

Included Analyses

- [K-Means Cluster Analysis for 9 Variables](#)

Results

K-Means Cluster Analysis

Introduction

A k-means cluster analysis was conducted to find similar groupings among the data.

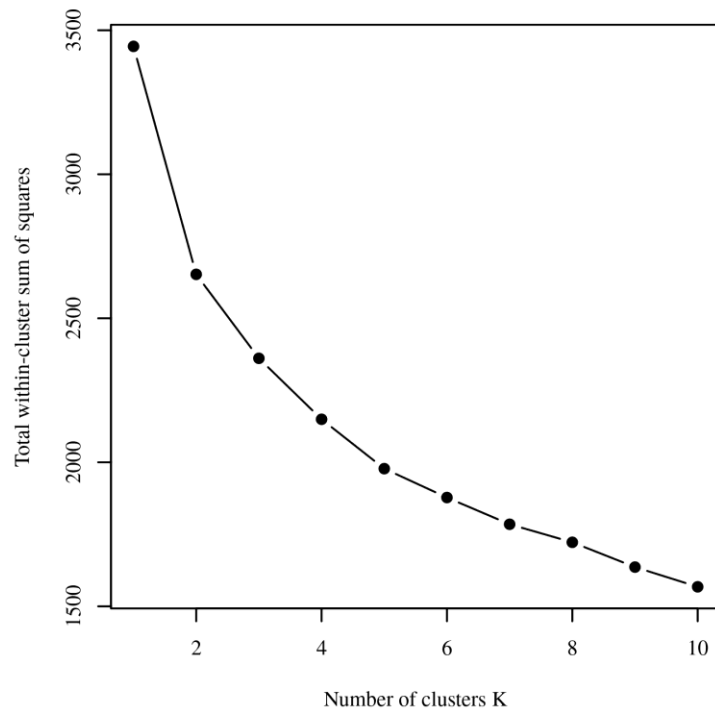
Results

First, the number of clusters, k , was chosen through graphical assessment. Next, the k-means cluster analysis was conducted and assessed through descriptive statistics and linear discriminant analysis (LDA).

Number of Clusters. The optimal number of clusters was chosen by plotting the total within-cluster sum of squares as a function of the number of clusters (Larson et al., 2005). This plot is presented in Figure 1. The optimal number of clusters is the point on the graph where the curve appears to flatten, indicating that additional clusters would have little effect on the total within-cluster sum of squares. For this analysis, 3 clusters were chosen.

Figure 1

Total Within-Clusters Sums of Squares for Number of Clusters K



Cluster Groups. The k-means cluster analysis was conducted with 3 clusters. For visual_1, Cluster 1 had a mean of 4.36 (SD = 1.09), Cluster 2 had a mean of 5.75 (SD = 0.96), and Cluster 3 had a mean of 4.78 (SD = 0.97). For visual_2, Cluster 1 had a mean of 5.75 (SD = 1.03), Cluster 2 had a mean of 6.75 (SD = 1.21), and Cluster 3 had a mean of 5.81 (SD = 1.03). For visual_3, Cluster 1 had a mean of 1.79 (SD = 1.00), Cluster 2 had a mean of 2.93 (SD = 0.99), and Cluster 3 had a mean of 2.11 (SD = 1.09). For textual_1, Cluster 1 had a mean of 2.43 (SD = 0.93), Cluster 2 had a mean of 4.15 (SD = 0.94), and Cluster 3 had a mean of 2.70 (SD = 0.78). For textual_2, Cluster 1 had a mean of 3.58 (SD = 1.14), Cluster 2 had a mean of 5.47 (SD = 0.77), and Cluster 3 had a mean of 4.09 (SD = 1.06). For textual_3, Cluster 1 had a mean of 1.61 (SD = 0.72), Cluster 2 had a mean of 3.24 (SD = 1.04), and Cluster 3 had a mean of 1.79 (SD = 0.68). For speed_1, Cluster 1 had a mean of 3.55 (SD = 0.84), Cluster 2 had a mean of 4.20 (SD = 1.01), and Cluster 3 had a mean of 4.94 (SD = 0.95). For speed_2, Cluster 1 had a mean of 4.87 (SD = 0.74), Cluster 2 had a mean of 5.62 (SD = 0.89), and Cluster 3 had a mean of 6.21 (SD = 0.92). For speed_3, Cluster 1 had a mean of 4.61 (SD = 0.74), Cluster 2 had a

mean of 5.76 (SD = 0.87), and Cluster 3 had a mean of 5.89 (SD = 0.87). For each cluster group, means (M), standard deviations (SD), and sample size (n) are shown in Table 1.

Table 1

Summary Statistics Table for Each Variable Grouped by Cluster

Variable	M	SD	n
visual_1			
Cluster 1	4.36	1.09	112
Cluster 2	5.75	0.96	96
Cluster 3	4.78	0.97	93
visual_2			
Cluster 1	5.75	1.03	112
Cluster 2	6.75	1.21	96
Cluster 3	5.81	1.03	93
visual_3			
Cluster 1	1.79	1.00	112
Cluster 2	2.93	0.99	96
Cluster 3	2.11	1.09	93
textual_1			
Cluster 1	2.43	0.93	112
Cluster 2	4.15	0.94	96
Cluster 3	2.70	0.78	93
textual_2			
Cluster 1	3.58	1.14	112
Cluster 2	5.47	0.77	96
Cluster 3	4.09	1.06	93
textual_3			
Cluster 1	1.61	0.72	112
Cluster 2	3.24	1.04	96
Cluster 3	1.79	0.68	93
speed_1			
Cluster 1	3.55	0.84	112
Cluster 2	4.20	1.01	96
Cluster 3	4.94	0.95	93
speed_2			
Cluster 1	4.87	0.74	112
Cluster 2	5.62	0.89	96

Cluster 3	6.21	0.92	93
speed_3			
Cluster 1	4.61	0.74	112
Cluster 2	5.76	0.87	96
Cluster 3	5.89	0.87	93

Note. '-' denotes the sample size is too small to calculate statistic.

Linear Discriminant Analysis. A LDA was conducted on the data by the clusters to determine if the clusters show separation in the data (Jombert et al., 2010). The LDA calculates components which are linear combinations of the data which exhibit the most separation in the data. The percentage of trace table is presented in Table 2. This identifies roughly what percentage of total separation occurs for each component of the analysis.

Table 2

Percentage of Trace Table for Each Linear Discriminant Component

LDA Component	Percentage of Trace
LDA1	68.05%
LDA2	31.95%

LDA Components by Clusters. For component 1 (LD1), Cluster 1 had a mean of -1.31 (SD = 1.04), Cluster 2 had a mean of 1.91 (SD = 1.07), and Cluster 3 had a mean of -0.39 (SD = 0.87). For component 2 (LD2), Cluster 1 had a mean of -0.81 (SD = 0.87), Cluster 2 had a mean of -0.38 (SD = 1.12), and Cluster 3 had a mean of 1.37 (SD = 1.01). Table 3 shows the results of the LDA. Figure 2 shows the plot of the LDA for the components that explain at least 5% of the percentage of trace.

Table 3

Summary Statistics Table for Each Variable Grouped by Cluster

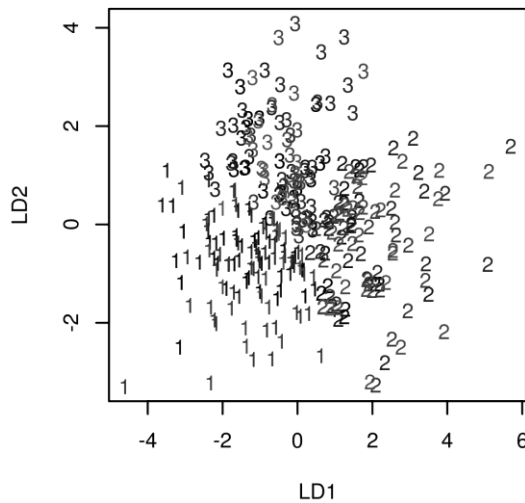
Variable	<i>M</i>	<i>SD</i>	<i>n</i>
LD1			
Cluster 1	-1.31	1.04	112
Cluster 2	1.91	1.07	96
Cluster 3	-0.39	0.87	93
LD2			
Cluster 1	-0.81	0.87	112

Cluster 2	-0.38	1.12	96
Cluster 3	1.37	1.01	93

Note. '-' denotes the sample size is too small to calculate statistic.

Figure 2

Plot of LD Components by Cluster Groups



LDA Components by Variable. Every component of the LDA has a coefficient and a correlation for each variable included in the analysis. The coefficients are the values used for the linear combinations to obtain the linear discriminant components. The coefficients for each variable and linear discriminant component can be seen in Table 4. Correlations were also calculated to determine which variables had a large ($r \geq .50$ or $r \leq -.50$) or moderate ($r \geq .30$ or $r \leq -.30$) contribution to the LDA components (Cohen, 1988). For LD1, visual_1 ($r = .62$), visual_3 ($r = .53$), textual_1 ($r = .80$), textual_2 ($r = .77$), and textual_3 ($r = .81$) had a strong contribution and visual_2 ($r = .46$) and speed_3 ($r = .49$) had a moderate influence on the component. For LD2, speed_1 ($r = .74$), speed_2 ($r = .74$), and speed_3 ($r = .63$) had a strong contribution and no variables had a moderate influence for the component. The variables that have large correlations for components that exhibit a significant percentage of trace contribute the most in separating the clusters. The Pearson correlations for each variable and linear discriminant component is shown in Table 5.

Table 4

Variable Coefficients for Each Linear Discriminant Component

Variable	LD1	LD2
visual_1	0.19	-0.09
visual_2	0.27	-0.12
visual_3	0.37	-0.13
textual_1	0.38	-0.23
textual_2	0.37	0.22
textual_3	0.40	-0.40
speed_1	0.04	0.56
speed_2	-0.01	0.55
speed_3	0.23	0.60

Table 5

Pearson Correlations Between Each Variable and Linear Discriminant Component

Variable	LD1	LD2
visual_1	.62	.01
visual_2	.46	-.13
visual_3	.53	.00
textual_1	.80	-.12
textual_2	.77	-.01
textual_3	.81	-.17
speed_1	.19	.74
speed_2	.27	.74
speed_3	.49	.63

References

Cohen, J. (1988). *Statistical power analysis for the behavior sciences* (2nd ed.). West Publishing Company.

Intellectus Statistics [Online computer software]. (2023). Intellectus Statistics.

<https://statistics.intellectus360.com>

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC genetics*, 11(1), 94. <https://doi.org/10.1186/1471-2156-11-94>

Larson, J. S., Bradlow, E. T., & Fader, P. S. (2005). An exploratory look at supermarket shopping paths. *International Journal of research in Marketing*, 22(4), 395-414. <https://doi.org/10.1016/j.ijresmar.2005.09.005>

Glossaries

K-Means Cluster Analysis

K-means cluster analysis is a multivariate statistical technique to explore and find similarities among numeric data. This analysis groups the numeric data into k clusters, where k is defined before the analysis is conducted. A cluster is a collection of data points that have similar attributes. The goal for the researcher is to find clusters that are meaningful and easy to interpret. For example, in marketing a researcher can use cluster analysis to determine different customer segments based on their needs and attitudes.

Percentage of Trace: The percentage of trace is calculated for each discriminant function from the ratio of the between-group and within-group variance. It can be roughly interpreted as the percentage of separation between the groups.

Cluster: A grouping of scale data that share similar attributes.

Linear Discriminant Analysis: A statistical method used to find linear combinations in the data that result in separating the data as much as possible.

Mean (M): The average value of a scale variable.

Sample Size (n): The frequency or count of a nominal or ordinal category.

Standard Deviation (SD): The spread of the data around the mean of a scale variable.

Within-Cluster Sum of Squares: The amount of variance inside an individual cluster, which is measured by the distance between each data point and the cluster center. The total within-cluster sum of squares is calculated by summing the within-cluster sum of squares for all clusters.

Raw Output

K-Means Cluster Analysis for 9 Variables and 3 Clusters

Summary Statistics Table for Each Variable Grouped by Cluster

Variable	M	SD	n
visual_1			
Cluster 1	4.365	1.091	112
Cluster 2	5.750	0.959	96

Cluster 3	4.783	0.972	93
visual_2			
Cluster 1	5.754	1.027	112
Cluster 2	6.747	1.211	96
Cluster 3	5.809	1.033	93
visual_3			
Cluster 1	1.786	1.000	112
Cluster 2	2.927	0.992	96
Cluster 3	2.112	1.090	93
textual_1			
Cluster 1	2.426	0.929	112
Cluster 2	4.149	0.939	96
Cluster 3	2.703	0.781	93
textual_2			
Cluster 1	3.576	1.145	112
Cluster 2	5.474	0.769	96
Cluster 3	4.091	1.060	93
textual_3			
Cluster 1	1.612	0.719	112
Cluster 2	3.237	1.039	96
Cluster 3	1.791	0.683	93
speed_1			
Cluster 1	3.547	0.836	112
Cluster 2	4.196	1.010	96
Cluster 3	4.945	0.946	93
speed_2			
Cluster 1	4.873	0.744	112
Cluster 2	5.624	0.891	96
Cluster 3	6.215	0.920	93
speed_3			
Cluster 1	4.614	0.737	112
Cluster 2	5.756	0.871	96
Cluster 3	5.895	0.868	93

Note. '-' denotes the sample size is too small to calculate statistic.

Percentage of Trace Table for Each Linear Discriminant Component

LDA Component	Percentage of Trace
LDA1	68.054%
LDA2	31.946%

Summary Statistics Table for LDA Component Grouped by Cluster

Variable	M	SD	n
LD1			
Cluster 1	-1.312	1.037	112
Cluster 2	1.912	1.068	96
Cluster 3	-0.394	0.874	93

LD2			
Cluster 1	-0.813	0.870	112
Cluster 2	-0.377	1.125	96
Cluster 3	1.369	1.010	93

Note. '-' denotes the sample size is too small to calculate statistic.

Variable Coefficients for Each Linear Discriminant Component

Variable	LD1	LD2
visual_1	0.187	-0.0869
visual_2	0.267	-0.118
visual_3	0.371	-0.128
textual_1	0.379	-0.234
textual_2	0.367	0.220
textual_3	0.397	-0.404
speed_1	0.0424	0.560
speed_2	-0.0126	0.555
speed_3	0.232	0.596

Pearson Correlations Between Each Variable and Linear Discriminant Component

Variable	LD1	LD2
visual_1	0.620	0.0138
visual_2	0.463	-0.129
visual_3	0.529	5.647×10^{-04}
textual_1	0.796	-0.122
textual_2	0.773	-0.0127
textual_3	0.806	-0.172
speed_1	0.188	0.739
speed_2	0.265	0.739
speed_3	0.488	0.629