



Les Déterminants de l'abus d'alcool

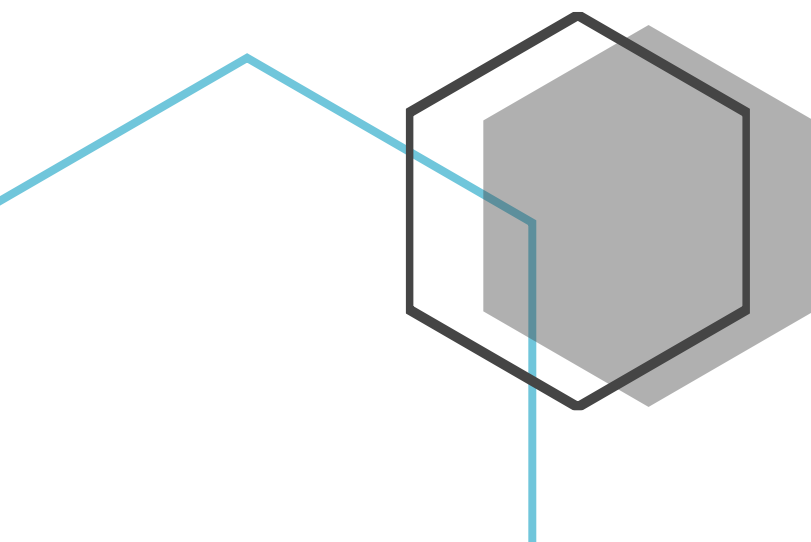
SEZESTRE Emilien

MI IEE

Professeure : KURTBEGU Enareta

Année universitaire : 2021 - 2022

Ce rapport s'intéresse notamment aux déterminants de l'abus d'alcool, via la base de donnée Wooldridge. Nous verrons notamment, le rôle prévalent de l'alcoolisme des parents, de l'état de santé et de l'éducation.



Les Déterminants de l'abus d'alcool

SEZESTRE Emilien

I – Introduction

L'alcool est un fléau des temps moderne, au même titre que la drogue. En effet, son coût social serait d'environ 120 milliard d'euros en 2010 selon un rapport de l'observatoire français des drogues et des toxicomanies.¹ La consommation d'alcool est notamment responsable de 49 000 décès par différentes maladies.²

Dans ce cadre l'alcool, fut la cible de nombreuses politiques publiques, ayant pour objectif de limiter la consommation d'alcool. La plus connus étant la prohibition aux Etats-Unis, une période allant de 1920 à 1933 qui interdit la vente, la fabrication et le transport des boissons alcoolisées³.

L'alcool est aujourd'hui encore sujet à beaucoup de politique en France. Sa consommation est très encadrée avec un arsenal législatif, relativement répressif (interdit aux moins de 18 ans, limité au volant...), il y a également une stratégie de prévention.

Nous allons, au cours de ce rapport se concentrer sur l'abus d'alcool et non sur la consommation d'alcool, ce qui a différentes implications en termes de définition. En effet, l'abus d'alcool prend 3 formes différentes selon l'ouvrage « Binge Drinking in Young Adults: Data, Definitions, and Determinants. » :

- Le « Binge drinking » : qui se définit comme étant, une alcoolisation ponctuelle forte (correspondant à 5 verres pour les hommes et 4 verres pour les femmes en moins de 2 heures) qui survient plus d'une fois en 6 mois.

- L'abus d'alcool : est un comportement inadapté vis-à-vis des boissons alcoolisées, manifesté par : - une consommation récurrente entraînant l'incapacité de travailler, d'aller à l'école, ou amenant à être exclue de l'école ou du lieu de travail.

Les effets de l'alcool sont diminués par une consommation récurrente, la personne souffre de syndrome de sevrage et de problème juridique lié à la consommation d'alcool

- La « dépendance » : cela correspond à l'alcoolisme, ce qui est une dépendance liée à l'alcool manifesté par d'autres comportements : la personne boit plus souvent en plus grande quantité. L'individu réduit également ses activités sociales, récréatives et ses occupations à cause de l'alcool. Pour finir, la consommation de boissons alcoolisées n'est pas modifiée malgré des problèmes de santé liés à la consommation d'alcool.⁴

A – Problématique

Nous avons donc vu qu'il y avait différentes composantes à prendre en compte concernant l'abus d'alcool. De plus, nous avons vu l'intérêt économique de ce sujet.

Nous avons donc vu que l'abus d'alcool entraînait des problèmes de santé et donc, il semble intéressant de se demander quels sont les déterminants de l'abus d'alcool ?

En effet, cette question nous permettra de connaître, quels sont les réels facteurs entraînant un mésusage des boissons alcooliques, afin d'agir plus efficacement en termes de politique publique pour réduire les comportements à risque.

L'objectif final étant de réduire le coût social liés à la consommation d'alcool.

¹ KOPP Pierre (2015). Le coût social des drogues en France. Paris : Observatoire français des drogues et des toxicomanies, 10 p

² IRDES (2019). La politique de lutte contre l'alcoolisme. Paris, 250 p

³ Prohibition. (2021, 17/12/2021). Dans Wikipédia. <https://fr.wikipedia.org/wiki/Prohibition>

⁴ COURTNEY Kelly, POLICH John (2009), Binge Drinking in Young Adults: Data, Definitions, and Determinants. Psychological Bulletin, Volume 135, n°1, p151

Donc nous pouvons dire, vis-à-vis de notre base de données que l'abus d'alcool et la « dépendance » rentre dans le cadre de la variable expliqué de notre modèle. Désormais, nous allons expliquer pourquoi la pratique du binge drinking peut être également intégré comme étant un abus d'alcool.

En effet, selon l'INSERM, cette pratique pourrait notamment avoir de graves effets sur la santé : « Les résultats de nombreux travaux sur les conséquences cérébrales du binge drinking sont alarmants (voir pour revue récente : Lisdahl et coll., 2013a). Les études d'imagerie cérébrale par résonance magnétique fonctionnelle (IRMf) ont mis en évidence de nombreuses atteintes morphologiques et fonctionnelles tant au niveau de la substance grise (neurones) que blanche »⁵ donc, cette pratique est belle et bien dangereuse pour la santé et peut donc être incluse dans notre définition.

B – Annonce du plan

Donc nous allons voir dans un premier temps, les principales revues de littératures que nous avons utilisé au cours de notre développement.

Ensuite, nous expliquerons, avec notre revue de littérature, les raisons de nos choix de variable.

Puis nous nous développerons vis-à-vis de notre base de données et ferons nos statistiques descriptives.

Pour finir, nous commenterons les différents résultats obtenus.

⁵ INSERM. Effets néfastes de l'alcool : impact du binge drinking sur le cerveau. p187-213

Sommaire :

I. Introduction	1
A - Problématique	1
B – Annonce du plan	2
II. Revue de la littérature	4
III. Modèle économique	4
A – Les variables	5
1- Alcoolisme des parents	6
2- Nombre d'année d'étude	6
3- L'âge	6
4- Etat de santé	6
5- Statut salarial	6
6- Milieu urbain	6
B - Hypothèses	5
IV - Méthodologie	4
A - Modèle Econométrique	5
B - Données	5
1- Présentation de la base de données	6
2- Description des variables	6
3- Statistiques descriptives	6
V - Résultats	4
A - Régressions	5
B - Tests	5
1- Test de Wald	6
2- Tableau de prévision	6

3- R ² de McFadden	6
4- Hosmer-Lemeshow.....	6

VI. Conclusions & Discussions	4
--	----------

VII. Limites	4
---------------------------	----------

VIII. Bibliographie	4
----------------------------------	----------

IX. Annexe	4
-------------------------	----------

A - Code Stata	5
----------------------	---

B - Autres.....	5
-----------------	---

II. Revue de la littérature

Dans cette partie, nous allons rapidement présenter les principaux ouvrages qui nous ont servi pour nous aiguiller, lors de la construction de notre modèle économique et économétrique. Nous avons donc 5 articles, nous expliquerons pour chacun, leurs apports principaux et leurs méthodologies.

La première étude se nomme « Déterminants de la consommation de boissons alcoolisées dans l'Étude NutriNet-Santé », c'est une enquête commanditée par santé publique France. C'est une étude qui se déroule en France et qui étudie uniquement la consommation d'alcool, et qui établit qu'une consommation supérieure à 10g/j correspond, pour un individu, équivaut à un verre de vin par jour, ce qui correspond à de l'abus d'alcool. C'est une étude réalisée sur 29 566 individus, 6 828 hommes et 22 738 femmes. De plus, cette étude avait pour objectifs de déterminer des actions de préventions pour réduire la consommation d'alcool, et communiquer de nombreuses informations sur les consommateurs d'alcool.

Cette étude met en lumière différents facteurs comme étant des déterminants de la consommation d'alcool. On y apprend notamment que le genre détermine énormément la consommation d'alcool, or, notre base de données ne nous donne pas le sexe des individus. Le rôle de l'âge est également pointé dans cette étude.

La seconde étude se nomme : « Les déterminants de la consommation d'alcool au Canada : Facteurs socio-démographiques et économiques », elle est réalisée par Marie-Rachelle Narcisse. Cette étude se fait au Canada et se base sur un échantillon de 12 155 individus. Dans cette étude, on met en avant l'effet du genre et de l'âge sur la consommation d'alcool.

Nous nous sommes notamment servis de cette étude pour parler du rôle du nombre d'année d'étude sur la consommation d'alcool.

Notre troisième source est la thèse de Marie Zins, qui se nomme « Déterminants sociaux des consommations d'alcool dans la cohorte gazel ». C'est une étude menée sur environ 20 000 individus, 15 000 hommes et 5 000 femmes employés d'EDF-GDF, et étudie différents facteurs qui pourraient déterminer la consommation d'alcool. Dans ce cadre, M. Zins s'intéresse notamment à l'environnement familial, de travail et les évolutions qui amène, plus ou moins, à consommer de l'alcool.

Nous nous sommes servis de cette étude pour parler des limites de notre travail que nous évoquerons à la fin de ce rapport dans notre conclusion et nos limites. Nous parlerons notamment de problème lié à notre base de données et de la manière dont ont été récupérées nos différentes données. Cette étude fut aussi très utile pour trouver d'autres sources de littérature, parlant des déterminants de l'abus d'alcool.

La quatrième étude, est un article dans le périodique « Devenir », co-écrit par Hiram E. Fitzgerald et Robert A. Zucker. Il se prénomme « Effets à court et à long terme de l'alcoolisme parental sur les enfants » et parle notamment des enfants de parents alcooliques aux États-Unis. Cet article cherche à expliquer pourquoi les enfants de parents alcooliques sont plus à même d'adopter des mauvais comportements vis-à-vis de l'alcool.

Selon nous, cet article est important car, pour nous, le fait d'avoir un père ou une mère alcoolique est la variable la plus explicative de notre modèle. En effet, c'est une des raisons les plus citées lorsque l'on s'intéresse à des ouvrages vis-à-vis des déterminants de l'alcoolisme.

La cinquième et dernière principale source de notre étude, est également un article d'une revue, la « Psychological bulletin », qui est une revue qui s'intéresse uniquement à la recherche dans le domaine de la psychologie. C'est une étude qui cherche à donner des précisions sur les définitions concernant le champ de recherche sur l'alcool. Il parle également des impacts que l'abus d'alcool a sur la santé.

Cet article nous permet de différencier les différentes définitions vis-à-vis de l'abus d'alcool, et à mieux comprendre la forme que prennent ces différentes formes de consommation. Nous nous sommes aussi servis de cet article dans nos limites car cet article parle notamment du rôle du genre.

III. Modèle économique

A – Les Variables

Notre revue de littérature a pour objectif de justifier la présence des différentes variables que nous allons utiliser pour faire notre modèle économique et économétrique. Nous allons donc nous intéresser aux différents déterminants de l'alcoolisme.

Nous allons, au cours de cette partie, parler des variables, et justifier leurs ajouts ou leurs suppressions de notre modèle.

I- Alcoolisme des parents

Tout d'abord, nous allons parler de la variable : « mothalc » et « fathalc », qui prend la valeur « 1 » si la mère ou le père est alcoolique. Cette variable est sans doute la variable la plus pertinente pour notre étude, et nous expliquerons pourquoi dans les limites de notre étude.

Pour corroborer le fait que cette variable est un déterminant de l'abus d'alcool, nous avons choisi une étude s'appelant « Effets à court et à long terme de l'alcoolisme parental sur les enfants », dans cette étude on y apprend notamment qu'il y a 3 raisons principales pour lesquelles des enfants ayant des parents alcooliques, risquent de reproduire ces comportements plus tard. Tout d'abord, parce que « Les enfants de parents alcooliques sont familiers avec un plus large éventail de boissons alcooliques à un âge plus précoce et ils développent des schèmes de consommation plus tôt dans leur vie. »⁶, donc les enfants de parents alcooliques boivent, en moyenne, de l'alcool plus tôt. Ensuite, il est dit que les enfants ont des attentes positives concernant la consommation d'alcool. Pour finir, il est dit que : « Chez les enfants de parents alcooliques de certains groupes ethniques, il existe un risque accru d'abus d'alcool, à cause de l'interaction entre les attentes liées à l'alcool et celles entretenues dans le groupe ethnique en cause. »⁷.

Via cette dernière citation, il semble qu'une autre variable est potentiellement intéressante, la variable « black » qui est une variable codée « 1 » si la personne est de couleur noire. Cependant, par manque d'étude sérieuse au sujet de cette variable, nous avons décidé de ne pas l'utiliser. Nous avons aussi retiré cette variable, car elle viendrait complexifier notre modèle économétrique.

2- Nombre d'année d'étude

Désormais, nous allons parler de la variable : « educ » qui correspond au nombre d'année d'éducation. Selon une étude, « » : cette variable est assez spéciale car il semble que les personnes n'ayant pas fait d'étude du tout sont les personnes qui sont le plus susceptible de boire de l'alcool, ensuite, si on considère seulement les individus ayant fait des études, les résultats sont contre intuitif, en effet, la consommation augmenterait avec le nombre d'année d'études.

Cependant l'étude concernée semble conclure que : « les coefficients associés à ces variables ne sont guère significatifs. La variable socio-démographique « éducation » n'aurait donc pas d'effet sur la probabilité de boire »⁸ donc on pourrait exclure cette variable. Cependant une autre étude viendrait valider le fait selon lequel, plus on a un niveau de diplôme élevé, plus on consomme de l'alcool. Cependant, nous ne sommes pas dans un cadre où cette consommation est forcément abusive. En effet, « Une consommation d'alcool plus élevée est plus fréquente chez les personnes ayant des revenus élevés. Chez les hommes, elle est plus fréquente chez les cadres. »⁹

Donc il semble que le niveau d'étude fait augmenter la consommation d'alcool. Cependant il est dit plus tard que : « Le Baromètre cancer 2010, publié récemment, montre que les moins diplômés et les plus

⁶ Fitzgerald, Hiram E, et Robert A. Zucker. « Effets à court et à long terme de l'alcoolisme parental sur les enfants », *Devenir*, vol. 14, no. 2, 2002, pp. 169-182.

⁷ Fitzgerald, Hiram E, et Robert A. Zucker. « Effets à court et à long terme de l'alcoolisme parental sur les enfants », *Devenir*, vol. 14, no. 2, 2002, pp. 169-182.

⁸ NARCISSE, Marie-Rachelle (2000). Les déterminants de la consommation d'alcool au Canada : Facteurs socio-démographiques et économiques [en ligne]. Montréal : Faculté des arts et des sciences - Département de Sciences Économiques. Disponible sur internet : <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/866> (consulté le 25/11/2021) p58

⁹ Mathilde Touvier, Nathalie Druet-Pecollo, Pilar Galan, Serge Hercberg, Paule Latino-Martel (2013). Déterminants de la consommation de boissons alcoolisées dans l'Étude NutriNet-Santé, France. *Bulletin Épidémiologique Hebdomadaire*, 2013, n°. 16-17-18, p.203

âgés sont moins bien informés sur les risques liés à l'alcool, en particulier les risques de cancers, et qu'il y a une forte prévalence d'opinions qui relativisent ces risques »¹⁰.

Donc il semble que, la consommation d'alcool augmente avec le niveau de diplôme, cependant, l'abus d'alcool semble être plus présent chez les personnes ayant un niveau d'étude plus bas. Pour résumer, plus gens font des études, plus il consomme de l'alcool, cependant leur consommation serait plus « responsable ». Cependant, une personne moins diplômée aurait moins de chance de consommer de l'alcool mais, parmi ces gens-là, ceux qui boivent de l'alcool ont plus de chance d'être des alcooliques. Donc, nous pouvons voir que cette variable est intéressante, nous avons donc décidé de l'intégrer à notre modèle économétrique.

3- L'âge

Ensuite, nous allons parler de la variable « age », il semble logique que l'abus d'alcool à plus de chance de survenir lorsque l'on vieillit. Nous allons voir que cette opinion préconçue est assez vraie dans les faits. En effet, selon une étude de santé publique France, : « Notre résultat concernant l'augmentation de la consommation d'alcool avec l'âge (avec une probabilité de dépasser 10 g/j maximale après 55 ans) est en accord avec l'observation d'une augmentation transitoire de la prévalence de forts consommateurs dans la période environnant le départ à la retraite dans l'étude longitudinale réalisée avec la cohorte française Gazel (10 023 hommes et 2 361 femmes dont la consommation d'alcool a été estimée annuellement et suivie cinq ans avant et après la retraite). »¹¹.

Cette étude permet d'appuyer le fait que, les personnes après 55 ans et après qu'ils aient pris leurs retraite, ont plus de chance d'abuser de l'alcool. En effet, selon cette même étude, le « le seuil de 10 g d'alcool/j, qui correspond environ à un verre standard, afin d'obtenir une variable binaire permettant de réaliser une régression logistique simple »¹² donc nous pouvons dire qu'une personne qui boit en moyenne plus d'un verre par jour, est quelqu'un qui abuse de l'alcool.

L'âge est en réalité une variable qui est beaucoup plus complexe à prendre en compte car, selon l'âge, l'individu a plus de chance d'abuser de l'alcool mais selon une définition différente. Ainsi, les jeunes auront plus tendance à s'alcooliser ponctuellement de manière plus importante, en effet, selon une étude : « La prévalence du mésusage de l'alcool varie suivant l'âge et les générations. Si les consommations ponctuelles importantes sont fréquentes chez les adolescents et jeunes adultes, la consommation quotidienne et le risque d'addiction sont particulièrement fréquents chez les adultes d'âge moyen. »¹³, donc on voit qu'à chaque étape de sa vie, l'individu a des probabilités assez semblable d'abuser de l'alcool, cependant, il semble que la forme dont prend cet abus, diffère selon l'âge.

Nous verrons dans la suite de notre analyse, ce que cette complexité implique en termes de modèle économétrique.

¹⁰ Mathilde Touver, Nathalie Druesne-Pecollo, Pilar Galan, Serge Hercberg, Paule Latino-Martel (2013). Déterminants de la consommation de boissons alcoolisées dans l'Étude NutriNet-Santé, France. Bulletin Épidémiologique Hebdomadaire, 2013, n°. 16-17-18, p.203

¹¹ Mathilde Touver, Nathalie Druesne-Pecollo, Pilar Galan, Serge Hercberg, Paule Latino-Martel (2013). Déterminants de la consommation de boissons alcoolisées dans l'Étude NutriNet-Santé, France. Bulletin Épidémiologique Hebdomadaire, 2013, n°. 16-17-18, p.203

¹² Mathilde Touver, Nathalie Druesne-Pecollo, Pilar Galan, Serge Hercberg, Paule Latino-Martel (2013). Déterminants de la consommation de boissons alcoolisées dans l'Étude NutriNet-Santé, France. Bulletin Épidémiologique Hebdomadaire, 2013, n°. 16-17-18, p.200

¹³ G. Airagnes, C. Lemogne, F. Limosin, M. Zins, M. Dosquet, N. Hoertel, M. Marie-Krause (2019). L'abus d'alcool à différents âges de la vie : données épidémiologiques récentes. French Journal of Psychiatry, volume 1, p.40-41.

4- Etat de santé

Désormais, nous allons parler des variables « exhealth », « vghealth », « goodhealth », « fairhealth » qui correspondent à l'état de santé de l'individu ; « exhealth » correspond à un état de santé excellent et « fairhealth » correspondant à une santé bonne. Nous allons voir que plus l'état de santé d'un individu se dégrade, plus la probabilité que l'individu consomme de l'alcool est forte.

Tout d'abord, il faut prendre en compte, qu'un individu qui abuse de l'alcool, est « naturellement » amené à avoir un état de santé moins bon, car de nombreuses maladies sont liées à la consommation d'alcool. Via « ameli.fr » le site de l'assurance maladie en France, on y apprend notamment qu'en 2015, 41 080 décès sont attribuables à l'alcool, donc 39% sont dus à des cancers (foie, œsophage...), 24% des maladies cardiovasculaires (hypertension artérielle, AVC...), 16,5% des maladies digestives (stéatose, hépatite aiguë alcoolique...)¹⁴. Donc, il semble apparaître un lien de causalité entre l'abus d'alcool et un état de santé moyen.

Ensuite, nous allons voir que l'abus d'alcool s'accompagne de d'autres comportements « à risque » c'est-à-dire, qui sont associés ou qui entraînent un mauvais état de santé. En effet, nous allons voir que le surpoids est corrélé positivement avec l'abus d'alcool, il en est de même pour les individus fumeurs ou ayant été ancien fumeur. On observe également que les individus mangeant des aliments avec des apports caloriques élevés sont plus à même d'être alcoolique. C'est ce qu'affirme l'étude de santé publique France qui dit notamment que :

« La probabilité de consommer au moins 10 g d'alcool/j [...] est plus élevée chez les fumeurs ou anciens fumeurs et plus faible chez les personnes maigres comparées à celles de poids normal. [...] Chez les hommes, la consommation d'alcool est plus élevée chez les sujets en surpoids ou obèses, alors que chez les femmes obèses, la consommation diminue par rapport à celles de poids normal. [...] Les relations entre consommation d'alcool et consommations alimentaires sont similaires chez les hommes et les femmes : globalement, des consommations d'alcool plus élevées sont associées à des apports plus élevés en énergie (totale et sans l'alcool), viande rouge et abats, charcuterie et jambon, poisson et fruits de mer, matières grasses et sauces, gâteaux, biscuits et viennoiseries, et fruits oléagineux et produits de l'apéritif »¹⁵

Si l'on devait faire une remarque sur cette citation, c'est notamment qu'il semble que les femmes en surpoids, et donc en mauvaise santé, consomment moins d'alcool que les femmes qui ne sont pas en surpoids. Cependant, comme nous n'avons pas la variable « sexe » et que cette spécificité est très particulière, nous ne la prendrons pas en compte.

Donc ces deux arguties nous permettent d'affirmer qu'il y a bien un lien de causalité entre un mauvais état de santé et l'abus d'alcool.

5- Statut salarial

Ensuite, la variable « status » qui est codée 1 si la personne est « out of workforce », ce qui veut dire, hors de la population active (c'est-à-dire, selon l'INSEE, « est définie aux fins de la Comptabilité nationale comme comprenant toutes les personnes des deux sexes au-dessus d'un âge déterminé qui fournissent, durant une période de référence spécifiée, la main-d'œuvre nécessaire aux activités de production (telles que retenues dans la définition de la production du système de comptabilité nationale, SCN). Elle comprend toutes les personnes qui remplissent les conditions pour être considérées comme personnes pourvues d'un emploi

¹⁴ ASSURANCE MALADIE (2021). Les maladies liées à l'alcool. Paris : Ameli, disponible sur internet : <https://www.ameli.fr/maine-et-loire/assure/sante/themes/alcool-sante/maladies-liees-alcool> (consulté le 26/12/2021)

¹⁵ Mathilde Touvier, Nathalie Druesne-Pecollo, Pilar Galan, Serge Hercberg, Paule Latino-Martel (2013). Déterminants de la consommation de boissons alcoolisées dans l'Étude NutriNet-Santé, France. Bulletin Épidémiologique Hebdomadaire, 2013, n°. 16-17-18, p.200

(salariés ou non-salariés) ou comme chômeurs. »¹⁶) il faut aussi prendre en compte que les étudiants sont considérés comme hors de la population active.

La variable est codée 2 lorsque la personne est au chômage et 3 lorsqu'elle a un emploi. Concernant cette variable, nous pouvons dire, selon une étude que : « Nevertheless, the unemployed drinkers drank more alcohol than those in employment, even after standardization for both age and social class (27.9 units vs 20.7 units per week). Binge drinking was common in both groups, but the proportion was higher among the unemployed (58.8% of the unemployed reported drinking more than eight units in any day in the previous week compared to 33.5% among the full-time workers) »¹⁷ donc, nous pouvons dire que les personnes qui sont au chômage ont plus de chance d'abuser de l'alcool, que les personnes qui travaillent, nous verrons dans notre étude si cela se valide. Cependant nous n'avons toujours pas d'information concernant les individus qui sont hors de la population active, ce qui concerne majoritairement, des étudiants, ou des retraités.

Cependant, notre base de données comporte uniquement des individus ayant entre 25 et 59 ans ce qui exclut beaucoup d'étudiants et de retraités. Ainsi s'intéresser aux étudiants et aux retraités pour étudier les personnes qui ne font pas partie de la population ne semble pas très pertinent ici. (Les deux catégories ayant des pratiques très particulières, en effet, les étudiants, boivent rarement mais s'alcoolisent très fortement par rapport à la moyenne (3.5 verres par jour où la personne boit) et les retraités boivent plus souvent mais ils boivent peu par rapport à la moyenne (1.3 verre par jour où la personne boit))¹⁸.

Donc nous ne connaissons que trop peu la composition de cette population, nous avons décidé de l'agréger à la modalité « unemployed », nous verrons dans la suite du rapport comment cela se présente dans notre modèle.

6- Milieu urbain

Nous allons désormais parler de la variable « centcity », qui est codée 1, si la personne vit au centre d'une ville du « MSA », ce qui correspond à « Metropolitan statistical area », ce qui signifie aux personnes vivant dans une région urbaine. Pour justifier cette variable nous nous sommes aidés d'un ouvrage de M. Dejoz qui s'appelle « La consommation d'alcool est-elle plus importante dans les régions rurales ? ».

Au cours de cette étude, l'auteur conclut que : « Il est particulièrement difficile de comparer la consommation d'alcool entre différents milieux. Les données sur la population générale (en Belgique) ne permettent pas de conclure à un usage problématique plus élevé au sein des régions rurales qu'au sein des régions urbaines. Chez les adolescents, il existe cependant une tendance modérée à une consommation plus élevée en milieu rural. »¹⁹

C'est pour cela que nous avons décidé d'ajouter la variable « outercity », en la croisant avec les jeunes, nous expliquerons dans la suite de notre rapport la méthode.

¹⁶ INSEE (2016). Population active (comptabilité nationale). Paris : INSEE

¹⁷ AMANDA J. LEE, IAN K. CROMBIE, WILLIAM C. S. SMITH, HUGH TUNSTALL-PEDOE, Alcohol consumption and unemployment among men: the Scottish Heart Health Study, British Journal of Addiction, volume 85, n°9, p1165

¹⁸ LOCK KUNZ, Jean GRAHAM, Kathryn (1998). Drinking patterns, psychosocial characteristics and alcohol consequences. Addiction, volume 93, Addiction Research Foundation, London, p1079-1090

¹⁹ Dejoz, Michael. « 37. La consommation d'alcool est-elle plus importante dans les régions rurales ? », Vincent Seutin éd., *L'alcool en questions*. Mardaga, 2015, pp. 190-194.

B - Hypothèses

Donc nous avons vu précédemment quelles variables nous allons utiliser pour constituer notre modèle économétrique ainsi que les justifications avec notre revue de littérature. Désormais nous allons clairement formuler les hypothèses associées à nos différentes variables. Nous avons au total 9 hypothèses :

Hypothèse 1 : Avoir un père ou une mère alcoolique influence très positivement la probabilité que la personne abuse de l'alcool.

Hypothèse 2 : Les personnes au chômage sont des individus qui ont plus de chance d'abuser de l'alcool.

Hypothèse 3 : Plus les personnes vieillissent, plus elles ont une probabilité forte d'abuser de l'alcool.

Hypothèse 4 : Les personnes ayant une mauvaise santé ont une probabilité plus forte d'abuser de l'alcool.

Hypothèse 5 : Les personnes ayant fait moins d'étude ont plus de chance d'abuser de l'alcool.

Hypothèse 6 : Les jeunes vivant à la campagne ont plus de chance d'abuser de l'alcool.

Nous verrons dans la suite de notre dossier, le modèle que nous avons choisi pour voir si, ces hypothèses sont validées ou non. Nous introduirons chaque variable et nous verrons quels sont les effets de chaque variable.

IV. Méthodologie

Concernant la variable « mothalc » et « fathalc », nous avons décidé de l'agréger en une seule variable, la variable « parent_alcoolique » ainsi, si l'individu à un de ses deux parents alcoolique, la variable prendra la valeur « 1 ».

Concernant la variable « age » nous aurions pu la séparer entre différent groupe, ainsi nous aurions la variable jeune qui correspond à des individus âgés de 0 à 25 ans, ensuite nous aurions les adultes, âgés de 26 à 60 ans puis nous aurions les retraités ayant 60 et plus, afin de voir, quelle catégorie boit le plus étant.

Cependant, notre base de données concerne des individus allant de 25 à 59 ans, ce qui rend impossible cette analyse en termes de classe d'âge.

Nous avons tout de même créé une variable « âge » correspondant aux individus ayant entre 25 et 30 ans, afin de la croiser avec la variable « outercity », qui est codé 1, si les individus vivent dans une région qui n'est pas une région urbaine. Nous verrons lors de nos résultats si cette variable est pertinente

Pour la variable « status », comme nous l'avons expliqué, nous avons agrégé les variable « unemployed » et « out of workforce », ce qui correspond en fait à la variable « employ » qui est codé 1 si la personne travail.

Les différentes variables « exhealth, vghealth, goodhealth, fairhealth » ont été regroupé en une seule variable « sante » afin de séparer les individus en 2 groupes : la variable prend 1 si l'individu à « vghealth » égale à 1 et « exhealth » égale à 1 et 0 sinon. Sachant que nous avons au préalable exclu les individus n'ayant aucune situation de santé (« exhealth, vghealth, goodhealth, fairhealth » étaient toute égale à 0)

Concernant le milieu urbain, nous avons décidé de la croiser avec le jeune, comme on a vu que les jeunes qui vivait en milieu rural avait plus de chance d'abuser de l'alcool. Donc nous avons utilisé la variable « outercity » qui est codé 1, si la personne vit hors d'une région urbaine.

Pour finir, concernant la variable « educ » nous avons repris la variable telle quel, sans rien modifier.

A - Modèle Econométrique

Notre modèle économétrique sera donc :

$$abuse = \beta_0 + \beta_1 * age + \beta_2 * sante + \beta_3 * employ + \beta_4 * parent_{alcoolique} + \beta_5 * educ + \beta_6 * (jeune * outercity)$$

Nous pouvons donc réaliser un tableau montrant les signes attendus, conformément à nos hypothèses :

VARIABLE	SIGNES ATTENDUS	SIGNIFICATIVITE
Age	+	*
Sante	-	**
Employ	-	**
Parent alcoolique	-	***
Educ	-	*
Jeune*Outercity	+	**

B - Données

Nous avons donc le modèle économétrique que nous avons retenus, désormais nous allons présenter la base de données, expliquer les variables que nous avons conservé. Pour finir, nous réaliserons des statistiques descriptives pour expliquer une partie de l'information.

I - Présentation de la base de données

Nous avons repris une base de données de Wooldridge, s'appellant « alcohol », cette base est tiré d'une revue économétrique appelé le « Journal of Applied Econometrics », dans un article appelé « Alcohol Abuse and Employment : A Second Look », elle est composé de 9822 observations et de 33 variables. Ces données sont collectées aux Etats-Unis

Nous avons au final retenue onze variables pour notre étude : abuse, age, educ, exhealth, vghealth, fairhealth, goodhealth, outercity, fathalc, mothacl, employ. Nous avons opéré différentes modifications sur ces variables, que nous avons expliqué plus tôt, donc on se retrouve seulement avec 7 variables : jeune, adulte, sante, employ, parent_alcoolique, educ, jeune*outercity.

Nous pouvons voir sur ce tableau 15 premiers individus de la base.

	abuse	age	sante	employ	parent~e	educ	jeunex~y
1.	0	37	0	1	0	12	0
2.	0	53	1	1	0	9	0
3.	0	59	1	1	0	11	0
4.	0	43	1	1	1	10	0
5.	0	38	1	1	0	10	0
6.	0	34	0	1	1	10	0
7.	0	31	1	1	0	12	0
8.	0	43	1	1	0	18	0
9.	1	46	1	1	1	12	0
10.	0	36	1	1	1	12	0
11.	0	29	1	1	0	16	0
12.	0	32	0	1	0	15	0
13.	0	37	0	1	0	12	0
14.	0	27	1	1	0	12	0
15.	0	55	0	1	0	11	0

Nous avons dû supprimer certains individus, et donc nous étudierons au total 9 591 individus.

2- Description des variables

Cette partie a pour objectif d'apporter une description des variables que nous allons utiliser dans notre étude.

Tout d'abord, la variable « abuse » est notre variable expliquée. C'est une variable binaire codé 1 si la personne abuse de l'alcool

La variable jeune est une variable binaire codé 1 si la personne est âgée de 0 à 30 ans, pour construire cette variable, nous nous sommes servis de la variable « age ».

La variable « sante » est une variable binaire codé 1 si la personne est en bonne santé. Nous considérons que les personnes en bonne santé sont les individus ayant une excellente santé (exhealth codé 1) et une très bonne santé (vrhealth codé 1).

La variable « employ » est une variable binaire codé 1 si la personne travaille, c'est-à-dire qu'elle n'est pas au chômage, ni hors de la population active.

La variable parent alcoolique est une variable binaire codé 1 si au moins un des deux parents est alcoolique.

La variable educ est une variable quantitative discrète, correspondant au nombre d'années d'étude. Les individus font entre 0 et 19 ans d'étude. La durée moyenne des études est de 13 ans, pour donner un ordre d'idée, 12 ans d'étude aux Etats-Unis sont équivalentes à un baccalauréat général en France.

La variable age est une variable quantitative discrète, qui correspond à l'âge des individus. Les individus de notre variable ont entre 25 et 59 ans et sont âgés en moyenne d'environ 39 ans.

La variable jeunexoutercity est une variable croisé binaire codé 1 si l'individu est un jeune qui ne vit pas dans un milieu urbain.

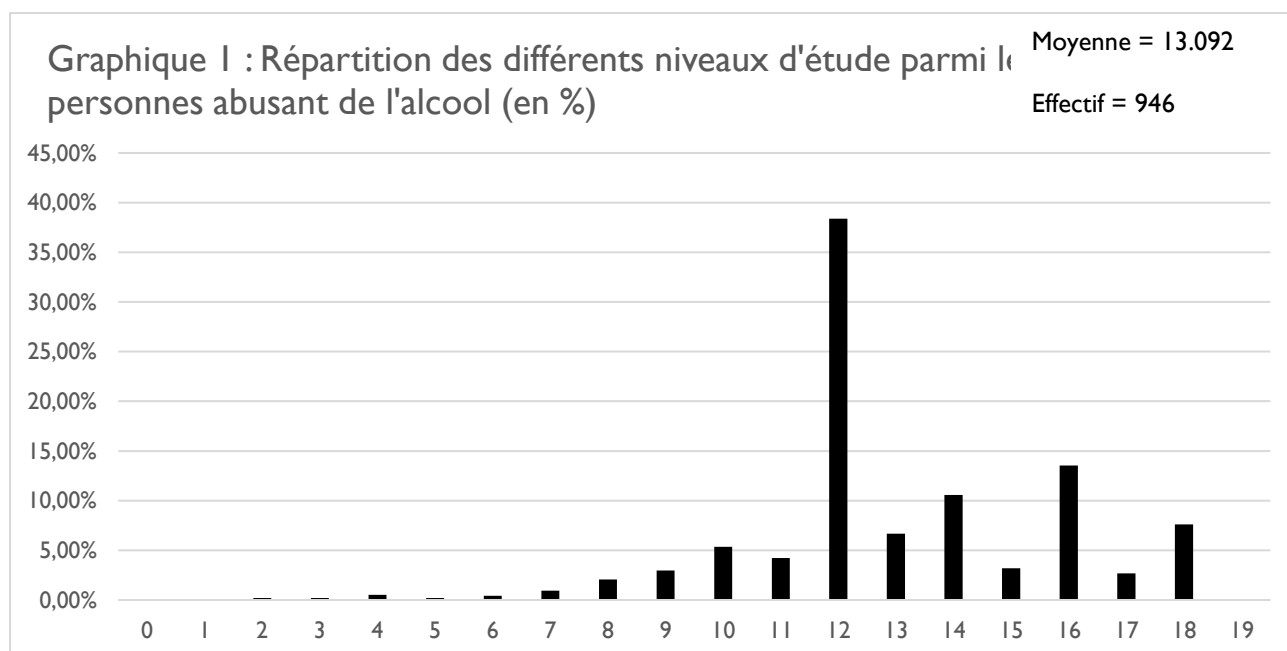
3- Statistiques descriptives

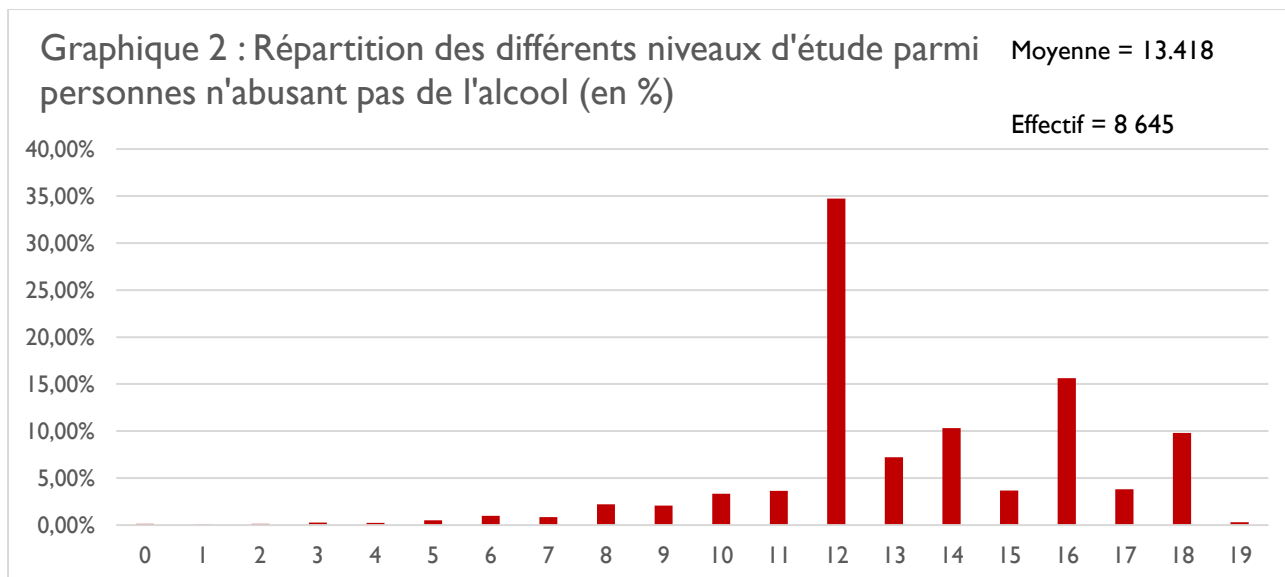
Dans cette partie nous allons réaliser les statistiques descriptives qui vont nous permettre d'expliquer une partie de l'information.

Individus	Effectif en nombre	Répartition en %
Alcoolique	946	9,86 %
Non Alcoolique	8645	90,14 %

Nous pouvons donc voir avec ce tableau que nous avons 9,86 % d'individus qui abusent de l'alcool.

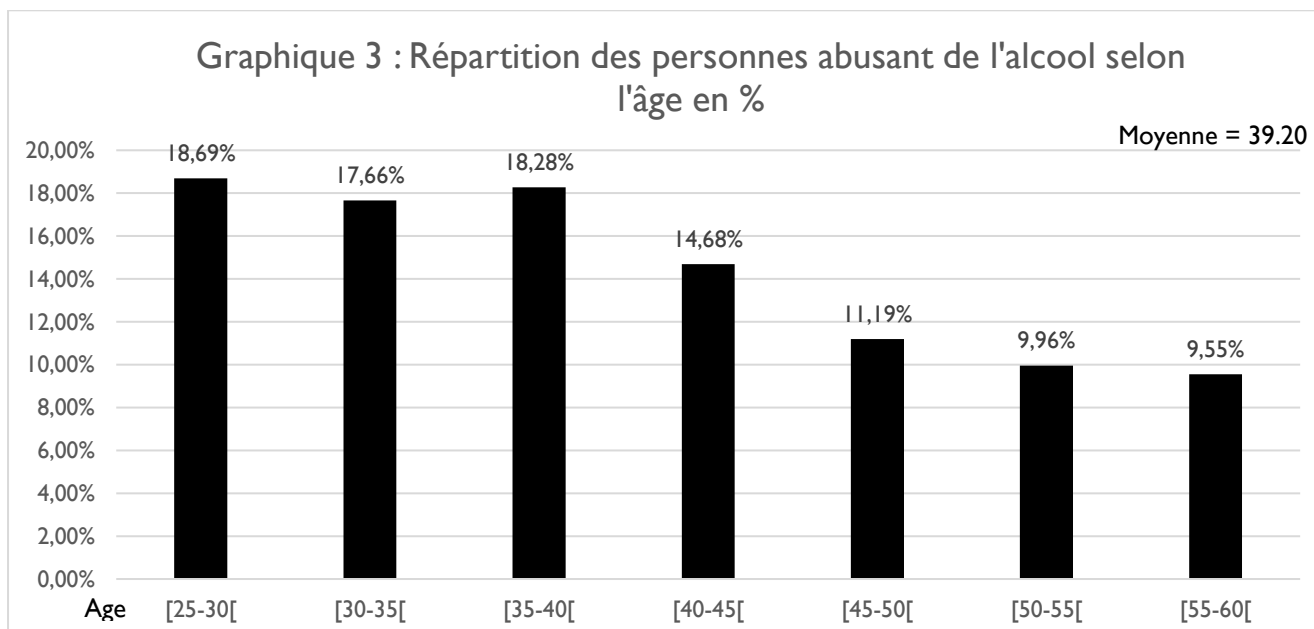
Nous avons ensuite réalisé deux histogrammes, un histogramme concernant uniquement les individus qui abuse de l'alcool et l'autre graphique concerne uniquement les individus qui n'abuse pas de l'alcool

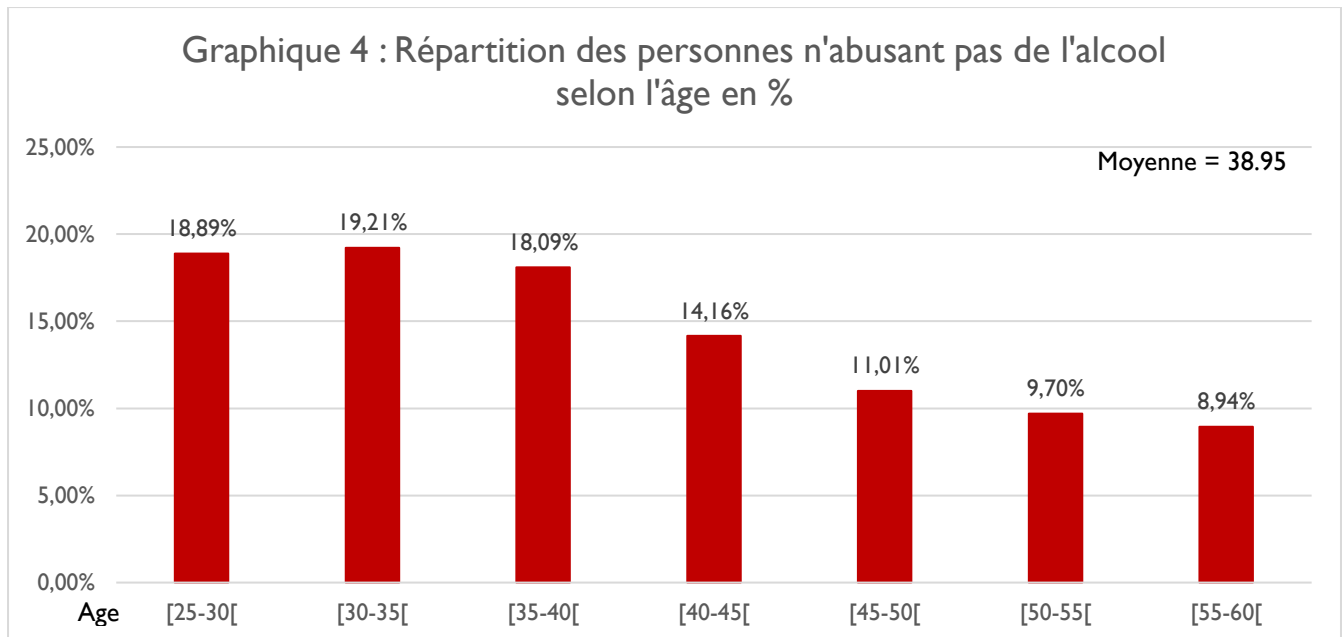




Nous pouvons voir avec ces 2 histogrammes ci-dessus qu'il ne semble pas y avoir de différence sensible entre les deux groupes. La moyenne entre les deux groupes est relativement la même. Nous pouvons donc nous demander si l'éducation a bien un impact sur la consommation d'alcool. Nous verrons dans la suite de notre rapport si cette variable est significative. Cependant on remarque que la moyenne du nombre d'année d'étude des personnes n'abusant pas de l'alcool est légèrement supérieure à la moyenne des individus abusant de l'alcool.

Nous allons désormais faire 2 histogrammes représentant la répartition des individus selon les 2 groupes d'individus : personne abusant de l'alcool et personne n'abusant pas de l'alcool.





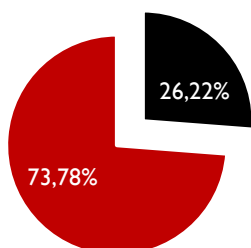
On voit sur ces 2 histogrammes au-dessus, qu'encore une fois, il ne semble pas y avoir de différence entre les individus les deux groupes (personnes abusant de l'alcool et personnes n'abusant pas de l'alcool). On voit que la moyenne des personnes abusant de l'alcool est légèrement supérieure à celle des individus n'abusant pas de l'alcool. Cependant encore une fois, nous ne pouvons pas dire si cette différence est réellement significative.

Ensuite nous avons réalisé un « tabulate » afin d'avoir les effectifs marginaux selon notre variable expliqué, et notre variable concernant l'alcoolisme des parents

	Parent alcoolique		Total
	0	1	
Abus d'alcool			
0	7 179	1 466	8 645
1	698	248	946
Total	7 877	1 714	9 591

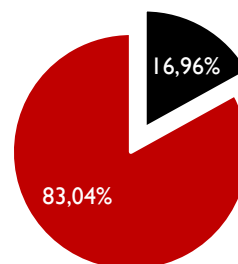
Puis nous avons réalisé deux graphiques en secteurs pour voir la proportion d'individus abusant de l'alcool selon qu'elle ai ou non des parents alcoolique

Graphique 5 : Répartition en % des personnes abusant de l'alcool et ayant ou non des parents alcoolique



■ Un des parents alcoolique ■ Aucun parent alcoolique

Graphique 6 : Répartition en % des personnes n'abusant pas de l'alcool et ayant ou non des parents alcoolique



■ Un des parents alcoolique ■ Aucun parent alcoolique

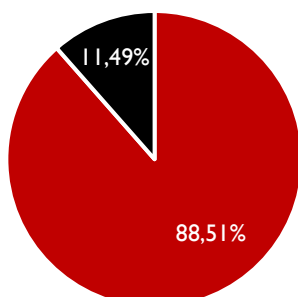
Nous pouvons voir à l'aide de ces graphiques en secteurs, qu'il y a une forte différence entre les 2 groupes. En effet, on voit qu'il y a une part beaucoup plus importante de parents alcoolique dans le groupe des personnes abusant de l'alcool.

Il y aurait environ 26,22% des personnes alcoolique qui ont eux même au moins un de leurs parents alcoolique. A contrario, il y aurait « seulement » 16,96 % des personnes n'étant pas alcoolique qui ont au moins un de leurs parents alcoolique.

Donc si l'on se fie uniquement à nos statistiques descriptives, il y aurait une corrélation positive entre les avoir des parents alcooliques et être soit même un alcoolique, ce qui valide notre première hypothèse.

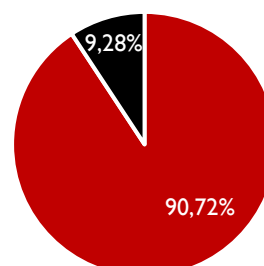
Ensuite nous avons décidé de faire la même chose concernant l'état de santé. Cependant dans ce cadre nous avons, dans un premier temps, constituer 2 groupes à partir de l'état de santé. En effet dans le premier cas, nous avons à faire aux groupes ayant une mauvaise santé et dans le second, les individus ayant une bonne santé.

Graphique 7 : Répartition en % des alcooliques parmi les individus ayant une mauvaise santé



■ N'abuse pas de l'alcool ■ Abuse de l'alcool

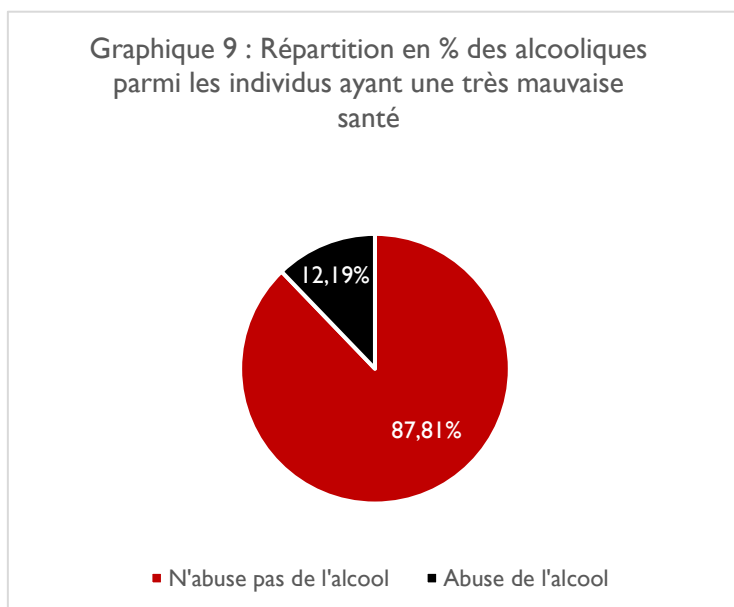
Graphique 8 : Répartition en % des alcooliques parmi les individus ayant une bonne santé



■ N'abuse pas de l'alcool ■ Abuse de l'alcool

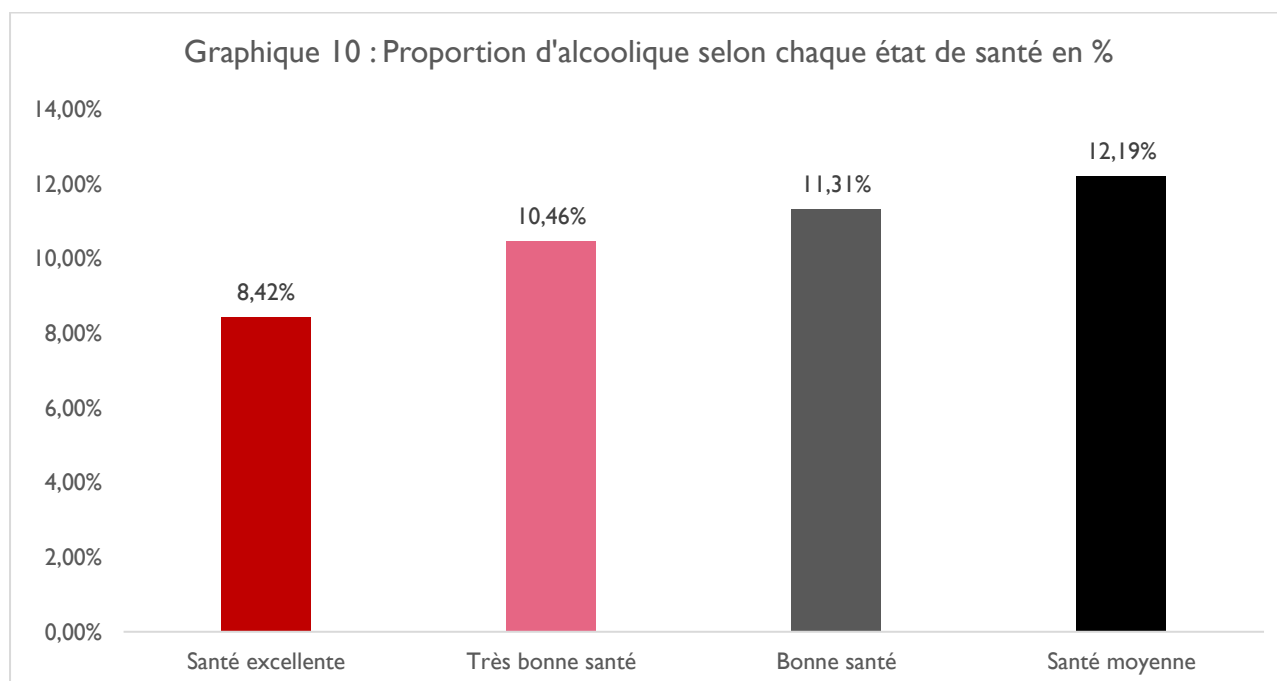
Nous pouvons voir que la proportion d'individus alcoolique est légèrement supérieure, lorsque les individus ont une santé moins bonne. Nous pouvons donc dire que l'hypothèse selon laquelle, plus les individus sont en mauvaise santé, plus ils ont une grande probabilité d'abuser de l'alcool semble vrai.

Pour renforcer cette hypothèse, nous avons réaliser un troisième graphique en secteur en prenant uniquement les individus ayant le pire état de santé. Donc nous avons uniquement pris les individus qui avait comme la variable : « fairhealth » codé 1.



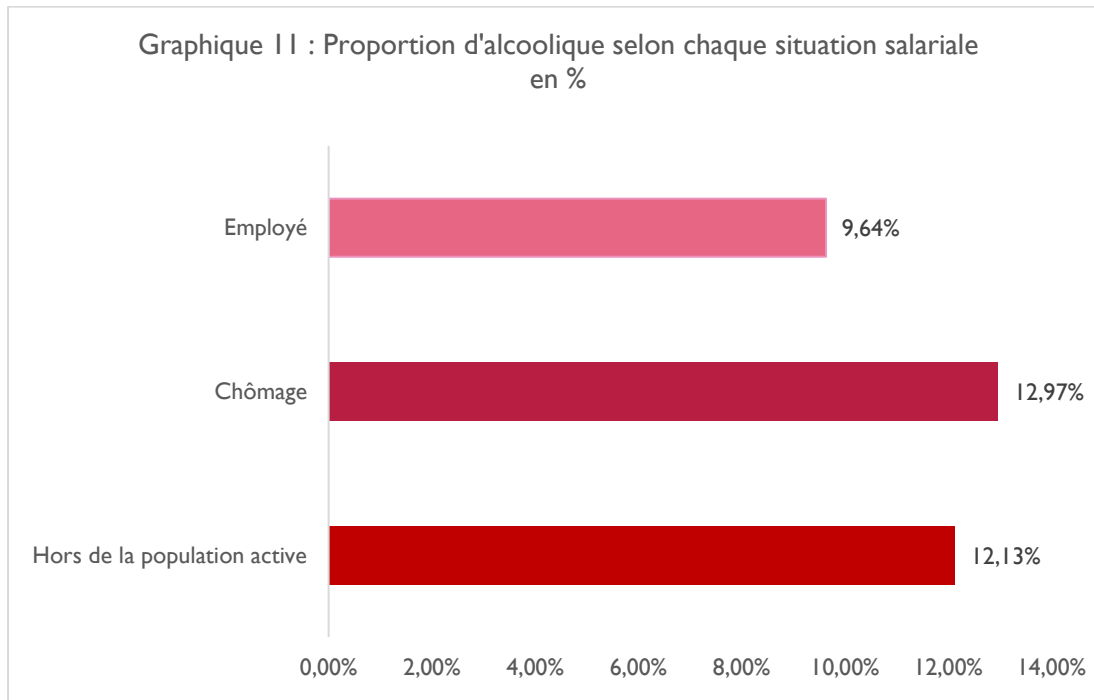
Nous pouvons donc voir que, plus l'état de santé se dégrade, plus il y a une proportion d'alcoolique forte.

Pour finir sur cette variable, nous avons décidé de faire un histogramme à bâton récapitulatif :



Ainsi nous pouvons clairement voir que la part d'alcoolique augmente clairement lorsque l'état de santé se dégrade.

Ensuite, nous avons décidé d'étudier la variable « employ », ce qui correspond à la situation salariale de l'individu. Nous avons repris la variable « status » pour réaliser ce graphique, car même si, dans notre revue de littérature, nous ne nous étions pas intéressés aux individus étant hors de la population active. Nous voulions tout de même voir, si être hors de la population active avait une incidence sur la probabilité d'abuser de l'alcool.



Nous pouvons donc voir sur ce graphique que la part des individus que la part des alcooliques, parmi les individus au chômage est beaucoup plus élevé que chez les individus étant employé. Nous pouvons également voir qu'il y a une légère différence entre les individus étant hors de la population active et les individus au chômage. Or comme cette différence est très faible, nous avons décidé de ne pas en prendre compte et de regrouper les 2 groupes (chômeur et individus hors de la population active) en 1 seul groupe avec la variable employ.

Nous pouvons également dire que notre hypothèse deux, semble être validé par ces statistiques descriptives.

Pour finir les statistiques descriptives, nous avons décidé d'étudier la variable croisée : jeune * outercity, ce qui correspond uniquement aux individus étant à la fois jeunes et vivant hors des régions urbaines.

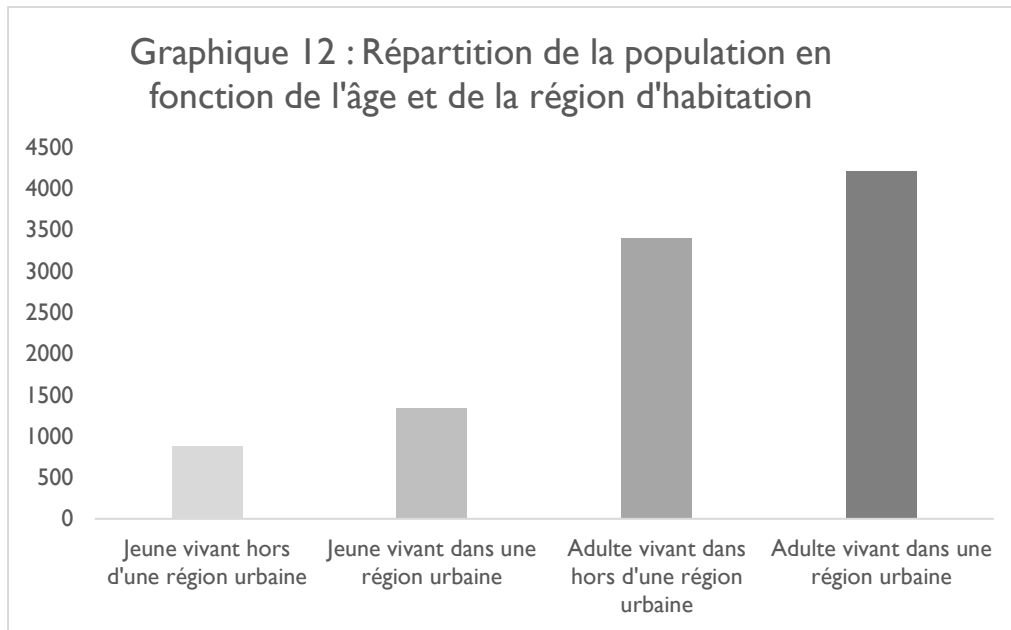
Tout d'abord, notre échantillon est réparti équitablement entre les personnes vivant en campagne et les personnes vivant dans les villes. Nous avons 5 385 habitants dans les villes et 4 206 individus vivant hors

des régions urbaines ce qui n'est pas très représentatif aux Etats-Unis, en effet, il y a environ 80 % des individus qui vivent dans des régions urbaines²⁰.

Donc pour étudier cette variable, nous avons commencer par découper notre échantillon en 4 groupes :

Les jeunes vivant hors d'une région urbaine, les jeunes vivant dans une région urbaine, les adultes vivant dans une région urbaine et les vieux ne vivant pas dans une région urbaine.

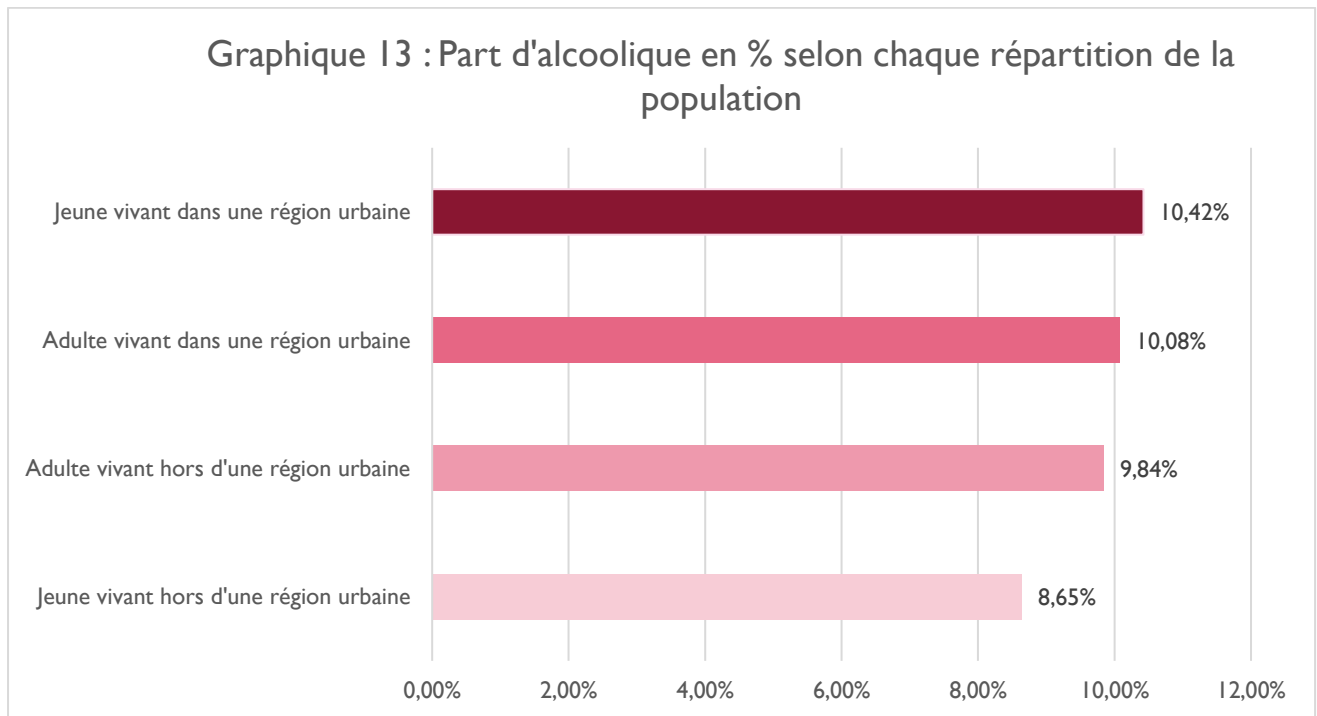
Ainsi on obtient une répartition de la population :



On peut voir que les jeunes ne représentent qu'une faible part de notre échantillon, contrairement aux adultes. En effet, les jeunes représentent 22,62% de notre échantillon, ce qui est assez faible, mais nous explorerons dans nos limites les raisons de cette mauvaise représentativité de notre échantillon.

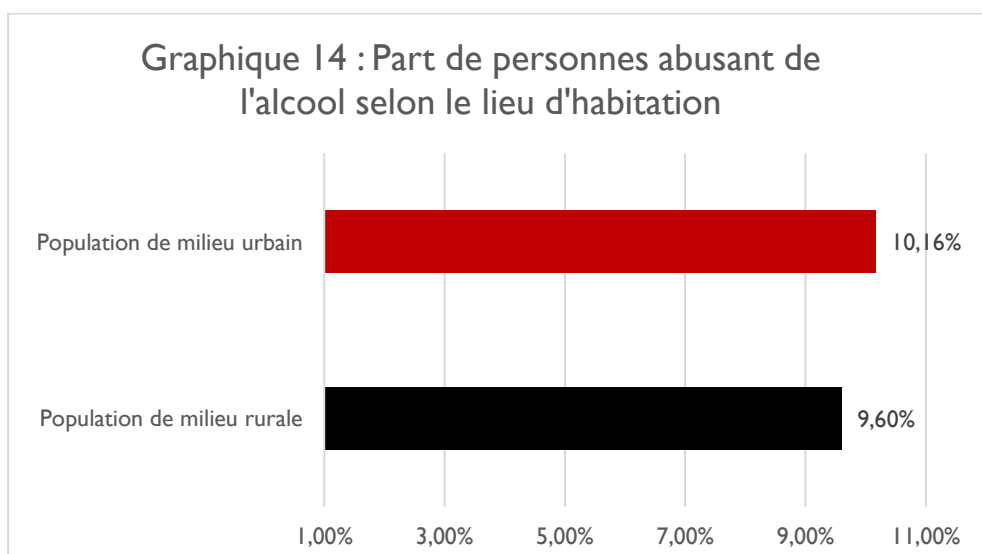
²⁰ SCOMMEGNA Paola (2003). Census Bureau to Tracj Both Metro and "Micropolitan" Areas [en ligne]. PRB. Disponible sur internet : <https://www.prb.org/resources/census-bureau-to-track-both-metro-and-micropolitan-areas/> : (consulté le 28/12/2021)

Ensuite, nous avons calculé, la proportion de personnes alcoolique selon chaque « strates » de l'échantillon que nous avons découpées préalablement. Nous obtenons donc ce graphique :



Nous voyons donc que nos résultats viennent complètement contredire notre revue de littérature, car il semble que, d'une part, les individus vivant dans une région urbaine sont plus à même de devenir alcoolique, alors qu'une étude prétendait l'inverse. Cependant les écarts entre les adultes vivant en région rurale et les individus vivant dans une région urbaine est fausse donc nous ferons un autre graphique pour uniquement étudier les 2 variables. De plus, on remarque que la part la plus faible de consommateur d'alcool semble être les jeunes venant d'une région rurale ce qui est le résultat inverse comparativement à notre hypothèse 6.

Nous allons donc voir un graphique comparant la consommation en milieu urbain et la consommation en milieu rurale, pour vérifier si la consommation est plus faible en milieu rurale qu'en milieu urbain.



Nous pouvons donc voir sur ce graphique qu'il y a donc qu'il y a plus de gens abusant de l'alcool dans les milieux urbains que dans les milieux ruraux, ce qui ne va pas à l'encontre de notre revue. En effet, M. DEJOZE concluait, avec les résultats inverse, que la différence n'était pas significative. Nous verrons, si, dans notre étude, ces résultats sont significatifs ou non.

Nous avons donc fini de parler de nos statistiques descriptives, nous allons désormais parler des résultats que nous avons obtenu avec notre analyse.

V. Résultats

A - Régressions

Pour notre régression, nous avons choisi un modèle probit car notre variable expliquée est une variable binaire. La régression de notre modèle nous donne le tableau ci-dessous, nous avons la valeur des coefficients associés à chaque variable, leurs significativités via les étoiles (plus une variable a d'étoile, plus la signification est bonne, une variable qui n'a pas d'étoiles n'est cependant pas significatif du tout). Nous avons également le nombre d'observation ainsi que le pseudo R².

Variable	model1
age	.00006599
sante	-.0842991**
employ	-.07857531
parentalco~e	.28148494***
educ	-.01489937**
jeunexoute~y	-.0814824
_cons	-1.0135243***
N	9591
pseudo_r2	0.0103

Tout d'abord, nous pouvons dire que nous avons 3 coefficients qui ne sont pas significativement différent de 0. Pour commencer : la variable age : n'est pas significative, ce qui est conforme avec ce que nous avons vu lors de nos statistiques descriptives. Nous allons donc omettre cette variable.

Ensuite, concernant la variable employ, nous avons essayé d'opérer d'autres modifications pour essayer de rendre le coefficient significativement différent de 0. Nous avons notamment supprimé les individus qui étaient membre de la population active, mais cela n'est pas suffisant pour rendre le coefficient significatif. Nous avons donc décidé d'omettre cette variable.

Pour finir, la variable jeunexoutercity est une variable qui n'est pas significative, ce qui avait été pressenti via nos statistiques descriptives. Nous avons donc décidé de modifier cette variable, en utilisant uniquement la variable « centcity » qui prend la modalité « 1 », si la personne vit au centre d'une région urbaine.

Ainsi nous avons spécifié le modèle une seconde fois et obtenu :

Variable	model2
sante	-.09060905**
parentalco~e	.27966889***
educ	-.0158234**
centcity	.10199626***
_cons	-1.1073312***

Legend: * p<.1; ** p<.05; *** p<.01

N	9591
pseudo_r2	0.0110

Ce nouveau modèle comporte uniquement des variables significatives au seuil de 5 %. Nous pouvons donc écrire la probabilité pour qu'un individu soit un individu qui abuse de l'alcool est :

$$abuse = -1,107 - 0,091 * sante + 0,280 * parentalcoolique - 0,016 * educ + 0,102 * centcity$$

Avec notre régression actuelle, la seule chose que nous pouvons interpréter est le signe des coefficients.

Ainsi nous pouvons dire :

- Une meilleure santé diminue la probabilité que la personne abuse de l'alcool.
- Avoir un père ou une mère alcoolique augmente la probabilité que l'individu abuse de l'alcool.
- Avoir fait plus d'années d'étude fait diminuer la probabilité que la personne soit alcoolique.
- Vivre dans une région urbaine augmente la probabilité que l'individu abuse de l'alcool. Nous pouvons noter que cette interprétation est la seule qui rentre en contradiction avec nos hypothèses.

Ensuite nous avons calculer les effets marginaux :

variable	dy/dx
sante*	-.0159233
parent~e*	.0536017
educ	-.0027047
centcity*	.0178316

Les effets marginaux permettent de calculer l'impact que chaque variable a sur la probabilité que l'individu soit alcoolique.

Ainsi, nous pouvons interpréter les coefficients de la manière suivante :

- Être une personne en bonne santé par rapport à la moyenne, diminue la probabilité de 1,59 % d'être une personne dépendante vis-à-vis de l'alcool, TCEPA.
- Être une personne avec au moins un des parents alcooliques augmente de 5,36 % la probabilité d'être une personne dépendante vis-à-vis de l'alcool, TCEPA.
- Une année supplémentaire d'éducation a pour conséquence de faire diminuer de 0,27 % la probabilité d'être une personne alcoolique, TCEPA.
- Une personne vivant dans une région urbaine augmente de 1,78 % la probabilité de devenir une personne dépendante vis-à-vis de l'alcool, TCEPA.

B - Tests

Nous avons donc vu qu'elles étaient les résultats de notre régression. Cette séquence est consacrée aux différents tests inhérents à notre régression. Ces tests servent à savoir si notre régression respecte les hypothèses inhérentes à nos méthodes de régression, ici le probit.

1- Test de Wald

Le test de Wald teste la significativité des coefficients de chacune de nos variables explicatives. Nous avons vu que certaines de nos variables ne passaient pas ce test, nous les avons donc modifiées et nous sommes arrivés à une situation où l'ensemble de nos variables sont, au moins significatives au seuil de 5%. Le test de Wald teste les hypothèses suivantes :

H0 : Les coefficients sont significativement différents de 0

H1 : Les coefficients ne sont pas significativement différents de 0

Nous avons accepté H0 au seuil de 5 % pour les variables sante et educ. Nous avons également accepté H0 au seuil de 10% pour les variables parent_alcoolique et centcity.

2- Tableau de prédiction

Nous allons ensuite mesurer la qualité de notre modèle via sa capacité prédictive. En effet, le tableau de prévision permet de visualiser, le pourcentage d'observation bien prédite par notre modèle pour les 2 modalités distinctes. Dans notre cas, nous allons prédire pour chaque individu, en fonction de notre modèle, si l'individu abuse ou non de l'alcool. Cela nous permettra d'estimer la précision de notre modèle.

Classified	True		Total
	D	~D	
+	0	0	0
-	946	8645	9591
Total	946	8645	9591
Correctly classified			90.14%

Notre modèle prédit correctement 90,14 % des modalités prises par les individus. Cependant il y a un problème principal, en effet, les individus tous ont été prédits comme n'étant pas alcoolique. Ainsi ce taux de prédiction est uniquement dû à la faible part d'individus alcoolique dans notre base de données. Donc nous pouvons conclure que notre modèle n'a pas une bonne capacité prédictive.

3- R^2 de McFadden

Nous allons désormais nous intéresser au R^2 de McFadden qui prend la valeur 0.0110 dans notre modèle. Cependant le R^2 de McFadden n'est pas interprétable comme le R^2 ajusté des MCO. On estime qu'un R^2 correcte pour une régression probit devrait être compris entre 0,2 et 0,4 pour avoir un modèle correct.

4- Hosmer-Lemeshow

Le test d'Hosmer-Lemeshow cherche à tester la qualité d'ajustement du modèle. Il a pour objectif d'estimer si le modèle s'ajuste aux données de façon adéquate.

H0 : La qualité d'ajustement de notre modèle est bonne ($\text{prob} > 0,5$)

H1 : La qualité d'ajustement de notre modèle n'est pas bonne ($\text{prob} < 0,5$)

```

Number of observations = 9,591
Number of covariate patterns = 2,185
Pearson chi2(2179) = 2062.24
Prob > chi2 = 0.9634
    
```

Nous pouvons voir qu'à la suite de la réalisation de ce test, il semble que nous avons un modèle avec une excellente qualité d'ajustement, car notre prob de chi2 est de 0,9634.

Cependant, vis-à-vis de ce que nous avons vu avec le R^2 de McFadden et le tableau de prévision, il ne semble pas que nous ayons un modèle correct.

VI. Conclusions & discussions

Cette partie a pour objectif de résumer les résultats obtenus, de les comparer à nos hypothèses et d'expliquer les raisons pour lesquelles notre étude n'est pas parfaite.

Dans un premier temps, nous avons vu que de nombreuses variables n'étaient pas significatives selon notre modèle, notamment : l'âge, l'emploi et la variable croisée : jeune*outercity.

La non-significativité de la variable âge est liée à la non représentation de différentes classes d'âge dans notre base de données. En effet, il semblerait, selon notre revue de littérature, que les différentes classes d'âge amènent des comportements différents vis-à-vis de l'alcool.

Concernant la non-significativité de l'emploi, l'explication est plus complexe. En effet, nous avons vu dans nos statistiques descriptives qu'il semblait y avoir une différence entre les individus étant au chômage et les individus employés. Cependant la présence des individus hors de la population active avec les chômeurs venait « compenser » ces différences. En revanche, nous avons essayé de supprimer les individus étant hors de la population active, mais, même en faisant cela, la variable n'était pas significative. Pour améliorer cette variable, il faudrait intégrer des informations supplémentaires : la durée du chômage par exemple.

Nous avons également vu que les résultats vis-à-vis de notre variable « outercity » semble aller à l'encontre de notre revue de littérature. Selon moi, cela s'explique par une différence culturelle. En effet, l'étude dont nous nous sommes servis pour l'ajout de cette variable se déroulait en France, contrairement à notre base de données qui est américaine. Et donc, la différence avec ce que l'on a obtenu pourrait s'expliquer par l'existence de différences vis-à-vis de la consommation d'alcool au Etats-Unis. Tout d'abord, l'âge minimum légal pour pouvoir consommer, où acheter de l'alcool, n'est pas le même aux Etats-Unis (21 ans au Etats-Unis et 18 ans en France). De plus, il semble que les Etats-Unis ont une politique qui encadre beaucoup plus la consommation d'alcool que la France, ce qui pourrait expliquer des différences de consommations selon les âges et le milieu d'habitation.

Donc pour finir, nous avons vu que nous avons validé les hypothèses suivantes :

1 : Avoir un père ou une mère alcoolique influence très positivement la probabilité que la personne abuse de l'alcool.

2 : Les personnes ayant une mauvaise santé ont une probabilité plus forte d'abuser de l'alcool.

3 : Les personnes ayant fait moins d'étude ont plus de chance d'abuser de l'alcool.

Cependant nous avons trouvé des résultats inverse pour l'hypothèse suivante :

Les jeunes vivant à la campagne ont plus de chance d'abuser de l'alcool.

Et Les deux autres hypothèses ont été réfuté par manque de significativité de nos variables. Nous allons voir dans la dernière partie, les limites qui expliquent la non significativité de nos variables et essayons de donner des pistes pour améliorer notre modèle.

VII. Limites

Nous avons donc conclu et discuter autour de notre sujet, et désormais nous allons voir, quels sont les différentes limites de notre rapport.

Tout d'abord, il y a une limite « technique » en effet, l'alcoolisme provient souvent d'une modification d'un statut : perte d'un emploi, divorce, perte d'un proche... ce qui n'est pas pris en compte par notre étude, comme notre base de données ne prend en compte que des profils à un instant donné, et pas de profil avec des évolutions. Il serait intéressant par exemple, de savoir si la personne est : célibataire, marié, divorcé, veuf etc... là où on a simplement une variable : marié ou non.

C'est pour cela que l'on a dit que la variable « mothalc / fathalc » était l'une des variables les plus intéressante, en effet, elle est pleinement expliquée par notre revue de littérature, là où notre variable « status » seraient plus intéressante si on pouvait savoir si, par exemple, la personne a perdu un emploi.

Une deuxième limite provient de notre base de données. En effet, nous ne disposons pas de la variable sexe, qui semble être assez déterminantes pour certaines variables : par exemple : une femme mariée aurait une probabilité plus faible de boire qu'une femme célibataire, Marie Zins dans : « les déterminants sociaux des consommations d'alcool dans la cohorte gazel », conclue que : « De nombreuses études ont montré l'existence chez les femmes d'une liaison entre le statut marital et la consommation d'alcool. Les résultats de ces travaux sont concordants à de légères différences près. Globalement il y est retrouvé une proportion plus élevée de grosses consommatrices chez les célibataires par rapport aux femmes mariées et veuves. Les différences concernent la consommation des femmes divorcées, supérieure à celles des femmes mariées dans certaines études, inférieure dans d'autres, mais toujours inférieure à celles des célibataires »²¹, donc, l'absence de la variable « sexe » nous fait perdre une partie de l'information. En effet on aurait pu créer une variable croisée avec entre les femmes et les personnes mariées.

Le sexe est également une variable explicative importante pour expliquer la consommation d'alcool, selon l'article : « Binge drinking in young adults : Data Definitions and Determinants » : « An additional factor is gender, because as many as 81% of all binge-drinking episodes are attributed to men [...], but men also demonstrate increased frequency of alcohol dependence »²². Donc cela montre que le sexe est également une variable importante pour expliquer l'abus d'alcool, notamment la pratique du « binge drinking » et l'alcoolisme.

Une troisième limite provient de notre base de données, en effet le manque d'information vis-à-vis des variables est un véritable problème. Notre variable explicative « abuse » n'est pas assez clair, et comme nous l'avons vu dans notre introduction, les pratiques concernant la consommation d'alcool diffèrent selon les individus. Pour faciliter notre analyse, nous avons décidé d'inclure tous ces pratiques comme étant de l'abus d'alcool, mais ce n'est pas très rigoureux sur le plan scientifique. Le manque d'information globale vis-à-vis de la description de nos variables pose un certain problème à la réalisation de cette étude.

Une autre limite concerne la variable âge, en effet nous avons vu que selon l'âge, il y avait différents types de consommation, et que certains événements, comme le départ en retraite pouvait entraîner de l'abus d'alcool. Or comme nous l'avons vu, nos individus sont âgés de 25 à 59 ans. Cela pose donc un problème car nous ne nous intéressons pas aux jeunes, ni aux personnes vieilles. Selon notre revue de littérature, nous pouvons dire qu'il y aurait une sorte de « cycle de vie », où selon l'âge de l'individu, il aurait plus ou moins de chance d'adopter un mésusage des boissons alcooliques et chaque classe d'âge à sa forme d'abus d'alcool.

Par exemple, si on prend l'exemple des jeunes, selon notre revue de littérature, les individus jeunes ont une forte probabilité de chance de pratiquer du binge drinking, et une faible probabilité de devenir dépendant des boissons alcooliques.

Pour conséquence à toutes ces limites, nous avons vu que nous avons un mauvais modèle prédictif. Pour améliorer la capacité prédictive de notre modèle il faudrait donc inclure la variable « sexe » qui est une des variables les plus déterminantes. De plus, l'ajout de différentes classes d'âge pourrait être intéressant pour expliquer plus d'information. Pour finir, l'ajout de variable « d'événement » pourrait être également quelque chose d'intéressant. Dans ce cadre il pourrait, par exemple y avoir une variable : perte d'un proche au cours des 6 derniers mois.

²¹ ZINS, Marie (2011). Déterminants sociaux des consommations d'alcool dans la cohorte gazel. Thèse de doctorat : Epidémiologie, Université Paris XI : 205 p

²² COURTNEY Kelly, POLICH John (2009), Binge Drinking in Young Adults: Data, Definitions, and Determinants. Psychological Bulletin, Volume 135, n°1, p142-156

VIII. Bibliographie

Fitzgerald, Hiram E, et Robert A. Zucker. « Effets à court et à long terme de l'alcoolisme parental sur les enfants », *Devenir*, vol. 14, no. 2, 2002, pp. 169-182.

NARCISSE, Marie-Rachelle (2000). Les déterminants de la consommation d'alcool au Canada : Facteurs socio-démographiques et économiques [en ligne]. Montréal : Faculté des arts et des sciences - Département de Sciences Economiques. Disponible sur internet : <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/866> (consulté le 25/11/2021)

Mathilde Touvier, Nathalie Druesne-Pecollo, Pilar Galan, Serge Hercberg, Paule Latino-Martel (2013). Déterminants de la consommation de boissons alcoolisées dans l'Étude NutriNet-Santé, France. *Bulletin Épidémiologique Hebdomadaire*, 2013, n°. 16-17-18, p. 199-204

G. Airagnes, C. Lemogne, F. Limosin, M. Zins, M. Dosquet, N. Hoertel, M. Marie-Krause (2019). L'abus d'alcool à différents âges de la vie : données épidémiologiques récentes. *French Journal of Psychiatry*, volume 1, p.40-41.

INSEE (2016). Population active (comptabilité nationale). Paris : INSEE

LOCK KUNZ, Jean GRAHAM, Kathryn (1998). Drinking patterns, psychosocial characteristics and alcohol consequences. *Addiction*, volume 93, Addiction Research Foundation, London, p1079-1090

Dejozé, Michael. « 37. La consommation d'alcool est-elle plus importante dans les régions rurales ? », Vincent Seutin éd., *L'alcool en questions*. Mardaga, 2015, pp. 190-194.

ZINS, Marie (2011). Déterminants sociaux des consommations d'alcool dans la cohorte gazel. Thèse de doctorat : Épidémiologie, Université Paris XI : 205 p

COURTNEY Kelly, POLICH John (2009), Binge Drinking in Young Adults: Data, Definitions, and Determinants. *Psychological Bulletin*, Volume 135, n°1, p142-156

KOPP Pierre (2015). Le coût social des drogues en France. Paris : Observatoire français des drogues et des toxicomanies, 10 p

Prohibition. (2021, 17/12/2021). Dans Wikipédia. <https://fr.wikipedia.org/wiki/Prohibition>

INSERM. Effets néfastes de l'alcool : impact du binge drinking sur le cerveau. p187-213

IRDES (2019). La politique de lutte contre l'alcoolisme. Paris, 250 p

SCOMMEGNA Paola (2003). Census Bureau to Track Both Metro and "Micropolitan" Areas [en ligne]. PRB. Disponible sur internet : <https://www.prb.org/resources/census-bureau-to-track-both-metro-and-micropolitan-areas/> : (consulté le 28/12/2021)

IX. Annexe

A - Code Stata

Vous pouvez télécharger notre base de données ici : [Alcohol.xlsx](#) > Fichier > Télécharger > Excel

Vous pouvez télécharger notre code ici : [Code Stata Alcohol](#)

```
import excel "C:\Users\emili\OneDrive\Bureau\Projet - Alcohol\alcohol.xlsx", sheet("alcohol") firstrow
```

```
destring abuse status unemrate age educ married famsize white exhealth vghealth goodhealth fairhealth  
northeast midwest south centcity outercity qrt1 qrt2 qrt3 beertax cigtax ethanol mothalc fathalc livealc inwf  
employ agesq beertaxsq cigtaxsq ethanolsq educsq, replace
```

*formule format, permet de changer le nombre de décimal

*I si la personne est alcoolique

*on va tester avec nos hypothèse de base : âge, nombre d'année d'éducation, santé de l'individu, alcoolisme de parents, situation pro

*appeler : age, educ, exhealth vghealth goodhealth fairhealth, mothalc fathalc, status

```
lab var educ "Nombre d'année d'éducation"
```

```
lab var abuse "abus d'alcool"
```

```
/*Création de la variable sante*/
```

```
gen sante = 0 if fairhealth == 1
```

```
replace sante = 0 if goodhealth == 1
```

```
replace sante = 1 if vghealth == 1
```

```
replace sante = 1 if exhealth == 1
```

```
replace sante = 3 if fairhealth == 0 & goodhealth == 0 & vghealth == 0 & exhealth == 0
```

```
drop if sante == 3
```

```
/*Création de la variable jeune*/
```

```
gen jeune = 1 if age <= 30
```

```
replace jeune = 0 if age > 30
```

```
/*Création de la variable adulte*/
```

```
gen adulte = 1 if age > 30
```

```
replace adulte = 0 if age <= 30
```

```
/*Création de la variable parent_alcoolique*/
```

```
gen parentalcoolique = 1 if fathalc == 1
```

```
replace parentalcoolique = 1 if mothalc == 1
```

```
replace parentalcoolique = 0 if fathalc == 0 & mothalc == 0
```

```
/*Création de la variable croisé*/
```

```
gen jeunexoutercity = 1 if jeune == 1 & outercity == 1
```

```
replace jeunexoutercity = 0 if jeune == 0
```

```
replace jeunexoutercity = 0 if outercity == 0
```

```
list abuse age sante employ parentalcoolique educ jeunexoutercity in 1/15
```

```
/*Drop des variables*/
```

```
drop exhealth fairhealth goodhealth vghealth fathalc mothalc
```

```
drop unemrate married famsize northeast midwest south centcity qrt1 qrt2 qrt3 beertax cigtax ethanol inwf  
agesq beertaxsq cigtaxsq ethanolsq
```

```
drop white
drop educsq
drop livealc
*drop if status == 1
```

```
lab var sante "bonne santé = 1"
```

```
summarize age
summarize educ
```

```
count if abuse == 0
count if abuse == 1
```

```
mean educ if abuse == 1
mean educ if abuse == 0
```

```
mean age if abuse == 1
mean age if abuse == 0
summarize educ if abuse == 1
```

```
graph box educ, over(abuse)
```

```
tabulate abuse parentalcoolique, row
tabulate abuse sante, row
```

```
tabulate outercity
```

```
probit abuse age sante employ parentalcoolique educ jeunexoutercity
```

```
estimates store model1
estimate table model1, star(.1 .05 .01)
estimate table model1, b(%7.2f) se(%7.2f) stats(N pseudo_r2)
```

```
gen age2 = age*age
gen age3= log(age)
```

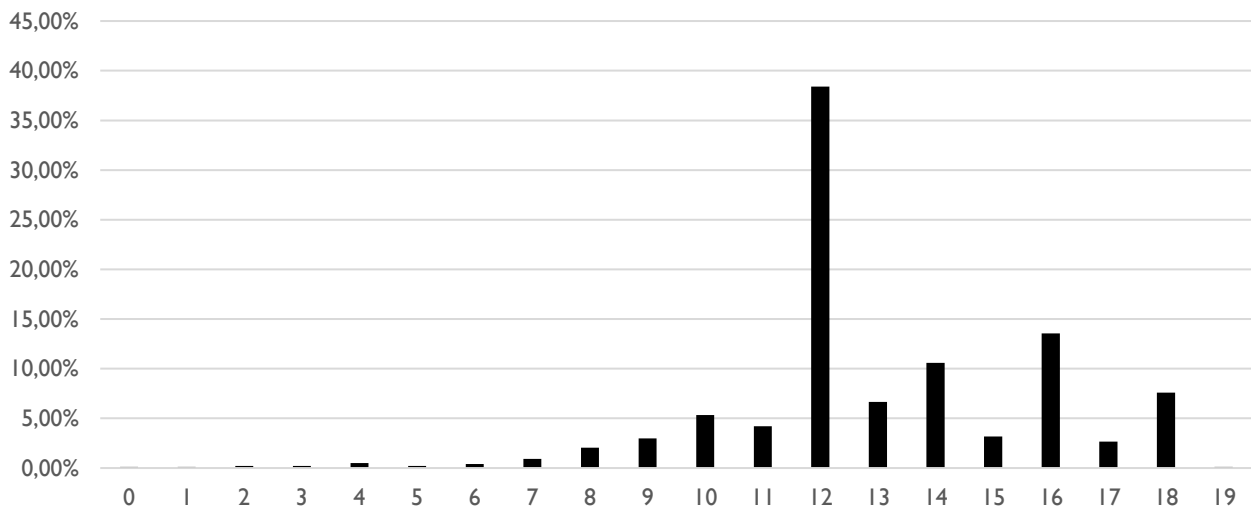
```
probit abuse sante parentalcoolique educ centcity
estimates store model2
estimate table model2, star(.1 .05 .01)
estimate table model2, b(%7.2f) se(%7.2f) stats(N pseudo_r2)
mfx compute
probit abuse sante parentalcoolique educ centcity
lstat
```

```
estat gof
```

B - Autres

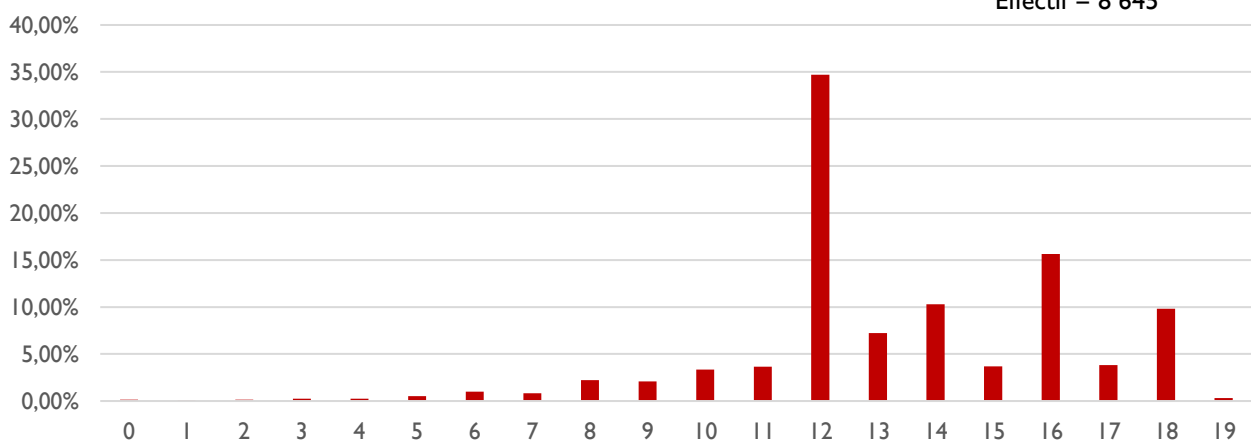
Graphique :

Graphique 1 : Répartition des différents niveaux d'étude parmi les personnes abusant de l'alcool (en %)

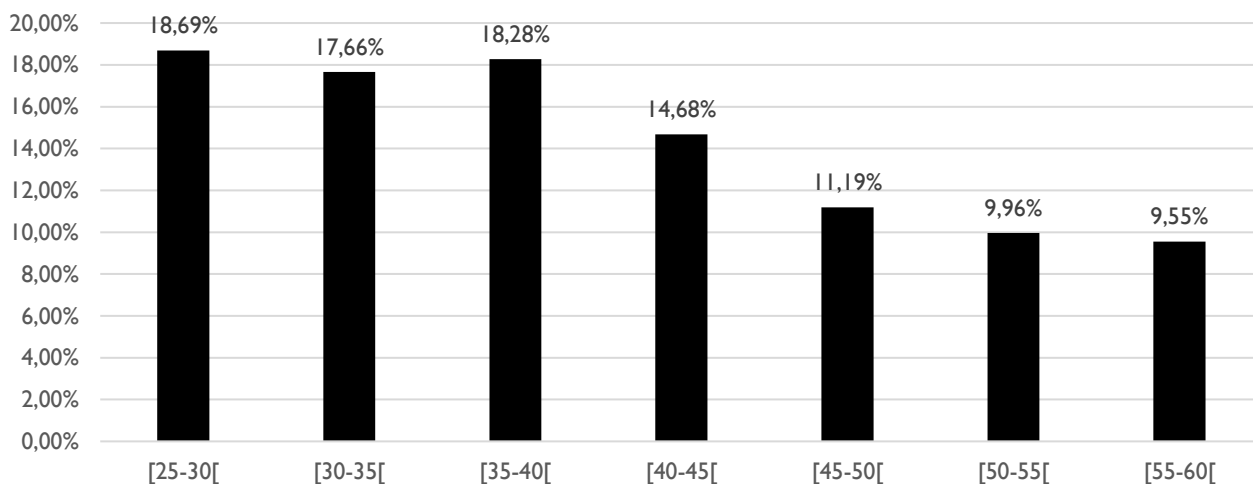


Graphique 2 : Répartition des différents niveaux d'étude parmi personnes n'abusant pas de l'alcool (en %) Moyenne = 13.418

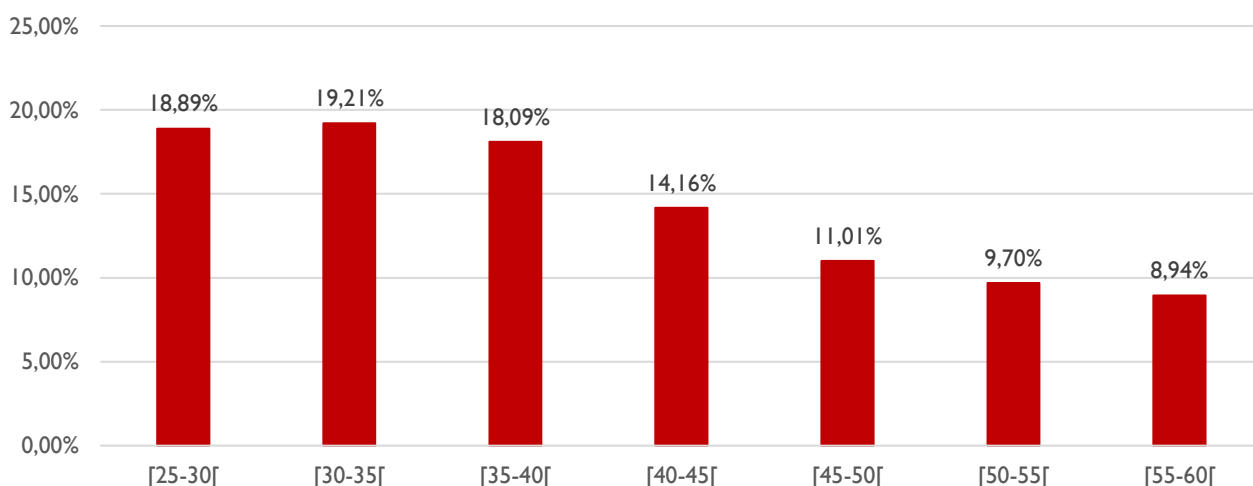
Effectif = 8 645



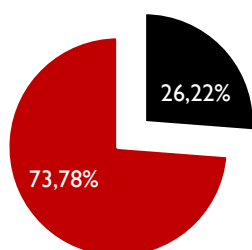
Graphique 3 : Répartition des personnes abusant de l'alcool selon l'âge en %



Graphique 4 : Répartition des personnes n'abusant pas de l'alcool selon l'âge en %

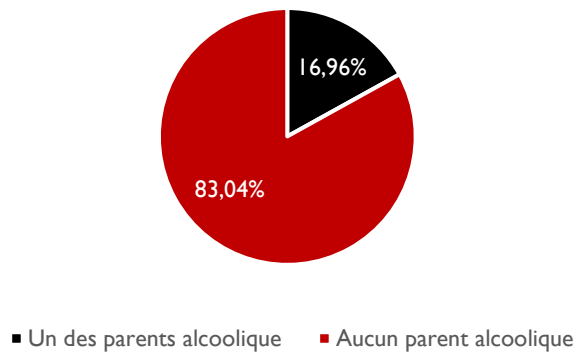


Graphique 5 : Répartition en % des personnes abusant de l'alcool et ayant ou non des parents alcoolique

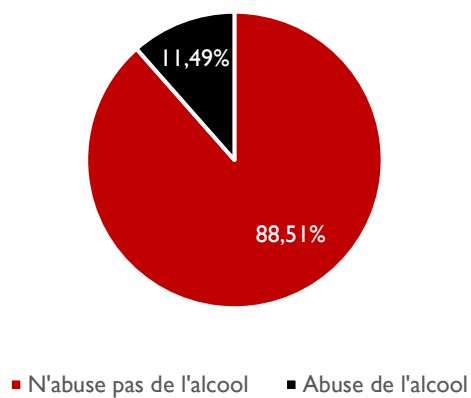


■ Un des parents alcoolique ■ Aucun parent alcoolique

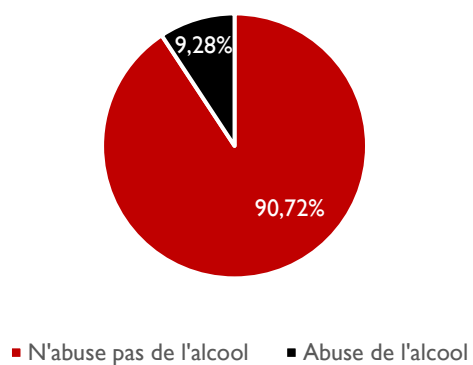
Graphique 6 : Répartition en % des personnes n'abusant pas de l'alcool et ayant ou non des parents alcoolique



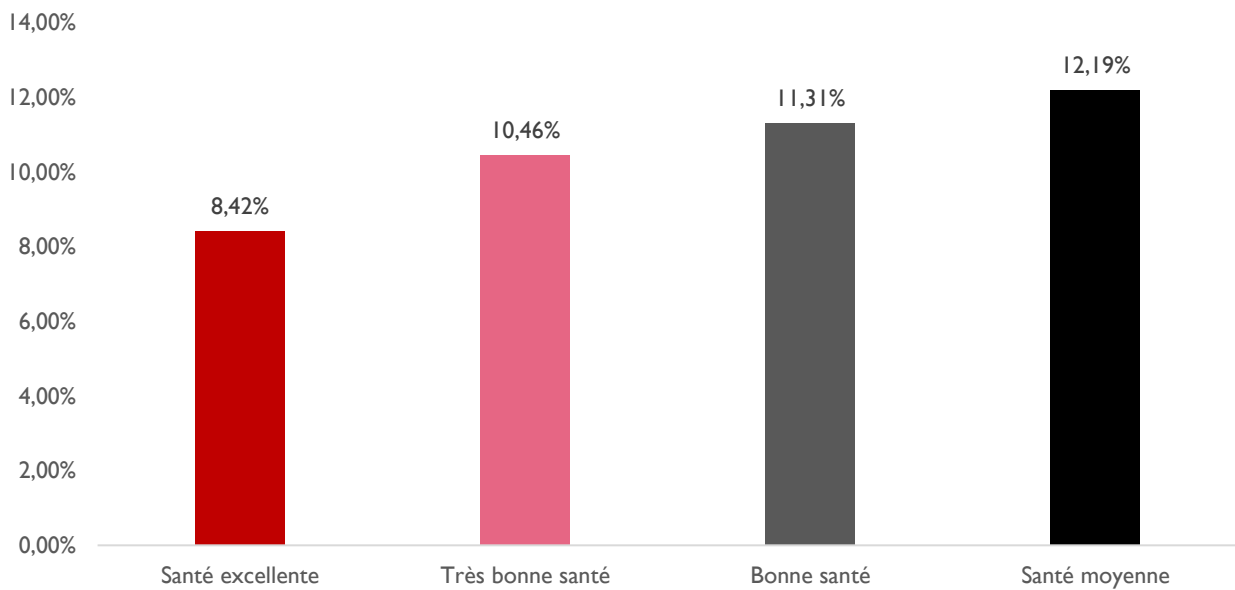
Graphique 7 : Répartition en % des alcooliques parmi les individus ayant une mauvaise santé



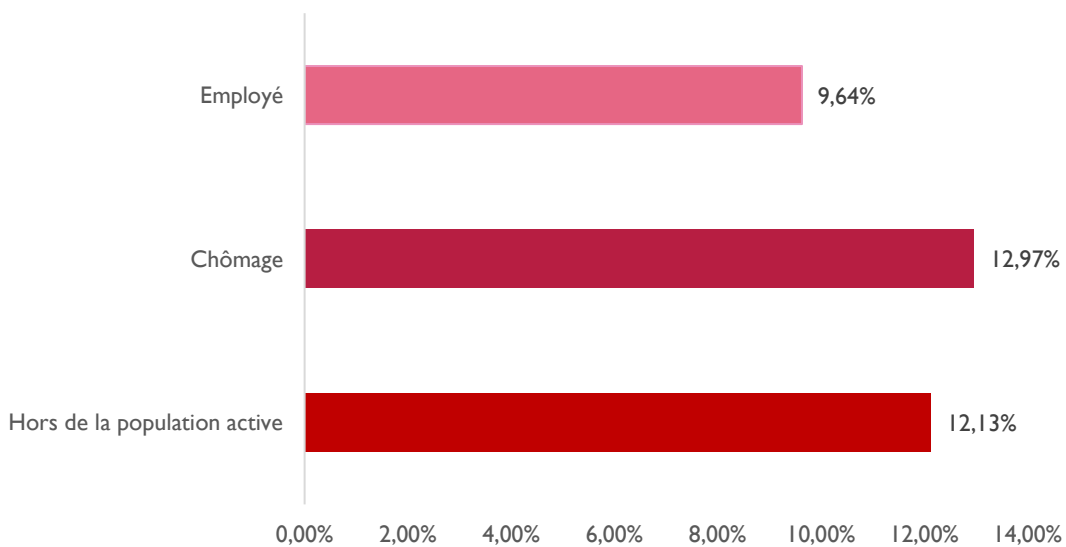
Graphique 8 : Répartition en % des alcooliques parmi les individus ayant une bonne santé



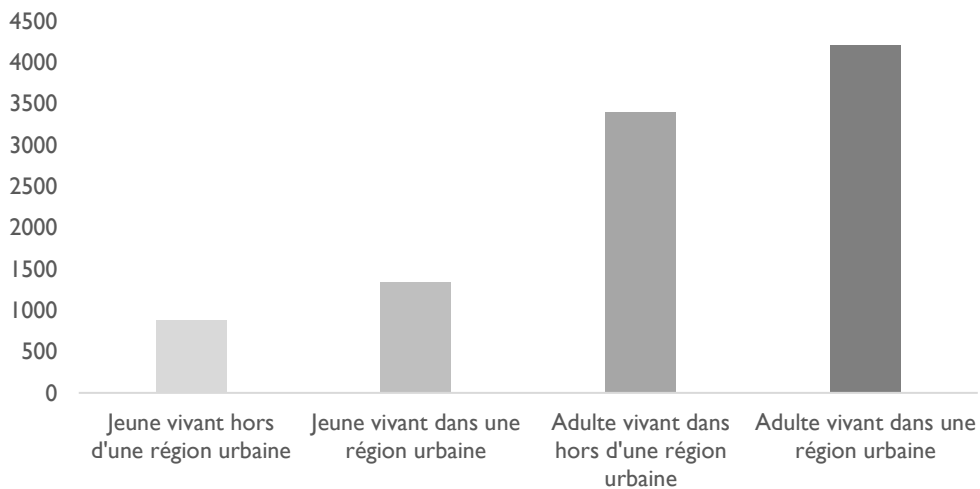
Graphique 10 : Proportion d'alcoolique selon chaque état de santé en %



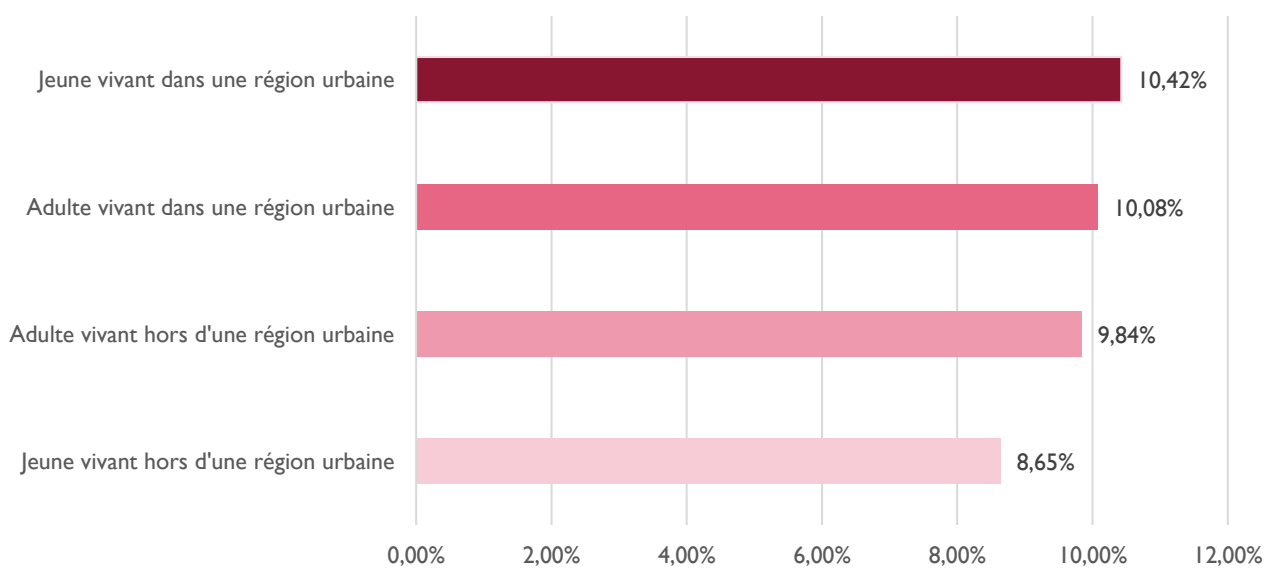
Graphique 11 : Proportion d'alcoolique selon chaque situation salariale en %

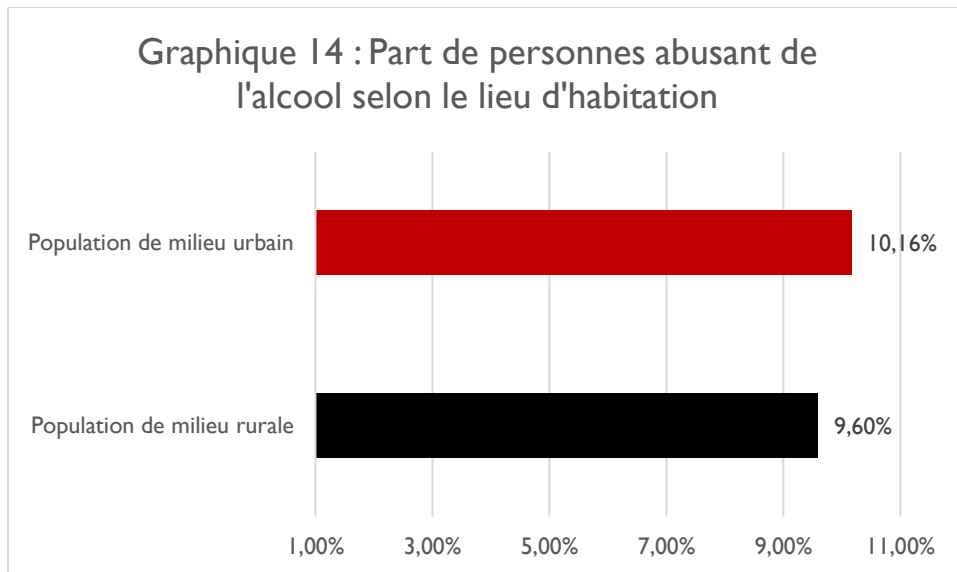


Graphique 12 : Répartition de la population en fonction de l'âge et de la région d'habitation



Graphique 13 : Part d'alcoolique en % selon chaque répartition de la population





Tableaux :

Tableaux

I :

VARIABLE	SIGNES ATTENDUS	SIGNIFICATIVITE
Age	+	*
Sante	-	**
Employ	-	**
Parent alcoolique	-	***
Educ	-	*
Jeune* <u>Outercity</u>	+	**

Tableaux 2 :

	abuse	age	sante	employ	parent~e	educ	jeunex~y
1.	0	37	0	1	0	12	0
2.	0	53	1	1	0	9	0
3.	0	59	1	1	0	11	0
4.	0	43	1	1	1	10	0
5.	0	38	1	1	0	10	0
6.	0	34	0	1	1	10	0
7.	0	31	1	1	0	12	0
8.	0	43	1	1	0	18	0
9.	1	46	1	1	1	12	0
10.	0	36	1	1	1	12	0
11.	0	29	1	1	0	16	0
12.	0	32	0	1	0	15	0
13.	0	37	0	1	0	12	0
14.	0	27	1	1	0	12	0
15.	0	55	0	1	0	11	0

Tableaux 3 :

Individus	Effectif en nombre	Répartition en %
Alcoolique	946	9,86 %
Non Alcoolique	8645	90,14 %

Tableaux 4 :

	Parent alcoolique		Total
Abus d'alcool	0	1	
0	7 179	1 466	8 645
1	698	248	946
Total	7 877	1 714	9 591

Tableaux 5 :

Variable	model1
age	.00006599
sante	-.0842991**
employ	-.07857531
parentalco~e	.28148494***
educ	-.01489937**
jeunexoute~y	-.0814824
_cons	-1.0135243***
<hr/>	
N	9591
pseudo_r2	0.0103

Tableaux 6 :

Variable	model2
sante	-.09060905**
parentalco~e	.27966889***
educ	-.0158234**
centcity	.10199626***
_cons	-1.1073312***

Legend: * $p < .1$; ** $p < .05$; *** $p < .01$

N	9591
pseudo_r2	0.0110

Tableaux 7 :

variable	dy/dx
sante*	- .0159233
parent~e*	.0536017
educ	- .0027047
centcity*	.0178316

Tableaux 8 :

Classified	True		Total
	D	~D	
+	0	0	0
-	946	8645	9591
Total	946	8645	9591
Correctly classified			90.14%

Tableaux 9 :

Number of observations = 9,591
 Number of covariate patterns = 2,185
 Pearson chi2(2179) = 2062.24
 Prob > chi2 = 0.9634