

# PROJET ANALYSE DE DONNÉES

## SONDAGE FSTG

---

Réalisée par :  
AHTOUTE WISSAM /IFA  
ELBCIR ABDELHADI/IRISI

Encadré par :  
OUAARAB AZIZ

# Sommaire

## **01** Introduction

## **02** Description du jeu de données

## **03** LA DÉMARCHE CHOISIE ET LES APPORTS DE L'ANALYSE

## **04** Analyse des correspondances multiples (ACM) Classification

## **05** Anova a un facteur Analyse de l'effet

## **06** Conclusion

# 01 Introduction

Ce projet vise à analyser le comportement des étudiants de la Faculté des Sciences et Techniques (FST) en utilisant un questionnaire. Les données recueillies comprennent des informations sur le profil démographique des étudiants, leurs pratiques d'études, leurs aspirations professionnelles et leurs perceptions personnelles.

Nous entreprendrons une analyse approfondie de ces données en suivant les étapes principales de la production d'un rapport d'analyse. Cela commencera par le nettoyage et l'importation des données brutes, suivi de leur présentation sous forme de tableaux appropriés. Ensuite, nous procéderons à une description détaillée des variables mesurées, de la population d'individus et des conditions de collecte des données.

Différentes méthodes d'analyse seront utilisées pour explorer les données, notamment une analyse descriptive pour dégager des tendances, des corrélations et des différences significatives entre les groupes d'étudiants. Nous utiliserons également des méthodes exploratoires multivariées telles que l'Analyse Factorielle des Correspondances (AFC), l'Analyse des Correspondances Multiples (ACM) et l'Analyse Factorielle Discriminante (AFD) en fonction des données et des objectifs d'analyse.

Une classification des étudiants sera effectuée en utilisant des méthodes de calcul de distances, et des groupes homogènes seront obtenus en appliquant l'AFC ou l'ACM. Enfin, nous analyserons la dimension de la personnalité des étudiants en utilisant des scores qui varient de manière croissante ou fixe, afin d'identifier les caractéristiques personnelles des étudiants.

Ces analyses permettront de fournir une description complète des données et de dégager des implications pertinentes pour mieux comprendre le comportement des étudiants de la FST.

## 02 Description du jeu du données

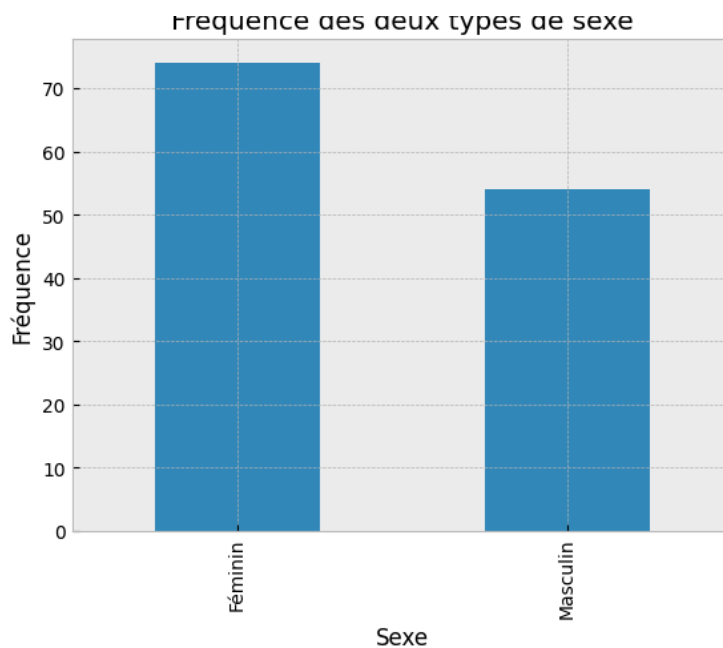
Le jeu de données comprend diverses informations sur les étudiants de la Faculté des Sciences et Techniques (FST), allant de leurs caractéristiques démographiques à leurs attitudes et croyances personnelles. Parmi les variables enregistrées, on retrouve des données telles que le genre, la formation suivie, le niveau d'études, la nationalité, les plans de carrière post-diplôme, ainsi que des éléments sur la fréquence des cours et le sentiment d'intégration à la communauté étudiante.

De plus, des questions portent sur les pratiques d'étude des étudiants, leur entourage social en termes d'âge et de niveau d'études, et leurs perceptions sur l'intelligence, l'apprentissage et l'effort personnel. Ces variables sont formulées sous forme de réponses vrai/faux pour explorer les attitudes des étudiants face au travail, aux défis, aux erreurs et à la notion de capacités intellectuelles.

En résumé, le jeu de données offre une vue détaillée sur les profils et les perceptions des étudiants de la FST, permettant ainsi une analyse approfondie de divers aspects de leur vie universitaire et de leurs perspectives personnelles.

- **Analyse des fréquences :**

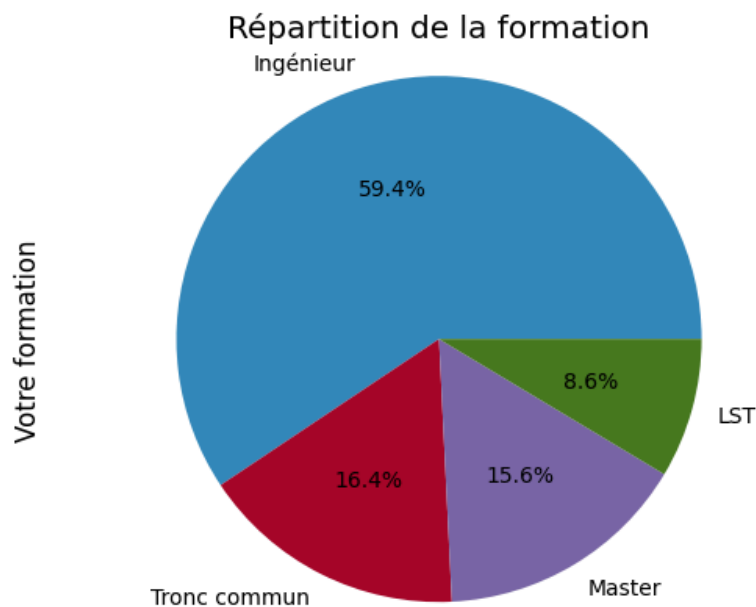
Dans cette étape, nous avons calculé les fréquences relatives des hommes et des femmes dans notre dataset, cela nous a permis de déterminer la répartition proportionnelle des deux sexe , le graphe suivant represente la repartition de sexe :



Cette analyse initiale nous a donné une idée de la représentativité de chaque groupe dans notre échantillon. Dans notre dataset , on remarque que les femmes sont plus nombreuses que les hommes

- Répartition des formations :

Nous avons ensuite examiné la répartition des formations au sein de notre échantillon. Pour cela, nous avons identifié les différentes catégories de formations présentes dans notre dataset, le graphe suivant représente la répartition de la formation :



Nous avons calculé les fréquences relatives de chaque catégorie de formation afin de comprendre comment elles étaient réparties parmi les individus de notre échantillon. Dans notre dataset , on remarque que la formation ingénieur domine , suivi par tron commun .

- **Test 2D (table de contingence) :**

Le test 2D, également connu sous le nom de table de contingence, est une étape permettant d'évaluer les relations entre les variables explicatives, nous avons construit 2 tables de contingence croisant les catégories de deux variables

### La relation entre la formation et le parcours après le diplôme des individus (en pourcentage )

Votre formation apres_diplome	Ingénieur	LST	Master \
Continuer dans un autre parcours de formation	0.342105	0.157895	0.131579
Créer mon entreprise	0.800000	0.066667	0.133333
Je ne sais pas	0.352941	0.117647	0.176471
Travailler dans le privé	0.795455	0.022727	0.181818
Travailler dans le public	0.714286	0.071429	0.142857

Votre formation	Tronc commun
...	
Créer mon entreprise	0.000000
Je ne sais pas	0.352941
Travailler dans le privé	0.000000
Travailler dans le public	0.071429

Cette table représente la répartition des individus de chaque formation selon leur choix après le diplôme.

#### Interprétations :

- Les individus ayant suivi une formation en ingénierie ont une forte association avec le choix de travailler dans le secteur privé (coefficient de contingence de 0.795455). Cela suggère que les étudiants en ingénierie sont plus enclins à chercher des opportunités d'emploi dans le secteur privé après l'obtention de leur diplôme. Cette tendance peut être due à la demande élevée de compétences techniques dans le secteur privé.
- Les individus ayant suivi une formation en LST ont une probabilité relativement faible de créer leur propre entreprise (coefficient de contingence de 0.066667). Cela indique que les diplômés en LST sont moins enclins à entreprendre et à se lancer dans des activités entrepreneuriales après l'obtention de leur diplôme. Ils peuvent préférer des carrières plus axées sur l'emploi salarié ou la poursuite de leurs études.



- **Test 2D (table de contingence) :**

Dans notre étude, nous avons utilisé un ensemble de questions fermées pour évaluer si la personnalité est perçue comme fixe ou croissante.

Les participants ont répondu à un total de 10 questions, chacune avec deux choix possibles : "vrai" ou "faux". Sur la base de ces réponses, nous avons défini des critères pour déterminer si la personnalité est considérée comme fixe ou croissante.

**Pour la perspective de la personnalité fixe:**

Réponses "vrai" pour les questions : 1, 3, 4,5, 7, 8,9

Réponses "faux" pour les questions : 2, 6, 10

**Pour la perspective de la personnalité croissante :**

Réponses "vrai" pour les questions : 2, 6, 10

Réponses "faux" pour les questions : 1, 3, 4,5, 7, 8,9

En analysant les réponses des participants, nous avons calculé un score basé sur leur tendance vers une perspective fixe ou croissante de la personnalité.

Si le score était supérieur à 5, cela indiquait une prédominance de la perspective de la personnalité fixe, tandis qu'un score inférieur ou égal à 5 révélait une tendance vers la perspective de la personnalité croissante.

Cette approche nous permet de catégoriser les participants en fonction de leur perception de la personnalité en tant que fixe ou croissante, en utilisant des questions spécifiques pour chaque perspective. Cela nous permettra d'explorer les corrélations entre ces croyances et d'autres variables d'intérêt dans notre étude, et d'approfondir notre compréhension de la manière dont les individus perçoivent le développement et le changement de la personnalité



- Test 2D (table de contingence) :

### La relation entre le sexe et assister au cours des individus (en pourcentage)

Coefficient de contingence :			
cours	Assez régulièrement	Rarement	Systématiquement
Sexe			
Féminin	0.148649	0.081081	0.770270
Masculin	0.314815	0.037037	0.648148

### Interprétations :

- Pour les personnes de sexe féminin, 11 d'entre elles suivent assez régulièrement des cours, 6 le font rarement et 57 le font systématiquement. Pour les personnes de sexe masculin, 17 d'entre elles suivent assez régulièrement des cours, 2 le font rarement et 35 le font systématiquement.
- Le coefficient de contingence mesure la relation entre les deux variables (sexe et fréquence des cours). Les valeurs entre 0 et 1 indiquent le degré de dépendance entre les variables. Plus la valeur est proche de 1, plus il y a une forte relation entre les variables.
- En utilisant le coefficient de contingence, on peut observer que la fréquence des cours est plus fortement liée au sexe chez les hommes que chez les femmes. Cela est suggéré par les coefficients de contingence plus élevés pour les hommes dans les catégories "Assez régulièrement" (0.314815) et "Systématiquement" (0.648148). Pour les femmes, les coefficients de contingence sont plus bas dans ces catégories, ce qui indique une relation moins forte entre le sexe et la fréquence des cours.

- **Test 2D (table de contingence) :**

Analyse de relation entre formation et la colonne personnalité :  
(en pourcentage)

```
Votre formation  Ingénieur  LST  Master  Tronc commun
personality
croissant          70    10    20         20
fixe                6     1     0         1
Chi-square statistic: 1.9150907571960203
P-value: 0.5902155658041921
Degrees of freedom: 3
Expected frequencies: [[71.25   10.3125  18.75   19.6875]
[ 4.75    0.6875  1.25    1.3125]]
```

### Interprétations :

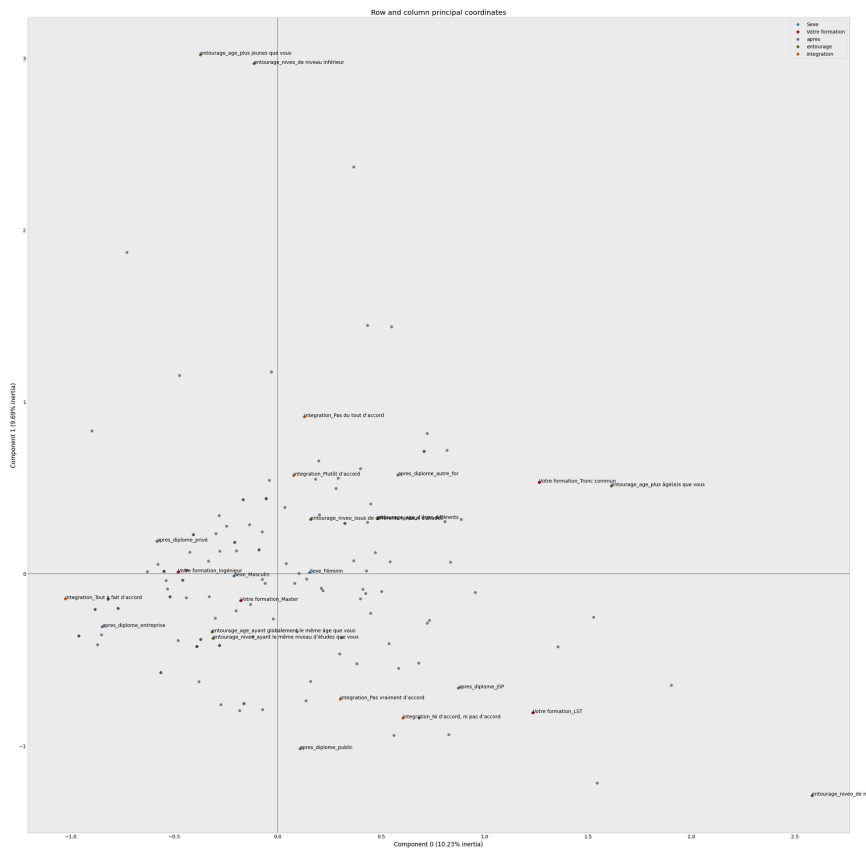
- L'interprétation des données de la table de contingence montre qu'il y a une concentration plus élevée de personnes ayant la formation d'Ingénieur et ayant une personnalité croissante. Avec 70 personnes dans cette catégorie, elle représente le groupe le plus important parmi tous les groupes de formation et de personnalité.
- En revanche, il y a seulement 6 personnes ayant la formation d'Ingénieur et ayant une personnalité fixe. Ce chiffre est beaucoup plus bas en comparaison aux autres combinaisons de formation et de personnalité.
- L'analyse du test du Chi-carré et la p-value associée indiquent qu'il n'y a pas suffisamment de preuves pour établir une relation significative entre la formation et la personnalité dans l'échantillon étudié. Cela signifie que la répartition observée des formations et des personnalités pourrait être due au hasard plutôt qu'à une relation véritable entre ces variables.

### • Analyse des correspondances multiples (ACM) :

- L'Analyse des Correspondances Multiples (ACM) est une technique d'analyse factorielle adaptée aux variables qualitatives. Elle permet d'étudier les relations entre les différentes variables explicatives et de visualiser ces relations dans un espace multidimensionnel.
- L'ACM nous a aidés à comprendre les regroupements, les tendances et les structures sous-jacentes dans nos données qualitatives.

Le graphe ci-dessous représente les individus et les variables dans un plan factoriel afin d'observer les associations et les différences entre les catégories de variables explicative :

- Examinant la relation entre formation , l'entourage , sexe , parcours après diplôme et intégration



### Interprétations :

- Les étudiants qui ont un entourage d'âge globalement du même âge ont également un entourage ayant globalement le même niveau.
- Généralement, les ingénieurs veulent continuer à travailler dans le secteur privé.
- Les étudiants qui sont totalement intégrés ont tendance à choisir de créer une entreprise après le diplôme.
- Les étudiants qui ne savent pas quoi faire après le diplôme ne sont pas vraiment intégrés.
- Les étudiants du tronc commun ont un entourage d'âge plus élevé.
- Les étudiants qui ont un entourage d'âge plus jeune ont un entourage de niveau inférieur

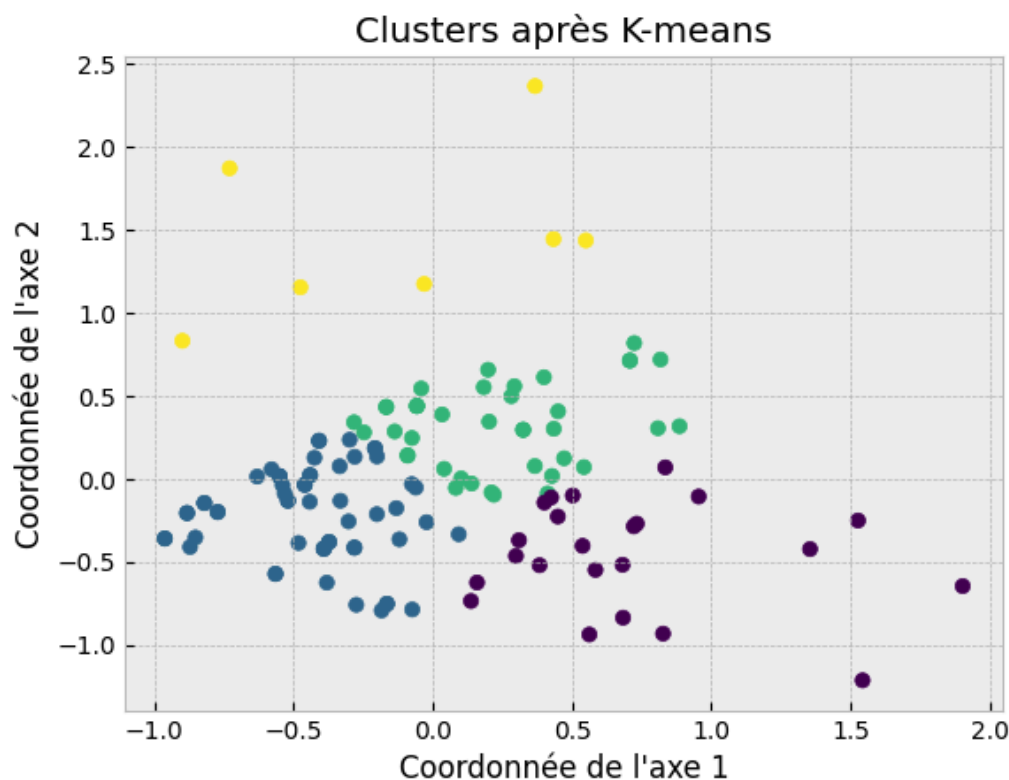
- **Classification :**

- classification nous a permis de regrouper les individus ayant des caractéristiques similaires, ce qui a facilité l'interprétation de nos résultats.

- **K-means :**

K-means est une méthode de partitionnement qui divise l'échantillon en K groupes distincts en minimisant la distance intra-cluster et maximisant la distance inter-cluster.

Nous avons effectué une classification des données en utilisant l'algorithme K-means. Il attribue chaque individu à l'un des 4 clusters en fonction de leurs coordonnées dans l'espace MCA. Les résultats sont ensuite affichés pour montrer le cluster auquel chaque individu est attribué.



L'utilisation des coordonnées des individus dans l'espace factoriel MCA dans la préparation pour la classification K-means permet de réduire la dimensionnalité, faciliter la visualisation et identifier des regroupements d'individus similaires. Cela peut aider à améliorer la performance et l'interprétation des résultats de la méthode K-means dans le contexte des données catégorielles

- **Analyse discriminante :**

L'analyse discriminante est l'étape finale de notre démarche. Elle consiste à construire une fonction discriminante qui maximise la séparation entre les différentes classes de notre variable cible. Cette fonction est basée sur les variables explicatives sélectionnées précédemment.

Les valeurs propres représentent la variance expliquée par chaque composante discriminante dans l'analyse discriminante linéaire (AFD). Dans notre cas, nous avons 4 valeurs propres : [0.46434055, 0.23458622, 0.15155463, 0.1495186]. Cela signifie que la première composante discriminante explique environ 46,43% de la variance, la deuxième composante discriminante explique environ 23,46% de la variance, la troisième composante discriminante explique environ 15,16% de la variance, et la quatrième composante discriminante explique environ 14,95% de la variance.

Les coefficients discriminants correspondent aux coefficients attribués à chaque variable dans chaque composante discriminante. Ils déterminent l'importance de chaque variable pour séparer les classes dans l'AFD.

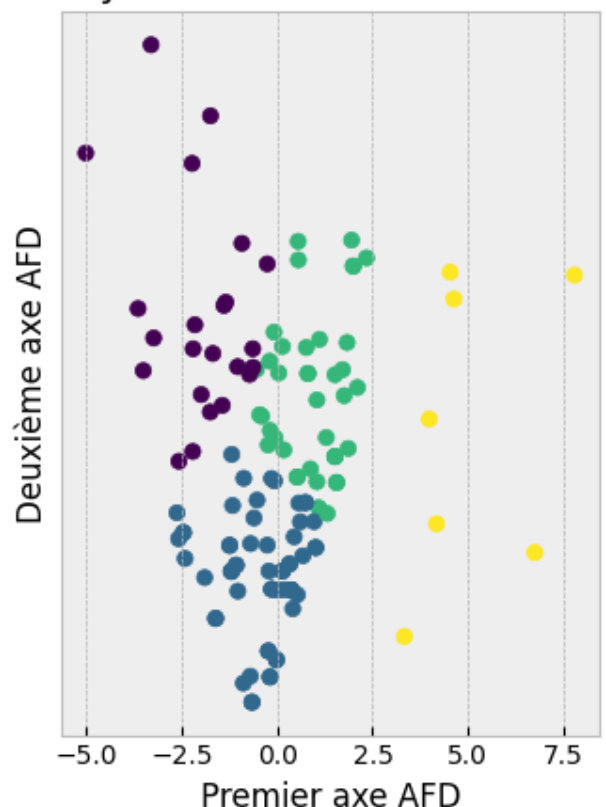
Dans notre cas, nous avons une matrice de coefficients discriminants de taille (4, 27). Chaque ligne représente une composante discriminante et chaque colonne représente une variable.

Par exemple, la première ligne de coefficients discriminants indique l'importance des variables dans la première composante discriminante

la deuxième ligne indique l'importance des variables dans la deuxième composante discriminante, et ainsi de suite.

Il est important de noter que les coefficients discriminants peuvent être difficiles à interpréter directement en raison de la transformation linéaire effectuée lors de l'AFD. Pour une interprétation plus approfondie des résultats, il peut être utile de visualiser les données à l'aide de graphiques ou de réaliser des analyses supplémentaires

Analyse Factorielle Discriminante

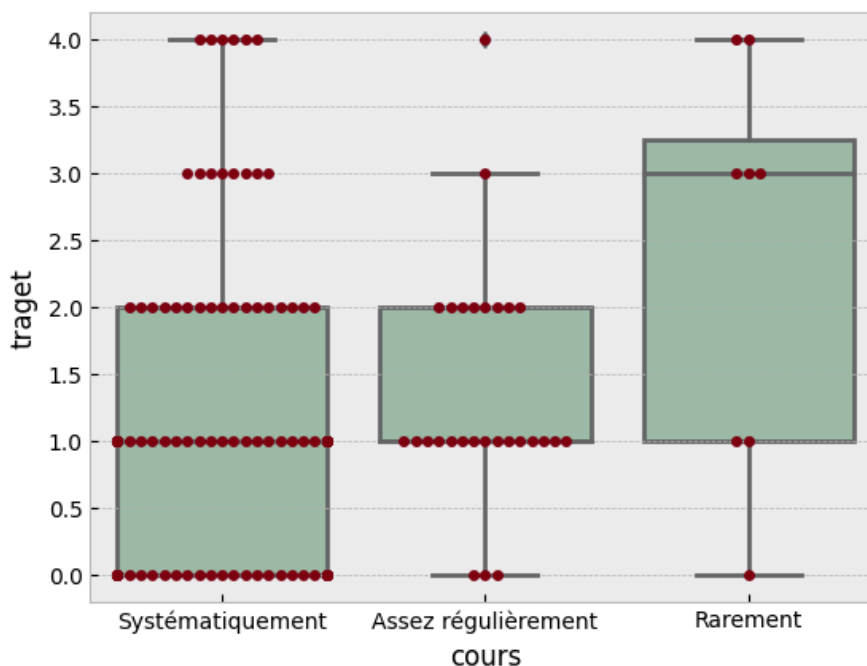


- Anova a un facteur :

L'analyse de variance à un facteur, également appelée ANOVA à un facteur, est une méthode statistique utilisée pour évaluer si la moyenne d'une variable dépend d'une variable catégorielle à un seul niveau, c'est-à-dire d'un seul facteur.

- Affichage de boîte à moustache :

La boîte à moustaches (boxplot) est un outil visuel utilisé dans l'analyse de variance (ANOVA) pour représenter graphiquement la distribution des données pour chaque groupe comparé. Dans le contexte de l'ANOVA, la boîte à moustaches permet de comparer les distributions des différentes catégories du facteur étudié. Chaque boîte représente la répartition des valeurs à l'intérieur d'un groupe, tandis que la ligne médiane de la boîte représente la médiane de chaque groupe.



- Analyse de l'effet

l'effet du facteur "présence(cours)" sur la variable "trajet" est significatif. La p-value associée à cet effet est de 0.028901, ce qui est inférieur au seuil de significativité couramment utilisé de 0,05. Par conséquent, on peut conclure qu'il y a un effet significatif du facteur "présence(cours)" sur la variable "target". Cela indique que les moyennes de la variable "trajet" diffèrent de manière statistiquement significative entre les différents niveaux du facteur "présence(cours)".

	df	sum_sq	mean_sq	F	PR(>F)
C(cours)	2.0	9.511986	4.755993	3.646279	0.028901
Residual	125.0	163.042702	1.304342	NaN	NaN



## **06 Conclusion**

**En résumé, ce projet d'analyse du comportement des étudiants de la Faculté des Sciences et Techniques (FST) à l'aide d'un questionnaire a permis d'explorer de manière approfondie divers aspects de leur vie académique, de leurs aspirations et de leurs perceptions personnelles.**

**Les données recueillies ont fourni une vision détaillée de la population étudiante et ont révélé des informations précieuses sur leurs pratiques d'études, leur intégration, leurs croyances et leurs réactions face aux défis. En conclusion, ce projet a permis d'obtenir une meilleure compréhension du comportement des étudiants de la FST et a fourni des indications précieuses pour améliorer leur expérience universitaire.**

**Ces informations peuvent servir de base solide pour des interventions et des politiques visant à favoriser leur épanouissement académique et personnel. De plus, les résultats de cette étude peuvent également constituer une référence importante pour de futures recherches dans le domaine de l'éducation et de la psychologie des étudiants.**

**Les résultats obtenus grâce à cette étude approfondie des données fournissent une description complète du comportement des étudiants de la FST. Ces résultats ont des implications importantes pour l'amélioration des pratiques pédagogiques, l'optimisation de l'environnement d'apprentissage et le soutien aux aspirations professionnelles des étudiants**

