SAS® Studio 3.3
User's Guide

Contents

Using This Book

Audience

This book is designed for all users of SAS Studio. SAS Studio was initially released with the first maintenance release for SAS 9.4. SAS Studio 3.3 is the latest release.

What's New

What's New in SAS Studio 3.3

Overview

SAS Studio 3.3 includes these new features and enhancements:

- support for the Windows operating environment in SAS Studio Basic

- new visual programming perspective, which enables you to create process flows

- new tool to create queries

- new and enhanced tasks

- new variable name and encoding options

Support for Windows Operating Environments in SAS Studio Basic

SAS Studio Basic is now supported in both Windows and UNIX operating environments. For more information, see *SAS Studio: Administrator's Guide*.

New Visual Programming Perspective

This new perspective enables you to create process flows. A process flow consists of one or more objects. Each object is represented by a node in the process flow. The process flow shows the relationship between two or more objects, such as a SAS program, a task, a query, and so on. For more information, see Chapter 4, "Working with Process Flows," on page 72.

New Query Feature

You can use the query feature to extract data from one or more tables according to criteria that you specify. You can create a query that is based on only one table, or you can join tables together. When you create a query, you specify the columns to include in the output, and you can filter the rows that are included based on values in the data. You can also perform a summary function on the columns in your data and sort your data by one or more columns. For more information, see Chapter 3, "Working with Queries," on page 48.

SAS Studio Tasks

New Tasks

- The new Analysis of Covariance task enables you to augment the analysis of variance model with a quantitative variable in order to reduce the variance of the error terms in the model. For more information, see "Analysis of Covariance Task" on page 348.

■ The new Bubble Plot task explores the relationship between three or more variables. In a bubble plot two variables determine the location of the bubble centers, and a third variable specifies the size of each bubble. For more information, see "Bubble Plot Task" on page 210.

■ There are six new Combinatorics and Probability tasks:

  ☐ The new Combinations task computes the possible combinations of the total number of objects into sets with the specified number in each set. For more information, see "Combinations Task" on page 131.

  ☐ The Same Birthday Probability task computes the probability that two or more people in a room have the same birthday. For more information, see "Same Birthday Probability Task" on page 266.

  ☐ The Coin Toss Simulation task simulates the tossing of 10 coins and calculates the probabilities for the possible outcomes. For more information, see "Coin Toss Simulation Task" on page 268.

  ☐ The Dice Roll Simulation task simulates the tossing of two dice and calculates the probabilities for the possible outcomes. For more information, see "Dice Roll Simulation Task" on page 270.

  ☐ The Poker Hand Probability task calculates the frequency and probability of poker hands. For more information, see "Poker Hand Probability Task" on page 272.

  ☐ The new Permutations task computes the possible permutations of a given number of objects. For more information, see "Permutations Task" on page 138.

■ The new Generalized Linear Models task enables you to create classical linear models with normal errors, logistic and probit models for binary data, and log-linear models for multinomial data. For more information, see "Generalized Linear Models" on page 396.

■ The new N-way ANOVA task performs an N-way analysis of variance. For more information, see "N-Way ANOVA Task" on page 343.

■ To perform a one-sample t test, paired sample t test, or two-sample t test, use the new T Test task. For more information, see "T Tests Task: One-Sample t Test" on

Enhanced Tasks

- In the High-Performance Generalized Linear Model task, you can now specify an offset variable to use as an offset to the linear predictor. The new **Dispersion parameter** option enables you to specify a fixed dispersion parameter for distributions that have a dispersion parameter. For more information, see "Generalized Linear Models" on page 248.

- The One-Way ANOVA task now includes these comparison methods: Dunnett two-tail, Dunnett lower one-tail, and Dunnett upper one-tail. For more information, see "One-Way ANOVA Task" on page 330.

Updates to the Custom Task Model

If you create custom tasks for your site, here are the changes to the Common Task Model (CTM) 3.3.

- The new `dualselector` input type creates a control that displays a list of default values for the option. However, the user can change the list of values by selecting from a list of values that appear in a separate window. The `dualselector` input type can use the `OptionChoice` element in the user interface.

- The new `outputdata` input type creates a text box where the user can specify the name of the output data set that is created by the task.

- The `required` attribute is now available for the `combobox`, `distinct`, `color`, and `numstepper` input types. Now for these input types, you can specify whether an option is required to run the task.

- For the `combobox` and `distinct` input types, the new `selectMessage` attribute specifies the message to display when a value is required for the combobox control, and no default value has been set. The default message is `Select a value`.

■ For the `multientry` input type, the new `reorderable` attribute specifies whether the user can reorder the values in a list.

For more information, see *SAS Studio: Developer's Guide*.

New Variable Name and Encoding Options

The new **SAS variable name policy** option enables you to specify the set of rules that you want to use for SAS variable names. The **Default text encoding** option specifies the character-set encoding that is used when text files are read or written. For more information, see "Setting General Preferences" on page 15.

Accessibility

For information about the accessibility of this product, see Accessibility Features of SAS Studio 3.3.

1

Introduction to SAS Studio

About SAS Studio

SAS Studio is a development application for SAS that you access through your web browser. With SAS Studio, you can access your data files, libraries, and existing programs, and you can write new programs. You can also use the predefined tasks in SAS Studio to generate SAS code. When you run a program or task, SAS Studio connects to a SAS server to process the SAS code. The SAS server can be a hosted server in a cloud environment, a server in your local environment, or a copy of SAS on your local machine. After the code is processed, the results are returned to SAS Studio in your browser.

SAS Studio supports multiple web browsers, such as Microsoft Internet Explorer, Apple Safari, Mozilla Firefox, and Google Chrome.

In addition to writing and running your own SAS programs, you can use the predefined tasks that are included with SAS Studio to analyze your data. The tasks are based on SAS System procedures and provide access to some of the most commonly used graph and analytical procedures. You can also use the default task template to write your own tasks.

SAS Studio includes two different perspectives: the SAS Programmer perspective and the Visual Programmer perspective. A *perspective* is a predetermined set of features that is customized to meet the needs of a specific user type. By selecting a specific perspective, you can narrow the choices that are available in the interface and focus on the features that you need to use regularly. By default, when you open SAS Studio, the SAS Programmer perspective is selected. After you open SAS Studio, you can change the perspective by using the perspectives menu on the toolbar. For more information, see "Understanding Perspectives" on page 16.

Using SAS Studio

About Using SAS Studio

When you sign on to SAS Studio, the main SAS Studio window appears with a blank program window so that you can start programming immediately. You also have access to all five sections of the navigation pane.

Note: To sign out of SAS Studio, click the **Sign Out** button on the toolbar. Do not use the Back button on your web browser.
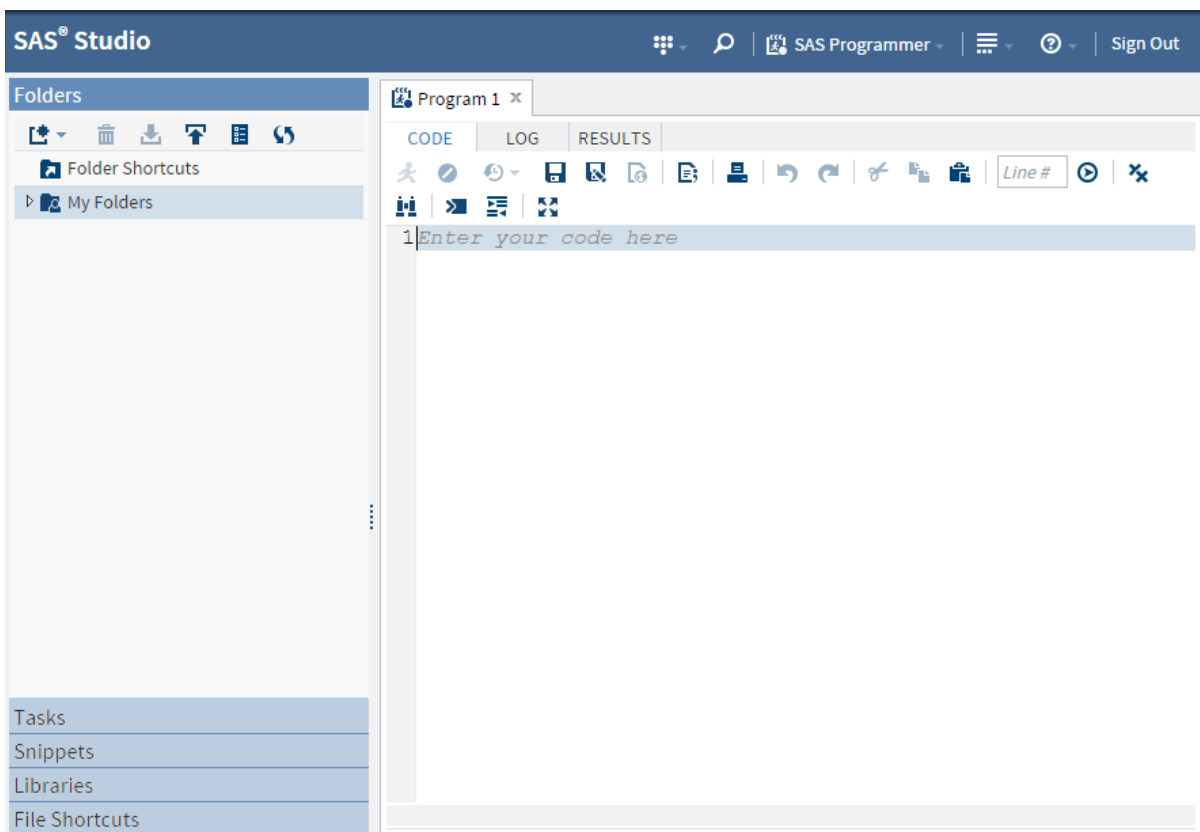
The main window of SAS Studio consists of a navigation pane on the left and a work area on the right. The navigation pane provides access to your folder shortcuts and

folders, your tasks and snippets, the libraries that you have access to, and your file shortcuts. The Folders section is displayed by default.

The work area is used to display your data, code, tasks, logs, and results. As you open these items, they are added to the work area as windows in a tabbed interface.

Using the Navigation Pane

About Using the Navigation Pane

You can expand the sections of the navigation pane by clicking the section that you want to view.

Working with Folders

The Folders section of the navigation pane enables you to access files and folders from the following locations:

- your SAS server

- any remote FTP server on which you have an account

Note: SAS Studio supports only FTP servers that use a UNIX style directory listing. Filenames in Asian languages are not supported on the FTP servers.

The type of SAS Studio deployment you are using determines the folders and files you can access from the Folders section and the locations to which you can assign libraries.

- If you are running the SAS Studio Mid-Tier (the enterprise edition) deployment, then your **My Folders** location is your home directory on the remote SAS server.

- If you are running the SAS Studio Basic deployment, then your **My Folders** location is your home directory on the remote SAS server. The folders list also includes the root folder of the UNIX file system on the remote SAS server.

- If you are running the SAS Studio Single-User deployment, then your SAS server is your local machine. Your **My Folders** location is your home directory on your local machine. The folders list also includes folders for your desktop, documents folder, and all currently mapped drives.

For more information, see *SAS Studio: Administrator's Guide*.

You can open files that are saved on the SAS server or the FTP server, such as SAS program files or program package files. You can also open SAS tables that are saved on the SAS server.

Note: You cannot open SAS tables from a remote FTP server.

You can use the Folders section to create folders and folder shortcuts, download and upload files, and create a new SAS program. From the folders tree, you can expand and collapse folders and open items in folders by double-clicking them or dragging them to the work area.

Note: Files and folders that are located on an FTP server and are accessible by using an FTP shortcut cannot contain any of the following characters in their names:

{ } [ ] , : ; " ' / \ | + < > ? @ # $ % ^ & ( ) + !

SAS Studio cannot access, move, rename, or delete files and folders on an FTP server whose names contain invalid characters.

To create a new folder shortcut:

1 Click **Folders** in the navigation pane. Then click and select **Folder Shortcut**. The New Folder Shortcut window opens.

2 In the **Name** box, enter the name of the folder.

3 From the **Folder Type** drop-down list, specify whether the shortcut refers to a SAS server folder or an FTP folder.

4 If you are creating a shortcut to a SAS server folder, enter the physical path for the directory in the **Directory** box.

If you are creating a shortcut to an FTP folder, enter the network address of the FTP host in the **Host Name** box as well as your user name and password. By default, the directory is the home directory of the FTP user. You can use the **Directory** box to specify another directory that is relative to the home directory. For example, if the home directory of the FTP user is `c:\homedir`, and you specify `data` in the **Directory** box, then the root directory of the shortcut is `c:\homedir\data` on the FTP server. You can validate your connection to the FTP server by clicking **Test**.

Note: If your mid-tier and FTP servers are running different operating systems, you must fully qualify the name of the FTP host in the **Host Name** box. If this name is not fully qualified, then the connection might fail.

New Folder Shortcut	✕
Name:	FTPshortcut
Folder Type:	FTP Folder ▾
Host Name:	ftpserver
User ID:	myuser
Password:	••••••••
Directory:	
	Test
	Save Cancel

5 Click **Save** to create the folder shortcut. The new shortcut is added to the list of folder shortcuts.

To create a new folder, select the folder in the Folders section in which you want to create the new folder. Click ⬒ and select **Folder**. The New Folder window opens. Enter the name of the new folder. The new folder is added to the list of folders.

To download a file, select the file that you want to download and click ⬇. You are prompted to open the file in the default application or save it to your local computer.

To upload one or more files from your local computer, select the folder to which you want to upload the files and click ⬆. The Upload Files window opens. Click **Choose Files** to browse for the files that you want to upload.

Working with Tasks

The Tasks section of the navigation pane enables you to access tasks in SAS Studio. Tasks are based on SAS procedures and generate SAS code and formatted results for you. SAS Studio is shipped with several predefined tasks that you can run. You can also edit a copy of these predefined tasks, and you can create your own new tasks.

To create a new task, click ⊞. SAS Studio creates a template in the work area that you can use to create custom tasks for your site. Custom tasks can be accessed from the My Tasks folder. For more information, see Chapter 7, "Understanding Tasks in SAS Studio," on page 115.

To edit a task that you have created, select the task from the My Tasks folder and click

. The XML code that is used to create the task is opened in the work area. If you want to edit a predefined task, you must first right-click the task and select **Add to My Tasks**. For more information, see "Edit a Predefined Task" on page 119.

Note: You can edit only the tasks that are in the My Tasks folder.

Working with Snippets

The Snippets section of the navigation pane enables you to access your code snippets. *Code snippets* are samples of commonly used SAS code that you can insert into your SAS program. SAS Studio is shipped with several predefined code snippets that you can use. You can also edit a copy of these snippets and create your own custom snippets. Your custom snippets can be accessed from the My Snippets folder. For more information, see Chapter 2, "Working with Programs," on page 22.

To edit a snippet that you have created, select the snippet from the My Snippets folder

and click . If you want to edit a predefined snippet, you must first right-click the snippet and select **Add to My Snippets**.

Note: You can edit only the snippets that are in the My Snippets folder.

Working with Libraries

The Libraries section of the navigation pane enables you to access your SAS libraries. SAS tables are stored in SAS libraries. From the Libraries section, you can open SAS tables and add them to your programs. You can use the Libraries section to expand a

table and view the columns in that table. The icon in front of the column name indicates the type.

Here are examples of common icons for the column types.

Icon	Type of Column
🔺	Character
123	Numeric
🗓	Date
🗓	Datetime

You can drag tables and columns from the Libraries section to a program, and SAS Studio adds code for the dragged items to your program. For more information, see "Opening and Creating Programs" on page 22.

You can also create new libraries and assign existing libraries.

To create a new library:

1   Click **Libraries** in the navigation pane and then click 📑. The New Library window appears.

2 In the **Name** box, enter the libref for the library. The libref must be eight characters or fewer.

3 In the **Path** box, enter the physical path where the library resides.

4 In the **Options** box, specify any configuration options that you need. For the appropriate options, see the documentation for your operating environment.

5 If you want to access this library each time you use SAS Studio, select **Re-create this library at start-up**.

6 Click **OK** to create the library. The new library is added to the list of libraries in the navigation pane.

To assign unassigned libraries, click ▦. If you want to access the selected libraries each time you use SAS Studio, select **Assign selected libraries at start-up**. If a library is unassigned, then you cannot access the tables in that library.

Using File Shortcuts

File shortcuts enable you to quickly access files that you specify. You can create a file shortcut to a file on your SAS server, via a URL, or on your FTP server.

Note: You can create a file shortcut to a file on an FTP server only if you have created a folder shortcut to an FTP folder.

To create a new file shortcut, click . You can define the shortcut by specifying a complete path and filename or by specifying a URL. If you want this shortcut to be available the next time you use SAS Studio, select **Re-create this file shortcut at start-up**.

You can open a file from a file shortcut by double-clicking it or dragging it to the work area.

Customizing the Navigation Pane

By default, all five sections of the navigation pane are displayed when you open SAS Studio. To customize which sections are displayed, click and select **View**. Select or clear any sections that you want to add or remove. The navigation pane is updated immediately.

Using the Work Area

About Using the Work Area

The work area is the main portion of the SAS Studio application for accessing programs and tasks and for viewing data. The work area is always displayed and cannot be minimized. When you open a program, task, or table, the windows open as new tabs in the work area. The code, log, and results that are associated with programs and tasks are grouped together under the main tab for the program or task.

Program 1 ×   SASHELP.CARS ×   List Data ×   *Bar Chart ×

Settings Code/Results Split | ⚡ ⊘ 🖫 🖺 ⛶ | Log Code

| DATA | OPTIONS | INFORMATION |

◢ DATA

SASHELP.CLASS ▼

▸ WHERE CLAUSE FILTER

◢ ROLES

*Category variable: *(1 item)* 🗑 ➕

   123 Age

Response variable: *(1 item)* 🗑 ➕

   123 Height

Group variable: *(1 item)* 🗑 ➕

   123 Age

URL variable: *(1 item)* 🗑 ➕

   Ⓐ Column

BY variable: *(1 item)* 🗑 ➕

   ◆ Column

▸ DIRECTION

▸ GROUP LAYOUT

▸ STATISTICS

| CODE | LOG | RESULTS |

🕘 ▾   📖 📄 🖥 📑 Line # ▶ ▮▮ Edit

```
12  *
13  */
14
15 /*--Set output size--*/
16 ods graphics / reset imagemap;
17
18 /*--SGPLOT proc statement--*/
19 proc sgplot data=SASHELP.CLASS noauto
20     /*--TITLE and FOOTNOTE--*/
21     title "Class Bar Chart";
22
23     /*--Bar chart settings--*/
24     vbar Age / response=Height group=
25         stat=Mean name='Bar';
26
27     /*--Category Axis--*/
28     xaxis discreteorder=data reverse;
29
30     /*--Response Axis--*/
31     yaxis grid;
32
33     /*--Legend Settings--*/
34     keylegend 'Bar' / title='Age:' lc
35 run;
36
37 ods graphics / reset;
38
```

Line 1, Column 1

Customizing the Work Area

By default, the work area is displayed beside the navigation pane, but you can maximize the work area and hide the navigation pane. You can also close all of the tabs in the work area at once.

To maximize the work area, click ▦ and select **Maximize View**.

Note: To reopen the navigation pane, click ▦ and select **Exit Maximized View**.
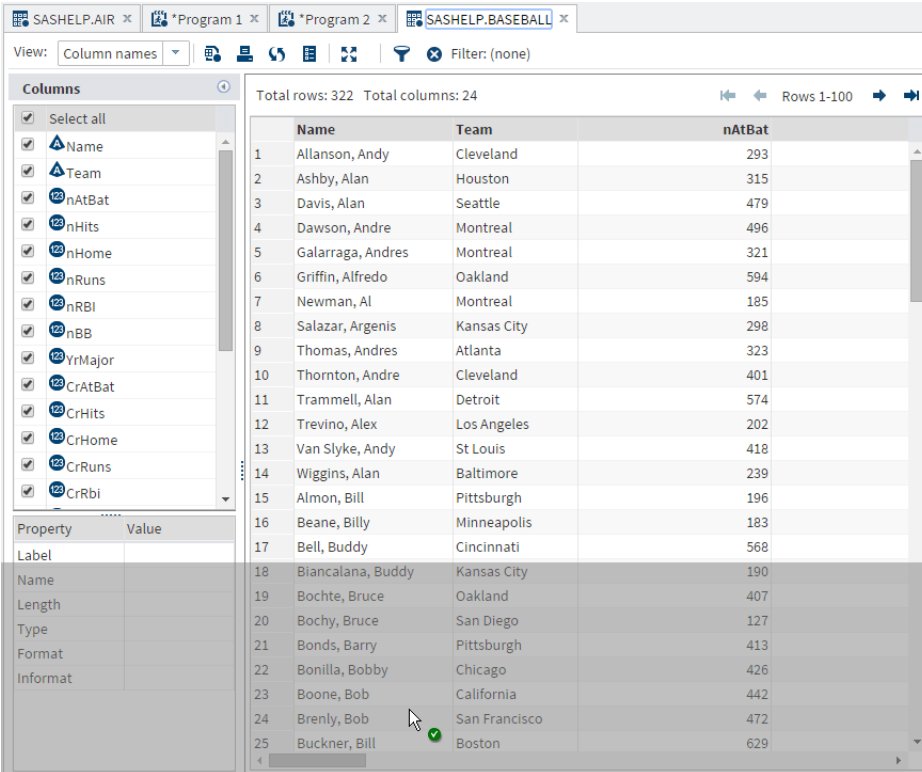
To close all tabs that are open in the work area, click and select **Close All Tabs**. You are prompted to save any unsaved programs or tasks.

Rearranging the Tabs in the Work Area

In the work area, you can rearrange the tabs by using a drag-and-drop operation to move them to the left or right. You can also dock a tab on the right side or bottom of the work area to view more than one tab at a time.

To rearrange a tab:

1   Select the tab that you want to move.

2   Move the tab icon to the location where you want to view this content. The icon indicates a valid location.

Note: The Results tab in Program 2 has also been moved to the right side of the program tab.
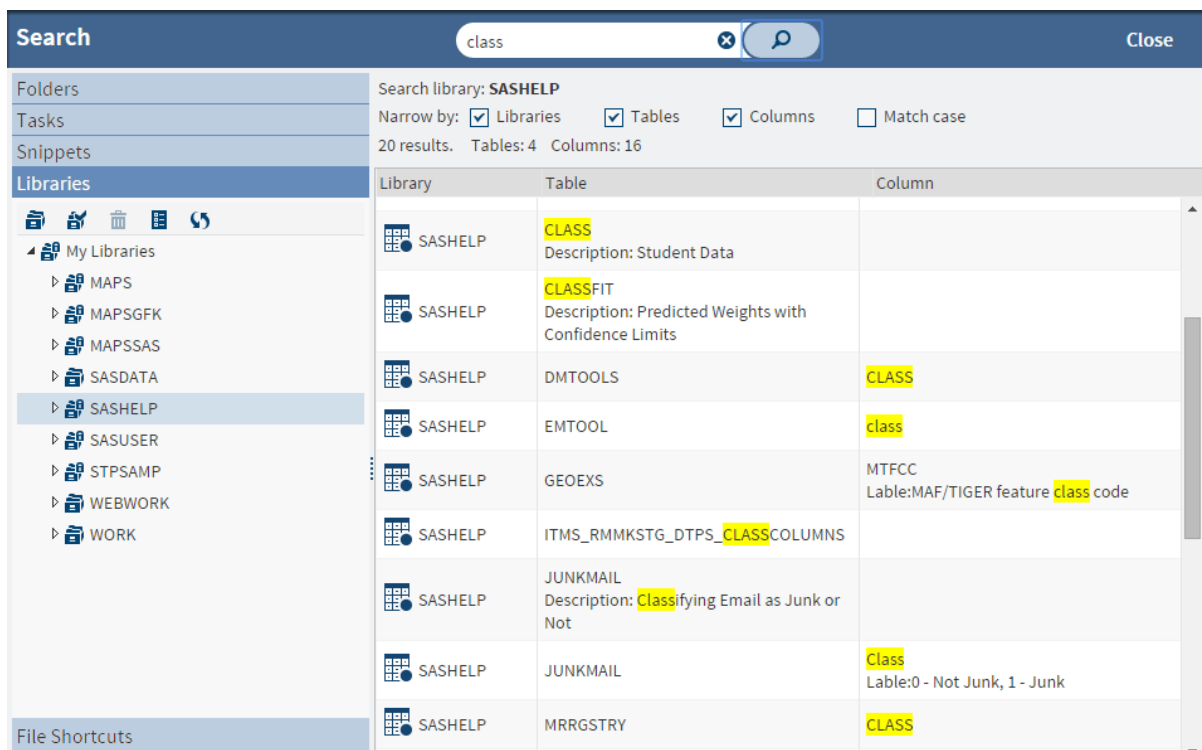
Searching in SAS Studio

You can use the Search feature to search all of the sections of the navigation pane. The types of items that you can search for depend on what is selected in the navigation pane. For example, if a folder or folder shortcut in the Folders section is selected, you can search for folders and files, and you can choose whether to include subfolders in the search. If a task category in the Tasks section is selected, you can search for task names, associated SAS procedures, and task descriptions.

To access the Search feature, click 🔎. The search box opens so that you can enter the text that you want to search for.

You can limit the scope of your search by selecting or clearing any of the **Narrow by** options. These options vary depending on what is selected in the navigation pane. By default, the search is case sensitive. To search for both uppercase and lowercase text, clear the **Match case** check box.

The following example shows a search for "class" in the Sashelp library. The search includes all tables and columns in the Sashelp library, and it is not case sensitive.

Search					Close
		class	✕ 🔍		
Folders	Search library: **SASHELP**				
Tasks	Narrow by: ☑ Libraries ☑ Tables ☑ Columns ☐ Match case				
Snippets	20 results. Tables: 4 Columns: 16				
Libraries	Library	Table		Column	
▲ My Libraries	SASHELP	CLASS Description: Student Data			
▷ MAPS	SASHELP	CLASSFIT Description: Predicted Weights with Confidence Limits			
▷ MAPSGFK					
▷ MAPSSAS	SASHELP	DMTOOLS		CLASS	
▷ SASDATA					
▷ SASHELP	SASHELP	EMTOOL		class	
▷ SASUSER	SASHELP	GEOEXS		MTFCC Lable:MAF/TIGER feature class code	
▷ STPSAMP					
▷ WEBWORK	SASHELP	ITMS_RMMKSTG_DTPS_CLASSCOLUMNS			
▷ WORK	SASHELP	JUNKMAIL Description: Classifying Email as Junk or Not			
	SASHELP	JUNKMAIL		Class Lable:0 - Not Junk, 1 - Junk	
File Shortcuts	SASHELP	MRRGSTRY		CLASS	

You can open an item from your results by double-clicking it. To return to your search results list, click 🔎 again.

To clear the search results, click ✕.

Setting General Preferences

The Preferences window enables you to change several options that affect SAS Studio.

To access the general options, click ⚏ and select **Preferences**. Click **General**.

Option	Description
Show generated code in the SAS log	displays the ODS statements, %LET statements, and any other code that is automatically generated by SAS in the log file. This option applies to both SAS tasks and SAS program files.
Include a Show Details button in error messages	adds a Show Details button to any error messages that SAS Studio generates.
Start new programs in interactive mode	opens new programs with the interactive mode on. This option is available only if you are running the first maintenance release for SAS 9.4. For more information, see "Working in Interactive Mode" on page 32.
Time-out interval: (hours)	specifies the amount of time that SAS Studio allows you to be logged on without any activity. The default value is one hour.

Option	Description
SAS variable name policy	enables you to specify one of the following sets of rules to apply to SAS variable names.
	■ ANY — specifies that the variable names can begin with or contain any characters, including blanks, must contain at least one character, and cannot contain any null bytes. Variable names can contain mixed-case letters as well as special and multi-byte characters. Names can be up to 32 bytes in length. This option is selected by default. Leading blanks are preserved, but trailing blanks are ignored.
	■ V7 — specifies that the variable names must begin with a letter of the Latin alphabet (A-Z, a-z) or the underscore character. They cannot contain blanks or special characters except for the underscore and cannot be assigned the names of special SAS automatic variables or variable list names. Variable names can contain mixed-case letters and can be up to 32 bytes in length.
Default text encoding	specifies the character-set encoding that is used when text files are read or written. The default value is UTF-8. This option is not available for SAS servers running on z/OS.
	Note: You can also specify the text encoding when you open a single file by holding down the Alt key and double-clicking the file. The Choose Text Encoding window opens and you can select the encoding. This option applies to SAS program files (*.SAS), text files (*.txt), comma-separated values files, (*.csv), and log (*.log) file types.

Understanding Perspectives

Because SAS Studio can be used by a variety of people and groups within an organization, you can choose to view a specific subset of features, or perspective, that

meets your needs best. Perspectives are sets of functionality that are customized to meet the needs of different types of users. SAS Studio includes two perspectives: the SAS Programmer perspective and the Visual Programmer perspective.

The SAS Programmer perspective is for users who intend to use SAS Studio mainly for writing and editing SAS programs. By default, the SAS Programmer perspective opens with a new program window and includes all five sections of the navigation pane so that you can easily manage all of your program files and code snippets.

The Visual Programmer perspective is designed for users who want to work with process flows in a project-based environment. You can use process flows in the Visual Programmer perspective to combine individual processes into one repeatable process flow that you can save, reuse, and share with other users. By default, the Visual Programmer perspective opens with a new process flow window and includes four sections in the navigation pane: Folders, Tasks, Snippets, and Libraries.

Note: If you open a process flow while you are using the SAS Programmer perspective, you are prompted to switch to the Visual Programmer perspective.

In both perspectives, you can run the predefined tasks that are shipped with SAS Studio, and you can create and edit tasks. You can also use both perspectives to create and run SAS programs and queries.

The differences between the perspectives can be viewed in the following table:

Element in SAS Studio	Available in SAS Programmer?	Available in Visual Programmer?
Navigation Pane sections		
▪ Folders	Yes	Yes
▪ Tasks	Yes	Yes
▪ Snippets	Yes	Yes
▪ Libraries	Yes	Yes
▪ File Shortcuts	Yes	Yes, but not displayed by default

Element in SAS Studio	Available in SAS Programmer?	Available in Visual Programmer?
SAS Programs	Yes	Yes
Queries	Yes	Yes
Process Flows	No	Yes

After you have started SAS Studio and selected a perspective, you can change the perspective that you are using by clicking the perspectives menu on the toolbar and selecting the perspective that you want to use.

You can specify which sections of the navigation pane are displayed in SAS Studio by clicking ▤ and selecting **View**.

Editing the Autoexec File

The autoexec.sas file includes SAS statements that run each time you start SAS Studio and connect to your SAS server. For example, you can use the autoexec.sas file to assign libraries that you want to be available every time you use SAS Studio in both interactive and noninteractive modes.

Note: If you create a new library by using the New Library window, you can select the **Re-create this library at start-up** option to automatically add the LIBNAME statement to the autoexec.sas file. For more information, see "Working with Libraries" on page 7.

To edit the autoexec.sas file:

1   Click ▤ and select **Edit Autoexec File**.

2   Enter the code that you want to include in the autoexec.sas file.

3   To validate your syntax, click **Run**. The **Log** tab opens so that you can view the log.

4   Click **Save** to save and close the autoexec file.

Changing Your SAS Workspace Server

If you have access to more than one SAS workspace server, you can change the server that SAS Studio connects to. To change the server, click ≡ and select **Change SAS Workspace Server**. Select the server that you want to use. When you change servers, any libraries and file shortcuts that you created are deleted. For more information, see *SAS Studio: Administrator's Guide*.

2

Working with Programs

About the Code Editor

SAS Studio includes a color-coded, syntax-checking editor for editing new or existing SAS programs. You can also edit SOURCE entries in SAS catalogs. The editor includes a wide variety of features such as autocomplete, automatic formatting, and pop-up syntax help. With the code editor, you can write, run, and save SAS programs. You can also modify and save the code that is automatically generated when you run a task.

SAS Studio also includes several sample code snippets that you can use to make programming common tasks easier.

Opening and Creating Programs

Opening a Program

You can open SAS programs from the Folders section of the navigation pane. To open a program, expand the appropriate folder and double-click the program that you want to open, or drag it into the work area. The program opens in a new tab in the work area.

Creating a New Program

You can create a new SAS program from the Folders section of the navigation pane. To create a new program, click and select **SAS Program**. A program window appears in a new tab in the work area.

Note: You can also click on the main application toolbar and select **New SAS Program**.

Running a Program

After you have written your program, you can run it by clicking .

If there are no errors, the results open automatically. If there are errors, the **Log** tab opens by default. You can expand the **Errors**, **Warnings,** and **Notes** sections to view the messages. When you click on a message, SAS Studio highlights it for you in the log so that you can see exactly where the message occurs in the log.
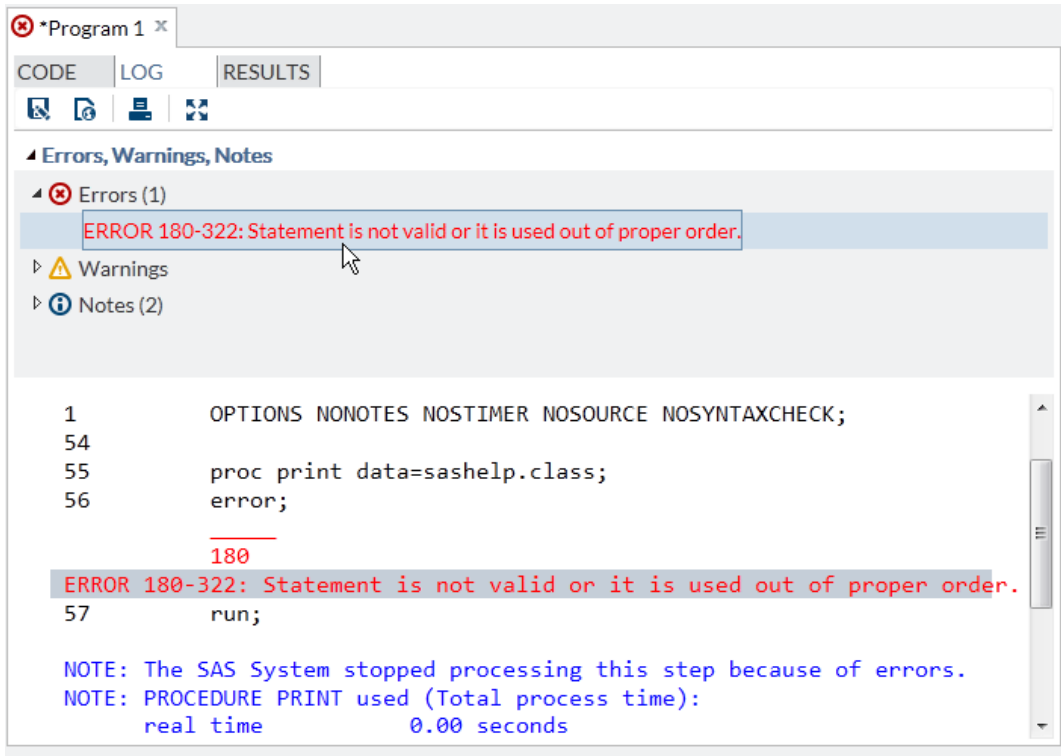
Note: Because you are working in a server environment, do not include the ENDSAS statement in your SAS programs. If you run a program that contains ENDSAS, reset your SAS session by clicking ▦ and selecting **Reset SAS Session**.

Using the Autocomplete Feature

About the Autocomplete Feature

The autocomplete, or code completion, feature in the code editor can predict the next word that you want to enter before you actually enter it completely. The autocomplete feature can complete keywords that are associated with SAS procedures, statements, macros, functions, CALL routines, formats, informats, macro variables, SAS colors,

style elements, style attributes, and statistics keywords, and various SAS statement and procedure options.

Note: The autocomplete feature is available only for editing SAS programs.

This example shows the keywords and help that appear when you enter `proc a` in the code editor.
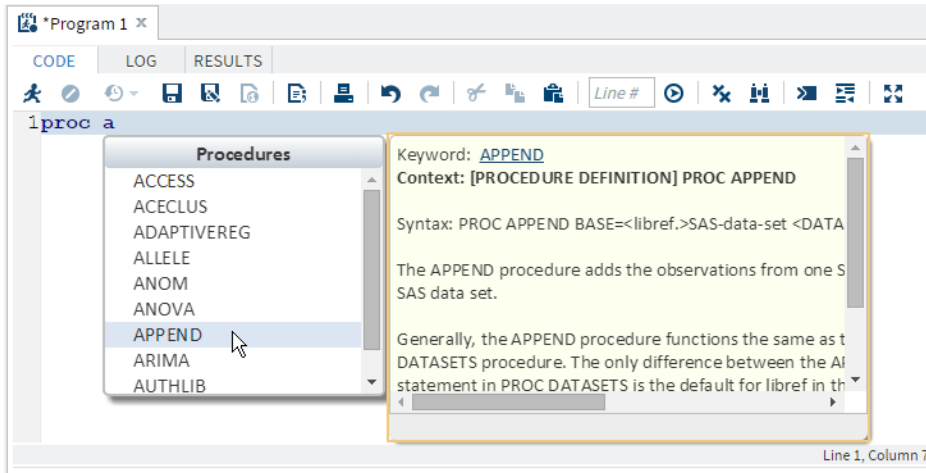
In this example, you select **APPEND** from the list of procedures, so that `proc append` appears in the code editor. When you enter a space, the code editor displays a list of options for the APPEND procedure.
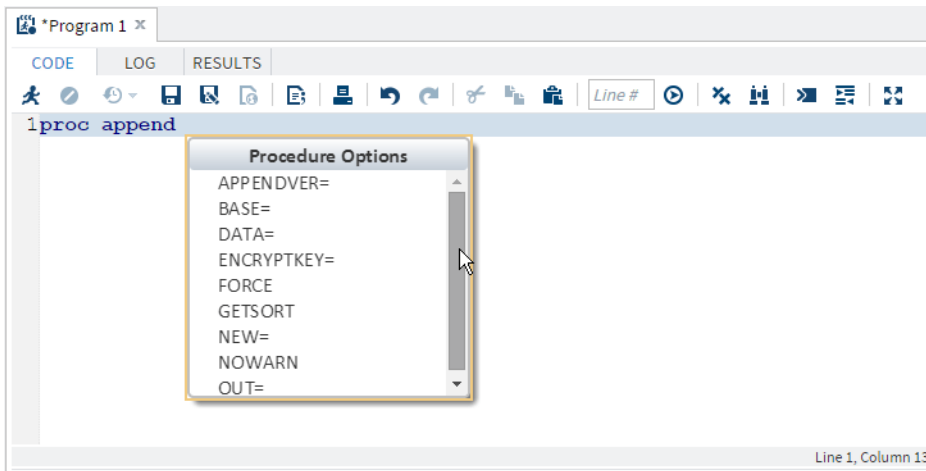
How to Use the Autocomplete Feature

To use the autocomplete feature:

1 How you open the autocomplete list depends on the keyword that you want to add.

   ▪ If you want to add a global statement, DATA step statement, CALL routine, procedure, macro statement, or automatic macro variable, enter the first one or more letters of the keyword that you want to use.

   A window opens with a list of suggested keywords that begin with those letters.
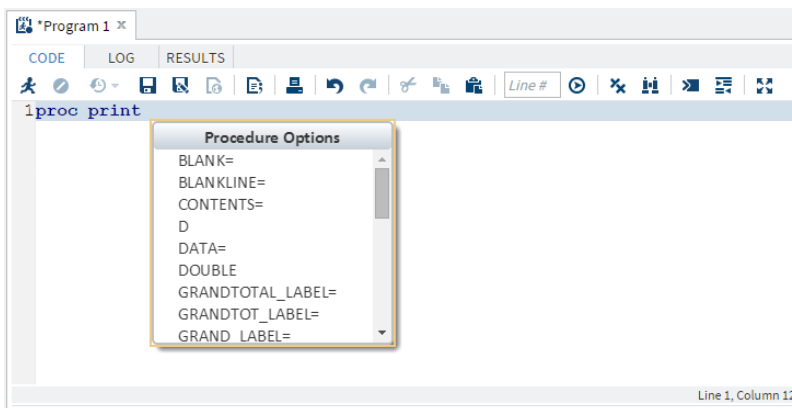
   ▪ If you want to specify colors, formats, informats, macro functions, SAS functions, statistics keywords, style elements, or style attributes, position your mouse pointer in a comment and press Ctrl+spacebar. To navigate through the list of options backward, press Ctrl+Shift+spacebar.

   Note: These shortcuts work even if you have deselected the **Enable autocomplete** option in the Preferences window. For more information, see "Customizing the Code Editor" on page 43.
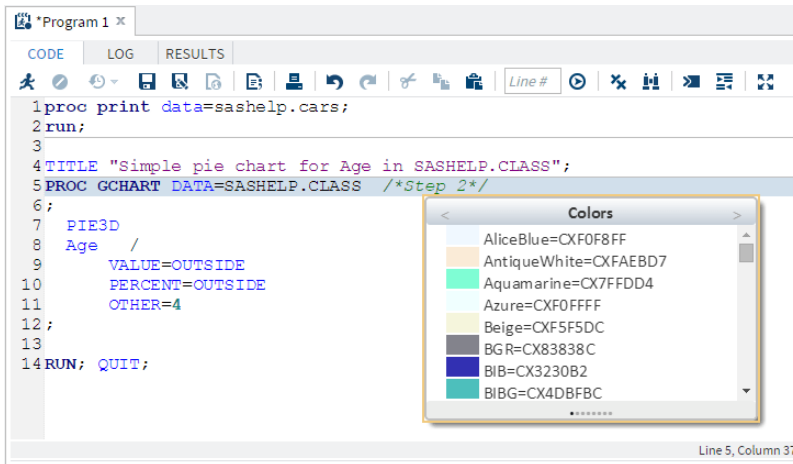
2   You can navigate to the keyword that you want to use in several ways:

■   Continue to type until the correct keyword is selected (because the matching improves as you type).

■   Scroll through the list by using the up and down arrow keys, the Page Up and Page Down keys, or your mouse.

3   You can add the keyword to your program by double-clicking the selected keyword or by pressing the Enter key.

Using the Syntax Help

The code editor displays brief SAS syntax documentation as you write and edit your programs. You can display the Help in the following ways:

■   Right-click a keyword in your program and select **Syntax Help**.

■   Start typing a valid SAS keyword, and then click a suggested keyword in the autocomplete window.

■   Position the mouse pointer over a valid SAS keyword in your program. This works only if you have selected the **Enable hint** option in the Editor preferences. For more information, see "Customizing the Code Editor" on page 43.

The SAS Product Documentation provides more comprehensive usage information about the SAS language, but the syntax help in the code editor can get you started with a hint about the syntax or a brief description of the keyword. You can get additional help by clicking links in the syntax help window as follows:

▪ Click the keyword link at the top of the window to search the support.sas.com website for the keyword.
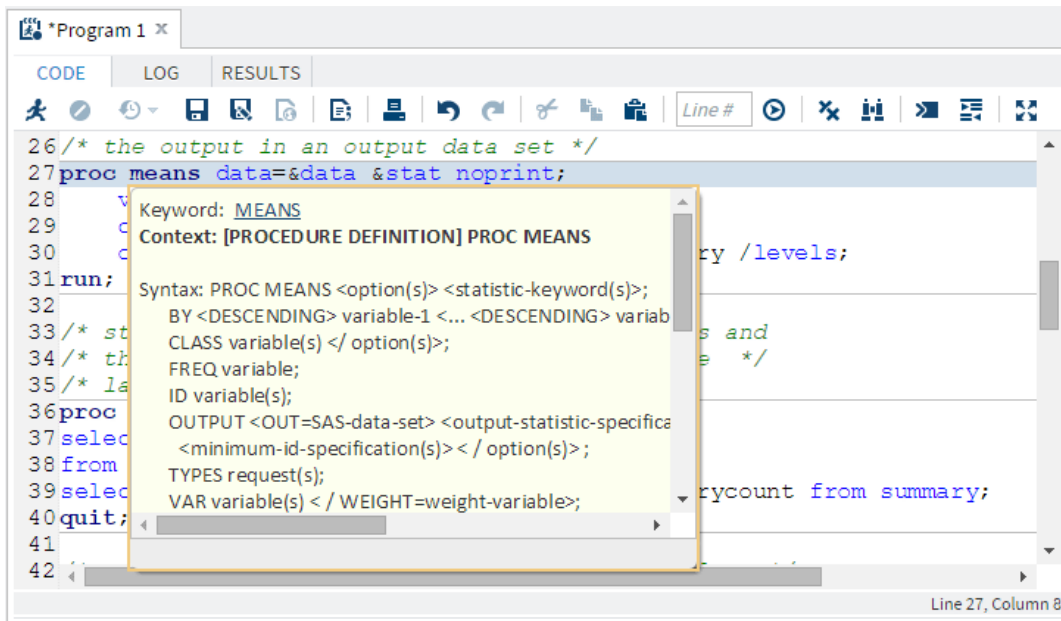
▪ Click the links at the bottom of the window to search for the keyword in the SAS Product Documentation, Samples and SAS Notes, and SAS Technical Papers.
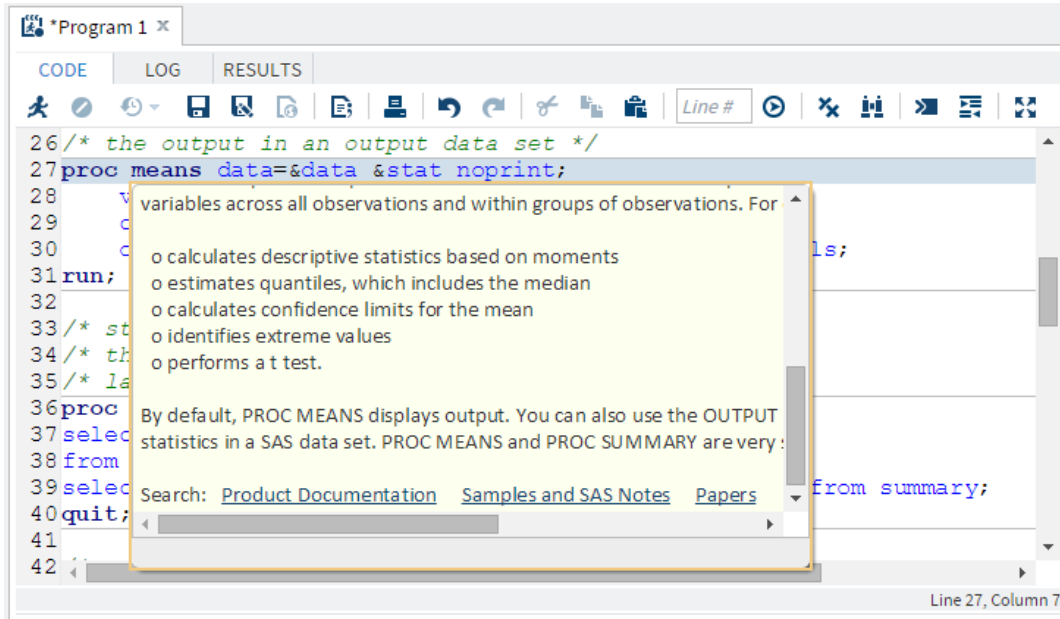
Adding Table Names and Column Names

From the Libraries section of the navigation pane, you can use a drag-and-drop operation to move table names and column names into the SAS code. For example, you can move the Sashelp.Cars table into the DATA option for the PRINT procedure. When you release the mouse, the fully qualified name for the table appears in your code.
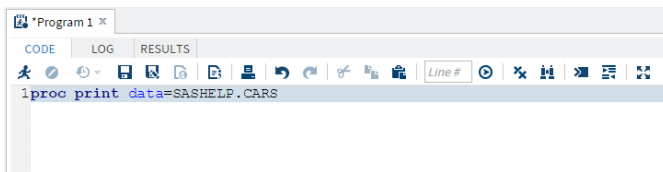
Editing the Code from a Task

You can edit the code that is generated automatically when you run a task and then run it with your modifications. When you edit the code, SAS Studio opens it in a separate program window. The code is no longer associated with the original task.

To edit a program generated by a task:

1   On the appropriate task tab in the work area, click **Code** to display the code that is associated with the task.

    Note: In order to edit the code that is associated with a task, you must first display the code with the task. If the task code is not displayed, click ☰ and select **Preferences**. Click **Tasks**, and then select **Show Task Code**.

2   On the toolbar, click **Edit**. The code is opened in a new program window.
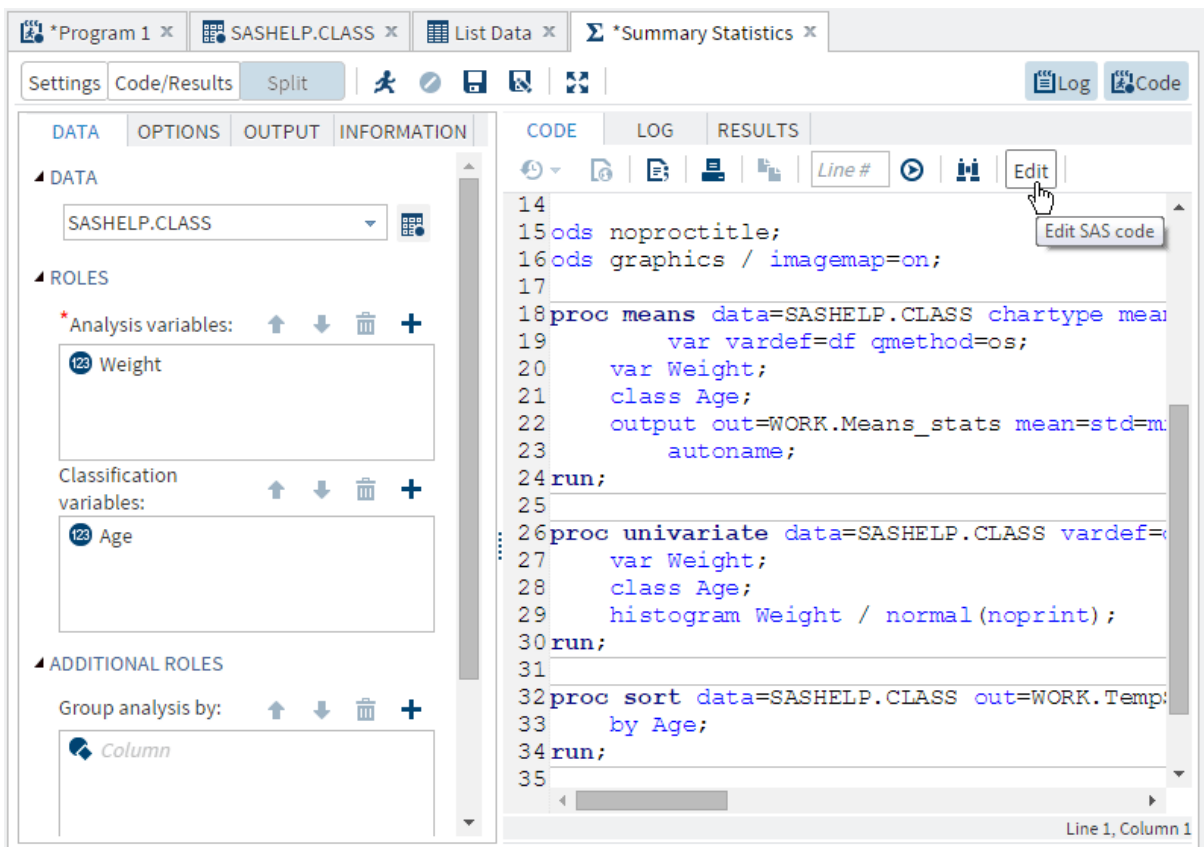
Creating a SAS Program Package

A SAS Program Package is a file that contains a snapshot of a SAS program along with its log and HTML results. You can create a program package from code that you have

written as well as code that is automatically generated when you run a task. When you open a program package in SAS Studio, you can access the code as well as the log and results without running the program again. If you make changes to the code and rerun it, the package is not automatically updated. You must save the package again to keep the changes.

Note: The program package does not include PDF or RTF results. You cannot create a program package if you are running your program in interactive mode.

To create a program package file, open the code that you want to use and click . Specify the file location and name, and then select `SAS Program Package` as the **Save as type** option.

Note: If you want to save the program or log individually or download the results as an HTML, PDF, or RTF file, you must resubmit the program after you open the program package file.

Creating a Program Summary

You can create a summary page for code that you have written as well as code that is automatically generated when you run a task. The Program Summary page is an HTML file that opens in a separate browser tab and includes information about the program execution, the complete SAS source code, the complete SAS log, and the results. To view the Program Summary page for a program, click .

Note: The Program Summary is available only after you have run the program.

You can also save a Program Summary file to a folder that you specify by clicking and selecting `SAS Program Summary` as the **Save as type** option. The Program Summary is saved as an HTML file.

Using Macro Variables

Macro variables can be used to add information that is obtained when a program or task is run, such as the name and version number of the application. You can reference these items within code, titles, or footnotes by preceding them with "&".

For example, you can use macros in a footnote like the following:

```
Generated with &_CLIENTAPP &_CLIENTAPPVERSION
```

Note: In addition to the following macro variables, you can use SAS macro functions and variables in your code. For information about SAS macro functions and variables, see *SAS Macro Language Reference*.

Macro Variable	Description
_CLIENTAPP	name of the application.
_CLIENTAPPVERSION	application version number.
_SASPROGRAMFILE	the full path and filename of the SAS program that is currently being run. This macro variable is available only for SAS program files that are saved on the same server on which your SAS Studio code is running.

Using Your Submission History

SAS Studio maintains a log with entries for each time you run a program or task. You can use this log, or submission history, to access prior versions of your submitted code. To view your submission history, click the **Code** tab in your program or task window. On the toolbar, click 🕐 and select the version that you want to open. The prior version of the program opens in a new window from which you can copy and paste the code as needed.

Note: The submission history is cleared when you sign off from SAS Studio.

Automatically Formatting Your SAS Code

You can use the code editor to make your programs easier to read by automatically formatting your code. When you automatically format your code, line breaks are added, and each line is correctly indented according to its nesting level. To format the code in the code editor, click 🔢.

For example, the following code is difficult to read because it lacks indention and logical line breaks:

```
data topn;
length rank 8; label rank="Rank";
set topn; by &category descending &measure;
if first.&category then rank=0; rank+1;
if rank le &n then output;
run;
```

After you use the automatic code-formatting feature, the program looks like this:

```
data topn;
    length rank 8;
    label rank="Rank";
    set topn;
    by &category descending &measure;

    if first.&category then
        rank=0;
    rank+1;

    if rank le &n then
        output;
run;
```

Working in Interactive Mode

Some SAS procedures are interactive, which means they remain active until you submit a QUIT statement, or until you submit a new PROC or DATA step. In SAS Studio, you can use the code editor to run these procedures, as well as other SAS procedures, in interactive mode.

By using interactive mode, you can run selected lines of code from your SAS program and use the results to determine your next steps. For example, the OPTMODEL procedure in SAS/OR enables you to model and solve mathematical programming models. By running this procedure interactively, you can quickly check results for parts of the program and determine whether you need to make any modifications without running the entire program.

Note: Interactive mode is available only if you are running the first maintenance release for SAS 9.4.

Note: When you use interactive mode, the default output library is Webwork. If you run a program that creates output data in interactive mode, you must refresh the Libraries section of the navigation pane to view the new data.

To run a program in interactive mode, click on the toolbar. To turn off interactive mode, click again. If you change modes while a program is open, the log and results for that program are cleared. You can also clear the log and results manually by clicking on the appropriate toolbar.

Note: If you create a new library when you are in interactive mode, the library might not be available when you turn off interactive mode. If you want the library to be available in both interactive and noninteractive modes, you can assign the library again in noninteractive mode, or you can add the library assignment to your autoexec.sas file. For more information, see "Editing the Autoexec File" on page 18.

When you run a program in interactive mode, SAS Studio does not add any automatically generated code, such as ODS and %LET statements, to your program. In addition, results are generated only in HTML. In interactive mode, the log and results are appended to the existing log and results. Previously submitted code remains active until you terminate it.

For example, suppose you have the following program:

```
proc sql;
   select * from sashelp.cars;

   select * from sashelp.class;
   quit;
```

In noninteractive mode, if you select the first two lines of code and submit them, the code runs successfully. If you then select the last two lines of code and submit them, the code fails because the PROC SQL statement is missing.

If you switch to interactive mode and follow the same steps, the last two lines of code run successfully because the PROC SQL statement is still active.

Note: For documentation about specific procedures, see the SAS Programmer's Bookshelf on support.sas.com.

Working with Code Snippets

Why Use Code Snippets?

Code snippets enable you to quickly insert SAS code into your program and customize it to meet your needs. SAS Studio is shipped with several code snippets. You can also create your own snippets and add snippets to your list of favorites.

Snippet Name	Description
Catalogs	
Edit a SOURCE Entry	enables you to specify the content for the SOURCE entry type in a catalog. The SOURCE entry can contain the same code as SAS programs.
List Catalogs	lists in a SAS table all of the catalogs in the specified library. By default, this table is saved in the Work library and with catalogs as the filename.
List Catalog Entries	lists all the catalog types in the specified catalog.
Print GRSEG Entry	prints the GRSEG entry to the open destinations specified by the SAS Output Delivery System (ODS output).
Data	

Snippet Name	Description
DS2 Code	provides a template for a DS2 program. DS2 is a SAS programming language that is appropriate for advanced data manipulation. DS2 is included with Base SAS and shares core features with the SAS DATA step. DS2 exceeds the DATA step by adding variable scoping, user-defined methods, ANSI SQL data types, and user-defined packages. The DS2 SET statement accepts embedded FedSQL syntax, and the run-time-generated queries can exchange data interactively between DS2 and any supported database. This allows SQL preprocessing of input tables, which effectively combines the power of the two languages. For more information, see *SAS DS2 Language Reference*.
DS2 Package	provides a template for a DS2 package. A package is similar to a DS2 program. The package body consists of a set of global declarations and a list of methods. The main syntactical differences are the PACKAGE and ENDPACKAGE statements. These statements define a block with global scope. For more information, see *SAS DS2 Language Reference*.
DS2 Thread	provides a template for a DS2 threaded program. Typically, DS2 code runs sequentially. That is, one process runs to completion before the next process begins. It is possible to run more than one process concurrently, using threaded processing. In threaded processing, each concurrently executing section of code is said to be running in a thread. For more information, see *SAS DS2 Language Reference*.
Generate CSV File	enables you to export SAS data as a comma-separated text file.
Generate PowerPoint Slide	enables you to stream Microsoft PowerPoint output to your web browser.
Generate XML File	enables you to export SAS data as an XML file that you can view in your web browser.
Import CSV File	enables you to import a comma-separated file and write the output to a SAS data set.

Snippet Name	Description
Import XLSX File	enables you to import a Microsoft XLSX file and write the output to a SAS data set.
Simulate Linear Regression Data	creates an input data source that you can use for linear regression analysis. Linear regression analysis tries to assign a linear function to your data by using the least squares method.
Simulate One-Way ANOVA Data	creates an input data source that considers one treatment factor with three treatment levels. When you analyze this data by using the One-Way ANOVA task, the goal is to test for differences among the means of the levels and to quantify these differences.
Descriptive	
Custom ODS Output	provides a template for creating HTML, PDF, and RTF output by using the SAS Output Delivery System. For more information, see *SAS Output Delivery System: User's Guide*.
PROC SQL	provides a template for writing SQL queries. For more information, see *SAS SQL Procedure User's Guide*.
Graph	
Note: For more information about the SGPLOT, SGPANEL, and SGSCATTER procedures, see *SAS ODS Graphics: Procedures Guide*.	
Bar Panel	uses the VBAR statement in the SGPANEL procedure and enables you to create multiple bar charts.
Box Panel	uses the VBOX statement in the SGPANEL procedure and enables you to create multiple box plots.
Comparative Scatter Plot	uses the COMPARE statement in the SGSCATTER procedure. This code snippet creates a comparative panel of scatter plots with shared axes.
Dot Plot	uses the DOT statement in the SGPLOT procedure. Dot plots summarize horizontally the values of a category variable. By default, each dot represents the frequency for each value of the category variable.

Snippet Name	Description
Fit Plot	uses the REG statement in the SGPLOT procedure. This code snippet produces a regression plot with a quadratic fit and includes confidence limits.
HBar Plot	uses the HBAR statement in the SGPLOT procedure. This code snippet creates a horizontal bar chart that summarizes the values of a category variable.
HighLow Plot	uses the HIGHLOW statement in the SGPLOT procedure. High-low charts show how several values of one variable relate to one value of another variable. Typically, each variable value on the horizontal axis has several corresponding values on the vertical axis.
Histogram Plot	uses the HISTOGRAM statement in the SGPLOT procedure. This code snippet produces a histogram with two density plots. In this snippet, one density plot uses a normal density estimate and the other density plot uses a kernel density estimate.
Scatter Plot Matrix	uses the MATRIX statement in the SGSCATTER procedure. This code snippet creates a scatter plot matrix.
VBox Plot	uses the VBOX statement in the SGPLOT procedure. A box plot summarizes the data and indicates the median, upper and lower quartiles, and minimum and maximum values. The plot provides a quick visual summary that easily shows center, spread, range, and any outliers. The SGPLOT and the SGPANEL procedures have separate statements for creating horizontal and vertical box plots.

Macro

Note: For more information about SAS macros, see *SAS Macro Language: Reference*.

Snippet Name	Description
SAS Macro Char Functions	provides several examples of these SAS macros that work with character values: ■ The %EVAL function evaluates arithmetic and logical expressions by using integer arithmetic. This function operates by converting its argument from a character value to a numeric or logical expression. After the expression is evaluated, the result is converted back to a character value. This function is useful because the SAS Macro Facility is basically a text generator. As a result, an arithmetic expression is first converted to a numeric expression. After this numeric expression is evaluated, it is converted back to an arithmetic expression. ■ The %INDEX function returns the position of the first character of a string. ■ The %LENGTH function returns the length of a string. ■ The %SCAN function searches for a word that is specified by its position in a string. ■ The %SUBSTR function produces a substring of a character string. ■ The %UPCASE function converts values to uppercase.
SAS Macro Do Statement	designates the beginning of a section of a macro definition that is treated as a unit until a matching %END statement is encountered. This macro section is called a %DO group. A simple %DO statement often appears in conjunction with %IF-%THEN-%ELSE statements to designate a section of the macro to be processed depending on whether the %IF condition is true or false. **Note:** SAS also provides a %DO iterative statement, which is different from the code that is generated by this snippet. For more information, see *SAS Macro Language: Reference*.
SAS Macro If Statement	conditionally processes a portion of a macro. The expression that is the condition for the %IF-%THEN-%ELSE statement can contain only operands that are constant text or text expressions that generate text.

Snippet Name	Description
SAS Macro Parameters	names one or more local macro variables whose values you specify when you invoke the macro. There are two types of macro variables: positional and keyword. Parameters are local to the macro that defines them. You must supply each parameter name. You cannot use a text expression to generate it. A parameter list can contain any number of macro parameters separated by commas. The macro variables in the parameter list are usually referenced in the macro.
SAS Macro	provides a basic template for working with SAS macros. Macros enable you to perform many tasks, including substituting text in a program. A SAS program can contain any number of macros, and you can invoke a macro multiple times in a single program. For more information, see *SAS Macro Language: Reference*.

Snippet Name	Description
SAS Macro Quoting	provides examples of macro functions that tell the macro processor to interpret special characters and mnemonics as text rather than as part of the macro language.

■ The %STR function masks special characters and mnemonic operators in constant text at macro compilation. This function masks these special characters and mnemonic operators:

```
 + - * / < > = ¬ ^ ~ ; , # blank
 AND OR NOT EQ NE LE LT GE GT IN
```

This function also masks these characters when they occur in pairs and when they are not matched and are marked by a preceding %:

```
 ' " ()
```

■ The %NRSTR function masks special characters and mnemonic operators in constant text at macro compilation. This function masks all of the special characters and mnemonic operators listed for the %STR function. In addition, the %NRSTR function masks these characters:

```
 & %
```

■ The %BQUOTE function masks special characters and mnemonic operators in a resolved value at macro execution. This function masks these special characters and mnemonic operators:

```
 ' " () + - * / < > = ¬ ^ ~ ; , # blank
 AND OR NOT EQ NE LE LT GE GT IN
```

Snippet Name	Description

SAS Macro Quoting (continued)	■ The %SUPERQ function masks all special characters and mnemonic operators at macro execution but prevents further resolution of the value. This function masks these special characters and mnemonic operators:
	
	`& % ' " () + - * / < > = ¬ ^ ~ ; , # blank`
	`AND OR NOT EQ NE LE LT GE GT IN`
	
	■ The %QSCAN function searches for a word and masks special characters and mnemonic operators.
	
	■ The %QSUBSTR function produces a substring and masks special characters and mnemonic operators.
	
	■ The %QUPCASE function converts a value to uppercase and returns a result that masks special characters and mnemonic operators.
	
	■ The %UNQUOTE function unmasks a value during macro execution so that any special characters and mnemonic operators are interpreted as macro language elements instead of text.
	
	For more information about macro complication and macro execution, see *SAS Macro Language: Reference*.

SAS Macro Variables	provides examples of how to create user-defined global and local macro variables. Macro variables are tools that enable you to dynamically modify the text in a SAS program through symbolic substitution. You can assign large or small amounts of text to macro variables. Then you can use that text by simply referencing the variable that contains the text. Macro variables that are defined by the macro programmer are called user-defined macro variables. Macro variables that are defined by the macro processor are called automatic macro variables. You can define and use macro variables anywhere in SAS programs, except within data lines.

IML

Note: These snippets are available only if your site licenses SAS/IML.

Snippet Name	Description
Find Roots of Nonlinear Equation	enables you to find the roots of a function of one variable. Finding the root (or zero) of a function enables you to solve nonlinear equations.
Integrate a Function	enables you to numerically integrate a one-dimensional function by using the QUAD subroutine in SAS/IML software. Use the QUAD subroutine to numerically find the definite integral of a function on a finite, semi-infinite, or infinite domain.
Generate a Bootstrap Distribution	uses the IML procedure to create and analyze a bootstrap distribution of the sample mean.
Fit by using Maximum Likelihood	uses maximum likelihood estimation to estimate parameters for the normal density estimate.
Simulate Multivariate Normal Data	simulates data from a multivariate normal distribution with a specified mean and covariance.

Create a Code Snippet

To create your own snippet:

1   Open your .sas file in SAS Studio and select the code that you want to save as a snippet.

2   On the **Code** tab, click . The Add to My Snippets dialog box appears.

3   Enter a name for the snippet and click **Save**.

This snippet is now available from the **My Snippets** folder.
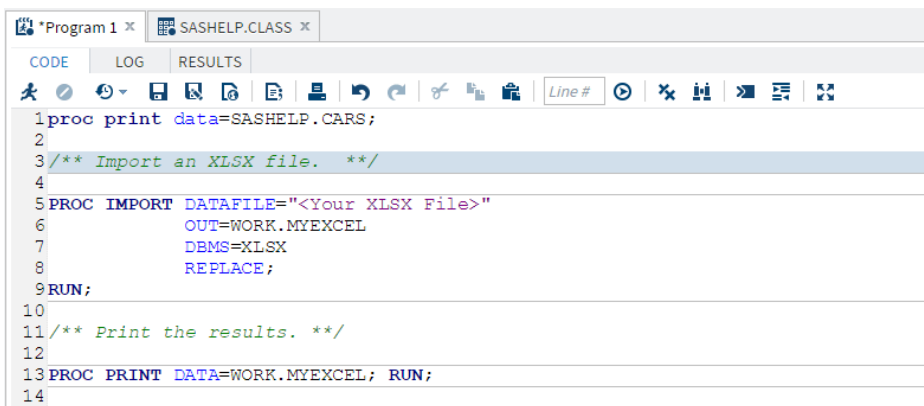
How to Insert a Code Snippet

To include a code snippet in your program:

1   Click the location in your program where you want to insert the snippet.

2   In the navigation pane, open the **Snippets** section.

3   You can add the snippet to your program in these ways:

■   use a drag-and-drop operation to move the snippet.

■   double-click the name of the snippet.

■   right-click the name of the snippet and select **Insert**. To select multiple snippets, use the Ctrl key. Then right-click and select **Insert**.

The following example shows the Import XLSX File snippet inserted into a program.

Customizing the Code Editor

The **Preferences** window enables you to change several options that affect the features in the code editor, including autocomplete and color coding.

To access the editor options, click ☰ and select **Preferences**. Click **Editor**.

Option	Description
Enable autocomplete	turns on the autocomplete feature of the code editor. This feature can predict the next keyword that you want to type before you actually type it completely. For more information, see "Using the Autocomplete Feature" on page 23.
Enable hint	displays the syntax help window when you position the mouse pointer over a valid SAS keyword in your program. If this option is not selected, then you can view the syntax help by right-clicking a keyword and selecting **Syntax Help**. This option is not selected by default.
Tab width	displays the number of spaces that are inserted into your text when you insert a tab character. The default value is four spaces for each tab character.
	Note: In Microsoft Internet Explorer and Apple Safari, spaces are used instead of Tab characters. If you are using those browsers, you must select the **Substitute spaces for tabs** check box in order for the value of the tab width to be used.
Substitute spaces for tabs	inserts the number of spaces listed in the **Tab width** box instead of a single tab character. This option applies to both text that you type in the code editor and text that you paste into the code editor.
Enable color coding	displays the text in the code editor in different colors to help you identify different elements in the syntax.
Show line numbers	displays line numbers in the leftmost column of the program and log windows.
Font size	specifies the font size of the text in the code editor and log window.

Option	Description
Enable autosave	automatically creates auto-saved copies of each previously saved program file that you are editing so that you can recover the files if your browser closes unexpectedly. New program files are not auto-saved until you save them first. The files are saved at the interval specified in the **Autosave Interval** option. The default interval is 30 seconds.

3

Working with Queries

What Is a Query?

A query enables you to extract data from one or more tables according to criteria that you specify. You can create a query that is based on only one table, or you can join tables together. When you create a query, SAS Studio generates Structured Query Language (SQL) code, which you can view. You can create queries in both the SAS Programmer and the Visual Programmer perspectives.

Creating a Query

Creating a New Query

On the main SAS Studio toolbar, click ⊞ and select **New Query**. A query window opens on a new tab in the work area.
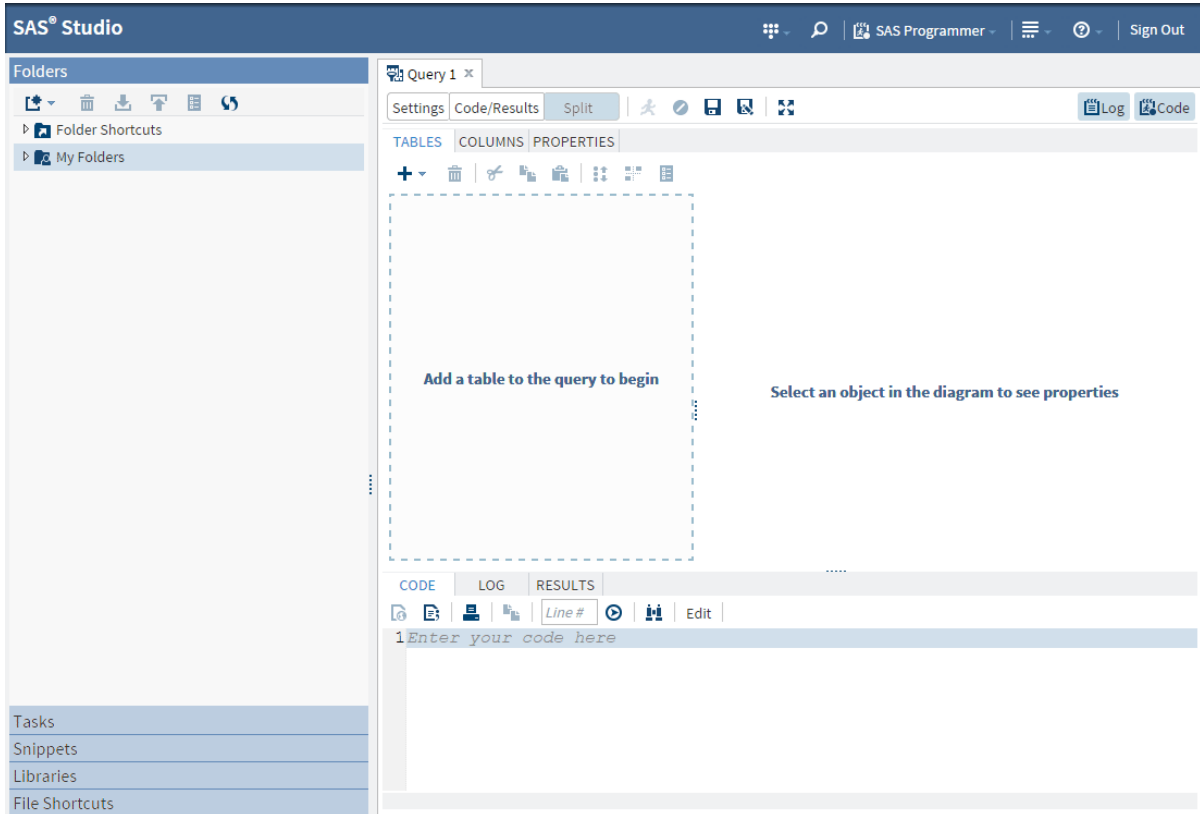
Adding Tables to a Query

On the Tables tab in the query window, click **+** and select **Table**. From the Choose a Table window, expand the appropriate library and select the table that you want to use. The table is added to the query.

Note: You can also add tables to the query by using the Libraries section of the navigation pane. Click **Libraries** and expand the appropriate library. Drag the table that you want to use to the Tables tab of the query window.

Understanding Joins

Joining Tables

When you create a query, you can join multiple tables together. SAS Studio can automatically join the tables together for you, or you can manually create the join. SAS Studio attempts to join tables by columns that have the same name and type. If no matches for column name and type are found, then you can specify the join criteria.

Note: If you have more than one table in your query and you do not specify join criteria, then your output data includes the Cartesian product, or every possible combination, of the data values.

Creating a Join

To add a table and automatically create a join:

From the Libraries section of the navigation pane, drag the table that you want to add to the query to the Tables tab. Next, drop that table on top of the first table in the query to join the two.

The Join window displays the join criteria. In the following example, the Classfit table is automatically joined to the Class table by using the Name column in both tables.
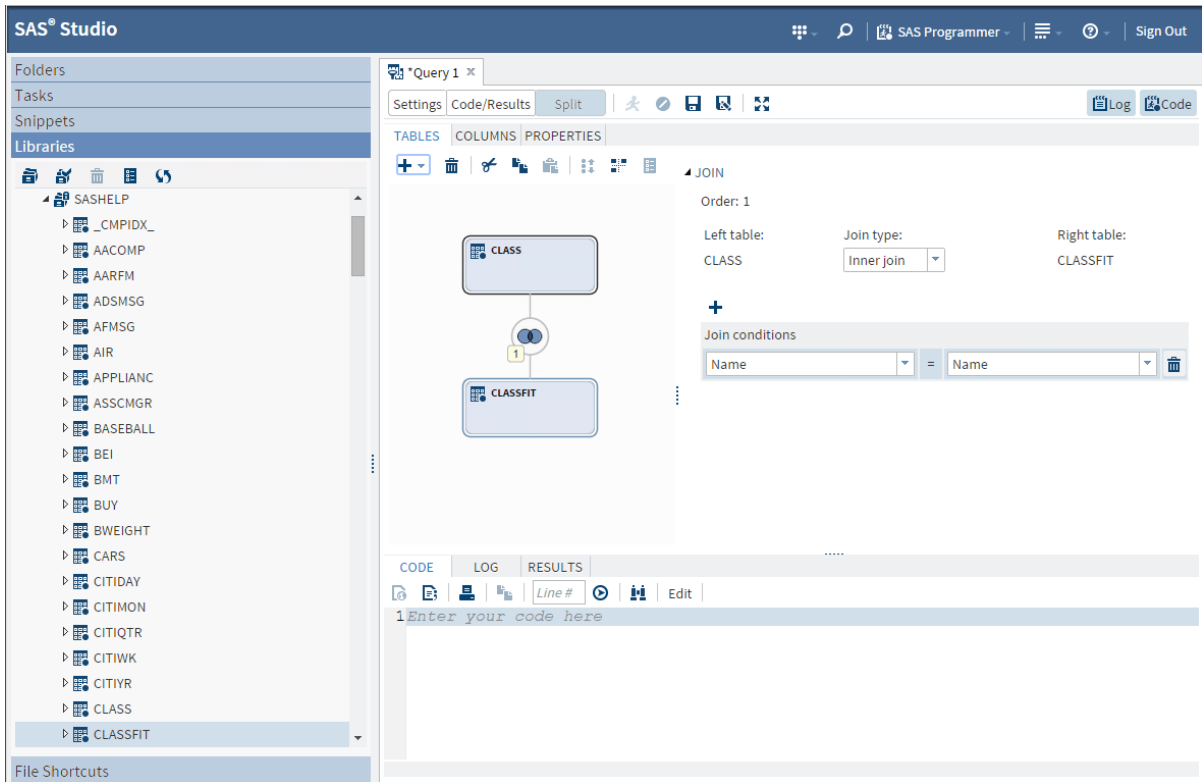
If a join cannot be created automatically, you can specify the join condition manually.

To manually create a join:

1   On the Tables tab of the query window, make sure that you can view the tables that you want to join.

2   Click **+** on the toolbar and select **Join**. The New Join window opens.
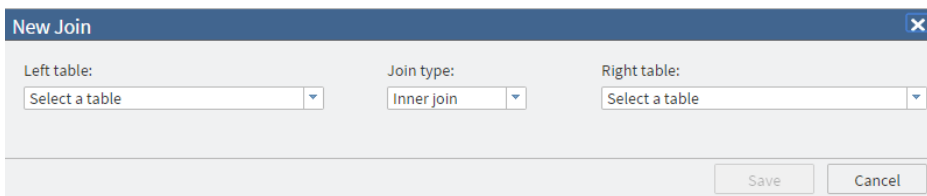
3   From the **Left table** drop-down list, select the table for the left side of the join.

4   From the **Join type** drop-down list, select the type of join that you want to use. The default join type is Inner join.

5   From the **Right table** drop-down list, select the table for the right side of the join.

6   Click **Save**. A join is created between the tables. If the tables include columns with matching names and data types, then a join condition is automatically created. If the tables do not include columns with matching names and data types, then you can select the columns for the join condition from the column drop-down lists.

7   To add another join condition to the join, click **+** and select the columns that you
    want to use from the column drop-down lists.

Understanding the Types of Joins

SAS Studio supports four different types of joins. You can select the type of join you
want by modifying an existing join.

You can select the join option that you want to use in the Join window.

SAS Studio Join Type	Join Icon	Description
Inner Join		The output rows include those for which the column in the first table matches the joining criterion of the column in the second table. Joins are inner joins by default.
Left Join		The output rows include all rows from the first table and the rows from the second table in which the joining criterion is met.
Right Join		The output rows include all rows from the second table and the rows from the first table in which the joining criterion is met.
Full Join		The output rows include all matching and nonmatching rows from both tables.

Modifying an Existing Join

You can modify an existing join by selecting a different type of join or by changing the columns that are used in the join condition. You can also add and remove join conditions or remove the entire join.

To modify a join:

1   On the Tables tab of the query window, click the join indicator that you want to modify. The join is displayed in the Join area.

2   To change the type of join, select a new type from the **Join type** drop-down list.

3   To add a new join condition, click ✚ and specify the columns to use in the join. To remove a join condition, click 🗑 next to the appropriate condition.

To delete the entire join, right-click the join indicator and select **Delete**.

Selecting Data

Specifying Columns in the Output

By default, no columns are included in the output. You must specify the columns that you want to appear in the output table. You can also specify an alias to use in place of the column name in the output table.

The order in which the columns are listed on the Select tab is the order in which they appear in the output table.

To select columns for the output table:

1   In the query window, click the **Columns** tab to view the list of columns from the tables in the query.

2   You can add one or more columns to the output data by dragging them from the
    columns list to the Select tab. You can also click **+** on the Select tab toolbar and
    select one or more columns from the Choose Column window.

To specify an alias for a column:

■   On the Select tab, enter the alias that you want to use for each column. The alias is
    used as the column heading for the output data.

Using Summary Functions

You can perform summary functions on any of the columns in your query. To perform a summary function, select the column on which you want to perform a summary function. Use the drop-down list in the Summary column to select the function that you want to use. By default, the query generates an output data set. The following example shows you how to find the average age of all of the students:

By default, the query generates an output table with the results:

By default, when you summarize a column, your output is grouped by all of the columns without summaries. For more information, see "Grouping Your Output" on page 65.

Filtering Data

Creating a Filter

When you query data, you might want to retrieve only rows that meet certain criteria, based on values of columns in the data. The process of telling SAS Studio which rows to retrieve is called setting a filter and is done on the Filter tab. This corresponds to using a WHERE clause in an SQL query.

1   In the query window, click the **Columns** tab to view the list of columns from the tables in the query.

2   You can add one or more columns to the filter by dragging them from the columns list to the Filter tab. You can also click **+** on the Filter tab toolbar and select one or more columns from the Choose Column window.

3   Select a comparison operator from the **Operator** drop-down list. The default value is **Equals**.

4 If the operator that you have selected requires a value, click ⬛ to enter or select a value in the **Select Value** window. To choose from a list of values, click ⊽ to expand the **Value** list. Select the values that you want to use and click **Add**.

Note: If you are selecting values for a character column, and you want the values to be enclosed in single quotation marks, select the **Enclose values in quotes** option. This option is selected by default. If you are using a macro variable or other value that is evaluated when the filter is run, you should clear this option.

If you want to choose from a list of columns, click ⊽ to expand the **Column** list. Select the column that you want to use and click **Add**.
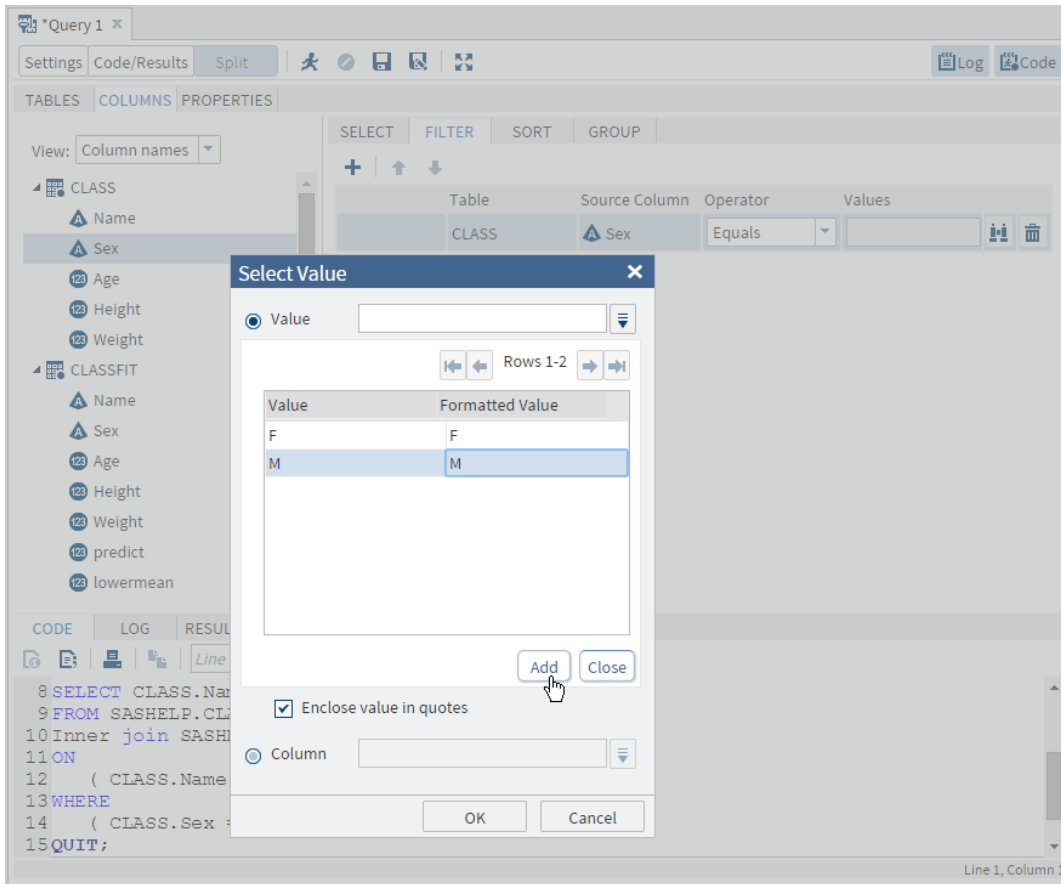
5   Click **OK** to add the values to the filter.

Changing the Relationship between Filters

You can use only one column in a filter, or you can use multiple columns to create several comparison expressions. If you create more than one comparison expression in your filter, then the default relationship between these filter elements is AND. You can change the relationship between filter elements from AND to OR, and you can group elements together.

To change the relationship between filters:

■   On the Filter tab, click the relationship value and select a new value.

Managing Output

Sorting Your Output

You can sort the output from your query by one or more columns from the tables that are used in the query.

Note: It is possible to sort the output table by columns that are not selected for the output.

To sort your output:

1 In the query window, click the **Sort** tab.

2 You can add one or more columns to the Sort tab by dragging them from the columns list to the Sort tab. You can also click **+** on the Sort tab toolbar and select one or more columns from the **Choose Column** window.

3 Click the **Sort** box for the column on which you want to sort the data. From the drop-down list, select **Ascending** or **Descending**. The default sort direction is **Ascending**.

4 If you are sorting by multiple columns, the output table is sorted first by the column that is listed first. Within each level of the first column, the rows are sorted by the second column in the list, and so on. You can change the sort order by selecting a column and clicking ↑ and ↓ to move the column up and down the list.

Eliminating Duplicate Rows in Output

Some types of queries output multiple, identical rows. Because these duplicate rows are generally not useful, SAS Studio enables you to keep only one of the identical rows and eliminate the duplicates.

To eliminate duplicate rows, click the **Select** tab and select the **Select distinct rows only** check box.

Grouping Your Output

If you have created a summarized column, you can choose to classify your data into groups based on the values in a column. This is equivalent to using the GROUP BY clause in an SQL query. For example, if you are calculating the average height of a group of students, you might want to group the results by age so that you can see the average height for each age group.

By default, when you perform a summary function on a column, your query is grouped by all columns without summaries. You can choose to edit the list of columns that the query is grouped by.

To group your output:

1   In the query window, click the **Group** tab.

2   You can add one or more columns to the Group tab by dragging them from the columns list to the Group tab. You can also click **+** on the Group tab toolbar and select one or more columns from the Choose Column window.

3   To change the order in which the columns are used to group the data, select the column that you want to move and click ⬆ and ⬇ .

The following example shows you how to find the average weight of students in each age group. First, add the Age and Weight columns to the Select tab, and then select the AVG summary function for the Weight column:

To see the average weight of students by age, you group by the Age column. The results show the average weight for each age group:

Note: By default, the query generates a table of the result. To generate a report of the results (which is displayed in **Results** tab), you must specify report as the output type for the query. For more information, see .

Saving Your Results

You can choose to generate your results in any one of three formats: data table, data view, or report.

If you save your results as a data table or data view, you can specify the library and filename that you want to use. If you don't specify the library and filename, the results are saved in the Work library.

To specify the results format:

1 In the query window, click the **Settings** tab.

2 Click the **Properties** tab. In the Results area, select the format that you want to use from the **Output type** drop-down list.

Report
    saves the query results as a report that you can download as an HTML, PDF, or RTF file. Query results in this format are not updated until you rerun the query. You cannot run SAS tasks against query results in this format.

Table
    saves the query results as a static data table against which you can run SAS tasks. Query results in this format are not updated until you rerun the query. By default, the data table is stored in the Work library.

View
    saves the query results as a dynamic data view against which you can run SAS tasks. Each time you open query results in the data view format, the results are updated with any changes to the data that is used in the query. By default, the data view is stored in the Work library.

To save your results to a specific location:

1 In the query window, click the **Settings** tab.

2 Click the **Properties** tab. In the Results area, enter the name of the library in which you want to save your results in the **Output location** box.

3 To specify a name for the results, enter the name that you want to use in the **Output name** box.

Running a Query

After you specify all the criteria for your query, you can generate your results by clicking ⊁ on the query window toolbar. The output data opens in the workspace on a separate tab.

4

Working with Process Flows

Understanding Process Flows

What Is a Process Flow?

You must be working in the Visual Programming perspective to work with process flows. For more information about the Visual Programming perspective, see "Understanding Perspectives" on page 16.

A process flow consists of one or more objects. Each object is represented by a node in the process flow. The process flow shows the relationship between two or more objects, such as a SAS program, a task, a query, and so on.

This sample process flow contains three branches.

- In the first branch, a query is created for the Sashelp.Classfit data set. As a result of this query, only observations where Sex=M are included in the output data set (called Webwork.Query). The Rank Data task ranks the Height and Weight values by Age.

- In the second branch, you write a SAS program that generates an output data set. After this program runs, a subflow that contains more program and task nodes runs. Then finally, SAS Studio tries to run the Bar Chart node. However, this task generates a warning.

- In the third branch, the SAS program generates an error. The information for the Table Attributes node is incomplete. SAS Studio cannot run the process flow until you provide the required data.
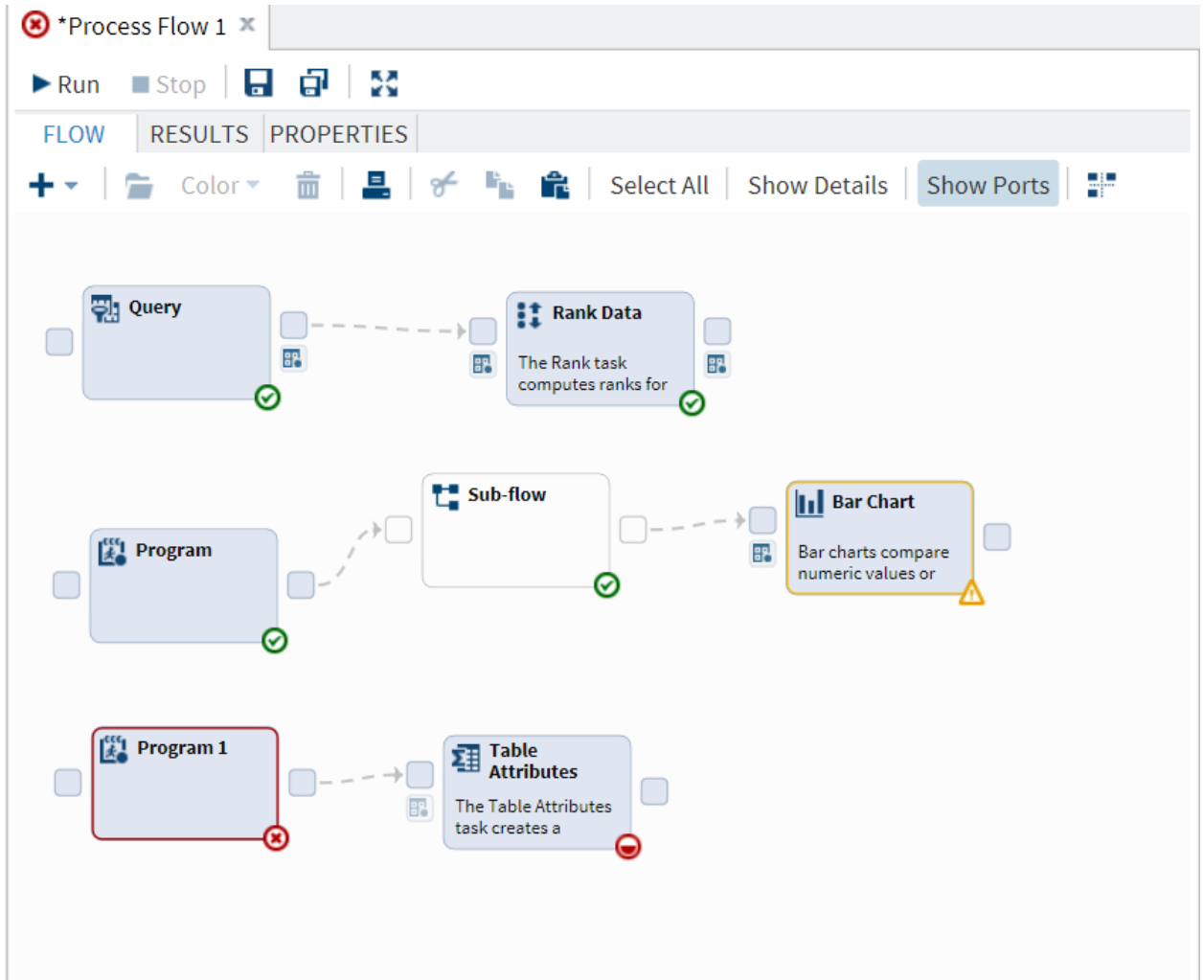
Creating Process Flows

In SAS Studio, you can have multiple process flows. These process flows run independently of each other. To create a new process flow, click [icon] and select **New Process Flow**. A new **Process Flow** tab appears in the interface.

What Are Ports?

In a process flow, you can have two types of ports:

control ports

In the process flow, these types of ports appear as ▢. You use these ports to specify the order in which nodes run in the process flow. For more information, see "Linking Nodes in a Process Flow" on page 91.

data ports

In the process flow, these types of ports appear as ▦.

When you run a task, you must specify an input data source. You specify the input data source in the task interface. From the process flow, you can determine the name of the input data source by positioning your mouse pointer over the input data port. To view the data source on a separate tab, double-click the data port icon.

In this example, the input data source for the Sort Data task is SASHELP.CLASSFIT.

Tasks from the Data category, such as the Sort Data task, have an output data source as well as an input data source. You can use the output data port to determine the name of the output data source. To view the data source on a separate tab, double-click the data port icon.

In this example, the output data source for the Sort Data task is WORK.SORT.
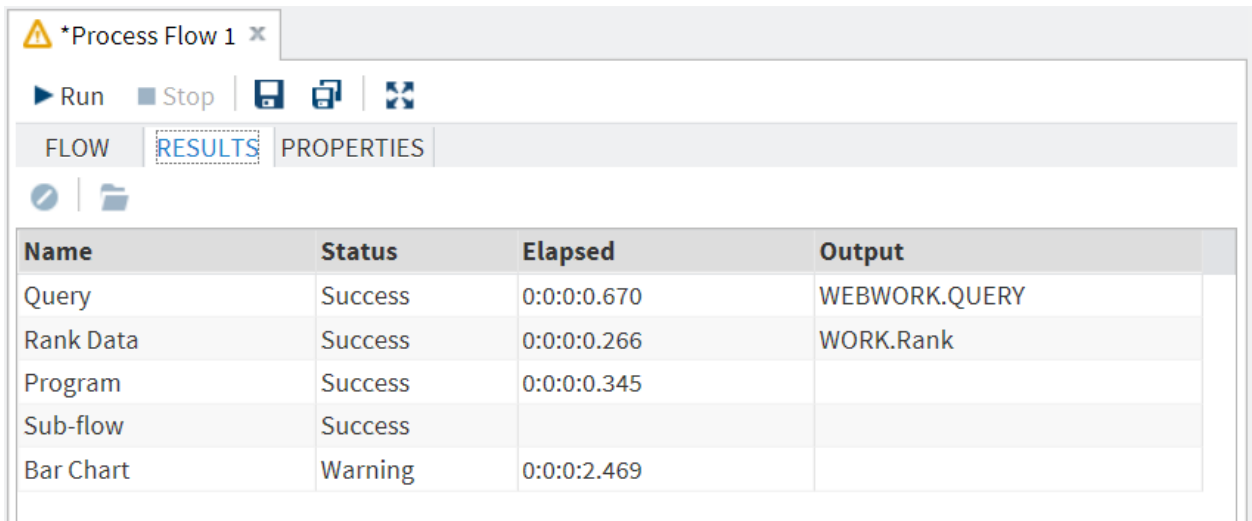
By default, ports are displayed in the process flow. To turn off the ports, click **Show Ports**.

Understanding the Status of Each Node

When you run a process flow, some nodes might run successfully, and others might not. To see the status of each node, how long it took to run the node, and whether any output data was created, click the **Results** tab.

Here are the contents of the **Results** tab for the previous process flow.

You can use these icons to determine the status of each node in the process flow:

Icon	Description
	specifies that more information is needed before SAS Studio can run the node. For example, if you see this icon on a task node, you must specify values for the required options in the task. This icon can also appear on an empty subflow node after SAS Studio tries to run it.
	specifies that the node is in the queue to run.
	specifies that the code for the node was submitted.

⚠	specifies that the code for the node generated a warning. Review the log for that node for more information.
⊗	specifies that the code for the node generated an error. Review the log for that node for more information.

Customizing a Process Flow

You can customize your process flow in these ways.

■ To specify a color for a node or a group of nodes, select the nodes and click **Color**. From the drop-down list, select the color that you want to use.

■ By default, you can arrange objects in the process flow any way you like. However, your process flow might become confusing if it contains many objects. Click ⊞ for SAS Studio to arrange the objects in your process flow. When arranging the nodes, SAS Studio considers any dependencies and the order in which the nodes were added to the process flow.

Viewing the Properties of a Process Flow

To view the properties of the current process flow, click the **Properties** tab. From the properties, you can specify the priority of execution of the nodes. The process flow runs the nodes in the order in which the nodes are added to the process flow. If node 1 is dependent on another node 2, node 2 must run completely before node 1 will run. You can also choose to run the nodes in parallel, which means that multiple workspace servers are used to run the nodes. As a result, the nodes do not share a common Work library, so SAS Studio saves any generated output to the Webwork library.

Adding a SAS Program to the Process Flow

Create a New SAS Program

To add a new SAS program to a process flow:

1 Click **+** and select **SAS program**. A node for the SAS program is added to the process flow.

2 Select the node and click 📁. (You can also select the node and press Enter.) The code editor appears.

3 Enter the code for your program.
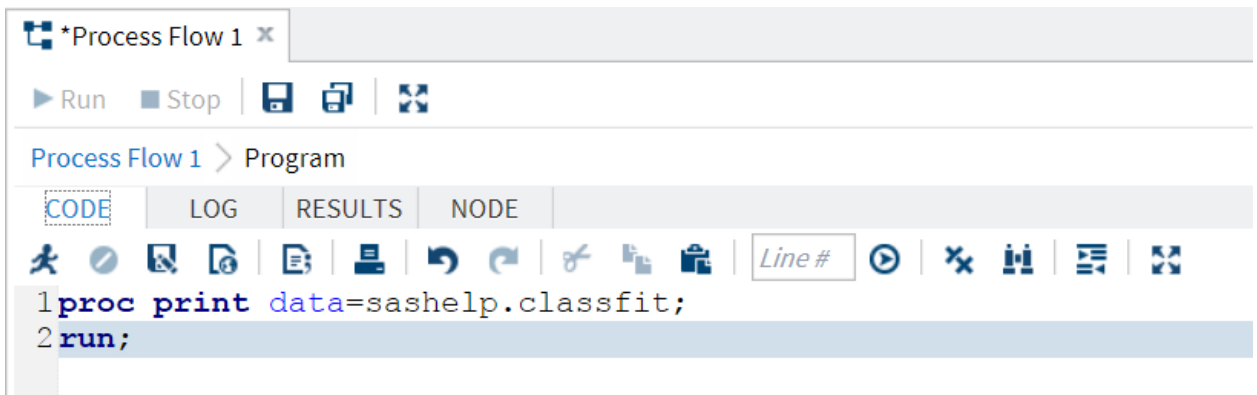
4 To name the program, to provide a brief description for the program, and to include any notes, click the **Node** tab.

*Process Flow 1 ✕

▶ Run   ■ Stop   💾 🗗   ⛶

Process Flow 1 ＞ PROC PRINT

| CODE | LOG | RESULTS | NODE |

◢ IDENTIFICATION

Label:

PROC PRINT

Description:

Prints the SASHELP.CLASSFIT data set

Created: 10/13/2014 1:35:38 PM

Modified: 10/13/2014 2:07:15 PM

◢ NOTES

Notes:

Run this program only once a day.

The SAS program node in the process flow now contains the name and description that you specified on the **Node** tab.
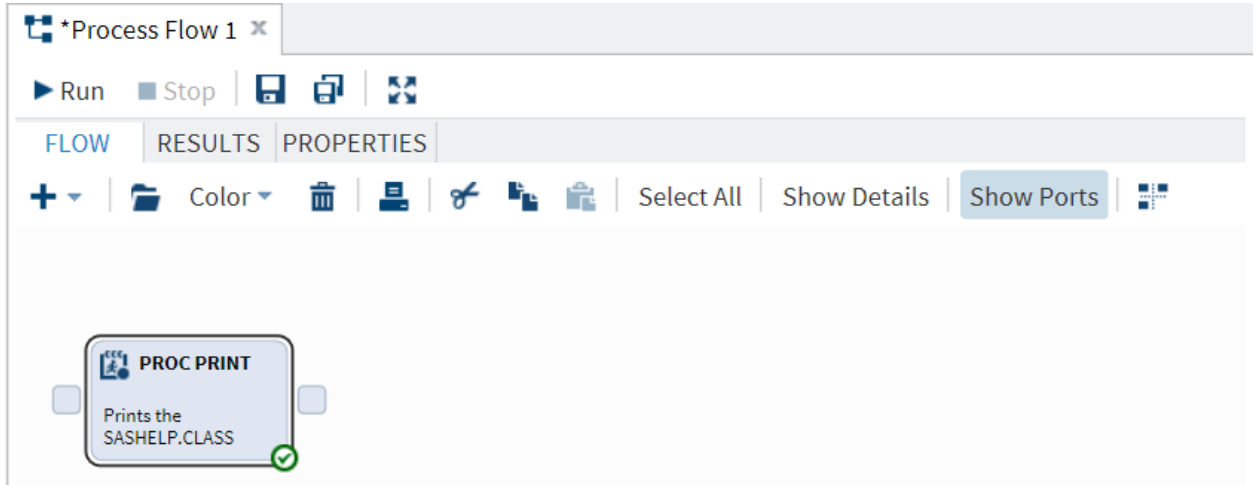
Add an Existing SAS Program

You might have already written a SAS program that you want to include in the process flow. To add this program:

1 In the navigation pane, click the **Folders** section.

2 Expand the folders in the **Folders** section until you find the program that you want to add.

3 Select the program that you want to add, and then drag it in the process flow. (A green check mark indicates that you can add this file to the process flow.)

In this example, you are adding the existing TopNCategories.sas file to the process flow in SAS Studio.
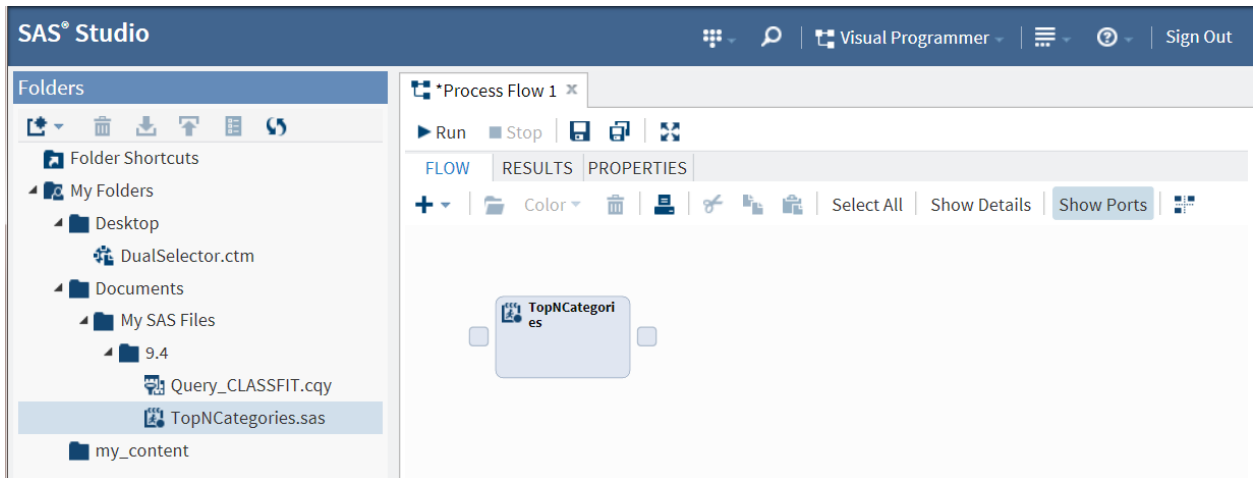
Add a Snippet

You can use snippets as the starting point for your SAS programs.

To add a snippet to a process flow:

1   In the navigation pane, select **Snippets**.

2   In the Snippets section, select the snippet that you want to add, and then drag it to the process flow.

In this example, the Import XLSX file snippet is added to the process flow.
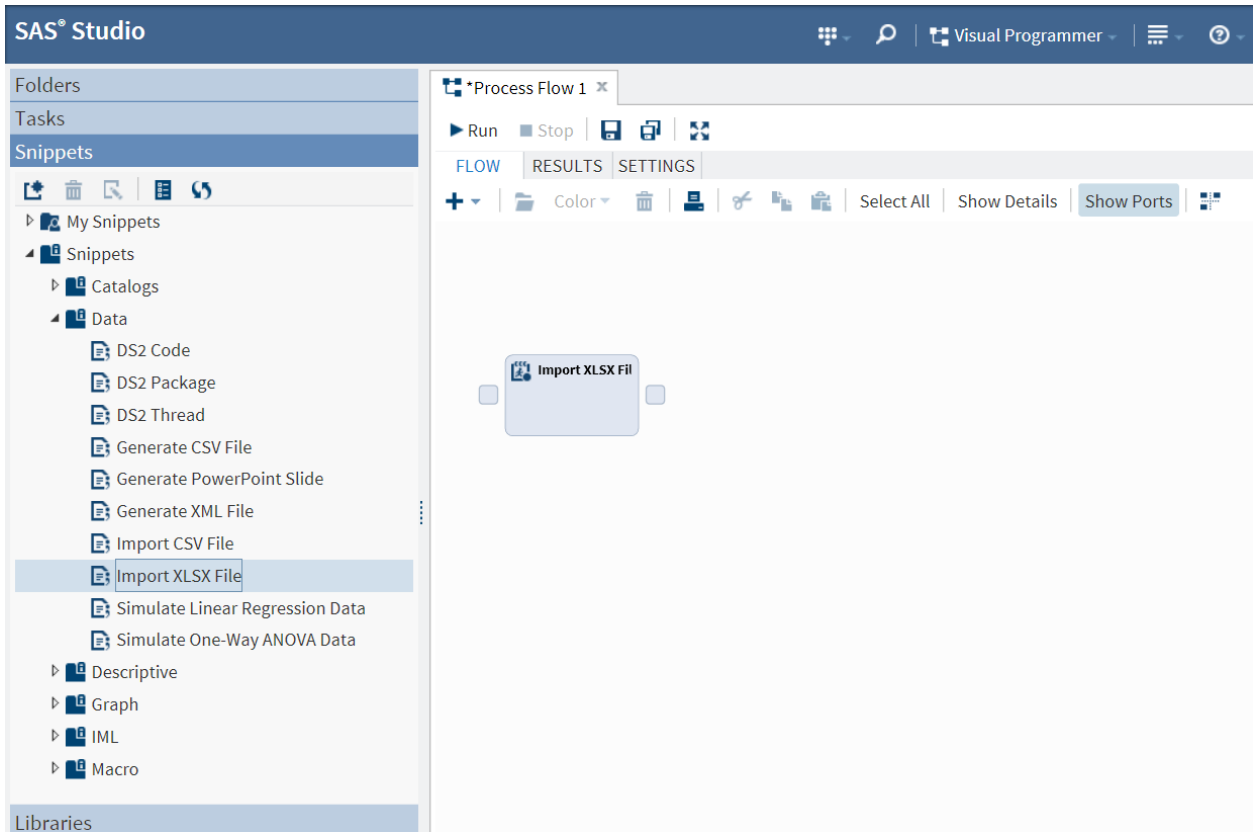
3   To view the code for the snippet, select the snippet node and click 📁. (You can also select the node and press Enter.) Edit the code to meet your needs.

Here is the code for the Import XLSX file snippet. In this example, specify in the code the location of the XLSX file that you want to import.

```
/** Import an XLSX file.  **/

PROC IMPORT DATAFILE="<Your XLSX File>"
                OUT=WORK.MYEXCEL
                DBMS=XLSX
                REPLACE;
RUN;

/** Print the results.  **/

PROC PRINT DATA=WORK.MYEXCEL;  RUN;
```

4   To name the program that you created, to provide a brief description for the program, and to include any notes, click the **Node** tab.

Process Flow 1 ✕

▶ Run   ■ Stop | 🖫   🗗 | ⛶

Process Flow 1 › Import XLSX File

| CODE | LOG | RESULTS | NODE |

◢ IDENTIFICATION

Label:

Import XLSX File

Description:

Imports the Sales.XLSX file

Created: 10/13/2014 2:16:27 PM

Modified: 10/13/2014 2:19:03 PM

◢ NOTES

Notes:

The Import XLSX File node in the process flow now includes the description that you provided.
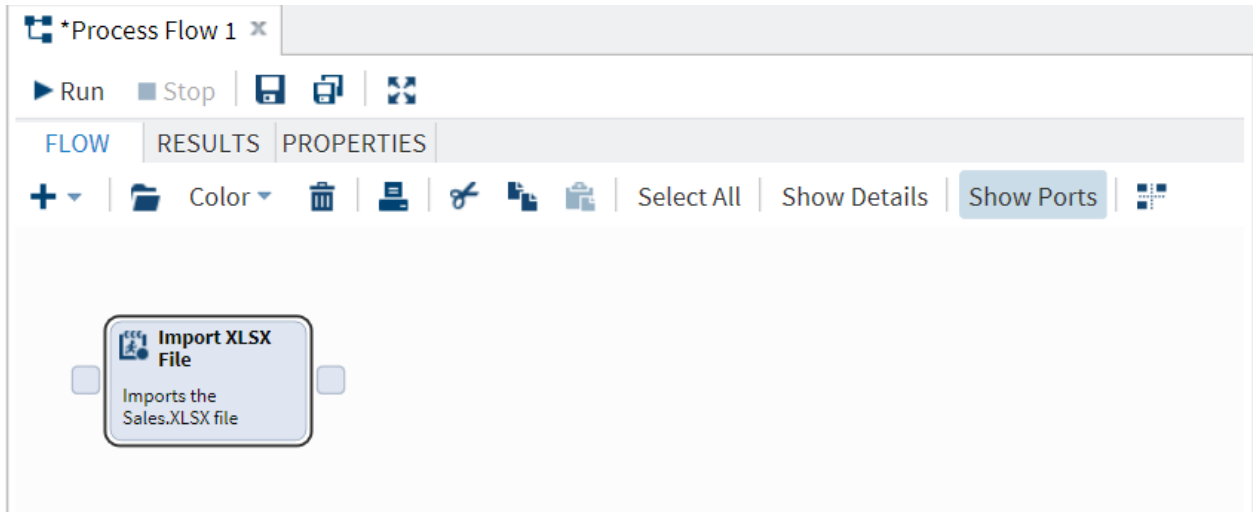
Adding a Query to a Process Flow

Create a New Query

To create a new query:

1 Click **+** and select **Query**. A node for the query is added to the process flow.

2 Select the node and click 📂. The Query Builder appears.

3 Use the Query Builder to define your query. For more information, see "Creating a New Query" on page 48.

4 To define the type of results for the query, use the **Properties** tab. For more information, see "Saving Your Results" on page 68.

5 To name the query node, to provide a brief description for the query, and to include any notes, click the **Node** tab.

   To return to the view of the process flow, select the name of the process flow in the breadcrumbs. The query node in the process flow now contains the name and description that you specified on the **Node** tab.
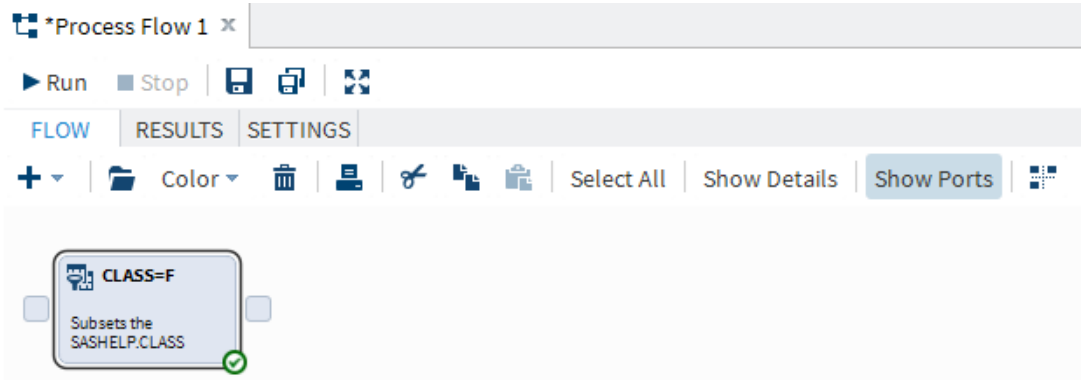
Add an Existing Query to a Process Flow

You might have already written a query that you want to include in the process flow. To add this query:

1 In the navigation pane, click the **Folders** section.

2 Expand the folders in the **Folders** section until you find the program that you want to add.

3 Select the query that you want to add, and then drag it in the process flow. (A green check mark indicates that you can add this file to the process flow.)

In this example, you are adding the existing Query_CLASSFIT file to the process flow in SAS Studio.
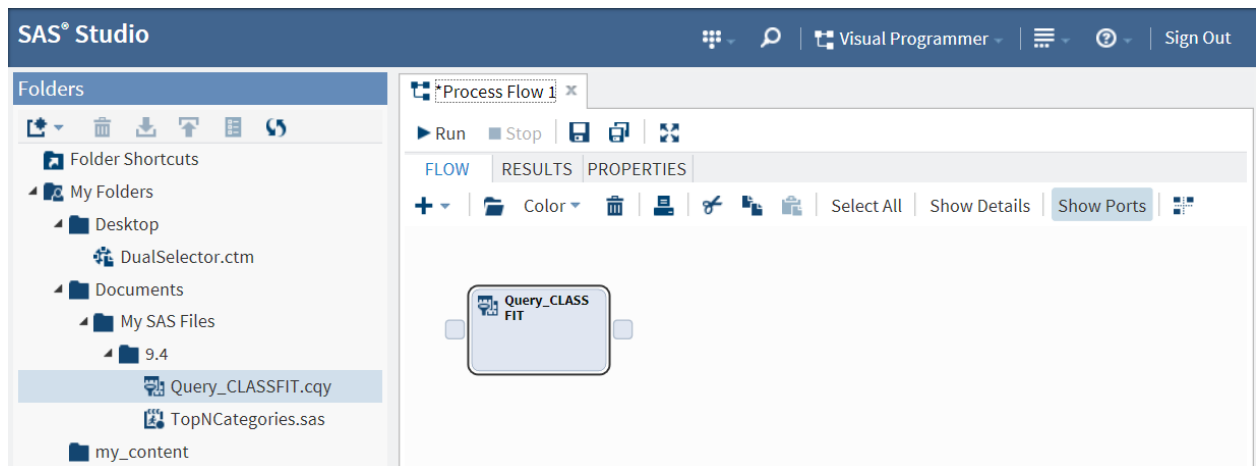
Add a Task to a Process Flow

You can add custom tasks and tasks that shipped with SAS Studio to your process flow.

1 In the navigation pane, click the **Tasks** section.

2 Select the task that you want to add, and then drag it in the process flow.

Here is an example of a process flow that contains the Sort Data task.
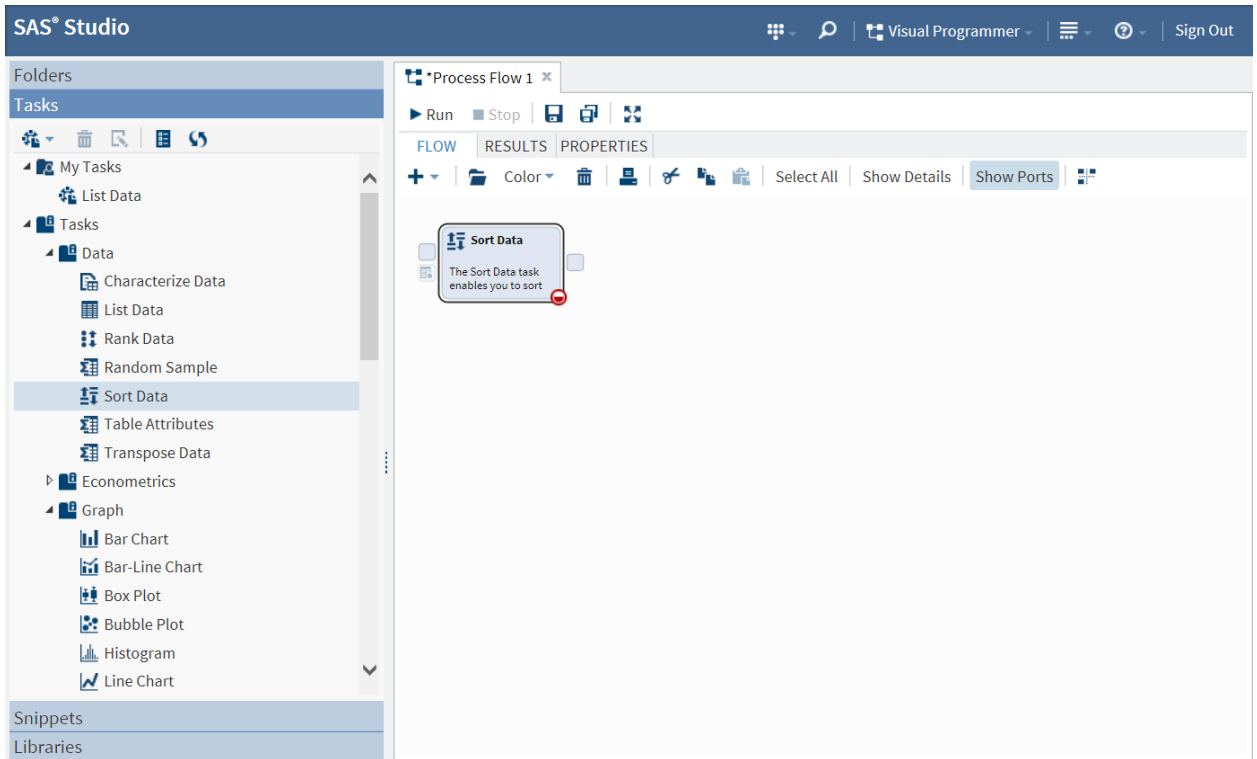
3   Select the task node and click 📁. To run the task, you must specify values for any
    required options.

Here is the user interface for the Sort Data task.

When you run the task, SAS Studio uses the values that you specified for the task options.

Understanding Subflows

What Is a Subflow?

Within a process flow, you could have subflows. These subflows contain one or more objects. The advantage to creating a subflow is that you can easily run the nodes within a subflow without running the entire process flow. If the subflow links to another node in the entire process flow, the subflow must run to completion before SAS Studio runs the subsequent node.

In this example, there are three nodes: a programming node called PROC PRINT, a node for a subflow, and a query node called CLASS=F. All of the nodes in the subflow node must run before SAS Studio runs the CLASS=F node. By default, the node for a subflow is white.
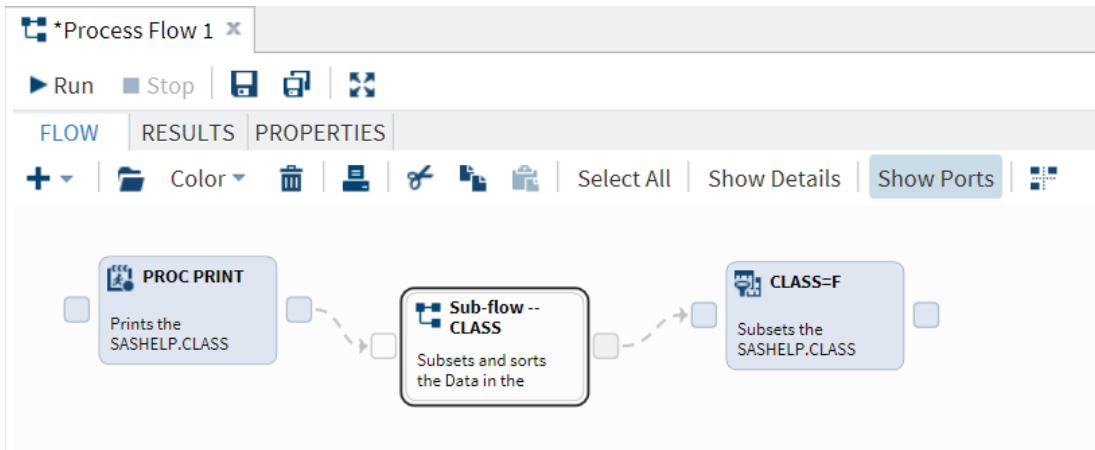
Note: SAS Studio cannot run a process flow that contains an empty subflow. An error message appears instead.

Create a New Subflow

To add a subflow to a process flow:

1 Click **+** and select **Sub-Flow**. A node for the subflow appears in the process flow.

2 Select the subflow node and click 🗀. Use the breadcrumbs to verify that you are working in the subflow.
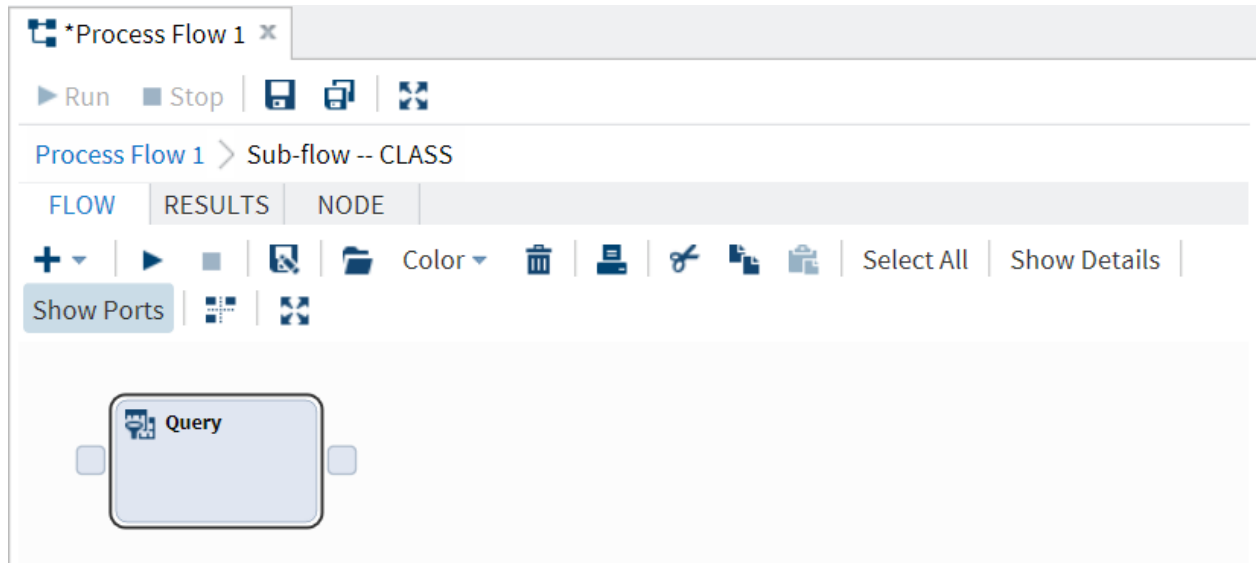
3   Add the content for the subflow. A subflow can contain other subflows.

4   To name the subflow node, to provide a brief description for the subflow, and to
    include any notes, click the **Node** tab.

    After you enter this information, the subflow node in the process flow contains the
    name and description that you specified on the **Node** tab.

Create a Subflow from an Existing Process Flow

You might want to add an existing process flow as a subflow to another process flow.

To add an existing process flow as a subflow:

1   In the navigation pane, click the **Folders** section.

2   Expand the folders in the **Folders** section until you find the program that you want to
    add.

3   Select the process that you want to add, and then drag it in the process flow. (A
    green check mark indicates that you can add this file to the process flow.)

In this example, you are adding the existing PF_Example.cpf file to the process flow in SAS Studio.
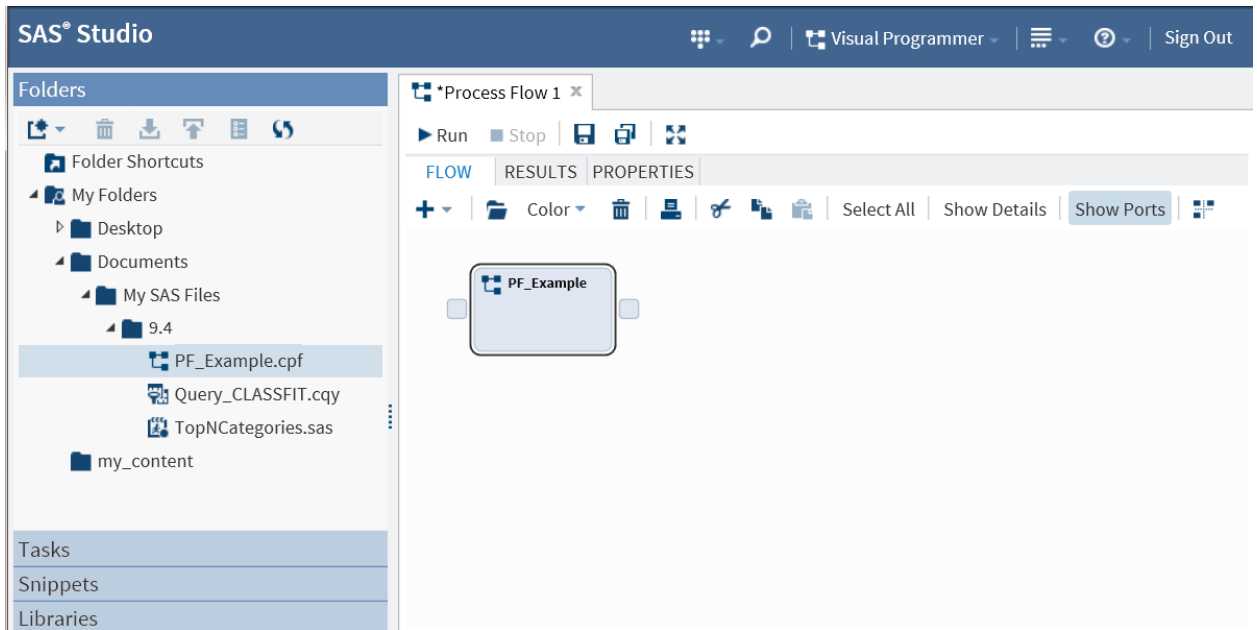
Linking Nodes in a Process Flow

In a process flow, you can specify the order in which each node runs. Generally, each node has an input port and an output port. To view the ports for the nodes in your process flow, click **Show Ports**.

To link nodes in the process flow:

1 Select a control port for the node that you want to link.
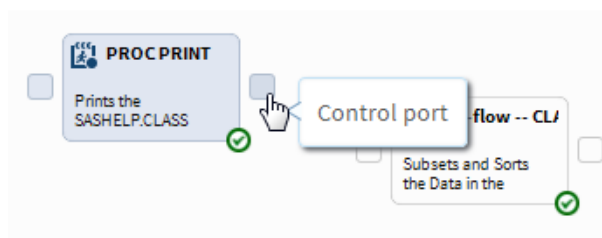
2 Draw an arrow from the control port of one node to the control port of the node that you want to link to.
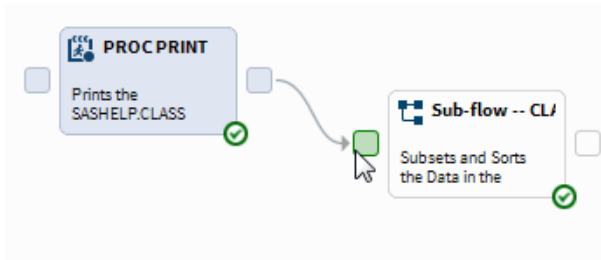
If a link is allowed between the two nodes, the two nodes are connected by a dotted line in the process flow.

Running a Process Flow

To run all the nodes in the process flow, click ▶ **Run**.

To run a select group of nodes, use the mouse to draw a box around the nodes that you want to run. Right-click your selection. From the pop-up menu, select **Run Selected**.

To stop running a process flow, click ■ **Stop**.

Save a Process Flow

To save the current process flow, click 🖫. The process flow is saved as a CPF file in **Folder Shortcuts** or **My Folders**.

5

Working with Data

About the Table Viewer

When you open a table in SAS Studio, you use the table viewer.

Note: The table viewer displays the first 100 rows of the table. If the structure or data values of the table change while the table is open, you must refresh the table viewer to see the changes. If the structure of the table changes and you do not refresh the table, the columns that are listed in the **Libraries** section of the navigation pane might be different from the columns that are displayed in the table viewer.

You can view the properties of the table and its columns by clicking ▦ on the toolbar.

SAS Table Properties ✕

| General | Columns | Extended Attributes | Column Extended Attributes |

Name:	CARS
Description:	2004 Car Data
Type:	Table
Location:	SASHELP.CARS
Rows:	428
Columns:	15
Date created:	Jun 12, 2013, 9:09:53 PM
Date modified:	Jun 12, 2013, 9:09:53 PM

Close

SAS Table Properties ✕

| General | Columns | Extended Attributes | Column Extended Attributes |

Column Name	Type:	Length:	Format:	Informat:	Label:
MAKE	Char	13			
MODEL	Char	40			
TYPE	Char	8			
ORIGIN	Char	6			
DRIVETRAIN	Char	5			
MSRP	Numeric	8	DOLLAR8.		
INVOICE	Numeric	8	DOLLAR8.		
ENGINESIZE	Numeric	8			Engine Size (L)
CYLINDERS	Numeric	8			
HORSEPOWE|	Numeric	8			
MPG_CITY	Numeric	8			MPG (City)
					MPG

Close

The extended attributes tabs enable you to associate additional user-defined characteristics with the table and columns in the table. For example, you could create extended attributes that contain a URL with information about your table or the formula that is used to create a column. For more information about creating extended attributes, see *Base SAS 9.4 Procedures Guide*.

Note: You can add extended attributes only to SAS tables that were created on a SAS 9.4 (or later) server.

Opening and Viewing Data

You can open files in SAS Studio in several ways:

- You can double-click a file in the **Folders** and **Libraries** sections.

- You can drag a file from the **Folders** and **Libraries** sections to the work area.

- You can search for a file and open it from the search results. You can open the file by double-clicking it or by dragging it to the work area.

- You can open a file by using a file shortcut in the **File Shortcuts** section. You can open the file by double-clicking it or by dragging it to the work area.

When you open a table, all of the columns in the table are displayed. You can use the Columns area to specify which columns you want to include in the table viewer. By default, the column names are displayed, but you can choose to display the column labels by selecting **Column labels** from the **View** drop-down list.

Viewing the Query Code That Is Used to Create a Table

While you select options and customize the table to look the way you want it to, SAS Studio is generating SAS code that you can use. To view the query code, click on the toolbar. A new program window appears with the code that was used to create the view of the table in the table viewer. The program is a copy of the query code and is no longer associated with the original query. Editing the code does not affect the data that is displayed in the table viewer, and modifying the table viewer does not affect the contents of the code.

Filtering and Sorting Data

In the table viewer, you can right-click a column heading to filter and sort the data by that column.

Total rows: 19 Total columns: 5 |◄ ← Rows 1-19 → ►|

	Name ▲	Sex	Age	Height
			Sort Ascending	
1	Alfred	M		69
2	Alice	F	Sort Descending	56.5
3	Barbara	F		65.3
			Add Filter	
4	Carol	F	14	62.8
5	Henry	M	14	63.5
6	James	M	12	57.3
7	Jane	F	12	59.8
8	Janet	F	15	62.5
9	Jeffrey	M	13	62.5
10	John	M	12	59
11	Joyce	F	11	51.3
12	Judy	F	14	64.3
13	Louise	F	12	56.3
14	Mary	F	15	66.5
15	Philip	M	16	72
16	Robert	M	12	64.8
17	Ronald	M	15	67
18	Thomas	M	11	57.5
19	William	M	15	66.5

The filter options vary depending on the type of column that you have selected. The Add Filter window for a numeric column enables you to specify one or two filter criteria for each column. To add a second filter criterion, click **+**.

The Add Filter window for a character column enables you to select one or more values in the column.

The Add Filter window for a date column enables you to select a date value from a pop-up calendar window.
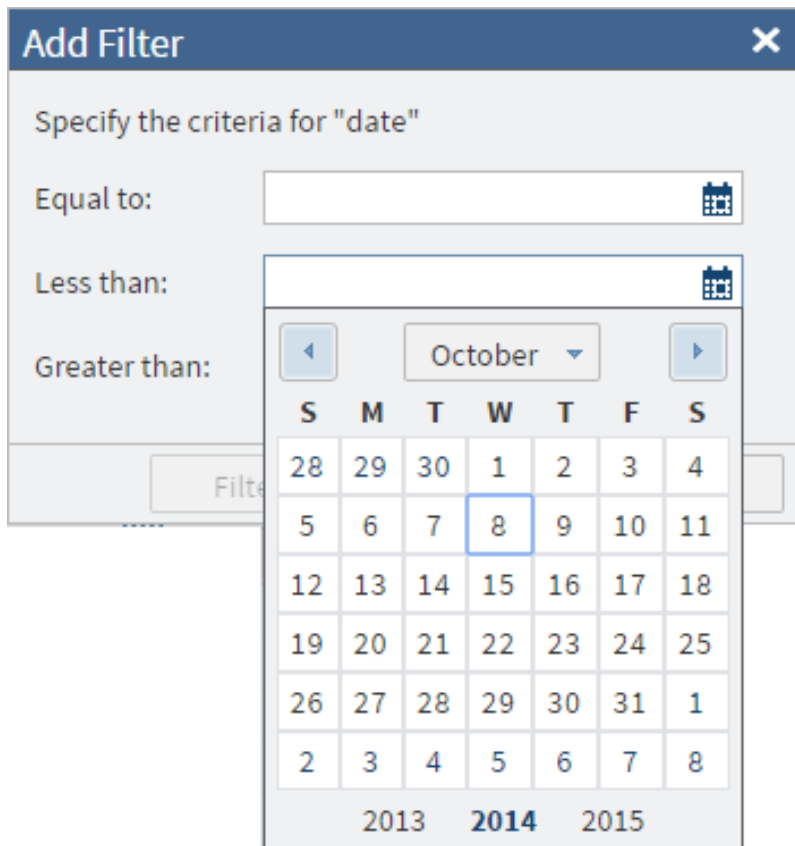
When you create a filter on your data, the filter criteria are displayed at the top of the workspace. You can click ▼ to edit the filter and ⊗ to delete the filter.

Exporting Data

You can use SAS Studio to export your data as another file type to a folder that you specify.

Note: You cannot export your data to an FTP folder.

To export your data:

1 Click **Libraries** in the navigation pane and browse to find the file that you want to export.

2 Right-click the file that you want to export and select **Export**. The Export Table window opens.

3 Select the folder in which you want to save the exported file.

4 In the **Filename** box, enter the name of the exported file.

5 From the **File format** drop-down list, select the format of the exported file.
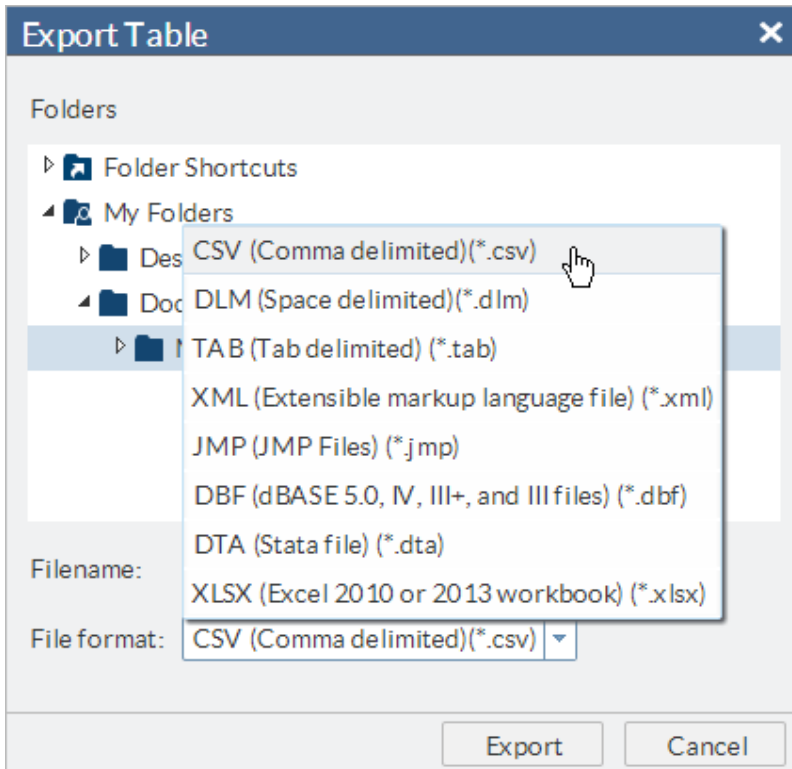
6 Click **Export** to export the file.

6

Working with Results

Viewing Results

When you run a task or a program in SAS Studio, the results are displayed in the work area. You can save the results as an HTML, PDF, or RTF file. You can also download any generated data.

SASHELP.CLASS ×   *List Data ×

| Settings | Code/Results | Split | | | Log | Code |

| DATA | OPTIONS | INFORMATION | | CODE | LOG | RESULTS |

◢ DATA

SASHELP.CARS ▾

◢ ROLES

List variables: ⬆ ⬇ 🗑 ✚

🅐 Model
🔢 MSRP
🔢 MPG_City

Group analysis by: ⬆ ⬇ 🗑 ✚

🅐 Make

Total of: ⬆ ⬇ 🗑 ✚

🔢 Column

Identifying label: ⬆ ⬇ 🗑 ✚

🔷 Column

List Data for SASHELP.CARS

Make=Acura

Row number	Model	MSRP	MPG (City)
1	MDX	$36,945	17
2	RSX Type S 2dr	$23,820	24
3	TSX 4dr	$26,990	22
4	TL 4dr	$33,195	20
5	3.5 RL 4dr	$43,755	18
6	3.5 RL w/Navigation 4dr	$46,100	18
7	NSX coupe 2dr manual S	$89,765	17

Make=Audi

Row number	Model	MSRP	MPG (City)
8	A4 1.8T 4dr	$25,940	22
9	A41.8T convertible 2dr	$35,940	23
10	A4 3.0 4dr	$31,840	20
11	A4 3.0 Quattro 4dr manual	$33,430	17
12	A4 3.0 Quattro 4dr auto	$34,480	18
13	A6 3.0 4dr	$36,640	20
14	A6 3.0 Quattro 4dr	$39,640	18
15	A4 3.0 convertible 2dr	$42,490	20
16	A4 3.0 Quattro convertible 2dr	$44,240	18
17	A6 2.7 Turbo Quattro 4dr	$42,840	18
18	A6 4.2 Quattro 4dr	$49,690	17
19	A8 L Quattro 4dr	$69,190	17
20	S4 Quattro 4dr	$48,040	14
21	RS 6 4dr	$84,600	15
22	TT 1.8 convertible 2dr (coupe)	$35,940	20
23	TT 1.8 Quattro 2dr (convertible)	$37,390	20
24	TT 3.2 coupe 2dr (convertible)	$40,590	21
25	A6 3.0 Avant Quattro	$40,840	18

Sending Your Results to Another User
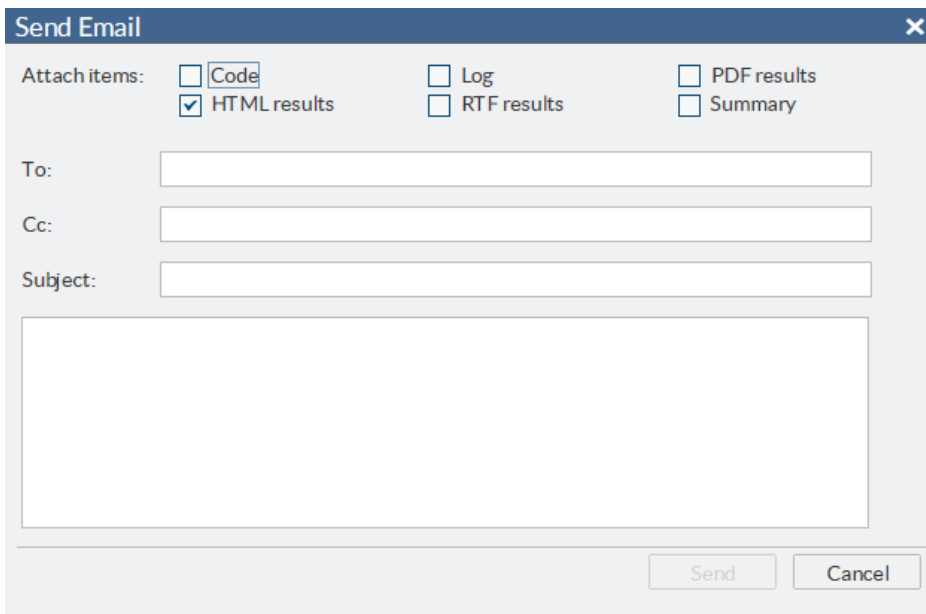
You can send a copy of your results and the associated code and log files to another user through electronic mail. Files that you can send include results in HTML, RTF, and PDF formats as well as the code and log files that are associated with the results. You can also send a Program Summary file, which includes information about the program execution, the complete SAS source code, the complete SAS log, and the results. The code, log, and program summary files are sent as HTML files. To send files through e-

mail, you need access to an SMTP server. For more information, contact your site administrator.

Note: If your SAS Studio e-mail messages are being marked as junk mail, see *SAS Studio: Administrator's Guide* for information about your configuration file.

To send results by e-mail:

1 On the toolbar for your results, click 📧. The Send Email window opens.

2 Select the items that you want to include as attachments to your e-mail. By default, the HTML results are selected.

3 In the **To** box, enter the e-mail addresses to which you want to send the files. Separate addresses with a semicolon.

4 If you want to send a copy of the e-mail to another address, enter the address in the **Cc** box.

5 In the **Subject** box, enter a subject for the e-mail. You can also add a message to include in the body of the e-mail.

6 Click **Send** to send the message and attachments.

About the SAS Output Delivery System

The SAS Output Delivery System (ODS) gives you greater flexibility in generating, storing, and reproducing SAS procedure and DATA step output along with a wide range of formatting options. ODS provides formatting functionality that is not available when using individual procedures or the DATA step without ODS.

SAS Studio uses very specific ODS options and the GOPTIONS statements so that the output is displayed properly in the web environment. To view all of the ODS options in

your code, click and select **Preferences**. In the Preferences window, click **General** and select the **Show generated code in the SAS log** option.

Note: To ensure that your output is displayed properly, do not change the settings of the ODS options or GOPTIONS statements in the generated code.

About SAS ODS Statistical Graphics

About SAS ODS Statistical Graphics

SAS ODS Statistical Graphics, more commonly referred to as SAS ODS Graphics, is an extension of the SAS Output Delivery System (ODS). ODS manages all output that is created by procedures and enables you to display the output in a variety of forms, including HTML and PDF.

Many SAS analytical procedures use ODS Graphics functionality to produce graphs as automatically as these procedures produce tables. ODS Graphics uses the Graph Template Language (GTL) syntax, which provides the power and flexibility to create many complex graphs. The GTL is a comprehensive language for defining statistical graphics.

In SAS Studio, you can use the ODS Graphics Designer to define these statistical graphics without knowing the GTL. After a graph definition is created, you can use that graph definition to create an ODS statistical graph in SAS Studio.

SAS ODS Graphics Designer

What Is the SAS ODS Graphics Designer?

The SAS ODS Graphics Designer is an interactive graphical application that you can use to create and design custom graphs. The designer creates graphs that are based on the Graph Template Language (GTL), which is the same language that is used by SAS analytical procedures and SAS ODS Graphics procedures. The ODS Graphics Designer provides a graphical user interface so that you can design graphs easily without knowing the details of templates and the GTL.

Using point-and-click interaction, you can create simple or complex graphical views of data for analysis. The ODS Graphics Designer enables you to design sophisticated graphs by using a wide array of plot types. You can design multi-cell graphs, classification panels, and scatter plot matrices. Your graphs can have titles, footnotes, legends, and other graphics elements. You can save the results as an image for inclusion in a report or as an ODS Graphics Designer file (SGD) that you can later edit.

For more information, see *SAS ODS Graphics Designer: User's Guide*, which is available from support.sas.com.

How to Install the SAS ODS Graphics Designer

If you have SAS Foundation installed on your machine, the SAS ODS Graphics Designer is already available. For example, if you are using the single-user edition of SAS Studio, the SAS ODS Graphics Designer is already installed because you are running SAS Foundation and SAS Studio on the same machine.

Note: Only the Windows version of the SAS ODS Graphics Designer is supported when you are running SAS Studio. The SAS ODS Graphics Designer must be installed on the same machine as the browser you are using to access SAS Studio. (SAS Studio might be installed on a different machine.) If multiple users are accessing SAS Studio and these users are on different machines, SAS ODS Graphics Designer must be installed on each user's machine.

To install the SAS ODS Graphics Designer:

1   Click ☰. Select **Tools** ▸ **Install ODS Graphics Designer**. The downloads and hot
    fixes page for Base SAS Software on support.sas.com opens.

2   Under the SAS 9.4M2 heading, click **SAS ODS Graphics Designer**.

3   From the list of download pages, click **Request download** for your Windows
    operating environment and follow the subsequent installation steps.

Open the SAS ODS Graphics Designer

After the SAS ODS Graphics Designer is installed, you can open it by using a menu

option in SAS Studio. To open SAS ODS Graphics Designer, click ☰ and select **Tools**.
Then select **ODS Graphics Designer**.
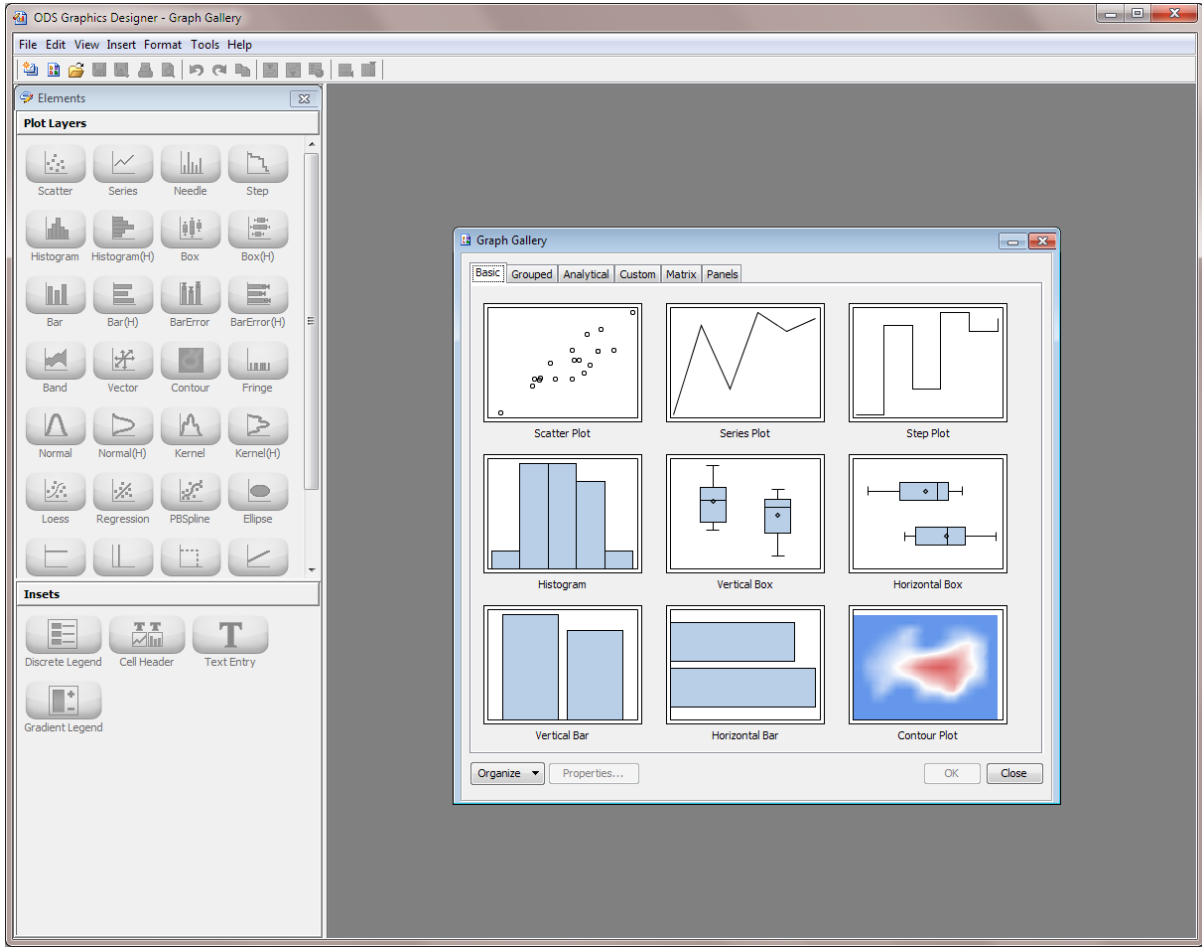
SAS ODS Graphics Editor

What Is the SAS ODS Graphics Editor?

The ODS Graphics Editor enables you to edit the various elements in the output graph while keeping the underlying data unchanged. In addition, you can annotate a graph by inserting text, lines, arrows, images, and other items in a layer above the graph. You can save the results of your customization as an ODS Graphics Editor (SGE) file and make incremental changes to the file. You can also save the results as a Portable Network Graphics (PNG) image file for inclusion in other documents.

For more information about the SAS ODS Graphics Editor, see *SAS ODS Graphics Editor: User's Guide*, which is available from support.sas.com.

How to Install the SAS ODS Graphics Editor

When you install the SAS ODS Graphics Editor, SAS Studio automatically creates the `~/Projects/ODSEditorFiles` directory.

Note: If you are running the single-user edition of SAS Studio, then the SAS ODS Graphics Editor is already installed.

To install the SAS ODS Graphics Editor:

1   Click ☰. Select **Tools ▶ Install ODS Graphics Editor**. The downloads and hot fixes page for Base SAS Software on support.sas.com opens.

2   For your release of SAS, click **ODS Graphics Editor**. (For example, if you are running on SAS 9.4, select ODS Graphics Editor under the SAS 9.4 heading.)

3   From the list of download pages, click **Request download** for your Windows operating environment and follow the subsequent installation steps.

How to Edit Your Graphics Output

1   Include this statement in your SAS code so that you can edit your graphics output:

```
ods listing sge=on gpath="{home}/Projects/ODSEditorFiles";
```

When you run this program, the graphical output is saved as an SGE file in your `~/Projects/ODSEditorFiles` directory.

2   In the **Folders** section of the navigation pane, expand the ODSEditorFiles folder.

Folders

netdocs

▲ ODSEditorFiles

　SGPanel.png

　SGPanel.sge

　SGPanel1.sge

　SGPanel2.png

　SGPlot.sge

　SGPlot1.sge

　SGPlot10.png

　SGPlot2.png

　SGPlot2.sge

　SGPlot3.sge

　SGPlot4.png

　SGPlot4.sge

　SGPlot6.png

　SGPlot8.png

　SGScatter.png

　SGScatter.sge
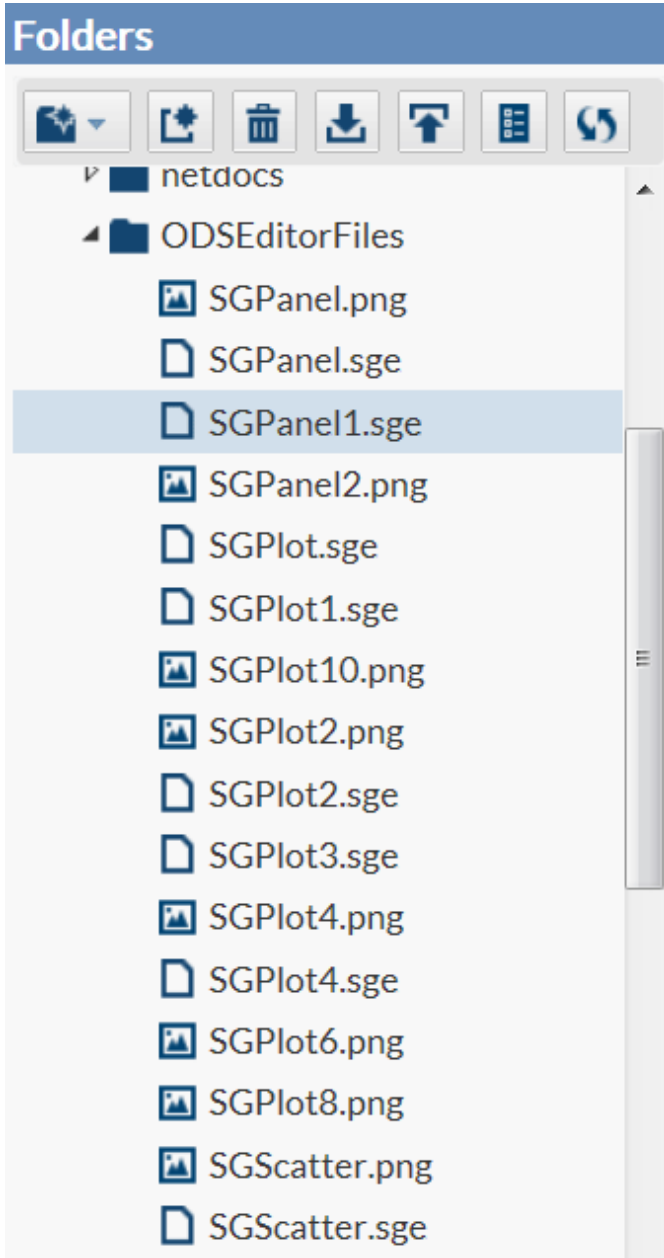
3 Double-click the filename to open the graph in the SAS ODS Graphics Editor.

For example, here is the SGPanel1.sge file in the SAS ODS Graphics Editor.
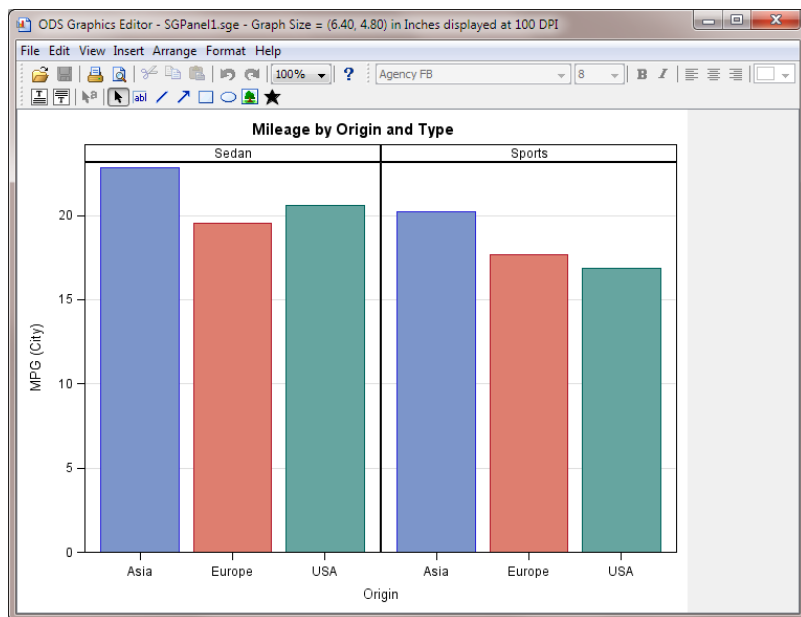
Note: The default list of files in your ODSEditorFiles folder is created by the code snippets in the **Snippets** section. For more information, see "Working with Code Snippets" on page 34.

Specifying the Style for Your Results

The Preferenceswindow enables you to change several options that affect how your results are displayed.

To access the editor options, click ▤ and select **Preferences**. Click **Results**.

Option	Description
Display warning if results are larger than n MB	displays a warning message when you attempt to open a results file that is larger than n megabytes (MB). The default value is 4 MB.

Option	Description
HTML output style	displays the style that is applied to results in HTML. To change the style that is applied to the results, select another style from the drop-down list.
Produce PDF output	generates results in PDF format. This option is selected by default.
PDF output style	displays the style that is applied to results in PDF. To change the style that is applied to the results, select another style from the drop-down list.
Generate the default table of contents	automatically creates a table of contents in the PDF file.
Produce RTF output	generates results in RTF format. This option is selected by default.
RTF output style	displays the style that is applied to results in RTF. To change the style that is applied to the results, select another style from the drop-down list.

7

Understanding Tasks in SAS Studio

What Is a Task?

A task is an XML and Apache Velocity code file that generates SAS code and formats results for you. Tasks include SAS procedures from simple data listings to complex analytical procedures. SAS Studio is shipped with several predefined tasks. You can edit a copy of these predefined tasks in order to customize the tasks for your site. You can also build your own tasks.

How to Run a Task

To run a predefined task:

1   In the navigation pane, click the **Tasks** section.

2   Expand the folder that contains the task.

3   Right-click the task name and select **Open**. Alternatively, you can double-click the task to open it.

    The task opens to the right of the work area.
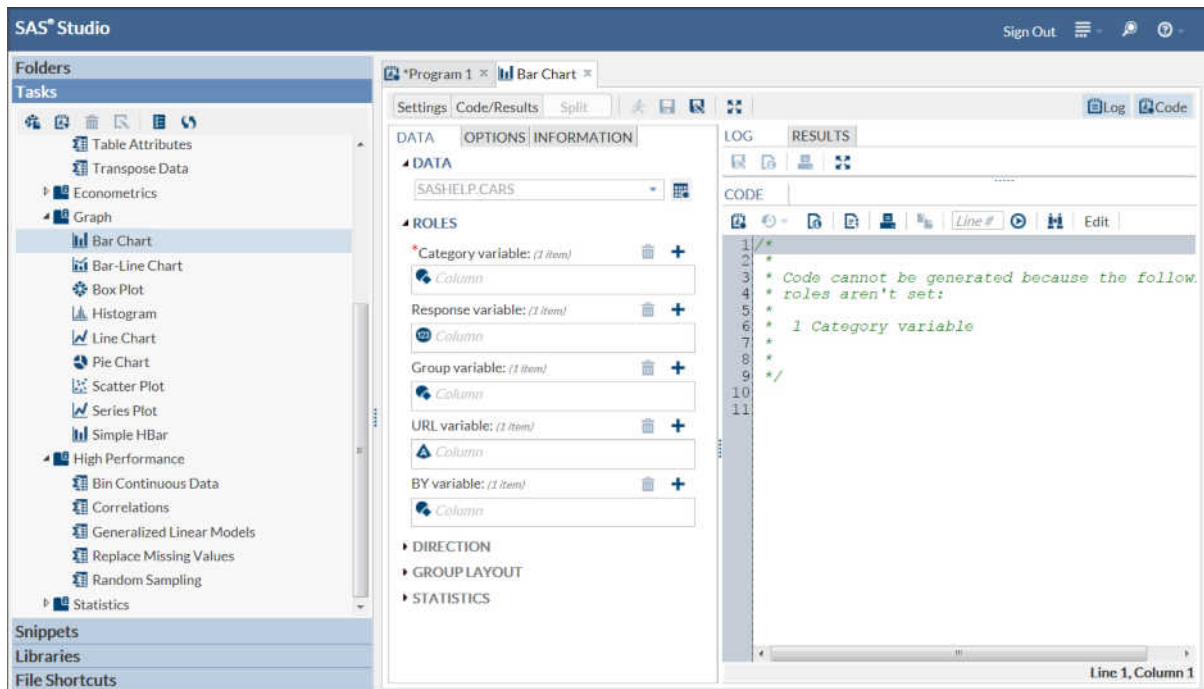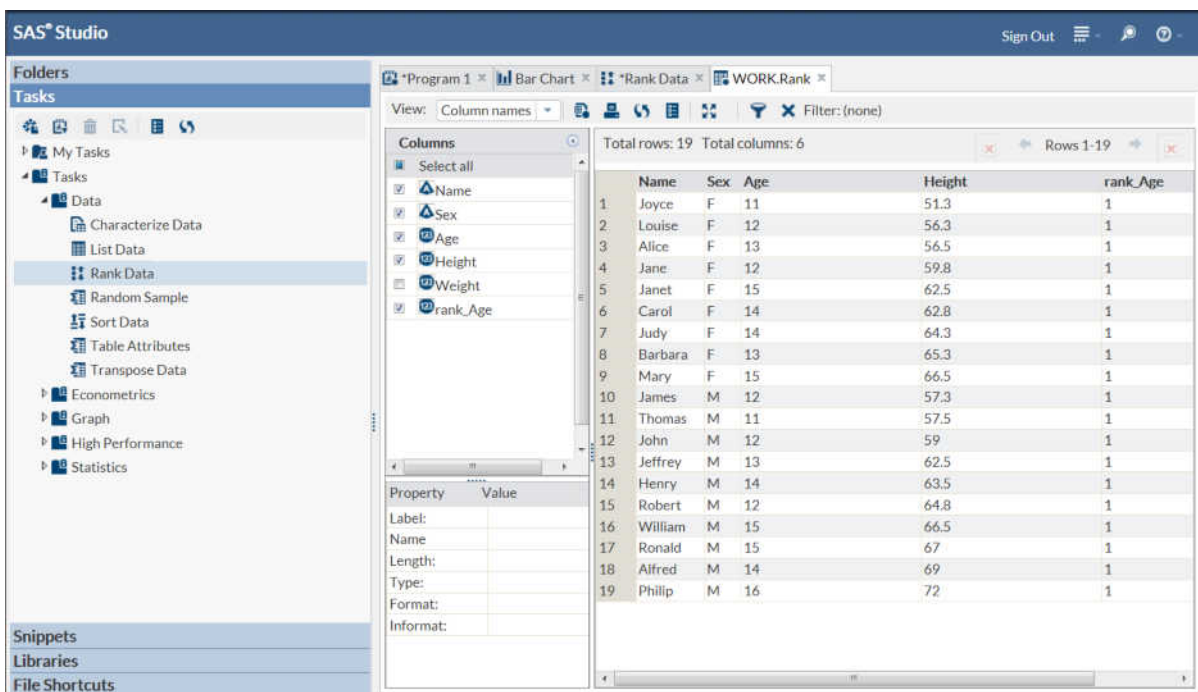
4   If the **Data** tab is available, specify an input data source and select columns for the roles in the data source. A role is a description of a variable's purpose in the task. To add a column to a role, click ➕. A list of available columns for that role appears. If only one column can be assigned to the role, you select a column and the list

disappears. If multiple columns can be assigned, you can press Ctrl or Shift to select multiple columns from the list and click **OK**.

5 On the remaining tabs, specify any other required options, which are denoted with a red asterisk. As you assign values to the task, the relevant SAS code is generated.

6 To run the task, click ![run icon].

If the task generates output data, the table opens in your work area.

If the task generates results, the output appears on the **Results** tab under the tab for the current task.

Save a Task and Its Option Settings

If you use a task frequently, you might want to save the task after you specify the input data source and the option settings. In SAS Studio, you can save a task as a CTK file in your **Folders** directory. The next time that you need to run the task, double-click the task in your **Folders** directory and the task appears with all of your previous settings.

Note: Before you can save a task, you must specify an input data set and all the options that are required to run the task.

To save a task:

1   Click ![icon]. The Save As window appears.

2   Select the location where you want to save the task file. You can save this file in the **Folders** section or in your **My Tasks** folder. Specify a name for this file. For the file type, select **CTK Files (*.CTK)**. Click **Save**.

In this example, the task is now available from the **Folders** section.
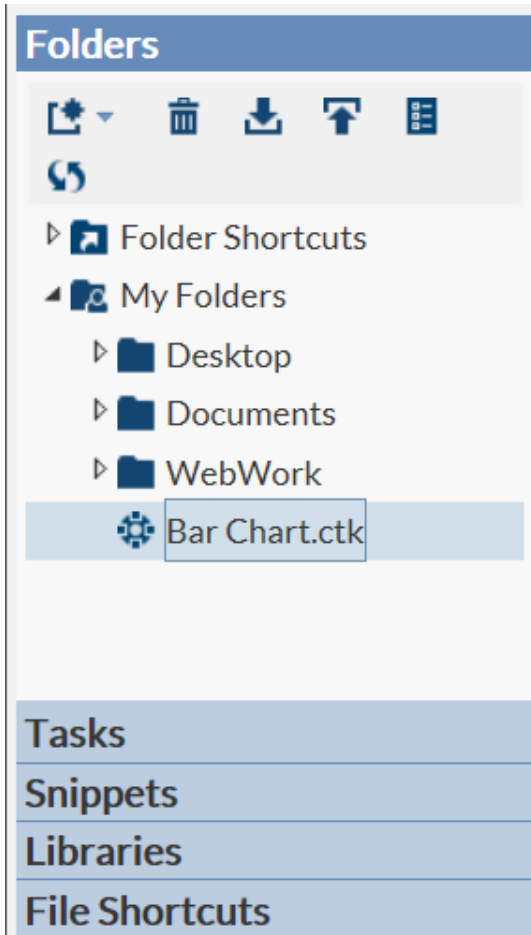
Note: In the **Tasks** section, you are still working with this task. If you save the task again, the CTK file in the **Folders** section is updated.

Edit a Predefined Task

To customize the predefined tasks for your site, you can edit the XML code that is used to create the task.

To edit a predefined task:

1   In the navigation pane, open the **Tasks** section.

2   Expand the folder that contains the task.

3   Right-click the name of the task that you want to edit and select **Add to My Tasks**. The Add to My Tasks window opens.

4   Specify a name and description for the task. By default, the name and description from the predefined task is used. Using the **Category** drop-down list, you can also specify where to save a copy of this task in the **My Tasks** folder. If you select **(none)**, the task is added directly to the **My Tasks** folder.

   Click **Add**.
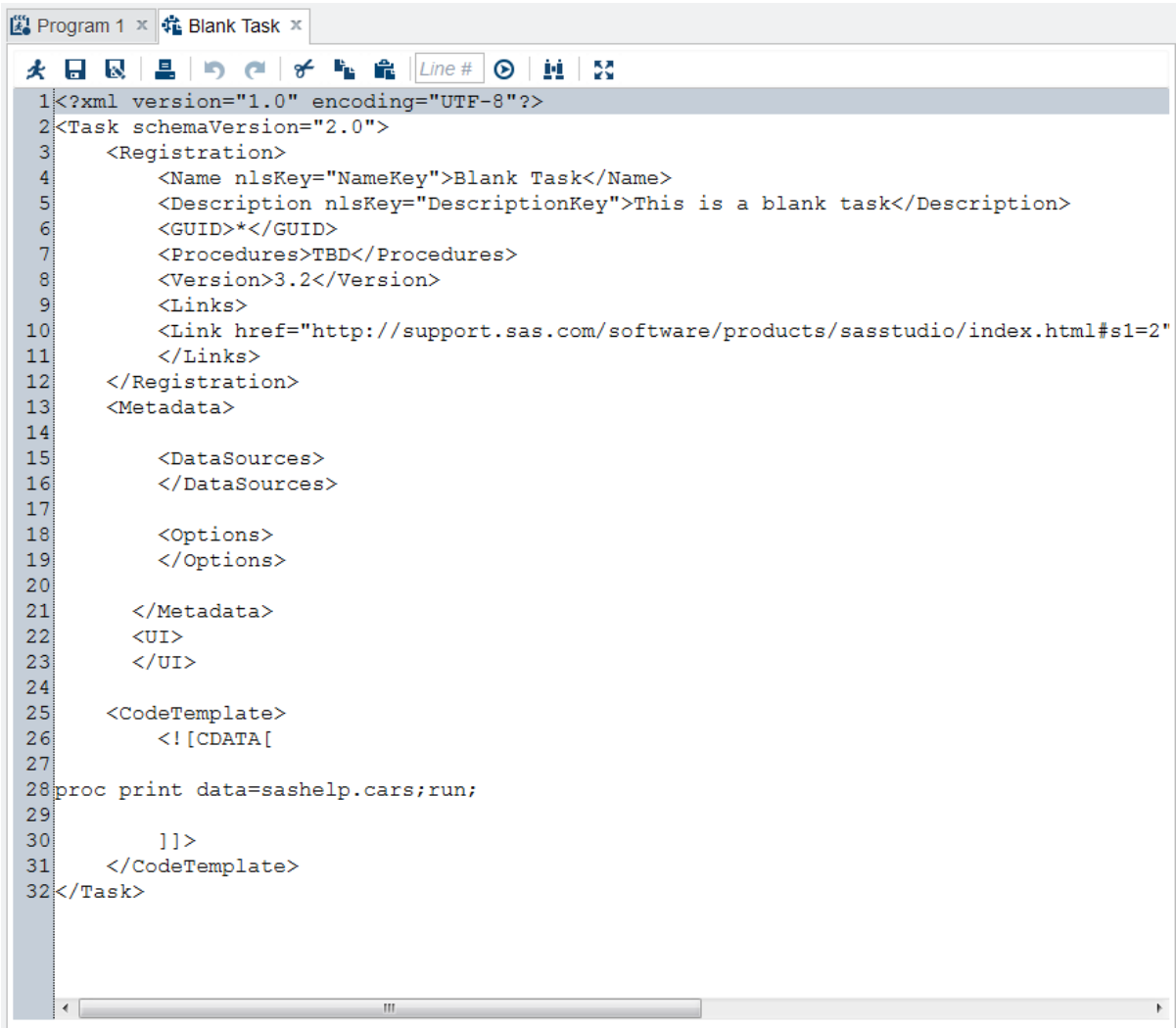
5   Open the **My Tasks** folder and select the copied task.

6   Click . The XML file for the task appears.

7   Edit the XML file and save your changes. To preview your changes, click .

Create a New Task

SAS Studio provides a template that you can use to create custom tasks for your site.

To create a custom task:

1   In the navigation pane, open the **Tasks** section.

2   Click and select **Blank Task**. A task template opens.

Program 1 ×   Blank Task ×

```
 1 <?xml version="1.0" encoding="UTF-8"?>
 2 <Task schemaVersion="2.0">
 3     <Registration>
 4         <Name nlsKey="NameKey">Blank Task</Name>
 5         <Description nlsKey="DescriptionKey">This is a blank task</Description>
 6         <GUID>*</GUID>
 7         <Procedures>TBD</Procedures>
 8         <Version>3.2</Version>
 9         <Links>
10         <Link href="http://support.sas.com/software/products/sasstudio/index.html#s1=2"
11         </Links>
12     </Registration>
13     <Metadata>
14
15         <DataSources>
16         </DataSources>
17
18         <Options>
19         </Options>
20
21     </Metadata>
22     <UI>
23     </UI>
24
25     <CodeTemplate>
26         <![CDATA[
27
28 proc print data=sashelp.cars;run;
29
30         ]]>
31     </CodeTemplate>
32 </Task>
```

3 Edit the code in the task template to create your task. To view the user interface for the task template, click . In the user interface for the task template, you can see examples of radio buttons, check boxes, combination boxes, and other types of options. For more information about this file, see *SAS Studio: Developer's Guide*.

4 Click .

Note: The name of the task cannot include these special characters:

[ ] { } , ? / \ * # " % + | < > :

Customizing the Task Code and the Task Layout in the Workspace

The Preferences window enables you to change several options that affect what and how the task code is displayed.

To access these options, click ⚏ and select **Preferences**. Click **Tasks**.

Trim all leading and trailing blank spaces in generated code	removes any blank spaces that appear before or after the generated code.
Generate header comments for task code	adds comments before the generated code for a SAS task.
Automatically format generated code	automatically formats any code that is generated by a task and displayed in the code editor.
View	specifies how to lay out the task options, task code, and task results in your workspace. You can choose from these options:
	■ **Split** — displays the task settings, the code, and the results for the task.
	■ **Settings** — displays only the options for the task in the workspace.
	■ **Code/Results** — displays the SAS code, the log, and any results in the workspace.
Show task code	specifies whether to display the SAS code for the task when you select the **Split** view or the **Code/Results** view.
Show task log	specifies whether to display the log that is generated when you run the task. This option is available only if you select the **Split** view or the **Code/Results** view.

Display task settings on right   displays the task option to the right of the SAS Studio workspace. By default, the task options are displayed to the left.

8

Data Tasks

Characterize Data Task

About the Characterize Data Task

The Characterize Data task creates a summary report, graphs, and frequency and univariate SAS data sets that describe the main characteristics of the data.

Example: Characterize Data Task

In this example, you want a better understanding of the contents in the Sashelp.Pricedata data set.

To create this example:

1 In the **Tasks** section, expand the **Data** folder and double-click **Characterize Data**. The user interface for the Characterize Data task opens.

2 On the **Data** tab, select the **SASHELP.PRICEDATA** data set.

3 To run the task, click ⚡.

Here is a sample of the results:

Summary of Character Variables for SASHELP.PRICEDATA
Limited to the 30 Most Frequent Distinct Values per Variable

Variable	Label	Value	Frequency Count	Percent of Total Frequency
		Line2	240	23.5294
		Line3	240	23.5294
productLine	Name of product line	Line4	240	23.5294
		Line1	180	17.6471
		Line5	120	11.7647

Variable	Label	Value	Frequency Count	Percent of Total Frequency
		Product1	60	5.8824
		Product10	60	5.8824
		Product11	60	5.8824
		Product12	60	5.8824
		Product13	60	5.8824
		Product14	60	5.8824

Summary of Numeric Variables for SASHELP.PRICEDATA

Variable	Label	N	NMiss	Total	Min	Mean	Median	Max	StdMean
cost	Unit Cost	1020	0	37419.00	16.400	36.685	29.7	78.0	0.55372

Variable	Label	N	NMiss	Total	Min	Mean	Median	Max	StdMean
discount	Price Discount	1020	0	11.50	0.000	0.011	0.0	0.2	0.00122

Variable	Label	N	NMiss	Total	Min	Mean	Median	Max	StdMean
line	Product Line ID	1020	0	2940.00	1.000	2.882	3.0	5.0	0.04003

Variable	Label	N	NMiss	Total	Min	Mean	Median	Max	StdMean
price	Unit Price	1020	0	80304.68	26.720	78.730	65.2	171.4	1.23613

Variable	Label	N	NMiss	Total	Min	Mean	Median	Max	StdMean
price1	Product 1 Unit Price	1020	0	52723.63	44.455	51.690	52.3	52.3	0.06410

Variable	Label	N	NMiss	Total	Min	Mean	Median	Max	StdMean
price10	Product 10 Unit Price	1020	0	57312.53	45.520	56.189	56.9	56.9	0.07740

Categorical Variable Value Counts for SASHELP.PRICEDATA
Variable=productLine
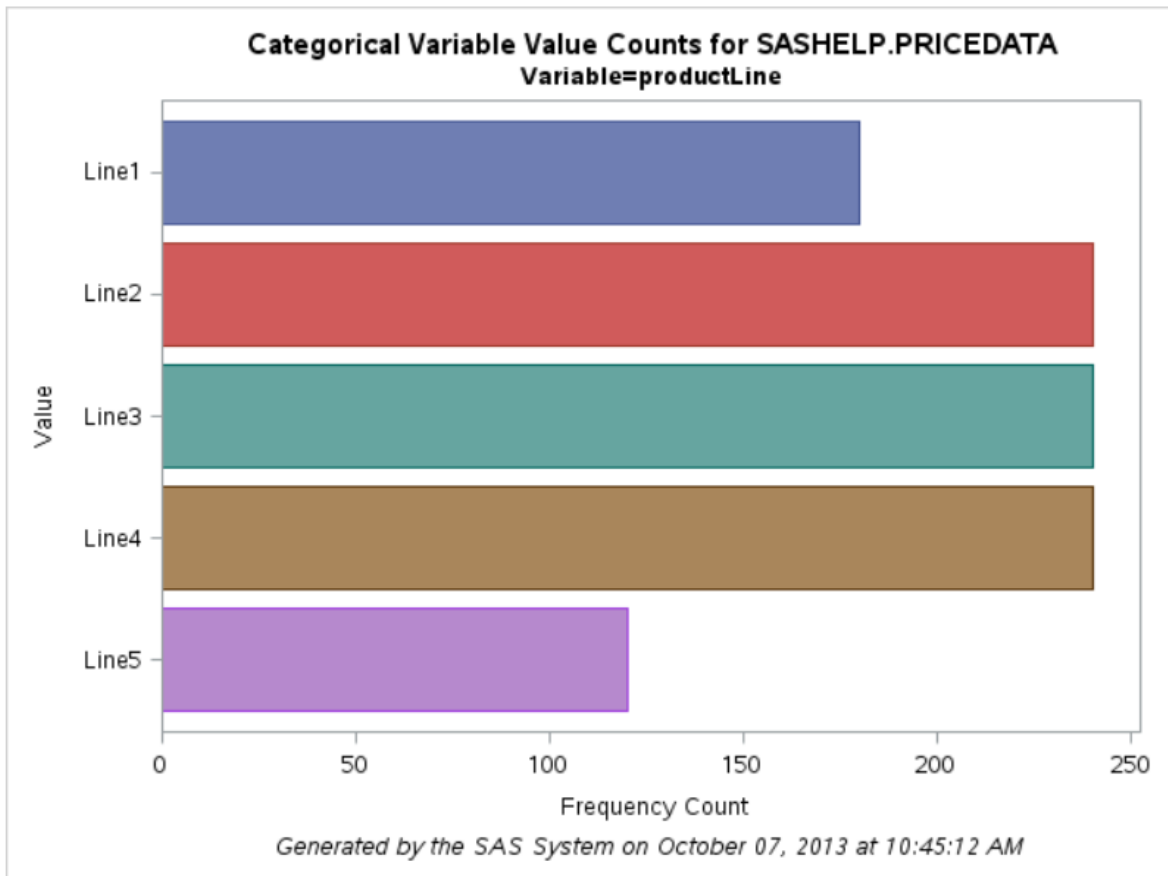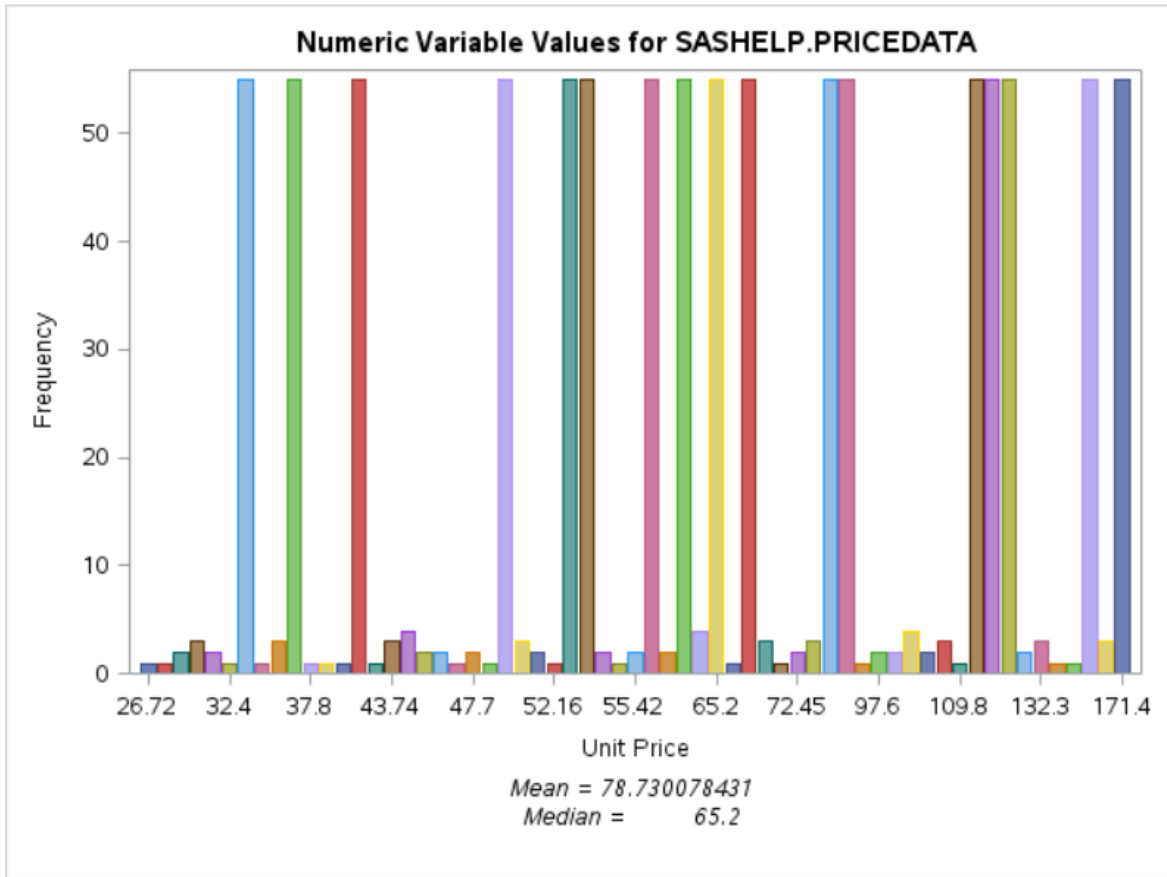
By default, the task also creates output data—a table with the frequency data and a table with the univariate data. Both of these tables are saved in the Work library.

Assigning Data to Roles

You must select a data source to run the Characterize Data task. However, no roles are available.

Setting Options

Option Name	Description
Output Options	You must select at least one output option. By default, a summary report, graphs, and output tables for the frequency data and univariate data are created.
Limit categorical values	Specifies the maximum number of categorical values to report per variable. By default, 30 values are reported. You can change this maximum value in the **Maximum number of unique values per variable** box.

Combinations Task

About the Combinations Task

The Combinations task computes the possible combinations of the total number of objects into sets with a specified number in each set.

Example: Calculating the Combinations of 52 Objects in 5 Sets

To create this example:
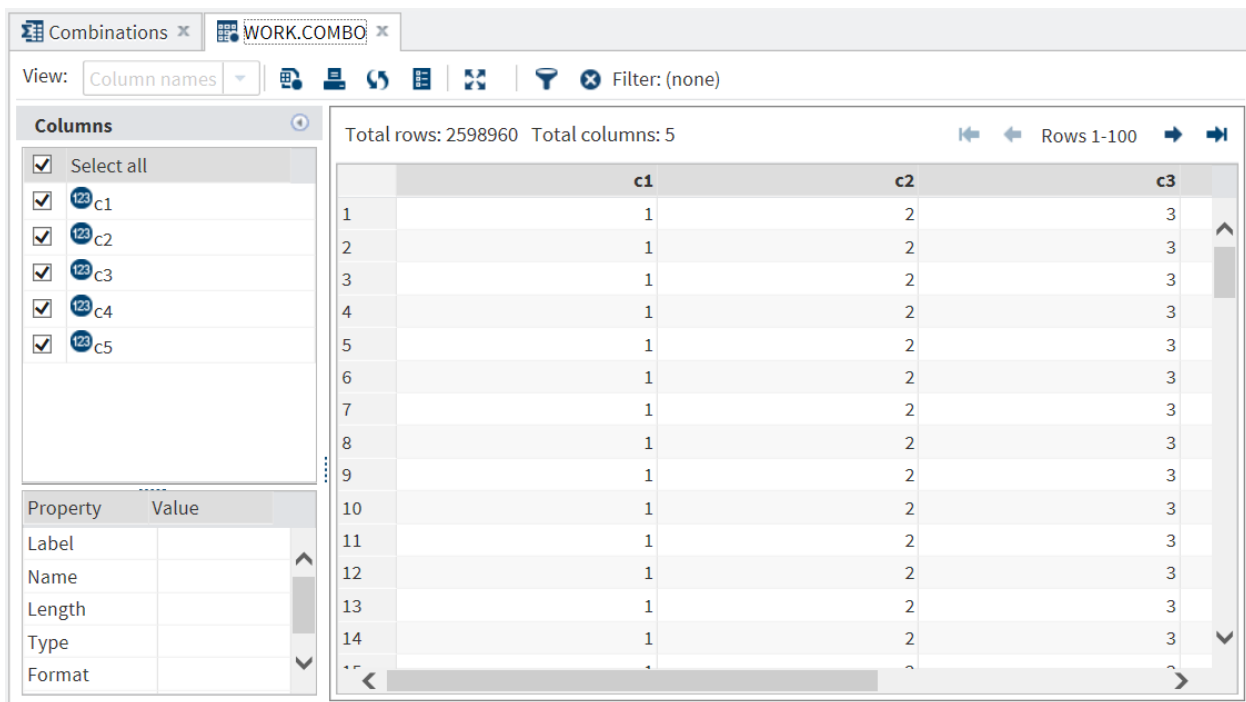
1 In the **Tasks** section, expand the **Combinatorics and Probability** folder and double-click **Combinations**. The user interface for the Combinations task opens.

2 On the **Options** tab, specify these options.

Assign columns to these roles:

Option Name	Value to Specify
Total number of objects	`52`
Number of objects in a set	`5`
Data set name	`Combo`

3 To run the task, click .

The Combinations task creates an output data set. In SAS Studio, this data set opens on the **WORK.Combo** tab.

Setting Options

All of these options are required to run the Combinations task.

Option Name	Description
Observations	
Total number of objects	specifies the number of objects.
Number of objects in a set	specifies the number of objects in a set.
Output Data Set	
Data set name	specifies the name of the output data set.

List Data Task

About the List Data Task

The List Data task displays the contents of a table as a report. For example, you can use the List Data task to create a report that sums the expenses and revenues for each sales region.

Example: Reports of Drive Train, MSRP, and Engine Size by Car Type

In this example, you want to create reports for each car type. Each report lists the drive train, MSRP, and engine size.

To create this example:

1   In the **Tasks** section, expand the **Data** folder and double-click **List Data**. The user interface for the List Data task opens.

2 On the **Data** tab, select the **SASHELP.CARS** data set.

3 Assign columns to these roles:

Role	Column Name
List variables	**DriveTrain**
	MSRP
	EngineSize
Group analysis by	**Type**

4 To run the task, click .

Here is a subset of the results:

List Data for SASHELP.CARS

Type=Hybrid

Row number	DriveTrain	MSRP	Engine Size (L)
1	Front	$20,140	1.4
2	Front	$19,110	2.0
3	Front	$20,510	1.5

Type=SUV

Row number	DriveTrain	MSRP	Engine Size (L)
4	All	$36,945	3.5
5	All	$37,000	3.0
6	All	$52,195	4.4
7	All	$37,895	4.2
8	Front	$26,545	3.4
9	Front	$52,795	5.3
10	Front	$46,995	4.6
11	Front	$42,735	5.3
12	All	$41,465	5.3
13	Front	$30,295	4.2
14	Front	$20,255	2.5
15	All	$32,235	4.7

Assigning Data to Roles

Role	Description
List variables	Prints the variables in the order in which they are listed.

Role	Description
Group analysis by	When you assign one or more variables to this role, the table is sorted by the selected variable or variables, and a listing is generated for each distinct value, or BY group, in the variable or combination of variables.
Total of	Prints the sum of the selected variable at the bottom of the listing report.
Identifying label	When you specify one or more variables in this role, the List Data task uses the formatted values of these variables to identify the rows, rather than observation numbers (designated in the results by the column heading "Obs").

Setting Options

Option Name	Description
Basic Options	
Display row numbers	Includes in the output a column that lists the row number for each observation.
	You can specify a label for this column in the **Column label** text box. By default, the name of this column is **Row number**.
Use column labels as column headings	Uses the column label instead of the column name as the column heading.
Display number of rows	Reports the number of rows in the table at the end of the output, or the number of rows in each BY group at the end of each BY group's output.

Option Name	Description
Round values before summing the variable	Rounds each numeric value to the number of decimal places in its format, or to two decimal places if no format is specified. If this option is specified, the List Data task performs the rounding before summing the variable.
Heading direction	Column headings can be printed horizontally or vertically, or you can select **Default** and let SAS determine the optimal arrangement for each column.

Column width	Specifies how the List Data task determines column widths:
	
	Default
	    determines the column widths on a per-page basis.
	
	Full
	    uses a format width (or default width if no format is specified) for all pages.
	
	Minimum
	    uses the smallest possible column width on a per-page basis.
	
	Uniform
	    reads the entire table to determine the appropriate column widths before generating output. When this option is not selected, different pages could have different widths for the same column.
	
	Uniform by
	    formats all columns uniformly within a BY group, using each variable's formatted width as its column width. If the variable does not have a format that explicitly specifies a field width, the task uses the widest data value as the column width.

Option Name	Description

Split labels	If the variable labels contain one of the split characters (*, !, @, #, $, %, ^, &, or +), the labels will be split at the split character or characters. For example, for a variable label that reads "This is*a label" and the * character is selected as the split character, the column heading will read
	
	`This is`
	`a label`
	
	You do not need to select both the **Use variable label as column headings** and **Split labels** options. The **Split labels** option implies that you want to use variable labels.

Rows to list	specifies the number of rows to list in the output. By default, all rows are listed.

Permutations Task

About the Permutations Task

The Permutations task computes the possible permutations of a given number of objects.

Example: Calculating the Permutations of Six Objects
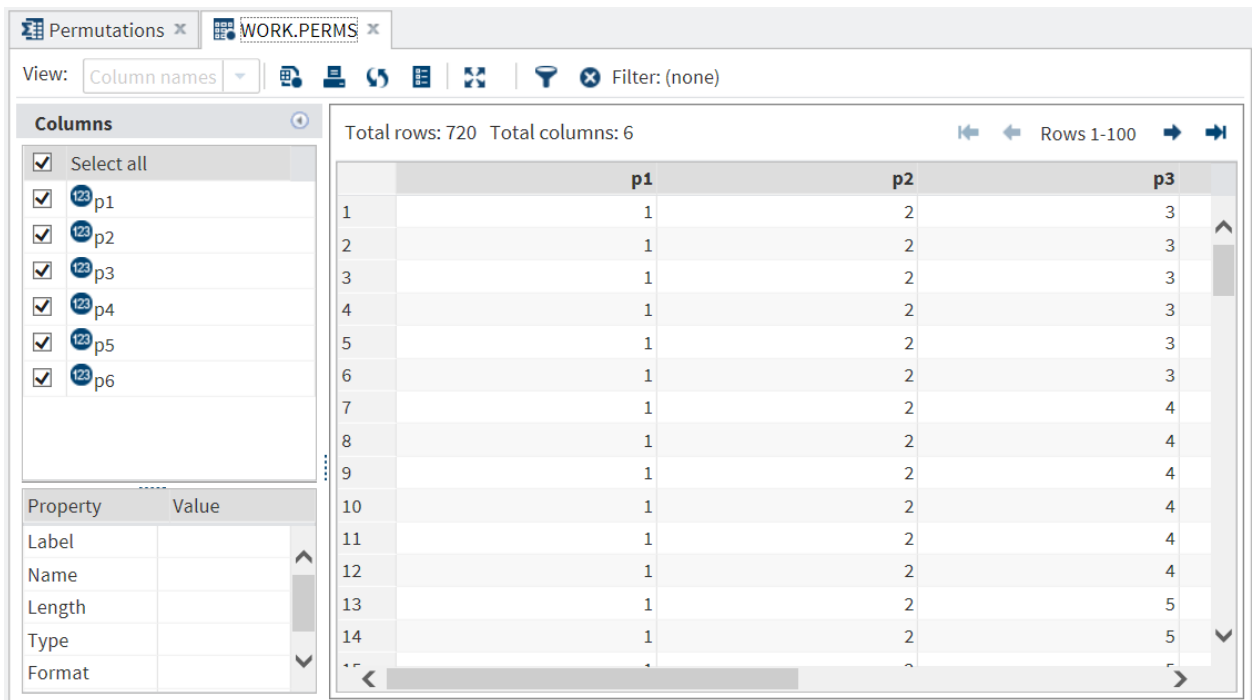
To create this example:

1   In the **Tasks** section, expand the **Combinatorics and Probability** folder and double-click **Permutations**. The user interface for the Permutations task opens.

2   On the **Options** tab, specify these options.

Assign columns to these roles:

Option Name	Value to Specify
Number of objects	6
Data set name	`Perms`

3 To run the task, click ⛷.

The Permutations task creates an output data set. In SAS Studio, this data set opens on the **WORK.Perms** tab.

Setting Options

All of these options are required to run the Permutations task.

Option Name	Description
Observations	
Number of objects	specifies the number of objects for which you want to compute permutations. This value can range from 1 to 10.
Output Data Set	
Data set name	specifies the name of the output data set.

Rank Data Task

About the Rank Data Task

The Rank Data task computes ranks for one or more numeric variables across the rows in a table and includes the ranks in an output table.

For example, you might want to rank the sales for each product that your company sells. In this case, the ranking variable would show the order of product sales. The product with the highest number of sales would be ranked first.

Example: Ranking Students by Age and Height

In this example, you want to rank the students in your class by age and height.

To create this example:

1 In the **Tasks** section, expand the **Data** folder and double-click **Rank Data**. The user interface for the Rank Data task opens.

2 On the **Data** tab, select the **SASHELP.CLASS** data set.

3 Assign columns to these roles:

Role	Column Name
Columns to rank	**Height**
Rank by	**Age**

4 To run the task, click ![run icon].

The Rank Data task creates an output data set. In SAS Studio, this data set opens on the **WORK.Rank** tab. This data set contains the additional rank_Height column, which shows where that student ranks within her age group. For example, in the 11-year-old age group, Joyce is ranked number one. In the 12-year-old age group, Louise is ranked number 1.

Program 1 ×   Rank Data 2 ×   WORK.Rank ×

View: Column names ▾ | 🖳 🖳 ↻ 🖽 | — | ▼ ✕ Filter: (none)

Columns ⊙

■ Select all
☑ Name
☑ Sex
☑ Age
☑ Height
☐ Weight
☑ rank_Height

Total rows: 19 Total columns: 6   Rows 1-19 ⬅ ➡

	Name	Sex	Age	Height	rank_Height
1	Joyce	F	11	51.3	1
2	Thomas	M	11	57.5	2
3	James	M	12	57.3	2
4	Jane	F	12	59.8	4
5	John	M	12	59	3
6	Louise	F	12	56.3	1
7	Robert	M	12	64.8	5
8	Alice	F	13	56.5	1
9	Barbara	F	13	65.3	3
10	Jeffrey	M	13	62.5	2
11	Alfred	M	14	69	4
12	Carol	F	14	62.8	1
13	Henry	M	14	63.5	2
14	Judy	F	14	64.3	3
15	Janet	F	15	62.5	1
16	Mary	F	15	66.5	2.5
17	Ronald	M	15	67	4
18	William	M	15	66.5	2.5
19	Philip	M	16	72	1

Property	Value
Label	Weight
Name	Weight
Length	8
Type	Numeric
Format	
Informat	

Assigning Data to Roles

To run the Rank Data task, you must assign a column to the **Columns to rank** role.

Role	Description
Columns to rank	Each column that is assigned to this role is ranked. You must assign at least one variable to this role. By default, the rankings column is given the name **rank_*column-name***, where *column-name* is the name of the original column.

Role	Description
Rank by	When you assign one or more columns to this role, the input table is sorted by the selected column or columns and rankings are calculated within each group.

Setting Options

You must select at least one output option.

Option Name	Description
Options	
Ranking method	specifies the method to use when ranking the data. Here are the valid values:
	None
	does not use a method to rank the data.
	Percentile ranks
	partitions the original values into 100 groups, in which the smallest values receive a percentile value of 0 and the largest values receive a percentile value of 99.
	Deciles
	partitions the original values into 10 groups, in which the smallest values receive a decile value of 0 and the largest values receive a decile value of 9.

Option Name	Description
Ranking method (continued)	**Quartiles** partitions the original values into four groups, in which the smallest values receive a quartile value of 0 and the largest values receive a quartile value of 3. **Group = n (NTILES)** partitions the original values into *n* groups, in which the smallest values receive a value of 0 and the largest values receive a value of *n*–1. Specify the value of *n* in the **Number of groups** box. **Fractional ranks with denominator = n** computes fractional ranks by dividing each rank by the number of observations that have nonmissing values of the ranking variable. **Fractional ranks with denominator = n+1** computes fractional ranks by dividing each rank by the denominator *n*+1, where *n* is the number of observations that have nonmissing values of the ranking variable. **Percents** divides each rank by the number of observations that have nonmissing values of the variable and multiplies the result by 100 to get a percentage.

Option Name	Description
Ranking method (continued)	**Normal scores (Blom formula)**, **Normal scores (Tukey formula)**, **Normal scores (van der Waerden formula)** computes normal scores from the ranks. The resulting variables appear normally distributed. Here are the formulas:

Blom formula

$$y_i = \Phi^{-1}\left(\frac{\left(r_i - \frac{3}{8}\right)}{\left(n + \frac{1}{4}\right)}\right)$$

Tukey formula

$$y_i = \Phi^{-1}\left(\frac{\left(r_i - \frac{1}{3}\right)}{\left(n + \frac{1}{3}\right)}\right)$$

van der Waerden

$$y_i = \Phi^{-1}\left(\frac{r_i}{(n + 1)}\right)$$

In these formulas, Φ^{-1} is the inverse cumulative normal (PROBIT) function, r_i is the rank of the *i*th observation, and *n* is the number of nonmissing observations for the ranking variable.

Note: If you set the **If values tie, use** option, the Rank Data task computes the normal score from the ranks based on non-tied values and applies the ties specification to the resulting score.

Savage scores (exponential)
computes Savage (or exponential) scores from the ranks.

Note: If you set the **If values tie, use** option, the Rank Data task computes the Savage score from the ranks based on non-tied values and applies the ties specification to the resulting score.

Option Name	Description
If values tie, use:	specifies how to compute normal scores or ranks for tied data values.
	Mean (Midrank) assigns the mean of the corresponding rank or normal scores
	High rank assigns the largest of the corresponding ranks or normal scores
	Low rank assigns the smallest of the corresponding ranks or normal scores
	Dense rank computes scores and ranks by treating tied values as a single-order statistic. For the default method, ranks are consecutive integers that begin with the number one and end with the number of unique, nonmissing values of the variable that is being ranked. Tied values are assigned the same rank.
Rank order	specifies whether to list the values from smallest to largest or from largest to smallest.
Results	
Location to save output data	specifies the location of the output table. By default, the table is saved in the temporary Work library.
Include ranked columns	specifies that the output table contains the original columns as well as the ranked columns. If you want to replace the original column with the ranked columns, deselect the **Include ranking columns** check box. By default, the ranked column is given the name `rank_`*column-name*, where *column-name* is the name of the original column.

Random Sample Task

About the Random Sample Task

The Random Sample task creates an output table that contains a random sample of the rows in the input table.

You might use this task when you need a subset of the data. For example, suppose you want to audit employee travel expenses in an effort to improve the expense reporting procedure and possibly reduce expenses. Because you do not have the resources to examine all expense reports, you can use statistical sampling to objectively select expense reports for audit.

Example: Creating a Random Sample of the Sashelp.Pricedata Data Set

In this example, you want to create a subset of the data in the Sashelp.Pricedata data set.

To create this example:

1   In the **Tasks** section, expand the **Data** folder and double-click **Random Sample**. The user interface for the Random Sample task opens.

2   On the **Data** tab, select the **SASHELP.PRICEDATA** data set.

3   To run the task, click .

Here are the tabular results:

The SURVEYSELECT Procedure

Selection Method	Simple Random Sampling

Input Data Set	PRICEDATA
Random Number Seed	496093001
Sample Size	10
Selection Probability	0.009804
Sampling Weight	102
Output Data Set	RANDOMSAMPLE

The task also creates a sample data set in the Work library. In SAS Studio, this data set opens on the **WORK.RandomSample** tab.

Assigning Data to Roles

For the Random Sample task, you must specify an input data source. No roles are required to run the task.

Role	Description
Output columns	specifies the variables to include in the output table. By default, all variables are included in the output table. However, you can select the variables to include in the output.

Role	Description
Strata columns	specifies the variables to use to partition the input table into mutually exclusive, nonoverlapping subsets that are known as strata. Each stratum is defined by a set of values of the strata variables, and each stratum is sampled separately. The complete sample is the union of the samples that are taken from all the strata.
	Note: If you do not assign any variables to this role, then the entire input table is treated as a single stratum.
	You can allocate the total sample size among the strata in proportion to the size of the stratum. For example, the variable GENDER has possible values of M and F, and the variable VOTED has possible values of Y and N. If you assign both GENDER and VOTED to the **Strata columns** role, then the input table is partitioned into four strata: males who voted, males who did not vote, females who voted, and females who did not vote.
	The input table contains 20,000 rows, and the values are distributed as follows:
	■ 7,000 males who voted
	■ 4,000 males who did not vote
	■ 5,000 females who voted
	■ 4,000 females who did not vote
	Therefore, the proportion of males who voted is 7,000/20,000=0.35 or 35%. The proportions in the sample should reflect the proportions of the strata in the input table. For example, if your sample table contains 100 observations, then 35% of the values in the sample must be selected from the males who voted stratum to reflect the proportions in the input table.

Setting Options

Option Name	Description
Sample size	specifies the sample size in the desired number of rows or in the desired percentage of input rows. For example, if you specify 3% of rows and there are 400 input rows, then the resulting sample has 12 rows.
	Note: If you assign variables to the **Strata columns** role, then the sample size specification that you make here applies to each stratum rather than to the entire input table.

Option Name	Description
Sample method	specifies the method to use when sampling the data. Here are the valid values:
	Simple (no duplicates) specifies the simple method when sampling the input data. When a row is selected, it is removed from eligibility for subsequent selections. This makes it impossible to select the same row more than once.
	Unrestricted (duplicates allowed) specifies the unrestricted method when sampling the input data. When a row is selected, it remains eligible for subsequent selections. This makes it possible to select the same row more than once. You can specify how multiple selections of the same row are recorded in the output table.
	You can choose from the following options:
	Show each observations once in output (exclude duplicates) a row that is selected *n* times occurs in the sample once. In the output, the NumberHits variable (which is calculated automatically by the Random Sample task) lists the number of times that the observation occurred in the input table.
	Show all observations in output (include duplicates) a row that is selected *n* times occurs in the sample *n* times.
Location of output data set	specifies the name and location for the output data. By default, the data is saved to the Work library.
Random seed number	specifies the initial seed for the generation of random numbers. If you do not specify a random seed number, then a seed that is based on the system clock will be used to produce the sample.

Option Name	Description
Generate a sample selection summary	generates a summary table that includes the seed that was used to produce the sample. By specifying this same seed later with the same input table, you can reproduce the same sample.

Sort Data Task

About the Sort Data Task

The Sort Data task enables you to sort the table by any of its columns. The result from this task is a sorted table in the Work library. No results or output data is displayed when you run this task.

Assigning Data to Roles

To run the Sort Data task, you must assign a column to the **Sort by** role.

Role	Description
Sort by	When you assign one or more variables to this role, the table is grouped by the selected variable or variables. The order in which the variables appear within this role determines which variable is the primary sort key, which variable is the secondary sort key, and so on. The primary sort key is always the first variable that is listed within the Sort by role.
Columns to drop	When you assign one or more variables to this role, the output that is generated does not contain the specified variables. You can assign a maximum of $(n - 1)$ variables to this role, where n is the total number of variables in the table.

Setting Options

Option Name	Description
Output Order	
Collating sequence	indicates what collating sequence to use when sorting character variables. You can use these collation standards: ■ sequence that is defined on the server (**Server default**) ■ the ASCII or EBCDIC collating sequences ■ the reverse collation order for character variables ■ a national standard, such as Danish, Finnish, Italian, Norwegian, Spanish, or Swedish ■ a custom-defined collating sequence that is defined by your installation site
Maintain original data order within 'Sort by' groupings	groups the data according to the order that you set for the **Sort by** role. If this option is not selected, then the output table is grouped in an undefined order within the sorted key groups.
Duplicate Records	
Keep all records	keeps all of the records that are in the output table, including all duplicates of records.

Option Name	Description
Keep only the first record for each 'Sort by' group	eliminates any duplicate observations that have the same values for the **Sort by** group. If the **Group data in the order of the Sort by variable** option is selected, then the observation that is retained for each **Sort by** group is the first one that is read from the original table. However, if the **Group data in the order of the Sort by variable** option is not selected, then the observation that is kept for each **Sort by** group cannot be predetermined.
Do not keep adjacent duplicate records	compares each record to the previous record in the output table. If an exact match is found, the duplicate record is not written to the output table. **Note:** If you do not assign all variables to the **Sort by** role, some duplicate records might not be removed because the records are not adjacent.
Advanced Sorting	
Memory for sorting	specifies the maximum amount of memory that can be used for the Sort Data task. You can specify the amount of memory in bytes (B), kilobytes (KB), megabytes (MB), or gigabytes (GB). You can also specify to use all of the available memory or to use the default amount of memory that has been allocated on the server.
Reduce temporary disk space requirements	indicates that during the Sort Data process, only the **Sort by** variables and the observation numbers are stored within temporary files, reducing the amount of storage necessary to perform the sort. In the final phase of the sort, the temporary file is used as an index to access the original table and then to send the data to the results table in the correctly sorted sequence.

Option Name	Description
Force a sort of indexed data	indicates that you want to sort all tables even if the table is already sorted in the desired sequence or the table contains a user-created index with keys that reflect those specified in the **Sort by** role. If you specify this option, the table is sorted regardless of the current order of the table or whether it contains an index.
Results	
Location to save output data	specifies the location for the output table. By default, this table is saved to the temporary Work library.

Table Attributes Task

About the Table Attributes Task

The Table Attributes task enables you to create these types of reports:

- a default report that includes the following data attributes: the date on which the table was created and last modified, the number of rows, the encoding, any engine-dependent or host-dependent information, and an alphabetic list of the variables and their attributes.

- an enhanced report displays the table and variable attributes. Unlike the default report, you can specify the order of the contents in the report. From this report, you can determine the table type, the date on which the table was created and modified, the number of observations, the variable labels, and the variable types.

Example: Table Attributes for the Sashelp.Pricedata Data Set

In this example, you want to view the table attributes for the Sashelp.Pricedata data set.

To create this example:

1 In the **Tasks** section, expand the **Data** folder and double-click **Table Attributes**. The user interface for the Table Attributes task opens.

2 On the **Data** tab, select the **SASHELP.PRICEDATA** data set.

3 On the **Options** tab, deselect the **Default report** check box.

4 To run the task, click 🏃.

Here is a subset of the results for the Table Attributes task. These results are the enhanced report for the Sashelp.Pricedata data set.

Table Name	Label	Type	Date Created	Date Modified	Number of Obs.	Char. Set	Password Protected	Data Set Type
PRICEDATA	Simulated monthly sales data with hierarchy of region, line, product	DATA	20JUN13:00:29:56	20JUN13:00:29:56	1020		---	

Name	Variable Number	Type	Format	Label	Length
cost	5	Numeric		Unit Cost	8
date	1	Numeric	MONYY	Order Date	8
discount	4	Numeric		Price Discount	8
line	27	Numeric		Product Line ID	8
price	3	Numeric		Unit Price	8
price1	6	Numeric		Product 1 Unit Price	8
price10	15	Numeric		Product 10 Unit Price	8
price11	16	Numeric		Product 11 Unit Price	8
price12	17	Numeric		Product 12 Unit Price	8
price13	18	Numeric		Product 13 Unit Price	8
price14	19	Numeric		Product 14 Unit Price	8
price15	20	Numeric		Product 15 Unit Price	8

An output data set also opens on the **WORK.TableAttributes** tab.

Setting Options

Option Name	Description
Default report	contains the output from the DATASETS procedure. This report includes the following data attributes: the date on which the table was created and last modified, the number of observations, the encoding, any engine-dependent or host-dependent information, and an alphabetic list of the variables and their attributes.

Option Name	Description
Enhanced report	contains the output from the DATASETS procedure. The REPORT procedure is used to create the enhanced report.
	This report displays the table and variable attributes. From this report, you can determine the table type, the date on which the table was created and modified, the number of observations, the variable labels, and the variable types.
Sort variables by	sorts the rows in the variable table by variable name, variable order in the table, variable type, variable format, or variable label.
	Note: This option affects only the enhanced report.
Order sequence	specifies whether to sort the rows in the table by ascending or descending order.
	Note: This option affects only the enhanced report.
Location to save output data	specifies the location of the output table. By default, this table is saved to the temporary Work library.

Transpose Data Task

About the Transpose Data Task

The Transpose Data task turns selected columns of an input table into the rows of an output table. If you do not use grouping variables, then each selected column is turned into a single row. If you use grouping variables, then the selected columns are divided into subcolumns based on the values of the grouping variables. Each subcolumn is turned into a row of the output table.

Assigning Data to Roles

To run the Transpose Data task, you must assign a column to the **Transpose variables** role.

Role	Description
Transpose variables	Each variable that you assign to this role becomes one or more rows of the output table. If you do not select any grouping variables, then an entire column is turned into a single row. If you select one or more grouping variables, then the grouping variables are used to segment each column into subcolumns, each of which is turned into a row. In this case, a column is transposed to the number of rows that is equal to the number of groups that are defined by the grouping variables.
	You must assign at least one column to the **Transpose variables** role. To select a grouping variable, assign a column to the **Group analysis by** role.
Copy variables	Each variable that you assign to this role is copied directly from the input table to the output table without being transposed. Because these columns are copied directly to the output table, the number of rows in the output table equals the number of rows in the input table. The output table is padded with missing values if the number of rows in the input table does not equal the number of variables that it transposes.
Group analysis by	Each variable that you assign to this role is used to segment the about-to-be-transposed columns into subcolumns that will be transposed separately. Each subcolumn, defined by a set of values of the grouping variables, becomes a row of the output table.

Role	Description
New column names	The variable that you assign to this role is used to name the transposed variables in the output data set. When the variable name is created for the transposed data set, all the nonmissing, formatted values are concatenated in the same order that the variables are assigned to this role. If you select the **Allow duplicate items** check box, the transposed output data set contains only the last observation for each BY group.
New column labels	The values of the variable that you assign to this role are used to label the variables in the **New column names** role.

Setting Options

Option Name	Description
Source Column	
Name	Each row of the output table includes the name of the variable in the input table to which the values in that output row belong. To specify a heading for the output column that contains these variable names, enter the heading in the **Name** box. The name can include special characters, leading numbers, and white space, but it cannot exceed 32 characters. The default name is Source.

Option Name	Description
Label	Each row of the output table includes the label of the variable in the input table to which the values in that output row belong. To specify a heading for the output column that contains these variable labels, enter the heading in the **Label** box. The label can include special characters, leading numbers, and white space, but it cannot exceed 32 characters. The default label is Label.
Column Name Prefix	
Use prefix	You can specify a prefix to use in constructing the names for the transposed variables in the output data set. When you use a prefix, the variable name begins with the prefix value and is followed by the value of the variable that you assigned to the **New column names** role.
Results	
Name of output table	You can designate a different name for the output table.

9

Econometrics Tasks

Count Data Regression Task

About the Count Data Regression Task

The Count Data Regression task performs a regression analysis of a continuous dependent variable. This variable takes a nonnegative integer value from distributions such as Poisson, negative binomial, zero-inflated Poisson, and zero-inflated negative binomial.

Note: The version of the task depends on what version of SAS/ETS is available at your site. For example, if your site is running SAS 9.4 or earlier, SAS Studio is running version 1 of the Count Data Regression task. If your site is running the first maintenance release for SAS 9.4 or later and SAS/ETS 13.1 or later is available, then SAS Studio is running version 2 of the Count Data Regression task. The difference between the two versions is the addition of new options in SAS/ETS 13.1 or later.

Example: Count Data Regression

To create this example:

1   Create the WORK.LONG97DATA data set. For more information, see "LONG97DATA Data Set" on page 412.

2   In the **Tasks** section, expand the **Econometrics** folder and double-click **Count Data Regression**. The user interface for the Count Data Regression task opens.

3   On the **Data** tab, select the **WORK.LONG97DATA** data set.

4   Assign columns to these roles:

Role	Column Name
Dependent variable	**art**
Continuous variables	**ment** **phd** **mar**
Categorical variables	**kid5**

5 To run the task, click [icon].

Here is a subset of the results:

Class Level Information		
Class	Levels	Values
kid5	4	0 1 2 3

Model Fit Summary	
Dependent Variable	art
Number of Observations	915
Data Set	WEBWORK.LONG97DATA
Model	Poisson
Log Likelihood	-1659
Maximum Absolute Gradient	4.52499E-9
Number of Iterations	5
Optimization Method	Newton-Raphson
AIC	3332
SBC	3366

Algorithm converged.

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	-0.570945	0.296411	-1.93	0.0541
ment	1	0.026211	0.002014	13.01	<.0001
phd	1	0.015683	0.026428	0.59	0.5529
mar	1	0.179880	0.062563	2.88	0.0040
kid5 0	1	0.726607	0.280831	2.59	0.0097
kid5 1	1	0.594124	0.283226	2.10	0.0359
kid5 2	1	0.451952	0.288913	1.56	0.1177
kid5 3	0	0	.	.	.

Assigning Data to Roles

To run the Count Data Regression task, you must assign a column to the **Dependent variable** role.

Role	Description
Dependent variable	specifies the numeric column that has nonnegative integer or count values.
	The **Distribution** option specifies the type of model to be analyzed. You can specify these types of models:
	■ Poisson regression model
	■ negative binomial regression model with a linear variance function
	■ negative binomial regression model with a quadratic variance function
	■ a zero-inflated Poisson model
	■ a zero-inflated negative binomial model
	If you select a zero-inflated distribution, you can specify the link function to use to compute the probability of zeros.
Continuous variables	specifies the independent covariates (regressors) for the regression model. If you do not specify a continuous variable, the task fits a model that contains only an intercept.
Categorical variables	specifies the variables to use to group data in the analysis.

Setting Options

Option	Description
Methods	

Option	Description
Type of covariances of the parameter estimates	specifies the type of covariance matrix of the parameter estimates.
	You can specify these types of matrices:
	■ the covariance from the inverse Hessian matrix
	■ the covariance from the outer product mix
	■ the covariance from the outer product and Hessian matrices (also called the quasi-maximum-likelihood-estimates)
Include the intercept in the model	specifies whether to include the intercept in the model.
Optimization	
Method	specifies the iterative minimization method to use.
Maximum number of iterations	specifies the maximum number of iterations for the selected method.

Statistics

You can specify whether to include in the results the default statistics that the task creates, the default statistics and any additional statistics that you select, or no statistics.

Here are the additional statistics that you can include in the results:

■ correlation matrix of the parameter estimates

■ covariance matrix of the parameter estimates

■ iteration history of the objective function and parameter estimates

Plots

Note: The plot options are available only if you are running the first maintenance release for SAS 9.4 or later.

Select plots to display	specifies whether to include in the results only the default plots, only the plots that you choose, or no plots.

Diagnostic Plots

Option	Description
Profile likelihood plot	produces the profile likelihood functions of the model parameters. The model parameter on the X axis is varied, whereas all other parameters are fixed at their estimated maximum likelihood estimates.
Overdispersion diagnostic plot	produces the overdispersion diagnostic plot.
Probability Plots	
Overall predictive probabilities plot	produces the overall predictive probabilities of the specified count levels.
Predictive probability profiles plot	produces the predictive probability profiles of specified count levels against model regressors. The regressor on the X axis is varied, whereas all other regressors are fixed at the mean of the observed data set.
Specified count levels	supplies the values of the response variable for the overall predictive probabilities plot and the predictive probability profiles plot. Each value should be a nonnegative integer. Nonintegers are rounded to the nearest integer. This value can also be a list in the form X TO Y BY Z. For example, COUNTS(0 1 2 TO 10 BY 2 15) creates a plot for counts 0, 1, 2, 4, 6, 8, 10, and 15.
Zero-Inflation Plots **Note:** The **Probability profiles plot of zero-inflation process selection** option is available only if you select **Zero-inflated Poisson** or **Zero-inflated negative binomial** as the distribution for the dependent variable.	
Probability profiles plot of zero-inflation process selection	produces the probability profiles of zero-inflation process selection and zero count prediction against model regressors. The regressor on the X axis is varied, whereas all other regressors are fixed at the mean of the observed data set.

Option	Description
Display as	specifies whether to display the plots in a panel or individually. This option applies to all distributions.

Heckman Selection Model Task

About the Heckman Selection Model Task

The Heckman two-step selection method provides a means of correcting for non-randomly selected samples. It is a two-stage estimation method. The first stage performs a probit analysis on a selection equation. The second stage analyzes an outcome equation based on the first-stage binary probit model.

Note: This task is available only if you are running SAS 9.4 (or later) and SAS/ETS 12.3 (or later).

Example: Heckman Selection Model Task

To create this example:

1   Create the Work.Mroz data set. For more information, see .

2   In the **Tasks** section, expand the **Econometrics** folder and double-click **Heckman Selection Model**. The user interface for the Heckman Selection Model task opens.

3   On the **Data** tab, select the **WORK.MROZ** data set.

4   Assign columns to these roles:

Role	Column Name
Selection Equation	
Dependent variable	**inlf**
Continuous variables	**nwifeinc**
	exper
	expersq
	age
	kidslt6
	kidsge6
Outcome Equation	
Dependent variable	**lwage**
Continuous variables	**exper**
	expersq
Categorical variables	**educ**

5 To run the task, click ![run icon].

Here is a subset of the results:

Summary Statistics of Continuous Responses								
Variable	N	Mean	Standard Error	Type	Lower Bound	Upper Bound	N Obs Lower Bound	N Obs Upper Bound
lwage	428	1.190173	0.723198	Regular				

Discrete Response Profile of inlf		
Index	Value	Total Frequency
1	0	325
2	1	428

Class Level Information		
Class	Levels	Values
educ	13	5 6 7 8 9 10 11 12 13 14 15 16 17

Heckman First Step Model Fit Summary	
Number of Endogenous Variables	1
Endogenous Variable	inlf
Number of Observations	753
Log Likelihood	-415.37436
Maximum Absolute Gradient	0.0001730
Number of Iterations	13
Optimization Method	Quasi-Newton
AIC	844.74872
Schwarz Criterion	877.11718

Goodness-of-Fit Measures		
Measure	Value	Formula
Likelihood Ratio (R)	199	2 * (LogL - LogL0)
Upper Bound of R (U)	1029.7	- 2 * LogL0
Aldrich-Nelson	0.209	R / (R+N)
Cragg-Uhler 1	0.2322	1 - exp(-R/N)
Cragg-Uhler 2	0.3116	(1-exp(-R/N)) / (1-exp(-U/N))

Assigning Data to Roles

To run the Heckman Selection Model task, you must assign columns to the **Dependent variable** roles for the selection and outcome equations.

Role	Column Name
Selection Equation	

Role	Column Name
Dependent variable	specifies a single numeric column that takes binary values. By default, the task uses samples where the dependent variable is equal to 1.
Continuous variables	specifies the independent columns (or regressors) to use in the model for the selection equation dependent variable.
Categorical variables	specifies how to group the values into levels.
Include the intercept	specifies whether to include the intercept in the selection equation.
Outcome Equation	
Dependent variable	specifies a single numeric column to use.
Continuous variables	specifies the independent columns (or regressors) to use in the model for the outcome equation dependent variable.
Categorical values	specifies how to group the values into levels.
Include the intercept	specifies whether to include the intercept in the selection equation.

Setting Options

Option	Description
Methods	
Variance estimation method	specifies whether to calculate the standard errors by using the corrected standard errors or the OLS standard errors.

Option	Description
Type of covariances of the parameter estimates	specifies the method to calculate the covariance matrix of parameter estimates. You can select the covariance from the outer product matrix, from the inverse Hessian matrix, or from the output product and Hessian matrices (the quasi-maximum likelihood estimates).
Optimization	
Method	specifies the iterative minimization method to use. By default, the Quasi-Newton method is used.
Maximum number of iterations	specifies the maximum number of iterations for the selected method.
Statistics	

You can specify whether the results include the statistics that the task creates by default, the default statistics and any additional statistics that you select, or no statistics.

Here is the information that you can include in the results:

- correlation matrix of the parameter estimates
- covariance matrix of the parameter estimates
- iteration history of the objective function and parameter estimates

Panel Data: Count Data Regression Task

About the Panel Data: Count Data Regression Task

The Panel Data: Count Data Regression task performs count data regression of a continuous dependent variable. This variable is a nonnegative integer value from Poisson or negative binomial distributed panel data.

Note: This task is available only if you are running the first maintenance release for SAS 9.4 (or later) and SAS/ETS 13.1 (or later).

Example: Count Data Regression with Panel Data

To create this example:

1   Create the WORK.LONG97DATA data set. For more information, see "LONG97DATA Data Set" on page 412.

2   In the **Tasks** section, expand the **Econometrics** folder and double-click **Panel Data: Count Data Regression**. The user interface for the Panel Data: Count Data Regression task opens.

3   On the **Data** tab, select the **WORK.LONG97DATA** data set.

4   Assign columns to these roles:

Role	Column Name
Dependent variable	**art**

Role	Column Name
Continuous variables	**ment**
	phd
	mar
Categorical variables	**kid5**
Cross-sectional ID	**fem**

5   To run the task, click ⨉.

Here is a subset of the results:

Class Level Information		
Class	**Levels**	**Values**
kid5	4	0 1 2 3

Model Fit Summary	
Dependent Variable	art
Number of Observations	915
Data Set	WORK.COUNTPANELREGDATA
Model	Poisson
Error Component	Random
Optimization Method	Newton-Raphson
Log Likelihood	-1654
Maximum Absolute Gradient	1.87227E-6
Number of Iterations	13
AIC	3323
SBC	3362

Algorithm converged.

Parameter Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**t Value**	**Approx Pr > |t|**
Intercept	1	-0.615254	0.305887	-2.01	0.0443
ment	1	0.025452	0.002024	12.57	<.0001
phd	1	0.013129	0.026437	0.50	0.6195

Assigning Data to Roles

To run the Count Panel Data Regression task, you must assign columns to the **Dependent variable** and **Cross-sectional ID** roles.

Role	Description
Dependent variable	specifies the numeric column that has nonnegative integer or count values. The **Distribution** option specifies the type of model to be analyzed. You can specify these types of models: ■ Poisson regression model ■ negative binomial regression model with a linear variance function ■ negative binomial regression model with a quadratic variance function
Continuous variables	specifies the independent covariates (regressors) for the regression model. If you do not specify a continuous variable, the task fits a model that contains only an intercept.
Categorical variables	specifies the variables to use to group data in the analysis.
Cross-sectional ID	specifies the cross-section for each observation. You can specify whether the error component model is fixed or random.

Setting Options

Option	Description
Methods	

Option	Description
Type of covariances of the parameter estimates	specifies the type of covariance matrix of the parameter estimates.
	You can specify these types of matrices:
	■ the covariance from the inverse Hessian matrix
	■ the covariance from the outer product matrix
	■ the covariance from the outer product and Hessian matrices (also called the quasi-maximum-likelihood-estimates)
Include the intercept in the model	specifies whether to include the intercept in the model.

Optimization	
Method	specifies the iterative minimization method to use.
Maximum number of iterations	specifies the maximum number of iterations for the selected method.

Statistics

You can specify whether to include the statistics that the task creates by default and any additional output tables in the results.

Here are the additional statistics that you can include in the results:

■ correlation matrix of the parameter estimates

■ covariance matrix of the parameter estimates

■ iteration history of the objective function and parameter estimates

Panel Data: Linear Regression

About the Panel Data: Linear Regression Task

The Panel Data: Linear Regression task analyzes a class of linear econometric models that commonly arise when time series and cross-sectional data are combined. This type of pooled data on time series cross-sectional bases is often referred to as panel data. Typical examples of panel data include observations over time on households, countries, firms, trade, and so on. For example, in the case of survey data on household income, the panel is created by repeatedly surveying the same households in different time periods (years).

Note: The version of the task depends on what version of SAS/ETS is available at your site. For example, if your site is running the second maintenance release for SAS 9.3, SAS/ETS 12.1 is available, and SAS Studio is running version 1 of the Panel Data: Linear Regression task. If your site is running SAS 9.4 or later and SAS/ETS 12.3 or later is available, then SAS Studio is running version 2 of the Panel Data: Linear Regression task. The difference between the two versions is the addition of new options in SAS/ETS 12.3 or later.

Example: Linear Regression with Panel Data

To create this example:

1   Create the WORK.GREENE data set. For more information, see "GREENE Data Set" on page 411.

2   In the **Tasks** section, expand the **Econometrics** folder and double-click **Panel Data: Linear Regression**. The user interface for the Panel Data: Linear Regression task opens.

3   On the **Data** tab, select the **WORK.GREENE** data set.

4 Assign columns to these roles:

Role	Column Name
Dependent variable	**cost**
Continuous variables	**production**
Cross-sectional ID	**firm**
Time series ID	**year**

5 To run the task, click ![run icon].

Wansbeek and Kapteyn Variance Components (RanOne)
Dependent Variable: cost

Model Description	
Estimation Method	RanOne
Number of Cross Sections	3
Time Series Length	4
HAC Kernel	Bartlett
HAC Bandwidth	Newey and West

Fit Statistics			
SSE	0.1541	DFE	8
MSE	0.0193	Root MSE	0.1388
R-Square	0.8774		

Variance Component Estimates	
Variance Component for Cross Sections	0.058961
Variance Component for Error	0.018214

Hausman Test for Random Effects		
DF	m Value	Pr > m
1	2.82	0.0930

Parameter Estimates					
Variable	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-2.16577	0.6317	-3.43	0.0090

Assigning Data to Roles

To run the Panel Data: Linear Regression task, you must assign columns to the **Dependent variable**, **Cross-sectional ID**, and **Time series ID** roles.

Role	Description
Dependent variable	specifies the numeric column that contains the count values. The dependent count variable should take on only nonnegative integer values in the input data set.
Continuous variables	specifies the independent covariates (regressors) for the regression model. If you do not specify a continuous variable, the task fits a model that contains only an intercept.
Categorical variables	specifies the variables to use to group data in the analysis.
Cross-sectional ID	specifies the cross section for each observation. The task verifies that the input data is sorted by the cross-sectional ID and by the time series ID within each cross section.
Time series ID	specifies the time period for each observation. The task verifies that the time series ID values are the same for all cross sections.

Setting the Model Options

Option	Description
Model	

Option	Description
Model type	specifies whether to create a random effects model or a fixed effects model.
	Note: The remaining options that are available in this section depend on whether you are creating a random or fixed effect.
Include the intercept in the model	specifies whether to include the model. This option applies whether you are creating a random effects model or a fixed effects model.
	Note: This option is available only if you are running on SAS 9.4 or later.
Random Effects	
Random effects	specifies whether a one-way or two-way random-effects model is estimated. By default, a one-way random-effects model is estimated.
Variance component estimation method	specifies the type of variance component estimate to use. For more information about the type of estimations, see the PANEL procedure in *SAS/ETS User's Guide*.
Test for Random Effects	
One-way Breusch-Pagan test **Two-way Breusch-Pagan test**	requests the Breusch-Pagan one-way or two-way test for random effects.
Fixed Effects	
Fixed effects	specifies whether a one-way or two-way fixed-effects model is estimated.
Display the fixed effects	specifies whether to include the fixed effects in the results.
	Note: This option is available only if you are running on SAS 9.4 or later.

Setting Options

Option	Description
Methods	
Covariance matrix estimator	specifies the estimator of the covariance matrix. You can select from these options:
	■ **Newey and West**
	Note: This option is available only if you are running on SAS 9.4 or later.
	■ **OLS estimator** specifies that the variance-covariance matrix is not corrected.
	■ **HCCME***n* specifies a heteroscedasticity-corrected covariance matrix. *n* is a value from 0–4.
Cluster correction for heteroscedasticity-consistent covariance matrix	specifies the cluster correction for the variance-covariance matrix.
	Note: This option is available only if you select one of the **HCCME0-3** options for the covariance matrix estimator.
Statistics	

You can specify whether the results include the statistics that the task creates by default, the default statistics and any additional statistics that you select, or no statistics.

Here is the information that you can include in the results:

■ correlation matrix of the parameter estimates

■ covariance matrix of the parameter estimates

■ iteration history of the objective function and parameter estimates

Plots	
Diagnostic Plots	

Option	Description
You can display these types of diagnostic plots: ▪ Plot of the predicted and actual values ▪ QQ plot of residuals ▪ Plot of residuals by observation ▪ Histogram of residuals	
Cross Sections Plots	
You can display these types of plots: ▪ Plot of actual values by time series ▪ Predicted values by time series ▪ Stacked residuals by time series ▪ Residuals by time series	
Number of cross sections in one time series plot	specifies the number of cross sections to be combined into one time series plot. **Note:** This option is available only if you select **Individual plots** for the **Display as** option.
Display as	specifies whether to display the plots in a panel or individually.

Binary Probit/Logit Regression Task

About the Binary Probit/Logit Regression Task

The Binary Probit/Logit Regression task performs a regression analysis of a binary dependent variable from normal or logistic distributed panel data.

Note: The version of the task depends on what version of SAS/ETS is available at your site. For example, if your site is running the second maintenance release for SAS 9.3,

SAS/ETS 12.1 is available, and SAS Studio is running version 1 of the Binary Probit/ Logit Regression task. If your site is running SAS 9.4, SAS/ETS 12.3 or later is available, and SAS Studio is running version 2 of the Binary Probit/Logit Regression task. The difference between the two versions is the addition of new options in SAS/ETS 12.3 or later.

Example: Binary Probit/Logit Regression Task

To create this example:

1   Create the Work.Mroz data set. For more information, see .

2   In the **Tasks** section, expand the **Econometrics** folder and double-click **Binary Probit/Logit Regression**. The user interface for the Binary Probit/Logit Regression task opens.

3   On the **Data** tab, select the **WORK.MROZ** data set.

4   Assign columns to these roles:

Role	Column Name
Dependent variable	**inlf**
Continuous variables	**nwifeinc**
	exper
	expersq
	age
	kidslt6
	kidsge6
Categorical variables	**educ**

5   To run the task, click ⚹.

Here is a subset of the results:

Discrete Response Profile of inlf		
Index	Value	Total Frequency
1	0	325
2	1	428

Class Level Information		
Class	Levels	Values
educ	13	5 6 7 8 9 10 11 12 13 14 15 16 17

Model Fit Summary	
Number of Endogenous Variables	1
Endogenous Variable	inlf
Number of Observations	753
Log Likelihood	-396.16371
Maximum Absolute Gradient	1.27229
Number of Iterations	157
Optimization Method	Quasi-Newton
AIC	830.32742
Schwarz Criterion	918.18466

Goodness-of-Fit Measures		
Measure	Value	Formula
Likelihood Ratio (R)	237.42	2 * (LogL - LogL0)
Upper Bound of R (U)	1029.7	- 2 * LogL0
Aldrich-Nelson	0.2397	R / (R+N)
Cragg-Uhler 1	0.2704	1 - exp(-R/N)
Cragg-Uhler 2	0.3629	(1-exp(-R/N)) / (1-exp(-U/N))
Estrella	0.3012	1 - (1-R/U)^(U/N)
Adjusted Estrella	0.255	1 - ((LogL-K)/LogL0)^(-2/N*LogL0)
McFadden's LRI	0.2306	R / U
Veall-Zimmermann	0.415	(R * (U+N)) / (U * (R+N))

Assigning Data to Roles

To run the Binary Probit/Logit Regression task, you must assign a column to the **Dependent variable** role.

Role	Description
Dependent variable	specifies the numeric column to use as the dependent variable for the regression analysis. Use the **Distribution** drop-down list to specify whether to create a normal or logistic model.
Continuous variables	specifies the numeric columns to use as the independent regressor (explanatory) variables for the regression model.
Categorical variables	specifies how to group values into levels.

Setting Options

Option	Description
Methods	
Type of covariances of the parameter estimates	specifies the type of covariance matrix of the parameter estimates. You can specify these types of matrices: ■ the covariance from the inverse Hessian matrix ■ the covariance from the outer product matrix ■ the covariance from the outer product and Hessian matrices (also called the quasi-maximum-likelihood-estimates)

Option	Description
Include the intercept in the model	specifies whether to include the intercept in the model.
Heteroscedasticity	
Analyze heteroscedasticity	displays the heteroscedasticity options.
Variables on the variance function	specifies the columns that are related to heteroscedasticity of the residuals and how these variables are used to model error variances. Here is the heteroscedastic regression model that is supported by this task: $y_i = x_i'\beta + \varepsilon_i$ $\varepsilon_i \sim N(0, \sigma_i^2)$
Form of variance function	specifies the link function to use. You can choose from these options: ▪ **Exponential** $\sigma_i^2 = \sigma^2(1 + \exp(z_i'\gamma))$ ▫ **Exponential with no constant** $\sigma_i^2 = \sigma^2 \exp(z_i'\gamma)$ ▪ **Linear** $\sigma_i^2 = \sigma^2(1 + z_i'\gamma)$ ▪ **Square of linear function** $\sigma_i^2 = \sigma^2\left(1 + \left(z_i'\gamma\right)^2\right)$
Optimization	
Method	specifies the iterative minimization method to use. By default, the Quasi-Newton method is used.
Maximum number of iterations	specifies the maximum number of iterations for the selected method.
Statistics	

Option	Description

You can specify whether to include any statistics in the results.

Here is the information that you can choose to include in the results:

- correlation matrix of the parameter estimates
- covariance matrix of the parameter estimates
- iteration history of the objective function and parameter estimates

Plots

Select plots to display	specifies whether to display the default plots created by the task, only the plots that you select, or no plots.

Diagnostic Plots

Error standard deviations by observed regressor	displays the error standard deviation versus observed regressors when you assign a column to the **Variables on the variance function** option.
Profiled log likelihood	displays the profiled log likelihood. Each profiled graph is obtained by setting all the parameters to their maximum likelihood estimate except for the profiling parameter. The profiling parameter takes values on a predefined grid that is determined by the maximum likelihood estimate of the corresponding standard deviation.

Output Plots

Predicted values by regressor	displays the model predicted values. Each contributing regressor is set equal to its mean, except for the parameter that is reported on the X axis.
Marginal effects by regressor	displays the marginal effects. Each contributing regressor is set equal to its mean, except for the parameter that is reported on the X axis.

Option	Description
Inverse Mills ratio by regressor	displays the inverse Mills ratio. Each contributing regressor is set equal to its mean, except for the parameter that is reported on the X axis.
Predicted response probability by regressor	displays the predicted response probability. Each contributing regressor is set equal to its mean, except for the parameter that is reported on the X axis.
Predicted probabilities for each level of the response by regressor	displays the predicted probabilities for each level of the response. Each contributing regressor is set equal to its mean, except for the parameter that is reported on the X axis.
Linear predictor values by regressor	displays the structural part on the right side of the model. Each contributing regressor is set equal to its mean, except for the parameter that is reported on the X axis.
Display as	specifies whether to display the plots in a panel or individually.

10

Graph Tasks

Bar Chart Task

About the Bar Chart Task

The Bar Chart task creates horizontal or vertical bar charts that compare numeric values or statistics between different values of a chart variable. Bar charts show the relative magnitude of data by displaying bars of varying height. Each bar represents a category of data.

Example: Bar Chart of Mean Sales for Each Product Line

For example, you can create a bar chart that compares the total amount of sales for each product line in the Sashelp.Pricedata data set. By default, the task calculates the mean of the response variable for each product line. This bar chart shows that Line 2 has the highest mean product sales.

To create this example:

1   In the **Tasks** section, expand the **Graph** folder and double-click **Bar Chart**. The user interface for the Bar Chart task opens.

2   On the **Data** tab, select the **SASHELP.PRICEDATA** data set.

3   Assign columns to these roles:

Role	Column Name
Category variable	**productLine**
Response variable	**sale**

4   To run the task, click .
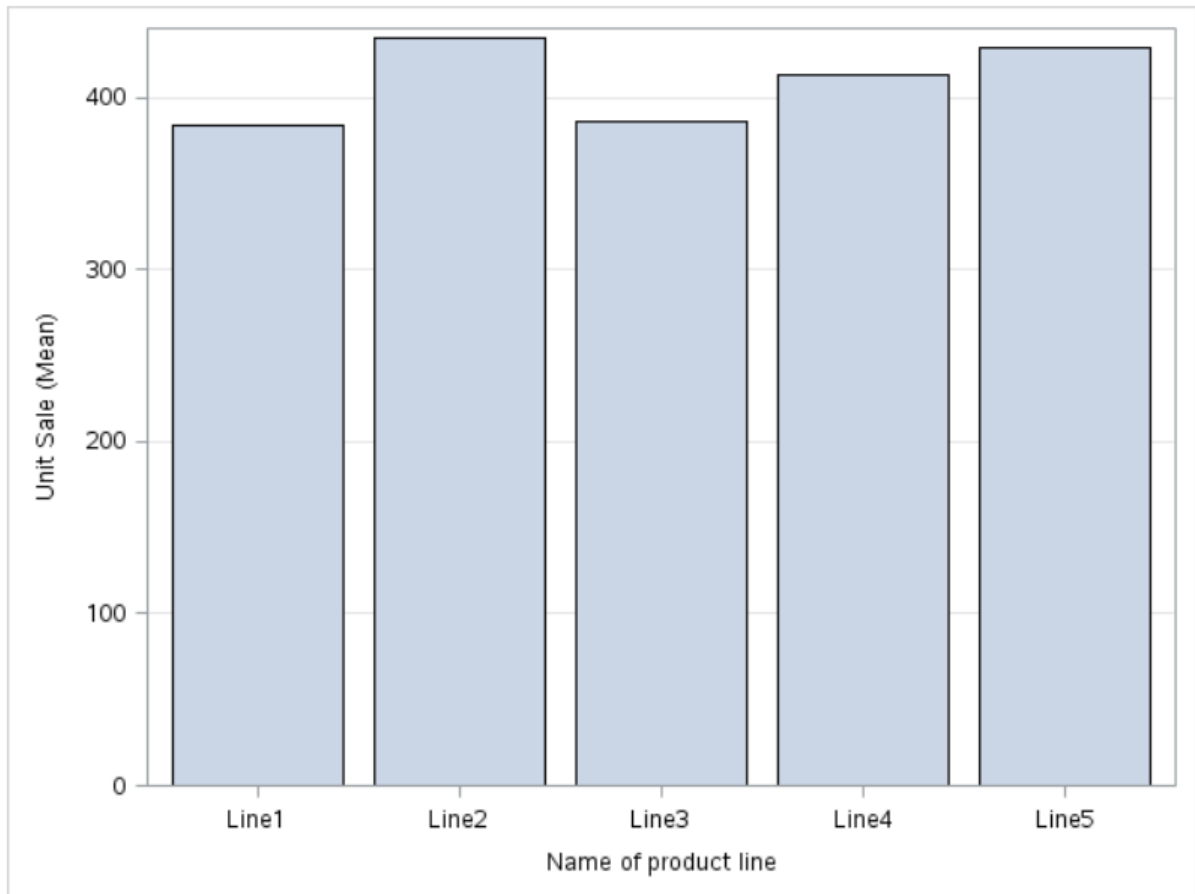
Here are the results:

Assigning Data to Roles

You can subset the data in the input data set by applying an SQL WHERE clause. In the **Where string** box, enter your valid SQL syntax. You must enclose character strings in single quotation marks. To view this WHERE clause with the resulting graph, select the **Include as footnote** check box.

To run the Bar Chart task, you must assign a column to the **Category variable** role.

Option Name	Description
Roles	
Category variable	specifies the variable that classifies the observations into distinct subsets.
Response variable	specifies a numeric response variable for the plot.
Group variable	specifies a variable that is used to group the data.
URL variable	specifies a character variable that contains URLs for web pages to be displayed when parts of the plot are selected within an HTML page.
BY variable	creates a separate graph for each BY group.
Direction	
You can create either a vertical or horizontal bar chart.	
Group Layout	
Cluster	displays group values as separate adjacent bars that replace the single category bar. Each set of group values is centered at the midpoint tick mark for the category.
Stack	overlays group values without any clustering. Each group is represented by unique visual attributes derived from the GraphData1... GraphData*n* style elements in the current style.

Option Name	Description
Statistics **Note:** The **Statistics** options are not available in these cases: ■ You did not assign a column to the **Response variable** role. If you do not assign a response variable, the default statistic is frequency. ■ You selected **Stack** for the group layout. In this case, the default statistic is sum.	
Mean	calculates the mean of the response variable.
Sum	calculates the sum of the response variable.
Limits	
Limits	specifies which limit lines to display. Limits are displayed as heavier line segments with a serif at the end that extends from each bar. Limit lines are displayed only if you select the **Mean** statistic.
Limit statistic	specifies the statistic for the limit lines.
Limit multiplier	specifies the number of standard units for the limit lines. By default, this value is 1.

Setting Options

Option Name	Description
Title and Footnote	
You can specify a custom title and footnote for the output. You can also specify the font size for this text.	
Bar Details	

Option Name	Description
Apply bar color	specifies the color for the bars when a column is not assigned to the **Group variable** role.
Transparency	specifies the degree of transparency for the plot. The range is 0 (completely opaque) to 1 (completely transparent).
Apply bar gradient	applies a gradient to each bar. **Note:** This option is available only if you are running the second maintenance release of SAS 9.4 or later.
Data skin	specifies a special effect to be used on all filled bars.
Bar Labels	
Show bar labels or statistics	displays the values of the calculated response as data labels.
Category Axis	
Reverse	specifies that the values for the tick marks are displayed in reverse (descending) order.
Show values in data order	places the discrete values for the tick marks in the order in which they appear in the data.
Show label	enables you to display a label for the axis. Enter this label in the **Custom label** box.
Response Axis	
Show grid	creates grid lines at each tick on the axis.
Drop statistics suffix	removes the name of the calculated statistic in the axis label. For example, if you are calculating the mean, the axis label could be Weight (Mean).

Option Name	Description
Show label	specifies whether to display the label for the response axis. By default, the axis label is the name of the variable. However, you can create a custom label.
Legend Details	
Legend location	specifies whether the legend is placed outside or inside the axis area.
Graph Size	

You can specify the width and height of the graph in inches, centimeters, or pixels.

Bar-Line Chart Task

About the Bar-Line Chart Task

The Bar-Line Chart task creates a vertical bar chart with a line chart overlay.

You can use this task to perform the following tasks:

- display and compare exact and relative magnitudes

- examine the contribution of each part to the whole

- determine trends and patterns in the data

Example: City and Highway Mileage by Origin

For example, you can create a bar-line chart that compares the number of miles per gallon (in the city and on the highway) that cars use depending on their country of origin. The task calculates the mean of the number of miles per gallon in the city and in

the highway for each country. This bar-line chart shows that cars from Asia tend to get the highest number of miles per gallon in city and highway driving.

To create this example:

1   In the **Tasks** section, expand the **Graph** folder and double-click **Bar-Line Chart**. The user interface for the Bar-LineChart task opens.

2   On the **Data** tab, select the **SASHELP.CARS** data set.

3   Assign columns to these roles:

Role	Column Name
Category variable	**Origin**
Bar response variable	**MPG_City**
Line response variable	**MPG_Highway**

4   To run the task, click .
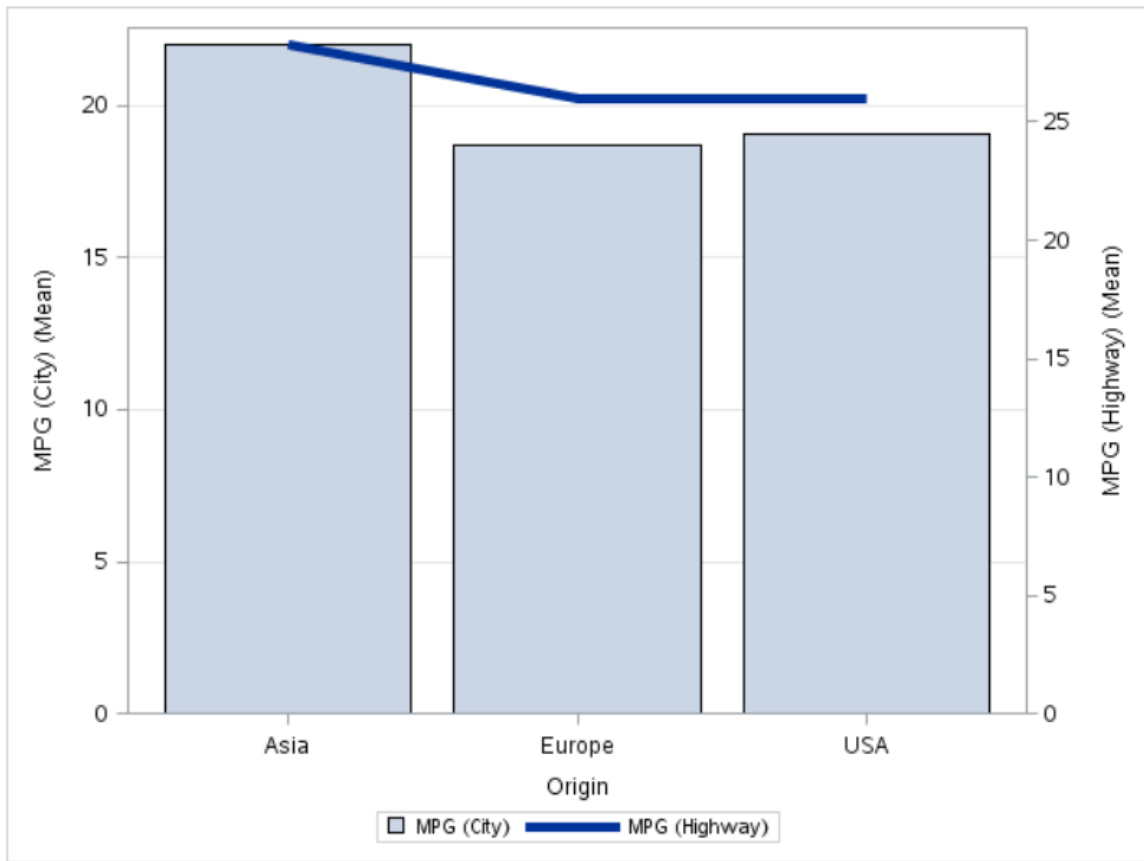
Assigning Data to Roles

You can subset the data in the input data set by applying an SQL WHERE clause. In the **Where string** box, enter your valid SQL syntax. You must enclose character strings in single quotation marks. To view this WHERE clause with the resulting graph, select the **Include as footnote** check box.

To run the Bar-Line Chart task, you must assign a column to the **Category variable**, **Bar response variable**, and **Line response variable** roles.

Option Name	Description
Roles	

Option Name	Description
Category variable	specifies the variable that classifies the observations into distinct subsets.
Bar response variable	specifies a numeric response variable for the bar chart.
Line response variable	specifies a numeric response variable for the line plot.
Group variable	specifies a variable that is used to group the data.
URL variable	specifies a character variable that contains URLs for web pages to be displayed when parts of the plot are selected within an HTML page.
Statistics	
Mean	calculates the mean of the response variables.
Sum	calculates the sum of the response variables.

Setting Options

Option Name	Description
Title and Footnote	

You can specify a custom title and footnote for the output. You can also specify the font size for this text.

Bar Details	
Apply bar color	specifies the color for the bars.

Option Name	Description
Transparency	specifies the degree of transparency for the plot. The range is 0 (completely opaque) to 1 (completely transparent).
Apply bar gradient	applies a gradient to each bar. **Note:** This option is available only if you are running the second maintenance release for SAS 9.4 or later.
Data skin	specifies a special effect to be used on all filled bars.
Line Details	
Apply line color	specifies the color for the line.
Line thickness	specifies the thickness (in pixels) of the line.
Transparency	specifies the degree of transparency for the plot. The range is 0 (completely opaque) to 1 (completely transparent).
Use solid line pattern	specifies a solid pattern for the line.
Category Axis	
Reverse	specifies that the values of the tick marks are displayed in reverse (descending) order.
Show values in data order	places the discrete values for the tick marks in the order in which they appear in the data.
Show label	enables you to display a label for the axis. Enter this label in the **Custom label** box.
Response Axes	
Use zero baseline	specifies whether to offset all lines from the discrete category values and all bars from category midpoints. By default, there is no offset.

Option Name	Description
Use uniform scale	uses the same scale for both response axes.
Show grid on left (bar) axis	creates grid lines at each tick on the axis for the bar chart.
Drop statistics suffix	removes the name of the calculated statistic in the axis label. For example, if you are calculating the mean, the axis label could be Weight (Mean).
Add plot prefix to axis labels	adds (Bar) and (Line) to the labels for the response axes.
Custom label for left (bar) axis	enables you to specify a custom label for the response axis in the bar chart. The default label is the name of the bar response variable.
Custom label for right (line) axis	enables you to specify a custom label for the response axis in the line chart. The default label is the name of the line response variable.
Legend Details	
Legend location	specifies whether the legend is placed outside or inside of the axis area.
Graph Size	

You can specify the width and height of the graph in inches, centimeters, or pixels.

Box Plot Task

About the Box Plot Task

The Box Plot task creates a single box plot, which represents numeric values measured as intervals. If you choose to categorize the values of the analysis variable, then multiple box plots are created.

Example: Box Plots Comparing MPG (City) for Cars

This example creates three box plots that compares how many miles per gallon (city) cars consume depending on their area of origin (Asia, Europe, and United States).

To create this example:

1 In the **Tasks** section, expand the **Graph** folder and double-click **Box Plot**. The user interface for the Box Plot task opens.

2 On the **Data** tab, select the **SASHELP.CARS** data set.

3 Assign columns to these roles:

Role	Column Name
Analysis variable	MPG_City
Category variable	Origin

4 To run the task, click 🏃.

Here is the resulting box plot:
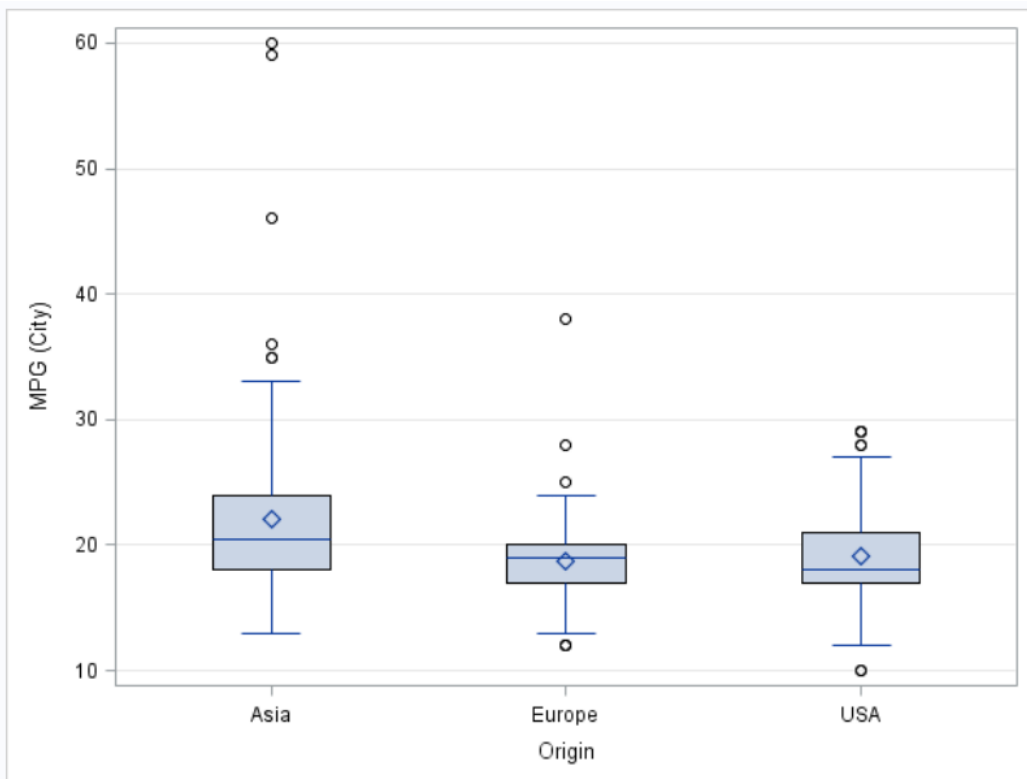
Assigning Data to Roles

You can subset the data in the input data set by applying an SQL WHERE clause. In the **Where string** box, enter your valid SQL syntax. You must enclose character strings in single quotation marks. To view this WHERE clause with the resulting graph, select the **Include as footnote** check box.

To run the Box Plot task, you must assign a column to the **Analysis variable** role. You can create either a vertical box plot or a horizontal box plot.

Role	Description
Analysis variable	specifies the analysis variable for the plot.

Role	Description
Category variable	creates a box plot for each distinct value of the category variable.
Group variable	specifies a variable that is used to group the data.
BY variable	creates a separate graph for each BY group.

Setting Options

Option Name	Description
Title and Footnote	

You can specify a custom title and footnote for the output. You can also specify the font size for this text.

Box Detail	
Box width	specifies the width of each box. Specify a value between 0.0 (0% of the available width) and 1.0 (100% of the available width).
Fill	specifies whether the boxes are filled with color. The default color is white.
Data skin	specifies a special effect to be used on the plot. The data skin affects all filled boxes. The effect that a data skin has on a filled area depends on the skin type, the graph style, and the color of the skinned element. Most of the skins work best with lighter colors over a medium to large filled area.
Transparency	specifies the degree of transparency for the plot. The default value is 0. However, valid values range from 0 (completely opaque) to 1 (completely transparent).

Option Name	Description
Set cap shape	specifies whether to display the cap lines for the whiskers. If you select this option, you can select the shape of the whisker cap lines. Here are the valid values:
	■ **Bracket** displays a straight line with brackets.
	■ **Line** displays a straight line.
	■ **Serif** displays a short straight line.
Notches	specifies that the boxes be notched. The endpoints of the notches are at the following computed locations: $median \pm 1.58\left(\frac{IQR}{\sqrt{N}}\right)$
Group Layout	
Group Order	specifies the orderi of boxes within a group. The groups can be displayed in ascending order, in descending order, or in the order in which they appear in the data.
Category Axis	
Reverse	specifies that the values for the tick marks are displayed in reverse (descending) order.
Show values in data order	places the discrete values for the tick marks in the order in which they appear in the data.
Show label	enables you to display a label for the axis. Enter this label in the **Custom label** box.
Analysis Axis	
Show grid	creates grid lines at each tick on the axis.
Show label	specifies whether to display the label for the analysis axis. By default, the axis label is the name of the variable. However, you can create a custom label.

Option Name	Description
Legend Details	
Legend location	specifies whether the legend is placed outside or inside the axis area.
Graph Size	
You can specify the width and height of the graph in inches, centimeters, or pixels.	

Bubble Plot Task

About the Bubble Plot Task

The Bubble Plot task explores the relationship between three or more variables. In a bubble plot, two variables determine the location of the bubble centers, and a third variable specifies the size of each bubble.

Example:

To create this example:

1   In the **Tasks** section, expand the **Graph** folder and double-click **Bubble Plot**. The user interface for the Bubble Plot task opens.

2   On the **Data** tab, select the **SASHELP.CLASS** data set.

3   Assign columns to these roles:

Role	Column Name
X variable	**Height**
Y variable	**Weight**

Role	Column Name
Size variable	Age

4 To run the task, click ![run icon].
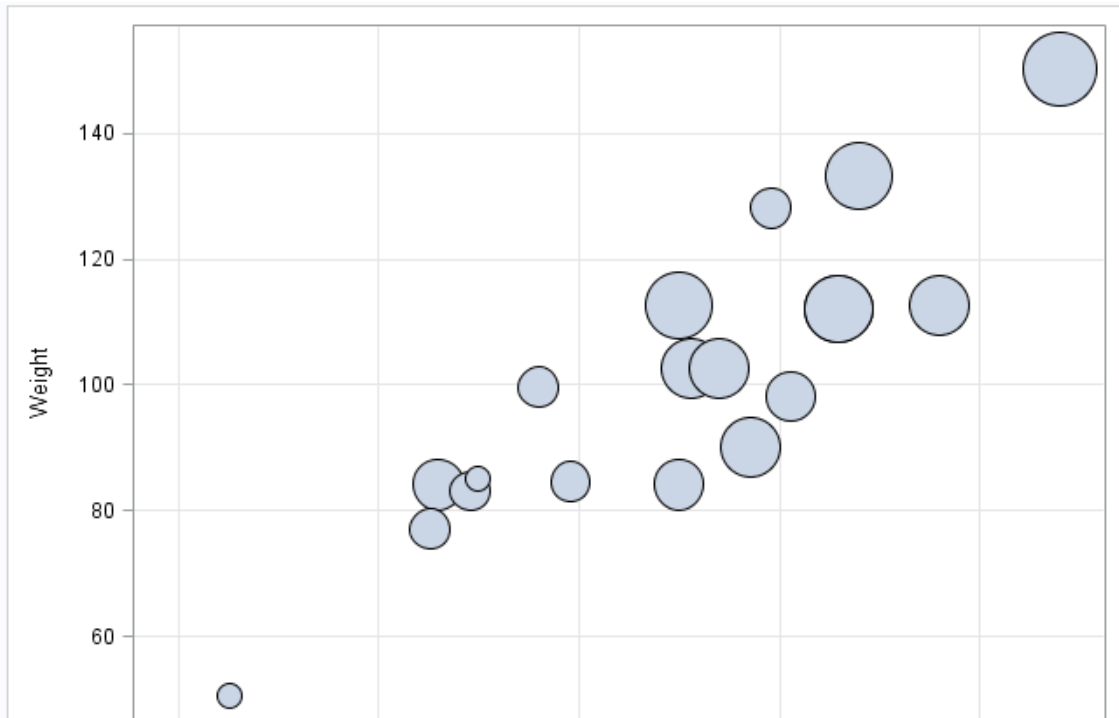
Assigning Data to Roles

You can subset the data in the input data set by applying an SQL WHERE clause. In the **Where string** box, enter your valid SQL syntax. You must enclose character strings in single quotation marks. To view this WHERE clause with the resulting graph, select the **Include as footnote** check box.

To run the Bubble Plot task, you must assign a column to the **X variable**, **Y variable**, and **Size variable** roles.

Option Name	Description
Roles	
X variable	specifies the variable for the X axis.
Y variable	specifies the variable for the Y axis.
Size variable	specifies a numeric variable that controls the size of the bubbles. The minimum and maximum values automatically provide the range that is used to determine bubble size. You can set these values on the **Options** tab.
Color response variable	specifies the numeric variable that is used to map colors to a gradient legend.
Group variable	specifies the variable that is used to group the data. The bubbles for each group are automatically distinguished by different colors.
	Note: If you also assign a variable to the **Color response variable** role, the group variable is ignored.
Label variable	specifies the values to use as labels for each data point. If you assign a variable to this role, the values of that variable are used for the data labels. If you do not assign a variable to this role, the values of the Y variable are used for the data labels.
URL variable	specifies a character variable that contains URLs for web pages to be displayed when parts of the plot are selected within an HTML page.

Setting Options

Option Name	Description
Title and Footnote	
You can specify a custom title and footnote for the output. You can also specify the font size for this text.	
Bubble Details	
Apply color	specifies the color for the bars when a column is not assigned to the **Group variable** role.
Minimum radius	specifies the radius of the smallest bubble.
Maximum radius	specifies the radius of the largest bubble.
Transparency	specifies the degree of transparency for the bubbles. The range is 0 (completely opaque) to 1 (completely transparent).
Data skin	specifies a special effect to be used on all filled bubbles.
Color Model	
If you assign a variable to the **Color response variable** role, you can specify three colors to use in the color map.	
Bubble Labels	
If you assign a variable to the **Label variable** role, you can determine the label color, the font size for the label text, and the label position.	
X Axis and Y Axis	
Show grid lines	creates grid lines at each tick on the axis.

Option Name	Description
Show label	specifies whether to display the label for the response axis. By default, the axis label is the name of the variable. However, you can create a custom label.
Legend Details	
Group legend location	specifies whether the group legend is located inside or outside the plot. This option is available when you assign a variable to the **Group variable** role.
Color legend position	specifies whether the color legend appears to the right or below the plot.
Graph Size	

You can specify the width and height of the graph in inches, centimeters, or pixels.

Histogram Task

About the Histogram Task

The Histograms task creates a chart that displays the frequency distribution of a numeric variable.

Example: Histogram of Stock Volume

To create this example:

1 In the **Tasks** section, expand the **Graph** folder and double-click **Histogram**. The user interface for the Histogram task opens.

2 In the **Data** tab, select the **SASHELP.STOCKS** data set.

3   To the **Analysis variable** role, assign the **Volume** column.

4   To run the task, click ⃗.
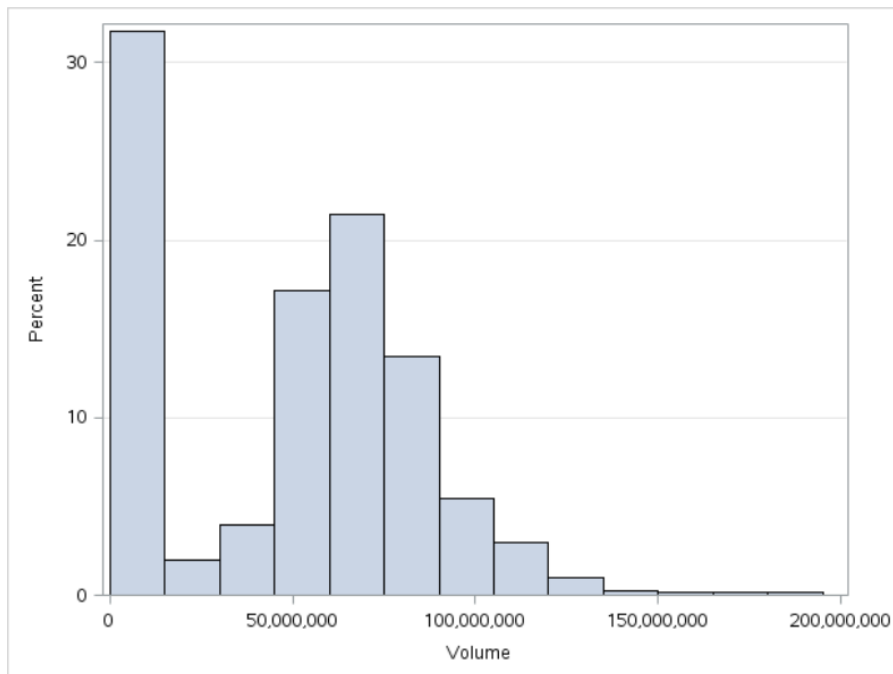
Here are the results:

Assigning Data to Roles

You can subset the data in the input data set by applying an SQL WHERE clause. In the **Where string** box, enter your valid SQL syntax. You must enclose character strings in single quotation marks. To view this WHERE clause with the resulting graph, select the **Include as footnote** check box.

To run the Histogram task, you must assign a column to the **Analysis variable** role.

You can specify whether to create a density curve that shows the distribution of values for a numeric variable. You can create density curves for normal and kernel distributions.

Setting Options

Option Name	Description
Title and Footnote	
You can specify a custom title and footnote for the output. You can also specify the font size for this text.	
Bin Details	
For the bins in the histogram, you can specify the color and the transparency. If you are running the second maintenance release for SAS 9.4 or later, you can also specify whether to apply a gradient to each bin.	
Horizontal Axis	
Interval axis	creates tick marks at regular intervals on the horizontal axis based on the minimum and maximum values of the analysis variable.
Bin axis	creates tick marks at the midpoints of the value bins on the horizontal axis.
Specify number of bins	enables you to specify the number of bins in the histogram. Valid values range from 2 to 20.
	The bins always span the range of data. The task tries to produce tick values that are easily interpreted (for example, 5, 10, 15, 20). Sometimes the location of the first bin and the bin width might be adjusted. By default, the task automatically determines the number of bins.
Show label	displays the label for the analysis variable along the horizontal axis. You can also enter a custom label.
Vertical Axis	

Option Name	Description
Specify axis scaling	specifies the scaling that is applied to the vertical axis. You can choose from these options:
	COUNT the axis displays the frequency count
	PERCENT the axis displays values as a percentage of the total.
	PROPORTION the axis displays values as proportions (0.0 to 1.0) of the total.
Show grid	specifies whether to show the grid lines for the vertical axis.
Show label	specifies whether to show the label for the type of axis scaling.
Legend Details	
Show legend	specifies whether to display a legend in the output.
Legend location	specifies whether the legend is placed outside or inside of the axis area.
Graph Size	

You can specify the width and height of the graph in inches, centimeters, or pixels.

Line Chart Task

About the Line Chart Task

The Line Chart task assumes that the values in the category variable are discrete. The task groups these values into distinct categories. If you assign a column from the input data source to the **Response variable** role, you can select the statistic (either mean or sum) for the response values. By default, the task calculates the mean of the values for the response variable. If no response variable is assigned, a frequency chart by category is created.

Example: Displaying the Mean Horsepower for Each Car Type

In this example, you want to display the mean horsepower for each car type in a line plot. The result shows that sports cars have the highest average horsepower and hybrid cars have the lowest average horsepower.

To create this example:

1 In the **Tasks** section, expand the **Graph** folder and double-click **Line Chart**. The user interface for the Line Chart task opens.

2 On the **Data** tab, select the **SASHELP.CARS** data set.

3 Assign columns to these roles:

Role	Column Name
Category variable	**Type**
Response variable	**Horsepower**

4 To run the task, click ![run icon].
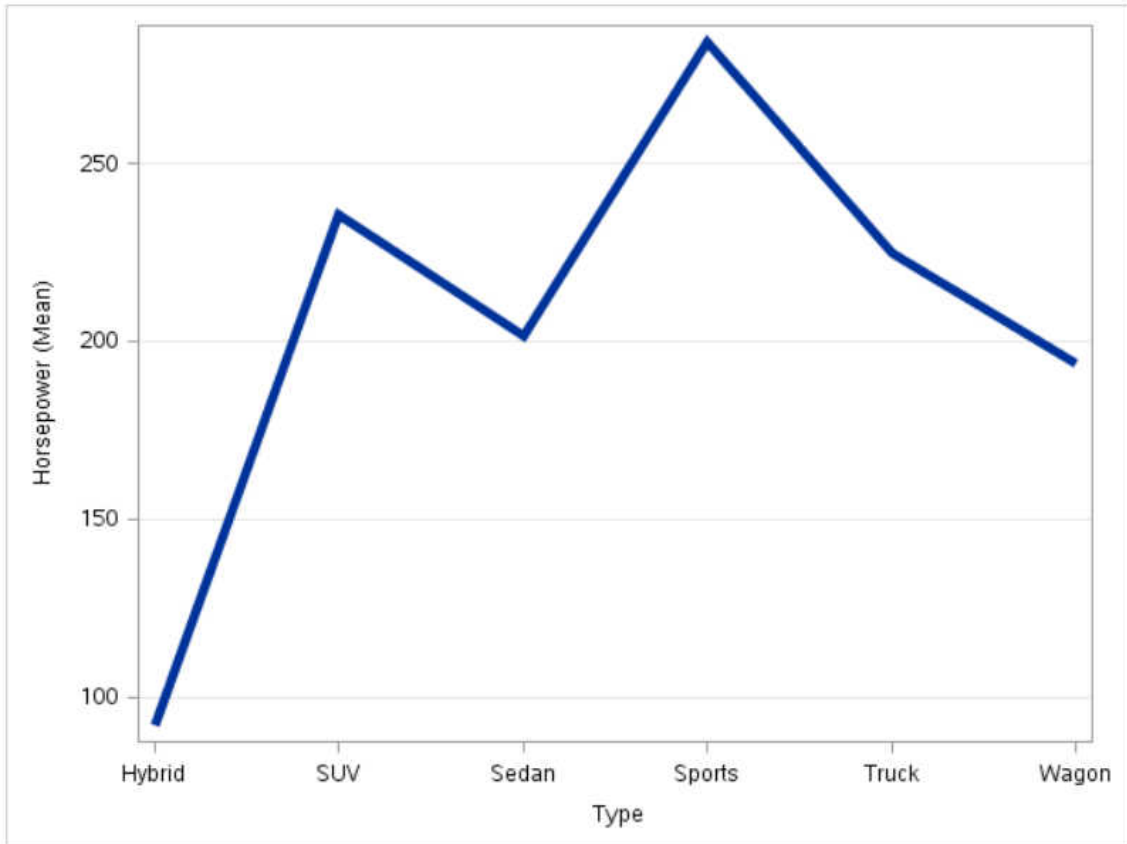
Assigning Data to Roles

You can subset the data in the input data set by applying an SQL WHERE clause. In the **Where string** box, enter your valid SQL syntax. You must enclose character strings in single quotation marks. To view this WHERE clause with the resulting graph, select the **Include as footnote** check box.

To run the Line Chart task, you must assign a column to the **Category variable** role.

Option Name	Description
Role	

Option Name	Description
Category variable	specifies the variable that classifies the observations into distinct subsets.
Response variable	specifies a numeric response variable for the plot.
Group variable	specifies a variable that is used to group the data.
URL variable	specifies a character variable that contains URLs for web pages to be displayed when parts of the plot are selected within an HTML page.
Statistics	
Mean	calculates the mean of the response variable.
Sum	calculates the sum of the response variable.

Setting Options

Option Name	Description
Title and Footnote	

You can specify a custom title and footnote for the output. You can also specify the font size for this text.

Line Details	
Apply line color	specifies the color for the line when you do not assign a column to the **Group variable** role.
Line thickness	specifies the thickness (in pixels) of the line.

Option Name	Description
Transparency	specifies the degree of transparency for the plot. The range is 0 (completely opaque) to 1 (completely transparent).
Use solid line	specifies a solid pattern for the line.
Line Labels	
Show line labels	displays the label from the response variable. If you assign a column to the **Group variable** role, each line is labeled with the group value.
Category Axis	
Reverse	specifies that the values of the tick marks are displayed in reverse (descending) order.
Show values in data order	places the discrete tick values in the order in which they appear in the data.
Show label	enables you to display a label for the axis. By default, the label is the variable name.
	To customize this label, enter this label in the **Custom label** box.
Response Axis	
Show grid	creates grid lines at each tick on the axis.
Drop statistics suffix	removes the name of the calculated statistic in the axis label. For example, if you are calculating the mean, the axis label could be Weight (Mean).
Show label	enables you to display a label for the axis. By default, the label is the variable name.
	To customize this label, enter this label in the **Custom label** box.
Legend Details	

Option Name	Description
Legend location	specifies whether the legend is placed outside or inside of the axis area.

Graph Size	
You can specify the width and height of the graph in inches, centimeters, or pixels.	

Pie Chart Task

About the Pie Chart Task

The Pie Chart task creates pie charts that represent the relative contribution of the parts to the whole by displaying data as wedge-shaped "slices" of a circle. Each slice represents a category of data. The size of a slice represents the contribution of the data to the total chart statistic.

Example: Pie Chart That Shows Total MSRP for Each Car Type by Region

In this example, you want to compare the manufacturer's suggested retail price (MSRP) for each car type grouped by region of origin. The resulting pie chart consists of six rings—one for each car type. The rings are then subset into the MSRP values for the three regions: Asia, Europe, and USA. Using this chart, you can compare the total MSRP values for each region. The ring for the SUV car type shows that the USA has the highest MSRP and that Europe has the lowest MSRP.

To create this example:

1 In the **Tasks** section, expand the **Graph** folder and double-click **Pie Chart**. The user interface for the Pie Chart task opens.

2 On the **Data** tab, select the **SASHELP.CARS** data set.

3 Assign columns to these roles:

Role	Column Name
Category variable	**Origin**
Response variable	**MSRP**
Group variable	**Type**
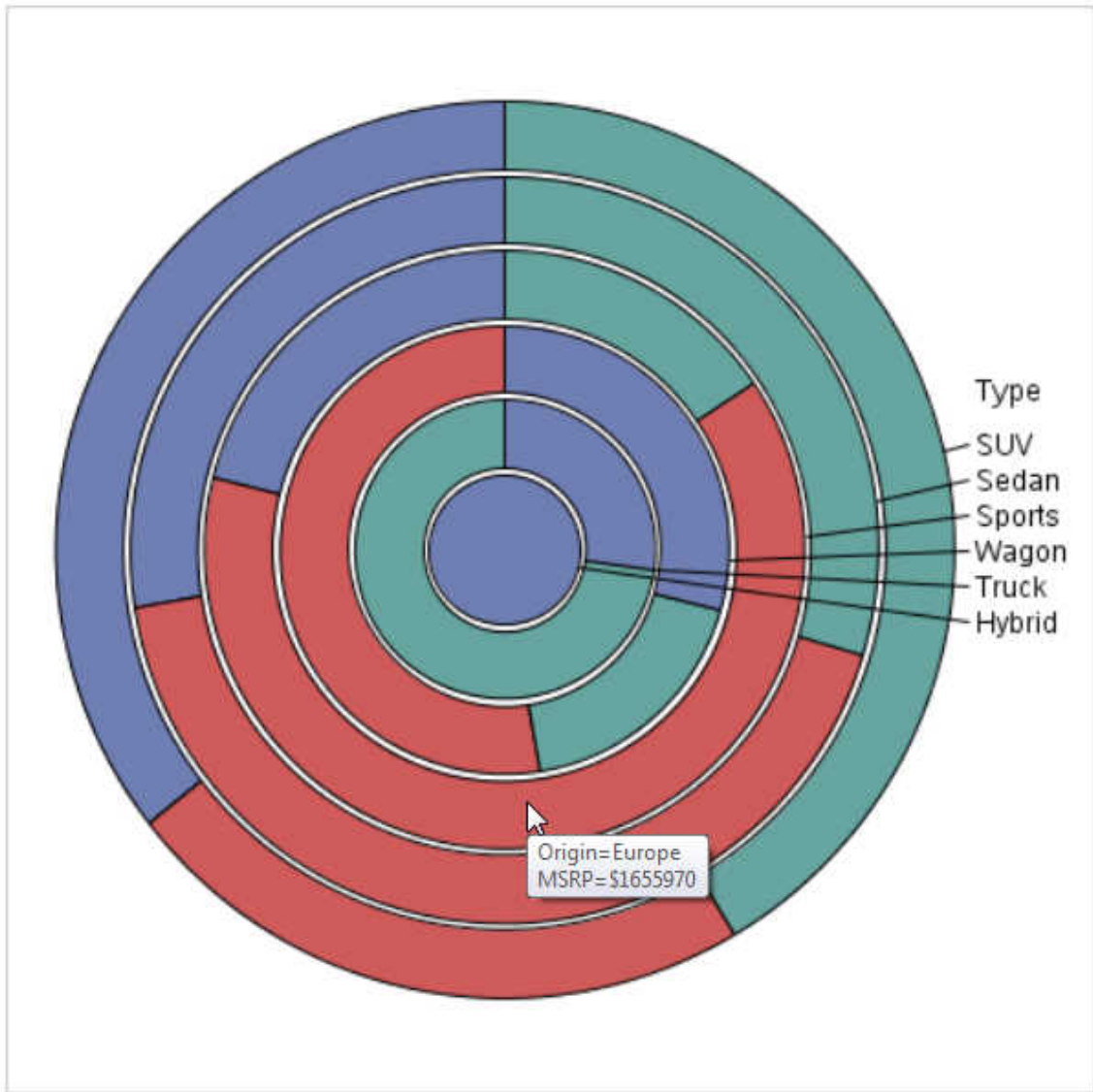
4 To run the task, click .

Here is the result:

Assigning Data to Roles

You can subset the data in the input data set by applying an SQL WHERE clause. In the **Where string** box, enter your valid SQL syntax. You must enclose character strings in

single quotation marks. To view this WHERE clause with the resulting graph, select the **Include as footnote** check box.

To run the Pie Chart task, you must assign a column to the **Category variable** role.

Option Name	Description
Role	
Category variable	specifies the variable that classifies the observations into distinct subsets.
Response variable	specifies a numeric response variable for the plot.
Group variable	specifies a variable that is used to group the data.
URL variable	specifies a character variable that contains URLs for web pages to be displayed when parts of the plot are selected within an HTML page. **Note:** If the task generates an "Other" slice in the pie chart, there is not a URL associated with this slice. Therefore, this slice does not contain a link.
Orientation	
Starting point	specifies where to create the first slice in the pie chart. The remaining slices appear counterclockwise.
Center the first slice	specifies whether to offset the first slice.

Setting Options

Option Name	Description
Title and Footnote	

Option Name	Description
You can specify a custom title and footnote for the output. You can also specify the font size for this text.	
Pie Details	
Fill transparency	specifies the degree of transparency for the plot. The range is 0 (completely opaque) to 1 (completely transparent).
Data skin	specifies a special effect to be used on all filled bars.
Pie Labels	
Location	specifies whether to display the label inside or outside the slice in the pie chart. By default, the Pie Chart task determines the best location for the slice.
Set label font size	enables you to specify the font size of the label for each slice.
Graph Size	
You can specify the width and height of the graph in inches, centimeters, or pixels.	

Scatter Plot Task

About the Scatter Plot Task

The Scatter Plot task creates plots that show the relationships between two or three variables by revealing patterns or concentrations of data points. For example, a two-dimensional scatter plot can display the heights and weights of all students in a class.

Example: Scatter Plot of Height versus Weight

In this example, you want to create a scatter plot of height versus weight.

To create this example:

1   In the **Tasks** section, expand the **Graph** folder and double-click **Scatter Plot**. The user interface for the Scatter Plot task opens.

2   On the **Data** tab, select the **SASHELP.CLASS** data set.

3   Assign columns to these roles:

Role	Column Name
X variable	**Height**
Y variable	**Weight**

4   To run the task, click .
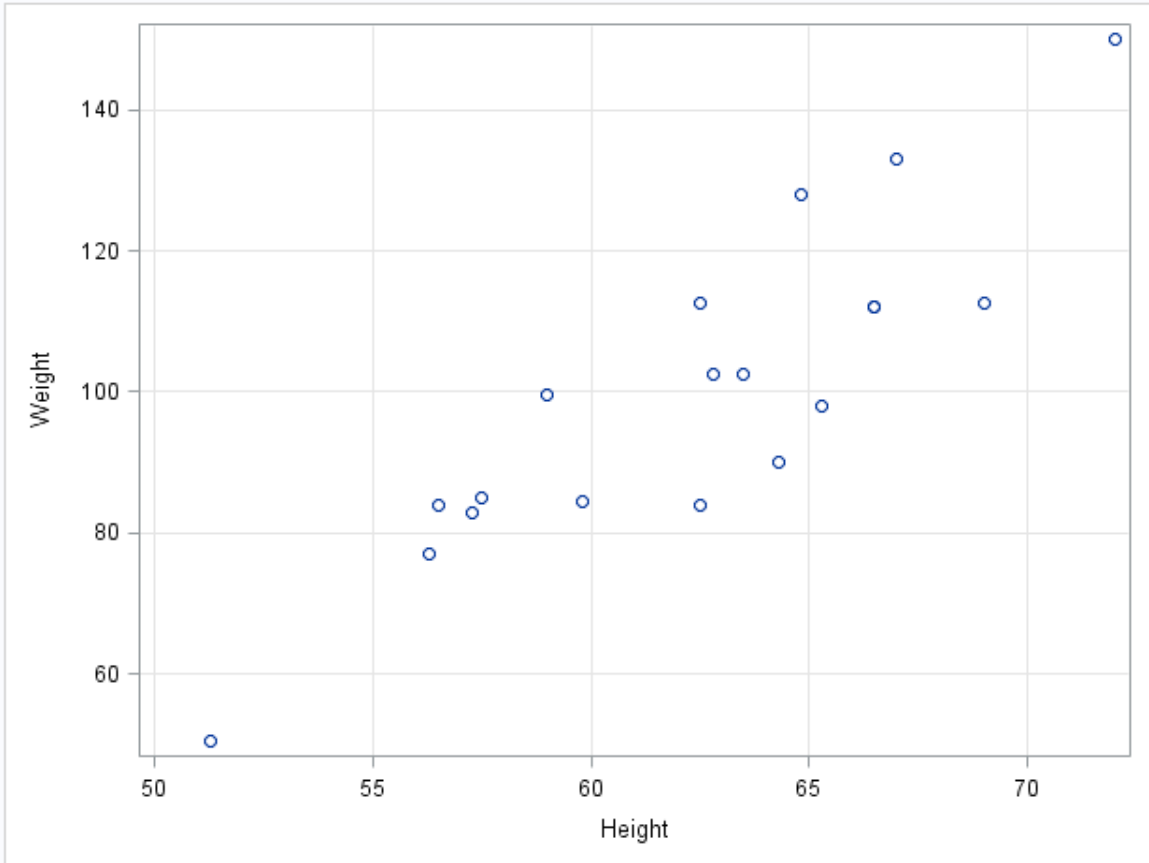
Assigning Data to Roles

You can subset the data in the input data set by applying an SQL WHERE clause. In the **Where string** box, enter your valid SQL syntax. You must enclose character strings in single quotation marks. To view this WHERE clause with the resulting graph, select the **Include as footnote** check box.

To run the Scatter Plot task, you must assign columns to the **X variable** and **Y variable** role.

Option Name	Description
Roles	

Option Name	Description
X variable	specifies the variable for the x axis.
Y variable	specifies the variable for the y axis.
Group variable	specifies a variable that is used to group the data. The plot elements for each group value are automatically distinguished by different visual attributes.
Marker label variable	displays a label for each data point. If you specify a variable, the values of that variable are used for the data labels. If you do not specify a variable, then the values of the Y variable are used for the data labels.
URL variable	specifies a character variable that contains URLs for web pages to be displayed when parts of the plot are selected within an HTML page.
Fit Plots **Note:** This is available only if you assign numeric variables.	
Regression	creates a plot with the fitted regression line. You can specify whether to include the confidence limits for means and the prediction limits for the individual predicted values. The **Alpha** option specifies the confidence level for the confidence limits. The **Degree** option specifies the degree of the polynomial fit.
Loess	creates a fitted loess curve. You can specify whether to include the confidence limits. The **Alpha** option specifies the confidence level for the confidence limits.
PBSpline	creates a fitted penalized B-spline curve. You can specify whether to include the confidence limits for means and the prediction limits for the individual predicted values. The **Alpha** option specifies the confidence level for the confidence limits.

Setting Options

Option Name	Description
Title and Footnote	
You can specify a custom title and footnote for the output. You can also specify the font size for this text.	
Marker Details	
You can specify the symbol type, color, and size of the markers. You can also specify the degree of transparency for the plot. The range is 0 (completely opaque) to 1 (completely transparent).	
Marker Labels	
Font size	specifies the appearance of the labels in the plot when you assign a variable to the **Marker label variable** role.
X Axis, Y Axis	
Show grid lines	creates grid lines at each tick on the axis.
Show label	displays the label for the axis. By default, the label is the variable name.
	To customize, enter this label in the **Custom label** box.
Legend Details	
Legend location	specifies whether the legend is placed outside or inside the axis area.
Graph Size	
You can specify the width and height of the graph in inches, centimeters, or pixels.	

Series Plot Task

About the Series Plot Task

The Series Plot task creates a line plot. Series plots display a series of line segments that connect observations of input data.

Example: Series Plot of Stock Trends

In this example, you want to create a series plot that shows stock trends.

To create this example:

1   In the **Tasks** section, expand the **Graph** folder and double-click **Series Plot**. The user interface for the Series Plot task opens.

2   On the **Data** tab, select the **SASHELP.STOCKS** data set.

3   Assign columns to these roles:

Role	Column Name
X variable	**Date**
Y variable	**Open**
Group variable	**Stock**

4   To run the task, click ![run icon].

The resulting series plot shows the stock values for three companies.
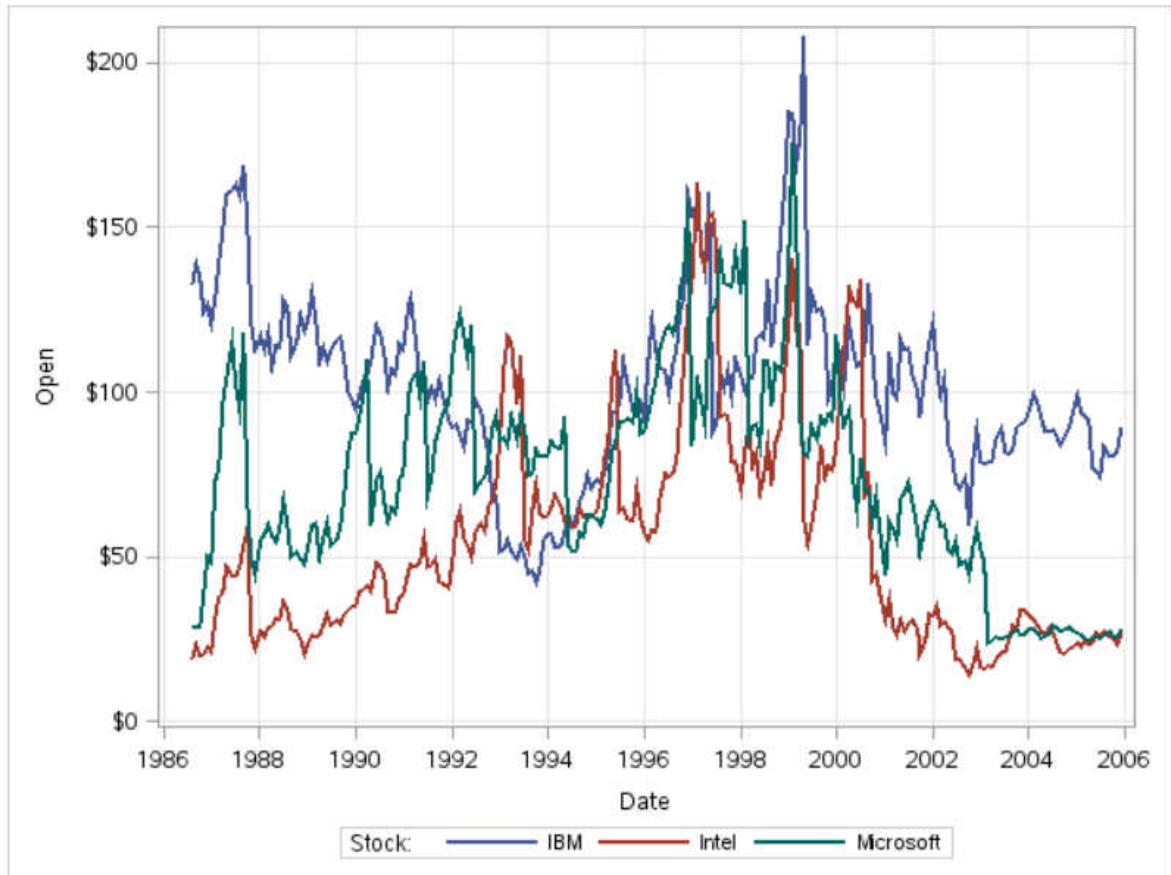
Assigning Data to Roles

You can subset the data in the input data set by applying an SQL WHERE clause. In the **Where string** box, enter your valid SQL syntax. You must enclose character strings in single quotation marks. To view this WHERE clause with the resulting graph, select the **Include as footnote** check box.

To run the Series Plot task, you must assign columns to the **X variable** and **Y variable** roles.

Role	Description
X variable	specifies the variable for the x axis.
Y variable	specifies the variable for the y axis.
Group variable	specifies a variable that is used to group the data.
URL variable	specifies a character variable that contains URLs for web pages to be displayed when parts of the plot are selected within an HTML page.

Setting Options

Option Name	Description
Title and Footnote	
You can specify a custom title and footnote for the output. You can also specify the font size for this text.	
Plot Details	
You can specify the symbol type, color, and size of the markers in the scatter plot. You can also specify the degree of transparency for the plot. The range is 0 (completely opaque) to 1 (completely transparent).	
Plot Labels	
Show plot labels	adds a label for the curve. You can also specify the size of this text.
X Axis, Y Axis	
Show grid lines	creates grid lines at each tick on the axis.

Option Name	Description
Show label	displays the label for the axis. By default, the label is the variable name.
	To customize, enter this label in the **Custom label** box.
Legend Details	
Legend location	specifies whether the legend is placed outside or inside of the axis area.
Graph Size	

You can specify the width and height of the graph in inches, centimeters, or pixels.

Simple HBar Task

About the Simple HBar Task

The Simple HBar task creates a simple horizontal bar chart. You can customize the title, footnotes, axes, and legends for the horizontal bar chart.

Example: Horizontal Bar Chart of Mileage by Origin and Type

To create this horizontal bar chart:

1   In the **Tasks** section, expand the **Graphs** folder and double-click **Simple HBar**. The user interface for the Simple HBar task opens.

2   On the **Data** tab, select the **SASHELP.CARS** data set.

3   Assign columns to these roles:

Role	Column Name
Category variable	**Origin**
Response variable	**MPG_City**
Group variable	**Type**
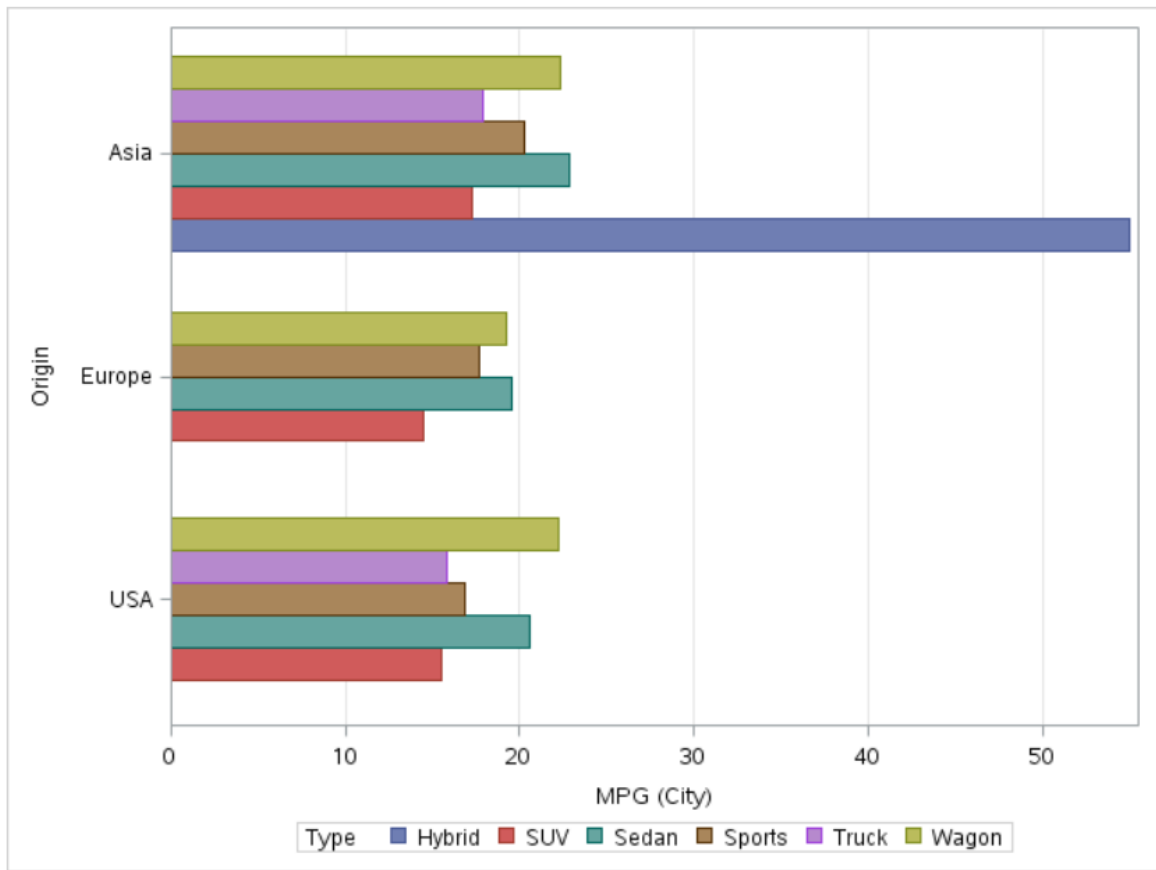
4 To run the task, click ![run icon].

Here are the results:

Assigning Data to Roles

You can subset the data in the input data set by applying an SQL WHERE clause. In the **Where string** box, enter your valid SQL syntax. You must enclose character strings in single quotation marks. To view this WHERE clause with the resulting graph, select the **Include as footnote** check box.

To run the Simple HBar task, you must assign a column to the **Category variable** role.

Option Name	Description
Role	
Category variable	specifies the variable that classifies the observations into distinct subsets.
Response variable	specifies a numeric response variable for the plot.
Group variable	specifies a variable that is used to group the data.
URL variable	specifies a character variable that contains URLs for web pages to be displayed when parts of the plot are selected within an HTML page.
Group Layout	
Cluster	displays group values as separate adjacent bars that replace the single category bar. Each set of group values is centered at the midpoint tick mark for the category.
Stack	overlays group values without any clustering. Each group is represented by unique visual attributes that are derived from the GraphData1... GraphData*n* style elements in the current style.
Statistics	

Option Name	Description
Mean	calculates the mean of the response variable.
Sum	calculates the sum of the response variable.

Setting Options

Option Name	Description
Title and Footnote	

You can specify a custom title and footnote for the output. You can also specify the font size for this text.

Bar Details	
Apply bar color	specifies the color for the bars when a column is not assigned to the **Group variable** role.
Transparency	specifies the degree of transparency for the plot. The range is 0 (completely opaque) to 1 (completely transparent).
Apply bar gradient	applies a gradient to each bar. **Note:** This option is available only if you are running the second maintenance release for SAS 9.4 or later.
Data skin	specifies a special effect to be used on all filled bars.
Bar Labels	
Show bar labels	displays the values of the calculated response as data labels.
Category Axis	

Option Name	Description
Reverse	specifies that the values of the tick marks are displayed in reverse (descending) order.
Show values in data order	places the discrete tick values in the order in which they appear in the data.
Show label	enables you to display a label for the axis. Enter this label in the **Custom label** box.
Response Axis	
Show grid	creates grid lines at each tick on the axis.
Drop statistics suffix	removes the name of the calculated statistic in the axis label. For example, if you are calculating the mean, the axis label could be Weight (Mean).
Custom Label	enables you to customize the label for the response axis. By default, the axis label is the name of the variable.
Legend Details	
Legend location	specifies whether the legend is placed outside or inside of the axis area.
Graph Size	

You can specify the width and height of the graph in inches, centimeters, or pixels.

11

High-Performance Statistics Tasks

About the High-Performance Tasks

The high-performance tasks are designed to be used with large data. Verify that you have adequate computing power before running these high-performance tasks on large data sets.

Bin Continuous Data Task

About the Bin Continuous Data Task

The Bin Continuous Data task is a data preparation task. This task divides the data values of a continuous variable into intervals and replaces the values for each interval with a single value that is representative of the interval.

Note: This task is available only if you are running SAS 9.4 or later.

Example: Winsorized Binning

In this example, the task provides the basic Winsorized statistical information for the input data.

To create this example:

1 To create the Work.Ex12 data set, enter this code into a **Program** tab:

```
data ex12;
   length id 8;
   do id=1 to 10000;
      x1 = ranuni(101);
      x2 = 10*ranuni(201);
      x3 = 100*ranuni(301);
      output;
   end;
run;
```

   Click ⚐.

2 In the **Tasks** section, expand the **High-Performance Statistics** folder and double-click **Bin Continuous Data**. The user interface for the Bin Continuous Data task opens.

3 On the **Data** tab, select the **WORK.EX12** data set.

4 To the **Variables to bin** role, assign the **x1** and **x2** columns.

5 Select the **Options** tab and set these options:

   ▪ In the **Number of bins** box, enter `10`.

   ▪ From the **Method** drop-down list, select **Winsorized binning**.

6 To run the task, click ⚐.

Here is a subset of the results:

Performance Information	
Execution Mode	Single-Machine
Number of Threads	4

Data Access Information			
Data	Engine	Role	Path
WEBWORK.EX12	V9	Input	On Client

Binning Information	
Method	Winsor Binning
Number of Bins Specified	10
Number of Variables	2

Mapping				
Variable	Binned Variable	Range	Frequency	Proportion
x1	BIN_x1	x1 < 0.137722	1405	0.14050000
		0.137722 <= x1 < 0.227865	849	0.08490000
		0.227865 <= x1 < 0.318007	897	0.08970000
		0.318007 <= x1 < 0.408150	864	0.08640000
		0.408150 <= x1 < 0.498293	906	0.09060000
		0.498293 <= x1 < 0.588435	899	0.08990000
		0.588435 <= x1 < 0.678578	935	0.09350000
		0.678578 <= x1 < 0.768720	901	0.09010000
		0.768720 <= x1 < 0.858863	948	0.09480000
		0.858863 <= x1	1396	0.13960000
x2	BIN_x2	x2 < 1.398500	1385	0.13850000
		1.398500 <= x2 < 2.301366	941	0.09410000
		2.301366 <= x2 < 3.204232	888	0.08880000
		3.204232 <= x2 < 4.107099	855	0.08550000
		4.107099 <= x2 < 5.009965	905	0.09050000
		5.009965 <= x2 < 5.912831	864	0.08640000
		5.912831 <= x2 < 6.815697	908	0.09080000
		6.815697 <= x2 < 7.718563	909	0.09090000

Assigning Data to Roles

To run the Bin Continuous Data task, you must assign a variable to the **Variables to bin** role.

Role	Description
Roles	
Variables to bin	specifies one or more variables as input variables for binning. The specified variables must be interval variables.
Additional Roles	
Frequency count	specifies a numeric variable that contains the frequency of occurrence for each observation. If the frequency value is less than 1 or is missing, the observation is not used in the analysis. If no variable is assigned to the **Frequency count** role, each observation is assigned a frequency of 1.

Setting Options

Option Name	Description
Methods	
Number of bins	specifies the global number of binning levels for all binning variables. This value can be any integer between 2 and 1,000, inclusive. The default number of binning levels is 16.

Option Name	Description
Method	specifies which binning method to use.
	■ **Bucket binning** creates equal-length bins and assigns the data to one of these bins. You can choose the number of bins during the binning. The default number of bins (the binning level) is 16.
	■ **Winsorized binning** is similar to bucket binning except that both tails are cut off to obtain a smooth binning result. This technique is often used to remove outliers during the data preparation stage.
	You must specify a value for the **Winsor rate** option. Valid values are from 0.0 to 0.5 (exclusive). The default value is 0.05.
	■ **Pseudo-quantile binning** mimics the results of the quantile binning method but is more efficient by consuming less CPU time and memory.
Statistics	
Select statistics to display	In the results, you can specify whether to include statistics.
	Here are the additional statistics that you can include:
	■ **Basic statistics** displays the mean, pseudo-median, standard deviation, minimum, maximum, and number of bins for each binning variable.
	■ **Quantile statistics** displays the estimated quantiles and extremes table.

Creating an Output Data Set

You can specify whether to save the results to an output data set. In the **Additional variables to include in the output data set** role, specify any columns from the input data set that you want to include in the output data set.

High-Performance Correlation Analysis Task

About the High-Performance Correlation Analysis Task

Correlation is a statistical procedure for describing the relationship between numeric variables. The relationship is described by calculating correlation coefficients for the variables. The High-Performance Correlation Analysis task computes Pearson statistics for investigating associations among variables. Correlations range from –1 to 1.

Note: This task is available only if you are running SAS 9.4 or later.

Example: Correlation between Weight, Oxygen, and Run Time

To create this example:

1   Create the Work.Fitness data set. For more information, see "FITNESS Data set" on page 407.

2   In the **Tasks** section, expand the **High-Performance Statistics** folder and double-click **Correlation Analysis**. The user interface for the High-Performance Correlation Analysis task opens.

3   On the **Data** tab, select the **WORK.FITNESS** data set.

4   To the **Analysis variables** role, assign the **Weight**, **Oxygen**, and **RunTime** columns.

5   To run the task, click ![run icon].

Here are the results:

Performance Information	
Execution Mode	Single-Machine
Number of Threads	4

Data Access Information			
Data	Engine	Role	Path
WEBWORK.FITNESS	V9	Input	On Client

3 Variables:	Weight Oxygen RunTime

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations			
	Weight	Oxygen	RunTime
Weight	1.00000 31	-0.15358 0.4264 29	0.20072 0.2965 29
Oxygen	-0.15358 0.4264 29	1.00000 29	-0.86843 <.0001 28
RunTime	0.20072 0.2965 29	-0.86843 <.0001 28	1.00000 29

Assigning Data to Roles

To run the High-Performance Correlation Analysis task, you must assign two columns to the **Analysis variables** role.

Role	Description
Roles	
Analysis variables	specifies the columns to use to calculate the correlation coefficients.
Additional Roles	
Frequency count	specifies a numeric column whose value represents the frequency of the observation.

Role	Description
Weight	specifies the weights to use in the calculation of Pearson weighted product-moment correlation.

Setting Options

Option Name	Description
Methods	
Missing values	specifies whether to include missing values in the calculations.
	■ If you select the **Use nonmissing values for all selected variables** option, any observations that have missing values are excluded from the analysis.
	■ If you select the **Use nonmissing values for pairs of variables** option, the data for an observation contributes to the correlation between two variables as long as both values are nonmissing. As a result, the correlations that are calculated for the analysis variable might be based on a different number of observations.
Statistics	

You can specify whether the results include only the statistics that the task automatically generates, the statistics that you selected, or no statistics. By default, only the correlations table is displayed in the results.

You can include these statistics in the results:

■ correlations

■ covariances

■ sum of squares and cross-products

■ corrected sum of squares and cross-products

■ descriptive statistics

Option Name	Description
Display p-values	specifies whether to display for each correlation coefficient the probability of observing a more extreme value than the observed coefficient.
Order correlations from highest to lowest	displays the ordered correlation coefficients for each variable. Correlations are ordered from highest to lowest in absolute value.

Creating an Output Data Set

You can specify whether to save the results to an output data set. By default, the output data set contains the correlations. You can also include covariances, sum of squares and cross-products, and corrected sum of squares and cross-products.

Generalized Linear Models

About the Generalized Linear Models Task

The Generalized Linear Models task is a high-performance task that provides model fitting and model building for generalized linear models. It fits models for standard distributions such as Normal, Poisson, and Tweedie in the exponential family. This task also fits multinomial models for ordinal and nominal responses. The task provides forward, backward, and stepwise selection methods.

Note: This task is available only if you are running SAS 9.4 or later.

Example: Model Selection

To create this example:

1. Create the Work.getStarted data set. For more information, see .

2. In the **Tasks** section, expand the **High-Performance Statistics** folder and double-click **Generalized Linear Models**. The user interface for the Generalized Linear Models task opens.

3. On the **Data** tab, select the **WORK.GETSTARTED** data set.

4. Assign columns to these roles:

Role or Option Name	Column Name
Distribution	Poisson
Response variable	Y
Classification variables	C1 C2 C3 C4 C5

5. Click the **Model** tab. In the **Variables** box, select **C1–C5**. Click **Add**.

6. Click the **Selection** tab. From the **Selection method** drop-down list, select **Forward selection**.

7. To run the task, click ![run icon].

Here is a subset of the results:

Performance Information	
Execution Mode	Single-Machine
Number of Threads	4

Data Access Information			
Data	Engine	Role	Path
WEBWORK.GETSTARTED	V9	Input	On Client

Model Information	
Data Source	WEBWORK.GETSTARTED
Response Variable	Y
Class Parameterization	GLM
Distribution	Poisson
Link Function	Log
Optimization Technique	Newton-Raphson with Ridging

Selection Information	
Selection Method	Forward
Select Criterion	Significance Level
Stop Criterion	Significance Level
Effect Hierarchy Enforced	Single
Entry Significance Level (SLE)	0.05
Stop Horizon	1

Number of Observations Read	100
Number of Observations Used	100

Class Level Information		
Class	Levels	Values
C1	4	0 1 2 3
C2	4	0 1 2 3
C3	4	0 1 2 3

Assigning Data to Roles

To run the Generalized Linear Models task, you must assign a column to the **Response variable** role.

Option Name	Description
Roles	

Option Name	Description
Response	
Distribution	specifies the distribution for your model. You can choose from these distributions: ■ Binomial ■ Gamma ■ Inverse Gaussian ■ Multinomial ■ Negative binomial ■ Normal ■ Poisson ■ Tweedie
Options for Binomial Distribution	
Response data consists of numbers of events and trials	specifies whether the data consists of a variable that specifies the number of positive responses (events) and another variable that specifies the number of trials.
Number of events	specifies the column that contains the number of events.
Number of trials	specifies the column that contains the number of trials.
Response	specifies the variable that contains response values. If you create a binomial response model, you can specify the first or last ordered category as the reference category by using the **Event of interest** option. You can also select a custom category. **Note:** This option is available only if you do not select the **Response data consists of numbers of events and trials** check box.
Options for All Distribution Types	

Option Name	Description
Response	specifies the variable that contains response values.
	If you create a binomial response model or a nominal multinomial model, you can specify the first or last ordered category as the reference category by using the **Event of interest** option. You can also select a custom category.
	■ To create a binomial response model, select **Binomial** as the distribution. For the binomial response model, specifying one response category as the reference is the same as specifying the other response category as the event category.
	■ To create a nominal multinomial model, select **Multinomial** as the distribution and select **Generalized logit** as the link function. For the generalized logit model, each logit contrasts a nonreference category with the reference category.

Option Name	Description
Link function	specifies the link function for your model. The functions that are available depend on the selected distribution.
	If you select **Default** for the link function, then the default link function for the model distribution is used.
	Here is the list of distributions with the corresponding default link function:
	■ Binomial distribution uses the logit link function.
	■ Gamma distribution uses the reciprocal link function.
	■ Inverse Gaussian distribution uses the reciprocal square link function.
	■ Multinomial distribution uses the cumulative logit link function.
	■ Negative binomial distribution uses the log link function.
	■ Normal distribution uses the identity link function.
	■ Poisson distribution uses the log link function.
	■ Tweedie distribution uses the log link function.
Explanatory Variables	
Classification variables	specifies the variables to use to group (classify) data in the analysis. Classification variables can be either character or numeric.
Parameterization of Effects	

Option Name	Description
Coding	specifies the parameterization method for the classification variable. Design matrix columns are created from the classification variables according to the selected coding scheme.
	You can select from these coding schemes:
	■ **GLM coding** specifies less-than-full-rank, reference-cell coding. This coding scheme is the default.
	■ **Reference coding** specifies reference-cell coding.

Treatment of Missing Values

An observation is excluded from the analysis when either of these conditions is met:

■ if any variable in the model contains a missing value

■ if any classification variable contains a missing value (regardless of whether the classification variable is used in the model)

Continuous variables	specifies the independent covariates (regressors) for the regression model. If you do not specify a continuous variable, the task fits a model that contains only an intercept.
Offset variable	specifies a variable to be used as an offset to the linear predictor. An offset plays the role of an effect whose coefficient is known to be 1. Observations that have missing values for the offset variable are excluded from the analysis.

Additional Roles

Frequency count	specifies the numeric column that contains the frequency of occurrence for each observation.
Weight variable	specifies the column to use as a weight to perform a weighted analysis of the data.

Building a Model

Requirements for Building a Model

By default, no effects are specified, which results in the task fitting an intercept-only model. To specify an effect, you must assign at least one variable to the **Classification variables** role or the **Continuous variables** role. You can select combinations of variables to create crossed, nested, factorial, or polynomial effects.

To create a model, use the model builder on the **Models** tab. After you create a model, you can specify whether to include the intercept in the model.

Create a Main Effect

1   Select the variable name in the **Variables** box.

2   Click **Add** to add the variable to the **Model effects** box.

Create Crossed Effects (Interactions)

1   Select two or more variables in the **Variables** box. To select more than one variable, press Ctrl.

2   Click **Cross**.

Create a Nested Effect

Nested effects are specified by following a main effect or crossed effect with a classification variable or list of classification variables enclosed in parentheses. The main effect or crossed effect is nested within the effects listed in parentheses. Here are examples of nested effects: B(A), C(B*A), D*E(C*B*A). In this example, B(A) is read "A nested within B."

1   Select the effect name in the **Model effects** box.

2   Click **Nest**. The Nested window opens.

3   Select the variable to use in the nested effect. Click **Outer** or **Nested within Outer** to specify how to create the nested effect.

> **Note:** The **Nested within Outer** button is available only when a classification variable is selected.

4   Click **Add**.

Create a Full Factorial Model

1   Select two or more variables in the **Variables** box.

2   Click **Full Factorial**.

For example, if you select the Height, Weight, and Age variables and then click **Full Factorial**, these model effects are created: Age, Height, Weight, Age*Height, Age*Weight, Height*Weight, and Age*Height*Weight.

Create an N-Way Factorial

1   Select two or more variables in the **Variables** box.

2   Click **N-way Factorial** to add these effects to the **Model effects** box.

For example, if you select the Height, Weight, and Age variables and then specify the value of N as 2, when you click **N-way Factorial**, these model effects are created: Age, Height, Weight, Age*Height, Age*Weight, and Height*Weight. If N is set to a value greater than the number of variables in the model, N is effectively set to the number of variables.

Create Polynomial Effects of the Nth Order

1   Select one variable in the **Variables** box.

2   Specify higher-degree crossings by adjusting the number in the **N** field.

3   Click **Polynomial Order=N** to add the polynomial effects to the **Model effects** box.

For example, if you select the Age and Height variables and then you specify 3 in the **N** field, when you click **Polynomial Order=N**, these model effects are created: Age, Age*Age, Age*Age*Age, Height, Height*Height, and Height*Height*Height.

Setting the Model Selection Options

Option	Description
Model Selection	
Selection method	specifies the selection method for the model. The task performs model selection by examining whether effects should be added to or removed from the model according to the rules that are defined by the selection method.
	Here are the valid values for the selection methods:
	■ **None** fits the full model.
	■ **Forward selection** starts with no effects in the model and adds effects based on the **Significance level to add an effect to the model** option.
Selection method (continued)	■ **Backward elimination** starts with all the effects in the model and deletes effects based on the value in the **Significance level to remove an effect from the model** option.
	■ **Stepwise selection** is similar to the forward selection model. However, effects that are already in the model do not necessarily stay there. Effects are added to the model based on the **Significance level to add an effect to the model** option and are removed from the model based on the **Significance level to remove an effect from the model** option.
Select best model by	specifies the criterion to use to identify the best-fitting model.
Details	

Option	Description
Selection process details	specifies how much information about the selection process to include in the results. You can display a summary, details for each step of the selection process, or all of the information about the selection process.
Maintain hierarchy of effects	specifies to maintain the hierarchy of effects.

Setting Options

Option	Description
Methods	
Dispersion	
Dispersion parameter	enables you to specify a fixed dispersion parameter for those distributions that have a dispersion parameter. By default, this parameter is estimated.
Optimization	
Method	specifies the optimization technique to use.
Maximum number of iterations	specifies the maximum number of iterations to perform for the selected optimization technique.
Statistics	

You can select the statistics to include in the output.

Here are the additional statistics that you can include:

- confidence limits for estimates
- correlations of parameter estimates
- covariances of parameter estimates

Setting the Output Options

You can specify whether to create an output data set. You can also specify whether to include predicted values, residuals, or any other variables in the output data set.

Replace Missing Values Task

About the Replace Missing Values Task

The Replace Missing Values task replaces missing values in a data set with an estimate of the missing value. The task also creates binary imputation indicators.

Assigning Data to Roles

Role	Description
Roles	
Replace missing values with the mean	replaces missing values with the mean for the variable.
Replace missing values with the pseudo-median	replaces missing values with the pseudo-median of the variable. If there is no nonmissing value, the pseudo-median is 0.
Replace missing values with a random number	replaces missing values with a random value that is drawn between the minimum and maximum of the variable. If there is no nonmissing value, the random value is 0.
Additional Roles	

Role	Description
Frequency count	specifies a numeric variable that contains the frequency of occurrence for each observation. If the frequency value is less than 1 or is missing, the observation is not used in the analysis. If no variable is assigned to the **Frequency count** role, each observation is assigned a frequency of 1.

Setting Options

You can specify whether to create an output data set. This output data set includes the data, imputation indicator variables (0 for not imputed or 1 for imputed), and imputed variables. You can also include any variables from the input data set.

Random Sampling Task

About the Random Sampling Task

The Random Sampling task is a high-performance procedure that performs either simple random sampling or stratified sampling. The output from this task includes an output data set and the sample data, a table with performance information, and a table with frequency information for the population and sample.

Assigning Data to Roles

If you want to perform stratified sampling, you must assign a column to the **Stratify by** role. Otherwise, the **Stratify by** role is optional.

Role	Description
Stratify by	specifies the variables to use to partition the input table into mutually exclusive, nonoverlapping subsets that are known as strata. Each stratum is defined by a set of values of the strata variables, and each stratum is sampled separately. The complete sample is the union of the samples that are taken from all the strata.
	Note: If you do not assign any variables to this role, then the entire input table is treated as a single stratum.
	You can allocate the total sample size among the strata in proportion to the size of the stratum. For example, the variable GENDER has possible values of M and F, and the variable VOTED has possible values of Y and N. If you assign both GENDER and VOTED to the **Stratify by** role, then the input table is partitioned into four strata: males who voted, males who did not vote, females who voted, and females who did not vote.
	The input table contains 20,000 rows, and the values are distributed as follows:
	■ 7,000 males who voted
	■ 4,000 males who did not vote
	■ 5,000 females who voted
	■ 4,000 females who did not vote

Role	Description
Stratify by (continued)	Therefore, the proportion of males who voted is 7,000/20,000=0.35 or 35%. The proportions in the sample should reflect the proportions of the strata in the input table. For example, if your sample table contains 100 observations, then 35% of the values in the sample must be selected from the males who voted stratum to reflect the proportions in the input table.

Creating the Output Data Set

You can select the numeric and character variables from the input data set to include in the output data. Select the **Include all input observations and a sampling indicator variable** to produce an output table with the same number of rows as the input table. The output table has an additional partition indicator (_PARTIND_) to indicate whether an observation is included in the sample (1) or not (0).

Setting Options

Option Name	Description
Methods	
Sample by	specifies the sample size in the desired number of rows or in the desired percentage of input rows. For example, if you specify 3% of rows and there are 400 input rows, then the resulting sample has 12 rows.
	Note: If you assign variables to the **Stratify by** role, then the sample size specification that you make here applies to each stratum rather than to the entire input table.

Option Name	Description
Random seed	specifies the initial seed for the generation of random numbers. If you set this value to zero or a negative number, then a seed that is based on the system clock is used to produce the sample.
Ignore case of character stratification values	distinguishes stratified variables that share the same normalized value when you perform stratified sampling. For example, if a target has three distinct values, "A", "B", and "b", and you want to treat "B" and "b" as different levels, you need to select this option. Otherwise, "B" and "b" are treated as the same level. The task normalizes a value as follows:

1 Leading blanks are removed.

2 The value is truncated to 32 characters.

3 Letters are changed from lowercase to uppercase.

12

Combinatorics and Probability Tasks

Same Birthday Probability Task

About the Same Birthday Probability Task

The Same Birthday Probability task computes the probability that two or more people in a room have the same birthday.

Example: Probability of Two or More People Sharing a Birthday in a Room of 145 People

To create this example:

1 In the **Tasks** section, expand the **Combinatorics and Probability** folder and double-click **Same Birthday Probability**. The user interface for the Same Birthday Probability task opens.

2 On the **Options** tab, specify these options.

Option Name	Value to Specify
Number in a room	`145`
Data set name	`Birthdays`

3 To run the task, click ![run icon].

Here is a subset of the results:

Same Birthday Probability

Number in a Room	Probability
2	0.00273972602740
3	0.00820416588478
4	0.01635591246655
5	0.02713557369979
6	0.04046248364911
7	0.05623570309598
8	0.07433529235167
9	0.09462383388917
10	0.11694817771108
11	0.14114137832173
12	0.16702478883806
13	0.19441027523243
14	0.22310251200497
15	0.25290131976369
16	0.28360400525285
17	0.31500766529656
18	0.34691141787179
19	0.37911852603154
20	0.41143838358058
21	0.44368833516521
22	0.47569530766255
23	0.50729723432399
24	0.53834425791453
25	0.56869970396946

Setting Options

All of these options are required to run the Same Birthday Probability task.

Option Name	Description
Observations	
Number in a room	specifies the number of people in the room.
Output Data Set	
Data set name	specifies the name of the output data set.

Coin Toss Simulation Task

About the Coin Toss Simulation Task

The Coin Toss Simulation task simulates the tossing of a specified number of coins. The results show the frequency and percentage of occurrences that the coin displays heads given a specified number of tosses.

Example: Probability of Outcomes for 10,000,000 Coin Tosses

To create this example:

1   In the **Tasks** section, expand the **Combinatorics and Probability** folder and double-click **Coin Toss Simulation**. The user interface for the Coin Toss Simulations task opens.

2   On the **Options** tab, specify these options.

Option Name	Value to Specify
Number of coins	`10`
Number of tosses	`10,000,000`
Data set name	`Coins`

3   To run the task, click ![runner icon].

Here are the results for this example:

Tossing 10 coins 10000000 times

Number of Heads	Frequency	Probability
0	9,724	0.000972
1	97,915	0.009792
2	439,805	0.043981
3	1,171,589	0.117159
4	2,050,285	0.205029
5	2,462,216	0.246222
6	2,049,961	0.204996
7	1,170,485	0.117049
8	440,501	0.044050
9	97,952	0.009795
10	9,567	0.000957
.	10,000,000	1.000000

Setting Options

All of these options are required to run the Coin Toss Simulation task.

Option Name	Description
Observations	
Number of coins	specifies the number of coins to toss.

Option Name	Description
Number of tosses	specifies the number of times to toss the coins.
Output Data Set	
Data set name	specifies the name of the output data set.

Dice Roll Simulation Task

About the Dice Roll Simulation Task

The Dice Roll Simulation task simulates rolling a specified number of dice. The results show the frequency and percentage of each possible roll given a specified number of throws.

Example: Probability of Outcomes for 100,000,000 Dice Rolls

To create this example:

1   In the **Tasks** section, expand the **Combinatorics and Probability** folder and double-click **Dice Roll Simulation**. The user interface for the Dice Roll Simulation task opens.

2   On the **Options** tab, specify these options.

Option Name	Value to Specify
Number of dice	2
Number of throws	`100,000,000`

Option Name	Value to Specify
Data set name	`Dice`

3   To run the task, click 🏃.

Here are the results:

Rolling 2 dice 100000000 times

Value Rolled	Frequency	Probability
2	2,776,093	0.027761
3	5,554,536	0.055545
4	8,335,512	0.083355
5	11,110,394	0.111104
6	13,888,936	0.138889
7	16,663,498	0.166635
8	13,891,321	0.138913
9	11,109,014	0.111090
10	8,336,640	0.083366
11	5,557,204	0.055572
12	2,776,852	0.027769
.	100,000,000	1.000000

Setting Options

All of these options are required to run the Dice Roll Simulation task.

Option Name	Description
Observations	
Number of dice	specifies the number of dice to roll.
Number of throws	specifies the number of times to roll the dice.
Output Data Set	
Data set name	specifies the name of the output data set.

Poker Hand Probability Task

About the Poker Hand Probability Task

The Poker Hand Probability task calculates the frequency and probability of poker hands. The input data set for the Poker Hand Probability task must be the output data set generated by the Computations task.

Example: Results from the Poker Hand Probability Task

To create this example:

1   In the **Tasks** section, expand the **Combinatorics and Probability** folder and double-click **Computations**. The user interface for the Computations task opens.

2   On the **Options** tab, specify these options.

Option Name	Value to Specify
Total number	52
Number in a set	5
Data set name	Combo

3   To run the task, click ![run icon]. The Work.Combo data set is created.

4   In the **Tasks** section, expand the **Combinatorics and Probability** folder and double-click **Poker Hand Probability**. The user interface for the Poker Hand Probability task opens.

5   For the input data set, select `WORK.COMBO`.

6   To run the task, click ![run icon] .

Here are the results:

Poker Hand Probability

Poker Hand	Frequency	Probability
Nothing	1,302,540	0.501177
One Pair	1,098,240	0.422569
Two Pair	123,552	0.047539
Three of a Kind	54,912	0.021128
Straight	10,200	0.003925
Flush	5,108	0.001965
Full House	3,744	0.001441
Four of a Kind	624	0.000240
Straight Flush	36	0.000014
Royal Flush	4	0.000002
Total	2,598,960	1.000000

13

Statistics Tasks

Data Exploration Task

About the Data Exploration Task

The Data Exploration task provides graphs that can be used to explore the relationships among selected variables.

Example: Exploring the SASHELP.CARS Data

To create this example:

1 In the **Tasks** section, expand the **Statistics** folder and double-click **Data Exploration**. The user interface for the Data Exploration task opens.

2 On the **Data** tab, select the **SASHELP.CARS** data set.

3 Assign columns to these roles and specify these options:

Role	Column Name
Continuous variables	**Horsepower** **MPG_City** **MPG_Highway**
Classification variables	**Type** **DriveTrain**

4 To run the task, click ![run icon].

Here is a subset of the results:

Scatterplot matrix of SASHELP.CARS data grouped by Type
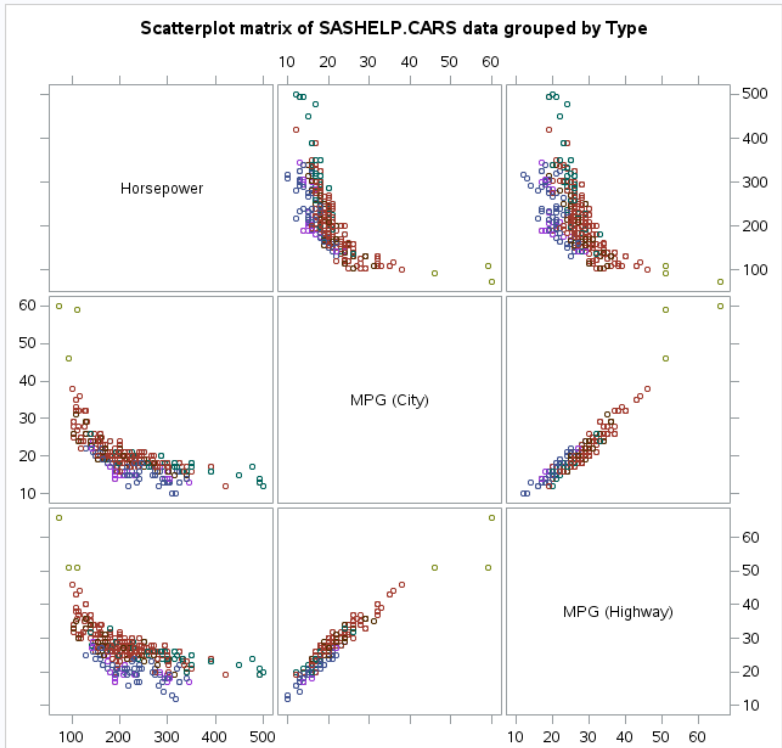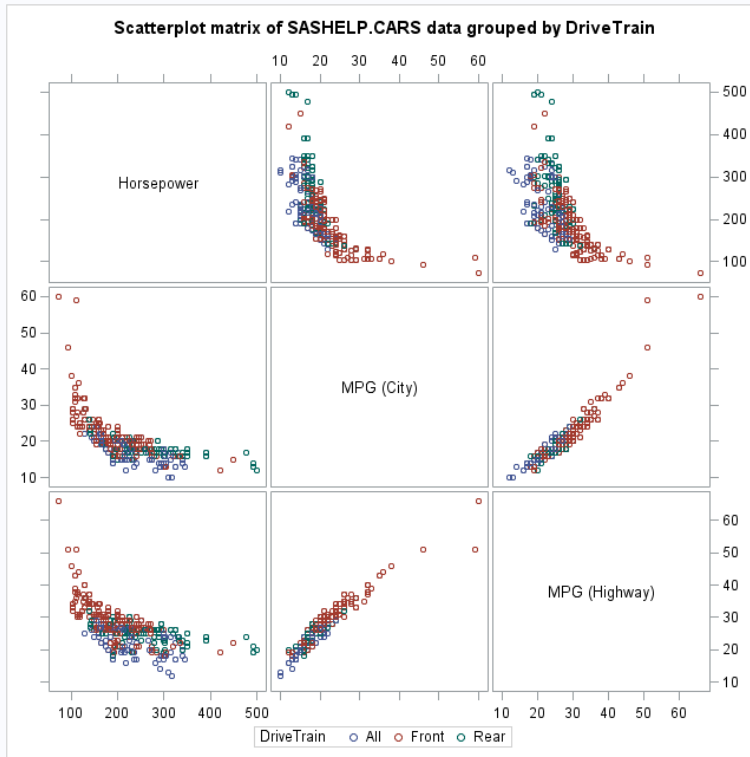
Assigning Data to Roles

To run the Data Exploration task, you must assign either two columns to the **Classification variables** role or one column to the **Continuous variables** role.

Role	Description
Roles	
Classification variables	specifies the classification variables to use to explore the data.
Continuous variables	specifies the continuous variables in the analysis.
Additional Roles	

Role	Description
Group analysis by	creates separate analyses based on the number of BY variables.

Setting the Plot Options

The plot options that are available depend on the columns that you assigned on the **Data** tab.

Option Name	Description
Histogram and Box Plot	
The combined histogram and box plot options are available when a column is assigned to the **Continuous variables** role, but no column is assigned to the **Classification variables** role.	
Scatter Plot Matrix	
The scatter plot matrix options are available when at least two columns are assigned to the **Continuous variables** role.	
Add histograms	adds histograms to the diagonal cells of the matrix. You can add a normal density curve and the kernel density estimate to these histograms.
Add prediction ellipses	adds a prediction ellipse to each cell that contains a scatter plot. You can specify the confidence level for the ellipses. Valid values are between 0 and 1.
Pairwise Scatter Plots	
The pairwise scatter plot options are available when at least two columns are assigned to the **Continuous variables** role.	

Option Name	Description
Pairwise scatter plots	plots the values of two or more variables and produces a separate cell for each combination of Y and X variables. That is, each Y*X pair is plotted on a separate set of axes.
Add a prediction ellipse	adds a prediction ellipse to each cell that contains a scatter plot. You can specify the confidence level for the ellipses. Valid values are between 0 and 1.

Regression Scatter Plots

The regression scatter plot options are available when at least two columns are assigned to the **Continuous variables** role.

Regression scatter plots	adds a regression fit to the scatter plot.
Select response variables	specifies the variables to use when fitting the regression line.
Add a fitted line	adds a regression fit to the scatter plot.
Add a loess fit	adds a loess fit to the scatter plot.
Add a fitted, penalized B-spline curve	adds a fitted, penalized B-spline curve to the scatter plot.

Mosaic Plot

Mosaic plot	creates a mosaic plot, which displays tiles that correspond to the crosstabulation table cells. The areas of the tiles are proportional to the frequencies of the table cells. The column variable is displayed on the X axis, and the tile widths are proportional to the relative frequencies of the column variable levels. The row variable is displayed on the Y axis, and the tile heights are proportional to the relative frequencies of the row levels within column levels.

Option Name	Description
Square mosaic plot	produces a square mosaic plot, where the height of the Y axis equals the width of the X axis. In a square mosaic plot, the scale of the relative frequencies is the same on both axes.
Specify colors of mosaic plot tiles	colors the mosaic plot tiles according to the values of residuals. You can also specify to color the tiles according to the Pearson or standardized residuals of the corresponding table cells.
Histogram	
Histogram	creates a histogram by using any numeric variables in the input data set.
Add normal density curve	adds a normal density curve to the histogram.
Add kernel density estimate	adds a kernel density estimate to the histogram.
Add inset statistics	adds a box or table of summary statistics directly in the histogram.
Box Plot	

The box plot options are available when at least one column is assigned to the **Classification variables** role.

Comparative box plot	creates a one-way box plot for each classification variable. This plot shows all continuous variables by the classification variable.

Summary Statistics Task

About the Summary Statistics Task

The Summary Statistics task provides descriptive statistics for variables across all observations and within groups of observations. You can also summarize your data in a graphical display, such as histograms and box plots.

For example, you could use this task to create a report on the number of new sales, arranged by product type and country.

Example: Summary Statistics of Unit Sales

In this example, you want to analyze unit sales. In addition to the tabular results, you choose to display a histogram of the distribution.

To create this example:

1 In the **Tasks** section, expand the **Statistics** folder and double-click **Summary Statistics**. The user interface for the Summary Statistics task opens.

2 On the **Data** tab, select the **SASHELP.PRICEDATA** data set.

3 To the **Analysis variables** role, assign the **sale** column.

4 On the **Options** tab, expand the **Plots** section and select the **Histogram** check box.

5 To run the task, click .

Here are the results:

Analysis Variable : sale Unit Sale				
Mean	Std Dev	Minimum	Maximum	N
408.5558824	73.0840041	203.0000000	747.0000000	1020

Distribution of sale
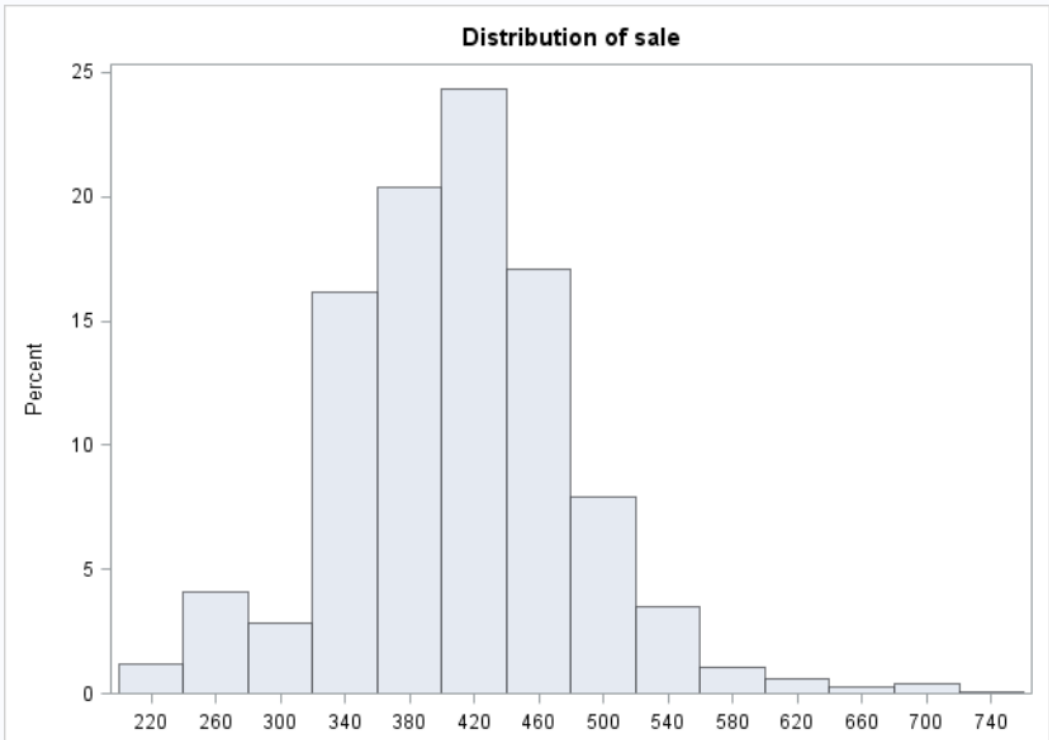
Assigning Data to Roles

To run the Summary Statistics task, you must assign a column to the **Analysis variables** role.

Role	Description
Roles	

Role	Description
Analysis variables	The variables that you assign to this role are the numeric variables for which you want statistics. You must assign at least one variable to this role.
Classification variables	The variables that you assign to this role are character or discrete numeric variables that are used to divide the input data into categories or subgroups. The statistics are calculated on all selected analysis variables for each unique combination of classification variables.
Additional Roles	
Group analysis by	The variables that you assign to this role are used to compute separate statistics for each distinct value or combination of values of the Group analysis by variables. The data is automatically sorted by the variables in this role before the statistics are computed.
Frequency count	When you assign a variable to this role, each observation in the table is assumed to represent *n* observations, where *n* is the value of the frequency count for that row. Statistics are calculated accordingly. You can assign a maximum of one variable to this role.
Weight variable	If you assign a variable to this role, the value of the variable for each observation is used to calculate weighted means, variances, and sums. You can assign a maximum of one variable to this role.

Setting Options

Option Name	Description
Statistics	
Basic Statistics	
Mean	is the arithmetic average, calculated by adding the values of an analysis variable and dividing this sum by the number of nonmissing observations.
Standard deviation	is a statistical measure of the variability of a group of data values. This measure, which is the most widely used measure of the dispersion of a frequency distribution, is equal to the positive square root of the variance.
Minimum value	is the smallest value for an analysis variable.
Maximum value	is the largest value for an analysis variable.
Median	is the middle value for an analysis variable.
Number of observations	is the total number of observations with nonmissing values.
Number of missing values	is the total number of observations with missing values.
Additional Statistics	
Standard error	is the standard deviation of the sample mean. The standard error is defined as the ratio of the sample standard deviation to the square root of the sample size.
	Note: This option is available only if **Degrees of freedom** is selected in the **Divisor for standard deviation and variance** drop-down list.

Option Name	Description
Variance	is a statistical measure of dispersion of data values. This measure is an average of the total squared dispersion between each observation and the sample mean.
Mode	is the most frequent value for the analysis variable.
Range	is the difference between the largest and smallest values in the data.
Sum	is the sum of all values in the analysis variable.
Sum of weights	is the sum of the numeric variable that is used to weight each observation. **Note:** You cannot compute the sum of the weights unless you assign a variable to the **Weight variable** role.
Confidence limits for the mean	is the two-sided confidence limits for the mean. A two-sided $100(1 - \alpha)\%$ confidence interval for the mean has the following upper and lower limits: $\bar{x} \pm t_{\left(1 - \frac{\alpha}{2}; n - 1\right)} \frac{s}{\sqrt{n}}$, where s is $\sqrt{\frac{1}{n-1} \Sigma (x_i - \bar{x})^2}$ and $t_{\left(1 - \frac{\alpha}{2}; n - 1\right)}$ is the $1 - \frac{\alpha}{2}$ of the Student's t statistics with $n - 1$ degrees of freedom.
Coefficient of variation	is a unitless measure of relative variability. This measure is defined as the ratio of the standard deviation to the mean expressed as a percentage. The coefficient of variation is meaningful only if the variable is measured on a ratio scale.
Skewness	is skewness, which measures the tendency of the deviations to be larger in one direction than in the other.

Option Name	Description
Kurtosis	is the kurtosis, which measures the heaviness of tails.
Percentiles	
1st, 5th, 10th, Lower quartile, Median, Upper quartile, 90th, 95th, 99th, Interquartile range	choose the percentiles and quantiles to compute.
Quantile method	specifies the method that is used to compute the quantiles, median, and percentiles. **Order statistics** reads all of the data into memory and sorts it by the unique values. **Piecewise-parabolic algorithm** approximates the quantile and is a less memory-intensive method. **Note:** If you assigned a variable to the **Weight variable** role, only the **Order statistics** method is available.
Plots	
Histogram	creates a graph that is used to determine the distribution of the data. If you add a normal density curve, the task uses the sample mean and sample standard deviation for μ and σ. If you add a kernel density curve, the task uses the AMISE method to compute the kernel density estimates. To include the statistics in the graph, select the **Add inset statistics** check box.

Option Name	Description
Comparative box plot	creates a graph that shows a measure of central location (the median), two measures of dispersion (the range and interquartile range), the skewness (from the orientation of the median relative to the quartiles), and potential outliers. Box plots are especially useful in comparing two or more sets of data.
	Note: The **Comparative box plot** option is available only when no column is assigned to the **Classification variable** role.
	You can choose to add the overall inset statistics to the graph or only the inset statistics for each group.
Histogram and box plot	displays the histogram and box plots together in a single panel, sharing common X axes. You can choose to add the overall inset statistics to the graph.
	Note: The **Histogram and box plot** option is available only when no column is assigned to the **Classification variable** role.

Details

Option Name	Description

Divisor for standard deviation and variance	specifies the divisor to use in the calculation of the variance and standard deviation. Here are the valid options:
	
	Degrees of freedom
	  $n - 1$
	
	  By default, the divisor for the variance is the degrees of freedom.
	
	Number of observations
	  n
	
	Sum of weights minus one
	  $(\Sigma_i w_i) - 1$
	
	Sum of weights
	  $\Sigma_i w_i$
	
	Note: The **Sum of weights minus one** and the **Sum of weights** options are available only if you assigned a variable to the **Weight variable** role.

Setting the Output Options

You can specify whether to save the statistics in an output data set.

Distribution Analysis Task

About the Distribution Analysis Task

Distribution analysis provides information about the distribution of numeric variables. A variety of plots such as histograms, probability plots, and quantile-quantile plots can be used in this analysis.

Example: Distribution Analysis of Sales for Each Region

In this example, you want to analyze the sales for each region. Because the data contains three regions, you get three sets of results.

To create this example:

1 In the **Tasks** section, expand the **Statistics** folder and double-click **Distribution Analysis**. The user interface for the Distribution Analysis task opens.

2 On the **Data** tab, select the **SASHELP.PRICEDATA** data set. Then assign the **sale** variable to the **Analysis variables** role.

3 Click the **Options** tab.

   a In the **Exploring Data** group, assign the **regionName** variable to the **Classification variables** role.

   b In the **Checking for Normality** group, select the **Histogram and goodness-of-fit tests** and **Normal quantile-quantile plot** options.

4 To run the task, click .

Here is a subset of the results:

Distribution of sale
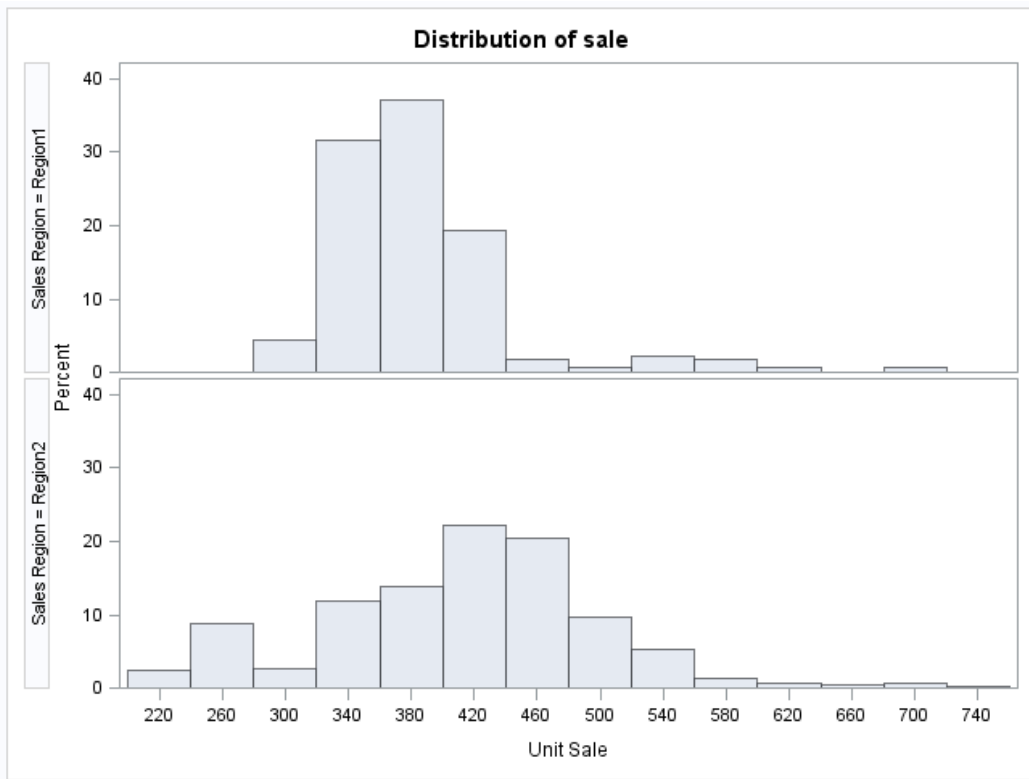
Fitted Normal Distribution for sale (Unit Sale)

Goodness-of-Fit Tests for Normal Distribution				
Test		Statistic	p Value	
Kolmogorov-Smirnov	D	0.04121623	Pr > D	<0.010
Cramer-von Mises	W-Sq	0.44027199	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	3.46017068	Pr > A-Sq	<0.005
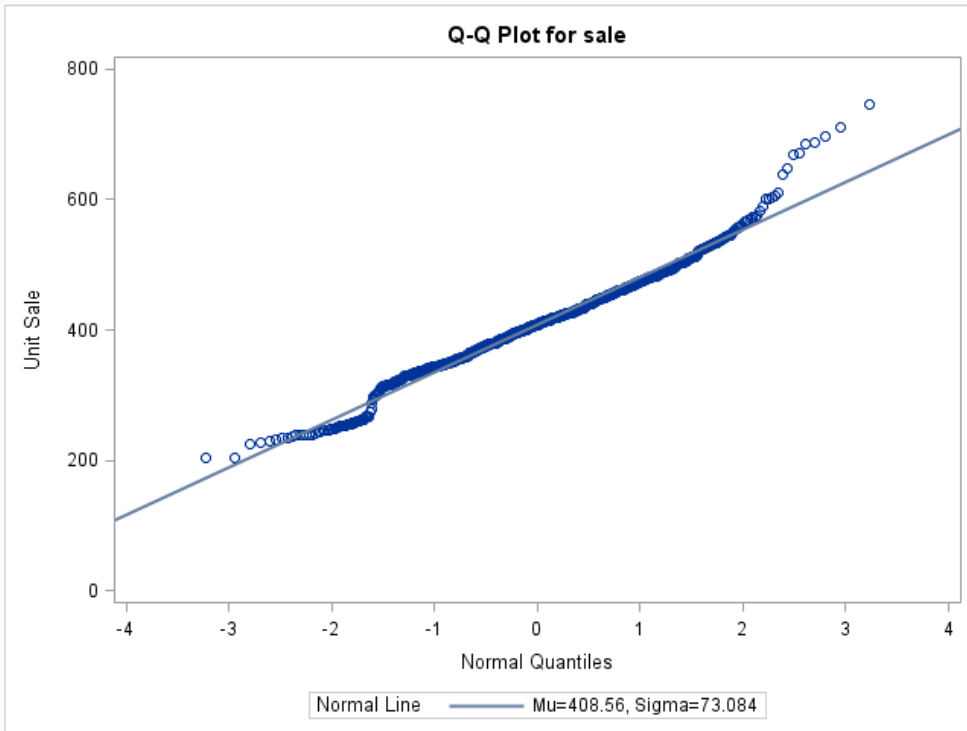
Assigning Data to Roles

To run the Distribution Analysis task, you must assign a column to the **Analysis variables** and select a plot or test on the **Options** tab.

Role	Description
Roles	
Analysis variables	specifies the analysis variables and their order in the results.

Role	Description
Additional Roles	
Frequency count	specifies a numeric variable whose value represents the frequency of the observation. The Distribution Analysis task assumes that each observation represents n observations, where n is the value of the variable.
Group analysis by	specifies the variables that the Distribution Analysis task uses to form groups.

Setting Options

Option Name	Description
Exploring Data	

By default, the task creates a histogram of the data. In the **Classification variables** role, specify the variables that are used to group the analysis variables into classification levels. You can assign a maximum of two columns to this role.

You can also specify whether to superimpose a kernel density estimate and the normal density curve on the histogram. Finally, you can specify whether to include an inset box of selected statistics in the graph.

Checking for Normality	

Note: If you select any of these options, you can also specify whether to include these inset statistics: number of observations, goodness-of-fit test, mean, median, standard deviation, variance, skewness, and kurtosis.

Histogram and goodness-of-fit tests	requests tests for normality that include a series of goodness-of-fit tests based on the empirical distribution function. The table provides test statistics and p-values for the Shapiro-Wilk test (provided the sample size is less than or equal to 2,000), the Kolmogorov-Smirnov test, the Anderson-Darling test, and the Cramér-von Mises test.

Option Name	Description
Normal probability plot	creates a probability plot, which compares ordered variable values with the percentiles of the normal distribution. If the data distribution matches the normal distribution, the points on the plot form a linear pattern. Probability plots are preferable for graphical estimation of percentiles.
	The distribution reference line on the plot is created from the maximum likelihood estimate for the parameter.
	You can also specify whether to include an inset box of selected statistics in the graph.
Normal quantile-quantile plot	creates quantile-quantile plots (Q-Q plots) and compares ordered variable values with quantiles of the normal distribution. If the data distribution matches the normal distribution, the points on the plot form a linear pattern. Q-Q plots are preferable for graphical estimation of distribution parameters.
	The distribution reference line on the plot is created from the maximum likelihood estimate for the parameter.
	You can also specify whether to include an inset box of selected statistics in the graph.
Fitting Distributions	
Note: If you select a plot option for any of these distributions, you can also specify whether to include these inset statistics: number of observations, mean, median, standard deviation, and variance.	
Beta	
Histogram and goodness-of-fit tests	fits beta distribution with threshold parameter θ, scale parameter σ, and shape parameters α and β.
Probability plot	specifies a beta probability plot for shape parameters α and β.

Option Name	Description
Quantile-quantile plot	specifies a beta Q-Q plot for shape parameters α and β.
Exponential	
Histogram and goodness-of-fit tests	fits exponential distribution with threshold parameter θ and scale parameter σ.
Probability plot	specifies an exponential probability plot.
Quantile-quantile plot	specifies an exponential Q-Q plot.
Gamma	
Histogram and goodness-of-fit tests	fits gamma distribution with threshold parameter θ, scale parameter σ, and shape parameter α.
Probability plot	specifies a gamma probability plot for shape parameter α.
Quantile-quantile plot	specifies a gamma Q-Q plot for shape parameter α.
Lognormal	
Histogram and goodness-of-fit tests	fits lognormal distribution with threshold parameter θ, scale parameter ζ, and shape parameter σ.
Probability plot	specifies a lognormal probability plot for shape parameter σ.
Quantile-quantile plot	specifies a lognormal Q-Q plot for shape parameter σ.
Weibull	
Histogram and goodness-of-fit tests	fits Weibull distribution with threshold parameter θ, scale parameter ζ, and shape parameter c.

Option Name	Description
Probability plot	specifies a two-parameter Weibull probability plot.
Quantile-quantile plot	specifies a two-parameter Weibull Q-Q plot.

One-Way Frequencies Task

About the One-Way Frequencies Task

The One-Way Frequencies task generates frequency tables from your data. You can also use this task to perform binomial and chi-square tests.

You might want to use this task to analyze the efficiency of a new drug. For example, suppose a group of medical researchers are interested in evaluating the efficacy of a new treatment for a skin condition. Dermatologists from participating clinics are trained to conduct the study and to evaluate the condition. After the training, two dermatologists examine patients with the skin condition from a pilot study and rate the same patients. The One-Way Frequencies task can be used to evaluate the agreement of the diagnoses.

Example: One-Way Frequencies of Unit Sales

In this example, you want to analyze unit sales for each sales region.

To create this example:

1 In the **Tasks** section, expand the **Statistics** folder and double-click **One-Way Frequencies**. The user interface for the One-Way Frequencies task opens.

2 On the **Data** tab, select the **SASHELP.PRICEDATA** data set.

3 Assign columns to these roles:

Role	Column Name
Analysis variables	**sale**
Group analysis by	**regionName**

4 To run the task, click 🏃.

Here is a subset of the results:

Frequency for SASHELP.PRICEDATA

Sales Region=Region1

		Unit Sale		
sale	Frequency	Percent	Cumulative Frequency	Cumulative Percent
298	1	0.56	1	0.56
300	1	0.56	2	1.11
301	1	0.56	3	1.67
307	1	0.56	4	2.22
308	1	0.56	5	2.78
314	1	0.56	6	3.33
316	1	0.56	7	3.89
318	1	0.56	8	4.44
320	1	0.56	9	5.00
321	1	0.56	10	5.56
322	2	1.11	12	6.67
323	1	0.56	13	7.22
324	2	1.11	15	8.33
328	1	0.56	16	8.89
331	3	1.67	19	10.56

Assigning Data to Roles

To run the One-Way Frequencies task, you must assign a column to the **Analysis variables** role.

Role	Description
Roles	
Analysis variables	specifies the variables to be analyzed. For each variable that you assign to this role, the task creates a one-way frequency table. You must assign at least one variable to this role.
Additional Roles	
Frequency count	specifies the variable to use as the frequency count. When you assign a variable to this role, each observation in the table is assumed to represent *n* observations. In this example, *n* is the value of the frequency count for that row. You can assign only one variable to this role.
Group analysis by	specifies one or more variables to sort the table by. Analyses are performed on each group.

Setting Options

Option Name	Description
Plots	

By default, plots are included in the results. Select the **Show frequencies table** check box to create the frequency and cumulative frequency plots. Select the **Asymptotic test** check box for the chi-square goodness-of-fit to create the deviation plot.

To suppress the plots from the results, select the **Suppress plots** check box.

Option Name	Description
Frequencies and Percentages	
Show frequencies table	specifies whether to create the frequencies table.
Include percentages	creates a table that contains the frequencies and percentages of total frequencies for each value of the analysis variable.
Include cumulative frequencies and percentages	creates a table that contains the frequencies and cumulative frequencies for each value of the analysis variable.

Statistics

Binomial Proportion

Specify whether to perform an asymptotic test. For binomial proportions, specify a null hypothesis proportion and a confidence level.

Chi-square goodness-of-fit

Specify whether to perform an asymptotic test.

To compute the Monte Carlo estimates of the exact *p*-values instead of directly computing the exact *p*-values, select the **Use Monte Carlo estimation** check box. Monte Carlo estimation can be useful for large problems that require a great amount of time and memory for exact computations but for which asymptotic approximations might be insufficient.

Exact Computations Methods

Note: This section appears if you select the **Exact test** check box for the binomial proportion or chi-square goodness-of-fit statistics.

Limit computation time	specifies the time limit (in seconds) for the computation of each *p*-value for each crosstabulation table. The default is 300 seconds (or 5 minutes).

Missing Values

Include in frequency table	includes missing values in the frequency tables.

Option Name	Description
Include in percentages and statistics	includes the frequencies of missing values in binomial or chi-square tests and in the calculations of percentages.

Correlation Analysis Task

About the Correlation Analysis Task

Correlation is a statistical procedure for describing the relationship between numeric variables. The relationship is described by calculating correlation coefficients for the variables. The correlations range from –1 to 1. The Correlation Analysis task provides graphs and statistics for investigating associations among variables.

Example: Correlations in the Sashelp.Cars Data Set

To create this example:

1 In the **Tasks** section, expand the **Statistics** folder and double-click **Correlation Analysis**. The user interface for the Correlation Analysis task opens.

2 On the **Data** tab, select the **SASHELP.CARS** data set.

3 Assign columns to these roles:

Role	Column
Analysis variables	**EngineSize** **Horsepower**
Correlate with	**Cylinders** **MPG_Highway**

4 To run the task, click 🏃.

Here are the results:

2 With Variables:	Cylinders MPG_Highway
2 Variables:	EngineSize Horsepower

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations	EngineSize	Horsepower
Cylinders	0.90800 <.0001 426	0.81034 <.0001 426
MPG_Highway MPG (Highway)	-0.71730 <.0001 428	-0.64720 <.0001 428

Assigning Data to Roles

To run the Correlation Analysis task, you must assign at least two columns to the **Analysis variables** role, or you must assign at least one column to the **Analysis variables** role and one column to the **Correlate with** role.

Roles	Description
Roles	
Analysis variables	lists the variables for which to compute correlation coefficients.
Correlate with	lists the variables with which the correlations of the analysis variables are to be computed.
Partial variables	removes the correlation of these variables from the analysis and correlates with variables before calculating the correlation.

Roles	Description
Additional Roles	
Frequency count	lists a numeric variable whose value represents the frequency of the observation. If you assign a variable to this role, the task assumes that each observation represents *n* observations, where *n* is the value of the frequency variable. If *n* is not an integer, SAS truncates it. If *n* is less than 1 or is missing, the observation is excluded from the analysis. The sum of the frequency variable represents the total number of observations.
Weight	lists the weights to use in the calculation of Pearson weighted product-moment correlation.
Group analysis by	enables you to obtain separate analyses of observations in groups that are defined by the BY variables.

Setting Options

Option Name	Description
Methods	
Missing values	specifies how to treat observations with missing values. If you select the **Use nonmissing values for all selected variables** option, all observations with missing values are excluded from the analysis. If you select the **Use nonmissing values for pairs of variables** option, the correlation statistics are computed using the nonmissing pairs of variables.
Statistics	

Option Name	Description

By default, the results contain a table with the correlations and *p*-values. You can also include these statistics:

Correlations
Selecting this option includes the correlations in the results. You can also specify probabilities that are associated with each correlation coefficient and whether to order the correlations from highest to lowest in absolute value.

Covariances
Selecting this option includes the variance and covariance matrix in the results. Also, the Pearson correlations are displayed. If you assign a column to the **Partial variables** role, the task computes a partial covariance matrix.

Sum of squares and cross-products
Selecting this option displays a table of the sums of squares and cross products in the results. The Pearson correlations are also included in the results. If you assign a column to the **Partial variables** role, the unpartial sums of squares and cross-products matrix is displayed.

Corrected sum of squares and cross-products
Selecting this option displays a table of the corrected sums of squares and cross products. The Pearson correlations are also included in the results. If you assign a column to the **Partial variables** role, the task computes both an unpartial and a partial corrected sum of squares and cross-products matrix.

Descriptive statistics
Selecting this option includes the simple descriptive statistics for each variable. Even if you do not select this option and you choose to create an output data set, the data set contains the descriptive statistics for the variables.

Fisher's z transformation
For a Pearson correlation, you can use the Fisher transformation options to request confidence limits and *p*-values under a specified alternative (null) hypothesis, $H_0: \rho = \rho_0$, for correlation coefficients that use Fisher's *z* transformation. If you select the **Fisher's z transformation** check box, you must specify a value in the **Null hypothesis** box.

You can choose from these types of confidence limits:

- **Two-sided confidence limits** requests two-sided confidence limits for the test of the null hypothesis, $H_0: \rho = \rho_0$. This is the default.

- **Lower confidence limit** requests a lower confidence limit for the test of the one-sided null hypothesis, $H_0: \rho \leq \rho_0$.

- **Upper confidence limit** requests an upper confidence limit for the test of the one-sided null hypothesis, $H_0: \rho \geq \rho_0$.

By default, the level of the confidence limits for the correlation is 95%.

Option Name	Description
Nonparametric Correlations	
Spearman's rank-order correlation	calculates Spearman rank-order correlation. This is a nonparametric measure of association that is based on the rank of the data values. The correlations range from –1 to 1.
Kendall's tau-b	calculates Kendall tau-b. This is a nonparametric measure of association that is based on the number of concordances and discordances in paired observations. Concordance occurs when paired observations vary together, and discordance occurs when paired observations vary differently. Kendall's tau-b ranges from –1 to 1.
Hoeffding's measure of dependence	calculates Hoeffding's measure of dependence, D. This is a nonparametric measure of association that detects more general departures from independence. This D statistic is 30 times larger than the usual definition and scales the range between –0.5 and 1 so that only large positive values indicate dependence.
Plots	

You can include either of these plots in your results:

- a scatter plot matrix for variables. You can also choose to include a histogram of the analysis variables in the symmetric matrix plot.

- a scatter plot for each applicable pair of distinct variables from the analysis variables. You can specify whether to display the prediction ellipses for new observations or the confidence ellipses for the mean.

You can also specify the number of variables to plot and the maximum number of points to plot.

Setting the Output Options

You can specify whether to create an output data set that contains the Pearson correlation statistics. This data set also includes means, standard deviations, and the number of observations.

You can also choose to include these statistics in the output data set:

■ **Correlations** – By default, the output data set contains the correlation coefficients with the corresponding _TYPE_ variable value of 'CORR'.

■ **Covariances** – When you select this option, the output data set contains the covariance matrix with the corresponding _TYPE_ variable value of 'COV'.

■ **Sum of squares and cross-products** – If you assign a column to the **Partial variables** role, the output data set does not contain a sum of squares and cross-products matrix.

■ **Corrected sum of squares and cross-products** — If you assign a column to the **Partial variables** role, the output data set contains a partial corrected sum of squares and cross-products matrix.

Table Analysis Task

About the Table Analysis Task

The Table Analysis task provides one-way to n-way frequency and contingency (crosstabulation) tables. This task also generates statistics about the association between rows and columns.

Example: Distribution of Type by DriveTrain

To create this example:

1   In the **Tasks** section, expand the **Statistics** folder and double-click **Table Analysis**. The user interface for the Table Analysis task opens.

2   On the **Data** tab, select the **SASHELP.CARS** data set.

3   Assign columns to these roles:

Role	Column
Row variable	**Type**
Column variable	**DriveTrain**

4   To run the task, click 🏃.

Here is a sample of the results:

Frequency	Table of Type by DriveTrain				
		DriveTrain			
	Type	All	Front	Rear	Total
	Hybrid	0	3	0	3
	SUV	38	22	0	60
	Sedan	28	179	55	262
	Sports	5	8	36	49
	Truck	12	0	12	24
	Wagon	9	14	7	30
	Total	92	226	110	428
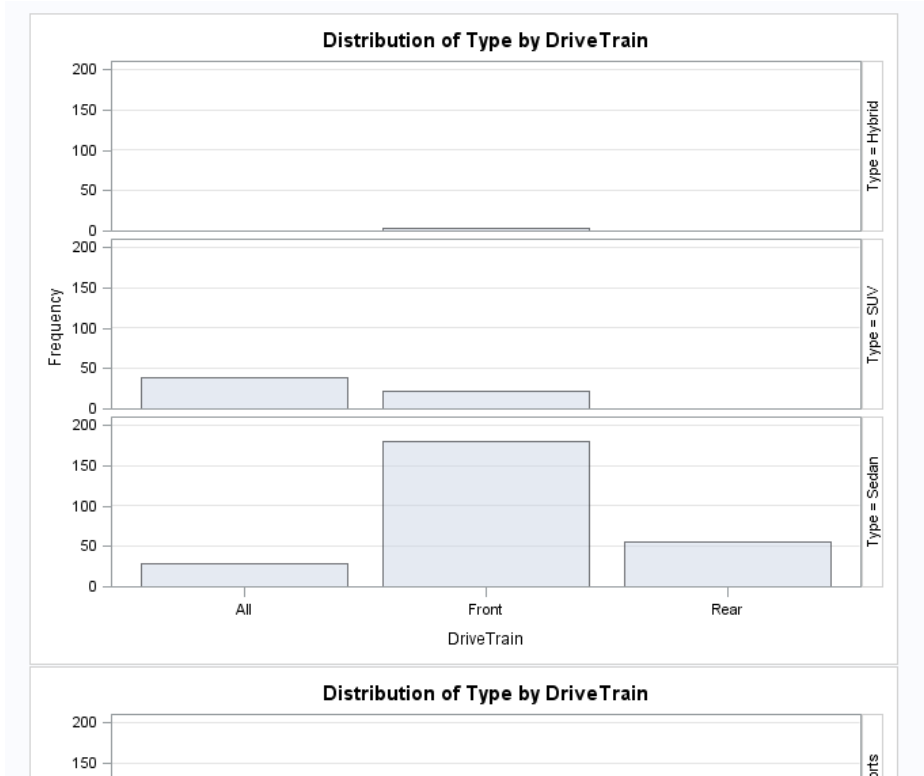
Assigning Data to Roles

To run the Table Analysis task, you must first assign at least one column to the **Row variables** or **Column variables** roles.

Roles	Description
Roles	
Row variables	specifies the row for one-way table analysis. If multiple variables are assigned to this role, the task performs multiple one-way table analyses.

Roles	Description
Column variables	creates the columns for one-way table analysis. If only column variables are assigned, the task performs multiple one-way table analyses.
Strata variables	creates the separate tables for n-way frequency and crosstabulation tables.
	Note: You must assign columns to both the **Row variables** or **Column variables** roles to use a strata variable.
Additional Roles	
Frequency count	specifies that each row in the table is assumed to represent *n* observations. In this example, *n* is the value of the frequency count for that observation.

Setting Options

Option Name	Description
Plots	

By default, plots are included in the results. To suppress these plots, select the **Suppress plots** check box.

Frequency Table	
Frequencies	
Observed	displays the frequency count for each cell.
Expected	displays the expected cell frequency for each cell.
Deviation	displays the deviation of the cell frequency from the expected value for each cell.

Option Name	Description
Percentages	
Cell	display of overall percentages in crosstabulation tables.
Row	display of row percentages in crosstabulation table cells.
Column	display of column percentages in crosstabulation table cells.
Cumulative	
Column percentages	displays the cumulative column percentage in each cell.
Frequencies and percentages	displays the cumulative frequencies and percentages in one-way frequency tables.
Cell contributions to the chi-square statistics	displays each table cell's contribution to the Pearson chi-square statistic in the crosstabulation table.
Statistics	
Chi-square statistics	requests chi-square tests of homogeneity or independence and measures of association that are based on the chi-square statistic. The tests include the Pearson chi-square, likelihood-ratio chi-square, and Mantel-Haenszel chi-square. For 2×2 tables, this test includes Fisher's exact test and the continuity-adjusted chi-square.
Measures of association	computes several measures of association and their asymptotic standard errors (ASE). The measures include gamma, Kendall's tau-*b*, Stuart's tau-*c*, Somers' D (*C*|*R*), Somers' D (*R*|*C*), the Pearson and Spearman correlation coefficients, lambda (symmetric and asymmetric), and uncertainty coefficients (symmetric and asymmetric).

Option Name	Description
Cochran-Mantel-Haenszel statistics	requests Cochran-Mantel-Haenszel statistics, which test for association between the row and column variables after adjusting for the remaining variables in a multiway table. These statistics include the CMH correlation statistic, the row mean scores (ANOVA), and the adjusted relative risks and odds ratios.
Measures of agreement (for square tables)	computes tests and measures of classification agreement for square tables. This option provides McNemar's test for 2×2 tables and Bowker's test of symmetry for tables with more than two response categories. It also produces the simple kappa coefficient, the weighted kappa coefficient, the asymptotic standard errors for the simple and weighted kappas, and the corresponding confidence limits. When there are multiple strata and two response categories, this option also computes Cochran's Q test.
Odds ratio and relative risk (for 2x2 tables)	requests relative risk measures and their asymptotic Walk confidence limits for 2x2 tables.
Binomial proportions and risk differences (for 2x2 tables)	requests risks (binomial proportions) and risk differences for 2x2 tables.
Exact Test	
Fisher's exact test	requests Fisher's exact test for tables that are larger than 2x2.
Methods	

Option Name	Description
Missing value treatment	specifies how to treat missing values:
	Exclude missing values specifies that an observation is excluded from a table if the observation has a missing value for any of the variables.
	Display missing value frequencies displays the frequencies of the missing values in the frequency and crosstabulation tables. These frequencies are not included in any computations of percentages, tests, or measures.
	Include missing values in calculations treats the missing values as valid for all variables.

T Tests Task: One-Sample t Test

About the One-Sample t Test Task

A one-sample *t* test compares the mean of the sample to the null hypothesis mean.

To compare an individual mean with a sample size of *n* to a value *m*, use $t = \dfrac{\bar{x} - m}{\frac{s}{\sqrt{n}}}$

where $\bar{x}$ is the sample mean of the observations and s^2 is the sample variance of the observations.

For example, you want to perform a one-sample *t* test on the horsepower values in the Sashelp.Cars data set. The null hypothesis is 300.

To run a one-sample *t* test, open the T Tests task. From the **T test** drop-down list, select **One-sample test**.

Example: One-Sample t Test for Horsepower

To create this example:

1   In the **Tasks** section, expand the **Statistics** folder and double-click **T Tests**. The user interface for the T Tests task opens.

2   On the **Data** tab, select the **SASHELP.CARS** data set.

3   From the **T test** drop-down list, select **One-sample test**.

4   To the **Analysis variable** role, assign the **Horsepower** column.

5   On the **Options** tab, enter 300 in the **Alternative hypothesis** field.

6   To run the task, click .

Here is a subset of the results:

Variable: Horsepower

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.949922	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.090516	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.589806	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	3.685805	Pr > A-Sq	<0.0050

Variable: Horsepower

N	Mean	Std Dev	Std Err	Minimum	Maximum
428	215.9	71.8360	3.4723	73.0000	500.0

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
215.9	209.1	222.7	71.8360	67.3244	77.0007

DF	t Value	Pr > |t|
427	-24.22	<.0001
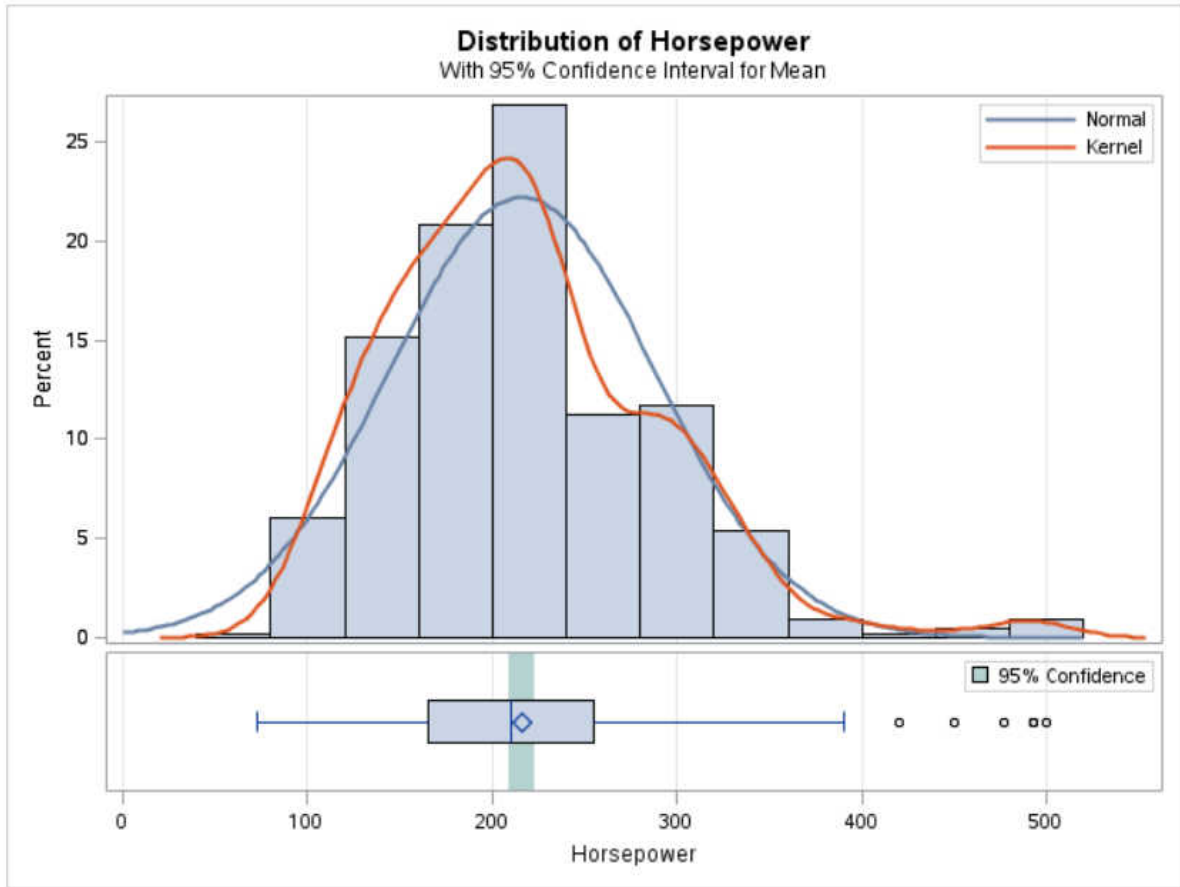
Assigning Data to Roles

To run a one-sample *t* test, select **One-sample test** from the **T test** drop-down list. Assign a numeric column to the **Analysis variable** role.

Setting Options

Option Name	Description
Test	

Option Name	Description
Tails	specifies the number of sides (or tails) and direction of the statistical tests and test-based confidence intervals. You can choose from these options:
	■ **Two-tailed test** specifies two-sided tests and confidence intervals for means.
	■ **Upper one-tailed test** specifies upper one-sided tests in which the alternative hypothesis indicates a mean greater than the null value, and upper one-sided confidence intervals between the lower confidence limit and infinity.
	■ **Lower one-tailed test** specifies lower one-sided tests in which the alternative hypothesis indicates a mean less than the null value, and lower one-sided confidence intervals between minus infinity and the upper confidence limit.
Alternative hypothesis	specifies the value of the null hypothesis. By default, the null hypothesis has a value of 0.
Normality Assumption	
Tests for normality	runs tests for normality that include a series of goodness-of-fit tests based on the empirical distribution function. The table provides test statistics and *p*-values for the Shapiro-Wilk test (provided the sample size is less than or equal to 2000), the Kolmogorov-Smirnov test, the Anderson-Darling test, and the Cramér-von Mises test.
Nonparametric Tests	

Note: This option is available only for a two-tailed test.

Option Name	Description
Sign test and Wilcoxon signed rank test	generates the results from these tests:
	▪ The sign test statistic is $M = (n^+ - n^-)/2$, where n^+ is the number of values that are greater than μ_0, and n^- is the number of values that are less than μ_0. Values equal to μ_0 are discarded.
	▪ The Wilcoxon signed rank statistic S is calculated as $$S = \sum_{i:\,|x_i - \mu_0| > 0} r_i^+ - \frac{n_t(n_t + 1)}{4}$$, where r^+_i is the rank of $x_i - \mu_0$ after discarding values of $x_i - \mu_0$, and n_t is the number of x_i values not equal to μ_0. Average ranks are used for tied values.
Plots	
Histogram and box plot	creates a histogram and box plot together in a single panel, sharing common X axes.
Normality plot	creates a normal quantile-quantile (Q-Q) plot.
Confidence interval plot	creates a plot of the confidence interval for the means.

T Tests Task: Paired t Test

About the Paired t Test Task

A paired *t* test compares the mean of the differences in the observations to a given number, the null hypothesis difference. The paired *t* test is used when the two samples are correlated, such as two measures of blood pressure from the same person.

To compare *n* paired differences to a value *m*, use $t = \dfrac{\bar{d} - m}{\frac{s_d}{\sqrt{n}}}$, where $\bar{d}$ is the sample

mean of the paired differences and $s^2{}_d$ is the sample variance of the paired differences.

To run a paired *t* test, open the T Tests task. From the **T test** drop-down list, select **Paired test**.

Example: Determining the Distribution of Price – Cost

In this example, you want to compare the means of differences in price and cost in the Sashelp.Pricedata data set. The null hypothesis for this test is 30.

To create this example:

1 In the **Tasks** section, expand the **Statistics** folder and double-click **T Tests**. The user interface for the T Tests task opens.

2 On the **Data** tab, select the **SASHELP.PRICEDATA** data set.

3 From the **T test** drop-down list, select **Paired test**.

4 Assign columns to these roles:

Role	Column Name
Group 1 variable	price
Group 2 variable	cost

5 On the **Options** tab, enter `30` in the **Alternative** field.

6 To run the task, click 🏃.

Here is a subset of the results:

Variable: _Difference_ (Difference: price - cost)

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.896986	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.1888	Pr > D	<0.0100
Cramer-von Mises	W-Sq	7.159388	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	39.28743	Pr > A-Sq	<0.0050

Difference: price - cost

N	Mean	Std Dev	Std Err	Minimum	Maximum
1020	42.0448	21.9813	0.6883	6.5700	93.4000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
42.0448	40.6942	43.3954	21.9813	21.0671	22.9791

DF	t Value	Pr > |t|
1019	17.50	<.0001

Distribution of Difference: price - cost
With 95% Confidence Interval for Mean
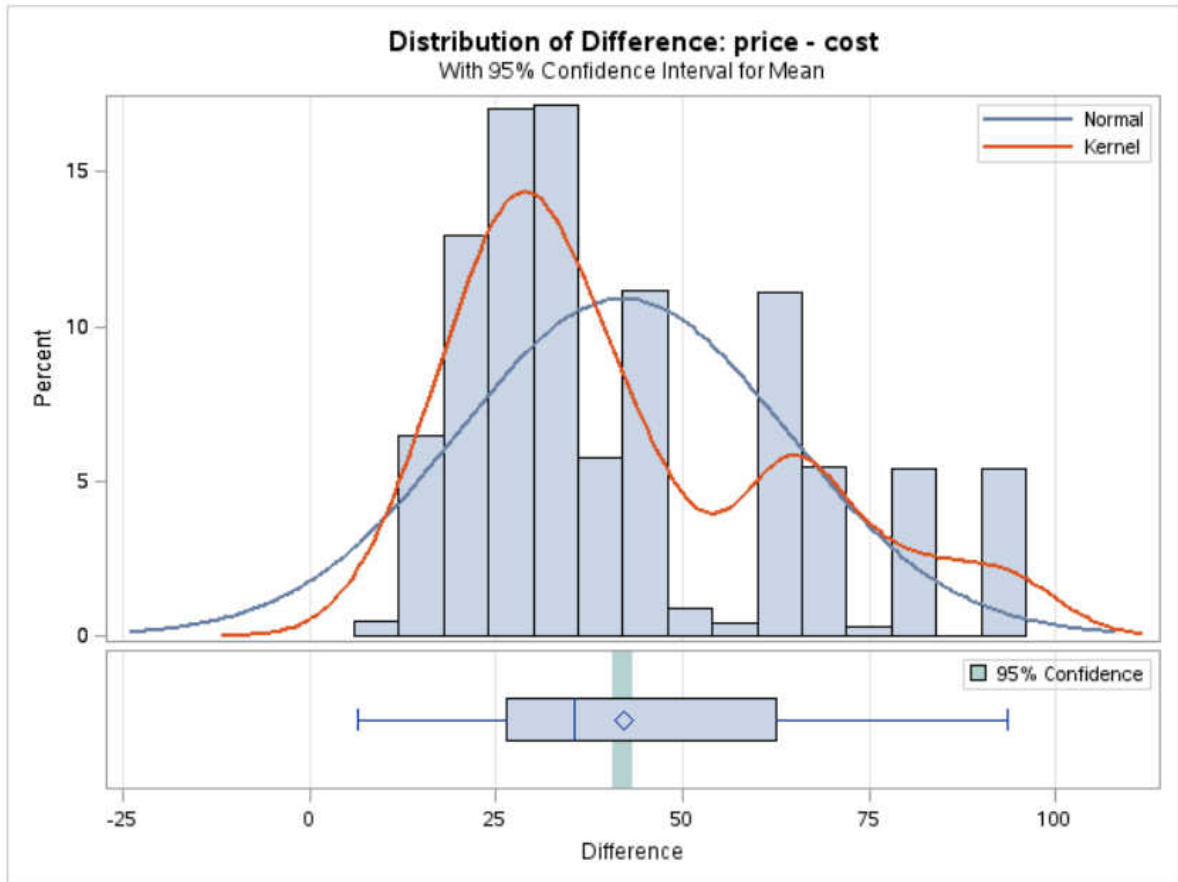
Assigning Data to Roles

To run a paired *t* test, select **Paired test** from the **T test** drop-down list. Assign columns to the **Group 1 variable** and **Group 2 variable** roles. The task compares these two variables. Because paired *t* tests are performed by subtracting each value of the **Group 2 variable** from the corresponding value of the **Group 1 variable**, the designation of the variables matters.

Setting Options

Option Name	Description
Test	
Tails	specifies the number of sides (or tails) and direction of the statistical tests and test-based confidence intervals. You can choose from these options:
	■ **Two-tailed test** specifies two-sided tests and confidence intervals for means.
	■ **Upper one-tailed test** specifies upper one-sided tests in which the alternative hypothesis indicates a mean greater than the null value. The upper one-sided confidence intervals range between the lower confidence limit and infinity.
	■ **Lower one-tailed test** specifies lower one-sided tests in which the alternative hypothesis indicates a mean less than the null value. The lower one-sided confidence intervals range between minus infinity and the upper confidence limit.
Alternative	specifies the value of the null hypothesis.
Normality Assumption	
Tests for normality	runs tests for normality that include a series of goodness-of-fit tests based on the empirical distribution function. The table provides test statistics and p-values for the Shapiro-Wilk test (provided the sample size is less than or equal to 2000), the Kolmogorov-Smirnov test, the Anderson-Darling test, and the Cramér-von Mises test.
Nonparametric Tests	

Note: This option is available only for a two-tailed test.

Option Name	Description
Sign test and Wilcoxon signed rank test	generates the results from these tests:
	▪ The sign test statistic is $M = (n^+ - n^-)/2$, where n^+ is the number of values that are greater than μ_0, and n^- is the number of values that are less than μ_0. Values equal to μ_0 are discarded.
	▪ The Wilcoxon signed rank statistic S is calculated as $$S = \sum_{i: \lvert x_i - \mu_0 \rvert > 0} r_i^+ - \frac{n_t(n_t + 1)}{4}$$, where r^+_i is the rank of $x_i - \mu_0$ after discarding values of $x_i - \mu_0$ and n_t is the number of x_i values not equal to μ_0. Average ranks are used for tied values.
Plots	
Histogram and box plot	creates a histogram and box plot together in a single panel, sharing common X axes.
Normality plot	creates a normal quantile-quantile (Q-Q) plot.
Agreement plot	plots the second response in each pair against the first response. The mean is shown as a large bold symbol. A diagonal line with slope=0 and y-intercept=1 is overlaid. The location of the points with respect to the diagonal line reveals the strength and direction of the difference or ratio. The tighter the clustering along the same direction as the line, the stronger the positive correlation of the two measurements for each subject. Clustering along a direction perpendicular to the line indicates negative correlation.

Option Name	Description
Response profile plot	creates a plot where a line is drawn for each observation from left to right that connects the first response to the second response. The mean first response and mean second response are connected with a bold line. The more extreme the slope, the stronger the effect. A wide spread of profiles indicates high between-subject variability. Consistent positive slopes indicate strong positive correlation. Widely varying slopes indicate lack of correlation. Consistent negative slopes indicate strong negative correlation.
Confidence interval plot	creates a plot of the confidence interval for the means.

T Tests Task: Two-Sample t Test

About the Two-Sample t Test Task

A two-sample *t* test compares the mean of the first sample minus the mean of the second sample to a given number, the null hypothesis difference.

To compare means from two independent samples with n_1 and n_2 observations to a value *m*, use $t = \dfrac{(\overline{x}_1 - \overline{x}_2) - m}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$. In this example, s^2 is the pooled variance

$s^2 = \dfrac{(n_1 - 1)s_1^2 + (n_1 - 1)s_2^2}{n_1 + n_2 - 2}$, and $s^2{}_1$ and $s^2{}_2$ are the sample variances of the two groups. The use of this *t* statistic depends on the assumption that $\sigma_1^2 = \sigma_2^2$, where σ_1^2 and σ_2^2 are the population variances of the two groups.

To run a two-sample *t* test, open the T Tests task. From the **T test** drop-down list, select **Two-sample test**.

Example: Two-Sample t Test

In this example, you want to analyze the height values for males and females in your class.

To create this example:

1   In the **Tasks** section, expand the **Statistics** folder and double-click **T Tests**. The user interface for the T Tests task opens.

2   On the **Data** tab, select the **SASHELP.CLASS** data set.

3   From the **T test** drop-down list, select **Two-sample test**.

4   Assign columns to these roles:

Role	Column Name
Analysis variable	**Height**
Groups variable	**Sex**

5   To run the task, click .

Here is a subset of the results:

Variable: Height
Sex = F

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.931231	Pr < W	0.4932
Kolmogorov-Smirnov	D	0.203889	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.049919	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.308707	Pr > A-Sq	>0.2500

Variable: Height
Sex = M

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.954758	Pr < W	0.7249
Kolmogorov-Smirnov	D	0.139972	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.025769	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.202753	Pr > A-Sq	>0.2500

Distribution of Height
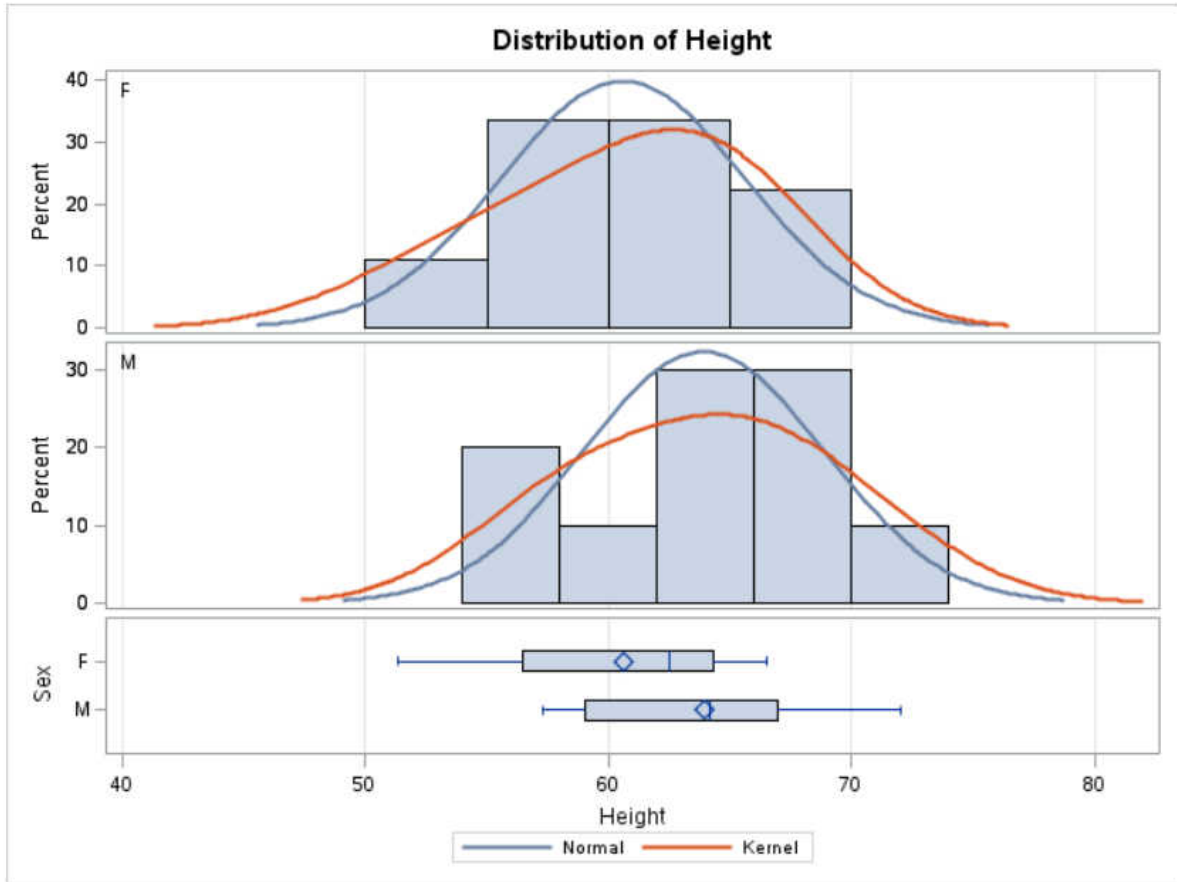
Assigning Data to Roles

To run a two-sample t test, select **Two-sample test** from the **T test** drop-down list.
Assign a column to each of these roles:

Role	Description
Analysis variable	specifies the column to use in the analysis.
Groups variable	specifies the column to use for grouping. This column must have only two levels.

Setting Options

Option Name	Description
Test	
Tails	specifies the number of sides (or tails) and direction of the statistical tests and test-based confidence intervals. You can choose from these options:
	■ **Two-tailed test** specifies two-sided tests and confidence intervals for means.
	■ **Upper one-tailed test** specifies upper one-sided tests in which the alternative hypothesis indicates a mean greater than the null value, and upper one-sided confidence intervals between the lower confidence limit and infinity.
	■ **Lower one-tailed test** specifies lower one-sided tests in which the alternative hypothesis indicates a mean less than the null value, and lower one-sided confidence intervals between minus infinity and the upper confidence limit.
Alternative hypothesis	specifies the value of the null hypothesis.

Option Name	Description
Cox and Cochran probability approximation for unequal variances	calculates the Cochran and Cox approximation. This approximation of the p-value of the t_u is the value of p such that $$t_u = \frac{\left(\dfrac{\dfrac{s_1^2}{n_1^*}}{\sum\limits_{i=1} f_{1i} w_{1i}}\right) t_1 + \left(\dfrac{\dfrac{s_2^2}{n_2^*}}{\sum\limits_{i=1} f_{2i} w_{2i}}\right) t_2}{\left(\dfrac{\dfrac{s_1^2}{n_1^*}}{\sum\limits_{i=1} f_{1i} w_{1i}}\right) + \left(\dfrac{\dfrac{s_2^2}{n_2^*}}{\sum\limits_{i=1} f_{2i} w_{2i}}\right)}$$. In this example, t_1 and t_2 are the critical values of the t distribution corresponding to a significance level of p and sample sizes n_1 and n_2, respectively. The degrees of freedom is undefined when $n_1 \neq n_2$. (Cochran and Cox 1950).
Normality Assumption	
Tests for normality	runs tests for normality that include a series of goodness-of-fit tests based on the empirical distribution function. The table provides test statistics and p-values for the Shapiro-Wilk test (provided the sample size is less than or equal to 2000), the Kolmogorov-Smirnov test, the Anderson-Darling test, and the Cramér-von Mises test.
Nonparametric Tests **Note:** This option is available only for a two-tailed test when the alternative hypothesis equals 0.	
Wilcoxon rank-sum test	generates an analysis of Wilcoxon scores. When there are two classification levels (samples), this option produces the Wilcoxon rank-sum test.

Option Name	Description
Plots	
Histogram and box plot	creates a histogram and box plot together in a single panel, sharing common X axes.
Normality plot	creates a normal quantile-quantile (Q-Q) plot.
Confidence interval plot	creates plots of the confidence interval for means. This plot is not created by default.
Wilcoxon box plot	creates a box plot of Wilcoxon scores. This plot is associated with the Wilcoxon analysis. This plot is not created by default.
	Note: This plot is available only for a two-tailed test when the alternative hypothesis equals 0.

One-Way ANOVA Task

About the One-Way ANOVA Task

The one-way analysis of variance (ANOVA) task tests and provides graphs for differences among the means of a single categorical variable on a single continuous dependent variable.

You might use the One-Way ANOVA task to do the following:

- study the effect of bacteria on the nitrogen content of red clover plants. The factor is the bacteria strain, and it has six levels.

- compare the life spans of three different brands of batteries. The factor is the brand, and it has three levels.

Example: Testing for Differences in the Means for MPG_Highway by Car Type

In this example, you want to study the differences in the means for the number of highway miles per gallon for six car types.

To create this example:

1   In the **Tasks** section, expand the **Statistics** folder and double-click **One-Way ANOVA**. The user interface for the One-Way ANOVA task opens.

2   On the **Data** tab, select the **SASHELP.CARS** data set.

3   Assign columns to these roles:

Role	Column Name
Dependent variable	**MPG_Highway**
Categorical variable	**Type**

4   To run the task, click .

Here is a subset of the results:

Class Level Information		
Class	**Levels**	**Values**
Type	6	Hybrid SUV Sedan Sports Truck Wagon

Number of Observations Read	428
Number of Observations Used	428

Dependent Variable: MPG_Highway MPG (Highway)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	6743.47900	1348.69580	77.64	<.0001
Error	422	7331.03268	17.37212		
Corrected Total	427	14074.51168			

R-Square	Coeff Var	Root MSE	MPG_Highway Mean
0.479127	15.52701	4.167987	26.84346

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Type	5	6743.478998	1348.695800	77.64	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Type	5	6743.478998	1348.695800	77.64	<.0001

Distribution of MPG_Highway
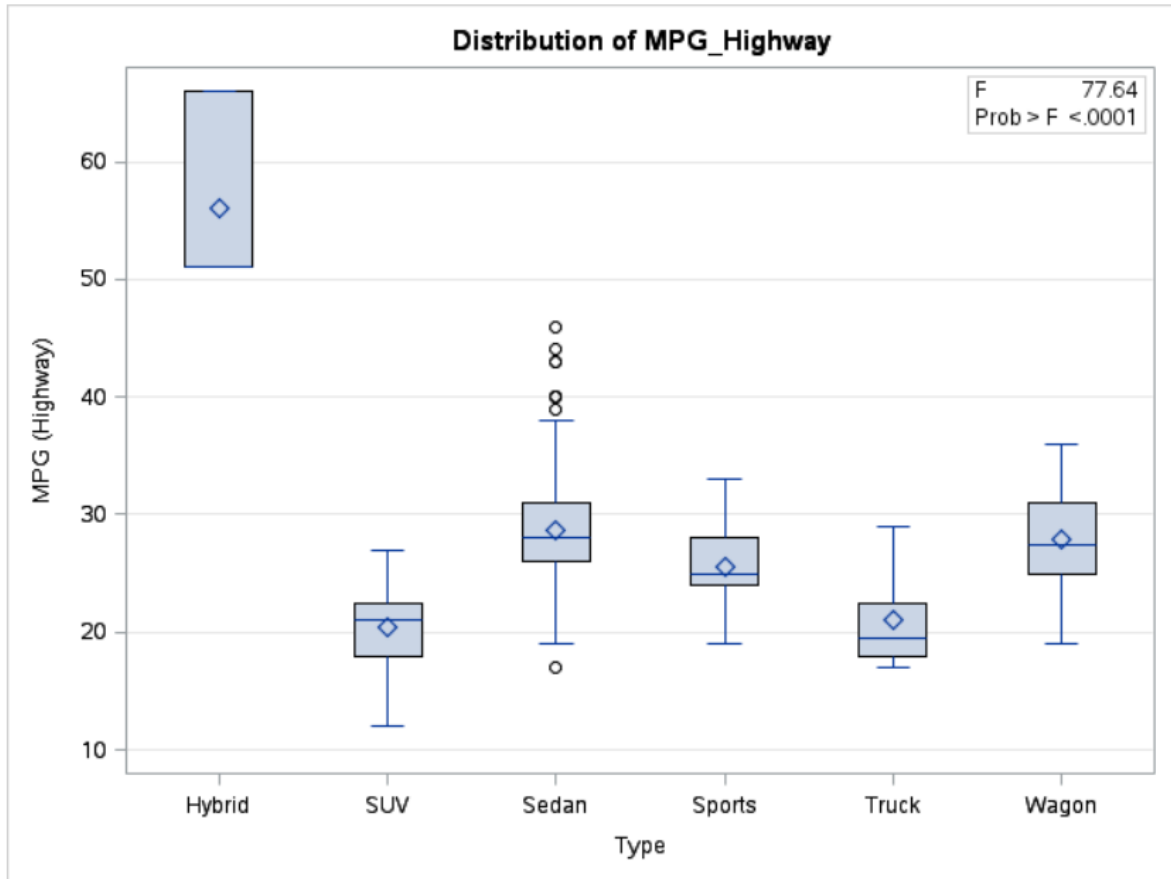
Assigning Data to Roles

To run the One-Way ANOVA task, you must assign columns to these roles:

Role Name	Description
Dependent variable	specifies a continuous numeric column.
Categorical variable	specifies a character or numeric column with values that specify the levels of the groups. The column that you assign to this role must have two or more distinct values.

Setting Options

Option Name	Description
Normality Assumption	
Tests for normality	runs tests for normality that include a series of goodness-of-fit tests based on the empirical distribution function. The table provides test statistics and *p*-values for the Shapiro-Wilk test (provided the sample size is less than or equal to 2,000), the Kolmogorov-Smirnov test, the Anderson-Darling test, and the Cramér-von Mises test.
Homogeneity of Variance	

Test	specifies the type of test to perform. Here are the valid values:
	
	None
	    specifies that no test is performed.
	
	Bartlett
	    computes accurate Type I error rates when the distribution of the data is normal.

Option Name	Description
Test (continued)	**Brown & Forsythe** is a variation of Levene's test. Equal variances are determined by using the absolute deviations from the group medians. Although this is a good test for determining variance differences, it can be resource intensive if your data contains several large groups. **Levene** computes the squared residuals to determine equal variance. Levene's test is considered to be the standard homogeneity of variance test. This is the default. **O'Brien** specifies O'Brien's test, which is a modification of Levene's test that uses squared residuals.
Welch's variance-weighted ANOVA	tests the group means by using a weighted variance. You can use this test if the assumption of equal variances is rejected.
Comparisons	

Option Name	Description

You can select from these comparison methods:

Bonferroni
performs Bonferroni *t* tests of differences between means for all means of the main effect.

Duncan multiple range
performs Duncan's multiple range test on all means of the main effect.

Dunnett two-tail
performs Dunnett's two-tailed *t* test, testing whether any treatments are significantly different from a single control for all main-effect means.

Dunnett lower one-tail
performs Dunnett's one-tailed *t* test, testing whether any treatment is significantly less than the control.

Dunnett upper one-tail
performs Dunnett's one-tailed *t* test, testing whether any treatment is significantly greater than the control.

Gabriel
performs Gabriel's multiple-comparison procedure on all means of the main effect.

Nelson
analyzes all the differences with the least squares means.

Ryan-Einot-Gabriel-Welsch
performs the Ryan-Einot-Gabriel-Welsch multiple range test on all means of the main effect.

Scheffé
performs Scheffé's multiple-comparison procedure on all means of the main effect.

Sidak
performs pairwise *t* tests on differences between means with levels adjusted according to Sidak's inequality for all means of the main effect.

Student-Newman-Keuls
performs the Student-Newman-Keuls multiple range test on all main effect means.

Least significant difference (LSD)
performs pairwise *t* tests for all means of the main effect. In the case of equal cell sizes, this test is equivalent to Fisher's least significant difference test.

Tukey
performs Tukey's studentized range test (HSD) on all means of the main effect. When the group sizes are different, this is the Tukey-Kramer test.

You can also specify the level of significance for the selected test.

Option Name	Description
Plots	

By default, the results include a box plot, a means plot, and a least squares mean difference plot. You can also specify to include any diagnostic plots, which can be displayed in a panel or as individual plots.

You can also specify the maximum number of points to include in these plots.

Setting the Output Options

You can specify whether to create an output data set. You can also specify the values to include in the output data set. You can include predicted values, residuals, standard errors, and influence statistics.

Nonparametric One-Way ANOVA Task

About the Nonparametric One-Way ANOVA Task

The Nonparametric One-Way ANOVA task consists of nonparametric tests for location and scale differences across a one-way classification. The task also provides a standard analysis of variance on the raw data and statistics based on the empirical distribution function.

Example: Wilcoxon Scores for MPG_Highway Classified by Origin

To create this example:

1 In the **Tasks** section, expand the **Statistics** folder and double-click **Nonparametric One-Way ANOVA**. The user interface for the Nonparametric One-Way ANOVA task opens.

2 On the **Data** tab, select the **SASHELP.CARS** data set.

3 Assign columns to these roles:

Role	Column Name
Dependent variable	**MPG_Highway**
Classification variable	**Origin**

4 To run the task, click .

Wilcoxon Scores (Rank Sums) for Variable MPG_Highway Classified by Variable Origin

Origin	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
Asia	158	37704.0	33891.00	1231.66801	238.632911
Europe	123	24687.0	26383.50	1155.00991	200.707317
USA	147	29415.0	31531.50	1211.97891	200.102041
Average scores were used for ties.					

Kruskal-Wallis Test

Chi-Square	9.5856
DF	2
Pr > Chi-Square	0.0083

Distribution of Wilcoxon Scores for MPG_Highway
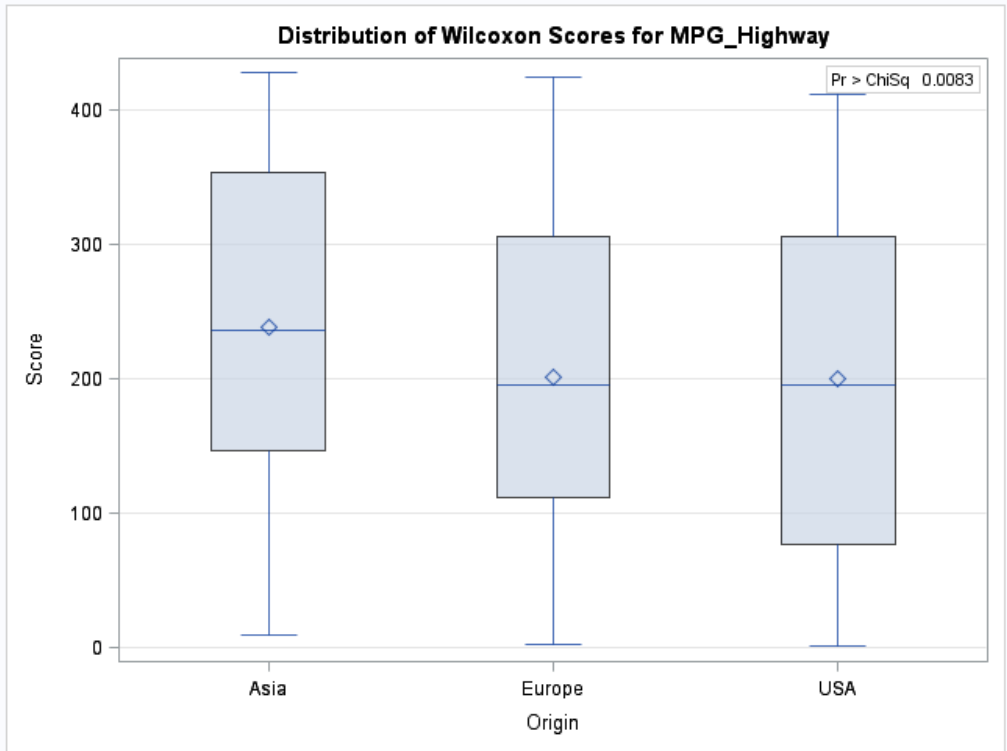
Assigning Data to Roles

To run the Nonparametric One-Way ANOVA task, you must assign columns to the **Dependent variable** and **Classification variable** roles.

Role Name	Description
Roles	
Dependent variable	specifies the column to use as the dependent variable.
Classification variable	defines the subgroups. Separate analyses are performed for each subgroup. You can specify whether to treat missing values as a valid level.
Additional Roles	
Frequency count	specifies that each row in the table is assumed to represent n observations. In this example, n is the value of the frequency count for that observation.
Group analysis by	sorts the table by these columns. The task performs analyses on each group.

Setting Options

Option Name	Description
Plots	

Option Name	Description

By default, plots are included in the results. These plots are determined by the options that you select. Here are some of the plots that you can create:

- By selecting the options in the **Location Differences** section, you can create a box plot of Wilcoxon scores, a stacked bar chart showing frequencies above or below the overall median, a box plot of Van der Waerden scores, and a box plot of Savage scores.
- By selecting the options in the **Scale Differences** section, you can create a box plot of Ansari-Bradley scores, a box plot of Klotz scores, a box plot of Mood scores, and a box plot of Siegel-Tukey scores.
- By selecting the options in the **Location and Scale Differences** section, you can create a box plot of Conover scores.
- By selecting the **Empirical distribution function tests, including Kolmogorov-Smirnov and Cramer-von Mises tests** option, you can create a plot of the empirical distribution test.

You can specify whether to display the *p*-values in the plot.

To suppress the plots from the results, select the **Suppress plots** check box.

Tests	
Tests	specifies whether to calculate only the asymptotic tests or both the asymptotic tests and exact tests for the various analyses.

Location Differences	
Wilcoxon scores	ranks of the observations.
Median scores	equals 1 for observations greater than the median and 0 otherwise.
Van der Waerden scores	the quantiles of a standard normal distribution. These scores are also known as quantile normal scores.
Savage scores	the expected values of order statistics from the exponential distribution with 1 subtracted to center the scores around 0.

Scale Differences	
Ansari-Bradley scores	similar to the Siegel-Tukey scores, but assigns the same scores to corresponding extreme ranks.
Klotz scores	the squares of the Van der Waerden (or quantile normal) scores.

Option Name	Description
Mood scores	the square of the difference between each rank and the average rank.
Siegel-Tukey scores	scores are computed as $a(1) = 1,\ a(n) = 2,\ a(n-1) = 3,\ a(2) = 4,\ a(3) = 5,\ a(n-2) = 6,\ ....$ The score values continue to increase in this pattern toward the middle ranks until all observations are assigned a score.
Location and Scale Differences	
Conover scores	based on the squared ranks of the absolution deviations from the sample means.
Additional Tests	
Empirical distribution function tests, including Kolmogorov-Smirnov and Cramer-von Mises tests	the empirical distribution function (EDF) statistics.
Pairwise multiple comparison analysis (asymptotic only)	computes the Dwass, Steel, Critchlow-Fligner (DSCF) multiple comparison analyses.
Methods	
Continuity Correction	
Continuity correction for two sample Wilcoxon and Siegel-Tukey tests	uses a continuity correction for the asymptotic two-sample Wilcoxon and Siegel-Tukey tests by default. The task incorporates this correction when computing the standardized test statistic *z* by subtracting 0.5 from the numerator $(S - E_0(S))$ if it is greater than zero. If the numerator is less than zero, the task adds 0.5.
Exact Statistics Computation	

Option Name	Description
Use Monte Carlo estimation	requests the Monte Carlo estimation of the exact *p*-values instead of using the direct exact *p*-value computation. You can also specify the level of the confidence limits for the Monte Carlo *p*-value estimates.
Limit computation time	specifies the time limit for calculating each exact *p*-value. Calculating exact *p*-values can consume a large amount of time and memory.

Creating an Output Data Set

You can specify whether to save the statistics to an output data set.

N-Way ANOVA Task

About the N-Way ANOVA Task

The N-Way ANOVA task tests and provides graphs for effects of one or more factors on the means of a single, continuous dependent variable.

Example: Analyzing the Sashelp.RevHub2 Data Set

To create this example:

1 In the **Tasks** section, expand the **Statistics** folder and double-click **N-Way ANOVA**. The user interface for the N-Way ANOVA task opens.

2 On the **Data** tab, select the **SASHELP.REVHUB2** data set.

3 Assign variables to these roles:

Role	Column Name
Dependent variable	Revenue
Factors	Source Type

4 On the **Model** tab, select **Source** and **Type.** Click **Full Factorial**.

5 To run the task, click ![run icon].

Here is a subset of the results:

Class Level Information

Class	Levels	Values
SOURCE	4	Freight Other Passenger Service
TYPE	3	Direct Indirect Other

Number of Observations Read	72
Number of Observations Used	72

Dependent Variable: REVENUE REVENUE

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	9.7259489E13	8.8417717E12	30.35	<.0001
Error	60	1.7480696E13	291344930348		
Corrected Total	71	1.1474018E14			

R-Square	Coeff Var	Root MSE	REVENUE Mean
0.847650	81.22575	539763.8	664523.0

Source	DF	Type I SS	Mean Square	F Value	Pr > F
SOURCE	3	3.2657409E13	1.0885803E13	37.36	<.0001
TYPE	2	3.1959368E13	1.5979684E13	54.85	<.0001
SOURCE*TYPE	6	3.2642711E13	5.4404519E12	18.67	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SOURCE	3	3.2657409E13	1.0885803E13	37.36	<.0001
TYPE	2	3.1959368E13	1.5979684E13	54.85	<.0001
SOURCE*TYPE	6	3.2642711E13	5.4404519E12	18.67	<.0001

Assigning Data to Roles

To run the N-Way ANOVA task, you must assign columns to the **Dependent variable** and **Factors** roles.

Role	Description
Dependent variable	specifies the dependent variable.
Factors	specifies the classification variables.

Building a Model

Requirements for Building a Model

By default, no effects are specified, which results in the task fitting an intercept-only model. To run the task, you must assign at least two variables to the **Factors** role. You can select combinations of variables to create crossed, nested, factorial, or polynomial effects.

To create a model, use the model builder on the **Model** tab. After you create your model, you can specify whether to include the intercept in the model.

Create a Main Effect

1   Select the variable name in the **Variables** box.

2   Click **Add** to add the variable to the **Model effects** box.

Create Crossed Effects (Interactions)

1   Select two or more variables in the **Variables** box. To select more than one variable, press Ctrl.

2   Click **Cross**.

Create a Nested Effect

Nested effects are specified by following a main effect or crossed effect with a classification variable or list of classification variables enclosed in parentheses. The main effect or crossed effect is nested within the effects listed in parentheses. Here are examples of nested effects: B(A), C(B*A), D*E(C*B*A). In this example, B(A) is read "A nested within B."

1 Select the effect name in the **Model effects** box.

2 Click **Nest**. The Nested window opens.

3 Select the variable to use in the nested effect. Click **Outer** or **Nested within Outer** to specify how to create the nested effect.

   Note: The **Nested within Outer** button is available only when a classification variable is selected.

4 Click **Add**.

Create a Full Factorial Model

1 Select two or more variables in the **Variables** box.

2 Click **Full Factorial**.

For example, if you select the Height, Weight, and Age variables and then click **Full Factorial**, these model effects are created: Age, Height, Weight, Age*Height, Age*Weight, Height*Weight, and Age*Height*Weight.

Create an N-Way Factorial

1 Select two or more variables in the **Variables** box.

2 Click **N-way Factorial** to add these effects to the **Model effects** box.

For example, if you select the Height, Weight, and Age variables and then specify the value of N as 2, when you click **N-way Factorial**, these model effects are created: Age, Height, Weight, Age*Height, Age*Weight, and Height*Weight. If N is set to a value greater than the number of variables in the model, N is effectively set to the number of variables.

Setting Options

Option	Description
Statistics	

You can choose to display only the default statistics, the default statistics and additional statistics, or no statistics in the output.

Here are the options for the additional statistics:

■ **Perform multiple comparisons** computes the least squares means for the specified effects. You can specify the method for adjustments for the *p*-values and confidence limits for the differences of the least squares means.

■ The **Sum of Squares** options enable you to display the sum of squares associated with Type I estimable functions for each effect and the sum of squares associated with Type III estimable functions for each effect.

Plots	

You can choose to display only the default plots, only selected plots, or no plots in your output. You can specify the maximum number of points to display in the plots.

Here are some plots that you can include in your results:

■ least squares means plot

■ mean difference plot

■ interaction plot (available only if there are two variables assigned to the **Factors** role)

■ analysis of means plot (available only if you select the **Nelson** method for adjustment)

■ diagnostic plots, which can be displayed individually or in a panel

Setting the Output Options

You can specify whether to create an output data set. You can also specify the values to include in the output data set. You can include predicted values, residuals, standard errors, and influence statistics.

Analysis of Covariance Task

About the Analysis of Covariance Task

The Analysis of Covariance task fits a linear model that combines the continuous and categorical predictors of a continuous dependent variable. This task also produces graphical output to interpret the results.

Example: Analyzing the Sashelp.Class Data Set

To create this example:

1 In the **Tasks** section, expand the **Statistics** folder and double-click **Analysis of Covariance**. The user interface for the Analysis of Covariance task opens.

2 On the **Data** tab, select the **SASHELP.CLASS** data set.

3 Assign variables to these roles:

Role	Column Name
Dependent variable	**Height**
Categorical variable	**Sex**
Continuous variables	**Weight**

4 To run the task, click ![run icon].

Here is a subset of the results:

Class Level Information		
Class	Levels	Values
Sex	2	F M

Number of Observations Read	19
Number of Observations Used	19

Dependent Variable: Height

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	369.1762676	123.0587559	17.75	<.0001
Error	15	103.9879429	6.9325295		
Corrected Total	18	473.1642105			

R-Square	Coeff Var	Root MSE	Height Mean
0.780229	4.223778	2.632970	62.33684

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Sex	1	52.2463216	52.2463216	7.54	0.0150
Weight*Sex	2	316.9299459	158.4649730	22.86	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Sex	1	2.7552010	2.7552010	0.40	0.5379
Weight*Sex	2	316.9299459	158.4649730	22.86	<.0001

Analysis of Covariance for Height
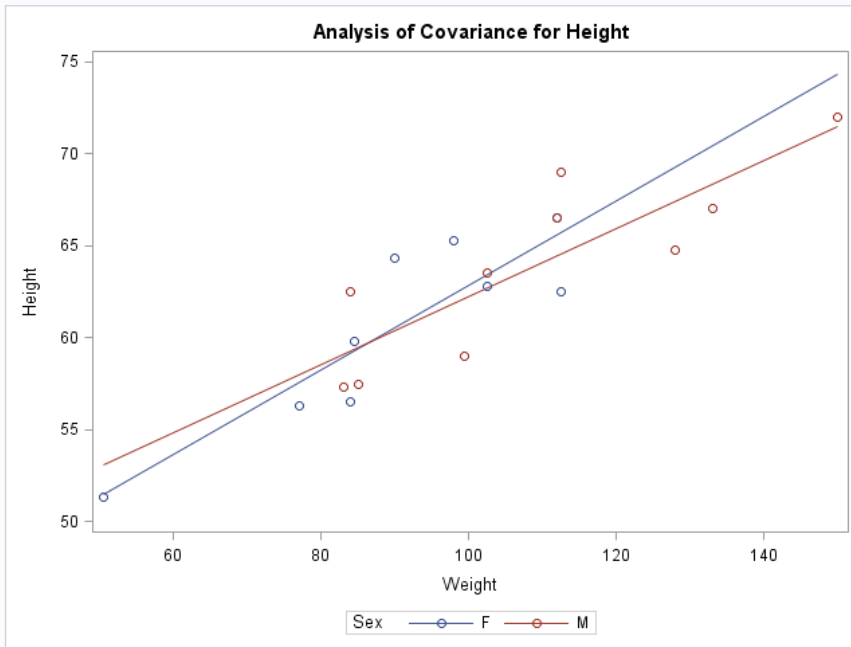
Assigning Data to Roles

To run the Analysis of Covariance task, you must assign columns to the **Dependent variable**, **Categorical variable**, and **Continuous covariate** roles.

Role	Description
Dependent variable	specifies a continuous numeric variable.
Categorical variable	specifies a character or numeric variable that specifies the levels of the groups.
Continuous covariate	specifies a continuous numeric variable that is related to the dependent variable. You can specify whether to center the covariate variable.

Setting Options

Option	Description
Model	
Intercepts	specifies whether to use the equal or unequal intercepts for each level of the categorical variable.
Slopes	specifies whether to use the equal or unequal slopes for each level of the categorical variable.
Show parameter estimates	produces a solution to the normal equations (parameter estimates). By default, the task displays a solution if your model does not include any classification variables. Select this option only if you want to see the solution for models with classification effects.
Multiple Comparisons	
Perform multiple comparisons	performs the least squares means for the categorical variable.
Covariate value	specifies the value to use in multiple comparisons. The covariate value can be either the mean value or a specified value.
Method	requests a multiple comparison adjustment for the *p*-values and confidence limits for the differences of LS-means. Here are the available methods: ■ Bonferroni ■ Dunnett ■ Nelson ■ Scheffe ■ Sidak ■ Tukey

Option	Description
Significance level	specifies the significance level for the comparisons. The default is 0.05.

Plots

You can choose to display only the default plots in your output, select the plots to display in the output, or display no plots in the output. The list of available plots depends on the method that you selected for multiple comparisons.

Setting the Output Options

You can specify whether to create an output data set. You can also specify the values to include in the output data set. You can include predicted values, residuals, standard errors, and influence statistics.

Linear Regression Task

About the Linear Regression Task

The Linear regression task fits a linear model to predict a single continuous dependent variable from one or more continuous or categorical predictor variables. This task produces statistics and graphs for interpreting the results.

Example: Predicting Weight Based on a Student's Height

In this example, you want to use regression analysis to find out how well you can predict a child's weight if you know the child's height.

To create this example:

1   In the **Tasks** section, expand the **Statistics** folder and double-click **Linear Regression**. The user interface for the Linear Regression task opens.

2 On the **Data** tab, select the **SASHELP.CLASS** data set.

3 Assign columns to these roles:

Role	Column Name
Dependent variable	**Weight**
Classification variables	**Sex**
Continuous variables	**Age** **Height**

4 Click the **Model** tab and create these models:

a Select the **Height** variable, and then press Ctrl and select the **Age** variable. Click **Add**.

b Select the **Height** variable, and then press Ctrl and select the **Sex** variable. Click **Cross**.

5 To run the task, click 🏃.

Here is a subset of the results:

Data Set	SASHELP.CLASS
Dependent Variable	Weight
Selection Method	None

Number of Observations Read	19
Number of Observations Used	19

Class Level Information		
Class	Levels	Values
Sex	2	F M

Dimensions	
Number of Effects	4
Number of Parameters	5

Least Squares Summary				
Step	Effect Entered	Number Effects In	Number Parms In	SBC
0	Intercept	1	1	120.6906
1	Age	2	2	108.5093
2	Height	3	3	98.4141*
3	Height*Sex	4	4	98.6423
	* Optimal Value of Criterion			

Least Squares Model (No Selection)

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	7498.06138	2499.35379	20.40	<.0001
Error	15	1837.67546	122.51170		
Corrected Total	18	9335.73684			

Root MSE	11.06850
Dependent Mean	100.02632
R-Square	0.8032
Adj R-Sq	0.7638

Model: MODEL1
Dependent Variable: Weight

Observed by Predicted for Weight

Fit Diagnostics for Weight
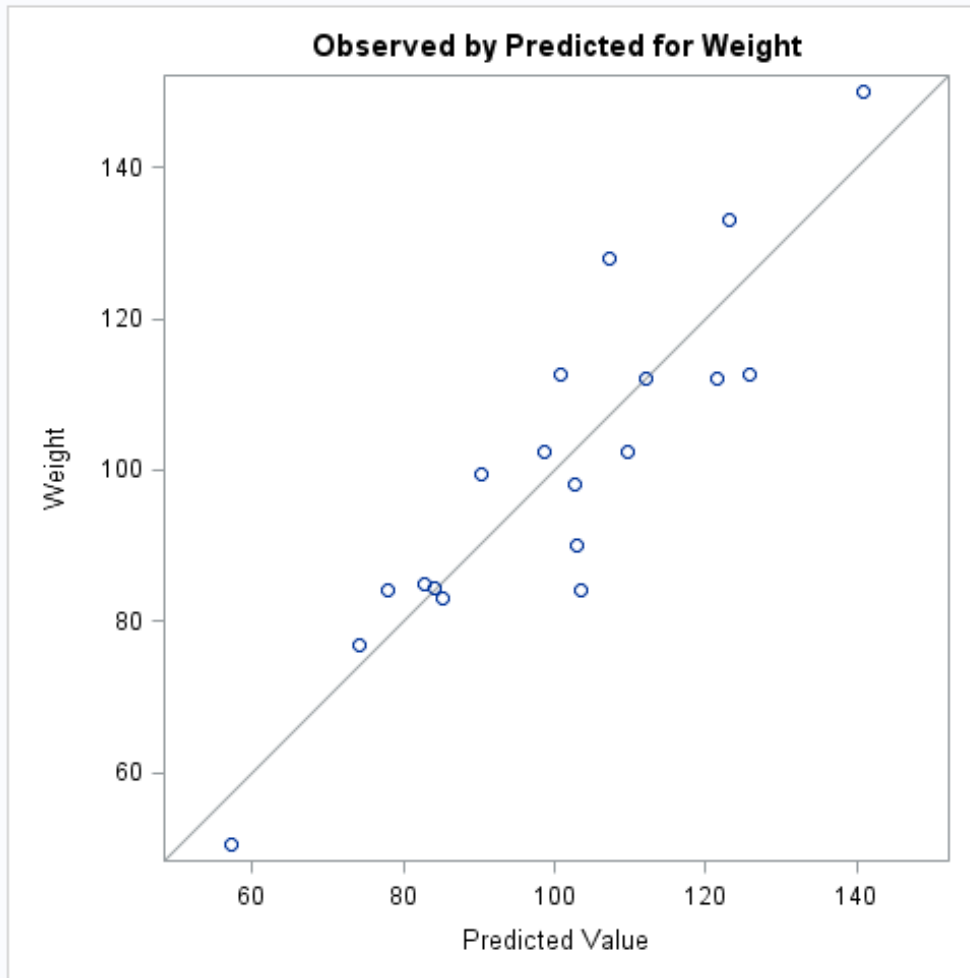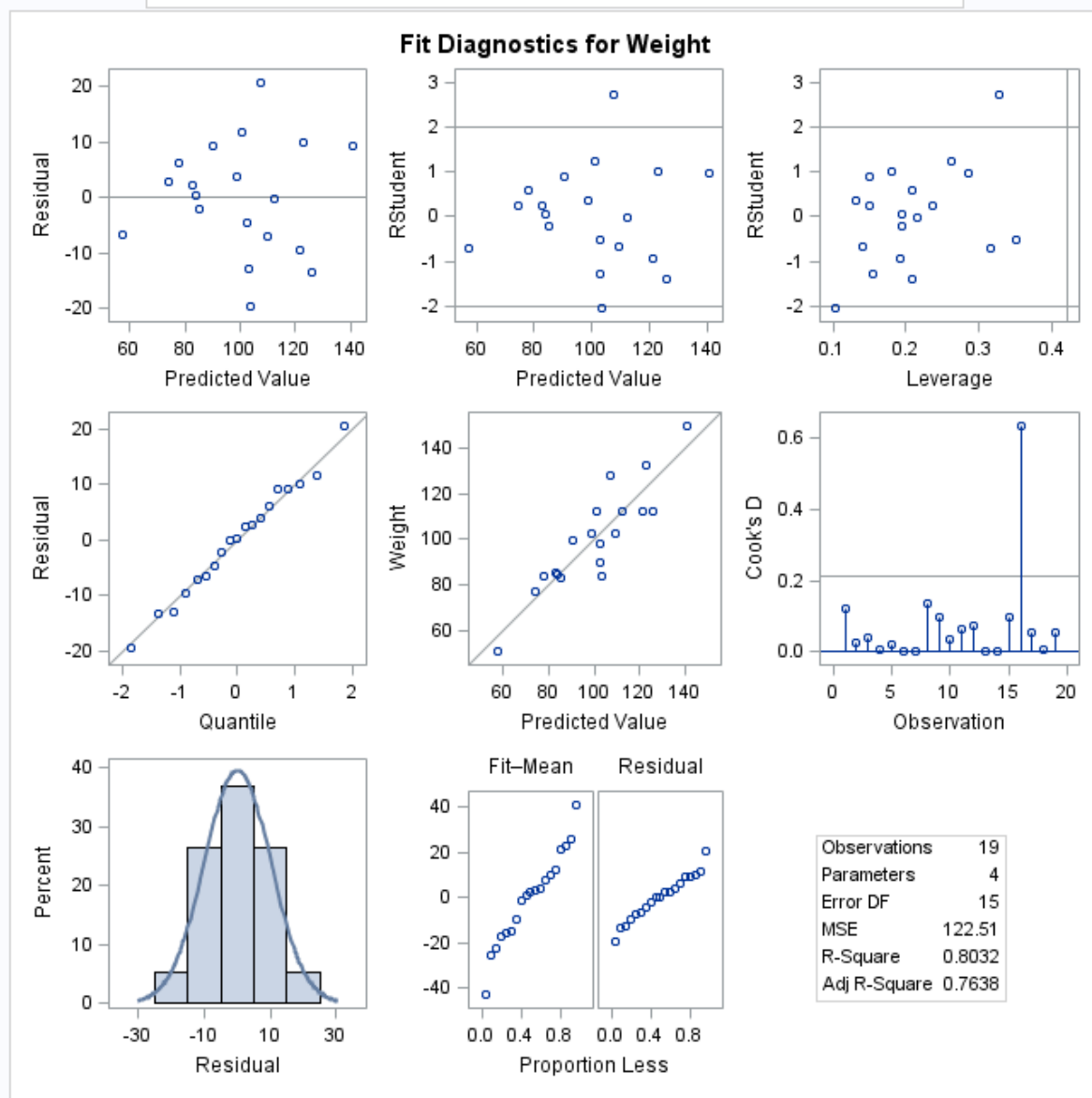
Assigning Data to Roles

To run the Linear Regression task, you must assign a column to the **Dependent variable** role and a column to the **Classification variables** role or the **Continuous variables** role.

Role	Description
Roles	
Dependent variable	specifies the numeric variable to use as the dependent variable for the regression analysis. You must assign a numeric variable to this role.
Classification variables	specifies categorical variables that enter the regression model through the design matrix coding.
Parameterization of Effects	
Coding	specifies the parameterization method for the classification variable. Design matrix columns are created from the classification variables according to the selected coding scheme.
	You can select from these coding schemes:
	■ **Effects coding** specifies effect coding.
	■ **GLM coding** specifies less-than-full-rank, reference-cell coding. This coding scheme is the default.
	■ **Reference coding** specifies reference-cell coding.
Treatment of Missing Values	

An observation is excluded from the analysis when either of these conditions is met:

■ if any variable in the model contains a missing value

■ if any classification variable contains a missing value (regardless of whether the classification variable is used in the model)

Role	Description
Continuous variables	specifies the numeric covariates (regressors) for the regression model.
Additional Roles	
Frequency count	lists a numeric variable whose value represents the frequency of the observation. If you assign a variable to this role, the task assumes that each observation represents *n* observations, where *n* is the value of the frequency variable. If *n* is not an integer, SAS truncates it. If *n* is less than 1 or is missing, the observation is excluded from the analysis. The sum of the frequency variable represents the total number of observations.
Weight	specifies the variable to use as a weight to perform a weighted analysis of the data.
Group analysis by	specifies to create a separate analysis for each group of observations.

Building a Model

Requirements for Building a Model

To specify an effect, you must assign at least one column to the **Classification variables** role or the **Continuous variables** role. You can select combinations of variables to create crossed, nested, factorial, or polynomial effects. You can also specify whether to include the intercept in the model.

To create the model, use the model builder on the **Model** tab.

Create a Main Effect

1 Select the variable name in the **Variables** box.

2 Click **Add** to add the variable to the **Model effects** box.

Create Crossed Effects (Interactions)

1   Select two or more variables in the **Variables** box. To select more than one variable, press Ctrl.

2   Click **Cross**.

Create a Nested Effect

Nested effects are specified by following a main effect or crossed effect with a classification variable or list of classification variables enclosed in parentheses. The main effect or crossed effect is nested within the effects listed in parentheses. Here are examples of nested effects: B(A), C(B*A), D*E(C*B*A). In this example, B(A) is read "A nested within B."

1   Select the effect name in the **Model effects** box.

2   Click **Nest**. The Nested window opens.

3   Select the variable to use in the nested effect. Click **Outer** or **Nested within Outer** to specify how to create the nested effect.

   Note: The **Nested within Outer** button is available only when a classification variable is selected.

4   Click **Add**.

Create a Full Factorial Model

1   Select two or more variables in the **Variables** box.

2   Click **Full Factorial**.

For example, if you select the Height, Weight, and Age variables and then click **Full Factorial**, these model effects are created: Age, Height, Weight, Age*Height, Age*Weight, Height*Weight, and Age*Height*Weight.

Create N-Way Factorial

1   Select two or more variables in the **Variables** box.

2   Click **N-way Factorial** to add these effects to the **Model effects** box.

For example, if you select the Height, Weight, and Age variables and then specify the value of N as 2, when you click **N-way Factorial**, these model effects are created: Age, Height, Weight, Age*Height, Age*Weight, and Height*Weight. If N is set to a value greater than the number of variables in the model, N is effectively set to the number of variables.

Create Polynomial Effects of the Nth Order

1 Select one variable in the **Variables** box.

2 Specify higher-degree crossings by adjusting the number in the **N** field.

3 Click **Polynomial Order=N** to add the polynomial effects to the **Model effects** box.

For example, if you select the Age and Height variables and then you specify 3 in the **N** field, when you click **Polynomial Order=N**, these model effects are created: Age, Age*Age, Age*Age*Age, Height, Height*Height, and Height*Height*Height.

Setting the Model Options

Option Name	Description
Methods	
Confidence level	specifies the significance level to use for the construction of confidence intervals.
Statistics	
You can choose to include the default statistics in the results or choose to include additional statistics.	
Additional available statistics	
Parameter Estimates	

Option Name	Description
Standardized regression coefficients	displays the standardized regression coefficients. A standardized regression coefficient is computed by dividing a parameter estimate by the ratio of the sample standard deviation of the dependent variable to the sample standard deviation of the regressor.
Confidence limits for estimates	displays the $100(1-\alpha)\%$ upper and lower confidence limits for the parameter estimates.
Sums of Squares	
Sequential sum of squares (Type I)	displays the sequential sums of squares (Type I SS) along with the parameter estimates for each term in the model.
Partial sum of squares (Type II)	displays the partial sums of squares (Type II SS) along with the parameter estimates for each term in the model.
Partial and Semipartial Correlations	
Squared partial correlations	displays the squared partial correlation coefficients computed by using Type I and Type II sums of squares.
Squared semipartial correlations	displays the squared semipartial correlation coefficients computed by using Type I and Type II sums of squares. This value is calculated as sum of squares divided by the corrected total sum of squares.
Diagnostics	
Analysis of influence	requests a detailed analysis of the influence of each observation on the estimates and the predicted values.

Option Name	Description
Analysis of residuals	requests an analysis of the residuals. The results include the predicted values from the input data and the estimated model, the standard errors of the mean predicted and residual values, the studentized residual, and Cook's *D* statistic to measure the influence of each observation on the parameter estimates.
Predicted values	calculates predicted values from the input data and the estimated model.
Multiple Comparisons	
Perform multiple comparisons	specifies whether to compute and compare the least squares means of fixed effects.
Select the effects to test	specifies the effects that you want to compare. You specified these effects on the **Model** tab.
Method	requests a multiple comparison adjustment for the *p*-values and confidence limits for the differences of the least squares means. Here are the valid methods: **Bonferroni**, **Nelson**, **Scheffé**, **Sidak**, and **Tukey**.
Significance level	requests that a *t* type confidence interval be constructed for each of the least squares means with a confidence level of 1 – number. The value of number must be between 0 and 1. The default value is 0.05.
Collinearity	
Collinearity analysis	requests a detailed analysis of collinearity among the regressors. This includes eigenvalues, condition indices, and decomposition of the variances of the estimates with respect to each eigenvalue.

Option Name	Description
Tolerance values for estimates	produces tolerance values for the estimates. Tolerance for a variable is defined as $1 - R^2$, where R square is obtained from the regression of the variable on all other regressors in the model.
Variance inflation factors	produces variance inflation factors with the parameter estimates. Variance inflation is the reciprocal of tolerance.
Heteroscedasticity	
Heteroscedasticity analysis	performs a test to confirm that the first and second moments of the model are correctly specified.
Asymptotic covariance matrix	displays the estimated asymptotic covariance matrix of the estimates under the hypothesis of heteroscedasticity and heteroscedasticity-consistent standard errors of parameter estimates.

Plots

Diagnostic and Residual Plots

By default, several diagnostic plots are included in the results. You can also specify whether to include plots of the residuals for each explanatory variable.

More Diagnostic Plots

Rstudent statistic by predicted values	plots studentized residuals by predicted values. If you select the **Label extreme points** option, observations with studentized residuals that lie outside the band between the reference lines $RSTUDENT = \pm 2$ are deemed outliers.

Option Name	Description
DFFITS statistic by observations	plots the DFFITS statistic by observation number. If you select the **Label extreme points** option, observations with a DFFITS statistic greater in magnitude than $2\sqrt{\frac{p}{n}}$ are deemed influential. The number of observations used is *n*, and the number of regressors is *p*.
DFBETAS statistic by observation number for each explanatory variable	produces panels of DFBETAS by observation number for the regressors in the model. You can view these plots as a panel or as individual plots. If you select the **Label extreme points** option, observations with a DFBETAS statistic greater in magnitude than $\frac{2}{\sqrt{n}}$ are deemed influential for that regressor. The number of observations used is *n*.
Label extreme points	identifies the extreme values on each different type of plot.
Scatter Plots	
Fit plot for a single continuous variable	produces a scatter plot of the data overlaid with the regression line, confidence band, and prediction band for models with a single continuous variable. The intercept is excluded. When the number of points exceeds the value for the **Maximum number of plot points** option, a heat map is displayed instead of a scatter plot.
Observed values by predicted values	produces a scatter plot of the observed values versus the predicted values.
Partial regression plots for each explanatory variable	produces partial regression plots for each regressor. If you display these plots in a panel, there is a maximum of six regressors per panel.
Maximum number of plot points	specifies the maximum number of points to include in each plot.

Setting the Model Selection Options

Option	Description
Model Selection	
Selection method	specifies the model selection method for the model. The task performs model selection by examining whether effects should be added to or removed from the model according to the rules that are defined by the selection method.
	Here are the valid values for the selection methods:
	■ **None** fits the full model.
	■ **Forward selection** starts with no effects in the model and adds effects based on the value of the specified criterion.
	■ **Backward elimination** starts with all the effects in the model and deletes effects based on the value of the specified criterion.
	■ **Stepwise selection** is similar to the forward selection model. However, effects that are already in the model do not necessarily stay there. Effects are added to the model based on the values of the specified criteria.
Criterion to add or remove effects	specifies the criterion to use to add or remove effects from the model.
Criterion to stop adding or removing effects	specifies the criterion to use to stop adding or removing effects from the model.
Select best model by	specifies the criterion to use to identify the best fitting model.
Selection Statistics	

Option	Description
Model fit statistics	specifies which model fit statistics are displayed in the fit summary table and the fit statistics tables. If you select **Default fit statistics**, the default set of statistics that are displayed in these tables includes all the criteria used in model selection.
	Here are the additional fit statistics that you can include in the results:
	■ Adjusted R-square
	■ Akaike's information criterion
	■ Akaike's information criterion corrected for small-sample bias
	■ Bayesian information criterion
	■ Mallows' Cp
	■ Press statistic, which specifies the predicted residual sum of squares statistic
	■ R-square
	■ Schwarz's Bayesian information criterion
Selection Plots	
Criterion plots	displays plots for these criteria: adjusted R-square, Akaike's information criterion, Akaike's information criterion corrected for small-sample bias, and the criterion used to select the best fitting model.
Coefficient plots	displays these plots:
	■ a plot that shows the progression of the parameter values as the selection process proceeds
	■ a plot that shows the progression of the criterion used to select the best fitting model
Details	

Option	Description
Selection process details	specifies how much information about the selection process to include in the results. You can display a summary, details for each step of the selection process, or all of the information about the selection process.

Creating Output Data Sets

You can specify whether to create an observationwise statistics data set. This data set contains the sum of squares and cross-products.

You can also choose to include these statistics in the output data set:

- Cook's *D* influence

- the standard influence of observation on covariance of betas

- the standard influence of an observation on predicted value (called DFFITS)

- leverage

- predicted values

- press statistic, which is the ith residual divided by $(1 - h)$, where h is the leverage, and where the model has been refit without the ith observation

- residual

- studentized residuals, which are the residuals divided by their standard errors

- studentized residual with current observation removed

Binary Logistic Regression Task

About the Binary Logistic Regression Task

The Binary Logistic Regression task is used to fit a logistic regression model to investigate the relationship between discrete responses with binary levels and a set of explanatory variables.

Example: Classifying E-Mail as Junk

To create this example:

1   In the **Tasks** section, expand the **Statistics** folder and double-click **Binary Logistic Regression**. The user interface for the Binary Logistic Regression task opens.

2   On the **Data** tab, select the **SASHELP.JUNKMAIL** data set.

3   Assign columns to these roles and specify these options:

Role	Column Name
Response	**Class**
Event of interest	**1**
Continuous variables	**Exclamation** **CapAvg**

4   Click the **Model** tab. Select the **Exclamation** and **CapAvg** variables, and click **Add**.

▲ MODEL EFFECTS

Model

Variables:		Model effects: ⬆ ⬇ 🗑
Exclamation CapAvg	**Single Effects** [Add] [Cross] [Nest] **Standard Models** [Full Factorial] [N-way Factorial] [Polynomial Order = N]	Exclamation CapAvg

☑ Include an intercept in the model

Offset variable: *(1 item)*   🗑 ✚

123 *Column*

5   To run the task, click 🏃.

Model Information		
Data Set	SASHELP.JUNKMAIL	Classifying Email as Junk or Not
Response Variable	Class	0 - Not Junk, 1 - Junk
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	4601
Number of Observations Used	4601

Response Profile		
Ordered Value	Class	Total Frequency
1	0	2788
2	1	1813

Probability modeled is Class=1.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	6172.153	5035.224
SC	6178.587	5054.526
-2 Log L	6170.153	5029.224

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	1140.9288	2	<.0001
Score	312.5004	2	<.0001
Wald	630.4776	2	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.8144	0.0652	773.5908	<.0001
Exclamation	1	1.8816	0.1165	261.0441	<.0001
CapAvg	1	0.3077	0.0189	264.8582	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
Exclamation	6.564	5.224	8.247

Assigning Data to Roles

To run the Binary Logistic Regression task, you must assign columns to the **Response variable** and a column to either the **Classification variables** role or the **Continuous variables** role.

Role	Description
Roles	
Response	
Response data consists of numbers of events and trials	specifies whether the response data consists of events and trials.
Number of events	specifies the variable that contains the number of events for each observation.
Number of trials	specifies the variable that contains the number of trials for each observation.
Response	specifies the variable that contains the response data. To perform a binary logistic regression, the response variable should have only two levels. Use the **Event of interest** drop-down list to select the event category for the binary response model.
Link function	specifies the link function that links the response probabilities to the linear predictors. Here are the valid values: ■ **Complementary log-log** is the complementary log-log function. ■ **Probit** is the inverse standard normal distribution function. ■ **Logit** is the log odds function.
Explanatory Variables	

Role	Description
Classification variables	specifies the classification variables to use in the analysis. A classification variable is a variable that enters the statistical analysis or model not through its values, but through its levels. The process of associating values of a variable with levels is termed levelization.
Parameterization of Effects	
Coding	specifies the parameterization method for the classification variable. Design matrix columns are created from the classification variables according to the selected coding scheme.

You can select from these coding schemes:

- **Effects coding** specifies effect coding.
- **GLM coding** specifies less-than-full-rank, reference-cell coding. This coding scheme is the default.
- **Reference coding** specifies reference-cell coding.

Treatment of Missing Values

An observation is excluded from the analysis when either of these conditions is met:

- if any variable in the model contains a missing value
- if any classification variable contains a missing value (regardless of whether the classification variable is used in the model)

Continuous variables	specifies the continuous variables to use as the explanatory variables in the analysis.
Additional Roles	
Frequency count	specifies the variables that contain the frequency of occurrence for each observation. The task treats each observation as if it appears *n* times, where *n* is the value of the variable for that observation.

Role	Description
Weight variable	specifies the how much to weight each observation in the input data set.
Group analysis by	creates separate analyses based on the number of BY variables.

Building a Model

Create a Main Effect

1 Select the variable name in the **Variables** box.

2 Click **Add** to add the variable to the **Model effects** box.

Create Crossed Effects (Interactions)

1 Select two or more variables in the **Variables** box. To select more than one variable, press Ctrl.

2 Click **Cross**.

Create a Nested Effect

Nested effects are specified by following a main effect or crossed effect with a classification variable or list of classification variables enclosed in parentheses. The main effect or crossed effect is nested within the effects listed in parentheses. Here are examples of nested effects: B(A), C(B*A), D*E(C*B*A). In this example, B(A) is read "A nested within B."

1 Select the effect name in the **Model effects** box.

2 Click **Nest**. The Nested window opens.

3 Select the variable to use in the nested effect. Click **Outer** or **Nested within Outer** to specify how to create the nested effect.

Note: The **Nested within Outer** button is available only when a classification variable is selected.

4 Click **Add**.

Create a Full Factorial Model

1 Select two or more variables in the **Variables** box.

2 Click **Full Factorial**.

For example, if you select the Height, Weight, and Age variables and then click **Full Factorial**, these model effects are created: Age, Height, Weight, Age*Height, Age*Weight, Height*Weight, and Age*Height*Weight.

Create N-Way Factorial

1 Select two or more variables in the **Variables** box.

2 Click **N-way Factorial** to add these effects to the **Model effects** box.

For example, if you select the Height, Weight, and Age variables and then specify the value of N as 2, when you click **N-way Factorial**, these model effects are created: Age, Height, Weight, Age*Height, Age*Weight, and Height*Weight. If N is set to a value greater than the number of variables in the model, N is effectively set to the number of variables.

Create Polynomial Effects of the Nth Order

1 Select one variable in the **Variables** box.

2 Specify higher-degree crossings by adjusting the number in the **N** field.

3 Click **Polynomial Order=N** to add the polynomial effects to the **Model effects** box.

For example, if you select the Age and Height variables and then you specify 3 in the **N** field, when you click **Polynomial Order=N**, these model effects are created: Age, Age*Age, Age*Age*Age, Height, Height*Height, and Height*Height*Height.

Setting the Model Options

Option	Description
Model	
Include an intercept in the model	specifies whether to include the intercept in the model.
Offset variable	specifies a variable to be used as an offset to the linear predictor. An offset plays the role of an effect whose coefficient is known to be 1. Observations that have missing values for the offset variable are excluded from the analysis.

Specifying the Model Selection Options

Option	Description
Model Selection	
Selection method	specifies the model selection method for the model. The task performs model selection by examining whether effects should be added to or removed from the model according to the rules that are defined by the selection method. Here are the valid values for the selection methods: ■ **None** fits the full model. ■ **Forward selection** starts with no effects in the model and adds effects based on the **Significance level to add an effect to the model** option. ■ **Backward elimination** starts with all the effects in the model and deletes effects based on the value in the **Significance level to remove an effect from the model** option.

Option	Description
Selection method (continued)	■ **Fast backward elimination** uses a computational algorithm of Lawless and Singhal (1978). This algorithm computes a first-order approximation to the remaining slope estimates for each subsequent elimination of a variable from the model. Variables are removed from the model based on these approximate estimates. This selection method is extremely efficient because the model is not refitted for every variable removed.
	■ **Stepwise selection** is similar to the forward selection model. However, effects that are already in the model do not necessarily stay there. Effects are added to the model based on the **Significance level to add an effect to the model** option and are removed from the model based on the **Significance level to remove an effect from the model** option.
	■ **Stepwise selection with fast backward elimination** uses a computational algorithm of Lawless and Singhal. This algorithm computes a first-order approximation to the remaining slope estimates for each subsequent elimination of a variable from the model. Variables are removed from the model based on these approximate estimates. This selection method is extremely efficient because the model is not refitted for every variable removed.
Details	
Display selection process details	specifies how much information about the selection process to include in the results. You can choose to display a summary, details for each step of the selection process, or all of the information about the selection process.

Option	Description
Maintain hierarchy of effects	specifies how the model hierarchy requirement is applied and that only a single effect or multiple effects can enter or leave the model at one time. For example, suppose you specify the main effects A and B and the interaction A*B in the model. In the first step of the selection process, either A or B can enter the model. In the second step, the other main effect can enter the model. The interaction effect can enter the model only when both main effects have already been entered. Also, before A or B can be removed from the model, the A*B interaction must first be removed.
	Model hierarchy refers to the requirement that, for any term to be in the model, all effects contained in the term must be present in the model. For example, in order for the interaction A*B to enter the model, the main effects A and B must be in the model. Likewise, neither effect A nor B can leave the model while the interaction A*B is in the model.

Setting Options

Option Name	Description
Statistics	
Note: In addition to the default statistics that are included in the results, you can select the additional statistics to include.	
Classification table	classifies the input binary response observations according to whether the predicted event probabilities are above or below the cut-point value z in the range. An observation is predicted as an event if the predicted event probability equals or exceeds z.

Option Name	Description
Partial correlation	computes the partial correlation statistic $\left(\beta_i\right)\sqrt{\dfrac{\chi_i^2 - 2}{-2\log L_0}}$ for each parameter *i*, where X^2_i is the Wald chi-square statistic for the parameter and *log L_0* is the log-likelihood of the intercept-only model (Hilbe 2009). If $X^2_i <$ 2, the partial correlation is set to 0.
Generalized R square	requests a generalized R square measure for the fitted model.
Goodness-of-fit and Overdispersion	
Deviance and Pearson goodness-of-fit	specifies whether to calculate the deviance and Pearson goodness-of-fit.
Aggregate by	specifies the subpopulations on which the Pearson chi-square test statistic and the likelihood ratio chi-square test statistic (deviance) are calculated. Observations with common values in the given list of variables are regarded as coming from the same subpopulation. Variables in the list can be any variables in the input data set.
Correct for overdispersion	specifies whether to correct for overdispersion by using the Deviance or Pearson estimate.

Option Name	Description
Hosmer & Lemeshow goodness-of-fit	performs the Hosmer and Lemeshow goodness-of-fit test (Hosmer and Lemeshow 2000) for the case of a binary response model. The subjects are divided into approximately 10 groups of approximately the same size based on the percentiles of the estimated probabilities. The discrepancies between the observed and expected number of observations in these groups are summarized by the Pearson chi-square statistic. This statistic is then compared to a chi-square distribution with t degrees of freedom, where t is the number of groups minus n. By default, $n = 2$. A small p-value suggests that the fitted model is not an adequate model.
Multiple Comparisons	
Perform multiple comparisons	specifies whether to compute and compare the least squares means of fixed effects.
Select the effects to test	specifies the effects that you want to compare. You specified these effects on the **Model** tab.
Method	requests a multiple comparison adjustment for the p-values and confidence limits for the differences of the least squares means. Here are the valid methods: **Bonferroni**, **Nelson**, **Scheffé**, **Sidak**, and **Tukey**.
Significance level	requests that a t type confidence interval be constructed for each of the least squares means with a confidence level of 1 – *number*. The value of *number* must be between 0 and 1. The default value is 0.05.
Exact Tests	
Exact test of intercept	calculates the exact test for the intercept.

Option Name	Description
Select effects to test	calculates exact tests of the parameters for the selected effects.
Significance level	specifies the level of significance α for $100(1 - \alpha)\%$ confidence limits for the parameters or odds ratios.

Parameter Estimates

You can calculate these parameter estimates:

- standardized estimates
- exponentiated estimates
- correlations of parameter estimates
- covariances of parameter estimates

You can specify the confidence intervals for parameters, confidence intervals for odds ratios, and the confidence level for these estimates.

Diagnostics

Influence diagnostics	displays the diagnostic measures for identifying influential observations. For each observation, the results include the sequence number of the observation, the values of the explanatory variables included in the final model, and the regression diagnostic measures developed by Pregibon (1981). You can specify whether to include the standardized and likelihood residuals in the results.

Plots

Option Name	Description

You can select whether to include plots in the results.

Here are the additional plots that you can include in the results:

- standardized DFBETA by observation number
- influence statistics by observation number
- influence on model fit and parameter estimates
- predicted probability plots
- effect plot
- odds ratio plot
- ROC plot

You can specify whether to display these plots in a panel or individually.

Label influence and ROC plots	specifies the variable from the input data that contains the labels for the influence and ROC plots.
Maximum number of plot points	specifies the maximum number of points to include in the plots. By default, 5,000 points are shown.

Methods

Optimization

Method	specifies the optimization technique for estimating the regression parameters. The Fisher scoring and Newton-Raphson algorithms yield the same estimates, but the estimated covariance matrices are slightly different except when the Logit link function is specified for binary response data.
Maximum number of iterations	specifies the maximum number of iterations to perform. If convergence is not attained in a specified number of iterations, the displayed output and all output data sets created by the task contain results that are based on the last maximum likelihood iteration.

Creating Output Data Sets

Option Name	Description
Output Data Sets	

You can create two types of output data sets. Select the check box for each data set that you want to create.

Create output data set
outputs a data set that contains the specified statistics.

Here are the statistics that you can include in the output data set:

- linear predictor
- predicted values
- confidence limits for predicted values
- Pearson residuals
- Deviance residuals
- Likelihood residuals
- standardized Pearson residuals
- standardized deviance residuals
- change in the chi-square goodness-of-fit from deleting the individual observation
- change in the deviance from deleting the individual observation
- leverage
- standardized DFBETA
- standard error of the linear predictor
- predicted probabilities for each response level

Create scored data set
outputs a data set that contains all the statistics in the output data set plus posterior probabilities.

Add SAS scoring code to the log	writes SAS DATA step code for computing predicted values of the fitted model either to a file or to a catalog entry. This code can then be included in a DATA step to score new data.

Predictive Regression Model

About the Predictive Regression Model

The task is predictive in that it selects the most influential effects based on observed data. This task enables you to logically partition your data into disjoint subsets for model training, validation, and testing. The Predictive Regression Model task focuses on the standard independently and identically distributed general linear model for univariate responses and offers great flexibility and insight into the model selection algorithm. This task can also create a scored data set. The results for this task make it easy to explore the selected model in more detail with other tasks, such as the Linear Regression task.

Example: Predicting a Baseball Player's Salary

To create this example:

1   In the **Tasks** section, expand the **Statistics** folder and double-click **Predictive Regression Model**. The user interface for the Predictive Regression Model task opens.

2   On the **Data** tab, select the **SASHELP.BASEBALL** data set.

3   Assign columns to these roles and specify these options:

Role	Column Name
Dependent variable	**logSalary**
Classification variables	**League** **Division**

Role	Column Name
Continuous variables	**nAtBat**
	nHits
	nHome
	nRuns
	nRBI
	nBB

4 Click the **Model** tab. Select the **nAtBat**, **nHits**, **nHome**, **nRuns**, **nRBI**, **nBB**. **League**, and **Division** variables, and then click **Add**.

5 To run the task, click .

Data Set	SASHELP.BASEBALL
Dependent Variable	logSalary
Selection Method	Stepwise
Select Criterion	SBC
Stop Criterion	SBC
Effect Hierarchy Enforced	Single

Number of Observations Read	322
Number of Observations Used	263

Class Level Information		
Class	Levels	Values
League	2	American National
Division	2	East West

Dimensions	
Number of Effects	9
Number of Parameters	11

	Stepwise Selection Summary				
Step	Effect Entered	Effect Removed	Number Effects In	Number Parms In	SBC
0	Intercept		1	1	-57.2041
1	nHits		2	2	-124.6362
2	nBB		3	3	-137.9363*
	* Optimal Value of Criterion				

Selection stopped at a local minimum of the SBC criterion.

Stop Details				
Candidate For	Effect	Candidate SBC		Compare SBC
Entry	nRBI	-137.2172	>	-137.9363
Removal	nBB	-124.6362	>	-137.9363

Partitioning Your Data

When you have sufficient data, you can partition your data into three parts: training data, validation data, and test data. During the selection process, models are fit on the training data, and the prediction error for the model is determined using the validation

data. This prediction error can be used to decide when to terminate the selection process or which effects to include as the selection process proceeds. Finally, after a model is selected, the test data can be used to assess how the selected model generalizes on data that played no role in selecting the model.

You can partition your data in either of these ways:

■ You can specify a proportion of the validation or test data. The proportions are used to divide the input data by sampling.

■ If the input data set contains a variable whose values indicate whether an observation is a validation or test case, you can specify the variable to use when partitioning the data. When you specify the variable, you also select the appropriate values for validation or test cases. The input data set is divided into partitions by using these values.

Assigning Data to Roles

To run the Predictive Regression Model task, you must assign a column to the **Dependent variable** role and a column to the **Classification variables** role or the **Continuous variables** role.

Role	Description
Roles	
Dependent variable	specifies the numeric variable to use as the dependent variable for the regression analysis.
Classification variables	specifies the variables to use to group (classify) data in the analysis. A classification variable is a variable that enters the statistical analysis or model through its levels, not through its values. The process of associating values of a variable with levels is termed levelization.
Parameterization of Effects	

Role	Description
Coding	specifies the parameterization method for the classification variable. Design matrix columns are created from the classification variables according to the selected coding scheme.
	You can select from these coding schemes:
	■ **Effects coding** specifies effect coding.
	■ **GLM coding** specifies less-than-full-rank, reference-cell coding. This coding scheme is the default.
	■ **Reference coding** specifies reference-cell coding.

Treatment of Missing Values

An observation is excluded from the analysis if any variable in the model contains a missing value. In addition, an observation is excluded if any classification variable specified earlier in this table contains a missing value, regardless if it is used in the model.

Continuous variables	specifies the independent covariates (regressors) for the regression model. If you do not specify a continuous variable, the task fits a model that contains only an intercept.

Additional Roles

Frequency count	lists a numeric variable whose value represents the frequency of the observation. If you assign a variable to this role, the task assumes that each observation represents n observations, where n is the value of the frequency variable. If n is not an integer, SAS truncates it. If n is less than 1 or is missing, the observation is excluded from the analysis. The sum of the frequency variable represents the total number of observations.
Weight	specifies the numeric column to use as a weight to perform a weighted analysis of the data.

Role	Description
Group analysis by	specifies to create a separate analysis for each group of observations.

Building a Model

Requirements for Building a Model

To specify an effect, you must assign at least one column to the **Classification variables** role or the **Continuous variables** role. You can select combinations of variables to create crossed, factorial, or polynomial effects.

To create a model, use the model builder on the **Model** tab. After you create a model, you can specify whether to include the intercept in the model.

Create a Main Effect

1 Select the variable name in the **Variables** box.

2 Click **Add** to add the variable to the **Model effects** box.

Create Crossed Effects (Interactions)

1 Select two or more variables in the **Variables** box. To select more than one variable, press Ctrl.

2 Click **Cross**.

Create a Nested Effect

Nested effects are specified by following a main effect or crossed effect with a classification variable or list of classification variables enclosed in parentheses. The main effect or crossed effect is nested within the effects listed in parentheses. Here are examples of nested effects: B(A), C(B*A), D*E(C*B*A). In this example, B(A) is read "A nested within B."

1 Select the effect name in the **Model effects** box.

2   Click **Nest**. The Nested window opens.

3   Select the variable to use in the nested effect. Click **Outer** or **Nested within Outer** to specify how to create the nested effect.

   Note: The **Nested within Outer** button is available only when a classification variable is selected.

4   Click **Add**.

Create a Full Factorial Model

1   Select two or more variables in the **Variables** box.

2   Click **Full Factorial**.

For example, if you select the Height, Weight, and Age variables and then click **Full Factorial**, these model effects are created: Age, Height, Weight, Age*Height, Age*Weight, Height*Weight, and Age*Height*Weight.

Create N-Way Factorial

1   Select two or more variables in the **Variables** box.

2   Click **N-way Factorial** to add these effects to the **Model effects** box.

For example, if you select the Height, Weight, and Age variables and then specify the value of N as 2, when you click **N-way Factorial**, these model effects are created: Age, Height, Weight, Age*Height, Age*Weight, and Height*Weight. If N is set to a value greater than the number of variables in the model, N is effectively set to the number of variables.

Create Polynomial Effects of the Nth Order

1   Select one variable in the **Variables** box.

2   Specify higher-degree crossings by adjusting the number in the **N** field.

3   Click **Polynomial Order=N** to add the polynomial effects to the **Model effects** box.

For example, if you select the Age and Height variables and then you specify 3 in the **N** field, when you click **Polynomial Order=N**, these model effects are created: Age, Age*Age, Age*Age*Age, Height, Height*Height, and Height*Height*Height.

Selecting a Model

Option Name	Description
Model Selection	
Selection method	By default, the complete model that you specified is used to fit the model. However, you can also use one of these selection methods:
Selection method (continued)	**Forward selection** specifies forward selection. This method starts with no effects in the model and adds effects. **Backward elimination** specifies backward elimination. This method starts with all effects in the model and deletes effects. **Stepwise regression** specifies stepwise regression, which is similar to the forward selection method except that effects already in the model do not necessarily stay there. **LASSO** specifies the LASSO method, which adds and deletes parameters based on a version of ordinary least squares where the sum of the absolute regression coefficients is constrained. If the model contains classification variables, these classification variables are split. **Adaptive LASSO** requests that adaptive weights be applied to each of the coefficients in the LASSO method. The ordinary least squares estimates of the parameters in the model are used in forming the adaptive weights.

Option Name	Description
Selection method (continued)	**Elastic net** specifies the elastic net method, which is an extension of LASSO. The elastic net method estimates parameters based on a version of ordinary least squares in which both the sum of the absolute regression coefficients and the sum of the squared regression coefficients are constrained. If the model contains classification variables, these classification variables are split. **Least angle regression** specifies least angle regression. This method starts with no effects in the model and adds effects. The parameter estimates at any step are "shrunk" when compared to the corresponding least squares estimates. If the model contains classification variables, these classification variables are split.
Criterion to add or remove effects	specifies the criterion to use to determine whether an effect should be added or removed from the model.
Criterion to stop adding or removing effects	specifies the criterion to use to determine whether effects should stop being added or removed from the model.
Select best model by	specifies the criterion to use to determine the best fitting model.
Selection Statistics	

Option Name	Description
Model fit statistics	specifies which model fit statistics are displayed in the fit summary table and the fit statistics tables. If you select **Default fit statistics**, the default set of statistics that are displayed in these tables includes all the criteria used in model selection.
	Here are the additional fit statistics that you can include in the results:
	■ Adjusted R-square
	■ Akaike's information criterion
	■ Akaike's information criterion corrected for small-sample bias
	■ Average square error
	■ Bayesian information criterion
	■ Mallows' Cp
	■ Press statistic, which specifies the predicted residual sum of squares statistic
	■ R-square
	■ Schwarz's Bayesian information criterion
Selection Plots	
Criterion plots	displays plots for these criteria: adjusted R-square, Akaike's information criterion, Akaike's information criterion corrected for small-sample bias, and the criterion used to select the best fitting model. You can choose to display these plots in a panel or individually.
Coefficient plots	displays these plots:
	■ a plot that shows the progression of the parameter values as the selection process proceeds
	■ a plot that shows the progression of the criterion used to select the best fitting model
Details	

Option Name	Description
Selection process details	specifies how much information about the selection process to include in the results. You can display a summary, details for each step of the selection process, or all of the information about the selection process.
Model Effects Hierarchy	
Model effects hierarchy	specifies how the model hierarchy requirement is applied and that only a single effect or multiple effects can enter or leave the model at one time. For example, suppose you specify the main effects A and B and the interaction A*B in the model. In the first step of the selection process, either A or B can enter the model. In the second step, the other main effect can enter the model. The interaction effect can enter the model only when both main effects have already been entered. Also, before A or B can be removed from the model, the A*B interaction must first be removed.
	Model hierarchy refers to the requirement that, for any term to be in the model, all effects contained in the term must be present in the model. For example, in order for the interaction A*B to enter the model, the main effects A and B must be in the model. Likewise, neither effect A nor B can leave the model while the interaction A*B is in the model.
Model effects subject to the hierarchy requirement	specifies whether to apply the model hierarchy requirement to the classification and continuous effects in the model or to only the classification effects.

Setting the Options for the Final Model

Option Name	Description
Statistics for the Selected Model	
You can choose to include the default statistics in the results or choose to include additional statistics, such as the standardized regression coefficients. A standardized regression coefficient is computed by dividing a parameter estimate by the ratio of the sample standard deviation of the dependent variable to the sample standard deviation of the regressor.	
Collinearity	
Collinearity analysis	requests a detailed analysis of collinearity among the regressors. This includes eigenvalues, condition indices, and decomposition of the variances of the estimates with respect to each eigenvalue.
Tolerance values for estimates	produces tolerance values for the estimates. Tolerance for a variable is defined as $1 - R^2$, where R square is obtained from the regression of the variable on all other regressors in the model.
Variance inflation factors	produces variance inflation factors with the parameter estimates. Variance inflation is the reciprocal of tolerance.
Plots for the Selected Model	
Diagnostic and Residual Plots	
You must specify whether to include the default diagnostic plots in the results. You can also specify whether to include plots of the residuals for each explanatory variable.	
More Diagnostic Plots	

Option Name	Description
Rstudent statistic by predicted values	plots studentized residuals by predicted values. If you select the **Label extreme points** option, observations with studentized residuals that lie outside the band between the reference lines $RSTUDENT = \pm 2$ are deemed outliers.
DFFITS statistic by observation number	plots the DFFITS statistic by observation number. If you select the **Label extreme points** option, observations with a DFFITS statistic greater in magnitude than $2\sqrt{\frac{p}{n}}$ are deemed influential. The number of observations used is *n*, and the number of regressors is *p*.
DFBETAS statistic by observation number for each explanatory variable	produces panels of DFBETAS by observation number for the regressors in the model. You can view these plots as a panel or as individual plots. If you select the **Label extreme points** option, observations with a DFBETAS statistic greater in magnitude than $\frac{2}{\sqrt{n}}$ are deemed influential for that regressor. The number of observations used is *n*.
Label extreme points	identifies the extreme values on each different type of plot.
Scatter Plots	
Observed values by predicted values	produces a scatter plot of the observed values versus the predicted values.
Partial regression plots for each explanatory variable	produces partial regression plots for each regressor. If you display these plots in a panel, there is a maximum of six regressors per panel.
Maximum number of plot points	specifies the maximum number of points to include in each plot.

Setting the Scoring Options

Option Name	Description
Scoring	
You can create a scored data set, which contains the predicted values and the residuals.	
Add SAS scoring code to the log	writes SAS DATA step code for computing predicted values of the fitted model either to a file or to a catalog entry. This code can then be included in a DATA step to score new data.

Generalized Linear Models

About the Generalized Linear Models Task

Generalized linear models are an extension of traditional linear models. In a generalized linear model, the mean of a population depends on a linear predictor through a nonlinear link function. The response probability distribution can be any member of the exponential family of distributions. Examples of generalized linear models include classical linear models with normal errors, logistic and probit models for binary data, and log-linear models for multinomial data. Other statistical models can be formulated as generalized linear models by the selection of an appropriate link function and response probability distribution.

The Generalized Linear Models task provides model fitting and model building for generalized linear models. It fits models for standard distributions such as Normal, Poisson, and Tweedie in the exponential family. This task also fits multinomial models for ordinal and nominal responses. The task provides forward, backward, and stepwise selection methods.

Example: Analyzing the Sashelp.Baseball Data Set

To create this example:

1   In the **Tasks** section, expand the **Statistics** folder and double-click **Generalized Linear Models**. The user interface for the Generalized Linear Models task opens.

2   On the **Data** tab, select the **SASHELP.BASEBALL** data set.

3   From the **Distribution** drop-down list, select **Poisson**. Assign columns to these roles:

Role	Column Name
Response	
Response variable	**nHome** From the **Link function** drop-down list, select **Logarithm**.
Explanatory Variables	
Classification variables	**League**
Continuous variables	**logSalary**

4   Click the **Model** tab. In the **Variables** box, select **League** and **logSalary**. Click **Add** to add these as main effects.

5   To run the task, click ⚉.

Here is a subset of the results:

Model Information		
Data Set	SASHELP.BASEBALL	1986 Baseball Data
Distribution	Poisson	
Link Function	Log	
Dependent Variable	nHome	Home Runs in 1986

Number of Observations Read	322
Number of Observations Used	263
Missing Values	59

Class Level Information		
Class	Levels	Values
League	2	American National

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	260	1442.0654	5.5464
Scaled Deviance	260	1442.0654	5.5464
Pearson Chi-Square	260	1449.1038	5.5735
Scaled Pearson X2	260	1449.1038	5.5735
Log Likelihood		4726.6264	
Full Log Likelihood		-1237.6660	
AIC (smaller is better)		2481.3320	
AICC (smaller is better)		2481.4247	
BIC (smaller is better)		2492.0485	

Assigning Data to Roles

To run the Generalized Linear Models task, you must assign a column to the **Response variable** role for all distribution types except binomial. If you select a binomial distribution, you must assign either a single response variable or a pair of variables to the **Number of events** and **Number of trials** roles.

Option Name	Description
Roles	

Option Name	Description
Response	
Distribution	specifies the distribution for your model. You can choose from these distributions: ■ Binomial ■ Gamma ■ Inverse Gaussian ■ Multinomial ■ Negative binomial ■ Normal ■ Poisson ■ Tweedie. If you select a Tweedie distribution, you can specify the Tweedie power parameter. This value can be 0, 1, or a value greater than 1.1 and less than or equal to 3.0. ■ Zero-inflated negative binomial ■ Zero-inflated Poisson
Options for Binomial Distribution	
Response data consists of numbers of events and trials	specifies that a pair of variables consists of response data for events and trials.
Number of events	specifies the column that contains the number of events.
Number of trials	specifies the column that contains the number of trials.

Option Name	Description
Response	specifies the single variable that contains response values.
	Use the **Event of interest** option to select a value of the response variable that represents the event that you want to model.
	Note: The **Response** role and the **Event of interest** option are available only if you do not select the **Response data consists of numbers of events and trials** check box.
Options for All Distribution Types	
Response	specifies the variable that contains the response data. For most distribution types, you specify a single numeric variable.
Link function	specifies the link function for your model. The functions that are available depend on the selected distribution.
Explanatory Variables	
Classification variables	specifies the variables to use to group (classify) data in the analysis. Classification variables can be either character or numeric. A classification variable is a variable that enters the statistical analysis or model through its levels, not through its values. The process of associating values of a variable with levels is termed levelization.
Parameterization of Effects	

Option Name	Description
Coding	specifies the parameterization method for the classification variable. Design matrix columns are created from the classification variables according to the selected coding scheme.
	You can select from these coding schemes:
	■ **Effect coding** specifies effect coding.
	■ **GLM coding** specifies less-than-full-rank, reference-cell coding. This coding scheme is the default.
	■ **Reference coding** specifies reference-cell coding.

Treatment of Missing Values

An observation is excluded from the analysis when either of these conditions is met:

■ if any variable in the model contains a missing value

■ if any classification variable contains a missing value (regardless of whether the classification variable is used in the model)

Continuous variables	specifies the independent covariates (regressors) for the regression model. If you do not specify a continuous variable, the task fits a model that contains only an intercept.
Offset variable	specifies a variable to be used as an offset to the linear predictor. An offset plays the role of an effect whose coefficient is known to be 1. Observations that have missing values for the offset variable are excluded from the analysis.

Additional Roles

Frequency count	specifies the numeric column that contains the frequency of occurrence for each observation.
Weight variable	specifies the numeric column to use as a weight to perform a weighted analysis of the data.

Option Name	Description
Group analysis by	specifies the column to use as the BY variable.

Building a Model

Requirements for Building a Model

By default, no effects are specified, which results in the task fitting an intercept-only model. To specify an effect, you must assign at least one variable to the **Classification variables** role or the **Continuous variables** role. You can select combinations of variables to create crossed, nested, factorial, or polynomial effects.

To create a model, use the model builder on the **Models** tab. After you create the model, you can specify whether to include the intercept in the model.

Create a Main Effect

1   Select the variable name in the **Variables** box.

2   Click **Add** to add the variable to the **Model effects** box.

Create Crossed Effects (Interactions)

1   Select two or more variables in the **Variables** box. To select more than one variable, press Ctrl.

2   Click **Cross**.

Create a Nested Effect

Nested effects are specified by following a main effect or crossed effect with a classification variable or list of classification variables enclosed in parentheses. The main effect or crossed effect is nested within the effects listed in parentheses. Here are examples of nested effects: B(A), C(B*A), D*E(C*B*A). In this example, B(A) is read "A nested within B."

1   Select the effect name in the **Model effects** box.

2   Click **Nest**. The Nested window opens.

3   Select the variable to use in the nested effect. Click **Outer** or **Nested within Outer** to specify how to create the nested effect.

> **Note:** The **Nested within Outer** button is available only when a classification variable is selected.

4   Click **Add**.

Create a Full Factorial Model

1   Select two or more variables in the **Variables** box.

2   Click **Full Factorial**.

For example, if you select the Height, Weight, and Age variables and then click **Full Factorial**, these model effects are created: Age, Height, Weight, Age*Height, Age*Weight, Height*Weight, and Age*Height*Weight.

Create N-Way Factorial

1   Select two or more variables in the **Variables** box.

2   Click **N-way Factorial** to add these effects to the **Model effects** box.

For example, if you select the Height, Weight, and Age variables and then specify the value of N as 2, when you click **N-way Factorial**, these model effects are created: Age, Height, Weight, Age*Height, Age*Weight, and Height*Weight. If N is set to a value greater than the number of variables in the model, N is effectively set to the number of variables.

Create Polynomial Effects of the Nth Order

1   Select one variable in the **Variables** box.

2   Specify higher-degree crossings by adjusting the number in the **N** field.

3   Click **Polynomial Order=N** to add the polynomial effects to the **Model effects** box.

For example, if you select the Age and Height variables and then you specify 3 in the **N** field, when you click **Polynomial Order=N**, these model effects are created: Age, Age*Age, Age*Age*Age, Height, Height*Height, and Height*Height*Height.

Specifying Model Effects for Zero-Inflated Models

These options are available if you selected **Zero-inflated negative binomial** or **Zero-inflated Poisson** as the distribution on the **Data** tab.

You must choose the type of model that you want to create:

- an intercept-only model.

- a model that includes effects from the main model. You define these model effects by using the model builder.

- a custom model. You specify these effects in the **Enter a custom model** text box. If you specify multiple effects, use a space between each effect.

If you choose to include effects in the zero-inflated models, specify the link function for those effects.

Setting Options

Option	Description
Methods	
Dispersion	
Adjust for overdispersion	adjusts the parameter covariance matrix and the likelihood function by a scale parameter. For the dispersion parameter, you can select a Pearson estimate or a deviance estimate. To define the subpopulations for calculating the Pearson and deviance chi-square goodness-of-fit tests, assign one or more variables to the role.
	Note: This option is available only for binomial and multinomial distributions.

Option	Description
Estimate dispersion parameter	enables you to specify a fixed dispersion parameter for those distributions that have a dispersion parameter. By default, this parameter is estimated.
	Note: This option is not available for binomial and multinomial distributions, but it is available for the other distribution types.

Optimization	
Maximum number of iterations	specifies the maximum number of iterations to perform for the selected optimization technique.

Statistics

You can select the statistics to include in the output.

Here are the additional statistics that you can include:

- type 1 (sequential) analysis
- type 3 analysis
- Wald statistics for Type 3 contrasts
- confidence intervals, such as Profile likelihood confidence intervals and Wald confidence intervals
- correlations of parameter estimates
- covariances of parameter estimates
- observation statistics, such as influence diagnostics, predicted values and confidence intervals, and residuals
- multiple comparisons for classification effects
- exact tests, which are available only for binomial distributions with a logit link function or a Poisson distribution with a log link function.

Plots

Option	Description
	You can select the plots to display in the output. If you choose to display multiple plots, you can display these plots individually or as a panel.
	Here are some plots that you can include in your results:
	■ predicted plots
	■ influence plots, such as Cook's D by observation number and DFBETA by observation number
	■ plots of residuals, deviance residuals, standardized deviance residuals, Pearson residuals, standardized Pearson residuals, standardized Pearson residuals, and likelihood residuals.

Setting the Output Options

You can specify whether to create an output data set. You can also specify the values to include in the output data set. You can include predicted values, residuals, influence statistics, and the standard error of the linear predictor in the output data set.

Appendix 1

Input Data Sets for Task Examples

About the Task Data Sets

To complete some of the examples in the task documentation, you might need to create one or more data sets. This appendix provides the SAS code that you need. To create these data sets, copy and paste this code (from the HTML version of *SAS Studio: User's Guide*) into a **Program** tab in SAS Studio and click 🏃.

FITNESS Data set

To create the Fitness data set, enter this code into a **Program** tab:

```
data Fitness;
   input Age Weight Oxygen RunTime @@;
   datalines;
44 89.47 44.609 11.37
40 75.07 45.313 10.07
44 85.84 54.297 8.65
42 68.15 59.571 8.17
38 89.02 49.874 .
47 77.45 44.811 11.63
40 75.98 45.681 11.95
43 81.19 49.091 10.85
44 81.42 39.442 13.08
38 81.87 60.055 8.63
44 73.03 50.541 10.13
45 87.66 37.388 14.03
45 66.45 44.754 11.12
47 79.15 47.273 10.60
54 83.12 51.855 10.33
49 81.42 49.156 8.95
51 69.63 40.836 10.95
51 77.91 46.672 10.00
48 91.63 46.774 10.25
49 73.37 . 10.08
57 73.37 39.407 12.63
54 79.38 46.080 11.17
52 76.32 45.441 9.63
50 70.87 54.625 8.92
51 67.25 45.118 11.08
54 91.63 39.203 12.88
51 73.71 45.790 10.47
57 59.08 50.545 9.93
49 76.32 . .
48 61.24 47.920 11.50
52 82.78 47.467 10.50
;
```

GETSTARTED Data Set

To create the getStarted data set, enter this code into a **Program** tab:

```
data getStarted;
   input C1-C5 Y Total;
   datalines;
0 3 1 1 3 2 28.361
```

```
2 3 0 3 1 2 39.831
1 3 2 2 2 1 17.133
1 2 0 0 3 2 12.769
0 2 1 0 1 1 29.464
0 2 1 0 2 1 4.152
1 2 1 0 1 0 0.000
0 2 1 1 2 1 20.199
1 2 0 0 1 0 0.000
0 1 1 3 3 2 53.376
2 2 2 2 1 1 31.923
0 3 2 0 3 2 37.987
2 2 2 0 0 1 1.082
0 2 0 2 0 1 6.323
1 3 0 0 0 0 0.000
1 2 1 2 3 2 4.217
0 1 2 3 1 1 26.084
1 1 0 0 1 0 0.000
1 3 2 2 2 0 0.000
2 1 3 1 1 2 52.640
1 3 0 1 2 1 3.257
2 0 2 3 0 5 88.066
2 2 2 1 0 1 15.196
3 1 3 1 0 1 11.955
3 1 3 1 2 3 91.790
3 1 1 2 3 7 232.417
3 1 1 1 0 1 2.124
3 1 0 0 0 2 32.762
3 1 2 3 0 1 25.415
2 2 0 1 2 1 42.753
3 3 2 2 3 1 23.854
2 0 0 2 3 2 49.438
1 0 0 2 3 4 105.449
0 0 2 3 0 6 101.536
0 3 1 0 0 0 0.000
3 0 1 0 1 1 5.937
2 0 0 0 3 2 53.952
1 0 1 0 3 2 23.686
1 1 3 1 1 1 0.287
2 1 3 0 3 7 281.551
1 3 2 1 1 0 0.000
2 1 0 0 1 0 0.000
0 0 1 1 2 3 93.009
0 1 0 1 0 2 25.055
1 2 2 2 3 1 1.691
0 3 2 3 1 1 10.719
3 3 0 3 3 1 19.279
2 0 0 2 1 2 40.802
```

```
2 2 3 0 3 3 72.924
0 2 0 3 0 1 10.216
3 0 1 2 2 2 87.773
2 1 2 3 1 0 0.000
3 2 0 3 1 0 0.000
3 0 3 0 0 2 62.016
1 3 2 2 1 3 36.355
2 3 2 0 3 1 23.190
1 0 1 2 1 1 11.784
2 1 2 2 2 5 204.527
3 0 1 1 2 5 115.937
0 1 1 3 2 1 44.028
2 2 1 3 1 4 52.247
1 1 0 0 1 1 17.621
3 3 1 2 1 2 10.706
2 2 0 2 3 3 81.506
0 1 0 0 2 2 81.835
0 1 2 0 1 2 20.647
3 2 2 2 0 1 3.110
2 2 3 0 0 1 13.679
1 2 2 3 2 1 6.486
3 3 2 2 1 2 30.025
0 0 3 1 3 6 202.172
3 2 3 1 2 3 44.221
0 3 0 0 0 1 27.645
3 3 3 0 3 2 22.470
2 3 2 0 2 0 0.000
1 3 0 2 0 1 1.628
1 3 1 0 2 0 0.000
3 2 3 3 0 1 20.684
3 1 0 2 0 4 108.000
0 1 2 2 1 1 4.615
0 2 3 2 2 1 12.461
0 3 2 0 1 3 53.798
2 1 1 2 0 1 36.320
1 0 3 0 0 0 0.000
0 0 3 2 0 1 19.902
0 2 3 1 0 0 0.000
2 2 2 1 3 2 31.815
3 3 3 0 0 0 0.000
2 2 1 3 3 2 17.915
0 2 3 2 3 2 69.315
1 3 1 2 1 0 0.000
3 0 1 1 1 4 94.050
2 1 1 1 3 6 242.266
0 2 0 3 2 1 40.885
2 0 1 1 2 2 74.708
```

```
2 2 2 2 3 2 50.734
1 0 2 2 1 3 35.950
1 3 3 1 1 1 2.777
3 1 2 1 3 5 118.065
0 3 2 1 2 0 0.000
;
```

GREENE Data Set

To create the Greene data set, enter this code into a **Program** tab:

```
data greene;
    input firm year production cost @@;
datalines;
1 1955   5.36598   1.14867  1 1960    6.03787    1.45185
1 1965   6.37673   1.52257  1 1970    6.93245    1.76627
2 1955   6.54535   1.35041  2 1960    6.69827    1.71109
2 1965   7.40245   2.09519  2 1970    7.82644    2.39480
3 1955   8.07153   2.94628  3 1960    8.47679    3.25967
;
```

IN Data Set

To create the In data set, enter this code into a **Program** tab:

```
data in;
   label q = "Quantity"
         p = "Price"
         s = "Price of Substitutes"
         y = "Income"
         u = "Unit Cost";
   drop i e1 e2;
   p = 0; q = 0;
   do i = 1 to 60;
      y = 1 + .05*i  + .15*rannor(123);
      u = 2              + .05*rannor(123) + .05*rannor(123);
      s = 4 - .001*(i-10)*(i-110) + .5*rannor(123);
      e1 = .15 * rannor(123);
      e2 = .15 * rannor(123);
      demandx = 1 + .3 * y + .35 * s + e1;
      supplyx = -1 - 1 * u + e2 - .4*e1;
```

```
        q = 1.4/2.15 * demandx + .75/2.15 * supplyx;
        p = ( - q + supplyx ) / -1.4;
        output;
    end;
run;
```

LONG97DATA Data Set

To create the In data set, enter this code into a **Program** tab:

```
data long97data;
    input fem ment phd mar kid5 art lnart;
datalines;
0    7.99999860    1.38000000  1  2  3    1.25276290
0    6.99999950    4.29000000  0  0  0   -0.69314720
0   47.00000760    3.84999990  0  0  4    1.50407740
0   19.00000190    3.58999990  1  1  1    0.40546510
0    0.00000000    1.80999990  1  0  1    0.40546510
0    6.00000050    3.58999990  1  1  1    0.40546510
0    9.99999900    2.11999990  1  1  0   -0.69314720
0    1.99999990    4.29000000  1  0  0   -0.69314720
0    1.99999990    2.57999990  1  2  3    1.25276290
0    3.99999900    1.80000000  1  1  3    1.25276290
0    0.00000000    4.29000000  1  2  1    0.40546510
0    3.00000000    2.76000000  1  1  0   -0.69314720
0    9.99999900    3.41000010  1  1  1    0.40546510
0    6.99999950    4.34000020  1  3  2    0.91629080
0   15.00000100    3.84999990  1  2  5    1.70474800
0    1.99999990    2.09999990  1  0  2    0.91629080
0   13.00000000    4.29000000  1  0  2    0.91629080
0   15.00000100    4.29000000  0  0  1    0.40546510
0    4.99999810    2.26000000  1  1  0   -0.69314720
0    6.00000050    2.09999990  0  0  0   -0.69314720
0   12.00000000    2.26000000  1  0  3    1.25276290
0   15.99999810    3.84999990  1  1  6    1.87180220
0    6.99999950    4.29000000  0  0  4    1.50407740
0    6.00000050    1.80000000  1  2  2    0.91629080
0    1.99999990    2.26000000  0  0  2    0.91629080
0    0.00000000    2.09999990  0  0  0   -0.69314720
0   30.00000190    4.29000000  1  0  4    1.50407740
0    9.99999900    4.29000000  1  2  1    0.40546510
0    1.99999990    2.09999990  1  0  1    0.40546510
0    0.99999990    3.58999990  1  0  7    2.01490310
0    3.00000000    3.42000010  1  1  2    0.91629080
```

```
0    9.99999900   4.29000000  1  2  2    0.91629080
0    9.99999900   4.29000000  0  0  2    0.91629080
0    0.99999990   3.33999990  1  2  0   -0.69314720
0    1.99999990   4.29000000  0  0  0   -0.69314720
0   10.99999710   4.29000000  1  0  1    0.40546510
0    4.99999810   3.61999990  1  0  4    1.50407740
0    0.00000000   4.29000000  1  3  1    0.40546510
0    3.99999900   4.34000020  1  1  1    0.40546510
0    1.99999990   1.25000000  1  1  2    0.91629080
0   19.00000190   4.34000020  0  0  7    2.01490310
0    3.00000000   1.67000000  1  3  1    0.40546510
0    0.00000000   3.47000000  0  0  0   -0.69314720
0    0.99999990   2.26000000  1  1  1    0.40546510
0    0.99999990   1.80000000  1  0  1    0.40546510
0   17.00000000   4.34000020  1  2  2    0.91629080
0    3.00000000   3.58999990  0  0  2    0.91629080
0    0.99999990   1.75000000  1  2  1    0.40546510
0    6.00000050   4.29000000  0  0  1    0.40546510
0    0.00000000   2.09999990  1  1  0   -0.69314720
0   15.00000100   4.29000000  1  2  0   -0.69314720
0    0.00000000   2.09999990  1  1  0   -0.69314720
0   26.99999810   3.31999990  1  2  2    0.91629080
0    4.99999810   4.34000020  1  0  2    0.91629080
0    6.99999950   3.41000010  0  0  4    1.50407740
0    0.00000000   4.29000000  1  0  1    0.40546510
0   10.99999710   3.19000010  1  0  2    0.91629080
0   13.00000000   4.29000000  1  0  2    0.91629080
0    3.99999900   1.74000000  1  2  1    0.40546510
0    3.99999900   2.76000000  0  0  1    0.40546510
0   26.99999810   3.58999990  1  1  7    2.01490310
0    9.99999900   1.80999990  1  0  4    1.50407740
0   13.00000000   4.29000000  1  1  2    0.91629080
0    0.99999990   4.29000000  1  1  1    0.40546510
0    6.00000050   2.76000000  0  0  1    0.40546510
0    6.00000050   3.47000000  0  0  6    1.87180220
0    4.99999810   2.50000000  1  2  2    0.91629080
0    1.99999990   1.25000000  1  0  5    1.70474800
0   13.99999710   3.58999990  1  1  3    1.25276290
0    0.00000000   2.09999990  1  1  0   -0.69314720
0   12.00000000   3.58999990  1  0  1    0.40546510
0    6.99999950   3.58999990  1  3  0   -0.69314720
0    3.00000000   1.75000000  1  0  1    0.40546510
0    1.99999990   1.75000000  1  2  1    0.40546510
0    1.99999990   3.58999990  1  1  1    0.40546510
0    1.99999990   4.29000000  0  0  1    0.40546510
0    0.00000000   4.29000000  0  0  0   -0.69314720
0    0.00000000   2.09999990  1  1  0   -0.69314720
```

```
0    0.00000000   2.60999990  1  0  3   1.25276290
0   30.00000190   4.29000000  1  0  5   1.70474800
0   21.00000000   1.74000000  1  0 16   2.80336050
0    4.99999810   2.76000000  1  0  1   0.40546510
0    9.00000000   4.29000000  0  0  0  -0.69314720
0    7.99999860   2.76000000  1  2  1   0.40546510
0   25.00000000   4.29000000  1  2  3   1.25276290
0    0.00000000   3.47000000  1  1  5   1.70474800
0    4.99999810   2.57999990  1  2  0  -0.69314720
0    0.99999990   2.14000010  1  0  0  -0.69314720
0    4.99999810   2.26000000  0  0  0  -0.69314720
0    0.00000000   4.29000000  1  2  3   1.25276290
0   15.00000100   4.29000000  1  0  3   1.25276290
0   30.00000190   4.29000000  0  0  3   1.25276290
0    1.99999990   2.20000000  1  0  0  -0.69314720
0    6.00000050   1.80000000  1  2  3   1.25276290
0    0.00000000   2.09999990  1  2  1   0.40546510
0   13.00000000   4.29000000  1  1  1   0.40546510
0    0.00000000   4.29000000  0  0  0  -0.69314720
0   12.00000000   2.09999990  1  1  0  -0.69314720
0   30.00000190   4.29000000  1  2  2   0.91629080
0    4.99999810   1.80999990  1  1  1   0.40546510
0    9.99999900   4.34000020  1  0  1   0.40546510
0    4.99999810   4.29000000  1  1  0  -0.69314720
0    3.99999900   2.50000000  1  2  1   0.40546510
0   13.00000000   2.05000000  1  2  4   1.50407740
0    7.99999860   3.47000000  1  0  3   1.25276290
0    6.00000050   2.60999990  1  1  1   0.40546510
0    6.00000050   4.29000000  1  2  1   0.40546510
0   25.00000000   4.29000000  0  0  2   0.91629080
0    1.99999990   4.29000000  1  1  2   0.91629080
0    9.00000000   4.34000020  1  0  6   1.87180220
0    9.99999900   2.11999990  1  1  0  -0.69314720
0    3.00000000   2.76000000  1  0  2   0.91629080
0    1.99999990   4.29000000  1  2  0  -0.69314720
0    0.00000000   2.50000000  1  0  1   0.40546510
0    6.00000050   4.34000020  1  0  5   1.70474800
0    7.99999860   2.76000000  1  1  2   0.91629080
0    9.99999900   3.19000010  1  1  2   0.91629080
0    7.99999860   4.61999990  0  0  3   1.25276290
0    6.00000050   3.15000010  1  2  0  -0.69314720
0   21.00000000   2.55000000  1  1  4   1.50407740
0    3.99999900   1.52000000  1  0  0  -0.69314720
0    1.99999990   1.72000000  1  2  4   1.50407740
0    0.99999990   1.78000000  1  1  2   0.91629080
0   17.00000000   2.85999990  1  1  1   0.40546510
0   30.00000190   4.61999990  1  2  0  -0.69314720
```

```
0    4.99999810    4.13999990   0  0   1    0.40546510
0   13.00000000    2.96000000   1  1   6    1.87180220
0   10.99999710    2.55000000   1  0   1    0.40546510
0   19.00000190    2.21000000   1  1   0   -0.69314720
0    4.99999810    3.08999990   1  0   3    1.25276290
0   66.00000760    4.54000000   1  2   4    1.50407740
0    0.00000000    1.78000000   1  0   3    1.25276290
0    3.00000000    2.21000000   1  3   0   -0.69314720
0    3.00000000    2.39000010   1  1   1    0.40546510
0    0.00000000    2.96000000   1  0   0   -0.69314720
0    7.99999860    2.51000000   1  1   0   -0.69314720
0    0.00000000    1.97000000   1  2   2    0.91629080
0    4.99999810    4.13999990   0  0   0   -0.69314720
0   29.00000000    4.25000000   1  1   4    1.50407740
0   10.99999710    2.85999990   1  2   0   -0.69314720
0   22.99999620    2.96000000   1  1   9    2.25129180
0   45.99999240    2.96000000   1  2   2    0.91629080
0    7.99999860    4.61999990   1  1   6    1.87180220
0    4.99999810    3.69000010   1  2   0   -0.69314720
0    0.99999990    3.15000010   1  1   0   -0.69314720
0    9.00000000    4.61999990   0  0   2    0.91629080
0    1.99999990    3.35999990   1  0   0   -0.69314720
0    6.99999950    3.69000010   0  0   4    1.50407740
0   10.99999710    3.54000000   1  0   1    0.40546510
0   56.99999620    2.96000000   1  1   4    1.50407740
0   15.99999810    2.55999990   1  1   1    0.40546510
0    0.00000000    2.31999990   1  0   0   -0.69314720
0    3.00000000    2.31999990   1  0   0   -0.69314720
0    0.99999990    0.92000000   1  2   0   -0.69314720
0    9.99999900    4.54000000   1  0   0   -0.69314720
0   10.99999710    4.54000000   1  3   0   -0.69314720
0   10.99999710    1.76000000   1  1   5    1.70474800
0   15.99999810    2.55999990   0  0   1    0.40546510
0    4.99999810    2.39000010   1  1   0   -0.69314720
0    9.00000000    3.40000010   1  0   2    0.91629080
0   19.99999620    2.86999990   1  2   2    0.91629080
0   21.00000000    4.54000000   1  2   4    1.50407740
0    4.99999810    2.82999990   1  0   4    1.50407740
0    6.00000050    1.67999990   1  0   0   -0.69314720
0   12.00000000    3.54000000   0  0   3    1.25276290
0    0.00000000    1.76000000   1  0   2    0.91629080
0   10.99999710    3.15000010   1  3   1    0.40546510
0    3.00000000    2.51000000   1  0   0   -0.69314720
0   15.99999810    3.69000010   1  1   0   -0.69314720
0   10.99999710    1.76000000   1  1   4    1.50407740
0    4.99999810    1.86000000   1  1  12    2.52572870
0    0.99999990    2.76000000   1  3   0   -0.69314720
```

0	15.99999810	4.61999990	1	1	1	0.40546510
0	12.00000000	4.25000000	1	0	5	1.70474800
0	10.99999710	2.54000000	1	0	4	1.50407740
0	0.00000000	2.20000000	1	1	0	-0.69314720
0	0.00000000	1.76000000	1	0	2	0.91629080
0	3.00000000	2.85999990	1	3	3	1.25276290
0	13.00000000	3.40000010	1	2	0	-0.69314720
0	45.00000000	4.54000000	1	1	1	0.40546510
0	47.00000760	1.86000000	1	1	9	2.25129180
0	6.99999950	1.52000000	1	0	0	-0.69314720
0	6.99999950	2.55999990	1	1	2	0.91629080
0	19.00000190	2.21000000	1	0	0	-0.69314720
0	9.00000000	3.69000010	1	0	7	2.01490310
0	76.99998470	1.78000000	1	1	1	0.40546510
0	0.00000000	1.17999990	1	1	0	-0.69314720
0	3.99999900	2.00000000	1	0	1	0.40546510
0	19.00000190	2.21000000	1	0	0	-0.69314720
0	12.00000000	4.13999990	1	2	0	-0.69314720
0	0.99999990	2.85999990	1	3	0	-0.69314720
0	17.00000000	2.85999990	1	1	1	0.40546510
0	6.00000050	2.54000000	0	0	7	2.01490310
0	3.99999900	2.85999990	1	1	0	-0.69314720
0	6.00000050	2.52000000	0	0	4	1.50407740
0	3.00000000	1.52000000	1	1	2	0.91629080
0	4.99999810	3.08999990	1	1	3	1.25276290
0	3.00000000	1.17999990	1	1	0	-0.69314720
0	3.00000000	1.42000000	1	0	0	-0.69314720
0	15.00000100	4.61999990	0	0	7	2.01490310
0	0.00000000	2.96000000	1	2	1	0.40546510
0	9.99999900	4.54000000	0	0	2	0.91629080
0	41.99999620	4.54000000	0	0	7	2.01490310
0	3.00000000	2.51000000	1	2	1	0.40546510
0	6.99999950	3.15000010	1	1	2	0.91629080
0	0.00000000	2.50000000	1	0	1	0.40546510
0	6.00000050	2.96000000	1	3	1	0.40546510
0	3.99999900	1.67999990	1	0	0	-0.69314720
0	0.00000000	1.22000000	1	1	1	0.40546510
0	1.99999990	1.52000000	1	0	1	0.40546510
0	4.99999810	2.21000000	1	1	0	-0.69314720
0	0.99999990	3.92000010	1	1	0	-0.69314720
0	13.00000000	4.54000000	1	0	5	1.70474800
0	0.00000000	1.17999990	1	0	2	0.91629080
0	26.00000000	3.69000010	1	0	3	1.25276290
0	0.99999990	1.72000000	1	0	2	0.91629080
0	25.00000000	2.57999990	1	1	5	1.70474800
0	3.00000000	1.52000000	0	0	3	1.25276290
0	47.00000760	1.86000000	1	1	4	1.50407740

```
0    3.99999900    2.50000000    1   0   5    1.70474800
0    0.99999990    4.61999990    1   0   1    0.40546510
0    6.99999950    1.40000000    1   2   0   -0.69314720
0    4.99999810    4.54000000    1   0   3    1.25276290
0   26.99999810    1.67999990    1   1   0   -0.69314720
0    0.99999990    2.82999990    1   0   2    0.91629080
0    4.99999810    3.35999990    0   0   1    0.40546510
0    0.00000000    1.97000000    1   1   7    2.01490310
0   12.00000000    3.40000010    0   0   1    0.40546510
0    3.99999900    1.74000000    1   1   1    0.40546510
0    0.00000000    2.96000000    0   0   0   -0.69314720
0   25.00000000    2.57999990    1   1   4    1.50407740
0    3.00000000    4.54000000    1   1   1    0.40546510
0    1.99999990    3.15000010    1   1   0   -0.69314720
0   12.00000000    2.96000000    1   2   0   -0.69314720
0   15.99999810    3.54000000    1   2   2    0.91629080
0    4.99999810    2.96000000    1   1   3    1.25276290
0   12.00000000    4.25000000    1   0   1    0.40546510
0    4.99999810    2.55999990    1   1   2    0.91629080
0    4.99999810    1.86000000    1   0   1    0.40546510
0    3.00000000    4.61999990    1   1   1    0.40546510
0    3.99999900    2.85999990    0   0   2    0.91629080
0    4.99999810    3.15000010    1   1   5    1.70474800
0    0.00000000    2.51000000    1   0   2    0.91629080
0   26.99999810    3.15000010    1   1   5    1.70474800
0    4.99999810    2.51000000    1   0   2    0.91629080
0    0.00000000    1.52000000    1   1   0   -0.69314720
0   17.99999810    4.29000000    0   0   6    1.87180220
0    4.99999810    4.29000000    1   1   4    1.50407740
0    7.99999860    4.29000000    1   2   2    0.91629080
0    3.99999900    2.09999990    1   0   2    0.91629080
0   35.00000760    4.29000000    1   1  12    2.52572870
0    4.99999810    4.29000000    0   0   2    0.91629080
0    9.00000000    3.58999990    1   1   1    0.40546510
0    6.00000050    4.29000000    1   0   1    0.40546510
0   24.00000190    4.29000000    1   0   2    0.91629080
0    0.00000000    2.09999990    1   1   0   -0.69314720
0    4.99999810    1.80999990    1   1   0   -0.69314720
0   19.00000190    4.29000000    0   0   7    2.01490310
0    3.00000000    4.29000000    0   0   1    0.40546510
0    1.99999990    4.29000000    1   1   1    0.40546510
0    3.99999900    1.25000000    1   0   3    1.25276290
0    9.99999900    3.58999990    1   0   1    0.40546510
0    7.99999860    2.09999990    1   1   1    0.40546510
0    1.99999990    4.29000000    1   0   4    1.50407740
0   19.00000190    4.29000000    1   1   3    1.25276290
0    0.99999990    3.47000000    1   1   1    0.40546510
```

```
0    3.00000000    3.19000010  1  0  1    0.40546510
0    4.99999810    3.19000010  1  2  1    0.40546510
0    0.00000000    2.09999990  1  0  1    0.40546510
0   21.00000000    3.58999990  1  1  5    1.70474800
0   13.99999710    4.29000000  1  1  1    0.40546510
0    0.00000000    2.09999990  1  0  0   -0.69314720
0    6.99999950    2.76000000  1  1  1    0.40546510
0    4.99999810    2.60999990  1  0  3    1.25276290
0   22.00000000    4.29000000  1  0  4    1.50407740
0   19.99999620    3.41000010  0  0  7    2.01490310
0   38.00000380    4.29000000  1  0  3    1.25276290
0   19.00000190    4.29000000  0  0  4    1.50407740
0    3.00000000    2.26000000  0  0  2    0.91629080
0   15.99999810    3.58999990  0  0  3    1.25276290
0    1.99999990    4.29000000  1  1  0   -0.69314720
0    3.00000000    2.05000000  0  0  1    0.40546510
0   12.00000000    4.29000000  1  2  1    0.40546510
0   35.00000760    4.29000000  0  0  0   -0.69314720
0   19.00000190    4.29000000  1  0  0   -0.69314720
0    1.99999990    4.29000000  0  0  0   -0.69314720
0    6.99999950    2.76000000  0  0  1    0.40546510
0    3.99999900    2.09999990  1  2  3    1.25276290
0   13.99999710    2.35999990  0  0  1    0.40546510
0    9.00000000    4.29000000  0  0  0   -0.69314720
0    7.99999860    3.58999990  0  0  0   -0.69314720
0   12.00000000    4.29000000  1  1  2    0.91629080
0    3.00000000    4.29000000  1  1  0   -0.69314720
0    0.99999990    4.29000000  1  0  1    0.40546510
0    9.99999900    4.29000000  0  0  1    0.40546510
0   21.00000000    3.41000010  0  0  4    1.50407740
0   13.00000000    4.29000000  1  1  0   -0.69314720
0   17.00000000    4.29000000  1  0  2    0.91629080
0    4.99999810    4.29000000  0  0  0   -0.69314720
0    1.99999990    2.14000010  0  0  0   -0.69314720
0    7.99999860    4.29000000  1  0  0   -0.69314720
0    1.99999990    2.50000000  1  1  2    0.91629080
0    4.99999810    2.60999990  1  1  3    1.25276290
0    0.99999990    1.80999990  1  2  0   -0.69314720
0    1.99999990    2.26000000  1  1  0   -0.69314720
0    3.99999900    3.61999990  1  2  1    0.40546510
0    9.00000000    4.29000000  1  2  1    0.40546510
0    9.00000000    4.29000000  0  0  0   -0.69314720
0   17.00000000    4.29000000  0  0  2    0.91629080
0   24.00000190    4.29000000  1  2  0   -0.69314720
0    3.99999900    3.47000000  1  0  4    1.50407740
0   13.99999710    4.29000000  1  1  2    0.91629080
0    4.99999810    3.58999990  1  1  1    0.40546510
```

```
0    9.99999900    1.80999990  0  0  1    0.40546510
0   17.99999810    4.29000000  1  1  3    1.25276290
0    0.99999990    4.29000000  1  1  0   -0.69314720
0    0.00000000    2.09999990  1  0  1    0.40546510
0    3.99999900    2.15000010  1  2  0   -0.69314720
0    7.99999860    4.29000000  1  0  0   -0.69314720
0    0.99999990    2.26000000  1  0  1    0.40546510
0    7.99999860    4.29000000  1  0  0   -0.69314720
0   13.00000000    4.29000000  1  2  0   -0.69314720
0    4.99999810    3.58999990  1  2  0   -0.69314720
0    6.99999950    3.41000010  1  2  0   -0.69314720
0    6.99999950    3.58999990  0  0  3    1.25276290
0    4.99999810    3.61999990  1  1  3    1.25276290
0    0.00000000    2.09999990  1  3  1    0.40546510
0   13.99999710    4.29000000  0  0  1    0.40546510
0    6.00000050    4.29000000  1  0  0   -0.69314720
0    4.99999810    2.26000000  1  1  1    0.40546510
0    7.99999860    2.76000000  1  0  3    1.25276290
0    3.99999900    2.60999990  1  1  1    0.40546510
0    9.00000000    4.29000000  0  0  2    0.91629080
0    3.00000000    2.09999990  1  1  2    0.91629080
0    1.99999990    3.47000000  1  1  0   -0.69314720
0    0.00000000    2.09999990  0  0  0   -0.69314720
0   10.99999710    4.29000000  0  0  1    0.40546510
0    9.00000000    4.29000000  0  0  1    0.40546510
0    7.99999860    2.96000000  1  0  1    0.40546510
0    1.99999990    2.96000000  1  0  2    0.91629080
0   17.99999810    4.61999990  1  0  3    1.25276290
0    3.00000000    1.42000000  1  0  3    1.25276290
0    1.99999990    4.54000000  1  0  1    0.40546510
0    6.99999950    2.52000000  1  0  0   -0.69314720
0    9.99999900    4.54000000  1  1  2    0.91629080
0   15.99999810    2.00000000  1  1  1    0.40546510
0    4.99999810    2.54000000  1  0  0   -0.69314720
0    0.00000000    2.50000000  1  0  2    0.91629080
0    7.99999860    4.54000000  0  0  0   -0.69314720
0    9.99999900    3.35999990  0  0  1    0.40546510
0   17.99999810    3.40000010  1  1  0   -0.69314720
0   12.00000000    1.67999990  1  1  1    0.40546510
0    9.00000000    2.00000000  0  0  1    0.40546510
0   39.00000000    2.85999990  1  0  1    0.40546510
0   17.99999810    4.61999990  1  0  1    0.40546510
0   15.00000100    4.13999990  1  0  1    0.40546510
0   15.99999810    4.13999990  1  0  2    0.91629080
0    1.99999990    2.96000000  0  0  0   -0.69314720
0    6.99999950    2.82999990  1  2  3    1.25276290
0   24.00000190    2.55000000  1  0  2    0.91629080
```

```
0    7.99999860    1.67999990   1   0   1     0.40546510
0    6.99999950    2.00000000   0   0   1     0.40546510
0   10.99999710    2.00000000   1   1   1     0.40546510
0    0.00000000    2.96000000   1   1   4     1.50407740
0    3.99999900    1.50500000   1   0   4     1.50407740
0   21.00000000    3.54000000   1   0   2     0.91629080
0    6.00000050    3.40000010   0   0   1     0.40546510
0    6.00000050    4.61999990   0   0   0    -0.69314720
0    7.99999860    2.82999990   0   0   3     1.25276290
0    3.99999900    2.54000000   1   0   1     0.40546510
0   12.00000000    2.86999990   0   0   1     0.40546510
0   15.00000100    1.86000000   0   0   1     0.40546510
0    0.00000000    3.92000010   1   1   1     0.40546510
0    4.99999810    3.69000010   1   1   1     0.40546510
0   15.00000100    2.85999990   1   2   4     1.50407740
0    4.99999810    4.54000000   0   0   0    -0.69314720
0    7.99999860    4.61999990   0   0   0    -0.69314720
0   13.00000000    2.85999990   1   1   0    -0.69314720
0    1.99999990    3.40000010   0   0   1     0.40546510
0    6.00000050    2.57999990   1   2   4     1.50407740
0   19.99999620    4.25000000   1   0   3     1.25276290
0    6.99999950    1.76000000   1   1   2     0.91629080
0    6.99999950    2.85999990   1   3   0    -0.69314720
0   15.99999810    3.69000010   1   2   4     1.50407740
0   13.00000000    3.40000010   1   0   0    -0.69314720
0    0.00000000    3.40000010   1   2   0    -0.69314720
0    0.99999990    4.54000000   1   2   0    -0.69314720
0   12.00000000    2.86999990   0   0   3     1.25276290
0    6.99999950    1.76000000   1   0   4     1.50407740
0    3.99999900    4.25000000   0   0   1     0.40546510
0    0.00000000    3.92000010   0   0   3     1.25276290
0    3.99999900    3.35999990   1   0   2     0.91629080
0    3.99999900    2.31999990   1   0   1     0.40546510
0   36.99999240    4.54000000   0   0   2     0.91629080
0   22.99999620    3.35999990   1   1   0    -0.69314720
0    7.99999860    2.00000000   1   3   2     0.91629080
0    7.99999860    3.92000010   1   1   1     0.40546510
0    1.99999990    3.92000010   0   0   1     0.40546510
0    0.00000000    3.35999990   0   0   1     0.40546510
0    0.99999990    1.78000000   1   0   1     0.40546510
0   21.00000000    3.54000000   1   1   1     0.40546510
0    9.99999900    3.92000010   0   0   2     0.91629080
0    7.99999860    2.31999990   0   0   3     1.25276290
0   33.99999240    1.67999990   1   0   1     0.40546510
0   13.99999710    3.08999990   0   0   2     0.91629080
0    6.00000050    2.57999990   1   1   0    -0.69314720
0   13.99999710    3.40000010   0   0   6     1.87180220
```

```
0   12.00000000    2.86999990   0   0   3    1.25276290
0   17.99999810    4.61999990   0   0   4    1.50407740
0    3.00000000    2.96000000   1   3   0   -0.69314720
0    6.00000050    1.86000000   1   2   2    0.91629080
0    1.99999990    1.22000000   1   1   0   -0.69314720
0   10.99999710    2.51000000   1   2   6    1.87180220
0   10.99999710    2.51000000   1   0   8    2.14006610
0    4.99999810    3.69000010   1   0   2    0.91629080
0    7.99999860    2.96000000   1   0   3    1.25276290
0    3.99999900    1.78000000   1   0   1    0.40546510
0    0.99999990    1.22000000   1   1   1    0.40546510
0    7.99999860    2.85999990   1   0   0   -0.69314720
0   19.00000190    3.69000010   0   0   7    2.01490310
0    1.99999990    2.11999990   0   0   0   -0.69314720
0    9.99999900    2.52000000   1   1   2    0.91629080
0    3.99999900    2.31999990   1   0   1    0.40546510
0    3.00000000    4.61999990   1   0   2    0.91629080
0    1.99999990    3.54000000   1   2   4    1.50407740
0    0.99999990    2.50000000   1   0   0   -0.69314720
0    4.99999810    1.67999990   0   0   0   -0.69314720
0    1.99999990    3.40000010   1   1   1    0.40546510
0    4.99999810    3.92000010   0   0   0   -0.69314720
0   52.99998090    4.54000000   1   1   5    1.70474800
0   54.99998860    4.54000000   0   0   2    0.91629080
0    0.00000000    2.50000000   0   0   0   -0.69314720
0   10.99999710    4.54000000   1   0   2    0.91629080
0   25.00000000    3.54000000   0   0   3    1.25276290
0    4.99999810    1.52000000   1   0   2    0.91629080
0    1.99999990    3.92000010   1   2   0   -0.69314720
0    7.99999860    4.61999990   1   0   2    0.91629080
0    0.00000000    3.92000010   1   0   3    1.25276290
0    4.99999810    2.31999990   1   2   0   -0.69314720
0    7.99999860    2.96000000   1   2   2    0.91629080
0    6.99999950    2.85999990   1   2   1    0.40546510
0   13.99999710    1.95000000   1   1   3    1.25276290
0    1.99999990    3.92000010   1   2   1    0.40546510
0    4.99999810    2.86999990   0   0   1    0.40546510
0   24.00000190    3.69000010   1   0   3    1.25276290
0    3.00000000    3.69000010   1   0   0   -0.69314720
0    3.99999900    2.39000010   0   0   2    0.91629080
0    3.00000000    1.95000000   1   2   3    1.25276290
0   10.99999710    3.35999990   0   0   5    1.70474800
0    3.99999900    2.39000010   1   1   2    0.91629080
0   26.00000000    3.69000010   0   0   2    0.91629080
0    3.99999900    2.96000000   1   1   5    1.70474800
0    1.99999990    3.21000000   1   0   0   -0.69314720
0    0.00000000    2.11999990   1   2   0   -0.69314720
```

```
0   52.99998090    4.54000000   1   1    2    0.91629080
0    3.99999900    2.31999990   1   0    3    1.25276290
0    6.00000050    2.54000000   1   1    3    1.25276290
0    1.99999990    2.85999990   1   1    1    0.40546510
0   13.99999710    3.47000000   1   0    3    1.25276290
0   15.00000100    2.86999990   0   0    4    1.50407740
0    3.99999900    2.31999990   1   0    0   -0.69314720
0   15.00000100    1.86000000   1   2    4    1.50407740
0    3.99999900    1.95000000   1   0    2    0.91629080
0    1.99999990    2.31999990   1   2    0   -0.69314720
0   19.99999620    4.25000000   1   2    1    0.40546510
0    3.99999900    1.97000000   0   0    6    1.87180220
0   41.99999620    1.86000000   1   0   19    2.97041440
0    7.99999860    3.69000010   1   1    0   -0.69314720
0    3.00000000    4.54000000   1   1    0   -0.69314720
0    9.00000000    3.54000000   1   0    4    1.50407740
0    9.99999900    4.54000000   1   1    1    0.40546510
0    6.99999950    2.85999990   1   2   11    2.44234700
0   25.00000000    3.35999990   1   0    0   -0.69314720
0    3.00000000    2.85999990   0   0    0   -0.69314720
0    3.99999900    2.96000000   1   2    0   -0.69314720
0   22.00000000    2.55999990   1   2    1    0.40546510
0    7.99999860    1.63000000   1   0    2    0.91629080
0   13.99999710    2.96000000   0   0    0   -0.69314720
0    0.00000000    2.96000000   1   0    1    0.40546510
0    7.99999860    1.63000000   1   1    1    0.40546510
0    0.00000000    2.96000000   0   0    0   -0.69314720
0   21.00000000    2.96000000   1   2    2    0.91629080
0    6.99999950    2.96000000   1   0    1    0.40546510
0   30.99998860    4.54000000   1   2    3    1.25276290
0    1.99999990    4.54000000   1   2    2    0.91629080
0    9.99999900    2.15000010   0   0    1    0.40546510
0    6.00000050    4.54000000   1   1    2    0.91629080
0   12.00000000    2.21000000   1   2    2    0.91629080
0    0.99999990    2.21000000   1   1    0   -0.69314720
0   29.00000000    4.54000000   0   0    2    0.91629080
0    9.99999900    2.21000000   1   1    3    1.25276290
0    6.00000050    2.21000000   1   0    0   -0.69314720
0   36.99999240    4.54000000   1   0    3    1.25276290
0    4.99999810    4.54000000   0   0    4    1.50407740
1    9.00000000    2.11999990   0   0    0   -0.69314720
1    1.99999990    2.11999990   1   0    0   -0.69314720
1    1.99999990    2.11999990   1   0    2    0.91629080
1    3.00000000    2.11999990   1   0    3    1.25276290
1    9.00000000    2.11999990   1   1    1    0.40546510
1   10.99999710    2.39000010   0   0    3    1.25276290
1    4.99999810    2.57999990   0   0    2    0.91629080
```

```
1    6.99999950    3.19000010  1  0  4    1.50407740
1    4.99999810    3.75000000  0  0  0   -0.69314720
1    3.00000000    3.58999990  0  0  1    0.40546510
1    1.99999990    2.57999990  0  0  2    0.91629080
1    0.00000000    3.75000000  0  0  3    1.25276290
1    4.99999810    3.75000000  1  0  0   -0.69314720
1    0.99999990    3.75000000  0  0  0   -0.69314720
1    3.99999900    1.22000000  1  2  2    0.91629080
1   13.99999710    3.75000000  0  0  0   -0.69314720
1    6.99999950    3.75000000  1  0  0   -0.69314720
1   26.00000000    3.75000000  0  0  0   -0.69314720
1    7.99999860    3.75000000  0  0  2    0.91629080
1    3.00000000    3.75000000  0  0  4    1.50407740
1    6.00000050    3.75000000  0  0  0   -0.69314720
1   21.00000000    3.75000000  0  0  1    0.40546510
1    7.99999860    3.75000000  1  0  2    0.91629080
1    3.99999900    3.75000000  0  0  0   -0.69314720
1    6.99999950    3.75000000  1  1  1    0.40546510
1    0.00000000    3.75000000  1  0  2    0.91629080
1    0.00000000    2.54000000  1  1  0   -0.69314720
1    0.99999990    2.54000000  1  0  0   -0.69314720
1    3.99999900    2.76000000  0  0  0   -0.69314720
1    1.99999990    4.13999990  0  0  0   -0.69314720
1    3.99999900    0.92000000  0  0  1    0.40546510
1    0.00000000    1.00500000  0  0  0   -0.69314720
1    6.00000050    3.08999990  0  0  2    0.91629080
1    3.00000000    3.08999990  1  1  2    0.91629080
1    0.00000000    1.79000000  0  0  0   -0.69314720
1    3.00000000    1.40000000  0  0  1    0.40546510
1    3.00000000    1.40000000  0  0  4    1.50407740
1    0.00000000    1.40000000  1  0  2    0.91629080
1    0.00000000    1.40000000  0  0  2    0.91629080
1   13.99999710    1.40000000  0  0  0   -0.69314720
1   12.00000000    1.40000000  0  0  1    0.40546510
1    1.99999990    1.40000000  0  0  1    0.40546510
1    6.00000050    2.00000000  0  0  1    0.40546510
1   10.99999710    4.34000020  1  0  2    0.91629080
1    6.00000050    4.34000020  1  0  4    1.50407740
1    0.00000000    1.75000000  0  0  0   -0.69314720
1   36.00000000    2.09999990  1  0  6    1.87180220
1    7.99999860    2.09999990  0  0  1    0.40546510
1    1.99999990    2.09999990  0  0  0   -0.69314720
1   10.99999710    3.58999990  1  0  2    0.91629080
1   10.99999710    3.58999990  1  2  2    0.91629080
1   17.99999810    3.58999990  0  0  0   -0.69314720
1    3.99999900    3.58999990  0  0  0   -0.69314720
1   17.99999810    3.58999990  1  0 10    2.35137530
```

```
1    7.99999860    3.58999990   0  0  1    0.40546510
1   19.00000190    3.41000010   1  0  3    1.25276290
1    3.99999900    3.41000010   1  0  4    1.50407740
1    4.99999810    3.41000010   1  2  0   -0.69314720
1    6.99999950    3.40000010   0  0  1    0.40546510
1    3.99999900    3.40000010   1  1  2    0.91629080
1    4.99999810    3.40000010   1  0  4    1.50407740
1   13.99999710    3.40000010   1  0  1    0.40546510
1    3.99999900    3.40000010   1  0  0   -0.69314720
1    3.99999900    2.52000000   0  0  3    1.25276290
1    0.00000000    2.52000000   0  0  1    0.40546510
1   15.00000100    3.69000010   1  1  2    0.91629080
1    0.99999990    3.69000010   1  0  1    0.40546510
1    3.00000000    3.69000010   1  1  0   -0.69314720
1   17.99999810    3.69000010   0  0  1    0.40546510
1   15.00000100    2.86999990   0  0  0   -0.69314720
1    4.99999810    2.86999990   1  0  1    0.40546510
1   17.99999810    2.86999990   1  0  1    0.40546510
1    9.00000000    2.86999990   0  0  2    0.91629080
1    6.00000050    2.86999990   0  0  4    1.50407740
1   15.00000100    2.86999990   1  0  1    0.40546510
1    3.99999900    3.35999990   0  0  2    0.91629080
1    1.99999990    3.35999990   1  2  0   -0.69314720
1    3.00000000    3.35999990   0  0  2    0.91629080
1    6.99999950    3.35999990   0  0  0   -0.69314720
1   15.99999810    4.54000000   1  0  4    1.50407740
1   48.00000000    4.54000000   1  2  2    0.91629080
1   15.00000100    4.54000000   1  0  5    1.70474800
1   36.99999240    4.54000000   0  0  1    0.40546510
1    9.99999900    4.54000000   1  0  2    0.91629080
1   13.00000000    4.54000000   0  0  2    0.91629080
1    6.00000050    4.54000000   1  2  0   -0.69314720
1   13.00000000    0.75500000   0  0  0   -0.69314720
1    6.99999950    4.54000000   1  1  2    0.91629080
1    9.00000000    4.54000000   1  0  2    0.91629080
1    4.99999810    4.54000000   0  0  2    0.91629080
1   15.00000100    4.54000000   0  0  6    1.87180220
1    3.00000000    4.54000000   0  0  3    1.25276290
1   13.99999710    4.54000000   0  0  0   -0.69314720
1    1.99999990    1.28000000   1  0  1    0.40546510
1    3.99999900    1.28000000   1  0  4    1.50407740
1    3.99999900    1.28000000   1  0  0   -0.69314720
1    0.00000000    2.50000000   1  1  0   -0.69314720
1   47.00000760    3.84999990   1  2  2    0.91629080
1   29.00000000    3.84999990   1  0  1    0.40546510
1    0.99999990    3.84999990   1  1  5    1.70474800
1   17.99999810    3.84999990   0  0  1    0.40546510
```

```
1  13.00000000   2.05000000  0  0  1   0.40546510
1  17.00000000   2.05000000  0  0  1   0.40546510
1   1.99999990   2.05000000  0  0  2   0.91629080
1   3.99999900   1.78000000  1  0  4   1.50407740
1   0.99999990   1.17999990  1  0  1   0.40546510
1   3.00000000   1.52000000  1  0  0  -0.69314720
1   9.00000000   1.48000000  1  0  2   0.91629080
1  12.00000000   4.29000000  0  0  0  -0.69314720
1  17.00000000   4.29000000  0  0  1   0.40546510
1   4.99999810   4.29000000  0  0  3   1.25276290
1   1.99999990   3.08999990  1  0  2   0.91629080
1   0.00000000   3.08999990  0  0  2   0.91629080
1   3.00000000   3.61999990  0  0  5   1.70474800
1   6.99999950   3.61999990  0  0  0  -0.69314720
1   7.99999860   3.61999990  0  0  0  -0.69314720
1   0.00000000   4.29000000  0  0  0  -0.69314720
1   6.00000050   2.60999990  1  1  1   0.40546510
1   6.99999950   2.60999990  1  0  1   0.40546510
1   9.00000000   2.09999990  0  0  0  -0.69314720
1   7.99999860   2.96000000  1  0  2   0.91629080
1   9.99999900   2.39000010  0  0  0  -0.69314720
1   3.00000000   1.95000000  1  0  1   0.40546510
1  13.99999710   3.41000010  0  0  2   0.91629080
1   0.99999990   4.29000000  0  0  2   0.91629080
1  10.99999710   3.58999990  0  0  2   0.91629080
1   9.99999900   4.61999990  0  0  1   0.40546510
1   6.00000050   2.14000010  0  0  1   0.40546510
1  10.99999710   2.85999990  0  0  0  -0.69314720
1   4.99999810   3.47000000  0  0  2   0.91629080
1  13.99999710   4.61999990  1  1  3   1.25276290
1   9.00000000   3.19000010  1  3  1   0.40546510
1   0.99999990   2.51000000  0  0  1   0.40546510
1   0.00000000   2.11999990  0  0  0  -0.69314720
1   3.00000000   3.19000010  0  0  0  -0.69314720
1   3.99999900   1.74000000  1  1  2   0.91629080
1   0.00000000   1.25000000  1  0  0  -0.69314720
1   6.99999950   3.69000010  1  0  0  -0.69314720
1   4.99999810   3.21000000  0  0  0  -0.69314720
1  15.00000100   4.61999990  1  1  0  -0.69314720
1  24.00000190   2.85999990  1  0  5   1.70474800
1   9.00000000   2.39000010  1  1  1   0.40546510
1   0.00000000   1.17999990  1  0  3   1.25276290
1   6.99999950   3.35999990  0  0  2   0.91629080
1   7.99999860   1.97000000  0  0  2   0.91629080
1  22.00000000   1.64000000  1  0  1   0.40546510
1   6.99999950   3.92000010  1  0  2   0.91629080
1   6.00000050   3.31999990  0  0  6   1.87180220
```

```
1    6.00000050    2.57999990   0  0  5    1.70474800
1   22.99999620    4.54000000   0  0  2    0.91629080
1    1.99999990    2.39000010   1  0  0   -0.69314720
1    9.00000000    3.58999990   0  0  4    1.50407740
1    6.99999950    3.69000010   0  0  2    0.91629080
1    9.99999900    3.19000010   0  0  2    0.91629080
1    3.99999900    2.31999990   0  0  2    0.91629080
1   10.99999710    3.47000000   1  1  2    0.91629080
1    3.99999900    3.69000010   0  0  2    0.91629080
1    4.99999810    2.31999990   1  1  3    1.25276290
1   10.99999710    3.19000010   0  0  1    0.40546510
1   19.00000190    4.54000000   0  0  1    0.40546510
1    3.00000000    3.35999990   1  1  0   -0.69314720
1    4.99999810    2.57999990   1  0  2    0.91629080
1    6.99999950    3.21000000   1  1  2    0.91629080
1    3.00000000    1.40000000   0  0  1    0.40546510
1    9.99999900    2.50000000   0  0  4    1.50407740
1    0.00000000    3.19000010   1  2  0   -0.69314720
1    3.00000000    3.35999990   0  0  0   -0.69314720
1    9.00000000    3.15000010   0  0  6    1.87180220
1    6.99999950    1.45000000   1  0  2    0.91629080
1    6.99999950    2.85999990   1  2  1    0.40546510
1   48.99999240    4.61999990   1  1  3    1.25276290
1    1.99999990    3.69000010   1  0  0   -0.69314720
1   19.00000190    2.96000000   1  0  1    0.40546510
1   12.00000000    3.08999990   1  1  1    0.40546510
1    0.99999990    3.08999990   0  0  0   -0.69314720
1   12.00000000    4.61999990   0  0  2    0.91629080
1   13.00000000    2.85999990   1  0  0   -0.69314720
1    1.99999990    3.21000000   0  0  0   -0.69314720
1    3.00000000    2.82999990   1  0  2    0.91629080
1   22.00000000    4.29000000   0  0  1    0.40546510
1   35.00000760    4.29000000   0  0  0   -0.69314720
1    0.99999990    3.08999990   1  0  2    0.91629080
1    3.99999900    3.69000010   0  0  0   -0.69314720
1    0.99999990    1.79000000   1  0  0   -0.69314720
1    1.99999990    3.35999990   1  1  0   -0.69314720
1   13.99999710    2.57999990   1  0  4    1.50407740
1   24.00000190    3.75000000   0  0  1    0.40546510
1    4.99999810    3.19000010   0  0  2    0.91629080
1    0.99999990    2.09999990   0  0  0   -0.69314720
1    7.99999860    3.58999990   1  1  0   -0.69314720
1    0.99999990    3.92000010   0  0  0   -0.69314720
1   24.00000190    3.31999990   1  1  1    0.40546510
1    1.99999990    2.00000000   0  0  0   -0.69314720
1    1.99999990    3.47000000   1  0  0   -0.69314720
1   13.99999710    3.21000000   1  0  4    1.50407740
```

```
1    4.99999810    2.05000000   0   0   3    1.25276290
1    0.99999990    2.52000000   0   0   0   -0.69314720
1    6.99999950    3.15000010   1   0   0   -0.69314720
1   38.00000380    1.86000000   1   2   6    1.87180220
1    3.00000000    2.85999990   0   0   2    0.91629080
1    3.99999900    4.29000000   1   1   4    1.50407740
1    0.00000000    1.25500000   1   2   0   -0.69314720
1   19.00000190    3.21000000   1   0   5    1.70474800
1    4.99999810    2.31999990   1   1   1    0.40546510
1    3.00000000    3.19000010   0   0   4    1.50407740
1    1.99999990    3.19000010   1   0   1    0.40546510
1   10.99999710    3.35999990   1   0   3    1.25276290
1   13.99999710    3.54000000   1   1   1    0.40546510
1    6.00000050    1.86000000   1   0   0   -0.69314720
1   13.00000000    1.50500000   0   0   0   -0.69314720
1    6.99999950    2.39000010   1   0   1    0.40546510
1   10.99999710    4.29000000   1   1   2    0.91629080
1    9.00000000    2.00000000   1   1   0   -0.69314720
1    3.99999900    3.92000010   0   0   1    0.40546510
1    6.00000050    4.29000000   1   0   2    0.91629080
1    6.00000050    3.35999990   0   0   1    0.40546510
1   13.99999710    4.61999990   1   0   3    1.25276290
1    6.00000050    2.00000000   0   0   1    0.40546510
1    4.99999810    3.58999990   1   0   2    0.91629080
1    6.00000050    2.86999990   0   0   1    0.40546510
1    3.99999900    2.96000000   0   0   0   -0.69314720
1   10.99999710    3.47000000   1   0   0   -0.69314720
1    7.99999860    3.19000010   1   0   4    1.50407740
1    3.00000000    2.85999990   1   1   0   -0.69314720
1   15.99999810    2.52000000   1   0   2    0.91629080
1    6.00000050    4.29000000   0   0   1    0.40546510
1    0.00000000    1.25500000   0   0   1    0.40546510
1    1.99999990    1.83000000   1   0   0   -0.69314720
1   12.00000000    4.29000000   0   0   3    1.25276290
1    7.99999860    2.96000000   1   0   0   -0.69314720
1    0.99999990    2.31999990   1   0   1    0.40546510
1   10.99999710    1.22000000   1   1   0   -0.69314720
1   10.99999710    4.29000000   1   0   1    0.40546510
1    1.99999990    4.25000000   0   0   1    0.40546510
1   36.00000000    2.55000000   0   0   2    0.91629080
1   10.99999710    1.95000000   1   0   1    0.40546510
1    1.99999990    3.69000010   0   0   0   -0.69314720
1   25.00000000    4.29000000   0   0   1    0.40546510
1    4.99999810    3.19000010   1   0   2    0.91629080
1   12.00000000    3.54000000   1   1   4    1.50407740
1    4.99999810    3.54000000   1   0   4    1.50407740
1    7.99999860    3.54000000   1   1   0   -0.69314720
```

```
1    0.99999990    1.86000000   1  1  3    1.25276290
1    3.00000000    4.61999990   0  0  4    1.50407740
1   15.99999810    4.61999990   1  0  1    0.40546510
1    9.99999900    4.61999990   1  0  3    1.25276290
1   13.00000000    4.54000000   1  0  1    0.40546510
1    0.99999990    3.47000000   1  0  0   -0.69314720
1   15.99999810    2.85999990   0  0  0   -0.69314720
1    6.99999950    2.00000000   1  0  0   -0.69314720
1    1.99999990    2.00000000   0  0  0   -0.69314720
1    3.00000000    2.60999990   1  0  0   -0.69314720
1    6.00000050    2.05000000   0  0  0   -0.69314720
1    6.00000050    2.05000000   0  0  0   -0.69314720
1    4.99999810    3.54000000   0  0  0   -0.69314720
1    0.00000000    0.92000000   1  0  2    0.91629080
1    0.00000000    1.79000000   1  0  0   -0.69314720
1    6.00000050    2.00000000   1  0  0   -0.69314720
1    7.99999860    3.15000010   0  0  4    1.50407740
1   15.99999810    2.26000000   1  0  5    1.70474800
1    3.99999900    2.26000000   0  0  1    0.40546510
1    0.99999990    4.29000000   1  0  6    1.87180220
1   21.00000000    4.29000000   1  0  1    0.40546510
1    9.00000000    4.54000000   0  0  1    0.40546510
1    3.00000000    3.35999990   1  1  0   -0.69314720
1    1.99999990    2.52000000   0  0  0   -0.69314720
1    9.99999900    4.29000000   0  0  3    1.25276290
1    9.00000000    4.29000000   0  0  0   -0.69314720
1   12.00000000    4.54000000   0  0  2    0.91629080
1    0.00000000    2.50000000   0  0  0   -0.69314720
1    1.99999990    2.76000000   0  0  1    0.40546510
1    0.00000000    2.55000000   1  1  0   -0.69314720
1    0.00000000    4.61999990   1  0  0   -0.69314720
1    6.00000050    1.63000000   0  0  2    0.91629080
1    3.00000000    3.47000000   1  0  1    0.40546510
1   30.99998860    3.41000010   1  0  2    0.91629080
1    0.99999990    4.29000000   0  0  0   -0.69314720
1    6.99999950    2.96000000   1  0  0   -0.69314720
1    0.00000000    4.61999990   1  0  5    1.70474800
1   12.00000000    3.58999990   1  0  0   -0.69314720
1   12.00000000    3.69000010   1  0  3    1.25276290
1    1.99999990    4.54000000   1  0  1    0.40546510
1   10.99999710    3.33999990   0  0  4    1.50407740
1    1.99999990    2.51000000   1  1  4    1.50407740
1    3.99999900    3.15000010   1  0  3    1.25276290
1    3.00000000    3.19000010   1  0  1    0.40546510
1   17.99999810    3.19000010   0  0  1    0.40546510
1    1.99999990    3.84999990   0  0  1    0.40546510
1    3.99999900    3.33999990   1  1  2    0.91629080
```

```
1   15.00000100   4.29000000   0   0   2    0.91629080
1    3.99999900   4.29000000   0   0   2    0.91629080
1   21.00000000   4.29000000   0   0   2    0.91629080
1    3.99999900   4.29000000   1   1   0   -0.69314720
1    0.00000000   2.96000000   1   0   0   -0.69314720
1    0.99999990   1.78000000   1   0   0   -0.69314720
1   17.00000000   3.54000000   1   2   1    0.40546510
1    3.00000000   3.69000010   1   0   0   -0.69314720
1    1.99999990   3.35999990   0   0   3    1.25276290
1    1.99999990   3.54000000   1   0   0   -0.69314720
1    0.99999990   3.54000000   1   1   0   -0.69314720
1    3.00000000   1.22000000   1   1   3    1.25276290
1    3.99999900   3.35999990   0   0   1    0.40546510
1    1.99999990   2.21000000   0   0   1    0.40546510
1    3.99999900   2.25000000   0   0   1    0.40546510
1   17.00000000   4.61999990   1   1   2    0.91629080
1    4.99999810   3.58999990   1   0   1    0.40546510
1    0.99999990   4.29000000   1   0   0   -0.69314720
1    1.99999990   2.11999990   0   0   2    0.91629080
1    3.00000000   2.26000000   0   0   2    0.91629080
1    3.00000000   2.26000000   1   0   4    1.50407740
1    4.99999810   4.29000000   0   0   0   -0.69314720
1    3.00000000   3.58999990   1   1   3    1.25276290
1   13.00000000   4.29000000   1   2   1    0.40546510
1    1.99999990   2.00000000   0   0   1    0.40546510
1    0.00000000   1.97000000   1   2   0   -0.69314720
1    7.99999860   3.92000010   1   0   0   -0.69314720
1   39.00000000   2.85999990   0   0   4    1.50407740
1   26.00000000   2.82999990   0   0   2    0.91629080
1    3.00000000   3.35999990   1   1   3    1.25276290
1   22.99999620   2.55999990   0   0   4    1.50407740
1    7.99999860   1.63000000   0   0   1    0.40546510
1   22.00000000   4.61999990   0   0   3    1.25276290
1    4.99999810   4.61999990   0   0   1    0.40546510
1   21.00000000   4.29000000   1   0   3    1.25276290
1    6.00000050   4.29000000   1   1   0   -0.69314720
1    3.99999900   4.29000000   1   1   2    0.91629080
1   24.00000190   4.29000000   1   0   6    1.87180220
1    6.00000050   4.29000000   1   0   1    0.40546510
1   33.99999240   3.35999990   0   0   1    0.40546510
1    0.00000000   3.21000000   0   0   2    0.91629080
1    9.99999900   2.00000000   0   0   1    0.40546510
1    1.99999990   3.21000000   1   2   2    0.91629080
1    3.00000000   2.57999990   1   2   2    0.91629080
1    3.00000000   2.57999990   1   0   0   -0.69314720
1    0.99999990   2.82999990   0   0   0   -0.69314720
1    6.00000050   3.19000010   0   0   1    0.40546510
```

```
1    3.00000000    3.47000000   0  0  1    0.40546510
1   22.99999620    4.61999990   1  0  2    0.91629080
1   13.00000000    4.25000000   1  0  4    1.50407740
1    4.99999810    1.86000000   1  0  2    0.91629080
1    9.00000000    4.29000000   1  1  0   -0.69314720
1    9.00000000    3.35999990   0  0  0   -0.69314720
1    9.99999900    1.80000000   1  0  2    0.91629080
1    0.00000000    1.65500000   0  0  1    0.40546510
1    6.00000050    2.85999990   0  0  0   -0.69314720
1   13.99999710    4.61999990   0  0  2    0.91629080
1    3.00000000    4.29000000   0  0  0   -0.69314720
1    4.99999810    2.35999990   1  0  4    1.50407740
1   10.99999710    1.80999990   1  0  1    0.40546510
1   32.00001140    3.58999990   1  0  1    0.40546510
1    9.99999900    1.76000000   0  0  0   -0.69314720
1    4.99999810    2.00000000   1  0  1    0.40546510
1   13.99999710    3.58999990   0  0  5    1.70474800
1    6.00000050    2.26000000   1  0  3    1.25276290
1    3.99999900    2.26000000   0  0  1    0.40546510
1    3.99999900    1.76000000   1  1  1    0.40546510
1    0.99999990    3.58999990   1  0  1    0.40546510
1    0.99999990    3.58999990   1  0  0   -0.69314720
1    4.99999810    2.82999990   1  1  1    0.40546510
1    1.99999990    2.57999990   0  0  1    0.40546510
1   17.99999810    4.61999990   1  0  1    0.40546510
1   15.00000100    2.31999990   0  0  1    0.40546510
1    0.00000000    2.39000010   1  0  0   -0.69314720
1   22.00000000    2.96000000   1  0  3    1.25276290
1   10.99999710    2.11999990   0  0  0   -0.69314720
1   17.99999810    4.61999990   1  0  1    0.40546510
1    3.00000000    2.25000000   1  0  1    0.40546510
1    7.99999860    1.76000000   1  0  1    0.40546510
1   12.00000000    4.54000000   0  0  6    1.87180220
1    9.99999900    3.69000010   1  1  2    0.91629080
1    1.99999990    1.25000000   0  0  0   -0.69314720
1    4.99999810    1.25000000   1  0  1    0.40546510
1    3.99999900    3.19000010   1  0  0   -0.69314720
1    0.00000000    2.57999990   1  2  0   -0.69314720
1    6.99999950    2.00000000   1  1  2    0.91629080
1    0.99999990    2.76000000   0  0  0   -0.69314720
1    3.99999900    2.54000000   0  0  2    0.91629080
1    0.00000000    3.19000010   0  0  0   -0.69314720
1    9.00000000    3.08999990   1  2  1    0.40546510
1    3.99999900    3.19000010   0  0  2    0.91629080
1    0.00000000    3.08999990   0  0  0   -0.69314720
1    7.99999860    2.26000000   0  0  2    0.91629080
1    9.00000000    2.26000000   1  1  1    0.40546510
```

```
1    7.99999860    3.35999990    1   0   2    0.91629080
1    9.00000000    3.15000010    1   0   1    0.40546510
1    3.99999900    4.54000000    1   0   0   -0.69314720
1    0.00000000    3.58999990    1   0   2    0.91629080
1    3.00000000    3.47000000    1   0   1    0.40546510
1    1.99999990    2.85999990    1   1   1    0.40546510
1    6.00000050    2.26000000    1   0   0   -0.69314720
1    3.99999900    3.47000000    1   1   2    0.91629080
1    0.00000000    2.76000000    1   0   0   -0.69314720
1    1.99999990    3.58999990    1   2   0   -0.69314720
1   13.00000000    3.75000000    0   0   5    1.70474800
1    1.99999990    2.57999990    1   2   1    0.40546510
1    1.99999990    3.58999990    1   1   2    0.91629080
1   15.99999810    1.89000000    1   2   3    1.25276290
1    9.00000000    3.15000010    0   0   7    2.01490310
1   12.00000000    2.86999990    0   0   5    1.70474800
1    0.99999990    2.11999990    1   2   0   -0.69314720
1    3.99999900    4.61999990    1   0   1    0.40546510
1    0.00000000    2.39000010    0   0   0   -0.69314720
1   10.99999710    2.11999990    0   0   0   -0.69314720
1    4.99999810    1.80999990    1   0   0   -0.69314720
1    7.99999860    2.31999990    0   0   4    1.50407740
1   15.00000100    2.39000010    0   0   0   -0.69314720
1    6.00000050    3.75000000    1   0   3    1.25276290
1    0.00000000    2.00000000    0   0   0   -0.69314720
1   13.99999710    3.58999990    1   1   0   -0.69314720
1    4.99999810    4.29000000    0   0   2    0.91629080
1    0.00000000    2.57999990    1   0   0   -0.69314720
1    4.99999810    4.61999990    1   0   0   -0.69314720
1    3.00000000    1.50500000    0   0   0   -0.69314720
1    3.99999900    3.75000000    1   0   2    0.91629080
1    3.99999900    3.75000000    1   2   0   -0.69314720
1    0.00000000    1.75000000    1   0   1    0.40546510
1    0.00000000    2.11999990    1   0   1    0.40546510
1    4.99999810    3.75000000    0   0   2    0.91629080
1    0.00000000    0.75500000    0   0   0   -0.69314720
1    6.99999950    2.25000000    0   0   1    0.40546510
1    3.00000000    3.19000010    0   0   1    0.40546510
;
```

MROZ Data Set

To create the Mroz data set, enter this code into a **Program** tab:

```
data mroz;
    input inlf nwifeinc educ exper expersq age kidslt6 kidsge6 lwage;
datalines;
1 10.91006    12 14   196   32  1   0   1.210154
1 19.49998    12 5    25    30  0   2   0.3285121
1 12.03991    12 15   225   35  1   3   1.514138
1 6.799996    12 6    36    34  0   3   0.0921233
1 20.10006    14 7    49    31  1   2   1.524272
1 9.859054    12 33   1089  54  0   0   1.55648
1 9.152048    16 11   121   37  0   2   2.12026
1 10.90004    12 35   1225  54  0   0   2.059634
1 17.305      12 24   576   48  0   2   0.7543364
1 12.925      12 21   441   39  0   2   1.544899
1 24.29995    12 15   225   33  0   1   1.401922
1 19.70007    11 14   196   42  0   1   1.524272
1 15.00001    12 0    0     30  1   2   0.7339532
1 14.6        12 14   196   43  0   2   0.8183691
1 24.63091    10 6    36    43  0   1   1.302831
1 17.53103    11 9    81    35  0   3   0.2980284
1 14.09998    12 20   400   43  0   2   1.16761
1 15.839      12 6    36    39  0   5   1.643839
1 14.1        12 23   529   45  0   0   0.6931472
1 10.29996    12 9    81    35  0   4   2.021932
1 22.65498    16 5    25    42  0   2   1.254248
1 8.090048    12 11   121   30  0   0   1.272958
1 17.479      13 18   324   48  0   0   1.178655
1 9.56        12 15   225   45  0   0   1.178655
1 8.274953    12 4    16    31  1   1   0.7675587
1 27.34999    17 21   441   43  0   2   1.331812
1 16          12 31   961   59  0   0   1.386294
1 16.99998    12 9    81    32  0   3   1.55327
1 15.10006    17 7    49    31  1   0   1.981815
1 15.69998    12 7    49    42  0   0   1.76936
1 5.11896     11 32   1024  50  0   0   0.4308079
1 16.75001    16 11   121   59  0   0   0.8997548
1 13.59993    13 16   256   36  0   2   1.76663
1 17.10005    12 14   196   51  0   1   1.272958
1 16.73405    16 27   729   45  0   3   1.336789
1 14.19698    11 0    0     42  0   1   0.9017048
1 10.31999    12 17   289   46  0   0   0.8651237
1 11.3841     10 28   784   46  0   1   1.511847
1 14.59408    14 24   576   51  0   0   1.726029
1 17.50044    17 11   121   30  0   0   2.683142
1 15.51       12 1    1     30  1   2   0.9852943
1 21.99998    12 14   196   57  0   0   1.365939
1 22.5        16 6    36    31  1   2   0.9450337
1 19.994      12 10   100   48  0   2   1.512376
```

1	14.13	12	6	36	30	0	3	0.6931472
1	5.000013	12	4	16	34	0	2	1.244788
1	21.1549	16	10	100	48	0	2	0.7011649
1	7.141946	12	22	484	45	0	0	1.519863
1	16.65007	12	16	256	51	0	0	0.8209686
1	6.352	12	6	36	30	0	2	0.9698315
1	27.31395	12	12	144	46	0	1	0.8285082
1	14.5	12	32	1024	58	0	0	0.0943096
1	16.25799	12	15	225	37	0	8	0.1625439
1	9.5	8	17	289	52	0	0	0.4700036
1	7.999956	10	34	1156	52	0	0	0.6292484
1	12.50003	16	9	81	31	0	0	1.39716
1	14.00003	14	37	1369	55	0	0	2.265444
1	20.80007	17	10	100	34	0	0	2.084541
1	19.38511	14	35	1225	55	0	0	1.525839
1	12.38699	12	6	36	39	0	2	0.7621601
1	28.5	14	19	361	40	0	3	1.481605
1	15.04991	12	10	100	43	0	4	1.262826
1	10.49998	8	11	121	48	0	0	0.9996756
1	11.81	12	15	225	47	0	0	1.832582
1	6.950073	12	12	144	41	0	4	2.479308
1	12.41997	8	12	144	36	0	0	1.279015
1	17.4	17	14	196	46	0	2	1.937936
1	15.5	12	11	121	34	0	0	1.070453
1	21.21704	12	9	81	41	0	3	1.123923
1	18	12	24	576	51	0	1	1.321756
1	11.89992	12	12	144	33	0	0	1.745
1	26.75196	12	13	169	52	0	0	1.301744
1	12.14996	9	29	841	58	0	0	1.641866
1	10.19999	10	11	121	34	2	4	2.10702
1	8.120015	12	13	169	31	0	1	1.467068
1	10.65996	12	19	361	48	0	1	1.605811
1	18.10001	12	2	4	32	0	2	-1.029739
1	8.599986	17	24	576	49	0	0	1.087686
1	13.665	15	9	81	32	2	2	0
1	32.34996	12	6	36	58	0	0	0.9382087
1	12.08501	6	22	484	50	0	0	-0.1505904
1	12.15	14	30	900	60	0	0	0
1	17.69502	12	10	100	50	0	1	1.073671
1	24.7	14	6	36	56	0	0	1.265848
1	2.133992	9	29	841	51	0	0	0.486369
1	20.95005	17	29	841	54	0	1	2.12026
1	10.50008	13	36	1296	59	0	0	1.129853
1	10.55	9	19	361	46	0	2	0.9932518
1	45.75	15	8	64	46	0	1	1.658628
1	13.63204	12	13	169	39	1	3	0.3474122
1	18.23894	12	16	256	44	0	2	1.568324

```
1 17.09        12 11   121    33  2   0    0.5108456
1 30.2349      12 15   225    33  1   2    0.1148454
1 28.7         12 6    36     48  0   2   -0.6931472
1 19.63        12 13   169    31  0   4   -0.3364523
1 12.82494     12 22   484    45  0   1    1.028226
1 23.8         12 24   576    45  0   1    1.580689
1 26.30003     13 2    4      32  0   2    0.5558946
1 20.69991     12 6    36     47  0   0    0.9014207
1 26           13 2    4      34  0   2    0.8843046
1 10.87702     12 2    4      37  0   1    0.4282046
1 25.61206     12 14   196    36  0   1    1.058415
1 20.98899     12 9    81     47  1   2    0.8783396
1 70.74993     16 11   121    48  0   1    1.654908
1 17.05        12 9    81     42  0   2    1.321756
1 21           13 6    36     33  0   3    0.3285121
1 8.12         11 19   361    46  0   0    1.386294
1 20.88599     12 26   676    47  0   3    1.172885
1 17.66892     12 19   361    44  0   1    1.224187
1 25.20003     12 3    9      36  0   4    0.2876571
1 14.24501     17 7    49     31  2   0    2.230262
1 14.3         14 28   784    55  0   0    1.504077
1 23.70001     16 13   169    45  0   1    1.531152
1 46           17 9    81     47  0   0    1.375158
1 42.9999      12 15   225    46  0   3    1.760269
1 14.749       11 20   400    49  0   0   -0.6931472
1 16.15005     12 29   841    49  0   0    1.406489
1 17.774       12 9    81     45  0   2    1.791759
1 91           17 1    1      38  1   3    1.299292
1 22.29993     10 8    64     47  0   0    1.351004
1 34.60001     13 19   361    54  0   3    1.016281
1 9.620002     11 23   529    41  0   0    1.075344
1 10.89995     12 3    9      43  0   2    1.478965
1 14.49994     16 13   169    31  1   1    1.689487
1 22.00002     17 8    64     47  0   0    2.288598
1 17.90008     12 17   289    35  0   2   -1.822631
1 23.67506     16 4    16     45  0   3   -0.9607652
1 11.79996     12 15   225    33  1   0    1.290994
1 16.14195     16 11   121    54  0   1    0.8648711
1 18.39997     8  7    49     35  0   4    1.540452
1 15.49995     12 0    0      31  1   2    0.6162121
1 17.324       12 0    0      55  0   0    1.648659
1 19.205       12 10   100    34  0   2    1.193498
1 21.30006     13 8    64     38  0   1    2.143976
1 23.56        11 2    4      45  0   1    0.7244036
1 20.85        12 4    16     47  0   1    0.9416075
1 26.15        12 6    36     39  0   2    0.7827594
1 17           14 18   324    36  1   0    1.832582
```

```
1 20.72       12 3    9      33  1  2   1.203963
1 17.00009    12 22   484    50  0  0   1.491645
1 16          12 33   1089   58  0  0   1.892133
1 19.50005    17 28   784    49  0  0   2.130895
1 12          14 23   529    41  0  2   1.480604
1 13.73191    12 27   729    51  0  1   0.8943313
1 27.19999    9  11   121    53  0  0   0.2025325
1 5.315       12 6    36     36  1  2   0.4855078
1 16          12 11   121    46  0  2   1.098612
1 27.87198    12 14   196    36  0  2   1.55327
1 40.00001    14 17   289    53  0  1   0.121598
1 15.90003    16 17   289    40  0  3   2.001804
1 27.49997    17 14   196    42  0  2   1.495037
1 17.02005    15 11   121    33  1  1   0.9052298
1 22.39494    12 7    49     43  0  3   0.6325476
1 11.1        16 8    64     31  1  0   1.386294
1 32.70001    17 6    36     47  0  0   2.102914
1 27.79996    17 8    64     54  0  0   1.959644
1 2.199994    12 4    16     33  1  3   0.5108456
1 19.72095    16 25   625    43  0  0   1.236924
1 9.999988    13 24   576    46  0  1   1.443313
1 13.19997    12 11   121    35  0  3   1.021659
1 12.70897    11 19   361    37  0  3   0.6361535
1 27.30005    16 9    81     37  0  2   1.616453
1 21.2        14 19   361    34  0  3   0.2231435
1 14.4        16 14   196    43  1  0   1.049807
1 20.57596    12 22   484    46  0  0   1.415052
1 12.49999    9  6    36     35  0  3   0.5753766
1 17.50022    17 23   529    46  0  0   2.606682
1 44.00004    14 15   225    46  0  0   1.517915
1 13.11895    12 6    36     43  0  2   0.7550416
1 14.00006    12 11   121    30  0  0   1.094972
1 9.645086    11 2    4      41  0  2   0.9421144
1 17.39705    12 22   484    54  0  1   1.724943
1 7.799889    12 10   100    31  0  1   1.031546
1 13.13398    10 14   196    44  0  0   0.4743691
1 25.6        12 12   144    32  0  1   0.8109302
1 13.90003    5  9    81     47  0  0   0.7092666
1 19.29794    17 13   169    46  0  1   1.710549
1 9.200016    11 18   324    37  0  0   0.4602689
1 37.99999    12 8    64     51  0  2   1.331812
1 44          12 11   121    49  0  1   1.098612
1 21.37202    14 9    81     36  0  4   2.157999
1 23.66802    11 9    81     39  0  1   1.437581
1 9           12 14   196    48  0  2   1.544899
1 25.19995    14 9    81     38  0  2   1.410597
1 21.22       12 2    4      40  0  2   3.218876
```

```
1 33.96991    10 12   144    39  1   5    0.9681619
1 17.07       16 15   225    37  0   0    1.791759
1 6.016024    13 11   121    49  0   1    1.68873
1 17.10001    12 7    49     33  0   3    -0.409172
1 8.237       12 9    81     30  0   0    0.2231435
1 13.30008    12 19   361    54  0   0    0.8221558
1 16.00002    11 11   121    39  0   4    1.241702
1 12.53999    12 8    64     43  0   3    1.427124
1 18.00004    9  13   169    31  0   3    1.497097
1 31.2        13 4    16     33  0   3    0.5596158
1 20.74991    12 7    49     40  0   3    1.300028
1 11.09992    12 19   361    36  0   1    1.88443
1 20.68       12 14   196    51  0   0    0.9555114
1 18.00001    13 14   196    44  0   1    1.582087
1 32.43007    16 3    9      42  0   3    1.755614
1 32.90003    12 9    81     40  0   1    1.513103
1 24.10001    16 7    49     34  1   1    2.251892
1 17.80039    17 7    49     30  0   0    2.364432
1 20.50002    12 14   196    54  0   0    0.1053505
1 10.4999     12 29   841    51  0   0    1.399729
1 10.43703    9  19   361    44  0   2    0.9884625
1 18.19499    12 14   196    43  0   1    1.090647
1 12.84508    12 16   256    34  0   1    1.154614
1 13.8        13 10   100    45  0   0    1.266948
1 22.2        12 12   144    39  0   0    2.885192
1 6.699941    12 24   576    50  0   0    1.22888
1 6.250016    12 6    36     52  0   0    1.203963
1 15.60001    12 9    81     41  0   2    1.35738
1 3.30001     10 14   196    59  0   0    0.8377236
1 3.670978    12 26   676    52  0   0    0.5369611
1 7.789997    16 7    49     46  0   0    0.7487238
1 18.27199    12 4    16     41  1   5    2.295873
1 10.95398    11 15   225    33  0   2    1.107803
1 13.49999    12 23   529    45  0   0    0.6208453
1 11.20001    10 1    1      36  1   2    -2.054164
1 20.99991    12 29   841    48  0   1    1.892012
1 25.7        12 9    81     47  0   1    1.729725
1 8.932994    12 6    36     45  0   0    0.4693784
1 19.15998    12 11   121    37  0   2    0.9808417
1 26.58999    16 17   289    46  0   4    2.069492
1 22.40001    17 6    36     43  0   3    1.675188
1 20.633      12 7    49     42  0   2    1.386294
1 28.20001    17 2    4      34  1   2    1.799215
1 28.8        12 24   576    52  0   0    1.832582
1 8.999997    12 4    16     37  0   3    1.090647
1 11.39994    12 11   121    37  0   1    1.443124
1 10.40001    8  25   625    52  0   0    1.25036
```

1	19.08006	12	11	121	30	1	0	1.602313
1	9.46604	13	2	4	31	0	1	1.018559
1	6.50006	12	19	361	38	0	1	1.297053
1	29.11701	12	7	49	43	0	3	1.685194
1	19.10302	8	2	4	49	0	1	-0.4209849
1	16.34997	12	20	400	55	0	0	1.562095
1	32.02502	17	10	100	38	0	2	2.146528
1	16.70006	17	19	361	52	0	0	2.347463
1	4.811038	12	17	289	48	0	0	0.9698315
1	24.62601	13	12	144	32	0	2	1.924146
1	17.40001	12	11	121	32	0	1	1.626728
1	13.02504	12	6	36	38	0	2	-0.0392607
1	19.00698	12	10	100	46	0	3	1.460149
1	14.03	12	4	16	40	0	3	1.955394
1	14.89991	9	2	4	31	0	4	0.9263599
1	25.00006	10	13	169	43	0	1	2.066192
1	10.70007	12	21	441	51	0	0	1.422843
1	24.25	16	9	81	30	1	0	2.101032
1	39.13997	13	4	16	52	0	0	2.261461
1	7.199973	8	2	4	30	1	5	0.7013138
1	31.811	16	19	361	51	0	0	2.031013
1	10.00005	13	4	16	31	0	2	1.162369
1	20.66	12	9	81	34	0	4	0.4700036
1	13.49998	11	14	196	49	0	0	1.410597
1	25.38	13	6	36	35	1	3	0.3930551
1	18.27498	12	24	576	53	1	0	1.290994
1	39.213	12	1	1	32	0	3	0
1	10.49994	10	13	169	38	0	3	0.9571255
1	34.857	12	3	9	54	0	0	0.5596158
1	28.502	17	10	100	47	0	1	1.568616
1	12.99996	15	16	256	45	0	1	1.710188
1	41.39991	16	9	81	47	0	1	1.410597
1	14.78	10	19	361	59	0	0	0.2231435
1	15.05	11	4	16	32	0	1	0.5108456
1	29.69998	12	10	100	45	0	1	1.332392
1	16.16502	12	5	25	40	0	4	0.8601859
1	25.20516	14	7	49	47	0	2	2.32278
1	14.2	16	3	9	36	1	2	1.919595
1	18.15897	14	38	1444	56	0	0	1.976107
1	28.98106	8	16	256	41	0	1	0.8954347
1	13.392	7	13	169	48	0	3	0.1812376
1	9.17502	12	1	1	36	1	2	0.4953058
1	27.03985	12	7	49	41	0	0	0.5777924
1	13.14995	14	15	225	41	0	0	1.078818
1	16.40007	12	10	100	36	0	3	1.603199
1	21.29999	12	2	4	37	0	3	0.6208453
1	17.20102	12	19	361	38	0	0	2.083894

```
1 8.560026    14 25  625    43  0   2   1.379169
1 6.49084     16 25  625    54  0   0   1.112384
1 12.49997    12 7   49     38  0   1   1.067122
1 27.00002    12 15  225    30  1   0   1.118807
1 53.50005    12 11  121    49  0   0   1.588541
1 52.49995    13 25  625    45  0   1   1.390311
1 38.39998    13 19  361    51  0   0   1.714806
1 13.89194    10 4   16     34  0   0   0.2010615
1 3.899993    12 14  196    34  0   2   0.987271
1 34.2        12 19  361    41  0   1   0.9835007
1 19.70008    12 18  324    49  0   1   2.233171
1 18.49995    12 14  196    32  0   0   1.143618
1 10.99998    14 11  121    32  0   0   -0.6113829
1 43.30001    17 4   16     32  0   2   2.153052
1 18.76001    10 29  841    47  0   0   1.299837
1 4.800096    9  21  441    39  0   1   0.8409204
1 21.5        12 24  576    49  0   0   1.058484
1 28.03994    12 19  361    37  0   3   1.152658
1 26          16 31  961    59  0   0   1.293576
1 27          12 28  784    50  0   0   1.832582
1 17.79969    17 15  225    32  0   1   2.32718
1 17.40195    12 27  729    46  0   0   1.166146
1 19.30999    17 13  169    43  0   2   2.034993
1 9.99998     11 4   16     37  0   3   0.6792511
1 11.17998    16 10  100    32  0   2   1.547137
1 18.85696    11 8   64     39  0   1   0.7530186
1 12.30002    13 4   16     34  0   2   0.8472836
1 13.67712    11 18  324    39  0   1   0.871126
1 9.559997    8  3   9      45  0   3   0.2282505
1 24.49998    11 11  121    50  0   0   0.0896578
1 23.15       12 8   64     40  0   1   1.321756
1 15.59088    10 10  100    30  0   1   1.196102
1 14.42092    17 33  1089   57  0   0   1.636119
1 17.45491    12 19  361    39  0   1   1.892012
1 9.800019    12 35  1225   53  0   0   1.518309
1 17.57446    17 21  441    48  0   1   2.472159
1 16.555      14 7   49     46  0   1   1.321756
1 13.29497    12 18  324    47  0   0   1.473641
1 11.844      12 4   16     43  0   1   1.369479
1 46.64506    12 12  144    47  0   0   1.203963
1 14.69999    12 16  256    47  0   1   1.198729
1 26.09008    12 14  196    47  0   0   1.27021
1 9.9         12 3   9      46  0   0   0.4700036
1 9.048026    9  1   1      34  0   4   0.7999817
1 30.75006    10 27  729    48  0   0   1.565946
1 8.49994     12 12  144    30  0   1   1.758978
1 22.24999    12 6   36     51  0   1   0.8580258
```

```
1 42.91       12 9    81    52  0  5   0.6931472
1 33.3        12 2    4     37  0  2   0.6418539
1 13.8199     12 6    36    32  0  2   1.63374
1 23.60001    17 9    81    36  0  2   1.703748
1 13.00007    12 16   256   35  0  2   1.844004
1 20.74994    17 22   484   45  0  0   1.966119
1 6.3         12 26   676   56  0  0   0.8649974
1 7.788925    10 11   121   40  0  2   0.9333052
1 10.47004    12 11   121   45  1  2   0.7792332
1 12          12 15   225   32  0  2   0.9555114
1 16.97992    12 13   169   45  0  0   1.316247
1 17.9        12 6    36    40  0  2   1.475906
1 15.53994    12 20   400   38  0  1   1.491397
1 9.883986    12 17   289   49  0  4   1.45575
1 28.59995    16 8    64    47  0  1   0.5108456
1 17.66001    13 13   169   52  0  0   1.180438
1 25.99992    13 15   225   34  0  1   1.688489
1 13.60201    12 14   196   44  0  2   0.7907275
1 15.8        16 14   196   36  0  3   1.401799
1 41.09999    17 6    36    50  0  0   -0.433556
1 10.77504    12 24   576   45  0  0   1.683172
1 9.000047    14 10   100   44  0  2   -1.766677
1 24.39899    12 2    4     57  0  2   3.155595
1 37.30009    17 9    81    35  0  0   2.259521
1 27.99995    12 23   529   46  0  0   1.306926
1 13.7        14 12   144   30  2  1   0.7984977
1 17.20994    12 8    64    42  0  3   0.5590442
1 14.00001    12 16   256   34  0  1   0.1479026
1 35.75502    17 10   100   45  0  2   1.944495
1 23.5        16 7    49    35  1  2   1.378338
1 31.99993    16 19   361   40  0  0   3.064745
1 17.15       12 2    4     32  0  1   -0.7419173
1 20.25002    9  9    81    54  0  0   0.7657004
1 5.485985    12 14   196   38  0  3   0.619393
1 25.07504    12 9    81    43  0  3   1.465452
1 18.21995    16 16   256   54  0  0   2.18926
1 26          14 7    49    39  0  3   1.021659
1 34.50007    12 6    36    37  0  1   0.9770095
1 12.4        12 22   484   46  0  2   0.9162908
1 10.78685    11 9    81    56  0  0   2.905096
1 16.32301    12 9    81    41  0  3   -0.1996712
1 30.5        16 14   196   45  0  1   0.6931472
1 51.29963    17 17   289   44  0  1   2.733393
1 33.04997    17 12   144   50  0  1   1.868335
1 34.75001    14 13   169   37  0  5   2.12026
1 16.40004    12 8    64    44  0  1   1.515193
1 19.70007    14 10   100   32  0  2   0.9146093
```

```
1 6.600003    12 16  256   34 1  1  1.499556
1 9.020008    10 1   1     32 0  2  0.8030772
1 10.40001    12 6   36    37 0  3  0.7280316
1 14.51999    13 4   16    44 0  1  0.51641
1 17.2        16 8   64    34 0  2  1.226448
1 43          12 4   16    33 1  3  0.9162908
1 13.87196    7  15  225   43 0  3  1.376471
1 -0.0290575  16 7   49    35 0  2  1.828975
1 16.76994    14 14  196   43 0  1  1.368283
1 7.8         12 16  256   34 0  0  1.064711
1 14.50006    10 15  225   36 0  3  1.406489
1 7.9         12 23  529   41 0  2  1.047319
1 79.80001    16 19  361   41 0  0  1.948093
1 7.17597     10 4   16    35 0  3  1.078001
1 17.50698    12 12  144   32 1  3  0.6539385
1 20.6        14 12  144   30 0  0  1.927892
1 18.55992    12 25  625   43 0  0  1.361028
1 9.3         6  14  196   54 0  0  0.6931472
1 5.120008    15 14  196   35 0  2  1.604687
1 14.50004    12 11  121   50 0  0  0.1839036
1 19.8        17 7   49    34 1  1  3.113515
1 18.29995    14 18  324   52 0  0  1.926829
1 33.99994    13 4   16    35 0  3  1.270126
1 11.62794    6  37  1369  55 0  0  0.6826927
1 11.80005    16 13  169   35 0  0  1.68107
1 39.09998    14 14  196   49 0  1  0.556296
1 18.43007    15 17  289   38 2  2  1.62822
1 21          14 5   25    42 0  2  0.9162908
1 59          8  2   4     48 0  1  1.341558
1 25.3        14 0   0     51 0  0  0
1 23.24899    12 3   9     43 0  2  1.122231
1 24.92809    12 21  441   43 0  1  0.5401708
1 14.78199    12 20  400   38 0  1  1.391506
1 18.90003    12 19  361   44 0  1  1.697174
1 21          12 4   16    36 1  3  3.218876
1 10.00001    12 19  361   38 0  0  0.8711678
1 29.30997    8  11  121   47 0  0  1.16733
1 13.14003    12 14  196   34 0  2  1.216988
1 25.08999    17 8   64    40 1  2  0.5753766
1 14.59993    12 13  169   31 0  1  1.151616
1 1.200001    12 24  576   46 0  0  0.9942513
1 32          14 1   1     36 0  3  0.5263249
1 16.11997    13 1   1     39 1  2  -1.543182
1 26.50002    17 3   9     36 0  2  1.912043
1 12.75006    8  4   16    37 0  4  0.5542873
1 12.9        12 21  441   39 0  4  0.9162908
1 10.69998    11 10  100   36 1  3  1.500939
```

```
1 14.43403    12 13   169    49  0   2   0.9446838
1 23.709      12 9    81     45  1   1   1.241269
1 15.1        17 14   196    32  2   0   1.564984
1 18.19998    10 2    4      36  0   5   0.8380265
1 22.64106    12 21   441    40  0   1   1.668857
1 21.64008    13 22   484    43  0   2   1.769429
1 23.99998    12 14   196    33  0   1   1.226448
1 16.00002    12 7    49     30  0   1   1.406489
0 21.025      12 2    4      49  0   1   .
0 23.6        16 5    25     30  2   0   .
0 22.8        12 12   144    30  1   0   .
0 35.91       12 1    1      41  0   4   .
0 21.7        12 12   144    45  0   1   .
0 21.823      12 4    16     43  0   5   .
0 31          13 9    81     42  0   1   .
0 15.3        12 9    81     60  0   0   .
0 12.925      12 6    36     57  0   0   .
0 15.83       10 5    25     38  0   2   .
0 30.2        12 5    25     56  0   0   .
0 16.6        12 8    64     32  0   3   .
0 11          7  2    4      49  0   1   .
0 15          12 6    36     55  0   0   .
0 20.528      9  0    0      36  1   1   .
0 13.126      12 3    9      44  0   3   .
0 15.55       10 7    49     44  0   1   .
0 18.01       14 3    9      35  1   2   .
0 18.874      14 10   100    44  2   3   .
0 24.8        12 3    9      45  0   1   .
0 17.5        12 2    4      34  1   0   .
0 16.15       17 12   144    30  2   0   .
0 15.189      8  15   225    39  0   1   .
0 6          12 5    25     36  0   2   .
0 37.25       17 4    16     38  0   2   .
0 27.76       12 10   100    53  0   0   .
0 9.09        12 1    1      36  0   2   .
0 14.5        12 8    64     32  1   1   .
0 19.7        9  20   400    51  0   3   .
0 16.788      11 4    16     38  0   0   .
0 18.52       12 7    49     33  2   0   .
0 20.95       12 10   100    54  0   0   .
0 7.574       9  3    9      38  0   3   .
0 10.027      11 5    25     30  2   2   .
0 5          12 10   100    34  2   3   .
0 7.04        9  0    0      34  0   1   .
0 40.8        12 3    9      50  0   2   .
0 16.05       17 10   100    30  2   0   .
0 33.1        12 2    4      38  0   2   .
```

```
0  33.856      14 10   100    54  0   0   .
0  20.5        12  4    16    30  1   2   .
0  28.6        12  0     0    55  0   0   .
0  18.75       10 10   100    51  0   1   .
0  20.3        12  5    25    44  0   1   .
0  13.42       12  0     0    53  0   0   .
0  18.4        10  0     0    42  0   2   .
0  16.682      12 19   361    38  0   2   .
0  32.685      13  2     4    38  1   3   .
0  7.05        12 12   144    41  1   4   .
0  10.867       8  5    25    35  0   3   .
0  18.22       12  5    25    33  1   2   .
0  26.613      13  5    25    48  0   0   .
0  25          12 10   100    47  0   0   .
0  15.7        12  0     0    34  0   5   .
0  40.25       13  4    16    33  2   1   .
0  73.6        13  3     9    31  3   1   .
0  10.592       8  2     4    58  0   0   .
0  8           12  1     1    49  0   0   .
0  13.4         8  0     0    55  0   1   .
0  23.7        14  1     1    44  0   0   .
0  18.9         9  1     1    44  0   0   .
0  48.3        16  6    36    36  0   3   .
0  24.47       12 12   144    38  0   3   .
0  28.63       16  6    36    37  0   3   .
0  25.32       12  9    81    47  0   0   .
0  13.53       12 14   196    47  0   3   .
0  14.8        12 13   169    32  1   1   .
0  17.4        12  8    64    43  1   2   .
0  15.98       11  0     0    42  1   4   .
0  16.576      12  1     1    56  0   0   .
0  21.85       13  3     9    38  0   5   .
0  14.6        12 13   169    52  0   2   .
0  21.6        12  3     9    50  0   0   .
0  24          16  8    64    33  0   0   .
0  20.883      16  8    64    44  0   2   .
0  19.5        12 18   324    41  0   1   .
0  42.8        12  2     4    45  0   1   .
0  41.5        14  3     9    53  0   0   .
0  18.965      14  5    25    53  0   0   .
0  16.1        12  2     4    42  0   1   .
0  14.7        13 10   100    32  2   0   .
0  18.8        12 30   900    56  0   0   .
0  14.75       11  1     1    37  1   3   .
0  21          12  5    25    40  1   2   .
0  35.4        15  8    64    54  0   3   .
0  10.7         7  0     0    53  0   0   .
```

```
0 24.5      12 4   16    48  0  1   .
0 17.045    12 2   4     36  1  2   .
0 18.8      12 30  900   57  0  0   .
0 14        12 25  625   51  0  0   .
0 18.214    13 3   9     33  0  4   .
0 20.177    12 20  400   52  0  0   .
0 8.3       10 20  400   56  0  0   .
0 14.2      12 0   0     36  1  2   .
0 21.768    14 15  225   36  1  0   .
0 29.553    12 10  100   46  0  1   .
0 4.35      10 4   16    31  0  3   .
0 24        11 3   9     52  0  0   .
0 18.3      12 10  100   46  0  2   .
0 17.2      12 9   81    35  2  0   .
0 16.476    12 7   49    59  0  0   .
0 13.4      8  12  144   36  0  1   .
0 44.988    7  0   0     51  1  3   .
0 18.2      16 16  256   31  1  0   .
0 28        14 4   16    31  0  2   .
0 11.55     12 7   49    32  1  1   .
0 28.45     16 7   49    35  1  2   .
0 15.096    12 14  196   40  0  3   .
0 8.009     10 2   4     33  1  2   .
0 10.04     7  20  400   54  0  0   .
0 16.7      12 5   25    36  1  1   .
0 8.4       10 10  100   50  0  1   .
0 13        8  20  400   54  0  0   .
0 17.97     11 10  100   48  0  1   .
0 18.45     15 8   64    41  0  4   .
0 31        12 11  121   50  0  4   .
0 24.135    12 3   9     46  0  2   .
0 31.7      13 6   36    42  0  1   .
0 10.19     9  4   16    31  1  2   .
0 21.574    12 4   16    53  0  0   .
0 26.68     12 9   81    51  0  1   .
0 17.7      12 10  100   47  0  1   .
0 29.4      12 3   9     50  0  1   .
0 22.159    6  2   4     37  0  1   .
0 35        12 2   4     30  2  2   .
0 8.63      12 0   0     49  0  0   .
0 17.08     12 8   64    52  0  2   .
0 32.5      12 6   36    47  0  2   .
0 16        12 15  225   49  0  0   .
0 18.85     12 15  225   44  0  4   .
0 17.5      8  9   81    53  0  0   .
0 19.392    12 8   64    30  1  0   .
0 14.45     12 18  324   54  0  2   .
```

```
0  21.8        7   3    9      47  1   1   .
0  7.7         15  10   100    56  0   0   .
0  31.8        12  6    36     49  0   1   .
0  17.258      6   20   400    48  0   0   .
0  13.399      12  8    64     49  0   1   .
0  16.073      12  3    9      56  0   1   .
0  23.26       12  4    16     46  0   0   .
0  37.3        12  13   169    45  0   2   .
0  11          12  4    16     32  0   2   .
0  13.075      12  17   289    43  1   1   .
0  13.7        12  4    16     34  1   1   .
0  25.1        12  0    0      30  1   1   .
0  18.6        17  15   225    38  2   0   .
0  29          16  11   121    33  1   1   .
0  19.237      12  23   529    52  0   0   .
0  19.855      11  1    1      43  0   3   .
0  9.45        12  5    25     33  1   1   .
0  30          10  1    1      45  0   0   .
0  15          10  5    25     36  2   1   .
0  24.701      12  3    9      34  1   1   .
0  15.9        14  3    9      37  0   2   .
0  16.24       10  19   361    46  0   1   .
0  21.1        12  20   400    47  0   0   .
0  23          16  5    25     31  2   1   .
0  6.34        5   0    0      57  0   0   .
0  42.25       12  3    9      30  1   1   .
0  14.694      12  3    9      30  0   0   .
0  21.417      12  7    49     44  0   3   .
0  20.2        13  7    49     53  0   0   .
0  12.09       8   1    1      51  0   0   .
0  24.76       12  13   169    39  1   3   .
0  23          8   0    0      52  0   0   .
0  19.365      8   0    0      46  0   4   .
0  5.55        12  12   144    47  0   5   .
0  68.035      8   0    0      52  0   2   .
0  29.3        12  5    25     45  0   2   .
0  18.5        11  45   2025   60  0   0   .
0  22.582      13  10   100    41  0   2   .
0  21.5        8   2    4      39  0   3   .
0  28.07       12  3    9      49  0   1   .
0  50.3        15  1    1      32  1   1   .
0  23.5        12  5    25     33  1   3   .
0  15.5        10  10   100    36  0   4   .
0  13.44       13  4    16     37  3   3   .
0  8.1         12  7    49     30  1   2   .
0  9.8         11  9    81     44  1   1   .
0  20.3        12  5    25     48  0   1   .
```

0	15	11	4	16	40	0	4	.
0	56.1	13	11	121	47	0	0	.
0	22.846	12	9	81	36	0	2	.
0	22.225	11	4	16	40	0	2	.
0	17.635	12	2	4	46	0	1	.
0	18.5	12	23	529	52	0	0	.
0	13.39	12	3	9	44	0	1	.
0	15.15	10	15	225	45	0	1	.
0	16.2	7	8	64	30	2	1	.
0	33.92	12	3	9	40	1	3	.
0	14	12	25	625	43	0	1	.
0	16.736	12	2	4	49	0	2	.
0	30.65	12	0	0	46	1	4	.
0	12.4	11	19	361	52	0	0	.
0	19.022	12	3	9	31	1	1	.
0	11.203	10	7	49	42	1	1	.
0	19.876	11	1	1	33	0	3	.
0	57	16	9	81	57	0	0	.
0	18.29	10	3	9	49	0	0	.
0	20.22	14	8	64	45	0	1	.
0	22.15	11	0	0	56	0	0	.
0	30.623	12	5	25	41	1	3	.
0	9.38	5	20	400	56	0	0	.
0	22	10	3	9	48	0	1	.
0	23.675	16	12	144	52	0	2	.
0	33.671	12	5	25	51	0	0	.
0	12.367	11	1	1	35	0	3	.
0	21.95	12	0	0	45	0	0	.
0	32	12	7	49	54	0	0	.
0	22.61	12	13	169	54	0	2	.
0	12.092	12	3	9	31	1	0	.
0	3.777	6	0	0	53	0	3	.
0	36	14	2	4	35	2	2	.
0	26.9	12	0	0	36	1	3	.
0	32.242	12	2	4	59	0	0	.
0	35.02	16	1	1	54	0	0	.
0	37.6	12	10	100	37	1	1	.
0	1.5	12	10	100	44	0	0	.
0	96	17	1	1	34	1	2	.
0	18.15	12	3	9	49	0	0	.
0	15.5	12	32	1024	49	0	0	.
0	14	9	0	0	60	0	0	.
0	14.756	12	7	49	51	0	0	.
0	22	12	5	25	30	1	1	.
0	24.466	12	2	4	47	0	2	.
0	24.4	12	5	25	36	0	4	.
0	24	12	3	9	35	1	3	.

```
0  15.5          12 25   625    58  0   0   .
0  30.8          14 0    0      41  1   3   .
0  10.66         10 3    9      51  0   1   .
0  13.35         12 10   100    47  0   0   .
0  10.09         9  10   100    45  1   2   .
0  55.6          14 7    49     60  0   0   .
0  25.7          16 5    25     30  1   1   .
0  29            11 15   225    55  0   0   .
0  7.286         12 1    1      32  1   2   .
0  37.752        12 5    25     36  0   2   .
0  13.072        12 9    81     55  0   0   .
0  7.044         12 18   324    47  0   0   .
0  18.2          12 1    1      47  0   1   .
0  27            11 0    0      37  0   1   .
0  30.3          12 6    36     50  0   2   .
0  12            12 1    1      30  0   3   .
0  31.5          17 2    4      48  0   1   .
0  27.092        10 15   225    43  0   2   .
0  20.968        11 25   625    48  1   0   .
0  27            14 1    1      41  1   2   .
0  11.225        12 0    0      50  0   0   .
0  37.7          8  0    0      58  0   0   .
0  28.2          13 0    0      38  0   5   .
0  34            12 8    64     37  0   1   .
0  63.2          16 22   484    50  0   0   .
0  7.5           8  5    25     42  0   4   .
0  17.41         9  10   100    37  1   3   .
0  51            16 1    1      41  0   2   .
0  12.916        12 1    1      31  0   2   .
0  21.9          12 6    36     51  0   0   .
0  17.64         12 4    16     36  1   2   .
0  20            15 6    36     54  0   0   .
0  15            12 0    0      49  0   0   .
0  14.06         9  1    1      48  1   1   .
0  15.825        9  3    9      42  0   2   .
0  16.51         12 15   225    41  1   2   .
0  13            16 33   1089   55  0   0   .
0  10            9  2    4      42  0   0   .
0  22            15 1    1      32  0   1   .
0  29.8          12 10   100    43  0   2   .
0  15            12 0    0      33  1   3   .
0  22.3          15 14   196    48  0   1   .
0  14.55         12 15   225    43  0   2   .
0  19.73         17 15   225    47  1   3   .
0  35            12 10   100    54  0   0   .
0  21.014        12 6    36     51  0   1   .
0  10.876        10 18   324    51  0   1   .
```

```
0 27.85       13 15   225     43   1   1    .
0 9.56        12 30   900     53   0   0    .
0 30.3        11 15   225     34   1   1    .
0 7.72        8  10   100     31   1   1    .
0 10.55       12 0    0       56   0   0    .
0 24.106      16 0    0       42   0   1    .
0 22.995      12 4    16      32   0   2    .
0 6           12 0    0       35   1   3    .
0 24.35       12 3    9       30   1   1    .
0 7.608       10 20   400     51   0   0    .
0 28.2        12 3    9       47   0   3    .
0 16.15       12 1    1       54   0   1    .
0 51.2        15 5    25      31   3   0    .
0 12.646      10 7    49      47   0   0    .
0 19          14 6    36      47   0   3    .
0 19          12 2    4       40   0   3    .
0 14.4        8  0    0       48   0   0    .
0 7.232       8  10   100     34   0   7    .
0 21.943      12 6    36      38   0   3    .
0 47.5        12 4    16      32   1   3    .
0 28.9        16 8    64      48   0   1    .
0 12.4        12 18   324     41   0   2    .
0 6.531       5  7    49      49   0   2    .
0 22.422      8  15   225     59   0   0    .
0 22.2        13 7    49      58   0   0    .
0 77          12 8    64      41   0   3    .
0 88          12 8    64      45   0   2    .
0 26.04       14 3    9       30   1   1    .
0 63.5        12 10   100     41   0   1    .
0 12.1        12 9    81      30   2   0    .
0 17.505      12 24   576     53   0   1    .
0 18          12 12   144     31   0   0    .
0 28.069      14 2    4       43   0   2    .
0 14          12 6    36      31   1   1    .
0 8.117       12 18   324     51   0   0    .
0 11.895      9  17   289     43   0   0    .
0 45.25       14 7    49      31   1   2    .
0 31.106      11 6    36      48   0   0    .
0 4           12 10   100     31   1   1    .
0 40.5        12 5    25      44   0   1    .
0 21.62       11 7    49      48   0   1    .
0 23.426      12 11   121     53   0   1    .
0 26          10 14   196     42   0   3    .
0 7.84        12 5    25      39   2   6    .
0 6.8         10 2    4       32   1   2    .
0 5.33        12 4    16      36   0   2    .
0 28.2        13 5    25      40   0   2    .
```

```
0 10          12 14  196   31  2   3   .
0 9.952       12 4   16    43  0   0   .
0 24.984      12 15  225   60  0   0   .
0 28.363       9 12  144   39  0   3   .
;
```

Appendix 2

References

Cochran, W. G., and G. M. Cox. 1950. *Experimental Designs*. New York, : Wiley.

Hilbe, J. M. 2009. *Logistic Regression Models*. London, England: Chapman & Hall/ CRC.

Hosmer, D. W. Jr., and S. Lemeshow. 2000. *Applied Logistic Regression*. 2 ed. New York, NY: John Wiley & Sons.

Lawless, J. F., and K. Singhal. 1978. "Efficient Screening of Nonnormal Regression Models." *Biometrics* (34): 318–327.

Pregibon, D. 1981. "Logistic Regression Diagnostics." *Annals of Statistics* (9): 705– 724.

Recommended Reading

- *Getting Started with Programming in SAS Studio*

- *The Little SAS Book: A Primer* (Buy)

- *Learning SAS by Example: A Programmer's Guide* (Buy)

- *SAS Statistics by Example* (Buy)

- *Elementary Statistics Using SAS* (Buy)

For a complete list of SAS books, go to support.sas.com/bookstore. If you have questions about which titles you need, please contact a SAS Book Sales Representative:

SAS Books
SAS Campus Drive
Cary, NC 27513-2414
Phone: 1-800-727-3228
Fax: 1-919-677-8166
E-mail: sasbook@sas.com
Web address: support.sas.com/bookstore

Index

Gain Greater Insight into Your SAS® Software with SAS Books.

Discover all that you need on your journey to knowledge and empowerment.