



UNIVERSITÉ DE NANTES



Master 1 Économétrie et Statistiques, parcours Économétrie Appliquée

Enquête sur les pratiques culturelles

Analyse économétrique sur 350 étudiants de Nantes

Auteurs

Laurianne MORICEAU

Diane THIERRY

Année universitaire 2019-2020

Abstract

This paper aims to understand the main determinants of the cultural consumption upon a questionnaire of 350 students from Nantes, through three different analyzes. Cultural consumption is measured by the frequencies of various types of cultural activities such as theatre, opera, library but also radio, video games, television etc. Firstly we studied the determinants of going or not to the theatre using a binary response model. We initially estimated a model which supposes homoscedasticity of errors and check all the hypothesis of logit validation model, while in a second time we also estimated a model which supposes now the contrary : the heteroscedasticity of errors. We can notice two facts: a student who goes at the museum and at festivals 3 to 10 times a year is more likely to go to the theatre compared to a person who never goes at that kind of activities. Moreover a student who has been sensibilized at culture during his childhood is more likely to go to the theatre, when a student who watch television is less likely to go. Secondly we studied the frequencies of cultural activities ordered from rare to most frequent, the aim was to highlight the differences of recurrence of every kind of cultural practices of Nantes students. Ours results shows that the sensibilization about culture, the budget and the level of education influence positively the high frequencies of cultural habits compared to a student in BTS, who didn't received a cultural education and without budget for culture.

Résumé

L'objectif de cette étude est d'analyser les différents déterminants de la consommation culturelle¹. Les données ont été récoltées auprès de 350 étudiants Nantais de tous horizons. Deux méthodes statistiques ont été utilisées. D'une part nous avons réalisés une analyse logit binaire sur la fréquentation ou non des théâtres par les étudiants. D'autre part nous avons réalisé un modèle multinomial ordonné sur la fréquence des pratiques culturelles (opéré, radio théâtre, etc.). Le premier modèle a été estimé deux fois : avec l'hypothèse d'homoscedasticité supposée et sans. Ainsi ,nous avons obtenu qu'un étudiant se rendant au musée et à des festivals entre 3 à 10 fois par an, ira plus probablement au théâtre que s'il ne pratiquait aucune de ces activités. De même que si l'étudiant a été sensibilisé à la culture durant l'enfance, les probabilités qu'il fréquente les théâtres est plus importante que pour un individu qui regarde régulièrement la télévision. Concernant la fréquence des pratiques culturelles, nous avons obtenu que la sensibilisation à la culture, le budget et le niveau d'éducation influençait positivement la variable.

¹ Le terme "*consommation culturelle*" est un abus de langage né de la sociologie.

Sommaire :

Partie 1 : Analyse économique -----	p.4
Partie 2 : Analyse économétrique-----	p.12
Partie 3 : Conclusion et discussion des résultats -----	p.50
Partie 4 : Annexes -----	p.54
Partie 5 : Bibliographie -----	p.70
Partie 6 : Table des matières -----	p.71

PARTIE 1 : Analyse économique

I- Introduction

*“Notre histoire”,
“Une ouverture, la découverte”,
“L’ennemi du capitalisme, l’ami de l’Homme cultivé”,
“Le développement du nombril narcissique afin de briller en société”,
“Un luxe”... ²*

Voilà des réponses antinomiques issues de notre enquête réalisé sur 350 étudiants Nantais de tous horizons, à la question : que représente l’art pour vous ?

La diversité des définitions témoigne de la polysémie du terme et de la confusion qu’il engendre quant à sa détermination. En effet, la culture englobe à la fois les connaissances, les traditions mais aussi les gestes et les idées qui caractérisent l’appartenance à une société.

Dans notre étude, nous nous concentrerons uniquement sur une définition de la culture comme pratique, perfectionnement de mode de connaissance ou d’expression, dans un domaine d’activité intellectuelle, artiste, etc. En outre, nous nous intéresserons uniquement aux pratiques culturelles des étudiants Nantais.

La culture sous toute ses formes a un caractère indispensable tant au plan collectif qu’individuel. En effet au niveau sociétal, la culture, au sens de civilisation, est quelque chose d’immatériel qui crée une cohésion en partageant une même histoire, des mêmes rites et valeurs entre les individus créant ainsi un sentiment d’appartenance à un groupe social. De même, au niveau individuel la culture permet aux individus de se forger une identité propre, de mieux connaître le monde, elle crée des chemins de réflexion et est ainsi vecteur de socialisation et d’intégration. C’est la raison pour laquelle l’existence de pratiques culturelles au sein des étudiants nous semble importante, peu importe la manière dont elle se

² Témoignages issues de notre enquête auprès des étudiants Nantais (2020)

matérialise, musiques, sorties, patrimoine, etc. Les formes d'expressions artistiques sont nombreuses et ne cessent d'évoluer avec les transformations sociétales. Le gothique a fait place au baroque, le fauvisme à l'expressionnisme, les transformations numériques ont révolutionnées nos manières de consommer la culture avec les ordinateurs, les jeux-vidéos... Historiquement, il existe six formes d'art que nous connaissons comme "arts supérieurs", c'est à dire créés pour être appréciés avec les sens : l'architecture, la sculpture, les arts visuels, la musique, la littérature et les arts de la scène. À cette classification a été ajouté l'art du cinéma en 1911 lorsque le théoricien et critique cinématographique Ricciotto Canudo emploie pour la première fois l'expression d'un 'septième art'. On parle aujourd'hui de huitième, neuvième voire dixième art en référence aux arts numériques, aux arts culinaires, à la photographie, la bande dessinée... Cela nous montre à quel point la société dans laquelle nous vivons est en perpétuel mouvement, et donc que la culture qui nous définit évolue elle aussi fortement. C'est pourquoi chaque forme de culture doit-être considérée, et non pas exclusivement les arts supérieurs ou culture dite « légitime ». D'autant plus que l'on observe aujourd'hui une sorte de dichotomie entre les pratiques culturelles plus traditionnelles et les pratiques plus jeunes, témoignant d'un intérêt et d'un accès divergent aux différents types de culture.

Ainsi, chacun doit pouvoir s'il le souhaite accéder à la culture et trouver la forme d'art qui lui correspond en fonction de ses affinités. L'accès à la culture et sa démocratisation font partie intégrante des politiques publiques liés au ministère de la culture. En effet, André Malraux alors nommé premier ministre d'Etat en charge des affaires culturelles, rédige le décret du 24 juillet 1959, dans lequel il définit les missions de son ministère : *«la mission est de rendre accessibles les œuvres capitales de l'humanité, et d'abord de la France, au plus grand nombre possible de Français, d'assurer la plus vaste audience à notre patrimoine culturel et de favoriser la création des œuvres l'art et de l'esprit qui l'enrichissent ».*

La mission de service publique est-elle alors remplie par nos gouvernements ?

L'enquête réalisée auprès des étudiants Nantais sur leurs pratiques culturelles tente alors de répondre aux questions liées à ces objectifs. Existe t-il des obstacles cognitifs qui empêcheraient certaines catégories de personnes de s'y intéresser tels que l'absence de sensibilisation dans l'enfance ? Les pratiques sont-elles les mêmes pour tous les étudiants ? ont-ils les mêmes ressources ?

En premier lieu nous identifierons les différents déterminants de ces pratiques culturelles que nous différencions en deux catégories : les déterminants inhérents à la personnes et donc liés aux caractéristiques socio-économiques de l'individu puis les déterminants extrinsèques liées à l'accessibilité de ces pratiques révélant ainsi l'efficacité ou non des politiques publiques. Puis, nous évoquerons les principaux faits stylisés révélés par notre enquête.

Dans un second temps, nous tenterons de comprendre les raisons qui poussent les étudiants à se rendre ou non au théâtre par un modèle de choix binaire puis nous tenterons d'évaluer la fréquence des pratiques culturelles par un modèle multinomial ordonné puis nous essayerons de comprendre la volonté des étudiants de diversifier ou d'améliorer leurs pratiques culturelles par un modèle de choix binaire. Les résultats nous permettront de conclure sur l'existence d'une potentielle dichotomie de la culture caractérisée par une culture dite « traditionnelle » et une néo-culture dite « jeune » grâce à un modèle multinomial non ordonné.

II- Analyse économique des variables

A) Les déterminants socio-économiques des pratiques culturelles

1- Les déterminants sociologiques

De nombreux sociologues se sont intéressés à la culture dans sa globalité, afin de comprendre, d'expliquer et d'analyser les différents rapports à la culture d'une société. En effet, les habitudes culturelles seraient intimement liées à la stratification sociale de celle-ci puisque ces pratiques seraient en fait des rituels d'identification de la vie sociale. En outre selon Bourdieu, « les classes sociales se distinguent ainsi les unes des autres par le partage et la transmission d'un certain nombre de traits culturels qui conditionnent les comportements individuels dans un grand nombre de domaines ». Par conséquent, les pratiques culturelles seraient le résultat d'un « *habitus* », qui renforcerait le sentiment d'appartenance à un groupe par rapport à une autre. Ainsi, des différences entre groupe sociaux résulte une scission entre les cultures, une culture dite « dominante » des classes supérieures qui imposent leurs goûts aux classes inférieures garantes d'une « sous-culture ». Ce capital culturel peut s'appréhender sous 3 formes ; le capital incorporé qui englobe l'*habitus*, le capital objectivé par la possession d'objets culturels et le capital institutionnalisé légitimés par les diplômes.

Nous savons alors que la classe sociale dont l'étudiant est issue est déterminante des habitudes de consommation culturelle, et que la sensibilisation à tout type de culture durant son enfance est prépondérante. Un individu sensibilisé dès le plus jeune âge par la pratique de la musique, de la lecture, habitué aux sorties culturelles (théâtre, conférences) sera plus enclin à reproduire ces habitudes dans le futur. De nombreuses enquêtes sociologiques ont démontré les résultats suivants, presque triviaux, sur les différences de pratiques culturelles selon la catégorie sociale professionnelle (CSP) d'appartenance :

➤ Les cadres et professions intellectuelles supérieures vont plus au théâtre, visitent des monuments historiques, vont à la médiathèque ou à la bibliothèque, pratiquent un instrument de musique, lisent davantage de livres et regardent 'Arte' au moins une fois par semaine.

- Les résultats montrent que la fréquence de ces pratiques décroît à mesure que les classes sont considérées comme inférieures relativement à leur revenu.
- En ce qui concerne les activités plus accessibles et plus répandues comme le cinéma ou les concerts de rock, les cadres arrivent également en tête.

Ainsi, dans notre étude, 14 variables représentent l'influence de l'*habitus* : la catégorie socio-professionnelle des deux parents, la sensibilisation à toute forme de culture, et comme conséquence de ces 3 variables ; les fréquences de sorties à divers lieux culturels (tels que musées, expositions, opéra). Les activités culturelles praticables chez soi telles que livres, et radios font également partie de notre analyse. Nous avons aussi inclu télévision et jeux vidéos dans le but de distinguer une culture traditionnelle et une culture jeune dans un modèle multinomial non ordonné.

Tableau 1: Question n°1 et 2 du questionnaire : variables à expliquer

Par an en moyenne, à quelle fréquence allez-vous ... ? *							Par an en moyenne, à quelle fréquence pratiquez-vous ces activités ? *						
	Jamais	1 à 2 fois par an	3 à 10 fois par an	1 fois par mois	Au moins 1 fois par semaine	Ne sait pas		Jamais	1 à 2 fois par an	3 à 10 fois par an	1 fois par mois	Au moins 1 fois par semaine	Ne sait pas
À l'opéra	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Radio	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Au théâtre	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Télévision	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
À des concerts ou festivals	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Jeux vidéos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
À des conférences	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Lectures en tout genre	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Au musée ou expo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>							
Au cinéma	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>							
À la bibliothèque ou médiathèque	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>							

Les 2 premières questions de notre questionnaire concernent donc la fréquence de pratique de certaines activités culturelles, volontairement diverses, avec des fréquences allant de 'jamais' à 'une fois par semaine'. Le choix de ces fréquences a été influencé par une analyse déjà réalisée sur le sujet en 2013 par des étudiants du campus de Toulon dans le cadre des rencontres culturelles organisées par leur université. Leur fréquence maximale était "*plus de 4 fois par an*". Ayant ajouté des pratiques plus 'contemporaines' visibles en deuxième question

du tableau n°1, nous avons allongé les fréquences en proposant jusqu'à 1 fois par semaine puisque nous nous doutions que les activités proposées induisent des '*consommations*' plus fréquentes. Pour rester précis et éviter d'avoir la majorité des répondants dans la modalité "plus de 4 fois par an" nous avons allongé les propositions de fréquences.

En relation avec les 11 variables issues de ces 2 tableaux, nous avons demandé l'intérêt qu'ont les individus à la pratique éventuelle des activités montrées en première question du tableau n°1. Ainsi, cette autre question avait pour but de souligner l'écart entre '*la soif*' de culture chez les étudiants et les pratiques effectives.

La proximité aux différents types de culture soulignée précédemment ne dépend pas uniquement de la CSP d'origine, mais aussi de l'âge, du niveau de diplôme, et du secteur d'étude. En effet, ces deux derniers peuvent conférer aux individus un nouveau statut social, indépendamment de la CSP d'origine par le rôle de l'école comme "*lisseur social*". Les études ont montré que dans ce cas, les pratiques et consommations culturelles s'alignent à la nouvelle position sociale des individus, alors plus élevée, s'appropriant ainsi la "*culture des héritiers*". Ainsi, si l'école remplit réellement son rôle de lissage des inégalités, la culture ne devrait pas être corrélée à l'origine sociale des parents, mais davantage au niveau de diplôme de l'individu. En outre, on s'attend à observer une relation positive entre le niveau de diplôme et la fréquence des activités culturelles : des étudiants très diplômés (master, doctorat par exemple) peuvent se rapprocher davantage d'une culture classique. Nous quantifierons cet effet grâce aux variables de l'âge, de la formation et du domaine d'étude.

2- Déterminants économiques

Certaines activités présentent un coût d'entrée relativement élevé, c'est le cas de l'opéra par exemple, ce qui peut parfois freiner les envies individuelles de pratiquer telle ou telle activité. Ainsi, selon le revenu disponible³, et après avoir dépensé pour ses besoins primaires, l'individu peut consacrer le reste de son budget au divertissement. La culture peut remplir ce rôle. Par conséquent plus le revenu disponible est élevé plus l'individu pourra

³ Le revenu disponible d'un ménage comprend les revenus d'activité (nets des cotisations sociales), les revenus du patrimoine, les transferts en provenance d'autres ménages et les prestations sociales (y compris les pensions de retraite et les indemnités de chômage), nets des impôts directs.

allouer d'argent aux sorties culturelles. Il est vrai que le revenu des étudiants (bourse, travail, versement des parents) est généralement faible mais nous voulions mesurer l'effet de celui-ci sur les différentes sorties culturelles. Ainsi, nous avons posé une question concernant le budget alloué aux sorties culturelles par mois en créant des tranches allant de "0€" à "plus de 100€", mais aussi des questions de type binaire pour savoir si les individus sont boursiers et s'ils travaillent en dehors du temps scolaire. En effet, si oui, cela peut refléter un manque de ressources qui est alors complété par un job étudiant ou par des subventions régionales. Nous savons que plus l'individu a de revenu disponible, moins la pratique d'une activité culturelle onéreuse aura d'impact sur son portefeuille. En outre, le coût à l'entrée de l'activité est moindre relativement à son revenu, comparé à d'autres dont le revenu serait plus modeste.

B) Les déterminants extrinsèques des pratiques culturelles

Nous avons ainsi vu comment le prix des activités peut être un frein à la consommation culturelle. Dans cette logique de comprendre les raisons qui poussent les étudiants Nantais à pratiquer des activités culturelles ou non, nous avons établi une liste de 6 facteurs regroupant les principaux freins aux pratiques culturelles. Aussi, on retrouve le coût des activités, le manque de temps, le manque de proximité, le manque d'informations sur les événements, le manque d'attraction de l'offre culturelle, et le sentiment que certaines pratiques sont réservées à une catégorie de personnes.

Le lieu d'habitation est non sans conséquence sur la consommation culturelle. En effet, ce sont dans les villes que se concentrent la plupart des institutions culturelles, musées, théâtres, opéra, conservatoires... L'offre culturelle y est alors plus importante et les possibilités d'activités sont parfois abondantes. La proximité du lieu d'habitat avec les lieux culturels est corrélée positivement à la fréquence des pratiques culturelles pour deux raisons évidentes : la première est le coût moindre engendré par le déplacement lorsque l'institution se trouve à proximité du lieu d'habitation, tandis que la seconde réside dans l'effort à fournir.

La dernière modalité de la variable concernant les freins à la consommation culturelle est intéressante puisqu'elle soulève un nouveau problème ; un aspect psychologique au fait de ne pas s'intéresser à la Culture. Nous pouvons aisément faire le lien avec des réponses à la question ouverte que nous avons posée concernant la définition de la Culture, telles que "Le

développement du nombril narcissique afin de briller en société” ou encore “La culture c’est soit pour les péteux, soit pour les personnes du même point de vue”. La scission entre la culture jeune et traditionnelle est d’une certaine manière ancrée dans les mentalités et peut expliquer le cercle vicieux qui éloigne de plus en plus la culture des classes supérieures à celle des classes inférieures. De plus, comme nous l’avons évoqué précédemment, l’école peut avoir un rôle à jouer dans l’atténuation des inégalités sociales préexistantes. Cependant, la culture enseignée à l’école est exclusivement celle de la culture dite “légitime”, valorisée par rapport à une “sous-culture”, plus accessible mais qualifiée seulement de divertissement. Selon Bourdieu, ces institutions imposent les normes actuelles de la société et ne permettent pas aux autres cultures d’être autant considéré. Dans notre étude nous prendrons en compte toutes sortes de cultures, incluant les arts nouveaux tels que les jeux-vidéos, le cinéma, la télévision, et les concerts tous styles confondus. L’objectif étant de mettre en évidence une potentielle dichotomie de ces pratiques dites “jeunes” et les pratiques plus anciennes.

Nous avons tenté de mesurer ce phénomène en demandant aux étudiants s’ils souhaiteraient améliorer ou diversifier leurs pratiques culturelles, et s’ils participaient aux activités culturelles proposées par leurs établissements. De plus, nous trouvions intéressant de voir si ce dernier était public ou privé de manière à noter ou non, une différence significative dans les pratiques culturelles des étudiants de ces 2 groupes.

De plus, le genre de l’étudiant peut également être pertinent puisqu’en 2017-2018, les 44 écoles nationales supérieures d’arts ont accueilli et formé 11 100 étudiants, dont 66 % de femmes⁴.

⁴. <https://www.cairn.info/chiffres-cles-statistiques-de-la-culture-2019--9782111399587-page-159.htm?contenu=resume> (p.156) consulté le 27/02/2020

PARTIE 2 : Analyse économétrique

Pour rappel nous avons récolté 350 réponses des étudiants nantais quant à leurs pratiques culturelles, pour un total de 34 variables. L'objectif de cette étude est de constater quelles sont les pratiques culturelles des étudiants de Nantes et d'en comprendre les déterminants. Pour cerner au mieux les types de consommation culturelle des jeunes et leurs fréquences nous procéderons à 3 analyses. Dans un premier temps nous étudierons les déterminants de la consommation de théâtre chez les étudiants à l'aide d'un modèle binaire, puis nous analyserons les fréquences de consommation de chaque pratique avec un modèle multinomial ordonné pour finir sur un modèle multinomial non ordonné qui nous permettra de voir si les jeunes ayant répondu à notre enquête consomment davantage une culture "jeune" ou une culture plus "traditionnelle", dont la différence a été soulignée dans la partie précédente.

Avant toute chose, il convient de nettoyer notre base de données pour la rendre exploitable pour la suite de l'analyse. Ainsi, dans un premier temps nous procéderons à une clarification de la base et au recodage de certaines réponses pour faciliter l'analyse, puis nous étudierons les éventuelles valeurs atypiques pour terminer sur les statistiques associées à chacune des variables.

I- Nettoyage de la base de données

A) Clarification de la base pour faciliter la lecture

Le nom associé à chaque variable correspond à l'intitulé complet de la question telle que nous l'avons posée lors du sondage. De manière à faciliter les manipulations nécessaires à l'analyse ainsi que ses interprétations, nous allons procéder au renommage de nos variables. Ainsi, par exemple la première variable de notre base "Par an en moyenne, à quelle fréquence allez-vous ... ? [À l'opéra]" devient "freqOpera" puisqu'il s'agit de la fréquence de sorties à l'opéra de chaque étudiant nantais ayant répondu à notre questionnaire.

De même, nous procédons au regroupement et au recodage des variables. En effet pour 3 des 34 questions posées, une réponse 'Autre' était proposée, il est donc nécessaire de trier

ces réponses soit en les regroupant avec les propositions existantes si elles s'y réfèrent, soit en créant une nouvelle modalité si plusieurs réponses 'Autre' induisent un même type de réponse. On commence ainsi avec la variable "Pour vous quels sont les principaux freins aux pratiques culturelles ?" pour laquelle nous avons reçu de nombreuses réponses telles que "la flemme", "pas intéressé", "pas envie" il est vrai que nous n'avons pas proposé de réponses s'y référant. Nous décidons ainsi de regrouper ces réponses en une nouvelle modalité "Le manque d'envie" qui englobe 6 observations au total.

Nous faisons de même pour les questions concernant les caractéristiques des individus, à savoir, la formation suivie et le domaine d'étude. Pour la formation nous regroupons "Ecole d'ingénieur", "Ecole de commerce" et les organismes privés (mentionnés en 'Autre') dans une modalité nommée "Ecole" puisque cette variable ne doit concerner que le type de formation suivie et non le domaine étudié, nous voyons ici encore une limite à notre questionnaire. Aussi, pour la variable "domaine d'étude" il y a beaucoup de réponses personnalisées dont de nombreuses concernaient des branches d'étude au sein de l'Économie comme "marketing", "gestion", "comptabilité", "finance" etc. Étant donné que nous avons diffusé le questionnaire sur les réseaux (Facebook) mais surtout grâce à la mail-liste de IAE de Nantes, nous aurions pu anticiper le fait que la plupart des répondants étudient l'économie et donc que la simple modalité "Economie" qui regroupe plus de 70% des effectifs n'est pas assez exhaustive. Par conséquent nous créons 2 nouvelles modalités : 'Gestion' et 'Finance' où la première regroupe à la fois les domaines de 'RH', 'Management', 'Gestion' et 'Communication', et où la deuxième englobe 'Comptabilité' comme 'Finance' ou encore 'Banque'. Avec respectivement 37 et 9 observations dans chaque modalité créée, le classement que nous faisons des répondants est désormais plus précis, en revanche il y a un manque à gagner puisque ces modalités n'étant pas proposées dans le questionnaire, il est possible que certains répondant se soient référés à la catégorie 'Economie' alors que leur domaine aurait pu être celui des modalités créées à posteriori.

Notre enquête n'est donc pas aussi juste que l'on aurait souhaité, la limite la plus importante concerne la variable 'domaine d'étude' puisque nous n'avons pas bien anticipé le fait que la plupart des étudiants répondants étudient l'économie qui représentait initialement une seule modalité. Nous continuons néanmoins l'analyse en codant les variables qui sont

maintenant regroupées ; nous attribuons un chiffre compris entre 0 et n, où 'n' correspond au nombre de modalités prises par la variable concernée. Le dictionnaire des variables est ainsi disponible en [annexe n°1](#) ; avec à la fois le chiffre pris par chaque modalité, et le nom attribué aux variables avec la question correspondante. Enfin, nous supprimons la question ouverte portant sur la définition que chacun se fait de la Culture en général, qui nous servait uniquement pour l'introduction et nous factorisons les variables qualitatives. Nous avons ainsi 33 variables dans la base de données finale⁵ dont 32 sont qualitatives et 1 quantitative (l'âge).

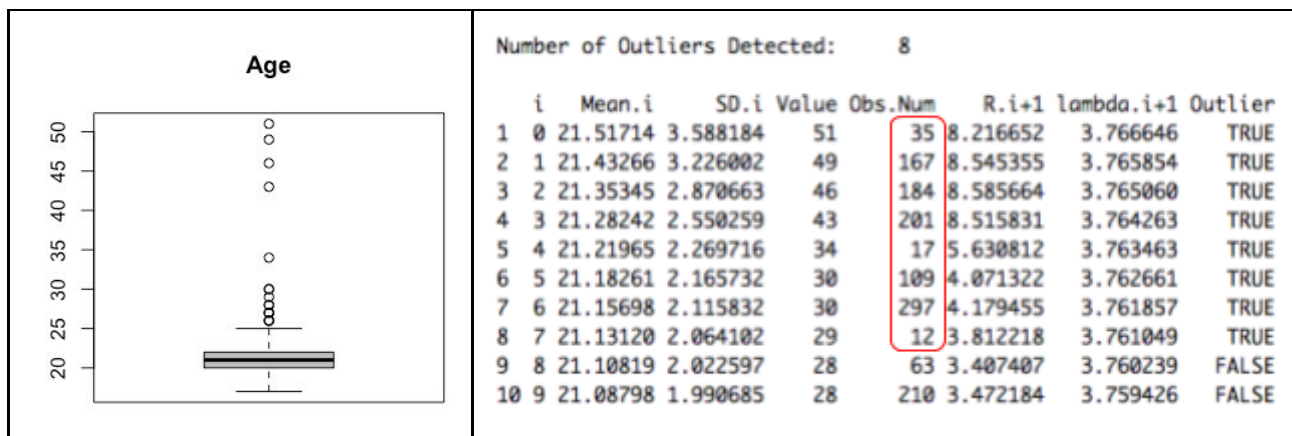
B) Détection des valeurs atypiques

Pour finir le nettoyage de notre base de données nécessaire à l'analyse, nous détectons les valeurs atypiques en 2 étapes : visuellement à l'aide des diagrammes en moustache puis statistiquement avec un test approprié au nombre de valeurs potentiellement atypiques détectées. Nous effectuons cette analyse pour notre seule variable quantitative qu'est l'âge (en années). À partir de la figure n°1 nous voyons 10 valeurs potentiellement atypiques pour lesquelles nous appliquons le test de Rosner dont l'hypothèse nulle est la suivante, H_0 : il n'y a pas de valeur atypique dans la série.

Les résultats du test de Rosner disponibles aussi en tableau n°2 nous indiquent que les répondants âgés de plus de 28 ans sont considérés comme atypiques, en ordonnant ces valeurs nous voyons qu'il s'agit des individus 12, 109, 297, 17, 201, 184, 167 et 35. En effet l'âge de ces personnes s'étendait de 29 à 51 ans, sachant que notre étude porte sur les étudiants nantais, il apparaît clair que ces individus soient considérés comme atypiques par leur âge élevé. Nous retirons donc ces 8 individus de la base et gardons une nouvelle base ("base1") composée de 342 observations et 33 variables, nous n'avons donc perdu que 2,3% de notre base initiale. Comme visible en [annexe n°2](#) sur les histogrammes de distribution, l'âge des répondants s'étend à présent de 17 à 18 ans où la plupart sont compris entre 17 et 22/23 ans.

⁵ Les réponses triées de l'enquête ainsi que la base recodée sont disponibles au lien suivant, en feuilles 1 et 2 : https://docs.google.com/spreadsheets/d/e/2PACX-1vT_9FgPFPyELZ0VXgW30I2IqEDca1RDY6snyM1foSUcaAAjYKHwZ1IBqE9MuQHwz4eGfRHvNONcOq2D/pubhtml

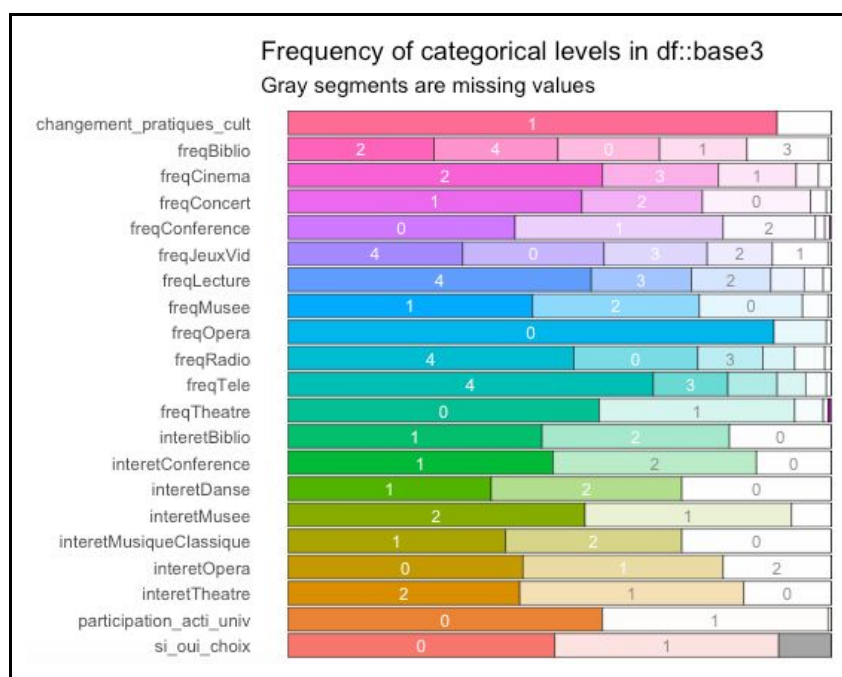
Tableau 2: Visuel et test de valeurs atypiques de la varige 'Age'



C) Statistiques descriptives

Dans notre enquête nous avons posé 34 questions pouvant être regroupées en 2 catégories. La première concerne les pratiques culturelles des étudiants, c'est à dire la fréquence et la diversité de ces dernières, quand la deuxième correspond aux caractéristiques propres à chaque individu, telles que genre, domaine d'étude, CSP des parents etc. Ainsi nous pouvons voir comment sont réparties les individus dans chaque modalité des variables, en regardant dans un premier temps leurs pratiques culturelles avec la base débarrassée des valeurs atypiques.

Tableau 3 : Répartition des pratiques culturelles des répondants dans chaque modalité



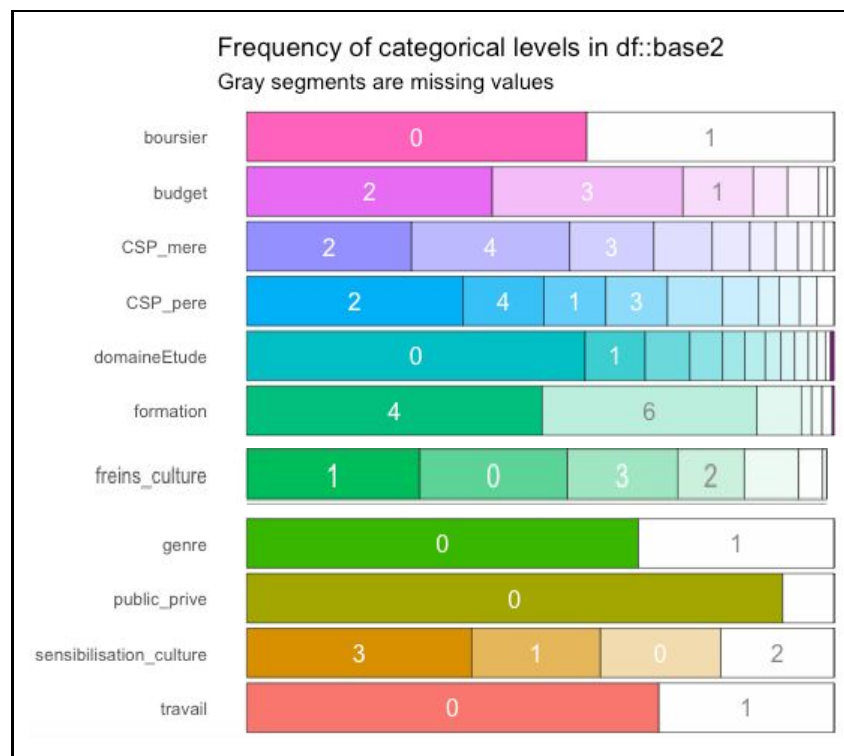
D'après le tableau n°3 et les statistiques des variables disponibles en [annexe n°3](#), on observe des répartitions différentes pour les variables se référant aux pratiques culturelles des individus. Par exemple on voit que 9 étudiants sur 10 ne sont pas satisfaits de leur consommation actuelle de culture, puisqu'ils souhaiteraient en majeure partie y consacrer plus de temps (55%) quand 45% aimerait diversifier les pratiques. Les fréquences de pratique des activités extérieures (théâtre, opéra, conférences...) comme intérieures (livre, radio, télévision...) diffèrent elles aussi grandement. Près de 90% des répondants ne vont jamais à l'opéra contre 57% pour le théâtre et 20% pour des concerts et festivals. L'activité pratiquée le plus par les étudiants de Nantes est le visionnage de télévision : 67,2% des étudiants regardent au moins une fois par semaine la télévision (modalité n°4 de la variable 'freqTele'), qui est suivie par la lecture en tout genre : 55,9% des étudiants lisent au moins une fois par semaine, et la radio : 52,6% des étudiants écoutent la radio au moins une fois par semaine.

En terme d'intérêt pour ces activités, on voit que les statistiques ne sont pas les mêmes que les pratiques réelles des étudiants. La majorité des répondants "aimerait beaucoup" (modalité n°3 des variables 'interet') aller au musée et au théâtre, avec respectivement 54,7% et 42,7% des étudiants. En revanche on retrouve une certaine cohérence pour l'activité 'Opéra' qui regroupe le moins de "J'aimerais beaucoup" (19,9%) et le plus de "Ça ne m'intéresse pas" (43,3%). La variable dont les réponses sont le plus équi-réparties est l'intérêt

pour la danse pour laquelle on observe 27,5% de “Ça ne m’intéresse pas”, “37,4%” de “Pourquoi pas” et 35,1% de “j’aimerais beaucoup”.

Ainsi, on observe un écart entre les pratiques effectives des étudiants et leurs attirances pour certaines activités : beaucoup ne vont jamais au musée ou au théâtre mais sont malgré tout attirés par ces activités. Ce décalage s’observe aussi à travers la variable ‘*changement_pratiques_cult*’ qui mesure la volonté de changer les pratiques culturelles actuelles, puisque l’on a vu que 90% des répondants ont répondu positivement à cette question. De la même manière, nous pouvons regarder les statistiques des variables qui servent à caractériser les individus à partir de la tableau n°4 et de [l’annexe n°4](#).

Tableau 4: Répartition des caractéristiques des répondants dans chaque modalité



Il est intéressant de noter d’après le tableau n°4 et l’[annexe n°4](#) que deux tiers des répondants sont des femmes, 70% ne travaillent pas en dehors du temps d’étude et 57,9% ne sont pas boursiers. On voit aussi que les modalités principales de la variable ‘formation’ sont la 4 et la 6 c’est à dire licence et master universitaires ce qui se retrouve dans les statistiques de

'*public_prive*' qui montrent que 91,2% des individus étudient dans un organisme public, on peut donc supposer une dépendance entre ces 2 variables qualitatives, dépendance que nous vérifierons dans la partie suivante. De plus, 80% des interrogés ont été sensibilisés à la culture, que ce soit par leurs parents ou par des cours extra-scolaires avec une répartition égale entre 'par les parents', 'par des cours extra-scolaires' et les deux. Enfin, on voit que les principaux freins aux pratiques culturelles sont selon les étudiants interrogés ; principalement le manque de temps (modalité n°1), mais aussi le coût des activités (modalité 0) et enfin, dans une moindre mesure le manque d'informations sur les événements (modalité n°3). Cette variable est très intéressante puisqu'elle permet de visualiser quels facteurs font que près de 90% des étudiants nantais ne vont pas à l'opéra par exemple.

À partir de ces informations nous pouvons choisir l'individu de référence, nous décidons de retenir un individu qui prend la modalité "0" pour chaque variable qualitative. Ainsi notre individu de référence est une femme étudiante en économie, ne travaillant pas pour financer ses études, n'ayant pas de bourse et un budget consacré aux activités culturelles nul. N'ayant pas été sensibilisée à la culture par ses parents agriculteurs ou par des cours extra-scolaires, elle ne pratique ni n'est attirée par aucune des activités culturelles. Cette dernière ne souhaite pas changer ses pratiques mais si elle avait le choix elle préférerait y consacrer plus de temps, enfin, elle n'a jamais participé à une activité proposée par son BTS.

D) Liens entre les variables

1- Détection des variables dépendantes

Afin de mieux cerner les pratiques culturelles des étudiants mais également l'adéquation entre les activités pratiquées et leurs envies, le questionnaire est divisé en 3 grandes parties. Une première plus empirique où nous nous renseignons sur la fréquence de chacune de celles-ci, une seconde sur les envies et affinité des étudiants à la culture, une dernière sur leurs caractéristiques socio-économiques. Ainsi certaines variables obtenues par ce biais peuvent-être corrélées et dépendantes les unes des autres. L'hypothèse d'indépendance est importante car elle peut entraîner des problèmes de multicolinéarité dans les différents modèles, sur lesquels nous reviendrons ultérieurement. Nos variables étant

essentiellement qualitatives, il convient de vérifier le test de Chi-2 qui suppose l'hypothèse nulle (H_0) : les deux variables sont indépendantes. Ainsi, lorsque H_0 est acceptée, nous pouvons affirmer que la variable n'a aucune influence sur l'autre. Après avoir testé bon nombre de variables qualitatives entre elles, nous obtenons les dépendances suivantes :

- fréquence opéra et fréquence radio
- fréquence opéra et fréquence théâtre
- toutes les variables de fréquences sont corrélées aux variables intérêts pour l'activité correspondant. Ainsi, fréquence musée est corrélée avec intérêt musée etc.
- sensibilisation à la culture et CSP de la mère
- domaine d'étude et CSP de la mère
- formation et public/privé (comme supposé précédemment)
- participation aux activités culturelles universitaires et établissement public/privé
- freins à la culture et budget
- changements de pratiques culturelles et fréquences musée

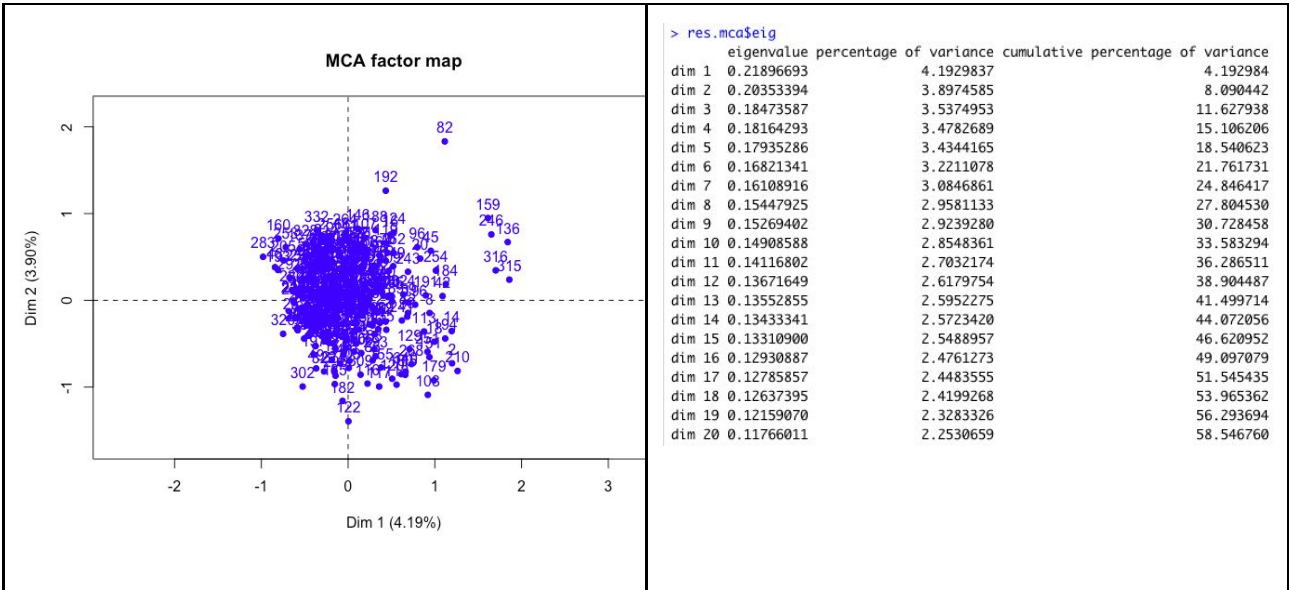
La dépendance entre les variables d'intérêts et les fréquences semble logique puisqu'un individu qui porte un intérêt à une activité sera plus enclin à la pratiquer. Dans la suite de notre étude nous décidons donc d'enlever les variables d'intérêt. De plus nous remarquons que la variable fréquence opéra est dépendante de celle de la radio et du théâtre. Ceci peut s'expliquer par le fait que l'opéra est une activité exigeante souvent pratiquée par des personnes dont les pratiques culturelles sont diverses. Nous faisons le choix de supprimer également cette variable. Nous supprimons également les variables CSP de la mère, établissement public/privé, freins à la culture, changements de pratiques culturelles.

2- Analyse des correspondances multiples

Pour mieux comprendre les liens entre nos individus et nos variables, nous procédons à une analyse des correspondances multiples (ACM), méthode factorielle d'identification des correspondances entre les variables mais aussi entre les individus. Ainsi un individu est considéré dans l'ensemble de ses modalités : deux individus se ressemblent s'ils choisissent

les mêmes modalités. En outre, l'ACM permet d'exprimer la variabilité des individus en fonction de leurs réponses liées aux fréquences de loisirs culturels.

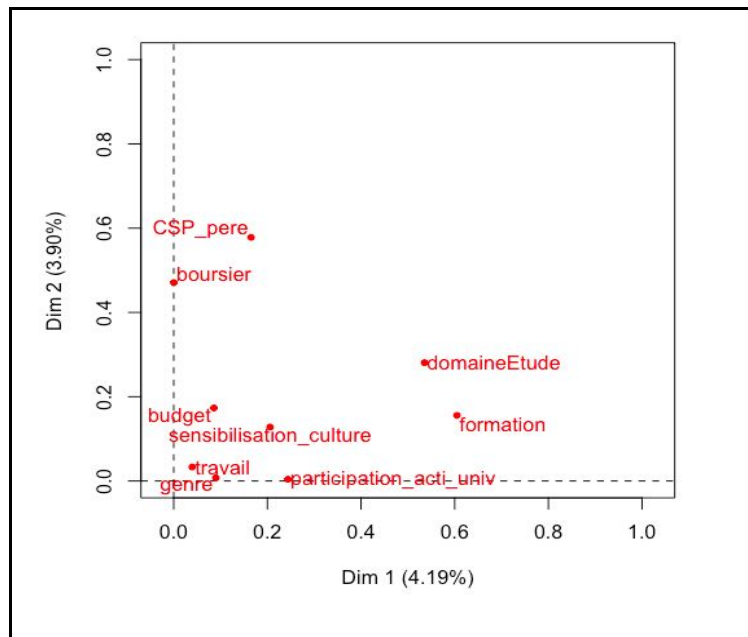
Tableau 5: Représentation graphique des individus et inertie associé aux axes



Nous pouvons voir sur le tableau n°5 qu'il n'y a pas de groupes d'individus bien distincts. Au contraire, on observe un nuage généralement concentré autour du centre et donc une faible distance entre les individus. De même, nous pouvons voir que les dix premiers axes de l'ACM représentent chacun entre 3 et 4% de l'inertie. On n'observe pas de décroissance importante entre deux facteurs. Cela peut être expliqué par une très faible diversité des profils d'activités culturelles parmi les interrogés.

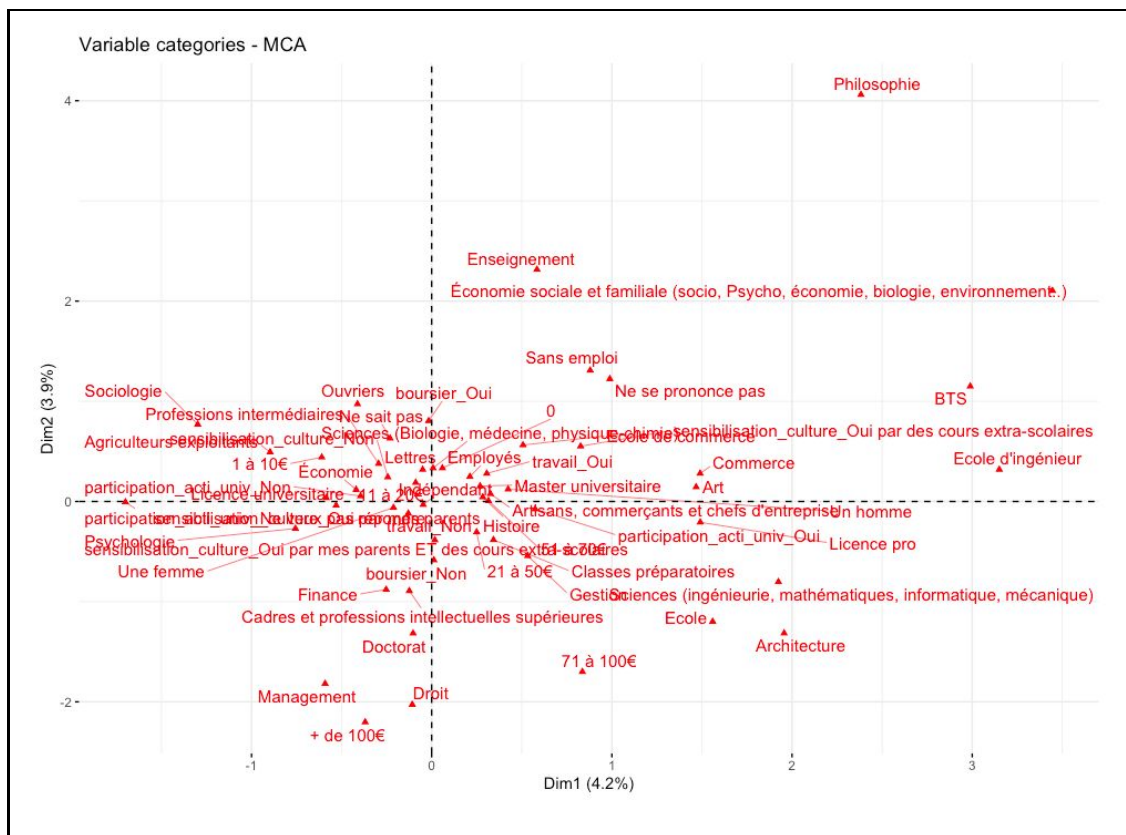
De plus, l'étude de l'ACM se fait en fonction des modalités. Ainsi, en fonction du nombre d'individus qui possèdent les mêmes caractéristiques *ie.* des modalités similaires, ces observations seront plus ou moins proches. En outre, des modalités rares peuvent influencer la détermination des axes et donc l'inertie des axes. Nous devons donc observer les graphiques avec attention car notre base est composée d'un nombre très restreint d'observations et ainsi nous avons des modalités avec des effectifs très différents les uns des autres.

Figure 1 : Graphique du carré des liaisons



D'après la figure n°1 on voit que les liaisons ne semblent à priori pas très forts entre les variables, puisque la plupart d'entre elles sont proches de l'origine. Nous pouvons relativiser ce constat puisque la base contient un nombre restreint d'individus, ainsi, les rapports de corrélation au carré peuvent-être significatifs. Également, les variables 'domaine d'étude' et 'formation' semblent proches de l'axe n°1, tandis que les variables 'CSP du père' et le statut 'boursier' semblent proches de l'axe n°2. Ainsi, l'axe 1 opposerait les caractéristiques académiques de l'individu, tandis que l'axe 2 représenterait les critères financiers.

Figure 2 : Représentation des variables dans le plan 1-2



Bien que l'inertie totale des facteurs 1-2 représente moins de 10% de l'inertie, la répartition des variables dans le plan factoriel permet de mieux comprendre comment les modalités sont liées entre elles. Sur l'axe 1 nous pouvons voir à gauche et à droites des modalités associées au domaine d'étude des étudiants. A gauche, davantage des domaines assimilées aux sciences sociales (Sociologie, Psychologie), à droite, des secteurs davantage scientifiques et professionnalisants tels que les écoles d'ingénieurs ou les sciences dites "dures". De plus, l'axe 2 semble opposer d'une part en haut les milieux plutôt modestes, où l'on retrouve les étudiants boursiers, les étudiants dans l'enseignement, les ouvriers, les sans emplois, etc. Tandis que l'on retrouve en bas des milieux plutôt aisés avec des budgets en dépenses culturelles supérieurs à 71 euros, des étudiants en droits, en management, des doctorants, la csp des cadres et professions supérieures. Ainsi, l'étude en ACM a permis d'opposer les étudiants modestes des étudiants aisés, et les milieux académiques plus littéraires aux sciences dures.

II- Modèle binaire

Le modèle binaire s'applique lorsque l'on analyse une variable qualitative dite '*binaire*' c'est à dire qui prendre 2 modalités : Oui et Non. La première condition pour effectuer une analyse binaire est d'avoir une répartition équivalente des observations dans les 2 modalités. Cette hypothèse forte nous a conduit à changer plusieurs fois la variable que nous cherchions à expliquer dans cette partie. En effet, nous voulions initialement prendre les réponses de la variable binaire issue de la question "*Voudriez-vous améliorer ou diversifier vos pratiques culturelles ?* ", mais nous sommes vite rendu compte que la plupart des étudiants répondent positivement à cette question. Il y a finalement près de 90% des répondants qui souhaitent changer leurs pratiques culturelles actuelles. Nous avons alors pensé à la variable "*Avez-vous été sensibilisé à la culture (arts plastiques, théâtre, musique etc.) ?*" initialement binaire, mais voyant que dès la diffusion d'un échantillon test de notre enquête, la grande majorité des répondants ont effectivement été sensibilisés, nous avons modifié les réponses proposées (Oui / Non) en les rendant plus précises :

- Oui par mes parents
- Oui par des cours extra-scolaires
- Oui par mes parents ET des cours extra-scolaires
- Non

Cette modification était nécessaire pour mieux cerner les raisons qui peuvent expliquer les différentes pratiques culturelles des étudiants. Comme nous l'avons vu à partir de la figure n°3 en partie précédente, seuls 20,5% des étudiants n'ont pas été sensibilisés à la culture, nous ne pouvions donc pas retenir cette variable pour l'étude d'un modèle binaire. Finalement, nous nous sommes intéressées aux fréquences de pratique des activités culturelles proposées en questions une et deux, il nous fallait trouver une variable dont la répartition était équivalente entre les premières et dernières modalités ("*Jamais*" et "*Ne sait pas*"), et les 4 autres (à savoir "*1 à 2 fois par an*", "*3 à 10 fois par an*", "*1 fois par mois*", "*Au moins une fois par semaine*").

La variable répondant à ces critères est celle de la fréquence de sortie au théâtre '*freqTheatre*' pour laquelle 57,6% des étudiants ne sont jamais allés, tandis que 42,4% y vont plus ou moins souvent, après recodage de la variable.

A) Hypothèses d'un modèle binaire

L'analyse d'une variable binaire peut s'effectuer de 2 manières : en utilisant un modèle **logit** ou un modèle **probit**. L'avantage de ces modèles par rapport à une régression linéaire, est qu'ils prennent en compte seulement 2 modalités de Y avec des probabilités comprises entre 0 et 1, ce qui n'est pas le cas d'un modèle linéaire qui est par conséquent non adapté à une telle étude. Dans cette partie nous décidons d'utiliser une méthode d'estimation "*logit*" qui suppose que les termes d'erreurs suivent une loi logistique.

Tout modèle binaire pour être valable doit vérifier certaines hypothèses :

- les variables explicatives ne sont pas corrélées entre elles : s'il y avait multicolinéarité cela signifierait que certaines variables mesurent le même phénomène.
- le modèle a un intérêt : au moins un des coefficients estimés par le modèle doit être significatif pour montrer que ce dernier permet d'expliquer en partie Y_t .
- le modèle est de bonne qualité : pour cela nous regarderons à la fois le R^2 de Mac Fadden, mais aussi les taux de spécificité, de prévision et de sensibilité.
- les observations ne sont pas influentes : la suppression d'une observation ne doit pas affecter la précision de l'estimation.
- les erreurs sont homoscédastiques : la variance du terme d'erreur doit être la même pour chaque observation, nous vérifierons cette hypothèse dans la partie C.

La vérification de ces hypothèses est fondamentale dans la validation du modèle et donc l'interprétation de ses coefficients de manière à pouvoir conclure sur le sujet étudié.

B) Estimation de modèles binaires sous hypothèse d'homoscédasticité des erreurs

La variable "*Theatre_binaire*" prend la modalité 1 si l'individu va au théâtre, 0 sinon. Dans cette partie nous allons estimer des modèles logit ayant pour but de comprendre les facteurs qui influencent le fait d'aller au théâtre ou non. Sachant que nous sommes dans le cas de modèles binaires, nous raisonnerons en terme de probabilité que l'événement se réalise, c'est à dire la probabilité de se rendre au théâtre. Cette variable est particulièrement intéressante, outre son équi-répartition entre les 2 modalités, puisqu'elle constitue une forme de culture "traditionnelle" mais qui ne semble pas inaccessible aux jeunes, comme peut l'être l'opéra par exemple. Elle semble ainsi refléter les différentes approches qu'ont les étudiants à la culture.

Comme nous avons pu le constater dans la partie sur les liens entre les variables, il existe une forte corrélation entre nos variables, ce qui nous oblige à devoir sélectionner seulement quelques unes d'entre elles, pour éviter les problèmes de multicolinéarité qui nous amèneraient à rejeter notre modèle. Il existe 3 méthodes fonctionnant par itérations en ajoutant ou retirant des variables à chaque estimation, de manière à trouver quel modèle est le plus pertinent en se basant sur 2 critères : la vraisemblance et la parcimonie. Ainsi les modèles sélectionnés par ces méthodes sont optimaux puisqu'ils maximisent leur qualité tout en minimisant leur complexité. Disponibles en [annexe n°5](#), les sorties des méthodes Backward, Forward et Both nous donnent approximativement les mêmes modèles. Les deux premières méthodes retiennent 5 variables dans l'explication de Y : la sensibilisation à la culture et les fréquences des concerts, musées, télévision et lecture. La méthode 'Both' quant à elle écarte la variable '*freqConcert*' des variables explicatives, mais étant donné que son critère d'Akaike est plus élevé que les 2 autres méthodes (403,6 contre 402,6) et que des 2 autres méthodes ressort un même modèle, nous décidons de garder le modèle composé des 5 variables : '*sensibilisation_culture*', '*freqLecture*', '*freqMusee*', '*freqTele*' et '*freqConcert*'.

Tableau 8: Modèle binaire sous hypothèse d'homoscédasticité des erreurs

```
> summary(logit1)

Call:
glm(formula = Theatre_binaire ~ freqConcert + freqMusee + freqTele +
    freqLecture + sensibilisation_culture, family = binomial(link = "logit"),
    data = base_binaire)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.5316  -0.8559  -0.2959   0.9680   2.1130

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -18.3464    1021.3695  -0.018  0.98567
freqConcert1     0.5056     0.3502   1.444  0.14886
freqConcert2     0.9526     0.4064   2.344  0.01907 *
freqConcert3     1.9274     1.0055   1.917  0.05524 .
freqConcert4    -1.0759     1.8215  -0.591  0.55474
freqMusee1       0.3261     0.3883   0.840  0.40103
freqMusee2       1.1437     0.4198   2.725  0.00644 **
freqMusee3       3.4255     1.1660   2.938  0.00331 **
freqMusee4      18.2502    2797.4420   0.007  0.99479
freqTele1       -0.1756     0.7551  -0.233  0.81614
freqTele2      -1.8120     0.7702  -2.353  0.01864 *
freqTele3       -0.1366     0.5684  -0.240  0.81013
freqTele4       -0.8448     0.4623  -1.827  0.06763 .
freqTele5       -0.1567     1.4691  -0.107  0.91507
freqLecture1     13.9209    1021.3700   0.014  0.98913
freqLecture2     16.7523    1021.3695   0.016  0.98691
freqLecture3     16.8317    1021.3694   0.016  0.98685
freqLecture4     16.6971    1021.3694   0.016  0.98696
freqLecture5     15.4889    1021.3701   0.015  0.98790
sensibilisation_culture1  1.2240     0.4355   2.810  0.00495 **
sensibilisation_culture2  0.7918     0.4527   1.749  0.08026 .
sensibilisation_culture3  1.3046     0.4048   3.223  0.00127 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 466.18  on 341  degrees of freedom
Residual deviance: 358.60  on 320  degrees of freedom
AIC: 402.6
```

Vérifions à présent si le modèle respecte les hypothèses énumérées en partie précédente.

1- Existe-t-il de la multicollinéarité entre les variables ?

Le premier critère consiste à vérifier l'existence ou non d'un problème de multicollinéarité entre les variables explicatives. Pour cela nous utilisons le critère du "Variance Inflation Factor". Si le VIF est supérieur à 10 alors il y a un problème de multicollinéarité. Ce dernier, s'il est élevé, implique de mauvaises estimations des coefficients car leurs signes peuvent être faussés, c'est pourquoi il est très important de vérifier sa valeur.

Tableau 9 : Résultat test VIF modèle binaire

<code>> vif(logit1)</code>			
	GVIF	Df	GVIF ^{1/(2*Df)}
freqConcert	1.317112	4	1.035030
freqMusee	1.298393	4	1.033179
freqTele	1.308163	5	1.027226
freqLecture	1.294104	5	1.026117
sensibilisation_culture	1.202881	3	1.031265

Le VIF de chacune des variables est inférieur à 10 et même à 5, il n'y a donc aucun problème de multicollinéarité, nous pouvons alors passer à la vérification de la deuxième hypothèse qu'est l'intérêt du modèle.

2- Le modèle est-il intéressant ?

La vérification de cette hypothèse s'effectue en regardant le log de vraisemblance. Ce dernier mesure la nullité des coefficients estimés et voit ainsi si au moins un des coefficients est utile dans l'explication de Y, donc si le modèle a un intérêt. Ce ratio est basé sur les mêmes hypothèses que le test de Fisher à savoir :

(1) H_0 : Tous les coefficients sont nuls

H_1 : Au moins un des coefficients est significatif

(2) Log de vraisemblance = $-2(\ln L_r - \ln L) \sim \chi_k^2$

(3) La p-value de ce test étant égale à $1,289 e^{-13}$ nous rejetons H_0 ; au moins un des coefficients estimés par le modèle est significatif, nous continuons donc notre analyse avec ce modèle.

3- La qualité du modèle est-elle bonne ?

Nous venons de vérifier que les variables ont bien un effet sur notre Y, il convient à présent de quantifier cet effet, pour cela regardons la qualité d'ajustement du modèle donnée par le R^2 de Mac Fadden. Cet indicateur, compris entre 0 et 1, indique le pourcentage de Y expliqué par les variables explicatives, ainsi plus il est proche de 1 (ou 100%) meilleure est la qualité du modèle.

Pour le modèle logit composé de 5 variables (4 fréquences et la sensibilisation à la culture), le coefficient de Mac Fadden s'élève à 0,231. Pour un modèle binaire ce chiffre est plutôt satisfaisant. Regardons à présent les taux énumérés en partie précédente ayant pour but de compléter l'analyse de qualité et de pertinence du modèle, grâce à la fonction hitmiss sous R studio.

Tableau 10: Taux de précision, de spécificité et de sensibilité du modèle logit

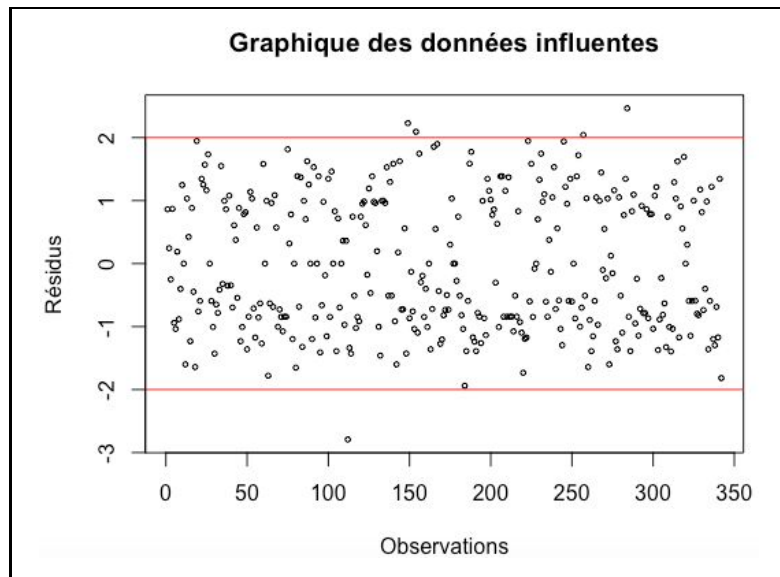
```
> hitmiss(logit1)
Classification Threshold = 0.5
      y=0 y=1
yhat=0 157  52
yhat=1  40  93
Percent Correctly Predicted = 73.1%
Percent Correctly Predicted = 79.7%, for y = 0
Percent Correctly Predicted = 64.14% for y = 1
Null Model Correctly Predicts 57.6%
[1] 73.09942 79.69543 64.13793
```

D'après le tableau n°10 on voit que la qualité de prévision du modèle est bonne : son taux d'erreur est de seulement 26,9% (il se trompe un peu plus d'une fois sur 4) et la qualité de prévision des modalités 0 et 1 du fait d'aller au théâtre sont respectivement 79,7% et 64,14% ce qui est satisfaisant. De plus, on remarque que le modèle prédit mieux le fait de ne pas aller au théâtre par rapport au fait d'y aller, cela peut s'expliquer par la répartition des observations dans chacune des modalités. Ayant 57,6% des étudiants dans la modalité 0, c'est à dire 'ne jamais aller au théâtre' contre 42,4% dans l'autre modalité, il est normal que la qualité de prévision soit meilleure pour la modalité où il y a le plus d'informations.

4- Les observations sont-elles influentes ?

Nous pouvons terminer cette partie en nous intéressant aux résidus des observations : pour que les données soient considérées comme non influentes il faut que ses résidus soient bondés entre -2 et 2. Effectivement, s'ils dépassent ces bornes alors cela signifie que certains individus influencent significativement l'estimation alors que dans un bon modèle, chaque observation est censée apporter la même information. Nous fixons la limite à 2,05, ainsi, tout résidu ne se situant pas dans l'intervalle $[-2,05; 2,05]$ sera considéré comme influent et sera donc retiré de la base (du moins l'observation correspondante).

Figure 3 : Graphique des résidus



D'après la figure n°3 il semble y avoir 4 données influentes puisque 4 points correspondant aux résidus sont hors de l'intervalle fixé graphiquement à 2. Par la fonction "sort" nous voyons qu'il s'agit des observations 112, 154, 149 et 284 que nous retirons de la base pour estimer par la suite un nouveau modèle logit. La base a maintenant 338 observations contre 350 au début, nous avons d'abord enlever 8 valeurs atypiques puis 4 données influentes. Le nouveau modèle "logit2" ainsi que ces hypothèses validées se trouve en [annexe n°6](#), on peut voir que la qualité d'ajustement est meilleure, que le critère d'Akaike est plus faible et que la qualité de prévision est plus proche d'une modalité à une autre de Y.

C) Vérification de l'homoscédasticité des erreurs

Nous avons estimé un modèle logit que nous avons corrigé de manière à ce qu'il valide toutes les hypothèses, il en reste encore une à valider : les erreurs sont homoscédastiques. Pour cela nous allons procéder par étape : nous estimerons premièrement un modèle qui suppose l'hétéroscédasticité des erreurs, en essayant chacune des 5 variables du modèle comme coupable potentiel de cette hétéroscédasticité et en ne retenant que les variables qui sont significatives au seuil de 5%. Finalement nous comparerons un modèle où seule la constante est coupable à un modèle où les variables significatives sont coupables. Cette

comparaison nous permettra de savoir quel modèle garder, et par conséquent, quel modèle interpréter.

Tableau 12: Modèle logit avec erreurs homoscedastiques et toutes les variables coupables

```
Call:
hetglm(formula = Theatre_binaire ~ freqConcert + freqMusee + freqTele + freqLecture + sensibilisation_culture | freqConcert +
  freqMusee + freqTele + freqLecture + sensibilisation_culture, data = base_binaire1, family = binomial(link = "logit"),
  control = hetglm.control(method = "BFGS", maxit = c(500, 25), hessian = FALSE, trace = FALSE), model = TRUE)
```

Deviance residuals:

	Min	1Q	Median	3Q	Max
	-1.7866	-1.0259	0.0000	1.1695	2.0772

Coefficients (binomial model with logit link):

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.363e+01	1.071e+02	-0.221	0.825
freqConcert1	1.933e+00	3.551e+01	0.054	0.957
freqConcert2	4.736e+00	7.753e+01	0.061	0.951
freqConcert3	3.785e+00	6.493e+01	0.058	0.954
freqConcert4	-1.537e+01	2.922e+04	-0.001	1.000
freqMusee1	9.886e-01	1.631e+01	0.061	0.952
freqMusee2	4.602e+00	7.455e+01	0.062	0.951
freqMusee3	3.199e+01	8.609e+02	0.037	0.970
freqMusee4	1.992e+01	5.329e+22	0.000	1.000
freqTele1	-1.420e+00	2.295e+01	-0.062	0.951
freqTele2	-2.213e+01	3.674e+02	-0.060	0.952
freqTele3	-1.895e+00	3.130e+01	-0.061	0.952
freqTele4	-2.807e+00	4.620e+01	-0.061	0.952
freqTele5	3.196e+00	1.732e+02	0.018	0.985
freqLecture1	-6.862e-01	4.764e+02	-0.001	0.999
freqLecture2	1.936e+01	8.134e+01	0.238	0.812
freqLecture3	1.995e+01	8.398e+01	0.238	0.812
freqLecture4	1.852e+01	7.835e+01	0.236	0.813
freqLecture5	-1.420e+00	8.923e+03	0.000	1.000
sensibilisation_culture1	2.178e+00	3.584e+01	0.061	0.952
sensibilisation_culture2	3.417e-01	6.133e+00	0.056	0.956
sensibilisation_culture3	1.942e+00	3.171e+01	0.061	0.951

Latent scale model coefficients (with log link):

	Estimate	Std. Error	z value	Pr(> z)
freqConcert1	-2.959e-01	3.580e+00	-0.083	0.93412
freqConcert2	-6.437e-00	4.432e+00	-1.452	0.14641
freqConcert3	-1.779e+01	4.474e+00	-3.976	7.01e-05 ***
freqConcert4	-7.699e-00	2.935e+03	-0.003	0.99791
freqMusee1	2.813e-01	6.282e-01	0.448	0.65428
freqMusee2	4.523e+00	2.662e+00	1.699	0.08929 .
freqMusee3	-8.728e-00	2.389e+01	-0.365	0.71488
freqMusee4	-7.998e-00	2.834e+00	-2.822	0.00477 **
freqTele1	-2.326e+01	1.078e+01	-2.158	0.03093 *
freqTele2	-4.511e+00	2.834e+00	-1.592	0.11146
freqTele3	-2.804e+00	2.821e+00	-0.994	0.32033
freqTele4	-3.134e+00	2.807e+01	-0.112	0.91112
freqTele5	-1.744e+01	1.537e+01	-1.135	0.25658
freqLecture1	-7.848e+00	1.524e+01	-0.515	0.60665
freqLecture2	5.214e+00	1.523e+01	0.342	0.73209
freqLecture3	5.316e+00	4.469e+02	0.012	0.99051
freqLecture4	4.181e+00	9.485e-01	4.408	1.04e-05 ***
freqLecture5	-8.075e-00	6.993e-01	-11.548	< 2e-16 ***
sensibilisation_culture1	1.033e+00	1.766e+00	0.585	0.55856
sensibilisation_culture2	-2.044e-01	3.358e+21	0.000	1.00000
sensibilisation_culture3	5.262e+00	2.164e+01	0.243	0.80789

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Lag-likelihood: -147.7 on 43 Df
 LR test for homoskedasticity: 37.03 on 21 Df, p-value: 0.01669
 Dispersion: 1
 Number of iterations in BFGS optimization: 126

D'après le tableau n°12 on constate que la variable '*sensibilisation_culture* n'est pas significative dans l'explication de l'hétéroscédasticité des erreurs, nous retirons donc cette variable et réestimons un modèle supposant l'hétéroscédasticité des erreurs. Les modèles suivants sont disponibles en [annexe n°7](#) ; après avoir retiré la variable '*sensibilisation_culture*' nous retirons la variable '*freqLecture*' qui est devenue non significative et réestimons le

modèle. Disponible aussi en [annexe n°7](#), nous voyons que désormais les 3 variables '*freqConcert*', '*freqMusee*' et '*freqTele*' sont significatives dans l'explication de l'hétéroscédasticité des erreurs. Nous notons aussi que la p-value du test de vraisemblance, dont l'hypothèse nulle est qu'il n'y a pas d'amélioration significative du log de vraisemblance entre le modèle supposant l'hétéroscédasticité des erreurs et celui supposant l'homoscédasticité des erreurs, s'élève à 0,0484. Cela signifie qu'au seuil d'erreur de 5% nous acceptons qu'il y ait une différence significative entre les 2 modèles, il faudrait donc retenir le modèle hétéroscédastique.

Cependant on voit que le VIF de la variable '*freqConcert*' est de 9,099, il est donc nécessaire d'enlever la variable et d'estimer à nouveau le modèle. Disponible en [annexe n°8](#), on voit que désormais c'est la variable '*freqMusee*' qui n'est pas significative en tant que variable explicative de l'hétéroscédasticité des erreurs. Encore une fois nous réestimons un modèle par la fonction *hetglm* en supprimant cette variable, il ne reste alors plus que la variable '*freqTele*' comme coupable. Aussi disponible en [annexe n°8](#), on constate que la dernière variable coupable n'est finalement plus significative ; nous la retirons des variables coupables, il n'y a plus aucun coupable - nous revenons donc dans le cas d'un modèle homoscédastique. Cela aurait pu se deviner car à plusieurs reprises, à force d'estimer des modèles sous hypothèse d'hétéroscédasticité des erreurs, nous voyions que la p-value associée au LR test était soit très proche de 0,05 comme c'était le cas pour le troisième modèle (avec une p-value=0,04842), soit supérieure au seuil de risque de 5% ce qui nous aurait amené dans ce cas à accepter l'hypothèse nulle qu'il n'y a pas de différence significative entre les 2 modèles.

Ainsi, la corrélation trop forte entre les variables ou la non significativité de certaines d'entre elles en tant que coupable de l'hétéroscédasticité, nous ont conduit petit à petit à retirer chaque variable du modèle pour finalement aboutir à un modèle homoscédastique c'est à dire sans coupable.

Nous pouvons maintenant interpréter le modèle logit estimé en partie précédente validant toutes les hypothèses, comme modèle final expliquant les déterminants du fait d'aller au théâtre ou non. Voici de nouveau en tableau n°5 l'estimation de ce modèle ainsi que les coefficients "odd ratios" dont nous allons nous servir pour interpréter les résultats de ce dernier. Il faut cependant prêter attention à ceux-ci ; si le coefficient estimé par le modèle est négatif la simple transformation en exponentiel ne suffit pas : il faut faire $\frac{1}{e^x}$.

Tableau 13: Modèle logit final : avec erreurs homoscédastiques, et odd ratios

<pre>> summary(logit2)</pre>				
Call: glm(formula = Theatre_binaire ~ freqConcert + freqMusee + freqTele + freqLecture + sensibilisation_culture, family = binomial(link = "logit"), data = base_binaire1)				
Deviance Residuals:				
Min	1Q	Median	3Q	Max
-1.92920	-0.84963	-0.00017	0.91910	2.07702
Coefficients:				
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-19.4262	1663.9719	-0.012	0.990685
freqConcert1	0.4687	0.3586	1.307	0.191251
freqConcert2	0.9367	0.4170	2.247	0.024665 *
freqConcert3	1.2088	1.0222	1.183	0.237000
freqConcert4	-15.0105	1326.0316	-0.011	0.990968
freqMusee1	0.5455	0.4117	1.325	0.185154
freqMusee2	1.3470	0.4402	3.060	0.002212 **
freqMusee3	31.6519	1888.8943	0.017	0.986631
freqMusee4	19.4355	4612.2020	0.004	0.996638
freqTele1	-0.3115	0.7754	-0.402	0.687863
freqTele2	-1.9554	0.7904	-2.474	0.013367 *
freqTele3	-0.2746	0.5933	-0.463	0.643557
freqTele4	-1.0208	0.4816	-2.120	0.034040 *
freqTele5	0.1846	1.6864	0.109	0.912849
freqLecture1	-0.2831	2109.9692	0.000	0.999893
freqLecture2	17.7419	1663.9719	0.011	0.991493
freqLecture3	17.8147	1663.9719	0.011	0.991458
freqLecture4	17.6667	1663.9718	0.011	0.991529
freqLecture5	-0.9777	3504.3436	0.000	0.999777
sensibilisation_culture1	1.3703	0.4554	3.009	0.002619 **
sensibilisation_culture2	0.8013	0.4739	1.691	0.090905 .
sensibilisation_culture3	1.4422	0.4221	3.417	0.000634 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
<pre>> exp(coef(logit1))</pre>				
(Intercept)	freqConcert1	freqConcert2	freqConcert3	freqConcert4
1.077064e-08	1.657906e+00	2.592552e+00	6.871792e+00	3.409843e-01
freqMusee1	freqMusee2	freqMusee3	freqMusee4	freqTele1
1.385502e+00	3.138253e+00	3.073654e+01	8.432201e+07	8.389780e-01
freqTele2	freqTele3	freqTele4	freqTele5	freqLecture1
1.633221e-01	8.723474e-01	4.296267e-01	8.549786e-01	1.111117e+06
freqLecture2	freqLecture3	freqLecture4	freqLecture5	sensibilisation_culture1
1.885516e+07	2.041339e+07	1.784230e+07	5.330058e+06	3.400718e+00
sensibilisation_culture2	sensibilisation_culture3			
2.207469e+00	3.686061e+00			

D'après le tableau n°13 on constate qu'au moins une modalité de chaque variable est significative au seuil de risque de 5%, excepté pour la variable de fréquence de lecture. De plus, pour les 3 autres variables concernant les fréquences on remarque que c'est toujours la modalité n°2 qui est significative, celle-ci correspond à '3 à 10 fois par an', il apparaît donc logique qu'elle soit la plus représentative du fait d'aller au théâtre, car c'est la fréquence la plus "sobre" entre ne jamais pratiquer telle activité (ou la pratiquer 1 à 2 fois par an, ce qui

reste très peu), ou au contraire la pratiquer chaque semaine. Le fait d'aller à des concerts ou des musées 3 à 10 fois par an augmente la probabilité de se rendre au théâtre, par rapport à une personne qui ne pratique jamais ces activités. À contrario, le fait de regarder la télévision 3 à 10 fois par an diminue la probabilité d'aller au théâtre (c'est à dire entrer dans la modalité 1 de la variable '*Theatre_binaire*') par rapport à la personne de référence à savoir, qui ne regarde jamais la télévision. On voit qu'il en est de même pour la modalité n°4 de la variable '*freqTele*' qui correspond à un visionnage au moins une fois par semaine, il apparaît donc évident que la relation entre ces 2 variables soit négative. Aussi, le fait d'avoir été sensibilisé à la culture par ses parents ou par ses parents ET des cours extra-scolaires augmente la probabilité de se rendre au théâtre par rapport à une personne qui n'aurait pas été sensibilisée à la culture. Il est intéressant de rappeler que le coefficient d'ajustement de ce modèle s'élève à 27,7% ce qui reste correct sachant qu'il est composé de seulement 5 des 33 variables initialement disponibles après récolte des réponses de notre enquête.

Il est intéressant de quantifier l'effet de chacune des variables énumérées ci-dessus en calculant les odd ratios. Ainsi on voit que les étudiants allant à des concerts ou festivals 3 à 10 fois par an ont environ 2,59 fois plus de chance de se rendre au théâtre qu'un étudiant n'allant jamais à de telles activités. De même, un étudiant allant au musée 3 à 10 fois par an a 3,14 fois plus de chance de sortir au théâtre par rapport à un étudiant qui ne va jamais au musée. Aussi, deux étudiants regardant la télévision 3 à 10 fois par an et au moins une fois par semaine ont respectivement $\frac{1}{0,1633} = 6,12$ et $\frac{1}{0,4296} = 2,33$ fois moins de chance de se rendre au théâtre que deux étudiants ne regardant jamais la télévision. Enfin, deux étudiants l'un sensibilisé à la culture seulement par ses parents, l'autre par ses parents et des cours extra-scolaires ont respectivement 3,4 et 3,69 fois plus de chance d'effectuer une sortie culturelle au théâtre que des étudiants qui n'ont été sensibilisés à aucune forme de culture.

Ces résultats semblent cohérents quant à la théorie sociologique sur le sujet ; comme nous l'avons vu dans la partie économique, le capital culturel apporté par les parents (que ce soit directement par eux ou par des cours extra-scolaires offerts à leurs enfants) détermine en bonne partie les pratiques culturelles aujourd'hui. De même, l'habitus mesuré par les fréquences de sorties aux différentes activités culturelles, influence lui aussi de manière significative le fait d'aller au théâtre pour les 350 étudiants nantais ayant répondu au questionnaire.

III- Modèle multinomial ordonné

Dans cette partie, nous tenterons de modéliser la fréquence des pratiques culturelles par un modèle multinomial ordonné. En effet, le nombre de modalités sera ici supérieur à 2, puisque nous tâcherons de créer des classes selon que les étudiants aient une pratique culturelle très fréquente, modérée, peu fréquente, ou inexistante. Ainsi les classes sont ordonnées selon la fréquence. En outre, la fréquence de sorties culturelles une fois par semaine est supérieure à une pratique une fois par an. Nous avons donc réalisé des groupes afin de faciliter la modélisation.

Ces classes ont été obtenues en associant à chaque fréquence un score plus ou moins élevé. Ainsi pour chaque activité pratiquée une fois par semaine on associe le score de 5, 1 fois par mois : 4, 3 à 10 fois par an : 3, 1 à 2 fois par an : 2, jamais : 1. En outre, nous avons additionné chaque fréquence à chacune des activités. Les modalités "*jamais*" et "*Ne sait pas*" ont également été regroupées.

Exemple : Laure ne va jamais à l'opéra, cependant elle va une fois par mois au théâtre, une fois par semaine en concert, 1 à 2 fois par an à des conférences, jamais au musée, ni au cinéma ou à la bibliothèque. Elle aura donc un score de : $0 + 4 + 5 + 2 + 0 + 0 + 0 = 11$. Ainsi elle sera catégorisée dans une des classes.

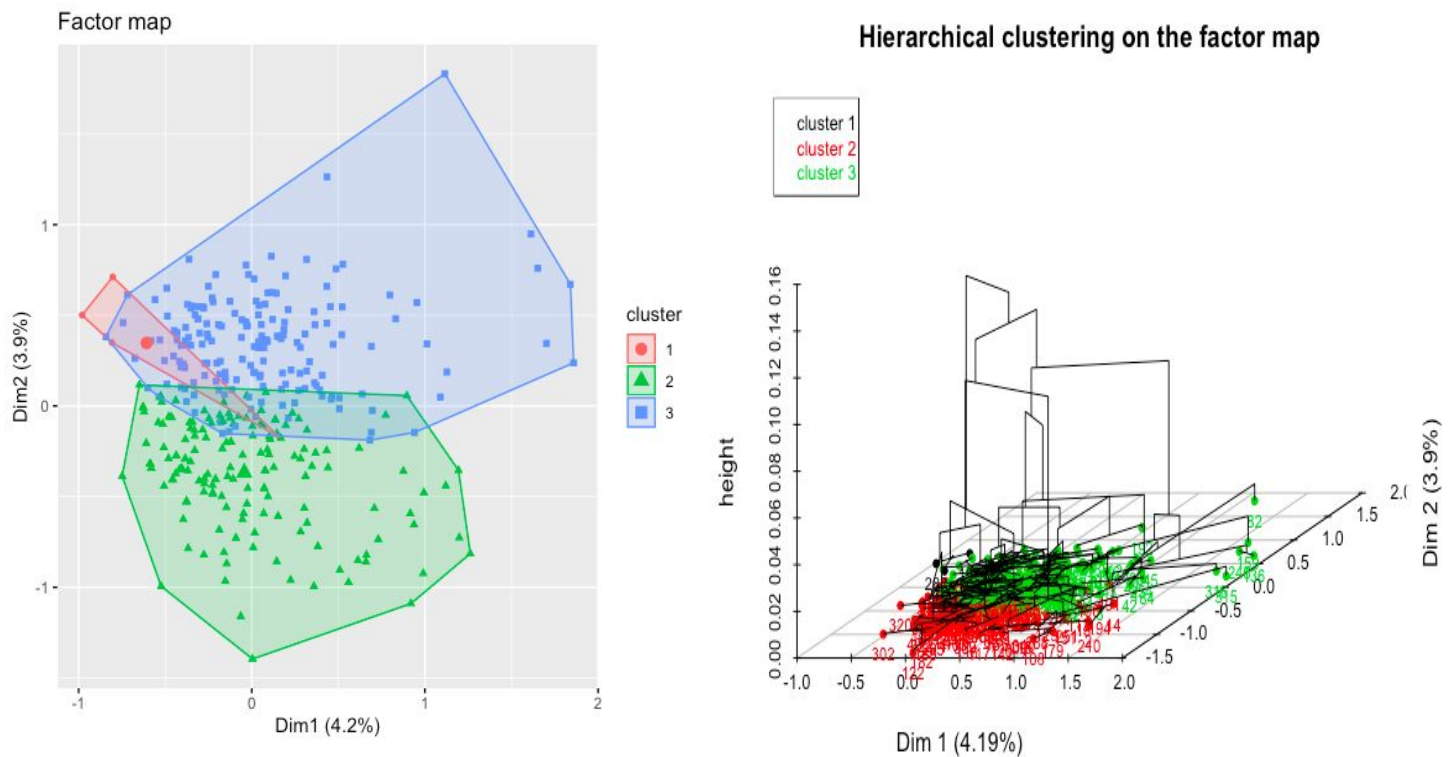
L'objectif de notre étude est de déterminer les variables quantitatives ou qualitatives qui jouent sur la fréquence de consommation culturelle. La fréquence culturelle se décline donc sous 4 modalités ordonnées, **identiques et mutuellement exclusives**. Ainsi, lorsque les termes d'erreur suivent une loi logistique, le modèle s'apparentera à un modèle logit multinomial ordonné tandis qu'elle s'apparentera à un modèle probit multinomial ordonné lorsque les erreurs suivent une loi normale centrée réduite. Cependant, ces modèles reposent sur les hypothèses essentielles d'égalité des pentes et d'hétéroscedasticité des erreurs, qu'il conviendra de vérifier ultérieurement.

1. Classes et significativité des seuils

Nous souhaitons donc trouver le nombre de classes et les seuils appropriés à l'analyse. La méthode des clusters et notamment de la méthode ward d2, consiste à minimiser la

variance intra-classe. Ainsi, la fonction Nbclust nous indique le nombre de clusters qui minimise cette variance. Le graphique ci-dessous montre que le dendrogramme construit à partir de l'ACM propose une répartition en trois dimensions.

Figure 4 : Dendrogramme et clusters de l'ACM



Cependant, comme nous pouvons le voir sur le plan à droite, les individus semblent ne pas clairement se distinguer puisque les groupes se superposent. La répartition faite par ward d2 ne semble pas très pertinente pour notre étude. Ainsi, nous décidons de créer nos classes manuellement. Cette décision a été confirmée par l'estimation du modèle global avec Y répartie par les clusters proposé par le dendrogramme. En outre, avec cette répartition, aucune des variables explicatives n'était significative ([annexe 12](#)).

La première approche que nous tentons consiste à faire des classes proportionnelles, même si nous savons que la 1ere et la dernière classes seront celles contenant le moins d'observations, nous ajusterons ensuite si les effectifs de celles-ci sont trop différents. A l'aide de la fonction Polr du package "forecats", nous pouvons voir que le découpage en catégorie

1/2/3/4 avec les seuils n'est pas pertinent puisque la première classe n'est pas significative au seuil de 5%. En outre, cela peut s'expliquer par le fait que la classe 1 représente 21 individus de l'effectif, la deuxième 94, la troisième 188 et la quatrième 39. Ainsi nous testons d'autres seuils plus pertinents de manière à ce que les effectifs soient plus équilibrés entre classes telle que chaque classe contient autour de 85 observations.

Tableau 15: Répartition des fréquences pour le logit ordonné

1. Absence ou très faible fréquence des pratiques culturelles	Score de 13 à 24
2. Faible pratiques culturelles	25 à 27
3. Fréquence de pratique modérée	28 à 30
4. Fréquence des pratiques culturelles conséquente	31 à 39

Cette fois-ci, nous obtenons des classes significatives au seuil de 10%. Avec des p-values, respectivement égales à 0,0006, 0, et 0, pour nos quatre classes. Ce sont celles-ci que nous utiliserons pour la suite de l'analyse.

2. Première modélisation

a. Estimation

Afin de pouvoir entreprendre sereinement notre modèle, il convient d'abord de vérifier l'hypothèse d'égalité des pentes. Le premier modèle avec toutes les variables explicatives a été estimé à l'aide de la fonction `vglm`. Nous obtenons ainsi les résultats suivants :

Tableau 16: Modèle 1 global

Variables explicatives		Modèle multinomial ordonné	Ecart-type
Constante	1	0.00388 **	1.09717
	2	8.92e-05 ***	1.10906
	3	6.03e-07 ***	1.12190
Participation activités université	0=Non		
	2=Oui	0.20676	0.22365
	3=Ne se prononce pas	0.82217	1.29361
Sensibilisation culture	0 =Non	Référence	
	1= Oui par mes parents	5.71e-05 ***	0.33217
	2=Oui par des cours extra-scolaire	0.08736 .	0.34854
	3=Oui pour les deux	5.47e-05 ***	0.30198
Budget	0= 0 euros	Référence	
	1= [1 ;10]	0.00520 **	0.61953
	2= [11 ;20]	0.00029 ***	0.56346
	3= [21 ;50]	4.34e-05 ***	0.57135
	4= [51 ;70]	6.74e-06 ***	0.71259
	5= [71 ;100]	0.10652	0.99998
	6= > 100 euros	0.13448	1.08615
Formation	0= BTS	Référence	
	1= CAP/BEP		
	2= Classes préparatoires		
	3= École	0.47445	0.91175
	4= Licence universitaire	0.05264 .	0.85622
	5= Licence professionnelle	0.00350 **	1.25052
	6= Master universitaire	0.03880 *	0.85690
	7= Doctorat	0.03160 *	1.22414

Domaine d'étude	0=Économie	Référence	
	1=Gestion	0.53343	0.37469
	2=Finance	0.64467	0.69351
	3=Sociologie	0.94690	0.97546
	4=Droit	0.54759	0.91002
	5=Histoire	0.34991	0.58054
	6=Lettres	0.19821	0.54729
	7=Psychologie	0.82345	1.09378
	8=Philosophie	0.50411	1.85964
	9=Commerce	0.26135	0.73408
	10=Architecture	NA	925.77937
	11=Sciences (ingénierie, mathématiques, informatique, mécanique)	0.21293	0.52137
	12=Sciences (Biologie, médecine, physique-chimie)	0.92311	0.41295
	13=Art	0.19580	
	0 = Femme		0.73179
Genre	1= Homme	Référence	
		0.15607	0.23308
Age	(en années)	0.20058	0.06800

Nous pouvons tirer dès lors, les premières conclusions de ce modèle. Tout d'abord, le fait d'avoir eu au préalable une sensibilisation à la culture par les parents et/ou par des cours extra-scolaire a un impact significatif sur la fréquence de pratique culturelle. De plus, un budget entre 1 et 70 euros impacte significativement la consommation culturelle, ce qui n'est pas le cas au-delà de 70 euros. Cela peut s'expliquer par le fait que les étudiants ayant un budget aussi conséquent destiné à la culture (relativement aux autres), n'ont pas d'arbitrage à faire quant au choix des activités. Au niveau de la formation, le fait d'être en classe préparatoire, en licence pro/universitaire, en master ou doctorat a également un impact significatif sur la fréquence des pratiques culturelles. Pour finir, le fait de faire des études d'architecture influence également la fréquence. Aussi, nous savons qu'un individu aura une probabilité de consommation culturelle d'autant plus élevée que le coefficient de la variable est faible. De plus, tous les coefficients sont positifs ce qui veut dire que ces variables influent

positivement sur la probabilité de pratiques culturelles. En outre, le fait d'être en classe préparatoire et d'avoir un budget destiné à la culture entre 0 et 70 euros, rend plus probable la pratique d'activités culturelles. Cependant, cette probabilité est plus élevée si l'étudiant a un budget entre 21 et 70 euros, que s'il a un budget inférieur à 21 euros.

De nombreuses variables qui semblaient théoriquement avoir un impact sur la fréquence des pratiques culturelles se révèlent être non significatives dans notre modèle, comme le genre ou le fait d'être boursier. De plus, nous suspectons de l'éventuelle multicollinéarité entre nos variables et ainsi de l'hétéroscédasticité, mais tout d'abord nous tenterons de vérifier l'hypothèse d'égalité des pentes. La fonction `vglm` permet pas à priori de vérifier cette hypothèse mais le logiciel nous indique que les valeurs ajustées sont proches de 0 ou 1. Pour cela, nous estimons à nouveau le modèle avec la fonction `clm` dont les résultats se trouvent en [annexe 14](#). Nous pouvons voir qu'à nouveau, seules les variables boursières, travail et genre ont pu valider cette hypothèse au seuil de 5% puisque la fonction n'a pas permis de calculer les p-value des autres variables. L'erreur est potentiellement due au nombre importants de modalités de nos variable. Cependant, nous ne souhaitons pas dichotomiser celles-ci car la perte d'information serait trop importante. L'hypothèse d'égalité des pentes n'est donc pas vérifiée pour le modèle global, nous interpréterons pas ses coefficients.

A présent, nous allons tout de même utiliser la méthode "*stepwise*" afin de trouver un modèle pertinent à notre analyse. La procédure *stepwise* tente de minimiser le critère AIC. Trois méthodes ont ainsi été testées : la méthode "*forward*", "*backward*" et "*both*". Ainsi, la méthode *forward* en [annexe 9](#) aboutit au modèle initial tandis que les méthodes *backward* et *both* statuent sur le même modèle à retenir dont l'AIC est minimisé et égal à 916,13 avec les trois variables explicatives, ci-dessous :

Tableau 17: Procédure stepwise

Step: AIC=916.13

somme4 ~ base1\$sensibilisation_culture + base1\$budget + base1\$formation

	Df	AIC
<none>		916.13
+ base1\$genre	1	916.18
+ base1\$boursier	1	916.60
+ base1\$travail	1	918.12
+ base1\$participation_acti_univ	2	918.23
- base1\$formation	6	921.50
+ base1\$domaineEtude	13	933.26
- base1\$sensibilisation_culture	3	933.31
- base1\$budget	6	933.48

Comme nous pouvons le voir, le modèle a été réduit à 3 variables explicatives. Notons que ces trois variables sont celles que nous avons identifiées comme statistiquement significatives précédemment grâce à la première estimation avec vglm, excepté pour la variable domaine d'étude. Ainsi, nous estimons un nouveau modèle afin de vérifier également la significativité de la répartition. Les résultats en [annexe 10](#) montrent que les classes sont également significatives. Nous retenons donc ce modèle (modèle 2) que nous pourrions comparer

Tableau 18: Modèle 2 vglm

Variable explicatives		Modèle multinomial ordonné	Ecart-type
Sensibilisation culture	0 =Non	Référence	
	1= Oui par mes parents	0.0001***	3.248346e-01
	2=Oui par des cours extra-scolaire	0.0541**	3.335087e-01
	3=Oui pour les deux	0.0000***	2.873600e-01
Budget	0= 0 euros	Référence	
	1= [1 ;10]	0.0085***	5.789521e-01
	2= [11 ;20]	0.0004***	5.230043e-01
	3= [21 ;50]	0.0000***	5.329133e-01
	4= [51 ;70]	0.0000***	6.837077e-01
	5= [71 ;100]	0.0518**	9.570248e-01
	6= > 100 euros	0.0953**	1.013132e+00
Formation	0= BTS	Référence	
	1= CAP/BEP	NA	
	2= Classes préparatoires	0.000***	4.280062e-07

3= École	0.6435	7.965842e-01
4= Licence universitaire	0.1191	7.182231e-01
5= Licence professionnelle	0.0029***	1.119350e+00
6= Master universitaire	0.1170	7.209390e-01
7= Doctorat	0.0698**	1.026705e+00

b. Interprétations

Tout d'abord, nous pouvons affirmer que le modèle a un intérêt puisque au moins une variable a un coefficient non nul, avec une p-value égale à $1.680093e-06 < 0.05$. On obtient alors pour le modèle 2, dont les coefficients se trouvent ci-dessous :

Ainsi, en supposant que l'hypothèse d'égalité des pentes est vérifiée et que les erreurs de notre modèle sont heteroscedastiques, nous pouvons procéder au calcul des odd-ratios pour interpréter les coefficients qui en ressortent :

Tableau 19: Odd-ratios du modèle 2

```
> exp(coef(modelord))
```

sensibilisation_culture1	sensibilisation_culture2	sensibilisation_culture3	budget1	budget2	budget3	budget4
3.693470e+00	1.901119e+00	3.337691e+00	4.592728e+00	6.423371e+00	9.607941e+00	2.176535e+01
budget5	budget6	formation2	formation3	formation4	formation5	formation6
6.433583e+00	5.418578e+00	4.287931e+07	1.445796e+00	3.063091e+00	2.803155e+01	3.095634e+00
formation7						
6.433196e+00						

La culture

Les étudiants ayant eu une sensibilisation à la culture par leurs parents ont 3.69 fois plus de chance d'avoir une consommation culturelle conséquente, par rapport aux personnes n'ayant pas été sensibilisées.

Les étudiants ayant eu une sensibilisation à la culture par des cours extra-scolaires ont 1.90 fois plus de chance de pratiquer très régulièrement une activité culturelle par rapport à ceux n'ayant pas été sensibilisés.

Les étudiants ayant été sensibilisé à la fois par leurs parents et des cours extra-scolaires ont 3.34 fois plus de chance d'avoir une consommation culturelle conséquente, par rapport à ceux qui n'ont pas été sensibilisés.

Le budget

Les étudiants ayant un budget destiné à leurs pratiques culturelles entre 1 et 10 euros ont 4.59 fois plus de chance de pratiquer très régulièrement une activité culturelle par rapport à ceux qui n'ont pas de budget pour. Les étudiants qui allouent 11 à 20 euros de leur budget à ces pratiques ont 6.42 fois plus de chance de les faire très régulièrement par rapport à ceux qui ne disposent pas de budget à allouer à la culture. Ainsi, globalement la probabilité augmente à mesure que le budget augmente jusqu'au seuil du budget de 50 euros puisque les personnes ayant un budget culture entre 21 et 50 euros ont 9.61 fois plus de chance de pratiquer une activité fréquemment par rapport aux personnes sans budget. On observe que les personnes ayant un budget entre 51 et 70 euros ont seulement 2.18 fois plus de chance de pratiquer des activités artistiques que les étudiants sans budget. De plus, on remarque que la relation budget-fréquence devient décroissante à partir de 71 euros. En outre, les personnes ayant un budget entre 71 et 100 euros ont 6.43 fois plus de chance de pratiquer une activité artistique fréquemment tandis qu'une personne ayant un budget supérieur à 100 euros a seulement 5.42 fois plus de chances.

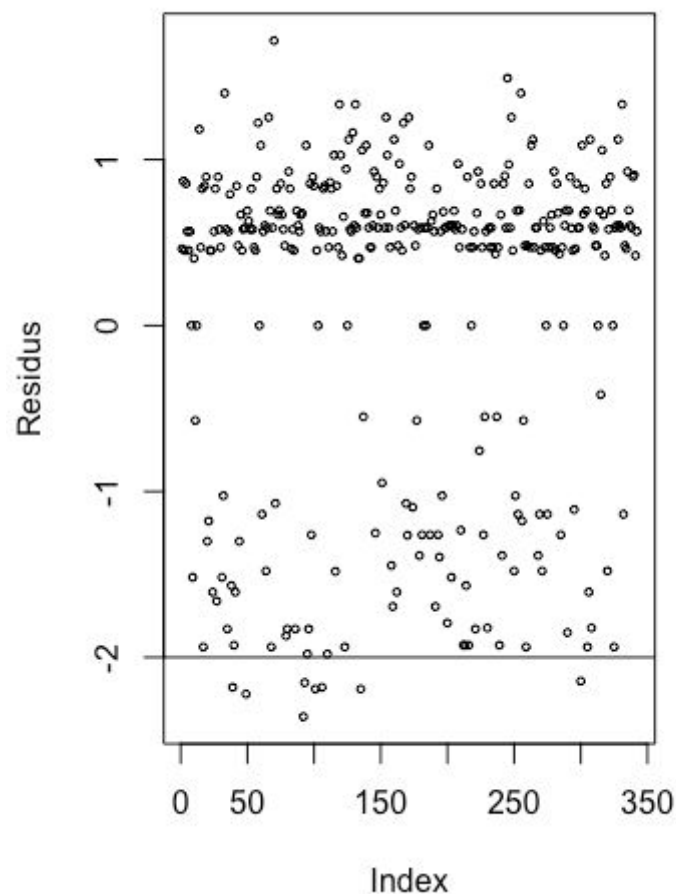
La formation

Un étudiant en classe préparatoire a 4.29 fois plus de chance de pratiquer très régulièrement et donc d'avoir une consommation culturelle conséquente par rapport à une personne en BTS. Paradoxalement, un étudiant d'école a seulement une probabilité 1.45 fois plus élevée d'avoir une pratique culturelle conséquente. Un étudiant de licence, 3.06 fois, un étudiant de licence pro 2.80 fois, un étudiant de master universitaire 3.09 fois, et un doctorant 6.43 fois plus de chance d'avoir des activités culturelles très régulièrement par rapport aux étudiants en BTS. Ces résultats semblent cohérents avec la partie théorique puisque les théories du capital culturel insistent sur l'influence du niveau d'étude sur la fréquence des pratiques culturelles.

c. Qualité du modèle

Concernant la qualité de notre modèle nous pouvons considérer trois indicateurs. La qualité de prévision, qui est ici égale à 23.1 ([annexe 10](#)) ce qui n'est pas vraiment satisfaisant puisque comme on peut le voir sur la matrice, beaucoup des valeurs sont mal prédites par le modèle (en dehors de la diagonale). Ensuite, la qualité d'ajustement calculé par le R^2 de Mc Fadden qui s'élève à 22% et l'intérêt du modèle calculé par le *likelihood ratio* test dont la p-value est inférieure à 0.05, ce qui nous permet d'affirmer que notre modèle à un intérêt au seuil de 1%. Nous devons également nous assurer qu'aucune observation n'influence les estimations de manière significative, pour cela nous étudions les résidus se trouvant dans l'intervalle $[-2;2]$, que nous pouvons apercevoir sur le graphique ci-dessous :

Figure 5: Résidus du modèle 2



Ainsi, comme nous pouvons le voir en bas du graphique, 8 observations semblent sortir de l'intervalle. Pour remédier à tout problème d'hétéroscédasticité, nous avons supprimé ces observations afin de réestimer le modèle.

3. Amélioration du modèle

Une fois après avoir supprimé ces observations, nous avons estimé un nouveau modèle sans la variable formation. En effet, les modalités liées à la variable formation n'étaient pas significatives dans le modèle. De plus, cette variable présente beaucoup de modalités très inégalement réparties. Un seul étudiant est représenté la deuxième modalité. Ainsi, il convient de supprimer la variable afin de potentiellement améliorer le modèle et pouvoir le comparer.

Tableau 20: Estimation du modèle 3 sans formation

```
Call:
vglm(formula = somme4 ~ sensibilisation_culture + budget, family = cumulative(parallel = TRUE,
  reverse = TRUE), data = basevglm, link = "logit")

Pearson residuals:
      Min      1Q  Median      3Q     Max
logitlink(P[Y>=2]) -3.804 -0.4587  0.2516  0.4846  1.605
logitlink(P[Y>=3]) -2.721 -0.5611  0.2925  0.8117  4.059
logitlink(P[Y>=4]) -1.556 -0.8092 -0.2551  0.8029  3.095

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept):1      -1.3995      0.5357  -2.612  0.008989 **
(Intercept):2      -2.5023      0.5479  -4.567  4.95e-06 ***
(Intercept):3      -3.6820      0.5612  -6.561  5.33e-11 ***
sensibilisation_culture1  1.2100      0.3137   3.857  0.000115 ***
sensibilisation_culture2   0.4595      0.3214   1.430  0.152834
sensibilisation_culture3  1.0687      0.2807   3.807  0.000140 ***
budget1              1.3614      0.5816   2.341  0.019245 *
budget2              1.7495      0.5310   3.295  0.000985 ***
budget3              2.1474      0.5385   3.988  6.67e-05 ***
budget4              2.9244      0.6707   4.360  1.30e-05 ***
budget5              1.7888      0.9564   1.870  0.061449 .
budget6              1.5748      1.0375   1.518  0.129067
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Vérification des hypothèses

Avant de pouvoir interpréter les coefficients de ce modèle, il convient tout d'abord de vérifier l'hypothèse d'égalité des pentes à partir des odd-ratios. Pour cela nous utilisons `vglm` de la librairie `VGAM`. Ainsi, en calculant la différence entre le modèle pour lequel l'argument 'parallèle' était vrai et celui pour lequel l'argument était faux, nous obtenons une p-value de 1 ([annexe 11.b](#)). En outre, l'hypothèse d'égalité des pentes est vérifiée pour toutes les variables.

Il convient ensuite de vérifier l'hypothèse d'homoscédasticité des erreurs que nous n'avons fait que supposer jusqu'à présent. La fonction `hetglm` nous donne les résultats suivants sur le modèle 3 :

Tableau 21: Test d'homoscédasticité des erreurs avec `hetglm`

```
Call:
hetglm(formula = somme4 ~ base1$sensibilisation_culture + base1$budget | base1$sensibilisation_culture + base1$budget, data = basevglm,
family = binomial(link = "logit"), control = hetglm.control(method = "BFGS", maxit = c(500, 25), hessian = FALSE, trace = FALSE),
model = TRUE)

Deviance residuals:
      Min       1Q   Median       3Q      Max
-2.2155  0.0000  0.6471  0.8136  1.9690

Coefficients (binomial model with logit link):
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.783e+00  1.077e+00  -1.656   0.0977
base1$sensibilisation_culture1  4.342e+01  2.357e+13   0.000   1.0000
base1$sensibilisation_culture2 -2.164e-01  5.644e+09   0.000   1.0000
base1$sensibilisation_culture3  1.596e+00  3.895e+09   0.000   1.0000
base1$budget1    2.000e+00  5.645e+09   0.000   1.0000
base1$budget2    2.000e+00  5.645e+09   0.000   1.0000
base1$budget3    2.000e+00  5.644e+09   0.000   1.0000
base1$budget4    2.000e+00  5.645e+09   0.000   1.0000
base1$budget5    4.357e+01  2.374e+13   0.000   1.0000
base1$budget6    6.124e-01  9.351e+09   0.000   1.0000

Latent scale model coefficients (with log link):
              Estimate Std. Error z value Pr(>|z|)
base1$sensibilisation_culture1  4.132e+00  5.663e+11   0.000   1.0000
base1$sensibilisation_culture2 -9.944e+00  2.823e+09   0.000   1.0000
base1$sensibilisation_culture3  8.690e-01  2.082e+10   0.000   1.0000
base1$budget1   -2.898e-01  2.608e+10   0.000   1.0000
base1$budget2   -6.507e-01  2.608e+10   0.000   1.0000
base1$budget3   -1.135e+00  2.608e+10   0.000   1.0000
base1$budget4   -1.149e+00  2.608e+10   0.000   1.0000
base1$budget5    1.383e+00  5.682e+11   0.000   1.0000
base1$budget6   -1.105e+01  7.988e+09   0.000   1.0000

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-likelihood: -170.5 on 19 Df
LR test for homoskedasticity: 9.44 on 9 Df, p-value: 0.3976
Dispersion: 1
Number of iterations in BFGS optimization: 242
```

Ainsi comme nous pouvons le voir, aucune variable ne présente aucune source d'hétéroscédasticité dans le modèle. De plus la p-value associée au LR test est supérieure à 0.05, ce qui prouve qu'il n'y a pas de multicollinéarité dans le modèle. En [annexe 15](#) nous pouvons voir que la variable formation ajoutée au modèle n'est pas significative. Ainsi, nous pouvons conclure que le modèle 3 a un intérêt puisque la p-value est également inférieure à 0.05%. La qualité de prévision est ici égale à 2.09 et le taux d'erreur de prédiction égal à 0,19, les informations liées à la qualité du modèle sont résumés dans le tableau ci-dessous :

Tableau 22 : Qualité des deux modèles

	Modèle 3	Modèle 2
Intérêt du modèle	6.73869e-09	1.680093e-06
Qualité de prévision	2,09	23.39
Taux d'erreur de prédiction	0.19	0.01
Sensibilité	0.93	0
Spécificité	0.24	0.98
Précision	0.54	0
Faux positif	0.76	0.018

Nous pouvons voir que la modèle 2 semble être de meilleur qualité puisqu'il a une qualité de prévision bien meilleure, ainsi qu'un taux d'erreur bien plus faible. Seulement, nous n'avons pas pu vérifier l'hypothèse d'égalité des pentes pour la variable formation. Comme nous l'avons dit tout à l'heure, il faudrait regrouper les modalités pour que l'hypothèse fonctionne. En effet, avec le logiciel R, il n'est pas possible d'estimer un modèle ne vérifiant pas cette hypothèse. Ainsi, nous conserverons seulement le modèle 3 qui respecte toutes les conditions et duquel nous discuterons en conclusion.

IV- Modèle multinomial non ordonné

Pour compléter l'analyse des déterminants des fréquences de pratique de certaines activités culturelles ainsi que ceux du fait d'effectuer une sortie culturelle au théâtre, nous aurions aimé étudier la dichotomie des pratiques culturelles des étudiants nantais. En effet, nous avons à plusieurs reprises parlé des différences de consommation culturelle qu'il existe : nous avons soulevé l'écart entre des pratiques traditionnelles qui se réfèrent aux 6 premiers arts énumérés en introduction, et des pratiques nouvelles, jeunes que sont les "arts nouveaux" pratiqués théoriquement par les jeunes et les classes inférieures.

Le but d'une étude par le biais d'un modèle multinomial non ordonné était de trier les activités culturelles des étudiants selon si elles étaient jeunes ou traditionnelles. Nous avons ainsi établi les modalités suivantes :

- Culture traditionnelle : opéra, théâtre, conférences, musées.
- Culture intermédiaire : concerts/festivals, bibliothèques, lecture et cinéma.
- Culture jeune : jeux vidéos, télévision, radio.

L'objectif était donc de classer chaque individu dans une de ces trois catégories en fonction des activités qu'il pratiquait. Cependant dans notre questionnaire nous avons posé la question des fréquences de pratiques de chacune des activités énumérées ci-dessus, nous devons donc procéder à un recodage des variables pour obtenir un classement clair des individus dans l'une des 3 classes de ce modèle.

Nous avons d'abord pensé à additionner les numéros associés à chaque modalité qui traduisent une fréquence croissante, de manière à obtenir un comptage de points par types de culture (traditionnelle, intermédiaire ou jeune) pour ainsi classer l'individu dans le type de culture qui a récolté le plus de points, c'est à dire le type de culture qu'il pratique le plus souvent. En prenant garde à changer le chiffre 5 attribué à la modalité '*ne sait pas*' concernant les fréquences en 0 c'est à dire avec la modalité '*jamais*'. Ainsi, nous aurions obtenu un score relevant des pratiques individuelles de consommation culturelle chez les jeunes. Cependant, il apparaît évident que la très grande majorité des étudiants auraient eu le score le plus élevé dans la case '*culture jeune*'. En effet, nous avons pu voir avec l'ACM que les individus n'étaient pas très différents les uns des autres et ainsi, que les modalités sont proches entre les observations. Même si un étudiant s'intéresse à la culture traditionnelle en se rendant de

temps en temps au théâtre, à des conférences etc, cette consommation reste marginale face à une écoute régulière de la radio et à un visionnage quotidien de la télévision. Suivant cette méthode de recodage nous n'aurions eu aucun individu classé en culture traditionnelle.

Nous avons donc pensé à une autre méthode qui puisse donner des résultats équi- répartis entre les 3 types de culture : prendre uniquement l'activité qui est la plus fréquente chez l'individu, la considérer comme activité principale et le classer selon cette dernière. Cela pallie au fait que l'addition de plusieurs activités et notamment la radio, la télévision et les jeux vidéos, peuvent vite atteindre un chiffre élevé difficilement atteignable par les 2 autres types d'activités, et qu'ainsi la culture jeune prenne le dessus chez bon nombre de répondants. Néanmoins en prenant l'activité la plus fréquente chez les individus il apparaît que les activités de la culture dite '*jeune*' arrivent en tête de classement et que l'on se retrouve avec le même problème de répartition des étudiants dans les 3 modalités non ordonnées d'un éventuel modèle multinomial.

Par conséquent la réalisation d'une analyse visant à opposer différents types de pratiques culturelles par l'utilisation d'un modèle multinomial non ordonné s'avère impossible. En effet, l'indispensable équi-répartition des répondants dans les différentes modalités nous empêche de pouvoir mener cette étude étant donné l'écart évident d'effectifs qu'il y aura entre les pratiques telles que radio, télévision, et celle du théâtre ou même des conférences. Sachant que nous avons porté l'analyse uniquement sur un public jeune (les étudiants de nantes), nous aurions pu deviner que l'opposition entre culture jeune et traditionnelle au sein des étudiants était impossible. Pour pallier à cela il aurait fallu étendre le public visé à une plus large tranche d'âge, en prenant en considération toutes les personnes majeures par exemple. On aurait ainsi certainement pu noter les différences de pratiques culturelles au sein de l'échantillon, pour voir par la suite si les pratiques jeunes sont à juste titre attribuées aux étudiants, comme les pratiques traditionnelles le sont aux personnes ayant plus d'âge.

Voilà un point intéressant qu'il nous faudra retenir pour la suite : avant de diffuser un questionnaire il est très important d'avoir ciblé parfaitement le type de répondants (nous savions qu'il s'agissait des étudiants mais n'avions pas pensé au manque "*d'opposition*" pour remplir les effectifs des autres modalités), et de savoir parfaitement quelle question servira à quel modèle, dans quel but et pour répondre à quelle problématique.

PARTIE 3 : Conclusion et discussion des résultats

Étant donné l'importance de la Culture dans notre société comme une histoire qui nous définit mais aussi comme un outil d'appartenance à un groupe social, l'objet de cette étude était de comprendre les causes, les facteurs des pratiques culturelles sur un public axé uniquement sur les étudiants nantais.

Dans cette étude nous avons été confrontées à plusieurs difficultés en commençant par le modèle binaire ; les questions initialement posées pour servir de variables à expliquer se sont avérées inexploitable car la répartition entre les 2 modalités était trop différente. Nous avons donc été contraintes de changer le Y de cette partie pour finalement retenir la fréquence de sorties au théâtre que nous avons recodée en 2 modalités ; se rendre ou ne pas se rendre au théâtre. Cette partie de l'analyse a été concluante et au vu des résultats il apparaît que se rendre au musée/expositions, à des concerts/festivals et le fait d'avoir été sensibilisé à n'importe quelle forme d'Art étant jeune augmente la probabilité de se rendre au théâtre. Cela peut alors s'expliquer par la curiosité d'esprit : si un individu fait déjà "l'effort" (qui n'en est peut être pas un) de se rendre à des musées ou des concerts, il a alors plus de chance d'effectuer une autre sortie au théâtre, par rapport à un individu qui reste chez lui pour regarder la télévision par exemple. Voilà pourquoi le fait de regarder la télévision diminue la probabilité d'aller au théâtre. De même, si un individu a été sensibilisé à l'art étant petit il gardera cette curiosité pour tout ce qui s'y réfère et sera alors plus enclin à reproduire les sorties culturelles de son enfance, par rapport à une personne qui n'a jamais approché la Culture.

Concernant l'étude des fréquences réalisée à travers un modèle multinomial ordonné, nous avons pu établir l'influence de la sensibilisation à la culture, du budget, ainsi que le niveau de formation sur la fréquence de pratique d'activités culturelles. Les résultats obtenus correspondant tout à fait aux effets supposés dans la partie économique. Ainsi, nous avons pu confirmer que l'éducation et la sensibilisation à une pratique culturelle avait un impact positif significatif sur la fréquence de celle-ci. Cet effet était d'autant plus important que l'étudiant ait été sensibilisé à la fois par ses parents et des cours extra-scolaires, puisque sa probabilité d'avoir une fréquence régulière était 3.34 fois plus élevée que pour un étudiant n'ayant pas

été du tout sensibilisé à la culture. De plus nous avons pu établir un lien croissant puis décroissant entre le budget et la fréquence de consommation culturelle, avec une rupture au niveau du budget entre 51 et 70 euros. Ainsi, selon notre modèle, jusqu'à 51 euros, plus le revenu augmente, plus la fréquence de consommation augmente. En revanche, après 70 euros, plus le budget culturel est conséquent, moins la fréquence des pratiques culturelles augmente. Ce phénomène peut s'expliquer tout d'abord par un effet revenu basique, plus le revenu augmente, plus l'individu aura d'argent à allouer au loisir. A contrario, notre échantillon étant composé exclusivement d'étudiants, un budget supérieur à 70 euros représente une somme relativement conséquente. Avec un tel budget, les possibilités de loisirs sont grandes et n'ont pas pu être toutes représentées dans notre étude. De plus, notre variable des fréquences ne distingue d'aucune manière la culture dite "légitime" de la "sous-culture". Nous savons que le budget culturel important d'un étudiant peut-être apparenté à son appartenance à une classe sociale supérieure. Les théories sociologiques ont également démontré l'existence d'une stratification sociale culturelle. En outre, un étudiant ayant pour habitude des pratiques coûteuses comme l'opéra ou le théâtre, mais n'ayant aucune familiarité avec le visionnage de la télé ou la pratique des jeux-vidéos aura un score modéré du point de vue de la fréquence (notre Y), ce qui peut expliquer que la fréquence décroît avec le revenu à partir de 70 euros de budget mensuel. Pour finir, nous avons également pu démontrer que la fréquence d'activités culturelles augmentait avec le niveau d'étude. La seule surprise peut éventuellement être le niveau de formation "école" qui peut être une piste d'amélioration de notre modèle. En effet, nous obtenons que une probabilité seulement 1.49 fois plus élevée que les étudiants en BTS. En effet, ce terme est bien trop large pour représenter la diversité des écoles en études supérieures (écoles d'hôtellerie, grandes écoles, école d'art etc.). *Idem*, les étudiants en classe préparatoire ont une probabilité élevée par rapport aux étudiants en BTS d'avoir une fréquence importante de pratiques culturelles, cependant, ce chiffre doit-être pris avec précaution car seulement un étudiant de notre enquête était issu de classes préparatoires. Malgré le nettoyage des observations atypiques, les faibles effectifs de certaines modalités ont pu influencer nos résultats.

Enfin, nous aurions aimé souligner l'écart entre les pratiques culturelles jeunes et traditionnelles grâce à un modèle multinomial non ordonné. Cependant nous avons été rapidement freinées dans cette analyse puisque le recodage des variables de manière à garder une répartition équivalente entre les 3 types de culture (traditionnelle, intermédiaire et

jeune), était irréalisable compte tenu de notre base de données. Il est certes impossible de prévoir les réponses des répondants au sondage, mais il est important d'essayer de cerner au mieux les réponses avant diffusion définitive du questionnaire. Nous avons ainsi vu comment les réponses à la question concernant le domaine d'étude par exemple, étaient incomplètes et imprécises étant donné le pourcentage de répondants étudiant l'économie. De même, le manque d'anticipation sur l'homogénéité des réponses concernant les différents types de pratiques culturelles nous a en quelques sortes bloquée dans notre analyse.

Il serait donc intéressant de compléter cette étude en élargissant le champ des répondants au plus de 18 ans par exemple, de manière à obtenir une plus grande variété des répondants et donc une meilleure répartition de ces derniers dans les différentes classes. De pair avec des connaissances plus pointues dans le recodage des variables, nous pourrions ainsi effectuer un plus grand nombre d'analyses. Aussi, sachant que c'était la première fois que chacune de nous réalisait une enquête et conscientes des limites voici un tableau récapitulatif des points forts et faibles de notre enquête. Les points forts nous servent d'exemple pour les prochaines quand les points faibles d'expérience pour plus tard.

Tableau 22: Points forts et points faibles de nos modèles

Points faibles	Points forts
<ul style="list-style-type: none"> – mauvaise anticipation de l'homogénéité du profils des répondants: trop de filles (66%)et trop de personnes en économie (68%). - le public visé ne permet pas de réaliser une des trois analyses voulues 	<ul style="list-style-type: none"> - un sujet d'étude intéressant - les 2 premières questions =donnent une double informations (fréquence et activité) - a permit de réaliser 2 analyses ayant abouti

Pour finir, nous pouvons conclure que la question de la démocratisation de l'accès à la culture reste une question importante du débat public puisqu'en effet, une part non négligeable de notre échantillon estimait n'avoir jamais été sensibilisée à la culture durant l'enfance. De fait, le modèle multinomial ordonné comme le modèle binaire a montré l'importance de cette variable dans les habitudes de pratiques culturelles. Ainsi, l'école peut jouer un rôle dans l'éducation afin de gommer les inégalités sociales préexistantes qui influencent les comportements. L'échantillon de données que nous avons récolté a montré que 89,4% des étudiants souhaitent diversifier/intensifier leurs pratiques culturelles

pourtant 23% considèrent encore aujourd’hui que certaines pratiques sont réservées à une catégorie de personnes. De même, le budget influe sur la consommation culturelle. 62% estimait que le coût des activités représentait un frein aux pratiques culturelles. Ainsi, pratiquer des tarifs étudiants, les initiatives tels que les “pass-culture” etc, sont donc nécessaires à la démocratisation des pratiques culturelles. Ainsi, chacun doit pouvoir s’il le souhaite accéder à la culture et trouver la forme d’art qui lui correspond en fonction de ses affinités.

PARTIE 4 : Annexes

Annexe n°1 : Dictionnaire des variables

freqOpera : Fréquence de sortie à l'opéra.

freqTheatre : Fréquence de sortie au théâtre.

freqConcert : Fréquence de sortie à des concerts ou des festivals.

freqConference : Fréquence de sortie à des conférences.

freqMusee : Fréquence de sortie à des musées ou des expositions.

freqCinema : Fréquence de sortie au cinéma.

freqBiblio : Fréquence de sortie à la bibliothèque ou à la médiathèque.

freqRadio : Fréquence d'écoute de la radio.

freqTele : Fréquence de visionnage de la télévision.

freqJeuxVid : Fréquence de consommation de jeux vidéos.

freqLecture : Fréquence de lecture en tous genres.

0	Jamais
1	1 à 2 fois par an
2	3 à 10 fois par an
3	1 fois par mois
4	Au moins 1 fois par semaine
5	Ne sait pas

participation_acti_univ : Participation aux activités proposées par l'établissement scolaire.

0	Non
1	Oui
3	Ne se prononce pas

freins_culture : Principaux freins aux pratiques culturelles.

0	Le coût des activités
---	-----------------------

1	Le manque de temps
2	Le manque de proximité
3	Le manque d'informations sur les événements
4	L'offre culturelle pour les étudiants ne me convient pas
5	Le sentiment que certaines pratiques sont réservées à une catégorie de personnes
6	Le manque d'envie

changement_pratiques_cult : Volonté d'améliorer ou diversifier les pratiques culturelles.

0	Non
1	Oui

si_oui_choix : Si oui, choix entre 'y consacrer plus de temps' et 'les diversifier'.

0	Consacrer plus de temps aux pratiques culturelles
1	Diversifier les pratiques culturelles

interetOpera : Intérêt pour l'opéra.

interetDanse : Intérêt pour la danse.

interetMusiqueClassique : Intérêt pour les concerts de musiques classique.

interetTheatre : Intérêt pour le théâtre.

interetMusee : Intérêt pour les musées et expositions.

interetConference : Intérêt pour les conférences.

interetBiblio : Intérêt pour les bibliothèques.

0	Ça ne m'intéresse pas
1	Pourquoi pas
2	J'aimerais beaucoup

sensibilisation_culture : Sensibilisation à la culture.

0	Non
1	Oui par mes parents

2	Oui par des cours extra-scolaires
3	Oui par les parents ET par des cours extra-scolaires

CSP_pere : Catégorie Socioprofessionnelle du père.

CSP_mere : Catégorie Socioprofessionnelle de la mère.

0	Agriculteurs exploitants
1	Artisans, commerçants et chefs d'entreprise
2	Cadres et professions intellectuelles supérieures
3	Professions intermédiaires
4	Employés
5	Indépendant
6	Ouvriers
7	Sans emploi
8	Ne sait pas
9	Ne se prononce pas

budget : Budget mensuel dédié aux sorties culturelles.

0	0€
1	1 à 10€
2	11 à 20€
3	21 à 50€
4	51 à 70€
5	71 à 100€
6	Plus de 100€

boursier : Le répondant est boursier.

0	Non
1	Oui

travail : Le répondant travaille en dehors de ses études.

0	Non
1	Oui

formation : Type de formation suivie par le répondant.

0	BTS
1	CAP/BEP
2	Classes préparatoires
3	Ecole
4	Licence universitaire
5	Licence pro
6	Master universitaire
7	Doctorat

domaineEtude : Domaine d'étude du répondant.

0	Économie
1	Gestion
2	Finance
3	Sociologie
4	Droit
5	Histoire
6	Lettres
7	Psychologie
8	Philosophie
9	Commerce
10	Architecture
11	Sciences (ingénierie, mathématiques, informatique, mécanique)

12	Sciences (Biologie, médecine, physique-chimie)
13	Art

public_prive : Nature de l'établissement du répondant.

0	Public
1	Privé

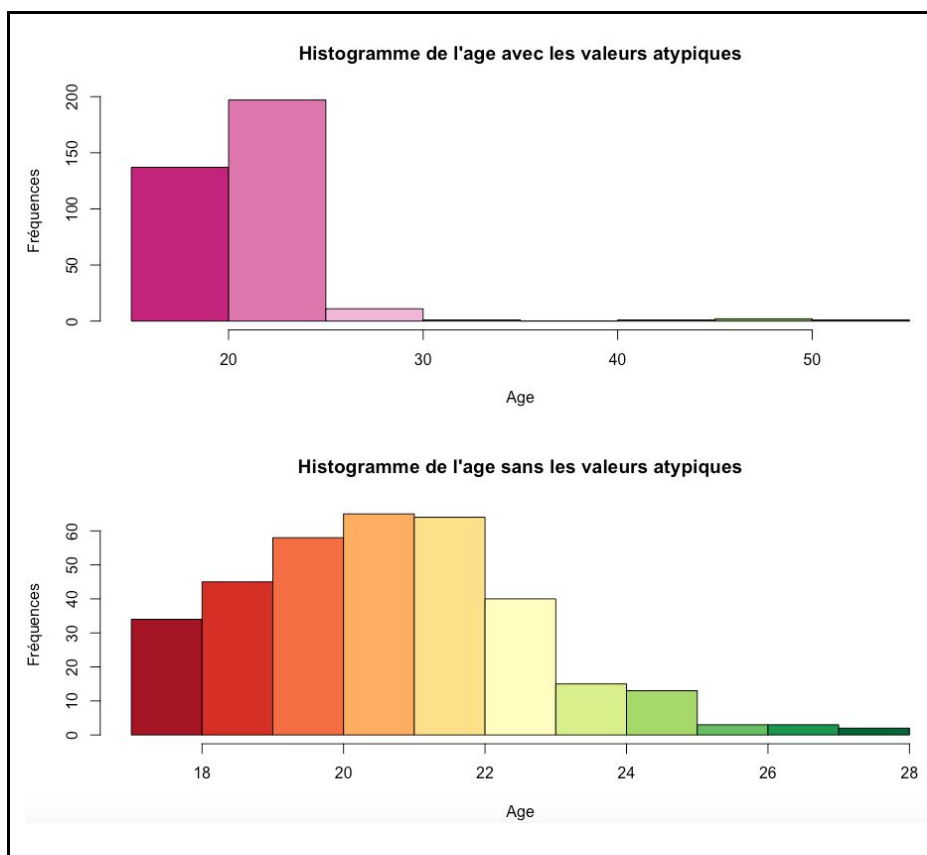
genre : Genre du répondant.

0	Femme
1	Homme

age : Âge du répondant.

...	Âge en années
-----	---------------

Annexe n°2 : Histogrammes de la variable 'age' avec et sans les valeurs atypiques



Annexe n°3 : Statistiques descriptives des pratiques culturelles des répondants

Data Frame Summary

base3

Dimensions: 342 x 21

Duplicates: 3

No	Variable	Stats / Values	Freqs (% of Valid)	Graph
1	freqOpera [factor]	1. 0 2. 1 3. 2	306 (89.5%) 33 (9.7%) 3 (0.9%)	
2	freqTheatre [factor]	1. 0 2. 1 3. 2 4. 3 5. 4 6. 5	196 (57.3%) 123 (36.0%) 18 (5.3%) 3 (0.9%) 1 (0.3%) 1 (0.3%)	
3	freqConcert [factor]	1. 0 2. 1 3. 2 4. 3 5. 4	68 (19.9%) 185 (54.1%) 76 (22.2%) 10 (2.9%) 3 (0.9%)	
4	freqConference [factor]	1. 0 2. 1 3. 2 4. 3 5. 4 6. 5	143 (41.8%) 131 (38.3%) 58 (17.0%) 6 (1.8%) 3 (0.9%) 1 (0.3%)	
5	freqMusee [factor]	1. 0 2. 1 3. 2 4. 3 5. 4 6. 5	65 (19.0%) 154 (45.0%) 105 (30.7%) 16 (4.7%) 2 (0.6%) 0 (0.0%)	
6	freqCinema [factor]	1. 0 2. 1 3. 2 4. 3 5. 4	14 (4.1%) 49 (14.3%) 198 (57.9%) 73 (21.3%) 8 (2.3%)	
7	freqBiblio [factor]	1. 0 2. 1 3. 2 4. 3 5. 4 6. 5	64 (18.7%) 55 (16.1%) 92 (26.9%) 51 (14.9%) 78 (22.8%) 2 (0.6%)	
8	freqRadio [factor]	1. 0 2. 1 3. 2 4. 3 5. 4 6. 5	78 (22.8%) 19 (5.6%) 20 (5.9%) 41 (12.0%) 180 (52.6%) 4 (1.2%)	

9	freqTele [factor]	1. 0 2. 1 3. 2 4. 3 5. 4 6. 5	31 (9.1%) 13 (3.8%) 18 (5.3%) 47 (13.7%) 230 (67.2%) 3 (0.9%)	
10	freqJeuxVid [factor]	1. 0 2. 1 3. 2 4. 3 5. 4 6. 5	89 (26.0%) 35 (10.2%) 41 (12.0%) 65 (19.0%) 110 (32.2%) 2 (0.6%)	
11	freqLecture [factor]	1. 0 2. 1 3. 2 4. 3 5. 4 6. 5	12 (3.5%) 21 (6.1%) 50 (14.6%) 63 (18.4%) 191 (55.9%) 5 (1.5%)	
12	participation_acti_univ [factor]	1. 0 2. 1 3. 2	198 (57.9%) 142 (41.5%) 2 (0.6%)	
13	changement_pratiques_cult [factor]	1. 0 2. 1	34 (9.9%) 308 (90.1%)	
14	si_oui_choix [factor]	1. 0 2. 1	168 (54.4%) 141 (45.6%)	
15	interetOpera [factor]	1. 0 2. 1 3. 2	148 (43.3%) 126 (36.8%) 68 (19.9%)	
16	interetDanse [factor]	1. 0 2. 1 3. 2	94 (27.5%) 128 (37.4%) 120 (35.1%)	
17	interetMusiqueClassique [factor]	1. 0 2. 1 3. 2	94 (27.5%) 137 (40.1%) 111 (32.5%)	
18	interetTheatre [factor]	1. 0 2. 1 3. 2	55 (16.1%) 141 (41.2%) 146 (42.7%)	
19	interetMusee [factor]	1. 0 2. 1 3. 2	25 (7.3%) 130 (38.0%) 187 (54.7%)	
20	interetConference [factor]	1. 0 2. 1 3. 2	47 (13.7%) 167 (48.8%) 128 (37.4%)	
21	interetBiblio [factor]	1. 0 2. 1 3. 2	64 (18.7%) 160 (46.8%) 118 (34.5%)	

Generated by [summarytools](#) 0.9.5 (R version 3.5.0)
2020-03-07

Annexe n°4 : Statistiques descriptives des caractéristiques des répondants

Data Frame Summary

base2

Dimensions: 342 x 11

Duplicates: 4

No	Variable	Stats / Values	Freqs (% of Valid)	Graph
1	sensibilisation_culture [factor]	1. 0 2. 1 3. 2 4. 3	70 (20.5%) 75 (21.9%) 66 (19.3%) 131 (38.3%)	
2	CSP_pere [factor]	1. 0 2. 1 3. 2 4. 3 5. 4 6. 5 7. 6 8. 7 9. 8 10. 9	12 (3.5%) 36 (10.5%) 126 (36.8%) 36 (10.5%) 47 (13.7%) 10 (2.9%) 32 (9.4%) 12 (3.5%) 10 (2.9%) 21 (6.1%)	
3	CSP_mere [factor]	1. 0 2. 1 < 3. 2 4. 3 5. 4 6. 5 7. 6 8. 7 9. 8 10. 9	6 (1.8%) 13 (3.8%) 96 (28.1%) 49 (14.3%) 92 (26.9%) 22 (6.4%) 8 (2.3%) 34 (9.9%) 7 (2.1%) 15 (4.4%)	
4	budget [factor]	1. 0 2. 1 3. 2 4. 3 5. 4 6. 5 7. 6	18 (5.3%) 41 (12.0%) 143 (41.8%) 111 (32.5%) 20 (5.9%) 5 (1.5%) 4 (1.2%)	
5	boursier [factor]	1. 0 2. 1	198 (57.9%) 144 (42.1%)	
6	travail [factor]	1. 0 2. 1	240 (70.2%) 102 (29.8%)	
7	formation [factor]	1. 0 2. 2 3. 3 4. 4 5. 5 6. 6 7. 7	6 (1.8%) 1 (0.3%) 26 (7.6%) 172 (50.3%) 6 (1.8%) 125 (36.5%) 6 (1.8%)	
8	domaineEtude [factor]	1. 0 2. 1 3. 10 4. 11 5. 12 6. 13 7. 2 8. 3 9. 4 10. 5 [4 others]	197 (57.6%) 35 (10.2%) 1 (0.3%) 19 (5.6%) 26 (7.6%) 8 (2.3%) 8 (2.3%) 5 (1.5%) 5 (1.5%) 12 (3.5%) 26 (7.6%)	
9	public_privé [factor]	1. 0 2. 1	312 (91.2%) 30 (8.8%)	
10	genre [factor]	1. 0 2. 1	228 (66.7%) 114 (33.3%)	
11	age [numeric]	Mean (sd) : 21.1 (2) min < med < max: 17 < 21 < 28 IQR (CV) : 2 (0.1)		12 distinct values

Annexe n°5 : Méthode stepwise de choix des variables pour le modèle binaire

```
> summary(modele.forward)
```

Call:

```
glm(formula = Theatre_binaire ~ freqMusee + freqLecture + sensibilisation_culture +  
    freqConcert + freqTele, family = binomial(logit), data = base_binaire)
```

Deviance Residuals:

```
      Min       1Q   Median       3Q      Max  
-2.5316 -0.8559 -0.2959  0.9680  2.1130
```

```
Null deviance: 466.18  on 341  degrees of freedom  
Residual deviance: 358.60  on 320  degrees of freedom  
AIC: 402.6
```

```
Number of Fisher Scoring iterations: 16
```

```
> summary(modele.backward)
```

Call:

```
glm(formula = Theatre_binaire ~ freqConcert + freqMusee + freqTele +  
    freqLecture + sensibilisation_culture, family = binomial(logit),  
    data = base_binaire)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.5316	-0.8559	-0.2959	0.9680	2.1130

Null deviance: 466.18 on 341 degrees of freedom

Residual deviance: 358.60 on 320 degrees of freedom

AIC: 402.6

Number of Fisher Scoring iterations: 16

```
> summary(modele.both)
```

Call:

```
glm(formula = Theatre_binaire ~ freqMusee + freqLecture + sensibilisation_culture +  
    freqTele, family = binomial(logit), data = base_binaire)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.3516	-0.9257	-0.3476	0.9639	2.5975

Null deviance: 466.18 on 341 degrees of freedom

Residual deviance: 367.60 on 324 degrees of freedom

AIC: 403.6

Number of Fisher Scoring iterations: 15

Annexe n°6 : Deuxième modèle logit (sans les 4 données influentes) + tests

```
> summary(logit2)
```

Call:
glm(formula = Theatre_binaire ~ freqConcert + freqMusee + freqTele +
freqLecture + sensibilisation_culture, family = binomial(link = "logit"),
data = base_binaire1)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.92920	-0.84963	-0.00017	0.91910	2.07702

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-19.4262	1663.9719	-0.012	0.990685
freqConcert1	0.4687	0.3586	1.307	0.191251
freqConcert2	0.9367	0.4170	2.247	0.024665 *
freqConcert3	1.2088	1.0222	1.183	0.237000
freqConcert4	-15.0105	1326.0316	-0.011	0.990968
freqMusee1	0.5455	0.4117	1.325	0.185154
freqMusee2	1.3470	0.4402	3.060	0.002212 **
freqMusee3	31.6519	1888.8943	0.017	0.986631
freqMusee4	19.4355	4612.2020	0.004	0.996638
freqTele1	-0.3115	0.7754	-0.402	0.687863
freqTele2	-1.9554	0.7904	-2.474	0.013367 *
freqTele3	-0.2746	0.5933	-0.463	0.643557
freqTele4	-1.0208	0.4816	-2.120	0.034040 *
freqTele5	0.1846	1.6864	0.109	0.912849
freqLecture1	-0.2831	2109.9692	0.000	0.999893
freqLecture2	17.7419	1663.9719	0.011	0.991493
freqLecture3	17.8147	1663.9719	0.011	0.991458
freqLecture4	17.6667	1663.9718	0.011	0.991529
freqLecture5	-0.9777	3504.3436	0.000	0.999777
sensibilisation_culture1	1.3703	0.4554	3.009	0.002619 **
sensibilisation_culture2	0.8013	0.4739	1.691	0.090905 .
sensibilisation_culture3	1.4422	0.4221	3.417	0.000634 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 459.90 on 337 degrees of freedom
Residual deviance: 332.48 on 316 degrees of freedom
AIC: 376.48

```
> vif(logit2)
```

	GVIF	Df	GVIF^(1/(2*Df))
freqConcert	2.192567	4	1.103111
freqMusee	2.279371	4	1.108478
freqTele	1.219915	5	1.020077
freqLecture	1.083012	5	1.008006
sensibilisation_culture	1.187835	3	1.029104

```
> print(pvalue)
[1] 3.051954e-17
```

```
> R2_MF
[1] 0.2770693
```

```
> hitmiss(logit2)
Classification Threshold = 0.5
      y=0 y=1
yhat=0 152 47
yhat=1 44 95
Percent Correctly Predicted = 73.08%
Percent Correctly Predicted = 77.55%, for y = 0
Percent Correctly Predicted = 66.9% for y = 1
Null Model Correctly Predicts 57.99%
[1] 73.07692 77.55102 66.90141
```

Annexe n°7 : 2eme et 3eme modèles logit supposant hétéroscédasticité des erreurs

Call:

```
hetglm(formula = Theatre_binaire ~ freqConcert +
freqMusee + freqTele + freqLecture +
sensibilisation_culture | freqConcert + freqMusee +
freqTele + freqLecture, data = base_binaire1, family
= binomial(link = "logit"), control =
hetglm.control(method = "BFGS", maxit = c(500, 25),
hessian = FALSE, trace = FALSE), model = TRUE)
```

```
Latent scale model coefficients (with log link):
      Estimate Std. Error z value Pr(>|z|)
freqConcert1 -2.715e-01 9.337e-01 -0.291 0.77123
freqConcert2 -2.255e+00 1.020e+00 -2.212 0.02699 *
freqConcert3 -8.484e+00 8.090e+02 -0.010 0.99163
freqConcert4 -1.380e+00 3.896e+04 0.000 0.99997
freqMusee1 3.291e-01 4.342e-01 0.758 0.44848
freqMusee2 1.988e+00 7.174e-01 2.771 0.00559 **
freqMusee3 -4.126e-01 3.630e+03 0.000 0.99991
freqMusee4 -7.439e-01 2.268e+01 -0.033 0.97384
freqTele1 -1.382e+01 3.055e+00 -4.525 6.05e-06 ***
freqTele2 1.113e+00 7.364e-01 1.512 0.13059
freqTele3 -1.862e-01 6.594e-01 -0.282 0.77771
freqTele4 4.048e-01 3.977e+04 0.000 0.99999
freqTele5 -4.960e+00 5.328e+03 -0.001 0.99926
freqLecture1 -2.606e+00 2.274e+01 -0.115 0.90876
freqLecture2 2.578e-01 2.272e+01 0.011 0.99095
freqLecture3 1.963e-01 2.273e+01 0.009 0.99311
freqLecture4 4.916e-01 7.165e+03 0.000 0.99995
freqLecture5 -9.151e-02 8.369e+17 0.000 1.00000
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-likelihood: -155 on 40 Df
LR test for homoskedasticity: 22.54 on 18 Df, p-value: 0.2089
Dispersion: 1
Number of iterations in BFGS optimization: 84
```

```
> vif(logit_homo1)
      GVIF Df GVIF^(1/(2*Df))
freqConcert 4.468603e+11 4 28.593773
freqMusee -7.150803e+27 4 NaN
freqTele 2.093014e+06 5 4.286247
freqLecture 7.494571e+17 5 61.302005
sensibilisation_culture 7.765477e+08 3 30.317591
```

Call:

```
hetglm(formula = Theatre_binaire ~ freqConcert +
freqMusee + freqTele + freqLecture +
sensibilisation_culture | freqConcert + freqMusee +
freqTele, data = base_binaire1, family = binomial(link
= "logit"), control = hetglm.control(method = "BFGS",
maxit = c(500, 25), hessian = FALSE, trace = FALSE),
model = TRUE)
```

```
Latent scale model coefficients (with log link):
      Estimate Std. Error z value Pr(>|z|)
freqConcert1 -3.699e-01 4.482e-01 -0.825 0.40918
freqConcert2 -2.555e+00 4.644e-01 -5.501 3.78e-08 ***
freqConcert3 -6.490e+00 9.534e+02 -0.007 0.99457
freqConcert4 -1.560e+00 8.008e+05 0.000 1.00000
freqMusee1 4.426e-01 4.102e-01 1.079 0.28064
freqMusee2 2.251e+00 4.538e-01 4.961 7.02e-07 ***
freqMusee3 -3.099e-01 5.744e+03 0.000 0.99996
freqMusee4 -8.713e-02 5.836e-01 -0.149 0.88131
freqTele1 -2.494e+01 1.827e+00 -13.649 < 2e-16 ***
freqTele2 1.987e+00 7.039e-01 2.823 0.00476 **
freqTele3 -8.083e-02 5.971e-01 -0.135 0.89233
freqTele4 4.031e-01 3.129e+03 0.000 0.99990
freqTele5 -3.559e+00 1.178e+18 0.000 1.00000
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-likelihood: -155 on 35 Df
LR test for homoskedasticity: 22.48 on 13 Df, p-value: 0.04842
Dispersion: 1
Number of iterations in BFGS optimization: 128
```

```
> vif(logit_homo2)
      GVIF Df GVIF^(1/(2*Df))
freqConcert 9.099206e+08 4 13.178787
freqMusee 2.624710e+33 4 15044.756375
freqTele 5.724458e+03 5 2.375601
freqLecture 2.991049e+18 5 70.401572
sensibilisation_culture -1.170016e+05 3 NaN
```

Annexe n°8 : 4eme et 5eme modèle logit supposant hétéroscédasticité des erreurs

<p>Call: <code>hetglm(formula = Theatre_binaire ~ freqMusee + freqTele + freqLecture + sensibilisation_culture freqMusee + freqTele, data = base_binaire1, family = binomial(link = "logit"), control = hetglm.control(method = "BFGS", maxit = c(500,25), hessian = FALSE, trace = FALSE), model = TRUE)</code></p>	<p>Call: <code>hetglm(formula = Theatre_binaire ~ freqMusee + freqTele + freqLecture + sensibilisation_culture freqTele, data = base_binaire1, family = binomial(link = "logit"), control = hetglm.control(method = "BFGS", maxit = c(500,25), hessian = FALSE, trace = FALSE), model = TRUE)</code></p>
<pre>Latent scale model coefficients (with log link): Estimate Std. Error z value Pr(> z) freqMusee1 6.622e-01 5.345e-01 1.239 0.215 freqMusee2 -9.726e-02 6.649e-01 -0.146 0.884 freqMusee3 -2.631e-01 3.781e+03 0.000 1.000 freqMusee4 -1.035e-04 4.058e+01 0.000 1.000 freqTele1 -1.192e+01 1.066e+00 -11.190 <2e-16 *** freqTele2 -3.576e-01 8.361e-01 -0.428 0.669 freqTele3 5.783e-01 5.746e-01 1.006 0.314 freqTele4 -1.939e-02 2.171e+04 0.000 1.000 freqTele5 -5.992e+00 1.685e+18 0.000 1.000 --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Log-likelihood: -163.7 on 27 Df LR test for homoskedasticity: 12.6 on 9 Df, p-value: 0.1816 Dispersion: 1 Number of iterations in BFGS optimization: 68</pre>	<pre>Latent scale model coefficients (with log link): Estimate Std. Error z value Pr(> z) freqTele1 -9.696e+00 2.072e+02 -0.047 0.963 freqTele2 7.499e-02 1.139e+00 0.066 0.948 freqTele3 2.581e-01 7.245e-01 0.356 0.722 freqTele4 1.272e-01 5.606e-01 0.227 0.821 freqTele5 -5.525e+00 1.797e+04 0.000 1.000 --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Log-likelihood: -164.4 on 23 Df LR test for homoskedasticity: 11.19 on 5 Df, p-value: 0.04779 Dispersion: 1 Number of iterations in BFGS optimization: 55</pre>
<pre>> vif(logit_homo3) GVIF Df GVIF^(1/(2*Df)) freqMusee 7.418714e+38 4 72242.219560 freqTele 1.847983e+02 5 1.685271 freqLecture -1.178001e+01 5 NaN sensibilisation_culture 1.350470e+02 3 2.265066</pre>	<pre>> vif(logit_homo4) GVIF Df GVIF^(1/(2*Df)) freqMusee 8.747929e+05 4 5.530166 freqTele 4.761221e+01 5 1.471539 freqLecture 2.261664e+03 5 2.164925 sensibilisation_culture 1.369401e+06 3 10.537925</pre>

Annexe n°9 : Sélection de variables modèle multinomial méthodes forward et backward

<p>Modèle forward</p> <pre>Start: AIC=938.61 somme4 ~ base1\$participation_acti_univ + base1\$sensibilisation_culture + base1\$budget + base1\$boursier + base1\$travail + base1\$formation + base1\$domaineEtude + base1\$genre Call: polr(formula = somme4 ~ base1\$participation_acti_univ + base1\$sensibilisation_culture + base1\$budget + base1\$boursier + base1\$travail + base1\$formation + base1\$domaineEtude + base1\$genre, data = base1, method = c("logistic")) Coefficients: base1\$participation_acti_univ1 base1\$participation_acti_univ2 base1\$sensibilisation_culture1 base1\$sensibilisation_culture2 base1\$sensibilisation_culture3 0.25446379 0.34285087 1.36091785 0.59508503 1.23750507 base1\$budget1 base1\$budget2 base1\$budget3 base1\$budget4 base1\$budget5 1.70428455 2.02550839 2.32292547 3.15811273 1.51989373 base1\$budget6 base1\$boursier1 base1\$travail1 base1\$formation2 base1\$formation3 1.52331886 -0.19890590 -0.06300551 17.94610032 0.61359884 base1\$formation4 base1\$formation5 base1\$formation6 base1\$formation7 base1\$domaineEtude1 1.68088628 3.46647155 1.57898920 2.29450943 0.25105900 base1\$domaineEtude10 base1\$domaineEtude11 base1\$domaineEtude12 base1\$domaineEtude13 base1\$domaineEtude2 -15.32681407 0.59813747 -0.06294742 0.84616472 -0.35158599 base1\$domaineEtude3 base1\$domaineEtude4 base1\$domaineEtude5 base1\$domaineEtude6 base1\$domaineEtude7 0.06617585 0.51713061 0.49288843 0.69645066 -0.38248371 base1\$domaineEtude8 base1\$domaineEtude9 base1\$genre1 1.03139585 0.80080607 0.33595828</pre>	<p>Modèle backward</p>
--	-------------------------------


```

Step: AIC=916.13
some4 ~ base1$sensibilisation_culture + base1$budget + base1$formation

              Df    AIC
<none>              916.13
- base1$formation    6  921.50
- base1$sensibilisation_culture  3  933.31
- base1$budget        6  933.48

Call:
polr(formula = some4 ~ base1$sensibilisation_culture + base1$budget +
      base1$formation, data = base1, method = c("logistic"))

Coefficients:
base1$sensibilisation_culture1 base1$sensibilisation_culture2 base1$sensibilisation_culture3
              1.3065664              0.6424425              1.2052791
base1$budget3              base1$budget4              base1$budget5
              2.2625899              3.0803193              1.8615316
base1$formation3              base1$formation4              base1$formation5
              0.3686598              1.1194245              3.3333305
base1$budget1              base1$budget2
              1.5244743              1.8599431
base1$formation2              base1$formation7
              17.5739000              1.8614714

Intercepts:
1|2      2|3      3|4
2.693075 3.832119 5.048715

```

Annexe n°10 : modèle 1 : multinomial ordonné après procédure stepwise

> Estimation du modèle multinomial

	Value	Std. Error	t value	pvalue
sensibilisation_culture1	1.3065664	3.248346e-01	4.022251e+00	0.0001
sensibilisation_culture2	0.6424425	3.335087e-01	1.926314e+00	0.0541
sensibilisation_culture3	1.2052791	2.873600e-01	4.194317e+00	0.0000
budget1	1.5244743	5.789521e-01	2.633161e+00	0.0085
budget2	1.8599431	5.230043e-01	3.556268e+00	0.0004
budget3	2.2625899	5.329133e-01	4.245699e+00	0.0000
budget4	3.0803193	6.837077e-01	4.505316e+00	0.0000
budget5	1.8615316	9.570248e-01	1.945124e+00	0.0518
budget6	1.6898335	1.013132e+00	1.667931e+00	0.0953
formation2	17.5739000	4.280062e-07	4.105992e+07	0.0000
formation3	0.3686598	7.965842e-01	4.628008e-01	0.6435
formation4	1.1194245	7.182231e-01	1.558603e+00	0.1191
formation5	3.3333305	1.119350e+00	2.977916e+00	0.0029
formation6	1.1299926	7.209390e-01	1.567390e+00	0.1170
formation7	1.8614714	1.026705e+00	1.813053e+00	0.0698
1 2	2.6930754	8.970853e-01	3.002028e+00	0.0027
2 3	3.8321190	9.074235e-01	4.223077e+00	0.0000
3 4	5.0487154	9.205628e-01	5.484379e+00	0.0000

> Qualité de prévision

```

      1  2  3  4
1 53 27 22 14
2  3  3  6  3
3 19 24 23 18
4 12 23 35 57
> qualite<-((mc[1,1]+mc[2,2]+mc[3,3])/sum(mc))*100
> print(qualite)
[1] 23.09942

```

> Qualité d'ajustement

> R2_Mc_Fadden

[1] 0.217734

Annexe 11 : Modèle 2 estimé par vglm

> Hypothèse d'égalité des pentes

> 1-pchisq(deviance(fit2)-deviance(fit5),df=df.residual(fit2)-df.resi

[1] 1

> Estimation du modèle vglm

Call:

```
vglm(formula = somme4 ~ sensibilisation_culture + budget + formation,  
     family = cumulative(parallel = TRUE, reverse = TRUE), data = baseclasse3,  
     link = "logit")
```

Pearson residuals:

	Min	1Q	Median	3Q	Max
logitlink(P[Y>=2])	-3.986	-0.4155	0.2416	0.4782	1.604
logitlink(P[Y>=3])	-2.756	-0.5497	0.2488	0.7659	2.613
logitlink(P[Y>=4])	-2.408	-0.7866	-0.2459	0.7732	3.235

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept):1	-2.6930	0.9470	-2.844	0.004459 **
(Intercept):2	-3.8321	0.9589	-3.996	6.44e-05 ***
(Intercept):3	-5.0486	0.9715	-5.197	2.03e-07 ***
sensibilisation_culture1	1.3066	0.3183	4.105	4.05e-05 ***
sensibilisation_culture2	0.6425	0.3303	1.945	0.051751 .
sensibilisation_culture3	1.2053	0.2876	4.190	2.79e-05 ***
budget1	1.5242	0.5925	2.572	0.010099 *
budget2	1.8597	0.5395	3.447	0.000566 ***
budget3	2.2623	0.5481	4.127	3.67e-05 ***
budget4	3.0801	0.6873	4.482	7.40e-06 ***
budget5	1.8612	0.9613	1.936	0.052863 .
budget6	1.6895	1.0431	1.620	0.105301
formation2	15.8311	568.8645	NA	NA
formation3	0.3689	0.8317	0.443	0.657424
formation4	1.1196	0.7689	1.456	0.145381
formation5	3.3336	1.2122	2.750	0.005958 **
formation6	1.1302	0.7725	1.463	0.143467
formation7	1.8615	1.1053	1.684	0.092131 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Annexe 11.b: Hypothèse d'égalité des pentes avec vglm

```
> 1-pchisq(deviance(fit2)-deviance(fit5),df=df.residual(fit2)-df.residual(fit5))
[1] 1
```

Annexe 12 : Modèle cluster

```
Call:
glm(formula = cluster ~ participation_acti_univ + sensibilisation_culture +
    CSP_mere + budget + boursier + travail + formation + domaineEtude +
    genre + age, family = binomial(logit), data = base_cluster2)
```

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-5.743e-06	2.293e-06	2.408e-06	2.514e-06	7.742e-06

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	7.363e+01	2.963e+05	0	1
participation_acti_univ1	1.078e-01	4.476e+04	0	1
participation_acti_univ2	-1.025e-01	2.572e+05	0	1
sensibilisation_culture1	-1.623e-03	6.249e+04	0	1
sensibilisation_culture2	-1.092e-01	6.620e+04	0	1
sensibilisation_culture3	3.562e-02	5.578e+04	0	1
CSP_mere1 <	1.051e+00	1.555e+05	0	1
CSP_mere2	9.598e-01	1.208e+05	0	1
CSP_mere3	1.183e+00	1.291e+05	0	1
CSP_mere4	9.540e-01	1.203e+05	0	1
CSP_mere5	9.798e-01	1.412e+05	0	1
CSP_mere6	1.061e+00	1.745e+05	0	1
CSP_mere7	9.776e-01	1.331e+05	0	1
CSP_mere8	4.003e-01	1.320e+05	0	1
CSP_mere9	8.725e-01	1.556e+05	0	1
budget1	-4.875e-03	1.042e+05	0	1
budget2	1.640e-02	9.421e+04	0	1
budget3	1.903e-02	9.504e+04	0	1
budget4	-9.491e-02	1.209e+05	0	1
budget5	6.394e-02	1.930e+05	0	1
budget6	-5.856e-02	2.073e+05	0	1
boursier1	9.017e-02	4.437e+04	0	1
travail1	-5.074e-02	4.549e+04	0	1
formation2	-5.019e+01	4.058e+05	0	1
formation3	-4.804e+01	2.096e+05	0	1
formation4	-4.812e+01	1.953e+05	0	1
formation5	-4.788e+01	2.530e+05	0	1
formation6	-4.807e+01	1.984e+05	0	1
formation7	-4.799e+01	2.606e+05	0	1
domaineEtude1	-1.575e-01	7.409e+04	0	1
domaineEtude10	3.387e-02	3.792e+05	0	1
domaineEtude11	-1.284e-01	1.007e+05	0	1
domaineEtude12	-1.421e-01	8.119e+04	0	1
domaineEtude13	-7.808e-03	1.410e+05	0	1
domaineEtude2	4.413e-02	1.335e+05	0	1
domaineEtude3	-5.086e+01	1.377e+05	0	1
domaineEtude4	-8.746e-02	1.732e+05	0	1
domaineEtude5	-2.346e-01	1.109e+05	0	1
domaineEtude6	-9.245e-02	1.052e+05	0	1
domaineEtude7	-2.083e-03	2.157e+05	0	1
domaineEtude8	4.067e-02	3.710e+05	0	1
domaineEtude9	-5.105e-01	1.285e+05	0	1
genre1	-3.691e-02	4.489e+04	0	1
age	8.943e-03	1.320e+04	0	1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 4.3541e+01 on 341 degrees of freedom
Residual deviance: 2.1250e-09 on 298 degrees of freedom
AIC: 88

Annexe 13 : Répartition des classes et tests des seuils

1. Absence ou très faible fréquence de consommation culturelle	Score de 13 à 20
2. Faible consommation culturelle	21 à 26
3. Fréquence de consommation modérée	27 à 32
5. Consommation culturelle conséquente	32 à 39

> Tests des seuils

1 2	1.51387284	1.026971e+00	1.474114e+00	0.1405
2 3	3.96564693	1.045121e+00	3.794438e+00	0.0001
3 4	7.18513877	1.093760e+00	6.569209e+00	0.0000

Annexe 14 : vérification de l'hypothèse d'égalité des pentes avec clm

Tests of nominal effects

formula: somme4 ~ base1\$participation_acti_univ + base1\$sensibilisation_culture + base1\$budget + base1\$boursier + base1\$travail + base1\$formation + base1\$domaineEtude + base1\$genre

	Df	logLik	AIC	LRT	Pr(>Chi)
<none>		-433.30	938.61		
base1\$participation_acti_univ					
base1\$sensibilisation_culture					
base1\$budget					
base1\$boursier	2	-432.92	941.84	0.77039	0.6803
base1\$travail	2	-432.39	940.79	1.81886	0.4028
base1\$formation					
base1\$domaineEtude					
base1\$genre	2	-433.21	942.43	0.18217	0.9129

Annexe 15: Estimation modèle 3 avec formation (vglm)

```
Call:
glm(formula = somme4 ~ basevglm0$sensibilisation_culture + basevglm0$budget +
  basevglm0$formation, family = binomial(link = "logit"), data = basevglm0)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.09577   0.00012   0.48563   0.65121   1.86669

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -3.2474     1.1375  -2.855 0.004306 **
basevglm0$sensibilisation_culture1  2.1635     0.4890   4.424 9.67e-06 ***
basevglm0$sensibilisation_culture2  0.6438     0.4110   1.567 0.117198
basevglm0$sensibilisation_culture3  1.8136     0.4004   4.529 5.93e-06 ***
basevglm0$budget1    1.8395     0.6991   2.631 0.008512 **
basevglm0$budget2    2.1578     0.6282   3.435 0.000592 ***
basevglm0$budget3    3.1444     0.6762   4.650 3.31e-06 ***
basevglm0$budget4    3.5257     1.0194   3.459 0.000543 ***
basevglm0$budget5   18.4992    1893.2436   0.010 0.992204
basevglm0$budget6    1.8290     1.3837   1.322 0.186228
basevglm0$formation2 17.0252    3956.1804   0.004 0.996566
basevglm0$formation3  -0.4602     0.9766  -0.471 0.637462
basevglm0$formation4  1.0042     0.9133   1.100 0.271504
basevglm0$formation5 18.7729    1307.7737   0.014 0.988547
basevglm0$formation6  1.2720     0.9252   1.375 0.169202
basevglm0$formation7 16.4052    1503.3549   0.011 0.991293
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 365.43  on 333  degrees of freedom
Residual deviance: 288.62  on 318  degrees of freedom
AIC: 320.62

Number of Fisher Scoring iterations: 16
```

PARTIE 5 : Bibliographie

BVA, "Étude sur la représentation culturelle", *Rapport d'études qualitatives*, 03/2020, p.17.

https://www.modernisation.gouv.fr/sites/default/files/fichiers-attaches/democratisation-culturelle_etude_mars2016.pdf

Caprais J., "Que sont les 7 arts", *Tout comment*, mis à jour le 21/03/2017.

<https://education.toutcomment.com/article/quels-sont-les-7-arts-2787.html>

Desouches O., "La culture : un bilan sociologique", *Idées économiques et sociales*, 01/2014 (n° 175), pp. 53-60.

<https://www.cairn.info/revue-idees-economiques-et-sociales-2014-1-page-53.htm>

LAROUSSE, définition de la culture, consultable en ligne :

<https://www.larousse.fr/dictionnaires/francais/culture/21072>

Vadelorge L., "Où va l'histoire culturelle ?", *Ethnologie française*, 02/2006 (Vol. 36), pp. 357-359.

<https://www.cairn.info/revue-ethnologie-francaise-2006-2-page-357.htm>

Desouches O., "La culture : un bilan sociologique", *Idées économiques et sociales* 2014/1 (N° 175), pp. 53-60

<https://www.cairn.info/revue-idees-economiques-et-sociales-2014-1-page-53.htm#re11no64> (consulté le 25/ 02/2020)

PARTIE 6 : Table des matières

Abstract	1
Résumé	2
PARTIE 1 : Analyse économique	4
I- Introduction	4
II- Analyse économique des variables	7
Les déterminants socio-économiques des pratiques culturelles	7
1- Les déterminants sociologiques	7
2- Déterminants économiques	9
Les déterminants extrinsèques des pratiques culturelles	10
PARTIE 2 : Analyse économétrique	12
I- Nettoyage de la base de données	12
Clarification de la base pour faciliter la lecture	12
Détection des valeurs atypiques	14
Statistiques descriptives	15
Liens entre les variables	18
1- Détection des variables dépendantes	18
2- Analyse des correspondances multiples	19
II- Modèle binaire	22
Hypothèses d'un modèle binaire	23
Estimation de modèles binaires sous hypothèse d'homoscédasticité des erreurs	24
Vérification de l'homoscédasticité des erreurs	28
III- Modèle multinomial ordonné	33
IV- Modèle multinomial non ordonné	46
PARTIE 3 : Conclusion et discussion des résultats	48
PARTIE 4 : Annexes	52
Annexe n°1 : Dictionnaire des variables	52
Annexe n°2 : Histogrammes de la variable 'age' avec et sans les valeurs atypiques	56
Annexe n°3 : Statistiques descriptives des pratiques culturelles des répondants	57
Annexe n°4 : Statistiques descriptives des caractéristiques des répondants	58
Annexe n°5 : Méthode stepwise de choix des variables pour le modèle binaire	58
Annexe n°6 : Deuxième modèle logit (sans les 4 données influentes) + tests	60
Annexe n°7 : 2eme et 3eme modèles logit supposant hétéroscédasticité des erreurs	61
Annexe n°8 : 4eme et 5eme modèle logit supposant hétéroscédasticité des erreurs	62

Annexe n°9 : Sélection de variables modèle multinomial méthodes forward et backward	62
Annexe n°10 : modèle 1 : multinomial ordonné après procédure stepwise	63
Annexe 11 : Modèle 2 estimé par vglm	64
Annexe 11.b: Hypothèse d'égalité des pentes avec vglm	65
Annexe 12 : Modèle cluster	65
Annexe 13 : Répartition des classes et tests des seuils	66
Annexe 14 : vérification de l'hypothèse d'égalité des pentes avec clm	66
Annexe 15: Estimation modèle 3 avec formation (vglm)	67
PARTIE 5 : Bibliographie	68
PARTIE 6 : Table des matières	69
PARTIE 7 : Table des illustrations	71

PARTIE 7 : Table des illustrations

1. Tableaux

Tableau 1: Question n°1 et 2 du questionnaire : variables à expliquer	7
Tableau 2: Visuel et test de valeurs atypiques de la varigé 'Age'	12
Tableau 3 : Répartition des pratiques culturelles des répondants dans chaque modalité	13
Tableau 4: Répartition des caractéristiques des répondants dans chaque modalité	14
Tableau 5: Représentation graphique des individus et inertie associé aux axes	17
Tableau 8: Modèle binaire sous hypothèse d'homoscédasticité des erreurs	22
Tableau 9 : Résultat test VIF modèle binaire	23
Tableau 10: Taux de précision, de spécificité et de sensibilité du modèle logit	25
Tableau 12: Modèle logit avec erreurs homoscédastiques et toutes les variables coupables	27
Tableau 13: Modèle logit final : avec erreurs homoscédastiques, et odd ratios	28
Tableau 15: Répartition des fréquences pour le logit ordonné	34
Tableau 16: Modèle 1 global	34
Tableau 17: Procédure stepwise	39
Tableau 18: Modèle 2 vglm	40
Tableau 19: Odd-ratios du modèle 2	41
Tableau 20: Estimation du modèle 3 sans formation	44
Tableau 21: Test d'homoscédasticité des erreurs avec hetglm	45
Tableau 22: Points forts et points faibles de nos modèles	49

2. Figures

Figure 1 : Graphique du carré des liaisonsFigure	18
Figure 2 : Représentation des variables dans le plan 1-2	19
Figure 3 : Graphique des résidus	25
Figure 4 : Dendrogramme et clusters de l'ACM	33
Figure 5: Résidus du modèle 2	42