HARISH SAIFI
GL6134
23DSMSA112

LAB 1 - PRINCIPAL COMPONENT ANALYSIS.

Objective

In this lab we are going to perform the factor reduction for the company of a diaper which has certain criteria(variables/factors) that they consider to define their product quality. For them the measures are (count, price, value, unisex, style, absorption(absorben)). These variables need to be lower down so that we can arrange 2 or more of them into simpler groups as per the relationship among them.

For that purpose we will use the Principal Component Analysis to reduce 6 variable into less variable(≤ 3)generally. Although it is not predefined to choose exactly 3 or less we will consider those variables which account for more variation explained by them.

Correlation Matrix									
	count	price	value	unisex	style	absorben	leakage	comfort	taping
count	1.000	.864	.692	.508	.459	.447	.412	.395	.321
price	.864	1.000	.730	.492	.464	.422	.387	.360	.321
value	.692	.730	1.000	.381	.346	.395	.357	.322	.248
unisex	.508	.492	.381	1.000	.894	.486	.486	.396	.324
style	.459	.464	.346	.894	1.000	.465	.460	.404	.317
absorben	.447	.422	.395	.486	.465	1.000	.942	.731	.540
leakage	.412	.387	.357	.486	.460	.942	1.000	.753	.576
comfort	.395	.360	.322	.396	.404	.731	.753	1.000	.654
taping	.321	.321	.248	.324	.317	.540	.576	.654	1.000

Correlation Matrix

- **High Correlations:** There are strong correlations between variables such as absorben and leakage (0.942), indicating that these two variables are highly related. Other high correlations include comfort and leakage (0.753) and price and value (0.730).
- **Moderate to Low Correlations:** Variables such as count and taping (0.321) show a lower correlation, suggesting weaker relationships compared to others.

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.803
Bartlett's Test of Sphericity	Approx. Chi-Square	2404.021
	df	36
	Sig.	.000

Kaiser-Meyer-Olkin Measure (KMO): The KMO value is 0.803, which is considered good and indicates that the sample size is adequate for PCA. **Bartlett's Test of Sphericity:** The chi-square statistic is significant ($p < 0.000$), which suggests that correlations between items are sufficient to proceed with PCA.

Communalities

- **Initial vs. Extraction Values:** Communalities represent the proportion of each variable's variance that is accounted for by the extracted components. Variables like unisex (0.944) and style (0.943) have high extraction values, meaning they are well represented by the components. Variables like taping (0.616) have lower communalities, indicating less variance is explained by the components.

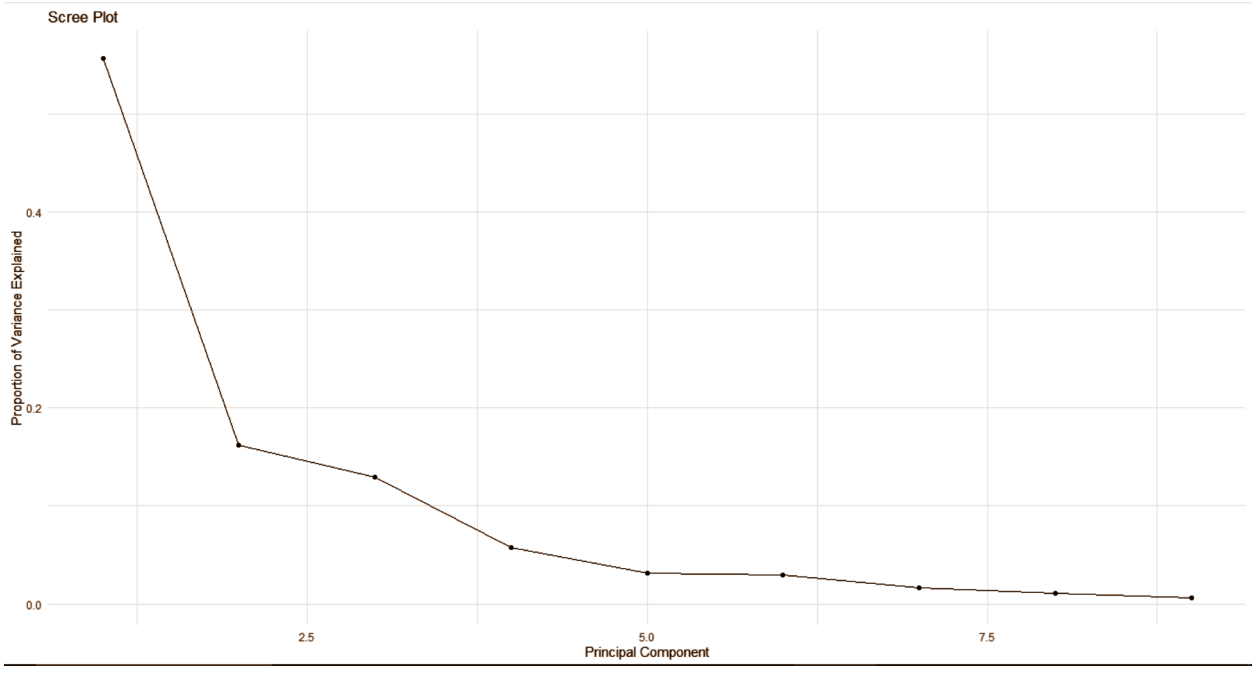
Communalities		
	Initial	Extraction
count	1.000	.862
price	1.000	.890
value	1.000	.786
unisex	1.000	.944
style	1.000	.943
absorben	1.000	.842
leakage	1.000	.871
comfort	1.000	.801
taping	1.000	.616

Total Variance Explained									
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loading		
	Total	% of Var.	Cumul. %	Total	% of Variance	Cumulative %	Total	% of Var.	Cumul. %
1	4.973	55.254	55.254	4.973	55.254	55.254	3.066	34.061	34.061
2	1.534	17.043	72.297	1.534	17.043	72.297	2.543	28.260	62.321
3	1.050	11.662	83.959	1.050	11.662	83.959	1.947	21.638	83.959

Total Variance Explained									
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loading		
4	.556	6.180	90.140						
5	.326	3.622	93.762						
6	.272	3.025	96.787						
7	.135	1.503	98.290						
8	.099	1.102	99.392						
9	.055	.608	100.000						

Total Variance Explained

- **Initial Eigenvalues:** The first component explains 55.25% of the variance, the second component explains 17.04%, and the third component explains 11.66%. Together, the first three components explain 83.96% of the variance.
- **Rotation Sums of Squared Loadings:** After rotation, the variance explained by the first three components is slightly different: 34.06%, 28.26%, and 21.64%, respectively. This suggests that the rotated components provide a clearer interpretation of the variance.



Scree Plot

From the above scree plot we can see that at the three dots(that is on the third principal component) we are getting lesser values afterward for the proportion of variance explained so we will consider only the first three principal components which are addressing about 84% variations of all remaining variables.

Component Matrix[a]

	Component		
	1	2	3
count	.761	.485	.219
price	.750	.524	.230
value	.661	.488	.334
unisex	.744	.162	-.604
style	.719	.137	-.638
absorben	.826	-.393	.082
leakage	.817	-.447	.067
comfort	.756	-.456	.147
taping	.636	-.424	.178

Extraction Method: Principal Component Analysis.[a]

a. 3 components extracted.

Rotated Component Matrix

	Component		
	1	2	3
count	.224	.865	.251
price	.193	.891	.243
value	.183	.862	.105
unisex	.244	.266	.902
style	.237	.220	.916
absorben	.850	.232	.256
leakage	.879	.182	.257
comfort	.863	.177	.157
taping	.768	.145	.079

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.[a]

a. Rotation converged in 4 iterations.

Component Matrix

- **Component Loadings:**
  - **Component 1:** Strong loadings for count, price, value, absorben, leakage, and comfort suggest that this component might represent overall product quality or functionality.
  - **Component 2:** Strong loadings for price, value, and weaker for unisex and style indicate a potential dimension related to product pricing and perceived value.
  - **Component 3:** Strong loadings for unisex and style suggest this component might represent product attributes related to gender neutrality and style.

Rotated Component Matrix

- **Component 1:** High loadings for absorben, leakage, and comfort indicate that this component is primarily associated with the performance attributes of the diaper (absorbency, leakage control, and comfort).
- **Component 2:** High loadings for price and value suggest it is related to the economic aspects of the product, such as cost and perceived value.
- **Component 3:** High loadings for unisex and style indicate that this component represents design features related to product appearance and target audience.

Summary

The PCA results suggest that the factors influencing the diaper product can be broadly categorized into three main components:

1. **Performance Attributes:** This includes variables like absorben, leakage, and comfort.
2. **Economic Factors:** This component relates to price and value.
3. **Design and Target Audience:** This involves unisex and style.

The analysis indicates that these components collectively explain a significant portion of the variance in the dataset, allowing for a clearer understanding of the underlying factors affecting the diaper product's attributes.