SPSS Inc. continues its tradition of regularly enhancing this family of powerful but easy-to-use statistical software products with the release of SPSS 16.0. Learn more about SPSS 16.0 in the present brochure.

TABLE OF CONTENTS

1  SPSS 16 what’s new in SPSS 16.0
2  SPSS 16 Base
3  SPSS 16 Advanced Models
4  SPSS 16 Categories
5  SPSS 16 Classification Trees
6  SPSS 16 Complex Samples
7  SPSS 16 Conjoint
8  SPSS 16 Data Preparation
9  SPSS 16 Exact Tests
10 SPSS 16 Missing Values
11 SPSS 16 Neural Networks
12 SPSS 16 Regression Models
13 SPSS 16 Tables
14 SPSS 16 Trends
15 SPSS 16 Server
16 SPSS Programmability Extension

For further information please contact:
SPSS (Schweiz) AG, Schneckenmannstrasse 25, 8044 Zürich
phone +41 44 266 90 30
fax +41 44 266 90 39
info@spss.ch
www.spss.ch
SPSS Inc. continues its tradition of regularly enhancing this family of powerful but easy-to-use statistical software products with the release of SPSS 16.0. Besides including features that customers like you have requested, SPSS 16.0 offers a new user interface, written completely in Java™.

The new interface makes SPSS even easier to use. You can resize dialog boxes to accommodate long variable names and lists, and quickly drag and drop variables from one pane to another to set up your analysis.

In addition to the new user interface, SPSS 16.0 offers:

- Significantly expanded analytical capabilities
- Enhanced data management and reporting capabilities
- Improved programmability
- Greater performance and scalability in enterprise applications

And, with this release, virtually the same add-on modules are available whether you use SPSS on a Microsoft® Windows®, Apple® Mac® OS® X, or Linux® platform.* Going forward, SPSS Inc. plans to make all new features and capabilities to the SPSS product family available on all three platforms simultaneously.

**A new, more flexible user interface**

The entire user interface has a new form and functionality in SPSS 16.0. User dialogs, the Data Editor, the Syntax Editor, the Output Viewer, the Help system, the Chart Editor, and the Pivot Table Editor—all have been rewritten in Java. Now, you’ll find it even easier to work with your data. You can instantly resize a dialog to see a more complete description of your variables. You can also quickly select and drag the variables you want to use in your analysis.

And if you work with data in multiple languages—as many organizations do—you’ll be able to process Unicode data, as well as treat text according to Unicode properties for tasks like sorting and case conversion.

**Introducing SPSS Neural Networks™**

A new add-on module, SPSS Neural Networks, provides a complementary approach to the statistical techniques available in SPSS Base and its modules. From the familiar SPSS interface, you can access SPSS Neural Networks and discover more complex relationships in your data. For example, businesses can forecast consumer demand for a particular set of products, calculate the likely rate of response to a marketing campaign, determine an applicant’s creditworthiness, or detect potentially fraudulent transactions.

* Amos™ and SPSS Exact Tests™ are only available on the Microsoft Windows platform, and the exchange of data with SPSS’ Dimensions™ family of survey research products is supported only on the version of SPSS 16.0 that operates on Windows.
Neural networks are non-linear data mining tools that consist of input and output layers plus one or more hidden layers. In a neural network, the connections between neurons have weights associated with them. These weights are iteratively adjusted by the training algorithm to minimize error and provide accurate predictions.

With the SPSS Neural Networks module, you can choose either the Multilayer Perceptron (MLP) or Radial Basis Function (RBF) procedure to explore your data in entirely new ways.

**New or enhanced statistical techniques**

SPSS 16.0 offers enhanced statistical techniques in SPSS Complex Samples™, SPSS Advanced Models™, Amos™, and through the SPSS Programmability Extension™.

SPSS Complex Samples now includes the Cox Regression technique for time-to-event data. If you have data based on a complex sample design, you can use this technique to accurately predict the timing to a specific event—how long a high-value customer remains active, for example, or how long people fitting a certain profile will survive a certain medical condition. SPSS Complex Samples Cox Regression (CSCOXREG) enables you to more easily analyze differences in subgroups as well as the effects of a set of predictors. The procedure takes the sample design into account when estimating variances and can handle data involving multiple cases, such as multiple patient visits, encounters, and observations.

SPSS Advanced Models offers additional enhancements to the generalized linear models (GENLIN) and generalized estimating equations (GEE) procedures introduced with SPSS 15.0. These procedures enable you to more accurately predict ordinal outcomes, such as customer satisfaction. Enhancements available in SPSS 16.0 enable analysts to predict outcomes that are a combination of discrete and continuous outcomes—such as claim amounts—using a Tweedie distribution.

Amos, SPSS Inc.’s powerful but easy-to-use tool for structural equation modeling (SEM), now offers latent class analysis and mixture modeling. This statistical method is particularly useful in market segmentation studies when estimating the probability that an individual belongs to a certain segment or cluster is important. This method also provides a useful alternative to k-means cluster analysis.
In the SPSS Programmability Extension, described elsewhere, the current integration plug-ins for Python® and the Microsoft.NET version of Visual Basic® are joined by an integration plug-in for R. This enables analysts to access the wealth of statistical routines created in R and use them within SPSS as part of SPSS syntax.

The SPSS Programmability Extension made possible the introduction in SPSS 16.0 of Partial Least Squares (PLS) regression as an alternative to Ordinary Least Squares (OLS) regression. PLS is a predictive technique that can handle many independent variables, even when they display multicollinearity. Choose PLS instead of OLS if you have a high number of variables relative to the number of cases—a situation that frequently occurs in survey research.

**Enhanced data management and reporting capabilities**

In addition to support for Unicode, as already mentioned, SPSS 16.0 includes many enhancements to data management that users have specifically requested. Now you'll have greater flexibility in how you work with, analyze, and save your data. Using SPSS 16.0 capabilities, you can:

- Change the string length or the data type of an existing variable, using syntax
- Define missing values and value labels for data strings of any length
- Choose either to round off or add decimal places to calculated dates when using the Date/Time Wizard
- Benefit from new capabilities in the Data Editor, including the ability to find and replace information, spell check value and variable labels, sort by variable name, type, or format, and more
- Find and replace text in the Output Viewer—for example, search for warnings to identify problems in your output
- Import/export data to and from Excel® 2007
- Suppress the number of active datasets in the user interface
- Set a permanent default working directory

As for reporting, a new, more powerful visualization engine replaces the Interactive Graph Properties (IGRAPH) feature, making graph editing faster and easier. (Existing IGRAPH syntax will continue to work.)

SPSS 16.0 introduces Python as the default front-end scripting language. Python supersedes SAX Basic as the scripting language for tasks such as automation of repetitive tasks and customization of output. As with SAX Basic, you can apply a “base” autoscript to all objects or to individual objects. Existing SAX Basic scripts will continue to work in SPSS 16.0

**Improved programmability**

The SPSS Programmability Extension enables you to enhance the capabilities of SPSS by using external programming languages such as Python. Applications written in Python and Visual Basic can also call upon the SPSS backend to conduct analysis or create reports. Integration plug-ins are available at the SPSS Developer Central Web site, as is the SPSS Programmability Extension SDK that allows users to create their own integration plug-ins.

SPSS continues to make the development of APIs easier for users with additional improvements to the Programmability Extension, and now allows the implementation of multiple integration plug-ins and multiple versions of a single integration plug-in.

An additional enhancement available through the SPSS Programmability Extension is the new data step procedure in the SPSS Python integration plug-in. This allows users to create a completely new SPSS data file including the simultaneous creation of defined variables and cases.

Visit SPSS Developer Central at [www.spss.com/devcentral](http://www.spss.com/devcentral) to share code, tools, and programming ideas.
Greater performance and scalability

SPSS 16.0 features several multithreaded procedures, which result in greater performance on machines containing multiple processors and multi-core processors. The following procedures are multithreaded: in SPSS Base, Linear Regression, Correlation, Partial Correlation, and Factor Analysis; and in SPSS Complex Samples, the SPSS Complex Samples Select procedure.

SPSS 16.0 also provides additional integration with SPSS Predictive Enterprise Services™. As organizations recognize the need to create more effective processes for managing and automating their analytic assets, providing an efficient, cost-effective way to manage and update these assets becomes increasingly important. SPSS Predictive Enterprise Services provides these capabilities for analytical assets created with SPSS—such as syntax, scripts, and output—as well as for assets created with other SPSS products such as the Clementine® data mining workbench.

Enhancements to the SPSS Adapter for Predictive Enterprise Services enable you to store and manage a variety of assets, including Python script files, and enjoy increased performance during retrieval and refresh processes.

To learn more, please visit www.spss.com/predictive_enterprise_services.

System requirements

**SPSS Base 16.0 for Windows**
- Operating System: Microsoft Windows XP (32-bit versions) or Vista™ (32-bit or 64-bit versions)
- Hardware:
  - Intel® or AMD x86 processor running at 1GHz or higher
  - Memory: 256MB RAM or more; 512MB recommended
  - Minimum free drive space: 450MB
  - CD-ROM drive
  - Super VGA (800x600) or higher-resolution monitor
  - For connecting with an SPSS Server, a network adapter running the TCP/IP network protocol
- Web browser: Internet Explorer 6

**SPSS Base 16.0 for MAC OS X**
- Operating system: Apple Mac OS X 10.4 (Tiger™)
- Hardware
  - PowerPC or Intel processor
  - Memory: 512MB RAM or more
  - Minimum free drive space: 800MB
  - CD-ROM drive
  - Super VGA (800x600) or higher-resolution monitor
- Web browser: Safari™ 1.3.1, Firefox 1.5, or Netscape 7.2
- Java Standard Edition 5.0 (J2SE 5.0)
**SPSS Base 16.0 for Linux**

- Operating system: any Linux OS that meets the following requirements**:
  - Kernel 2.4.33.3 or higher
  - glibc 2.3.2 or higher
  - XFree86-4.0 or higher
  - libstdc++5

- Hardware:
  - Processor: Intel or AMD x86 processor running at 1GHz or higher
  - Memory: 256MB RAM or more; 512MB recommended
  - Minimum free drive space: 450MB
  - CD-ROM drive
  - Super VGA (800x600) or a higher-resolution monitor

- Web browser: Konqueror 3.4.1, Firefox 1.0.6, or Netscape 7.2

**Note: SPSS 16.0 was tested on and is supported only on Red Hat Enterprise Linux 4 Desktop and Debian 3.1**

**SPSS add-on modules**

All SPSS 16.0 add-on modules require SPSS Base 16.0. No other system requirements are necessary.

**Amos 16.0**

- Operating system: Windows XP or Windows Vista

- Hardware:
  - Memory: 256MB RAM minimum
  - 125MB or more available hard-drive space
  - Web browser: Internet Explorer 6.0

**SPSS Server 16.0**

- Operating system: Windows Server 2003 (32-bit or 64-bit); Sun™ Solaris™ (SPARC) 9 and later (64-bit only); IBM® AIX® 5.3 and later; or Red Hat® Enterprise Linux® ES4 and later; HP-UX IIi (64-bit Itanium)

- Hardware:
  - Minimum CPU: Two CPUs recommended, running at 1GHz or higher
  - Memory: 256MB RAM per expected concurrent user
  - Minimum free drive space: 300MB
  - Required temporary disk space: Calculate by multiplying 2.5 x number of users x expected size of dataset in megabytes

**SPSS Adapter for SPSS Predictive Enterprise Services**

- Requires SPSS Base 16.0 and SPSS Predictive Enterprise Services

---

**Version comparison chart: new features added to SPSS by version number and by area**

<table>
<thead>
<tr>
<th>New feature</th>
<th>Version number</th>
<th>16.0</th>
<th>15.0</th>
<th>14.0</th>
<th>13.0</th>
<th>12.0</th>
<th>11.5</th>
<th>11.0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Desktop versions available on Windows, Mac, or Linux</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Resizable dialogs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Drag-and-drop in dialogs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td><strong>Programmability</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Addition of Python as a “front-end” cross-platform scripting language</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Ability to create a data source, including variables and cases, without having to import the active data source into SPSS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Control the flow of your syntax jobs or create your own user-defined algorithms using external programming languages (through the SPSS Programmability Extension)</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Python programming language included on the SPSS CD</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to create first-class, user-defined procedures</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syntax control of output files</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Features subject to change based on final product release.
## Version comparison chart: new features added to SPSS by version number and by area

<table>
<thead>
<tr>
<th>New feature</th>
<th>Version number</th>
<th>16.0</th>
<th>15.0</th>
<th>14.0</th>
<th>13.0</th>
<th>12.0</th>
<th>11.5</th>
<th>11.0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Predictive Enterprise</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Several multithreaded procedures for improved performance and scalability</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPSS Adapter for SPSS Predictive Enterprise Services (added in SPSS 14.0.1)</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Updated PMML to include transformations</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single administration utility for SPSS Server, Clementine, and SPSS Predictive Enterprise Services platforms</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stripe temporary files over multiple disks for increased performance (in SPSS Server)</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conversion-free/copy-free data access in SQL DBMS (in SPSS Server)</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Data-free client (in SPSS Server)</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Support for Open SSL (in SPSS Server)</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-database data preparation (sort and aggregate) to improve performance (in SPSS Server)</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Score data using PMML models created with SPSS, Clementine, and AnswerTree® (in SPSS Server)</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictor Selection and Naïve Bayes algorithms (in SPSS Server)</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Data access and data management</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Improved Data Editor:</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to customize variable view</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spell checking for value labels and variable labels</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to sort by variable name, type, format, etc.</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unicode support</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Import/export Excel 2007 data</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syntax to change string length and basic data type of existing variables</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creation of value labels and missing values on strings of any length</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to set a permanent default working directory</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Define variable properties tool</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Date and Time Wizard</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Export to Database Wizard</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct Microsoft Excel interface</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Identify Duplicate Cases tool</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clone dataset command</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to open multiple datasets within a single SPSS session</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Export data to recent versions of Excel and SAS®</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Long variable names (up to 64 bytes)</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very long text strings (up to 32,767 bytes)</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long value labels (up to 120 bytes)</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Custom Attributes for user-defined meta data in the SPSS Data Editor</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Read recent SAS files</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Read/write Stata® files</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Export to Dimensions Data Model</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLE DB data access (Windows only)</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Restructure Data Wizard</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Visual Binner to easily bin data (for example, break income into “bands” of $10,000)</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal Binning (in SPSS Data Preparation add-on module)</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subset variable views</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Features subject to change based on final product release.*
<table>
<thead>
<tr>
<th>New feature</th>
<th>Version number</th>
<th>16.0</th>
<th>15.0</th>
<th>14.0</th>
<th>13.0</th>
<th>12.0</th>
<th>11.5</th>
<th>11.0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Analysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPSS Neural Networks add-on module</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complex Samples Cox Regression added to SPSS Complex Samples</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latent Class Analysis in Amos</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partial Least Squares regression**</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Support for R algorithms**</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CATPCA and PROXSCAL (in SPSS Categories™ add-on module)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Multiple correspondence analysis (in SPSS Categories add-on module)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Preference scaling (in SPSS Categories add-on module)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>TwoStep cluster analysis (in SPSS Base)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Descriptive ratio statistics</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Linear mixed models (also known as hierarchical linear models) (in SPSS Advanced Models add-on module)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Generalized linear models (in SPSS Advanced Models add-on module)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Generalized estimating equations (in SPSS Advanced Models add-on module)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Multinomial logistic regression (in SPSS Regression Models add-on module)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Ordinal regression to model ordinal outcomes (in SPSS Base)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Receiver-operating characteristic (ROC) analysis (in SPSS Base)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>SPSS Complex Samples add-on module</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Complex samples general linear model and logistic regression (in SPSS Complex Samples add-on module)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Complex samples ordinal regression (in SPSS Complex Samples add-on module)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>SPSS Classification Trees™ add-on module</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Validate Data procedure (in SPSS Data Preparation add-on module)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Anomaly Detection for multivariate outliers (in SPSS Data Preparation add-on module)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Enhanced SPSS Trends™ add-on module with Expert Modeler</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Bayesian estimation—MCMC algorithm (in Amos structural equation modeling software)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Data imputation, including multiple imputation (in Amos structural equation modeling software)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Estimation and imputation of ordered-categorical and censored data (in Amos structural equation modeling software)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Run significance tests on multiple response variables, excluding categories used in subtotal calculations (in SPSS Tables™ add-on module)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Features subject to change based on final product release.

**Available at SPSS Developer Central; requires the SPSS Programmability Extension**
## Version comparison chart: new features added to SPSS by version number and by area

<table>
<thead>
<tr>
<th>New feature</th>
<th>Version number</th>
<th>16.0</th>
<th>15.0</th>
<th>14.0</th>
<th>13.0</th>
<th>12.0</th>
<th>11.5</th>
<th>11.0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Graphs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presentation graphics system</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Chart Builder user interface for graphics</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Support for SPSS Inc.’s Graphics Production Language (GPL)</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dual-Y axis and overlay charts</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enhanced process control charts</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-D line charts (both axes can be scale axes) and charts for multiple response sets</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population pyramids (also called mirror charts or dual charts), 3-D bar charts, and dot charts (also called dot density charts)</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Additional chart display features/options, including paneled charts and error bars on categorical charts</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Output</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Find and Replace feature in the Output Viewer</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enhanced SPSS Tables module with table preview builder and inferential statistics</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Export output to Microsoft Excel and Word</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Export output to Microsoft PowerPoint®</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Export output to PDF</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Output Management System (turn pivot table output, such as SPSS data files, XML, and HTML, into data/input)</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interactive interface for the output management system</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Switch output language</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Table to graph</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td><strong>Licensing improvements</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Network license reservations and priority settings</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Network commuter license</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>License manager redundancy</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Help</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interactive case studies</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>“SPSS Manuals on CD,” featuring manuals in PDF format for SPSS Base and all add-on modules</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistical Coach™</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tutorial</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chart tutorial</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“What’s This?” (context-sensitive help)</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Features subject to change based on final product release.
Solve business and research problems using SPSS, a statistical and data management package for analysts and researchers. Compared to other data analysis packages, SPSS is easier to use, has a lower total cost of ownership, and comprehensively addresses the entire analytical process. SPSS Base is an integral part of this process, providing functionality for data access, data management and preparation, data analysis, and reporting. It enables you to work confidently with add-on modules and other products in the SPSS Family, which provide capabilities for planning, data collection, and deployment, and add incremental functionality to areas SPSS Base also addresses. This release features a new, Java™-based interface, which makes working with your data even easier. And most features and add-on modules are available on your choice of platforms: Microsoft® Windows®, Apple® Mac®, or Linux®.

Access and analyze massive datasets quickly
SPSS makes it easy for you to quickly access, manage, and analyze any kind of dataset, including survey data, corporate databases, or data downloaded from the Web. In addition, SPSS Base can process Unicode data. This eliminates variability in data due to language-specific encoding and enables your organization to view, analyze, and share data written in multiple languages. Make your analysis as fast as possible—let your server do the heavy computation work. Just combine SPSS with the optional SPSS Server.

Prepare your data for analysis quickly and easily
Before you can analyze your data, you need to prepare them for analysis. Numerous techniques and features built into SPSS Base enable easy data preparation. Following are summaries of just a few SPSS data management highlights.

With SPSS Base, you can easily set up data dictionary information (for example, value labels and variable types) and prepare your data for analysis more quickly using the Define Variable Properties tool. SPSS presents a list of values and counts of those values so you can add this information. Once the data dictionary is set up, you can apply it using the Copy Data Properties tool. The data dictionary acts as a template, so you can apply it to other data files and to other variables within the same file.

SPSS makes it easy for you to identify duplicate cases, so you can eliminate them prior to your analysis. Use the Identify Duplicate Cases tool to set parameters and flag duplicates so that you can keep track of them for the record.

Additionally, SPSS makes it easy to prepare continuous-level data for analysis. The Visual Binner enables you to easily (for example, break income into “bands” of 10,000 or break ages into groups). A data pass provides a histogram that enables you to specify cutpoints in an intelligent manner. You can then automatically create value labels from the specified cutpoints (for example, “21-30”).
Create your own dictionary information for variables with Custom Attributes. For example, create a custom attribute that represents the full text of a survey question when a code name such as “demo01” is used as the variable name. You can also create custom attributes describing transformations for a derived variable with information explaining how you transformed the variable.

You can open multiple datasets within a single SPSS session. This enables you to save time and condense steps when merging data files. It also helps you maintain consistency when copying data dictionary information between multiple files. Or, if you prefer, you can suppress the number of active datasets.

SPSS enables you to restructure your data files to prepare them for analysis. For example, take a data file that has multiple cases per subject and restructure the data to put all data for each subject into a single record. SPSS gives you the flexibility to complete the reverse action—you can take a data file that has a single case per subject and spread the data across multiple cases.

Use the Date and Time Wizard to make calculations with dates and times, create date/time variables from strings containing date variables (such as “03/29/06”), and bring date/time data from a variety of sources into SPSS. You can also parse individual date/time units, such as year, from date/time variables to apply filters.

Build charts more easily with sophisticated reporting capabilities

Create commonly used charts, such as SPLOMs (scatterplot matrices), histograms, and population pyramids, more easily with Chart Builder. This highly visual chart creation interface enables you to create a chart by dragging variables and elements onto a chart creation canvas. Optionally, use a shortcut method based on an existing chart in the Gallery. You will see a limited preview of the chart as it is being built. Advanced users can attain a broader range of chart and option possibilities by using the Graphics Production Language (GPL).

The presentation graphics system gives you control at both the creation and edit stages, to help ease your workload in a production setting. Create a chart once, and then use your specifications to create hundreds more just like it.

Present your best results with report OLAP

OLAP technology transforms the way you create and share information. Report OLAP in SPSS provides you with a fast, flexible way to create, distribute, and manipulate information for ad hoc decision making. Create tables, graphs, and report cubes that feature unique, award-winning pivoting technology and enable you to discover new insights into your data. Swap rows, columns, and layers of report cubes—or quickly change information and statistics in graphs—for new levels of understanding. You can even convert a table to a graph with just a few mouseclicks.

Maximize the impact of your analytic results

To gain the greatest value from your analyses, you need to manage analytical assets and efficiently share results with others. Using SPSS Server in conjunction with SPSS Predictive Enterprise Services, you can do this. The SPSS Adapter for SPSS Predictive Enterprise Services enables you to store and retrieve a variety of file types, including data and output files, syntax, chart templates, and plan files created with SPSS Complex Samples™. Using SPSS Predictive Enterprise Services, you can schedule jobs, automate refreshes, and deploy models and results. Learn more by visiting www.spss.com/predictive_enterprise_services/.
Features

General operations

- Apply splitters through the Data Editor to more quickly and easily understand wide and long datasets
- Select the customizable toolbar feature to:
  - Assign procedures, scripts, or other software products
  - Select from standard toolbar icons or create your own
- Work with multidimensional pivot tables/report cubes to:
  - Rearrange columns, rows, and layers by dragging icons for easier ad hoc analyses
  - Toggle between layers by clicking on an icon for easier comparison between subgroups
  - Enable online statistical help for choosing statistical procedures or chart types and interpreting results; realistic application examples are included
- Change text attributes such as fonts, colors, bolding, italics, and others
- Change table attributes such as number formats, line styles, line width, column alignments, background/foreground shading, enable or disable lines, and more
- Selectively display or hide rows, columns, or labels to highlight important findings
- Enable task-oriented help with step-by-step instructions:
  - View case studies that show you how to use selected statistics and interpret results
  - Select the Statistics Coach™, which helps you choose the best statistical procedure or graph
  - Work through tutorials
  - Select “Show Me” buttons, which link to the tutorial for more in-depth help when you need it
  - Use “What’s This?” help, which provides pop-up definitions of statistical terms and rules of thumb

- Use formatting capabilities for output to:
  - Transform a table into a graph for more visually compelling communication
  - Show correlation coefficients together with their significance level (as well as n) in correlations using the default output display
  - Control whether, upon activation, a table is opened in place or in its own window
  - Stamp date and time into the journal file for easy reference
  - Right-click on an SPSS syntax file icon to run a command file without needing to go through production mode
  - Use drop-down lists for easier access to different layers
  - Set permanent page settings
  - Set a column width for all pivot tables and define text wrapping
  - Choose whether to use scientific notation to display small numbers
  - Control number of digits of precision in presentations
  - Interact with reports and use models and code created by others in your organization with the optional addition of SPSS Predictive Enterprise Services.
  - Add footnotes and annotations
  - Reorder categories within a table to display results most effectively
  - Group or ungroup multiple categories in rows or columns under a single heading that spans the rows or columns
  - Use one of 16 pre-formatted TableLooks™ for quick and consistent formatting of results
  - Create and save customized formats as TableLooks for your own personalized style
  - Display values or labels
  - Rotate table labels
  - Use drop-down lists for easier access to different layers

- Work with the Viewer to organize, view, and move through results
  - Keep a record of your work using the “append” default in journal files
  - Use outline representation to quickly determine output location
  - Select an icon in the outline and see corresponding results displayed in the content pane
  - Reorder charts, tables, and other objects by dragging icons in the outline
  - Selectively collapse or expand the outline to view or print selected results
  - Contain tables, charts, and objects in a single content pane for easy review and access
  - Right-justify, left-justify, or center output
  - Search and replace information in the Viewer of the contents pane, the outline pane, or both

- Create and save analysis specifications for repetitive tasks or unattended processing

- Use the enhanced production mode facility with dialog interface and macros for easier periodic reporting

- Have full control over table splitting with improved pagination and printing

- Select the print preview option

- Enter your own commands, if you wish, via a command line input window

- Refer to explanations of statistical terms through the on-screen statistical glossary

- Work with data more easily, thanks to:
  - Resizable dialog boxes
  - Drag-and-drop in dialogs
  - Export output to Microsoft Word
    - Convert pivot tables to Word tables with all formatting saved
    - Convert graphics into static pictures
  - Export output to PowerPoint® (Windows only)
    - Convert pivot tables to tables in PowerPoint with all formatting saved
    - Convert graphics into static pictures

Features subject to change based on final product release. □ Symbol indicates a new feature.
- Export output to Excel®
  - Put tables on the same sheet or on separate sheets within one Excel workbook file
  - Export only the current view or all layers of an SPSS pivot table
  - Place each pivot table layer on the same sheet or on separate sheets within one Excel workbook
- Export SPSS output to PDF
  - Choose to optimize the PDF for Web viewing
  - Control whether PDF-generated bookmarks correspond to Navigator Outline entries in the Output Viewer. Bookmarks facilitate navigation of large documents.
  - Control whether fonts are embedded in the document. Embedded fonts ensure that the reader of your document sees the text in its original font, preventing font substitution.
- Easily open/save and create new output files through syntax
- Receive wheel mouse support for Output Viewer scroll
- Switch output languages (for example, switch between Japanese and English)
- Use the scripting facility to:
  - Create, edit, and save scripts
  - Build customized form interfaces
  - Assign scripts to toolbar icons or menus
  - Automatically execute scripts whenever certain events occur
  - Support Python 2.5 to make scripting easier and more reliable
- Use automation to:
  - Integrate SPSS with other desktop applications
  - Build custom applications using Visual Basic®, PowerBuilder®, and C++
  - Integrate SPSS into larger custom applications (such as Word or Excel)
- Use the HOST command to take advantage of the operating system functionality in SPSS. This command enables applications to "escape" to the operating system and execute other programs in sync with the SPSS session.

- Prevent syntax jobs from breaking when you create a common or main project directory that enables you to include transformations for multiple projects
- Better manage multiple projects, syntax files, and datasets
- Specify interactive syntax rules using the INSERT command

**Graphic capabilities**

- Categorical charts
  - 3-D Bar: Simple, cluster, and stacked
  - Bar: Simple, cluster, stacked, drop-shadow, and 3-D
  - Line: Simple, multiple, and drop-line
  - Area: Simple and stacked
  - Pie: Simple, exploding, and 3-D effect
  - High-low: High-low-close, difference area, and range bar
  - Boxplot: Simple and clustered
  - Error bar: Simple and clustered
  - Error bars: Add error bars to bar, line, and area charts; confidence level; standard deviation; and standard error
  - Dual-Y axis and overlay

- Scatterplots
  - Simple, grouped, scatterplot matrix, and 3-D
  - Fit lines: Linear, quadratic, or cubic regression, and Lowess smoother; confidence interval control for total or subgroup; and display spikes to line
  - Bin points by color or marker size to prevent overlap

- Density charts
  - Population pyramids: Mirrored axis to compare distributions; with or without normal curve
  - Dot charts: Stacked dots show distribution; symmetric, stacked, and linear
  - Histograms: With or without normal curve; custom binning options

- Quality control charts
  - Pareto
  - X-Bar
  - Range
  - Sigma
  - Individuals
  - Moving range
  - Control chart enhancements include automatic flagging of points that violate Shewhart rules, the ability to turn off rules, and the ability to suppress charts

- Diagnostic and exploratory charts
  - Caseplots and time-series plots
  - Probability plots
  - Autocorrelation and partial autocorrelation function plots
  - Cross-correlation function plots
  - Receiver-Operating Characteristics (ROC)

- Multiple use charts
  - 2-D line charts (both axes can be scale axes)
  - Charts for multiple response sets

- Custom charts
  - Graphics Production Language (GPL), a custom chart creation language, enables advanced users to attain a broader range of chart and option possibilities than the interface supports

- Editing options
  - Automatically reorder categories in differing order (descending or ascending) or by different sort methods (value, label, or summary statistic)
  - Create data value labels
  - Drag to any position on your chart, add connecting lines, and match font color to subgroup
  - Select and edit specific elements directly within a chart: Colors, text, and styles
  - Choose from a wide range of line styles and weights
  - Display gridlines, reference lines, legend ends, titles, footnotes, and annotations
  - Include an Y=X reference line

- Layout options
  - Paneled charts: Create a table of subcharts, one panel per level or condition, showing multiple rows and columns
  - 3-D effects: Rotate, modify depth, and display backplanes

- Chart templates
  - Save selected characteristics of a chart and apply them to others automatically.
  - You can apply the following attributes at creation or editing time: Layout, titles, footnotes and annotations, chart element styles, data element styles, axis scale range, axis scale settings, fit and reference lines, and scatterplot point binning
  - Tree-view layout and finer control of template bundles

- Graph export: BMP, EMF, EPS, JPG, PCT, PNG, TIF, and WMF

*Features subject to change based on final product release.*  
*Symbol indicates a new feature.*
OLAP cubes enable you to:

- Quickly estimate changes in the mean or sum between any two related variables using percent change. For example, easily see how sales increase from quarter to quarter.
- Create case summaries
- Create report summaries
- Generate presentation-quality reports using numerous formatting options
- Generate case listing and case summary reports with statistics on break groups

Frequency tables: Frequency counts, percent, valid percent, and cumulative percent
Option to order your output by analysis or by table
More compact output tables by eliminating extra lines of text where they’re not needed
Central tendency: Mean, median, mode, and sum
Dispersion: Maximum, minimum, range, standard deviation, standard error, and variance
Distribution: Kurtosis and skewness
Z scores: Compute and save as new variables
Display order: Ascending or descending order on means and variable name
Confidence intervals for mean
Descriptives: Interquartile range, kurtosis, kurtosis standard error, median, mean, maximum, minimum, range, skewness, skewness standard error, standard deviation, standard error, variance, five percent trimmed mean, and percentages
M-estimators: Andrew’s wave estimator, Hampel’s M-estimator, Huber’s M-estimator, and Tukey’s biweight estimator
Extreme values and outliers identified
Grouped frequency tables: Bin center, frequency, percent, valid, and cumulative percent
Plots: Construct plots with uniform scale or dependence on data values
- Boxplots: Dependent variables and factor levels together
- Descriptive: Histograms and stem-and-leaf plots
- Normality: Normal probability plots and detrended probability plots with Kolmogorov-Smirnov and Shapiro-Wilk statistics
- Spread versus level plots using Levene’s test: Power estimation, transformed, or untransformed
- Shapiro-Wilk test of normality in EXAMINE allows for 5,000 cases when weights are not specified

Descriptive ratio statistics
- Help for understanding your data using:
  - Coefficient of dispersion
  - Coefficient of variation
  - Price-related differential (PRD)
  - Average absolute deviance
Compare means

Means
- Create better models with harmonic and geometric means
- Cells: Count, mean, standard deviation, sum, and variance
- All-ways totals
- Measure of analysis with Eta and Eta²
- Test of linearity with R and R²
- Results displayed in report, crosstabular, or tree format
- Statistics computed for total sample

\( t \) test
- One sample \( t \) test to compare sample mean to a reference mean of your choice
- Independent sample statistics: Compare sample means of two groups for both pooled and separate-variance estimates with Levene’s test for equal variances
- Paired sample statistics: Correlation between pairs, difference between means, and two-tailed probability for test of no difference and for test of zero correlation between pairs
- Statistics: Confidence intervals, counts, degrees of freedom, mean, two-tailed probability, standard deviation, standard errors, and \( t \) statistic

One-way ANOVA
- Contrasts: Linear, quadratic, cubic, higher-order, and user-defined
- Range tests: Duncan, LSD, Bonferroni, Student-Newman-Keuls, Scheffé, Tukey’s alternate test, and Tukey’s HSD
- Post hoc tests: Student-Newman-Keuls, Tukey’s honestly significant difference, Tukey’s \( b \), Duncan’s multiple comparison procedure based on the Studentized range test, Scheffé’s multiple comparison \( t \) test, Dunnett’s two-tailed \( t \) test, Dunnett’s one-tailed \( t \) test, Bonferroni \( t \) test, least significant difference \( t \) test, Sidak \( t \) test, Hochberg’s GT2, Gabriel’s pairwise comparisons test based on the Studentized maximum modulus test, Ryan-Einot-Gabriel-Welsch’s multiple stepdown procedure based on an \( F \) test, Ryan-Einot-Gabriel-Welsch’s multiple stepdown procedure based on the Studentized range test, Tamhane’s T2, Tamhane’s T3, Games and Howell’s pairwise comparisons test based on the Studentized range test, Dunnett’s C, and Waller-Duncan \( t \) test
- ANOVA statistics: Between- and within-groups sums of squares, degrees of freedom, mean squares, \( F \) ratio, and probability of \( F \)
- Fixed-effects measures: Standard deviation, standard error, and 95 percent confidence intervals
- Random effects measures: Estimate of variance components, standard error, and 95 percent confidence intervals
- Group descriptive statistics: Maximum, mean, minimum, number of cases, standard deviation, standard error, and 95 percent confidence interval
- Homogeneity of variance test: Levene’s test
- Read and write matrix materials
- Equality of means: Reach accurate results when variances and sample sizes vary across different groups
  - Brown-Forsythe test
  - Welch test
- ANOVA models—simple factorial
  - Create custom models without limits on maximum order of interaction
  - Work faster because you don’t have to specify ranges of factor levels
  - Choose the right model using four types of sum of squares
  - Increase certainty with better data handling in empty cells
  - Perform lack-of-fit tests to select your best model
  - Choose from one of two designs: Balanced or unbalanced
  - Use analysis of covariance with up to 10 covariate methods: Classic experimental, hierarchical, and regression
  - Enter covariates control: Before, with, or after main effects
  - Set interaction to: None, 2-, 3-, 4-, or 5-way
  - Select from the following statistics:
    - ANOVA, means and counts table, multiple classification analysis, unstandardized regression coefficients, and n-way cell means
    - Choose up to 10 independent variables
    - Reach predicted values and deviations from the mean in MCA table

Correlate*

Bivariate
- Pearson \( r \), Kendall’s Tau-b, and Spearman
- One- and two-tailed probabilities
- Means, number of non-missing cases, and standard deviations
- Cross-product deviations and covariances
- Coefficients displayed in matrix or serial format

Partial*
- One- and two-tailed probabilities
- Mean, number of non-missing cases, and standard deviation
- Zero-order correlations
- Up to 100 control variables
- Up to five order values
- Correlations displayed in matrix or serial string format, lower triangular, or rectangular correlation matrix

Distances
- Compute proximities between cases or variables
- Dissimilarity measures
  - Interval measure: Euclidean and squared Euclidean distance, Chebychev distance metric, city-block or Manhattan distance, distance in an absolute Minkowski power metric, and customized
  - Counts measures: Chi-square and Phi-square
  - Binary measures: Euclidean and squared Euclidean distance; size, pattern, and shape difference; variance dissimilarity measure; and Lance and Williams nonmetric
- Similarity measures
  - Interval measures: Pearson correlation and cosine
  - Binary measures: Russell and Rao; simple matching; Jaccard; dice (or Czekanowski or Sorenson); Rodgers and Tanimoto; Sokal and Sneath 1 through 5; Kulczynski 1 and 2; Hamann; Goodman and Krusal Lambda; Anderberg’s D; Yule’s coefficient of colligation; Yule’s Q; Ochiai; dispersion similarity measure; and fourfold point correlation
- Standardize data values: Z scores, range of -1 to 1, range of 0 to 1, maximum magnitude of 1, mean of 1, and standard deviation of 1
Multithreaded algorithm, resulting in improved performance and scalability on multiprocessor or multicore machines.

Features subject to change based on final product release. Symbol indicates a new feature.

Regression—linear regression*

Methods: Backward elimination, forced entry, forced removal, forward entry, forward stepwise selection, and R² change/test of significance

Equation statistics: Akaike information criterion (AIC), Ameniya’s prediction criterion, ANOVA tables (F, mean square, probability of F, regression, and residual sum of squares), change in R², F at step, Mallow’s Cp, multiple R, probability of F, R², adjusted R², Schwarz Bayesian criterion (SBC), standard error of estimate, sweep matrix, and variance-covariance matrix

Descriptive statistics: Correlation matrix, covariance matrix, cross-product deviations from the mean, means, number of cases used to compute correlation coefficients, one-tailed probabilities of correlation coefficients, standard deviations, and variances

Independent variable statistics: Regression coefficients, including B, standard errors of coefficients, standardized regression coefficients, approximate standard error of standardized regression coefficients, and t; tolerances; zero-order; part and partial correlations; and 95 percent confidence interval for unstandardized regression coefficient

Variables not in equation: Beta or minimum tolerance

Durbin-Watson

Collinearity diagnostics: Condition indexes, eigenvalues, variance inflation factors, variance proportions, and tolerances

Plots: Casewise, histogram, normal probability, de-trended normal, partial, outlier, and scatterplots

Create and save variables:
- Prediction intervals: Mean and individual
- Predicted values: Unstandardized, standardized, adjusted, and standard error of mean
- Distances: Cook’s distances, Mahalanobis’ distance, and leverage values
- Residuals: Unstandardized, standardized, Studentized, deleted, and Studentized deleted
- Influence statistics: dbetas, standardized dbetas, dffits, standardized dffits, and covariance ratios

Option controls: F-to-enter, F-to-remove, probability of F-to-enter, probability of F-to-remove, suppress the constant, regression weights for weighted least-squares model, confidence intervals, maximum number of steps, replace missing values with variable mean, and tolerance

Regression coefficients displayed in user-defined order

System files can contain parameter estimates and their covariance and correlation matrices through the OUTFILE command

Solutions can be applied to new cases or used in further analysis

Decision making can be further improved throughout your organization when you export your models via XML

Ordinal regression—PLUM*

Predict ordinal outcomes
- Seven options to control the iterative algorithm used for estimation, to specify numerical tolerance for checking singularity, and to customize output
- Five link functions to specify the model: Cauchit, complementary log-log, logit, negative log-log, and probit
- Location subcommand to specify the location model: Intercept, main effects, interactions, nested effects, multiple-level nested effects, nesting within an interaction, interactions among nested effects, and covariates
- Print: Cell information, asymptotic correlation matrix of parameter estimates, goodness-of-fit statistics, iteration history, kernel of the log-likelihood function, test of parallel lines assumption, parameter statistics, and model summary
- Save casewise post-estimation statistics into the active file: Expected probabilities of classifying factor/covariate patterns into response categories and response categories with the maximum expected probability for factor/covariate patterns
- Customize your hypotheses tests by directly specifying null hypotheses as linear combinations of parameters using the TEST subcommand (syntax only)

Curve estimation
- Eleven types of curves are available for specification
- Regression summary displays: Curve type, R² coefficient, degrees of freedom, overall F test and significance level, and regression coefficients
- Trend-regression models available: Linear, logarithmic, inverse, quadratic, cubic, compound, power, S, growth, exponential, and logistic

Nonparametric tests
- Chi-square: Specify expected range (from data or user-specified) and frequencies (all categories equal or user-specified)
- Binomial: Define dichotomy (from data or output) and specify test proportion
- Runs: Specify cutpoints (median, mode, mean, or specified)
- One sample: Kolmogorov-Smirnov, uniform, normal, and Poisson
- Two independent samples: Mann-Whitney U, Kolmogorov-Smirnov Z, Moses extreme, and Wald-Wolfowitz runs
- k-independent samples: Kruskal-Wallis H and median
- 2-related samples: Wilcoxon, sign, and McNemar
- k-related samples: Friedman, Kendall’s W, and Cochran’s Q
- Descriptives: Maximum, mean, minimum, number of cases, and standard deviation

Multiple response
- Crosstabulation tables: Cell counts, cell percentages based on cases or responses, column and row, and two-way table percentages
- Frequency tables: Counts, percentage of cases, or responses
- Both multiple-dichotomy and multiple-response groups can be handled

Data reduction
- Factor*
- Number of cases and variable labels for analysis can be displayed
- Input from correlation matrix, factor, loading matrix, covariance matrix, or raw data case file
- Output of correlation matrix or factor matrix

*Multithreaded algorithm, resulting in improved performance and scalability on multiprocessor or multicore machines.
- Seven extraction methods available for use when analysis is performed on correlation matrices or raw data files: Principal component, principal axis, Alpha factoring, image factoring, maximum likelihood, unweighted least squares, and generalized least squares.
- Rotation methods: Varimax, equamax, quartimax, promax, and oblimin.
- Display: Initial and final communalities, eigenvalues, percent variance, unrotated factor loadings, rotated factor pattern matrix, factor transformation matrix, factor structure, and correlation matrix (oblique rotations only).
- Covariance matrices can be analyzed using three extraction methods: Principal component, principal axis, and image.
- Factor scores: Regression, Bartlett, and Anderson-Rubin.
- Factor scores saved as active variables.
- Statistics available: Univariate correlation matrix, determinant and inverse of correlation matrix, anti-image correlation and covariance matrices, Kaiser-Meyer-Olkin measure of sampling adequacy, Bartlett's test of sphericity, factor pattern matrix, revised communalities, eigenvalues and percent variance by eigenvalue, reproduced and residual correlations, and factor score coefficient matrix.
- Plots: Scree plot and plot of variables in factor space.
- Matrix input and output.
- Post-rotational calculated through sums-of-squares loadings.
- Solutions applied to new cases or to use in further analysis with the SELECT subcommand.
- Factor score coefficient matrix exported to score new data (syntax only).

Classify
TwoStep cluster analysis
- Group observations into clusters based on a nearness criterion. This procedure uses a hierarchical agglomerative clustering procedure in which individual cases are successively combined to form clusters whose centers are far apart. This algorithm is designed to cluster large numbers of cases. It passes the data once to find cluster centers and again to assign cluster memberships. Cluster observations by building a data structure called the CF Tree, which contains the cluster centers. The CF Tree is grown during the first stage of clustering and values are added to its leaves if they are close to the cluster center of a particular leaf.
  - Categorical-level and continuous-level data can be used.
  - Distance measures: Euclidean distance and the likelihood distance.
  - Criteria command tunes the algorithm so that:
    - The initial threshold can be specified to grow a CF Tree.
    - The maximum number of child nodes, a leaf node may have can be set.
    - The maximum number of levels a CF Tree may have can be set.
  - HANDLENOISE subcommand enables you to treat outliers in a special manner during clustering. The default value of noise percent is zero, equivalent to no noise handling. The value can range between zero and 100.
  - INFILE subcommand allows the algorithm to update a cluster model in which a CF Tree is saved as an XML file using the OUTFILE subcommand.
  - MEMALLOCATE subcommand specifies the maximum amount of memory in megabytes (MB) that the cluster algorithm is allowed to use.
  - Missing data: Exclude both user-missing and system-missing values, or let user-missing values be treated as valid.
  - Option to standardize continuous-level variables or leave them at the original scale.
  - Ability to specify the number of clusters, specify the maximum number of clusters, or let the number of clusters be chosen automatically.
  - Algorithms available for determining the number of clusters: BIC or AIC.
  - Output written to a specified filename as XML.
  - Final model output saved, or use an option that updates the model later with more data.
- Plots:
  - Bar chart of frequencies for each cluster.
  - Pie chart showing observation percentages and counts within each cluster.
  - Importance of each variable within each cluster: The output is sorted by the importance rank of each variable.
- Print options:
  - Comparisons (one plot per cluster or one plot per variable).
  - Measure of variable importance (parametric or non-parametric).
  - Ability to specify Alpha level when considering importance.

Cluster
- Use one of six linkage methods to determine clusters: Single linkage (nearest neighbor), average linkage between groups, centroid (average linkage within groups), complete linkage (farthest neighbor), median, and Ward.
  - Provide the same set of similarity and dissimilarity measures as in proximity.
  - Save cluster memberships as new variables.
  - Save distance matrices for use in other procedures.
  - Display: Agglomeration schedules, cluster membership, and distance matrices.
  - Use proximities between variable matrices for improved scalability.
  - Choose from the following plots: Horizontal and vertical icicle plots and dendrogram plots of cluster solutions.
  - Specify case identifiers for tables and plots.
  - Have the ability to accept matrix input and produce matrix output.

Quick cluster
- Squared Euclidean distance.
- Centers selected by widely spaced cases, first K cases, or direct specification.
- Cluster membership saved as a variable.
- Two methods provided for updating cluster centers.
- K-means clustering algorithms.

Features subject to change based on final product release. Symbol indicates a new feature. * Multithreaded algorithm, resulting in improved performance and scalability on multiprocessor or multicore machines.
Discriminant
- Variable selection methods: Direct entry, Wilks’ Lambda minimization, Mahalanobis’ distance, smallest F ratio, minimization of sum of unexplained variation for all pairs, and largest increase in Rao’s V
- Statistics:
  - Summary: Eigenvalues, percent and cumulative percent of variance, canonical correlations, Wilks’ Lambda, and Chi-square tests
  - At each step: Wilks’ Lambda, equivalent F, degrees of freedom, and significance of F for each step; F-to-remove; tolerance; minimum tolerance; F-to-enter; and value of statistic for each variable not in equation
- Final: Standardized canonical discriminant function coefficients, structure matrix of discriminant functions, and functions evaluated within group means
- Optional: Means, standard deviations, univariate F ratios, pooled within-groups covariance and correlation matrices, matrix of pairwise F ratios, Box’s M test, group and total covariance matrices, unstandardized canonical discriminant functions, classification results table, and classification function coefficients
- Rotation of coefficient (pattern) and structure matrices
- Output displayed step by step and/or in summary form
- In classification stage: Prior probabilities, equal, proportion of cases, or user-specified
- All groups, cases, territorial maps, and separate groups plotted
- Casewise results saved to system file for further analysis
- Matrix files read/written, including additional statistics: Counts, means, standard deviations, and Pearson correlation coefficients
- Solutions applied to new cases or for use in further analysis
- Jackknife estimates provided for misclassified error rate
- Decision making further improved by exporting your models throughout your organization via XML

Matrix operations
- Write your own statistical routines in the compact language of matrix algebra

Data management
- Prepare continuous-level data for analysis with the Visual Binner
  - Specify cutpoints in an intelligent manner using a histogram created through a data pass
  - Automatically create value labels based on your cutpoints
- Create your own custom programs with the Output Management System (OMS). Turn output from SPSS procedures into data (SPSS data files, XML, or HTML) and create your programs for bootstrapping, jacknifing and leaving-one-out methods, and Monte Carlo simulations
- Create custom programs in SPSS, even if you have little or no experience with SPSS syntax, using the Output Management System Control Panel
- Easily clean your data when you identify duplicate records through the user interface with the Identify Duplicate Cases tool
- Make sense and keep track of your data files by adding notes to them with the Data File Comments command
- Prevent the accidental destruction of data by making the dataset read-only
- Easily set up all of your value labels to prepare your data for analysis using the Define Variable Properties tool
  - Set up data dictionary information, including value labels and variable types
  - Intelligently add labels because a data pass made first enables SPSS to present a list of values and counts of those values
  - Save time by being able to enter data and value labels directly onto the grid rather than having to use nested dialogs
- Save work by easily copying dictionary information from one variable to another and from one dataset to another using the Copy Data Properties tool
- Copy dictionary information (such as variable and value labels) between variables and datasets using the template facility
- Receive a ready means of cloning dictionaries
- Analyze more data, more efficiently—file size considerations are practically eliminated (especially when used

Features subject to change based on final product release. Symbol indicates a new feature.
- Enter, edit, and browse data in the Data Editor's spreadsheet format
- Easily work with dates and times using the Date and Time Wizard
  - Create a date/time variable from a string containing a date/time variable
  - Create a date/time variable from variables that include individual date units, such as month or year
  - Parse individual date/time units from date/time variables
  - Calculate with dates and times
    - Round instead of truncating date/time information, if desired
    - Add decimal places to time data, if desired
- Display values or value labels in Data Editor cells
- With a right mouseclick, receive direct access to variable information within dialog boxes
- Rename and reorder variables
- Sort cases
- Choose from several data formats: Numeric, comma, dot, scientific notation, date, dollar, custom currency, and string
- Set an option to show currency as comma- or decimal-delimited
- Choose system missing and up to three user-defined missing values per variable
- Create value labels of up to 120 characters (double that of versions prior to SPSS 13.0)
- Create variable labels of up to 256 characters
- Insert and delete variables and cases
- Search for values of a selected variable
- Transpose working files
- Clone or duplicate datasets
- Apply an extended Variable Properties command to customize properties for individual users
- Aggregate data using an extensive set of summary functions
  - Save aggregated values directly to your active file
  - Aggregate by string for source variables (within the interface)
    - Allow the use of long strings as a break variable (e.g., if gender is the break variable, then males and females aggregate separately)
    - Allow the use of strings as the aggregated variable
- Split files to apply analyses and operations to subgroups
- Select cases either permanently or temporarily
- Process first n cases

- Select random samples of cases for analysis
- Select subsets of cases for analysis
- Weigh cases by values of a selected variable
- Specify random number seeds
- Rank data
- Use neighboring observations for smoothing, averaging, and differentiating fast Fourier transformations and their inverse
- More accurately describe your data using longer variable names (up to 64 bytes)
  - Work more easily with data from databases and spreadsheets that include longer variable names than allowed in versions earlier than SPSS 12.0
- Ensure data containing longer text strings (up to 32,767 bytes) is not truncated or lost when working with open-ended question responses, data from other software that allows long text strings, or other types of long text strings
- Find and replace information using the Data Editor
- Save time with spell checking of value labels and variable labels
- Easily inspect data dictionary information in the Variable View of the Data Editor, since you can configure (show only certain attributes) and sort by Variable name, by Type, by Format, etc.
- Easily navigate the Data View in the Data Editor by going directly to a variable
- Add missing values and value labels for strings of any length
- Change string length and variable type through syntax

File management
- Use Unicode when working with multi-lingual data, thus eliminating variability in data due to language-specific encodings.
  - Save the data file either as a Unicode file or as a codepage file (for backwards compatibility with earlier versions of SPSS).
- Truly minimize data handling with conversion-free/copy-free data access in SQL databases. Save time by not needing to convert data into SPSS format (especially when used in conjunction with the optional SPSS Server)
- Set a permanent default starting folder
- Easily write back to databases from SPSS by using the Database Wizard. For example, you can:
  - Create a new table and export it to your database
  - Add new rows to an existing table
- Export data to existing columns in a table
- Import data (including compound documents) from current versions of Excel without needing the Database Wizard
- Read columns that contain mixed data types without any loss of data
- Automatically read columns with mixed data types as string variables and read all values as valid string variables
- Open multiple datasets within a single SPSS session
- Supress the number of datasets in the user interface
- Directly import data from Dimensions™ products, including mrInterview™, and traditional market research products, including Quanvert™
- Export data from SPSS to Dimensions products
- Import from OLE DB data sources without having to go through ODBC
- Read/write Stata® files
- Work more efficiently as you run multiple sessions on one desktop. For example, on lengthy jobs, you can use SPSS in another session as long as the licenses are available.
- Easily read and define ASCII data using a Text Wizard similar to the one provided in Excel
  - Use text qualifiers to make reading in data even easier
- Increase the accuracy and repeatability of your syntax files with search and replace enhancements
- Read database tables using the Database Wizard
  - Drag-and-drop join support
- Export tables and text as ASCII output
- Save tables as HTML and charts as JPG formats to post SPSS results on the Internet or your intranet
- Gain quick access to the SPSS Developer Central Web site through the SPSS Help menu
  - Read/write Excel 2007 files
  - Translate files to and from Excel, Lotus® 1-2-3®, and dBASE®
  - Read and write data to and from fixed, free-field, or tab-delimited ASCII files
  - Write data to fixed-format or tab-delimited ASCII files
  - Read complex file structures: Hierarchical files, mixed record types, repeating data, and non-standard file structures
  - Read and write SPSS/PC++ system files

Features subject to change based on final product release. □ Symbol indicates a new feature. **Supported only on SPSS for Windows
Features subject to change based on final product release. □ Symbol indicates a new feature.

- Merge files
- Display and apply data definitions from an SPSS data file to a working file
- Update master files using transaction files
- Read and write data matrices
- Save many intermediate results for further analysis
- Read recent versions of SAS® files
- Export data files to SAS
- Export data files to current versions of Excel
- Save comma-separated value (CSV) text files from SPSS data files

Transformations
- Compute new variables using arithmetic, cross-case, date and time, logical, missing-value, random-number, and statistical or string functions
- Count occurrences of values across variables
- Recode string or numeric values
- Automatically convert string variables to numeric variables using the autorecode command
  - Use an autorecode template to append existing recode schemes
  - Recode multiple variables simultaneously
  - Autorecode blank strings so that they are defined as "user-missing"
- Create conditional transformations using do if, else if, else, and end if structures
- Use programming structures such as do repeat-end repeat, loop-end loop, and vectors
- Make transformations permanent or temporary
- Execute transformations immediately, in batch mode, or on demand
- Easily find and replace text strings in your data using the find/replace function
- Use cumulative distribution, inverse cumulative distribution, and random number generator functions: Beta, Cauchy, Chi-square, Exponential, F, Gamma, Laplace, logistic, lognormal, Normal, Pareto, Student t, uniform, and Weibull
  - Standard bivariate normal distribution with correlation r, Half Normal, inverse Gaussian, Studentized range, and Studentized maximum modulus
- Work with cumulative distribution and the random number generator for discrete distribution functions: Bernoulli, binomial, geometric, hypergeometric, negative binomial, and Poisson
- Use cumulative distribution for non-central distribution: Non-central Beta, non-central Chi-square, non-central F, and non-central T
- Use density/probability functions for:
  - Continuous distributions: Beta, standard bivariate normal with correlation R, Cauchy, Chi-square, exponential, F, Gamma, half normal random, inverse Gaussian, Laplace, logistic, lognormal, normal, Pareto, Student t, uniform, and Weibull
  - Discrete distributions: Bernoulli, binomial, geometric, hypergeometric, negative binomial, and Poisson
- Use non-central density/probability functions for: Non-central Beta, non-central Chi-square, non-central F distribution, and non-central t distribution
- Select two-tail probabilities: Chi-square & F
- Use auxiliary function: Logarithm of the complete Gamma function

System requirements

- **SPSS Base 16.0 for Windows**
  - Operating System: Microsoft Windows XP (32-bit versions) or Vista™ (32-bit or 64-bit versions)
  - Hardware:
    - Intel® or AMD x86 processor running at 1GHz or higher
    - Memory: 512MB RAM or more; 1GB recommended
    - Minimum free drive space: 450MB
    - CD-ROM drive
    - Super VGA (800x600) or higher-resolution monitor
  - For connecting with an SPSS Server, a network adapter running the TCP/IP network protocol
  - Web browser: Internet Explorer 6

- **SPSS Base 16.0 for MAC OS X**
  - Operating system: Apple Mac OS X 10.4 (Tiger™)
  - Hardware:
    - PowerPC or Intel processor
    - Memory: 512MB RAM; 1GB recommended
    - Minimum free drive space: 800MB
    - CD-ROM drive
    - Super VGA (800x600) or higher-resolution monitor
  - Web browser: Safari™ 1.3.1, Mozilla® Firefox® 1.5, or Netscape® 7.2
  - Java Standard Edition 5.0 (J2SE 5.0)

- **SPSS Base 16.0 for Linux**
  - Operating system: any Linux OS that meets the following requirements***:
    - Kernel 2.4.33.3 or higher
    - glibc 2.3.2 or higher
    - XFree86-4.0 or higher
    - libstdc++5
  - Hardware:
    - Processor: Intel or AMD x86 processor running at 1GHz or higher
    - Memory: 512MB RAM or more; 1GB recommended
    - Minimum free drive space: 450MB
    - CD-ROM drive
    - Super VGA (800x600) or higher-resolution monitor
  - Web browser: Konqueror 3.4.1, Firefox® 1.0.6, or Netscape 7.2

***Note: SPSS 16.0 was tested on and is supported only on Red Hat® Enterprise Linux® 4 Desktop and Debian 3.1

Enterprise products

- **SPSS Server**
  SPSS Server enables SPSS users in your organization to work with large data files for better decision making. The client/server version combines SPSS for Windows with SPSS Server and a wide range of add-on modules to deliver enterprise-strength scalability and enhanced performance.

- **SPSS Adapter for SPSS Predictive Enterprise Services™**
  Enterprise users gain powerful capabilities to manage their analytical assets and processes with the SPSS Adapter. The SPSS Adapter enables SPSS for Windows to integrate into the SPSS Predictive Enterprise Services platform. This enterprise-level application provides you with a centralized, secure, auditable repository for data and models. With it, for example, your organization can:
  - Institutionalize analytics and models and schedule jobs
  - Standardize the use of SPSS transformations and models throughout your organization
  - Regularly refresh information for models and scoring databases
  - Audit analysis conducted for regulatory compliance
SPSS Family
Add more analytical power, as you need it, with optional add-on modules and stand-alone software from the SPSS Family. Unless otherwise noted, the products described below require you to use the corresponding version of SPSS Base to operate.

SPSS Programmability Extension™
Expanded programmability functionality helps make SPSS one of the most powerful statistical development platforms. You can use the external programming language Python® to develop new procedures and applications, including those written in R. You’ll enjoy improved tools for adding these procedures, namely a new user interface and the ability to deliver results to pivot tables in the SPSS Output Viewer. Visit SPSS Developer Central at www.spss.com/devcentral to share code, tools, and programming ideas.

SPSS Regression Models
Predict behavior or events when your data go beyond the assumptions of linear regression techniques. Perform multinomial or binary logistic regression and nonlinear regression, techniques. Perform multinomial or binary logistic regression and nonlinear regression, weighted least squares, two-stage least squares, logistic regression and nonlinear regression, and more.

SPSS Advanced Models
SPSS Advanced Models’ powerful multivariate techniques include generalized linear models (GENLIN), generalized estimating equations (GEE), mixed level models, general linear models (GLM), variance component estimation, MANOVA, Kaplan-Meier estimation, Cox regression, hiloglinear, loglinear, and survival analysis.

SPSS Tables™
Use SPSS Tables to present survey, customer satisfaction, polling, and compliance reporting results. Features such as a table builder preview, included inferential statistics, and data management capabilities make it easy to clearly communicate your results.

SPSS Classification Trees™
Create highly visual classification and decision trees directly within SPSS for segmentation, stratification, prediction, data reduction and variable screening, interaction identification, category merging, and discretizing continuous variables. Highly visual trees enable you to present results in an intuitive manner.

SPSS Exact Tests™ (Windows Only)
SPSS Exact Tests always provides you with correct p-values, regardless of your data structure, even if you have a small number of cases, have subset your data into fine breakdowns, or have variables where 80 percent or more of the responses are in one category.

SPSS Categories™
Unleash the full potential of your categorical data through perceptual maps with optimal scaling and dimension reduction techniques. This add-on module provides you with everything you need to analyze and interpret multivariate data and their relationships more completely.

SPSS Trends™
Improve forecasting with complete time-series analyses, including multiple curve-fitting and smoothing models and methods for estimating autoregressive functions. Use the Expert Modeler to automatically determine which ARIMA (autoregressive integrated moving average) process or exponential smoothing model best fits your time-series and independent variables, eliminating selection through trial and error.

SPSS Conjoint™
SPSS Conjoint helps market researchers develop successful products. By performing conjoint analysis, you learn what product attributes are important in the consumer’s mind and what the most preferred attribute levels are, and can perform pricing studies and brand equity studies.

SPSS Missing Value Analysis™
If values are missing from your data, this procedure may find some relationships between the missing values and other variables. In addition, the missing values procedure can estimate what the value would be if data weren’t missing.

SPSS Data Preparation™
With SPSS Data Preparation, you gain several procedures that facilitate the data preparation process. This add-on module enables you to easily identify suspicious and invalid cases, variables, and data values; view patterns of missing data; summarize variable distributions to get your data ready for analysis; and more accurately work with algorithms designed for nominal attributes. (This add-on module was previously called SPSS Data Validation™.)

SPSS Neural Networks™
Use the new SPSS Neural Networks module to model complex relationships between inputs and outputs or to discover patterns in your data. Choose from algorithms that can be used for classification (categorical outcomes) and prediction (numerical outcomes). The two available algorithms are Multilayer Perceptron and Radial Basis Function.

SPSS Complex Samples™
Incorporate complex sample designs into data analysis for more accurate analysis of complex sample data. SPSS Complex Samples, with specialized planning tools and statistics, reduces the risk of reaching incorrect or misleading inferences for stratified, clustered, or multistage sampling.

Amos™ (Windows only)
Support your research and theories by extending standard multivariate analysis methods when using this stand-alone software package for structural equation modeling (SEM). Build attitudinal and behavioral models that more realistically reflect complex relationships, because any numeric variable, whether observed or latent, can be used to predict any other numeric variable.

SPSS Text Analysis for Surveys™
SPSS Text Analysis for Surveys is a stand-alone software package that offers a combination of linguistic technologies and manual techniques to categorize responses to open-ended questions. To enhance your quantitative analysis, you can export the results as CSV or use Excel.

SPSS Data Entry™ and Dimensions products
SPSS Inc. offers a variety of stand-alone products that help you enter and capture data for survey research. SPSS Data Entry products provide you with options for desktop- or Web-based data entry, useful when networking multiple stations. Dimensions gives you the ability to automatically capture data online, by telephone, through handheld devices, or when using paper forms that you scan. All of these products work with SPSS 15.0 for Windows, enabling you to seamlessly analyze your survey data.
More Accurately Analyze Complex Relationships

Make your analysis more accurate and reach more dependable conclusions with statistics designed to fit the inherent characteristics of data describing complex relationships. SPSS Advanced Models provides a powerful set of sophisticated univariate and multivariate analytical techniques for real-world problems, such as:

- Medical research—Analyze patient survival rates
- Manufacturing—Assess production processes
- Pharmaceutical—Report test results to the FDA
- Market research—Determine product interest levels

Access a wide range of powerful models

In addition to the general linear models (GLM) and mixed models procedures, SPSS Advanced Models now offers the generalized linear models (GENLIN) and generalized estimating equations (GEE) procedures. GENLIN include widely used statistical models, such as linear regression for normally distributed responses, logistic models for binary data, and loglinear models for count data. This procedure also offers many useful statistical models through its very general model formulation, such as ordinal regression, Tweedie regression, Poisson regression, Gamma regression, and negative binomial regression. GEE procedures extend generalized linear models to accommodate correlated longitudinal data and clustered data.

GENLIN and GEE provide a common framework for the following outcomes:

- Numerical: Linear regression, analysis of variance, analysis of covariance, repeated measures analysis, and Gamma regression
- Count data: Loglinear models, logistic regression, probit regression, Poisson regression, and negative binomial regression
- Ordinal data: Ordinal regression
- Event/trial data: Logistic regression
- Claim data: Inverse Gaussian regression
- Combination of discrete and continuous outcomes: Tweedie regression
- Correlated responses within subjects: GEE or correlated response models
Get more accurate predictive models when working with nested-structure data

The linear mixed models procedure expands upon the models used in the GLM procedure so that you can analyze data that exhibit correlation and non-constant variability. This procedure enables you to model not only means but also variances and covariances in your data.

The procedure’s flexibility allows you to formulate a wide variety of models, including fixed effects ANOVA models, randomized complete blocks designs, split-plot designs, purely random effects models, random coefficient models, multilevel analyses, unconditional linear growth models, linear growth models with person-level covariates, repeated measures analyses, and repeated measures analyses with time-dependent covariates. Work with repeated measures designs, including incomplete repeated measurements in which the number of observations varies across subjects.

Build flexible models

The GLM procedure enables you to describe the relationship between a dependent variable and a set of independent variables. Models include linear regression, ANOVA, ANCOVA, MANOVA, and MANCOVA. GLM also includes capabilities for repeated measures, mixed models, post hoc tests and post hoc tests for repeated measures, four types of sums of squares, and pairwise comparisons of expected marginal means, as well as the sophisticated handling of missing cells, and the option to save design matrices and effect files.

Apply more sophisticated models

Use SPSS Advanced Models when your data do not conform to the assumptions required by simpler techniques. SPSS Advanced Models has loglinear and hierarchical loglinear analysis for modeling multiway tables of count data. The general loglinear analysis procedure helps you analyze the frequency counts of observations falling into each cross-classification category in a crosstabulation or contingency table. You can select up to 10 factors to define the cells of a table. Model information and goodness-of-fit statistics are shown automatically. Display a variety of statistics and plots, or save residuals and predicted values in the working data file.

Analyze event history and duration data

You can examine lifetime or duration data to understand terminal events, such as part failure, death, or survival. SPSS Advanced Models includes Kaplan-Meier and Cox regression, state-of-the-art survival procedures. Use Kaplan-Meier estimations to gauge the length of time to an event; use Cox regression to perform proportional hazard regression with time-to-response or duration response as the dependent variable. These procedures, along with life tables analyses, provide a flexible and comprehensive set of techniques for working with your survival data.
Features

GENLIN and GEE

GENLIN procedures provide a unifying framework that includes classical linear models with normally distributed dependent variable, logistic, and probit models for binary data, and loglinear models for count data, as well as various other nonstandard regression-type models. GEE procedures extend the generalized linear model to correlated longitudinal data and clustered data. More particularly, GEE procedures model correlations within subjects.

Users benefit from having a common framework for the following outcomes:
- Continuous outcomes: Linear regression, analysis of variance, analysis of covariance, repeated measures analysis, and Gamma regression
- Ordinal data: Ordinal regression
- Count data: Loglinear models, logistic regression, Poisson regression, and negative binomial regression
- Event/trial data: Logistic regression
- Claim data: Inverse Gaussian regression
- Combination of discrete and continuous outcomes: Tweedie regression
- Correlated responses within subjects: GEE or correlated response models

The MODEL subcommand is used to specify model effects, an offset or scale weight variable if either exists, the probability distribution, and the link function
- Offers an option to include or exclude the intercept
- Specifies an offset variable or fixes the offset at a number
- Specifies a variable that contains Omega weight values for the scale parameter
- Enables users to choose from the following probability distributions: Binomial, Gamma, inverse Gaussian, negative binomial, normal, multinomial ordinal, Tweedie, and Poisson
- Offers the following link functions: Complementary log-log, identity, log, log complement, logit, negative binomial, negative log-log, odds power, probit, cumulative logit, and power

The CRITERIA subcommand controls statistical criteria for GENLIN and specifies numerical tolerance for checking singularity. It provides options to specify the following:
- The type of analysis for each model effect: Type I, Type III, or both
- A value for starting iteration for checking complete and quasi-complete separation
- The confidence interval level for coefficient estimates and estimated marginal means
- Parameter estimate covariance matrix: Model-based estimator or robust estimator
- The Hessian convergence criterion
- Initial values for parameter estimates
- Log-likelihood convergence criterion
- Form of the log-likelihood function
- Maximum number of iterations for parameter estimation and log-likelihood
- Maximum number of steps in step-halving method
- Model parameters estimation method: Fisher scoring method or Newton-Raphson method
- Parameter convergence criterion
- Method of fitting the scale parameter: Maximum likelihood, deviance, Pearson Chi-square, or fixed at a number
- Tolerance value used to test for singularity

The REPEATED subcommand specifies the working correlation matrix structure used by GEE to model correlations within subjects, and controls statistical criteria in the non-likelihood-based iterative fitting algorithm. It provides options to specify the following:
- The within-subject or time effect
- Correlation matrix structure: Independent working correlation matrix, AR(1) working correlation matrix, exchangeable working correlation matrix, fixed working correlation matrix, m-dependent working correlation matrix, and unstructured working correlation matrix
- Whether to adjust the working correlation matrix estimator by the number of non-redundant parameters
- Whether to use the robust or the model-based estimator or the parameter estimate covariance matrix for generalized estimating equations
- The Hessian convergence criterion for the generalized estimating equations
- Maximum iterations
- Relative or absolute parameter convergence criterion
- The number of iterations between updates of the working correlation matrix
- To display estimated marginal means of the dependent variable for all level combinations of a set of factors

The EMMEANS subcommand displays estimated marginal means of the dependent variable for all level combinations of a set of factors. It offers the option to specify the following:
- The cells for which estimated marginal means are displayed
- The covariate values to use when computing the estimated marginal means
- Whether to compute estimated marginal means based on the original scale of the dependent variable or on the link function transformation
- The factor or set of crossed factors, the levels or level combinations which are compared using the contrast type specified by using the CONTRAST keyword
- The type of contrast to use for the levels of the factor, or level combinations of the crossed factors, by using the COMPARE keyword. The following contrast types are available: Pairwise, deviation, difference, Helmert, polynomial, repeated, and simple.
- The method of adjusting the significance level used in tests of the contrasts: Least significant difference, Bonferroni, Sequential Bonferroni, Sidak, and Sequential
- The MISSING subcommand specifies how missing values are handled

Features subject to change based on final product release. □ Symbol indicates a new feature.
The PRINT subcommand offers options to display the following: Correlation matrix for parameter estimates, covariance matrix for parameter estimates, case processing summary, descriptive statistics, goodness of fit, general estimable function, iteration history, Lagrange multiplier test, set of contrast coefficient (L) matrices, model information, parameter estimates and corresponding statistics, model summary statistics, and working correlation matrix.

The SAVE subcommand offers options to save the following to the working data file: Predicted value of the linear predictor, estimated standard error of the predicted value of the linear predictor, predicted value of the mean of the response, confidence interval for the mean of the response, leverage value, raw residual, Pearson residual, deviance residual, standardized Pearson residual, standardized deviance residual, likelihood residual, and Cook’s distance.

The OUTFILE subcommand offers options to save the following to an external file: The parameter correlation matrix and other statistics to an SPSS dataset, the parameter covariance matrix and other statistics to an SPSS dataset, and the parameter estimates and the parameter covariance matrix to an XML file.

MIXED
Expands the general linear model used in the GLM procedure so that data can exhibit correlation and non-constant variability.

Fit the following types of models:
- Fixed effects ANOVA model, randomized complete blocks design, split-plot design, purely random effects model, random coefficient model, multilevel analysis, unconditional linear growth model, linear growth model with person-level covariate, repeated measures analysis, and repeated measures analysis with time-dependent covariate
- Opt to apply frequency weights or regression weights
- Use one of six covariance structures offered:
  - First-Order autoregressive
  - Compound asymmetry
  - Huynh-Feldt
  - Identity
  - Unstructured
  - Variance components
- Depending on the covariance type specified, random effects specified in one RANDOM subcommand may be correlated
- Use one of two estimation methods:
  - Maximum likelihood
  - Restricted maximum likelihood
- Select from a variety of print options:
  - Asymptotic correlation matrix of the fixed effects parameter estimates, asymptotic covariance matrix of the fixed-effects parameter estimates, case processing summary, descriptive statistics, estimated covariance matrix of random effects, iteration history, estimable functions, estimated covariance matrix of residual, solution for fixed-effects and random-effects parameters, and tests for covariance parameters
- Use the REPEATED subcommand to specify the residual covariance matrix in the mixed effects model: Identify the subjects and covariance structure (first-order autoregressive, compound symmetry, Huynh-Feldt, identity, unstructured, and variance components)
- Save fixed predicted values, predicted values, and residuals
- Use the TEST subcommand to customize hypotheses tests by directly specifying null hypotheses as linear combinations of parameters
- Supply divisor for coefficients of random effects
- Save standard error of prediction
- Means subcommand for fixed effects, which displays the dependent variable’s estimated marginal means in the cells and its standard errors for the specified factors

GLM
Describe the relationship between a dependent variable and a set of independent variables.

Select univariate and multivariate lack-of-fit tests
- Regression model
- Fixed effect ANOVA, ANCOVA, MANOVA, and MANCOVA
- Random or mixed ANOVA and ANCOVA
- Doubly multivariate design
- Four types of sums of squares
- Full-paramterization approach to estimate parameters in the model
- General linear hypothesis testing for parameters in the model
- Write a covariance or correlation matrix of the parameter estimates in the model in a matrix data file
- Plots: Spread vs. level, residual, and profile
Features (continued)

- Post hoc tests for observed cell means:
  - Student-Newman-Keuls, Tukey’s honestly significant difference, Tukey’s b, Duncan’s multiple comparison procedure based on the Studentized range test, Scheffé’s multiple comparison t test, Dunnett’s one-tailed t test (compares if the mean at any level is smaller than that of the reference category), Dunnett’s two-tailed t test (compares if the mean at any level is larger than that of the reference category), Bonferroni t test, least significant difference t test, Sidak t test, Hochberg’s GT2, Gabriel’s pairwise comparisons test based on the Studentized maximum modulus test, Ryan-Einot-Gabriel-Welsch’s multiple stepdown procedure based on an F test, Ryan-Einot-Gabriel-Welsch’s multiple stepdown procedure based on the Studentized range test, Tamhane’s T2, Dunnett’s T3, Games and Howell’s pairwise comparisons test based on the Studentized range test, Dunnett’s C, and Waller-Duncan t test
- User-specified error term in post hoc analysis
- Estimated population marginal means for predicted cell means
- Save variables to the active file:
  - Unstandardized predicted values, weighted unstandardized predicted values, unstandardized residuals, weighted unstandardized residuals, deleted residuals, standardized residuals, Studentized residuals, standard errors of predicted value, Cook’s distance, and centered leverage values
- Fractional numbers in LMATRIX, MMATRIX, and KMATRIX subcommands
- Pairwise comparisons of expected marginal means
- Linear hypothesis testing of an effect vs. a linear combination of effects
- Option to save design matrices
- Contrasts: Deviations, simple, difference, Helmert, polynomial, repeated, and special
- Print: Descriptive statistics, tests of homogeneity of variance, parameter estimates, partial Eta², general estimable function table, lack-of-fit tests, observed power for each test, and a set of contrast coefficient (L) matrices

**VARCOMP**

- Variance component estimation
- Estimation methods: ANOVA MINQUE, maximum likelihood (ML), and restricted maximum likelihood (REML)
- Type I and Type III sums of squares for the ANOVA method
- Choices of zero-weight or uniform-weight methods
- Choices of ML and REML calculation methods: Fisher’s scoring method or Newton-Raphson method
- Save variance components estimates and covariance matrices
- Criteria specification: Iterations, convergence, and Epsilon value used as tolerance in checking singularity
- Print: Expected mean squares, iteration history, and sums of squares

**SURVIVAL**

- Analysis of life tables
- Life tables for individual groups
- Interval variable lengths
- Plots: Cumulative survival distribution on log or linear scale, hazard function, and density function
- Comparisons of subgroups
- Plots of the one-minus survival function
- Status variables to indicate if the terminal event occurred for the observation
- Print life tables
- Calculate comparisons of the subgroups: Exact, approximate, conditional, painwise, and compare
- Option to write survival table data records and label records files

**LOGLINEAR**

- General models of multiway contingency tables (syntax only)
  - ML estimation
  - Models: Saturated, hierarchical, or nonhierarchical single degree of freedom partitions and logit
  - Observed and expected frequencies
  - Raw and standardized residuals
  - Parameter estimates
  - Cell weight and structural zero specification
  - Plots of adjusted residual vs. observed/expected counts

- Normal and de-trended probability plots of adjusted residuals
- Likelihood ratio and Pearson Chi-squares
- Contrasts: Deviation, difference, Helmert, simple, repeated, polynomial, and special

**HILOGLINEAR**

- Hierarchical loglinear models for multiway contingency tables
- Simultaneous entry and backward elimination methods
- Print: Frequencies and residuals
- Parameter estimates and partial associations for saturated models
- Criteria specification: Convergence, maximum iterations, probability of Chi-square for model, and maximum steps
- Specified cell weights and maximum order of terms
- Plots of standardized residuals vs. observed and expected counts
- Normal probability plots of standardized residuals
- Pivot table output

**GENLOG**

- Fit loglinear and logit models to count data by means of a generalized linear model approach
- Model fit, using ML estimation under Poisson loglinear model and multinomial loglinear models
- Exponential of the Beta
- GLM approach handles “messy data”
- Cell structure specification
- Model designs are specified through GLM model syntax
- Accommodate structural zeros
- Print Chi-square goodness-of-fit statistics
- Generalized log-odds ratio facility tests whether the specific generalized log-odds ratios are equal to zero, and can print confidence intervals
- Cell statistics include expected cell counts, residual, standardized, adjusted, and deviance residual
- Include generalized residuals facility
- Diagnostic plots include high-resolution scatterplots and normal probability plots of residual statistics
- Print parameter estimates, along with correlations and covariances of the estimates

---

Features subject to change based on final product release. Symbol indicates a new feature.
Save residuals, standardized residuals, adjusted residuals, deviance residuals, and predicted values
Criteria specification: Confidence interval, iterations, convergence, Delta, and Epsilon values used as tolerance in checking singularity

KAPLAN-MEIER
Estimates the length of time to an event using Kaplan-Meier estimation methods
Define factors and strata
Plots: Cumulative hazard functions, cumulative, and log survival
Display censored cases
Save variables to a file: Cumulative number of events, hazard, standard error, and survival function
Statistical display: Cumulative events and survival, mean and median survival times with standard errors, number at risk, requested percentiles, and standard error
Tests for equality of survival distributions: Breslow, Tarone, and logrank
Specify a trend component for factor levels having a metric
Include plots of the one-minus survival function
Status variables to indicate if the terminal event occurred for the observation
Specify strata (subgroups) within categories of factors
Compare the survival distributions for different levels of the factor: Compare all factor levels in a single test, compare each pair of factors, pool the test statistic across all strata, and compare the factor levels for each stratum

COX REGRESSION
Proportional hazards with time-dependent covariates
Contrasts: Deviations, simple, difference, Helmert, polynomial, repeated, special, and indicator
Define strata to estimate separate baseline functions
Methods: Backward and forward stepwise and direct entry
Plots: Cumulative survival, hazard, and log-minus-log plots for each stratum
Removal of variables: Change in likelihood ratio, conditional, and Wald
Save variables to files: Baseline survival and hazard functions and their standard errors, cumulative hazard function, dfbeta, log-minus-log of survival function, residuals, and survival function
Include plots of the one-minus survival function
Status variables to indicate if the terminal event occurred for the observation
Specify ordinal or nominal predictors
Print: Full regression output including overall model statistics for variables in the equation and variables not in the equation, summary information, correlation/covariance matrix of the parameter estimates for the variables in the model, baseline table, and confidence intervals for exponential of Beta
Criteria: Change in parameter estimates for terminating iteration; maximum number of iterations; percentage of change in log-likelihood ratio for terminating iteration; probability of score statistic for variable entry; and probability of Wald, likelihood ratio (LR), or conditional LR statistic to remove a variable
Specify the pattern of covariate values to be used for requested plots and coefficient tables
Write to external SPSS data files: Coefficients in the final model and survival table

System requirements
Software: SPSS Base 16.0
Other system requirements vary according to platform

Features subject to change based on final product release.
Symbol indicates a new feature.
Reveal Relationships in Categorical Data

Unleash the full potential of your data through perceptual mapping, optimal scaling, preference scaling, and dimension reduction techniques. SPSS Categories provides you with all the tools you need to obtain clear insight into complex categorical and high-dimensional data.

With SPSS Categories, you can visually interpret data and see how rows and columns relate in large tables of counts, ratings, or rankings. This gives you the ability to:

- Work with and understand ordinal and nominal data using procedures similar to conventional regression, principal components, and canonical correlation
- Perform regression using nominal or ordinal categorical predictor or outcome variables

For example, use SPSS Categories to understand which characteristics consumers relate most closely to your product or brand, or to determine customer perception of your products compared to other products that you or your competitors offer.

Turn your qualitative variables into quantitative ones

The advanced procedures available in SPSS Categories enable you to perform additional statistical operations on categorical data.

Use SPSS Categories’ optimal scaling procedures to assign units of measurement and zero-points to your categorical data. This opens up a new set of statistical functions by allowing you to perform analyses on variables of mixed measurement levels—on nominal, ordinal, and numeric variables, for example.

SPSS Categories’ ability to perform correspondence and multiple correspondence analyses helps you numerically evaluate similarities between two or more nominal variables in your data.

And, with its principal components analysis procedure, you can summarize your data according to important components. Or incorporate variables of different measurement levels into sets and then analyze them by using nonlinear canonical correlation analysis.
Graphically display underlying relationships
Whatever types of categories you study—market segments, subcultures, political parties, or biological species—optimal scaling procedures free you from the restrictions associated with two-way tables, placing the relationships among your variables in a larger frame of reference. You can see a map of your data—not just a statistical report.

SPSS Categories’ dimension reduction techniques enable you to go beyond unwieldy tables. Instead, you can clarify relationships in your data by using perceptual maps and biplots.

- Perceptual maps are high-resolution summary charts that graphically display similar variables or categories close to each other. They provide you with unique insight into relationships between more than two categorical variables.
- Biplots enable you to look at the relationships among cases, variables, and categories. For example, you can define relationships between products, customers, and demographic characteristics.

By using the preference scaling procedure, you can further visualize relationships among objects. The breakthrough algorithm on which this procedure is based enables you to perform non-metric analyses for ordinal data and obtain meaningful results.

How you can use SPSS Categories
The following procedures are available to add meaning to your data analyses.

Categorical regression (CATREG) predicts the values of a nominal, ordinal, or numerical outcome variable from a combination of categorical predictor variables that the procedure quantifies through optimal scaling techniques.

You can use regression with optimal scaling to describe, for example, how job satisfaction relates to job category, geographic region, and the amount of work-related travel.

Correspondence analysis (CORRESPONDENCE) enables you to analyze two-way tables that contain some measurement of correspondence between the rows and columns. A very common type of correspondence table is a crosstabulation in which the cells contain frequency counts.

SPSS Categories displays relationships among nominal variables in a perceptual map, a visual presentation that also shows the relationships among the categories of the variables.

Multiple correspondence analysis (MULTIPLE CORRESPONDENCE) is used to analyze multivariate categorical data. It differs from correspondence analysis in that it allows you to use more than two variables in your analysis. With this procedure, all the variables are analyzed at the nominal level (unordered categories).
For example, you can use multiple correspondence analysis to explore relationships between favorite television show, age group, and gender. By examining a low-dimensional map created with SPSS Categories, you could see which groups gravitate to each show while also learning which shows are most similar.

**Categorical principal components analysis (CATPCA)** uses optimal scaling to generalize the principal components analysis procedure so that it can accommodate variables of mixed measurement levels. It is similar to multiple correspondence analysis, except that you are able to specify an analysis level on a variable-by-variable basis.

For example, you can display the relationships between different brands of cars and characteristics such as price, weight, fuel efficiency, etc. Alternatively, you can describe cars by their class (compact, midsize, convertible, SUV, etc.), and CATPCA uses these classifications to group the points for the cars. SPSS Categories displays results in a low-dimensional map that makes it easy to understand relationships.

**Nonlinear canonical correlation analysis (OVERALS)** uses optimal scaling to generalize the canonical correlation analysis procedure so that it can accommodate variables of mixed measurement levels. This type of analysis enables you to compare multiple sets of variables to one another in the same graph, after removing the correlation within sets.

For example, you might analyze characteristics of products, such as soups, in a taste study. The judges represent the variables within the sets while the soups are the cases. OVERALS averages the judges’ evaluations, after removing the correlations, and combines the different characteristics to display the relationships between the soups. Alternatively, each judge may have used a separate set of criteria to judge the soups. In this instance, each judge forms a set and OVERALS averages the criteria, after removing the correlations, and then combines the scores for the different judges.

**Multidimensional scaling (PROXSCAL)** performs multidimensional scaling of one or more matrices containing similarities or dissimilarities (proximities). Alternatively, you can compute distances between cases in multivariate data as input to PROXSCAL.

PROXSCAL displays proximities as distances in a map in order for you to gain a spatial understanding of how objects relate. In the case of multiple proximity matrices, PROXSCAL analyzes the commonalities and plots the differences between them.

For example, you can use PROXSCAL to display the similarities between different cola flavors preferred by consumers in various age groups. You might find that young people emphasize differences between traditional and new flavors, while adults emphasize diet versus non-diet colas.

**Preference scaling (PREFSCAL)** visually examines relationships between variables. Preference scaling performs multidimensional unfolding on two sets of objects in order to find a common quantitative scale.

This enables you to find clusters among variables. For example, if a group of drivers rated 26 models of cars on ten attributes on a six-point scale, you could find clusters showing which models were similar, and which attributes were associated with them.
Better Understand Consumer Perceptions

Market researchers in South Australia wanted to better understand how consumers perceived six brands of iced coffee. They asked consumers to rate each of the brands (denoted AA to FF in Figure 1) on 16 different categorical attributes. The 96-cell table that resulted made it difficult for analysts to clearly see the relationships between the brands and the perceived attributes.

The market researchers used the correspondence procedure in SPSS to identify the two strongest underlying factors in the relationships between the brands and attributes. By assigning each brand and attribute a specific number within each dimension, the information was displayed in an easily understood chart, commonly called a perceptual map. For example, it is clear from Figure 1 that Brand AA is the brand most closely identified by the market with the “popular” attribute. Similarly, researchers can quickly identify that consumers who are interested in healthy and low-fat products perceive CC and DD more positively, while FF is perceived as a rich, sweet premium brand.*

Features
Statistics
CATREG
- Categorical regression analysis through optimal scaling
  - Specify the optimal scaling level at which you want to analyze each variable.
    Choose from: Spline ordinal (monotonic), spline nominal (nonmonotonic), ordinal, nominal, multiple nominal, or numerical.
  - Discretize continuous variables or convert string variables to numeric integer values by multiplying, ranking, or grouping values into a preselected number of categories according to an optional distribution (normal or uniform), or by grouping values in a preselected interval into categories. The ranking and grouping options can also be used to recode categorical data.
  - Specify how you want to handle missing data. Impute missing data with the variable mode or with an extra category, or use listwise exclusion.
  - Specify objects to be treated as supplementary
  - Specify the method used to compute the initial solution
  - Control the number of iterations
  - Specify the convergence criterion
- Plot results, either as:
  - Specify the number of dimensions in the solution
  - Choose from five types of standardization: Remove row means, remove column means, remove row-and-column means, equalize row totals, or equalize column totals
  - Five types of normalization: Symmetrical, principal, row principal, column principal, and customized
  - Print results, including:
    - Correspondence table
    - Summary table: Singular values, inertia, proportion of inertia accounted for by the dimensions, cumulative proportion of inertia accounted for by the dimensions, confidence statistics for the maximum number of dimensions, row profiles, and column profiles
    - Overview of row and column points: Mass, scores, inertia, contribution of the points to the inertia of the dimensions, and contribution of the dimensions to the inertia of the points
    - Row and column confidence statistics: Standard deviations and correlations for active row and column points
- Tables for fit and model parameters: ANOVA table with degrees of freedom according to optimal scaling level; model summary table with adjusted R² for optimal scaling, t values, and significance levels; a separate table with the zero-order, part and partial correlation, and the importance and tolerance before and after transformation
  - Correlations of the transformed predictors and eigenvalues of the correlation matrix
  - Correlations of the original predictors and eigenvalues of the correlation matrix
  - Category quantifications
  - Write discretized and transformed data to an external data file

CORRESPONDENCE
- Correspondence analysis
  - Input data as a case file or directly as table input
  - Specify the number of dimensions in the solution
  - Choose from two distance measures: Chi-square distances for correspondence analysis or Euclidean distances for biplot analysis types
  - Choose from five types of standardization: Remove row means, remove column means, remove row-and-column means, equalize row totals, or equalize column totals
  - Five types of normalization: Symmetrical, principal, row principal, column principal, and customized
  - Print results, including:
    - Correspondence table
    - Summary table: Singular values, inertia, proportion of inertia accounted for by the dimensions, cumulative proportion of inertia accounted for by the dimensions, confidence statistics for the maximum number of dimensions, row profiles, and column profiles
    - Overview of row and column points: Mass, scores, inertia, contribution of the points to the inertia of the dimensions, and contribution of the dimensions to the inertia of the points
    - Row and column confidence statistics: Standard deviations and correlations for active row and column points

MULTIPLE CORRESPONDENCE
- Multiple correspondence analysis (replaces HOMALS, which was included in versions prior to SPSS Categories 13.0)
  - Specify variable weights
  - Discretize continuous variables or convert string variables to numeric integer values by multiplying, ranking, or grouping values into a preselected number of categories according to an optional distribution (normal or uniform), or by grouping values in a preselected interval into categories. The ranking and grouping options can also be used to recode categorical data.
  - Specify how you want to handle missing data. Exclude only the cells of the data matrix without valid value, impute missing data with the variable mode or with an extra category, or use listwise exclusion.
  - Specify objects and variables to be treated as supplementary (full output is included for categories that occur only for supplementary objects)
  - Specify the number of dimensions in the solution
  - Specify a file containing the coordinates of a configuration and fit variables in this fixed configuration
  - Choose from five normalization options: Variable principal (optimizes associations between variables), object principal (optimizes distances between objects), symmetrical (optimizes relationships between objects and variables), independent, or customized (user-specified value allowing anything in between variable principal and object principal normalization)
  - Control the number of iterations
  - Specify convergence criterion
  - Print results, including:
    - Model summary
    - Iteration statistics and history

Features subject to change based on final product release.
- Add transformed variables and object scores to the working data file
- Write discretized data, transformed data, and object scores to an external data file

**CATPCA**
- Categorical principal components analysis through optimal scaling
  - Specify the optimal scaling level at which you want to analyze each variable.
    Choose from: Spline ordinal (monotonic), spline nominal (nonmonotonic), ordinal, nominal, multiple nominal, or numerical.
  - Specify variable weights
  - Discretize continuous variables or convert string variables to numeric integer values by multiplying, ranking, or grouping values into a preselected number of categories according to an optional distribution (normal or uniform), or by grouping values in a preselected interval into categories. The ranking and grouping options can also be used to recode categorical data.
  - Specify how you want to handle missing data. Exclude only the cells of the data matrix without valid value, impute missing data with the variable mode or with an extra category, or use listwise exclusion.
  - Specify objects and variables to be treated as supplementary (full output is included for categories that occur only for supplementary objects)
  - Specify the number of dimensions in the solution
  - Specify a file containing the coordinates of a configuration and fit variables in this fixed configuration
  - Choose from five normalization options: Variable principal (optimizes associations between variables), object principal (optimizes distances between objects), symmetrical (optimizes relationships between objects and variables), independent, or customized (user-specified value allowing anything in between variable principal and object principal normalization)
  - Control the number of iterations
  - Specify convergence criterion
  - Print results, including:
    - Model summary
    - Iteration statistics and history
    - Descriptive statistics (frequencies, missing values, and mode)
    - Variance accounted for by variable and dimension
    - Component loadings
    - Category quantifications and category coordinates (vector and/or centroid coordinates) for each dimension
    - Correlations of the transformed variables and the eigenvalues of the correlation matrix

- Correlations of the original variables and the eigenvalues of the correlation matrix
- Object (component) scores
- Plot results, creating:
  - Category plots: Category points, transformation (optimal category quantifications against category indicators), residuals for selected variables, and joint plot of category points for a selection of variables
  - Object scores
  - Object contributions: Mass, inertia, contribution of the objects to the inertia of the dimensions, and contribution of the dimensions to the inertia of the objects
- Print weights, initial configurations, and fixed coordinates
- Optionally transform proximities with linear, ordinal, smooth ordinal, or spline functions
- Specify multidimensional unfolding with identity, weighted Euclidean, or generalized Euclidean models
- Specify fixed row and column coordinates to restrict the configuration
- Specify initial configuration (classical triangle, classical Spearman, Ross-Cliff, correspondence, centroids, random starts, or custom), iteration criteria, and penalty parameters
- Specify plots for multiple starts, initial common space, stress per dimension, final common space, space weights, individual spaces, scatterplot of fit, residuals plot, transformation plots, and Shepard plots
- Specify output that includes the input data, multiple starts, initial common space, iteration history, fit measures, stress decomposition, final common space, space weights, individual spaces, fitted distances, and transformed proximities
- Write common space coordinates, individual weights, distances, and transformed proximities to a file

**PREFSCAL** (syntax only)
- Visually examine relationships between variables in two sets of objects in order to find a common quantitative scale
- Read one or more rectangular matrices of proximities
- Read weights, initial configurations, and fixed coordinates
- Optionally transform proximities with linear, ordinal, smooth ordinal, or spline functions
- Specify multidimensional unfolding with identity, weighted Euclidean, or generalized Euclidean models
- Specify fixed row and column coordinates to restrict the configuration
- Specify initial configuration (classical triangle, classical Spearman, Ross-Cliff, correspondence, centroids, random starts, or custom), iteration criteria, and penalty parameters
- Specify plots for multiple starts, initial common space, stress per dimension, final common space, space weights, individual spaces, scatterplot of fit, residuals plot, transformation plots, and Shepard plots
- Specify output that includes the input data, multiple starts, initial common space, iteration history, fit measures, stress decomposition, final common space, space weights, individual spaces, fitted distances, and transformed proximities
- Write common space coordinates, individual weights, distances, and transformed proximities to a file

**System requirements**
- Software: SPSS Base 16.0
- Other system requirements vary according to platform

To learn more, please visit www.spss.com.
For SPSS office locations and telephone numbers, go to www.spss.com/worldwide.

SPSS is a registered trademark and the other SPSS products named are trademarks of SPSS Inc. All other names are trademarks of their respective owners. © 2007 SPSS Inc. All rights reserved.
Easily Identify Groups and Predict Outcomes

The SPSS Classification Trees add-on module creates classification and decision trees directly within SPSS to help you better identify groups, discover relationships between groups, and predict future events. You can use classification and decision trees for segmentation, stratification, prediction, data reduction and variable screening, interaction identification, category merging, and discretizing continuous variables.

Highly visual diagrams enable you to present categorical results in an intuitive manner—so you can more clearly explain categorical results to non-technical audiences. These trees enable you to explore your results and visually determine how your model flows. Visual results can help you find specific subgroups and relationships that you might not uncover using more traditional statistics. Because classification trees break the data down into branches and nodes, you can easily see where a group splits and terminates.

Use SPSS Classification Trees in a variety of applications, including:

- Database marketing
  - Choose a response variable to segment your customer base (for example, responders/non-responders in a test mailing; high-, medium-, and low-profit customers; or recruits who have extended service versus those who haven’t)
  - Profile groups based on other attributes, such as demographics or customer activity
  - Customize new promotions to focus on a specific subgroup, help reduce costs, and improve return on investment (ROI)
- **Market research**
  - Perform customer, employee, or recruit satisfaction surveys
  - Choose a variable that measures satisfaction (for example, on a “1-5” scale)
  - Profile satisfaction levels according to responses to other questions
  - Change factors, such as work environment or product quality, that can affect satisfaction
- **Credit risk scoring**
  - Determine risk groups (high, medium, or low)
  - Profile risk groups based on customer information, such as account activity
  - Offer the right credit line to the right applicants based on risk group
- **Program targeting**
  - Choose a variable with a desirable versus undesirable outcome (for example, successful completion of a welfare-to-work program)
  - Reveal the factors that lead to success, based on applicant information
  - Customize new programs to satisfy the needs of more people
- **Marketing in the public sector**
  - Choose a response variable for segmenting your customer base (for example, potential college applicants who actually applied versus those who haven’t)
  - Profile groups based on other attributes, such as demographics or customer activity
  - Customize new promotions to focus on a specific subgroup, help reduce costs, and improve ROI

Use the highly visual trees to discover relationships that are currently hidden in your data. SPSS Classification Trees’ diagrams, tables, and graphs are easy to interpret.

Use tree model results to score cases directly in SPSS.
Choose from four decision tree algorithms
SPSS Classification Trees includes four established tree-growing algorithms:

- **CHAID**—A fast, statistical, multi-way tree algorithm that explores data quickly and efficiently, and builds segments and profiles with respect to the desired outcome
- **Exhaustive CHAID**—A modification of CHAID that examines all possible splits for each predictor
- **Classification & regression trees (C&RT)**—A complete binary tree algorithm that partitions data and produces accurate homogeneous subsets
- **QUEST**—A statistical algorithm that selects variables without bias and builds accurate binary trees quickly and efficiently

With four algorithms, you have the ability to try different tree-growing methods and find the one that best fits your data.

Extend your results with further analysis within SPSS
Since you use SPSS Classification Trees within the SPSS interface, you can create classification trees directly in SPSS and conveniently use the results to segment and group cases directly within the data. There is no back and forth between SPSS and other software. Additionally, you can generate selection or classification/prediction rules in the form of SPSS syntax, SQL statements, or simple text (through syntax). You can display these rules in the Viewer and save them to an external file for later use to make predictions about individual and new cases. If you’d like to use your results to score other data files, you can write information from the tree model directly to your data or create XML models for use in SPSS Server.
**Features**

**Trees**
- Display tree diagrams, tree maps, bar graphs, and data tables
- Easily build trees using the comprehensive interface, which enables the setup of:
  - Measurement level (nominal, ordinal, and continuous)
  - Independent variables
  - Dependent variables
  - Influence variables
  - Growing method
  - Output setup, which includes trees, statistics, charts, and rules
  - Split sample validation or cross-validation
  - Stopping criteria
  - Saved variables, including predicted values, probability, and XML models
- Choose from four tree-growing methods
- View nodes using one of several methods: Show bar charts or tables of your target variables, or both, in each node
- Collapse and expand branches, and change other cosmetic properties, such as fonts and colors
- View and print trees
- Specify the exact zoom percentage for viewing visual tree models in the interface
- Automate tree building using the production mode
  - Automatically generate syntax from the interface
- Force one predictor into the model
- Specify prior probabilities, misclassification costs, revenues, expenses, and scale scores

**Tree-growing algorithms**
- Perform analysis using one of four powerful tree-growing algorithms:
  - CHAID by Kass (1980)
  - Exhaustive CHAID by Biggs, de Ville, and Suen (1991)
  - Classification & regression trees (C&RT) by Breiman, Friedman, Olshen, and Stone (1984)
  - QUEST by Loh and Shih (1997)
- Handle missing predictor data using one of two methods: Assign to a category or impute using a surrogate
- Discretize continuous predictor variables according to the number of categories specified
- Have pruning capabilities for C&RT and QUEST
- Randomly sample source data for split sample validation or use a variable to split the sample

**Model evaluation**
- Generate risk and classification tables
- Summarize node performance with evaluation graphs and tables to help identify the best segments:
  - Gains
  - Index (lift)
  - Response
  - Mean
  - Average profit
  - ROI
- Partition data between training and test data to verify accuracy
- Display summary graphs or classification rules for selected nodes using the node summary window

**Deployment**
- Export:
  - Tree diagrams, charts, and tables. Export formats include: HTML, text, Word/RTF, and Excel® files, and PDF
  - Save information from the model as variables in the working data file
  - Export decision rules that define selected segments in SQL to score databases, as SPSS syntax to score SPSS files, or as simple text (through syntax)
  - Export trees as XML models for use with SPSS Server and SmartScore® to score new cases or data files
  - Publish trees as images and tables as static or interactive tables to SmartViewer® Web Server™
  - For additional insight, select interesting segments in the working data file via tree nodes, and run more analyses

**System requirements**
- Software: SPSS Base 16.0
- Other system requirements vary according to platform
When you conduct sample surveys, use a statistics package dedicated to producing correct estimates for complex sample data. SPSS Complex Samples provides specialized statistics that enable you to correctly and easily compute statistics and their standard errors from complex sample designs. You can apply it to:

- Survey research—Obtain descriptive and inferential statistics for survey data
- Market research—Analyze customer satisfaction data
- Health research—Analyze large public-use datasets on public health topics such as health and nutrition or alcohol use and traffic fatalities
- Social science—Conduct secondary research on public survey datasets
- Public opinion research—Characterize attitudes on policy issues

SPSS Complex Samples provides you with everything you need for working with complex samples. It includes:

- An intuitive Sampling Wizard that guides you step by step through the process of designing a scheme and drawing a sample
- An easy-to-use Analysis Preparation Wizard to help prepare public-use datasets that have been sampled, such as the National Health Inventory Survey data from the Centers for Disease Control and Prevention (CDC)
- Numerical outcome prediction through the Complex Samples General Linear Model (CSGLM)
- Ordinal outcome prediction through Complex Samples Ordinal Regression (CSORDINAL)
- Categorical outcome prediction through Complex Samples Logistic Regression (CSLOGISTIC)
- Time to an event prediction through Complex Samples Cox Regression (CSCOXREG)

From the planning stage and sampling through the analysis stage, SPSS Complex Samples makes it easy to obtain accurate and reliable results. Since SPSS Complex Samples takes up to three states into account when analyzing data from a multistage design, you’ll end up with more accurate analyses. In addition to giving you the ability to assess your design’s impact, SPSS Complex Samples also produces a more accurate picture of your data because subpopulation assessments take other subpopulations into account.

You can use the following types of sample design information with SPSS Complex Samples:

- Stratified sampling—Increase the precision of your sample or ensure a representative sample from key groups by choosing to sample within subgroups of the survey population. For example, subgroups might be a specific number of males or females, or contain people in certain job categories or people of a certain age group.
- Clustered sampling—Select clusters, which are groups of sampling units, for your survey. Clusters can include schools, hospitals, or geographic areas with sampling units that might be students, patients, or citizens. Clustering often helps make surveys more cost-effective.
- Multistage sampling—Select an initial or first-stage sample based on groups of elements in the population, then create a second-stage sample by drawing a subsample from each selected unit in the first-stage sample. By repeating this option, you can select a higher-stage sample.
More confidently reach results
As a researcher, you want to be confident about your results. Most conventional statistical software assumes your data arise from simple random sampling. Simple random sampling, however, is generally neither feasible nor cost-effective in most large-scale surveys. Analyzing such sample data with conventional statistics risks incorrect results. For example, estimated standard errors of statistics are often too small, giving you a false sense of precision. SPSS Complex Samples enables you to achieve more statistically valid inferences for populations measured in your complex sample data because it incorporates the sample design into survey analysis.

Work efficiently and easily
Only SPSS Complex Samples makes understanding and working with your complex sample survey results easy. Through the intuitive interface, you can analyze data and interpret results. When you’re finished, you can publish datasets and include your sampling or analysis plans. Each plan acts as a template and allows you to save all the decisions made when creating it. This saves time and improves accuracy for yourself and others who may want to use your plans with the data, either to replicate results or pick up where you left off.

A grocery store wants to determine if the frequency with which customers shop is related to the amount spent, controlling for gender of the customer and incorporating a sample design. First, the store specifies the sample design used in the Analysis Preparation Wizard (top). Next, the store sets up the model in the Complex Samples General Linear Model (bottom).
To begin your work in SPSS Complex Samples, use the wizards, which prompt you for the many factors you must consider. If you are creating your own samples, use the Sampling Wizard to define the sampling scheme. If you’re using public-use datasets that have been sampled, such as those provided by the CDC, use the Analysis Preparation Wizard to specify how the samples were defined and how to estimate standard errors. Once you create a sample or specify standard errors, you can create plans, analyze your data, and produce results (see the diagram above for workflow).

Accurate analysis of survey data is easy in SPSS Complex Samples. Start with one of the wizards (which one to select depends on your data source) and then use the interactive interface to create plans, analyze data, and interpret results.

SPSS Complex Samples makes it easy to learn and work quickly. Use the online help system, explore the interactive case studies, or run the online tutorial to learn more about using your data with the software. SPSS Complex Samples enables you to:

- Reach correct point estimates for statistics such as totals, means, and ratios
- Obtain the standard errors of these statistics
- Produce correct confidence intervals and hypothesis tests
- Predict numerical outcomes
- Predict ordinal outcomes
- Predict categorical outcomes
- Predict time to an event
Features

Complex Samples Plan (CSPLAN)
This procedure provides a common place to specify the sampling frame to create a complex sample design or analysis specification used by companion procedures in the SPSS Complex Samples add-on module. CSPLAN does not actually extract the sample or analyze data. To sample cases, use a sample design created by CSPLAN as input to the CSSELECT procedure (described on the following pages). To analyze sample data, use an analysis design created by CSPLAN as input to the CSDESCRIPTIONS, CSTABULATE, CSGLM, CSLOGISTIC, or CSORDINAL procedures (described on the following pages).

- Create a sample design: Use to extract sampling units from the active file
- Create an analysis design: Use to analyze a complex sample
- When you create a sample design, the procedure automatically saves an appropriate analysis design to the plan file. A plan file is created for designing a sample, and therefore, can be used for both sample selection and analysis.
- Display a sample design or analysis design
- Specify the plan in an external file
- Name planwise variables to be created when you extract a sample or use it as input to the selection or estimation process with the PLANVARS subcommand
  - Specify final sample weights for each unit to be used by SPSS Complex Samples analysis procedures in the estimation process
  - Indicate overall sample weights that will be generated when the sample design is executed in the CSSELECT procedure
  - Select weights to be used when computing final sampling weights in a multistage design
- Control output from the CSPLAN procedure with the PRINT subcommand
  - Display a plan specifications summary in which the output reflects your specifications at each stage of the design
  - Display a table showing MATRIX specifications
- Signal stages of the design with the DESIGN subcommand. You can also use this subcommand to define stratification variables and cluster variables or create descriptive labels for particular stages.

- Specify the sample extraction method using the METHOD subcommand. Select from a variety of equal- and unequal-probability methods, including simple and systematic random sampling. Methods for sampling with probability proportionate to size (PPS) are also available. Units can be drawn with replacement (WR) or without replacement (WOR) from the population.
  - SIMPLE_WOR: Select units with equal probability. Extract units without replacement.
  - SIMPLE_WR: Select units with equal probability. Extract units with replacement.
  - SIMPLE_SYSTEMATIC: Select units at a fixed interval throughout the sampling frame or stratum. A random starting point is chosen within the first interval.
  - SIMPLE_CHROMY: Select units sequentially with equal probability. Extract units without replacement.
  - PPS_WOR: Select units with probability proportional to size. Extract units without replacement.
  - PPS_WR: Select units with probability proportional to size. Extract units with replacement.
  - PPS_SYSTEMATIC: Select units by systematic random sampling with probability proportional to size. Extract units without replacement.
  - PPS_CHROMY: Select units sequentially with probability proportional to size. Extract units without replacement.
  - PPS_BREWER: Select two units from each stratum with probability proportional to size. Extract units without replacement.
  - PPS_MURTHY: Select two units from each stratum with probability proportional to size. Extract units without replacement.
  - PPS_SAMPFORD: Extends Brewer's method to select more than two units from each stratum with probability proportional to size. Extract units without replacement.
- Control for the number or percentage of units to be drawn: Set at each stage of the design. You can also choose output variables, such as stagewise sampling weights, which are created upon the sample design execution.
  - Estimation methods: With replacement, equal probability without replacement in the first stage, and unequal probability without replacement
- You can choose whether to include the finite population correction when estimating the variance under simple random sampling (SRS)
- Unequal probability estimation without replacement: Request in the first stage only
- Variable specification: Specify variables for input for the estimation process, including overall sample weights and inclusion probabilities

- Specify the number of sampling units drawn at the current stage using the SIZE subcommand
- Specify the percentage of units drawn at the current stage. For example, specify the sampling fraction using the RATE subcommand.
- Specify the minimum number of units drawn when you specify RATE. This is useful when the sampling rate for a particular stratum is very small due to rounding.
- Specify the maximum number of units to draw when you specify RATE. This is useful when the sampling rate for a particular stratum is larger than desired due to rounding.
- Specify the measure of size for population units in a PPS design. Specify a variable that contains the sizes or request that sizes be determined when the CSSELECT procedure scans the sample frame.
- Obtain stagewise sample information variables when you execute a sample design using the STAGEVARS subcommand. You can obtain:
  - The proportion of units drawn from the population at a particular stage using stagewise inclusion (selection) probabilities
  - Prior stages using cumulative sampling weight for a given stage
  - Uniquely identified units that have been selected more than once when your sample is done with replacement, with a duplication index for units selected in a given stage
  - Population size for a given stage
  - Number of units drawn at a given stage
  - Stagewise sampling rate
  - Sampling weight for a given stage

Features subject to change based on final product release. □ Symbol indicates a new feature.
- Choose an estimation method for the current stage with the ESTIMATOR subcommand. You can indicate:
  - Equal selection probabilities without replacement
  - Unequal selection probabilities without replacement
  - Selection with replacement
- Specify the population size for each sample element with the POPSIZE subcommand
- Specify the proportion of units drawn from the population at a given stage with the INCLPROB subcommand

**Complex Samples Descriptive**

*COMPLEX SAMPLES DESCRIPTIVES*

CSDESCRIPTIVES estimates means, sums, and ratios, and computes their standard errors, design effects, confidence intervals, and hypothesis tests for samples drawn by complex sampling methods. The procedure estimates variances by taking into account the sample design used to select the sample, including equal probability and PPS methods, and WR and WOR sampling procedures. Optionally, CSDESCRIPTIVES performs analyses for subpopulations.

- Specify the name of a plan file, which is written by the CSPLAN procedure, containing analysis design specifications with the PLAN subcommand
- Specify joint inclusion probabilities file names
- Specify the analysis variables used by the MEAN and SUM subcommands using the SUMMARY subcommand
- Request that means and sums be estimated for variables specified on the SUMMARY subcommand through the MEAN and SUM subcommands
- Request t tests of the population mean(s) and sums and give the null hypothesis value(s) through the TTEST keyword. If you define subpopulations using the SUBPOP subcommand, then null hypothesis values are used in the test(s) for each subpopulation, as well as for the entire population.
- Request that ratios be estimated for variables specified on the SUMMARY subcommand through the RATIO subcommand
- Request t tests of the population ratios and give the null hypothesis value(s) through the TTEST keyword

**Complex Samples Selection (CSSELECT)**

CSSELECT selects complex, probability-based samples from a population. It chooses units according to a sample design created through the CSPLAN procedure.

- Control the scope of execution and specify a seed value with the CRITERIA subcommand
- Control whether or not user-missing values of classification (stratification and clustering) variables are treated as valid values with the CLASSMISSING subcommand
- Use the most updated Mersenne Twister random number generator to select the sample
- Specify general options concerning input and output files with the DATA subcommand
  - Opt to rename existing variables when the CSSELECT procedure writes sample weight variables and stagewise output variables requested in the plan file, such as inclusion probabilities
- Write sampled units to an external file using an option to keep/drop specified variables
- Automatically save first-stage joint inclusion probabilities to an external file when the plan file specifies a PPS, WR sampling method
- Opt to generate text files containing a rule that describes characteristics of selected units
- Control output display through the PRINT subcommand
  - Summarize the distribution of selected cases across strata. Information is reported per design stage.
  - Produce a case-processing summary

**Complex Samples Tabulate (CSTABULATE)**

CSTABULATE displays one-way frequency tables or two-way crosstabulations and associated standard errors, design effects, confidence intervals, and hypothesis tests for samples drawn by complex sampling methods. The procedure estimates variances by taking into account the sample design used to select the sample, including equal probability and PPS methods, and WR and WOR sampling procedures. Optionally, CSTABULATE creates tables for subpopulations.

- Specify the name of an XML file, written by the CSPLAN procedure, containing analysis design using the PLAN subcommand
- Specify the joint inclusion probabilities file name
- Use the following statistics within the table:
  - Population size: Estimate the population size for each cell and marginal in a table
  - Standard error: Calculate the standard error for each population size estimate
- Row and column percentages: Express the population size estimate for each cell in a row or column as a percentage of the population size estimate for that row or column. This functionality is available for two-way crosstabulations.
- Table percentages: Express the population size estimate in each cell of a table as a percentage of the population size estimate for that table
- Coefficient of variation
- Design effects
- Square root of the design effects
- Confidence interval: Specify any number between zero and 100 as the confidence interval
- Unweighted counts: Use unweighted counts as the number of valid observations in the dataset for each population size estimate
- Cumulative population size estimates: Use cumulative population size estimates for one-way frequency tables only
- Cumulative percentages: Use cumulative percentages corresponding to the population size estimates for one-way frequency tables only
- Expected population size estimates: Use expected population size estimates if the population size estimates of each cell in the two variables in the crosstabulation are statistically independent. This functionality is available for two-way crosstabulations only.
- Residuals: Show the difference between the observed and expected population size estimates in each cell. This functionality is available for two-way crosstabulations only.
- Pearson residuals: This functionality is available for two-way crosstabulations only
- Adjusted Pearson residuals: This functionality is available for two-way crosstabulations only
- Specify how to handle missing data
  - Base each table on all valid data for the tabulation variable(s) used in creating the table. You may base tables for different variables on different sample sizes.
  - Use only cases with valid data for all tabulation variables in creating the tables. Always base tables for different variables on the same sample size.
  - Exclude user-missing values among the strata, cluster, and subpopulation variables
  - Include user-missing values among the strata, cluster, and subpopulation variables. Treat user-missing values for these variables as valid data.

**Complex Samples General Linear Model (CSGLM)**

This procedure enables you to build linear regression, analysis of variance (ANOVA), and analysis of covariance (ANCOVA) models for samples drawn using complex sampling methods. The procedure estimates variances by taking into account the sample design used to select the sample, including equal probability and PPS methods, and WR and WOR sampling procedures. Optionally, CSGLM performs analyses for subpopulations.

- **Models**
  - Main effects
  - All n-way interactions
  - Fully crossed
  - Custom, including nested terms
- **Statistics**
  - Model parameters: Coefficient estimates, standard error for each coefficient estimate, t test for each coefficient estimate, confidence interval for each coefficient estimate, design effect for each coefficient estimate, and square root of the design effect for each coefficient estimate
  - Population means of dependent variable and covariates
  - Model fit
  - Sample design information
- **Hypothesis tests**
  - Test statistics: Wald F test, adjusted Wald F test, Wald Chi-square test, and adjusted Wald Chi-square test
  - Adjustment for multiple comparisons: Least significant difference, Bonferroni, sequential Bonferroni, Sidak, and sequential Sidak
- **Sampling degrees of freedom:** Based on sample design or fixed by user
- **Estimated means:** Requests estimated marginal means for factors and interactions in the model
- **Contrasts:** Simple, deviation, Helmert, repeated, or polynomial
- Model variables can be saved to the active file and/or exported to external files that contain parameter matrices
- **Variables:** Predicted values and residuals
- Parameter covariance matrix and its other statistics, as well as parameter correlation matrix and its other statistics, can be exported as an SPSS data file
- Parameter estimates and/or the parameter covariance matrix can be exported to an XML file

**Complex Samples Ordinal (CSORDINAL)**

CSORDINAL performs regression analysis on a binary or ordinal polytomous dependent variable using the selected cumulative link function for samples drawn by complex sampling methods. The procedure estimates variances by taking into account the sample design used to select the sample, including equal probability and PPS methods, as well as...
WR and WOR sampling procedures. Optionally, CSORDINAL performs analyses for a subpopulation.

- **Models**
  - Main effects
  - All n-way interactions
  - Fully crossed
  - Custom, including nested terms

- **Statistics:**
  - Model parameters: Coefficient estimates, exponentiated estimates, standard error for each coefficient estimate, t test for each coefficient estimate, confidence interval for each coefficient estimate, design effect for each coefficient estimate, square root of the design effect for each coefficient estimate, covariances of parameter estimates, and correlations of the parameter estimates
  - Model fit: Pseudo $R^2$ and classification
  - Parallel lines tests: Wald tests of equal slopes, parameter estimates for generalized (unequal slopes) model, and covariances of parameter estimates for generalized (unequal slopes) model
  - Summary statistics for model variables
  - Sample design information

- **Hypothesis tests**
  - Test statistics: Wald F test, adjusted Wald F test, Wald Chi-square test, and adjusted Wald Chi-square test
  - Adjustment for multiple comparisons: Least significant difference, Bonferroni, sequential Bonferroni, Sidak, and sequential Sidak
  - Sampling degrees of freedom: Based on sample design or fixed by user

- **Output**
  - Sample design information (such as strata and PSUs)
  - Summary information about the dependent variable, covariates, and factors
  - Summary information about the sample, including the unweighted count and the population size
  - Confidence limits for parameter estimates and user-specified confidence levels
  - Model summary statistics
  - Wald F test, adjusted Wald F test, Wald Chi-square, and adjusted Wald Chi-square for model effects
  - Design effects
  - Classification table
  - Set of contrast coefficients (L) matrices
  - Variance-covariance matrix of regression coefficients
  - General estimable function table
  - Correlation matrix for regression coefficients
  - Missing data handling
  - Listwise deletion of missing values
  - Other
  - User-specified denominator, df, used in computing $p$ values for all test statistics
  - Collinearity diagnostics
  - Fits model for a subpopulation

- **Complex Samples Logistic Regression (CSLOGISTIC)**

This procedure performs binary logistic regression analysis, as well as multinomial logistic regression (MLR) analysis, for samples drawn by complex sampling methods. CSLOGISTIC estimates variances by taking into account the sample design used to select the sample, including equal probability and PPS methods, and WR and WOR sampling procedures. Optionally, CSLOGISTIC performs analyses for subpopulations.

- **Models**
  - Main effects
  - All n-way interactions

- **Statistics**
  - Model parameters: Coefficient estimates, exponentiated estimates, standard error for each coefficient estimate, t test for each coefficient estimate, confidence interval for each coefficient estimate, design effect for each coefficient estimate, square root of the design effect for each coefficient estimate, covariances of parameter estimates, and correlations of the parameter estimates
  - Model fit: Pseudo $R^2$ and classification table
  - Summary statistics for model variables
  - Sample design information

- **Output**
  - Sample design information (such as strata and PSUs)
  - Summary information about the dependent variable, covariates, and factors
  - Summary information about the sample, including the unweighted count and population size
  - Confidence limits for parameter estimates and user-specified confidence levels
  - Model summary statistics
  - Wald F test for model effects
  - Design effects
  - Classification table
  - Set of contrast coefficients (L) matrices
  - Variance-covariance matrix of regression coefficients
– Root mean square error
– Covariance and correlation matrices for regression coefficients
– Missing data handling
  – Listwise deletion of missing values
– Other
  – User-specified denominator, df, used in computing p values for all test statistics
  – Collinearity diagnostics
  – Model can be fitted for subpopulations

Complex Samples Cox Regression (CSCOXREG)
This procedure applies Cox proportional hazards regression to analysis of survival times—that is, the length of time before the occurrence of an event for samples drawn by complex sampling methods. CSCOXREG supports continuous and categorical predictors, which can be time-dependent. CSCOXREG provides an easy way of considering differences in subgroups as well as analyzing effects of a set of predictors. Also, the procedure handles data where there are multiple cases (such as patient visits, encounters, and observations) for a single subject.

Time and Event: specify survival time variables and values that indicate that the event of interest has occurred
– Survival time
  – Start of interval (onset of risk)
    – Time 0
    – Varies by subject
  – End of interval
  – Event as individual values or a range of values

Predictors:
– Factors
– Covariates
– Time-dependent predictors

Subgroups: stratify the analysis and/or limit it to a particular subpopulation.

Models
– Main effects
– All n-way interactions
– Custom, including nested terms

Statistics:
– Sample design information
– Event and censoring summary
– Risk set at event time
– Model parameters: Coefficient estimates, exponentiated estimates, standard error for each coefficient estimate, test for each coefficient estimate, confidence interval for each coefficient estimate, design effect for each coefficient estimate, square root of the design effect for each coefficient estimate, covariances of parameter estimates, and correlations of the parameter estimates
– Model assumptions
  – Test of proportional hazards
  – Parameter estimates for alternative model
  – Covariance matrix for alternative model
  – Baseline survival and cumulative hazard functions

Plots:
– Survival function
– Hazard function
– Log minus log of the survival function
– One minus survival function
– Option to display confidence intervals
– Plot factors and covariates at specified levels

Hypothesis tests
– Test Statistics: F test, Adjusted F test, Chi-square test, Adjusted Chi-square test
– Adjustment for multiple comparisons: Least significant difference, Bonferroni, Sequential Bonferroni, Sidak, and sequential Sidak
– Sampling degrees of freedom: based on sample design or fixed by user

Save model variables to the active file and/or export external files that contain parameter matrices
– Variables: Survival function, lower bound of confidence interval for survival function, upper bound of confidence interval for survival function, cumulative hazard function, lower bound of confidence interval for cumulative hazard function, upper bound of confidence interval for cumulative hazard function, predicted value of linear predictor, Schoenfeld residual (one variable per model parameter), Martingale residual, deviance residual, Cox-Snell residual, score residual (one variable per model parameter), DFBeta residual (one variable per model parameter), aggregated Martingale residual, aggregated deviance residual, aggregated Cox Snell residual, aggregated Score residual (one variable per model parameter), and aggregated DFBETA residual (one variable per model parameter)

Export the model and/or the survival function
– Export as SPSS data file
– Export survival function as SPSS data file
– Export model as XML file

Options to specify estimation criteria, methods for computing survival functions and confidence intervals, and handling of user-missing values
– Estimation: Maximum iterations, maximum step halving, limit iterations based on change in parameter estimates, limit iterations based on change in log-likelihood, display iteration history, and tie breaking method for parameter estimation (Efron or Breslow)
– Survival functions: method for estimating baseline survival functions (Efron, Breslow or product-limit), and confidence intervals for survival functions (transformed or original units)
– Specify level of confidence interval
– Missing Data Handling (treat as valid or invalid)

System requirements
– Software: SPSS Base 16.0
– Other system requirements vary according to platform
In the real world, buyers do not make decisions based on a single attribute, such as price or brand name. Instead, they examine a range of products, all with different combinations of features and attributes, and perform a complex series of trade-offs before reaching a decision. Conjoint analysis is the research tool used to model the consumer’s decision-making process. Using SPSS Conjoint can increase your understanding of consumer preferences, enabling you to more effectively design, price, and market successful products.

Conjoint analysis enables you to measure the value consumers place on individual attributes or features that define products and services. Armed with this knowledge, your company can design products that include the features most important to your target market, set prices based on the value the market assigns to the product’s attributes, and focus messages on the points most likely to appeal to target buyers.

Even as competitors, products, and pricing change over time in the market, you can continue to use the results from SPSS Conjoint to develop market simulation models that incorporate changes, along with your proposed responses. This enables you to predict the response to your proposed actions before spending valuable resources on product development and marketing programs.

SPSS Conjoint provides answers to your critical questions
- Which features or attributes of a product or service drive the purchase decision?
- Which feature combinations will have the most success?
- What market segment is most interested in the product?
- What marketing messages will most appeal to that segment?
- What feature upgrades will most affect consumer preference and increase sales?
- What is the optimal price to charge consumers for a product or service?
- Can the price be increased without a significant loss in sales?
- Are product levels too close together?
SPSS Conjoint gives you all the tools you need
The three procedures in SPSS Conjoint enable you to plan, implement, and efficiently analyze results from conjoint studies. Following is a summary of these procedures.

- Generate designs easily — Orthoplan produces an orthogonal array of product attribute combinations, which dramatically reduces the number of questions you must ask while ensuring that you have enough information to perform a full analysis.

- Print “cards” to elicit respondents’ preferences — Plancards quickly generates cards that respondents can use to easily sort and rank product attribute combinations.

- Get informative results — The conjoint procedure performs a specially tailored version of regression on your response rankings. You’ll receive results you can act on, such as which product attributes are important and at what levels consumers most prefer them. You can also perform simulations to determine the market share of preference for any combination of attributes.

Four ways to make your product launch a success
1. Right product — Design your product with the feature set for which the market has the greatest need.
2. Right price — Price your product based on the value your target audience assigns to it.
3. Right place — Predict how your product/price combination will perform in the market before committing valuable development and launch resources.
4. Right promotion — Focus your marketing on the individual features that most interest your target audience.
Offer options consumers prefer: A real-life study

A software company planned to develop training programs that differed from its traditional instructor-led training. Since many options were available, the company decided to perform a conjoint study to evaluate the proposed product. The company believed six key attributes would influence consumer preference: method of delivery, video content, example types, certification test, method of asking questions remotely, and price. Four of these attributes had two levels, while two others had three. The resulting full factorial design would have had 144 alternative product bundles (2x2x2x2x3x3), making for an unfeasibly large study. Using orthoplan, the research department reduced the number of hypothetical product bundles (such as those shown in Figure 1) to 16, while ensuring that the department received all the information needed to perform a complete analysis. A researcher then printed the 16 product bundles using plancards and gave them to a sample of target users who ranked them in order of preference.

A researcher analyzed the preference rankings with SPSS Conjoint, and the results are shown in Figure 2. Two attributes stand out as very important—inclusion of video and price—while test and example types are relatively unimportant. The Utility Estimate and Standard Error columns in Figure 2 show the relative preference for each level of each attribute. Within question, Instant Message is the most preferred attribute level and No Support is the least preferred.
Features

Orthoplan
- Generate orthogonal main effects fractional factorial designs; orthoplan is not limited to two-level factors
- Specify variable list, optional variable labels, a list of values for each variable, and optional value labels
- Specify the desired number of cards for the plan; orthoplan will try to generate a plan in the desired minimum number of runs
- Generate holdout cards to test the fitted conjoint model
- Mix the training and holdout cards or stack the holdout cards after the training cards
- Save the plan file as an SPSS system file
- Display output in pivot tables

Plancards
- Use this utility procedure to produce printed cards for a conjoint experiment; the printed cards are used as stimuli to be sorted, ranked, or rated by the subjects
- Specify the variables to be used as factors and the order in which their labels are to appear in the output
- Choose a format
  - Listing-file format: Differentiate holdout cards from experimental cards, and then list simulation cards
  - Card format: Holdout cards are not differentiated and simulation cards are not produced
- Write the cards to an external file or the listing file
- Specify optional title and footer
- Specify pagination so that each new card in single-card format begins on a new page
- Display output in pivot tables

Conjoint
- Perform an ordinary least-squares analysis of preference or rating data with this procedure
- Work with the plan file generated by plancards, or a plan file input by the user using a data list
- Work with individual level rank or rating data
- Provide individual level and aggregate results
- Treat the factors in any of a number of ways; conjoint indicates reversals
  - Discrete: Factor levels are categorical
  - Linear: Scores or ranks are linearly related to the factor
  - Ideal: A quadratic relationship is expected between the scores or ranks and the factor; this method assumes that there is an ideal level for the factor, and that distance from the ideal point in either direction is associated with decreasing preference
  - Antideal: A quadratic relationship is expected between the scores or ranks and the factor; this method assumes that there is a worst level for the factor, and that distance from this point in either direction is associated with increasing preference
- Work with experimental cards that have one of three scenarios
  - Training
  - Holdout
  - Simulation
- Select from three conjoint simulation methods
  - Max utility
  - Bradley-Terry-Luce (BTL)
  - Logit
- Print controls
  - Print only the results of the experimental (training and holdout) data analysis
  - Print only the results of the conjoint simulation
  - Print results of both the experimental data analysis and the conjoint simulation
- Write utilities to an external file
- Show print results with:
  - Attribute importance
  - Utility (part-worth) and standard error
  - Graphical indication of most to least preferred levels of each attribute
  - Counts of reversals and reversal summary
  - Pearson R for training and holdout data
  - Kendall’s Tau for training and holdout data
  - Simulation results and simulation summary
- Display output in pivot tables

System requirements
- Software: SPSS Base 16.0
- Other system requirements vary according to platform

Features subject to change based on final product release.
All researchers have to prepare their data prior to analysis. While SPSS Base includes tools for data preparation, sometimes you need more specialized techniques to get your data ready. With the SPSS Data Preparation* add-on module, you can easily identify suspicious or invalid cases, variables, and data values; view patterns of missing data; summarize variable distributions; and more accurately work with algorithms designed for nominal attributes. This streamlines the data preparation process—so that you can get ready for analysis faster and reach more accurate conclusions.

**Perform data checks**
Data validation has typically been a manual process. You might run a frequency on your data, print the frequencies, circle what needs to be fixed, and check for case IDs. Needless to say, this is time consuming. And since every analyst in your organization could use a slightly different method, maintaining consistency from project to project may be a challenge.

To eliminate manual checks, use the Validate Data procedure. This procedure enables you to apply rules to perform data checks based on each variable’s measure level (whether categorical or continuous). For example, if you’re analyzing survey data that has variables on a five-point Likert scale, use the Validate Data procedure to apply a rule for five-point scales and flag all cases that have values outside of the 1-5 range. You can receive reports of invalid cases as well as summaries of rule violations and the number of cases affected. You can specify validation rules for individual variables (such as range checks) and cross-variable checks (for example, “pregnant males”).

With this knowledge you can determine data validity and remove or correct suspicious cases at your discretion prior to analysis.

*SPSS Data Preparation was previously called SPSS Data Validation™.*
**Quickly find multivariate outliers**
Prevent outliers from skewing analyses when you use the Anomaly Detection procedure. This procedure searches for unusual cases based upon deviations from similar cases and gives reasons for such deviations. You can flag outliers by creating a new variable. Once you have identified unusual cases, you can further examine them and determine if they should be included in your analyses.

**Preprocess data prior to model building**
In order to use algorithms that are designed for nominal attributes (such as Naïve Bayes and logit models), you must bin your scale variables prior to model building. If scale variables aren’t binned, algorithms such as multinominal logistic regression will take an extremely long time to process or they might not converge. This is especially true if you have a large dataset. In addition, the results you receive may be difficult to read or interpret.

Optimal Binning, however, enables you to determine cutpoints to help you reach the best possible outcome for algorithms designed for nominal attributes.

With this procedure, you can select from three types of binning for preprocessing data prior to model building:
- **Unsupervised**: Create bins with equal counts
- **Supervised**: Take the target variable into account to determine cutpoints. This method is more accurate than unsupervised; however, it is also more computationally intensive.
- **Hybrid approach**: Combines the unsupervised and supervised approaches. This method is particularly useful if you have a large amount of distinct values.
Features

Validate data
Use the Validate Data procedure to validate data in the working data file
- Basic checks: Specify basic checks to apply to variables and cases in your file. For example, obtain reports that identify variables with a high percentage of missing values or empty cases.
  - Maximum percentage of missing values
  - Maximum percentage of cases in a single category
  - Maximum percentage of cases with a count of 1
  - Minimum coefficient of variation
  - Minimum standard deviation
  - Flag incomplete IDs
  - Flag duplicate IDs
  - Flag empty cases
- Standard rules: Describe the data, view single variable rules, and apply them to analysis variables
  - Description of data:
    - Distribution: Shows a thumbnail-size bar chart for categorical variables or a histogram for scale variables
    - Minimum and maximum data values are shown
  - Single-variable rules:
    - Apply rules to individual variables to identify missing or invalid values, such as values outside a valid range
    - User-defined single-variable rules are also possible
  - Custom rules: Define cross-variable rule expressions in which respondents’ answers violate logic (“pregnant males,” for example)
- Output: Reports describing invalid data
  - Casewise report, which lists the validation rule violations by case
  - Specify the minimum number of violations needed for a case to be included in the report
  - Specify the maximum number of cases in the report
- Standard validation rules reports
  - Summarize violations by analysis variable
  - Summarize violations by rule
  - Display descriptive statistics
- Save: Enables you to save variables that record rule violations and use them to help clean data and filter out bad cases
  - Summary variables:
    - Empty case indicator
    - Duplicate ID indicator
    - Incomplete ID indicator
    - Validation rule violation (total count)
  - Indicator variables that record all validation rule violations

Identify unusual cases
The Anomaly Detection procedure searches for unusual cases, based upon deviations from their peer group, and gives reasons for such deviations
- Specify variables to be used by the procedure with the VARIABLES subcommand. Specify categorical, continuous, and ID variables (to identify cases), and list variables that are excluded from the analysis.
- The HANDLEMISSING subcommand specifies the methods of handling missing values in this procedure
  - Apply missing value handling. If this option is selected, grand means are substituted for missing values of continuous variables, and missing categories of categorical variables are combined and treated as a valid category. The processed variables are then used in the analysis. If this option is not selected, cases with missing values are excluded from the analysis.
  - Create an additional Missing Proportion Variable and use it in the analysis. If chosen, an additional variable called the Missing Proportion Variable that represents the proportion of missing variables in each record is created, and this variable is used in the analysis. If it is not chosen, the Missing Proportion Variable is not created.

The CRITERIA subcommand specifies the following settings:
- Minimum and maximum number of peer groups
- Adjustment weight on the measurement level
- Number of reasons in the anomaly list
- Percentage of cases considered as anomalies and included in the anomaly list
- Number of cases considered as anomalies and included in the anomaly list
- Cutpoint of the anomaly index to determine whether a case is considered an anomaly

Save additional variables to the working data file with the SAVE subcommand
- Anomaly index
- Peer group ID
- Peer group size
- Peer group size in percentage
- The variable, associated with a reason
- The variable impact measure, associated with a reason
- The variable value, associated with a reason
- The norm value, associated with a reason
- Write the model to a specified filename as XML with the OUTFILE subcommand

Control the display of the output results with the PRINT subcommand. You can print:
- Case-processing summary
- The anomaly index list, the anomaly peer ID list, and the anomaly reason list
- The Continuous Variable Norms table, if any continuous variable is used in the analysis, and the Categorical Variable Norms, if any categorical variable is used in the analysis
- Anomaly Index Summary
- Reason Summary Table for each reason
- Suppress all displayed output except the notes table and any warnings
Optimal Binning

Preprocess data using Optimal Binning, which categorizes one or more continuous variables by distributing the values of each variable into bins. This procedure is useful for reducing the number of values in the given binning input variables, which can greatly improve the performance of algorithms. When using certain Optimal Binning methods, a guide variable helps you determine the cutpoints, thereby maximizing the relationship between the guide variable and the binned variable.

- **Specify the following criteria:**
  - How to define the minimum cutpoint for each binning input variable
  - How to define the maximum cutpoint for each binning input variable
  - How to define the lower limit of an interval
  - Whether to force merging of sparsely populated bins
  - Whether missing values are handled using listwise or pairwise deletion

- **Save the following:**
  - New variables containing binned values
  - Syntax to an SPSS syntax file

- **Control output results display with the PRINT subcommand.** You can print:
  - The binning input variables’ cutpoint sets
  - Descriptive information for all binning input variables
  - Model entropy for binned variables

**System requirements**

- **Software:** SPSS Base 16.0
- **Other system requirements vary according to platform**

---

To learn more, please visit www.spss.com. For SPSS office locations and telephone numbers, go to www.spss.com/worldwide.

SPSS is a registered trademark and the other SPSS products named are trademarks of SPSS Inc. All other names are trademarks of their respective owners. © 2007 SPSS Inc. All rights reserved.

`SDP16/SPCA4-0707`
More Accurately Analyze Small Datasets

To determine if a relationship between variables exists, researchers often first look at p values in crosstabulations and nonparametric tests. Traditional methods for computing them are fine if your data meet the underlying assumptions. However, if you have a small number of case variables with a high percentage of responses in one category, or have to subset your data into fine breakdowns, traditional tests could be incorrect. SPSS Exact Tests eliminates this risk.

You'll find SPSS Exact Tests particularly useful if you perform data mining or database marketing analysis for direct marketing, survey research, medical research, biostatistics, social science research, market research, or to conduct any type of experiment.

Keep your original categories
Don’t lose valuable information by collapsing categories to meet the assumptions of traditional tests. With SPSS Exact Tests, you can keep your original design or natural categories—for example, regions, income, or age groups—and analyze what you intend to analyze.

Easily interpret and apply exact tests
Exact tests are easy to run. You can calculate them anytime with just a press of a button—during your original analysis or when you rerun it. With SPSS Exact Tests, there is no steep learning curve because you don’t need to learn any new statistical theories or procedures. You simply interpret the exact tests results the same way you already interpret the results in SPSS Base. And, you’ll always have the right statistical test for your data situation. More than 30 exact tests cover the entire spectrum of nonparametric and categorical data problems for small or large datasets.

Use small samples credibly
If securing a large sample size is impossible or costly, SPSS Exact Tests enables you to use small samples and still feel confident about the results. With the money saved by using smaller sample sizes, you can conduct surveys or test direct marketing programs more often. Stay ahead of your competition by using these resources to find new opportunities.

Obtain more value from your data
With SPSS Exact Tests, you can “slice and dice” your data into breakdowns, which can be as fine as you want, so you learn more by extending your analysis to subgroups. You aren’t limited by required expected counts of five or more per cell for correct results. And you can even rely on SPSS Exact Tests when you’re searching for rare occurrences within large datasets.

Obtain more value from your data
With SPSS Exact Tests, you can “slice and dice” your data into breakdowns, which can be as fine as you want, so you learn more by extending your analysis to subgroups. You aren’t limited by required expected counts of five or more per cell for correct results. And you can even rely on SPSS Exact Tests when you’re searching for rare occurrences within large datasets.
Features
The following tests and statistics are available with SPSS Exact Tests. SPSS Base includes the asymptotic versions of these tests. All results are produced as SPSS pivot tables/report cubes.

### Pearson Chi-square test
- Exact 1-tailed and 2-tailed p values for 2x2 table
- Exact 2-tailed p value for general RxC table
- Monte Carlo 2-tailed p value and confidence intervals (CIs) for general RxC table

### Likelihood ratio test
- Exact 1-tailed and 2-tailed p values for 2x2 table
- Exact 2-tailed p value for general RxC table
- Monte Carlo 2-tailed p value and CIs for general RxC table

### Fisher’s exact test
- Exact 1-tailed and 2-tailed p values for 2x2 table
- Exact 2-tailed p value for general RxC table
- Monte Carlo 2-tailed p value and CIs for general RxC table

### Linear-by-linear association test
- Exact 1-tailed and 2-tailed p values and exact point probability
- Monte Carlo 1-tailed and 2-tailed p values and CIs

### Contingency coefficient
- Exact 2-tailed p value
- Monte Carlo 2-tailed p value and CIs

### Phi
- Exact 2-tailed p value
- Monte Carlo 2-tailed p value and CIs

### Cramer’s V
- Exact 2-tailed p value
- Monte Carlo 2-tailed p value and CIs

### Goodman and Kruskal Tau
- Exact 2-tailed p value
- Monte Carlo 2-tailed p value and CIs

### Uncertainty coefficient—symmetric or asymmetric
- Exact 2-tailed p value
- Monte Carlo 2-tailed p value and CIs

### Kappa
- Exact 2-tailed p value
- Monte Carlo 2-tailed p value and CIs

### Gamma
- Exact 2-tailed p value
- Monte Carlo 2-tailed p value and CIs

### Kendall’s Tau-b and Tau-c
- Exact 2-tailed p value
- Monte Carlo 2-tailed p value and CIs

### Somers’ D—symmetric and asymmetric
- Exact 2-tailed p value
- Monte Carlo 2-tailed p value and CIs

### Pearson’s R
- Exact 2-tailed p value
- Monte Carlo 2-tailed p value and CIs

### Spearman correlation
- Exact 2-tailed p value
- Monte Carlo 2-tailed p value and CIs

### McNemar test
- Exact 1-tailed and 2-tailed p values and point probability

### Sign test
- Exact 1-tailed and 2-tailed p values and point probability
- Monte Carlo 1-tailed and 2-tailed p values and CIs

### Wilcoxon signed-rank test
- Exact 1-tailed and 2-tailed p values and point probability
- Monte Carlo 1-tailed and 2-tailed p values and CIs

### Marginal homogeneity test
- Asymptotic, exact, Monte Carlo 1-tailed and two 2-tailed p values, and point probability

### 2-Sample Kolmogorov-Smirnov test
- Exact 2-tailed p value and point probability
- Monte Carlo 2-tailed p value and CIs

### Mann-Whitney U or Wilcoxon rank-sum W test
- Exact 1-tailed and 2-tailed p values and point probability
- Monte Carlo 1-tailed and 2-tailed p values and CIs

### Wald-Wolfowitz runs test
- Exact 1-tailed p value and point probability
- Monte Carlo 1-tailed p value and CIs

### Cochran’s Q test
- Exact 2-tailed p value and point probability
- Monte Carlo 2-tailed p value and CIs

### Friedman test
- Exact 2-tailed p value and point probability
- Monte Carlo 2-tailed p value and CIs

### Kendall’s coefficient of concordance
- Exact 2-tailed p value and point probability
- Monte Carlo 2-tailed p value and CIs

### Kruskal-Wallis test
- Exact 2-tailed p value and point probability
- Monte Carlo 2-tailed p value and CIs

### Median test
- Exact 2-tailed p value and point probability
- Monte Carlo 2-tailed p value and CIs

### Jonckheere-Terpstra test
- Asymptotic, exact, Monte Carlo 1-tailed and 2-tailed p values, and point probability

### 1-Sample Chi-square test
- Exact 2-tailed p value and point probability
- Monte Carlo 2-tailed p value and CIs

### 1-Sample Kolmogorov-Smirnov test
- Exact 2-tailed p value and point probability
- Monte Carlo 2-tailed p value and CIs

### 1-Sample Wald-Wolfowitz runs test
- Exact 2-tailed p value and point probability
- Monte Carlo 2-tailed p value and CIs

### Binomial test
- Both exact 1-tailed and 2-tailed p values and point probability

### System requirements
- Software: SPSS Base 16.0
- Requires Microsoft® Windows®

to learn more, please visit www.spss.com. for SPSS office locations and telephone numbers, go to www.spss.com/worldwide.

SPSS is a registered trademark and the other SPSS products named are trademarks of SPSS Inc. All other names are trademarks of their respective owners. © 2007 SPSS Inc. All rights reserved. SET16SPCA4-0607
Build Better Models When You Fill in the Blanks

When you ignore or exclude missing data, you risk finding invalid and insignificant results. Use SPSS Missing Value Analysis to impute your missing data and draw more valid conclusions. SPSS Missing Value Analysis is a critical tool for anyone concerned about data validity. You can easily examine your data from several angles using six diagnostic reports to uncover missing data patterns. Then, estimate summary statistics and impute missing values through statistical algorithms.

Quickly and easily diagnose your missing data
You can quickly diagnose a serious missing data problem using the data patterns report, which provides a case-by-case overview of your data. This report helps you determine the extent of missing data; it displays a snapshot of each type of missing value and any extreme values for each case.

Use the flexible separate variance t test and crosstabulation of categorical variables tables to discover if significant differences exist between respondents and non-respondents. These reports help you decide if missing data might cause problems in your analysis.

Receive a summary of missing data patterns and highlights of the variable sets that comprise the patterns with the tabulated pattern reports. For example, in a consumer survey, quickly notice that 98 out of 100 people didn’t respond to the question set on preferred features and price.

Improve survey questions that you’ve identified as possibly confusing based on observed missing data patterns. You can even determine if missing values for one variable are related to missing values of another with the percent mismatch of patterns table. For example, respondents who skip a question on income might also bypass a question about education level. Use this information to enhance the quality of your surveys in the future.

Receive better summary statistics
Since summary statistics are often the starting point for other analyses, SPSS Missing Value Analysis allows you to adjust for missing data when working with them. Choose from four methods: Listwise deletion, pairwise deletion, expectation maximization (EM), and covariance matrix.

Reach more statistically significant results
Replace missing values with estimates and increase the chance of receiving statistically significant results. Remove hidden bias from your data by replacing missing values with estimates to include all groups in your analysis—even those with poor responsiveness. Use the powerful EM or regression algorithm to predict missing values based on data you already have.
**Features**

**Analyze patterns**
- Display missing data and extreme cases for all cases and all variables using the data patterns table
  - Display system-missing and three types of user-defined missing values
  - Sort in ascending or descending order
  - Display actual values for specified variables
- Display patterns of missing values for all cases that have at least one missing value using the missing patterns table
  - Group similar missing value patterns together
  - Sort by missing patterns and variables
  - Display actual values for specified variables
- Determine differences between missing and non-missing groups for a related variable with the separate variance t test table
  - t test, degrees of freedom, mean, p value, and count
- Show differences between present and missing data for categorical variables using the distribution of categorical variables table
  - Produce crosstabs showing product and missing data for each category of one variable by the other variables
- Assess how much missing data for one variable relates to the missing data of another variable using the percent mismatch of patterns table
  - Sort matrices by missing value patterns or variables
- Identify all unique patterns with the tabulated patterns table, which summarizes each missing data pattern and displays the count for each pattern plus means and frequencies for each variable
  - Display count and averages for each missing value pattern using the summary of missing value patterns table

**Statistics**
- Univariate: Compute count, mean, standard deviation, and standard error of mean for all cases, excluding those containing missing values, counts, percent of missing values, and extreme values for all variables
- Listwise: Compute mean, covariance matrix, and correlation matrix for all quantitative variables for cases excluding missing values
- Pairwise: Compute frequency, mean, variance, covariance matrix, and correlation matrix
- EM algorithm
  - Estimate the means, covariance matrix, and correlation matrix of quantitative variables with missing values, assuming normal distribution, t distribution with degrees of freedom, or a mixed-normal distribution with any mixture proportion and any standard deviation ratio
  - Impute missing data and save the completed data as a file

**Regression algorithm**
- Estimate the means, covariance matrix, and correlation matrix of variables set as dependent; set the number of predictor variables; and set random elements as normal, t, residuals, or none
- Impute missing data and save the completed data as a file

**Data management**
- Handle all character variables as categorical variables
- Use the first eight characters of a string variable when it is defined as categorical
- Save the completed data matrix as an external file

**System requirements**
- Software: SPSS Base 16.0
- Other system requirements vary according to platform

---

Features subject to change based on final product release.

Missing data (top left) can leave you with invalid or erroneous results. Mean substitution (top right) and a fairly simple regression (bottom left) show that these methods provide an inaccurate or insignificant way to impute missing values. SPSS Missing Value Analysis (bottom right) provides the best method for imputing missing values. As shown here, it provides a scatterplot of YMISS and Y with imputed missing values.
New Tools for Building Predictive Models

Your organization needs to find patterns and connections in the complex and fast-changing environment you work in so that you can make better decisions at every turn. You may be using SPSS and one or more of its add-on modules to help you do this. If so, you know the power and versatility you have at your fingertips. But there’s even more you can do.

You can explore subtle or hidden patterns in your data, using SPSS Neural Networks. This new add-on module offers you the ability to discover more complex relationships in your data and generate better performing predictive models. The result? Deeper insight and better decision-making.

The procedures in SPSS Neural Networks complement the more traditional statistics in SPSS Base and its modules. Find new associations in your data with SPSS Neural Networks and then confirm their significance with traditional statistical techniques.

Why use a neural network?

A computational neural network is a set of non-linear data modeling tools consisting of input and output layers plus one or two hidden layers. The connections between neurons in each layer have associated weights, which are iteratively adjusted by the training algorithm to minimize error and provide accurate predictions. You set the conditions under which the network “learns” and can finely control the training stopping rules and network architecture, or let the procedure automatically choose the architecture for you.

You can combine SPSS Neural Networks with other statistical procedures to gain clearer insight in a number of areas. In market research, for example, you can create customer profiles and discover customer preferences. In database marketing, you can segment your customer base and optimize marketing campaigns.

In financial analysis, you can use SPSS Neural Networks to analyze applicants’ creditworthiness and to detect possible fraud. In operational analysis, use this new tool to manage cash flow and improve logistics planning. Scientific and healthcare applications include forecasting treatment costs, performing medical outcomes analysis, and predicting the length of a hospital stay.

Control the process from start to finish

With SPSS Neural Networks, you select either the Multilayer Perceptron (MLP) or Radial Basis Function (RBF) procedure.

Both of these are supervised learning techniques—that is, they map relationships implied by the data. Both use feed-forward architectures, meaning that data moves in only one direction, from the input nodes through the hidden layer of nodes to the output nodes. Your choice of procedure will be influenced by the type of data you have and the level of complexity you seek to uncover. While the MLP procedure can find more complex relationships, the RBF procedure is generally faster.
With either of these approaches, you divide your data into training, testing, and holdout sets. The training set is used to estimate the network parameters. The testing set is used to prevent overtraining. The holdout set is used to independently assess the final network, which is applied to the entire dataset and to any new data.

You specify the dependent variables, which may be scale, categorical, or a combination of the two. If a dependent variable has scale measurement level, then the neural network predicts continuous values that approximate the “true” value of some continuous function of the input data. If a dependent variable is categorical, then the neural network is used to classify cases into the “best” category based on the input predictors.

You adjust the procedure by choosing how to partition the dataset, what sort of architecture you want, and what computation resources will be applied to the analysis. Finally, you choose to display results in tables or graphs, save optional temporary variables to the active dataset, and export models in XML-file formats to score future data.

In an MLP network like the one shown here, the data feeds forward from the input layer through one or more hidden layers to the output layer.

The results of exploring data with neural network techniques can be shown in a variety of graphic formats. This simple bar chart is one of many options.

From the Multilayer Perceptron (MLP) dialog, you select the variables that you want to include in your model.
Features
Multilayer Perceptron (MLP)
The MLP procedure fits a particular kind of neural network called a multilayer perceptron. The multilayer perceptron is a supervised method using feedforward architecture. It can have multiple hidden layers. One or more dependent variables may be specified, which may be scale, categorical, or a combination. If a dependent variable has scale measurement level, then the neural network predicts continuous values that approximate the “true” value of some continuous function of the input data. If a dependent variable is categorical, then the neural network is used to classify cases into the “best” category based on the input predictors.

- Predictors
  - Factors
  - Covariates
- The EXCEPT subcommand lists any variables that the MLP procedure should exclude from the factor or covariate lists on the command line. This subcommand is useful if the factor or covariate lists contain a large number of variables.
- The RESCALE subcommand is used to rescale covariates or scale dependent variables
  - Dependent variable (if scale): standardized, normalized, adjusted normalized, or none
  - Covariates: standardized, normalized, adjusted normalized, or none
- The PARTITION subcommand specifies the method of partitioning the active dataset into training, testing, and holdout samples. The training sample comprises the data records used to train the neural network. The testing sample is an independent set of data records used to track prediction error during training in order to prevent overtraining. The holdout sample is another independent set of data records used to assess the final neural network. You can specify:
  - The relative number of cases in the active dataset to randomly assign to the training sample
  - The relative number of cases in the active dataset to randomly assign to the testing sample
  - The relative number of cases in the active dataset to randomly assign to the holdout sample
  - A variable that assigns each case in the active dataset to the training, testing, or holdout sample
- The ARCHITECTURE subcommand is used to specify the neural network architecture. You can specify:
  - Whether to use the automatic architecture or, if automatic is not used:
    - The number of hidden layers in the neural network
    - The activation function to use for all units in the hidden layers (Hyperbolic tangent or Sigmoid)
    - The activation function to use for all units in the output layer (Identity, Hyperbolic tangent, Sigmoid, or Softmax)
- The CRITERIA subcommand specifies the computational and resource settings for the MLP procedure. You can specify the training type, which determines how the neural network processes training data records: batch training, online training, mini-batch training. You can also specify:
  - The number of training records per mini-batch (if selected as the training method)
  - The maximum number of cases to store in memory when automatic architecture selection and/or mini-batch training is in effect
  - The optimization algorithm used to determine the synaptic weights: Gradient descent, Scaled conjugate gradient
  - The initial learning rate for the gradient descent optimization algorithm
  - The lower boundary for the learning rate when gradient descent is used with online or mini-batch training
  - The momentum rate for the gradient descent optimization algorithm
  - The initial lambda, for the scaled conjugate gradient optimization algorithm
  - The initial sigma, for the scaled conjugate gradient optimization algorithm
  - The interval \([a_0-a, a_0+a]\) in which weight vectors are randomly generated when simulated annealing is used
The STOPPINGRULES subcommand specifies the rules that determine when to stop training the neural network. You can specify:
- The number of steps \( n \) to allow before checking for a decrease in prediction error
- Whether the training timer is turned on or off and the maximum training time
- The maximum number of epochs allowed
- The relative change in training error criterion
- The training error ratio criterion

The MISSING subcommand is used to control whether user-missing values for categorical variables—that is, factors and categorical dependent variables—are treated as valid values.

The PRINT subcommand indicates the tabular output to display and can be used to request a sensitivity analysis. You can choose to display:
- The case processing summary table
- Information about the neural network, including the dependent variables, number of input and output units, number of hidden layers and units, and activation functions
- A summary of the neural network results, including the average overall error, the stopping rule used to stop training and the training time
- A classification table for each categorical dependent variable
- The synaptic weights; that is, the coefficient estimates, from layer \( i-1 \) unit \( j \) to layer \( i \) unit \( k \)
- A sensitivity analysis, which computes the importance of each predictor in determining the neural network

The PLOT subcommand indicates the chart output to display. You can display:
- Network diagram
- A predicted by observed value chart for each dependent variable
- A residual by predicted value chart for each scale dependent variable
- ROC (Receiver Operating Characteristic) curves for each categorical dependent variable. It also displays a table giving the area under each curve.
- Cumulative gains charts for each categorical dependent variable
- Lift charts for each categorical dependent variable

The SAVE subcommand writes optional temporary variables to the active dataset.

The OUTFILE subcommand saves XML-format files containing the synaptic weights.

The RBF procedure trains the network in two stages:
1. The procedure determines the radial basis functions using clustering methods. The center and width of each radial basis function are determined.
2. The procedure estimates the synaptic weights given the radial basis functions. The sum-of-squares error function with identity activation function for the output layer is used for both prediction and classification. Ordinary Least Squares regression is used to minimize the sum-of-squares error.

Due to this two-stage training approach, the RBF network is in general trained much faster than MLP.

Subcommands listed for the MLP procedure perform similar functions for the RBF procedure, with the following exceptions:
- When using the ARCHITECTURE subcommand, users can specify the Gaussian radial basis function used in the hidden layer: either Normalized RBF or Ordinary RBF
- When using the CRITERIA subcommand, users can specify the computation settings for the RBF procedures, specifying how much overlap occurs among the hidden units.

Features subject to change based on final product release.
Improve Predictions with Regression Software

Use SPSS Regression Models’ wide range of nonlinear modeling procedures to apply more sophisticated models to your data. For example, you can use SPSS Regression Models for:
- Market research—Study consumer buying habits
- Medical research—Study response to dosages
- Loan assessment—Analyze good and bad credit risks
- Institutional research—Measure academic achievement tests
- And much more

Predict categorical outcomes with more than two categories
With multinomial logistic regression (MLR), you are free from constraints such as yes/no answers. For example, you can model which factors predict if the customer buys product A, product B, or product C.

Easily classify your data into two groups
Use binary logistic regression to predict dichotomous variables such as buy or not buy and vote or not vote. This procedure offers stepwise methods to select the main and interaction effects that best predict your response variable.

Control your model
Have more control over your model and your model expression by using constrained and unconstrained nonlinear regression procedures. These procedures provide two methods for estimating parameters of nonlinear models. The Levenberg-Marquardt algorithm analyzes unconstrained models. The sequential quadratic programming algorithm enables you to specify constraints on parameter estimates, provide your own loss function, and get bootstrap estimates of standard errors.

Use alternative procedures to meet assumptions
When your data do not meet the statistical assumptions for ordinary least squares, use weighted least squares (WLS) or two-stage least squares (2SLS). Give more weight to measurements within a series by using WLS. 2SLS helps control for correlations between predictor variables and error terms that often occur with time-based data.

Find the best stimuli
Perform probit and logit response modeling to analyze the potency of responses to stimuli, such as medicine doses, prices, or incentives. Probit evaluates the value of the stimuli using a logit or probit transformation of the proportion responding.
Features

Multinomial logistic regression (MLR)
Regresses a categorical dependent variable with more than two categories on a set of independent variables
- Control the values of the algorithm-tuning parameters using the CRITERIA subcommand
- Include interaction terms
- Customize hypotheses by directly specifying null hypotheses as linear combinations of parameters, using the TEST subcommand
- Specify the dispersion scaling value by using the SCALE subcommand
- Build equations with or without a constant
- Use a confidence interval for odds ratios
- Save the following statistics: Predicted probability, predicted response category, probability of the predicted response category, and probability of the actual response category
- Specify the reference category in the dependent variables
- Handle very large problems
- Find the best predictor from dozens of possible predictors using stepwise functionality
- Find predictors using forward entry, backward elimination, forward stepwise, or backward stepwise
- Opt to select a rule for effect entry or removal from the analysis
- Base entry or removal on satisfying the hierarchy requirement for all effects, for factor-only effects, or for satisfying the containment requirement for all effects
- Optionally, perform entry or removal without satisfying the hierarchy or containment requirement for any effects in the model
- Use Score and Wald methods, which help you more quickly reach results if you have a large number of predictors
- Assess model fit using Akaike information criterion (AIC) and Bayesian information criterion (BIC; also called Schwarz Bayesian Criterion, or SBC)
- Choose from the following diagnostics for the classification table:
  - Percent concordance
  - Percent ties
  - Percent discordance
  - C-value for logistic mode
  - Somer's D
  - Gamma
  - Tau-a statistics

Binary logistic regression
Regresses a dichotomous dependent variable on a set of independent variables
- Use forward/backward stepwise and forced entry modeling
- Transform categorical variables by using deviation contrasts, simple comparison, difference (reverse Helmert) contrasts, Helmert contrasts, polynomial contrasts, comparison of adjacent categories, user-defined contrasts, or indicator variables
- Select criteria for model building:
  - Probability of score statistic for entry, probability of Wald, or likelihood ratio statistic for removal
  - Save the following statistics: Predicted probability and group, residuals, deviance residuals, leverage value, analog of Cook's influence statistic, and difference in Beta
- Export the model using XML

Constrained nonlinear regression (CNLR)
Uses linear and nonlinear constraints on any combination of parameters
- Save predicted values, residuals, and derivatives
- Choose numerical or user-specified derivatives

Nonlinear regression (NLR)
Estimates models with arbitrary relationships between independent and dependent variables using iterative estimation algorithms
- Specify loss function options
- Use bootstrap estimates of standard errors

Weighted least squares (WLS)
Gives more weight to measurements within a series
- Calculate weights based on source variable and Delta values or apply from an existing series
- Select output for calculated weights: Log-likelihood functions for each value of Delta; R, R^2, adjusted R^2, standard errors, analysis of variance, and t tests of individual coefficient for Delta value with maximized log-likelihood function
- Display output in pivot tables

Two-stage least squares (2SLS)
Helps control for correlations between predictor variables and error terms
- Use structural equations and instrumental variables
- Set control for correlations between predictor variables and error terms
- Display output in pivot tables

Probit
Evaluates the value of stimuli using a logit or probit transformation of the proportion responding
- Transform predictors: Base 10, natural, or user-specified base (including none)
- Allow for natural response rate estimates or specify them yourself
- Use algorithm control parameters: Convergence, iteration limit, and heterogeneity criterion probability
- Select from the following statistics:
  - Frequencies, fiducial confidence intervals, relative median potency, test of parallelism, plots of observed probits, or logits
- Display output in pivot tables

System requirements
- Software: SPSS Base 16.0
- Other system requirements vary according to platform

Features subject to change based on final product release. □ Symbol indicates a new feature.

To learn more, please visit www.spss.com. For SPSS office locations and telephone numbers, go to www.spss.com/worldwide.

SPSS is a registered trademark and the other SPSS products named are trademarks of SPSS Inc. All other names are trademarks of their respective owners. © 2007 SPSS Inc. All rights reserved. SIMS65FCA6-0707
Create Custom Tables in No Time

Easily analyze and communicate your results with SPSS Tables, an add-on module for the SPSS product line.
Share analytical results with decision makers

You often report the results of your analyses to decision makers, colleagues, clients, grant committees, or others. Building tabular reports, however, can be a time-consuming, trial-and-error process. SPSS Tables 16.0 enables you to view your tables as you build them, so you can create polished, accurate tables in less time.

SPSS Tables, an add-on module for the SPSS product line, makes it easy for you to summarize your data in different styles for different audiences. The module’s build-as-you-go interface updates in real time, so you always know what your tables will look like. You can add descriptive and inferential test statistics, for example, and then customize the table so your audience can easily understand the information. When your tables are complete, you can export them to Microsoft® Word, Excel®, and PowerPoint®, and HTML.

SPSS Tables is ideal for anyone who creates and updates reports on a regular basis, including people who work in survey or market research, the social sciences, database or direct marketing, and institutional research.

SPSS Tables 16.0 is optimized for use with SPSS 16.0. It includes such frequently requested features as:

- An interactive table builder that enables you to preview your tables as you create them
- Category management capabilities that enable you to exclude specific categories, display missing value cells, and add subtotals to your table
- Three significance tests: Chi-square test of independence, comparison of column means (t test), or comparison of column proportions ($z$ test)
- Easily export tables to Word or Excel for use in reports

Preview tables as you build them

SPSS Tables’ intuitive graphical user interface takes the guesswork out of building tables. The drag-and-drop capabilities and preview pane enable you to see what your tables will look like before you click “OK.”

You can interact with the variables on your screen, identify variables as categorical or scale, and know immediately how your data are structured.

To create a table, just drag your desired variables into the table preview builder. You don’t have to write complicated syntax or work with dialog boxes. And you can move variables easily from row to column for precise positioning. The table preview builder updates after every change you make, so you can see the formatting effect immediately. You can also add, swap, and nest variables, or hide statistic labels, directly from within the table preview builder. And you can collapse large, complex tables for a more concise view, and still see your variables.

Customize your tables

Display information the way you want to with the category management features in SPSS Tables. Create totals and subtotals without changing your data file. You can combine several categories into a single category, for example, for frequent top- and bottom-box analyses. You can also sort categories within your table without affecting the subtotal calculation.
Make your tables more precise, as you create them, by changing variable types or excluding categories. You can display or exclude categories with no counts for clearer and more concise output. Or sort and rank categories based on cell values for a neater, more informative table.

**Get in-depth analyses**

You can use SPSS Tables as an analytical tool to understand your data better and create tables that present your results most effectively. Give your readers reports that enable them to delve into the information and make more informed decisions.

Highlight opportunities or problem areas in your results when you include inferential statistics. Using inferential test statistics with SPSS Tables enables you to compare means or proportions for demographic groups, customer segments, time periods, or other categorical variables. You can also identify trends, changes, or major differences in your data.

A market researcher at a major publishing company, for example, studies student ratings of college textbooks. He notices a potential relationship between students at private universities and low ratings for math textbooks. The researcher runs a column proportions test with SPSS Tables. The test shows, at a 95 percent confidence level, that there is a difference in math textbook ratings between students at private and public universities. Knowing that the confidence level for this difference is high, and that it’s unlikely that the relationship is due to chance, the researcher recommends that the publishing company explore the reasons for the difference in ratings.

You can also select summary statistics, which include everything from simple counts for categorical variables to measures of dispersion. Summary statistics for categorical variables and multiple response sets include counts and a wide variety of percentage calculations, including row, column, subtable, table, and valid N percentages. Summary statistics for scale variables and custom total summaries for categorical variables include mean, median, percentiles, sum, standard deviation, range, and minimum and maximum values. To focus on specific results, you can sort categories by any summary statistic you used.

**“It is a real timesaver to have an analytical tool like SPSS Tables that can not only quickly produce tables, but produce them in a report-quality format that I need. I can simply insert the tables into a client report without the need for additional formatting.”**

— Brian Robertson, PhD
Director of Research
Market Decisions
Control your table output

Many features in SPSS Tables help you create tables with the look you want and the time-saving capabilities you need:

- Add titles and captions
- Use table expressions in titles
- Use SPSS Base features such as TableLooks™ and scripts to automate formatting and redundant tasks
- Specify minimum and maximum column widths for individual tables during table creation

Share results more easily with others

Once you have results, you need to put them in the hands of those who need them. SPSS Tables enables you to create results as interactive pivot tables, for export to Word or Excel. This not only improves your workflow, it saves time because you don’t have to reconfigure your tables in Word or Excel. No editing is required after you export your tables. You can, however, insert descriptive content if you choose to.

Save time and effort by automating frequent reports

Do you regularly create reports that have the same structure? Do you spend a lot of time updating reports that you built in the past? Use syntax and automation in SPSS Tables to run frequently needed reports, such as the compliance reports required for grant funding, in production mode.

When you create a table, SPSS Tables records every click you make and saves your actions as syntax. To run an automated report, you simply paste the relevant syntax into a syntax window, then just click and go. With syntax and automation, your report is ready without tedious and time-consuming production.

Syntax in SPSS Tables 16.0 uses a more natural language than in earlier versions, so it’s easier to understand. Syntax created in earlier versions is still usable, however. To take advantage of features such as inferential statistics in SPSS Tables 16.0, simply use the included syntax converter to translate the original syntax to new command syntax.

Create multiple types of output

SPSS Tables can produce a wide variety of customized tables. Here are examples of three common table types you may want to use when analyzing and describing your data.

Two-dimensional crosstabulation: This example shows the relationship between two categorical variables, Age and Gender. Using Age as the row variable and Gender as the column variable, you can create a two-dimensional crosstab that shows the number of males and females in each age category.

Multiple response set: Multiple response sets use multiple variables to record responses to questions for which the respondent can give more than one answer. When asked the question, “Which of the following sources do you rely on for news,” respondents could select any combination of five possible choices. Notice that the percentages total more than 100 percent because each respondent may choose more than one answer. You can also perform significance tests on multiple response variables.

Shared response categories (comperimeter tables): Surveys often contain many questions that have a common set of possible responses. For example, the questions in this survey concern confidence in public and private institutions and services, and all have the same set of response categories: 1 = A great deal, 2 = Only some, and 3 = Hardly any. Use stacking to display these related variables in the same table—and display the shared response categories in the columns of the table.
Create high-quality tables from SPSS data

With SPSS Tables 16.0’s interactive table builder, creating professional-quality tables is easy to do. This diagram shows you how.

**Steps**

1. Drag your desired variables to the table builder. As shown in this screen, you can preview the category list (lower left corner) before dragging the desired categories to the table.

2. Define the summary statistics or categories and totals. You can choose from more than 40 summary statistics.

3. All results are produced as SPSS pivot tables. You can apply TableLooks to your output for a more polished appearance. In addition, you can export output to Word, Excel, PowerPoint, or HTML.
Features

Graphical user interface

- Simple, drag-and-drop table builder interface enables you to preview tables as you select variables and options
- Single, unified table builder, instead of multiple menu choices and dialog boxes for different table types, makes building tables easier

Control contents

- Create tables with up to three display dimensions: Rows (stub), columns (banner), and layers
- Nest variables to any level in all dimensions
- Crosstabulate multiple independent variables in the same table
- Display frequencies for multiple variables side by side with tables of frequencies
- Display all categories when multiple variables are included in a table, even if a variable has a category without responses
- Display multiple statistics in rows, columns, or layers
- Place totals in any row, column, or layer
- Create subtotals for subsets of categories of a categorical variable
- Customize your control over category display order and selectively show or hide categories
- Better control how you display your data using expanded category options:
  - Sort categories by any summary statistic in your table
  - Hide the categories that comprise subtotals—you can remove a category from the table without removing it from the subtotal calculation

Test statistics

- Select from these summary statistics:
  - Count, count row %, count column %, count table %, count subtable %, layer %, count table row %, count table column %, valid N row %, valid N column %, valid N table %, valid N subtable %, valid N layer %, valid N table row %, valid N table column %, total N row %, total N column %, total N table %, total N subtable %, total N layer %, total N table row %, total N table column %, maximum, mean, median, minimum, missing, mode, percentile, percentile 05, percentile 25, percentile 75, percentile 95, percentile 99, range, standard error (SE) mean, standard deviation (SD), sum, total N, valid N, variance, sum row %, sum column %, sum table %, sum subtable %, sum layer %, sum table row %, and sum table column %
- Calculate statistics for each cell, subgroup, or table
- Calculate percentages at any or all levels for nested variables
- Calculate counts and percentages for multiple response variables based on the number of responses or the number of cases
- Select percentage bases for missing values to include or exclude missing responses
- Exclude subtotal categories from significance tests
- Run significance tests on multiple response variables

Formatting controls

- Directly edit any table element, including formatting and labels
- Sort tables by cell contents in ascending or descending order
- Automatically display labels instead of coded values
- Specify minimum and maximum width of table columns (overrides TableLooks)
- Show a name, label, or both for each table variable
- Display missing data as blank, zero, ’.‘ or any other user-defined term, such as “missing”
- Use the global break command to produce a table for each value of a variable when the variable is used in a series of tables
- Set titles for pages and tables to be multiple lines with left, right, or center justification
- Add captions for pages or tables
- Specify corner labels
- Customize labels for statistics
- Display the entire label for variables, values, and statistics
- Format numerical results: Commas, date/time, dollars, F (standard numeric), negative parentheses, “N=,” parentheses (around numbers of percentages), percentages, and customized formats
- Apply preformatted TableLooks to results
- Define the set of variables that is related to multiple response data and save it with your data definition for subsequent analysis
- Accepts both long- and short-string elementary variables
- Imposes no limit on the number of sets that can be defined or the number of variables that can exist in a set
- All results are produced as SPSS pivot tables so you can explore your results more easily with the pivot feature
- Rearrange columns, rows, and layers by dragging icons for easier ad hoc analysis
- Toggle between layers by clicking on an icon for easier comparison between subgroups
- Reset a table to its original organization with a simple menu choice
- Rotate even the outermost nests in the stub, banner, and layer to uncover information that can be easily hidden in large reports

Syntax

- Syntax converter translates syntax created in versions earlier than SPSS Tables 11.5 into CTABLES syntax

Printing formats

- Print more than one table per page
- Specify page layout: Top, bottom, left, and right margins; page length
- Use the global break command to produce a table for each value of a variable when the variable is used in a series of tables

System requirements

- Software: SPSS Base 16.0
- Other system requirements vary according to platform

Features subject to change based on final product release.
Build Expert Forecasts—in a Flash

Forecasts provide a solid foundation for your organization’s planning. SPSS Trends offers a number of capabilities that enable both novice and experienced users to quickly develop reliable forecasts using time-series data. SPSS Trends is a fully integrated module of SPSS, so you have all of SPSS’ capabilities at your disposal, plus features specifically designed to support forecasting.

Because they help you develop and manage plans affecting a number of operational areas, forecasts have a major impact on profits. They enable your organization to better anticipate revenues; control staffing, inventory, and other costs; and manage other business processes more precisely—all improvements that lead to a healthier bottom line. However, working with the time-series data needed to develop forecasts can be challenging.

SPSS Trends has the advanced techniques you need without the drawbacks of traditional forecasting methods. Unlike spreadsheet programs, SPSS Trends enables you to use advanced statistical methods in creating forecasts. But you don’t need expert statistical knowledge to do so.

People new to forecasting can create sophisticated forecasts that take into account multiple variables. And experienced forecasters can use SPSS Trends to validate their models. You get the information you need faster because SPSS Trends makes building forecasts easy.

Efficiently generate and update models

Instead of laboriously building forecasts by re-setting parameters and re-estimating models, variable by variable, you can speed through the process with SPSS Trends. You’ll save hours, even days, of valuable time, with no compromise in the quality or reliability of your forecasts.

With SPSS Trends, you can:
- Develop reliable forecasts quickly, no matter how large the dataset or how many variables are involved
- Reduce forecasting error by automating the selection of the appropriate models and their parameters
- Update and manage forecasting models efficiently, so you can devote more time to exploring why some models diverge from the norm
- Allow a broader group of people in your organization to generate forecasts
- Give experienced forecasters control over choices affecting models, parameters, and output
- Deliver understandable and useful information to your organization’s decision makers
You have tremendous flexibility in creating forecasts. For example, you can easily convert transactional data into time-series data, and convert existing time-series data to the time intervals that best suit your organization’s planning needs, with SPSS.

You can create forecasts at exactly the level of detail you need—for example, for each product line, for individual products, and across geographic and functional areas.

Then, using SPSS Trends’ Expert Modeler, you can:
- Automatically determine the best-fitting ARIMA or exponential smoothing model for your time-series data
- Model hundreds of different time series at once, rather than having to run the procedure for one variable at a time

You can also:
- Save models to an XML file so that when data changes, forecasts can be updated without having to re-set parameters or re-estimate the model
- Write scripts so that updates can be performed automatically

**Guides novice forecasters**
If you’re new to modeling time-series data, or create time-series models only occasionally, you benefit from SPSS Trends’ ability to choose the appropriate model for your data and guide you through the model-building process. Using SPSS Trends, you can:
- Generate reliable models, even if you’re not sure how to choose exponential smoothing parameters or ARIMA orders, or how to achieve stationarity
- Automatically test your data for seasonality, intermittency, and missing values, and select appropriate models
- Detect outliers and prevent them from influencing parameter estimates
- Generate graphs of your data, showing confidence intervals and the model’s goodness of fit

After your models are created and validated, you can share them with others by incorporating them into Microsoft® Office applications. Or, by using the SPSS Output Management System (OMS), write output in HTML or XML formats for posting on corporate intranets. You can also save models as SPSS data files. This enables you to continue exploring the files for characteristics such as each model’s goodness of fit.

**Provides control for experts**
If you are an experienced forecaster, you benefit from SPSS Trends because you can create models from time-series data more efficiently while still controlling key aspects of the process.

For example, you can limit the choice of models to ARIMA models only, or to exponential smoothing models only, through SPSS Trends’ Expert Modeler. You can opt out of the Expert Modeler and choose every parameter of the model yourself. Alternatively, use the Expert Modeler recommendations as a starting point for your selections, or to check your work.

You can limit output so that you see only the worst-fitting models—those that require further examination. This enables you to uncover problems with your data or models quickly and efficiently.
Greg is an inventory manager for a major retailer. He has responsibility for more than 5,000 products and uses SPSS Trends to develop inventory forecasts three months out for each product. Because SPSS Trends automates the modeling of thousands of variables, the process of generating initial forecasts takes only a few hours, instead of several days. And the process of updating models also can be done efficiently.

His company’s data warehouse is refreshed monthly with actual sales data, so Greg runs forecasts as a batch job once a month. By doing so, he incorporates the new data and extends his forecast horizon one more month into the future. He is able to do this without re-estimating his models, which speeds the process.

To check model performance, Greg includes SPSS command syntax in the batch job to identify any series having time points with observed sales outside the confidence intervals established by the original model. For these series, he runs another batch job to identify a new model that better fits the revised data.

By using SPSS Trends, Greg conducts sophisticated forecasting efficiently and improves his company’s ability to plan effectively.
Features

**TSMODEL**

Model a set of time-series variables by using the Expert Modeler or by specifying the structure of autoregressive integrated moving average (ARIMA) or exponential smoothing models.

- Allow Expert Modeler to select the best-fitting predictor variables and models
  - Limit search space to ARIMA models only, or to exponential smoothing models only
  - Treat independent variables as events
- Specify custom ARIMA models, which produce maximum likelihood estimates for seasonal and non-seasonal univariate models
  - Work with general or constrained models specified by autoregressive or moving average order, order of differencing, seasonal autoregressive or moving average order, and seasonal differencing
  - Use two dependent variable transformations: square root and natural log
  - Automatically detect or specify outliers: additive, level shift, innovation, transient, seasonal additive, local trend, and additive patch
- Specify seasonal and non-seasonal numerator, denominator, and difference transfer function orders and transformations for each independent variable
- Specify custom exponential smoothing models
  - Four non-seasonal model types: simple, Holt’s linear trend, Brown’s linear trend, and damped trend
  - Three seasonal model types: simple seasonal, Winters’ additive, and Winters’ multiplicative
  - Two dependent variable transformations: Square root and natural log
- Display forecasts, fit measures, Ljung-Box statistic, parameter estimates, and outliers by model
- Generate tables and plots to compare statistics across all models
- Choose from eight available goodness-of-fit measures: R², stationary R², root mean square error, mean absolute percentage error, mean absolute error, maximum absolute percentage error, maximum absolute error, and normalized Bayes information criterion (BIC)
- Create tables and plots of residual autocorrelation function (ACF) and partial autocorrelation function (PACF)
- Plot observed values, forecasts, fit values, and confidence intervals for forecasts, and fit values for each series
- Filter output to a fixed number or percentage of best- or worst-fitting models
- Specify forecast period, treatment of user-missing values, and confidence intervals
- Export models to an XML file for later use by TSAPPLY

**SEASON**

Estimate multiplicative or additive seasonal factors for periodic time series.

- Choose either a multiplicative or an additive model
- Calculate moving averages, ratios, seasonal and seasonal adjustment factors, seasonally adjusted series, smoothed trend-cycle components, and irregular components

**SPECTRA**

Decompose a time series into its harmonic components, a set of regular periodic functions at different wavelengths or periods.

- Produce/plot univariate or bivariate periodogram and spectral density estimates
- Produce/plot bivariate spectral analyses
- Smooth periodogram values with weighted moving averages
- Smooth, using available spectral data windows: Tukey-Hamming, Tukey, Parzen, Bartlett, equal weight, no smoothing, and user-specified weights
- Produce high-resolution charts: Periodogram, spectral and cospectral density estimate, squared coherency, quadrature spectrum estimate, phase spectrum, cross amplitude, and gain

**TSAPPLY**

Apply saved models to new or updated data.

- Simultaneously apply models from multiple XML files created with TSMODEL
- Re-estimate model parameters and goodness-of-fit measures from the data, or load them from the saved model file
- Selectively choose saved models to apply
- Override the periodicity (seasonality) of the active dataset
- Choose from the same output, fit measure, statistics, and options as TSMODEL
- Export re-estimated models to an XML file

**System requirements**

- Software: SPSS Base 16.0
- Other system requirements vary according to platform

Features subject to change based on final product release.
SPSS is a suite of products for statistical analysis and data management. It enables analysts and researchers to access, prepare, manage, analyze, and report on data in order to solve business and research problems in the context of the analytical process. SPSS Server features client/server architecture, which is the combination of two powerful products:

- SPSS for Windows®, SPSS Inc.’s flagship software for in-depth data exploration, analytical reporting, and modeling
- SPSS Server, server-based software that delivers enterprise-strength scalability, additional tools, security, and enhanced performance

When you combine the strength of world-class analytical tools and techniques with the flexibility and speed of server functionality, you have a powerful solution for supporting better decision making throughout your enterprise.

**Increase productivity**

SPSS Server uses server-grade resources to perform data preparation and analysis faster than desktop resources can. It also removes the need to keep a copy of the data on the desktop, freeing local resources and network bandwidth, and allowing you to perform additional tasks while using these resources. Administrative controls in SPSS Server enable you to increase performance for high-priority users and also allocate server resources where they are needed most—rather than on a first-come, first-served basis.

Additional features unique to the server version of SPSS increase the tools available for preparing data and creating reports. These features enable you to:

- Sort and aggregate data inside the database prior to its retrieval for analysis
- Open multiple XML models created by SPSS, Clementine®, or AnswerTree® and score new data
- Filter large amounts of irrelevant data to obtain only features relevant for modeling with the Predictor Selection algorithm
- Predict classification of cases by treating each variable as independent and equal by using the Naïve Bayes algorithm
- Reduce network traffic and improve performance with the data-free client feature. Your administrator can confine user-viewing rights to the data dictionary only, while allowing users to conduct all data preparation and analysis from the Data Editor. Users can view, manipulate, and deploy output through the Output Viewer.
- Process large data preparation or analysis tasks when using the SPSS Batch Facility (SPSSB)

SPSSB (included with SPSS Server) also enables you to process repeated analytical jobs, especially when server use is low (such as overnight). The SPSS Batch Facility runs without requiring an active or connected client machine and can deliver output in text, HTML, and XML formats. It can also save data preparation work to the SPSS (SAV) file format.
Analyze massive data files faster
SPSS provides a scalable client/server version for your enterprise-wide analytical solution. By analyzing data on a server rather than your desktop, you can analyze massive datasets quickly. You can handle large data analysis problems easily because SPSS Server practically eliminates data size limitations. You also receive better performance because less temporary file space is needed when using procedures that change or add more data (such as recoding variables, saving regression residuals, or saving cluster memberships). SQL pushback enables SPSS Server to sort and aggregate data in the database prior to its retrieval. With this capability, you can take advantage of more scalable data preparation abilities at the DBMS level.

To significantly speed up data preparation tasks, seamlessly connect to third-party multithreaded sorting applications. You can also gain greater speed in reading and writing large temporary files when you stripe these files over multiple disks.

Streamline and speed data access
With SPSS Server, data remain on the server and don’t need to be copied to the client machine or converted into an SPSS file. You dramatically reduce data handling processing times with conversion-free, copy-free data access from SQL databases when using the SPSS Data Access Pack (included with SPSS Server). Because data remain on the server, you increase the speed of your analysis and reduce network traffic. And, because you don’t have to convert data into SPSS format, you can save even more time.

Make better use of existing hardware
SPSS Server is a flexible product that enables you to create a solution customized to your organization’s unique environment. You can run SPSS Server on a wide variety of popular server platforms, which communicate seamlessly with the desktop version of SPSS* and enable your IT staff to use resources they may already have. By providing a native 64-bit analytical solution that directly contributes to increased productivity and improved decision making, 64-bit versions of SPSS Server will enable your organization to more easily realize a positive return from its 64-bit hardware investment.

Access your data remotely—even from outside a firewall
Do you or others need to access data outside of your company’s firewall? SPSS Server uses various port-forwarding schemes so that remote analysts with permissions can securely tunnel through firewalls and access datasets. And because data are processed on the server, remote employees do not have to download entire datasets for processing—an important consideration for users with limited bandwidth. Network administrators can create a connection that uses Secure Sockets Layer (SSL) encryption between the server and the client. This ensures that communication between the SPSS client and SPSS Server is unreadable by third parties if intercepted.

* Currently, this is possible only in the Microsoft® Windows desktop version of SPSS.

---

**SPSS Server**

SPSS Server enables you to analyze data on your server for better decision making throughout your enterprise. Data can reside on the same machine as SPSS Server or on a remote data server.
Features

Client/server architecture
- Reduce network traffic and improve performance with the data-free client feature. Administrators can limit users' viewing rights to the data dictionary when they're connected to SPSS Server.
- Run server-based "back end" processes such as data access, aggregation, transformations, and statistical analysis using SPSS command syntax language
- Reduce network traffic because data reside on the server and are not brought down to users' machines for analysis
- Reduce the amount of temporary space required for many processes
- Analyze massive datasets faster using server-grade hardware
- Increase the speed of your analyses by letting your server do the heavy computation work, freeing your desktop for other activity
- Work with a separate analytical server framework and receive:
  - Performance improvements, including the increased size of messages (which increases client/server communication speed and the optimized variable sort, especially on wide datasets) and faster data loading
  - The ability to run multiple instances of the SPSSB while the server framework manages the processes
  - Additional tools to increase productivity and performance

Copy-free data access in SQL DBMS
- Perform analysis without the need to convert data to SPSS format (data must be at the same level as the current ODBC)
- Sort and aggregate data inside the database prior to its retrieval for analysis
- Easily read-in data tables with the SPSS Data Access Pack
- Read data stored in SPSS (SAV) file format

Ability to launch multiple sessions
- Run multiple sessions of SPSS simultaneously on the same desktop
- Access multiple datasets simultaneously by running multiple sessions from a single desktop of SPSS client

Security
- Work efficiently within your vendor’s security framework
  - Require password protection when clients access SPSS Server
  - Set security levels and require passwords to access data sources
- Receive support for Open SSL

Communications framework between client and server
- Move client freely between server and local mode
- Work in a multi-platform environment (for example, use a Windows client with a UNIX® server)
- Work in multiple locations (for example, Japanese and French SPSS clients can be attached to a single English version of SPSS Server)

SPSSB
- Automate production of SPSS data preparation and statistical reports through command syntax files in a UNIX script or Windows batch files without requiring an active and connected SPSS client
- Use the following output formats: Text, HTML, and XML
- Save prepared data to the SPSS (SAV) file format
- Run more efficiently in a production environment using return codes
- Create any SPSS chart type (except SPSS Maps™ and interactive graphics) and export it in HTML format

Tunneling protocol
- Enable remote users to analyze data from off-site locations while keeping the data and SPSS Server safely behind a firewall. Modern internationalized communications protocols are included with SPSS Server to enable users to connect to SPSS Server using:
  - Point-to-Point Tunneling Protocol (PPTP)
  - Level 2 Tunneling Protocol (L2TP)
  - Network Address Translation (NAT)

Administrator controls
- Work with a utility that assists the SPSS administrator in monitoring and managing SPSS usage on the server
  - Start/stop user sessions
  - Start/stop server processes
  - Log events
  - Tune use of the server
  - Assign priority to individual users
  - Assign a unique disk to individual users for temporary files
  - Disconnect users
- Use a single Administrative Utility for working with SPSS, Clementine, and SPSS Predictive Enterprise Services™. With this utility, you can administer any combination of these three products installed at your site, eliminating the need for multiple administration tools.

Features subject to change based on final product release.
Maintain and increase functionality beyond that of the desktop version
■ Leverage the same functionality as the client-only product: Graphical user interface (GUI) or syntax-driven capabilities are available in all areas, including statistics, graphics, OLAP report cubes, and data transformations
■ Increase the tools available for preparing data and creating reports using tools unique to the server version
  – In-database sort and aggregate control
  – Open multiple XML models from SPSS, Clementine, or AnswerTree, and score new data using a scoring engine
■ Receive support for OLE DB
■ Filter irrelevant data to obtain only features relevant for modeling by using the Predictor Selection algorithm. This algorithm supports categorical and continuous independent and dependent variables, and accepts very large sets of predictors (up to 100,000).

Predict classification of cases by treating each variable as independent and equal with the Naïve Bayes algorithm. Predictors can be continuous or categorical, and the algorithm is best used when you have fewer than 200 predictors.
■ Seamlessly utilize third-party multithreaded sorting applications to significantly speed up data preparation tasks

Ability to maximize use of enterprise-level servers
□ Harness the full power of your multiple CPU servers using multithreaded ODBC to access your data
□ Take advantage of your powerful 64-bit machines with the 64-bit version of SPSS Server (contact your representative about availability for your operating system)
□ Stripe temporary files over multiple disks based on the administrator’s settings. This provides much greater speed in reading and writing large temporary files, which are often associated with time-consuming tasks such as sorting and aggregation.

System requirements
□ Operating system: Windows Server 2003 (32-bit or 64-bit); Sun™ Solaris™ (SPARC) 9 and later (64-bit only); IBM® AIX® 5.3 and later; or Red Hat® Enterprise Linux® ES4 and later, HP-UX II 64 bit (Itanium)
□ Memory: 256MB RAM per expected concurrent user
□ Minimum free drive space: 300MB (includes SPSS Server install, SPSS Server Administrator install, and the stand-alone SPSS Batch Facility [SPSSB] install)
□ Minimum CPU: Two CPUs recommended, Pentium class
□ Required temporary disk space: Calculate by multiplying 2.5 x number of users x expected size of dataset in megabytes

Features subject to change based on final product release.
The SPSS Programmability Extension dramatically increases the power, capabilities, and usability of SPSS Base and modules. Developers and end-users can use this feature to extend the SPSS command syntax language, introduce additional statistical functionality, and access the SPSS engine from external applications.

With the SPSS Programmability Extension, you can:

- Use external programming languages from within the SPSS command syntax by using the BEGIN PROGRAM and END PROGRAM commands
  - The external language for which you have installed integration support is invoked via BEGIN PROGRAM
  - Statements between BEGIN PROGRAM and END PROGRAM are written in the external programming language you have chosen, and are executed entirely by the external language’s processor
  - Different supported languages can be called in separate programs within SPSS command syntax

- Gain programmatic access to the SPSS analytical engine through an application program interface (API). APIs provide programs with:
  - Direct access to the active dataset’s variables, variable properties and attributes (name, format, labels, measurement level, type, and user-defined attributes), case count, and case data
  - Access to an in-memory, XML version of the data dictionary and procedure output
  - An XPath evaluation engine that allows access to and navigation of the in-memory XML workspace
  - A method for queuing and executing SPSS command syntax
  - Direct access to the last error code and message

- Develop your own procedures—including those for statistical analyses not included in SPSS
  - Define new syntax in SPSS style via an XML schema and have SPSS handle parsing and error checking
  - The procedure can send results into an SPSS pivot table or into text blocks—essentially extending the analytical capabilities of SPSS
**Benefits**
- Extend SPSS functionality. The SPSS Programmability Extension enables you to add functionality not included in SPSS.
- Write generalized and more flexible jobs. Create generalized jobs by controlling logic based on the Variable Dictionary, procedure output (XML or datasets), case data, and environment. Reusable code means data is not tied to a single program.
- Handle errors with generated exceptions. The SPSS Programmability Extension makes it easy to check whether a long syntax job worked. Hundreds of standard modules for Python are available.
- React to results and metadata
- Build SPSS functionality into other applications

**Programming capabilities**
Combining backend processor APIs with an external programming or scripting language opens up a limitless set of new possibilities from within SPSS syntax jobs.

For example, use the SPSS Programmability Extension to control the flow of your SPSS command syntax jobs through conditional execution control statements (such as “If/Then/Else”) and looping control statements (such as “For” and “While”) found in the external programming language’s syntax.

Use scripts written in external programming languages to conditionally execute or make decisions about which syntax is executed based on a particular condition, such as:
- The value of the variable attributes in the data dictionary
- Values in the output
- Values in the active dataset
- Error-level return codes from SPSS procedures

In short, you can create reusable code that speeds the process of turning data into decisions.

Additionally, take advantage of all your external programming language’s non-SPSS-related capabilities in your scripts. For example, have a production job trigger an e-mail notification once your job has successfully completed.

**SPSS Syntax Job Flow**

- Command 1
- Check state of dictionary, output, or return code
- Command 2
- Command 3

Control the flow of your SPSS syntax jobs. In this example, command 1 is executed. Then if the dictionary, output, or return code passes, command 2 is performed. If it fails, then command 3 is performed instead.

**Take advantage of procedures created and shared by other users through SPSS Developer Central**

**How to get started—integration plug-ins**
Since the SPSS Programmability Extension is included with SPSS Base 16.0, you can get started quickly. SPSS Programmability Integration Plug-Ins are available online.
Before installing the SPSS-Python Integration Plug-In, you will need to install Python. The version of Python recommended for your version of SPSS is included on the SPSS installation CD.

**SPSS-.NET Integration Plug-In**

The SPSS-.NET Integration Plug-In is a complete, freeware example plug-in for integrating the .NET** version of Microsoft Visual Basic with the SPSS Programmability Extension.

The SPSS-.NET Integration Plug-In includes:
- An installer that configures itself for use with SPSS
- A native .NET package, which contains a library of functions that interact with the SPSS backend processor API
- Complete documentation with examples

The SPSS-.NET Integration Plug-In allows you to drive the SPSS analytical engine from an external application.

Before installing the SPSS-.NET Integration Plug-In, you will need to download and install a copy of the .NET Framework from the Microsoft Download Center at www.microsoft.com/downloads.

* SPSS Inc. is not the owner or licensor of the Python software. All Python users must agree to the terms of the Python license agreement located on the Python Web site. SPSS does not make any statement about the quality of the Python program. SPSS fully disclaims all liability associated with your use of the Python program. For more information on Python, visit www.python.org.

** SPSS Inc. is not the owner or licensor of the .NET Framework. All .NET users must agree to the terms of the license agreement located on the Microsoft Web site. SPSS does not make any statement about the quality of the .NET Framework. SPSS fully disclaims all liability associated with your use of .NET Framework. For more information on .NET, visit www.microsoft.com/net.

New Programmability Integration Plug-Ins are being developed by SPSS Inc., and will be available to download at SPSS Developer Central as soon as they are ready.

**SPSS-Python Integration Plug-In**

The SPSS-Python Integration Plug-In is a complete, freeware example plug-in for integrating the open source Python* programming language with the SPSS Programmability Extension.

The SPSS-Python Integration Plug-In includes:
- An installer that configures itself for use with SPSS
- A native Python package, which contains a library of functions that interact with the SPSS backend processor API
- Complete documentation with examples

The SPSS-Python Integration Plug-In enables you to use the BEGIN PROGRAM and END PROGRAM syntax commands to extend SPSS syntax with Python programming. You can also use this plug-in to access and drive the SPSS backend processor from an external application.

An SPSS Programmability Integration Plug-In provides the crucial link and configuration instructions that enable an SPSS syntax job to take advantage of a specific external programming language or dynamic link library (DLL).

Also available for download is the SPSS Programmability Extension SDK. This provides software developers with the information needed to develop an SPSS Programmability Integration Plug-In for a programming language’s use with the SPSS Programmability Extension. In addition to providing documentation for creating a new plug-in, it includes the full source code for the example SPSS-Python Integration Plug-In.

The SPSS-Python Integration Plug-In is a complete, freeware example plug-in for integrating the open source Python* programming language with the SPSS Programmability Extension.

The SPSS-Python Integration Plug-In includes:
- An installer that configures itself for use with SPSS
- A native Python package, which contains a library of functions that interact with the SPSS backend processor API
- Complete documentation with examples

The SPSS-Python Integration Plug-In enables you to use the BEGIN PROGRAM and END PROGRAM syntax commands to extend SPSS syntax with Python programming. You can also use this plug-in to access and drive the SPSS backend processor from an external application.
**SPSS Developer Central**

SPSS Developer Central can be found at www.spss.com/devcentral/. It is the online resource for end users and software developers interested in SPSS-related programming and development. From this Web site, you can download programmability extensions and sample code, access forums and participate in discussions on programmability practices, and read in-depth articles on SPSS programmability topics.

At SPSS Developer Central, you’ll also find many example libraries and syntax jobs for use with plug-ins such as the SPSS-Python Integration Plug-In. Some examples of Python resources include:

- Functions for simplifying the calls to the SPSS backend processor API
- Functions for working with the SPSS Viewer
- Bootstrap regression
- Poisson regression

Another great resource for programmability in SPSS is **SPSS Programming and Data Management: A Guide for SPSS and SAS® Users, Fourth Edition**. This book documents the wealth of functionality beneath the SPSS user interface. It includes detailed examples of command syntax, the Output Management System (OMS), and extending command syntax with the Python® programming language.

---

**SPSS-R Integration Plug-In**

The SPSS-R Integration Plug-In is a complete, freeware example plug-in for integrating the R*** programming language with the SPSS Programmability Extension.

The SPSS-R Integration Plug-In includes:

- An installer that configures itself for use with SPSS
- An integrated R package, which contains a library of functions that interact with the SPSS backend processor API
- Complete documentation with examples

The SPSS-R Integration Plug-In enables you to use the BEGIN PROGRAM and END PROGRAM syntax commands to extend SPSS syntax with R programming.

Before installing the SPSS-R Integration Plug-In, you will need to download and install a copy of the R language from [www.r-project.org/](http://www.r-project.org/).

*** **SPSS Inc.** is not the owner or licensor of R. All R users must agree to the terms of the license agreement located on the R project web site. SPSS does not make any statement about the quality of R. SPSS fully disclaims all liability associated with your use of R. For more information on R, visit [http://www.r-project.org/](http://www.r-project.org/).