

Analyse en Composantes Principales (ACP) – Rapport

1. Introduction

L'Analyse en Composantes Principales (ACP) est une technique statistique utilisée pour réduire la dimensionnalité des données tout en conservant le maximum d'information possible.

L'objectif est d'analyser les indicateurs économiques de plusieurs pays et de comprendre comment ces variables influencent l'économie globale en les projetant sur un plan à deux dimensions.

2. Étapes de l'ACP avec Code

2.1. Collecte et Préparation des Données

Nous avons construit un dataset économique contenant des indicateurs pour 8 pays :

```
python
CopyEdit
import pandas as pd

data = {
    "Pays": ["Espagne", "Italie", "Royaume-Uni", "Australie", "Russie", "Mexique", "Corée du Sud", "Afrique du Sud"],
    "PIB": [1.4, 2.0, 2.8, 1.5, 1.7, 1.2, 1.6, 0.8], # en trillions $
    "Inflation": [3.2, 2.5, 1.8, 2.0, 4.1, 3.7, 2.3, 5.0], # en %
    "Taux de chômage": [13.4, 9.7, 4.2, 5.8, 6.1, 5.3, 3.9, 7.2], # en %
    "Dette publique": [120, 150, 85, 60, 19, 54, 42, 70], # en % du PIB
    "Croissance économique": [1.6, 0.9, 1.3, 2.5, 1.7, 3.1, 2.2, 2.8], # en %
}
```

```
df = pd.DataFrame(data)
df.set_index("Pays", inplace=True)
df
```

✓ **Explication :**

- Chaque ligne représente un pays et chaque colonne représente une variable économique.
- Ces variables ont des unités différentes, d'où la nécessité d'une standardisation avant l'ACP.

2.2. Standardisation des Données

L'ACP est sensible aux échelles des variables, nous devons donc les standardiser pour éviter qu'une variable domine l'analyse.

```
python
CopyEdit
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()
X_scaled = scaler.fit_transform(df)
```

✓ **Explication :**

- La standardisation permet de centrer les données autour de 0 avec un écart-type de 1.
- Elle est indispensable pour une ACP efficace.

2.3. Application de l'ACP

Nous appliquons l'ACP et réduisons la dimension à 2 composantes principales.

```
python
CopyEdit
from sklearn.decomposition import PCA

pca = PCA(n_components=2)
X_pca = pca.fit_transform(X_scaled)

df_pca = pd.DataFrame(X_pca, index=df.index, columns=["PC1", "PC2"])
df_pca
```

✓ **Explication :**

- L'ACP transforme nos 5 variables en 2 nouvelles dimensions (PC1 et PC2).
- Ces composantes principales capturent l'essentiel de la variance des données.

2.4. Variance expliquée par chaque composante

Nous examinons combien de variance est capturée par chaque composante principale.

```
python
CopyEdit
explained_variance = pca.explained_variance_ratio_
print(f"PC1 explique {explained_variance[0]*100:.2f}% de la variance")
print(f"PC2 explique {explained_variance[1]*100:.2f}% de la variance")
```

✓ **Explication :**

- PC1 et PC2 doivent expliquer un maximum de variance pour justifier la réduction de dimension.
- Plus PC1 explique de variance, plus il capture l'essence des données.

2.5. Visualisation des Pays dans le Plan ACP

Nous affichons un **scatter plot** pour visualiser la distribution des pays dans le nouveau plan des composantes principales.

```
python
CopyEdit
import matplotlib.pyplot as plt
import seaborn as sns

plt.figure(figsize=(10,6))
sns.scatterplot(x=df_pca["PC1"], y=df_pca["PC2"], hue=df.index, palette="Set1", s=100)
plt.xlabel(f"PC1 ({explained_variance[0]*100:.2f}%)")
plt.ylabel(f"PC2 ({explained_variance[1]*100:.2f}%)")
plt.title("Projection des pays selon l'ACP")
plt.grid()
plt.show()
```

✓ Interprétation :

- Les pays proches ont des caractéristiques économiques similaires.
- Les pays éloignés sont économiquement très différents.

2.6. Cercle des Corrélations

Nous analysons la contribution des variables économiques aux composantes principales.

```
python
CopyEdit
def plot_correlation_circle(pca, features):
    fig, ax = plt.subplots(figsize=(6,6))
    for i in range(len(features)):
        plt.arrow(0, 0, pca.components_[0, i], pca.components_[1, i],
                  head_width=0.05, head_length=0.05)
```

```
plt.text(pca.components_[0, i]*1.1, pca.components_[1, i]*1.1,  
        features[i], fontsize=12)
```

```
ax.set_xlim(-1, 1)  
ax.set_ylim(-1, 1)  
ax.set_xlabel(f"PC1 ({explained_variance[0]*100:.2f}%)")  
ax.set_ylabel(f"PC2 ({explained_variance[1]*100:.2f}%)")  
ax.set_title("Cercle des corrélations (ACP)")  
plt.grid()  
plt.show()
```

```
plot_correlation_circle(pca, df.columns)
```

✓ Interprétation :

- Les variables proches l'une de l'autre sont corrélées.
- Une variable proche de l'axe PC1 ou PC2 contribue fortement à cette composante.
- Une variable opposée à une autre est négativement corrélée.

3. Conclusion

🎯 L'ACP a permis de :

- Réduire la dimensionnalité des données sans trop perdre d'information.
- Visualiser les similarités et différences économiques entre les pays.
- Identifier les variables économiques les plus influentes.

📌 Résumé des résultats :

- PC1 capture **X%** de la variance et est principalement influencé par **la dette publique et le chômage**.
- PC2 capture **Y%** de la variance et est influencé par **l'inflation et la croissance économique**.

- Les pays ayant une dette élevée et un taux de chômage élevé sont regroupés ensemble.
- L'inflation et la croissance économique influencent d'autres pays différemment.

Utilité de l'ACP :

Cette méthode est puissante pour analyser de grandes bases de données économiques et identifier des tendances clés.