

## **EXEMPLE D'ANALYSE FACTORIELLE EN COMPOSANTE PRINCIPALE**

### **Introduction**

L'Analyse en Composantes Principales permet d'analyser des tableaux de données numériques quantitatives pour en réduire la dimensionnalité aux principaux facteurs d'interaction entre les variables et en représenter graphiquement les interrelations.

Toutes les colonnes d'un jeu de données initial présentent des données potentiellement intéressantes. L'ACP crée un autre jeu de données artificiel avec un nombre de dimension égal à celui du premier la seule différence est que la première dimension de l'ACP concentre la majeure partie de l'information

Dans le monde de l'ACP, la quantité d'information est appelée inertie. Les dimensions sont appelées facteurs ou axe.

L'ACP permet de ressortir une quantité d'information intéressante à partir d'un jeu de données qui peut sembler complexe.

### **LE CAS ETUDIE**

On souhaite classifier le comportement de consommateurs en fonction de leurs attitudes vis à vis des dentifrices. Dans le cadre d'un pré-test, une étude quantitative a été réalisée sur un échantillon de 20 personnes.

Les consommateurs ont exprimé leur degré d'accord avec les affirmations suivantes sur une échelle en 7 points (1 = n'est pas du tout d'accord – 7 = est d'accord) :

- V1 = Il est important d'utiliser un dentifrice qui prévient la formation des caries
- V2 = Un dentifrice doit rendre les dents brillantes
- V3 = Un dentifrice doit renforcer les gencives
- V4 = Un dentifrice doit rafraîchir l'haleine
- V5 = La prévention contre les caries n'est pas un avantage important du dentifrice
- V6 = Un dentifrice doit, avant tout, donner de belles dents

Une fois les données recueillies, il vous est demandé de réaliser une analyse typologique afin de distinguer des groupes d'individus puis de décrire les caractéristiques des groupes obtenus.

Pour cela, les données recueillies seront traitées à partir du logiciel SPSS.

## Les résultats de l'ACP

SPSS affiche ensuite la matrice des corrélations ainsi que les coefficients de signification.

Une corrélation mesure l'intensité d'une relation entre deux variables. Une corrélation peut prendre une valeur située entre -1 et 1. Une valeur proche de 1 indique que les variables sont fortement liées et évoluent dans le même sens. Une corrélation de 0 indique une absence de lien. Si plusieurs variables sont corrélées ( $> 0.5$ ), la factorisation est possible. Si non, la factorisation n'a pas de sens et n'est donc pas conseillée.

**Matrice de corrélation**

	Il est important d'utiliser un dentifrice qui prévient la formation des caries	Un dentifrice doit rendre les dents brillantes	Un dentifrice doit renforcer les gencives	Un dentifrice doit rafraîchir l'haleine	La prévention contre les caries n'est pas un avantage important du dentifrice	Un dentifrice doit, avant tout, donner de belles dents
Corrélation	1,000	-,053	,873	-,086	-,858	,004
Il est important d'utiliser un dentifrice qui prévient la formation des caries						
Un dentifrice doit rendre les dents brillantes	-,053	1,000	-,155	,572	,020	,640
Un dentifrice doit renforcer les gencives	,873	-,155	1,000	-,248	-,778	-,018
Un dentifrice doit rafraîchir l'haleine	-,086	,572	-,248	1,000	-,007	,640
La prévention contre les caries n'est pas un avantage important du dentifrice	-,858	,020	-,778	-,007	1,000	-,136
Un dentifrice doit, avant tout, donner de belles dents	,004	,640	-,018	,640	-,136	1,000

Dans notre exemple, plusieurs variables sont corrélées entre elles :

La variable V1 (Il est important d'utiliser un dentifrice qui prévient la formation des caries) est fortement corrélée à la variable V3 (Un dentifrice doit renforcer les gencives), avec une valeur de coefficient de corrélation linéaire de 0,873. En revanche, cette même variable 1 est négativement corrélée à la variable V5 (La prévention contre les caries n'est pas un avantage important du dentifrice), avec un coefficient de corrélation linéaire de -0,858.

La variable V2 (Un dentifrice doit rendre les dents brillantes) est significativement corrélée à la variable V4 (Un dentifrice doit rafraîchir l'haleine), avec une valeur de coefficient de corrélation linéaire de 0,572 ainsi qu'à la variable V6 (Un dentifrice doit, avant tout, donner de belles dents) avec une valeur de coefficient de corrélation linéaire de 0,640.

La matrice des coefficients de corrélations montre aussi que les variables 1, 3 et 5 ne sont pas corrélées aux variables 2, 4 et 5.

Comme plusieurs variables sont corrélées ( $> 0.5$ ), la factorisation est possible. Les résultats suivants peuvent donc être analysés.

## L'indice KMO et le test de sphéricité de Bartlett

Dans un deuxième temps, il faut observer l'indice de KMO (Kaiser-Meyer-Olkin) qui doit tendre vers 1. si ce n'est pas le cas, la factorisation n'est pas conseillée. Pour juger de l'indice de KMO, on peut utiliser l'échelle suivante :

- 0,50 et moins est misérable
- entre 0,60 et 0,70, c'est médiocre
- entre 0,70 et 0,80 c'est moyen
- entre 0,80 et 0,90 c'est méritoire
- et plus 0,9 c'est merveilleux.

Enfin, on utilise le test de sphéricité de Bartlett. : si la signification (Sig.) tend vers 0.000, c'est très significatif, inférieur à 0.05 significatif, entre 0.05 et 0.10 acceptable et au dessus de 0.10, on rejette. Si l'ACP satisfait à ces trois conditions, on peut continuer l'analyse des résultats.

Mesure de précision de l'échantillonnage de Kaiser-Meyer- Olkin.		,660
Test de sphéricité de Bartlett	Khi-deux approximé	111,314
	ddl	15
	Signification de Bartlett	,000

Dans le cas de notre analyse, la valeur de l'indice KMO est de 0,660. Même si on peut considérer cette valeur comme « médiocre », elle est supérieure au seuil de 0,5.

De même, la significativité du test de sphéricité de Bartlett. (Sig.) est très inférieure au seuil de 0,500 avec une valeur de 0.000.

Les deux conditions de seuils des valeurs des test de Kaiser-Meyer- Olkin et de Bartlett satisfont aux conditions. La factorisation est applicable aux données analysées et l'analyse des résultats peut être poursuivie.

## Qualité de représentation

Pour choisir les variables à éliminer, on observe leur qualité de représentation : plus la valeur associée à la ligne « Extraction » est faible, moins la variable explique la variance.

Si la valeur d' « extraction » associée à une variable est inférieure à 0,5, il est recommandé de procéder à une nouvelle ACP en retirant la variable correspondante.

	Initial	Extraction
Il est important d'utiliser un dentifrice qui prévient la formation des caries	1,000	,926
Un dentifrice doit rendre les dents brillantes	1,000	,723
Un dentifrice doit renforcer les gencives	1,000	,894
Un dentifrice doit rafraîchir l'haleine	1,000	,739
La prévention contre les caries n'est pas un avantage important du dentifrice	1,000	,878
Un dentifrice doit, avant tout, donner de belles dents	1,000	,790

Méthode d'extraction : Analyse en composantes principales.

Toutes les variables analysées présentent des valeurs d' « extraction » supérieures au seuil de 0,5. Chacune d'entre elles contribue fortement à la formation des composantes extraites de l'analyse.

## Combien d'axes retenir

Trois règles sont applicables :

- **1<sup>ère</sup> règle** : la règle de Kaiser qui veut qu'on ne retienne que les facteurs aux valeurs propres supérieures à 1.
- **2<sup>ème</sup> règle** : on choisit le nombre d'axe en fonction de la restitution minimale d'information que l'on souhaite. Par exemple, on veut que le modèle restitue au moins 80% de l'information.

Pour ces deux premières règles, on examine le tableau « Variance Totale Expliquée ».

Dans notre cas, 2 composantes (axes) ont une valeur propre supérieure à 1 pour une variance totale expliquée de 82,488%.

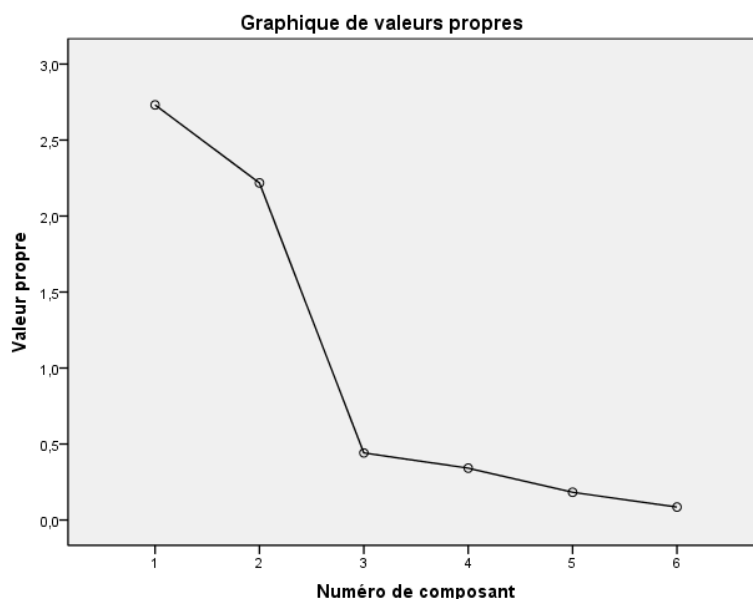
Variance totale expliquée

Composante	Valeurs propres initiales			Extraction Sommes des carrés des facteurs		
	Total	% de la variance	% cumulés	Total	% de la variance	% cumulés
1	2,731	4	45,520	2,731	45,520	45,520
2	2,218	3	82,488	2,218	36,969	82,488
3	,442	7,360	89,848			
4	,341	5,688	95,536			
5	,183	3,044	98,580			
6	,085	1,420	100,000			

Méthode d'extraction : Analyse en composantes principales.

**3<sup>ème</sup> méthode** : le « Scree-test » ou test du coude. On observe le graphique des valeurs propres et on ne retient que les valeurs qui se trouvent à gauche du point d'inflexion. Graphiquement, on part des composants qui apportent le moins d'information (qui se trouvent à droite), on relie par une droite les points presque alignés et on ne retient que les axes qui sont au dessus de cette ligne.

Ici, le point d'inflexion est entre les composantes 2 et 3. On retient donc 2 axes.



## Matrice des composantes

Ici, l'objectif consiste à donner un sens à un axe grâce à une recherche lexicale (ou recherche de mots) à partir des coordonnées des variables et des individus. Ce sont les éléments extrêmes qui concourent à l'élaboration des axes.

**Matrice des composantes<sup>a</sup>**

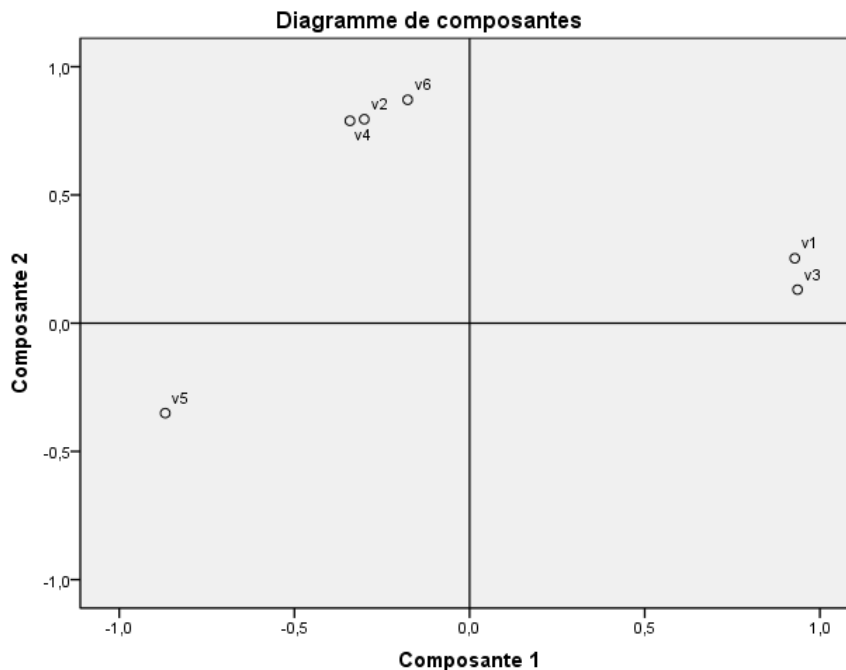
	Composante	
	1	2
Il est important d'utiliser un dentifrice qui prévient la formation des caries	,928	,253
Un dentifrice doit rendre les dents brillantes	-,301	,795
Un dentifrice doit renforcer les gencives	,936	,131
Un dentifrice doit rafraîchir l'haleine	-,342	,789
La prévention contre les caries n'est pas un avantage important du dentifrice	-,869	-,351
Un dentifrice doit, avant tout, donner de belles dents	-,177	,871

a. 2 composantes extraites.

Dans notre exemple :

- Les variables V1 (Il est important d'utiliser un dentifrice qui prévient la formation des caries) et V3 (Un dentifrice doit renforcer les gencives) sont celles qui concourent le plus à la formation de l'axe 1 pour sa portion positive et la variable V5 (La prévention contre les caries n'est pas un avantage important du dentifrice) pour sa partie négative.
- Les variables V2 (Un dentifrice doit rendre les dents brillantes), V4 (Un dentifrice doit rafraîchir l'haleine) et V6 (Un dentifrice doit, avant tout, donner de belles dents) contribuent à la formation de l'axe 2.

Le diagramme des composantes donne une représentation graphique des axes et des variables qui contribuent à leurs formations.



La partie positive de l'axe 1 représente l'effet protecteur du dentifrice. La partie négative de l'axe 1 représente le désintérêt pour l'effet protecteur du dentifrice.

La partie positive de l'axe 2 représente l'intérêt pour les effets esthétiques du dentifrice.

**Matrice des coefficients des coordonnées des composantes**

	Composante	
	1	2
Il est important d'utiliser un dentifrice qui prévient la formation des caries	,340	,114
Un dentifrice doit rendre les dents brillantes	-,110	,359
Un dentifrice doit renforcer les gencives	,343	,059
Un dentifrice doit rafraîchir l'haleine	-,125	,356
La prévention contre les caries n'est pas un avantage important du dentifrice	-,318	-,158
Un dentifrice doit, avant tout, donner de belles dents	-,065	,393

Méthode d'extraction : Analyse en composantes principales.

La matrice des co-variances indique que les deux composantes sont orthogonales.

**Matrice de covariance des composantes**

Composante	1	2
1	1,000	,000
2	,000	1,000

Méthode d'extraction : Analyse en composantes principales.