



Supplementary Material 6.5

Clustering

Objective

The objective of this cluster analysis is to assign observations to groups (“clusters”) so that observations within each group are similar to one another with respect to variables or attributes of interest and the groups themselves stand apart from one another. In other words, the objective is to divide the observations into homogeneous and distinct groups.

Target Group

This supplementary material is intended to support superiors and senior technicians to manage their teams for the development of a maintenance plan and to coordinate the work.

1 Working procedure

- Select a measure of similarity.
- Decision is to be made on the type of clustering technique to be used.
- Decision regarding the number of clusters.

2 Select a measure of similarity

A great deal of subjectivity is involved in the choice of similarity measures. Important considerations are the nature of the variables: discrete, continuous, binary or scales of measurement. For example: material (steel, ductile iron, PE, etc.); purpose (main pipe, distribution pipe); sector (downtown, suburb, commercial sector, hydraulic DMA, etc.); leak rate; material age; risk (hospital); transport volume; non-revenue water; damage rate, etc.

3 Decision on the type of clustering technique to be used

There are many kinds of clustering. For this exercise are used

- Disjoint cluster where every object appears in single cluster.
- Hierarchical cluster where one cluster can be completely contained in another cluster, but no other kind of overlap is permitted.

4 Decision regarding the number of clusters

Clustering begins with a series of successive merges of successive divisions. Consider a natural process of grouping.

For example:

- Cluster 1
 - Non-revenue water $> 0.2 \text{ m}^3/\text{h*km}$
 - Damages > 0.5 per km and year
- Cluster 2
 - Non-revenue water between $0.1\text{-}0.2 \text{ m}^3/\text{h*km}$
 - Damages between $0.1 - 0.5$ per km and year
- Cluster 3
 - Non-revenue water between $< 0.1 \text{ m}^3/\text{h*km}$
 - Damages between < 0.1 per km and year.