

STAR USER MANUAL

**Beta Version
2012**

Crop Research Informatics Laboratory
INTERNATIONAL RICE RESEARCH INSTITUTE

Table of Contents

1. Introduction	1
2. Overview of STAR.....	2
2.1. Starting STAR.....	2
Upon launching the <i>STAR</i> , the main window will appear.	2
2.2. STAR Menu Bar.....	2
2.3. Project Explorer	4
3. Data Manipulation	5
3.1. Insert Row(s)	5
3.2. Delete Row(s).....	5
3.3. Inserting Column(s)	5
3.4. Deleting Column(s).....	6
3.5. Create New Variables	6
3.6. Edit Variable Information.....	7
3.7. Sorting	8
3.8. Merge Datasets	10
3.9. Append Datasets	12
3.10. Creating Data Subset.....	13
3.11. Reshaping	14
3.11.1. Reshaping Data from Long (Serial) to Wide (Parallel) Format	14
3.11.2. Reshape Data from Wide to Long Format	16
3.12. Aggregate	18
4. Design Menu	20
4.1. Completely Randomized Design (CRD).....	20
4.2. Single-Factor Randomized Complete Block Design (RCBD).....	21
4.2.1. Two-Factor Randomized Complete Block Design (RCBD)	24
4.3. Split-Plot Design in RCBD	26
4.4. Strip-Plot Design.....	27
4.5. Alpha Lattice Design	29
4.6. Augmented Design in RCB	30
5. Descriptive Statistics	32
6. Cross Tabulation	37
7. T-test.....	41
7.1. One Sample <i>t</i> -tests	41
7.2. Paired Sample <i>t</i> -tests.....	42
7.3. Independent Sample <i>t</i> -tests	45
8. Analysis of Variance	50
8.1. One Factor Randomized Complete Block	50
8.2. Two Factor Randomized Complete Block	58
8.3. Split Plot Design in Randomized Complete Block	62
8.4. Strip Plot Design	67
8.5. Partitioning Sum of Squares	71
8.5.1. Group Comparison.....	71
8.5.2. Trend Comparison.....	73
9. Correlation Analysis.....	75
10. Linear Regression Analysis	80
11. Normality Test	87
12. Heterogeneity Test	91
13. Nonparametric Tests	93

13.1. Test for Several Independent Samples	93
14. Appendix	95
14.1. Data Description	95
14.1.1. Insecticides.csv	95
14.1.2. Maize.csv	95
14.1.3. SeedingRate.csv	96

1. Introduction

Statistical Tool for Agricultural Research (STAR) is a computer program for data management and basic statistical analysis of experimental data. It was created using the Eclipse Rich Client Platform (RCP) and uses the R language and environment for statistical computing and graphics. It has a user-friendly graphical interface where items are accessible via pull-down menus.

STAR has been developed primarily for the analysis of data from agricultural field trials, but many of the features can be used for analysis of data from other sources.

The main modules and facilities are

1. Data management with a spreadsheet
2. Descriptive statistics
3. Analysis of variance for basic experimental designs
4. Correlation and linear regression analysis
5. Graphics
6. Utilities for randomization and electronic field book
7. Non-parametric data analysis

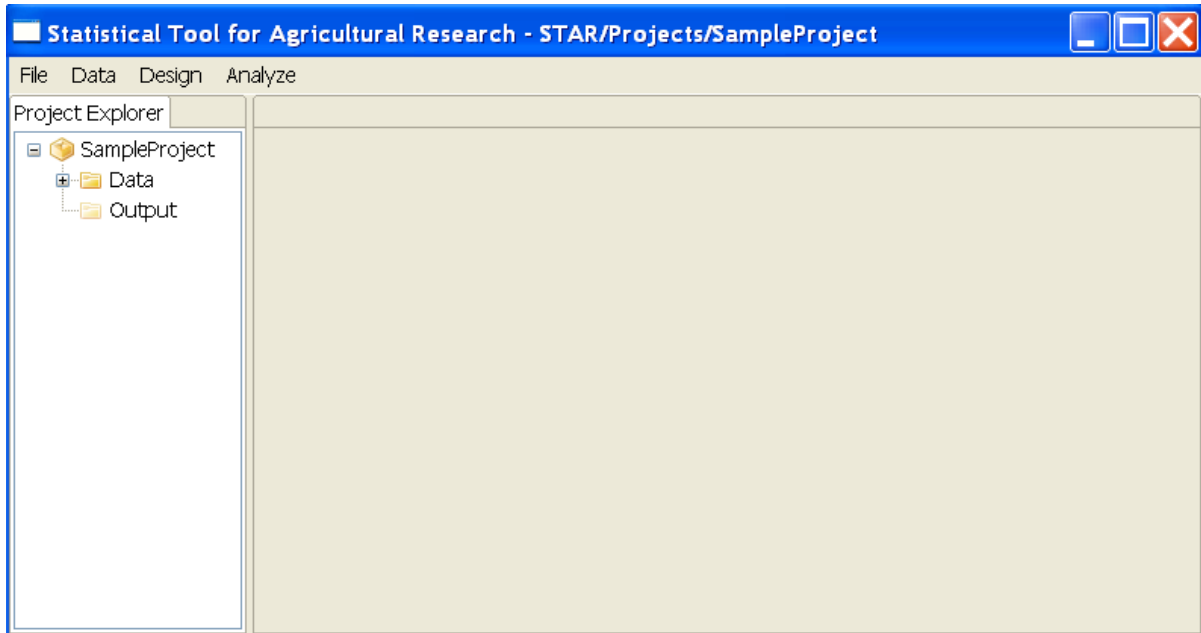
The STAR user's manual provides step-by-step instructions on how to perform certain tasks of interest to users. Screen images have been included as deemed helpful.

Menu items, names of dialogs and form controls are in **bold** letters. Filenames, variable names and directories are *italicized*. Menus appear in the form **File | Open**, which means "choose **Open** from the **File** menu". All of the examples used in this manual are included when the STAR package is installed.

2. Overview of STAR

2.1. Starting STAR

Upon launching the *STAR*, the main window will appear.



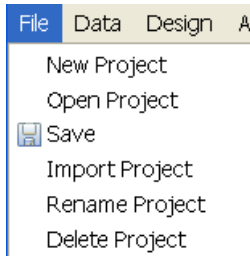
2.2. STAR Menu Bar

Most of the tasks that you perform in STAR start with menu selections. STAR has a menu bar (Figure 3.2 - 1) and five tab pages (Figure 3.2 - 2). The same main menu is displayed in all tab pages.

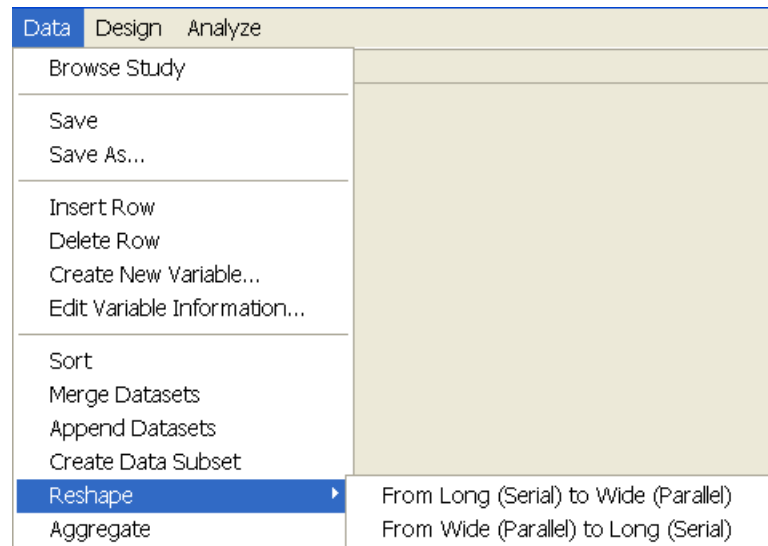
File Data Design Analyze

There are four pull down menus available in STAR and this are:

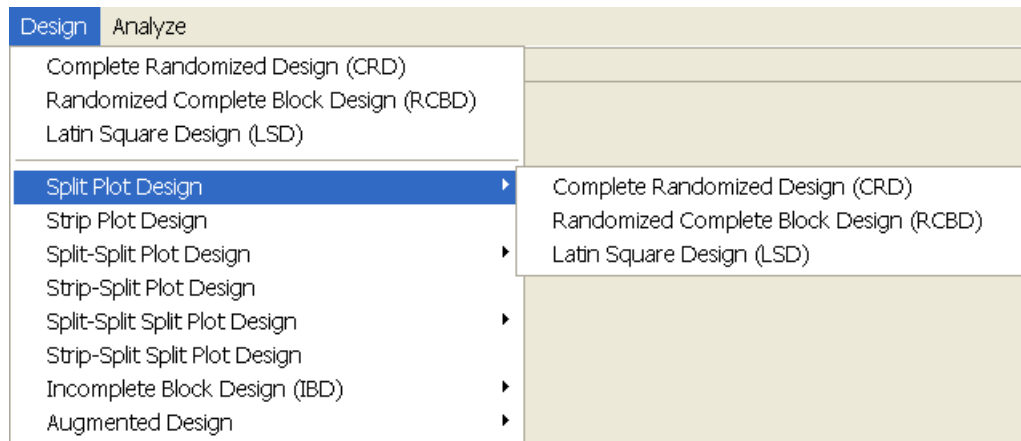
File Menu contains creating new project, open existing project, saving, importing project, renaming project and deleting project. It contains the following submenu items:



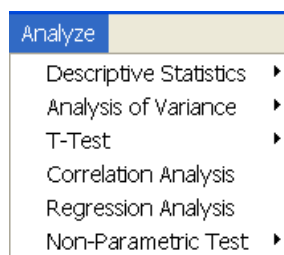
Data Menu contain items for reading and manipulating datasets. It contains the following submenu items:



Design Menu contain items for performing randomization. It contains the following submenu items:

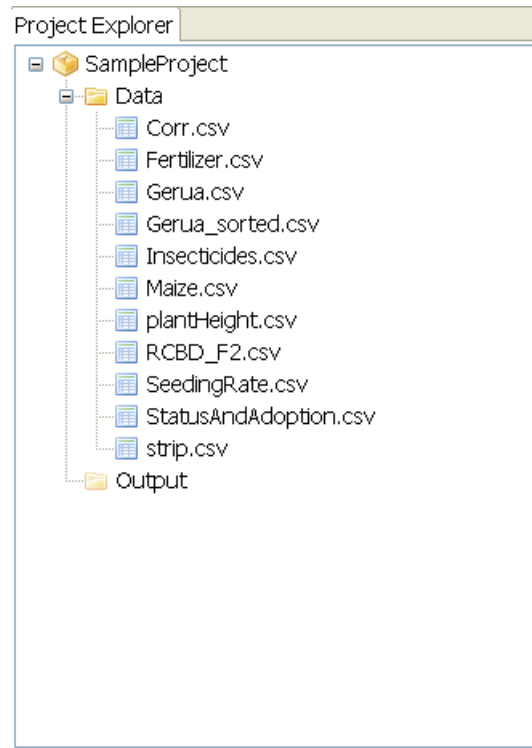


Analyze Menu contain items to perform statistical analysis. It contains the following submenu items:




2.3. Project Explorer

Project Explorer can be seen at the right-hand side of the main window. For first-time users, a default project named *SampleProject* is available with *Data* folder and *Output* folder inside it. The *Data* folder contains sample datasets that will be used in this tutorial. The *Output* folder, on the other hand, will contain all output files that will be created when an analysis is performed.




3. Data Manipulation


3.1. Insert Row(s)

- On the Project Explorer, locate the *data1.csv* file from the *Data* folder of the project named *SampleProject*. Double-click the file to view it in the Data Viewer.
- Select any row(s) or any cell in the row(s) where the user wants to insert new row(s) above it. To select several cells/rows, click any cell/row, then hold the Ctrl key and click on another cell/row. The number of selected rows is the number of row that will be inserted.
- To insert row, click **Data | Insert Row** from the main window of STAR or click on the Insert Row icon  in the Data Viewer tool bar.
- If there is no row selected before executing the Insert Row function, the user will be prompted if the user wants to insert a row after the last row of the data.

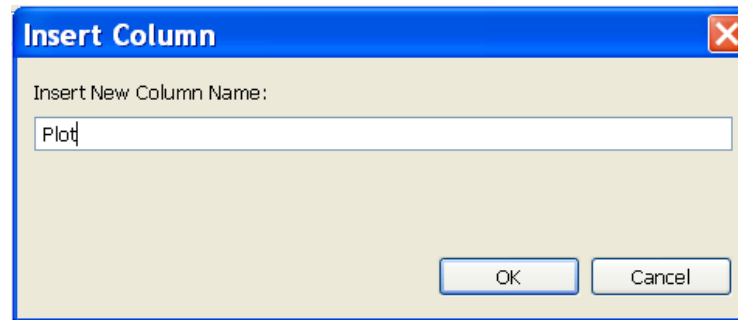
3.2. Delete Row(s)

- On the Project Explorer, locate the *data1.csv* file from the *Data* folder of the project named *SampleProject*. Double-click the file to view it in the Data Viewer.
- Select any row(s) or any cell in the row(s) that the user wants to delete. To select several cells/rows, click any cell/row, then hold the Ctrl key and click on another cell/row. The number of selected rows is the number of row that will be deleted.
- To delete row(s), click **Data | Delete Row** from the main window of STAR or click on the Delete Row icon  in the Data Viewer tool bar.

3.3. Inserting Column(s)


- On the Project Explorer, locate the *data1.csv* file from the *Data* folder of the project named *SampleProject*. Double-click the file to view it in the Data Viewer.
- Select a column where the user wants a new column to be inserted before it by clicking on the column name. A check icon will appear on the column header.
- To insert column(s), click **Data | Insert Column** from the main window of STAR or click on the Insert Column icon  in the Data Viewer tool bar.

The user will be ask to specify the name of the new column, say, *Plot*.




- Click OK.

3.4. Deleting Column(s)

- On the Project Explorer, locate the data1.csv file from the Data folder of the project named SampleProject. Double-click the file to view it in the Data Viewer.
- Select the column or one of the columns to be deleted by clicking on the column name. A check icon will appear on the column header.
- To delete column(s), click Data | Delete Column or click on the Insert Column icon  in the Data Viewer tool bar.
- Click OK.

3.5. Create New Variables

- On the Project Explorer, locate the *data1.csv* file from the *Data* folder of the project named *SampleProject*. Double-click the file to view it in the Data Viewer.
- To create new variables, click **Data | Create New Variable** from the main window of STAR or click on the Create New Variable icon  in the Data Viewer tool bar.
- Specify the required fields and appropriate options.

Name

This is the name of the new variable to be created.

Values

There are two options to determine the values of the new variable. The user can either transform any existing numeric variables or concatenate the values of any variables in the data set.

If transformation is selected, the user needs to specify the following:

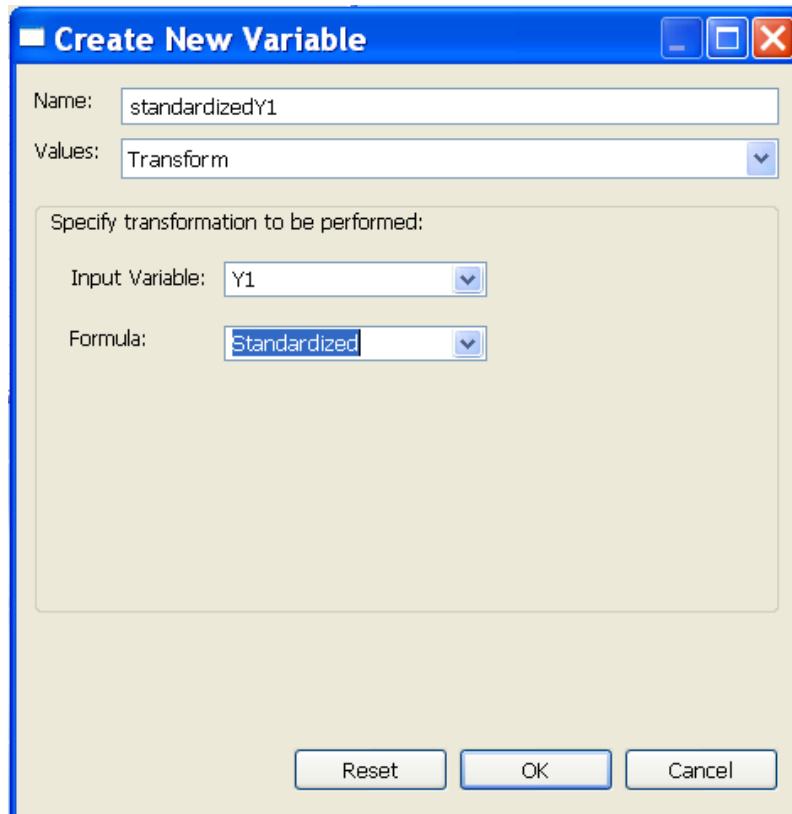
Input Variable

This is variable where the transformation will be performed.

Formula

Transformation options available are logarithm, natural logarithm, square root, power, exponential and standardized.


For the example, select the options as shown below:



If concatenation of columns is selected, the user needs to specify the columns to concatenate.

- Click OK.

3.6. Edit Variable Information

- On the Project Explorer, locate the *data1.csv* file from the *Data* folder of the project named *SampleProject*. Double-click the file to view it in the Data Viewer.
- To edit variable information, click **Data | Edit Variable Information** from the main window of STAR or click on the Edit Variable Information icon  in the Data Viewer tool bar.

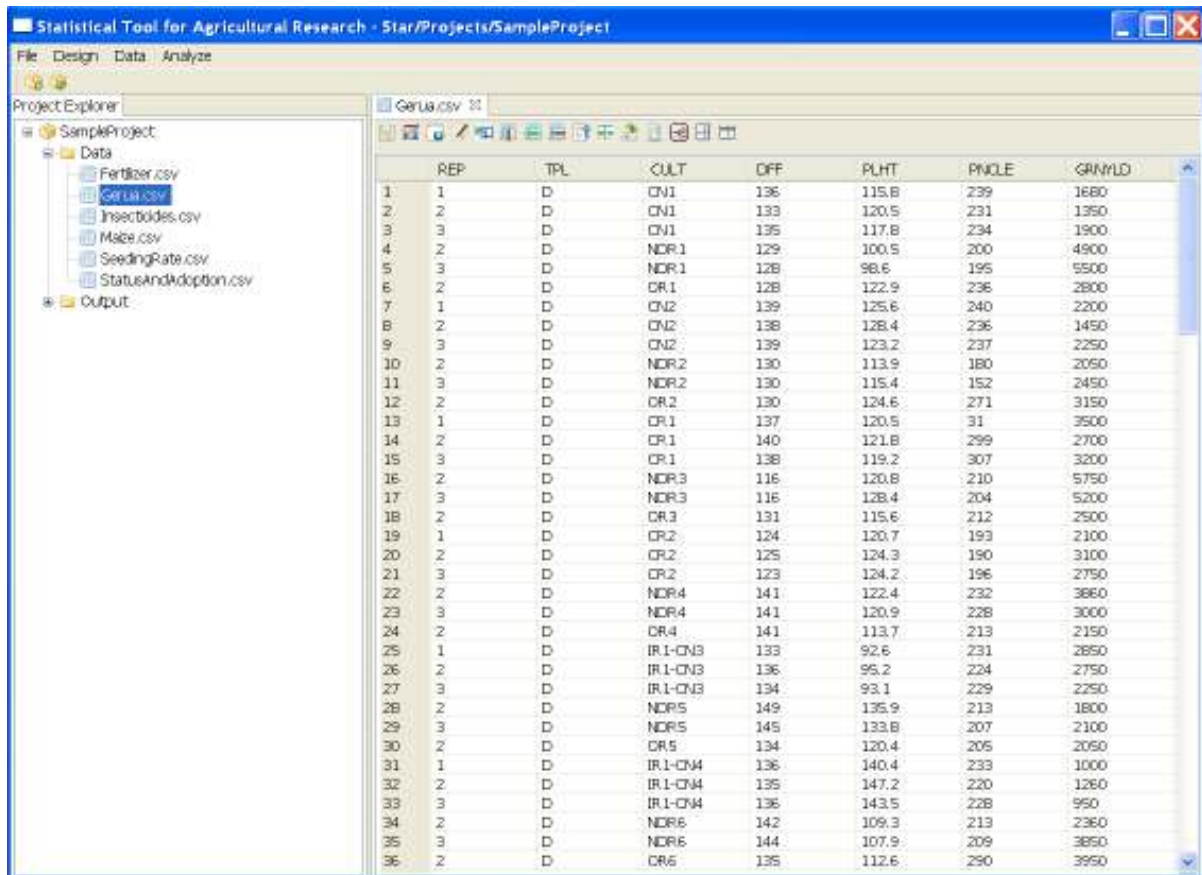
- The user can modify the variable type of the variables in the data set.
- Click OK.

3.7. Sorting


Re-arranging or sorting the rows or cases of the data file is often useful and sometimes necessary for certain types of analysis. The **Sort Cases Menu** can be used to sort rows based on the value of one or more sorting variables.

The steps to re-arrange the data set based on the value of one or more sorting variable are listed below:

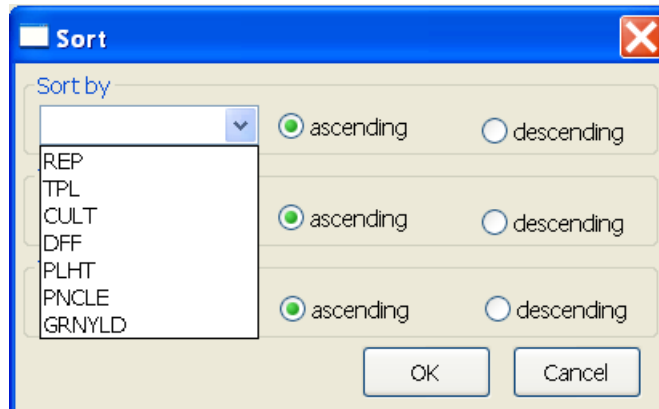
- On the Project Explorer, locate the dataset *Gerua.csv* from the *Data* folder of the project named *SampleProject*. Double-click the file to open it and view it in the Data Viewer. The data file is arranged by *TPL*, *CULT* then by *REP*.



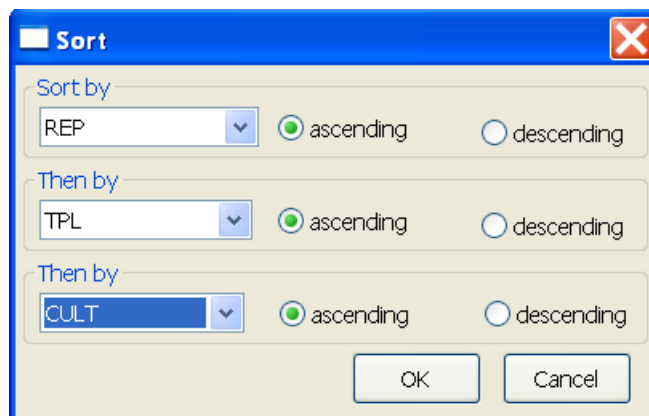
	REP	TPL	CULT	OFF	PLHT	PINOLE	GRN/LD
1	1	D	CN1	136	115.8	239	1680
2	2	D	CN1	133	120.5	231	1350
3	3	D	CN1	135	117.8	234	1900
4	2	D	NDR1	129	100.5	200	4900
5	3	D	NDR1	128	98.6	195	5500
6	2	D	CR1	128	122.9	236	2800
7	1	D	CN2	139	125.6	240	2200
8	2	D	CN2	138	128.4	236	1450
9	3	D	CN2	139	123.2	237	2250
10	2	D	NDR2	130	113.9	180	2050
11	3	D	NDR2	130	115.4	152	2450
12	2	D	CR2	130	124.6	271	3150
13	1	D	CR1	137	120.5	31	3500
14	2	D	CR1	140	121.8	299	2700
15	3	D	CR1	138	119.2	307	3200
16	2	D	NDR3	116	120.8	210	5750
17	3	D	NDR3	116	128.4	204	5200
18	2	D	CR3	131	115.6	212	2500
19	1	D	CR2	124	120.7	193	2100
20	2	D	CR2	125	124.3	190	3100
21	3	D	CR2	123	124.2	196	2750
22	2	D	NDR4	141	122.4	232	3860
23	3	D	NDR4	141	120.9	228	3000
24	2	D	CR4	141	113.7	213	2150
25	1	D	IR1-CN3	133	92.6	231	2850
26	2	D	IR1-CN3	136	95.2	224	2750
27	3	D	IR1-CN3	134	93.1	229	2250
28	2	D	NDR5	149	135.9	213	1800
29	3	D	NDR5	145	133.8	207	2100
30	2	D	CR5	134	120.4	205	2050
31	1	D	IR1-CN4	136	140.4	233	1000
32	2	D	IR1-CN4	135	147.2	220	1260
33	3	D	IR1-CN4	136	143.5	228	950
34	2	D	NDR6	142	109.3	213	2360
35	3	D	NDR6	144	107.9	209	3850
36	2	D	CR6	135	112.6	290	3950

- Click **Data | Sort** from the main window of STAR or click on the Sort icon  in the Data Viewer tool bar. The **Sort** dialog box will appear.
- The user can specify up to three variables as the basis for sorting and the order of sorting.

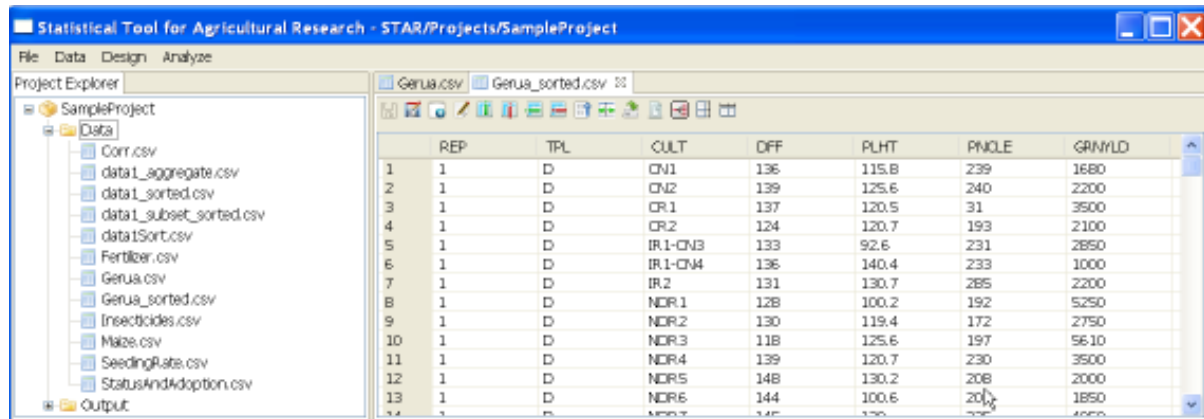
Click the drop-down list box to identify the sorting variables. The rows can be re-arranged or sorted in ascending or descending order, with ascending order as the default option. If multiple variables are selected, rows are sorted for each variable within categories of the preceding variables. For character variables, uppercase letters precede their lowercase counterparts. Further, the rows can be re-arranged or sorted in ascending or descending order, with ascending order as the default option.



Suppose we want to re-arranging the content of the data *Gerua.csv*, by *REP*, *TPL* and *CULT*. The completed dialog box should appear as illustrated below:



- Click the **Ok** button to sort the active data. The **Sort** dialog box is closed and STAR activates the new dataset in the Data Viewer tab.

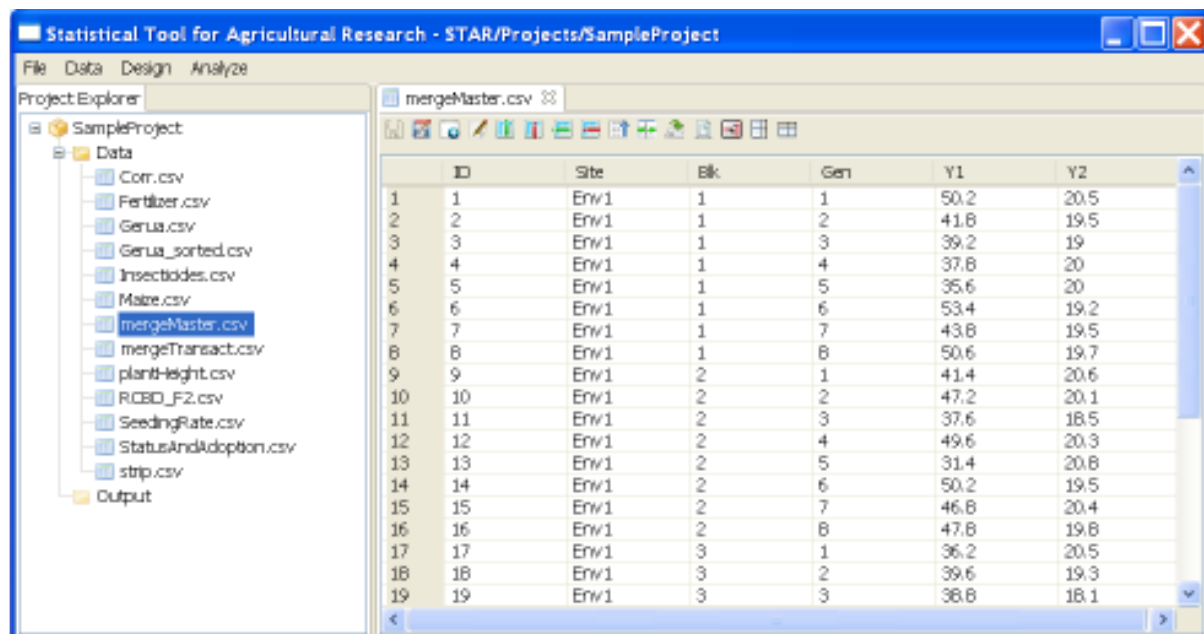


	REP	TPL	CULT	OFF	PLHT	PNQLE	GRN/WD
1	1	D	CN1	136	115.8	239	1680
2	1	D	CN2	139	125.6	240	2200
3	1	D	CR1	137	120.5	31	3500
4	1	D	CR2	124	120.7	193	2100
5	1	D	IR1-CN3	133	92.6	231	2850
6	1	D	IR1-CN4	136	140.4	233	1000
7	1	D	IR2	131	130.7	285	2200
8	1	D	NDR1	128	100.2	192	5250
9	1	D	NDR2	130	119.4	172	2750
10	1	D	NDR3	118	125.6	197	5610
11	1	D	NDR4	139	120.7	230	3500
12	1	D	NDR5	148	130.2	208	2000
13	1	D	NDR6	144	100.6	20	1850
14	1	D	NDR7	145	130	208	4000


3.8. Merge Datasets

The **Merge** submenu can be used to combine the active data file (known as the master data) with another data file (referred as the transaction data) that contains the same cases or rows but different variables.

- On the Project Explorer, locate the *mergeMaster.csv* file from the *Data* folder of the project named *SampleProject*. Double-click the file to view it in the Data Viewer.



	ID	Site	Blk	Gen	Y1	Y2
1	1	Enw1	1	1	50.2	20.5
2	2	Enw1	1	2	41.8	19.5
3	3	Enw1	1	3	39.2	19
4	4	Enw1	1	4	37.8	20
5	5	Enw1	1	5	35.6	20
6	6	Enw1	1	6	53.4	19.2
7	7	Enw1	1	7	43.8	19.5
8	8	Enw1	1	8	50.6	19.7
9	9	Enw1	2	1	41.4	20.6
10	10	Enw1	2	2	47.2	20.1
11	11	Enw1	2	3	37.6	18.5
12	12	Enw1	2	4	49.6	20.3
13	13	Enw1	2	5	31.4	20.8
14	14	Enw1	2	6	50.2	19.5
15	15	Enw1	2	7	46.8	20.4
16	16	Enw1	2	8	47.8	19.8
17	17	Enw1	3	1	36.2	20.5
18	18	Enw1	3	2	39.6	19.3
19	19	Enw1	3	3	38.8	18.1

- Choose **Data | Merge Datasets** or click on the Merge Datasets icon  in the Data Viewer tool bar.
- Specify the required fields and appropriate options.

Transaction File Name

Specify this file by selecting from the project *Data* folder using the drop-down combo box or by locating it using the **Browse** button. For the example, select *data2_merge.csv* file from the drop-down combo box.

Text File Delimiter

If the transaction file selected is a text file, the delimiter should be specified.

Observations to Include

The options available are to include common observations, all observations in the active data, all observations in the transaction data and all observations.

Key Variables

These are pair(s) of variables that are present in the active and transaction files and will be the basis for the merging. Variables can be paired by clicking one variable in the active data and clicking one variable in the transaction data then clicking the **Pair** button.

Other

These are the other variables the user wants to include in the merge data set.

For the *example*, the completed dialog box should appear as illustrated below:

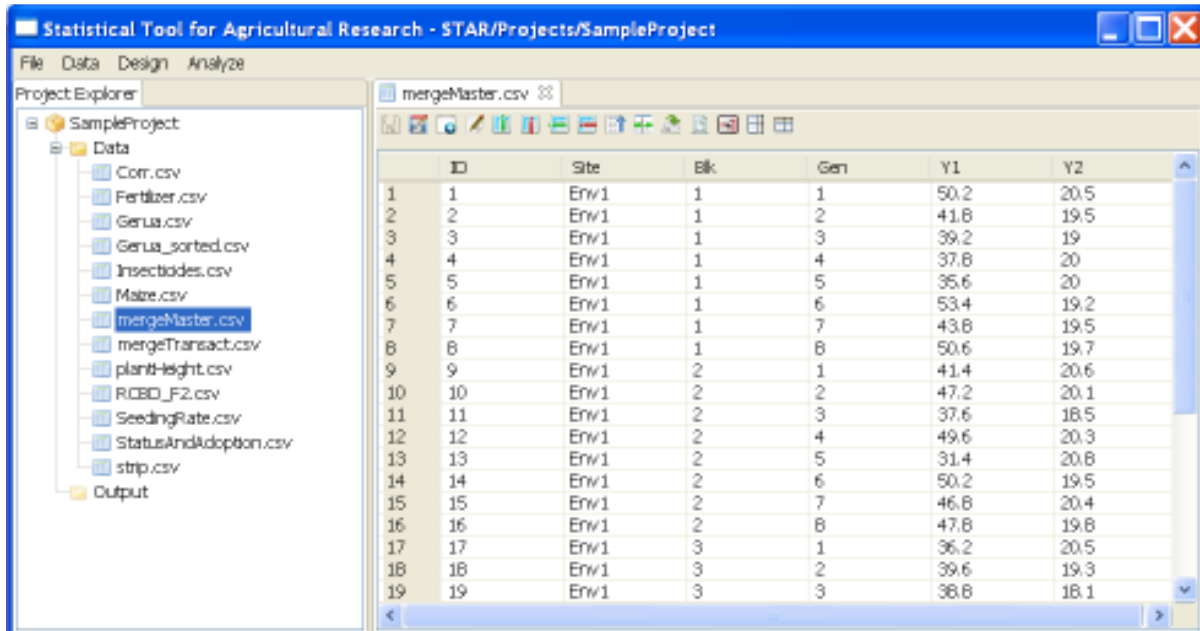
The screenshot shows the 'Merge Data' dialog box. The 'Transaction File Name' field contains 'mergeTransact.csv' and a 'Browse' button is to its right. The 'Text File Delimiter' is set to 'comma'. The 'Observations to Include' dropdown is set to 'common observations'. Under 'Variables in the...', the 'Active Data' list contains 'Y2' and the 'Transaction Data' list is empty. The 'New Working Data File:' section has a 'Key Variables' list containing 'ID & PlotNum', 'Site & Env', 'Blk & Rep', and 'Gen & Gen'. Below this list is an 'Add' button. The 'Other' section contains a list with 'Y1' and 'Y3'. A 'Pair' button is located between the 'Active Data' and 'Transaction Data' lists. At the bottom of the dialog are 'Reset', 'Ok', and 'Cancel' buttons.

- Click OK. The merged data is saved in the *Data* folder and displayed in the Data Viewer.

3.9. Append Datasets


The **Append** submenu can be used combine two data files with the same variables but different cases.

- On the Project Explorer, locate the *mergeMaster.csv* file from the *Data* folder of the project named *SampleProject*. Double-click the file to view it in the Data Viewer.



The screenshot shows the STAR software interface. On the left, the Project Explorer displays a tree structure for 'SampleProject' with a 'Data' folder containing various CSV files. 'mergeMaster.csv' is selected. On the right, the Data Viewer displays the contents of 'mergeMaster.csv' as a table with 7 columns: ID, Site, Blk, Gen, Y1, and Y2. The table contains 19 rows of data.

ID	Site	Blk	Gen	Y1	Y2
1	1	Erw1	1	50.2	20.5
2	2	Erw1	1	41.8	19.5
3	3	Erw1	1	39.2	19
4	4	Erw1	1	37.8	20
5	5	Erw1	1	35.6	20
6	6	Erw1	1	53.4	19.2
7	7	Erw1	1	43.8	19.5
8	8	Erw1	1	50.6	19.7
9	9	Erw1	2	41.4	20.6
10	10	Erw1	2	47.2	20.1
11	11	Erw1	2	37.6	18.5
12	12	Erw1	2	49.6	20.3
13	13	Erw1	2	31.4	20.8
14	14	Erw1	2	50.2	19.5
15	15	Erw1	2	46.8	20.4
16	16	Erw1	2	47.8	19.8
17	17	Erw1	3	36.2	20.5
18	18	Erw1	3	39.6	19.3
19	19	Erw1	3	38.8	18.1

- Choose **Data | Append Datasets** or click on the Append Datasets icon  in the Data Viewer tool bar.
- Specify the required fields and appropriate options.

Transaction File Name

Specify this file by selecting from the project *Data* folder using the drop-down combo box or by locating it using the **Browse** button. For the example, select *data2_merge.csv* file from the drop-down combo box.

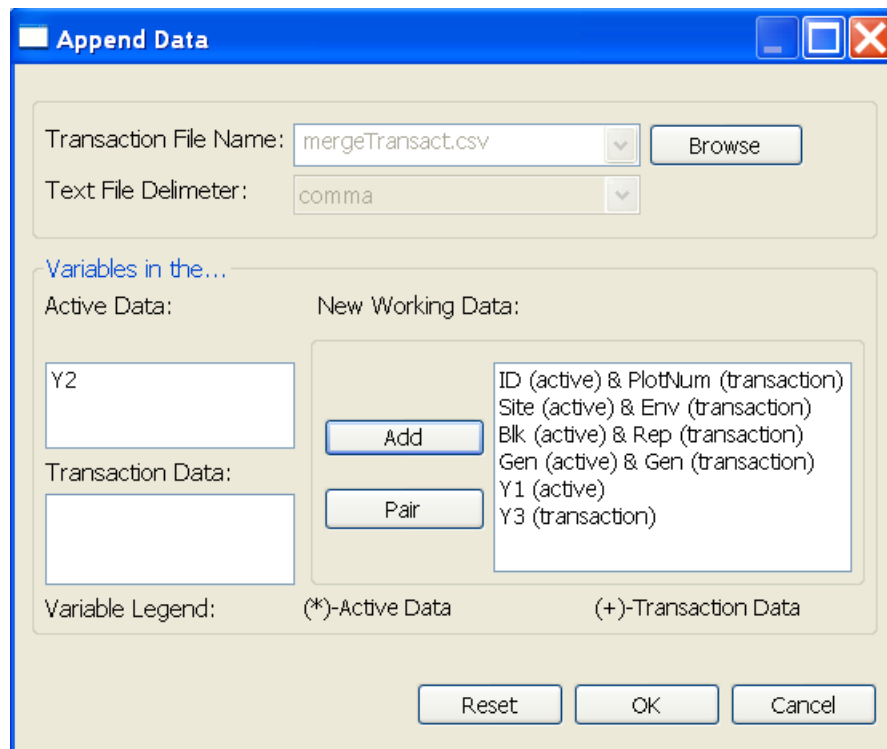
Text File Delimiter

If the transaction file selected is a text file, the delimiter should be specified.

New Working Data

This is the list of variables or paired variables that will be included in the new data set. Variables can be paired by clicking one variable in the active data and clicking one variable in the transaction data then clicking the **Pair** button. Other variables can be added to this list by clicking the variable then clicking the **Add** button.

For the *example*, the completed dialog box should appear as illustrated below:



- Click OK. The new data is saved in the *Data* folder and displayed in the Data Viewer.

3.10. Creating Data Subset

- On the Project Explorer, locate the *mergeMaster.csv* file from the Data folder of the project named *SampleProject*. Double-click the file to view it in the Data Viewer.


The screenshot shows the STAR software interface with the 'mergeMaster.csv' file selected in the Project Explorer and its data displayed in the Data Viewer.

Project Explorer:

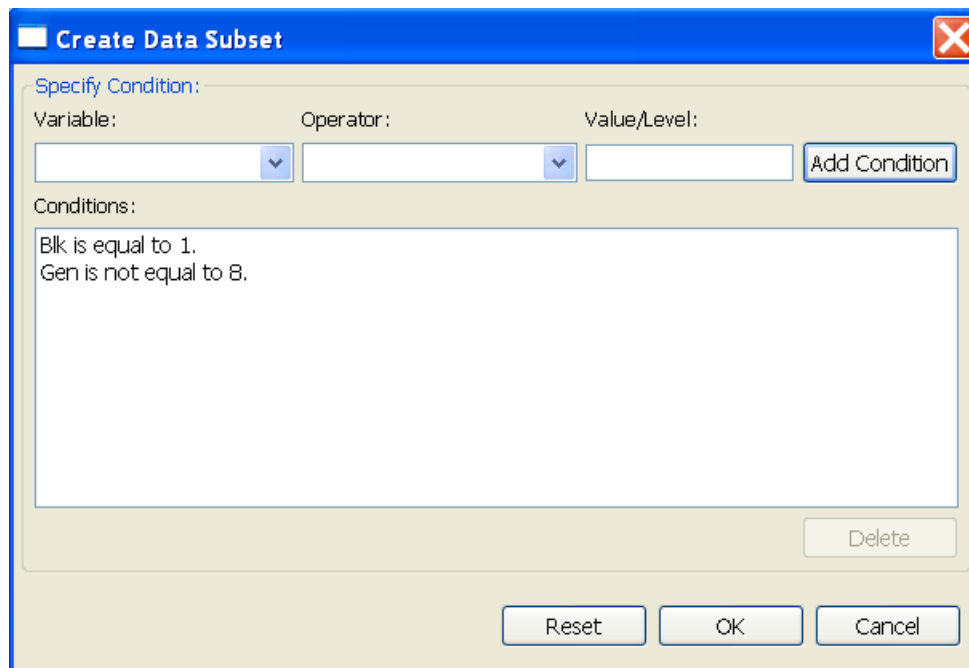
- SampleProject
 - Data
 - Corr.csv
 - Fertilizer.csv
 - Genus.csv
 - Genus_sorted.csv
 - Insecticides.csv
 - Maize.csv
 - mergeMaster.csv (selected)
 - mergeTransact.csv
 - plantHeight.csv
 - RCBD_F2.csv
 - SeedingRate.csv
 - StatusAndAdoption.csv
 - strip.csv
 - Output

Data Viewer (mergeMaster.csv):

	ID	Site	Blk	Gen	Y1	Y2
1	1	Env1	1	1	50.2	20.5
2	2	Env1	1	2	41.8	19.5
3	3	Env1	1	3	39.2	19
4	4	Env1	1	4	37.8	20
5	5	Env1	1	5	35.6	20
6	6	Env1	1	6	53.4	19.2
7	7	Env1	1	7	43.8	19.5
8	8	Env1	1	8	50.6	19.7
9	9	Env1	2	1	41.4	20.6
10	10	Env1	2	2	47.2	20.1
11	11	Env1	2	3	37.6	18.5
12	12	Env1	2	4	49.6	20.3
13	13	Env1	2	5	31.4	20.8
14	14	Env1	2	6	50.2	19.5
15	15	Env1	2	7	46.8	20.4
16	16	Env1	2	8	47.8	19.8
17	17	Env1	3	1	36.2	20.5
18	18	Env1	3	2	39.6	19.3
19	19	Env1	3	3	38.8	18.1

- Choose **Data | Create Data Subset** or click on the Create Data Subset icon  in the Data Viewer tool bar.
- Suppose the user wants to select only the observations from Env2. This can be done by selecting variable Site, operator equals (==), typing Env2 under Value/Level and clicking the Add Condition button. Do the same for the condition “Gen not equal to 8”.

For the example, the completed dialog box should appear as illustrated below:



- Click OK. The new data is saved in the *Data* folder and displayed in the Data Viewer.

3.11. Reshaping

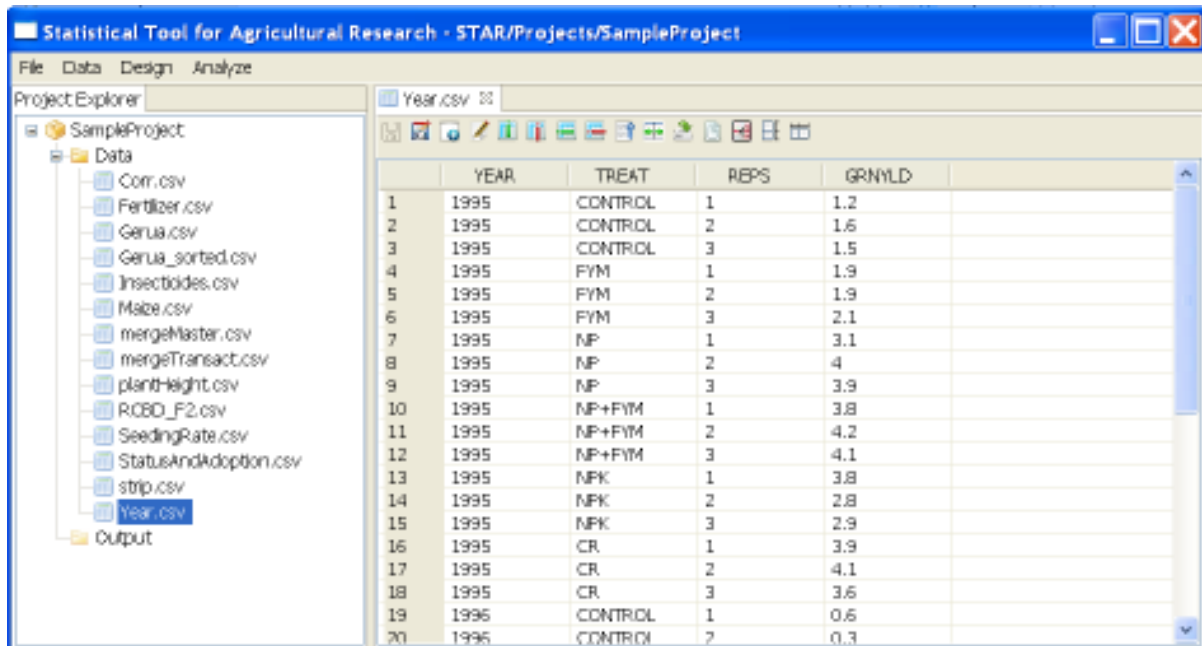
Data files are not always organized in the ideal form for specific needs of the users. The **Reshape subenu** can be used to restructure or reshape your active data set to follow the structure of the data set appropriate for the STAR procedure you want to use. This menu can also be used to combine data files, change the unit of analysis by grouping cases together or select a subset of observations to be used for the analysis.

3.11.1. Reshaping Data from Long (Serial) to Wide (Parallel) Format


Reshaping data from long (serial) to wide (parallel) involves re-arranging a data file, such that, repeated measurements are in separate columns.

The steps for reshaping the data from long to wide format are listed below:

- On the Project Explorer, locate the *Year.csv* file from the Data folder of the project named *SampleProject*. Double-click the file to view it in the Data Viewer.



	YEAR	TREAT	REPS	GRNYLD	
1	1995	CONTROL	1	1.2	
2	1995	CONTROL	2	1.6	
3	1995	CONTROL	3	1.5	
4	1995	FYM	1	1.9	
5	1995	FYM	2	1.9	
6	1995	FYM	3	2.1	
7	1995	NP	1	3.1	
8	1995	NP	2	4	
9	1995	NP	3	3.9	
10	1995	NP+FYM	1	3.8	
11	1995	NP+FYM	2	4.2	
12	1995	NP+FYM	3	4.1	
13	1995	NPK	1	3.8	
14	1995	NPK	2	2.8	
15	1995	NPK	3	2.9	
16	1995	CR	1	3.9	
17	1995	CR	2	4.1	
18	1995	CR	3	3.6	
19	1996	CONTROL	1	0.6	
20	1996	CONTROL	2	0.3	

- Click **Data | Reshape | Long (Serial) to Wide (Parallel) Format** from the main window of STAR or click on the Rshape to Wide icon  in the Data Viewer tool bar. The **Manipulate Data: from Long (Serial) to Wide (Parallel)** dialog box will automatically appear as shown below.
- Specify the required fields and appropriate options.

Variable(s) to be Reshape

This pertains to the list of variable(s) which is to be divided into two or more columns and saved to different variables in terms of the levels of the index factor.

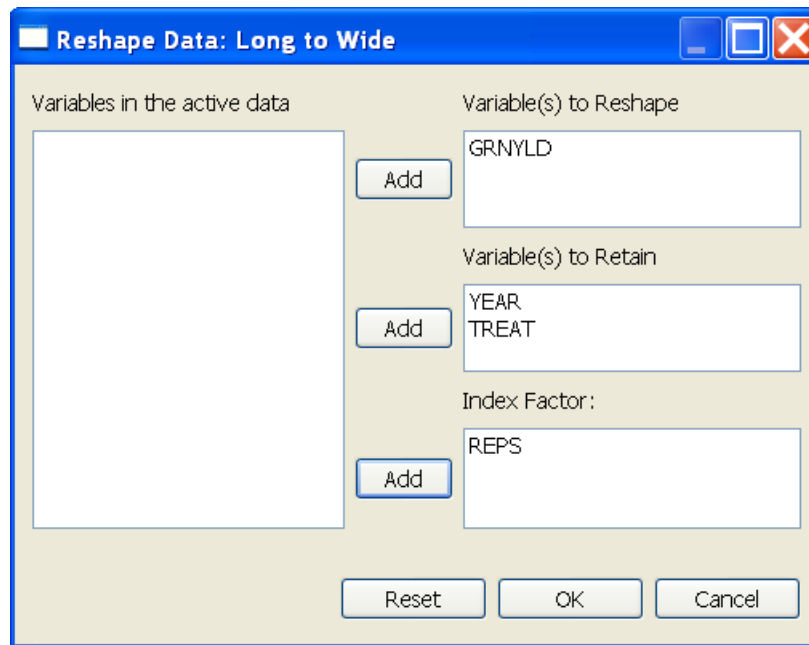
Variables(s) to be Retained

This pertains to variable(s) to be retained to describe the individual cases or observations.

Index Factor

This pertains to variable(s) which will determine the groupings of the values of the variable(s) to be transposed

For the example, the completed **Reshape Data: Long to Wide** dialog box should now look like this:



- Click the **Ok** button to save the restructured data set. The new data is saved in the Data folder and displayed in the Data Viewer tab.


3.11.2. Reshape Data from Wide to Long Format

Manipulating data from wide (parallel) to long (serial) refers to re-arranging a multivariate into a univariate data. For instance, different columns representing measurements taken for a response variable over a period of time can be combined to form a single column, with the time variable used as an additional identifier variable.

The steps for reshaping the data from wide to long format are listed below:

- On the Project Explorer, locate the *Year_ReshapeToWide.csv* file (the generated file from section 3.11.1) from the Data folder of the project named *SampleProject*. Double-click the file to view it in the Data Viewer.

	YEAR	TREAT	GRNYLD.1	GRNYLD.2	GRNYLD.3	
1	1995	CONTROL	1.2	1.6	1.5	
2	1995	FYM	1.9	1.9	2.1	
3	1995	NP	3.1	4	3.9	
4	1995	NP+FYM	3.8	4.2	4.1	
5	1995	NPK	3.8	2.8	2.9	
6	1995	CR	3.9	4.1	3.6	
7	1996	CONTROL	0.6	0.3	0.3	
8	1996	FYM	0.6	1.1	0.3	
9	1996	NP	1	1.4	1.4	
10	1996	NP+FYM	1.5	2	1.6	
11	1996	NPK	1.8	1.3	1.2	
12	1996	CR	1.3	2.9	1.6	

- Click **Data | Reshape | From Wide (Parallel) to Long (Serial) Format** from the main window of STAR or click on the Reshape to Long icon  in the Data Viewer tool bar. The **Reshape Data: Wide to Long** dialog box will appear.
- Specify the required fields and appropriate options,

Variable(s) to be Reshape

This pertains to variables which are to be combined into a single column.

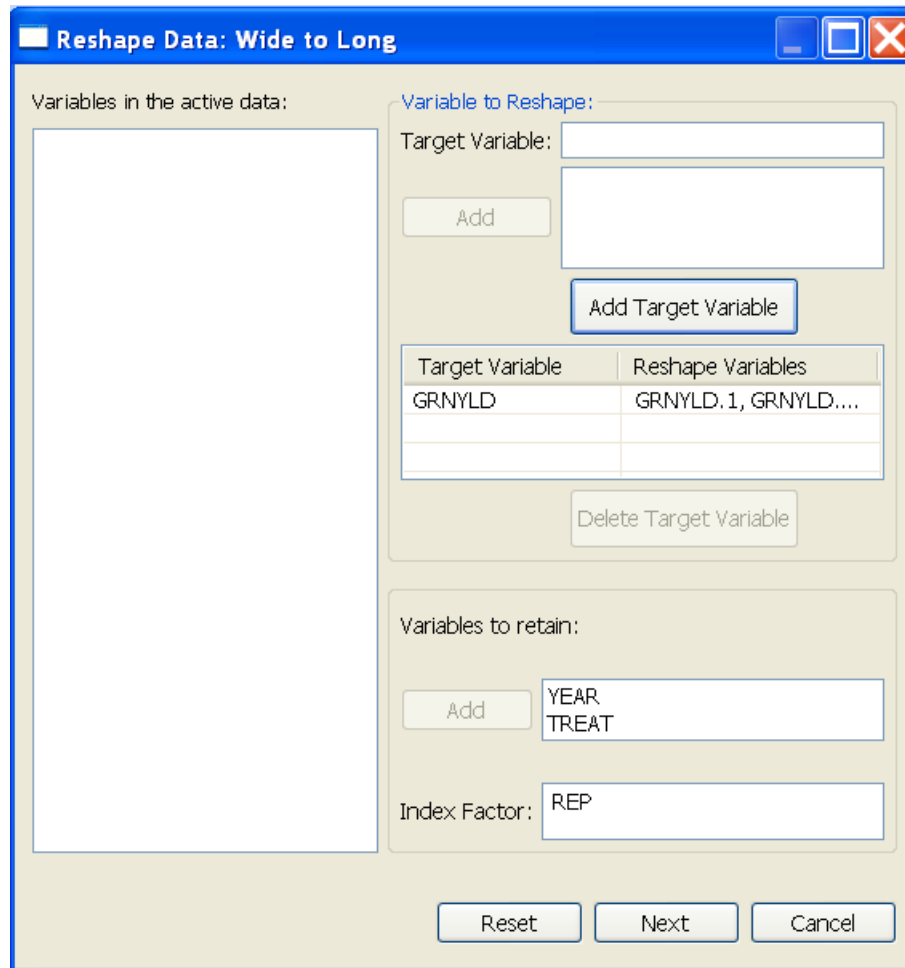
Variable(s) to be Retained

This pertains to variable(s) that describe the individual cases.

Index Factor

This is the new variable which will be create that pertains to the column where the particular observation came from.

For the *example*, the completed dialog box should appear as illustrated below:



The dialog box is titled "Reshape Data: Wide to Long". It contains the following elements:

- Variables in the active data:** A large empty rectangular box on the left.
- Variable to Reshape:** A section on the right containing:
 - Target Variable:** A text input field.
 - Add:** A button below the Target Variable field.
 - Add Target Variable:** A button below the Add button.
 - Table:** A table with two columns: "Target Variable" and "Reshape Variables".

Target Variable	Reshape Variables
GRNYLD	GRNYLD.1, GRNYLD....
 - Delete Target Variable:** A button below the table.
- Variables to retain:** A section containing:
 - Add:** A button.
 - Input field:** A text box containing "YEAR" and "TREAT" on separate lines.
- Index Factor:** A text input field containing "REP".
- Buttons:** "Reset", "Next", and "Cancel" buttons at the bottom.

- Click **Ok** to save the restructured data set. It can now be viewed in the **Data View** tab.

3.12. Aggregate

The **Aggregate submenu** aggregates group of cases into single cases and creates a new, aggregated file or creates new variables in the working data file that contain the aggregated data. Rows are aggregated based on the value of one or more grouping variables.

The steps for reshaping the data from long to wide format are listed below:

- On the Project Explorer, locate the *Year.csv* file from the Data folder of the project named *SampleProject*. Double-click the file to view it in the Data Viewer.
- Choose **Data | Aggregate** or click on the Aggregate Data icon in the Data Viewer tool bar.

Specify the required fields and appropriate options.

Variable(s) to Aggregate

This is the list of variables whose values will be summarized using the selected functions.

Grouping Factor(s)

Specify this grouping factor, if the user wants to summarize the values per level of a grouping factor.

Functions

Available functions are minimum, maximum, mean, median, sum, variance and standard deviation.

For the *example*, the completed dialog box should appear as illustrated below:

Aggregate Data

Available Variable(s):
YEAR
REPS
Add

Variable(s) to aggregate:
GRNYLD

Available Factor(s):
Add

Grouping Factor(s):
TREAT

Available Function(s):
maximum
sum
variance
Add

Functions:
mean
standard deviation
median

Save
☒ Add aggregated variables to the active data
☐ Create new data file containing the aggregated variables o

Reset OK Cancel

- Click OK. The active data is saved with the additional columns and displayed in the Data Viewer.

4. Design Menu

The **Design menu** allows user to generate randomized field plans for single- and multi-factor designs. This menu will generate a field book saved in csv format.

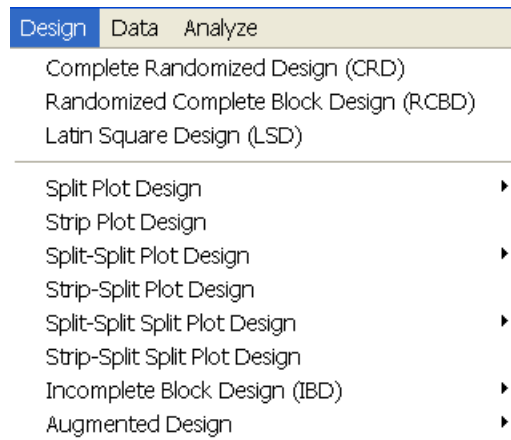
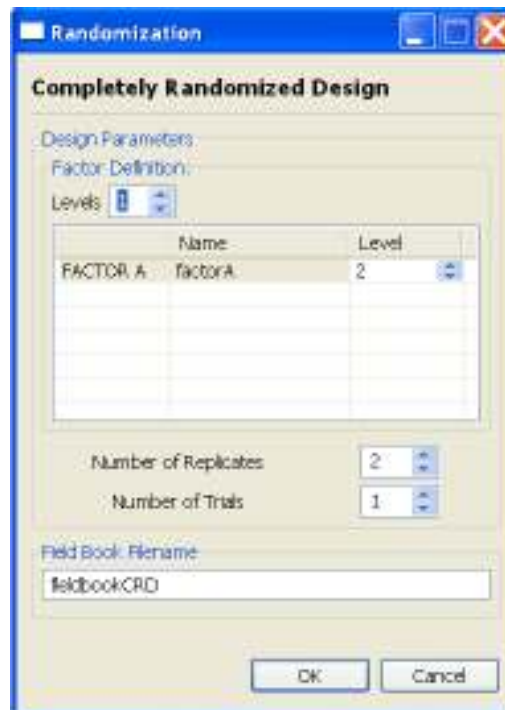


Figure 6 – 1. The Design Menu.

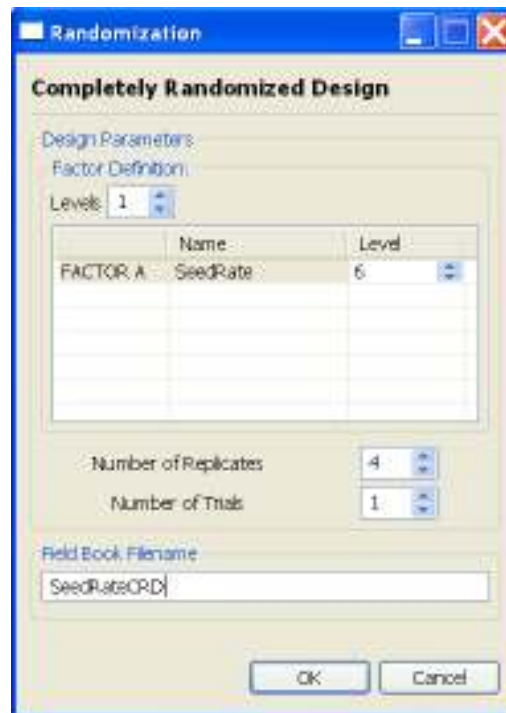
4.1. Completely Randomized Design (CRD)

To perform randomization, click **Design | Completely Randomized Design (CRD)**. The **Randomization** dialog box will appear.



The **Field Book Filename** text box lets the user specify the name of the data file that will be created. The default filename is *fieldbookCRD*.

Specify the required field. For the *example*, the completed dialog box should appear as illustrated below:



The screenshot shows a Windows-style dialog box titled "Randomization". Inside, the "Completely Randomized Design" section is active. Under "Design Parameters", the "Factor Definition" is set to "Levels: 1". A table below lists factors with columns for "Name" and "Level". The first row shows "FACTOR A: SeedRate" with a level of "6". Below the table, "Number of Replicates" is set to "4" and "Number of Trials" is set to "1". The "Field Book Filename" field contains the text "SeedRateCRD". At the bottom are "OK" and "Cancel" buttons.

Name	Level
FACTOR A: SeedRate	6

Sample Problem:

An experiment to compare the effects of six different rates of seeding (25, 50, 75, 100, 125,

4.2. Single-Factor Randomized Complete Block Design (RCBD)

Suppose we want to generate a randomization for an experiment whose aim is to compare the effects of six different rates of seeding on grain yield which will be conducted in an RCBD with four replicates.

The steps to generate randomization for a Single-Factor in Randomized Complete Block Design are listed below:

- Click **Design | Randomized Complete Block Design (RCBD)** from the main window. The **Randomization** dialog box will appear.

Randomization

Randomized Complete Block Design

Design Parameters

Factor Definition:

Levels

	Name	Level	
FACTOR A	factorA	2	

Number of Blocks

Number of Trials

Field Book Filename

OK Cancel

- Specify the required fields and appropriate options.

Levels Spinbox

Define the number of factors. The number of factors specified will define the number of rows of the table inside the **Factor Definition** frame.

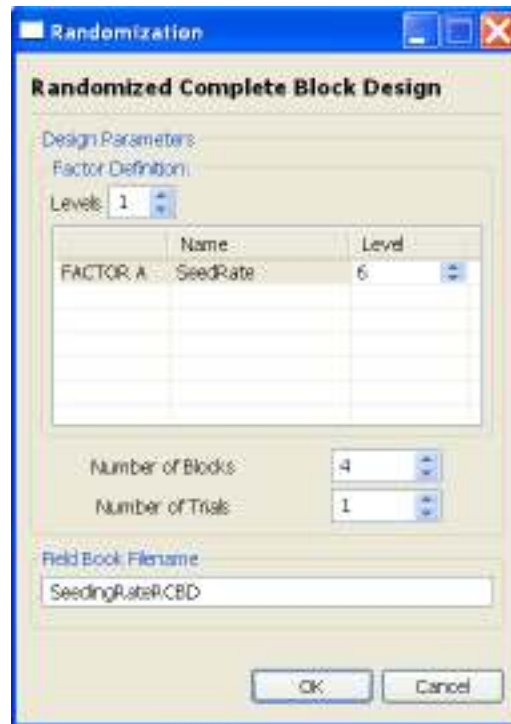
Factor Definition Frame

This table is required. It lets the user define the name of the factor(s) and its levels.

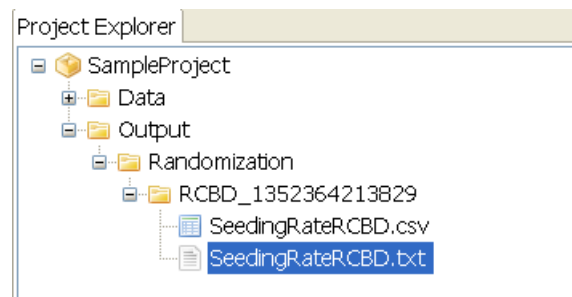
Field Book Filename Textbox

This field is required and it specifies the name where the data file will be created. The default filename is *fieldbookRCBD*.

For this example, the completed dialog box should appear as illustrated below:



- Click the **OK** button to generate the randomization. The **Randomization** dialog box will be minimized and STAR open the data file created.



- Sample data file that was created is shown below:

SeedingRateRCBD.csv

	Trial	Block	SeedRate	PlotNum
1	1	1	SeedRate5	1
2	1	1	SeedRate2	2
3	1	1	SeedRate4	3
4	1	1	SeedRate3	4
5	1	1	SeedRate6	5
6	1	1	SeedRate1	6
7	1	2	SeedRate1	1
8	1	2	SeedRate4	2
9	1	2	SeedRate2	3
10	1	2	SeedRate6	4
11	1	2	SeedRate5	5
12	1	2	SeedRate3	6
13	1	3	SeedRate4	1
14	1	3	SeedRate2	2
15	1	3	SeedRate5	3
16	1	3	SeedRate1	4
17	1	3	SeedRate6	5
18	1	3	SeedRate3	6
19	1	4	SeedRate4	1
20	1	4	SeedRate5	2
21	1	4	SeedRate2	3
22	1	4	SeedRate6	4
23	1	4	SeedRate3	5
24	1	4	SeedRate1	6

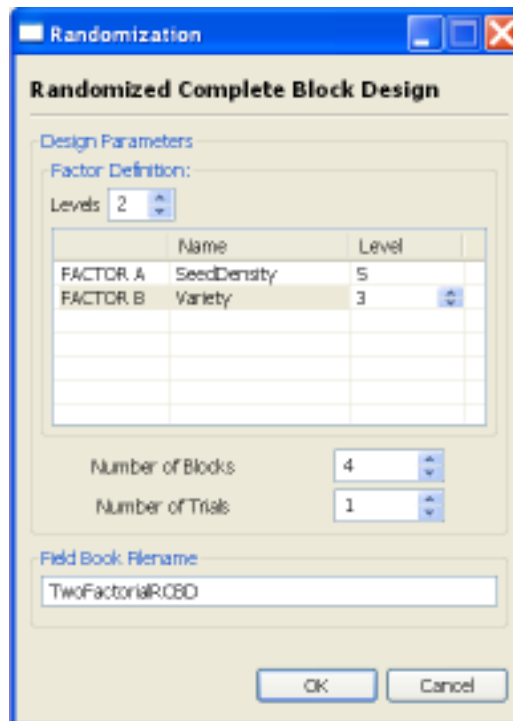
4.2.1. Two-Factor Randomized Complete Block Design (RCBD)

Suppose we want to generate a randomization for an experiment whose aim is to compare the effects of five seeding densities on the on grain yield of three rice varieties which will be conducted in an RCBD with four replicates

The steps to generate randomization for a Two-Factor in Randomized Complete Block Design are listed below:

- Click **Design | Randomized Complete Block Design (RCBD)** from the main window. The **Randomization** dialog box will appear.
- Specify the required fields and appropriate options

For this example, the completed dialog box should appear as illustrated below:



- Click the **OK** button to generate the randomization. The **Randomization** dialog box will be minimized and STAR open the data file created.
- Sample data file created is shown below:

TwoFactorialRCBD.csv					
	Trial	Block	SeedDensity	Variety	PlotNum
1	1	1	SeedDensity4	Variety1	1
2	1	1	SeedDensity1	Variety2	2
3	1	1	SeedDensity4	Variety3	3
4	1	1	SeedDensity2	Variety3	4
5	1	1	SeedDensity5	Variety3	5
6	1	1	SeedDensity2	Variety1	6
7	1	1	SeedDensity5	Variety2	7
8	1	1	SeedDensity3	Variety3	8
9	1	1	SeedDensity2	Variety2	9
10	1	1	SeedDensity5	Variety1	10
11	1	1	SeedDensity3	Variety1	11
12	1	1	SeedDensity3	Variety2	12
13	1	1	SeedDensity1	Variety3	13
14	1	1	SeedDensity4	Variety2	14
15	1	1	SeedDensity1	Variety1	15
16	1	2	SeedDensity2	Variety3	1
17	1	2	SeedDensity3	Variety1	2
18	1	2	SeedDensity3	Variety2	3
19	1	2	SeedDensity5	Variety1	4
20	1	2	SeedDensity1	Variety3	5
21	1	2	SeedDensity4	Variety3	6
22	1	2	SeedDensity2	Variety1	7
23	1	2	SeedDensity1	Variety1	8

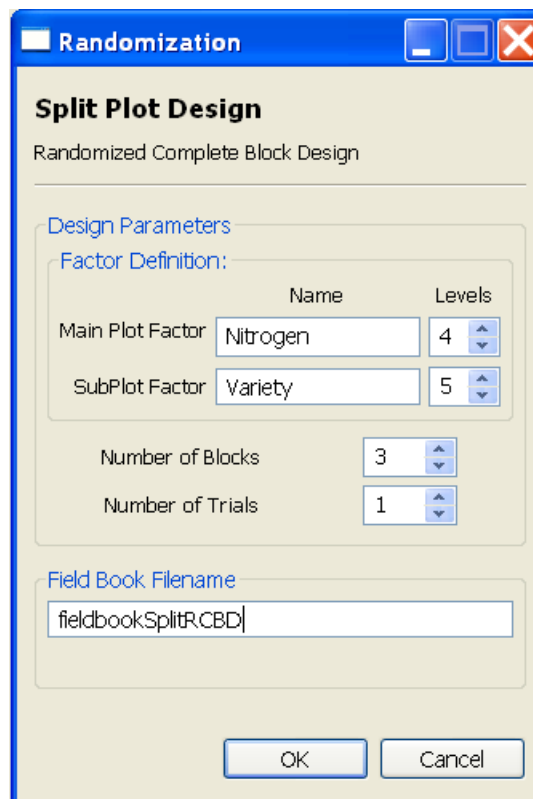
4.3. Split-Plot Design in RCBD

Suppose we want to generate a randomization for an experiment whose aim is to compare the effects of four nitrogen rates on the on grain yield of five rice varieties which will be conducted in RCBD using Split Plot with four replicates

The steps to generate randomization for Split Plot Design in Randomized Complete Block Design are listed below:

- Click **Design | Split Plot Design | Randomized Complete Block Design (RCBD)** from the main window. The **Randomization** dialog box will appear.
- Specify the required field and appropriate options.

For this example, the completed dialog box should appear as illustrated below:

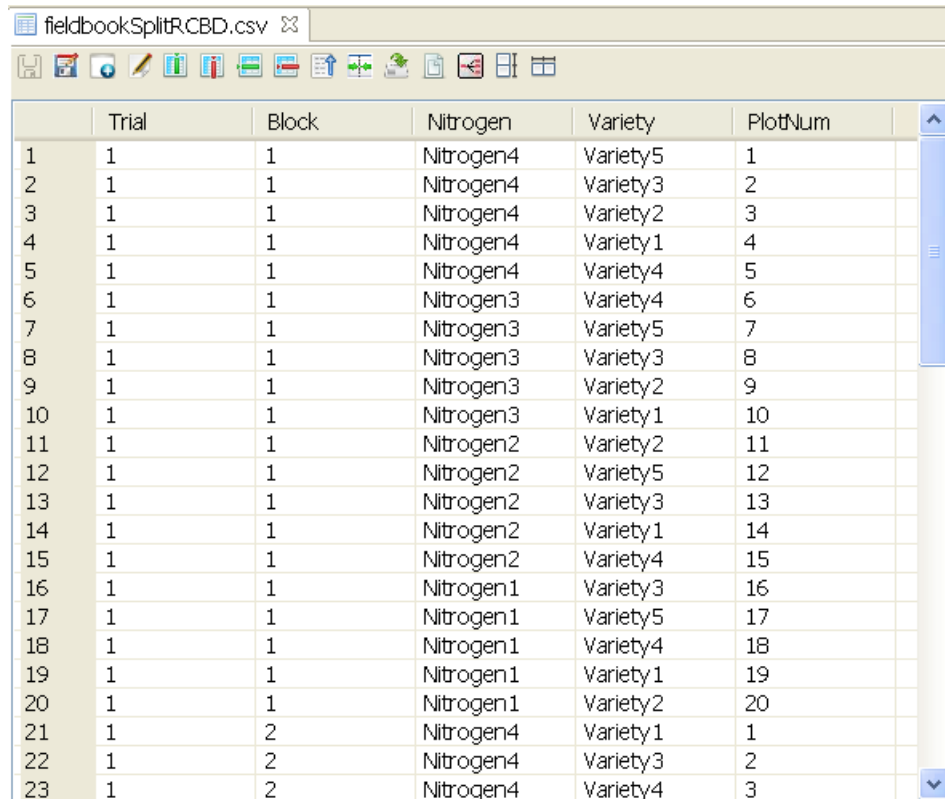


The screenshot shows a dialog box titled "Randomization" with a subtitle "Split Plot Design" and "Randomized Complete Block Design". It contains the following fields and controls:

- Design Parameters**
 - Factor Definition:**

	Name	Levels
Main Plot Factor	Nitrogen	4
SubPlot Factor	Variety	5
 - Number of Blocks: 3
 - Number of Trials: 1
- Field Book Filename:** fieldbookSplitRCBD
- Buttons:** OK, Cancel

- Click the **OK** button to generate the randomization. The **Randomization** dialog box will be minimized and STAR activates the Data Viewer tab.
- Sample data file created is shown below:



	Trial	Block	Nitrogen	Variety	PlotNum
1	1	1	Nitrogen4	Variety5	1
2	1	1	Nitrogen4	Variety3	2
3	1	1	Nitrogen4	Variety2	3
4	1	1	Nitrogen4	Variety1	4
5	1	1	Nitrogen4	Variety4	5
6	1	1	Nitrogen3	Variety4	6
7	1	1	Nitrogen3	Variety5	7
8	1	1	Nitrogen3	Variety3	8
9	1	1	Nitrogen3	Variety2	9
10	1	1	Nitrogen3	Variety1	10
11	1	1	Nitrogen2	Variety2	11
12	1	1	Nitrogen2	Variety5	12
13	1	1	Nitrogen2	Variety3	13
14	1	1	Nitrogen2	Variety1	14
15	1	1	Nitrogen2	Variety4	15
16	1	1	Nitrogen1	Variety3	16
17	1	1	Nitrogen1	Variety5	17
18	1	1	Nitrogen1	Variety4	18
19	1	1	Nitrogen1	Variety1	19
20	1	1	Nitrogen1	Variety2	20
21	1	2	Nitrogen4	Variety1	1
22	1	2	Nitrogen4	Variety3	2
23	1	2	Nitrogen4	Variety4	3

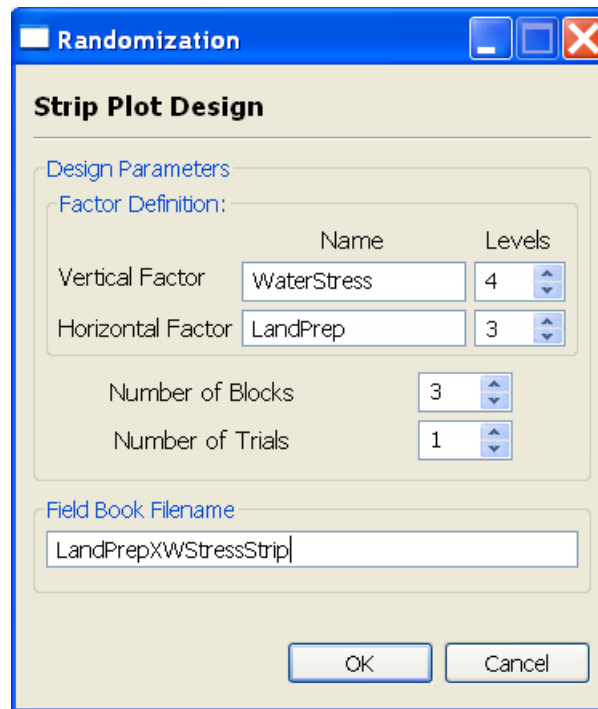
4.4. Strip-Plot Design

Suppose we want to generate a randomization for an experiment to be conducted using Strip Plot, with land preparation (with 3 levels) as the horizontal factor and water stress (with four levels) as the vertical factor. Three replicates will be used.

The steps to generate randomization for a Strip Plot Design are listed below:

- Choose **Design | Strip Plot Design** from the main window. The **Randomization** dialog box will appear.
- Specify the required fields and appropriate options.

For this example, the completed dialog box should appear as illustrated below:



- Click the OK button to generate the randomization. The Randomization dialog box will be minimized and STAR activates the Data Viewer tab.
- Sample data file created is shown below:

Statistical Tool for Agricultural Research - STAR/Projects/SampleProject

File Data Design Analyze

Project Explorer

- SampleProject
 - Data
 - Corr.csv
 - data1_aggregate.csv
 - data1_sorted.csv
 - data1_subset_sorted.csv
 - data1Sort.csv
 - Fertilizer.csv
 - Genus.csv
 - Genus_sorted.csv
 - Insecticides.csv
 - Maize.csv
 - SeedingRate.csv
 - StatusAndAdoption.csv
 - Output
 - Randomization
 - StripPlot_1352704237743
 - LandPrepXWStressStrip.csv
 - LandPrepXWStressStrip.txt

LandPrepXWStressStrip.csv

	Trial	Block	WaterStress	LandPrep	PlotNum
1	1	1	WaterStress2	LandPrep3	1
2	1	1	WaterStress2	LandPrep1	2
3	1	1	WaterStress2	LandPrep2	3
4	1	1	WaterStress1	LandPrep3	4
5	1	1	WaterStress1	LandPrep1	5
6	1	1	WaterStress1	LandPrep2	6
7	1	1	WaterStress3	LandPrep3	7
8	1	1	WaterStress3	LandPrep1	8
9	1	1	WaterStress3	LandPrep2	9
10	1	1	WaterStress4	LandPrep3	10
11	1	1	WaterStress4	LandPrep1	11
12	1	1	WaterStress4	LandPrep2	12
13	1	2	WaterStress2	LandPrep1	1
14	1	2	WaterStress2	LandPrep2	2
15	1	2	WaterStress2	LandPrep3	3
16	1	2	WaterStress1	LandPrep1	4
17	1	2	WaterStress1	LandPrep2	5
18	1	2	WaterStress1	LandPrep3	6
19	1	2	WaterStress4	LandPrep1	7
20	1	2	WaterStress4	LandPrep2	8
21	1	2	WaterStress4	LandPrep3	9

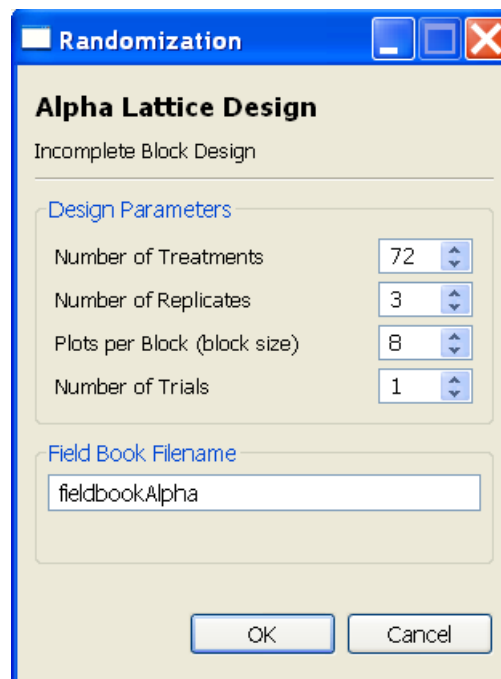
4.5. Alpha Lattice Design

Suppose we want to generate a randomization for an experiment that will be conducted using Alpha Lattice design with 72 entries grouped into 8 with three replicates.

The steps to generate randomization for Alpha Lattice Design are listed below:

- Click **Design | Incomplete Block Design | Alpha Lattice Design** from the main window. The **Randomization** dialog box will appear.

For this example, the completed dialog box should appear as illustrated below:



The screenshot shows a Windows-style dialog box titled "Randomization" with a blue title bar. Inside, the "Alpha Lattice Design" section is active, showing "Incomplete Block Design". Under "Design Parameters", there are four spinners: "Number of Treatments" set to 72, "Number of Replicates" set to 3, "Plots per Block (block size)" set to 8, and "Number of Trials" set to 1. Below this, the "Field Book Filename" text box contains "fieldbook.Alpha". At the bottom are "OK" and "Cancel" buttons.

- Click the **OK** button to generate the randomization. The **Randomization** dialog box will be minimized and STAR activates the Data Viewer tab.
- Sample data file created is shown below:

	Trial	replication	block	cols	Treatment	plots
1	1	1	1	1	T55	1
2	1	1	1	2	T42	2
3	1	1	1	3	T63	3
4	1	1	1	4	T23	4
5	1	1	1	5	T72	5
6	1	1	1	6	T18	6
7	1	1	1	7	T22	7
8	1	1	1	8	T24	8
9	1	1	2	1	T4	9
10	1	1	2	2	T12	10
11	1	1	2	3	T39	11
12	1	1	2	4	T27	12
13	1	1	2	5	T29	13
14	1	1	2	6	T69	14
15	1	1	2	7	T6	15
16	1	1	2	8	T52	16
17	1	1	3	1	T21	17
18	1	1	3	2	T62	18
19	1	1	3	3	T44	19
20	1	1	3	4	T43	20
21	1	1	3	5	T31	21
22	1	1	3	6	T3	22
23	1	1	3	7	T45	23
24	1	1	3	8	T19	24
25	1	1	4	1	T7	25
26	1	1	4	2	T61	26
27	1	1	4	3	T60	27
28	1	1	4	4	T54	28
29	1	1	4	5	T37	29
30	1	1	4	6	T20	30

4.6. Augmented Design in RCB

Suppose we want to generate a randomization for an experiment which will be conducted using Augmented design in RCB involving four replicates. The experiment will use six replicated treatment and 60 unreplicated treatment (test entries).

The steps to generate randomization for a Augmented Design in Randomized Complete Block are listed below:

- Click **Design | Augmented Design | Randomized Complete Block Design (RCBD)** from the main window. The **Randomization** dialog box will appear.

For this example, the completed dialog box should appear as illustrated below:

Randomization

Augmented Design

Randomized Complete Block Design

Design Parameters

Number of Replicated Treatments: 6

Number of Replicates: 4

Number of Unreplicated Treatments: 60

Number of Trials: 1

Field Book Filename

fieldbookAugRCBD

OK Cancel

- Click the **OK** button to generate the randomization. The **Randomization** dialog box will be minimized and STAR activates the Data Viewer tab.
- Sample data file created is shown below:

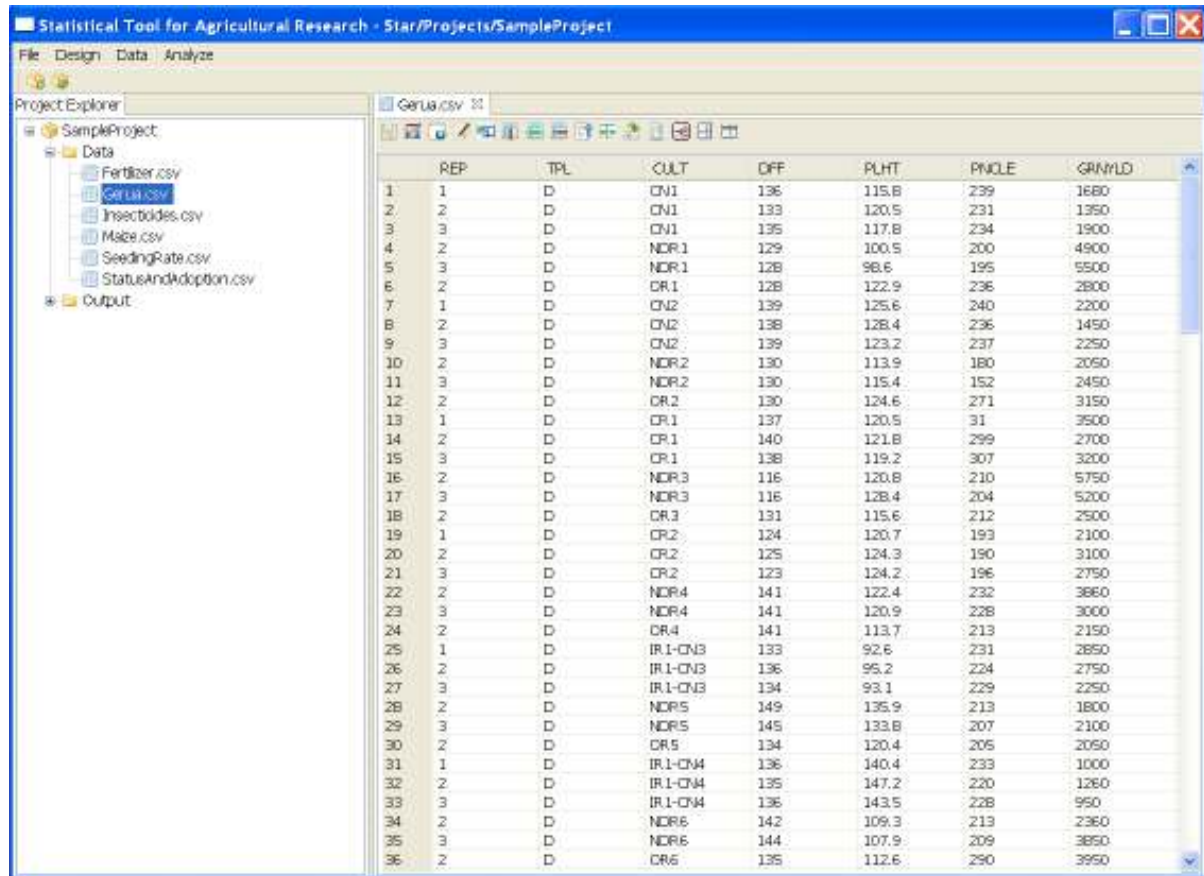
	Trial	Block	Treatment	PlotNum
1	1	1	new20	1
2	1	1	new45	2
3	1	1	new12	3
4	1	1	check6	4
5	1	1	new41	5
6	1	1	check2	6
7	1	1	new7	7
8	1	1	check3	8
9	1	1	new14	9
10	1	1	new8	10
11	1	1	check4	11
12	1	1	new13	12
13	1	1	new4	13
14	1	1	new33	14
15	1	1	new44	15
16	1	1	check5	16
17	1	1	new39	17
18	1	1	new49	18
19	1	1	check1	19
20	1	1	new22	20
21	1	1	new19	21
22	1	2	new37	1
23	1	2	new27	2
24	1	2	check4	3

5. Descriptive Statistics

The **Descriptive** submenu of STAR performs univariate summary statistics for several variables and/or per levels of grouping variables.

The steps to perform Descriptive are listed below:

- On the Project Explorer, locate the dataset *Gerua.csv* from the *Data* folder of the project named *SampleProject*. Double-click the file to open it and view it in the Data Viewer.



	REP	TPL	CULT	OFF	PLHT	PINOLE	GRN/WD
1	1	D	CN1	136	115.8	239	1680
2	2	D	CN1	133	120.5	231	1350
3	3	D	CN1	135	117.8	234	1900
4	2	D	NDR1	129	100.5	200	4900
5	3	D	NDR1	128	98.6	195	5500
6	2	D	CR1	128	122.9	236	2800
7	1	D	CR2	139	125.6	240	2200
8	2	D	CR2	138	128.4	236	1450
9	3	D	CR2	139	123.2	237	2250
10	2	D	NDR2	130	113.9	180	2050
11	3	D	NDR2	130	115.4	152	2450
12	2	D	CR2	130	124.6	271	3150
13	1	D	CR1	137	120.5	31	3500
14	2	D	CR1	140	121.8	299	2700
15	3	D	CR1	138	119.2	307	3200
16	2	D	NDR3	116	120.8	210	5750
17	3	D	NDR3	116	128.4	204	5200
18	2	D	CR3	131	115.6	212	2500
19	1	D	CR2	124	120.7	193	2100
20	2	D	CR2	125	124.3	190	3100
21	3	D	CR2	123	124.2	196	2750
22	2	D	NDR4	141	122.4	232	3860
23	3	D	NDR4	141	120.9	228	3000
24	2	D	CR4	141	113.7	213	2150
25	1	D	IR1-CN3	133	92.6	231	2850
26	2	D	IR1-CN3	136	95.2	224	2750
27	3	D	IR1-CN3	134	93.1	229	2250
28	2	D	NDR5	149	135.9	213	1800
29	3	D	NDR5	145	133.8	207	2100
30	2	D	CR5	134	120.4	205	2050
31	1	D	IR1-CN4	136	140.4	233	1000
32	2	D	IR1-CN4	135	147.2	220	1260
33	3	D	IR1-CN4	136	143.5	228	950
34	2	D	NDR6	142	109.3	213	2360
35	3	D	NDR6	144	107.9	209	3850
36	2	D	CR6	135	112.6	290	3950

- To generate descriptive statistics, from the main window click **Analyze | Descriptive Statistics | Descriptive**. The **Descriptive Statistics** dialog box will appear.
- Specify the required field and appropriate options for the analysis.

Variable Description Tab

Variable(s)

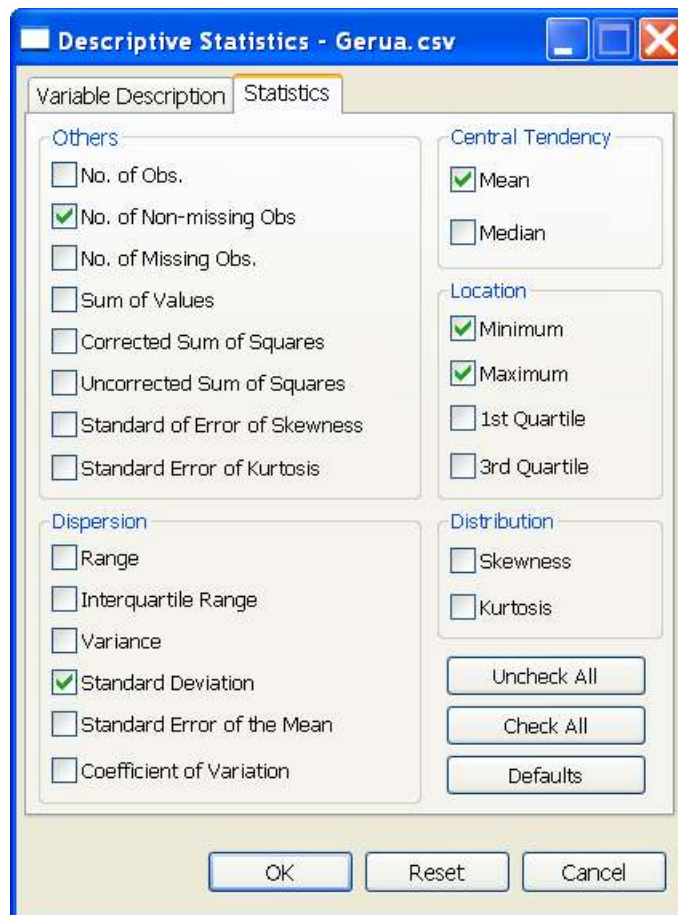
This field is required and at least one entry is need for the analysis to proceed.

By Variable(s)

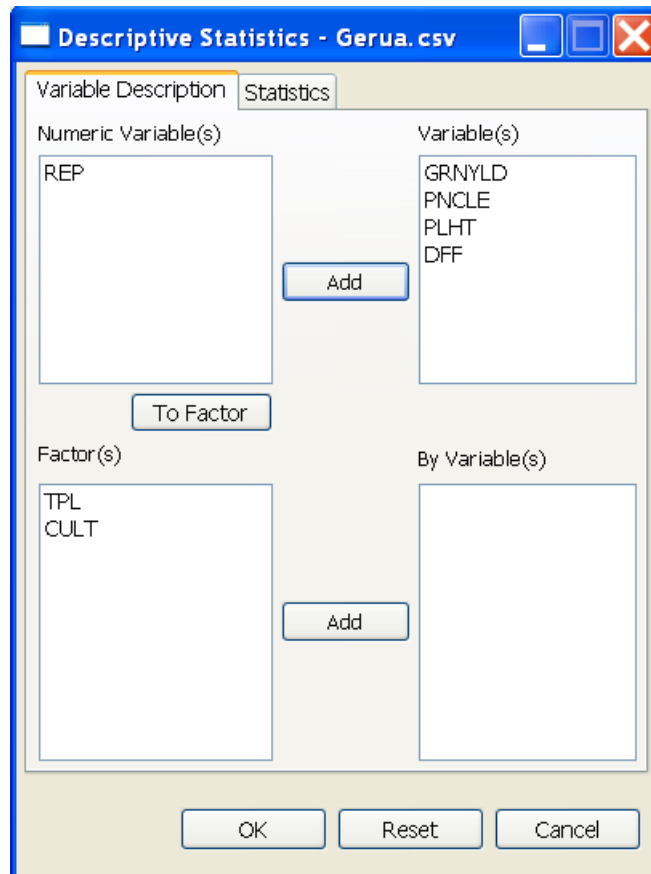
This field is optional. If a **By Variable(s)** is specified, it will generate descriptive statistic per category of the **By Variable**.

Statistics Tab

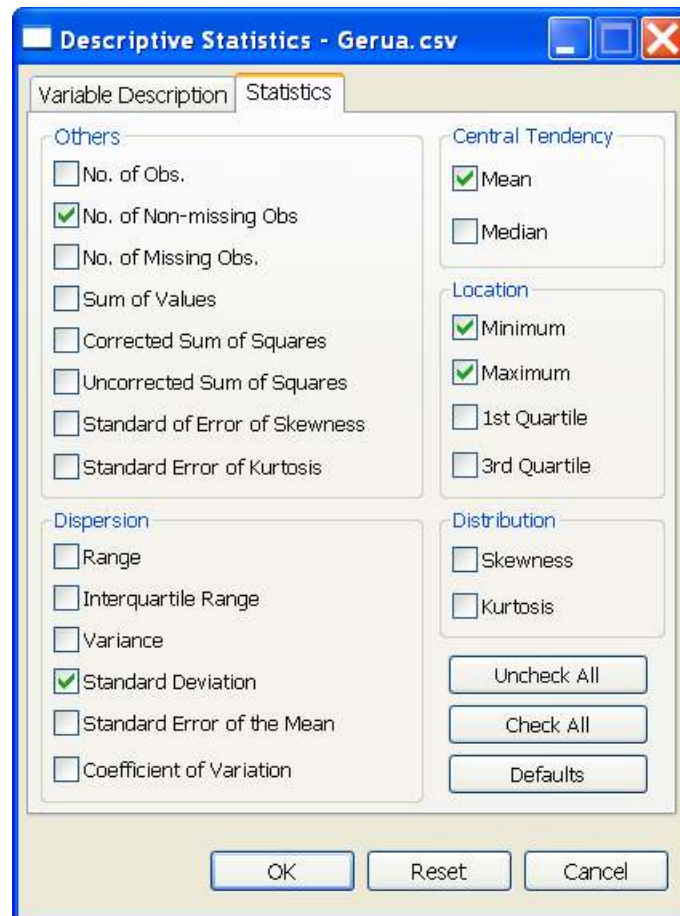
This tab lets the user identify the numerical descriptive measures to be displayed. By default, the procedure will display five statistics, namely: number of non-missing observations (NNMissObs), minimum (min), maximum (max), mean, and standard deviation (stdev). To request for all statistics, click the **Check All** button of the **Statistics** tab. You may click on the **Uncheck All** button to remove previously selected items and start with a new selection.



Suppose we want to compute the descriptive statistics for variables *DFF*, *PLHT*, *PNCLE* and *GRNYLD*. The completed **Variable Description** tab should appear as illustrated below:



- The completed **Statistics** tab should appear as illustrated below:

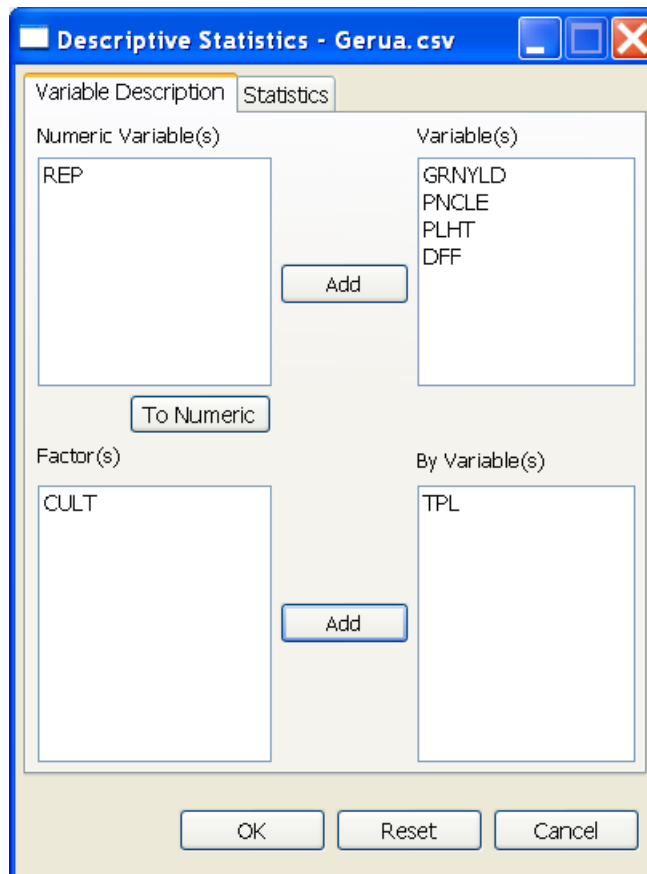


- Click the **Ok** button to generate the summary statistics. The **Descriptive Statistics** dialog box will be minimize and STAR activates the **Output Viewer** tab.
- Sample output of the descriptive statistics is shown below:

DESCRIPTIVE STATISTICS

Variable	N_NonMissObs	Min	Max	Mean	StdDev
GRNYLD	156	950.00	6500.00	3759.20	1323.36
PNCLE	156	29.00	336.00	230.12	46.19
PLHT	156	92.60	163.50	128.48	15.29
DFF	156	14.00	151.00	125.26	14.65

- Suppose we want to compute the descriptive statistics for variables *DFF*, *PLHT*, *PNCLE* and *GRNYLD* per level of *TPL*. The completed **Variable Description** tab should appear as illustrated below:



- Click the **Ok** button to generate the summary statistics. Sample output of the descriptive statistics is shown below:

DESCRIPTIVE STATISTICS

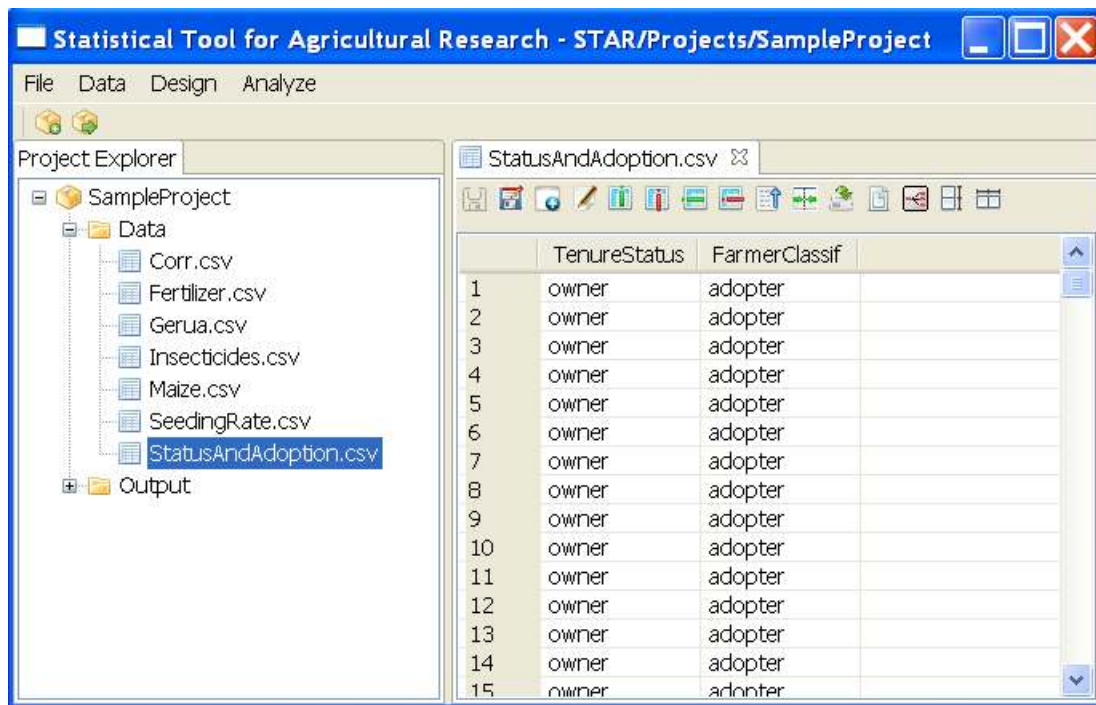
Variable	TPL	N_NonMissObs	Min	Max	Mean	StdDev
GRNYLD	D	78	950.00	5750.00	2919.68	1226.98
GRNYLD	N	78	2900.00	6500.00	4598.72	769.37
PNCLE	D	78	31.00	307.00	212.19	42.49
PNCLE	N	78	29.00	336.00	248.05	42.87
PLHT	D	78	92.60	149.30	120.05	11.38
PLHT	N	78	115.20	163.50	136.90	14.05
DFF	D	78	14.00	151.00	133.85	15.64
DFF	N	78	106.00	128.00	116.67	6.14

6. Cross Tabulation

The cross tabulation forms two-way tables and provides a variety of tests and measures of association for a two-way tables.

The steps to perform Cross Tabulation are listed below:

- Locate the dataset *StatusAndAdoption.csv* from the *Data* folder of the project named *SampleProject* in the Project Explorer. Double-click the file to open it and view it in the Data Viewer.



- To generate cross table, from the main window click **Analyze | Descriptive Statistics | Cross Tabulation**. The **Cross Tabulation** dialog box will appear.
- Specify the required field and the appropriate options for the analysis.

Variable Description Tab

For the analysis to proceed, the **Row Variable(s)** list box and the **Column Variable(s)** list box from the **Variable Description** tab should both have at least one entry and at least one checkbox should be tick from the **Option** tab.

Options Tab

For the analysis to proceed, at least one checkbox should be tick in this tab.

Cell Display

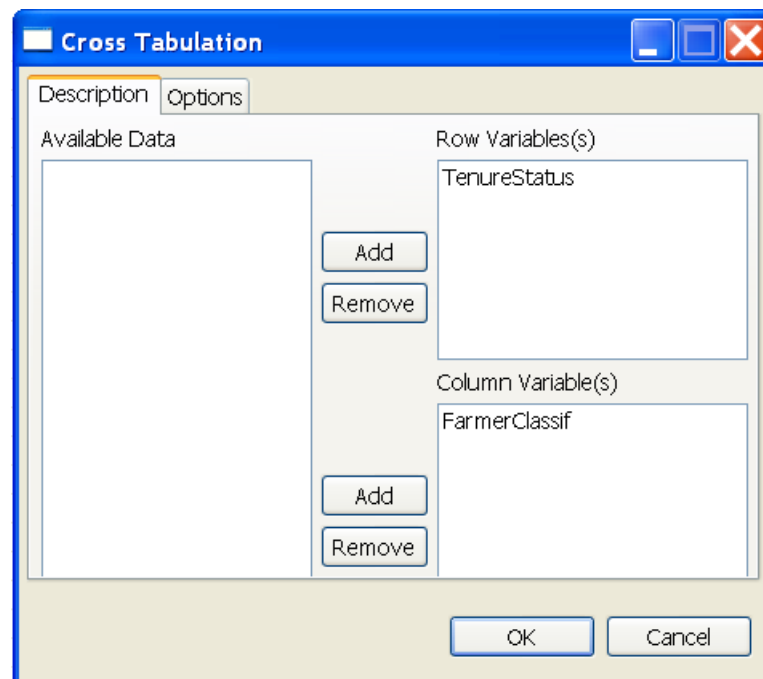
The Crosstabs procedure displays the observed and expected frequency. Each cell of the table may contain counts and percentages.

- **Counts**
Display the number of observed frequency and number of expected frequency. By default, the observed frequency is displayed.
- **Percentages**
This field is optional. It may display the row, column or the total percentages.
- **Residuals**

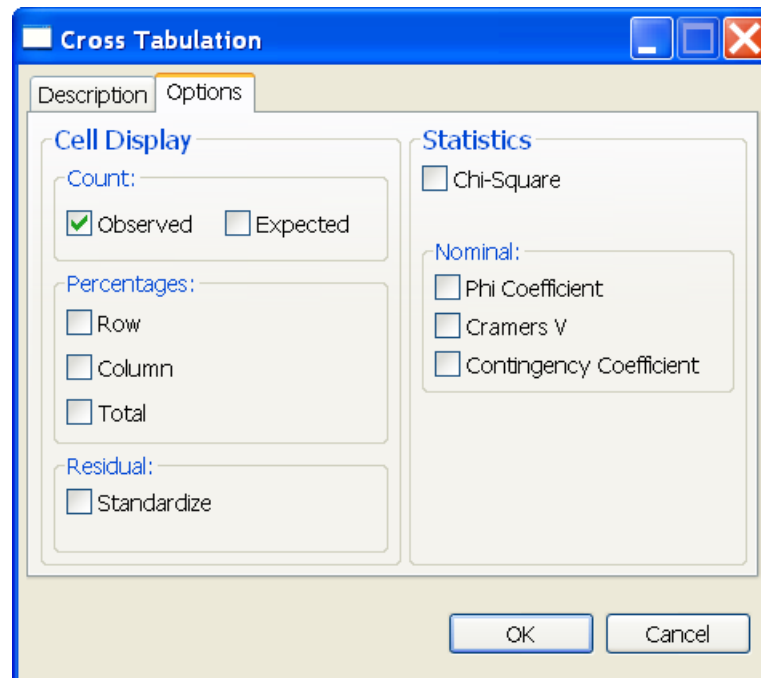
Statistics

- **Chi-square**
Select Chi-square to calculate the Chi-square Test of Independence and the likelihood-ratio chi-square.
- **Nominal**
For nominal data, three measure of association based on chi-square (Phi Coefficient, Cramer's V and Contingency Coefficient) may be requested.

For the example, the completed **Variable Description** tab should appear as illustrated below:



The completed **Options** tab should appear as illustrated below:



- Click the **Ok** button. Sample output of the analysis using the default option is shown below:

CROSS TABULATION

Table of Observed Frequency: TenureStatus by FarmerClassif

	FarmerClassif		
TenureStatus	adopter	nonadopter	Total
fixed-rent	4	3	7
owner	102	26	128
share-rent	42	10	52
Total	148	39	187

- Sample output of the analysis if all checkbox is tick is shown below:

CROSS TABULATION

Table of TenureStatus by FarmerClassif

		FarmerClassif		
TenureStatus		adopter	nonadopter	Total
fixed-rent	Obs Freq	4	3	7
	Expected Freq	5.5401	1.4599	7.0000

Statistical Tools for Agricultural Research (STAR) Tutorial

owner	Obs Freq		102	26	128
	Expected Freq		101.3048	26.6952	128.0000
-----+-----+-----+-----+-----+					
share-rent	Obs Freq		42	10	52
	Expected Freq		41.1551	10.8449	52.0000
-----+-----+-----+-----+-----+					
Total	Obs Freq		148	39	187
	Expected Freq		148.0000	39.0000	187.0000
-----+-----+-----+-----+-----+					

Statistics for Table TenureStatus by FarmerClassif

Statistics	DF	Value	Prob

Pearson Chi-Square	2	2.1589	0.3398
Likelihood Ratio Chi-Square	2	1.8236	0.4018
Phi Coefficient		0.1074	
Contingency Coefficient		0.1068	
Cramer's V		0.1074	

* Cells with Expected Frequency < 5: 2 of 6 (33.33%)

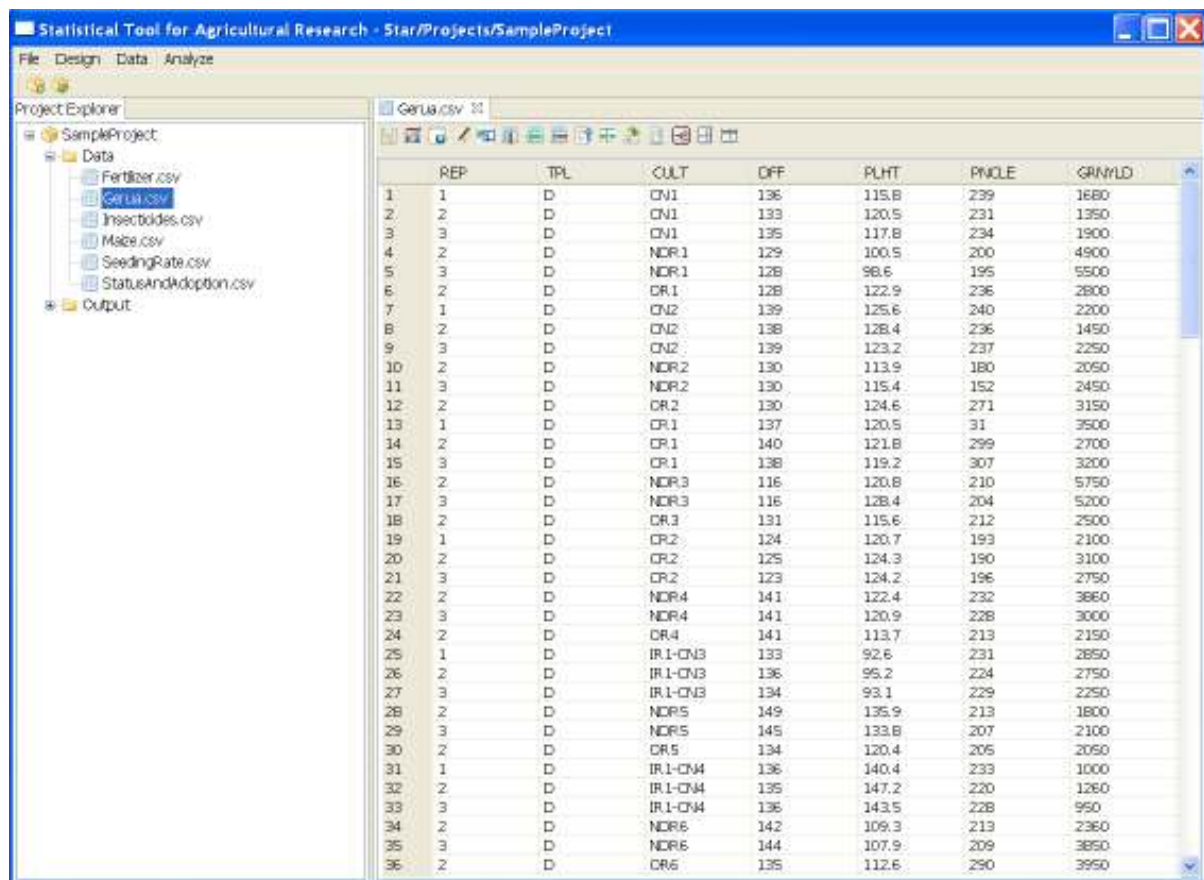
7. T-test

7.1. One Sample t-tests

The One Sample *t*-test procedure determine whether the mean of one variable differ from the hypothesized value. It assumes that the samples are randomly taken from the population of interest and that the observations are obtained from a normal distribution.

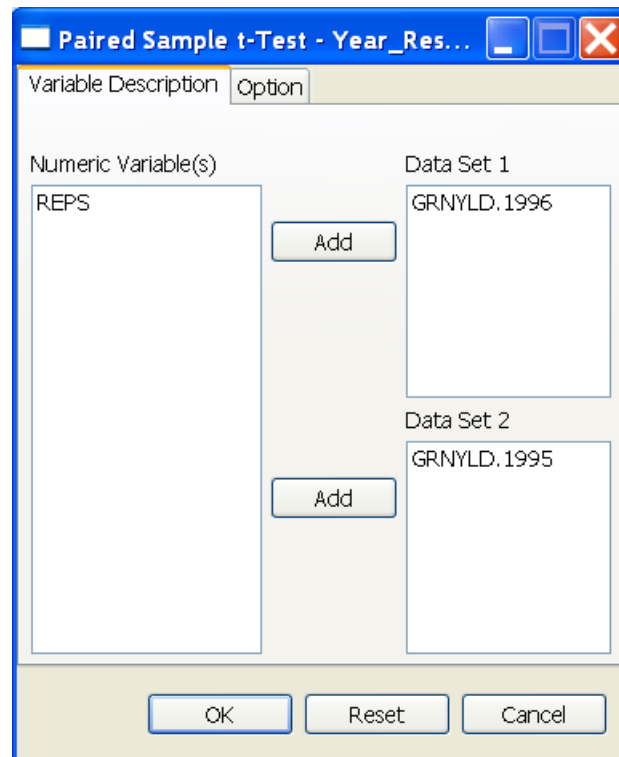
The steps to perform Heterogeneity test are listed below:

- Locate the dataset *Gerua.csv* from the *Data* folder of the project named *SampleProject* in the Project Explorer.



	REP	TPL	CULT	OFF	PLHT	PNOLE	GRN/ID
1	1	D	CN1	136	115.8	239	1680
2	2	D	CN1	133	120.5	231	1950
3	3	D	CN1	135	117.8	234	1900
4	2	D	NDR1	129	100.5	200	4900
5	3	D	NDR1	128	98.6	195	5500
6	2	D	CR1	128	122.9	236	2800
7	1	D	CN2	139	125.6	240	2200
8	2	D	CN2	138	128.4	236	1450
9	3	D	CN2	139	123.2	237	2250
10	2	D	NDR2	130	113.9	180	2050
11	3	D	NDR2	130	115.4	152	2450
12	2	D	CR2	130	124.6	271	3150
13	1	D	CR1	137	120.5	31	3500
14	2	D	CR1	140	121.8	299	2700
15	3	D	CR1	138	119.2	307	3200
16	2	D	NDR3	116	120.8	210	5750
17	3	D	NDR3	116	128.4	204	5200
18	2	D	CR3	131	115.6	212	2500
19	1	D	CR2	124	120.7	193	2100
20	2	D	CR2	125	124.3	190	3100
21	3	D	CR2	123	124.2	196	2750
22	2	D	NDR4	141	122.4	232	3860
23	3	D	NDR4	141	120.9	228	3000
24	2	D	CR4	141	113.7	213	2150
25	1	D	IR1-CN3	133	92.6	231	2850
26	2	D	IR1-CN3	136	95.2	224	2750
27	3	D	IR1-CN3	134	93.1	229	2250
28	2	D	NDR5	149	135.9	213	1800
29	3	D	NDR5	145	133.8	207	2100
30	2	D	CR5	134	120.4	205	2050
31	1	D	IR1-CN4	136	140.4	233	1000
32	2	D	IR1-CN4	135	147.2	220	1260
33	3	D	IR1-CN4	136	143.5	228	950
34	2	D	NDR6	142	109.3	213	2360
35	3	D	NDR6	144	107.9	209	3850
36	2	D	CR6	135	112.6	290	3950

- Click **Analyze | t-Test | One Sample**. The **One Sample** dialog box will appear.
- Opening the data for the first time, *Rep* field is regarded by R as numerical variables, they need to be changed as factors. Choose the variable and click on the **To Factor** button.
- Specify the required field and appropriate options for the analysis.

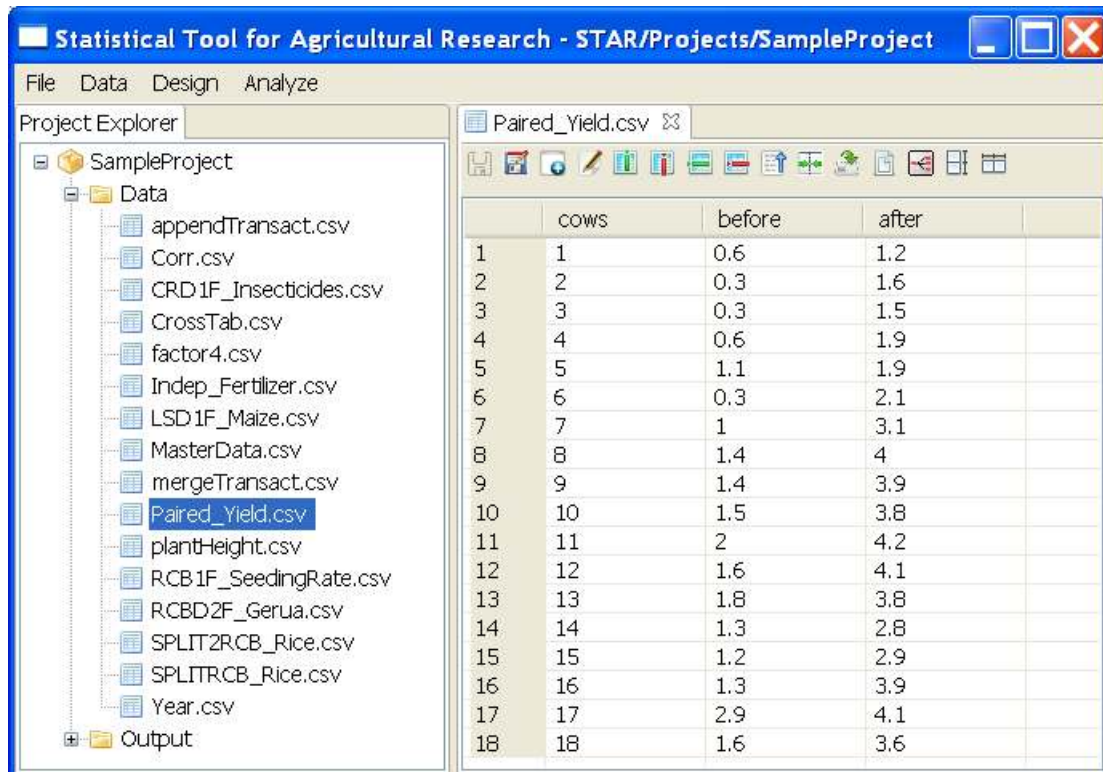


7.2. Paired Sample t -tests

The paired sample t -test is used to compares the means of paired or related samples. It assumes that the samples are randomly taken from the population of interest and the population of differences is normally distributed.

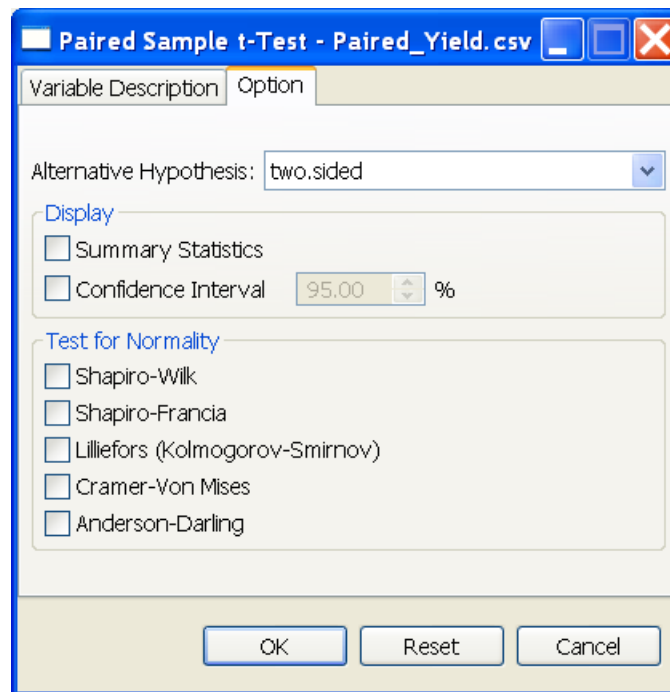
The steps to perform paired t -test are listed below:

- Locate the dataset *Paired_.csv* from the *Data* folder of the project named *SampleProject* in the Project Explorer.



- Click **Analyze | t-Tests | Paired Samples**. The **Paired Samples** dialog box will appear.
- Specify the required field and appropriate options for the analysis.

Options Tab



Alternative Hypothesis

User can select a one-tailed or two-tailed test. The default alternative hypothesis used is that for the two-sided test.

Level of Significance

This will be used to determine if the two groups have equal variance. Its default value is 0.05. The user can change this value by specifying a numeric value from zero to 1.

Display Summary Statistics

If this option is selected, a summary table with number of observations, minimum, maximum, mean and standard deviation will be displayed.

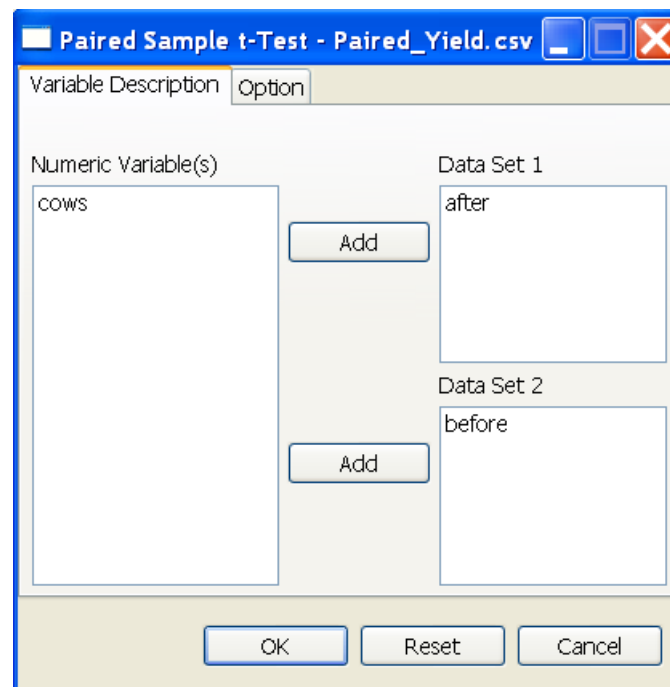
Display Confidence Interval

If this option is selected, a confidence interval of the mean will be displayed. If this is selected, by default, a 95% confidence interval is displayed. Valid value is between 90% to 99%.

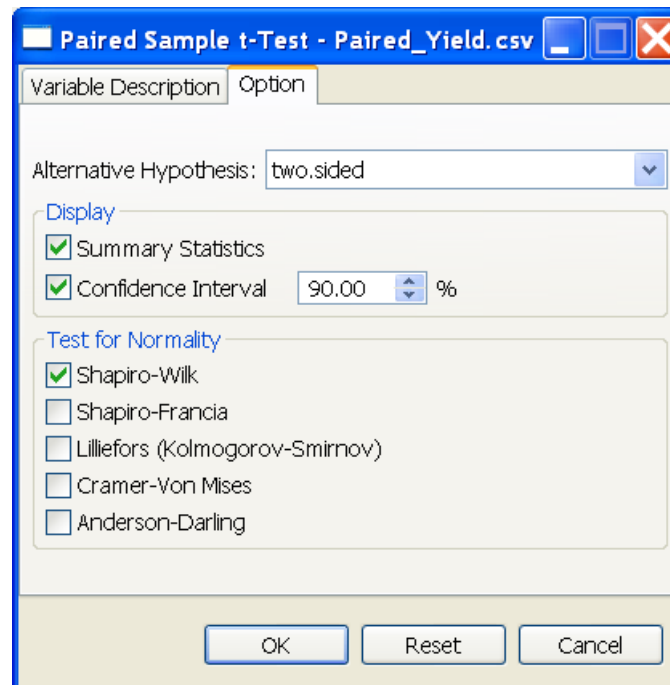
Test Procedure

If at least one option is selected, if display the test for normality test for each level of the **Grouping Variable**. There are five available test for normality procedure.

For the example, the completed **Variable Description** tab should appear as illustrated below:



The completed **Option** tab should appear as illustrated below:



- Click the **OK** button to perform the analysis.
- Sample output of the analysis is shown below.

Test for Normality

Difference	Method	Stat	Value	p Value
after - before	Shapiro-Wilk	W	0.9420	0.3132

Descriptive Statistics

Difference	N	Lower CI*	Mean	Upper CI*	StdDev	SE_Mean
after - before	18	1.53	1.79	2.04	0.6230	0.1468

* At 90% Confidence Level.

PAIRED SAMPLE t-TEST, $h_0 = 0$

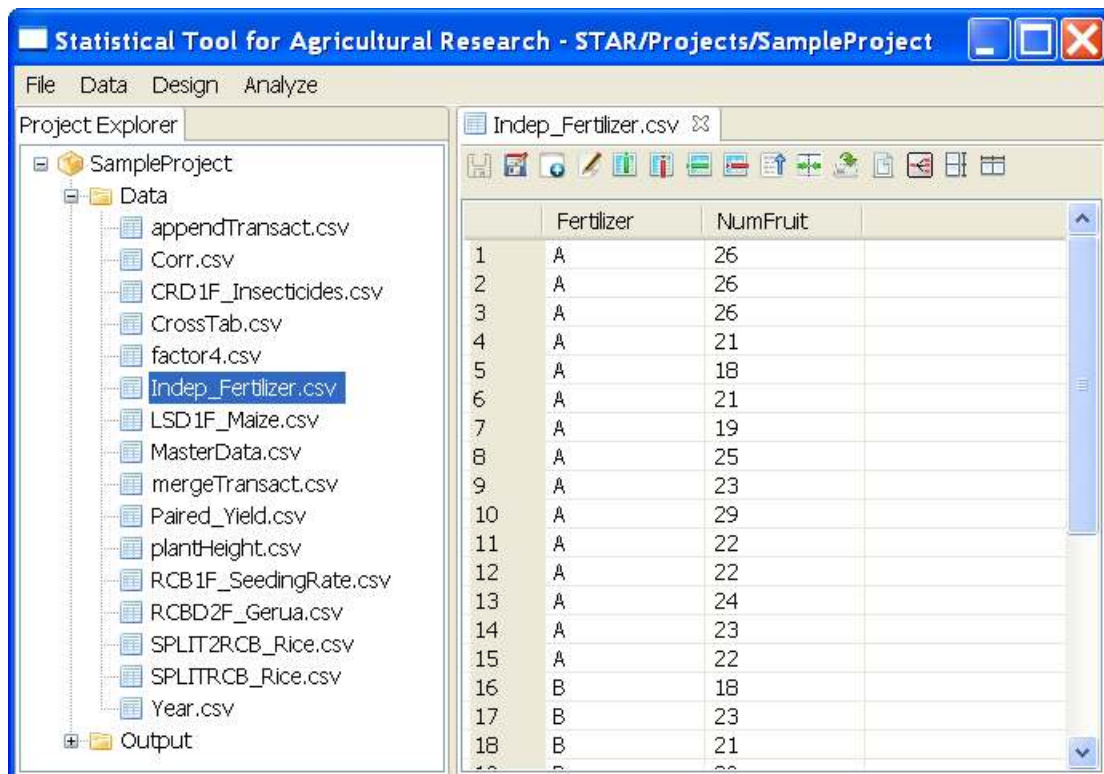
Difference	DF	t Value	Pr > t
after - before	17	12.18	0.0000

7.3. Independent Sample t-tests

The independent sample *t*-test is used to determine whether the means between two groups or populations are equal. It assumes that (i) the samples are drawn independently from each population; (ii) the samples from each population are randomly taken; and (iii) the observations from each population are obtained from a normal distribution.

The steps to perform *t*-Test for two independent samples are listed below:

- On the Project Explorer, locate the dataset *Indep_Fertilizer.csv* from in the *Data* folder of the project named *SampleProject*. Double-click the file to open and view it in the Data Viewer. The file contains the number of fruits (*NumFruit*) from trees sprayed by two fertilizers (*Fertilizer – A and B*). Suppose we want to test the hypothesis that the mean numbers of fruits from trees sprayed by the two fertilizers are the same.



	Fertilizer	NumFruit
1	A	26
2	A	26
3	A	26
4	A	21
5	A	18
6	A	21
7	A	19
8	A	25
9	A	23
10	A	29
11	A	22
12	A	22
13	A	24
14	A	23
15	A	22
16	B	18
17	B	23
18	B	21

- To perform *t*-test on two independent samples, from the main window of STAR click **Analyze | t-test | Independent Sample**. The **Two Independent Sample** dialog box will appear.

Variable Description Tab

Test Variable(s)

This field is required for the analysis to proceed. This field need at least one entry and should come from the **Numeric Variable(s)** list box.

Grouping Variable(s)

This field is required and will only accept one entry at a time. The entry must came from the **Factor(s)** list box. The entry should have only two levels.

Options Tab

Alternative Hypothesis

User can select a one-tailed or two-tailed test. The default alternative hypothesis used is that for the two-sided test.

Level of Significance

This will be used to determine if the two groups have equal variance. Its default value is 0.05. The user can change this value by specifying a numeric value from zero to 1.

Display Summary Statistics

If this option is selected, a summary table with number of observations, minimum, maximum, mean and standard deviation will be displayed.

Display Confidence Interval

If this option is selected, a confidence interval of the mean will be displayed. If this is selected, by default, a 95% confidence interval is displayed. Valid value is between 90% to 99%.

Test Procedure

If at least one option is selected, if display the test for normality test for each level of the **Grouping Variable**. There are five available test for normality procedure.

For the example, the completed **Variable Description** tab should appear as illustrated below:

The screenshot shows a dialog box titled "t-Test Independent Sample - Fertilizer.csv". It has two tabs: "Variable Description" and "Option", with "Option" being the active tab. The dialog is divided into four main sections. The top-left section, labeled "Numeric Variable(s)", is empty. The top-right section, labeled "Test Variable(s)", contains the text "NumFruit". Below this section is an "Add" button. The bottom-left section, labeled "Factor(s)", is empty. The bottom-right section, labeled "Grouping Variable", contains the text "Fertilizer". Below this section is a "Remove" button. Between the "Numeric Variable(s)" and "Factor(s)" sections is a "To Factor" button. At the bottom of the dialog are three buttons: "OK", "Reset", and "Cancel".

The completed **Option** tab should appear as illustrated below:

This screenshot shows the same dialog box as the previous one, but with the "Option" tab fully configured. The "Alternative Hypothesis" is set to "two.sided" in a dropdown menu. The "Level of Sig. (Testing Variance Equality)" is set to "0.05" in a text box with up and down arrows. There are two sections with blue headers. The first section, "Display", contains two checkboxes: "Summary Statistics" and "Confidence Interval". The "Confidence Interval" checkbox is checked, and next to it is a text box with "90.00" and a percentage sign. The second section, "Test for Normality", contains five checkboxes: "Shapiro-Wilk", "Shapiro-Francia", "Lilliefors (Kolmogorov-Smirnov)", "Cramer-Von Mises", and "Anderson-Darling". The "Shapiro-Wilk" checkbox is checked. At the bottom are the "OK", "Reset", and "Cancel" buttons.

- Click the **OK** button to perform the analysis.
- Sample output of the analysis is shown below. The default output of the t-Test for two independent samples includes the test for equality of variances. Depending on the result of the equality of variances, Pooled t-Test or a t-test with satterwaite adjustment is computed.

TEST FOR NORMALITY

Grp	Level	Variable	Method	Stat	Value	p Value
Fertilizer A		NumFruit	Shapiro-Wilk	W	0.9726	0.8950
Fertilizer B		NumFruit	Shapiro-Wilk	W	0.9198	0.1916

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
NumFruit	Folded F	14	14	1.47	0.4820

TWO INDEPENDENT SAMPLE t-TEST, $h_0 = 0$

Variable	Method*	Variances	DF	t Value	Pr > t
NumFruit	Pooled	Equal	28	1.91	0.0663

* At 0.05 level of significance.

8. Analysis of Variance

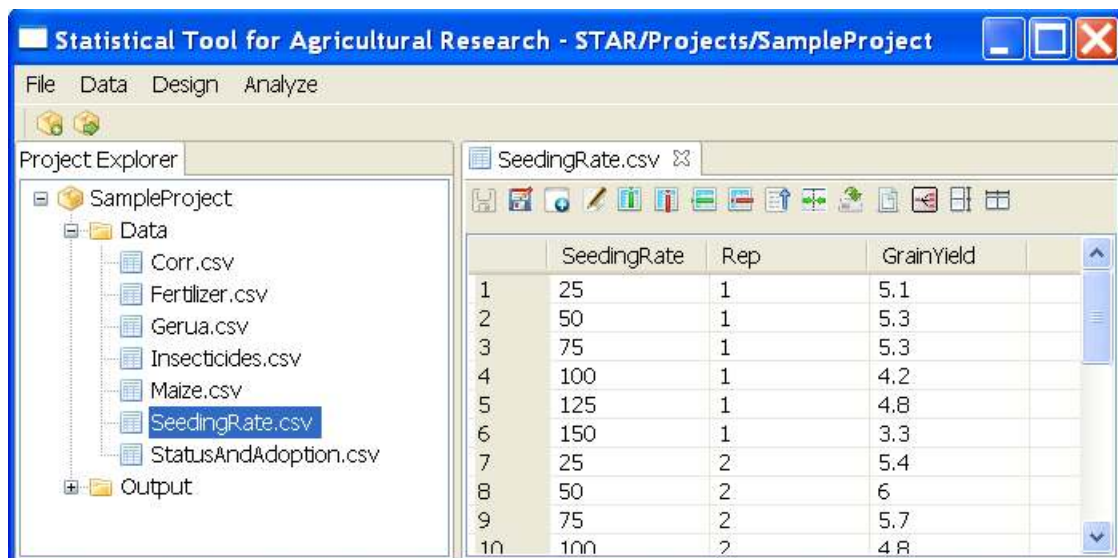
The **Analysis of Variance submenu** can be used to perform analysis of variance for data from experimental designs with single error (e.g., RCB) and with 2 or more errors (e.g., Split Plot).

Pairwise mean comparisons and contrast analyses are available to evaluate differences among specific treatment/treatment combination.

8.1. One Factor Randomized Complete Block

The steps to perform the Analysis of Variance in Randomized Complete Block Design are listed below:

- On the Project Explorer the dataset *SeedingRate.csv* located in the *Data* folder of the project named *SampleProject*. Double-click the file to open and view it in the Data Viewer. The example file contains data from an experiment conducted to investigate the effect of six different rates of seeding (*SeedingRate*, in kg seed/ha with levels 25, 50, 75, 100, 125 and 150) on the grain yield (*GrainYield*, in kg/ha) of rice a variety using Randomized Complete Block (RCB) field design with four blocks (*Rep* with levels 1, 2, 3 and 4).



The screenshot shows the STAR software interface. On the left, the Project Explorer displays a tree structure for 'SampleProject' with a 'Data' folder containing several CSV files. 'SeedingRate.csv' is selected. On the right, the Data Viewer displays the contents of 'SeedingRate.csv' as a table with columns 'SeedingRate', 'Rep', and 'GrainYield'.

	SeedingRate	Rep	GrainYield
1	25	1	5.1
2	50	1	5.3
3	75	1	5.3
4	100	1	4.2
5	125	1	4.8
6	150	1	3.3
7	25	2	5.4
8	50	2	6
9	75	2	5.7
10	100	2	4.8

- From the main window of STAR click **Analyze | Analysis of Variance | Randomized Complete Block Design (RCBD)**. The **Randomized Complete Block Design (RCBD)** dialog box will appear.
- Opening the data for the first time, *SeedingRate* and *Rep* fields in the data file are regarded by R as numerical variables; they need to be changed as factors. Choose the variable and click on the **To Factor** button.
- Specify the required field and appropriate options for the analysis.

Model Specification Tab

Response Variable(s)

This field is required. For the analysis to proceed, at least one item should be entered in the list box. Only Items from the **Numeric Variable(s)** can be added in this dialog box.

Factor(s)

This field is required. At least one item can be added in this list box and should come from the **Factor(s)** list box.

Block

This field is required. Only one item can be added in this field and should come from the **Factor(s)** list box.

Options Tab

Display Descriptive Statistics

If this option is selected, a summary table with number of non-missing observations, minimum, maximum, mean and standard deviation of the response variable will be displayed.

Shapiro-Wilk Test

If this option is selected, test for normality of residuals using Shapiro-Wilk will be displayed.

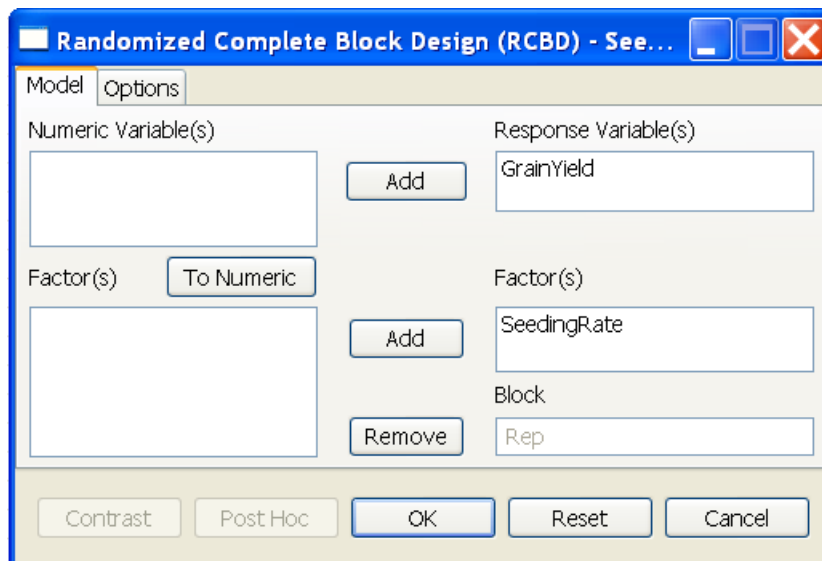
Bartlett's Test

If this option is selected, test for homogeneity of variances using Bartlett's Test will be displayed.

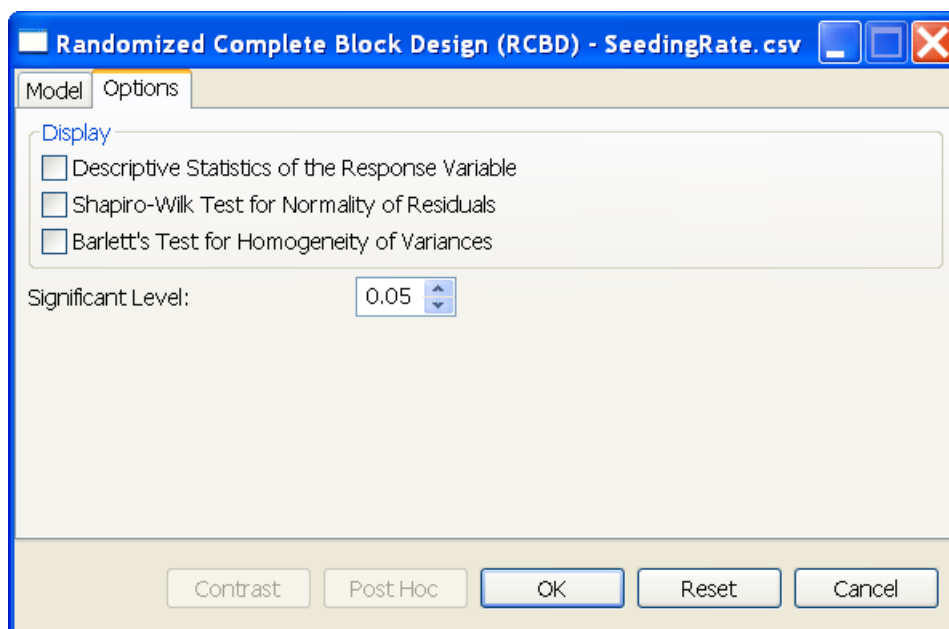
Significance Level

This will be used to determine if the treatment effects and interaction effects are significant in Analysis of Variance. This is also the level of significance to be used when pairwise mean comparison is performed. Its default value is 0.05. The user can change this value by specifying a numeric value from zero to 1.

For the example, the completed **Model** Specification tab should appear as illustrated below:



The completed **Options** tab should appear as illustrated below:



- Click **OK** button to perform the analysis. The **Randomized Complete Block Design (RCBD)** dialog box will be minimized and STAR activates the **Output Viewer** tab and the **Graph Viewer** tab.

Output Viewer

This tab contains the text output which begins with the descriptive statistics and the output for the tests for homogeneity of variances and test for normality, whichever was specified in the **Options** tab. The next part is the Analysis of Variance table. If there is only one factor considered in the analysis, and if the factor is not significant a table of treatment means will be displayed. If there are 2 or more factors in the analysis and if

the highest interaction is not significant, table of means for all levels of the treatment combinations will be displayed. Then, depending on the result of the analysis of variance, pairwise comparison of means will be automatically generated using one applicable pairwise comparison procedure. If at most 5 treatment levels are to be compared, the default pairwise comparison procedure used in STAR is the Least Significant Difference (LSD) test; otherwise, the Honestly Significant Difference (HSD) or Tukey's test is used.

Graph Viewer

The tab displays two diagnostics plots. The left plot, entitled *Residuals vs. Fitted* is for assessing the distribution of the residuals. An ideal plot should reveal that the points are distributed around 0. The residuals should not be affected by the size of the fitted values. Problematic data may result to a plot wherein the points increase as the fitted values increase, thus forming a “funnel-shaped” distribution. This will indicate a violation of the homogeneity of variances requirement for analysis of variance. The *Normal Q-Q* plot is used for assessing normality. The points should lie about the line.

- Sample text output of the analysis displayed in the Output Viewer tab is shown below:

```
Analysis of Variance
Randomized Complete Block Design (RCBD)
```

CLASS INFORMATION

Class Level Information

```
-----
FACTOR          NO. OF LEVELS LEVELS
-----
SeedingRate      6 25, 50, ..., 150
Rep              4 1, 2, 3, 4
-----
```

Number of Observations Read and Used: 24

Bartlett's Test for Homogeneity of Variances

```
-----
Method    DF    Test Stat    p Value
-----
bartlett    5          4.76    0.4462
-----
```

TEST FOR NORMALITY

```
-----
Variable          Method          Stat    Value    p Value
-----
GrainYield_resid  Shapiro-Wilk  W          0.9346    0.1235
-----
```

ANOVA TABLE

Response Variable: GrainYield

```
-----
Source          DF    Sum of Square    Mean Square    F Value    Pr > F
-----
Rep              3          1.7667          0.5889          4.87    0.0147
-----
```


SeedingRate	5	10.1800	2.0360	16.84	0.0000
Error	15	1.8133	0.1209		
Total	23	13.7600			

Summary Statistics

Coef Var	GrainYield Mean
7.48	4.65

Standard Errors

Effects	StdErr
Rep	0.2007
SeedingRate	0.2459

Pairwise Mean Comparison of SeedingRate

Tukeys's Honest Significant Difference (HSD) Test

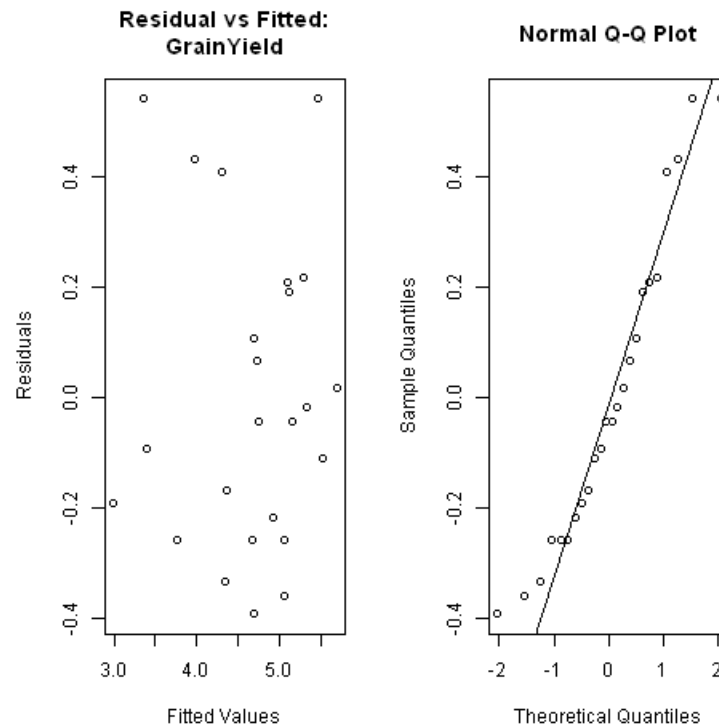
Alpha	0.05
Error Degrees of Freedom	15
Error Mean Square	0.1209
Critical Value	4.5947
Test Statistics	0.7988

Summary of the Result:

SeedingRate	means	N	group
100	4.35	4	b
125	4.67	4	ab
150	3.38	4	c
25	5.12	4	ab
50	5.08	4	ab
75	5.30	4	a

* Means with the same letter are not significantly different.

- Sample graphical output of the analysis displayed in the Graph Viewer tab is shown below:



Additional Options

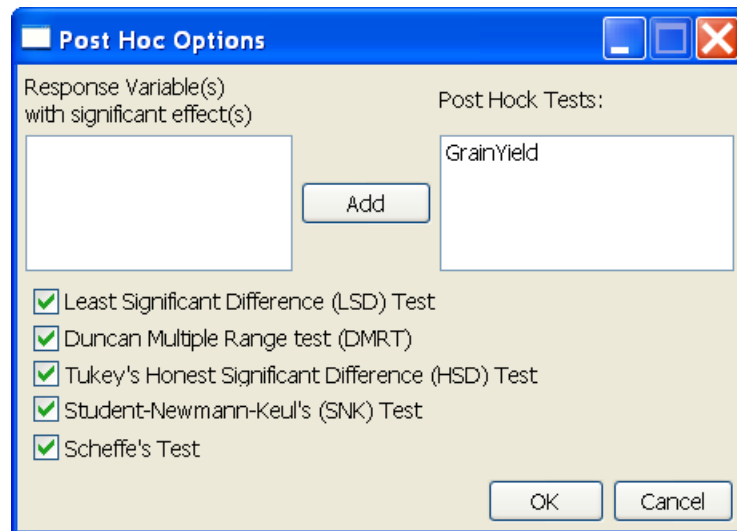
If there are still modifications to be made, the following option buttons may be used:

- **Post Hoc**

Use this option if other pairwise comparison procedures, other than what was presented in the **Output Viewer** tab page, are desired. There are five pairwise mean comparison procedure available, namely: Least Significant Difference Test, Duncan Multiple Range Test, Tukey's Honest Significant Test, Student Newmann-Keul's Test and the Scheffe's Test. This button will be disabled if *F*-tests conducted using ANOVA yielded results that are not significant.

Suppose additional pairwise mean comparison is requested.

- Click the **Post Hoc Option** button. The **Post Hoc Option** dialog box will appear.
- Specify the at least one response variable where pairwise mean comparison will be perform and choose at least one pairwise mean comparison procedure.
- The completed **Post Hoc Option** dialog box should appear as shown below:



- Click the **OK** button to perform the pairwise mean comparison. The **Post Hoc Option** dialog box will be closed and STAR activates the **Output Viewer** tab.
- Sample result of the pairwise mean comparison requested is shown below:

Result of Pairwise Comparison

Response Variable: GrainYield

Pairwise Mean Comparison of SeedingRate

Alpha 0.05
 Error Degrees of Freedom 15
 Error Mean Square 0.1209

	LSD	HSD	scheffe
Critical Value	2.1314	4.5947	2.9013
Test Statistic	0.5240	0.7988	0.9364

Summary:

SeedingRate	means	N	std.err	LSD	HSD	scheffe
100	4.35	4	0.1708	c	b	b
125	4.67	4	0.0946	bc	ab	ab
150	3.38	4	0.2287	d	c	c
25	5.12	4	0.1548	ab	ab	a
50	5.08	4	0.3705	ab	ab	a
75	5.30	4	0.2160	a	a	a

* Means with the same letter are not significantly different

Duncan's Multiple Range Test (DMRT)

Alpha 0.05
 Error Degrees of Freedom 15

Error Mean Square	0.1209
-------------------	--------

Number of Means	2	3	4	5	6
Tabular Value	3.01	3.16	3.25	3.31	3.36
Test Statistics	0.52	0.55	0.57	0.58	0.58

Summary of the Result:

SeedingRate	means	N	group
100	4.35	4	c
125	4.67	4	bc
150	3.38	4	d
25	5.12	4	ab
50	5.08	4	ab
75	5.30	4	a

* Means with the same letter are not significantly different.

Student Newman Keuls (SNK) Test

Alpha	0.05
Error Degrees of Freedom	15
Error Mean Square	0.1209

Number of Means	2	3	4	5	6
Critical Value	3.0143	3.6734	4.0760	4.3670	4.5947
Test Statistics	0.5240	0.6386	0.7086	0.7592	0.7988

Summary of the Result:

SeedingRate	means	N	group
100	4.35	4	b
125	4.67	4	ab
150	3.38	4	c
25	5.12	4	a
50	5.08	4	a
75	5.30	4	a

* Means with the same letter are not significantly different.

- **Contrast**

This button will be enabled if there is at least one factor that has more than two levels. Use of this button will be discussed in Partitioning Sum of Squares section of this module.

8.2. Two Factor Randomized Complete Block

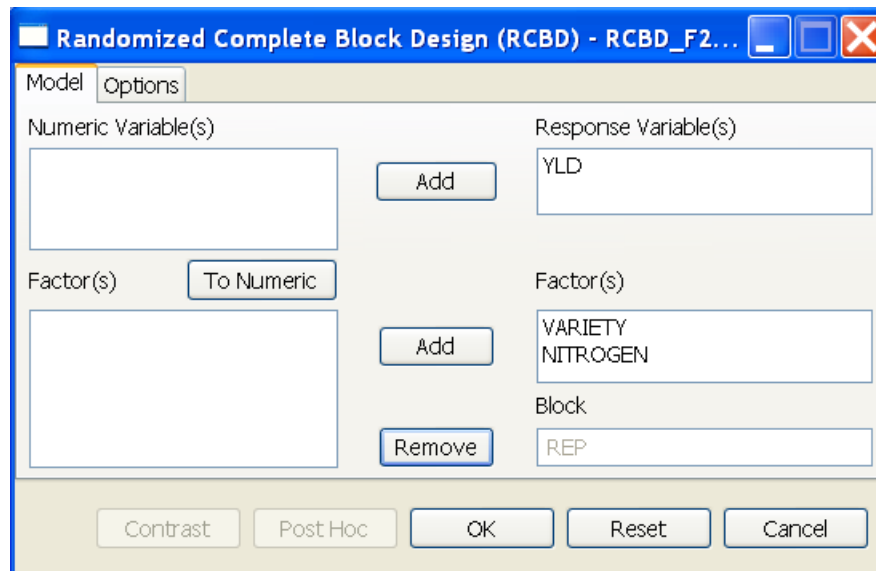
The steps to perform the Analysis of Variance in Randomized Complete Block Design are listed below:

- On the Project Explorer the dataset *RCBD_F2.csv* located in the *Data* folder of the project named *SampleProject*. Double-click the file to open and view it in the Data Viewer. The example file contains data from an experiment conducted to compare the effects of three rice variety (*VARIETY*, with levels *V1*, *V2* and *V3*) and five nitrogen levels (*NITROGEN*, with levels *N0*, *N1*, *N2*, *N3* and *N4*) on grain yield (*YLD*, in t/ha) using Randomized Complete Block (RCB) field design with four blocks (*REP* with levels 1, 2, 3 and 4).

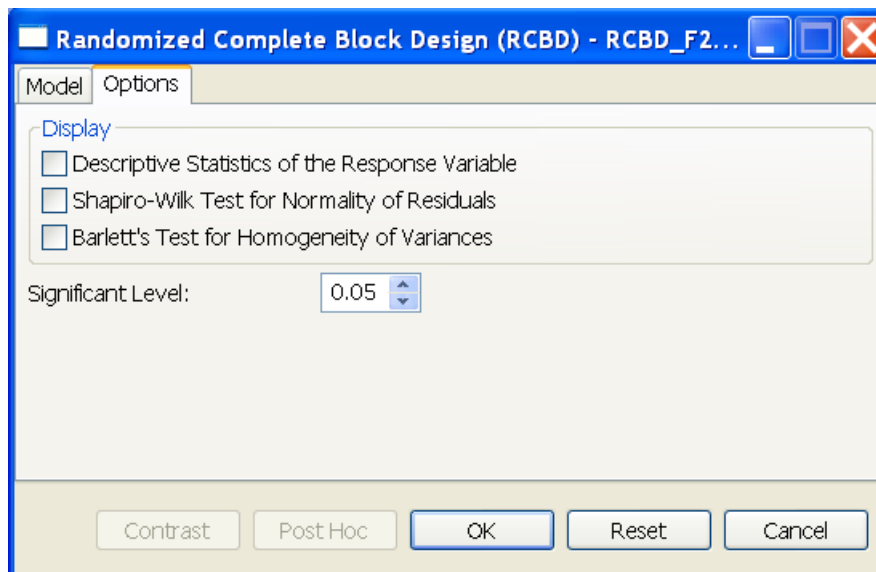
	REP	VARIETY	NITROGEN	YLD
1	1	V1	N0	3.582
2	1	V1	N1	4.788
3	1	V1	N2	4.576
4	1	V1	N3	6.034
5	1	V1	N4	5.874
6	1	V2	N0	2.846
7	1	V2	N1	4.956
8	1	V2	N2	5.928
9	1	V2	N3	5.644
10	1	V2	N4	5.458
11	1	V3	N0	4.192
12	1	V3	N1	5.25
13	1	V3	N2	5.822
14	1	V3	N3	5.888
15	1	V3	N4	5.864

- From the main window of STAR click **Analyze | Analysis of Variance | Randomized Complete Block Design (RCBD)**. The **Randomized Complete Block Design (RCBD)** dialog box will appear.
- Opening the data for the first time, *REP* field in the data file is regarded by R as numerical variable; it need to be changed as factors. Choose the variable and click on the **To Factor** button.
- Specify the required field and appropriate options for the analysis.

For the example, the completed **Model** Specification tab should appear as illustrated below:



The completed **Options** tab should appear as illustrated below:



- Click **OK** button to perform the analysis. The **Randomized Complete Block Design (RCBD)** dialog box will be minimized and STAR activates the **Output Viewer** tab and the **Graph Viewer** tab.
- Sample text output of the analysis displayed in the Output Viewer tab is shown below:

```
Analysis of Variance
Randomized Complete Block Design (RCBD)
```

```
CLASS INFORMATION
Class Level Information
```

```
-----
FACTOR      NO. OF LEVELS  LEVELS
-----
```

```
REP                4 1, 2, 3, 4
VARIETY            3 V1, V2, V3
NITROGEN           5 N0, N1, N2, N3, N4
-----
```

Number of Observations Read and Used: 60

ANOVA TABLE

Response Variable: YLD

```
-----
Source              DF    Sum of Square    Mean Square    F Value    Pr > F
-----
REP                  3          2.5018          0.8339         5.73    0.0022
VARIETY              2          1.1527          0.5763         3.96    0.0266
NITROGEN             4         42.0064         10.5016        72.12    0.0000
VARIETY:NITROGEN     8          2.3943          0.2993         2.06    0.0627
Error               42          6.1157          0.1456
Total               59         54.1709
-----
```

Summary Statistics

```
-----
Coef Var    YLD Mean
-----
      7.71      4.95
-----
```

Standard Errors

```
-----
Effects              StdErr
-----
REP                  0.1393
VARIETY              0.1207
NITROGEN             0.1558
VARIETY:NITROGEN     0.2698
-----
```

Table of Means

```

      NITROGEN      N0      N1      N2      N3      N4
VARIETY
V1          3.0565  4.7235  4.4595  5.7170  5.8230
V2          3.5480  4.8060  5.4360  5.7345  5.6805
V3          3.7780  4.7535  5.3200  5.5525  5.8865
```

Pairwise Mean Comparison of VARIETY

Least Significant Difference (LSD) Test

```

Alpha                      0.05
Error Degrees of Freedom    42
Error Mean Square           0.1456
Critical Value               2.0181
Test Statistics              0.2435
```

Summary of the Result:

```
-----  
VARIETY      means      N group  
-----  
V1           4.76      20   b  
V2           5.04      20   a  
V3           5.06      20   a  
-----
```

* Means with the same letter are not significantly different.

Pairwise Mean Comparison of NITROGEN

Least Significant Difference (LSD) Test

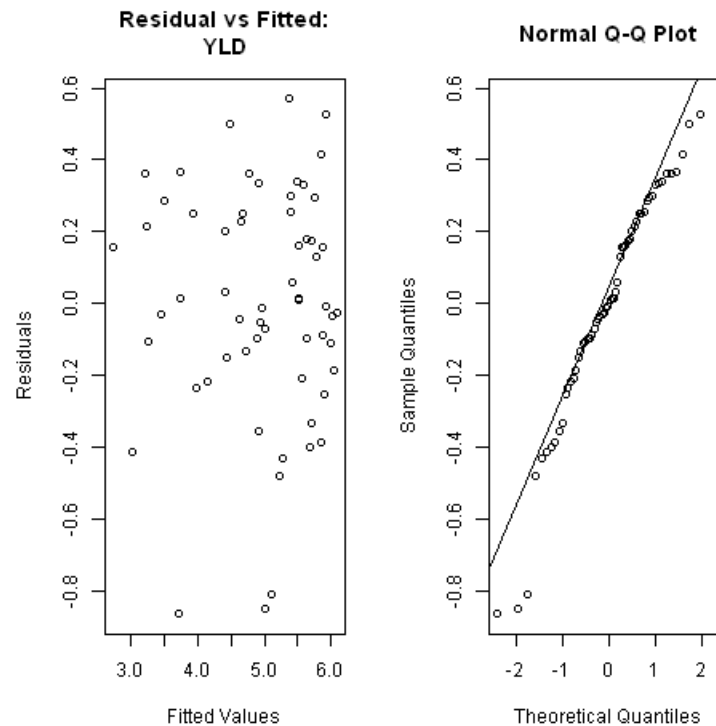
```
Alpha                      0.05  
Error Degrees of Freedom    42  
Error Mean Square           0.1456  
Critical Value              2.0181  
Test Statistics              0.3144
```

Summary of the Result:

```
-----  
NITROGEN      means      N group  
-----  
N0            3.46      12   c  
N1            4.76      12   b  
N2            5.07      12   b  
N3            5.67      12   a  
N4            5.80      12   a  
-----
```

* Means with the same letter are not significantly different.

- Sample graphical output of the analysis displayed in the Graph Viewer tab is shown below:



8.3. Split Plot Design in Randomized Complete Block

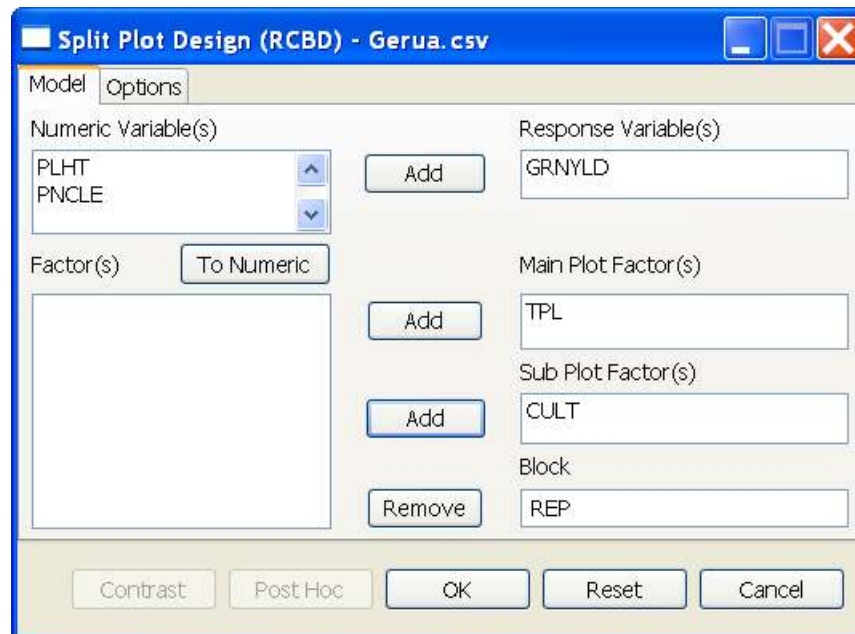
The steps to perform the Analysis of Variance using Split Plot Design in Randomized Complete Block are listed below:

- On the Project Explorer the dataset *Gerua.csv* located in the *Data* folder of the project named *SampleProject*. Double-click the file to open and view it in the Data Viewer. The example file contains data from an experiment conducted using Split Plot Design with transplanting method (*TPL*) as the main plot factor and cultivar (*CULT*) as the subplot factor in RCB. The experiment was done to compare the effects of two transplanting method (with levels *T* and *N*) using 25 cultivars with three blocks (*REP* with levels 1, 2, and 3) on grain yield (*GRNYLD*, in kg/ha) as well as other variables: plant height (*PLHT*, in cm) and panicle number (*PNCLE*).

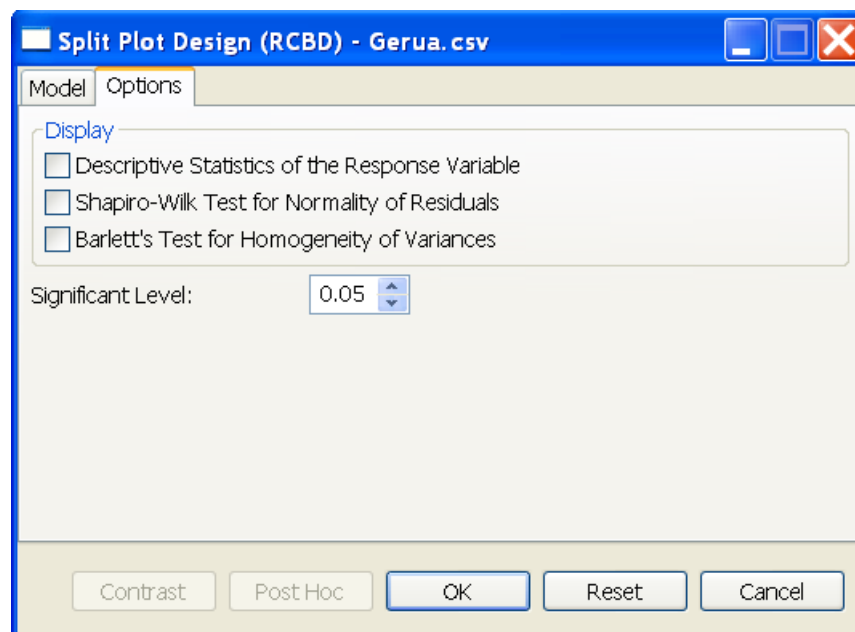
	REP	TPL	CULT	OFF	PLHT	PNQLE	GRN/LD
1	1	D	CN1	136	115.8	239	1680
2	2	D	CN1	133	120.5	231	1350
3	3	D	CN1	135	117.8	234	1900
4	2	D	NDR1	129	100.5	200	4900
5	3	D	NDR1	128	98.6	195	5500
6	2	D	CR1	128	122.9	236	2800
7	1	D	CN2	139	125.6	240	2200
8	2	D	CN2	138	128.4	236	1450
9	3	D	CN2	139	123.2	237	2250
10	2	D	NDR2	130	113.9	180	2050
11	3	D	NDR2	130	115.4	152	2450
12	2	D	CR2	130	124.6	271	3150
13	1	D	CR1	137	120.5	31	3500
14	2	D	CR1	140	121.8	299	2700
15	3	D	CR1	138	119.2	307	3200
16	2	D	NDR3	116	120.8	210	5750
17	3	D	NDR3	116	128.4	204	5200
18	2	D	CR3	131	115.6	212	2500
19	1	D	CR2	124	120.7	193	2100
20	2	D	CR2	125	124.3	190	3100
21	3	D	CR2	123	124.2	196	2750
22	2	D	NDR4	141	122.4	232	3860
23	3	D	NDR4	141	120.9	228	3000
24	2	D	CR4	141	113.7	213	2150
25	1	D	IR1-CN3	133	92.6	231	2850
26	2	D	IR1-CN3	136	95.2	224	2750
27	3	D	IR1-CN3	134	93.1	229	2250
28	2	D	NDR5	149	135.9	213	1800
29	3	D	NDR5	145	133.8	207	2100
30	2	D	CR5	134	120.4	205	2050
31	1	D	IR1-CN4	136	140.4	233	1000
32	2	D	IR1-CN4	135	147.2	220	1260
33	3	D	IR1-CN4	136	143.5	228	950
34	2	D	NDR6	142	109.3	213	2360
35	3	D	NDR6	144	107.9	209	3850
36	2	D	CR6	135	112.6	250	3950

- From the main window of STAR click **Analyze | Analysis of Variance | Split Plot Design | Randomized Complete Block Design (RCBD)**. The **Split Plot Design (RCBD)** dialog box will appear.
- Opening the data for the first time, *REP* field in the data file is regarded by R as numerical variable; it need to be changed as factors. Choose the variable and click on the **To Factor** button.
- Specify the required field and appropriate options for the analysis.

For the example, the completed **Model** Specification tab should appear as illustrated below:



The completed **Options** tab should appear as illustrated below:



- Click **OK** button to perform the analysis. The **Split Plot Design (RCBD)** dialog box will be minimized and STAR activates the **Output Viewer** tab and the **Graph Viewer** tab.
- Sample text output of the analysis (only results from *response variable = GRNYLD*) displayed in the Output Viewer tab is shown below:

```
Analysis of Variance
Split Plot Design in RCBD
```

CLASS INFORMATION

Class Level Information

FACTOR	NO. OF LEVELS	LEVELS
REP	3	1, 2, 3
TPL	2	D, N
CULT	26	CN1, CN2, ..., S2

Number of Observations Read and Used: 156

ANOVA TABLE

Response Variable: GRNYLD

Source	DF	Sum of Square	Mean Square	F Value	Pr > F
REP	2	58473.3974	29236.6987	0.35	0.7425
TPL	1	109947636.0577	109947636.0577	1303.92	0.0008
Error(a)	2	168641.3462	84320.6731		
CULT	25	134144970.6731	5365798.8269	51.02	0.0000
TPL:CULT	25	16611234.7756	664449.3910	6.32	0.0000
Error(b)	100	10517968.5897	105179.6859		
Total	155	271448924.8397			

Summary Statistics

Coef Var(a)	Coef Var(b)	GRNYLD Mean
7.72	8.63	3759.20

Comparison of TPL at each level of CULT

Least Significant Difference (LSD) Test

Alpha	0.05
Error Degrees of Freedom	101
Error Mean Square	104377.4162
Critical Value	1.9837
Test Statistic	523.2683

Summary:

TPL	N	CULT = CN1 group	CULT = CN2 group	...	CULT = S2 group
D	3	1643.33 b	1966.67 b	...	2163.33 b
N	3	3683.33 a	3763.33 a	...	4260.00 a

* Means with the same letter are not significantly different

Comparison of CULT at each level of TPL

Tukeys's Honest Significant Difference (HSD) Test

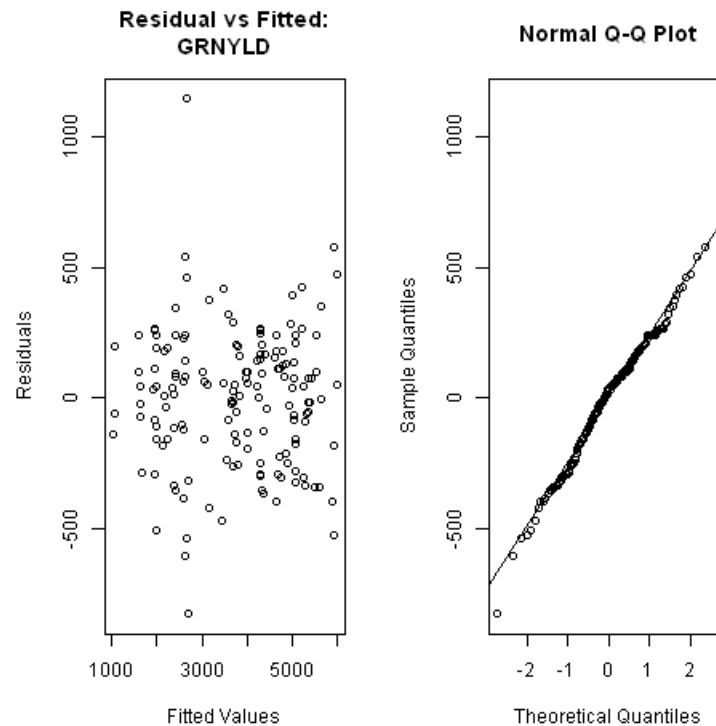
Alpha 0.05
 Error Degrees of Freedom 100
 Error Mean Square 105179.6859
 Critical Value 5.3519
 Test Statistic 1002.1128

Summary:

CULT	N	TPL = D group		TPL = N group	
CN1	3	1643.33	jk	3683.33	hij
CN2	3	1966.67	ijk	3763.33	ghij
CR1	3	3133.33	efgh	5033.33	abcde
CR2	3	2650.00	ghi	4270.00	efghi
IR1-CN3	3	2616.67	ghij	5050.00	abcde
IR1-CN4	3	1070.00	k	3026.67	j
IR2	3	1950.00	ijk	4260.00	efghi
NDR1	3	5216.67	ab	5890.00	ab
NDR2	3	2416.67	hij	4926.67	bcdef
NDR3	3	5520.00	a	5956.67	a
NDR4	3	3453.33	defg	4690.00	cdefg
NDR5	3	1966.67	ijk	3986.67	fghij
NDR6	3	2686.67	fghi	4830.00	cdef
NDR7	3	4783.33	abc	5333.33	abcd
OR1	3	2583.33	ghij	5250.00	abcde
OR2	3	2616.67	ghij	4326.67	efghi
OR3	3	2416.67	hij	4400.00	defghi
OR4	3	1966.67	ijk	4000.00	fghij
OR5	3	2216.67	hij	4653.33	cdefgh
OR6	3	3800.00	cde	5033.33	abcde
OR7	3	5066.67	ab	5616.67	abc
OR8	3	4316.67	bcd	5386.67	abcd
RA1	3	1633.33	jk	3763.33	ghij
RA2	3	3673.33	def	4616.67	cdefgh
S1	3	2385.00	hij	3560.00	ij
S2	3	2163.33	hij	4260.00	efghi

* Means with the same letter are not significantly different

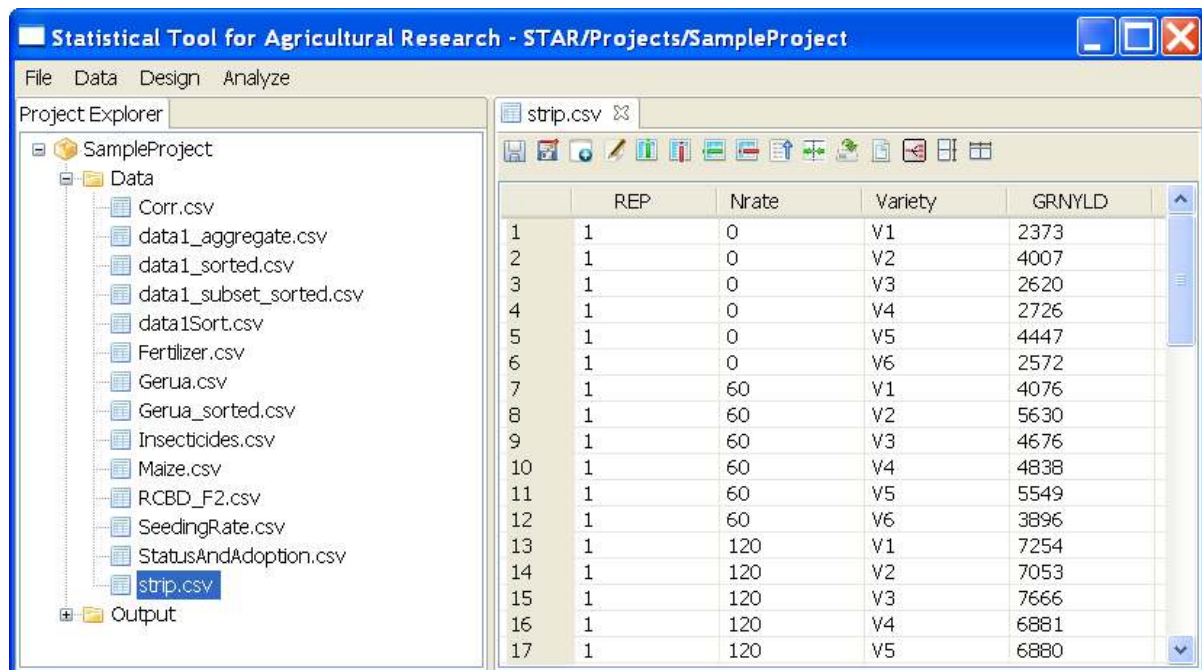
- Sample graphical output of the analysis displayed in the Graph Viewer tab is shown below:



8.4. Strip Plot Design

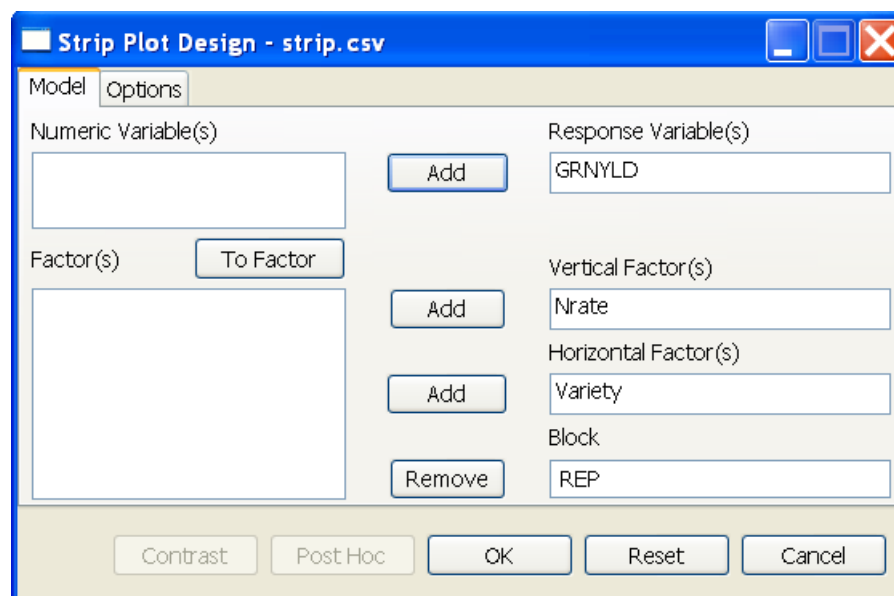
The steps to perform the Analysis of Variance in Randomized Complete Block Design are listed below:

- On the Project Explorer the dataset *strip.csv* located in the *Data* folder of the project named *SampleProject*. Double-click the file to open and view it in the Data Viewer. The example file contains data from an experiment conducted to compare the effects of six rice variety (*Variety*, with levels *V1*, *V2*, *V3*, *V4*, *V5* and *V6*) and three rates of nitrogen (*Nrate*, in kg/ha with levels *0*, *60* and *120*) on grain yield (*GRNYLD*, in kg/ha). A strip plot design with four blocks (*REP* with levels *1*, *2*, and *3*) was used with rice variety as the horizontal factor and nitrogen rates as the vertical factor.

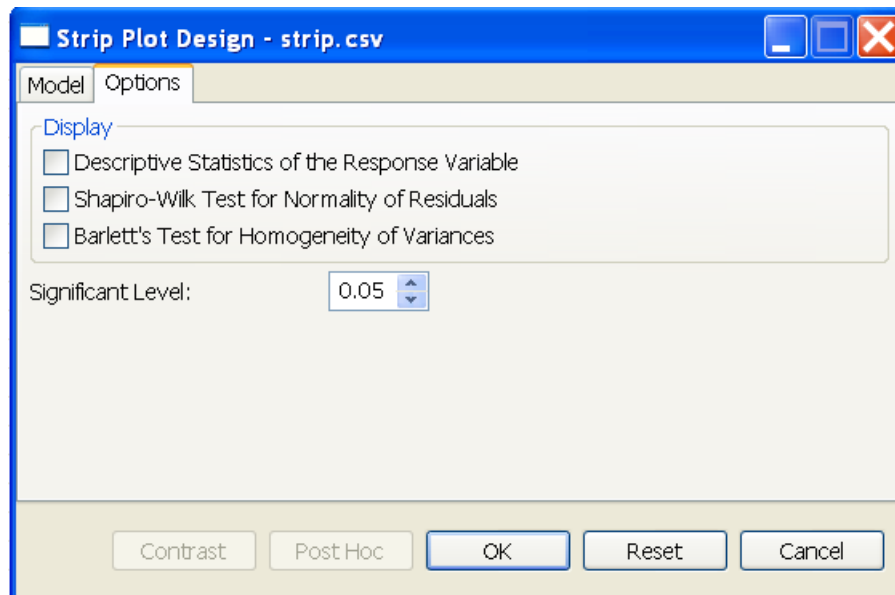


- From the main window of STAR click **Analyze | Analysis of Variance | Strip Plot Design**. The **Strip Plot** dialog box will appear.
- Opening the data for the first time, *REP* and *Nrate* fields in the data file are regarded by R as numerical variables; they need to be changed as factors. Choose the variable and click on the **To Factor** button.
- Specify the required field and appropriate options for the analysis.

For the example, the completed **Model** Specification tab should appear as illustrated below:



The completed **Options** tab should appear as illustrated below:



- Click **OK** button to perform the analysis. The **Strip Plot Design** dialog box will be minimized and STAR activates the **Output Viewer** tab and the **Graph Viewer** tab.
- Sample text output of the analysis displayed in the Output Viewer tab is shown below:

```
Analysis of Variance
Strip Plot Design
```

```
CLASS INFORMATION
Class Level Information
```

```
-----
FACTOR      NO. OF LEVELS  LEVELS
-----
REP          3  1, 2, 3
Nrate        3  0, 60, 120
Variety      6  V1, V2, ..., V6
-----
```

```
Number of Observations Read and Used: 54
```

```
ANOVA TABLE
Response Variable: GRNYLD
```

```
-----
Source          DF    Sum of Square    Mean Square    F Value    Pr > F
-----
REP              2      9215077.1481      4607538.5741      6.19    0.0596
Nrate            2      50672299.5926      25336149.7963     34.05    0.0031
Error(a)         4       2975911.9630       743977.9907
Variety          5      57101029.4259      11420205.8852      7.65    0.0034
Error(b)        10      14922379.9630       1492237.9963
Nrate:Variety    10      23874076.8519       2387407.6852      5.80    0.0004
-----
```


Statistical Tools for Agricultural Research (STAR) Tutorial

Error(c)	20	8230550.9259	411527.5463
Total	53	166991325.8704	

Summary Statistics

Coef Var(a)	Coef Var(b)	Coef Var(c)	GRNYLD Mean
16.31	23.09	12.13	5289.76

Comparison of Nrate at each level of Variety

Least Significant Difference (LSD) Test

Alpha	0.05
Error Degrees of Freedom	22
Error Mean Square	466935.9537
Critical Value	2.0716
Test Statistic	1155.8252

Summary:

Nrate	N	Variety = V1 group	Variety = V2 group	...	Variety = V6 group
0	3	3571.67 c	4934.33 b	...	3207.33 ab
120	3	7548.00 a	7211.33 a	...	2492.00 b
60	3	5132.00 b	6713.67 a	...	3714.33 a

* Means with the same letter are not significantly different

Comparison of Variety at each level of Nrate

Tukeys's Honest Significant Difference (HSD) Test

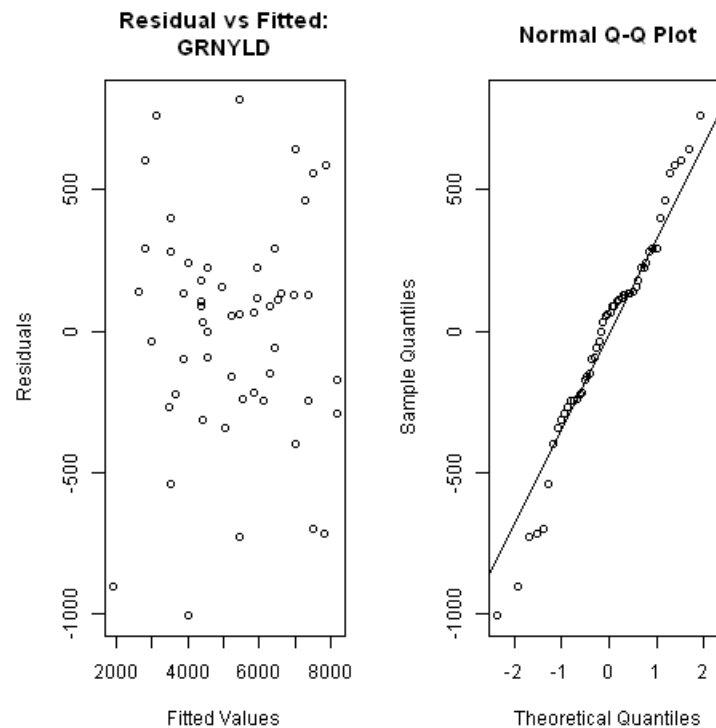
Alpha	0.05
Error Degrees of Freedom	10
Error Mean Square	1492237.9963
Critical Value	4.9120
Test Statistic	3464.3214

Summary:

Variety	N	Nrate = 0 group	Nrate = 60 group	Nrate = 120 group
V1	3	3571.67 a	5132.00 a	7548.00 a
V2	3	4934.33 a	6713.67 a	7211.33 a
V3	3	4249.67 a	6122.33 a	7868.33 a
V4	3	4059.00 a	5553.67 a	7094.33 a
V5	3	4101.67 a	5630.00 a	6012.00 a
V6	3	3207.33 a	3714.33 a	2492.00 b

* Means with the same letter are not significantly different

- Sample graphical output of the analysis displayed in the Graph Viewer tab is shown below:



8.5. Partitioning Sum of Squares

The **Contrast** option is used in partitioning sum of squares that decompose the variability of the response(s) into various components to have a better view of the nature of variability. There are different selections in partitioning sum of squares such as, “Compare with Control”, “User Specified Contrast (Group Comparison)” and “Orthogonal Polynomial Contrast”.

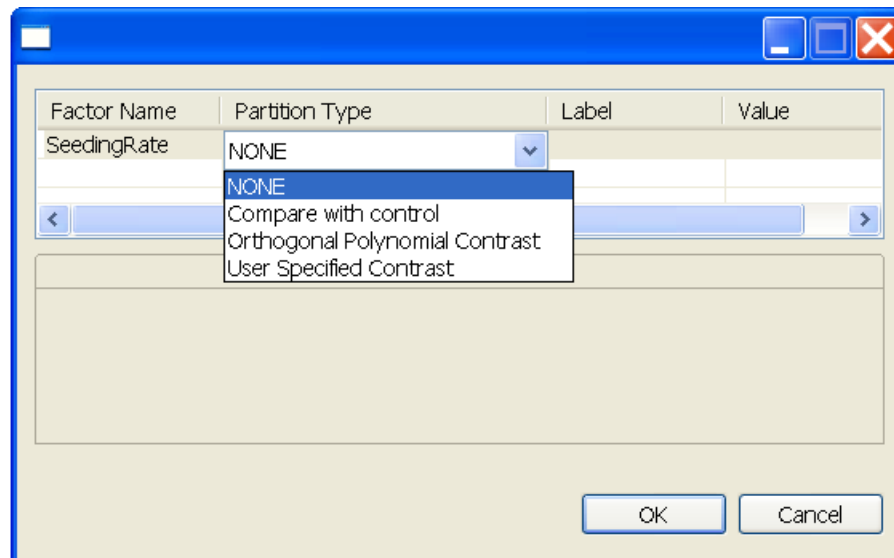
8.5.1. Group Comparison

The steps to perform the Group Comparison are listed below:

- On the Project Explorer the dataset *SeedingRate.csv* located in the *Data* folder of the project named *SampleProject*. Double-click the file to open and view it in the Data Viewer.
- Perform the Analysis of Variance. Follow the discussion in **Randomized Complete Block Design (RCBD)** of this module.
- After performing the Analysis of Variance, the **Contrast** button will be enabled if there is at least one factor in the analysis with more than two levels.

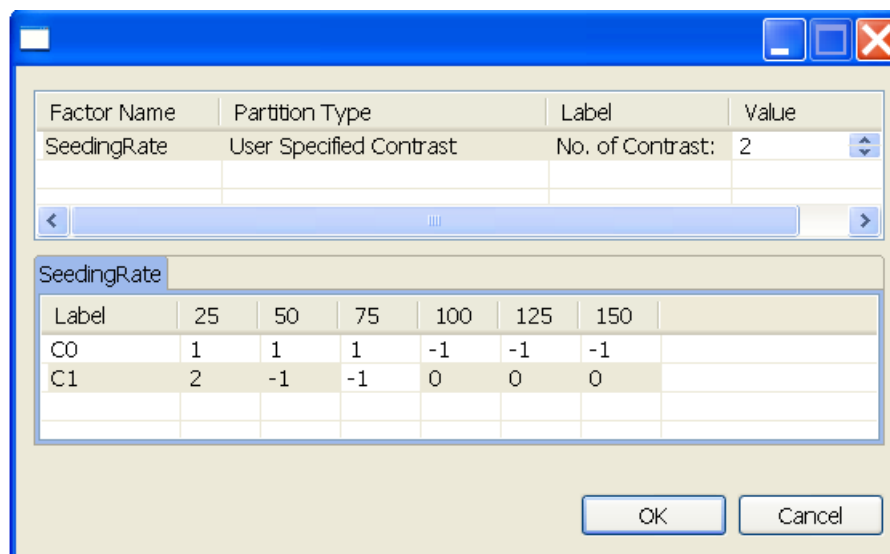
Suppose we want to determine the following:

- i. If the mean of the *SeedingRate* 25, 50, and 75 is significantly different from *SeedingRate* 100, 125 and 150
 - ii. If the mean of the *SeedingRate* 25, is significantly different from *SeedingRate* 50, and 75
- Click the **Contrast** button. A dialog box will appear.



- Specify the required fields and appropriate options for the analysis. In the **Partition Type** column, click the drop-down menu and choose *User Specified Contrast*. Specify the number of contrast. A tab will be displayed in the lower part of the Contrast Analysis dialog box.

For the example, the completed dialog box should appear as illustrated below:



- Click the **OK** button to perform Group Comparison. The Contrast Analysis dialog box will be closed and STAR activates the Output Viewer tab.
- Sample output of the analysis:

8.5.2. Trend Comparison

This procedure is applicable and available if the levels of the factor are quantitative.

The steps to perform the Group Comparison are listed below:

- On the Project Explorer the dataset *SeedingRate.csv* located in the *Data* folder of the project named *SampleProject*. Double-click the file to open and view it in the Data Viewer.
- Perform the Analysis of Variance. Follow the discussion in **Randomized Complete Block Design (RCBD)** of this module.
- After performing the Analysis of Variance, the **Contrast** button will be enabled if there is at least one factor in the analysis with more than two levels.

Suppose we also want to know the degree of the relationship between yield response and Seeding Rate.

- Click the **Contrast** button. A dialog box will appear.
- Specify the required fields and appropriate options for the analysis. In the **Partition Type** column, click the drop-down menu and choose *Orthogonal Polynomial Contrast*. Specify the highest degree polynomial that will be fitted.

For the example, the completed dialog box should appear as illustrated below:

Factor Name	Partition Type	Label	Value
SeedingRate	Orthogonal Polynomial Contrast	Degree:	3

- Click the **OK** button to perform Trend Comparison. The Contrast Analysis dialog box will be closed and STAR activates the Output Viewer tab.
- Sample output of the analysis:

Result of Contrast Analysis

ANOVA TABLE

Response Variable: GrainYield

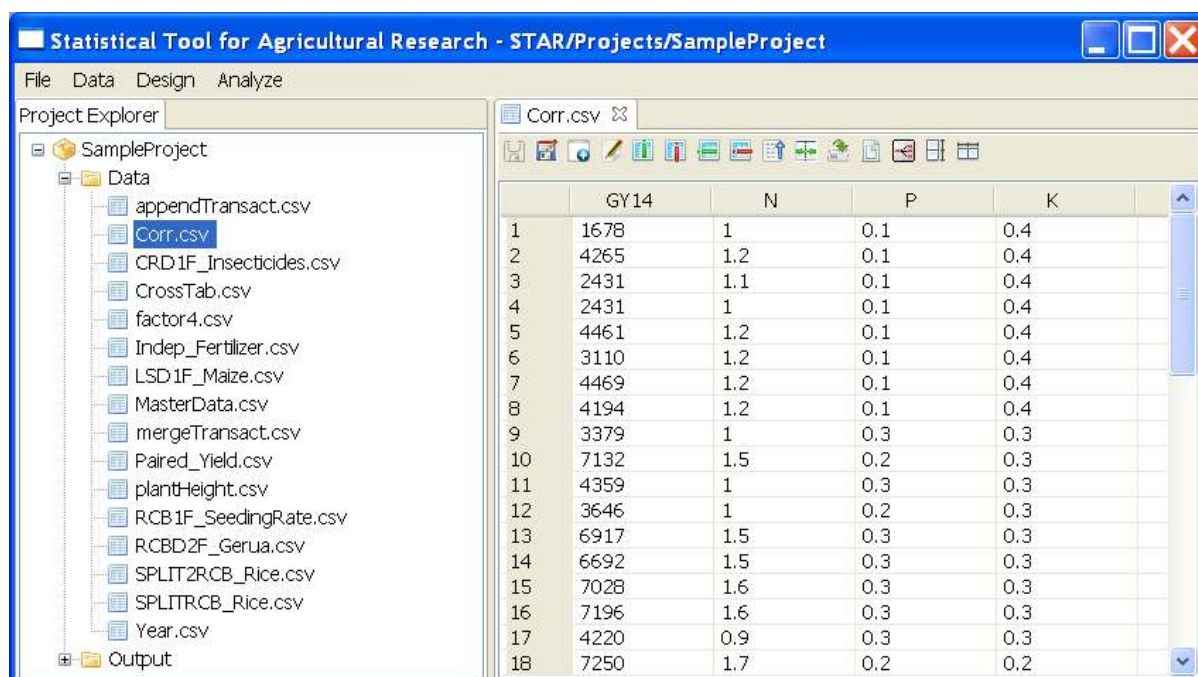
Source	DF	Sum of Square	Mean Square	F Value	Pr > F
Rep	3	1.7667	0.5889	4.87	0.0147
SeedingRate	5	10.1800	2.0360	16.84	0.0000
SeedingRate: 1	1	6.7891	6.7891	56.16	0.0000
SeedingRate: 2	1	1.6296	1.6296	13.48	0.0023
SeedingRate: 3	1	0.1027	0.1027	0.85	0.3712
Error	15	1.8133	0.1209		
Total	23	13.7600			

9. Correlation Analysis

The **Correlation Analysis** is used to determine the strength of linear relationship between variables. The **Correlation Analysis** submenu performs bivariate correlation using the pearson's correlation coefficient, spearman's rho and kendall tau. By default, it uses the pearson's correlation coefficient.

The steps to perform Correlation Analysis are listed below:

- On the Project Explorer, locate the dataset *Corr.csv* in the *Data* folder of the project named *SampleProject*. Double-click the file to open and view it in the Data Viewer tab. The example file contains data from an experiment conducted to investigate the relationship between grain yield (GY14, in kg/ha), the grain's nitrogen (N), phosphorus (P) and potassium (K) content in percent. The data were collected from 48 experimental plots with grains adjusted to 14% moisture content.



	GY14	N	P	K
1	1678	1	0.1	0.4
2	4265	1.2	0.1	0.4
3	2431	1.1	0.1	0.4
4	2431	1	0.1	0.4
5	4461	1.2	0.1	0.4
6	3110	1.2	0.1	0.4
7	4469	1.2	0.1	0.4
8	4194	1.2	0.1	0.4
9	3379	1	0.3	0.3
10	7132	1.5	0.2	0.3
11	4359	1	0.3	0.3
12	3646	1	0.2	0.3
13	6917	1.5	0.3	0.3
14	6692	1.5	0.3	0.3
15	7028	1.6	0.3	0.3
16	7196	1.6	0.3	0.3
17	4220	0.9	0.3	0.3
18	7250	1.7	0.2	0.2

- Click **Analyze | Correlation Analysis** from the main window of STAR. The **Correlation Analysis** dialog box will appear.
- Specify the required field and appropriate options for the analysis:

Variable Description Tab

Test Variable(s)

This field is required. For the analysis to proceed, this list box should have at least two entry.

Correlation Coefficient

For the analysis to proceed, at least one of the methods should be tick. By default, Pearson's correlation coefficient is computed. For quantitative and normally distributed variables, choose the Pearson correlation coefficient. If the data are not normally distributed or ordinal level, choose Kendall's tau or Spearman, which measures the association between ranks. The correlation coefficient ranges from -1 to +1.

Option Tab

Alternative Hypothesis

User can select a one-tailed or two-tailed test. The default alternative hypothesis used is that for the two-sided test.

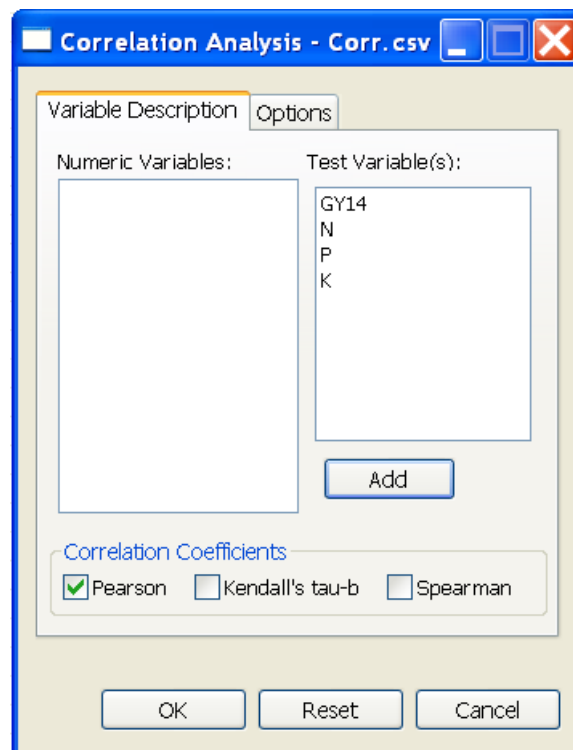
Display Summary Statistics

If this option is selected, a summary table with number of observations, minimum, maximum, mean and standard deviation will be displayed.

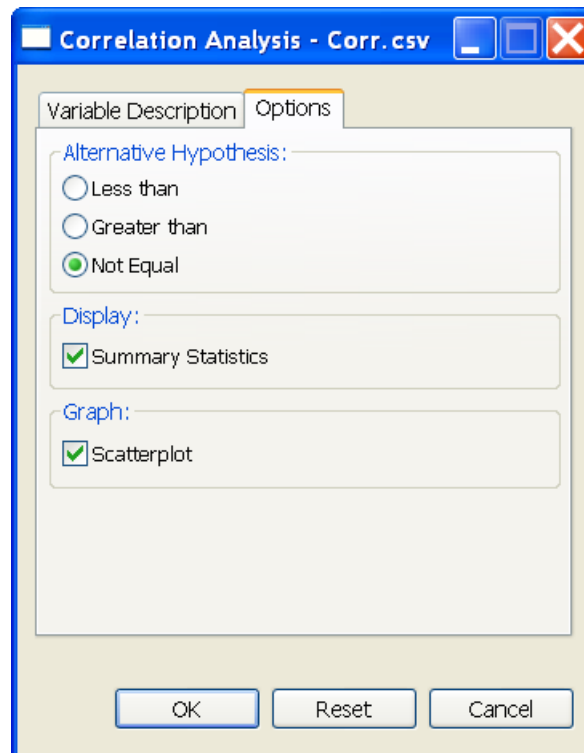
Scatterplot

If this option is selected, scatter plot for all pair of variables listed in the **Test Variable(s)** list box will be created.

For the example, the completed **Variable Description** tab should appear as illustrated below:



The completed **Option** tab should appear as illustrated below:



- Once all options have been specified, click the **OK** button. The Correlation Analysis dialog box will be minimized and STAR activates the **Output Viewer** which shows the results of the analysis.
- Sample output of the Correlation Analysis using Pearson's product-moment correlation is shown below.

Result of Correlation Analysis

DESCRIPTIVE STATISTICS

Variable	N_NonMissObs	Min	Max	Mean	StdDev
GY14	48	1678.00	7860.00	4788.67	1552.27
N	48	0.10	1.80	1.23	0.33
P	48	0.10	0.30	0.23	0.08
K	48	0.20	0.40	0.31	0.06

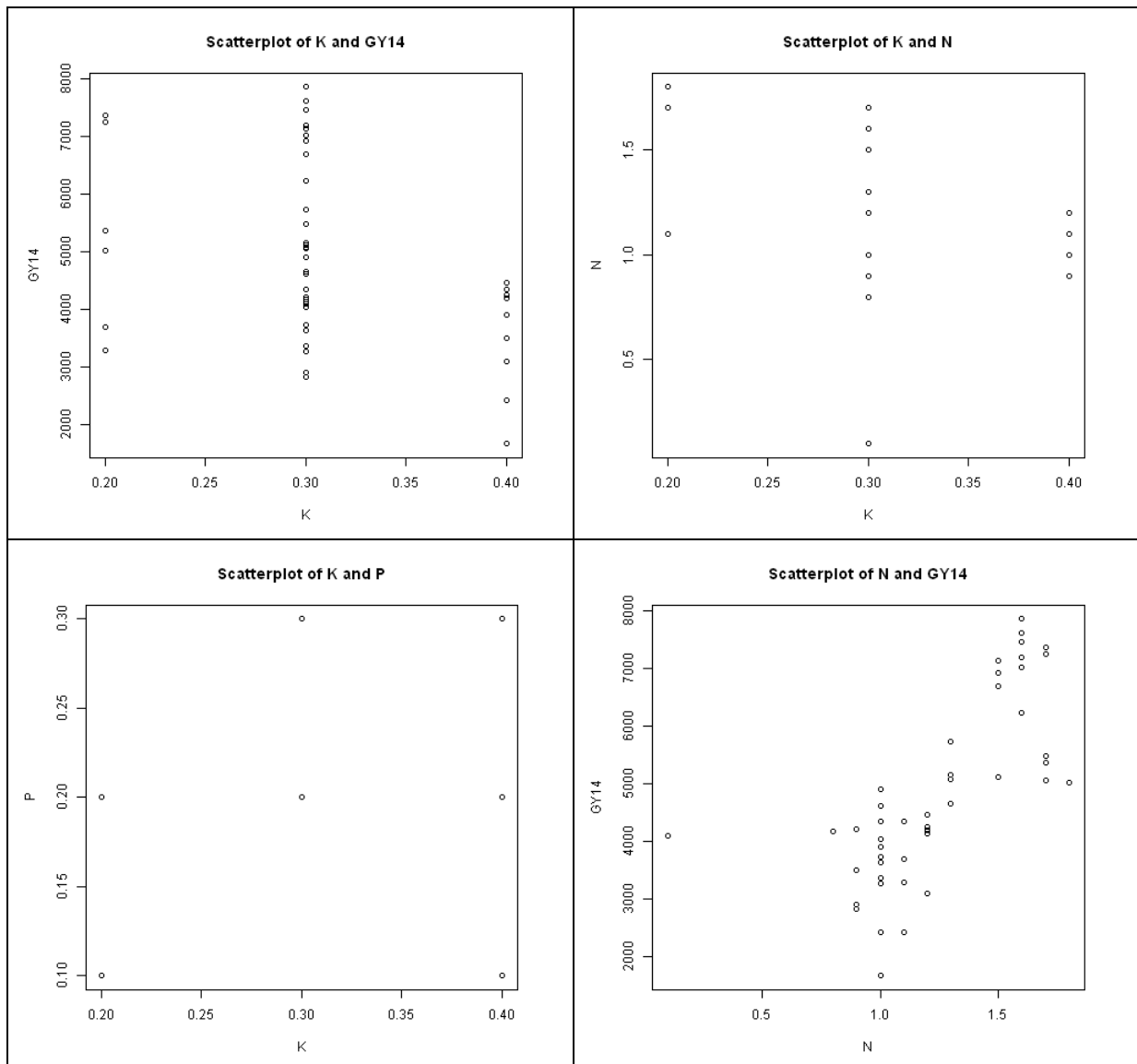
CORRELATION ANALYSIS

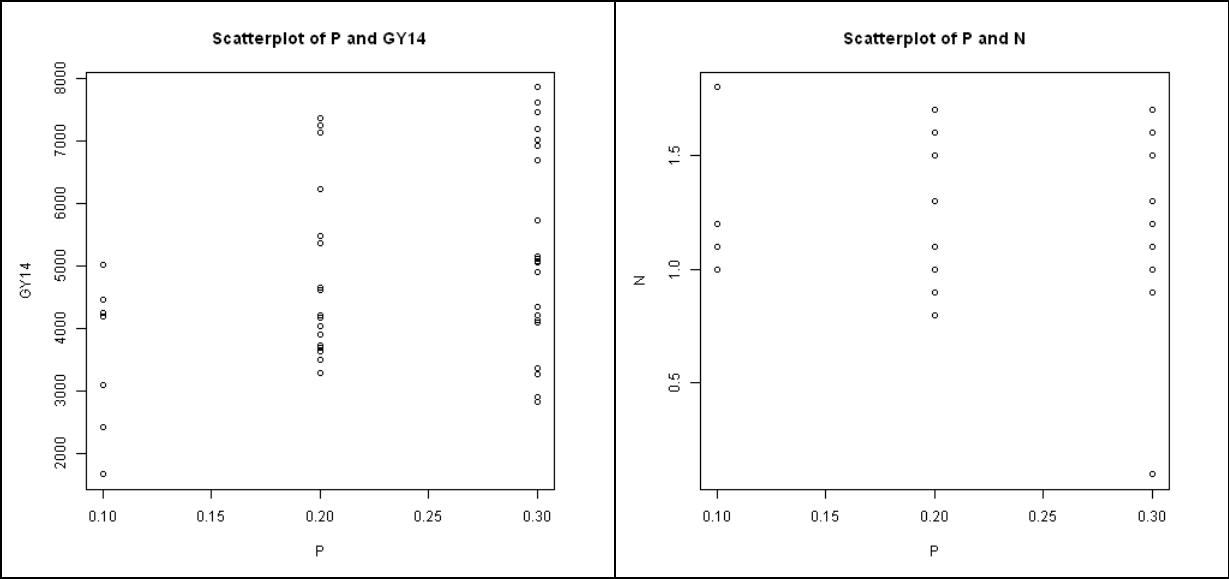
Pearson's product-moment correlation, Prob > |r|

		GY14	N	P	K
GY14	coef	1.00	0.72	0.38	-0.40
	p-value		0.0000	0.0080	0.0053
	n	48	48	48	48
N	coef	0.72	1.00	0.02	-0.34
	p-value	0.0000		0.8686	0.0180
	n	48	48	48	48

P	coef	0.38	0.02	1.00	-0.35
	p-value	0.0080	0.8686		0.0161
	n	48	48	48	48
K	coef	-0.40	-0.34	-0.35	1.00
	p-value	0.0053	0.0180	0.0161	
	n	48	48	48	48

- If a scatter plot is requested, the graph can be viewed in the **Graph Viewer**. Sample generated graphs are shown below:





10. Linear Regression Analysis

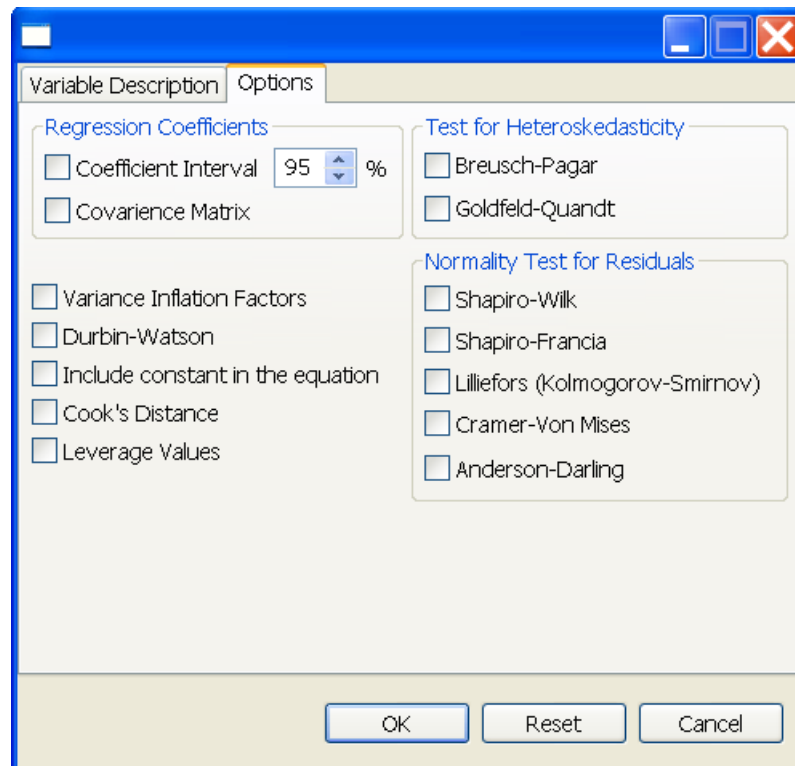
Regression Analysis is the statistical technique used to determine the strength of linear relationship between variables.

The steps to perform Linear Regression Analysis are listed below:

- On the Project Explorer, locate the dataset *plantHeight.csv* in the *Data* folder of the project named *SampleProject*. Double-click the file to open and view it in the Data Viewer tab. The file contains the data of number of days after seeding (DAS) as independent variable and plant height as dependent variable. Based on this data, simple linear regression analysis is to be conducted to assess how the days after seeding linearly affects plant height, on the average.

[illegible]

- Click **Analyze | Regression Analysis** from the main window of STAR. The **Regression** dialog box will appear.



- Specify the required field and appropriate options for the analysis:

Variable Description Tab

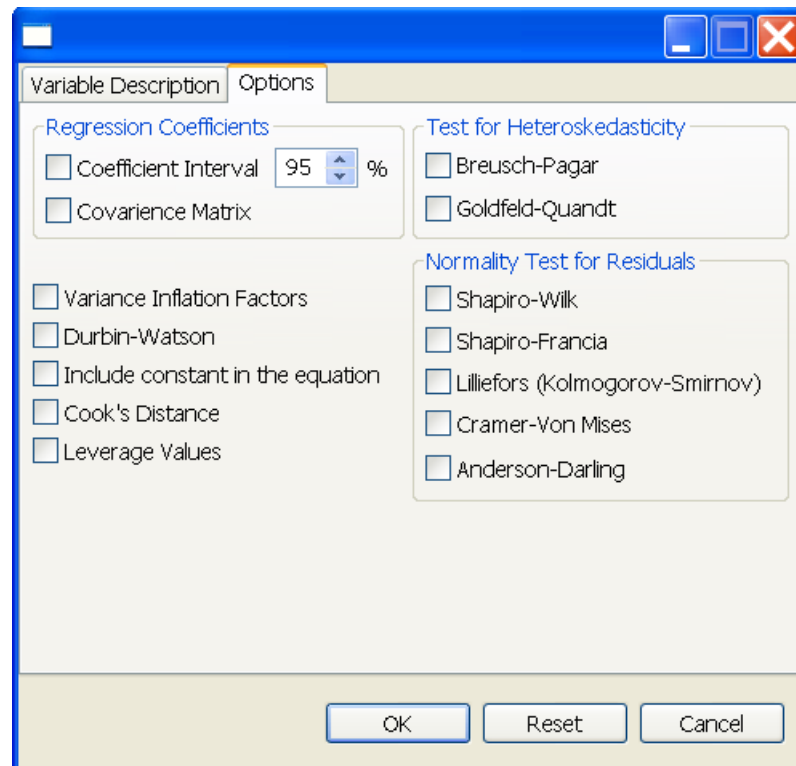
Dependent Variable

This field is required. For the analysis to proceed, this list box should have at least one entry.

Independent Variables

This field is required. For the analysis to proceed, this list box should have at least one entry.

Option Tab



Regression Coefficients

User can choose to display the confidence interval and the covariance matrix. If any of this is selected, the results is embedded in the table of parameter estimates. For the confidence interval, the default value is 95%. Valid value is between 90% to 99%.

Test for Heteroskedasticity

There are two available procedures namely: Breusch-Pagar and Goldfeld-Quandt.

Normality Test

If this option is selected, it will perform normality test for the residuals. There are five procedures available, namely: Shapiro-Wilk, Shapiro-Francia, Lilliefors, Cramer-Von Mises and Anderson-Darling.

Variance Inflation Factors

If this is selected, it displays the variance inflation factors (VIF) which measures the Collinearity (or multicollinearity).

Durbin-Watson

If this is selected, it displays the Durbin-Watson test for serial correlation of the residuals.

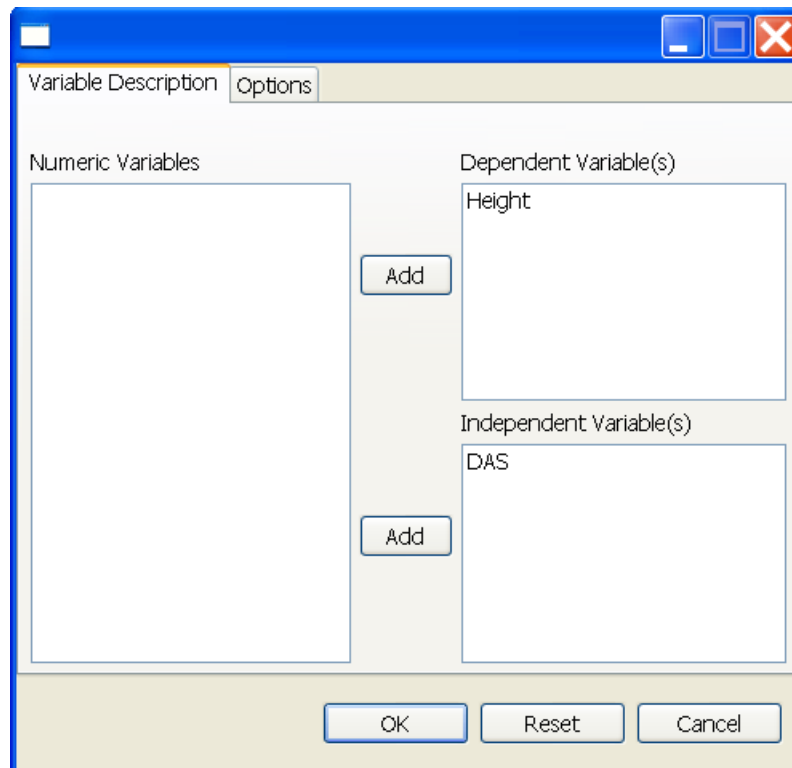
Include constant in the equation

By default, the regression model includes a constant term in the equation. Deselecting this option forces regression through the origin, this is rarely done.

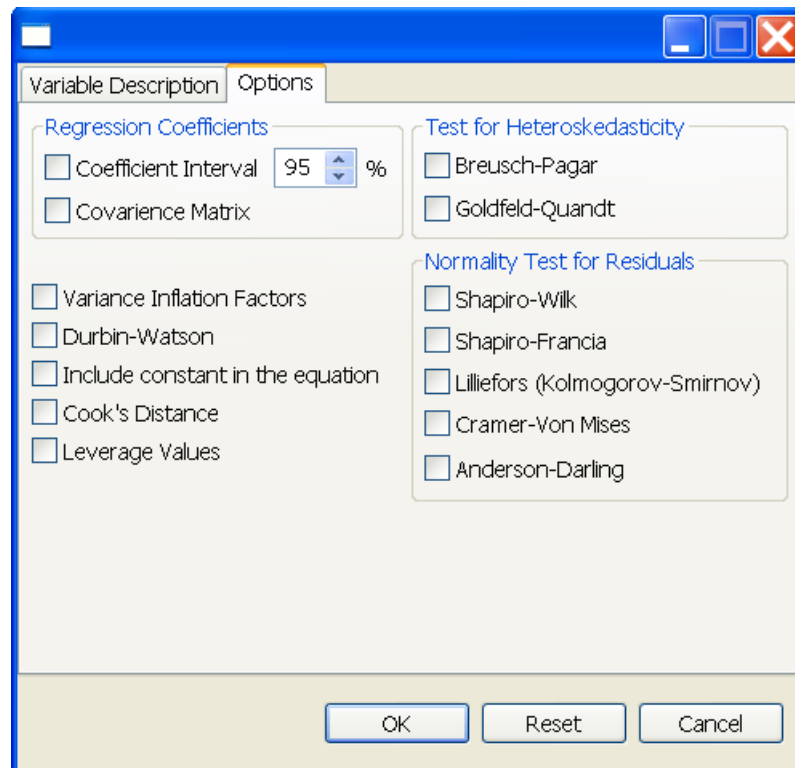
Cook's Distance

Leverage Values

For the example, the completed **Variable Description** tab should appear as illustrated below:



The completed **Option** tab should appear as illustrated below:



- Once all options have been specified, click the **OK** button. The Regression Analysis dialog box will be minimized and STAR activates the **Output Viewer** which shows the results of the analysis.

Output Viewer

The default output of the Regression Analysis procedure in STAR includes the Analysis of Variance table, a table of model summary and the parameter estimated. The output starts with the descriptive statistics, if this option was specified in the **Options** tab.

Graph Viewer

The tab displays two diagnostics plots. The left plot, entitled *Residuals vs. Fitted* is for assessing the distribution of the residuals. An ideal plot should reveal that the points are distributed around 0. The residuals should not be affected by the size of the fitted values. Problematic data may result to a plot wherein the points increase as the fitted values increase, thus forming a “funnel-shaped” distribution. This will indicate a violation of the homogeneity of variances requirement for analysis of variance. The *Normal Q-Q* plot is used for assessing normality. The points should lie about the line.

- Sample output of the Regression Analysis is shown below.

```

DESCRIPTIVE STATISTICS
-----
Variable    N_NonMissObs    Mean    StdDev    SE_Mean
-----
Height              5    51.40    45.99    20.57
    
```

DAS	5	38.00	37.01	16.55
-----	---	-------	-------	-------

LINEAR REGRESSION ANALYSIS
Model Fitted: Height ~ DAS

Analysis of Variance Table

Source	DF	Sum of Square	Mean Square	F Value	Pr > F
Model	1	8201.3898	8201.3898	95.44	0.0023
DAS	1	8201.3898	8201.3898	95.44	0.0023
Error	3	257.8102	85.9367		
Total	5	16660.5898			

Model Summary:

Root MSE	Height Mean	Coeff Var	R-Square	Adj R-Sq
9.27	51.40	89.47	0.9695	0.9594

Parameter Estimates:

Variable	Estimate	Std. Error	t value	LL CI*	UL CI*	Pr(> t)
Intercept	4.91	6.31	0.78	-15.17	25.00	0.4931
DAS	1.22	0.13	9.77	0.82	1.62	0.0023

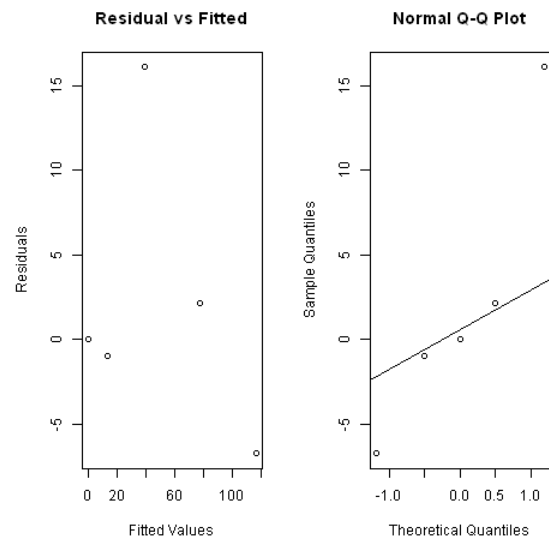
* At 95% Confidence Interval

TEST FOR NORMALITY

Variable	Method	Stat	Value	p Value
residual	Shapiro-Wilk	W	0.7532	0.0319

```
Durbin Watson Test for Autocorrelation
lag Autocorrelation D-W Statistic p-value
1 -0.1143968 2.037653 0.494
Alternative hypothesis: rho != 0
```

- The graph can be viewed in the **Graph Viewer**. Sample generated graphs are shown below:

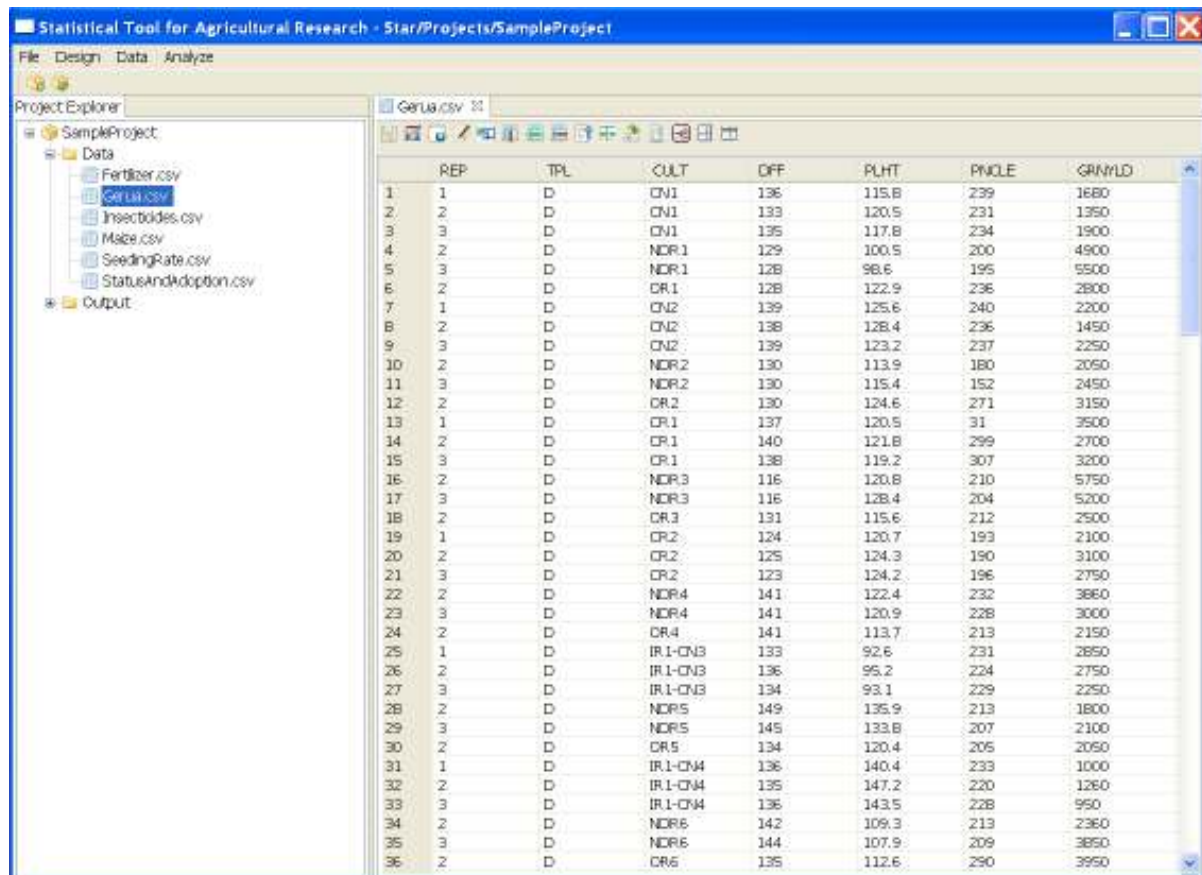


11. Normality Test

The Test for Normality menu display normality test for at least one numeric variable. There are five procedures available in this menu, namely: Shapiro-Wilk, Shapiro-Francia, Lilliefors (Kolmogorov-Smirnov), Cramer-Von Mises and the Anderson-Darling test for normality.

The steps to perform Normality Test are listed below:

- On the Project Explorer, locate the dataset *Gerua.csv* from the *Data* folder of the project named *SampleProject* in. Double-click the file to open and view it in the Data Viewer tab.



	REP	TPL	CULT	OFF	PLHT	PINOLE	GRN/YLD
1	1	D	CN1	136	115.8	239	1680
2	2	D	CN1	133	120.5	231	1350
3	3	D	CN1	135	117.8	234	1900
4	2	D	NDR1	129	100.5	200	4900
5	3	D	NDR1	128	98.6	195	5500
6	2	D	CR1	128	122.9	236	2800
7	1	D	CN2	139	125.6	240	2200
8	2	D	CN2	138	128.4	236	1450
9	3	D	CN2	139	123.2	237	2250
10	2	D	NDR2	130	113.9	180	2050
11	3	D	NDR2	130	115.4	152	2450
12	2	D	CR2	130	124.6	271	3150
13	1	D	CR1	137	120.5	31	3500
14	2	D	CR1	140	121.8	299	2700
15	3	D	CR1	138	119.2	307	3200
16	2	D	NDR3	116	120.8	210	5750
17	3	D	NDR3	116	128.4	204	5200
18	2	D	CR3	131	115.6	212	2500
19	1	D	CR2	124	120.7	193	2100
20	2	D	CR2	125	124.3	190	3100
21	3	D	CR2	123	124.2	196	2750
22	2	D	NDR4	141	122.4	232	3860
23	3	D	NDR4	141	120.9	228	3000
24	2	D	CR4	141	113.7	213	2150
25	1	D	IR1-CN3	133	92.6	231	2850
26	2	D	IR1-CN3	136	95.2	224	2750
27	3	D	IR1-CN3	134	93.1	229	2250
28	2	D	NDR5	149	135.9	213	1800
29	3	D	NDR5	145	133.8	207	2100
30	2	D	CR5	134	120.4	205	2050
31	1	D	IR1-CN4	136	140.4	233	1000
32	2	D	IR1-CN4	135	147.2	220	1260
33	3	D	IR1-CN4	136	143.5	228	950
34	2	D	NDR6	142	109.3	213	2360
35	3	D	NDR6	144	107.9	209	3850
36	2	D	CR6	135	112.6	290	3950

- Click **Analyze | Descriptive Statistics | Normality Test**. The **Normality Test** dialog box will appear.
- Opening the data for the first time, *Rep* field is regarded by R as numerical variables, they need to be changed as factors. Choose the variable and click on the **To Factor** button.
- Specify the required field and appropriate options for the analysis.

Variable Description Tab

Variables

This field is required. At least one item should be specified for the analysis to proceed.

By Variables

This field is optional. If a **By Variable(s)** is specified, it will perform test for normality per level of the **By Variable(s)**.

Options Tab

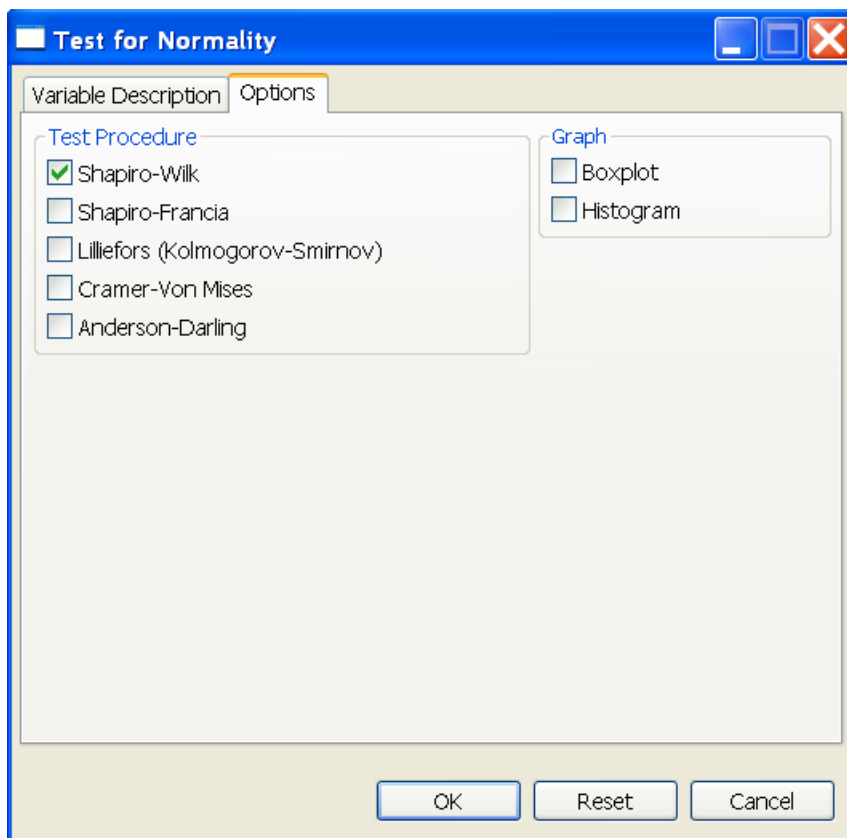
Test Procedure

This option is required. At least one test procedure should be tick for the analysis to proceed. By default, the Shapiro-Wilk test is performed.

Graph

The user has the option to display the box plot and the histogram.

For the example, the complete dialog box should appear as shown below:



- Click **Ok** button to perform the test. The **Normality Test** dialog box will be minimized and STAR activates the **Output Viewer** tab.

- Sample output of the normality test for variable *GRNYLD*.

TEST FOR NORMALITY

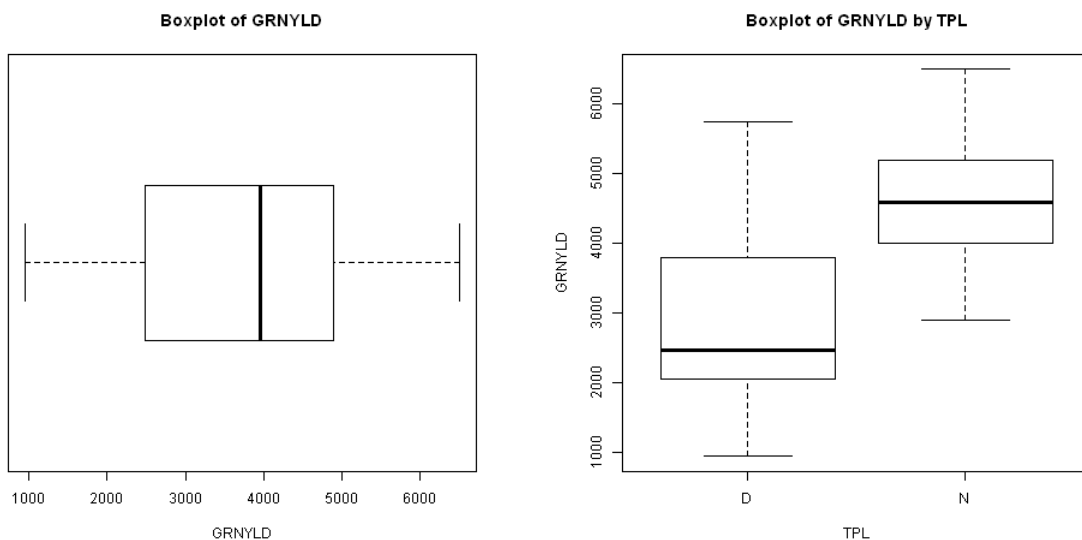
Variable	Method	Stat	Value	p Value
GRNYLD	Shapiro-Wilk	W	0.9624	0.0003
PNCLE	Shapiro-Wilk	W	0.9294	0.0000
PLHT	Shapiro-Wilk	W	0.9580	0.0001
DFF	Shapiro-Wilk	W	0.8137	0.0000

- Sample output of the normality test for variable *GRNYLD* by *TPL*.

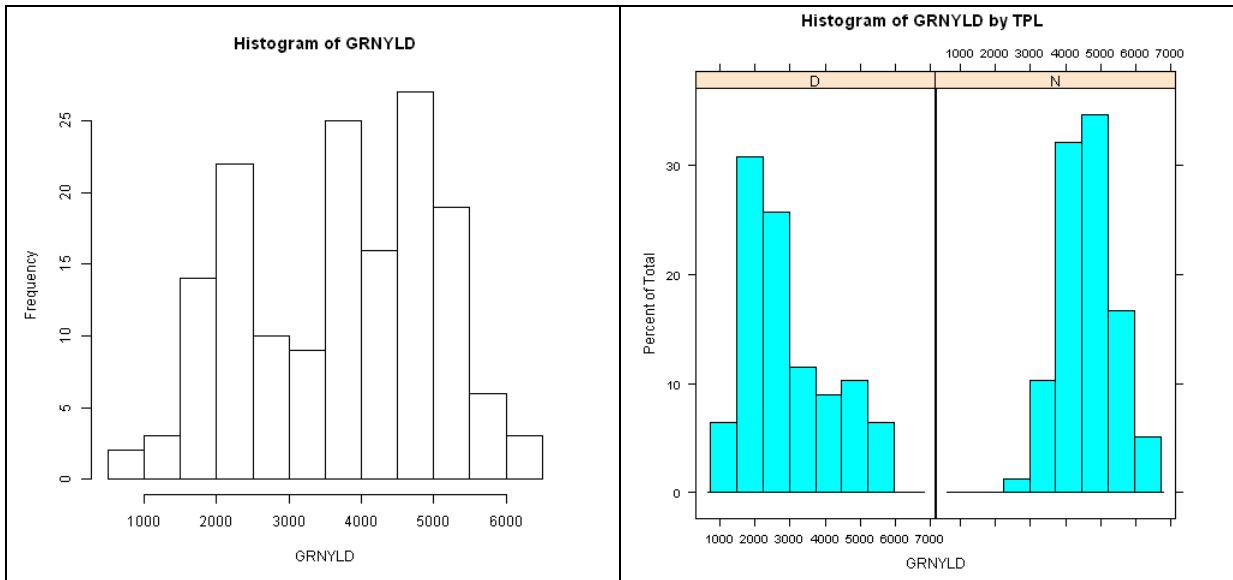
TEST FOR NORMALITY

Grp	Level	Variable	Method	Stat	Value	p Value
TPL	D	GRNYLD	Shapiro-Wilk	W	0.9176	0.0001
TPL	N	GRNYLD	Shapiro-Wilk	W	0.9908	0.8517

- If at least one of the graphs is chosen it can be viewed in the **Graph Viewer** tab. Below is the sample output if the boxplot was chosen:



- Below is the sample output if the histogram was chosen:

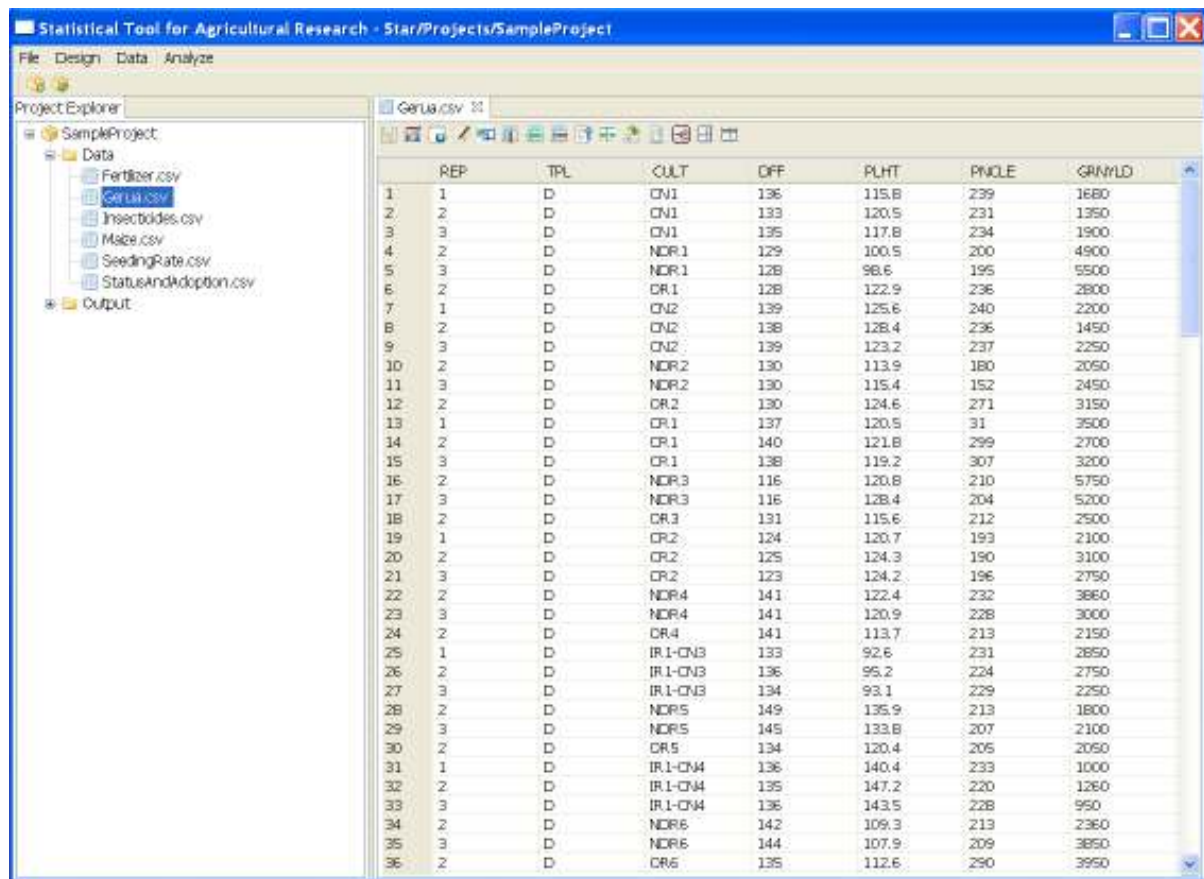


12. Heterogeneity Test

The Heterogeneity Test menu perform test for equality of variances. There are two procedures available in this menu, namely: Bartlett's test and the Levene's test.

The steps to perform Heterogeneity test are listed below:

- Locate the dataset *Gerua.csv* from the *Data* folder of the project named *SampleProject* in the Project Explorer.



	REP	TPL	CULT	OFF	PLHT	PNOLE	GRN/LD
1	1	D	CN1	136	115.8	239	1680
2	2	D	CN1	133	120.5	231	1950
3	3	D	CN1	135	117.8	234	1900
4	2	D	NDR1	129	100.5	200	4900
5	3	D	NDR1	128	98.6	195	5500
6	2	D	CR1	128	122.9	236	2800
7	1	D	CN2	139	125.6	240	2200
8	2	D	CN2	138	128.4	236	1450
9	3	D	CN2	139	123.2	237	2250
10	2	D	NDR2	130	113.9	180	2050
11	3	D	NDR2	130	115.4	152	2450
12	2	D	CR2	130	124.6	271	3150
13	1	D	CR1	137	120.5	31	3500
14	2	D	CR1	140	121.8	299	2700
15	3	D	CR1	138	119.2	307	3200
16	2	D	NDR3	116	120.8	210	5750
17	3	D	NDR3	116	128.4	204	5200
18	2	D	CR3	131	115.6	212	2500
19	1	D	CR2	124	120.7	193	2100
20	2	D	CR2	125	124.3	190	3100
21	3	D	CR2	123	124.2	196	2750
22	2	D	NDR4	141	122.4	232	3860
23	3	D	NDR4	141	120.9	228	3000
24	2	D	CR4	141	113.7	213	2150
25	1	D	IR1-CN3	133	92.6	231	2850
26	2	D	IR1-CN3	136	95.2	224	2750
27	3	D	IR1-CN3	134	93.1	229	2250
28	2	D	NDR5	149	135.9	213	1800
29	3	D	NDR5	145	133.8	207	2100
30	2	D	CR5	134	120.4	205	2050
31	1	D	IR1-CN4	136	140.4	233	1000
32	2	D	IR1-CN4	135	147.2	220	1260
33	3	D	IR1-CN4	136	143.5	228	950
34	2	D	NDR6	142	109.3	213	2360
35	3	D	NDR6	144	107.9	209	3850
36	2	D	CR6	135	112.6	250	3950

- Click **Analyze | Descriptive Statistics | Heterogeneity Test**. The **Heterogeneity Test** dialog box will appear.
- Opening the data for the first time, *Rep* field is regarded by R as numerical variables, they need to be changed as factors. Choose the variable and click on the **To Factor** button.
- Specify the required field and appropriate options for the analysis.

Variable Description Tab

Variables

This field is required. At least one item should be specified for the analysis to proceed.

By Variables

This field is required. All items specified here should contain two or more levels.

Options Tab

Test Procedure

This option is required. At least one test procedure should be tick for the analysis to proceed. By default, the Bartlett's test for homogeneity of variances is performed.

Graph

The user has the option to display the box plot and the histogram.

For the example, the complete dialog box should appear as shown below:

Suppose we want to determine whether the variances are equal among the levels of *TPL* for variables *DFF*, *PLHT*, *PNCLE* and *GRNYLD*. For the *example*, the completed **Variable Description** tab should appear as illustrated below:

The completed **Option** tab should appear as illustrated below:

- Click **Ok** button to perform the test. The **Normality Test** dialog box will be minimized and STAR activates the **Output Viewer** tab.
- Sample output of the normality test for variable *GRNYLD*.

13. Nonparametric Tests

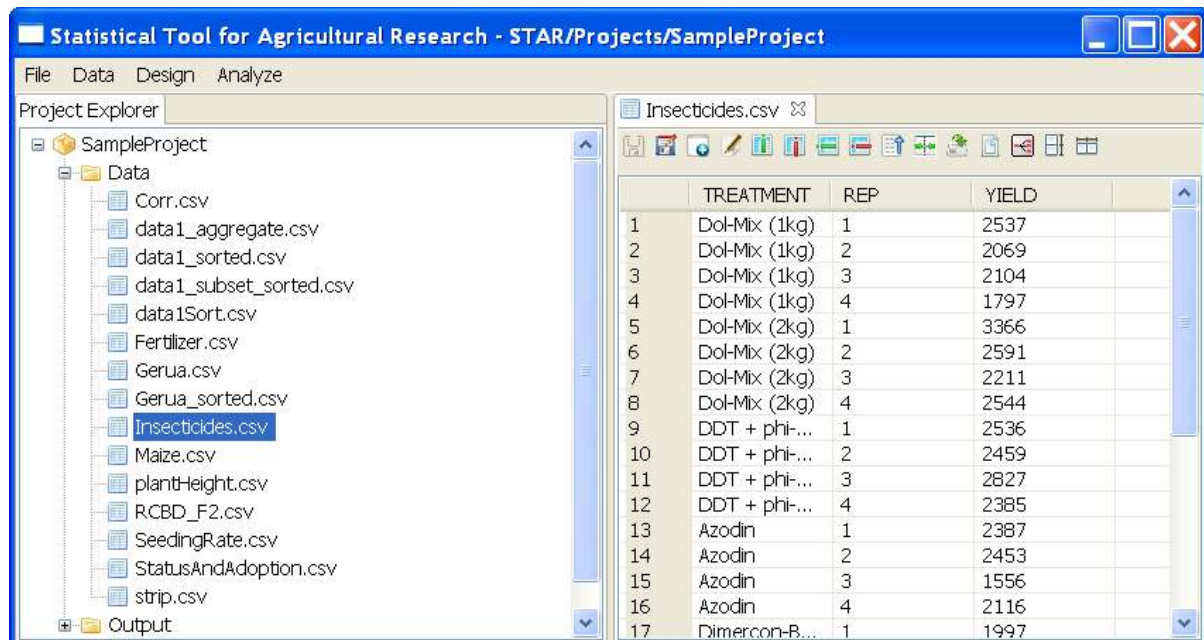
Nonparametric Tests procedure provides several tests that do not require assumptions about the shape of the underlying distribution.

13.1. Test for Several Independent Samples

This procedure perform Kruskal Wallis test.

The steps to perform Kruskal Wallis test are listed below:

- Locate the dataset *Insecticides.csv* from the *Data* folder of the project named *SampleProject* in the Project Explorer.



- Click **Analyze | Non-Parametric Test | Test for Several Independent Samples**. The **Test for Several Independent Samples** dialog box will appear.
- Specify the required field and appropriate options for the analysis. The completed dialog box should appear as shown below:

Test Several Independent Samples - Insectici...

Variable Description

Numeric Variable(s): REP

Test Variable(s): YIELD

Factor(s):

Grouping Variable: TREATMENT

Buttons: Add, To Factor, OK, Reset, Cancel

- Sample output is shown below:

Rank Sums for Variable YIELD
Classified by Variable TREATMENT

TREATMENT	N	Sum of Ranks	Mean of Ranks
Azodin	4	64	16.00
Control	4	12	3.00
DDT + phi-BHC	4	91	22.75
Dimercon-Boom	4	41	10.25
Dimercon-Knap	4	35	8.75
Dol-Mix (1kg)	4	66	16.50
Dol-Mix (2kg)	4	97	24.25

* Average scores are used for ties.

Kruskal-Wallis rank sum test

Chi-Square	20.8522
DF	6
Pr > Chi-Square	0.0020

14. Appendix

14.1. Data Description

14.1.1. Insecticides.csv

Description

Grain yield of rice resulting from use of different foliar and granular insecticides for the control of brown planthoppers and stem borers from a Completely Randomized Design Experiment with Four Replicates and 7 treatments. The data contains 28 rows and 3 columns.

Formats

This data contains the following columns:

Treatment	a factor with 7 levels (Azodin, DDT + γ -BHC, Dimercon-Boom, Dimercon-Knap, Dol-Mix (1 kg), Dol-Mix (2 kg) and Control) denoting the Different Foliar and Granular Insecticides
Rep	a factor with 4 levels (1 to 4) denoting the replicates
GrainYield	a numeric vector of grain yield in kg/ha

Source

Gomez, Kwanchai A. and Arturo A. Gomez. (1984) *Statistical Procedures for Agricultural Research 2nd Edition*.

14.1.2. Maize.csv

Description

Grain yield of three promising maize hybrids and a check variety from an experiment using Latin Square design. The data contains 24 rows and 3 columns.

Formats

This data contains the following columns:

Maize	a factor with 4 levels denoting the Maize Hybrids (A, B, and D) and a Check Variety (C)
Row	a factor with 4 levels (1 to 4) denoting the row blocking factor
Column	a factor with 4 levels (1 to 4) denoting the column blocking factor
GrainYield	a numeric vector of grain yield in kg/ha

Source

Gomez, Kwanchai A. and Arturo A. Gomez. (1984) *Statistical Procedures for Agricultural Research 2nd Edition*.

14.1.3. SeedingRate.csv

Description

Grain Yield of Rice Variety IR8 with Six Different Rates of Seeding, from a Randomized Complete Block Design Experiment with Four Replicates. The data contains 24 rows and 3 columns.

Formats

This data contains the following columns:

SeedingRate	Seeding Rate in kg seed/ha with 6 levels (25, 50, 75, 100, 125 and 150) denoting the treatment
Rep	block with 4 levels (1 to 4)
GrainYield	a numeric vector of grain yield in kg/ha

Source

Gomez, Kwanchai A. and Arturo A. Gomez. (1984) *Statistical Procedures for Agricultural Research 2nd Edition*.