

文章

[姚鑫](#) · 一月 3, 2022 阅读大约需 4 分钟

第十二章 SQL聚合函数 VARIANCE, VAR_SAMP, VAR_POP

第十二章 SQL聚合函数 VARIANCE, VAR_SAMP, VAR_POP

返回数据集统计方差的聚合函数。

大纲

```
VARIANCE([ALL | DISTINCT [BY(col-list)]] expression [%FOREACH(col-list)] [%AFTERHAVING])
```

```
VAR_SAMP([ALL | DISTINCT [BY(col-list)]] expression [%FOREACH(col-list)] [%AFTERHAVING])
```

```
VAR_POP([ALL | DISTINCT [BY(col-list)]] expression [%FOREACH(col-list)] [%AFTERHAVING])
```

参数

- ALL - 可选-指定统计方差函数返回表达式中所有值的方差。
如果没有指定关键字，这是默认值。
- DISTINCT - 可选- DISTINCT子句，指定统计方差函数返回不同(唯一)表达式值的方差。
DISTINCT可以指定BY(col-list)子句，其中col-list可以是单个字段，也可以是用逗号分隔的字段列表。
- expression - 任何有效的表达式。
通常是包含要分析方差的数据值的列的名称。
- %FOREACH(col-list) - 可选-列名或以逗号分隔的列名列表。
- %AFTERHAVING - 可选-应用在HAVING子句中的条件。

这些函数返回NUMERIC数据类型，除非表达式是数据类型DOUBLE。
如果expression为DOUBLE，则返回DOUBLE。

描述

这三个方差聚合函数在丢弃NULL值后返回表达式值的统计方差。
也就是说，从数据集的平均值变化的量，表示为一个正数。
返回值越大，值的数据集的变化就越大。
SQL还提供聚合函数来返回对应于每个方差函数的标准偏差。

在这种统计差异的推导过程中有一些细微的差异：

- 方差:如果数据集中所有的值都有相同的值(无可变性)，则返回0。
如果数据集只包含一个值(没有可能的可变性)，则返回0。
如果数据集没有值，则返回NULL。

方差计算为：

$$\frac{(\text{SUM}(\text{expression}^{**2}) * \text{COUNT}(\text{expression})) - \text{SUM}(\text{expression}^{**2})}{\text{COUNT}(\text{expression}) * (\text{COUNT}(\text{expression}) - 1)}$$

- VARSAMP:样本方差。
如果数据集中的所有值都具有相同的值(无可变性), 则返回0。
如果数据集只包含一个值(没有可能的可变性), 则返回NULL。
如果数据集没有值, 则返回NULL。
使用与方差相同的变量计算。
- VARPOP:总体方差。
如果数据集中的所有值都具有相同的值(无可变性), 则返回0。
如果数据集只包含一个值(没有可能的可变性), 则返回0。
如果数据集没有值, 则返回NULL。

VARPOP的计算是:

$$\frac{(\text{SUM}(\text{expression}^{**2}) * \text{COUNT}(\text{expression})) - (\text{SUM}(\text{expression})^{**2})}{(\text{COUNT}(\text{expression})^{**2})}$$

这些方差聚合函数可以在引用表或视图的SELECT查询或子查询中使用。
它们可以在SELECT列表或HAVING子句中与普通字段值一起出现。

这些方差聚合函数不能在WHERE子句中使用。
它们不能在JOIN的ON子句中使用, 除非SELECT是子查询。

这些方差聚合函数返回数据类型NUMERIC的值, 精度为36, 刻度为17, 除非表达式是数据类型DOUBLE, 在这种情况下, 函数返回数据类型DOUBLE。

这些方差聚合函数通常应用于具有数值的字段或表达式。
它们将非数值值(包括空字符串(""))计算为零(0)。

这些方差聚合函数忽略数据字段中的NULL值。
如果查询没有返回行, 或者返回的所有行的数据字段值为NULL, 则返回NULL。

与所有聚合函数一样, 统计方差函数可以采用一个可选的DISTINCT子句。
方差(DISTINCT col1)返回那些不同(唯一)的col1字段值的方差。
方差(DISTINCT BY(col2) col1)返回记录中col1字段值的方差, 其中col2值是不同的(唯一的)。
但是请注意, 不同的col2值可能包含一个单独的NULL值。

当前事务期间所做的更改

与所有聚合函数一样, 方差函数总是返回数据的当前状态, 包括未提交的更改, 而不管当前事务的隔离级别如何。

示例

下面的示例使用方差返回 sample.employee 中雇员年龄的方差, 以及由一个或多个雇员表示的不同年龄的方差:

```
SELECT VARIANCE(Age) AS AgeVar, VARIANCE(DISTINCT Age) AS PerAgeVar
FROM Sample.Employee
```

下面的示例使用 VARPOP 返回 sample.employee

中雇员年龄的总体差异，以及由一个或多个雇员表示的不同年龄的差异：

```
SELECT VAR_POP(Age) AS AgePopVar, VAR_POP(DISTINCT Age) AS PerAgePopVar
FROM Sample.Employee
```

[#SQL](#) [#Caché](#)

源

URL:

<https://cn.community.intersystems.com/post/%E7%AC%AC%E5%8D%81%E4%BA%8C%E7%AB%A0-sql%E8%81%9A%E5%90%88%E5%87%BD%E6%95%B0-variance-varsamp-varpop>