

VARSAAMP Function

Computes the variance among all values in a column using the sample statistical method. Input column can be of Integer or Decimal. If no numeric values are detected in the input column, the function returns 0.

The **variance** of a set of values attempts to measure the spread in values around the mean. A variance of zero means that all values are the same, and a small variance means that the values are closely bunched together. A high value for variance indicates that the numbers are spread out widely. Variance is always a positive value.

If a row contains a missing or null value, it is not factored into the calculation.

NOTE: This function applies to a sample of the entire population. More information is below.

Relevant terms:

Term	Description
Population	Population statistical functions are computed from all possible values. See https://en.wikipedia.org/wiki/Statistical_population .
Sample	Sample-based statistical functions are computed from a subset or sample of all values. See https://en.wikipedia.org/wiki/Sampling_(statistics) . These function names include SAMP in their name.

NOTE: Statistical sampling has no relationship to the samples taken within the product. When statistical functions are computed during job execution, they are applied across the entire dataset. Sample method calculations are computed at that time.

In the following computation, the sample method computes variances with $N - 1$ as the divisor.

$$\text{Var}(X) = [\text{Sum}((X - \text{mean}(X))^2)] / (\text{Count}(X) - 1)$$

The square root of variance is standard deviation, which is used to measure variance under the assumption of a bell curve distribution. See *STDEV Function*.

For a version of this function computed over a rolling window of rows, see *ROLLINGVAR Function*.

Wrangle vs. SQL: This function is part of Wrangle, a proprietary data transformation language. Wrangle is not SQL. For more information, see *Wrangle Language*.

Basic Usage

```
varsamp(myRating)
```

Output: Returns the variance of the group of values from the `myRating` column using the sample method of calculation.

Syntax and Arguments

```
varsamp(col_ref) [group:group_col_ref] [limit:limit_count]
```

Argument	Required?	Data Type	Description
function_col_ref	Y	string	Name of column to which to apply the function

For more information on the `group` and `limit` parameters, see *Pivot Transform*.

For more information on syntax standards, see *Language Documentation Syntax Notes*.

col_ref

Name of the column whose values you wish to use in the calculation. Column must be a numeric (Integer or Decimal) type.

- Literal values are not supported as inputs.
- Multiple columns and wildcards are not supported.

Usage Notes:

Required?	Data Type	Example Value
Yes	String (column reference)	myValues

Examples

Tip: For additional examples, see *Common Tasks*.

This example shows some of the statistical functions that use the sample method of computation. These include:

- `STDEVSAMP` - computes standard deviation using the sample method. See *STDEVSAMP Function*.
- `VARSAAMP` - computes variance using the sample method. See *VARSAAMP Function*.
- `STDEVSAMPIF` - computes standard deviation based on a condition and using the sample method. See *STDEVSAMPIF Function*.
- `VARSAMPIF` - computes standard deviation based on a condition and using the sample method. See *VARSAMPIF Function*.

Source:

Students took tests on three consecutive Saturdays:

Student	Date	Score
Andrew	11/9/19	81
Bella	11/9/19	84
Christina	11/9/19	79
David	11/9/19	64
Ellen	11/9/19	61
Fred	11/9/19	63
Andrew	11/16/19	73
Bella	11/16/19	88

Christina	11/16/19	78
David	11/16/19	67
Ellen	11/16/19	87
Fred	11/16/19	90
Andrew	11/23/19	76
Bella	11/23/19	93
Christina	11/23/19	81
David	11/23/19	97
Ellen	11/23/19	97
Fred	11/23/19	91

Transformation:

You can use the following transformations to calculate standard deviation and variance across all dates using the sample method. Each computation has been rounded to three digits.

Transformation Name	New formula
Parameter: Formula type	Single row formula
Parameter: Formula	round(stdevsamp(Score), 3)
Parameter: New column name	'stdevSamp'

Transformation Name	New formula
Parameter: Formula type	Single row formula
Parameter: Formula	round(varsamp(Score), 3)
Parameter: New column name	'varSamp'

You can use the following to limit the previous statistical computations to the last two Saturdays of testing:

Transformation Name	New formula
Parameter: Formula type	Single row formula
Parameter: Formula	round(stdevsampilf(Score, Date != '11\9\2019'), 3)
Parameter: New column name	'stdevSampIf'

Transformation Name	New formula
Parameter: Formula type	Single row formula
Parameter: Formula	round(varsampilf(Score, Date != '11\9\2019'), 3)
Parameter: New column name	'varSampIf'

Results:

Student	Date	Score	varSamplf	stdevSamplf	varSamp	stdevSamp
Andrew	11/9/19	81	94.515	9.722	131.673	11.475
Bella	11/9/19	84	94.515	9.722	131.673	11.475
Christina	11/9/19	79	94.515	9.722	131.673	11.475
David	11/9/19	64	94.515	9.722	131.673	11.475
Ellen	11/9/19	61	94.515	9.722	131.673	11.475
Fred	11/9/19	63	94.515	9.722	131.673	11.475
Andrew	11/16/19	73	94.515	9.722	131.673	11.475
Bella	11/16/19	88	94.515	9.722	131.673	11.475
Christina	11/16/19	78	94.515	9.722	131.673	11.475
David	11/16/19	67	94.515	9.722	131.673	11.475
Ellen	11/16/19	87	94.515	9.722	131.673	11.475
Fred	11/16/19	90	94.515	9.722	131.673	11.475
Andrew	11/23/19	76	94.515	9.722	131.673	11.475
Bella	11/23/19	93	94.515	9.722	131.673	11.475
Christina	11/23/19	81	94.515	9.722	131.673	11.475
David	11/23/19	97	94.515	9.722	131.673	11.475
Ellen	11/23/19	97	94.515	9.722	131.673	11.475
Fred	11/23/19	91	94.515	9.722	131.673	11.475