  
**Statistical Working Group  
Statistical Guidelines**

**CL\_STATISTICAL\_OPERATION  
Code list for statistical operation performed on data**

**Version 1.0 – 09/06/2023**

# **Name**

Code list for concept “Statistical operation” (ID “STATISTICAL\_OPERATION”).

# **Description**

This code list represents common statistical operations for calculated values.

# **Explanatory notes**

It may be used as a dimension to have multiple series showing the result of different operations, or as an attribute to describe which operation was applied to the underlying data. It can describe the calculation of primary data into an aggregate. For example, the population of a country that is calculated as a sum of individuals.

# **Established international standard(s) used as input for the code list**

None

# Recommended Codes

|  |  |  |
| --- | --- | --- |
| **Code Id** | **Code Name** | **Code Description** |
| MEAN | Arithmetic mean | Value representing the arithmetic mean of a distribution. Example “The mean age of persons in a country” |
| MEANG | Geometric mean | Value representing the geometric mean of a distribution. |
| MEANW | Weighted mean | Value representing the weighted arithmetic mean |
| VALUEW | Weighted value | Value representing the weighted value |
| MEDIAN | Median | Value representing the median of a distribution. |
| OBS | Observed | Observed value without applying any statistical operation applied |
| COUNT | Count | Value representing the Count of a distribution.  Example: the number of enterprises in a country having a specific economic activity. |
| INDEX | Index | Represents an index value for example for prices.  Example: the index of industrial production. |
| RATE | Rate | Value computed through a rate of values.  Example: Unemployment rate |
| Q1 | Quintile 1 | Value representing the first quintile cut point of a distribution. |
| Q2 | Quintile 2 | Value representing the second quintile cut point of a distribution. |
| Q3 | Quintile 3 | Value representing the third quintile cut point of a distribution. |
| Q4 | Quintile 4 | Value representing the fourth quintile cut point of a distribution. |
| D1 | Decile 1 | Value representing the first decile cut point of a distribution. |
| D2 | Decile 2 | Value representing the second decile cut point of a distribution. |
| D3 | Decile 3 | Value representing the third decile cut point of a distribution. |
| D4 | Decile 4 | Value representing the fourth decile cut point of a distribution. |
| D5 | Decile 5 | Value representing the fifth decile cut point of a distribution. |
| D6 | Decile 6 | Value representing the sixth decile cut point of a distribution. |
| D7 | Decile 7 | Value representing the seventh decile cut point of a distribution. |
| D8 | Decile 8 | Value representing the eighth decile cut point of a distribution. |
| D9 | Decile 9 | Value representing the ninth decile cut point of a distribution. |
| BOUNDL | Lower bound | Value representing the lower bound of a distribution. |
| BOUNDU | Upper bound | Value representing the lower bound of a distribution. |
| D9\_5 | Decile 9/5 | Value representing the ratio of the ninth decile to the fifth decile. |
| D9\_1 | Decile 9/1 | Value representing the ratio of the ninth decile to the first decile. |
| D5\_1 | Decile 5/1 | Value representing the ratio of the fifth decile to the first decile. |
| QRT1 | Quartile 1 | Value representing the first quartile cut point of a distribution. |
| QRT2 | Quartile 2 | Value representing the second quartile cut point of a distribution. |
| QRT3 | Quartile 3 | Value representing the third quartile cut point of a distribution. |
| SD | Standard deviation | Value representing the standard deviation of a distribution. |
| VAR | Variance | Value representing the variance of a distribution. |
| KUR | Kurtosis | Value representing the “tailedness” of the probability distribution of a real-valued random variable. |
| SKEW | Skewness | Value representing the measure of the lack of symmetry of a distribution |
| ADMEAN | Mean absolute deviation | Value representing the mean absolute deviation of a distribution. |
| ADMEDIAN | Median absolute deviation | Value representing the median absolute deviation of a distribution. |
| SUM | Raw sum | Value computed as a sum of values.  Example: the gross domestic product of a country. |
| SUMW | Weighted sum | Value computed as a weighted sum of values. |
| RANGEIQ | Interquartile range | Value representing the interquartile range of a distribution. |
| RANGE | Range | Value representing the difference between the largest and the smallest values of a distribution. |
| BOUNDLCI | Confidence interval lower bound | Value representing the confidence interval lower bound of a distribution |
| BOUNDUCI | Confidence interval upper bound | Value representing the confidence interval upper bound of a distribution |
| MIN | Minimum | Value representing the minimum of a distribution. |
| MAX | Maximum | Value representing the maximum of a distribution |
| SCORE | Score | Value as a result of a scoring function |
| SE | Standard error | Value representing the standard error of the mean of a distribution |
| RANK | Rank | Value as a result of a ranking function |
| \_Z | Not applicable​​ |  |

## **Remarks**

## Correspondence to VTL pre-defined operators

VTL has corresponding pre-defined operators for some operations in this codelist. For example:

|  |  |  |
| --- | --- | --- |
| **CL\_STATISTICAL\_OPERATION** | | **VTL 2.0[[1]](#footnote-1) operator** |
| **Code Id** | **Code Name** |
| MEAN | Arithmetic mean | avg |
| MEDIAN | Median | median |
| COUNT | Count | count |
| SUM | Raw sum | sum |
| MIN | Minimum | min |
| MAX | Maximum | max |
| RANK | Rank | rank |

All of the operations in this codelist can be replicated in VTL using expressions.

***Quantiles***

The quantiles (quartiles, quintiles, deciles) in this codelist are defined as the cut points (not groups) in the distribution, therefore there is one fewer quantile than the number of groups. This follows the definition and examples at <https://en.wikipedia.org/wiki/Quantile>.

#### Example of even-sized population quartiles[[2]](#footnote-2)

Consider an ordered population of 10 data values [3, 6, 7, 8, 8, 10, 13, 15, 16, 20]. What are the "quartiles" of this dataset?

|  |  |  |
| --- | --- | --- |
| **Quartile** | **Calculation** | **Result** |
| First quartile | The rank of the first quartile is 10×(1/4) = 2.5, which rounds up to 3, meaning that 3 is the rank in the population (from least to greatest values) at which approximately 1/4 of the values are less than the value of the first quartile. The third value in the population is 7. | 7 |
| Second quartile | The rank of the second quartile (same as the median) is 10×(2/4) = 5, which is an integer, while the number of values (10) is an even number, so the average of both the fifth and sixth values is taken—that is (8+10)/2 = 9, though any value from 8 through to 10 could be taken to be the median. | 9 |
| Third quartile | The rank of the third quartile is 10×(3/4) = 7.5, which rounds up to 8. The eighth value in the population is 15. | 15 |

So the first, second and third quartiles of the dataset are [7, 9, 15].

#### Example of odd-sized population quartiles[[3]](#footnote-3)

Consider an ordered population of 11 data values [3, 6, 7, 8, 8, 9, 10, 13, 15, 16, 20]. What are the "quartiles" of this dataset?

|  |  |  |
| --- | --- | --- |
| **Quartile** | **Calculation** | **Result** |
| First quartile | The first quartile is determined by 11×(1/4) = 2.75, which rounds up to 3, meaning that 3 is the rank in the population (from least to greatest values) at which approximately 1/4 of the values are less than the value of the first quartile. The third value in the population is 7. | 7 |
| Second quartile | The second quartile value (same as the median) is determined by 11×(2/4) = 5.5, which rounds up to 6. Therefore, 6 is the rank in the population (from least to greatest values) at which approximately 2/4 of the values are less than the value of the second quartile (or median). The sixth value in the population is 9. | 9 |
| Third quartile | The third quartile value for the original example above is determined by 11×(3/4) = 8.25, which rounds up to 9. The ninth value in the population is 15. | 15 |

So the first, second and third quartiles of the dataset are [7, 9, 15].

## Recommended code values

Further recommended code values for expressing general statistical concepts can be found in section “Generic codes” of the "Guidelines for the creation and management of SDMX Cross-Domain Code Lists"[[4]](#footnote-4).

1. [https://sdmx.org/wp-content/uploads/VTL-2.0-Reference-Manual-20180712-final.pdf](https://sdmx.org/wp-content/uploads/VTL-2.0-Reference-Manual-20180712-final.pdf%20%20)  section VTL-ML - Aggregate and Analytic operators. All other operations can be done using VTL expressions. [↑](#footnote-ref-1)
2. Example source: <https://en.wikipedia.org/wiki/Quantile> [↑](#footnote-ref-2)
3. Example source: <https://en.wikipedia.org/wiki/Quantile> [↑](#footnote-ref-3)
4. <https://sdmx.org/?page_id=4345%20-%20CodeListGuideline#CodeListGuideline> [↑](#footnote-ref-4)