CORREL Function

Computes the correlation coefficient between two columns. Source values can be of Integer or Decimal type.

The **correlation coefficient** measures the relationship between two sets of values. You can use it as a measurement for how changes in one value affect changes in the other.

- Values range between -1 (negative correlation) and +1 (positive correlation).
  - Negative correlation means that the second number tends to decrease when the first number increases.
  - Positive correlation means that the second number tends to increase when the first number increases.
  - A correlation coefficient that is close to 0 indicates a weak or non-existent correlation.

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.

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

```
correl(initialInvestment,ROI)
```

Output: Returns the correlation coefficient between the values in the `initialInvestment` column and the `ROI` column.

Syntax and Arguments

```
correl(function_col_ref1,function_col_ref2) [group:group_col_ref] [limit:limit_count]
```

Argument	Required?	Data Type	Description
function_col_ref1	Y	string	Name of column that is the first input to the function
function_col_ref2	Y	string	Name of column that is the second input to 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*.

function_col_ref1, function_col_ref2

Name of the column the values of which you want to calculate the correlation. Column must contain Integer or Decimal values.

- 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)	`myInputs`

Examples

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

This example illustrates statistical functions that can be applied across two columns of values.

Functions:

Item	Description
CORREL Function	Computes the correlation coefficient between two columns. Source values can be of Integer or Decimal type.
COVAR Function	Computes the covariance between two columns using the population method. Source values can be of Integer or Decimal type.
COVARSAMP Function	Computes the covariance between two columns using the sample method. Source values can be of Integer or Decimal type.
ROUND Function	Rounds input value to the nearest integer. Input can be an Integer, a Decimal, a column reference, or an expression. Optional second argument can be used to specify the number of digits to which to round.

Source:

The following table contains height in inches and weight in pounds for a set of students.

Student	heightIn	weightLbs
1	70	134
2	67	135
3	67	147
4	67	160
5	72	136
6	73	146
7	71	135
8	63	145
9	67	138

10	66	138
11	71	161
12	70	131
13	74	131
14	67	157
15	73	161
16	70	133
17	63	132
18	64	153
19	64	156
20	72	154

Transformation:

You can use the following transformations to calculate the correlation co-efficient, the covariance, and the sampling method covariance between the two data columns:

Transformation Name	`New formula`
Parameter: Formula type	`Single row formula`
Parameter: Formula	`round(correl(heightIn, weightLbs), 3)`
Parameter: New column name	`'corrHeightAndWeight'`

Transformation Name	`New formula`
Parameter: Formula type	`Single row formula`
Parameter: Formula	`round(covar(heightIn, weightLbs), 3)`
Parameter: New column name	`'covarHeightAndWeight'`

Transformation Name	`New formula`
Parameter: Formula type	`Single row formula`
Parameter: Formula	`round(covarsamp(heightIn, weightLbs), 3)`
Parameter: New column name	`'covarHeightAndWeight-Sample'`

Results:

Student	heightIn	weightLbs	covarHeightAndWeight-Sample	covarHeightAndWeight	corrHeightAndWeight
1	70	134	-2.876	-2.732	-0.074
2	67	135	-2.876	-2.732	-0.074
3	67	147	-2.876	-2.732	-0.074

4	67	160	-2.876	-2.732	-0.074
5	72	136	-2.876	-2.732	-0.074
6	73	146	-2.876	-2.732	-0.074
7	71	135	-2.876	-2.732	-0.074
8	63	145	-2.876	-2.732	-0.074
9	67	138	-2.876	-2.732	-0.074
10	66	138	-2.876	-2.732	-0.074
11	71	161	-2.876	-2.732	-0.074
12	70	131	-2.876	-2.732	-0.074
13	74	131	-2.876	-2.732	-0.074
14	67	157	-2.876	-2.732	-0.074
15	73	161	-2.876	-2.732	-0.074
16	70	133	-2.876	-2.732	-0.074
17	63	132	-2.876	-2.732	-0.074
18	64	153	-2.876	-2.732	-0.074
19	64	156	-2.876	-2.732	-0.074
20	72	154	-2.876	-2.732	-0.074