



UNIVERSITÉ
CAEN
NORMANDIE

Réputation du programme Erasmus et citoyenneté européenne chez les élèves de lycées technologique et professionnel

Master 2 Statistiques Appliquées et Analyse Décisionnelle
Licence 3 Mathématiques Informatiques Appliquées aux Sciences
Humaines et Sociales

Octobre 2018 – Mars 2019

Rapport réalisé par : CAPLAN Mateo, CHALONY Pauline, CRIGNON Valentin, FADIL
Issam, FLAUX Pauline, LAHUTTE Maud

Programme Erasmus +



Erasmus + est un programme visant à soutenir des actions sur divers domaines comme, l'enseignement scolaire, l'enseignement supérieur, la formation professionnelle, l'éducation des adultes, la jeunesse (éducation non formelle) et le sport, tout en sachant que le programme Erasmus + soutient financièrement cette large gamme d'activités.

Ce programme vise à donner aux lycéens, étudiants, stagiaires, d'une manière générale aux jeunes de moins de 30 ans diplômés ou non ainsi qu'au personnel, la possibilité de séjourner à l'étranger pour renforcer leurs compétences et par la même occasion d'accroître leur employabilité.

Il existe différents types de projets concernant Erasmus + qui se distinguent en trois actions clés :

- La mobilité à des fins d'apprentissage, clé action 1.
- La coopération pour l'innovation et l'échange de bonnes pratiques, clé action 2.
- Le soutien à la réforme politique, clé action 3.

Présentation du lycée des métiers Institut Lemonnier :

L'Institut Lemonnier est un lycée technologique professionnel et agricole privé se situant à Caen, en Normandie. Créé en 1962 après avoir été un orphelinat et un institut professionnel, il a aujourd'hui 1300 élèves de 14 à 20 ans.

Concernant l'enseignement, il regroupe trois lycées :

Lycée général et technologique :

- Seconde générale et technologique.

Première S Sciences de l'ingénieur.

Première Sciences et technologies de l'industrie et du développement durable (STI 2D) Énergie et environnement.

Première Sciences et technologies de l'industrie et du développement durable (STI 2D) Systèmes d'information et numérique.

Terminale S Sciences de l'ingénieur.

Terminale Sciences et technologies de l'industrie et du développement durable (STI 2D) Énergie et environnement.

Terminale Sciences et technologies de l'industrie et du développement durable (STI 2D) Systèmes d'information et numérique.

BTS :

Conception et réalisation de systèmes automatiques.

Électrotechnique.

Fluides, énergies, environnements, option Génie frigorifique.

Informatique et réseaux pour l'industrie et les services techniques.

Maintenance industrielle.

Licence professionnelle :

Systèmes informatiques embarqués.

Conception et supervision des systèmes automatisés.

Le lycée professionnel :

- Troisième Prépa professionnelle

CAP

Ébénisterie

Ébénisterie - parcours adapter

Maintenance des matériels de parcs et jardins

Menuisier fabricant de menuiserie, mobilier et agencement

Peinture en carrosserie

Réparation des carrosseries

Bac Professionnel

Électrotechnique, énergie, équipements communicants

Maintenance des équipements Industriels

Maintenance des véhicules automobiles, option Voitures particulières

Technicien du froid et du conditionnement de l'air

Technicien menuisier agencier

Brevet des métiers d'arts option ébéniste

Le lycée agricole :

- Quatrième d'orientation professionnelle

Troisième d'orientation professionnelle

CAP Travaux paysagers

Seconde professionnelle

Technicien conseil-vente

Aménagements paysagers

Première professionnelle

Technicien conseil-vente

Aménagements paysagers

B.T.S. Technico-Commercial, option Produits alimentaires et boissons

1. Contexte et objectifs :

Le projet porte sur la citoyenneté européenne au lycée Lemonnier, notamment avec le projet Erasmus.

Autrement dit, cela correspond à la valeur qu'à l'Union européenne à nos yeux, ainsi qu'à mesurer et répertorier les relations et les actions menées par les élèves de différents pays, notamment pour rester en contact les uns avec les autres.

Pour cela, l'objectif est d'aller interroger les élèves et de récolter des données. Une enquête ayant déjà eu lieu, cela va permettre de faire des comparaisons ainsi que d'affiner les questions afin d'avoir des résultats plus précis. Avoir des résultats précis peut permettre au projet d'avoir plus de poids et plus d'importance, permettant à plus d'élèves de participer ainsi que des subventions plus élevées pour les différents projets.

Le projet Erasmus actuel, EEVEE (Engineering Energy Vehicule for European Environnement), au lycée Lemonnier, les élèves en Bac pro carrosserie ont pour objectif de faire des mathématiques et des sciences pour créer par eux-mêmes une voiture. Ce projet est en collaboration avec un lycée en Suède et un lycée en Espagne. Des lycéens des trois lycées participent à ce projet, ils doivent donc interagir les uns avec les autres, permettant la création du sentiment d'appartenance à l'union européenne.

Cette étude sert à différentes échelles. En premier lieu, elle va servir au personnel du lycée qui va pouvoir mesurer l'importance et l'engagement des élèves vis à vis du projet Erasmus, ainsi que soutenir les demandes de subventions à l'avenir afin de réaliser plus de projet selon le nombre d'élèves intéressés. Ensuite, elle sert à l'échelle régionale, aidant la réputation de la région et les subventions versées. Puis cela va servir aux différents organismes organisant les projets Erasmus qui vont pouvoir savoir l'importance et la valeur du projet pour les élèves.

2. Elaboration du questionnaire :

Au début de ce projet, Mme Jenvrin, la coordinatrice du projet pour le lycée, nous a contactés afin de fixer un premier rendez-vous. Lors de ce rendez-vous, elle nous a transmis toutes les informations à savoir concernant le lycée et le travail attendu. Elle nous a fait part d'une enquête réalisée sur le même sujet l'année précédente au lycée Lemonnier également.

Ces informations nous ont permis de mettre au point le nouveau questionnaire, dans lequel nous avons toutefois effectué des changements importants sur le fond et sur la forme.

L'objectif cette année est notamment d'interroger plus de 200 élèves. Alors que l'ancien questionnaire était entièrement en format papier, nous avons réalisé un questionnaire informatisé sur la plate-forme Google Form. Le format informatique permet notamment d'être sûr que les élèves répondront à toutes les questions obligatoires, ainsi que de surveiller plus facilement à ce que les élèves répondent individuellement au questionnaire.

En première attention, nous avons ajouté une définition brève du projet Erasmus afin que même les élèves qui ne connaissent pas le projet puissent répondre aux questions. Afin que certains élèves ne soient pas directement réticents à donner leurs noms et prénoms, nous avons décidé de demander ces informations à la fin, en précisant à l'oral que cela est juste à titre indicatif et ce quel que soit leurs réponses, cela n'aura aucune conséquence future pour eux.

Par la suite, nous avons également choisi de fragmenter nos réponses, afin que les réponses possibles soient plus nuancées et ne se résument pas à oui et non. Certaines questions ont été reformulées afin qu'elles soient plus compréhensibles et plus précises, notamment celles à propos des moyens de communication sur le projet ainsi que la question à propos du développement durable.

Nous avons également ajouté certaines questions, notamment sur les carburants alternatifs à l'essence. Cette question est intéressante car certains élèves participent à un projet Erasmus et doivent réaliser une voiture respectueuse de l'environnement. Il est donc intéressant de savoir le nombre d'élèves au courant de ce qui se passe dans le lycée ainsi que les progrès récents en matière de carburants.

Nous avons également ajouté une question par rapport aux financements du projet Erasmus, pour faire savoir aux élèves qu'il existe des aides financières pour participer au projet Erasmus, ainsi que les organismes qui versent les subventions.

Notre collecte de données sur le terrain a contribué à diffuser le programme européen auprès de ces lycéens.

Toutefois, en comparaison à l'ancienne enquête, nous avons décidé de rajouter des questions et nous ne nous sommes pas limité à des questions générales pour ainsi récolter des réponses plus précises et de laisser aux élèves des questions ouvertes.

Pour autant, nous ne demandons pas beaucoup d'informations en plus, car nous nous intéressons principalement à la précision des données afin d'affiner nos analyses.

3. Récolte des données :

Afin de récolter les données, il a fallu organiser des venues dans le lycée. Pour faciliter notre récolte et optimiser notre temps, nous avons préalablement réalisé une visite du lycée en compagnie de Mme. Jenvrin.

Nous avons ainsi pu voir les principales parties du lycée ainsi que quelques salles de classes spécifiques aux différentes filières.

La récolte des données s'est organisée en deux passages car les disponibilités des Master 2 et Licence 3 ne concordaient pas.

La première, début janvier, a été réalisée par les trois élèves en deuxième année de Master. Lors de cette matinée, ils ont été interrogés maintes classes du lycée général et technologique du lycée. Lors de cette matinée, plus de 200 questionnaires ont été récoltés, tous ou presque sous format informatique.

Dans un second temps, fin janvier, les trois étudiantes en troisième année de licence sont elles aussi allées récolter des données pour l'enquête. Cette fois, les élèves interrogés ont été ceux du lycée professionnel. Les élèves des 6 classes visitées se sont montrés à l'écoute, cependant à la place de la majorité des enquêtes informatiques du lycée général, nous avons dû utiliser le format papier.

Durant cette récolte, beaucoup d'élèves n'ont pas compris le sens de la question à propos de la connaissance de carburants alternatifs, ce qui s'est traduit par des réponses manquantes à la question ouverte. Certains aussi, ne comprenaient pas, malgré la définition, en quoi constituait le projet Erasmus, il a donc fallu les guider afin qu'ils puissent remplir le questionnaire aussi précisément que possible.

A la suite de ces deux matinées, nous avons pu récolter 441 réponses, ce qui est deux fois plus que l'objectif visé et ceci nous permettra de réaliser des analyses plus fines et plus précises.

4. Statistiques descriptives

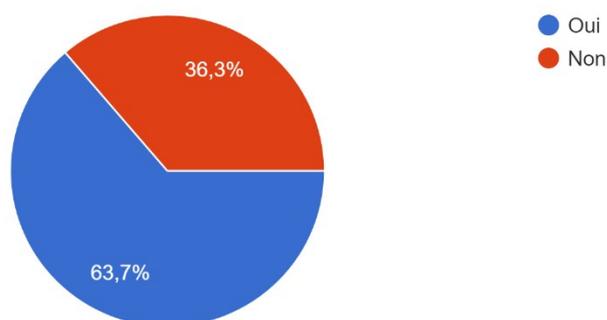
Avec ces données, nous pouvons en premier lieu faire un tri à plat dû au fait que nos données sont plus quantitatives que qualitatives. Nous savons par la première question, que sur les 441 élèves interrogés, 24 élèves ne connaissent pas le projet Erasmus, nous supposons par leur nombre peu élevé que ce sont des secondes où des élèves venant d'intégrer le lycée.

La plupart des élèves qui ont entendus parler du projet l'ont su de plusieurs manières, la plus récurrente étant par un professeur avec 302 réponses. Aussi étonnant que cela puisse paraître, la deuxième réponse la plus donnée n'est pas sur les réseaux sociaux ou par un camarade, mais par une affiche dans les couloirs avec 174 élèves. Alors que 94,6 % des élèves ont entendu parler du projet Erasmus en général, seuls 63,7 % d'entre eux sont au courant de l'existence du projet Erasmus au sein du lycée Lemonnier, mais seulement 10 % des élèves se disent très intéressés par le projet, soit 44 élèves.

Voici un graphique illustrant la réponse à la question 3 qui reprend les informations énumérés ci-dessus.

3 - As-tu entendu parler des projets Erasmus réalisés par l'Institut Lemonnier ?

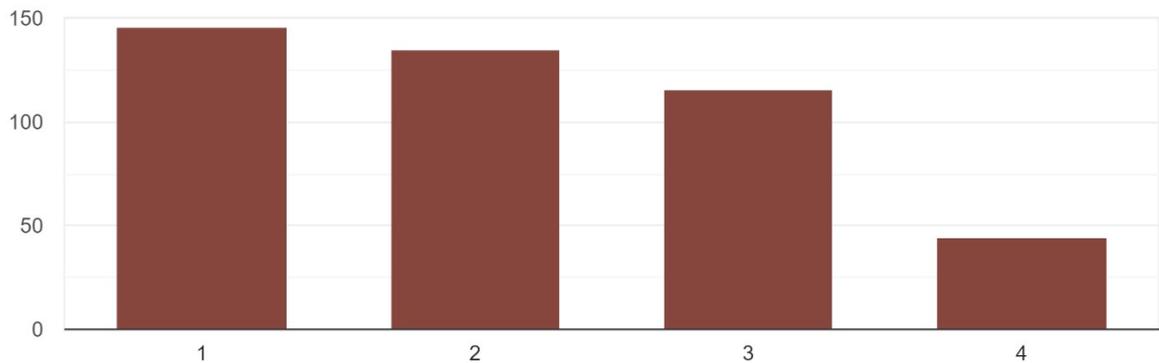
441 réponses



Le graphique suivant montre le pourcentage de réponses à l'intéressement d'un projet Erasmus. Comme dit plus haut, les élèves avaient la possibilité de répondre par 1, 2, 3 ou 4 en fonction de l'intensité de cet intérêt :

4 - Serais-tu intéressé par un projet Erasmus ?

441 réponses



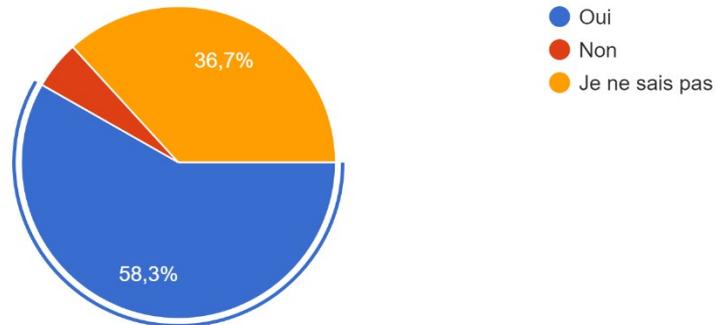
Ce nombre est très faible par rapport aux élèves se disant très peu intéressés par le projet dont le nombre est de 146 élèves, soit plus du triple des élèves très intéressés. Vu les réponses obtenues sur l'auto-évaluation de leur niveau d'anglais, nous pouvons supposer que certains élèves manquent de confiance en eux.

En effet, beaucoup ne se sentent pas capable d'avoir une conversation en anglais, mais pense que participer au projet Erasmus leur permettrait d'améliorer leur niveau d'anglais. L'anglais étant jugé par 72,6 % des élèves comme enrichissant pour la candidature à un futur emploi, ceci est expliqué par le fait que 90 % des élèves pensent qu'il est nécessaire aujourd'hui de parler anglais dans le monde du travail.

A la question sur le Développement Durable, nous avons la preuve que malgré la reformulation de la question, beaucoup d'élèves n'ont pas forcément compris le sens de cette question.

10 - Penses-tu que le Développement Durable est un facteur important à prendre en compte pour un projet Erasmus ?

441 réponses



Nous pouvons supposer que les élèves ne comprennent pas la question car 36,7 % des élèves ont répondu « je ne sais pas » à cette question. De même, 64,2 % des élèves ont répondu « oui » à la question à propos des carburants alternatifs, soit 283 élèves, or, à la question suivante, où il a fallu en citer, nous n'avons reçu que 271 réponses.

Cela signifie qu'au moins 12 élèves ayant dit connaître des carburants alternatifs n'en connaissaient finalement pas, où ne le disaient pas. Ensuite, nous avons demandé aux élèves s'ils seraient prêts à voyager à l'étranger pour un futur métier et si oui, combien de temps.

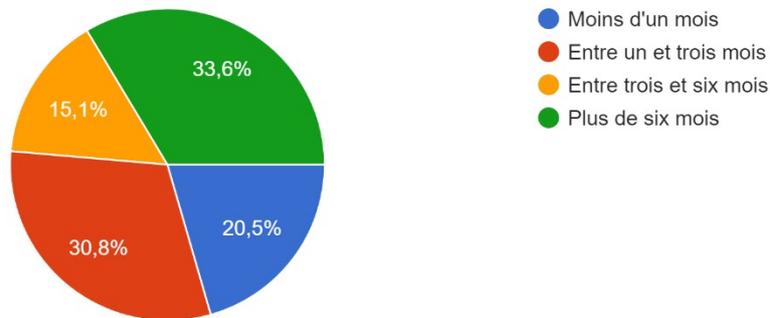
Nous nous rendons compte que les élèves sont perspicaces concernant l'importance de la mobilité pour un futur emploi, en effet 84 % des élèves sont prêts à passer plusieurs jours à l'étranger pour leur futur métier, cependant nous nous rendons compte que les durées envisagées sont très différentes selon les élèves.

Il n'y a pas de durée qui ressort majoritairement, un peu plus de 30 % des élèves sont prêts à partir entre un et trois mois, ce qui est presque autant que le nombre d'élèves prêts à partir plus de 6 mois.

Les élèves restants sont répartis entre les 15 % des élèves prêts à partir entre trois et six mois, et les 20 % des élèves prêts à partir moins d'un mois. On voit donc qu'il y a une grande disparité de réponses. Pourtant, 87 % des élèves sont déjà partis à l'étranger, la plupart du temps ils ont visité entre 1 et 3 pays. Seuls 30 % des élèves sont partis dans plus de 4 pays, et parmi ceux-là, un tiers dans plus de six pays.

16 - Combien de temps serais-tu prêt à partir à l'étranger pour tes études ?

438 réponses



Nous pouvons souligner que nous retrouvons le même cas de figure pour la question concernant les études ou le stage à l'étranger. Néanmoins, pour cette question, la durée qui ressort est de 1 mois avec 40 % des élèves ainsi que 25 % des élèves qui ne sont prêts à partir que deux semaines.

Cela nous permet donc de comprendre que malgré le fait que les élèves sachent l'importance de l'anglais et des langues étrangères en général, peu d'élèves sont prêts à partir longtemps loin de chez eux pour les études ou pour un stage.

Il est tout de même important de se remémorer que les élèves interrogés sont des lycéens, vivant chez leurs parents. Il est donc compréhensible que ce soit ces durées qui ressortent.

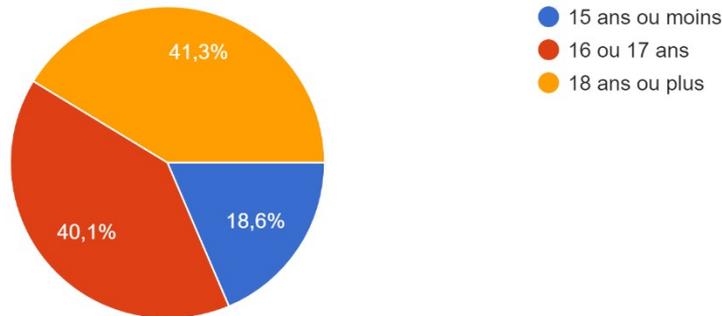
Concernant le financement du projet Erasmus, presque la moitié des élèves ne savent pas qu'il existe des aides pour participer au projet, ils ne savent donc pas d'où elles peuvent provenir.

Nous pouvons donc en conclure que même si les élèves connaissent l'existence du projet, que cela soit dans le lycée ou non, ils ne savent pas comment se déroule le projet, ni comment celui-ci a été créé et est maintenant dans les établissements, notamment au lycée Lemonnier.

Voici un graphique illustrant la réponse à la question :

Quel âge as-tu ?

441 réponses



Comme on peut le voir, 40 % des élèves ont 18 ans ou plus, ce qui peut expliquer certaines réponses sur le fait de partir pendant des durées plus longues à l'étranger. Il y a également 40 % des élèves qui ont 16 ou 17 ans, expliquant les réponses plus mitigées. Les réponses de ceux ne connaissant pas le projet Erasmus peuvent quant à elles provenir des 18 % des élèves ayant 15 ans ou moins.

Effectivement, les élèves de 15ans ou moins sont au nombre de 79, et les élèves n'ayant pas connaissance du projet Erasmus sont au nombre de 24, cela peut donc correspondre. Parmi les élèves interrogés, 44 % viennent du lycée technologique, 35 % du lycée professionnel et 21 % de BTS.

L'explication quant à cette dispersion au niveau des élèves peut être expliquée par des résultats parfois nuancés, nous pouvons notamment supposer que les élèves étant les plus favorables à partir pendant de longues durées à l'étranger sont en BTS, étant plus âgés, donc plus susceptible d'être plus indépendants.

5. Analyses statistiques

Nous avons pris l'initiative de réaliser 3 analyses statistiques différentes. D'une part nous avons réalisé deux régressions. Tout d'abord, une régression linéaire multiple puis une régression logistique. Par la suite, nous avons réalisé une Analyse en Composante Principale (ACP). Ainsi, dans cette partie, nous allons expliquer plus en détail chaque analyse statistique réalisée.

5.1. Préparation des données

Nous avons exporté les données depuis Google Form vers Excel et avons transformé chaque variable qualitative en variable quantitative.

Par exemple, concernant la question « Combien de temps serais-tu prêt à partir à l'étranger pour tes études ? » qui contient 4 modalités :

- Moins d'un mois
- Entre un et trois mois
- Entre trois et six mois
- Plus de six mois

Nous avons créé 4 nouvelles variables binaires correspondant à chaque modalité comme on peut le voir ci-dessous.

16 - Combien de temps serais-tu prêt à partir à l'étranger pour tes études ?	16 - Moins d'un mois	16 - Entre un et trois mois	16 - Entre trois et six mois	16 - Plus de six mois
Entre trois et six mois	0	0	1	0
Entre un et trois mois	0	1	0	0
Entre un et trois mois	0	1	0	0
Entre un et trois mois	0	1	0	0
Entre trois et six mois	0	0	1	0
Plus de six mois	0	0	0	1
Moins d'un mois	1	0	0	0

Figure 1 - Aperçu du recodage de la question 16

Ainsi, nous avons réalisé ce recodage pour toutes les variables qualitatives pour ne conserver dans notre jeu de données, que des variables quantitatives ce qui nous servira notamment dans la partie Analyse en Composantes Principales.

Lors de la conception du questionnaire, nous avons créé une question « classe » en question libre. Cela nous a posé des problèmes lors des futures analyses puisque chaque lycéen écrivait sa classe de manière différente et cela était très difficile de recoder cette variable. Nous avons donc pris l'initiative de ne pas conserver cette variable dans notre jeu de données.

5.2. Régression Linéaire Multiple (RLM)

5.2.1. Construction du modèle

Nous avons cherché à expliquer la variable « Serais-tu intéressé par effectuer un projet Erasmus ? » qui correspond à la question numéro 4 de notre questionnaire (Q4 qui correspond à notre Y). Celle-ci nous paraissait la plus importante à étudier afin de répondre aux demandes de l'Institut Lemonnier.

Pour trouver un modèle qui explique au mieux cette variable, nous avons décidé dans un premier temps d'inclure toutes les variables explicatives dans notre modèle pour ne pas faire de choix arbitraire sur les variables. Le modèle que nous avons obtenu n'était pas satisfaisant en terme de qualité du modèle. En effet, nous avons obtenu un R carré ajusté proche de 0.2 ce qui est très faible.

L'approche qui consiste à éliminer les variables non significatives d'un seul coup n'est pas bonne car certaines variables peuvent être corrélées à d'autres ce qui peut masquer leur influence sur Y.

Pour améliorer la qualité explicative de notre modèle, nous avons décidé de réaliser une sélection de variables pour n'avoir finalement dans le modèle, que les variables qui expliquent au mieux notre variable Y (Q4).

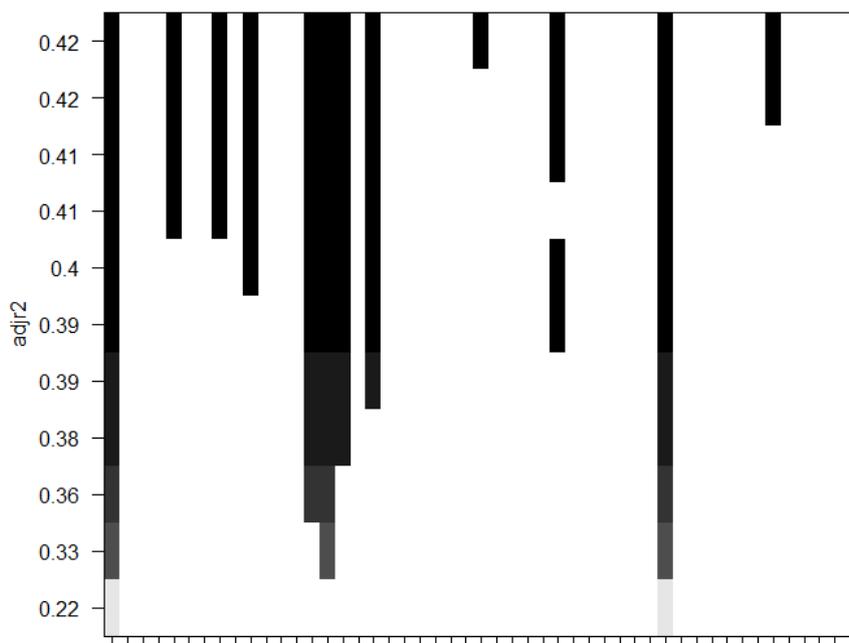


Figure 1 - Sélection des variables RLM

Comme on peut le voir sur l'image ci-dessus, on retient le modèle qui possède le plus grand R carré ajusté qui est ici de 0.42 correspondant aux « cases » noircies. Ce R carré ajusté est relativement faible ce qui veut dire que la qualité du modèle est moyenne. Néanmoins, au vu des résultats obtenus, nous observons que cette régression linéaire est très adaptée à nos données puisque nous obtenons une p value au test global de Fisher hautement significative de $2.2e-16$ donc bien inférieur à 0.001.

Ainsi, nous avons obtenu le modèle suivant :

```

reg_m = lm(X4...Serais.tu.intéressé.par.un.projet.Erasmus.. ~
X9...Penses.tu.que.participer.à.un.projet.Erasmus.peut.enrichir.ta.candidature.pour.un.futur.emploi.. +
X2...Un.professeur.m.en.a.parlé +
X3...As.tu.entendu.parler.des.projets.Erasmus.réalisés.par.l.Institut.Lemonnier..+
X5...Email+
X7...Penses.tu.que.le.projet.Erasmus.te.permettrait.d.améliorer.ton.niveau.d.anglais.. +
X8...Penses.tu.que.le.projet.Erasmus.te.permettrait.d.améliorer.tes.connaissances.en.Sciences.et.Technologies.. +
X11...Connais.tu.des.carburants.alternatifs.au.pétrole.pour.les.voitures..+
X18...1.à.3 +
X16...Moins.d.un.mois +
X19...Serais.tu.intéressé.par.effectuer.des.études.ou.un.stage.à.l.étranger..)

```

Figure 2 - Modèle RLM sélectionné

5.2.2. Validation des hypothèses standards

Dans le choix de notre modèle de régression linéaire multiple, nous avons posé les hypothèses suivantes :

- Indépendance des résidus avec les variables explicatives (Indépendance de ϵ et X_1, \dots, X_p)
 - Indépendances des résidus $\epsilon_1, \dots, \epsilon_n$
 - Normalité des résidus $\epsilon_1, \dots, \epsilon_n$
 - Egalité des variances $V(\epsilon_1) = \dots = V(\epsilon_n)$

1. Indépendance de ϵ et X_1, \dots, X_p

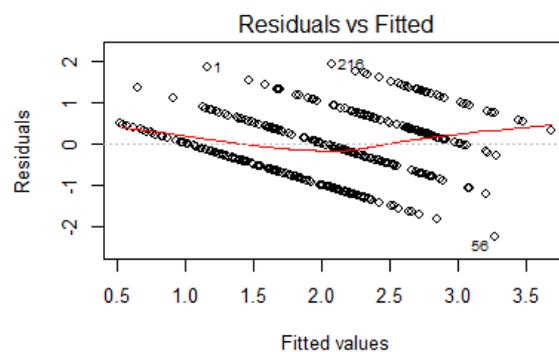


Figure 3 - indépendance des résidus avec les variables explicatives

Comme on peut voir sur le graphique ci-dessus, on constate que le nuage de point n'est pas ajustable par une droite ce qui confirme l'hypothèse standard d'indépendance des résidus avec les variables explicatives. On suppose donc que le lien linéaire entre Y et les variables explicatives X_p est justifié. Afin d'en être sûr, nous avons réalisé en parallèle, le test de Rainbow qui permet de conclure à la linéarité ou non du modèle, nous avons obtenu une p value de 0.256 qui est > 0.05 donc on admet que le modèle est linéaire.

2. Indépendance des résidus $\epsilon_1, \dots, \epsilon_n$

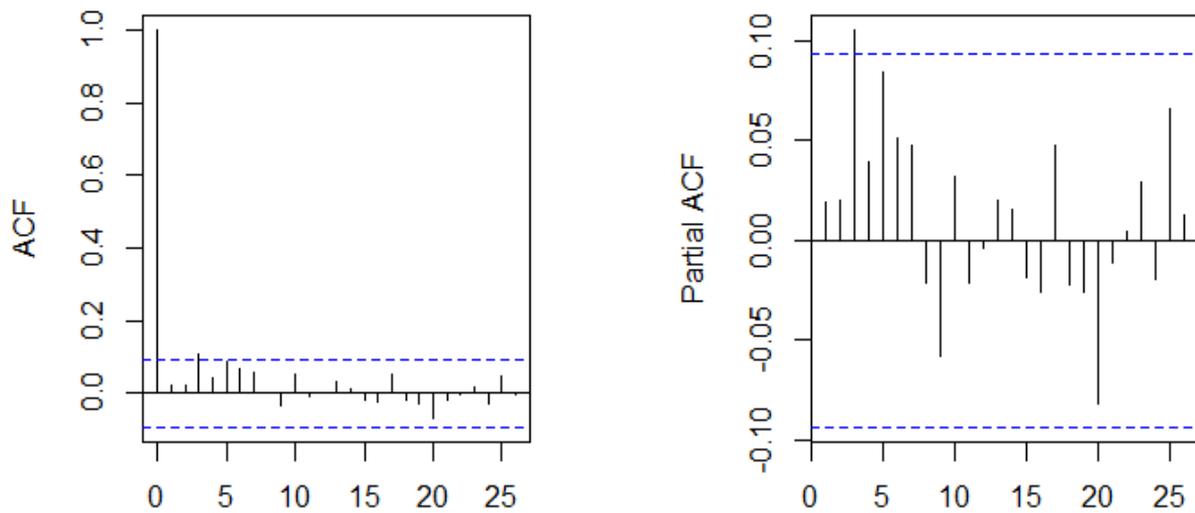


Figure 4 - Indépendance des résidus

On constate sur les deux graphiques qu'il n'y a pas de structure particulière dans la taille et dans l'orientation des bâtons. Par ailleurs, très peu de bâtons dépassent les bornes donc on peut dire qu'il y a indépendance des résidus.

Nous avons réalisé le test de Ljung-Box (test du portemanteau) et nous avons obtenu une p value de 0.684 ce qui est > 0.05 donc cela confirme l'indépendance des résidus.

Normalité des résidus $\varepsilon_1, \dots, \varepsilon_n$

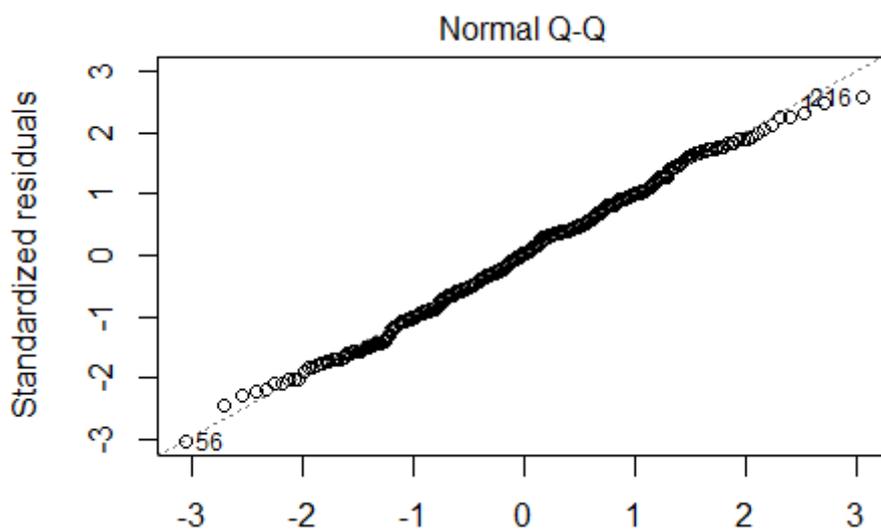


Figure 5 - Normalité des résidus

On remarque ici que tous les points sont bien alignés sur la droite d'équation $y = x$ ce qui démontre bien la normalité des résidus.

Nous avons par ailleurs réalisé un test de Shapiro-Wilk et avons obtenu une p value de 0.367 ce qui est encore une fois > 0.05 donc on admet la normalité des résidus.

Egalité des variances $V(\varepsilon_1) = \dots = V(\varepsilon_n)$

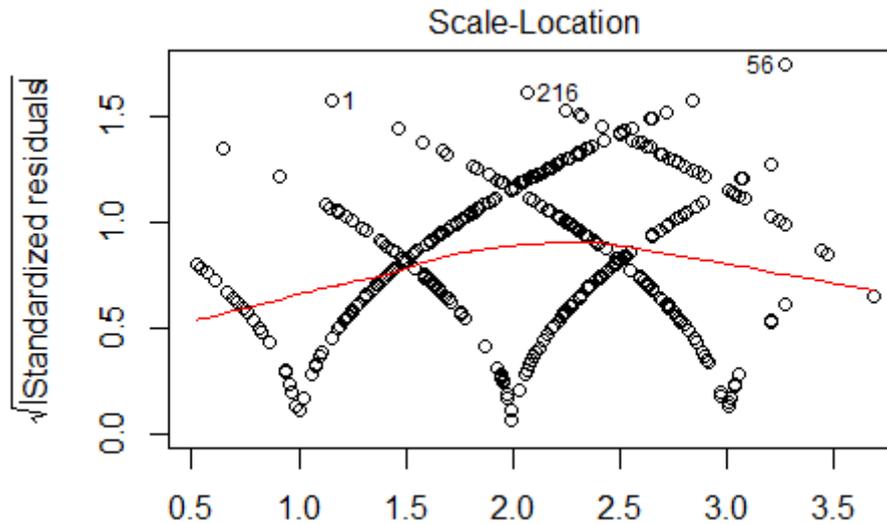


Figure 6 - Egalité des variances

On peut voir ici que les points ne sont pas disposés dans une forme « mégaphone » ni en « bandeau », cependant, il semble il y avoir une structure particulière en forme de « V ».

Au vu de la forme très particulière de ce graphique, nous avons réalisé un test statistique complémentaire, le test de Breusch-Pagan et avons obtenu une p value de 0.002 ce qui est < 0.05 donc on rejette l'égalité des variances. Il y a de l'hétéroscédasticité.

Nous aurions pu traiter ce problème avec deux méthodes :

- En essayant de transformer judicieusement Y (par exemple $\log(Y)$) mais cela est compliqué de réussir à trouver une bonne transformation de Y de façon directe.

En utilisant la méthode des MCQG : Moindres Carrés Quasi-Généralisé qui consiste à intégrer une matrice Omega dans le modèle, ce qui change la façon d'estimer les coefficients Beta.

Détection des valeurs anormales

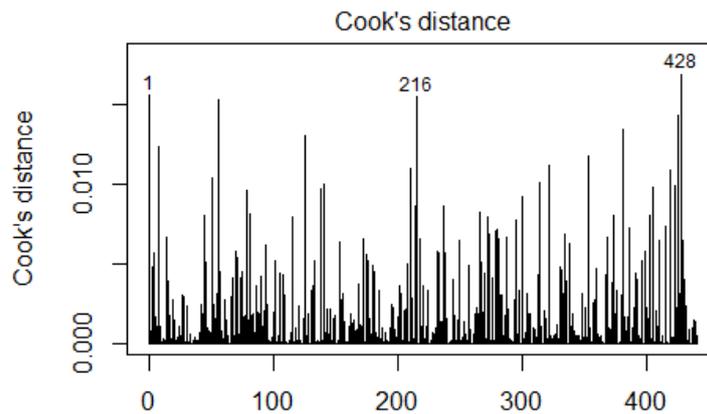


Figure 7 - Détection des valeurs anormales

On peut voir ci-dessus qu'aucune valeur ne dépasse 1 ce qui montre qu'il n'y a pas de valeurs anormales. On peut confirmer cela avec un test statistique, néanmoins, étant donné que les variances des résidus ne sont pas égales, nous avons décidé d'arrêter l'analyse.

Ainsi, nous avons vu dans cette partie que toutes nos hypothèses standards ont été vérifiées sauf celle sur l'égalité des variances des résidus. On conclut donc que ce modèle est faussé par l'hétéroscédasticité et n'est donc pas utilisable en tant que tel. Comme nous l'avons dit précédemment, il existe des méthodes de transformations de la variable Y afin d'essayer de corriger ce problème. Néanmoins, nous n'avons pas poursuivi nos recherches dans ce sens.

5.3. Régression Logistique

5.3.1. Construction du modèle :

Suite à nos résultats mitigés sur notre régression linéaire multiple, nous avons choisi de réaliser une régression logistique dans le but de prédire la variable « Serais-tu capable d'aller à l'étranger pour le travail ? » en fonction des cinq autres variables suivantes qui nous semblaient les plus pertinentes.

- L'âge de la personne.
- La filière d'études.
- La capacité à dialoguer en anglais.
- Le ressenti sur l'amélioration de son niveau d'anglais suite à un projet Erasmus.
- Le nombre de pays de l'Union Européenne déjà visité.

5.3.2. Pertinence du modèle :

Nous avons dans un premier temps réalisé un test statistique permettant de valider le choix d'une régression logistique. Nous avons effectué le test du rapport de vraisemblance, et nous obtenons une p.valeur <0.05 très faible. Afin d'aller dans ce sens, nous avons également effectué le test de Hosmer-Lemeshow pour évaluer la pertinence du modèle (on utilise la loi du Chi²). Toutes les variables

choisies ont un effet significatif sur la variable à expliquer. Ces tests nous permettent donc d'admettre que la régression logistique est tout à fait adapté à notre modèle.

Au départ, nous avons choisi de mettre une variable supplémentaire dans notre modèle : la variable « As-tu déjà voyagé dans un ou plusieurs pays de l'Union Européenne ». Notre régression donnait de très bonnes prédictions mais cela était biaisé par le fait que plusieurs variables étaient corrélées entre elles. Pour nous en rendre compte nous avons étudié la multi colinéarité entre les variables avec la méthode des « vifs ». Nous nous sommes donc rendu compte qu'il fallait enlever une variable pour corriger cela. A présent notre modèle ne contient plus que des vifs < 5 ce qui signifie que les variables ne sont pas colinéaires.

5.3.3. Qualité du modèle :

Pour mesurer la qualité de notre modèle nous avons tracé la courbe ROC ci-dessous.

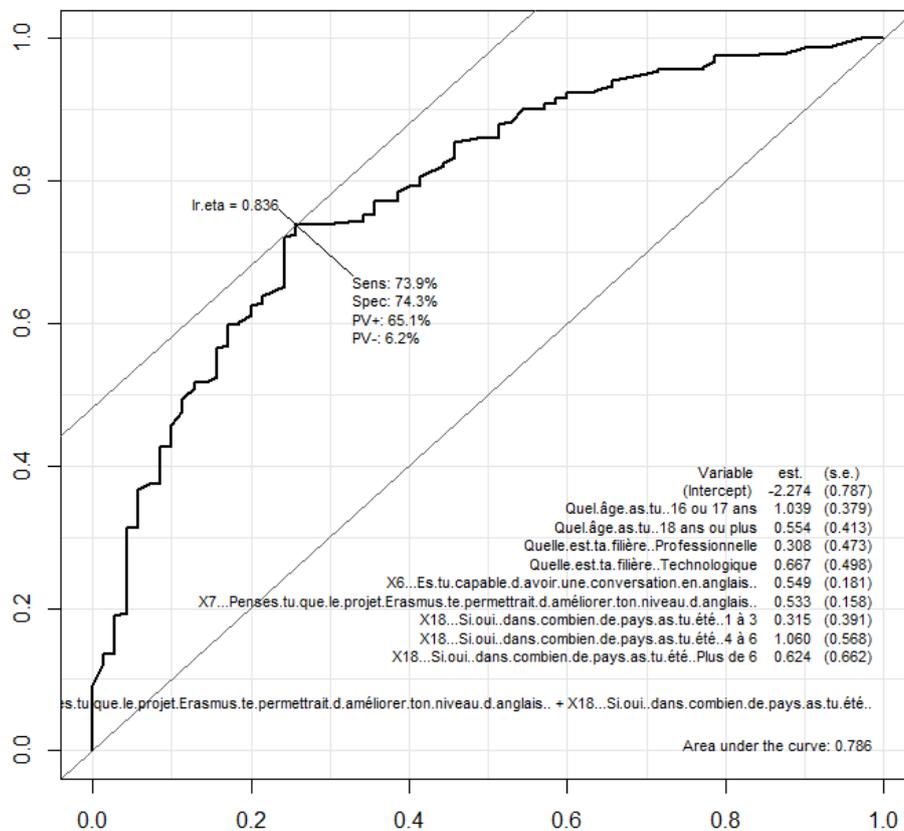


Figure 2 - Courbe ROC

Nous pouvons voir sur ce graphique que l'aire sous la courbe est assez importante (AUC = 0.786) c'est à dire qu'elle est comprise entre 0.75 et 1 ce qui signifie que notre modèle a une bonne qualité prédictive. Ainsi si nous interrogeons un nouvel élève sur les 5 questions que nous avons sélectionné, nous pourrions prédire avec peu de chance de se tromper si l'élève serait intéressé ou pas par partir à l'étranger pour le travail.

Nous avons donc à partir de nos 5 variables prédit pour chaque individus le fait ou non que celui-ci veuille partir à l'étranger. En effet, chaque élève se voit attribué une probabilité d'appartenir au groupe des personnes qui ont répondu « Oui ». Ainsi nous avons pu distinguer 2 groupes distinct, tous les individus qui ont obtenu une probabilité supérieure à 0.5 ont plus de chance d'appartenir au groupe des « oui » donc nous les avons mis dans ce groupe. Au contraire, ceux qui ont une probabilité inférieure à 0.5 ont été mis dans le groupe des gens qui répondent « Non ».

Afin d'évaluer le pouvoir prédictif de notre modèle, nous avons comparé les prédictions obtenues à l'aide de notre régression logistique avec les réponses réelles obtenues à la question étudiée. Ainsi

nous avons affiché la matrice de confusion qui permet de montrer les individus qui ont bien été attribué au bon groupe et les erreurs de prédiction qui ont eu lieu.

Le taux d'erreur de prédiction est de 15,2% ce qui est assez faible et nous montre une très bonne qualité prédictive de notre modèle.

5.4. Analyse en Composantes Principales (ACP)

Nous avons réalisé une ACP pour distinguer des groupes d'individus selon les réponses apportées au questionnaire. Pour cela, nous avons tout d'abord retiré toutes les variables qualitatives du jeu de données (classe d'âge, classe et filière). Nous avons conservé au total 15 variables pour réaliser notre ACP que l'on retrouve ci-dessous :

- 1 - As-tu déjà entendu parler du programme Erasmus ?
- 3 - As-tu entendu parler des projets Erasmus réalisés par l'Institut Lemonnier ?
- 4 - Serais-tu intéressé par un projet Erasmus ?
- 6 - Es-tu capable d'avoir une conversation en anglais ?
- 7 - Penses-tu que le projet Erasmus te permettrait d'améliorer ton niveau d'anglais ?
- 8 - Penses-tu que le projet Erasmus te permettrait d'améliorer tes connaissances en Sciences et Technologies ?
- 9 - Penses-tu que participer à un projet Erasmus peut enrichir ta candidature pour un futur emploi ?
- 10 - Penses-tu que le Développement Durable est un facteur important à prendre en compte pour un projet Erasmus ?
- 11 - Connais-tu des carburants alternatifs au pétrole pour les voitures ?
- 13 - Penses-tu qu'aujourd'hui, il est nécessaire de parler anglais dans le monde du travail ?
- 14 - Penses-tu que le projet Erasmus te permettrait d'améliorer ton niveau d'anglais professionnel ?
- 15 - Serais-tu capable de passer plusieurs jours à l'étranger pour ton futur métier ?
- 17 - As-tu déjà voyagé dans un ou plusieurs pays de l'Union Européenne ?
- 19 - Serais-tu intéressé par effectuer des études ou un stage à l'étranger ?
- 21 - Sais-tu qu'il existe des aides financières permettant de partir en programme Erasmus ?

5.4.1. Choix des dimensions

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	3.8346110	25.564073	25.56407
Dim.2	1.4315298	9.543532	35.10760
Dim.3	1.2169428	8.112952	43.22056
Dim.4	1.0971340	7.314226	50.53478
Dim.5	0.9590039	6.393359	56.92814
Dim.6	0.9251956	6.167971	63.09611
Dim.7	0.8868297	5.912198	69.00831
Dim.8	0.8171671	5.447781	74.45609
Dim.9	0.7685327	5.123551	79.57964
Dim.10	0.7444161	4.962774	84.54242
Dim.11	0.6472928	4.315286	88.85770
Dim.12	0.5095240	3.396827	92.25453
Dim.13	0.4912046	3.274698	95.52923
Dim.14	0.4122932	2.748622	98.27785
Dim.15	0.2583226	1.722151	100.00000

Figure 8 - Valeur propre et variance expliquée

On peut voir sur le tableau ci-dessus l'ensemble des valeurs propres obtenues (eigenvalue) triées par ordre décroissant ainsi que le pourcentage cumulé de variance expliquée (cumulative.variance.percent).

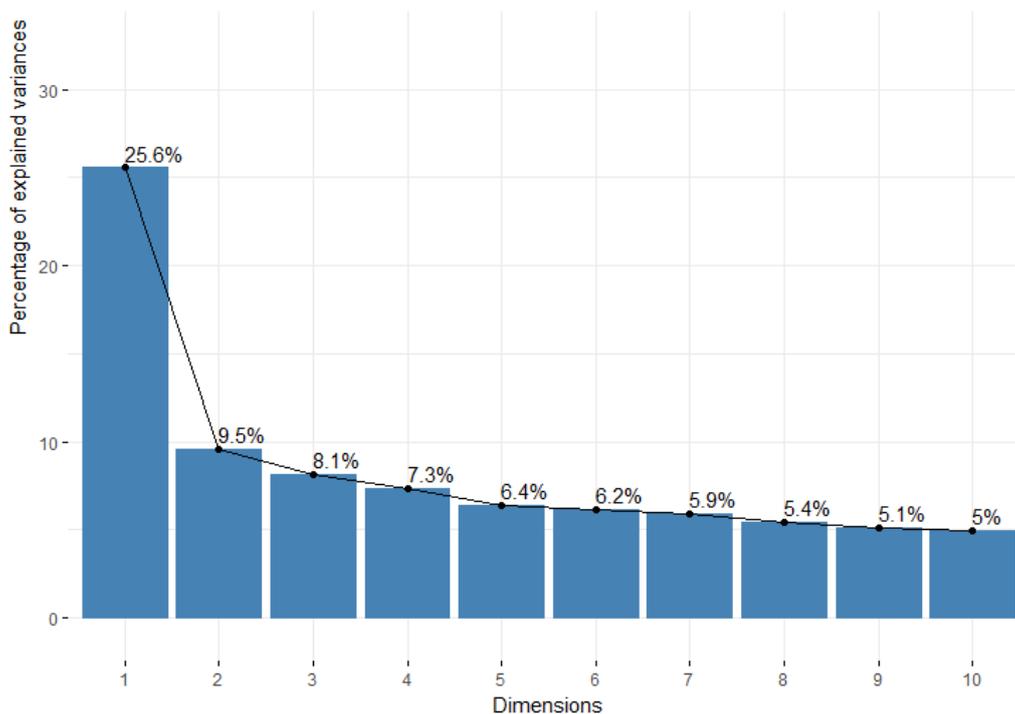


Figure 9 - Eboulis des valeurs propres

Nous avons décidé de conserver les 4 premières dimensions (critère de Kaiser) qui consiste à ne conserver que les valeurs propres > 1 , malgré le fait que 4 dimensions ne représentent que 50.5 % de la variance expliquée ce qui est assez faible. Si nous avons retenu le critère du coude, nous aurions choisi 2 dimensions car c'est à cet endroit-là qu'on observe un décrochement (coude) suivi d'une décroissance régulière.

5.4.2. Choix des classes et interprétation

Nous avons décidé de coupé notre arbre en 3 classes et nous obtenons le graphique suivant :

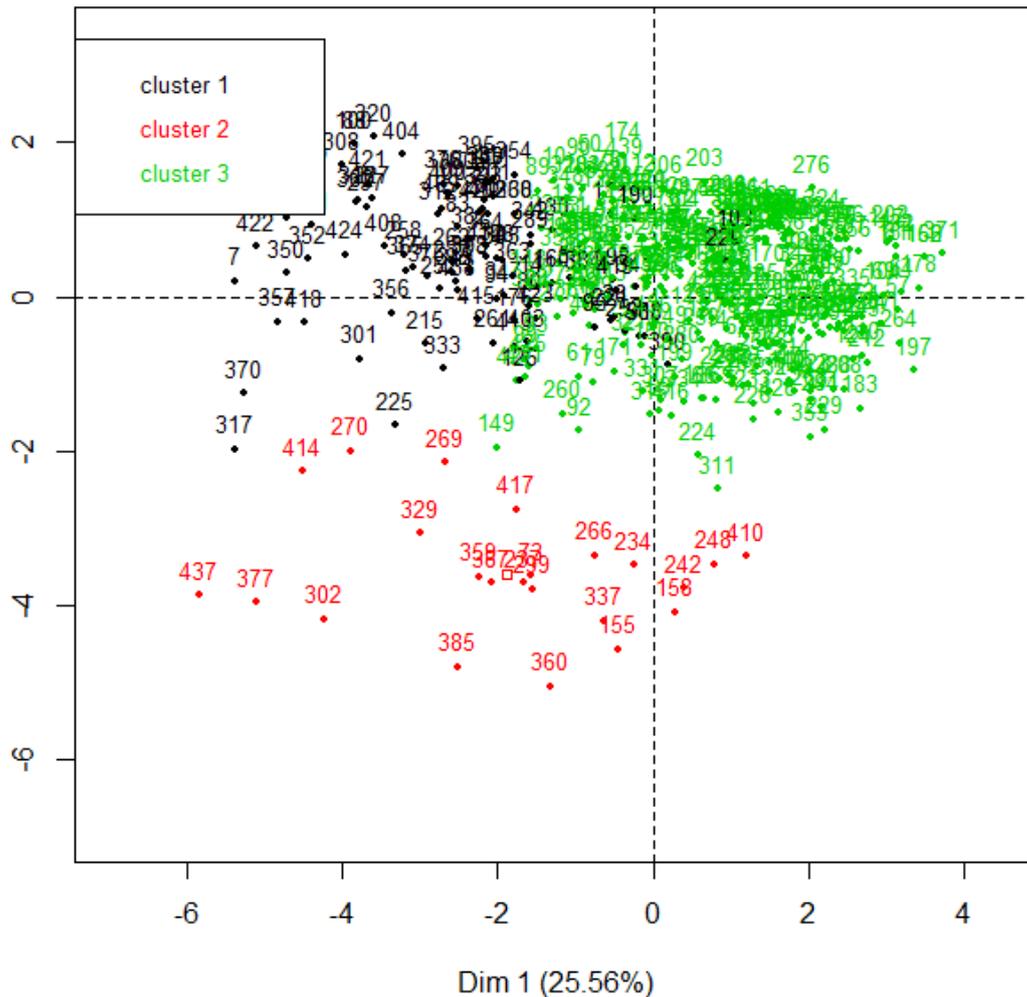


Figure 10 - Nuage de points des individus projetés

Nous avons par la suite cherché à interpréter chaque classe et nos deux axes (en pointillés). Afin d'interpréter les groupes, nous avons utilisé le jeu de données des 15 variables en ajoutant en plus, une variable de groupe (qui vaut 1, 2 ou 3) ce qui nous a permis de faire un filtre sur les groupes et de voir quelles étaient leurs caractéristiques.

- Groupe 1 (en noir) – 209 élèves :

Ce sont des élèves qui ont tous entendu parlé du programme Erasmus. Néanmoins, ils sont peu intéressés par celui-ci. Ils ont un niveau d'anglais moyen. En effet, on retrouve majoritairement des « pas du tout » et « pas vraiment » pour la question « Es-tu capable d'avoir une conversation en anglais ? ».

L'anglais reste tout de même nécessaire pour eux dans le monde professionnel. Enfin, ce sont des élèves qui ont majoritairement voyagé dans plusieurs pays de l'Union Européenne.

- Groupe 2 (en rouge) – 27 élèves :

Tous les élèves qui ne connaissent pas le programme Erasmus sont dans ce groupe (24 élèves sur les 27). Par ailleurs, l'anglais ne semble pas être important pour eux et leur niveau est relativement faible (majoritairement des « pas du tout » pour la question « Es-tu capable d'avoir une conversation en anglais ? »). Ils ne sont pas intéressés par partir à l'étranger et s'ils doivent partir pour leurs études, ce serait pour moins d'un mois.

- Groupe 3 (en vert) – 205 élèves :

Ce groupe est assez ressemblant au premier mais il se démarque notamment par un niveau d'anglais bien supérieur et par l'envie de s'investir dans un projet Erasmus. En effet, on retrouve majoritairement des « Oui beaucoup » pour la question « Es-tu capable d'avoir une conversation en anglais ? ». Ils pensent aussi que l'anglais est important pour enrichir une candidature et aussi pour réaliser un projet professionnel. Ils se sentent également capable de partir à l'étranger pour un futur emploi ou pour un stage ou même pour des études

Ainsi, en vue des résultats obtenus, nous avons interpréter les axes de la manière suivante :

Par rapport à la moyenne d'un individu, plus un élève se retrouve à droite, meilleur est son niveau d'anglais. Au contraire, plus il est à gauche, plus son niveau d'anglais est faible. Nous avons également constaté que plus les individus sont à droites et plus ils ont envie de partir à l'étranger pour une longue durée.

Par ailleurs, plus un élève se retrouve en haut de l'axe des abscisses et plus il aura connaissance des projets Erasmus et plus il aura envie de s'y investir. A contrario, plus il est en bas de l'axe des abscisses et moins il aura connaissance des projets ainsi que l'envie d'y participer par rapport à la moyenne des individus.

Conclusion

Au fil de ce projet, nous avons eu l'opportunité de travailler sur une enquête dans sa globalité. C'est-à-dire que nous avons pris contact avec le client afin de comprendre les besoins et objectifs de l'enquête. Nous avons par la suite réalisé le questionnaire en collaboration avec Mme Jenvrin, enseignante en mathématiques et coordinatrice du projet Erasmus. Par ailleurs, nous sommes allés sur le terrain pour mener l'enquête auprès des lycéens concernés par le sujet afin de récolter les réponses. Il en fallait au moins 200 et nous en avons obtenu environ de 440. Enfin, nous avons travaillé sur les données afin de remettre nos résultats dans les temps impartis.

Ce projet professionnalisant nous a permis de découvrir le monde de l'enquête statistique dont nous avons eu des connaissances théoriques au cours du Master 1 et que nous avons pu mettre en application au cours de ce projet.

Par ailleurs, ce projet nous aura permis de travailler en équipe, d'être autonome et rigoureux.

Nous espérons que nos travaux permettront le rayonnement de l'Institut Lemmonier au près d'Erasmus.