

TP 1 : Analyse en composantes principales (ACP)

Zaineb Smida

2024-10-17

Le fichier étudié concerne 25 pays de l'Union Européenne (source : Eurostat 2002). Les variables considérées sont les suivantes :

- région (en 3 catégories : 1 = pays de l'est, 2 = pays du sud, 3 = autres)
- espérance de vie à la naissance pour un homme (en années)
- espérance de vie à la naissance pour une femme (en années)
- population (en milliers d'habitants)
- taux d'activité (en pourcentage)% : %population active/population en âge de travailler
- produit intérieur brut par habitant (en standards de pouvoir d'achat)
- taux d'inflation (en pourcentage)
- taux d'emploi (en pourcentage) : actifs occupés/population en âge de travailler
- taux de chômage (en pourcentage): chômeurs/population active
- taux de chômage longue durée (en pourcentage)
- nombre de mariages (pour 1000 personnes)
- nombre d'abonnés aux services de téléphonie mobile (en milliers)
- variable indicatrice qui vaut 1 pour les pays de l'Europe des 15.

Packages nécessaires :

```
library("FactoMineR")  
library("ggcorrplot")
```

1 Question

Importer le fichier `pays-eu.txt` sous **R**. Vous pouvez l'appeler `pays`

Conseils

Utiliser la fonction `read.table()`. Lorsque la première ligne du fichier txt contient le nom des variables, il faut ajouter l'option `header = T` pour l'indiquer. Vous pouvez visualiser que l'importation s'est correctement effectuée en faisant `View(pays)`. Il peut être utile d'ajouter des noms de lignes en utilisant la fonction `row.names()`.

Solution

2 Question

Réaliser une étude univariée rapide des données à l'aide d'indicateurs numériques et de graphiques. Vous pouvez également représenter les nuages de points entre les paires de variables.

Conseils

Voici quelques fonctions **R** de base permettant de faire de l'analyse univariée :

- `summary()` retourne plusieurs indicateurs comme le minimum, le maximum, la moyenne, la médiane, les quartiles,
- `hist()` prend une variable quantitative en entrée et retourne un histogramme,
- `boxplot()` prend une variable quantitative en entrée et retourne une boîte à moustache,
- `pairs()` prend en entrée plusieurs variables quantitatives (sous forme d'un `data.frame` et retourne toutes les nuages de points possibles entre paires de variables)
- la fonction `plot()` s'utilise principalement sur deux variables quantitatives

Solution

3 Question

Donner la matrice des corrélations et les nuages de points associés. Commenter.

Conseils

- `cor()` prend entrée plusieurs variables quantitatives et retourne la matrice des corrélations,
- la fonction `ggcorrplot()` du package **ggcorrplot** retourne une matrice de corrélation avec une couleur de palette divergente

Solution

4 Question

L'ACP vous paraît-elle justifiée ?

Solution

5 Question

Déterminer le nombre de composantes principales à retenir pour cette ACP. Justifier votre réponse.

Conseils

- La fonction `PCA()` du package **FactoMineR** permet de réaliser une ACP,
- La fonction `str()` appliquée sur le résultat de l'ACP permet d'identifier les informations retournées. On accède aux éléments de cet objet avec le symbole `$`

Solution

6 Question

Interpréter les composantes principales retenues à l'aide des variables initiales. Donner un (des) graphique(s) permettant de visualiser l'interprétation.

Conseils

- On pourra accéder aux coordonnées des variables à partir de l'objet retourné par la fonction `PCA()`. Il pourra être utile de regarder aussi les contributions. La fonction `plot()` appliqué sur les résultats de l'ACP, en ajoutant l'option `choix = "var"` permet de représenter le graphique des variables

Solution

7 Question

Quels sont les pays bien représentés sur chacun des axes retenus ? Justifier votre réponse.

Conseils

- On pourra accéder aux cosinus carré des individus à partir de l'objet retourné par la fonction `PCA()`.

Solution

8 Question

Commenter les contributions des pays aux premiers axes.

Conseils

- On pourra accéder aux contributions des individus à partir de l'objet retourné par la fonction `PCA()`.

Solution

9 Question

Réaliser le(s) graphique(s) des pays et commenter l'ACP.

Conseils

- On pourra accéder aux coordonnées des individus à partir de l'objet retourné par la fonction `PCA()`. Il pourra être utile de regarder aussi les contributions. La fonction `plot()` appliqué sur les résultats de l'ACP, en ajoutant l'option `choix = "ind"` permet de représenter le graphique des variables

Solution

10 Question

Refaire l'analyse en utilisant la fonction `PCAshiny()` du package **Factoshiny**:

Solution