

The risks of using spreadsheets for statistical analysis

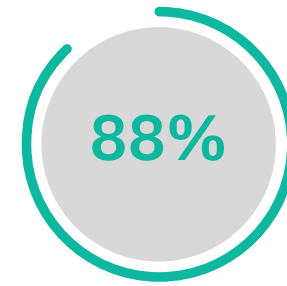
Why spreadsheets have their limits,
and what you can do to avoid them

Let's go 

Introduction

Spreadsheets are widely used for statistical analysis; and while they are incredibly useful tools, they are useful only to a certain point. When used for a task they're not designed to perform, or for a task at or beyond the limit of their capabilities, spreadsheets can be somewhat risky.

This paper presents some points you should consider if you use, or plan to use, a spreadsheet to perform statistical analysis. It also describes an alternative that in many cases will be more suitable.



of all spreadsheets contain at least one error

Introduction

Why spreadsheets are popular

An alternative to spreadsheets

The learning curve with IBM
SPSS Statistics

Software licensing
the SPSS way

Conclusion



Why spreadsheets are popular

A spreadsheet is an attractive choice for performing calculations because it's easy to use. Most of us know (or think we know) how to use one. Plus, spreadsheet programs come as a standard desktop computer resource, so they're already available.

A spreadsheet is a wonderful invention and an excellent tool—for certain jobs. All too often, however, spreadsheets are called upon to perform tasks that are beyond their capabilities.

Moreover, the perception that a spreadsheet is easy to use is, to some extent, an illusion. It is always easy to get an answer out of a spreadsheet—but it's not necessarily easy to get the correct answer.

However, the decision to use something else—an unfamiliar technology or tool—is not always an easy one. When considering an alternative, two questions come to mind: How useful is this tool? And how hard is it to learn?

The answer to the first question depends on the scale and the complexity of your data analysis. A typical spreadsheet will have a restriction on the number of records it can handle, so if the scale of the job is large, a tool other than a spreadsheet may be very useful.

1	Employee Code	Gender	Date of Birth	Educational Level (years)	Employment Category	Current Salary	Beginning Salary	Months since Hire
2	1	m	3-Feb-52	15	3	\$57,000.00	\$27,000.00	98
3	2	m	23-May-58	16	1	\$40,200.00	\$18,750.00	98
4	3	f	26-Jul-29	12	1	\$21,450.00	\$12,000.00	98
5	4	f	15-Apr-47	8	1	\$21,900.00	\$13,200.00	98
6	5	m	9-Feb-55	15	1	\$45,000.00	\$21,000.00	98
7	6	m	22-Aug-58	15	1	\$32,100.00	\$13,500.00	98
8	7	m	26-Apr-56	15	1	\$36,000.00	\$18,750.00	98
9	8	f	6-May-66	12	1	\$21,900.00	\$9,750.00	98
10	9	f	23-Jan-46	15	1	\$27,900.00	\$12,750.00	98
11	10	f	13-Feb-46	12	1	\$24,000.00	\$13,500.00	98
12	11	f	7-Feb-50	16	1	\$30,300.00	\$16,500.00	98
13	12	m	11-Jan-66	8	1	\$28,350.00	\$12,000.00	98
14	13	m	17-Jul-60	15	1	\$27,750.00	\$14,250.00	98
15	14	f	26-Feb-49	15	1	\$35,100.00	\$16,800.00	98
16	15	m	29-Aug-62	12	1	\$27,300.00	\$13,500.00	97
17	16	m	17-Nov-64	12	1	\$40,800.00	\$15,000.00	97
18	17	m	18-Jul-62	15	1	\$46,000.00	\$14,250.00	97
19	18	m	20-Mar-56	16	3	\$103,750.00	\$27,510.00	97
20	19	m	19-Aug-62	12	1	\$42,300.00	\$14,250.00	97
21	20	f	23-Jan-40	12	1	\$26,250.00	\$11,550.00	97
22	21	f	19-Feb-63	16	1	\$38,850.00	\$15,000.00	97
23	22	m	24-Sep-40	12	1	\$21,750.00	\$12,750.00	97
24	23	f	15-Mar-65	15	1	\$24,000.00	\$11,100.00	97
25	24	f	27-Mar-33	12	1	\$16,950.00	\$9,000.00	97
26	25	f	1-Jul-42	15	1	\$21,150.00	\$9,000.00	97
27	26	m	8-Nov-66	15	1	\$31,050.00	\$12,600.00	96
28	27	m	19-Mar-54	19	3	\$60,375.00	\$27,480.00	96
29	28	m	11-Apr-63	15	1	\$32,550.00	\$14,250.00	96
30	29	m	28-Jan-44	19	3	\$135,000.00	\$79,980.00	96
31	30	m	17-Sep-61	15	1	\$31,200.00	\$14,250.00	96
32	31	m	24-Feb-64	12	1	\$36,150.00	\$14,250.00	96
33	32	m	20-Jan-54	10	3	\$110,000.00	\$45,000.00	96

Introduction

Why spreadsheets are popular

1 • 2 • 3 • 4 • 5
6 • 7 • 8 • 9

An alternative to spreadsheets

The learning curve with IBM
SPSS Statistics

Software licensing
the SPSS way

Conclusion



Spreadsheets can be useful for statistical analysis; but when used for tasks they're not designed to perform, they have some limitations.

As for complexity, if you only need a superficial review of your data, a spreadsheet may be a suitable tool. But if you suspect that there is valuable information in your data that isn't immediately obvious, or if you need to perform a detailed analysis or find hidden patterns, a spreadsheet will not give you the functionality you need.

Another factor to consider is the degree of accuracy required. Spreadsheet results can be unreliable when working with large datasets and/or performing complex calculations. If absolute accuracy is required, a spreadsheet may not suffice. Instead, a different, more reliably accurate tool should be considered.

The effort required in using a spreadsheet is also a consideration. An analyst using spreadsheet for performing statistical analysis will have to spend more

time preparing data because of lack of specialized data preparation features.

Finally, if the task is simply to analyze a limited quantity of historical data, a spreadsheet will do. But if you want to make reliable forecasts or draw trends, especially if they involve large datasets, then there are much better tools.

This paper will return to consider the answer to the second question—how hard is it to learn?—when it looks at an *alternative to spreadsheets* for statistical calculations. Before continuing, however, it is worth noting that people use spreadsheets for tasks other than numerical calculation. For example, spreadsheets are frequently used as if they were databases, to create and manage lists. Again, the principles of scale and complexity apply. Beyond certain limits, a proper database, with built-in rules for structuring data, maintaining data integrity, developing audit trails and so on, is far more suitable.

Two things to remember about spreadsheets

Users need to keep two important considerations in mind when working with spreadsheets: spreadsheets can be

Introduction

Why spreadsheets are popular

1 • 2 • 3 • 4 • 5
6 • 7 • 8 • 9

An alternative to spreadsheets

The learning curve with IBM SPSS Statistics

Software licensing the SPSS way

Conclusion



complex to create, and they can be error prone.

Spreadsheets are really computer programs

When you design a spreadsheet layout, you are writing a computer program. Spreadsheet programs such as Microsoft Excel use what is known as a “nonprocedural programming language.” Although it is also possible to write procedural programs for Excel in Visual Basic, the everyday business of typing formulas into cells is an exercise in nonprocedural programming.

Creating a spreadsheet can be as complex as computer programming.

Nonprocedural programming is just as full of decisions, complexity and chances

Introduction

Why spreadsheets are popular

1 • 2 • 3 • 4 • 5

6 • 7 • 8 • 9

An alternative to spreadsheets

The learning curve with IBM
SPSS Statistics

Software licensing
the SPSS way

Conclusion



for mistakes as all but the simplest procedural program.

With standard software development methodology, procedural computer programs are double- and triple-checked. By contrast, a spreadsheet, although it may be vitally important to the operation of a company, is usually the work of one person. It is almost never checked or tested in detail, and quite often goes into production with little or no verification. Yet important management decisions—revenue forecasts and plans for future investment, for example—are based on the numbers that it produces.

Spreadsheets are prone to errors

A number of studies have been made concerning the frequency of errors in spreadsheets. Based on these, it seems that 88 percent of all sheets contain at least one error. The studies were carried out by making visual inspections of mission-critical spreadsheets, so it is possible that many other errors were not found. Also, it was found that attempts to correct errors often introduced new ones.

Examples of costly spreadsheet errors from around the world include:

- “...a cut and paste error in a spreadsheet **cost TransAlta \$24M**”¹
- “...a simple spreadsheet formatting error made MI5, the British Intelligence Agency **bug more than one thousand wrong phones**”²
- “..Harvard University economists Carmen Reinhart and Kenneth Rogoff have acknowledged making a spreadsheet calculation mistake in a research paper, ‘Growth in a Time of Debt’ (PDF), which has been **widely cited to justify budget-cutting**”³
- “...London Olympics ended up selling **10,000 tickets for non-existent seats** because of a spreadsheet error”⁴

Types of spreadsheet errors

Spreadsheet errors can be divided into three main types.

The “friendliest” type is what you could call functional errors. These errors are the easiest to find because they simply stop the spreadsheet from working. Instead of giving you wrong numbers, they give error messages, or nothing at all.

Introduction

Why spreadsheets are popular

1 • 2 • 3 • 4 • 5
6 • 7 • 8 • 9

An alternative to spreadsheets

The learning curve with IBM
SPSS Statistics

Software licensing
the SPSS way

Conclusion



There are a number of stories about how spreadsheet errors have had embarrassing consequences.

One dates from 2013, when a graduate student uncovered a number of errors in the spreadsheets from a highly influential 2010 research paper used by many government policy makers to justify budget cutting.⁶ While Harvard University economists Carmen Reinhart and Kenneth Rogoff have acknowledged their mistake, they maintain that their conclusions about the association between higher government debt and slower economic growth is still valid.⁷

Another story is about a \$100 million “oops.” Companies are bought and sold all the time. And once the parties agree on a price, that’s usually it. However, in October 2014, the number of outstanding shares of Tibco Software was misstated because of one small spreadsheet error. As a result, the sale price of the company—a price all parties had agreed to—was overstated by \$100 million. When the transaction was completed, the acquired company’s shareholders received \$100 million less than they expected. The only folks smiling here were the litigators.⁶

Then there are outlier errors. With these, the spreadsheet appears to work, but the numbers aren’t right. Often, such errors are spotted by someone who has an idea of what the results should be and draws attention to the fact that the results don’t match expectations.

Types of spreadsheet errors include functional errors, outlier errors and stealth errors, which range in severity from low to high.

The most serious are what might be called stealth errors. These produce incorrect results, but nobody believes they’re incorrect. They pass inspection and are accepted as the truth. Stealth errors occur either because nobody knows what the correct result should be (which is often the case with statistical calculations), or the numbers are only slightly different from expectations and seem reasonable.⁵

Introduction

Why spreadsheets are popular

1 • 2 • 3 • 4 • 5

6 • 7 • 8 • 9

An alternative to spreadsheets

The learning curve with IBM
SPSS Statistics

Software licensing
the SPSS way

Conclusion



Accidentally inserting a number into a cell already containing a formula will overwrite the equation and turn the cell content into a constant.

Causes of spreadsheet errors

Spreadsheet users should know what factors cause errors. Unfortunately, there are too many causes to list here, but the major ones are:

- **Mistakes in logic:** These can be something simple, such as calling the wrong function, subtracting instead of adding, or the omission of parenthesis in formula creation. These sorts of errors can also be caused by the implied relationship of the cells in the spreadsheet.
- **Incorrectly copied formulas:** Typing in an equation while reading it from another location often leads to errors, as does cutting and pasting. Copying existing equations to new locations commonly changes the referenced cells, making it important to review your results for accuracy.
- **Accidentally overwritten formulas:** A cell that contains an equation looks like a number—all the user sees on first glance is the result. So accidentally inserting a number into a cell already containing a formula will overwrite the equation and turn the contents of the cell into a constant. If other formulas rely on the results from this cell, the error can be compounded significantly.

- **Misuse of built-in functions:** The wrong function can be used—for example, using AVERAGEA, which evaluates text and false entries to zero, instead of AVERAGE, which ignores them. This type of mistake is, unfortunately, a very easy mistake to make.
- **Omitted factors:** It is very easy to simply leave something out. This could be an equation, data or both. Errors of this type occur quite often when new data is added to a previously completed spreadsheet. It could be that not all the data is entered, or the some of the new cells aren't included in all the relevant equations.
- **Data input errors:** If you are fortunate, data input errors will lead to outlier errors; but this is not always the case. For example, if 3,5 is entered instead of 3.5, the spreadsheet assumes it is a string rather than a number. The result is that the value zero is used in any formulas referencing that cell. Another mistake would be to enter 3/5, which becomes a date—a huge number in any calculation.

Introduction

Why spreadsheets are popular

1 • 2 • 3 • 4 • 5
6 • 7 • 8 • 9

An alternative to spreadsheets

The learning curve with IBM
SPSS Statistics

Software licensing
the SPSS way

Conclusion



There are a plenty of other possibilities. For example, if you sort a column that contains both figures and equations, you will sort the equations in addition to the figures, which can result in computation errors.

Spreadsheet errors occur for many reasons, including logic mistakes, incorrect formulas, accidentally overwritten formulas and misuse of built-in functions.

Finally, there is the issue of the reliability of spreadsheet software. Spreadsheet developers continually issues patches and corrections to their software. Several studies show that spreadsheets are not nearly accurate enough for complicated mathematical procedures or large data sets, even when they are coded correctly.



Because it is so widely used, even in the teaching of statistics, numerous articles have detailed the errors in statistical procedures in Excel, and more than a few websites highlight its shortcomings with respect to advanced analytics. (See citations at the end of this paper.) To summarize their findings, “serious analysis of Excel indicates that it is not well suited for large or complex data sets, as it may require you to compromise the accuracy of your results to some degree.”

Other issues related to using spreadsheets

Spreadsheet users may also encounter limitations when it comes to working with special types of data, making predictions and managing data.

Accommodating special types of data

Several types of data that are common to many types of research require special accommodation. One frequently occurring issue is how to handle missing data values. When working with spreadsheets, you have to be careful when dealing with missing values. If you were to assign a zero value to such data, doing so would distort, for example, the average of a range of values. If you enter a string in a cell to indicate a missing value, some equations will ignore the string while others will evaluate it as zero. Because zero is a valid value in some cases, you need another way to indicate a missing value.

Introduction

Why spreadsheets are popular

1 • 2 • 3 • 4 • 5
6 • 7 • 8 • 9

An alternative to spreadsheets

The learning curve with IBM
SPSS Statistics

Software licensing
the SPSS way

Conclusion



When doing this, however, not only must you take care to use the missing value designation consistently when entering data, you must also document your approach carefully so that later modifications to the spreadsheet don't invalidate the data conventions you have used. And note that none of the above approaches provides an approved method for reliably imputing missing values.

When recording survey results using spreadsheets, it can be especially difficult to accurately represent missing or categorical data.

Another special situation arises when dealing with categorical data (often encountered in survey results). For example, suppose the four values 1, 2, 3 and 4 are assigned to represent the answers "Yes," "No," "Don't know" and "Refuse to answer" in a survey question. If you use a spreadsheet to store this type of data, you have to make a special effort to document the values and what they mean, to ensure that the data is entered correctly (the correct value assigned to each answer) and that it is then processed meaningfully. Otherwise, the meaning will be lost as soon as the person who developed the spreadsheet moves on.



Introduction

Why spreadsheets are popular

1 • 2 • 3 • 4 • 5
6 • 7 • 8 • 9

An alternative to spreadsheets

The learning curve with IBM
SPSS Statistics

Software licensing
the SPSS way

Conclusion



Projecting the future

Typically, spreadsheets are used to extract information and relationships about past events. Increasingly, however, organizations want to know what is likely to happen in the future. Recent versions of Excel have dedicated functions—for example, FORECAST, TREND and GROWTH—for predicting new values based on existing data and a range of plug-in programs are also available. What is in question, however, is the reliability and accuracy of these capabilities, which, in any case, provide none of the tests that serious mathematicians would use to check the validity of the results.

Data management

With their cell-level focus, spreadsheets present a number of challenges in managing data. A conceptually simple change—such as modifying a start time, adding new members or changing a formula—can require dozens, even hundreds, of other changes.

Even one simple modification can require cell/row/column insertions or deletions, editing or copying formulas across a range of cells, or re-configuring the entire spreadsheet. These operations are not only time consuming, they can actually lead to more errors.

Almost invariably, new data has to be added to a finished spreadsheet. But how should the new figures be accommodated? One way is to set the spreadsheet to extend equations to include all new data: the problem with this is that some equations could inadvertently be extended to include data that shouldn't be in them. On the other hand, if the sheet is set not to extend automatically, some data that should be included is likely to be left out. Either way, it is unlikely that the spreadsheet will produce correct results unless the changes are checked carefully.

Though some spreadsheet applications contain functions for predicting future trends and outcomes, these methods are often unreliable and inaccurate.

Introduction

Why spreadsheets are popular

1 • 2 • 3 • 4 • 5
6 • 7 • 8 • 9

An alternative to spreadsheets

The learning curve with IBM
SPSS Statistics

Software licensing
the SPSS way

Conclusion



An alternative to spreadsheets

So far, this paper has provided an overview of situations in which spreadsheets may prove inadequate, or at least cumbersome, for statistical analysis. This is not to say that they are without value. If the task is to carry out simple tests on a small number of variables, then a spreadsheet is as good a tool as any.

That said, a spreadsheet program is, as stated previously, general purpose software. With or without plug-ins, the range of analytical tools is limited and the spreadsheet program's algorithms are not as rigorously designed or tested as those in software programs specifically designed for statistical analysis.

IBM SPSS Statistics offers organizations the ability to do robust, in-depth statistical information analysis without doing any programming.

Anyone wishing to do robust and in-depth analysis should use a tool especially built for the job. One of these is IBM® SPSS® Statistics.

SPSS Statistics has been continuously developed and tested since 1968. Over that period, many forms of statistical analysis have been embedded in the software and the algorithms that execute the equations have been tested by both developers and users in academia, in laboratories and in virtually every type of business. As a result, users can have confidence that the software has been thoroughly tested and its results found to be reliable.

Without doing any programming, users can run a very broad range of statistical analyses. In addition, as users' understanding progresses, they are immediately able to apply more advanced methods because they are already there, in the software.



Introduction

Why spreadsheets are popular

An alternative to spreadsheets

1 • 2 • 3 • 4 • 5

The learning curve with IBM SPSS Statistics

Software licensing the SPSS way

Conclusion



Naturally, SPSS Statistics is optimized to handle statistical calculations in a way that a spreadsheet could never be. In fact, the software is optimized for statistical work at every point, from data entry through to the creation of reports for decision makers.

Data entry the SPSS Statistics way

With SPSS Statistics, the data entry process starts with definitions of the data types that are going to be used. These are quite detailed. For example, every data type has both a long and a short name. (The name that is the best fit is the one used to annotate tables and graphs.) In addition, the type of data that can be entered—numbers or text, to give a simple example—can be specified. At this point, the first level of error checking takes place. The data must fit the characteristics of the defined type, or it won't be accepted. Neither the data type nor any other characteristics of the layout can be modified accidentally. Nor can you alter relationships among the data. Data entry is simply data entry: it is in no way intermixed with programming.

The data validation and error checking mechanisms supplied with SPSS Statistics are really quite comprehensive. It allows you to compare two data files or two data sets either by entire document metadata or case-by-case comparison of selected variable values to identify

any discrepancies between them. Automatic procedures locate values that appear to be out of line – which takes care of locating most typos. If, however, the value happened to be within range but somehow abnormal when compared to the other numbers entered, SPSS Statistics would spot it and ask about it.

SPSS Statistics won a 2018 Red Dot Design award for its new user interface.

age	sex	empl.	with.	income	educ.	cred.	othd.	default	prod.	prod.	N	N	O	P	Q	R	S
41	1	27	12	170.00	6.20	11.20	1.01	0	0.00000	0.70000	0.21204						
27	1	10	6	31.00	97.00	1.20	4.00	0	0.00000	0.12000	0.43000						
40	1	10	14	95.00	6.00	9.20	2.27	0	0.02000	0.00000	0.14100						
42	1	12	14	120.00	5.90	2.60	0.01	0	0.02000	0.02000	0.03000						
28	2	2	0	20.00	17.20	1.70	2.00	2	0.78100	0.78100	0.42000						
41	2	5	5	25.00	10.20	0.20	2.10	0	0.13171	0.00000	0.23000						
39	1	20	9	67.00	30.40	0.00	0.01	0	0.00000	0.27000	0.01000						
43	1	11	11	80.00	9.40	0.10	1.10	0	0.01470	0.00000	0.11000						
34	1	3	4	14.00	14.40	1.30	2.20	1	0.74000	0.00000	0.00000						
36	1	0	13	30.00	30.70	2.70	0.10	0	0.00000	0.79123	0.01000						
27	1	0	1	16.00	1.70	0.10	0.00	0	0.00000	0.00000	0.00000						
25	1	4	0	22.00	0.20	0.00	0.04	0	0.22000	0.11000	0.12000						
32	1	14	14	62.00	10.00	0.00	2.47	0	0.00000	0.20000	0.20000						
37	1	4	9	19.00	10.20	1.20	0.01	0	0.34440	0.00000	0.40000						
46	1	22	10	100.00	0.20	0.70	0.40	0	0.01100	0.00000	0.20000						
36	1	0	0	40.00	0.00	0.00	0.00	1	0.00000	0.11000	0.10000						
34	0	13	0	61.00	10.40	0.00	0.01	1	0.11000	0.17000	0.00000						
40	1	27	19	72.00	9.60	1.10	0.20	0	0.01000	0.00000	0.07000						
39	1	6	9	41.00	0.70	0.04	0.01	0	0.00000	0.00000	0.10000						
41	0	0	21	26.00	1.70	0.10	0.00	0	0.00000	0.00000	0.00000						
39	1	22	3	62.00	0.20	1.10	0.01	0	0.00000	0.00000	0.00000						
47	1	17	11	83.00	9.60	0.00	0.01	0	0.00000	0.00000	0.10000						
28	1	7	4	24.00	10.00	0.00	0.10	0	0.25000	0.20000	0.20000						
29	1	9	0	27.00	0.00	0.40	2.24	0	0.00000	0.00000	0.00000						
21	2	1	2	14.00	10.00	0.10	2.64	1	0.63700	0.00000	0.00000						
25	4	0	2	32.00	17.60	2.10	3.49	0	0.00000	0.77000	0.00000						
40	2	9	20	49.00	0.70	0.70	0.01	0	0.00000	0.00000	0.10000						
40	1	0	0	60.00	10.70	0.00	0.74	0	0.00000	0.00000	0.41000						
33	0	11	0	10.00	10.40	0.00	0.00	0	0.25000	0.21000	0.42000						
26	0	0	1	17.00	10.20	0.00	0.00	0	0.00000	0.47000	0.10000						
40	1	1	10	10.00	0.20	0.10	0.10	0	0.00000	0.00000	0.00000						
30	1	1	10	22.00	10.00	1.10	1.17	0	0.42000	0.40000	0.24000						
27	0	2	7	26.00	0.00	0.70	0.00	0	0.00000	0.23000	0.10000						
20	1	0	4	17.00	14.40	1.00	2.07	0	0.23100	0.27000	0.00000						
20	0	0	1	22.00	0.00	0.00	0.00	0	0.00000	0.00000	0.00000						

Introduction

Why spreadsheets are popular

An alternative to spreadsheets

1 • 2 • 3 • 4 • 5

The learning curve with IBM SPSS Statistics

Software licensing the SPSS way

Conclusion



Unlike spreadsheets, SPSS Statistics includes powerful tools for data preparation prior to analysis, including dealing with missing data and checking for outliers that would interfere with analysis.

Preparing data for analysis: the SPSS Statistics approach

As noted above, it often happens that the data available for analysis is incomplete. In a survey, for example, some people may miss or choose not to answer a question. As was pointed out earlier, handling incomplete data in a spreadsheet poses numerous difficulties. With SPSS Statistics, researchers can examine the available data and calculate values for missing items (a process called “imputation”). They can examine data using one of six diagnostic reports to uncover missing data patterns.

Or they can estimate summary statistics and impute missing values, using an automated procedure that chooses the most suitable imputation method based on the characteristics of the data. Then the analysis can be performed just as if all the data were present—which, in a very real and mathematically valid sense, it is.

Other data steps in preparing data for analysis include looking at the distribution of data, checking for outliers and organizing or “binning” data so that the algorithms you plan to use—such as Naïve Bayes or logit models—operate efficiently. SPSS Statistics performs these data preparation steps—something no spreadsheet program is designed to do.

Statistical analysis with SPSS Statistics

When SPSS Statistics switches to analysis mode and performs the actions necessary to produce an output, the data is not modified: it is used only as input to the process, and the output—available in various formats including an impressive array of plots and graphs—is displayed in a separate window.

Data analyzed with SPSS Statistics may be output to various formats including an impressive array of plots and graphs.

In addition, when any type of analysis is performed, the software automatically writes a program in the form of syntax that can be saved and run time after time on different datasets without the need to change it. (It can be changed if required, though.)

Introduction

Why spreadsheets are popular

An alternative to spreadsheets

1 • 2 • 3 • 4 • 5

The learning curve with IBM
SPSS Statistics

Software licensing
the SPSS way

Conclusion



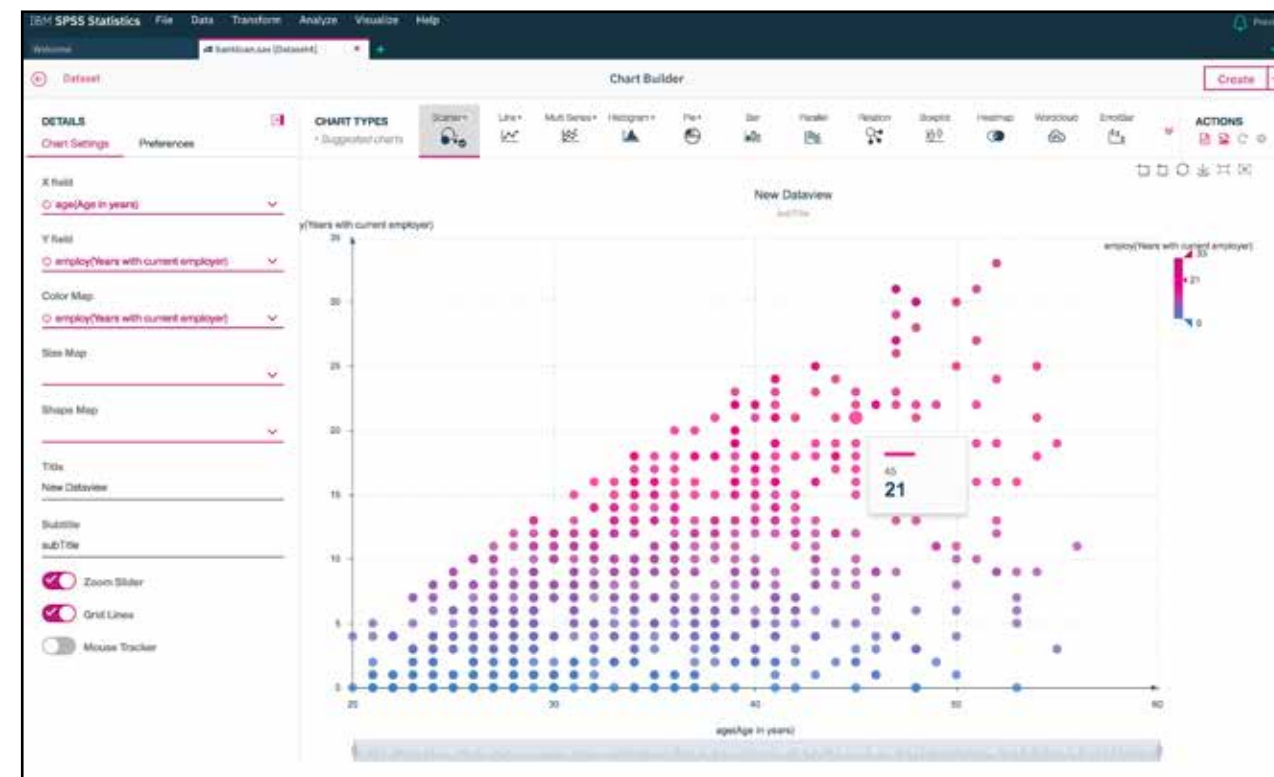
SPSS Statistics also has the advantage of allowing advanced users to implement new procedures and functionality through its Programmability Extension. This advanced feature allows users that are comfortable with the R statistical programming language, Python, .NET or Java, to embed new algorithms or functions directly into the product. They can even create a native GUI for the new feature they've created to give access to it to non-programmers, who can then perform the analyses on their own, rapidly and efficiently.

Present your best results with SPSS Statistics

SPSS Statistics users can create commonly used charts, such as scatterplot matrices (SPLOMs), histograms and population pyramids more easily with the Chart Builder. Those working with statistical process control charts can request rule-checking on both primary and secondary control charts, which provides greater accuracy and a better understanding of whether a process is operating normally. The presentation graphics system in SPSS Statistics provides control at both the creation and edit stages, to help ease workload in a production setting.

Many different types of visualizations can be created, including choropleth maps (color maps), maps with mini-charts and overlay maps, to help users plan, forecast and target more effectively.

You can make your SPSS Statistics output ready for presentation by creating tables directly from procedure dialogs with conditional formatting in place. You can also automate common edits to your output document.



Introduction

Why spreadsheets are popular

An alternative to spreadsheets

1 • 2 • 3 • 4 • 5

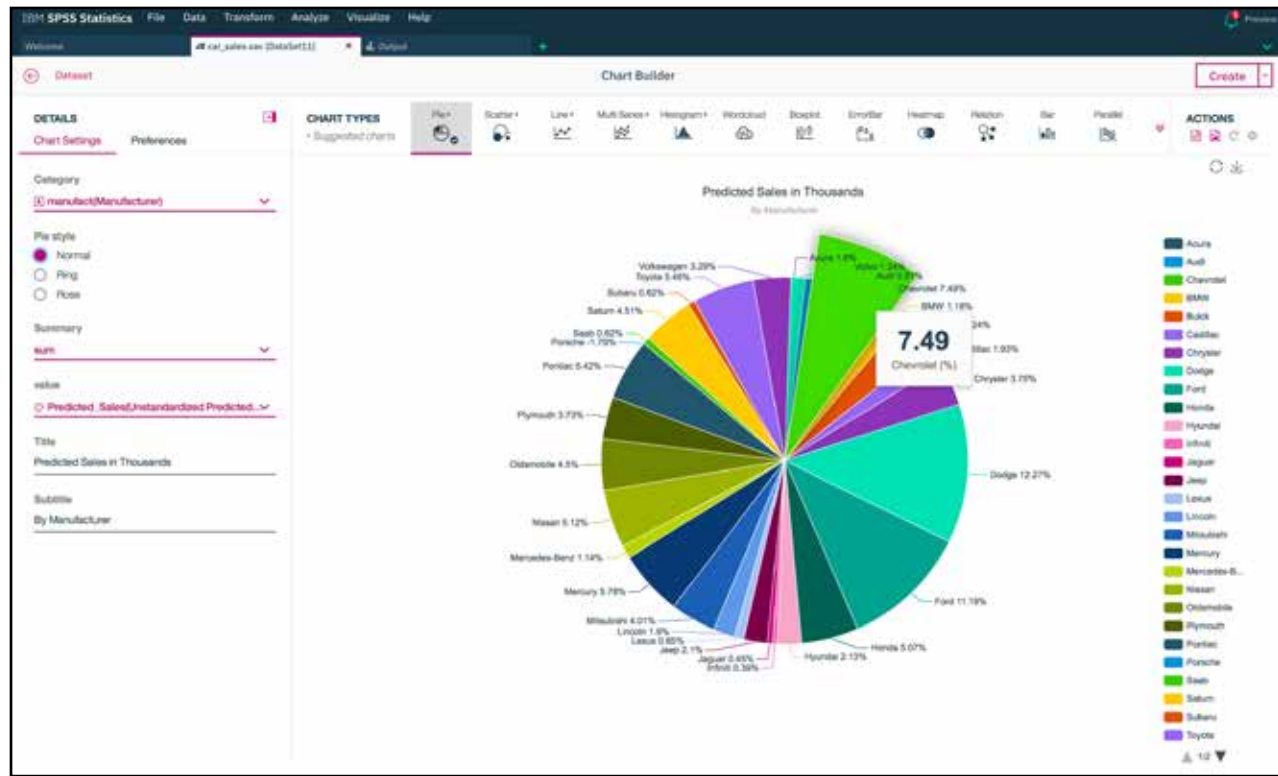
The learning curve with IBM SPSS Statistics

Software licensing the SPSS way

Conclusion



Though spreadsheets can be used to make some kinds of projections, you need a tool such as SPSS Statistics to factor in complex variables or situations.



Charts and tables can be viewed anytime or anywhere on a range of operating systems and desktop and smart devices.

Looking into the future with IBM SPSS Statistics
Spreadsheets are frequently used to make projections—to estimate future events based on historical data or uncertain inputs. For example, a common business application would be to forecast the next two quarters’ revenues based on the previous year’s results. Although it is possible to carry out such a calculation using a spreadsheet, accounting for factors such as seasonality in a business, what-if analysis or developing scenarios based on multiple other variables is only feasible with mathematically robust software such as SPSS Statistics.

Introduction

Why spreadsheets are popular

An alternative to spreadsheets

1 • 2 • 3 • 4 • 5

The learning curve with IBM SPSS Statistics

Software licensing the SPSS way

Conclusion



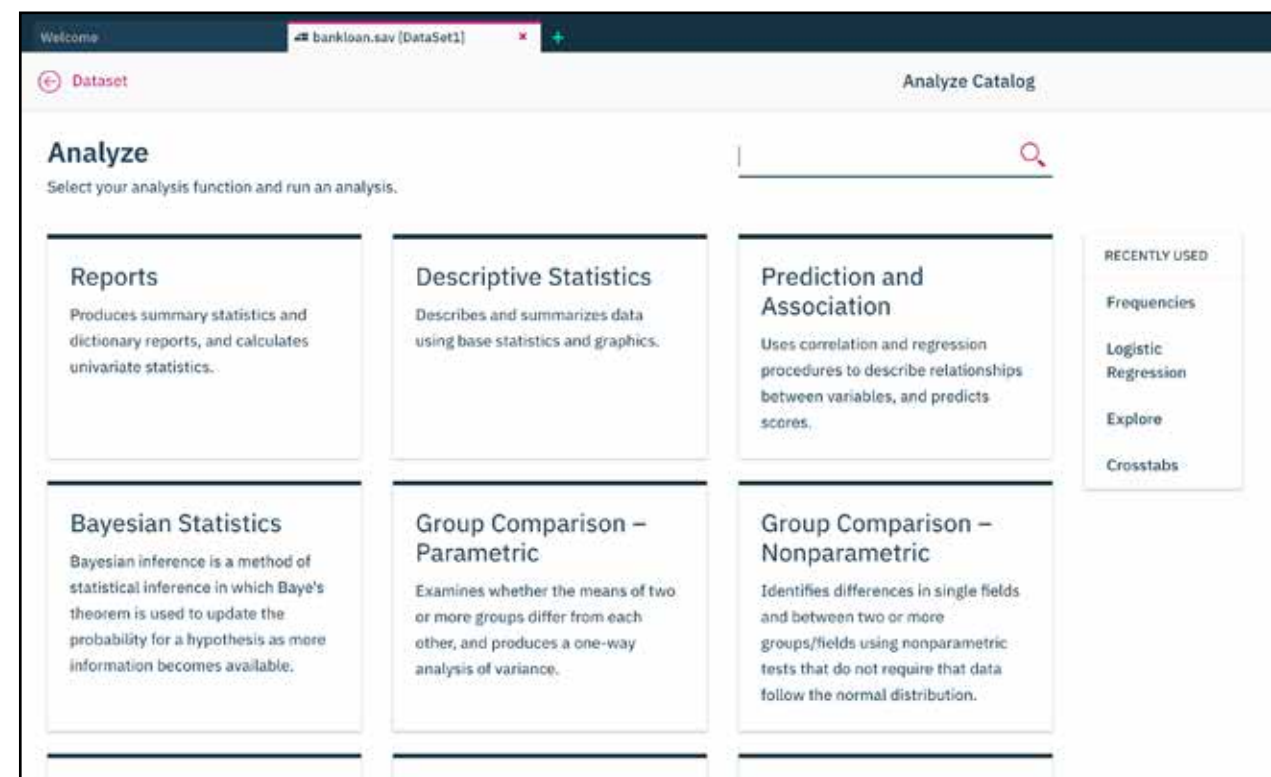
The learning curve with IBM SPSS Statistics

At the beginning of this paper, we posed a question that often rears its head when people are considering using new software: How hard is it to learn?

In the case of SPSS Statistics, the answer is: “Not hard at all.” Like a spreadsheet, it has a WYSIWIG interface, so everything is laid out clearly to view and its features are accessed through familiar Menu and Toolbar layouts. The program’s statistical functions are logically grouped: When one is selected, the relevant options appear in a pop-up window and the calculation is performed by choosing the required options and clicking the “OK” or “Run” button.

In addition, SPSS Statistics comes with very comprehensive tutorials, how-to videos, extremely detailed Help files and case studies. Together, these can take a user from statistical novice to competent analyst fairly quickly. The company, of course, offers a number of training options, including on-demand web-based training. Also, because of its long history of use among analysts in all types of settings, additional learning resources are available from the community, including online discussion boards with tips from other users, instructional books and videos, textbooks and workbooks.

SPSS Statistics is an easy-to-learn tool that helps your organization move beyond spreadsheets to mathematically robust analysis of complex data.



Introduction

Why spreadsheets are popular

An alternative to spreadsheets

The learning curve with IBM SPSS Statistics

Software licensing the SPSS way

Conclusion



Software licensing the SPSS Statistics way



Since different organizations have different preferences and needs when it comes to supporting software installations, SPSS Statistics offers a variety of licensing options to fit any organization. The choices available include perpetual, annual and subscription-based options. That means you can purchase full-time access to the software or month-to-month subscriptions, or both—whichever best meets the needs of your users.

Introduction

Why spreadsheets are popular

An alternative to spreadsheets

The learning curve with IBM
SPSS Statistics

**Software licensing
the SPSS way**

Conclusion



Conclusion

In writing this paper, the author discovered several things worth bringing up at this point: First, spreadsheets are used far more widely than is generally realized—often without seeking out other solutions. Second, the error rate in spreadsheet usage can be higher than the acceptable error rate in other forms of computing. Third, spreadsheets are called upon to address a wide variety of problem types—some of which may not be at all suited to the programs’ capabilities.

Your dataset is unique and so is the way you might use a spreadsheet.

For you to understand whether a spreadsheet is sufficient for your needs or if you might benefit from a specialized tool such as SPSS Statistics, it’s best to see for yourself how each program works with your data, carrying out the analytical tasks you normally require.

It’s easy to test drive SPSS Statistics; you can contact IBM or download a free trial copy of the software at ibm.biz/GetStats. If your data is already in a spreadsheet, SPSS Statistics can easily import it. And once your data is in, you can evaluate the types of analysis available and see if there are advantages, in your case or in some situations, to using a tool designed for statistical analysis, in place of a general- purpose spreadsheet program.

Try SPSS Statistics now

To start your 14-day free trial of SPSS Statistics, visit: ibm.com/spssstatistics/free

To learn more

Please visit ibm.com/products/spss-statistics to learn more about the features of IBM SPSS Statistics and your purchasing options.

Introduction

Why spreadsheets are popular

An alternative to spreadsheets

The learning curve with IBM
SPSS Statistics

Software licensing
the SPSS way

Conclusion



© Copyright IBM Corporation 2018

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the United States of America
November 2018

IBM, the IBM logo, ibm.com, and SPSS are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at ibm.com/legal/copytrade.shtml

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS” WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

- ¹ <https://www.theglobeandmail.com/report-on-business/human-error-costs-transalta-24-million-on-contract-bids/article18285651/>
- ² <https://www.theguardian.com/government-computing-network/2011/jul/01/mi5-data-collection-errors>
- ³ <https://www.bloomberg.com/news/articles/2013-04-18/faq-reinhart-rogooff-and-the-excel-error-that-changed-history>
- ⁴ <https://www.theguardian.com/sport/2012/jan/04/london-2012-olympic-swimming-tickets-oversold>
- ⁵ <https://www.bloomberg.com/news/articles/2013-04-18/faq-reinhart-rogooff-and-the-excel-error-that-changed-history>
- ⁶ <https://www.bbc.com/news/magazine-22223190>
- ⁷ <https://blogs.wsj.com/moneybeat/2014/10/16/spreadsheet-mistake-costs-tibco-shareholders-100-million/>

YTW03240-USEN-04

Introduction

Why spreadsheets are popular

An alternative to spreadsheets

The learning curve with IBM SPSS Statistics

Software licensing the SPSS way

Conclusion
