Logo

Description automatically generated

Academia de Studii Economice din București

**Dezvoltare Software pentru analiza datelor**

**Tema:**

**Analiza componentelor principale**

**Profesor: Student:**

Conf. Univ. Dr. Vințe Claudiu Baltariu Andrei

**Introducere:**

Analiza în componente principale (ACP) este cea mai utilizată metodă de analiză a datelor. A fost propusă de Hotteling în 1938, dar necesitând numeroase și laborioase calcule s-a impus în practică abia începând cu anii 70 odată cu apariția calculatoarelor. ACP sintetizează informația conținută în tabelele de date cantitative de mari dimensiuni, cu un mare număr de instanțe și de variabile. O colectivitate statistică descrisă printr-un număr mare de variabile este greu de studiat.

Componentelor principalesunt construite sub formă de combinație liniară de variabile inițiale, având o varianță cât mai mare.Prima componentă principală are rolul de a prelua maximul din varianța variabilelor originale. A doua componentă preia maximul varianței rămase, neexplicată de prima componenta, a treia procedează la fel, etc. Scopul analizei în componente principale este de a sintetiza varianța existentă la nivelul întregului set de date și de a scoate în evidență ceea ce este semnificativ.

1. **Definirea datelor:**

In cadrul acestui proiect, mi-am propus sa fac o analiza asupra mai multor tari, analizand fiecare tara folosind anumiti indicatori considerati de mine relevanti pentru a evidentia forta de dezvoltare si nivelul la care a ajuns tara respectiva. In total am ales 50 de tari.

* - Populatia
* - PIB
* - Cheltuielile suportate pentru sanitate
* - Somajul persoanelor cu studii superioare
* - Suprafata tarii in km^2
* - Forta de munca
* - Rata taxarii
* - Consumul final
* - Economiile brute

**2. Prezentarea rezultatelor**

În urmă efectuării analizei componentelor, vor rezulta 12 componente principale.

In vederea evaluării rezultatelor modelului, am ales câteva criterii de evaluare pe care le-am salvat în fișierul dataOUT.

Pentru analizarea rezultatelor obținute se vor studia graficele generate.

**2.1 Contribuția observațiilor:**

Chart, scatter chart

Description automatically generated

Figură 1 Distribuția observațiilor în spațiul C1,C2

Table

Description automatically generated with medium confidence

Figură 2 Corelograma contribuției observațiilor la varianta axelor componentelor

**2.2 Analiza factorilor de corelație:**

Chart

Description automatically generated

Figură 3 Corelatia dintre variabilele initiale si C1, C2

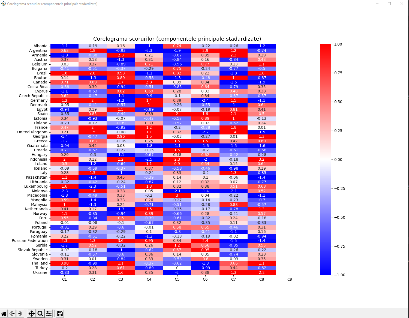
Chart

Description automatically generated

Figură 4 Corelograma factorilor de corelație

**2.3 Scorurile:**

- componente principale standardizate



Figură 5 Corelograma scorurilor

**2.4 Calitatea observațiilor:**

Graphical user interface, application, table, calendar

Description automatically generated

Figură 6 Corelograma calitatii reprezentarii observatiilor

**2.5 Comunalitățile:**

- cantitatea de varianță explicată în comun de către un grup de componente principale

Calendar

Description automatically generated with low confidence

Figură 7 Corelograma comunitatilor

**3. Concluzii**

În concluzie, deoarece analiza componentelor principale este una dintre cele mai utilizate metode de analiză factorială multidimensională, am putut scoate în evidență un sistem de axe factoriale care concentrează informația conținută în tabelul inițial pentru o mai buna vizualizare a acestuia. Aceasta metoda determină un număr mai mic de variabile noi, numite componente principale, care rezuma informația, variabilitatea prezenta la nivelul colectivității studiate și se utilizează ca instrument de analiză statistică în multe domenii de activitate. Folosind aceasta metoda de analiza am putut determina diferite legaturi intre tarile alese in setul de date, astfel putand sa aducem o opinie obiectiva asupra unor elemente legate de dezvoltarea si forta anumitor state.

**4. Surse**

Sursa datelor: **EUROSTAT, WORLDOMETER**

URL: <https://ec.europa.eu/eurostat/en/web/main/data/database>

URL: https://www.worldometers.info/gdp/gdp-by-country/