

# L'ACP sous SPSS

À Propos de ce document.....	1
Introduction .....	1
La démarche à suivre sous SPSS .....	2
1. « Descriptives... » .....	2
2. « Extraction... » .....	2
3. « Rotation... » .....	3
4. « Scores... » .....	3
5. « Options... » .....	4
Analyse des résultats .....	4
1. Les données sont-elles factorisables ? .....	4
2. Combien de facteurs retenir ? .....	5
3. Interprétation des résultats .....	6

## ***À Propos de ce document***

---

Ce document a été créé dans le but d'aider toute personne qui débute dans SPSS, logiciel très puissant mais très peu sympathique.

Ce document se base sur la version 11.0 Base de SPSS, en version anglaise. La plupart des exemples sont issus des dictatiers du programme SPSS en lui-même.

Toutes les remarques, tant sur le fond que sur la forme, sont les bienvenues. N'hésitez pas à me contacter à l'adresse suivante : <lemoal@lemoal.org> ou à venir visiter mon site internet : <http://www.lemoal.org/spss/>

Merci.

## ***Introduction***

---

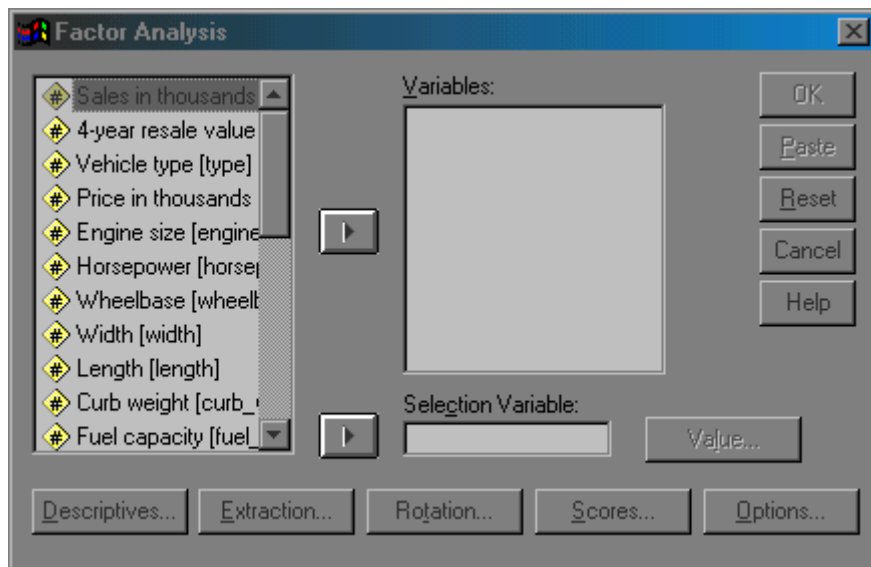
L'Analyse en Composante Principale (ACP) fait partie des analyses descriptives multivariées. Le but de cette analyse est de résumer le maximum d'informations possibles en en perdant le moins possible pour :

- Faciliter l'interprétation d'un grand nombre de données initiales
- Donner plus de sens aux données réduites

L'ACP permet donc de réduire des tableaux de grandes tailles en un petit nombre de variables (2 ou 3 généralement) tout en conservant un maximum d'information. Les variables de départ sont dites 'métriques'.

## La démarche à suivre sous SPSS

Aller dans Analyze > Data Reduction > Factor... La boîte de dialogue suivante apparaît alors :

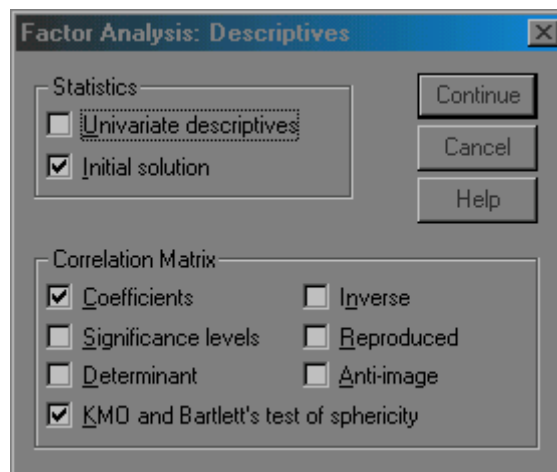


On choisit les variables qui nous paraissent les mieux adaptées à l'analyse en les sélectionnant dans la partie de droite puis en cliquant sur la flèche qui pointe vers la droite.

Cinq boîtes de dialogue d'options s'offrent maintenant à nous : 1. Descriptives... 2. Extraction... 3. Rotation... 4. Scores... 5. Options... que nous allons maintenant examiner une à une.

### **1. « Descriptives... »**

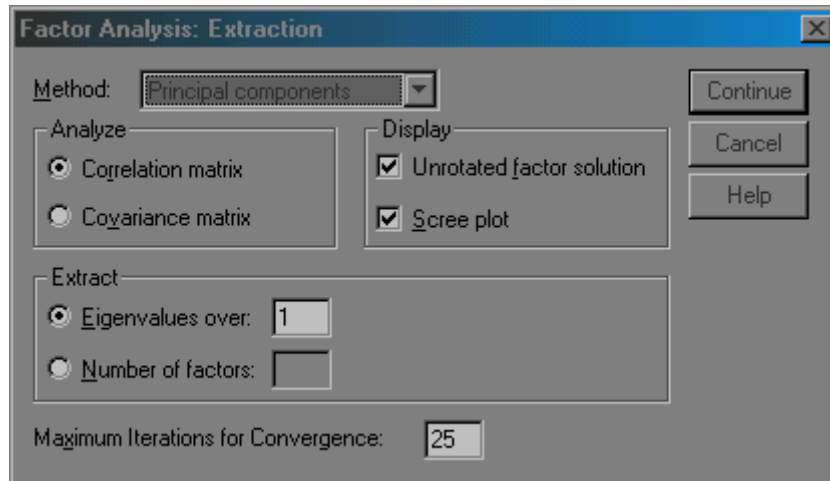
La boîte de dialogue « Factor Analysis : Descriptives » apparaît.



Dans « Correlation Matrix », cliquer sur « Coefficients » et « KMO and Bartlett's test of sphericity ».

### **2. « Extraction... »**

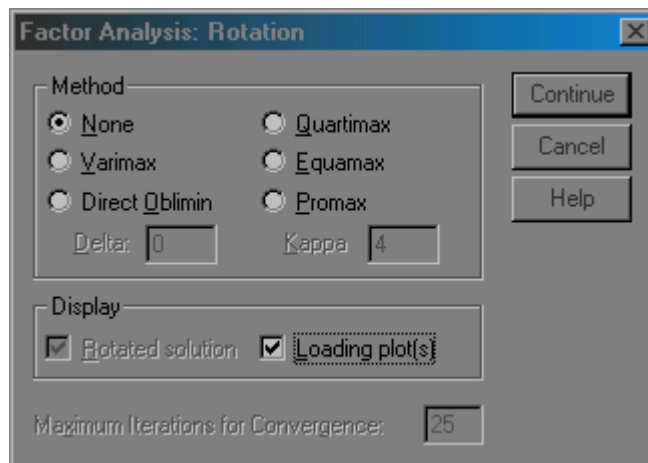
La boîte de dialogue « Factor Analysis : Extraction » apparaît.



Cliquer sur « Scree Plot » (Graphique des valeurs propres). Ne pas toucher aux autres options.

### **3. « Rotation... »**

La boîte de dialogue « Factor Analysis : Rotation » apparaît.

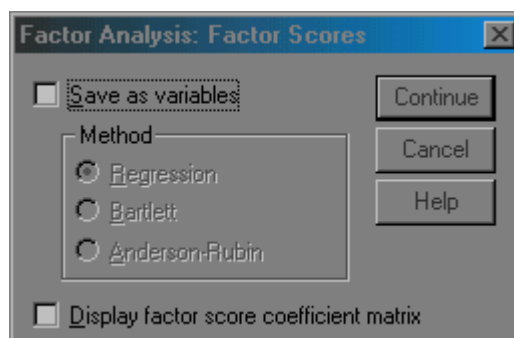


Pour l'instant, il ne faut rien toucher dans « Method ». L'option « Varimax » pourra être choisie si les résultats ne sont pas suffisants dans un premier temps.

Par contre, cocher l'option « Loading plot(s) » (Carte(s) factorielle(s)). Cette option permet d'avoir une représentation des différents axes.

### **4. « Scores... »**

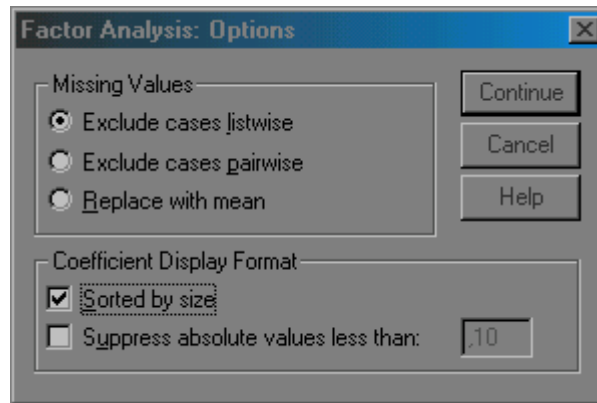
La boîte de dialogue « Factor Analysis : Factor Scores » apparaît.



Pour l'instant, il ne faut toucher à rien. L'option « Save as variables » (enregistrer dans des variables) permettra d'attribuer à chaque individu ses coordonnées factorielles une fois l'analyse terminée.

## 5. « Options... »

La boîte de dialogue « Factor Analysis : Options » apparaît.



Choisir l'option « Sorted by size » (Classement des variables par taille) dans Affichage des projections.

## Analyse des résultats

Analyser les résultats d'une ACP, c'est répondre à trois questions :

1. Les données sont-elles factorisables ?
2. Combien de facteurs retenir ?
3. Comment interpréter les résultats ?

### 1. Les données sont-elles factorisables ?

Pour répondre à cette question, dans un premier temps, il convient d'observer la matrice des corrélations (« Correlation Matrix »). 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.

Correlation Matrix

	Vehicle type	Price in thousands	Engine size	Horsepower	Wheelbase	Width	Length	Curb weight	Fuel capacity	
Correlation	Vehicle type	1,000	-,040	,268	,013	,391	,251	,141	,524	,599
	Price in thousands	-,040	1,000	,623	,838	,106	,323	,150	,526	,424
	Engine size	,268	,623	1,000	,836	,470	,688	,537	,760	,667
	Horsepower	,013	,838	,836	1,000	,283	,536	,387	,610	,504
	Wheelbase	,391	,106	,470	,283	1,000	,682	,840	,651	,654
	Width	,251	,323	,688	,536	,682	1,000	,709	,721	,658
	Length	,141	,150	,537	,387	,840	,709	1,000	,627	,565
	Curb weight	,524	,526	,760	,610	,651	,721	,627	1,000	,864
	Fuel capacity	,599	,424	,667	,504	,654	,658	,565	,864	1,000

Dans notre exemple, plusieurs variables sont corrélées entre elles :

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.

**KMO and Bartlett's Test**

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,810
Bartlett's Test of Sphericity	Approx. Chi-Square	1212,128
	df	28
	Sig.	,000

Si l'ACP satisfait à au moins deux de ces trois conditions, on peut continuer.

**2. Combien de facteurs retenir ?**

Trois règles sont applicables :

- 1ere règle : la règle de Kaiser qui veut qu'on ne retienne que les facteurs aux valeurs propres supérieures à 1.
- 2eme 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 « Total Variance Explained ».

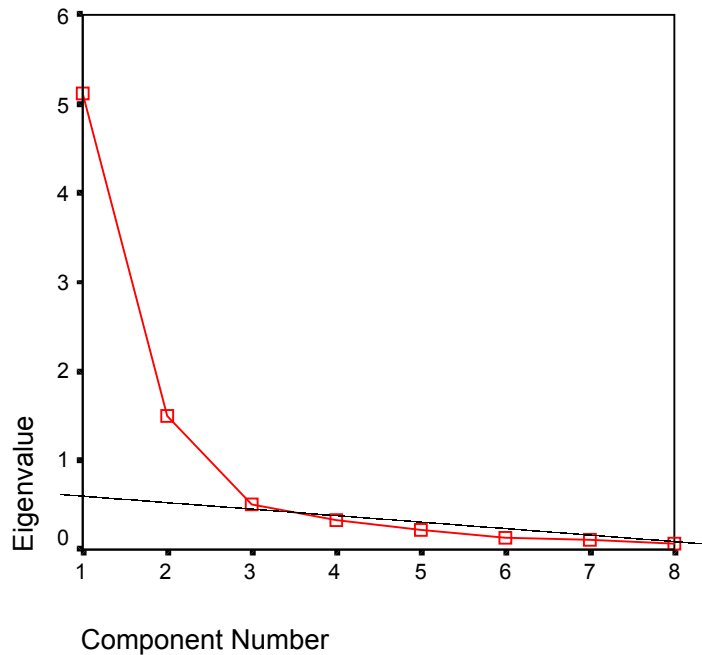
**Total Variance Explained**

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5,121	64,008	64,008	5,121	64,008	64,008
2	1,510	18,874	82,882	1,510	18,874	82,882
3	,496	6,205	89,087			
4	,328	4,100	93,187			
5	,223	2,793	95,980			
6	,141	1,757	97,736			
7	,115	1,433	99,169			
8	6,645E-02	,831	100,000			

Extraction Method: Principal Component Analysis.

- 3eme 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.

## Scree Plot



Dans notre exemple, nous ne retenons que les deux premiers axes.

### **3. Interprétation des résultats**

C'est la phase la plus délicate de l'analyse. On donne 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.

**Component Matrix<sup>a</sup>**

	Component	
	1	2
Curb weight	,912	-2,57E-02
Engine size	,878	,265
Fuel capacity	,847	-,109
Width	,843	-,221
Horsepower	,772	,554
Length	,760	-,487
Wheelbase	,742	-,569
Price in thousands	,606	,715

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

Dans notre exemple, ce sont les variables « Curb Weight » et « Engine Size » qui concourent le plus à la construction de l'axe 1. Si la recherche lexicale à partir des variables ne donne rien, il faut alors donner un sens à l'axe en s'appuyant là aussi aux individus qui ont les coordonnées extrêmes.

**Note :** Des axes peuvent très bien ne pas avoir de sens, auquel cas il ne faut pas les retenir.

Si la variance expliquée est trop faible, on peut choisir d'exclure certaines variables. 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.

### Communalities

	Initial	Extraction
Price in thousands	1,000	,879
Engine size	1,000	,841
Horsepower	1,000	,903
Wheelbase	1,000	,874
Width	1,000	,759
Length	1,000	,815
Curb weight	1,000	,832
Fuel capacity	1,000	,729

Extraction Method: Principal Component Analysis.

Il faut également tenir compte du positionnement de chaque variable sur chaque axe : les variables à éliminer sont les variables qui sont

- Soit proches du centre sur l'ensemble des axes retenus.
- Soit au milieu d'un quart de cercle sur les axes retenus.
- Soit les variables qui forment un axe à elles toute seule.