

## K-Means Clustering Analysis

With this procedure, one first specifies the number of groups to be formed among the objects. The procedure uses a procedure to load each of the k groups with one object in a somewhat random manner. The procedure then iteratively adds or subtracts objects from each group based on an error measure of the distance between the objects in the group. The procedure ends when subsequent iterations do not produce a lower value or the number of iterations has been exceeded.

In this example, we loaded the cansas.LAZ file to group the 20 subjects into four groups. The results may be compared with the other cluster methods of this chapter.

The main grid should contain data values representing variables measured on the objects to be clustered (rows.) Enter the desired number of clusters, select the variables to use in clustering and select the options desired.

No. of Desired Clusters

No. of Iterations

Analysis Options

- ☒ Standardize Variables
- ☐ Replace Grid Values
- ☐ Descriptive Statistics

Available Variables

Selected Variables

Reset Cancel Compute Return

**Figure 1. The K-Means Clustering Form**

Results are:

K-Means Clustering. Adapted from AS 136 APPL. STATIST. (1979) VOL.28, NO.1

File = C:\Documents and Settings\Owner\My Documents\Projects\Clanguage\OpenStat4\cansas.OS4  
No. Cases = 20, No. Variables = 6, No. Clusters = 4

NUMBER OF SUBJECTS IN EACH CLUSTER

Cluster = 1 with 1 cases.  
Cluster = 2 with 7 cases.  
Cluster = 3 with 9 cases.  
Cluster = 4 with 3 cases.

#### PLACEMENT OF SUBJECTS IN CLUSTERS

##### CLUSTER SUBJECT

1	14
2	2
2	6
2	8
2	1
2	15
2	17
2	20
3	11
3	12
3	13
3	4
3	5
3	9
3	18
3	19
3	10
4	7
4	16
4	3

#### AVERAGE VARIABLE VALUES BY CLUSTER

##### VARIABLES

CLUSTER	1	2	3	4	5	6
1	0.11	1.03	-0.12	-0.30	-0.02	-0.01
2	-0.00	0.02	-0.02	-0.19	-0.01	-0.01
3	-0.02	-0.20	0.01	0.17	0.01	0.01
4	0.04	0.22	0.05	0.04	-0.00	0.01

#### WITHIN CLUSTER SUMS OF SQUARES

Cluster 1 = 0.000  
Cluster 2 = 0.274  
Cluster 3 = 0.406  
Cluster 4 = 0.028