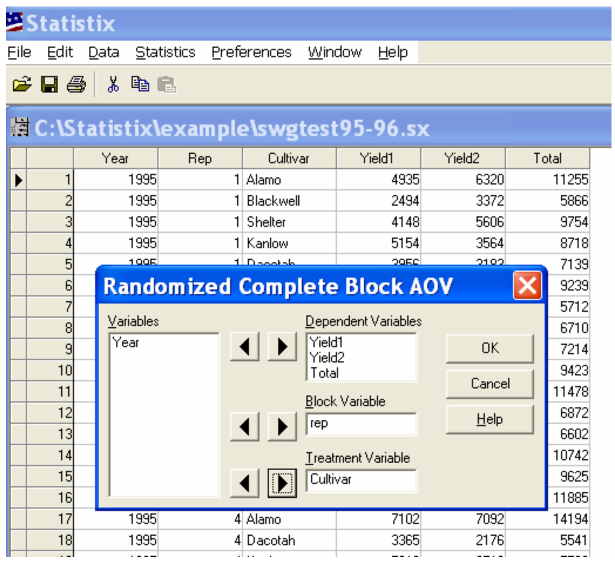
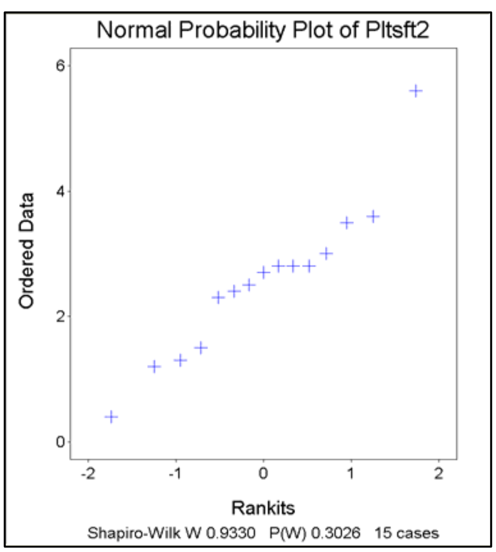
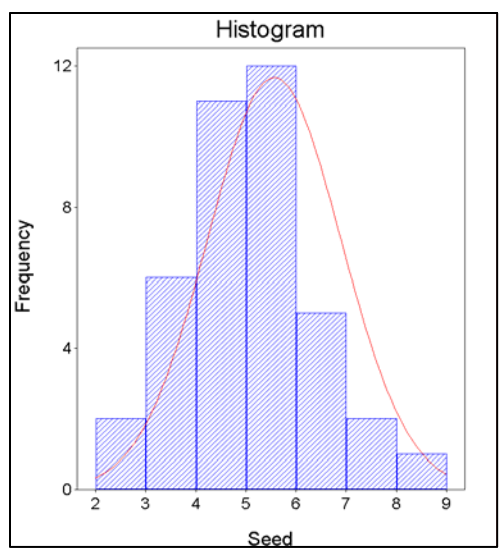
Statistix 8 User Guide
for the
Plant Materials Program







United States Department of Agriculture
Natural Resources Conservation Service

March 2007

Statistix 8 User Guide Version 2.0

Introduction

The Plant Materials Program generates a tremendous amount of data each and every year. Statistics is a great TOOL to help you make sense of your data. This User's Guide is built in the spirit of helping you use statistics effectively.

Statistics is daunting to some because it has a jargon all of its own. This User's Guide restates much of the "statistics jargon" in terms that Plant Materials scientists can better understand. Call out boxes, colored lettering, and circles are used to point out critical jargon and explain what these numbers, codes, and values really mean.

The User's Guide is not meant to be a substitute for the manual. The User's Guide focuses on information that is most useful to you, a plant scientist.

Statistix 8 is a commercial software package. It was chosen by a team of Plant Materials scientists based upon a number of factors. Ease of use and ability to import data from spreadsheets are two of the factors that weighed in heavily during the selection process.

Statistix 8 provides you with almost all of the statistical analysis techniques that the Plant Materials Program will ever use. If you require more robust, very highly specialized data analysis, there are other software programs available. However, as statistical complexity increases—so does the opportunity to misinterpret your data. Always consult with a qualified statistician before you embark on a study that generates complex data sets and equally complex statistical results.

Members of the Plant Materials Statistics Committee included: Paul Salon, Plant Materials Specialist, Syracuse, New York; Joel Douglas, Plant Materials Specialist, Central National Technology Support Center, Fort Worth, Texas; Ramona Garner, Manager, Tucson Plant Materials Center, Tucson, Arizona; Jim Stevens, Manager, East Texas Plant Materials Center, Nacogdoches, Texas; Mike Owsley, Manager, Jimmy Carter Plant Materials Center, Americus, Georgia; John Englert, Manager, National Plant Materials Center, Beltsville, Maryland; Mark Stannard, Team Leader, Pullman Plant Materials Center, Pullman, Washington.

Conventions

Menu items and button names are referred to in **boldface**:

Data, Insert, Variables
OK

Names of dialog boxes are ***boldface and italicized:***

Insert Variables

Item names within a dialog box are *italicized* and include: list and drop down list, text box and text area, radio button, and check box:

Variables list

Text to be typed are within “quotes”, but the quotes should not be typed:

“rep”

A Note: is not a step but a further explanation of a step or a caution of what to avoid.

An Example uses sample information to show what you can type in or what your result could look like.

Text indented in line with the steps describes the results you will see.

Figures are (usually) screen shots of what you will see on the screen; they are numbered according to their chapter.

Tables are marked discussions of output; they are numbered according to their chapter.

STATISTIX 8 USER’s MANUAL

Chapter 1: Data Entry and File Management

I. Entering Data into Statistix 8

A. Keyboard Data Entry

When you first run *Statistix 8*, an empty, untitled spreadsheet is displayed with the main menu bar appearing above it (Fig. 1-1). The main menu, or spreadsheet menu, is visible whenever the spreadsheet dialog box is the active dialog box. When you place your cursor on an item in the main menu, a drop down list of items appears, each of which is a submenu. These submenus offer a variety of data management and statistical procedure options.

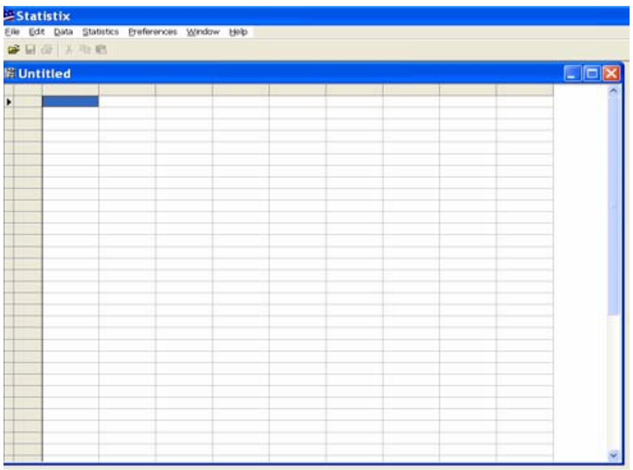


Figure 1-1

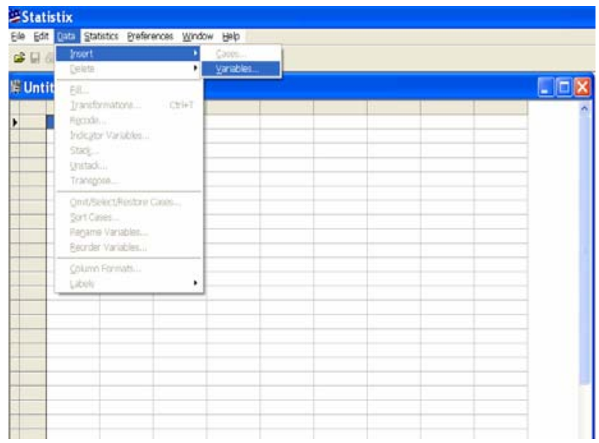


Figure 1-2

To create variables in the columns of the spreadsheet:

1. From the main menu, select **Data, Insert, Variables** (Fig. 1-2)
2. In the ***Insert Variables*** dialog box, in the *New Variable Names* list, enter the names of the variables, using a comma to separate the variable names (Fig. 1-3).

Example: "rep, cultivar, location, cover1, cover2, cover3, biomass"

response variables: biomass, cover1, cover2, cover3
treatment variables: cultivar, location

Note: A variable name is one to nine characters in length, must begin with a letter, and can only consist of letters, digits and the underscore character. You should assign meaningful variable names to describe the variables of interest. You cannot use such words as CASE, M, PI and RANDOM because these words have been reserved for other tasks.

3. Click **OK**.

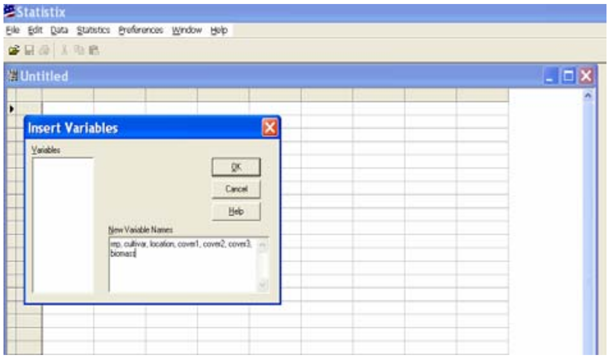


Figure 1-3

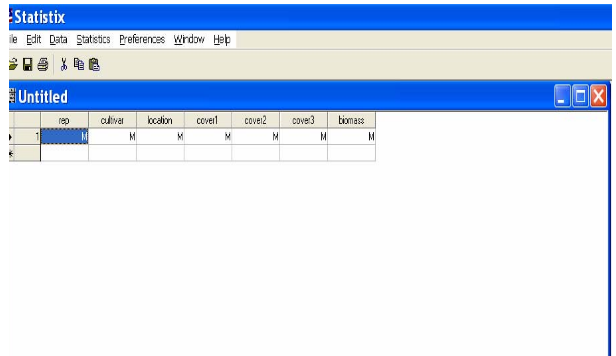


Figure 1-4

The variable names appear in separate columns followed by the letter "M" below the variable name (Fig. 1-4).

Note: The letter "M" represents missing data. Since no data has been entered in the spreadsheet then there will be "missing data" until data is added. *Statistix 8* requires that cases be added to the spreadsheet so you can enter data.

"Cases" are the total number of observations in the experiment. For example, if you have 2 cultivars of hairy vetch replicated 4 times at 3 test locations then you would enter 24 cases into the spreadsheet (2 x 4 x 3 = 24). They are represented by the rows.

To add cases (rows) to the spreadsheet:

1. From the main menu, select **Data, Insert, Cases** (Fig. 1-5).
2. In the ***Insert Cases*** dialog box, enter the number "1" in the *First New Case Number* text box to represent the first case (Fig. 1-6).
3. Enter the additional number of cases to achieve your total number of cases in the *Number of Cases to Insert* text box. (total – first new = number-of-cases-to-insert).

Note: The number of cases to insert is determined by: (Total number of cases – First new case number = Number of cases to insert). So for a total of 25 cases, (25 - 1 = 24):

enter "1" in the *First New Case Number* text box
enter "24" in the *Number of Cases to Inser*" text box

4. Click **OK**.

Once variable names have been entered and the proper number of cases identified, begin populating the *Statistix 8* spreadsheet with numerical values representing treatment variables.

Examples: cultivar =1, 2, etc; location = 1, 2, 3, 4, etc. (Fig. 1-7 and 1-8)

Note: If you prefer to use accession numbers, use only the last four digits rather than the NRCS 9 million number when entering data into the *Statistix 8* spreadsheet.

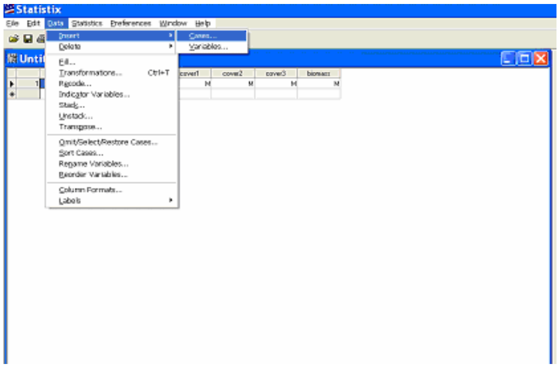


Figure 1-5

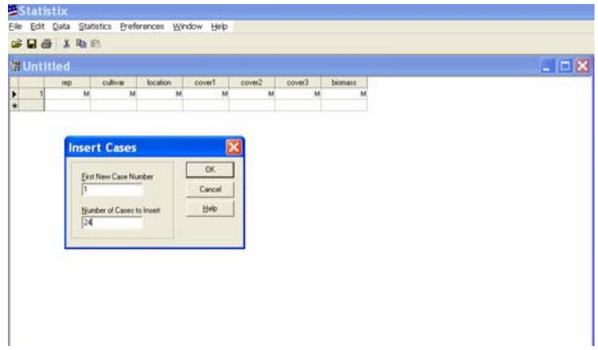


Figure 1-6

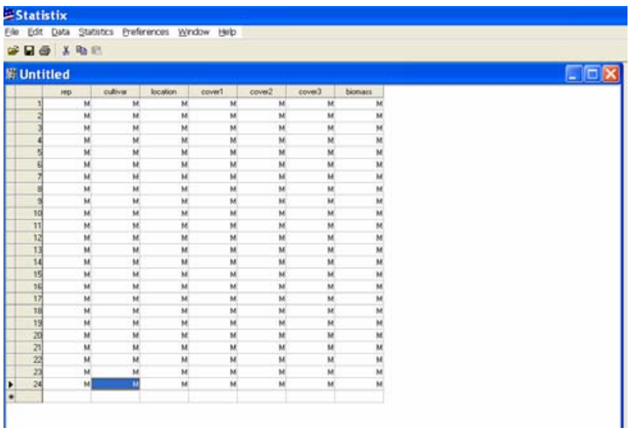


Figure 1-7

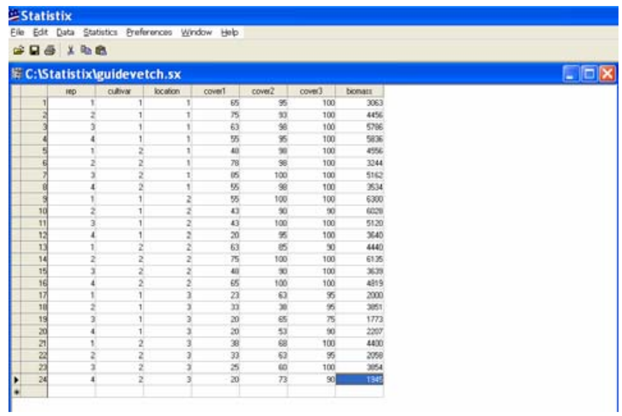


Figure 1-8

B. Entering Alphanumeric Data in Variable Name Columns

There are situations where alphanumeric data such as a cultivar name (e.g. 'Haskell', 'Shelter', 'Plateau') or location of study (e.g. PMC, Tribune, Altus) is preferred rather than designating a numerical value for the variable names (e.g. 1 = Haskell, 2 = Shelter).

The string (s) data type is used to enter alphanumeric data in the variable column(s) in a *Statistix 8* spreadsheet.

To enter alphanumeric data:

1. From the main menu, select **Data, Insert**, **Variables** (Fig. 1-2)**.**
2. In the ***Insert Variables*** dialog box, in the *New Variable Names* list (Fig. 1-3), type the variable name, followed by parentheses containing: the letter "s" (without quotes) followed by a number representing the maximum character length for the column, and end-parentheses.

Examples: cultivar (s7), Haskell (s7) (Fig. 1-9 and 1-10)

Note: Review the list of variable names to ensure that you are assigning the appropriate character length to the column. Remember the maximum length for a column variable name is 9 characters.

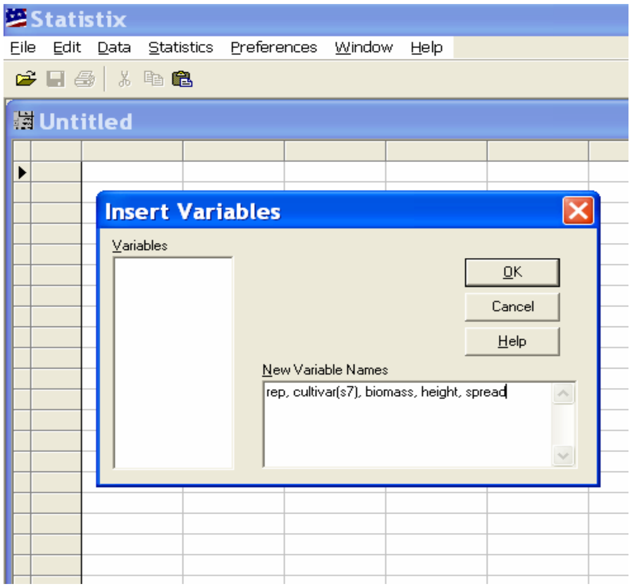
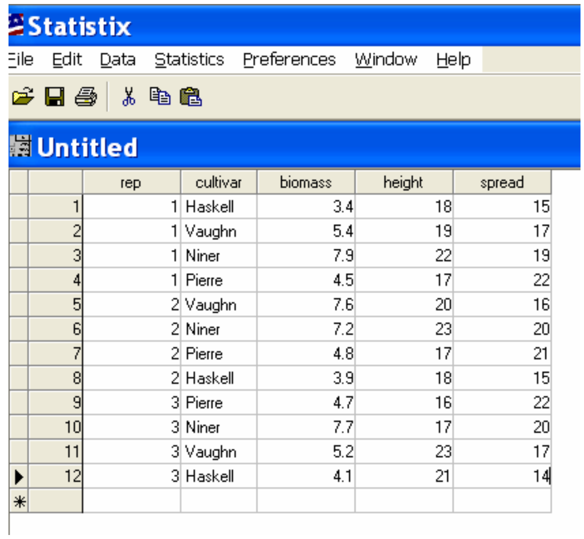


Figure 1-9



	rep	cultivar	biomass	height	spread
1	1	Haskell	3.4	18	15
2	1	Vaughn	5.4	19	17
3	1	Niner	7.9	22	19
4	1	Pierre	4.5	17	22
5	2	Vaughn	7.6	20	16
6	2	Niner	7.2	23	20
7	2	Pierre	4.8	17	21
8	2	Haskell	3.9	18	15
9	3	Pierre	4.7	16	22
10	3	Niner	7.7	17	20
11	3	Vaughn	5.2	23	17
12	3	Haskell	4.1	21	14

Figure 1-10

C. Renaming Variables

There are situations after data entry where you may want to rename variables to better reflect what data is being collected. For example, rather than “cover1”, “cover2”, and “cover3” you may want to change them to “canopy1”, “canopy2” and “canopy3”. To make this change:

1. From the main menu, select **Data**, **Rename Variables**.
2. In the ***Rename Variables*** dialog box, where you see the old variable name in the *Old Name* column, type the new variable name in the *New Name* column (Fig.1-11).

Example:

Old Name	New Name
cover1	canopy1
cover2	canopy2
cover3	canopy3

3. Click **OK.**

The new names will have replaced the old names and will appear in the spreadsheet (Fig. 1-12).

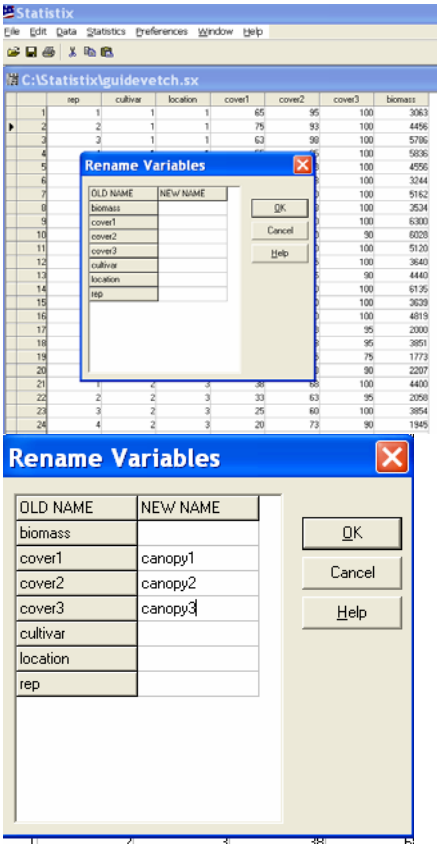


Figure 1-11

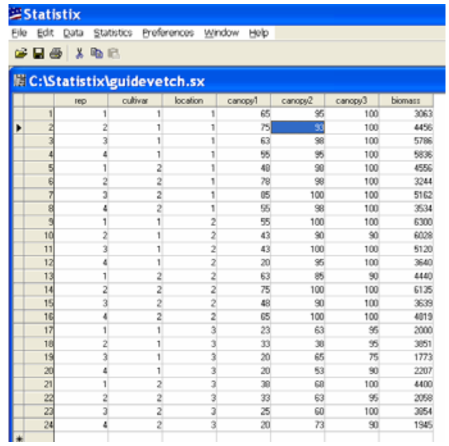


Figure 1-12

D. Inserting Additional Cases into *Statistix 8* Spreadsheet

Statistix 8 allows the user to insert more cases into an existing *Statistix 8* spreadsheet. For example, you wish to insert 5 additional cases into an existing *Statistix 8* spreadsheet which currently has 25 cases.

To add new cases:

1. From main menu, select **Data**, **Insert Cases.**
2. In the ***Insert Cases*** dialog box, *First New Case Number* text box, enter "26" to specify that the new case will be the 26th case, given that 25 cases already exist. In the *Number of New Cases to Insert* text box, type in the number "5" to add the additional 5 cases to the data set to total 30 cases (Fig. 1-13).

If you view the spreadsheet, you will see that there are now a total of 30 rows available for data input.

3. Click **OK**.

Once the new cases have been added, you can begin populating the data set (Fig. 1-14).

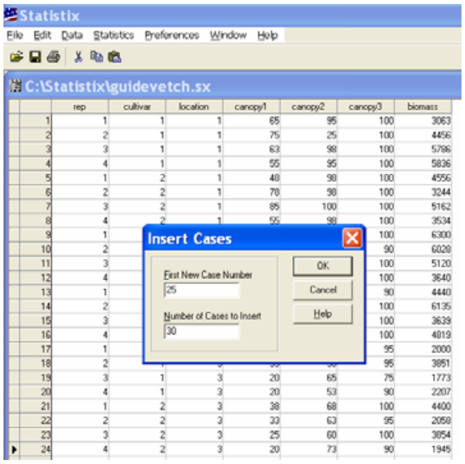


Figure 1-13

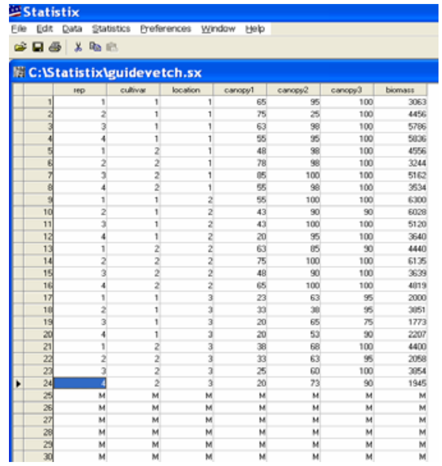


Figure 1-14

II. Moving Data from Microsoft Excel Spreadsheet into Statistix 8 Spreadsheet

If you prefer to copy data from an Excel spreadsheet into a *Statistix 8* spreadsheet for analysis, use the following steps.

1. Open the Excel spreadsheet containing the data.
2. In Excel, click in the upper-left hand corner where the beige row meets the beige column of the spreadsheet. The entire spreadsheet dataset will be highlighted.

This will highlight the entire spreadsheet (Fig.1-15).

3. From the Excel main menu, select **Edit, Copy.**
4. Open a *Statistix 8* spreadsheet, and click inside the upper left hand corner cell of the spreadsheet to select it.
5. From the *Statistix 8* main menu, select **Edit**, **Paste.**

The data set will be pasted in the *Statistix 8* spreadsheet (Fig.1-16).

<u>Note</u>: The maximum number of letters in the variable names is 9; thus only “replicati” appear in the column.

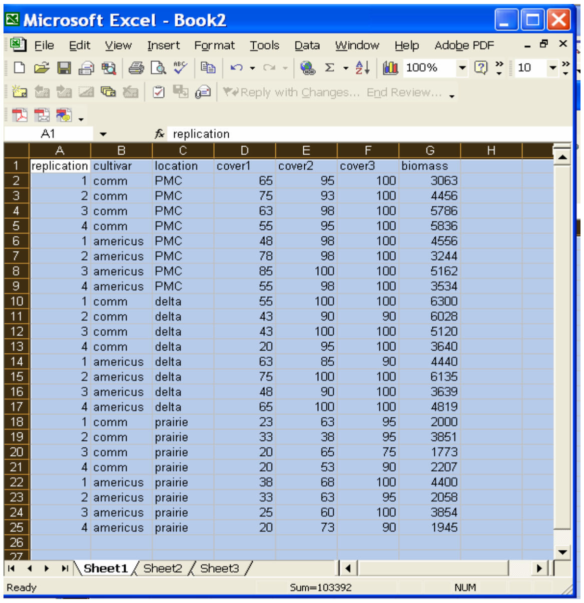


Figure 1-15

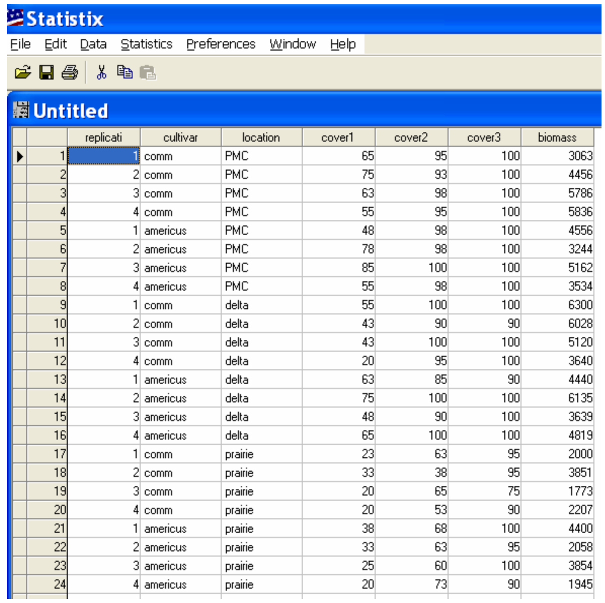


Figure 1-16

III. File Management

In this example, we will name the file "hairy vetch" which will be saved in a default folder named "Statistix".

To save a *Statistix 8* file:

1. From the main menu, select **File**, **Save As.**
2. Within the ***Save As*** dialog box, navigate to the folder you want to save the file into using the arrow to the right of the *Save in* drop down list box. Select the drive and folder your subfolder is located within. Double-click to enter subfolders until the desired folder is reached. Double-click the desired folder to open it.
3. In the ***Save As*** dialog box, enter the name of the file in the *File name* text box. In this case, the we entered filename "hairy vetch.sx"
4. Click the **Save** button.

<u>Note</u>: The filename will be given the ".sx" extension, which designates it as a *Statistix 8* file. Inn this case the filename appears as hairy vetch.sx (Fig.17).

You may want to save your data sets and analyses in specific folders other than the *Statistix* folder. For example, you may want to maintain all of your data sets on cover crops in a folder named "cover crop evaluations" or in folders titled after your study plan.

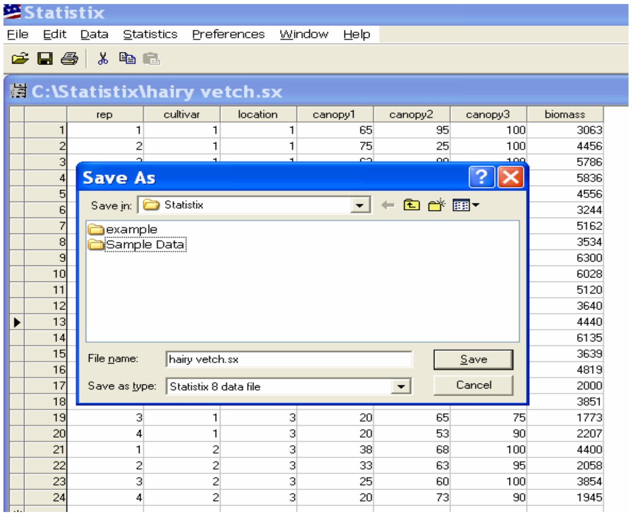


Figure 1-17

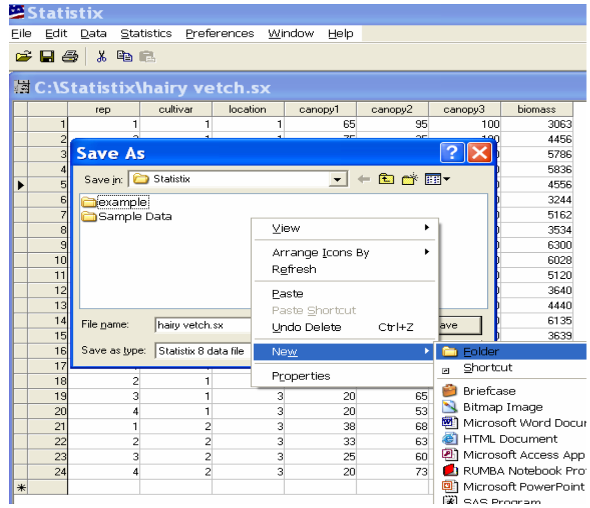


Figure 1-18

To create a new folder, one way is to:

1. After selecting **File, Save As** from the main menu, place the cursor inside the ***Save As*** dialog box and right click on the mouse.
2. From the shortcut menu, select **New, Folder** (Fig. 1-18).

A folder icon will appear in the directory you are viewing, with the words “New Folder” highlighted.

3. Type the name of the folder and it will replace the words “New Folder.” For this example, we named the new folder “cover crop evaluations” (Fig.1-19).
4. You can now save the file “hairy vetch.sx” into the “cover crop evaluation” folder (Fig. 1-20).

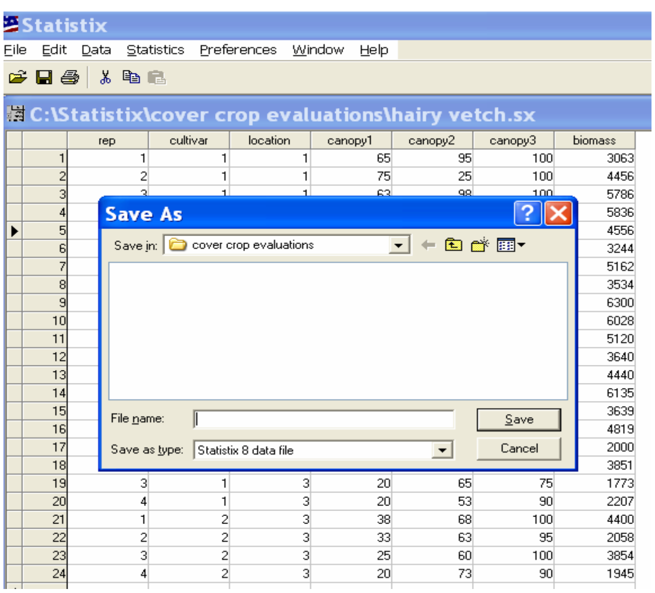


Figure 1-19

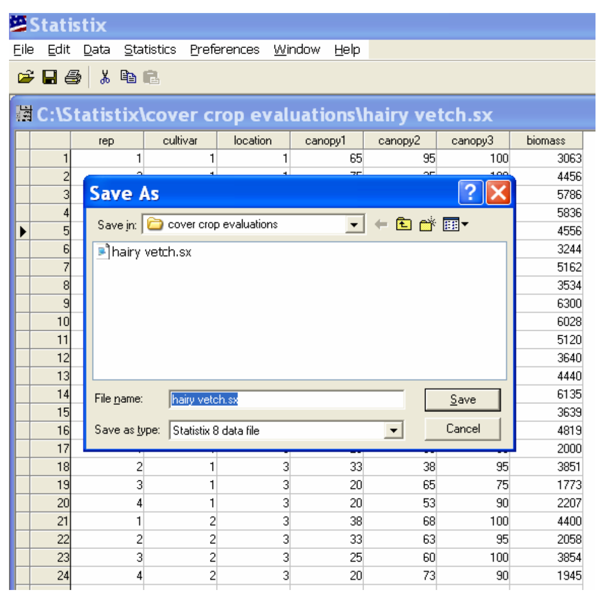


Figure 1-20

You can move a saved file from the “Statistix” folder into a subfolder by dragging:

1. After having saved the file, select **File, Save As** from the main menu.
2. Within the ***Save As*** dialog box, navigate to the folder that contains your file, using the arrow to the right of the *Save in* drop down list box. Select the drive and folder your subfolder is located within. Double-click to enter subfolders until the desired folder is reached.
3. Double-click the desired folder to open it.
4. Highlight the file.
5. At the same time, keep the left mouse button held down and drag the file into the chosen subfolder.

The file will have been moved to the new subfolder.

You can also move a saved file into another folder by cutting/copying and pasting:

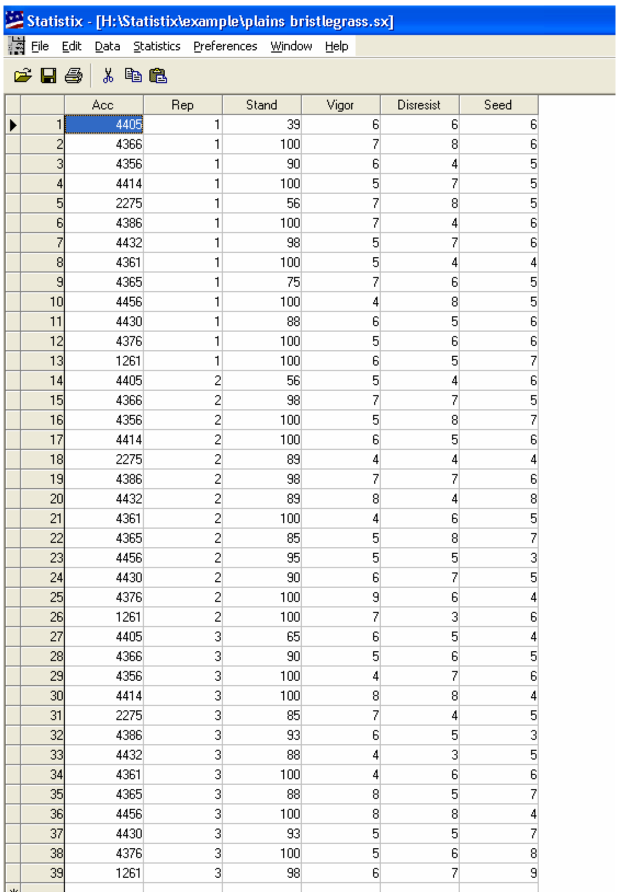
1. After having saved the file, select **File, Save As** from the main menu.
2. Within the ***Save As*** dialog box, navigate to the folder that contains your file, using the arrow to the right of the *Save in* drop down list box. Select the drive and folder your subfolder is located within. Double-click to enter subfolders until the desired folder is reached.
3. Double-click the desired folder to open it.
4. Highlight the file.
5. Right-click the mouse so that a shortcut menu appears, and select **Cut** or **Copy** from the shortcut menu.
6. Navigate to the folder you want to place the file in using the arrow to the right of the “Save in” list box as in step #2.
7. Click on the right mouse button so that a shortcut menu appears, and select **Paste** from the shortcut menu to paste the file into the new folder.

Chapter 2: Descriptive Statistics

Descriptive statistics such as mean, maximum/minimum, standard deviation, and standard error are popular analyses used to reduce the set of measurements to a few summary measures that provide a rough picture of the original measurements. One of the simplest yet most powerful ways of summarizing data is by means of tables and graphs.

Statistix 8 provides the user with many different descriptive statistics options. In this example we will focus on a few of the descriptive statistics commonly used in exploratory data analyses.

I. Tables



	Acc	Rep	Stand	Vigor	Disresist	Seed
1	4405	1	39	6	6	6
2	4366	1	100	7	8	6
3	4356	1	90	6	4	5
4	4414	1	100	5	7	5
5	2275	1	56	7	8	5
6	4386	1	100	7	4	6
7	4432	1	98	5	7	6
8	4361	1	100	5	4	4
9	4365	1	75	7	6	5
10	4456	1	100	4	8	5
11	4430	1	88	6	5	6
12	4376	1	100	5	6	6
13	1261	1	100	6	5	7
14	4405	2	56	5	4	6
15	4366	2	98	7	7	5
16	4356	2	100	5	8	7
17	4414	2	100	6	5	6
18	2275	2	89	4	4	4
19	4386	2	98	7	7	6
20	4432	2	89	8	4	8
21	4361	2	100	4	6	5
22	4365	2	85	5	8	7
23	4456	2	95	5	5	3
24	4430	2	90	6	7	5
25	4376	2	100	9	6	4
26	1261	2	100	7	3	6
27	4405	3	65	6	5	4
28	4366	3	90	5	6	5
29	4356	3	100	4	7	6
30	4414	3	100	8	8	4
31	2275	3	85	7	4	5
32	4386	3	93	6	5	3
33	4432	3	88	4	3	5
34	4361	3	100	4	6	6
35	4365	3	88	8	5	7
36	4456	3	100	8	8	4
37	4430	3	93	5	5	7
38	4376	3	100	5	6	8
39	1261	3	98	6	7	9

Figure 2-1

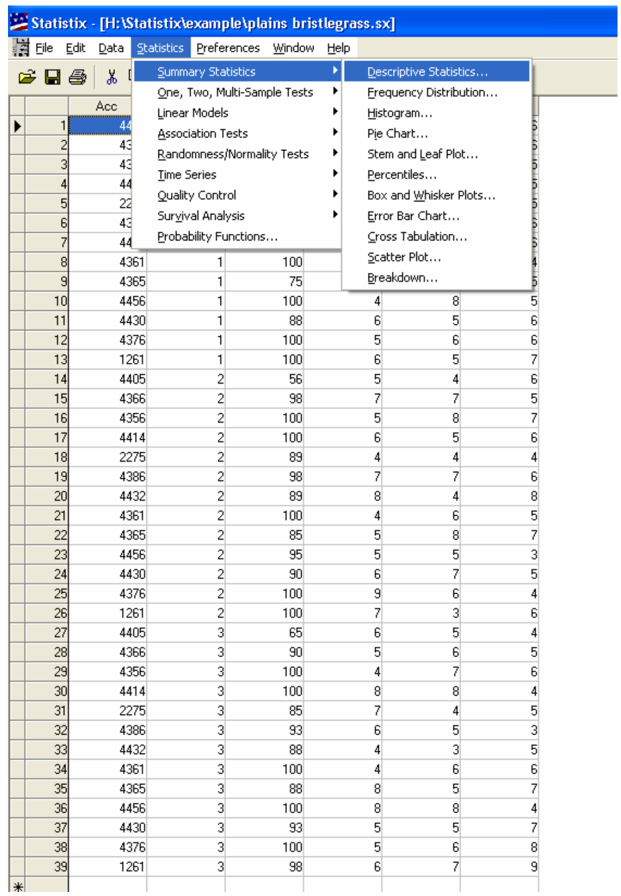


Figure 2-2

For this example, we will use subjective ratings for percent stand (“Stand”), plant vigor (“Vigor”), disease resistance (“Disresist”) and seed production (“Seed”) collected on 13 accessions of plains bristlegrass planted in a randomized complete block design with 3 replications (Fig. 2-1).

Note: Use only the last four digits of an accession number rather than the NRCS 9 million number when entering data into *Statistix 8* spreadsheet.

To analyze the data:

1. From the main menu, select **Statistics**, **Summary Statistics**, **Descriptive Statistics** (Fig. 2-2).
2. In the ***Descriptive Statistics*** dialog box (Fig. 2-3), move the variables "Stand", "Vigor", "Disresist", and "Seed" from the *Variables* list in the left column to the *Descriptive Variables* list one at a time by highlighting the variables and clicking on the right-arrow button to the left of the *Descriptive Variables* list.

Note: Each variable is moved one at a time from the *Variables* list to the *Descriptive Variables* list.

3. Move the variable "Acc" from the *Variables* list in the left column to the *Grouping Variable* list by highlighting the variable and clicking on the right-arrow button to the left of the *Grouping Variable* list. (Fig. 2-4).

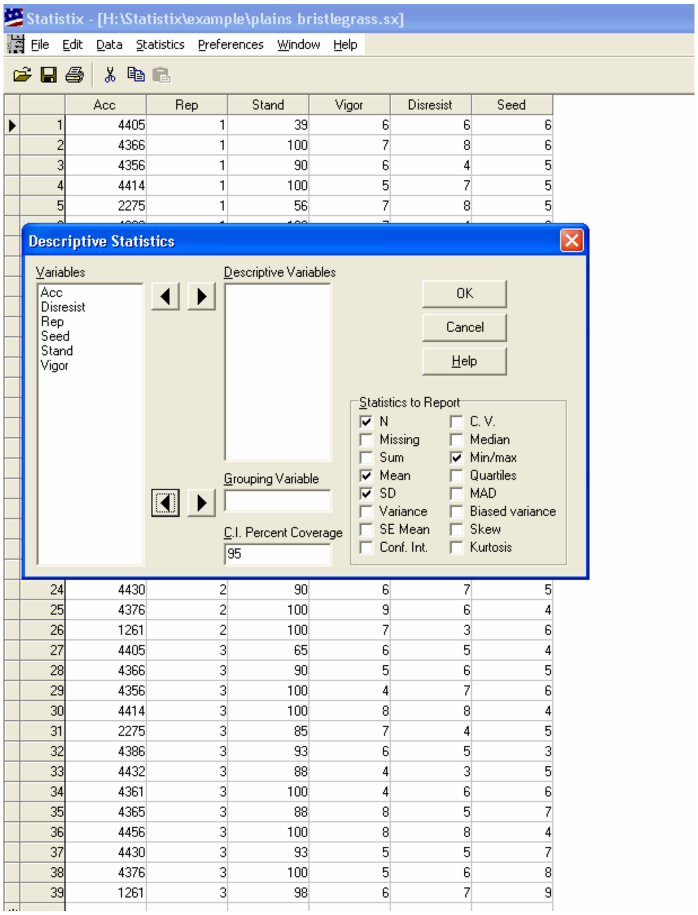


Figure 2-3

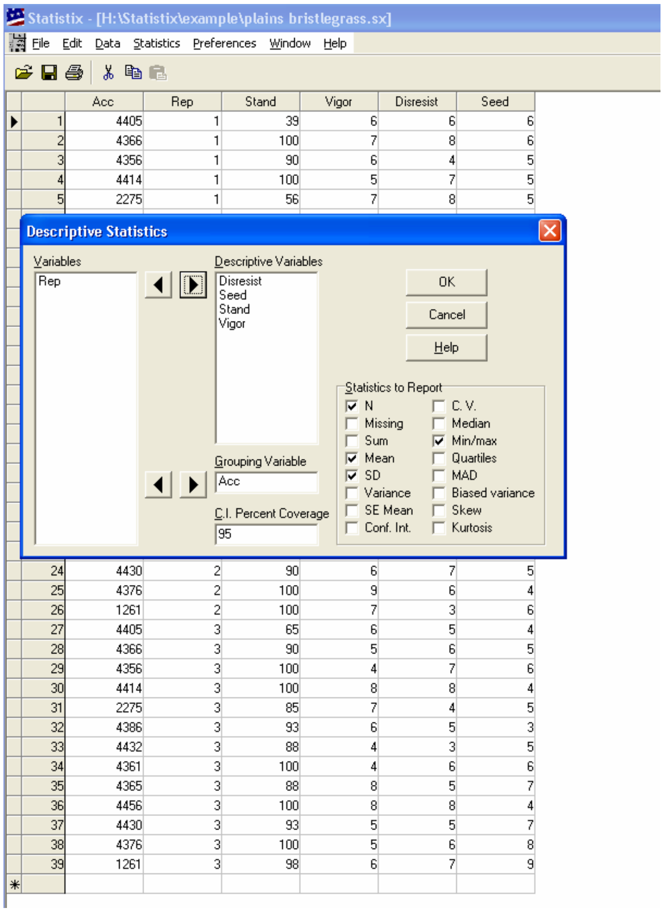


Figure 2-4

Notice in the “Statistics to Report” grouping in the ***Descriptive Statistics*** dialog box that options “N”, “Mean”, “SD”, and “Min/max” have been selected by default (Fig. 4-4). These defaults represent:

N = number of observations
Mean = mean of the sum of measurements divided by the total number of measurements
SD = standard deviation. SD is a measure of variability and is defined as the positive square root of the variances.
Min/max = minimum and maximum measurement

To add or remove a descriptive statistic:

1. Move appropriate variables from the *Variables* list into the *Descriptive Variables* or *Grouping Variable* lists by highlighting the variable clicking on the right-arrow button to the left of the new variable list.
2. Enter the appropriate confidence interval in the *C.I. Percent Coverage* text box, which is set to 95 by default.
3. Select only the tests you desire in the *Statistics to Report* group.
4. Click **OK.**

The results of the analysis on the 13 accessions are presented below (Table 2-1).

Table 2-1

Descriptive Statistics for Acc = 1261

Variable	N	Mean	SD	Minimum	Maximum
Disresist	3	5.0000	2.0000	3.0000	7.0000
Seed	3	7.3333	1.5275	6.0000	9.0000
Stand	3	99.333	1.1547	98.000	100.00
Vigor	3	6.3333	0.5774	6.0000	7.0000

Descriptive Statistics for Acc = 2275

Variable	N	Mean	SD	Minimum	Maximum
Disresist	3	5.3333	2.3094	4.0000	8.0000
Seed	3	4.6667	0.5774	4.0000	5.0000
Stand	3	76.667	18.009	56.000	89.000
Vigor	3	6.0000	1.7321	4.0000	7.0000

Descriptive Statistics for Acc = 4356

Variable	N	Mean	SD	Minimum	Maximum
Disresist	3	6.3333	2.0817	4.0000	8.0000
Seed	3	6.0000	1.0000	5.0000	7.0000
Stand	3	96.667	5.7735	90.000	100.00
Vigor	3	5.0000	1.0000	4.0000	6.0000

Descriptive Statistics for Acc = 4361

Variable	N	Mean	SD	Minimum	Maximum
Disresist	3	5.3333	1.1547	4.0000	6.0000
Seed	3	5.0000	1.0000	4.0000	6.0000
Stand	3	100.00	0.0000	100.00	100.00
Vigor	3	4.3333	0.5774	4.0000	5.0000

Descriptive Statistics for Acc = 4365

Variable	N	Mean	SD	Minimum	Maximum
Disresist	3	6.3333	1.5275	5.0000	8.0000
Seed	3	6.3333	1.1547	5.0000	7.0000
Stand	3	82.667	6.8069	75.000	88.000
Vigor	3	6.6667	1.5275	5.0000	8.0000

Descriptive Statistics for Acc = 4366

Variable	N	Mean	SD	Minimum	Maximum
Disresist	3	7.0000	1.0000	6.0000	8.0000
Seed	3	5.3333	0.5774	5.0000	6.0000
Stand	3	96.000	5.2915	90.000	100.00
Vigor	3	6.3333	1.1547	5.0000	7.0000

Descriptive Statistics for Acc = 4376

Variable	N	Mean	SD	Minimum	Maximum
Disresist	3	6.0000	0.0000	6.0000	6.0000
Seed	3	6.0000	2.0000	4.0000	8.0000
Stand	3	100.00	0.0000	100.00	100.00
Vigor	3	6.3333	2.3094	5.0000	9.0000

Descriptive Statistics for Acc = 4386

Variable	N	Mean	SD	Minimum	Maximum
Disresist	3	5.3333	1.5275	4.0000	7.0000
Seed	3	5.0000	1.7321	3.0000	6.0000
Stand	3	97.000	3.6056	93.000	100.00
Vigor	3	6.6667	0.5774	6.0000	7.0000

Descriptive Statistics for Acc = 4405

Variable	N	Mean	SD	Minimum	Maximum
Disresist	3	5.0000	1.0000	4.0000	6.0000
Seed	3	5.3333	1.1547	4.0000	6.0000
Stand	3	53.333	13.204	39.000	65.000
Vigor	3	5.6667	0.5774	5.0000	6.0000

Descriptive Statistics for Acc = 4414

Variable	N	Mean	SD	Minimum	Maximum
Disresist	3	6.6667	1.5275	5.0000	8.0000
Seed	3	5.0000	1.0000	4.0000	6.0000
Stand	3	100.00	0.0000	100.00	100.00
Vigor	3	6.3333	1.5275	5.0000	8.0000

Descriptive Statistics for Acc = 4430

Variable	N	Mean	SD	Minimum	Maximum
Disresist	3	5.6667	1.1547	5.0000	7.0000
Seed	3	6.0000	1.0000	5.0000	7.0000
Stand	3	90.333	2.5166	88.000	93.000
Vigor	3	5.6667	0.5774	5.0000	6.0000

Descriptive Statistics for Acc = 4432

Variable	N	Mean	SD	Minimum	Maximum
Disresist	3	4.6667	2.0817	3.0000	7.0000
Seed	3	6.3333	1.5275	5.0000	8.0000
Stand	3	91.667	5.5076	88.000	98.000
Vigor	3	5.6667	2.0817	4.0000	8.0000

Descriptive Statistics for Acc = 4456

Variable	N	Mean	SD	Minimum	Maximum
Disresist	3	7.0000	1.7321	5.0000	8.0000
Seed	3	4.0000	1.0000	3.0000	5.0000
Stand	3	98.333	2.8868	95.000	100.00
Vigor	3	5.6667	2.0817	4.0000	8.0000

II. Graphs

A scatter plots, normality plots and histogram are among the most popular and useful graphics for presentation of data. They allow the user to see the features of the data in a glance.

Statistix 8 can display data in a scatter plot, histogram or normal probability plot and test for normality using Shapiro-Wilk normality test.

For our examples, we will use the plain bristlegrass dataset. For graphing ease and interpretation of the scatter plot the accession numbers were assigned with a different numbering system (e.g. 4405 =1; 4366 = 2, 4356=3); thus, a new variable was created and inserted into the original dataset using the insert new variable procedure.

To enter a new variable:

1. From the main menu, select **Data**, **Insert**, **Variables**.
2. Enter “acc2” into the ***Insert Variables*** dialog box, in the *New Variable Names* text box.

3. Click OK.

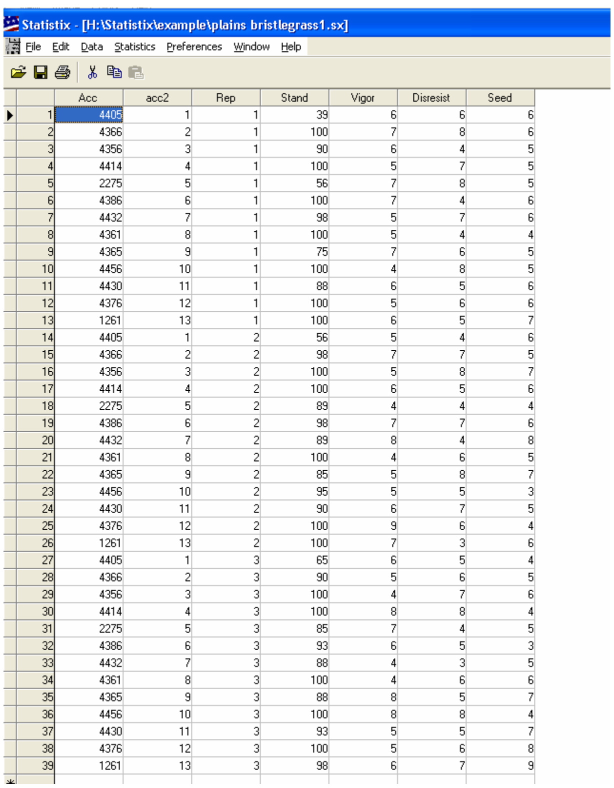
The new variable "acc2" is added to the list of variables in the dataset (Fig. 2-5).

A. Scatter Plot

To construct a scatter plot of the variable "acc2" against the variable "Seed":

1. From the main menu, select **Statistics**, **Summary Statistics**, **Scatter Plot** (Fig. 2-6).
2. In the ***Scatter Plot*** dialog box (Fig. 2-7), move the variable "acc2" from the *Variables* list in the left column to the *X-Axis Variables* list by highlighting the variable and clicking the right-arrow button to the left of the *X-Axis Variables* list (Fig. 2-8).
3. Move the variable "Seed" from the *Variables* list in the left column to the *Y-Axis Variables* list in the middle column by highlighting the variable and clicking the right-arrow button to the left of the *Y-Axis Variables* list.
4. Click **OK**.

Statistix 8 produces a scatter plot of the data (Fig. 2-9).



	Acc	acc2	Rep	Stand	Vigor	Disresist	Seed
1	4405	1	1	39	6	6	6
2	4366	2	1	100	7	8	6
3	4356	3	1	90	6	4	5
4	4414	4	1	100	5	7	5
5	2275	5	1	56	7	8	5
6	4386	6	1	100	7	4	6
7	4432	7	1	98	5	7	6
8	4361	8	1	100	5	4	4
9	4365	9	1	75	7	6	5
10	4456	10	1	100	4	8	5
11	4430	11	1	88	6	5	6
12	4376	12	1	100	5	6	6
13	1261	13	1	100	6	5	7
14	4405	1	2	56	5	4	6
15	4366	2	2	98	7	7	5
16	4356	3	2	100	5	8	7
17	4414	4	2	100	6	5	6
18	2275	5	2	89	4	4	4
19	4386	6	2	98	7	7	6
20	4432	7	2	89	8	4	8
21	4361	8	2	100	4	6	5
22	4365	9	2	85	5	8	7
23	4456	10	2	95	5	5	3
24	4430	11	2	90	6	7	5
25	4376	12	2	100	9	6	4
26	1261	13	2	100	7	3	6
27	4405	1	3	65	6	5	4
28	4366	2	3	90	5	6	5
29	4356	3	3	100	4	7	6
30	4414	4	3	100	8	8	4
31	2275	5	3	85	7	4	5
32	4386	6	3	93	6	5	3
33	4432	7	3	88	4	3	5
34	4361	8	3	100	4	6	6
35	4365	9	3	88	8	5	7
36	4456	10	3	100	8	8	4
37	4430	11	3	93	5	5	7
38	4376	12	3	100	5	6	8
39	1261	13	3	98	6	7	9

Figure 2-5

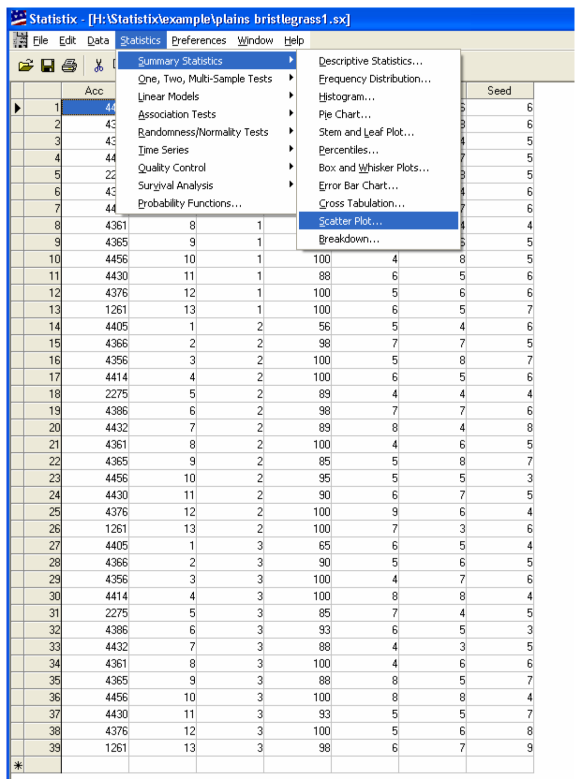


Figure 2-6

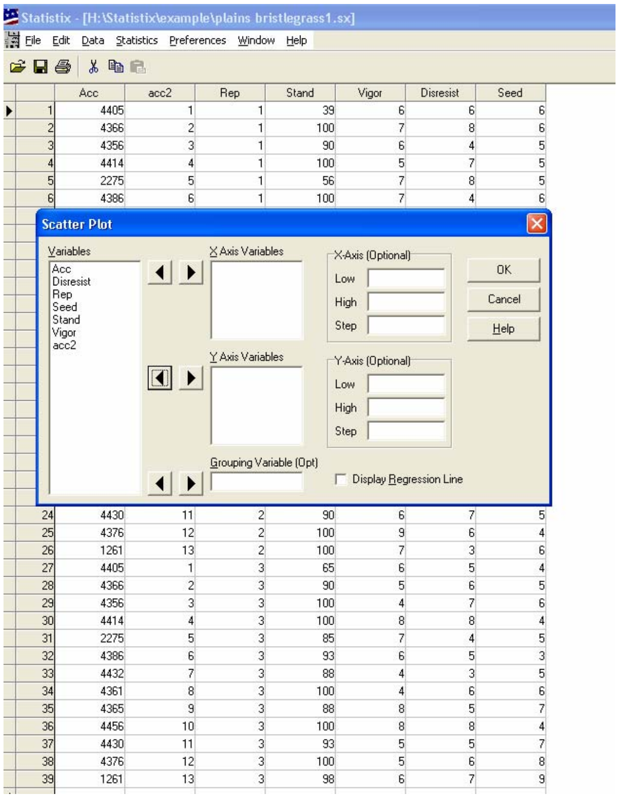


Figure 2-7

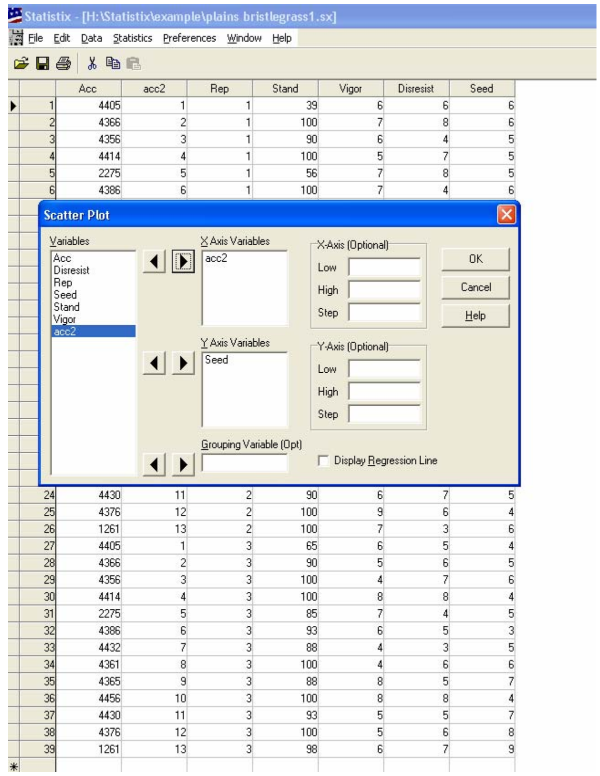


Figure 2-8

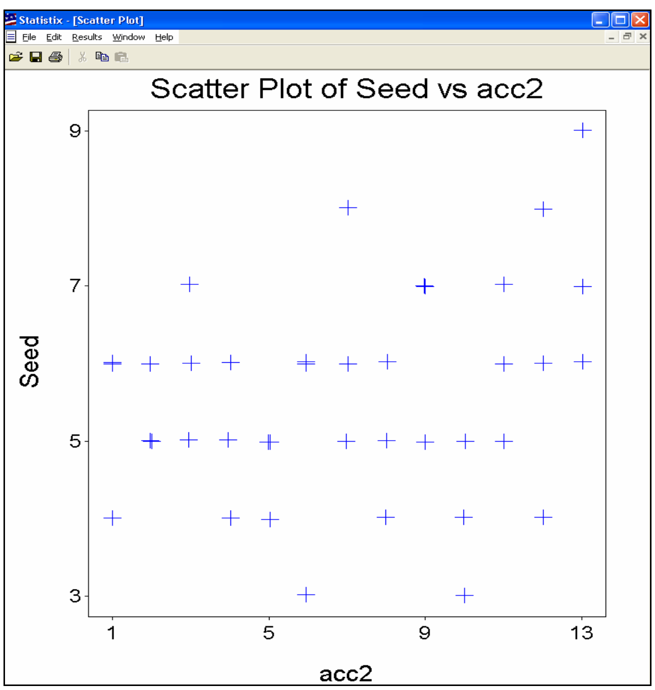


Figure 2-9

In reviewing the scatter plot, it appears that most of the accessions produced seed production rating of 5 and 6. Some accessions had above and below average seed ratings on some plots which may warrant further investigation, if deem necessary, to determine the possible cause(s).

B. Histogram

To construct a histogram of the variable “Seed”:

1. From the main menu, select, **Statistics**, **Summary Statistics**, **Histogram** (Fig. 2-10).
2. In the ***Histogram*** dialog box (Fig. 2-11), move the variable “Seed” from the *Variables* list to the *Histogram Variables* list by highlighting the variable and clicking the right-arrow button to the left of the *Histogram Variables* list (Fig. 2-12).
3. Check the *Display Normal Curve* check box to superimpose a normal curve over the bars of the histogram.
4. Click **OK**.

Statistix 8 produces a histogram of the “Seed” data (Fig. 2-13).

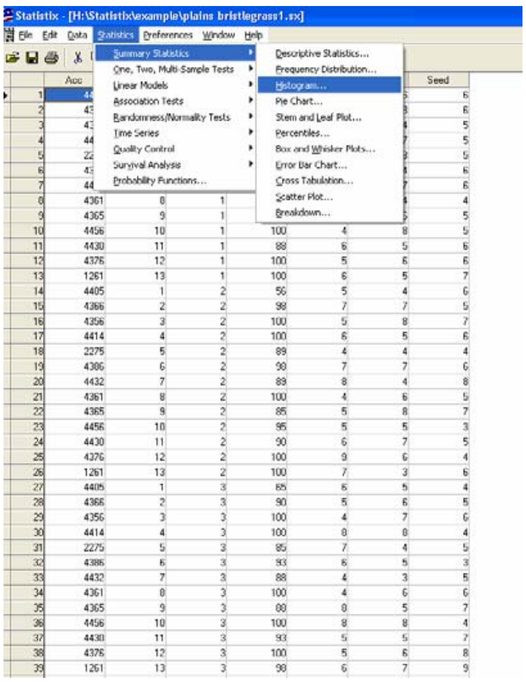


Figure 2-10

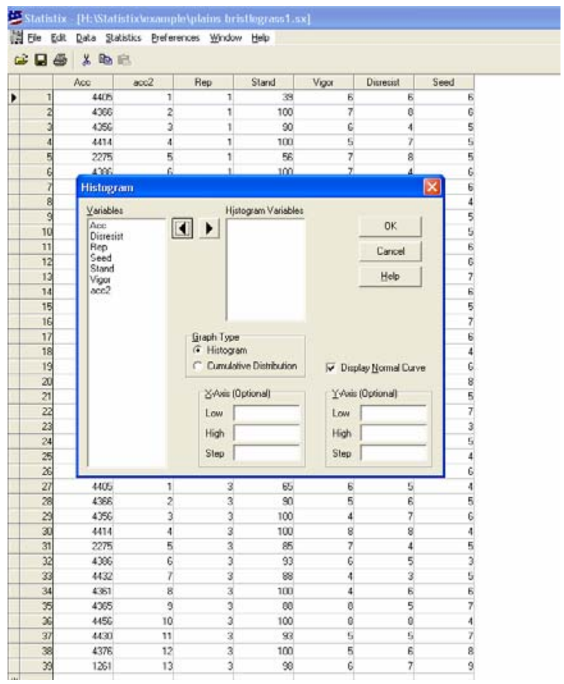


Figure 2-11

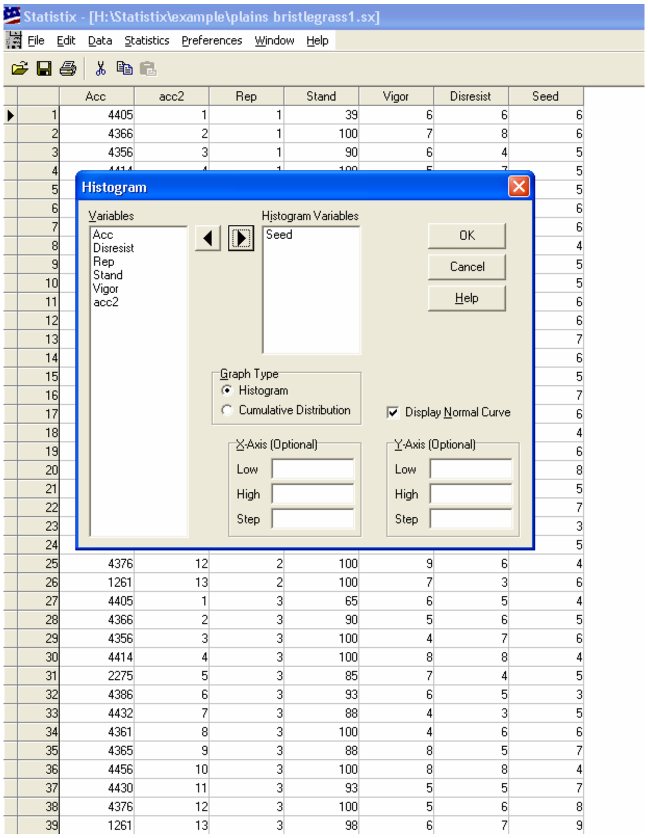


Figure 2-12

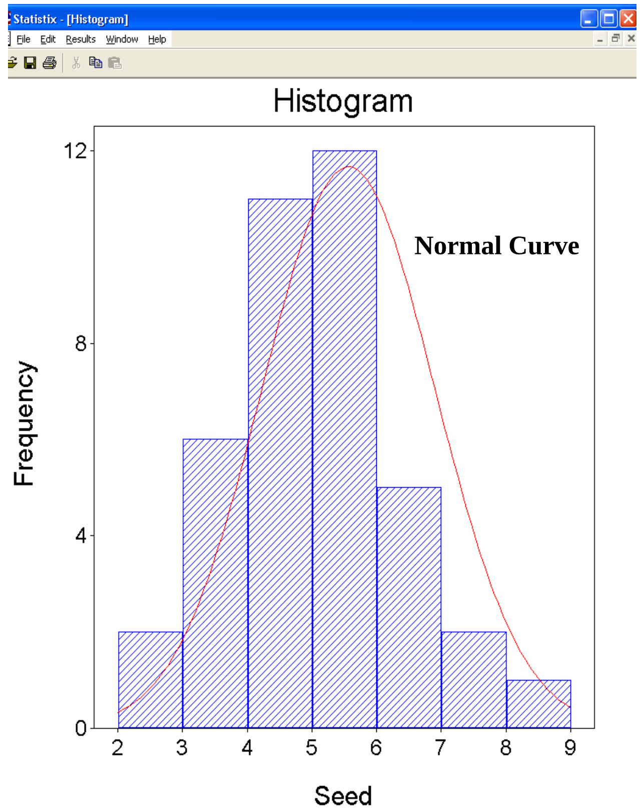


Figure 2-13

The histogram of the seed variable did not show abnormal distribution of the data. Extreme deviations beyond the normal curve would be of concern but small deviations would not be problematic.

C. Normal Probability Plot and Normality Test

A normality plot is an excellent tool to test the assumption that the data you are working with is normality distributed. The normality probability plot, plots the data against the corresponding rankits. When the plotted data are drawn from a normal population, the points appear to fall on a straight line. The Shapiro-Wilk Test, which examines whether the data conforms to a normal distribution, appears at the bottom of the graph.

For this example, we will apply the normal probability plot to data collected from a seeding rate trial of Florida paspalum (Fig. 2-14).

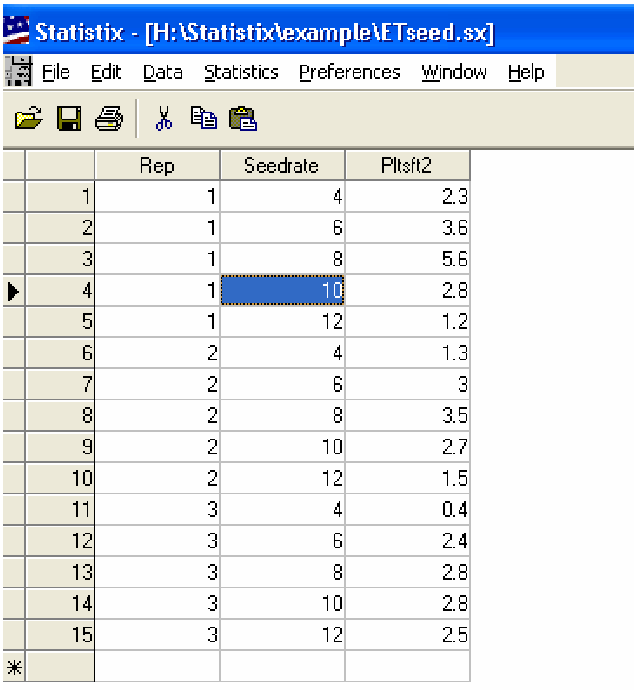


Figure 2-14

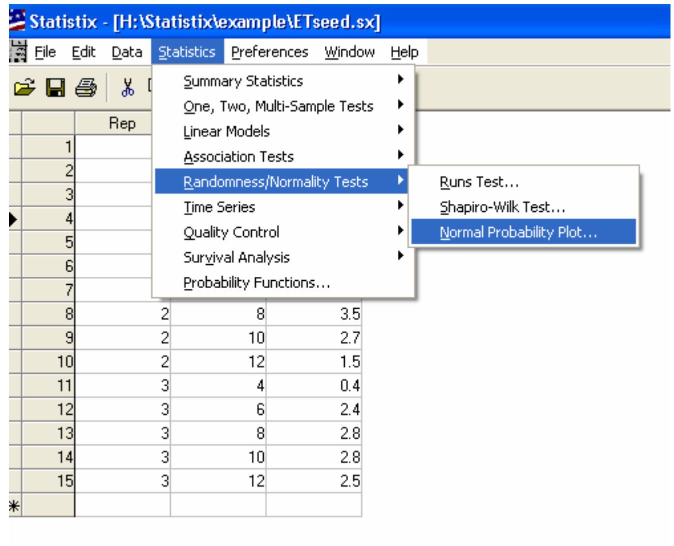


Figure 2-15

To construct a normal probability plot**:**

1. From the main menu**,** select **Statistics**, **Randomness/Probability Plot**, **Normal Probability Plot** (Fig.2-15).
2. In the ***Normal Probability Plot*** dialog box (Fig. 2-16), move the variable "Pltsft2" from the *Variables* list to the *Plot Variable* list by highlighting the variable and clicking the right-arrow button to the left of the *Plot Variable* list (Fig. 2-17).
3. Click **OK**.

Statistix 8 produces a normal probability plot, which includes a Shapiro-Wilks Test (Fig. 2-18).

The results of the normal probability plot suggest that the data conforms to a normal distribution based on the apparent straightness of the line. Furthermore, since the Shapiro-Wilk test for the W statistic (underlined in red) is approaching 1 (0.9330) for normally distributed data and the P value of 0.3026 is greater than p= 0.05, then we conclude that the data meets the assumption of normality.

There are cases when data may not be normally distributed. When this occurs, one may need to increase sample size, transform data or analyze the data using non-parametric statistical analyses.

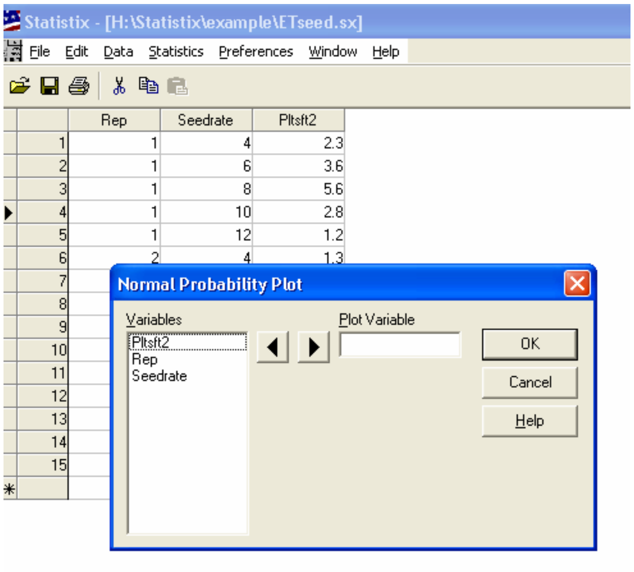


Figure 2-16

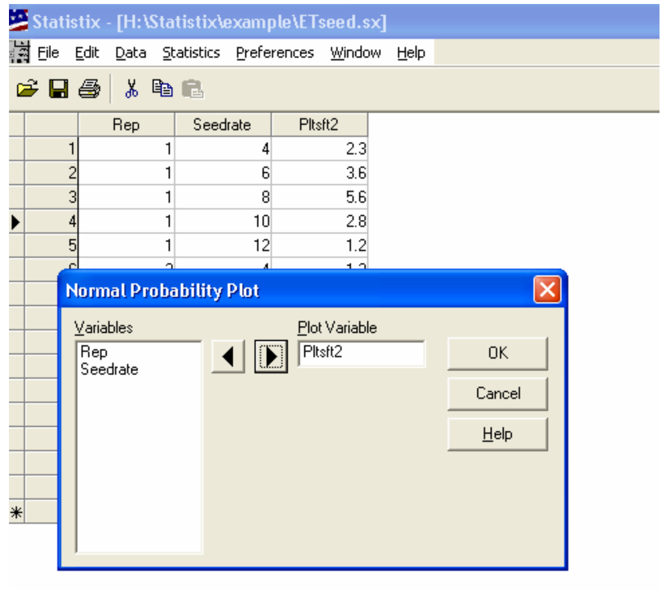


Figure 2-17

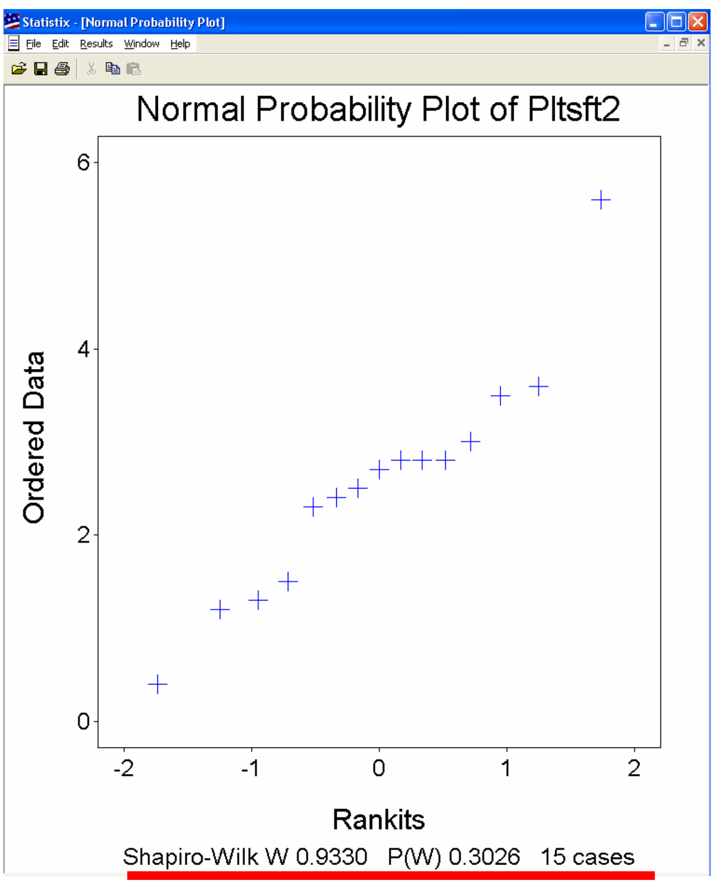


Figure 2-18

Chapter 3: Analysis of Variance

I. Randomized Complete Block Design

You have been collecting yield data on 5 switchgrass cultivars that were planted in a randomized complete block design with four replications. Two harvests were made in 1995 ("Yield1" and "Yield2") and 1996, respectively. The two harvests were summed ("Yield1"+"Yield2") for the season total yield in 1995 and 1996 ("Total"). Data was entered into the *Statistix 8* spreadsheet (Fig. 3-1). You wish to analyze the yield data to determine if there are any statistically significant differences between cultivars at the 5% level of probability, or p = 0.05.

A. Analyzing the 1995 Switchgrass Cultivar Data

In your first data analyses you want to analyze the 1995 cultivar yields ("Yield1", "Yield2" and "Total"). Since the 1996 data is included in the dataset you will need to disregard or omit the data in cases 21-40 so they will not be included in the analyses.

To omit data:

1. Click on the empty beige cell to the left of the cell marked with the fist case number that you want to delete. In this example, we selected the beige cell to the left of case "21".

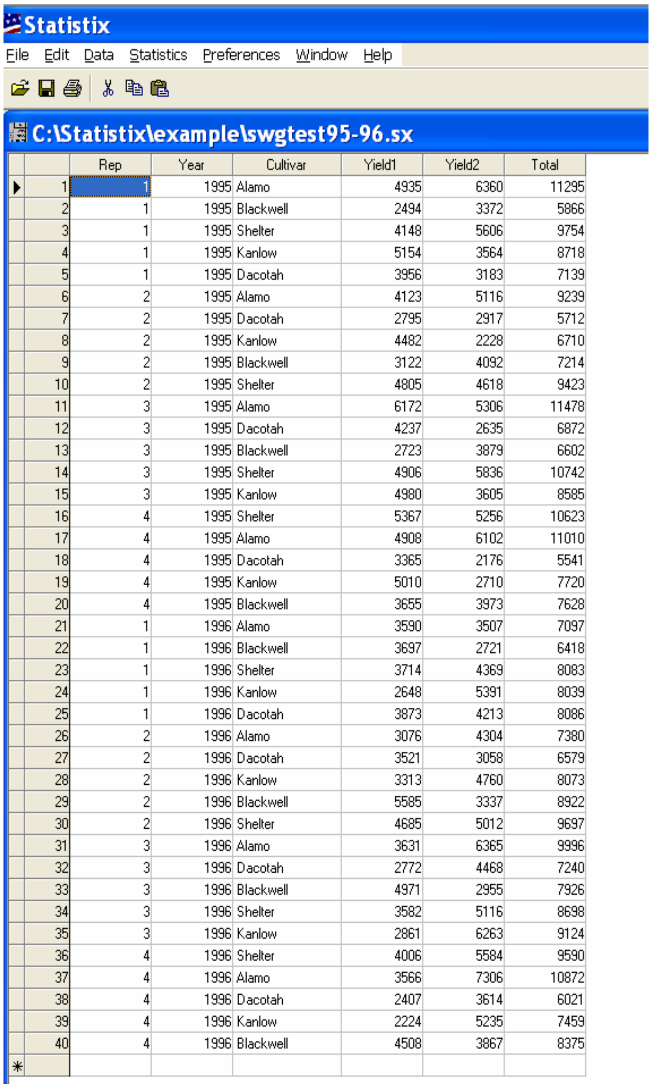
A triangle will appear in the beige cell to the left of the case number, and the entire row for the case will be highlighted.

2. Click on the left mouse button and hold, and drag the blue highlighted area to the row with the last case that you want to omit. In this example, drag to the row for case "40."

The rows for cases 21 through 40 will be highlighted.

3. From the main menu, select **Edit, Omit Highlighted Cases** (Fig. 3-2)**.**

Cases 21-40 will fade to a gray tint indicating that they have been omitted from the dataset and will not be included in the analyses.



	Rep	Year	Cultivar	Yield1	Yield2	Total
1	1	1995	Alamo	4935	6360	11295
2	1	1995	Blackwell	2494	3372	5866
3	1	1995	Shelter	4148	5606	9754
4	1	1995	Kanlow	5154	3564	8718
5	1	1995	Dacotah	3956	3183	7139
6	2	1995	Alamo	4123	5116	9239
7	2	1995	Dacotah	2795	2917	5712
8	2	1995	Kanlow	4482	2228	6710
9	2	1995	Blackwell	3122	4092	7214
10	2	1995	Shelter	4805	4618	9423
11	3	1995	Alamo	6172	5306	11478
12	3	1995	Dacotah	4237	2635	6872
13	3	1995	Blackwell	2723	3879	6602
14	3	1995	Shelter	4906	5836	10742
15	3	1995	Kanlow	4980	3605	8585
16	4	1995	Shelter	5367	5256	10623
17	4	1995	Alamo	4908	6102	11010
18	4	1995	Dacotah	3365	2176	5541
19	4	1995	Kanlow	5010	2710	7720
20	4	1995	Blackwell	3655	3973	7628
21	1	1996	Alamo	3590	3507	7097
22	1	1996	Blackwell	3697	2721	6418
23	1	1996	Shelter	3714	4369	8083
24	1	1996	Kanlow	2648	5391	8039
25	1	1996	Dacotah	3873	4213	8086
26	2	1996	Alamo	3076	4304	7380
27	2	1996	Dacotah	3521	3058	6579
28	2	1996	Kanlow	3313	4760	8073
29	2	1996	Blackwell	5585	3337	8922
30	2	1996	Shelter	4685	5012	9697
31	3	1996	Alamo	3631	6365	9996
32	3	1996	Dacotah	2772	4468	7240
33	3	1996	Blackwell	4971	2955	7926
34	3	1996	Shelter	3582	5116	8698
35	3	1996	Kanlow	2861	6263	9124
36	4	1996	Shelter	4006	5584	9590
37	4	1996	Alamo	3566	7306	10872
38	4	1996	Dacotah	2407	3614	6021
39	4	1996	Kanlow	2224	5235	7459
40	4	1996	Blackwell	4508	3867	8375

Figure 3-1

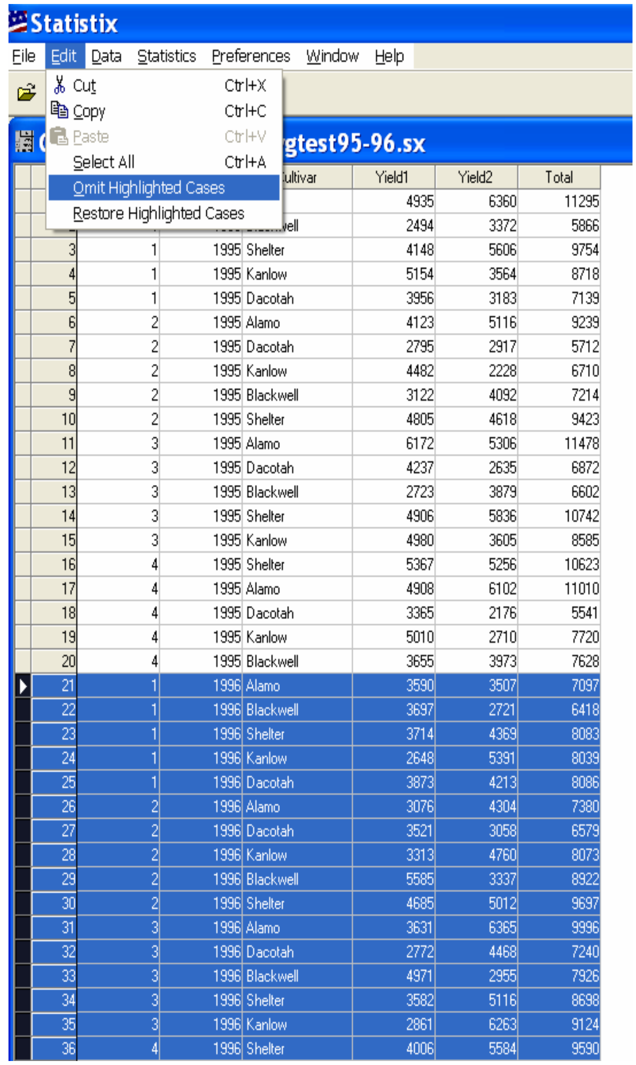


Figure 3-2

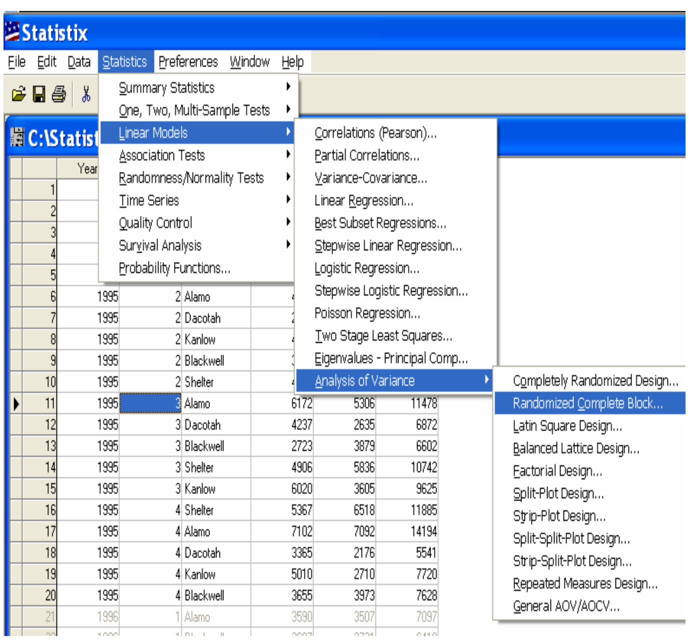


Figure 3-3

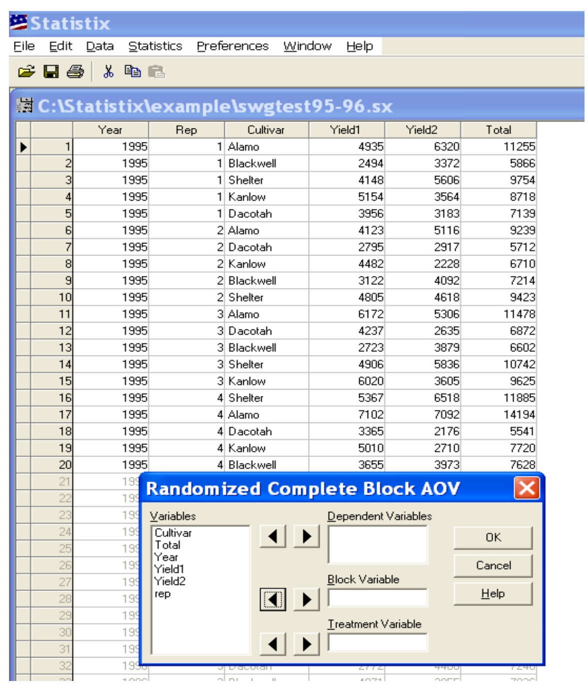


Figure 3-4

To analyze the 1995 data, which consists of cases 1-20:

1. From the main menu, select **Statistics, Linear Models, Analysis of Variance, Randomized Complete Block AOV** (Fig. 3-3).
2. In the ***Randomized Complete Block AOV*** dialog box, move the variables you want to designate as dependent variables from the *Variables* list in the left column to the *Dependent Variables* list in the right column by highlighting the variables (in this example, "Yield1", "Yield2", and "Total") and clicking on the right-arrow button to the left of the *Dependent Variables* list (Fig. 3-4).
3. Move the variables you want to designate as block variables (in this example, "rep") from the "Variables" list in the left column to the *Block Variable* list in the right column by highlighting the variable and clicking on the right-arrow button to the left of the *Block Variable* list.
4. Move the variables you want to designate as treatment variables (in this example, "cultivar") from the *Variables* list in the left column to the *Treatment Variable* list in the right column by highlighting the variable and clicking on the right-arrow button to the left of the *Treatment Variable* list (Fig. 3-5).
5. Click **OK**.
6. To return to the data set, from the main menu, select **File**, **Close.**

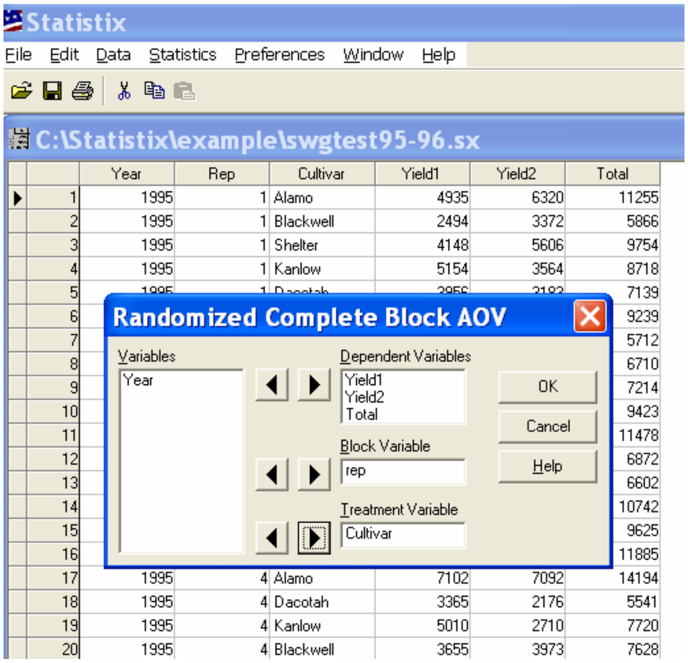


Figure 3-5

B. Results and Interpretation of the Analyses

The following statistical analysis (Table 3-1) was performed by *Statistix 8* on the 1995 data. Rather than discuss all of the numbers generated in the analysis of variance table, we will focus on numbers of greatest interest to the user for Yield 1. Numbers have been highlighted in red followed by an interpretation of the number and its significance to the user.

Table 3-1

Yield1

Randomized Complete Block AOV Table for YIELD1

Source	DF	SS	MS	F	P
REP	3	1645071	548357		
CULTIVAR	4	1.343E+07	3358870	11.41	0.0005
Error	12	3532694	294391		
Total	19	1.861E+07			

Since P value of 0.0005 is less than p= 0.05 then it has been determined that there is statistical differences between switchgrass cultivars for Yield1.

Grand Mean 4266.9 **CV 12.72**

The grand mean was determined by adding the means of Yield1 for cultivars and dividing the total by 5 (5 cultivars).

The coefficient of variation (CV) indicates the degree of precision with which the treatments are compared and is a good indicator of reliability of the experiment. It expresses the experimental error as a percentage. The lower the CV value the greater the reliability of the experiment. CV values vary depending on crop grown, type of experiment and measurements.

Tukey's 1 Degree of Freedom Test for Nonadditivity

Source	DF	SS	MS	F	P
Nonadditivity	1	174979	174979	0.57	0.4649
Remainder	11	3357715	305247		

Statistix 8 conducts a **Tukey's One Degree of Freedom Test for Nonadditivity** for a Randomized Complete Block (RCB) analysis of variance. This test is useful when the experimental design only permits an additive model to be fitted to the data and you suspect that interaction is present. In this example, there is no suggestion of nonadditivity (p=0.4649). If the P value is less than 0.05 then nonadditivity may be present and you should consider transforming the data in an effort to remove it.

Relative Efficiency, RCB 1.11

The purpose of using a RCB design is to reduce the error mean square. The relative efficiency indicates the magnitude to which blocking succeeded in reducing experimental error. In this example, the relative efficiency of using the RCB design over the completely randomized design is1.11 which is ~ 11 % increase in precision.

Means of YIELD1 for CULTIVAR

CULTIVAR	Mean
Alamo	5034.5
Blackwell	2998.5
Dacotah	3588.2
Kanlow	4906.5
Shelter	4806.5

Mean Yield1 for each cultivar was determined by adding the yields of each cultivar by replication and dividing the total by 4 (replications).

```
Observations per Mean          4
Standard Error of a Mean  271.29
Std Error (Diff of 2 Means) 383.66
```

Yield2 (**only the interpretation of the P value is provided for Yield 2**)

Randomized Complete Block AOV Table for YIELD2

Source	DF	SS	MS	F	P
REP	3	1087597	362532		
CULTIVAR	4	2.897E+07	7242894	28.33	0.0000
Error	12	3067458	255622		
Total	19	3.313E+07			

Since P value of 0.0000 is less than p= 0.05 then it has been determined that there is statistical difference between switchgrass cultivars for Yield2.

Grand Mean 4126.7   CV 12.25

Tukey's 1 Degree of Freedom Test for Nonadditivity

Source	DF	SS	MS	F	P
Nonadditivity	1	56135	56135	0.21	0.6595
Remainder	11	3011324	273757		

Relative Efficiency, RCB 1.04

Means of YIELD2 for CULTIVAR

CULTIVAR	Mean
Alamo	5721.0
Blackwell	3829.0
Dacotah	2727.8
Kanlow	3026.8
Shelter	5329.0

```
Observations per Mean          4
Standard Error of a Mean  252.80
Std Error (Diff of 2 Means) 357.51
```

<u>Total</u> Yield (**only the interpretation of the P value for total yield is provided**)

Randomized Complete Block AOV Table for TOTAL

Source	DF	SS	MS	F	P
REP	3	3952591	1317530		
CULTIVAR	4	6.238E+07	1.559E+07	27.59	0.0000
Error	12	6783563	565297		
Total	19	7.311E+07			

Since P value of 0.0000 is less than p= 0.05 then it has been determined that there are statistical differences between switchgrass cultivars for total yield.

Grand Mean 8393.6   CV 8.96

Tukey's 1 Degree of Freedom Test for Nonadditivity

Source	DF	SS	MS	F	P
Nonadditivity	1	1072457	1072457	2.07	0.1785
Remainder	11	5711106	519191		

Relative Efficiency, RCB 1.18

Means of TOTAL for CULTIVAR

CULTIVAR	Mean
Alamo	10756
Blackwell	6828
Dacotah	6316
Kanlow	7933
Shelter	10136

Observations per Mean 4
Standard Error of a Mean 375.93
Std Error (Diff of 2 Means) 531.65

C. Saving the Analysis of Variance Table

To save the analysis of variance table for future reference, follow the procedure outlined in "File Management" in Chapter 1, Section III, except *Statistix 8* will save as *Rich Text Files*. Thus, the filename will be given an ".rtf "extension.

To retrieve the analysis:

1. From the main menu, select to click on **File, View Text File**.
2. In the ***Open*** dialog box, navigate to the folder where the file was saved, and highlight the filename. Click **Open**.

To copy the file contents and paste them into an MS-Word document:

1. Highlight the text in the file you want to copy. Click on the *right* mouse button and select **Copy** from the shortcut menu.
2. Open an MS-Word document, click inside the document, and click on the *right* mouse button and select **Paste** from the shortcut menu.

D. Performing Mean Separation Test on Switchgrass Cultivars

In our example, we determined that there were significant differences between cultivars for "Yield1", "Yield 2", and season "Total" yield based on the analysis of variance, which was performed at the 5% level of probability. Now, we want to know which cultivars are significantly different. For our example, we will perform a mean separation test on the cultivars for "Yield1" only (Fig. 3-6).

1. From the main menu, select **Results, Multiple Comparison**, **All-Pairwise Comparison** (Fig. 3-7).
2. In the ***All-pairwise Comparison*** dialog box, select the comparison method of your choice from the *Comparison Method* group (Fig. 3-8). In this example, we select the *LSD* radio button to conduct the least significance difference (LSD) test.
3. Keep the default alpha level, which is 0.05 in the *Alpha* text box, or enter 0.01 or 0.10.
4. In the *Report Format* group, keep the default selection *Homogenous groups*.

Note: The homogenous groups format will give the means followed by letters (A, B, C, etc) which makes it easier to interpret whether the means are similar or different.

5. Click **OK.**

Statistix 8 performs a means separation test using the LSD test at the 5% level of probability (Fig. 3-9).

6. To return to the data set, from the main menu, select **File**, **Close.**

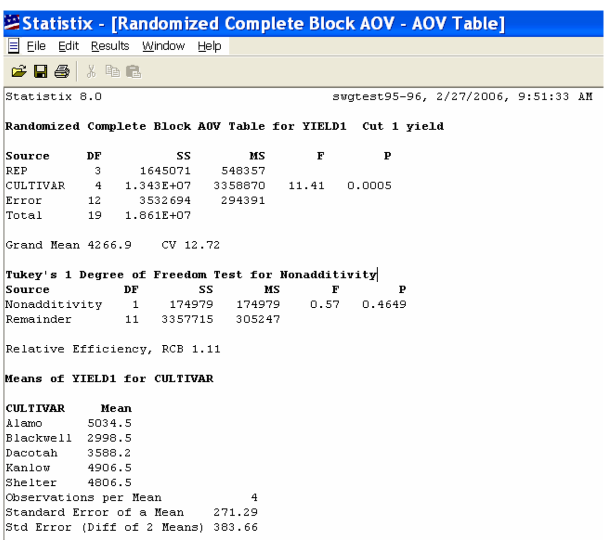


Figure 3-6

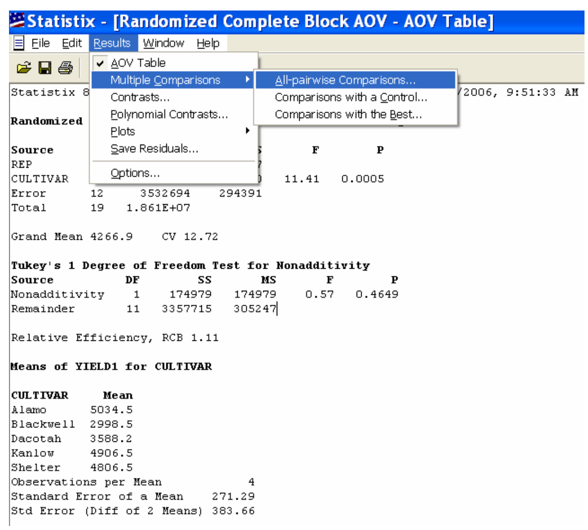


Figure 3-7

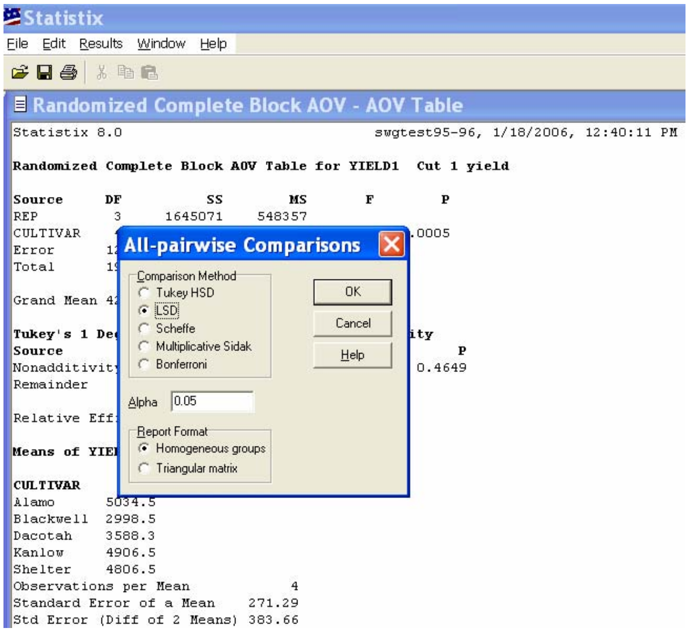


Figure 3-8

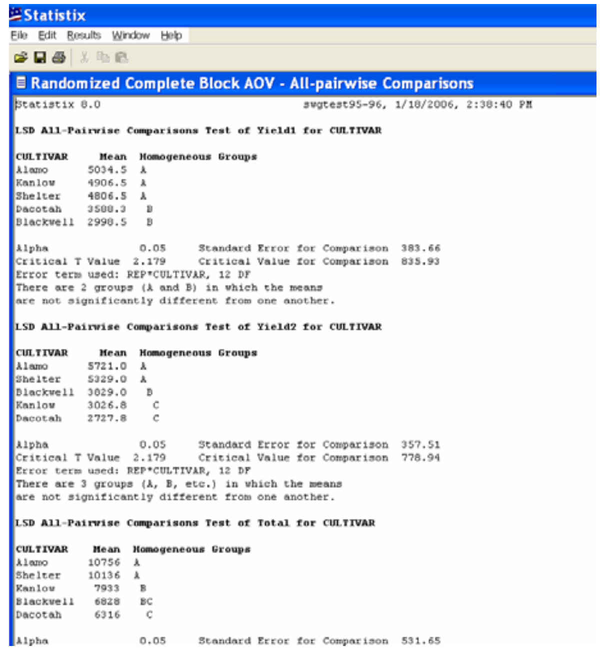


Figure 3-9

E. Saving the Mean Separation Test Results

To save the mean separation test for future reference follow the procedure outlined in Chapter 1-III File Management.

To retrieve the analysis or to copy the file contents and paste them into an MS-Word document, follow the procedure outlined in "Saving the Analysis of Variance Table" in Chapter 3-I-C Saving the Analysis of Variance Table.

F. General Interpretations and Presentation of Results

Table 3-2

Yield1

```
LSD All-Pairwise Comparisons Test of Yield1 for CULTIVAR

CULTIVAR      Mean  Homogeneous Groups
Alamo       5034.5  A
Kanlow      4906.5  A
Shelter     4806.5  A
Dacotah     3588.3   B
Blackwell   2998.5   B

Alpha               0.05     Standard Error for Comparison  383.66
Critical T Value   2.179     Critical Value for Comparison  835.93
Error term used: REP*CULTIVAR, 12 DF
There are 2 groups (A and B) in which the means are not significantly different
from one another.
```

Interpretation: In the first cutting, we found that Alamo, Kanlow, and Shelter switchgrass cultivars produced significantly higher (P<0.05) yields than Dacotah or Blackwell.

Yield2

```
LSD All-Pairwise Comparisons Test of Yield2 for CULTIVAR

CULTIVAR     Mean  Homogeneous Groups
Alamo      5721.0  A
Shelter    5329.0  A
Blackwell  3829.0   B
Kanlow     3026.8    C
Dacotah    2727.8    C

Alpha              0.05     Standard Error for Comparison  357.51
Critical T Value  2.179     Critical Value for Comparison  778.94
Error term used: REP*CULTIVAR, 12 DF
```

Interpretation: In the second cutting, Alamo and Shelter were the highest yielding cultivars. Their yields were significantly higher (P<0.05) than the other cultivars in the trial.

Total Yield

```
LSD All-Pairwise Comparisons Test of Total for CULTIVAR

CULTIVAR    Mean  Homogeneous Groups
Alamo      10756  A
Shelter    10136  A
Kanlow      7933   B
Blackwell   6828   BC
Dacotah     6316    C

Alpha              0.05     Standard Error for Comparison  531.65
Critical T Value  2.179     Critical Value for Comparison  1158.4
Error term used: REP*CULTIVAR, 12 DF
```

Interpretation: Season total yield of Alamo was higher than Shelter but the yield difference was not significant (P<0.05). However, Alamo and Shelter produced significantly (P<0.05) higher season total yield than Kanlow, Blackwell and Dacotah.

G. Using an LSD Value for Comparing Significant Means

Some users may prefer to use an LSD value for comparison of significant means rather than letters when presenting results in data tables (Fig. 3-10 and 3-11). See Gomez and Gomez (1984) for calculating LSD values.

H. Presentation of Results with letters and LSD value

Comparison of switchgrass cultivars **using letters** to denote differences in cultivars for yield.

	Harvest		
Cultivar	Yield 1	Yield 2	Season Total
	--------------------lb/acre----------------------		
Alamo	5035 a*	5721 a	10 756 a
Shelter	4907 a	5329 a	10 136 a
Kanlow	4807 a	3027 b	7933 b
Blackwell	2999 b	3829 c	6828 bc
Dacotah	3588 b	2728 c	6316 c
Mean	4267	4127	8394

Figure 3-10
* Means within column followed by the same letters are not significantly different as determined by least significant difference test at P<0.05.

Comparison of switchgrass cultivars **using LSD** value to determine yield differences between cultivars.

	Harvest		
Cultivar	Yield 1	Yield 2	Season Total
	---------------lb/acre----------------------		
Alamo	5035	5721	10 756
Shelter	4907	5329	10 136
Kanlow	4807	3027	7933
Blackwell	2999	3829	6828
Dacotah	3588	2728	6316
Mean	4267	4127	8394
LSD (0.05)*	836	779	1158

Figure 3-11
Least significant difference at P<0.05.

Note: Two treatment means are determined to be significantly different at a prescribed level of significance if their difference is greater than the LSD value; other wise they are not significant. For example, the difference between season total yield of Alamo (10 756 lb/acre) compared to Kanlow (7933 lb/acre) is 2823 lb/acre, which is greater than the computed LSD value of 1158 (lb/acre); thus, there is differences in yields between Alamo and Kanlow.

To analyze the 1996 switchgrass cultivar data, follow the procedure as the 1995 analysis.

I. Analyzing all of the Dataset

At the beginning of this example, you may remember that we were only interested in analyzing cultivar yields ("Yield1", "Yield2" and "Total") for 1995. Because the 1996 data was included in the original dataset we "omitted" the data in cases 21-40 so they would not be included in the analyses (Fig. 3-2 in Chapter 3-I-A Analyzing the 1995 Switchgrass Cultivar Data). Notice that the omitted cases have a gray tint (Fig. 3-12). Now, we want to include the entire dataset in the final analysis.

To restore the 1996 data, or cases 21 to 40:

1. Click on the beige cell to the left of the cell marked with the case number. In this example, we selected the beige cell to the left of "21".

A triangle will appear in the beige cell to the left of the case number. The entire row for the case will be highlighted.

2. Click on the left mouse button and hold, and drag the blue highlighted area to the row with the last case that you want to omit. In this example, drag to the row for case "40."

The rows for cases 21 through 40 will be highlighted.

3. From the main menu, select **Edit**, **Restore Highlighted Cases** (Fig. 3-13)**.**

Cases 21-40 will return to normal number tint indicating that they have been restored and can be included in the analyses (Fig. 3-14).

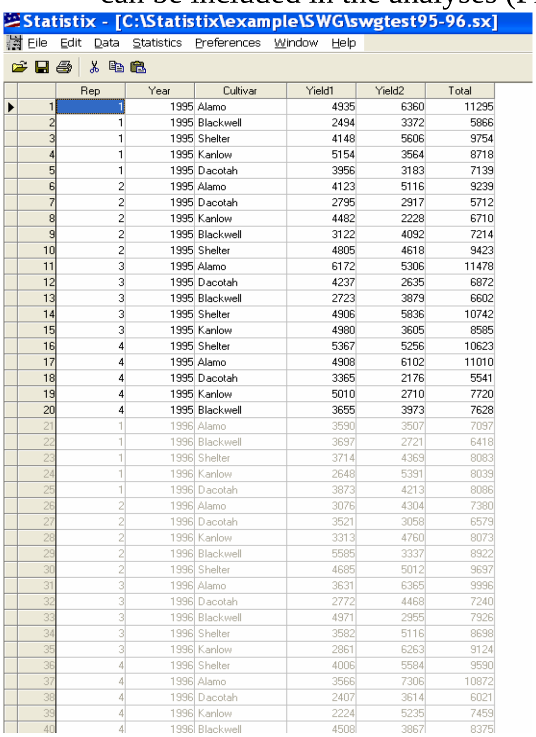


Figure 3-12

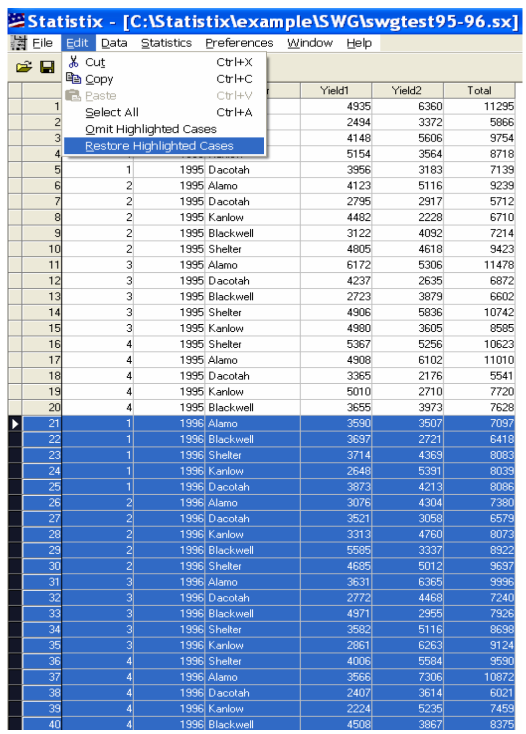
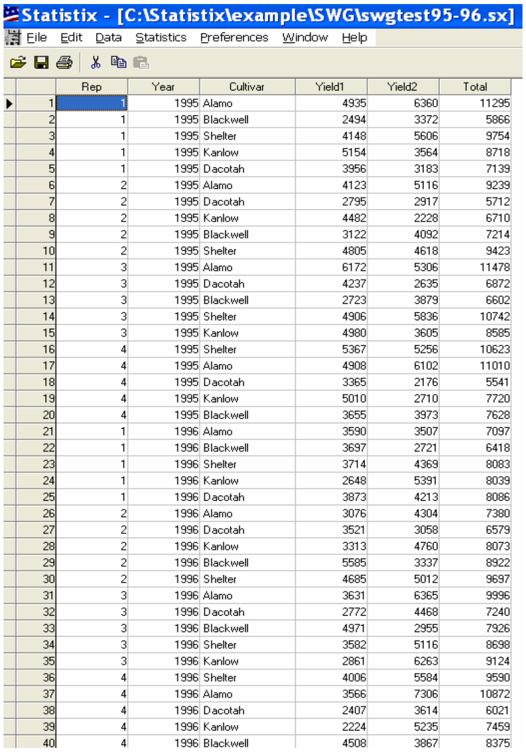


Figure 3-13



Statistix - [C:\Statistix\example\SWG\swgtest95-96.sx]

	Rep	Year	Cultivar	Yield1	Yield2	Total
1	1	1995	Alamo	4935	6360	11295
2	1	1995	Blackwell	2494	3372	5866
3	1	1995	Shelter	4148	5606	9754
4	1	1995	Kanlow	5154	3564	8718
5	1	1995	Dacotah	3956	3183	7139
6	2	1995	Alamo	4123	5116	9239
7	2	1995	Dacotah	2795	2917	5712
8	2	1995	Kanlow	4482	2228	6710
9	2	1995	Blackwell	3122	4092	7214
10	2	1995	Shelter	4805	4618	9423
11	3	1995	Alamo	6172	5306	11478
12	3	1995	Dacotah	4237	2635	6872
13	3	1995	Blackwell	2723	3879	6602
14	3	1995	Shelter	4906	5836	10742
15	3	1995	Kanlow	4980	3605	8585
16	4	1995	Shelter	5367	5256	10623
17	4	1995	Alamo	4908	6102	11010
18	4	1995	Dacotah	3365	2176	5541
19	4	1995	Kanlow	5010	2710	7720
20	4	1995	Blackwell	3655	3973	7628
21	1	1996	Alamo	3590	3507	7097
22	1	1996	Blackwell	3697	2721	6418
23	1	1996	Shelter	3714	4369	8083
24	1	1996	Kanlow	2648	5391	8039
25	1	1996	Dacotah	3873	4213	8086
26	2	1996	Alamo	3076	4304	7380
27	2	1996	Dacotah	3521	3058	6579
28	2	1996	Kanlow	3313	4760	8073
29	2	1996	Blackwell	5585	3337	8922
30	2	1996	Shelter	4685	5012	9697
31	3	1996	Alamo	3631	6365	9996
32	3	1996	Dacotah	2772	4468	7240
33	3	1996	Blackwell	4971	2955	7926
34	3	1996	Shelter	3582	5116	8698
35	3	1996	Kanlow	2861	6263	9124
36	4	1996	Shelter	4006	5584	9590
37	4	1996	Alamo	3566	7306	10872
38	4	1996	Dacotah	2407	3614	6021
39	4	1996	Kanlow	2224	5235	7459
40	4	1996	Blackwell	4508	3867	8375

Figure 3-14

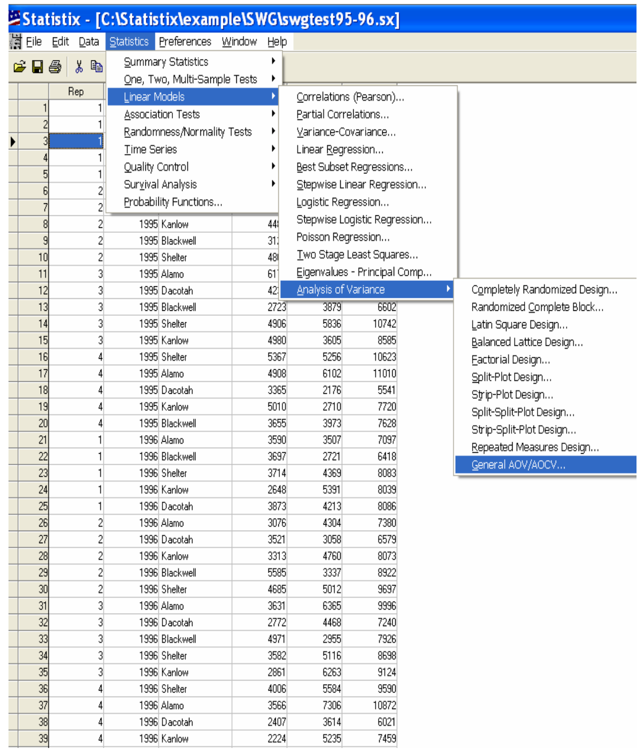


Figure 3-15

J. Including Year as a Variable

The variable "Year" can be included in the overall analyses if year (1995 or 1996) may have an effect on switchgrass cultivar yield; 1995 was considered a wet year and 1996 a dry year. *Statistix 8* allows the user to write math statements or models to analyze datasets containing multiple variables. In this case, we are interested in knowing if there is a significant "Year*Cultivar**"** interaction effect in "Total" yield, which may give insight on which cultivar(s) performed the best under wet or dry conditions.

To perform the analyses to determine the "Year*Cultivar**"** interaction effect in "Total" yield:

1. From the main menu, select **Statistics**, **Linear Models, Analysis of Variance, General AOV/AOCV** (Fig. 3-15).
2. In the ***General AOV/AOCV*** dialog box (Fig. 3-16), move the variables you want to designate as dependent variables from the *Variables* list in the left column to the *Dependent Variables* list in the right column by highlighting the variable and clicking on the right-arrow button to the left of the *Dependent Variables* list. In this example, we selected the variable "Total".
3. In the *AOV/Model Statemen*" text area, enter the statistical statement.

Example: "Rep Year Cultivar Year*Cultivar" without quotes (Fig. 3-17).

<u>Note</u>: The statement "Rep Year Cultivar Year*Cultivar" tells *Statistix 8* to perform an analysis of variance using the variables "Rep", "Year", "Cultivar", with the "Year* Cultivar" interaction effect as the sources of variation.

4. Click **OK**.

Statistix 8 will compute the error term for the model. The results of the analysis of variance (Table 3-3) is discussed below.

Note: Other model statements for general analysis of variance can be found in Chapter 7.

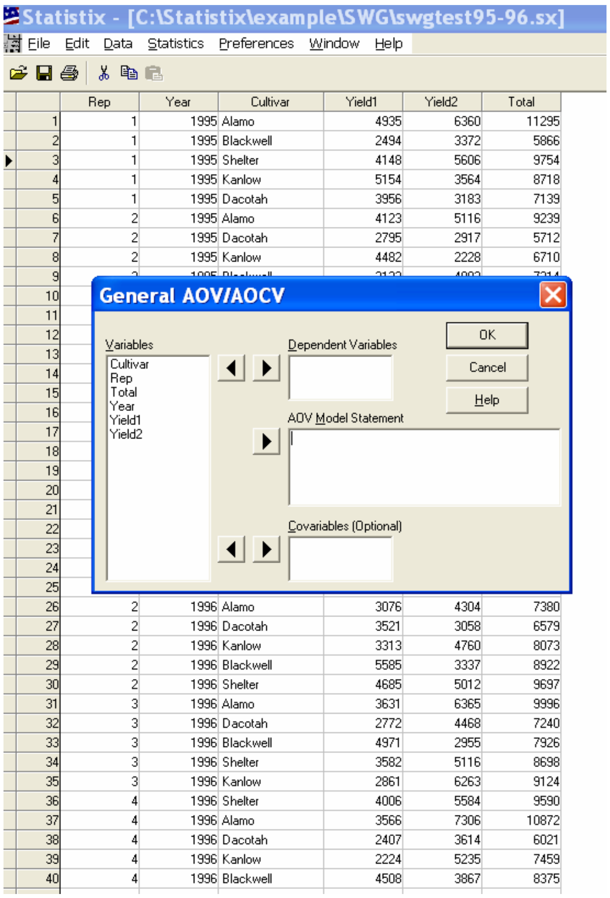


Figure 3-16

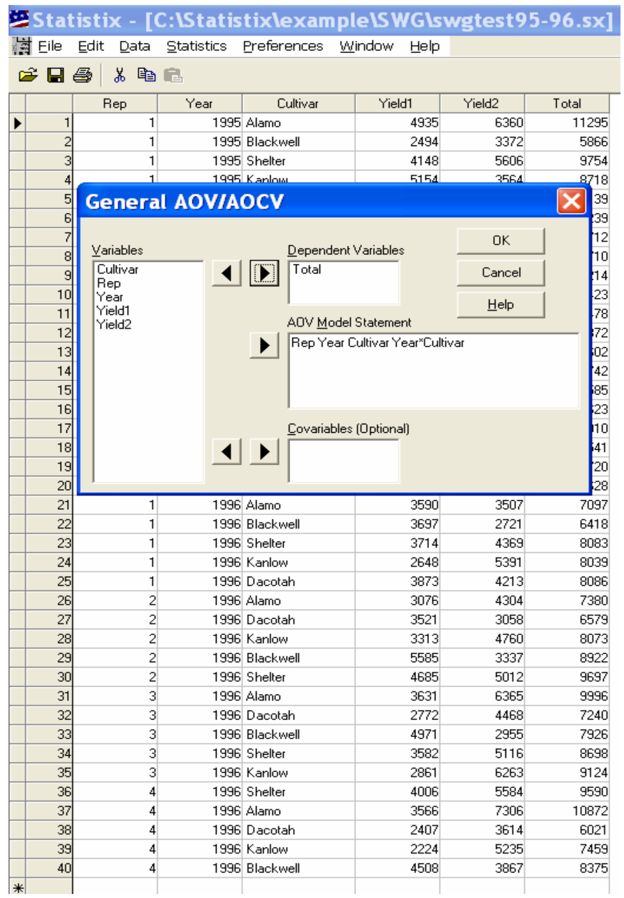


Figure 3-17

K. Interpretation of the Analysis

Table 3-3

```
Analysis of Variance Table for TOTAL

Source          DF        SS          MS        F        P
REP              3    4418919     1472973     1.53   0.2294
YEAR             1     440160      440160     0.46   0.5047
CULTIVAR         4  6.016E+07   1.504E+07    15.62   0.0000
YEAR*CULTIVAR    4  1.277E+07     3193780     3.32   0.0247
Error           27  2.600E+07      962859
Total           39  1.037E+08

Grand Mean 8288.7     CV 11.84
```

Interpretation: The analysis of variance determined that there were no significant differences in years (p=0.5047) but there were differences in cultivars (p=0.0000). Because there was a significant "Year*Cultivar" interaction effect (p=0.0247) we will ignore the cultivar differences and focus only on the interaction (Table 3-3).

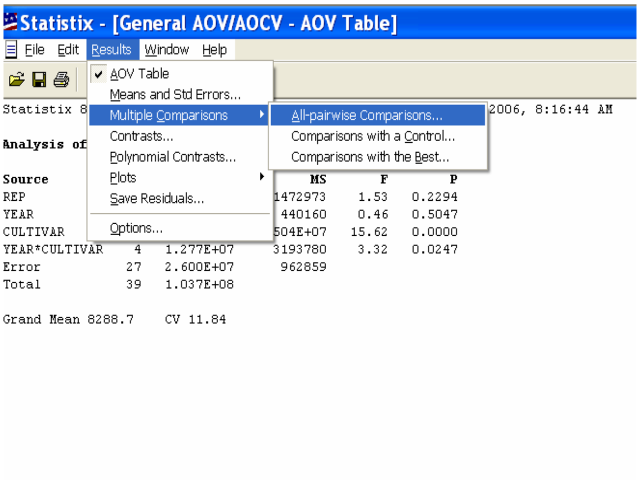


Figure 3-18

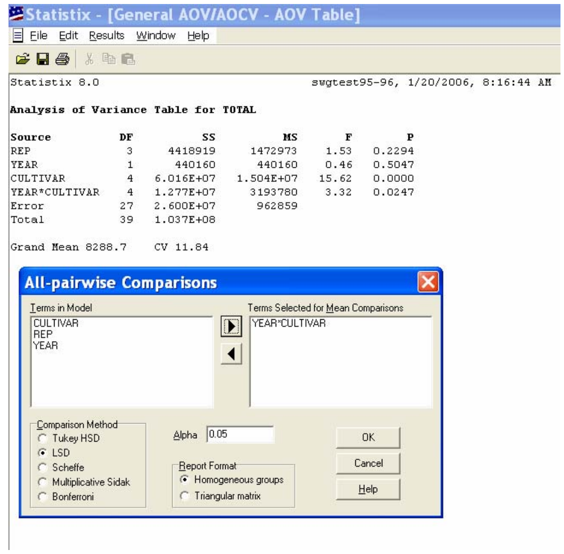


Figure 3-19

L. Comparing Means of "Year*Cultivar" Interaction

Statistix 8 can perform an All-Pairwise Comparison Test on the simple effects means of the "Year*Cultivar" interaction effect, which may help to explain the interaction.

1. Once the result of the analysis of variance table is in the active dialog box, from the main menu, select **Results, Multiple Comparison, All-Pairwise Comparison** (Fig. 3-18).
2. In the ***All-pairwise Comparison*** dialog box, highlight "Year*Cultivar" from the *Terms in Model* list in the left column, then press the right-arrow button to move it to the *Terms Selected for Mean Comparison* list in the right column(Fig. 3-19).
3. In the ***All-pairwise Comparison*** dialog box, select the preferred method in the *Comparison Method* group. In this example, we selected *LSD*.
4. Keep the default alpha level, which is 0.05, in the *Alpha* text box or enter 0.01 or 01.0.
5. In the *Report Format* group, keep the default section *Homogeneous groups*.
6. Once the information is in place, click on **OK.**

Note: The homogenous groups format will give the means followed by letters (A, B, C, etc) which makes it easier to interpret whether the means are similar or different and may aid in the explanation of the analysis.

Statistix 8 will perform a mean comparison test on the simple effect means (Table 3-4).

Table 3-4

```
LSD All-Pairwise Comparisons Test of Total for YEAR*CULTIVAR

YEAR CULTIVAR    Mean  Homogeneous Groups
1995 Alamo      10756  A
1995 Shelter    10136  AB
1996 Shelter     9017   BC
1996 Alamo       8836   BC
1996 Kanlow      8174    CD
1995 Kanlow      7933    CD
1996 Blackwell   7910    CD
1996 Dacotah     6982     DE
1995 Blackwell   6828     DE
1995 Dacotah     6316      E

Alpha              0.05     Standard Error for Comparison  693.85
Critical T Value  2.052     Critical Value for Comparison  1423.7
Error term used: REP*YEAR*CULTIVAR, 27 DF
There are 5 groups (A, B, etc.) in which the means
are not significantly different from one another.
```

M. Interpreting the Interaction

Earlier in the example, we stated that 1995 was a wet year and 1996 was a dry year. Because of the differences in moisture conditions we wanted to determine how the switchgrass cultivars would perform. Obviously, with the "Year*Cultivar" interaction there is indication that one or more of the cultivars responded to the different moisture conditions in 1995 and 1996.

Notice that *Statistix 8* ranks the performance of the cultivars based on high to low yields. Shelter, Kanlow, Blackwell and Dacotah performed similarly under both moisture regimes. Alamo experienced the greatest yield loss (1,920 lb/acre) when moisture became limited in 1996, and appears to have caused the significant "Year*Cultivar" interaction effect. Results suggest Alamo may require supplemental water to maximize yields in a dry year.

N. Graphing the Interaction Effect

Some may prefer to graph the simple effect means in order to have a better understanding of what may be causing the interaction. A graph of the "Year*Cultivar" interaction effect is illustrated in Fig. 3-20.

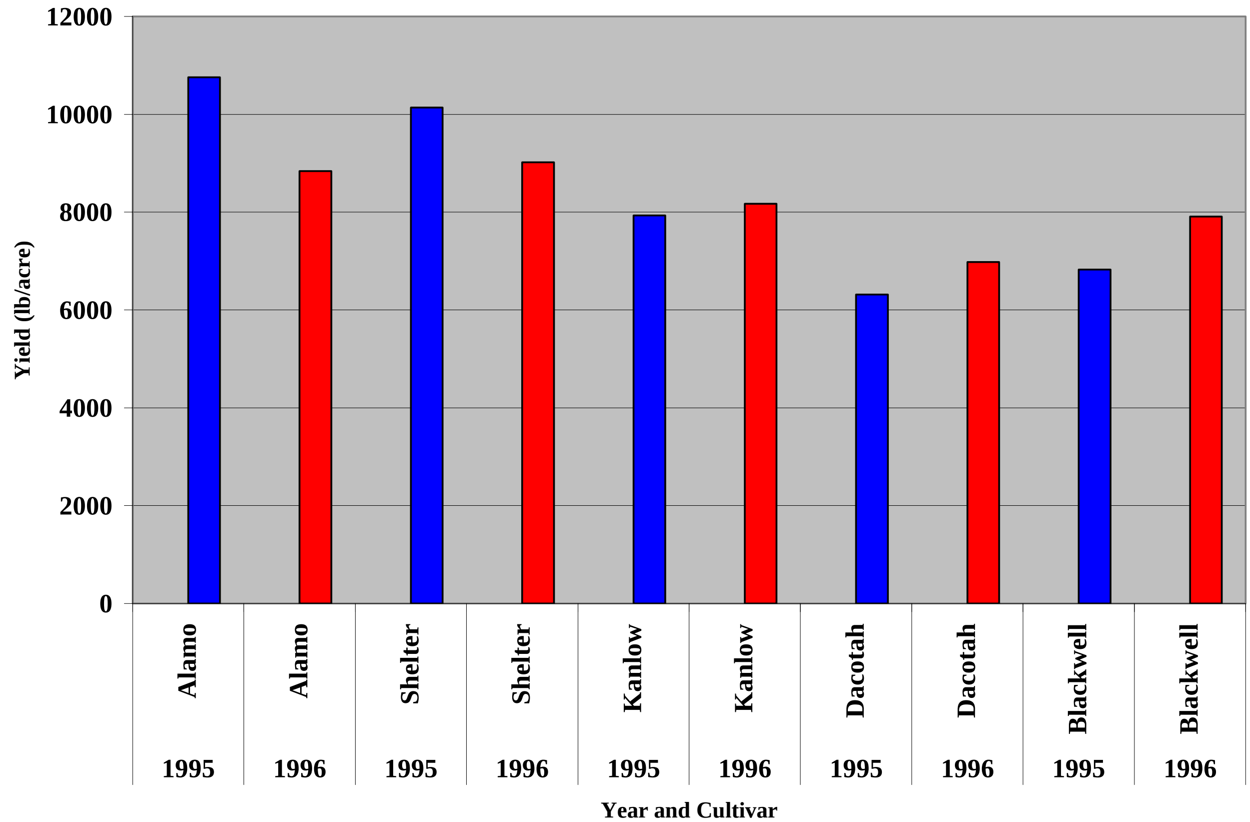


Figure 3-20

II. Completely Randomized Design

A. Analyzing Basin Wildrye Seed Germination Trial

Statistix 8 can analyze data from a completely randomized design using a one-way analysis of variance or by the completely randomized design analysis of variance model built into the program.

For this example, we will use germination data from four populations of basin wildrye replicated 10 times and arranged in a completely randomized design by randomly rotating the Petri dishes inside the germinator on a daily basis.

To determine if there are significant differences in the basin wildrye populations and identify the one(s) with superior germination:

1. Enter data into a *Statistix 8* spreadsheet (Fig. 3-21).
2. From the main menu, select **Statistics**, **Linear Models**, **Analysis of Variance**, **Completely Randomized Design** (Fig. 3-22).
3. In the ***Completely Randomized Design AOV*** dialog box (Fig. 3-23), move the chosen variable from the *Variables* list in the left column to the *Dependent Variables* list in the right column by highlighting the variable name ("Germ" in this example) and clicking on the right-arrow button to the left of the *Dependent Variables* list.

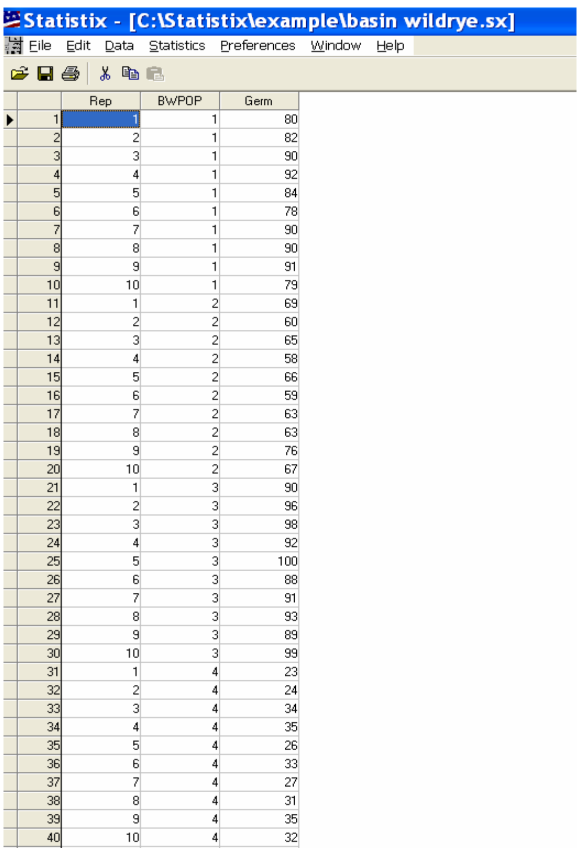
4. Move the chosen variable from the *Variables* list in the left column to the *Treatment Variable* list in the right column by highlighting the variable ("BWPOP" in this example) and clicking the right-arrow button to the left of the *Treatment Variable* list (Fig. 3-24)
5. Click **OK**

Statistix 8 performs the data analyses (Fig. 3-25).

6. To return to the data set, from the main menu, select **File**, **Close.**

B. Results and Interpretation of the Analyses

Statistix 8 performed the following Analysis of Variance for a completely randomized design. Rather than discuss all of the numbers generated in the analysis of variance, we will focus on numbers of greatest interest. Numbers have been highlighted in red followed by an interpretation of the number and its significance to the user (Table 3-5).



	Rep	BWPOP	Germ
1	1	1	80
2	2	1	82
3	3	1	90
4	4	1	92
5	5	1	84
6	6	1	78
7	7	1	90
8	8	1	90
9	9	1	91
10	10	1	79
11	1	2	69
12	2	2	60
13	3	2	65
14	4	2	58
15	5	2	66
16	6	2	59
17	7	2	63
18	8	2	63
19	9	2	76
20	10	2	67
21	1	3	90
22	2	3	96
23	3	3	98
24	4	3	92
25	5	3	100
26	6	3	88
27	7	3	91
28	8	3	93
29	9	3	89
30	10	3	99
31	1	4	23
32	2	4	24
33	3	4	34
34	4	4	35
35	5	4	26
36	6	4	33
37	7	4	27
38	8	4	31
39	9	4	35
40	10	4	32

Figure 3-21

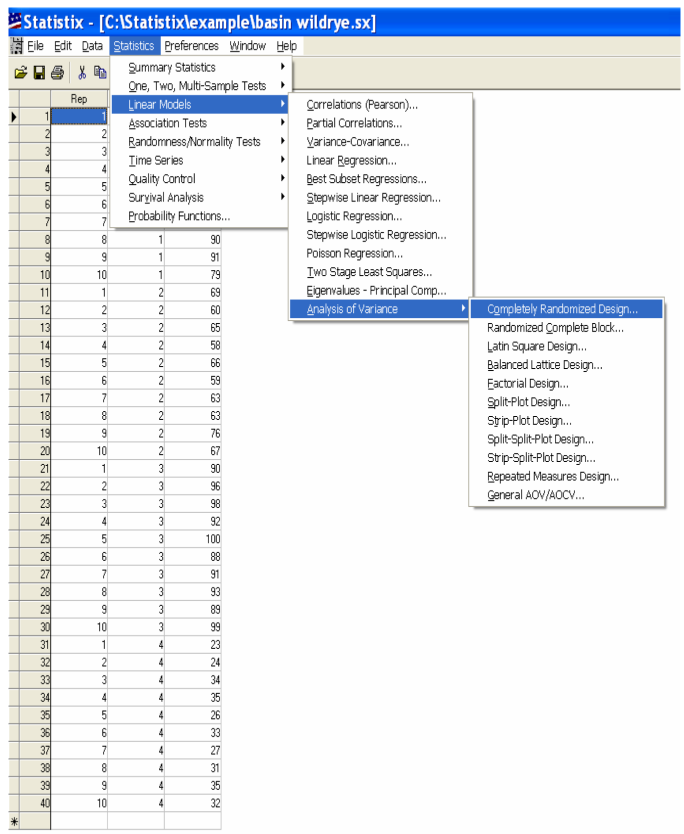


Figure 3-22

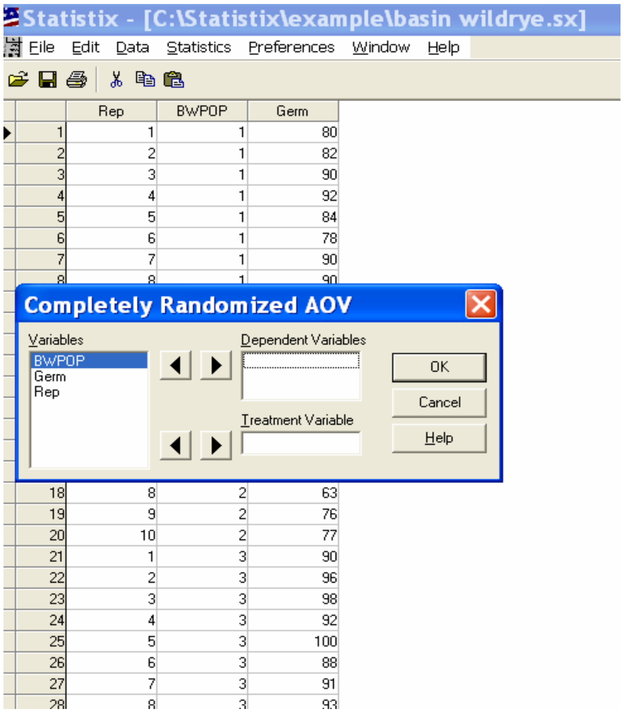


Figure 3-23

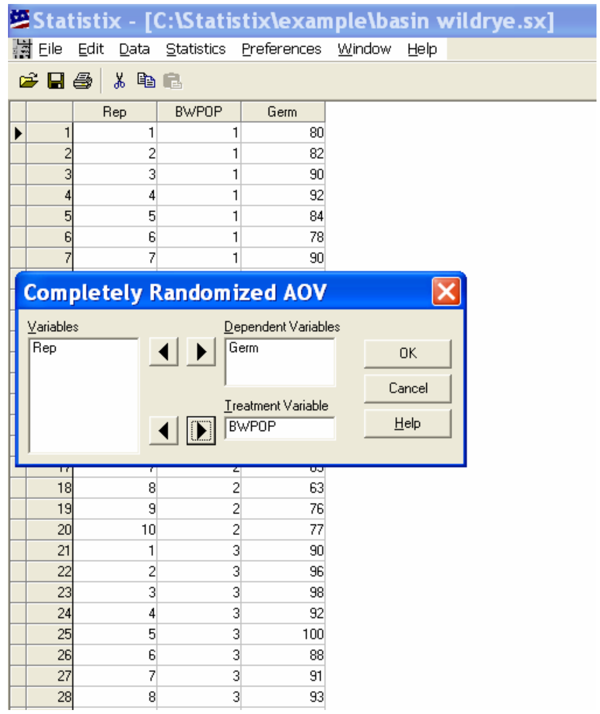
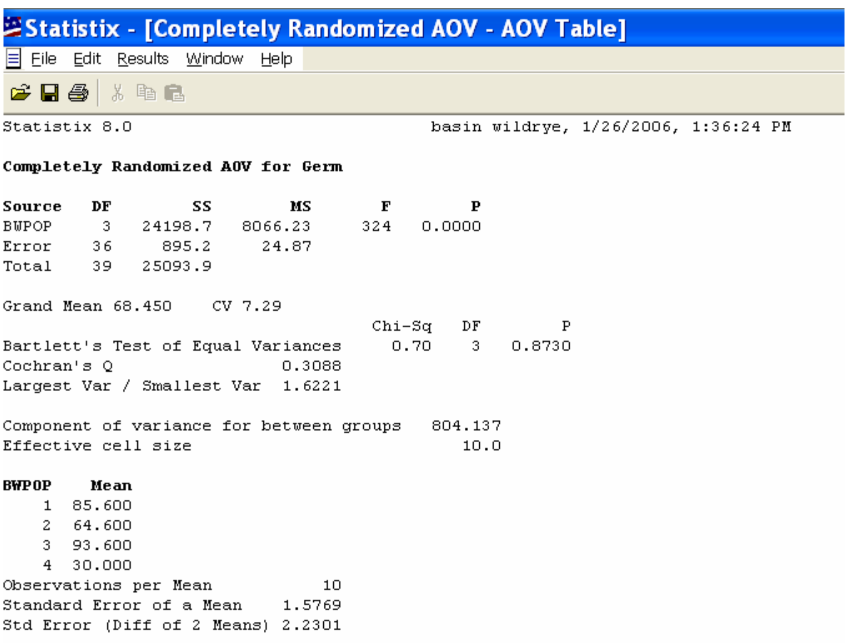


Figure 3-24




```
Statistix 8.0                              basin wildrye, 1/26/2006, 1:36:24 PM

Completely Randomized AOV for Germ

Source    DF       SS       MS       F       P
BWPOP      3  24198.7  8066.23     324  0.0000
Error     36    895.2    24.87
Total     39  25093.9

Grand Mean 68.450    CV 7.29
                                     Chi-Sq   DF        P
Bartlett's Test of Equal Variances     0.70    3   0.8730
Cochran's Q                      0.3088
Largest Var / Smallest Var       1.6221

Component of variance for between groups    804.137
Effective cell size                            10.0

BWPOP     Mean
    1   85.600
    2   64.600
    3   93.600
    4   30.000
Observations per Mean           10
Standard Error of a Mean    1.5769
Std Error (Diff of 2 Means) 2.2301
```

Figure 3-25

Table 3-5

```
Completely Randomized AOV for Germ

Source    DF        SS        MS       F        P
BWPOP      3   24198.7   8066.23     324   0.0000
Error     36     895.2     24.87
Total     39   25093.9

Grand Mean 68.450    CV 7.29   Chi-Sq   DF      P
Bartlett's Test of Equal Variances     0.70    3   0.8730
Cochran's Q                      0.3088
Largest Var/Smallest Var   1.6221

Component of variance for between groups   804.137
Effective cell size                           10.0

BWPOP     Mean
    1   85.600
    2   64.600
    3   93.600
    4   30.000
```

Since P value of 0.0000 is less than p= 0.05 then it has been determined that there are statistical differences between basin wildrye population for germination.

Since P value 0.8730 is greater than p= 0.05 then Bartlett's test for equal variances is not significant meaning that the variances are equal and satisfies the assumption of the F test which assumes within group variances are the same.

Mean germination for each population of basin wildrye was determined by adding the germination of each population by replication and dividing the total by 4 (replications).

C. Performing Mean Separation Test on Basin Wildrye Populations

In this example, we determined that there were significant differences between basin wildrye populations for germination based on the analysis of variance which was performed at the 5% level of probability.

To determine which basin wildrye germination means are significant:

1. From the main menu, select **Results**, **Multiple Comparison**, **All-Pairwise Comparison** (Fig. 3-26).
2. In the ***All-Pairwise Comparison*** dialog box (Fig. 3-27), select the LSD (least significance difference test) option from the *Comparison Method* group.
3. Keep the default alpha level, which is 0.05, in the *Alpha* text box, or enter 0.01 or 0.10.
4. In the *Report Format* group, keep the default selection *Homogenous groups*.

Note: The homogenous groups format will give the means followed by letters (A, B, C, etc) which makes it easier to interpret whether the means are similar or different and may aid in the explanation of the analysis.

5. Click **OK.**

Statistix 8 performs a means separation test using the LSD test at the 5% level of probability (Fig. 3-28).

6. To return to the data set, from the main menu, select **File**, **Close.**

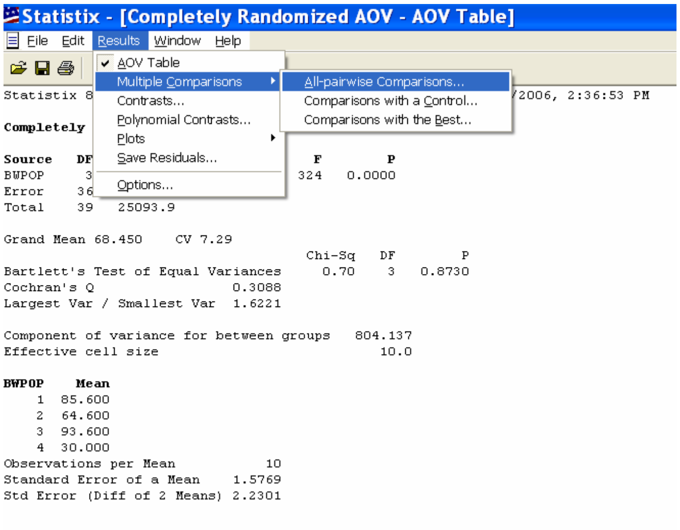


Figure 3-26

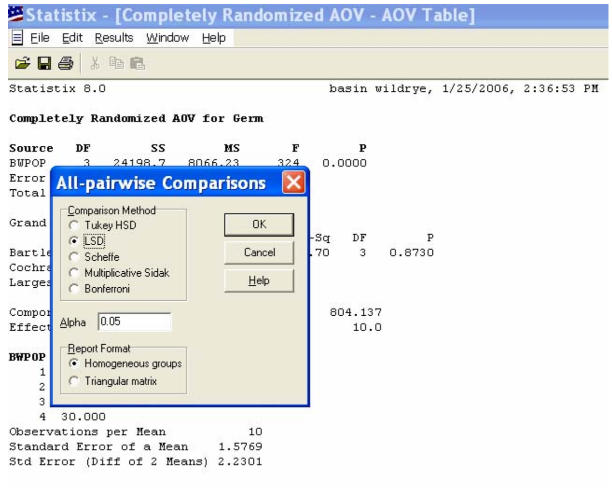


Figure 3-27

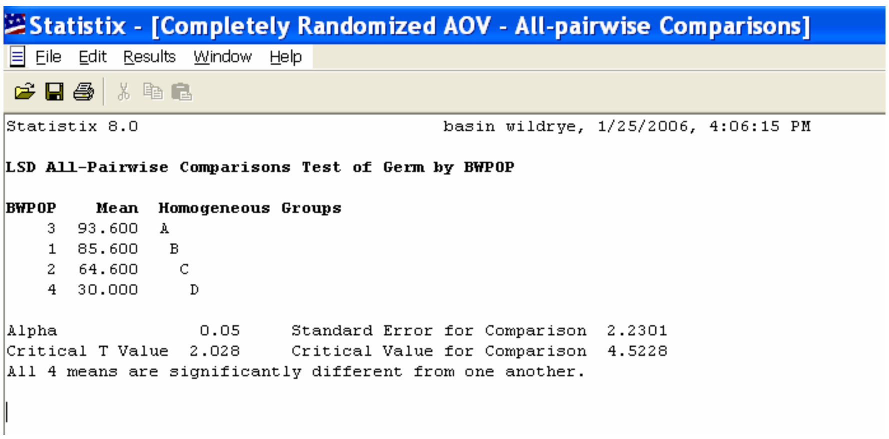


Figure 3-28

D. General Interpretations

Table 3-6

```
LSD All-Pairwise Comparisons Test of Germ by BWPOP

BWPOP    Mean  Homogeneous Groups
    3  93.600  A
    1  85.600   B
    2  64.600    C
    4  30.000     D

Alpha                 0.05     Standard Error for Comparison   2.2301

Critical T Value   2.028     Critical Value for Comparison   4.5228

All 4 means are significantly different from one another.
```

Interpretation: The analysis of variance determined that statistical differences exist in basin wildrye populations for germination characteristics. Using the LSD test, we found that population 3 had significantly higher germination ($P<0.05$) rate than the other populations.

III. Split Plot Design (Without Interaction Effect)

A. Analyzing Clipping Frequency of Eastern Gamagrass Cultivars

The split plot design is often used in agricultural research when two or more treatments are applied in factorial combinations. *Statistix 8* has the capability to perform an analysis of variance from data collected from studies arranged as a split plot design.

For this example, we will evaluate the effects of clipping frequency on yield of eastern gamagrass cultivars. Clipping frequency of 30, 45, and 60 days were assigned to the main plot and cultivars of Highlander, Jackson, Medina, and San Marcos to the subplots. The total yield by clipping frequency and cultivar was entered into *Statistix 8* spreadsheet (Fig. 3-29).

To analyze the data:

1. From the main menu, select **Statistics**, **Linear Models**, **Analysis of Variance**, **Split-Plot Design** (Fig. 3-30).
2. In the ***Split Plot Design AOV*** dialog box (Fig. 3-31), move the variable “Yld” from the *Variables* list in the left column the to the *Dependent Variables* list in the right column by highlighting “Yld” and clicking on the right-arrow button to the left of *Dependent Variables* list.
3. Move the variable “Rep” from the *Variables* list in the left column to the *Replication Variable* list in the right column by highlighting “Rep” and clicking on the right-arrow button to the left of the *Replication Variable* list.
4. Move the variable “Clipping” from the *Variables* list in the left column to the *Main-Plot Factor* list in the right column by highlighting “Clipping” and clicking on the right-arrow button to the left of the *Main-Plot Factor* list.
5. Move the variable “Cultivars” from the *Variables* list in the left column to the *Subplot Factor* list in the right column by highlighting “Cultivars” and clicking on the right-arrow button to the left of the *Subplot Factor* list (Fig. 3-32).

Note: For a greater degree of precision for Factor B than for Factor A, assign B to the subplot and Factor A to the main-plot. In our example, we are more interested in the cultivars than the clipping frequency so cultivars were assigned to the subplot and clipping frequency to main-plot.

6. Click **OK**.

Statistix 8 performs the data analyses.

7. To return to the data set, from the main menu, select **File**, **Close.**

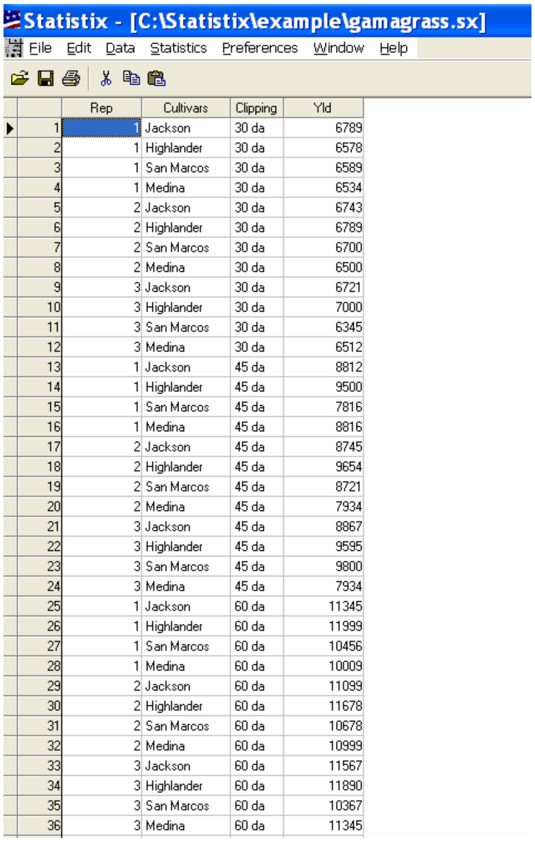


Figure 3-29

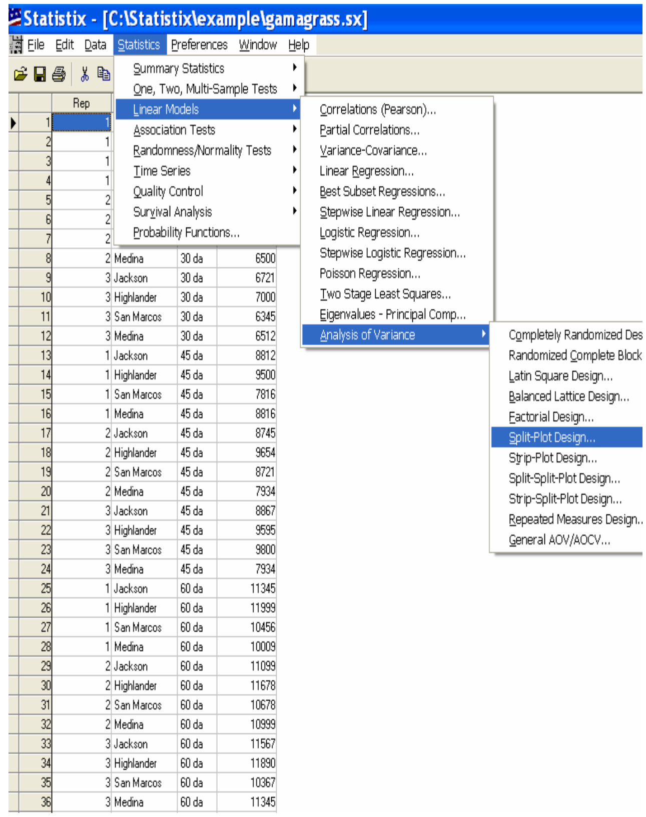


Figure 3-30

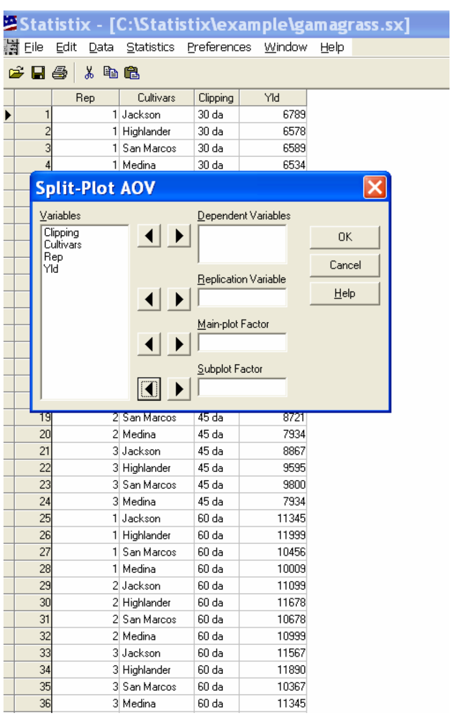


Figure 3-31

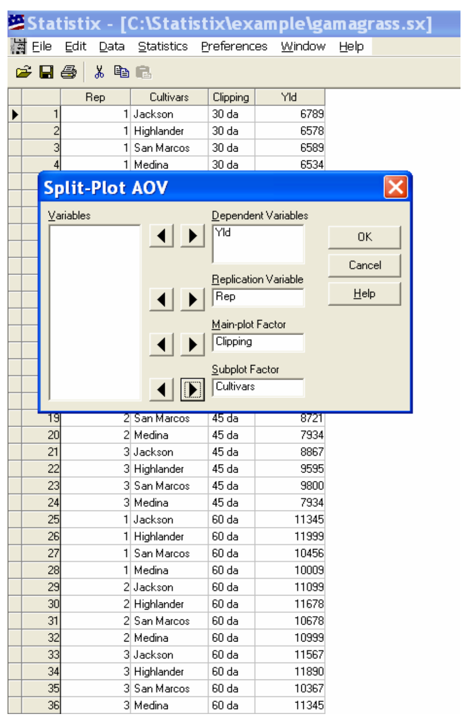
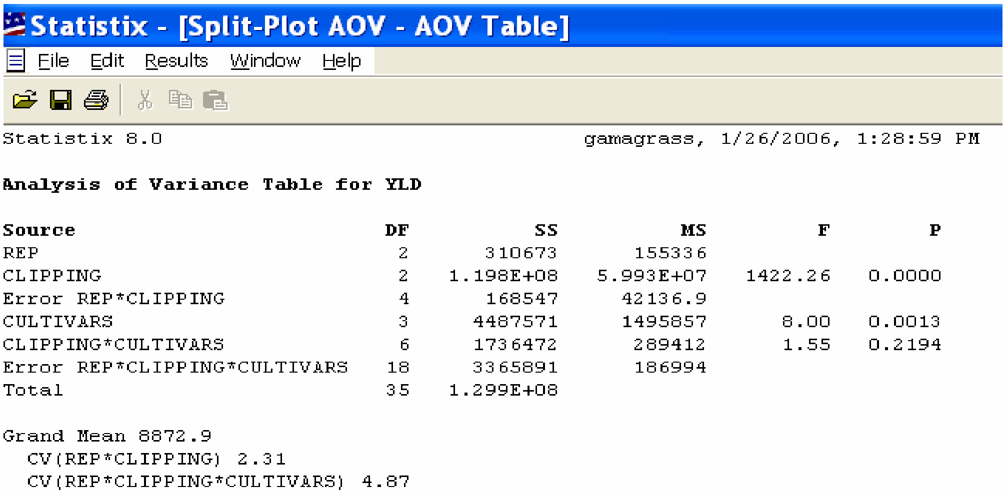


Figure 3-32

B. Results and Interpretation of the Analyses

The following statistical analysis was performed by *Statistix 8* (Fig. 3-33). Rather than discuss all of the numbers generated in the analysis of variance, we will focus on numbers of greatest interest to the user. Interested numbers have been highlighted in red followed by an interpretation of the number and its significance to the user (Table 3-7).


Statistix - [Split-Plot AOV - AOV Table]

File Edit Results Window Help

Statistix 8.0 gamagrass, 1/26/2006, 1:28:59 PM

Analysis of Variance Table for YLD

Source	DF	SS	MS	F	P
REP	2	310673	155336		
CLIPPING	2	1.198E+08	5.993E+07	1422.26	0.0000
Error REP*CLIPPING	4	168547	42136.9		
CULTIVARS	3	4487571	1495857	8.00	0.0013
CLIPPING*CULTIVARS	6	1736472	289412	1.55	0.2194
Error REP*CLIPPING*CULTIVARS	18	3365891	186994		
Total	35	1.299E+08			

Grand Mean 8872.9
CV(REP*CLIPPING) 2.31
CV(REP*CLIPPING*CULTIVARS) 4.87

Figure 3-33

Table 3-7

Analysis of Variance Table for YLD

Source	DF	SS	MS	F	P
REP	2	310673	155336		
CLIPPING	2	1.198E+08	5.993E+07	1422.26	**0.0000**

Since P value of 0.0000 is less than p= 0.05 then it has been determined that there are significant differences between clipping frequency for yield.

Source	DF	SS	MS	F	P
*** Error REP*CLIPPING**	4	168547	42136.9		
CULTIVARS	3	4487571	1495857	8.00	**0.0013**

Since P value of 0.0013 is less than p= 0.05 then it has been determined that there are significant differences between cultivars for yield.

Source	DF	SS	MS	F	P
CLIPPING*CULTIVARS	**6**	**1736472**	**289412**	**1.55**	**0.2194**

Since P value of 0.2194 is greater than p= 0.05 then it has been determined that there is not a significant clipping * cultivar interaction effect for yield.

Source	DF	SS	MS	F	P
*** Error REP*CLIPPING*CULTIVARS**	18	3365891	186994		
Total	35	1.299E+08			

Grand Mean 8872.9
CV(REP*CLIPPING) 2.31
CV(REP*CLIPPING*CULTIVARS) 4.87

<u>Note</u>: Notice that the analysis of variance for a split plot uses two error terms, one for testing factor A and one for testing factor B.

C. Performing Mean Separation Test on Clipping and Cultivars

In this example, we determined that there were significant differences in clipping frequency and cultivars for yield based on the analysis of variance which was performed at the 5% level of probability.

To determine which clipping frequency and cultivars are significantly different:

1. From the main menu, select **Results**, **Multiple Comparison**, **All-Pairwise Comparison** (Fig. 3-34).
2. In the ***All-pairwise Comparison*** dialog box (Fig. 3-35), highlight "Clipping" and "Cultivars" in the *Terms in Model* list in the left column, and click on the right-arrow button to the left of the *Terms Selected for Mean Comparison* list in the right column (Fig. 3-36).
3. From the *Comparison Method* group, select *LSD* (least significance difference test).
4. Keep the default alpha level, which is 0.05, in the *Alpha* text box, or enter 0.01 or 0.10.
5. In the *Report Format* group, keep the default *Homogeneous groups*.

<u>Note</u>: The homogenous groups format will give the means followed by letters (A, B, C, etc) which makes it easier to interpret whether the means are similar or different and may aid in the explanation of the analysis.

6. Click **OK.**

Statistix 8 performs a means separation test on "Clipping" and "Cultivars" using the LSD test at the 5% level of probability (Fig. 3-37, Table 3-8).

7. To return to the data set, from the main menu, select **File**, **Close.**

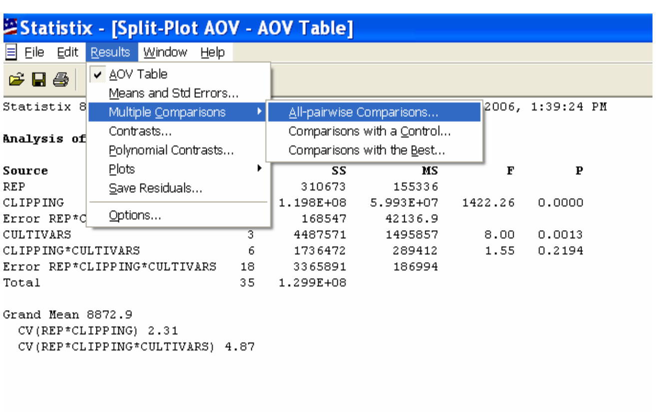


Figure 3-34

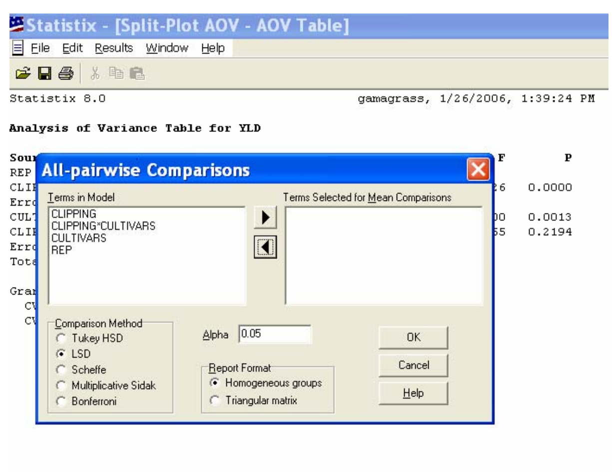


Figure 3-35

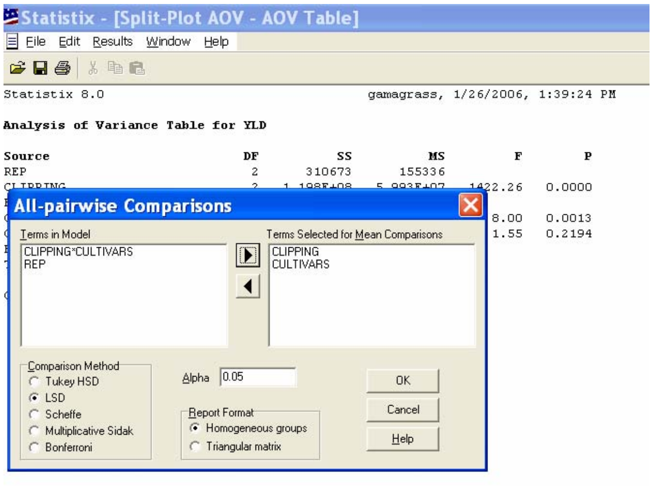


Figure 3-36

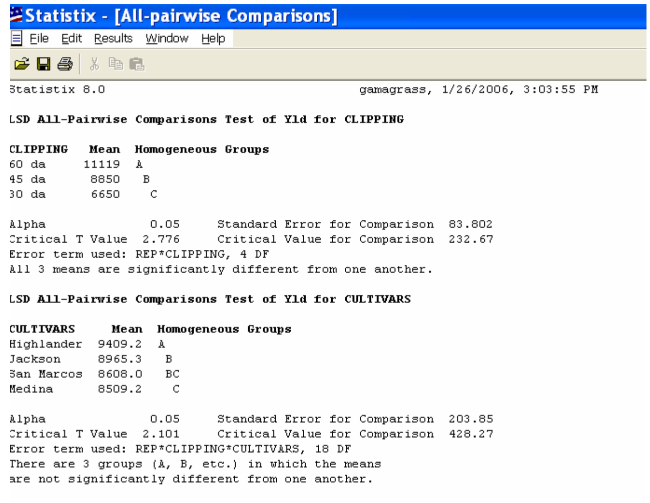


Figure 3-37

D. General Interpretations

Table 3-8

LSD All-Pairwise Comparisons Test of Yld for CLIPPING

CLIPPING	Mean	Homogeneous Groups
60 da	11119	A
45 da	8850	B
30 da	6650	C

Alpha 0.05 Standard Error for Comparison 83.802
Critical T Value 2.776 Critical Value for Comparison **232.67**
Error term used: REP*CLIPPING, 4 DF
All 3 means are significantly different from one another.

LSD All-Pairwise Comparisons Test of Yld for CULTIVARS

CULTIVARS	Mean	Homogeneous Groups
Highlander	9409.2	A
Jackson	8965.3	B
San Marcos	8608.0	BC
Medina	8509.2	C

Alpha 0.05 Standard Error for Comparison 203.85
Critical T Value 2.101 Critical Value for Comparison **428.27**
Error term used: REP*CLIPPING*CULTIVARS, 18 DF

IV. Split Plot Design (With Interaction Effect)

A. Analyzing Clipping Frequency in Eastern Gamagrass Cultivars

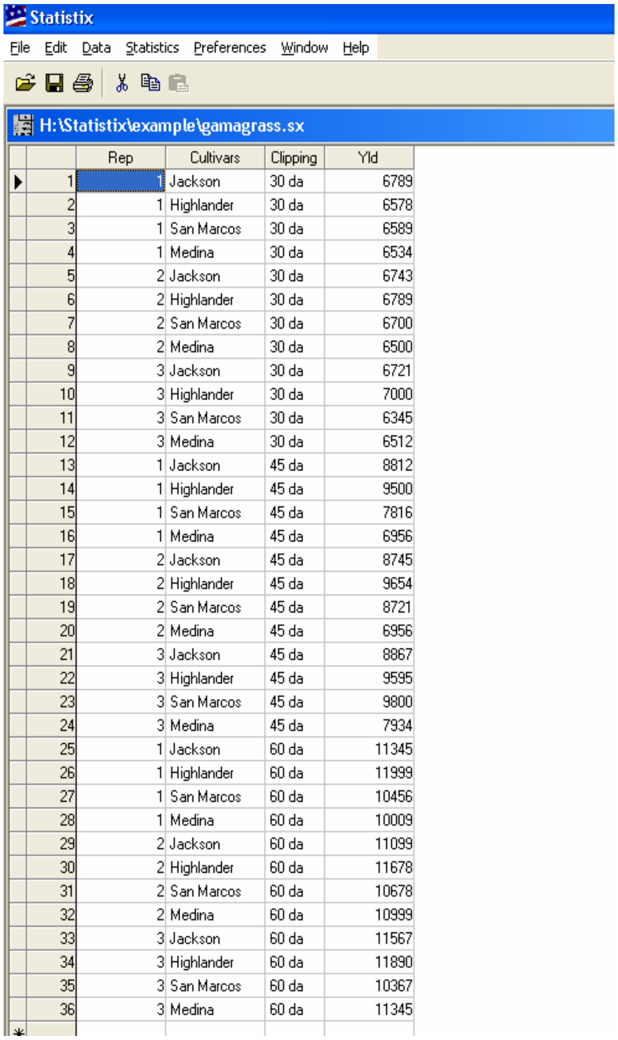
Anytime two or more factors are included in an experiment interaction between factors may occur. Experiments designed with two factors are conducted in anticipation of finding a significant interaction effect between factors; thus, providing a better understanding of how changing the level of one factor affects the change in another factor in the experiment.

For our example, we will use another clipping frequency and eastern gamagrass cultivar dataset (Fig. 3-38). To analyze the data:

1. From the main menu, select **Statistics**, **Linear Models**, **Analysis of Variance**, **Split-Plot Design** (Fig. 3-39).
2. In the ***Split Plot Design AOV*** dialog box (Fig. 3-40), move the variable "Yld" from the *Variables* list in the left column to the *Dependent Variables* list in the right column by highlighting "Yld" and clicking on the right-arrow button to the left of the *Dependent Variables* list.
3. Move the variable "Rep" from the *Variables* list in the left column to the *Replication Variable* list in right column by highlighting "Rep" and clicking on the right-arrow button to the left of the *Replication Variable* list.
4. Move the variable "Clipping" from the *Variables* list in the left column to the *Main-Plot Factor* list in the right column by highlighting "Clipping" and clicking on the right-arrow button to the left of the *Main-Plot Factor* list.
5. Move the variable "Cultivars" from the *Variables* list in the left column to the *Subplot Factor* list in the right column by highlighting "Cultivars" and clicking on the right-arrow button to the left of the *Subplot Factor* list (Fig. 3-41).
6. Click **OK**.

Statistix 8 performs the data analyses.

7. To return to the data set, from the main menu, select **File, Close.**



	Rep	Cultivars	Clipping	Yld
1	1	Jackson	30 da	6789
2	1	Highlander	30 da	6578
3	1	San Marcos	30 da	6589
4	1	Medina	30 da	6534
5	2	Jackson	30 da	6743
6	2	Highlander	30 da	6789
7	2	San Marcos	30 da	6700
8	2	Medina	30 da	6500
9	3	Jackson	30 da	6721
10	3	Highlander	30 da	7000
11	3	San Marcos	30 da	6345
12	3	Medina	30 da	6512
13	1	Jackson	45 da	8812
14	1	Highlander	45 da	9500
15	1	San Marcos	45 da	7816
16	1	Medina	45 da	6956
17	2	Jackson	45 da	8745
18	2	Highlander	45 da	9654
19	2	San Marcos	45 da	8721
20	2	Medina	45 da	6956
21	3	Jackson	45 da	8867
22	3	Highlander	45 da	9595
23	3	San Marcos	45 da	9800
24	3	Medina	45 da	7934
25	1	Jackson	60 da	11345
26	1	Highlander	60 da	11999
27	1	San Marcos	60 da	10456
28	1	Medina	60 da	10009
29	2	Jackson	60 da	11099
30	2	Highlander	60 da	11678
31	2	San Marcos	60 da	10678
32	2	Medina	60 da	10999
33	3	Jackson	60 da	11567
34	3	Highlander	60 da	11890
35	3	San Marcos	60 da	10367
36	3	Medina	60 da	11345

Figure 3-38

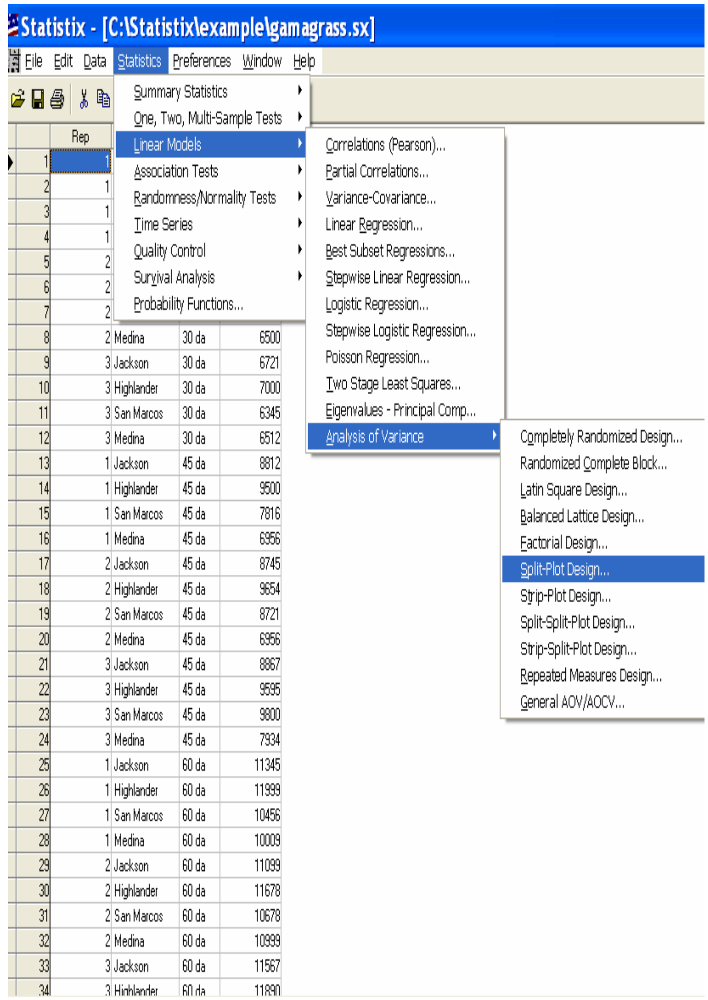


Figure 3-39

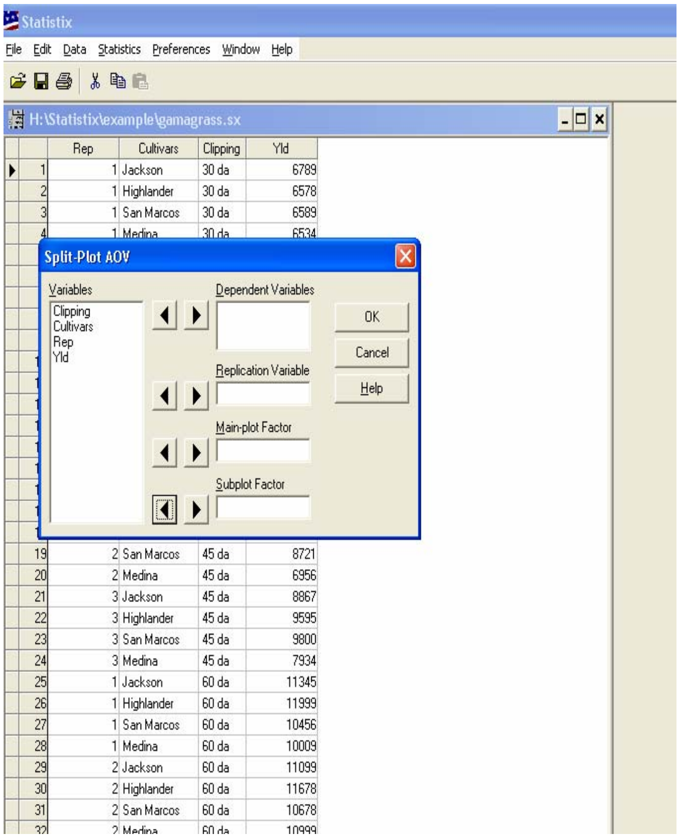


Figure 3-40

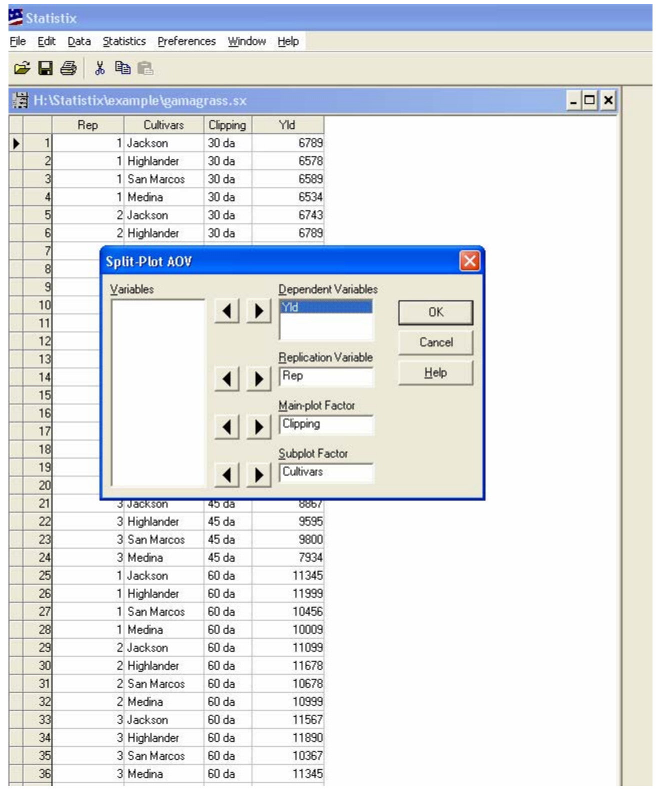
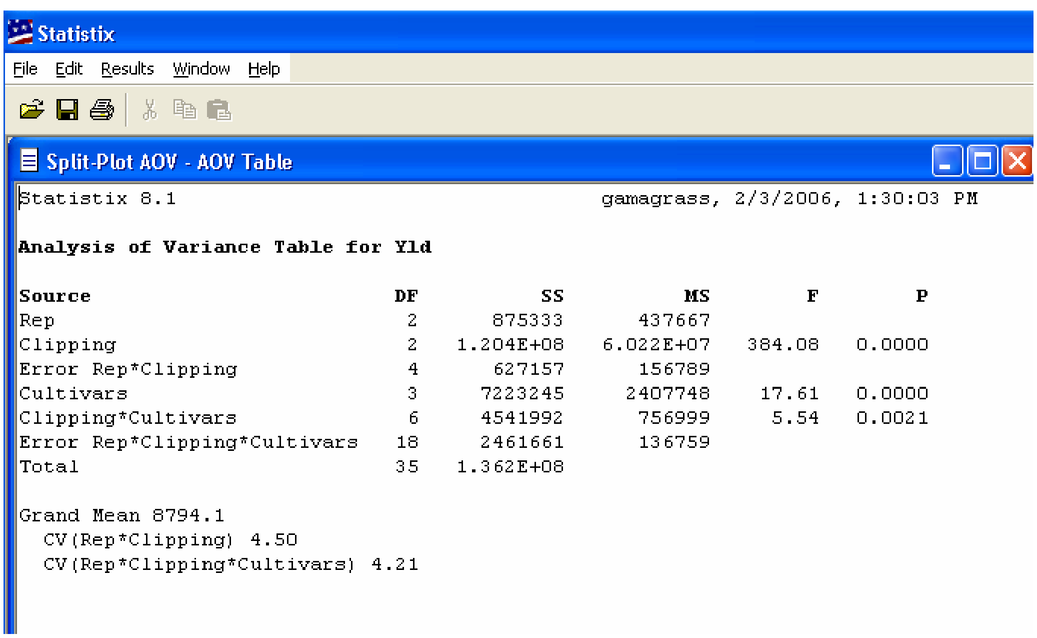


Figure 3-41

B. Results and Interpretation of the Analyses

The following statistical analysis was performed by *Statistix 8* (Fig. 3-42). Rather than discuss all of the numbers generated in the analysis of variance, we will focus on numbers of greatest interest to the user (Table 3-9). Interested numbers have been highlighted in red followed by an interpretation of the number and its significance to the user.



Statistix 8.1 gamagrass, 2/3/2006, 1:30:03 PM

Analysis of Variance Table for Yld

Source	DF	SS	MS	F	P
Rep	2	875333	437667		
Clipping	2	1.204E+08	6.022E+07	384.08	0.0000
Error Rep*Clipping	4	627157	156789		
Cultivars	3	7223245	2407748	17.61	0.0000
Clipping*Cultivars	6	4541992	756999	5.54	0.0021
Error Rep*Clipping*Cultivars	18	2461661	136759		
Total	35	1.362E+08			

Grand Mean 8794.1
CV(Rep*Clipping) 4.50
CV(Rep*Clipping*Cultivars) 4.21

Figure 3-42

Table 3-9

Analysis of Variance Table for Yld

Source	DF	SS	MS	F	P
Rep	2	875333	437667		
Clipping	2	1.204E+08	6.022E+07	384.08	**0.0000**
Error Rep*Clipping	4	627157	156789		
Cultivars	3	7223245	2407748	17.61	**0.0000**
Clipping*Cultivars	6	4541992	756999	5.54	**0.0021**
Error Rep*Clipping*Cultivars	18	2461661	136759		
Total	35	1.362E+08			

Since P value of 0.0000 is less than p= 0.05 then it has been determined that there are significant differences between clipping frequency for yield.

Since P value of 0.0000 is less than p= 0.05 then it has been determined that there are significant differences between cultivars for yield.

Since P value of 0.0021 is less than p= 0.05 then it has been determined that there is a significant clipping * cultivar interaction effect for yield.

Grand Mean 8794.1
CV(Rep*Clipping) 4.50
CV(Rep*Clipping*Cultivars) 4.21

Note: When a significant interaction effect is detected in the analysis of variance, the user will ignore the main effect means of the treatment variables and analyze the simple effects means of the interaction.

Interpretation: In reviewing the analysis of variance table we find that there are significant differences in "Clipping" and "Cultivars" for total yield**.** Because there was a significant "Clipping*Cultivar" interaction effect (p=0.0021), we will ignore the significant "Clipping" and "Cultivars" differences and focus on the "Clipping*Cultivar" interaction effect.

C. Comparing Means of Clipping *Cultivar Interaction

Statistix 8 performs an all-pairwise comparison Test on the simple effects means of the "Clipping*Cultivars" interaction, which may help to explain what may be causing the interaction.

To perform a mean comparison test on the simple effect means:

1. Make sure the result of the analysis of variance table is in the active window.
2. From the main menu, select **Results, Multiple Comparison, All-Pairwise Comparison** (Fig. 3-43).
3. In the ***All-pairwise Comparison*** dialog box, move the statement "Clipping*Cultivar" from the *Terms in Model* list in the left column to the *Terms Selected for Mean Comparison* in the right column by highlighting the statement and clicking on the right-arrow button to the left of the *Terms Selected for Mean Comparison* (Fig. 3-44).
4. In the *Comparison Method* group, select *LSD* or some other preferred method.
5. Keep the default alpha level, which is 0.05, in the *Alpha* text box, or enter 0.01 or 0.10.
6. In the *Report Format* group, select the default *Homogeneous groups*.

Note: The homogenous groups format will give the means followed by letters (A, B, C, etc). *The letters assigned to the simple effects means by the All Pairwise Comparison test have little or no impact on explaining the interaction.*

7.Click **OK.**

Statistix 8 performs a mean comparison test on the simple effect means (Table 3-10).

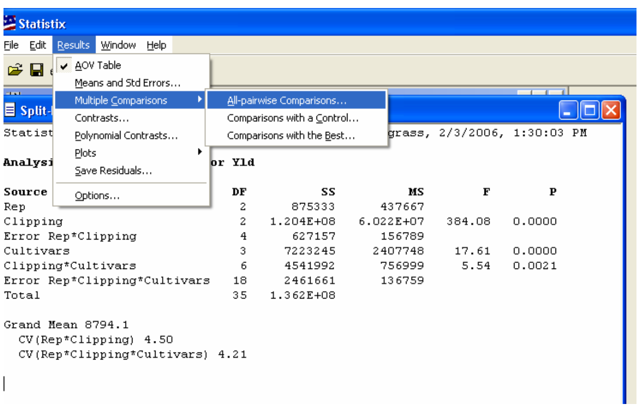


Figure 3-43

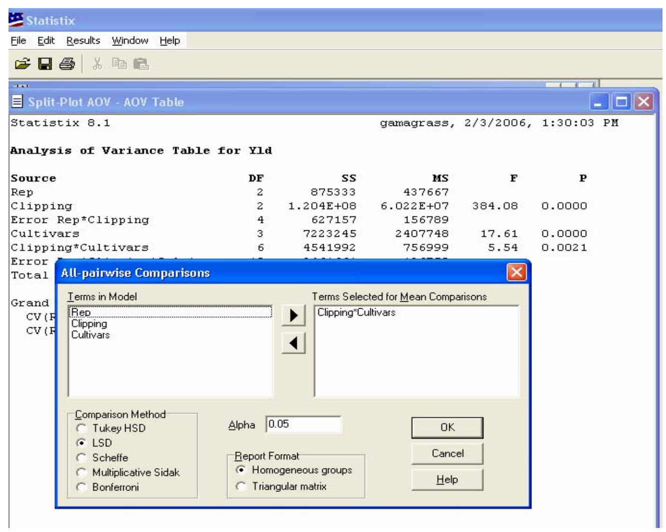


Figure 3-44

Table 3-10

```
LSD All-Pairwise Comparisons Test of Yld for Clipping*Cultivars

Clipping Cultivars    Mean  Homogeneous Groups
60 da    Highlander  11856  A
60 da    Jackson     11337  AB
60 da    Medina      10784   BC
60 da    San Marcos  10500    C
45 da    Highlander   9583     D
45 da    Jackson      8808      E
45 da    San Marcos   8779      E
45 da    Medina       7282       F
```

```
30 da    Highlander   6789       FG
30 da    Jackson      6751       FG
30 da    San Marcos   6545        G
30 da    Medina       6515        G

Comparisons of means for the same level of Clipping
  Alpha              0.05     Standard Error for Comparison  301.95
  Critical T Value    2.1     Critical Value for Comparison  634.37
  Error term used: Rep*Clipping*Cultivars, 18 DF
Comparisons of means for different levels of Clipping
  Alpha              0.05     Standard Error for Comparison  307.43
  Critical T Value    2.3     Critical Value for Comparison  703.30
  Error terms used: Rep*Clipping and Rep*Clipping*Cultivars
```

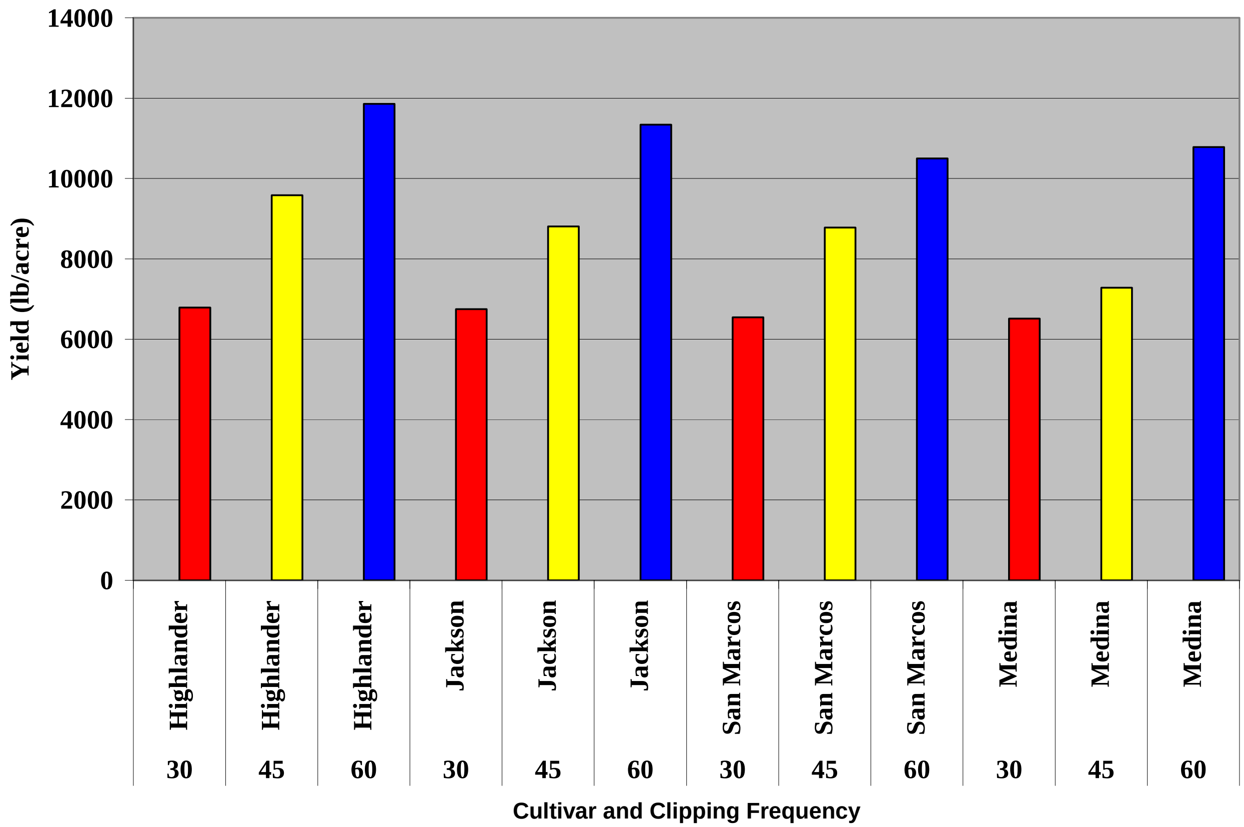


Figure 3-45

Chapter 4: Regression Analysis

I. Linear Regression

A. Eastern Gamagrass Nitrogen Rate Trial

Statistix 8 performs simple and multiple linear regression analysis. The menu driven options make it easy to analyze data from gradient experiments such nitrogen (N) fertilizer rates, seeding depths, seeding rates, row spacing, and chemical concentration experiments.

The focus of a gradient treatment design is to investigate the response relationship. To do that, one should plot the response (Y) against the treatment level (X) and look for an equation describing the relationship between X and Y.

For example, you wish to determine if there is a relationship in the response of eastern gamagrass to increased rates of nitrogen fertilizer. Your study design is a randomized complete block with 3 replications. Nitrogen rates (“Nrate”) are applied in increments of 120 lbs beginning with 0 lb/acre up to 480 lb/acre (0, 120, 240, 360 and 480 lb/acre). Yield is taken ever 45 days during the growing season and reported in tons/acre. Data is entered into the *Statistix 8* spreadsheet (Fig. 4-1).

To execute the regression analyses:

1. From the main menu, select **Statistics**, **Linear Models**, **Linear Regression** (Fig. 4-2).
2. In the ***Linear Regression*** dialog box (Fig. 4-3), move the variable “TotalYld” from the *Variables* list in the left column to the *Dependent Variable* list in the right column by highlighting the variable and clicking on the right-arrow button to the left of the *Dependent Variable* text area.
3. Move the variable “Nrate” from the *Variables* list in the left column to the *Independent Variable* list in the right column by highlighting the variable and clicking on the right-arrow to the left of the *Independent Variable* list (Fig. 4-4).
4. Click **OK.**

Statistix 8 performs the linear regression analyses and provides a table of the coefficients (Fig. 4-5, Table 4-1).

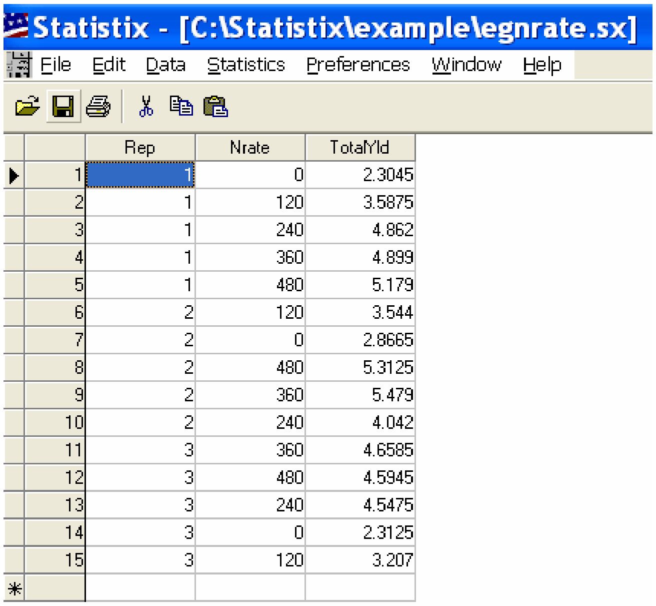


Figure 4-1

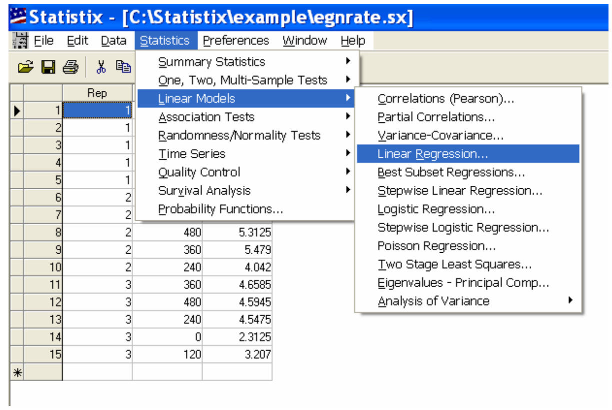


Figure 4-2

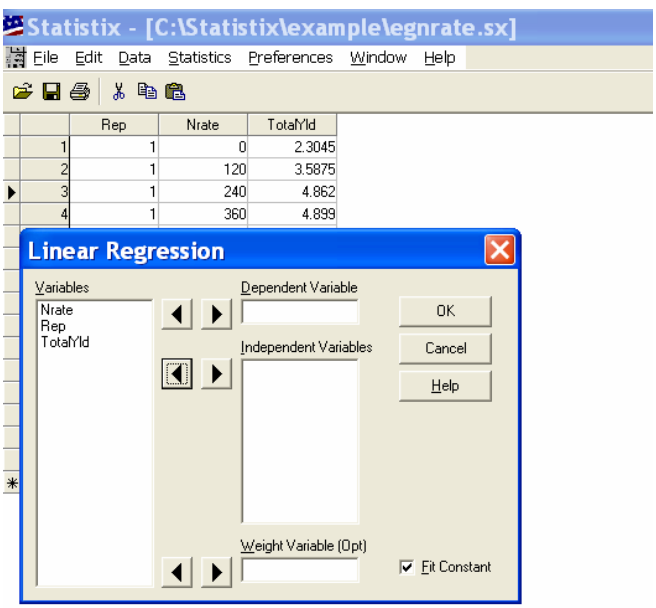


Figure 4-3

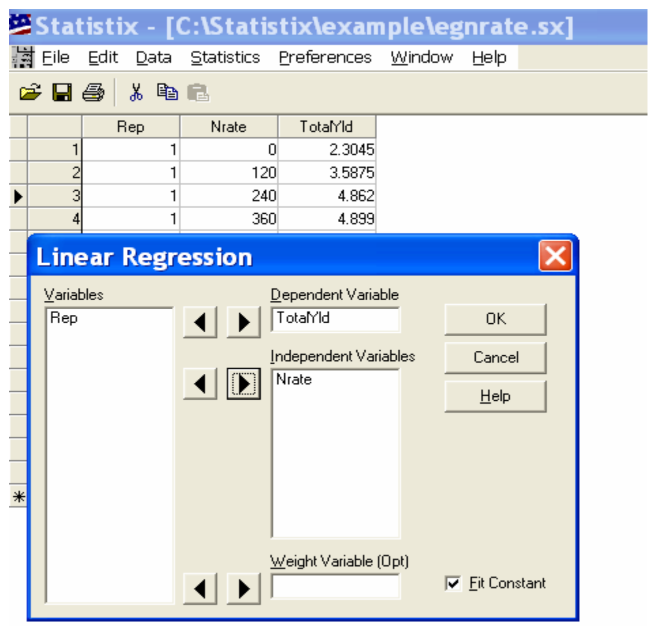
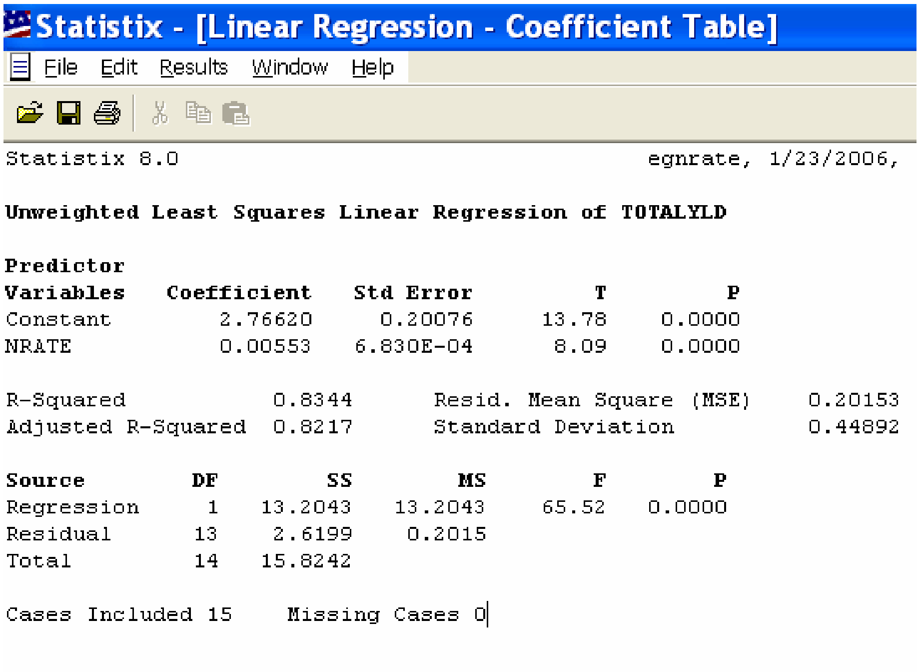


Figure 4-4



```
Statistix - [Linear Regression - Coefficient Table]
File  Edit  Results  Window  Help

Statistix 8.0                                        egnrate, 1/23/2006,

Unweighted Least Squares Linear Regression of TOTALYLD

Predictor
Variables    Coefficient    Std Error        T          P
Constant         2.76620     0.20076      13.78     0.0000
NRATE            0.00553    6.830E-04      8.09     0.0000

R-Squared              0.8344      Resid. Mean Square (MSE)     0.20153
Adjusted R-Squared     0.8217      Standard Deviation           0.44892

Source          DF        SS          MS          F         P
Regression       1    13.2043     13.2043      65.52    0.0000
Residual        13     2.6199      0.2015
Total           14    15.8242

Cases Included 15     Missing Cases 0
```

Figure 4-5

B. Interpreting the Regression Analyses

Table 4-1

```
Linear Regression of TOTALYLD

Predictor
Variables    Coefficient    Std Error        T          P

Constant         2.76620      0.20076     13.78     0.0000
```

The constant coefficient (2.76620) is the point where the regression line intercepts or crosses the Y axis. In this example, the line intercepts Y at 2.8 (rounded) which would be 2.8 tons/acre.

```
NRATE            0.00553    6.830E-04      8.09     0.0000
```

The 0.00553 is the linear regression coefficient which is the slope of the line or the amount of change in Y for each unit change in X. In this example, for every one lb of N applied we see a 0.00553 tons/acre yield increase in gamagrass. For a better interpretation of the number we will convert 0.00553 tons/acre to lbs/acre so we find that for every 1 lb of N applied we get an 11 lb/acre increase in yield.

```
R-Squared            0.8344       Resid. Mean Square (MSE)     0.20153
```

The R-squared (R^2) value of 0.8344 or 83% is the total variation in yield that can be accounted for by linear function of the independent variable X, which is N fertilization rate. The higher the R^2 value the more important the regression equation is in characterizing Y.

```
Adjusted R-Squared  0.8217      Standard Deviation         0.44892
```

Source	DF	SS	MS	F	P
Regression	1	13.2043	13.2043	65.52	0.0000
Residual	13	2.6199	0.2015		
Total	14	15.8242			

```
Cases Included 15    Missing Cases 0
```

Since the P value is less than p=0.05 then we can conclude that the analysis of variance for the overall model is significant.

Note: If the R^2 value is low, even if the P value is significant for the model, the regression equation may not be meaningful. For example, an R^2 is .30, even if the model is significant, indicates that only 30% of the variation in the dependent variable Y is explained by the linear function of the independent variables considered. In other word, 70% of the variation in Y cannot be accounted for by the regression. With such low level of influence, the estimated regression equation would not be useful in estimating, much less predicting, the value of Y.

C. Model Defined

The model for this example is defined as linear

$$Y = a + bX$$

Where: Y is estimated yield; *a* is intercept and b is slope.

In our example, the model would be written with the following coefficients:

$$\text{Yield} = 2.7660 + 0.00553\,(X)$$

D. Application and Presentation of Regression Equation

Our question at the beginning of this example was whether or not there is a relationship in yield of eastern gamagrass as a function of N fertilizer. The linear model was significant and N fertilizer explained 83% of the variation in yield (Fig. 4-6). We can conclude that there is a positive relationship in yields of eastern gamagrass with increased rates of N fertilization.

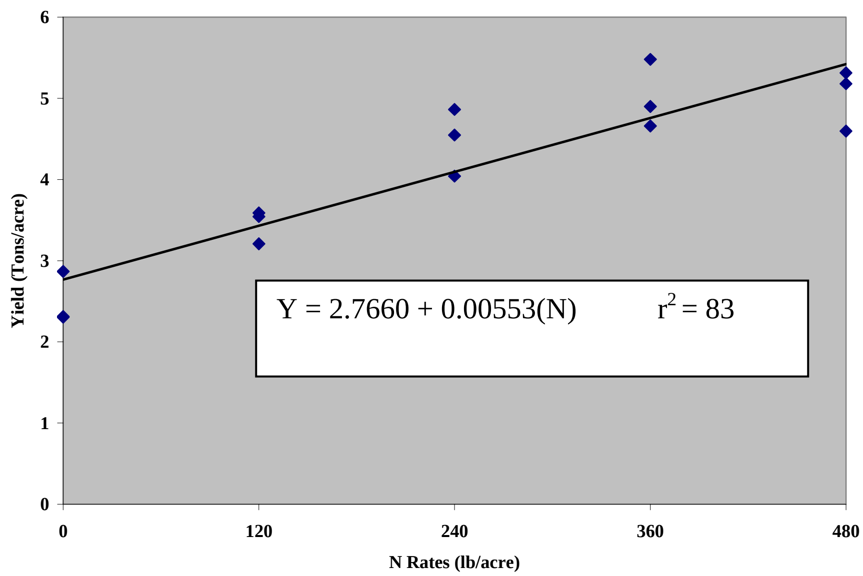


Figure 4-6

We can use the regression equation to predict yield at treatment levels not included in the data. For example, you want to predict what the yield of eastern gamagrass would have been if 100 of N was applied. Since 100 lbs is between 0-480 lbs, which was in the range of fertilizer rates applied, we are able to predict the yield using the coefficients calculated by *Statistix 8*.

Using the equation, Y = 2.7660 + 0.00553(N), and inserting 100 for N, the equation is written as:

Yield = 2.7660 + 0.00553(100)

To estimate the yield of eastern gamagrass multiply 0.00553 by 100, then add 2.7660. The calculated result is 3.319. Thus, the predicted yield of eastern gamagrass fertilized with 100 lbs of N would be approximately 3.3 tons/acre.

II. Quadratic Regression

A. Eastern Gamagrass Nitrogen Rate Trial

Statistix 8 can also perform quadratic regression analysis or second-degree polynomials. In order to perform this task we will need to transform the data using the built-in mathematical formulas in the *Statistix 8* software program.

B. Regression Analysis (Quadratic) and Transformations

For our example, we will use the data set from the pervious N fertilization experiment example (Fig. 4-1). In order to describe eastern gamagrass response to N fertilization using a quadratic regression equation we will need to “square” the independent variable X, which in our case is “Nrate”, and create a new variable called “N2”.

To create a new variable and transform the data:

1. From the main menu, select **Data**, **Insert, Variable** (Fig. 4-7).
2. In the ***Insert Variable*** dialog box, enter the new variable "N2" without quotes in the *New Variable Names* text area (Fig. 4-8).
3. Click **OK**.

The new variable "N2" is inserted into the data set (Fig. 4-9).

Note: Notice that the column contains the letter "M" which indicates missing data.

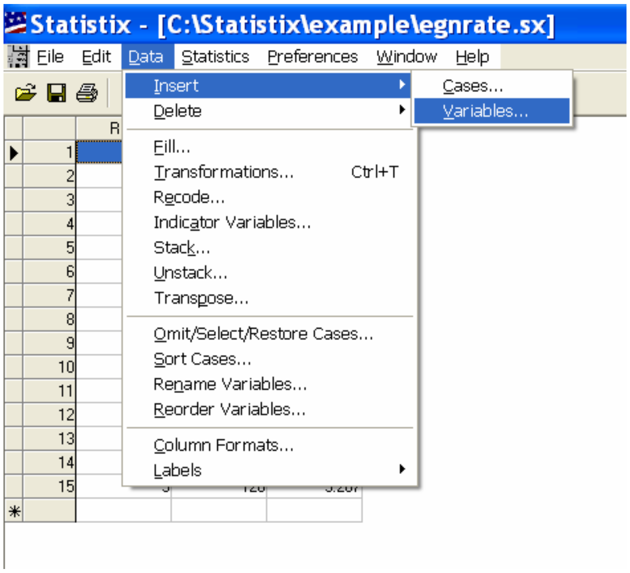


Figure 4-7

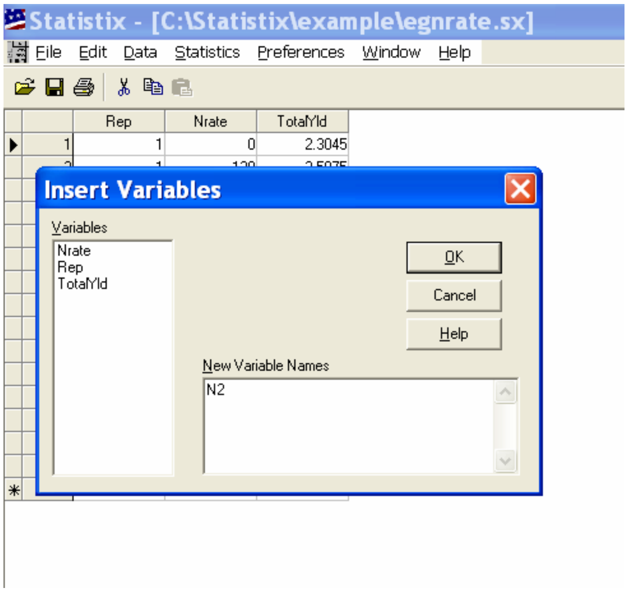


Figure 4-8

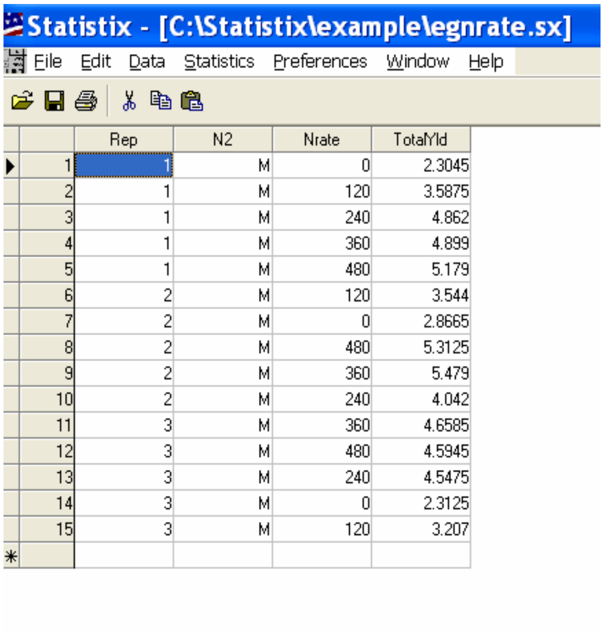


Figure 4-9

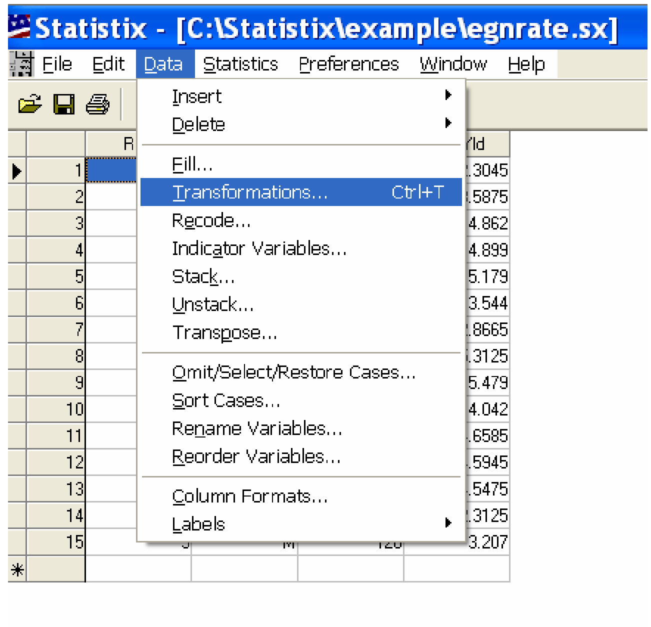


Figure 4-10

To calculate the square of the variable "Nrate":

1. From the main menu, select **Data**, **Transformations** (Fig. 4-10).
2. In the ***Tranformations*** dialog box (Fig. 4-11), move the variable "N2" from the *Variables* list in the left column to the *Transformation Expression* text area by highlighting the variable and clicking on the right-arrow button to the left of the *Transformation Expression* text area.
3. Insert an equals sign (=) after the "N2" variable (e.g. N2=).
4. Insert the square function into the *Transformation Expression* text area by scrolling down the *Functions* list in the right column, highlighting the function "Sqr()", and clicking on the left-arrow button to the left of the "Functions" list (Fig. 4-12). The equation in the *Transformation Expression* text area should be:

   N2=Sqr( )

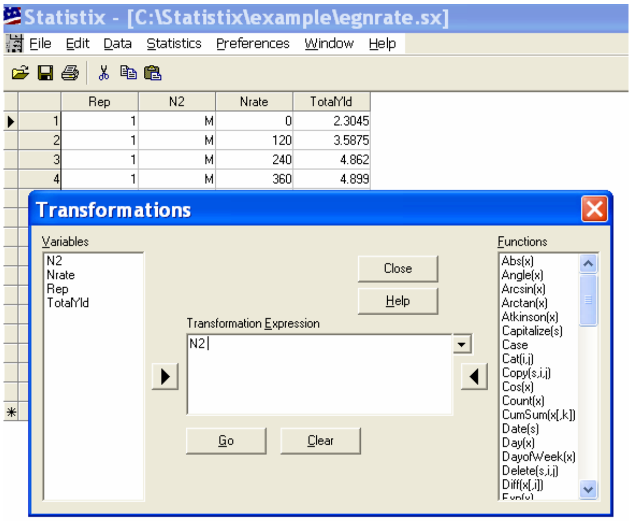


Figure 4-11

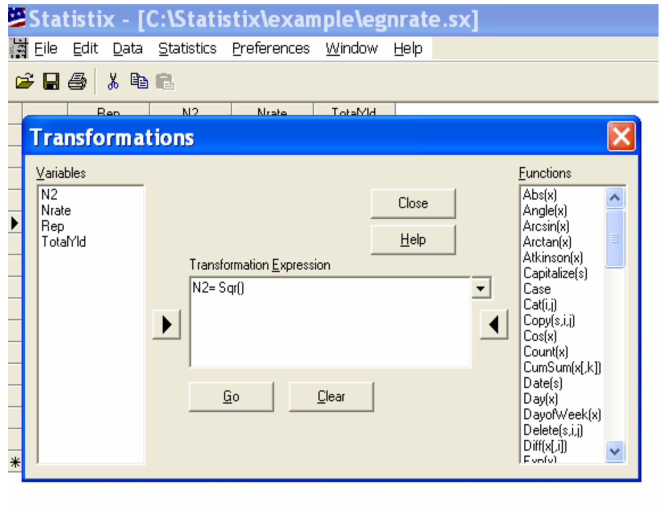


Figure 4-12

To take the square of the “Nrate”:

1. Move the variable “Nrate” from the *Variables* list in the left column to the *Transformation Expression* text area by highlighting the variable and clicking the right-arrow button to the left of the *Transformation Expression* text area.

The expression “N2=Sqr(Nrate)” appears in the *Transformation Expression* text area (Fig. 4-13).

2. Click **Go.**

The transformation is executed and the new values are added into the *N2* column (Fig 4-14).

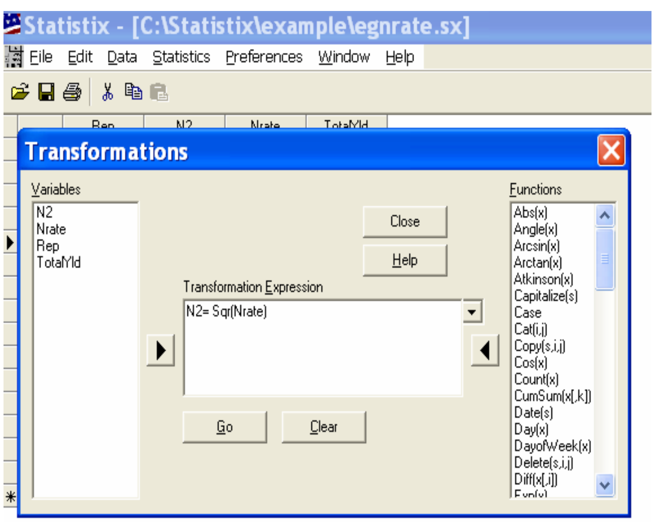
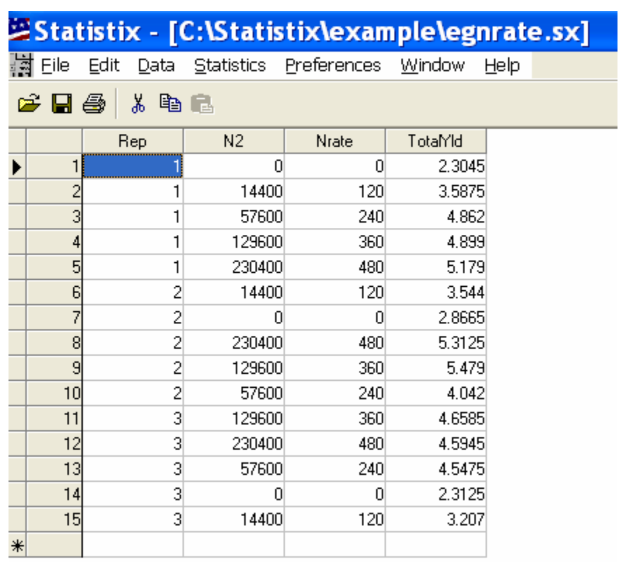


Figure 4-13



	Rep	N2	Nrate	TotalYld
1	1	0	0	2.3045
2	1	14400	120	3.5875
3	1	57600	240	4.862
4	1	129600	360	4.899
5	1	230400	480	5.179
6	2	14400	120	3.544
7	2	0	0	2.8665
8	2	230400	480	5.3125
9	2	129600	360	5.479
10	2	57600	240	4.042
11	3	129600	360	4.6585
12	3	230400	480	4.5945
13	3	57600	240	4.5475
14	3	0	0	2.3125
15	3	14400	120	3.207

Figure 4-14

C. Quadratic Regression

To execute the regression analyses:

1. From the main menu, select **Statistics**, **Linear Models**, **Linear Regression** (Fig. 4-15)
2. In the ***Linear Regression*** dialog box (Fig. 4-16), move the variable “TotalYld” from the “Variables” list in the left column to the *Dependent Variable* list in the right column by highlighting the variable and clicking n the right-arrow button to the left of the *Dependent Variable* list.
3. Move the variables “Nrate” and “N2” from the *Variables* list in the left column to the *Independent Variable* list in the right column by highlighting each variable one at a time and clicking on the right-arrow button to the left of the *Independent Variable* list.
4. Click **OK.**

Statistix 8 performs the linear regression analyses and provides a table of the coefficients (Fig. 4-17, Table 4-2).

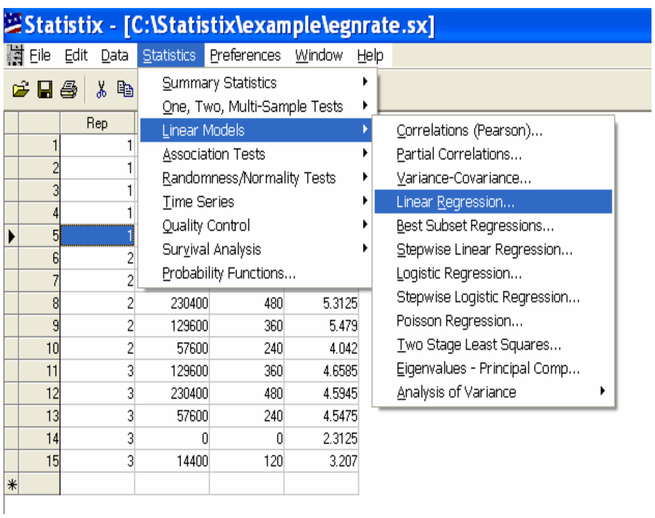


Figure 4-15

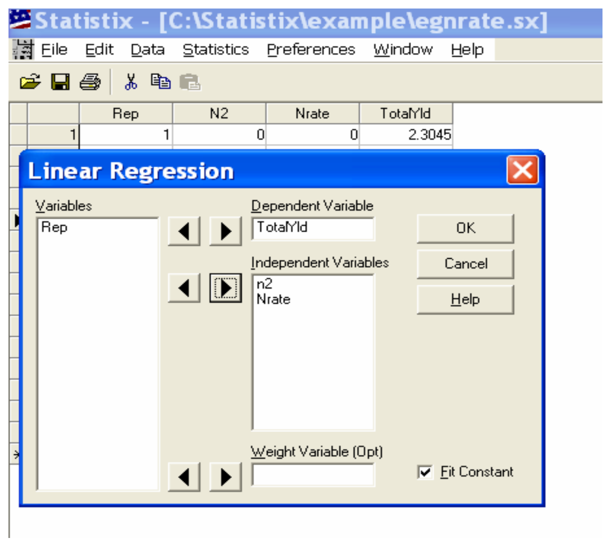
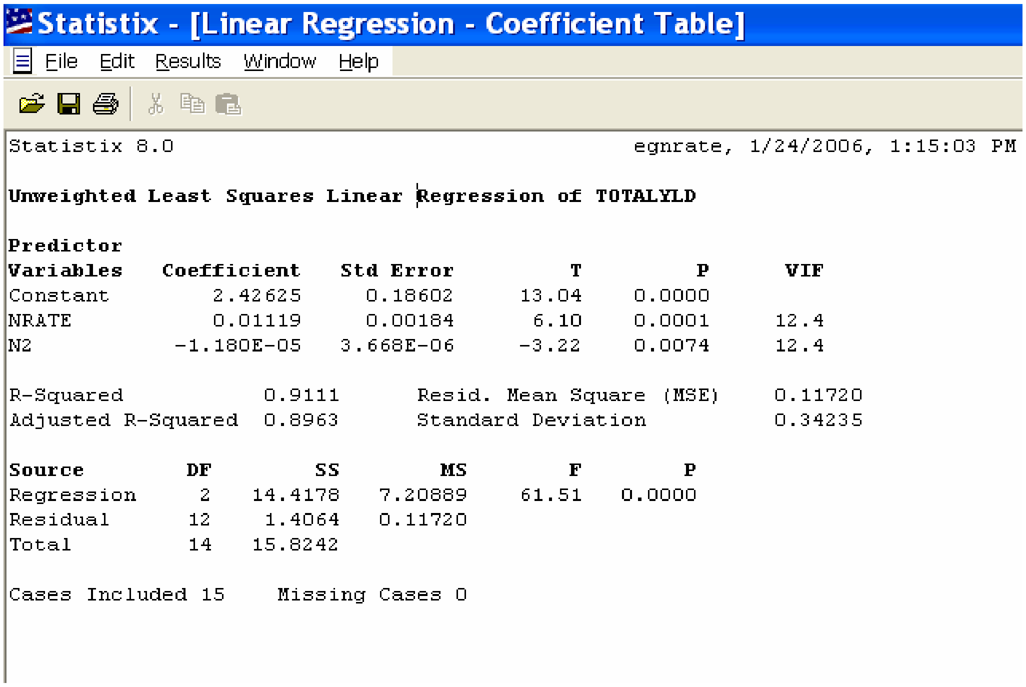


Figure 4-16




```
Statistix 8.0                                    egnrate, 1/24/2006, 1:15:03 PM

Unweighted Least Squares Linear Regression of TOTALYLD

Predictor
Variables    Coefficient    Std Error        T        P       VIF
Constant         2.42625      0.18602    13.04   0.0000
NRATE            0.01119      0.00184     6.10   0.0001      12.4
N2            -1.180E-05    3.668E-06    -3.22   0.0074      12.4

R-Squared              0.9111      Resid. Mean Square (MSE)    0.11720
Adjusted R-Squared     0.8963      Standard Deviation          0.34235

Source         DF        SS         MS        F        P
Regression      2   14.4178    7.20889    61.51   0.0000
Residual       12    1.4064    0.11720
Total          14   15.8242

Cases Included 15    Missing Cases 0
```

Figure 4-17

D. Interpretation of the Regression Analysis

Table 4-2

```
Regression of TOTALYLD

Predictor
Variables    Coefficient    Std Error        T        P       VIF
Constant         2.42625      0.18602    13.04   0.0000
```

Constant or Intercept is 2.43

```
NRATE            0.01119      0.00184     6.10   0.0001      12.4
```

Coefficient for Nrate is 0.01

```
N2            -1.180E-05    3.668E-06    -3.22   0.0074      12.4
```

Coefficient for N2 is -1.18E-05 or -0.0000118

```
R-Squared              0.9111      Resid. Mean Square (MSE)    0.11720
```

The R-squared (R^2) value is .91 or 91% which in this case indicates a good relationship.

```
Adjusted R-Squared     0.8963      Standard Deviation          0.34235

Source         DF        SS         MS        F        P
Regression      2   14.4178    7.20889    61.51   0.0000
Residual       12    1.4064    0.11720
Total          14   15.8242

Cases Included 15    Missing Cases 0
```

Since the P value is less than p=0.05 then we can conclude that the analysis of variance for the quadratic regression model is significant.

E. Model Defined

The model for this example is defined as:

$$Y = a + B_1X + B_2X^2$$

Where: Y is estimated yield; *a* is intercept and B_1 and B_2 are partial regression coefficients.

In our example, the quadratic regression equation would be written with the following coefficients:

$$\text{Yield} = 2.42 + 0.01\ N + -\ 1.18E\text{-}05\ N^2$$

F. Application and Presentation of Regression Equation

A quadratic equation fit the real data on which this example was based better than the linear equation we used earlier (91% vs. 83%). We can conclude that eastern gamagrass evaluated in this study produced a curvilinear response to N fertilization (Fig. 4-18).

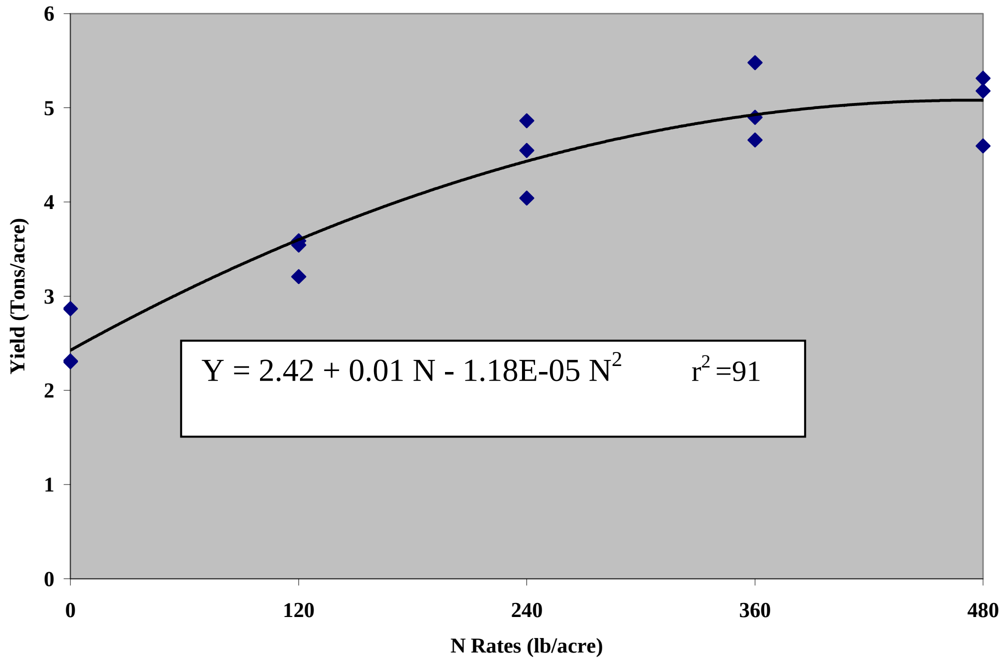


Figure 4-18

We can also use the quadratic equation to predict the yield of eastern gamagrass at a treatment level not included in the data. For example, we wish to estimate the yield of eastern gamagrass fertilized with 200 lbs/acre N. Insert 200 in the partial regression coefficients (N and N^2) and calculate the estimated yield.

$$\text{Yield} = 2.42 + 0.01\ (200)\ -1.18E\text{-}05\ (200)^2$$

The estimated yield of eastern gamagrass fertilized with 200 lbs of N/acre is 3.9 tons/acre.

The quadratic equation can also be useful in determining at which treatment becomes cost-effective or the treatment level associated with the maximum or minimum response. For example, we wish to determine the N rate at maximum yield. To make this calculation we will take a derivative of the regression equation and set it equal to 0 (zero) and solve for N (nitrogen).

The equation becomes:

$B_1 + 2(B^2\text{N}) = 0$

Solving for Nitrogen = $\dfrac{B_1}{-2B_2}$

Nitrogen = $\dfrac{0.01}{-2(0.0000118)}$

Nitrogen = 424

The N rate at maximum yield occurred at 424 lb/acre.

Chapter 5: Missing Data

I. Missing Data

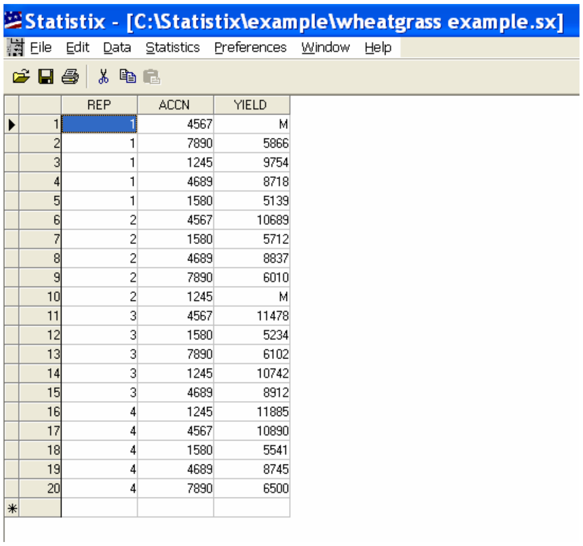
A. Description

There is a difference between missing data and the number zero (0). For example, if cultivar A was used as standard of comparison in an adaptation trial in Pullman, WA and was winterkilled, then it would be reported as a 0 because the cultivar was not adapted. Conversely, if you have 10 clover cultivars planted in a randomized complete block with four replications and deer forages on cultivar A in replication 4, and no production data can be collected, then replication 4 would be reported as missing data.

There are times when you may encounter missing data in your experiment. There are techniques to estimate missing data that are based on data for the treatment with the missing value that are available and data for treatments in the surrounding plots (see Gomez and Gomez, 1984). However, the question is can statistical analyses be performed on data collected from unequal observations? In other words, can data collected from "accession A" from 3 observations (replications) be compared to data collected on accessions from 4 replications without estimating a missing value for accession A? The answer to this question is yes. *Statistix 8* handles missing data for a treatment observation without requiring that you estimate a missing value for the treatment.

For example, you collected yield data on 5 accessions of wheatgrasses that were planted in a randomized complete block design replicated 4 times (Fig. 5-1). Unfortunately, you were unable to collect data on accession 4567 in replication 1 and accession 1245 in replication 2 due to loss of the sample. In the dataset, the missing data is represented by the letter "M".

Perform the analysis of variance for a randomized complete block design with the missing data in the dataset (Fig. 5-2).



	REP	ACCN	YIELD
1	1	4567	M
2	1	7890	5866
3	1	1245	9754
4	1	4689	8718
5	1	1580	5139
6	2	4567	10689
7	2	1580	5712
8	2	4689	8837
9	2	7890	6010
10	2	1245	M
11	3	4567	11478
12	3	1580	5234
13	3	7890	6102
14	3	1245	10742
15	3	4689	8912
16	4	1245	11885
17	4	4567	10890
18	4	1580	5541
19	4	4689	8745
20	4	7890	6500
*			

Figure 5-1

```
Randomized Complete Block AOV Table for YIELD

Source    DF        SS          MS          F        P
REP        3     1148484      382828
ACCN       4   9.005E+07   2.251E+07    117.25   0.0000
Error     10     1920020      192002
Total     17

Note: SS are marginal (type III) sums of squares

Grand Mean 8401.5    CV 5.22

Tukey's 1 Degree of Freedom Test for Nonadditivity
Source          DF       SS        MS        F        P
Nonadditivity    1    401596    401596    2.38   0.1573
Remainder        9   1518424    168714

Relative Efficiency, RCB 1.13

Means of YIELD for ACCN

ACCN   N    Mean     SE
1245   3   10797   252.98
1580   4    5407   219.09
4567   3   10882   252.98
4689   4    8803   219.09
7890   4    6120   219.09
```

Figure 5-2

B. Results and Interpretation of the Analyses

Rather than discuss all of the numbers in the analysis of variance we will discuss those of interest to the user (Table 5-1).

Table 5-1

```
Randomized Complete Block AOV Table for YIELD

Source    DF          SS          MS          F        P
REP        3     1148484      382828
ACCN       4   9.005E+07   2.251E+07     117.25   0.0000
Error     10     1920020      192002
Total     17
```

Since the P value is less than 0.05 then we conclude that there are significant differences between wheatgrasses.

Note: SS are marginal (type III) sums of squares

A Type III sums of squares test the same hypothesis that would have been tested if the cell size would have been equal that is, if there was no missing data.

```
Grand Mean 8401.5    CV 5.22


Tukey's 1 Degree of Freedom Test for Nonadditivity
Source          DF        SS        MS        F        P
Nonadditivity    1    401596    401596     2.38   0.1573
Remainder        9   1518424    168714

Relative Efficiency, RCB 1.13
```

```
Means of YIELD for ACCN

ACCN   N    Mean      SE
1245   3   10797  252.98
1580   4    5407  219.09
4567   3   10882  252.98
4689   4    8803  219.09
7890   4    6120  219.09
```

The mean for ACCN 1245 and 4567 was determined by 3 replications (N=observations) rather than 4 replications as in the other accessions (ACCN).

Since there were significant differences in wheatgrass accessions as determined by the analysis of variance (p= 0.0000), you may perform a mean separation test to determine significant means and report the results.

Chapter 6: Data Transformations

I. Introduction

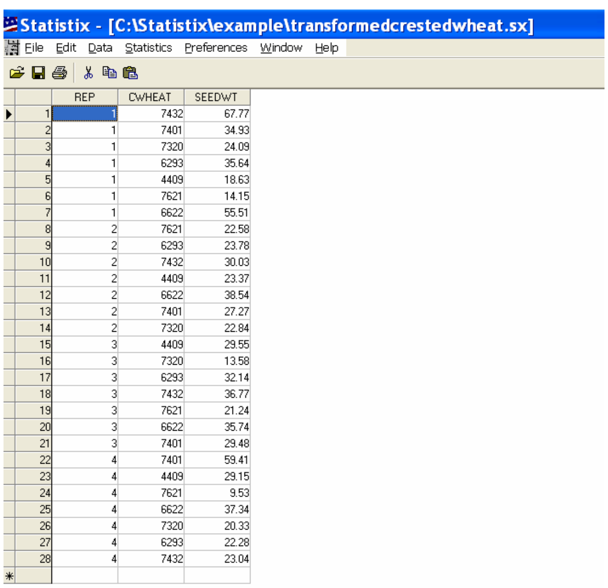
The analysis of variance is valid when certain assumptions are made such as *additive effects*, (treatment and environmental effects are additive) *independence of errors* (experimental errors are independent), *homogeneity of variance* (experimental errors have common variance) and *normal distribution* (experimental errors are normally distributed). When these assumptions are not valid (e.g. when there is a significant treatment x replication), data transformations may be used. Some of the more popular data transformations are square root, logarithmic, and arc sine. In another chapter, we used the square root transformation of N rate to create a new variable named N2.

II. Tukey's One Degree of Freedom Test of Non-additivity

A. Description

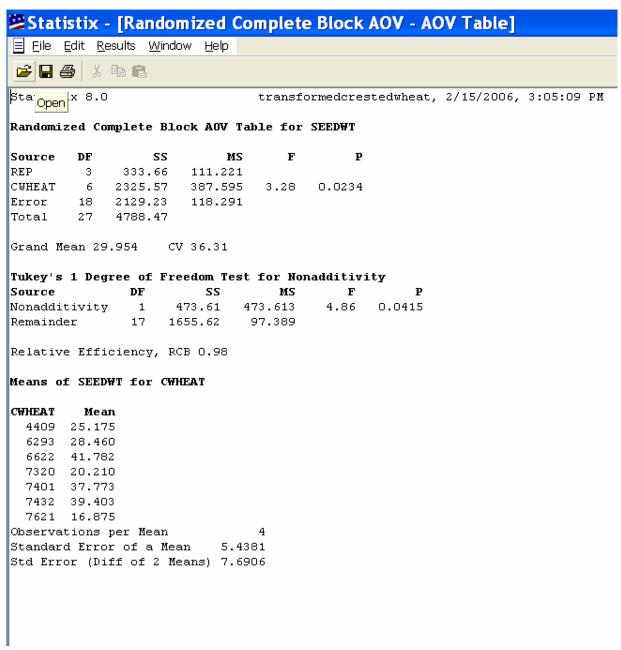
Statistix 8 conducts a Tukey's One Degree of Freedom Test for Non-additivity for a randomized complete block design analysis of variance. In the example given in chapter 2, we found that there was no suggestion of non-additivity in the Yield1 data of the switchgrass cultivar trial (p=0.4649). If the P value had been less than p=0.05, which would have indicated the likely presence of non-additivity, we would have transformed the data in an attempt to reduce the non-additive effect. The question is how do we transform the data to remove the non-additive effect and how is it reported?

For this example, we will use seed weight data of 7 accessions of crested wheatgrass harvested from plots arranged in a randomized complete block design with four replications (Fig. 6-1). You perform the analysis of variance (Fig. 6-2).



	REP	CWHEAT	SEEDWT
1	1	7432	67.77
2	1	7401	34.93
3	1	7320	24.09
4	1	6293	35.64
5	1	4409	18.63
6	1	7621	14.15
7	1	6622	55.51
8	2	7621	22.58
9	2	6293	23.78
10	2	7432	30.03
11	2	4409	23.37
12	2	6622	38.54
13	2	7401	27.27
14	2	7320	22.84
15	3	4409	29.55
16	3	7320	13.58
17	3	6293	32.14
18	3	7432	36.77
19	3	7621	21.24
20	3	6622	35.74
21	3	7401	29.48
22	4	7401	59.41
23	4	4409	29.15
24	4	7621	9.53
25	4	6622	37.34
26	4	7320	20.33
27	4	6293	22.28
28	4	7432	23.04

Figure 6-1




```
Statistix 8.0                    transformedcrestedwheat, 2/15/2006, 3:05:09 PM

Randomized Complete Block AOV Table for SEEDWT

Source    DF        SS        MS       F        P
REP        3    333.66   111.221
CWHEAT     6   2325.57   387.595    3.28   0.0234
Error     18   2129.23   118.291
Total     27   4788.47

Grand Mean 29.954    CV 36.31

Tukey's 1 Degree of Freedom Test for Nonadditivity
Source            DF        SS         MS        F        P
Nonadditivity      1    473.61    473.613     4.86   0.0415
Remainder         17   1655.62     97.389

Relative Efficiency, RCB 0.98

Means of SEEDWT for CWHEAT

CWHEAT    Mean
  4409  25.175
  6293  28.460
  6622  41.782
  7320  20.210
  7401  37.773
  7432  39.403
  7621  16.875
Observations per Mean           4
Standard Error of a Mean   5.4381
Std Error (Diff of 2 Means) 7.6906
```

Figure 6-2

B. Results and Interpretation of the Analyses

Results of the analysis of variance (Table 6-1) indicate that there are significant differences in crested wheatgrass accessions as indicated by the P value of 0.0234. However, we also find that there is evidence of non-additivity effect as determined by Tukey's 1 Degree of Freedom Test for Non-additivity (p=0.0415). The next step is to see if we can reduce the non-additivity by transforming the data.

Table 6-1

Randomized Complete Block AOV Table for SEEDWT

Source	DF	SS	MS	F	P
REP	3	333.66	111.221		
CWHEAT	6	2325.57	387.595	3.28	0.0234
Error	18	2129.23	118.291		
Total	27	4788.47			

Grand Mean 29.954   CV 36.31

Tukey's 1 Degree of Freedom Test for Nonadditivity

Source	DF	SS	MS	F	P
Nonadditivity	1	473.61	473.613	4.86	0.0415
Remainder	17	1655.62	97.389		

Relative Efficiency, RCB 0.98

Means of SEEDWT for CWHEAT

CWHEAT	Mean
4409	25.175
6293	28.460
6622	41.782
7320	20.210
7401	37.773
7432	39.403
7621	16.875

Observations per Mean 4
Standard Error of a Mean 5.4381
Std Error (Diff of 2 Means) 7.6906

III. Transforming the Data Using Logarithmic

A. Transformation

To transform the seed weight data into a logarithmic scale:

1. From the main menu, select **Data, Transformation** (Fig. 6-3).
2. In the ***Transformation*** dialog box (Fig. 6-4), in the *Transformation Expression* text box, enter "NEWT=" without quotes, to represent new weight (Fig. 6-5).
3. In the *Functions* list, scroll down and select "Log(x)" and click on the right arrow button to insert the Log (x) function into the *Transformation Expression* text area (Fig. 6-6).

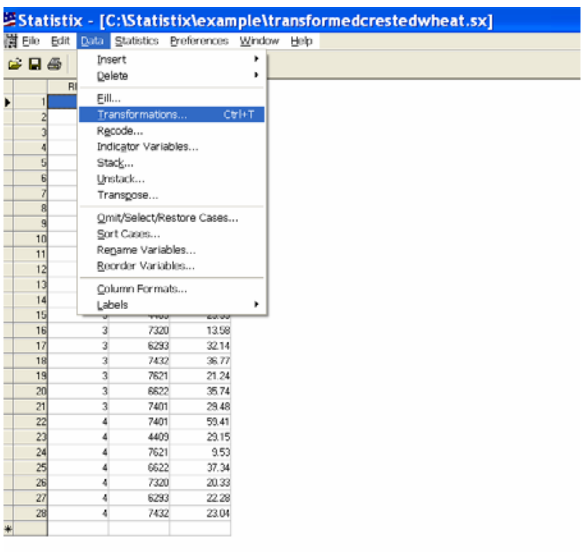


Figure 6-3

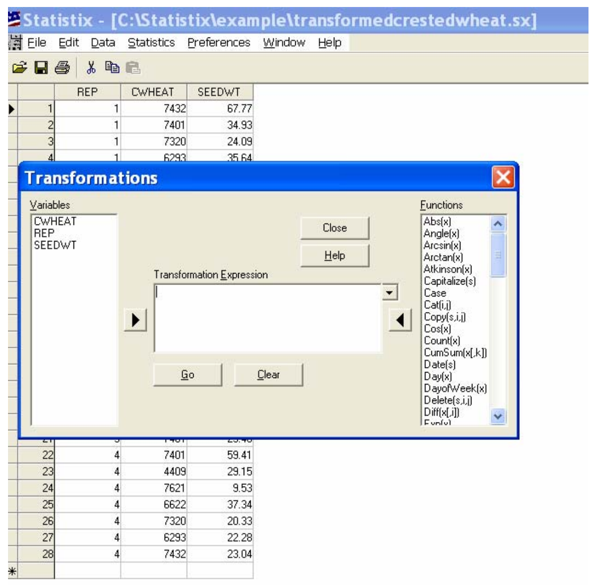


Figure 6-4

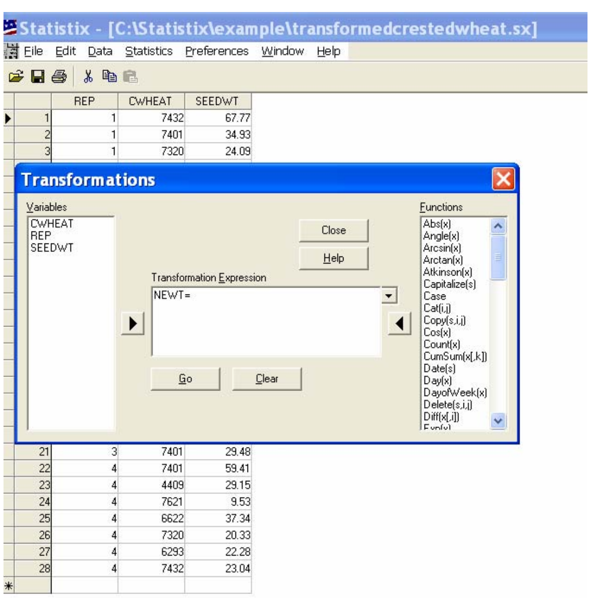


Figure 6-5

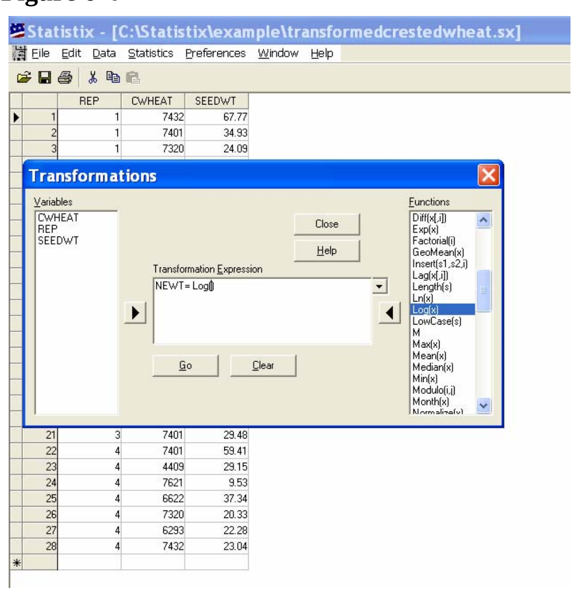


Figure 6-6

The seed weight variable upon which the data transformation will be performed is now added to the mathematical equation.

4. Insert “SEEDWT” within the parenthesis by highlighting the variable in the *Variables* list and clicking on the right-arrow button to the left of the *Transformation Expression* text area, to achieve (Fig. 6-7):

    (NEWT=Log(SEEDWT)

5. Click **Go.**

A new variable named “NEWT” plus the logarithmic of the seed weight data is inserted into the original data set (Fig.6-8).

To determine if transforming the data removed the nonadditivity effect, we will perform an analysis of variance on the “NEWT” variable.

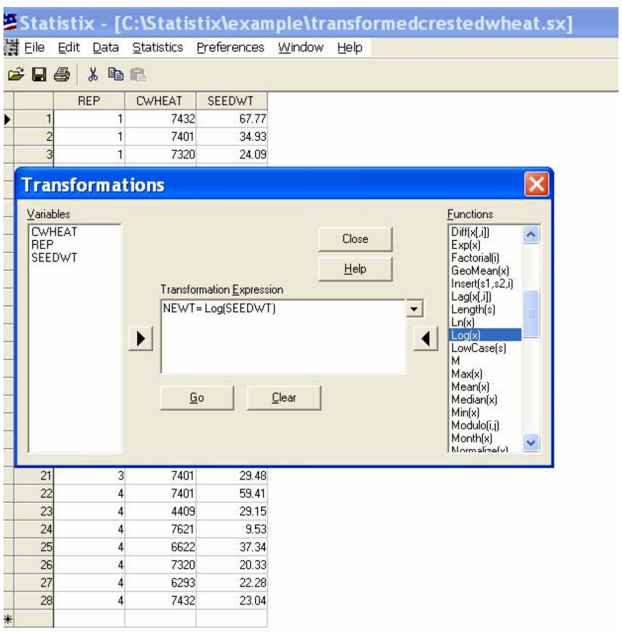
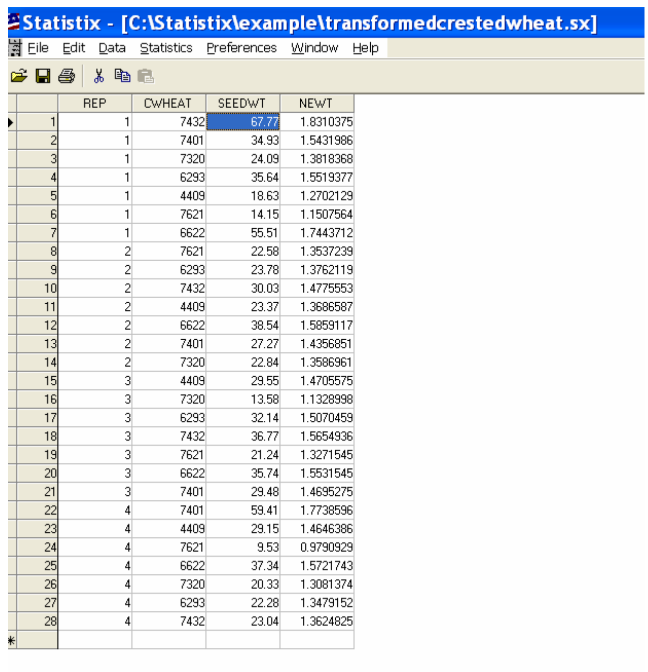


Figure 6-7



	REP	CWHEAT	SEEDWT	NEWT
1	1	7432	67.77	1.8310375
2	1	7401	34.93	1.5431986
3	1	7320	24.09	1.3818368
4	1	6293	35.64	1.5519377
5	1	4409	18.63	1.2702129
6	1	7621	14.15	1.1507564
7	1	6622	55.51	1.7443712
8	2	7621	22.58	1.3537239
9	2	6293	23.78	1.3762119
10	2	7432	30.03	1.4775553
11	2	4409	23.37	1.3686587
12	2	6622	38.54	1.5859117
13	2	7401	27.27	1.4356851
14	2	7320	22.84	1.3586961
15	3	4409	29.55	1.4705575
16	3	7320	13.58	1.1328998
17	3	6293	32.14	1.5070459
18	3	7432	36.77	1.5654936
19	3	7621	21.24	1.3271545
20	3	6622	35.74	1.5531545
21	3	7401	29.48	1.4695275
22	4	7401	59.41	1.7738596
23	4	4409	29.15	1.4646386
24	4	7621	9.53	0.9790929
25	4	6622	37.34	1.5721743
26	4	7320	20.33	1.3081374
27	4	6293	22.28	1.3479152
28	4	7432	23.04	1.3624825

Figure 6-8

B. Results and Interpretation of the Analyses

Since the P value for the Tukey’s 1 Degree Test for Nonadditivity is greater (p=0.3887) than p=0.05 then we can conclude that transforming the seed weight data using the logarithmic transformation was successful in removing the nonadditivity effect (Table 6-2).

Table 6-2

Randomized Complete Block AOV Table for NEWT

Source	DF	SS	MS	F	P
REP	3	0.03514	0.01171		
CWHEAT	6	0.54876	0.09146	4.54	0.0057
Error	18	0.36228	0.02013		
Total	27	0.94617			

Grand Mean 1.4380 CV 9.87

Tukey's 1 Degree of Freedom Test for Nonadditivity

Source	DF	SS	MS	F	P
N onadditivity	1	0.01594	0.01594	0.78	0.3887
Remainder	17	0.34633	0.02037		

Relative Efficiency, RCB 0.94

Means of NEWT for CWHEAT

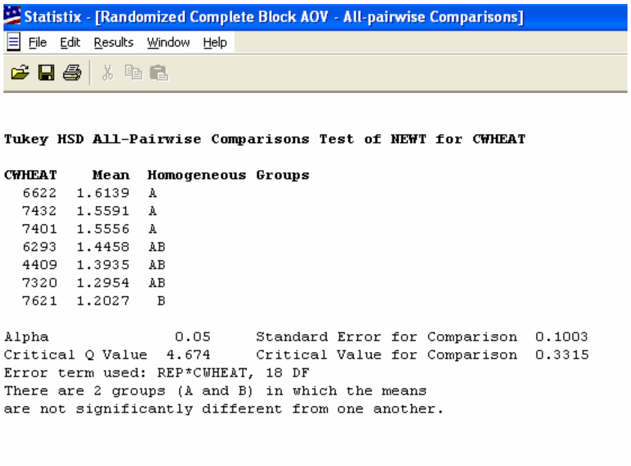
CWHEAT	Mean
4409	1.3935
6293	1.4458
6622	1.6139
7320	1.2954
7401	1.5556
7432	1.5591
7621	1.2027

Observations per Mean 4
Standard Error of a Mean 0.0709
Std Error (Diff of 2 Means) 0.1003

C. Presenting the Results

Now that we have successfully transformed the data and removed the nonadditivity effect, how is the transformed data used and how do you present the results? Since the analysis of variance determine that there were significant differences in the crested wheatgrass accessions (p=0.0057; Table 6-2) then perform an all Pairwise Comparison Test (e.g. Tukey's HSD) on the transformed means to determine differences between accessions (Fig. 6.9).

Like any other data presentation, report the original seed weight data of the 7 crested wheatgrass accessions (means are found in Table 6-1) in a table **but use mean separation test results from the transformed data** (Fig. 6.9)**.** Results are presented in Table 6-3. If the paper is peer viewed for symposia proceedings or journal article, you will need to report in the materials and methods section that the seed weight data was transformed (e.g. logarithmic, square root, etc.). It would also be advisable to report it in the PMC ATR.




```
Tukey HSD All-Pairwise Comparisons Test of NEWT for CWHEAT

CWHEAT    Mean  Homogeneous Groups
  6622  1.6139  A
  7432  1.5591  A
  7401  1.5556  A
  6293  1.4458  AB
  4409  1.3935  AB
  7320  1.2954  AB
  7621  1.2027   B

Alpha              0.05     Standard Error for Comparison  0.1003
Critical Q Value  4.674     Critical Value for Comparison  0.3315
Error term used: REP*CWHEAT, 18 DF
There are 2 groups (A and B) in which the means
are not significantly different from one another.
```

Figure 6-9

Table 6-3

Presenting crested wheatgrass seed weight using original means in Table 6-1 and the means separation test using Tukey's HSD in Fig 6-9 that was performed on the transformed data.

Accession	Seed weight
	-----------g/plot----------
6622	41.782 a*
7432	39.403 a
7401	37.773 a
6293	28.460 ab
4409	25.175 ab
7320	20.210 ab
7621	16.875 b
Mean	29.954

* Means followed by the same letters are not significantly different as determined by Tukey's HSD at P<0.05.

Chapter 7: Other Model Statements

I. Other Model Statements

Listed below are examples of model statements that *Statistix 8* recognizes and will perform the appropriate analyses given the model statement has been correctly entered.

In most situations, neither the order in which terms are specified in the model nor the order in which variables are specified within terms has any influence on the analysis. The only exception occurs in certain models with multiple error terms.

Example 1: Completely randomized design, also called the one-way design. If the treatment factor is A, the model is specified simply as:

A

Example 2: Randomized complete block design. BLK is the factor for blocks and A if the treatment factor.

BLK A

Example 3: Single-factor Latin square design. The variables ROW and COL are the row- and column blocking factors, and A is the treatment factor.

ROW COL A

Example 4: Three-factor factorial in a completely randomized design with all two factor interactions.

A B C A*B A*C B*C

Note: The example factorial design above could be entered more concisely using the ALL2 keyword.

ALL2(A B C)

Example 5: Split-plot design. The variable REP is the factor for replication, A is the main-plot factor, and B is the subplot factor.

REP A REP*A(E) B A*B

Note: the interaction REP*A is an error term. The three factor interaction term REP*A*B is also an error term, but was omitted above because Statistix 8 always adds the high order interaction term automatically.

Example 6: Strip-plot design. The variable REP is the factor for replication, A is the main-plot factor, and B is the subplot factor.

REP A REP*A(E) B REP*B(E) A*B

Example 7: One-way repeated measures design. SUBJ is the subjects factor and A is the within-subjects factor.

SUBJ A

Example 8: Two-factor repeated measures design with a between-subjects factor A and a within-subjects factor B.

A SUBJ*A(E) B A*B

Examples 1 - 8 above are all models that could be more easily specified using the specific AOV procedures discussed earlier in this chapter. But the General AOV/AOCV procedure allows you to specify other models, including variations of the above models.

Example 9: A two-factor nested model with factor B nested within factor A.

A B*A

Compare the nested model with the two-factor cross-classified (factorial) model:

A B A*B

Example 10: A three-factor factorial experiment in a split-plot design. The factors A and B are both main-plot factors, and C is the subplot factor.

REP A B A*B REP*A*B(E) C A*C B*C A*B*C

Chapter 8: Converting MSTAT-C Files to Statistix 8 Files

Many PMCs and PMSs used MSTAT-C to analyze performance data of plant accessions and cultivars. The question is how to convert MSTAT-C data files into Statistix 8 files without re-entering the data? The following steps provide a procedure on how to convert MSTAT-C files to Statistix 8 files:

Step 1 Open the file in MSTAT-C.

Step 2 Select 41 (SEDIT). Select which variables that you want to convert (i.e., 1 = rep; 4 = seed trt; 5 = cultivar, etc.

Step 3 Select 36 (Printlist) and chose all cases

Step 4 Enter the variables of interest (i.e., 1, 4, 5)
left justified = No
wide paper = No
paginagation = No
variable description = No

Step 5 Save output to disk

Step 6 Exit MSTAT-C

Step 7 Open OUTPUT file in Microsoft Word (Excel had problems copying alphanumeric variables).

Step 8 Open *Statistix 8*

Step 9 From the main menu bar in *Statistix 8* select **Data, Insert, Variables**
Enter the name of the variables you want to move over from the MSTAT-C file. For example, rep, seedtrt, cultivar

Step 10 Copy data from OUTPUT to Statistix 8 spreadsheet (*shift-alt to prevent copying the "case no." column*)

Step 11 Save file