

Forming clusters abiding to the rules for statistical disclosure control

When calculating statistics in geographical areas, there are requirements for how many households should be included in the calculation as a minimum. The requirements are either 50, 100 or 150 households depending on the statistical variable. The list of standard variables shows the minimum household requirement for each variable. You can find the list on our homepage here:

<https://www.dst.dk/en/TilSalg/produkter/geodata/kvadratnet>

This means that grid cells with a smaller number of households than the minimum requirement must be added together with other cells before statistics for this geographical area can be disclosed. The merged cells are called clusters.

How are the clusters formed?

For a given household number requirement (K), the clusters are formed as follows:

1. All square grid cells are uniquely placed in one municipality. If a cell belongs to several municipalities, the municipality with the most households in the cell is selected.
2. The clustering described in points 3 and 4 below is then carried out separately for each municipality.
3. All cells with a minimum K number of households will form an independent cluster and are therefore not added to other cells.
4. The remaining cells in the municipality (with less than K households) are added according to the following principle:
 - a. Starting with the cell in the southwest corner of the remaining cells.
 - b. This cell is added to the closest cell in bird flight line. If there is the same distance to several cells, the westernmost of these cells is selected.
 - c. Next, the number of households in this new pooled cluster is calculated.
 - d. If the number of households is still below K , the new geographical center of the clusters (averages of their X , Y coordinates) is calculated and then points b-d above are repeated until the cluster consists of a minimum K households.
 - e. The new cluster is given a unique cluster ID consisting of the municipality code and a serial number.
 - f. If there are a minimum of K households remaining in the municipality, points a-e above are repeated.

If there are fewer than K households remaining in the municipality, the remaining households will be merged with the most recently formed cluster.

What does data look like?

A file is provided which contains both statistical information at the cluster level and an overview of which square grid cells belong to which clusters. The statistical information includes an average value for each cluster as well as the distribution of households at different intervals according to the statistical variable.

The following table shows statistics on each cluster.

Cluster ID	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Number of households	Average
1	6	7	123	23	8	11	178	5,2
2	2	7	99	13	6	3	130	6,2
3	9	8	183	18	18	3	239	4,1
4	1	2	78	19	25	16	141	7,3
5	7	8	77	12	6	0	110	6,7
6	5	2	81	91	3	11	193	4,8
...								

The following table shows which clusters the individual square grid cells belong to.

Municipality	Cluster ID	Cell ID	Number of households in cell	Number of households in cluster
XX	1	1	92	178
XX	1	2	86	178
XX	2	3	76	130
XX	2	4	54	130
YY	3	5	85	239
YY	3	6	73	239
YY	3	7	81	239
YY	4	8	78	141
YY	4	9	63	141
...				

Cluster Formation – an example

Below is an example of how square grid cells are merged into clusters. The number of households is seen in the cells. In the example, the minimum household requirement is 100. Square grid cells without households are not clustered.

1: The number of households in cells is identified.

There is one cell (cell C2) in the area with a minimum of 100 households - it can become an independent cluster (cluster # 1 marked with light green). The remaining cells (marked in gray) are individually too small to form an independent cluster.

6	0	0	0	0	20
5	0	0	10	0	0
4	0	20	0	40	0
3	0	20	15	50	0
2	0	50	120	0	0
1	0	0	0	0	0
	A	B	C	D	E

2: Cluster # 2 is created

A new cluster (cell B2 marked dark green) is created according to the principle described above. Since it is not large enough to be an independent cluster, it must be added to several cells selected from the remaining cells (highlighted in gray).

6	0	0	0	0	20
5	0	0	10	0	0
4	0	20	0	40	0
3	0	20	15	50	0
2	0	50	120	0	0
1	0	0	0	0	0
	A	B	C	D	E

3: Cluster # 2 is finalized

Cell B2 is joined with the closest cells according to the principle described above. Cluster # 2 is now complete.

6	0	0	0	0	20
5	0	0	10	0	0
4	0	20	0	40	0
3	0	20	15	50	0
2	0	50	120	0	0
1	0	0	0	0	0
	A	B	C	D	E

4: Cluster # 3 is created

Now the next of the remaining cells is selected (cell D3). Cluster # 3 is created (marked in blue).

6	0	0	0	0	20
5	0	0	10	0	0
4	0	20	0	40	0
3	0	20	15	50	0
2	0	50	120	0	0
1	0	0	0	0	0
	A	B	C	D	E

5: Cluster # 3

Cell D3 is added to the nearest cells. Cluster # 3 is now complete.

6	0	0	0	0	20
5	0	0	10	0	0
4	0	20	0	40	0
3	0	20	15	50	0
2	0	50	120	0	0
1	0	0	0	0	0
	A	B	C	D	E

6: Cluster # 3 is finalized

Since there are now less than 100 households that are not assigned to a cluster, cell E6 is merged with cluster # 3, which will look like this.

6	0	0	0	0	20
5	0	0	10	0	0
4	0	20	0	40	0
3	0	20	15	50	0
2	0	50	120	0	0
1	0	0	0	0	0
	A	B	C	D	E

Contact information

Questions about the above can be answered by contacting

Allan Hansen

e-mail: hnn@dst.dk

tel .: +45 39 17 31 68