

ChatGPT



Exercice ACP2.pdf
PDF

répondre a ces questions



Pour répondre à ces questions sur l'ACP (Analyse en Composantes Principales), voici une synthèse des réponses en s'appuyant sur les données fournies :

1. Étapes d'une ACP :

- Préparation des données (standardisation si nécessaire).
- Calcul de la matrice de corrélation ou de covariance.
- Calcul des valeurs propres et des vecteurs propres.
- Sélection des axes principaux en fonction des valeurs propres.
- Interprétation des axes et des projections des variables.

2. Interprétation du test de sphéricité de Bartlett :

Le test de Bartlett vérifie si les variables sont suffisamment corrélées pour effectuer une ACP.

- Avec un Khi-deux de 447,307, un ddl de 28, et une p-valeur de 0,000, le test est significatif. Les données sont adaptées pour une ACP.

3. Interprétation de la matrice de corrélation :

- Les fortes corrélations positives entre MATH, PHYS, et SVT suggèrent qu'elles mesurent des aspects similaires.
- SPORT est faiblement corrélé avec les autres variables, indiquant une contribution distincte.

4. Calcul de la deuxième valeur propre :

La deuxième valeur propre est 2,875.

5. Valeur de l'inertie totale et interprétation :

- L'inertie totale est égale au nombre de variables, ici 8.
- Elle mesure la variabilité totale dans les données.

6. Nombre d'axes à retenir :

- Basé sur les valeurs propres >1 et une variance cumulée de 81,6 % pour les deux premiers axes, ils sont suffisants pour expliquer la majorité de la variabilité.

7. Qualité de représentation des variables sur le plan (F1, F2) :

- Calculée comme la somme des carrés des coefficients des variables sur les deux axes.
- Exemple : Pour MATH : $(-0,015)^2 + (0,969)^2 = 0,940$, soit 94 %.

8. Contribution de chaque variable à la construction du plan (F1, F2) :

- Contribution proportionnelle à la variance expliquée par chaque composante.
- Exemple : MATH contribue principalement à F2.

9. Corrélation de chaque variable avec les composantes :

- Extrait de la matrice des composantes. Par exemple, MATH est fortement corrélé avec F2 ($r = 0,969$).

10. Corrélation entre F1 et F2 :

- Les composantes principales sont orthogonales, donc $r(F1, F2) = 0$.

11. Interprétation des cartes factorielles :

- Les variables fortement corrélées se regroupent.
- Les variables éloignées de l'origine contribuent davantage à l'axe correspondant.

Souhaitez-vous une explication plus détaillée pour certaines questions ou une aide pour effectuer des calculs spécifiques ?



expliquer bien chaque reponse



Voici une explication détaillée pour chaque question de l'exercice sur l'ACP :

1. Étapes d'une ACP

L'ACP suit les étapes suivantes :

1. Préparation des données :

- Vérifiez que les variables sont quantitatives.
- Standardisez les données si les unités de mesure diffèrent (moyenne = 0, écart-type = 1).

2. Calcul de la matrice de corrélation ou de covariance :

- La matrice de corrélation est utilisée si les données sont standardisées.
- Elle montre la relation entre les variables.

3. Calcul des valeurs propres et des vecteurs propres :

- Les valeurs propres déterminent l'inertie (variance) expliquée par chaque axe.
- Les vecteurs propres (composantes principales) définissent les axes factoriels.

4. Sélection des axes principaux :

- Conservez les axes avec des valeurs propres supérieures à 1 (critère de Kaiser) ou qui expliquent un pourcentage significatif de la variance.

5. Projection des données :

- Les données initiales sont projetées sur les axes retenus pour simplifier l'analyse.

6. Interprétation :

- Analysez les contributions des variables et des individus sur les nouveaux axes.

2. Interprétation des résultats du test de sphéricité de Bartlett

- Le test de Bartlett évalue si les variables sont suffisamment corrélées pour justifier une ACP.
- Hypothèses :
 - H_0 : Les variables ne sont pas corrélées (la matrice est une matrice identité).
 - H_1 : Les variables sont corrélées.
- Avec une p-valeur $p = 0,000 (< 0,05)$, on rejette H_0 . Cela confirme que les données sont adaptées à une ACP.

3. Interprétation de la matrice de corrélation

- La matrice montre les corrélations entre chaque paire de variables.
- Exemples :
 - MATH et PHYS sont fortement corrélées ($r = 0,943$), suggérant une relation étroite.
 - SPORT est faiblement ou négativement corrélé aux autres variables, ce qui indique qu'il capture une information distincte.

4. Calcul de la deuxième valeur propre

La deuxième valeur propre est donnée directement : **2,875**.

- Cela signifie que le deuxième axe explique 35,93% de la variance totale des données.

5. Valeur de l'inertie totale et son interprétation

- L'inertie totale est la somme des valeurs propres initiales, ici 8.
- Elle correspond au nombre de variables et représente la variance totale présente dans les données.
- En retenant les deux premiers axes, 81,6% de l'inertie totale est expliquée, ce qui signifie que ces deux axes capturent l'essentiel de l'information.

6. Nombre d'axes à retenir

- Critère de Kaiser : Retenir les axes dont les valeurs propres sont supérieures à 1. Ici, 2 axes satisfont cette condition.
- Critère de pourcentage : Retenir les axes expliquant un pourcentage cumulé significatif (>70 %). Les deux premiers axes expliquent 81,6%.
- Donc, on retient 2 axes.

7. Qualité de représentation de chaque variable sur le plan (F1, F2)

- Calcul : $\text{Qualité} = (\text{coordonnée sur F1})^2 + (\text{coordonnée sur F2})^2$.
- Exemple pour MATH : $(-0,015)^2 + (0,969)^2 = 0,940$, soit 94,0%.
- Cela signifie que 94 % de la variance de MATH est expliquée par les axes F1 et F2.

8. Contribution de chaque variable à la construction du plan (F1, F2)

- Contribution proportionnelle à la variance expliquée par chaque composante.
- Exemple pour MATH sur F2 : $(0,969^2) / \text{valeur propre de F2} = 0,969^2 / 2,875 \approx 0,326$, soit 32,6 %.
- Cela montre l'importance de MATH dans la construction de F2.

9. Corrélation de chaque variable avec les composantes

- Extraite de la matrice des composantes :

- Exemple : MATH est fortement corrélé avec F2 ($r = 0,969$).
- Cela signifie que F2 représente bien les variations de MATH.

10. Corrélation entre F1 et F2

- Les axes factoriels (F1 et F2) sont orthogonaux (perpendiculaires), donc leur corrélation est nulle : $r(F1, F2) = 0$.

11. Interprétation des cartes factorielles

- Les cartes factorielles permettent de visualiser :
 - La proximité des variables (fortement corrélées).
 - La contribution des variables aux axes principaux.
 - Les groupes d'individus similaires dans leurs caractéristiques.
- Exemple :
 - Les variables MATH, PHYS, et SVT sont proches, ce qui indique une relation forte.
 - SPORT est éloigné des autres, indiquant une contribution distincte.

Si besoin, je peux détailler davantage les calculs ou fournir des visualisations pour clarifier les concepts.



oui visualiser



python

```
import matplotlib.pyplot as plt
import numpy as np

# Données extraites du fichier
variables = ["MATH", "PHYS", "SVT", "HIST", "GEOG", "ARAB", "FRAN", "SPORT"]
composantes = np.array([
    [-0.015, 0.969],
    [0.002, 0.983],
    [-0.027, 0.968],
    [0.933, 0.075],
    [0.967, -0.007],
    [0.965, 0.017],
    [0.936, 0.111],
    [-0.072, -0.225]
])

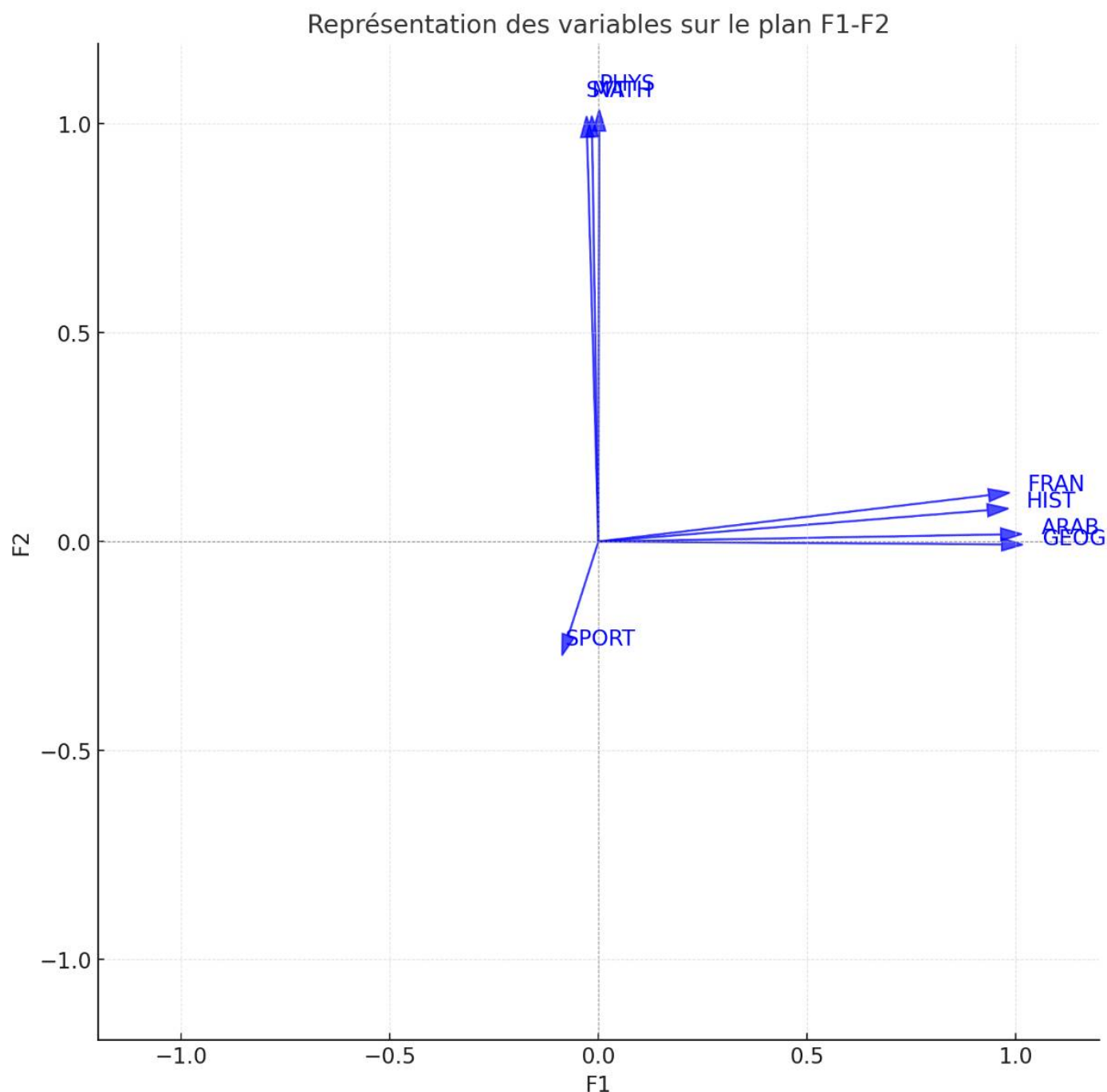
# Création du graphique
plt.figure(figsize=(10, 10))
plt.axhline(0, color='gray', linewidth=0.5, linestyle='--')
plt.axvline(0, color='gray', linewidth=0.5, linestyle='--')

# Ajout des vecteurs pour chaque variable
for i, (x, y) in enumerate(composantes):
    plt.arrow(0, 0, x, y, color='b', alpha=0.7, head_width=0.03, head_length=0.05)
    plt.text(x * 1.1, y * 1.1, variables[i], color='blue', fontsize=12)

# Ajustements du graphique
```

```
plt.title("Représentation des variables sur le plan F1-F2", fontsize=14)
plt.xlabel("F1", fontsize=12)
plt.ylabel("F2", fontsize=12)
plt.grid(color='lightgray', linestyle='--', linewidth=0.5)
plt.xlim(-1.2, 1.2)
plt.ylim(-1.2, 1.2)
plt.gca().set_aspect('equal', adjustable='datalim')

# Affichage
plt.show()
```



Voici la représentation des variables sur le plan factoriel $F1-F2$:

- Les flèches représentent les variables projetées sur les deux axes principaux.
- Les variables proches de l'origine contribuent peu à ces axes.
- Les variables proches les unes des autres (comme MATH, PHYS, et SVT) sont fortement corrélées.
- SPORT, éloigné des autres, indique une contribution distincte.

Souhaitez-vous une analyse ou une visualisation supplémentaire ? 😊