

Analyse en Composantes Principales (ACP)

Application sur les données de dépenses d'un État

Jordan Matin

13 février 2026

Table des matières

1	Introduction et exploration des données	2
1.1	Contexte	2
1.2	Statistiques descriptives	2
1.3	Analyse des corrélations	3
2	Réalisation de l'ACP	3
2.1	Méthodologie	3
2.2	Valeurs propres et variance expliquée	4
2.3	Nombre d'axes retenus	4
3	Interprétation des axes retenus	5
3.1	Corrélations entre variables et axes principaux	5
3.2	Interprétation de PC1 (45.21% de variance)	5
3.3	Interprétation de PC2 (18.63% de variance)	6
3.4	Interprétation de PC3 (11.72% de variance)	7
4	Analyse des biplots	7
4.1	PC1 vs PC2	7
4.2	PC1 vs PC3	8
4.3	PC2 vs PC3	9
5	Conclusion	9

1 Introduction et exploration des données

1.1 Contexte

On considère 11 pôles de dépenses d'un État (en pourcentages du budget total) sur 24 années successives. Les variables étudiées sont les suivantes :

TABLE 1 – Pôles de dépenses étudiés

Abréviation	Pôle de dépense
PVP	Pouvoirs publics
AGR	Agriculture
CMI	Commerce et industrie
TRA	Travail
LOG	Logement et aménagement du territoire
EDU	Éducation
ACS	Action sociale
ACO	Anciens combattants
DEF	Défense
DET	Dette
DIV	Divers

L'objectif est de réaliser une **Analyse en Composantes Principales** (ACP) afin de :

- Identifier les principaux axes de variation des dépenses publiques
- Réduire la dimensionnalité des données
- Interpréter les tendances structurelles des dépenses au fil du temps

1.2 Statistiques descriptives

Nous commençons par examiner les statistiques élémentaires des données brutes :

TABLE 2 – Statistiques descriptives des données initiales

Variable	Moyenne	Écart-type	Étendue	Q25	Q75
PVP	12.21	2.19	10.4	10.575	13.425
AGR	1.99	1.64	5.7	0.800	2.650
CMI	3.93	4.48	16.4	0.400	7.350
TRA	8.32	2.46	10.8	6.675	9.150
LOG	3.95	4.18	15.3	0.675	6.200
EDU	9.94	5.22	21.7	7.325	10.600
ACS	4.81	3.40	10.8	1.800	6.800
ACO	4.27	4.15	13.4	0.000	5.450
DEF	30.25	7.30	23.6	25.925	37.025
DET	19.14	12.19	38.1	6.350	26.450
DIV	1.18	1.02	3.0	0.000	2.025

Observation clé : Deux variables se distinguent nettement — **DEF** et **DET** — tant par leur étendue que par leur écart-type élevé. Cette hétérogénéité d'échelles justifie la **standardisation** des données avant l'ACP.

1.3 Analyse des corrélations

Pour renforcer cette décision, nous examinons trois représentations des dépendances entre variables.

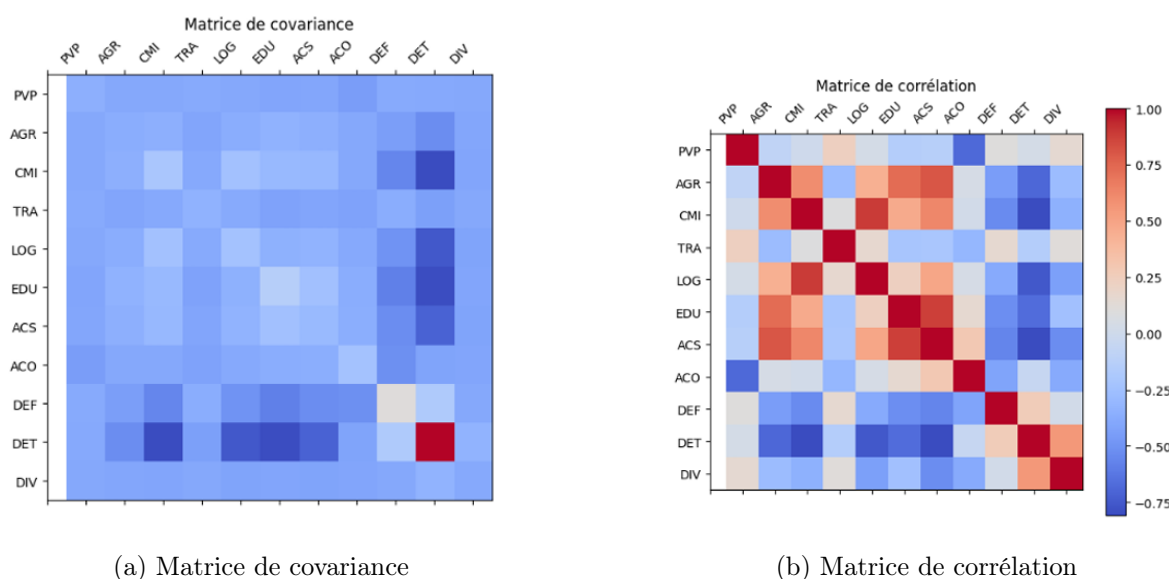


FIGURE 1 – Matrices de covariance et de corrélation

FIGURE 2 – Matrice des coefficients de corrélation (valeurs numériques)

Constats principaux :

- La matrice de covariance (Figure 1) est insuffisante à elle seule pour analyser les dépendances, car elle est dominée par les variables à forte variance (DEF, DET).
- Les matrices de corrélation (Figures 2 & 3) offrent une vision normalisée et plus riche.
- **Corrélation positive forte** entre : AGR, CMI, LOG, EDU, ACS (pôles « sociaux et économiques »).
- **Corrélation négative forte** entre ce groupe et : DET, DIV (dette et divers).
- **DEF** présente un comportement relativement indépendant.
- **ACO** est négativement corrélé à PVP et positivement corrélé à certaines dépenses sociales.

2 Réalisation de l'ACP

2.1 Méthodologie

L'ACP est réalisée sur la **matrice standardisée** (centrée-réduite) afin de neutraliser les effets d'échelle. On obtient 11 composantes principales (autant que de variables).

Pour la réalisation complète de l'ACP en Python : voir le fichier `.py` joint.

2.2 Valeurs propres et variance expliquée

TABLE 3 – Valeurs propres et pourcentages de variance expliquée

Composante	Valeur Propre	% Variance	% Cumulé
1	4.9734	45.21	45.21
2	2.0499	18.63	63.84
3	1.2899	11.72	75.57
4	0.9933	9.03	84.60
5	0.7083	6.43	91.04
6	0.5587	5.07	96.12
7	0.2042	1.85	97.97
8	0.1253	1.13	99.11
9	0.0619	0.56	99.68
10	0.0350	0.31	100.00
11	0.0000	0.00	100.00

2.3 Nombre d'axes retenus

Nous retenons **3 axes principaux**, pour les raisons suivantes :

1. **Critère de Kaiser** : Seules les 3 premières composantes possèdent une valeur propre strictement supérieure à 1 ($\lambda_1 = 4.97$, $\lambda_2 = 2.05$, $\lambda_3 = 1.29$). La quatrième ($\lambda_4 = 0.99$) est trop proche de 1 pour être retenue avec certitude.
2. **Variance cumulée** : Les 3 premiers axes expliquent **75.57%** de la variance totale, ce qui constitue un seuil acceptable pour une analyse exploratoire.
3. **Règle du coude** : Le graphique des pourcentages de variance expliquée (Figure 3) montre un coude net après la 3^{ème} composante. Au-delà, les gains marginaux deviennent faibles.

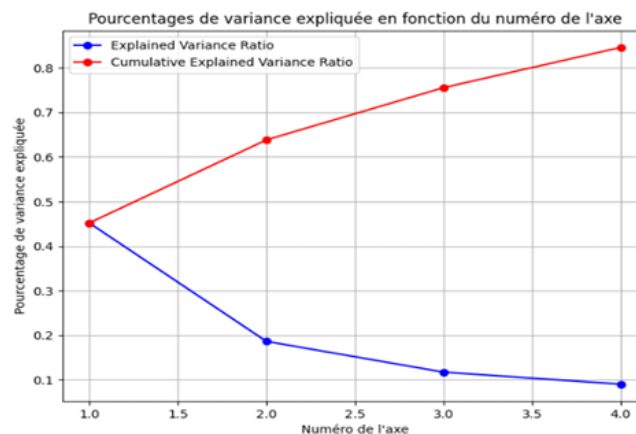


FIGURE 3 – Pourcentages de variance expliquée en fonction du numéro de l'axe

3 Interprétation des axes retenus

3.1 Corrélations entre variables et axes principaux

TABLE 4 – Corrélations entre les variables d'origine et les 3 premiers axes

Variable	PC1	PC2	PC3
PVP	−0.1735	−0.7397	+0.3414
AGR	+0.8185	−0.0059	+0.3664
CMI	+0.8335	−0.3402	−0.1408
TRA	−0.1371	−0.6305	−0.3760
LOG	+0.7216	−0.3977	−0.3851
EDU	+0.7868	+0.1365	+0.4246
ACS	+0.9332	+0.1004	+0.1661
ACO	+0.2889	+0.8075	−0.3745
DEF	−0.6123	−0.2161	−0.2599
DET	−0.8887	+0.3014	+0.1608
DIV	−0.5481	−0.1123	+0.5365

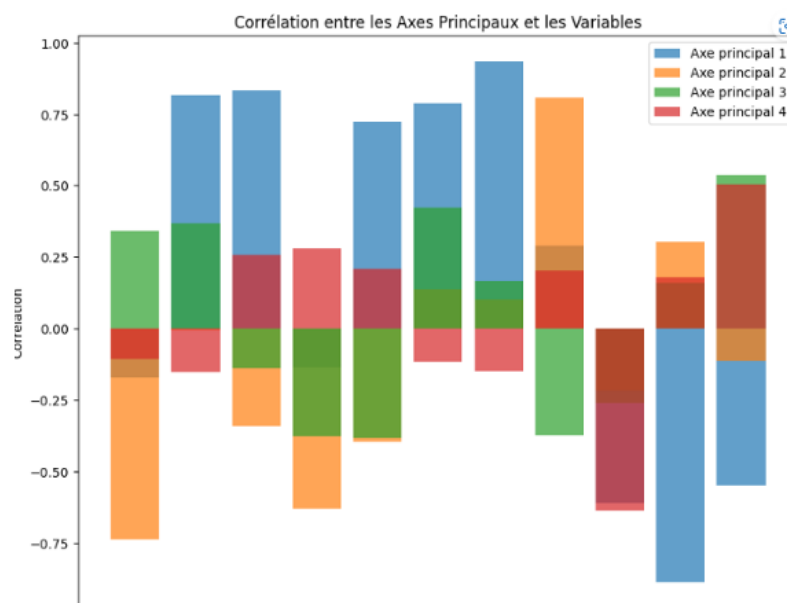


FIGURE 4 – Corrélations entre les axes principaux et les variables

3.2 Interprétation de PC1 (45.21% de variance)

Le premier axe principal est **l'axe le plus structurant** des données. Il oppose deux groupes de variables :

TABLE 5 – Structure de PC1

Côté positif (PC1 > 0)	Côté négatif (PC1 < 0)
AGR (+0.82)	DET (−0.89)
CMI (+0.83)	DEF (−0.61)
LOG (+0.72)	DIV (−0.55)
EDU (+0.79)	
ACS (+0.93)	

Interprétation : PC1 mesure la **balance entre dépenses sociales/économiques** (éducation, agriculture, action sociale, commerce, logement) d’un côté, et **dépenses militaires/dette** (défense, dette) de l’autre.

- Un score PC1 **positif** indique une année où l’État investit davantage dans les secteurs sociaux et économiques.
- Un score PC1 **négatif** indique une année dominée par les dépenses d’armement et le remboursement de la dette.

PC1 peut être interprété comme un axe de **progression lente dans le temps** : les premières années se situent du côté négatif (dette et défense élevées), puis migrent progressivement vers le côté positif (investissements sociaux croissants).

3.3 Interprétation de PC2 (18.63% de variance)

TABLE 6 – Structure de PC2

Côté positif (PC2 > 0)	Côté négatif (PC2 < 0)
ACO (+0.81)	PVP (−0.74)
DET (+0.30)	TRA (−0.63)
	CMI (−0.34)
	LOG (−0.40)

Interprétation : PC2 capture des **fluctuations relativement rapides** dans le temps, à l’inverse de PC1 qui mesure une tendance lente.

- Les observations se regroupent par **périodes de 5 ans environ** le long de cet axe, formant des « vagues migratoires » pour chaque lustre.
- **ACO** (anciens combattants) est un secteur dont le financement dépend du contexte géopolitique ; sa forte corrélation avec PC2 reflète ces changements conjoncturels.
- PC2 est positivement corrélé à ACO et modérément à DEF, et inversement corrélé à TRA et PVP.

PC2 mesure donc la **fluctuation conjoncturelle** des dépenses, notamment l’arbitrage entre dépenses liées aux conflits (anciens combattants, défense) et dépenses de fonctionnement courant (travail, pouvoirs publics).

3.4 Interprétation de PC3 (11.72% de variance)

TABLE 7 – Structure de PC3

Côté positif (PC3 > 0)	Côté négatif (PC3 < 0)
DIV (+0.54)	TRA (−0.38)
EDU (+0.42)	LOG (−0.39)
AGR (+0.37)	ACO (−0.37)
PVP (+0.34)	

Interprétation : PC3 présente une corrélation modérée avec les secteurs TRA, ACO et LOG d'un côté, et une corrélation négative avec AGR, EDU, DIV et PVP de l'autre.

- L'interprétation est plus subtile : les secteurs du domaine social semblent être **inversement corrélés** le long de cet axe.
- PC3 mesure vraisemblablement l'**équilibre des investissements au sein d'un même secteur** (social), c'est-à-dire les arbitrages internes entre postes de dépenses proches.

4 Analyse des biplots

4.1 PC1 vs PC2

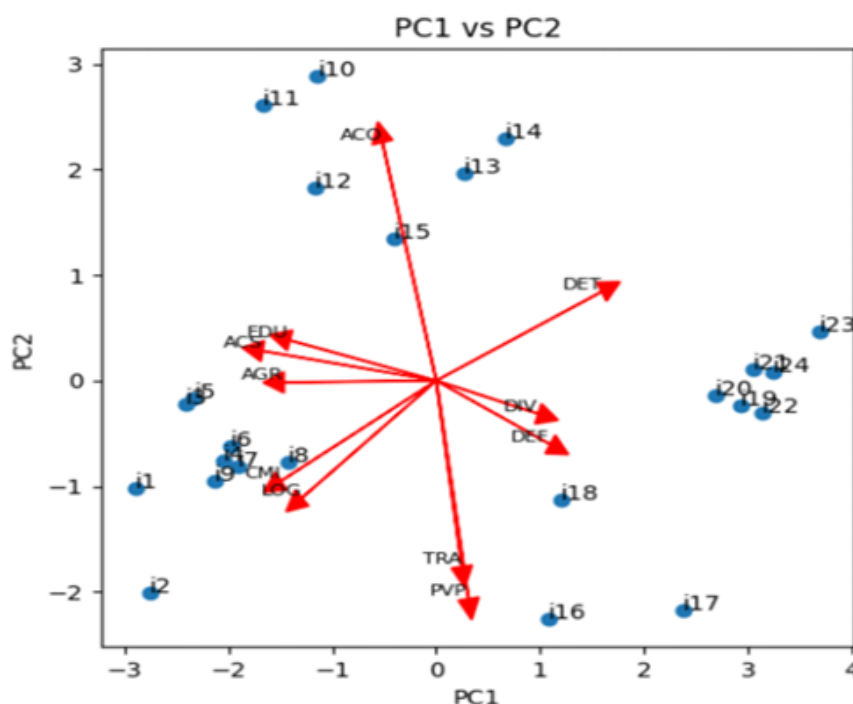


FIGURE 5 – Biplot PC1 vs PC2

Le biplot PC1 vs PC2 révèle :

- **Cluster A** = {EDU, ACS, AGR, CMI, LOG} : secteurs sociaux et économiques, fortement corrélés positivement à PC1.
- **Cluster B** = {DET, DIV, DEF} : dette et défense, inversement corrélés au Cluster A.

- Les deux clusters sont éloignés, confirmant que des dépenses accrues dans l'un entraînent une baisse dans l'autre.
- On observe une **migration temporelle** des observations : les premières années (i1–i8) sont concentrées du côté négatif de PC1, puis migrent progressivement vers le côté positif, suggérant un changement structurel des priorités budgétaires.

4.2 PC1 vs PC3

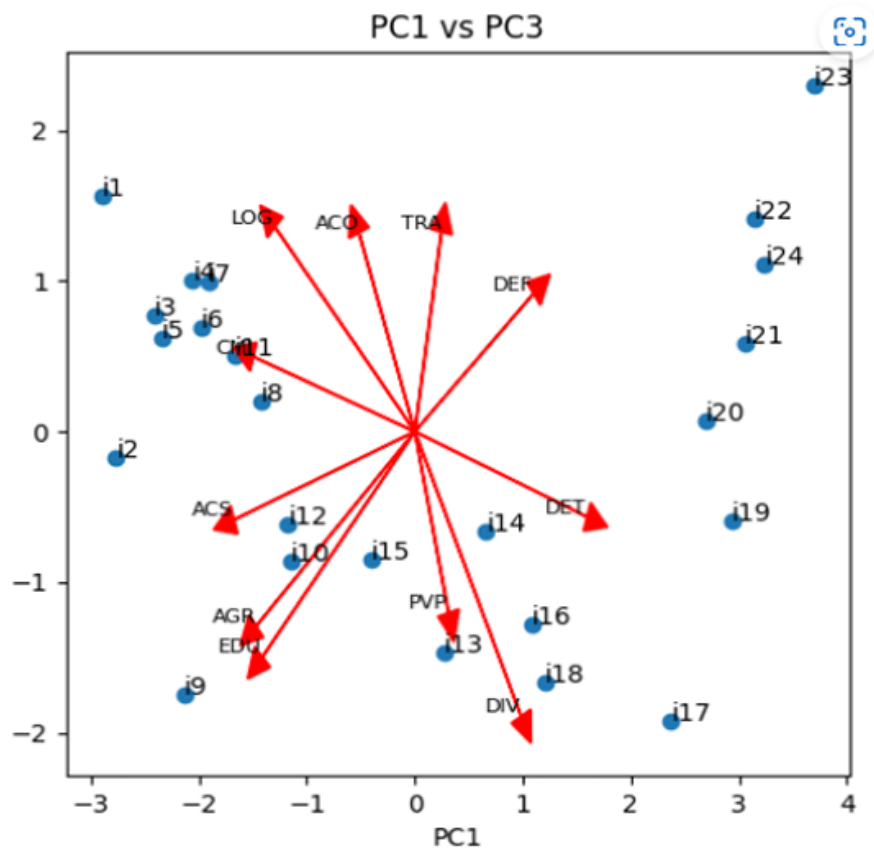


FIGURE 6 – Biplot PC1 vs PC3

4.3 PC2 vs PC3

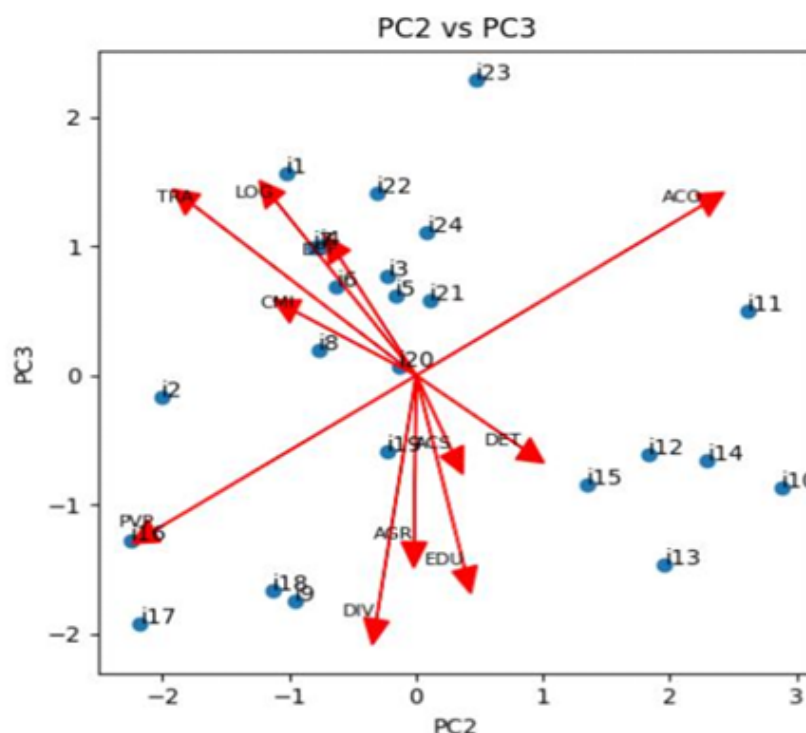


FIGURE 7 – Biplot PC2 vs PC3

5 Conclusion

L'Analyse en Composantes Principales réalisée sur les données de dépenses publiques a permis de dégager trois axes structurants, expliquant **75.57%** de la variance totale :

TABLE 8 – Synthèse des axes retenus

Axe	Interprétation	% Variance
PC1	Balance dépenses sociales vs militaires/dette	45.21%
PC2	Fluctuations conjoncturelles (conflits vs fonctionnement)	18.63%
PC3	Arbitrages internes au sein des dépenses sociales	11.72%

Enseignements principaux :

1. Les données révèlent une **transition structurelle** au fil des décennies : d'un budget dominé par la dette et la défense vers des investissements croissants en éducation, action sociale et agriculture.
2. Le deuxième axe capture des **cycles conjoncturels** d'environ 5 ans, vraisemblablement liés aux contextes géopolitiques et à l'évolution du poste « anciens combattants ».
3. Le troisième axe, plus subtil, met en lumière les **arbitrages internes** entre différents postes du domaine social.