Complete end-to-end analysis with SPSS 11.0

The end-to-end analytical process: A standard for improved results
Executive summary

Good, accurate data analysis is essential for most organizations. With data analysis, decision makers can make well-informed decisions about the future of their organization. But the actual data analysis isn’t the only step in the analytical research process. To get the most accurate analysis and report it to decision makers, you must be able to plan the data analysis project, collect data, access and manage data easily, analyze data and then share your results. The analytical process is comprised of each of these steps.

In order to perform an end-to-end analysis, analysts need a set of tools that don’t hinder their work or slow them down. Unfortunately, many analysts lack an analytical solution that gives them broad, flexible analytics, much less one that covers stages of the analytical process.

One way for analysts to have a solution that covers stages of the analytical process is to build it themselves using a variety of specialty software from different vendors. The problem with this approach is that analysts run into incompatibility problems, proprietary file formats, and spend too much time creating work-around solutions, not finding results. Analysts experience wasted time and resources, and the process takes longer to do than it should.

A better recommendation for analysts may be to choose an analytical solution that is expandable through add-on modules or stand-alone software programs that have different or specialized functions.

SPSS 11.0 from SPSS Inc. is analytical software that provides analysts with the foundation for project planning, data collection, data access, data management and preparation, data analysis, reporting and deployment. Tightly integrated add-on modules and stand-alone software that work seamlessly with SPSS 11.0 enhance SPSS 11.0’s capabilities. Because SPSS 11.0 and its modules are tightly integrated, analysts do not run into the same incompatibility problems common with cobbled-together solutions. The result: instead of creating work-around solutions, duplicating work and resolving incompatible file formats, analysts can easily perform all the stages of the analytical process using a single software solution.
Table of contents

Executive summary ................................................................. 2
Choosing the right analytical software product ................................. 5

Stages of the analytical process ................................................ 6
1. Planning stage ....................................................................... 6
2. Data collection stage ........................................................... 7
3. Data access stage ............................................................... 7
4. Data preparation and management stage .................................. 7
5. Data analysis: descriptive and inferential ................................ 7
6. Reporting stage ..................................................................... 8
7. Deployment stage ............................................................... 9

Multi-vendor solution versus single-vendor solution ......................... 9
Multi-vendor approach .............................................................. 9
End-to-end, modular solution ..................................................... 10

SPSS 11.0 is the most thorough solution for end-to-end analysis .......... 11
SPSS Base 11.0 ................................................................. 11
SPSS add-on modules and stand-alone software .............................. 11

Planning stage ......................................................................... 12
Determine correct size of your sample data with SamplePower ........ 12
Determine customer preferences with SPSS Conjoint .................... 14

Data collection stage ............................................................... 14
Create interactive forms and surveys with SPSS Data Entry .............. 14

Data access stage ..................................................................... 15
Access data easily with the Database Wizard ................................. 16
Access data easily with the Text Wizard ...................................... 16
Work with SAS datasets or transport files .................................... 16

Data preparation stage ............................................................. 16
Prepare data quickly with the Data Editor ..................................... 16
Transform data easily with the Data Restructure Wizard ................. 17
Other data transformation techniques ......................................... 17
Estimate missing data with SPSS Missing Value Analysis ............... 18
Data analysis .................................................................19
  Work with large datasets using client/server architecture ..........19
  Benefits to network administrators ...................................20
  Describe your data ......................................................20
  Predict numerical outcomes .........................................21
  Segmentation and classification to identify groups .................23
  Forecasting with time-series data ....................................27
Reporting stage ................................................................28
  Communicate results in a tabular format with SPSS Tables ..........28
  Create reports using SPSS data with SPSS Report Writer ..........28
  Automate report production with scheduling software ..............29
  Interactive tables and graphs .........................................29
Deployment stage ..........................................................31
  Share results with decision makers using SmartViewer Web Server .....31
  Score live data in real time with SmartScore .........................32
Conclusion ..................................................................32
About the SPSS Business Intelligence division ..........................32
Appendix 1: product matrix ..............................................33
Appendix 2: contact SPSS for more information .......................34
  SPSS software for teaching ............................................34
Choosing the right analytical software product

Data analysis can be a time-consuming process, but it is essential for almost any organization. With data analysis, decision makers in your organization can set long-term organizational goals, better target new customers or retain existing ones, and better allocate resources throughout your organization. If results from analysis are inaccurate, your organization could experience:

- Lost revenue-generating opportunities
- Wasted resources and money
- Missed sales objectives
- Poor service levels
- Decreased customer satisfaction
- Increased customer turnover
- High levels of fraudulent activity
- Lower student test scores
- Reduced classroom enrollment
- Poor response to new public programs
- Drops in patients’ perceived quality of health care
- Low levels in job satisfaction

However, data analysis is not the only step in the analytical research process. To deliver informative, cost-conscious analyses to decision makers in your organization, you spend much of your time:

- Planning the process
- Gathering data
- Preparing data for analysis
- Displaying data analyses in a format that decision makers can use to explore results and act upon to make informed decisions

You need to have the right tools to help you through the analytical process from beginning to end. But what tools should you look for and use to help you reach your goals? And what combination of tools is right for your projects? This white paper summarizes the analytical process, the types of tools that are available, and why SPSS 11.0 provides you with the most compelling solution to perform end-to-end analysis.
Stages of the analytical process

The analytical process can be divided into seven stages, as shown in the graphic below:

Below are outlined the general goals you want to reach within each stage of the analytical process:

1. **Planning stage**

   Analysts need to take the time to plan analytical projects in the beginning to reduce costs and wasted resources in the long run. It’s important to map a course of action — you should not dive into analysis. Planning ensures that you:

   - **Define the research objective** — it is important to determine a clear objective of the analysis project right from the beginning. This will minimize wasted time and effort, and will provide focus and direction for your research project.

   Once a clear objective is determined, you can:

   - **Define the target population** — by defining your research objective at the beginning, you can also better choose the data you want to use for analysis.

   - **Select research design, data collection and analytical techniques**

   - **Establish a project budget** — mapping an action plan helps you justify your study and budget. Before beginning a study, many organizations must prove research has a viable chance of producing results. For example, a research organization applying for a grant may need to produce a power analysis with easy-to-understand charts and graphs. The charts and graphs will show funding agencies that the study in question will produce beneficial results, and thus deserves to be funded.

   - **Estimate required sample sizes** — if the sample size is too small, you risk not finding any significant results and will have to start over. On the other hand, if the sample size is too big, you will have to take time to sort through the data, resulting in wasted money because a smaller sample size would be sufficient.
2. Data collection stage

_If you already have your data you can skip the data collection stage._ You do not need to perform data collection if you already have data or use pre-purchased data. If you don't have your data collected, you need to collect clean, unbiased and up-to-date data in an efficient manner. The goals of the data collection stage depend on your project. For instance, if you are developing a customer survey analysis, you need to maximize your desired respondent pool and design questions that are not only relevant, but elicit the responses you are seeking.

3. Data access stage

The goal of the data access stage is to read the data into analytical software for further processing. Data can come from a variety of sources, including databases, transactional systems, Web logs, surveys and Web downloads.

4. Data preparation and management stage

The goals of the data preparation and management stage include:

- **Clean the data for analysis** — cleansed data help ensure your analytical results are accurate. For example, if you have gender data in which “1” is for male and “2” is for female, your data shouldn’t have “3” as a response.

- **Set up metadata** — metadata are any variable definition information, including variable names, variable formats, and descriptive variable labels (data such as gender or income level) and value labels (numbers assigned to data, such as “1” is for male, “2” is for female). Metadata are stored in the data “dictionary.”

- **Transform your data** — get your data in the structure and form needed for analysis

- **Set up multiple-item indices and scales** — combine multiple indices into a single, multiple-item index for projects such as customer surveys. This can provide a more reliable measurement of the concept of interest.

- **Fill in missing data** — replace missing data values with estimates to ensure better summary statistics using expectation-maximization or regression algorithms.

5. Data analysis: descriptive and inferential

Before running data through models, you first need to understand the data you’re going to process. This is _descriptive data analysis_. In descriptive data analysis, you want to understand your data by asking questions such as, “What’s the distribution of my data?” or, “What does the average person in my study score?” The data analysis stage can be time-intensive, and you want a solution that provides breadth of features and flexibility, but also the ability to dig deep into data.

The goals of descriptive data analysis are to:

- Summarize your data

- Get an accurate description of the variables of interest
Once you have performed descriptive data analysis, you perform *inferential data analysis*. Inferential data analysis includes:

- **Predicting numerical outcomes** — sometimes your outcomes are numerical. For example, you might want to predict income, customer lifetime value, standardized test scores or patient length of stay. You might want answers to questions such as “How well can an outcome be predicted by a set of predictors?” or “What are the best predictors of an outcome?”

- **Working with related multiple dependent variables** — do you need a flexible procedure that works simultaneously to provide regression analysis with related multiple dependent variables? You may need procedures that provide flexible design and contrast options to estimate means and variances, as well as to test and predict means. For example, a financial services company might want to know what predicts account balance for various account types.

- **Measuring the same people over time** — do you need to measure the same people over time, for example, to measure how perceived quality of health care increases or decreases? Being able to make the same measurement a number of times on individual subjects or cases can give you long-term perspective on your products or services.

- **Estimating nonlinear equations** — are you working with models that have nonlinear relationships? For example, are you predicting coupon redemption as a function of time and the number of coupons distributed?

- **Identifying and classifying groups** — sometimes you want to predict what groups people and things belong in. For example, to better target groups of people based on their purchasing patterns, you might want to group people into similar buying groups (do they buy frequently, for special occasions only or once per year?).

- **Forecasting future events** — the goal of forecasting is to analyze your organization’s historical information — such as sales figures, student enrollments, crime rates — and build models on this data, and then use them to predict future events.

6. **Reporting stage**

The target audience for your reports is generally not as analytical or technically savvy as an analyst. Therefore, the goal of the reporting stage is to create easy-to-understand results from your data analysis for decision makers who can quickly understand and act upon your information. Reports can be generated as Word documents, presentations, tables or through the Web on your organization’s intranet.
7. Deployment stage
The goals of the deployment stage are to:

- **Increase the return on your analytical investment** — bring new knowledge and information to decision makers in your organization so that they may better plan the future of your organization.

- **Tailor reports for the target audience** — make sure your reports take into account your readership. Don’t dump information in your audience’s lap. Most readers of your data analysis results won’t be analysts or necessarily as knowledgeable as you about the analytical results at hand.

- **Deliver results into the hands of decision makers** — this includes distributing results via the Web and creating interactive reports.

Multi-vendor solution versus single-vendor solution
The analytical process is extensive, and analysts need a thorough set of tools that can help them move from one stage of the process to the next. When choosing your analytical toolset, you can:

- Build an end-to-end software solution using a variety of tools from different vendors that perform part of the analytical process, for example, one of the stages, or part of a stage.

- Use a modular, but tightly integrated, end-to-end software solution from a single vendor.

Multi-vendor approach
The benefits of combining various software products to build a system are:

- Choosing tools that deliver the specialized features needed for a particular project.

- Not being tied to a single software vendor.

However, analysts who analyze data using different products often run into several issues:

- **Not enough statistical horsepower** — the software used, such as spreadsheets, may not have enough statistical tools for analysts to use; and may use all of a PC’s computations cycles for large projects leaving analysts unable to use their workstations for hours at a time.

- **Inaccurate software means inaccurate results** — general purpose “office suite” software, such as spreadsheets, often are not as accurate for specialized statistical techniques, resulting in inaccurate analyses.

- **Niche software wastes resources** — specialized software purchased for one project may be inappropriate for other projects, meaning you might have to purchase additional software.

- **Difficult to work with other software in solution** — analysts who use a solution built from many different software tools will find that the various software don’t always work well together, and this negatively affects the analytical process in general. Analysts may have to format and reformat the same data in order to access it within different software, thereby wasting time and resources. Or, analysts may find that creating reports from data analyses, or sharing results over the Web, is difficult and time-consuming.
For instance, when collecting data, performing analysis on data, and then creating a report, an analyst using products from multiple vendors may have to re-enter variable definition information, including variable names, formats, and descriptive variable and value labels. Plus, to share the analysis, the analyst may have to hand-code HTML pages in order to effectively distribute the results to decision makers in the organization. This wastes valuable time and effort, and is inconvenient, especially during crunch time.

- **Inadequate customer service support** — a multi-vendor approach means that to get adequate customer service for your solution, you will have to contact different vendors for each software product. Each company has different support and consulting services policies, and you will have to navigate these policies to get the service needed to perform your job.

**End-to-end, modular solution**

Instead of cobbling together a variety of software tools from different vendors, you can choose to use a modular, integrated analytical product that encompasses all stages of the analytical process. In this type of solution, a single product provides you with the framework for planning, data collection, data access, data management and preparation, data analysis, reporting and deployment. Add-on modules and stand-alone software are available for specialty needs and additional capabilities.

The benefits of a modular, integrated solution include:

- **Data format compatibility** — by using an end-to-end analytical product from a single vendor, you won’t experience data format incompatibility problems, common with the multi-vendor approach. By spending less time and resources on data formats, you can focus on actual data analysis and reporting.

- **Seamless workflow for increased time savings** — a modular, integrated solution from a single vendor allows you to work quickly and seamlessly from one stage to the next. Because you don’t have to worry about data formatting or duplicating work, you can streamline the entire analytical process. Because all software works together, analysts can smoothly move from data planning and data access and management to reporting and sharing results.

- **Integrated customer service options** — in contrast to a multi-vendor approach, an end-to-end modular solution provides you with a single point for all your customer support needs. A single vendor makes it easier for you to get technical support, training and other consulting services you may need. Furthermore, having one place to contact for each piece of your analytical software means that the vendor’s technical support group and consulting services understand how each product and module integrates with the other.
SPSS® 11.0 is the most thorough solution for end-to-end analysis

SPSS 11.0 is a modular analytical product for all stages of the analytical process, from planning, data collection, data access, data management and preparation, to reporting and deployment.

SPSS 11.0 is a modular solution that allows you to plug in add-on modules and integrate stand-alone software which enhance the core SPSS 11.0 system with specialized capabilities. Because SPSS and its modules work seamlessly together, you don't have to duplicate work, such as reformatting data, that's often required when using products from a variety of vendors. For instance, once you enter variable labels in SPSS Data Entry™, you won't have to set up variable labels again throughout the process.

The SPSS product line can be divided into three product categories:

- SPSS Base
- SPSS add-on modules
- Stand-alone software that also tightly integrates with SPSS Base

**SPSS Base 11.0**

SPSS Base 11.0 takes you from planning all the way through to deployment of results and models. Core capability is flexible data analysis, which includes a broad range of statistical and procedural capabilities for descriptive analysis, numerical prediction and group identification. You can also enhance the SPSS Base 11.0 system via macros, matrices and published Application Programming Interfaces (API).

**SPSS add-on modules and stand-alone software**

When you combine SPSS Base 11.0 with any SPSS add-on module or stand-alone product, your analytical capabilities grow. Add-on modules and stand-alone products provide you with enhanced tools necessary for planning, data collection, data access, data management and preparation, data analysis, reporting and deployment. All SPSS modules and stand-alone software programs work seamlessly with SPSS Base 11.0 and also share common data sources (including dictionary information and metedata), so you don't have to duplicate work or worry about different data formats.

You can add as much analytical capability to your system as you need using modules and stand-alone software:

- **Add-on modules** easily plug into and enhance SPSS Base 11.0 with added capabilities. For instance, SPSS Trends™ provides you with powerful forecasting tools to build models for time-series data and predict future events.

- **Stand-alone software programs** add capabilities that complement SPSS 11.0, but also work outside SPSS Base 11.0 software. For instance, AnswerTree® is a stand-alone software product that enables you to build decision trees for segmentation.
The illustration below shows the analytical process and related SPSS modules and stand-alone products. Modules and stand-alone products are described in greater detail throughout this paper. All stages of the analytical process make use of the SPSS Base 11.0 software.

Planning stage

The planning stage is where you take the time to map out your project, design the research objective, define the target population and select your data collection techniques. Mapping an action plan not only helps you determine your targets, but also allows you to justify your project and estimated budget.

Determine correct size of your sample data with SamplePower®

In order to improve your chances of finding significant effects where they exist without wasting precious resources, you need to conduct a power analysis. A power analysis determines the likelihood that your study will yield a statistically significant result if one actually exists. Any organization about to conduct research, or that is planning research, can take advantage of power analyses and increase its chances of getting funding for its projects. Such projects include:

- Survey research
- Experimental researching settings
Clinical trials

Any research organization applying for funding (grants or contracts)

SamplePower; a stand-alone software program, performs power analyses. SamplePower determines the proper sample size that meets your research criteria so you don’t waste time or money because:

- Too small a sample size yields inaccurate results, leading to wasted time and effort and possible Type 2 Errors (meaning, possible positive results exist, but you aren’t able to find them)
- Too large a sample size wastes money

With SamplePower, you can specify aspects of your test situation, such as:

- Effect size
- Alpha level
- One- or two-tailed tests

1 Developed by Michael Borenstein, Ph.D., Hannah Rothstein, Ph.D., and Jacob Cohen, Ph.D.
SamplePower’s capabilities allow you to plan for the following types of analyses:

- ANOVA and ANCOVA
- Correlations
- Equivalence tests
- Logistic regression
- Means and their differences
- Proportions and their differences
- Regression
- Survival analysis

**Determine customer preferences with SPSS Conjoint™**

The SPSS Conjoint module helps you during product or service planning. With SPSS Conjoint, you can better allocate your resources by determining your customer preferences and the products and services they want or need. SPSS Conjoint empowers you to:

- Automatically select the minimal number of attribute combinations (such as price, package design, etc.) needed to get all the information to make a good decision. This is done using fractional factorial designs and it isn’t limited to only two-level factors.
- Produce printed cards for a conjoint experiment (PLANCARDS); the printed cards are used as stimuli to be sorted, ranked or rated by the subjects.
- Perform an ordinary least squares analysis of preference or rating data (CONJOINT) on group, subgroup and individual levels.

**Data collection stage**

Once your analytical plan is set, it may be important to collect clean, unbiased and up-to-date data in an efficient manner. The goals of the data collection stage depend on your project. For instance, if you are developing a customer survey analysis, you need to maximize your desired respondent pool, and design questions that are not only relevant, but that elicit the types of responses you are looking for.

**Create interactive forms and surveys with SPSS Data Entry**

SPSS Data Entry is a stand-alone software family. Comprised of SPSS Data Entry Builder, SPSS Data Entry Station and SPSS Data Entry Enterprise Server, the SPSS Data Entry family works in conjunction with SPSS Base. SPSS Data Entry enables you to easily collect data for your analysis by helping you build interactive forms and surveys that:

- **Minimize data cleaning by encouraging valid responses** — SPSS Data Entry helps researchers write questionnaires that encourage valid answers and minimize data cleaning. For example, you can design your surveys so that people cannot enter illegitimate responses, such as entering “3” for gender (where “1” is male and “2” is female).

SPSS Data Entry’s Rules Wizard further encourages valid results by helping you define data entry and collection rules to check the validity of your survey responses and the accuracy of entered data and to encourage complete answers from online respondents.

---

1 If you already have your data, such as prepurchased data, you can skip the data collection stage.
For example, use checking and validation rules to verify you received valid responses. You can also use skip-and-fill rules to speed respondents past unnecessary parts of the survey based on their answers.

- **Reduce time spent setting up variable and value labels** — SPSS Data Entry allows you to set up the variable and value labels you want in analysis and reporting when designing your surveys. By setting up labels, you can use the same labels globally throughout the entire analytical process — thus reducing redundant work.

- **Minimize turnaround time and reduce errors by deploying surveys on the Web** — SPSS Data Entry Builder enables you to export your surveys and forms for Web or intranet deployment using SPSS Data Entry Enterprise Server. Your turnaround time is significantly reduced when respondents directly enter data either through a Web browser or using SPSS Data Entry Station for high-volume data entry projects. SPSS Data Entry conveniently collects responses to a central location, saving them directly into the SPSS data file format by default, or any ODBC-compliant database by design. Response rules and requirements built into forms and surveys ensure the appropriate data are collected and are ready for analysis with minimum preparation. Because SPSS Data Entry can save data in the SPSS native data file format, you can set automatic production routines that automatically analyze and publish interactive results to the Web, based on the data collected. For example, you can set SPSS 11.0 to analyze data collected from SPSS Data Entry Enterprise Server and publish results nightly to SmartViewer Web Server (discussed on page 31) on the organization’s intranet. Decision makers can then access and interact with the results the next day using their Web browser. For more information on automating production, see “Automate report production in any format with scheduling software” on page 29.

- **Secure forms through SSL** — SPSS Data Entry uses SSL encryption between the respondents’ browsers and your SPSS Data Entry Server.

**Data access stage**

If you use multiple software products from different vendors, there’s a chance one or more of the products won’t have the correct driver for your data. And, without the correct data driver, you can’t access your data. SPSS 11.0 includes drivers for many ODBC-compliant databases, including Oracle, SQL Server, DB2 UDB, Microsoft Access, Sybase and Siebel (through an ODBC-compliant driver), which means you can access your data hassle-free throughout the analytical process. With the right drivers, you can connect to any ODBC-compliant database, such as NCR — resulting in minimal data handling using conversion-free/copy-free data access.
Access data easily with the Database Wizard

SPSS 11.0's built-in Database Wizard enables you to easily access massive amounts of data from numerous database sources (supported databases are discussed on page 15, “Data access stage”) without having to write code or complex syntax. The Database Wizard guides you through the process of accessing data and generates the proper code in the background.

Access data easily with the Text Wizard

Read text data — such as survey data, purchased data or downloaded data — in a variety of formats using the Text Wizard. For example, purchased data comes in fixed, free-field or tab-delimited ASCII formats. Other software requires you to use complex syntax just to access the data. With the Text Wizard, you can quickly read data just by clicking a mouse.

Work with SAS® datasets or transport files

SPSS Base 11.0 has a built-in GET SAS command that helps you build SPSS-format working data files from SAS datasets or SAS transport files (Version 8 or earlier). SPSS 11.0 automatically adjusts SAS variables and values for SPSS formats. For example, SAS variable names are converted to valid SPSS variable names.

Data preparation stage

Data preparation can be a time-consuming process if you use software products from different vendors. With SPSS 11.0 Base, you can prepare your data quickly and easily using built-in wizards and commands.

Prepare data quickly with the Data Editor

SPSS 11.0's Data Editor makes data preparation efficient so you can get to analysis faster. The Data Editor is a centralized area where you can set up information about data and copy and paste attributes between variables. Information about data that you can set up includes:

- Variable labels
- Value labels
- Missing values
- Variable types
- Formats

Use the Data Editor in conjunction with data transformation techniques to assure data fit your analytical procedure.
Transform data easily with the Data Restructure Wizard

Do you work with data that requires painstaking program data transformations in order to analyze it? The Data Restructure Wizard in SPSS 11.0 enables you to take a data file that has multiple records per subject and restructure them — without setting up vectors or loops — so data for each subject are in single records. This is particularly helpful if you work with transactional data. If you prefer, SPSS 11.0 also enables technically proficient analysts to work directly with vectors and loops. In SPSS 11.0, you can also do the reverse action — take data from a single record and spread it across multiple cases for techniques such as LINEAR MIXED MODELS.

Other data transformation techniques

Other transformation techniques that help get data ready for analysis empower you to:

- Compute new variables using arithmetic, cross-case, data and time, logical, missing-value, random-number, statistical or string functions
- Recode string or numeric value
- Recode values into consecutive integers
- Create conditional transformations using DO IF, ELSE IF, ELSE and END IF statements
- Use programming structures, such as do repeat-end repeat, loop-end loop and vectors
- Get a clearer understanding of your continuous data's distribution as you automatically divide variables into categories via NTILES and create new variable(s) for each category
- Count occurrences of values across variables
- Make transformations permanent or temporary
- Execute transformations immediately, batched or on demand
- Cumulative distribution, inverse cumulative distributions and random number generator functions: beta, Cauchy, chi-square, Exponential, F, gamma, laplace, logistic, lognormal, Normal, Pareto, student t, uniform, Weibull
– Standard bivariate normal variates with correlation \( r \), Half Normal, inverse Gaussian, Studentized range, Studentized maximum modulus

- Cumulative distribution and random number generator for discrete distribution functions: Bernoulli, Binomial, Geometric, Hypergeometric, Negative Binomial, Poisson

- Cumulative distribution for Non-central distribution: Non-central Beta, Non-central Chi-square, Non-central F, Non-central T

- 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, Weibull
  - Discrete distributions: Bernoulli, binomial, geometric, hypergeometric, negative binomial, Poisson

- Non-central density/probability functions for: non-central beta, non-central chi-square, non-central F distribution, non-central t distribution

- Tail probabilities: Chi-square, F

- Auxiliary function: logarithm of the complete Gamma function

**Estimate missing data with SPSS Missing Value Analysis™**

Missing data can seriously affect your results. When you ignore missing data or assume that excluding missing data from calculations is sufficient, you risk getting invalid results.

<table>
<thead>
<tr>
<th>Case</th>
<th>Age</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>Female</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>Male</td>
</tr>
<tr>
<td>3</td>
<td>39</td>
<td>Male</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>Female</td>
</tr>
<tr>
<td>5</td>
<td>42</td>
<td>Male</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>Female</td>
</tr>
<tr>
<td>7</td>
<td>37</td>
<td>Male</td>
</tr>
<tr>
<td>8</td>
<td>39</td>
<td>Male</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case</th>
<th>Age</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>21</td>
<td>Female</td>
</tr>
<tr>
<td>2</td>
<td>22</td>
<td>Male</td>
</tr>
<tr>
<td>3</td>
<td>39</td>
<td>Male</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>Female</td>
</tr>
<tr>
<td>5</td>
<td>42</td>
<td>Male</td>
</tr>
<tr>
<td>6</td>
<td>18</td>
<td>Female</td>
</tr>
<tr>
<td>7</td>
<td>37</td>
<td>Male</td>
</tr>
<tr>
<td>8</td>
<td>39</td>
<td>Male</td>
</tr>
</tbody>
</table>

Figure 6: missing data can seriously impact the conclusions you draw from your data. The table on the left shows how missing data can impact your results. In this example, when you calculate the mean and ignore the missing data the average age is 39. However, if you impute the missing data the mean age will be 29 — a difference of a generation.

SPSS Missing Value Analysis module is a critical tool for anyone concerned about the validity of data, including survey researchers, social scientists, data miners and market researchers. With SPSS Missing Value Analysis, you can easily examine data from several different angles with six diagnostic reports to uncover missing data patterns. You can then estimate summary statistics and impute missing values through statistical algorithms.

SPSS Missing Value Analysis helps you to:

- Diagnose if you have a serious missing data problem
- Replace missing values with estimates, for example, impute your missing data with EM or regression algorithms
Data analysis

Every piece of your acquired data has intrinsic value. The key is extracting this value. By accessing and analyzing massive data tables thoroughly, analysts can enhance their own understanding of their data. Analysts can increase their customer awareness by analyzing the most intimate customer details.

As a solution for end-to-end analysis, SPSS 11.0 provides you with the tools you need to progress seamlessly from one stage to the next. Since you don’t have to worry about file formats, data formatting, or importing and exporting data files, SPSS 11.0 lets you concentrate on its core capabilities, which are:

- Deep, flexible data analysis
- A broad range of statistics and techniques, which includes a wide variety of techniques for descriptive analysis, numerical prediction and group identification and forecasting. These techniques are described in detail on the following pages.

SPSS 11.0's analytical depth and breadth and long-term applicability to a wide range of current and future problems means you will realize a long-term return on your investment. Tightly integrated add-on modules and stand-alone applications that work seamlessly with SPSS Base 11.0 give you even more analytical horsepower. For example, when using SPSS Advanced Models, you can now more accurately analyze nested-structure data using Linear Mixed Models.

Work with large datasets using client/server architecture

You can leverage your IT investment when using a larger server machine with a client/server version of SPSS 11.0 installed. With SPSS Server 11.0, you can analyze terabytes of data on a server, and manipulate that data using SPSS 11.0 for Windows on a desktop client.

SPSS Server 11.0’s client/server architecture means you process data where they can be handled more effectively. SPSS Server 11.0 takes advantage of a server’s faster processing and scalability. This means you can perform number-crunching operations, such as all statistical calculations, on the server. For example, if you have multiple users, you can analyze data on the server rather than running data on users’ individual machines, and then easily publish data results to decision makers throughout the organization.
Other benefits of SPSS Server 11.0’s client/server architecture are:

- Faster data reads because data remains on the server and don’t need to be copied to the client machine or converted into an SPSS file
- Reduced data handling processing times with conversion-free/copy-free data access from SQL DBMS database with the SPSS Data Access Pack (included with SPSS 11.0 Server)
- Tackling large datasets easily because SPSS 11.0 Server practically eliminates file size limitations
- Getting better performance because less temporary file space is needed when using techniques (such as recoding variables, saving regression residuals or saving cluster membership) that change or add more data
- Using a desktop client to interact with, or manipulate, large datasets on a server

Benefits to network administrators

SPSS Server 11.0 makes it easier for your IT department to configure and maintain your SPSS deployment. With SPSS Server 11.0, your administrators can:

- Deploy SPSS Server 11.0 on a variety of popular server platforms, including Windows® NT, Sun Solaris, IBM AIX and HP-UX
- Upgrade a new, large database once, rather than individually modify each client. Point the SPSS Server 11.0 to the new database. Because all clients have access to SPSS Server 11.0, they can get instant access to the new database.
- Enable remote users to tunnel through firewalls to access data securely using various port-forwarding schemes

Describe your data

Before you run your data through models or create reports and summaries, you want to understand your data. You might ask questions such as “What’s the distribution?” or “What does the average person score?” SPSS Base 11.0 gives you a variety of ways to summarize your data and accurately describe variables of interest:

- Data displays, such as frequency distributions and histograms, help you understand data distribution and identify outliers so you can determine if you need to re-enter or recode your data
- Measures of central tendency, such as mean and median
- Estimates of variance, such as standard deviation

You can also use a variety of comprehensive significance tests to immediately know if your results are meaningful or if differences are random. Results can usually be accompanied by significance tests (Z-tests, t-tests and F-tests) and confidence intervals with a mouse-click.
Predict numerical outcomes

SPSS can help you predict numerical data. For example, you might want to predict income, customer lifetime value, standardized test scores or patient length of stay. SPSS 11.0 and its supporting software have a variety of techniques that let you answer these questions and more.

- **Get accurate predictive models using nested-structure data with Linear Mixed Models** — SPSS 11.0’s Linear Mixed Models technique, which is a more accurate extension of regression and ANOVA, enables you to measure not only means but also variances and covariances in your data. Analysts can measure data that display correlation and non-constant variability, such as students nested within classrooms or consumers nested within families. For example, educators can discover how different teaching methods affect student test scores across individual classrooms in the district. Linear Mixed Models’ flexibility means you can formulate dozens of models, including Fixed Effects ANOVA, Split-Plot Design, Multi-Level Models with Fixed-Effects Covariance and Randomized Complete Blocks Design. By taking a hierarchical structure of your data into its modeling, SPSS 11.0 gives you more accurate predictive models.

You can also use Linear Mixed Models if you are working with repeated measures data, including situations in which there are different numbers of repeated measurements or different intervals for different cases or both. For example, a health care organization that tests patients recovering from an illness might not have uniformly collected data. Ideally, the organization would test individuals at regularly scheduled intervals (for example, every 90 days). But people often miss appointments or are tested earlier or later than the optimal time. More standard techniques will discard data that don’t meet the test conditions. Linear Mixed Models, however, will use all your data so you get a more accurate picture of patient recovery.

- **Work with related multiple dependent variables using General Linear Models (GLM)** — the General Linear Models multivariate technique is a flexible technique that works simultaneously to provide regression analysis with related multiple dependent variables. GLM provides flexible design and contrast options to estimate means and variances and to test and predict means. For example, a financial services company might want to know what predicts account balance for various account types. You can also mix and match categorical and continuous predictors to build models. Because GLM multivariate doesn’t limit you to one type of predictor variable, you have options that give you a wealth of model-building possibilities.

For example, a manufacturer might want to examine the deal sizes between products and types of customers. It can test two types of effects using GLM multivariate:

- **Main effects** — this includes questions such as, do public and private sector customers differ, on average, in terms of deal size regardless of whether customer purchases product A or product B?

- **Interaction** — does the difference between deal sizes for product A and product B differ by type of customer? For example, the private sector deal size may be greater for product A than product B; in the public sector, deal size is greater for B than A.
- **Measure the same people over time using GLM** — SPSS 11.0 provides you with the flexibility to mix and match categorical and continuous-level predictors, including interactions. Using the GLM repeated measures technique, you can make the same measurement a fixed number of times on individual subjects or cases.

You can also test three types of effects using GLM repeated measures. For example, a hospital might want to study these relationships:

- **Between subjects** — does perceived quality of care vary by hospital department?
- **Within subjects** — is perceived quality of care increasing or decreasing over time?
- **Interaction** — is perceived quality of care increasing or decreasing at different rates for different hospital departments?

- **Estimate nonlinear equations using Nonlinear Regression and Constrained Nonlinear Regression** — SPSS 11.0 enables you to estimate nonlinear equations using one of two SPSS techniques:

  - **Nonlinear Regression (NLR)** — for unconstrained problems, NLR enables you to estimate models with arbitrary relationships between independent and dependent variables using iterative estimation algorithms.

  - **Constrained Nonlinear Regression (CNLR)** — for both constrained and unconstrained problems, CNLR gives you more options so you can customize your models. CNLR also empowers you to:
    - Use linear and nonlinear constraints on any combination of parameters
    - Estimate parameters by minimizing any smooth loss function (objective function)
    - Compute bootstrap estimates of parameter standard errors and correlations

- **Identify latent variables in your data with Amos™**

  With Amos, you can get better estimates by combining information from multiple measures into broader concepts. Amos enables you to use multiple indicators to obtain a more complete understanding of broader latent concepts, or concepts you cannot see. For example, you might want to examine the relationship between factors, such as promptness of delivery, ease of return and level of customer service assistance received that affect customer satisfaction. You can quickly verify how variables affect each other and why they are affected that way using Amos.

  Models built using Amos stand-alone software reflect relationships more accurately because any variable, whether observed or hidden, can be used to predict other variables. Amos complements SPSS Base 11.0 by enabling you to analyze multiple groups or data from several populations at the same time.

  Amos makes structural equation modeling (SEM) easy with its rich, visual framework. Unlike traditional SEM methods, you simply build graphical models in Amos to illustrate the dependencies and relationships found in your data.

  Developed by James Arbuckle, Ph.D.
Other numerical outcome procedures — SPSS also uses the following techniques for predicting numerical outcomes:

- **Linear Regression** — enables you to explore the relationships between predictors and what you want to predict. For example, predict sales using price and customer type.

- **Weighted Least Squares Regression for weight estimation** — appropriate when the variance in the dependent variable isn’t constant within the population under study.

- **Two-Stage Least Squares** — appropriate when the predictor and outcome have reciprocal effects on each other.

- **Survival Analysis procedures** — enables you to examine the distribution of times between two events, such as length of customer retention, even when the second event isn’t recorded. For example, customers who are still loyal, or people who are still employed. Survival Analysis techniques include:
  - Cox Regression with time-dependent covariates
  - Kaplan-Meier
  - Life Tables

Segmentation and classification to identify groups
Sometimes you want to predict what groups people and things belong in. For example, you might want to group people into similar buying groups — for example, those who buy frequently, those who buy for special occasions only or those who buy once a year — to target promotions at purchasers. SPSS 11.0’s techniques for identifying groups fall under two types: segmentation/dimension reduction and classification.

Uncover important relationships through segmentation with segmentation/dimension reduction — segmentation allows you to discover groups you might not have considered or groups you didn’t know existed. When using segmentation techniques, the algorithm divides subjects, objects or variables into a number of relatively homogeneous groups. Your classes are not predefined; the segmentation procedure you select will define groups for you.

SPSS 11.0 enables you to segment your customer, market and research data using powerful data reduction techniques, including:

- **Factor Analysis and Principal Components Analysis** — Factor Analysis and Principal Components Analysis (PCA) techniques have a high degree of flexibility and give you a number of methods for factor extraction, rotation and factor score computation. Factor Analysis is used to identify underlying variables or factors that explain the correlations within a set of observed variables. For example, you can use it in data reduction to identify a small number of factors that explain most of the variance observed in a much larger number of manifest variables.

The purpose of PCA is to form linear combinations of the observed variables that account for the maximum amount of their variance in as few components as possible (subject to being uncorrelated with each other).
Multidimensional Scaling — Multidimensional Scaling is appropriate when you want to find a structure in a set of distance measures between objects and cases. For example, you could identify dimensions that describe consumers' perceptions between different cars, using similarity ratings between different makes and models.

Segmentation

- Use cluster analysis to group cases in smaller-sized datasets through segmentation — with Hierarchical Cluster Analysis in SPSS 11.0, you can take clusters from a single record and form groups until all clusters are merged. You can choose from over 40 measures of similarity or dissimilarity, standardize data using several methods and cluster cases or variables. You can also:
  - Analyze raw variables or choose from a variety of standardizing transformations
  - Generate distance or similarity measures using the proximities technique
  - Display statistics at each stage to help you select the best solution

While Hierarchical Cluster Analysis is impractical for clustering very large numbers of objects, you are not required to specify the number of clusters you want to analyze in advance. This technique is recommended for datasets that are smaller in number; for example, focus group lists.

For example, a research group could use Hierarchical Cluster Analysis to identify types of television shows that attract similar audiences for each show type. The organization could cluster TV shows into homogeneous groups based on viewer characteristics to identify segments for advertising.

- Group data from larger datasets through segmentation — you typically use K-means Cluster Analysis if you want to group larger datasets, such as customer mailing lists. This technique assumes data fall into a known number of clusters. Given this number, the technique will assign cases to clusters. You can select one of two methods to classify cases — either update cluster centers iteratively or classify only.

The research group might want to cluster cities into homogeneous groups so that it can select comparable cities to test marketing strategies using K-means Cluster Analysis.

Classification

- Target the right people more effectively with classification — when using classification techniques, you know the outcome you want to predict. You use predictor fields to predict a categorical target field, such as:
  - Who will respond to your offer?
  - Is the applicant a good credit risk?
  - Who will graduate?
  - Who will vote?
Target the right people with AnswerTree® — AnswerTree stand-alone software complements SPSS 11.0 by empowering you to segment and group using scalable decision trees. Four powerful decision tree algorithms enable you to detect segments and patterns in your data:

- CHAID
- Exhaustive CHAID
- Classification and Regression Tree (C&RT)
- Quest

AnswerTree groups your data with respect to a desired outcome. For example, with AnswerTree, you might discover moderate-profit customers are more likely to respond to personalized Web offerings, or students who miss more than 45 days of school a year are twice as likely to drop out. Decision trees are represented graphically in an easy-to-read tree diagram so you can quickly see the important segments and patterns in your data. You can even have AnswerTree generate tree diagrams automatically, yet still maintain the size of the tree.

In addition, AnswerTree’s unique evaluation graphs provide at-a-glance summaries of selected segments — giving you a clear picture of your results. Once you have results, you can use SmartScore software to write information to a database or SPSS files using SQL or SPSS syntax. Then, use scores to make decisions about your customers.

Group people with respect to a desired outcome with Binary Logistic Regression — the Binary Logistic Regression techniques in SPSS 11.0 empowers you to select the predictive model for dichotomous dependent variables. Predictors can be data that are continuous and categorical.

Binary Logistic Regression gives you depth and flexibility to specify models and to choose predictor order inclusion. You can use six types of forward- or backward-stepwise methods to select variables:

- Work forward — tell the model to select the strongest variables until there are no more significant predictors in the dataset or until only significant predictors remain
- Work backwards — at each step, the technique removes the least significant predictor in the dataset until only significant predictors remain

You can also set inclusion or exclusion criteria. The technique produces a report telling you the action it took at each step to determine your variables.
Classify people into two or more groups with Multinomial Logistic Regression —
do you need to classify individuals into two or more groups? When a dependent variable
includes two or more categories, the Multinomial Logistic Regression technique in SPSS
11.0 gives you what's needed to accurately predict group membership within key groups.
For example, a telecommunications company can build a model to predict if a customer
will most likely order caller ID, voice mail, three-way calling. If the model predicts the
customer is likely to order caller ID, then it can send direct mail emphasizing caller ID
to that customer. This means the company won't waste resources advertising products
or services that are not likely to interest its customer. Like Binary Logistic Regression,
your predictors can be continuous or categorical data.

Build models for ordinal data using PLUM
Do you want to rank people in terms of their likelihood to take a specific action? For
example, you might want to know the probability (such as, low, medium or high) a
person will enter a program.

When you need to predict ordinal-dependent variables, SPSS 11.0's PoLytomous Universal
Models (PLUM) gives you a versatile method to build models for ordinal data. When you
choose different link functions, you get the flexibility to fit ordinal logistic regression,
ordinal probit models and ordinal Cauchit models. You can also model both the location
and scale of the underlying distribution.

PLUM gives you options to save predicted probabilities for all dependent variable categories
back to your data. Like the other classification techniques in SPSS 11.0, your predictors
can be continuous or categorical data.

Other classification techniques — other techniques you can use for classification include:

- Loglinear/Logit Models — used to analyze the frequency counts of observations
falling into each cross-classification category into a crosstabulation or a contingency
table. For example, a market research group could use data from a product preference
report to determine the relationship between exposure to an advertisement and
product placement.

- Model Selection Loglinear Analysis — appropriate when you want to find out
which categorical variables have an association. You can use both forced entry and
backward elimination methods to build models; if you are building saturated models,
you can request parameter estimates and tests of partial association.

- General Loglinear Analysis — used to analyze the frequency counts of observations
falling into each cross-classification category in a crosstabulation or contingency table

- Logit Loglinear Analysis — used to analyze the relationship between categorical
dependent (or response) variables and independent (or explanatory) variables

- Probit Models — appropriate when the outcome is dichotomous, predictors are
the level of a stimulus (such as a drug dosage level) or you want to measure the
relationship between the strength of a stimulus and the proportion of cases exhibiting
a certain response to the stimulus
– **Categorical Regression** — used to quantify categorical data by assigning numerical values to categories, resulting in an optimal linear regression equation for transformed variables. You could use Categorical Regression to describe how customer satisfaction depends on ease of purchase, price and quality. The resulting equation could be used to predict customer satisfaction for any combination of the three independent variables.

– **SPSS Exact Tests™** — analyze small datasets and still get correct results. Use SPSS Exact Tests to determine if a relationship between variables exists when you have a small number of cases, variables with a high percentage of responses in one category or have to subset your data into fine breakdowns. You can also use SPSS Exact Tests when you’re searching for rare occurrences (such as sales above $1 million).

**Forecasting with time-series data**

Accurate, long-term forecasting is generally thought of as a time-consuming and tedious process — but essential for almost any organization in today’s fast-moving marketplace. Decision makers in your organization need to be able to set long-term goals — and know how to achieve them — based on your organization’s past performance and tacit knowledge of your industry.

- **Analyze historical data and predict events using SPSS Trends** — time-series analysis is the most powerful tool you have for analyzing historical information, building models and forecasting future events. Whatever data you are examining — sales figures, student enrollments, crime rates — the SPSS Trends add-on module gives you an easy-to-use graphical interface to analyze historical information and predict future events.

You can use SPSS Trends to:
- Monitor quality standards
- Manage forecasting systems performance
- Run sales forecasts
- Study public opinion

Select from SPSS Trends’ estimation techniques to:
- Use ARIMA models for seasonal or nonseasonal time-series data
- Access techniques performing seasonal decompositions
- Estimate up to four parameters in 12 different models for exponential smoothing
- Use different regression methods: trend regressions, regression models with first-order autoregressive errors
- Decompose a time series into its harmonic components

Figure 9: chart created with SPSS Trends illustrates housing starts, raw and seasonally differenced over a ten-year period. Using seasonal difference helps to clarify your relationships.
Reporting stage

The challenge of reporting your data analysis is making sure the report is tailored to your audience, and thus, easier to understand. If you use one or more pieces of software to crunch data, and another to generate reports, you may have difficulty getting the data analysis into the report-generating software. You might have to export your data, tweak formats and then import into the report software. A single-source solution, like SPSS 11.0, enables you to analyze your data and generate reports without having to repeatedly export and reformat all your data.

The most important thing to remember when creating a report is that you are creating it for your audience: not all decision makers are analysts, and many don’t have your technical skills or knowledge.

Communicate results in a tabular format with SPSS Tables™

SPSS Tables module enables analysts to summarize and communicate their results in a tabular format, including complex stub-and-banner tables, contingency tables and display listings of survey data.

SPSS Tables is very useful for people, such as survey or market researchers, who update reports, like customer surveys or marketing reports, on a regular basis. With SPSS Tables, analysts can create new reports by recalculating new data, and inserting the new data into established templates. The SPSS Tables module includes special capabilities for survey analysis, such as layering, nesting, stacking and multiple response analysis, so you can generate reports on surveys, customer satisfaction and polling results.

For instance, market researchers can condense the results of multiple-response questions, such as answers to “check all that apply,” into a single table to make it easy to present this information. SPSS Tables also distinguishes between the different types of missing values, such as “do not know,” “didn’t answer,” “not applicable,” “refused to answer” or any other missing value you specify. This enables report readers to see the full picture in regards to the type of responses you receive.

Once a report is created using SPSS Tables, you can easily print your results, publish them to the Web using SmartViewer® Web Server (see page 31 for more information on SmartViewer Web Server) or insert them in Microsoft® Office programs, such as PowerPoint, for presentations.

Create reports using SPSS data with SPSS Report Writer™

The SPSS Report Writer stand-alone software program uses SPSS value and variable labels to make your reports more descriptive — and saves you time because you do not have to re-enter this content.

After you have completed your analysis, you can open your SPSS files in SPSS Report Writer to create professional-looking, presentation-quality reports, form letters, or mailing labels from your SPSS data. SPSS Report Writer allows you to display your SPSS data on a Microsoft Word-like page layout. SPSS Report Writer gives you a variety of formatting capabilities so you can present your data in a professional way.
Automate report production in any format with scheduling software

SPSS 11.0 graphical user interface (GUI) generates syntax behind the scenes so you can save time by automating repetitive techniques in production mode.

For example, if you need to generate a monthly report on data collected from customer surveys, you can use SPSS Data Entry to collect the data and write to an SPSS file at the data warehouse.

You can then create the first report in the SPSS GUI. After the first report is generated, you can paste the syntax into the Production Facility to set up a report that runs on a regular basis. You even can have collected data published in HTML and hosted on your organization’s intranet or extranet by using SmartViewer Web Server.

The Production Facility uses filters to pull only the data needed from the data warehouse. It runs unattended and terminates after executing the last command, so you can still be productive on your PC. You can schedule recurring production jobs using scheduling software, such as Windows NT Scheduler. This means you can run a big job overnight and have the results waiting for you in the morning.

Interactive tables and graphs

SPSS 11.0 includes a number of graphical features and chart types so you can provide visuals that complement text in technical papers, journals, reports and slide presentations. A variety of chart types means you can display results just as you want. Graphing features help your audience better understand your results. Templates and other features enable you to save selected characteristics of a chart and apply them to others automatically, making it easy for you to prepare high-quality reports.

- **Explore your data with interactive tables with SPSS report OLAP cubes** — SPSS 11.0 features report OLAP cubes, interactive tables that enable you to explore all kinds of data from every angle and aspect. SPSS report OLAP cubes include SPSS statistic capabilities, and can work right from your SPSS data files to include value and variable labels. SPSS’ report OLAP cubes are easier to use than OLAP cubes found in other solutions. Unlike other solutions, you can set up SPSS’ OLAP cubes yourself without having to involve your organization’s IT department. Also, SPSS 11.0 can give meaning to millions of rows by aggregating the data within seconds in an OLAP cube.

- **View your data with interactive graphs** — charting and graphing enable you to clearly see differences within your data through color variations, size, patterns, and textures. SPSS 11.0’s graphing and charting capabilities go far beyond typical spreadsheet capabilities. With SPSS 11.0, you can transform information and watch your chart update on the fly. For example, you can change a chart that shows revenue by store location to show revenue by day of the week. You can even take a complex chart and simplify it by splitting it into multiple categories for a side-by-side comparison.
# Reporting options in SPSS

<table>
<thead>
<tr>
<th>If you need to…</th>
<th>SPSS Base: REPORT or LIST command</th>
<th>SPSS Base: OLAP cubes</th>
<th>SPSS Base: Crosstabulations</th>
<th>SPSS Base + SPSS Tables</th>
<th>SPSS Base + SPSS Report Writer</th>
<th>SPSS Base + Microsoft Office programs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Report a list of cases in an SPSS data file, without analysis or aggregation</td>
<td>Create interactive reports with pivoting capabilities, including data aggregation and statistics</td>
<td>Explore and report the relationships between categorical variables, such as customer type and geographic region</td>
<td>Analyze and report survey or market research data. This includes data aggregation, statistics, and special capabilities for survey analysis, such as layering, nesting, stacking and multiple response analysis.</td>
<td>Create presentation-quality paper reports, form letters or mailing labels from your SPSS data, including aggregation after report creation.</td>
<td>Create financial reports with custom calculations using results of SPSS analytical output. Or, share results electronically with a small number of non-SPSS users.</td>
<td></td>
</tr>
<tr>
<td><strong>Primary application</strong></td>
<td>Create reports that list sales revenue, products, employees, students and citizens.</td>
<td>Interactive reporting according to hierarchies, such as product sales or sales by region.</td>
<td>Survey and market researchers who need to analyze customer/citizen demographics and preferences.</td>
<td>Same as Crosstabulations, but with superior ability to arrange data, such as layering, stacking and nesting.</td>
<td>Create professional-looking sales, revenue, product, inventory, employee, student and citizen reports.</td>
<td>Use SPSS variables or results of analysis in management or client reports or presentations.</td>
</tr>
<tr>
<td><strong>Analytics</strong></td>
<td>None</td>
<td>23 statistics, including sum, count, mean and standard deviation.</td>
<td>15+ statistics, including significance and hypothesis testing, such as Chi Square and Correlations testing.</td>
<td>35+ stats, including counts, mean, mode, median, row percent and column percent.</td>
<td>Count, subtotal, sum, average, standard deviation, max, min.</td>
<td>The results of any SPSS analysis, plus basic statistical, financial, mathematical, other user-defined functions, written text.</td>
</tr>
</tbody>
</table>

Figure 11: This table describes some of the reporting options for SPSS. Find your project and application on the reporting options table above to help determine which reporting option is right for you.

- **Create compelling, print-quality graphs using SigmaPlot**— with SigmaPlot, you can create high-quality graphs right from SPSS 11.0. Graphs created using SigmaPlot are suitable for technical papers and journals, formal reports, slide presentations and more. SigmaPlot enables you to customize every detail of your graph to make it look just the way you want, including:
  - Adding text and images anywhere on your graph
  - Specifying exact thickness of lines and objects
  - Scaling axes

Figure 12: Create high-quality graphs right from SPSS 11.0 using SigmaPlot and ensure your graphs are publishable.
Attach geographic data to maps using SPSS Maps™ — SPSS Maps, an add-on module, enables you to turn geographic data into visual output. With SPSS Maps, you can include geographic information in maps and generate customized analyses by geographic areas or boundaries. You can also export maps as standard graphic files for distribution to many mediums, including the Web.

Deployment stage
Once data analysis is performed in SPSS 11.0, you can distribute the results to your intranet or Web site without having to hand-code HTML. By sharing your results throughout your organization, everyone in the decision-making process can see how their decisions can affect your organization’s long-term results.

Share results with decision makers using SmartViewer Web Server
To ensure decision makers in your organization use data results, SPSS 11.0 and SmartViewer Web Server enable you to easily share your results with the people who can use them. Using SmartViewer Web Server, analysts can publish output from SPSS 11.0 to a secure Web-based system where decision makers can view your results using a Web browser.

SmartViewer Web Server centrally stores all of your organization’s analyses — including pivot tables, graphs and charts, report OLAP cubes, and output from third-party products — and organizes it all within a convenient home page for each user so they don’t have to search through files to find reports.

Unlike static reports, SmartViewer Web Server makes information more meaningful by enabling people to interact with key findings. Everyone, from managers, analysts and remote workers, can:

- Drill down through tables for more detail
- Use pivoting capabilities to pivot rows, columns and report layers
- Access only the information you specify to secure confidential reports from the public

Figure 13: SmartViewer Web Server enables your audience to view and interact with results online when you publish them quickly, easily and directly from the SPSS Output Viewer.
Score live data in real time with SmartScore®

SmartScore delivers accurate profiles by scoring your data, one of the most efficient data mining techniques available. Using SmartScore, raw data is translated into information that can provide a rational basis for decisions.

SmartScore is a software development kit that you can use to score cases using models deployed in SPSS 11.0, as well as in AnswerTree and other SPSS products. The SmartScore development kit allows you to deploy models in real time so that decision makers or decision-dependent computer systems can make better decisions about data as they receive it.

SmartScore works seamlessly within your workflow either interactively or in a batch mode. Uses for SmartScore range from identifying potential problem customers to upselling new products and/or services to existing customers.

Conclusion

SPSS Base 11.0, add-on modules and stand-alone products greatly streamline the end-to-end data analysis process. Because SPSS 11.0 is modular, you can actually grow the system by incorporating other SPSS software products, such as AnswerTree and SPSS Data Entry, and add-on modules, such as SPSS Regression Models, enhancing the base software with specialized capabilities. Furthermore, because SPSS Base and the add-on modules all work together, analysts don’t have to redo work, such as reformatting data, often required when using products from a variety of vendors.

About the SPSS Business Intelligence division

The SPSS Business Intelligence division helps people solve business problems using statistics and data mining. This predictive technology enables our customers in the commercial, higher education and public sectors to make better decisions and improve results. The SPSS Business Intelligence division software and services are used successfully in a wide range of applications, including customer attraction and retention, cross-selling, survey research, fraud detection, enrollment management, Web site performance, forecasting and scientific research. The SPSS Business Intelligence division’s market-leading products include SPSS®, Clementine®, AnswerTree®, DecisionTime® and SigmaPlot®. For more information, visit our Web site at www.spssbi.com.
Appendix 1: product matrix

Need help deciding which SPSS products you need for specific applications? Listed by product and function, this diagram guides you to the right products — so you can start getting results. Visit [www.spssbi.com/spss](http://www.spssbi.com/spss) to view specifications for each product listed below.

<table>
<thead>
<tr>
<th>Survey/market research and direct marketing</th>
<th>SPSS Base 11.0</th>
<th>SamplePower</th>
<th>SPSS Complete</th>
<th>SPSS Data Entry</th>
<th>SPSS Marketing</th>
<th>SPSS Regression</th>
<th>SPSS Advanced Modules</th>
<th>Amos</th>
<th>SPSS Complex</th>
<th>SPSS Exact Tests</th>
<th>AnswerTree</th>
<th>SPSS Trends</th>
<th>SPSS Tables</th>
<th>SPSS Report Writer</th>
<th>SPSS Map</th>
<th>SPSS Models</th>
<th>SPSS Viewer</th>
<th>SPSS Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer satisfaction surveys</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Pricing and promotion analysis</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Market segmentation studies</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Demographic studies and opinion polling</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Database marketing/direct marketing</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

| Academic                                     |               |             |               |                 |               |                 |                       |     |             |                 |            |            |            |                 |        |             |             |             |
| Teaching                                    | ✓              | ✓           | ✓             | ✓               | ✓             | ✓               | ✓                     | ✓   | ✓           | ✓               | ✓          | ✓          | ✓          | ✓                |        | ✓          | ✓          | ✓          |
| Administration                               |              |             |               |                 |               |                 |                       |     |             |                 |            |            |            |                 |        |             |             |             |

| Administrative research and human resources/resource planning |               |             |               |                 |               |                 |                       |     |             |                 |            |            |            |                 |        |             |             |             |
| Program effectiveness                        | ✓              | ✓           | ✓             | ✓               | ✓             | ✓               | ✓                     | ✓   | ✓           | ✓               | ✓          | ✓          | ✓          | ✓                |        | ✓          | ✓          | ✓          |
| Employee attitude/satisfaction surveys       | ✓              | ✓           | ✓             | ✓               | ✓             | ✓               | ✓                     | ✓   | ✓           | ✓               | ✓          | ✓          | ✓          | ✓                |        | ✓          | ✓          | ✓          |
| Applicant selection and testing              | ✓              | ✓           | ✓             | ✓               | ✓             | ✓               | ✓                     | ✓   | ✓           | ✓               | ✓          | ✓          | ✓          | ✓                |        | ✓          | ✓          | ✓          |
| Compensation and employment analysis         | ✓              | ✓           | ✓             | ✓               | ✓             | ✓               | ✓                     | ✓   | ✓           | ✓               | ✓          | ✓          | ✓          | ✓                |        | ✓          | ✓          | ✓          |

| Medical, scientific and social science research |               |             |               |                 |               |                 |                       |     |             |                 |            |            |            |                 |        |             |             |             |
| Treatment outcome analysis                   | ✓              | ✓           | ✓             | ✓               | ✓             | ✓               | ✓                     | ✓   | ✓           | ✓               | ✓          | ✓          | ✓          | ✓                |        | ✓          | ✓          | ✓          |
| Behavioral and biomedical research           | ✓              | ✓           | ✓             | ✓               | ✓             | ✓               | ✓                     | ✓   | ✓           | ✓               | ✓          | ✓          | ✓          | ✓                |        | ✓          | ✓          | ✓          |
| Environmental impact studies                 | ✓              | ✓           | ✓             | ✓               | ✓             | ✓               | ✓                     | ✓   | ✓           | ✓               | ✓          | ✓          | ✓          | ✓                |        | ✓          | ✓          | ✓          |
| Criminal justice studies                     | ✓              | ✓           | ✓             | ✓               | ✓             | ✓               | ✓                     | ✓   | ✓           | ✓               | ✓          | ✓          | ✓          | ✓                |        | ✓          | ✓          | ✓          |
| Outcomes management                          | ✓              | ✓           | ✓             | ✓               | ✓             | ✓               | ✓                     | ✓   | ✓           | ✓               | ✓          | ✓          | ✓          | ✓                |        | ✓          | ✓          | ✓          |
| Project management                           | ✓              | ✓           | ✓             | ✓               | ✓             | ✓               | ✓                     | ✓   | ✓           | ✓               | ✓          | ✓          | ✓          | ✓                |        | ✓          | ✓          | ✓          |
| Research and development engineering         | ✓              | ✓           | ✓             | ✓               | ✓             | ✓               | ✓                     | ✓   | ✓           | ✓               | ✓          | ✓          | ✓          | ✓                |        | ✓          | ✓          | ✓          |

| Planning and forecasting                     |               |             |               |                 |               |                 |                       |     |             |                 |            |            |            |                 |        |             |             |             |
| Sales/marketing forecasting and budgeting    | ✓              | ✓           | ✓             | ✓               | ✓             | ✓               | ✓                     | ✓   | ✓           | ✓               | ✓          | ✓          | ✓          | ✓                |        | ✓          | ✓          | ✓          |
| Resource requirements and forecasting        | ✓              | ✓           | ✓             | ✓               | ✓             | ✓               | ✓                     | ✓   | ✓           | ✓               | ✓          | ✓          | ✓          | ✓                |        | ✓          | ✓          | ✓          |
| New product forecasting                      | ✓              | ✓           | ✓             | ✓               | ✓             | ✓               | ✓                     | ✓   | ✓           | ✓               | ✓          | ✓          | ✓          | ✓                |        | ✓          | ✓          | ✓          |
| Econometrics                                 |               |             |               |                 |               |                 |                       |     |             |                 |            |            |            |                 |        |             |             |             |
| Enrollment/admissions forecasting            | ✓              | ✓           | ✓             | ✓               | ✓             | ✓               | ✓                     | ✓   | ✓           | ✓               | ✓          | ✓          | ✓          | ✓                |        | ✓          | ✓          | ✓          |
| Financial account balance                    |               |             |               |                 |               |                 |                       |     |             |                 |            |            |            |                 |        |             |             |             |

| Quality improvement                          |               |             |               |                 |               |                 |                       |     |             |                 |            |            |            |                 |        |             |             |             |
| Productivity and service quality             | ✓              | ✓           | ✓             | ✓               | ✓             | ✓               | ✓                     | ✓   | ✓           | ✓               | ✓          | ✓          | ✓          | ✓                |        | ✓          | ✓          | ✓          |
| Manufacturing and productivity analysis       | ✓              | ✓           | ✓             | ✓               | ✓             | ✓               | ✓                     | ✓   | ✓           | ✓               | ✓          | ✓          | ✓          | ✓                |        | ✓          | ✓          | ✓          |
| Predictive maintenance planning               | ✓              | ✓           | ✓             | ✓               | ✓             | ✓               | ✓                     | ✓   | ✓           | ✓               | ✓          | ✓          | ✓          | ✓                |        | ✓          | ✓          | ✓          |
| Utilization studies                          | ✓              | ✓           | ✓             | ✓               | ✓             | ✓               | ✓                     | ✓   | ✓           | ✓               | ✓          | ✓          | ✓          | ✓                |        | ✓          | ✓          | ✓          |

| Reporting and ad-hoc decision making          |               |             |               |                 |               |                 |                       |     |             |                 |            |            |            |                 |        |             |             |             |
| Fraud detection/compliance                    | ✓              | ✓           | ✓             | ✓               | ✓             | ✓               | ✓                     | ✓   | ✓           | ✓               | ✓          | ✓          | ✓          | ✓                |        | ✓          | ✓          | ✓          |
| Risk management/credit management             |               |             |               |                 |               |                 |                       |     |             |                 |            |            |            |                 |        |             |             |             |
| Reports to or from governmental agencies      |               |             |               |                 |               |                 |                       |     |             |                 |            |            |            |                 |        |             |             |             |
| Industry trend analysis                       |               |             |               |                 |               |                 |                       |     |             |                 |            |            |            |                 |        |             |             |             |
| Program evaluation and funding decisions      | ✓              | ✓           | ✓             | ✓               | ✓             | ✓               | ✓                     | ✓   | ✓           | ✓               | ✓          | ✓          | ✓          | ✓                |        | ✓          | ✓          | ✓          |
| Institutional research                        |               |             |               |                 |               |                 |                       |     |             |                 |            |            |            |                 |        |             |             |             |
| Property and tax assessment                   | ✓              | ✓           | ✓             | ✓               | ✓             | ✓               | ✓                     | ✓   | ✓           | ✓               | ✓          | ✓          | ✓          | ✓                |        | ✓          | ✓          | ✓          |
Appendix 2: contact SPSS for more information

For more information on pricing and licensing, visit www.spss.com to contact your local SPSS office or visit our online store. SPSS also offers a full-range of services to help you get the most out of your SPSS 11.0 analytical software:

- **Technical Support** (www.spss.com/tech/spssdefault.htm) provides you with solutions to problems or questions you may have.

- **Consulting Services** (www.spss.com/spssbi/consulting) save you time and money by helping you automate your SPSS procedures through customized scripts.

- **Training Services** (www.spss.com/spssbi/training) help you to expand your knowledge of SPSS 11.0 software through courses held at SPSS locations, on-site training or Web-based and computer-based distance learning.

**SPSS software for teaching**

If you’re an instructor who teaches statistics or classes involving research methods, SPSS has a number of options for your courses. Options vary by country, but can include full copies of SPSS software for student purchase and licensing for on-campus computer labs. Contact your local office to discuss your options.