

Statistical functions and tools in Microsoft Excel

Contents:

Listing of all statistical worksheet functions	1
Detailed description of all statistical functions	4
Statistical analysis tools *)	90
Listing of all worksheet functions by category	94

*) About statistical analysis tools:

Microsoft Excel provides a set of data analysis tools— called the *Analysis ToolPak*— 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 macro functions and then displays the results in an output table. Some tools generate charts in addition to output tables.

Related worksheet functions Excel provides many other statistical, financial, and engineering worksheet functions. Some of the statistical functions are built-in and others become available when you install the *Analysis ToolPak*.

Accessing the data analysis tools The *Analysis ToolPak* includes the tools described below. To access these tools, click *Data Analysis* on the *Tools* menu. If the *Data Analysis* command is not available, you need to load the *Analysis ToolPak* add-in program.

Trondheim, 28.10.2004

Arvid Aakre
arvid.aakre@ntnu.no

Statistical functions in Excel

[AVEDEV](#) Returns the average of the absolute deviations of data points from their mean

[AVERAGE](#) Returns the average of its arguments

[AVERAGEA](#) Returns the average of its arguments, including numbers, text, and logical values

[BETADIST](#) Returns the beta cumulative distribution function

[BETAINV](#) Returns the inverse of the cumulative distribution function for a specified beta distribution

[BINOMDIST](#) Returns the individual term binomial distribution probability

[CHIDIST](#) Returns the one-tailed probability of the chi-squared distribution

[CHIINV](#) Returns the inverse of the one-tailed probability of the chi-squared distribution

[CHITEST](#) Returns the test for independence

[CONFIDENCE](#) Returns the confidence interval for a population mean

[CORREL](#) Returns the correlation coefficient between two data sets

[COUNT](#) Counts how many numbers are in the list of arguments

[COUNTA](#) Counts how many values are in the list of arguments

[COUNTBLANK](#) Counts the number of blank cells within a range

[COUNTIF](#) Counts the number of nonblank cells within a range that meet the given criteria

[COVAR](#) Returns covariance, the average of the products of paired deviations

[CRITBINOM](#) Returns the smallest value for which the cumulative binomial distribution is less than or equal to a criterion value

[DEVSQ](#) Returns the sum of squares of deviations

[EXPONDIST](#) Returns the exponential distribution

[FDIST](#) Returns the F probability distribution

[FINV](#) Returns the inverse of the F probability distribution

[FISHER](#) Returns the Fisher transformation

[FISHERINV](#) Returns the inverse of the Fisher transformation

[FORECAST](#) Returns a value along a linear trend

[FREQUENCY](#) Returns a frequency distribution as a vertical array

[FTEST](#) Returns the result of an F-test

[GAMMADIST](#) Returns the gamma distribution

[GAMMAINV](#) Returns the inverse of the gamma cumulative distribution

[GAMMALN](#) Returns the natural logarithm of the gamma function, $\Gamma(x)$

[GEOMEAN](#) Returns the geometric mean

[GROWTH](#) Returns values along an exponential trend

[HARMEAN](#) Returns the harmonic mean

[HYPGEOMDIST](#) Returns the hypergeometric distribution

[INTERCEPT](#) Returns the intercept of the linear regression line

[KURT](#) Returns the kurtosis of a data set

[LARGE](#) Returns the k-th largest value in a data set

[LINEST](#) Returns the parameters of a linear trend

[LOGEST](#) Returns the parameters of an exponential trend

[LOGINV](#) Returns the inverse of the lognormal distribution

[LOGNORMDIST](#) Returns the cumulative lognormal distribution

[MAX](#) Returns the maximum value in a list of arguments

[MAXA](#) Returns the maximum value in a list of arguments, including numbers, text, and logical values

[MEDIAN](#) Returns the median of the given numbers

[MIN](#) Returns the minimum value in a list of arguments

[MINA](#) Returns the smallest value in a list of arguments, including numbers, text, and logical values

[MODE](#) Returns the most common value in a data set

[NEGBINOMDIST](#) Returns the negative binomial distribution

[NORMDIST](#) Returns the normal cumulative distribution

[NORMINV](#) Returns the inverse of the normal cumulative distribution

[NORMSDIST](#) Returns the standard normal cumulative distribution

[NORMSINV](#) Returns the inverse of the standard normal cumulative distribution

[PEARSON](#) Returns the Pearson product moment correlation coefficient

[PERCENTILE](#) Returns the k-th percentile of values in a range

[PERCENTRANK](#) Returns the percentage rank of a value in a data set

[PERMUT](#) Returns the number of permutations for a given number of objects

[POISSON](#) Returns the Poisson distribution

[PROB](#) Returns the probability that values in a range are between two limits

[QUARTILE](#) Returns the quartile of a data set

[RANK](#) Returns the rank of a number in a list of numbers

[RSQ](#) Returns the square of the Pearson product moment correlation coefficient

[SKEW](#) Returns the skewness of a distribution

[SLOPE](#) Returns the slope of the linear regression line

[SMALL](#) Returns the k-th smallest value in a data set

[STANDARDIZE](#) Returns a normalized value

[STDEV](#) Estimates standard deviation based on a sample

[STDEVA](#) Estimates standard deviation based on a sample, including numbers, text, and logical values

[STDEVPA](#) Estimates standard deviation based on the entire population

[STDEV](#) Estimates standard deviation based on a sample, including numbers, text, and logical values

[STDEVPA](#) Estimates standard deviation based on the entire population, including numbers, text, and logical values

[STEYX](#) Returns the standard error of the predicted y-value for each x in the regression

[TDIST](#) Returns the Student's t-distribution

[TINV](#) Returns the inverse of the Student's t-distribution

[TREND](#) Returns values along a linear trend

[TRIMMEAN](#) Returns the mean of the interior of a data set

[TTEST](#) Returns the probability associated with a Student's t-test

[VAR](#) Estimates variance based on a sample

[VARA](#) Estimates variance based on a sample, including numbers, text, and logical values

[VARP](#) Estimates variance based on the entire population

[VARPA](#) Estimates variance based on the entire population, including numbers, text, and logical values

[WEIBULL](#) Returns the Weibull distribution

[ZTEST](#) Returns the one-tailed probability-value of a z-test

AVEDEV

[See Also](#)

Returns the average of the absolute deviations of data points from their mean. AVEDEV is a measure of the variability in a data set.

Syntax

AVEDEV(number1,number2,...)

Number1, number2, ... are 1 to 30 arguments for which you want the average of the absolute deviations. You can also use a single array or a reference to an array instead of arguments separated by commas.

Remarks

- The arguments must either be numbers or be names, arrays, or references that contain numbers.
- If an array or reference argument contains text, logical values, or empty cells, those values are ignored; however, cells with the value zero are included.
- The equation for average deviation is:

$$\frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

AVEDEV is influenced by the unit of measurement in the input data.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	
1	Data
2	4
3	5
4	6
5	7
6	5
7	4
8	3

Formula	Description (Result)
<u>=AVEDEV(A2:A8)</u>	Average of the absolute deviations of the numbers above from their mean (1.020408)

AVERAGE

[See Also](#)

Returns the average (arithmetic mean) of the arguments.

Syntax

AVERAGE(number1,number2,...)

Number1, number2, ... are 1 to 30 numeric arguments for which you want the average.

Remarks

- The arguments must either be numbers or be names, arrays, or references that contain numbers.
- If an array or reference argument contains text, logical values, or empty cells, those values are ignored; however, cells with the value zero are included.

▼ Tip

When averaging cells, keep in mind the difference between empty cells and those containing the value zero, especially if you have cleared the **Zero values** check box on the **View** tab (**Options** command, **Tools** menu). Empty cells are not counted, but zero values are.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	
1	Data
2	10
3	7
4	9
5	27
6	2

Formula	Description (Result)
<u>=AVERAGE(A2:A6)</u>	Average of the numbers above (11)
<u>=AVERAGE(A2:A6, 5)</u>	Average of the numbers above and 5 (10)

AVERAGEA

[See Also](#)

Calculates the average (arithmetic mean) of the values in the list of arguments. In addition to numbers, text and logical values such as TRUE and FALSE are included in the calculation.

Syntax

AVERAGEA(value1,value2,...)

Value1, value2, ... are 1 to 30 cells, ranges of cells, or values for which you want the average.

Remarks

- The arguments must be numbers, names, arrays, or references.
- Array or reference arguments that contain text evaluate as 0 (zero). Empty text ("") evaluates as 0 (zero). If the calculation must not include text values in the average, use the AVERAGE function.
- Arguments that contain TRUE evaluate as 1; arguments that contain FALSE evaluate as 0 (zero).

Tip

When averaging cells, keep in mind the difference between empty cells and those containing the value zero, especially if you have cleared the **Zero values** check box on the **View** tab (**Options** command, **Tools** menu). Empty cells are not counted, but zero values are.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[How?](#)

A	
1 Data	
2	10
3	7
4	9
5	2
6	Not available
7	
Formula	Description (Result)
=AVERAGEA(A2:A6)	Average of the numbers above, and the text "Not Available". The cell with the text "Not available" is used in the calculation. (5.6)
=AVERAGEA(A2:A5,A7)	Average of the numbers above, and the empty cell. (7)

BETADIST

[See Also](#)

Returns the beta cumulative distribution function. The beta distribution is commonly used to study variation in the percentage of something across samples, such as the fraction of the day people spend watching television.

Syntax

BETADIST(x,alpha,beta,A,B)

X is the value between A and B at which to evaluate the function.

Alpha is a parameter of the distribution.

Beta is a parameter of the distribution.

A is an optional lower bound to the interval of x.

B is an optional upper bound to the interval of x.

Remarks

- If any argument is nonnumeric, BETADIST returns the #VALUE! error value.
- If $\alpha \leq 0$ or $\beta \leq 0$, BETADIST returns the #NUM! error value.
- If $x < A$, $x > B$, or $A = B$, BETADIST returns the #NUM! error value.
- If you omit values for A and B, BETADIST uses the standard cumulative beta distribution, so that $A = 0$ and $B = 1$.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[How?](#)

A		B	
1 Data		Description	
2	2	Value at which to evaluate the function	
3	8	Parameter of the distribution	
4	10	Parameter of the distribution	
5	1	Lower bound	
6	3	Upper bound	
Formula	Description (Result)		
=BETADIST(A2,A3,A4,A5,A6)	Cumulative beta probability density function, for the above parameters (0.685470581)		

BETAINV

[See Also](#)

Returns the inverse of the cumulative distribution function for a specified beta distribution. That is, if probability = BETADIST(x,...), then BETAINV(probability,...) = x. The beta distribution can be used in project planning to model probable completion times given an expected completion time and variability.

Syntax

BETAINV(probability,alpha,beta,A,B)

Probability is a probability associated with the beta distribution.

Alpha is a parameter of the distribution.

Beta is a parameter the distribution.

A is an optional lower bound to the interval of x.

B is an optional upper bound to the interval of x.

Remarks

- If any argument is nonnumeric, BETAINV returns the #VALUE! error value.
- If alpha ≤ 0 or beta ≤ 0, BETAINV returns the #NUM! error value.
- If probability ≤ 0 or probability > 1, BETAINV returns the #NUM! error value.
- If you omit values for A and B, BETAINV uses the standard cumulative beta distribution, so that A = 0 and B = 1.

Given a value for probability, BETAINV seeks that value x such that BETADIST(x, alpha, beta, A, B) = probability. Thus, precision of BETAINV depends on precision of BETADIST. BETAINV uses an iterative search technique. If the search has not converged after 100 iterations, the function returns the #N/A error value.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B
1 Data	Description
2 0.685470581	Probability associated with the beta distribution
3 8	Parameter of the distribution
4 10	Parameter of the distribution
5 1	Lower bound
6 3	Upper bound
Formula	Description (Result)
=BETAINV(A2,A3,A4,A5,A6)	Inverse of the cumulative beta probability density function for the parameters above (2)

BINOMDIST

[See Also](#)

Returns the individual term binomial distribution probability. Use BINOMDIST in problems with a fixed number of tests or trials, when the outcomes of any trial are only success or failure, when trials are independent, and when the probability of success is constant throughout the experiment. For example, BINOMDIST can calculate the probability that two of the next three babies born are male.

Syntax

BINOMDIST(number_s,trials,probability_s,cumulative)

Number_s is the number of successes in trials.

Trials is the number of independent trials.

Probability_s is the probability of success on each trial.

Cumulative is a logical value that determines the form of the function. If cumulative is TRUE, then BINOMDIST returns the cumulative distribution function, which is the probability that there are at most number_s successes; if FALSE, it returns the probability mass function, which is the probability that there are number_s successes.

Remarks

- Number_s and trials are truncated to integers.
- If number_s, trials, or probability_s is nonnumeric, BINOMDIST returns the #VALUE! error value.
- If number_s < 0 or number_s > trials, BINOMDIST returns the #NUM! error value.
- If probability_s < 0 or probability_s > 1, BINOMDIST returns the #NUM! error value.
- The binomial probability mass function is:

$$b(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}$$

where:

$$\binom{n}{x}$$

is COMBIN(n,x).

The cumulative binomial distribution is:

$$B(x; n, p) = \sum_{y=0}^x b(y; n, p)$$

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B
1 Data	Description
2 6	Number of successes in trials
3 10	Number of independent trials
4 0.5	Probability of success on each trial
Formula	Description (Result)
=BINOMDIST(A2,A3,A4,FALSE)	Probability of exactly 6 of 10 trials being successful (0.205078)

CHIDIST

[See Also](#)

Returns the one-tailed probability of the chi-squared distribution. The χ^2 distribution is associated with a χ^2 test. Use the χ^2 test to compare observed and expected values. For example, a genetic experiment might hypothesize that the next generation of plants will exhibit a certain set of colors. By comparing the observed results with the expected ones, you can decide whether your original hypothesis is valid.

Syntax

CHIDIST(x,degrees_freedom)

X is the value at which you want to evaluate the distribution.

Degrees_freedom is the number of degrees of freedom.

Remarks

- If either argument is nonnumeric, CHIDIST returns the #VALUE! error value.
- If x is negative, CHIDIST returns the #NUM! error value.
- If degrees_freedom is not an integer, it is truncated.
- If degrees_freedom < 1 or degrees_freedom $\geq 10^{10}$, CHIDIST returns the #NUM! error value.
- CHIDIST is calculated as $\text{CHIDIST} = P(X > x)$, where X is a χ^2 random variable.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B
1 Data	Description
2 18.307	Value at which you want to evaluate the distribution
3 10	Degrees of freedom
Formula	Description (Result)
=CHIDIST(A2,A3)	One-tailed probability of the chi-squared distribution, for the above terms (0.050001)

CHIINV

[See Also](#)

Returns the inverse of the one-tailed probability of the chi-squared distribution. If probability = CHIDIST(x,...), then CHIINV(probability,...) = x. Use this function to compare observed results with expected ones in order to decide whether your original hypothesis is valid.

Syntax

CHIINV(probability,degrees_freedom)

Probability is a probability associated with the chi-squared distribution.

Degrees_freedom is the number of degrees of freedom.

Remarks

- If either argument is nonnumeric, CHIINV returns the #VALUE! error value.
- If probability < 0 or probability > 1, CHIINV returns the #NUM! error value.
- If degrees_freedom is not an integer, it is truncated.
- If degrees_freedom < 1 or degrees_freedom $\geq 10^{10}$, CHIINV returns the #NUM! error value.

Given a value for probability, CHIINV seeks that value x such that CHIDIST(x, degrees_freedom) = probability. Thus, precision of CHIINV depends on precision of CHIDIST. CHIINV uses an iterative search technique. If the search has not converged after 100 iterations, the function returns the #N/A error value.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B
1 Data	Description
2 0.05	Probability associated with the chi-squared distribution
3 10	Degrees of freedom
Formula	Description (Result)
=CHIINV(A2,A3)	Inverse of the one-tailed probability of the chi-squared distribution (18.30703)

CHITEST

[See Also](#)

Returns the test for independence. CHITEST returns the value from the chi-squared (χ^2) distribution for the statistic and the appropriate degrees of freedom. You can use χ^2 tests to determine whether hypothesized results are verified by an experiment.

Syntax

CHITEST(actual_range,expected_range)

Actual_range is the range of data that contains observations to test against expected values.

Expected_range is the range of data that contains the ratio of the product of row totals and column totals to the grand total.

Remarks

- If actual_range and expected_range have a different number of data points, CHITEST returns the #N/A error value.
- The χ^2 test first calculates a χ^2 statistic using the formula:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(A_{ij} - E_{ij})^2}{E_{ij}}$$

where:

A_{ij} = actual frequency in the i-th row, j-th column

E_{ij} = expected frequency in the i-th row, j-th column

r = number of rows

c = number of columns

- A low value of χ^2 is an indicator of independence. As can be seen from the formula, χ^2 is always positive or 0, and is 0 only if $A_{ij} = E_{ij}$ for every i,j.
- CHITEST returns the probability that a value of the χ^2 statistic at least as high as the value calculated by the above formula could have happened by chance under the assumption of independence. In computing this probability, CHITEST uses the χ^2 distribution with an appropriate number of degrees of freedom, df. If $r > 1$ and $c > 1$, then $df = (r - 1)(c - 1)$. If $r = 1$ and $c > 1$, then $df = c - 1$ or if $r > 1$ and $c = 1$, then $df = r - 1$. $r = c = 1$ is not allowed and #N/A is returned.
- Use of CHITEST is most appropriate when E_{ij} 's are not too small. Some statisticians suggest that each E_{ij} should be greater than or equal to 5.

Example

The example may be easier to understand if you copy it to a blank worksheet.

► [How?](#)

A	B	C
1 Men (Actual)	Women (Actual)	Description
2 58	35	Agree
3 11	25	Neutral
4 10	23	Disagree
5 Men (Expected)	Women (Expected)	Description
6 45.35	47.65	Agree
7 17.56	18.44	Neutral
8 16.09	16.91	Disagree
Formula	Description (Result)	
=CHITEST(A2:B4,A6:B8)	The χ^2 statistic for the data above is 16.16957 with 2 degrees of freedom (0.000308)	

CONFIDENCE

[See Also](#)

Returns a value that you can use to construct a confidence interval for a population mean. The confidence interval is a range of values. Your sample mean, \bar{x} , is at the center of this range and the range is $\bar{x} \pm \text{CONFIDENCE}$. For example, if \bar{x} is the sample mean of delivery times for products ordered through the mail, $\bar{x} \pm \text{CONFIDENCE}$ is a range of population means. For any population mean, μ_0 , in this range, the probability of obtaining a sample mean further from μ_0 than \bar{x} is greater than alpha; for any population mean, μ_0 , not in this range, the probability of obtaining a sample mean further from μ_0 than \bar{x} is less than alpha. In other words, assume that we use \bar{x} , standard_dev, and size to construct a two-tailed test at significance level alpha of the hypothesis that the population mean is μ_0 . Then we will not reject that hypothesis if μ_0 is in the confidence interval and will reject that hypothesis if μ_0 is not in the confidence interval. The confidence interval does not allow us to infer that there is probability $1 - \alpha$ that our next package will take a delivery time that is in the confidence interval.

Syntax

CONFIDENCE(alpha,standard_dev,size)

Alpha is the significance level used to compute the confidence level. The confidence level equals $100 \times (1 - \alpha)\%$, or in other words, an alpha of 0.05 indicates a 95 percent confidence level.

Standard_dev is the population standard deviation for the data range and is assumed to be known.

Size is the sample size.

Remarks

- If any argument is nonnumeric, CONFIDENCE returns the #VALUE! error value.
- If $\alpha \leq 0$ or $\alpha \geq 1$, CONFIDENCE returns the #NUM! error value.
- If $\text{standard_dev} \leq 0$, CONFIDENCE returns the #NUM! error value.
- If size is not an integer, it is truncated.
- If $\text{size} < 1$, CONFIDENCE returns the #NUM! error value.
- If we assume alpha equals 0.05, we need to calculate the area under the standard normal curve that equals $(1 - \alpha)$, or 95 percent. This value is ± 1.96 . The confidence interval is therefore:

$$\bar{x} \pm 1.96 \left(\frac{\sigma}{\sqrt{n}} \right)$$

Example

Suppose we observe that, in our sample of 50 commuters, the average length of travel to work is 30 minutes with a population standard deviation of 2.5. With $\alpha = .05$, CONFIDENCE(.05, 2.5, 50) returns 0.69291. The corresponding confidence interval is then $30 \pm 0.69291 =$ approximately [29.3, 30.7]. For any population mean, μ_0 , in this interval, the probability of obtaining a sample mean further from μ_0 than 30 is more than 0.05. Likewise, for any population mean, μ_0 , outside this interval, the probability of obtaining a sample mean further from μ_0 than 30 is less than 0.05.

The example may be easier to understand if you copy it to a blank worksheet.

► [How?](#)

A	B
1 Data	Description
2 0.05	Significance level
3 2.5	Standard deviation of the population
4 50	Sample size
Formula	Description (Result)
=CONFIDENCE(A2,A3,A4)	Confidence interval for a population mean. In other words, the confidence interval for the underlying population mean for travel to work equals 30 ± 0.692951 minutes, or 29.3 to 30.7 minutes. (0.692951)

CORREL

[See Also](#)

Returns the correlation coefficient of the array1 and array2 cell ranges. Use the correlation coefficient to determine the relationship between two properties. For example, you can examine the relationship between a location's average temperature and the use of air conditioners.

Syntax

CORREL(array1,array2)

Array1 is a cell range of values.

Array2 is a second cell range of values.

Remarks

- If an array or reference argument contains text, logical values, or empty cells, those values are ignored; however, cells with the value zero are included.
- If array1 and array2 have a different number of data points, CORREL returns the #N/A error value.
- If either array1 or array2 is empty, or if s (the standard deviation) of their values equals zero, CORREL returns the #DIV/0! error value.
- The equation for the correlation coefficient is:

$$\text{Correl}(X, Y) = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

where x and y are the sample means AVERAGE(array1) and AVERAGE(array2).

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B
1 Data1	Data2
2 3	9
3 2	7
4 4	12
5 5	15
6 6	17
Formula	Description (Result)
=CORREL(A2:A6,B:B)	Correlation coefficient of the two data sets above (0.997054)

COUNT

[See Also](#)

Counts the number of cells that contain numbers and also numbers within the list of arguments. Use COUNT to get the number of entries in a number field that's in a range or array of numbers.

Syntax

COUNT(value1,value2,...)

Value1, value2, ... are 1 to 30 arguments that can contain or refer to a variety of different types of data, but only numbers are counted.

Remarks

- Arguments that are numbers, dates, or text representations of numbers are counted; arguments that are error values or text that cannot be translated into numbers are ignored.
- If an argument is an array or reference, only numbers in that array or reference are counted. Empty cells, logical values, text, or error values in the array or reference are ignored. If you need to count logical values, text, or error values, use the COUNTA function.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	
1 Data	
2 Sales	
3 12/8/2008	
4	
5 19	
6 22.24	
7 TRUE	
8 #DIV/0!	
Formula	Description (Result)
=COUNT(A2:A8)	Counts the number of cells that contain numbers in the list above (3)
=COUNT(A5:A8)	Counts the number of cells that contain numbers in the last 4 rows of the list (2)
=COUNT(A2:A8,2)	Counts the number of cells that contain numbers in the list, and the value 2 (4)

COUNTA

[See Also](#)

Counts the number of cells that are not empty and the values within the list of arguments. Use COUNTA to count the number of cells that contain data in a range or array.

Syntax

COUNTA(value1,value2,...)

Value1, value2, ... are 1 to 30 arguments representing the values you want to count. In this case, a value is any type of information, including empty text ("") but not including empty cells. If an argument is an array or reference, empty cells within the array or reference are ignored. If you do not need to count logical values, text, or error values, use the COUNT function.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	
1	Data
2	Sales
3	12/8/2008
4	
5	19
6	22.24
7	TRUE
8	#DIV/0!

Formula	Description (Result)
=COUNTA(A2:A8)	Counts the number of nonblank cells in the list above (6)
=COUNTA(A5:A8)	Counts the number of nonblank cells in the last 4 rows of the list (4)
=COUNTA(A1:A7,2)	Counts the number of nonblank cells in the list above and the value 2 (7)
=COUNTA(A1:A7,"Two")	Counts the number of nonblank cells in the list above and the value "Two" (7)

COUNTBLANK

[See Also](#)

Counts empty cells in a specified range of cells.

Syntax

COUNTBLANK(range)

Range is the range from which you want to count the blank cells.

Remark

Cells with formulas that return "(empty text)" are also counted. Cells with zero values are not counted.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B
1	Data
2	
3	6
4	
5	4

Formula	Description (Result)
=COUNTBLANK(A2:B)	Counts empty cells in the range above. The formula returns empty text. (4)

COUNTIF

[See Also](#)

Counts the number of cells within a range that meet the given criteria.

Syntax

COUNTIF(range,criteria)

Range is the range of cells from which you want to count cells.

Criteria is the criteria in the form of a number, expression, or text that defines which cells will be counted. For example, criteria can be expressed as 32, "32", ">32", "apples".

Remark

Microsoft Excel provides additional functions that can be used to analyze your data based on a condition. For example, to calculate a sum based on a string of text or a number within a range, use the SUMIF worksheet function. To have a formula return one of two values based on a condition, such as a sales bonus based on a specified sales amount, use the IF worksheet function.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B
1 Data	Data
2 apples	32
3 oranges	54
4 peaches	75
5 apples	8
Formula	Description (Result)
=COUNTIF(A2A5,"apples")	Number of cells with apples in the first column above (2)
=COUNTIF(BB,">55")	Number of cells with a value greater than 55 in the second column above (2)

COVAR

[See Also](#)

Returns covariance, the average of the products of deviations for each data point pair. Use covariance to determine the relationship between two data sets. For example, you can examine whether greater income accompanies greater levels of education.

Syntax

COVAR(array1,array2)

Array1 is the first cell range of integers.

Array2 is the second cell range of integers.

Remarks

- The arguments must either be numbers or be names, arrays, or references that contain numbers.
- If an array or reference argument contains text, logical values, or empty cells, those values are ignored; however, cells with the value zero are included.
- If array1 and array2 have different numbers of data points, COVAR returns the #N/A error value.
- If either array1 or array2 is empty, COVAR returns the #DIV/0! error value.
- The covariance is:

$$Cov(X, Y) = \frac{\sum (x - \bar{x})(y - \bar{y})}{n}$$

where x and y are the sample means AVERAGE(array1) and AVERAGE(array2), and n is the sample size.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B
1 Data1	Data2
2 3	9
3 2	7
4 4	12
5 5	15
6 6	17
Formula	Description (Result)
=COVAR(A2A6, BB)	Covariance, the average of the products of deviations for each data point pair above (5.2)

CRITBINOM

[See Also](#)

Returns the smallest value for which the cumulative binomial distribution is greater than or equal to a criterion value. Use this function for quality assurance applications. For example, use CRITBINOM to determine the greatest number of defective parts that are allowed to come off an assembly line run without rejecting the entire lot.

Syntax

CRITBINOM(trials,probability_s,alpha)

Trials is the number of Bernoulli trials.

Probability_s is the probability of a success on each trial.

Alpha is the criterion value.

Remarks

- If any argument is nonnumeric, CRITBINOM returns the #VALUE! error value.
- If trials is not an integer, it is truncated.
- If trials < 0, CRITBINOM returns the #NUM! error value.
- If probability_s is < 0 or probability_s > 1, CRITBINOM returns the #NUM! error value.
- If alpha < 0 or alpha > 1, CRITBINOM returns the #NUM! error value.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B
1 Data	Description
2 6	Number of Bernoulli trials
3 0.5	Probability of a success on each trial
4 0.75	Criterion value
Formula	Description (Result)
=CRITBINOM(A2,A3,A4)	Smallest value for which the cumulative binomial distribution is greater than or equal to a criterion value (4)

DEVSQ

[See Also](#)

Returns the sum of squares of deviations of data points from their sample mean.

Syntax

DEVSQ(number1,number2,...)

Number1, number2, ... are 1 to 30 arguments for which you want to calculate the sum of squared deviations. You can also use a single array or a reference to an array instead of arguments separated by commas.

Remarks

- The arguments must be numbers, or names, arrays, or references that contain numbers.
- If an array or reference argument contains text, logical values, or empty cells, those values are ignored; however, cells with the value zero are included.
- The equation for the sum of squared deviations is:

$$\text{DEVSQ} = \sum (x - \bar{x})^2$$

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	
1 Data	
2 4	
3 5	
4 8	
5 7	
6 11	
7 4	
8 3	
Formula	Description (Result)
=DEVSQ(A2:A8)	Sum of squares of deviations of data above from their sample mean (48)

EXPONDIST

[See Also](#)

Returns the exponential distribution. Use **EXPONDIST** to model the time between events, such as how long an automated bank teller takes to deliver cash. For example, you can use **EXPONDIST** to determine the probability that the process takes at most 1 minute.

Syntax

EXPONDIST(x,lambda,cumulative)

X is the value of the function.

Lambda is the parameter value.

Cumulative is a logical value that indicates which form of the exponential function to provide. If cumulative is TRUE, **EXPONDIST** returns the cumulative distribution function;if FALSE, it returns the probability density function.

Remarks

- If x or lambda is nonnumeric, **EXPONDIST** returns the #VALUE! error value.
- If x < 0, **EXPONDIST** returns the #NUM! error value.
- If lambda ≤ 0, **EXPONDIST** returns the #NUM! error value.

• The equation for the probability density function is:

$$f(x; \lambda) = \lambda e^{-\lambda x}$$

• The equation for the cumulative distribution function is:

$$F(x; \lambda) = 1 - e^{-\lambda x}$$

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B
1 Data	Description
2 0.2	Value of the function
3 10	Parameter value
Formula	Description (Result)
=EXPONDIST(A2,A3,TRUE)	Cumulative exponential distribution function (0.84665)
=EXPONDIST(0.2,10,FALSE)	Probability exponential distribution function (1.353353)

FDIST

[See Also](#)

Returns the F probability distribution. You can use this function to determine whether two data sets have different degrees of diversity. For example, you can examine the test scores of men and women entering high school and determine if the variability in the females is different from that found in the males.

Syntax

FDIST(x,degrees_freedom1,degrees_freedom2)

X is the value at which to evaluate the function.

Degrees_freedom1 is the numerator degrees of freedom.

Degrees_freedom2 is the denominator degrees of freedom.

Remarks

- If any argument is nonnumeric, **FDIST** returns the #VALUE! error value.
- If x is negative, **FDIST** returns the #NUM! error value.
- If degrees_freedom1 or degrees_freedom2 is not an integer, it is truncated.
- If degrees_freedom1 < 1 or degrees_freedom1 ≥ 10¹⁰, **FDIST** returns the #NUM! error value.
- If degrees_freedom2 < 1 or degrees_freedom2 ≥ 10¹⁰, **FDIST** returns the #NUM! error value.
- **FDIST** is calculated as $FDIST=P(F>x)$, where F is a random variable that has an F distribution with degrees_freedom1 and degrees_freedom2 degrees of freedom.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B
1 Data	Description
2 15.20675	Value at which to evaluate the function
3 6	Numerator degrees of freedom
4 4	Denominator degrees of freedom
Formula	Description (Result)
=FDIST(A2,A3,A4)	F probability distribution for the terms above (0.01)

FINV

[See Also](#)

Returns the inverse of the F probability distribution. If $p = \text{FDIST}(x, \dots)$, then $\text{FINV}(p, \dots) = x$.

The F distribution can be used in an F-test that compares the degree of variability in two data sets. For example, you can analyze income distributions in the United States and Canada to determine whether the two countries have a similar degree of income diversity.

Syntax

FINV(probability,degrees_freedom1,degrees_freedom2)

Probability is a probability associated with the F cumulative distribution.

Degrees_freedom1 is the numerator degrees of freedom.

Degrees_freedom2 is the denominator degrees of freedom.

Remarks

- If any argument is nonnumeric, FINV returns the #VALUE! error value.
- If probability < 0 or probability > 1, FINV returns the #NUM! error value.
- If degrees_freedom1 or degrees_freedom2 is not an integer, it is truncated.
- If degrees_freedom1 < 1 or degrees_freedom1 $\geq 10^{10}$, FINV returns the #NUM! error value.
- If degrees_freedom2 < 1 or degrees_freedom2 $\geq 10^{10}$, FINV returns the #NUM! error value.

FINV can be used to return critical values from the F distribution. For example, the output of an ANOVA calculation often includes data for the F statistic, F probability, and F critical value at the 0.05 significance level. To return the critical value of F, use the significance level as the probability argument to FINV.

Given a value for probability, FINV seeks that value x such that $\text{FDIST}(x, \text{degrees_freedom1}, \text{degrees_freedom2}) = \text{probability}$. Thus, precision of FINV depends on precision of FDIST. FINV uses an iterative search technique. If the search has not converged after 100 iterations, the function returns the #N/A error value.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B
1 Data	Description
2 0.01	Probability associated with the F cumulative distribution
3 6	Numerator degrees of freedom
4 4	Denominator degrees of freedom
Formula	Description (Result)
<u>$=\text{FINV}(\text{A2},\text{A3},\text{A4})$ Inverse of the F probability distribution for the terms above (15.20675)</u>	

FISHER

[See Also](#)

Returns the Fisher transformation at x. This transformation produces a function that is normally distributed rather than skewed. Use this function to perform hypothesis testing on the correlation coefficient.

Syntax

FISHER(x)

X is a numeric value for which you want the transformation.

Remarks

- If x is nonnumeric, FISHER returns the #VALUE! error value.
- If $x \leq -1$ or if $x \geq 1$, FISHER returns the #NUM! error value.
- The equation for the Fisher transformation is:

$$z' = \frac{1}{2} \ln \left(\frac{1+x}{1-x} \right)$$

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B
1 Formula	Description (Result)
2 <u>$=\text{FISHER}(0.75)$</u>	Fisher transformation at 0.75 (0.972955)

FISHERINV

[See Also](#)

Returns the inverse of the Fisher transformation. Use this transformation when analyzing correlations between ranges or arrays of data. If $y = \text{FISHER}(x)$, then $\text{FISHERINV}(y) = x$.

Syntax

FISHERINV(y)

Y is the value for which you want to perform the inverse of the transformation.

Remarks

- If y is nonnumeric, FISHERINV returns the #VALUE!error value.
- The equation for the inverse of the Fisher transformation is:

$$x = \frac{e^{2y} - 1}{e^{2y} + 1}$$

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B
1 Formula	Description (Result)
2 =FISHERINV(0.972955)	Inverse of the Fisher transformation at 0.972955 (0.75)

FORECAST

[See Also](#)

Calculates, or predicts, a future value by using existing values. The predicted value is a y-value for a given x-value. The known values are existing x-values and y-values, and the new value is predicted by using linear regression. You can use this function to predict future sales, inventory requirements, or consumer trends.

Syntax

FORECAST(x,known_y's,known_x's)

X is the data point for which you want to predict a value.

Known_y's is the dependent array or range of data.

Known_x's is the independent array or range of data.

Remarks

- If x is nonnumeric, FORECAST returns the #VALUE!error value.
- If known_y's and known_x's are empty or contain a different number of data points, FORECAST returns the #N/A error value.
- If the variance of known_x's equals zero, then FORECAST returns the #DIV/0!error value.
- The equation for FORECAST is $a+bx$, where:

$$a = \bar{y} - b\bar{x}$$

and:

$$b = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}$$

and where \bar{x} and \bar{y} are the sample means $\text{AVERAGE}(\text{known_x's})$ and $\text{AVERAGE}(\text{known_y's})$.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B
1 Known Y	Known X
2 6	20
3 7	28
4 9	31
5 15	38
6 21	40
Formula	Description (Result)
=FORECAST(30,A2:A6,B2:B6)	Predicts a value for y given an x value of 30 (10.60725)

FREQUENCY

[See Also](#)

Calculates how often values occur within a range of values, and then returns a vertical array of numbers. For example, use FREQUENCY to count the number of test scores that fall within ranges of scores. Because FREQUENCY returns an array, it must be entered as an array formula.

Syntax

FREQUENCY(data_array,bins_array)

Data_array is an array of or reference to a set of values for which you want to count frequencies. If data_array contains no values, FREQUENCY returns an array of zeros.

Bins_array is an array of or reference to intervals into which you want to group the values in data_array. If bins_array contains no values, FREQUENCY returns the number of elements in data_array.

Remarks

- FREQUENCY is entered as an array formula after you select a range of adjacent cells into which you want the returned distribution to appear.
- The number of elements in the returned array is one more than the number of elements in bins_array. The extra element in the returned array returns the count of any values above the highest interval. For example, when counting three ranges of values (intervals) that are entered into three cells, be sure to enter FREQUENCY into four cells for the results. The extra cell returns the number of values in data_array that are greater than the third interval value.
- FREQUENCY ignores blank cells and text.
- Formulas that return arrays must be entered as array formulas.

Example

This example assumes all test scores are integers.

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B
1 Scores	Bins
2 79	70
3 85	79
4 78	89
5 85	
6 50	
7 81	
8 95	
9 88	
10 97	
Formula	Description (Result)
=FREQUENCY(A2:A10,B2:B5)	Number of scores less than or equal to 70 (1)
	Number of scores in the bin 71-79 (2)
	Number of scores in the bin 80-89 (4)
	Number of scores greater than or equal to 90 (2)

Note The formula in the example must be entered as an array formula. After copying the example to a blank worksheet, select the range A13:A16 starting with the formula cell. Press F2, and then press CTRL+SHIFT+ENTER. If the formula is not entered as an array formula, the single result is 1.

FTEST

[See Also](#)

Returns the result of an F-test. An F-test returns the one-tailed probability that the variances in array1 and array2 are not significantly different. Use this function to determine whether two samples have different variances. For example, given test scores from public and private schools, you can test whether these schools have different levels of test score diversity.

Syntax

FTEST(array1,array2)

Array1 is the first array or range of data.

Array2 is the second array or range of data.

Remarks

- The arguments must be either numbers or names, arrays, or references that contain numbers.
- If an array or reference argument contains text, logical values, or empty cells, those values are ignored; however, cells with the value zero are included.
- If the number of data points in array1 or array2 is less than 2, or if the variance of array1 or array2 is zero, FTEST returns the #DIV/0! error value.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B
1 Data1	Data2
2 6	20
3 7	28
4 9	31
5 15	38
6 21	40
Formula	Description (Result)
=FTEST(A2:A6,B2:B6)	F-test for the data sets above (0.648318)

GAMMADIST

[See Also](#)

Returns the gamma distribution. You can use this function to study variables that may have a skewed distribution. The gamma distribution is commonly used in queuing analysis.

Syntax

GAMMADIST(x,alpha,beta,cumulative)

X is the value at which you want to evaluate the distribution.

Alpha is a parameter to the distribution.

Beta is a parameter to the distribution. If beta = 1, GAMMADIST returns the standard gamma distribution.

Cumulative is a logical value that determines the form of the function. If cumulative is TRUE, GAMMADIST returns the cumulative distribution function; if FALSE, it returns the probability density function.

Remarks

- If x, alpha, or beta is nonnumeric, GAMMADIST returns the #VALUE!error value.
- If x < 0, GAMMADIST returns the #NUM!error value.
- If alpha ≤ 0 or if beta ≤ 0, GAMMADIST returns the #NUM!error value.
- The equation for the gamma probability density function is:

$$f(x; \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}$$

The standard gamma probability density function is:

$$f(x; \alpha) = \frac{x^{\alpha-1} e^{-x}}{\Gamma(\alpha)}$$

- When alpha = 1, GAMMADIST returns the exponential distribution with:

$$\lambda = \frac{1}{\beta}$$

- For a positive integer n, when alpha = n/2, beta = 2, and cumulative = TRUE, GAMMADIST returns (1 - CHIDIST(x)) with n degrees of freedom.
- When alpha is a positive integer, GAMMADIST is also known as the Erlang distribution.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B
1 Data	Description
2 10	Value at which you want to evaluate the distribution
3 9	Alpha parameter to the distribution
4 2	Beta parameter to the distribution
Formula	Description (Result)
=GAMMADIST(A2,A3,A4,FALSE)	Probability gamma distribution with the terms above (0.032639)
=GAMMADIST(A2,A3,A4,TRUE)	Cumulative gamma distribution with the terms above (0.068094)

GAMMAINV

[See Also](#)

Returns the inverse of the gamma cumulative distribution. If p = GAMMADIST(x,...), then GAMMAINV(p,...) = x.

You can use this function to study a variable whose distribution may be skewed.

Syntax

GAMMAINV(probability,alpha,beta)

Probability is the probability associated with the gamma distribution.

Alpha is a parameter to the distribution.

Beta is a parameter to the distribution. If beta = 1, GAMMAINV returns the standard gamma distribution.

Remarks

- If any argument is nonnumeric, GAMMAINV returns the #VALUE!error value.
- If probability < 0 or probability ≥ 1, GAMMAINV returns the #NUM!error value.
- If alpha ≤ 0 or if beta ≤ 0, GAMMAINV returns the #NUM!error value.

Given a value for probability, GAMMAINV seeks that value x such that GAMMADIST(x, alpha, beta, TRUE) = probability. Thus, precision of GAMMAINV depends on precision of GAMMADIST. GAMMAINV uses an iterative search technique. If the search has not converged after 100 iterations, the function returns the #N/A error value.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B
1 Data	Description
2 0.068094	Probability associated with the gamma distribution
3 9	Alpha parameter to the distribution
4 2	Beta parameter to the distribution
Formula	Description (Result)
=GAMMAINV(A2,A3,A4)	Inverse of the gamma cumulative distribution for the above terms (10)

GAMMALN

[See Also](#)

Returns the natural logarithm of the gamma function, $\Gamma(x)$.

Syntax

GAMMALN(x)

X is the value for which you want to calculate GAMMALN.

Remarks

- If x is nonnumeric, GAMMALN returns the #VALUE! error value.
- If $x \leq 0$, GAMMALN returns the #NUM! error value.
- The number e raised to the GAMMALN(i) power, where i is an integer, returns the same result as $(i - 1)!$.
- GAMMALN is calculated as follows:

$$GAMMALN = LN(\Gamma(x))$$

where:

$$\Gamma(x) = \int_0^{\infty} e^{-u} u^{x-1} du$$

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B	
1	Formula	Description (Result)
2	<u>=GAMMALN(4)</u>	Natural logarithm of the gamma function at 4 (1.791759)

GEOMEAN

[See Also](#)

Returns the geometric mean of an array or range of positive data. For example, you can use GEOMEAN to calculate average growth rate given compound interest with variable rates.

Syntax

GEOMEAN(number1,number2,...)

Number1, number2, ... are 1 to 30 arguments for which you want to calculate the mean. You can also use a single array or a reference to an array instead of arguments separated by commas.

Remarks

- The arguments must be either numbers or names, arrays, or references that contain numbers.
- If an array or reference contains text, logical values, or empty cells, those values are ignored; however, cells with the value zero are included.
- If any data point ≤ 0 , GEOMEAN returns the #NUM!error value.
- The equation for the geometric mean is:

$$GM_{\bar{y}} = \sqrt[n]{y_1 y_2 y_3 \dots y_n}$$

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	
1	Data
2	4
3	5
4	8
5	7
6	11
7	4
8	3
Formula	Description (Result)
<u>=GEOMEAN(A2:A8)</u>	Geometric mean of the data set above (5.476987)

GROWTH

[See Also](#)

Calculates predicted exponential growth by using existing data. GROWTH returns the y-values for a series of new x-values that you specify by using existing x-values and y-values. You can also use the GROWTH worksheet function to fit an exponential curve to existing x-values and y-values.

Syntax

GROWTH(known_y's,known_x's,new_x's,const)

Known_y's is the set of y-values you already know in the relationship $y = b \cdot m^x$.

- If the array known_y's is in a single column, then each column of known_x's is interpreted as a separate variable.
- If the array known_y's is in a single row, then each row of known_x's is interpreted as a separate variable.
- If any of the numbers in known_y's is 0 or negative, GROWTH returns the #NUM!error value.

Known_x's is an optional set of x-values that you may already know in the relationship $y = b \cdot m^x$.

- The array known_x's can include one or more sets of variables. If only one variable is used, known_y's and known_x's can be ranges of any shape, as long as they have equal dimensions. If more than one variable is used, known_y's must be a vector (that is, a range with a height of one row or a width of one column).
- If known_x's is omitted, it is assumed to be the array {1,2,3,...} that is the same size as known_y's.

New_x's are new x-values for which you want GROWTH to return corresponding y-values.

- New_x's must include a column (or row) for each independent variable, just as known_x's does. So, if known_y's is in a single column, known_x's and new_x's must have the same number of columns. If known_y's is in a single row, known_x's and new_x's must have the same number of rows.
- If new_x's is omitted, it is assumed to be the same as known_x's.
- If both known_x's and new_x's are omitted, they are assumed to be the array {1,2,3,...} that is the same size as known_y's.

Const is a logical value specifying whether to force the constant b to equal 1.

- If const is TRUE or omitted, b is calculated normally.
- If const is FALSE, b is set equal to 1 and the m-values are adjusted so that $y = m^x$.

Remarks

- Formulas that return arrays must be entered as array formulas after selecting the correct number of cells.
- When entering an array constant for an argument such as known_x's, use commas to separate values in the same row and semicolons to separate rows.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[How?](#)

This example uses the same data as the LOGEST example. The first formula shows corresponding values to the known values. The second formula predicts the next months values, if the exponential trend continues.

A	B	C
1	Month Units	Formula (Corresponding Units)
2	11 33,100	=GROWTH(B2:B7,A2:A7)
3	12 47,300	
4	13 69,000	
5	14 102,000	
6	15 150,000	
7	16 220,000	
	Month Formula (Predicted Units)	
17	=GROWTH(B2:B7,A2:A7, A9:A10)	
18		

Note The formula in the example must be entered as an array formula. After copying the example to a blank worksheet, select the range C2:C7 or B9:B10 starting with the formula cell. Press F2, and then press CTRL+SHIFT+ENTER. If the formula is not entered as an array formula, the single results are 32618.20377 and 320196.7184.

HARMEAN

[See Also](#)

Returns the harmonic mean of a data set. The harmonic mean is the reciprocal of the arithmetic mean of reciprocals.

Syntax

HARMEAN(number1,number2,...)

Number1, number2, ... are 1 to 30 arguments for which you want to calculate the mean. You can also use a single array or a reference to an array instead of arguments separated by commas.

Remarks

- The arguments must be either numbers or names, arrays, or references that contain numbers.
- If an array or reference argument contains text, logical values, or empty cells, those values are ignored; however, cells with the value zero are included.
- If any data point ≤ 0 , HARMEAN returns the #NUM!error value.
- The harmonic mean is always less than the geometric mean, which is always less than the arithmetic mean.
- The equation for the harmonic mean is:

$$\frac{1}{H_y} = \frac{1}{n} \sum \frac{1}{Y_j}$$

Example

The example may be easier to understand if you copy it to a blank worksheet.

[How?](#)

A	
1	Data
2	4
3	5
4	8
5	7
6	11
7	4
8	3
	Formula Description (Result)
	=HARMEAN(A2:A8) Harmonic mean of the data set above (5.028376)

HYPGEOMDIST

[See Also](#)

Returns the hypergeometric distribution. HYPGEOMDIST returns the probability of a given number of sample successes, given the sample size, population successes, and population size. Use HYPGEOMDIST for problems with a finite population, where each observation is either a success or a failure, and where each subset of a given size is chosen with equal likelihood.

Syntax

HYPGEOMDIST(sample_s,number_sample,population_s,number_population)

Sample_s is the number of successes in the sample.

Number_sample is the size of the sample.

Population_s is the number of successes in the population.

Number_population is the population size.

Remarks

- All arguments are truncated to integers.
- If any argument is nonnumeric, HYPGEOMDIST returns the #VALUE!error value.
- If sample_s < 0 or sample_s is greater than the lesser of number_sample or population_s, HYPGEOMDIST returns the #NUM! error value.
- If sample_s is less than the larger of 0 or (number_sample - number_population + population_s), HYPGEOMDIST returns the #NUM!error value.
- If number_sample < 0 or number_sample > number_population, HYPGEOMDIST returns the #NUM!error value.
- If population_s < 0 or population_s > number_population, HYPGEOMDIST returns the #NUM!error value.
- If number_population < 0, HYPGEOMDIST returns the #NUM!error value.
- The equation for the hypergeometric distribution is:

$$P(X = x) = h(x; n, M, N) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$$

where:

x = sample_s

n = number_sample

M = population_s

N = number_population

HYPGEOMDIST is used in sampling without replacement from a finite population.

Example

A sampler of chocolates contains 20 pieces. Eight pieces are caramels, and the remaining 12 are nuts. If a person selects 4 pieces at random, the following function returns the probability that exactly 1 piece is a caramel.

The example may be easier to understand if you copy it to a blank worksheet.

► [How?](#)

A	B
1 Data	Description
2 1	Number of successes in the sample
3 4	Sample size
4 8	Number of successes in the population
5 20	Population size
Formula	Description (Result)
=HYPGEOMDIST(A2,A3,A4,A5)	Hypergeometric distribution for sample and population above (0.363261)

INTERCEPT

[See Also](#)

Calculates the point at which a line will intersect the y-axis by using existing x-values and y-values. The intercept point is based on a best-fit regression line plotted through the known x-values and known y-values. Use the INTERCEPT function when you want to determine the value of the dependent variable when the independent variable is 0 (zero). For example, you can use the INTERCEPT function to predict a metal's electrical resistance at 0°C when your data points were taken at room temperature and higher.

Syntax

INTERCEPT(known_y's,known_x's)

Known_y's is the dependent set of observations or data.

Known_x's is the independent set of observations or data.

Remarks

- The arguments should be either numbers or names, arrays, or references that contain numbers.
- If an array or reference argument contains text, logical values, or empty cells, those values are ignored; however, cells with the value zero are included.
- If known_y's and known_x's contain a different number of data points or contain no data points, INTERCEPT returns the #N/A error value.
- The equation for the intercept of the regression line, a, is:

$$a = \bar{y} - b\bar{x}$$

where the slope, b, is calculated as:

$$b = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}$$

and where x and y are the sample means AVERAGE(known_x's) and AVERAGE(known_y's).

Example

The example may be easier to understand if you copy it to a blank worksheet.

► [How?](#)

A	B
1 Known y	Known x
2 2	6
3 3	5
4 9	11
5 1	7
6 8	5
Formula	Description (Result)
=INTERCEPT(A2:A6, B2:B6)	Point at which a line will intersect the y-axis by using the x-values and y-values above (0.0483871)

KURT

[See Also](#)

Returns the kurtosis of a data set. Kurtosis characterizes the relative peakedness or flatness of a distribution compared with the normal distribution. Positive kurtosis indicates a relatively peaked distribution. Negative kurtosis indicates a relatively flat distribution.

Syntax

KURT(number1,number2,...)

Number1, number2, ... are 1 to 30 arguments for which you want to calculate kurtosis. You can also use a single array or a reference to an array instead of arguments separated by commas.

Remarks

- The arguments must be either numbers or names, arrays, or references that contain numbers.
- If an array or reference argument contains text, logical values, or empty cells, those values are ignored; however, cells with the value zero are included.
- If there are fewer than four data points, or if the standard deviation of the sample equals zero, KURT returns the #DIV/0! error value.
- Kurtosis is defined as:

$$\left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x_j - \bar{x}}{s} \right)^4 \right\} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

where:

s is the sample standard deviation.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

	A
1	Data
2	3
3	4
4	5
5	2
6	3
7	4
8	5
9	6
10	4
11	7

Formula	Description (Result)
<u><code>=KURT(A2:A11)</code></u>	<u>Kurtosis of the data set above (-0.1518)</u>

LARGE

[See Also](#)

Returns the k-th largest value in a data set. You can use this function to select a value based on its relative standing. For example, you can use LARGE to return the highest, runner-up, or third-place score.

Syntax

LARGE(array,k)

Array is the array or range of data for which you want to determine the k-th largest value.

K is the position (from the largest) in the array or cell range of data to return.

Remarks

- If array is empty, LARGE returns the #NUM! error value.
- If $k \leq 0$ or if k is greater than the number of data points, LARGE returns the #NUM! error value.

If n is the number of data points in a range, then LARGE(array,1) returns the largest value, and LARGE(array,n) returns the smallest value.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

	A	B
1	Data	Data
2	3	4
3	5	2
4	3	4
5	5	6
6	4	7

Formula	Description (Result)
<u><code>=LARGE(A2:B6,3)</code></u>	<u>3rd largest number in the numbers above (5)</u>
<u><code>=LARGE(A2:B6,7)</code></u>	<u>7th largest number in the numbers above (4)</u>

LINEST

[See Also](#)

Calculates the statistics for a line by using the "least squares" method to calculate a straight line that best fits your data, and returns an array that describes the line. Because this function returns an array of values, it must be entered as an array formula.

The equation for the line is:

$$y = mx + b \text{ or}$$

$$y = m_1x_1 + m_2x_2 + \dots + b \text{ (if there are multiple ranges of x-values)}$$

where the dependent y-value is a function of the independent x-values. The m-values are coefficients corresponding to each x-value, and b is a constant value. Note that y, x, and m can be vectors. The array that LINEST returns is {mn,mn-1,...,m1,b}. LINEST can also return additional regression statistics.

Syntax

LINEST(known_y's,known_x's,const,stats)

Known_y's is the set of y-values you already know in the relationship $y = mx + b$.

- If the array known_y's is in a single column, then each column of known_x's is interpreted as a separate variable.
- If the array known_y's is in a single row, then each row of known_x's is interpreted as a separate variable.

Known_x's is an optional set of x-values that you may already know in the relationship $y = mx + b$.

- The array known_x's can include one or more sets of variables. If only one variable is used, known_y's and known_x's can be ranges of any shape, as long as they have equal dimensions. If more than one variable is used, known_y's must be a vector (that is, a range with a height of one row or a width of one column).
- If known_x's is omitted, it is assumed to be the array {1,2,3,...} that is the same size as known_y's.

Const is a logical value specifying whether to force the constant b to equal 0.

- If const is TRUE or omitted, b is calculated normally.
- If const is FALSE, b is set equal to 0 and the m-values are adjusted to fit $y = mx$.

Stats is a logical value specifying whether to return additional regression statistics.

- If stats is TRUE, LINEST returns the additional regression statistics, so the returned array is {mn,mn-1,...,m1,b;sen,se1,...,se1,seb;r2,sey;F,df;ssreg,ssresid}.
- If stats is FALSE or omitted, LINEST returns only the m-coefficients and the constant b.

The additional regression statistics are as follows.

Statistic	Description
se1,se2,...,sen	The standard error values for the coefficients m1,m2,...,mn.
seb	The standard error value for the constant b (seb = #N/A when const is FALSE).
r2	The coefficient of determination. Compares estimated and actual y-values, and ranges in value from 0 to 1. If it is 1, there is a perfect correlation in the sample—there is no difference between the estimated y-value and the actual y-value. At the other extreme, if the coefficient of determination is 0, the regression equation is not helpful in predicting a y-value. For information about how r2 is calculated, see "Remarks" later in this topic.
sey	The standard error for the y estimate.
F	The F statistic, or the F-observed value. Use the F statistic to determine whether the observed relationship between the dependent and independent variables occurs by chance.
df	The degrees of freedom. Use the degrees of freedom to help you find F-critical values in a statistical table. Compare the values you find in the table to the F statistic returned by LINEST to determine a confidence level for the model. For information about how df is calculated, see "Remarks" later in this topic. Example 4 below shows use of F and df.
ssreg	The regression sum of squares.
ssresid	The residual sum of squares. For information about how ssreg and ssresid are calculated, see "Remarks" later in this topic.

The following illustration shows the order in which the additional regression statistics are returned.

	A	B	C	D	E	F
1	m _n	m _{n-1}	...	m ₂	m ₁	b
2	se _m	se _{m-1}	...	se ₂	se ₁	seb
3	r ²	sey				
4	F	df				
5	ssreg	ssresid				

Remarks

- You can describe any straight line with the slope and the y-intercept:

Slope (m):

To find the slope of a line, often written as m, take two points on the line, (x1,y1) and (x2,y2); the slope is equal to (y2 - y1)/(x2 - x1).

Y-intercept (b):

The y-intercept of a line, often written as b, is the value of y at the point where the line crosses the y-axis.

The equation of a straight line is $y = mx + b$. Once you know the values of m and b, you can calculate any point on the line by plugging the y- or x-value into that equation. You can also use the TREND function.

- When you have only one independent x-variable, you can obtain the slope and y-intercept values directly by using the following formulas:

Slope:

$$=INDEX(LINEST(known_y's,known_x's),1)$$

Y-intercept:

$$=INDEX(LINEST(known_y's,known_x's),2)$$

- The accuracy of the line calculated by LINEST depends on the degree of scatter in your data. The more linear the data, the more accurate the LINEST model. LINEST uses the method of least squares for determining the best fit for the data. When you have only one independent x-variable, the calculations for m and b are based on the following formulas:

$$m = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}$$

$$b = \bar{y} - m\bar{x}$$

where x and y are sample means, i.e., $x = AVERAGE(\text{known } x\text{'s})$ and $y = AVERAGE(\text{known } y\text{'s})$.

- The line- and curve-fitting functions LINEST and LOGEST can calculate the best straight line or exponential curve that fits your data. However, you have to decide which of the two results best fits your data. You can calculate TREND (known_y's,known_x's) for a straight line, or GROWTH(known_y's, known_x's) for an exponential curve. These functions, without the new_x's argument, return an array of y-values predicted along that line or curve at your actual data points. You can then compare the predicted values with the actual values. You may want to chart them both for a visual comparison.
- In regression analysis, Microsoft Excel calculates for each point the squared difference between the y-value estimated for that point and its actual y-value. The sum of these squared differences is called the residual sum of squares, ssresid. Microsoft Excel then calculates the total sum of squares, sstotal. When const = TRUE, or omitted, the total sum of squares is the sum of the squared differences between the actual y-values and the average of the y-values. When const = FALSE, the total sum of squares is the sum of the squares of the actual y-values (without subtracting the average y-value from each individual y-value). Then regression sum of squares, ssreg, can be found from: $ssreg = sstotal - ssresid$. The smaller the residual sum of squares is, compared with the total sum of squares, the larger the value of the coefficient of determination, r2, which is an indicator of how well the equation resulting from the regression analysis explains the relationship among the variables. $r2$ equals $ssreg/sstotal$.
- In some cases, one or more of the X columns (assume that Y's and X's are in columns) may have no additional predictive value in the presence of the other X columns. In other words, eliminating one or more X columns might lead to predicted Y values that are equally accurate. In that case these redundant X columns should be omitted from the regression model. This phenomenon is called "collinearity" because any redundant X column can be expressed as a sum of multiples of the non-redundant X columns. LINEST checks for collinearity and removes any redundant X columns from the regression model when it identifies them. Removed X columns can be recognized in LINEST output as having 0 coefficients as well as 0 se's. If one or more columns are removed as redundant, then df is affected because df depends on the number of X columns actually used for predictive purposes. For details on the computation of df, see Example 4 below. If df is changed because redundant X columns are removed, values of sey and F are also affected. Collinearity should be relatively rare in practice. However, one case where it is more likely to arise is when some X columns contain only 0's and 1's as indicators of whether a subject in an experiment is or is not a member of a particular group. If const = TRUE or omitted, LINEST effectively inserts an additional X column of all 1's to model the intercept. If you have a column with a 1 for each subject if male, or 0 if not, and you also have a column with a 1 for each subject if female, or 0 if not, this latter column is redundant because entries in it can be obtained from subtracting the entry in the "male indicator" column from the entry in the additional column of all 1's added by LINEST.
- df is calculated as follows when no X columns are removed from the model due to collinearity: if there are k columns of known_x's and const = TRUE or omitted, then $df = n - k - 1$. If const = FALSE, then $df = n - k$. In both cases, each X column removed due to collinearity increases df by 1.
- Formulas that return arrays must be entered as array formulas.
- When entering an array constant such as known_x's as an argument, use commas to separate values in the same row and semicolons to separate rows. Separator characters may be different depending on your locale setting in **Regional Settings** or **Regional Options** in **Control Panel**.
- Note that the y-values predicted by the regression equation may not be valid if they are outside the range of the y-values you used to determine the equation.

Example 1 Slope and Y-Intercept

The example may be easier to understand if you copy it to a blank worksheet.

▶ How?

A	B
1 Known y	Known x
2 1	0
3 9	4
4 5	2
5 7	3
Formula	Formula
=LINEST(A2:A5,B2:B5,,FALSE)	

Note The formula in the example must be entered as an array formula. After copying the example to a blank worksheet, select the range A7:B7 starting with the formula cell. Press F2, and then press CTRL+SHIFT+ENTER. If the formula is not entered as an array formula, the single result is 2.

When entered as an array, the slope (2) and the y-intercept (1) are returned.

Example 2 Simple Linear Regression

The example may be easier to understand if you copy it to a blank worksheet.

▶ How?

A	B
1 Month	Sales
2 1	3100
3 2	4500
4 3	4400
5 4	5400
6 5	7500
7 6	8100
Formula	Description (Result)
=SUM(LINEST(B2:B7, A2:A7)*{9,1}) Estimate sales for the ninth month (11000)	

In general, $SUM(\{m,b\}*x,1)$ equals $mx + b$, the estimated y-value for a given x-value. You can also use the TREND function.

Example 3 Multiple Linear Regression

Suppose a commercial developer is considering purchasing a group of small office buildings in an established business district.

The developer can use multiple linear regression analysis to estimate the value of an office building in a given area based on the following variables.

Variable Refers to the

y	Assessed value of the office building
x1	Floor space in square feet
x2	Number of offices
x3	Number of entrances
x4	Age of the office building in years

This example assumes that a straight-line relationship exists between each independent variable (x1, x2, x3, and x4) and the dependent variable (y), the value of office buildings in the area.

The developer randomly chooses a sample of 11 office buildings from a possible 1,500 office buildings and obtains the following data. "Half an entrance" means an entrance for deliveries only.

The example may be easier to understand if you copy it to a blank worksheet.

▶ How?

A	B	C	D	E
1 Floor space (x1)	Offices (x2)	Entrances (x3)	Age (x4)	Assessed value (y)

2	2310	2	2	20	142,000
3	2333	2	2	12	144,000
4	2356	3	1.5	33	151,000
5	2379	3	2	43	150,000
6	2402	2	3	53	139,000
7	2425	4	2	23	169,000
8	2448	2	1.5	99	126,000
9	2471	2	2	34	142,900
10	2494	3	3	23	163,000
11	2517	4	4	55	169,000
12	2540	2	3	22	149,000

Formula

=LINEST(E2:E12,A2:D12,TRUE,TRUE)

Note The formula in the example must be entered as an array formula. After copying the example to a blank worksheet, select the range A14:E18 starting with the formula cell. Press F2, and then press CTRL+SHIFT+ENTER. If the formula is not entered as an array formula, the single result is -234.2371645.

When entered as an array, the following regression statistics are returned. Use this key to identify the statistic you want.

	A	B	C	D	E	F
1	m_n	m_{n-1}	...	m_2	m_1	b
2	se_n	se_{n-1}	...	se_2	se_1	se_b
3	r^2	se_y				
4	F	df				
5	ss_{reg}	ss_{resid}				

The multiple regression equation, $y = m1*x1 + m2*x2 + m3*x3 + m4*x4 + b$, can now be obtained using the values from row 14:

$$y = 27.64*x1 + 12,530*x2 + 2,553*x3 - 234.24*x4 + 52,318$$

The developer can now estimate the assessed value of an office building in the same area that has 2,500 square feet, three offices, and two entrances and is 25 years old, by using the following equation:

$$y = 27.64*2500 + 12530*3 + 2553*2 - 234.24*25 + 52318 = \$158,261$$

Or you can copy the following table to cell A21 of the example workbook.

Floor space (x1)	Offices (x2)	Entrances (x3)	Age (x4)	Assessed value (y)
2500	3	2	25	=D14*A22 + C14*B22 + B14*C22 + A14*D22 + E14

You can also use the TREND function to calculate this value.

Example 4 Using the F and r2 Statistics

In the previous example, the coefficient of determination, or r^2 , is 0.99675 (see cell A17 in the output for LINEST), which would indicate a strong relationship between the independent variables and the sale price. You can use the F statistic to determine whether these results, with such a high r^2 value, occurred by chance.

Assume for the moment that in fact there is no relationship among the variables, but that you have drawn a rare sample of 11 office buildings that causes the statistical analysis to demonstrate a strong relationship. The term "Alpha" is used for the probability of erroneously concluding that there is a relationship.

F and df in LINEST output can be used to assess the likelihood of a higher F value occurring by chance. F can be compared with critical values in published F-distribution tables or Excel's FDIST can be used to calculate the probability of a larger F value occurring by chance. The appropriate F distribution has $v1$ and $v2$ degrees of freedom. If n is the number of data points and $const = TRUE$ or omitted, then $v1 = n - df - 1$ and $v2 = df$. (If $const = FALSE$, then $v1 = n - df$ and $v2 = df$.) Excel's FDIST(F,v1,v2) will return the probability of a higher F value occurring by chance. In Example 4, $df = 6$ (cell B18) and $F = 459.753674$ (cell A18).

Assuming an Alpha value of 0.05, $v1 = 11 - 6 - 1 = 4$ and $v2 = 6$, the critical level of F is 4.53. Since $F = 459.753674$ is much higher than 4.53, it is extremely unlikely that an F value this high occurred by chance. (With Alpha = 0.05, the hypothesis that there is no relationship between known_y's and known_x's is to be rejected when F exceeds the critical level, 4.53.) Using Excel's FDIST you can obtain the probability that an F value this high occurred by chance. FDIST(459.753674, 4, 6) = 1.37E-7, an extremely small probability. You can conclude, either by finding the critical level of F in a table or by using Excel's FDIST, that the regression equation is useful in predicting the assessed value of office buildings in this area. Remember that it is critical to use correct values of $v1$ and $v2$ computed in the previous paragraph.

Example 5 Calculating the t-Statistics

Another hypothesis test will determine whether each slope coefficient is useful in estimating the assessed value of an office

building in example 3. For example, to test the age coefficient for statistical significance, divide -234.24 (age slope coefficient) by 13.268 (the estimated standard error of age coefficients in cell A15). The following is the t-observed value:

$$t = m4 \div se4 = -234.24 \div 13.268 = -17.7$$

If the absolute value of t is sufficiently high, it can be concluded that the slope coefficient is useful in estimating the assessed value of an office building in Example 3. The table below shows the absolute values of the 4 t-observed values.

If you consult a table in a statistics manual, you will find that t-critical, two tailed, with 6 degrees of freedom and Alpha = 0.05 is 2.447. This critical value can also be found using Excel's TINV function. TINV(0.05,6) = 2.447. Because the absolute value of t, 17.7, is greater than 2.447, age is an important variable when estimating the assessed value of an office building. Each of the other independent variables can be tested for statistical significance in a similar manner. The following are the t-observed values for each of the independent variables.

Variable	t-observed value
Floor space	5.1
Number of offices	31.3
Number of entrances	4.8
Age	17.7

These values all have an absolute value greater than 2.447; therefore, all the variables used in the regression equation are useful in predicting the assessed value of office buildings in this area.<SPAN FPRev

LOGEST

[See Also](#)

In regression analysis, calculates an exponential curve that fits your data and returns an array of values that describes the curve. Because this function returns an array of values, it must be entered as an array formula.

The equation for this curve is:

$$y = b * m^x \text{ or}$$

$$y = (b * (m1^x1) * (m2^x2) * \dots) \text{ (if there are multiple x-values)}$$

where the dependent y-value is a function of the independent x-values. The m-values are bases corresponding to each exponent x-value, and b is a constant value. Note that y, x, and m can be vectors. The array that LOGEST returns is {mn,mn-1,...,m1,b}.

Syntax

LOGEST(known_y's,known_x's,const,stats)

Known_y's is the set of y-values you already know in the relationship $y = b * m^x$.

- If the array known_y's is in a single column, then each column of known_x's is interpreted as a separate variable.
- If the array known_y's is in a single row, then each row of known_x's is interpreted as a separate variable.

Known_x's is an optional set of x-values that you may already know in the relationship $y = b * m^x$.

- The array known_x's can include one or more sets of variables. If only one variable is used, known_y's and known_x's can be ranges of any shape, as long as they have equal dimensions. If more than one variable is used, known_y's must be a range of cells with a height of one row or a width of one column (which is also known as a vector).
- If known_x's is omitted, it is assumed to be the array {1,2,3,...} that is the same size as known_y's.

Const is a logical value specifying whether to force the constant b to equal 1.

- If const is TRUE or omitted, b is calculated normally.
- If const is FALSE, b is set equal to 1, and the m-values are fitted to $y = m^x$.

Stats is a logical value specifying whether to return additional regression statistics.

- If stats is TRUE, LOGEST returns the additional regression statistics, so the returned array is {mn,mn-1,...,m1,b;sen,sen-1,...,se1,seb;r 2,sey; F,df;ssreg,ssresid}.
- If stats is FALSE or omitted, LOGEST returns only the m-coefficients and the constant b.

For more information about additional regression statistics, see LINEST.

Remarks

- The more a plot of your data resembles an exponential curve, the better the calculated line will fit your data. Like LINEST, LOGEST returns an array of values that describes a relationship among the values, but LINEST fits a straight line to your data; LOGEST fits an exponential curve. For more information, see LINEST.
- When you have only one independent x-variable, you can obtain the slope (m) and y-intercept (b) values directly by using the following formulas:

Slope (m):
INDEX(LOGEST(known_y's,known_x's),1)

Y-intercept (b):
INDEX(LOGEST(known_y's,known_x's),2)

You can use the $y = b * m^x$ equation to predict future values of y, but Microsoft Excel provides the GROWTH function to do this for you. For more information, see GROWTH.

- Formulas that return arrays must be entered as array formulas.
- When entering an array constant such as known_x's as an argument, use commas to separate values in the same row and semicolons to separate rows. Separator characters may be different depending on your locale setting in **Regional Settings or Regional Options in Control Panel**.
- You should note that the y-values predicted by the regression equation may not be valid if they are outside the range of y-values you used to determine the equation.

Example 1 m-coefficients and the constant b

The example may be easier to understand if you copy it to a blank worksheet.

► [How?](#)

A	B
1 Month	Units

2	11	33,100
3	12	47,300
4	13	69,000
5	14	102,000
6	15	150,000
7	16	220,000

Formula **Formula**
 =LOGEST(B2:B7,A2:A7, TRUE, FALSE)

Note The formula in the example must be entered as an array formula. After copying the example to a blank worksheet, select the range A9:B9 starting with the formula cell. Press F2, and then press CTRL+SHIFT+ENTER. If the formula is not entered as an array formula, the single result is 1.463275628.

When entered as an array, the m-coefficients and the constant b are returned.

$y = b * m_1^{x-1}$ or using the values from the array:

$$y = 495.3 * 1.4633^x$$

You can estimate sales for future months by substituting the month number for x in this equation, or you can use the GROWTH function.

Example 2 Full statistics

The example may be easier to understand if you copy it to a blank worksheet.

► [How?](#)

A	B
1	Month
2	11
3	12
4	13
5	14
6	15
7	16

B	Units
2	33,100
3	47,300
4	69,000
5	102,000
6	150,000
7	220,000

Formula
 =LOGEST(B2:B7,A2:A7, TRUE, TRUE)

Note The formula in the example must be entered as an array formula. After copying the example to a blank worksheet, select the range A9:B13 starting with the formula cell. Press F2, and then press CTRL+SHIFT+ENTER. If the formula is not entered as an array formula, the single result is 1.463275628.

When entered as an array, the following regression statistics are returned. Use this key to identify the statistic you want.

	A	B	C	D	E	F
1	m _n	m _{n-1}	...	m ₂	m ₁	b
2	se _n	se _{n-1}	...	se ₂	se ₁	se _b
3	r ²	se _v				
4	F	d _f				
5	ss _{reg}	ss _{resid}				

You can use the additional regression statistics (cells A10:B13 in the above output array) to determine how useful the equation is for predicting future values.

Important The methods you use to test an equation using LOGEST are similar to the methods for LINEST. However, the additional statistics LOGEST returns are based on the following linear model:

$$\ln y = x_1 \ln m_1 + \dots + x_n \ln m_n + \ln b$$

You should keep this in mind when you evaluate the additional statistics, especially the se_i and se_b values, which should be compared to ln m_i and ln b, not to m_i and b. For more information, consult an advanced statistics manual.

LOGINV

[See Also](#)

Returns the inverse of the lognormal cumulative distribution function of x, where ln(x) is normally distributed with parameters mean and standarddev. If p = LOGNORMDIST(x,...) then LOGINV(p,...) = x.

Use the lognormal distribution to analyze logarithmically transformed data.

Syntax

LOGINV(probability,mean,standard_dev)

Probability is a probability associated with the lognormal distribution.

Mean is the mean of ln(x).

Standarddev is the standard deviation of ln(x).

Remarks

- If any argument is nonnumeric, LOGINV returns the #VALUE! error value.
- If probability < 0 or probability > 1, LOGINV returns the #NUM! error value.
- If standarddev <= 0, LOGINV returns the #NUM! error value.
- The inverse of the lognormal distribution function is:

$$\text{LOGINV}(p, \mu, \sigma) = e^{[\mu + \sigma \times \text{NORMSINV}(p)]}$$

Example

The example may be easier to understand if you copy it to a blank worksheet.

► [How?](#)

A	B
1	Data
2	0.0398
3	3.5
4	1.2
	Description
	Probability associated with the lognormal distribution
	Mean of ln(x)
	Standard deviation of ln(x)
	Formula
	=LOGINV(A2, A3, A4)
	Description (Result)
	Inverse of the lognormal cumulative distribution function for the terms above (1.00001)

LOGNORMDIST

[See Also](#)

Returns the cumulative lognormal distribution of x, where ln(x) is normally distributed with parameters mean and standarddev. Use this function to analyze data that has been logarithmically transformed.

Syntax

LOGNORMDIST(x,mean,standard_dev)

X is the value at which to evaluate the function.

Mean is the mean of ln(x).

Standarddev is the standard deviation of ln(x).

Remarks

- If any argument is nonnumeric, LOGNORMDIST returns the #VALUE! error value.
- If $x \leq 0$ or if $\text{standarddev} \leq 0$, LOGNORMDIST returns the #NUM! error value.
- The equation for the lognormal cumulative distribution function is:

$$\text{LOGNORMDIST}(x, \mu, \sigma) = \text{NORMSDIST}\left(\frac{\ln(x) - \mu}{\sigma}\right)$$

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B
1 Data	Description
2 4	Value at which to evaluate the function (x)
3 3.5	Mean of ln(x)
4 1.2	Standard deviation of ln(x)
Formula	Description (Result)
=LOGNORMDIST(A2,A3,A4)	Cumulative lognormal distribution at 4 with the terms above (0.0399)

MAX

[See Also](#)

Returns the largest value in a set of values.

Syntax

MAX(number1,number2,...)

Number1, number2, ... are 1 to 30 numbers for which you want to find the maximum value.

Remarks

- You can specify arguments that are numbers, empty cells, logical values, or text representations of numbers. Arguments that are error values or text that cannot be translated into numbers cause errors.
- If an argument is an array or reference, only numbers in that array or reference are used. Empty cells, logical values, or text in the array or reference are ignored. If logical values and text must not be ignored, use MAX instead.
- If the arguments contain no numbers, MAX returns 0 (zero).

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	
1 Data	
2 10	
3 7	
4 9	
5 27	
6 2	
Formula	Description (Result)
=MAX(A2:A6)	Largest of the numbers above (27)
=MAX(A2:A6, 30)	Largest of the numbers above and 30 (30)

MAXA

[See Also](#)

Returns the largest value in a list of arguments. Text and logical values such as TRUE and FALSE are compared as well as numbers.

MAX is similar to MINA. For more information, see the examples for MINA.

Syntax

MAXA(value1,value2,...)

Value1, value2, ... are 1 to 30 values for which you want to find the largest value.

Remarks

- You can specify arguments that are numbers, empty cells, logical values, or text representations of numbers. Arguments that are error values cause errors. If the calculation must not include text or logical values, use the MAX worksheet function instead.
- If an argument is an array or reference, only values in that array or reference are used. Empty cells and text values in the array or reference are ignored.
- Arguments that contain TRUE evaluate as 1; arguments that contain text or FALSE evaluate as 0 (zero).
- If the arguments contain no values, MAX returns 0 (zero).

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	
1	Data
2	0
3	0.2
4	0.5
5	0.4
6	TRUE
Formula	Description (Result)
=MAX(A2:A6)	Largest of the numbers above. TRUE evaluates to 1 (1)

MEDIAN

[See Also](#)

Returns the median of the given numbers. The median is the number in the middle of a set of numbers; that is, half the numbers have values that are greater than the median, and half have values that are less.

Syntax

MEDIAN(number1,number2,...)

Number1, number2, ... are 1 to 30 numbers for which you want the median.

Remarks

- The arguments should be either numbers or names, arrays, or references that contain numbers. Microsoft Excel examines all the numbers in each reference or array argument.
- If an array or reference argument contains text, logical values, or empty cells, those values are ignored; however, cells with the value zero are included.
- If there is an even number of numbers in the set, then MEDIAN calculates the average of the two numbers in the middle. See the second formula in the example.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	
1	Data
2	1
3	2
4	3
5	4
6	5
7	6
Formula	Description (Result)
=MEDIAN(A2:A6)	Median of the first 5 numbers in the list above (3)
=MEDIAN(A2:A7)	Median of all the numbers above, or the average of 3 and 4 (3.5)

MIN

[See Also](#)

Returns the smallest number in a set of values.

Syntax

MIN(number1,number2,...)

Number1, number2, ... are 1 to 30 numbers for which you want to find the minimum value.

Remarks

- You can specify arguments that are numbers, empty cells, logical values, or text representations of numbers. Arguments that are error values or text that cannot be translated into numbers cause errors.
- If an argument is an array or reference, only numbers in that array or reference are used. Empty cells, logical values, or text in the array or reference are ignored. If logical values and text should not be ignored, use MINA instead.
- If the arguments contain no numbers, MIN returns 0.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	
1	Data
2	10
3	7
4	9
5	27
6	2
	Formula Description (Result)
	=MIN(A2:A6) Smallest of the numbers above (2)
	=MIN(A2:A6,0) Smallest of the numbers above and 0 (0)

MINA

[See Also](#)

Returns the smallest value in the list of arguments. Text and logical values such as TRUE and FALSE are compared as well as numbers.

Syntax

MINA(value1,value2,...)

Value1, value2, ... are 1 to 30 values for which you want to find the smallest value.

Remarks

- You can specify arguments that are numbers, empty cells, logical values, or text representations of numbers. Arguments that are error values cause errors. If the calculation must not include text or logical values, use the MIN worksheet function instead.
- If an argument is an array or reference, only values in that array or reference are used. Empty cells and text values in the array or reference are ignored.
- Arguments that contain TRUE evaluate as 1; arguments that contain text or FALSE evaluate as 0 (zero).
- If the arguments contain no values, MINA returns 0.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	
1	Data
2	FALSE
3	0.2
4	0.5
5	0.4
6	0.8
	Formula Description (Result)
	=MINA(A2:A6) Smallest of the numbers above. FALSE evaluates to 0 (0)

MODE

[See Also](#)

Returns the most frequently occurring, or repetitive, value in an array or range of data. Like MEDIAN, MODE is a location measure.

Syntax

MODE(number1,number2,...)

Number1, number2, ... are 1 to 30 arguments for which you want to calculate the mode. You can also use a single array or a reference to an array instead of arguments separated by commas.

Remarks

- The arguments should be numbers, names, arrays, or references that contain numbers.
- If an array or reference argument contains text, logical values, or empty cells, those values are ignored; however, cells with the value zero are included.
- If the data set contains no duplicate data points, MODE returns the #N/A error value.

In a set of values, the mode is the most frequently occurring value; the median is the middle value; and the mean is the average value. No single measure of central tendency provides a complete picture of the data. Suppose data is clustered in three areas, half around a single low value, and half around two large values. Both AVERAGE and MEDIAN may return a value in the relatively empty middle, and MODE may return the dominant low value.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	
1	Data
2	5.6
3	4
4	4
5	3
6	2
7	4
	Formula Description (Result)
	=MODE(A2:A7) Mode, or most frequently occurring number above (4)

NEGBINOMDIST

[See Also](#)

Returns the negative binomial distribution. NEGBINOMDIST returns the probability that there will be number_f failures before the number_s success, when the constant probability of a success is probability_s. This function is similar to the binomial distribution, except that the number of successes is fixed, and the number of trials is variable. Like the binomial, trials are assumed to be independent.

For example, you need to find 10 people with excellent reflexes, and you know the probability that a candidate has these qualifications is 0.3. NEGBINOMDIST calculates the probability that you will interview a certain number of unqualified candidates before finding all 10 qualified candidates.

Syntax

NEGBINOMDIST(number_f,number_s,probability_s)

Number_f is the number of failures.

Number_s is the threshold number of successes.

Probability_s is the probability of a success.

Remarks

- Number_f and number_s are truncated to integers.
- If any argument is nonnumeric, NEGBINOMDIST returns the #VALUE! error value.
- If probability_s < 0 or if probability_s > 1, NEGBINOMDIST returns the #NUM! error value.
- If number_f < 0 or number_s < 1, NEGBINOMDIST returns the #NUM! error value.
- The equation for the negative binomial distribution is:

$$nb(x; r, p) = \binom{x+r-1}{r-1} p^r (1-p)^x$$

where:

x is number_f, r is number_s, and p is probability_s.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B
1	Data Description
2	10 Number of failures
3	5 Threshold number of successes
4	0.25 Probability of a success
	Formula Description (Result)
	=NEGBINOMDIST(A2,A3,A4) Negative binomial distribution for the terms above (0.05509)

NORMDIST

[See Also](#)

Returns the normal distribution for the specified mean and standard deviation. This function has a very wide range of applications in statistics, including hypothesis testing.

Syntax

NORMDIST(x,mean,standard_dev,cumulative)

X is the value for which you want the distribution.

Mean is the arithmetic mean of the distribution.

Standarddev is the standard deviation of the distribution.

Cumulative is a logical value that determines the form of the function. If cumulative is TRUE, NORMDIST returns the cumulative distribution function; if FALSE, it returns the probability mass function.

Remarks

- If mean or standarddev is nonnumeric, NORMDIST returns the #VALUE! error value.
- If standarddev ≤ 0, NORMDIST returns the #NUM! error value.
- If mean = 0, standarddev = 1, and cumulative = TRUE, NORMDIST returns the standard normal distribution, NORMSDIST.
- The equation for the normal density function (cumulative = FALSE) is:

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{x-\mu}{2\sigma^2}\right)}$$

- When cumulative = TRUE, the formula is the integral from negative infinity to x of the given formula.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B
1 Data	Description
2 x	Value for which you want the distribution
3 μ	Arithmetic mean of the distribution
4 1.5	Standard deviation of the distribution
Formula	Description (Result)
=NORMDIST(A2,A3,A4,TRUE)	Cumulative distribution function for the terms above (0.989)
=NORMDIST(A2,A3,A4,FALSE)	Probability mass function for the terms above (0.109905)

NORMSDIST

[See Also](#)

Returns the standard normal cumulative distribution function. The distribution has a mean of 0 (zero) and a standard deviation of one. Use this function in place of a table of standard normal curve areas.

Syntax

NORMSDIST(z)

Z is the value for which you want the distribution.

Remarks

- If z is nonnumeric, NORMSDIST returns the #VALUE! error value.
- The equation for the standard normal density function is:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B
1 Formula	Description (Result)
2 =NORMSDIST(1.333333)	Normal cumulative distribution function at 1.333333 (0.909)

NORMSINV

[See Also](#)

Returns the inverse of the standard normal cumulative distribution. The distribution has a mean of zero and a standard deviation of one.

Syntax

NORMSINV(probability)

Probability is a probability corresponding to the normal distribution.

Remarks

- If probability is nonnumeric, NORMSINV returns the #NUM! error value.
- If probability < 0 or if probability > 1, NORMSINV returns the #NUM! error value.

Given a value for probability, NORMSINV seeks that value z such that NORMSDIST(z) = probability. Thus, precision of NORMSINV depends on precision of NORMSDIST. NORMSINV uses an iterative search technique. If the search has not converged after 100 iterations, the function returns the #A error value.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B
1 Formula	Description (Result)
2 =NORMSINV(0.989)	Inverse of the standard normal cumulative distribution, with a probability of 0.989 (1.3333)

PEARSON

[See Also](#)

Returns the Pearson product moment correlation coefficient, r, a dimensionless index that ranges from -1.0 to 1.0 inclusive and reflects the extent of a linear relationship between two data sets.

Syntax

PEARSON(array1,array2)

Array1 is a set of independent values.

Array2 is a set of dependent values.

Remarks

- The arguments must be either numbers or names, array constants, or references that contain numbers.
- If an array or reference argument contains text, logical values, or empty cells, those values are ignored; however, cells with the value zero are included.
- If array1 and array2 are empty or have a different number of data points, PEARSON returns the #A error value.
- The formula for the Pearson product moment correlation coefficient, r, is:

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

where x and y are the sample means AVERAGE(array1) and AVERAGE(array2).

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B
1 Independent values	Dependent values
2 9	10
3 7	6
4 5	1
5 3	5
6 1	3
Formula	Description (Result)
=PEARSON(A2:A6,B:B)	Pearson product moment correlation coefficient for the data sets above (0.6979)

PERCENTILE

[See Also](#)

Returns the kth percentile of values in a range. You can use this function to establish a threshold of acceptance. For example, you can decide to examine candidates who score above the 90th percentile.

Syntax

PERCENTILE(array,k)

Array is the array or range of data that defines relative standing.

K is the percentile value in the range 0..1, inclusive.

Remarks

- If array is empty or contains more than 255 data points, PERCENTILE returns the #NUM!error value.
- If k is nonnumeric, PERCENTILE returns the #VALUE!error value.
- If k is <0 or if k >1, PERCENTILE returns the #NUM!error value.
- If k is not a multiple of 1/(n -1), PERCENTILE interpolates to determine the value at the kth percentile.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	
1	Data
2	1
3	3
4	2
5	4
Formula	Description (Result)
=PERCENTILE(A2:A5,0.3)	30th percentile of the list above (1.9)

PERCENTRANK

[See Also](#)

Returns the rank of a value in a data set as a percentage of the data set. This function can be used to evaluate the relative standing of a value within a data set. For example, you can use PERCENTRANK to evaluate the standing of an aptitude test score among all scores for the test.

Syntax

PERCENTRANK(array,x,significance)

Array is the array or range of data with numeric values that defines relative standing.

X is the value for which you want to know the rank.

Significance is an optional value that identifies the number of significant digits for the returned percentage value. If omitted, PERCENTRANK uses three digits (0.xxx).

Remarks

- If array is empty, PERCENTRANK returns the #NUM!error value.
- If significance <1, PERCENTRANK returns the #NUM!error value.
- If x does not match one of the values in array, PERCENTRANK interpolates to return the correct percentage rank.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	
Data	
13	
12	
11	
8	
4	
3	
2	
1	
1	
1	
Formula	Description (Result)
=PERCENTRANK(A2:A11,2)	Percent rank of 2 in the list above (0.333, because 3 values in the set are smaller than 2, and 6 are larger than 2; 3/(3+6)=0.333)
=PERCENTRANK(A2:A11,8)	Percent rank of 8 in the list above (0.555)
=PERCENTRANK(A2:A11,8)	Percent rank of 8 in the list above (0.666)
=PERCENTRANK(A2:A11,5)	Percent rank of 5 in the list above (0.58, one-quarter of the way between the PERCENTRANK of 4 and the PERCENTRANK of 6)

Note To view the number as a percentage, select the cell and click **Cells** on the **Format** menu. Click the **Number** tab, and then click **Percentage** in the **Category** box.

PERMUT

[See Also](#)

Returns the number of permutations for a given number of objects that can be selected from number objects. A permutation is any set or subset of objects or events where internal order is significant. Permutations are different from combinations, for which the internal order is not significant. Use this function for lottery-style probability calculations.

Syntax

PERMUT(number,number_chosen)

Number is an integer that describes the number of objects.

Number_chosen is an integer that describes the number of objects in each permutation.

Remarks

- Both arguments are truncated to integers.
- If number or number_chosen is nonnumeric, PERMUT returns the #NUM! error value.
- If number ≤ 0 or if number_chosen < 0, PERMUT returns the #NUM! error value.
- If number < number_chosen, PERMUT returns the #NUM! error value.
- The equation for the number of permutations is:

$$P_{k,n} = \frac{n!}{(n-k)!}$$

Example

Suppose you want to calculate the odds of selecting a winning lottery number. Each lottery number contains three numbers, each of which can be between 0 (zero) and 9 inclusive. The following function calculates the number of possible permutations:

The example may be easier to understand if you copy it to a blank worksheet.

► [How?](#)

A	B
1 Data	Description
2 100	Number of objects
3 3	Number of objects in each permutation
Formula	Description (Result)
=PERMUT(A2,A3)	Permutations possible for the terms above (90200)

POISSON

[See Also](#)

Returns the Poisson distribution. A common application of the Poisson distribution is predicting the number of events over a specific time, such as the number of cars arriving at a toll plaza in 1 minute.

Syntax

POISSON(x,mean,cumulative)

X is the number of events.

Mean is the expected numeric value.

Cumulative is a logical value that determines the form of the probability distribution returned. If cumulative is TRUE, POISSON returns the cumulative Poisson probability that the number of random events occurring will be between zero and x inclusive; if FALSE, it returns the Poisson probability mass function that the number of events occurring will be exactly x.

Remarks

- If x is not an integer, it is truncated.
- If x or mean is nonnumeric, POISSON returns the #NUM! error value.
- If x < 0, POISSON returns the #NUM! error value.
- If mean ≤ 0, POISSON returns the #NUM! error value.

- POISSON is calculated as follows.

For cumulative = FALSE:

$$POISSON = \frac{e^{-\lambda} \lambda^x}{x!}$$

For cumulative = TRUE:

$$CUMPOISSON = \sum_{k=0}^x \frac{e^{-\lambda} \lambda^k}{k!}$$

Example

The example may be easier to understand if you copy it to a blank worksheet.

► [How?](#)

A	B
1 Data	Description
2 2	Number of events
3 5	Expected mean
Formula	Description (Result)
=POISSON(A2,A3,TRUE)	Cumulative Poisson probability with the terms above (0.12652)
=POISSON(A2,A3,FALSE)	Poisson probability mass function with the terms above (0.0027)

PROB

[See Also](#)

Returns the probability that values in a range are between two limits. If upperlimit is not supplied, returns the probability that values in xrange are equal to lowerlimit.

Syntax

PROB(x_range,prob_range,lower_limit,upperlimit)

x_range is the range of numeric values of x with which there are associated probabilities.

prob_range is a set of probabilities associated with values in xrange.

lowerlimit is the lower bound on the value for which you want a probability.

upperlimit is the optional upper bound on the value for which you want a probability.

Remarks

- If any value in prob_range ≤ 0 or if any value in prob_range > 1 , PROB returns the #NUM! error value.
- If the sum of the values in prob_range $\neq 1$, PROB returns the #NUM! error value.
- If upperlimit is omitted, PROB returns the probability of being equal to lowerlimit.
- If xrange and prob_range contain a different number of data points, PROB returns the #N/A error value.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B
1 x	Probability
2 0	0.2
3 1	0.3
4 2	0.1
5 3	0.4
Formula	Description (Result)
=PROB(A2:A5,B:2)	Probability that x is 2 (0.1)
=PROB(A2:A5,B:1,3)	Probability that x is between 1 and 3 (0.7)

QUARTILE

[See Also](#)

Returns the quartile of a data set. Quartiles often are used in sales and survey data to divide populations into groups. For example, you can use QARTIE to find the top 25 percent of incomes in a population.

Syntax

QUARTILE(array,quart)

Array is the array or cell range of numeric values for which you want the quartile value.

Quart indicates which value to return.

If quart equals QUARTILE returns

0	Minimum value
1	First quartile (25th percentile)
2	Median value (50th percentile)
3	Third quartile (75th percentile)
4	Maximum value

Remarks

- If array is empty, QARTIE returns the #NUM! error value.
- If quart is not an integer, it is truncated.
- If quart < 0 or if quart > 4 , QARTIE returns the #NUM! error value.
- MIN, MEDIAN, and MAX return the same value as QARTIE when quart is equal to 0 (zero), 2, and 4 respectively.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	
1 Data	
2 1	
3 2	
4 4	
5 7	
6 8	
7 9	
8 10	
9 12	
Formula	Description (Result)
=QARTIE(A2:A9,1)	First quartile (25th percentile) of the data above (3.5)

RANK

[See Also](#)

Returns the rank of a number in a list of numbers. The rank of a number is its size relative to other values in a list. (If you were to sort the list, the rank of the number would be its position.)

Syntax

RANK(number,ref,order)

Number is the number whose rank you want to find.

Ref is an array of, or a reference to, a list of numbers. Nonnumeric values in ref are ignored.

Order is a number specifying how to rank number.

- If order is 0 (zero) or omitted, Microsoft Excel ranks number as if ref were a list sorted in descending order.
- If order is any nonzero value, Microsoft Excel ranks number as if ref were a list sorted in ascending order.

Remarks

- RANK gives duplicate numbers the same rank. However, the presence of duplicate numbers affects the ranks of subsequent numbers. For example, in a list of integers sorted in ascending order, if the number 10 appears twice and has a rank of 5, then 11 would have a rank of 7 (no number would have a rank of 6).
- For some purposes one might want to use a definition of rank that takes ties into account. In the previous example, one would want a revised rank of 5.5 for the number 10. This can be done by adding the following correction factor to the value returned by RANK. This correction factor is appropriate both for the case where rank is computed in descending order (order = 0 or omitted) or ascending order (order = nonzero value).

Correction factor for tied ranks=[COUNT(ref) + 1 - RANK(number, ref, 0) - RANK(number, ref, 1)]/2.

In the following example, RANK(A2,A1:A5,1) equals 3. The correction factor is (5 + 1 - 2 - 3)/2 = 0.5 and the revised rank that takes ties into account is 3 + 0.5 = 3.5. If number occurs only once in ref, the correction factor will be 0, since RANK would not have to be adjusted for a tie.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	
1	Data
2	7
3	3.5
4	3.5
5	1
6	2
Formula	Description (Result)
=RANK(A3,A2:A6,1)	Rank of 3.5 in the list above (3)
=RANK(A2,A2:A6,1)	Rank of 7 in the list above (5)

RSQ

[See Also](#)

Returns the square of the Pearson product moment correlation coefficient through data points in known_y's and known_x's. For more information, see PEARSON. The r-squared value can be interpreted as the proportion of the variance in y attributable to the variance in x.

Syntax

RSQ(known_y's,known_x's)

Known_y's is an array or range of data points.

Known_x's is an array or range of data points.

Remarks

- The arguments must be either numbers or names, arrays, or references that contain numbers.
- If an array or reference argument contains text, logical values, or empty cells, those values are ignored; however, cells with the value zero are included.
- If known_y's and known_x's are empty or have a different number of data points, RSQ returns the #A error value.
- The equation for the Pearson product moment correlation coefficient, r, is:

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

where x and y are the sample means AVERAGE(known_x's) and AVERAGE(known_y's).

RSQ returns r², which is the square of this correlation coefficient.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B	
1	Known y	Known x
2	2	6
3	3	5
4	9	11
5	1	7
6	8	5
7	7	4
8	5	4
Formula	Description (Result)	
=RSQ(A2:A8,B2:B8)	Square of the Pearson product moment correlation coefficient through data points above (0.09)	

SKEW

[See Also](#)

Returns the skewness of a distribution. Skewness characterizes the degree of asymmetry of a distribution around its mean. Positive skewness indicates a distribution with an asymmetric tail extending toward more positive values. Negative skewness indicates a distribution with an asymmetric tail extending toward more negative values.

Syntax

SKEW(number1,number2,...)

Number1, number2 ... are 1 to 30 arguments for which you want to calculate skewness. You can also use a single array or a reference to an array instead of arguments separated by commas.

Remarks

- The arguments must be either numbers or names, arrays, or references that contain numbers.
- If an array or reference argument contains text, logical values, or empty cells, those values are ignored; however, cells with the value zero are included.
- If there are fewer than three data points, or the sample standard deviation is zero, SKEW returns the #DIV/0! error value.
- The equation for skewness is defined as:

$$\frac{n}{(n-1)(n-2)} \sum \left(\frac{x_j - \bar{x}}{s} \right)^3$$

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

	A
1	Data
2	3
3	4
4	5
5	2
6	3
7	4
8	5
9	6
10	4
11	7

Formula	Description (Result)
---------	----------------------

=SKEW(A2:A11)	Skewness of a distribution of the data set above (0.39)
---------------	---

SLOPE

[See Also](#)

Returns the slope of the linear regression line through data points in known_y's and known_x's. The slope is the vertical distance divided by the horizontal distance between any two points on the line, which is the rate of change along the regression line.

Syntax

SLOPE(known_y's,known_x's)

Known_y's is an array or cell range of numeric dependent data points.

Known_x's is the set of independent data points.

Remarks

- The arguments must be either numbers or names, arrays, or references that contain numbers.
- If an array or reference argument contains text, logical values, or empty cells, those values are ignored; however, cells with the value zero are included.
- If known_y's and known_x's are empty or have a different number of data points, SLOPE returns the #N/A error value.
- The equation for the slope of the regression line is:

$$b = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}$$

where x and y are the sample means AVERAGE(known_x's) and AVERAGE(known_y's).

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B	
1	Known y	Known x
2	2	6
3	3	5
4	9	11
5	1	7
6	8	5
7	7	4
8	5	4
Formula	Description (Result)	
=SLOPE(A2:A8;B2:B8)	Slope of the linear regression line through the data points above (0.305)	

SMALL

[See Also](#)

Returns the k-th smallest value in a data set. Use this function to return values with a particular relative standing in a data set.

Syntax

SMALL(array,k)

Array is an array or range of numerical data for which you want to determine the k-th smallest value.

K is the position (from the smallest) in the array or range of data to return.

Remarks

- If array is empty, SMALL returns the #NUM!error value.
- If $k \leq 0$ or if k exceeds the number of data points, SMALL returns the #NUM!error value.
- If n is the number of data points in array, SMALL(array,1) equals the smallest value, and SMALL(array,n) equals the largest value.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

	A	B
1	Data	Data
2	3	1
3	4	4
4	5	8
5	2	3
6	3	7
7	4	12
8	6	5
9	4	8
10	7	23

Formula	Description (Result)
---------	----------------------

=SMALL(A2:A10,1) 1st smallest number in first column (3)

=SMALL(B2:B10,2) 2nd smallest number in the second column (3)

STANDARDIZE

[See Also](#)

Returns a normalized value from a distribution characterized by mean and standarddev.

Syntax

STANDARDIZE(x,mean,standard_dev)

X is the value you want to normalize.

Mean is the arithmetic mean of the distribution.

Standarddev is the standard deviation of the distribution.

Remarks

- If standarddev ≤ 0 , STANDARDIZE returns the #NUM!error value.
- The equation for the normalized value is:

$$Z = \frac{X - \mu}{\sigma}$$

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B	
1	Data	Description
2	2	Value to normalize
3	0	Arithmetic mean of the distribution
4	1.5	Standard deviation of the distribution
Formula		Description (Result)
=STANDARDIZE(A2,A3,A4)		Normalized value of 2 for the terms above (1.333333)

STDEV

[See Also](#)

Estimates standard deviation based on a sample. The standard deviation is a measure of how widely values are dispersed from the average value (the mean).

Syntax

STDEV(number1,number2,...)

Number1, number2, ... are 1 to 30 number arguments corresponding to a sample of a population. You can also use a single array or a reference to an array instead of arguments separated by commas.

Remarks

- STDEV assumes that its arguments are a sample of the population. If your data represents the entire population, then compute the standard deviation using STDEVP.
- The standard deviation is calculated using the "unbiased" or "n-1" method.
- STDEV uses the following formula:

$$\sqrt{\frac{\sum (x - \bar{x})^2}{(n-1)}}$$

where x is the sample mean AVERAGE(number1,number2,...) and n is the sample size.

- Logical values such as TRUE and FALSE and text are ignored. If logical values and text must not be ignored, use the STDEVA worksheet function.

Example

Suppose 10 tools stamped from the same machine during a production run are collected as a random sample and measured for breaking strength.

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	
1	Strength
2	135
3	1301
4	1368
5	1322
6	1310
7	130
8	1318
9	130
10	1303
11	129
	Formula Description (Result)
	=STDEV(A2:A11) Standard deviation of breaking strength (273.92)

STDEVA

[See Also](#)

Estimates standard deviation based on a sample. The standard deviation is a measure of how widely values are dispersed from the average value (the mean). Text and logical values such as TRUE and FALSE are included in the calculation.

Syntax

STDEVA(value1,value2,...)

Value1, value2, ... are 1 to 30 values corresponding to a sample of a population. You can also use a single array or a reference to an array instead of arguments separated by commas.

Remarks

- STDEVA assumes that its arguments are a sample of the population. If your data represents the entire population, you must compute the standard deviation using STDEVPA.
- Arguments that contain TRUE evaluate as 1; arguments that contain text or FALSE evaluate as 0 (zero). If the calculation must not include text or logical values, use the STDEV worksheet function instead.
- The standard deviation is calculated using the "unbiased" or "n-1" method.
- STDEVA uses the following formula:

$$\sqrt{\frac{\sum (x - \bar{x})^2}{(n-1)}}$$

where x is the sample mean AVERAGE(value1,value2,...) and n is the sample size.

Example

Suppose 10 tools stamped from the same machine during a production run are collected as a random sample and measured for breaking strength.

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	
1	Strength
2	135
3	1301
4	1368
5	1322
6	1310
7	130
8	1318
9	130
10	1303
11	129
	Formula Description (Result)
	=STDEVA(A2:A11) Standard deviation of breaking strength for all the tools (273.92)

STDEVP

[See Also](#)

Calculates standard deviation based on the entire population given as arguments. The standard deviation is a measure of how widely values are dispersed from the average value (the mean).

Syntax

STDEVP(number1,number2,...)

Number1, number2, ... are 1 to 30 number arguments corresponding to a population. You can also use a single array or a reference to an array instead of arguments separated by commas.

- Text and logical values such as TRUE and FALSE are ignored. If logical values and text must not be ignored, use the STDEVPA worksheet function.

Remarks

- STDEVP assumes that its arguments are the entire population. If your data represents a sample of the population, then compute the standard deviation using STDEV.
- For large sample sizes, STDEV and STDEVP return approximately equal values.
- The standard deviation is calculated using the "biased" or "n" method.
- STDEVP uses the following formula:

$$\sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

where x is the sample mean AVERAGE(number1,number2,...) and n is the sample size.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	
1	Strength
2	135
3	1301
4	1368
5	1322
6	1310
7	130
8	1318
9	130
10	1303
11	129

Formula	Description (Result)
=STDEVP(A2:A11)	Standard deviation of breaking strength, assuming only 10 tools are produced (26.08)

STDEVPA

[See Also](#)

Calculates standard deviation based on the entire population given as arguments, including text and logical values. The standard deviation is a measure of how widely values are dispersed from the average value (the mean).

Syntax

STDEVPA(value1,value2,...)

Value1, value2, ... are 1 to 30 values corresponding to a population. You can also use a single array or a reference to an array instead of arguments separated by commas.

Remarks

- STDEVPA assumes that its arguments are the entire population. If your data represents a sample of the population, you must compute the standard deviation using STDEVA.
- Arguments that contain TRUE evaluate as 1; arguments that contain text or FALSE evaluate as 0 (zero). If the calculation must not include text or logical values, use the STDEVP worksheet function instead.
- For large sample sizes, STDEVA and STDEVPA return approximately equal values.
- The standard deviation is calculated using the "biased" or "n" method.
- STDEVPA uses the following formula:

$$\sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

where x is the sample mean AVERAGE(value1,value2,...) and n is the sample size.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	
1	Strength
2	135
3	1301
4	1368
5	1322
6	1310
7	130
8	1318
9	130
10	1303
11	129

Formula	Description (Result)
=STDEVPA(A2:A11)	Standard deviation of breaking strength, assuming only 10 tools are produced (26.08)

STEYX

[See Also](#)

Returns the standard error of the predicted y-value for each x in the regression. The standard error is a measure of the amount of error in the prediction of y for an individual x.

Syntax

STEYX(known_y's,known_x's)

Known_y's is an array or range of dependent data points.

Known_x's is an array or range of independent data points.

Remarks

- The arguments must either be numbers or be names, arrays, or references that contain numbers.
- If an array or reference argument contains text, logical values, or empty cells, those values are ignored; however, cells with the value zero are included.
- If known_y's and known_x's are empty or have a different number of data points, STEYX returns the #N/A error value.
- The equation for the standard error of the predicted y is:

$$\sqrt{\frac{1}{(n-2)} \left[\sum (y - \bar{y})^2 - \frac{[\sum (x - \bar{x})(y - \bar{y})]^2}{\sum (x - \bar{x})^2} \right]}$$

where x and y are the sample means AVERAGE(known_x's) and AVERAGE(known_y's), and n is the sample size.

Example

The example may be easier to understand if you copy it to a blank worksheet.

▶ [How?](#)

A	B
1 Known y	Known x
2 2	6
3 3	5
4 9	11
5 1	7
6 8	5
7 7	4
8 5	4
Formula	Description (Result)
=STEYX(A2:A8,B2:B8)	Standard error of the predicted y-value for each x in the regression (3.3079)

TDIST

[See Also](#)

Returns the Percentage Points (probability) for the Student t-distribution where a numeric value (x) is a calculated value of t for which the Percentage Points are to be computed. The t-distribution is used in the hypothesis testing of small sample data sets. Use this function in place of a table of critical values for the t-distribution.

Syntax

TDIST(x,degrees_freedom,tails)

X is the numeric value at which to evaluate the distribution.

Degrees_freedom is an integer indicating the number of degrees of freedom.

Tails specifies the number of distribution tails to return. If tails = 1, TDIST returns the one-tailed distribution. If tails = 2, TDIST returns the two-tailed distribution.

Remarks

- If any argument is nonnumeric, TDIST returns the #VALUE! error value.
- If degrees_freedom < 1, TDIST returns the #NUM! error value.
- The degrees_freedom and tails arguments are truncated to integers.
- If tails is any value other than 1 or 2, TDIST returns the #NUM! error value.
- If x < 0, then TDIST returns the #NUM! error value.
- If tails = 1, TDIST is calculated as TDIST = P(X > x), where X is a random variable that follows the t-distribution. If tails = 2, TDIST is calculated as TDIST = P(X > x) = P(X < -x).
- Since x < 0 is not allowed, to use TDIST when x < 0, note that TDIST(-x,df,1) = 1 - TDIST(x,df,1) = P(X > -x) and TDIST(-x,df,2) = TDIST(x,df,2) = P(X > x).

Example

The example may be easier to understand if you copy it to a blank worksheet.

▶ [How?](#)

A	B
1 Data	Description
2 1.0	Value at which to evaluate the distribution
3 60	Degrees of freedom
Formula	Description (Result)
=TDIST(A2,A3,2)	Two-tailed distribution (0.0697 or 6.97 percent)
=TDIST(A2,A3,1)	One-tailed distribution (0.0232263 or 2.3 percent)

Note To view the number as a percent, select the cell and, on the **Format** menu, click **Cells**. Click the **Number** tab, and then click **Percentage** in the **Category** box.

TINV

[See Also](#)

Returns the t-value of the Student's t-distribution as a function of the probability and the degrees of freedom.

Syntax

TINV(probability,degrees_freedom)

Probability is the probability associated with the two-tailed Student's t-distribution.

Degreesfreedom is the number of degrees of freedom with which to characterize the distribution.

Remarks

- If either argument is nonnumeric, TINV returns the #VALUE!error value.
- If probability < 0 or if probability > 1, TINV returns the #NUM!error value.
- If degreesfreedom is not an integer, it is truncated.
- If degreesfreedom < 1, TINV returns the #NUM!error value.
- TINV returns that value t, such that $P(X > t) = \text{probability}$ where X s a random variable that follows the t-distribution and $P(X < -t) = P(X > t)$.
- A one-tailed t-value can be returned by replacing probability with 2*probability. For a probability of 0.05 and degrees of freedom of 10, the two-tailed value is calculated with TINV(0.05,10), which returns 2.2839. The one-tailed value for the same probability and degrees of freedom can be calculated with TINV(2*0.05,10), which returns 1.8262.

Note In some tables, probability is described as (1-p).

Given a value for probability, TINV seeks that value x such that TDIST(x, degreesfreedom, 2) = probability. Thus, precision of TINV depends on precision of TDIST. TINV uses an iterative search technique. If the search has not converged after 100 iterations, the function returns the #N/A error value.

-

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B
1 Data	Description
2 0.05	Probability associated with the two-tailed Student's t-distribution
3 60	Degrees of freedom
Formula	Description (Result)
=TINV(A2,A3)	t-value of the Student's t-distribution for the terms above (1.8262)

TREND

[See Also](#)

Returns values along a linear trend. Fits a straight line (using the method of least squares) to the arrays knowny's and knownx's. Returns the y-values along that line for the array of newx's that you specify.

Syntax

TREND(known_y's,known_x's,new_x's,const)

Knowny's is the set of y-values you already know in the relationship $y = mx + b$.

- If the array knowny's is in a single column, then each column of knownx's is interpreted as a separate variable.
- If the array knowny's is in a single row, then each row of knownx's is interpreted as a separate variable.

Knownx's is an optional set of x-values that you may already know in the relationship $y = mx + b$.

- The array knownx's can include one or more sets of variables. If only one variable is used, knowny's and knownx's can be ranges of any shape, as long as they have equal dimensions. If more than one variable is used, knowny's must be a vector (that is, a range with a height of one row or a width of one column).

- If knownx's is omitted, it is assumed to be the array {1,2,3,...} that is the same size as knowny's.

Newx's are new x-values for which you want TREND to return corresponding y-values.

- Newx's must include a column (or row) for each independent variable, just as knownx's does. So, if knowny's is in a single column, knownx's and newx's must have the same number of columns. If knowny's is in a single row, knownx's and newx's must have the same number of rows.

- If you omit newx's, it is assumed to be the same as knownx's.

- If you omit both knownx's and newx's, they are assumed to be the array {1,2,3,...} that is the same size as knowny's.

Const is a logical value specifying whether to force the constant b to equal 0.

- If const is TRUE or omitted, b is calculated normally.
- If const is FALSE, b is set equal to 0 (zero), and the m-values are adjusted so that $y = mx$.

Remarks

- For information about how Microsoft Excel fits a line to data, see LINES.
- You can use TREND for polynomial curve fitting by regressing against the same variable raised to different powers. For example, suppose column A contains y-values and column B contains x-values. You can enter x^2 in column C, x^3 in column D, and so on, and then regress columns B through D against column A.
- Formulas that return arrays must be entered as array formulas.
- When entering an array constant for an argument such as knownx's, use commas to separate values in the same row and semicolons to separate rows.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

The first formula shows corresponding values to the known values. The second formula predicts the next months values, if the linear trend continues.

A	B	C
1 Month	Cost	Formula (Corresponding Cost)
2 1	\$33,000	=TREND(B2:B13,A2:A13)
3 2	\$37,000	
4 3	\$39,000	
5 4	\$37,000	
6 5	\$38,300	
7 6	\$39,000	
8 7	\$39,000	
9 8	\$41,120	
10 9	\$41,000	
11 10	\$3,230	
12 11	\$4,000	
13		

12	\$20
Month Formula (Predicted Cost)	
13	=TREND(B2:B13,A2:A13,A15A1\$
14	
15	
16	
17	

Note The formula in the example must be entered as an array formula. After copying the example to a blank worksheet, select the range C2:C13 or B15:B19 starting with the formula cell. Press F2, and then press CTRL+SHIFT+ENTER. If the formula is not entered as an array formula, the single results are 1339.3333 and 1417.52.

TRIMMEAN

[See Also](#)

Returns the mean of the interior of a data set. TRIMMEAN calculates the mean taken by excluding a percentage of data points from the top and bottom tails of a data set. You can use this function when you wish to exclude outlying data from your analysis.

Syntax

TRIMMEAN(array,percent)

Array is the array or range of values to trim and average.

Percent is the fractional number of data points to exclude from the calculation. For example, if percent = 0.2, 4 points are trimmed from a data set of 20 points (20 x 0.2): 2 from the top and 2 from the bottom of the set.

Remarks

- If percent < 0 or percent > 1, TRIMMEAN returns the #NUM!error value.
- TRIMMEAN rounds the number of excluded data points down to the nearest multiple of 2. If percent = 0.1, 10 percent of 30 data points equals 3 points. For symmetry, TRIMMEAN excludes a single value from the top and bottom of the data set.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	
1	Data
2	4
3	5
4	6
5	7
6	2
7	3
8	4
9	5
10	1
11	2
12	3

Formula	Description (Result)
=TRIMMEAN(A2:A12,0.2)	Mean of the interior of a data set above, with 20 percent excluded from calculation (3.7)

TTEST

[See Also](#)

Returns the probability associated with a Student's t-Test. Use TTEST to determine whether two samples are likely to have come from the same two underlying populations that have the same mean.

Syntax

TTEST(array1,array2,tails,type)

Array1 is the first data set.

Array2 is the second data set.

Tails specifies the number of distribution tails. If tails = 1, TTEST uses the one-tailed distribution. If tails = 2, TTEST uses the two-tailed distribution.

Type is the kind of t-Test to perform.

If type equals This test is performed

1	Paired
2	Two-sample equal variance (homoscedastic)
3	Two-sample unequal variance (heteroscedastic)

Remarks

- If array1 and array2 have a different number of data points, and type = 1 (paired), TTEST returns the #A error value.
- The tails and type arguments are truncated to integers.
- If tails or type is nonnumeric, TTEST returns the #VALUE!error value.
- If tails is any value other than 1 or 2, TTEST returns the #NUM!error value.
- TTEST uses the data in array1 and array2 to compute a non-negative t-statistic. If tails=1, TTEST returns the probability of a higher value of the t-statistic under the assumption that array1 and array2 are samples from populations with the same mean. The value returned by TTEST when tails=2 is double that returned when tails=1 and corresponds to the probability of a higher absolute value of the t-statistic under the "same population means" assumption.

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

	A	B
1	Data 1	Data 2
2	3	6
3	4	19
4	5	3
5	8	2
6	9	14
7	1	4
8	2	5
9	4	17
10	5	1
	Formula	Description (Result)
	=TTEST(A2:A10,B2:B10,2,1)	Probability associated with a Student's paired t-Test, with a two-tailed distribution (0.10016)

VAR

[See Also](#)

Estimates variance based on a sample.

Syntax

VAR(number1,number2,...)

Number1, number2, ... are 1 to 30 number arguments corresponding to a sample of a population.

Remarks

- VAR assumes that its arguments are a sample of the population. If your data represents the entire population, then compute the variance using VARP.
- Logical values such as TRUE, FALSE, and text are ignored. If logical values and text must not be ignored, use the VARA worksheet function.
- VAR uses the following formula:

$$\frac{\sum (x - \bar{x})^2}{(n - 1)}$$

where x is the sample mean AVERAGE(number1,number2,...) and n is the sample size.

Example

Suppose 10 tools stamped from the same machine during a production run are collected as a random sample and measured for breaking strength.

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

	A	
1	Strength	
2	135	
3	1301	
4	1368	
5	1322	
6	1310	
7	130	
8	1318	
9	130	
10	1303	
11	129	
	Formula	Description (Result)
	=VAR(A2:A11)	Variance for the breaking strength of the tools (266666)

VARA

[See Also](#)

Estimates variance based on a sample. In addition to numbers, text and logical values such as TRUE and FALSE are included in the calculation.

Syntax

VARA(value1,value2,...)

Value1, value2, ... are 1 to 30 value arguments corresponding to a sample of a population.

Remarks

- VARA assumes that its arguments are a sample of the population. If your data represents the entire population, you must compute the variance using VARPA.
- Arguments that contain TRUE evaluate as 1; arguments that contain text or FALSE evaluate as 0 (zero). If the calculation must not include text or logical values, use the VAR worksheet function instead.
- VARA uses the following formula:

$$\frac{\sum (x - \bar{x})^2}{(n - 1)}$$

where x is the sample mean AVERAGE(value1,value2,...) and n is the sample size.

Example

Suppose 10 tools stamped from the same machine during a production run are collected as a random sample and measured for breaking strength.

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	
1	Strength
2	135
3	1301
4	1368
5	1322
6	1310
7	130
8	1318
9	130
10	1303
11	129

Formula	Description (Result)
=VARA(A2:A11)	Estimates the variance for the breaking strength (6666667)

VARP

[See Also](#)

Calculates variance based on the entire population.

Syntax

VARP(number1,number2,...)

Number1, number2, ... are 1 to 30 number arguments corresponding to a population.

Remarks

- VARP assumes that its arguments are the entire population. If your data represents a sample of the population, then compute the variance using VAR.
- The equation for VARP is:

$$\frac{\sum (x - \bar{x})^2}{n}$$

where x is the sample mean AVERAGE(number1,number2,...) and n is the sample size.

- Logical values such as TRUE and FALSE and text are ignored. If logical values and text must not be ignored, use the VARPA worksheet function.

Example

Suppose all 10 tools stamped from the same machine during a production run are collected and measured for breaking strength.

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	
1	Strength
2	135
3	1301
4	1368
5	1322
6	1310
7	130
8	1318
9	130
10	1303
11	129

Formula	Description (Result)
=VARP(A2:A11)	Variance of breaking strengths for all the tools, assuming that only 10 tools are produced (6666667)

VARPA

[See Also](#)

Calculates variance based on the entire population. In addition to numbers, text and logical values such as TRUE and FALSE are included in the calculation.

Syntax

VARPA(value1,value2,...)

Value1, value2, ... are 1 to 30 value arguments corresponding to a population.

Remarks

- VARPA assumes that its arguments are the entire population. If your data represents a sample of the population, you must compute the variance using VARA.
- Arguments that contain TRUE evaluate as 1; arguments that contain text or FALSE evaluate as 0 (zero). If the calculation must not include text or logical values, use the VARP worksheet function instead.
- The equation for VARPA is :

$$\frac{\sum (x - \bar{x})^2}{n}$$

where x is the sample mean AVERAGE(value1,value2,...) and n is the sample size.

Example

Suppose all 10 tools stamped from the same machine during a production run are collected and measured for breaking strength.

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

	A	B
1	Strength	
2	135	
3	1301	
4	1368	
5	1322	
6	1310	
7	130	
8	1318	
9	130	
10	1303	
11	129	
	Formula	Description (Result)
	=VARPA(A2:A11)	Variance of breaking strengths for all the tools, assuming that only 10 tools are produced (63)

WEIBULL

[See Also](#)

Returns the Weibull distribution. Use this distribution in reliability analysis, such as calculating a device's mean time to failure.

Syntax

WEIBULL(x,alpha,beta,cumulative)

X is the value at which to evaluate the function.

Alpha is a parameter to the distribution.

Beta is a parameter to the distribution.

Cumulative determines the form of the function.

Remarks

- If x, alpha, or beta is nonnumeric, WEIBULL returns the #VALUE! error value.
- If x < 0, WEIBULL returns the #NUM! error value.
- If alpha ≤ 0 or if beta ≤ 0, WEIBULL returns the #NUM! error value.
- The equation for the Weibull cumulative distribution function is:
$$F(x; \alpha, \beta) = 1 - e^{-(x/\beta)^\alpha}$$
- The equation for the Weibull probability density function is:
$$f(x; \alpha, \beta) = \frac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha}$$
- When alpha = 1, WEIBULL returns the exponential distribution with:

$$f = \frac{1}{\beta}$$

Example

The example may be easier to understand if you copy it to a blank worksheet.

[▶ How?](#)

A	B	
1	Data	Description
2	105	Value at which to evaluate the function
3	20	Alpha parameter to the distribution
4	100	Beta parameter to the distribution
	Formula	Description (Result)
	=WEIBULL(A2,A3,A4,TRUE)	Weibull cumulative distribution function for the terms above (0.98)
	=WEIBULL(A2,A3,A4,FALSE)	Weibull probability density function for the terms above (0.03)

ZTEST

[See Also](#)

Returns the one-tailed probability-value of a z-test. For a given hypothesized population mean, μ_0 , ZTEST returns the probability that the sample mean would be greater than the average of observations in the data set (array)— that is, the observed sample mean.

To see how ZTEST can be used in a formula to compute a two-tailed probability value, see "Remarks" below.

Syntax

ZTEST(array, μ_0 , sigma)

Array is the array or range of data against which to test μ_0

μ_0 is the value to test.

Sigma is the population (known) standard deviation. If omitted, the sample standard deviation is used.

Remarks

- If array is empty, ZTEST returns the #A error value.
- ZTEST is calculated as follows when sigma is not omitted:

$$ZTEST(array, \mu_0) = 1 - NORMSDIST((\bar{x} - \mu_0) / (sigma / \sqrt{n}))$$

or when sigma is omitted:

$$ZTEST(array, \mu_0) = 1 - NORMSDIST((\bar{x} - \mu_0) / (s / \sqrt{n}))$$

where \bar{x} is the sample mean AVERAGE(array); s is the sample standard deviation STDEV(array); and n is the number of observations in the sample COUNT(array).

- ZTEST represents the probability that the sample mean would be greater than the observed value AVERAGE(array), when the underlying population mean is μ_0 . From the symmetry of the Normal distribution, if AVERAGE(array) < μ_0 , ZTEST will return a value greater than 0.5
- The following Excel formula can be used to calculate the two-tailed probability that the sample mean would be further from μ_0 (in either direction) than AVERAGE(array), when the underlying population mean is μ_0 :

=2 *MIN(ZTEST(array, μ_0 , sigma), 1 - ZTEST(array, μ_0 , sigma)).

Example

The example may be easier to understand if you copy it to a blank worksheet.

► [How?](#)

A		
1	Data	
2	3	
3	6	
4	7	
5	8	
6	6	
7	5	
8	4	
9	2	
10	1	
11	9	
	Formula	Description (Result)
	=ZTEST(A2:A11,9)	One-tailed probability-value of a z-test for the data set above, at the hypothesized population mean of 9 (0.000)
	=2 *MIN(ZTEST(A2:A11,9), 1 - ZTEST(A2:A11,9))	Two-tailed probability-value of a z-test for the data set above, at the hypothesized population mean of 9 (0.1819)
	=ZTEST(A2:A11,6)	One-tailed probability-value of a z-test for the data set above, at the hypothesized population mean of 6 (0.8303)

=2 *MIN(ZTEST(A2:A11,6), 1 - ZTEST(A2:A11,6))

Two-tailed probability-value of a z-test for the data set above, at the hypothesized population mean of 6 (0.2393)

About statistical analysis tools

Microsoft Excel provides a set of data analysis tools— called the Analysis ToolPak— 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 macro functions and then displays the results in an output table. Some tools generate charts in addition to output tables.

Related worksheet functions Excel provides many other statistical, financial, and engineering worksheet functions. Some of the statistical functions are built-in and others become available when you install the Analysis ToolPak.

Accessing the data analysis tools The Analysis ToolPak includes the tools described below. To access these tools, click **Data Analysis** on the **Tools** menu. If the **Data Analysis** command is not available, you need to load the Analysis ToolPak **add-in** (add-in: A supplemental program that adds custom commands or custom features to Microsoft Office.) program.

▼ 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.

Anova: Single Factor 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. If there were only two samples, the worksheet function, TTEST, could equally well be used. With more than two samples, there is no convenient generalization of TTEST and the Single Factor Anova model can be called upon instead.

Anova: Two-Factor With Replication 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:

- Whether heights of plants for the different fertilizer brands are drawn from the same underlying population; temperatures are ignored for this analysis.
- Whether heights of plants for the different temperature levels are drawn from the same underlying population; fertilizer brands are ignored for this analysis.
- 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.

Input range		
	Group 1	Group 2
Trial 1	78	98
	68	98
	71	61
	75	60
Trial 2	88	82
	78	60
	68	59
	68	98

Anova: Two-Factor Without Replication 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.

▼ Correlation

The CORREL and PEARSON worksheet functions both calculate the correlation coefficient between two 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.) The Correlation analysis tool is particularly useful when there are more than two measurement variables for each of N subjects. It provides an output table, a correlation matrix, showing the value of CORREL (or PEARSON) 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).

▼ Covariance

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 Covariance tool computes the value of the worksheet function, COVAR, for each pair of measurement variables. (Direct use of COVAR rather than the Covariance tool is a reasonable alternative when there are only two measurement variables, i.e. N=2.) 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 as calculated by the worksheet function, VARP.

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).

▼ Descriptive Statistics

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.

▼ Exponential Smoothing

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 α , 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.

▼ F-Test Two-Sample for Variances

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 ≤ 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 ≤ 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.

▼ Fourier Analysis

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.

Input range		Output table	
Time	Frequency		
Domain	Domain		
Data	Output		
1	3		
1	1.707106769	-1	7.071067691
	-1		
0	0.292893231	+0	2.92893231i
0			

▼ Histogram

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.

▼ Moving Average

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. Each forecast value is based on the following formula.

$$F_{(t+i)} = \frac{1}{N} \sum_{j=1}^N A_{t-j+1}$$

where:

- N is the number of prior periods to include in the moving average
- A_j is the actual value at time j

- F_j is the forecasted value at time j

▼ Random Number Generation

The Random Number Generation analysis tool fills a range with independent random numbers drawn from one of several distributions. You can characterize subjects in a population with a probability distribution.

For example, you might use a normal distribution to characterize the population of individuals' heights, or you might use a Bernoulli distribution of two possible outcomes to characterize the population of coin-flip results.

▼ Rank and Percentile

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. This tool uses the worksheet functions, RANK and PERCENTRANK. RANK does not account for tied values. If you wish to account for tied values, use the worksheet function, RANK, together with the correction factor suggested in the help file for RANK.

▼ Regression

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.

The Regression tool uses the worksheet function, LINEST.

▼ Sampling

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.

▼ 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 <= t) one-tail" gives the probability that a value of the t-Statistic would be observed that is more negative than t . If $t >= 0$, "P(T <= 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 <= 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.

t-Test: Two-Sample Assuming Equal Variances 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. It is referred to as a homoscedastic t-test. You can use this t-test to determine whether the two samples are likely to have come from distributions with equal population means.

t-Test: Two-Sample Assuming Unequal Variances 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. It is referred to as a heteroscedastic t-test. 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.

The following formula is used to determine the statistic value t .

$$t' = \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}}$$

The following formula is used to calculate the degrees of freedom, df . Because the result of the calculation is usually not an integer, the value of df is rounded to the nearest integer to obtain a critical value from the t table. The Excel worksheet function, TTEST, uses the calculated df value without rounding since it is possible to compute a value for TTEST with a non-integer df . Because of these different approaches to determining degrees of freedom, results of TTEST and this t-Test tool will differ in the Unequal Variances case.

$$df = \frac{\left(\frac{S_1^2}{m} + \frac{S_2^2}{n}\right)^2}{\frac{(S_1^2/m)^2}{m-1} + \frac{(S_2^2/n)^2}{n-1}}$$

t-Test: Paired Two Sample For Means 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, derived from the following formula.

$$S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}$$

▼ z-Test

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. If variances are not known, the worksheet function, ZTEST, should be used instead.

When using the z-Test tool, one should be careful to understand the output. "P(Z <= z) one-tail" is really $P(Z >= \text{ABS}(z))$, 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 $P(Z >= \text{ABS}(z))$ or $Z <= -\text{ABS}(z)$, 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 2. 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.

Worksheet functions listed by category

▼ Database

Microsoft Excel includes worksheet functions that analyze data stored in lists or databases. Each of these functions, referred to collectively as the Dfunctions, uses three arguments: database, field, and criteria. These arguments refer to the worksheet ranges that are used by the function.

[DAVERAGE](#) Returns the average of selected database entries

[DCOUNT](#) Counts the cells that contain numbers in a database

[DCOUNTA](#) Counts nonblank cells in a database

[DGET](#) Extracts from a database a single record that matches the specified criteria

[DMAX](#) Returns the maximum value from selected database entries

[DMIN](#) Returns the minimum value from selected database entries

[DPRODUCT](#) Multiplies the values in a particular field of records that match the criteria in a database

[DSTDEV](#) Estimates the standard deviation based on a sample of selected database entries

[DSTDEVP](#) Calculates the standard deviation based on the entire population of selected database entries

[DSUM](#) Adds the numbers in the field column of records in the database that match the criteria

[DVAR](#) Estimates variance based on a sample from selected database entries

[DVARP](#) Calculates variance based on the entire population of selected database entries

[GETPIVOTDATA](#) Returns data stored in a PivotTable

▼ Date and Time

[DATE](#) Returns the serial number of a particular date

[DATEVALUE](#) Converts a date in the form of text to a serial number

[DAY](#) Converts a serial number to a day of the month

[DAYS360](#) Calculates the number of days between two dates based on a 360-day year

[EDATE](#) Returns the serial number of the date that is the indicated number of months before or after the start date

[EOMONTH](#) Returns the serial number of the last day of the month before or after a specified number of months

[HOUR](#) Converts a serial number to an hour

[MINUTE](#) Converts a serial number to a minute

[MONTH](#) Converts a serial number to a month

[NETWORKDAYS](#) Returns the number of whole workdays between two dates

[NOW](#) Returns the serial number of the current date and time

[SECOND](#) Converts a serial number to a second

[TIME](#) Returns the serial number of a particular time

[TIMEVALUE](#) Converts a time in the form of text to a serial number

[TODAY](#) Returns the serial number of today's date

[WEEKDAY](#) Converts a serial number to a day of the week

[WEEKNUM](#) Converts a serial number to a number representing where the week falls numerically with a year

[WORKDAY](#) Returns the serial number of the date before or after a specified number of workdays

[YEAR](#) Converts a serial number to a year

[YEARFRAC](#) Returns the year fraction representing the number of whole days between start_date and end_date

▼ External

These functions are loaded with [add-in](#) programs

[EUROCONVERT](#) Converts a number to euros, converts a number from euros to a euro member currency, or converts a number from one euro member currency to another by using the euro as an intermediary (triangulation)

[SQL.REQUEST](#) Connects with an external data source and runs a query from a worksheet, then returns the result as an array without the need for macro programming

▼ Engineering

[BESSELI](#) Returns the modified Bessel function $I_n(x)$

[BESSELJ](#) Returns the Bessel function $J_n(x)$

[BESSELK](#) Returns the modified Bessel function $K_n(x)$

[BESSELY](#) Returns the Bessel function $Y_n(x)$

[BIN2DEC](#) Converts a binary number to decimal

[BIN2HEX](#) Converts a binary number to hexadecimal

[BIN2OCT](#) Converts a binary number to octal

[COMPLEX](#) Converts real and imaginary coefficients into a complex number

[CONVERT](#) Converts a number from one measurement system to another

[DEC2BIN](#) Converts a decimal number to binary

[DEC2HEX](#) Converts a decimal number to hexadecimal

[DEC2OCT](#) Converts a decimal number to octal

[DELTA](#) Tests whether two values are equal

[ERF](#) Returns the error function

[ERFC](#) Returns the complementary error function

[GESTEP](#) Tests whether a number is greater than a threshold value

[HEX2BIN](#) Converts a hexadecimal number to binary

[HEX2DEC](#) Converts a hexadecimal number to decimal

[HEX2OCT](#) Converts a hexadecimal number to octal

[IMABS](#) Returns the absolute value (modulus) of a complex number

[IMAGINARY](#) Returns the imaginary coefficient of a complex number

[IMARGUMENT](#) Returns the argument theta, an angle expressed in radians

[IMCONJUGATE](#) Returns the complex conjugate of a complex number

[IMCOS](#) Returns the cosine of a complex number

[IMDIV](#) Returns the quotient of two complex numbers

[IMEXP](#) Returns the exponential of a complex number

[IMLN](#) Returns the natural logarithm of a complex number

[IMLOG10](#) Returns the base-10 logarithm of a complex number

[IMLOG2](#) Returns the base-2 logarithm of a complex number

[IMPOWER](#) Returns a complex number raised to an integer power

[IMPRODUCT](#) Returns the product of two complex numbers

[IMREAL](#) Returns the real coefficient of a complex number

[IMSIN](#) Returns the sine of a complex number

[IMSQRT](#) Returns the square root of a complex number

[IMSUB](#) Returns the difference between two complex numbers

[IMSUM](#) Returns the sum of complex numbers

[OCT2BIN](#) Converts an octal number to binary

[OCT2DEC](#) Converts an octal number to decimal

[OCT2HEX](#) Converts an octal number to hexadecimal

▼ Financial

[ACCRINT](#) Returns the accrued interest for a security that pays periodic interest

[ACCRINTM](#) Returns the accrued interest for a security that pays interest at maturity

[AMORDEGRC](#) Returns the depreciation for each accounting period by using a depreciation coefficient

[AMORLINC](#) Returns the depreciation for each accounting period

[COUPDAYBS](#) Returns the number of days from the beginning of the coupon period to the settlement date

[COUPDAYS](#) Returns the number of days in the coupon period that contains the settlement date

[COUPDAYSNC](#) Returns the number of days from the settlement date to the next coupon date

[COUPNCD](#) Returns the next coupon date after the settlement date

[COUPNUM](#) Returns the number of coupons payable between the settlement date and maturity date

[COUPPCD](#) Returns the previous coupon date before the settlement date

[CUMIPMT](#) Returns the cumulative interest paid between two periods

[CUMPRINC](#) Returns the cumulative principal paid on a loan between two periods

[DB](#) Returns the depreciation of an asset for a specified period using the fixed-declining balance method

[DDB](#) Returns the depreciation of an asset for a specified period using the double-declining balance method or some other method you specify

[DISC](#) Returns the discount rate for a security

[DOLLARDE](#) Converts a dollar price, expressed as a fraction, into a dollar price, expressed as a decimal number

[DOLLARFR](#) Converts a dollar price, expressed as a decimal number, into a dollar price, expressed as a fraction

[DURATION](#) Returns the annual duration of a security with periodic interest payments

[EFFECT](#) Returns the effective annual interest rate

[FV](#) Returns the future value of an investment

[FVSCHEDULE](#) Returns the future value of an initial principal after applying a series of compound interest rates

[INTRATE](#) Returns the interest rate for a fully invested security

[IPMT](#) Returns the interest payment for an investment for a given period

[IRR](#) Returns the internal rate of return for a series of cash flows

[ISPMT](#) Calculates the interest paid during a specific period of an investment

[MDURATION](#) Returns the Macauley modified duration for a security with an assumed par value of \$100

[MIRR](#) Returns the internal rate of return where positive and negative cash flows are financed at different rates

[NOMINAL](#) Returns the annual nominal interest rate

[NPER](#) Returns the number of periods for an investment

[NPV](#) Returns the net present value of an investment based on a series of periodic cash flows and a discount rate

[ODDFPRICE](#) Returns the price per \$100 face value of a security with an odd first period

[ODDFYIELD](#) Returns the yield of a security with an odd first period

[ODDLPRICE](#) Returns the price per \$100 face value of a security with an odd last period

[ODDLYIELD](#) Returns the yield of a security with an odd last period

[PMT](#) Returns the periodic payment for an annuity

[PPMT](#) Returns the payment on the principal for an investment for a given period

[PRICE](#) Returns the price per \$100 face value of a security that pays periodic interest

[PRICEDISC](#) Returns the price per \$100 face value of a discounted security

[PRICEMAT](#) Returns the price per \$100 face value of a security that pays interest at maturity

[PV](#) Returns the present value of an investment

[RATE](#) Returns the interest rate per period of an annuity

[RECEIVED](#) Returns the amount received at maturity for a fully invested security

[SLN](#) Returns the straight-line depreciation of an asset for one period

[SYD](#) Returns the sum-of-years' digits depreciation of an asset for a specified period

[TBILLEQ](#) Returns the bond-equivalent yield for a Treasury bill

[TBILLPRICE](#) Returns the price per \$100 face value for a Treasury bill

[TBILLYIELD](#) Returns the yield for a Treasury bill

[VDB](#) Returns the depreciation of an asset for a specified or partial period using a declining balance method

[XIRR](#) Returns the internal rate of return for a schedule of cash flows that is not necessarily periodic

[XNPV](#) Returns the net present value for a schedule of cash flows that is not necessarily periodic

[YIELD](#) Returns the yield on a security that pays periodic interest

[YIELDDISC](#) Returns the annual yield for a discounted security; for example, a Treasury bill

[YIELDMAT](#) Returns the annual yield of a security that pays interest at maturity

▼ Information

[CELL](#) Returns information about the formatting, location, or contents of a cell

[ERROR.TYPE](#) Returns a number corresponding to an error type

[INFO](#) Returns information about the current operating environment

[ISBLANK](#) Returns TRUE if the value is blank

[ISERR](#) Returns TRUE if the value is any error value except #N/A

[ISERROR](#) Returns TRUE if the value is any error value

[ISEVEN](#) Returns TRUE if the number is even

[ISLOGICAL](#) Returns TRUE if the value is a logical value

[ISNA](#) Returns TRUE if the value is the #N/A error value

[ISNONTEXT](#) Returns TRUE if the value is not text

[ISNUMBER](#) Returns TRUE if the value is a number

[ISODD](#) Returns TRUE if the number is odd

[ISREF](#) Returns TRUE if the value is a reference

[ISTEXT](#) Returns TRUE if the value is text

[N](#) Returns a value converted to a number

[NA](#) Returns the error value #N/A

[TYPE](#) Returns a number indicating the data type of a value

▼ Logical

[AND](#) Returns TRUE if all its arguments are TRUE

[FALSE](#) Returns the logical value FALSE

[IF](#) Specifies a logical test to perform

[NOT](#) Reverses the logic of its argument

[OR](#) Returns TRUE if any argument is TRUE

[TRUE](#) Returns the logical value TRUE

▼ Lookup and Reference

[ADDRESS](#) Returns a reference as text to a single cell in a worksheet

[AREAS](#) Returns the number of areas in a reference

[CHOOSE](#) Chooses a value from a list of values

[COLUMN](#) Returns the column number of a reference

[COLUMNS](#) Returns the number of columns in a reference

[HLOOKUP](#) Looks in the top row of an array and returns the value of the indicated cell

[HYPERLINK](#) Creates a shortcut or jump that opens a document stored on a network server, an intranet, or the Internet

[INDEX](#) Uses an index to choose a value from a reference or array

[INDIRECT](#) Returns a reference indicated by a text value

[LOOKUP](#) Looks up values in a vector or array

[MATCH](#) Looks up values in a reference or array

[OFFSET](#) Returns a reference offset from a given reference

[ROW](#) Returns the row number of a reference

[ROWS](#) Returns the number of rows in a reference

[RTD](#) Retrieves real-time data from a program that supports COM automation

[TRANSPOSE](#) Returns the transpose of an array

[VLOOKUP](#) Looks in the first column of an array and moves across the row to return the value of a cell

▼ Math and Trigonometry

[ABS](#) Returns the absolute value of a number

[ACOS](#) Returns the arccosine of a number

[ACOSH](#) Returns the inverse hyperbolic cosine of a number

[ASIN](#) Returns the arcsine of a number

[ASINH](#) Returns the inverse hyperbolic sine of a number

[ATAN](#) Returns the arctangent of a number

[ATAN2](#) Returns the arctangent from x- and y-coordinates
[ATANH](#) Returns the inverse hyperbolic tangent of a number
[CEILING](#) Rounds a number to the nearest integer or to the nearest multiple of significance
[COMBIN](#) Returns the number of combinations for a given number of objects
[COS](#) Returns the cosine of a number
[COSH](#) Returns the hyperbolic cosine of a number
[DEGREES](#) Converts radians to degrees
[EVEN](#) Rounds a number up to the nearest even integer
[EXP](#) Returns e raised to the power of a given number
[FACT](#) Returns the factorial of a number
[FACTDOUBLE](#) Returns the double factorial of a number
[FLOOR](#) Rounds a number down, toward zero
[GCD](#) Returns the greatest common divisor
[INT](#) Rounds a number down to the nearest integer
[LCM](#) Returns the least common multiple
[LN](#) Returns the natural logarithm of a number
[LOG](#) Returns the logarithm of a number to a specified base
[LOG10](#) Returns the base-10 logarithm of a number
[MDETERM](#) Returns the matrix determinant of an array
[MINVERSE](#) Returns the matrix inverse of an array
[MMULT](#) Returns the matrix product of two arrays
[MOD](#) Returns the remainder from division
[MROUND](#) Returns a number rounded to the desired multiple
[MULTINOMIAL](#) Returns the multinomial of a set of numbers
[ODD](#) Rounds a number up to the nearest odd integer
[PI](#) Returns the value of pi
[POWER](#) Returns the result of a number raised to a power
[PRODUCT](#) Multiplies its arguments
[QUOTIENT](#) Returns the integer portion of a division
[RADIANS](#) Converts degrees to radians
[RAND](#) Returns a random number between 0 and 1
[RANDBETWEEN](#) Returns a random number between the numbers you specify
[ROMAN](#) Converts an arabic numeral to roman, as text
[ROUND](#) Rounds a number to a specified number of digits
[ROUNDDOWN](#) Rounds a number down, toward zero
[ROUNDUP](#) Rounds a number up, away from zero
[SERIESSUM](#) Returns the sum of a power series based on the formula
[SIGN](#) Returns the sign of a number
[SIN](#) Returns the sine of the given angle
[SINH](#) Returns the hyperbolic sine of a number
[SQRT](#) Returns a positive square root
[SQRTPI](#) Returns the square root of (number * pi)
[SUBTOTAL](#) Returns a subtotal in a list or database
[SUM](#) Adds its arguments
[SUMIF](#) Adds the cells specified by a given criteria
[SUMPRODUCT](#) Returns the sum of the products of corresponding array components
[SUMSQ](#) Returns the sum of the squares of the arguments

[SUMX2MY2](#) Returns the sum of the difference of squares of corresponding values in two arrays
[SUMX2PY2](#) Returns the sum of the sum of squares of corresponding values in two arrays
[SUMXMY2](#) Returns the sum of squares of differences of corresponding values in two arrays
[TAN](#) Returns the tangent of a number
[TANH](#) Returns the hyperbolic tangent of a number
[TRUNC](#) Truncates a number to an integer
 ▼ Statistical
[AVEDEV](#) Returns the average of the absolute deviations of data points from their mean
[AVERAGE](#) Returns the average of its arguments
[AVERAGEA](#) Returns the average of its arguments, including numbers, text, and logical values
[BETADIST](#) Returns the beta cumulative distribution function
[BETAINV](#) Returns the inverse of the cumulative distribution function for a specified beta distribution
[BINOMDIST](#) Returns the individual term binomial distribution probability
[CHIDIST](#) Returns the one-tailed probability of the chi-squared distribution
[CHIINV](#) Returns the inverse of the one-tailed probability of the chi-squared distribution
[CHITEST](#) Returns the test for independence
[CONFIDENCE](#) Returns the confidence interval for a population mean
[CORREL](#) Returns the correlation coefficient between two data sets
[COUNT](#) Counts how many numbers are in the list of arguments
[COUNTA](#) Counts how many values are in the list of arguments
[COUNTBLANK](#) Counts the number of blank cells within a range
[COUNTIF](#) Counts the number of nonblank cells within a range that meet the given criteria
[COVAR](#) Returns covariance, the average of the products of paired deviations
[CRITBINOM](#) Returns the smallest value for which the cumulative binomial distribution is less than or equal to a criterion value
[DEVSQ](#) Returns the sum of squares of deviations
[EXPONDIST](#) Returns the exponential distribution
[FDIST](#) Returns the F probability distribution
[FINV](#) Returns the inverse of the F probability distribution
[FISHER](#) Returns the Fisher transformation
[FISHERINV](#) Returns the inverse of the Fisher transformation
[FORECAST](#) Returns a value along a linear trend
[FREQUENCY](#) Returns a frequency distribution as a vertical array
[FTEST](#) Returns the result of an F-test
[GAMMADIST](#) Returns the gamma distribution
[GAMMAINV](#) Returns the inverse of the gamma cumulative distribution
[GAMMALN](#) Returns the natural logarithm of the gamma function, $\Gamma(x)$
[GEOMEAN](#) Returns the geometric mean
[GROWTH](#) Returns values along an exponential trend
[HARMEAN](#) Returns the harmonic mean
[HYPGEOMDIST](#) Returns the hypergeometric distribution
[INTERCEPT](#) Returns the intercept of the linear regression line
[KURT](#) Returns the kurtosis of a data set
[LARGE](#) Returns the k-th largest value in a data set
[LINEST](#) Returns the parameters of a linear trend
[LOGEST](#) Returns the parameters of an exponential trend
[LOGINV](#) Returns the inverse of the lognormal distribution
[LOGNORMDIST](#) Returns the cumulative lognormal distribution

[MAX](#) Returns the maximum value in a list of arguments

[MAXA](#) Returns the maximum value in a list of arguments, including numbers, text, and logical values

[MEDIAN](#) Returns the median of the given numbers

[MIN](#) Returns the minimum value in a list of arguments

[MINA](#) Returns the smallest value in a list of arguments, including numbers, text, and logical values

[MODE](#) Returns the most common value in a data set

[NEGBINOMDIST](#) Returns the negative binomial distribution

[NORMDIST](#) Returns the normal cumulative distribution

[NORMINV](#) Returns the inverse of the normal cumulative distribution

[NORMSDIST](#) Returns the standard normal cumulative distribution

[NORMSINV](#) Returns the inverse of the standard normal cumulative distribution

[PEARSON](#) Returns the Pearson product moment correlation coefficient

[PERCENTILE](#) Returns the k-th percentile of values in a range

[PERCENTRANK](#) Returns the percentage rank of a value in a data set

[PERMUT](#) Returns the number of permutations for a given number of objects

[POISSON](#) Returns the Poisson distribution

[PROB](#) Returns the probability that values in a range are between two limits

[QUARTILE](#) Returns the quartile of a data set

[RANK](#) Returns the rank of a number in a list of numbers

[RSQ](#) Returns the square of the Pearson product moment correlation coefficient

[SKEW](#) Returns the skewness of a distribution

[SLOPE](#) Returns the slope of the linear regression line

[SMALL](#) Returns the k-th smallest value in a data set

[STANDARDIZE](#) Returns a normalized value

[STDEV](#) Estimates standard deviation based on a sample

[STDEVA](#) Estimates standard deviation based on a sample, including numbers, text, and logical values

[STDEVP](#) Calculates standard deviation based on the entire population

[STDEVPA](#) Calculates standard deviation based on the entire population, including numbers, text, and logical values

[STEYX](#) Returns the standard error of the predicted y-value for each x in the regression

[TDIST](#) Returns the Student's t-distribution

[TINV](#) Returns the inverse of the Student's t-distribution

[TREND](#) Returns values along a linear trend

[TRIMMEAN](#) Returns the mean of the interior of a data set

[TTEST](#) Returns the probability associated with a Student's t-test

[VAR](#) Estimates variance based on a sample

[VARA](#) Estimates variance based on a sample, including numbers, text, and logical values

[VARP](#) Calculates variance based on the entire population

[VARPA](#) Calculates variance based on the entire population, including numbers, text, and logical values

[WEIBULL](#) Returns the Weibull distribution

[ZTEST](#) Returns the one-tailed probability-value of a z-test

▼ Text and Data

[ASC](#) Changes full-width (double-byte) English letters or katakana within a character string to half-width (single-byte) characters

[BAHTTEXT](#) Converts a number to text, using the ฿ (baht) currency format

[CHAR](#) Returns the character specified by the code number

[CLEAN](#) Removes all nonprintable characters from text

[CODE](#) Returns a numeric code for the first character in a text string

[CONCATENATE](#) Joins several text items into one text item

[DOLLAR](#) Converts a number to text, using the \$ (dollar) currency format

[EXACT](#) Checks to see if two text values are identical

[FIND](#) Finds one text value within another (case-sensitive)

[FIXED](#) Formats a number as text with a fixed number of decimals

[JIS](#) Changes half-width (single-byte) English letters or katakana within a character string to full-width (double-byte) characters

[LEFT](#) Returns the leftmost characters from a text value

[LEN](#) Returns the number of characters in a text string

[LOWER](#) Converts text to lowercase

[MID](#) Returns a specific number of characters from a text string starting at the position you specify

[PHONETIC](#) Extracts the phonetic (furigana) characters from a text string

[PROPER](#) Capitalizes the first letter in each word of a text value

[REPLACE](#) Replaces characters within text

[REPT](#) Repeats text a given number of times

[RIGHT](#) Returns the rightmost characters from a text value

[SEARCH](#) Finds one text value within another (not case-sensitive)

[SUBSTITUTE](#) Substitutes new text for old text in a text string

[T](#) Converts its arguments to text

[TEXT](#) Formats a number and converts it to text

[TRIM](#) Removes spaces from text

[UPPER](#) Converts text to uppercase

[VALUE](#) Converts a text argument to a number