

Chapter 9 ***Data Analysis***

Using Scenarios, Goal Seek, Solver, others

This PDF is designed to be read onscreen, two pages at a time. If you want to print a copy, your PDF viewer should have an option for printing two pages on one sheet of paper, but you may need to start with page 2 to get it to print facing pages correctly. (Print this cover page separately.)

Copyright

This document is Copyright © 2007–2009 by its contributors as listed in the section titled **Authors**. You may distribute it and/or modify it under the terms of either the [GNU General Public License](#), version 3 or later, or the [Creative Commons Attribution License](#), version 3.0 or later.

All trademarks within this guide belong to their legitimate owners.

Authors

Jean Hollis Weber

Feedback

Please direct any comments or suggestions about this document to: authors@user-faq.openoffice.org

Acknowledgments

Some of this chapter is based on articles originally written by Bruce Byfield and first published on the *Linux Journal* website. Some examples are taken from the OOo Help.

Publication date and software version

Published 15 May 2009. Based on OpenOffice.org 3.1.



*You can download
an editable version of this document from
<http://oooauthors.org/en/authors/userguide3/published/>*

Contents

Copyright.....	2
Introduction.....	4
Consolidating data.....	4
Creating subtotals.....	6
Using “what if” scenarios.....	8
Creating scenarios.....	9
Settings.....	10
Working with scenarios using the Navigator.....	11
Providing alternate versions of formulas.....	12
Multiple operations in columns or rows.....	13
Calculating with one formula and one variable.....	13
Calculating with several formulas simultaneously.....	15
Multiple operations across rows and columns.....	16
Calculating with two variables.....	16
Working backwards using Goal Seek.....	17
Goal Seek example.....	18
Using the Solver.....	19

Introduction

Once you are familiar with functions and formulas, the next step is to learn how to automate the processes and perform useful analyses of the data.

Calc includes several tools to help you manipulate functions and formulas, ranging from features for copying and reusing data, to creating subtotals automatically, to varying information to help you find the answers you need. These tools are divided between the Tools and Data menus.

If you are a newcomer to spreadsheets, these tools can be overwhelming at first. However, they become simpler if you remember that they all depend on input from either a cell or a range of cells that contain the data with which you are working.

You can always enter the cells or range manually, but in many cases, you can also use the Shrink/Maximize icon beside a field to change temporarily the size of the tool's window while you select the cells with the mouse. Sometimes, you may have to experiment with which data goes into one field, but, once you have found out, the rest is simply setting a selection of options, many of which can be ignored in any given case. Just keep the basic purpose of each tool in mind, and you should have little trouble with Calc's function tools.

You don't need to learn them, especially if your spreadsheet use is simple, but as your manipulation of data becomes more sophisticated, they can save time in making calculations, especially as you start to deal with hypothetical situations. Just as importantly, they can allow you to preserve your work and to share it with other people—or yourself at a later session.

One function tool not mentioned here are DataPilots (also known as *pivot tables*), but they are a topic that is sufficiently interesting and complex to require an chapter to themselves. (See Chapter 8.)

Consolidating data

Data > Consolidate provides a graphical interface for copying data from one range of cells to another, then running one of a dozen functions on the data. During consolidation, the contents of cells from several sheets can be combined in one place.

- 1) Open the document that contains the cell ranges to be consolidated.

- 2) Choose **Data > Consolidate** to open the Consolidate dialog (Figure 1).

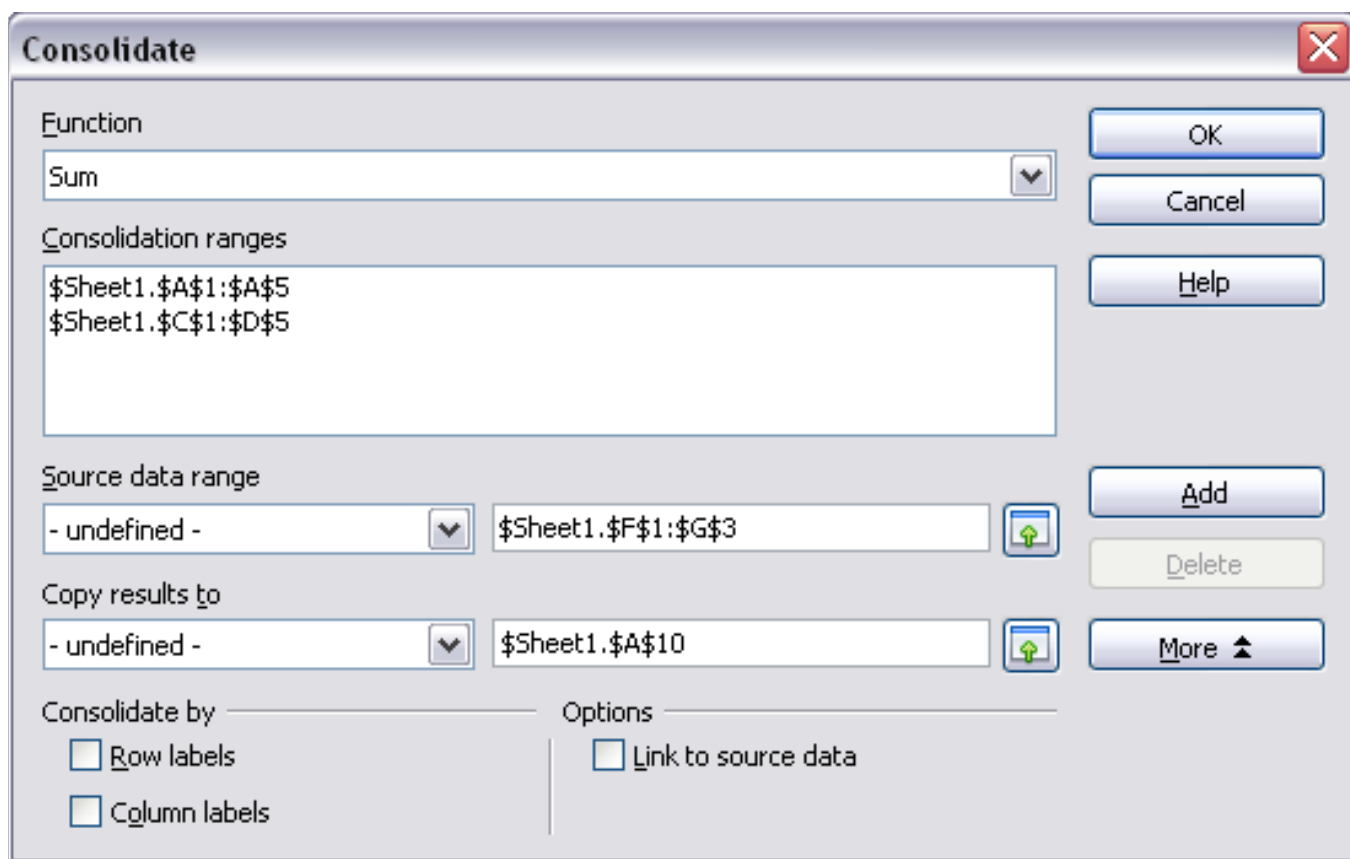


Figure 1: Defining the data to be consolidated

- 3) If the **Source data range** list contains named ranges, you can select a source cell range to consolidate with other areas. If the source range is not named, click in the field to the right and either type a reference for the first source data range or use the mouse to select the range on the sheet. (You may need to move the Consolidate dialog to reach the required cells.)
- 4) Click **Add**. The selected range now appears on the Consolidation ranges list.
- 5) Select additional ranges and click **Add** after each selection.
- 6) Specify where you want to display the result by selecting a target range from the **Copy results to** box. If the target range is not named, click in the field next to **Copy results to** and enter the reference of the target range or select the range using the mouse or position the cursor in the top left cell of the target range.
- 7) Select a function from the Function list. The function specifies how the values of the consolidation ranges are linked. The Sum function is the default setting.

Most of the available functions are statistical (such as AVERAGE, MIN, MAX, STDEV), and the tool is most useful when you are working with the same data over and over.

- 8) Optionally click **More** in the Consolidate dialog to display additional settings.
 - Select **Link to source data** to insert the formulas that generate the results in the target range, rather than the actual results. If you link the data, any values modified in the source range are automatically updated in the target range. The corresponding cell references in the target range are inserted in consecutive rows, which are automatically ordered and then hidden from view. Only the final result, based on the selected function, is displayed.
 - Under **Consolidate by**, select either *Row labels* or *Column labels* if the cells of the source data range are not to be consolidated corresponding to the identical position of the cell in the range, but instead according to a matching row label or column label. To consolidate by row labels or column labels, the label must be contained in the selected source ranges. The text in the labels must be identical, so that rows or columns can be accurately matched. If the row or column label does not match any that exist in the target range, it will be appended as a new row or column.
- 9) Click **OK** to consolidate the ranges.

If you are continually working with the same range, then you probably want to use **Data > Create Range** to give it a name.

The data from the consolidation ranges and target range are saved when you save the document. If you later open a document in which consolidation has been defined, this data will again be available.

Creating subtotals

SUBTOTAL is a function listed under the Mathematical category when you use the Function Wizard (**Insert > Function**). Because of its usefulness, the function has a graphical interface accessible from **Data > Subtotals**.

As the name suggests, SUBTOTAL totals data arranged in an array—that is, a group of cells with labels for columns and/or rows. Using the Subtotals dialog, you can select arrays, then choose a statistical function to apply to them. For efficiency, you can choose up to three

groups of arrays to which to apply a function. When you click **OK**, Calc adds subtotals and grand totals to the selected arrays, using the Result and Result2 cell styles for them.

To insert subtotal values into a sheet:

- 1) Ensure that the columns have labels.
- 2) Select the range of cells that you want to calculate subtotals for, and then choose **Data > Subtotals**.
- 3) In the Subtotals dialog (Figure 2), in the **Group by** box, select the column that you want to add the subtotals to. If the contents of the selected column change, the subtotals are automatically recalculated.
- 4) In the **Calculate subtotals for** box, select the columns containing the values that you want to subtotal.
- 5) In the **Use function** box, select the function that you want to use to calculate the subtotals.
- 6) Click **OK**.

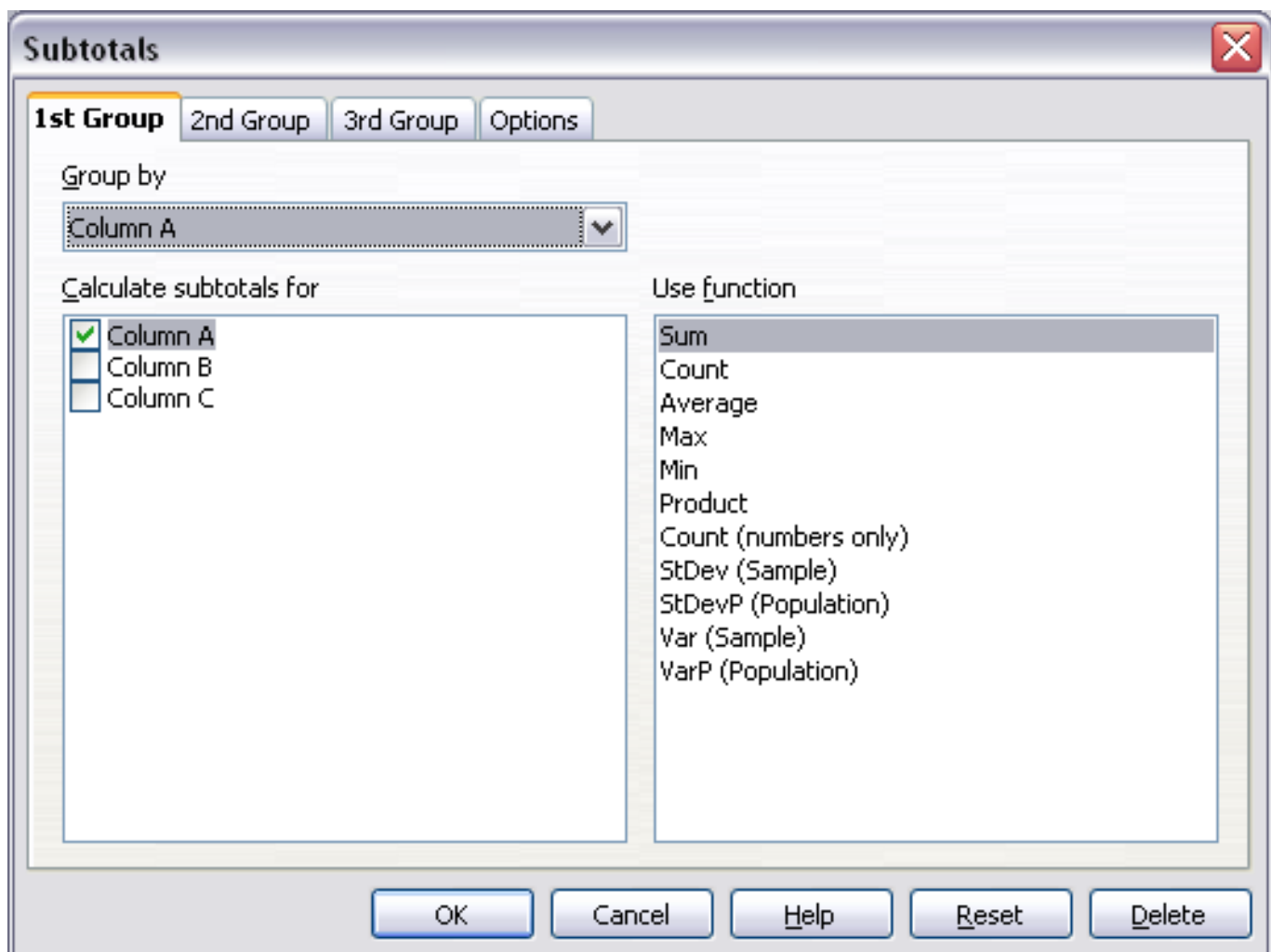


Figure 2: Setting up subtotals

If you use more than one group, then you can also arrange the subtotals according to choices made on the dialog's *Options* page (Figure 3), including ascending and descending order or using one of the predefined custom sorts defined in **Tools > Options > OpenOffice.org Calc > Sort Lists**.

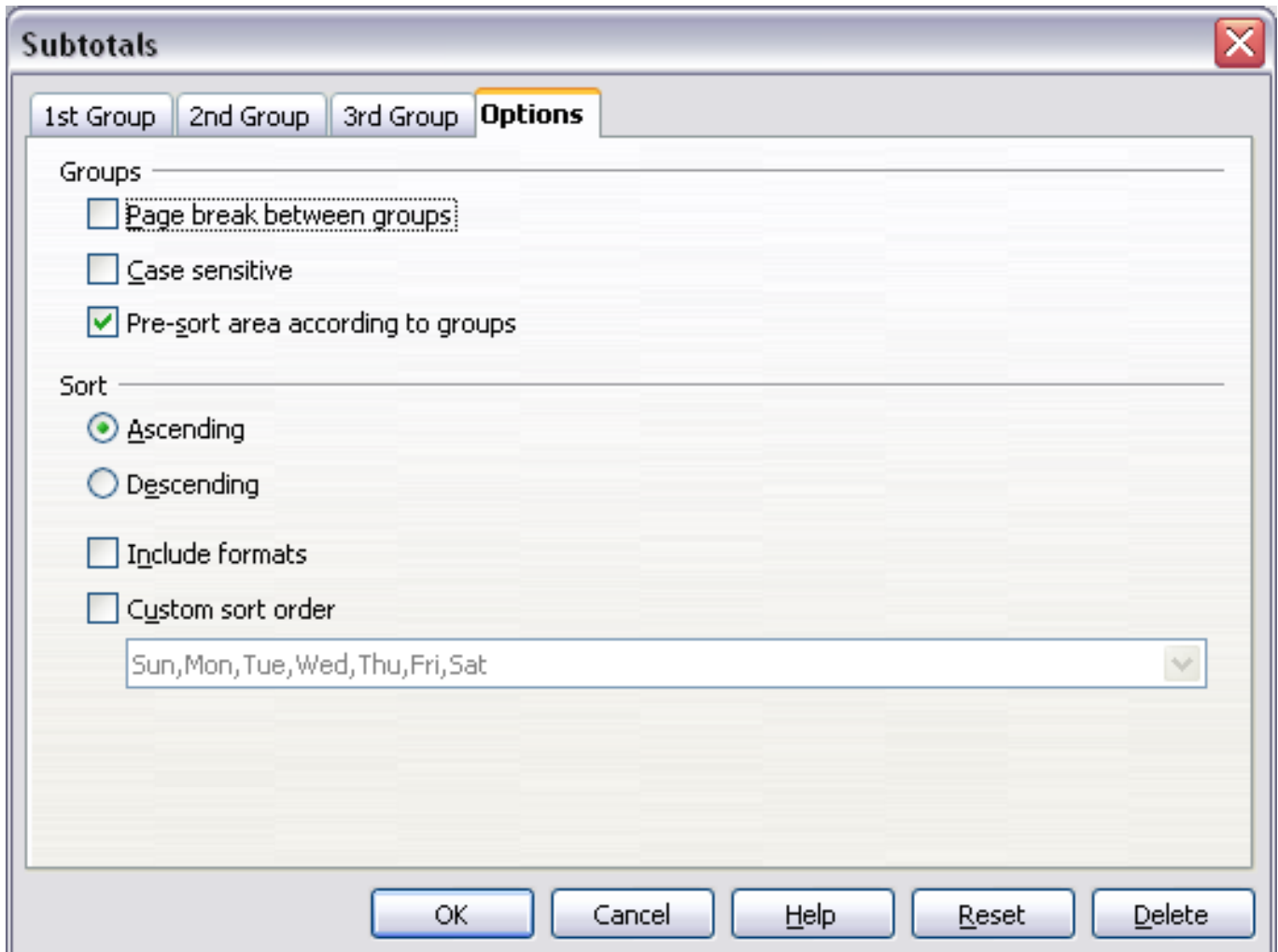


Figure 3: Choosing options for subtotals

Using “what if” scenarios

Scenarios are a tool to test “what-if” questions. Use **Tools > Scenarios** to enter variable contents—scenarios—in the same cell. Each scenario is named, and can be edited and formatted separately, and chosen from a drop-down list in the Navigator and the title bar of the scenario. When you print the spreadsheet, only the contents of the currently active scenario is printed.

By adding a scenario, you can quickly change the arguments of a formula and view the new results. For example, if you wanted to calculate different interest rates on an investment, you could add a

scenario for each interest rate, and quickly view the results. If you had another formula that calculated your yearly income and included the result of the interest rate formula as an argument, it would also be updated. If all your sources of incomes used scenarios, you could efficiently build a complex model of your possible income.

Creating scenarios

To create a scenario, select all the cells that provide the data for the scenario.

- 1) Select the cells that contain the values that will change between scenarios. To select multiple cells, hold down the *Ctrl* key as you click each cell.
- 2) Choose **Tools > Scenarios**.
- 3) On the Create Scenario dialog (Figure 4), enter a name for the new scenario. It's best to use a name that clearly identifies the scenario, not the default name as shown in the illustration. This name is displayed in the Navigator and on the title bar of the scenario on the sheet itself.

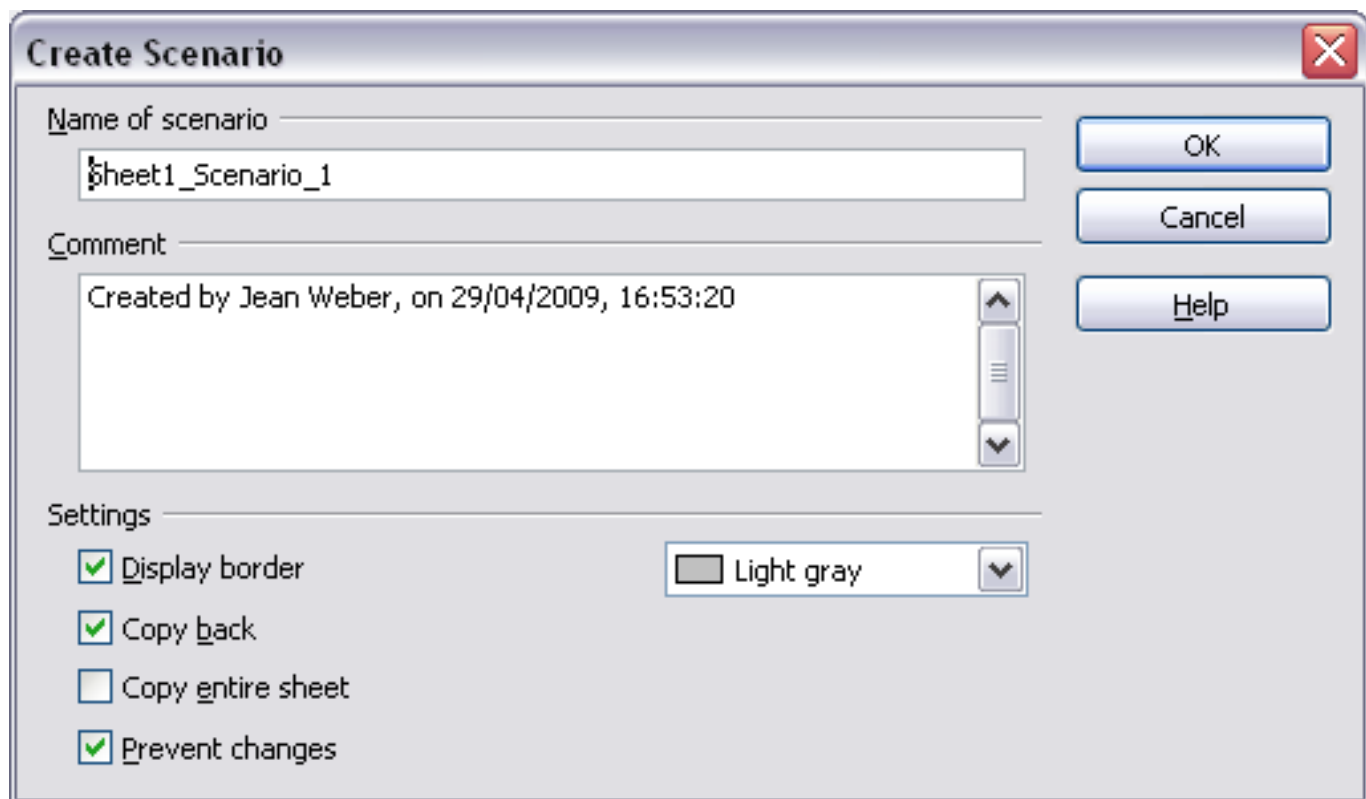


Figure 4: Creating a scenario

- 4) Optionally add some information to the **Comment** box. The example shows the default comment. This information is

displayed in the Navigator when you click the Scenarios icon and select the desired scenario.

- 5) Optionally select or deselect the options in the *Settings* section. See below for more information about these options.
- 6) Click **OK** to close the dialog. The new scenario is automatically activated.

Settings

The lower portion of the Create Scenario dialog contains several options. In most cases the default settings (shown selected in the example) are suitable.

Display border

Highlights the scenario in your table with a border. The color for the border is specified in the field to the right of this option. The border has a title bar displaying the name of the last scenario. The button on the right of the scenario border offers you an overview of all the scenarios in this area, if several have been defined. You can choose any of the scenarios from this list without restrictions.

Copy back

Copies the values of cells that you change into the active scenario. If you do not select this option, the scenario is not changed when you change cell values. The behavior of the **Copy back** setting depends on the cell protection, the sheet protection, and the **Prevent changes** settings.

Copy entire sheet

Copies the entire sheet into an additional scenario sheet.

Prevent changes

Prevents changes to the active scenario.

- You can only change the scenario properties if the **Prevent changes** option is not selected and if the sheet is not protected.
- You can only edit cell values if the **Prevent changes** option is selected, if the **Copy back** option is not selected, and if the cells are not protected.
- You can only change scenario cell values and write them back into the scenario if the **Prevent changes** option is not selected, if the **Copy back** option is selected, and if the cells are not protected.

Working with scenarios using the Navigator

After scenarios are added to a spreadsheet, you can jump to a particular scenario by using the Navigator, then selecting a scenario from the list. You can also color code scenarios to make them easier to distinguish from one another.

To select a scenario in the Navigator, click the **Scenarios** icon in the Navigator. The defined scenarios are listed, with the comments that were entered when the scenarios were created.

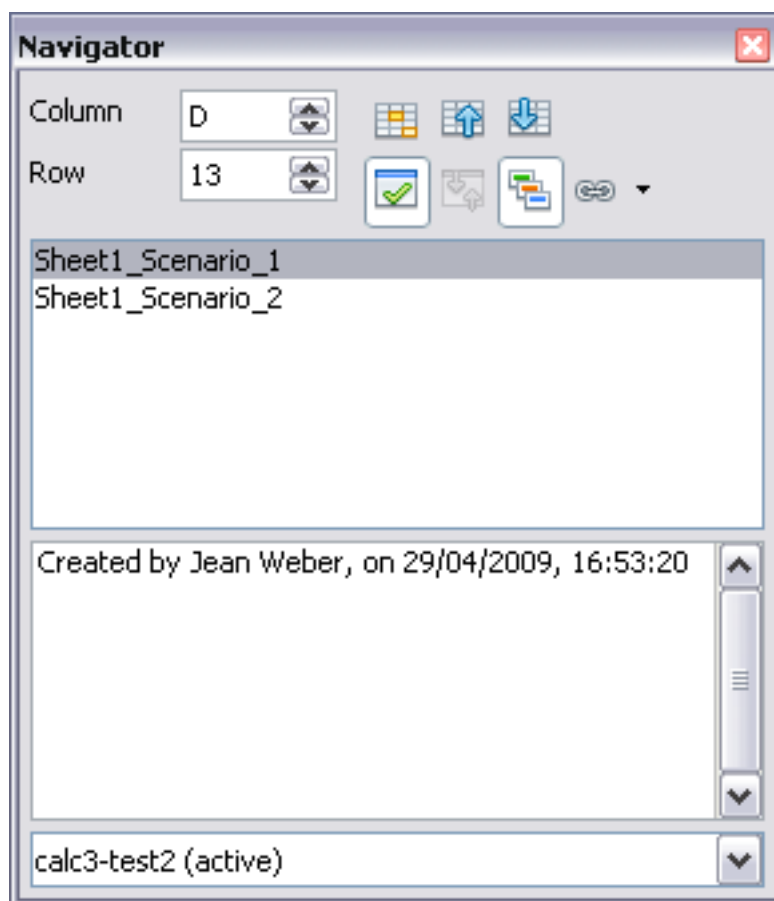


Figure 5: Scenarios in the Navigator

Double-click a scenario name in the Navigator to apply that scenario to the current sheet.

To delete a scenario, right-click the name in the Navigator and choose **Delete**.

To edit a scenario, including its name and comments, right-click the name in the Navigator and choose **Properties**. The Edit Properties dialog is the same as the Create Scenario dialog (Figure 4).

To learn which values in the scenario affect other values, choose **Tools > Detective > Trace Dependents**. Arrows point to the cells that are directly dependent on the current cell.

Providing alternate versions of formulas

Like scenarios, **Data > Multiple Operations** is a planning tool for “what if” questions. Unlike a scenario, the Multiple Operations tool does not present the alternate versions in the same cells or with a drop-down list. Instead, the Multiple Operations tool creates a formula array: a separate set of cells that give all the alternative results for the formulas used. Although the tool is not listed among the functions, it is really a function that acts on other functions, allowing you to calculate different results without having to enter and run them separately.

To use the Multiple Operations tool, you need two arrays of cells. The first array contains the original or default values and the formulas applied to them. The formulas must be in a range.

The second array is the formula array. It is created by entering variables—alternative values—for one or two of the original values. Once the variable values are created, you use the Multiple Operations tool to specify the formulas you are using, as well as the original values altered by the variables. The result is the outcome of the formulas using each variable.

The Multiple Operations tool can use any number of formulas, but only one or two variables. With one variable, you add different values for the variables as column or row labels. You then add a column to the right of the column labels or a row below the row labels (whichever you are using) for the results of each formula using the variable values. With two variables, the values for one variable are used as column labels, and the values for the other variable as row labels.

Setting up multiple operations can be confusing at first. For example, when using two variables, you need to select them carefully, so that they form a meaningful table. Not every pair of variables is useful to add to the same formula array. Yet, even when working with a single variable, a new or tired user can easily make mistakes or forget the relation between cells in the original array and cells in the formula array. In these situations, **Tools > Detective** can help to clarify the relations.

You can also make formula arrays easier to work with if you apply some simple design logic. Place the original and the formula array close together on the same sheet, and use labels for the rows and columns in both. These small exercises in organizational design will make working with the formula array much less painful, particularly when you are correcting mistakes or adjusting results.

Note

If you export a spreadsheet containing multiple operations to Microsoft Excel, the location of the cells containing the formula must be fully defined relative to the data range.

Multiple operations in columns or rows

In your spreadsheet, enter a formula to calculate a result from values that are stored in other cells. Then, set up a cell range where you enter some fixed values, and the Multiple Operations command will calculate the results depending on the formula.

In the Formula field, enter the cell reference to the formula that applies to the data range. In the Column input cell/Row input cell field, enter the cell reference to the corresponding cell that is part of the formula. This can be explained best by examples. (Cell references correspond to those in the following figures.)

Let's say you produce toys that you sell for \$10 each (cell B1). Each toy costs \$2 to make (cell B2), in addition to which you have fixed costs of \$10,000 per year (cell B3). How much profit will you make in a year if you sell a particular number of toys?

Calculating with one formula and one variable

- 1) To calculate the profit, first enter any number as the quantity (items sold); in this example, 2000 (cell B4). The profit is found from the formula *Profit=Quantity * (Selling price - Direct costs) - Fixed costs*. Enter this formula in B5: `=B4*(B1-B2) -B3`.
- 2) In column D enter given annual sales, one below the other; for example, 500 to 5000, in steps of 500.
- 3) Select the range D2:E11, and thus the values in column D and the empty cells alongside in column E.
- 4) Choose **Data > Multiple Operations**.
- 5) With the cursor in the *Formulas* field of the Multiple operations dialog, click cell B5.
- 6) Set the cursor in the *Column input cell* field and click cell B4. This means that B4, the quantity, is the variable in the formula, which is replaced by the selected column values. Figure 6 shows the worksheet and the Multiple operations dialog.
- 7) Click **OK**. The profits for the different quantities are now shown in column E. See Figure 7.

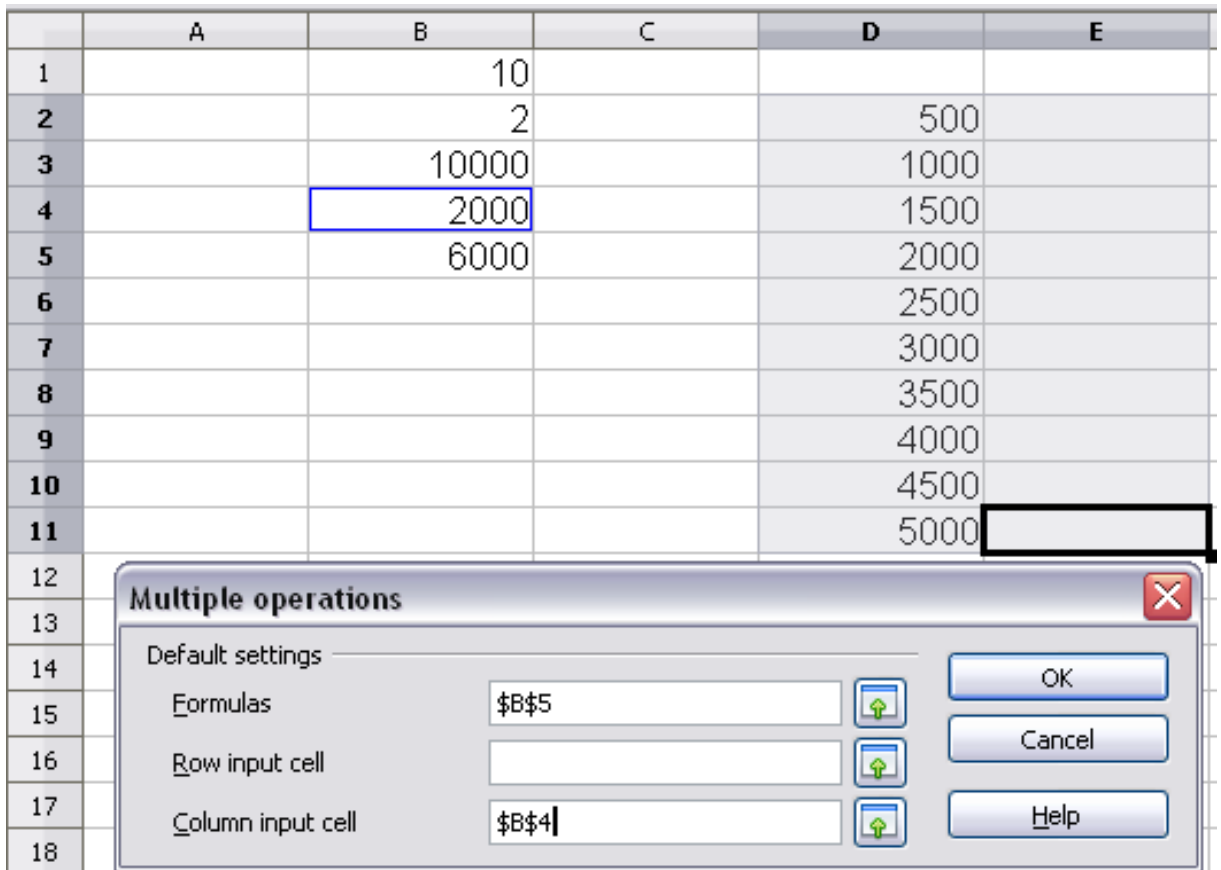


Figure 6: Sheet and Multiple operations dialog showing input

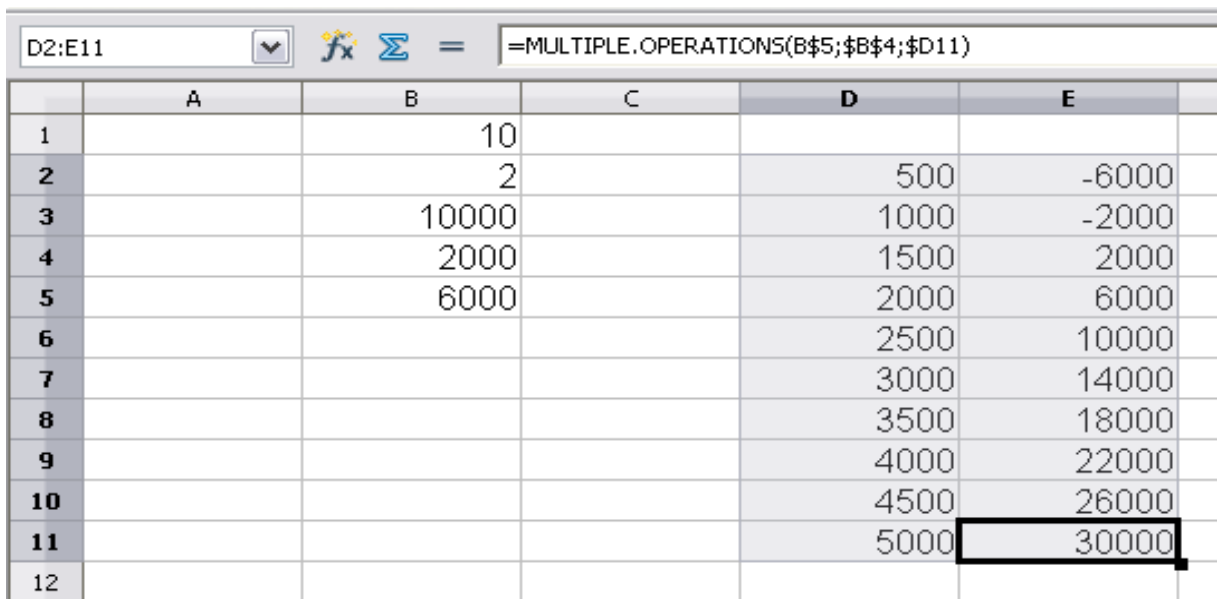


Figure 7: Sheet showing results of multiple operations calculations

Tip You may find it easier to mark the required reference in the sheet if you click the Shrink icon to reduce the Multiple operations dialog to the size of the input field. The icon then changes to the Maximize icon; click it to restore the dialog to its original size.

Calculating with several formulas simultaneously

- 1) In the sheet from the previous example, delete the contents of column E.
- 2) Enter the following formula in C5: $=B5/B4$. You are now calculating the annual profit per item sold.
- 3) Select the range D2:F11, thus three columns.
- 4) Choose **Data > Multiple Operations**.
- 5) With the cursor in the *Formulas* field of the Multiple operations dialog, select cells B5 thru C5.
- 6) Set the cursor in the *Column input cell* field and click cell B4. Figure 8 shows the worksheet and the Multiple operations dialog.

	A	B	C	D	E	F
1		10				
2		2		500		
3		10000		1000		
4		2000		1500		
5		6000	3	2000		
6				2500		
7				3000		
8				3500		
9				4000		
10				4500		
11				5000		

Field	Value
Formulas	\$B\$5:\$C\$5
Row input cell	
Column input cell	\$B\$4

Figure 8: Sheet and dialog showing input

- 7) Click **OK**. Now the profits are listed in column E and the annual profit per item in column F. See Figure 9. (You might want to format column F to show 2 decimal points, for easier comparison of the amounts.)

	A	B	C	D	E	F
1		10				
2		2		500	-6000	-12
3		10000		1000	-2000	-2
4		2000		1500	2000	1.33
5		6000	3	2000	6000	3
6				2500	10000	4
7				3000	14000	4.67
8				3500	18000	5.14
9				4000	22000	5.5
10				4500	26000	5.78
11				5000	30000	6
12						

Figure 9: Results of multiple operations calculations

Multiple operations across rows and columns

You can carry out joint multiple operations for columns and rows in so-called cross-tables. The formula cell must refer to both the data range arranged in rows and the one arranged in columns. Select the range defined by both data ranges and call the multiple operation dialog. Enter the reference to the formula in the Formula field. The Row input cell and the Column input cell fields are used to enter the reference to the corresponding cells of the formula.

Calculating with two variables

You now want to vary not just the quantity produced annually, but also the selling price, and you are interested in the profit in each case.

Expand the table shown above. D2 thru D11 contain the numbers 500, 1000 and so on, up to 5000. In E1 through H1 enter the numbers 8, 10, 15 and 20.

- 1) Select the range D1:H11.
- 2) Choose **Data > Multiple Operations**.
- 3) With the cursor in the *Formulas* field of the Multiple operations dialog, click cell B5 (profit).
- 4) Set the cursor in the *Row input cell* field and click cell B1. This means that B1, the selling price, is the horizontally entered variable (with the values 8, 10, 15 and 20).
- 5) Set the cursor in the *Column input cell* field and click cell B4. This means that B4, the quantity, is the vertically entered variable.

	A	B	C	D	E	F	G	H
1		10		8	10	15	20	
2		2		500	-6000	-12		
3		10000		1000	-2000	-2		
4		2000		1500	2000	1.33		
5		6000	3	2000	6000	3		
6				2500	10000	4		
7				3000	14000	4.67		
8				3500	18000	5.14		
9				4000	22000	5.5		
10				4500	26000	5.78		
11				5000	30000	6		

Multiple operations

Default settings

Formulas:

Row input cell:

Column input cell:

Figure 10: Sheet and dialog showing input

- Click **OK**. The profits for the different selling prices are now shown in the range E2:H11.

D1:H11								
=MULTIPLE.OPERATIONS(\$B\$5;\$B\$4;\$D11;\$B\$1;H\$1)								
	A	B	C	D	E	F	G	H
1		10		8	10	15	20	
2		2		500	-7000	-6000	-3500	-1000
3		10000		1000	-4000	-2000	3000	8000
4		2000		1500	-1000	2000	9500	17000
5		6000	3	2000	2000	6000	16000	26000
6				2500	5000	10000	22500	35000
7				3000	8000	14000	29000	44000
8				3500	11000	18000	35500	53000
9				4000	14000	22000	42000	62000
10				4500	17000	26000	48500	71000
11				5000	20000	30000	55000	80000

Figure 11: Results of multiple operations calculations

Working backwards using Goal Seek

Tools > Goal Seek reverses the usual order for a formula. Usually, you run a formula to get the result when certain arguments are entered. By contrast, with Goal Seek, you work with a completed formula to see what values you need in an argument to get the results that you want.

To take a simple example, imagine that the Chief Financial Officer of a company is developing sales projections for each quarter of the forthcoming year. She knows what the company's total income must be for the year to satisfy stockholders. She also has a good idea of the

company's income in the first three quarters, because of the contracts that are already signed. For the fourth quarter, however, no definite income is available. So how much must the company earn in Q4 to reach its goal? To answer, the CFO enters the projected earnings for each of the other three quarters and the projection for the entire year. Then she runs a goal seek on the cell for Q4 sales, and receives her answer.

Other uses of goal seek may be more complicated, but the method remains the same. To run a goal seek, at least one of the values for an argument must be a referenced cell or range. Only one argument can be altered in a single goal seek. After you get the result of a goal seek, you can replace the original value in the referenced cell with the result, or record the result elsewhere for later use, possibly as a scenario.

With the help of Goal Seek you can calculate a value that, as part of a formula, leads to the result you specify for the formula. You thus define the formula with several fixed values and one variable value and the result of the formula.

Goal Seek example

To calculate annual interest (I), create a table with the values for the capital (C), number of years (n), and interest rate (i). The formula is $I = C * n * i$.

Let us assume that the interest rate i of 7.5% and the number of years n (1) will remain constant. However, you want to know how much the investment capital C would have to be modified in order to attain a particular return I. For this example, calculate how much capital C would be required if you want an annual return of \$15,000.

Enter each of the values for Capital C (an arbitrary value like \$100,000), number of years n (1), and interest rate i (7.5%) in one cell each. Enter the formula to calculate the interest I in another cell. Instead of C, n, and i use the reference to the cell with the corresponding value. In our example, these are B1, B2, and B3.

- 1) Place the cursor in the formula cell (the cell containing the interest I), and choose **Tools > Goal Seek**.
- 2) On the Goal Seek dialog, the correct cell is already entered in the *Formula cell* field.
- 3) Place the cursor in the *Variable cell* field. In the sheet, click in the cell that contains the value to be changed, in this example it is the cell with the capital value C.

- 4) Enter the desired result of the formula in the *Target value* field. In this example, the value is 15000. Figure 12 shows the cells and fields.

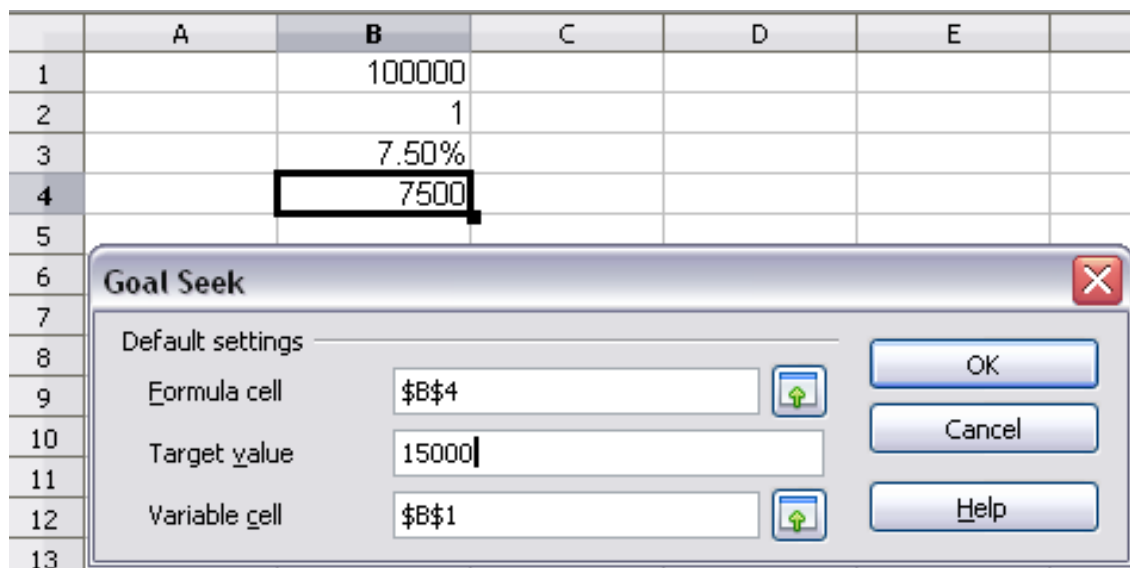


Figure 12: Example setup for goal seek

- 5) Click **OK**. A dialog appears informing you that the Goal Seek was successful. Click **Yes** to enter the result in the cell with the variable value. The result is shown in Figure 13.

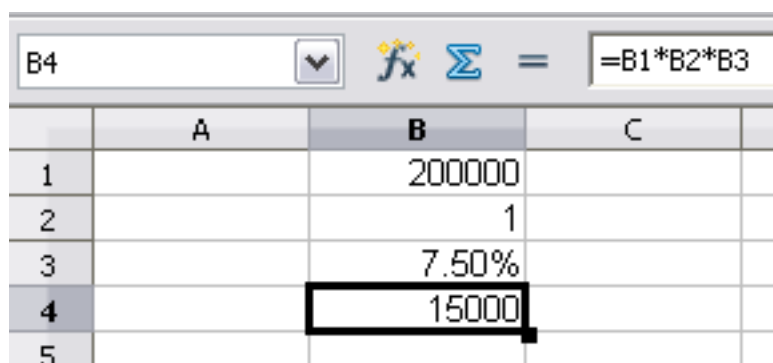


Figure 13: Result of goal seek operation

Using the Solver

Tools > Solver amounts to a more elaborate form of Goal Seek. The difference is that the Solver deals with equations with multiple unknown variables. It is specifically designed to minimize or maximize the result according to a set of rules that you define.

Each of these rules sets up whether an argument in the formula should be greater than, lesser than, or equal to the figure you enter. If you want the argument to remain unchanged, you enter a rule that the cell that contains it should be equal to its current entry. For arguments that

you would like to change, you need to add two rules to define a range of possible values: the limiting conditions. For example, you can set the constraint that one of the variables or cells must not be bigger than another variable, or not bigger than a given value. You can also define the constraint that one or more variables must be integers (values without decimals), or binary values (where only 0 and 1 are allowed). Once you have finished setting up the rules, you can adjust the argument and the results by clicking the **Solve** button.

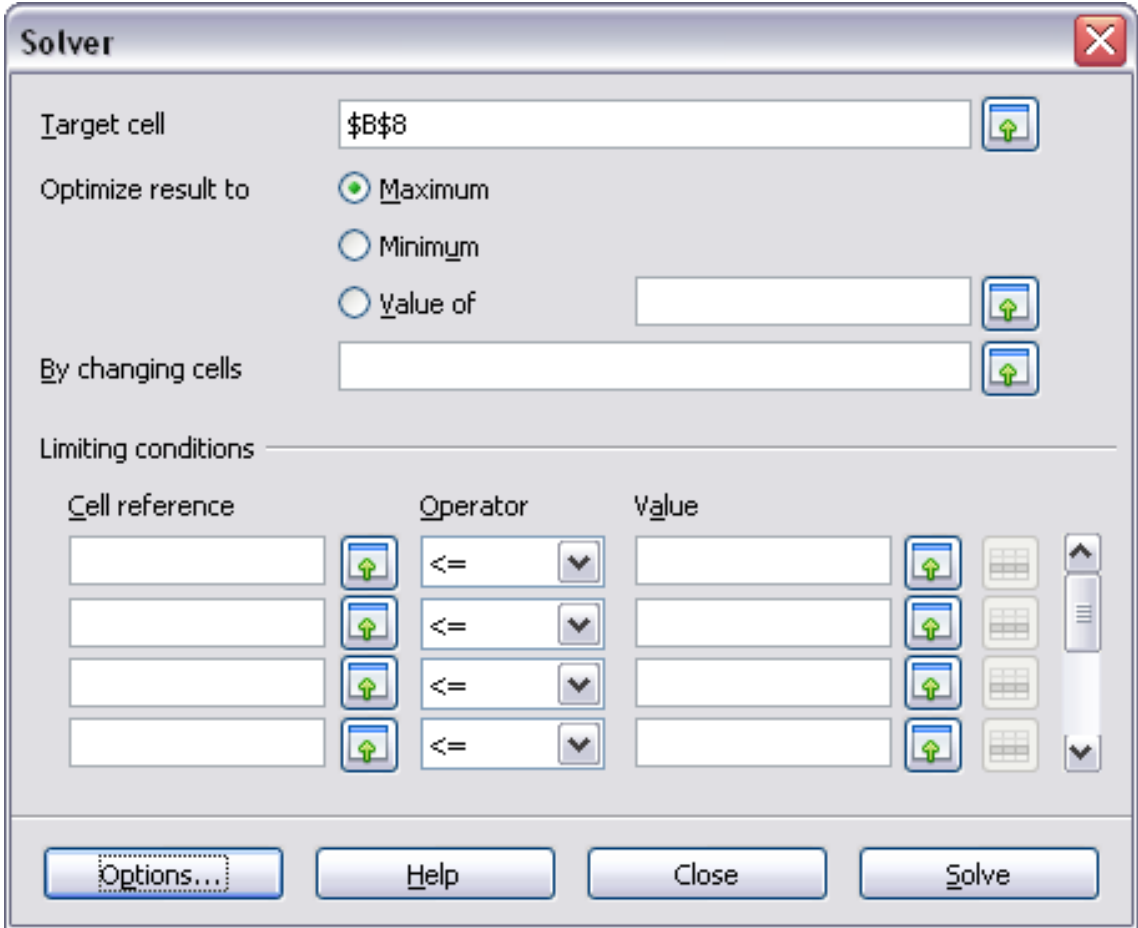


Figure 14: The Solver dialog

Note The default solver supports only linear equations. For nonlinear programming requirements, try the EuroOffice Solver or Sun’s Solver for Nonlinear Programming [Beta]. Both are available from the OpenOffice.org extensions repository. (For more about extensions, see Chapter 14 (Setting up and Customizing Calc.)
