

Data Analysis and Solver Plugins for KSpread

# **USER'S MANUAL**

*Tomasz Maliszewski*  
*tmaliszewski@wp.pl*

## Table of Content

---

<b>CHAPTER 1: INTRODUCTION.....</b>	<b>3</b>
1.1. ABOUT DATA ANALYSIS PLUGIN .....	3
1.3. ABOUT SOLVER PLUGIN.....	3
1.3. SOFTWARE REQUIREMENTS .....	3
<b>CHAPTER 2: INSTALLATION.....</b>	<b>4</b>
2.1. INSTALL THE DATA ANALYSIS PLUGIN .....	4
2.2. INSTALL THE SOLVER PLUGIN.....	4
<b>CHAPTER 3: DATA ANALYSIS TOOLS DESCRIPTION.....</b>	<b>6</b>
3.1. ANOVA.....	6
3.1.1. Anova: Single Factor .....	6
3.1.3. Anova: Two factor With Replication .....	7
3.1.3. Anova: Two factor Without Replication .....	8
3.3. CORRELATION.....	9
3.3. COVARIANCE .....	10
3.4. DESCRIPTIVE STATISTICS .....	11
3.5. EXPONENTIAL SMOOTHING.....	12
3.6. F-TEST TWO-SAMPLE FOR VARIANCES.....	13
3.7. FOURIER ANALYSIS .....	14
3.8. HISTOGRAM .....	15
3.9. MOVING AVERAGE .....	16
3.10. RANK AND PERCENTILE .....	17
3.11. REGRESSION .....	17
3.13. SAMPLING.....	19
3.13. T-TEST .....	20
3.13.1. t-Test: Two-Sample Assuming Equal Variances .....	20
3.13.3. t-Test: Two-Sample Assuming Unequal Variances .....	21
3.13.3. t-Test: Paired Two Sample For Means .....	22
3.14. Z-TEST: TWO SAMPLE FOR MEANS .....	24
<b>CHAPTER 4: SOLVER TOOL DESCRIPTION .....</b>	<b>26</b>
4.1. SOLVER PARAMETERS.....	26
4.2. ADD CONSTRAINT.....	27
4.3. CHANGE CONSTRAINT .....	27
4.4. SOLVER OPTIONS .....	28

## Chapter 1: Introduction

### 1.1. About Data Analysis Plugin

---

This software is Data Analysis Plugin for KSpread. It provides a set of data analysis tools – that you can use to save steps when you develop complex statistical or engineering analyses. You provide the data and parameters for each analysis; the tool uses the appropriate statistical or engineering functions and then displays the results in an output table. Some tools generate charts in addition to output tables.

Accessing the data analysis tools Data Analysis Plugin includes the tools described in this manual. To access these tools, click Data Analysis on the Tools menu. If the Data Analysis command is not available, the installation of the Data Analysis Plugin which is described in this manual was not successful.

### 1.3. About Solver Plugin

---

Solver is part of commands sometimes called what-if analysis (what-if analysis: A process of changing the values in cells to see how those changes affect the outcome of formulas on the worksheet. For example, varying the interest rate that is used in an amortization table to determine the amount of the payments.) tools. With Solver, you can find an optimal value for a formula (formula: A sequence of values, cell references, names, functions, or operators in a cell that together produce a new value. A formula always begins with an equal sign (=) in one cell – called the target cell – on a worksheet. Solver works with a group of cells that are related directly to the formula in the target cell. It adjusts the values in the changing cells you specify – called the adjustable cell – to produce the result you specify from the target cell formula. You can apply constraints (constraints: The limitations placed on a Solver problem. You can apply constraints to adjustable cells, the target cell, or other cells that are directly related to the target cell.) to restrict the values Solver can use in the model, and the constraints can refer to other cells that affect the target cell formula. Formulas in target cell and constraints must be linear mathematical relations.

Use Solver to determine the maximum or minimum value of one cell by changing other cells – for example, you can change the amount of your projected advertising budget and see the affect on your projected profit amount.

To access Solver tool, click Solver on the Tools menu. If the Solver command is not available, the installation of the Solver Plugin which is described in this manual was not successful.

### 1.3. Software Requirements

---

Data Analysis Plugin for KSpread requires KSpread and KChart from KOffice Suite in version 1.5.1 of the source code distribution. Solver Plugin requires only KSpread of this suite. All the external software required by KOffice is also required by this application.

Note: There is no guarantee that forward versions of KOffice will be compatible with this software.

## Chapter 2: Installation

### 2.1. Install the Data Analysis Plugin

---

To install this package do the following:

1. Install the programs and any data files and documentation from KOffice's package source code as it is described in "INSTALL" file in the root directory of the package. Installation of KSpread and KChart must be successful. DO NOT follow step 5 described in the file by typing 'make clean' and/or 'make distclean'.
2. Copy all directories from the plugin's package source code to KOffice's package by copying them to the following directory:

```
[KOffice_Root_Directory]/kspread/plugins/
```

3. Modify "configure" file in the root directory of the KOffice package. To do this find in the file the following line:

...

```
ac_config_files="$ac_config_files kspread/plugins/insertcalendar/Makefile"
```

...

And after this line add the following line:

```
ac_config_files="$ac_config_files kspread/plugins/dataAnalysis/Makefile"
```

4. Type './configure' in the root directory of the KOffice package to run modified "configure" file.
5. Type 'cd ./kspread/plugins/dataAnalysis' from the root directory of the KOffice package.
6. Type 'make' to compile plugin's package.
7. Type 'make install' to install plugin's binaries.
8. You can remove the program binaries and object files by following step 5 described in mentioned "INSTALL" file.

Note: Solver Plugin and Data Analysis Plugin can be installed at the same time.

### 2.2. Install the Solver Plugin

---

The installation of Solver Plugin is very similar to the installation of Data Analysis Plugin. To install this package do the following:

1. Install the programs and any data files and documentation from KOffice's package source code as it is described in "INSTALL" file in the root directory of the package. Installation of KSpread and KChart must be successful. DO NOT follow step 5 described in the file by typing 'make clean' and/or 'make distclean'.

2. Copy all directories from the plugin's package source code to KOffice's package by copying them to the following directory:

*[KOffice\_Root\_Directory]/kspread/plugins/*

3. Modify "configure" file in the root directory of the KOffice package. To do this find in the file the following line:

...

```
ac_config_files="$ac_config_files kspread/plugins/insertcalendar/Makefile"
```

...

And after this line add the following line:

```
ac_config_files="$ac_config_files kspread/plugins/solver/Makefile"
```

4. Type './configure' in the root directory of the KOffice package to run modified "configure" file.
5. Type 'cd ./kspread/plugins/solver' from the root directory of the KOffice package.
6. Type 'make' to compile plugin's package.
7. Type 'make install' to install plugin's binaries.
8. You can remove the program binaries and object files by following step 5 described in "INSTALL" file, which was mentioned above.

Note: Solver Plugin and Data Analysis Plugin can be installed at the same time.

## Chapter 3: Data Analysis Tools Description

### 3.1. Anova

---

The Anova analysis tools provide different types of variance analysis. The tool to use depends on the number of factors and the number of samples you have from the populations you want to test.

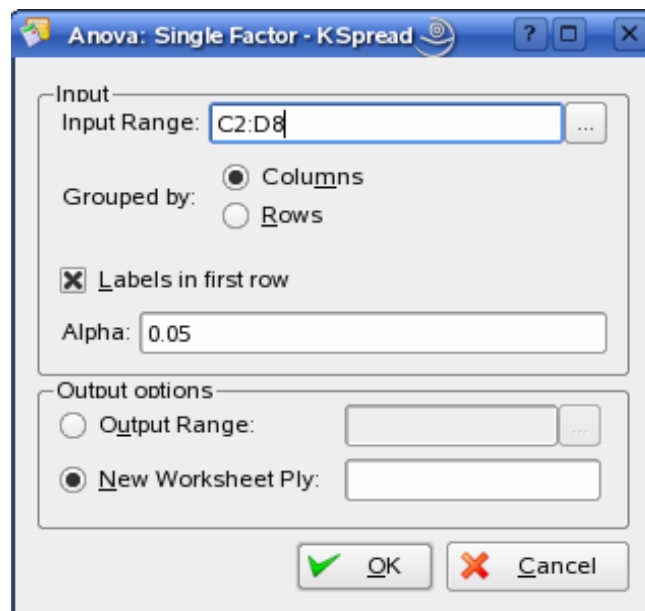
#### 3.1.1. Anova: Single Factor

---

**Description:**

This tool performs a simple analysis of variance on data for two or more samples. The analysis provides a test of the hypothesis that each sample is drawn from the same underlying probability distribution against the alternative hypothesis that underlying probability distributions are not the same for all samples.

**Dialog Box:**



**Parameters:**

- Input Range – Enter the cell reference for the range of data you want to analyze. The reference must consist of two or more adjacent ranges of data arranged in columns or rows.
- Grouped By – To indicate whether the data in the input range is arranged in rows or in columns, click Rows or Columns.
- Labels in first row/Labels in first column – If the first row of your input range contains labels, select the Labels in First Row check box. If the labels are in the first column of your input range, select the Labels

in First Column check box. This check box is clear if your input range has no labels; application generates appropriate data labels for the output table.

- **Alpha** – Enter the level at which you want to evaluate critical values for the F statistic. The alpha level is a significance level related to the probability of having a type I error (rejecting a null hypothesis).
- **Output Range** – Enter the reference for the upper-left cell of the output table. Application automatically determines the size of the output area and displays a message if the output table will replace existing data or extend beyond the bounds of the worksheet.
- **New Worksheet Ply** – Click to insert a new worksheet in the current workbook and paste the results starting at cell A1 of the new worksheet. To name the new worksheet, type a name in the box.

### 3.1.3. Anova: Two factor With Replication

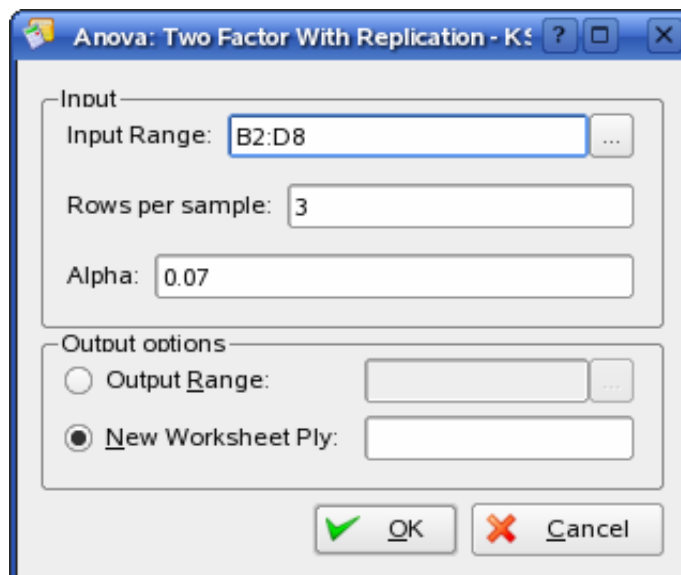
---

#### **Description:**

This analysis tool is useful when data can be classified along two different dimensions. For example, in an experiment to measure the height of plants, the plants may be given different brands of fertilizer (for example, A, B, C) and might also be kept at different temperatures (for example, low, high). For each of the 6 possible pairs of {fertilizer, temperature} we have an equal number of observations of plant height. Using this Anova tool we can test:

1. Whether heights of plants for the different fertilizer brands are drawn from the same underlying population; temperatures are ignored for this analysis.
2. Whether heights of plants for the different temperature levels are drawn from the same underlying population; fertilizer brands are ignored for this analysis.
3. Whether having accounted for the effects of differences between fertilizer brands found in step 1 and differences in temperatures found in step 2, the 6 samples representing all pairs of {fertilizer, temperature} values are drawn from the same population. The alternative hypothesis is that there are effects due to specific {fertilizer, temperature} pairs over and above differences based on fertilizer alone or on temperature alone.

#### **Dialog Box:**



#### **Parameters:**

- **Input Range** – Enter the cell reference for the range of data you want to analyze. The reference must consist of two or more adjacent ranges of data arranged in columns or rows.
- **Rows per sample** – Enter the number of rows contained in each sample. Each sample must contain the same number of rows, because each row represents a replication of the data.

- **Alpha** – Enter the level at which you want to evaluate critical values for the F statistic. The alpha level is a significance level related to the probability of having a type I error (rejecting a true hypothesis).
- **Output Range** – Enter the reference for the upper-left cell of the output table. Application automatically determines the size of the output area and displays a message if the output table will replace existing data or extend beyond the bounds of the worksheet.
- **New Worksheet Ply** – Click to insert a new worksheet in the current workbook and paste the results starting at cell A1 of the new worksheet. To name the new worksheet, type a name in the box.

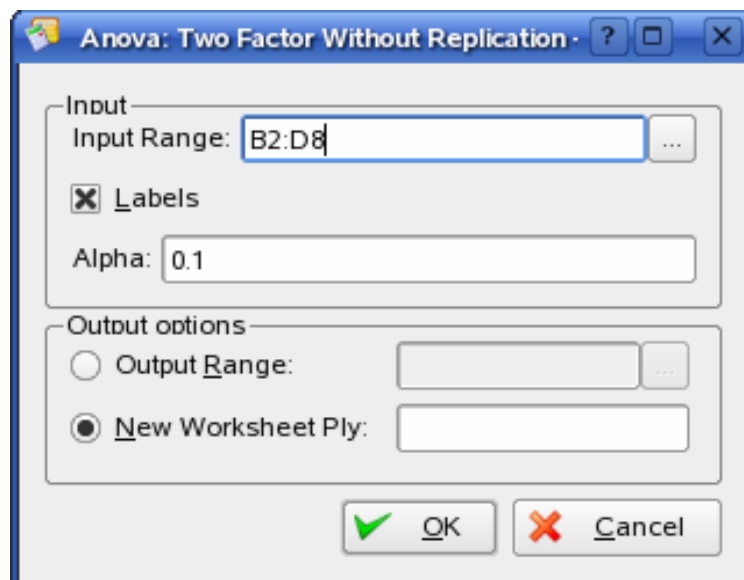
### 3.1.3. Anova: Two factor Without Replication

---

**Description:**

This analysis tool is useful when data are classified on two different dimensions as in the Two-Factor case With Replication. However, for this tool we assume that there is only a single observation for each pair (for example, each {fertilizer, temperature} pair in the example above. Using this tool we can apply the tests in steps 1 and 2 of the Anova: Two-Factor With Replication case but do not have enough data to apply the test in step 3.

**Dialog Box:**



**Parameters:**

- **Input Range** – Enter the cell reference for the range of data you want to analyze. The reference must consist of two or more adjacent ranges of data arranged in columns or rows.
- **Labels** – This check box is clear if your input range has no labels; application generates appropriate data labels for the output table.
- **Alpha** – Enter the level at which you want to evaluate critical values for the F statistic. The alpha level is a significance level related to the probability of having a type I error (rejecting a true hypothesis).
- **Output Range** – Enter the reference for the upper-left cell of the output table. Application automatically determines the size of the output area and displays a message if the output table will replace existing data or extend beyond the bounds of the worksheet.
- **New Worksheet Ply** – Click to insert a new worksheet in the current workbook and paste the results starting at cell A1 of the new worksheet. To name the new worksheet, type a name in the box.



### 3.3. Correlation

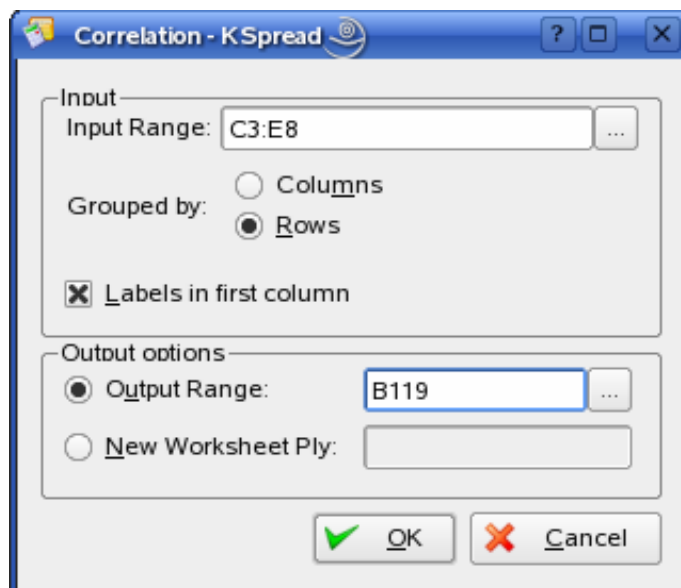
#### Description:

The Correlation analysis tool is useful when there is a need to calculate the correlation coefficient between measurement variables when measurements on each variable are observed for each of N subjects. (Any missing observation for any subject causes that subject to be ignored in the analysis.) It provides an output table, a correlation matrix, showing the value of correlation applied to each possible pair of measurement variables.

The correlation coefficient, like the covariance, is a measure of the extent to which two measurement variables “vary together.” Unlike the covariance, the correlation coefficient is scaled so that its value is independent of the units in which the two measurement variables are expressed. (For example, if the two measurement variables are weight and height, the value of the correlation coefficient is unchanged if weight is converted from pounds to kilograms.) The value of any correlation coefficient must be between -1 and +1 inclusive.

You can use the correlation analysis tool to examine each pair of measurement variables to determine whether the two measurement variables tend to move together – that is, whether large values of one variable tend to be associated with large values of the other (positive correlation), whether small values of one variable tend to be associated with large values of the other (negative correlation), or whether values of both variables tend to be unrelated (correlation near zero).

#### Dialog Box:



#### Parameters:

- **Input Range** – Enter the cell reference for the range of data you want to analyze. The reference must consist of two or more adjacent ranges of data arranged in columns or rows.
- **Grouped By** – To indicate whether the data in the input range is arranged in rows or in columns, click Rows or Columns.
- **Labels in first row/Labels in first column** – If the first row of your input range contains labels, select the Labels in First Row check box. If the labels are in the first column of your input range, select the Labels in First Column check box. This check box is clear if your input range has no labels; application generates appropriate data labels for the output table.
- **Output Range** – Enter the reference for the upper-left cell of the output table. Application populates only half of the table because correlation between two ranges of data is independent of the order in which the ranges are processed. Cells in the output table with matching row and column coordinates contain the value 1 because each data set correlates exactly with itself.

- New Worksheet Ply – Click to insert a new worksheet in the current workbook and paste the results starting at cell A1 of the new worksheet. To name the new worksheet, type a name in the box.

### 3.3. Covariance

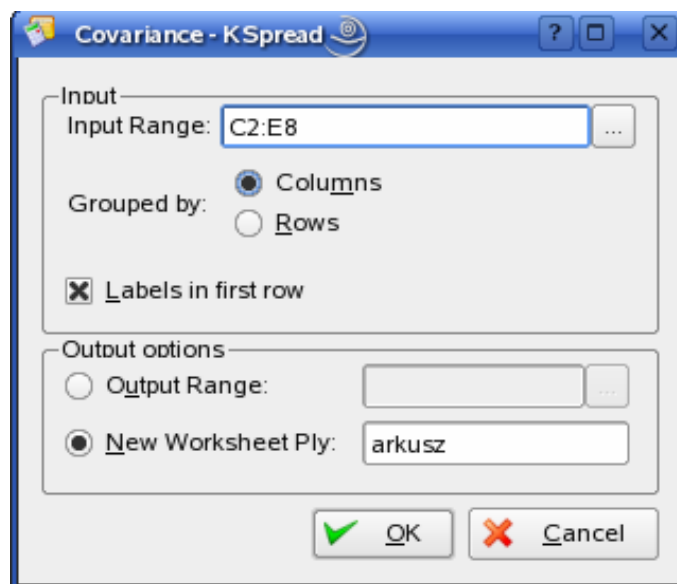
**Description:**

The Correlation and Covariance tools can both be used in the same setting, when you have N different measurement variables observed on a set of individuals. The Correlation and Covariance tools each give an output table, a matrix, showing the correlation coefficient or covariance, respectively, between each pair of measurement variables. The difference is that correlation coefficients are scaled to lie between -1 and +1 inclusive, Corresponding covariances are not scaled. Both the correlation coefficient and the covariance are measures of the extent to which two variables “vary together.”

The entry on the diagonal of the Covariance tool's output table in row i, column i is the covariance of the i-th measurement variable with itself; this is just the population variance for that variable.

You can use the covariance tool to examine each pair of measurement variables to determine whether the two measurement variables tend to move together – that is, whether large values of one variable tend to be associated with large values of the other (positive covariance), whether small values of one variable tend to be associated with large values of the other (negative covariance), or whether values of both variables tend to be unrelated (covariance near zero).

**Dialog Box:**



**Parameters:**

- Input Range – Enter the cell reference for the range of data you want to analyze. The reference must consist of two or more adjacent ranges of data arranged in columns or rows.
- Grouped By – To indicate whether the data in the input range is arranged in rows or columns, click Rows or Columns.
- Labels in first row/Labels in first column – If the first row of your input range contains labels, select the Labels in First Row check box. If the labels are in the first column of your input range, select the Labels in First Column check box. This check box is clear if your input range has no labels; application generates appropriate data labels for the output table.
- Output Range – Enter the reference for the upper-left cell of the output table. Application populates only half of the table because covariance between two ranges of data is independent of the order in which the ranges are processed. The table's diagonal contains the variance of each range.

- **New Worksheet Ply** – Click to insert a new worksheet in the current workbook and paste the results starting at cell A1 of the new worksheet. To name the new worksheet, type a name in the box.

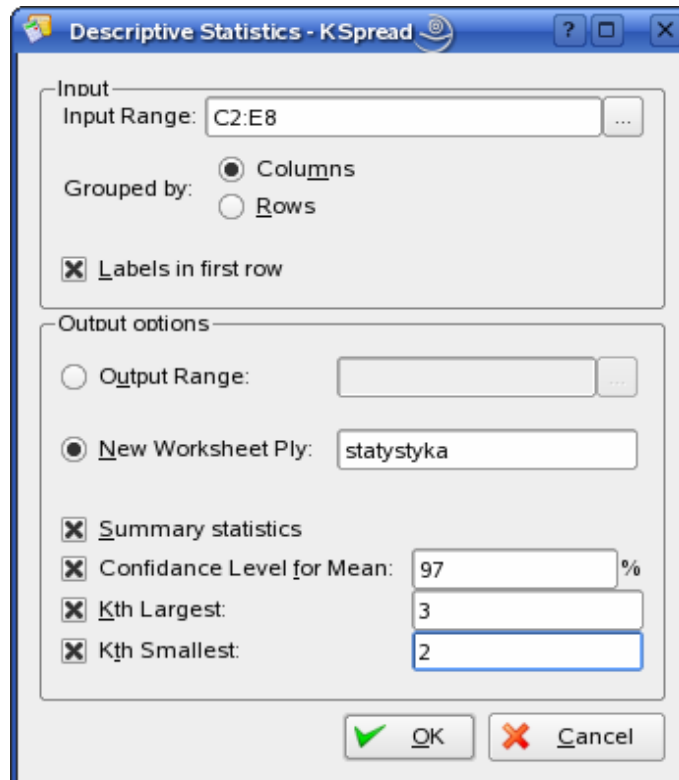
### 3.4. Descriptive Statistics

---

**Description:**

The Descriptive Statistics analysis tool generates a report of univariate statistics for data in the input range, providing information about the central tendency and variability of your data.

**Dialog Box:**



**Parameters:**

- **Input Range** – Enter the cell reference for the range of data you want to analyze. The reference must consist of two or more adjacent ranges of data arranged in columns or rows.
- **Grouped By** – To indicate whether the data in the input range is arranged in rows or in columns, click Rows or Columns.
- **Labels in first row/Labels in first column** – If the first row of your input range contains labels, select the Labels in First Row check box. If the labels are in the first column of your input range, select the Labels in First Column check box. This check box is clear if your input range has no labels; application generates appropriate data labels for the output table.
- **Confidence Level for Mean** – Select if you want to include a row in the output table for the confidence level of the mean. In the box, enter the confidence level you want to use. For example, a value of 95 percent calculates the confidence level of the mean at a significance of 5 percent.
- **Kth Largest** – Select if you want to include a row in the output table for the kth largest value for each range of data. In the box, enter the number to use for k. If you enter 1, this row contains the maximum of the data set.

- **Kth Smallest** – Select if you want to include a row in the output table for the kth smallest value for each range of data. In the box, enter the number to use for k. If you enter 1, this row contains the minimum of the data set.
- **Output Range** – Enter the reference for the upper-left cell of the output table. This tool produces two columns of information for each data set. The left column contains statistics labels, and the right column contains the statistics. Application writes a two-column table of statistics for each column or row in the input range, depending on the Grouped By option selected.
- **New Worksheet Ply** – Click to insert a new worksheet in the current workbook and paste the results starting at cell A1 of the new worksheet. To name the new worksheet, type a name in the box.
- **Summary statistics** – Select if you want Application to produce one field for each of the following statistics in the output table: Mean, Standard Error (of the mean), Median, Mode, Standard Deviation, Variance, Kurtosis, Skewness, Range, Minimum, Maximum, Sum, Count, Largest (#), Smallest (#), and Confidence Level.

### 3.5. Exponential Smoothing

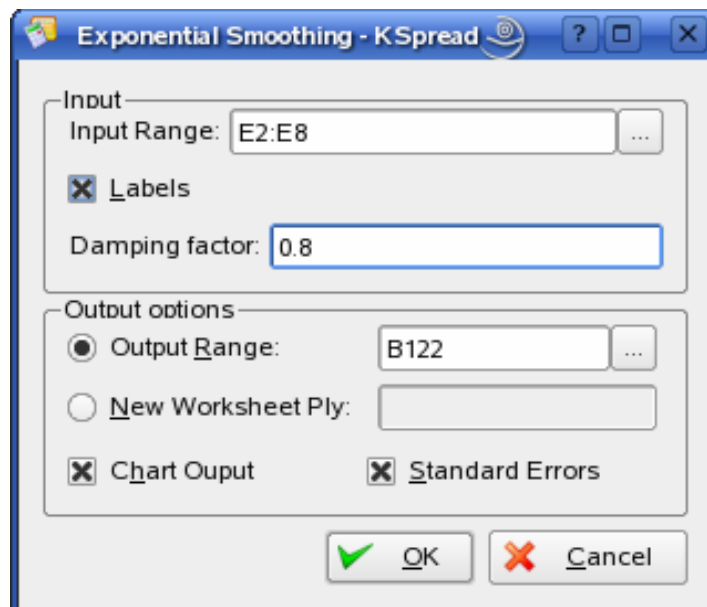
---

**Description:**

The Exponential Smoothing analysis tool predicts a value based on the forecast for the prior period, adjusted for the error in that prior forecast. The tool uses the smoothing constant  $a$ , the magnitude of which determines how strongly forecasts respond to errors in the prior forecast.

Note: Values of 0.2 to 0.3 are reasonable smoothing constants. These values indicate that the current forecast should be adjusted 20 to 30 percent for error in the prior forecast. Larger constants yield a faster response but can produce erratic projections. Smaller constants can result in long lags for forecast values.

**Dialog Box:**



**Parameters:**

- **Input Range** – Enter the cell reference for the range of data you want to analyze. The range must contain a single column or row with four or more cells of data.
- **Damping factor** – Enter the damping factor you want to use as the exponential smoothing constant. The damping factor is a corrective factor that minimizes the instability of data collected across a population. The default damping factor is 0.3.

Note: Values of 0.2 to 0.3 are reasonable smoothing constants. These values indicate that the current forecast should be adjusted 20 to 30 percent for error in the prior forecast. Larger constants yield a faster response but can produce erratic projections. Smaller constants can result in long lags for forecast values.

- **Labels** – Select if the first row and column of your input range contain labels. Clear this check box if your input range has no labels; application generates appropriate data labels for the output table.
- **Output Range** – Enter the reference for the upper-left cell of the output table. If you select the Standard Errors check box, Application generates a two-column output table with standard error values in the right column. If there are insufficient historical values to project a forecast or calculate a standard error, Application returns the #N/A error value.
- Note: The output range must be on the same worksheet as the data used in the input range. For this reason, the New Worksheet Ply and New Workbook options are unavailable.
- **Chart Output** – Select to generate an embedded chart for the actual and forecast values in the output table.
- **Standard Errors** – Select if you want to include a column that contains standard error values in the output table. Clear if you want a single-column output table without standard error values.

### 3.6. F-Test Two-Sample for Variances

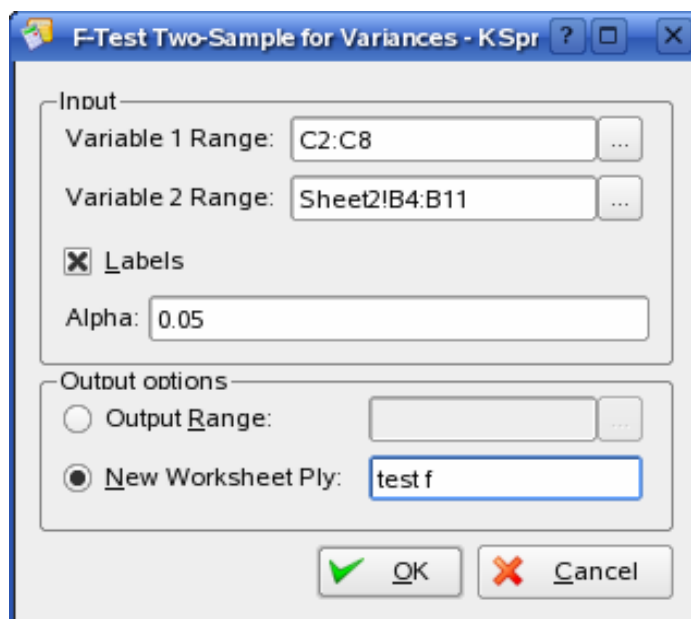
#### **Description:**

The F-Test Two-Sample for Variances analysis tool performs a two-sample F-test to compare two population variances.

For example, you can use the F-test tool on samples of times in a swim meet for each of two teams. The tool provides the result of a test of the null hypothesis that these two samples come from distributions with equal variances against the alternative that the variances are not equal in the underlying distributions.

The tool calculates the value  $f$  of an F-statistic (or F-ratio). A value of  $f$  close to 1 provides evidence that the underlying population variances are equal. In the output table, if  $f < 1$  “P(F  $\leq$  f) one-tail” gives the probability of observing a value of the F-statistic less than  $f$  when population variances are equal and “F Critical one-tail” gives the critical value less than 1 for the chosen significance level, Alpha. If  $f > 1$ , “P(F  $\leq$  f) one-tail” gives the probability of observing a value of the F-statistic greater than  $f$  when population variances are equal and “F Critical one-tail” gives the critical value greater than 1 for Alpha.

#### **Dialog Box:**



**Parameters:**

- Variable 1 Range – Enter the reference for the first column or row of data you want to analyze.
- Variable 2 Range – Enter the reference for the second column or row of data you want to analyze.
- Labels – Select if the first row or column of your input range contains labels. Clear this check box if your input range has no labels; application generates appropriate data labels for the output table.
- Alpha – Enter the confidence level for the test. This value must be in the range 0...1. The alpha level is a significance level related to the probability of having a type I error (rejecting a true hypothesis).
- Output Range – Enter the reference for the upper-left cell of the output table. Application automatically determines the size of the output area and displays a message if the output table will replace existing data.
- New Worksheet Ply – Click to insert a new worksheet in the current workbook and paste the results starting at cell A1 of the new worksheet. To name the new worksheet, type a name in the box.
- New Workbook – Click to create a new workbook and paste the results on a new worksheet in the new workbook.

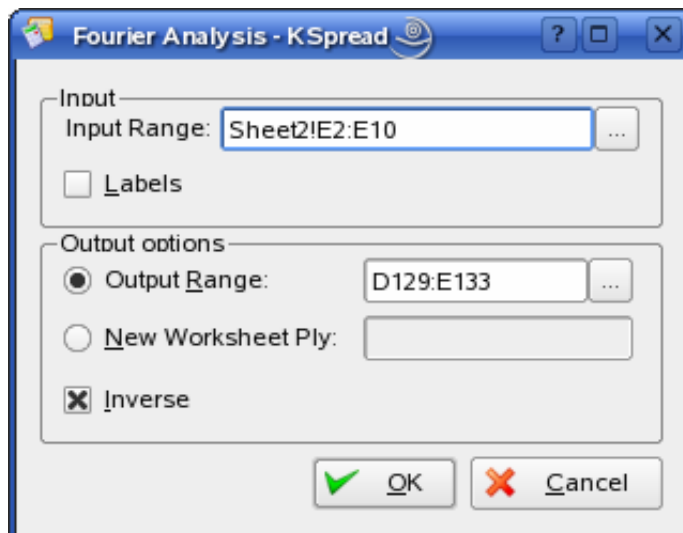
### 3.7. Fourier Analysis

---

**Description:**

The Fourier Analysis tool solves problems in linear systems and analyzes periodic data by using the Fast Fourier Transform (FFT) method to transform data. This tool also supports inverse transformations, in which the inverse of transformed data returns the original data.

**Dialog Box:**



**Parameters:**

- Input Range – Enter the reference for the range of real type data you want to transform. The number of input range values must be a power of two.
- Labels – Select if the first row of your input range contains labels. Clear if your input range has no labels; application generates appropriate data labels for the output table.
- Output Range – Enter the reference for the upper-left cell of the output table. Application automatically determines the size of the output area and displays a message if the output table will replace existing data.

- **New Worksheet Ply** – Click to insert a new worksheet in the current workbook and paste the results starting at cell A1 of the new worksheet. To name the new worksheet, type a name in the box.
- **Inverse** – If selected, the data in the input range is considered transformed and an inverse transformation is performed, returning the original inputs. If cleared, the data in the input range is transformed in the output table.

### 3.8. Histogram

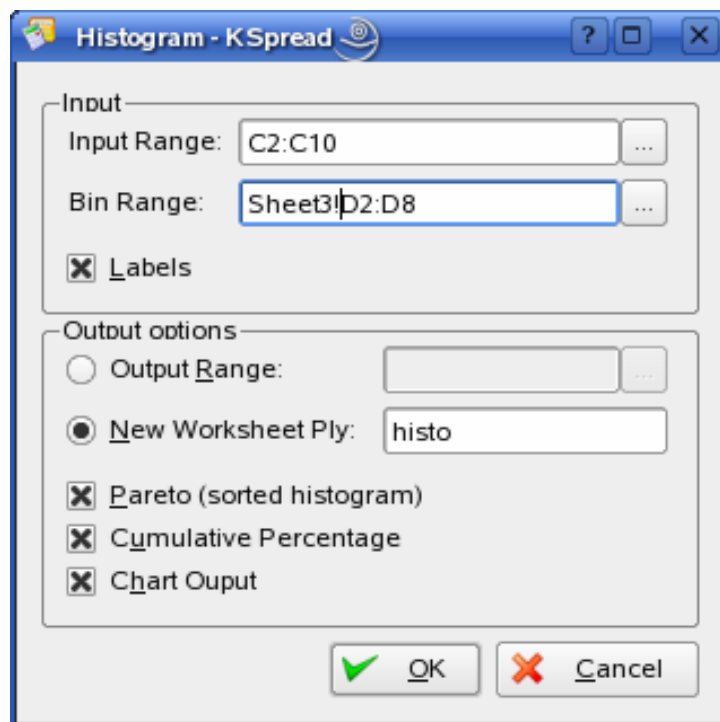
---

**Description:**

The Histogram analysis tool calculates individual and cumulative frequencies for a cell range of data and data bins. This tool generates data for the number of occurrences of a value in a data set.

For example, in a class of 20 students, you could determine the distribution of scores in letter-grade categories. A histogram table presents the letter-grade boundaries and the number of scores between the lowest bound and the current bound. The single most-frequent score is the mode of the data.

**Dialog Box:**



**Parameters:**

- **Input Range** – Enter the cell reference for the range of data you want to analyze.
- **Bin Range** (optional) – Enter the cell reference to a range that contains an optional set of boundary values that define bin ranges. These values should be in ascending order. Application counts the number of data points between the current bin number and the adjoining higher bin, if any. A number is counted in a particular bin if it is equal to or less than the bin number down to the last bin. All values below the first bin value are counted together, as are the values above the last bin value.

If you omit the bin range, application creates a set of evenly distributed bins between the data's minimum and maximum values.

- **Labels** – Select if the first row or column of your input range contains labels. Clear this check box if your input range has no labels; application generates appropriate data labels for the output table.

- **Output Range** – Enter the reference for the upper-left cell of the output table. Application automatically determines the size of the output area and displays a message if the output table will replace existing data.
- **New Worksheet Ply** – Click to insert a new worksheet in the current workbook and paste the results starting at cell A1 of the new worksheet. To name the new worksheet, type a name in the box.
- **Pareto (sorted histogram)** – Select to present data in the output table in descending order of frequency. If this check box is cleared, application presents the data in ascending order and omits the three rightmost columns that contain the sorted data.
- **Cumulative Percentage** – Select to generate an output table column for cumulative percentages and to include a cumulative percentage line in the histogram chart. Clear to omit the cumulative percentages.
- **Chart Output** – Select to generate an embedded histogram chart with the output table.

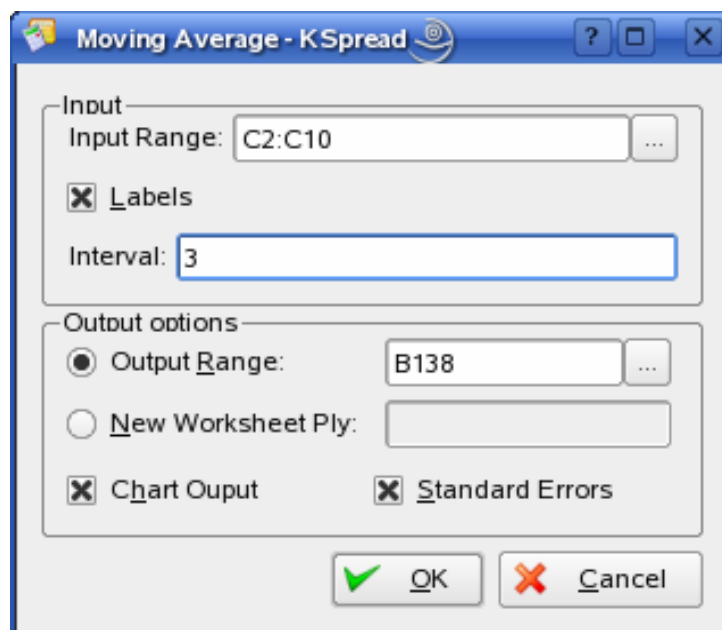
### 3.9. Moving Average

---

**Description:**

The Moving Average analysis tool projects values in the forecast period, based on the average value of the variable over a specific number of preceding periods. A moving average provides trend information that a simple average of all historical data would mask. Use this tool to forecast sales, inventory, or other trends.

**Dialog Box:**



**Parameters:**

- **Input Range** – Enter the cell reference for the range of data you want to analyze. The range must consist of a single column that contains four or more cells of data.
- **Labels** – Select if the first row of your input range contains labels. Clear this check box if your input range has no labels; application generates appropriate data labels for the output table.
- **Interval** – Enter the number of values you want to include in the moving average. The default interval is 3.
- **Output Range** – Enter the reference for the upper-left cell of the output table. If you select the Standard Errors check box, application generates a two-column output table with standard error



values in the right column. If there are insufficient historical values to project a forecast or calculate a standard error, application returns the #N/A error value.

- **Chart Output** – Select to generate an embedded histogram chart with the output table.
- **Standard Errors** – Select if you want to include a column that contains standard error values in the output table. Clear if you want a single-column output table without standard error values.

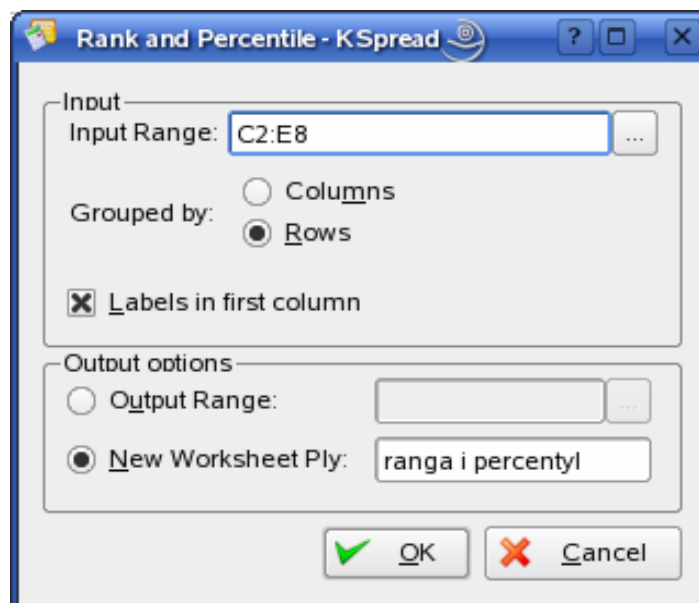
### 3.10. Rank and Percentile

---

**Description:**

The Rank and Percentile analysis tool produces a table that contains the ordinal and percentage rank of each value in a data set. You can analyze the relative standing of values in a data set.

**Dialog Box:**



**Parameters:**

- **Input Range** – Enter the cell reference for the range of worksheet data you want to analyze.
- **Grouped By** – To indicate whether the data in the input range is arranged in rows or columns, click Rows or Columns.
- **Labels in first row/Labels in first column** – If you click Columns and the first row of your input range contains labels, select the Labels in First Row check box. If you click Rows and the first column of your input range contains labels, select the Labels in First Column check box. Clear the check box if your input range has no labels; application generates appropriate row and column labels for the output table.
- **Output Range** – Enter the reference for the upper-left cell of the output table. Application generates one output table for each data set in the input range. Each output table contains four columns: the data point number, the data point value, the data point rank, and the data point percentage rank, sorted in ascending rank order.
- **New Worksheet Ply** – Click to insert a new worksheet in the current workbook and paste the results starting at cell A1 of the new worksheet. To name the new worksheet, type a name in the box.

### 3.11. Regression

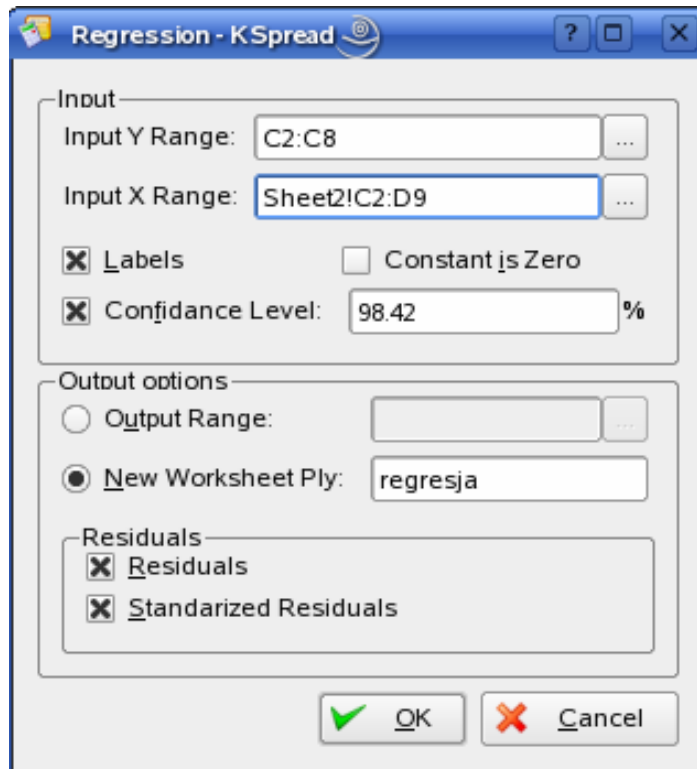
---

**Description:**

The Regression analysis tool performs linear regression analysis by using the "least squares" method to fit a line through a set of observations. You can analyze how a single dependent variable is affected by the values of one or more independent variables.

For example, you can analyze how an athlete's performance is affected by such factors as age, height, and weight. You can apportion shares in the performance measure to each of these three factors, based on a set of performance data, and then use the results to predict the performance of a new, untested athlete.

**Dialog Box:**



**Parameters:**

- Input Y Range – Enter the reference for the range of dependent data. The range must consist of a single column of data.
- Input X Range – Enter the reference for the range of independent data. Application orders independent variables from this range in ascending order from left to right.
- Labels – Select if the first row or column of your input range or ranges contains labels. Clear if your input has no labels; Application generates appropriate data labels for the output table.
- Confidence Level – Select to include an additional level in the summary output table. In the box, enter the confidence level you want applied in addition to the default 95 percent level.
- Constant is Zero – Select to force the regression line to pass through the origin.
- Output Range – Enter the reference for the upper-left cell of the output table. Allow at least seven columns for the summary output table, which includes an anova table, coefficients, standard error of y estimate, r2 values, number of observations, and standard error of coefficients.
- New Worksheet Ply – Click to insert a new worksheet in the current workbook and paste the results starting at cell A1 of the new worksheet. To name the new worksheet, type a name in the box.
- Residuals – Select to include residuals in the residuals output table.
- Standardized Residuals – Select to include standardized residuals in the residuals output table.

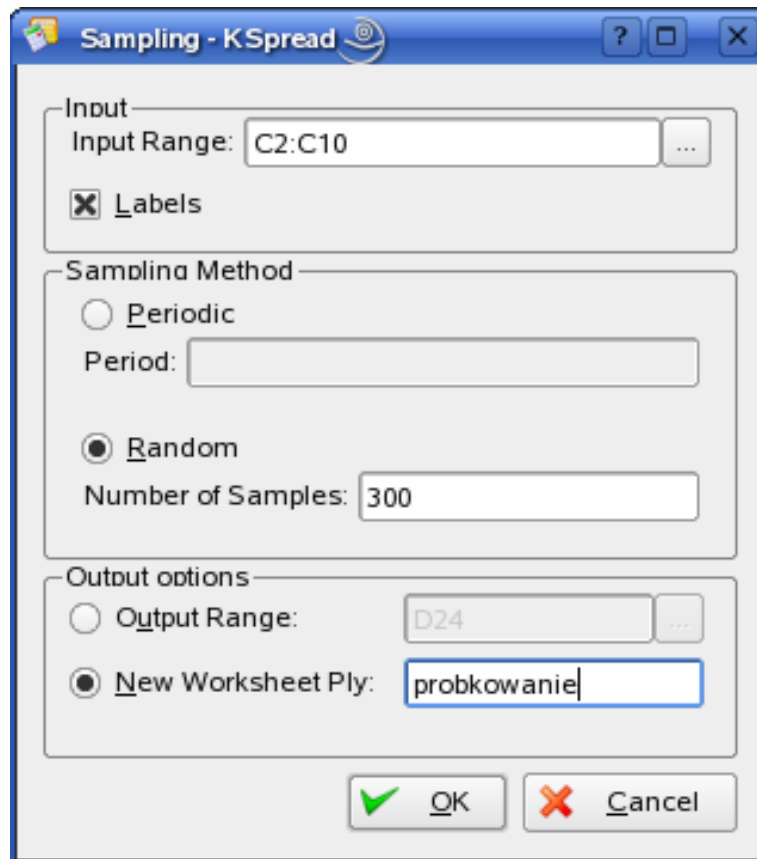
### 3.13. Sampling

**Description:**

The Sampling analysis tool creates a sample from a population by treating the input range as a population. When the population is too large to process or chart, you can use a representative sample. You can also create a sample that contains only values from a particular part of a cycle if you believe that the input data is periodic.

For example, if the input range contains quarterly sales figures, sampling with a periodic rate of four places values from the same quarter in the output range.

**Dialog Box:**



**Parameters:**

- Input Range – Enter the references for the range of data that contains the population of values you want to sample. Application draws samples from the first column, then the second column, and so on.
- Labels – Select if the first row or column of your input range contains labels. Clear if your input range has no labels; Application generates appropriate data labels for the output table.
- Sampling Method – Click Periodic or Random to indicate the sampling interval you want.
- Period – Enter the periodic interval at which you want sampling to take place. The period-th value in the input range and every period-th value thereafter is copied to the output column. Sampling stops when the end of the input range is reached.
- Number of Samples – Enter the number of random values you want in the output column. Each value is drawn from a random position in the input range, and any number can be selected more than once.

- Output Range – Enter the reference for the upper-left cell of the output table. Data is written in a single column below the cell. If you select Periodic, the number of values in the output table is equal to the number of values in the input range, divided by the sampling rate. If you select Random, the number of values in the output table is equal to the number of samples.
- New Worksheet Ply – Click to insert a new worksheet in the current workbook and paste the results starting at cell A1 of the new worksheet. To name the new worksheet, type a name in the box.

### 3.13. t-Test

---

The Two-Sample t-Test analysis tools test for equality of the population means underlying each sample. The three tools employ different assumptions: that the population variances are equal, that the population variances are not equal, and that the two samples represent before treatment and after treatment observations on the same subjects.

For all three tools below, a t-Statistic value,  $t$ , is computed and shown as “t Stat” in the output tables. Depending on the data, this value,  $t$ , can be negative or non-negative. Under the assumption of equal underlying population means, if  $t < 0$ , “P(T  $\leq$  t) one-tail” gives the probability that a value of the t-Statistic would be observed that is more negative than  $t$ . If  $t \geq 0$ , “P(T  $\leq$  t) one-tail” gives the probability that a value of the t-Statistic would be observed that is more positive than  $t$ . “t Critical one-tail” gives the cutoff value so that the probability of observing a value of the t-Statistic greater than or equal to “t Critical one-tail” is Alpha.

“P(T  $\leq$  t) two-tail” gives the probability that a value of the t-Statistic would be observed that is larger in absolute value than  $t$ . “P Critical two-tail” gives the cutoff value so that the probability of an observed t-Statistic larger in absolute value than “P Critical two-tail” is Alpha.

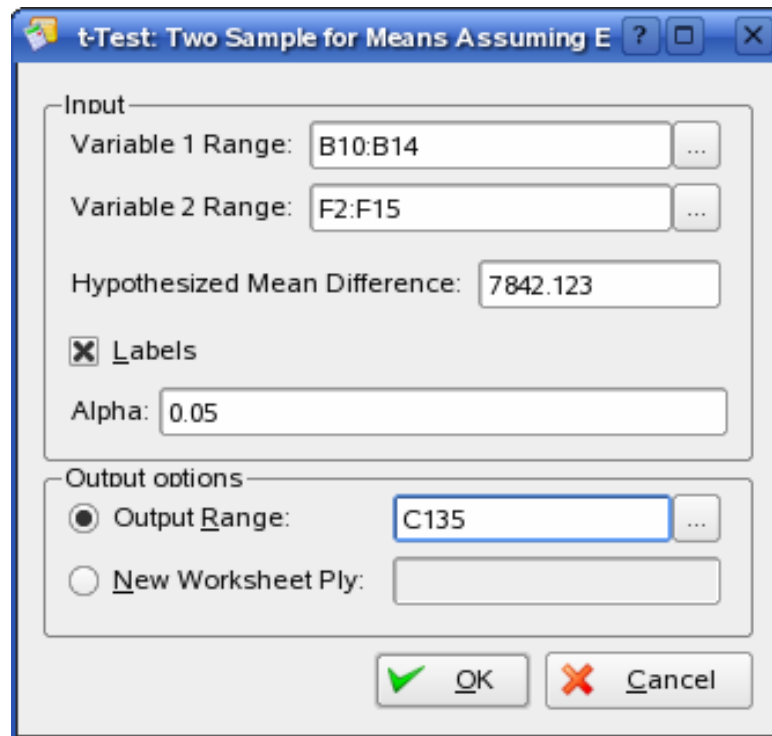
#### 3.13.1. t-Test: Two-Sample Assuming Equal Variances

---

**Description:**

This analysis tool performs a two-sample student's t-test. This t-test form assumes that the two data sets came from distributions with the same variances. You can use this t-test to determine whether the two samples are likely to have come from distributions with equal population means.

**Dialog Box:**



**Parameters:**

- **Variable 1 Range** – Enter the cell reference for the first range of data you want to analyze. The range must consist of a single column or row and contain the same number of data points as the second range.
- **Variable 2 Range** – Enter the cell reference for the second range of data you want to analyze. The range must consist of a single column or row and contain the same number of data points as the first range.
- **Hypothesized Mean Difference** – Enter the number you want for the shift in sample means. A value of zero indicates that the sample means are hypothesized to be equal.
- **Labels** – Select if the first row or column of your input ranges contains labels. Clear this check box if your input ranges have no labels; application generates appropriate data labels for the output table.
- **Alpha** – Enter the confidence level for the test. This value must be in the range 0...1. The alpha level is a significance level related to the probability of having a type I error (rejecting a true hypothesis).
- **Output Range** – Enter the reference for the upper-left cell of the output table. Application automatically determines the size of the output area and displays a message if the output table will replace existing data.
- **New Worksheet Ply** – Click to insert a new worksheet in the current workbook and paste the results starting at cell A1 of the new worksheet. To name the new worksheet, type a name in the box.

---

### 3.13.3. t-Test: Two-Sample Assuming Unequal Variances

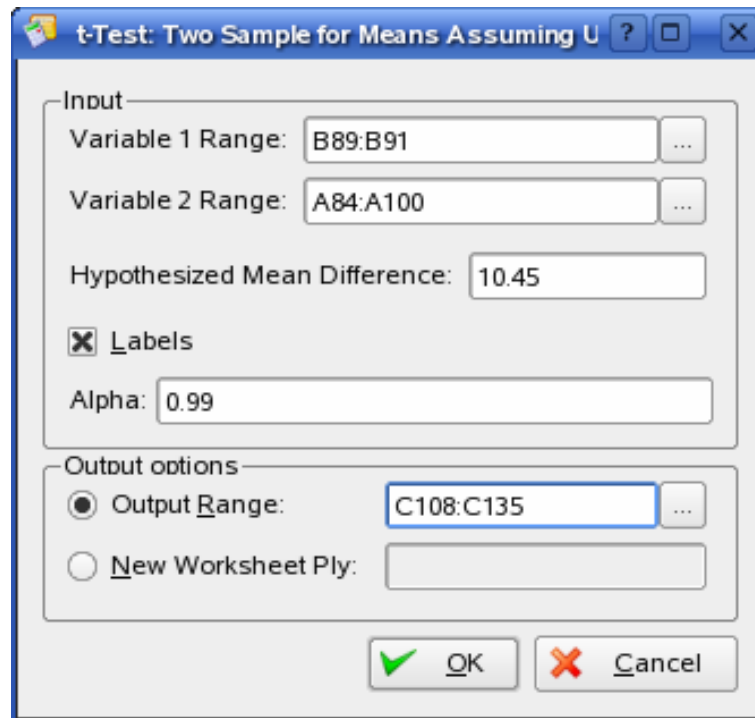
---

**Description:**

This analysis tool performs a two-sample student's t-test. This t-test form assumes that the two data sets came from distributions with unequal variances. As with the Equal Variances case above, you can use this t-test to determine whether the two samples are likely to have come from distributions with equal

population means. Use this test when there are distinct subjects in the two samples. Use the Paired test, described below, when there is a single set of subjects and the two samples represent measurements for each subject before and after a treatment.

**Dialog Box:**



**Parameters:**

- Variable 1 Range – Enter the cell reference for the first range of data you want to analyze. The range must consist of a single column or row of data.
- Variable 2 Range – Enter the cell reference for the second range of data you want to analyze. The range must consist of a single column or row of data.
- Hypothesized Mean Difference – Enter the number you want for the shift in sample means. A value of zero indicates that the sample means are hypothesized to be equal.
- Labels – Select if the first row or column of your input ranges contain labels. Clear this check box if your input ranges have no labels; application generates appropriate data labels for the output table.
- Alpha – Enter the confidence level for the test. This value must be in the range 0...1. The alpha level is a significance level related to the probability of having a type I error (rejecting a true hypothesis).
- Output Range – Enter the reference for the upper-left cell of the output table. Application automatically determines the size of the output area and displays a message if the output table will replace existing data.
- New Worksheet Ply – Click to insert a new worksheet in the current workbook and paste the results starting at cell A1 of the new worksheet. To name the new worksheet, type a name in the box.

---

### 3.13.3. t-Test: Paired Two Sample For Means

---

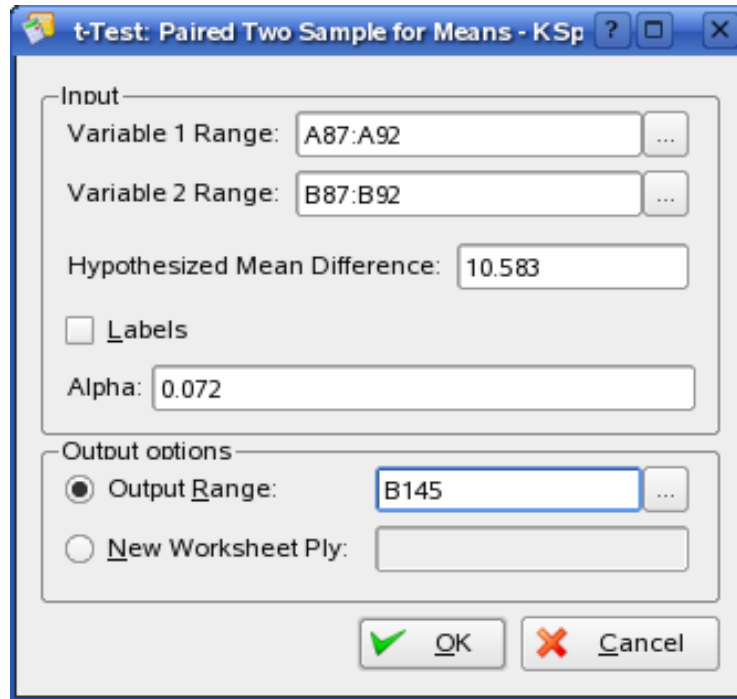
**Description:**

You can use a paired test when there is a natural pairing of observations in the samples, such as when a sample group is tested twice – before and after an experiment. This analysis tool and its formula perform a

paired two-sample student's t-test to determine whether observations taken before a treatment and observations taken after a treatment are likely to have come from distributions with equal population means. This t-test form does not assume that the variances of both populations are equal.

Note: Among the results generated by this tool is pooled variance, an accumulated measure of the spread of data about the mean.

**Dialog Box:**



**Parameters:**

- **Variable 1 Range** – Enter the cell reference for the first range of data you want to analyze. The range must consist of a single column or row and contain the same number of data points as the second range.
- **Variable 2 Range** – Enter the cell reference for the second range of data you want to analyze. The range must consist of a single column or row and contain the same number of data points as the first range.
- **Hypothesized Mean Difference** – Enter the number you want for the shift in sample means. A value of zero indicates that the sample means are hypothesized to be equal.
- **Labels** – Select if the first row or column of your input ranges contains labels. Clear this check box if your input ranges have no labels; application generates appropriate data labels for the output table.
- **Alpha** – Enter the confidence level for the test. This value must be in the range 0...1. The alpha level is a significance level related to the probability of having a type I error (rejecting a true hypothesis).
- **Output Range** – Enter the reference for the upper-left cell of the output table. Application automatically determines the size of the output area and displays a message if the output table will replace existing data.
- **New Worksheet Ply** – Click to insert a new worksheet in the current workbook and paste the results starting at cell A1 of the new worksheet. To name the new worksheet, type a name in the box.

### 3.14. z-Test: Two Sample for Means

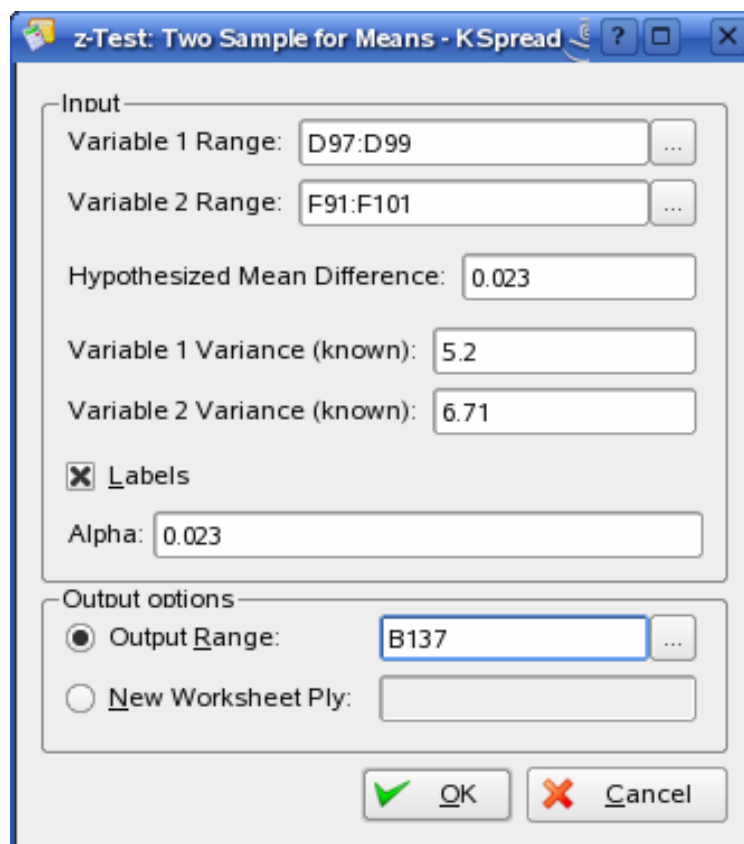
**Description:**

The z-Test: Two Sample for Means analysis tool performs a two-sample z-test for means with known variances. This tool is used to test the null hypothesis that there is no difference between two population means against either one-sided or two-sided alternative hypotheses .

When using the z-Test tool, one should be careful to understand the output. “P(Z <= z) one-tail” is really P(Z >= ABS(z)), the probability of a z-value further from 0 in the same direction as the observed z value when there is no difference between the population means. “P(Z <= z) two-tail” is really P(Z >= ABS(z) or Z <= -ABS(z)), the probability of a z-value further from 0 in either direction than the observed z-value when there is no difference between the population means. The two-tailed result is just the one-tailed result multiplied by 3. The z-Test tool can also be used for the case where the null hypothesis is that there is a specific non-zero value for the difference between the two population means.

For example, you can use this test to determine differences between the performances of two car models.

**Dialog Box:**



**Parameters:**

- Variable 1 Range – Enter the cell reference for the first range of data you want to analyze. The range must consist of a single column or row of data.
- Variable 2 Range – Enter the cell reference for the second range of data you want to analyze. The range must consist of a single column or row of data.
- Hypothesized Mean Difference – Enter the number you want for the shift in sample means. A value of zero indicates that the sample means are hypothesized to be equal.
- Variable 1 Variance (known) – Enter the known population variance for the Variable 1 input range.
- Variable 2 Variance (known) – Enter the known population variance for the Variable 2 input range.

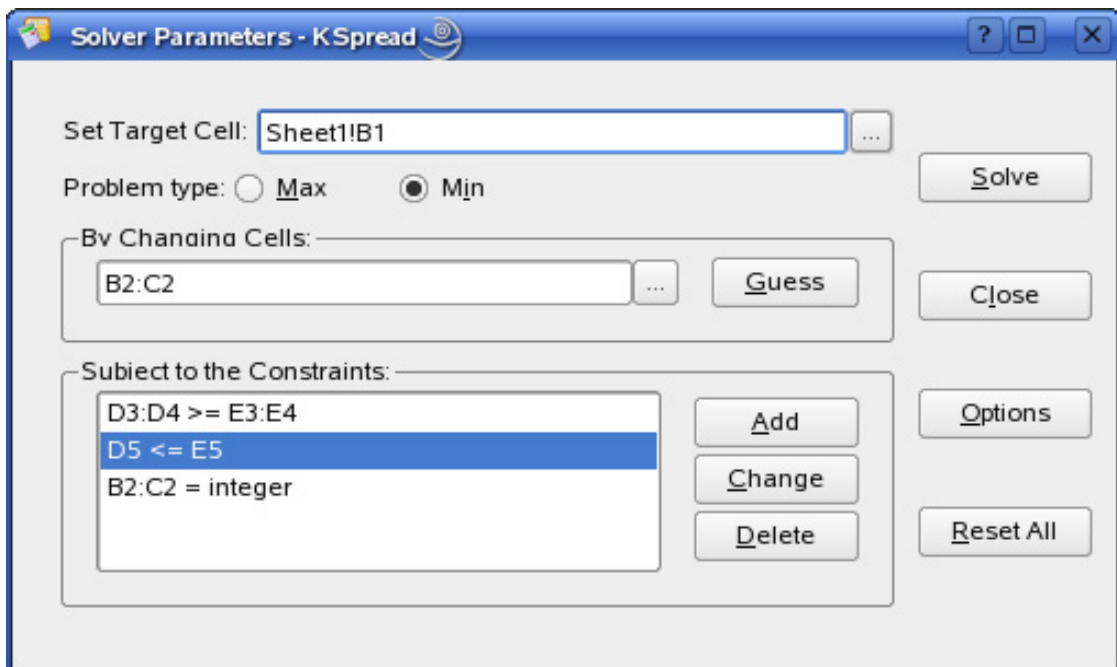


- Labels – Select if the first row or column of your input ranges contains labels. Clear this check box if your input ranges have no labels; Application generates appropriate data labels for the output table.
- Alpha – Enter the confidence level for the test. This value must be in the range 0...1. The alpha level is a significance level related to the probability of having a type I error (rejecting a true hypothesis).
- Output Range – Enter the reference for the upper-left cell of the output table. Application automatically determines the size of the output area and displays a message if the output table will replace existing data.
- New Worksheet Ply – Click to insert a new worksheet in the current workbook and paste the results starting at cell A1 of the new worksheet. To name the new worksheet, type a name in the box.

## Chapter 4: Solver Tool Description

### 4.1. Solver Parameters

**Dialog Box:**



**Options:**

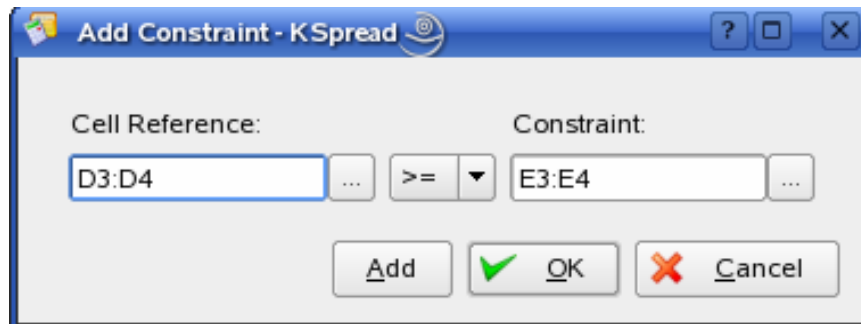
- Set Target Cell – Specifies the target cell that you want to set to a certain value or that you want to maximize or minimize. This cell must contain a formula.
- Problem Type – Specifies whether you want the target cell to be maximized or minimized.
- By Changing Cells – Specifies the cells that can be adjusted until the constraints in the problem are satisfied and the cell in the Set Target Cell box reaches its target. The adjustable cells must be related directly or indirectly to the target cell.
- Guess – Guesses all nonformula cells referred to by the formula in the Set Target Cell box, and places their references in the By Changing Cells box.
- Subject to the Constraints – Lists the current restrictions on the problem.
- Add – Displays the **Add Constraint** dialog box.
- Change – Displays the **Change Constraint** dialog box.
- Delete – Removes the selected constraint.

- Solve – Starts the solution process for the defined problem.
- Close – Closes the dialog box without solving the problem. Retains any changes you made by using the Options, Add, Change, or Delete buttons.
- Options – Displays the **Solver Options** dialog box, where you can control advanced features of the solution process.
- Reset All – Clears the current problem settings, and resets all settings to their original values.

## 4.2. Add Constraint

---

*Dialog Box:*



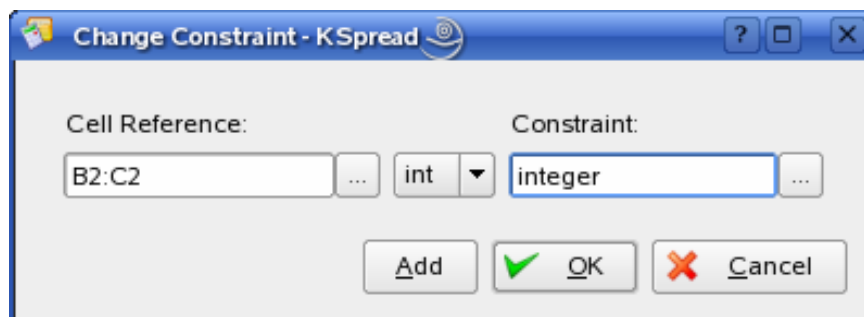
*Options:*

- Cell reference – Specifies the cell or range of cells where you want to constrain the values.
- Constraint – Specifies a restriction on the contents of the Cell reference box. Select the relationship you want to add or change ( < =, =, > = or Int ) between the referenced cell and the constraint. Then enter the constraint – a number, cell or range reference, or formula – in the box to the right.
- Add – Click to add another constraint without returning to the **Solver Parameters** dialog box.

## 4.3. Change Constraint

---

*Dialog Box:*



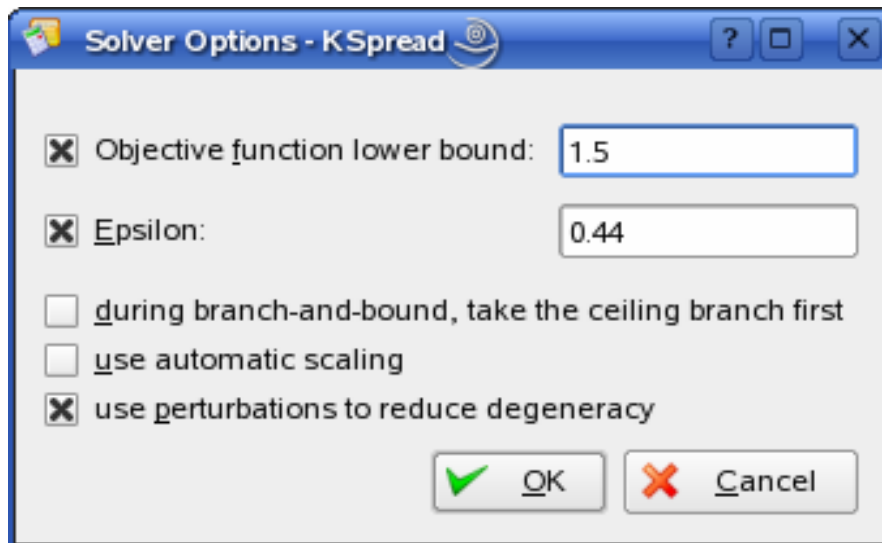
*Options:*

- Cell reference – Specifies the cell or range of cells where you want to constrain the values.
- Constraint – Specifies a restriction on the contents of the Cell reference box. Select the relationship you want to add or change ( < =, =, > = or Int ) between the referenced cell and the constraint. Then enter the constraint – a number, cell or range reference, or formula – in the box to the right.
- Add – Click to add another constraint without returning to the **Solver Parameters** dialog box.

## 4.4. Solver Options

---

**Dialog Box:**



**Options:**

Objective function lower bound – Specifies a lower bound for the objective function to the program. If close enough, may speed up the calculations.

Epsilon – Specifies the epsilon which is used to determine whether a floating point number is in fact an integer. Should be  $< 0.5$ .

during branch-and-bound, take the ceiling branch first – Specifies which branch to take first in branch-and-bound algorithm. This can influence solving times considerably. Depending on the model one rule can be best and for another model another rule.

use automatic problem scaling – Select to use automatic scaling when inputs and outputs have large differences in magnitude – for example, when maximizing the percentage of profit based on million-dollar investments..

use perturbations to reduce degeneracy – Specifies if special handling must be done to reduce degeneracy/cycling while solving. Selecting this can avoid cycling, but can also increase numerical instability.