

TP3 : Analyse en Composantes Principales

A. L. N'Guessan, V. Roca & W. Heyse

Le 24 Septembre 2025

Contents

On souhaite réaliser une ACP du jeu de données `Country-data.csv`, ce jeu de données comporte 10 variables résumées décrites ici :

Column.Name	Description
country	Name of the country
child_mort	Death of children under 5 years of age per 1000 live births
exports	Exports of goods and services. Given as %age of the Total GDP
health	Total health spending as %age of Total GDP
imports	Imports of goods and services. Given as %age of the Total GDP
Income	Net income per person
Inflation	The measurement of the annual growth rate of the Total GDP
life_expec	The average number of years a new born child would live if the current mortality patterns are to remain
total_fer	The number of children that would be born to each woman if the current age-fertility rates remain the sa
gdpp	The GDP per capita. Calculated as the Total GDP divided by the total population.

1. Chargez les données et proposez une brève analyse descriptive des variables.
2. On propose de réaliser l'ACP centrée-réduite de ce jeu de données à l'aide des packages `FactoMineR` et `factoextra`. A quoi sert l'argument `scale.unit` dans la fonction `PCA` ?
3. Déterminez le nombre d'axes retenir. *Indication* : Vous pouvez utiliser la fonction `fviz_eig` pour représenter la part d'inertie de chaque axe.
4. Déterminez les individus les mieux représentés et les plus contributeurs aux dimensions que vous avez retenus. *Indication* : Vous pouvez utiliser les fonctions `fviz_pca_ind`, `fviz_contrib` et `fviz_cos2`.
5. Déterminez les variables les mieux représentées et les plus contributrices aux dimensions que vous avez retenus.
6. Que pouvez-vous dire de l'information portée par les variables `child_mort` et `total_fer` ?
7. A l'aide des questions précédantes et de la fonction `fviz_pca_biplot`, interprétez les axes.
8. On propose de faire à nouveau l'ACP des mêmes données mais en utilisant `scale.unit=FALSE`. Que remarquez-vous sur cette ACP, pour quelle raison observez-vous ces différences ? *Indication* : On pourra utiliser la fonction `explor` du package éponyme pour analyser les résultats de l'ACP.
9. L'ACP *n'étant qu'une histoire de projections*, il est possible de construire l'espace des composantes principales sans certaines variables/individus puis de les projeter après coup, qu'on appelle variables/individus supplémentaires. La fonction `PCA` permet de faire cela avec les arguments `ind.sup` et `quanti.sup`. Réalisez une ACP non-normée avec les variables `gdpp` et `income` en variables supplémentaires.
10. En suivant le déroulé habituel de l'ACP, proposez une analyse des structures latentes des données.