

HARISH SAIFI
GL6134 23DSMSA112

LAB 4 – CLUSTER ANALYSIS

Objective

To perform a cluster analysis of consumers based on their attitudes toward shopping, identifying distinct consumer segments according to six key attitudinal variables. This analysis aims to group consumers with similar shopping behaviors and attitudes, facilitating targeted marketing strategies by understanding patterns in consumer preferences, such as enjoyment of shopping, budget concerns, and price sensitivity.

Data Overview

The data contains feedback on shopping attitudes and behaviors across 20 individuals. Each row represents a different individual, with six variables measuring aspects like enjoyment of shopping, budget concerns, and bargain-seeking behavior. These variables are rated on a scale, capturing diverse perspectives on shopping. The dataset is structured to allow for the identification of patterns in shopping attitudes through cluster analysis, helping to reveal distinct groups or "shopping types" among the respondents. This structured format will facilitate the clustering process by highlighting similarities and differences in shopping-related responses across the cases.

Case Number	V1	V2	V3	V4	V5	V6
1	6	4	7	3	2	3
2	2	3	1	4	5	4
3	7	2	6	4	1	3
4	4	6	4	5	3	6
5	1	3	2	2	6	4
6	6	4	6	3	3	4
7	5	3	6	3	3	4
8	7	3	7	4	1	4
9	2	4	3	3	6	3
10	3	5	3	6	4	6
11	1	3	2	3	5	3
12	5	4	5	4	2	4
13	2	2	1	5	4	4
14	4	6	4	6	4	7
15	6	5	4	2	1	4
16	3	5	4	6	4	7
17	4	4	7	2	2	5
18	3	7	2	6	4	3
19	4	6	3	7	2	7
20	2	3	2	4	7	2

Case Processing Summary ^{a,b}					
Cases					
Valid		Missing		Total	
N	Percent	N	Percent	N	Percent
20	100.0	0	.0	20	100.0
a. Squared Euclidean Distance used					
b. Ward Linkage					

The analysis included all 20 cases with no missing data. Clustering used **Ward's linkage method** to group cases, minimizing variance within clusters, and **Squared Euclidean Distance** as the measure of similarity, emphasizing larger differences between cases

Ward Linkage

Agglomeration Schedule						
Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	14	16	1.000	0	0	6
2	6	7	2.000	0	0	7
3	2	13	3.500	0	0	15
4	5	11	5.000	0	0	11
5	3	8	6.500	0	0	16
6	10	14	8.167	0	1	9
7	6	12	10.500	2	0	10
8	9	20	13.000	0	0	11
9	4	10	15.583	0	6	12
10	1	6	18.500	0	7	13
11	5	9	23.000	4	8	15
12	4	19	27.750	9	0	17
13	1	17	33.100	10	0	14
14	1	15	41.333	13	0	16
15	2	5	51.833	3	11	18
16	1	3	64.500	14	5	19
17	4	18	79.667	12	0	18
18	2	4	172.667	15	17	19
19	1	2	328.600	16	18	0

Interpretation of Ward Linkage Output:

- Early Stages:** Similar cases merge (e.g., Stage 1: Cases 14 and 16 merge at a low coefficient of 1.000).
- Intermediate Stages:** Gradual increases in coefficients indicate less similar clusters combining (e.g., Stage 11: Clusters 5 and 9 merge at 23.000).
- Final Stages:** Significant jumps in coefficients (e.g., 172.667 to 328.600 in Stages 18-19) suggest distinct clusters merging, indicating possible stopping points.

Cluster Solutions:

- 4 Clusters:** Meaningful divisions with Cluster 4 as an isolated group (e.g., case 18).
- 3 Clusters:** Cluster 4 merges with Cluster 3.

- **2 Clusters:** Cluster 1 merges with Cluster 3.

Key Insight: The 4-cluster solution is likely optimal, balancing meaningful groupings and distinctiveness.

Cluster Membership

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

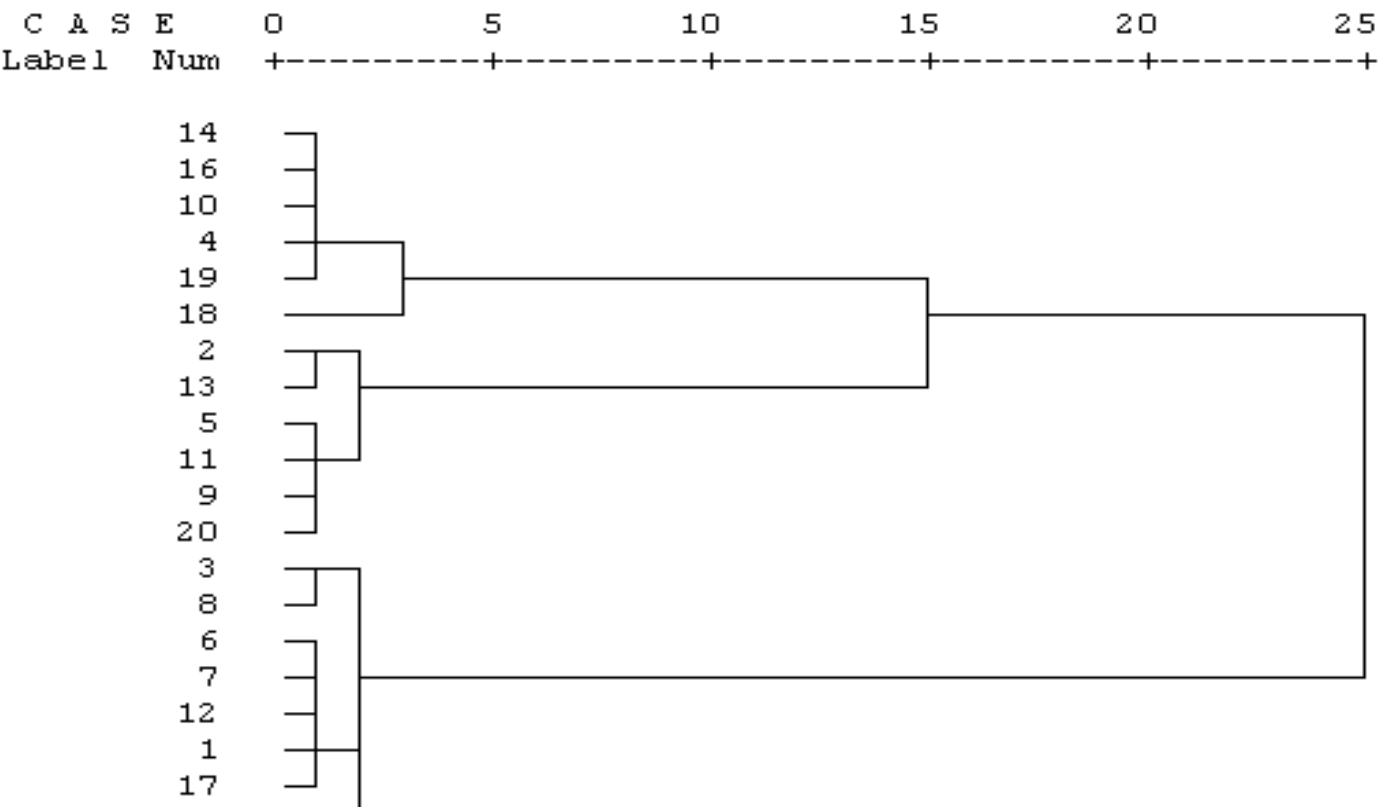
Interpretation of Cluster Membership Table:

- **4-Cluster Solution:**
 - **Cluster 1:** Cases 1, 3, 6, 7, 8, 12, 15
 - **Cluster 2:** Cases 2, 5, 9, 11, 13, 20
 - **Cluster 3:** Cases 4, 10, 14, 16, 19
 - **Cluster 4:** Case 18 (isolated)
- **3-Cluster Solution:**
 - Cluster 4 merges with Cluster 3, forming three distinct groups.
- **2-Cluster Solution:**
 - Clusters 1 and 3 merge, leaving two main clusters.

The **4-cluster solution** provides the most distinct and balanced grouping.

Dendrogram
***** H I E R A R C H I C A L C L U S T E R
A N A L Y S I S *****

Dendrogram using Ward Method
Rescaled Distance Cluster Combine



Interpretation of Dendrogram:

- **Distinct Clusters:** The dendrogram shows cases grouped based on similarity, with lower-level branches representing closer similarities.
- **Cut-off Point:** A cut around 15-20 on the horizontal axis would yield 3-4 clusters, consistent with the prior analysis.
- **Cluster Isolation:** Case 18 stands out, merging with others only at a higher distance, indicating it's distinct.

This supports a **3 or 4-cluster solution** for optimal grouping.