

HDDA Tutorial: Clustering

Department of Econometrics and Business Statistics, Monash University

Tutorial 4

Concepts

1. Why might a researcher use cluster analysis?
2. Distinguish between hierarchical and non-hierarchical cluster techniques. In what situations would you use each of these techniques?

Application

This section uses the `mtcars` dataset which is available with R.

1. Cluster the data using Ward's D2 method (with squared distances) and Euclidean distance.
2. Produce a dendrogram for the above analysis
3. Discuss whether the following choices for the number of clusters are suitable or not
 - `One-cluster`
 - `Two-cluster`
 - `Three-cluster`
 - `Four-cluster`
4. For the 2-cluster solution, store the cluster membership in a new variable.
5. Repeat part 4 using:
 - (a) Average Linkage
 - (b) Centroid Method
 - (c) Complete Linkage Method
6. Find the adjusted Rand Index between your answer to question 4 and
 - (a) Your answer to 5 (a)
 - (b) Your answer to 5 (b)
 - (c) Your answer to 5 (c)