**CO15:** Distribution of number of distinct conditions persons have

This query is used to provide summary statistics for the number of different distinct conditions (condition\_concept\_id) of all persons: the mean, the standard deviation, the minimum, the 25th percentile, the median, the 75th percentile, the maximum and the number of missing values. No input is required for this query.

**Sample query:**

with ranked as (

SELECT

num\_of\_conditions,

row\_number() over (order by num\_of\_conditions asc) AS rownumasc

FROM(

select

person\_id,

count(distinct condition\_concept\_id) AS num\_of\_conditions

FROM condition\_occurrence

where person\_id!=0

GROUP BY person\_id

)

),

other\_stat AS (

SELECT

count(num\_of\_conditions) as condition\_dist\_num\_count,

min(num\_of\_conditions) as condition\_dist\_num\_min,

max(num\_of\_conditions) as condition\_dist\_num\_max,

avg(num\_of\_conditions) as condition\_dist\_num\_averege,

stddev(num\_of\_conditions) as condition\_dist\_num\_stddev

FROM (

SELECT

count(distinct condition\_concept\_id) AS num\_of\_conditions,

person\_id

FROM condition\_occurrence

WHERE person\_id!=0

GROUP BY person\_id

)

)

SELECT

(SELECT count(distinct person\_id) FROM condition\_occurrence WHERE person\_id!=0 and condition\_occurrence\_id is null) AS condition\_null\_count,

\*

FROM

other\_stat,

( SELECT distinct

num\_of\_conditions as condition\_dist\_num\_25percentile

FROM

(select \*,(select count(\*) from ranked) as rowno from ranked)

WHERE

(rownumasc=cast (rowno\*0.25 as int) and mod(rowno\*25,100)=0) or

(rownumasc=cast (rowno\*0.25 as int) and mod(rowno\*25,100)>0) or

(rownumasc=cast (rowno\*0.25 as int)+1 and mod(rowno\*25,100)>0)

) AS condition\_end\_date\_25percentile,

( SELECT distinct

num\_of\_conditions as condition\_dist\_num\_median

FROM

(select \*,(select count(\*) from ranked) as rowno from ranked)

WHERE

(rownumasc=cast (rowno\*0.50 as int) and mod(rowno\*50,100)=0) or

(rownumasc=cast (rowno\*0.50 as int) and mod(rowno\*50,100)>0) or

(rownumasc=cast (rowno\*0.50 as int)+1 and mod(rowno\*50,100)>0)

) AS condition\_end\_date\_median,

( SELECT distinct

num\_of\_conditions as condition\_dist\_num\_75percentile

FROM

(select \*,(select count(\*) from ranked) as rowno from ranked)

WHERE

(rownumasc=cast (rowno\*0.75 as int) and mod(rowno\*75,100)=0) or

(rownumasc=cast (rowno\*0.75 as int) and mod(rowno\*75,100)>0) or

(rownumasc=cast (rowno\*0.75 as int)+1 and mod(rowno\*75,100)>0)

) AS condition\_end\_date\_75percentile;

**Input:**

None

**Output:**

|  |  |
| --- | --- |
| **Field** | **Description** |
| condition\_null\_count | Number of condition occurrences where condition\_occurrence\_id is null |
| condition\_count | Number of condition occurrences |
| condition\_dist\_num\_min | The lowest number of distinct condition occurrences |
| condition\_dist\_num\_max | The highest number of distinct condition occurrences |
| condition\_dist\_num\_averege | The avarege number of distinct condition occurrences |
| condition\_dist\_num\_stddev | The standard deviation of distinct condition occurence numbers |
| condition\_dist\_num\_25percentile | A distinct condition occurrence number where 25 percent of the other numbers are lower |
| condition\_dist\_num\_median | A distinct condition occurrence number where half of the other numbers are lower and half are higher |
| condition\_dist\_num\_75percentile | A distinct condition occurrence number where 75 percent of the other numbers are lower |

**Sample output record:**

|  |  |
| --- | --- |
| **Field** | **Description** |
| condition\_null\_count | 0 |
| condition\_count | 4395019 |
| condition\_dist\_num\_min | 1 |
| condition\_dist\_num\_max | 327 |
| condition\_dist\_num\_averege | 17 |
| condition\_dist\_num\_stddev | 16.94 |
| condition\_dist\_num\_25percentile | 6 |
| condition\_dist\_num\_median | 12 |
| condition\_dist\_num\_75percentile | 23 |