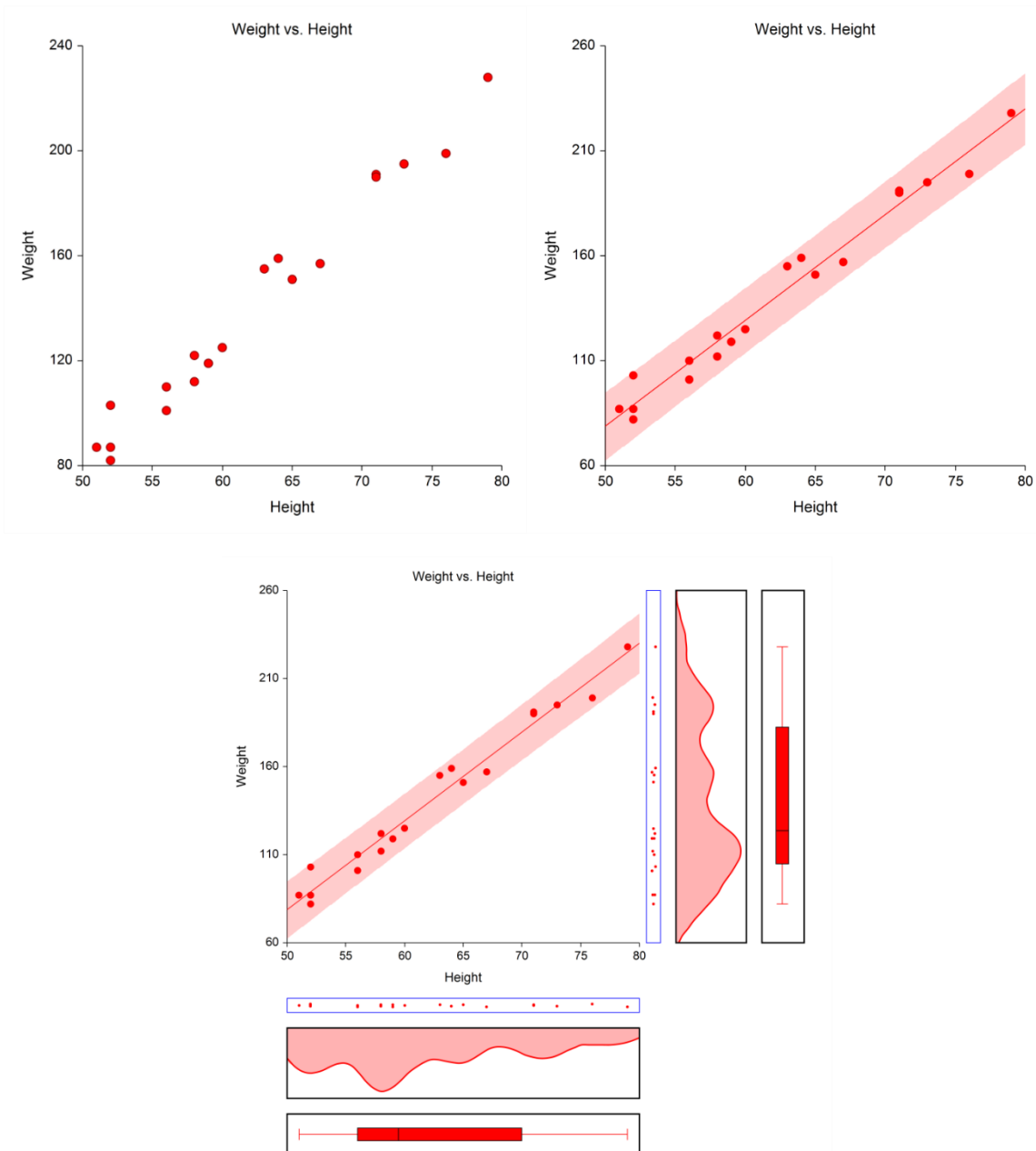


Chapter 161

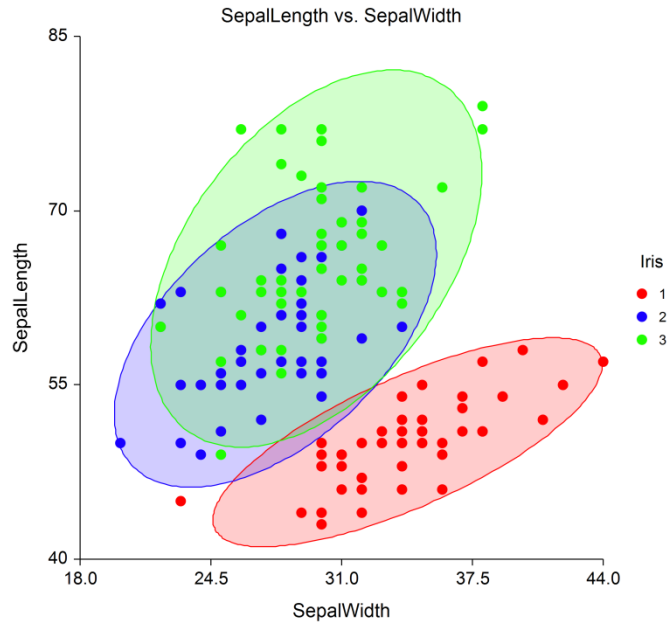
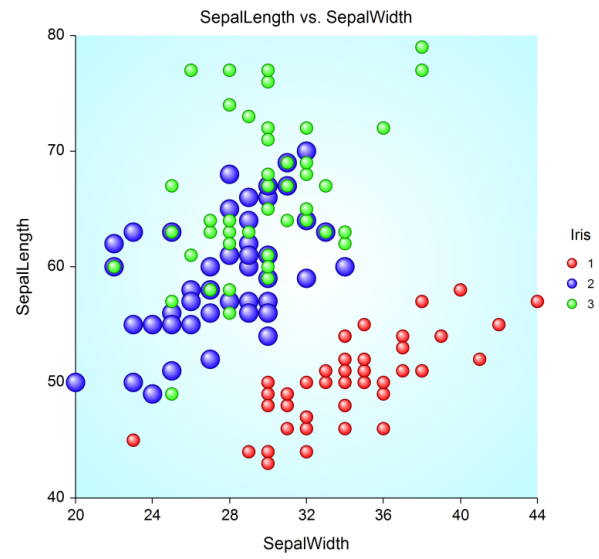
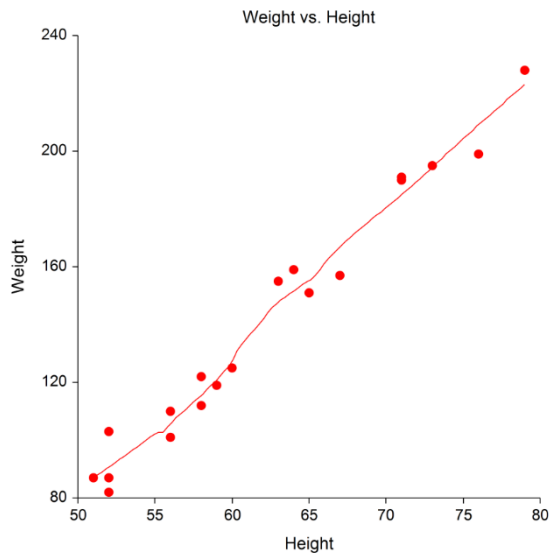
Scatter Plots

Introduction

The x-y scatter plot is one of the most powerful tools for analyzing data. **NCSS** includes a host of features to enhance the basic scatter plot. Some of these features are trend lines (least squares) and confidence limits, polynomials, splines, loess curves, border box plots, and sunflower plots. Following are examples of scatter plots that you can create with this procedure. Instructions on how to create each are given later.



Scatter Plots



Data Structure

A scatter plot is constructed from two variables. A third variable may be used to divide the first two variables into groups (e.g., age group or gender). No other constraints are made on the input data. Note that rows with missing values in one of the selected variables are ignored.

Scatter Plot Window Options

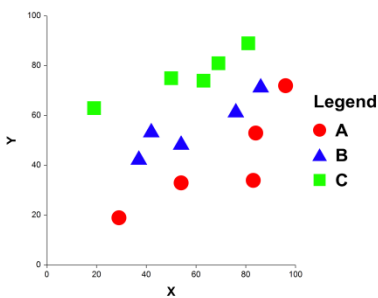
This section describes the specific options available on the Scatter Plot window, which is displayed when the Scatter Plot button is clicked. Common options, such as axes, labels, legends, and titles are documented in the Graphics Components chapter.

Scatter Plot Tab

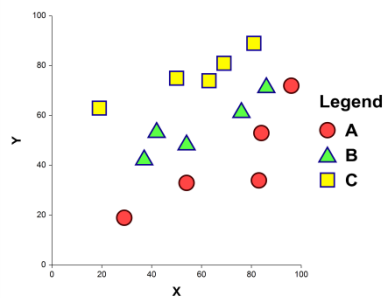
Symbols Section

You can modify the shape, color, and size of the plot symbols. To change them, click the Symbol Format button to display the Symbol Format window. Here are some of the graphics effects that can easily be achieved.

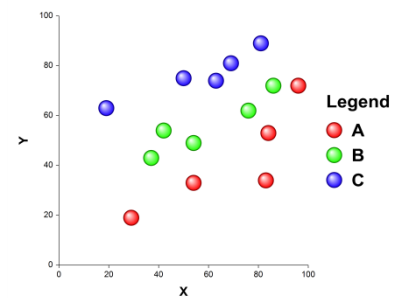
Various Shapes and Colors



Symbols with Borders



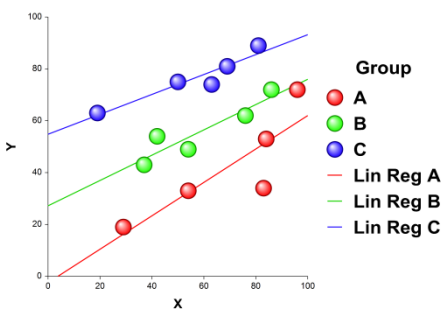
Symbols with Gradient Fill



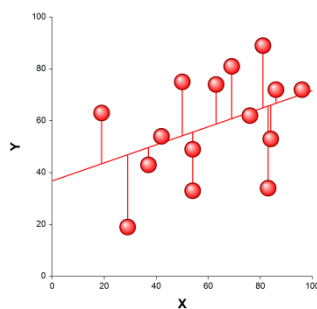
Linear Regression Section

You can add regression lines, residuals, confidence limits, confidence bands, prediction limits, and probability ellipses to the plot using the options available in this section. The technical details of how these items are calculated are presented in the Linear Regression and Correlation chapter.

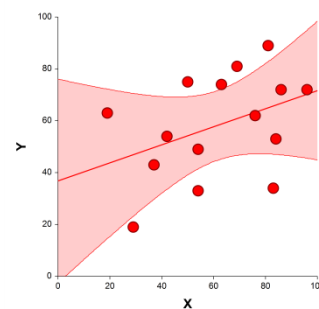
Regression Lines



Residuals

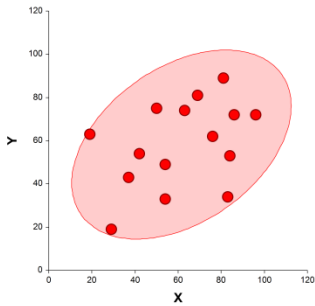


Confidence Band



Scatter Plots

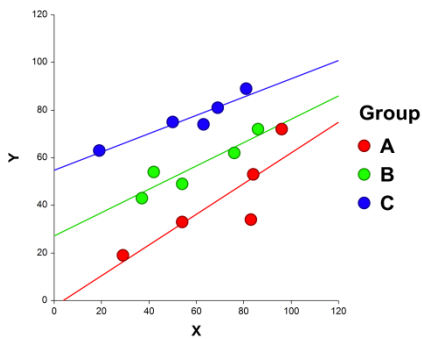
Probability Ellipse



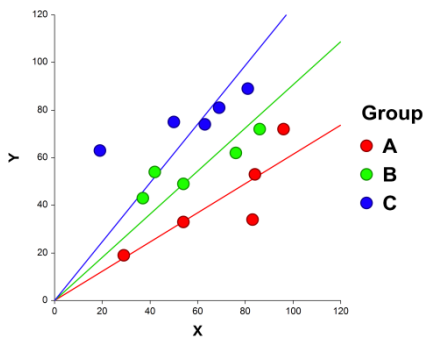
Estimation Section

With these options, you can set the resolution of any curved lines (such as confidence limits and polynomial fits). You can also force the regression line to pass through the origin.

Regular Regression



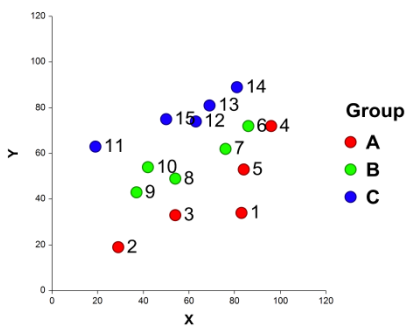
Regression Thru Origin



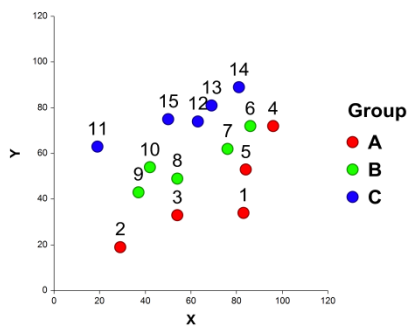
Data Point Labels Section

You can display data labels next to the symbols.

Data Labels - Right



Data Labels - Above

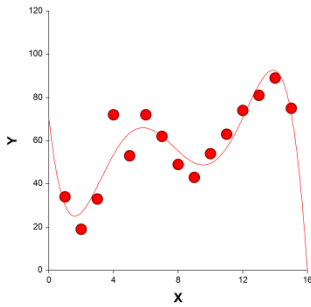


More Lines Tab

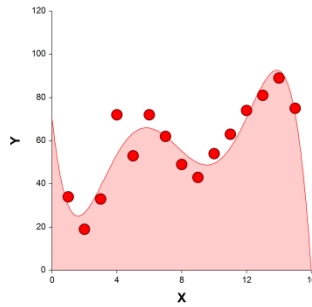
Alternative Linear Regression Lines Section

You can add other types of regression lines to the plot using the options available in this section. A popular alternative to linear regression is curvi-linear regression. Other possible lines are orthogonal (principal components) regression, robust regression through the medians, and robust regression through the quartiles.

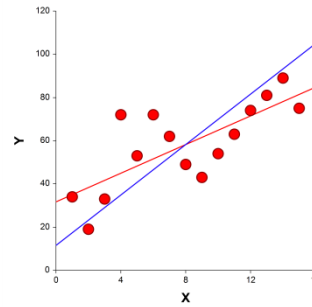
5th Order Polynomial



Polynomial with Fill



Blue Orthogonal Line



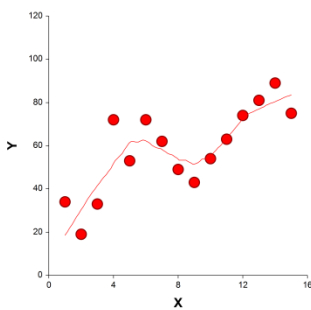
Smooth Lines Section

You can add loess smooth and median smooth lines using the options available in this section.

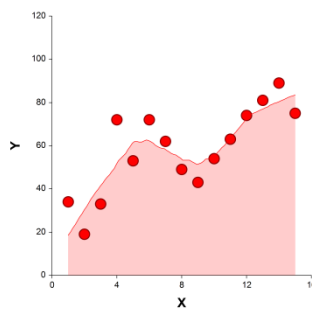
The loess smooth line is documented in the Linear Regression and Correlation chapter.

The median smooth is analogous to the moving average. It is computed using running medians instead of running means.

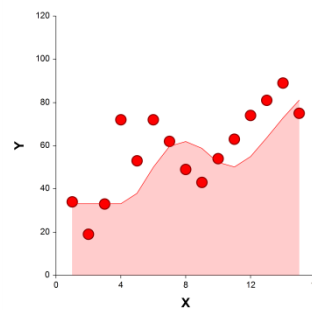
Loess Line



Loess Line with Fill



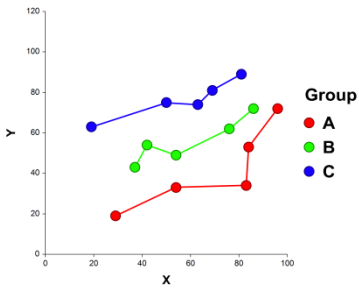
Median Smooth Line



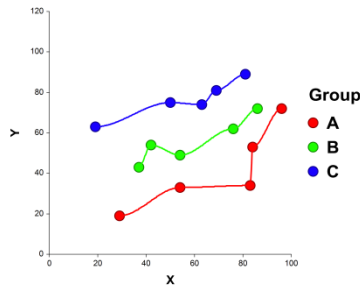
Connect Data Points Section

You can connect the symbols with a line. Along with setting the format of the connecting line, this section also lets you specify the order in which the points are connected. The connection can be done with a straight line or a curved spline.

Connection Line



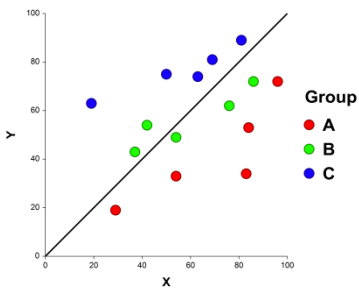
Spline



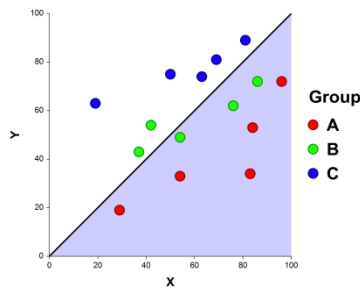
45° Line

You can connect the symbols with a line. Along with setting the format of the connecting line, this section also lets you specify the order in which the points are connected. The connection can be done with a straight line or a curved spline.

45° Line



45° Line with Fill



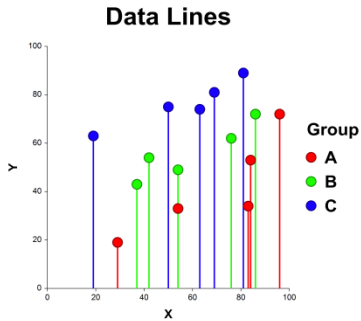
Points-to-Axis Tab

Tick Marks for each Data Point Section

You can display a tick mark along each axis for each point.

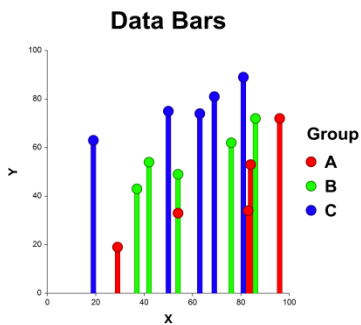
Lines from each Data Point Section

You can display a line between the data point and an axis.



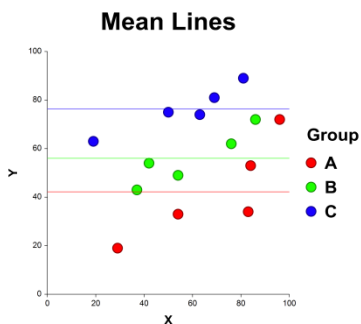
Bars from each Data Point Section

You can display bars between data points and an axis.



Mean Extension Lines Section

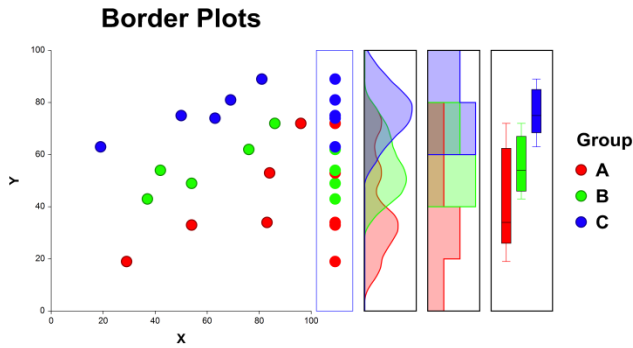
You can display horizontal and vertical lines at the group means.



Border Plots Tab

X and Y Axis Sections

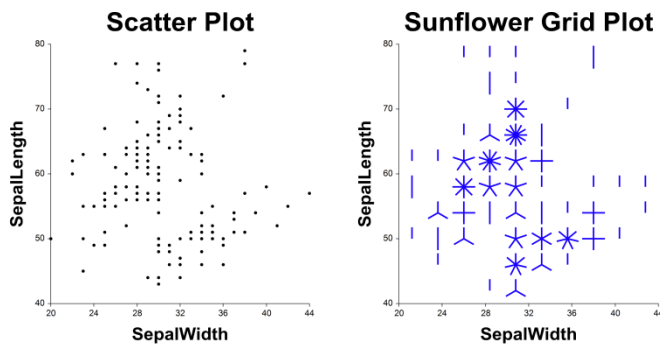
You can display box plots, density plots, dot plots, and histograms in the border of each axis.



Sunflower Tab

Sunflower Density Section

You can display a sunflower grid over the points on the plot. This technique was developed for when you have many, many overlapping points on the plot. Instead of showing the points, flowers are shown in which the number of petals is proportional to the number of points falling in the grid cell.



Titles, Legend, X Axis, Y Axis, Grid Lines, and Background Tabs

Details on setting the options in these tabs are given in the Graphics Components chapter.

Example 1 – Creating a Scatter Plot

This section presents an example of how to generate a simple scatter plot. The data used are from the Size dataset. We will create a scatter plot of variables *Weight* versus *Height*.

Setup

To run this example, complete the following steps:

1 Open the Size example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Size** and click **OK**.

2 Specify the Scatter Plots procedure options

- Find and open the **Scatter Plots** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Y (Vertical) Variable(s).....**Weight**

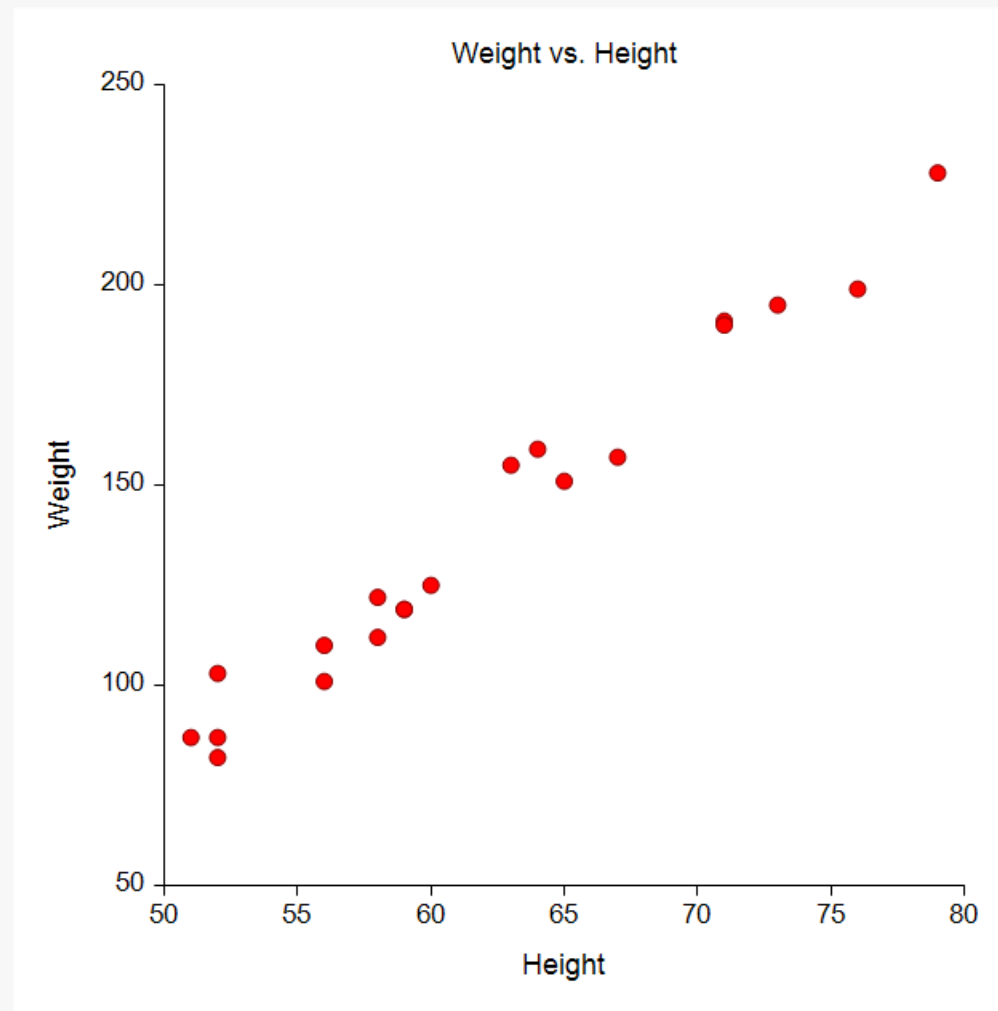
X (Horizontal) Variable(s)**Height**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Scatter Plot Output

Scatter Plot Section



Example 2 – Scatter Plot with a Regression Line and Prediction Limits

This section presents an example of how to generate a simple scatter plot with a regression line and prediction limits. The data used are from the Size dataset. We will create a scatter plot of variables *Weight* (Y) versus *Height* (X).

Setup

To run this example, complete the following steps:

1 Open the Size example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Size** and click **OK**.

2 Specify the Scatter Plots procedure options

- Find and open the **Scatter Plots** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Y (Vertical) Variable(s).....**Weight**

X (Horizontal) Variable(s)**Height**

Scatter Plot Format (*Click the Button*)

Scatter Plot Tab

Regression Line.....**Checked**

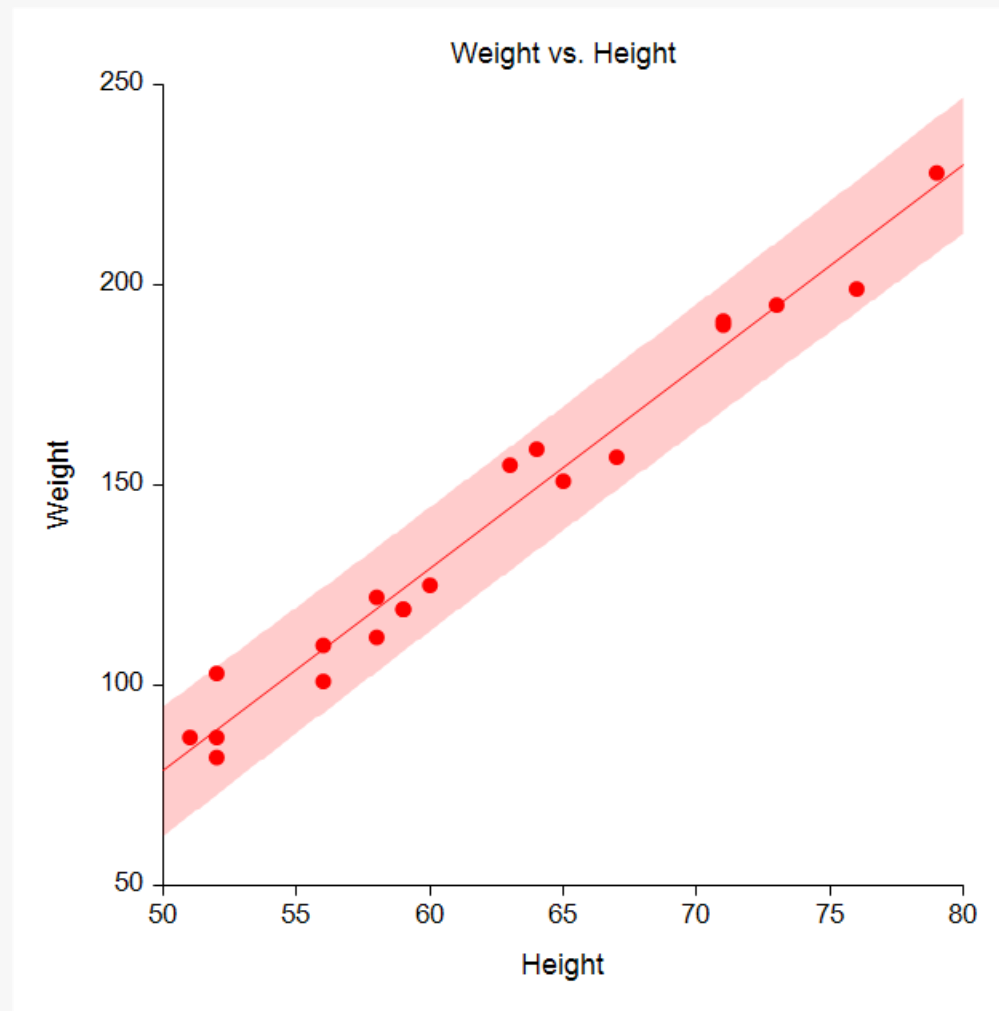
Prediction Limits – Fill.....**Checked**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Scatter Plot Output with Regression Line and Prediction Limits

Scatter Plot Section



Example 3 – Scatter Plot with a Regression Line, Prediction Limits, and Border Plots

This section presents an example of how to generate a simple scatter plot with a regression line, prediction limit, and border plots (box plots and dot plots). The data used are from the Size dataset. We will create a scatter plot of variables *Weight* (Y) versus *Height* (X).

Setup

To run this example, complete the following steps:

1 Open the Size example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Size** and click **OK**.

2 Specify the Scatter Plots procedure options

- Find and open the **Scatter Plots** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 3** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Y (Vertical) Variable(s).....**Weight**
 X (Horizontal) Variable(s)**Height**

Scatter Plot Format (*Click the Button*)

Scatter Plot Tab

Regression Line.....**Checked**
 Prediction Limits – Fill.....**Checked**

Border Plots Tab

X Axis

Box Plot.....**Checked**
 Density.....**Checked**
 Dot Plot.....**Checked**

Y Axis

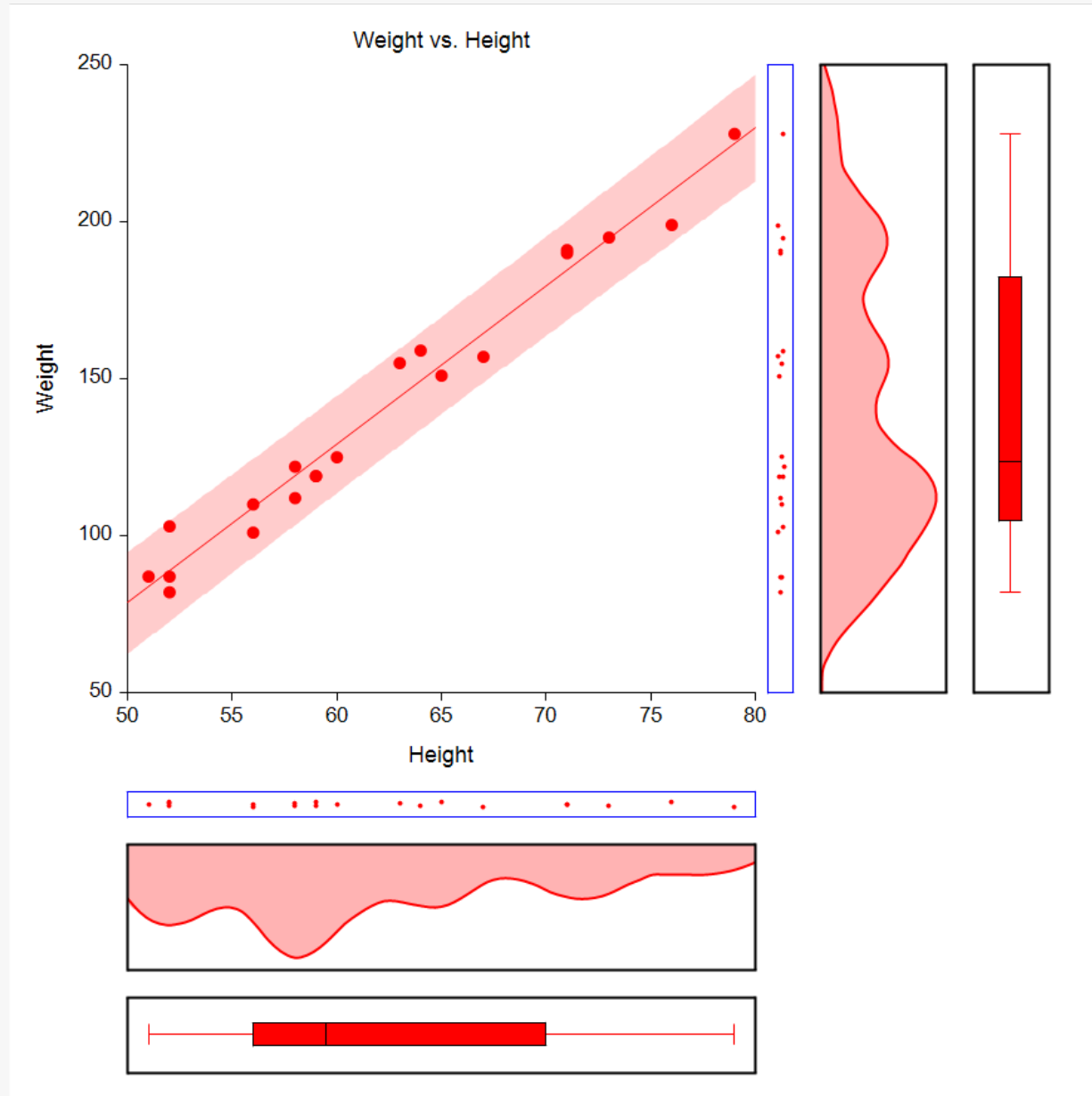
Box Plot.....**Checked**
 Density.....**Checked**
 Dot Plot.....**Checked**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Scatter Plot Output with Regression Line, Prediction Limits, and Border Plots

Scatter Plot Section



Example 4 – Scatter Plot with a Loess Line

This section presents an example of how to generate a simple scatter plot with a Loess line. The data used are from the Size dataset. We will create a scatter plot of variables *Weight* (Y) versus *Height* (X).

Setup

To run this example, complete the following steps:

1 Open the Size example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Size** and click **OK**.

2 Specify the Scatter Plots procedure options

- Find and open the **Scatter Plots** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 4** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Y (Vertical) Variable(s).....**Weight**

X (Horizontal) Variable(s)**Height**

Scatter Plot Format (*Click the Button*)

More Lines Tab

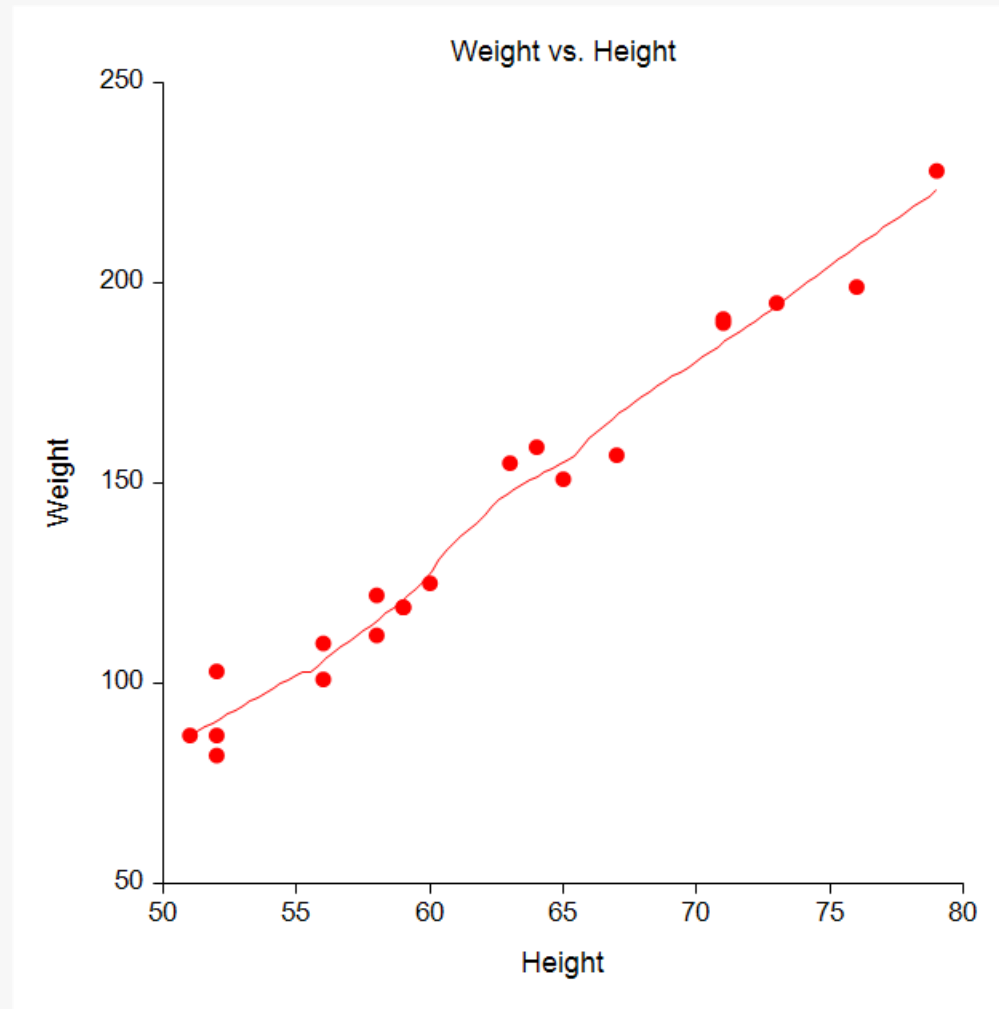
Loess**Checked**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Scatter Plot Output with Loess Line

Scatter Plot Section



Example 5 – Scatter Plot with 3D Symbols of Varying Sizes and Gradient Interior

This section presents an example of how to generate a scatter plot with 3D symbols of varying sizes and a gradient-filled interior background. The data used are from the FISHER dataset. We will create a scatter plot of variables *SepalLength* (Y) versus *SepalWidth* (X) by *Iris* (G).

Setup

To run this example, complete the following steps:

1 Open the Fisher example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Fisher** and click **OK**.

2 Specify the Scatter Plots procedure options

- Find and open the **Scatter Plots** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 5** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Y (Vertical) Variable(s).....**SepalLength**
 X (Horizontal) Variable(s)**SepalWidth**
 Grouping (Symbol) Variable.....**Iris**

Scatter Plot Format (*Click the Button*)

Scatter Plot Tab

Symbol Format (*Click the Button*)

Click on the right side of **symbol 1** (the red circle) to display the Preset Symbols.

- Select the red 3D symbol in row **7** column **3**.
- Click on the right side of symbol **2** (the blue circle).
- Select the blue 3D symbol in row **7** column **8**.
- Click on the right side of symbol **3** (the green circle).
- Select the green 3D symbol in row **7** column **6**.
- Click the **Change All** button to bring up the **Change All Symbol Properties** window.
- Check the **Symbol Size** and **Border Width** boxes to display those options.
- Change the **Symbol Size** to **135**.
- Change the **Border Width** to **1**.
- Click **OK**.

Click **symbol 2** (the blue 3D circle) to display the Symbols Format (Group 2) window.

- Change the **Symbol Size** to **200**.
- Click **OK**.

Scatter Plots

Background Tab

Interior Fill Format (*Click the Button*)

Click on the right side of **symbol 1** (the red circle) to display the Preset Symbols.

- Click on the **Quick Fills** button at the top of the window.
- Select the light blue gradient in row **6** column **7**.
- The **Gradient Stops** options will now be displayed.
- Click on the right side of the **Color** button (it is light blue).
- Select the light blue color in row **1** column **7**.
- Click **OK**.

Report Options (*in the Toolbar*)

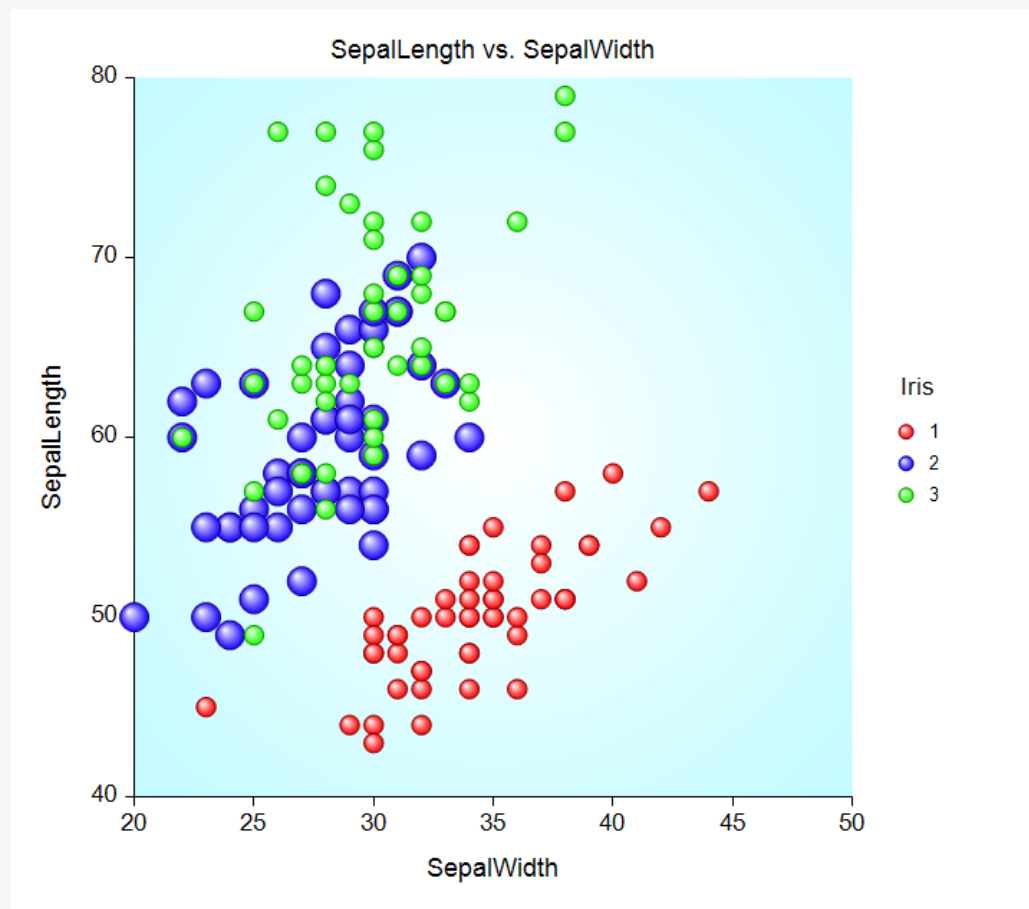
Variable Labels.....**Column Names**
 Data Labels.....**Data Values**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Scatter Plot Output with 3D Symbols and Gradient Background

Scatter Plot Section



Example 6 – Scatter Plot with Probability Ellipse

This section presents an example of how to generate a scatter plot with overlaid probability ellipses. The data used are from the FISHER dataset. We will create a scatter plot of variables *SepalLength* (Y) versus *SepalWidth* (X) by *Iris* (G).

Setup

To run this example, complete the following steps:

1 Open the Fisher example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Fisher** and click **OK**.

2 Specify the Scatter Plots procedure options

- Find and open the **Scatter Plots** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 6** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Y (Vertical) Variable(s).....**SepalLength**

X (Horizontal) Variable(s)**SepalWidth**

Grouping (Symbol) Variable.....**Iris**

Scatter Plot Format (*Click the Button*)

Scatter Plot Tab

Probability Ellipse**Checked**

Probability Ellipse – Fill.....**Checked**

Report Options (*in the Toolbar*)

Variable Labels.....**Column Names**

Data Labels.....**Data Values**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Scatter Plot Output with Probability Ellipses

Scatter Plot Section

