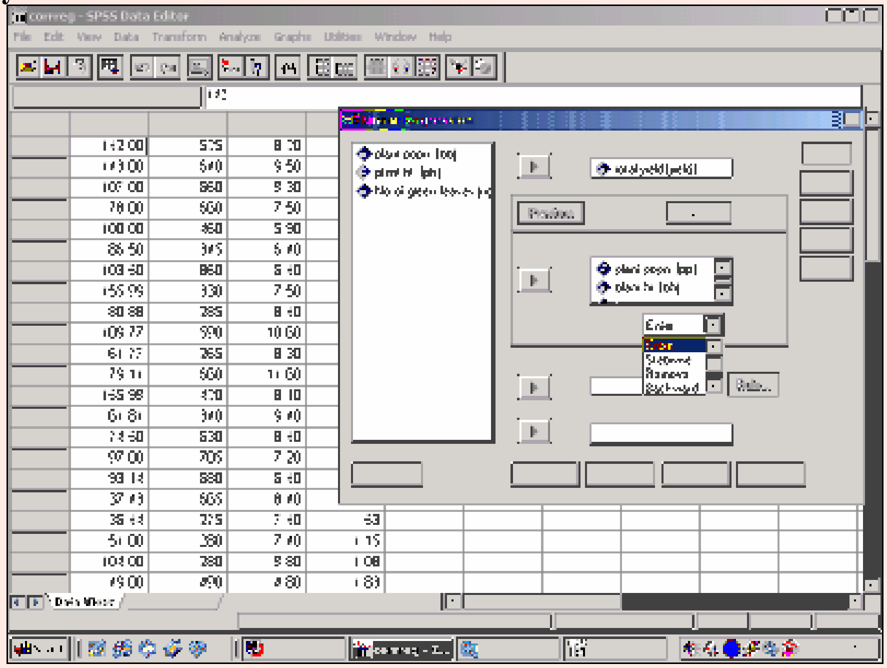


Exercise 6: The following data pertains to Jowar crop on yield and biometrical characters. The biometrical characters are average Plant Population (PP), average Plant Height (PH), average Number of Green Leaves (NGL) and Yield (Kg./plot).

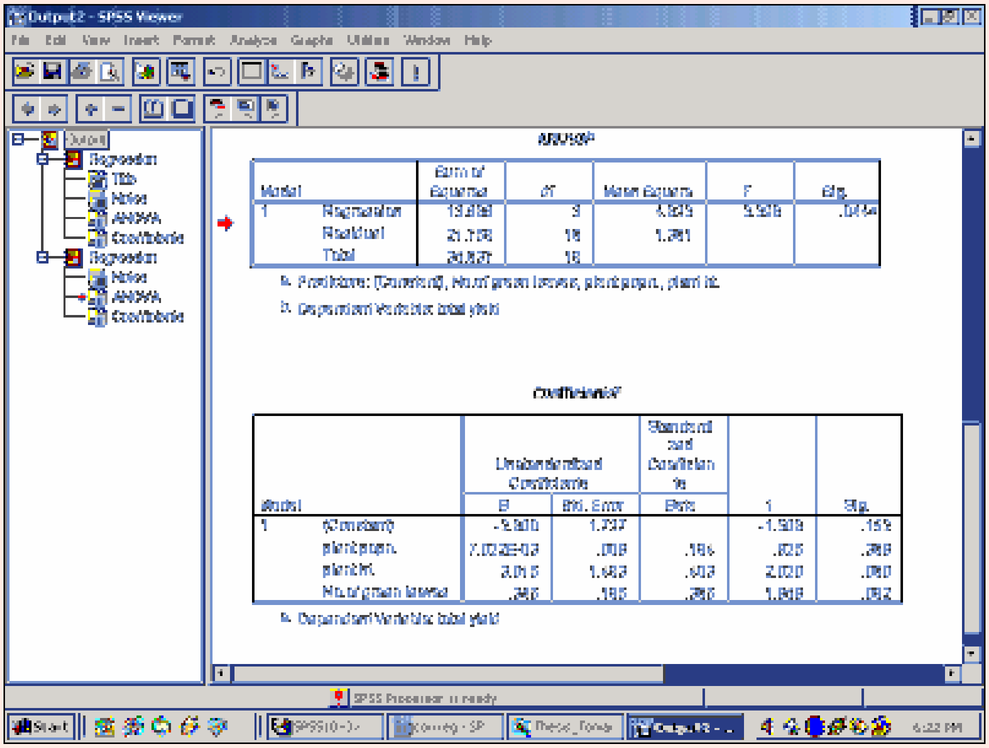
No.	PP	PH	NGL	Yield
1	142.00	0.5250	8.20	2.470
2	143.00	0.6400	9.50	4.760
3	107.00	0.6600	9.30	3.310
4	78.00	0.6600	7.50	1.970
5	100.00	0.4600	5.90	1.340
6	86.50	0.3450	6.40	1.140
7	103.50	0.8600	6.40	1.500
8	155.99	0.3300	7.50	2.030
9	80.88	0.2850	8.40	2.540
10	109.77	0.5900	10.60	4.900
11	61.77	0.2650	8.30	2.910
12	79.11	0.6600	11.60	2.760
13	155.99	0.4200	8.10	0.590
14	61.81	0.3400	9.40	0.840
15	74.50	0.6300	8.40	3.870
16	97.00	0.7050	7.20	4.470
17	93.14	0.6800	6.40	3.310
18	37.43	0.6650	8.40	1.570
19	36.44	0.2750	7.40	0.530
20	51.00	0.2800	7.40	1.150

Fit a multiple linear regression equation by taking yield as dependent variable and biometrical characters as explanatory variables.

Data Entry



Output


SPSS Viewer
File Edit View Insert Format Analyze Graphs Utilities Window Help
Regression
Title
Notes
ANOVA
Coefficients
Model
Sum of Squares
df
Mean Square
F
Sig.
Regression
Residual
Total
Unstandardized Coefficients
Standardized Coefficients
B
Std. Error
Beta
(Constant)
SPSS Processor is ready