

Assignment #8: Cluster Analysis (50 points)

Data: The data for this assignment come from the metadata relating to time-use surveys conducted across various countries in the mid 1970s. Along with a few demographic descriptors for the twenty-eight groups, these data represent the average number of minutes spent in a fixed-time observation period of 40 hours. Activities are grouped into ten general categories. We call this study the “time-use case.” The original data were presented in 1974 in the French journal *Revue de Statistique Appliquée*, available online at http://www.numdam.org/article/RSA_1976__24_2_45_0.pdf

Assignment Instructions:

In this assignment we summarize the time-use case study data using cluster analysis.

(1) Review of Prior Research

- Provide an executive summary of previous work completed with these case data.

(2) Distance Measures and Input Matrices

- Compute alternative distance measures that could be used in identifying clusters of time-use variables/activities. Prepare distance matrices for input to cluster analysis.
- Transpose the input data matrix and compute alternative distance measures that could be used in identifying clusters of demographic groups. Prepare distance matrices for input to cluster analysis.

(3) Clusters of Activities

- Utilize both partitioning (K-Means) and hierarchical cluster analysis methods to obtain clusters of time-use variables/activities. Pick the cluster solution that you find easiest to interpret, and summarize the results.

(4) Clusters of Demographic Groups

- Utilize both partitioning (K-Means) and hierarchical cluster analysis methods to obtain clusters of demographic groups. Pick the cluster solution that you find easiest to interpret, and summarize the results.

(5) Model Comparison and Recommendation

- Imagine that management is interested in understanding consumer groups for future target marketing. Which clustering solutions would you present to management, and why?

Assignment Documents:

All assignment reports should conform to the standards and style of the report template provided to you. Results should be presented and discussed in an organized manner with the discussion in close proximity of the results. The report should not contain unnecessary results or information. The document should be submitted in pdf format. Name your file **Assignment8_YourLastName.pdf**

The assignment pdf file and accompanying plain text files for Python programs should be included in a zip archive using standard zip compression with the name **Assignment8_YourLastName.zip**

Here is a reasonable section outline for this assignment report.

Section 1: Review of Prior Research

Section 2: Distance Measures and Input Matrices

Section 3: Clusters of Activities

Section 4: Clusters of Demographic Groups

Section 5: Model Comparison and Recommendation