

RAPPORT- SAE PLAN D'EXPÉRIENCE

SAE 2024/2025

Axelle AKARE NZE

Kinaza ALI

Hajar CKAKIR

Ines DRIS



METHODOLOGIE DU RAPPORT

<u>Introduction</u>	3
<u>La méthodologie utilisée</u>	4
<u>Principaux enseignements/résultats</u>	5
<u>Conclusion</u>	8

I-INTRODUCTION

Dans le cadre de la **SAE "Plan d'expérience"**, un traitement et une analyse sur les données de l'enquête de 2024 a été effectué via un redressement afin d'améliorer la précision des résultats de l'année précédente. Nous avons choisi de nous concentrer sur la problématique suivante : **les usages de l'intelligence artificielle (IA) en fonction du département en comparant les étudiants et les enseignants au sein de l'IUT.**

L'objectif ici est de reprendre notre enquête afin d'avoir des résultats représentatifs de l'ensemble de la population d'intérêt.

Pour mener à bien cette étude, nous avons structuré notre analyse autour des questions suivantes :

- *Dans quel département étudiez vous ou enseignez vous principalement ? (variable département).*
- *Quel est votre statut au sein de l'IUT ? (étudiant ou enseignant).*
- *À quelle fréquence utilisez vous l'IA ?*
- *Quel(s) type(s) d'Intelligence Artificielle avez-vous déjà expérimentés ?*
- *Quel est votre niveau de confiance par rapport à l'utilisation de l'IA ?*

Ces questions permettront de recueillir des informations précises sur les pratiques d'utilisation de l'IA, tout en tenant compte des spécificités liées aux rôles et départements sur l'échelle réelle de l'IUT en utilisant des poids. À travers cette étude, notre objectif est de mieux comprendre comment l'IA est intégrée dans les activités académiques et pédagogiques et d'identifier les éventuelles disparités ou besoins en formation.

Ce rapport présentera notre méthodologie, les traitements des données réalisés avec SAS et Excel, ainsi que les principaux résultats issus de l'analyse. Nous espérons ainsi contribuer à une meilleure compréhension des usages de l'IA dans l'enseignement supérieur.

II-LA MÉTHODOLOGIE UTILISÉE

Pour répondre à notre problématique concernant les usages de l'IA en fonction des départements et des statuts au sein de l'IUT, nous avons adopté une méthodologie structurée en plusieurs étapes :

1.Exploration des données disponibles :

Examinassions des données mises à disposition, notamment :

- Le questionnaire administré aux étudiants et enseignants, afin de comprendre les questions posées et leur pertinence dans le cadre de notre problématique.
- Les données théoriques relatives aux effectifs des étudiants et personnels par département et filière pour l'année académique 2023-2024.

2.Sélection des variables d'intérêt :

Nous avons sélectionné les variables correspondant aux 5 questions sélectionnées ci-dessus. (*cf. introduction*).

3.Nettoyage et préparation des données :

Un traitement des données a été réalisé pour garantir leur qualité :

- Suppression des doublons ou réponses incomplètes.
- Vérification de la cohérence des réponses (par exemple, alignement entre le statut déclaré et le type d'utilisation d'IA).
- Codage des variables catégoriques pour les analyses statistiques et la procédure de redressement.

4.Redressement des données :

Afin d'ajuster les données de l'enquête à la population théorique de l'IUT, nous avons utilisé la macro Calmar sous SAS. Cette étape a impliqué :

- La construction de variables de redressement basées sur des critères comme le statut et le département (critère à notre disposition dans les effectifs théoriques).
- La vérification de la qualité du redressement à l'aide d'indicateurs tels que le rapport de poids et l'efficacité du redressement.

5.Analyses statistiques :

Une fois les données préparées et redressées, nous avons procédé aux analyses statistiques :

- Fréquences et répartitions pour observer les tendances générales.
- Comparaison des réponses entre les étudiants et les enseignants selon les départements.
- Mise en évidence des types d'IA les plus utilisés et des différences dans leur adoption selon les groupes.

6.Logiciels et outils utilisés :

- SAS : pour le redressement des données et les analyses statistiques avancées.
- Excel : pour le traitement préliminaire des données et la visualisation graphique des résultats.

Cette méthodologie rigoureuse garantit des résultats fiables et représentatifs de la population cible. Dans la section suivante, nous détaillerons les principaux enseignements issus de ces analyses.

III-PRINCIPAUX ENSEIGNEMENTS/RÉSULTATS

Pour ce projet, nous disposons au préalable d'une table **Données** ainsi que des effectifs de l'IUT. Nous avons également créé une table **marges** sur Excel basée sur la proportion d'étudiants et d'enseignants, ainsi que sur la répartition des départements.

Le processus débute par l'importation des deux fichiers Excel contenant des informations essentielles pour le bon fonctionnement du redressement. Le fichier "**marge.xlsx**" est chargé dans une table SAS appelée Marges. Ce fichier contient les marges de pondération utilisées pour ajuster les données. Le fichier "**donnees.xlsx**" est importé dans une table SAS nommée Données, qui contient les données brutes collectées lors de l'enquête.

Une fois les données importées, nous avons créé deux nouvelles variables dans la table Données. La première variable, **dep**, est ajoutée pour catégoriser les départements. En fonction de la valeur de la variable département (par exemple, "CS", "GEA", "Info", etc.), une valeur numérique est assignée à **dep**. Cela est nécessaire pour le redressement. La deuxième variable **statut**, est créée pour identifier le statut des individus, en fonction de la variable **fonction**. Si fonction contient le mot "etu", la valeur de statut est définie comme 2 (pour un étudiant, qu'il soit en formation initiale ou en alternance). Si fonction contient "enseignant", statut prend la valeur 1 (pour un enseignant, qu'il soit vacataire ou chercheur).

Exemple de table en sortie :

#	dep	#	statut
	6		2
	2		2
	6		2
	6		2
	5		2

Le code supprime ensuite les lignes où les valeurs des variables département et fonction sont manquantes, garantissant ainsi que seules les lignes contenant des informations complètes sont conservées pour l'analyse. Une procédure **PROC FREQ** est utilisée pour générer des tableaux de fréquences des variables **dep** et **statut**, permettant ainsi de vérifier la distribution des départements et des statuts avant toute modification ou traitement supplémentaire.

On procède ensuite au **redressement des données**, ou **calage**, en utilisant une macro SAS nommée **calmar2**. Cette macro modifie les pondérations des informations en se basant sur les marges présentes dans le fichier Marges. Le calage offre la possibilité de :

- Adapter les pondérations des observations dans la table Données pour qu'elles soient en adéquation avec les marges spécifiques de la population visée.
- Employer un algorithme de redressement fondé sur la technique du sinus hyperbolique, permettant d'évaluer la dimension de la population et du test, ainsi que les indicateurs liés au redressement (rapport de poids, efficacité).
- Restreindre le nombre d'itérations à 50 et se servir de paramètres précis pour déterminer les pondérations.
- Évaluer la progression du redressement.

LES RÉSULTATS OBTENUS :

Taille de l'échantillon (brut)	Taille de la population (pondéré)	Poids minimum	Poids maximum	Poids moyen	Rapport de poids	Efficacité (sans poids d'entrée)
750	3209	1.4051442	20.81805	4.3580607	14.209489	09.41133

Itérations et poids négatifs

Numéro d'itération	Valeur du critère d'arrêt	Nombre d'observations avec un poids négatif
1	3.5276	0
2	0.9008	0
3	0.1680	0
4	0.0296	0
5	0.0010	0
6	0.0000	0

Comparaison échantillon / marges

VAR	TYPE	MOD	Effectif brut	Effectif théorique (MARGE)	Effectif initial	Ecart initial en %	Effectif final	Ecart final en %	Marge en %	% initial	Ecart initial en points	% final	Ecart final en points	ERREUR
DEP	C	1	120	447	523	16.8	447	-0.0	13.70	16.00	2.30	13.70	-0.00	
DEP	C	2	113	755	492	-34.8	755	0.0	23.10	15.07	-8.03	23.10	-0.00	
DEP	C	3	138	509	601	17.9	509	-0.0	15.60	18.40	2.80	15.60	-0.00	
DEP	C	4	88	405	296	-28.9	405	-0.0	12.40	9.07	-3.33	12.40	-0.00	
DEP	C	5	171	290	745	156.2	290	0.0	8.90	22.80	13.90	8.90	0.00	
DEP	C	6	129	657	952	-14.4	657	-0.0	20.10	17.20	-2.90	20.10	-0.00	
DEP	C	7	11	202	47	-79.3	202	0.0	6.20	1.47	-4.73	6.20	-0.00	
STATUT	C	1	72	591	313	-47.0	591	0.0	18.10	9.60	-8.50	18.10	0.00	
STATUT	C	2	678	2 677	2 955	10.4	2 677	0.0	81.90	90.40	8.50	81.90	-0.00	

RAPPORT DE CONVERGENCE	
Statut	CALAGE ATTEINT
Nombre d'écarts aux objectifs	0

Une fois la macro exécuté , la table Données est pondérée en fonction des marges ajustées. Une nouvelle analyse des fréquences est effectuée pour la variable **statut**, en tenant compte des poids ajustés. La procédure **PROC FREQ** applique les pondérations calculées précédemment (stockées dans la variable **POIDS**) afin d'obtenir des résultats plus représentatifs de la population cible.

La procédure FREQ				
dep	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulé
1	120	16.00	120	16.00
2	113	15.07	233	31.07
3	138	18.40	371	49.47
4	88	9.07	439	58.53
5	171	22.80	610	81.33
6	129	17.20	739	98.53
7	11	1.47	750	100.00

La procédure FREQ				
statut	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulé
1	72	9.60	72	9.60
2	678	90.40	750	100.00

La procédure FREQ				
dep	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulé
1	447.853	13.70	447.853	13.70
2	755.139	23.10	1202.992	36.80
3	509.984	15.60	1712.956	52.40
4	405.356	12.40	2118.312	64.80
5	290.941	8.90	2409.253	73.70
6	657.009	20.10	3066.322	93.80
7	202.678	6.20	3269	100.00

La procédure FREQ				
statut	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulé
1	591.669	18.10	591.669	18.10
2	2677.311	81.90	3269	100.00

Enfin, une fois toutes les transformations et analyses effectuées, la table Données, désormais pondérée et traitée, est exportée dans un fichier Excel nommé "**donnees_final.xlsx**". Ce fichier peut ensuite être utilisé pour des analyses supplémentaires ou pour partager les résultats via des graphiques et visualisations.

V- CONCLUSION

Ce projet nous a permis d'avoir un regard critique sur la représentativité des échantillons. Ainsi, grâce au redressement par la macro Calmar, nous avons pu mieux comprendre le processus des techniques de sondage et la méthodologie d'enquête, en reprenant une enquête réalisée ultérieurement. Cela nous permet donc d'affiner notre projet et de le rendre plus précis et recevable.

Nos analyses ont mis en évidence des tendances significatives concernant l'utilisation de l'intelligence artificielle (IA) au sein de l'IUT. Les étudiants se distinguent par une utilisation plus variée et fréquente des outils d'IA, tandis que les enseignants privilégient des applications adaptées à leurs pratiques pédagogiques. Nous avons également constaté des écarts dans le niveau de confiance envers l'IA et des différences notables selon les départements, confirmant que les usages de l'IA varient selon les rôles et contextes académiques.

Les visualisation et résultats seront visibles lors de l'oral.

Ce projet nous a permis de mieux comprendre les usages et perceptions de l'IA dans les activités académiques, tout en consolidant nos compétences en analyse et traitement de données.