

# CLUSTER ANALYSIS

Seth Howells  
Concepts of Statistics II  
Week# 8 Project – Cluster Analysis  
08/29/20

## TABLE OF CONTENTS

<b><i>Overview</i></b>	<b><i>2</i></b>
<b><i>Figure 9.11</i></b>	<b><i>3</i></b>
<b><i>Figure 9.12</i></b>	<b><i>4</i></b>
<b><i>Table 9.9</i></b>	<b><i>5</i></b>

## OVERVIEW

Cluster analysis was performed on 5 metric variables on the HBAT.xls file with 4 metric dependent variables to assess the criterion validity. Below are the variables used for each corresponding FIGURE and TABLE. The figures and tables provide visual understanding of the multivariate relationships.

FIG 9.11

<u>ID</u>	<u>Variable</u>	<u>Measurement</u>	<u>Description</u>
$x_6$	Product Quality	metric	Performance
$x_8$	Technical Support	metric	Performance
$x_{12}$	Salesforce Image	metric	Performance
$x_{15}$	New Products	metric	Performance
$x_{18}$	Delivery Speed	metric	Performance

FIG 9.12

<u>ID</u>	<u>Variable</u>	<u>Measurement</u>	<u>Description</u>
$x_6$	Product Quality	metric	Performance
$x_8$	Technical Support	metric	Performance
$x_{12}$	Salesforce Image	metric	Performance
$x_{15}$	New Products	metric	Performance
$x_{18}$	Delivery Speed	metric	Performance

TABLE 9.9

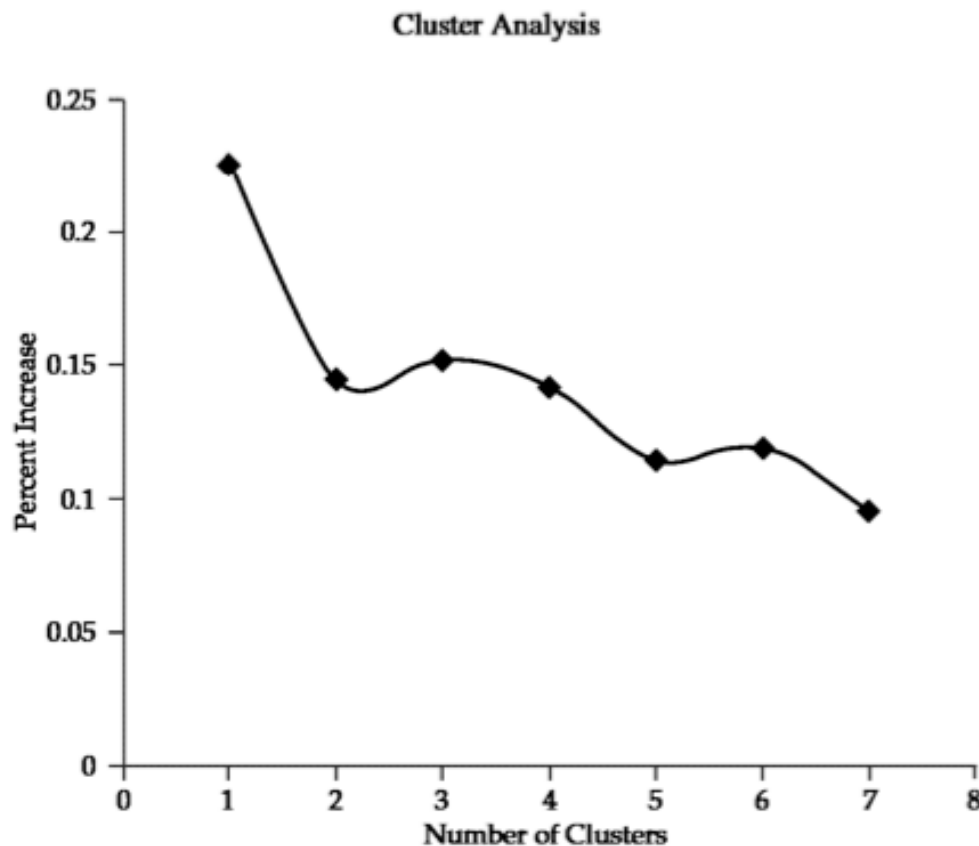
<u>ID</u>	<u>Variable</u>	<u>Measurement</u>	<u>Description</u>	<u>Type</u>
$x_{19}$	Satisfaction	metric	Relationship	Dependent
$x_{20}$	Likelihood of Recommendation	metric	Relationship	Dependent
$x_{21}$	Likelihood of Future Purchase	metric	Relationship	Dependent
$x_{22}$	Current Purchase / Usage Level	metric	Relationship	Dependent

**FIG 9.11**

In a previous step, the hierarchical Ward's agglomerative method was performed on the variables under investigation. The agglomeration coefficient was useful to find the best number of clusters. The smaller coefficients indicated more homogenous clusters were emerging, while larger coefficients indicated that the clusters lacked homogeneity.

Another useful method is to assess the percent increase in homogeneity, where each stage is assessed. The table found that at stage 96, homogeneity reached 22.6% while stage 97 resulted in 1 cluster with 100% homogeneity. Although a higher percentage may indicate that the observations in the cluster are similar, one cluster is meaningless in the analysis because the researcher is seeking to examine the differences in clusters which is not possible with only 1 cluster.

FIG 9.11 is a graph that displays the percent of homogeneity increase on the y-axis and the number of clusters on the x-axis. This figure allows a quick, visual understanding of the relationship between the two assessments. 3 to 5 clusters seem to be the most likely to be used with precautions over the increasing coefficients. Because of this, 4 clusters will be used in further analysis.



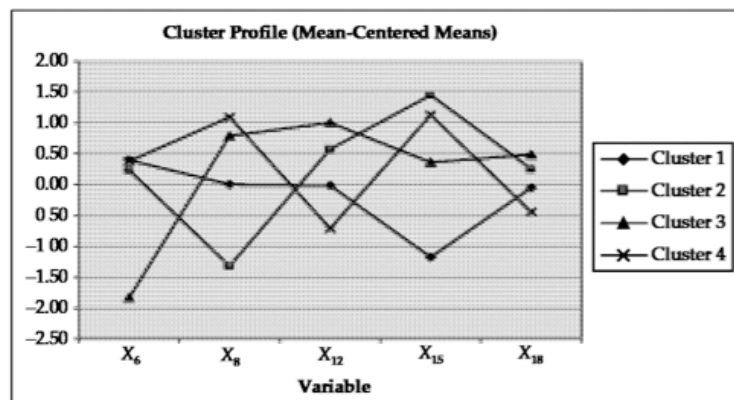
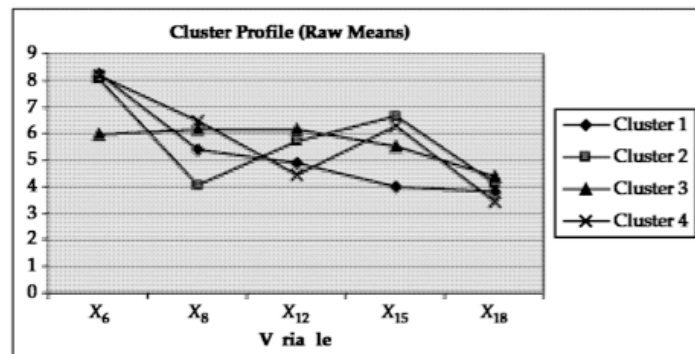
**FIGURE 11 Percent Change in Heterogeneity**

**FIG 9.12**

Hierarchical method was used in the previous figure; however, this figure approaches the cluster analysis through a non-hierarchical method. Recall that hierarchical methods are additive and as such, once observations are entered, they cannot be reentered. In contradistinction, non-hierarchical methods allow for optimization; that is, optimizing has an advantage over the hierarchical method's inability to reassign.

FIG 9.12 utilizes the results yielded from the hierarchical Ward's method to determine the number of clusters but takes advantage of non-hierarchical methods to optimize through reassignment based on the determined clusters in the previous figure.

Means from Hierarchical Cluster Analysis										
Variable	Mean Values Cluster Number:				Mean-Centered Values Cluster Number:				F	Sig
	1	2	3	4	1	2	3	4		
X <sub>6</sub> Product Quality	8.21	8.04	5.97	8.18	0.40	0.23	-1.84	0.37	14.56	0.000
X <sub>8</sub> Technical Support	5.37	4.04	6.16	6.47	0.00	-1.33	0.78	1.09	12.64	0.000
X <sub>12</sub> Salesforce Image	4.91	5.69	6.12	4.42	-0.02	0.57	1.00	-0.72	11.80	0.005
X <sub>15</sub> New Products	3.97	6.63	5.51	6.28	-1.18	1.45	0.36	1.13	62.74	0.000
X <sub>18</sub> Delivery Speed	3.83	4.14	4.37	3.45	-0.06	0.25	0.48	-0.44	5.49	0.002
Cluster Sample Sizes	49	18	14	17	49	18	14	17		



**FIGURE 12** Profile of Four Clusters from Hierarchical Cluster Analysis

TABLE 9.9

TABLE 9.9 assesses the predictive validity onto 4 metric dependent variables with the determined clusters in previous steps as the independent variables. Because the 4 metric variables are relationship measures, the analysis is more of a theoretical approach than the previous computational understandings. A multivariate analysis of variance (MANOVA) model was performed because the dependent variables are already known to have to correlate with each other.

The *Multivariate F\** is 2.23 and the corresponding significance level is below 0.05, which suggests that the variables can be predicted. In real-life terms, this means that the model can predict which segment of the HBAT file that the customers may belong to. All the *Univariate F\** and its corresponding significance levels are statistically significant. The goal is to predict key outcomes based on customer profiles. An example of interpretation can be seen in cluster 2 where it possesses the highest cluster mean in each of the 4 metric dependent variables (relationship variables). Since the 4 variables are geared toward sales and revenue, the researcher may theoretically approach this Table with an understanding of business, while other variables used may need to be approached with a supply chain understanding (delivery speed, etc.)

**TABLE 9 Multivariate F Results Assessing Cluster Solution Criterion Validity**

Variable	Cluster Number	Cluster Mean	Multivariate $F^*$	Univariate $F^*$	Sig.
			2.23		0.01
$X_{19}$ Satisfacti n	1	6.76		5.98	0.001
	2	7.44			
	3	7.39			
	4	6.34			
$X_{20}$ Likely to Recommend	1	6.89		3.06	0.032
	2	7.46			
	3	7.14			
	4	6.68			
$X_{21}$ Likely to Purchase	1	7.74		3.53	0.018
	2	8.09			
	3	7.83			
	4	7.33			
$X_{22}$ Purchase Level	1	58.70		6.21	0.001
	2	62.17			
	3	60.92			
	4	53.17			

\*Multivariate  $F$  has 12,241 degrees of freedom and univariate  $F$ s each have 3,94 degrees of freedom.