

# Analyse en Composantes Principales des Données Académiques :

## Table des matières

<b>Rapport sur l'Analyse en Composantes Principales (ACP)</b>	<b>1</b>
Table des matières	1
Introduction	2
Étapes du Processus	2
1. Chargement des Bibliothèques	2
2. Chargement des Données	2
3. Préparation des Données	2
4. Réalisation de l'ACP	3
6. Identification et Suppression des Outliers	5
7. Positionnement Personnel sur le Nuage ACP	6

# Introduction

L'Analyse en Composantes Principales (ACP) est une méthode d'analyse multivariée permettant de résumer l'information contenue dans des données quantitatives en identifiant les dimensions principales. Ce rapport documente les étapes suivies pour effectuer une ACP sur des données académiques, en vue de mieux comprendre les profils des étudiants

## Étapes du Processus

### 1. Chargement des Bibliothèques

Pour mener à bien cette analyse, nous avons utilisé plusieurs bibliothèques R :

- `readxl` pour lire les fichiers Excel.
- `FactoMineR` et `factoextra` pour effectuer et visualiser l'ACP.
- `dplyr` pour manipuler les données.

### 2. Chargement des Données

Les données sont chargées depuis deux fichiers Excel :

- `pv_deliberation.xlsx` : Contient les informations académiques des étudiants (notes, rangs, etc.) durant l'années 1CS.
- `pv_affectation.xlsx` : indique l'affectation des étudiants.

Les fichiers ont été fusionnés à l'aide de la variable commune "Matricule" pour créer un jeu de données complet.

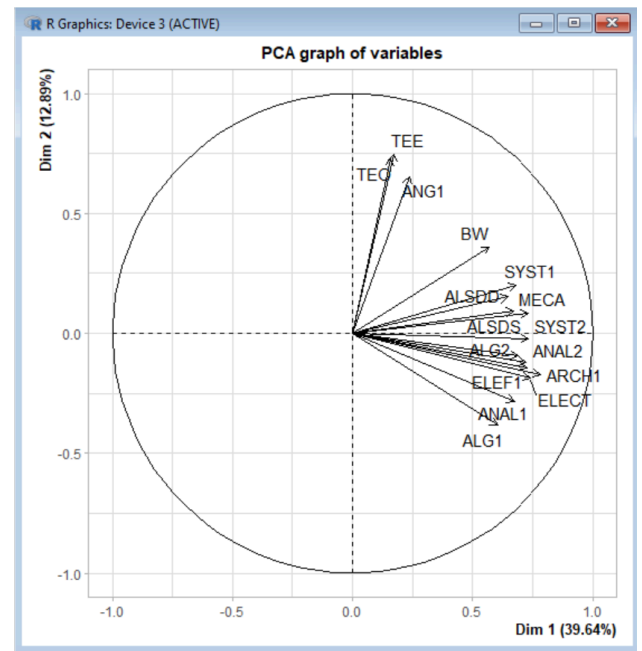
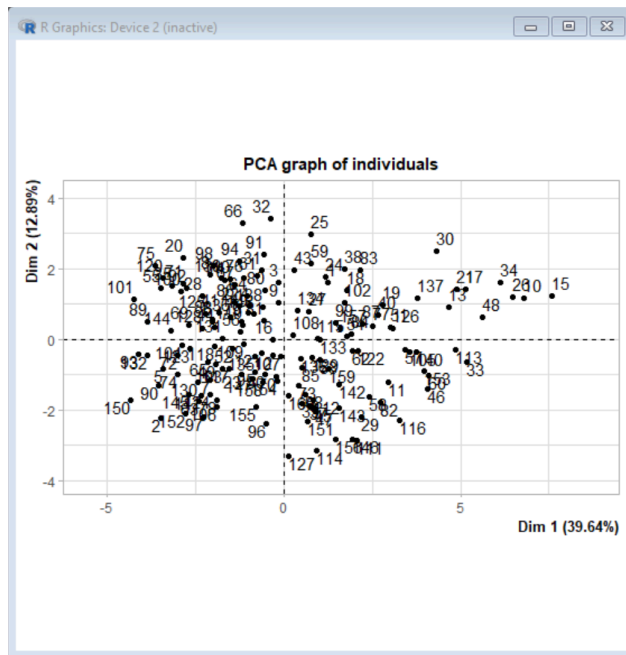
### 3. Préparation des Données

La préparation des données est cruciale pour garantir la pertinence de l'analyse :

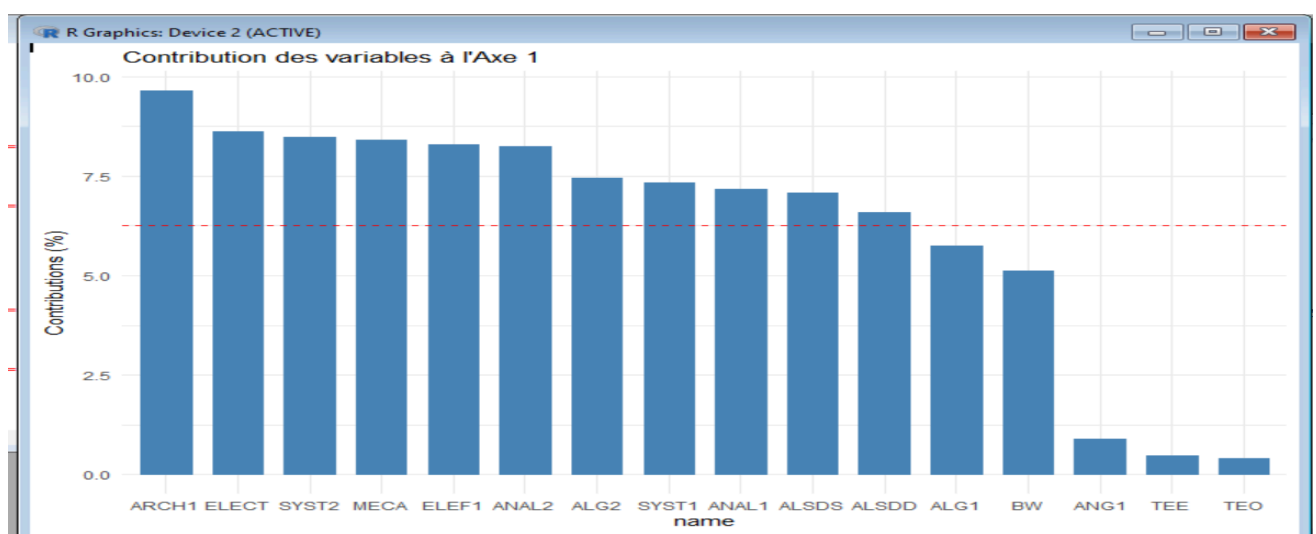
- Retrait des colonnes inutiles : Les colonnes liées aux notes, rangs, et crédits ont été supprimées car elles ne contribuent pas directement à l'ACP.
- Suppression de variables qualitatives : La colonne "Affectation" a été temporairement retirée, car l'ACP nécessite des variables quantitatives.
- Retrait d'un individu spécifique : L'étudiant ayant le matricule "21/0061" a été isolé pour être projeté ultérieurement sur le nuage des individus.

#### 4. Réalisation de l'ACP

L'ACP est réalisée sur les données nettoyées en utilisant la fonction **PCA** du package **FactoMineR** :



Afin d'interpréter les axes 1 et 2 on a affiché les contribution de chacune des variables et on a gardé les variable qui avait une contribution supérieure à  $1/16=6,25\%$

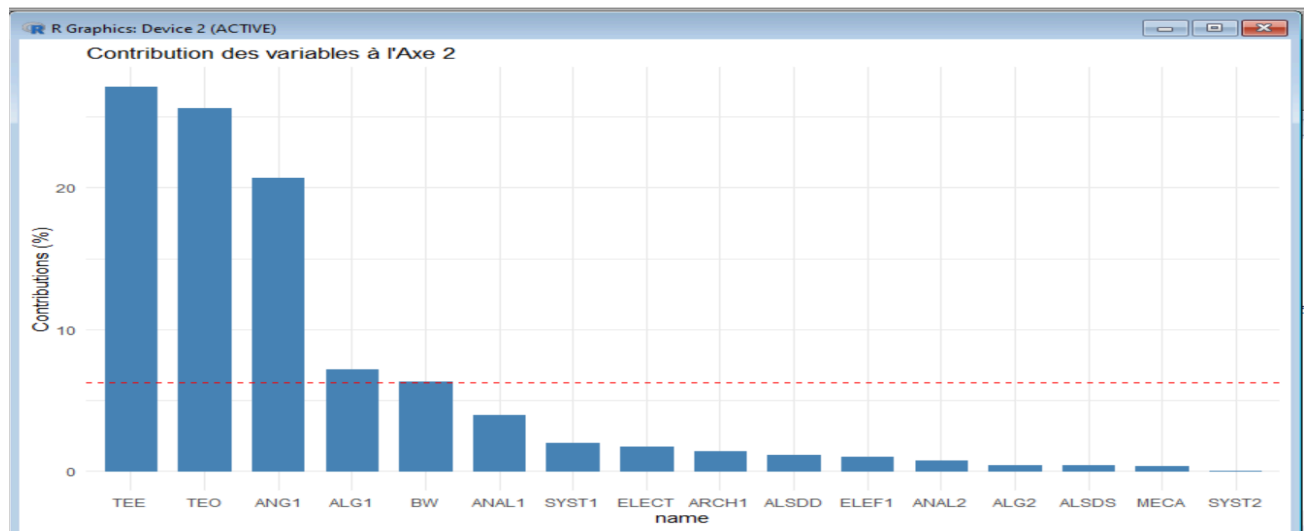


L'axe 1 représente 39,64% de l'information. Il est fortement influencé par les variables ARCH1, ELECT, SYST2, MECA, ELEF1, ANAL2, ALG2, SYST1, ANAL1, ALSDDS, et ALSDD. Ces

contributions suggèrent que cet axe mesure principalement les performances académiques des étudiants dans des matières techniques fondamentales liées à :

- L'architecture informatique (ARCH1),
- L'électricité (ELECT),
- Les systèmes d'exploitation (SYST1, SYST2),
- La mécanique et l'électronique fondamentale (MECA, ELEF1),
- Les mathématiques appliquées (ANAL1, ANAL2, ALG1, ALG2),
- Et les structures de données (ALSDS, ALSDD).

L'axe 1 mesure donc les compétences techniques générales des étudiants dans les disciplines liées à l'informatique et à l'ingénierie.

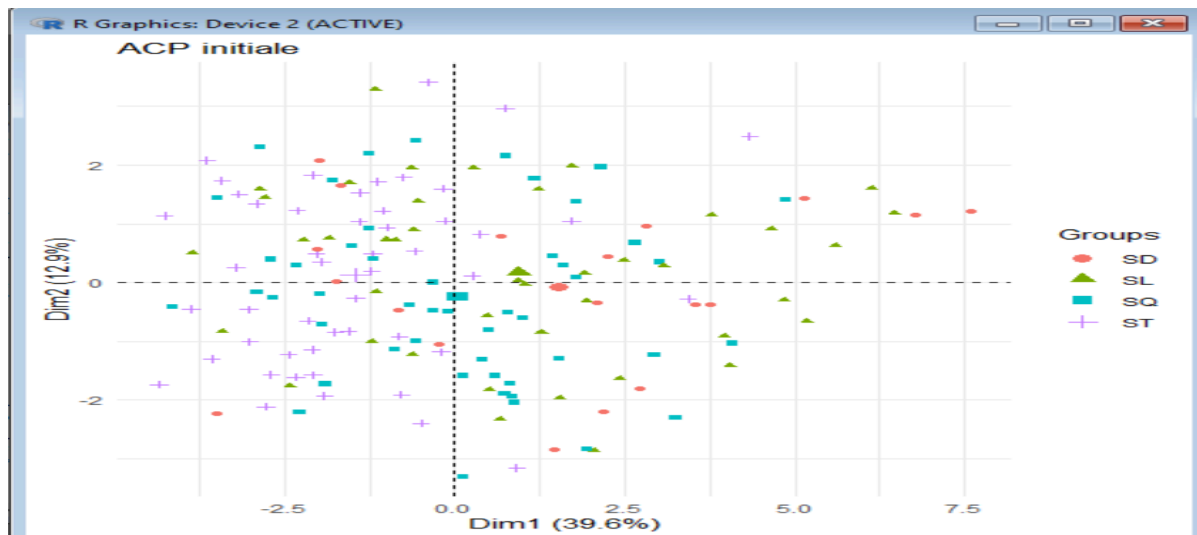


L'axe 2 représente 12,89% de l'information. Il est principalement influencé par TEE, TEO, et ANG1, qui sont des matières moins liées aux compétences techniques directes et davantage aux compétences transversales comme :

- Les Techniques d'Expression Écrite (TEE) et Orale (TEO),
- L'anglais technique (ANG1).

L'axe 2 mesure les compétences en communication et linguistiques, séparant les étudiants qui excellent dans ces aspects de ceux qui ont des résultats plus faibles.

Les résultats initiaux ont été visualisés sous forme de nuages de points, où les individus ont été colorés en fonction de leur affectation. Cela a permis d'observer les regroupements et les tendances générales des données.



Les étudiants du groupe SIT se distinguent par des performances plus faibles dans des modules tels que ALSDD, ALSDS, Analyse, et Algèbre, comme le montre leur position sur le graphique.

Les étudiants SIL obtiennent de meilleures performances dans les mêmes modules (ALSDD, ALSDS, Analyse, Algèbre), ce qui indique une certaine maîtrise dans ces disciplines.

Le groupe SID est très dispersé sur le graphique, ce qui signifie que ses étudiants ont des résultats très variés. Certains peuvent exceller dans certains modules, tandis que d'autres sont moins performants.

À l'instar des SID, les SIQ montrent également une forte dispersion, avec des étudiants répartis sur toute la carte.

Les deux groupes SIL et SIT montrent une opposition claire, avec les SIL excellant dans les modules théoriques et fondamentaux, tandis que les SIT ont des performances plus faibles dans ces mêmes modules.

## 6. Identification et Suppression des Outliers

Les outliers (ou valeurs aberrantes) sont des observations qui se situent loin de la majorité des autres données dans un ensemble. Dans le contexte de l'Analyse en Composantes Principales (ACP), les outliers peuvent fausser les résultats et doivent donc être identifiés et supprimés. Une méthode courante pour détecter les outliers dans des données multivariées est la distance de Mahalanobis.

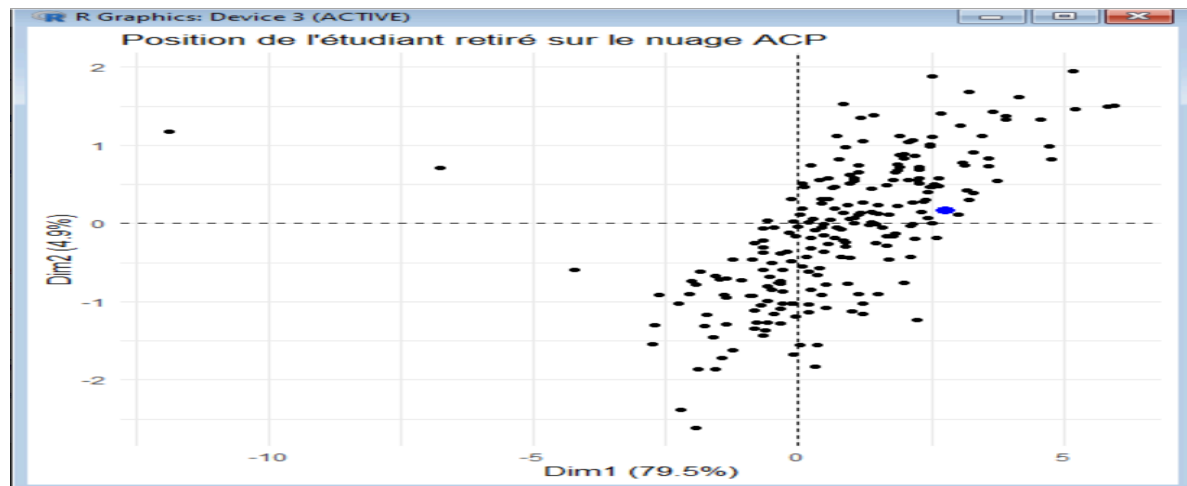
La distance de Mahalanobis est une mesure qui permet de déterminer à quelle distance un point est de la moyenne d'un ensemble de données, en tenant compte de la variance-covariance des données. Contrairement à la distance Euclidienne, elle prend en compte la structure des données (c'est-à-dire les corrélations et les variances des différentes variables).

La distance de Mahalanobis suit une distribution du chi-deux ( $\chi^2$ ). En général, une distance de Mahalanobis élevée indique que l'observation est loin de la distribution des autres données, donc un outlier potentiel. On calcule donc le seuil de chi-deux pour un niveau de confiance de 97.5%. Et on identifie et supprime des observations qui ont une distance de Mahalanobis supérieure au seuil, donc les outliers.

## 7. Positionnement Personnel sur le Nuage ACP

D'après la position de l'étudiant sur le graph on peut dire qu'il excelle dans les modules transversales (TEO TEE anglais) et qu'il a des compétences moyennes dans les disciplines liées à l'informatique et à l'ingénierie.

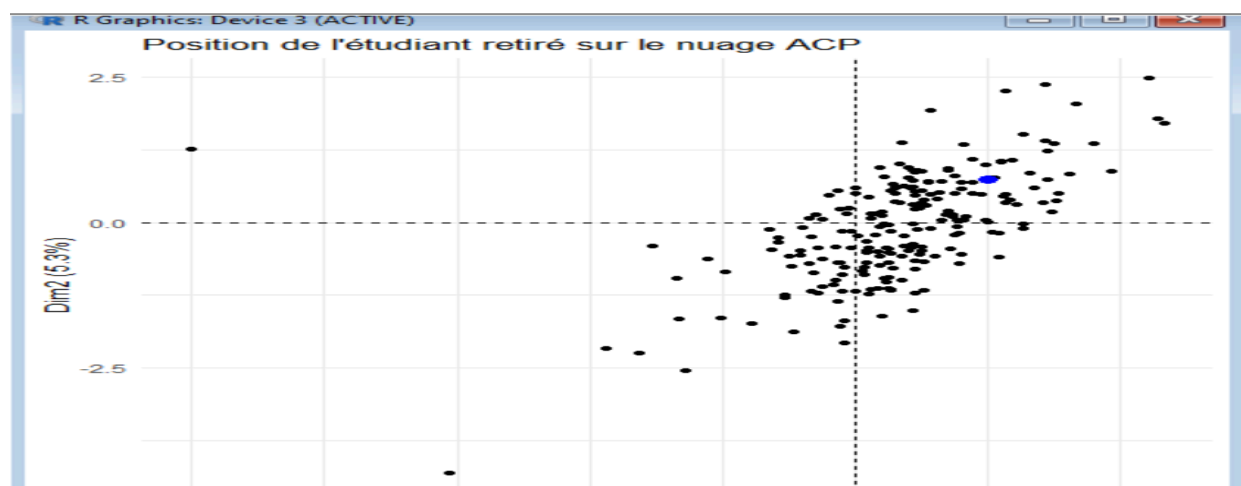
On a aussi projeté l'étudiant au matricule 21/0061 sur le nuage d'individu obtenu après une ACP effectué sur les résultats des étudiants en 2cp. L'axe 1, qui représente 79,46% de l'information, évalue les compétences techniques fondamentales, notamment en programmation orientée objets, architecture 2, optique, algèbre 3, systèmes d'information, analyse 3, électronique 2, structure de fichier et de données, et probabilités 1. Cet axe traduit la maîtrise des concepts clés indispensables en informatique et sciences exactes. L'axe 2, qui explique 4,92% de l'information, reflète les compétences transversales et interdisciplinaires des étudiants, incluant leur performance en probabilités 2, anglais 2 et 3, projet



pluridisciplinaire, et logique mathématique.

D'après la position de l'étudiant, il est plutôt moyen en anglais proba 2 projet pluridisciplinaire et logique mathématique. Et il est bon en programmation orientée objets, architecture 2, optique, algèbre 3, systèmes d'information, analyse 3, électronique 2, structure de fichier et de données, et probabilités 1.

L'étudiant au matricule 21/0061 a été projeté sur le nuage d'individus obtenu après une ACP sur les résultats des étudiants en 1CS. L'axe 1, qui représente 80,1% de l'information globale, évalue les compétences techniques clés, notamment en IGL, architecture, MCSI, système centralisé, organisation et analyse numérique, et réseau 2, reflétant la maîtrise des bases essentielles en sciences et technologies de l'informatique. L'axe 2, qui explique 5,3% de l'information globale, met en lumière des compétences plus spécialisées et interdisciplinaires,



incluant les performances en projet 1CS, théorie des langages de programmation, conduite de projet, organisation et RO, anglais, et système 2.

A partir du graph on peut dire que l'étudiant 21/0061 est moyennement bon. Que ce soit dans les modules mesurés par l'axe 1: IGL, architecture, MCSI... ou les modules mesuré par l'axe 2: théorie des langages de programmation conduite de projet.

À travers les projections obtenues sur les différentes ACP, on peut conclure que l'étudiant 21/0061 a des performances globalement moyennes tout au long de son parcours universitaire. Cela correspond à son appartenance au groupe SID, caractérisé par une hétérogénéité des résultats, ce qui reflète des compétences solides dans certaines matières, mais avec un besoin d'amélioration dans d'autres.

## Conclusion :

L'Analyse en Composantes Principales (ACP) réalisée sur les données académiques des étudiants a permis de dégager des profils distincts et d'identifier des tendances générales concernant les compétences des étudiants dans différents domaines. L'ACP a révélé des différences notables entre les groupes d'étudiants, notamment en termes de performances dans les matières techniques fondamentales et les compétences transversales. Les étudiants des groupes SIL se distinguent par leurs compétences solides dans les disciplines théoriques, tandis que ceux du groupe SIT montrent des résultats plus faibles dans les mêmes domaines. Les groupes SID et SIQ, quant à eux, affichent une forte dispersion, reflétant la diversité des performances au sein de ces groupes.

L'analyse de l'étudiant au matricule 21/0061, projeté sur les nuages des individus, a mis en évidence une performance académique globalement moyenne, marquée par des compétences fortes dans certaines matières techniques, mais avec une nécessité d'amélioration dans les compétences transversales et interdisciplinaires. Cela correspond à son appartenance au groupe SID, caractérisé par une grande hétérogénéité des résultats.

Enfin, la suppression des outliers a permis une légère amélioration de la qualité des données, renforçant ainsi la validité des résultats obtenus. Cette analyse ouvre des pistes intéressantes pour une meilleure compréhension des profils d'étudiants et peut être utilisée pour guider des stratégies pédagogiques ciblées visant à améliorer les performances dans les domaines où des faiblesses sont observées.