Guide to Using

StatTools

Statistics Add-In for Microsoft[®] Excel

> Version 5.7 September, 2010

Palisade Corporation 798 Cascadilla St Ithaca, NY 14850 (607) 277-8000 http://www.palisade.com

Copyright Notice

Copyright © 2010, Palisade Corporation.

Trademark Acknowledgments

Microsoft, Excel and Windows are registered trademarks of Microsoft, Inc. IBM is a registered trademark of International Business Machines, Inc. Palisade, TopRank, BestFit and RISKview are registered trademarks of Palisade Corporation.

Welcome to StatTools for Excel

Welcome

StatTools gives Microsoft Excel - the industry-standard data analysis and modeling tool - a new, powerful statistics toolset! StatTools is a Microsoft Excel statistics add-in, allowing you to analyze data in Excel worksheets and work in the familiar Microsoft Office environment. By combining a powerful data manager, along with analyses that rival the best statistics packages available, StatTools brings you the best of two worlds: Microsoft Office ease-of-use and reporting, and robust statistical power.

Work Where You're Comfortable

If you know Excel, you'll know StatTools! StatTools works just as Excel does, with toolbars, menus and custom worksheet functions, all inside of Excel. Unlike stand-alone statistics software, there's no steep learning curve and upfront training costs with StatTools, because you work just as you are used to working in Excel. Your data and variables are in Excel spreadsheets. You can utilize standard Excel formulas for calculations and transformations, along with Excel sorting and pivot tables. Reports and charts from your statistical analyses are in standard Excel format and can utilize all of Excel's built-in formatting capabilities.

Robust Statistics Inside Excel

StatTools replaces Excel's built-in statistics with its own robust and fast calculations. The accuracy of Excel's built-in statistics calculations has often been questioned, and StatTools uses none of them! Even Excel's worksheet statistics functions – such as STDEV() – are replaced by new, robust StatTools versions – such as StatSTDEV(). StatTools statistics calculations meet the highest tests for accuracy, with performance optimized through the use of C++ .DLLs, not macro calculations.

StatTools Analyses

StatTools covers the range of the most commonly used statistical procedures, and offers unprecedented capabilities for adding new, custom analyses. A total of 36 wide-ranging statistical procedures plus 8 built-in data utilities cover the most widely used statistical analyses. Statistical functions provided include descriptive statistics, normality tests, group comparisons, correlation, regression analysis, quality control, forecasts and more. Add to this a library of custom procedures (written by your staff or other experts in the field) and you've got a comprehensive and customizable statistics toolset, right inside of Excel!

StatTools features live, "hot-linked" statistics calculations! If you change a value in Excel, you expect your worksheet to recalculate and give you a new answer. Well, the same thing happens in StatTools! Change a value in your dataset and your statistics report automatically updates. StatTools uses a powerful set of custom worksheet functions to insure that the statistics displayed in your reports are always up-to-date with your current data.

StatTools Data Management

StatTools provides a comprehensive dataset and variable manager right in Excel, just as you would expect from a stand-alone statistics package. You can define any number of datasets, each with the variables you want to analyze, directly from your data in Excel. StatTools intelligently assesses your blocks of data, suggesting variable names and data locations for you. Your datasets and variables can reside in different workbooks and worksheets, allowing you to organize your data as you see fit. Then, you run statistical analyses that refer to your variables, instead of re-selecting your data over and over again in Excel. And StatTools variables aren't limited in size to a single column of data in an Excel worksheet – you can use the same column across up to 255 worksheets for a single variable! (65,535 X 255, or over 16 million cases in StatTools Industrial; 10,000 cases in StatTools Professional Edition)

StatTools Reporting

Excel is great for reports and graphs, and StatTools makes the most of this. StatTools uses Excel-format graphs, which can be easily customized for new colors, fonts and added text. Report titles, number formats and text can be changed just as is any standard Excel worksheet. Drag and drop tables and charts from StatTools reports straight into your own documents in other applications. Charts and tables stay linked to your data in Excel, so whenever your analysis reports change, your document is automatically updated.

Data Access and Sharing

Excel has great data import features, so bringing your existing data into StatTools is easy! Use standard Excel capabilities to read in data from Microsoft SQL Server, Oracle, Microsoft Access, or any other ODBC compliant database. Load data from text files or other applications – if you can read it into Excel, you can use it with StatTools!

StatTools saves all its results and data in Excel workbooks. Just like any other Excel file, you can send your StatTools results and data to colleagues anywhere. Sharing couldn't be easier!

StatTools Industrial

StatTools Industrial includes a complete, object-oriented, programming interface, where custom statistical procedures may be added using Excel's built-in VBA programming language. These custom procedures can utilize StatTools' built-in data management, charting and reporting tools, all accessible via StatTools custom controls, functions and methods. Your custom procedures can even be displayed on the StatTools menu for easy access!

So what if you're not going to write your own statistical procedures? StatTools Industrial still provides a great benefit to you, as you can use custom procedures that are written by others, right off the standard StatTools menu! Experts in the field are constantly adding to the library of new, custom procedures that are built with StatTools. Simply copy a workbook with a new procedure into your StatTools directory on your PC and it instantly shows up on the StatTools menu. Run it and you'll see all the standard StatTools data management tools, combined with the new statistical analysis you need!

Table of Contents

Chapter 1: Getting Started	1
Introduction	3
Checking Your Package What the Package Includes About This Version Working with your Operating Environment If You Need Help StatTools System Requirements	3 3 4 6
Installation Instructions	7
General Installation Instructions Setting Up the StatTools Icons or Shortcuts The DecisionTools Suite	7 8 9
Software Activation	11
Chapter 2: An Overview to StatTools	15
Overview	17
StatTools Menu and Toolbar Data Sets and the Data Manager StatTools Reports and Charts	17 17 20
Chapter 3: StatTools Reference Guide	23
Introduction StatTools VBA Macro Language and Developer's Toolkit	27 27
Reference: StatTools Icons	29
StatTools Toolbar	29
Reference: StatTools Menu Commands	33
Introduction	33

Add-in Analysis Packs	33
Command Listing	35
StatTools Menu – DataSets	43
Data Set Manager Command	43
Data Utilities Menu	49
Stack Command	49
Unstack Command	
Transform Command	
Lag Command	54
Interaction Command	
Combination Command	
Dummy Command	
Random Sample Command	64
Summary Statistics Menu	67
One Variable Summary Command	
Correlations and Covariance Command	70
Summary Graphs Menu	73
Histogram Command	73
Scatter Plot Command	76
Box-Whisker Plot Command	78
Statistical Inference Menu	81
Confidence Interval - Mean/ Std. Deviation Command	81
Confidence Interval - Proportions Command	84
Hypothesis Test - Mean/ Std. Deviation Command	87
Hypothesis Test - Proportions Command	90
Sample Size Selection Command	
One-Way ANOVA Command	
Iwo-way ANOVA Command	
	100
Normality Tests Menu	103
Chi-Square Normality Test Command	
Lilliefors Test Command	107
Q-Q Normal Plot Command	110
Time Series and Forecasting Menu	113
Time Series Graph Command	
Autocorrelation Command	
Runs Test for Randomness Command	118

Forecasting Command	120
Regression and Classification Menu	125
Regression Command Logistic Regression Command Discriminant Analysis Command	126 131 135
Quality Control Menu	139
Pareto Chart Command X/R Charts Command P Chart Command C Chart Command U Chart Command	140 143 147 150 153
Nonparametric Tests Menu	157
Sign Test Command Wilcoxon Signed-Rank Test Command Mann-Whitney Test Command	159 162 165
Utilities Menu	169
Application Settings Command Delete Data Sets Command Clear Dialog Memory Command Unload StatTools Add-in Command	169 175 175 175
Help Menu	177
StatTools Help Online Manual License Activation Command About Command	177 177 177 177
Reference: StatTools Functions	179
Introduction	179
StatTools Functions vs. Excel Functions Distribution Functions "Live" Reports	179 180 181
Reference: Listing of Statistics Functions	183
Table of Available Functions Detailed Function Descriptions	183 187
Index	203

Chapter 1: Getting Started

Introduction	3
Checking Your Package	3
What the Package Includes	3
About This Version	3
Working with your Operating Environment	4
If You Need Help	4
StatTools System Requirements	6
Installation Instructions	7
General Installation Instructions	7
Setting Up the StatTools Icons or Shortcuts	8

Introduction

This introduction describes the contents of your StatTools package and shows you how to install StatTools and attach it to your copy of Microsoft Excel 2000 for Windows or higher.

Checking Your Package

Your StatTools package should contain:

Learning Statistics with StatTools, a statistics book that uses StatTools written by Dr. S. Christian Albright of Indiana University

The StatTools or DecisionTools Suite CD-ROM including:

- StatTools Program
- StatTools Tutorial
- The StatTools Users Guide (this book) in .PDF format

The StatTools Licensing Agreement

If your package is not complete, please call your StatTools dealer or supplier or contact Palisade Corporation directly at (607) 277-8000.

What the Package Includes

StatTools may be purchased on its own and also ships with the DecisionTools Suite Professional and Industrial versions. The StatTools CD-ROM contains the StatTools Excel add-in, several StatTools examples, and a fully-indexed StatTools on-line help system. The DecisionTools Suite Professional and Industrial versions contain all of the above plus additional applications.

About This Version

This version of StatTools can be installed as a 32-bit program for Microsoft Excel 2000 or higher.

Working with your Operating Environment

This User's Guide assumes that you have a general knowledge of the Windows operating system and Excel. In particular:

- You are familiar with your computer and using the mouse.
- You are familiar with terms such as icons, click, double-click, menu, window, command and object.
- You understand basic concepts such as directory structures and file naming.

If You Need Help

Technical support is provided free of charge for all registered users of StatTools with a current maintenance plan, or is available on a per incident charge. To ensure that you are a registered user of StatTools, **please register online at**

http://www.palisade.com/support/register.asp.

If you contact us by telephone, please have your serial number and User's Guide ready. We can offer better technical support if you are in front of your computer and ready to work.

Before Calling Before contacting technical support, please review the following checklist:

- *Have you referred to the on-line help?*
- *Have you checked this User's Guide and reviewed the on-line multimedia tutorial?*
- *Have you read the README.WRI file? It contains current information on StatTools that may not be included in the manual.*
- *Can you duplicate the problem consistently? Can you duplicate the problem on a different computer or with a different model?*
- Have you looked at our site on the World Wide Web? It can be found at http://www.palisade.com. Our Web site also contains the latest FAQ (a searchable database of tech support questions and answers) and StatTools patches in our Technical Support section. We recommend visiting our Web site regularly for all the latest information on StatTools and other Palisade software.

Contacting Palisade

Palisade Corporation welcomes your questions, comments or suggestions regarding StatTools. Contact our technical support staff using any of the following methods:

- *Email us at support@palisade.com.*
- Telephone us at (607) 277-8000 any weekday from 9:00 AM to 5:00 PM, EST. Follow the prompt to reach technical support.
- Fax us at (607) 277-8001.
- *Mail us a letter at:*

Technical Support Palisade Corporation 798 Cascadilla St. Ithaca, NY 14850 USA

If you want to contact Palisade Europe:

- *Email us at support@palisade-europe.com.*
- Telephone us at +44 1895 425050 (UK).
- Fax us at +44 1895 425051 (UK).
- *Mail us a letter at:*

Palisade Europe 31 The Green West Drayton Middlesex UB7 7PN United Kingdom

If you want to contact Palisade Asia-Pacific:

- Email us at support@palisade.com.au
- *Telephone us at* + 61 2 9252 5922 (AU).
- Fax us at + 61 2 9252 2820 (AU).
- *Mail us a letter to:*

Palisade Asia-Pacific Pty Limited Suite 404, Level 4 20 Loftus Street Sydney NSW 2000 Australia

Regardless of how you contact us, please include the product name, version and serial number. The exact version can be found by selecting the Help About command on the StatTools menu in Excel.

Student Versions Telephone support is not available with the student version of StatTools. If you need help, we recommend the following alternatives:

- Consult with your professor or teaching assistant.
- Log on to http://www.palisade.com for answers to frequently asked questions.
- Contact our technical support department via e-mail or fax.

StatTools System Requirements

System requirements for StatTools for Microsoft Excel for Windows include:

- Pentium PC or faster with a hard disk.
- Microsoft Windows 2000 SP4, Windows XP or higher.
- Microsoft Excel 2000 or higher.

Installation Instructions

General Installation Instructions

The Setup program copies the StatTools system files into a directory you specify on your hard disk. To run the Setup program in Windows 2000 or higher:

- 1) Insert the StatTools or DecisionTools Suite CD-ROM in your CD-ROM drive
- 2) Click the Start button, click Settings and then click Control Panel
- 3) Double-click the Add/Remove Programs icon
- 4) On the Install/Uninstall tab, click the Install button
- 5) Follow the Setup instructions on the screen

If you encounter problems while installing StatTools, verify that there is adequate space on the drive to which you're trying to install. After you've freed up adequate space, try rerunning the installation.

Removing StatTools from Your Computer If you wish to remove StatTools from your computer, use the Control Panel's Add/Remove Programs utility and select the entry for StatTools.

Setting Up the StatTools Icons or Shortcuts

Creating the Shortcut in the Windows Taskbar In Windows, setup automatically creates a StatTools command in the Programs\Palisade DecisionTools menu of the Taskbar. However, if problems are encountered during Setup, or if you wish to do this manually another time, follow these directions. Note that the directions given below are for Windows XP Professional. Instructions for other operating systems may vary.

- 1) Click the Start button, and then point to Settings.
- 2) Click Taskbar and Start Menu, and then click the Start Menu tab.
- 3) Click Customize, click Add, and then click Browse.
- 4) Locate the file StatTools.EXE, click it and then click OK.
- 5) Click Next, and then double-click the menu on which you want the program to appear.
- 6) Type the name "StatTools", and then click Finish.
- 7) Click OK on all opened dialogs.

The DecisionTools Suite

StatTools is part of the DecisionTools Suite, a set of products for risk and decision analysis available from Palisade Corporation. The default installation procedure of StatTools puts StatTools in a subdirectory of a main "Program Files\Palisade" directory. This is quite similar to how Excel is often installed into a subdirectory of a "Microsoft Office" directory.

One subdirectory of the Program Files\Palisade directory will be the StatTools directory (by default called StatTools5). This directory contains the Evolver add-in program file (STATTOOLS.XLA) plus example models and other files necessary for StatTools to run. Another subdirectory of Program Files\Palisade is the SYSTEM directory which contains files needed by every program in the DecisionTools Suite, including common help files and program libraries.

Software Activation

Activation is a one time license verification process that is required in order for your StatTools software to run as a fully licensed product. An **activation code** is on your printed/emailed invoice and may resemble a dash separated sequence like "19a0-c7c1-15ef-1be0-4d7f-cd". If you enter your Activation code during installation, then your software is activated the first time the software is run and no further user action is required. If you wish to activate your software after installation, select the StatTools Help menu License Activation command and enter your activation code in the displayed **Palisade License Activation** dialog box.

@RISK Industrial for Excel	<u>P</u> urchase
- Serial Number:	Extend Maintenance
Status: Trial, 17 days remaining	Upgrade
	View EULA
Activate Software Activation ID 19a0-c7c1-15ef-1be0-4d7f-cd	Automatic via Internet Manual via Email
Activate Software Activation ID 19a0-c7c1-15ef-1be0-4d7f-cd	Automatic via Internet Manual via Email
Activate Software Activation ID 19a0-c7c1-15ef-1be0-4d7f-cd Peactivate Software Evifilment ID 1/0	Automatic via Internet Manual via Email Automatic via Internet

Frequently Asked Questions

1) What if my software is not activated?

If you do not enter an activation code during installation or you are installing a trial version, your software will run as a trial version with time and/or number of uses limitations and must be activated with an activation code in order to run as a fully licensed product.

2) How long can I use the product before I have to activate it?

Software that is not activated may be run for fifteen days. All of the product's features are present, but the License Activation dialog will appear each time the program is launched to remind you to activate and to indicate the time remaining. If the 15 day trial period expires, the software will require activation in order to run.

3) How do I check my activation status?

The License Activation dialog box is viewed through the StatTools Help menu License Activation command. Activated software shows a status of **Activated** and trial version software shows a status of **Not Activated**. If the software is not activated, the remaining time that the software is allowed to run is displayed.

4) How do I activate my software?

If you do not have an activation code you may obtain one by clicking the Puchase button in the License Activation dialog. An online purchase will be immediately given an activation code and an optional link to download the installer should reinstallation become necessary. To purchase by phone call the local Palisade office given in the **Contacting Palisade** section of this chapter.

Activation may be done over the Internet or via email:

• Activation if you have Internet Access

In the Palisade License Activation dialog box, type or paste the activation code and press "Automatic via Internet". A success message should appear after a few seconds and the License Activation dialog box will reflect the software's activated status.

• Activation if you do not have Internet Access

Automated activation by email requires a few steps:

- 1. **Click "Manual via Email"** to display the request.xml file which you may save to disk or copy to the Windows clipboard. (It is recommended you note the location on your computer of the request.xml file.)
- 2. **Copy or attach the XML file** to an email and send it to *activation@palisade.com*. You should receive an automatic response to the return address in your email shortly.
- 3. **Save the response.xml attachment** in the response email to your hard drive.
- 4. **Click on the Process button** that is now in the Palisade License Activation dialog box and navigate to the response.xml file. Select the file and click OK.

A success message should appear and the License Activation dialog will reflect the software's activated status.

5) How do I transfer my software license to another machine?

Transfer of a license, or **rehosting**, may be peformed through the Palisade License Activation dialog box as a two step procedure: *deactivation* on the first machine and *activation* on the second machine. A typical use of rehosting is to transfer your copy of StatTools from your office PC to your laptop. To rehost a license from *Machine1* to *Machine2*, make sure both machines have the software installed and are connected to the Internet during the deactivation/activation rehosting.

- 1. On *Machine1*, click deactivate **Automatic via Internet** in the License Activation dialog. Wait for the success message.
- 2. On *Machine2*, click activate **Automatic via Internet**. Wait for the success message.

If the machines do not have Internet access then you may follow the similar instructions above for rehosting by the automated email process.

6) I have Internet Access but I am still unable to Activate/Deactivate automatically.

Your firewall must be set to allow TCP access to the licensing server. For single user (non network installations) ths is http://service.palisade.com:8888 (TCP port 8888 on http://service.palisade.com).

Chapter 2: An Overview to StatTools

Overview	17
StatTools Menu and Toolbar	17
Data Sets and the Data Manager	
Multi-Range Data	19
Stacked and Unstacked Data	19
Treatment of Missing Values	19
StatTools Reports and Charts	20
Use of Formulas Versus Values	20
Use of Comments in Cells	22

Overview

StatTools provides you with powerful statistical capabilities in an environment that you are familiar with - Microsoft Excel. StatTools procedures - such as creating scatterplots, testing a variable for normality, and running a regression analysis - can be run on your data in Excel and the reports and charts from your analyses are created in Excel.

StatTools Menu and Toolbar

Once you have installed StatTools in Excel 2003 and earlier, its menu and commands will be included as part of the Excel menu bar. There will also be a StatTools toolbar displayed. In Excel 2007 and later, a StatTools ribbon will be displayed.

	- (- (B+) +				Boo	1 - Microsof	t Excel							. = X
	nome Insert	Page Layout	Pormulas	Data Ra	niew View	Add-Ins	StatTools							- = x
Data Set Da Manager Utilit	ta bes - A Statistical	Statistics * 🔺 Graphs * 🖄 Inference * 🛃	Normality Tes Time Series as Regression as	ts + nd Forecasting + nd Classification	Quality C ▲ Nonpara	ontrol + meteric Tests +	of Unities -							
Data			An	alyses :			Tools							
A1	• (c	Ja	1			-								¥
A	8 C	D	E	F	G H		1	ĸ	L	M	N	0	P	0
1														
2														
3														
5														
6														
7														
8														
9														
10														
11														
HITH	heet1 /5						14	-		-	-			*
Ready										1		00% 🕣		- 🕀

Data Sets and the Data Manager

StatTools is analogous to most stand-alone statistical software packages in that it is structured around variables. For most analyses it is required that you work with a data set, or a set of statistical variables, often located in contiguous columns with variable names in the first row of the data set. The StatTools **Data Set Manager** allows you to define your data sets and variables. You can then use these predefined variables in your statistical analyses, without re-selecting the data you wish to analyze over and over.

0.	3 . 6	* @*) *	One	Variable	Summary.xls	[Compatibility N	lode] - Mi	crosoft Excel				- 3	n X
9	🔁 Home	Insert Pag	e Layout 🛛 🛛 F	ormulas	Data R	leview View	Add-Ins	StatTools			0	-	a x
Data Se Manag	et Data er Utilities * Data	Summary States	tics * 🚮 No hs * 🔀 Tim ence * 📈 Rep	rmality Test le Series ar gression ar Ani	is * id Forecasting id Classification alyses	 Quality Co Nonparam 	ntrol * eteric Tests *	 Utilities Help * Tools 					
	A14	- (9	and the second	Data Cat	11	Westelle Com							*
A	B	C D	Stat Loois -	Data Set	Manager [U	ne variable sum	imary.xisj			J	K		
1			New	Stacked									
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22	StatTou Use the One salaries of a Running the Note: The Or This example Zappe. Cop S30,200 S30,200 S30,200 S30,200 S33,200 S3,200 S33,	DIS Example Variable Analysis an sample of receivable test generates the state was adapted from D variable Summary was adapted from D variable Summary was adapted from D variable Summary was adapted from D variable Summary M F F M F F M F	Delete Data Set - Name Excel Range Variables Layout: G Excel C B15:82 2 C15:C2	Stacke B14:C Ø Apply Qolumns Data Range 04	d Cell Formatting Cal Formatting Variable Nar Salary Gender	ne Excel Rat ST_Salar ST_Gend	rge Name y_2 er	Multo	tain: Itain: Itain: Drm. Chri	sthe			
23	\$25,900	F	Differichier	100 Data 0	alla Dan Mariahla		_	_					
24 25 26	\$27,900 \$22,800 \$33,800	M M M	2 variables,	190 Data C	ens Per Variable			OK Can	cel				
27	\$22,300	F	Sum Let Quartile		\$2826400.0	0 \$2828400.00 \$27500.00							
29	\$22,900	F	Brd Quartile		\$31900.00	\$32500.00							
30	\$33,200	F	nterquartile R	ange	\$5200.00	\$5000.00							
31	\$30,000	М	1.00%		\$17100.00	\$22400.00							*
H 4 F	H Stacked	Data 🦯 😏 🦯					14	12	-00	0			F
Ready										100% 😑	U		(Ŧ) .::,

Each variable in a data set has a name and a range of Excel cells associated with it. A typical variable layout is **One Variable Per Column**, but variables may also be laid out by row. A data set can include multiple blocks of cells, allowing you to put data on different sheets in the same workbook.

When you are defining a data set, StatTools attempts to identify the variables in a block of cells surrounding the current selection in Excel. This makes it quick and easy to set up a data set with variable names in the top row and variables laid out by column.

The lengths of the columns in your data set do not necessarily have to be equal. For example, you could have two variables, *Weight_Men* and *Weight_Women*, with different numbers of observations. However, for many analyses, StatTools will treat the blank cells in the shorter columns as missing data.

Multi-Range Data	A single column in an Excel 2003 or earlier worksheet can hold up to 65536 data points for a variable. If your variables have more values than this and you choose not to adopt Excel 2007, StatTools allows multiple cell ranges to be assigned to a single data set. For example, you could "repeat" a data set across multiple sheets, assigning the same columns in different worksheets to hold all the values for a data set. You could also use this capability to assign different blocks of cells on the same worksheet to a single data set. This is helpful if your data is scattered about a single worksheet, but you want to combine it all into a single data set.
Stacked and Unstacked Data	StatTools supports both Stacked and Unstacked data. With some statistical procedures it easier to work with Stacked data, and with others, Unstacked data. For example, if we are comparing mean household incomes in several different neighborhoods, then in unstacked form there would be a separate <i>Income</i> variable (or column) for each neighborhood. In stacked form there would be a value variable <i>Income</i> and a category variable <i>Neighborhood</i> indicating which neighborhood each household is in.
	StatTools Variable Stacking utility allows you to "stack" your variables into two columns: a value column, <i>Income</i> , and a category column, <i>Neighborhood</i> . Depending on the type of analysis, the stacked data set might be easier to work with than the unstacked version.
Treatment of Missing Values	If your data set has missing values (a common occurrence in statistical analysis) StatTools deals with them in an appropriate way, depending on the task. For example, summary measures such as means and standard deviations ignore missing values. As another example, a regression analysis involving three variables uses only the rows of the data set that have no missing values for any of the three variables. (This is called "listwise", or "casewise", deletion.) As a third example, a scatterplot of two variables plots only those points where both of the variables have nonmissing values.
	Note: Not all StatTools procedures allow missing values. Check the Reference section of this manual to see how each procedure deals with missing values.

StatTools Reports and Charts

Whenever StatTools creates numerical output, such as a report from a regression analysis or a table of summary statistics, it gives a set of options for the placement of the report. These include:

- In a New Workbook, where a new workbook is created (if necessary) and each report is placed on a sheet in that workbook.
- **In New Worksheet in the Active Workbook**, where each report is placed in a new sheet in the active workbook.
- Starting After Last Used Column, where each report is placed on the active sheet to the right of the last used column.
- **Starting at Cell,** where you have the opportunity to select a cell where the top-left corner of the report or graph will be placed.

Whenever StatTools creates one or more charts, it places them with the reports. Charts are created in Excel format and may be customized using standard Excel chart commands.

Use of Formulas Versus Values

By default, StatTools tries to make the results as "live" as possible. That is, whenever it is practical, reports have formulas that link to the original data. For example, suppose you have a variable *Weight* and you want summary measures on *Weight*, such as its mean and standard deviation. The Summary Statistics procedure names the range of weights as *Weight*, and then it enters formulas in the output cells: **=StatMean(Weight)** and **=StatStdDev(Weight)**. StatMean and StatStdDev are built in StatTools functions for calculating mean and standard deviation. These replace the standard built-in Excel functions for the same statistics.

0	1-2-6-1	⊠ •) •	Bo	ook2 - Micro	soft Excel				-	•	x
C	💛 💽 Home	Insert Page	Layout Formula:	5 Data	Review	View	Add-Ins	StatTools	0 -	5	×
		Summary Statisti Summary Graph	cs * 🚹 Normality '	Fests + and Forecasti	ing ≁ 🔺 N	uality Cor Ionparame	ntrol * eteric Tests *	of Help +	*		
Ma	nager Utilities *	Statistical Infere	nce - 🥂 Regression	and Classifica	ation -						
	Data			Analyses				Tools			
	A1	- (°	fx StatTools								×
	А	В	С	D	E		F	G	Н	2	
1	StatTools	(Core Analysis Pa	ck)								
2	Analysis	One Variable Sun	imary								
3	Performed By:	: Test									
4	Date	: Sunday, February	15, 2009								
5	Updating	: Live									-
0											
0	One Veriable Granman	Salary (F)	Salary (M)								
0	Maan	\$20//11 67	\$20099 26								
10	Variance	14737192 98	12661821.09								
11	Std Day	\$3838 91	\$3558.35								
12	Skewnerr	-0.3034	0.0493								
13	Kurtosis	3,1872	2.4343								
14	Median	\$29700.00	\$29900.00								
15	Mean Abs. Dev.	\$3050.35	\$2893.16								
16	Minimum	\$17100.00	\$22400.00								=
17	Maximum	\$36900.00	\$38200.00								
18	Range	\$19800.00	\$15800.00								
19	Count	96	94								
20	Sum	\$2826400.00	\$2828400.00								
21	1st Quartile	\$26700.00	\$27500.00								
22	3rd Quartile	\$31900.00	\$32500.00								
23	Interquartile Range	\$5200.00	\$5000.00								
24	1.00%	\$17100.00	\$22400.00								
25	2.50%	\$22300.00	\$23400.00								
26	5.00%	\$22800.00	\$23900.00								
27	10.00%	\$24500.00	\$25200.00								
28	20.00%	\$26500.00	\$26900.00								
29	80.00%	\$32400.00	\$33800.00								
30	90.00%	\$34600.00	\$35000.00								
31	95.00%	\$36200.00	\$35900.00								
32	97.50%	\$36400.00	\$36800.00								
33	99.00%	\$36900.00	\$38200.00								-
14	One Var Su	mmary 🧷 💭 🖊			1	(•	I
Rea	dy							100% 😑		(+) .:

Formulas are used in reports for two reasons. First, it helps you to learn statistical procedures and the StatTools functions in Excel. You don't just see a numerical result; you see how it is formed. Second, it has the practical advantage that if your data change, the results change automatically, so that you don't have to rerun the procedure.

There are times when it is not practical to do this. The prime example is regression. StatTools does not provide the formulas that are used to create regression output; it provides only the numerical results. In such cases, if your data change, you will have to rerun the procedures.

StatTools also gives you the option to turn off live updating. This is useful if Excel recalculation time becomes an issue as data changes.

Use of Comments in Cells

A feature of Excel is the ability to include a "pop-up" comment in any cell. You can tell that there is a comment in a cell by noticing a small red triangle in the upper right corner of the cell. You can read the comment by placing the cursor over the cell. StatTools has taken advantage of these comments to insert some context-sensitive help. You can think of these as the most "online" of all online help!

By the way, if you ever have a spreadsheet where the comments won't go away, that is, they always appear in front of your data, select the Tools/Options menu item, click on the View tab, and click on the Comment Indicator Only button.

Chapter 3: StatTools Reference Guide

Introduction	27
StatTools VBA Macro Language and Developer's Toolkit	27
Reference: StatTools Icons	29
StatTools Toolbar	29
Reference: StatTools Menu Commands	33
Introduction	33
Add-in Analysis Packs	33
Command Listing	35
StatTools Menu – DataSets	43
Data Set Manager Command	43
Data Utilities Menu	49
Stack Command	49
Unstack Command	51
	52
Transform Command	
Transform Command Lag Command	
Transform Command Lag Command Difference Command	54
Transform Command Lag Command Difference Command Interaction Command Combination Command	54 56 58 60
Transform Command Lag Command Difference Command Interaction Command Combination Command Dummy Command	54 56 58 60 62
Transform Command Lag Command Difference Command Interaction Command Combination Command Dummy Command Random Sample Command	54 56 60 62 64
Transform Command Lag Command Difference Command Interaction Command Combination Command Dummy Command Random Sample Command Summary Statistics Menu	54 56 58 60 62 64 64
Transform Command Lag Command Difference Command Interaction Command Combination Command Dummy Command Random Sample Command Summary Statistics Menu One Variable Summary Command	54 56 58 60 62 64 67

Summary Graphs Menu	73
Histogram Command	73
Scatter Plot Command	76
Box-Whisker Plot Command	78
Statistical Inference Menu	81
Confidence Interval - Mean/ Std. Deviation Command	81
Confidence Interval - Proportions Command	84
Hypothesis Test - Mean/ Std. Deviation Command	90
Hypothesis Test - Proportions Command	93
Sample Size Selection Command	95
One-Way ANOVA Command	95
Two-Way ANOVA Command	98
Chi-square Independence Test Command	100
Normality Tests Menu	103
Chi-Square Normality Test Command	103
Lilliefors Test Command	107
Q-Q Normal Plot Command	110
Time Series and Forecasting Menu	113
Time Series Graph Command	113
Autocorrelation Command	116
Runs Test for Randomness Command	118
Forecasting Command	120
Regression and Classification Menu	125
Regression Command	126
Logistic Regression Command	131
Discriminant Analysis Command	135
Quality Control Menu	139
Pareto Chart Command	140
X/R Charts Command	143
P Chart Command	147
C Chart Command	150
U Chart Command	153
Nonparametric Tests Menu	157
Sign Test Command	159
Wilcoxon Signed-Rank Test Command	162
Mann-Whitney Test Command	165
Utilities Menu	169
Application Settings Command	169
Delete Data Sets Command	175

Clear Dialog Memory Command Unload StatTools Add-in Command	175 175
Help Menu	177
StatTools Help Online Manual License Activation Command About Command	
Reference: StatTools Functions	179
Introduction	179
StatTools Functions vs. Excel Functions	179
Reference: Listing of Statistics Functions	
Table of Available Functions Detailed Function Descriptions	
Introduction

The StatTools Reference Guide chapter describes the icons, commands, and statistics functions used by StatTools. This chapter is divided into four sections:

- 1) Reference: StatTools Icons
- 2) Reference: Summary of StatTools Procedures
- 3) Reference: StatTools Commands
- 4) Reference: StatTools Functions

StatTools VBA Macro Language and Developer's Toolkit

StatTools also includes a powerful VBA-based macro language which can be used for:

- 1) Automating StatTools analyses
- 2) Developing new statistical analyses which use the StatTools Data Set Manager, reports and graphs. These custom calculations can provide analyses not found in the built-in StatTools procedures. These analyses can show up on the StatTools menu and toolbars if desired.

For more information on StatTools VBA Macro Language and Developer's Toolkit, see the on-line documentation supplied with the product.

Reference: StatTools Icons

StatTools Toolbar

StatTools icons are used to define data sets and variables and then run statistical procedures on those variables. StatTools icons appear on the Excel toolbar (i.e., as a custom toolbar in Excel) in Excel 2003 and earlier and on a ribbon in Excel 2007. This section briefly describes each icon, outlining the functions they perform and the menu command equivalents associated with them.

The following icons are shown on the StatTools toolbar in Excel 2003 and earlier.

lcon	Function Performed and Command Equivalent
Ê	Define a data set and variables, or edit or delete an existing data set and variables
	Command equivalent: Data Set Manager command
₽\$	Run a data utility
	Command equivalent: Data Utilities command
\overline{x}	Run a summary statistics procedure
	Command equivalent: Summary Statistics command
alle.	Create summary graphs for variables
	Command equivalent: Summary Graphs command
	Run a statistical inference procedure
	Command equivalent: Statistical Inference command
	Run a normality test on variables
	Command equivalent: Normality Tests command

謎	Run a time series or forecasting procedure
	Command equivalent: Time Series & Forecasting command
/	Run a regression or classification procedure
	Command equivalent: Regression & Classification command
<u></u>	Run a quality control procedure
	Command equivalent: Quality Control command
	Run a nonparametric test
	Command equivalent: Nonparametric Tests command
4	Display StatTools Utilities
	Command equivalent: Utilities commands
	Display StatTools help file
	Command equivalent: Help command

The following icons are shown on the StatTools ribbon in Excel 2007.

lcon	Function Performed and Command	Equivalent

			SIZ.	
		Ŧ	219	
		=	Ħ	
	_		_	
D	at	a l	Se	t
		_		
IV	lar	۱a	ge	er

Define a data set and variables, or edit or delete an existing data set and variables

Command equivalent: Data Set Manager command



Command equivalent: Data Utilities command

Run a data utility

E Summary Statistics *	Run a summary statistics procedure
	Command equivalent: Summary Statistics command
Lummary Graphs *	Create summary graphs for variables
	Command equivalent: Summary Graphs command
Statistical Inference +	Run a statistical inference procedure
	Command equivalent: Statistical Inference command
A homality Tests *	Run a normality test on variables
	Command equivalent: Normality Test command
The Series and Forecasting 1	Run a time series or forecasting procedure
	Command equivalent: Time Series & Forecasting command
Engenier int Christiane -	Run a regression or classification procedure
	Command equivalent: Regression & Classification command
C Quality Control -	Run a quality control procedure
	Command equivalent: Quality Control command
Nonparameteric Tests *	Run a nonparametric test
	Command equivalent: Nonparametric Tests command
of Utilities 🔹	Display StatTools Utilities
	Command equivalent: Utilities commands
🕜 Help 🔹	Display StatTools help file
	Command equivalent: Help command

Reference: StatTools Menu Commands

Introduction

This section of the Reference Guide details the available StatTools commands as they appear on the StatTools menu in Excel 2003 or earlier and on the StatTools ribbon in Excel 2007 and later. Commands are discussed as they appear on the menu, starting with the Data Set Manager command and subsequently moving down. StatTools icons can be used to perform many of the available commands. The **Reference: StatTools Icons** section of this chapter gives the command equivalents for each **StatTools** icon.

Microsoft Excel - Book1				
In Die Edit wen peser Format India Bata Worden Die Gel 2, on 2, 17 12, k in 25 - √	StatTools Help Qata Set Manager Data Utilities	z 😡 📄 į Arial	- 10 - B Z U E E	oe a question for help 🕴 . 🖷 🗙 E 🖩 🕮 I S 🐁 I 🔟 - 🔍 📲
HID A B C D E 1 6 6 5 6 7 8 6	Summary Statistics Summary Graphs Statistical Inference Simulary Trais Ime Series & Forecasting Bergression & Classification Quality Control Ngroparametric Tests	JKI	. M N	0 P Q 7
10 11 12 13 14 15 16 16 16 16 16 16 16 16 16 16	Ullites	¢		
EDcen * R AutoShapes * N N 🗆 O 🖆 🗐 📲 🛟 🤱 Ready			1.1	

Several **StatTools** commands are also available in a pop-up floating menu that is displayed when the right mouse button is clicked in Excel.

Add-in Analysis Packs

StatTools Industrial can use "analysis packs", or add-in packages of new analyses not found in the core StatTools product. Commands found in these packages will be added to the StatTools menu. Thus, if you are using add-in analysis packs your menu may not look exactly as is shown here. For more information on developing add-on procedures for StatTools, see the on-line documentation for the StatTools VBA Macro Language and Developer's Toolkit.

Command Listing

The procedures available in StatTools come in natural groups. For each group there is a item on a StatTools menu. If a group has more than one item, there is a submenu listing the items in this group. This section provides a brief description of each procedure in each group. More detailed information on each procedure is provided in the section of this chapter titled **Reference: StatTools Commands**.

Procedure	Description	Missing Data?	Live vs. Static Reports	Data Requirement	Multi- Range Data?	Invalid Data	# of Vars.
Summary Statistics							
One Variable Summary Command	Generates Summary Statistics including the usual measures, such as average, median, and standard deviation, plus options such as quartiles and percentiles.	Allowed at beginning, middle and end of data	Live	Stacked and unstacked data Up to 16m cases allowed	Yes	Ignored	1-100
Correlations and Covariance Command	Creates a table of correlations and/or covariances for a set of variables that you select.	Allowed at beginning, middle and end of data	Live	Unstacked data only Up to 16m cases allowed	Yes	Ignored	1-250
Summary Graphs							
Histogram Command	Creates a Histogram for each variable you select. It gives you the option of defining the histogram's categories or "bins".	Allowed at beginning, middle and end of data	Partially Live – data changes update graph when data is within the graph's X-axis range	Stacked and unstacked data Up to 16m cases allowed	Yes	Ignored	1-100
Scatter Plot Command	Creates a Scatter plot for each pair of variables you select.	Allowed at beginning, middle and end of data	Live	Unstacked data only Up to 32,000 cases allowed	No	Not allowed	1-10
Box-Whisker Plot Command	Creates a single boxplot (if you select a single variable) or side-by-side boxplots (if you select several variables).	Allowed at beginning, middle and end of data	Live	Stacked and unstacked data Up to 16m cases allowed	No	Ignored	1-10

Procedure	Description	Missing Data?	Live vs. Static Reports	Data Requirement	Multi- Range Data?	Invalid Data	# of Vars.
Statistical Inference							
Confidence Interval - Mean/ Std. Deviation Command	Calculates a confidence interval for the mean and standard deviation of single variables, or the differences between the means for pairs of variables. The confidence intervals can be calculated using a One-Sample Analysis, a Two-Sample Analysis, or a Paired Sample Analysis.	Allowed at beginning, middle and end of data	Live	Stacked and unstacked data Up to 16m cases allowed	Yes	Ignored	1-250 (one- sample analysis) Exactly 2 (two- sample analysis; paired sample analysis)
Confidence Interval - Proportion Command	Analyzes the proportion of items in a sample that belong to a given category (One-Sample Analysis), or compares two samples with regard to the proportion of items in a given category (Two- Sample Analysis).	Allowed at beginning, middle and end of data	Live	Stacked and unstacked data Up to 16m cases allowed	Yes	Ignored	1-250 (one- sample analysis) Exactly 2 (two- sample analysis; paired sample analysis)
<u>Hypothesis Test</u> <u>- Mean/ Std.</u> <u>Deviation</u> <u>Command</u>	Performs hypothesis tests for the mean and standard deviation of single variables, or the differences between the means for pairs of variables. The hypothesis tests can be performed using a One-Sample Analysis, a Two-Sample Analysis, or a Paired Sample Analysis.	Allowed at beginning, middle and end of data	Live	Stacked and unstacked data Up to 16m cases allowed	Yes	Ignored	1-250 (one- sample analysis) Exactly 2 (two- sample analysis; paired sample analysis)
Hypothesis Test - Proportion <u>Command</u>	Analyzes the proportion of items in a sample that belong to a given category (One-Sample Analysis), or compares two samples with regard to the proportion of items in a given category (Two- Sample Analysis).	Allowed at beginning, middle and end of data	Live	Stacked and unstacked data Up to 16m cases allowed	Yes	Ignored	1-250 (one- sample analysis) Exactly 2 (two- sample analysis; paired sample analysis)

Procedure	Description	Missing Data?	Live vs. Static Reports	Data Requirement	Multi- Range Data?	Invalid Data	# of Vars.
Sample Size Command	Determines the sample size (or sample sizes) required to obtain a confidence interval with a prescribed half-length. It does this for confidence intervals for a mean, a proportion, the difference between two means, and the difference between two proportions.	n/a	n/a	n/a	n/a	n/a	n/a
One-Way ANOVA Command	An extension of the two- sample analysis for comparing two population means. It tests whether two or more means are all equal.	Allowed at beginning, middle and end of data	Live	Stacked and unstacked data Up to 16m cases allowed	Yes	Ignored	2-50
Two-Way ANOVA Command	Performs a two-way analysis of variance. This is usually done in the context of an experimental design where there are two "factors" that are each set at several "treatment levels.	Not allowed	Live	Stacked data Up to 16m cases allowed Experiment must be balanced	Yes	Not allowed	2 category variable s, 1 value variable
Chi-square Independence Test Command	Uses a chi-square test to check whether row and column attributes in a contingency table are statistically independent.	No	Live (as long as table size does not change)	n/a	n/a	n/a	n/a

Procedure	Description	Missing Data?	Live vs. Static Reports	Data Requirement	Multi- Range Data?	Invalid Data	# of Vars.		
Normality Tests									
Chi-Square Normality Test Command	Runs a chi-square test of normality for any variable you select.	Allowed at beginning, middle and end of data	Partially Live (bin positioning will not change but occupation and graphs will)	Stacked and unstacked data Up to 16m cases allowed	Yes	Ignored	1		
Lilliefors Test Command	Provides a more powerful test for normality than the chi-square goodness-of-fit test. (More powerful means that it is more likely to detect nonnormality if it exists)	Allowed at beginning, middle and end of data	Live	Stacked and unstacked data Up to 16m cases allowed	Yes	Ignored	1-10		
Q-Q Normal Plot Command	Creates a quantile-quantile (Q-Q) plot for a selected variable. It provides an informal test of normality.	Allowed at beginning, middle and end of data	Live	Stacked and unstacked data Up to 16m cases allowed	No	Ignored	1		
Time Series & Forecasting	Time Series & Forecasting								
Time Series Graph Command	Creates a time series plot of one or more time series variable(s), all on the same chart.	Allowed at beginning, middle and end of data	Live	Unstacked data Up to 32,000 cases allowed	No	Not allowed	1-100		

Procedure	Description	Missing Data?	Live vs. Static Reports	Data Requirement	Multi- Range Data?	Invalid Data	# of Vars.
Autocorrelation Command	Calculates any number of autocorrelations for a time series variable, indicates which (if any) are significantly nonzero, and (optionally) provides a bar chart (called a correlogram) of the autocorrelations.	Allowed at beginning or end of data	Live	Unstacked data Up to 32,000 cases allowed	No	Not allowed	1-10
Runs Test for Randomness Command	Performs a runs test to check whether a variable (usually a time series variable) is random.	Allowed at beginning or end of data	Live	Unstacked data Up to 16m cases allowed	Yes	Not allowed	1 or more
Forecasting Command	Forecasts time series data using the moving averages method, simple exponential smoothing, Holt's exponential smoothing method for capturing trend, and Winters' exponential smoothing method for capturing seasonality.	Allowed at beginning of data only	Live	Unstacked data Up to 32,000 cases allowed	Yes	Not allowed	1 or more
Regression & Classification							
Regression Command	Runs a variety of regression analyses including Simple Multiple, Stepwise, Forward, Backward and Block.	Allowed at beginning, middle and end of data	Static	Unstacked data Up to 16m cases allowed	Yes	Not allowed	1 depende nt; 1-250 indepen dent
Logistic Regression Command	Performs a logistic regression analysis on a data set. This is essentially a nonlinear type of regression analysis where the response variable is binary: 0 or 1.	Allowed at beginning, middle and end of data	Static	Stacked data Up to 16m cases allowed	Yes	Not allowed	1 depende nt; 1-250 indepen dent
Discriminant Analysis Command	Performs a discriminant analysis on a data set. There should be a "category" variable that specifies which of two or more groups each observation is in, plus one or more explanatory variables that can be used to predict group membership.	Allowed at beginning, middle and end of data	Static	Stacked data Up to 16m cases allowed	Yes	Not allowed except in depend ent variable s	1 depende nt; 1-250 indepen dent

Procedure	Description	Missing Data?	Live vs. Static Reports	Data Requirement	Multi- Range Data?	Invalid Data	# of Vars.
Quality Control							
Pareto Chart Command	Produces Pareto chart that allows you to see the relative importance of categorized data.	Allowed at beginning, middle and end of data	Static	Unstacked data	Yes	Ignored	1 category , or 1 value and 1 category
X/R Charts Command	Produces X and R charts that allow you to see whether a process is in statistical control.	Not allowed	Static	Unstacked data Up to 32,000 cases allowed	No	Not allowed	2-25
P Chart Command	Produces P charts that allow you to see whether a process is in statistical control.	Not allowed	Static	Unstacked data Up to 32,000 cases allowed	No	Not allowed	1 variable 1 size variable
C Chart Command	Produces C charts that allow you to see whether a process is in statistical control.	Not allowed	Static	Unstacked data Up to 32,000 cases allowed	No	Not allowed	1
U Chart Command	Produces U charts that allow you to see whether a process is in statistical control.	Not allowed	Static	Unstacked data Up to 32,000 cases allowed	No	Not allowed	1
Non- Parametric Tests							
Sign Test Command	Performs hypothesis tests for the median of a single variable or for the median of differences for a pair of variables	Allowed at beginning, middle and end of data	Live	Stacked and unstacked data Up to 16m cases allowed	Yes	Ignored	1-250 (one- sample analysis) Exactly 2 (two- sample analysis; paired sample analysis)

Procedure	Description	Missing Data?	Live vs. Static Reports	Data Requirement	Multi- Range Data?	Invalid Data	# of Vars.
Wilcoxon Signed-Rank Test command	Performs hypothesis tests as does the Sign Test but assumes that the probability distribution is symmetric	Allowed at beginning, middle and end of data	Live	Stacked and unstacked data Up to 16m cases allowed	Yes	Ignored	1-250 (one- sample analysis) Exactly 2 (two- sample analysis; paired sample analysis)
Mann-Whitney Test Command	Performs a hypothesis test on two samples	Allowed at beginning, middle and end of data	Live	Stacked and unstacked data Up to 16m cases allowed	Yes	Ignored	1-250 (one- sample analysis) Exactly 2 (two- sample analysis; paired sample analysis)
Data Utilities							
Stack Command	Takes a data set with separate variables for each group in separate columns, and allows you to "stack" them into two columns: a "category" column, and a "value" column. Depending on the type of analysis, the stacked data set might be easier to work with than the unstacked version.	Yes – anywhere in variable	Static	Unstacked data only Up to 65535 cases allowed	No	n/a	1-100
Unstack Command	Does the exact opposite of the Stack procedure.	Yes – anywhere in variable	n/a	Stacked data only Up to 16m cases allowed	Yes	n/a	1-32
Dummy Command	Creates dummy (0-1) variables based on existing variables.	Yes – anywhere in variable	Live	Unstacked data only Up to 16m cases allowed	Yes	n/a	1

Procedure	Description	Missing Data?	Live vs. Static Reports	Data Requirement	Multi- Range Data?	Invalid Data	# of Vars.
Combination Command	Creates a new variable from a pair of numeric variables, a category and numeric variable or two category variables	Yes – anywhere in variable	Live	Unstacked data only Up to 16m cases allowed	Yes	Not allowed	2-32 from same dataset
Interaction Command	Creates a new variable by taking the product, sum, average, min, max or min- max range from one or more variables.	Yes – anywhere in variable	Live	Unstacked data only Up to 16m cases allowed	Yes	n/a	2 value vars., or 1 value and 1 category var, or 2 category var
Lag Command	Creates a new lagged variable based on an existing variable. A lagged variable is simply a version of the original variable, "pushed down" by a number of rows equal to the lag.	Yes – anywhere in variable	Live	Unstacked data only Up to 16m cases allowed	Yes	Ignored	1
Transform Command	Applies any of four nonlinear transformations to any variables you select - natural logarithm, square, square root, or reciprocal - to create a new variable	Yes – anywhere in variable	Live or static	Unstacked data only Up to 16m cases allowed	Yes	Ignored	1-100
Difference Command	Creates any number of difference variables from an original variable.	Yes – anywhere in variable	Live	Unstacked data only Up to 16m cases allowed	Yes	n/a	1
Random Sample Command	Allows you to generate any number of random samples from a given data set, where sampling is with or without replacement.	Yes – anywhere in variable	Static	Stacked data only Up to 16m cases allowed	Yes	Ignored	1-32

StatTools Menu – DataSets

Data Set Manager Command

Defines StatTools data sets and variables, or edits or deletes an existing data set and variables

The **Data Set Manager** command allows you to define your data sets and variables. Once data sets and variables are defined, they may be analyzed in StatTools procedures. The Data Set Manager dialog box allows you to add or remove data sets, name a data set, specify the layout of the variables in a data set, and name the variables in a data set.

What Are Data
Sets and
Variables?StatTools is analogous to most stand-alone statistical software
packages in that it is structured around variables. For most analyses
it is required that you work with a data set, or a set of statistical
variables, often located in contiguous columns with variable names in
the first row of the data set. You can then use these predefined
variables when you run statistical analyses, without re-selecting the
data you wish to analyze over and over.

Each variable in a data set has a name and a range of Excel cells associated with it. The selected **Layout** specifies how variables are located within a data set. A typical variable layout is **Columns** with one variable per column, but variables may also be laid out by **Rows**. A data set can include multiple blocks of cells, allowing you to put data on different sheets in the same workbook.

When you are defining a data set, StatTools attempts to identify the variables in a block of cells surrounding the current selection in Excel. This can make it quick and easy to set up a data set with variable names in the top row and variables laid out by column.

The lengths of the columns in your data set do not necessarily have to be equal. For example, you could have two variables, *Weight_Men* and *Weight_Women*, with different numbers of observations. However, for many analyses, StatTools will treat the blank cells in the shorter columns as missing data.

Data Set Manager Dialog Box

StatTools - D	ata Set Ma	anager [One Varia	ble Summary.xls] 🛛 🛛
Ne <u>w</u>	Stacked			
Data Set	Stacked B14:C204	Il Formatting		_ Multiple
Variables	<u>C</u> olumns (<u>R</u> ows	V	Names in <u>F</u> irst Row
Excel Dat B15:B204 2 C15:C204	ta Range 1 4	Variable Name Salary Gender	Excel Range Name ST_Salary_2 ST_Gender	Output Format auto
2 Variables 10	0 Data Cells	Per Variable	-	
2 variables, 19	o Data Cells	Per variable		

The **Data Set** options in the Data Set Manager dialog box include:

- New, Delete adds a new data set, or deletes an existing one.
- Name specifies the name of the data set.
- **Excel Range** specifies the Excel Range associated with a data set. If multiple cell ranges have been assigned to a data set this entry will be prefaced by the label **Multiple**.
- **Apply Cell Formatting** adds a grid and colors that identify your data sets.
- **Multiple** clicking the **Multiple** button in the Data Set Manager dialog box displays the **Multiple Range Selector** dialog. This dialog allows the entry of individual cell ranges that comprise the multiple cell range data set.

Multiple Range Data Sets

- StatTools allows multiple cell ranges to be assigned to a single data set. A multiple range data set can be used when:
- 1) Each variable in a data set has more than 65536 data points (in Excel 2003 or earlier), requiring the data set to extend across multiple worksheets in the same workbook,
- 2) The data for a variable is located in multiple blocks scattered throughout the worksheets in a workbook.

Stat	Tools - Multiple Range Selector		×				
	Data Set Ranges		<u>C</u> lear All				
. •	'Data1'!A:C						
2	'Data2'!A1:C34466		<u>A</u> uto Fill				
3			DEE Coloret				
4			E Select				
5							
6							
7							
8							
9							
10							
11							
12							
13		-					
Secondary Ranges Have Variable Names In First Row							
		ОК	Cancel				

The options in the Multiple Range Selector dialog include:

- Clear All Clears all entered ranges.
- **Auto Fill** Applies the first range entered (in row 1) to all visible worksheets in the active workbook, and enters these *SheetName!CellRange* references in the grid
- **Select** Displays a selector for highlighting a block of cells to be used as a Data Set Range.
- Secondary Ranges Have Variable Names in the First Column (Row) - Multiple range data sets can have variable names labeling each column (or row, depending on the variable layout selected) in each range listed in the dialog, or variable names labeling the column or row in just the first selected range. The first selected range is the range entered in row 1 of the Multiple Range Selector dialog.

The **Variables** options in the Data Set Manager dialog box include:

- **Layout** specifies how variables are structured in the Excel range that holds the data set. The options for Layout include:
 - **Columns**. This is the typical layout where each column in the data set's Excel range has the data for a variable. Often the names of the variables will be entered at the top of each column.



- **Rows**. With this layout, each row of the data set holds the data for a variable. This layout often is used for time series data in Excel.



 Names in First Column (or Row) – select this when you have included the names of the variables in a data set in cells at the top of columns (or in the leftmost cells when variable layout is Rows).

Variables Options

Each row in the grid in the Data Set Manager dialog box lists the variables in a data set, including the name of each variable, the Excel range that holds the data points for a variable, and the Excel range name used to identify the data for the variable in Excel formulas.

- Excel Range Name The shown range name will be used in Excel formulas that are created in StatTools reports and graphs. These formulas allow your reports to be "live" that is, update automatically when a variable's data changes. Having range names that are understandable to you helps to make formulas more readable.
- Output Format specifies the format for values shown for a variable in reports from StatTools analyses. The Auto entry specifies that StatTools will select a "best" format based on the numeric formatting applied to the cells containing the variable's values in Excel. By clicking the arrow next to the Output Format entry, you can alternatively select a specific format to be used:

Number Format						
Output Number Format Type						
Decimal Digits 4						
OK Cancel						

General is equivalent to Excel's General numeric format. **Fixed** displays precision using the entered number of **Decimal Digits**. **Currency** is equivalent to Excel's Currency numeric format, and displays precision using the entered number of **Decimal Digits**.

Note: The desired output format can also be entered directly in the Data Set Manager dialog, using the notation *selectedFormat(#decimalDigits)*, such as *Currency(4)*.

Data Set and Variable Capacities

In a single session, StatTools allows:

- Up to 256 data sets, located in a single workbook.
- Up to 256 variables per data set. All the data for a single data set must be located in the same workbook.
- Up to 16,777,216 data points per variable.

Actual data capacities may be less than shown above depending on the system configuration and version of Excel in use. Specific StatTools analyses may have different limitations. Memory limitations of Excel itself may also affect data capacities.

Note: the Data Set Manager dialog box lists all data sets and variables in the active workbook (this is the workbook listed in the caption of the Data Set Manager dialog). To list data sets in other workbooks, activate the desired workbook in Excel and display the Data Set Manager dialog.

Data Utilities Menu

Stack Command

Converts a set of variables to stacked format from unstacked format

The Stack command allows you to convert data from "unstacked" form, where a data set includes at least two **value** variables, to a "stacked" form, in where the data set includes a **category** variable and a value variable. For example, if we are comparing mean household incomes in several different neighborhoods, then in unstacked form there would be a separate *Income* variable (or column) for each *Neighborhood*. These columns would not need to be of equal length, that is, each neighborhood could have a different sample size. In stacked form there would be a value variable *Income* and a category variable *Neighborhood* indicating which neighborhood each household is in.

Essentially, this procedure allows you to "stack" your variables into two columns: a value column, *Income*, and a category column, *Neighborhood*. Depending on the type of analysis, the stacked data set might be easier to work with than the unstacked version.

	Pm P	R Summary	Statistics *	Normality Test	1.	9	Quality Con	trol =	P Utilities -	
l	Set Data	La Summary	Graphs	Time Series an	d Forecast	ing 🔹 🔺	Nonparame	teric Tests *	W Help *	
a	ger Utilities -	Statistica	Inference -	Regression an	d Classific	ation -			Teste	
	126	•6	f.	: ADI	illites				60.013	
	A	B	C	0	F		F	ß	н	1 1 1
2	eported Fa	mily Incomes	(Unstacked):	~				Reported Far	nily Incomes (Stacke
		and the second	A contraction of	Concernance of the					-	1.0
	ast Side V	Vest Side N	orth Side S	outh Side					Neighborhoo	d Income
	\$43,108	\$28,650	\$46,809	\$73,115					East Side	\$43,107.60
	\$4,524	\$5,046	\$45,812	\$35,747					East Side	\$4,524.44
	\$55,628	52,702	549,132	\$129,974					East Side	555,628,28
	\$46,488	\$175,693	\$15,459	534,920					East Side	546,488.11
	\$80,831	\$120,050	\$39.503	\$113,469					East Side	580,830,63
	545 000	\$129,321	540 000	\$230,111					East Side	556,703.17
	\$65 627	\$107,754	\$43,000	\$104 391					East Side	544,555.54
	200,027	505,053	\$2 1201	5104,361					East Side	500,020,70
	\$10,042	670 704	1251 66	003 0012					East Side	640,002,51
	\$26 312	\$75 454	\$23,115	\$176 962					East Side	\$26 242 72
	\$28,457	\$125,464	\$22 550	\$234 325					East Side	\$38,457,09
	58 241	582 216	\$15 975	0204,020					East Side	58 240 64
	\$52 476	\$81 951	\$716						East Side	\$52 476 04
	\$88 298	582 222	\$27.065						East Side	588 297 65
	\$92 347	\$147,837	\$17 569						East Side	\$92 346 89
	\$59,968	\$147.412	\$2,020						East Side	\$59,968,28
	\$69.574	\$175,167	\$53,883						East Side	\$69 573 56
	\$3.055	\$159,937	\$56,984						East Side	\$3.055.42
	\$28,375	\$187,547	\$29,041						East Side	\$28,375.30
	\$94,943	\$137,448							East Side	\$94,942.85
	\$54,889	\$115,696							East Side	\$54,889.06
	\$86,583	\$97,260	1						East Side	\$86,583.40
	\$56,665								East Side	\$56,665.00
	\$28,078								East Side	\$28,077.61
	\$71,380								East Side	\$71,379,90
									West Side	\$28,649.50
									West Side	\$5,046.37
									West Side	\$2,701.70
									West Side	\$175,693.35
									West Side	\$126,655.85
									West Side	\$129,321.41

Stacked and Unstacked Variables Variables are stacked using the **Stacking Utility** dialog box:

StatTools - Stacking Utility										
Variables (Select Two or More)										
Data Set	a Set #1		_	<u>F</u> ormat						
Nam	e		Address							
East Side		A4:A29								
🔲 West Side		B4:B29								
🔲 North Side		C4:C29								
South Side		D4:D29								
Stacked Variable 1	Names									
<u>C</u> ategory Name	Category									
<u>V</u> alue Name	Value									
0 🖬 🛂			ОК	Cancel						

At least two or more variables must be selected for stacking. The selected data set is always initially treated as unstacked data. Variables can be from different data sets.

The options in the Variable Stacking Utility dialog include:

• Stacked Variable Names – specifies the name of the category and value variables that will comprise the two variable stacked data set. These names will appear at the top of the columns for the category and value variables.

When OK is clicked, the variables are stacked and a new data set is created for the stacked data.

What are Category and Value Variables?	Category and Value variables are required for a stacked data set. The Category variable (sometimes referred to as the "code" variable) is simply a descriptive identifier for a related set of value variable(s). The Category variable is often a text label. Value variables, on the other hand, (sometimes referred to as "measurement" variables), are standard numeric variables which can be analyzed in statistical procedures.
<i>Number of Variables for Analysis in Stacked Format</i>	If a StatTools procedure imposes limits on the number of variables to select for analysis, in stacked format that limit applies to the number of categories in the category variable. In stacked format one typically selects a single category and a single value variable, thereby specifying multiple variables for analysis, one corresponding to each category in the category variable.

Stacking Utility Dialog Box

Unstack Command

Converts a set of variables from stacked format to unstacked format

The Unstack command does the exact opposite of the Stack command. For example, if you start with a category variable *Gender* and a value variable *Weight*, this command unstacks them into separate *Weight_Men* and *Weight_Women* columns.

Unstacking Utility Dialog Box Variables are unstacked using the Unstacking Utility dialog box:

StatTools - Unstacking Utility 🛛 🛛 🔀							
Variables (Select One Category and	One or More Value)						
Data Set Data Set #2	✓ <u>Format</u>						
Cat Val Name	Address						
Neighborhood	H4:H107						
Income	I4:I107						
	OK Cancel						

The selected data set is always initially treated as stacked data. At least two or more variables must be selected for unstacking. One of these variables is identified as the **Category** variable (by checking **Cat**) and one or more variables are identified as **Value** variables (by checking **Val**). Variables can be from different data sets.

When OK is clicked, the variables are unstacked and one or more new data set(s) are created for the unstacked data.

Transform Command

Transforms one or more variables to new variables and values based on an entered transformation function

The Transform command allows you to transform any variable with one of four possible transformations: natural logarithm, square, square root, or reciprocal. In addition, it allows you to enter a formula which will be used to calculate a transformed variable value.

If there are missing values for the variable on which the transformed variable is based, there will be corresponding missing values in the transformed variable.

Variables are transformed using the **Transformation Utility** dialog box:

StatTools - Trans	formatio	n Utility		
Variables (One or M	lore From a	Single Data Set)		
Data Set Data	Set #1		•	<u>E</u> ormat
Nam	e		Address	
East Side		A4:A29		
✓ West Side		B4:B29		
🔲 North Side		C4:C29		
South Side		D4:D29		
Transformation —				
• Function	Log	•	<u>S</u> hift 0	
C F <u>o</u> rmula =	StatLN(vari	able)		
0 🖬 🛂			ОК	Cancel

The selected data set is always unstacked data. Only variables from one data set at a time can be transformed.

Transformation Utility Dialog Box The options in the Transformation Utility dialog box include:

• **Transformation Function** – or the mathematical operation that will be performed on each value for the selected variables when generating the new transformed value. Built-in transformation functions include natural logarithm, square, square root, and reciprocal. A custom **Formula** can also be entered that calculates a new variable value based on a mathematical expression, such as:

(Variable*1.5)^2

Note that in the equation the keyword "Variable" is used as a placeholder for the actual value of the variable to be transformed.

9	3	× •	 Neighbor 	hoodData.xls	[Compatibility N	/lode] - Microsof	t Excel			- =	X
\sim	- 💌 Home	Insert	Page Layou	it Formulas	Data R	eview View	Add-Ins	StatTools	0	- 0	
I	TH TLH	👔 Summa	ry Statistics *	A Normality 1	ests *	Quality Con	trol 👻	🐇 Utilities	*		
II II		🔒 Summa	ry Graphs +	H Time Series	and Forecasting	Nonparame	teric Tests -	🕜 Help 🕆			
Dat	ta Set Data	A Statisti	cal Inference *		and Classification						
Ma	Data	All Statistic	cur interence	No regression	Anaburar			Tools			
<u> </u>	Data	1			Andryses			10013	9		
	E3	- (C	Ĵx.	Log(East Side)						×
	A	В	С	D	E	F	G		Н		-
1	Reported Far	nily Income	es (Unstacke	d):							. П
2	F 011 W			0 1 011	- (F		1				
3	East Side W	est Side I	North Side	South Side	Log(East Side)	Log(West Side)					
4	\$43,100	\$20,000	546,609	\$/3,115 C2C 747	10.67 145453	0 50640540					_
0	94,924	\$5,040	\$45,01Z	\$35,747	0.41/240424	0.020420103					-
7	555,620 CAC 400	\$2,70Z	049,132 016 460	\$125,574	10.32044702	12 07640642					
0	\$90,931	\$126 656	\$10,403	\$113 /69	11 3001113	11 7/022881					
0	\$58 703	\$120,000	\$9.547	\$238 111	10 98024907	11 7700561					
10	\$45,000	\$107 704	\$40,990	\$2,30,111	10.30024307	11 5970709/					
11	\$65 527	\$80 853	\$41,866	\$104,381	11 00021206	11 /0592612					
12	\$58 842	\$84 903	\$3 129	\$88.441	10 98261748	11 3/19262/10					
13	\$12,067	\$72 781	\$23,125	\$139,580	9 398189002	11 19521161					
14	\$26 313	\$25 464	\$47,886	\$176,853	10 1778081/	10 14502471					
15	\$38 457	\$125 144	\$22,550	\$234 325	10 55729842	11 73722094					
16	\$8 241	\$82 216	\$15 975	Q204,020	9.016833511	11 31710069					
17	\$52 476	\$81 951	\$716		10 86811192	11 31387578					
18	\$88,298	\$82 222	\$27 065		11 38846876	11.31718228					
19	\$92 347	\$147,837	\$17,569		11,43330734	11,9038642					
20	\$59,968	\$147,412	\$2,020		11.00157097	11,90098642					
21	\$69,574	\$175,167	\$53,883		11.15013987	12.07349657					
22	\$3,055	\$159,937	\$56,984		8.024671101	11.98253782					
23	\$28,375	\$187,547	\$29,041		10.25327429	12.14178264					
24	\$94,943	\$137,448			11.46103044	11.83099844					
25	\$54,889	\$115,696			10.91306932	11.65872537	·				
26	\$86,583	\$97,260			11.36886334	11.48513882					
27	\$56,665				10.94491198	3					
28	\$28,078				10.24272788	3					
29	\$71,380				11.17577156	5	1				
30							-				
31											×
14 4	Sheet:	1/0/								•	
Rea	dy		Average: 1	0.89466951	Count: 54 Sum: 5	33.8388061	100%	. 😑		(Ð .::

Data Set with New Transformed Variables

Lag Command

Creates a new lagged variable based on an existing variable

The Lag command allows you to create a new lagged variable based on an existing variable. A lagged variable is simply a version of the original variable, "pushed down" by a number of rows equal to the lag. For example, the lag 3 version of sales in November 1998 is sales three months earlier, in August 1998.

Lag Utility Dialog Box Variables are lagged using the Lag Utility dialog box:

StatTools - L	ag Utility			
-Variables (Sele	ect Exactly One) -			
<u>D</u> ata Set	Coca-Cola Data			▼ Eormat
	Name		Address	
Quarter		B13:B73		
Sales		C13:C73		
Options				
Number of La	gs 1			
0 🖬 🛂			ОК	Cancel

The selected data set is always unstacked data. Only one variable at a time can be lagged.

The options in the Variable Lag Utility dialog box include:

• **Number of Lags** – or the number of time periods to lag values when creating the new variables. A new variable is created for each of the lags up to the entered Number of Lags.

Lagged Variable to Right of Source Data Set

F	1 21 6			Forecast.xls [Compatib	ility Mode] - Micros	soft Excel					>
9	Home	Insert	Page Layout	Formulas Data	Review	View	Add-Ins	StatTools		۲	- 7	1
m di		T Summary	Statistics -	Normality Tests *	4	Quality Cor	ntrol *	J Utiliti	es *			
		A Summan	Granbs *	Time Series and Forecasti		Nonnarami	eteric Tests	- R Help				
Data Se	t Data	A Chakinking		Dependent and Classifier	- ·	tonparam	counciles as	() neip				
Manage	er Utilities *	Statistica	in interence • 12	Regression and Classifica	uon •							
	Data			Analyses				10015				
_	D13	- (?	fx									
A	В	С	D	E F			G	Н			J	_
2	StatTo	ols Exar	nple : For	ecast								
3												
1	The StatToo	Is Forecast Ana	alysis allows you t	o project the future values of	a time-series	svariable ba	ased on its v	alues in the pas	it.			
2	In this examp	ple, quarterly sa	ales figures (in mi	llions) of Coca-Cola are show	n. Usingthe	forecasting	ganalysis, w	e can project s	ales figur	res into the		
5	future.											
	This example	e was adapted	from Data Analys	is and Decision Making with N	licrosoft Exc	elby S. Chr	ristian Albrig	ght, Wayne L. W	inston, a	and Christophe		
5	Zappe. Cop	yright 2003 by E	Brooks/Cole Publ	ishing Company.								
1												
1												
0	Quester	Calaa	Land/Calan)									
2	Quarter	5ales	Lag (Sales)	Farman Kara Caratan								
3	01-00	1/34.03	4724.02	Forecasting Constan	ts (Optimiz	£(I)	4 000	-				
4	Q2-86	2244.96	1/34.83	Level (Alpha)			1.000					
5	Q3-86	2533.8	2244.96	Trend (Beta)			0.000					
5	Q4-86	2154.96	2533.8	Season (Gamma)			0.000					
1	Q1-87	1547.82	2154.96			1000						
8	Q2-87	2104.41	1547.82			Es	stimation	Holdouts	•			
9	Q3-87	2014.36	2104.41	Winters' Exponential		-	Period	Period				
0	Q4-87	1991.75	2014.36	Mean Abs Err			120.20	203.39				
1	Q1-88	1869.05	1991.75	Root Mean Sq Err			162.66	218.25				
2	Q2-88	2313.63	1869.05	Mean Abs Per% Err			3.95%	4.1/%				
3	Q3-88	2128.32	2313.63									
4	Q4-88	2026.83	2128.32		Fore	cast and	Original	Observation	S			
5	Q1-89	1910.6	2026.83		I OIC	cust unu	ongina	Observation	3			
6	Q2-89	2331.16	1910.6									
7	Q3-89	2206.55	2331.16	7000.00								
8	Q4-89	2173.97	2206.55	6000.00					12			
.9	Q1-90	2148.28	2173.97	0000.00				1	AA			
0	Q2-90	2739.31	2148.28	5000.00 -			A	NAN	N V			
1	Q3-90	2792.75	2739.31				AN	VYVV	1.			
2	Q4-90	2556.01	2792.75	4000.00 -		F	N	A A		Sales	٦.	
3	Q1-91	2480.97	2556.01			N	1			- Carda		
4	Q2-91	3039.52	2480.97	3000.00 -	al	V					L	
5	Q3-91	3172.12	3039.52	2000 00 A-	M							
6	Q4-91	2879	3172.12	2000.00 7 7	-							
7	Q1-92	2772	2879	1000.00 -								
8	Q2-92	3550	2772									
9	Q3-92	3508	3550	0.00								
	04-92	3243.86	3508	387	989	992	996	996 998	ē			
0												
1	Q1-93	3056	3243.86	2-18	1-1:	3-19	4-15	3-1	1-2			
.0 .1 2	Q1-93 Q2-93	3056 3899	3243.86 3056	01-18 02-18	04-19	02-19 03-19	Q4-18 Q1-18	02-1 03-1 04-1	01-2			
0 1 2	Q1-93 Q2-93	3056 3899 t 🐑	3243.86 3056	02-18	04-19 04-19	02-19 4 03-19	Q4-15	02-1 03-1	01-2		•	I

Difference Command

Creates any number of difference variables from an original variable

The Difference command allows you to create any number of difference variables from an original variable. It is used primarily for time series variables. A variable to be differenced is selected along with the number of differences (usually 1 or 2). The procedure then creates this many new difference variables. Each difference variable contains differences of the selected variable. For example (for monthly data), the March 1997 difference value is the original March 1997 value minus the original February 1997 value. Similarly, the second difference variable (if requested) contains the differences of the first differences.

Differencing is often used in time series analysis when the original variable is not "stationary" through time. For example, a time series with an upward trend is not stationary. Differencing often achieves stationarity. Sometimes second differencing is useful, but it is less common. Third differencing (or differencing beyond the third) is almost never necessary.

Difference variables are created using the **Differences Utility** dialog box:

StatTools - D	ifference Util	ity			×
-Variables (Sele	ect Exactly One) -				
<u>D</u> ata Set	Coca-Cola Data			-	<u>F</u> ormat
	Name		Address		
D Quarter		B13:B73			
Sales		C13:C73			
- Options					
<u>N</u> umber of Di	fferences 1				
0 🖬 🛂			ОК		Cancel

The selected data set is always unstacked data. Only one variable at a time can be used for creating difference variables.

Difference Utility Dialog Box

The options in the Differences Utility dialog box include:

• Number of Differences – or the number of differences to create.

🖬 47 · (H · 🞯 ·) = Forecast.xls [Compatibility Mode] - Microsoft Excel = x 9 Home Insert Page Layout Formulas Data Review Add-Ins StatTools -X View Quality Control * A Utilities 🕱 Summary Statistics * 👫 Normality Tests * \rightarrow 🛔 Summary Graphs 🐐 🔀 Time Series and Forecasting 🛀 🔺 Nonparameteric Tests 👻 (Help Data Set Data Manager Utilities 🛦 Statistical Inference 🛪 📈 Regression and Classification 🛪 Data Analyses D13 f_x + () * A D В C F E G H 23 StatTools Example : Forecast 4 The StatTools Forecast Analysis allows you to project the future values of a time-series variable based on its values in the past 567 In this example, quarterly sales figures (in millions) of Coca-Cola are shown. Using the forecasting analysis, we can project sales figures in to the future This example was adapted from Data Analysis and Decision Making with Microsoft Excel by S. Christian Albright, Wayne L. Winston, and Christoph Zappe. Copyright 2003 by BrooksCole Publishing Company. 8 9 10 11 12 13 14 01-86 1734.83 Forecasting Constants (Optimized) 2244 96 02-86 510.13 Level (Alpha) 14 15 16 17 0 000 03-86 2533.8 288 84 Trend (Beta) Q4-86 2154.96 -378.84 Season (Gamma) 0.000 1547.82 Q1-87 -607.14 18 Q2-87 2104.41 556.59 Estimation Holdouts Period 19 Q3-87 2014.36 -90.05 Winters' Exponential Period 20 21 22 23 24 25 Q4-87 1991.75 -22.61 Mean Abs Err 120.20 203.39 Q1-88 1869 05 -122.7 Root Mean Sq Err 162 66 218.25 02-88 2313 63 444 58 Mean Abs Per% Err 3 95% 4 17% -185.31 Q3-88 2128.32 Q4-88 2026.83 -101.49 Forecast and Original Observations -116.23 Q1-89 1910.6 26 27 28 29 30 31 32 33 Q2-89 2331.16 420.56 Q3-89 2206.55 -124.61 7000.00 Q4-89 2173 97 -32.58 MMM 6000.00 Q1-90 2148 28 -25.69 02-90 2739 31 591 03 5000.00 Q3-90 2792.75 53.44 Q4-90 2556.01 -236.74 4000.00 - Sales Q1-91 2480.97 -75.04 Forecast 34 35 36 37 38 39 Q2-91 3039.52 558.55 3000.00 Q3-91 3172.12 132.6 2000.00 Q4-91 2879 -293.12 01-92 2772 -107 1000.00 778 Q2-92 3550 Q3-92 3508 -42 0.00 40 Q4-92 3243.86 -264.14 Q4-1989 Q3-1998 1986 Q2-1987 Q3-1988 1991 Q2-1992 993 Q4-1994 Q1-1996 1997 666 Q1-2001 41 Q1-93 3056 -187.86 03-1^c 0-7 5 02-1 04-1 42 Q2-93 3899 843 H + + H Forecast 82 100% (+)

Data Set with Difference Variable

Interaction Command

Creates an interaction variable from one or more original variables

The Interaction command allows you to create an interaction variable from one or more original variables. Interaction variables can be created using two numeric variables, one numeric and one category variable and two category variables.

If both variables are numerical (non-categorical), it creates their product. If one variable is numerical and the other is categorical, it creates the products of the numerical variable with each dummy corresponding to the categories of the categorical variable. Finally, if both variables are categorical, it creates products of all pairs of dummies from the two categorical variables.

Interaction Utility Dialog Box

Interaction variables are created using the **Interaction Utility** dialog box:

StatTools - Interaction Uti	ility		×					
Interaction Between Two Numeric Variables								
Data Set Data Set #1			▼ Eormat					
Val Name		Address						
East Side	A4:A29							
West Side	B4:B29							
North Side	C4:C29							
South Side	D4:D29							
0 🖬 🛂		ОК	Cancel					

The selected data set is always unstacked data. One or more variables at a time can be used for creating an interaction variable.

The options in the Interaction Utility dialog box include:

• Interaction Between – selects the type of each variable to be selected; Two Numeric Variables, One Numeric and One Category Variable or Two Category Variables.

How is an Interaction Variable Created?

Data Set with Interaction Variable

An interaction variable is formed from the two variables you select in the dialog. There are three basic options for these two variables. First, they can both be numerical "measurement" variables. Then the interaction variable is their product. Second, one variable can be a numerical "measurement" variable and the other can be a categorical variable. Then StatTools internally creates dummy variables for each category of the categorical variable and multiplies *each* dummy by the numerical variable. Third, both variables can be categorical variables. Then StatTools internally creates dummy variables for each category of each categorical variable and multiplies each dummy for the first by each dummy for the second. For example, if the two categorical variables have 2 and 5 categories, respectively, then StatTools will create 2x5=10 interaction variables.

-													
0	1 1 1	(* • 🔊 •)	• Nei	ghborhoodDa	ata.xls [C	ompatibil	ity Mode] - Micros	oft Excel		-		x
	🤊 💽 Home	e Insert	Page Layout	Formulas	Data	Review	View	Add-Ins	StatTo	ols	0 -	•	×
	ni nui	😨 Summa	ary Statistics *	A Normality 1	Fests *			ty Control *		4 U	tilities *		
		🚮 Summa	ary Graphs *	Time Series	and Fore	casting *	A Nonp	arameteric T	ests -	🕜 H	lelp –		
Da Ma	ta Set Data mager Utilities *	A Statist	ical Inference *		and Class	ification -							
	Data				Analyses					Т	ools		
	E29	- (0	f_x										*
	A	В	С	D		E			F			G	F
1	Reported Fa	mily Incom	es (Unstacke	d):									T
2													
3	East Side V	Vest Side	North Side	South Side	Interaction	on(East S	ide ,Wes	t Side)					
4	\$43,108	\$28,650	\$46,809	\$73,115			1235	011297					
5	\$4,524	\$5,046	\$45,812	\$35,747			2283	1999.87					-
6	\$55,628	\$2,702	\$49,132	\$129,974			1502	91111.6					
0	\$40,400	\$175,693	\$15,459	\$34,920			10227	051230					
0	\$58,703	\$120,000	\$9.547	\$238 111			7591	676941					
10	\$45,000	\$107 794	\$49 880	\$55 385			4850	735781					
11	\$65.527	\$89 853	\$41 866	\$104,381			5887	745950					
12	\$58,842	\$84,903	\$3,129	\$88,441			4995	882828					
13	\$12,067	\$72,781	\$23,115	\$139,580			8782	13781.3					
14	\$26,313	\$25,464	\$47,886	\$176,853			6700	29868.9					
15	\$38,457	\$125,144	\$22,550	\$234,325			4812	2677259					
16	\$8,241	\$82,216	\$15,975				677	7509545					
17	\$52,476	\$81,951	\$716				4300	0459405					
18	\$88,298	\$82,222	\$27,065				7260	0039078					
19	\$92,347	\$147,837	\$17,569				13652	2268520					
20	\$59,968	\$147,412	\$2,020				8840	0041398					
21	\$69,574	\$1/5,16/	\$53,883				1218/	75249 4					
22	\$3,055	0109,937	\$20,904				4000	15340.1					
21	\$94 943	\$137 //8	925,041				13040	672572					
25	\$54,889	\$115 696					6350	470184					
26	\$86,583	\$97,260					8421	065163					
27	\$56,665												
28	\$28,078												
29	\$71,380												
30													-
14	Sheet	t1 🖉 🖊				14		Ш					1
		Av	erage: 56543139	961 Count: 27	7 Sum: 1	.30049E+1	1 00	100%	0]	ŧ	1 .:

Combination Command

Creates a combination variable from one or more original variables

The Combination command allows you to create a combination variable from one or more original variables. Taking the product, sum, average, min, max or min-max range from one or more variables creates a combination variable.

Combination variables are created using the **Combination Utility** dialog box:

StatTools - Combination Utility										
Variables (Sel	Variables (Select Two or More)									
<u>D</u> ata Set	Data Set #1				•	<u>F</u> ormat				
Val	Name			Address						
🔲 Pair		B10:B44								
Husband		C10:C44								
Vife Wife		D10:D44								
	Type									
• Product	.,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,		C <u>M</u> in							
C <u>S</u> um			С Ма <u>х</u>							
C <u>A</u> verage			C <u>R</u> ange	e (Max - Min)						
0 🖬 🛂]			OK		Cancel				

The selected data set is always unstacked data. One or more variables at a time can be used for creating a combination variable.

The options in the Combination Utility dialog box include:

• **Options** – or mathematical operation to be performed on the selected variables when creating the interaction variable. This can be a product, sum, average, min, max or min-max range.

Combination

Utility Dialog

Box

Data Set with Combination Variable

Image: Second secon		1000	- 🛛 - E	₹ C	onfidence Interv	/al1.xls [0	Compatibi	ility Mode]	- Microso	ft Excel	-		x
Pair Husband Wife Product/Husband Wife Product/Husband Wife Product/Husband Wife Product/Husband 10 1 6 7 8 5 40 11 2 7 8 5 40 12 3 8 5 40 13 4 6 4 24 14 5 8 5 40 13 4 6 4 24 14 6 3 18 5 13 4 6 7 4 26 14 5 8 5 40 15 15 6 7 42 18 18 14 5 8 5 40 17 15 6 7 42 18 27 18 5 20 11 6 3 18 20 21 21 25 4 20 21 12 5 4 20 22 13 8 5 <td></td> <td>🐻 Home</td> <td>Insert</td> <td>Page Layou</td> <td>rt Formulas</td> <td>Data</td> <td>Review</td> <td>View</td> <td>Add-Ins</td> <td>StatTools</td> <td>0 -</td> <td>5</td> <td>×</td>		🐻 Home	Insert	Page Layou	rt Formulas	Data	Review	View	Add-Ins	StatTools	0 -	5	×
Data St. Data Manager Utilities - Data Statistical Inference × Regression and Classification × Analyses Nonparameteric Tests × Tools @ Help × Tools E10 × A B C D E F G H J 7 8 9 Pair Husband Wife Product/Husband Wifel 4 18 × × Tools 9 Pair Husband Wife Product/Husband Wifel 4 10 1 J × </td <td>m</td> <td>a ma</td> <td>🕱 Summa</td> <td>ry Statistics 👻</td> <td>A Normality To</td> <td>ests 👻</td> <td></td> <td>Quality</td> <td>Control *</td> <td>🐇 Uti</td> <td>lities *</td> <td></td> <td></td>	m	a ma	🕱 Summa	ry Statistics 👻	A Normality To	ests 👻		Quality	Control *	🐇 Uti	lities *		
Data Statistical Inference × kergerssion and Classification × Analyses Tools E10 ✓ k 18 × A B C D E F G H J 01 1 6 3 18 × × × 01 1 6 3 18 × × × 11 2 7 8 5 40 13 4 6 4 24 12 3 8 5 40 40 40 41 5 8 5 40 13 4 6 4 24 40 7 8 5 40 17 8 5 40 3 16 7 42 18 6 7 42 3 16 17 18 16 19 10 7 5 35 35 30 30			📥 Summa	ry Graphs 👻	Time Series	and Foreca	sting *	📐 Nonpar	ameteric Test	s 🐐 🕜 He	lp *		
Data Analyses Tools E10 ✓ fs 18 ✓ A B C D E F G H J 0 1 6 3 18 ✓ F G H J J 10 1 6 3 18 ✓ F G H J J 11 2 7 8 6 7 18 5 40 12 3 8 5 40 40 15 6 7 6 42 16 7 8 5 40 18 18 9 7 8 5 40 17 8 6 7 42 18 9 7 8 5 40 19 10 7 5 355 20 11 6 3 18 21 12 5 4 20 22 13 8 5 400 23 21 13 8 </td <td>Mana</td> <td>ger Utilities *</td> <td>🛕 Statistic</td> <td>al Inference</td> <td>Regression</td> <td>and Classif</td> <td>fication -</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>	Mana	ger Utilities *	🛕 Statistic	al Inference	Regression	and Classif	fication -						
E10 f_{a} 18 A B C D E F G H J 7 8 C D E F G H J 10 1 6 3 18 18 18 11 2 7 8 56 10 11 2 7 8 56 11 2 7 8 540 13 46 4 24 15 6 7 6 42 16 7 8 5 40 17 8 6 7 42 18 9 7 8 5 10 7 5 35 35 20 11 6 3 18 21 12 5 4 20 22 13 8 5 40 23 14 7 8 56 10 10 11 11 11 11 11 11 11 11 11 11 11 1		Data			A	nalyses				To	ols		
A B C D E F G H I J 7 7 8 56 18 18 18 11 2 7 8 56 11 2 7 8 56 40 13 4 6 4 24 12 3 8 5 40<		E10	- ()	f_x	18								*
7 8 9 Pair Husband Wife Product(Husband Wife) 10 1 6 3 18 11 2 7 8 56 12 3 8 5 40 13 4 6 4 24 16 7 8 5 40 17 8 6 7 42 16 7 8 56 19 10 7 5 35 20 11 6 3 18 9 7 8 56 19 10 7 5 35 21 12 5 4 20 22 13 8 5 40 23 14 7 8 56 24 16 7 6 42 26 17 6 5 30 29 20 9 10 90 30 21 7 9	-	A B	С	D	E		F	G	Н	1	J		-
9 Pair Husband Wife Product(Husband Wife) 10 1 6 3 18 11 2 7 8 56 12 3 8 5 40 13 4 6 4 24 14 5 8 5 40 15 6 7 6 42 16 7 8 5 40 17 8 6 7 42 18 9 7 8 56 19 10 7 5 35 22 13 8 5 40 23 14 7 8 56 24 15 7 6 42 25 16 7 6 42 26 17 6 5 30 29 20 9 10 90 30 31 22 9 6 54 30 32 23 6	8											_	
10 1 6 3 13 11 2 7 8 56 12 3 8 5 40 13 4 6 4 24 14 5 8 5 40 15 6 7 6 42 16 7 8 5 40 17 8 6 7 42 18 9 7 8 56 19 10 7 5 35 20 11 6 3 18 21 12 5 4 20 22 13 8 5 40 23 14 7 8 56 24 16 7 6 42 26 17 6 5 30 29 20 9 10 90 30 31 22 9 6 54 30 32 23 6 5	9	Pair	Husband	Wife I	Product(Husban	d.Wife)							
11 2 7 8 56 12 3 8 5 40 13 4 6 4 24 14 5 8 5 40 15 6 7 6 42 16 7 8 5 40 17 8 6 7 42 18 9 7 8 56 20 11 6 3 18 21 12 5 4 20 22 13 8 56 21 12 5 4 20 22 13 8 56 24 16 7 5 35 25 16 7 6 30 27 18 5 40 20 28 19 6 5 30 30 21 7 9 63 30 31 22 9 6 5 30 <t< td=""><td>10</td><td>1</td><td>6</td><td>3</td><td></td><td>18</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></t<>	10	1	6	3		18							
12 3 0 5 40 13 4 6 4 24 14 5 8 5 40 15 6 7 6 42 16 7 8 5 40 17 8 6 7 42 18 9 7 8 56 20 11 6 3 18 21 12 5 4 20 22 13 8 56 24 16 7 5 35 25 16 7 6 42 23 14 7 8 56 24 15 7 5 35 25 16 7 6 30 27 18 5 30 30 29 20 9 10 90 30 21 7 9 63 31 22 23 6 5 30	11	2	7	8		56							
14 5 8 5 40 15 6 7 6 42 16 7 8 5 40 17 8 6 7 42 18 9 7 8 56 19 10 7 5 35 20 11 6 3 18 21 12 5 4 20 22 13 8 56 24 15 7 5 35 25 16 7 6 42 26 17 6 5 30 27 18 5 4 20 28 19 6 5 30 30 21 7 9 63 31 22 23 6 5 30 33 24 6 4 24 34 25 6 5 30 35 26 8 5 40 35 36 <td>12</td> <td>4</td> <td>6</td> <td>2</td> <td></td> <td>24</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>	12	4	6	2		24							
15 6 7 6 42 16 7 8 5 40 17 8 6 7 42 18 9 7 8 56 19 10 7 5 35 20 11 12 5 4 20 21 12 5 4 20 22 23 14 7 8 56 24 15 7 5 35 25 16 7 6 42 26 17 6 5 30 27 18 5 4 20 28 19 6 5 30 29 20 9 10 90 30 21 7 9 63 31 22 9 6 54 32 23 6 5 30 33 24 6 4 24 34 25 6 5<	14	5	8	5		40							
16 7 8 5 40 17 8 6 7 42 18 9 7 8 56 19 10 7 5 35 20 11 6 3 18 21 12 5 4 20 22 13 8 6 40 23 14 7 8 56 24 15 7 5 35 25 16 7 6 42 26 17 6 5 30 27 18 5 4 20 28 19 6 5 30 29 20 9 10 90 30 21 7 9 63 31 22 9 6 54 32 23 6 5 30 33 24 6 4 24 34 25 6 5 30	15	6	7	6		42							
18 9 7 8 56 19 10 7 5 35 20 11 6 3 18 21 12 5 4 20 22 13 8 5 40 22 13 8 5 40 23 14 7 5 35 24 15 7 5 35 25 16 7 6 42 26 17 6 5 30 27 18 5 4 20 28 19 6 5 30 29 20 9 10 90 30 21 7 9 63 31 22 9 6 54 32 23 6 5 30 33 24 6 4 24 34 25 6 5 30 35 26 8 5 40	16	7	8	5		40							
19 10 7 5 36 20 11 6 3 18 21 12 5 4 20 22 13 8 5 40 23 14 7 8 56 24 16 7 6 42 26 17 6 5 30 27 18 5 4 20 28 19 6 5 30 29 20 9 10 90 30 21 7 9 63 31 22 9 6 54 32 23 6 5 30 33 24 6 4 24 34 26 8 5 40 35 26 8 5 40 36 26 8 5 40 36 26 8 5 40 36 26 8 5 40	18	9	7	8		56							
20 11 6 3 18 21 12 5 4 20 22 13 8 5 40 23 14 7 8 56 24 16 7 6 42 26 17 6 5 30 27 18 5 4 20 28 19 6 5 30 29 20 9 10 90 30 21 7 9 63 31 22 9 6 54 32 23 6 5 30 33 24 6 4 24 34 26 8 5 40 36 26 8 5 40 36 26 8 5 40 37 7 5 35 30 37 7 5 35 4 4 4 5 4 4	19	10	7	5		35							
21 12 5 4 20 22 13 8 5 40 23 14 7 8 56 24 15 7 5 35 25 16 7 6 42 26 17 6 5 30 27 18 5 4 20 28 19 6 6 30 29 20 9 10 90 30 21 7 9 63 31 22 9 6 54 32 23 6 5 30 33 24 6 4 24 34 25 6 5 30 35 26 8 5 40 36 27 9 7 5 35 14 + Paired Sample Auto Two Sample Motors One Sample S 4	20	11	6	3		18							=
23 14 7 8 56 24 15 7 5 35 25 16 7 6 42 26 17 6 5 30 27 18 5 4 20 28 19 6 5 30 29 20 9 10 90 30 21 7 9 63 31 22 9 6 54 32 23 6 5 30 33 24 6 4 24 34 25 6 5 30 35 26 8 5 40 36 27 9 7 5 37 28 7 5 35 4 + Paired Sample Auto Two Sample Motors One Sample S 4	21	12	5	4		20							11
24 15 7 5 36 25 16 7 6 42 26 17 6 5 30 27 18 5 4 20 28 19 6 5 30 29 20 9 10 90 30 21 7 9 63 31 22 9 6 54 32 23 6 5 30 33 24 6 4 24 34 25 6 5 30 35 26 8 5 40 36 27 9 7 5 37 9 7 5 35 4 + Paired Sample Auto Two Sample Motors One Sample S 4	23	14	7	8		56							
25 16 7 6 42 26 17 6 5 30 27 18 5 4 20 28 19 6 5 30 29 20 9 10 90 30 21 7 9 63 31 22 9 6 54 32 23 6 5 30 33 24 6 4 24 34 25 6 5 30 35 26 8 5 40 36 27 9 7 63 37 28i 7 5 35 14 + Paired Sample Auto Two Sample Motors One Sample S 4	24	15	7	5		35							
20 1 0 5 30 27 18 5 4 20 28 19 6 5 30 29 20 9 10 90 30 21 7 9 63 31 22 9 6 54 32 23 6 5 30 33 24 6 4 24 34 25 6 5 30 35 26 8 5 40 36 27 9 7 63 37 28 7 5 35 14 + Paired Sample Auto Two Sample Motors One Sample S	25	16	7	6		42							
28 19 6 5 30 29 20 9 10 90 30 21 7 9 63 31 22 9 6 54 32 23 6 5 30 33 24 6 4 24 34 25 6 5 30 35 26 8 5 40 36 27 9 7 63 37 28 7 5 35 14 + Paired Sample Auto Two Sample Motors One Sample S 4	20	18	5	4		20							
29 20 9 10 90 30 21 7 9 63 31 22 9 6 54 32 23 6 5 30 33 24 6 4 24 34 25 6 5 30 35 26 8 5 40 36 27 9 7 63 37 28 7 5 35 14 + > Paired Sample Auto Two Sample Motors One Sample S 4	28	19	6	5		30							
30 21 7 9 63 31 22 9 6 54 32 23 6 5 30 33 24 6 4 24 34 25 6 5 30 35 26 8 5 40 36 27 9 7 63 37 28 7 5 35 14 + H Paired Sample Auto Two Sample Motors One Sample S 4	29	20	9	10		90							
31 22 5 0 54 32 23 6 5 30 33 24 6 4 24 34 25 6 5 30 35 26 8 5 40 36 27 9 7 63 37 28 7 5 36 14 + H Paired Sample Auto Two Sample Motors One Sample Sile	30	21	7	9		63							
33 24 6 4 24 34 25 6 5 30 35 26 8 5 40 36 27 9 7 63 37 28 7 5 35 It ↓ ▶ Paired Sample Auto Two Sample Motors One Sample Single Single	32	22	9	5		30							
34 25 6 5 30 35 26 8 5 40 36 27 9 7 63 37 28 7 5 35 № 4 → ▶ Paired Sample Auto Two Sample Motors One Sample Sile	33	24	6	4		24							
35 26 8 5 40 36 27 9 7 63 37 28 7 5 36 It ↓ → Paired Sample Auto Two Sample Motors One Sample Site ■	34	25	6	5		30							
37 28 7 5 35 14 + H Paired Sample Auto Two Sample Motors One Sample Si +	35	26	8	5		40							
H 4 + M Paired Sample Auto / Two Sample Motors / One Sample Si 4	37	21	7	5		35							¥
	14 4 1	Paired	Sample Aut	to / Two S	iample Motors	One Sa	mple S			0		>	1

Dummy Command

Creates dummy (0-1) variables based on existing variables

The Dummy command creates dummy (0-1) variables based on existing variables. There are two options:

- You can create a dummy variable for each category of a categorical variable. For example, if you have a categorical variable *Origin* (for automobiles) with categories *US*, *Europe*, and *Asia*, it creates three dummies with variable names *Origin_US*, *Origin_Europe*, and *Origin_Asia*.
- 2) You can create a single dummy variable from a numerical variable, based on a cutoff value you select. For example, if you have a variable *Weight*, you could create dummies for the condition *Weight* <= 160. In this case, a new variable would be added that had the value 0 when *Weight* > 160 and 1 when *Weight* <= 160.</p>

Dummy variables are created using the **Dummy Utility** dialog box:

Dummy Utility Dialog Box

Stat	StatTools - Dummy Variable Utility							
Var	iables (Sel	ect Exactly One)						
<u>D</u> ata	a Set	Data Set #1			-	<u>F</u> ormat		
Cat		Name		Address				
	Pair		B10:B44					
	Husband		C10:C44					
	Wife		D10:D44					
	otions							
0	Create On	e Dummy Variable	for <u>E</u> ach Distir	nct Category				
۲	Create a Single Dummy Variable from Numeric Variable							
<u>C</u> o	ndition:	< • / /						
0				ОК		Cancel		

The selected data set must be unstacked data. One variable at a time can be used for creating dummy variables.
The options in the Dummy Utility dialog box include:

• **Options** – selects the manner in which dummy variables will be created – either 1) a dummy variable for each category of a categorical variable or 2) single dummy variable from a numerical variable. The **Cutoff value** specifies the cutoff to use when assigning a numeric variable to a 0-1 dummy variable.





Random Sample Command

Generates any number of random samples from selected variables

The Random Samples command allows you to generate any number of random samples from selected variable(s). You specify the number of samples and the sample size for each sample, and StatTools generates the samples from the selected variables. Multiple variables may be sampled independently or dependently, and sampling may be done with or without replacement.

Random Sample Utility Dialog Box

Random Samples are generated using the **Random Sample Utility** dialog box:

StatTools - Random Sampl	e Utility 🛛 🔀
Variables (Select One or More)	
Data Set Data Set #1	▼ <u>F</u> ormat
Name	Address
East Side	A4:A29
West Side	B4:B29
North Side	C4:C29
South Side	D4:D29
r	
Options Number of Samples	Sample with <u>R</u> eplacement
Sample Size 20	Sample Multiple Variables Independently
	Include <u>B</u> lanks in Sample
0 🖬 🛂	OK Cancel

The selected data set is always unstacked data. One or more variables at a time can be used for generating random samples.

The options in the Random Sample Utility dialog box include:

- Number of Samples and Sample Size. The selected number of samples will be generated for each selected variable, and each sample will have a number of elements equal to the sample size.
- Sample with Replacement. Indicates that a value "goes back" to the original population after it is sampled, allowing it to be sampled again. Otherwise, if Sampling with Replacement is not selected (i.e., Sampling Without Replacement is used) a value is not returned and cannot be sampled again.
- Sample Multiple Variables Independently. Selects to have an independent draw used for each sampled value for each variable. Otherwise, for each sampled value, the same sampled index (a number between 1 to # of values in the variable) will be used for all variables.

9	К но	ome Inser	t Page Li	ayout i	Formulas	Data	Review	View	Add-Ins	StatTool	s 🕜 –	⊐ x
Dat Mar	a Set Data nager Utilitie	a es v	immary Statis immary Grap atistical Infer	stics * 🔏 hs * 🔀 rence * 🔀	Normality Time Serie Regression	Tests ▼ s and Fo n and Cli	recasting * assification	N N	uality Control onparameteri	↓ c Tests ↓		Ť
	Data		6	f. Eac	+ Sido (1)	Analyse	5				Tools	×
	AI	•	0	Jx Eas	t Side (1)		E	~	H			~
	A	D		U	E		F	G	п		J	-
1	Last Side	VVest Side	North Side	(1)	0							
2	4524.437	00000 24	40011.07									
1	60573 56	175167 3	53882 03									
4	38457 00	125144 1	225/0 07									
6	65526 7	80852 63	41866 45									
7	26312 73	25/6/ 1	47886 15									
8	58703 17	129321 4	9546 604									
9	12066 51	72781 1	23115 28									
10	46488 11	175693.4	15459 41									
11	58842 37	84902 81	3129.054									-
12	43107.6	28649.5	46808.76									
13	92346.89	147836.8	17569.43									
14	8240.642	82215.63	15974.75									
15	28375.3	187546.6	29041.25									
16	59968.28	147412	2020.283									
17	55628.28	2701.703	49131.74									
18	44999.94	107794.3	49879.77									
19	3055.416	159937.4	56983.64									
20	52476.04	81950.92	716.2596									
21	80830.63	126655.8	39502.79									
22												
14 4	► H Rar	ndom Sam	ple / Shee	t1 / 🔁	/		14		00			
					6.	11		(m)	100%	0	U	(+) .

Generated Random Samples

Summary Statistics Menu

The commands on the Summary Statistics Menu allow you to calculate several numerical summary measures for single variables or pairs of variables. Note that there are no contingency tables in StatTools' procedures. Excel already provides this capability with pivot tables.

One Variable Summary Command

Calculates summary statistics for variables

The **One Variable Summary** command provides summary data for any number of selected numerical variables. These include the mean, median, standard deviation, variance, minimum, maximum, range, first quartile, third quartile, interquartile range, mean absolute deviation, skewness, kurtosis, count, sum, and selected percentiles.

One Variable Summary Statistics Dialog Box This analysis is set up using the **One Variable Summary Statistics** dialog box:

StatTools - One-Variable Summary Statistics											
Variables (Select One Cate	ory and One Value) —										
Data Set Stacked		_	Eormat								
Cat Val Name		Address									
🗖 🗹 Salary	B15:B204										
Gender	C15:C204										
- Summary Statistics to Boo	ort										
		_									
Mean	Minimum	V Other P	ercentiles:								
Variance	Maximum	1.00%	95.00%								
Standard Deviation	Range	2.50%	97.50%								
Kewness	Count	5.00%	99.00%								
I Kurtosis	I∕ Sum	10.00%									
E		20.00%									
Median	First Ouartile	80.00%									
Mean Abs. Deviation	M Third Ouartile	90.00%									
I™ Mode	I Interguartile										
0 2 2		ОК	Cancel								

One or more variables can be selected for analysis. The selected data set can be stacked or unstacked data. Variables can be from different data sets.

The options in the One Variable Summary Statistics dialog include:

• **Summary Statistics to Report** – Selects the desired statistics to include in the report. Percentiles can be added by typing in the desired values.

The One Variable Summary Report uses StatTools Stat functions (such as **StatSkewness**) to allow hot-linking to data. The report is placed in the location specified using the Settings command.

🖳 Home Ir	isert Page Lay	out Formulas	Data	Review	View	Add-Ins	StatTools		-
T I	Summary Statisti	ics 🐐 🚹 Normalii	ty Tests 👻		Qu	ality Contro	•	🐇 Utiliti	es -
	Summary Graph	s * 🛛 🕅 Time Ser	ies and Fore	casting *	🔺 No	nparameteri	c Tests *	🕜 Help	Ŧ
Data Set Data	Statistical Infere	nce - 📈 Regressi	on and Clas	sification	-				
Data		<u></u>	Analyses					Tools	5
A1	→ (°	fx StatTools							
A	В	С	D		E		F	G	
StatTools	(Core Analysis Pa	ck)							
Analysis	One Variable Sum	nmary							
Performed By	: Test								
Date	: Sunday, February	15, 2009							
Updating	: Live								
7	r-1- (r)	C-1 (0.0)							
One Variable Summary	Salary (F)	Salary (IVI)							
Mean	\$29441.67	\$30089.36							
Variance	14737192.98	12661821.09							
1 Std. Dev.	\$3838.91	\$3558.35							
2 Skewness	-0.3034	0.0493							
3 Kurtosis	3.1872	2.4343							
4 Median	\$29700.00	\$29900.00							
5 Mean Abs. Dev.	\$3050.35	\$2893.16							
6 Minimum	\$17100.00	\$22400.00							
7 Maximum	\$36900.00	\$38200.00							
8 Range	\$19800.00	\$15800.00							
9 Count	96	94							
0 Sum	\$2826400.00	\$2828400.00							
1 1st Quartile	\$26700.00	\$27500.00							
2 3rd Quartile	\$31900.00	\$32500.00							
3 Interquartile Range	\$5200.00	\$5000.00	:						
4 1.00%	\$1/100.00	\$22400.00							
5 2.50%	\$22300.00	\$23400.00							
7 10 00%	\$22800.00	\$25200.00							
8 20 00%	\$26500.00	\$26900.00							
80.00%	\$32400.00	\$33800.00							
90.00%	\$34600.00	\$35000.00							
1 95.00%	\$36200.00	\$35900.00							
2 97.50%	\$36400.00	\$36800.00							
3 99.00%	\$36900.00	\$38200.00							
	1								

One Variable Summary Report

- **Missing Data -** This procedure allows missing data in a casewise manner. That is, for each variable, the missing data for that variable are ignored when calculating the summary measures. (This is Excel's default method anyway. For example, if you use the AVERAGE function on a range, it will average only the *numerical* values in the range.)
- Link to Data All of the summary measures are calculated by formulas that are linked to the data. Therefore, if any of the data change, the summary measures change automatically.

Correlations and Covariance Command

Produces a table of correlations and/or a table of covariances between variables

The **Correlations and Covariance** command produces a table of correlations and/or a table of covariances between any set of selected numerical variables. Because both of these tables are symmetric (e.g., the correlation between X and Y is the same as the correlation between Y and X), you can choose to have (1) only the correlations (or covariances) *below* the diagonal show, (2) only those *above* the diagonal show, or (3) those below *and* above the diagonal show.

This analysis is set up using the **Correlations and Covariance** dialog box:

StatTools - Correlation and Covariance Variables (Select Two or More) Data Set #1 -Format Data Set Address Name Salary B19:B118 Culture C19:C118 Sports D19:D118 Dining E19:E118 Tables to Create Table Structure Table of Correlations Symmetric Table of Covariances C Entries Above the Diagonal C Entries Below the Diagonal Only OK Cancel

Two or more variables must be selected for analysis. The selected data set must be unstacked data. Variables can be from different data sets.

Correlations and Covariance

Dialog Box

The options in the Correlations and Covariance dialog include:

- **Tables to Create** Selects the desired correlation and/or covariance table
- **Table Structure** Specifies the structure of the table(s) to be generated:
 - **Symmetric** correlations (or covariances) below *and* above the diagonal show
 - Entries Above the Diagonal Only correlations (or covariances) *above* the diagonal show
 - Entries Below the Diagonal Only correlations (or covariances) *below* the diagonal show

Correlations and Covariance Report

The Correlations and Covariance Report uses StatTools Stat functions (such as **StatCorrelationCoeff**) to allow hot-linking to data. The report is placed in the location specified using the Settings command.

0	0.6	• 🖗 •) =	Во	ok3 - Microso	ft Excel		-	= x
0	🥙 💽 Home	Insert Page	Layout Form	ulas Data	Review View	Add-Ins Sta	tTools 🥝 –	, a x
Da Ma	ta Set Data inager Utilities * Data	X Summary Sta	itistics * 👫 No aphs * 🔀 Tim ference * 🔀 Re	rmality Tests * ne Series and For gression and Cla Analyses	ecasting * 🔺 N ssification *	Quality Control * Ionparameteric Te:	sts * () Help	ies + +
_	A1	• (9	<i>f</i> _≭ StatTo	ols				*
	А	В	C	D	E	F	G	
2 3 4 5	Analysis: Performed By: Date: Updating:	Correlation and Co Test Sunday, February 1 Live	variance 5, 2009					
6 7 8	Correlation Table	Salary Data Set #1	Culture Data Set #1	Sports Data Set #1	Dining Data Set #1			
9	Salary	1.000	0.506	-0.081	0.558	=		-
10	Culture	0.506	1.000	-0.520	0.170			
11	Sports	-0.081	-0.520	1.000	0.266			
12	Dining	0.558	0.170	0.266	1.000			
14 15	Covariance Table	Salary Data Set #1	Culture Data Set #1	Sports Data Set #1	Dining Data Set #1			
16	Salary	91130278.79	1105845.25	-221238.79	2590600.81	-		
17	Culture	1105845.25	52315.39	-33946.99	18938.02			
18	Sports	-221238.79	-33946.99	81427.23	36829.58			
19	Dining	2590600.81	18938.02	36829.58	236187.67			
20	11123							-
14	Correla	tion and Covar	iance 🦉		14	110		× I
Rea	ady					100% 😑		🕂

- **Missing Data** Missing data are allowed, and they are treated in a pairwise manner. That is, to obtain the correlation (or covariance) between any pair of variables, all cases with missing data on *either* of the two variables are ignored.
- Link to Data The correlations and covariances are calculated by formulas that are linked to the data. Therefore, if any of the data change, these summary measures update automatically.

Summary Graphs Menu

The commands on the Summary Graphs Menu enable you to create charts that are very useful in statistical analysis and not terribly easy (or possible) to produce with Excel's chart wizard. Of course, Excel's charting capabilities are extensive, so StatTools tries not to duplicate things that Excel already does well.

Histogram Command

Creates histograms for variables

The Histogram command creates a histogram for each variable you select. It gives you the option of defining the histogram's categories (often called "bins"), and it shows these clearly on the chart. It also creates a frequency table that each histogram is based on.

Histogram Dialog Box

This graph type is set up using the Histogram dialog box:

Stat	Tools - Histogra	am									
Vari	Variables (Select One or More)										
<u>D</u> ata	Set Unstad	ked			-	<u>F</u> ormat					
	Name		Address								
	Salary		B15:B204								
- Op	tions										
Nu	mber of Bins	auto	-	<u>X</u> -Axis	Numeric	-					
His	togram Minimum	auto	•	<u>Y</u> -Axis	Prob. Dens	ity 💌					
His	togram <u>M</u> aximum	auto	•								
۲				[ОК	Cancel					

One or more variables can be selected for graphing. The selected data set can be stacked or unstacked data. Variables can be from different data sets. The options in the Histogram dialog include:

- Number of Bins. Sets the number of histogram intervals calculated across the range of a graph. The value entered must be in the range 1 to 200. The setting Auto calculates the best number of bins to use for your data based on an internal heuristic.
- **Histogram Minimum**. Sets the minimum value where histogram bins start. **Auto** specifies that StatTools will start the histogram bins based on the minimum of the data graphed.
- **Histogram Maximum**. Sets the maximum value where histogram bins end. **Auto** specifies that StatTools will end the histogram bins based on the maximum of the data graphed.
- **X-axis**. Selects **Categorical** or **Numeric**. A categorical x-axis simply labels each bin with the midpoint of the bin. A numeric x-axis has a "readable" x-axis minimum and maximum and can be rescaled using Excel's standard rescaling options.
- **Y-axis.** Selects **Frequency**, **Rel. Frequency** or **Prob. Density** as the unit of measure reported on the Y-axis. Frequency is the actual number of observations in a bin. Relative Frequency is the probability of a value in the range of a bin occurring (observations in a bin/total observations). Density is the relative frequency value divided by the width of the bin, insuring that Y-axis values stay constant as the number of bins is changed.

Histogram of a Single Variable

0		5-6	- 🔊 -)	÷	Во	ok4 - Microso	oft Excel				-	-	x
	26	Home	Insert	Page Layou	Formu	ılas Data	Review	View	Add-Ins	StatTools	0 -	•	x
	1174	ШЩ	x Summ	ary Statistics	- ANO	rmality Tests 👻		4.0	uality Contro	- I	of Utilit	ies -	
	🛄 🌃 Summary Graphs * 🔀 Time Series and Forecasting * 🔺 Nonparameteric Tests * 🎯											*	
Da	ata Set Data												
IVId	Data Analyses Tools												
	Di	Data Analyses 1001s											
	A1 • StatTools ¥												Ŷ
		A	В		С	D	E		F		G		
1	Sta	tTools	(Core Analys	is Pack)									
2		Analysis:	Histogram										
3	Pe	rformed By:	Test Sundau Enhr	uppy 15, 2009									
5		Updating:	Live	uary 15, 2005									
6													
7						Salary /	Unstacked						
8	Histog	ram	Bin Min	Bi	n Max	Bin Midpoint	Fre	q.	Rel. Freq.	Prb	Density		
9	Bin #1		\$17100.	00 \$19	144.44	\$18272.22	1	\$	0.0053	0.0	000002	-	
10	Bin #2		\$19444.	44 \$21	788.89	\$20616.67	1		0.0053	0.0	000002		
11	Bin #3		\$21788.	89 \$24	133.33	\$22961.11	10	D	0.0526	0.0	000022		
12	Bin #4		\$24133.	33 \$26	177.78	\$25305.56	18	3	0.0947	0.0	000040		
13	Bin #5		\$26477.	78 \$28	322.22	\$27650.00	47	7	0.2474	0.0	000106		
14	Bin #6		\$28822.	22 \$31	166.67	\$29994.44	45	5	0.2368	0.0	00101		
15	Bin #7		\$31166.	57 \$33	511.11	\$32338.89	35	5	0.1842	0.0	000079		
16	Bin #8		\$33511.	11 \$35 56 ¢30	355.56	\$34683.33	23	5	0.1211	0.0	000052		
1/	Bin #9		\$35855.	56 538	200.00	\$3/02/./8	10)	0.0526	0.0	00022		
10													
20				HISLOGIA	1 OI Salary	// Unstacked							
21		0.00012											
22													
23		0.0001 -											
24		1.111											
25		0.00008											
26	usity	0.00000				<u> г</u>							
27	De												
28	file	0.00006 -											
29	obal							1					
30	7	0.00004 -											
31													
32		0.00002 -		-	- 1								
33		0.00002											
25		12											
36		0 +			1				1				
37		1500	0.00 20	00.00	25000.00	30000.00	35000	.00	40000.00				-
14	• •	Histog	ram 🖉	2			1.43						I
Rea	dy								100%	0	U	(+)) .:
-								and the second s	and the second s	-			

- **Missing Data -** Missing data are allowed. All rows with missing data on the selected variables are ignored.
- Link to Data Histograms are partially linked to data. When data changes and new values fall within the range of the bins of the original histogram, the graph will automatically update. Data changes that require re-binning will not update on the graph.

Scatter Plot Command

Creates scatter plots between pairs of variables

The Scatter Plot command creates a scatterplot for each pair of variables you select. Scatter plots can be created with Excel's XY-plot option but Excel automatically puts the first (or leftmost) variable on the horizontal axis, which might not be what you want. StatTools allows you to choose which of the variables will be placed on the horizontal axis. Each scatterplot shows the correlation between the two variables with the corresponding plot.

Scatter Plot Dialog Box This graph type is set up using the **Scatter Plot** dialog box:



Two or more variables can be selected for graphing. At least one Xaxis and one Y-axis variable are required. If more than two variables are selected multiple scatter plots are created. The selected data set must be unstacked data. Variables can be from different data sets. The options in the **Scatter Plot** dialog include:

• **Display Correlation Coefficient**. Specifies that the correlation coefficient between the graphed variables will be displayed



Scatter Plot Example

- **Missing Data -** Missing data are allowed. All rows with missing data on either of the two selected variables in any given pair are ignored.
- Link to Data The scatterplots are linked to the original data. If the data change, so do the scatterplots. However, the scales of the axes might need to be updated manually if the ranges of the selected variables change significantly.

Box-Whisker Plot Command

Creates Box-Whisker plots for variables

The Box-Whisker plot command creates a single Box-Whisker plot (if you select a single variable) or side-by-side Box-Whisker plots (if you select several variables). It also creates a sheet that shows the summary statistics (quartiles, interquartile range, etc.) that are used to form the Box-Whisker plot(s).

Box-Whisker Plot Dialog Box

This graph type is set up using the **Box-Whisker Plot** dialog box:

StatTools - Box-Whisker Plot	X
Variables (Select One Category and Data Set Actor Data	One Value)
Cat Val Name Name Name Gender Salary	Address 'Stacked Data'!B19:B84 'Stacked Data'!C19:C84 'Stacked Data'!D19:D84
Options Image: Construction of the second	ents OK Cancel

One or more variables can be selected for graphing. The selected data set can be stacked or unstacked data. Variables can be from different data sets.

The options in the **Box-Whisker Plot** dialog include:

• **Include Key Describing Plot Elements**. Specifies that a separate key describing chart elements will be displayed below the graph.





- **Missing Data -** Missing data are allowed. All rows with missing data on any of the selected variables for the chart are ignored.
- Link to Data The Box-Whisker plots that are created are linked to the original data; if the data change, so do the Box-Whisker plots. However, the scale of the horizontal axis might need to be updated manually if the scale of the data changes significantly

Statistical Inference Menu

The commands on the Statistical Inference Menu perform the most common statistical inference analyses: confidence intervals and hypothesis tests, along with one and two-way ANOVA.

Confidence Interval - Mean/ Std. Deviation Command

Calculates confidence intervals for mean and standard deviation of variables

The Confidence Interval for Mean/ Std. Deviation command calculates a confidence interval for the mean and standard deviation of single variables, or the differences between the means for pairs of variables. The confidence intervals can be calculated using a **One-Sample Analysis**, a **Two-Sample Analysis**, or a **Paired Sample Analysis**.

This analysis is set up using the **Confidence Interval for Mean/ Std. Deviation** dialog box:

StatTools - Confidence Interve	l for Mean/Std. Deviation 🛛 🛛 🔀
Analysis Type Two-Sample Analys	Dine Value)
Cat Val Name Name Sender Salary	Address B18:B83 C18:C83 D18:D83
Confidence Intervals to Calculate - ✓ For the Difference of Means ✓ For the Standard Deviation	Confidence Level 95% Confidence Level 95% OK Cancel

The number of variables selected depends on the Analysis Type used. A One-Sample Analysis requires one or more variables, while a Two-Sample Analysis and a Paired Sample Analysis require two variables. The selected data set can be stacked or unstacked data. Variables can be from different data sets.

Confidence Interval for Mean/ Std. Deviation Dialog Box The options in the Confidence Interval dialog include:

- **Analysis Type.** Selects the type of analysis performed. Options include:
 - **One-Sample Analysis**. Calculates confidence intervals for a single numerical variable.
 - **Two-Sample Analysis**. Calculates confidence interval for the difference between means from two independent populations.
 - **Paired Sample Analysis**. This is basically the same as the two-sample analysis, but it is appropriate when the two variables are naturally paired in some way. It essentially runs a one-sample analysis on the differences between pairs.
- **Confidence Intervals to Calculate**. Specifies the confidence intervals that will be calculated on the selected variables. Options change based on the analysis type selected:
 - **One-Sample Analysis**. Selects to calculate confidence intervals on the mean and/or standard deviation, and the confidence level (0 to 100%) for each.
 - **Two-Sample Analysis** or **Paired Sample Analysis**. Selects to calculate the confidence interval for the difference between means for two variables, and specifies the confidence level (0 to 100%) desired.

Confidence Interval Report

P . C	- 🖉 -) =	Во	ok7 - Microsof	t Excel		_	-	= x
Home 🕑	Insert Page	e Layout 🛛 Formu	ulas Data	Review Vi	ew Add-Ins	StatTools	- 10	a x
Data Set Data Manager Utilities *	Summary St	atistics * 🥼 No raphs * 🔀 Tim nference * 🔀 Reg	rmality Tests + te Series and For- gression and Clas Anabyses	ecasting *	Quality Contro	ol * ic Tests *	 Utiliti Help 	es * *
0.1	6	f ChatTa	-la				10013	~
AI	•6	Jx Statio	JIS				196.0	~
4	4	В	С	D	E		F	-
1 2 3 4 5	StatTools Analysis: Performed By: Date: Updating:	(Core Analysis Pack Confidence Interva Test Sunday, February 1 Live	k) 11 15, 2009					
6 7 8 Sample Summaries		Salary (F) Actor Data	Salary (M) Actor Data					
9 Sample Size		18	48					
10 Sample Mean 11 Sample Std Dev		3.737	5.509					
13 14 Conf. Intervals (Diff	erence of Means)	Equal Variances	Unequal Variances					
15 Confidence Level 16 Sample Mean Diffe 17 Standard Error of D	rence	95.0% -3.424 1.409296073	95.0% -3.424 1.18671913					
18 Degrees of Freedor 19 Lower Limit	n	64 -6.239003668	45 -5.813786132					
20 Upper Limit 21 22		-0.608218554	-1.03343609					
23 Equality of Variance	es Test							
24 Ratio of Sample Van 25 p-Value 26	riances	0.4602 0.0827						
II + H Confide	ence Interval	2		14				N I
Ready				Æ	100%	0	Ū	+

- **Missing Data -** Missing data are allowed. All rows with missing data on the selected variables are ignored.
- Link to Data All of the reports are calculated with formulas that are linked to the data. If the values of the selected variable change, the outputs change automatically.

Confidence Interval - Proportion Command

Calculates confidence intervals for proportions

The Confidence Interval for Proportion command allows you to analyze the proportion of items in a sample that belong to a given category (**One-Sample Analysis**), or to compare two samples with regard to the proportion of items in a given category (**Two-Sample Analysis**). There are three Data Types supported by this procedure: **Population Sample, Summary Table with Counts**, and **Summary Table with Proportions**.

This analysis is set up using the **Confidence Interval - Proportions** dialog box:

StatTools - C	onfidence Interv	al for Propo	rtion		X				
<u>A</u> nalysis Type	One-Sample Analys	iis	•						
<u>D</u> ata Type	Summary Table with	h Counts	•						
Select One Variable with Categories and One or More Variables with Counts									
<u>D</u> ata Set	Simple-Counts			•	<u>F</u> ormat				
Cat Ct	Name		Address		Ĩ				
🗹 🗖 Cate	gory	B18:B20							
Coun	t	C18:C20							
Categories to	Analyze (Select One	or More) ——							
Democrat	t								
	ก								
3 categories in	n selected variable.								
Options									
Confidence Lev	/el 95%	•							
<u>F</u> irst Sample Si	ze	150							
Second Sample	Size	150							
0 2 4			ОК		Cancel				

The number of variables selected depends on the Analysis Type used. A **One-Sample** analysis requires one or more samples for analysis; a **Two-Sample** analysis requires two samples. One selects variables with sample information using column labeled **Ct** (Count), % (Proportion), or column with no label for the **Population Sample** Data Type. (If the Population Sample data is stacked, the samples are selected in columns labeled **C1** and **C2**, where C1 contains the stacked

Confidence Interval for Proportion Dialog Box categories.) If the data is in the form of a table with counts or proportions, there is an additional **Cat** column for selecting one variable with category names.

With the **Population Sample** Data Type variables can be from different data sets.

The options in the **Confidence Interval** dialog include:

- **Analysis Type.** Selects the type of analysis performed. Options include:
 - **One-Sample Analysis**. Calculates confidence intervals for the proportion of items in a sample that belong to a given category.
 - **Two-Sample Analysis**. Calculates confidence interval for two samples with regard to the proportion of items in a given category
- Data Type. Specifies the type of data to be analyzed, either Population Sample, Summary Table with Counts, or Summary Table with Proportions.
- **Options**. Options available change with the Analysis Type and Data Type. Options include:
 - **Confidence Level**. Selects the confidence level (0 to 100%) for the analysis.
 - **First Sample Size and Second Sample Size**. For the **Summary Table with Proportions Data Type**, selects the size of the first sample and second sample (Two-Sample Analysis only).

Confidence Interval Report

0		•	Bo	ok3 - Mi	crosoft Exc	el			-		x
	💛 💽 Home Insert	Page Layou	ut Formulas	Data	Review	View	Add-Ins	StatTools	0	- 🕫	×
Da Ma	ta Set Data nager Utilities - Statis	nary Statistics * nary Graphs * stical Inference *	Normality Tess	ts ▼ nd Forecast nd Classific	ting * 🔺 I	Quality Cor Nonparam	ntrol * eteric Tests *	 Utilities Help + 			
	Data	5	An.	alyses				lools	9		×
	AI V	Jx	Statiools			-	-				*
	А	в	C	D		E	F	G		1	-
5 6 7	Updating	Live	Count								-
9	Category	Democrat	Republican								
10	Sample Size	1040	1040								
11	Sample Proportion	0.538	0.433								
12	Confidence Level	95.0%	95.0%								
13	Standard Error of Proportion	0.015	0.015								_
14	Lower Limit	0.508	0.403								
15 16 17	Upper Limit	0.569	0.463								
Rez	dv	. interval / t					(min)	100%		G	. (I
								100010 G			

- **Missing Data -** Missing data are allowed. All rows with missing data on the selected variables are ignored.
- Link to Data All of the reports are calculated with formulas that are linked to the data. If the values of the selected variable change, the outputs change automatically.

Hypothesis Test - Mean/ Std. Deviation Command

Performs a hypothesis test for mean and standard deviation of variables

The Hypothesis Test for Mean/ Std. Deviation command performs hypothesis tests for the mean and standard deviation of single variables, or the differences between the means for pairs of variables. The hypothesis tests can be performed using a **One-Sample Analysis**, a **Two-Sample Analysis**, or a **Paired Sample Analysis**.

Hypothesis Test for Mean/ Std. Deviation Dialog Box This analysis is set up using the **Hypothesis Test for Mean/ Std. Deviation** dialog box:

StatTools - Hypothesis Tes	t for Mean/Std. Deviation 🛛 🛛 🔀
Analysis Type One-Sample An	nalysis 💌
Variables (Select One or More)	
Data Set Part Data	✓ Eormat
Name	Address
Part #	B18:B107
 Hypothesis Tests to Perform 	
Mean	
<u>N</u> ull Hypothesis Value	10
Alternative <u>Hypothesis</u>	Not Equal to Null Value (Two-Tailed Test)
Null Hunsthesis Value	0.01
Null Hypothesis Value	
Alternative Hypothesis	Not Equal to Null Value (Two-Tailed Test)
Image: A state of the state	OK Cancel

The number of variables selected depends on the Analysis Type used. A One-Sample Analysis requires one or more variables, while a Two-Sample Analysis and a Paired Sample Analysis require two variables. The selected data set can be stacked or unstacked data. Variables can be from different data sets. The options in the Hypothesis Test dialog include:

- **Analysis Type.** Selects the type of analysis performed. Options include:
 - **One-Sample Analysis**. Performs hypothesis tests for a single numerical variable.
 - **Two-Sample Analysis**. Performs hypothesis tests for the difference between means from two independent populations.
 - **Paired Sample Analysis**. This is basically the same as the two-sample analysis, but it is appropriate when the two variables are naturally paired in some way. It essentially runs a one-sample analysis on the differences between pairs.
- Hypothesis Tests to Perform. Specifies the hypothesis tests that will be performed on the selected variables. Options change based on the analysis type selected. A One-Sample Analysis selects hypothesis tests on the mean and/or standard deviation. A Two-Sample Analysis or Paired Sample Analysis selects hypothesis tests for the difference between means for two variables. For each hypothesis test selected, options include:
 - **Null Hypothesis Value**, or the value of the population parameter under the null hypothesis.
 - Alternative Hypothesis Type, or the alternative to the Null Hypothesis Value that will be evaluated during the analysis. The Alternative Hypothesis Type can be either "one-tailed" (that is, greater or less than the null hypothesis) or "two-tailed" (that is, not equal to the null hypothesis).

Hypothesis Tests Report

0		;	Book8 - Mic	rosoft Excel				x
	🕈 📧 Home Insert	Page Layout	Formulas Da	ta Review	View Add-In	s StatTools	0 - =	x
Da Ma	ta Set Data inager Utilities - Data	ny Statistics * 👔 ny Graphs * 🛃 al Inference * 🛃	Normality Tests Time Series and Regression and Analy	▼ Forecasting ▼ Classification ▼ rses	Quality Contro	ol • 4	Utilities * Help * Tools	
	A1 🗸 💽	<i>f</i> _≭ Sta	tTools					×
	А	В	С	D	E	F	G	
1 2 3 4 5	StatTools Analysis: Performed By: Date: Updating:	(Core Analysis Pao Hypothesis Test Test Sunday, February Live	:k) 15, 2009					
6 7 8 9	Hypothesis Test (One-Sample)	Width Part Data 90	-					
10	Sample Mean	9.999256						
11	Sample Std Dev	0.009728						
12	Hypothesized Mean	10						
13	Alternative Hypothesis	<>10						
14	Standard Error of Mean	0.001025464						=
15	Degrees of Freedom	0 7260						
10	t-Test Statistic	-0.7200						
10	p-Value	0.4098						
10	Null Hypoth. at 10% Significance	Don't Reject						
20	Null Hypoth, at 5% Significance	Don't Reject						
21	Hypothesized Std Dev	0.01						
22	Alternative Hynothesis	<> 0.01						
23	Ratio of Std Devs	0.9728						
24	Degrees of Freedom	89						
25	Chi-square Test Statistic	84.2312						
26	p-Value	0.7139						
27	Null Hypoth. at 10% Significance	Don't Reject						
28	Null Hypoth. at 5% Significance	Don't Reject						
29	Null Hypoth. at 1% Significance	Don't Reject						
30		-						
31	-							-
14	Hypothesis Test	2		14	luna III.			
Rea	adv				田田田 100	%	0 (÷
						0		J 199

- **Missing Data -** Missing data are allowed. All rows with missing data on the selected variables are ignored.
- Link to Data All of the reports are calculated with formulas that are linked to the data. If the values of the selected variable change, the outputs change automatically.

Hypothesis Test - Proportion Command

Performs a hypothesis test for proportions

The Hypothesis Test for Proportion command analyzes the proportion of items in a sample that belong to a given category (**One-Sample Analysis**), or compares two samples with regard to the proportion of items in a given category (**Two-Sample Analysis**). There are three Data Types supported by this procedure: **Population Sample, Summary Table with Counts**, and **Summary Table with Proportions**.

Hypothesis Test for Proportion Dialog Box This analysis is set up using the **Hypothesis Test for Proportion** dialog box:

StatTools - Hypothesis Test for Proportion									
Analysis Type One-Sam	ple Analysis	-							
Data Type Summary Table with Counts									
Select One Variable with Categories and One or More Variables with Counts									
Data Set Table with	Counts	Format							
Cat Ct Name	Ad	dress							
Category	B18:B19								
Count	C18:C19								
Categories to Analyze (S	elect One or More)								
Satisfied									
2 categories in selected v	ariable								
- Hypotheses About Propr	artion								
Hypotheses About Prope									
Null Hypothesis Value	0.075								
Alternative <u>Hypothesis</u>	Less Than Null Value (One-Ta	iled Test) 🔹							
Options									
First Sample Size	400								
	-								
Second sample Size	100								
0 2		OK Cancel							

The number of variables selected depends on the Analysis Type used. A **One-Sample** analysis requires one or more samples for analysis; a **Two-Sample** analysis requires two samples. One selects variables with sample information using column labeled **Ct** (Count), % (Proportion), or column with no label for the **Population Sample** Data Type. (If the Population Sample data is stacked, the samples are selected in columns labeled **C1** and **C2**, where C1 contains the stacked categories.) If the data is in the form of a table with counts or proportions, there is an additional **Cat** column for selecting one variable with category names.

With the **Population Sample** Data Type variables can be from different data sets.

The options in the Hypothesis Test for Proportion dialog include:

- Analysis Type. Selects the type of analysis performed. Options include:
 - **One-Sample Analysis**. Performs hypothesis test for the proportion of items in a sample that belong to a given category.
 - **Two-Sample Analysis**. Performs hypothesis test for two samples with regard to the proportion of items in a given category
- Data Type. Specifies the type of data to be analyzed, either Population Sample, Summary Table with Counts, or Summary Table with Proportions.
- **Hypothesis Tests to Perform**. Specifies the hypothesis tests that will be performed on the selected proportion. Options change based on the analysis type selected. Options include:
 - **Null Hypothesis Value**, or the value of the population parameter under the null hypothesis.
 - Alternative Hypothesis Type, or the alternative to the Null Hypothesis Value that will be evaluated during the analysis. The Alternative Hypothesis Type can be either "one-tailed" (that is, greater or less than the null hypothesis) or "two-tailed" (that is, not equal to the null hypothesis).
- First Sample Size and Second Sample Size. For the Summary Table with Proportions Data Type, selects the size of the first sample and second sample (Two-Sample Analysis only).

Hypothesis Tests Report

0	🕞 🕼 🕐 · 🕲 · 😝 · 🗧 Book4 - Microsoft Excel 💷 📼 🗙										
0	🥐 🕼 Home Insert	Page Layout	Formulas	Data	Review	View	Add-Ins	StatToo	Is	🥝 –	⇒ x
Da Ma	ta Set Data inager Utilities + Data	Statistics * 🔒 No Graphs * 🔀 Tri Inference * 🔀 Re	ormality Test me Series an egression an Ana	s * d Forecasi d Classific hyses	ting *	Quality Cor Nonparame	ntrol * eteric Tests *	/ Utilit	ties * , *		
-	Δ1	fr StatTo	ols	ij ses				100	<u> </u>		×
	Δ.	B	C		D	F		F	G		н
1 2 3 4 5	StatTool Analysi Performed B Dat Updatin	Core Analysis Pac Hypothesis Test for Test Monday, February Live	k) or Proportion (16, 2009								
6 7 8	Hypothesis Test (Proportion)	Count Table with Counts									
9	Category	Unsatisfied									_
11	Sample Proportion	0.058									1
12	Hypothesized Proportion	0.075									
13	Alternative Hypothesis	< 0.075									
14	Standard Error of Sample Proportion	-1 3288									
16	p-Value	0.0920									
17	Null Hypoth. at 10% Significance	Reject									
18	Null Hypoth. at 5% Significance	Don't Reject									
19 20 21	Null Hypoth. at 1% Significance	Don't Reject									
14	Proportion Hypoth.	Test 🖉									
Rea	ady								00% 😑 📉		- 🕂 .:

- **Missing Data -** Missing data are allowed. All rows with missing data on the selected variables are ignored.
- Link to Data All of the reports are calculated with formulas that are linked to the data. If the values of the selected variable change, the outputs change automatically.

Sample Size Selection Command

Determines the sample size required to calculate confidence intervals

The Sample Size Selection command determines the sample size (or sample sizes) required to obtain a confidence interval with a prescribed half-length. It does this for confidence intervals for a mean, a proportion, the difference between two means, and the difference between two proportions. No data sets or variables are required, as the sample size is information you typically need *before* you collect data. You need to specify the confidence level, the desired half-length, and any other parameters necessary for determining sample size.

This analysis is set up using the **Sample Size Selection** dialog box:

Sample Size Selection Dialog Box

StatTools - Sample Size Selection									
Parameter to Estimate	Confidence Interval Speci	fication							
	Confidence Level	95% 💌							
C Proportion	Half-Length of Interval	0.1							
C Difference of Means	Estimated <u>S</u> td Dev	2							
C Difference of Proportions									
	ОК	Cancel							

The options in the Sample Size Selection dialog include:

- **Parameter to Estimate -** Selects the type of parameter that will be estimated from the sample (whose size you are determining). Options include **Mean**, **Proportion** (values between 0 and 1), **Differences of Means**, and **Differences of Proportions**.
- **Confidence Interval Specification** These options vary by the selected Parameter to Estimate, as shown:
 - When Parameter to Estimate is Mean and Difference of Means. First, enter the desired Confidence Level (usually between 90% and 100%), the Half Length of the Interval (the "plus or minus" component of the interval) and the Estimated Standard Deviation of the population. Note: Confidence Level and Interval Length are related, as a higher confidence level requires a longer confidence interval length.

When Parameter to Estimate is **Proportion** and **Difference of Proportions.** First, enter the desired **Confidence Level**(usually between 90% and 100%), the **Half Length of the Interval** (the "plus or minus" component of the interval) and the **Estimated Proportion** (a value between 0 and 1). If **Difference of Proportions** is being estimated, provide an **Estimated Proportion** for each population.



Estimate Report

Sample Size

- **Missing Data -** Not relevant.
- Link to Data Not relevant.

One-Way ANOVA Command

Performs a One-Way ANOVA on variables

The One-Way ANOVA command is a generalization of the twosample procedure for comparing means between two populations. With One-Way ANOVA the means from *at least two* (usually more than two) populations are compared. This is done with an ANOVA (analysis of variance) table. This table actually compares two sources of variation: the variation *within* each population against the variation *among* sample means from the different populations. If the latter variation is large relative to the former, as measured by an F test, then there is evidence of differences between population means.

The key value in the ANOVA table is the p-value. A small p-value is evidence of different population means. Besides the ANOVA table, it is informative to look at confidence intervals for all differences between pairs of means. Confidence intervals that do *not* include 0 are evidence of means that are *not* equal. StatTools provides the option of several types of confidence intervals, each based on a slightly different method.

This analysis is set up using the One-Way ANOVA dialog box:

itatTools - One-Way ANOVA										
Variables (Select One Category and C	Dne Value)									
Data Set Stacked										
Cat Val Name	Address									
🗹 🗖 Location E	B19:B253									
Response (C19:C253									
Confidence Interval Methods										
✓ No Correction	Confidence Level 95% 💌									
Bonferroni Correction										
Tukey Correction										
Scheffe Correction										
0 🚽 🛂	OK Cancel									

Two or more variables need to be selected for analysis. The selected data sets can be stacked or unstacked data. Variables can be from different data sets.

ANOVA Dialog Box

The options in the One-Way ANOVA dialog include:

- Confidence Interval Methods Select one or more from the following methods for correcting confidence intervals for individual variables. No Correction, or no correction performed, and Bonferroni, Tukey, and Scheffe correction methods.
- **Confidence Level** This is a "simultaneous" confidence level for the results for all variables. That is, it is the confidence you want to have that *all* of the confidence intervals contain their respective population mean differences. For technical reasons, the *actual* overall confidence level will typically be less than this specified level for the "no correction" confidence intervals. This is the reason for the "correction" methods above. They correct (expand) the lengths of the confidence intervals so that the overall confidence level is the one specified.

0	1 - 1- 14	- 🞯 -) =			Book6 -	Microsoft Excel					-		x
C	🔊 💽 Home	Insert	Page Lay	out Formulas	Data Revie	w View	Add-Ins	StatTools			0.	_ =	×
	and man	F Summan	Statistics	• M Normality Te	ists *		ol *	Jutilities -					
	\rightarrow	die Commen	Caraba	It's Time Carlos	and Free sections in	- Normania	de Teste e	(i) Halana					
Da	ta Set Data	Summary	/ Graphs +	[Not time series of	and Porecasting *	Nonparamete	inc rests *	Meip *					
Ma	nager Utilities *	A Statistica	al Inference	e * Kegression	and Classification *								
	Data			A	nalyses			Tools					
	A1	+ ()	f_x	StatTools									×
		А		В	С	D	E		F	G	1	н	-
1		Sta	tTools	(Core Analysis Pack)									
2			Analysis:	One-Way ANOVA									
3		Per	formed By:	Test									
4			Date:	Monday, February 16	, 2009								
5			Updating:	Live									
6													
7	1000000000												
8	ANOVA Summary			0.05									
9	Total Sample Size			235									
10	Grand Mean			5.383									
11	Pooled Std Dev			1.970									
12	Pooled Variance			5.904									
14	Confidence Level			95.00%									
15	Confidence Level			33.00%									
16				Parnance (Midwart)	Parnance (Northeast)	Paragana (South)	Response IS	outhwart) Por	papers (Mart)				
17	ANOVA Sample Stats			Stacked	Stacked	Stacked	Stac	ked	Stacked				
18	Sample Size			55	43	40	4	7	50				
19	Sample Mean			5.400	4.140	5,600	6.7	45	4.980				
20	Sample Std Dev			2.469	1.820	2.073	1.6	87	1.635				=
21	Sample Variance			6.096	3.313	4.297	2.8	46	2.673				
22	Pooling Weight			0.2348	0.1826	0.1696	0.20	000	0.2130				
23													
24				Sum of	Degrees of	Mean	C D	41-	a Malua				
25	OneWay ANOVA Tab	le		Squares	Freedom	Squares	r-ne	luo	p-value				
26	Between Variation			163.653	4	40.913	10.4	180 .	< 0.0001				
27	Within Variation			897.879	230	3.904							
28	Total Variation			1061.532	234								
29													
30				Difference	No Corre	ection							
31	Confidence Interval 1	Tests		of Means	Lower	Upper							
32	Response (Midwest)	-Response (No	rtheast)	1.260	0.467996986	2.052933246							
33	Response (Midwest)	-Kesponse (Sou	utn)	-0.200	-1.008974445	0.008974445							
34	Response (Midwest)	-nesponse (Sou	utnwest)	-1.343	-2.11/991/88	1 190699646							
36	Response (Northeast)	Response (We	escy auth)	-1 460	-2 315648321	-0.605281011							
37	Response (Northeast	Response (Sc	outhwest!	-2.605	-3.426674007	-1.783617838							
38	Response (Northeast	-Response /W	(est)	-0.840	-1.650132388	-0.030797844							
39	Response (South)-Re	sponse (South	west)	-1.145	-1.98214272	-0.307218982							
40	Response (South)-Re	sponse (West)		0.620	-0.205829889	1.445829889							
41	Response (Southwes	t)-Response (V	Vest)	1.765	0.973754079	2.555607623							
42			-				-						
14 .	One-Way	ANOVA	67									•	
Rea	dy									00% 😑		(9 .;

One-Way ANOVA Report

In the One-Way ANOVA report here, summary statistics for each population (in this case, each plant) appear at the top. After the summary statistics, a table of sample statistics for each variable is included. The ANOVA table appears next. In this example, the very small p-value indicates without a doubt that the mean scores from the five plants are not all equal. To see which means are different from which others, we look at the confidence intervals at the bottom of the report. Those pairs with values in bold have significantly different means.

- **Missing Data -** Missing data are allowed. All rows with missing data on the selected variables are ignored.
- Link to Data All of the reports are calculated with formulas that are linked to the data. If the values of the selected variable change, the outputs change automatically.

Two-Way ANOVA Command

Performs a Two-Way ANOVA on variables

The Two-Way ANOVA command performs a two-way analysis of variance. This is usually done in the context of an experimental design where there are two "factors" that are each set at several "treatment levels." For example, in a study of golf ball performance, the two factors might *Brand* and *Outside Temperature*. Then the treatment levels for *Brand* would be "A" through "E", and the treatment levels for Temperature would be the "*Cool*", "*Mild*" and "*Warm*". The value variable would be *Yards Driven*, and observations on this variable would be collected for a number of balls of each *Brand* /*Temperature* combination. The purpose of the study is to see whether there are significant mean differences among the various treatment level combinations.

The data for two-way ANOVA *must* be in stacked form. That is, there must be two "category" variables (corresponding to *Brand* and *Outside Temperature* in the above example), and there must be a "value" variable (corresponding to *Yards Driven* above). Also, the data set must be "balanced," meaning that there should be an equal number of observations in each treatment level combination. It is certainly possible to analyze an unbalanced design, but this design is best analyzed with regression (with dummy variables).

ANOVA Dialog Box

This analysis is set up using the Two-Way ANOVA dialog box:

StatTools - Two-Way ANOVA												
Variables (Select Two Category and One Value Variable)												
<u>D</u> ata Set	<u>F</u> ormat											
C 1 C 2 Val	Name	Address										
	Brand	B19:B318										
	Temp	C19:C318										
	Yards	D19:D318										
0 🖬 🖳	a	ОК	Cancel									

Two category variables (**C1** and **C2**) and one value variable (**Val**) need to be selected for analysis. The selected data set needs to be stacked data.
Two-Way ANOVA Report

0	Home In	sert Page Lay	out Formula	s Data i	Review View	Add-Ins	StatTools	0 -	
Da	ta Set Data nager Utilities +	ummary Statistics ummary Graphs + tatistical Inference	 Mormality Time Serie Regression 	Tests + s and Forecasting n and Classificatio	← Quality C * ▲ Nonpara n *	Control * meteric Tests *	 Utilities * Help * 		
	Data			Analyses			Tools		
	A1 -	$-(\circ f_x)$	StatTools						
	A	В	С	D	E	F	G	Н	
	StatTools	(Core Analysis Pack)	í.						
	Analysis:	Two-Way ANOVA							
	Performed By:	lest Monday February 1	6 2009						
	Updating:	Live	0,2005						
	ANOVA Sample Sizes	Cool	Mild	Warm	Totals				
	A	20	20	20	60				
C	В	20	20	20	60				
1	c	20	20	20	60				
2	D	20	20	20	60				
3	E	20	20	20	60				
4	Totals	100	100	100					
5	Balanced	TRUE							
6									
7			L SIGNET						
8	ANOVA Sample Means	Cool	Mild	Warm	Totals				
9	A	218.82	236.45	258.44	237.90				
0	B	224.15	245.13	258.27	242.52				
1	-	228.00	242.72	203.04	244.36				
2	5	215.00	237.02	250.11	230.24				
D A	t Totale	224.75	233.75	270.54	230.45				
5	Totals	222.15	243.33	201.30					
6									
7	ANOVA Sample Std Dev	Cool	Mild	Warm	Totals				
8	A	10.90	8.83	11.01	19.22				
9	в	11.70	9.80	8.93	17.36				
0	c	10.85	14.25	7.08	18.15				
1	D	13.64	10.18	12.13	20.69				
2	E	10.67	10.96	9.05	21.84				
3	Totals	12.28	12.78	10.98					
4						1.12			
5		Sum of	Degrees of	Mean	E-Ratio	n-Value			
6	TwoWay ANOVA Table	Squares	Freedom	Squares		p saue			
7	Brand	7702.44	4	1925.61	16.47	< 0.0001			
0	Temp	77086.00	2	38543.00	329.58	< 0.0001			
0	Interaction	1999.97	8	250.00	2.14	0.0325			
9			005						
9	Error	33329.13	285	116.94					

The top three items are summary measures (sample sizes, sample means, and sample standard deviations) for the various treatment level combinations. The bottom part of the output shows the ANOVA table. There are three important p-values in this table: two for "main effects" and one for "interactions." The main effects indicate whether there are significant mean differences across levels of either factor, averaged over the levels of the other factor. For example, the Temperature main effect indicates whether the values in cells B17 to D17 are significantly different. (They are, as indicated by the very small p-value for *Temperature* in the ANOVA table.)

- **Missing Data** Because of the requirement for a balanced design, there should not be any missing data.
- Link to Data All of the ANOVA formulas are linked to the data. If the data change, the results change automatically. The balance of the experiment is verified when the analysis is run. Changes in the data may affect the balance and cause invalid results.

Chi-square Independence Test Command

Tests for independence between the row and column attributes of a contingency table

The Chi-square Independence Test command tests for independence between the row and column attributes of a contingency table. For example, if the contingency table lists counts of people in different drinking and smoking categories, the procedure tests whether smoking habits are independent of drinking habits. The contingency table (also called cross-tabs) could be an Excel pivot table.

This procedure is somewhat different from most StatTools procedures. For this analysis all that is required is a rectangular contingency table. Each cell in this table should be a count of observations in a particular row/column combination (nondrinkers and heavy smokers, for example). The table can have row and column labels (headings) and/or row and column totals, but these are not necessary and are only used for clarity in StatTools reports.

Chi-square Independence Test Dialog Box

This analysis is set up using the **Chi-square Independence Test** dialog box:

StatTools - Chi-Square Test for Independence 🛛 🛛 🔀						
Contingency Table Range C16:G20						
Row and Column He	Row and Column Headers and Titles					
Table Includes R	ow and Column <u>H</u> eaders					
Columns Title	Drinking					
Rows Title	Rows Title Smoking					
OK Cancel						

The options in the Chi-square Independence Test dialog include:

• Row and Column Headers and Titles - Select one or more from the following: Table Includes Row and Column Headers or headers in the leftmost column and topmost row of the table, Columns Title or the title you want to use to represent the columns in the table, Rows Title or the title you want to use to represent the rows in the table.

Chi-square Independence Test Report

6	100.00	ج (- ک	Book	8 - Microsoft E	xcel		-	= x
U	Home Ir	nsert Page La	vout Formula	s Data Re	view View	Add-Ins St	atTools 🕜 –	σx
		Cummen Chefint			A- 0	Combact a	Service -	
		Summary Statisti	I Normain	ly rests *		y control *	1 Ouncies	
Da	ta Set Data	Summary Graphs	 Time Ser 	ies and Forecastin	g 👻 📐 Nonpa	arameteric Tests *	🕜 Help *	
Ma	nager Utilities * 🛛 🗥	Statistical Inferen	nce 🔹 🔀 Regressi	on and Classificat	ion *			
	Data			Analyses			Tools	
	A1	- (°)	StatTools					*
	۸	в	C	D	F	E	G	-
	ChatTaala	1	U U	U	-		0	-
1	Statiools	Core Analysis Pa	:k)					
2	Analysis: Decformed By:	Chi-Square Indep	endence Test					
4	Dates	Monday, Februar	16, 2009					
5	Updating	: Live						
6								
7			Rows: S	moking / Columns: [Drinking			
8	Original Counts	None	Light	Moderate	Heavy	Total	=7)	
9	None	20	15	10	10	55		
10	Occasional	10	15	12	15	52		
11	Moderate	8	12	15	20	55		
12	Heavy	0	10	15	25	219		
14	Iotal	44	52	52	70	210		
15			Rowers	moking / Columns: I	rinking			
16	Percentage of Rows	None	Light	Moderate	Heavy			
17	None	36.36%	27.27%	18.18%	18.18%	100.00%	e.)	
18	Occasional	19.23%	28.85%	23.08%	28.85%	100.00%		
19	Moderate	14.55%	21.82%	27.27%	36.36%	100.00%		
20	Heavy	10.71%	17.86%	26.79%	44.64%	100.00%		
21	Contract Contract							-
22			Rows: Smoking /	Columns: Drinking				
23	Percentage of Columns	None	Light	Moderate	Heavy	=12		
24	None	45.45%	28.85%	19.23%	14.29%			
20	Occasional	10 100/	28.85%	23.08%	21.43%			
20	Home	13 64%	19 23%	28.85%	35 71%			
28	Treavy	100.00%	100.00%	100.00%	100.00%			
29	-							
30			Rows: Smoking /	Columns: Drinking				
31	Expected Counts	None	Light	Moderate	Heavy	- 100		
32	None	11.1009	13.1193	13.1193	17.6606	- 12		
33	Occasional	10.4954	12.4037	12.4037	16.6972			
34	Moderate	11.1009	13.1193	13.1193	17.6606			
35	Heavy	11.3028	13.3578	13.3578	17.9817			
30								
32	Distance from Expected	Nono	Kows: Smoking /	Modorato	Homer			
39	None	7,1340	0.2696	0.7416	3,3229			
40	Occasional	0.0234	0.5435	0.0131	0.1725			
41	Moderate	0.8662	0.0955	0.2696	0.3099			
42	Heavy	2.4878	0.8441	0.2019	2.7393			
43								
44								
45	Chi-Square Statistic		-					
46	Chi-Square	20.0349						
47	p-Value	0.0177						-
14 4	Chi-Sq Inde	ependence Tes	t/10/			III		
Rea	ady					🛛 🛄 100% 🕞		- (+) ,;;

The report above shows the basic result of the test, a p-value. If this p-value is small (as here) we can conclude that the row and column attributes are *not* independent. We can study the numbers on this sheet to understand better how smoking and drinking are related.

- **Missing Data -** There should not be any missing data in the cells of the contingency table.
- Link to Data The formulas in the Chi-square Independence Test Report are linked to the data. So if the counts in the original contingency table change, the outputs on this sheet change also.

Normality Tests Menu

Because so many statistical procedures assume that a set of data is normally distributed, it is useful to have methods for checking this assumption. StatTools provides three commonly used checks, as described in this section.

Chi-Square Normality Test Command

Tests if observed data for a variable is normally distributed

The Chi-Square Normality Test procedure uses a chi-square goodness-of-fit test to test whether the observed data in a specified variable could have come from a normal distribution. To do so, it creates a histogram of this variable, using the categories you specify, and it superimposes a histogram for a normal distribution on the histogram from the data. If the two histograms have essentially the same shape, we cannot reject the null hypothesis of a normal fit.

The formal test is performed by comparing the observed counts in the various categories with the expected counts that are based on a normality assumption. Actually, the procedure allows you to test several variables (separately) for normality. A histogram is created for each variable you select, and the chi-square test is run on each of them.

The only requirement for the Chi-Sq Normality Test is that there must be at least one numerical variable. Beyond this, most analysts suggest that there should be at least 100 observations -- the more, the better. This analysis is set up using the Chi-Sq Normality Test dialog box:

Chi-Sq Normality Test Dialog Box

StatTools Chi-Square Normality Test						
Variables (Select Exactly One)						
Data Set Data Set #1	Eormat					
Name	Address					
Part #	B15:B104					
Vidth	C15:C104					
Binning Options						
# <u>B</u> ins auto ▼						
Minimum auto 💌 🗖	Extend to -Infinity					
M <u>a</u> ximum auto ▼ Γ	Extend to +Infinity					
	OK Cancel					

One variable can be selected for testing. The selected data set must be unstacked data.

The options in the Chi-Sq Normality Test dialog include:

- **# Bins -** Specifies a fixed number of bins or, alternatively, specifies that the number of bins will be automatically calculated for you.
- Minimum and Maximum auto specifies that the minimum and maximum of your data set will be used to calculate the minimum and maximum of equal interval bins. First and last bins, however, may be added using the Extend to -Infinity and Extend to +Infinity options. If Auto is not selected, you can enter a specific Minimum and Maximum value where your bins will start and end. This allows you to enter a specific range where binning will be performed without regard to the minimum and maximum values in your data set.
- Extend to -Infinity indicates that the first bin used will stretch from the specified minimum to -Infinity. All other bins will be of equal length. In certain circumstances, this improves testing for data sets with unknown lower bounds.
- Extend to +Infinity indicates that the last bin used will stretch from the specified maximum to +Infinity. All other bins will be of equal length. In certain circumstances, this improves testing for data sets with unknown upper bounds.

Chi-Sq Normality Test Report



The results of the test are shown in the report above. The p-value of 0.4776 is good evidence that the amounts are normally distributed. More evidence to this effect appears in the histograms in Figure 4 and the frequency data. However, be aware of two things. First, if there are too few observations (well less than 100, say), then the chi-square test is not good at distinguishing normality from nonnormality. The effect is that the p-value is usually not small enough to reject the normality hypothesis. In essence, almost *everything* tends to look normal with small data sets. On the other hand, if the data set is really large (several hundred observations, say), then the p-value will usually be small, indicating nonnormality. The reason is that with large data sets, every little "bump" in the curve is likely to create a small p-value. In this case, the real test is a *practical* one: Do the histograms really differ that much for all practical purposes?

- **Missing Data -** Missing data are allowed. All rows with missing data on the selected variables are ignored.
- Link to Data The histogram and all formulas for the test are linked to the original data. So if the data change, the histogram and the test results change automatically.

Lilliefors Test Command

Tests if observed data for a variable is normally distributed

The Lilliefors Test procedure provides a more powerful test for normality than the more familiar chi-square goodness-of-fit test. (More powerful means that it is more likely to detect non-normality if it exists.) It is based on a comparison of the "empirical cdf" and a normal cdf, where "cdf" stands for cumulative distribution function, showing the probability of being less than or equal to any particular value.

The empirical cdf is based on the data. For example, if there are 100 observations and the 13th smallest is 137, then the empirical cdf, evaluated at 137, is 0.13. The Lilliefors test finds the maximum vertical distance between the empirical cdf and the normal cdf, and it compares this maximum to tabulated values (that are based on sample size). If the observed maximum vertical distance is sufficiently large, then we have evidence that the data do *not* come from a normal distribution.

Lilliefors Test Dialog Box This analysis is set up using the Lilliefors Test dialog box:



One or more variables can be selected for testing. The selected data set needs to be unstacked data. Variables can be from different data sets.

Lilliefors Test Report



The results of the test are shown in the report above. There is no pvalue (as in most hypothesis tests), but we see from the statement that the maximum vertical distance is sufficiently large to cast doubt on the normality assumption. More evidence to this effect appears in the cdf's in the included chart. Actually, the fit between the two curves appears to be "pretty good," and it might be good enough for all practical purposes. That is, we might conclude that these data are "close enough" to being normally distributed for our purposes.

- **Missing Data -** Missing data are allowed. All rows with missing data on the selected variables are ignored.
- Link to Data The CDFs and all formulas for the test are linked to the original data. So if the data change, the graph and the test results change automatically.

Q-Q Normal Plot Command

Tests if observed data for a variable is normally distributed

The Q-Q Normal Plot command creates a quantile-quantile (Q-Q) plot for a single variable. It provides an informal test of normality. Although the details are somewhat complex, the objective is fairly simple: to compare the quantiles (or percentiles) for the data to the quantiles from a normal distribution. If the data are essentially normal, then the points on the Q-Q plot should be close to a 45-degree line. However, obvious curvature in the plot is an indication of some form of non-normality (skewness, for example).

Q-Q Normal Plot Dialog Box This analysis is set up using the Q-Q Normal Plot dialog box:

StatTools -	Q-Q Normal Pl	ot	
Variables (Se	elect Exactly One)		
<u>D</u> ata Set	Data Set #1		▼ <u>F</u> ormat
	Name	Address	
Part #		B15:B104	
Width		015-0104	
1 Widdi		015:0104	
Options		<u>C13:C104</u>	
Options -		013:0104	
Options	g Standardized Q-	Values	ce Line

One variable can be selected for plotting. The selected data set needs to be unstacked data.

This options in the Q-Q Normal Plot dialog box include:

• Plot Using Standardized Q-Values - Specifies to use a standardized Q-Value, instead of Q-Q data, on the Y-axis of the graph. This makes comparisons of the Y-axis values between Q-Q Normal plots possible.

Q-Q Normal Plot Report



As stated earlier, this is an *informal* test of normality. It is difficult to say "how close" to a 45-degree line the plot should be to accept a normality assumption. Typically, we look for obvious curvature in the plot, and none is apparent in the plot here.

- **Missing Data -** Missing data are allowed. All rows with missing data on the selected variables are ignored.
- Link to Data The plots and all formulas are linked to the original data. So if the data change, the plot changes automatically.

Time Series and Forecasting Menu

The procedures on the Time Series and Forecasting menu deal with the analysis of data collected over time, with applications for forecasting and quality control. The forecasting methods provided include the moving averages method, simple exponential smoothing, Holt's exponential smoothing method for capturing trend, and Winters' exponential smoothing method for capturing seasonality.

Time Series Graph Command

Creates a time series graph for variables

The Time Series Graph command plots one or more time series variables all on the same plot. If two variables are selected, you have the option of using the same or different scales on the Y-axis for the two variables. The latter option is useful when the ranges of values for the two variables are considerably different. However, if more than two variables are plotted, they must all share the same vertical scale.

There must be at least one numerical variable in the data set. There can also be a "date" variable, but if it is to be used to label the horizontal axis of the chart, it must be selected as the "label" variable.

This graph type is set up using the **Time Series Graph** dialog box:

Time Series Graph Dialog Box

StatTools - Time Series Graph							
<u>G</u> raph Format	Graph Format Time Series (without Label)						
Variables (Sele	Variables (Select One or More)						
<u>D</u> ata Set	Data Set #1		•	Eormat			
Val	Name		Address				
Week		B11:B52					
Sales		C11:C52					
Options							
☑ Plot All Variables on a Single Graph							
☑ Use Two Y-Axes (Graphs of Two Variables Only)							
0 🖬 🛂			ОК	Cancel			

One or more variables can be selected for graphing. The selected data set must be unstacked data. Variables can be from different data sets. The Label variable (Lbl checkbox) appears on the X-axis.

The options in the **Time Series Graph** dialog include:

- **Plot All Variables on a Single Graph**. Selects to plot all variables in one graph.
- **Use Two Y-Axes**. Select to display a separate Y-axis for each variable in a two variable graph. Units and values for each variable can then be displayed on the graph.

Time Series Graph of Two Variables



- **Missing Data -** Missing data are allowed. All rows with missing data on the selected variables are ignored.
- Link to Data Graphs are linked to data so if the data changes the graph automatically updates.

Autocorrelation Command

Calculates the autocorrelations for variables

The Autocorrelation Command calculates the autocorrelations for any selected numeric variable. Typically, this variable will be a time series variable, although StatTools will do the calculations for any variable. You can select the number of autocorrelations you want (i.e., the number of lags). You can also request a chart of the autocorrelations, called a correlogram. The output shows which of the autocorrelations, if any, are significantly different from 0.

Autocorrelation Dialog Box This graph type is set up using the Autocorrelation dialog box:

StatTools - Autocorrelation						
Variables (Select One or More)						
<u>D</u> ata	Set Data Set #1		T	<u>F</u> ormat		
	Name		Address			
	Week	B15:B56				
	Sales	C15:C56				
— Ор	tions	_				
Nur	mber of Lags auto	•	Create Autocorrela	tion Chart		
0			ОК	Cancel		

One or more variables can be selected for analysis. The selected data set must be unstacked data.

The options in the Autocorrelation dialog include:

- Number of Lags. The number of time periods to lag when calculating autocorrelations. If **Auto** is selected, StatTools determines the appropriate number of lags to test. If you enter a specific # of lags, the maximum number of lags you can request is 25% of the number of observations in the series. For example, if you have 80 monthly values, you can request up to 20 lags.
- **Create Autocorrelation Chart.** Creates a bar chart with the height of each bar equal to the corresponding autocorrelation.

Autocorrelation Report



The Autocorrelation report appears above. For each lag, the corresponding autocorrelation appears, along with an *approximate* standard error.

- **Missing Data -** This procedure allows missing data at the beginning of the time series, but none in the middle or at the end of the series.
- Link to Data StatTools ties the output to the data. Therefore, if the data change, the autocorrelations (and the correlogram) change automatically.

Runs Test for Randomness Command

Performs a runs test to check whether a variable is random

The Runs Test for Randomness command allows you to check the "randomness" of a sequence of values in a variable, usually a time series variable. It indicates how many "runs" there are in the sequence, where a run is a consecutive number of values on one side or the other of some cutoff point (such as the mean or median of the sequence). For a random sequence, we would expect neither too few runs nor too many. The runs test counts the number of runs and then reports a p-value for the test. If this p-value is small, we can conclude that sequence is probably not random, i.e., there are either too many or too few runs.

This analysis is set up using the **Runs Test for Randomness** dialog box:

Runs Test for Randomness Dialog Box

Stat	StatTools - Runs Test for Randomness						
Vari	Variables (Select One or More)						
<u>D</u> ata	Set Data	a Set #1		-	<u>E</u> ormat		
	Nam	e		Address			
	Week		B15:B56				
	Sales		C15:C56				
- Cu	itoff Value for D	efining Runs	I				
۲	<u>M</u> ean of Series						
C	Median of <u>S</u> eries	s					
0	<u>C</u> ustom Cutoff						
0				ОК	Cancel		

One or more variables can be selected for analysis. The selected data set must be unstacked data. Variables can be from different data sets.

The options in the Runs Test for Randomness dialog include:

• Cutoff Value for Defining Runs - A runs test is always based on runs above or below some Cutoff Value. This can be the Mean of Series, the Median of Series, or any other Custom Cutoff Value you want to enter.

Runs Test for Randomness Report

0	3 - 6	- @-) -	Book15	- Microsoft	Excel				-	= x
C	🥙 💽 Home	Insert	Page Layou	ut Formulas	Data R	eview	View Ad	d-Ins S	atTools		⇒ x
Da Ma	ta Set Data mager Utilities + Data	😨 Sumi	mary Statistics mary Graphs * stical Inference	 Mormality Time Serie Regressio 	Tests + s and Forecast n and Classific Analyses	ing + 🔺	Quality Co	ntrol + eteric Tests	• 0	Utilities * Help * Tools	
	A1	• (f _x	StatTools							×
	А		В	С	D		E	F		G	
4 5 6 7 8	Perform Up	Date: M dating: Li	sst onday, February ve Sales Data Sot #1	16, 2009							
9	Observations	mess	42	-01							
10	Below Mean		17								
11	Above Mean		25								
12	Number of Runs		9								
13	Mean		68.43	,							
14	E(R)		21.2381								1
15	StdDev(R)		3.0810								
10 17 18	P-Value (two-tailed)	•	< 0.0001								
14	Runs Te	st ⁄ 🔁	/					-III.	_		1
Rea	ady						#0U	100% (=)	0	+ .:

The report above shows the number of runs and the expected number of runs under randomness E(R). As 9 is significantly below 21.2381, this series is not entirely random; i.e., sales do not "zigzag" as much as a random series.

- **Missing Data** This procedure allows missing data at the beginning and at the end of the time series, but none in the middle of the series.
- Link to Data StatTools ties the output to the data. Therefore, if the data change, the reports change automatically.

Forecasting Command

Generates forecasts for time series variables

The Forecasting command provides you with a number of methods for forecasting a time series variable. These methods include the moving averages method, simple exponential smoothing, Holt's exponential smoothing method for capturing trend, and Winters' exponential smoothing method for capturing seasonality. The Forecasting command also allows you to deseasonalize the data first, using the ratio-to-moving-averages method and a multiplicative seasonality model. Then you can use any of the forecasting methods (other than Winters' method) to forecast the deseasonalized data, and finally "reseasonalize" the forecasts to get back to original units.

The forecast reports include a set of columns to show the various calculations (for example, the smoothed levels and trends for Holt's method, the seasonal factors from the ratio-to-moving-averages method, and so on), the forecasts, and the forecast errors. Summary measures are also included (MAE, RMSE, and MAPE) for tracking the fit of the model to the observed data. (When you use exponential smoothing methods, you have the option of using optimization to find the smoothing constant(s) that minimize RMSE.)

Finally, several time series plots are available, including a plot of the original series, a plot of the series with forecasts superimposed, and a plot of the forecast errors. In case of deseasonalizing, these plots are available for the original series and the deseasonalized series.

Forecasting Dialog Box

Forecasts are set up using the **Forecasting** dialog box:

StatTools - Forecast					
Variables (Select Exactly One)					
Data Set Coca-Cola Data	✓ Eormat				
Name	Address				
Quarter	B13:B73				
Sales	C13:C73				
Forecast Settings Time Scale	Graphs to Display				
Number of Forecasts 4	Qptimize Parameters				
Number of Holdouts 8					
- Method	Parameters				
C Moving Average	Level (a) 0.1				
C Exponential Smoothing (Sin	nple) Irend (b) 0.1				
C Exponential Smoothing (Ho	lte) Seasonality (g) 0.1				
C Exponential shood ling (<u>no</u>	icas geosphancy (g)				
 Exponential Smoothing 					
	OK Cancel				

One variable can be selected for analysis. The selected data set must be unstacked data.

The **Forecasting** options in this dialog specify the forecasting method used and the settings for the selected method. Options include:

- **Number of Forecasts**. Specifies the number of *future* periods to provide forecasts for.
- Number of Holdouts. Specifies the number of observations to "hold out", or not use in, the forecasting model. You can choose to use all of the observations for estimating the forecasting model (0 Holdouts), or you can hold out a few for validation. Then the model is estimated from the observations not held out, and it is used to forecast the held-out observations.
- **Optimize Parameters (exponential smoothing methods only)**. Finds the smoothing constant that minimizes the RMSE (for the non-holdout period). Optimization requires Parameters shown in the dialog to be between 0 and 1. If you edit parameter values directly in a Forecasting report, make sure to enter values in this range.
- **Deseasonalize**. Selects to deseasonalize data before forecasting. For seasonal data, that is, data suspected of having a seasonal pattern, there are two options. You can use Winters' method,

Forecasting Options -Forecasting Dialog Box which deals with seasonality directly, or you can select this option to deseasonalize the data first, using the ratio-to-moving-averages method for deseasonalizing. Then *any* method can be used to forecast the deseasonalized series.

- **Method** selects the forecasting method to be used; either **Moving Average** or **Simple**, **Holt's**, or **Winters'** exponential smoothing methods.
- **Parameters** specifies the parameters to be used for the selected forecasting method:
 - **Span parameter** (Moving Average method only), or the number of consecutive observations used in each moving average.
 - **Level parameter** (all Exponential Smoothing methods), a smoothing parameter that can take any value between 0 and 1 (the default value is 0.1).
 - **Trend parameter** (Holt's and Winter's Exponential Smoothing methods), a second smoothing parameter that can take any value between 0 and 1 (the default value is 0.1).
 - **Seasonality parameter** (Winter's Exponential Smoothing method only), a third smoothing parameter that can take any value between 0 and 1 (the default value is 0.1).

Note: If Optimize Parameters is selected, Level, Trend and Seasonality parameters cannot be set, as these are the parameters whose values are being optimized.

Time Scale Options -Forecasting Dialog Box The Time Scale options specify the timing and time scale labeling for the analyzed variable. Options include:

- Seasonal Period. Specifies the type of time series data; either Annual, Quarterly, Monthly, Weekly, Daily or None. This is used for seasonalizing data and for labeling.
- **Label Style.** Specifies how the time scale will be labeled on any generated graphs.
- **Starting Label**. Specifies the entry for the first time scale label on the graph.

Graph Options -Forecasting Dialog Box

The Graph options specify the forecast graphs that will be generated. Available graphs include:

- 1) **Forecast Overlay**, or the time series graph of the data values generated by the forecast
- 2) **Original Series**, or the time series graph of the actual data
- 3) **Forecast Errors**, or the error between the forecast and actuals
- 4) **Deseasonalized Forecast Overlay**, or the time series graph of the data values generated by the forecast after the original data is deseasonalized
- 5) **Deseasonalized Original Series**, or the time series graph of the actual data after it is deseasonalized
- 6) **Deseasonalized Forecast Errors**, or the error between the forecast and actuals after the original data is deseasonalized

Example Forecasting Report



- **Missing Data -** Missing data are allowed at the *beginning* of the time series, but not in the middle or at the end.
- Link to Data Due to the lengthy calculations required, forecasts are not linked to data. If changes are made to the original data, the procedure should be rerun.

Regression and Classification Menu

The commands on the Regression and Classification Menu perform regression and classification analyses. Available regression analyses include Simple Multiple, Stepwise, Forward, Backward and Block. Other analyses on the Regression and Classification menu include discriminant analysis and logistic regression.

Regression Command

Runs regression analyses on a set of variables

	The Regression Command runs a variety of regression analyses including Simple Multiple, Stepwise, Forward, Backward and Block. Reports from each analysis include summary measures of each regression equation run, an ANOVA table for each regression, and a table of estimated regression coefficients, their standard errors, their t- values, their p-values, and 95% confidence intervals for them for each regression.
	In addition, you have the option of creating two new variables, the fitted values and the residuals, and creating a number of diagnostic scatterplots.
Available Regression Types	The available regression types include Simple Multiple , Stepwise , Forward , Backward or Block . The Simple Multiple regression procedure builds an equation all at once, using the selected explanatory variables. The other procedures allow variables (or blocks of variables) to enter or leave the equation sequentially. Specifically, the stepwise procedure allows variables to enter one at a time. The next variable to enter is the one most highly correlated with the unexplained part of the response variable. However, the stepwise option also allows variables to leave once they have entered if they no longer contribute significantly. The forward procedure is the same as the stepwise procedure, except that variables are not allowed to leave once they enter. The backward procedure starts with all potential explanatory variables in the equation and then deletes them one at a time if they do not contribute significantly. Finally, the block procedure allows blocks of explanatory variables to enter or not enter <i>as a block</i> in a specified order. If one block is not significant and does not enter, then no later blocks are considered for entry.

Regression Dialog Box

These analyses are set up using the **Regression** dialog box:

StatTools - Regression					
Regression Type Multiple Variables (Select One or More Independents and One Dependent)					
Data Set Data Set #1	▼ <u>F</u> ormat				
I D Name	Address				
Month	C20:C55				
ProdRuns	D20:D55				
	E20;E33				
Parameters	Graphs				
Use p-Values O Use F-Values	Fitted Values vs Actual <u>Y</u> -Values				
p-Value to Enter 0.05	Fitted Values vs X-Values				
p-Value to Leave 0.1	Residuals vs Fitted Values				
Set Constant to Zero (Origin)	Residuals vs X- <u>V</u> alues				
Advanced Options					
Indude Detailed Step Information					
Include Prediction for Data Set Data Set #1					
Confidence Level 95%	▼				
	OK Cancel				

One dependent variable (D) and one or more independent variables (I) need to be selected for analysis except in the case of a Block regression. With Block regression, one dependent variable (D) and one to seven blocks (B1 to B7) need to be selected. The selected data set must be unstacked data. Variables can be from different data sets.

The options in the **Regression** dialog include:

• **Regression Type**. Selects the type of regression to perform - Simple Multiple, Stepwise, Forward, Backward or Block.

Regression Parameters change with the selected Analysis Type. Options include:

- Use p-Values Affects how variables are added or removed from the regression equation. When selected, you can specify a p to Enter and/or a p to Leave, depending on the regression method in use. The lower the p-value, the more significant a variable must be to enter or leave the regression equation. The default values that are shown are usually acceptable. Just remember two things. First, the p-value to enter cannot be larger than the p-value to leave. Second, to make it easier for variables to enter (and harder to leave), use *larger* p-values. To make it harder for variables to enter (and easier to leave), use *smaller* p-values. Typical p-values are in the range .01 to .1.
- Use F-Values As with p-Values, affects how variables are added or removed from the regression equation. Allows you to specify a F to Enter and/or a F to Leave, depending on the regression method in use. Typical values are in the range 2.5 to 4.

StatTools can create several optional scatterplots, as shown in the selected **Graphs** options. These include:

- Fitted Values vs. Actual Y-Values
- Fitted Values vs. X-Values
- Residuals vs. Fitted Values
- Residuals vs. X-Values

These plots are typically used in "residual analysis" to check whether the regression assumptions are satisfied. Probably the most useful plot is the one of the residuals (on the vertical axis) versus the fitted (or predicted) values of the response variable. Advanced Options for a regression analysis include:

- **Include Detailed Step Information**, reporting r-squared and standard error statistics for the independent variable at each intermediate step of the regression.
- Include Prediction, where predicted values for the dependent variable are generated for the independent variable values in a second data set. This prediction data set must have the same variable names as the original data set that the regression is analyzing. Typically, in the prediction data set, you will have sets of values for independent variables for which you wish to predict the value for the dependent variable. The regression equation calculated from the first data set is used to make the predictions. The predicted values for the dependent variable will be entered directly in the prediction data set; filling the column (or row) for the dependent variable with the predicted values. The Confidence Level specifies the lower and upper bounds that will be generated for the predicted values.

Stepwise Regression Report

0	0 - 6	- 🖉 -) =		Book17 - Micro	soft Excel				x
C	🥙 💽 Home	Insert Pa	ge Layout 🛛 Fo	rmulas Data	Review V	iew Add-Ins	StatTools @	_ =	x
Da Ma	ta Set Data anager Utilities ~ Data	Summary St	atistics * 🔒 No raphs * 🕌 Tin Iference * 🔀 Re	rmality Tests + ne Series and Fore gression and Clas Analyses	casting • 🔺 N sification •	Quality Control 👻 Ionparameteric Te	sts + Otilities Help + Tools		
	A1	- (9			≽				
	А	В	С	D	E	F	G	Н	
1 2 3 4 5	StatTools Analysis: Performed By: Date: Updating:	(Core Analysis Pac Regression Test Monday, February Static	k) 16, 2009						
6 7 8 9	Summary	Multiple R 0.9308	R-Square	Adjusted R-Square 0.8583	StErr of Estimate 4108.99309	Ū			
10 11 12	ANOVA Table	Degrees of Freedom	Sum of Squares	Mean of Squares	F-Ratio	p-Value			I
13 14 15	Explained Unexplained	2 33	3614020661 557166199.1	1807010330 16883824.22	107.0261	< 0.0001			
16 17	Regression Table	Coefficient	Standard Error	t-Value	p-Value	Confidence I Lower	Interval 95% Upper		
18	Constant	3996.678209	6603.650932	0.6052	0.5492	-9438.550632	17431.90705		
19	MachHrs	43.53639812	3.5894837	12.1289	< 0.0001	36.23353862	50.83925761		
20	ProdRuns	883.6179252	82.25140753	10.7429	< 0.0001	716.2761784	1050.959672		
21 22 23		s	catterplot of Fit	vs Overhead					
24	140000.0								
25	120000.0	-			1000				
27	i fa conserva				+				
28	100000.0			44 44	## # "				
29				· · · · ·	叶十				
30	80000.0			+ + *					
31	Ē								
32	60000.0	- 3							
33	40000 0								
34	40000.0								
35	20000.0								
36									
37	0.0			r.					
38		0.0 20000.0	40000.0 60000	.0 80000.0 100	0000.0 120000.0	140000.0			
39 40			o	verhead					
11	Regres	sion 🕅				10		*	
Rea	ady					田口田 100%	0 0	(+)	.:

- **Missing Data -** If there are missing values, then any row with missing values for *any* of the selected variables is ignored.
- Link to Data There is no link to the original data. If the data change, you must rerun the analysis.

Logistic Regression Command

Runs a logistic regression on a set of variables

The Logistic Regression command performs a logistic regression analysis on a set of variables. This is essentially a nonlinear type of regression analysis where the response variable is binary: 0 or 1. There should be a 0-1 response variable that specifies whether each observation is a "success" or "failure", plus one or more explanatory variables that can be used to estimate the probability of success.

A second option for logistic regression is to have a "count" variable that specifies the number of "trials" observed at each combination of explanatory variables. Then the response variable should indicate the number of trials resulting in "success." The result of the logistic regression is a regression equation that is similar to a regular multiple regression equation. However, it must be interpreted somewhat differently, as explained below.

StatTools' logistic regression procedure relies on optimization to find the regression equation. This optimization must use a complex nonlinear algorithm, so the procedure can take quite a while, depending on the speed of your PC. This analysis is set up using the **Logistic Regression** dialog box:

Logistic Regression Dialog Box

StatTools - Logistic Regression									
Analysis Type Samples with no Co Variables (Select One Dependent an Data Set Data Set #1	d One or More Independents)								
I D Name Person Income Income InvestAmt InvestAmt V	Address B16:899 C16:C99 D16:D99 E16:E99								
Options Include Classification Summary Include Classification Results Include Prediction Include Prediction	OK Cancel								

One dependent or response variable (D) and one or more independent variables (I) need to be selected for analysis. The data must:

- Be in "stacked" form, and there must either be a 0-1 response variable that specifies whether each observation is a "success" or "failure". This is referred to here as Samples with No Count Variable.
- 2) Have a "count" variable and an integer response variable. This is referred to here as **Summary of Samples (with Count Variable).**

Variables can be from different data sets. If the Analysis Type is set to **Summary of Samples (with Count Variable)**, an additional count variable needs to be selected.

The options in the Logistic Regression dialog include:

• Analysis Type. Selects the type of logistic regression to perform – Samples with No Count Variable or Summary of Samples (with Count Variable).

Samples with No Count Variable have a 0-1 response variable that specifies whether each observation is a "success" or "failure", plus one or more explanatory variables that can be used to estimate the probability of success.

Summary of Samples (with Count Variable) have a "count" variable that specifies the number of "trials" observed at each combination of explanatory variables. Then the response variable should indicate the number of trials resulting in "success." For this analysis type a separate column C appears in the Variable Selector, allowing the count variable to be selected.

- Include Classification Summary with the regression report.
- Include Classification Results with the regression report.
- Include Prediction, where predicted values for the dependent variable are generated for the independent variable values in a second data set. This prediction data set must have the same variable names as the original data set that the regression is analyzing. Typically, in the prediction data set, you will have sets of values for independent variables for which you wish to predict the value for the dependent variable. The regression equation calculated from the first data set is used to make the predictions. The predicted values for the dependent variable will be entered directly in the prediction data set; filling the column (or row) for the dependent variable with the predicted values.

Logistic Regression Report

0	0 - 0 - 0	- @-) =				Book19 - N	/licrosoft Exc	el				-		x
C	🤊 💽 Home	Insert	Page Lay	out	Formulas	Data Review	View	Add-Ins	StatTools			0 -	•	x
	mil mail	T Summary	Statistics	· .	Normality Tests	- 1	Quality Con	trol -	Ju Utilities					
		L Summary	Graphs *	128	Time Series and	Forecasting *	Nonparame	teric Tests *	Help *					
Da	ta Set Data	A Statistics	Inference	- 1.·/	Pagrarrian and	Classification x	•							
Ma	Data	Jun Statistica	minerence	N.	Regression and	classification			Teels					
_	Data Analyses Tools												-	
	A1	- (C	fx	Stat	Tools								_	*
	A		В		С	D	E	F		G	Н	I		
1	S	StatTools	(Core Analy:	sis Pack	<)									
2		Analysis:	Logistic Reg	ression	1									
3		Performed By:	Test	21070000										
4		Undating:	Monday, Fe Static	bruary	16,2009									
6		oposeing:	ordin.										-	
7														=
8	Summary Measures	5												
9	Null Deviance		105.4941	1869										
10	Model Deviance		26.83613	3587										
11	Improvement		78,6580	5103										
12	n-Value		< 0.00	01										
13														
14					Standard	Wald		Low	er	Joper				
15	Regression Coefficie	ents	Coefficie	ent	Error	Value	p-Value	Lim	it 🇨	Limit	Exp(Coef)			
16	Constant		-7.93043	8797	3.10005134	-2.558163697	0.0105	-14.006	53942 -1.85	4338171	0.000359629			
17	Income		-9.19596	E-05	5.30141E-05	-1.734625165	0.0828	-0.0001	95867 1.1	948E-05	0.999908045			
18	InvestAmt		0.000351	1829	9.60677E-05	3.662306752	0.0002	0.00016	3537 0.00	0540122	1.000351891			
19														
20			1		0	Percent								
21	Classification Matr	rix				Correct	40							
22	1		23		4	85.19%								
23	0		3		54	94.74%								
24														
25			Percer	nt										
26	Summary Classifica	ation	Percer											
27	Correct		91.67	%										
28	Base		67.86	%										
29	Improvement		74.07	%										
30														
31					11011111111		Analysis	Origi	nal					
32	Probabilities and C	lassifications	Incom	e	InvestAmt	Probability	Class	Clas	s					
33	1		6640	0	26900	1.02%	0	0						
34	2		6800	U	/100	0.00%	0	0						
35	3		5490	0	21500	0.44%	0	0						
36	4		5060	0	19300	0.30%	0	0						
3/	5		5410	0	16/00	0.09%	0	0						
38	6	Degrees'	/820	U	31900	1.99%	0	0						
	Logisti	c Regression	1/24/										-	U.
Rea	ady										100%		÷	

The report above includes the original data plus data used for classification. The predicted classifications, in column E, are based on whether the estimated probabilities of "success", in column D, are above or below a cutoff value of 0.5, or 50%. The report lists summary statistics for the regression (somewhat similar to R-square for multiple regression), detailed information about the regression equation, and summary results of the classification procedure. (There are plenty of cell comments to help you interpret the results). In this example, we see that 90.5% of the observations are classified correctly. Of course, we are hoping to make this percentage as large as possible. Generally, the only way to improve the results is to use more (or better) explanatory variables. The values in column H **Exp (Coeff)** are generally used to interpret the regression equation. They indicate the estimated change in the odds of "success" when any explanatory variable increases by 1 unit.

- **Missing Data -** If there are missing values, then any row with missing values for *any* of the selected variables is ignored.
- Link to Data There is no link to the original data. If the data change, you must rerun the analysis.
Discriminant Analysis Command

Runs a discriminant analysis on a set of variables

The Discriminant Analysis command performs a discriminant analysis on a data set. In this analysis there is a "category" variable that specifies which of two or more groups an observation is in, plus one or more explanatory variables that can be used to predict group membership. There are two ways to predict group membership. The more general way, valid for any number of groups, is to calculate the "statistical distance" of each observation to the mean of each group and to classify the observation according to the smallest statistical distance. A second method, used for the case of two groups, is to calculate a discriminant function (a linear expression of the explanatory variables) and to classify each observation according to whether its discriminant value is less than or greater than some cutoff value. This second method also allows you to specify prior probabilities of group membership, as well as misclassification costs. Then the classification procedure is equivalent to minimizing the expected cost of misclassification.

Discriminant Analysis Dialog Box This analysis is set up using the Discriminant Analysis dialog box:

StatTools - Discriminant Analysis 🛛 🛛 🔀										
Variables (Select One Dependent and One or More Independents)										
Data Set #1 Eormat										
I D Name	Address									
Person	B14:B97									
Income	C14:C97									
InvestAmt	D14:D97									
WSJSubscriber	E14:E97									
Options										
✓ Include Classification Summary										
✓ Include <u>Variances</u> and Covariance	es									
✓ Include <u>Classification Results</u>										
✓ Use <u>M</u> isclassification Table (Two	Categories Only)									
Include Prediction	Include Prediction									
	OK Cancel									

One dependent variable (D) and one or more independent variables (I) need to be selected for analysis. The data must be in "unstacked" form. Variables can be from different data sets.

The options in the Discriminant Analysis dialog include:

- Include Classification Summary with the regression report.
- Include Variances and Covariances with the regression report.
- Include Classification Results with the regression report.
- **Use Misclassification Table**, selected when you wish to change prior probabilities or misclassification costs.
- Include Prediction, where predicted values for the dependent variable are generated for the independent variable values in a second data set. This prediction data set must have the same variable names as the original data set that the regression is analyzing. Typically, in the prediction data set, you will have sets of values for independent variables for which you wish to predict the value for the dependent variable. The regression equation calculated from the first data set is used to make the predictions. The predicted values for the dependent variable will be entered directly in the prediction data set; filling the column (or row) for the dependent variable with the predicted values.

Misclassification
Costs Dialog BoxIf there are exactly two groups possible for the dependent category
variable (as in this example) and the Use Misclassification Table
option is selected, a dialog box appears that allows you to specify
prior probabilities and/or misclassification costs. The default settings
are that each group is equally likely and that the misclassification
costs are equal, but you can override these settings.

rameters 🛛 🔀
0.5
1
1
Cancel

Discriminant Analysis Report

C. 2	- 🔯 - 🖯 -		Book20 -	Microsoft Excel				-	
🙂 💽 Home	Insert Page Lay	out Formulas	Data R	eview View	Add-Ins	StatTools		🥝 –	6
	T Summary Statistics	• A Normality T	ests *	Cuality Co	ontrol *	J Utilities *			
÷	Lummony Graphs x	UN Time Caries	and Encoraction	A Monnaran	antonic Tacto v	() Halp r			
ata Set Data	Summary Graphs	I Z	and rorecasting	A Northaran	ietene rests	(neip			
anager Utilities *	Statistical Inference	Regression	and Classification	n -					
Data			Analyses			Tools			
A1	• (* fx	StatTools							
A	В	С	D	E	F	G	Н	1	
StatT	Core Analysis Pa	ck)							
. F	nalysis: Discriminant								
Perior	Date: Monday, Februar	y 16, 2009							
U	odating: Static	Sector Contents							
i									
	Prior	Misclassifi	cation Cost						
Setup Variables	Probability	No	Yes	c.					
0 10	0.5	1	0						
1	0.5	-	U						
2	Sample	Mean	Mean						
3 Sample Summary	Size	Income	InvestAmt						
4 No	57	66042.10526	24952.63158						
.5 Yes	27	80485.18519	53000						
.6									
7									
.8 Discriminant Functi	on Coefficient	-							
19 Income	6.586E-05								
20 InvestAmt	-0.000352733								
22									
23 Classification Matri	x No	Yes	Correct						
24 No	52	5	91.2%						
25 Yes	2	25	92.6%						
26									
27									
8 Summary Classifican	ion	-							
29 Correct	91.7%								
U Base	7/ 1%								
1 improvement	/4.1/0								
33									
A Matrix of Vars and C	ovars Income	InvestAmt							
5 No	101 C 70 C 2000	0.0000.0000							
6 Income	209308552.6	79303458.65							
7 InvestAmt	79303458.65	99485394.74							
88 Yes									
9 Income	118875156.7	64698846.15							
II Declad	64698846.15	80470769.23							
PUOIED	180634549	74672727 96							
2 Income	74672727.86	93456367.14							
2 Income 3 InvestAmt									
2 Income 3 InvestAmt									
2 Income 3 InvestAmt 4			StatDist	StatDist		Original	Analysis		
2 Income 3 InvestAmt 4 5 5 5 5 5 5 5 5 5 5 5 5 5	Income	InvestAmt	StatDist No	StatDist Yes	Discrim	Original Class	Class		
12 Income 13 InvestAmt 14 15 16 Statistical Distance 17	Income 66400	InvestAmt 26900	StatDist No 0.229005023	StatDist Yes 2.769086119	Discrim	Original Class . NO	Class No	- 11	
12 Income 13 InvestAmt 14 15 16 Statistical Distance 17 1 18 2	Income 66400 68000	InvestAmt 26900 7100	StatDist No 0.229005023 2.363424601	StatDist Yes 2.769086119 5.232611294	Discrim -5.115462701 1.974062536	Original Class NO NO	Class No No	-	
12 Income 13 InvestAmt 14 15 15 5 16 Statistical Distance 17 1 18 2 19 3	Income 66400 68000 54900	InvestAmt 26900 7100 21500	StatDist No 0.229005023 2.363424601 0.84174592	StatDist Yes 2.769086119 5.232611294 3.258632671	Discrim -5.115462701 1.974062536 -3.968084673	Original Class NO NO NO	Analysis Class No No No	6	

The Discriminant Analysis report includes the original data plus data used for classification. The predicted classifications are based on whether the discriminant values shown are below or above a cutoff value. If the prior probabilities and misclassification costs are left at their default values, then this classification procedure is equivalent to basing classification on the smaller of the two statistical distances. In fact, if there were more than two groups, then the discriminant values would not appear, and classification would be based on the smallest of the statistical distances. The report also shows descriptive statistics for the groups and the coefficients of the discriminant function (it does this only when there are two groups), the prior probabilities, misclassification costs, and cutoff value for misclassification (again, only if there are two groups), and the summary results of the classification procedure (with cell comments to help you interpret the results). In this example, we see that 89% of the observations are classified correctly. Of course, we are hoping to make this percentage as large as possible. Generally, the only way to improve the results is to use more (or better) explanatory variables

Missing Data and Link to Data

- **Missing Data -** If there are missing values, then any row with missing values for *any* of the selected variables is ignored.
- Link to Data There is no link to the original data. If the data change, you must rerun the analysis.

Quality Control Menu

The procedures on the Quality Control menu deal with the analysis of data collected over time, with applications in quality control.

The Pareto chart displays the relative importance of categorized data.

The four types of control charts plot time series data and allow you to see whether a process is in statistical control. We can see whether the data stay within the control limits on the chart, and we can also check for other nonrandom behavior such as long runs above or below the centerline.

Pareto Chart Command

Creates Pareto chart for categorized variable

Pareto charts are useful for determining the most significant items in a group of categorized data, as well as conveying a quick visual representation of their relative importance. Typically Pareto charts are used in the area of Quality Assurance to determine the few factors which have the most significance (Pareto's 80/20 rule).

For example, a manufacturer of machine parts has decided to investigate why customers have been rejecting a particular product. When each batch is returned, a reason ("wrong size", "incorrect surface finish", etc.) is entered. After several months of data have been collected, a Pareto chart is plotted. Action is taken to address the largest sources of problems.

StatTools allows you to create Pareto charts based on data in one of two formats – Category Only, or Category and Value. A Category Only variable will typically contain one entry for each reading. In the example above, each cell would correspond to the reason a batch of parts was returned. A cell value might be "incorrect surface finish" and there would likely be many duplicated cells. StatTools will count the number of times each entry appears in the variable and create the corresponding Pareto chart. When Category and Value is selected, the variables you specify are the categories and each corresponding count.

The axes of the Pareto chart are constructed as follows:

- Categories are placed along the horizontal axis
- Frequency (or count) is placed along the left vertical axis
- Cumulative percentage is placed along the right vertical axis

Pareto Chart Dialog Box

StatTools	- Pareto Chart			
Data <u>T</u> ype	Category and Value	•		
Variables (Select One Category and	One Value)		
<u>D</u> ata Set	Data Set #1		-	
Cat Val	Name	Add	ress	
🗹 🗆 R	eason for Failed Delivery	B14:B19		
🗆 🗹 Ir	stances of Failure	C14:C19		
Category	Options			
• None				
C Include	Additional Category with	Fixed Value of		
C Merge	All Categories with Values	Less than or Equal to		
0	2		ОК	Cancel

The options in the Pareto Chart dialog include:

• **Data Type**. Selects the type of data used to construct the Pareto chart – Category and Value or Category Only.

Category Options in the Pareto Chart dialog include:

- **None** each distinct category will be represented by a bar in the Pareto chart.
- Include Additional Category with Fixed Value of a bar labeled "Misc" will be added at the extreme right side of the Pareto chart with a frequency equal to the specified value.
- Merge All Categories with Values Less Than or Equal to all categories whose frequency is less than or equal to the specified value will be combined into a category labeled "Misc" and placed at the extreme right side of the Pareto chart.



Missing Data and Link to Data

- **Missing Data -** If there are missing values, then any row with missing values for *any* of the selected variables is ignored.
- Link to Data There is no link to the original data. If the data change, you must rerun the analysis.

X/R Charts Command

Creates X and R control charts for time series variables

This analysis produces X-bar and R charts for time series data. It assumes that data have been collected in small subsamples over time. For example, an operator might collect measurements on the widths of four randomly selected parts every half hour. The subsample size is then 4. If data are collected for 50 half-hour periods, then the data should be arranged in four adjacent columns and 50 adjacent rows, with variable headings such as SubSamp1 through SubSamp4 above the first row of data.

The purpose of the procedure is to check whether the process that is generating the data is in statistical control. To do so, the procedure first calculates an X-bar and an R for each row in the dataset. X-bar is the average of the observations in that row, and R is the range (maximum minus minimum) for the observations in that row.

The X-bars and R's are charted in separate time series plots around centerlines. The centerline for the X-bar chart is the average of the X-bars (sometimes called X-double-bar), and the centerline for the R chart is R-bar, the average of the R's. A simple way to check whether the process is in control is to see whether any of the X-bars or R's fall outside their respective upper and lower control limits (UCL and LCL), which are approximately plus or minus 3 standard deviations from the centerlines. The charts show these control limits, so that it is easy to spot any extreme values.

The procedure also allows you to check for other possible out-ofcontrol behavior, including 8 or more points in a row above or below the centerline, 8 or more points in a row in an uphill or downhill direction, at least 4 of 5 points in a row more than one standard deviation from the centerline, and at least 2 of 3 points in a row more than two standard deviations from the centerline.

These graphs are set up	using the XBar and R	Control Charts dialog
box:		

Stat	StatTools - XBar and R Control Charts								
<u>C</u> hart	Chart Type X-Bar/R Chart Variables (Select Two or More)								
<u>D</u> ata	a Set Data Set #1		- Eorma	at					
	Name		Address						
	Sample	A4:A73							
	Obs1	B4:B73							
	Obs2	C4:C73							
	Obs3	D4:D73							
	Obs4	E4:E73							
	Obs5	F4:F73		-					
	Graph Options Image: Sigma 1 Control Limits Image: Sigma 2 Control Limits								
Control Limit Calculations Based On C All Observations C Observations In Range C Previous Data									
0			OK Cance	ł					

Two or more variables can be selected for analysis. The selected data set must be unstacked data. Variables can be from different data sets.

Graph Options in the XBar and R Control Charts dialog include:

- Sigma 1 and 2 Control Limits Adds control limit lines at one and/ or two sigmas from mean line. These extra lines allow you to check for other types of out-of-control behavior (the so-called "zone" rules).
- Zone A and B Analysis number of points beyond Zone A (2 sigma) and Zone B (1 sigma)
- Runs Up/Down and Runs Above/Below Analyses sequential up or down moves of length 8 or greater
- Limit Graph Range From Index Limits the points on the graph to a range of data points for a variable (i.e., range from starting index to ending index)

XBar and R Control Charts Dialog Box **Control Limit Calculations Based On** options in the XBar and R Control Charts dialog determine the data on which control limit calculations will be based, including:

- All Observations use all available data in control limit calculations
- Observations in Range use data between Start Index and Stop Index in control limit calculations
- **Previous Data** creates control limits from previously observed data. Simply enter the **Subsample Size**, **Average R** and **Average X-Bar** that was calculated from the previous data.





Example R Chart

Note: Page Layert Famely Layert	2 50 1100						Carl Contract Contract of the	PC#I						
Image: Section of Content of Con	real roome	Iniert Pag	pe Layout 7	Formulas Data	Review	View Add-b	ns Statiools						w -	
A1 - (m A Surfaces A B C D E F G H I J K L M R (bit1 - bits) K L D E F G H I J K L M a1 a2 b2 b b b b J K L M a2 a3 a3 a5	ta Set Data nager Utatise - Data	 X) Summary State Summary Grap Statistical Infe 	nations * 🔺 Nam phos * 🔡 The prence * 🔡 Re	rmality Tests + ne Series and Fores gression and Class Analyses	anting * 🔺 No Reation *	uality Control + Imparameteric Test	y ^p Usakes hr * W Help * Toolk							
A B C D E P G H I J K L MI # B C D E P G H I J K L MI # B C D E P G H I J K L MI # B C D E P G H I J K L MI ## B C D E P G H I J K L MI ## D<	AI	• (*	∫₂ StatTo	ols										
$\frac{1}{10000000} \frac{1}{10000000000000000000000000000000000$	A	8	c	D	£		6	н	1	1	ĸ	L	M	
Tark Annue 0 Sector Processing Control Control Upgeter Lineer Upgeter Lineer Right Control Right Control <th cols<="" th=""><th></th><th>A. Mark</th><th>μW</th><th>M</th><th></th><th>nnar Line nnisi Linux</th><th></th><th></th><th></th><th></th><th></th><th></th><th></th></th>	<th></th> <th>A. Mark</th> <th>μW</th> <th>M</th> <th></th> <th>nnar Line nnisi Linux</th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th>		A. Mark	μW	M		nnar Line nnisi Linux							
Dightment manual programme manuprogramme manual programme manual programme manual pro	0 10	20 30	40 50	60 70	80									
Alsent/Retext Field 1 Date Range Cantor Upper Losser Control Contro Contro Control	0 10	20 30	40 50	60 70	80									
Name Range Losset Upper Losset Register	0 10 Test Results Up/Down Runs	20 30	40 50	60 70	80									
Josef Josef Line Online Data	0 0 10 Test Results Up/Down Runs Above/Below Runs	20 30 0 1	40 50	60 70	80									
1 12.0404 0.145 12.0457714 11.0505666 12.1181479 11.5454635 12.118471 12.00034021 12.0195647 0.235525571 0 0.44975 12.04074 0.248 12.0457714 11.0505686 12.1181479 11.5454635 12.118471 12.00034021 12.0195647 0.23552571 0 0.44975 12.04074 0.148 12.0457714 11.0505686 12.1181479 11.54546359 12.118471 12.00034021 12.0195647 0.23552571 0 0.44975 12.04074 0.148 12.0457714 11.0505686 12.1181479 11.54546359 12.118471 12.00034021 12.0195647 0.23552571 0 0.44975 12.04074 0.148 12.0457714 11.0505686 12.1181479 11.54546359 12.118471 12.00034021 12.0915647 0.23552571 0 0.44975 12.04074 0.147 12.0457714 11.0505686 12.1181479 11.54546359 12.118471 12.00034021 12.0915647 0.23552571 0 0.44975 12.00140 0.148 12.0457714 11.0505686 12.1181479 11.54546359 12.118471 12.00034021 12.0915647 0.23552571 0 0.44975 12.00140 0.148 12.0457714 11.0505686 12.1181479 11.54546359 12.118471 12.00034021 12.0915647 0.23552571 0 0.44975 12.00140 0.148 12.0457714 11.0505686 12.1181479 11.54546359 12.118471 12.00034021 12.0915647 0.23552571 0 0.44975 12.00140 0.148 12.0457714 11.0505686 12.1181479 11.54546359 12.118471 12.00034021 12.0915647 0.23552571 0 0.44975 12.00140 0.148 12.0457714 11.0505686 12.1181479 11.54546359 12.118471 12.00034021 12.0915647 0.23552571 0 0.44975 12.00140 0.148 12.0457714 11.0505686 12.1181479 11.54546359 12.118471 12.00034021 12.0915647 0.23552571 0 0.44975 12.00140 0.148 12.0457714 11.0505686 12.1181479 11.54546359 12.118471 12.00034021 12.0915647 0.23552571 0 0.44975 12.00140 0.148 12.0457714 11.0505686 12.1181479 11.15456859 12.118471 12.00034021 12.0015674 0.23552571 0 0.44975 12.00140 0.148 12.0457714 11.0505686 12.1181479 11.15456859 12.118471 12.00034021 12.0015674 0.23552571 0 0.44975 12.00140 0.148 12.045774 11.0505686 12.1181479 0.114479 0.114477 0.114677 0.23552571 0 0.44975 12.00140 0.114979 0.133 0.016774 0.114679 0.114477 0.114677 0.114677 0.12003407 0.21505571 0 0.44975	0 10 Test Results Up/Down Runs Above/Below Runs	20 30	40 50	60 70 Center	50 Lower	Upper	Lower	Upper	Lower	Upper	RCenter	RLower	RUpper	
1 12.044 0.288 12.04577174 11.95596816 12.1814779 11.95496129 12.181477 12.00014021 12.09195407 0.239255571 0 0.48975 1 12.0052 0.189 12.0057714 11.95956861 12.1814779 11.95496129 12.181477 12.00014021 12.09195407 0.23925571 0 0.48975 1 12.00517 11.95956861 12.1814771 11.95496129 12.181477 12.00014021 12.09195407 0.23925571 0 0.48975 1 12.0051771 11.95956861 12.1814771 11.95496129 12.181477 12.00014021 12.09195407 0.23925571 0 0.48975 1 12.0051771 11.95956861 12.1814771 11.95496129 12.181471 12.00014021 12.09195407 0.23925571 0 0.48975 1 12.00517714 11.95956861 12.1814771 11.95496129 12.181471 12.00014021 12.09195407 0.23925571 0 0.48975 1 12.0051 12.00517714	0 10 Test Results Up/Down Runs Above/Below Runs Centrel Data	20 30 0 1 Xillar	40 50 Range	60 70 Center Line	E0 Lower Ctri Limit	Upper Ctol Limit	Löwer Sig2 Limit	Upper Sig2 Limit	Lower Sig1 Limit	Upper Sig1Limit	R Center Line	R Lower Ctri Limit	R Upper Cerl Limit	
1 1.0.0002 0.1.79 1.0.0057/17.4 1.1.0000800 1.0.104771 1.0.000871/1.4 0.0.00080 1.0.000871/1.4 0.0.00080 1.0.000871/1.4 0.0.000802 1.0.00080200 1.0.0008020 1.0.0	0 0 10 Trat Results Up/Down Runs Above/Below Runs Eentre/Date	20 30 0 1 XBar 12,0404	40 50 Range 0.185	60 70 Canter Line 12.04571714	E0 Lower Ctrl Limit 11.90958636	Upper Ctri Limit 12.18184793	Löwer Fig2 Limit 11.95496329	Upper Sig2 Limit 12.136471	Lower Sigt Limit 12.00034021	Upper 5ig11mit 12:09109407	R Center Line 0.235928571	RLOwer Ctr/Limit 0	R Upper Ctrl Limit 0.49975	
Lossme Lossme <thlossme< th=""> <thlossme< th=""> <thlossme< td="" th<=""><td>0 10 Fest Results Up/Down Runs Above/Below Runs Control Date 1 2</td><td>20 30 0 1 12.0404 12.0404</td><td>40 50 Range 0.185 0.268 0.268</td><td>60 70 Cantar Line 12.04571714 12.04571714</td><td>Lower Criticant 11.9055636 11.9055836</td><td>Upper Cert Limit 12.18184793 12.18184793</td><td>Lower Eg21init 11.95496329 11.95496329</td><td>Upper 5g2 Limit 12.136471 12.136471</td><td>Lower Sigt Limit 12.00034021 12.00034021</td><td>Upper Sig11imit 12.09109407 12.09109407</td><td>RCenter Line 0.235928571 0.235928571</td><td>RLower CertLimit 0 0</td><td>R Upper Cerl Limit 0.49875 0.49875</td></thlossme<></thlossme<></thlossme<>	0 10 Fest Results Up/Down Runs Above/Below Runs Control Date 1 2	20 30 0 1 12.0404 12.0404	40 50 Range 0.185 0.268 0.268	60 70 Cantar Line 12.04571714 12.04571714	Lower Criticant 11.9055636 11.9055836	Upper Cert Limit 12.18184793 12.18184793	Lower Eg21init 11.95496329 11.95496329	Upper 5g2 Limit 12.136471 12.136471	Lower Sigt Limit 12.00034021 12.00034021	Upper Sig11imit 12.09109407 12.09109407	RCenter Line 0.235928571 0.235928571	RLower CertLimit 0 0	R Upper Cerl Limit 0.49875 0.49875	
2000000 000007 21.00007174 0.100007021 12.00007021 12.00007021 0.00000701	0 20 Test Results Up/Down Runs Makove/Below Runs Centrel Date 1 2 3	20 30 0 1 XBar 12.0404 12.044 12.0502	40 50 Range 0.185 0.268 0.179	60 70 Canter Line 12.04571714 12.04571714 12.04571714	Lower Col Limit 11.90956636 11.90956636 11.90956636	Upper 12.18184793 12.18184793 12.18184793	Lawer Bg2 Linest 11.55456329 11.55456329 11.55456329	Upper <u>Fig2 Limit</u> 12.136471 12.136471 12.136471	Lower Sigt Limit 12.00034021 12.00034021	Upper 52,9510=407 12.09109407 12.09109407	R Center Line 0.235928571 0.235928571 0.235928571	RLower CtriLimit 0 0 0	R Upper Cert Limit 0.49875 0.49875 0.49875	
12.0082 0.239 12.04571714 11.9595959 12.118579139 12.106971 12.0001940 12.01919897 0.239938971 0 0.49975 12.014 0.144 12.04571714 11.95959591 12.118577 11.95959519 12.118477 11.95959519 12.118477 11.95959519 12.118477 11.959591957 11.959591957 094875 11 12.065 0.119759774 11	0 0 10 Test Results Up/Down Runs Above/Below Runs Control Dato 1 2 3 4	0 1 XBar 12.0404 12.0404 12.0502 12.0592	40 50 Range 0.185 0.268 0.179 0.186 0.268	60 70 Center Line 12.04571714 12.04571714 12.04571714 12.04571714	Lower Cot Limit 11.9055636 11.9055636 11.9055636 11.9055636	Upper Ctrl Limit 12.18184793 12.18184793 12.18184793 12.18184793	Löwer Fig2 Limit 11.95496329 11.95496329 11.95496329 11.95496329	Upper 5ig2 Limit 12.136471 12.136471 12.136471 12.136471	Louar Sig1Limit 12.00034021 12.00034021 12.00034021 12.00034021	Upper <u>Sig11mm</u> 12.09109407 12.09109407 12.09109407 12.09109407	RCentar Line 0.235928571 0.235928571 0.235928571 0.235928571 0.235928571	R Louver Ctri Linvit 0 0 0 0	R Upper Cert Limit 0.49875 0.49875 0.49875 0.49875	
12.0014 0.144 12.0457/174 11.95956956 12.1814793 11.55496929 12.18477 12.00034021 12.09199407 0.235928571 0 0.49875 9 12.035 0.166 12.04577174 11.959568956 12.1814793 11.54949529 12.18477 12.00034021 12.0919407 0.235928571 0 0.49875 10 12.0996 0.166 12.04577174 11.959568165 12.1814789 12.184787 12.15449167 12.00034021 12.0919407 0.235928571 0 0.49875 11 12.0457 12.04517174 11.959568165 12.1814789 11.64491579 12.164471 12.00034021 12.0919407 0.235928571 0 0.49875 11 12.045 0.120177174 11.95956816 12.1814793 11.6449739 12.164717 12.00034021 12.0919407 0.235928571 0 0.49875 11 12.0457774 11.95956816 12.1814793 11.54497395 12.181471 12.00034021 12.000194021 0.235928571 0 0.49875 <	0 0 20 Test Results Up/Down Runs Above/Below Runs Control Data 2 3 4 5 5	0 1 12.0404 12.0404 12.0404 12.0502 12.0994 12.0102 12.0274	40 50 Range 0.185 0.268 0.179 0.186 0.357 0.187	60 70 Center Line 12.04571714 12.04571714 12.04571714 12.04571714 12.04571714	Lower Col Level 11.9055636 11.9055636 11.9055636 11.9055636 11.9055636	Upper Col Linit 12.18184793 12.18184793 12.18184793 12.18184793 12.18184793 12.18184793	Löveter Fig2 Lavit 11.55496329 11.55496329 11.55496329 11.55496329 11.55496329 11.55496329	Upper 5g2 Limit 12.136471 12.136471 12.136471 12.136471 12.136471 12.136471	Louar Sigliati 12.0034021 12.00034021 12.00034021 12.00034021 12.00034021 12.00034021	Upper Sq11mit 12.09109407 12.09109407 12.09109407 12.09109407 12.09109407	RCenter Line 0.235928571 0.235928571 0.235928571 0.235928571 0.235928571 0.235928571	R Lower Ctri Limit 0 0 0 0 0	R Upper Cert Limit 0,49875 0,49875 0,49875 0,49875 0,49875	
12.035 0.166 12.0457/174 11.95959359 12.1814773 11.55959375 12.100014021 12.0	0 0 10 Test Results Uty/Down Runs Above/Below Runs Control Date 1 2 3 4 5 5 6 7	0 1 12.0404 12.0404 12.0404 12.0502 12.0594 12.0102 12.0724 12.0102	40 50 Range 0.185 0.268 0.179 0.185 0.357 0.187 0.293	60 70 Canter Line 12.04571714 12.04571714 12.04571714 12.04571714 12.04571714 12.04571714	Lower Cort Limit 11.90558636 11.90558636 11.90558636 11.90558636 11.90558636 11.90558636	Upper Col Limit 12.18184793 12.18184793 12.18184793 12.18184793 12.18184793 12.18184793	Lawer Eg2 Limit 11.55496329 11.55496329 11.55496329 11.55496329 11.55496329 11.55496329	Upper 5g2 Limit 12.136471 12.136471 12.136471 12.136471 12.136471 12.136471 12.136471	Lower Sigt Limit 12.00034021 12.00034021 12.00034021 12.00034021 12.00034021 12.00034021	Upper Sp21 Limit 12.09109407 12.09109407 12.09109407 12.09109407 12.09109407 12.09109407	R Center Line 0.235928571 0.235928571 0.235928571 0.235928571 0.235928571 0.235928571	R Lower Ctri Limit 0 0 0 0 0 0 0 0 0 0	R Upper Ctri Limit 0.45875 0.45875 0.45875 0.45875 0.45875 0.45875	
10 12.0596 0.366 12.04571714 11.95956936 12.18144793 11.55459039 12.184471 12.0004021 12.09109407 0.2355928571 0 0.49875 11 1.02055 0.322 12.04571714 11.959586936 12.1814979 11.55459529 12.181471 12.0004021 13.00010001 0.315001001 0.048975 11.0410714 10.041071 0.041071 0.0410710 0.041071 0.041071 0.0004001 13.00010001 0.315001001 0.01000101 0.0100010	0 0 10 Test Results Up/Down Runs Above/Delow Runs Control Date 1 2 3 4 5 6 7 8	0 1 12.0404 12.0404 12.0404 12.0502 12.0794 12.0102 12.0724 12.0032 12.0724	40 50 Range 0.185 0.268 0.179 0.186 0.357 0.187 0.293 0.164	60 70 Center Line 12.04571714 12.04571714 12.04571714 12.04571714 12.04571714 12.04571714 12.04571714	Lower Certiant 11.90358636 11.90358636 11.90358636 11.90358636 11.90358636 11.90358636 11.90358636	Upper Cert Limit 12:18184793 12:18184793 12:18184793 12:18184793 12:18184793 12:18184793 12:18184793	Lawer Teg2 Linet 11.55496329 11.55496329 11.55496329 11.55496329 11.55496329 11.55496329 11.55496329	Upper SetLimit 12.136471 12.136471 12.136471 12.136471 12.136471 12.136471 12.136471	Louser Sigt Limit 12.00034021 12.00034021 12.00034021 12.00034021 12.00034021 12.00034021 12.00034021	Upper Stat Limit 12.09109407 12.09109407 12.09109407 12.09109407 12.09109407 12.09109407 12.09109407	#Center Line 0.235928571 0.235928571 0.235928571 0.235928571 0.235928571 0.235928571 0.235928571	R.Lower Cer/Limit 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	R Upper Carl Limit 0.45875 0.45875 0.45875 0.45875 0.45875 0.45875 0.45875	
11 12.065 0.322 12.04571714 11.9955636 12.18184793 11.95496329 12.136471 12.0034021 12.09109407 0.235928571 0 0.49875	0 0 10 Terr Results Up/Down Runs Above/Below Runs Control Data 1 2 3 4 5 6 7 8 9	0 1 30 30 30 30 30 30 30 30 30 30 30 30 30	40 50 Range 0.185 0.268 0.179 0.186 0.357 0.187 0.187 0.189 0.293 0.66	60 70 Line 12.04571714 12.04571714 12.04571714 12.04571714 12.04571714 12.04571714 12.04571714	Lower Certilent 11.9055636 11.9055636 11.9055636 11.9055636 11.9055636 11.9055636 11.9055636 11.9055636	Upper Ctri Lient 12.18184793 12.18184793 12.18184793 12.18184793 12.18184793 12.18184793 12.18184793 12.18184793 12.18184793	Linux Fig2 Linux 11.55496329 11.55496329 11.55496329 11.55496329 11.55496329 11.55496329 11.55496329	Upper 12.136471 12.136471 12.136471 12.136471 12.136471 12.136471 12.136471 12.136471 12.136471 12.136471	Louar Sigl Limit 12.00034021 12.00034021 12.00034021 12.00034021 12.00034021 12.00034021 12.00034021 12.00034021 12.00034021	Upper Sig11emit 12.09109407 12.09109407 12.09109407 12.09109407 12.09109407 12.09109407 12.09109407	8 Center Une 0.235928571 0.235928571 0.235928571 0.235928571 0.235928571 0.235928571 0.235928571 0.235928571	RLower CertLivit 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	RUpper Cert Limit 0.45975 0.45975 0.45975 0.45975 0.45975 0.45975 0.45975 0.45975	
10 01 0 000 0 010 0 010 0 000 0 0 0 000 0 0 0 0 0 0 0 0 0 0 0	0 20 Test Results UkyDown Rans Above/Telow Rans Above/Telow Rans 2 2 3 4 5 5 6 7 8 9 10	0 1 38w 12.0404 12.0594 12.0594 12.0594 12.002 12.0724 12.0024 12.0014 12.0014 12.0696	40 50 Range 0.185 0.265 0.357 0.187 0.357 0.187 0.293 0.357 0.164 0.366	Cantar Line 12.04571714 12.04571714 12.04571714 12.04571714 12.04571714 12.04571714 12.04571714	Lover Crittani 11.9055636 11.9055636 11.9055636 11.9055636 11.9055636 11.9055636 11.9055636	Upper Coti Limit 12.18154793 12.18154793 12.18154793 12.18154793 12.18154793 12.18154793 12.18154793 12.18154793 12.18154793	Lower <u>Bg2 Linkt</u> 11.55496329 11.55496329 11.55496329 11.55496329 11.55496329 11.55496329 11.55496329 11.55496329	Upper Fig2 Linit 12.136471 12.136471 12.136471 12.136471 12.136471 12.136471 12.136471 12.136471 12.136471 12.136471	Lower Sigt1 Limit 12.00034021 12.00034021 12.00034021 12.00034021 12.00034021 12.00034021 12.00034021 12.00034021 12.00034021	Upper Set Limit 12.09103407 12.09103407 12.09103407 12.09103407 12.09103407 12.09103407 12.09103407 12.09103407	R Center Line 0.235928571 0.235928571 0.235928571 0.235928571 0.235928571 0.235928571 0.235928571 0.235928571 0.235928571	RLower CertLinit 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	R Upper Ctri Limit 0.49875; 0.49875; 0.49875; 0.49875; 0.49875; 0.49875; 0.49875; 0.49875;	
	0 0 20 Text Results Mover/Below Runs Above/Below Runs Construct Dato 1 2 3 4 5 6 7 8 9 10 11	0 1 20 30 1 2.0404 12.0404 12.0592 12.0094 12.0092 12.0094 12.0095 12.0696	40 50 Range 0.185 0.268 0.179 0.186 0.357 0.293 0.166 0.366 0.366 0.322	60 70 Exeter Exe 12.64571714 12.04571714 12.04571714 12.04571714 12.04571714 12.04571714 12.04571714 12.04571714 12.04571714	00 00 01 00 01 00 056636 11.0055636 11.0055636 11.0055636 11.0055636 11.0055636 11.0055636 11.0055636 11.0055636 11.0055636	Upper Ortlinet 12.18154793 12.18154793 12.18154793 12.18154793 12.18154793 12.18154793 12.18154793 12.18154793 12.18154793	Lower <u>Fig Linen</u> <u>1135496329</u> 11.55496329 11.55496329 11.55496329 11.55496329 11.55496329 11.55496329 11.55496329	Upper 5g2 Linet 12.136471 12.136471 12.136471 12.136471 12.136471 12.136471 12.136471 12.136471 12.136471 12.136471	Louwer Sigt Limit 12.00034021 12.00034021 12.00034021 12.00034021 12.00034021 12.00034021 12.00034021 12.00034021 12.00034021 12.00034021	Viger 941Um 12.05405407 12.05405407 12.05105407 12.05105407 12.05105407 12.05105407 12.05105407 12.05105407	R Cuntar Lisse 0.235928571 0.235928571 0.235928571 0.235928571 0.235928571 0.235928571 0.235928571 0.235928571 0.235928571 0.235928571	RLower CertLinit 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	R Upper Ctrl Limit 0.45875; 0.45875; 0.45875; 0.45875; 0.45875; 0.45875; 0.45875; 0.45875; 0.45875; 0.45875;	

Missing Data and Link to Data

- **Missing Data -** Missing data are not allowed.
- Link to Data Graphs are not linked to data.

P Chart Command

Creates P charts for time series variables

P charts are for "attribute" data. With attribute data each observation indicates the number (or fraction) of items that do not conform to specifications from a sample of items. For example, a process might produce a certain number of items each half hour, some of which are nonconforming. Then a P chart would plot each half hour's fraction of items that are nonconforming. As always, the purpose is to see if the process is in control.

This procedure requires a data set with at least one of the following: a variable that contains the *number* of nonconforming items in each sample or a variable that contains the *fraction* of nonconforming items in each sample. Optionally, there can be a variable that contains the sample sizes. If there is no sample size variable, then you must enter a sample size, which is assumed to be constant across all samples. If there is a sample size variable, however, the sample sizes are not required to be equal.

This graph is set up using the **P Control Charts** dialog box:



P Control Charts Dialog Box One value variable and optionally, a size variable, are selected for analysis. The selected data set must be unstacked data. Variables can be from different data sets.

Input Data options in the P Control Charts dialog include:

- Numbers of Non-Conforming Items Specifies that the Value variable gives the actual number of non-conforming items out of the total sample
- **Fractions of Non-Conforming Items** Specifies that the Value variable gives the fraction of non-conforming items in the sample

Sample Size options in the P Control Charts dialog include:

- **Use Size Variable** Specifies that a size variable is used to give the total size of each sample
- **Use Common Size** Specifies that no size variable is used as each sample is the entered size.

Graph Options in the P Control Charts dialog include:

- Sigma 1 and 2 Control Limits Adds control limit lines at one and/ or two sigmas from mean line. These extra lines allow you to check for other types of out-of-control behavior (the so-called "zone" rules).
- **Zone A and B Analysis** number of points beyond Zone A (2 sigma) and Zone B (1 sigma)
- Runs Up/Down and Runs Above/Below Analyses sequential up or down moves of length 8 or greater
- Limit Graph Range From Index Limits the points on the graph to a range of data points for a variable (i.e., range from starting index to ending index)

Control Limit Calculations Based On options in the P Control Charts dialog determine the data on which control limit calculations will be based, including:

- All Observations use all available data in control limit calculations
- Observations in Range use data between Start Index and Stop Index in control limit calculations
- **Previous Data** creates control limits from previously observed data. Simply enter the **Subsample Size** and **Average P** that was calculated from the previous data.



Example P Chart

- Missing Data and Link to Data
- Missing Data Missing data are not allowed.
- Link to Data Graphs are not linked to data.

C Chart Command

Creates C charts for time series variables

C charts are used to plot the number of defects for items of a constant size. For example, suppose that car doors are produced in batches of 50. In each batch we could count the number of defects (a paint blemish or a rough edge, for example). These counts are then plotted on a chart. As always, the purpose is to check whether the process is in control.

The data set for a C chart must include a variable that contains the count of defects for each item. It is assumed that the item size is equal for each observation. For example, if an "item" is a batch of car doors, then we would assume that each batch has the same number of doors.

This graph is set up using the C Control Charts dialog box:

C Control Charts Dialog Box

Stat	Tools - C Control Char	ts		X
Vari	iables (Select Exactly One)			
<u>D</u> ata	Set Data Set #1		_	<u>F</u> ormat
Val	Name		Address	
	Unit	A3:A52		
	SqFt	B3:B52		
	Blemishes	C3:C52		
Gr	aph Options			
Г	Sigma <u>1</u> Control Limits		🔲 Zone <u>A</u> Analysis	
	Sigma <u>2</u> Control Limits		Zone <u>B</u> Analysis	
	<u>R</u> uns Up / Down Analysis		Runs Above / Below Ana	lysis
	Limit Graph Range From Ind	ex:	to	
C Co	ontrol Limit Calculations Base	d On		
œ	All <u>O</u> bservations			
0	Observations In Range			
0	Previous Data			
0			ОК	Cancel

One or more variables are selected for analysis. The selected data set must be unstacked data. Variables can be from different data sets.

Graph Options in the C Control Charts dialog include:

- Sigma 1 and 2 Control Limits Adds control limit lines at one and/ or two sigmas from mean line. These extra lines allow you to check for other types of out-of-control behavior (the so-called "zone" rules).
- **Zone A and B Analysis** number of points beyond Zone A (2 sigma) and Zone B (1 sigma)
- Runs Up/Down and Runs Above/Below Analyses sequential up or down moves of length 8 or greater
- Limit Graph Range From Index Limits the points on the graph to a range of data points for a variable (i.e., range from starting index to ending index)

Control Limit Calculations Based On options in the C Control Charts dialog determine the data on which control limit calculations will be based, including:

- All Observations use all available data in control limit calculations
- Observations in Range use data between Start Index and Stop Index in control limit calculations
- **Previous Data** creates control limits from previously observed data. Simply enter the **Average C** that was calculated from the previous data.

Example C Chart



Missing Data and Link to Data

- Missing Data Missing data are not allowed.
- Link to Data Graphs are not linked to data.

U Chart Command

Creates U charts for time series variables

U charts are similar to C charts, but now we plot the rate of defects. Using the car door example, suppose the batch sizes are not necessarily equal, that is, different batches have different numbers of car doors. Then in a U chart, we would plot the rate of defects per car door, that is, the number of defects in a batch divided by the number of doors in the batch. As always, the purpose is to check whether the process is in control.

This procedure requires a data set with at least one of the following: a variable that contains the *number* of defects for each observation or a variable that contains the *rate* of defects in each observation. Optionally, there can be a variable that contains the item sizes. If there is no size variable, then you must enter an item size, which is assumed to be the constant item size for all observations. If there is a size variable, however, the item sizes are not required to be equal.

U Control Charts Dialog Box

This graph is set up using the U Control Charts dialog box:

StatTools - U Control Charts	×
Variables (Select One Size and One	Value)
Data Set Data Set #1	Eormat
Siz Val Name	Address
Unit Unit	A3:A52
I SqFt	B3:B52
□ □ □ Pate	D3:D52
	03.032
Input Data	Sample Size
Numbers of Defects	Use Size Variable
C Rates of Defects	C Use Common Size
-	
Graph Options	
Sigma <u>1</u> Control Limits	Zone <u>A</u> Analysis
Sigma <u>2</u> Control Limits	Zone <u>B</u> Analysis
Runs Up / Down Analysis	Runs Above / Below Analysis
Limit Graph Range From Index:	to
Control Limit Calculations Based Or	1
All Observations	
C Observations In Range	
C Previous Data	
0 2	OK Cancel

One value variable and optionally, a size variable, are selected for analysis. The selected data set must be unstacked data. Variables can be from different data sets.

Input Data options in the U Control Charts dialog include:

- **Numbers of Defects** Specifies that the Value variable gives the actual number of non-conforming items out of the total sample.
- **Rates of Defects** Specifies that the Value variable gives the fraction of non-conforming items in the sample.

Sample Size options in the U Control Charts dialog include:

- **Use Size Variable–** Specifies that a size variable is used to give the total size of each sample
- **Use Common Size** Specifies that no size variable is used as each sample is the entered size.

Graph Options in the U Control Charts dialog include:

- Sigma 1 and 2 Control Limits Adds control limit lines at one and/ or two sigmas from mean line. These extra lines allow you to check for other types of out-of-control behavior (the so-called "zone" rules).
- **Zone A and B Analysis** number of points beyond Zone A (2 sigma) and Zone B (1 sigma)
- Runs Up/Down and Runs Above/Below Analyses sequential up or down moves of length 8 or greater
- Limit Graph Range From Index Limits the points on the graph to a range of data points for a variable (i.e., range from starting index to ending index)

Control Limit Calculations Based On options in the U Control Charts dialog determine the data on which control limit calculations will be based, including:

- All Observations use all available data in control limit calculations
- Observations in Range use data between Start Index and Stop Index in control limit calculations
- **Previous Data** creates control limits from previously observed data. Simply enter the **Subsample Size** and **Average U** that was calculated from the previous data.



Example U Chart

- Missing Data and Link to Data
- Missing Data Missing data are not allowed.
- Link to Data Graphs are not linked to data.

Nonparametric Tests Menu

"Nonparametric" tests are statistical procedures applied to samples of data to test hypotheses about underlying probability distributions. "Parametric" hypothesis tests are more familiar and widely used; however, the nonparametric alternative offers advantages that make it the more suitable choice in many situations.

Parametric hypothesis tests make assumptions about the type of the underlying distribution (typically, that it is normal), and estimate the parameters of that type of distribution (typically, the mean and standard deviation). In many applications the normality assumption would be incorrect. For example, the numbers of calls per hour to a customer service center and the waiting time at a checkout in a supermarket are not normally distributed. Nonparametric tests do not require any assumptions about the type of the underlying distribution. Some of them make certain general assumptions about the shape of the distribution: in this pack the Wilcoxon Signed-Rank Test assumes that the distribution is symmetric. The other two tests in the pack (the Sign Test and the Mann-Whitney Test) do not assume anything about the distribution shape.

With small sample sizes the nonparametric alternative is often more appropriate. If the sample is large, a normality test can be applied. If the assumption that the distribution is normal turns out to be justified, a parametric test can be used. However, for small sample sizes normality tests have little power to differentiate between the normal and other distributions. Nonparametric tests can provide a way out of the dilemma.

Ordinal Data	For certain types of data, parametric tests cannot be used, while some nonparametric ones can. One such case is ordinal data where observations are described in terms of numbers that express places in a ranking; however, the difference between two such numbers is not meaningful. For example, levels of educational attainment can be coded as 0 (less than high school), 1 (some high school), 2 (high school degree), 3 (some college), 4 (college degree), and 5 (post college). When this scale is used, there is no implication that the difference in educational attainment between having a "high school degree" and having "less than high school" is equivalent to the difference between being in the categories "post college" and "some college ", even though in both cases the difference between the ranks is equal 2. This pack includes tests that can be applied to such data - the Sign Test (One- Sample analysis type) and Mann-Whitney Test.
Summary of Uses	In summary, nonparametric tests are applicable in the following situations, in which parametric ones are not:
	 when there is little information about the underlying probability distribution, when the sample size is too small to reliably test the normality assumption, when the data is ordinal.

Sign Test Command

Performs sign test on variables

The Sign Test command performs hypothesis tests for the median of a single variable (**One-Sample Analysis**) or for the median of differences for a pair of variables (**Paired-Sample Analysis**). The test does not make any assumptions about the shape of the distribution (and in particular does not assume that it is normal). The One-Sample Analysis can be used with **ordinal data** as described in the Overview.

Sign Test Dialog Box This analysis is set up using the **Sign Test** dialog box:

2	StatTools - Nonparametric Tests - Sign Test								
ļ	Analysis Type Paired-Sample Analysis								
Γ	Variables (Select Exactly Two)								
	<u>D</u> ata	Set	Grades	✓ Eormat					
			Name	Address					
	₽	2002		B19:B42					
	☑	2003		C19:C42					
Γ	- Me	edian of Di	fferences						
	<u>N</u> u	ll Hypothe	sis Value	0					
	Alt	ernative <u>H</u>	lypothesis	Less Than Null Value (One-Tailed Test)					
ļ	0]	OK Cancel					

The number of variables selected depends on the Analysis Type used. A One-Sample Analysis requires one or more variables, while a Paired-Sample Analysis requires two variables. For a One-Sample Analysis the selected variables can be stacked or unstacked data; for a Paired-Sample Analysis they have to be unstacked. Variables can be from different data sets. The options in the **Sign Test** dialog include:

- Analysis Type. Selects the type of analysis performed. Options include:
 - **One-Sample Analysis**. Performs hypothesis tests for a single numerical variable.
 - **Paired-Sample Analysis**. This analysis is appropriate when two variables are naturally paired. It is equivalent to a one-sample analysis on the differences between pairs.
- Median (or Median of Differences).
 - **Null Hypothesis Value** or the value of the population parameter under the null hypothesis.
 - Alternative Hypothesis or the alternative to the Null Hypothesis Value that will be evaluated during the analysis. The Alternative Hypothesis can be either "onetailed" (that is, greater or less than the null hypothesis) or "two-tailed" (that is, not equal to the null hypothesis).



Sign Test Report

Missing Data and Link to Data

- **Missing Data -** Missing data are allowed. All rows with missing data on the selected variables are ignored.
- Link to Data All of the reports are calculated with formulas that are linked to the data. If the values of the selected variable change, the outputs change automatically.

Wilcoxon Signed-Rank Test Command

Performs Wilcoxon Signed-Rank tests on variables

The Wilcoxon Signed-Rank Test command performs hypothesis tests for the median of a single variable (**One Sample Analysis**) or for the median of differences for a pair of variables (**Paired-Sample Analysis**). The test assumes that the probability distribution is symmetric (but it does not assume that it is normal).

This analysis is set up using the **Wilcoxon Signed-Rank Test** dialog box:

StatTools - N	onparametric	Tests - Wilcoxo	on Signed-Ran	ık Test 🔀
<u>A</u> nalysis Type	One-Sample Ar	nalysis	•	
Variables (Sele	ect One or More)			
<u>D</u> ata Set	Wilcoxon Signed	-Rank Example	-	<u>F</u> ormat
	Name		Address	
I Temperat	ure	B20:B43		
Median				
Null Hypothes	is Value	0		
Alternative <u>Hy</u>	ypothesis	Not Equal to Null Va	alue (Two-Tailed '	Test) 💌
Correct for 1	Ties in Computation	on of Normal Approx	imation (recomme	ended)
0 🖬 🋂			ОК	Cancel

The number of variables selected depends on the Analysis Type used. A One-Sample Analysis requires one or more variables, while a Paired-Sample Analysis requires two variables. For a one-sample analysis the selected variables can be stacked or unstacked data; for a paired-sample analysis they have to be unstacked. Variables can be from different data sets.

Wilcoxon Signed-Rank Test Dialog Box The options in the Wilcoxon Signed-Rank Test dialog include:

- **Analysis Type.** Selects the type of analysis performed. Options include:
 - **One-Sample Analysis**. Performs hypothesis tests for a single numerical variable.
 - **Paired-Sample Analysis**. This type of analysis is appropriate when two variables are naturally paired. It is equivalent to a one-sample analysis on the differences between pairs.
- Median (or Median of Differences).
 - **Null Hypothesis Value** or the value of the population parameter under the null hypothesis.
 - Alternative Hypothesis or the alternative to the Null Hypothesis Value that will be evaluated during the analysis. The Alternative Hypothesis can be either "onetailed" (that is, greater or less than the null hypothesis) or "two-tailed" (that is, not equal to the null hypothesis).
- **Tie Correction**. A recommended selection that corrects for tied ranks in the test only when the normal approximation is used. The correction involves counting the numbers of elements in groups of tied ranks and reducing the variance accordingly. The correction for ties will always increase the value of the z statistic, if tied ranks are present. (Note: The tie correction will produce no change in the variance when there are no ties.)

Wilcoxon Signed-Rank Test Report

	Ŧ	Book27	- Micros	oft Excel				-		x
Home Insert	Page Lay	out Formulas	Data	Review	View	Add-Ins	StatTool	s 🕜 –	•	×
Data Set Data	ary Statistics ។ ary Graphs *	Normality Tests	s + d Forecas	≤; (ting + ▲ !	Quality Con Nonparame	itrol * teric Tests	 ✓ Utiliti (i) Help 	es * *		
Manager Utilities * 🛛 🗥 Statist	Vanager Utilities * 🔺 Statistical Inference * 📈 Regression and Classification *									
Data		Ana	lyses				Tools			_
A1 - 🤇	fx	StatTools								3
A		В		С	D		E	F		
StatTools (Core Analysis Pack) 2 Analysis: Wilcowo Signed-Rank Test 3 Performed By: Test 4 Date: Monday, February 16, 2009 5 Updating: Live										
6 7 8 Signed-Rank Test (One-Sample)	* w	Temperature ilcoxon Signed-Rank Exa	mple							
9 Sample Statistics 10 Sample Size 11 Sample Mean 12 Sample Median		24 -6.34 13.14 5.75								
Hypotheses Hypothesized Median (HM) Alternative Hypothesis		0 ◇0								
17 Sample Size Adjustment 18 Number Of Values = HM 19 Num. of Values < or > HM (Adjustment)	usted Size)	0 24								.=
20 Ranking Information 21 Number of Tied Values 22 Sum of Negative Ranks		2 226.5								
23 Sum of Positive Ranks (Test Sta 24 p-Value Computation 25 Normal Approximation (NA) U:	tistic) sed	73.5 Yes								
26 Ties Present, but Not Correcte 27 Mean for NA 28 Std. Dev. for NA with Tie Corre	d For	No 150 34.95711659								
29 z-Statistic for NA with Tie Corre 30 p-Value	ection	-2.1741 0.0297								
S1 Significance Levels Null Hypoth. at 10% Significance Null Hypoth. at 5% Significance Null Hypoth. at 1% Significance	e	Reject Reject Don't Reject								
35 WSR Test 22 Ready				[▲]		100)% (=)	0) (+	

Note: In this report, the p-value is computed using normal approximation when the sample size is greater than 15.

Missing Data and Link to Data

- **Missing Data -** Missing data are allowed. All rows with missing data on the selected variables are ignored.
- Link to Data All of the reports are calculated with formulas that are linked to the data. If the values of the selected variable change, the outputs change automatically.

Mann-Whitney Test Command

Performs Mann-Whitney test on variables

The Mann-Whitney Test command performs a hypothesis test on two samples. In one version of the test (the Median Version) the hypothesis states that the medians of the two populations are identical. In this version the probability distributions are assumed to have the same shape. In the other version (the General Version) this assumption is not made, and the hypothesis denies that either probability distribution tends to yield smaller values than the other (more precisely, it states that P[X1>X2] = P[X2>X1], where P[X1>X2] is the probability that an observation from population 1 is greater than an observation from population 2). Note that the Mann-Whitney test can be used to reject the hypothesis that two samples are generated by the same probability distribution. The Mann-Whitney test is often also referred to as the Wilcoxon Rank-Sum test.

This analysis is set up using the **Mann-Whitney Test** dialog box:

Mann-Whitney Test Dialog Box

Stat	Tools - N	lonparametric	Tests - Mann-White	ney Test	X	
<u>A</u> naly	ysis Type	General Version) (has no assumptions ab	out distributior	n shapes) 💌	
Variables (Select Exactly Two)						
Data Set Temperature Data			Eormat			
		Name	Ado	dress		
	Observer	1	B21:B49			
	Observer	2	C21:C49			
Hy	potheses					
Null Hypothesis Neither distribution tends to yield smaller values than other Alternative Hypothesis 1st distribution smaller (One-Tailed Test)						
☑ Correct for Ties in Computation of Normal Approximation (recommended)						
0				ОК	Cancel	

The analysis requires two variables. They can be stacked or unstacked, and can come from different data sets.

The options in the Mann-Whitney Test dialog include:

- **Analysis Type**. Selects the formulation of the null hypothesis and the alternatives. Options include:
 - **General Version**. Performs a hypothesis test to see if one probability distribution tends to yield smaller values than the other.
 - Hypotheses.
 - Null Hypothesis: Denies that either probability distribution tends to yield smaller values that the other. More precisely, it states that P[X1>X2] = P[X2>X1], where P[X1>X2] is the probability that an observation from population 1 is greater than an observation from population 2, and P[X2>X1] has analogous interpretation. For continuous distributions, this is equivalent to saying that both of these probabilities are 0.5 (P[X1>X2] = P[X2>X1] = 0.5).
 - **Alternative Hypothesis**: The Alternative Hypothesis can be either "one-tailed" (that is, one probability is greater or less than the other) or "two-tailed" (that is, the two probabilities are not equal).
 - **Median Version**. Performs a hypothesis test to see if the median of one population is the same or different from the median of the other population. Assumes the two distributions have the same shape.
 - Hypotheses.
 - **Null Hypothesis**: States that the two medians are equal.
 - **Alternative Hypothesis**: The Alternative Hypothesis can be either "one-tailed" (that is, the median of the first population is greater or less than that of the second) or "two-tailed" (that is, the medians are not equal).

Note: The calculations performed in the General and Median versions of the test are the same; the versions differ only with regard to the presence of the assumptions of equal distribution shapes, and with regard to the null hypothesis. The point of having these two versions is to make it clear that the Mann-Whitney test can be applied even if one cannot assume that the distributions have approximately identical shapes, as long as one considers an appropriate null hypothesis. Specifically, if you run the test when the two distributions clearly have different shapes, and the test rejects the null hypothesis, it could be because the medians are different, but it could also be because the variances are different or other reasons.

• **Tie Correction**. A recommended selection that corrects for tied ranks in the test only when the normal approximation is used. The correction involves counting the numbers of elements in groups of tied ranks and reducing the variance accordingly. The correction for ties will always increase the value of the z statistic, if tied ranks are present. (Note: The tie correction will produce no change in the variance when there are no ties.)



In the report above, the *p- value* is computed using normal approximation when the size of one of the two samples is greater than 10 (except when one of the sizes is equal to 11 or 12, while the other is equal to 3 or 4).

Mann-Whitney Test Report

Missing Data and Link to Data

- **Missing Data -** Missing data are allowed. All rows with missing data on the selected variables are ignored.
- Link to Data All of the reports are calculated with formulas that are linked to the data. If the values of the selected variable change, the outputs change automatically.

Utilities Menu

Application Settings Command

Specifies settings for StatTools reports, graphs, utilities, datasets and analyses

The **Application Settings** command allows you to specify general settings for StatTools reports, graphs, utilities, datasets and analyses. These settings apply across analyses and datasets. Other analysis-specific settings are defined in the dialog box for each analysis.

StatTools - Application Settings						
-	General Settings					
	Show Welcome Screen	False		•		
Ξ	Reports					
	Placement	New Workbo	ook			
	- Reuse Same New Workbook	False				
	Updating Preference	Live - Linked	l to Input Data			
	Display Comments					
	- Notes and Warnings	True				
	- Educational Comments False					
Ξ	Utilities					
	New Variable Preference	Insert in Source Data Set				
	Updating Preference	Static				
Ξ	Data Set Defaults					
	Apply Cell Formatting	False				
	Variable Layout	Columns				
	Names In First Row					
	- Primary Range True					
	- Secondary Range True					
Ξ	Analyses					
	Warning Messages When					
	- Ignoring Missing Data	True				
	- Ignoring Non-Numeric Data True					
	Dialog Memory	Remember Last Used Values				
	Percentile Calculations	Automatic (Based on Input Data)				
6			ОК	Cancel		

Reports Settings The **Reports Settings** specify options for reports and graphs created by StatTools analyses. It includes the following:

- **Placement** selects the location in Excel for new reports and graphs, including:
 - Active Workbook, where a new worksheet is created for each report.

- **In New Workbook**, where a new StatTools report workbook is created (if necessary) and each report is placed on a sheet in that workbook.
- After Last Used Column in Active Sheet, where each StatTools report is placed on the active sheet to the right of the last used column.
- **Query for Starting Cell,** where, after running an analysis, you have the opportunity to select a cell where the top-left corner of the report or graph will be placed.
- **Reuse Same New Workbook,** where, if a new workbook is created, that same new workbook will be used for all reports
- **Updating Preference** specifies how results will change when variable data is changed. Options for results updating include:
 - Live Values Change With Input Data, where reports update automatically as input data changes.
 - **Static Values are Fixed,** or reports do not change with changing input data. Statistics are fixed based on the input data values when the procedure was run.

Results are made live in StatTools through the use of Excel formulas and custom StatTools functions. For example, the formula:

=StatMean('Confidence Interval.xls'!Pair)

calculates the mean of the variable *Pair* (which uses data from the Excel range name "Pair" located in the workbook Confidence Interval.xls). As data in the Pair range changes, the value returned by the StatMean function will update.

Reports and graphs from all StatTools procedures can be updated live, with the following exceptions:

- 1) Regression
- 2) Discriminant Analysis
- 3) Logistic Regression
- 4) Forecast

These procedures require lengthy recalculations that would cause Excel to become less responsive during live updating.
• **Display Comments** – specifies which categories of StatTools messages will be included in reports. Notes, warnings and Educational notes may be displayed as pop-up notes on cells in reports, as shown here:

n n	nome in	seit	Page Layout	t Formulas Data Review View	v Add-Ins StatT	Tools	1					0 -	
11. 13	19 Es	ummary	Statistics *	🚮 Normality Tests • 🖄 Quality	Control - + U	tilities	•						
	و 🛓 🖬	ummary	Graphs -	Time Series and Forecasting - 🔺 Nonpar	ameteric Tests = 🕡 H	elp -							
Set D	ities -	tatistica	Inference -	Regression and Classification *									
Data				Analyses	Т	ipols							
A20	0 -	. (*	fu .										
1	В	С	D	E	F		G	Н		1		J	
St		Evan	ople · Ma	ann Whitney Test (General)	fersion Linsta	cke	d)						
04	attools	Aan	ipie . m	and windley rest (General v	reision, onsta	CRE	u)						
The	StalTools Mann	-Whitne	ry Test analysis	s is used to check if a set of sample data provides ev	ridence for accepting or rej	jecting v	various hypot	heses.					
distr	ibutions are ass	umedic	have the same	e shape. In the other version (the General Version)	this assumption is not mad	fe. and	the hypothes	Soachty					
deni	ies that either p	robabilit	y distribution to	ends to yield smaller values than the other (more pre	cisely, the hypothesis state	es that is	P[X1aX2]=						
sdx5	exal where PD	(19/2])	s the probabil	ty that an observation from population 1 is greater th	an an observation from pag	pusition.	n2)						
Note	This procedur	e suppo	its both stacke	d and unstacked data. This example uses unstacked	ed data.								
The	data satshown	is hand	Herent observ	ent'measurements of temperatures at a local weath	erstation Runnings Man	er. 15.8%	her Test/Ces	(and					
Vers	ion) analysis or	whethe	in the first pers	on's observations tend to be smaller than those of th	e other's gives the results s	shown.	Usinga one	tailed					
alter	native, we find:	that we d	can reject the n	null hypothesis at 10% significance, but not at 5% or	1% significance.								
This	example was o	reated b	v Palisade Co	moration.									
6													
_													
Obs	erver 1 C	baerve			Observer 1		Observer 2						
Obs	anver 1 C	29.	75	Sample Statistics	Observer 1 Temperature Data	Ten	Observer 2 nperature D	lata					
Obs	anver 1 0 29.4 25.28	29. 37.	75	Sample Statistics Sample Size	Observer 1 Temperature Data 29	Ten	Observer 2 nperature 0 24	lata					
Obs	29.4 25.28 34.1	29. 37. 32.	75 88 59	Sample Statistics Sample Site Sample Man	Observer 1 Temperature Data 29 29 292	(Tem	Observer 2 nperature D 24 32 017	Jata					
Obs	erver 1 0 29.4 25.28 34.1 36.81	29. 37. 32. 33.	75 88 59	Sample Statistics Sample Site Sample Man Sample Man Dev.	Observer 1 Temperature Data 29 29 292 4 671	(Tem	Observer 2 nperature D 24 32 017 5.076	lata					
Obs	erver 1 0 29.4 25.28 34.1 36.81 26.54	29. 37. 32. 33. 29.	75 88 59 54 08	Sample Statistics Sample Site Sample Mean Sample Median	Observer 1 Temperature Data 29 28 292 4 671 23.400	(Tem	Observer 2 nperature 0 24 32 017 5 076 31.195	Data					
Obs	29.4 25.28 34.1 36.81 26.54 24.55	29. 37. 32. 33. 29. 29.	75 75 88 59 54 08 03	Sample Statistics Sample Site Sample Mon Sample McDev. Sample Median	Observer 1 Temperature Data 29 29 292 4 671 29.400	(Tem	Observer 2 nperature 0 24 32 017 5 076 31.195	Data					
Obs	29.4 25.28 34.1 36.81 26.54 24.55 25.33	29. 37. 32. 33. 29. 29. 29. 29.	75 75 88 59 54 08 03 64	Sample Statistics Sample Site Sample Mean Sample Median Sample Median Mann-Whitney Test (General Version)	Observer 1 Temperature Data 29 29 292 4 671 29.400	(Tem	Observer 2 nperature 0 24 32 017 5.076 31.195	ata					
Obs	29.4 25.28 34.1 36.81 26.54 24.55 26.33 32.8	29. 37. 32. 33. 29. 29. 29. 29. 29. 29.	75 88 59 54 08 03 64 55	Sample Statistics Sample Size Sample Moan Sample Modan Sample Median Mann-Whitney Test (General Version) Typotheas	Observer 1 Temperature Data 29 29 292 4 671 29.400	Tem	Observer 2 nperature 0 24 32 017 5.076 31.195	Jata					
Obs	29.4 25.28 34.1 36.81 26.54 24.55 25.33 32.8 21.24	29 37 32 33 29 29 29 29 29 23 30	75 88 59 64 08 03 64 55 47	Sample Statistics Sample Site Sample Man Sample Mc Dev. Sample Median Mann-Whitney Test (General Version) Hypothesis	Observer 1 Temperature Data 29 29 4 671 29.400 Neither Dist. Smalle	Tem	Observer 2 neerature D 24 32 017 5 076 31.195 StatTools M	iata Soles					
Obs	29.4 25.28 34.1 36.81 24.55 25.33 32.8 21.24 31.84	29 37 32 33 29 29 29 29 29 30 30 38	75 75 88 69 54 08 03 64 55 64 55 47 34	Sample Statistics Sample Site Sample Site Sample Man Sample Median Mann-Whitney Test (General Version) Typothesa Null Hypothesis Atternative Hypothesis	Observer 1 Temperature Data 29 292 4 671 29.400 Neither Dist. Smaller Observer 1 Smaller	Tem	Observer 2 nperature D 24 32 017 5 076 31.195 StatTools M The General W	lata Materi Fersion hypo	othesis de	enies the	at		
Obs	29.4 25.28 34.1 26.54 24.55 26.33 32.8 21.24 31.84 19.75	29. 37. 32. 33. 29. 29. 29. 29. 29. 29. 29. 29. 30. 38. 36.	75 75 88 54 06 03 64 55 64 55 7 7 34 69	Sample Statistics Sample Site Sample Man Sample Median Mann-Whitney Test (General Version) Hypothesis Hull Hypothesis Atternative Hypothesis Anternative Hypothesis	Observer 1 Temperature Data 29 28 292 4 671 29.400 Neither Dist. Smaller Observer 1 Smaller	Tem	Observer 2 nperature D 24 32.017 5.076 31.195 StatTools M The General V ether distributions the other the other	lata Viter Son Hyper Terson Hyper Titon tends a v (2012) S21	othesis de to yield se	enies tha	at alues here		
Obs	anner 1 C 29.4 25.28 34.1 36.81 26.54 24.55 26.33 3.2.8 21.24 31.84 19.75 31.35	29 37 32 33 29 29 29 29 29 30 38 36 39	75 88 69 54 00 03 64 47 34 69 33	Sample Statistics Sample Site Sample Men Sample Median Mann-Whitney Test (General Version) Hypothesis Natternative Hypothesis Atternative Hypothesis Ranking information Rumber of Values in Renking	Observer 1 Temperature Data 29 29 292 4 671 29,400 Neither Dist. Smaller 0bserver 1 Smaller 53	Tem	Observer 2 nperature D 24 32 017 5 076 31.195 StatToosh Hi The General V other distribution the other V(L1xX2) and	lata lata letson hypo tion tends t er (P(1)x2) obbility tha	othesis dr to yield se] = P[X2: t observ.	enies the maler va >X1], with fro	at alues here om		
Obs	enver 1 0 29.4 25.28 34.1 36.81 26.54 24.55 25.33 32.8 21.24 31.84 19.76 31.35 27.74	29. 37. 32. 29. 29. 29. 29. 29. 30. 38. 36. 39. 29.	75 75 54 59 54 00 00 00 00 64 47 34 69 33 24	Sample Statistics Sample Site Sample Mean Sample Median Mann-Whitney Test (General Version) Hypothesis Atternative Typothesis Anternative Typothesis Anternative Typothesis Mumber of Values in Branking Number of Values in Branking Number of Values in Branking	Observer 1 Temperature Data 29 29 292 4 671 23.400 Neither Dist. Smaller Observer 1 Smaller 53 2	Tem S T U U P D	Observer 2 nperature D 24 32,017 5,076 31,195 StatTools Re The General W other distribu- than the other (DLTX2): proposition 13	iata i iata i iata i iata i iata i iata i iata i iata i iata i i iata i iata i iata i i iata i i iata i i iata i i i iata i i iata i i i i i i i i i i i i i i i i i i i	othesis de to yield se] = P[X2: t observ. pop. 2).	enies tha mailer va >X1], wh abon fro	at alues here om	-	
Obs	erver 1 223.4 25.28 33.4 36.81 26.54 24.55 25.33 3.2.8 21.24 31.84 19.75 31.35 27.74 35.85	29 37 32 29 29 29 29 29 29 30 38 36 39 39 29 34	75 88 59 54 00 03 04 55 54 69 33 47 69 33 24 76	Sample Statistics Sample Site Sample Men Sample Median Mann-Whitney Test (General Version) Myootheas Null Hypothesis Attentave Ingothesis Ranking Information Rumber of Valaes in Ranking Number of Valaes in Ranking Number of Valaes in Ranking Number of Valaes	Observer 1 Temperature Data 29 29 292 4 671 29 400 Neither Dist. Smaller 53 2 800.5	Tem	Observer 2 nperature 0 24 32 017 5 076 31.195 Stations fit Other General V either distribution for the P(N1>x2); provide 0 1 > x20; provide 0 1 >	lata Velta: itersion hypothesion hypothesion obability thas > obs. from	othesis dr to yield st] = P[X2: t observ. pop. 2).	enies tha maller va >X 1), wh abon fro	at alues here om		
Obs	229.4 25.28 34.1 36.81 26.54 24.55 25.33 32.8 21.24 31.35 21.24 31.35 27.74 35.85 27.74	29. 37. 32. 29. 29. 29. 29. 29. 29. 29. 29. 29. 2	75 75 88 69 54 60 03 64 47 34 69 33 24 76 27	Sample Statistics Sample Site Sample Mon Sample Median Mann-Whitney Tust (General Version) Hypothese Null Hypothesis Alternative Hypothesis Ranking Information Number of Values in Ranking Number of Values in Ranking Ranks Sum for Vanable 1 (Test Statistic) Rank Sum for Vanable 1 (Test Statistic)	Observer 1 Temperature Data 29 29 292 4 671 29,400 Neither Dist. Smaller 53 2 690.5 740.5	Tem	Observer 2 nperature 0 24 32.017 5.076 31.195 StatTools M The General V Han General V Han General V P(1>x2) ero population 1.3	lata Peter Person hypo prison hypo hypo h	othesis de to yield sr] = P[X2: st observ. pop. 2).	enies this nalier va ×X1], vition fro	at alues here om		

Utilities SettingsThe Utilities Settings specify options for new variables
created by the StatTools Data Utilities. These are found on
the Data Utilities menu. They create new variables by
transforming, combining or otherwise processing existing
variables.

- **New Variable Preference** Specifies the desired location for new variables that are created by data utilities. The options include:
 - **Insert in Source Data Set**, where each new variable created is inserted at the right-hand side (or bottom) of the data set containing the original variable.
 - Create New Data Set, where new variables will be placed in a new data set

There are, however, cases when the selected New Variable Preference may not be followed. These include **Stack and Unstack** (where new variables will always be placed in a new data set) and when the **original data is from a multi-range data set** (where new variables can only be inserted in the source data set)

- **Updating Preference** specifies how the values for new variables created by a utility will change when original variable data is changed. Options for updating include:
 - Live Values Change With Input Data, where new variable values update automatically as input data changes.
 - **Static Values are Fixed**, where new variable values do not change with changing input data. New variable values are fixed based on the input data values when the procedure was run.

There are, however, cases when the selected Updating Preference will not be followed and new variables will always use the **Static – Values are Fixed** preference. These include **Stack and Unstack** and **Random Samples**. Live updating is not applicable to these utilities.

Data Sets Settings	The Data Sets Settings specify options for new data sets created using the Data Set Manager command. These settings are just defaults that appear when you create a new data set. They can be changed if desired in the Data Set Manager dialog box.
•	Apply Cell Formatting – Specifies if the data set will be formatted by StatTools.
•	Layout – Selects the default variable layout (row-wise or column-wise)
•	Names in First Column/Row (Primary Range) – Specifies if names are entered in the first column or row for the first range defined.
•	Names in First Column/Row (Secondary Range) – Specifies if names are entered in the first column or row of the second and all subsequent ranges for a multi-range data set.
Analyses Settings	The Analyses Settings specify the default entries displayed in dialog boxes used to set up analyses. These are just the entries that appear when you first display a dialog box for an analysis. They can be changed if desired in each dialog box.
•	Warning Messages – Selects whether warning messages will be displayed when running an analysis if StatTools detects missing data in a variable or if StatTools detects non-numeric data.
•	Dialog Memory – Specifies the default entries in dialog boxes used to set up analyses. Options include:
	- Remember Last Used Values (by Workbook) , where a displayed dialog box shows the entries that were made in that dialog the last time it was displayed for the active workbook. If a dialog is displayed for the first time, it shows the saved system default settings for the analysis.
	- Always Use System Default Values, where a

Always Use System Default Values, where a displayed dialog box shows the saved system default settings for the analysis.

- **Percentile Calculations** Selects the method to be used for calculating percentiles. Depending on the nature of your data, different methods can give better answers. Available methods (and the type of data they are suited for) include:
 - 1) Automatic (Based on Input Data)
 - 2) Interpolated with Asymmetric Endpoints (Continuous)
 - 3) Interpolated with Symmetric Endpoints (Continuous)
 - 4) Excel Percentile Function (Continuous)
 - 5) Closest Observation (Discrete)
 - 6) Empirical Dist. Function (Discrete)
 - 7) Empirical Dist. Function with Averaging (Discrete)

Delete Data Sets Command

Deletes StatTools data sets in the active workbook

The **Delete Data Sets** command deletes all defined data sets from the active workbook. The actual data in Excel is not deleted; just the definition of the data sets.

Clear Dialog Memory Command

Clears all memory of entries in analysis dialog boxes

The **Clear Dialog Memory** command clears all "memory" of entries for analysis dialog boxes. Subsequent displayed dialog boxes will initially show the saved system default settings for each analysis.

Unload StatTools Add-in Command

Unloads the StatTools Add-in

The Unload StatTools Add-in Command unloads StatTools, closing all StatTools windows.

Help Menu

StatTools Help

Opens on-line help file for StatTools

The Help menu StatTools Help command opens the main help file for StatTools. All of StatTools's features and commands are described in this file.

Online Manual

Opens online manual for StatTools

The Help menu Online Manual command opens this manual in PDF format. You must have Adobe Acrobat reader installed to view the online manual.

License Activation Command

Displays licensing information for StatTools and allows the licensing of trial versions

The Help menu License Activation command displays the License Activation dialog box, listing the version and licensing information for your copy of StatTools. Using this dialog box you can also convert a trial version of StatTools into an licensed copy.

For more information on licensing your copy of StatTools, see **Chapter 1: Getting Started** in this manual.

About Command

Displays version and copyright information about StatTools

The Help menu About command displays the About dialog box, listing the version and copyright information for your copy of StatTools.

Reference: StatTools Functions

Introduction

Custom worksheet functions are used by StatTools to return calculated statistics to Excel formulas. These functions allow:

- 1) Statistics calculations to be embedded in worksheet formulas, just as are standard Excel functions.
- 2) Statistics to be "live", that is, results change when original data changes.

If you look at the formulas in the cells in a StatTools report, you will see StatTools worksheet functions. All StatTools functions begin with the prefix "Stat", such as **StatMean()** or **StatStdDev()**. All StatTools functions are displayed in the Excel Insert Function dialog for ease of entry.

StatTools Functions vs. Excel Functions

In some cases StatTools replaces Excel's built-in statistics with its own robust and fast calculations. The accuracy of Excel's built-in statistics calculations has often been questioned, and StatTools uses none of them! Even Excel's worksheet statistics functions – such as STDEV() – are replaced by new, robust StatTools versions – such as StatSTDEV(). StatTools statistics calculations meet the highest tests for accuracy, with performance optimized through the use of C++ .DLLs, not macro calculations.

StatTools functions, as opposed to the built-in Excel functions, support the use of stacked data. The StatTools **StatDestack** function automatically unstacks data from a stacked data set (for a category you specify). It then passes this data to a StatTools statistics function for analysis.

StatTools functions also support the analysis of data that resides on different worksheets. Multi-sheet data sets allow more than 65535 points per variable. They are entered using the **Multiple** button in the Data Set Manager dialog.

Distribution Functions

StatTools includes a set of distribution functions (such as **StatBinomial**) which replace Excel's built-in distribution functions (such as **BinomDist**). Unlike Excel's distribution functions, the StatTools distribution functions can return a number of different values from a probability distribution. The value returned is set by the **statistic** argument (the second to last argument in the function). This argument can be a value 1 to 12 or a string that indicates the statistic you want to get for the entered distribution:

Entered Value or String	Returned Statistic
1 or "mean"	mean
2 or "stddev"	standard deviation
3 or "variance"	variance
4 or "skewness"	skewness
5 or "kurtosis"	kurtosis
6 or "mode"	mode
7 or "discrete mean"	discrete mean (or the value closest to the true mean that actually could occur)
8 or "x to y"	x to y (the distribution y-value for an entered x-value)
9 or "x to p"	x to p (the distribution p-value for an entered x-value)
10 or "p to x"	p to x (the distribution x-value for an entered p-value)
11 or "x to q"	x to q (the distribution q-value for an entered x-value)
12 or "q to x"	q to x (the distribution x-value for an entered q-value)

For example, the StatTools distribution function:

StatNormal(10,1,"*x* to *p*", 9.5)

Returns the p value associated with the x value of 9.5 in a normal distribution with a mean of 10 and a standard deviation of 1.

List of Possible Returned Statistics

"Live" Reports

StatTools uses custom functions to make the results as "live" as possible. That is, whenever it is practical, reports have formulas that link to the original data. For example, suppose you have a variable *Weight* and you want summary measures on *Weight*, such as its mean and standard deviation. The Summary Statistics procedure names the range of weights as Weight, and then it enters formulas in the output cells: **=StatMean(Weight)** and **=StatStdDev(Weight)**. StatMean and StatStdDev are built in StatTools functions for calculating mean and standard deviation. These replace the standard built-in Excel functions for the same statistics. Because of these functions, when your data change, the results change automatically, so that you don't have to rerun the procedure.

	(m.) .			BOOKAS	MICroso	IT EXCEL				-	
Home Home	Inseit Page	Layout Formul	as Data	Review	View	Add-Ins	StatTools			w	
ata Set Data	Summary Statist	ics • 🔺 Normality s = 🖄 Time Seri	Tests + es and Foreca	sting = 1	Quality Co Nonparam	itrol = eteric Tests =	📌 Utilities *				
Data		M. Sector	Anabasa				Tools				
B9	• (*	fe =StatMean[!	StatDestack	('One Variat	ole Sumn	ary.xls'IST	Salary_2,'One V	ariable Sum	mary.xls'IST_	Gender,"F"	.))
A	В	C	D	E		F	G	н	1	L	
Analysi Performed B Data Updatin	s: One Variable Sun y: Test es Monday, Februar g: Live	nmary ny 16, 2009									
	Salary (F)	Salary (M)									
One Variable Summary	Stacked	Stacked									
Mean	\$29441.67	\$30089.36									
Variance	14737192.98	12661821.09									
Std. Dev.	\$3838.91	\$3558.35									
Skewness	-0.3034	0.0493									
Kurtosis	3.1872	2.4343									
Median	\$29700.00	\$29900.00									
Mean Abs. Dev.	\$3050.35	\$2893.16									
Minimum	\$17100.00	\$22400.00									
Maximum	\$36900.00	\$38200.00									
Baore	\$19800.00	\$15800.00									
Count	96	94									
Sum	\$2826400.00	\$2828400.00									
1st Chartile	\$26700.00	\$27500.00									
Bed Quartile	\$31900.00	\$32500.00									
Internuartile Banne	\$5200.00	\$5000.00									
1.00%	\$17100.00	\$22400.00									
5 5.0%	\$22300.00	\$23400.00									
Te nom	\$22800.00	\$23900.00									
10.00%	\$24500.00	\$25200.00									
20.00%	\$26500.00	\$26900.00									
80.00%	\$32400.00	\$33800.00									
50.00%	\$34600.00	\$35000.00									
SE 00%	\$36200.00	\$35900.00									
97.50%	\$36400.00	\$36800.00									
99.00%	\$36900.00	\$38200.00									
-											
and the second second	0	20									-
1 P One Var N						10 A					

There are times when it is not practical to link results to data. The prime example is regression. StatTools does not provide the formulas that are used to create regression output; it provides only the numerical results. In such cases, if your data change, you will have to rerun the procedures.

You can use the **Reports Settings command Static option** to not have your reports linked to your data. This is useful if Excel recalculation time becomes an issue as data changes.

Reference: Listing of Statistics Functions

Table of Available Functions

This table lists the custom functions that are added to Excel by StatTools. When used, all functions are preceded by the entry **Stat**.

Function	Returns
AUTOCORRELATION(data	Calculates the autocorrelation for the values
,numLags)	in a data set
<u>AVEDEV</u> (Data1,Data2,DataN)	Calculates the average absolute deviation of
	the data from their mean. The arguments
	can be numbers, arrays, or ranges.
<u>BINOMIAL</u> (N,P,statistic,value)	Calculates the <i>statistic</i> for the entered
	binomial distribution
CATEGORYINDICIES(range,	Gets the cell indices for a specified category
category_name)	
CATEGORYNAMES(range)	Gets the names of the categories in a range
CATEGORYOCCURRENCECO	Calculates the number of cells in a range in a
<u>UNT</u> (range,category_name)	specified category
CHISQ(deg_freedom,statistic,value	Calculates the <i>statistic</i> for the entered one-
)	tailed chi-squared distribution
<u>CORRELATIONCOEFF(data1,</u>	Calculates the correlation coefficient
data2)	between 2 data sets
<u>COUNT(</u> Data1,Data2,DataN)	Calculates the number of elements in its
	arguments, which can be numbers, arrays, or
	ranges
COUNTCATEGORIES(range)	Counts the number of the categories in a
	range
<u>COUNTCELLSBYTYPE(</u> range,	Calculates the number of cells of a specified
type)	type in a range
<u>COUNTRANGE</u> (Range,MinValu	Calculates the number of values in the <i>Range</i>
e,MaxValue	that fall between <i>MinValue</i> and <i>MaxValue</i>
,IncludeMin,IncludeMax)	
COVARIANCE(data1,data2)	Calculates the sample covariance between 2
	data sets
<u>COVARIANCEP</u> (<i>data1</i> , <i>data2</i>)	Calculates the population covariance
	between 2 data sets. Any missing numbers
	cause a blank to be returned.

Function	Returns
DESTACK(data_range,	Extracts data in a specified category from
categories_range_1,category_1,	stacked data
categories_range_2,category_2)	
DURBINWATSON(data)	Calculates the Durbin-Watson statistic for
	the values in a data set
<u>F</u> (deg_freedom1,deg_freedom2,statis	Calculates the <i>statistic</i> for the entered F
tic, value)	distribution for 2 data sets
GETCELLVALUES(range)	Gets the values for all the cells of a specified
KURTOSIS/Data1 Data2	Calculates the sample kurtosis of its
DataN)	arguments, which can be numbers, arrays, or
Dutury	ranges
KURTOSISP(Data1,Data2,	Calculates the population kurtosis of its
DataN)	arguments, which can be numbers, arrays, or
	ranges
$\underline{LN}(x)$	Calculates the natural logarithm of a
	positive, real number
<u>MAX</u> (Data1,Data2,DataN)	Calculates the maximum of its arguments,
	which can be numbers, arrays, or ranges
<u>MEAN</u> (Data1,Data2,DataN)	Calculates the arithmetic mean (average) of
	its arguments, which can be numbers,
MEANABS (Data1 Data2 DataN	Calculates the arithmetic mean (average) of
)	the absolute values of its arguments, which
/	can be numbers, arrays, or range
MEDIAN(data,discrete flag)	Calculates the median of a data set
MIN(Data1,Data2,DataN)	Calculates the minimum of its arguments,
	which can be numbers, arrays, or ranges
<u>MODE</u> (Data1,Data2,DataN)	Calculates the mode of its arguments, which
NORMAL (magn and day statistic	Calculates the statistic for the entered normal
<u>NORMAL</u> (<i>meun,stu_ueo, stutistic,</i>	(Gaussian) distribution
PAIRCOUNT (Data1 Data2)	Counts the number of pairs of cells for
<u>I AIRCOUNT</u> (Dutu1, Dutu2)	which each of the cells in the pair is numeric
PAIRMEAN(Data1.Data2)	Calculates the mean of the differences
(==)==)	between pairs of cells
PAIRMEDIAN(Data1,Data2)	Calculates the median of the differences
	between pairs of cells
<u>PAIRSTDDEV</u> (Data1,Data2)	Calculates the sample standard deviation of
DEDCENITH E (1-1	the differences between pairs of cells
<u>rekcentile</u> (aata, p, discrete Elac)	Calculates the p-th percentile of a data set
BODUCT/Data1 Data2 DataN	Calculator the product of its arguments
$\frac{\mathbf{r}_{NODUC1}(Dutu1,Data2,\ldotsDataN)}{NODUC1}$	which can be numbers arrays or ranges
)	which can be numbers, arrays, or ranges

Function	Returns
<u>QUARTILE</u> (data,q,discrete flag)	Calculates the specified quartile of a data set
RAND()	Returns a random number in the range $0 - 1$.
<u>RANGE (</u> Data1,Data2,DataN)	Calculates the range (maximum - minimum) of its arguments, which can be numbers, arrays, or ranges
RUNSTEST (data, cutoff)	Calculates the runs statistic for the values in a data set
<u>SKEWNESS(</u> Data1,Data2,Data N)	Calculates the sample skewness of its arguments, which can be numbers, arrays, or ranges
<u>SKEWNESSP</u> (Data1,Data2,Data N)	Calculates the population skewness of its arguments, which can be numbers, arrays, or ranges
<u>STANDARDIZE(</u> x,mean,std_dev)	Calculates a normalized value from a distribution with the specified mean and standard deviation
STDDEV	Calculates the sample standard deviation of
(Data1,Data2,DataN)	its arguments, which can be numbers, arrays, or ranges
<u>STDDEVP</u> (Data1,Data2,DataN)	Calculates the population standard deviation of its arguments, which can be numbers, arrays, or ranges
<u>SUM</u> (Data1,Data2,DataN)	Calculates the sum of its arguments, which can be numbers, arrays, or ranges
SUMDEVSQ(Data1,Data2,Data N)	Calculates the sum of the square of the deviation from the mean of its arguments, which can be numbers, arrays, or ranges
<u>SUMSQ</u> (Data1,Data2,DataN)	Calculates the sum of the square of its arguments, which can be numbers, arrays, or ranges
<u>STUDENT</u> (deg_freedom,statistic, value)	Calculates the <i>statistic</i> for the entered Student's t-distribution
<u>VARIANCE</u> (Data1,Data2,Data N)	Calculates the sample variance of its arguments, which can be numbers, arrays, or ranges
VARIANCEP(Data1,Data2,Data N)	Calculates the population variance of its arguments, which can be numbers, arrays, or ranges

Detailed Function Descriptions

Statistics functions are listed here with their required arguments.

AUTOCORRELA TION	Description	Autocorrelation (<i>data,numLags</i>) calculates the autocorrelation for the values in a data set data using the number of lags <i>numLags. data</i> is the array or range of data to calculate the autocorrelation for. <i>numLags</i> is the number of lags to use.
	Examples	StatAutocorrelation (<i>C1:C100,1</i>) returns the autocorrelation for the data in the range <i>C1:C100</i> with 1 <i>lag</i> .
	Guidelines	<i>numLags</i> must greater than or equal to 1.
AVEDEV	Description	AVEDEV (<i>Data1,Data2,DataN</i>) calculates the average absolute deviation of <i>Data1,Data2,DataN</i> from their mean.
	Examples	StatAveDev (<i>1</i> , <i>2</i> , <i>5</i>) calculates the average absolute deviation of the 1,2 and 5 from their mean.
	Guidelines	<i>Data1,Data2,DataN</i> arguments can be numbers, arrays, or ranges.
BINOMIAL	Description	BINOMIAL (<i>N</i> , <i>P</i> , <i>statistic</i> , <i>value</i>) calculates the <i>statistic</i> for the binomial distribution which has the specified <i>N</i> and <i>P</i> values.
	Examples	StatBinomial (2,.3, " <i>x</i> to p ", 1) calculates the p-value for an x-value of 1 in the binomial distribution with N=2 and P = .3
	Guidelines	<i>N</i> is the number of trials or events and must be an integer >0
		<i>P</i> is the probability and must be $\geq =0$ and $\leq =1$.
		<i>statistic</i> is an integer value 1 to 12 or a string indicating the statistic to be returned. For more information see the section Distribution Functions at the start of this chapter.
		<i>value</i> must be ≥ 0 (when an x value is entered) or ≥ 0 and ≤ 1 (when a p value is entered).

CATEGORYINDI CIES	Description	CATEGORYINDICES (<i>range, category_name</i>) gets the indices for the cells containing the specified <i>category_name</i> with the Excel <i>range</i> . This is an array function and the returned indices are index numbers (between 1 and # of cells in <i>range</i>) giving the positions of the cells containing <i>category_name</i> within the <i>range</i> .
	Examples	StatCategoryIndices (C1:C100,"Male") returns the indexes (between 1 and 100) of the cells containing the string Male.
	Guidelines	<i>range</i> is a valid Excel range <i>category_name</i> is a string, value or cell reference specifying the category to find
CATEGORYNA MES	Description	CATEGORYNAMES (<i>range</i>) gets the names of the categories in the specified Excel <i>range</i> . This is an array

	function and the number of returned names is between 1 and # of cells in <i>range</i> .
Examples	StatCategoryNames (C1:C100) returns the names of the categories in the range C1:C100.
Guidelines	range is a valid Excel range

CATEGORYOCC
URRENCECOUN
Т

Description	CATEGORYOCCURRENCECOUNT (<i>range, category_name</i> returns the number of cells containing the specified <i>category_name</i> in the Excel <i>range</i> .	
Examples	StatCategoryOccurrenceCount (C1:C100,"Male") returns the number of cells in the range C1:C100 that contain "Male".	
Guidelines	<pre>range is a valid Excel range category_name is a string, value or cell reference specifying the category to find</pre>	

Description	CHISQ (<i>deg_freedom,statistic,value</i>) calculates the <i>statistic</i> for the one-tailed chi-squared distribution using the specified degrees of freedom <i>deg_freedom</i> .
Examples	StatChiDist (2, " <i>x to p</i> ", 5) calculates the one-tailed chi- squared distribution at the value 5 with 2 degrees of freedom
Guidelines	the number of degrees of freedom <i>deg_freedom</i> must be in the range 1-32767
	<i>statistic</i> is an integer value 1 to 12 or a string indicating the statistic to be returned. For more information see the section Distribution Functions at the start of this chapter.
	<pre>value must be >= 0 (when an x value is entered) or >=0 and <=1 (when a p value is entered)</pre>

Description	CORRELATIONCOEFF (<i>data1,data2</i>) calculates the correlation coefficient between two data sets <i>data1</i> and <i>data2</i> .
Examples	StatCorrelationCoeff (<i>A</i> 1: <i>A</i> 100, <i>B</i> 1: <i>B</i> 100) calculates the correlation coefficient between two data sets located in <i>A</i> 1: <i>A</i> 100 and <i>B</i> 1: <i>B</i> 100.
Guidelines	<i>data1</i> and <i>data2</i> must have the same number of elements.

COUNT

CORRELATION

COEFF

CHISQ

Description	COUNT (<i>data1,data2,dataN</i>) calculates the number of elements in <i>data1,data2</i> through <i>dataN</i> , which can be numbers, arrays, or ranges.
Examples	StatCount (<i>A1:A100,B1:B100</i>) calculates the number of elements in the two data sets located in <i>A1:A100</i> and <i>B1:B100</i> .
Guidelines	<i>data1,data2,dataN</i> are 1 to 30 arguments, which can be numbers, arrays, or ranges.

COUNTCATEGO RIES	Description	COUNTCATEGORIES (<i>range</i>) returns the number of categories in the specified Excel <i>range</i> .
	Examples	StatCountCategories (C1:C100) returns the number of the categories in the range C1:C100.
	Guidelines	range is a valid Excel range

COUNTCELLSB YTYPE

Description	COUNTCELLSBYTYPE (<i>range, type</i>) calculates the number of elements in the entered <i>range</i> which are of the specified <i>type</i> .
Examples	StatCountCellByType (<i>A1:A100,1</i>) calculates the number of elements in the data set located in <i>A1:A100</i> that are numeric.
Guidelines	<i>range</i> is a valid Excel range <i>type</i> is 1=numeric, 2=non-empty, 3=non-empty, non- numeric, 4=empty. Note: StatTools considers a cell containing only spaces to be an empty cell.

COUNTRANGE

Description	COUNTRANGE (<i>range,minValue,maxValue,</i> <i>includeMin,includeMax</i>) calculates number of values in the <i>range</i> that fall between <i>minValue</i> and <i>maxValue</i> . Values equaling <i>minValue</i> and <i>MaxValue</i> can be included by setting <i>IncludeMin</i> and/or <i>IncludeMax</i> to TRUE.
Examples	StatCountRange (<i>A</i> 1: <i>A</i> 100,1,10, <i>TRUE</i> , <i>TRUE</i>) calculates the number of values in the data set located in <i>A</i> 1: <i>A</i> 100 that fall in the range >=1 and <=10.
Guidelines	<pre>range is the range of cells from which you want to count values. minValue is the minimum value in the range. maxValue is the maximum value in the range. includeMin is a boolean value indicating if the MinimumValue should be included in the count. Default is TRUE. includeMax is a boolean value indicating if the MaximumValue should be included in the count. Default is TRUE.</pre>

COVARIANCE

Description	COVARIANCE (<i>data1,data2</i>) calculates the sample covariance between the data sets <i>data1</i> and <i>data2</i> .	
Examples	StatCovariance (<i>A1:A100,B1:B100</i>) calculates the sample covariance between two data sets located in <i>A1:A100</i> and <i>B1:B100</i> .	
Guidelines	<i>data1</i> and <i>data2</i> can be arrays or ranges. <i>data1</i> and <i>data2</i> must have the same number of elements.	

COVARIANCEP

Description	COVARIANCEP (<i>data1, data2</i>) calculates the population covariance between 2 data sets <i>data1</i> and <i>data2</i> .	
Examples	StatCovarianceP (<i>A1:A100,B1:B100</i>) calculates the population covariance between two data sets located in <i>A1:A100</i> and <i>B1:B100</i> .	
Guidelines	data1 and data2 can be arrays or ranges.	

DESTACK

Description	DESTACK (<i>data_range,categories_range_1,category_1,cate gories_range_2,category_2</i>) extracts data in a specified category <i>category_1</i> from the stacked data in <i>data_range</i> . This function allows other StatTools statistics functions to take stacked data as input. Thus, the StatDestack function will only be seen as "embedded" in other functions, as shown in the example here. It returns an array of data for the specified category, extracted from <i>data_range</i> .
Examples	StatMean(StatDestack (B1:B100,A1:A100,"Male")) calculates the mean for the values in the range B1:B100 where the corresponding category range A1:A100 has the value "Male".
Guidelines	<pre>data_range is the range with the stacked data. categories_range_1 is the range with the first category name. category_1 is the first category for which to get the data. categories_range_2 (optional) is the range with the second category name. category_2 (optional) is the second category for which to get the data.</pre>

DURBINWATSON

Description	DURBINWATSON (<i>data</i>) calculates the Durbin- Watson statistic for the values in the data set <i>data</i> .
Examples	StatDurbinWatson (<i>A1:A100</i>) calculates the Durbin-Watson statistic for the range of data located in A1:A100.
Guidelines	data can be an array or range of data.

Description	F (<i>deg_freedom1,deg_freedom2,statistic,value</i>) calculates the <i>statistic</i> for the F distribution using the numerator degrees of freedom <i>deg_freedom1</i> and the denominator degrees of freedom <i>deg_freedom2</i> .
Examples	StatF (1,1,"x to p",1.5) calculates the F Distribution p value for an x value of 1.5 with a numerator degrees of freedom =1 and a denominator degrees of freedom =1.
Guidelines	<pre>deg_freedom1 and deg_freedom2 must be an integer > 0. statistic is an integer value 1 to 12 or a string indicating the statistic to be returned. For more information see the section Distribution Functions at the start of this chapter. value must be >= 0.</pre>

GETCELLVALU

E	S	

F

Description	GETCELLVALUES (<i>range,typeOfCell</i>) gets the values for all the cells of a specified type in the range
Examples	StatGetCellValues (<i>A1:A100,2</i>) gets the values for non- empty cells in the range located in <i>A1:A100</i> .
Guidelines	<i>range</i> is the range of cells from which you want to get the values. <i>typeOfCell</i> is 0 for all, 1 for numeric, 2 for non-empty, 3 for non-empty, non-numeric, and 4 for empty cells.

KURTOSIS

Description	KURTOSIS (<i>Data1,Data2,DataN</i>) calculates the sample kurtosis of the data specified in <i>Data1,Data2,DataN</i> . Note: StatKurtosis calculated on normally distributed data returns the value 3.
Examples	StatKurtosis (<i>A</i> 1: <i>A</i> 100,{1;2;3;2.4}) calculates the sample kurtosis of the data set located in <i>A</i> 1: <i>A</i> 100 and <i>the values</i> 1,2,3 <i>and</i> 2.4.
Guidelines	Data1,Data2,DataN can be numbers, arrays, or ranges.

KURTOSISP

Description	KURTOSISP (<i>Data1,Data2,DataN</i>) calculates the population kurtosis of the data specified in <i>Data1,Data2,DataN</i> . Note: StatKurtosisP calculated on normally distributed data returns the value 3.
Examples	StatKurtosisP (<i>A</i> 1: <i>A</i> 100,{1;2;3;2.4}) calculates the population kurtosis of the data set located in <i>A</i> 1: <i>A</i> 100 and <i>the values</i> 1,2,3 <i>and</i> 2.4.
Guidelines	Data1,Data2,DataN can be numbers, arrays, or ranges.

Description	LN(x) calculates the natural logarithm.
Examples	StatLN(4.5) calculates natural logarithm of 4.5.
Guidelines	<i>x</i> must be a positive, real number.

MAX

LN

Description	MAX (<i>Data1,Data2,DataN</i>) calculates the maximum of the data specified in <i>Data1,Data2,DataN</i> .
Examples	StatMax (<i>A1:A100,</i> { <i>1;2;3;2.4</i> }) calculates the maximum value in the data set located in <i>A1:A100</i> and the values <i>1,2,3</i> and <i>2.4</i> .
Guidelines	<i>Data1,Data2,DataN</i> can be numbers, arrays, or ranges.

MEAN

Description	MEAN (<i>Data1,Data2,DataN</i>) calculates the mean of the data specified in <i>Data1,Data2,DataN</i> .
Examples	StatMean (<i>A</i> 1: <i>A</i> 100,{1;2;3;2.4}) calculates the mean value in the data set located in <i>A</i> 1: <i>A</i> 100 and the values 1,2,3 and 2.4.
Guidelines	Data1,Data2,DataN can be numbers, arrays, or ranges.

MEANABS

Description	MEANABS (<i>Data1,Data2,DataN</i>) calculates the mean of the absolute value of the data specified in <i>Data1,Data2,DataN</i> .
Examples	StatMeanAbs (<i>A1:A100,{1;2;3;2.4}</i>) calculates the mean of the absolute values of the data set located in <i>A1:A100</i> and the values <i>1,2,3</i> and <i>2.4</i> .
Guidelines	Data1,Data2,DataN can be numbers, arrays, or ranges.

MEDIAN

Description	MEDIAN (<i>data,calcFlag</i>) calculates the median of the values located in <i>data</i> . It may be calculated using any of five alternative methods, as optionally specified by <i>calcFlag</i> .
Examples	StatMedian (<i>A1:A100,1</i>) calculates the median value in the data set located in <i>A1:A100</i> . Data is continuous.
Guidelines	<i>data</i> is an Excel range.
	<i>calcFlag</i> is an optional argument that can take an integer value in the range -1 to 5. This value corresponds with the desired method for calculating the percentile.
	-1 or not specified) Automatic (Based on Input Data)
	0) Same as Excel's Percentile Function (Continuous)
	1) Interpolated with Asymmetric Endpoints (Continuous)
	2) Closest Observation (Discrete)
	3) Empirical Dist. Function (Discrete)
	4) Interpolated with Symmetric Endpoints (Continuous)
	5) Empirical Dist. Function with Averaging (Discrete)

MIN

Description	MIN (<i>Data1,Data2,DataN</i>) calculates the minimum of the data specified in <i>Data1,Data2,DataN</i> .
Examples	StatMin (<i>A1:A100,</i> { <i>1;2;3;2.4</i> }) calculates the minimum value in the data set located in <i>A1:A100</i> and the values <i>1,2,3</i> and <i>2.4</i> .
Guidelines	Data1,Data2,DataN can be numbers, arrays, or ranges

MODE

Description	MODE (<i>Data,Is_discrete</i>) calculates the mode of a data set.
Examples	StatMode (<i>A1:A100,FALSE</i>) calculates the mode of the data set located in A1:A100. Data is continuous.
Guidelines	<i>Data</i> is the array or range of data to calculate mode for. Is_ <i>discrete</i> is an optional argument, and specifies whether the data is to be treated as discrete (true), or continuous (false). If missing, it is automatically determined from the data.

NORMAL

Description	NORMAL (<i>mean,std_dev,statistic,value</i>) calculates the <i>statistic</i> for the normal distribution specified by <i>mean</i> and <i>std_dev</i> .
Examples	StatNormal (2,1, " <i>x</i> to <i>p</i> ",3) calculates the p value for an x value of 3 in the normal distribution with mean=2 and std dev =1
Guidelines	mean is the arithmetic mean of the distribution
	std_dev is the standard deviation of the distribution. It must be > 0.
	<i>statistic</i> is an integer value 1 to 12 or a string indicating the statistic to be returned. For more information see the section Distribution Functions at the start of this chapter.
	<i>value</i> must be >=0 and <=1 when a p value is entered.

PAIRCOUNT

Description	PAIRCOUNT (Data1,Data2) counts the number of pairs of cells in Data1 and Data. Only pairs of numeric cells are counted. A missing value in either range will not be counted.
Examples	StatPairCount (<i>A1:A100,B1:B100</i>) counts the number of pairs of numeric cells in the data sets located in <i>A1:A100</i> and <i>B1:B100</i> .
Guidelines	<i>Data1</i> and <i>Data2</i> must be equal sized Excel ranges. Pairs of cells from <i>Data1</i> and <i>Data2</i> are selected on a row-wise basis, starting from the top left.

PAIRMEAN

Description	PAIRMEAN (<i>Data1,Data2</i>) calculates the mean of the differences between pairs of cells in <i>Data1</i> and <i>Data2</i> . Calculations are only made for pairs of numeric cells. A missing value in either range will not be counted.
Examples	StatPairMean (<i>A1:A100,B1:B100</i>) calculates the mean of the differences between pairs of numeric cells in the data sets located in A1:A100 and B1:B100.
Guidelines	<i>Data1</i> and <i>Data2</i> must be equal sized Excel ranges. Pairs of cells from <i>Data1</i> and <i>Data2</i> are selected on a row-wise basis, starting from the top left.
	0
Description	PAIRMEDIAN (<i>Data1,Data2,method_Flag</i>) calculates the median of the differences between pairs of cells in <i>Data1</i> and <i>Data2</i> . Calculations are only made for pairs of numeric cells. A missing value in either range will not be counted. Median may be calculated using any of five alternative methods, as optionally specified by <i>method_Flag</i> .

PAIRMEDIAN

	for the busis, suiting from the top fert.
Description	 PAIRMEDIAN(<i>Data1,Data2,method_Flag</i>) calculates the median of the differences between pairs of cells in <i>Data1</i> and <i>Data2</i>. Calculations are only made for pairs of numeric cells. A missing value in either range will not be counted. Median may be calculated using any of five alternative methods, as optionally specified by <i>method_Flag</i>.
Examples	StatPairMedian (<i>A1:A100,B1:B100</i>) calculates the median of the differences between pairs of numeric cells in the data sets located in A1:A100 and B1:B100.
Guidelines	Data1 and Data2 must be equal sized Excel ranges.
	Pairs of cells from <i>Data1</i> and <i>Data2</i> are selected on a row-wise basis, starting from the top left.
	<i>method_Flag</i> is an optional argument that can take an integer value in the range -1 to 5. This value corresponds with the desired method for calculating the median.
	-1 or not specified) Automatic (Based on Input Data)
	0) Same as Excel's Percentile Function (Continuous)
	1) Interpolated with Asymmetric Endpoints (Continuous)
	2) Closest Observation (Discrete)
	3) Empirical Dist. Function (Discrete)
	4) Interpolated with Symmetric Endpoints (Continuous)
	5) Empirical Dist. Function with Averaging (Discrete)

PAIRSTDDEV

Description	PAIRSTDDEV (<i>Data1,Data2</i>) calculates the sample standard deviation of the differences between pairs of cells in <i>Data1</i> and <i>Data2</i> . Calculations are only made for pairs of numeric cells. A missing value in either range will not be counted.
Examples	StatPairStdDev (<i>A</i> 1: <i>A</i> 100, <i>B</i> 1: <i>B</i> 100) calculates the sample standard deviation of the differences between pairs of numeric cells in the data sets located in A1:A100 and B1:B100.
Guidelines	<i>Data1</i> and <i>Data2</i> must be equal sized Excel ranges. Pairs of cells from <i>Data1</i> and <i>Data2</i> are selected on a row-wise basis, starting from the top left.

PERCENTILE

Description	PERCENTILE (<i>data,p,method_Flag</i>) calculates the <i>p</i> -th percentile of <i>data</i> . Percentiles may be calculated using any of five alternative methods, as optionally specified by <i>method_Flag</i> .
Examples	StatPercentile (<i>A1:A100,.15,0</i>) calculates the 15th percentile for the data located in <i>A1:A100</i> . The data is continuous and uses the weighted average percentile calculation method.
Guidelines	<i>p</i> must be 0-1, inclusive
	<i>method_Flag</i> is an optional argument that can take an integer value in the range -1 to 5. This value corresponds with the desired method for calculating the percentile.
	-1 or not specified) Automatic (Based on Input Data)
	0) Same as Excel's Percentile Function (Continuous)
	1) Interpolated with Asymmetric Endpoints (Continuous)
	2) Closest Observation (Discrete)
	3) Empirical Dist. Function (Discrete)
	4) Interpolated with Symmetric Endpoints (Continuous)
	5) Empirical Dist. Function with Averaging (Discrete)

PRODUCT

Description	PRODUCT (<i>Data1,Data2,DataN</i>) calculates the product of the data specified in <i>Data1,Data2,DataN</i> .
Examples	StatProduct (<i>A</i> 1: <i>A</i> 10,{1;2;3;2.4}) calculates the product of all value in the data set located in <i>A</i> 1: <i>A</i> 10 and the values 1,2,3 and 2.4.
Guidelines	<i>Data1,Data2,DataN</i> can be numbers, arrays, or ranges.

QUARTILE

Description	QUARTILE (<i>data,q, method_Flag</i>) calculates the specified quartile of <i>data</i> . Quartiles may be calculated using any of five alternative methods, as optionally specified by <i>method_Flag</i> .
Examples	StatQuartile (<i>A1:A100,1,FALSE</i>) calculates the 1 st quartile for the data located in A1:A100. The data is continuous.
Guidelines	data must be an Excel range
	<i>Q</i> is the quartile; 0=minimum, 1=1st quartile, 2=2nd quartile (median), 3=3rd quartile, 4=maximum.
	<i>method_Flag</i> is an optional argument that can take an integer value in the range -1 to 5. This value corresponds with the desired method for calculating the percentile.
	-1 or not specified) Automatic (Based on Input Data)
	0) Same as Excel's Percentile Function (Continuous)
	1) Interpolated with Asymmetric Endpoints (Continuous)
	2) Closest Observation (Discrete)
	3) Empirical Dist. Function (Discrete)
	4) Interpolated with Symmetric Endpoints (Continuous)
	5) Empirical Dist. Function with Averaging (Discrete)

RAND

Description	RAND () returns a random number in the range 0 to 1. This function uses the random number generator from Palisade's @RISK product and not Excel's built-in random number generator.
Examples	StatRand () returns a random number in the range 0 to 1.

RANGE D	escription	RANGE (<i>Data1,Data2,DataN</i>) calculates the range (maximum - minimum) of the data specified in <i>Data1,Data2,DataN</i> .
E	xamples	StatRange (<i>A1:A100, {1;2;3;2.4}</i>) calculates the range (maximum - minimum) of the data located in <i>A1:A100</i> and <i>the values 1,2,3 and 2.4</i> .
G	uidelines	<i>Data1,Data2,DataN</i> can be numbers, arrays, or ranges.
RUNSTEST D	escription	RUNSTEST (<i>data,cutoff</i>) calculates the runs statistic for the values in <i>data</i> using <i>cutoff</i> value.
E	xamples	StatRunsTest (<i>A1:A100,StatMean</i> (<i>A1:A100</i>)) calculates the runstest statistic on the data in <i>A1:A100</i> using the mean of the data as the cutoff value.
G	uidelines	data must be an Excel range.
SKEWNESS	escription	SKEWNESS (<i>Data1,Data2,DataN</i>) calculates the sample skewness of the data specified in <i>Data1,Data2,DataN</i> .
E	xamples	StatSkewness (<i>A1:A10, {1;2;3;2.4}</i>) calculates the sample skewness of all value in the data set located in <i>A1:A10</i> and <i>the values 1,2,3 and 2.4</i> .
G	uidelines	Data1,Data2,DataN can be numbers, arrays, or ranges.
SKEWNESSP	escription	SKEWNESSP (<i>Data1,Data2,DataN</i>) calculates the population skewness of the data specified in <i>Data1,Data2,DataN</i> .
E	xamples	StatSkewnessP (<i>A</i> 1: <i>A</i> 10,{1;2;3;2.4}) calculates the population skewness of all value in the data set located in <i>A</i> 1: <i>A</i> 10 and the values 1,2,3 and 2.4.

STANDARDIZE

Description	STANDARDIZE (<i>x</i> , <i>mean</i> , <i>std_dev</i>) calculates a normalized value from a distribution with the specified <i>mean</i> and standard deviation <i>std_dev</i> . <i>x</i> is the value to be normalized.
Examples	StatStandardize (2,1,3) calculates a normalized value at the value 2 from a distribution with a mean of 1 and a standard deviation of 3.
Guidelines	x is the value to be normalized <i>mean</i> is the arithmetic mean of the distribution. std_dev is the standard deviation of the distribution. It must be > 0.

STDDEV

Description	STDDEV (<i>Data1,Data2,DataN</i>) calculates the sample standard deviation of the data specified in <i>Data1,Data2,DataN</i> .
Examples	StatStdDev (<i>A1:A10,{1;2;3;2.4}</i>) calculates the sample standard deviation of all value in the data set located in <i>A1:A10</i> and the values <i>1,2,3</i> and <i>2.4</i> .
Guidelines	<i>Data1,Data2,DataN</i> can be numbers, arrays, or ranges.

STDDEVP

1	population standard deviation of the data specified in <i>Data1,Data2,DataN</i> .
Examples	StatStdDevP (<i>A1:A10,{1;2;3;2.4}</i>) calculates the population standard deviation of all value in the data set located in <i>A1:A10</i> and the values <i>1,2,3</i> and <i>2.4</i> .
Guidelines	<i>Data1,Data2,DataN</i> can be numbers, arrays, or ranges.

SUM

Description	SUM (<i>Data1,Data2,DataN</i>) calculates the sum of the data specified in <i>Data1,Data2,DataN</i> .
Examples	StatSum (<i>A1:A10,</i> { <i>1;2;3;2.4</i> }) calculates the sum of all value in the data set located in <i>A1:A10</i> and <i>the values 1,2,3 and 2.4</i> .
Guidelines	<i>Data1,Data2,DataN</i> can be numbers, arrays, or ranges.

SUMDEVSQ

Description	SUMDEVSQ (<i>Data1,Data2,DataN</i>) calculates the sum of the square of the deviation from the mean of its arguments, which can be numbers, arrays, or ranges. Any missing numbers cause a blank to be returned.
Examples	StatSumDevSq (<i>A</i> 1: <i>A</i> 10,{1;2;3;2.4}) calculates the sum of the square of the deviation from the mean of all values in the data set located in <i>A</i> 1: <i>A</i> 10 and <i>the values</i> 1,2,3 and 2.4.
Guidelines	<i>Data1,Data2,DataN</i> can be numbers, arrays, or ranges.

SUMSQ

Description	SUMSQ (<i>Data1,Data2,DataN</i>) calculates the sum of the square of its arguments, which can be numbers, arrays, or ranges. Any missing numbers cause a blank to be returned.
Examples	StatSumSq (<i>A</i> 1: <i>A</i> 10,{1;2;3;2.4}) calculates the sum of the square of all values in the data set located in <i>A</i> 1: <i>A</i> 10 and <i>the values</i> 1,2,3 <i>and</i> 2.4.
Guidelines	<i>Data1,Data2,DataN</i> can be numbers, arrays, or ranges.

STUDENT

Description	STUDENT (<i>deg_freedom, statistic,value</i>) calculates the <i>statistic</i> for the entered Student's t-distribution.
Examples	StatStudent (5,1, " <i>x to p</i> ",2) calculates the p-value from the Student's t-distribution with 5 degrees of freedom at the x-value of 2.
Guidelines	<i>deg_freedom</i> is an integer indicating the number of degrees of freedom. It must be in the range 1-32767. <i>statistic</i> is an integer value 1 to 12 or a string indicating the statistic to be returned. For more information see the section Distribution Functions at the start of this chapter.
	distribution. It must be ≥ 0 .

VARIANCE

Description	VARIANCE (<i>Data1,Data2,DataN</i>) calculates the sample variance of the data specified in <i>Data1,Data2,DataN</i> .
Examples	StatVariance (<i>A</i> 1: <i>A</i> 10,{1;2;3;2.4}) calculates the sample variance of all value in the data set located in <i>A</i> 1: <i>A</i> 10 and <i>the values</i> 1,2,3 <i>and</i> 2.4.
Guidelines	<i>Data1,Data2,DataN</i> can be numbers, arrays, or ranges.

VARIANCEP

Description	VARIANCEP (<i>Data1,Data2,DataN</i>) calculates the population variance of the data specified in <i>Data1,Data2,DataN</i> .
Examples	StatVarianceP (<i>A</i> 1: <i>A</i> 10,{1;2;3;2.4}) calculates the sample variance of all value in the data set located in <i>A</i> 1: <i>A</i> 10 and <i>the values</i> 1,2,3 <i>and</i> 2.4.
Guidelines	Data1,Data2,DataN can be numbers, arrays, or ranges.

Index

About Command, 177 Application Settings Command, 169 Authorization, 177 Authorization Command, 177 Autocorrelation Command, 116 Box-Whisker Plot Command, 78 C Charts Command, 150 Chi-square Independence Test Command, 100 Chi-Square Normality Test Command, 103 Clear Dialog Memory Command, 175 Combination Command, 60 Confidence Interval - Mean/ Std. Deviation Command, 81 **Confidence Interval - Proportions** Command, 84 Correlations and Covariance Command, 70 Data Set Manager Command, 43 Delete Data Sets Command, 175 Difference Command, 56 Discriminant Analysis Command, 135 Dummy Command, 62 Forecasting Command, 120 Histogram Command, 73 Hypothesis Test - Mean/ Std. Deviation Command, 87 Hypothesis Test - Proportions Command, 90 Icons Desktop, 8 StatTools, 29 Installation Instructions, 7–8 Interaction Command, 58 Lag Command, 54 Lillefors Test Command, 107

Logistic Regression Command, 131 Mann-Whitney Test Command, 165 Menus Data Utilities Menu, 49 Help Menu (Model Window), 177 Normality Tests Menu, 103 **Regression and Classification** Menu. 125 Statistical Inference Menu, 81 Summary Graphs Menu, 73 Summary Statistics Menu, 67 Time Series and Forecasting Menu, 113, 139, 157 Utilities Menu, 169 One Variable Summary Command, 67 One-Way ANOVA Command, 95 P Charts Command, 147 Palisade Corporation, 5 Pareto Chart Command, 140 O-O Normal Plot Command, 110 Random Sample Command, 64 Regression Command, 126 Runs Test for Randomness Command, 118 Sample Size Selection Command, 93 Scatter Plot Command, 76 Sign Test Command, 159 Stack Variables Command, 49 StatTools Functions, 179 Alphabetical List of, 183 StatAutocorrelation, 187 StatAveDev, 187 StatBinomial. 187 StatCategoryIndices, 188 StatCategoryNames, 188

StatCategoryOccurrenceCou nt, 188 StatChiDist, 189 StatCorrelationCoeff, 189 StatCount, 189 StatCountCategories, 189 StatCountCellByType, 190 StatCountRange, 190 StatCovariance. 190 StatCovarianceP, 191 StatDestack, 191 StatDurbinWatson, 191 StatF, 192 **StatGetCellValues**, 192 StatKurtosis. 192 StatKurtosisP, 192 **StatLN**, 193 StatMax. 193 StatMean. 193 StatMeanAbs, 193 StatMedian. 194 StatMin, 194 StatNormal, 195 StatPairCount, 195 StatPairMean, 196 StatPairMedian, 196 StatPairStdDev. 197 StatPercentile, 197 StatProduct, 198 StatQuartile, 198 StatRand, 198

StatRange, 199 StatRunsTest, 199 StatSkewness. 199 StatSkewnessP, 199 StatStandardize, 200 StatStdDev, 200 StatStdDevP, 200 StatStudent, 201 StatSum. 200 StatSumDevSq, 201 StatSumSq, 201 StatVariance. 201 StatVarianceP, 202 System Requirements, 6 Time Series Graph Command, 113 Toolbars StatTools, 29 Transform Variables Command, 52 Two-Way ANOVA Command, 98 U Charts Command, 153 Uninstalling StatTools, 7 Unload StatTools Add-in Command, 175 Unstack Variables Command, 51 Wilcoxon Signed-Rank Test Command, 162 X/R Charts Command, 143