# JMP STATISTICAL DISCOVERY SOFTWARE: AN OVERVIEW

**Rajender Parsad, Manoj Kumar Khandelwal and RS Tomar**
**I.A.S.R.I., Library Avenue, Pusa, New Delhi - 110 012**
**rajender@iasri.res.in, khandelwalmanoj85@gmail.com; tomar@iasri.res.in**

## 1. Introduction

JMP Statistical Discovery Software is a comprehensive and interactive statistical package. It dynamically links data with graphics for interactive exploration, understanding, and visualization of the data. This allows one to click on any point in a graph, and see the corresponding data point highlighted in the data table, and other graphs. It can work with a variety of data formats, such as text files, Microsoft Excel files, SAS datasets, and ODBC-compliant databases. It supports Windows, Macintosh and Linux operating systems. It has a flexible working environment: user-friendly menu-based interface for the new user and allows for custom programming and script development via JSL, "**JMP Scripting Language**". JSL is an interpreted scripting language which is executed at runtime, and provides for manipulating JMP application platform objects in a coherent and coordinated way. It is a SAS Product and was founded by John Sall in 1989. It is generally pronounced as "jump". The latest version is JMP 9. More details about JMP can be seen at www.jmp.com. The other JMP products are JMP Clinical and JMP Genomics.

## 2. Getting Started with JMP Statistical Discovery Software

To open JMP Software, Go to **Start → Programs→ JMP → JMP 8** or Just double-click the icon  on your desktop. Initial view of JMP is a menu bar, a toolbar, the Tip of the Day window, and the JMP Starter window. See Figure 2.1

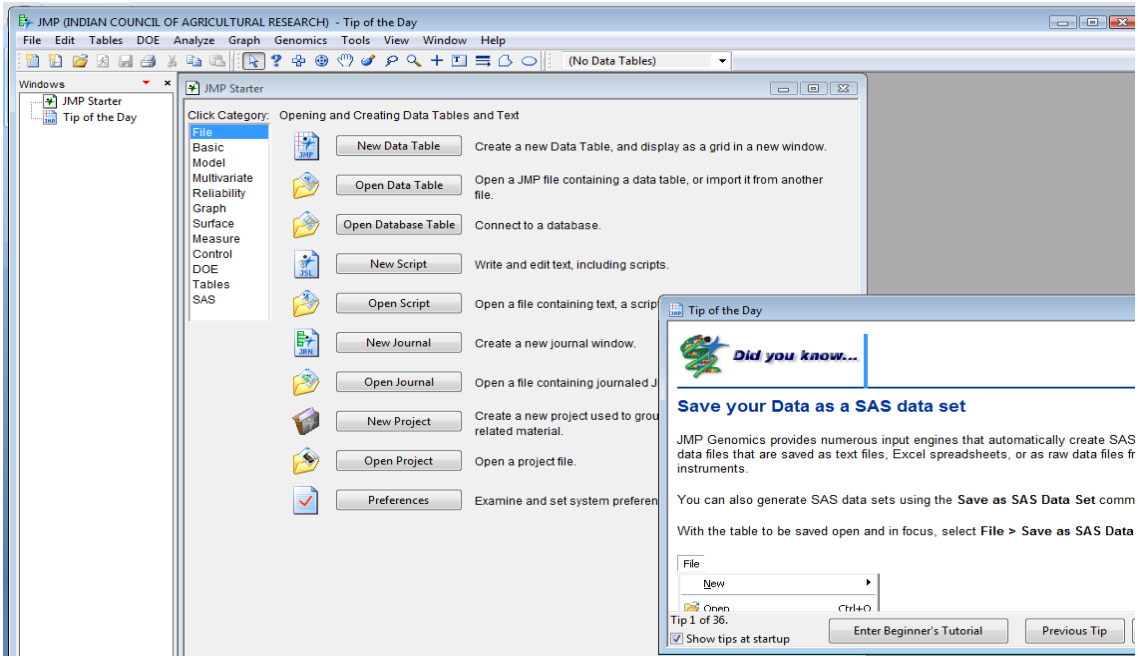
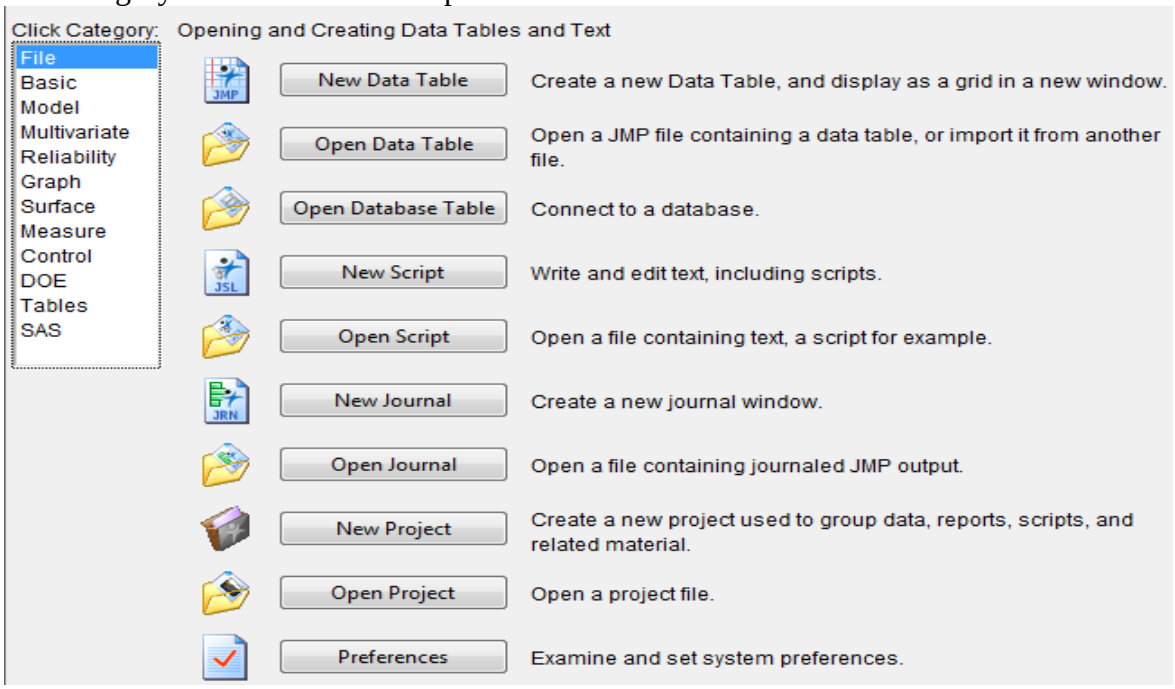



Figure 2.1: First view of JMP Window

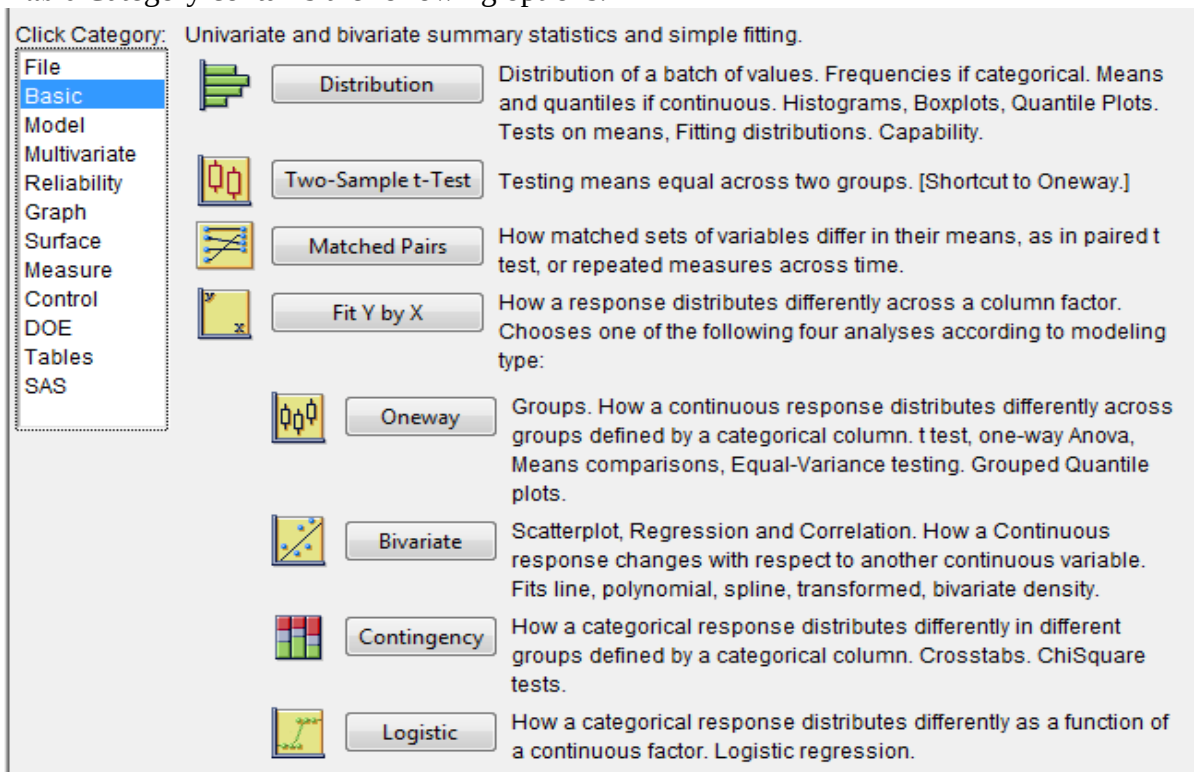
Tip of the Day window provides tips about using JMP that you might not know. It can be closed by selecting close button. To view it again, select **Help → Tip of the Day**.

JMP Starter window is helpful in navigating in JMP. It presents a collection of starting points, grouped by categories that organize platforms and commands by task with descriptions to guide ones selections. One can browse the various options available through **Click Category**. These categories include File, Basic, Model, Multivariate, Reliability, Graph, Surface, Measure, Control, DOE, Tables and SAS. In case, JMP Starter Window gets closed, one can always return to the JMP Starter window by opening it from the View menu (**View → JMP Starter**). A brief description of these click categories is given in the sequel.

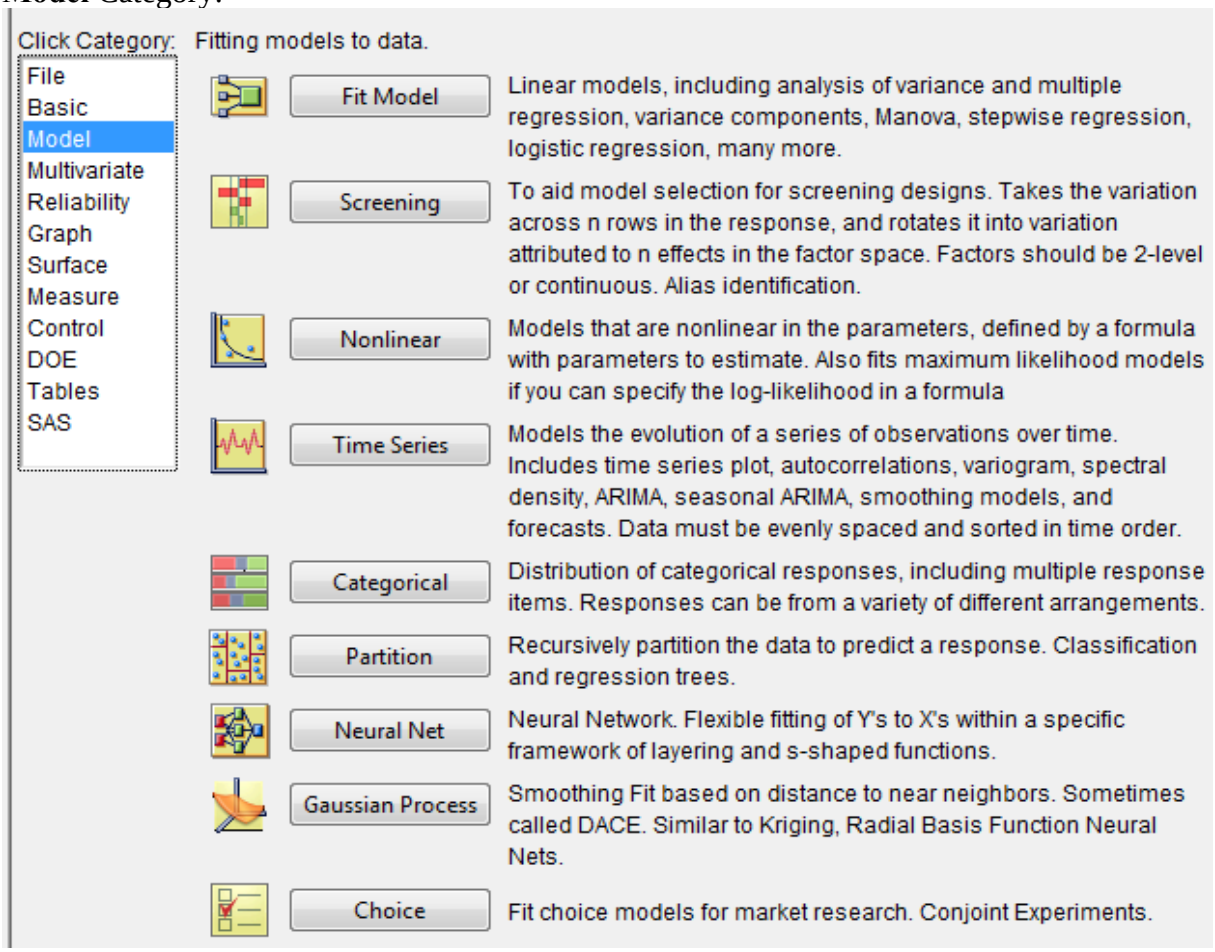
**File** category consists of various options such as



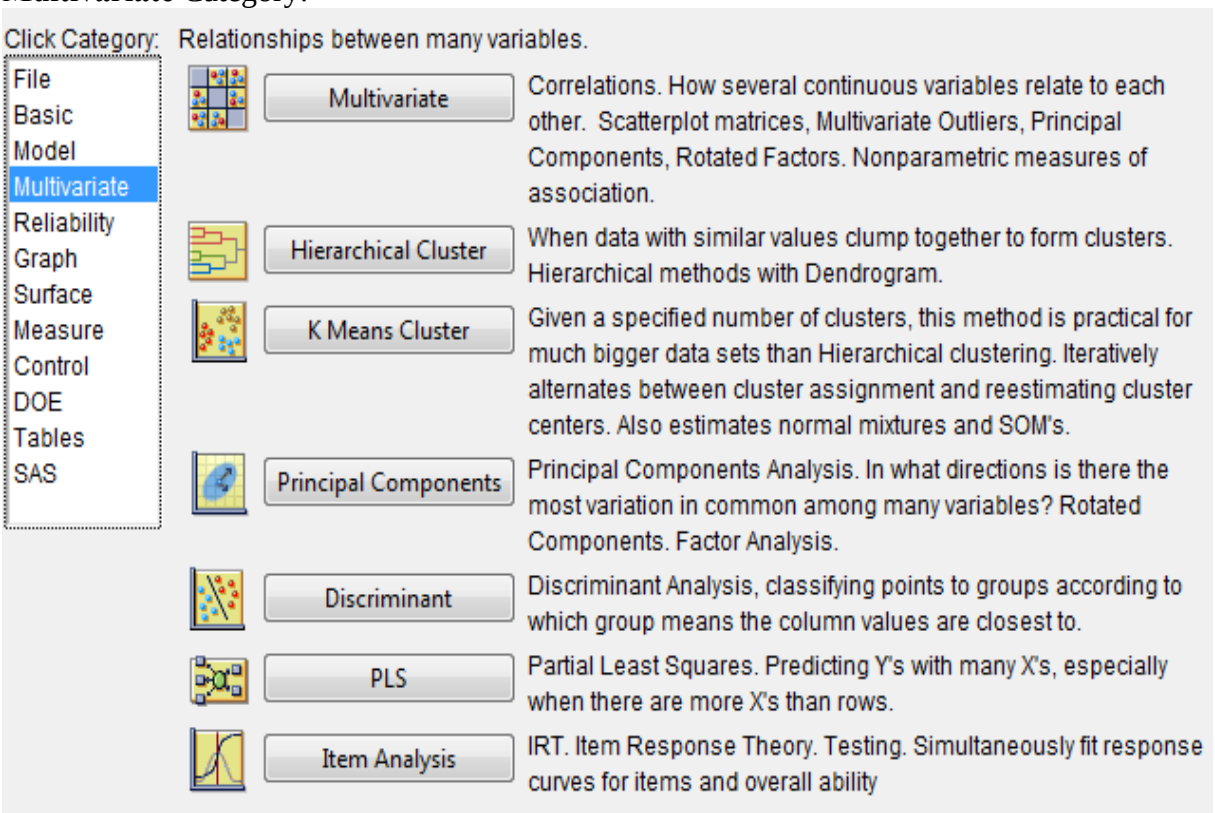

**Basic** Category contains the following options:



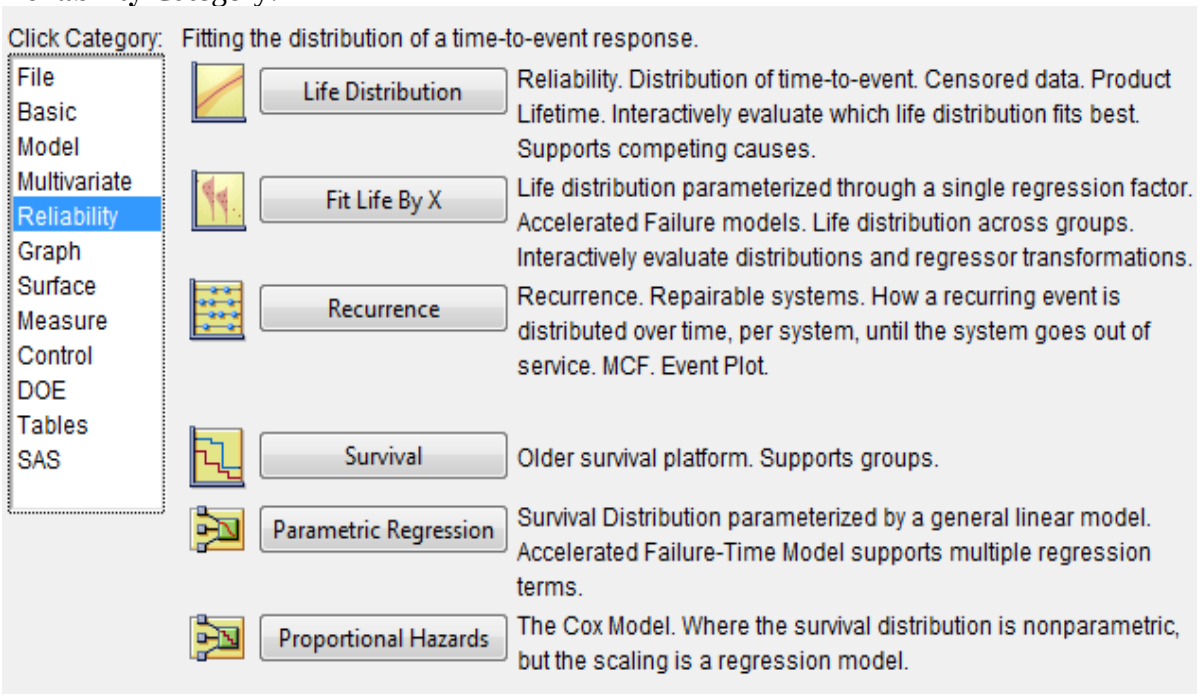
**Model** Category:



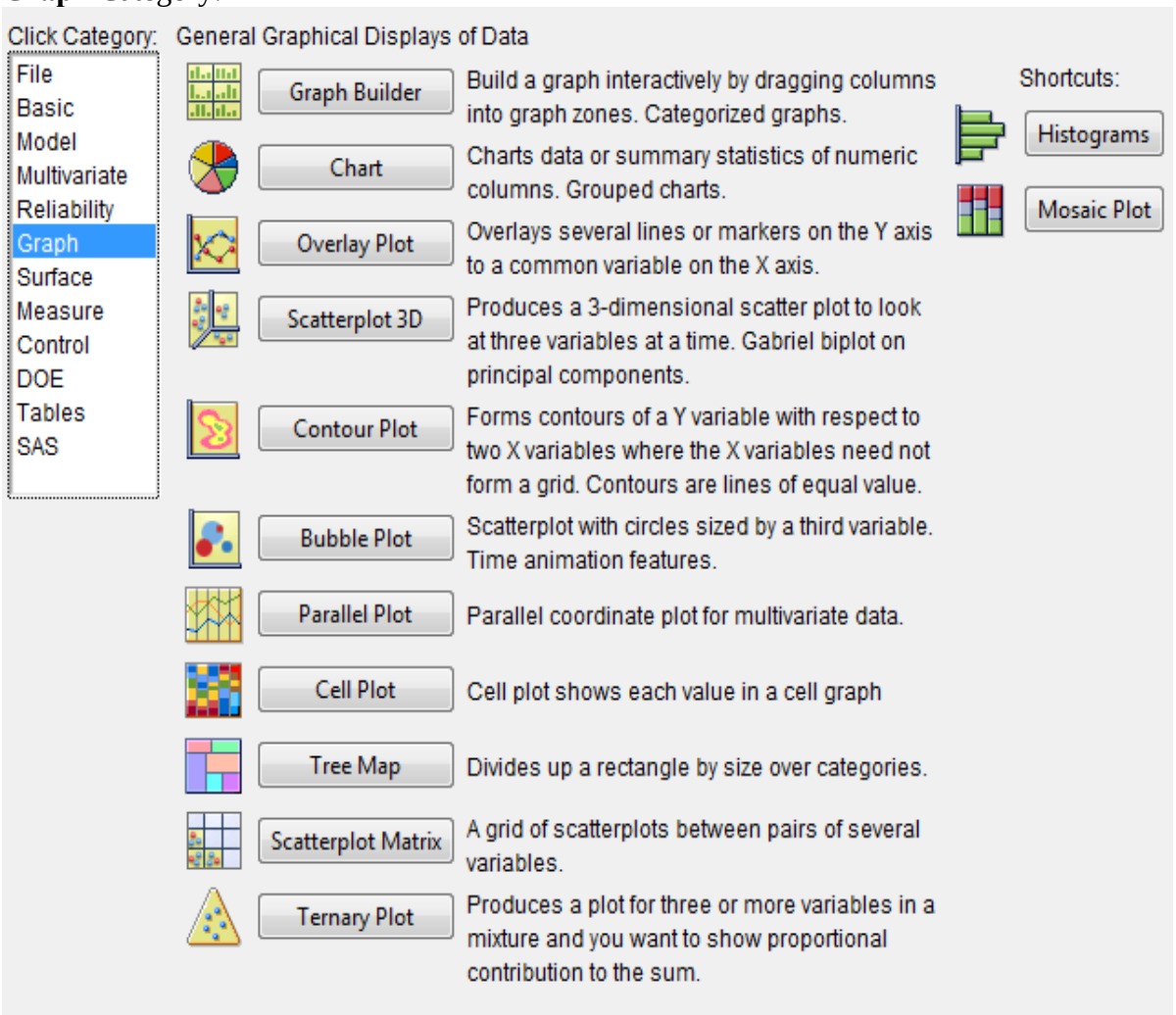

**Multivariate** Category:



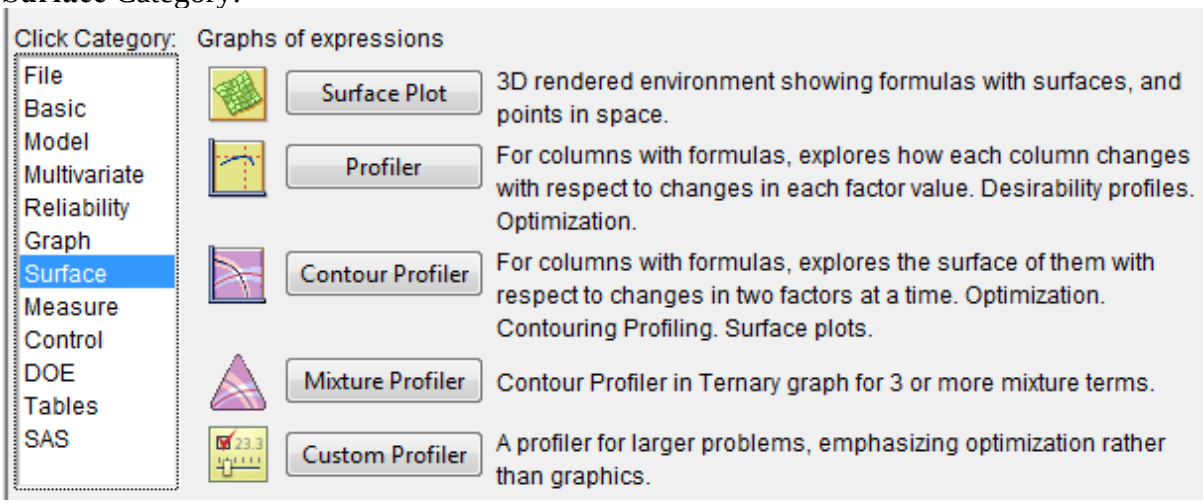
**Reliability** Category:



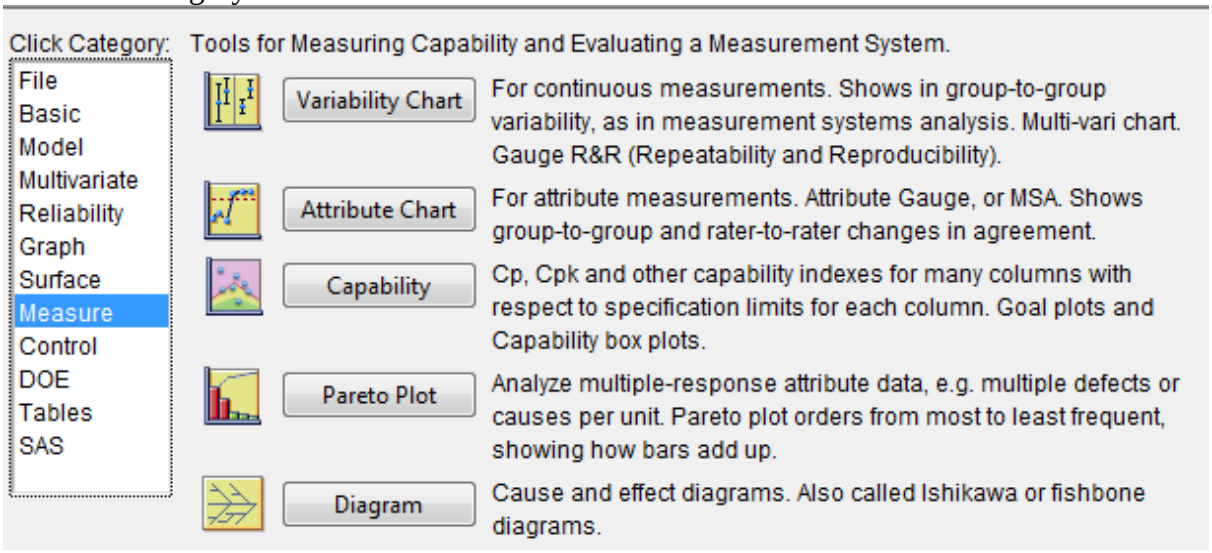

**Graph** Category:



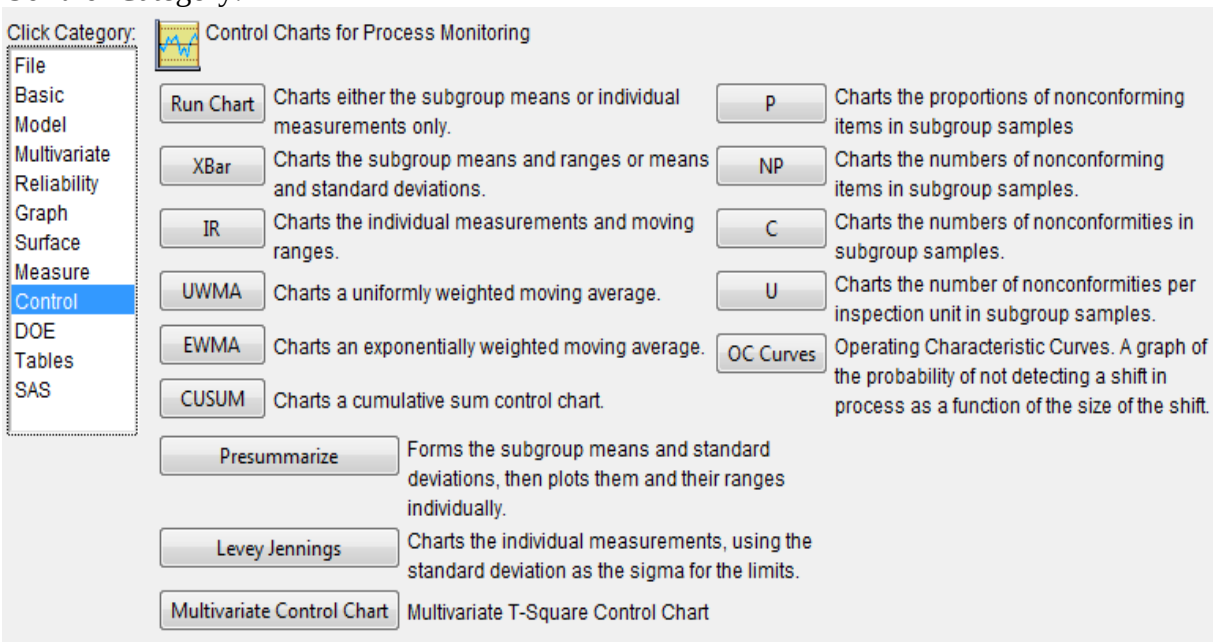
**Surface** Category:



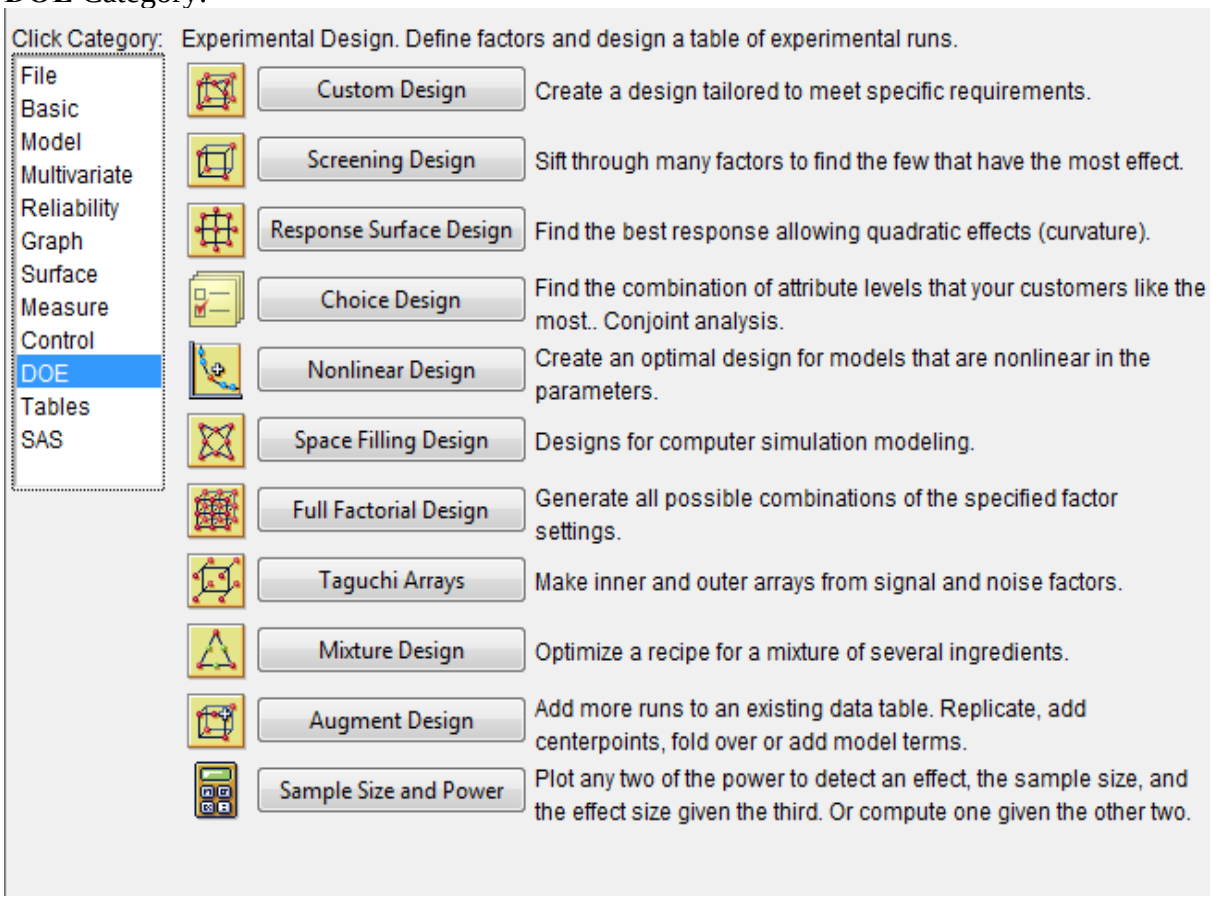

**Measure** Category:



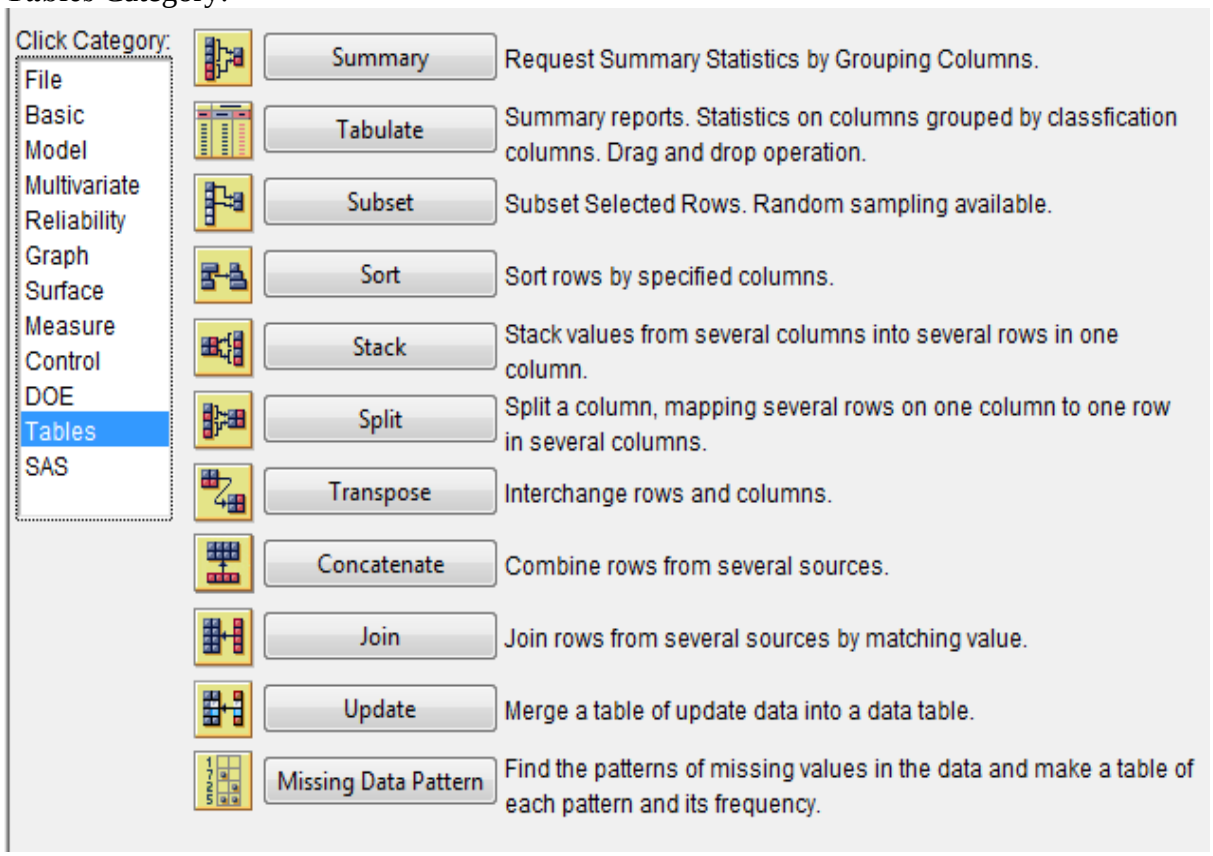

**Control** Category:



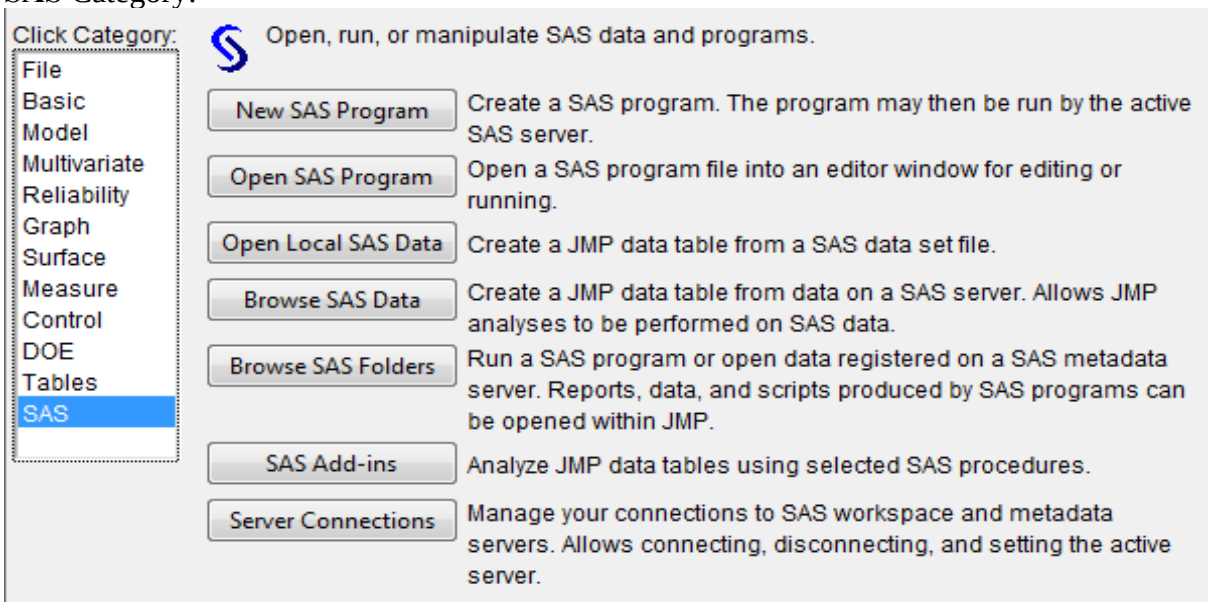
**DOE** Category:




**Tables** Category:



**SAS** Category:

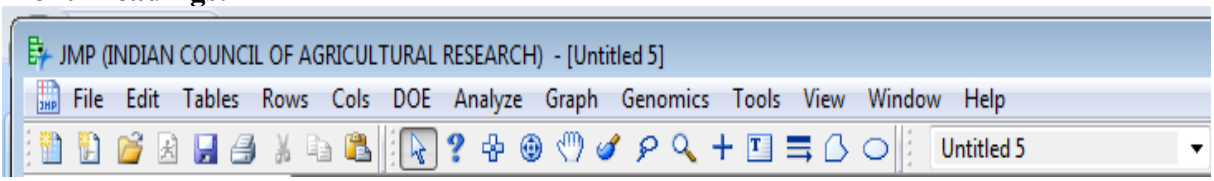



From **JMP Starter window**, one can also select **Preferences.** The choices available are:

- general operation and appearance of JMP.
- background color of windows and graphs.
- type, style, and size of fonts.
- graphic formats for copy and drag results from RTF and HTML files.
- communications settings.
- default directory paths for file locations.
- results initially presented by each analysis or graph platform.
- settings for importing and exporting data to suit our needs or situation.

The JMP Statistical Discovery Software has the following Menu bar.

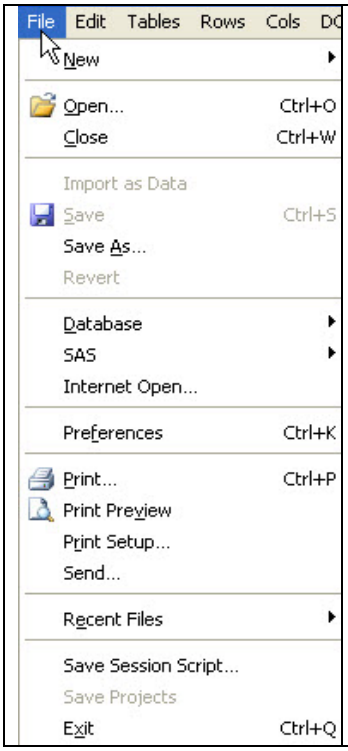
**Menu Headings:**




| | |
|---|---|
| **File** | performs most routine file functions, such as opening, closing and saving. |
| **Edit** | performs most common editing functions such as cutting and pasting. |
| **Tables** | performs table functions, such as sort, subset, join, transpose and merge. |
| **Rows** | performs row operations (JMP treats rows as observations). |
| **Cols** | performs column operations (JMP treats columns as variables). |
| **DOE** | facilitates the Design of the Experiment. |
| **Analyze** | performs most statistical analyses. |
| **Graph** | generates a variety of plots. |
| **Tools** | displays analysis window tools. |
| **View** | appears only under the Windows operating system environment. |
| **Window** | selects among currently opened windows and performs window operations. |
| **Help** | accesses the main help features in JMP. |

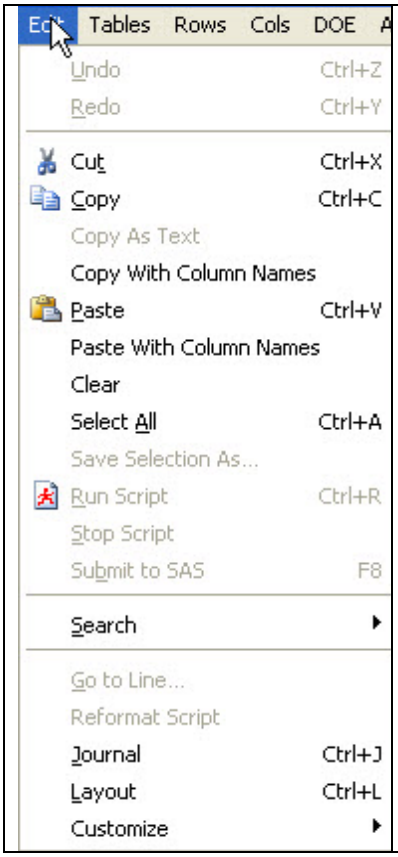
One can see below the detailed description of all the options available in above-mentioned menu headings.

**FILE MENU**



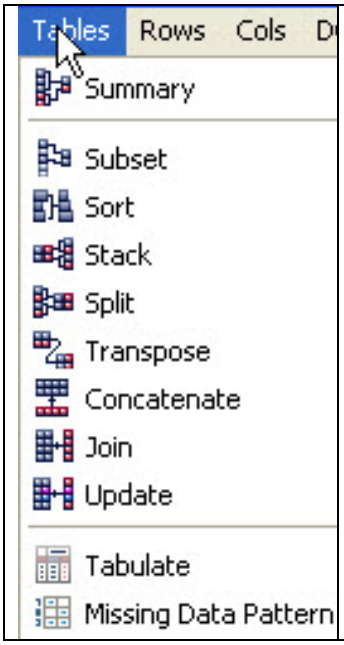

- builds the new data table, new script window, new project and new Journal.
- opens an existing JMP data table.
- closes the active window.
- writes an open text file to a JMP data table.
- writes the active data table, journal, or layout to a file.
- restores the current data table to its condition when it was last saved.
- links to data base at a different location.
- lets one to open an internet browser within JMP.
- selects default preferences.
- printing options.
- previews ready to print window.
- selects desired print format.
- reveals a submenu that lists the JMP tables, scripts, and journals most recently opened.
- saves the script of the executed analysis.
- saves the executed analysis and data table as a project.
- exits JMP software.

**EDIT MENU**



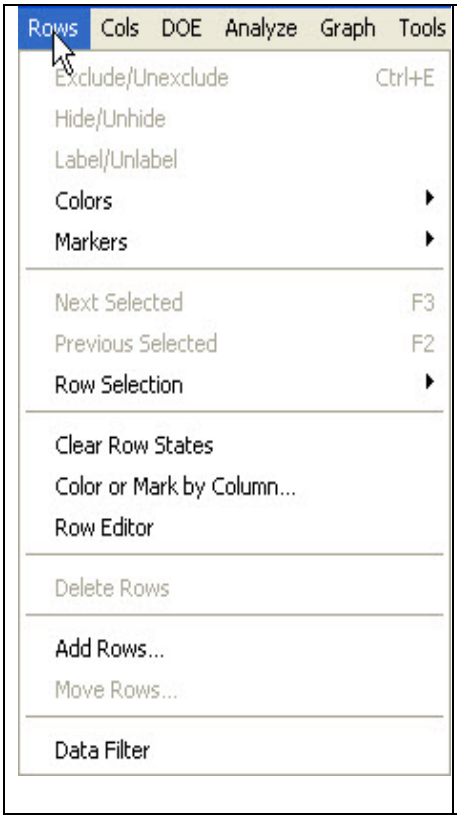

- undo the last action if possible.
- redo the last action if possible.
- cuts selection and keeps it in clipboard.
- copies selection.
- copies selection only in text format.
- preserves the data table's column labels in the copied image
- uses the first line of information on the clipboard as column headers.
- clears the data at the end of the current data table.
- selects all data in data table.
- saves selection in desired format.
- runs script if there is one in the current window.
- submits the JMP scripts as a SAS program to SAS server.
- gives the ability to find and replace text in data tables and scripts.
- finds the line in the data table for observations that meet our criteria.
- saves a report just as it appears in the report window.
- reveals a submenu to customize menus and toolbars. Revert to factory defaults resets the menus and toolbars to the arrangement when we first installed JMP.

## TABLES MENU




- request summary statistics by grouping columns.
- creates a new data table that is a subset of the active data table.
- sorts a JMP data table by one or more columns.
- creates a new data table from the active table by stacking specified columns into a single new column.
- creates a new data table from the active table by dividing one or more columns to form multiple columns.
- creates a new data table by interchanging rows and columns.
- creates a new data table from two or more open tables by combining them end to end.
- creates a new data table by merging (joining) two tables side by side.
- updates one data table with values from a second table.
- displays descriptive statistics in tabular format.
- creates a new data table showing the pattern that the missing data in original data table creates.

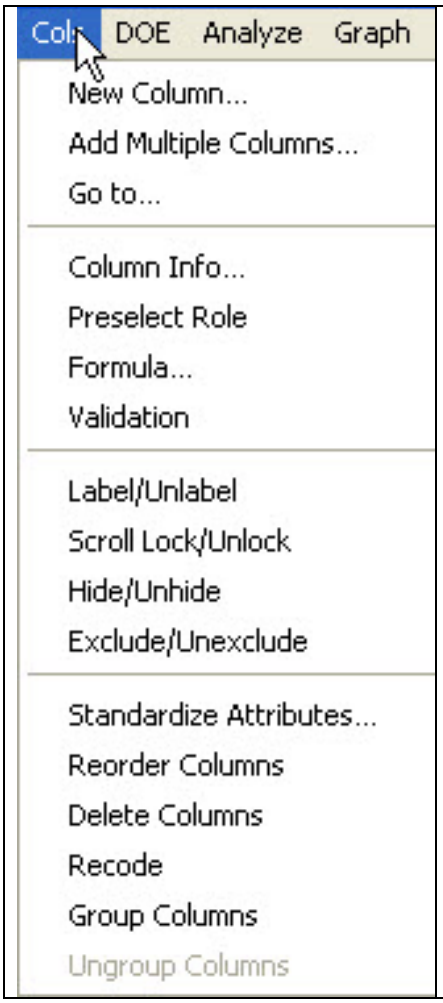
## ROWS MENU




* Recall that JMP treats rows as observations

- excludes selected rows from statistical analyses.
- suppresses (hides) rows so they do not appear in plots and graphs.
- labels or identifies points on all scatter plots.
- changes highlighted points in all scatter plots to the colors one select.
- lists markers.
- searches for observations meeting our criteria.
- returns to last selection.
- utilities on row selection.
- clears all active row states in the data table.
- lets one color or mark points in plots.
- lets one browse or edit cells.
- deletes all selected rows from a JMP data table.
- adds new rows (default=20)
- moves highlighted rows to the location one specify in the move rows window.
- lets select rows, create subsets and animate selected rows.

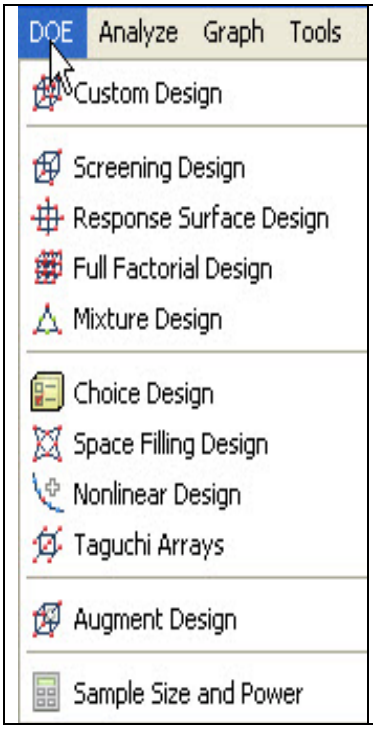
## COLS MENU




*Recall that JMP treats columns as variables

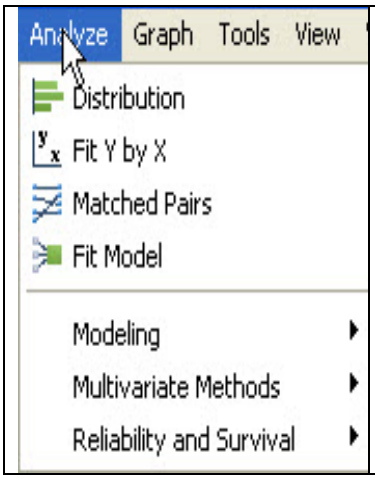
- creates a new column.
- add more than one column at a time to a table.
- highlights a specific column in the table.
- opens the column information for a selected column.
- assigns a role to the selected column and saves the role with the data table.
- displays the column's formula editor to create a formula that computes column values.
- lets one set up a column so that it only accepts certain values.
- is a toggle command that labels or identifies points on all scatter plots.
- locks column into left-most position in the data grid
- is a toggle command that suppresses (hides) columns so they do not appear in plots and graphs.
- is a toggle command used to exclude selected columns from statistical analyses.
- lets one apply attributes and properties to multiple columns.
- lets one move columns according to the selection from its submenu:
- deletes selected columns from the data table.
- lets one quickly recode data that is coded incorrectly.

## DOE MENU




- create a design tailored to meet specific requirements.
- allows one to specify continuous factors, and two- and three-level categorical factors.
- define a number of continuous factors.
- specify a set of continuous and categorical factors with arbitrary numbers of levels.
- define a set of factors that are ingredients in a mixture.
- creates a design by spreading the design points out to the maximum distance possible between two points.
- create an optimal design for models that are nonlinear in the parameters.
- make inner and outer arrays from signal and noise factors.
- optimize a recipe for a mixture of several ingredients.
- add more runs to an existing data table. Replicate, add center points, fold over or add model terms.
- computes power, sample size, or the effect size one want to detect for a given alpha and error standard deviation.

**ANALYZE MENU**




- investigates the distribution of values in each column.
- studies the relationship of two variables.
- models one or more response variables with one or more predictor variables.
- fits one or more y variables to a model of x variables.
- performs nonlinear modeling, time series analysis and neural network analysis.
- explores correlation, scatterplot matrix and performs cluster, principal components, discriminant and item analysis.
- performs reliability, survival and recurrence analysis.

Here is a brief description of the **Analyze** platforms:

**Distribution** describes the distribution of each column one choose for analysis with histograms, box plots and normal plots for continuous columns, and divided (mosaic) bar charts for nominal and ordinal columns. One can also compare the computed mean and standard deviation of the distribution to a constant and examine the fit of a variety of different distributions.

**Fit Y by X** describes each pair of X and Y columns one specify. The displays and reports vary depending upon the modeling types (continuous, nominal, ordinal) one assign to the X and Y columns. The four combinations of response and factor modeling types lead to the four analyses: bivariate analysis, one-way analysis of variance, contingency table analysis, and logistic regression.

**Matched Pairs** analyzes the special situation where multiple measurements are taken on a single subject, and a paired t-test or repeated measures analysis is needed.

**Fit Model** gives the general fitting platform for fitting one or more Y columns by all the effects one create. The general fitting platform fits multiple regression models, models with complex effects, response surface models, and multivariate models including discriminant and canonical analysis. Leverage plots, least-squares means plots, and contour plots help one visualize the whole model and each effect. Special dialogs lets one request contrasts, custom tests, and power details. The Fit Model platform also has special effect screening tools such as the cube plots, a prediction profiler, and a contour profiler, for models with multiple responses.

**Modeling** is for advanced models that are nonlinear in their parameters, or models that have correlated terms. In some cases, the form of the model is not important.

**Screening** provides analysis of screening designs.

**Nonlinear** fits the response (Y) as a function of a nonlinear formula of the specified x variable and parameters. Nonlinear fits models that are nonlinear in their parameters. Nonlinear fitting begins with a formula one can build in a column using the formula editor. This formula has parameters to be estimated by the nonlinear platform. On launching Nonlinear, one can interact with a control panel to do the fitting and plotting of results.

**Partition** fits classification and regression trees.
**Neural Net** implements a simple one-layer neural net.
**Time Series** performs simple forecasting and fits ARIMA models.
**Gaussian Process** fits no-error-term models. These are common in areas like computer simulations, where a given input always results in the same output.
**Categorical** does tabulation and summarization of categorical response data.
**Choice** performs analysis of choice experiments, and for making probabilistic statements about human choice. Useful in marketing research and product design.

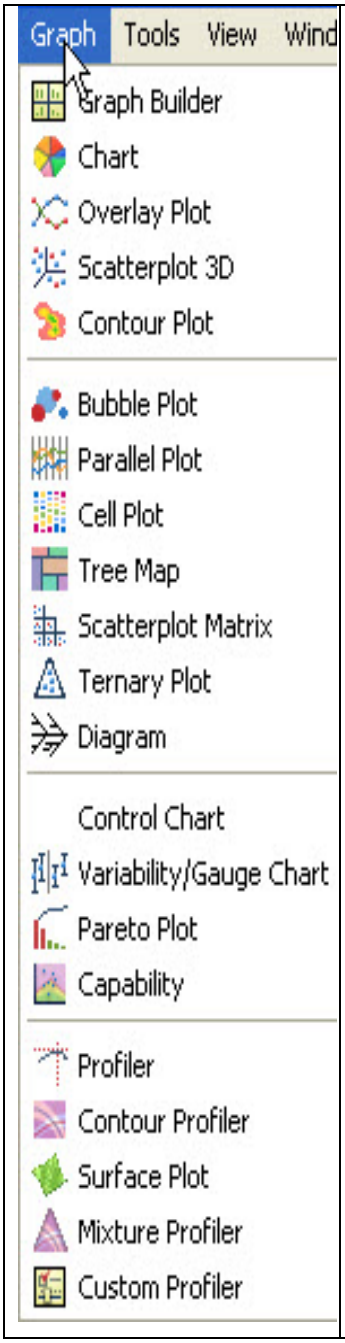
**Multivariate Methods** provide analysis methods for multivariate data. Cluster analysis, correlations, principal components, and discriminant analysis are the purveyance of this set of tools.

**Multivariate** describes relationships among response variables with correlations, a scatterplot matrix, a partial correlation matrix, the inverse correlation matrix, a multivariate outlier plot or a jackknifed outlier plot, and principal components analysis. One can also display the covariance matrix and see a three-dimensional spinning plot of the principal components.
**Cluster** clusters rows of a JMP table using the hierarchical, k-means, or EM (expectation maximization) method. The hierarchical cluster method displays results as a tree diagram of the clusters followed by a plot of the distances between clusters. The k-means cluster option, suitable for larger tables, iteratively assigns points to the number of clusters one specify. The EM mixture clustering method is for mixtures of distributions and is similar to the k-means method.
**Principal Components** performs principal component analysis. The purpose is to derive linear combinations of variables that describe the variability in the original variables.
**Discriminant** predicts classification variables (nominal or ordinal) based on a known continuous response. Discriminant analysis can be regarded as inverse prediction from a multivariate analysis of variance.
**PLS** analyzes data using partial least squares.
**Item Analysis** fits response curves from Item Response Theory (IRT) for analyzing survey or test data.

**Reliability and Survival** allows analysis of univariate or multivariate survival and reliability data.

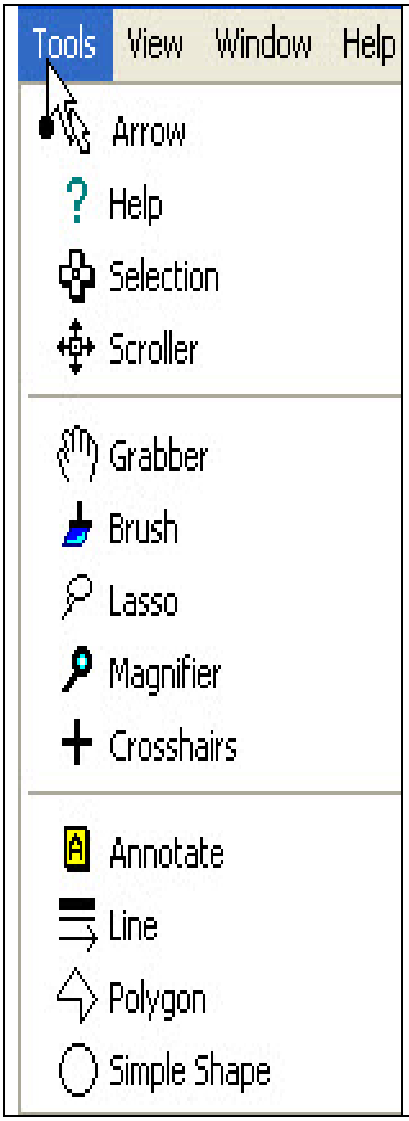
**Life Distribution** performs analysis of time-to-event data, including distributional fitting.
**Fit Life by X** provides for accelerated life-testing analysis.
**Survival** analyzes survival data using product-limit (Kaplan-Meier) survival computations.
**Fit Parametric Survival** launches a personality of the Fit Model platform to accomplish parametric fitting of censored data.
**Fit Proportional Hazards** launches a personality of the Fit Model platform for proportional hazard regression analysis that fits a Cox model.
**Recurrence Analysis** performs recurrence analysis.

## GRAPH MENU



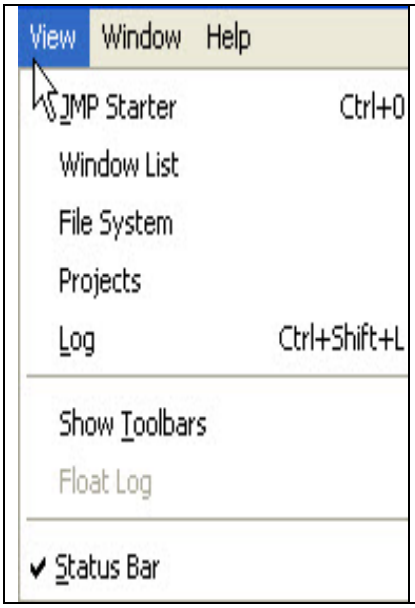

- produces bar and pie charts.
- produces overlay of a single numeric or categorical x column and all specified numeric y variables.
- produces a three-dimensional rotatable display of values from any three numeric columns in the active data table.
- constructs a contour plot for one or more response variables, y, for the values of two x variables.
- is a scatter plot which represents its points as circles and see up to five dimensions at once (x position, y position, size, color, and time).
- draws a parallel coordinate plot, which shows connected line segments representing each row of a data table.
- produces a rectangular array of cells drawn with a one-to-one correspondence to data table values.
- displays tree maps.
- allows quick production of scatterplot matrices.
- constructs a plot using triangular coordinates
- used to construct Ishikawa Charts, also called Fishbone Charts, or cause-and-effect diagrams
- produces quality control charts.
- Variability or Continuous Gauge charts are for responses whose values can be measured on a continuous scale. Attribute Gauge charts are for responses whose values are binary or categorical.
- produces Pareto Charts.
- used in quality control, measures the conformance of a process to given specification limits.
- is available for tables with columns whose values are computed from model prediction formulas.
- works the same as the Profiler command.
- plots surfaces and points in three dimensions based on formulas or data.
- is available for tables with columns whose values are computed from model prediction formulas.

## TOOLS MENU



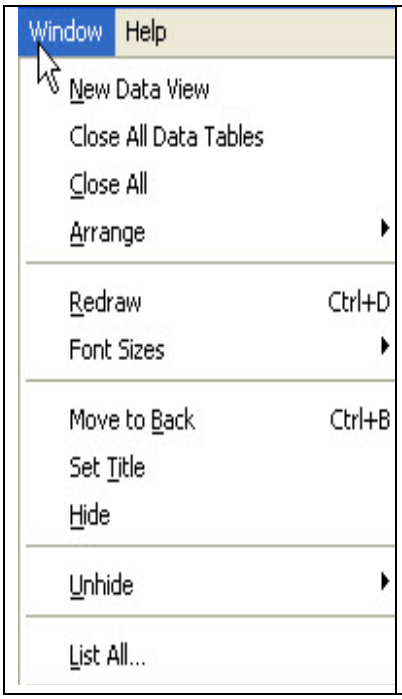

- **Arrow:** A default tool used to identify, highlight, and magnify points. Click on a point to highlight it. Click and hold on a point to identify the point. Shift-Click to extend a selection
- **Help:** To access JMP Help. Select the help tool and then click graphs, plots, or tables to see help windows.
- **Selection:** To select rows and columns in the data table or areas of a report. Use Shift-Click to extend the selection. To deselect click in the selected area.
- **Scroller:** To grab a report and scroll by dragging.
- **Grabber:** To direct manipulation of plots, charts, axes, and formula components.
- **Brush:** To highlight an area of points in plots. When one clicks, a rectangle appears. Move the rectangle over points to highlight them. Alt-Click to change the size of the selection rectangle and also extend the selection.
- **Lasso:** To highlight an irregular area of points in plots.
- **Magnifier:** To zoom in on any area of a plot. Alt-Click to restore the original plot.
- **Crosshairs:** A movable set of axes to measure points and distances in graphical displays.
- **Annotate:** To add editable text notes to a JMP report, journal, or layout window.
- **Line:** To draw thin, thick, or dashed lines which can have arrows on the ends.
- **Polygon:** To draw any shaped polygon. May be spline smoothed.
- **Simple Shape:** To draw either oval shapes or rectangles. May be filled or raised for a three-dimensional effect.

## VIEW MENU




- opens the JMP Starter Window.
- displays a pane at the left side of the JMP window that lists the name of each window opened in JMP.
- displays a pane at the left side of the JMP window that shows PC's file system.
- displays a pane at the left side of the JMP window that lists all open projects.
- displays a pane at the bottom of the JMP window that monitors JSL statements (JSL scripts) as they execute.
- displays a window that lists all available toolbars with a checkbox to show or hide them.
- detach or re-attach the log window to the bottom of the screen, right-click log and select float log window
- hides or shows the status bar at the bottom window edge.

**WINDOWS MENU**




- displays a duplicate view of an open data table.
- closes all data tables when the active window is a data table.
- closes all open windows.
- organizes the open windows within JMP.

- redraws active window.
- quick way to change the font size JMP uses.
- moves the active window behind all other windows generated by the current JMP session, leaving the next window in the sequence showing.
- changes the name of an active JMP window.
- suppresses the display of the active window but does not close it.
- displays a list of all hidden windows
- allows one to select the window one want to be the active window.

**HELP MENU**

To access the main help features from the help menu in JMP.

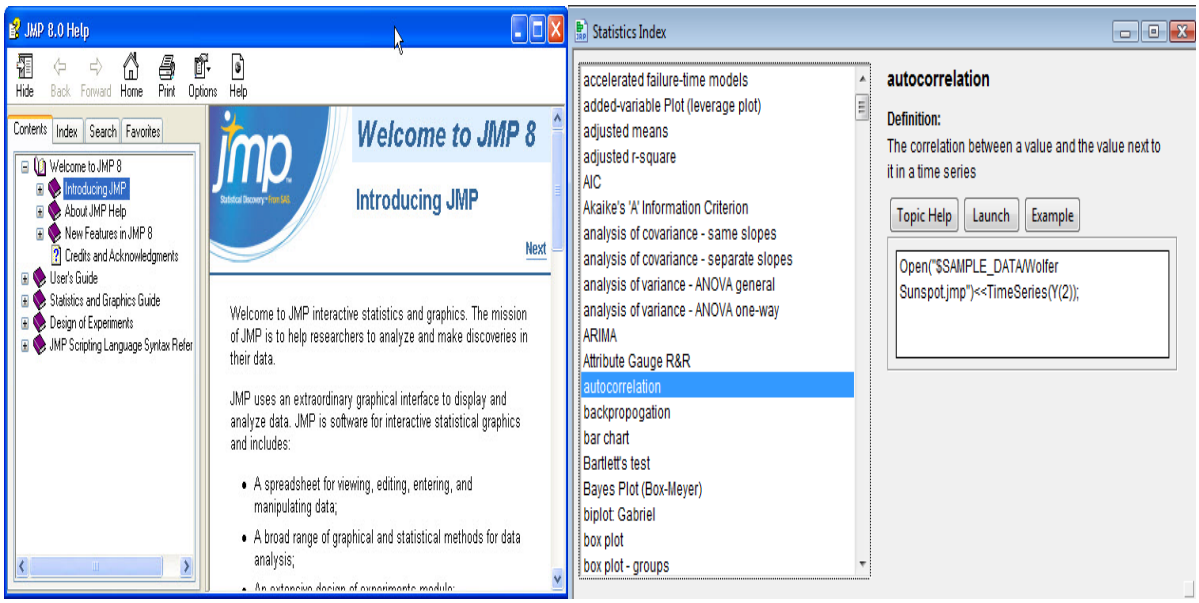



Figure 2.2: JMP Help Window and Statistics Index Window

One can read books on JMP in *.pdf format using **Help → Books →** (JMP Introductory Guide, JMP User Guide, JMP Stat and Graph Guide, JMP DOE Guide, JMP Scripting Guide, JMP Menu Card, JMP Quick Reference Card etc.). Also there is a **Statistics Index** window available in **Help → Indexes → Statistics**. From this window one can get help about various statistical terminologies with Topic help and an example on it. See on the right of Figure 2.2.

## 3. Working with JMP

### Importing Data

If one have data that exists in a format other than a *.jmp file, one can import it and save it as a JMP data table. The list below gives the file types which can be imported into JMP.

- Microsoft Excel (.xls), Microsoft Excel 2007 (.xlsm, .xlsx, .xlsb) on Windows.
- Text (.txt).
- Text with comma-separated values (.csv).
- Tabbed separated values (.tsv).
- SAS transport (.xpt, .stx) files.
- Minitab files (.mtw, .mtp).
- FACS (.fcs).
- Microsoft Access Database (.mdb) on Windows.
- Database (dBASE) (.dbf, .ndx, .mdx) on Windows.
- Data (.dat) files.
- HTML (.htm, .html).
- SAS versions 6-9 (.sd2, .sd5, .sd7, .sas7bdat) on Windows.

**3.1 Open a JMP Data Table :** There are several ways to open a data table:

- Go to the Sample Data directory in its default location. For example:

  D:\Program Files (x86)\SAS\JMP\8\Support Files English\Sample Data

The Sample data can also be viewed from menu **Help → Sample data**. It opens a screen which look like as Figure 3.1.1. One can open a number of sample JMP tables from various listed disciplines.

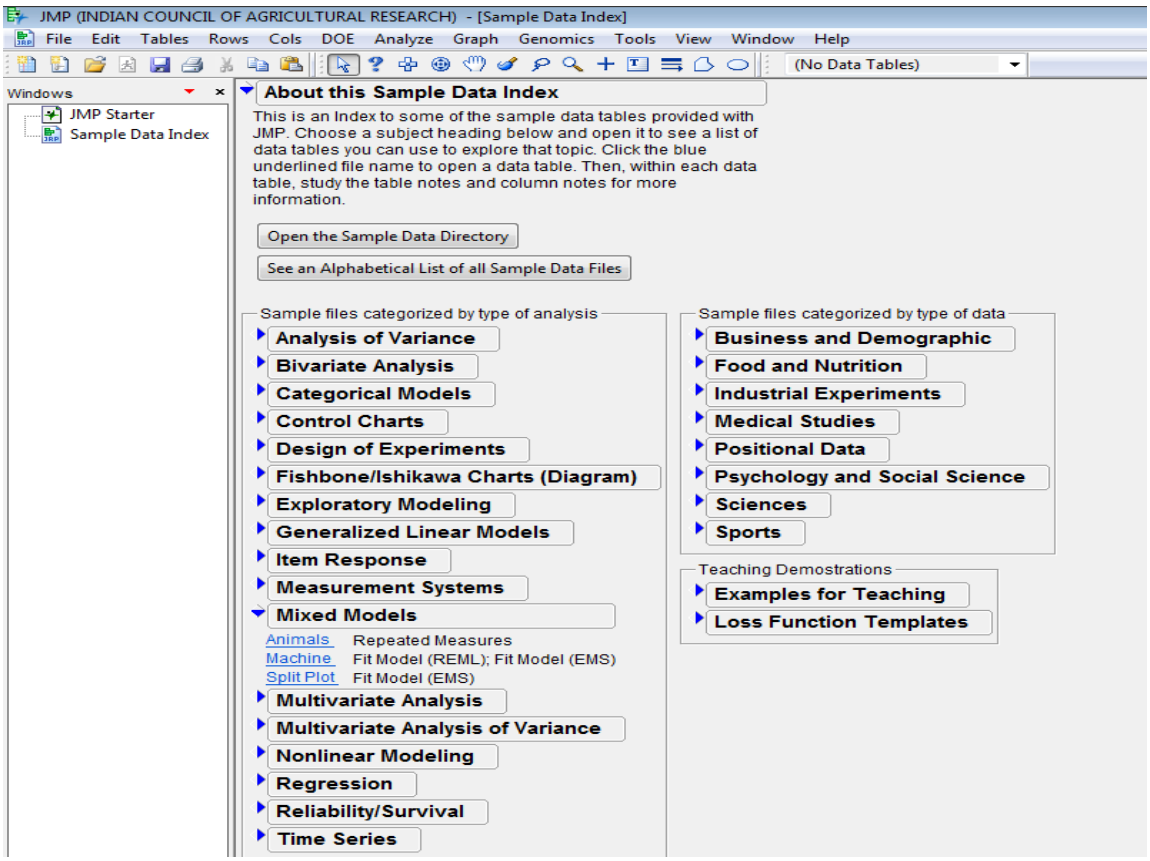



Figure 3.1.1: Sample Data Index

### 3.2 Open an Excel file in JMP

- Selecting **File → Open** (or clicking the Open Data Table button on the JMP Starter window) presents a file selection window with a list of existing tables.

  By clicking the drop-down menu just above the Open Tab, displays a list of files with extensions which one can open in JMP. One can open an EXCEL file by selecting *.xls, *.xlsx files. Alternatively, On clicking **All files** select the required Excel file, one gets 3 options for Always enforce Excel Row1 as labels, as "Best Guess", "Always" and "Never". It is preferred to use Best Guess option to open an Excel file (See Figure 3.2.1)

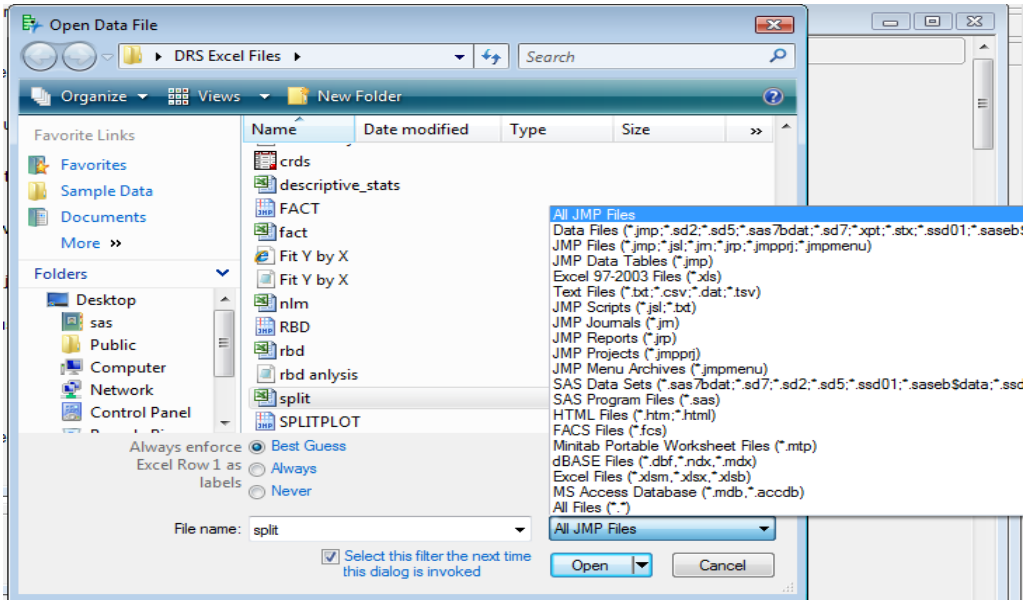



Figure 3.2.1: Opening a Excel file through Open Data File dialog box

### 3.3 Open a CSV (Comma Separated Value) file in JMP :

- Go to **File → Open**. It opens the Open Data File dialog box. Now click **Text files** from the drop-down button just above Open tab. Select the required CSV file by browsing then it displays 4 options. Select any of the first three options to open the file.

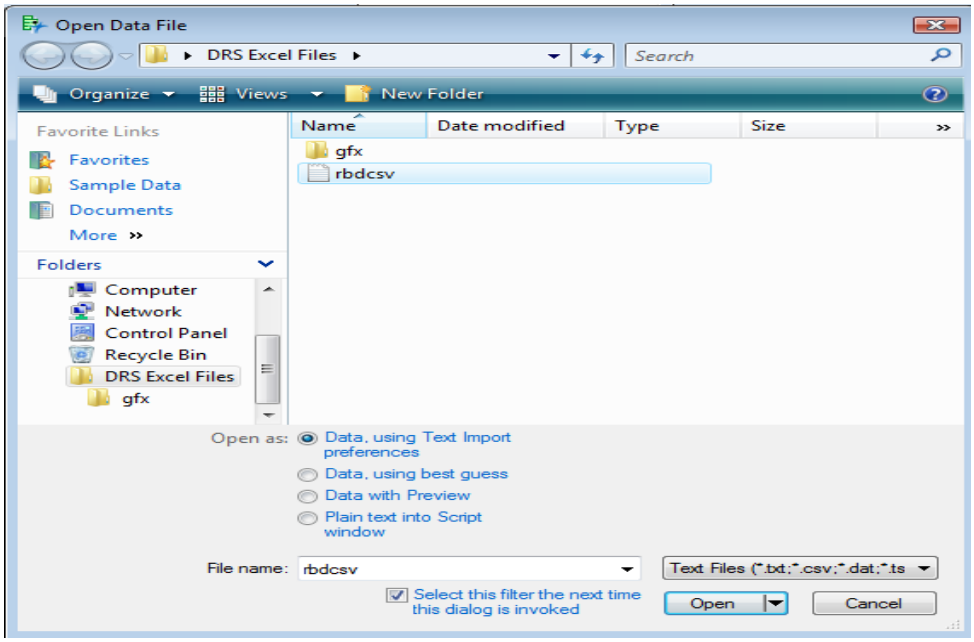



Figure 3.3.1: Opening a CSV file through Open Data File dialog box

### 3.4 Creating a New Data table

JMP data tables (Figure 3.4.1) have two parts: 1) data grid containing rows and columns of data, and 2) three data panels for the whole table, columns, and rows. The counts of table rows and columns appear in the corresponding column and rows data panel. In the data grid, a row number identifies each row, and each column has a column name. Within the data grid's red triangles for rows and columns, specific data can be selected based on variable attributes, and rows can be marked by specific column information, and column variables can be pre-designated as X, Y, weights, or frequency. Within the data panel red triangle for the whole table, summary statistics can be computed, missing data patterns can be analyzed, and table transformations (sort, transpose) are possible. These same options are available on a full menu bar under *Table*, *Columns*, or *Rows*.

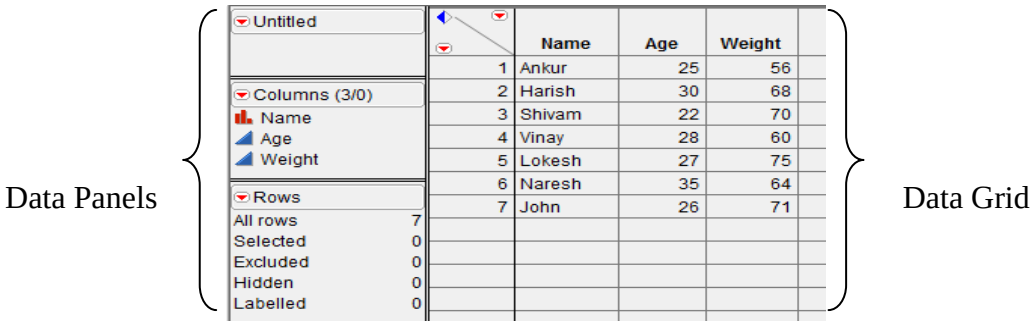



Figure 3.4.1: A Data Table

To create a new data table Select **File → New → Data Table**. To edit any column properties like column name, data type, modeling type and format etc. just double-click that column and do the required changes. It's very simple to enter the data just like MS-EXCEL. One can enter the values one by one and press enter key. If one wants to add multiple rows at once then right-click in the blank space say, say after seventh row then a Add Rows dialog box opens as in Figure 3.4.2. One can specify the number of rows to add. Similarly to add a new column, right-click on the blank space next to the last column. JMP data table is saved with *.jmp extension.

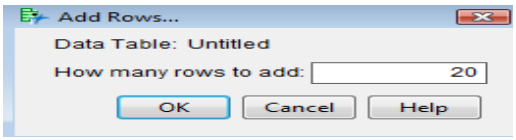



Figure 3.4.2: Add Rows Dialog Box

**Specifying the Values' Type :**

The small icon to the left of the column name in the columns panel can be used to declare the modeling type of the variable. JMP uses three modeling types to determine how to analyze the column's values:

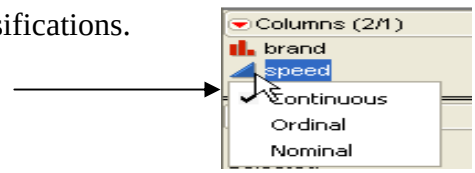
**Continuous ( )** Values are numeric measurements.

**Ordinal ( )** Values are ordered categories, which can have either numeric or character values.

**Nominal ( )** Values are numeric or character classifications.

To assign a different modeling type to a variable

1. Click the icon next to the variable name.
2. Select the appropriate modeling type.




The cursor changes to a hand when you move the mouse over a red triangle icon (also called **Hot-spot button**) ( ) and diamond-shaped **disclosure button** ( ). Click the red triangle to reveal the menu and select a menu icon. Clicking the disclosure button opens and closes sections of the output report.

# 4. Data Management in JMP

## 4.1 Creating a Subset Data Table

One can create a new data table that is a subset of all rows and columns, only highlighted rows and columns, or randomly selected rows from the active data table.

**Steps:** First select the rows or columns that one wants to take in new data table. Select **Tables** → **Subset**. A Subset window appears as in Figure 4.1.1

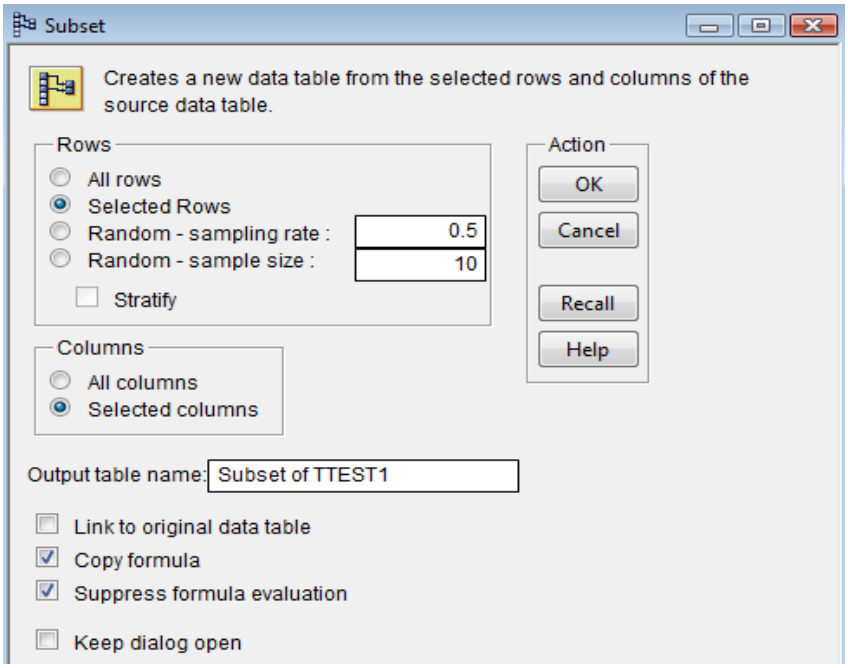



Figure 4.1.1: Subset Window

**Options in Subset Window**

One can select various options from Rows & Columns list. Some important options are discussed below:

1. **Random - sampling rate:** Creates a subset table whose data is a random proportion of the active data table. Enter the proportion of the sample in the text box. For example, if one wants a random 50% of the data to be included in the new table, enter 0.5 in the text box.
2. **Random - sample size:** Creates a subset table whose data is a random sample of the active data table. Enter the size of the sample in the text box. For example, if one want 16 random rows to be included in the new table, enter 16 into the text box.
3. **Output table name:** One can give the name of output table name. After clicking **OK**, a new data table gets created.
4. **Link to original data table:** To keep the subset table and any plot or graph of that subset table linked to the original table, check this box..
5. **Copy formula:** To include formulas from the original table in the output columns, check this box.
6. **Suppress formula evaluation:** To prevent JMP from evaluating columns' formulas when the new table is created, check this box.
7. **Keep dialog open:** To keep this window open after clicking **OK**, check this box.

**Stratified Subsets**

If one specify a sample size and add stratification columns, the sample size represents the size per stratum, rather than the size of the whole subset. See Figure 4.1.2.

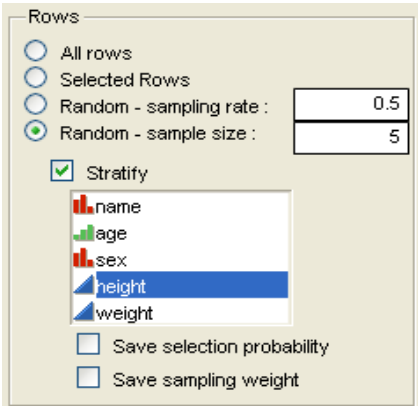



Figure 4.1.2: Stratified Subset option

There are also two columns that can be saved for stratified random subsets, **Selection Probability** and **Sampling Weight**. Check the corresponding check box to save these columns.

**Creating a Subset Data Table from a Report**

The following given two methods produce linked subsets table from a data table.

1) **Using a Histogram**

   Open TTEST.xls file. Select **Analyze** → **Distribution**. Put fw in **Y, Columns** → **OK**. One can use the histogram to create a new data table containing the data in the histogram's highlighted bars. To create a subset, **double-click** a highlighted bar. Or, right-click anywhere in the histogram and select **Subset** from the menu. The subset table appears, as shown in Figure 4.1.3.

2) **Using a Pareto Plot**

   From the output that contains a Pareto Plot (by selecting **Graph** → **Pareto Plot**), one can use the Pareto Plot to create a new data table containing the data in the Pareto Plot's highlighted bars. To create a subset, **double-click** a highlighted bar.

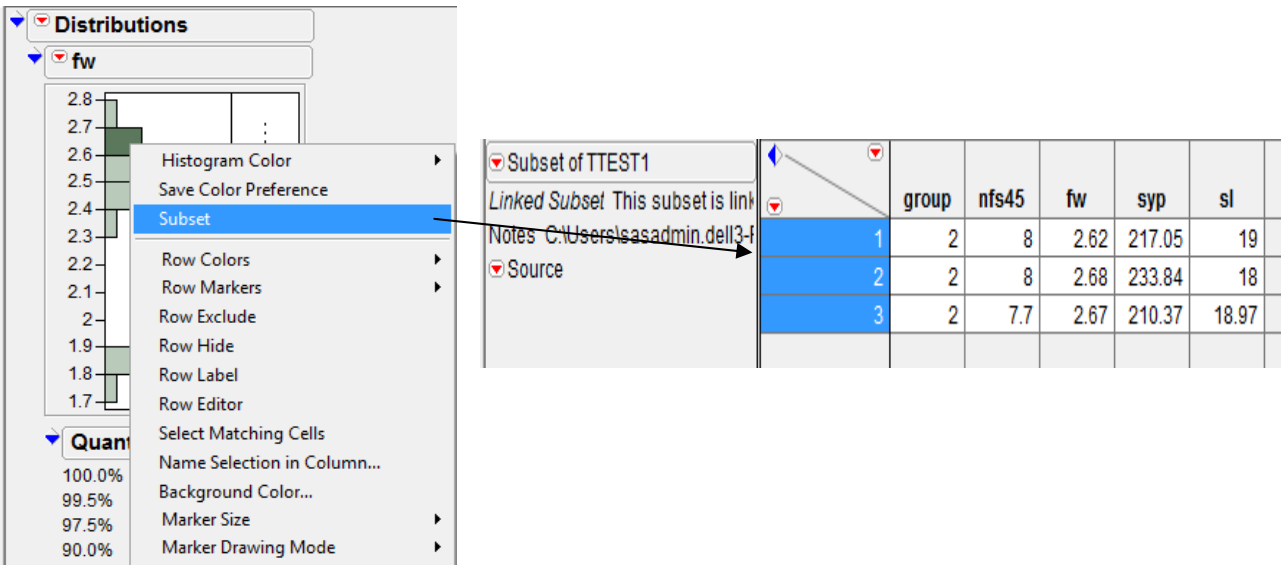



Figure 4.1.3: Subset created from a Histogram of variable fw from TTEST data table

### 4.2 Sorting Data Tables

One can sort a JMP data table by columns in either ascending or descending order. By default, columns sort in ascending order. One can either create a new table that contains the sorted values, or replace the original table with the sorted table. By default, it creates a new data table as Untitled.

**Steps:** Open TTEST.xls file. Select **Tables** → **Sort**. Highlight the name of the column by which one would like to sort, say **nfs45**. See the Sort window as in Figure 4.2.1. Click the **By** button to add them to the sort list The columns one add to the list establish the order of precedence for sorting. To replace the original data table with the sorted table instead of creating a new table with the sorted values, click the box beside **Replace Table**. Click **OK**.

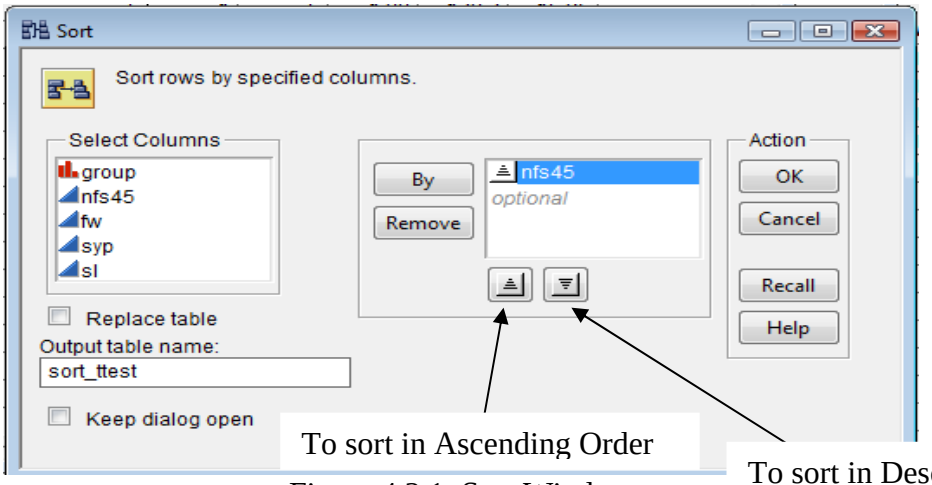



Figure 4.2.1: Sort Window

### 4.3 Splitting Columns

One can create a new data table from the active table by splitting one column into several new columns. This column is split according to the values found in another column, referred to as the **Split By** column. One can also split columns according to the values of one or more grouping variables.

**Splitting a Column: Basic Example**

**Steps:** Open **Data_Mgmt.jmp** file. The variety column shows that there were two Varieties, Arpita and Narangi. The objective of this example is to split the Antioxidant column into two new columns, one for each variety. See the table as in Figure

Data_Mgmt
Notes C:\Documents and Se
Columns (4/0)
Variety
Treatment
Rep
Antioxidant
Rows
All rows 24
Selected 0
Excluded 0
Hidden 0
Labelled 0

| | Variety | Treatment | Rep | Antioxidant |
|---|---|---|---|---|
| 1 | Narangi | DGN | 1 | 3.516883 |
| 2 | Narangi | DGN | 2 | 2.338451 |
| 3 | Narangi | DGN | 3 | 2.817189 |
| 4 | Narangi | CGN | 1 | 2.780363 |
| 5 | Narangi | CGN | 2 | 2.596233 |
| 6 | Narangi | CGN | 3 | 1.270497 |
| 7 | Narangi | BGN | 1 | 3.13021 |
| 8 | Narangi | BGN | 2 | 2.633059 |
| 9 | Narangi | BGN | 3 | 2.215 |
| 10 | Narangi | DFN | 1 | 1.086367 |
| 11 | Narangi | DFN | 2 | 1.981 |
| 12 | Narangi | DFN | 3 | 1.603 |
| 13 | Arpita | DGN | 1 | 2.743537 |
| 14 | Arpita | DGN | 2 | 2.017 |
| 15 | Arpita | DGN | 3 | 2.375277 |
| 16 | Arpita | CGN | 1 | 2.301625 |
| 17 | Arpita | CGN | 2 | 2.320038 |
| 18 | Arpita | CGN | 3 | 2.596233 |
| 19 | Arpita | BGN | 1 | 3.074971 |
| 20 | Arpita | BGN | 2 | 3.093384 |
| 21 | Arpita | BGN | 3 | 2.264799 |
| 22 | Arpita | DFN | 1 | 1.142 |
| 23 | Arpita | DFN | 2 | 1.105 |
| 24 | Arpita | DFN | 3 | 1.62 |

Figure 4.3.1: Original Data_Mgmt.jmp data table

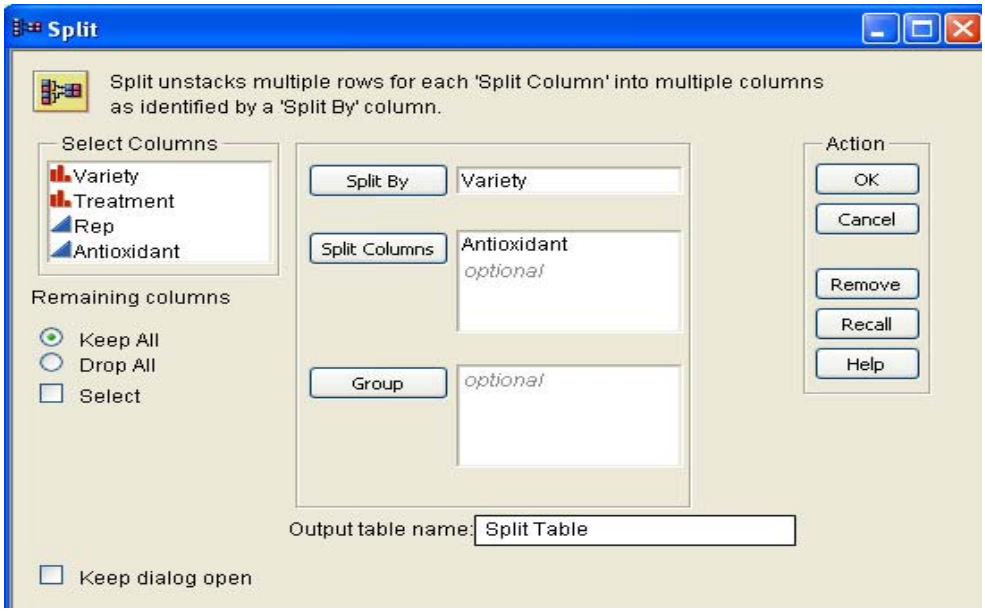



Figure 4.3.2: Split Window

Select **Tables → Split**. Add Antioxidant → **Split Columns** and Variety → **Split By** Column (See Figure 4.3.2.). The default is (**Drop All**) to omit any columns that are not in the **Split By**, **Split Columns**, or **Group** fields, so select the option **Keep All** choice to include these columns in the new table. Give "Split table" in **Output table name** field. Click **OK**. A new data table is created that looks like Figure 4.3.3. The values from the original antioxidant column are now split into the new columns named Arpita and Narangi. One can rename the new columns so the names are meaningful. Also, notice that the columns other than variety and antioxidant are exactly the same as they were in the original table.

Split Table

Split Table
Source
Columns (4/0)
Treatment
Rep
Arpita
Narangi
Rows
All rows 12

| | Treatment | Rep | Arpita | Narangi |
|---|---|---|---|---|
| 1 | DGN | 1 | 2.743537 | 3.516883 |
| 2 | DGN | 2 | 2.017 | 2.338451 |
| 3 | DGN | 3 | 2.375277 | 2.817189 |
| 4 | CGN | 1 | 2.301625 | 2.780363 |
| 5 | CGN | 2 | 2.320038 | 2.596233 |
| 6 | CGN | 3 | 2.596233 | 1.270497 |
| 7 | BGN | 1 | 3.074971 | 3.13021 |
| 8 | BGN | 2 | 3.093384 | 2.633059 |
| 9 | BGN | 3 | 2.264799 | 2.215 |
| 10 | DFN | 1 | 1.142 | 1.086367 |
| 11 | DFN | 2 | 1.105 | 1.981 |
| 12 | DFN | 3 | 1.62 | 1.603 |

The names of these new columns are values from the variety column, and the values in the new columns are from the antioxidant column.

Figure 4.3.3: New Table created by splitting Antioxidant column by Variety column

### 4.4 Stacking Columns

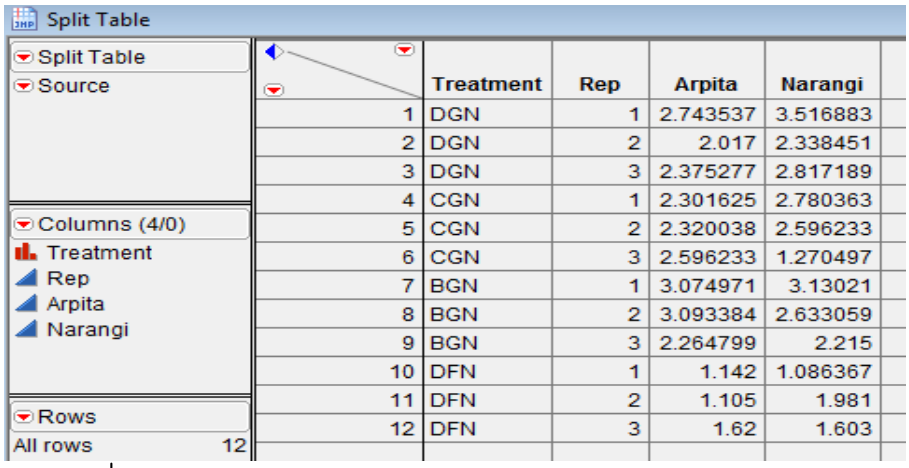
One can rearrange data table by stacking two or more columns into a single new column, preserving the values from the other columns. Or, one can stack a set of columns into multiple groups.

**Steps:** Use **Split Table.jmp** data table. In the data table, there are two columns **Arpita** and **Narangi** (See Figure 4.4.1). If one wants these two columns to be stacked into a single column and call this new column as **Antioxidant**. Go to **Tables** → **Stack**. Select Arpita and Narangi → **Stack Columns**. One can give the Output table name say 'Stacked Table'. Enter Antioxidant as **Stacked data Column** → **OK**. See Figure 4.4.2 for Stack Window and below it is the New data table (Stacked Table, Figure 4.4.3). A new column named **Variety** gets created.

### Options in Stack Window

Some important options are discussed below:

1. **Multiple series stack:** To stack selected columns into two or more columns, check this box. Specify the number of columns into which one wants the selected columns to be stacked by entering the number into the **Number of Series** box. This box appears only when **Multiple series stack** box is checked.
2. **Eliminate missing rows:** To eliminate missing data from the new table, check this box.
3. **Drop non-stacked columns:** To include only the stacked columns in the new table, check this box.




| | Treatment | Rep | Arpita | Narangi |
|---|---|---|---|---|
| 1 | DGN | 1 | 2.743537 | 3.516883 |
| 2 | DGN | 2 | 2.017 | 2.338451 |
| 3 | DGN | 3 | 2.375277 | 2.817189 |
| 4 | CGN | 1 | 2.301625 | 2.780363 |
| 5 | CGN | 2 | 2.320038 | 2.596233 |
| 6 | CGN | 3 | 2.596233 | 1.270497 |
| 7 | BGN | 1 | 3.074971 | 3.13021 |
| 8 | BGN | 2 | 3.093384 | 2.633059 |
| 9 | BGN | 3 | 2.264799 | 2.215 |
| 10 | DFN | 1 | 1.142 | 1.086367 |
| 11 | DFN | 2 | 1.105 | 1.981 |
| 12 | DFN | 3 | 1.62 | 1.603 |

Figure 4.4.1: Original Table (Split Table)

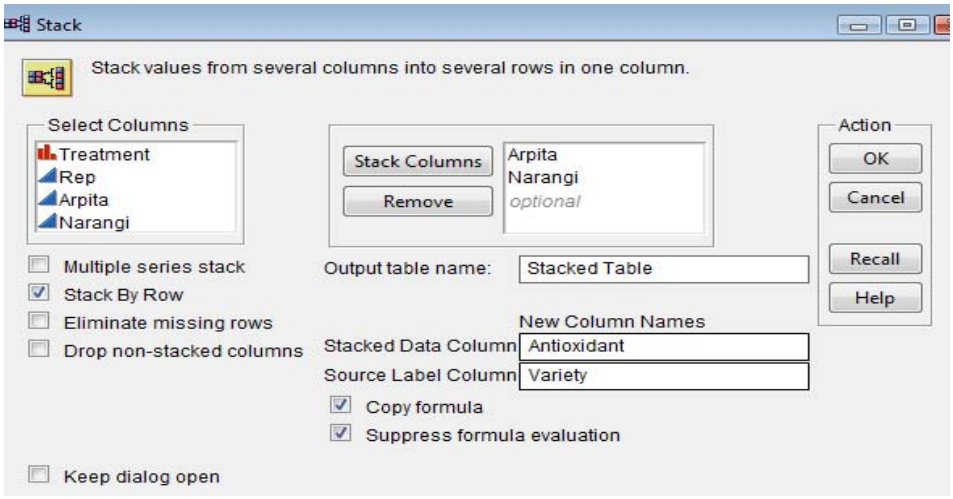
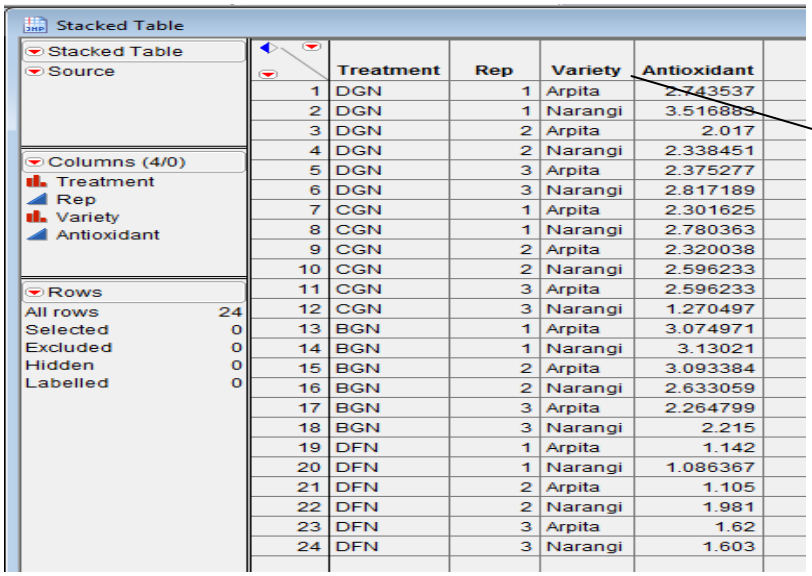



Figure 4.3.2: Stack Window



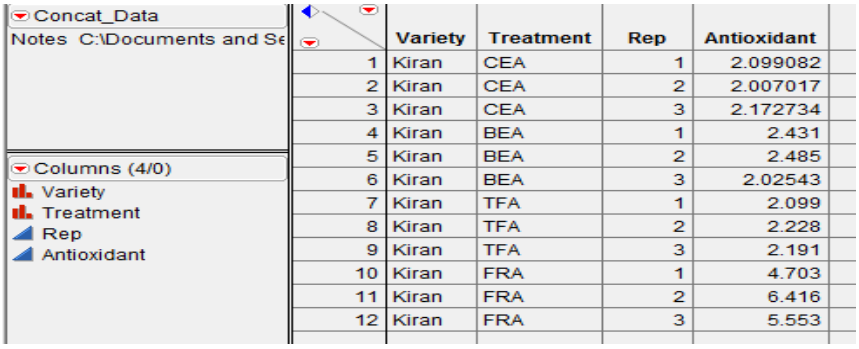

| | Treatment | Rep | Variety | Antioxidant |
|---|---|---|---|---|
| 1 | DGN | 1 | Arpita | 2.743537 |
| 2 | DGN | 1 | Narangi | 3.516883 |
| 3 | DGN | 2 | Arpita | 2.017 |
| 4 | DGN | 2 | Narangi | 2.338451 |
| 5 | DGN | 3 | Arpita | 2.375277 |
| 6 | DGN | 3 | Narangi | 2.817189 |
| 7 | CGN | 1 | Arpita | 2.301625 |
| 8 | CGN | 1 | Narangi | 2.780363 |
| 9 | CGN | 2 | Arpita | 2.320038 |
| 10 | CGN | 2 | Narangi | 2.596233 |
| 11 | CGN | 3 | Arpita | 2.596233 |
| 12 | CGN | 3 | Narangi | 1.270497 |
| 13 | BGN | 1 | Arpita | 3.074971 |
| 14 | BGN | 1 | Narangi | 3.13021 |
| 15 | BGN | 2 | Arpita | 3.093384 |
| 16 | BGN | 2 | Narangi | 2.633059 |
| 17 | BGN | 3 | Arpita | 2.264799 |
| 18 | BGN | 3 | Narangi | 2.215 |
| 19 | DFN | 1 | Arpita | 1.142 |
| 20 | DFN | 1 | Narangi | 1.086367 |
| 21 | DFN | 2 | Arpita | 1.105 |
| 22 | DFN | 2 | Narangi | 1.981 |
| 23 | DFN | 3 | Arpita | 1.62 |
| 24 | DFN | 3 | Narangi | 1.603 |

Figure 4.3.3: New Table (Stacked Table)

## 4.5 Concatenating Data Tables

One can concatenate data tables in JMP to combine rows from two or more data tables. One can create a new data table or append rows to the first data table. If a column name is the same in the data tables one want to concatenate, the column in the new data table lists the values from all data tables in the order of concatenation. If the two original data tables have columns with different names, those columns are included in the new data table showing missing values.

**Steps:** Open **Data_Mgmt.jmp** and **Concat_Data.jmp** (See Figure 4.5.1). Select **Tables → Concatenate. Add** Data_Mgmt & Concat_Data tables to concatenate as shown in Figure 4.5.2. Give **New Concat Table** name in the **Output table name** field → **OK**. See the output in Figure 4.5.3.




| | Variety | Treatment | Rep | Antioxidant |
|---|---|---|---|---|
| 1 | Kiran | CEA | 1 | 2.099082 |
| 2 | Kiran | CEA | 2 | 2.007017 |
| 3 | Kiran | CEA | 3 | 2.172734 |
| 4 | Kiran | BEA | 1 | 2.431 |
| 5 | Kiran | BEA | 2 | 2.485 |
| 6 | Kiran | BEA | 3 | 2.02543 |
| 7 | Kiran | TFA | 1 | 2.099 |
| 8 | Kiran | TFA | 2 | 2.228 |
| 9 | Kiran | TFA | 3 | 2.191 |
| 10 | Kiran | FRA | 1 | 4.703 |
| 11 | Kiran | FRA | 2 | 6.416 |
| 12 | Kiran | FRA | 3 | 5.553 |

Figure 4.5.1: Concat_Data.jmp data table

**Options in Concatenate Window:**

Some important options are discussed below:

1. **Save and evaluate formulas:** To request that JMP include all formulas. If one do not select this option, no formulas are included in the new data table. **Note:** If columns with the same name have different formulas, the formula from the first data table is saved in the concatenated data table.
2. **Create source column:** To add a column called Source Table to the new data table. This column identifies the name of the source data table in the corresponding rows.
3. **Append to first table:** To append rows to the data table listed first in the **Data Tables to be Concatenated** field instead of creating a new data table.

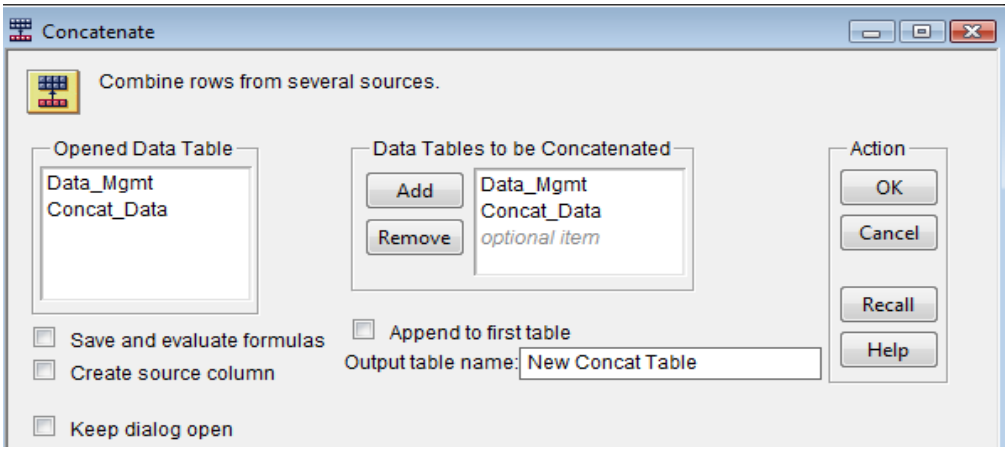
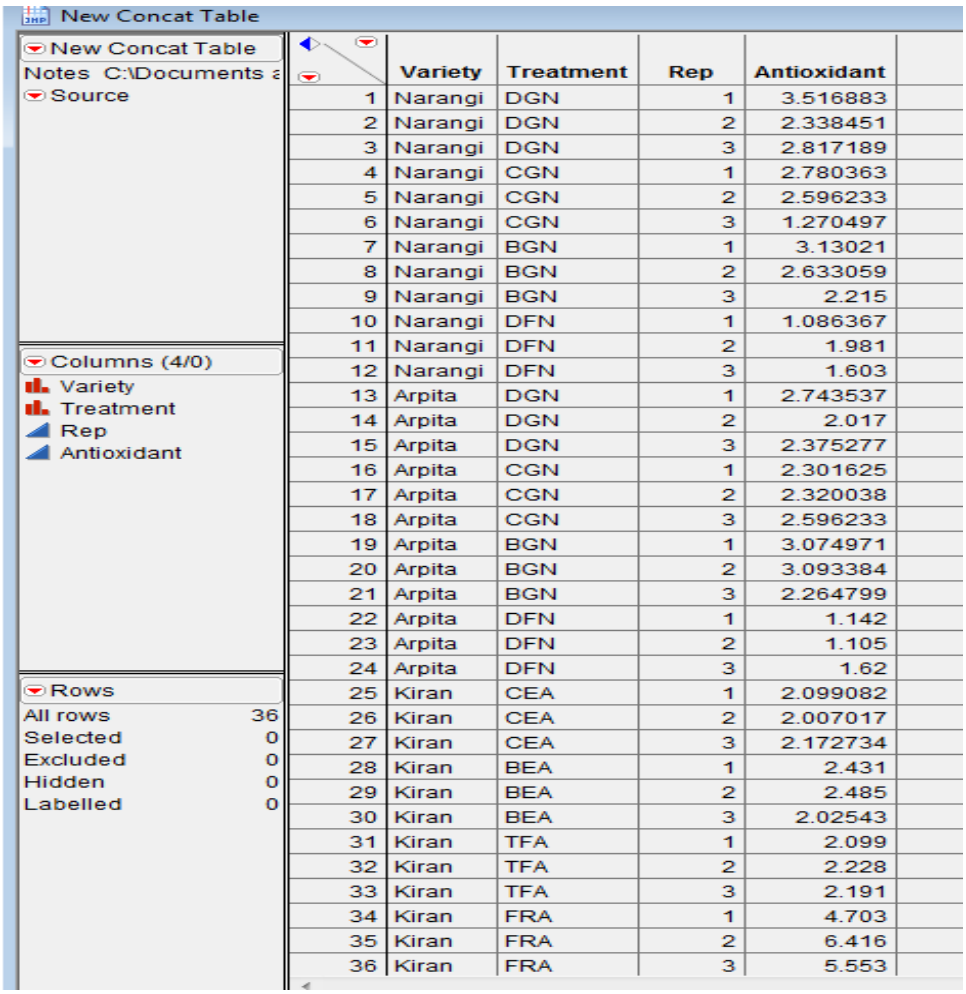



Figure 4.5.2: Concatenate Window




| | Variety | Treatment | Rep | Antioxidant |
|---|---|---|---|---|
| 1 | Narangi | DGN | 1 | 3.516883 |
| 2 | Narangi | DGN | 2 | 2.338451 |
| 3 | Narangi | DGN | 3 | 2.817189 |
| 4 | Narangi | CGN | 1 | 2.780363 |
| 5 | Narangi | CGN | 2 | 2.596233 |
| 6 | Narangi | CGN | 3 | 1.270497 |
| 7 | Narangi | BGN | 1 | 3.13021 |
| 8 | Narangi | BGN | 2 | 2.633059 |
| 9 | Narangi | BGN | 3 | 2.215 |
| 10 | Narangi | DFN | 1 | 1.086367 |
| 11 | Narangi | DFN | 2 | 1.981 |
| 12 | Narangi | DFN | 3 | 1.603 |
| 13 | Arpita | DGN | 1 | 2.743537 |
| 14 | Arpita | DGN | 2 | 2.017 |
| 15 | Arpita | DGN | 3 | 2.375277 |
| 16 | Arpita | CGN | 1 | 2.301625 |
| 17 | Arpita | CGN | 2 | 2.320038 |
| 18 | Arpita | CGN | 3 | 2.596233 |
| 19 | Arpita | BGN | 1 | 3.074971 |
| 20 | Arpita | BGN | 2 | 3.093384 |
| 21 | Arpita | BGN | 3 | 2.264799 |
| 22 | Arpita | DFN | 1 | 1.142 |
| 23 | Arpita | DFN | 2 | 1.105 |
| 24 | Arpita | DFN | 3 | 1.62 |
| 25 | Kiran | CEA | 1 | 2.099082 |
| 26 | Kiran | CEA | 2 | 2.007017 |
| 27 | Kiran | CEA | 3 | 2.172734 |
| 28 | Kiran | BEA | 1 | 2.431 |
| 29 | Kiran | BEA | 2 | 2.485 |
| 30 | Kiran | BEA | 3 | 2.02543 |
| 31 | Kiran | TFA | 1 | 2.099 |
| 32 | Kiran | TFA | 2 | 2.228 |
| 33 | Kiran | TFA | 3 | 2.191 |
| 34 | Kiran | FRA | 1 | 4.703 |
| 35 | Kiran | FRA | 2 | 6.416 |
| 36 | Kiran | FRA | 3 | 5.553 |

Figure 4.5.3: Result of Concatenating Two Data Tables

If there are two columns that do not match between the data tables say, antioxidant1 and antioxidant2, so the new concatenated data table has both antioxidant1 and antioxidant2 columns. These columns have missing values for rows from the data table in which the column did not exist.

### 4.6 Joining Data Tables

One can combine two data tables into one new table by selecting Tables Join. Tables can be joined in three different ways:

1. By combining them according to row number.
2. In a Cartesian fashion, where one form a new table consisting of all possible combinations of the rows from two original tables.
3. By matching the values in one or more columns that exist in both data tables.

#### 4.6.1 Join by Row Number

Joining tables by row number joins the two tables side by side, and the new table has all columns from both tables (unless one specify to include only certain columns).

**a) Joining Tables with an Unequal Number of Rows** - If one want to join two tables with unequal number of rows, the new table will have values for rows found in both tables.

**Steps**: Open **Variety.jmp** and **Season.jmp** data table. Make Variety as current data table by highlighting it. Select Tables → **Join**. Select Season in **Join..With** Box. Select **By Row Number** in **Matching Specialization** Box. Give **Output table name** as 'Join table'. Click **OK** (See Figure 4.6.1.1 for Output).

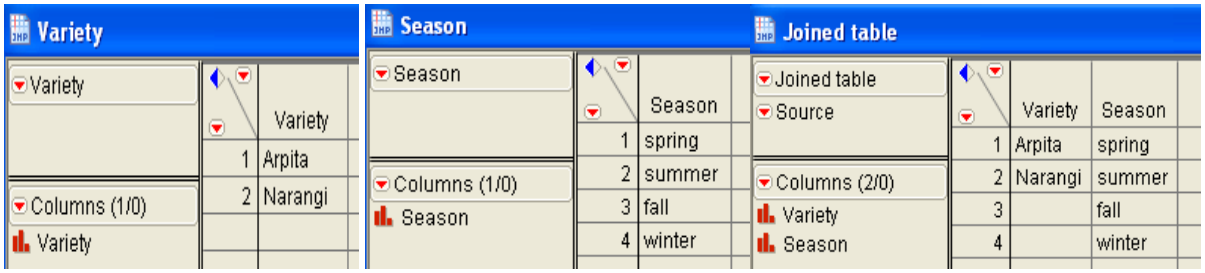



Figure 4.6.1.1: Joining Tables by Row Number

**b) Joining Columns with the Same Name -** If the two tables have same column names , the names of these columns in the new table appear as "column name of table name." For example, if one joined tables named **Animal Data** and **Reptile Data**, and both tables contained a column named **gender**, the new table contains columns named 'gender of Animal Data' and 'gender of Reptile Data' , as shown in Figure 4.6.1.2.

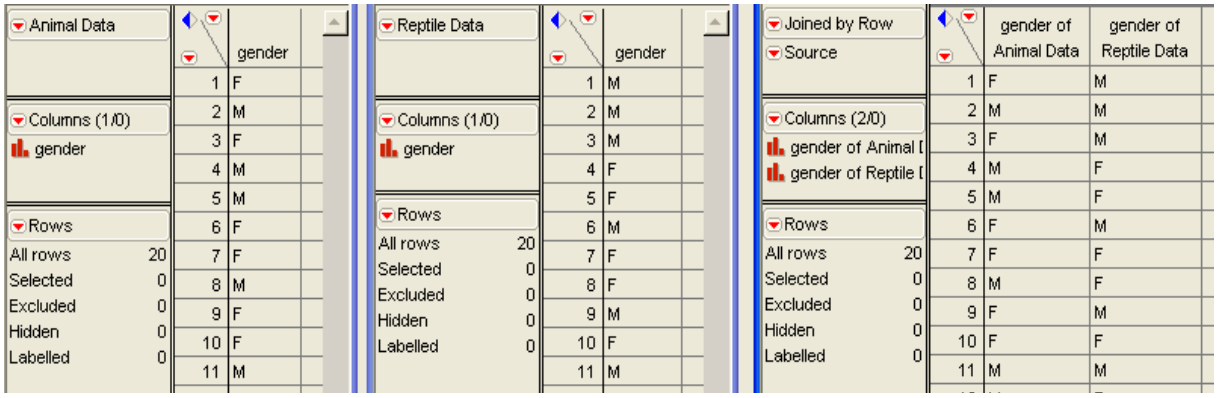



Figure 4.6.1.2: Joining Tables by Row Number

**c) Joining All Columns** - If one want to combine rows from each data table so that new data table contains all columns from two data tables.

**Steps:** Open **Concat_Data.jmp** and **Join_all.jmp**. Select **Tables** → **Join**. Select Join_all in **Join..With** Box, See Figure 4.6.1.4 for Join Window. Select **By Row Number** in **Matching Specialization** Box. Give **Output table name** as 'New Joined Table'. Click **OK**. See Figure 4.6.1.5 for Output.

Join_all

Columns (4/0)
- Variety
- Treatment
- Rep
- Antioxidant

Rows

| | |
|---|---|
| All rows | 12 |
| Selected | 0 |
| Excluded | 0 |
| Hidden | 0 |
| Labelled | 0 |

| | Variety | Treatment | Rep | Antioxidant |
|---|---|---|---|---|
| 1 | Kiran | CEA | 1 | 1.804 |
| 2 | Kiran | CEA | 2 | 2.707 |
| 3 | Kiran | CEA | 3 | 2.413 |
| 4 | Kiran | BEA | 1 | 2.431 |
| 5 | Kiran | BEA | 2 | 2.136 |
| 6 | Kiran | BEA | 3 | 1.423 |
| 7 | Kiran | TFA | 1 | 2.744 |
| 8 | Kiran | TFA | 2 | 2.017 |
| 9 | Kiran | TFA | 3 | 2.375 |
| 10 | Kiran | FRA | 1 | 2.302 |
| 11 | Kiran | FRA | 2 | 2.32 |
| 12 | Kiran | FRA | 3 | 2.596 |

Figure 4.6.1.3: Join_all.jmp data table

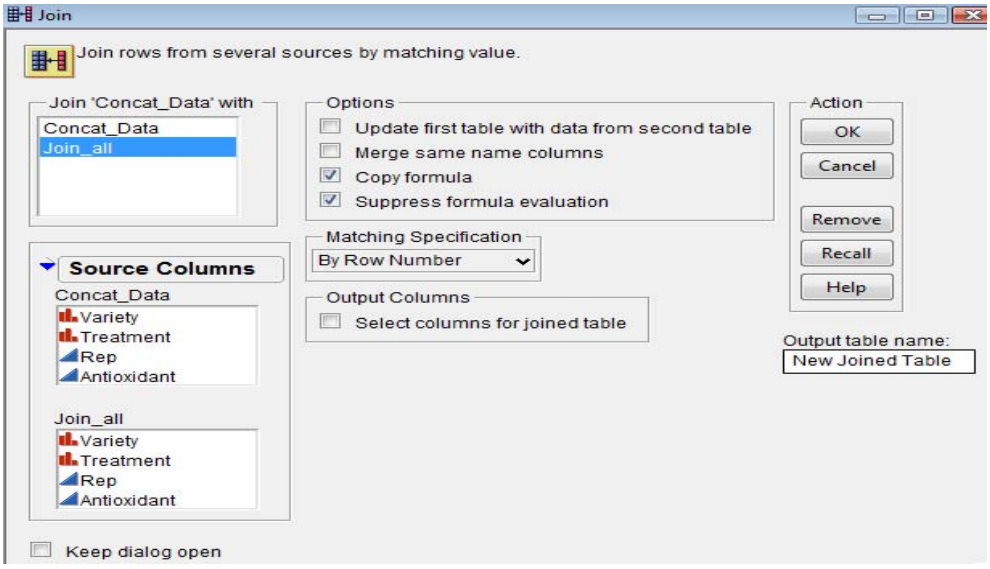



Figure 4.6.1.4: Join Window

New Joined Table

New Joined Table
Notes C:\Documents ar
Source

Columns (8/0)
- Variety of Concat_Da
- Treatment of Concat
- Rep of Concat_Data
- Antioxidant of Concat
- Variety of Join_all
- Treatment of Join_al
- Rep of Join_all

| | Variety of Concat_Data | Treatment of Concat_Data | Rep of Concat_Data | Antioxidant of Concat_Data | Variety of Join_all | Treatment of Join_all | Rep of Join_all | Antioxidant of Join_all |
|---|---|---|---|---|---|---|---|---|
| 1 | Kiran | CEA | 1 | 2.099082 | Kiran | CEA | 1 | 1.804 |
| 2 | Kiran | CEA | 2 | 2.007017 | Kiran | CEA | 2 | 2.707 |
| 3 | Kiran | CEA | 3 | 2.172734 | Kiran | CEA | 3 | 2.413 |
| 4 | Kiran | BEA | 1 | 2.431 | Kiran | BEA | 1 | 2.431 |
| 5 | Kiran | BEA | 2 | 2.485 | Kiran | BEA | 2 | 2.136 |
| 6 | Kiran | BEA | 3 | 2.02543 | Kiran | BEA | 3 | 1.423 |
| 7 | Kiran | TFA | 1 | 2.099 | Kiran | TFA | 1 | 2.744 |
| 8 | Kiran | TFA | 2 | 2.228 | Kiran | TFA | 2 | 2.017 |
| 9 | Kiran | TFA | 3 | 2.191 | Kiran | TFA | 3 | 2.375 |
| 10 | Kiran | FRA | 1 | 4.703 | Kiran | FRA | 1 | 2.302 |
| 11 | Kiran | FRA | 2 | 6.416 | Kiran | FRA | 2 | 2.32 |
| 12 | Kiran | FRA | 3 | 5.553 | Kiran | FRA | 3 | 2.596 |

Figure 4.6.1.5: The Joined table

**d) Joining Only Specified Columns** - If one doesn't want all columns from the original data tables to be in the joined table say, only want Variety, Treatment, Rep & Antioxidant from 'Concat_Data' and only Antioxidant from 'Join_all' to be in the new joined table.

**Steps:** Select **Tables → Join**. Select Join all col JMP in **Join…With** Box. Select **By Row Number** in **Matching Specialization** Box. Click **Select columns for joined table** in the **Output Columns** area to specify the subset of columns one want. Select all the columns one want from both tables in the **Source Columns** list and click **Select**. In this example, select Variety, Treatment, Rep & Antioxidant from 'Concat_Data' and only Antioxidant from 'Join_all' list. The box in the **Output Columns** area lists the columns one want in the new table. The tables **Concat_Data.jmp** and **Join_all.jmp** (See Figure 4.6.1.5) have identical data in the Variety, Treatment & Rep columns, so only one of them is needed in the joined table. Give **Output table name** as 'Joined selected'. Click **OK**. The new table has only the selected columns. See Figure 4.6.1.6 for output.

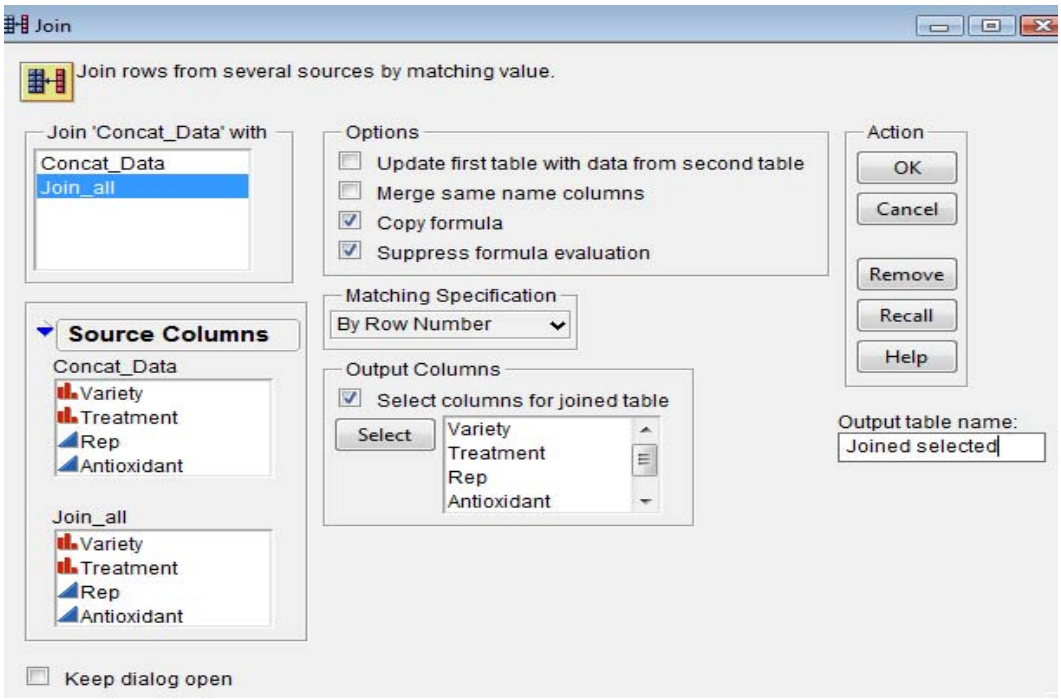
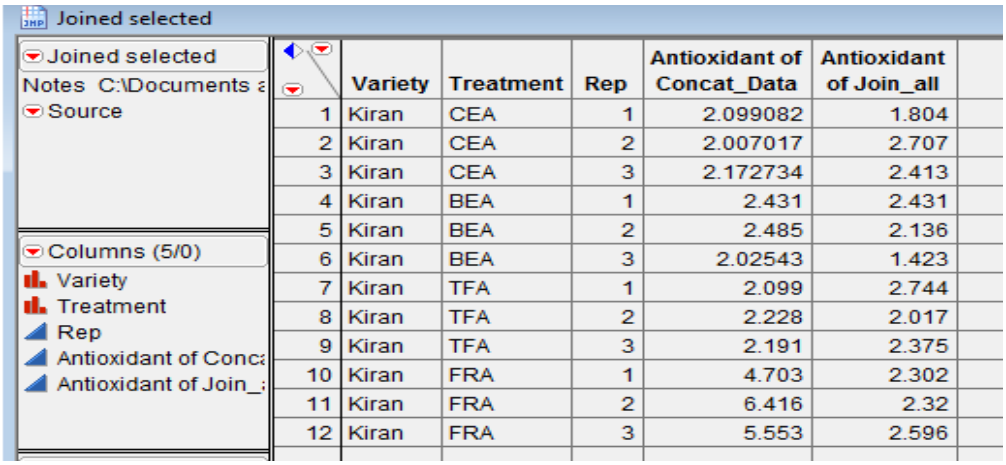



Figure 4.6.1.5: Joining Only Specified Column




| | Variety | Treatment | Rep | Antioxidant of Concat_Data | Antioxidant of Join_all |
|---|---|---|---|---|---|
| 1 | Kiran | CEA | 1 | 2.099082 | 1.804 |
| 2 | Kiran | CEA | 2 | 2.007017 | 2.707 |
| 3 | Kiran | CEA | 3 | 2.172734 | 2.413 |
| 4 | Kiran | BEA | 1 | 2.431 | 2.431 |
| 5 | Kiran | BEA | 2 | 2.485 | 2.136 |
| 6 | Kiran | BEA | 3 | 2.02543 | 1.423 |
| 7 | Kiran | TFA | 1 | 2.099 | 2.744 |
| 8 | Kiran | TFA | 2 | 2.228 | 2.017 |
| 9 | Kiran | TFA | 3 | 2.191 | 2.375 |
| 10 | Kiran | FRA | 1 | 4.703 | 2.302 |
| 11 | Kiran | FRA | 2 | 6.416 | 2.32 |
| 12 | Kiran | FRA | 3 | 5.553 | 2.596 |

Figure 4.6.1.6: New Joined Table

**4.6.2 A Cartesian Join -** When doing a Cartesian join, JMP joins two tables in a Cartesian fashion, where it forms a new table consisting of all possible combinations of the rows from two original tables. This creates cases in the output table so there will be one case for each combination of column values. For example, as Figure 4.6.7 shows, JMP crosses the data in table Variety with the data in table Season to display all combinations of the values in each set (the table named Joined table).

**Steps:** Open **Variety.jmp** & **Season.jmp**. Select **Tables** → **Join**. Select Season in **Join… With** Box. Select **By Cartesian Join** in **Matching Specialization** Box. Give **Output table name** as 'Joined Table'. Click **OK**. See Figure 4.6.2.1 for Output.

Variety.jmp Season.jmp Joined table.jmp

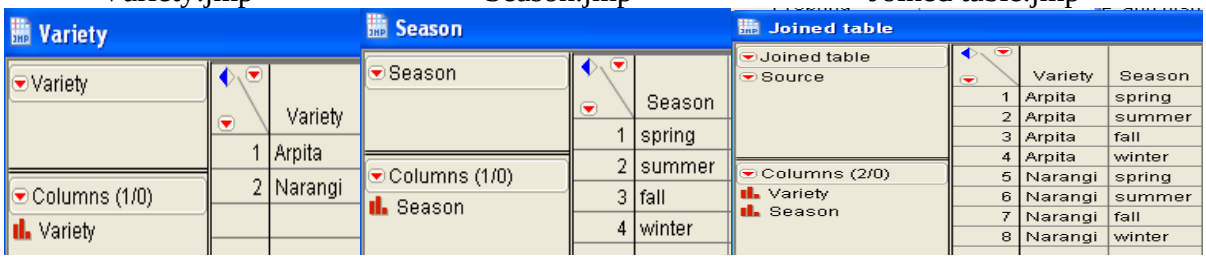



Figure 4.6.2.1: Joining tables using Cartesian Join

**4.6.3 Join by Matching Columns** - If one wants to join by matching columns, JMP finds specified column(s) values that exist in both tables and combines all values associated with that value into a new data table. **Note:** To join by matching columns, the columns must have the same data type (numeric, character, or row state).

**a) Joining Tables with the Same Rows in Different Order**

Suppose one have one data table containing students' names, ages, and sexes & another data table containing their names, height, and weight. Instead of working with two separate tables, one would like to combine the tables into one data table containing students' name, age, sex, height and weight.

**Steps:** Open **Students1.jmp** & **Students2.jmp** in the Sample Data Directory. Select **Tables** → **Join**. Select Students2 in **Join...With** Box. Select **By Matching Columns** in **Matching Specialization** Box. Highlight name from **Students1**'s list and name from **Student2**'s list → **Match**. One want the new table to contain only one row for each name, so check the **Drop multiples** boxes for both tables (Figure 4.6.3.1). Give 'joining by matching' as **Output table name**. Then, click **OK**. See partial output in Figure 4.6.3.2.

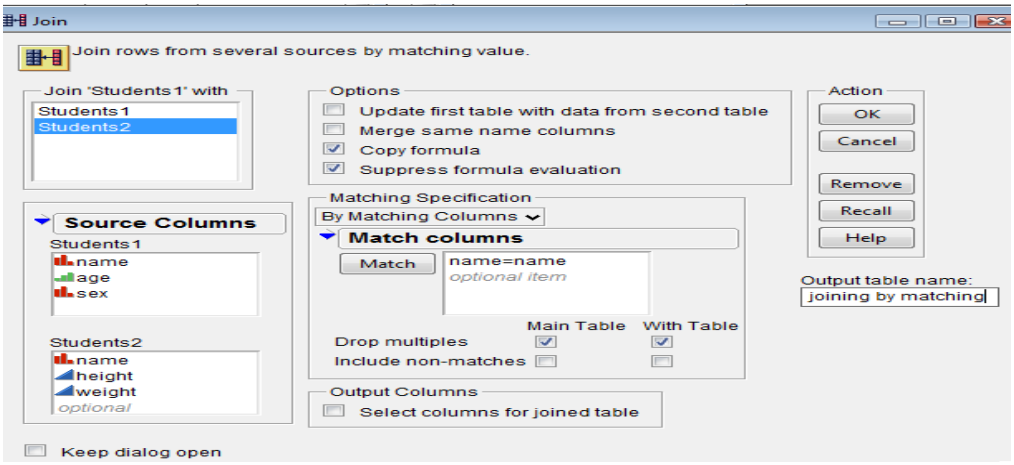



Figure 4.6.3.1: Joining by matching Columns Window

| | name of Students1 | age | sex | name of Students2 | height | weight |
|---|---|---|---|---|---|---|
| 1 | ALFRED | 11 | M | ALFRED | 59 | 87 |
| 2 | ALICE | 11 | F | ALICE | 56 | 84 |
| 3 | AMY | 11 | M | AMY | 55 | 74 |
| 4 | BARBARA | 11 | F | BARBARA | 53 | 64 |
| 5 | CAROL | 11 | M | CAROL | 60 | 95 |
| 6 | CHRIS | 11 | M | CHRIS | 58 | 85 |
| 7 | CLAY | 12 | F | CLAY | 63 | 108 |
| 8 | DANNY | 12 | F | DANNY | 59 | 89 |
| 9 | DAVID | 11 | F | DAVID | 56 | 73 |
| 10 | EDWARD | 11 | M | EDWARD | 58 | 101 |
| 11 | ELIZABETH | 11 | M | ELIZABETH | 61 | 87 |
| 12 | FREDERICK | 11 | M | FREDERICK | 58 | 75 |
| 13 | HENRY | 11 | M | HENRY | 55 | 70 |
| 14 | JACLYN | 11 | F | JACLYN | 62 | 104 |

Figure 4.6.3.2: Joining Students1.jmp with Students2.jmp by matching columns

If one want only name, age, sex, height & weight in the new data table then check **Select columns for joined table** and select required columns from **Source Columns** from Join window in Figure 4.6.3.1.

**b) Joining Tables with Different Numbers of Rows and Different Column Names**

Suppose one have the following two data tables with different number of rows and different column names as in Figure 4.6.3.3. In the new table, one only want one column for Variety, one column for Treatment, one column for Rep and two columns for Antioxidant—Antioxidant from Join_all and Antioxidant from Join_diff as in Figure 4.6.3.4.

Join_all.jmp

Join_all
Columns (4/0)
Variety
Treatment
Rep
Antioxidant

| | Variety | Treatment | Rep | Antioxidant |
|---|---|---|---|---|
| 1 | Kiran | CEA | 1 | 1.804 |
| 2 | Kiran | CEA | 2 | 2.707 |
| 3 | Kiran | CEA | 3 | 2.413 |
| 4 | Kiran | BEA | 1 | 2.431 |
| 5 | Kiran | BEA | 2 | 2.136 |
| 6 | Kiran | BEA | 3 | 1.423 |
| 7 | Kiran | TFA | 1 | 2.744 |
| 8 | Kiran | TFA | 2 | 2.017 |
| 9 | Kiran | TFA | 3 | 2.375 |
| 10 | Kiran | FRA | 1 | 2.302 |
| 11 | Kiran | FRA | 2 | 2.32 |
| 12 | Kiran | FRA | 3 | 2.596 |

Join_diff.jmp

Join_diff
Source
Columns (4/0)
Variety
Trt
Replication
Antioxidant

| | Variety | Trt | Replication | Antioxidant |
|---|---|---|---|---|
| 1 | Kiran | CEA | 1 | 1.804 |
| 2 | Kiran | CEA | 2 | 2.707 |
| 3 | Kiran | CEA | 3 | 2.413 |
| 4 | Kiran | BEA | 1 | 2.431 |
| 5 | Kiran | BEA | 2 | 2.136 |
| 6 | Kiran | BEA | 3 | 1.423 |

Figure 4.6.3.3: Join_all.jmp & Join_diff.jmp data tables

Join all_diff

Join all_diff
Source
Columns (5/0)
Variety
Treatment
Rep
Antioxidant of Join_
Antioxidant of Join_
Rows
All rows 12
Selected 0
Excluded 0
Hidden 0

| | Variety | Treatment | Rep | Antioxidant of Join_all | Antioxidant of Join_diff |
|---|---|---|---|---|---|
| 1 | Kiran | BEA | 1 | 2.431 | 2.431 |
| 2 | Kiran | BEA | 2 | 2.136 | 2.136 |
| 3 | Kiran | BEA | 3 | 1.423 | 1.423 |
| 4 | Kiran | CEA | 1 | 1.804 | 1.804 |
| 5 | Kiran | CEA | 2 | 2.707 | 2.707 |
| 6 | Kiran | CEA | 3 | 2.413 | 2.413 |
| 7 | Kiran | FRA | 1 | 2.302 | • |
| 8 | Kiran | FRA | 2 | 2.32 | • |
| 9 | Kiran | FRA | 3 | 2.596 | • |
| 10 | Kiran | TFA | 1 | 2.744 | • |
| 11 | Kiran | TFA | 2 | 2.017 | • |
| 12 | Kiran | TFA | 3 | 2.375 | • |

Figure 4.6.3.4: Table Joined by Matching Columns

**Steps:** Open **Join_all.jmp** & **Join_diff.jmp**. Select **Tables** → **Join**. Select Join_diff in **Join...With** box. Select **By Matching Columns** in the **Matching Specification** area. Highlight Variety, Treatment, and Rep from Join_all's list. Highlight Variety, Trt, and Replication from Join_diff's list. Click **Match**. Check the **Include Non Matches** boxes for both tables. Check the box beside **Select columns for joined table**. Highlight Variety, Treatment, Rep and Antioxidant from **Join_all** list → **Select**. Highlight Antioxidant from **Join_diff** list → **Select**. Give 'join all_diff' as **Output table name**. The Join Window should look like as Figure 4.6.3.5 and click **OK**.

**Note:** The yield column from Little.jmp (Antioxidant of Join_diff) has missing values whenever there were no matching values in Join_all.

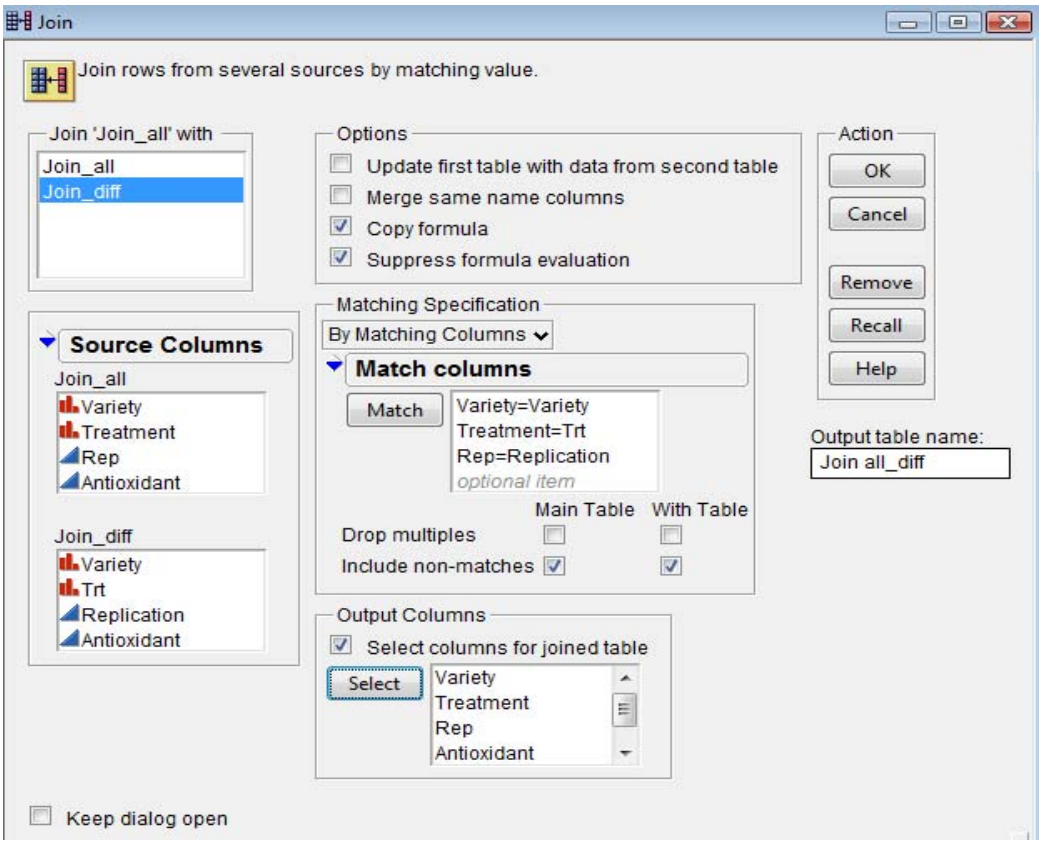



Figure 4.6.3.5: Join by Matching Columns Window

### 4.7 Missing Data Pattern

If data table contains missing data, one might want to determine whether there is a pattern that the missing data creates. The pattern might help one make discoveries about the data.

**Steps:** Open **Missing_Data.jmp**. (See data table shown to the left of Figure 4.7.2). Select **Tables** → **Missing Data Pattern**. Highlight the columns from which one would like to find missing data. Select Trial1, Trial2, Trial3 and Trial4 from **Select Columns** list. Click **Add Columns** → **OK** (See Figure 4.7.1). By Default, **Output table name** is 'Missing Data Pattern'. See the Missing Data Pattern Table as shown to the right of Figure 4.7.2.

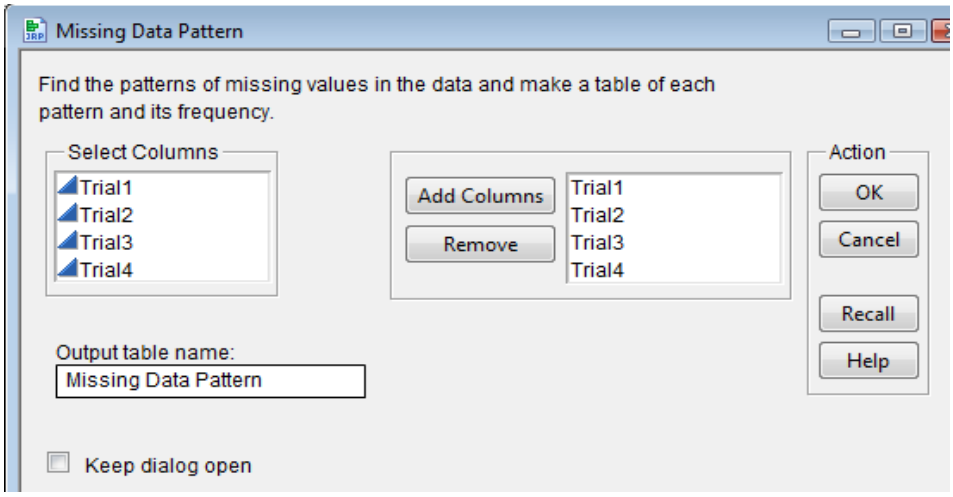



Figure 4.7.1: Missing Data Pattern Window

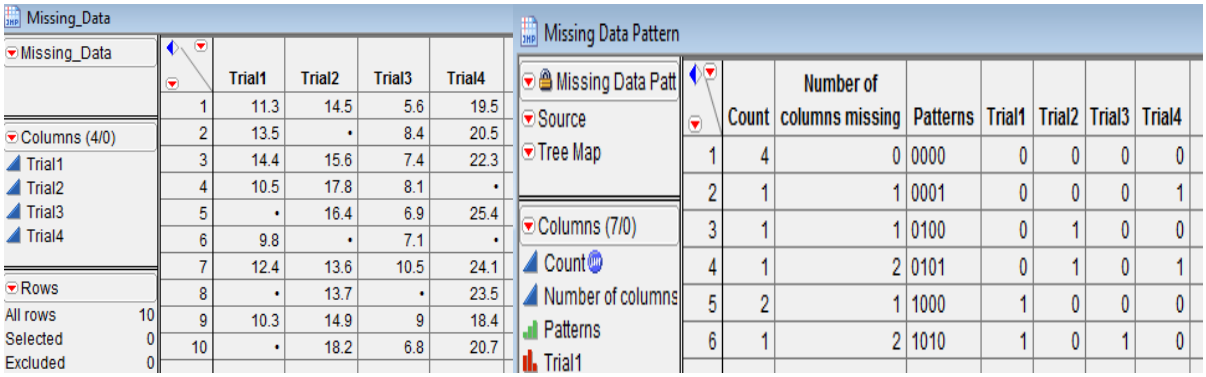



Figure 4.7.2: Missing_Data.jmp table & Missing Data Pattern data table

Figure 4.7.2 shows the following patterns:

- Row 1 shows that there are four instances where all rows in Trial 1, Trial 2, Trial 3, and Trial 4 have no missing values.
- Row 2 shows that there is one row in the source table whose one missing value is in the Trial 4 column.
- Row 3 shows that there is one row in the source table whose one missing value is in the Trial 2 column.
- Row 4 shows that there is one row in the source table whose two missing values are in the Trial 2 and Trial 4 columns.
- Row 5 shows that there are two instances where Trial1 column have two missing values.
- Row 6 shows that there is one row in the source table whose two missing values are in the Trial 2 and Trial 4 columns.

### 4.7 Generating Random Data

**Steps:** Open a New Data Table from JMP Starter Window. Go to **Rows** → **Add Rows** or one can right click in rows area or go to hotspot button. Add 20 rows. Right Click on **Column1** → **Formula...** See Figure 4.7.1. In JMP Formula Editor Window, from **Functions (grouped)**, scroll down to **Random**. Here, we will select **Random Normal** (there are many distributions to choose from). Click **OK**. JMP will populate the new column with simulated standard normal data. See the Standard Normal data filled in Column1 column shown to the right of Figure 4.7.1.

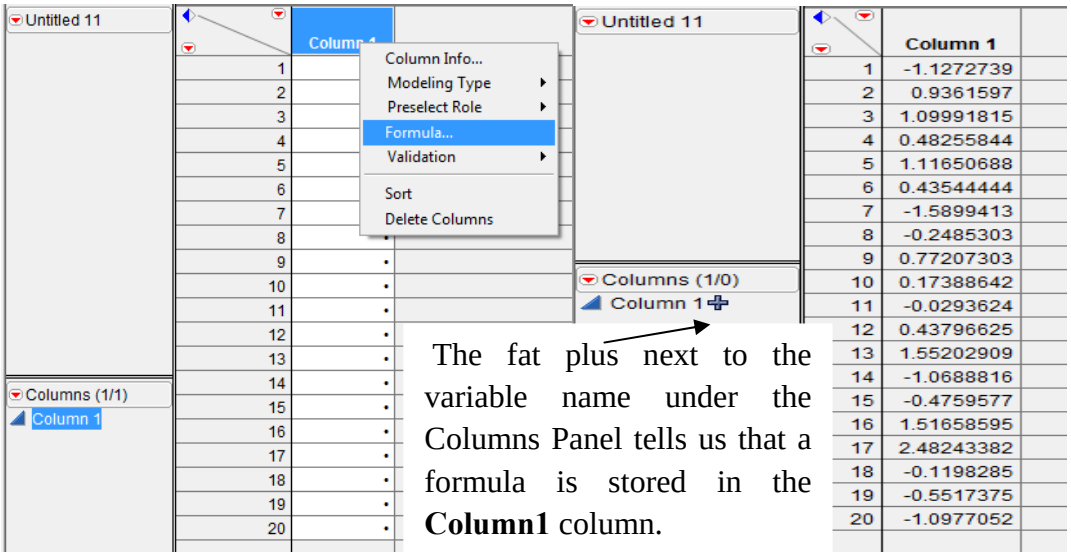



Figure 4.7.1: New Data Table

### 4.8 Selecting specific rows

Go to **Rows → Row Selection** and one can see various options available as shown in Figure 4.8.1. There are various options available which are discussed below:

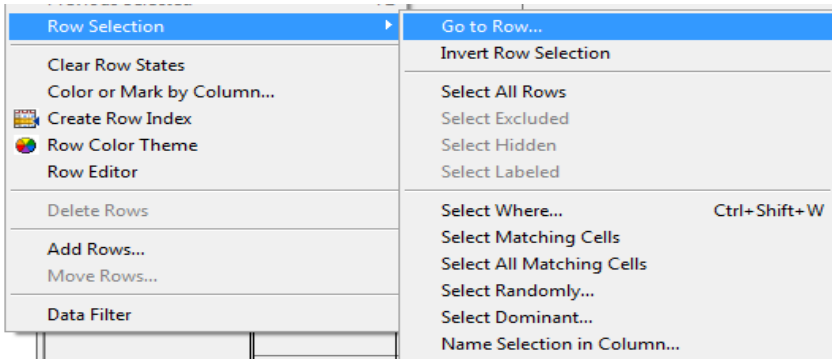



Figure 4.8.1: Row Selection options

1. **Go to Row... :** Go to specific row number.
2. **Invert Row Selection:** Select all deselected rows in data table.
3. **Select All Rows:** Selects all rows in data table.
4. **Select Excluded:** Selects all excluded rows regardless of their current selection status and deselects any other previously selected rows
5. **Select Hidden:** Selects all hidden rows regardless of their current selection status and deselects any other previously selected rows
6. **Select Labeled:** Selects all labeled rows regardless of their current selection status and deselects any other previously selected rows
7. **Select Where...:** Select rows based on some criteria that one enter.
8. **Select Matching Cells:** Select rows in the active data table with values that are similar to the highlighted row(s).
9. **Select All Matching Cells:** Select rows in all open data table with values that are similar to the highlighted row(s).
10. **Select Randomly:** Selects rows randomly.
11. **Select Dominant:** Useful for Pareto charts: selects the high or low values for a column.
12. **Name Selection in Column:** Use current selection to add a column using 1s and 0s to indicate the selection.

# 5. Graphics in JMP

**5.1 Mosaic Plot** displays a mosaic bar chart for each nominal or ordinal response variable. A mosaic plot is a stacked bar chart where each segment is proportional to its group's frequency count. A two-way frequency table can be graphically portrayed in a mosaic plot. The plot is divided into small rectangles such that the area of each rectangle is proportional to a frequency count of interest.

**Steps:** Open **Big Class.jmp** data table from sample data folder. Go to **Analyze → Fit Y by X**. Put age to **Y, Response** and sex to **X, Factor**. See the Figure 5.1.1 for output.

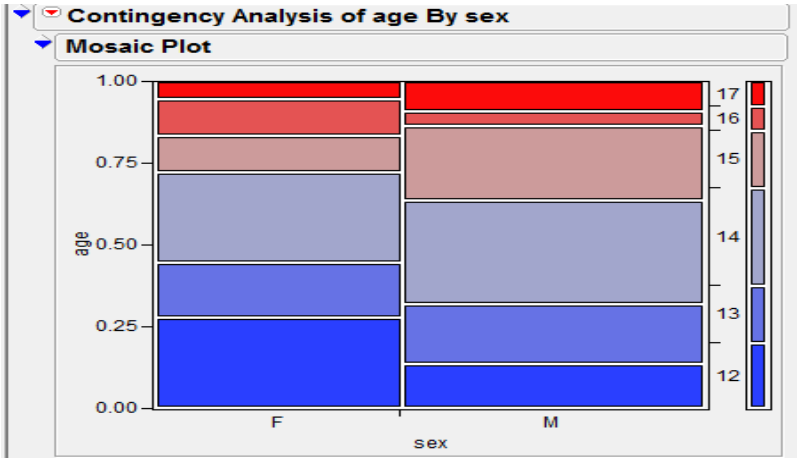



Figure 5.1.1: Contingency Analysis Mosaic Plot

**Note:** When one click on a section in the mosaic plot, the section is highlighted and corresponding data table rows are selected.

**5.2 Bar, Line and Pie Chart** Graph → Chart creates a bar for each level of a categorical X variable, and charts a count or statistic (if requested). One can have up to two X variables as categories. The X values are always treated as discrete values.

**Steps**: Open **Data_Mgmt.jmp** data table. Go to **Graph → Chart**. Select Antioxidant from **Select Columns** list, Click **Statistics**. There are various options on which one can plot analysis variable. If one want Standard Error Bars to be shown then Select **Mean** from this list. Error bars can be added to charts that involve means of variables. Check **Add Error Bars to Mean**. One can plot variety of Error bars like Range, Standard Error, Standard Deviation, Confidence Interval. As of now select Standard Error from this list. The number of standard errors can be specified. On X-axis If we want treatments to be shown within variety then select first **Variety** and then **Treatment** and take it to **Categories, X, Levels**. The Chart dialog box will look like as in Figure 5.2.1. The output is shown on the right of Figure 5.2.1.

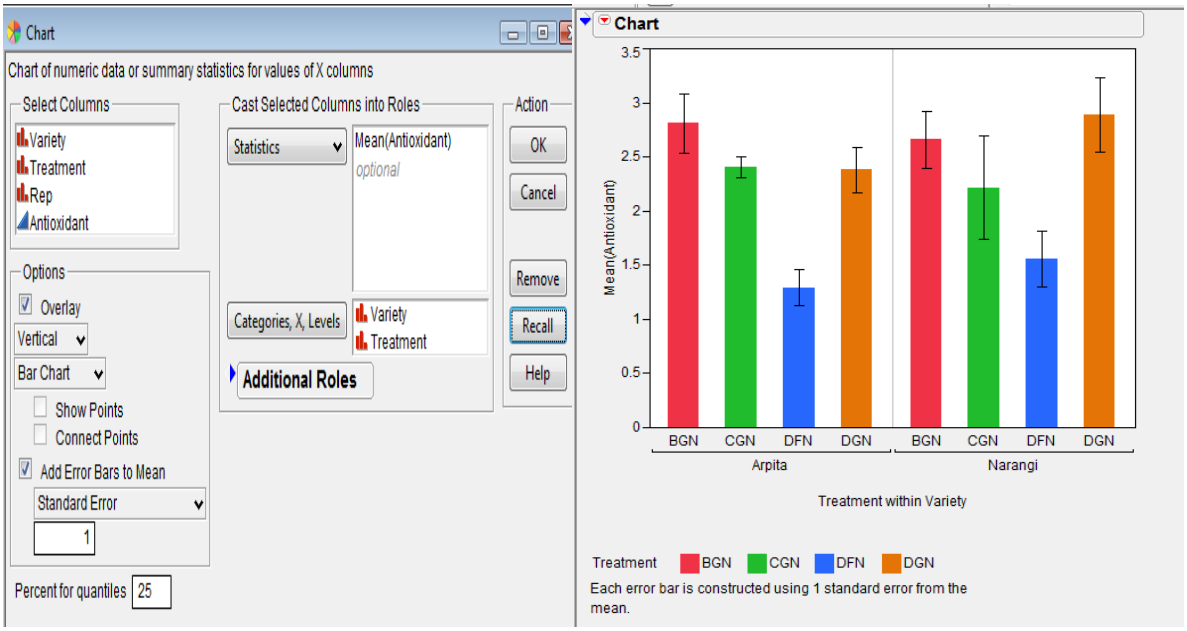



Figure 5.2.1: Chart Dialog Box and Bar Chart output with Standard error bars.

One can place the legend inside the Chart using Grabber tool . One can also change the legend settings by double-clicking on the legend. Right click on the chart, select **Chart Options → Line Chart** and one can see the output of Line chart with Standard error bars in Figure 5.2.2. Go to the hot spot button of Chart and select **Stack Bars**. Again go to hot spot button select **Label options → Label by percent of total value**. See the output on the right of Figure 5.2.2.

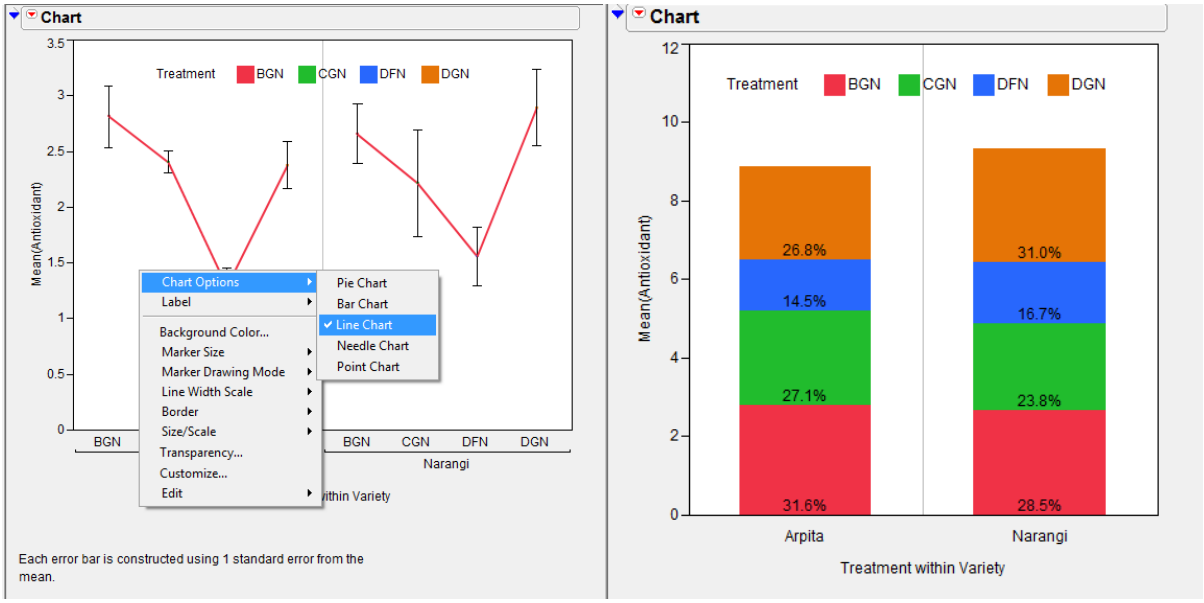



Figure 5.2.2: Line Chart with standard error bars and Stacked Bar Chart

One can also change the Stacked bar chart to Pie chart by going to the hotspot button and selecting the required option.

**5.2 Contour Plot** constructs contours of a response in a rectangular coordinate system. To create a contour plot, one need two variables for the x- and y-axes and at least one more variable for contours (although optionally, one can have several y-variables).

**Steps**: Open the **Little Pond.jmp** data table in the Sample Data folder. Go to **Graph** → **Contour Plot**. For this example, the coordinate variables X and Y → **X** role for the plot, and Z (which is the pond depth) → **Y**, (contour variable) → **OK**. See Figure 5.2.1 for launch dialog box of contour plot. The function Z should be a function of exactly two variables. Those variables should be the x-variables entered in the contour plot launch dialog. See Figure 5.2.2 for the output.

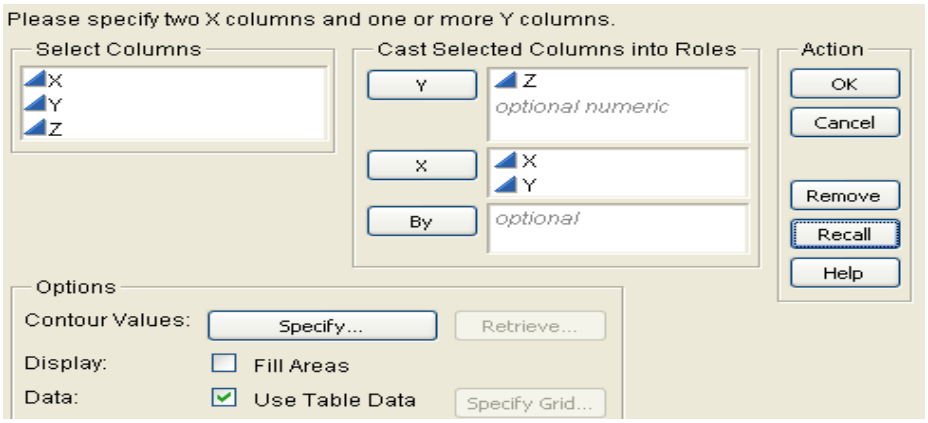



Figure 5.2.1: Launch Dialog for Contour Plot

By default, the contour levels are values computed from the data. One can specify own number of levels and level increments with options **Contour Values: Specify** in the Contour Plot Launch dialog, or options in the hotspot button on the Contour Plot title bar **Change Contours** → **Specify Contours**. One can also label the contours from the hotspot button **Label Contours**.

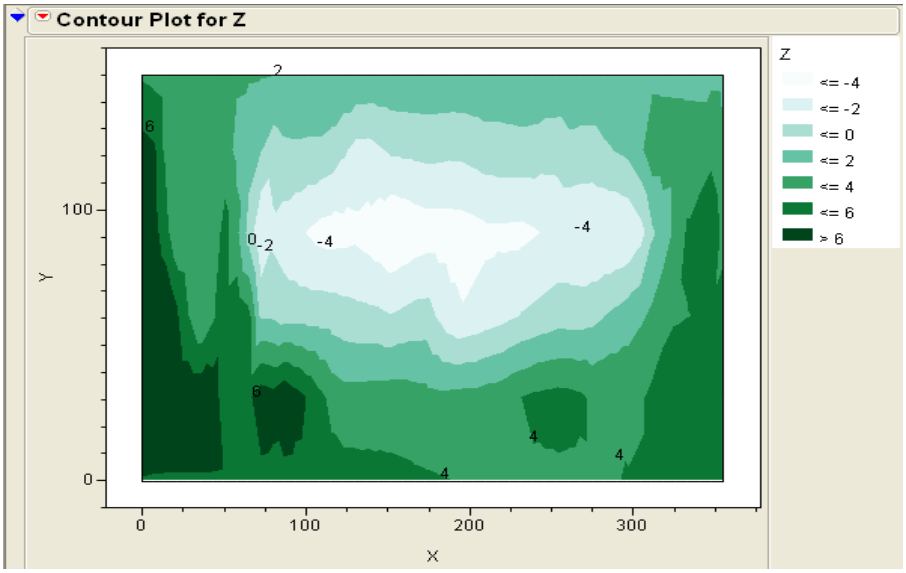



Figure 5.2.2 Contour Plot Output with legend and contour labels

**5.3 Overlay Plot** produces plots of a single X column and one or more numeric Y's. The Overlay Plot platform does not accept non-numeric variables for the y-axis.

**Steps:** Open S**pring.jmp** data table in the sample data folder. The column called April is the numeric day of the month, and the remaining columns are various weather statistics. Now go to **Graph → Overlay Plot**. The values in the column called April are the days of the month. Put **April → X** role. Daily humidity measures at 1:00 PM and 5:00 PM, **Humid1:PM** and **Humid5:PM** are assigned as → **Y** variables. The **Sort X** option causes the points to be connected in order of ascending X-values. Otherwise, the points are connected in row order. Click **OK**. See Figure 5.3.1 for the output.

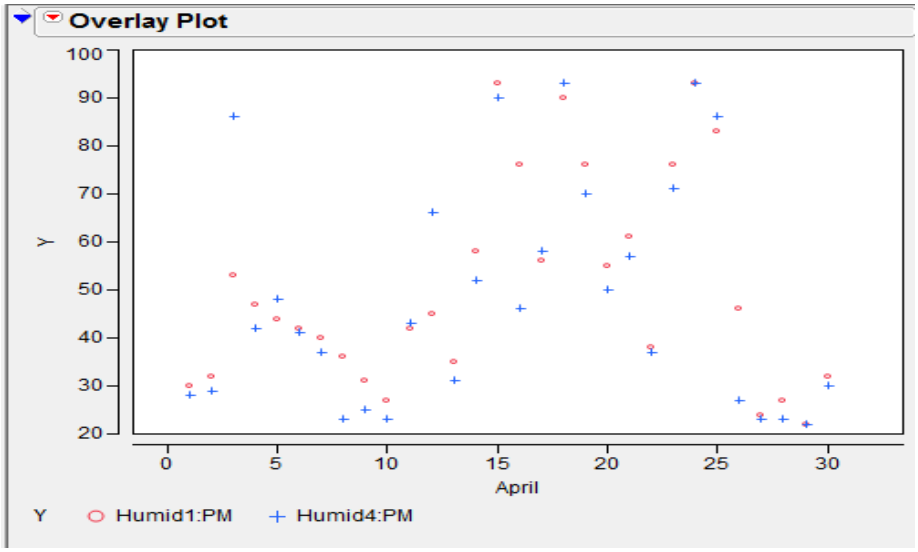



Figure 5.3.1: Overlay Plot with legend

Go to hot spot button of Overlay Plot and select **Y Options → Connect Points**. Adjacent points are connected for each Y variable (See Figure 5.3.2). One can also try various options such as **Needle**, **Step** and **Range plot**.

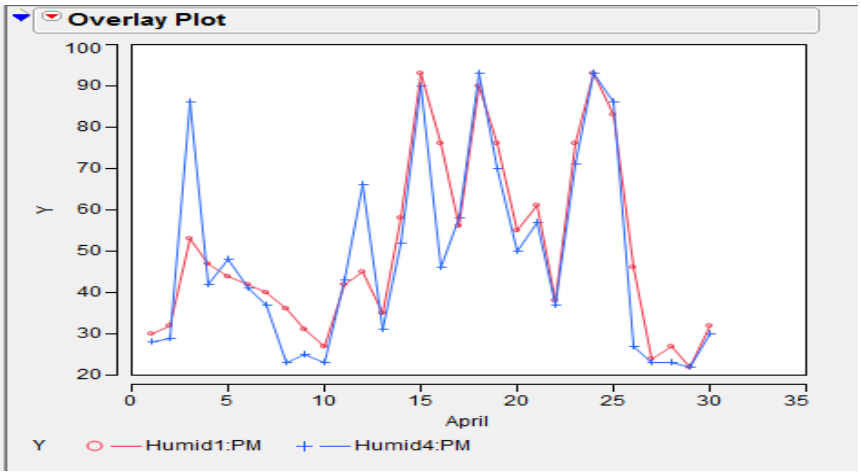



Figure 5.3.2: Overlay Plot with Connected Points

- **Range Plot** connects the lowest and highest points at each x value with a line with bars at each end.
- **Needle** draws a vertical line from each point to the x-axis.
- **Step** joins the position of the points with a discrete step by drawing a horizontal line from each point to the x value of the following point, and then a vertical line to that point.
- **Function Plot** plots a formula (stored in the Y column) as a smooth curve. To use this function, store a formula in a column that is a function of a single X column. Assign the formula to the Y role.

Now, Go to hot spot button **Overlay Plot → No Overlay**. This option does not overlay any Y's or groups. It creates a separate plot for each Y and Group.
**Separate Axes** lets one associate each plot with its own set of XY-axes. If Separate Axes is off, the vertical axis is shared across the same row of plots and the horizontal axis is shared on the same column of plots. See Figure 5.3.3 for the output.

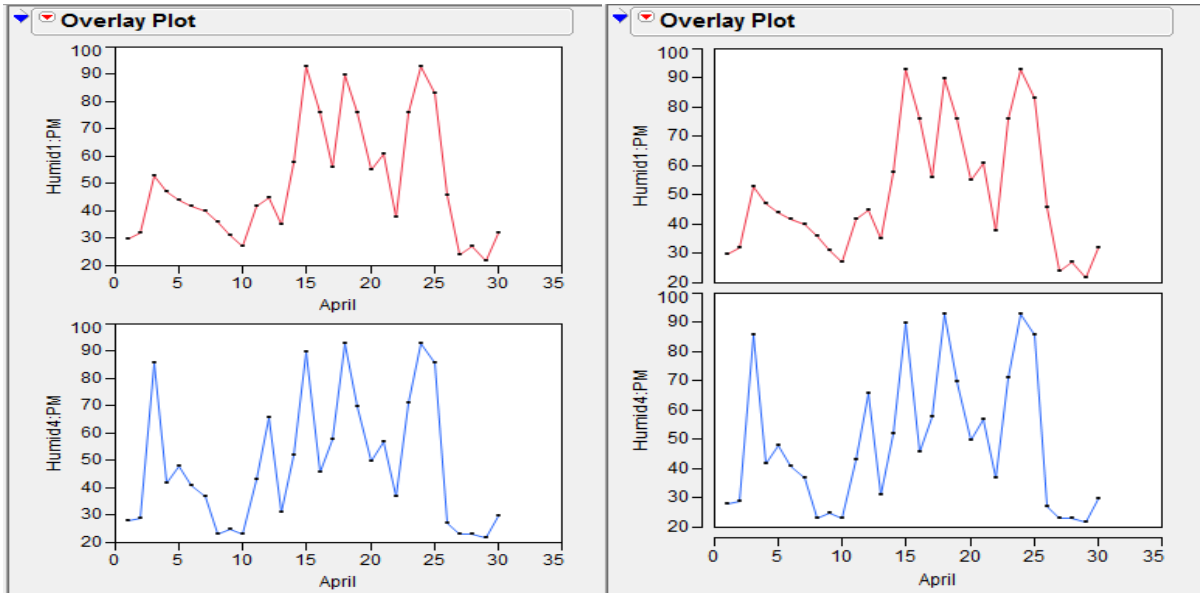



Figure 5.3.3: Separate X axis and Shared X axis Overlay plot.

**Plotting Two or More Variables with a Second Y-axis**

Go to **Graph → Overlay plot**. Assign Humid1:PM, Humid5:PM and Temp as **Y** role. Select **Temp** and click **Left Scale/Right Scale** tab. Humid1:PM & Humid5:PM variables are given as left axis and Temp variable is given right axis to display in plot. The screen will look as shown to the left of Figure 5.3.4. Click **OK**. See the output shown to the right of Figure 5.3.4

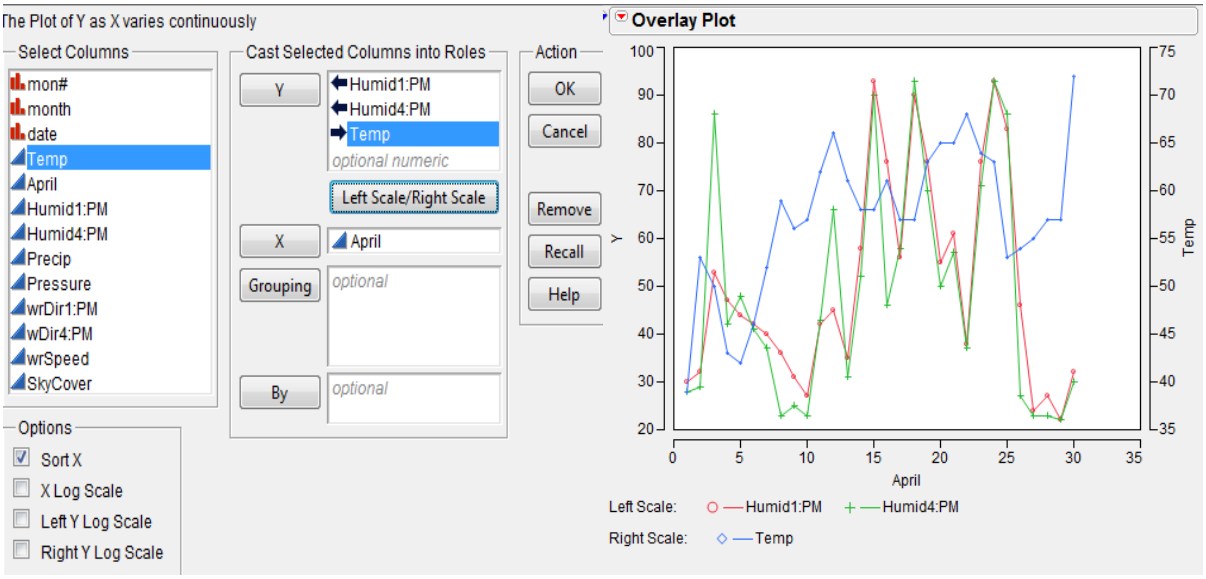



Figure 5.3.4: Overlay Plot Dialog box & Overlay plot output.

If one wants to connect points only for one of the variables say, Temp then go to the bottom of Overlay Plot in the area Right Scale and Right-click **Temp** and click **Connect Points**. One

can see the format of temp variable. It changed to Bold and Italic format. The legends can be repositioned with the hand tool .

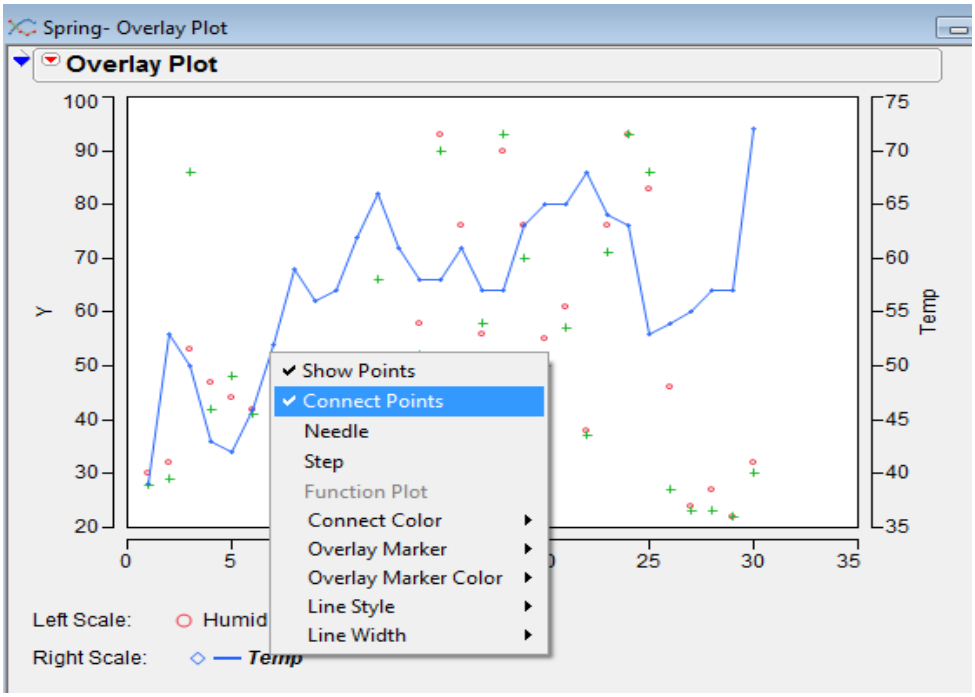



Figure 5.3.6: Overlay Plot output of connected points for Temp variable only

**5.4 Bubble Plot** is a scatter plot which represents its points as circles (bubbles). It displays the plot upto five dimensions at once (x position, y position, size, color, and time).

The following roles are used to generate the bubble plot.

- **Y, X** columns become the (x, y) coordinates of the bubbles in the plot. These values can be continuous or categorical, where the bubbles are positioned by the category indices.
- **Sizes** controls the size of the bubbles. The area of the bubble is proportional to the **Size** column's value. If **Size** is left blank, the default bubble size is proportional to the number of rows in that combination of **Time** and **ID**.
- **ID** variables are optional and used to identify rows that should be aggregated and displayed as a single bubble. The default coordinates of each bubble are the averaged x- and y-values, and the default size of each bubble is the sum of the sizes of all aggregated members. A second **ID** variable provides a hierarchy of categories, but the bubbles are not split by the second category until they are selected and split interactively. If a second **ID** variable is specified, then **Split** and **Combine** buttons appear on the graph for this use. For example, one may specify a country as the first **ID** variable, resulting in a separate aggregated bubble for each country. A second **ID** variable, perhaps designating regions within each country, would further split each country when the interactive **Split** button under the graph is pressed.
- **Time** columns cause separate coordinates, sizes, and colors to be maintained for each unique time period. The bubble plot then shows these values for a single time period.
- **Coloring** causes the bubbles to be colored. If the **Coloring** variable is categorical, each category is colored distinctly. If the **Coloring** variable is numeric, a continuous gradient of colors from blue to grey to red is used.

The Bubble Plot platform is used in one of two modes
• *static mode*, when one specify only **X**, **Y**, and **Size** variables.
• *dynamic mode*, where **Time** and **ID** variables are additionally specified.
Open **SATByYear.jmp** data table in the sample data folder. Go to **Graph → Bubble Plot**. Select **ACT Score (2004) → Y**, **SAT Math → X** and **Region & State → ID**. See Figure 5.4.1 for the Buubel plot output with splitting and combining a bubble.

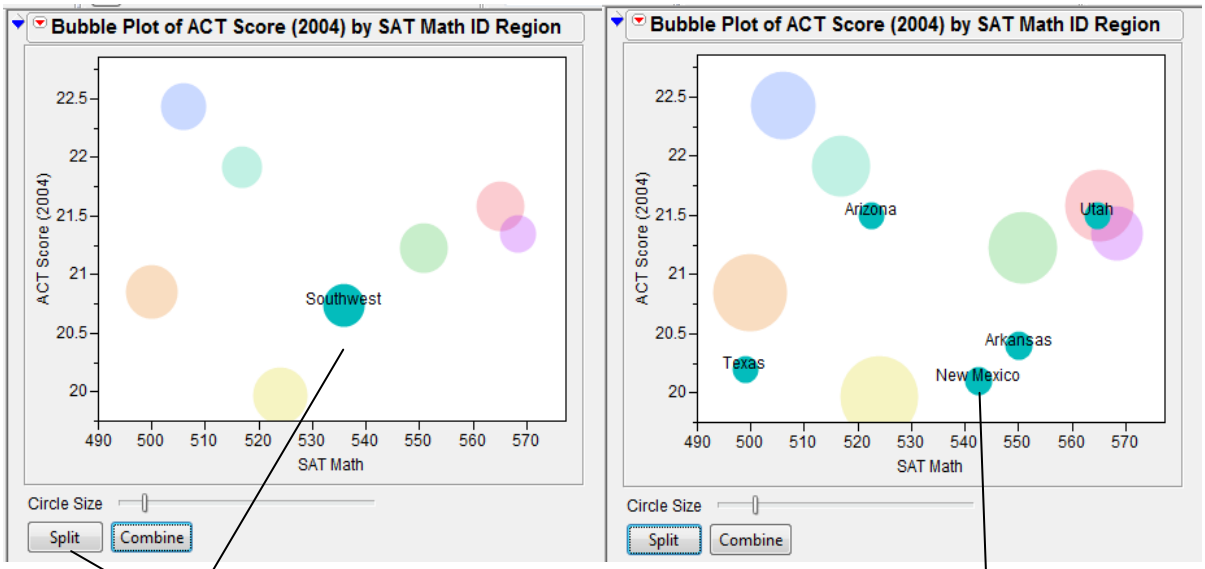



Figure 5.4.1: Splitting a Bubble

Select this bubble and click Split

The bubble vanishes and explodes into its second ID variable

When one click **Combine**, the splitted bubbles again get combined into one single bubble (as per first **ID** variable).

**Static Example:** Go to **Graph → Bubble Plot**. Put **SAT Verbal → Y** and **SAT Math → X** column. Take **State → ID** and **% Taking (2004) → Sizes**.

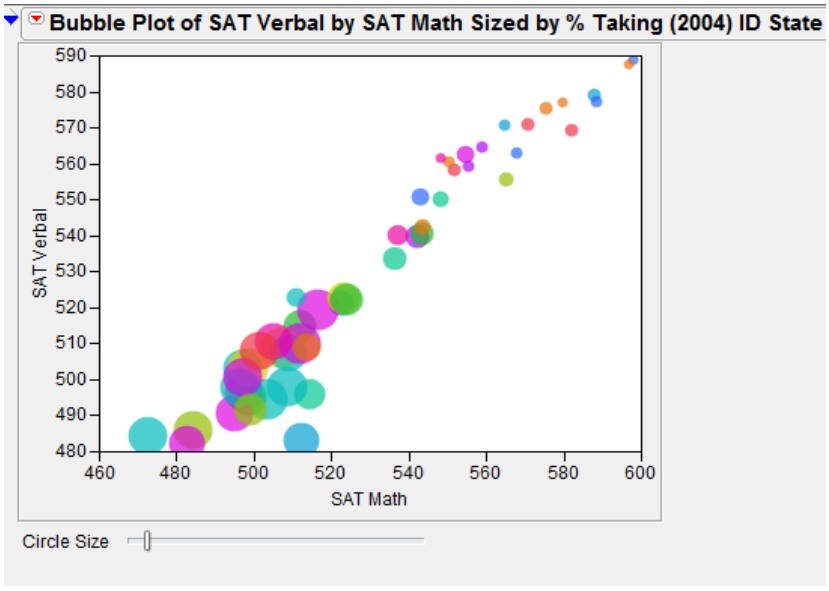



Figure 5.4.2: Static Bubble plot output

From the above graph (See Figure 5.4.2) one can see that

- there is a strong correlation between verbal and math scores
- States that have a large **% Taking** (large bubbles) are grouped together in the lower left of the graph.

This data can also be shown according to **Region**, only revealing the state-level information when needed. The following graph (See Figure 5.4.3) is identical to the previous one, except there are two **ID** variables: **Region**, then **State**.

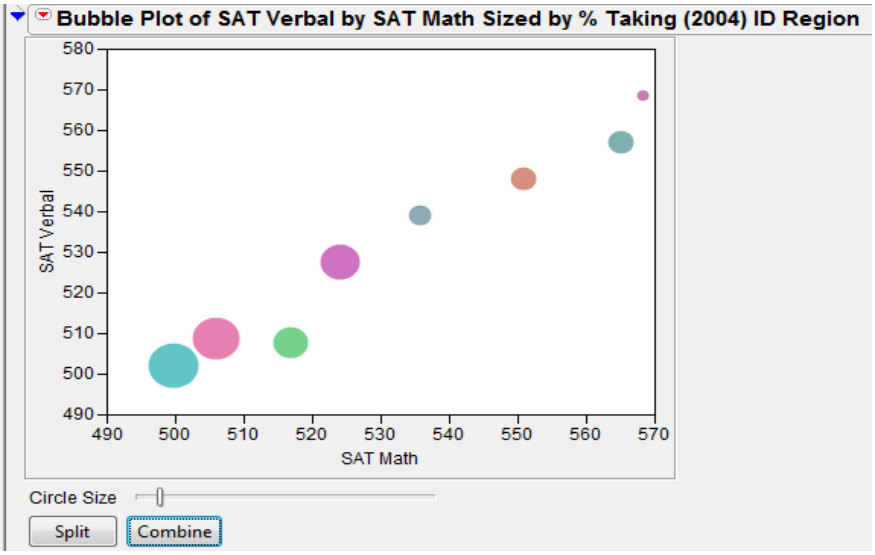



Figure 5.4.3: Static Bubble plot Output with two ID variables

**Dynamic Example** Assign the variables as in following dialog box See Figure 5.4.4

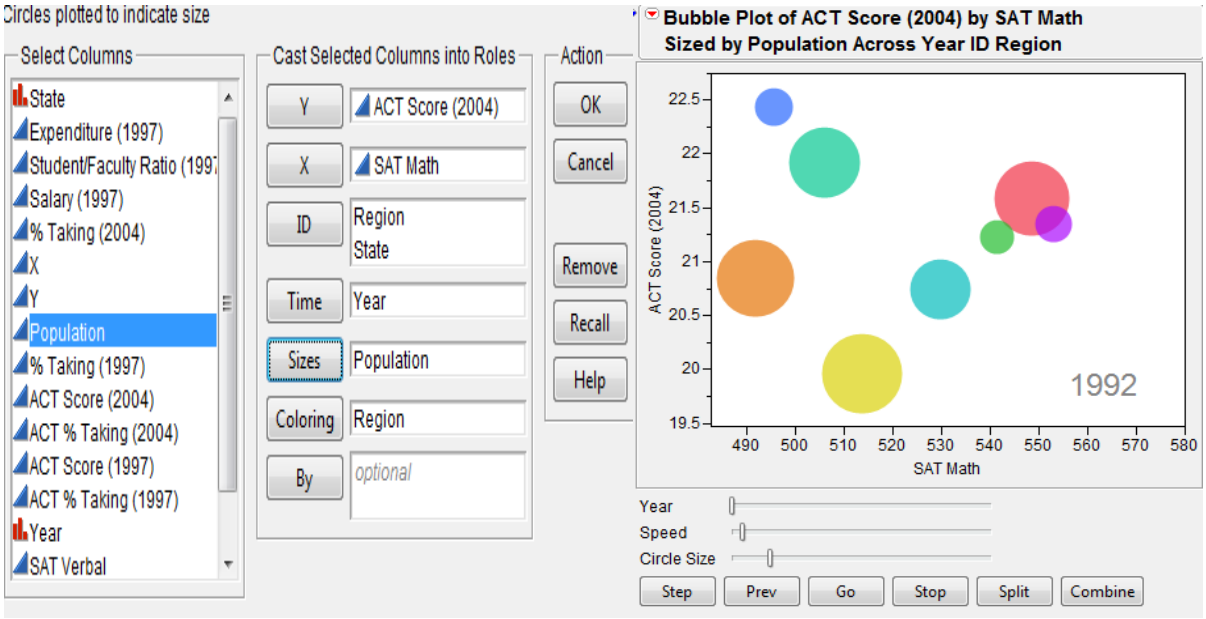



Figure 5.4.4: Dynamic Bubble plot dialog box and output

When one click **OK**, the following report appears See on the right of Figure 5.4.4. Since there is a Time variable, controls appear at the bottom of the report for progressing through the time levels. Click **Step** to move forward by one year. Click **Go** to see the animated, dynamic report. It's interesting to note that as time progresses, the SAT Math Score increases.

## 5. Statistical Analysis through JMP

All the examples and exercise are taken from "Analysis of data" (http://www.iasri.res.in/design/Analysis%20of%20data/Analysis%20of%20Data.html) and also see Design Resources Server (http://www.iasri.res.in/design/)

### 5.1 Descriptive Statistics

**Example 5.1:** An experiment was conducted to study the hybrid seed production of bottle gourd **(***Lagenaria siceraria (Mol) Standl)* Cv. Pusa hybrid-3 under open field conditions during Kharif-2005 at Indian Agricultural Research Institute, New Delhi . The main aim of the investigation was to compare natural pollination under field conditions. The data were collected on 10 randomly selected plants from each of natural pollination and hand pollination on number of fruit set for the period of 45 days, fruit weight (kg), seed yield per plant (g) and seedling length (cm). The data obtained is as given below:

| Group | No. of fruit Set (45days) | Fruit weight (kg) | Seed yield/plant (g) | Seedling length (cm) |
|---|---|---|---|---|
| 1 | 7.0 | 1.85 | 147.70 | 16.86 |
| 1 | 7.0 | 1.86 | 136.86 | 16.77 |
| 1 | 6.0 | 1.83 | 149.97 | 16.35 |
| 1 | 7.0 | 1.89 | 172.33 | 18.26 |
| 1 | 7.0 | 1.80 | 144.46 | 17.90 |
| 1 | 6.0 | 1.88 | 138.30 | 16.95 |
| 1 | 7.0 | 1.89 | 150.58 | 18.15 |
| 1 | 7.0 | 1.79 | 140.99 | 18.86 |
| 1 | 6.0 | 1.85 | 140.57 | 18.39 |
| 1 | 7.0 | 1.84 | 138.33 | 18.58 |
| 2 | 6.3 | 2.58 | 224.26 | 18.18 |
| 2 | 6.7 | 2.74 | 197.50 | 18.07 |
| 2 | 7.3 | 2.58 | 230.34 | 19.07 |
| 2 | 8.0 | 2.62 | 217.05 | 19.00 |
| 2 | 8.0 | 2.68 | 233.84 | 18.00 |
| 2 | 8.0 | 2.56 | 216.52 | 18.49 |
| 2 | 7.7 | 2.34 | 211.93 | 17.45 |
| 2 | 7.7 | 2.67 | 210.37 | 18.97 |
| 2 | 7.0 | 2.45 | 199.87 | 19.31 |
| 2 | 7.3 | 2.44 | 214.30 | 19.36 |

{Here 1 denotes natural pollination and 2 denotes the hand pollination}

1. Obtain mean, standard deviation, median, coefficient of skewness, coefficient of kurtosis, minimum and maximum values of all the characters.
2. Obtain mean, standard deviation, median, coefficient of skewness, coefficient of kurtosis, minimum and maximum values of all the characters for each group separately.
3. Test whether the data follows a normal distribution or not for all the characters? Do it separately for each of the two groups.
4. Prepare a discrete frequency table for all the characters for the above data on group.

5. Prepare 2-way frequency table between group and fruit set after 45 days.
6. Create a stem and leaf plot and box plot for all the characters.
7. Create a stem and leaf plot and box plot for all the characters for each group separately.
8. Create Normal Quantile Plot and CDF Plot for each characters

**Analysis using JMP**

Here Group is "classification variable".

1. First open the DESCRIPTIVE_STATS.xls file in JMP using **File → Open** dialog box.
2. The first step is to make sure that grouping factor/classificatory variable is nominal. Click on the blue triangle to the left of Factor (Group) and select nominal.
3. For Question 1, Go to **Analyze → Distribution**, (Select all the characters fs45, sw, syp and sl) from **Select Columns → Y, Columns → O**K. In the Output Window, click on the red triangle (hotspot) on the left of **Distributions** and Select **Stack**. Now click on the red triangle on the left of fs45 and select **Display Options → More Moments**. For all the characters repeat it for sw, syp and sl also and one will get the output as in Figure 5.1.1
4. For Question 2, Repeat Step 2 and in addition to it, Select group from **Select Columns → By** and one will get the descriptive statistics for each group seperately.
5. For Question 3, There are various ways to check the normality of the data. One way is go to the output of Step 3, click the hot spot button for each character and Go to **Continuous Fit → Normal →** (In the output window) **Fitted Normal → Goodness of Fit**. One can also get the **Diagnostic plot**. See the Figure 5.1.2 for Goodness of Fit test for variable fs45.
6. For Question 4, First change the scale of variables fs45, sw and syp from continuous to either nominal or ordinal. Go to **Analyze → Distribution**. Select all the characters fs45, sw, syp and sl from **Select Columns → Y, Columns → OK**. See the output in Figure 5.1.3
7. For Question 5, Go to **Analyze → Fit Y by X**. Select group from **Select Columns** box **→ X factor** and fs45 to **Y, Columns → OK**.
8. For Question 6 and 8, In the Output of Step 3, click the red triangle button of characters fs45, fw, syp and sl and select Stem and Leaf, Normal Quantile and CDF Plot option to get the required output.
9. For Question 7, Stem and Leaf plot for each group separately See Step 4 and do the require steps.

When one click any of blue triangle then it expand/contract the output.

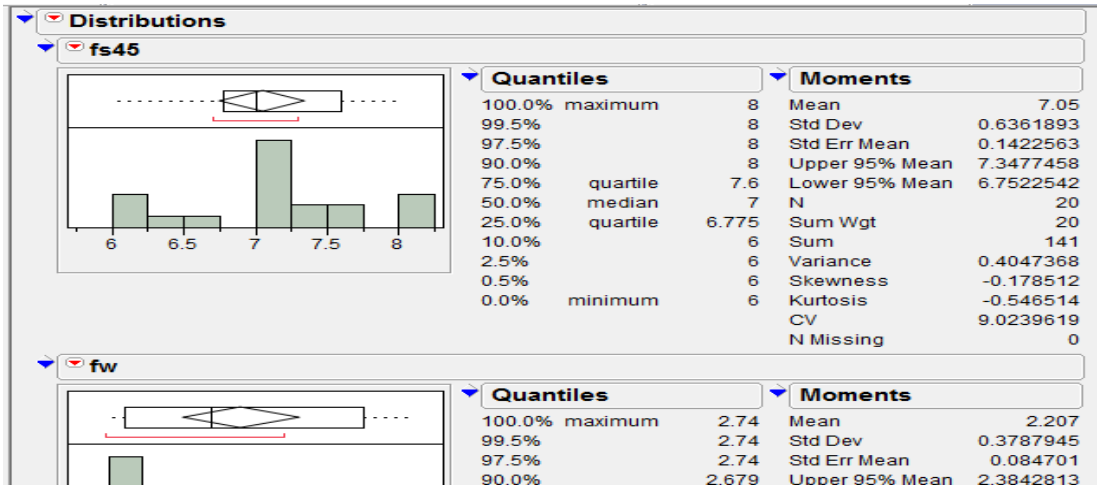



Figure 5.1.1: Descriptive Statistics Output

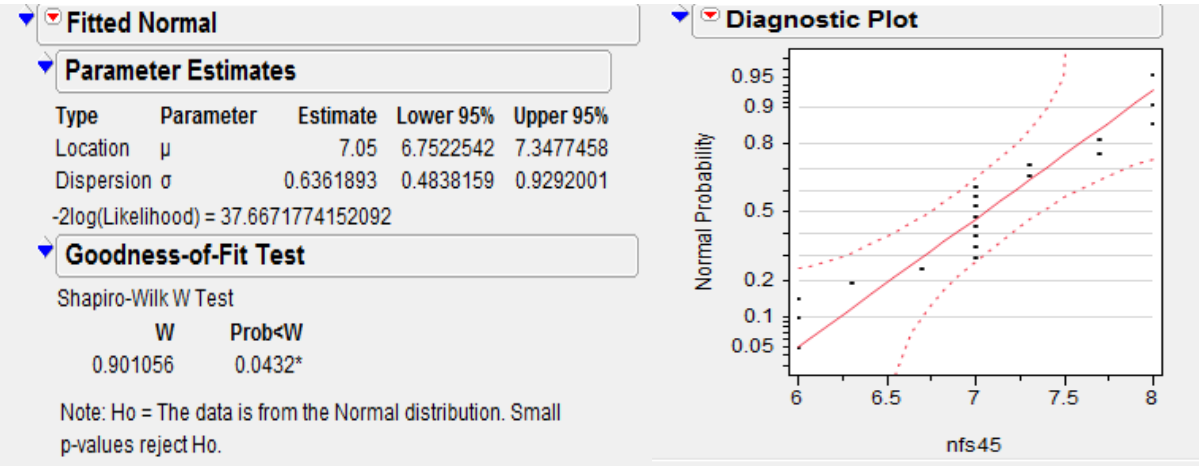
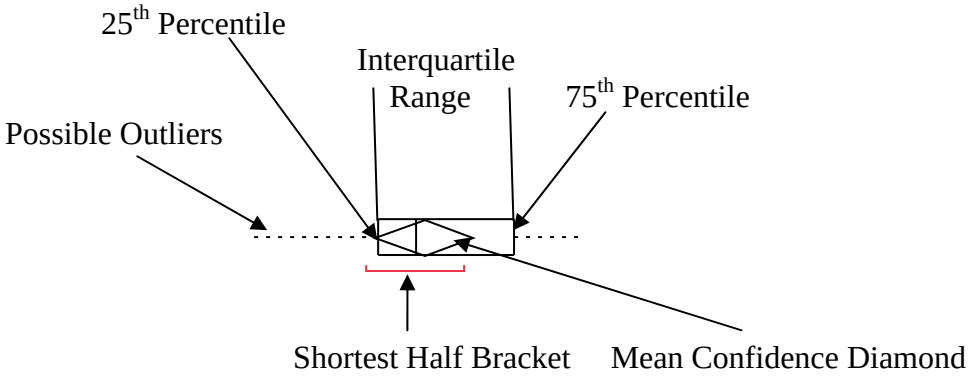



Figure 5.1.2: Normality check for character fs45

**Outlier Box Plot**




Right-clicking on the box plot reveals a menu that allows one to toggle on and off the **Mean Confid Diamond** and **Shortest Half Bracket**.

**Quantile Box Plot and Quantiles Table**

One can get the Quantile Box plot by clicking hotspot button in the output window for variable sl and select **Quantile Box Plot**.

| | | |
|---|---|---|
| 100.0% | maximum | 19.36 |
| 99.5% | | 19.36 |
| 97.5% | | 19.36 |
| 90.0% | | 19.286 |
| 75.0% | quartile | 18.9425 |
| 50.0% | median | 18.22 |
| 25.0% | quartile | 17.5625 |
| 10.0% | | 16.779 |
| 2.5% | | 16.35 |
| 0.5% | | 16.35 |
| 0.0% | minimum | 16.35 |

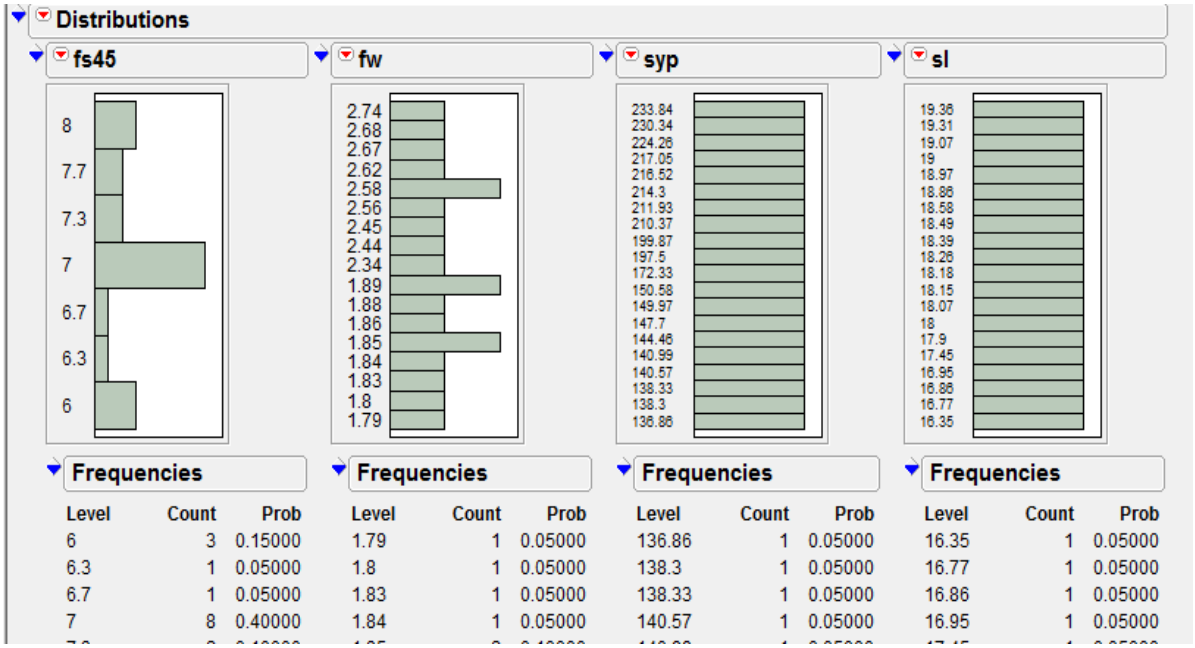



Figure 5.1.3: Discrete Frequency Table Output

One can always save the output by clicking **File → Save As** and select required format to save it as RTF file(*.rtf), Word document(*.doc), JPEG image(*.jpg) or HTML(*.html) file etc. as in Figure 5.1.4. By default JMP Output/Reports are saved with .jrp extension.

To copy a portion of a report, get the selection tool ( ) and click on the area(s) one want to select. Then goto Word document and paste it.

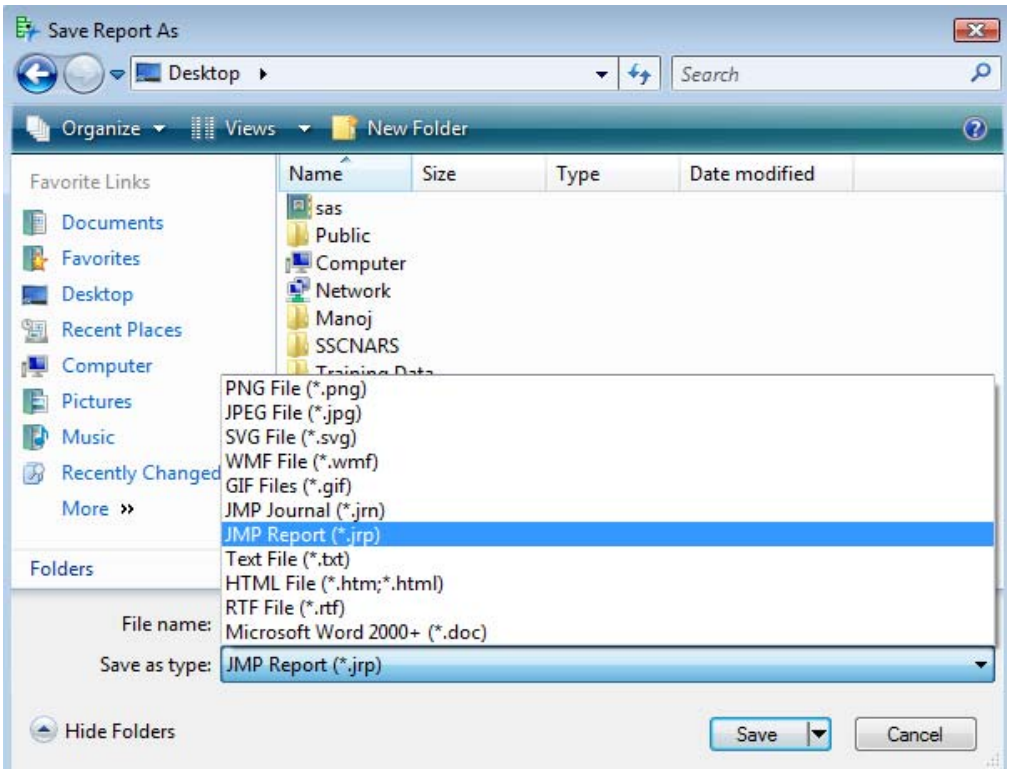



Figure 5.1.4: Save Report As dialog box

### 5.2 Tests of Significance Based on t - distribution

**Example 5.2:** Using the data from Example 5.1 test

1. Whether the mean of the population of Seed yield/plant (g) is 200 or not?
2. Whether the natural pollination and hand pollination under open field conditions are equally effective or are significantly different?
3. Whether hand pollination is better alternative in comparison to natural pollination?

**Analysis using JMP**

1. First open the TESTSIG.xls file in JMP using **File → Open** dialog box.
2. Make sure variable 'group' is nominal.
3. For Question 1, Go to **Analyze → Distribution**. Select the variable syp from **Select Columns → Y, Columns → OK**. Now in the output, click on the red triangle on the left of **syp** and select **Test Mean** option. In the Test mean dialog box, put 200 as specify hypothesized mean. If one wants the Nonparametric **Wilcoxon Signed Rank** test then check the required box. See the Figure 5.2.1 for output.
4. For Question 2, Go to **Analyze → Fit Y by X**. Put nfs45, fw, syp and sl variables into **Y, Response** and group **→ X, Factor** box **→ OK**. First we test that whether variances are equal or not? So in the output window, click the hotspot button on the left of **Oneway Analysis** and select **UnEqual Variances**. (See the result shown to the left of Figure 5.2.2). Now If variances are equal then we select **Means/Anova/Pooled t** to get independent samples t test. For a test without the assumption of equal variances i.e. Assuming unequal variances select **t test** instead from hotspot button of Oneway Analysis. See the output shown to the right of Figure 5.2.2.
5. To answer the Question 3, one has to perform the one tail t-test. The easiest way to convert a two-tailed test into a one-tailed test is take half of the p-value provided in the output of 2-tailed test output for drawing inferences.

**Output:**

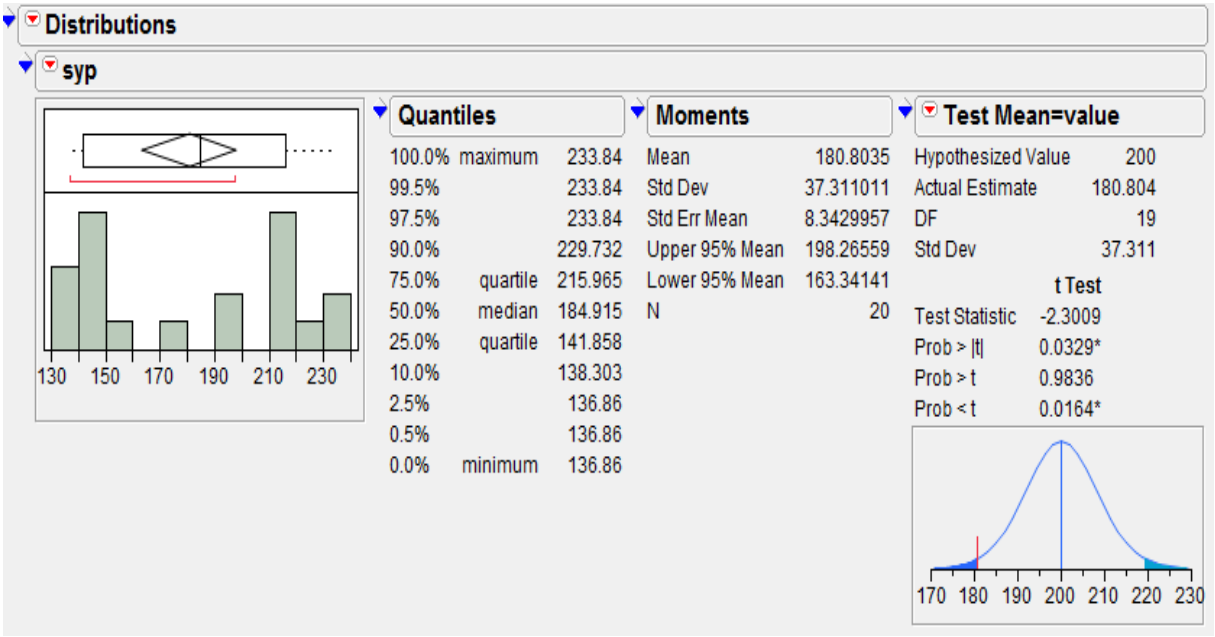



Figure 5.2.1: One Sample t-test output

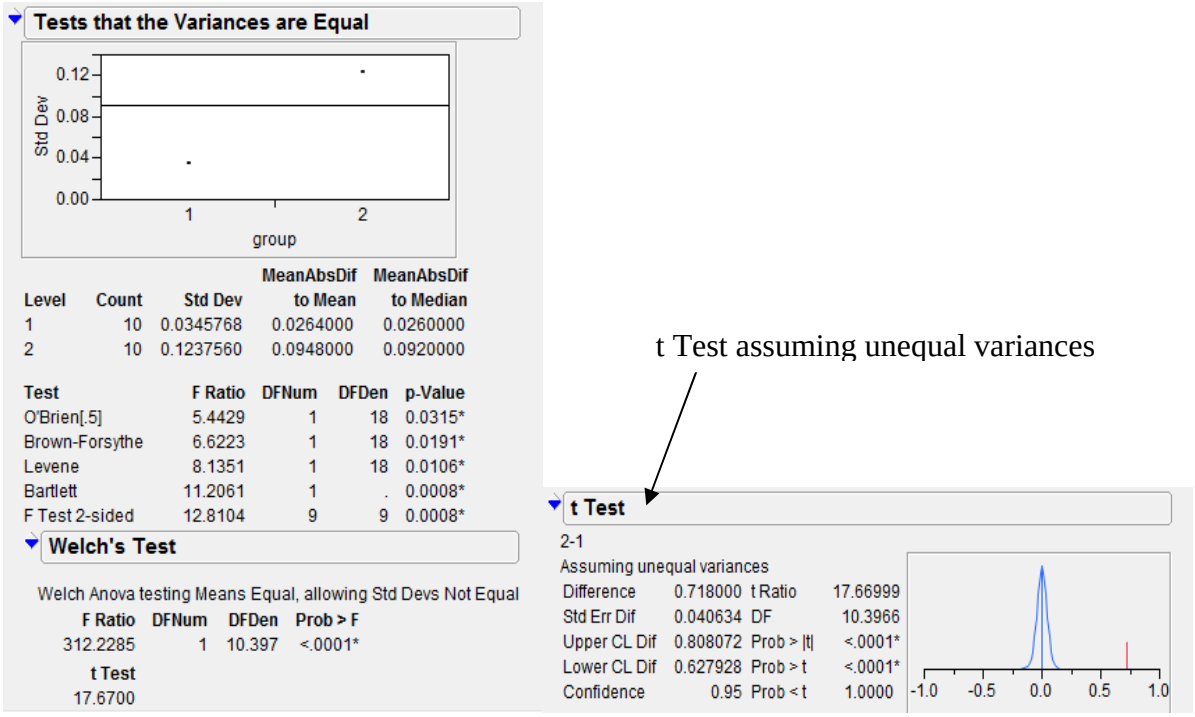



Figure 5.2.2: Two Sample t-test output

**Note:** If one wants Sum of Squares in ANOVA table to be displayed upto four decimal points then double click in the output of Sum of Squares. A **Column Numeric Format** dialog box appears as in Figure 5.2.3. The **Format** is Fixed Dec. Change the **Dec** to four and Click **OK**.

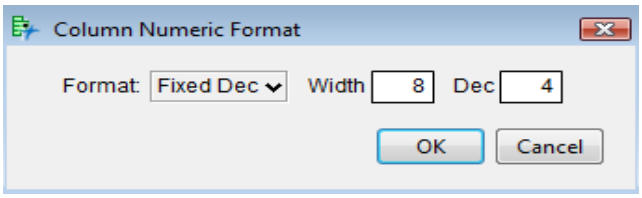



Figure 5.2.3: Column Numeric Format Dialog box

Also one can sort the columns in the output. Right click in the output area and Select **Sort by Column...** It displays all the numeric columns on which one can sort in ascending or descending order.. See Figure 5.2.4.

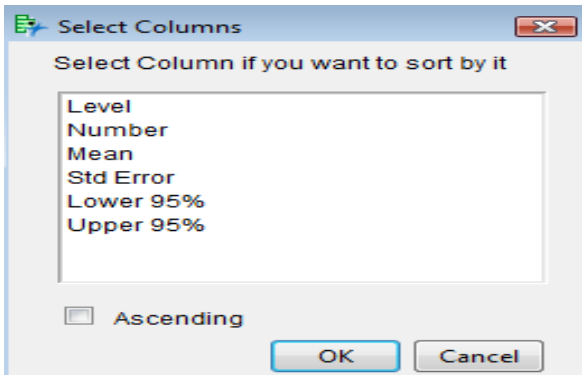



Figure 5.2.4: Sorting columns in the output.

### 5.3 Correlation and Regression

**Example 5.3:** The following data was collected through a pilot sample survey on Hybrid Jowar crop on yield and biometrical characters. The biometrical characters were average Plant Population (PP), average Plant Height (PH), average Number of Green Leaves (NGL) and Yield (kg/plot).

| S.No. | PP | PH | NGL | Yield | S.No. | PP | PH | NGL | Yield |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 142.00 | 0.525 | 8.2 | 2.470 | 24 | 55.55 | 0.265 | 5.0 | 0.430 |
| 2 | 143.00 | 0.640 | 9.5 | 4.760 | 25 | 88.44 | 0.980 | 5.0 | 4.080 |
| 3 | 107.00 | 0.660 | 9.3 | 3.310 | 26 | 99.55 | 0.645 | 9.6 | 2.830 |
| 4 | 78.00 | 0.660 | 7.5 | 1.970 | 27 | 63.99 | 0.635 | 5.6 | 2.570 |
| 5 | 100.00 | 0.460 | 5.9 | 1.340 | 28 | 101.77 | 0.290 | 8.2 | 7.420 |
| 6 | 86.50 | 0.345 | 6.4 | 1.140 | 29 | 138.66 | 0.720 | 9.9 | 2.620 |
| 7 | 103.50 | 0.860 | 6.4 | 1.500 | 30 | 90.22 | 0.630 | 8.4 | 2.000 |
| 8 | 155.99 | 0.330 | 7.5 | 2.030 | 31 | 76.92 | 1.250 | 7.3 | 1.990 |
| 9 | 80.88 | 0.285 | 8.4 | 2.540 | 32 | 126.22 | 0.580 | 6.9 | 1.360 |
| 10 | 109.77 | 0.590 | 10.6 | 4.900 | 33 | 80.36 | 0.605 | 6.8 | 0.680 |
| 11 | 61.77 | 0.265 | 8.3 | 2.910 | 34 | 150.23 | 1.190 | 8.8 | 5.360 |
| 12 | 79.11 | 0.660 | 11.6 | 2.760 | 35 | 56.50 | 0.355 | 9.7 | 2.120 |
| 13 | 155.99 | 0.420 | 8.1 | 0.590 | 36 | 136.00 | 0.590 | 10.2 | 4.160 |
| 14 | 61.81 | 0.340 | 9.4 | 0.840 | 37 | 144.50 | 0.610 | 9.8 | 3.120 |
| 15 | 74.50 | 0.630 | 8.4 | 3.870 | 38 | 157.33 | 0.605 | 8.8 | 2.070 |
| 16 | 97.00 | 0.705 | 7.2 | 4.470 | 39 | 91.99 | 0.380 | 7.7 | 1.170 |
| 17 | 93.14 | 0.680 | 6.4 | 3.310 | 40 | 121.50 | 0.550 | 7.7 | 3.620 |
| 18 | 37.43 | 0.665 | 8.4 | 1.570 | 41 | 64.50 | 0.320 | 5.7 | 0.670 |
| 19 | 36.44 | 0.275 | 7.4 | 0.530 | 42 | 116.00 | 0.455 | 6.8 | 3.050 |
| 20 | 51.00 | 0.280 | 7.4 | 1.150 | 43 | 77.50 | 0.720 | 11.8 | 1.700 |
| 21 | 104.00 | 0.280 | 9.8 | 1.080 | 44 | 70.43 | 0.625 | 10.0 | 1.550 |
| 22 | 49.00 | 0.490 | 4.8 | 1.830 | 45 | 133.77 | 0.535 | 9.3 | 3.280 |
| 23 | 54.66 | 0.385 | 5.5 | 0.760 | 46 | 89.99 | 0.490 | 9.8 | 2.690 |

1. Obtain correlation coefficient between each pair of the variables PP, PH, NGL and yield.
2. Obtain partial correlation between NGL and yield after removing the linear effect of PP and PH.
3. Give a scatter plot of the variable PP.

4. Fit a multiple linear regression equation by taking yield as dependent variable and biometrical characters as explanatory variables. Print the matrices used in the regression computations.
5. Obtain the predicted values corresponding to each observation in the data set.
6. Check for the linear relationship among the biometrical characters, i.e., multi-colinearity in the data.
7. Fit the multiple linear regression model without intercept.

**Analysis using JMP**

1. First open the CORREG.xls file in JMP using **File → Open** dialog box.
2. Make sure variable 'sn' is nominal.
3. For Question 1, Go to **Analyze → Multivariate Methods → Multivariate**. Select all the characters PP, PH, NGL and YLD from **Select Columns → Y, Columns → OK**. See the correlation coefficient output as in Figure 5.3.1.
4. For Question 2, Click the hotspot button of **Multivariate → Partial Correlations**. The partial correlation between NGL & YLD after adjusting for PP & PH is 0.1920. See Figure 5.3.2. Click the hotspot button of **Scatterplot Matrix → Show Histogram → Horizontal**. See the output as shown to the left of Figure 5.3.3. One can also see the descriptive statistics by clicking red button of **Multivariate → Simple Statistics → Univariate Simple Statistics**.
5. For Question 3, Go to **Graph → Scatterplot Matrix**. Put PP in **Y, Columns** and YLD in **X**. See the output as shown to the right of Figure 5.3.3.
6. For Question 4, Go to **Analyze → Fit Model**. Put the response variable Yield into **Y, Response** box and **Add** PP, PH and NGL factor to **Construct Model Effects → Run Model**. If one wants the model without Intercept then check **No Intercept** in the Fit Model dialog box. See Figure 5.3.4 for output.
7. One can also save predicted values by clicking hot spot button of **Response Yield → Save Columns**. It will save the required values to the data table.
8. One can also get the Durbin-Watson Statistic by clicking hot spot button of **Response Yield → Row Diagnostics → Durbin Watson Test**. See Figure 5.3.5.

**Output:**

**Multivariate**

**Correlations**

| | pp | ph | ngl | yld |
|---|---|---|---|---|
| pp | 1.0000 | 0.2396 | 0.2853 | 0.3859 |
| ph | 0.2396 | 1.0000 | 0.0887 | 0.3323 |
| ngl | 0.2853 | 0.0887 | 1.0000 | 0.2788 |
| yld | 0.3859 | 0.3323 | 0.2788 | 1.0000 |

The correlations are estimated by REML method.

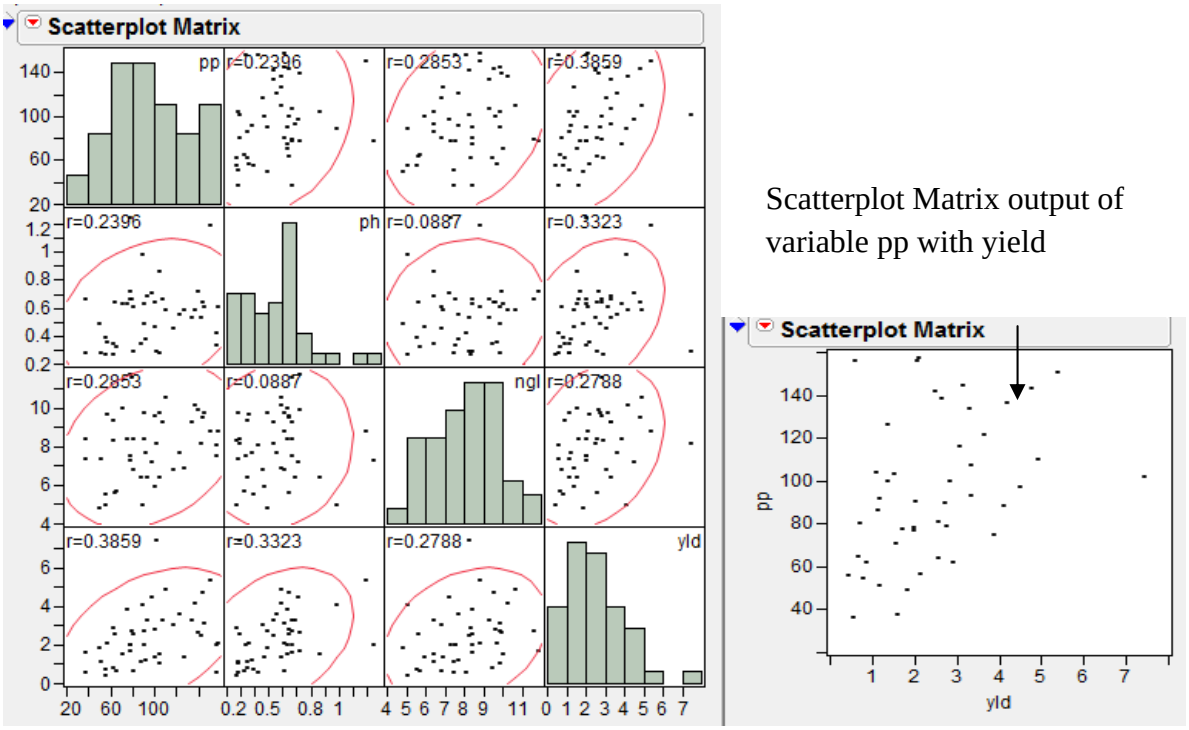
Figure 5.3.1: Correlation output of pp, ph, ngl and yield

**Partial Corr**

| | pp | ph | ngl | yld |
|---|---|---|---|---|
| pp | . | 0.1315 | 0.2028 | 0.2825 |
| ph | 0.1315 | . | -0.0309 | 0.2686 |
| ngl | 0.2028 | -0.0309 | . | 0.1920 |
| yld | 0.2825 | 0.2686 | 0.1920 | . |

partialed with respect to all other variables

Figure 5.3.2: Partial Correlation Output partialed with respect to all other variables




Scatterplot Matrix output of variable pp with yield

Figure 5.3.3: Scatterplot Matrix output

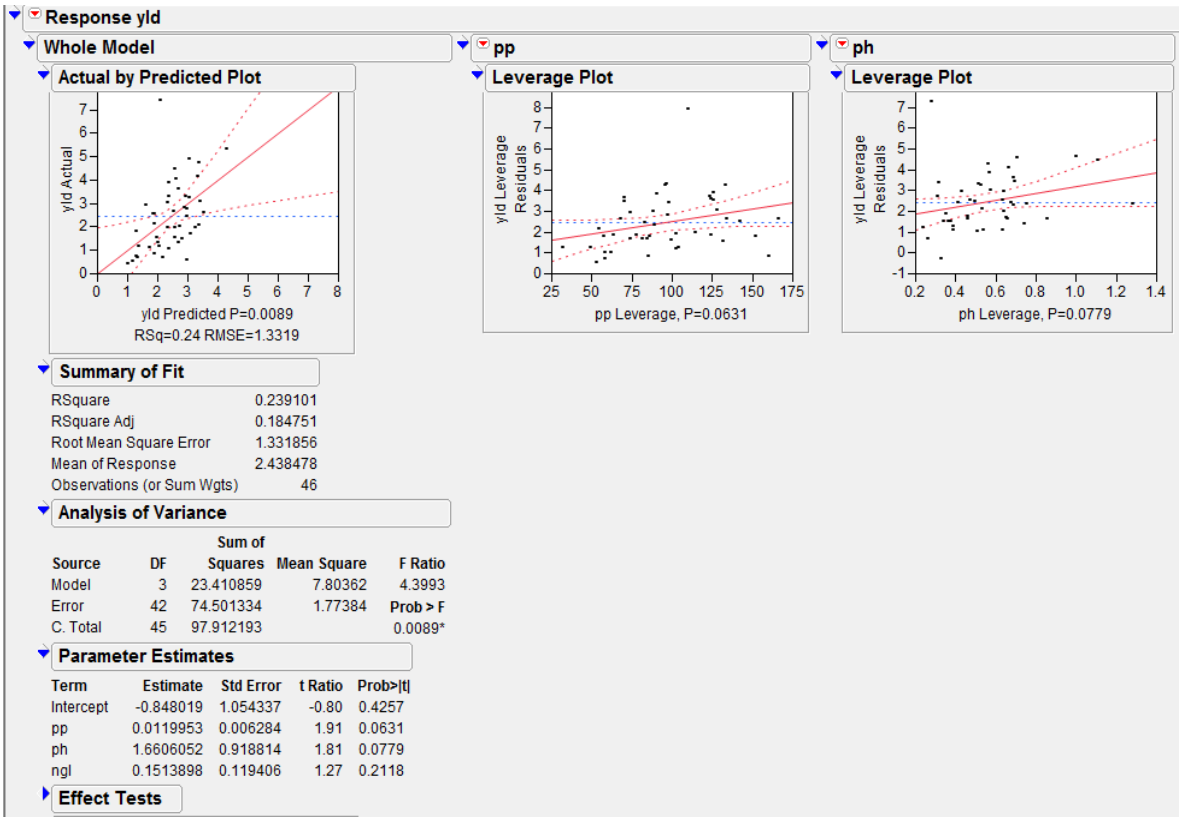



Figure 5.3.4: Multiple Regression Output

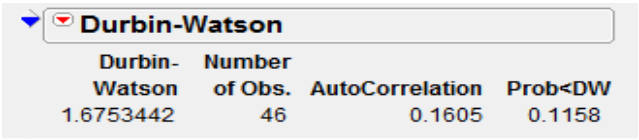



Figure 5.3.5: Durbin-Watson Statistic Output

## 5.4 Principal Components Analysis

**Example 5.4:** Use the data from Example 5.3 and Perform Principal Component Analysis to the variables PP, PH, NGL and yield..

**Analysis using JMP**

1. First open the Excel file PCA.xls using **File → Open** dialog box.
2. Go to **Analyze → Multivariate Methods → Principal Components**. Select the variables pp, ph, ngl and yield from **Select Columns** box → **Y, Columns**. See the output as in Figure 5.4.2.
3. One can see the Eigenvectors, Scree plot, Score plot, Loading plot and various other options available by clicking hot spot button of **Principal Components/Factor Analysis** and select the required option. See Figure 5.4.3 for output.
4. By Default, It computes principal component scores based on correlation matrix. If one wants to compute it, based on covariance matrix then go to hot spot button of **Multivariate → Principal Components → On Covariances.**
5. If one wants to save the principal component scores then go to the hotspot button of **Principal Components/Factor Analysis → Save Principal Components**. The uncorrelated principal component scores get saved in the current data table. See the output as in Figure 5.4.4.

The hotspot button in the output popup menu (See Figure 5.4.1) on the **Principal Components/ Factor Analysis** title bar gives us options to see the results of rotating principal components and saving principal component scores.

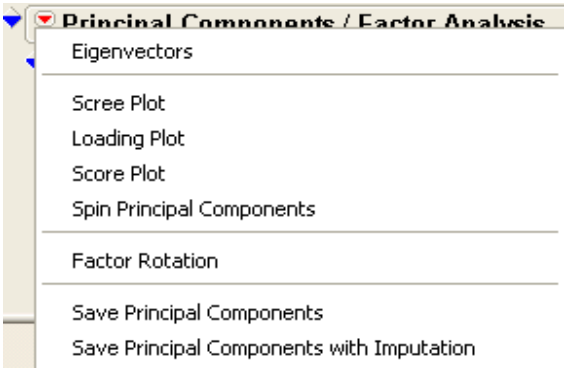



Figure 5.4.1 Popup Menu of Principal Components/Factor Analysis

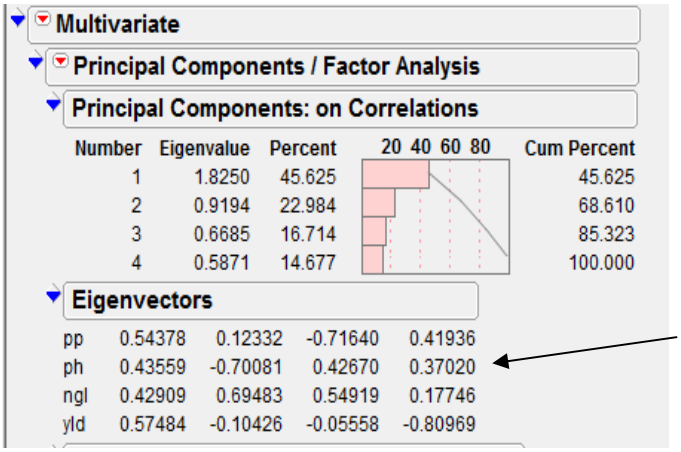
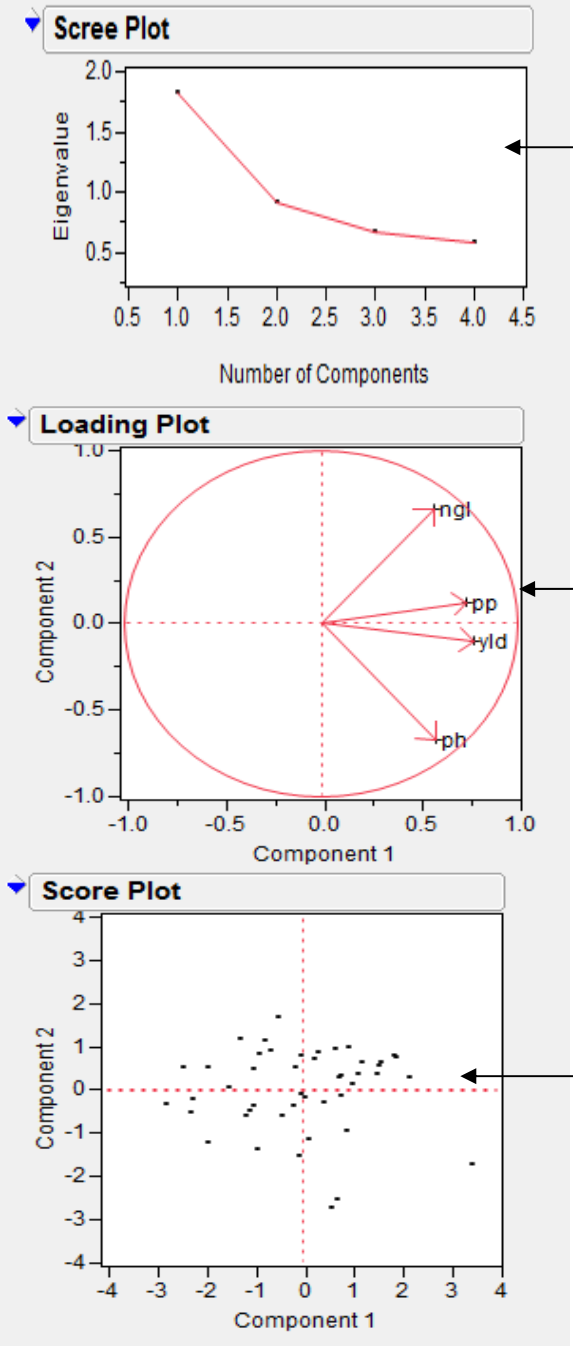



Figure 5.4.2: Principal Component Analysis Output

**Eigenvectors** shows columns of values that correspond to the eigenvectors for each of the principal components, in order, from left to right. Using these coefficients to form a linear combination of the original variables produces the principal component variables.




**Scree Plot** shows or hides a scree plot of the eigenvalues vs. the number of components. This plot is useful for visualizing the dimensionality of the data space.

**Loading Plot** shows or hides a matrix of two-dimensional representations of factor loadings. After being prompted for the number of loadings to show, a matrix of plots showing all possible pairs of these loadings appears.

**Score Plot** shows or hides a two-dimensional matrix representation of the scores for each pair of principal components. After being prompted for the number of principal components to show, a matrix of plots showing all possible pairs of these component scorings appears.

Figure 5.4.3: Scree Plot, Loading Plot and Score plot Output

| | sn | pp | ph | ngl | yld | Prin1 | Prin2 | Prin3 | Prin4 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 142 | 0.525 | 8.2 | 2.47 | 0.73604783 | 0.30871995 | -0.9984327 | 0.52494553 |
| 2 | 2 | 143 | 0.64 | 9.5 | 4.76 | 2.19102214 | 0.3090301 | -0.4740373 | -0.3954952 |
| 3 | 3 | 107 | 0.66 | 9.3 | 3.31 | 1.03691083 | 0.13723274 | 0.31805947 | -0.033055 |
| 4 | 4 | 78 | 0.66 | 7.5 | 1.97 | -0.3965837 | -0.5945362 | 0.41304368 | 0.15889639 |
| 5 | 5 | 100 | 0.46 | 5.9 | 1.34 | -1.0754135 | -0.4808776 | -0.9188731 | 0.28123782 |
| 6 | 6 | 86.5 | 0.345 | 6.4 | 1.14 | -1.4717505 | 0.04627343 | -0.6876058 | 0.08355872 |
| 7 | 7 | 103.5 | 0.86 | 6.4 | 1.5 | -0.0504924 | -1.5384145 | -0.0740925 | 0.95310319 |
| 8 | 8 | 155.99 | 0.33 | 7.5 | 2.03 | 0.23483092 | 0.72438712 | -1.8734443 | 0.5440477 |
| 9 | 9 | 80.88 | 0.285 | 8.4 | 2.54 | -0.6393527 | 0.91655267 | -0.1033575 | -0.6498371 |
| 10 | 10 | 109.77 | 0.59 | 10.6 | 4.9 | 1.88556417 | 0.77585349 | 0.47675101 | -0.8549505 |

Figure 5.4.4: Uncorrelated Principal Component Scores

## 5.5 Cluster Analysis

**Example:** The given below are the adjusted means of 8 characters observed in an experiment to evaluate 110 genotypes of Lentil conducted using an alpha-design in 3 replications with block size 10.

| Genotypes | Y1 | Y2 | Y3 | Y4 | Y5 | Y6 | Y7 | Y8 |
|---|---|---|---|---|---|---|---|---|
| 1 | 42.43 | 7.51 | 78.32 | 128.24 | 45.26 | 6.51 | 7.74 | 7.66 |
| 2 | 43.83 | 7.91 | 71.20 | 114.71 | 82.90 | 6.51 | 6.34 | 7.05 |
| 3 | 40.77 | 9.77 | 85.47 | 129.50 | 189.22 | 6.64 | 6.91 | 10.09 |
| 4 | 43.35 | 8.18 | 68.25 | 113.16 | 176.59 | 6.77 | 6.57 | 8.91 |
| 5 | 45.74 | 7.97 | 83.52 | 123.77 | 80.53 | 6.57 | 6.83 | 7.34 |
| 6 | 45.43 | 8.11 | 84.45 | 132.23 | 77.30 | 6.47 | 6.81 | 7.35 |
| 7 | 42.68 | 7.68 | 92.47 | 128.09 | 36.70 | 6.55 | 7.16 | 6.43 |
| 8 | 39.18 | 6.64 | 73.50 | 122.04 | 49.24 | 6.25 | 7.20 | 6.75 |
| 9 | 45.98 | 8.10 | 67.41 | 123.39 | 45.25 | 6.20 | 8.01 | 6.68 |
| 10 | 43.64 | 8.56 | 84.44 | 138.81 | 90.98 | 6.64 | 6.28 | 7.04 |
| 11 | 44.68 | 8.11 | 91.71 | 125.12 | 65.78 | 6.40 | 6.49 | 6.51 |
| 12 | 45.90 | 7.50 | 70.85 | 122.81 | 54.94 | 6.36 | 8.85 | 8.48 |
| 13 | 42.61 | 7.57 | 75.78 | 120.91 | 85.98 | 6.57 | 7.02 | 6.91 |
| 14 | 42.56 | 8.21 | 94.64 | 134.48 | 111.13 | 6.65 | 6.60 | 8.06 |
| 15 | 45.86 | 7.78 | 84.67 | 123.80 | 82.93 | 6.67 | 6.75 | 6.89 |
| 16 | 41.70 | 8.00 | 95.02 | 137.69 | 116.10 | 6.63 | 6.58 | 7.45 |
| 17 | 43.25 | 7.78 | 82.50 | 129.05 | 107.23 | 6.68 | 6.83 | 7.46 |
| 18 | 43.05 | 8.10 | 73.76 | 120.28 | 203.81 | 6.64 | 6.65 | 10.18 |
| 19 | 40.24 | 7.48 | 74.66 | 121.99 | 88.66 | 6.69 | 6.53 | 7.46 |
| 20 | 44.43 | 8.32 | 83.58 | 122.86 | 83.86 | 6.58 | 6.58 | 8.57 |
| 21 | 44.34 | 7.81 | 75.01 | 129.42 | 74.00 | 6.61 | 7.53 | 8.53 |
| 22 | 44.67 | 7.98 | 75.55 | 123.79 | 102.31 | 6.62 | 6.72 | 7.85 |
| 23 | 43.54 | 7.65 | 94.30 | 134.73 | 77.17 | 6.53 | 6.93 | 7.26 |
| 24 | 45.10 | 8.01 | 91.40 | 134.84 | 86.93 | 6.52 | 7.32 | 7.12 |
| 25 | 45.03 | 10.15 | 85.85 | 133.02 | 73.07 | 6.63 | 6.32 | 6.85 |
| 26 | 46.82 | 7.54 | 85.22 | 130.92 | 65.22 | 6.63 | 6.59 | 7.10 |
| 27 | 46.52 | 7.65 | 82.51 | 125.34 | 66.98 | 6.59 | 6.60 | 6.90 |
| 28 | 42.33 | 7.20 | 65.85 | 120.38 | 66.24 | 6.61 | 7.41 | 7.17 |
| 29 | 42.98 | 8.06 | 84.04 | 128.80 | 94.25 | 6.65 | 6.91 | 8.18 |
| 30 | 46.19 | 7.60 | 94.02 | 137.15 | 97.35 | 6.59 | 6.26 | 7.38 |
| 31 | 45.69 | 8.03 | 92.94 | 136.41 | 68.96 | 6.47 | 6.41 | 6.54 |
| 32 | 44.46 | 7.79 | 85.94 | 132.44 | 83.46 | 6.65 | 6.74 | 7.15 |
| 33 | 46.50 | 7.92 | 82.59 | 133.78 | 56.92 | 6.39 | 7.11 | 6.43 |
| 34 | 46.45 | 8.30 | 81.20 | 134.81 | 92.09 | 6.43 | 8.09 | 7.52 |
| 35 | 43.05 | 7.98 | 84.54 | 131.97 | 88.59 | 6.62 | 6.59 | 7.78 |
| 36 | 43.64 | 7.49 | 66.53 | 114.51 | 107.24 | 6.78 | 7.11 | 8.29 |
| 37 | 44.57 | 7.76 | 91.55 | 135.29 | 49.93 | 6.48 | 6.94 | 6.74 |
| 38 | 44.34 | 8.00 | 94.67 | 140.19 | 103.65 | 6.48 | 6.50 | 7.40 |
| 39 | 43.65 | 7.66 | 66.88 | 119.78 | 127.28 | 6.59 | 7.13 | 8.70 |
| 40 | 43.58 | 7.92 | 80.75 | 133.31 | 147.71 | 6.79 | 6.40 | 8.58 |
| 41 | 44.45 | 8.06 | 84.76 | 121.68 | 85.58 | 6.10 | 8.22 | 7.75 |
| 42 | 47.87 | 8.12 | 86.94 | 125.43 | 81.97 | 6.39 | 7.61 | 8.11 |
| 43 | 45.23 | 7.22 | 79.38 | 126.39 | 51.82 | 6.33 | 9.44 | 7.25 |
| 44 | 43.53 | 7.55 | 82.47 | 131.00 | 35.56 | 6.04 | 10.36 | 7.59 |
| 45 | 43.13 | 7.52 | 67.45 | 116.49 | 133.48 | 6.70 | 6.59 | 9.20 |
| 46 | 41.95 | 7.34 | 71.86 | 121.80 | 95.91 | 6.74 | 6.52 | 7.58 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 47 | 41.06 | 7.28 | 69.43 | 120.32 | 74.69 | 6.65 | 6.78 | 7.18 |
| 48 | 39.71 | 7.24 | 64.23 | 114.25 | 130.62 | 6.76 | 6.59 | 8.83 |
| 49 | 41.30 | 7.18 | 64.53 | 113.78 | 93.68 | 6.78 | 6.52 | 7.71 |
| 50 | 41.95 | 7.30 | 66.53 | 115.44 | 139.30 | 6.71 | 6.32 | 7.96 |
| 51 | 43.48 | 7.38 | 66.36 | 115.12 | 106.66 | 6.74 | 6.24 | 8.01 |
| 52 | 43.22 | 7.56 | 71.29 | 115.29 | 159.95 | 6.68 | 6.44 | 8.54 |
| 53 | 40.15 | 7.31 | 67.65 | 115.35 | 140.49 | 6.77 | 6.69 | 8.48 |
| 54 | 44.30 | 7.73 | 63.26 | 117.41 | 144.21 | 6.71 | 6.59 | 8.42 |
| 55 | 38.25 | 7.24 | 63.71 | 113.24 | 104.85 | 6.76 | 6.69 | 7.46 |
| 56 | 44.07 | 7.53 | 64.34 | 113.68 | 96.62 | 6.75 | 6.48 | 8.01 |
| 57 | 43.79 | 7.75 | 65.22 | 116.27 | 128.92 | 6.70 | 6.55 | 8.58 |
| 58 | 43.11 | 7.68 | 62.48 | 115.50 | 143.87 | 6.79 | 6.52 | 8.97 |
| 59 | 40.87 | 7.55 | 66.39 | 114.95 | 141.49 | 6.79 | 6.43 | 8.31 |
| 60 | 42.98 | 7.36 | 68.64 | 115.32 | 115.23 | 6.72 | 6.46 | 8.06 |
| 61 | 47.40 | 7.76 | 65.37 | 115.10 | 134.99 | 6.64 | 6.60 | 7.41 |
| 62 | 39.00 | 7.67 | 64.94 | 113.05 | 122.66 | 6.72 | 6.49 | 7.71 |
| 63 | 44.38 | 7.41 | 68.30 | 115.99 | 128.34 | 6.76 | 6.66 | 8.65 |
| 64 | 42.13 | 7.19 | 68.88 | 119.77 | 90.78 | 6.67 | 7.21 | 7.90 |
| 65 | 42.68 | 7.40 | 65.26 | 118.73 | 115.82 | 6.79 | 7.03 | 8.35 |
| 66 | 40.62 | 7.85 | 65.17 | 113.06 | 134.02 | 6.79 | 6.38 | 8.44 |
| 67 | 44.43 | 7.40 | 67.14 | 117.82 | 115.09 | 6.73 | 7.37 | 8.47 |
| 68 | 41.56 | 6.94 | 69.03 | 115.53 | 93.68 | 6.63 | 6.79 | 7.23 |
| 69 | 41.07 | 7.00 | 63.97 | 115.42 | 91.20 | 6.83 | 7.53 | 7.92 |
| 70 | 41.10 | 7.71 | 63.98 | 113.52 | 144.02 | 6.86 | 6.65 | 11.09 |
| 71 | 42.45 | 7.12 | 65.92 | 117.29 | 79.98 | 6.75 | 7.18 | 7.15 |
| 72 | 42.12 | 7.35 | 60.95 | 108.99 | 128.10 | 6.77 | 6.40 | 8.13 |
| 73 | 41.00 | 7.33 | 65.33 | 113.44 | 130.96 | 6.77 | 6.37 | 7.97 |
| 74 | 43.67 | 7.64 | 62.95 | 118.32 | 119.09 | 6.72 | 7.15 | 9.26 |
| 75 | 46.49 | 7.97 | 88.06 | 126.87 | 75.97 | 6.58 | 6.53 | 6.78 |
| 76 | 42.98 | 7.39 | 66.57 | 119.79 | 118.84 | 6.65 | 7.04 | 8.54 |
| 77 | 41.01 | 7.02 | 59.90 | 113.64 | 104.40 | 6.70 | 7.09 | 8.36 |
| 78 | 48.85 | 6.84 | 45.32 | 104.53 | 66.53 | 6.75 | 8.27 | 7.69 |
| 79 | 49.60 | 7.17 | 59.37 | 110.36 | 82.16 | 6.76 | 7.93 | 7.80 |
| 80 | 49.50 | 7.40 | 62.24 | 113.10 | 144.37 | 6.50 | 6.64 | 9.01 |
| 81 | 44.53 | 7.63 | 65.14 | 113.71 | 140.34 | 6.76 | 6.74 | 8.31 |
| 82 | 46.59 | 7.47 | 85.08 | 123.53 | 72.37 | 6.67 | 6.37 | 6.74 |
| 83 | 44.78 | 7.47 | 85.72 | 126.54 | 113.46 | 6.69 | 6.59 | 6.92 |
| 84 | 42.22 | 7.33 | 69.77 | 115.38 | 105.00 | 6.70 | 6.67 | 8.11 |
| 85 | 37.10 | 7.13 | 80.79 | 122.32 | 64.18 | 6.44 | 6.29 | 6.69 |
| 86 | 44.42 | 6.94 | 66.76 | 120.34 | 49.67 | 6.33 | 8.35 | 7.28 |
| 87 | 45.53 | 7.52 | 67.79 | 114.47 | 146.09 | 6.71 | 6.54 | 8.13 |
| 88 | 42.50 | 7.48 | 62.49 | 114.59 | 130.72 | 6.79 | 6.61 | 8.24 |
| 89 | 46.06 | 7.67 | 86.69 | 125.51 | 75.84 | 6.44 | 6.44 | 7.12 |
| 90 | 36.44 | 7.45 | 71.74 | 114.81 | 137.22 | 6.73 | 6.28 | 8.23 |
| 91 | 42.67 | 7.36 | 70.64 | 121.17 | 43.06 | 6.71 | 6.79 | 6.13 |
| 92 | 40.44 | 7.06 | 59.99 | 107.81 | 99.93 | 6.86 | 7.11 | 8.53 |
| 93 | 39.35 | 7.51 | 55.26 | 120.88 | 91.51 | 6.63 | 6.86 | 6.98 |
| 94 | 45.41 | 7.08 | 67.92 | 118.80 | 64.25 | 6.63 | 7.93 | 7.61 |
| 95 | 43.19 | 7.32 | 65.39 | 118.82 | 98.70 | 6.67 | 7.07 | 8.18 |
| 96 | 45.43 | 7.52 | 65.12 | 118.90 | 95.82 | 6.79 | 7.61 | 8.21 |
| 97 | 44.34 | 7.48 | 63.66 | 113.72 | 110.46 | 6.79 | 7.02 | 7.97 |
| 98 | 42.76 | 7.25 | 63.02 | 109.59 | 104.84 | 6.84 | 7.08 | 9.16 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 99 | 45.07 | 8.56 | 92.05 | 135.80 | 94.57 | 6.68 | 6.55 | 7.60 |
| 100 | 40.96 | 7.39 | 63.33 | 112.40 | 96.36 | 6.76 | 6.90 | 7.56 |
| 101 | 42.15 | 6.97 | 61.06 | 114.76 | 118.46 | 6.78 | 6.71 | 8.42 |
| 102 | 40.36 | 7.00 | 69.93 | 120.29 | 53.84 | 6.23 | 8.44 | 7.15 |
| 103 | 42.68 | 7.81 | 62.51 | 113.08 | 152.97 | 6.75 | 6.83 | 8.49 |
| 104 | 39.78 | 7.00 | 69.38 | 121.18 | 76.18 | 6.57 | 8.58 | 9.46 |
| 105 | 41.20 | 7.06 | 68.03 | 118.67 | 69.91 | 6.61 | 6.48 | 6.91 |
| 106 | 40.65 | 7.16 | 65.63 | 118.89 | 92.78 | 6.71 | 7.03 | 7.99 |
| 107 | 43.32 | 7.78 | 65.56 | 118.91 | 128.72 | 6.82 | 6.92 | 8.63 |
| 108 | 50.53 | 7.07 | 68.18 | 118.80 | 89.23 | 6.72 | 7.84 | 8.66 |
| 109 | 44.89 | 7.82 | 65.60 | 114.99 | 124.45 | 6.77 | 6.67 | 8.40 |
| 110 | 46.23 | 7.72 | 67.76 | 120.83 | 111.15 | 6.72 | 7.52 | 9.12 |

**Note**-Adjusted means have been subjected to change of origin and change of scale so as to retain the rights of original data in the experiment.

1. Using the data, perform hierarchical cluster analysis with unweighted pair-group method using arithmetic averages (UPGMA) method.
2. Construct the dendrogram.

**Analysis using JMP**

1. First open the Excel file CLUSTER.xls using **File → Open** dialog box.
2. Make sure variable 'Treatments' is nominal.
3. Go to **Analyze → Multivariate Methods → Cluster** Select the variables y1, y2, y3, y4, y5, y6, y7 and y8 from **Select Columns** box → **Y, Columns**. Give **Treatments → Label** variable. Select **Method → Average**. See Clustering dialog box as in Figure 5.5.1. Uncheck **Standardize Data → OK**. The Distance is RMS (Root Mean Square) distance shown in the partial Clustering history output in Figure 5.5.4.
4. Go to hotspot button on the left of **Hierarchical Clustering → Mark Clusters** and **Color Clusters**. To change the number of clusters to three, go to hotspot button and click **Number of Clusters** and give value as 3 or one can also change the number of clusters with the help of **diamond slider** in Dendrogram output. See Figure 5.5.2.
5. One can also save the clusters using hotspot button on clicking **Save Clusters**. A new data column gets created in the data table containing cluster numbers.

Standardizes the variables by subtracting Mean and dividing by Std Deviation.

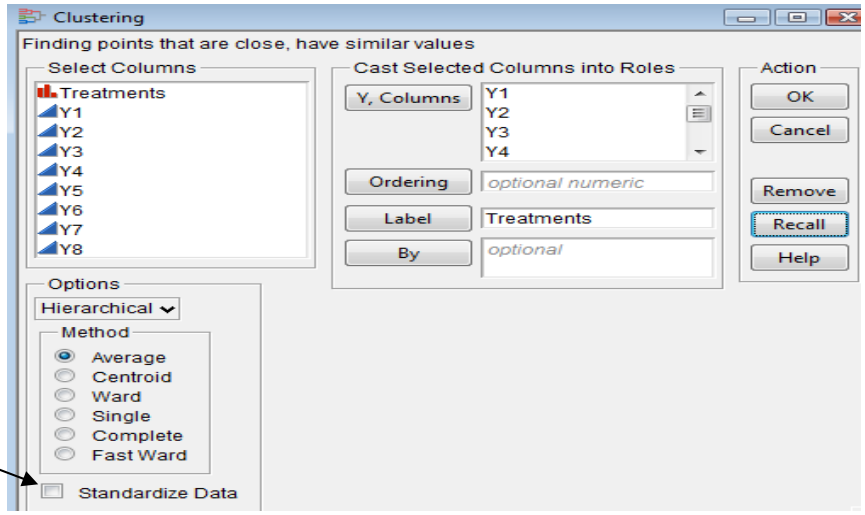
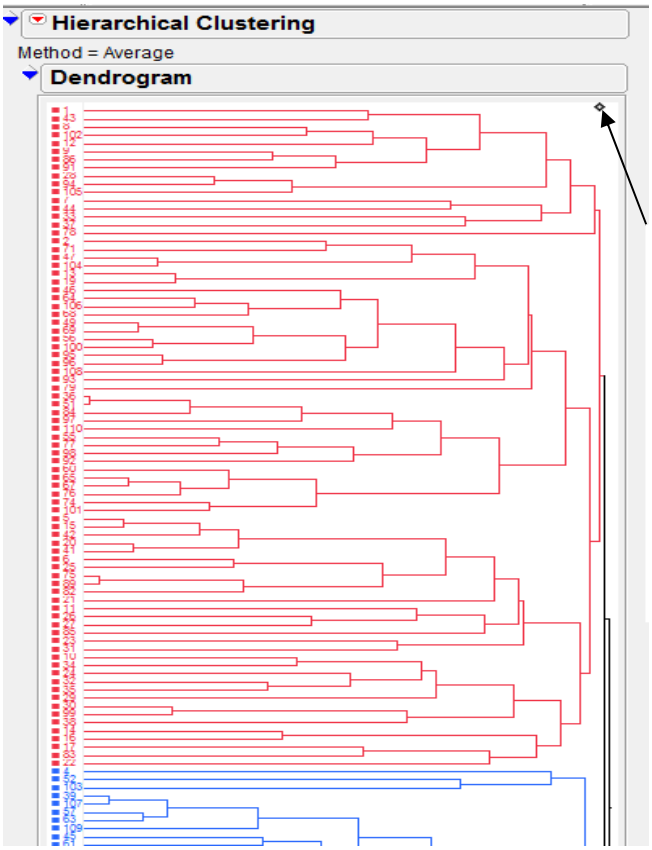



Figure 5.5.1: Clustering Dialog Box




Identify *n* clusters: drag slider by upper or lower diamond to intersect *n* cluster stems or use the **Number of Clusters** popup menu option.

Figure 5.5.2: Dendogram

The scree plot beneath the dendrogram has a point for each cluster join. The ordinate is the distance that was bridged to join the clusters at each step. Often there is a natural break where the distance jumps up suddenly. These breaks suggest natural cutting points to determine the number of clusters. See Figure 5.5.3.

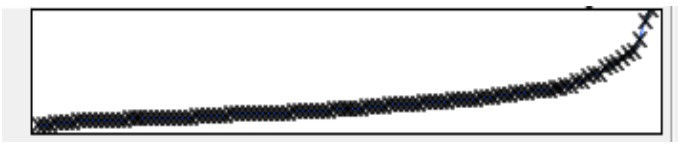


Figure 5.5.3: Scree Plot

Hierarchical Clustering

Clustering History

RMS Distance

| Number of Clusters | Distance | Leader | Joiner |
|---|---|---|---|
| 109 | 1.2696456 | 36 | 51 |
| 108 | 1.8494594 | 53 | 59 |
| 107 | 2.0400000 | 75 | 89 |
| 106 | 2.1076053 | 48 | 73 |
| 105 | 2.1904566 | 39 | 107 |
| 104 | 2.4707286 | 54 | 58 |
| 103 | 2.5254010 | 50 | 53 |
| 102 | 2.7116047 | 5 | 15 |
| 101 | 2.8445562 | 65 | 67 |
| 100 | 3.0665290 | 20 | 41 |
| 99 | 3.2102336 | 49 | 69 |
| 98 | 3.2226542 | 57 | 63 |
| 97 | 3.4053928 | 48 | 88 |
| 96 | 3.5772615 | 56 | 100 |
| 95 | 3.6226786 | 47 | 104 |

Figure 5.5.4.: Clustering History Output

## Analysis of data from Designed Experiments

### 5.6 Completely Randomised Design

**Example 5.6**:.A feeding trial with 3 feeds namely (i) Pasture(control), (ii) Pasture and Concentrates and (iii) Pasture, Concentrates and Minerals was conducted at the Yellachihalli Sheep Farm, Mysore, to study their effect on wool yield of Sheep. For this purpose twenty-five ewe lambs were allotted at random to each of the three treatments and the three treatments and the weight records of the total wool yield (in gms) of first two clipping were obtained. The data for two lambs for feed 1 {Pasture (control)}, three for feed 2 {Pasture and Concentrates} and one for feed 3 {Pasture, Concentrates and Minerals} are missing. The details of the experiment are given below:

**Yield (in gms)**

| FEED 1 | FEED 2 | FEED 3 |
|---|---|---|
| 850.50 | 510.30 | 992.25 |
| 453.60 | 963.90 | 850.50 |
| 878.85 | 652.05 | 1474.20 |
| 623.70 | 1020.60 | 510.30 |
| 510.30 | 878.85 | 850.50 |
| 765.45 | 567.00 | 793.80 |
| 680.40 | 680.40 | 453.60 |
| 595.35 | 538.65 | 935.55 |
| 538.65 | 567.00 | 1190.70 |
| 850.50 | 510.30 | 481.95 |
| 850.50 | 425.25 | 623.70 |
| 793.80 | 567.00 | 878.85 |
| 1020.60 | 623.70 | 1077.30 |
| 708.75 | 538.65 | 850.50 |
| 652.05 | 737.10 | 680.40 |
| 623.70 | 453.60 | 737.10 |
| 396.90 | 481.95 | 737.10 |
| 822.15 | 368.55 | 708.75 |
| 680.40 | 567.00 | 708.75 |
| 652.05 | 595.35 | 652.05 |
| 538.65 | 567.00 | 567.00 |
| 850.50 | 595.35 | 453.60 |
| 680.40 | . | 652.05 |
| . | . | 567.00 |
| . | . | . |

1. Perform analysis of variance of the data to test whether there is any difference between treatment effects.
2. Perform all possible pair wise treatment comparisons and identify the best treatment i.e. the treatment giving highest yield.

**Analysis using JMP**

Here factor is "Classification Variable" (treatment). Response is "Dependent Variable" (yield)

1. First open the CRD.xls file in JMP using **File → Open** dialog box.
2. Make sure the Factor (trt) is nominal.
3. Go to **Analyze → Fit Y by X**.
4. Put Factor into the **X, Factor** box, Response into the **Y, Response** box **→ OK**.
5. A box will pop up with a plot of the data. Click the red triangle on this box and Select **Means/Anova**. This will give the basic ANOVA output. (See Figure 5.6.1 & 5.6.2)
6. For all possible pairwise treatment comparisons, click the red triangle and select **Compare Means → Each pair, Student's t**.

**Output:**

Each multiple comparison test begins with a comparison circles plot, which is a visual representation of group mean comparisons. The plot shows the alignment of comparison circles with the confidence intervals of their respective group means for the Student's t comparison. Other comparison tests widen or shorten the radii of the circles.

**Overlap marks** show for each diamond and are computed as (group mean ± ($\sqrt{2}$ * CI)/4).

Overlap marks in one diamond that are closer to the mean of another diamond than that diamond's overlap marks indicate that those two groups are not different at the 95% confidence level.

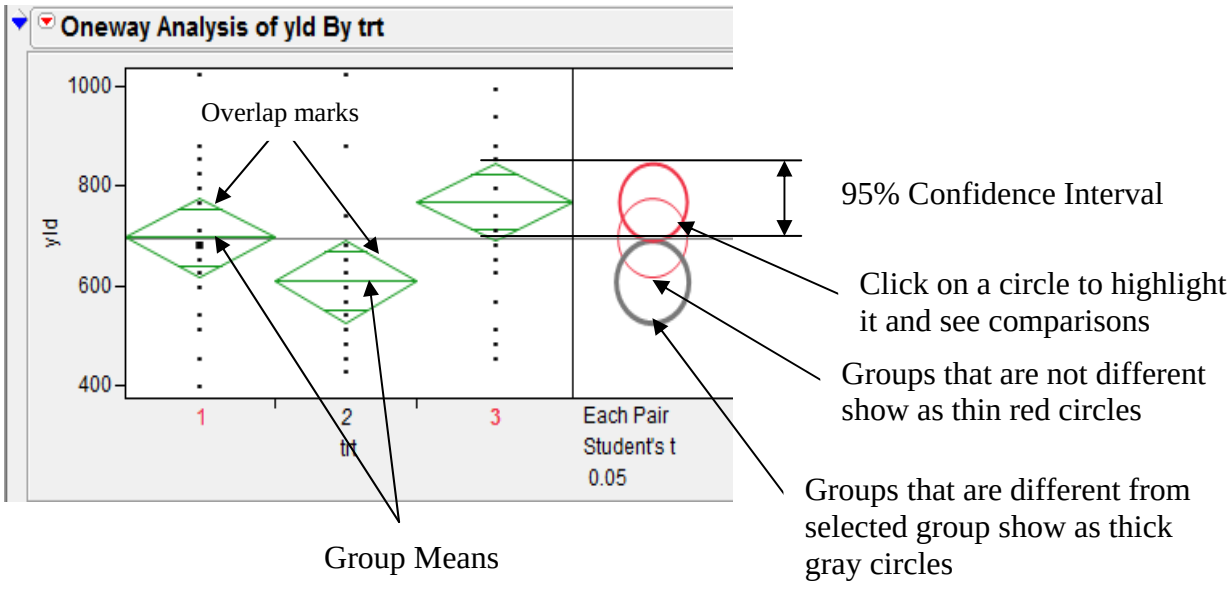
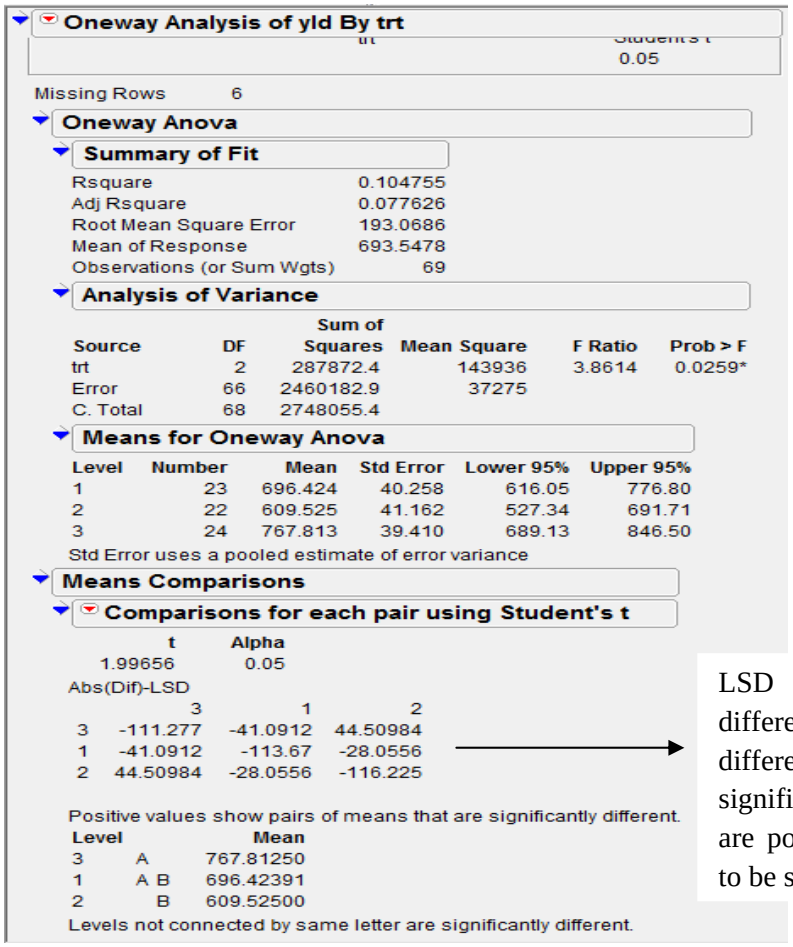



Figure 5.6.1: Visual Comparisons of Group Means



**Oneway Analysis of yld By trt**

trt — Student's t 0.05

Missing Rows 6

**Oneway Anova**

**Summary of Fit**

| | |
|---|---|
| Rsquare | 0.104755 |
| Adj Rsquare | 0.077626 |
| Root Mean Square Error | 193.0686 |
| Mean of Response | 693.5478 |
| Observations (or Sum Wgts) | 69 |

**Analysis of Variance**

| Source | DF | Sum of Squares | Mean Square | F Ratio | Prob > F |
|---|---|---|---|---|---|
| trt | 2 | 287872.4 | 143936 | 3.8614 | 0.0259* |
| Error | 66 | 2460182.9 | 37275 | | |
| C. Total | 68 | 2748055.4 | | | |

**Means for Oneway Anova**

| Level | Number | Mean | Std Error | Lower 95% | Upper 95% |
|---|---|---|---|---|---|
| 1 | 23 | 696.424 | 40.258 | 616.05 | 776.80 |
| 2 | 22 | 609.525 | 41.162 | 527.34 | 691.71 |
| 3 | 24 | 767.813 | 39.410 | 689.13 | 846.50 |

Std Error uses a pooled estimate of error variance

**Means Comparisons**

**Comparisons for each pair using Student's t**

| t | Alpha |
|---|---|
| 1.99656 | 0.05 |

Abs(Dif)-LSD

| | 3 | 1 | 2 |
|---|---|---|---|
| 3 | -111.277 | -41.0912 | 44.50984 |
| 1 | -41.0912 | -113.67 | -28.0556 |
| 2 | 44.50984 | -28.0556 | -116.225 |

Positive values show pairs of means that are significantly different.

| Level | | Mean |
|---|---|---|
| 3 | A | 767.81250 |
| 1 | A B | 696.42391 |
| 2 | B | 609.52500 |

Levels not connected by same letter are significantly different.

LSD Threshold Matrix is the difference between the absolute difference and the LSD that would be significantly different. If the values are positive, the difference is judged to be significant.

Figure 5.6.2: One way ANOVA and Means Comparisons Output

## 5.7 Randomized Complete Block Design

**Example 5.7:** An initial varietal trial (Late Sown, irrigated) was conducted to study the performance of 20 new strains of mustard vis-a-vis four checks (Swarna Jyoti: ZC; Vardan: NC; Varuna: NC; and Kranti: NC) using a Randomized complete Block Design (RCB) design at Bhatinda with 3 replications. The seed yield in kg/ha was recorded. The details of the experiment are given below:

**Yield in kg/ha**

| Strain | Code | Replications | | |
|---|---|---|---|---|
| | | 1 | 2 | 3 |
| RK-04-3 | MCN-04-110 | 1539.69 | 1412.35 | 1319.73 |
| RK-04-4 | MCN-04-111 | 1261.85 | 1065.05 | 1111.36 |
| RGN-124 | MCN-04-112 | 1389.19 | 1516.54 | 1203.97 |
| HYT-27 | MCN-04-113 | 1192.39 | 1215.55 | 1157.66 |
| PBR-275 | MCN-04-114 | 1250.27 | 1203.97 | 1366.04 |
| HUJM-03-03 | MCN-04-115 | 1296.58 | 1273.43 | 1308.16 |
| RGN-123 | MCN-04-116 | 1227.12 | 1018.74 | 937.71 |
| BIO-13-01 | MCN-04-117 | 1273.43 | 1157.66 | 1088.20 |
| RH-0115 | MCN-04-118 | 1180.82 | 1203.97 | 1041.90 |
| RH-0213 | MCN-04-119 | 1296.58 | 1458.65 | 1250.27 |

| | | | | |
|---|---|---|---|---|
| NRCDR-05 | MCN-04-120 | 1122.93 | 1065.05 | 1018.74 |
| NRC-323-1 | MCN-04-121 | 1250.27 | 926.13 | 1030.32 |
| RRN-596 | MCN-04-122 | 1180.82 | 1053.47 | 717.75 |
| RRN-597 | MCN-04-123 | 1146.09 | 1180.82 | 856.67 |
| CS-234-2 | MCN-04-124 | 1574.42 | 1412.35 | 1597.57 |
| RM-109 | MCN-04-125 | 914.55 | 972.44 | 659.87 |
| BAUSM-2000 | MCN-04-126 | 891.40 | 937.71 | 798.79 |
| NPJ-99 | MCN-04-127 | 1227.12 | 1203.97 | 1389.19 |
| **SWARNA JYOTI(ZC)** | **MCN-04-128** | 1389.19 | 1180.82 | 1273.43 |
| **VARDAN(NC)** | **MCN-04-129** | 1331.31 | 1157.66 | 1180.82 |
| PR-2003-27 | MCN-04-130 | 1250.27 | 1250.27 | 1296.58 |
| **VARUNA(NC)** | **MCN-04-131** | 717.75 | 740.90 | 578.83 |
| PR-2003-30 | MCN-04-132 | 1169.24 | 1157.66 | 1111.36 |
| **KRANTI-(NC)** | **MCN-04-133** | 1203.97 | 1296.58 | 1250.27 |

**Note**: Strains of mustard in bold are the four checks.

1. Perform the analysis of variance of the data to test whether there is any difference between treatment effects.
2. Perform all possible pair wise treatment comparisons and identify the best treatment i.e. the treatment giving highest yield. Also identify the other treatments which are non-significantly different from this treatment.
3. The varieties Swarna Jyoti (MCN-04-128), Vardan (MCN-04-129), Varuna (MCN-04-131) and Kranti (MCN-04-133) were check varieties and rest of them were strains. Test whether the performance of check varieties is significantly different from strains.

**Analysis using JMP**

1. As in the one-way ANOVA, the first step is to make sure that both the blocking and treatment factors are nominal.
2. Open the RBD.xls file in JMP using **File → Open** dialog box.
3. Select **Analyze → Fit Model**.
4. Put the response variable Yield → **Y** and **Add** the blocking (rep) & treatment (trtn) factor to **Construct Model Effects → Run Model**. This will give the ANOVA table, treatment means and block means. (See Figure 5.7.1)
5. One can perform Tukey's procedure on the treatment, Click the red triangle besides the treatment factor (in the output window) and select **LSMeans Tukey HSD**. (See Figure 5.7.1)
6. For Contrast Analysis, click the red triangle and select **LSMeans Contrast...** and proceed as in the instruction. Click on + for strain varieties and - for check varieties and click **Done**. (See Figure 5.7.3)

**Output:**

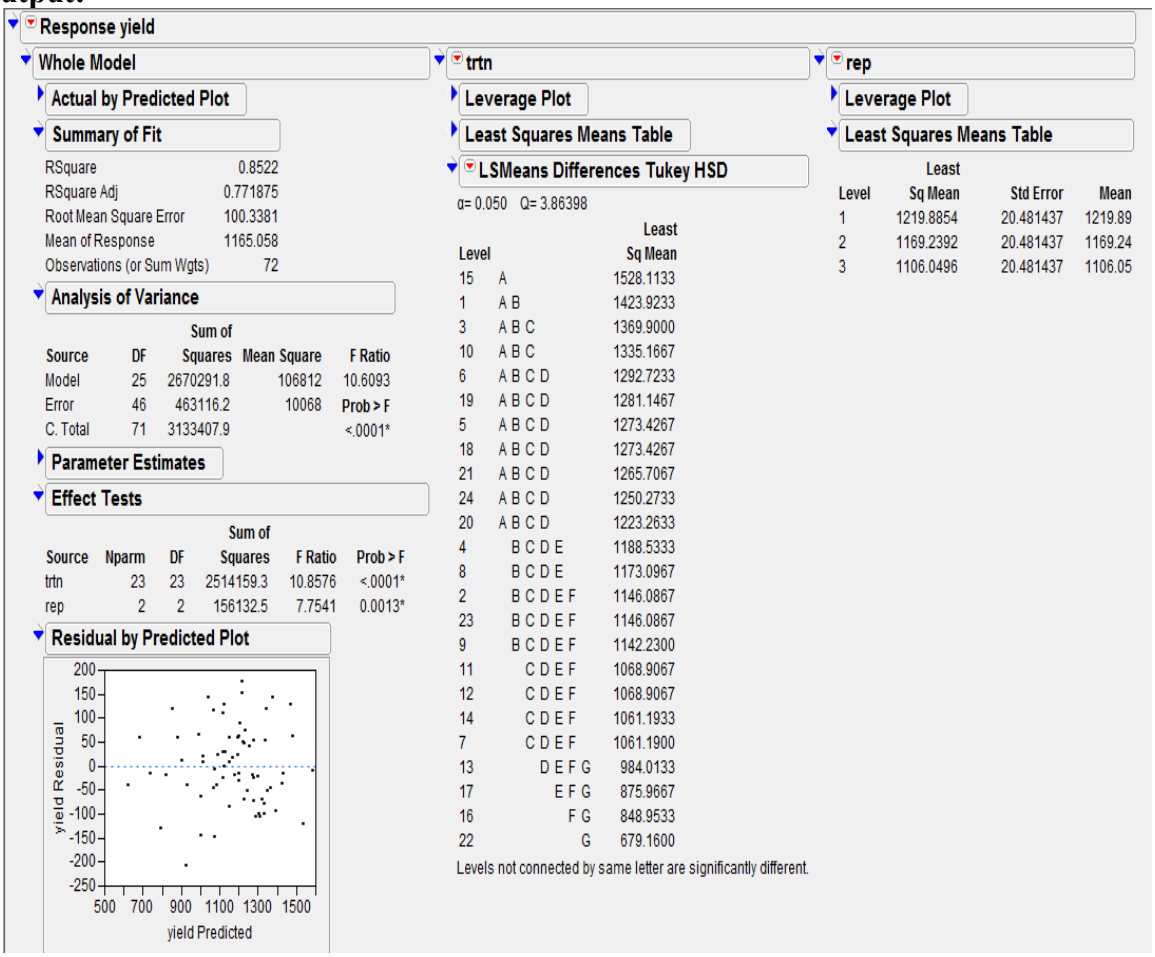



Figure 5.7.1: ANOVA table and Tukey Grouping Output

**Leverage Plot** - The graphical display of an effect's significance test is called a **Leverage Plot**. This kind of plot shows for each point what the residual would be both with and without that effect in the model. The distance from a point to the line of fit shows the actual residual. The distance from the point to the horizontal line of the mean shows what the residual error would be without the effect in the model.

The leverage plots are shown with C**onfidence Curves**. These indicate whether the test is significant at the 5% level by showing a confidence region for the line of fit. If the confidence region between the curves contains the horizontal line, then the effect is not significant. If the curves cross the line, the effect is significant.

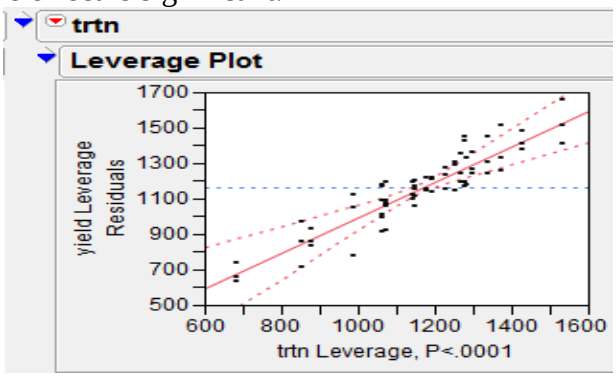



Figure 5.7.2: Leverage Plot for Treatments

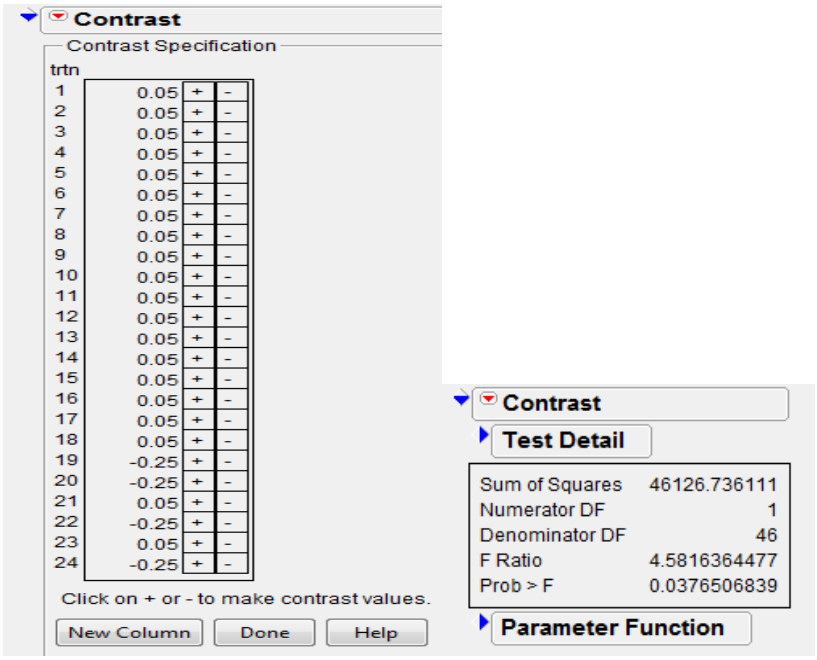



Figure 5.7.3: Contrast Analysis Output

## 5.8 Factorial RCB Design

**Example 5.8:** An experiment was conducted at Crop Research Center, G.B.P.U.A.T., Pantnagar, Uttar Pradesh on bengal gram in rabi season 2003 using a factorial experiment with three factors *viz.*, Farmyard Manure (2 levels:0 and 50 q/ha), Phosphorus (3 levels:0, 20 and 40 kg/ha) and Phosphorus Solublizing Bacteria (2 levels: control and 20 gm/kg of seed as seed inoculation). The main objective of the experiment was to study the effect of Farmyard Manure (FYM), Phosphorus (P) and Phosphate Solublizing Bacteria (PSB) on productivity of bengal gram. The experiment was conducted in a randomized complete block design in 4 replications with 12 plots per replication of net plot size as $5.00 \times 1.80 m^2$. The yield (in kg/plot) are given as below:

**Yield in kg/plot**

| **Replication 1** | | | | **Replication 2** | | | |
|---|---|---|---|---|---|---|---|
| **FYM** | **P** | **PSB** | **Yield** | **FYM** | **P** | **PSB** | **Yield** |
| 1 | 1 | 1 | 0.70 | 1 | 1 | 1 | 0.98 |
| 1 | 1 | 2 | 1.13 | 1 | 1 | 2 | 1.13 |
| 1 | 2 | 1 | 1.23 | 1 | 2 | 1 | 1.18 |
| 1 | 2 | 2 | 1.25 | 1 | 2 | 2 | 1.13 |
| 1 | 3 | 1 | 1.25 | 1 | 3 | 1 | 1.26 |
| 1 | 3 | 2 | 1.25 | 1 | 3 | 2 | 1.25 |
| 2 | 1 | 1 | 0.83 | 2 | 1 | 1 | 0.93 |
| 2 | 1 | 2 | 1.23 | 2 | 1 | 2 | 0.88 |
| 2 | 2 | 1 | 1.18 | 2 | 2 | 1 | 1.50 |
| 2 | 2 | 2 | 0.88 | 2 | 2 | 2 | 1.30 |
| 2 | 3 | 1 | 1.63 | 2 | 3 | 1 | 1.38 |
| 2 | 3 | 2 | 1.48 | 2 | 3 | 2 | 1.43 |

| Replication 3 | | | | Replication 4 | | | |
|---|---|---|---|---|---|---|---|
| **FYM** | **P** | **PSB** | **Yield** | **FYM** | **P** | **PSB** | **Yield** |
| 1 | 1 | 1 | 0.90 | 1 | 1 | 1 | 0.73 |
| 1 | 1 | 2 | 1.10 | 1 | 1 | 2 | 1.25 |
| 1 | 2 | 1 | 1.10 | 1 | 2 | 1 | 1.43 |
| 1 | 2 | 2 | 0.88 | 1 | 2 | 2 | 1.25 |
| 1 | 3 | 1 | 1.35 | 1 | 3 | 1 | 1.10 |
| 1 | 3 | 2 | 1.35 | 1 | 3 | 2 | 1.75 |
| 2 | 1 | 1 | 1.10 | 2 | 1 | 1 | 0.98 |
| 2 | 1 | 2 | 1.03 | 2 | 1 | 2 | 1.38 |
| 2 | 2 | 1 | 1.30 | 2 | 2 | 1 | 1.35 |
| 2 | 2 | 2 | 0.88 | 2 | 2 | 2 | 1.43 |
| 2 | 3 | 1 | 1.38 | 2 | 3 | 1 | 1.30 |
| 2 | 3 | 2 | 1.43 | 2 | 3 | 2 | 1.50 |

Here levels of different factors have been coded as

| **Factor** | **Original** | **Coded Level** |
|---|---|---|
| Farmyard Manure | 0 q/ha | 1 |
| | 50 q/ha | 2 |
| Phosphorus | 0 kg/ha | 1 |
| | 20 kg/ha | 2 |
| | 40 kg/ha | 3 |
| Phosphate Solublizing Bacteria | 0 gm/ha | 1 |
| | 20 gm/ha | 2 |

(For performing the analysis using original levels, change coded levels to original level throughout the steps discussed in sequel).

1. Perform the analysis of variance of the data to test the significance of the main effects of farmyard manure, phosphorus, phosphorus solublizing bacteria and their 2-factor and 3-factor interactions.
2. Identify the best levels of farmyard manure, phosphorus and phosphorus solublizing bacteria and combinations of two factors.
3. Perform the analysis of variance for testing the significance of treatment combinations and identify the best treatment combination.
4. Compare the treatment combinations with and without farmyard manure.
5. Compare the treatments with and without phosphorus solublizing bacteria.

**Analysis using JMP**

1. As in the two-way ANOVA, the first step is to make sure that both the blocking (REP) and treatment (FYM, P, PSB, trt) factors are nominal.
2. Open the FACT.xls file in JMP using **File → Open** dialog box.
3. Select **Analyze → Fit Model**.
4. Put the response variable, Yield → **Y**. **Add** REP from **Select Columns List → Construct Model Effects**, select FYM, P and PSB and click on **Macros → Full Factorial →** (See Figure 5.8.1) → **Run Model**. (See Output in Figure 5.8.2). One can also select **Emphasis → Minimal Report** to get output without Effect Leverage Plot.
5. Click the red triangle of P main effect and FYM*P interaction effect and select **LSMeans plot**. See Figure 5.8.3 for the output.
6. To answer Question 3, again select **Analyze → Fit Model**, Put the response variable Yield → **Y**. **Add** REP and TRT from **Select Columns List → Construct Model Effects → Run Model**. Alternatively, go to Fit Model in windows list and remove all the effects except REP and add TRT from **Select Columns List → Construct Model Effects → Run Model**.
7. One can perform Tukey's procedure on the treatment, click the red triangle besides the treatment factor and select **LSMeans Tukey HSD** (See Figure 5.8.4)
8. To answer Question 4, for contrast analysis click the red triangle and select **LSMeans Contrast...** See the Fact table, value of without FYM is 1, corresponding treatment values are from 1-6 and value of with FYM is 2, corresponding treatment values are from 7-12 Now click on + for 1-6 and click - for 7-12. One will get the output as in Figure 5.8.5(a).
9. Similarly proceed to answer Question 5, and one will get the output for contrast analysis of with and without PSM as in Figure 5.8.5(b).

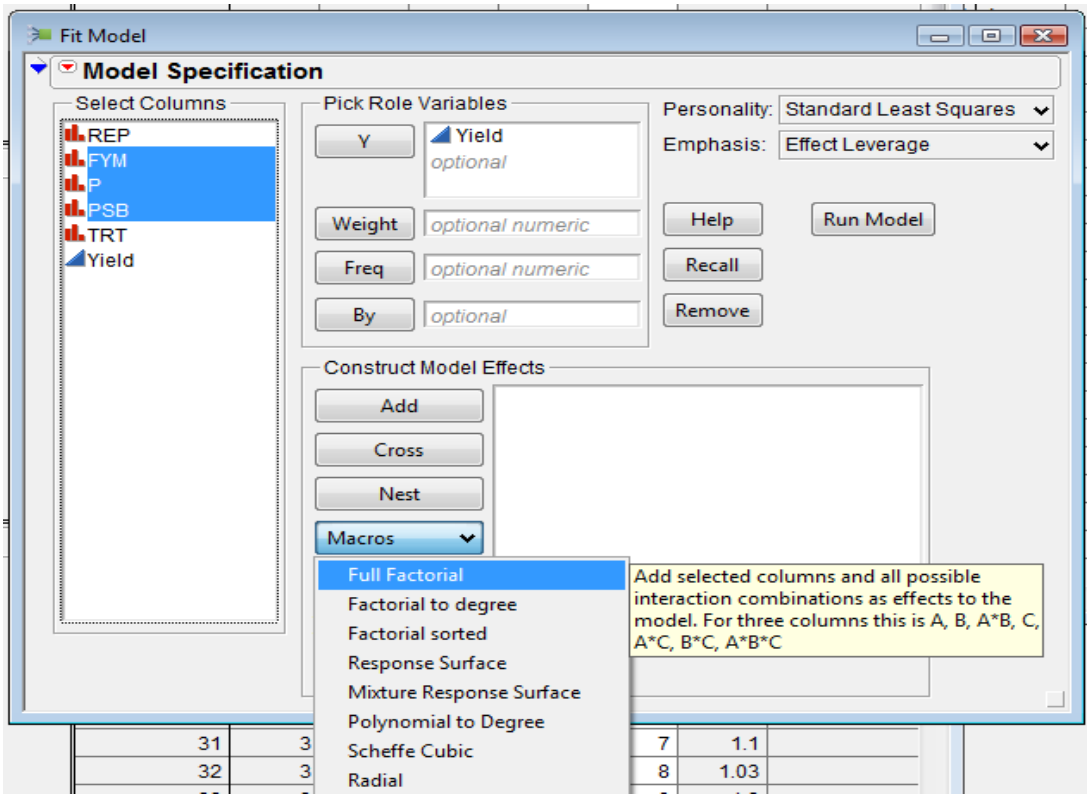
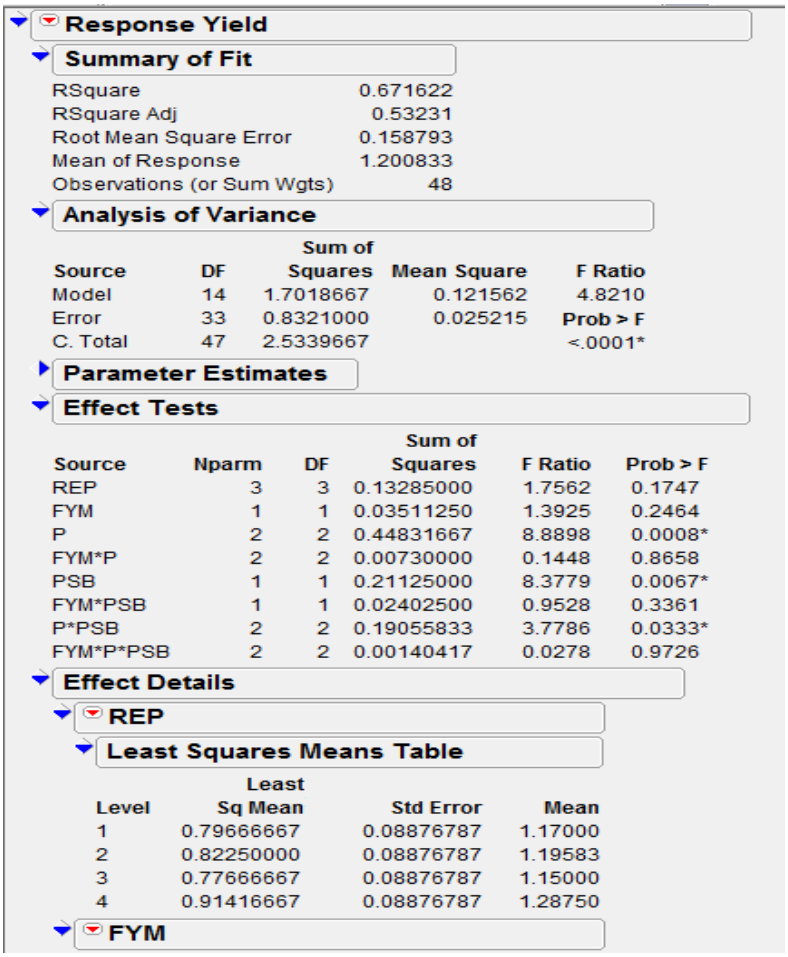



Figure 5.8.1: Fit Model dialog box



**Response Yield**

**Summary of Fit**

| | |
|---|---|
| RSquare | 0.671622 |
| RSquare Adj | 0.53231 |
| Root Mean Square Error | 0.158793 |
| Mean of Response | 1.200833 |
| Observations (or Sum Wgts) | 48 |

**Analysis of Variance**

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Model | 14 | 1.7018667 | 0.121562 | 4.8210 |
| Error | 33 | 0.8321000 | 0.025215 | Prob > F |
| C. Total | 47 | 2.5339667 | | <.0001* |

**Parameter Estimates**

**Effect Tests**

| Source | Nparm | DF | Sum of Squares | F Ratio | Prob > F |
|---|---|---|---|---|---|
| REP | 3 | 3 | 0.13285000 | 1.7562 | 0.1747 |
| FYM | 1 | 1 | 0.03511250 | 1.3925 | 0.2464 |
| P | 2 | 2 | 0.44831667 | 8.8898 | 0.0008* |
| FYM*P | 2 | 2 | 0.00730000 | 0.1448 | 0.8658 |
| PSB | 1 | 1 | 0.21125000 | 8.3779 | 0.0067* |
| FYM*PSB | 1 | 1 | 0.02402500 | 0.9528 | 0.3361 |
| P*PSB | 2 | 2 | 0.19055833 | 3.7786 | 0.0333* |
| FYM*P*PSB | 2 | 2 | 0.00140417 | 0.0278 | 0.9726 |

**Effect Details**

**REP**

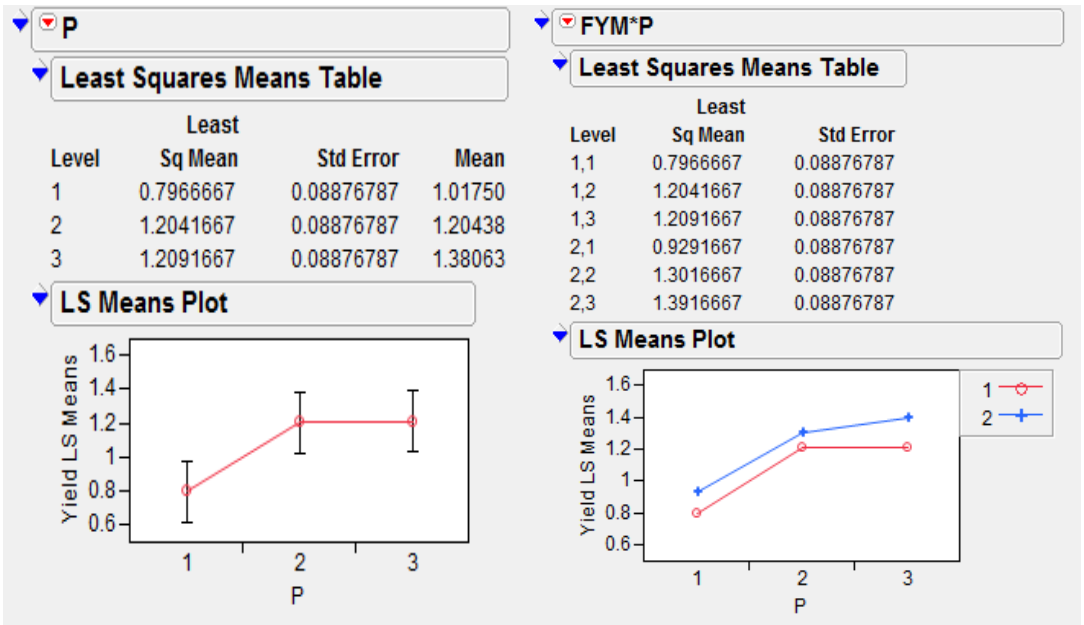
**Least Squares Means Table**

| Level | Least Sq Mean | Std Error | Mean |
|---|---|---|---|
| 1 | 0.79666667 | 0.08876787 | 1.17000 |
| 2 | 0.82250000 | 0.08876787 | 1.19583 |
| 3 | 0.77666667 | 0.08876787 | 1.15000 |
| 4 | 0.91416667 | 0.08876787 | 1.28750 |

**FYM**

Figure 5.8.2: Factorial RCB partial output

**LSMeans Plot** option plots least squares means (LSMeans) plots for nominal and ordinal main effects and two-way interactions.



**P**

**Least Squares Means Table**

| Level | Least Sq Mean | Std Error | Mean |
|---|---|---|---|
| 1 | 0.7966667 | 0.08876787 | 1.01750 |
| 2 | 1.2041667 | 0.08876787 | 1.20438 |
| 3 | 1.2091667 | 0.08876787 | 1.38063 |

**LS Means Plot**



**FYM*P**

**Least Squares Means Table**

| Level | Least Sq Mean | Std Error |
|---|---|---|
| 1,1 | 0.7966667 | 0.08876787 |
| 1,2 | 1.2041667 | 0.08876787 |
| 1,3 | 1.2091667 | 0.08876787 |
| 2,1 | 0.9291667 | 0.08876787 |
| 2,2 | 1.3016667 | 0.08876787 |
| 2,3 | 1.3916667 | 0.08876787 |

**LS Means Plot**



Figure 5.8.3: LSMeans plot for main effect P and interaction effect FYM*P

TRT

Least Squares Means Table

LSMeans Differences Tukey HSD

α= 0.050 Q= 3.51108

| Level | | Least Sq Mean |
|---|---|---|
| 12 | A | 1.4600000 |
| 11 | A | 1.4225000 |
| 6 | A | 1.4000000 |
| 9 | A B | 1.3325000 |
| 5 | A B | 1.2400000 |
| 3 | A B | 1.2350000 |
| 2 | A B C | 1.1525000 |
| 8 | A B C | 1.1300000 |
| 4 | A B C | 1.1275000 |
| 10 | A B C | 1.1225000 |
| 7 | B C | 0.9600000 |
| 1 | C | 0.8275000 |

Levels not connected by same letter are significantly different

Figure 5.8.4 Tukey's Multiple Comparisons for all Treatments

Contrast

Contrast Specification

| TRT | | | |
|---|---|---|---|
| 1 | 0.1667 | + | - |
| 2 | 0.1667 | + | - |
| 3 | 0.1667 | + | - |
| 4 | 0.1667 | + | - |
| 5 | 0.1667 | + | - |
| 6 | 0.1667 | + | - |
| 7 | -0.167 | + | - |
| 8 | -0.167 | + | - |
| 9 | -0.167 | + | - |
| 10 | -0.167 | + | - |
| 11 | -0.167 | + | - |
| 12 | -0.167 | + | - |

Click on + or - to make contrast values.

New Column Done Help

Contrast

Test Detail

| | |
|---|---|
| Sum of Squares | 0.0660083333 |
| Numerator DF | 1 |
| Denominator DF | 33 |
| F Ratio | 2.6178043504 |
| Prob > F | 0.115190625 |

Parameter Function

Figure 5.8.5(a): Contrast Analysis Output for with and without FYM

Contrast

Contrast Specification

| TRT | | | |
|---|---|---|---|
| 1 | 0.1667 | + | - |
| 2 | -0.167 | + | - |
| 3 | 0.1667 | + | - |
| 4 | -0.167 | + | - |
| 5 | 0.1667 | + | - |
| 6 | -0.167 | + | - |
| 7 | 0.1667 | + | - |
| 8 | -0.167 | + | - |
| 9 | 0.1667 | + | - |
| 10 | -0.167 | + | - |
| 11 | 0.1667 | + | - |
| 12 | -0.167 | + | - |

Click on + or - to make contrast values.

New Column Done Help

Contrast

Test Detail

| | |
|---|---|
| Sum of Squares | 0.046875 |
| Numerator DF | 1 |
| Denominator DF | 33 |
| F Ratio | 1.859001322 |
| Prob > F | 0.1819701217 |

Parameter Function

Figure 5.8.5(b): Contrast Analysis Output for with and without PSM

## 5.9 Split-Plot Design

**Example 5.9:** An experiment was conducted at Crop Research Center, G.B.P.U.A.T., Pantnagar, Uttar Pradesh on wheat crop in 2006 Rabi season using a Split-plot (Main-A and sub-B) experiment to study the effects of organic, inorganic and micronutrients on grain yield of wheat under Paddy-Wheat cropping system. The experiment was conducted using a split plot design with method of sowing in paddy in Kharif (of Paddy in Kharif (transplanting and direct seeding) in main plots and six sub-plot treatments consisting of organic, inorganic fertilizers and micronutrients (B1: 150 kg/ha of N as Urea+60 kg/ha of $P_2O_5$ as Super+40 kg/ha of $K_2O$ as Murate of Potash. as recommended inorganic fertilizer, B2: B1 + 150 q/ha of FYM, B3=B1+residual effect of Green manure (Sesbania), B4=B1+$MnSO_4$ @ 0.5% as foliar spray, B5= B1 + 150 q/ha of FYM+ $MnSO_4$ @ 0.5% as foliar spray and B6= B1+residual effect of Green manure (Sesbania)+ $MnSO_4$ @ 0.5% as foliar spray. There were 3 replications, and the data of wheat yield in kg/ha is:

| | | Manurial Treatments | | | | | |
|---|---|---|---|---|---|---|---|
| **Replication** | **Method of Sowing** | **B1** | **B2** | **B3** | **B4** | **B5** | **B6** |
| **I** | **A1** | 4940 | 4810 | 5150 | 5090 | 5130 | 5140 |
| | **A2** | 4900 | 4920 | 5070 | 4890 | 5150 | 5070 |
| **II** | **A1** | 4830 | 5110 | 4920 | 4900 | 4880 | 4930 |
| | **A2** | 5020 | 5110 | 5230 | 5120 | 5160 | 5200 |
| **III** | **A1** | 5080 | 5160 | 5180 | 5190 | 5160 | 5280 |
| | **A2** | 5090 | 5130 | 4980 | 5200 | 4290 | 5250 |

Perform the analysis of the data to test the significance of effects due to residual effect method of sowing in paddy, direct effect of manurial treatments applied in wheat and interaction between method of sowing and manurial treatments.

**Analysis using JMP :**

1. As in the Factorial RCB design, the first step is to make sure that all blocking: rep (Replication), whole-plot: ms (Method of Sowing) and sub-plot: mt (Manurial Treatments) factors are nominal.
2. Open the SPLIT.xls file in JMP using **File → Open** dialog box.
3. Select **Analyze → Fit Model**.
4. Put the response variable Yield → **Y**. **Add** rep, ms, rep*ms, mt & ms*mt variables from **Select Columns List → Construct Model Effects → Emphasis → Minimal Report → Run Model**. One can take the interaction effect rep*ms by highlighting both rep & ms in **Select Column List → Cross**.
5. One can see in the output that the F-Ratio for testing for a difference between ms (Method of Sowing) uses the incorrect error term.
6. To resolve it, again go to Fit Model dialog box through windows list. Select rep*ms from construct model effects and click on the red arrow to the right of **Attributes → Random Effect**. Also change the method from **REML (Recommended) → EMS (Traditional) → Run Model**. The Model Specification dialog box should look like as in Figure 5.9.1. This should produce the output below as in Figure 5.9.2.
7. To see Effect Details click on blue arrow on right of it. Click on red arrow on the right of ms and select **LSMeans Student's t**. See Figure 5.9.2. Similarly, one can find the LSMeans differences Student's t for mt and ms*mt interaction also.

**Output :**

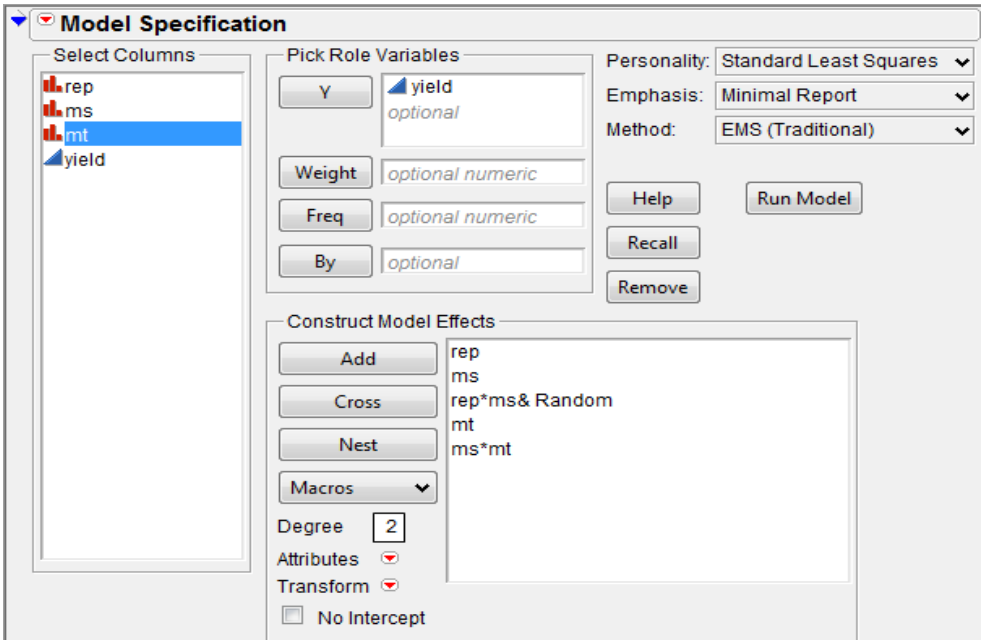
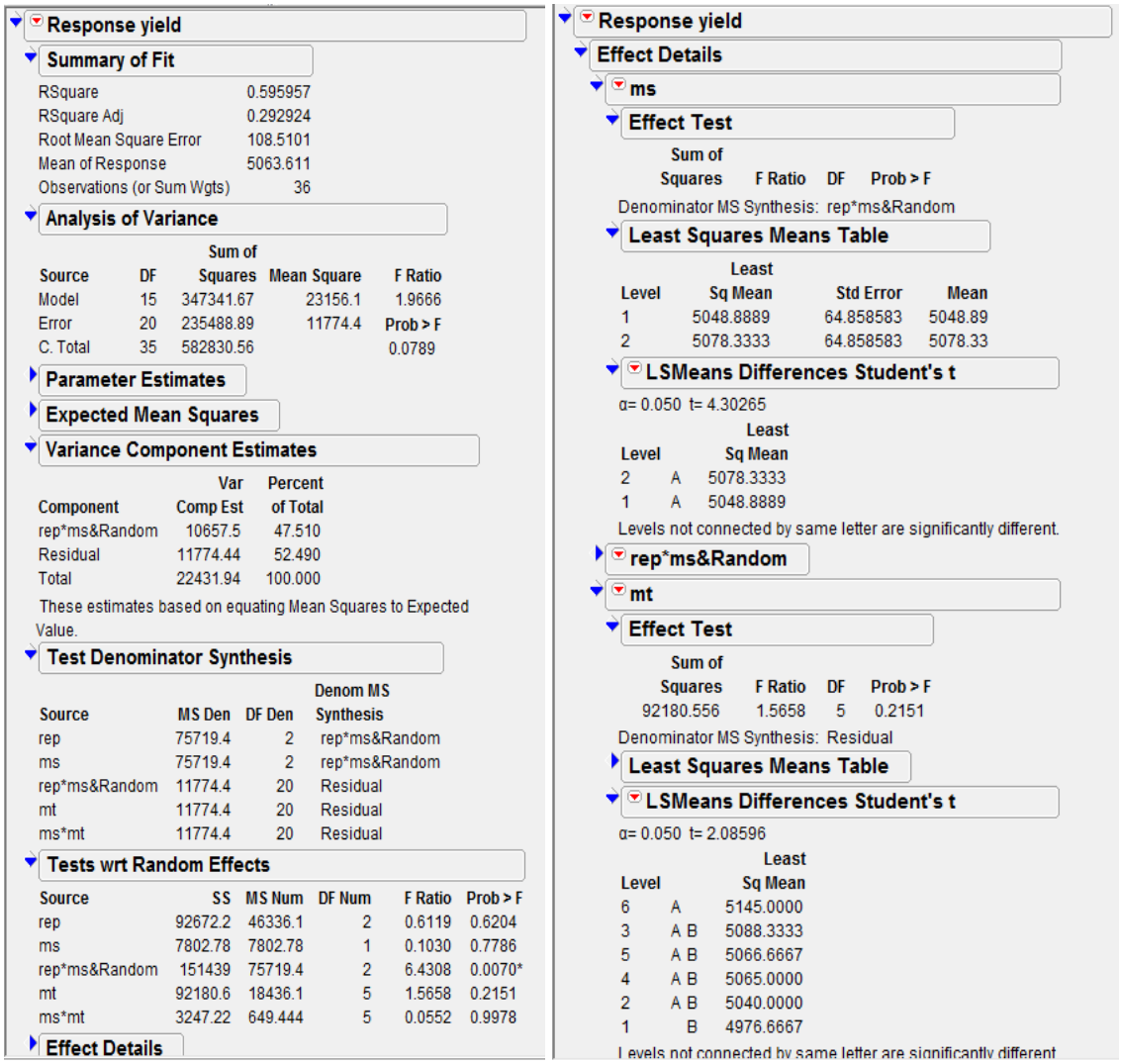



Figure 5.9.1: Model Specification Dialog Box



Response yield

Summary of Fit

| | |
|---|---|
| RSquare | 0.595957 |
| RSquare Adj | 0.292924 |
| Root Mean Square Error | 108.5101 |
| Mean of Response | 5063.611 |
| Observations (or Sum Wgts) | 36 |

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Model | 15 | 347341.67 | 23156.1 | 1.9666 |
| Error | 20 | 235488.89 | 11774.4 | Prob > F |
| C. Total | 35 | 582830.56 | | 0.0789 |

Parameter Estimates

Expected Mean Squares

Variance Component Estimates

| Component | Var Comp Est | Percent of Total |
|---|---|---|
| rep*ms&Random | 10657.5 | 47.510 |
| Residual | 11774.44 | 52.490 |
| Total | 22431.94 | 100.000 |

These estimates based on equating Mean Squares to Expected Value.

Test Denominator Synthesis

| Source | MS Den | DF Den | Denom MS Synthesis |
|---|---|---|---|
| rep | 75719.4 | 2 | rep*ms&Random |
| ms | 75719.4 | 2 | rep*ms&Random |
| rep*ms&Random | 11774.4 | 20 | Residual |
| mt | 11774.4 | 20 | Residual |
| ms*mt | 11774.4 | 20 | Residual |

Tests wrt Random Effects

| Source | SS | MS Num | DF Num | F Ratio | Prob > F |
|---|---|---|---|---|---|
| rep | 92672.2 | 46336.1 | 2 | 0.6119 | 0.6204 |
| ms | 7802.78 | 7802.78 | 1 | 0.1030 | 0.7786 |
| rep*ms&Random | 151439 | 75719.4 | 2 | 6.4308 | 0.0070* |
| mt | 92180.6 | 18436.1 | 5 | 1.5658 | 0.2151 |
| ms*mt | 3247.22 | 649.444 | 5 | 0.0552 | 0.9978 |

Effect Details

Response yield

Effect Details

ms

Effect Test

| Sum of Squares | F Ratio | DF | Prob > F |
|---|---|---|---|

Denominator MS Synthesis: rep*ms&Random

Least Squares Means Table

| Level | Least Sq Mean | Std Error | Mean |
|---|---|---|---|
| 1 | 5048.8889 | 64.858583 | 5048.89 |
| 2 | 5078.3333 | 64.858583 | 5078.33 |

LSMeans Differences Student's t

α= 0.050 t= 4.30265

| Level | | Least Sq Mean |
|---|---|---|
| 2 | A | 5078.3333 |
| 1 | A | 5048.8889 |

Levels not connected by same letter are significantly different.

rep*ms&Random

mt

Effect Test

| Sum of Squares | F Ratio | DF | Prob > F |
|---|---|---|---|
| 92180.556 | 1.5658 | 5 | 0.2151 |

Denominator MS Synthesis: Residual

Least Squares Means Table

LSMeans Differences Student's t

α= 0.050 t= 2.08596

| Level | | Least Sq Mean |
|---|---|---|
| 6 | A | 5145.0000 |
| 3 | A B | 5088.3333 |
| 5 | A B | 5066.6667 |
| 4 | A B | 5065.0000 |
| 2 | A B | 5040.0000 |
| 1 | B | 4976.6667 |

Levels not connected by same letter are significantly different

Figure 5.9.2: Split-plot Analysis and LSMeans output

## FORMULAE FOR THE STANDARD ERRORS (SPLIT PLOT DESIGN)

MSE - Mean Sum of Square,
r - number of replication
MSE(a) and MSE(b) denotes the main-plot error and sub-plot error respectively in split plot experiment.
S.E.(d) denotes the Estimate of Standard Error of the difference between the means of two treatments.
Where p is the number of main-plot treatments and q is the number of sub-plot treatments.

1) S.E.(d) between two main plot treatment means = $\sqrt{\frac{2 \times MSE(a)}{r \times q}}$

2) S.E.(d) between two sub-plot treatment means = $\sqrt{\frac{2 \times MSE(b)}{r \times p}}$

3) S.E.(d) between two main plot treatments means at the same or different levels of subplot treatment means = $\sqrt{\frac{2\{(q-1)MSE(b) + MSE(a)\}}{r \times q}}$

4) S.E.(d) between two sub-plot treatment means at the same level of main plot treatment = $\sqrt{\frac{2 \times MSE(b)}{r}}$

C.D. = S.E.(d) × t at 5% for error degree of freedom (d.f.)
In case (3), above formula for calculating C.D. cannot be used , we use the following formula

$$t^{*} = \frac{(q-1)MSE(b)\, t_b + MSE(a)\, t_b}{(q-1)MSE(b) + MSE(a)}$$

where ta and tb are the t-values at main-plot error and sub-plot error d.f. respectively at 5%.
C.D. for (c) = S.E.(d) × t*

## Standard Errors and Critical Differences using JMP Scripting Language (JSL).

```
fitm=Fit Model(
        Y( :yield ),
        Effects( :rep, :ms, :rep * :ms & Random, :mt, :ms * :mt ),
        NoBounds( 0 ),
        Personality( Standard Least Squares ),
        Method( EMS ),
        Emphasis( Minimal Report ),
        Run Model(
                :yield << {Lack of Fit( 0 ), Plot Actual by Predicted( 0 ),
                Plot Regression( 0 ), Plot Residual by Predicted( 0 ),
                Plot Effect Leverage( 0 )},
        )
);
```

```
rfitm=fitm<<report;
Model Dialog[1]<<close window;  //To close Fit Model Dialog Box Window.
mpeanova=rfitm[tablebox(6)][columnbox("MS Num")] << get as matrix;
mdf=rfitm[tablebox(6)][columnbox("DF Num")] << get as matrix;
sdf=rfitm[tablebox(2)][columnbox("DF")] << get as matrix;
speanova=rfitm[tablebox(2)][columnbox("Mean Square")] << get as matrix;
mainerr=mpeanova[3,1]; //Main Plot Error
suberr=speanova[2,1]; //Sub Plot Error
r1=mdf[1,1];
a1=mdf[2,1];
ea1=mdf[3,1]; //Main Plot error df
b1=mdf[4,1];
eb1=sdf[2,1]; //Sub Plot error df

//Levels of Replication, Main & Sub Plots
r=r1+1;
a=a1+1;
b=b1+1;

/*calculations for standard errors
S.E. of diff. b/w 2 main plot trt means*/
se_a = round(sqrt((2*mainerr)/(r*b)),4);
/*S.E. of diff. b/w 2 sub-plot trt means*/
se_b = round(sqrt((2*suberr)/(r*a)),4);
/*S.E. of diff. b/w 2 sub-plot at same level of main plot trt means*/
se_b_a = round(sqrt((2*suberr)/r),4);
/*S.E. of diff. b/w 2 main plot at same or diff level of sub-plot*/
se_a_b = round(sqrt((2*((b-1)*suberr + mainerr))/(r*b)),4);
std_error=se_a|/se_b|/se_b_a|/se_a_b; //Append for standard errors

//calculations for CD values at 5% level of sig.
ta = Students t Quantile(0.95, ea1);
tb = Students t Quantile(0.95, eb1);
pooled_t = (((b-1)*suberr*tb)+(mainerr*ta))/(((b-1)*suberr)+mainerr);
cd_a = round(se_a * ta,4);
cd_b = round(se_b * tb,4);
cd_b_a = round(se_b_a * tb,4);
cd_a_b = round(se_a_b * pooled_t,4);
cd=cd_a|/cd_b|/cd_b_a|/cd_a_b;  /*Append for cd values |/for vertical concatenation */

//calculations for CD values at 1% level of sig.
ta1 = Students t Quantile(0.99, ea1);
tb1 = Students t Quantile(0.99, eb1);
pooled_t1 = (((b-1)*suberr*tb1)+(mainerr*ta1))/(((b-1)*suberr)+mainerr);
cd_a1 = round(se_a * ta1,4);
cd_b1 = round(se_b * tb1,4);
cd_b_a1 = round(se_b_a * tb1,4);
cd_a_b1 = round(se_a_b * pooled_t1,4);
cd1=cd_a1|/cd_b1|/cd_b_a1|/cd_a_b1; /*Append for cd values |/for vertical concatenation*/
```

```
/*calculations for Tukey's HSD values at 5% level of sig.*/
qa = Tukey HSD Quantile(0.95,a,ea1);
qb = Tukey HSD Quantile(0.95,b,eb1);
pooled_q = (((b-1)*suberr*qb)+(mainerr*qa))/(((b-1)*suberr)+mainerr);
hsd_a = round(se_a * qa ,4);
hsd_b = round(se_b * qb ,4);
hsd_b_a = round(se_b_a * qb ,4);
hsd_a_b = round(se_a_b * pooled_q ,4);
hsd=hsd_a|/hsd_b|/hsd_b_a|/hsd_a_b;  /*Append for tukey's HSD values |/for vertical
concatenation*/

/*calculations for Tukey's HSD values at 1% level of sig.*/
qa1 = Tukey HSD Quantile(0.99,a,ea1);
qb1 = Tukey HSD Quantile(0.99,b,eb1);
pooled_q1 = (((b-1)*suberr*qb1)+(mainerr*qa1))/(((b-1)*suberr)+mainerr);
hsd_a1 = round(se_a * qa1 * (1/sqrt(2)),4);
hsd_b1 = round(se_b * qb1 * (1/sqrt(2)),4);
hsd_b_a1 = round(se_b_a * qb1 * (1/sqrt(2)),4);
hsd_a_b1 = round(se_a_b * pooled_q1 * (1/sqrt(2)),4);
hsd1=hsd_a1|/hsd_b1|/hsd_b_a1|/hsd_a_b1; /*Append for tukey's HSD values |/for vertical
concatenation (row wise)*/

/*display the SD & CD values in a new new window*/
name={"b/w 2 main plot","b/w 2 sub-plot","b/w 2 sub-plot at same level of main plot","b/w 2
main plot at same or diff level of sub-plot"};
se_cd_table=New Window("Split plot Standard Error and Critical Differences",
                outlinebox("Standard Error, Tukey's HSD and CD",
                tablebox(stringcolbox("Error Type",name),
                                Numbercolbox("SE",std_error),
                                Numbercolbox("HSD(5%)",hsd),
                                Numbercolbox("HSD(1%)",hsd1),
                                Numbercolbox("CD(5%)",cd),
                                Numbercolbox("CD(1%)",cd1)
        )       )       );
```

**Output:** See the Windows list and click on the JMP Report "Split Plot Standard Error and Critical Differences"

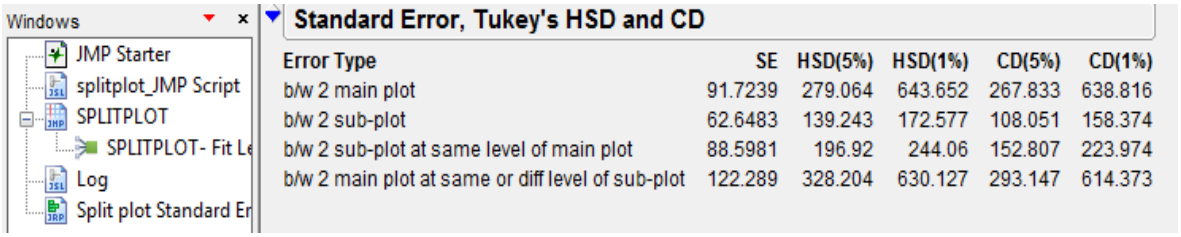



Figure 5.9.3: Split plot Standard Error and Critical Differences

**Note**: The above script can be used for any split plot data by changing the variable names in the above script according to the given situation. Also to use the above script **Replication and Main plot factor i.e. rep & ms must be in nominal scale (As ordinal terms can't be specified as random effects).**

## 5.10 Analysis of Covariance

**Example 5.10:** A trial was designed to evaluate 15 rice varieties grown in soil with a toxic level of iron. The experiment was in a RCB design with three replications. Guard rows of a susceptible check variety were planted on two sides of each experimental plot. Scores for tolerance for iron toxicity were collected from each experimental plot as well as from guard rows. For each experimental plot, the score of susceptible check (averaged over two guard rows) constitutes the value of the covariate for that plot. Data on the tolerance score of each variety (Y variable) and on the score of the corresponding susceptible check (X variable) are shown below:

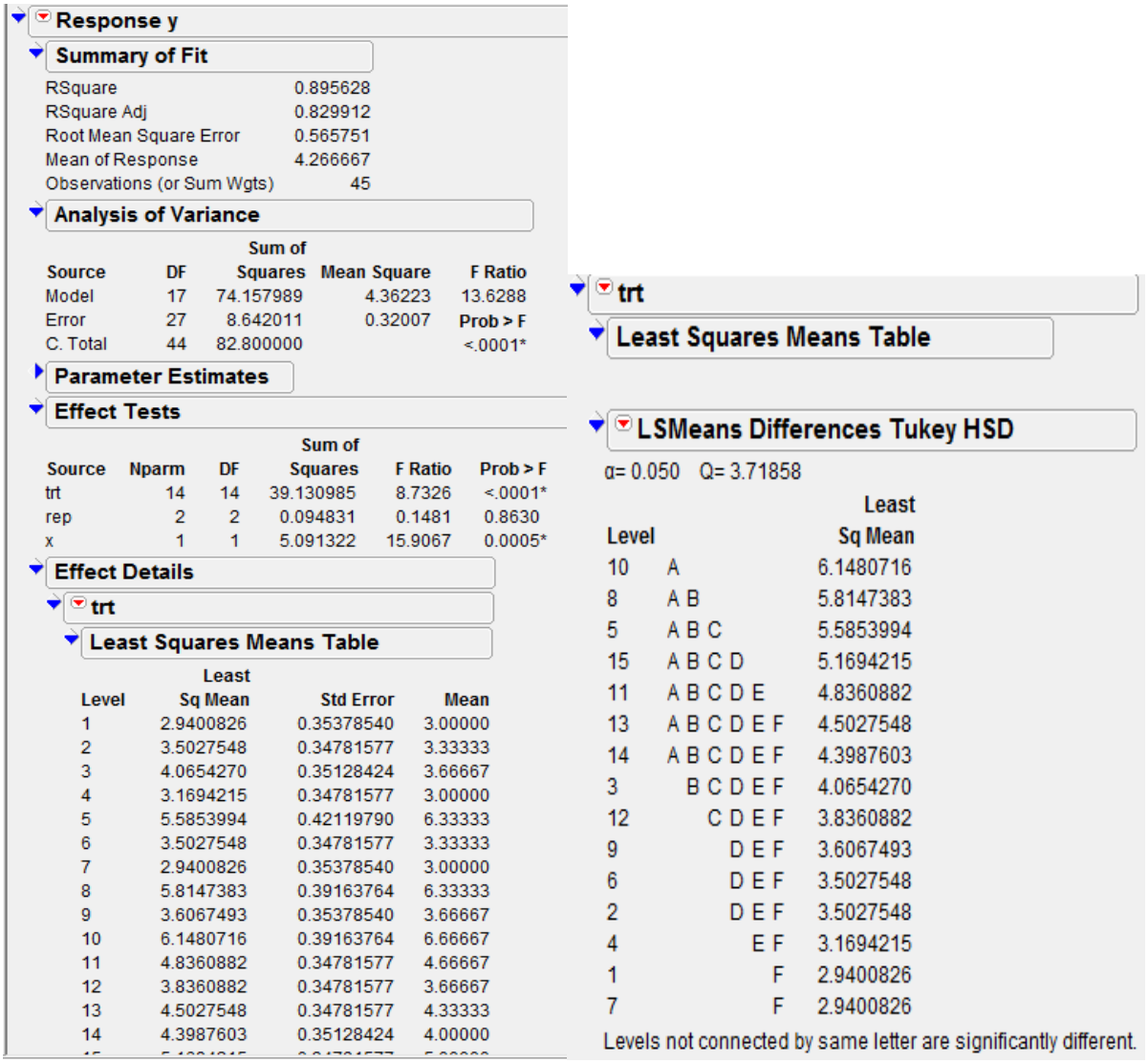
**Scores for tolerance for iron toxicity (Y) of *15* rice varieties and those corresponding guard rows of a susceptible check variety (X) in a RCB trial**

| Variety Number | Replication I | | Replication II | | Replication III | |
|---|---|---|---|---|---|---|
| | X | Y | X | Y | X | Y |
| 1 | 5 | 2 | 6 | 3 | 6 | 4 |
| 2 | 6 | 4 | 5 | 3 | 5 | 3 |
| 3 | 5 | 4 | 5 | 4 | 5 | 3 |
| 4 | 6 | 3 | 5 | 3 | 5 | 3 |
| 5 | 7 | 7 | 7 | 6 | 6 | 6 |
| 6 | 6 | 4 | 5 | 3 | 5 | 3 |
| 7 | 6 | 3 | 5 | 3 | 6 | 3 |
| 8 | 6 | 6 | 7 | 7 | 6 | 6 |
| 9 | 7 | 4 | 5 | 3 | 5 | 4 |
| 10 | 7 | 7 | 7 | 7 | 5 | 6 |
| 11 | 6 | 5 | 5 | 4 | 5 | 5 |
| 12 | 6 | 5 | 5 | 3 | 5 | 3 |
| 13 | 5 | 4 | 5 | 4 | 6 | 5 |
| 14 | 5 | 5 | 5 | 4 | 5 | 3 |
| 15 | 5 | 4 | 5 | 5 | 6 | 6 |

1. Perform analysis of covariance by taking tolerance score of each variety (Y) as dependent variable and score of the corresponding susceptible check (X) as covariate.
2. Perform all possible pair wise variety comparisons and identify the best variety.

**Analysis using JMP**

1. As in the Split plot, the first step is to make sure that rep (Replication) and trt (Treatment) factors are nominal.
2. Open the ANCOVA.xls file in JMP using **File → Open** dialog box.
3. Select **Analyze → Fit Model**.
4. Put the response variable y → **Y**. **Add** rep, trt and x variables from **Select Columns List → Construct Model Effects → Run Model**. (See Figure 5.10.1). One can also see Tukey's Means Comparison by selecting **LSMeans Tukey's HSD**. (See Figure 5.10.1)



**Response y**

**Summary of Fit**

| | |
|---|---|
| RSquare | 0.895628 |
| RSquare Adj | 0.829912 |
| Root Mean Square Error | 0.565751 |
| Mean of Response | 4.266667 |
| Observations (or Sum Wgts) | 45 |

**Analysis of Variance**

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Model | 17 | 74.157989 | 4.36223 | 13.6288 |
| Error | 27 | 8.642011 | 0.32007 | Prob > F |
| C. Total | 44 | 82.800000 | | <.0001* |

**Parameter Estimates**

**Effect Tests**

| Source | Nparm | DF | Sum of Squares | F Ratio | Prob > F |
|---|---|---|---|---|---|
| trt | 14 | 14 | 39.130985 | 8.7326 | <.0001* |
| rep | 2 | 2 | 0.094831 | 0.1481 | 0.8630 |
| x | 1 | 1 | 5.091322 | 15.9067 | 0.0005* |

**Effect Details**

**trt**

**Least Squares Means Table**

| Level | Least Sq Mean | Std Error | Mean |
|---|---|---|---|
| 1 | 2.9400826 | 0.35378540 | 3.00000 |
| 2 | 3.5027548 | 0.34781577 | 3.33333 |
| 3 | 4.0654270 | 0.35128424 | 3.66667 |
| 4 | 3.1694215 | 0.34781577 | 3.00000 |
| 5 | 5.5853994 | 0.42119790 | 6.33333 |
| 6 | 3.5027548 | 0.34781577 | 3.33333 |
| 7 | 2.9400826 | 0.35378540 | 3.00000 |
| 8 | 5.8147383 | 0.39163764 | 6.33333 |
| 9 | 3.6067493 | 0.35378540 | 3.66667 |
| 10 | 6.1480716 | 0.39163764 | 6.66667 |
| 11 | 4.8360882 | 0.34781577 | 4.66667 |
| 12 | 3.8360882 | 0.34781577 | 3.66667 |
| 13 | 4.5027548 | 0.34781577 | 4.33333 |
| 14 | 4.3987603 | 0.35128424 | 4.00000 |

**trt**

**Least Squares Means Table**

**LSMeans Differences Tukey HSD**

α= 0.050 Q= 3.71858

| Level | | Least Sq Mean |
|---|---|---|
| 10 | A | 6.1480716 |
| 8 | A B | 5.8147383 |
| 5 | A B C | 5.5853994 |
| 15 | A B C D | 5.1694215 |
| 11 | A B C D E | 4.8360882 |
| 13 | A B C D E F | 4.5027548 |
| 14 | A B C D E F | 4.3987603 |
| 3 | B C D E F | 4.0654270 |
| 12 | C D E F | 3.8360882 |
| 9 | D E F | 3.6067493 |
| 6 | D E F | 3.5027548 |
| 2 | D E F | 3.5027548 |
| 4 | E F | 3.1694215 |
| 1 | F | 2.9400826 |
| 7 | F | 2.9400826 |

Levels not connected by same letter are significantly different.

Figure 5.10.1: ANCOVA Analysis and LSMeans Output

## 5.11 Multivariate Analysis of Variance (MANOVA)

**Example 5.11:** A researcher randomly assigns 33 subjects to one of three groups. The first group receives technical dietary information interactively from an on-line website. Group 2 receives the same information in from a nurse practitioner, while group 3 receives the information from a video tape made by the same nurse practitioner. The researcher looks at three different ratings of the presentation, difficulty, useful and importance, to determine if there is a difference in the modes of presentation. In particular, the researcher is interested in whether the interactive website is superior because that is the most cost-effective way of delivering the information. The following data is taken from (http://www.ats.ucla.edu/stat/sas/dae/manova1.htm).

| group | useful | difficulty | importance |
|---|---|---|---|
| 1 | 19.6 | 5.1500001 | 9.5 |
| 1 | 15.4 | 5.75 | 9.1000004 |
| 1 | 22.299999 | 4.3499999 | 3.3 |
| 1 | 24.299999 | 7.5500002 | 5 |

| | | | |
|---|---|---|---|
| 1 | 22.5 | 8.5 | 6 |
| 1 | 20.5 | 10.25 | 5 |
| 1 | 14.1 | 5.9499998 | 18.799999 |
| 1 | 13 | 6.3000002 | 16.5 |
| 1 | 14.1 | 5.4499998 | 8.8999996 |
| 1 | 16.700001 | 3.75 | 6 |
| 1 | 16.799999 | 5.0999999 | 7.4000001 |
| 2 | 17.1 | 9 | 7.5 |
| 2 | 15.7 | 5.3000002 | 8.5 |
| 2 | 14.9 | 9.8500004 | 6 |
| 2 | 19.700001 | 3.5999999 | 2.9000001 |
| 2 | 17.200001 | 4.0500002 | 0.2 |
| 2 | 16 | 4.4000001 | 2.5999999 |
| 2 | 12.8 | 7.1500001 | 7 |
| 2 | 13.6 | 7.25 | 3.2 |
| 2 | 14.2 | 5.3000002 | 6.1999998 |
| 2 | 13.1 | 3.0999999 | 5.5 |
| 2 | 16.5 | 2.4000001 | 6.5999999 |
| 3 | 16 | 4.5500002 | 2.9000001 |
| 3 | 12.5 | 2.6500001 | 0.7 |
| 3 | 18.5 | 6.5 | 5.3000002 |
| 3 | 19.200001 | 4.8499999 | 8.3000002 |
| 3 | 12 | 8.75 | 9 |
| 3 | 13 | 5.1999998 | 10.3 |
| 3 | 11.9 | 4.75 | 8.5 |
| 3 | 12 | 5.8499999 | 9.5 |
| 3 | 19.799999 | 2.8499999 | 2.3 |
| 3 | 16.5 | 6.5500002 | 3.3 |
| 3 | 17.4 | 6.5999999 | 1.9 |

Perform Multivariate Analysis of variance (MANOVA) on the above data.

**Analysis using JMP**

1. Open MANOVA.xls file in JMP using **File** → **Open** dialog box.
2. Make sure 'group' variable is nominal.
3. Select **Analyze** → **Fit Model**.
4. Put the response variable useful, difficulty & importance → **Y** and **Add** group to **Construct Model Effects**. Select **Manova** from **Personality** → **Run Model**. See Figure 5.11.1 for Initial Manova fit.
5. Click **Choose Response** → **Identity** → **Run**. See the Figure 5.11.1.

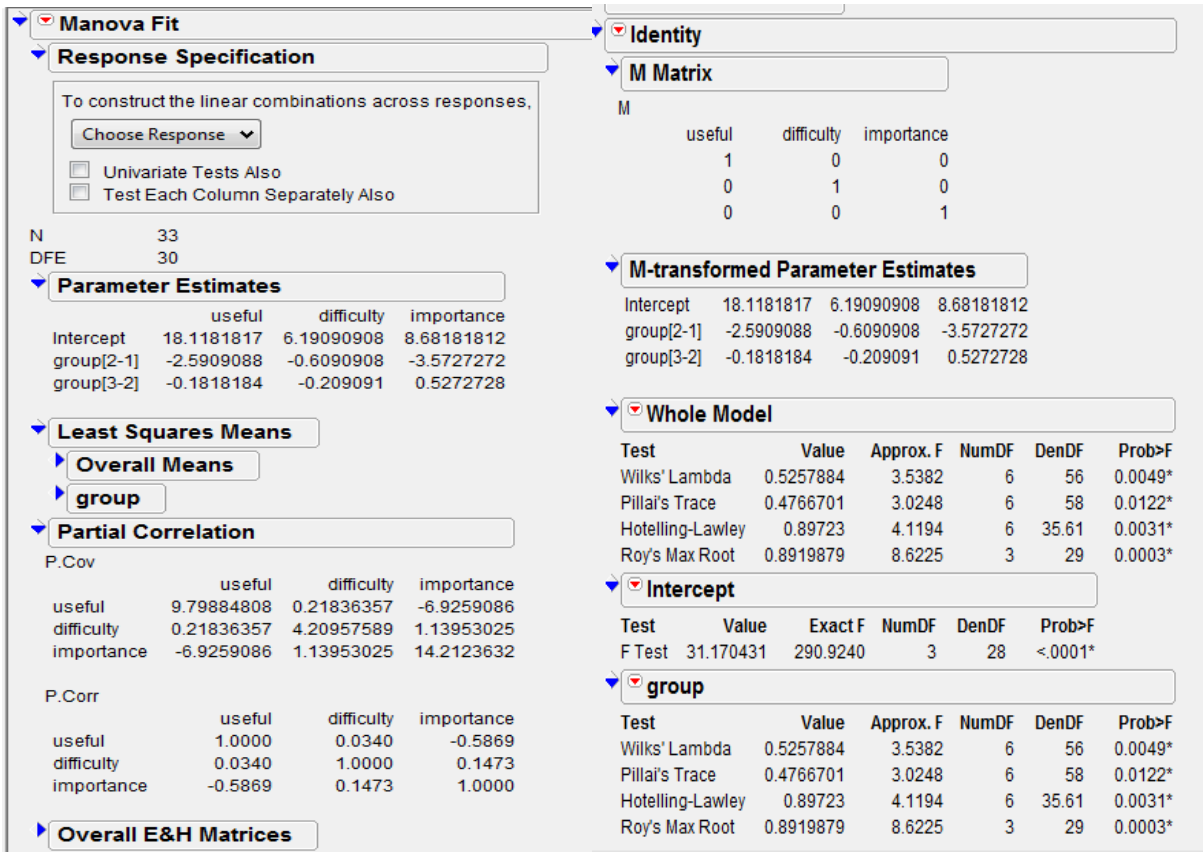



Figure 5.11.1: MANOVA Analysis Output

## 5.12 α-Designs

**Example 5.12:** An initial varietal trial was conducted to study the performance of 21 new strains of Toria vis-a-vis 3 checks using an alpha design at Pantnagar with three replications. The seed yield in kg/ha was recorded. The details of strains, design adopted and data obtained are given as under.

| Treatment | Treatment No. | Treatment | Treatment No. | Treatment | Treatment No. |
|---|---|---|---|---|---|
| RAU DT-01-03 | 1 | TK-06-1 | 9 | RH-0304 | 17 |
| RAU DT-01-02 | 2 | TK-06-2 | 10 | TH-0302 | 18 |
| BAUSM-92-24 | 3 | TL-2027 | 11 | JMT-05 | 19 |
| RGN 186 | 4 | TL-2013 | 12 | **PT-303(NC)** | **20** |
| EJ-17 | 5 | JMT-02-6 | 13 | **Zonal Check** | **21** |
| NPJ-112 | 6 | NDT 05-5 | 14 | PTC-99-14 | 22 |
| VLT-4 | 7 | NDRE 200216 | 15 | **JD-6(check)** | **23** |
| RRN-612 | 8 | PT-2004-3 | 16 | ORT 17-6-16 | 24 |

**Note: strains of toria in boldface are the three checks, i.e., treatment numbers 20, 21 and 23 are checks**

## Replication 1

| | | | | | | |
|---|---|---|---|---|---|---|
| **Block 1** | 1 (1555.6) | 5 (1160.5) | 9 (1308.6) | 13 (1382.7) | 17 (987.7) | 21 (1135.8) |
| **Block 2** | 2 (1284.0) | 6 (1086.4) | 10 (1284.0) | 14 (1111.1) | 18 (938.3) | 22 (1308.6) |
| **Block 3** | 3 (1234.6) | 7 (419.8) | 11 (1308.6) | 15 (963.0) | 19 (963.0) | 23 (987.7) |
| **Block 4** | 4 (1234.6) | 8 (987.7) | 12 (1284.0) | 16 (913.6) | 20 (1160.5) | 24 (790.1) |

## Replication 2

| | | | | | | |
|---|---|---|---|---|---|---|
| **Block 1** | 1 (1481.5) | 6 (1086.4) | 11 (1308.6) | 16 (1284.0) | 19 (1111.1) | 22 (1185.2) |
| **Block 2** | 2 (987.7) | 7 (308.6) | 12 (1234.6) | 13 (1308.6) | 20 (765.4) | 23 (938.3) |
| **Block 3** | 3 (1012.3) | 8 (864.2) | 9 (1234.6) | 14 (938.3) | 17 (913.6) | 24 (864.2) |
| **Block 4** | 4 (1135.8) | 5 (987.7) | 10 (987.7) | 15 (740.7) | 18 (963.0) | 21 (1135.8) |

## Replication 3

| | | | | | | |
|---|---|---|---|---|---|---|
| **Block 1** | 1 (1284.0) | 7 (333.3) | 12 (1135.8) | 15 (839.5) | 18 (814.8) | 24 (888.9) |
| **Block 2** | 2 (1135.8) | 8 (913.6) | 9 (1456.8) | 16 (1037) | 19 (938.3) | 21 (1037.0) |
| **Block 3** | 3 (963.0) | 5 (1209.9) | 10 (1259.3) | 13 (1234.6) | 20 (963.0) | 22 (1111.1) |
| **Block 4** | 4 (1086.4) | 6 (765.4) | 11 (1111.1) | 14 (1037.0) | 17 (938.3) | 23 (938.3) |

*Figures in the parenthesis gives the seed yield in kg/ha.*

1. Perform the analysis of variance of the data to test whether there is any difference between treatment effects.
2. Obtain the adjusted treatment means.
3. Test whether there is any significant different treatment from the best performing test.

**Analysis using JMP**

1. First open **Alpha_des.sas7bdat** SAS dataset.
2. Make sure variables rep, blk, trt should be in nominal scale.
3. Select **Analyze → Fit model**.
4. Put the response variable, syield → **Y**. **Add** trt, rep (outside effect) and blk (nested effect) variables from **Select Columns** List → **Construct Model Effects**. Now, Select rep in **Select Columns** List and select blk in **Construct Model Effects** and click **Nest → Run Model**. See Fit Model dialog box as in Figure 5.12.1. See the output as in Figure 5.12.2.

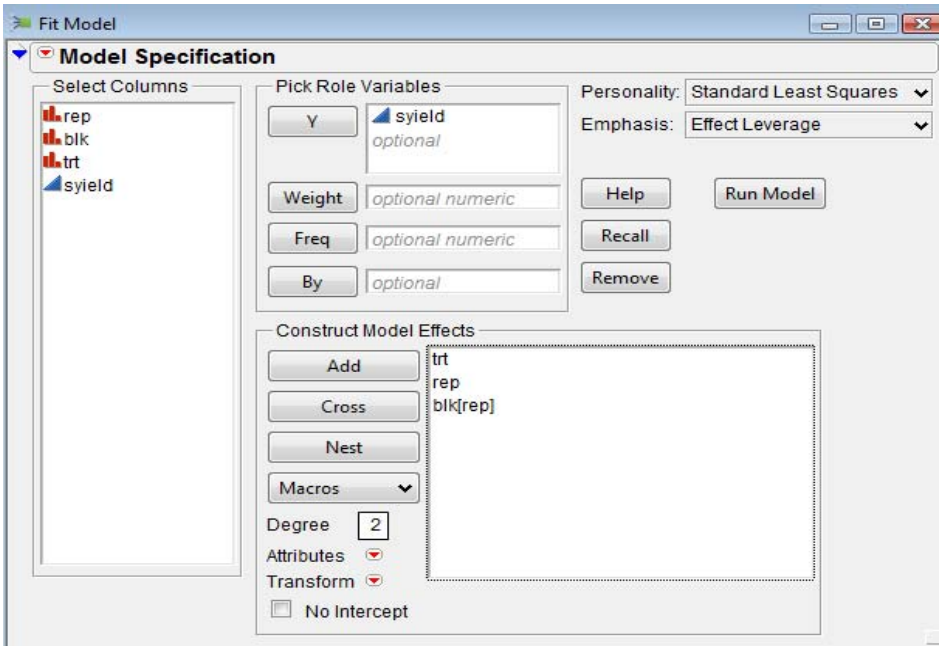



Figure 5.12.1: Fit Model dialog box

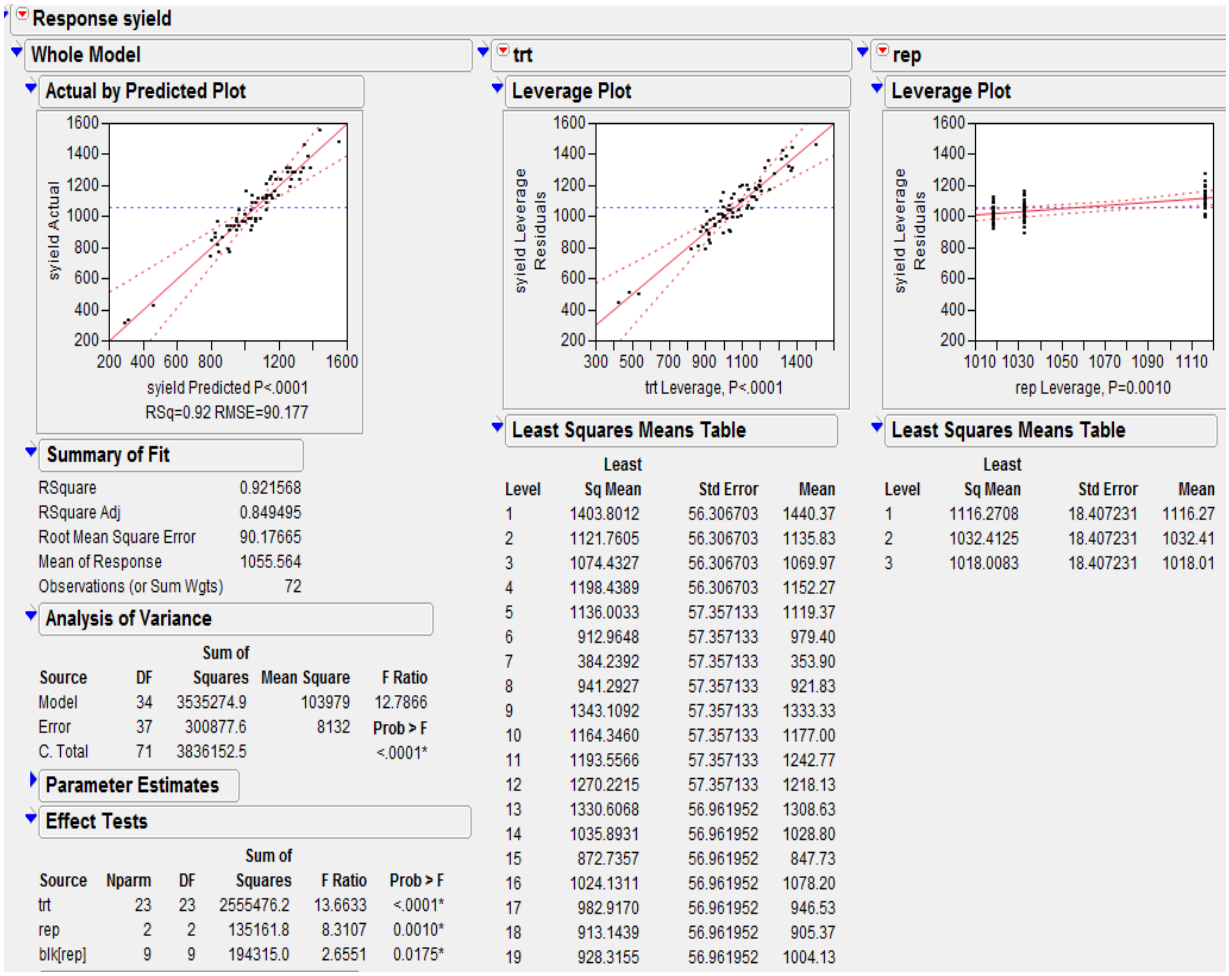



Figure 5.12.2: Alpha-Design Analysis Output

## 5.13 Response Surface Design

**Example 5.13:** An experiment was conducted at Division of Agronomy, Indian Agricultural Research Institute, New Delhi to obtain the optimum combination of Nitrogen (4 levels: 0, 50, 100, 150 kg/ha) and Sulphur (4 levels: 0, 20, 40, 60 kg/ha) for maximizing the yield of paddy crop. The plot size used was $4 \times 2.2$ $m^2$.The experiment was conducted using a RCB design in 3 replications. The analysis of variance of the data revealed that replications are not significantly different, i.e. the replication mean square was small in comparison to error mean square. The main objective of the experiment was to obtain the optimum combination of nitrogen and sulphur that maximizes the yield. The treatment combinations tried and average yield of paddy in kg/ha are:

| Nitrogen | Sulphur | Yield | | Nitrogen | Sulphur | Yield |
|---|---|---|---|---|---|---|
| 0 | 0 | 4121.21 | | 100 | 0 | 6761.36 |
| 0 | 20 | 4678.03 | | 100 | 20 | 6916.67 |
| 0 | 40 | 4742.42 | | 100 | 40 | 6852.27 |
| 0 | 60 | 4727.27 | | 100 | 60 | 6810.61 |
| 50 | 0 | 6083.33 | | 150 | 0 | 6174.24 |
| 50 | 20 | 6041.67 | | 150 | 20 | 7022.73 |
| 50 | 40 | 6223.49 | | 150 | 40 | 7003.79 |
| 50 | 60 | 6715.91 | | 150 | 60 | 6943.18 |

1. Fit a second order response surface using the above data.
2. Obtain the co-ordinates of the stationary point.
3. Also find the nature of the stationary point.

**Analysis using JMP**

1. Open RSD_NS.xls file in JMP using **File → Open** dialog box.
2. Select **Analyze → Fit Model**.
3. Put yield → **Y**. Select N and S from **Select Column** list, click **Macros → Response Surface**. ***An important thing to note that by default, JMP does the coding on quantitative variables. To avoid that click on hotspot button on the left of Model Specification and uncheck Center Polynomials***. The window will look as shown in Figure 5.13.1. Click **Run Model**. See the output as shown in Figure 5.13.2.
4. If one wants to get Surface Profiler plot then go to hotspot button on the left of **Response Yield → Factor Profiling → Surface Profiler**. See Figure 5.13.3 for surface profiler output.
5. If one wants to get the Prediction Formula then go to hotspot button **Response Yield → Estimates → Show Prediction Expression**. One can also save the predicted values from **Response Yield → Save Columns → Prediction Formula**. A new data table column called **Pred Formula yield** saves the prediction formula using the coefficients from the Parameter Estimates table.

Click the Red button and uncheck **Center Polynomials.**

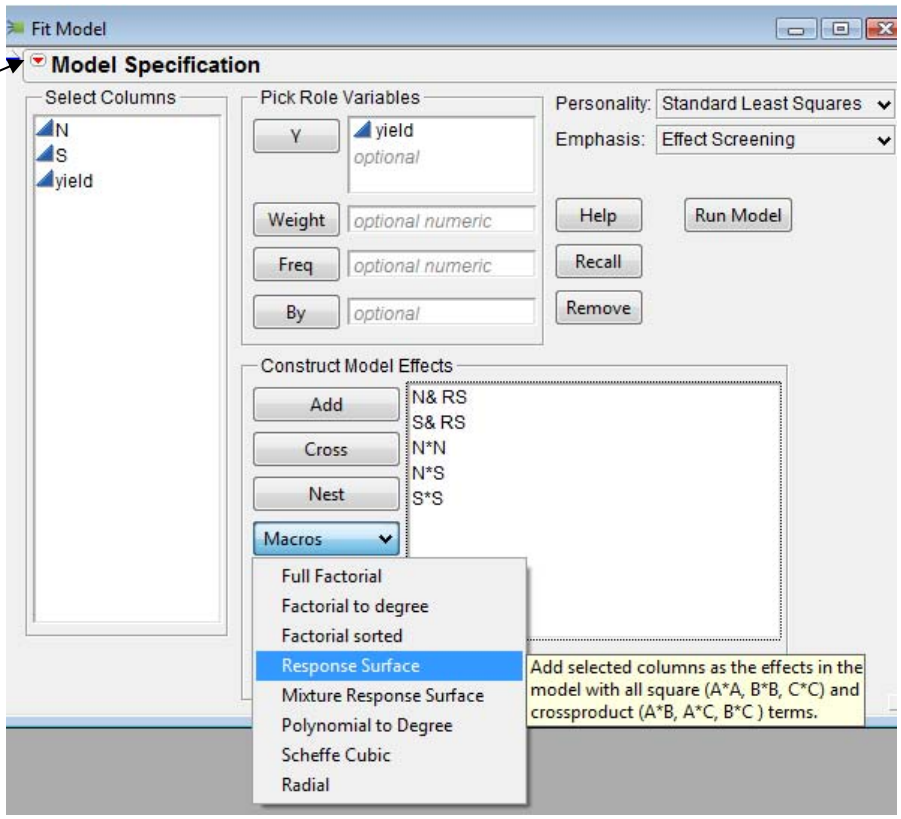



Figure 5.13.1: Response Surface Dialog Box

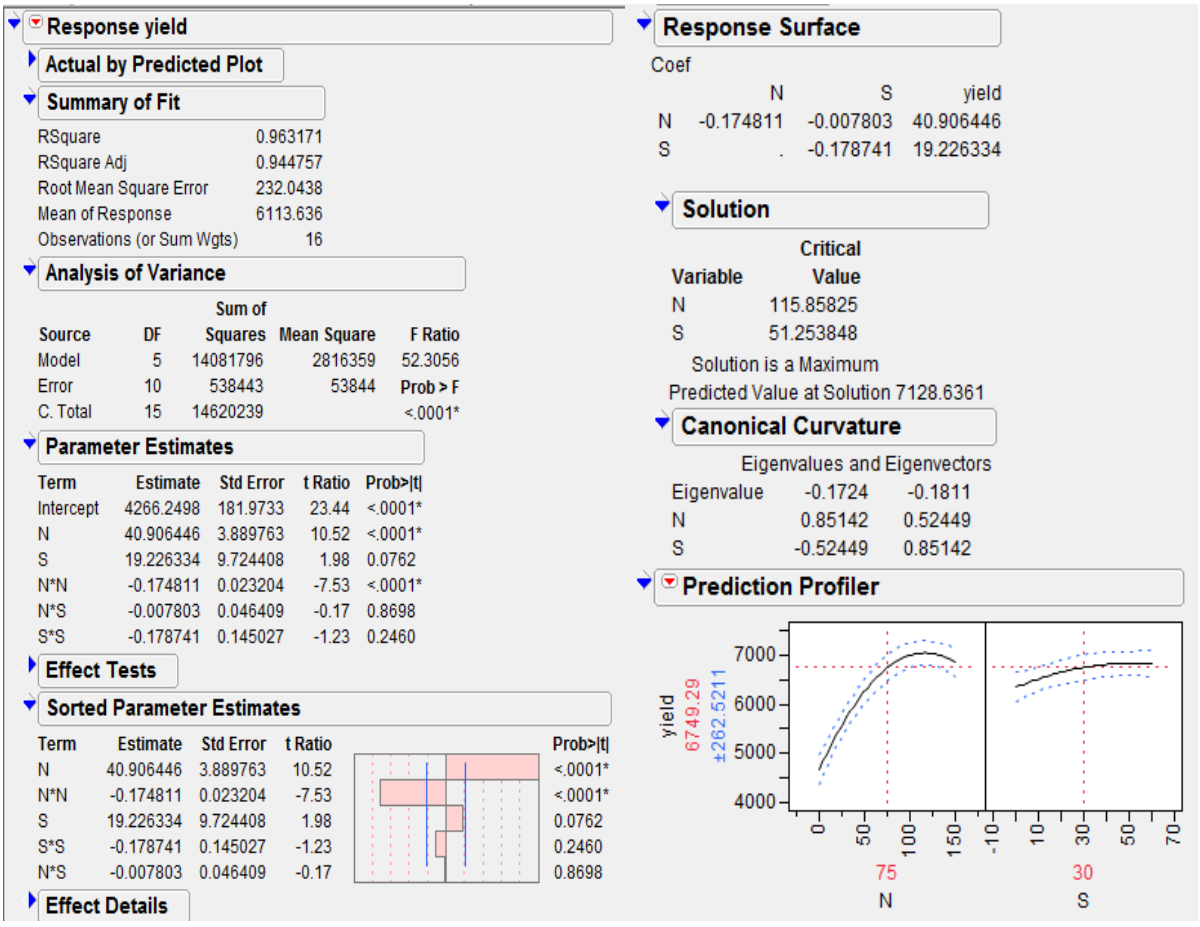



Figure 5.13.2: Response Surface Design Output

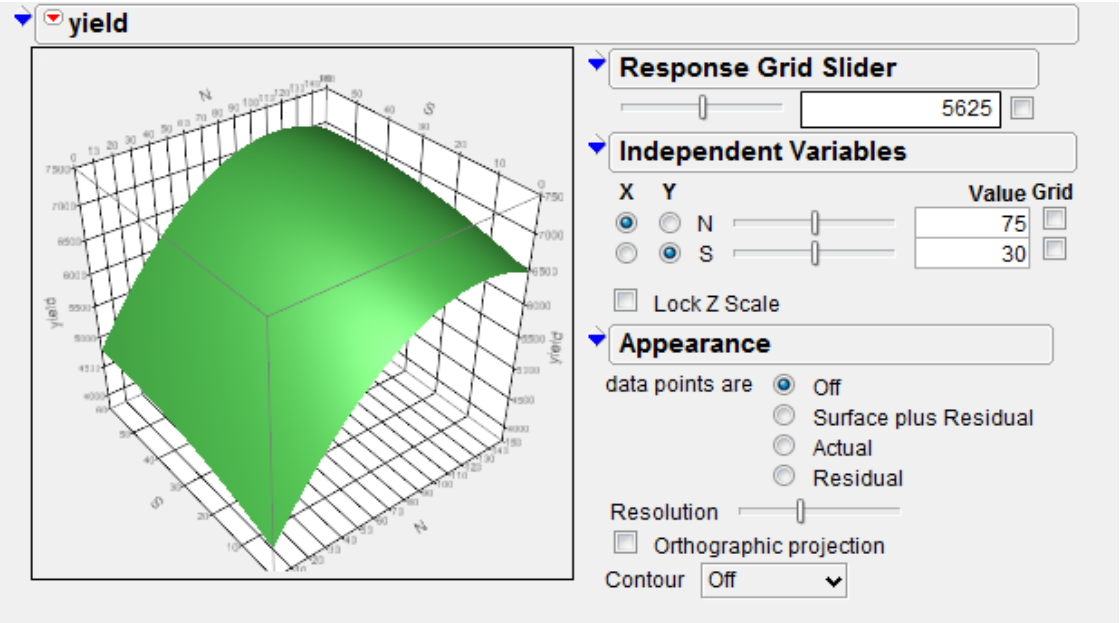



Figure 5.13.3: Surface Profiler Output

## 5.14 NonParametric Tests

### a) One Sample

**Example 5.14a:** Following data is related to the length(in cm) of the ear-head of a wheat variety 9.3, 18.8, 10.7, 11.5, 8.2, 9.7, 10.3, 8.6, 11.3, 10.7, 11.2, 9.0, 9.8, 9.3, 10.3, 10, 10.1 9.6, 10.4. Test the data that the median length of ear-head is 9.9 cm.

**Analysis using JMP**

1. Open NP_ONESAMPLE.xls file in JMP using **File** → **Open** dialog box.
2. Go to **Analyze** → **Distribution**. Put Length → **Y, Columns**. Click the hotspot button on the left of Length (cm) and select **Test Mean** option. In the Test mean dialog box, put 9.9 as specify hypothesized mean. Check the box for **Wilcoxon Signed Rank**. Click **OK**.

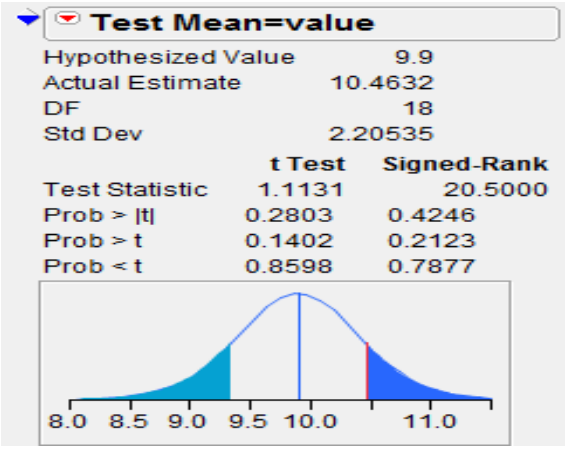



Figure 5.14a.1: One sample Wilcoxon Signed-Rank Test Output

**b) k-Independent Samples Test**

**Excersise 5.14b:** An experiment was conducted with 21 animals to determine if the four different feeds have the same distribution of Weight gains on experimental animals. The feeds 1, 3 and 4 were given to 5 randomly selected animals and feed 2 was given to 6 randomly selected animals. The data obtained is presented in the following table.

| Feeds | Weight gains (kg) | | | | | |
|---|---|---|---|---|---|---|
| 1 | 3.35 | 3.8 | 3.55 | 3.36 | 3.81 | |
| 2 | 3.79 | 4.1 | 4.11 | 3.95 | 4.25 | 4.4 |
| 3 | 4 | 4.5 | 4.51 | 4.75 | 5 | |
| 4 | 3.57 | 3.82 | 4.09 | 3.96 | 3.82 | |

**Analysis using JMP**

1. Open NPKRUS.xls file in JMP using **File** → **Open** dialog box.
2. Make sure variable 'feed' is nominal.
3. Go to **Analyze** → **Fit Y by X**. Put wt → **Y, Response** and feed → **X, Factor** box → **OK**. Now in the output, click on the hotspot button on the left of **Oneway Analysis** → **Nonparametric** → **Wilcoxon Test**. If there are two groups it displays a **Wilcoxon Rank Sums Test** and a **Kruskal-Wallis** nonparametric one-way analysis of variance if there are more than two groups. Note that the Wilcoxon test is equivalent to the Mann-Whitney U-statistic. See the output in Figure 15.4b.1.

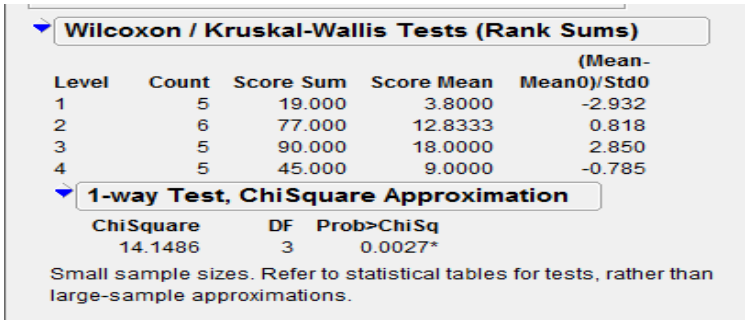



Figure 5.14b.1: k-Independent samples nonparametric Kruskal-Wallis Test

**c) Two Sample Median Test**

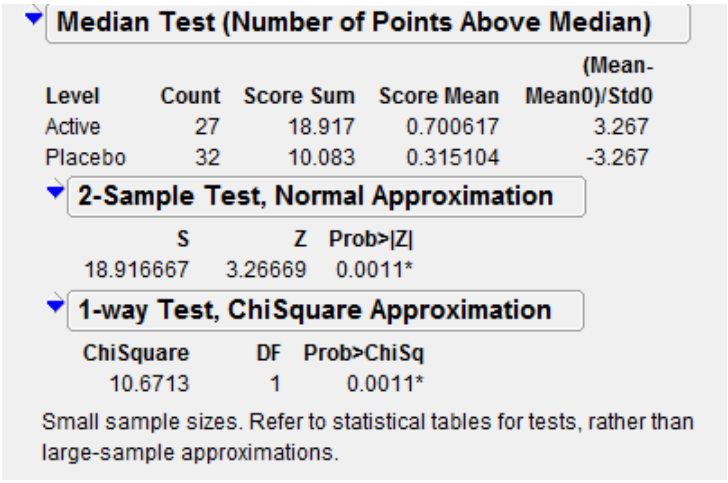
**Excercise 5.14c:** Fifty-nine female patients with rheumatoid arthritis who participated in a clinical trial were assigned to two groups, active and placebo. The response status (excellent=5, good=4, moderate=3, fair=2, poor=1) of each patient was recorded.

| Treatment | Response | Frequency |
|---|---|---|
| Active | 5 | 5 |
| Active | 4 | 11 |
| Active | 3 | 5 |
| Active | 2 | 1 |
| Active | 1 | 5 |
| Placebo | 5 | 2 |
| Placebo | 4 | 4 |
| Placebo | 3 | 7 |
| Placebo | 2 | 7 |
| Placebo | 1 | 12 |

Test for difference in location using Median Test.

**Analysis using JMP**

1. Open MEDIAN_TEST.xls file in JMP using **File** → **Open** dialog box.
2. Make sure variable 'Treatment' is nominal.
3. Go to **Analyze** → **Fit Y by X**. Put Response → **Y, Response**, Treatment → **X, Factor** box and Frequency → **Freq** → **OK**. Now in the output, click on the hotspot button on the left of **Oneway Analysis** → **Nonparametric** → **Median Test**. For k-samples it displays **Brown-Mood median test** to compare group means.



**Median Test (Number of Points Above Median)**

| Level | Count | Score Sum | Score Mean | (Mean-Mean0)/Std0 |
|---|---|---|---|---|
| Active | 27 | 18.917 | 0.700617 | 3.267 |
| Placebo | 32 | 10.083 | 0.315104 | -3.267 |

**2-Sample Test, Normal Approximation**

| S | Z | Prob>\|Z\| |
|---|---|---|
| 18.916667 | 3.26669 | 0.0011* |

**1-way Test, ChiSquare Approximation**

| ChiSquare | DF | Prob>ChiSq |
|---|---|---|
| 10.6713 | 1 | 0.0011* |

Small sample sizes. Refer to statistical tables for tests, rather than large-sample approximations.

Figure 5.14c.1: Two Sample Median Test Output

## 6. Some Differences Between JMP and Excel

There are a number of important differences between JMP and Excel or other spreadsheet applications.

| JMP | Excel |
|---|---|
| Formulas are applied only to entire columns. | Formulas are applied to individual cells. |
| Column names are not part of the grid. | Column names are part of the grid. |
| There are no rows and columns beyond the existing data. The grid is only as big as the data. | Numbered rows and labeled columns extend past the data |
| A column is either numeric or character. If a column contains both character and numeric data, the entire column's data type is character, and the numbers are treated as character data. | Numeric and character data reside in the same column |
| It does not have the concept of worksheets. Each data table is a separate with separate name with extension * .jmp file and appears in a separate window. | A single spreadsheet contains several tables, or worksheets. Each sheet have different name. By default Sheet1, Sheet2 ... etc, which can be changed. |
| Data always begins in row 1 and column 1. | Data can be located anywhere in the data grid. |
| Results appear in a separate window. | All data, analyses, and graphs are placed inside the data grid. |

## 7. JMP Journals

The journal is a copy of the report window. It consists of JMP graphs and reports, graphics, text, link to web page and files. Save platform reports for future viewing by creating a journal of the report window. One can edit or append additional reports to an existing journal. The journal is not connected to the data table. A journal is an easy way to save the results from several report windows in a single report window that you can share with others. Journals has several advantages:

- It saves relevant graphs and reports and then arranges the contents as we want.
- It stores information presentation that one can edit and interact with contents such as scrolling an axis or changing marker size.
- It is an intermediate format to export to Microsoft Word, HTML or other format.
- It stores data tables and reports from a session, close and reopen JMP and then open the journal file instantly.

### Creating a Journal

1. **File→ Edit→ Journal or** from the JMP Starter window, select New Journal. A blank Journal window will get opened.
2. Open data table go to top menu select **Edit→ Journal** . This data table automatically added to the journal.
3. To create a special report one can select the part or full portion of report by using selection tool and then copy and paste the same in the journal.

### Create and Modify a Report and Save It to a Journal

1) By using descriptive _stat.xls file as given in Example 5.1, let us create a descriptive statistics report:
2) Open the descriptive_stat.xls or descriptive_stat.jmp data table.
3) Select **Analyze →Distribution**.
4) Select both **fs45** and click **Y, Columns**.
5) Click **OK**.
6) From the red triangle menu for **fs45**, select **stem and leaf and CDF Plot**.
7) From the red triangle menu for syp, select **stem and leaf.**

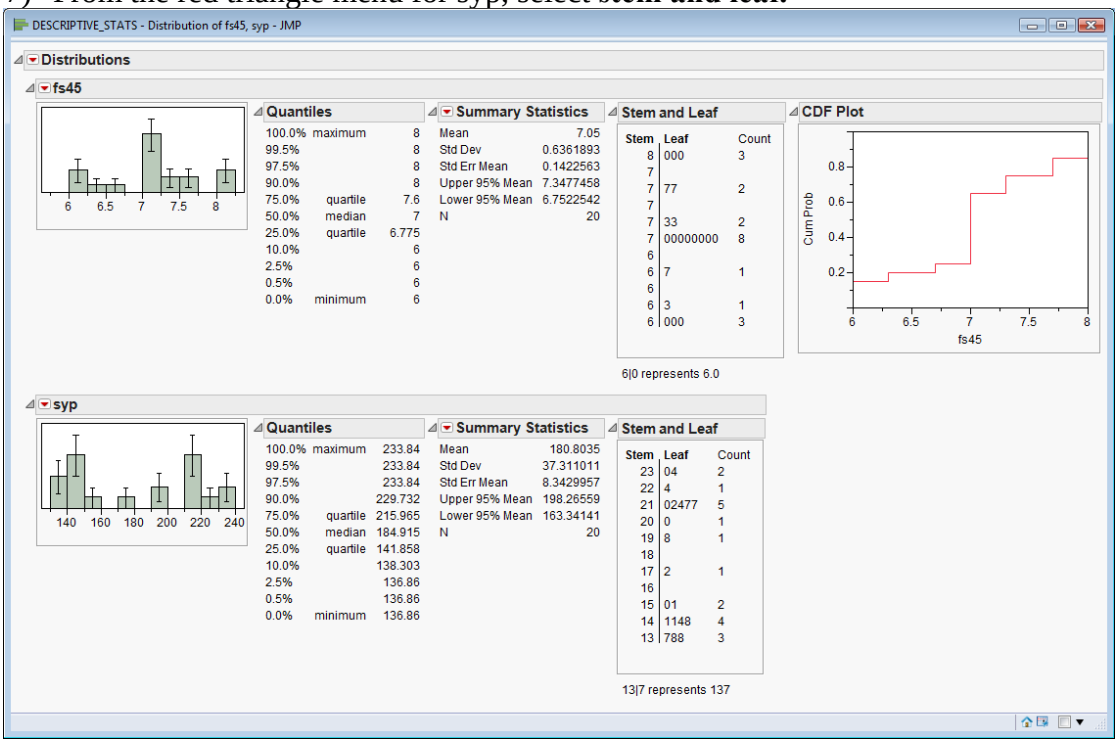



Figure 7.1: Report on Descriptive Statistics

a) Open the data table from log window, Select **Edit →Journal** to journal these results. A journal window will open in which data of data table get copied automatically.
b) Go to top menu of data table tool, select the cross mark ( called selection tool)
c) Now open the descriptive statistics report created just above, click the tool ( cross mark) to select the report or portion of report we want to add to journal.
d) Press Ctrl +C to copy the report.

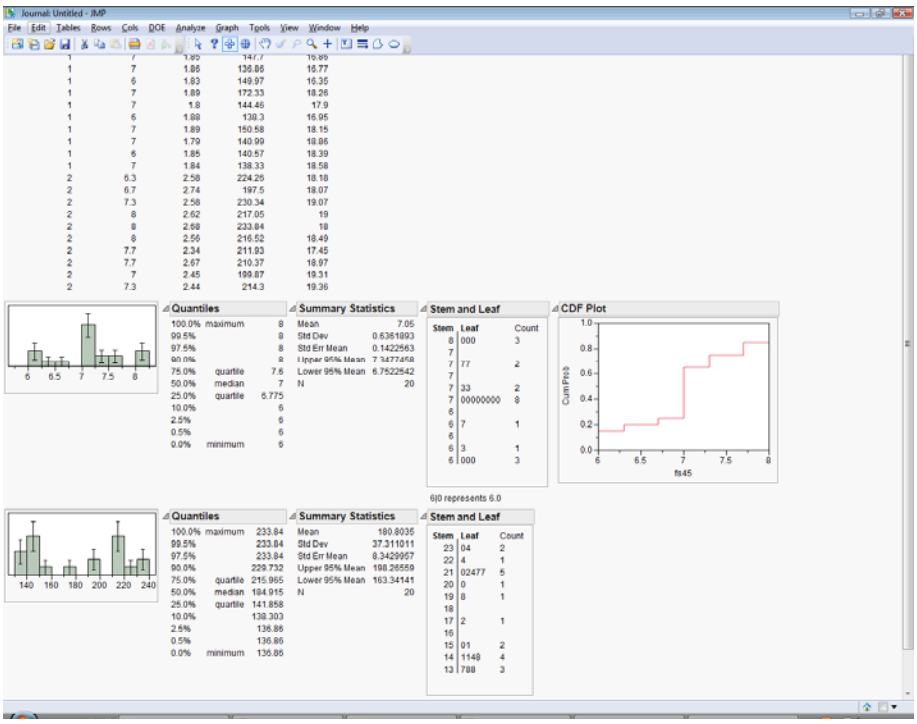



Figure 7.2: Journal on Descriptive Statistics

To save the journal, Click in **File→ Save as→ Select the location to save (Desktop)→ Give Journal name (des_stat→ Click Save**. Journal will saved with extension *.jrn. see snapshot given below

## 8. Integrating JMP and SAS

To use SAS through JMP, one must have access to SAS, either on local machine or on a server. Using JMP, one can interact with SAS as follows:

- Write or create SAS code in JMP.
- Submit SAS code and view the results in JMP.
- Connect to a SAS Metadata Server or a SAS Server on a remote machine.
- Connect to SAS on your local machine.
- Open and browse SAS data sets.
- Retrieve and view data sets generated by SAS.

### Example 8.1.1: Creating SAS Code

Create a data table RBD and follow the analysis steps as explained in Example 5.7 and then for creating SAS Code use the following steps

Figure 8.1.1: Data Table

From the red triangle menu for Model Specification, select **Create SAS Job** (see snapshot given left side below). In response, a window appears on the desktop (as shown in right side snap shot ) in which SAs code get generated automatically. Select 'Submit Button' to obtain the results.

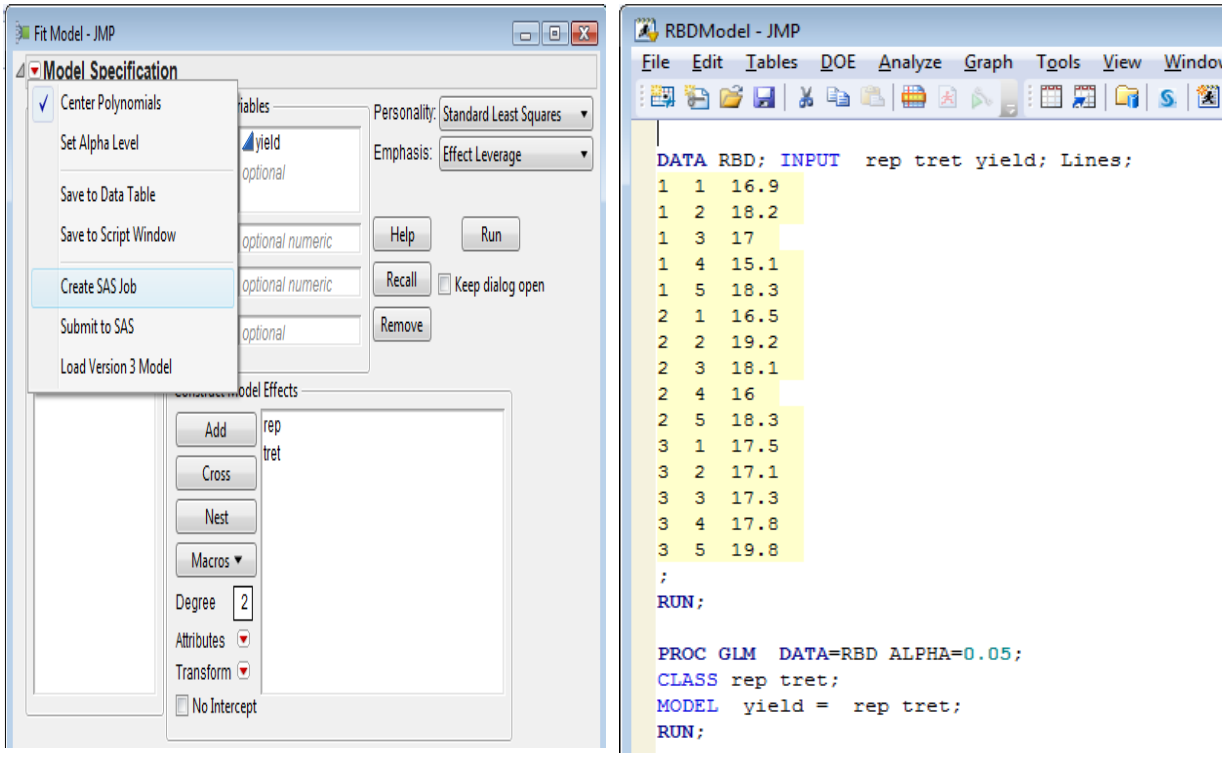



Figure 8.1.1: Showing the SAS Code Generated

**Example 8.1.2: Submitting SAS Code**

Follow all the steps of Examples 5.7, from the red triangle menu for Model Specification, select **Submit to SAS**. Select **Connect to SAS on this machine** if SAS is installed otherwise, select **Connect to SAS Server** window choose a method to connect to SAS .

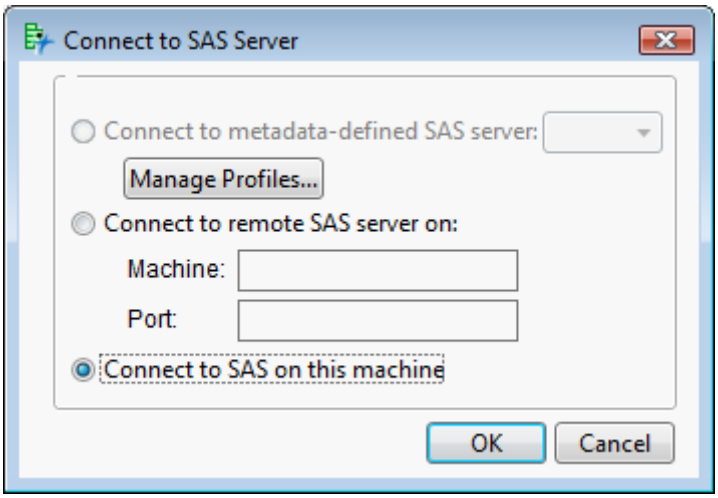



Figure 8.1.2: Connecting to SAS

Select OK. Result will be displayed SAS format. If we want that our result should appear in RTF, HTML or PDF format than before performing steps of analysis as given above, follow

the following steps: **File → Preference**. Following window will appear, Select **SAS integration** from Preference Group and then Select the **RTF** from Store process Result

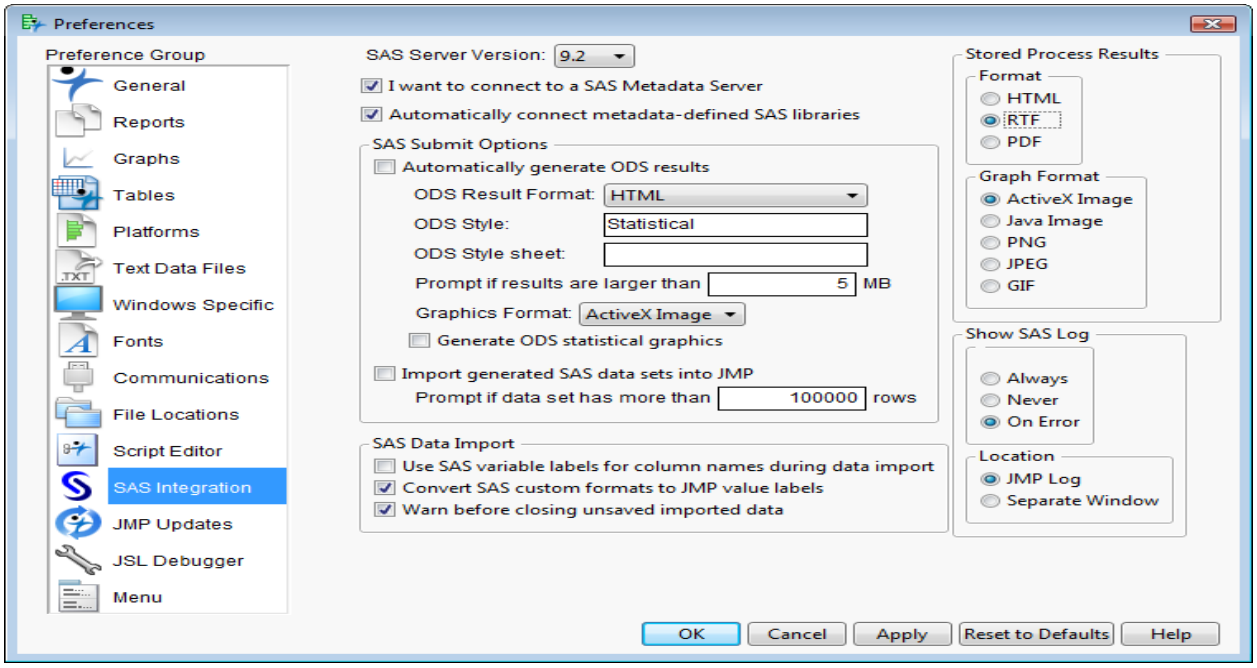



Figure 8.1.3: Screen showing selection of preferences for output

On submitting the SAS code generated we obtain the desired results. Partial result of analysis of data is as under:

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| **Model** | 6 | 13.20400000 | 2.20066667 | 2.52 | 0.1133 |
| **Error** | 8 | 6.99200000 | 0.87400000 | | |
| **Corrected Total** | 14 | 20.19600000 | | | |

| R-Square | Coeff Var | Root MSE | yield Mean |
|---|---|---|---|
| 0.653793 | 5.329987 | 0.934880 | 17.54000 |

| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| **rep** | 2 | 1.64800000 | 0.82400000 | 0.94 | 0.4289 |
| **tret** | 4 | 11.55600000 | 2.88900000 | 3.31 | 0.0705 |

## 9. Using JMP 10:

- **Activating**

Generally Icon of JMP software get prepared automatically just after installtion and appears on the desk top of the computer. To start working with this software one can double click the icon present on the desktop of the computer. In case icon is not present then Go to start button Click All programme→ JMP10 → Double click JMP10 icon . Starting JMP 10 just after installation ( First time) , a personalization window get activated automatically (shown in the snap shot below) and appears on the desktop. One can opt Check All or Check None from option Menu. Genarally we opt Check all so that all possible statistical analysis option will remain on the top menu while working with this software.

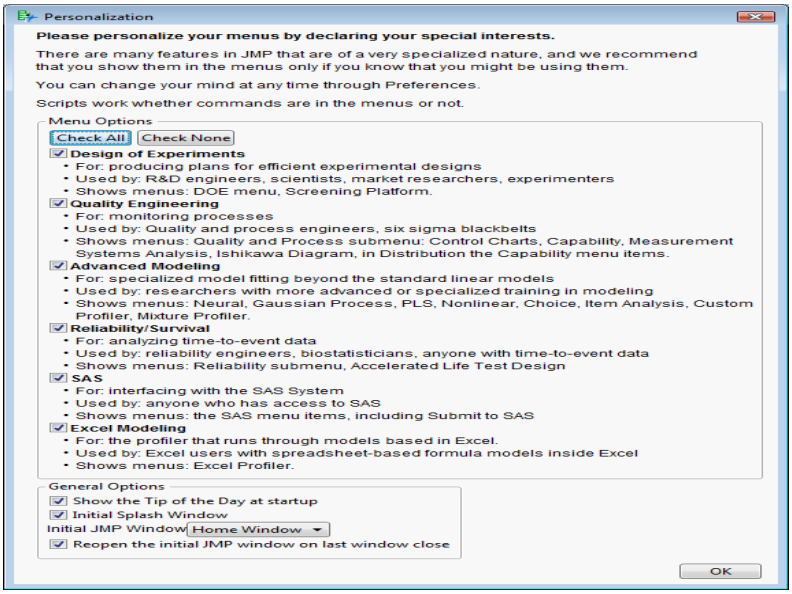



Figure 9.1: Starter Screen of JMP 10

Usually, on the double click of JMP 10 icon following window of appears with the tip of the Day window. One can close Tip of the Day window. The usual window of JMP 10 is different from the previous version of this software. Window is divided into four parts: Recent Files, Window List, Recent Help and Project. One can use the option displayed under the sub window Recent help or one can use it any time when in need.

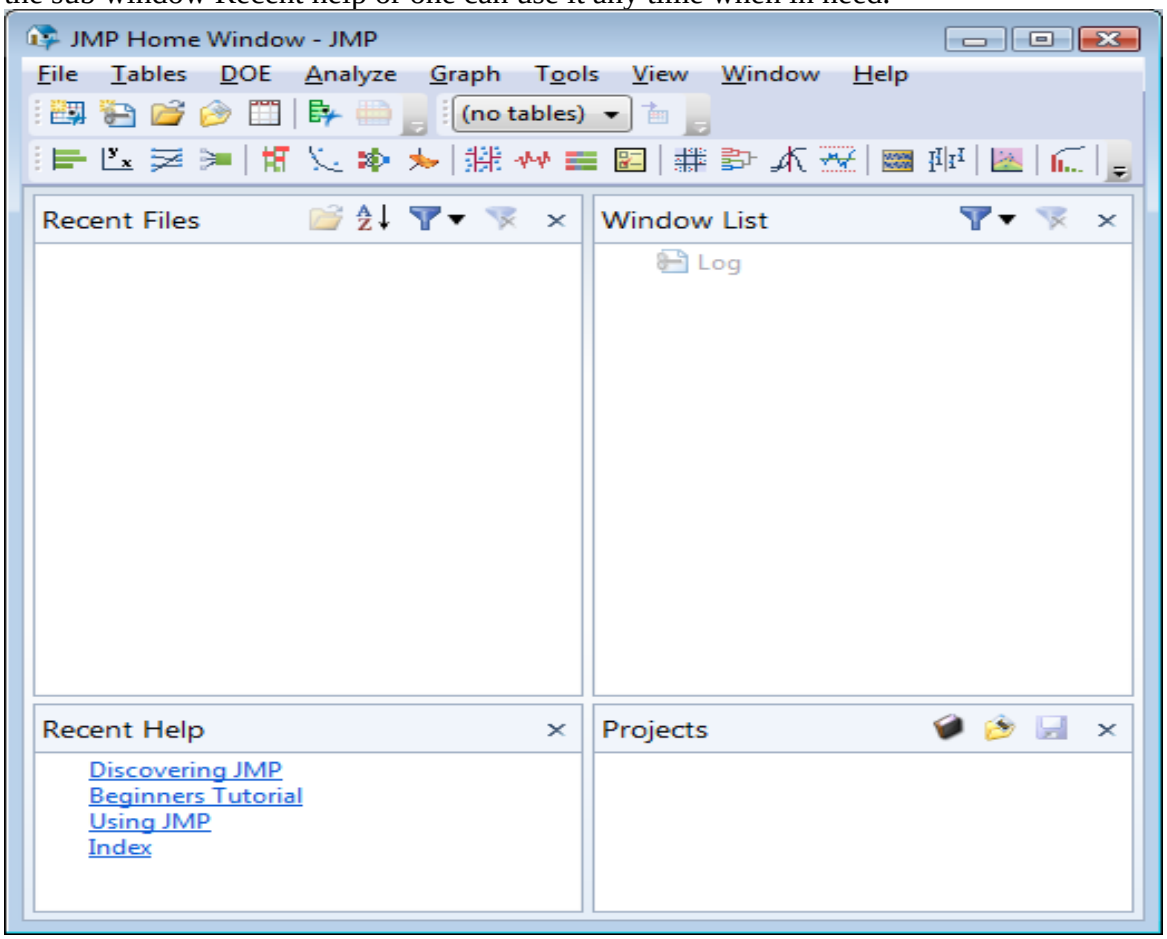



Figure 9.2: Home Window in JMP 10

If this window does not appear automatically, then one can go to View in the Menu Bar and select Home Window option.

**References**

Parsad, R., Gupta, V.K., Umrao, A.K. and Kole, B. Analysis of Data (http://iasri.res.in/design/Analysis%20of%20data/Contact%20us.html). *Design Resources Server*. Indian Agricultural Statistics Research Institute (ICAR), New Delhi 110 012, India. www.iasri.res.in/design (Accessed lastly on 18.11.2010).

Introduction to SAS. UCLA: Academic Technology Services, Statistical Consulting Group. from http://www.ats.ucla.edu/stat/sas/notes2/ (accessed September 09, 2010).

www.sas.com

www.support.sas.com

www.jmp.com