

TP-Analyse des résultats ACP

1. Statistiques descriptives

Statistiques descriptives

	Moyenne	Ecart-type	analyse
PVP	12,213	2,238	24
AGR	1,996	1,681	24
CMI	3,979	4,551	24
TRA	8,321	2,521	24
LOG	4,000	4,242	24
EDU	9,942	5,336	24
ACS	4,817	3,482	24
ANC	4,275	4,244	24
DEF	30,258	7,467	24
DET	19,142	12,456	24
DIV	1,183	1,048	24

Ce tableau donne les moyennes, les écarts-type de toutes les variables.

2. Test de Bartlett**Indice KMO et test de Bartlett(a)****Indice KMO et test de Bartlett^a**

Mesure de précision de l'échantillonnage de Kaiser-Meyer-Olkin.		,172
Test de sphéricité de Bartlett	Khi-deux approximé	317,514
	ddl	55
	Signification de Bartlett	,000

a. Basé sur les corrélations

- Le test de Sphéricité de Bartlett permet de tester l'hypothèse nulle qui est : la matrice de corrélation est une matrice identité. Le résultat du test de sphéricité de Bartlett est significatif ($p < 0,0005$). On peut donc rejeter l'hypothèse nulle ; et donc la matrice de corrélation n'est pas une matrice identité.

3. Valeur propre et variance totale expliquée

Variance totale expliquée

Composante		Valeurs propres initiales ^a			Extraction Sommes des carrés des facteurs retenus		
		Total	% de la variance	cumulés %	Total	% de la variance	% cumulés
Non normée	1	213,162	65,891	65,891	213,162	65,891	65,891
	2	58,245	18,004	83,895	58,245	18,004	83,895
	3	22,771	7,039	90,934			
	4	17,065	5,275	96,209			
	5	5,757	1,78	97,989			
	6	3,362	1,039	99,028			
	7	1,163	0,36	99,388			
	8	1,082	0,335	99,722			
	9	0,646	0,2	99,922			
	10	0,249	0,077	99,999			
	11	0,004	0,001	100			
Redimensionné	1	213,162	65,891	65,891	4,689	42,624	42,624
	2	58,245	18,004	83,895	1,462	13,29	55,914
	3	22,771	7,039	90,934			
	4	17,065	5,275	96,209			
	5	5,757	1,78	97,989			
	6	3,362	1,039	99,028			
	7	1,163	0,36	99,388			
	8	1,082	0,335	99,722			
	9	0,646	0,2	99,922			
	10	0,249	0,077	99,999			
	11	0,004	0,001	100			

Méthode d'extraction : Analyse en composantes principales.

a. Lors de l'analyse d'une matrice de covariance, les valeurs propres initiales des solutions directe et centrée sont les mêmes.

- Les valeurs propres sont données dans la colonne « **total** » ; elles représentent l'inertie portées par les axes principaux.
- L'inertie totale est égale à la somme des valeurs propres ; ainsi

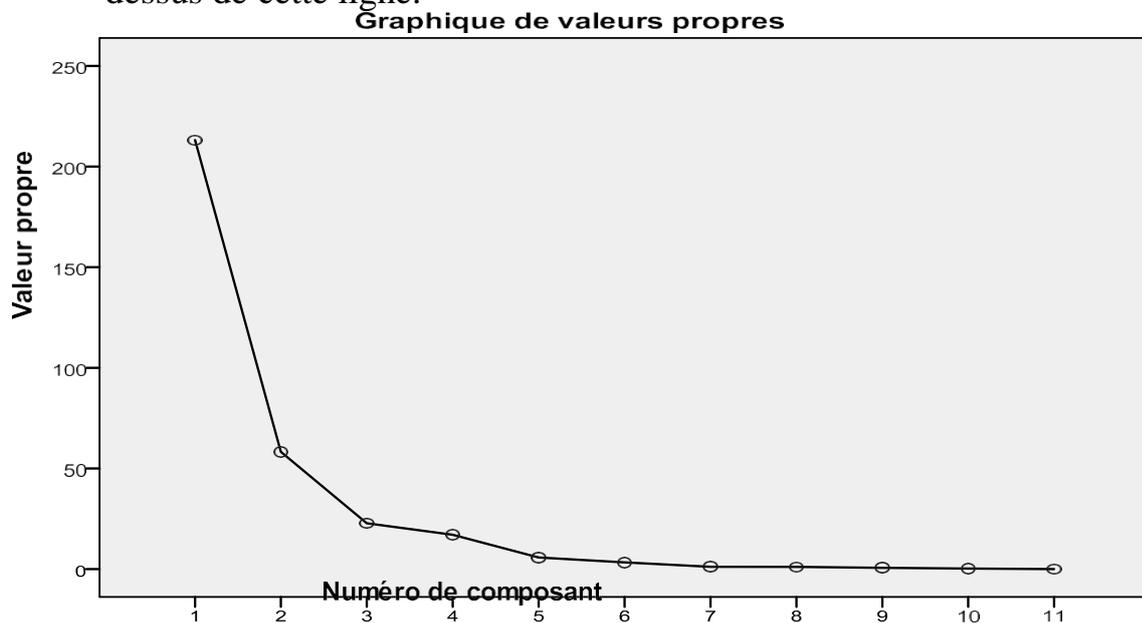
$$I=213,586+\dots+,000=324,025.$$

- La contribution de chaque axe principal à l'inertie totale est donnée par le pourcentage d'inertie, qui est le rapport entre la valeur propre et l'inertie totale, soit λ/I .
- La plus grande valeur propre 213,586 est associé à l'axe principale D1. Cet axe explique 65,917 de la variabilité. De même, la valeur propre 58,252 est associée à l'axe principal D2 qui explique 17,978 de la variance. Le plan constitué des deux axes D1 et D2 explique 83,895 de la variabilité.

4. Nombres d'axes principaux à retenir :

Pour choisir le nombre d'axes principaux à retenir, deux règles sont applicables :

- **Première 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.
- **Deuxième règle** : 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.



Dans notre exemple, on ne retient que les deux premiers axes qui permettent de prendre en compte environ 83,895 de l'inertie totale.

5. Qualité de représentation

Qualité de représentation

	Non normée		Redimensionné	
	Initial	Extraction	Initial	Extraction
PVP	5,01	0,303	1	0,061
AGR	2,827	1,648	1	0,583
CMI	20,709	15,482	1	0,748
TRA	6,355	0,822	1	0,129
LOG	17,998	10,978	1	0,61
EDU	28,469	18,029	1	0,633
ACS	12,125	9,995	1	0,824
ANC	18,013	5,65	1	0,314
DEF	55,752	53,513	1	0,96
DET	155,151	154,665	1	0,997
DIV	1,098	0,32	1	0,292

Méthode d'extraction : Analyse en composantes principales.

La Qualité de représentation permet de voir dans quelle mesure les variables initiales sont-elles prises en compte par les nouvelles variables extraites (i.e. les composantes principales).

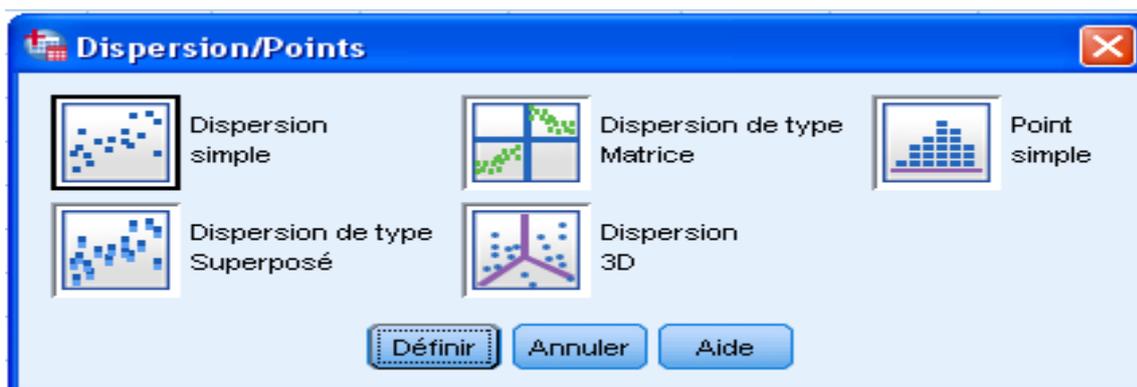
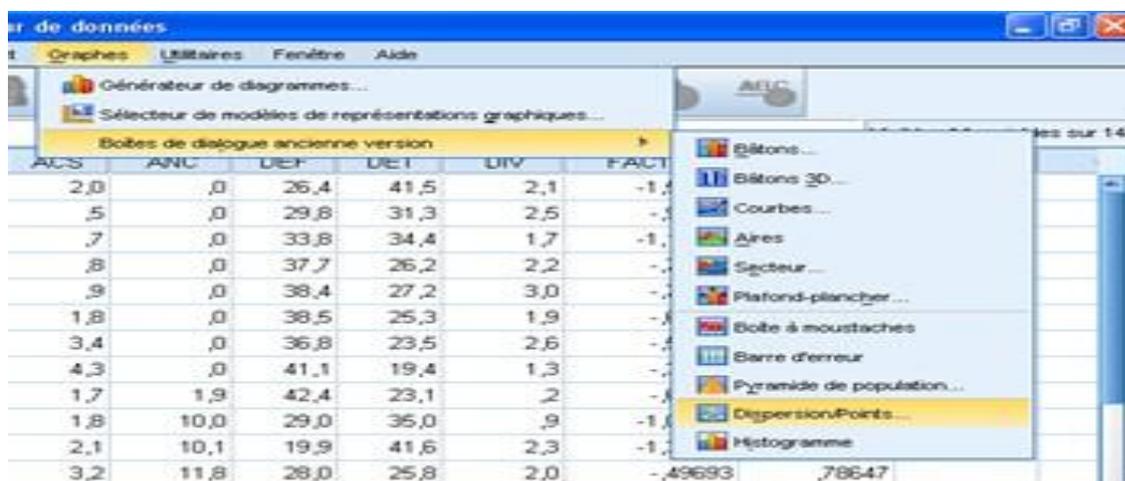
Par **exemple**, la qualité de représentation de la variable « **det** » est 0.997 ; cela veut dire que 99.7% de la variance de la variable est prise en compte par l'un des deux axes principaux. Par contre la variable « **pvp** » est mal représentée par les deux axes, et donc sera mal prise en compte par les nouvelles variables.

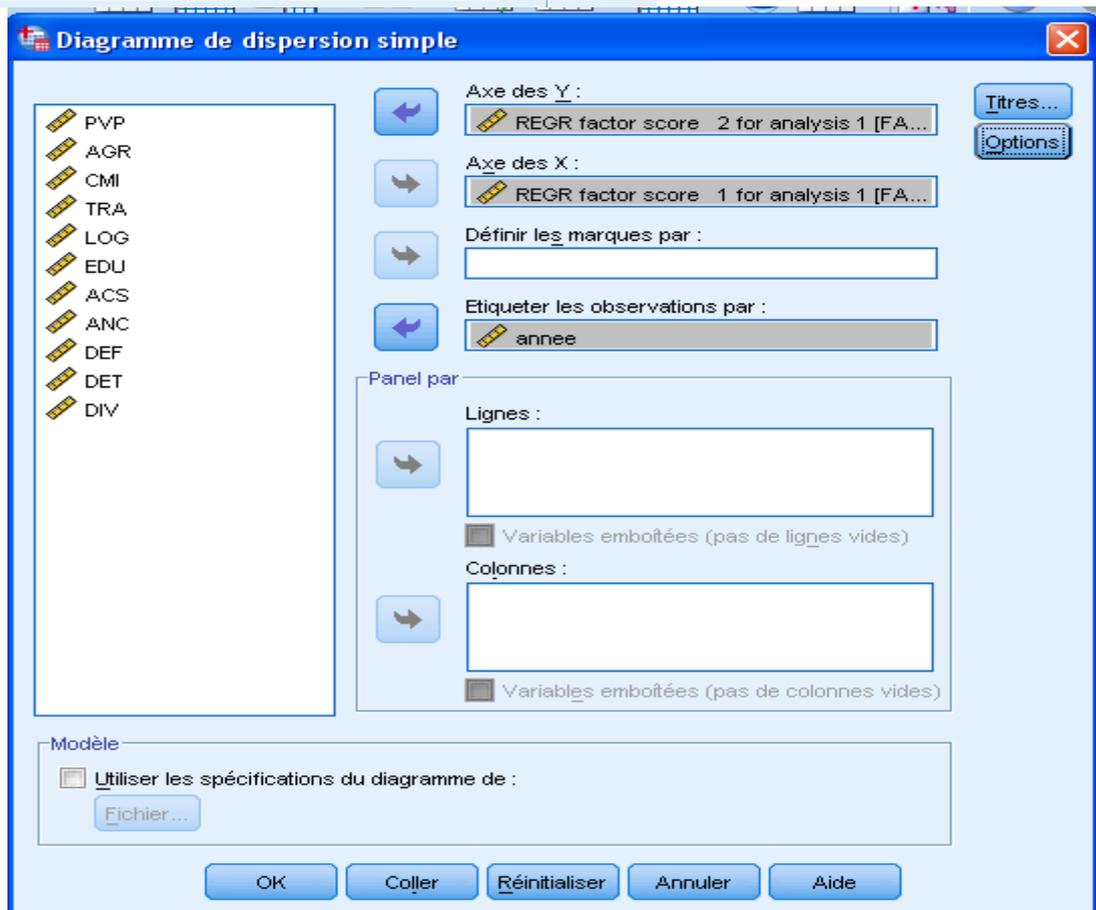
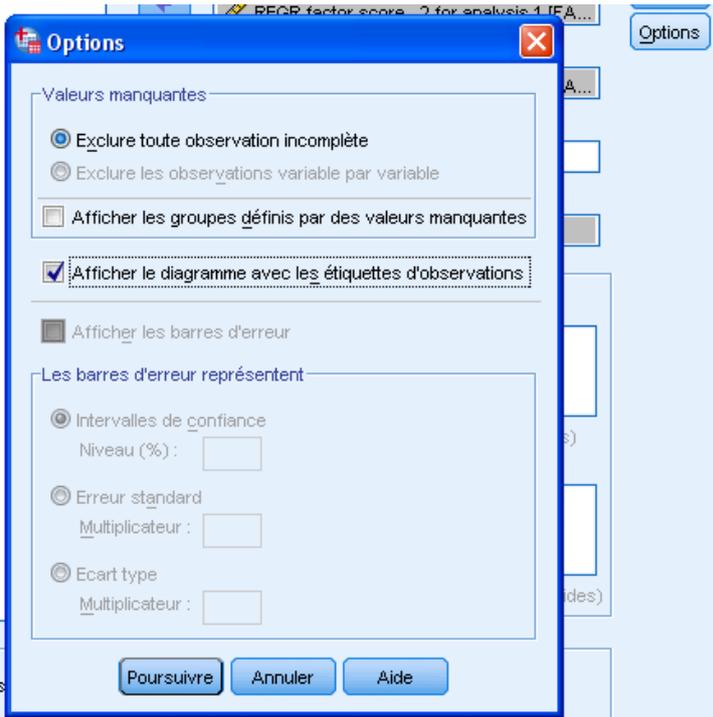
6. Coordonnées des individus dans les axes principaux :

Dans l'éditeur des données, on retrouve les coordonnées des individus dans le premier axe principal et dans le deuxième axe principal.

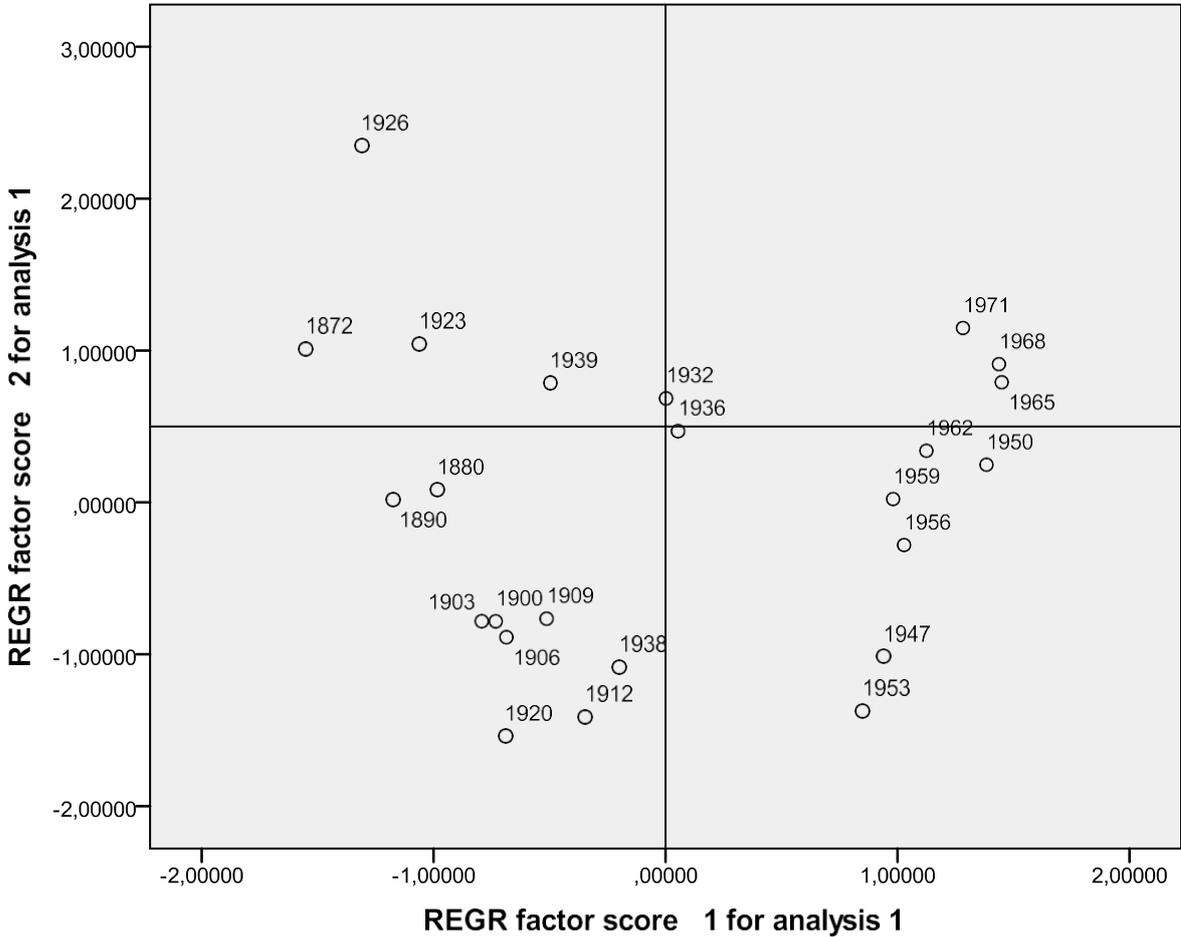
	annee	PVP	AGR	CMI	TRA	LOG	EDU	ACS	ANC	DEF	DET	DIV	FAC1_1	FAC2_1
1	1872	18,0	,5	,1	6,7	,5	2,1	2,0	,0	26,4	41,5	2,1	-1,55063	1,01329
2	1880	14,1	,8	,1	15,3	1,9	3,7	,5	,0	29,8	31,3	2,5	-,98450	,08468
3	1890	13,6	,7	,7	6,8	,6	7,1	,7	,0	33,8	34,4	1,7	-1,17421	,01662
4	1900	14,3	1,7	1,7	6,9	1,2	7,4	,8	,0	37,7	26,2	2,2	-,73179	-,78309
5	1903	10,3	1,5	1,4	9,3	,6	8,5	,9	,0	38,4	27,2	3,0	-,79146	-,78240
6	1906	13,4	1,4	,5	8,1	,7	8,6	1,8	,0	38,5	25,3	1,9	-,68606	-,88562
7	1909	13,5	1,1	,5	9,0	1,6	9,0	3,4	,0	36,8	23,5	2,6	-,51438	-,76615

- Pour la représentation graphique, on clique sur «**Graphes+Diagramme de dispersion+simple**».





- En cliquant deux fois sur le graphe et en ajoutant les deux axes des **X** et **Y** dans le graphe, on ale graphe suivant



7. Coordonnées des variables :

Les coordonnées des variables du tableau sont données par la matrice des corrélations variables/facteurs.

Matrice des composantes^a

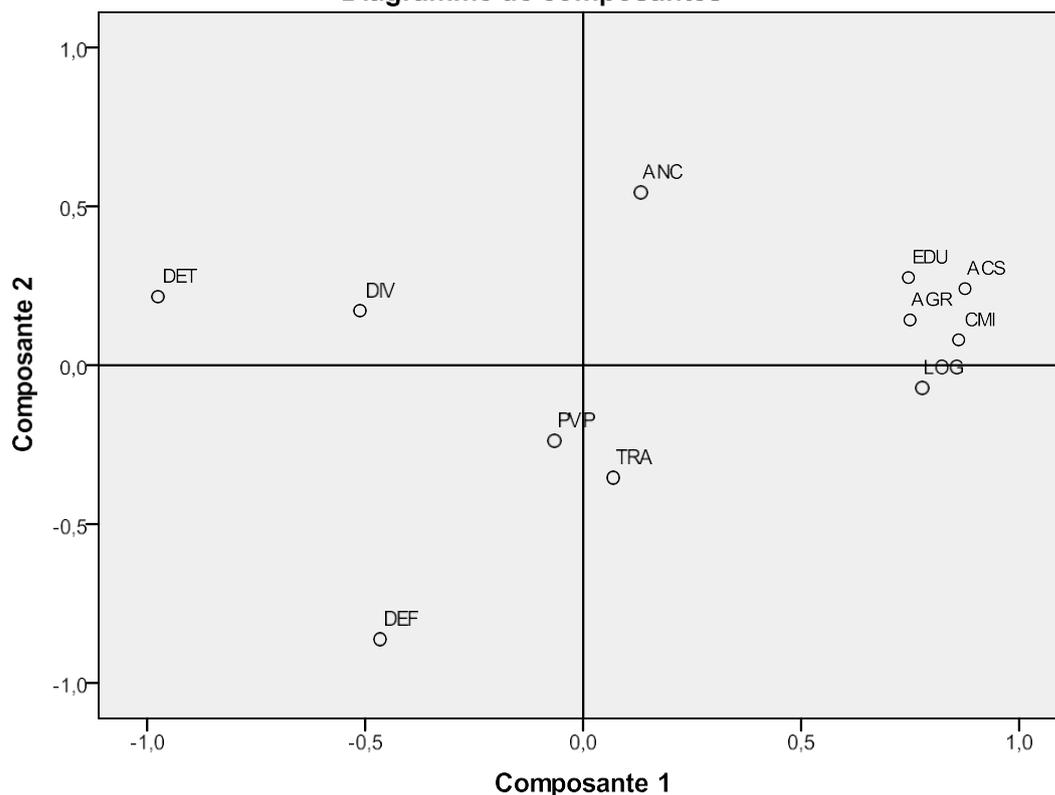
	Non normée		Redimensionné	
	Composante		Composante	
	1	2	1	2
DET	-12,142	2,690	-0,975	0,216
ACS	3,048	0,838	0,875	0,241
CMI	3,918	0,366	0,861	0,080
LOG	3,300	-0,300	0,778	-0,071
AGR	1,261	0,240	0,750	0,143
EDU	3,983	1,471	0,747	0,276
DIV	-0,537	0,180	-0,512	0,172
DEF	-3,478	-6,436	-0,466	-0,862
ANC	0,561	2,310	0,132	0,544
TRA	0,174	-0,890	0,069	-0,353
PVP	-0,147	-0,531	-0,066	-0,237

Méthode d'extraction : Analyse en composantes principales.

a. 2 composantes extraites.

Représentations graphiques

Diagramme de composantes



- Il y a de fortes corrélations entre le premier axe et la Dette (négative), le Commerce, le Logement et l'Agriculture (positives). L'axe ϕ_2 est lui très lié à la variable DEF.
- La proximité des variables EDU, ACS, AGR, CMI et LOG sur le graphe permet de conclure qu'il existe un lien fort entre ces variables mais cette interprétation ne vaut que parce ces variables sont bien représentées dans le plan 1-2 (les points correspondants sont proches du cercle). La proximité des variables PVP et TRA ne peut pas être interprétée : ces variables sont mal représentées puisque les points correspondants sont éloignés du cercle.

