**OP19:** Distribution of observation period length, stratified by age.

This query is used to provide summary statistics for the observation period length across all observation period records stratified by age: the mean, the standard deviation, the minimum, the 25th percentile, the median, the 75th percentile, the maximum and the number of missing values. The length of an is defined as the difference between the start date and the end date. The age value is defined at the time of the observation date. All existing age values are summarized.

**Sample query:**

SELECT

age,

count(\*) AS observation\_periods\_cnt ,

min( period\_length ) AS min\_period ,

max( period\_length ) AS max\_period ,

round( avg( period\_length ), 2 ) AS avg\_period ,

round( stdDev( period\_length ), 1 ) AS stdDev\_period ,

percentile\_25,

median,

percentile\_75

FROM (

SELECT

person\_id,

age,

period\_length,

PERCENTILE\_DISC(0.25) WITHIN GROUP( ORDER BY period\_length ) over(partition by age) AS percentile\_25 ,

PERCENTILE\_DISC(0.5) WITHIN GROUP (ORDER BY period\_length ) over(partition by age) AS median ,

PERCENTILE\_DISC(0.75) WITHIN GROUP (ORDER BY period\_length ) over(partition by age) AS percentile\_75

FROM /\* person, age \*/ (

SELECT

person\_id ,

extract( YEAR FROM first\_observation\_date ) - year\_of\_birth AS age

FROM (

SELECT

person\_id ,

min( observation\_period\_start\_date ) AS first\_observation\_date

FROM observation\_period

GROUP BY person\_id

)

JOIN person USING( person\_id )

WHERE year\_of\_birth IS NOT NULL

)

JOIN (

SELECT

person\_id ,

observation\_period\_end\_date - observation\_period\_start\_date + 1 AS period\_length

FROM observation\_period

) USING( person\_id )

)

GROUP BY

age,

percentile\_25 ,

median ,

percentile\_75;

**Input:**

None

**Output:**

|  |  |
| --- | --- |
| **Field** | **Description** |
| age | Stratification age |
| observation\_period\_cnt | Number of observation periods |
| min\_period | Minimum number of observation periods grouped by age |
| max\_period | Maximum number of observation periods grouped by age |
| avg\_period | Average number of observation periods grouped by age |
| stdDev\_period | Standard deviation of observation periods grouped by age |
| percentile\_25 | 25th percentile of observation periods stratified by age |
| median | Median of observation periods stratified by age |
| percentile\_75 | 75th percentile of observation periods stratified by age |

**Sample output record:**

|  |  |
| --- | --- |
| **Field** | **Description** |
| age | 1 |
| observation\_period\_cnt | 49990 |
| min\_period | 1 |
| max\_period | 2372 |
| avg\_period | 571.28 |
| stdDev\_period | 40.60 |
| percentile\_25 | 365 |
| median | 366 |
| percentile\_75 | 730 |