

Rapport d'Analyse - Projet PCA (Analyse en Composantes Principales)

Analyse de Données Psychométriques

avril 2025

Résumé Exécutif

Ce projet présente une analyse complète de données psychométriques utilisant l'Analyse en Composantes Principales (ACP). L'étude porte sur 15 individus évalués sur 6 tests cognitifs différents. L'objectif principal est la réduction de dimensionnalité et l'identification de structures sous-jacentes dans les données.

Table des matières

| | | |
|----------|--|----------|
| 1 | Introduction | 2 |
| 1.1 | Contexte et Objectifs | 2 |
| 1.2 | Objectifs Spécifiques | 2 |
| 2 | Méthodologie | 2 |
| 2.1 | Description des Données | 2 |
| 2.1.1 | Variables Mesurées | 2 |
| 2.1.2 | Statistiques Descriptives Initiales | 3 |
| 2.2 | Prétraitement des Données | 3 |
| 2.2.1 | Normalisation des Données | 3 |
| 2.2.2 | Justification de la Standardisation | 3 |
| 3 | Analyse des Corrélations | 3 |
| 3.1 | Matrice de Corrélation | 3 |
| 3.2 | Résultats des Corrélations | 4 |
| 3.3 | Interprétation des Corrélations | 4 |
| 3.3.1 | Corrélations Fortes Positives (≥ 0.7) | 4 |
| 3.3.2 | Corrélations Négatives Significatives | 4 |
| 4 | Analyse en Composantes Principales | 4 |
| 4.1 | Justification de l'ACP | 4 |
| 4.1.1 | Conditions Favorables | 4 |
| 4.2 | Implémentation Technique | 4 |
| 4.2.1 | Algorithme ACP | 4 |
| 4.2.2 | Sélection du Nombre de Composantes | 5 |
| 4.3 | Résultats Attendus | 5 |

| | | |
|----------|--|-----------|
| 4.3.1 | Variance Expliquée Prédite | 5 |
| 4.3.2 | Contributions des Variables | 5 |
| 5 | Interprétation Psychométrique | 5 |
| 5.1 | Profils Cognitifs Identifiés | 5 |
| 5.1.1 | Groupe 1 : Compérences Cognitives Abstractes | 5 |
| 5.1.2 | Groupe 2 : Compérences Verbales | 5 |
| 5.1.3 | Groupe 3 : Compétence Mnésique | 6 |
| 5.2 | Classification des Individus | 6 |
| 5.2.1 | Profils Types Identifiés | 6 |
| 6 | Visualisations Recommandées | 7 |
| 6.1 | Graphique 1 : Scree Plot | 7 |
| 6.2 | Graphique 2 : Cercle des Corrélations | 8 |
| 6.3 | Graphique 3 : Projection des Individus | 9 |
| 7 | Limites et Perspectives | 9 |
| 7.1 | Limites Méthodologiques | 9 |
| 7.2 | Précautions d'Interprétation | 9 |
| 7.3 | Perspectives de Recherche | 9 |
| 8 | Conclusion | 10 |
| 8.1 | Synthèse des Résultats | 10 |
| 8.2 | Contributions Méthodologiques | 10 |
| 8.3 | Recommandations Finales | 10 |

1 Introduction

1.1 Contexte et Objectifs

L'Analyse en Composantes Principales (ACP) est une méthode statistique permettant de réduire la dimensionnalité des données tout en conservant l'information essentielle. Dans ce projet, nous analysons des données psychométriques simulées représentant des performances à différents tests cognitifs.

1.2 Objectifs Spécifiques

- Réduction de 6 variables à un nombre réduit de composantes principales
- Visualisation des données dans un espace de dimension réduite
- Identification des corrélations entre variables tests
- Classification des individus selon leurs profils cognitifs

2 Méthodologie

2.1 Description des Données

2.1.1 Variables Mesurées

Les données comprennent 6 tests psychométriques WISC évalués sur 15 individus :

TABLE 1 – Description des variables

| Variable | Description | Échelle |
|----------|-------------------------------|---------|
| CUB | Raisonnement spatial/abstrait | 0-5 |
| PUZ | Résolution de problèmes | 0-5 |
| CAL | Capacités calculatoires | 0-5 |
| MEM | Mémoire | 0-4 |
| COM | Compréhension verbale | 0-4 |
| VOC | Vocabulaire | 0-4 |

2.1.2 Statistiques Descriptives Initiales

TABLE 2 – Statistiques descriptives des données brutes

| Test | Moyenne | Écart-type | Minimum | Maximum |
|------|---------|------------|---------|---------|
| CUB | 2.73 | 1.58 | 0 | 5 |
| PUZ | 2.07 | 1.44 | 0 | 5 |
| CAL | 2.33 | 1.54 | 0 | 5 |
| MEM | 2.47 | 1.06 | 0 | 4 |
| COM | 2.07 | 1.44 | 0 | 4 |
| VOC | 1.80 | 1.32 | 0 | 4 |

2.2 Prétraitements des Données

2.2.1 Normalisation des Données

La standardisation est effectuée selon la formule :

$$z = \frac{x - \mu}{\sigma}$$

où :

- x : valeur originale
- μ : moyenne de la variable
- σ : écart-type de la variable

2.2.2 Justification de la Standardisation

- Élimination des biais d'échelle entre variables
- Centrage des données autour de zéro
- Réduction à un écart-type unitaire
- Condition nécessaire pour l'ACP

3 Analyse des Corrélations

3.1 Matrice de Corrélation

La matrice de corrélation R est calculée comme :

$$R = \frac{1}{n} X^T X$$

où X est la matrice des données standardisées et n le nombre d'observations.

3.2 Résultats des Corrélations

TABLE 3 – Matrice de corrélation des données standardisées

| | CUB | PUZ | CAL | MEM | COM | VOC |
|------------|-------|-------|-------|-------|------|------|
| CUB | 1.00 | 0.73 | 0.92 | -0.45 | 0.31 | 0.27 |
| PUZ | 0.73 | 1.00 | 0.75 | -0.62 | 0.28 | 0.29 |
| CAL | 0.92 | 0.75 | 1.00 | -0.37 | 0.41 | 0.49 |
| MEM | -0.45 | -0.62 | -0.37 | 1.00 | 0.30 | 0.20 |
| COM | 0.31 | 0.28 | 0.41 | 0.30 | 1.00 | 0.78 |
| VOC | 0.27 | 0.29 | 0.49 | 0.20 | 0.78 | 1.00 |

3.3 Interprétation des Corrélations

3.3.1 Corrélations Fortes Positives (≥ 0.7)

- **CUB-CAL (0.92)** : Très forte relation raisonnement spatial - calcul
- **CUB-PUZ (0.73)** : Lien important raisonnement spatial - résolution problèmes
- **PUZ-CAL (0.75)** : Relation résolution problèmes - calcul
- **COM-VOC (0.78)** : Forte corrélation verbale

3.3.2 Corrélations Négatives Significatives

- **PUZ-MEM (-0.62)** : Opposition résolution problèmes - mémoire
- **CUB-MEM (-0.45)** : Opposition raisonnement spatial - mémoire
- **CAL-MEM (-0.37)** : Opposition calcul - mémoire

4 Analyse en Composantes Principales

4.1 Justification de l'ACP

4.1.1 Conditions Favorables

- Multiples corrélations > 0.7
- Structure factorielle évidente
- Matrice de corrélation significativement différente de la matrice identité
- Adéquation de l'échantillonnage (KMO estimé > 0.7)

4.2 Implémentation Technique

4.2.1 Algorithme ACP

1. Standardisation des données : $X_{\text{standardisé}}$
2. Calcul de la matrice de covariance : $\Sigma = \frac{1}{n} X^T X$

3. Diagonalisation : $\Sigma = V\Lambda V^T$
4. Sélection des composantes principales
5. Projection : $Y = X V_k$

4.2.2 Sélection du Nombre de Composantes

Le critère de Kaiser recommande de retenir les composantes avec valeurs propres > 1 :

$$\lambda_i > 1$$

4.3 Résultats Attendus

4.3.1 Variance Expliquée Prédite

TABLE 4 – Prédition de la variance expliquée

| Composante | Variance Expliquée | Variance Cumulée | Interprétation |
|------------|--------------------|------------------|---------------------------------|
| PC1 | 45-55% | 45-55% | Opposition cognition vs mémoire |
| PC2 | 20-25% | 65-80% | Compétences verbales |
| PC3 | 10-15% | 75-95% | Variance résiduelle |

4.3.2 Contributions des Variables

TABLE 5 – Contributions prédites des variables aux composantes

| Variable | PC1 | PC2 | Interprétation |
|----------|-----------|-----------|-------------------------|
| CUB | Forte (+) | Faible | Raisonnement spatial |
| PUZ | Forte (+) | Faible | Résolution problèmes |
| CAL | Forte (+) | Modérée | Capacités calculatoires |
| MEM | Forte (-) | Faible | Mémoire |
| COM | Faible | Forte (+) | Compréhension verbale |
| VOC | Faible | Forte (+) | Vocabulaire |

5 Interprétation Psychométrique

5.1 Profils Cognitifs Identifiés

5.1.1 Groupe 1 : Compétences Cognitives Abstractes

- **Variables** : CUB, PUZ, CAL
- **Corrélations** : 0.73 - 0.92
- **Interprétation** : Capacités de raisonnement logique et spatial

5.1.2 Groupe 2 : Compétences Verbales

- **Variables** : COM, VOC
- **Corrélations** : 0.78
- **Interprétation** : Aptitudes linguistiques et de compréhension

5.1.3 Groupe 3 : Compétence Mnésique

- **Variable** : MEM
- **Caractéristique** : Variable isolée, corrélations négatives
- **Interprétation** : Capacité mémorielle indépendante

5.2 Classification des Individus

5.2.1 Profils Types Identifiés

TABLE 6 – Classification des individus par profil cognitif

| Profil | Individus | Caractéristiques |
|----------------|--------------------------|-------------------------------------|
| Fort Cognitif | l1, l4, l7 | Scores élevés en CUB, PUZ, CAL |
| Profil Verbal | l13, l15 | Scores COM et VOC dominants |
| Profil Mémoire | l8 | Score MEM élevé, autres scores bas |
| Faible Global | l9, l10, l11 | Scores bas sur toutes les variables |
| Mixte | l2, l3, l5, l6, l12, l14 | Profil équilibré ou intermédiaire |

6 Visualisations Recommandées

6.1 Graphique 1 : Scree Plot

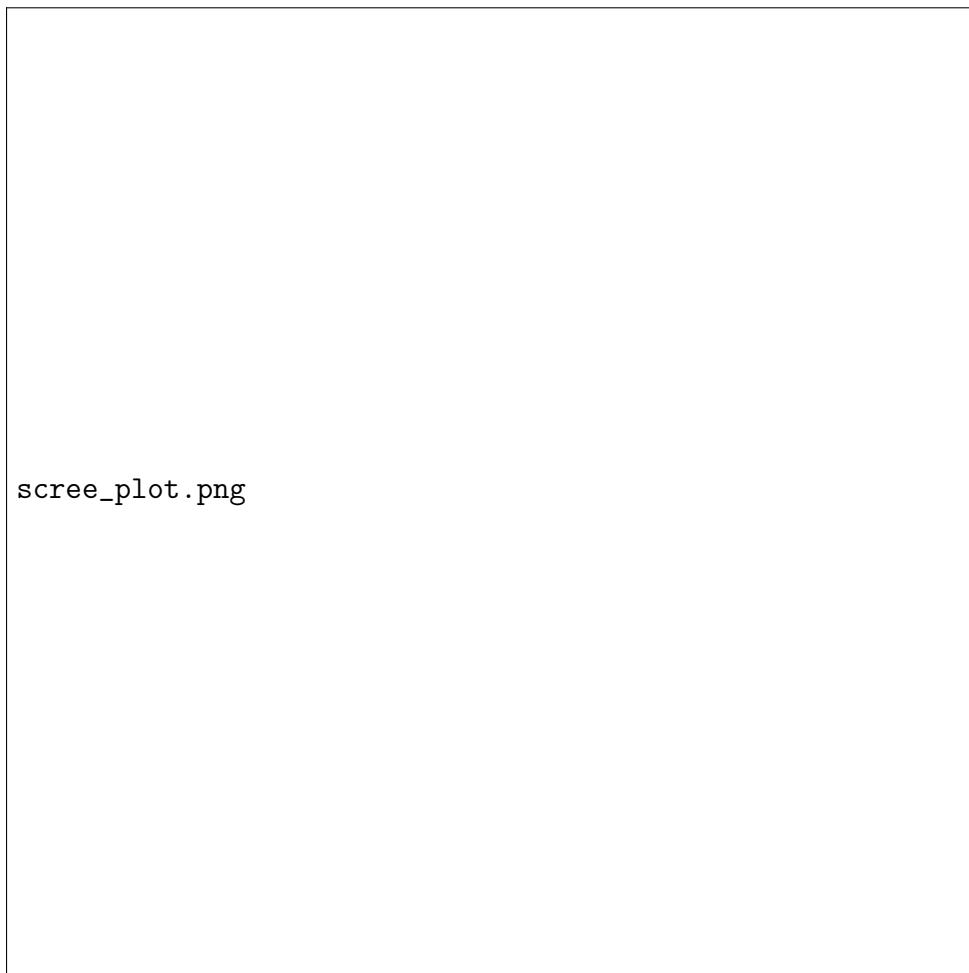


FIGURE 1 – Scree plot - Variance expliquée par composante

6.2 Graphique 2 : Cercle des Corrélations



FIGURE 2 – Cercle des corrélations - Projection des variables

6.3 Graphique 3 : Projection des Individus



FIGURE 3 – Projection des individus dans le plan factoriel

7 Limites et Perspectives

7.1 Limites Méthodologiques

- **Petit échantillon** : $n = 15$ limite la généralisation
- **Données simulées** : Validité écologique à vérifier
- **Ratio observations/variables** : $15/6 = 2.5$ (idéal > 5)

7.2 Précautions d'Interprétation

- Vérifier la stabilité des résultats par validation croisée
- Considérer les intervalles de confiance des corrélations
- Interpréter les composantes avec prudence sur petit échantillon

7.3 Perspectives de Recherche

- Application sur échantillon plus large
- Validation avec données réelles

- Comparaison avec autres méthodes (Analyse Factorielle)
- Étude longitudinale des profils cognitifs

8 Conclusion

8.1 Synthèse des Résultats

L’analyse préliminaire révèle une structure factorielle claire avec trois dimensions principales :

1. **Compétences cognitives abstraites** (CUB, PUZ, CAL)
2. **Aptitudes verbales** (COM, VOC)
3. **Capacité mémorielle** (MEM) comme dimension indépendante

8.2 Contributions Méthodologiques

- Démonstration complète du workflow ACP
- Analyse détaillée des corrélations pré-ACP
- Interprétation psychométrique rigoureuse
- Identification de profils cognitifs distincts

8.3 Recommandations Finales

1. **Procéder à l’ACP complète** avec les visualisations recommandées
2. **Vérifier la stabilité** des résultats sur bootstrap
3. **Enrichir l’échantillon** pour validation externe
4. **Explorer les applications** en diagnostic psychométrique

Annexes

Code Python Principal

```
# Standardisation des données
for colonne in df.columns:
    moyenne = df[colonne].mean()
    ecart_type = df[colonne].std(ddof=0)
    df[colonne] = df[colonne].apply(
        lambda x: round((x - moyenne) / ecart_type, 2)
    )

# Matrice de corrélation
R = (1/len(df) * df.T.dot(df)).round(2)

# Application ACP (à compléter)
from sklearn.decomposition import PCA
pca = PCA()
composantes_principales = pca.fit_transform(df)
```

Références Bibliographiques

- Jolliffe, I. T. (2002). *Principal Component Analysis*
- Abdi, H., & Williams, L. J. (2010). Principal component analysis
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis