le cnam
Paris

Session I : Février 2024

PROJET STA101
ANALYSE DE DONNÉES : MÉTHODES DESCRIPTIVES

# Etude du paysage français des établissements d'enseignement supérieur en matière de ressources (année scolaire 2020-2021)

Hamza GUIZANI & Fatiha KATEB

## RÉSUMÉ

Le présent projet porte sur l'étude des établissements supérieurs français et leurs sources de revenus, associées à des critères quantitatifs sur les effectifs étudiants et personnel ainsi que sur la nature du personnel permanent. Parmi les revenus, on distingue deux sortes : la dotation de l'état (SCSP) et les revenus provenant de subventions diverses ou autres ressources. Un jeu de données totalisant 97 établissements pour 23 variables a été élaboré à partir d'informations disponibles sur la plateforme de diffusion de données publiques de l'état français, associées à des ressources diffusées par le syndicat national de l'enseignement supérieur (SNESUP).

Il semble de prime abord, intuitif que les revenus devraient être proportionnels aux effectifs étudiants, ainsi qu'au nombre de personnel permanent. L'application de méthodes uni-, bi- et multivariées nous a permis de mettre en évidence des disparités au sein des différents types d'établissement supérieurs sur plusieurs de ces variables et de faire ressortir trois groupes au sein de l'ensemble des établissements étudiés :

- le premier uniquement composée d'écoles, parmi lesquelles, ce qu'on appelle communément les « grandes écoles »., avec une dotation rapporté au nombre d'étudiants, plus importante, une proportion de femmes chez les enseignants / enseignants chercheurs plus faible, pour une proportion de professeurs et assimilés plus élevée.
- le second comprend plusieurs types d'établissements, dont l'ensemble des universités Sciences Humaines, presque l'ensemble des universités pluridisciplinaires hors secteur santé et une grande partie des universités pluridisciplinaires avec secteur santé mais de taille relativement petite.
- le troisième correspondant à de grands établissements en terme d'effectifs étudiants composé uniquement d'universités pluridisciplinaires avec secteur santé, dont la dotation par étudiant est faible, mais avec de plus importants revenus par ailleurs.

Table des matières

## LISTE DES ABBRÉVIATIONS

ACP : Analyse en composantes principales
AM2D : Enseignants du secondaire et Arts et Métiers
ANR : Agence Nationale de la Recherche
BIATSS : Bibliothécaires, Ingénieurs, Administratifs, Techniciens, personnels Sociaux et de Santé
CAH : Classification Ascendante Hiérarchique
EPTP : Emplois à Temps Plein Travaillés
ITRF : Ingénieurs et personnels Techniques de Recherche et de Formation
MCF : Maitre de conférences
NA : données manquantes
PR : Professeur
Q-Q plot = Quantile-Quantile plot
SCSP : Subventions pour Charges de Service Public
USH : Universités Sciences Humaines

## INTRODUCTION

La spécificité française du paysage de l'enseignement supérieur réside dans l'existence de différents types d'établissements supérieurs. Ainsi, en plus des universités, il existe de nombreuses écoles sous tutelle publique (souvent accessibles par concours). Pour ce projet, nous nous sommes concentrés sur les établissements pour lesquels les données étaient disponibles, soit l'ensembles des universités françaises et une partie des écoles.
Nous avons choisi de nous intéresser aux spécificités de ces établissements en matière de nombre d'étudiants, de personnel (nombre, nature) et en matière de sources de revenus (nature, montant). Les différentes variables utilisées seront décrites dans le chapitre suivant, ainsi que la source de ces données.

### Présentation des données

- Présentation des variables

  Les données utilisées dans ce projet correspondent à celles de l'année universitaire 2020-2021 et ont été récoltées à partir des sources suivantes.
  - Le site du Syndicat National de l'Enseignement SUPérieur (SNESUP)[1] : pour les variables concernant le nombre d'étudiants, la masse salariale, ainsi que la dotation et le nombre d'équivalents temps.
  - La plateforme ouverte des données publiques françaises (data.gouv)[2] : pour les autres subventions, ressources, ainsi que pour les données concernant le nombre et les différents types de personnel.

  Les variables utilisées sont toutes quantitatives à l'exception, de deux variables qualitatives :
  - Une relative à l'académie d'appartenance de l'établissement concerné
  - L'autre relative à la nature de l'établissement concerné.

  Les variables quantitatives sont pour la plupart, relatives au montant des revenus des établissements supérieurs. Les ressources des établissements supérieurs sont en effet de natures différentes mais proviennent également de différentes sources, les principales sources étant :
  - La dotation de l'État ou Subventions pour Charges de Service Public (SCSP) relative aux salaires et au fonctionnement. Le montant de celle-ci, ne répond à aucune règle connue et peut présenter de grandes disparités en fonction de l'établissement concerné, indépendamment du nombre d'étudiants inscrits dans l'établissement. Ici, nous utiliserons la SCSP rapportée au nombre d'étudiants (SCSP_ETU), pour nous affranchir de ce paramètre.

  - Les ressources propres :
    - Les ressources provenant d'appels à projets déposés à l'Agence Nationale de la Recherche (ANR).
    - Les droits d'inscription (DROITS_INS) : les frais d'inscription varient en fonction du diplôme proposé et de l'origine des étudiants.
    - La taxe d'apprentissage (APPRENTIS)
    - Les ressources provenant de l'offre de formation continue (FORM_CONTINUE) : diplômes propres et Validation des Acquis de l'Expérience (VAE).

[1] https://www.snesup.fr/article/inegalites-de-dotation-quels-sont-les-taux-dencadrement-et-les-budgets-par-etudiant-des-universites-et-etablissements-denseignement-superieur-publics

[2] https://data.enseignementsup-recherche.gouv.fr/pages/explorer/?sort=modified&q=

- Autres (RESS_AUTRES) : valorisation, prestation de recherche, locations de salles.

- Les subventions :
  - Européennes (SUB_EUR)
  - Régionales (SUB_REGION)
  - Autres (SUB_AUTRES) : collectivités (département, communes), organismes nationaux ou internationaux

Les autres variables quantitatives utilisées sont celles caractéristiques de ce type d'établissements et que nous avons jugé pouvant être liées aux précédentes. Il s'agit :

- Du montant de la masse salariale (MAS_SAL)
- Du nombre d'étudiants inscrits (ETU)
- Du nombre d'Equivalents Temps Plein Travaillés rapporté à 100 étudiants (ETPT_100ETU)
- Du nombre d'enseignants / enseignants chercheurs permanents (ENS) et de la nature de ceux-ci :
  - Proportion de professeurs et assimilés (X.PR)
  - Proportion de maîtres de conférences et asismilés (X.MCF)
  - Proportion d'enseignants du secondaire et Arts et Métiers (X.AM2D)
  - Proportion de femmes parmi l'ensemble (X.FEMMES)

- Du nombre de Bibliothécaires, Ingénieurs, Administratifs, Techniciens, personnels Sociaux et de Santé (BIATSS) et de la nature de ceux-ci :
  - Proportion d' Ingénieurs et personnels Techniques de Recherche et de Formation (X.ITRF)
  - Proportion de femmes parmi l'ensemble (X.FEMMES_BIATSS)

- Du taux de permanents (TAUX_PERM)

Le tableau 1 résume l'ensemble des variables utilisées dans ce projet.

| VARIABLE | DESCRIPTION | | NATURE |
|---|---|---|---|
| **ACAD** $ | Académie de rattachement de l'établissement | | quali |
| **TYPE_U** $ | Type d'établissements | | quali |
| *Type d'universités* | UPavS | Universités pluridisciplinaires avec secteur santé | |
| | Uphs | Universités pluridisciplinaires hors secteur santé | |
| | USTS | Universités scientifiques et/ou médicales | |
| | UTALLSHS | Universités tertiaires Arts Lettres Langues Sciences Humaines et Sociale | |
| | UTDEG | Universités tertiaires Droit Economie Gestion | |
| *Type d'écoles* | EI | Écoles d'ingénieurs | |
| | GE | Grandes écoles | |
| | Autres | Autres | |
| **MAS_SAL** * | Masse salariale 2020-2021 (votée fin 2020) en euros | | quanti |
| **ETU** * | Nombre d'étudiants inscrits en 2020-2021 | | quanti |
| **SCSP_ETU** * | SCSP par étudiant | | quanti |
| **ETPT_100ETU** * | Nombre d'équivalent temps plein travaillé pour 100 étudiants | | quanti |
| **ENS** $ | Nombre total d'enseignants ou enseignants-chercheurs titulaires | | quanti |
| **X.AM2D** $ | Proportion d'Enseignants du 2nd degré et "Arts et métiers" parmi le nombre total d'ENS | | quanti |
| **X.MCF** $ | Proportion de maitre de conférences et assimilés parmi le nombre total d'ENS | | quanti |
| **X.PR** $ | Proportion de professeurs et assimilés parmi le nombre total d'ENS | | quanti |
| **X.FEMMES** $ | Proportion de femmes parmi le nombre total d'ENS | | quanti |
| **BIATSS** $ | Nombre de BIATSS permanents | | quanti |
| **X.ITRF** $ | Proportion d'ITRF parmi le nombre total de BIATSS | | quanti |
| **X.FEMMES_BIATSS** $ | Proportion de femmes parmi le nombre total de BIATSS | | quanti |
| **TAUX_PERM** $ | Pourcentage de permanents | | quanti |
| **ANR** $ | Ressources provenant de l'Agence Nationale de la Recherche en euros | | quanti |
| **DROITS_INS** $ | Ressources correspondant aux droits d'inscription en euros | | quanti |
| **APPRENTIS** $ | Ressources provenant de la taxe d'apprentissage en euros | | quanti |
| **FORM_CONTINUE** $ | Ressources provenant de la formation continue en euros | | quanti |
| **SUB_EUR** $ | Subventions européennes en euros | | quanti |
| **SUB_REGION** $ | Subventions de la région en euros | | quanti |
| **SUB_AUTRES** $ | Autres subventions en euros | | quanti |
| **RESS_AUTRES** $ | Autres ressources (valorisation, prestations,…) en euros | | quanti |

* données récupérées sur le site du SNESUP ; $ Données récupérées sur la plateforme de données publiques française : data.gouv

***Tableau 1*** *: Description des variables de l'étude. Les variables correspondant à des proportions ont été calculées à partir des données récupérées sur la plateforme de données publiques, à l'exception du taux de permanents, qui était disponible tel quel dans les fichiers téléchargés sur le site.*

❖ Présentation des individus

Les individus utilisés dans cette étude, correspondent à l'ensemble des universités et écoles pour lesquelles, les données étaient disponibles. Le tableau disponible dans l'annexe 1, précise le nom complet des établissements étudiés, ainsi que la nature de ceux-ci.
A noter que l'université Rennes 1, apparaissant dans les différents fichiers téléchargés, a été retirée de l'étude parce qu'elle correspondait au regroupement de plusieurs universités et écoles également utilisées (ENS Rennes, INSA Rennes, ENSC Rennes et Sciences-Po Rennes).
Un total de 97 individus pour 23 variables a ainsi été obtenu.

### Objectif de l'étude

L'objectif de cette étude est de mettre en pratique les méthodes d'analyses descriptives uni-, bi- et multivariées abordées dans le cadre de l'enseignement du module STA 101, afin de mettre en évidence des groupes d'établissements d'enseignement supérieur aux profils similaires et d'identifier les variables sur lesquelles reposent ces ressemblances / différences. Enfin, nous explorerons également l'effet de la nature des établissements (types d'universités ou école) sur ces ressemblances.
Pour cela, après avoir décrit l'ensemble des variables à l'aide d'indicateurs simples, nous procéderons à l'analyse des liens entre chaque paire de variables (analyse bivariée), avant de procéder à une Analyse en Composantes Principales (ACP). En effet, notre jeu de données ne comportant que deux variables qualitatives, dont une géographique, cette dernière méthode est la mieux adaptée.

## DESCRIPTION ÉLÉMENTAIRE

### Prétraitement des données

Dans un premier temps, il est à noter que dans les ressources utilisées pour construire le jeu de données, les universités UPC, Lyon1 et Toulouse 3 sont affichées comme appartenant aux USTS (Universités scientifiques et/ou médicales) alors qu'elles apparaissent sur leur site comme étant des universités pluridisciplinaires avec secteur santé (UPavS). Nous avons donc modifié en conséquence, la variable TYPE_U pour ces trois établissements.
Par ailleurs, la répartition des établissements toujours en fonction de la variable TYPE_U, met en évidence un faible nombre d'universités UTALLSHS et UTDEG (tableau 2a). Et dans la mesure où ces disciplines sont souvent rassemblées dans une même faculté au sein des universités pluridisciplinaires, nous avons choisi de les regrouper en une seule catégorie : USH (Université Sciences Humaines). De même, le déséquilibre entre les effectifs des écoles en fonction de leur nature, nous a conduit à faire le choix de les regrouper sous une même modalité : ECOLES. Le tableau 2b, correspond à la répartition finale de la variable TYPE_U.

| | | |
|---|---|---|
| UNIVERSITES | UPavS | 34 |
| | UPhs | 21 |
| | UTALLSHS | 8 |
| | UTDEG | 4 |
| ECOLES | AUTRES | 4 |
| | EI | 21 |
| | GE | 5 |

***Tableau 2a** : Répartition initiale des établissements selon la variable TYPE_U*

| | | |
|---|---|---|
| UNIVERSITES | UPavS | 34 |
| | UPhs | 21 |
| | USH | 12 |
| ECOLES | ECOLES | 30 |

***Tableau 2b** : Répartition finale des établissements selon la variable TYPE_U.*

## Analyse univariée

### Variables qualitatives

Les occurrences associées à la variable TYPE_U, ont été déterminées dans le paragraphe précédent (Tableaux 2a et 2b). Une répartition plus ou moins uniforme pour l'ensemble des modalités est observée après regroupement des modalités mentionnées dans le paragraphe précédent.

Pour ce qui est de la variable ACAD, le tableau 3, présente la répartition des établissements en fonction de l'académie de rattachement.

Dans la mesure où, le nombre d'établissements par académie est très inégal, avec plusieurs d'académies ne comportant qu'un seul établissement, cette variable ne sera pas analysée par la suite.

| | | | | |
|---|---|---|---|---|
| AIX-MARSEILLE | 3 | | MONTPELLIER | 5 |
| AMIENS | 2 | | NANCY-METZ | 1 |
| BESANÇON | 2 | | NANTES | 4 |
| BORDEAUX | 3 | | NICE | 2 |
| CLERMONT-FERRAND | 2 | | NOUVELLE-CALEDONIE | 1 |
| CORSE | 1 | | ORLEANS-TOURS | 3 |
| CRETEIL | 4 | | PARIS | 11 |
| DIJON | 1 | | POITIERS | 3 |
| GRENOBLE | 3 | | POLYNESIE FRANÇAISE | 1 |
| GUADELOUPE | 1 | | REIMS | 2 |
| GUYANE | 1 | | RENNES | 6 |
| LA REUNION | 1 | | ROUEN | 5 |
| LILLE | 5 | | STRASBOURG | 3 |
| LIMOGES | 1 | | TOULOUSE | 6 |
| LYON | 6 | | VERSAILLES | 8 |

***Tableau 3*** *: Répartition des établissements selon la variable ACAD.*

### Variables quantitatives

Dans un premier temps, les variables sont passées en revue à l'aide d'une représentation graphique sous formes d'histogrammes (Annexe 2). Cette première inspection des données nous permet de voir que pour la plupart des données, l'hypothèse d'une distribution normale ne semble pas vérifiée. En effet, seules les variables correspondant à des proportions (X.AM2D, X.MCF, X.PR, X_FEMMES_BIATSS et dans une moindre mesure, X.FEMMES) semblent suivre une loi normale.

Pour l'ensemble des autres variables, nous observons une grande fréquence pour les valeurs les plus faibles, à l'exception des variables TAUX_PERM, correspondant au pourcentage de permanents parmi l'ensemble du personnel des établissements, et X.ITRF, correspondant à la proportion d'ITRF parmi les BIATSS. Pour ces deux variables, de grandes fréquences sont observées pour les plus grandes valeurs de ces variables.

Ceci est confirmé à l'aide des principaux indicateurs de position et de dispersion déterminés pour chacune des variables (tableau 4). On constate ainsi que si la valeur minimale observée est de 18,09 %, 50 % des établissements ont un taux de permanents compris entre 78,46 % (= Q1) et 83,55 % (= Q3), très proche du maximum observé (87,6 %). De même pour la variable X.ITRF, 50 % des établissements présentent une proportion d'ITRF parmi les BIATSS comprise entre 74 % (= Q1) et 89 % (= Q3), avec une valeur minimale de 9 % et une valeur maximale de 100 %.

| Variable | N | Min | Q1 | Médiane | Moyenne | Q3 | Max | Ecart-type |
|---|---|---|---|---|---|---|---|---|
| MAS_SAL | 97 | 4412403 | 36425602 | 78524163 | 113305642 | 151833439 | 469173152 | 110555701 |
| ETU | 97 | 469 | 3089 | 13201 | 18554,9 | 27092 | 73161 | 17930,76 |
| SCSP_ETU | 97 | 3647,66 | 6028,04 | 6929,83 | 9992,43 | 10183,54 | 67187 | 9585,57 |
| ETPT_100ETU | 97 | 4,05 | 7,5 | 8,66 | 13,33 | 13,25 | 116,28 | 16,44 |
| ENS | 97 | 81 | 445 | 811 | 1061,55 | 1456 | 3931 | 838,37 |
| X.AM2D | 97 | 0 | 0,16 | 0,2 | 0,21 | 0,26 | 0,38 | 0,08 |
| X.MCF | 97 | 0,33 | 0,49 | 0,51 | 0,52 | 0,55 | 0,67 | 0,06 |
| X.PR | 97 | 0,07 | 0,22 | 0,27 | 0,28 | 0,33 | 0,59 | 0,08 |
| X.FEMMES | 97 | 0,13 | 0,31 | 0,39 | 0,37 | 0,41 | 0,64 | 0,1 |
| BIATSS | 90 | 3 | 347,8 | 654 | 953,99 | 1189 | 3979 | 973,98 |
| X.ITRF | 90 | 0,09 | 0,74 | 0,81 | 0,79 | 0,89 | 1 | 0,15 |
| X.FEMMES_BIATSS | 90 | 0,45 | 0,61 | 0,65 | 0,64 | 0,67 | 1 | 0,07 |
| TAUX_PERM | 95 | 18,09 | 78,46 | 81,75 | 80,11 | 83,55 | 87,6 | 7,52 |
| ANR | 96 | 0 | 744844,2 | 1999963 | 7800879 | 5080552 | 101602434 | 15882479 |
| DROITS_INS | 96 | 73884 | 914438,2 | 2394965 | 3338365 | 4412244 | 14488573 | 3175152 |
| APPRENTIS | 92 | 871 | 271781,8 | 558493,5 | 824753,1 | 923261,2 | 10850922 | 1233645 |
| FORM_CONTINUE | 94 | 0 | 855493,5 | 3353818 | 5976335 | 9092012 | 38938157 | 6925123 |
| SUB_EUR | 95 | 0 | 425781,5 | 1489900 | 2421340 | 3066524 | 14254802 | 2823677 |
| SUB_REGION | 93 | 0 | 660511 | 1532367 | 2773716 | 3654740 | 15443323 | 3228572 |
| SUB_AUTRES | 95 | 112426 | 2362971 | 5087500 | 8739531 | 10276264 | 51844385 | 9930356 |
| RESS_AUTRES | 97 | 496524 | 2436060 | 5516092 | 8135002 | 9212361 | 64536777 | 9423019 |

***Tableau 4*** *: Principaux indicateurs de position (**N** = nombre de valeurs ; **Min** = valeur minimale observée ; **Q1** = 1er quartile ; **Médiane** ; **Moyenne** ; **Q3** = 3ème quartile ; **Max** = valeur maximale observée), et de dispersion (**Ecart-type**), des différentes variables quantitatives.*

Certains établissements sont caractérisés par des valeurs aberrantes pour l'ensemble de ces variables, comme l'attestent les boîtes à moustache présentées en Annexe 3. Seule la variable X.AM2D présente une boîte à moustaches bien symétrique, sans « outlier ». Les différentes politiques de fusion des universités menées récemment, peuvent potentiellement expliquer ces valeurs. Nous avons fait le choix de les conserver, en considérant qu'une partie de l'information intéressante est contenue dans les valeurs prises par ces « outliers », un des objectifs de l'étude étant justement de voir si certains établissements, en fonction de leur nature, sont caractérisés par des profils particuliers.
Par ailleurs, les unités des différentes variables n'étant pas les mêmes, avec des valeurs des indicateurs très différents selon la variable, il apparait nécessaire de centrer-réduire les données avant l'analyse multivariée.
On peut également remarquer que le nombre d'observations pour les différentes variables n'est pas le même, mettant en évidence la présence de données manquantes (NA). Au vu du faible nombre de celles-ci (pourcentage de NA compris entre 0 et 7 % selon les variables), nous avons fait le choix de conserver tous les individus et d'imputer les valeurs manquantes à l'aide du package *missMDA*[3] pour l'analyse multivariée.

### Analyse bivariée

Comme mentionné dans l'analyse univariée, la variable ACAD étant une variable géographique avec plusieurs d'académies ne comportant qu'un seul établissement, le lien de cette variable avec l'autre variable qualitative ou les variables quantitatives, ne sera pas étudié.

[3] http://factominer.free.fr/missMDA/index_fr.html

### Corrélation entre paires de variables quantitatives

L'étude du lien entre les variables quantitatives commence par une exploration graphique des nuages de points pour les différentes paires de variables. Dans un souci de clarté, seules quelques représentations graphiques sont exposées dans ce rapport (Figure 1), le nombre total de variables étant relativement important.

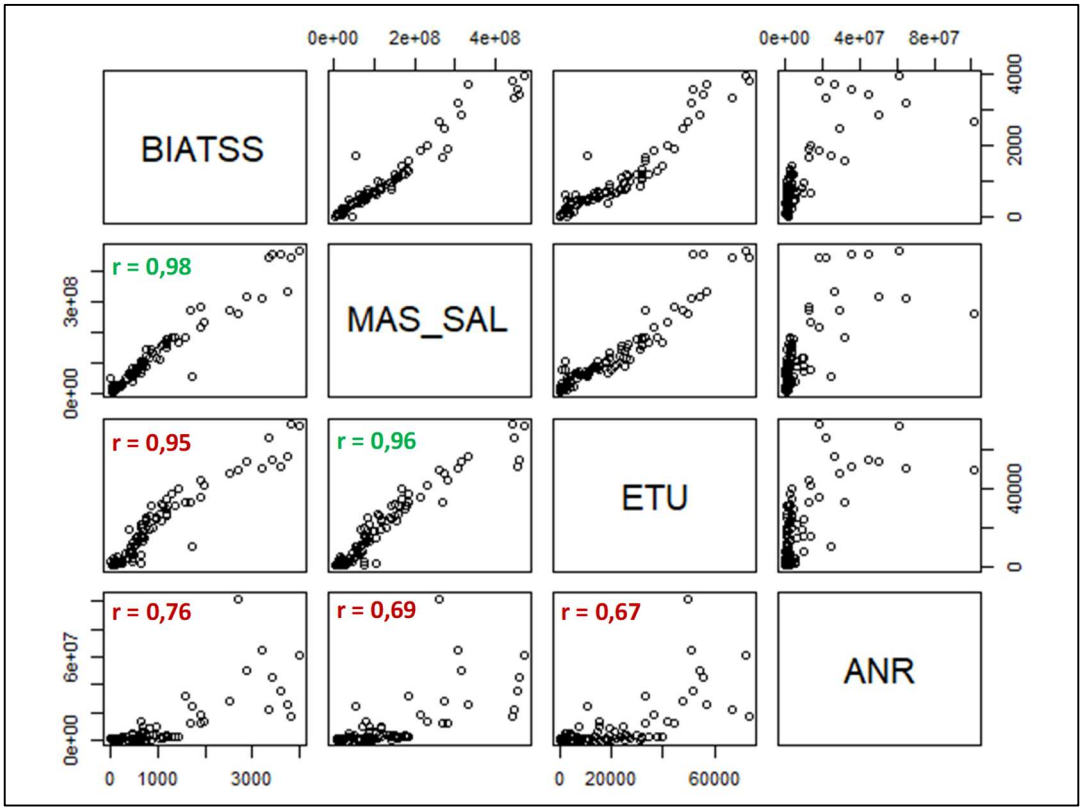



***Figure 1 :*** *Matrice de la représentation graphique des nuages de points de certaines paires de variables. Le coefficient de corrélation de Pearson est indiqué dans la partie basse de la matrice (en vert, pour une dépendance d'apparence linéaire, et en rouge pour les paires de variables entre lesquelles un lien linéaire ne semble pas avéré).*

La représentation graphique utilisée dans la figure 1 montre que les variables BIATSS et MAS_SAL, ou ETU et MAS_SAL, semblent bien présenter une liaison linéaire, en accord avec la valeur élevée du coefficient de corrélation de Pearson calculé pour ces différentes paires de variables. Ce n'est pas le cas des autres paires de variables représentées dans la figure 1 pour lesquelles le nuage de points ne permet pas de conclure sans ambiguïté, à une liaison linéaire, alors que les coefficients de Pearson correspondants sont relativement élevés.

Par ailleurs, il est important de souligner que le nombre d'individus de l'étude reste relativement faible et que quelques établissements présentent des valeurs extrêmes (souvent plus élevées) pour certaines variables. L'utilisation du coefficient de Pearson étant sensibles aux valeurs aberrantes, cela pourrait conduire à un biais et fausser l'interprétation.

D'autre part, l'analyse univariée a montré qu'une grande partie des variables étudiées, ne présentaient pas une distribution « normale ». Une comparaison visuelle de la distribution de nos variables avec une distribution « normale » est représentée graphiquement sous la forme de Q-Q plot (Annexe 4) et permet de mieux se rendre compte de l'écart à la normalité de la majorité de nos variables.

Si pour la variable X.AM2D, l'ensemble des points semble bien suivre la distribution théorique, certaines variables, présentent quant à elles, des distributions plutôt bien éloignées de la distribution « normale » avec des queues clairement en dehors de ce qui serait attendu si elles avaient une distribution « normale » : soit beaucoup plus de points aux extrémités. Certaines variables présentent

une situation intermédiaire : à savoir des points ne s'éloignant que légèrement de la distribution « normale » (les variables TAUX-PERM et X.PR)
Pour illustrer ces trois cas de figure, les Q-Q plot de trois variables montrant respectivement une bonne, moyenne et mauvaise adéquation avec une distribution « normale, sont représentés dans la figure 2 ci-après.

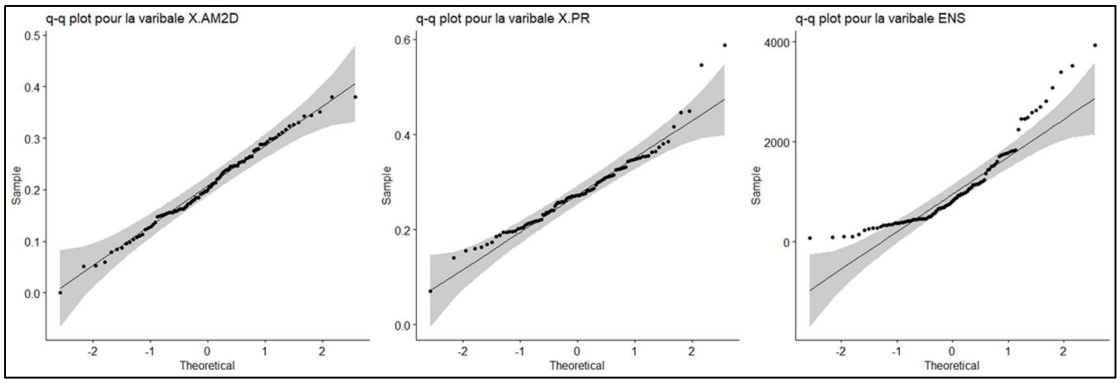



***Figure 2 :*** *Q-Q plot des variables X.AM2D (gauche), X.PR (milieu) et ENS (droite), illustrant respectivement une bonne, moyenne et mauvaise adéquation avec une distribution « normale ».*

Pour confirmer cette observation, un test de normalité a été effectué (test de Shapiro-Wilk). Les p-valeurs obtenues sont répertoriées dans le tableau 5 ci-après et confirment l'absence de normalité de la distribution des variables à un risque d'erreur de 5 %, pour toutes les variables à l'exception de X.AM2D (p-valeur non significative, *i.e,* supérieure à la valeur seuil de 5 %).

| test de Shapiro | p-valeurs |
|---|---|
| MAS_SAL | 4,5E-10 |
| ETU | 9,0E-08 |
| SCSP_ETU | 2,9E-16 |
| ETPT_100ETU | 0,0E+00 |
| ENS | 5,3E-08 |
| X.AM2D | 9,0E-01 |
| X.MCF | 8,0E-03 |
| X.PR | 1,1E-03 |
| X.FEMMES | 1,5E-03 |
| BIATSS | 3,8E-10 |
| X.ITRF | 1,1E-06 |
| X.FEMMES_BIATSS | 1,6E-06 |
| TAUX_PERM | 4,6E-16 |
| ANR | 2,6E-16 |
| DROITS_INS | 1,6E-08 |
| APPRENTIS | 0,0E+00 |
| FORM_CONTINUE | 2,9E-10 |
| SUB_EUR | 9,3E-11 |
| SUB_REGION | 2,0E-10 |
| SUB_AUTRES | 2,7E-11 |
| RESS_AUTRES | 4,9E-13 |

***Tableau 5 :*** Résultats du test de normalité de Shapiro-Wilk : p-valeurs obtenues pour chaque variable.

L'ensemble de ces points, ajouté au fait que nous n'ayons pas observé de variation non monotone du nuage de points, nous a conduit à faire le choix de l'utilisation de la corrélation de Spearman, plus

robuste et dont le test de significativité ne repose pas sur l'hypothèse de normalité des distributions des variables.
La figure 3, montre les résultats obtenus pour l'ensemble de nos variables.
Ainsi, pour ce qui est des corrélations positives, on observe que les variables MAS_SAL, ETU et BIATSS présentent un lien monotone fort avec des coefficients de corrélation, ρ, proches de 1 (0,97 et 0,99).
De même, la dotation et le nombre d'équivalents temps plein semblent être très liées (*i.e.*, variables SCSP_ETU et ETPT_100ETU) avec un coefficient de corrélation de 0,98.
Les variables associées aux revenus des établissements (ANR, DROITS_INS, APPRENTIS, FORM_CONTINUE, SUB_EUR, SUB_REGION, SUB_AUTRES et RESS_AUTRES) semblent présenter un lien assez fort les unes avec les autres (coefficients de corrélation compris entre 0,49 et 0,85), ainsi qu'avec les variables MAS_SAL et ETU (ρ = 0,84 et 0,85).
Des corrélations négatives sont également observées entre les variables SCSP_ETU / EPTP_100ETU et la variable ETU (avec un coefficient de corrélation respectivement égal à - 0,59 et - 0,53), mais avec un lien moins fort en valeur absolue. Ce qui peut s'expliquer par le fait que ces variables sont obtenues en divisant la dotation globale (SCSP) / le nombre d'équivalents temps plein, par le nombre d'étudiants / 100 étudiants. Ces deux variables sont également liées de façon négative, à la proportion de femmes dans le personnel permanent d'enseignement et/ou de recherche (X.FEMMES, avec ρ = - 0,68 et - 0,66 respectivement). Enfin, les variables X.AM2D et X.PR/X.MCF semblent négativement corrélées avec un ρ = - 0,73 et - 0,53 respectivement.
Pour l'ensemble de ces variables liées, nous avons fait le choix de n'en exclure aucune. En effet, elles ne mesurent pas la même chose et peuvent être importantes pour la mise en évidence de profils similaires / différents.

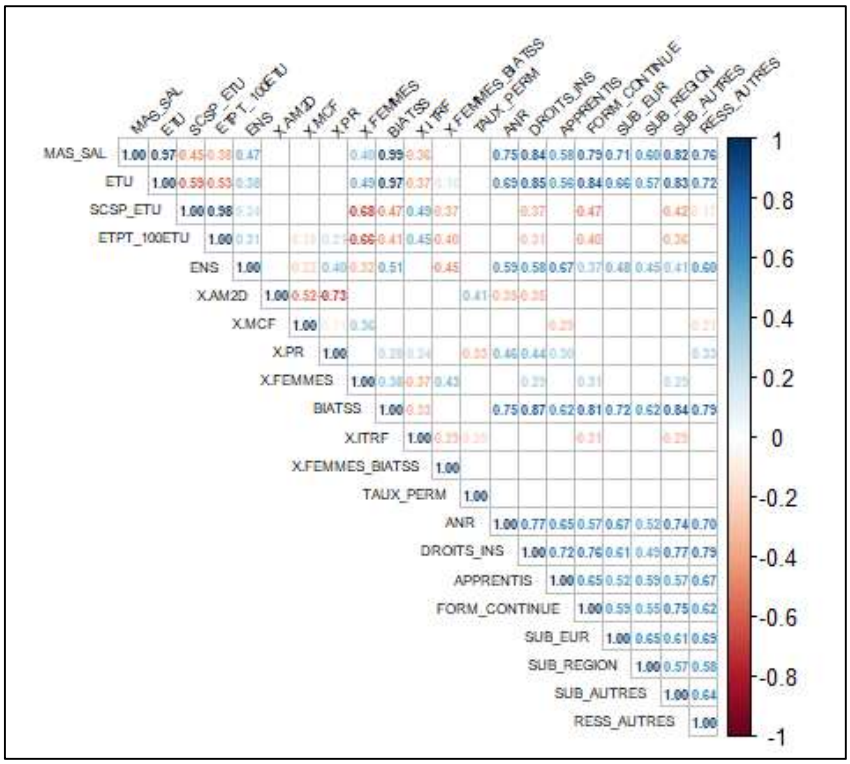
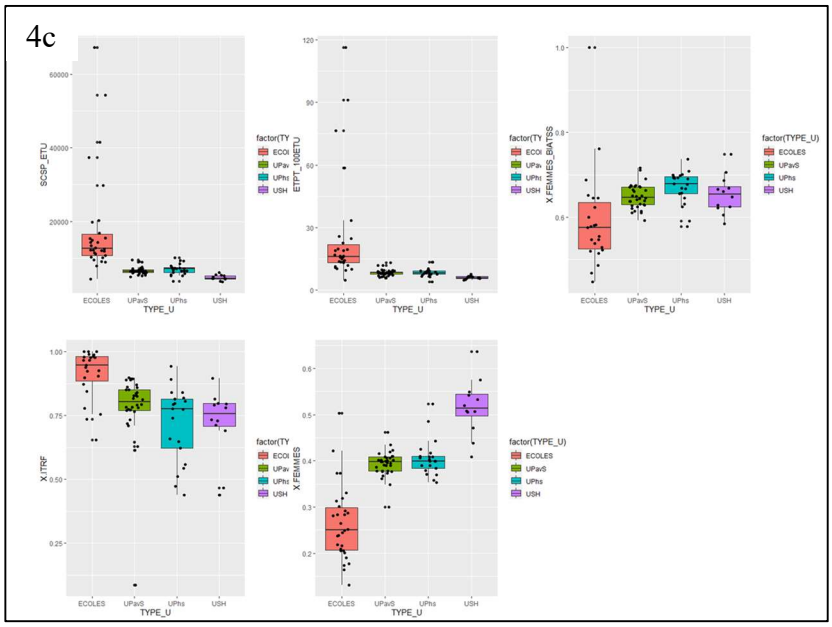


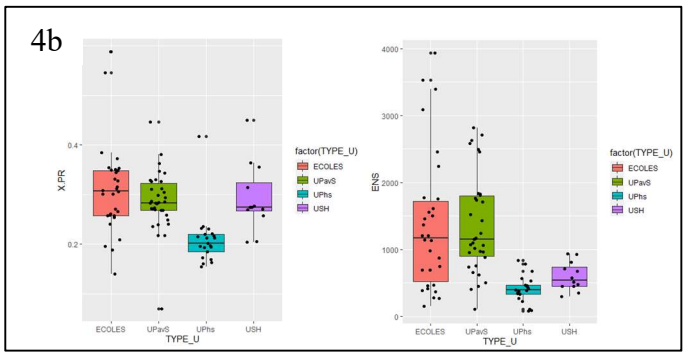

***Figure 3 :*** *Carte des coefficients de corrélation de Spearman entre les paires de variables. La détermination de ces coefficients s'est faite en éliminant les individus présentant des données manquantes à l'aide de la commande « use = "complete.obs" ». Seuls les coefficients significatifs au seuil d'erreur de 5 %, apparaissent dans cette table. La liaison entre deux variables est d'autant plus forte que les couleurs des nombres sont prononcées.*

### Corrélation entre la valeur qualitative TYPE_U et les variables quantitatives

La mesure de la corrélation entre la variable qualitative TYPE_U et l'ensemble des variables quantitatives pourrait théoriquement s'effectuer à l'aide de la détermination du rapport de corrélation η, suivi de tests de nullité. Cependant, la majorité de ces tests font l'hypothèse de l'égalité des variances et d'une distribution normale des données, ce qui n'est pas le cas ici. Nous avons donc choisi d'étudier graphiquement cette corrélation, à l'aide de boîtes à moustaches (Figures 4).

L'inspection des graphiques montre que certaines variables comme TAUX_PERM, X.AM2D, X.MCF ou APPRENTIS, semblent présenter des distributions similaires, peu dépendantes de la nature de l'établissement (Fig. 4a). Le nombre d'enseignants / enseignants chercheurs permanents ou la proportion de professeurs parmi ceux-ci, semblent plus faibles dans les établissements pluridisciplinaires hors secteur santé (Fig 4b). D'un autre côté, les écoles présentent également quelques variables dont les valeurs semblent différentes de celles des autres établissements (Fig. 4c) : par exemple, la proportion de femmes parmi les enseignants / enseignants chercheurs ou parmi les BIATSS permanents (variables X.FEMMES et X.FEMMES_BIATSS) qui semblent plus faibles dans ce cas. Enfin, les universités pluridisciplinaires avec secteur santé, semblent présenter des valeurs particulières de certaines variables : principalement, celles associées aux ressources (Fig. 4d).



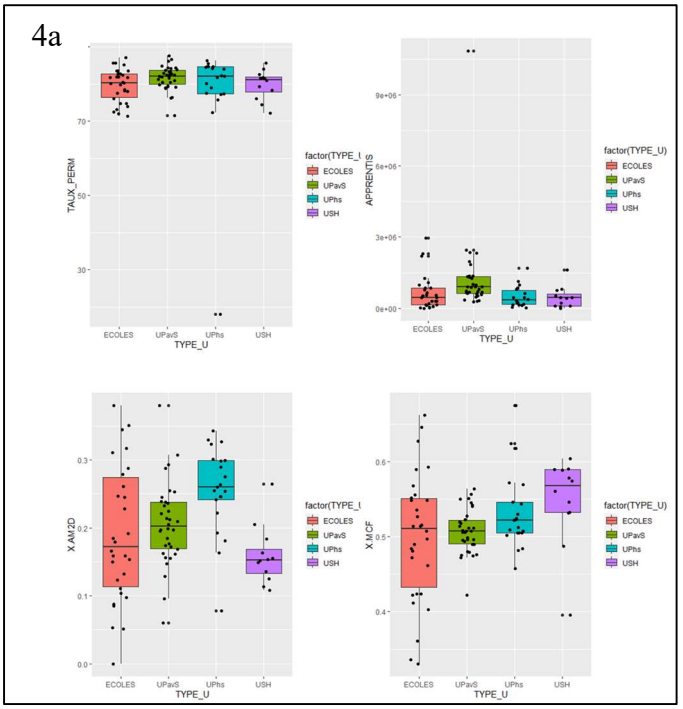
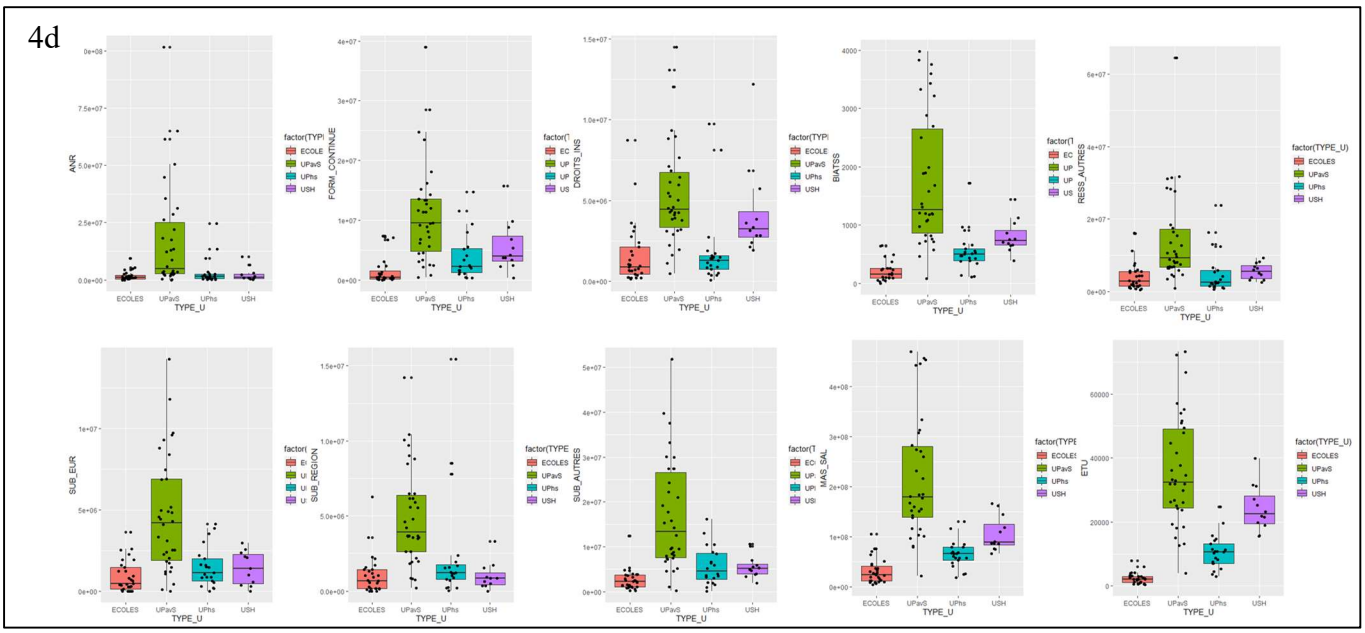




***Figure 4 :*** *Boîtes à moustache représentant la distribution des variables quantitatives en fonction de la variable TYPE_U.* ***Fig. 4a*** *: Variables de distributions similaires selon la nature de l'établissement ;* ***Fig. 4b*** *: Variables pour lesquelles les UPhs semblent présenter des valeurs particulières (plus basses) ;* ***Fig. 4c*** *: Variables pour lesquelles les ECOLES semblent présenter des valeurs particulières ;* ***Fig. 4d :*** *Variables pour lesquelles les UPavS semblent présenter des valeurs particulières.*



## ANALYSE MULTIVARIÉE : ANALYSE EN COMPOSANTE PRINCIPALE (ACP)

### Mise en œuvre de l'ACP

#### Choix des variables

L'objectif de l'étude multivariée est d'identifier des profils similaires entre différents établissements en fonction de la source de leurs revenus, mais également en fonction de caractéristiques propres à ce type d'établissements, à savoir : le nombre d'étudiants, le nombre de personnel permanent, et la répartition du personnel en fonction du sexe et des différentes catégories associées.
Pour ce qui est des variables à sélectionner, nous remarquons, qu'à l'exception de l'établissement PSL, pour lequel nous observons un taux de permanents particulièrement faible (18,1 %), l'ensemble des valeurs pour les autres établissements varie dans une très faible gamme.
La variable MAS_SAL représente, le montant en euros de la masse salariale, soit un coût plutôt qu'un revenu. Dans la mesure où, nous avons par ailleurs le nombre d'enseignants / enseignants chercheurs et de BIATSS, cette variable ne nous est pas apparue indispensable.
Nous avons donc décidé de conserver ces deux variables mais en tant que variables illustratives.
De même, les deux variables qualitatives seront utilisées comme variables illustratives.

#### Prétraitement des données

Les données recueillies n'ont pas toutes les mêmes unités. Elles seront donc centrées et réduites avant l'Analyse en Composantes Principales (ACP).
Par ailleurs, quelques variables n'étaient pas renseignées pour certains établissements. Dans la mesure où le nombre de ces données manquantes est faible, nous avons fait le choix de compléter le jeu de données à l'aide d'une méthode d'imputation plutôt que de supprimer les individus incomplets. L'imputation a été effectuée à l'aide du package *missMDA*[4] qui repose sur l'utilisation d'ACP itératives et permet de ne pas apporter de poids aux valeurs imputées dans la détermination des axes factoriels en ACP.
Les lignes de codes utilisées pour l'ACP et pour la classification ont été récupérées à partir de l'application factoshiny[5].

#### Choix du nombre de dimensions

Le résumé de l'ensemble de nos données à l'aide de l'ACP a conduit à 19 axes factoriels. Le nombre d'axes à retenir pour l'interprétation peut être effectué de différentes manières. Bien qu'une seule règle suffirait, deux règles seront abordées ici pour vérifier la convergence des méthodes dans le choix du nombre d'axes.

- Règle de Kaiser
  Les données ayant été centrées et réduites, le choix du nombre d'axes peut se faire en ne conservant que ceux dont l'inertie (valeur propre) est supérieure à 1. Ainsi, les composantes retenues auront une valeur explicative supérieure à celui d'une variable initiale.
  Dans notre cas, cela revient à ne retenir que les 4 premiers axes (Tableau 6).

[4] http://factominer.free.fr/missMDA/index_fr.html
[5] http://factominer.free.fr/graphs/factoshiny-fr.html

| Résultats ACP | valeurs propres | % de variance | % variance cumulées |
|---|---|---|---|
| comp 1 | 7,42 | 39,05 | 39,05 |
| comp 2 | 2,97 | 15,64 | 54,69 |
| comp 3 | 1,68 | 8,86 | 63,55 |
| comp 4 | 1,47 | 7,72 | 71,27 |
| comp 5 | 0,94 | 4,97 | 76,23 |
| comp 6 | 0,86 | 4,53 | 80,77 |
| comp 7 | 0,71 | 3,74 | 84,51 |
| comp 8 | 0,56 | 2,94 | 87,45 |
| comp 9 | 0,53 | 2,80 | 90,25 |
| comp 10 | 0,42 | 2,18 | 92,44 |
| comp 11 | 0,41 | 2,18 | 94,62 |
| comp 12 | 0,33 | 1,72 | 96,34 |
| comp 13 | 0,26 | 1,38 | 97,72 |
| comp 14 | 0,16 | 0,85 | 98,57 |
| comp 15 | 0,15 | 0,77 | 99,34 |
| comp 16 | 0,09 | 0,47 | 99,82 |
| comp 17 | 0,03 | 0,14 | 99,96 |
| comp 18 | 0,01 | 0,04 | 100,00 |
| comp 19 | 0,00 | 0,00 | 100,00 |

***Tableau 6 :*** *Tableau récapitulatif des valeurs propres, pourcentage de variance et pourcentage de variance cumulée expliqués par chacune des composantes de l'ACP obtenue. En vert, les axes sélectionnées pour l'interprétation.*

- Règle du coude

  Le diagramme des éboulis correspondant au pourcentage d'inertie expliquée par chaque axe est représenté sur la figure 5. On observe bien une cassure entre les axes 3 et 4 (et dans une moindre mesure entre les axes 5 et 6).

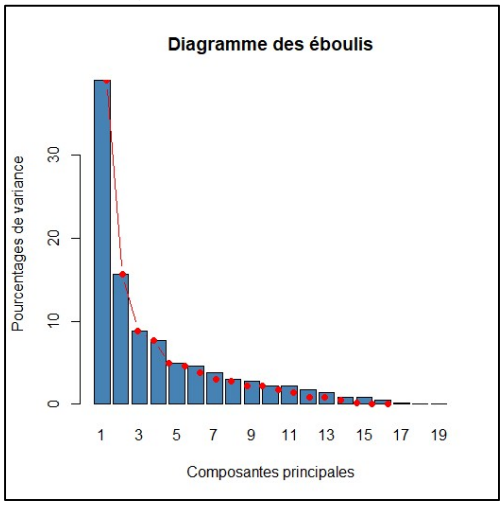



***Figure 5 :*** *Représentation du diagramme des éboulis.*

Les deux méthodes conduisent à la même conclusion. Nous retiendrons donc les 4 premiers axes pour l'interprétation des résultats. Nous avons pu ainsi résumer environ 71 % de l'information à l'aide de 4 composantes uniquement.

Avant d'interpréter les différents axes, il est nécessaire de sélectionner les variables les mieux projetées sur les différents plans à l'aide de différents paramètres tels que le coefficient de corrélation entre chaque variable quantitative et les différents axes (tableau 7a) et le rapport de corrélation entre

les variables qualitatives illustratives et les coordonnées des individus sur les différents axes (tableau 7b). Même si la p-valeur associée au test de corrélation des variables avec les différents axes n'est pas interprétables pour les variables actives, le tri des valeurs qu'elle prend permet de sélectionner les variables les plus importantes.
Les résultats de la variable ACAD sont affichés, mais ne seront pas interprétés par la suite en raison du faible nombre d'individus par modalité.

| QUANTI. | cor.1 | cor.2 | cor.3 | cor.4 |
|---|---|---|---|---|
| ETU | 0,92* **(3)** | -0,16 | -0,03 | 0,12 |
| SCSP_ETU | -0,32* **(17)** | 0,80* **(1)** | 0,05 | 0,35 |
| ETPT_100ETU | -0,27* **(16)** | 0,78* **(2)** | 0,02 | 0,40 |
| ENS | 0,54* **(12)** | 0,53* **(6)** | -0,09 | -0,17 |
| X.AM2D | -0,26* **(15)** | -0,05 | -0,91* **(1)** | 0,25 |
| X.MCF | -0,07 | -0,60* **(4)** | 0,28* **(4)** | -0,40 |
| X.PR | 0,30* **(13)** | 0,51* **(5)** | 0,69* **(2)** | 0,06 |
| X.FEMMES | 0,28* **(14)** | -0,63* **(3)** | 0,26* **(7)** | 0,42 |
| BIATSS | 0,97* **(1)** | 0,02 | 0,00 | 0,04 |
| X.ITRF | -0,14 | 0,40* **(7)** | 0,11 | -0,54 |
| X.FEMMES_BIATSS | 0,05 | -0,40* **(8)** | 0,25* **(6)** | 0,61 |
| ANR | 0,78* **(9)** | 0,13 | 0,04 | -0,04 |
| DROITS_INS | 0,85* **(5)** | 0,03 | 0,07 | 0,05 |
| APPRENTIS | 0,61* **(11)** | 0,18 | -0,13 | -0,17 |
| FORM_CONTINUE | 0,80* **(8)** | -0,04 | -0,04 | 0,07 |
| SUB_EUR | 0,80* **(7)** | 0,04 | -0,12 | 0,05 |
| SUB_REGION | 0,63* **(10)** | -0,01 | -0,31* **(5)** | -0,16 |
| SUB_AUTRES | 0,92* **(4)** | 0,03 | -0,05 | 0,03 |
| RESS_AUTRES | 0,81* **(6)** | 0,12 | -0,04 | 0,02 |
| MAS_SAL | 0,94* **(2)** | - | - | - |
| TAUX_PERM | - | - | -0,35* **(3)** | - |

***Tableau 7a :*** *Tableau récapitulatif des coefficients de corrélation entre chaque variable quantitative et les différents axes retenus dans l'analyse.* ***Gris :*** *variables quantitatives supplémentaires.* ******* *coefficients de corrélation pour lesquels une p-valeur a été obtenue. Le classement des variables en fonction de la p-valeur apparait en rouge entre parenthèses.*

| VAR = TYPE_U | $\eta_1$ = 0,46 | $\eta_2$ = 0,54 | $\eta_3$ = 0,17 |
|---|---|---|---|
| QUALI. | Estim.dim1 | Estim.dim2 | Estim.dim3 |
| TYPE_U=UPavS | 2,59 | - | - |
| ACAD=NANCY-METZ | 7,72 | - | - |
| TYPE_U=UPhs | -0,96 | -0,97 | -0,74 |
| TYPE_U=ECOLES | -1,66 | 2,05 | - |
| TYPE_U=USH | - | -1,23 | 1,00 |
| ACAD=PARIS | - | - | 1,84 |
| ACAD=CORSE | - | - | -2,17 |

***Tableau 7b :***
*Tableau récapitulatif des corrélations significatives entre les variables qualitatives (rapport de corrélation* η*) ou modalités de celles-ci (estimation moyenne) et chaque axe. La significativité repose sur une v-test > 2).*

La qualité de représentation peut être évaluée également à l'aide du cos² lié à la projection des variables sur les différents axes (Tableau 8), ou encore à l'aide des coordonnées de celles-ci sur chaque dimension.

| QUANTI. | cos2.1 | cos2.2 | cos2.3 | cos2.4 | contrib.1 | contrib.2 | contrib.3 | contrib.4 |
|---|---|---|---|---|---|---|---|---|
| ETU | 0,86 | 0,02 | 0,00 | 0,01 | 11,53 | 0,83 | 0,06 | 0,94 |
| SCSP_ETU | 0,10 | 0,64 | 0,00 | 0,12 | 1,38 | 21,48 | 0,14 | 8,28 |
| ETPT_100ETU | 0,07 | 0,62 | 0,00 | 0,16 | 1,01 | 20,69 | 0,03 | 11,11 |
| ENS | 0,29 | 0,28 | 0,01 | 0,03 | 3,95 | 9,55 | 0,48 | 1,86 |
| X.AM2D | 0,07 | 0,00 | 0,83 | 0,06 | 0,88 | 0,07 | 49,45 | 4,29 |
| X.MCF | 0,00 | 0,37 | 0,08 | 0,16 | 0,06 | 12,29 | 4,57 | 11,11 |
| X.PR | 0,09 | 0,26 | 0,48 | 0,00 | 1,24 | 8,80 | 28,29 | 0,26 |
| X.FEMMES | 0,08 | 0,40 | 0,07 | 0,18 | 1,05 | 13,33 | 3,99 | 12,13 |
| BIATSS | 0,94 | 0,00 | 0,00 | 0,00 | 12,63 | 0,01 | 0,00 | 0,12 |
| X.ITRF | 0,02 | 0,16 | 0,01 | 0,30 | 0,27 | 5,26 | 0,78 | 20,20 |
| X.FEMMES_BIATSS | 0,00 | 0,16 | 0,06 | 0,37 | 0,04 | 5,38 | 3,79 | 25,27 |
| ANR | 0,61 | 0,02 | 0,00 | 0,00 | 8,20 | 0,58 | 0,10 | 0,12 |
| DROITS_INS | 0,73 | 0,00 | 0,01 | 0,00 | 9,84 | 0,02 | 0,33 | 0,19 |
| APPRENTIS | 0,37 | 0,03 | 0,02 | 0,03 | 4,99 | 1,04 | 0,94 | 1,87 |
| FORM_CONTINUE | 0,64 | 0,00 | 0,00 | 0,01 | 8,61 | 0,06 | 0,10 | 0,36 |
| SUB_EUR | 0,64 | 0,00 | 0,01 | 0,00 | 8,62 | 0,06 | 0,89 | 0,14 |
| SUB_REGION | 0,40 | 0,00 | 0,10 | 0,02 | 5,42 | 0,00 | 5,80 | 1,65 |
| SUB_AUTRES | 0,85 | 0,00 | 0,00 | 0,00 | 11,42 | 0,03 | 0,18 | 0,06 |
| RESS_AUTRES | 0,66 | 0,02 | 0,00 | 0,00 | 8,88 | 0,52 | 0,09 | 0,03 |

***Tableau 8 :*** *Tableau récapitulatif des cos² représentant la projection de celles-ci sur les différents axes et contributions des variables aux axes. Les variables les mieux représentées, sont celles correspondant à une valeur élevée de ce paramètre sur l'ensemble du plan considéré.*

## Analyses des résultats de l'ACP

### Premier plan (dimensions 1 et 2)

Le premier plan obtenu résume à lui seul, près de 55 % de la variance initiale (Tableau 6). Sachant que nous avons utilisé 19 variables actives pour l'ACP, ce pourcentage est à mettre en regard de celui qu'on aurait attendu avec deux dimensions avant l'ACP, soit 2 * (1/19) ≈ 10 %. L'ACP nous a donc permis de représenter plus de 50 % de l'information initiale sur un seul plan, cela peut venir du fait que les variables sélectionnées sont suffisamment corrélées entre elles pour nous permettre de réduire de façon importante le nombre de dimensions à utiliser pour résumer l'information.

- Analyse des variables

  Le cercle de corrélation des variables quantitatives correspondant aux deux premières dimensions de l'ACP est représenté sur la figure 6.

  On observe ainsi que le premier axe (F1) est caractérisé par de grandes valeurs pour plusieurs variables : BIATSS, ETU, SUB_AUTRES, DROITS_INS, RESS_AUTRES, FORM_CONTINUE, SUB_EUR, ANR, et SUB_REGION.

  Le second axe (F2), est quant à lui caractérisé par de grandes valeurs de SCSP_ETU et ETPT_100ETU.

  Les variables ENS et X.FEMMES sont dans une moindre mesure bien représentés également sur le plan, avec une somme de cos² = 0.57 et 0,48 respectivement (tableau 8). Elles sont caractérisées par une contribution modérée sur F1, et plus importante sur F2, avec de grandes valeurs de la variable ENS le long de F2, et au contraire une corrélation négative entre la variable X.FEMMES et F2 (tableau 8).

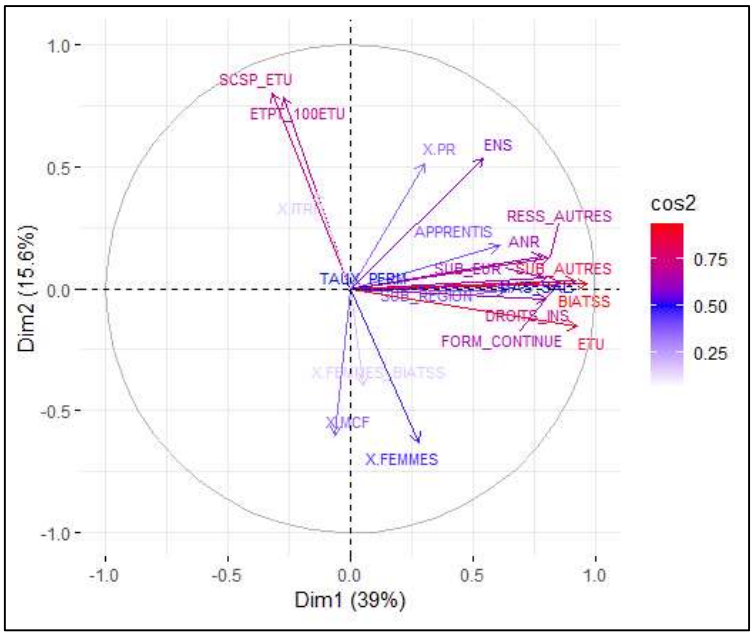



***Figure 6 :*** *Cercle de corrélation des variables pour le premier plan factoriel. La couleur des flèches est modulée par l'intensité de la corrélation de la variable avec l'axe concerné (voir légende sur le côté)*

On note également que la variable MAS_SAL, qui est une variable illustrative et pour laquelle la p-valeur associée au test de corrélation correspondant est particulièrement significative (p-valeur = $1{,}82.\ 10^{-45}$), est très corrélée positivement à F1.

Pour ce qui est des variables qualitatives (illustratives), seule la variable TYPE_U présente un rapport de corrélation significatif et relativement important pour ce plan, avec $\eta_1 = 0{,}46$ et $\eta_1 = 0{,}54$ (tableau 7b). Au sein de cette variable, les modalités UPavS, UPhs et ECOLES affichent en moyenne une coordonnée significativement positive le long de F1 pour UPavS et significativement, négatives pour UPhs et ECOLE. Pour F2, les modalités UPhs et USH prennent en moyenne des coordonnées négatives, alors que les individus associés à la modalité ECOLES ont en moyenne une coordonnée d'environ 2.
Pour la seconde variable qualitative, ACAD, seule la modalité NANCY-METZ affiche en moyenne de grande coordonnées le long de F1.

❖ Analyse des individus

Les tableaux 9 répertorient les valeurs de cos² entre les différents individus et les axes 1 et 2. Par souci de clarté, seuls les premiers individus de cos² plus élevés sont indiqués (l'ensemble de ces coefficients est disponible en Annexe 5a).

| ***Tableau 9a*** | Dim,1 | Dim,2 |
|---|---|---|
| Lille | 0,86 | 0,01 |
| Lorraine | 0,84 | 0,01 |
| AixMarseille | 0,81 | 0,01 |
| UPC | 0,79 | 0,03 |
| Lyon1 | 0,79 | 0,00 |
| Nantes | 0,74 | 0,00 |
| SorbonneU | 0,74 | 0,02 |
| Montpellier | 0,73 | 0,02 |
| Toulouse3 | 0,73 | 0,03 |
| INP Clermont | 0,72 | 0,05 |
| Clermont | 0,71 | 0,01 |
| Strasbourg | 0,70 | 0,00 |
| Saclay | 0,57 | 0,00 |
| Bordeaux | 0,57 | 0,02 |
| Grenoble | 0,56 | 0,01 |
| ENSC Montpellier | 0,56 | 0,04 |
| INSA val de loire | 0,53 | 0,04 |
| Antilles | 0,51 | 0,27 |

| ***Tableau 9b*** | Dim,1 | Dim,2 |
|---|---|---|
| Rennes2 | 0,13 | 0,65 |
| ENS Rennes | 0,24 | 0,61 |
| ENS Saclay | 0,14 | 0,58 |
| Lyon2 | 0,00 | 0,55 |
| Toulouse-Jaures | 0,01 | 0,52 |
| ENS Lyon | 0,07 | 0,51 |
| Montpellier3 | 0,01 | 0,50 |
| Nîmes | 0,27 | 0,44 |

*Tableaux récapitulatif des cos² représentant la projection des individus sur les différents axes. Les individus les mieux représentés, sont ceux correspondant à une valeur élevée de ce paramètre sur l'ensemble du plan.* ***Tableau 9a*** *: Etablissements les mieux représentés sur l'axe F1.* ***Tableau 9b*** *: Etablissements les mieux représentés sur l'axe F2.*

On voit ainsi que des individus tel que les universités de Lille, Lorraine, Aix-Marseille ou encore Lyon1, sont bien représentés sur l'axe F1.
De plus, la figure 7, nous indique que ces universités ont également de grandes coordonnées sur ce même axe.
D'un autre côté, les écoles normales supérieures (ENS Saclay, Lyon, Rennes et dans une moindre mesure Paris) sont quant à elles mieux représentées sur F2 et présentent de très grandes coordonnées sur ce même axe.
L'université de Nîmes est également mieux représentée sur F2, mais présente des coordonnées négatives sur ce même axe.

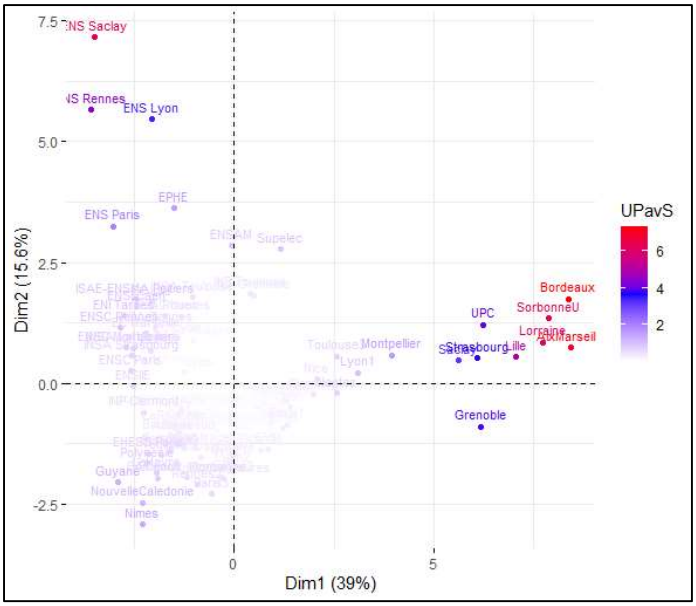



***Figure 7 :*** *Graphique des individus obtenu pour le premier plan factoriel. Les individus apparaissent avec une couleur modulée par l'intensité de la contribution à l'axe (voir légende sur le côté).*

❖ Synthèse du premier plan

En reprenant l'analyse des variables les plus contributives à ce plan, on peut en conclure que les individus associés à de grande coordonnées sur l'axe F1, sont ceux qui présentent de grandes valeurs pour l'ensemble des variables fortement corrélées (positivement) à cet axe. A savoir : MAS_SAL, BIATSS, ETU, SUB_AUTRES, DROITS_INS, RESS_AUTRES, FORM_CONTINUE, SUB_EUR, ANR, et SUB_REGION. Ce sont donc des établissements caractérisés par un nombre important de BIATSS et d'étudiants et qui ont l'avantage de présenter des revenus importants. Cela est vérifié sur le fichier source, avec notamment les universités correspondant aux points les plus à droite sur la figure 8 étant les 8 premières universités en matière de MAS_SAL, ETU, BIATSS.

Les écoles normales sont quant à elles caractérisées par de grandes valeurs de SCSP_ETU et ETPT_100ETU et plutôt de faibles valeurs pour les variables citées précédemment, ainsi que pour X.MCF. Elles semblent donc bénéficier de plus de dotation de l'état et de plus d'équivalents temps plein travaillés. Le graphique suggère également plus d'enseignants (nombre compris entre 265 et 3090), mais avec une proportion de maître de conférences plus faible (comprise entre 0,34 et 0,42) et à l'inverse, une proportion de professeur plus grande (comprise entre 0,26 et 0,38), le tout pour moins d'étudiants inscrits et moins de revenus alternatifs par ailleurs. Par ailleurs, elles semblent également caractérisées par une proportion de femmes parmi le personnel enseignant, plus faible (compris entre 0,17 et 0,37 pour un premier quartile associé à cette variable égal à 0,31) contrairement à l'université de Nîmes (X.FEMMES = 0,52, ENS = 86, X.PR = 0,16 et X.MCF = 0,67).

### Deuxième plan (dimensions 3 et 4)

Le second plan explique environ 17 % de variance supplémentaire (tableau 6), soit toujours plus qu'attendu avec deux dimensions sans ACP.

❖ Analyse des variables

Les variables les mieux projetées (corrélation significative) sur la dimension 3 (F3) sont les variables suivantes : X.AM2D, X.FEMMES_BIATSS, X.PR et dans une moindre mesure, X.MCF, X.FEMMES, X.ITRF et SUB_REGION.

Le cercle de corrélation des variables correspondant montre que les variables les mieux projetées sur ce plan sont celles qui ont également les plus grandes contributions (Figure 8).

La dimension 3 est principalement caractérisée par de fortes valeurs de X.PR et de faibles valeurs de X.AM2D. La dimension 4 (F4) est quant à elle associée à fortes valeurs de X.FEMMES_BIATSS, X.FEMMES et de faibles valeurs de X.ITRF et X.MCF notamment, mais cette corrélation n'est pas significative.

Parmi les variables illustratives, seule la variable TAUX_PERM présente une corrélation significative avec F3, dimension pour laquelle elle est négativement corrélée (corrélation moyenne = -0,35).

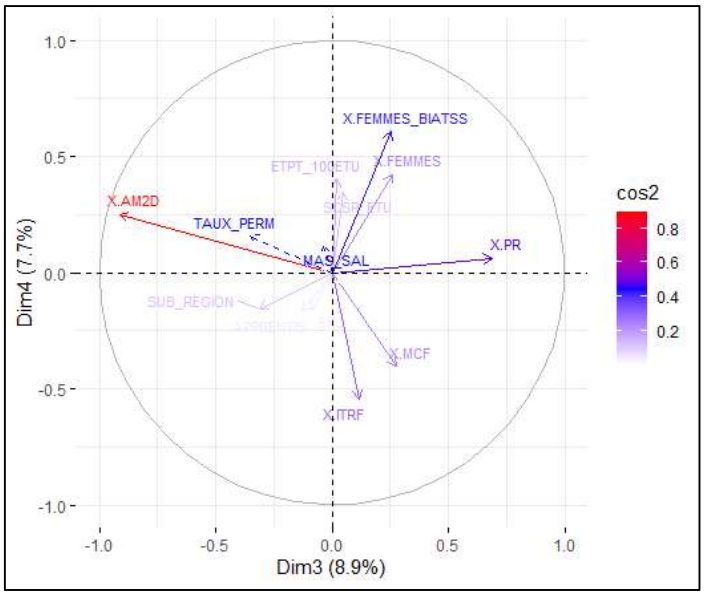



***Figure 8 :*** *Cercle de corrélation des variables pour le second plan factoriel. La couleur des flèches est modulée par l'intensité de la corrélation de la variable avec l'axe concerné (voir légende sur le côté)*

Pour ce qui est des variables qualitatives (illustratives), bien que la variable TYPE_U présente un rapport de corrélation significatif avec la troisième dimension, celui-ci reste faible ($\eta_3$ = 0,17) (Tableau 7b). Au sein de cette variable, ce sont les modalités USH et UPhs qui affichent en moyenne une coordonnée respectivement positive et négative le long de F3.

❖ Analyse des individus

Les tableaux 10 répertorient les valeurs de cos² entre les différents individus et les axes 3 et 4. Par souci de clarté, seuls les premiers individus de cos² plus élevés sont indiqués (l'ensemble de ces coefficients est disponible en Annexe 5b).

| ***Tableau 10a*** | Dim,3 | Dim,4 |
|---|---|---|
| Orleans | 0,50 | 0,00 |
| Bretagne_occ | 0,45 | 0,10 |
| EHESS Paris | 0,45 | 0,07 |
| PSL | 0,40 | 0,09 |
| versailles | 0,40 | 0,00 |
| Bretagnesud | 0,40 | 0,04 |
| INSA Strasbourg | 0,37 | 0,21 |
| ENSAM | 0,36 | 0,19 |
| ToulouseCapitole | 0,34 | 0,12 |
| polytech_HautsFrance | 0,34 | 0,00 |
| ENI Tarbes | 0,31 | 0,02 |

*Tableaux récapitulatif des cos² représentant la projection des individus sur axes 3 et 4. Les individus les mieux représentés, sont ceux correspondant à une valeur élevée de cos² sur l'ensemble du plan.*

***Tableau 10a*** *: Etablissements les mieux représentés sur l'axe F3.*

| ***Tableau 10b*** | Dim,3 | Dim,4 |
|---|---|---|
| ENSIIE | 0,05 | 0,47 |
| EC Nantes | 0,12 | 0,44 |
| EC Lyon | 0,01 | 0,43 |
| INSA Rennes | 0,02 | 0,38 |
| ISAE-ENSMA Poitiers | 0,01 | 0,30 |

***Tableau 10b*** *: Etablissements les mieux représentés sur l'axe F4.*

Sur ce plan (voir Fig.9), les individus les mieux représentés et ayant les plus grandes coordonnées sont : EHESS PARIS, INSA Strasbourg et ENSAM pour F3 et ENSIIE et EC Nantes pour F4. Trois des ENS présentent également de grandes coordonnées sur l'axe F4, mais sont moins bien représentées (cos² compris entre 0,17 et 0,21).

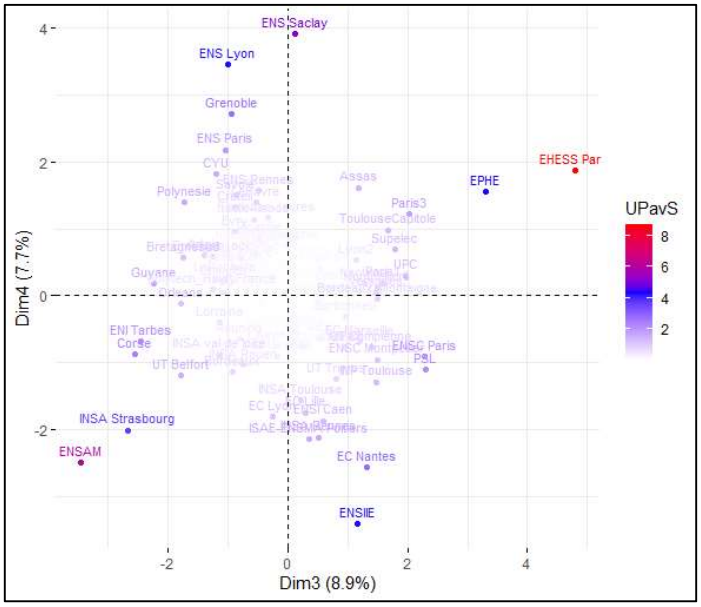



***Figure 9 :*** *Graphique des individus obtenu pour le second plan factoriel. Les individus apparaissent avec une couleur modulée par l'intensité de la contribution à l'axe (voir légende sur le côté).*

❖ Synthèse du second plan

En reprenant l'analyse des variables les plus contributives à ce plan, on peut en conclure que l'EHESS Paris est associée à une grande proportion de professeurs et de femmes et une faible proportion d'AM2D parmi le personnel enseignant et d'ITRF parmi les BIATSS. A l'inverse, l'ENSAM et INSA Strasbourg semblent être caractérisés par de fortes valeurs des proportions d'AM2D, et un plus grand taux de permanents. Il est difficile de se prononcer sur l'EC Nantes et l'ENSIIE qui semblent être associées à de grandes valeurs de proportions de MCF parmi les enseignants et d'ITRF parmi les BIATSS du fait de l'absence de p-valeurs pour la corrélation entre ces variables et F4.

Les écoles normales semblent quant à elles être caractérisées par de grandes valeurs de X.FEMMES sur F4, ce qui peut sembler contradictoirement avec l'observation faite sur le premier plan. Cependant, ici aussi, le test ne nous ayant pas renvoyé de p-valeurs pour la dimension 4, cette variable n'est sans doute pas significativement corrélée à cet axe. De même, le cercle de corrélation des variables pour ce plan (figure 8) semblent indiquer que les variables X.ITRF et X.MCF (X.FEMMES_BIATSS) sont particulièrement faibles (fortes) pour ENS Lyon et ENS Saclay. Après vérification des données initiales, il s'avère que nous n'avions pas de données concernant les variables X.ITRF et X.FEMMES_BIATSS pour l'ENS Saclay (celles-ci ayant été imputées lors de l'ACP), alors que pour l'ENS Lyon, les valeurs initiales correspondantes étaient respectivement de 0,78 et 0,65, valeurs proches de la moyenne.

## Conclusion sur l'ACP

Nous avons mis en évidence quelques caractéristiques spécifiques pour certains établissements.

Si l'on reprend le premier plan de l'ACP, nous avons vu que la variable qualitative TYPE_U était significativement corrélée à ce plan, avec un rapport de corrélation relativement élevé de près de 0,6. La figure 10 reprend le graphique des individus de ce plan, en les colorant selon cette variable.

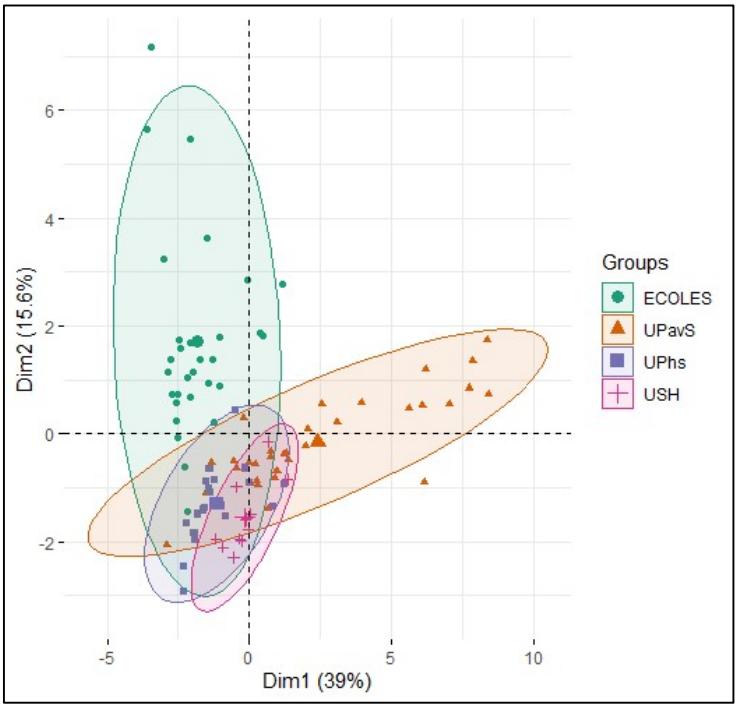



***Figure 10 :*** *Graphique des individus obtenu pour le premier plan factoriel où la couleur des individus est corrélée à la nature de l'établissement défini par la variable TYPE_U (voir légende sur le côté).*

Il semble se dessiner quelques profils particuliers qui pourraient être rassemblés en trois groupes, un premier constitué des individus situés dans le premier quart en haut à gauche), un second pour ceux localisés dans le second quart (en haut à droite) et un troisième pour le reste des individus. Nous allons tenter de mettre en évidence l'existence de groupes à l'aide d'une classification non supervisée.

### CLASSIFICATION NON SUPERVISÉE

La classification a été effectuée à partir de l'application factoshiny[6], sur les composantes de l'analyse factorielle.

## Choix de la méthode : Analyse TANDEM

Nous avons fait le choix de l'utilisation d'une classification ascendante hiérarchique effectuée à partir des axes factoriels issus de l'ACP. Cette méthode permet en effet de pouvoir « nettoyer » les données en excluant les derniers axes correspondant à du bruit. Nous nous sommes fixé une valeur seuil de 90 % de variance cumulée expliquée pour la sélection des axes à retenir, ce qui représente dans le cas présent, les 9 premiers axes de l'ACP.

La méthode choisie combine l'utilisation d'une distance euclidienne pour tenir compte de l'inertie des axes, associée au critère d'agrégation de Ward (plus adapté à l'utilisation de la distance euclidienne) qui permet d'éviter l'effet de chaîne qu'on peut avoir en agrégeant de proche en proche.

La partition obtenue à l'issue de cette étape n'étant pas forcément optimale, une seconde étape de consolidation a été effectuée afin d'homogénéiser les classes.

Le nombre d'observations n'étant pas très important, il n'a pas été nécessaire de réaliser une première classification à l'aide d'un kmeans.

[6] http://factominer.free.fr/graphs/factoshiny-fr.html

La figure 11, représente le dendrogramme obtenu, ainsi que l'éboulis du gain d'inertie, sur lequel on peut voir une grande variation d'inertie quand on passe de 4 à 3 classes, de 3 à 2 et de 2 à 1 classe. Le nombre d'individus n'étant pas important, un grand nombre de classes ne serait pas souhaitable. Par ailleurs, la figure 10 semble plutôt indiquer un regroupement de l'ensemble des individus en 3 groupes. Il nous apparait donc raisonnable de choisir un nombre de classes égal à 3.

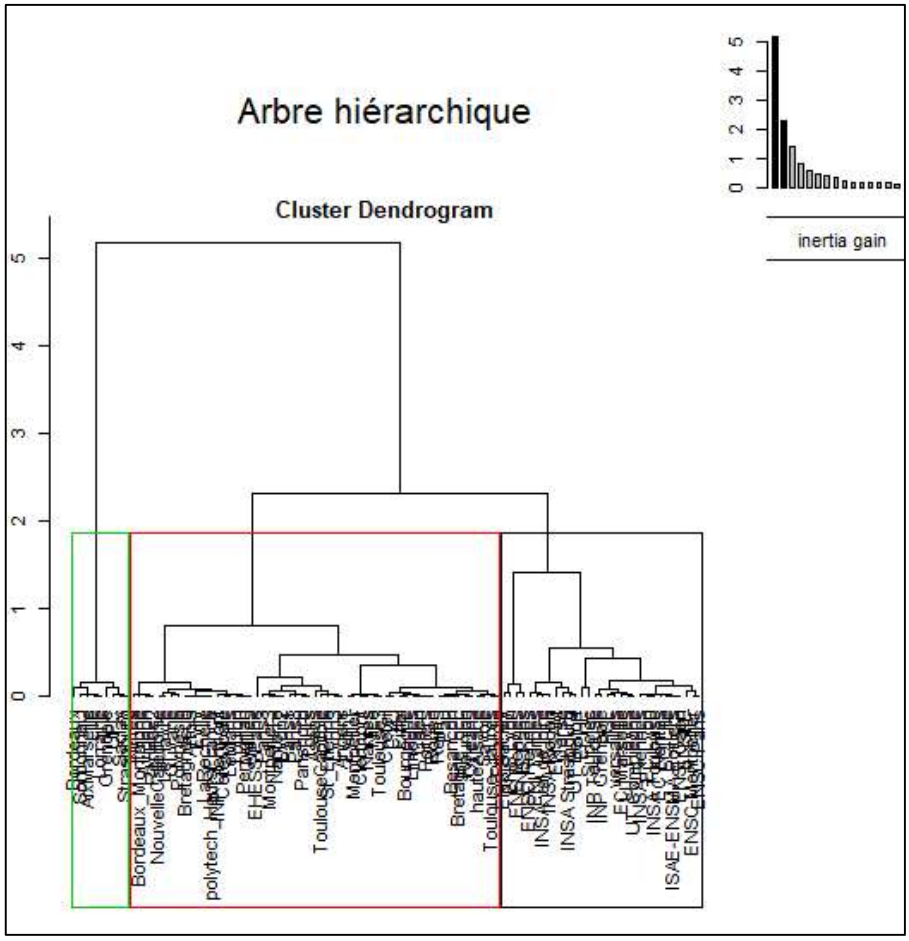



***Figure 11 :*** *Arbre hiérarchique obtenu à la suite d'une CAH sur les 9 premières composantes de l'ACP, suivie d'une étape de consolidation. Diagramme des éboulis de gain d'inertie (en haut à droite).*

## Analyse des classes obtenues

Pour décrire ces classes, nous allons nous intéresser aux résultats obtenus avec l'ACP. La figure 12, montre la répartition des individus au sein des différentes classes sur le premier plan de l'ACP (la ségrégation entre les classes sur le deuxième plan, n'étant pas probante).

Le premier groupe (en bleu sur la figure 12) correspond aux individus associés d'une part à de grandes coordonnées positives sur l'axe F2, soit de fortes valeurs de SCSP_ETU, ETPT_100ETU, X.ENS et X.PR, de faibles valeurs de X.FEMMES ; et d'autre part, à des coordonnées plutôt négatives selon F1, soit de faibles valeurs pour les variables suivantes : MAS_SAL, BIATSS, ETU, SUB_AUTRES, DROITS_INS, RESS_AUTRES, FORM_CONTINUE, SUB_EUR, ANR, et SUB_REGION. Il s'agit donc d'établissements bénéficiant de dotation par étudiant et d'équivalent temps plein importants, ainsi que de nombreux enseignants / enseignants-chercheurs, parmi lesquels la proportion de professeurs et d'hommes sont plus élevés, alors que le nombre d'étudiants et autres sources de revenus sont plus faibles.

Le tableau 11 met en évidence la répartition des individus en fonction de la variable TYPE_U et des classes. Il apparait que la classe 1 regroupe une grande partie des écoles (28 sur 30), notamment les ENS.

La classe 2 est principalement caractérisée par de fortes valeurs pour les variables X.FEMMES, X.MCF et dans une moindre mesure, X.AM2D ; ainsi que de faibles valeurs pour SCSP_ETU, ETPT_100ETU, ENS et X.PR. Il s'agit donc d'établissements avec une faible dotation par étudiant, et un faible nombre d'enseignants parmi lesquels, la proportion de femmes, de MCF et de AM2D est plus importante. Cette classe regroupe notamment l'ensemble des universités en Sciences Humaines et pluridisciplinaires hors secteur santé à une exception près (tableau 11), comme l'université de Nîmes par exemple.

Enfin, le groupe 3 est quant à lui, caractérisé par de fortes coordonnées sur l'axe F1, soit de fortes valeurs des variables suivantes : MAS_SAL, BIATSS, ETU, SUB_AUTRES, DROITS_INS, RESS_AUTRES, FORM_CONTINUE, SUB_EUR, ANR, et SUB_REGION. Il n'est pas affecté par les variables associées à F2. Il s'agit de établissements avec beaucoup d'étudiants, de personnel BIATSS et ayant plus de revenus autres que la dotation de l'État. Le tableau 11 indique qu'il s'agit d'établissements pluridisciplinaires avec secteur santé, tels que Saclay, UPC et Sorbonne U.

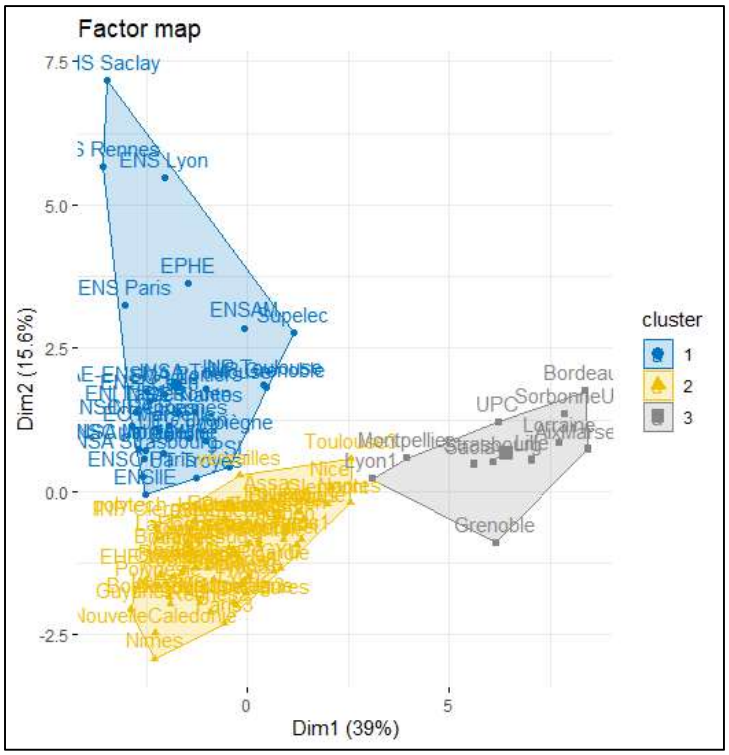



***Figure 12 :*** Répartition des individus au sein des différentes classes *sur le premier plan de l'ACP.*

| Répartition | classe 1 | classe 2 | classe 3 |
|---|---|---|---|
| ECOLES | 28 | 2 | 0 |
| UPavS | 0 | 23 | 11 |
| UPhs | 1 | 20 | 0 |
| USH | 0 | 12 | 0 |

***Tableau 11 :*** *Répartition des individus en fonction de la variable TYPE_U et des classes.*

### Conclusion sur la classification

La méthode de partitionnement utilisée nous a permis de mettre en évidence 3 groupes au sein de l'ensemble des établissements. Deux de ces groupes ne sont constitués que d'un seul type d'établissements comme défini par la variable TYPE_U. Le groupe 1 ne rassemblant que des écoles et le groupe 3 ne rassemblant que des universités pluridisciplinaires avec secteur santé.

Nous avions vu au cours de l'analyse de l'ACP, que cette même variable était significativement corrélée au premier plan obtenu. Les résultats obtenus avec la CAH viennent conforter cette observation.

## Conclusion

L'ensemble de ces méthodes a permis de faire ressortir le fait que les établissements supérieurs français ne sont pas similaires en terme de revenus ou d'effectifs étudiants et personnel. La classification a mis en évidence l'existence de trois classes :

- La première est composée uniquement d'écoles qui semblent mieux dotées par rapport au nombre d'étudiants inscrits, des revenus par ailleurs moins important, avec une proportion de femmes chez les enseignants / enseignants chercheurs plus faible, pour une proportion de professeurs et assimilés plus élevée.
- La seconde comprend plusieurs types d'établissements, dont l'ensemble des universités Sciences Humaines, presque l'ensemble des universités pluridisciplinaires hors secteur santé et une grande partie des universités pluridisciplinaires avec secteur santé mais de taille relativement petite et plutôt hors région parisienne.
- La troisième correspond à de grands établissements en terme d'effectifs étudiants comme attendu dans le cas d'universités pluridisciplinaires avec secteur santé qui composent ce groupe, dont la dotation par étudiant est faible, mais avec de plus importants revenus par ailleurs.

Ce projet nous a permis de rassembler et d'appliquer les méthodes d'analyse uni-, bi- et multivariées vues en cours sur un jeu de données réel que nous avons élaboré à partir de plusieurs sources.

# ANNEXES

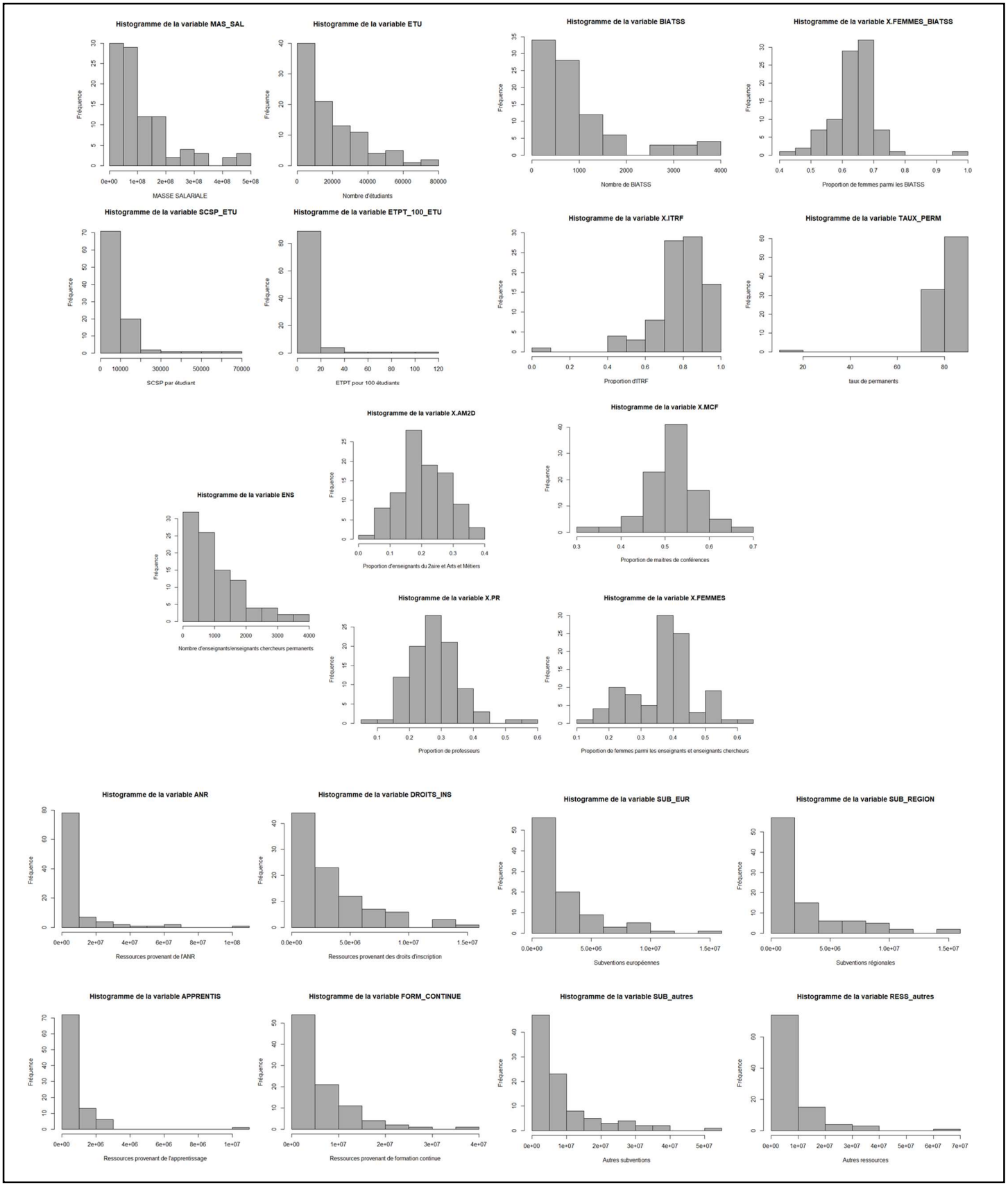
L’ensemble du projet est le fruit de séances de travail en binôme et chacun des deux auditeurs a contribué de façon égale à l’ensemble du projet : de la récupération des données au rapport final, en passant par la mise au point des méthodes et l’analyse des sorties.

***Annexe 1 :*** *Nom complet des universités et écoles étudiés.*

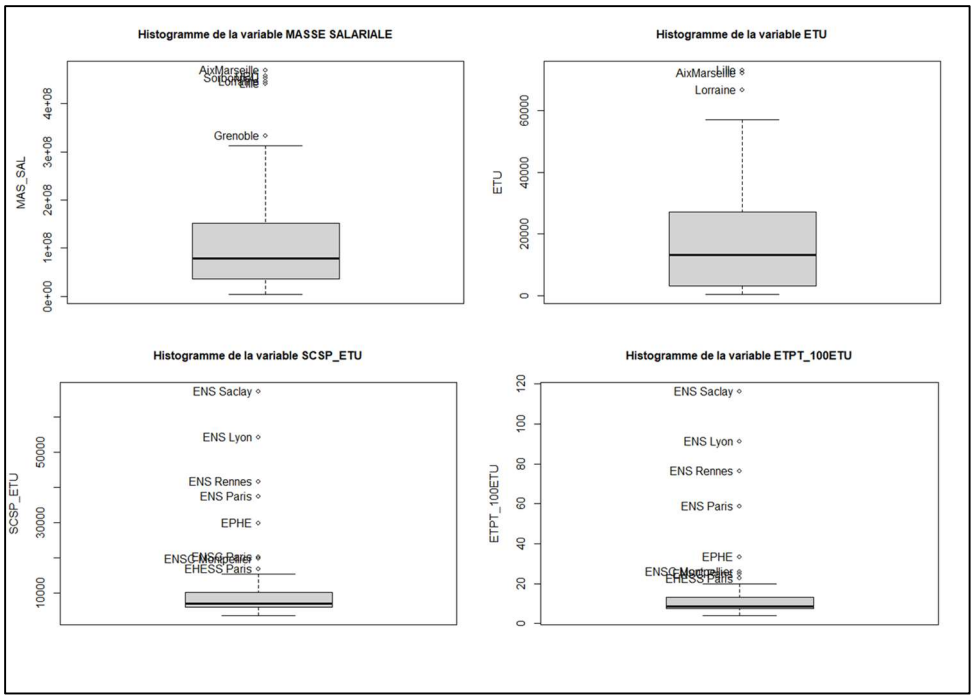
| Université | Nom complet | TYPE_U | Université | Nom complet | TYPE_U |
|---|---|---|---|---|---|
| Angers | U. d'Angers | UPavS | PSL | U. Paris Sciences et Lettres | UPhs |
| Picardie | U. Picardie Jules- Verne | UPavS | Artois | U. d'Artois | UPhs |
| Rouen | U. Rouen Normandie | UPavS | CoteOpale | U. du littoral Côte d'Opale | UPhs |
| Creteil | U. Paris-Est Créteil | UPavS | Savoie | U. Savoie Mont Blanc | UPhs |
| Reunion | U. de la Réunion | UPavS | LeMans | Le Mans U. | UPhs |
| Saclay | U. Paris-Saclay | UPavS | Toulon | U. de Toulon | UPhs |
| St_Etienne | U. Jean Monnet | UPavS | Pau | U. de Pau et des pays de l'Adour | UPhs |
| Guyane | U. Guyane | UPavS | Orleans | U. d'Orléans | UPhs |
| Nice | U. Côte d'azur | UPavS | Perpignan | U. de Perpignan via Domitia | UPhs |
| Caen | U. de Caen Normandie | UPavS | Avignon | Avignon U. | UPhs |
| Nantes | U. de Nantes | UPavS | hauteAlsace | U. de haute Alsace | UPhs |
| Bretagne_occ | U. de Bretagne occidentale | UPavS | LaRochelle | La Rochelle U. | UPhs |
| Bourgogne | U. de Bourgogne | UPavS | NouvelleCaledonie | U. de la nouvelle Calédonie | UPhs |
| Montpellier | U. de Montpellier | UPavS | LeHavre | U. Le Havre Normandie | UPhs |
| Parisnord | U. Sorbonne Paris nord | UPavS | Corse | U. de Corse Pasquale Paoli | UPhs |
| Tours | U. de Tours | UPavS | Polynesie | U. de la Polynésie française | UPhs |
| Strasbourg | U. de Strasbourg | UPavS | Evry | U. d'Evry-val-d'Essone | UPhs |
| Reims | U. de Reims Champagnne-Ardenne | UPavS | Bretagnesud | U. de Bretagne sud | UPhs |
| Grenoble | U. Grenoble Alpes | UPavS | Lyon1 | U. Claude Bernard - Lyon 1 | UPavS |
| polytech_HautsFrance | U. polytechnique Hauts-de-France | UPavS | UPC | UPC | UPavS |
| Limoges | U. de Limoges | UPavS | Toulouse3 | U. Toulouse III - Paul Sabatier | UPavS |
| Lille | U. de Lille | UPavS | Montpellier3 | U. Paul Valéry - Montpellier 3 | UTALLSHS |
| Bordeaux | U. de Bordeaux | UPavS | Lyon2 | U. Lumière - Lyon 2 | UTALLSHS |
| Clermont | U. Clermont-Auvergne | UPavS | Rennes2 | U. de Rennes 2 | UTALLSHS |
| Poitiers | U. de Poitiers | UPavS | Toulouse-Jaures | U. Toulouse - Jean Jaurès | UTALLSHS |
| Antilles | U. des Antilles | UPavS | Bordeaux_Montaigne | U. Bordeaux Montaigne | UTALLSHS |
| AixMarseille | Aix-Marseille U. | UPavS | Paris8 | Université Paris 8 - Vincennes - Saint Denis | UTALLSHS |
| Besancon | U. de Franche-Comté | UPavS | Nanterre | U. Paris Nanterre | UTALLSHS |
| Lorraine | U. de Lorraine | UPavS | Paris3 | U. Sorbonne Nouvelle - Paris 3 | UTALLSHS |
| versailles | U. de versailles Saint-Quentin en Yvelines | UPavS | Lyon3 | U. Jean Moulin - Lyon 3 | UTDEG |
| SorbonneU | Sorbonne U. | UPavS | Assas | U. Panthéon-Assas | UTDEG |
| Nimes | U. de Nîmes | UPhs | ToulouseCapitole | U. Toulouse Capitole | UTDEG |
| Eiffel | U. Gustave Eiffel | UPhs | Paris1 | U. Paris 1 - Panthéon Sorbonne | UTDEG |
| CYU | CergY U. | UPhs | | | |

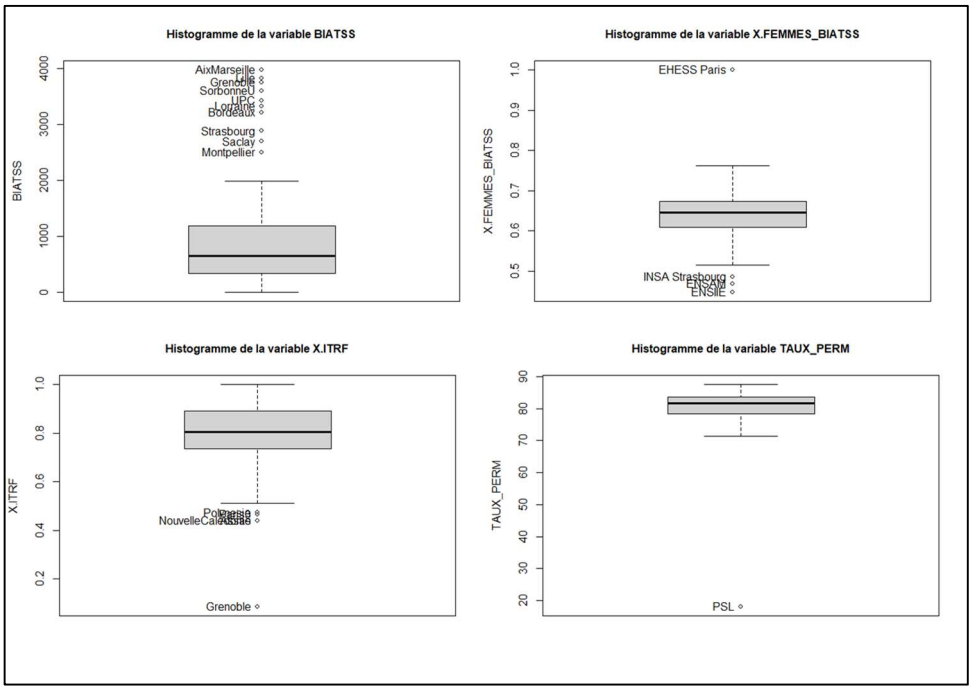
| Ecoles | Nom complet | TYPE_U | Ecoles | Nom complet | TYPE_U |
|---|---|---|---|---|---|
| Supelec | Centrale-Supélec | GE | INSA Rennes | Institut National des Sciences Appliquées de Rennes | EI |
| EC Lille | Ecole de Commerce de Lille | EI | INSA Rouen | Institut National des Sciences Appliquées de Rouen | EI |
| EC Lyon | Ecole de Commerce de Lyon | EI | INSA Strasbourg | Institut National des Sciences Appliquées de Strasbourg | EI |
| EC Marseille | Ecole de Commerce de Marseille | EI | INSA Toulouse | Institut National des Sciences Appliquées de Toulouse | EI |
| EC Nantes | Ecole de Commerce de Nantes | EI | ISAE-ENSMA Poitiers | Ecole Nationale Supérieure de Mécanique et d'Aérotechnique | EI |
| ENI Tarbes | Ecole Nationale d'Ingénieurs de Tarbes | EI | INP Clermont | Institut National Polytechnique de Grenoble | EI |
| ENSAM | Ecole Nationale Supérieure d'Arts et Métiers | GE | UT Belfort | Université Technologique de Belfort | EI |
| ENSC Montpellier | Ecole Normale Supérieure de Chimie de Montpellier | EI | UT Compiègne | Université Technologique de Compiègne | EI |
| ENSC Paris | Ecole Normale Supérieure de Chimie de Paris | EI | UT Troyes | Université Technologique de Troyes | EI |
| ENSC Rennes | Ecole Normale Supérieure de Chimie de Rennes | EI | EHESS Paris | Ecole des Hautes Etudes en Sciences Sociales | GE |
| ENSI Caen | Ecole Nationale Supérieure d'Ingénieurs de Caen | EI | ENS Paris | Ecole Normale Supérieure de Paris | AUTRES |
| ENSIIE | Ecole Nationale Supérieure d'Informatique pour l'Industrie et l'Entreprise | EI | ENS Lyon | Ecole Normale Supérieure de Lyon | AUTRES |
| INP Grenoble | Institut National Polytechnique de Grenoble | GE | ENS Saclay | Ecole Normale Supérieure Paris-Saclay | AUTRES |
| INP Toulouse | Institut National Polytechnique de Toulouse | EI | ENS Rennes | Ecole Normale Supérieure de Rennes | AUTRES |
| INSA val de loire | Institut National des Sciences Appliquées Val de Loire | EI | EPHE | | GE |

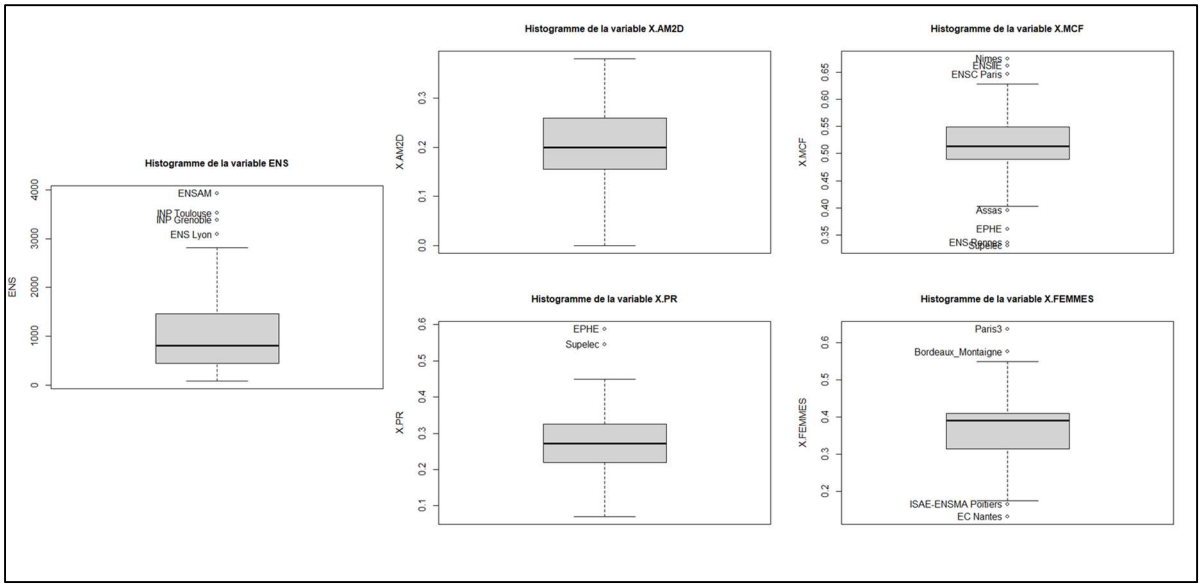
***Annexe 2 :*** *Histogrammes de l'ensemble des variables quantitatives de l'étude.*

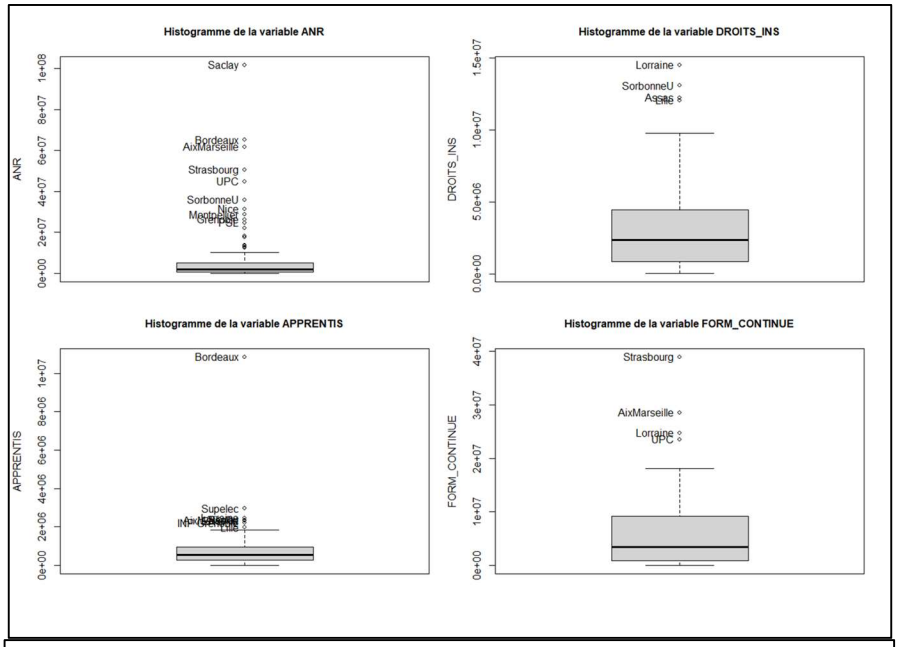


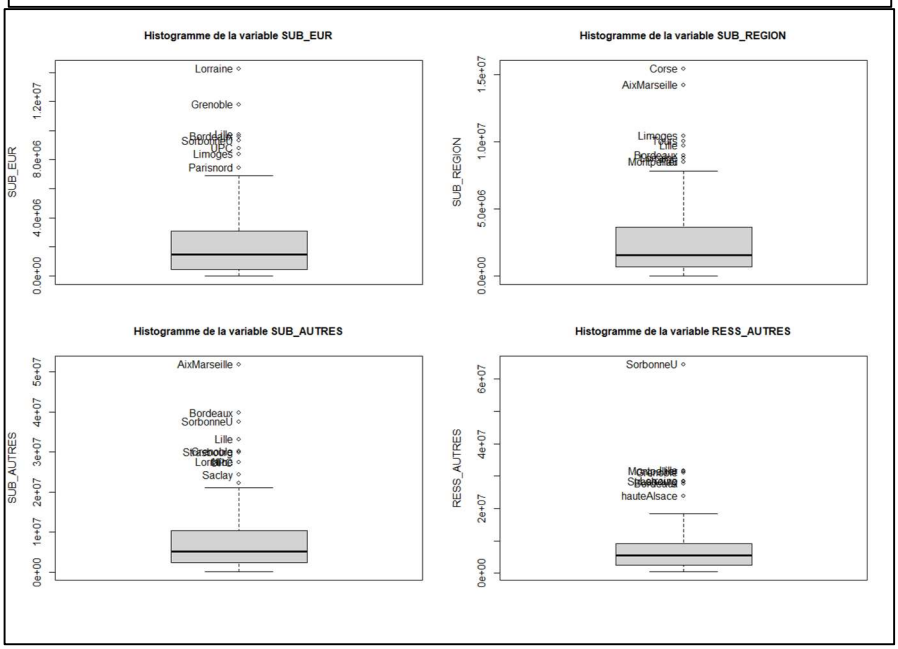
***Annexe 3 :*** *Représentation sous forme de boîtes à moustaches des variables quantitatives étudiées*.



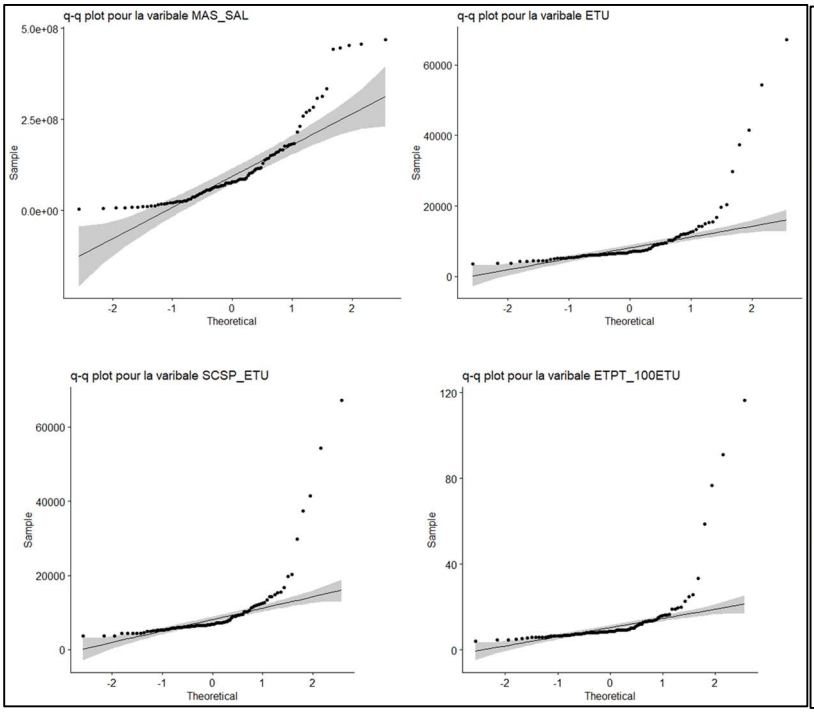



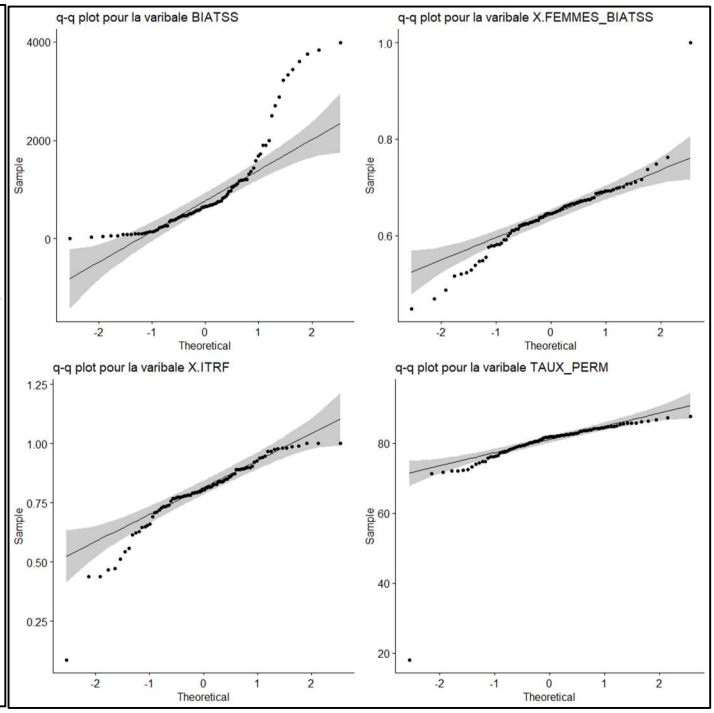

Histogramme de la variable ENS
ENS
ENSAM
INP Toulouse
INP Grenoble
ENS Lyon
Histogramme de la variable X.AM2D
X.AM2D
Histogramme de la variable X.MCF
X.MCF
Nimes
ENSIIE
ENSC Paris
Assas
EPHE
Histogramme de la variable X.PR
X.PR
EPHE
Supelec
Histogramme de la variable X.FEMMES
X.FEMMES
Paris3
Bordeaux_Montaigne
ISAE-ENSMA Poitiers
EC Nantes

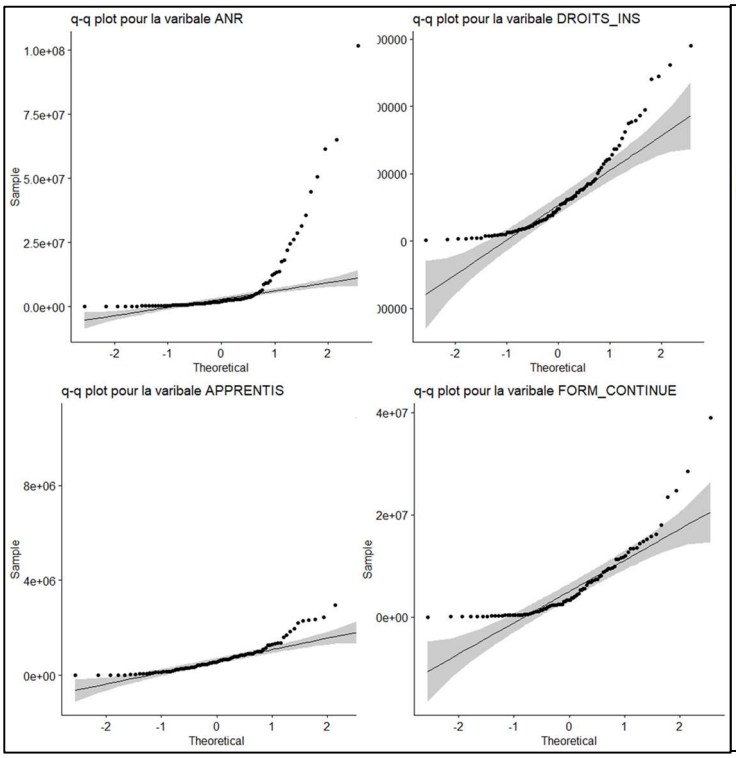

Histogramme de la variable ANR
ANR
Saclay
Bordeaux
AixMarseille
Strasbourg
UPC
SorbonneU
Nice
Histogramme de la variable DROITS_INS
DROITS_INS
Lorraine
SorbonneU
Assas
Lille
Histogramme de la variable APPRENTIS
APPRENTIS
Bordeaux
Supelec
Histogramme de la variable FORM_CONTINUE
FORM_CONTINUE
Strasbourg
AixMarseille
Lorraine
UPC

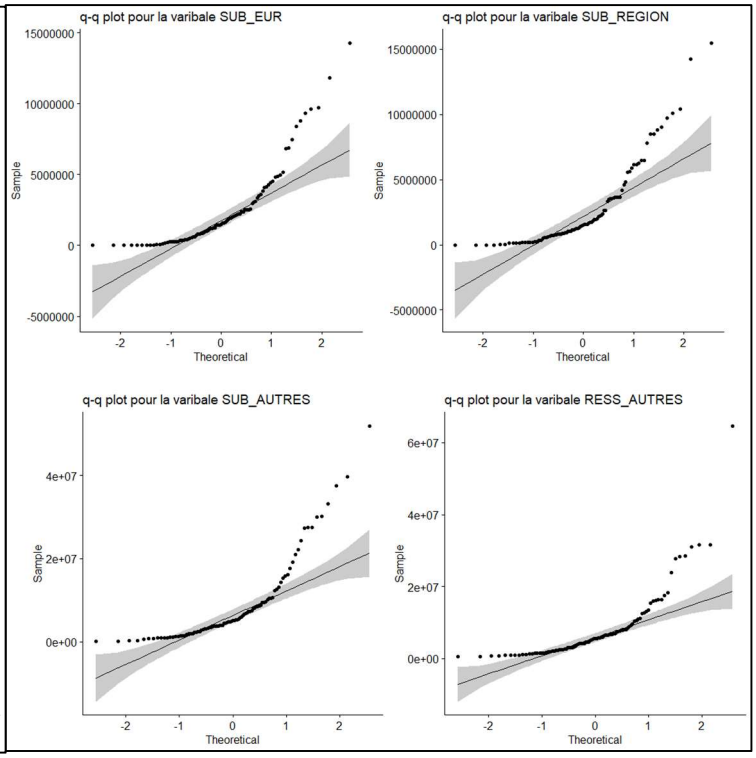

Histogramme de la variable SUB_EUR
SUB_EUR
Lorraine
Grenoble
Limoges
Parisnord
Histogramme de la variable SUB_REGION
SUB_REGION
Corse
AixMarseille
Limoges
Histogramme de la variable SUB_AUTRES
SUB_AUTRES
AixMarseille
Bordeaux
SorbonneU
Lille
Saclay
Histogramme de la variable RESS_AUTRES
RESS_AUTRES
SorbonneU
hauteAlsace

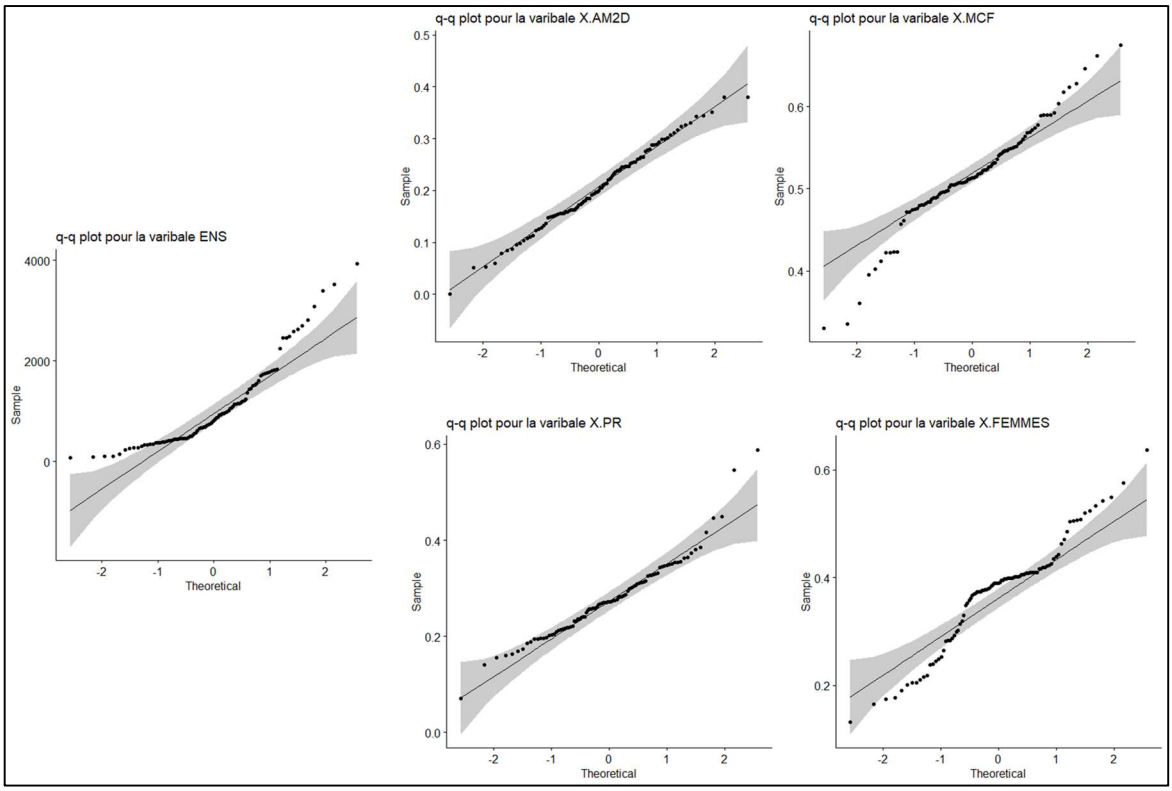
***Annexe 4 :*** *Q-Q plot de l'ensemble des variables étudiées (les variables sont indiquées dans le titre du graphique).*




q-q plot pour la varibale ENS
Sample
Theoretical
4000
2000
0
-2
-1
0
1
2
q-q plot pour la varibale X.AM2D
Sample
Theoretical
0.5
0.4
0.3
0.2
0.1
0.0
q-q plot pour la varibale X.MCF
Sample
Theoretical
0.6
0.5
0.4
q-q plot pour la varibale X.PR
Sample
Theoretical
0.6
0.4
0.2
0.0
q-q plot pour la varibale X.FEMMES
Sample
Theoretical
0.6
0.4
0.2

***Annexe 5a :*** *Tableau recensant les valeurs de cos² entre les individus et les deux premiers axes factoriels.*

| établissement | dim1 | dim2 | établissement | dim1 | dim2 |
|---|---|---|---|---|---|
| Lille | 0,86 | 0,01 | Bourgogne | 0,26 | 0,02 |
| Lorraine | 0,84 | 0,01 | ENSIIE | 0,26 | 0,00 |
| AixMarseille | 0,81 | 0,01 | UT Troyes | 0,25 | 0,01 |
| UPC | 0,79 | 0,03 | Bretagnesud | 0,25 | 0,15 |
| Lyon1 | 0,79 | 0,00 | ENS Rennes | 0,24 | 0,61 |
| Nantes | 0,74 | 0,00 | Artois | 0,24 | 0,26 |
| SorbonneU | 0,74 | 0,02 | Paris1 | 0,21 | 0,08 |
| Montpellier | 0,73 | 0,02 | Eiffel | 0,16 | 0,09 |
| Toulouse3 | 0,73 | 0,03 | Limoges | 0,14 | 0,01 |
| INP Clermont | 0,72 | 0,05 | ENS Saclay | 0,14 | 0,58 |
| Clermont | 0,71 | 0,01 | Bordeaux_Montaigne | 0,13 | 0,36 |
| Strasbourg | 0,70 | 0,00 | Rennes2 | 0,13 | 0,65 |
| Saclay | 0,57 | 0,00 | UT Compiègne | 0,12 | 0,09 |
| Bordeaux | 0,57 | 0,02 | INSA Toulouse | 0,12 | 0,34 |
| Grenoble | 0,56 | 0,01 | EC Nantes | 0,12 | 0,13 |
| ENSC Montpellier | 0,56 | 0,04 | Tours | 0,11 | 0,09 |
| INSA val de loire | 0,53 | 0,04 | hauteAlsace | 0,09 | 0,18 |
| Antilles | 0,51 | 0,27 | Picardie | 0,09 | 0,38 |
| Toulon | 0,48 | 0,33 | EHESS Paris | 0,09 | 0,04 |
| Perpignan | 0,47 | 0,35 | Savoie | 0,08 | 0,29 |
| LeMans | 0,46 | 0,24 | ENS Lyon | 0,07 | 0,51 |
| EC Marseille | 0,46 | 0,10 | Creteil | 0,06 | 0,02 |
| ENSI Caen | 0,45 | 0,19 | EPHE | 0,06 | 0,35 |
| ENSC Rennes | 0,43 | 0,07 | Corse | 0,05 | 0,03 |
| CoteOpale | 0,43 | 0,33 | St_Etienne | 0,05 | 0,11 |
| Guyane | 0,42 | 0,21 | CYU | 0,05 | 0,13 |
| Nice | 0,41 | 0,00 | Reunion | 0,05 | 0,04 |
| ISAE-ENSMA Poitiers | 0,40 | 0,19 | Reims | 0,04 | 0,29 |
| ENI Tarbes | 0,39 | 0,10 | Supelec | 0,04 | 0,24 |
| LaRochelle | 0,38 | 0,13 | ToulouseCapitole | 0,03 | 0,12 |
| INSA Rouen | 0,38 | 0,24 | Angers | 0,02 | 0,23 |
| ENSC Paris | 0,38 | 0,00 | Bretagne_occ | 0,02 | 0,21 |
| polytech_HautsFrance | 0,38 | 0,06 | INP Grenoble | 0,02 | 0,26 |
| Avignon | 0,37 | 0,40 | Paris3 | 0,02 | 0,28 |
| Caen | 0,37 | 0,05 | PSL | 0,02 | 0,01 |
| EC Lille | 0,36 | 0,24 | Assas | 0,02 | 0,00 |
| ENS Paris | 0,34 | 0,39 | Montpellier3 | 0,01 | 0,50 |
| INSA Strasbourg | 0,34 | 0,02 | INP Toulouse | 0,01 | 0,22 |
| LeHavre | 0,34 | 0,31 | Toulouse-Jaures | 0,01 | 0,52 |
| INSA Rennes | 0,33 | 0,11 | versailles | 0,01 | 0,01 |
| UT Belfort | 0,32 | 0,03 | Orleans | 0,00 | 0,06 |
| Rouen | 0,32 | 0,16 | Lyon3 | 0,00 | 0,36 |
| Polynesie | 0,31 | 0,18 | Nanterre | 0,00 | 0,30 |
| Poitiers | 0,29 | 0,06 | Parisnord | 0,00 | 0,14 |
| NouvelleCaledonie | 0,29 | 0,34 | Lyon2 | 0,00 | 0,55 |
| Evry | 0,27 | 0,06 | ENSAM | 0,00 | 0,25 |
| Nimes | 0,27 | 0,44 | Paris8 | 0,00 | 0,29 |
| EC Lyon | 0,27 | 0,12 | Pau | 0,00 | 0,12 |
| | | | Besancon | 0,00 | 0,11 |

***Annexe 5b :*** *Tableau recensant les valeurs de cos² entre les individus et les axes factoriels 3 et 4.*

| établissement | Dim,3 | Dim,4 | établissement | Dim,3 | Dim,4 |
|---|---|---|---|---|---|
| Orleans | 0,50 | 0,00 | Reims | 0,06 | 0,02 |
| Bretagne_occ | 0,45 | 0,10 | Assas | 0,05 | 0,10 |
| EHESS Paris | 0,45 | 0,07 | Parisnord | 0,05 | 0,00 |
| PSL | 0,40 | 0,09 | ENSIIE | 0,05 | 0,47 |
| versailles | 0,40 | 0,00 | Nimes | 0,05 | 0,03 |
| Bretagnesud | 0,40 | 0,04 | Picardie | 0,04 | 0,15 |
| INSA Strasbourg | 0,37 | 0,21 | Perpignan | 0,04 | 0,00 |
| ENSAM | 0,36 | 0,19 | ENS Paris | 0,04 | 0,18 |
| ToulouseCapitole | 0,34 | 0,12 | Caen | 0,04 | 0,00 |
| polytech_HautsFrance | 0,34 | 0,00 | Eiffel | 0,03 | 0,04 |
| ENI Tarbes | 0,31 | 0,02 | hauteAlsace | 0,03 | 0,13 |
| ENSC Paris | 0,29 | 0,05 | Avignon | 0,03 | 0,00 |
| Paris1 | 0,29 | 0,00 | Nice | 0,03 | 0,04 |
| EPHE | 0,29 | 0,06 | ENSI Caen | 0,03 | 0,27 |
| Artois | 0,29 | 0,05 | Nantes | 0,03 | 0,00 |
| Montpellier3 | 0,29 | 0,00 | LeHavre | 0,03 | 0,17 |
| Besancon | 0,28 | 0,06 | INSA Rennes | 0,02 | 0,38 |
| Guyane | 0,25 | 0,00 | Rouen | 0,02 | 0,03 |
| Corse | 0,24 | 0,03 | St_Etienne | 0,02 | 0,14 |
| Lyon2 | 0,23 | 0,05 | Lorraine | 0,02 | 0,00 |
| UT Belfort | 0,23 | 0,11 | ENS Lyon | 0,02 | 0,21 |
| UT Compiègne | 0,22 | 0,07 | CoteOpale | 0,02 | 0,06 |
| Paris3 | 0,22 | 0,08 | Rennes2 | 0,02 | 0,00 |
| Bordeaux_Montaigne | 0,21 | 0,00 | LeMans | 0,02 | 0,02 |
| Nanterre | 0,20 | 0,00 | Toulouse-Jaures | 0,01 | 0,18 |
| Polynesie | 0,19 | 0,13 | Tours | 0,01 | 0,02 |
| LaRochelle | 0,19 | 0,01 | Grenoble | 0,01 | 0,11 |
| Poitiers | 0,18 | 0,00 | SorbonneU | 0,01 | 0,00 |
| ENSC Montpellier | 0,17 | 0,07 | Toulon | 0,01 | 0,00 |
| Pau | 0,17 | 0,00 | EC Lyon | 0,01 | 0,43 |
| INP Toulouse | 0,14 | 0,10 | Clermont | 0,01 | 0,00 |
| EC Marseille | 0,14 | 0,04 | ISAE-ENSMA Poitiers | 0,01 | 0,30 |
| Paris8 | 0,12 | 0,08 | Montpellier | 0,01 | 0,00 |
| EC Nantes | 0,12 | 0,44 | EC Lille | 0,01 | 0,26 |
| Evry | 0,11 | 0,13 | Bordeaux | 0,01 | 0,01 |
| INSA val de loire | 0,11 | 0,07 | NouvelleCaledonie | 0,01 | 0,03 |
| Reunion | 0,11 | 0,07 | Antilles | 0,01 | 0,06 |
| Bourgogne | 0,10 | 0,05 | Strasbourg | 0,01 | 0,00 |
| UT Troyes | 0,10 | 0,25 | INSA Toulouse | 0,01 | 0,26 |
| CYU | 0,10 | 0,24 | ENS Rennes | 0,00 | 0,05 |
| Supelec | 0,10 | 0,01 | INP Grenoble | 0,00 | 0,05 |
| Limoges | 0,10 | 0,00 | Lille | 0,00 | 0,01 |
| Savoie | 0,09 | 0,27 | INP Clermont | 0,00 | 0,05 |
| Creteil | 0,08 | 0,19 | Toulouse3 | 0,00 | 0,03 |
| Lyon3 | 0,08 | 0,00 | AixMarseille | 0,00 | 0,01 |
| UPC | 0,08 | 0,00 | Lyon1 | 0,00 | 0,01 |
| Angers | 0,07 | 0,05 | Saclay | 0,00 | 0,00 |
| INSA Rouen | 0,07 | 0,14 | ENS Saclay | 0,00 | 0,17 |
| | | | ENSC Rennes | 0,00 | 0,01 |