

COVAR Function

Computes the covariance between two columns using the population method. Source values can be of Integer or Decimal type.

Covariance measures the joint variation between two sets of values. The sign of the covariance tends to show the linear relationship between the two datasets; positive covariance indicates that the numbers tend to increase with each other.

- The magnitude of the covariance is difficult to interpret, as it varies with the size of the source values.
- The normalized version of covariance is the correlation coefficient, in which covariance is normalized between -1 and 1. For more information, see *CORREL Function*.

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. |

- This function is calculated across the entire population.
- For more information on a sampled version of this function, see *COVAR_SAMP 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

```
covar(squareFootage, purchasePrice)
```

Output: Returns the covariance between the values in the `squareFootage` column and the `purchasePrice` column.

Syntax and Arguments

```
covar(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 covariance. 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 the following two-column statistical functions:

- **CORREL** - Correlation co-efficient between two columns. See *CORREL Function*.
- **COVAR** - Calculates the covariance between two columns. See *COVAR Function*.
- **COVARSAAMP** - Calculates the covariance between two columns using the sample population method. See *COVARSAAMP Function*.

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 |