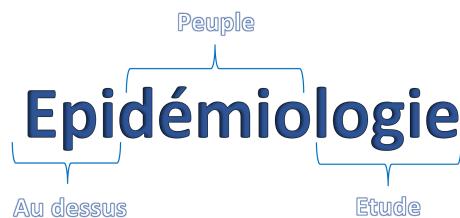

Étude de l'impact de l'activité professionnelle sur la santé perçue
des sujets de la cohorte EPP3.

INSERM U970 PARCC (Paris Centre Cardiovasculaire)
E04 : Integrative Epidemiology of Cardiovascular Diseases

01/04/2019 - 31/08/2019



Odélia Guedj
M1 Mathématiques en Interaction
Université Evry Val d'Essonne

Dr Jean Philippe Empana
Research Director, MD, PhD
INSERM U970 PARCC

Table des matières

Introduction	1
Remerciements	1
Citations	1
1 Contexte	2
1.1 L'Inserm U970 Équipe 4	2
1.2 L'étude Parisienne Prospective 3 (EPP3)	2
2 Bref aperçu de l'épidémiologie	4
2.1 Histoire et définition	4
2.2 Notions de base, vocabulaire	5
2.3 Principaux biais de l'épidémiologie	7
3 Description et explication des données	7
3.1 Variable d'exposition : Activité Professionnelle	7
3.1.1 Codage manuel	8
3.1.2 ACM et CAH	10
3.1.3 Comparaisons des deux codages	17
3.2 Variable d'intérêt : Self Rated Health (SRH)	22
3.3 Variables d'ajustement et facteurs de confusion	23
3.4 Données Manquantes	27
4 Statistiques	28
4.1 Flowchart	28
4.2 Analyses univariées	28
4.3 Analyses multivariées	33
4.3.1 Sélection des variables	33
4.3.2 Diagnostique du modèle	36
4.4 Analyse en sous-groupes	38
4.5 Analyse de sensibilité	38
5 Perspectives : Ce que j'ai fait, découvert, appris	40
6 Annexes	42
6.1 Annexe 1 : Questionnaire IPC	42
6.2 Annexe 2 : Codes travail IPC	50
6.3 Annexe 3 : Classification des codes IPC par activité professionnelle	54
6.4 Annexe 4 : Graphes des résidus du modèle de sélection de variables	55
6.5 Annexe 5 : Fiche d'appréciation de l'employeur	56

Remerciements

Je tiens à remercier l'ensemble des personnes ayant contribué au succès de ce stage.

Tous d'abord, merci au Dr Jean Philippe Empana de m'avoir fait confiance en m'accueillant dans son équipe. J'y ai énormément appris tant sur le plan statistique que sur le plan humain. J'ai découvert avec beaucoup de plaisir le monde de la recherche où j'espère un jour me faire une place ! Merci d'avoir toujours pris le temps de répondre à mes questions et surtout de m'avoir laissé assez de liberté pour que je découvre des choses par moi-même.

Je tiens également à remercier Marie Lê Hoang, Ingénieur d'étude à l'INSERM qui a gentiment accepté de faire circuler mon CV dans l'unité et grâce à qui j'ai eu la chance d'avoir 3 propositions de stages.

Un grand merci à toute l'équipe 4 de l'unité pour ces 5 mois passés ensemble : des heures de discussions méthodologie, statistiques, éthique et médecine. Des heures à dompter des bouts de codes capricieux qui ne font pas ce qu'on leur demande. Merci pour cette ambiance de travail dans la bonne humeur : Bamba Gaye (MD, Phd), Marie-Aude Penet (MD, MSc), Prunelle Getten (MD, MSc), Willy Sutter (MD, PhD), Delphine Lavignasse (PhD), Anouk Asselin (MSc), Marie-Cécile Perrier(MSc), Lucile Ofredo (MSc), Radia B(ARC) ainsi que tous les apprentis chercheurs de passage.

Enfin, merci à Mme Agathe Guilloux et Mme Marie Luce TAUPIN, Enseignants Chercheurs à l'Université d'Évry Val d'Essonne pour leur aide durant cette année un peu particulière. Merci pour votre compréhension et pour le temps que vous m'avez accordé.

Citations

Nous ne devons pas laisser croire que tout progrès scientifique peut être réduit à des mécanismes, des machines, des rouages, quand bien même de tels mécanismes ont eux aussi leur beauté.

Madame Curie, Ève Curie, éd. Gallimard, 1938

The new form of the problem can be described in terms of a game which we call the 'imitation game.'

It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex.

The interrogator stays in a room apart front the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either "X is A and Y is B" or "X is B and Y is A."

The interrogator is allowed to put questions to A and B...

We now ask the question, "What will happen when a machine takes the part of A in this game?" Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman?

These questions replace our original, "Can machines think?"

Mechanical Intelligence : Collected Works of A.M. Turing

Introduction

Les retraités : leur nombre, leur âge, le montant de leur pension sont des sujets récurrents des derniers mandats présidentiels français.

Les récentes législations[1] concernant le recul de l'âge légal du départ à la retraite se sont vues opposer de violentes résistances qui posent la question de la santé des individus actifs par rapport à celle des retraités[6] .

La santé perçue, Self Rated Health en anglais, est une mesure subjective de la santé des individus souvent utilisée en santé publique.

Cette mesure est effectuée grâce à une question directe aux individus : Sur une échelle de 0 à 10, comment évaluez vous votre santé (avec 0 : très mauvaise et 10 excellente).

La santé perçue est décrite dans la littérature comme un indicateur consistant de mortalité[10] [4] ayant des liens fort avec la santé mentale [9] et le contexte socio-économique des individus [8] [2]. Si la santé perçue reflète assez bien la santé objective d'individus [15] , il est important de souligner que sa nature subjective en fait un outils plus puissant encore puisqu'il permet de prendre en compte des critères psychologiques et sociaux comme les habitudes alimentaires, l'historique familial ou les disposition à la sur/sous évaluation d'un risque.

Cet outils a été utilisé dans l'étude de la cohorte GAZEL pour mettre en évidence l'hypothèse suivante[14] : les individus actifs voient leur santé perçue augmenter après avoir pris leur retraite. Pour ce faire les investigateurs disposaient de plusieurs points de mesure avant et après la retraite des sujets de la cohorte GAZEL.

L'objectif de mon stage était de m'inspirer des hypothèses de l'étude décrite ci-dessus pour montrer que la santé perçue des individus de l'Etude Parisienne Prospective 3 [5] est impactée par leur activité professionnelle.

1 Contexte

1.1 L'Inserm U970 Équipe 4

L'INSERM : Institut National de la Santé Et de la Recherche Médicale est un établissement public de recherche créé en 1964 par le ministre de la santé Raymond Marcellin et dont la santé Publique et l'épidémiologie sont un des domaines de spécialité.

L'INSERM est constitué de plusieurs dizaines d'unités réparties sur l'ensemble du territoire français et dont le siège se trouve rue Tolbiac dans le 13 ème arrondissement de Paris.

J'ai effectué mon stage au PARCC (Paris Centre Cardiovasculaire) : une unité mixte de recherche de l'INSERM (U970) située dans le bâtiment recherche de l'hôpital George Pompidou (Rue Leblanc Paris 15) depuis 2009.

Cette unité est composée d'une dizaine d'équipes ayant une "vocation¹ internationale dans le domaine de la recherche fondamentale et translationnelle sur les maladies cardiovasculaires, à partir d'approches multiples combinant biochimie, biologie cellulaire et moléculaire, imagerie moléculaire, physiologie intégrative, pharmacologie, génétique et épidémiologie".

Durant mon stage j'ai intégré l'équipe 4 : Integrative Epidemiology of Cardiovascular Diseases, co-dirigée par le Dr Jean Philippe Empana et le Dr Xavier Jouven dont les sujets de recherche peuvent être résumés en quatre points :

1. CEMS : Centre d'Expertise de la Mort Subite. Il s'agit de la création d'une base de données recensant de façon exhaustive les événements de mort subite de Paris et de la petite couronne.
2. Approche multi-marqueurs à la détection de nouveaux facteurs de risque aux maladies cardiovasculaires.
3. Épidémiologie des maladies cardiovasculaires dans les pays en voie de développement.
4. The Paris Transplant Group : Epidémiologie de l'immuno-atherosclerosis.

Ma recherche est inclue dans le deuxième point ci dessus.

L'équipe est composée de médecins, de statisticiens, d'attachés de recherche clinique (ARC) et d'étudiants (thèse de science, stage M1/M2, post-doc).

On y trouve une grande diversité de nationalités, de parcours, de spécialités donnant une vue d'ensemble sur le monde de la recherche en santé/bio-statistiques et créant une émulsion qui m'a beaucoup profité.

1.2 L'étude Parisienne Prospective 3 (EPP3)

EPP3 [5] est une étude prospective de cohorte en population générale comptant n = 10157 sujets. Ces derniers ont été recrutés dans des centres IPC (centres d'examens de santé conventionnés par l'assurance maladie) entre Juin 2008 et Décembre 2011.

1. Site Web de l'Hôpital Européen George Pompidou

Pour entrer dans l'étude, les sujets doivent avoir entre 50 et 75 ans.

Les données sont récoltées via des questionnaires envoyés tous les deux ans qui, à l'exception du questionnaire d'inclusion sont élaborés à l'INSERM U970 équipe 4.

Le questionnaire d'inclusion² provient de l'IPC. Il contient une partie socio-administrative, des questions sur l'environnement professionnel des sujets ainsi que sur leurs habitudes de vie (alimentation, tabac, alcool). Une autre partie du questionnaire traite des antécédents médicaux des sujets, tant familiaux que personnels, de leur état de santé actuel et de leurs prescriptions médicamenteuses. La dernière partie traite du bien-être des sujets : on y trouve des questions sur leur stress perçu, leur équilibre mental ainsi que leur nutrition.

Le but de l'étude est la recherche de facteurs de risque pour les maladies cardiovasculaires.

Dans chacun des questionnaires il est demandé aux sujets de déclarer leurs hospitalisations en détaillant le motif d'hospitalisation, le nom de l'hôpital, du service où ils ont été traités ...

Chaque hospitalisation déclarée est appelée évènement.

Pour s'assurer de la validité des déclarations des sujets, un protocole de validation d'événement a été mis en place : régulièrement un certain nombre d'événements sont extraits de la base. Pour chaque évènement, on contacte l'hôpital pour qu'il transmette au responsable de l'étude les comptes-rendus hospitaliers (CRH) de l'évènement en question. Ensuite les CRH sont lus par un médecin qui valide, invalide ou corrige le diagnostique déclaré par le sujet.

L'étude EPP3 est une étude longue, il est prévu que le suivi dure 20 ans (10 questionnaires à raison d'un tous les deux ans).

Elle a déjà donné lieu à la publication de plus d'une vingtaine d'articles dans des revues prestigieuses comme le JACC (Journal of the American College of Cardiology) ou le JAMA (Journal of the American Medical Association).

2. Annexe 1

2 Bref aperçu de l'épidémiologie

2.1 Histoire et définition

Épidémiologie vient des mots grecs "epi" : au-dessus, "demio" : peuple. Ainsi la définition étymologique de l'épidémiologie est l'étude des peuples.

Si l'on tente de donner une définition plus précise de ce qu'est l'épidémiologie on se retrouve vite enseveli sous le nombre de possibilités.

Le dictionnaire Larousse propose la définition suivante : *Science qui étudie, au sens des populations (humaines, animales voir végétales), la fréquence et la répartition des problèmes de santé dans le temps et dans l'espace, ainsi que le rôle des facteurs qui les déterminent.*

Cependant deux autres définition méritent d'être citées pour leur pertinence (préférence subjective bien entendu). Celle Mac Mahon : *Étude de la distribution et des déterminants d'une maladie dans des populations humaines, et application des résultats de cette étude dans la lutte contre cette maladie*, et celle de Jeniceck : *L'épidémiologie est un raisonnement et une méthode propres au travail objectif en médecine et dans d'autres sciences de la santé, appliqués à la description des phénomènes de santé, à l'explication de leur étiologie, et à la recherche des méthodes d'intervention les plus efficaces.*

Ce cocasse problème de définition est cependant assez sérieux pour que des épidémiologistes décident de mener leur enquête [7] sur les définitions données à leur discipline entre 1978 et 2017. Ils en répertorient 102 et identifient 5 termes présent dans plus de la moitié des définitions répertoriées : "population", "étude", "santé", "maladie" et "distribution".

Ils observent également une augmentation des mots "contrôle" et "santé" dans les définitions entre 1978 et 2017. Il est à noter que les définitions autorisées à être incluses dans l'étude sont soit en anglais soient des définitions nationales traduites en anglais.

Les 5 mots cités précédemment rendent assez bien compte de ce qu'est réellement la discipline, preuve que l'étude menée sur ses définitions n'était pas vainne puisqu'elle permet de comprendre en profondeur ce qu'est l'épidémiologie (santé, maladie, étude), en quoi elle se détache de la médecine (population, distribution) et pourquoi elle est utile (contrôle, maladie).

Forts de ces éclaircissements, nous pouvons sans trop nous tromper, faire l'hypothèse que l'épidémiologie est aussi vieille que la médecine. Depuis la discipline a beaucoup évolué, s'adaptant aux nouveaux défis posés par l'apparition de maladies.

Dans un article de médecine science publié en 2016 [3], Philippe Bizouarn³ identifie 4 aires distinctes de l'épidémiologie.

La première appelée statistiques sanitaires repose sur la théorie des miasmes (émanations fétides du sol, de l'air, de l'eau) et occupe la deuxième moitié du XIX^{ème} siècle.

La seconde, ère des maladies infectieuses, s'étend de la deuxième moitié du XIX^{ème} siècle à la

³. Épidémiologue, Praticien hospitalier - Service d'Anesthésie-Réanimation de l'Hôpital G et R Laennec, CHU de Nantes

deuxième moitié du XX^{ème} siècle. Elle rejette la théorie des miasmes au profit de la théorie des germes avec la découverte du microbe.

Un exemple historique d'étude épidémiologique de cette ère est celle que J.Snow effectue en 1954 afin de mettre en évidence le mode de propagation du choléra. Il découvre un lien fort entre le nombre de cas de choléra dans une zone et la distribution d'eau qui la dessert. Il arrive à prouver qu'une des pompes d'eau occasionne le plus grand nombre de cas de choléra, la fait retirer et endigue ainsi l'épidémie. L'eau est alors communément admise comme voie de propagation de la maladie.

La troisième ère de l'épidémiologie identifiée par P.Bizouarn (deuxième moitié du XX^{ème} siècle jusqu'au début du XXI^{ème}) est celle des maladies chroniques dont une des études les plus connues est celle de Dull et Hill datant de 1950 sur le cancer bronchopulmonaire.

Enfin il appelle la dernière ère, l'ère de l'éco-épidémiologie caractérisée par la recherche et l'analyse des associations entre plusieurs "facteurs de risques" (par exemple l'exposition à une maladie, le tabac, l'alcool...) et les "issues" (maladies) sans rechercher systématiquement de "lien causal" entre un des facteurs et les issues.

2.2 Notions de base, vocabulaire

L'épidémiologie utilise des méthodes statistiques afin de décrire une maladie ou d'en rechercher les causes. Cependant, le vocabulaire utilisé diffère quelque peu de celui utilisé en mathématiques. Ainsi la variable à expliquer, Y , est souvent appelée variable d'intérêt.

De plus, en épidémiologie, les variables explicatives, ou covariables, sont classées dans plusieurs groupes :

- Il y a d'une part la variable d'exposition c'est à dire la variable explicative principale.
- Les facteurs de confusion : ce sont les covariables qui sont incluses de façon (quasi) systématiques dans les modèles statistiques pour des raisons cliniques. Ces variables sont décrites dans la littérature et varient selon la branche médicale. On y retrouve souvent l'âge, le sexe, l'indice de masse corporelle des sujets ainsi que leur habitudes de vie : tabac, alcool, score de dépression, statut marital...
Les facteurs de confusion sont des variables à la fois liées au facteur de risque et à la variable d'intérêt.
- Enfin les autres variables explicatives sont appelées variables d'ajustement : il s'agit d'autres facteurs pouvant participer, avec la variable d'exposition, à l'explication de la variable d'intérêt.

L'épidémiologie a deux champs principaux d'application :

- *L'épidémiologie classique* qui décrit et mesure des phénomènes de santé dans une population. Elle a pour but l'élaboration de stratégies de santé publique.
- *L'épidémiologie clinique* dont les études se concentrent sur des populations de patients en vue d'améliorer les connaissances d'une maladie ou de tester l'efficacité d'un traitement.

L'épidémiologie se décompose en trois branches :

1. *L'épidémiologie descriptive* qui consiste en l'étude de la fréquence et de la répartition de phénomènes de santé. Elle permet de formuler des hypothèses quant aux causes de ces phénomènes.
2. L'épidémiologie analytique dont le but est de vérifier les hypothèses susmentionnées. Elle recherche les liens entre l'exposition à un facteur et la survenue d'un phénomène de santé.
3. *L'épidémiologie évaluative* qui mesure l'efficacité d'une intervention (thérapeutique ou de prévention) sur le phénomène de santé.

Une étude épidémiologique peut se faire deux manières :

- En observant les effets sur une population de l'exposition de facteurs de risques : on parle *d'observation*.
- En contrôlant les conditions d'exposition à ces facteurs de risque : on parle *d'étude expérimentale*.

L'étude est dite *randomisée* si le hasard seul est responsable de l'appartenance d'un sujet à un groupe. On parle d'une *étude ouverte* quand le traitement est connu des sujets et des investigateurs, *en simple aveugle* quand seuls les investigateurs connaissent le traitement et *en double aveugle* lorsque ni les sujets ni les investigateurs ne connaissent le traitement.

Cependant les études expérimentales sont presque toujours irréalisables car non éthiques : une étude expérimentale sur le lien entre le tabac et l'apparition d'un cancer du poumon supposerait en effet d'exposer une partie de la population étudiée au tabac puis d'étudier les conséquences sur leur poumon.

Lorsque l'on fait de l'épidémiologie descriptive on peut faire deux sortes d'étude :

- Étudier la *prévalence* d'un phénomène de santé, c'est à dire la fréquence de survenue de ce phénomène dans une certaine population à un temps donné. Dans ce type d'étude l'évolution temporelle n'existe pas, c'est une mesure effectuée à un temps donné : on parle *d'étude transversale*.
- Étudier l'*incidence* d'un phénomène de santé, c'est à dire étudier les modifications de l'état de santé d'un ou de plusieurs groupes de sujets sur une période donnée. Ici c'est bien l'évolution dans le temps d'un état qui importe : on parle *d'étude longitudinale*.

Par ailleurs lorsque le but de l'étude épidémiologique est à visée étiologique on dénombre deux grandes familles d'étude :

- *L'étude de cohorte* : Étude d'un groupe de personnes étant ou pouvant être exposé à un facteur de risque. C'est l'étude épidémiologique la plus exhaustive.
- *L'étude cas-témoin* : elle étudie la comparaison entre deux groupes : les malades appelés cas et les non-malades appelés témoins. Le but est de déterminer si la réponse à l'exposition à un facteur est similaire dans les deux sous-groupes et ainsi d'établir un lien facteur/maladie.

Enfin, une étude peut être *prospective* si l'information concernant l'exposition des sujets à un facteur est recueillie avant la survenu d'un phénomène de santé, ou *rétrospective* si l'information concernant l'exposition des sujets à un facteur est recueillie après la survenu d'un phénomène de santé chez certains sujets.

2.3 Principaux biais de l'épidémiologie

Un biais est une erreur de méthodologie commise le plus souvent durant l'inclusion des sujets dans l'étude et conduisant à une mauvaise estimation des paramètres étudiés.

Il en existe 3 principaux :

- Le *biais de sélection* : : Intervient lors de la constitution de l'échantillon d'enquête. C'est le biais induit par la manière dont les sujets sont choisis au sein de la population.
- Le *biais de mesure* : due à une mauvaise mesure du facteur d'exposition.
- Le *biais de confusion* : due à la mauvaise analyse de la relation d'un facteur avec une issue. Un tel facteur est appelé facteur de confusion.

3 Description et explication des données

Le but de cette analyse est d'établir l'existence d'un lien entre l'activité professionnelle des sujets d'EPP3 et la manière dont ils perçoivent leur santé. Si cette dernière fait l'objet d'une question claire dans le questionnaire d'inclusion, ce n'est pas le cas de l'activité professionnelle. J'ai ainsi dû créer une nouvelle variable en extrayant de l'information de plusieurs questions grâce à des méthodes de classifications non supervisées.

La description des méthodes utilisées à la mise en forme de l'ensemble des variables utiles à l'analyse fait l'objet de cette section.

3.1 Variable d'exposition : Activité Professionnelle

Dans le questionnaire d'inclusion, 3 questions traitent de l'activité professionnelle des sujets (Question 1,2,3 du questionnaire IPC en annexe). Ces 3 questions ont été codées en 4 variables catégorielles :

- Adm12 : Êtes-vous
 - 6NNNNN contrat emploi-solidarité, intérim, CDD
 - N5NNNN chômeur depuis + de 6 mois
 - NN4NNN chômeur depuis - de 6 mois
 - NNN3NN à la recherche d'un emploi
 - NNNN2N jeune en cours de formation
 - NNNNN1 étudiant
- Adm12a : Êtes-vous
 - JXXX en formation professionnelle
 - X9XX au foyer
 - XX8X retraité(e)
 - XXX7 pré-retraité(e)
- Adm11 : Depuis quand n'exercez-vous plus d'activité professionnelle ?
 - 0 en activité

- 1 moins d'un an
 - 2 1 an
 - 3 2 ans
 - 4 3 ans ou +
 - 5 jamais travaillé
- Adm10 : Si vous travaillez quelle est votre profession ?
La réponse à cette question est du texte libre. Après la récupération des questionnaires par l'IPC, un code à deux chiffres est attribué à chaque grand groupe de profession.

L'encodage de la variable Adm10 introduit une première source potentielle d'erreur du fait de la difficulté d'interpréter du texte libre d'une part et des possibles erreurs de "classification humaine" d'autre part. Il faut également noter qu'un des sujets a un code travail de 88 et que ce code ne correspond à aucune des professions de la liste IPC⁴.

Une autre difficulté est due à la possibilité qu'ont les sujet de cocher plusieurs réponses par question. Il en résulte un grand nombre de classes dans chaque variable ce qui a compliqué le codage de la variable activité professionnelle.

3.1.1 Codage manuel

L'objectif du projet étant d'étudier l'impact de l'activité professionnelle sur la santé perçue des sujets, il faut tout d'abord résumer l'information contenue dans les 3 variables citées précédemment en une seule variable.

J'ai donc créer une variable catégorielle qui comporte 5 classes :

- R pour retraités
- T pour travailleur
- C pour chômeur
- I pour inactif
- NSP pour ne sais pas (cas "inclassables")

Pour cela j'ai fait le choix de me baser de manière successive sur les variables Adm12a puis Adm11 puis Adm12 puis Adm10. La raison en est simple : c'est la variables Adm12a qui propose la réponse "Retraités", or c'est la catégorie qu'il m'intéresse le plus d'étudier. Ensuite la variable Adm11 nous indique si le sujet travaille encore OU depuis combien de temps il a cessé de travailler. La variable Adm12 discrimine les chômeurs. Enfin, la variable Adm10 est utile pour vérifier la cohérence des différentes réponse des sujets ou bien de trancher dans des cas où les autres variables ne fournissent pas d'informations suffisantes.

Par exemple, il y a 57 individus pour lesquels les variables Adm12 et Adm12a ne sont pas renseignées et qui déclarent ne plus travailler depuis moins d'un an ou plus (c'est à dire qu'ils ont cochés les cases 1,2,3,4 ou 5 du questionnaire). Il est possible que ces sujets soient au chômage, à la retraite ou qu'ils n'aient jamais travaillé.

4. Annexe 2

Règle de décision :

Les sujets indiquant qu'ils sont à la retraite et qu'ils travaillent sont classés en tant que travailleurs et ce, même si leur code travail (c'est à dire la variable Adm10) est de type "Ancien X" (codes 71 à 78 de la liste IPC) car il est possible de percevoir une pension de retraite et de continuer à travailler (par exemple en tant qu'expert).

Or ce qui est intéressant pour l'étude est la santé perçue des sujets ne travaillant plus du tout. Il vaut donc mieux classer ceux qui complètent leur retraite avec un emploi en tant que travailleurs qu'en tant que retraités.

Ainsi, sont étiquetés inactifs les sujets n'ayant jamais travaillé, ou déclarant être au foyer ou ayant un code travail parmi les deux suivants : 85/86. Ces codes correspondent à des individus inactifs non retraités respectivement de moins de 60 ans et de plus de 60 ans.

Si les variables Adm12 OU Adm10 indiquent que le sujet est chômeur (en incluant la simple recherche d'emplois), il est classé en chômeur.

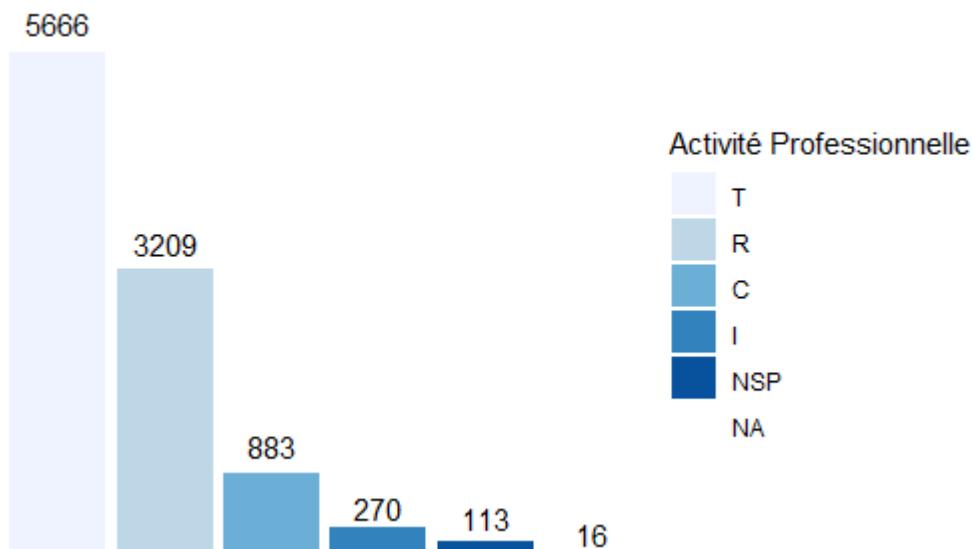
En effet, on peut raisonnablement faire l'hypothèse que la précarité induite par la recherche d'un emploi occasionnera une plus mauvaise santé perçue.

On classe retraités tous les sujets n'ayant aucun indicateur de chômage ou d'activité professionnelle et n'étant pas inactifs.

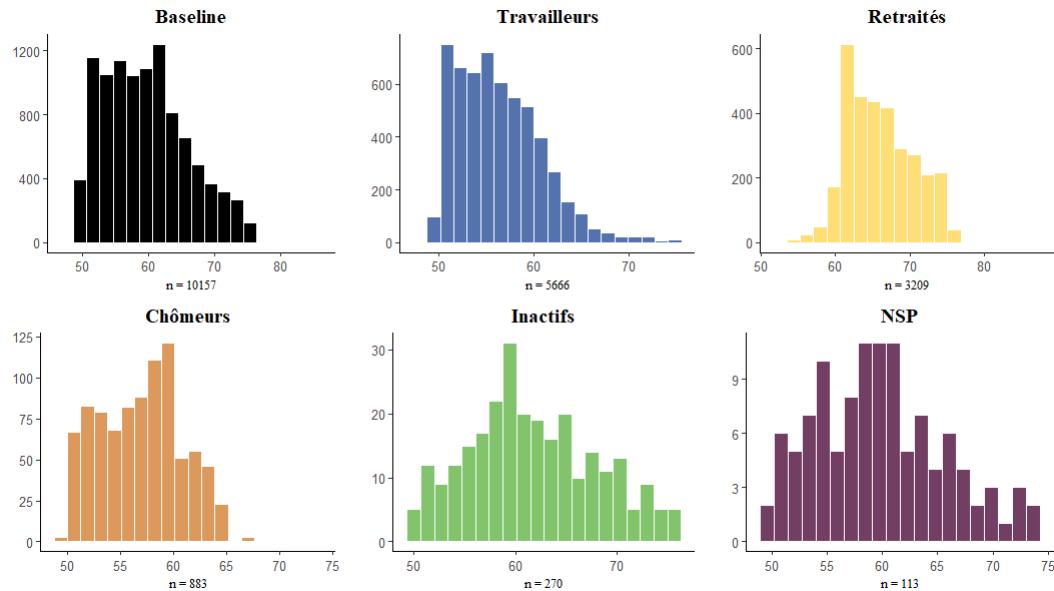
Enfin, le reste est classé comme travailleurs.

Malgré toutes les étapes de la classification et la minutie des vérifications, il existe un certain nombre de cas "impossibles" à classer car les informations d'un sujet pour les différentes variables sont contradictoires. Ces sujets sont étiquetés NSP.

Finalement on obtient la répartition suivante :



Pour vérifier la cohérence des classes obtenues on peut tracer les distributions de l'âge dans chaque classe.



On remarque que 1116 sujets classés Travailleurs ont moins de 60 ans (on peut raisonnablement supposer que l'âge de la retraite est 60 ans car les sujets de la cohorte ont, à l'inclusion, un âge compris entre 50 et 75 ans). De même, 134 sujets sont classés en tant que Retraités et ont moins de 60 ans.

Ceci peut s'expliquer soit par une erreur de classification : ayant effectué cette dernière à la main il y a un risque non négligeable que je n'ai pas appliqué exactement le même critère de jugement pour chacun des cas. Il peut aussi ne pas s'agir d'une erreur, dans ce cas les données sont ainsi et on veillera simplement à garder cela à l'esprit lorsque nous interpréterons les résultats des tests statistiques ultérieurs.

3.1.2 ACM et CAH

Pour vérifier la cohérence des résultats présentés au paragraphe précédent, j'ai effectué une Classification Ascendante Hiérarchique sur les résultats d'une Analyse des Correspondances Multiples appliquée à une base de données constituée des quatre variables utilisées pour la classification "manuelle" : Adm12a, Adm11, Adm12 et Adm10.

J'ai également veillé à supprimer les individus ($n = 29$) ayant une valeur manquante pour l'une de ces quatre variables.

Les résultats de l'ACM sont rassurants, dans le sens où l'emplacement dans le plan factoriel des quatre variables qui nous intéressent correspond à l'intuition qu'on en avait à savoir :

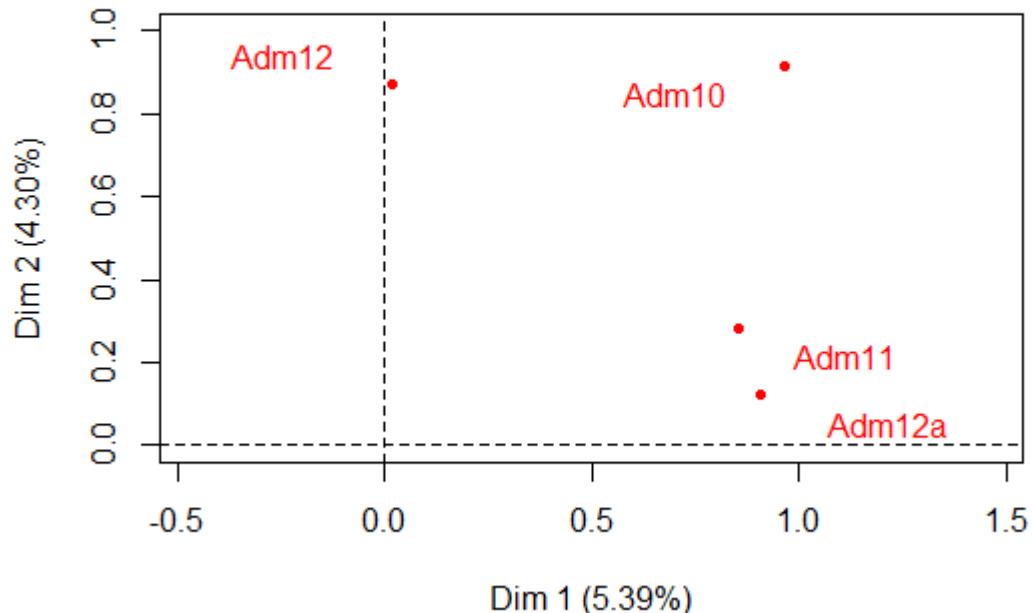
- Les variables Adm12a et Adm11 apportent globalement la même information, cohérent puisque Adm12a et Adm11 discriminent les retraités et les travailleurs.
- Adm12 et Adm10 sont sur le même plan horizontal : elles discriminent les chômeurs (avec les

codes travail 91 à 96).

— Adm12a, Adm10 sont dans le même plan vertical : elles discriminent les inactifs.

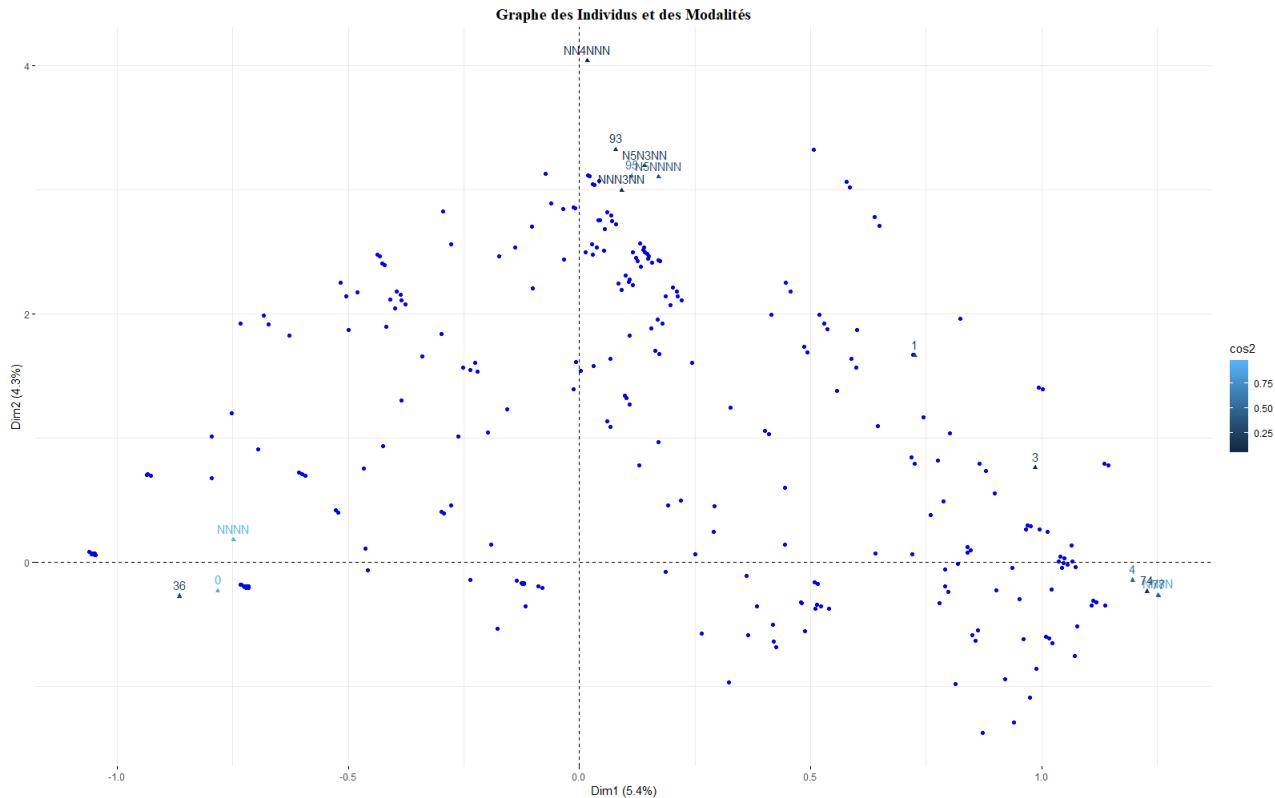
On s'attend donc à ce que la CAH effectuée sur les résultats de l'ACM produise une classification proche de celle effectuée à la main.

Graphe des Variables



NB : Les faibles pourcentages d'inertie sont tout à fait normaux dans le cadre d'une ACM. En effet, comme l'explique Jérôme Pagès dans son livre "Analyse factorielle multiple avec R", si les variables étaient toutes identiques dans le cas d'une ACP la première dimension aurait 100% d'inertie alors que dans une ACM la première dimension aurait au maximum $\frac{100}{(nb_{modalités}-1)}$ % d'inertie.

Sur le graphe suivant on affiche les individus ainsi que les 15 modalités ayant la plus grande contribution. Ces 15 modalités sont affichées selon un gradient de couleur en fonction de leur \cos^2 .



On voit de façon assez claire trois groupes distincts : le premier en bas à droite est certainement celui des sujets retraités, on y retrouve les modalités 1,3,4 de Adm11 (ne travaille plus depuis resp, moin de 1 an, deux ans, trois ans et plus), 74 de Adm 10 (Ancien cadre), NN8N de Adm12a (Retraité).

Le groupe en haut correspond sans doute aux chômeurs puisque les modalités NNN3NN, NN4NNN et N5N3NN de Adm12 (resp. A la recherche d'un emploi , Chômeur depuis moins de 6 mois et A la recherche d'un emplois + Chômeur depuis 6 mois) et 93 de Adm10 (cadres et professions intellectuelles chômeurs).

Enfin le dernier groupe est celui en bas à gauche, il représente sûrement les travailleurs puisque la modalité 0 de Adm11 s'y trouve (Travaille) ainsi que 36 de Adm10 (Cadres) et NNNN de Adm12a. La présence de cette dernière modalité fait sens puisqu'elle représente les individus n'ayant rien coché dans la variable Adm12a doit les choix étaient pré-retraités, retraités, personne au foyer, en formation professionnelle. Or si ces sujets travaillent, aucune de ces catégories ne les concerne.

Il reste deux choses qu'il me semble pertinent de relever : l'ACM semble particulièrement discriminer les cadres, on les retrouve dans les chômeurs, les retraités et les travailleurs. Ceci s'explique probablement par le fait que les modalités 36, 74 et 93 (resp Cadres d'entreprises, Anciens cadres

et Cadres et professions intellectuelles chômeurs) représentent à elles seules 3848 sujets (sur 10157 sujets dans la base globale et 10128 dans la base ayant permis l'ACM).

La deuxième chose à noter est l'absence d'un quatrième groupe qui aurait représenté les inactifs. Une explication possible est la faible proportion de ces sujets dans la base. En effet lors de la classification manuelle on n'avait étiqueté "que" 270 sujets comme inactifs.

La question que je me suis ensuite posée est celle de l'allure qu'aurait ma classification si elle était effectuée par un algorithme. De plus, il serait intéressant d'analyser la classification des individus que je n'ai pas su classer. Pour cela j'ai effectué une Classification Ascendante Hiérarchique avec le package FactoMineR de R en choisissant la distance du χ^2 et le critère d'agrégation de Ward.

Distance du χ^2 :

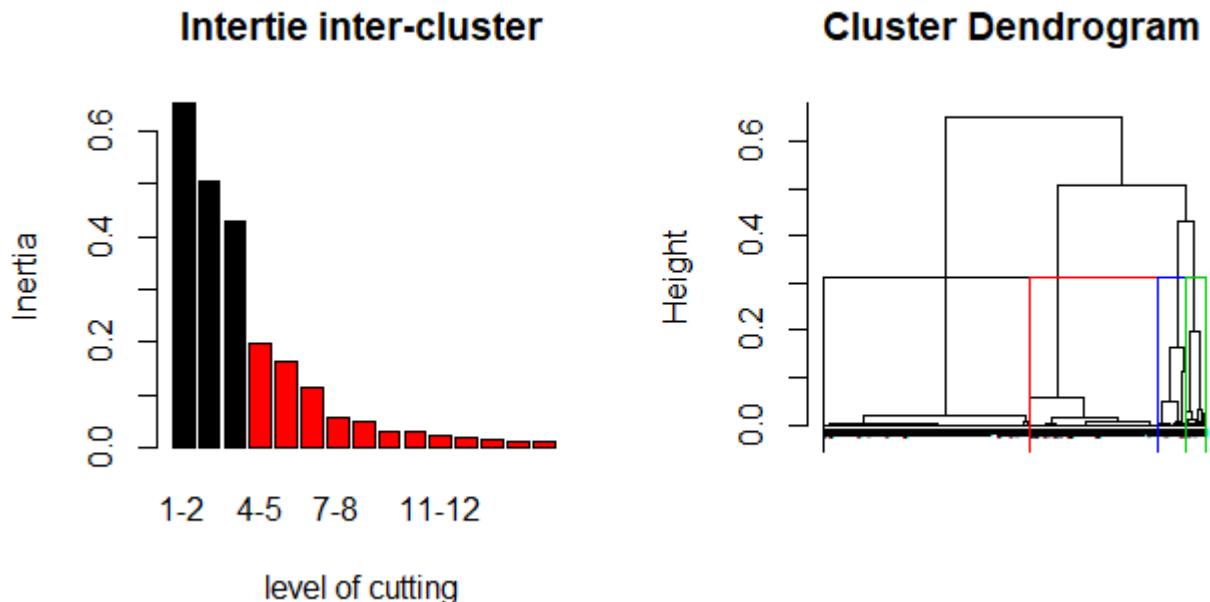
La distance entre un individu i et un individu j représentés par p variables ayant m_1, m_2, m_P modalités est donnée par :

$$d_{\chi^2}(i, j) = \sqrt{\sum_k \frac{np}{n_{.k}} \left(\frac{x_{ik} - x_{jk}}{p} \right)^2}$$

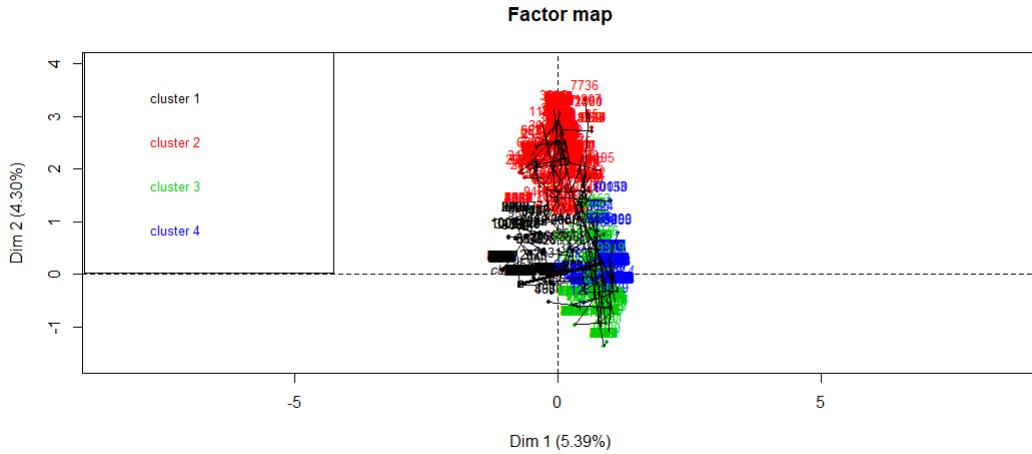
La distance du χ^2 traduit le fait que deux individus ayant en commun une modalité rare sont plus proches que deux individus ayant en commun une modalité fréquente.

Méthode de Ward :

C'est la méthode d'agrégation la plus courante. Elle consiste à choisir la réunion de deux groupes qui fera le moins baisser l'inertie entre les groupes, le but étant d'obtenir des groupes les plus "séparés" possible.



Le graphe d'inertie semble indiquer que le nombre idéal de clusters est 4 ce qui est bon signe puisque l'idée était de répartir tous les sujets en 4 groupes : travailleurs, retraités, chômeurs et inactifs.



Pour pouvoir interpréter les Cluster obtenus par la CAH il est indispensable de comprendre quelles sont les modalités qui contribuent le plus à chaque Cluster.

Pour cela on cherche à déterminer si la forte présence d'une modalité dans un Cluster est due au hasard. Dans le cas contraire cela signifie que la modalité est représentative du Cluster.

On effectue le test suivant pour chaque modalité dans chaque Cluster :

H_0 : La proportion de la modalité m dans le Cluster c est due au hasard.

H_1 : La proportion de la modalité m est anormalement élevée/basse dans le Cluster c .

Ainsi sous H_0 on a :

$$\frac{n_{mc}}{n_c} = \frac{n_m}{n}$$

où :

n_{mc} est le nombre de sujet du Cluster c présentant la modalité m ,

n_c est le nombre de sujets dans le Cluster c ,

n_m est le nombre de sujets présentant la modalité m ,

n est le nombre total d'individus.

Sous H_0 : $n_{mc} = \frac{n_m \cdot n_c}{n}$

Or le rapport $\frac{n_{mc}}{n}$ suit une loi hypergéométrique.

Soit N_{mc} la variable aléatoire représentant le nombre d'individus de la modalité m dans le Cluster c .

$$N_{mc} \underset{H_0}{\sim} \mathcal{H}(n, n_c, \frac{n_m}{m})$$

La fonction catdes() du package FactoMineR nous résume les résultats de ce test pour chaque modalité dans chacun des Clusters en classant les modalités dans l'ordre décroissant de leur contribution

au Cluster. Seuls les test significatifs sont affichés.

Dans les résultats affichés juste après, la première colonne correspond au nom de la modalité, la deuxième calcule le rapport $\frac{n_{mc}}{n_m}$, la troisième le rapport $\frac{n_m}{n_c}$, la quatrième le rapport $\frac{n_{mc}}{n}$. La cinquième colonne est le calcul de la p-valeur :

$$\mathbb{P}_{H_0}(N_{mc} \geq n_{mc,obs}) = \mathbb{P}_{\mathcal{H}(n,n_c, \frac{n_m}{n})}(N_{mc} \geq n_{mc,obs}) \quad (1)$$

Enfin, la dernière colonne est la valeur de la statistique de test calculée. Lorsqu'elle est positive la modalité à laquelle elle est associée est sur représentée dans le Cluster et lorsqu'elle est négative le modalité en question est sous représentée dans le Cluster.

Par souci de lisibilité nous n'afficheront que les modalités sur-représentées dans chaque Cluster :

Cluster 1	Cla/Mod	Mod/Cla	Global	p.value	v.test
Adm12a=NNNN	87.25	99.91	61.73	0	Inf
Adm11=0	92.41	99.29	57.92	0	Inf
Adm10=36	99.88	46.21	24.94	0	Inf
Adm10=54	99.62	14.41	7.80	3.80e-217	31.45
Adm12=NNNNNN	57.94	97.60	90.82	7.30e-155	26.51
Adm10=51	100	9.03	4.87	1.12e-137	24.98
Adm10=32	99.78	8.19	4.42	3.78e-122	23.50
Adm10=61	100	6.01	3.24	9.09e-91	20.20
Adm10=56	100	3.86	2.08	3.58e-58	16.08
Adm10=66	100	3.59	1.94	4.94e-54	15.48
Adm10=47	99.28	2.53	1.37	2.75e-36	12.58
Adm10=48	100	2.20	1.18	3.43e-33	12.00
Adm10=55	100	1.41	0.76	1.70e-21	9.52
Adm12=6NNNNNN	80.52	2.27	1.52	4.19e-12	6.93
Adm10=41	100	0.77	0.41	4.99e-12	6.91
Adm10=46	100	0.51	0.28	2.97e-08	5.54

Les modalités de la variable Adm10 sur représentées dans le **Cluster 1** correspondent toutes à des codes travail de sujets qui travaillent.

La modalité la plus présente dans ce Cluster est Adm12a = NNNN, c'est à dire les sujets n'ayant coché aucune modalité de cette variable. Cela fait sens puisqu'aucune des modalités possibles ne traitaient d'une activité professionnelle.

Enfin les modalités de la variable Adm12 représentées dans le Cluster 1 sont NNNNNNN c'est à dire ceux n'ayant rien coché et 6NNNNNN : en contrat emplois-solidarité / intérim / CDD.

Cluster 2	Cla/Mod	Mod/Cla	Global	p.value	v.test
Adm12=N5NNNN	99.74	45.49	3.79	0	Inf
Adm10=95	98.52	71.14	6.00	0	Inf
Adm10=93	100	17.93	1.49	1.42 e-169	27.76
Adm12=N5N3NN	100	17.81	1.48	2.05 e-168	27.66
Adm12a=NNNN	12.59	93.47	61.73	1.89 e-108	22.12
Adm11=1	37.58	28.38	6.28	1.39 e-103	21.61
Adm12=NNN3NN	90.65	11.52	1.06	1.38 e-94	20.63
Adm12=NN4NNN	100	7.13	0.59	2.13 e-66	17.21
Adm10=96	82.95	8.67	0.87	7.22 e-65	17.01
Adm12=NN43NN	100	4.99	0.41	1.65 e-46	14.32
Adm12a=NNN7	55.56	5.94	0.89	6.43 e-31	11.56
Adm11=2	26.68	13.18	4.11	3.20 e-30	11.42
Adm11=3	24.17	14.61	5.03	5.04 e-29	11.18
Adm10=94	100	1.31	0.11	1.23 e-12	7.10
Adm12=65NNNN	78.57	1.31	0.14	3.63 e-10	6.27
Adm12=6NN3NN	70	0.83	0.099	2.66 e-06	4.70
Adm11=4	10.34	32.54	26.16	1.61 e-05	4.31
Adm12=65N3NN	100	0.48	0.04	4.75 e-05	4.07
Adm12=6NNNNN	17.53	3.21	1.52	2.14 e-04	3.70
Adm12=6N43NN	100	0.36	0.03	5.73 e-04	3.44
Adm12a=JNNN	66.67	0.24	0.03	2.01 e-02	2.32

Dans le **Cluster 2**, les codes travail présents sont 94/93/94 : ils correspondent aux sujets chômeurs. De plus la modalité la plus représentée dans ce Cluster est Adm12 = N5NNNN : sujets chômeurs depuis plus de 6 mois.

Toutes les autres modalités de Adm12 présentent dans le Cluster concernent également des chômeurs. Concernant la variable Adm11, les modalités présentes sont 1,2,3,4 c'est à dire les sujets ne travaillant plus depuis moins d'un an ou plus.

On remarque aussi que certains sujets en formation professionnelle (Adm12a = JNNN) sont inclus dans le Cluster.

Cluster 3	Cla/Mod	Mod/Cla	Global	p.value	v.test
Adm12a=N9NN	99.41	81.49	3.37	0.00 e+00	Inf
Adm10=85	98.25	67.55	2.82	0.00 e+00	Inf
Adm10=86	85.00	28.61	1.38	3.76 e-149	26.01
Adm11=5	94.23	11.78	0.51	1.36 e-65	17.10
Adm11=4	8.72	55.53	26.16	2.07 e-38	12.96
Adm12a=NNN7	36.67	7.93	0.89	2.67 e-23	9.94
Adm12=NNNNNN	4.47	98.80	90.82	2.73 e-12	6.99
Adm12a=N9N7	100.00	1.68	0.07	1.88 e-10	6.37
Adm12a=N98N	66.67	1.44	0.09	3.57 e-07	5.09

Dans le **Cluster 3** la modalité la plus significativement représentée est Adm12a = N9NN : les personnes au foyer. On trouve également des sujet se déclarant en pré retraite (Adm12a = NNN7/N9N7)

ou à la retraite(Adm12a = N98N).

Les seules modalités de la variables Adm10 dans ce cluster sont 85/86 ce qui semble indiquer que le Cluster correspond à celui des inactifs.

Enfin on remarque que plus de 94 % des personnes déclarant n'avoir jamais travaillé sont dans le Cluster (Adm11 = 5).

Cluster 4	Cla/Mod	Mod/Cla	Global	p.value	v.test
Adm12a=NN8N	99.09	99.50	33.81	0.00 e+00	Inf
Adm11=4	79.84	62.02	26.16	0.00 e+00	Inf
Adm10=77	98.62	62.73	21.42	0.00 e+00	Inf
Adm10=74	99.57	34.19	11.56	0.00 e+00	Inf
Adm12=NNNNNN	36.99	99.77	90.82	1.99 e-159	26.90
Adm11=3	72.89	10.88	5.06	7.49 e-77	18.55
Adm11=2	69.23	8.45	4.11	1.43 e-51	15.11
Adm11=1	57.70	10.76	6.28	1.32 e-37	12.82
Adm10=78	100.00	2.05	0.69	5.01 e-34	12.16
Adm10=75	100.00	0.79	0.27	1.60 e-13	7.38

Enfin pour le **Cluster 4** on s'attends à trouver des sujets retraités. En effet, la modalité la plus significativement représentée est Adm12a = NN8N : sujets retraités.

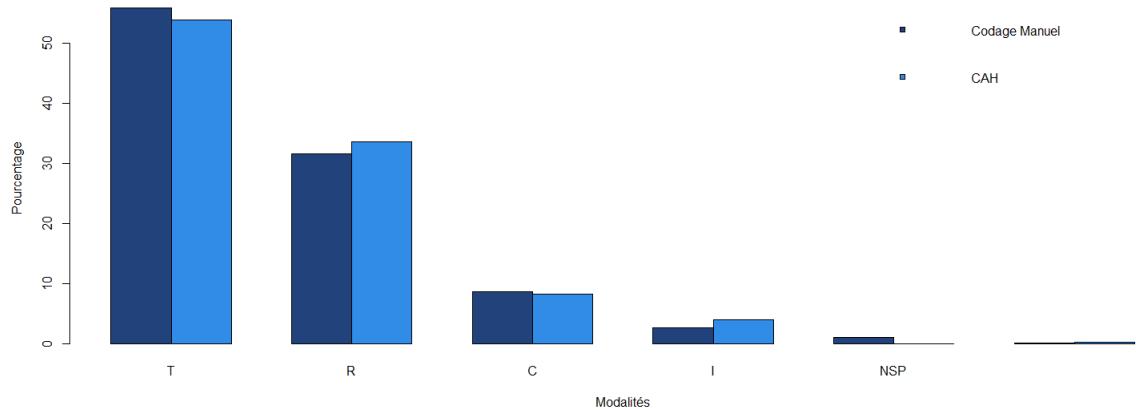
Les codes travail intervenant (variable Adm10) sont 77/74/78/75 : uniquement des codes de la forme "Ancien *nom de la profession*". En outre on ne trouve, dans ce Cluster, aucun individu déclarant travailler (Adm11 = 0 ou Adm12 = 6NNNNN).

On peut alors conclure que **le Cluster 1 correspond aux travailleurs, le Cluster 2 aux chômeurs, le Cluster 3 aux inactifs et le Cluster 4 aux retraités.**

3.1.3 Comparaisons des deux codages

Étudions les différences entre la classification manuelle et la Classification Ascendante Hiérarchique. Tout d'abord on classe comme NA tous les sujets qui avaient une valeurs manquante pour au moins une des quatre variables ayant servies à la classification (n = 29).

On a alors :



	T	R	C	I	NSP	NA
activpro	5666	3209	883	270	113	16
activproCAH	5460	3410	842	416	0	29
Différence	206	-201	41	-146	113	-13

Il y a 509 sujets pour lesquels les deux classifications ne correspondent pas. Sur un total de 10157 cela représente 5,01 % des sujets.

On remarque que la classification Ascendante Hiérarchique a tendance à classer plus facilement les sujets en Retraités et en Inactifs et la classification manuelle en Travailleurs. Une raison simple à cette différence est sans doute la volonté d'être le plus restrictifs possible quant à la classification manuelle des retraités.

Le nombre de valeurs manquantes n'est pas le même pour les deux classifications pour une raison simple : les 16 NA de la classification manuelle sont des sujets pour lesquels TOUTES les variables nécessaires à la classification (Adm12a, Adm11, Adm12 et Am10) étaient manquantes alors que j'ai dû, pour la CAH, classer en NA les sujets pour lesquels AU MOINS UNE des 4 variables étaient manquantes afin de pouvoir réaliser l'ACM.

Il peut également être intéressant d'étudier la manière dont les sujets classés NSP ($n = 113$) par la classification manuelle ont été classés par la CAH :

R	I	C	T	Total
5	77	0	31	113

Sur quels critères la CAH s'est-elle basée pour classer ces 113 sujets ? Pour le comprendre on s'intéresse aux trois modalités dans lesquels elle a répartis les NSP (I,R,T) en construisant des Tableaux de Burt.

Par souci de lisibilité j'ai choisis de n'inclure dans les tableaux de Burt que les modalités contenant un nombre de sujets non nul.

Les cases en rouge correspondent simplement aux tableaux de contingences usuels (croisement des variables deux à deux).

Le tableau est évidemment "symétrique", j'ai choisi de n'en remplir que la partie inférieure.

INACTIFS

	N98N	N9NN	0	4	NNN3NN	NNNNN	77	85	86
N98N	1	∅							
N9NN	∅	76							
0	0	76	76	∅					
4	1	0	∅	1					
NNN3NN	1	0	0	1	1	∅			
NNNNNN	0	76	76	0	∅	76			
77	0	2	2	0	0	2	2	∅	∅
85/86	1	74	74	1	1	74	∅	58	17

1 sujet déclare être au foyer et à la retraite (Adm12a = "N98N), ne plus travailler depuis au moins 3 ans (Adm11 = 4) et a un code travail correspondant à un inactif non retraité (Adm10 + 85/86). Le classer en Inactifs a donc du sens.

2 sujets déclarent être au foyer (Adm12a = N9NN), travailler (Adm11 = 0) et ont un code travail correspondant à un retraité (Adm10 = 77 : Ancien employé). Toutes les variables sont contradictoires. Classer ces sujets en Inactifs n'est pas pertinent.

74 sujets présentent les mêmes caractéristiques que ceux expliqués précédemment à l'exception de la variable Adm10 qui indique que les sujets sont inactifs non retraités. De même, les classer en inactifs n'est pas pertinent.

De manière générale on remarque que déclarer être "au foyer" est très discriminant pour la classification en tant qu'inactifs surtout si cette modalité est couplée avec un code travail 85 ou 86.

RETRAITES

	N98N	NN87	NN8N	0	4	NNNNNN	32/36	77
N98N	1	∅	∅					
NN87	∅	1	∅					
NN8N	∅	∅	3					
0	1	1	0	2	∅			
4	0	0	3	∅	3			
NNNNNN	1	1	3	2	3	5		
32/36	0	0	3	0	3	1	2	∅
77	1	1	0	2	0	2	∅	2

1 déclare être à la retraite et au foyer (N98N) avec un code de travail correspondant bien à un retraité (77). Cependant il déclare également travailler (0). La CAH le classe en retraité malgré tout mais ce choix n'est pas judicieux.

3 sujets se déclarent à la retraite (NN8N), ne travaillant plus (Adm11 = 3) et ont un code travail qui correspond à des travailleurs (Adm10 = 32/36). Leur classification en Retraité n'est pas aberrante. Il est possible que ces sujets n'aient pas lu la liste des codes travail jusqu'à la fin (la où ceux concernant les retraités se trouvent) ce qui expliquerait qu'ils aient indiqué par ce code qu'ils travaillent (32/36).

1 sujet déclare être retraité et pré-retraité (NN87), travailler (Adm11 = 0) et a un code travail de 77. Il y a une certaine forme de cohérence à l'ensemble des choix de ce sujet mais on ne peut pas le classer en tant que Retraité à cause de Adm11 = 0.

TRAVAILLEURS

	NNN7	NNNN	0	1	3	4	NNNNNN	32/36/47/48/ 51/54/55/56	85/86	88
NNN7 NNNN	1 \emptyset	\emptyset 30								
0	0	12	12	\emptyset	\emptyset	\emptyset				
1	0	1	\emptyset	1	\emptyset	\emptyset				
3	0	1	\emptyset	\emptyset	1	\emptyset				
4	1	16	\emptyset	\emptyset	\emptyset	17				
NNNNNN	1	30	12	1	1	171	31			
32/36/47/48/ 51/54/55/56 85/86	1	18	0	1	1	17	19	19	\emptyset	\emptyset
88	0	11	11	0	0	0	11	\emptyset	11	\emptyset
	0	1	1	0	0	0	1	\emptyset	\emptyset	1

1 sujet se déclare pré-retraité (NNN7), travaille (Adm11 = 0) et a un code travail correspondant à quelqu'un ayant encore une activité professionnelle. On peut accepter la classification en travailleur de ce sujet par la CAH.

11 sujets ne cochent aucune case pour Adm12a (soit ils ont oublié cette question soit aucun des choix ne les satisfont ce qui correspondrait à des personnes qui travaillent puisque ce choix n'est pas proposé dans Adm12a). En outre ils déclarent travailler (Adm11 = 0) mais leur code travail est incohérent : 85/86 correspond à des inactifs.

1 sujet travaille et a un code travail qui vaut 88. Ce code n'est pas présent dans la liste, il s'agit probablement d'une faute de frappe.

Enfin, 18 sujets déclarent ne plus travailler (Adm11 = 1/3/4) mais ont un code travail indiquant qu'ils ont encore une activité professionnelle. Ces sujets ne peuvent pas être classés comme des

travailleurs.

A présent que les deux classifications sont établies et clairement expliquées, il y a lieu de se demander si utiliser l'une ou l'autre dans un futur modèle changera les résultats.

Autrement dit construisons un test permettant de vérifier si les deux distributions de la variable activité professionnelle sont similaires.

Test du Khi deux d'ajustement

Il s'agit de vérifier si les écarts d'effectifs entre le codage manuel et la CAH sont "trop" importants pour chacune des classes.

Étant donné que la CAH classe TOUS les individus, la classe NSP de la CAH est vide. Or le test du χ^2 d'ajustement n'est applicable que si toutes les classes contiennent au moins 5 sujets.

Puisque nous nous intéressons particulièrement aux Travailleurs, Retraités et Chômeurs, on rassemble les inactifs, les NSP et les NA dans une classe "AUTRE".

On a alors :

	T	R	C	Autre
activpro	5666	3209	883	299
activproCAH	5460	3410	842	445

Notations :

k : nombre de modalités.

ϕ_i : probabilité qu'un individus soit dans la classe i de la distribution observée.

ϕ_{hi} : probabilité qu'un individu soit dans la classe i de la distribution théorique.

O_i : nombre d'individus observés dans la modalité i .

A_i : nombre d'individus attendus dans la modalité i .

On a : $\forall i \in \llbracket 1, k \rrbracket, A_i = n\phi_{hi}$

Hypothèses :

$$H_0 : \forall i \in \llbracket 1, k \rrbracket, \phi_i = \phi_{hi}$$

$$H_1 : \exists i \in \llbracket 1, k \rrbracket \text{ tq } \phi_i \neq \phi_{hi}$$

Statistique de test :

$$Q = \sum_{i=1}^k \frac{(O_i - A_i)^2}{A_i}$$

D'après le Théorème de Pearson :

$$\begin{aligned} & Si \forall i \in \llbracket 1, k \rrbracket, A_i \geq 5 : \\ & Q \xrightarrow[H_0]{} \chi^2(k-1), n \rightarrow +\infty \end{aligned}$$

Application numérique :

$Q_{obs} = 69.52$ Au niveau de confiance 95 % le χ^2 à 3 degrés de liberté vaut 12.84.

On rejette donc H_0 .

On ne peut pas conclure que les deux variables ont une distribution similaire. Cependant après avoir observé la manière dont les sujets "inclassables" ont été classés par la CAH il semble préférable de se baser sur la classification manuelle de l'activité professionnelle des sujets de l'étude. En analyse de sensibilité nous feront tourner le modèle final sur les données classifiées par la machine.

3.2 Variable d'intérêt : Self Rated Health (SRH)

La santé perçue est décrite dans la littérature comme un bon indicateur de la mortalité. Il s'agit d'une mesure subjective puisque le sujet s'auto-évalue : il est son propre témoin.

Dans le questionnaire d'inclusion il est demandé aux sujets d'indiquer par une note comprise entre 0 et 10 leur état de santé tel qu'il le ressent : 0 pour mauvais et 10 pour excellent.

Globalement les sujets de la cohorte EPP3 s'estiment en bonne santé avec une médiane à 7.

En choisissant ce cut off on crée une variable binaire de santé perçue :

0	1	NA
4390	5741	26

Test du χ^2 d'indépendance entre la santé perçue et l'activité professionnel des sujets de la cohorte EPP3.

On note :

X la v.a qualitative à l modalités.

Y le v.a qualitative à c modalités.

$\forall (i, j) \in \llbracket 0, 5 \rrbracket \times \llbracket 0, 11 \rrbracket :$

$n_{i,j}$ = Nombre de sujets présentant la modalité i de la variable X et la modalité j de la variable Y .

$n_{i..} = \sum_{j=1}^c n_{ij}$: nb total individus dans la classe i de la v.a X .

$n_{..j} = \sum_{i=1}^l n_{ij}$: nb total d'individus dans la classe j de la v.a Y .

$n = \sum_{i=1}^l \sum_{j=1}^c n_{ij}$: nb total d'individus.

On teste $H_0 : X \perp Y$ contre $H_1 = \overline{H_0}$

Statistique de test :

$$Q = \frac{\sum_{i=1}^l \sum_{j=1}^c (n_{ij} - \frac{n_{i..}n_{..j}}{n})^2}{\frac{n_{i..}n_{..j}}{n}}$$

On a :

$$Q \underset{H_0}{\sim} \chi^2((l-1)(c-1))$$

On pose :

$$e_{ij} = \frac{n_{i..}n_{..j}}{n}$$

C'est l'effectif théorique sous H_0 c'est à dire le nombre de sujets présentant les modalités i et j (resp. des v.a X et Y) auxquelles on s'attend sous H_0 .

Règle de décision :

Si $Q_{calc} \geq Q_{th}$ on rejette H_0 (où Q_{th} est le quantile de la loi du χ^2 à $(l-1)(c-1)$ degrés de libertés pour un niveau de confiance α fixé).

Le test du χ^2 d'indépendance n'est applicable que lorsque $\forall (i, j) \in [1, l] \times [1, c] : e_{ij} \geq 5$ et $n_{ij} \geq 50$.

Application numérique :

Le tableau de contingence des variables activpro et srh présente des cases dont les effectifs sont inférieurs à 5. Pour effectuer le test on se base donc sur la version binarisée de la variable de santé perçue.

On pose :

X v.a qualitative à 11 modalités représentant la santé perçue des sujets.

Y v.a qualitative à 5 modalités représentant l'activité professionnelle des sujets.

On trouve $Q_{calc} = 31.71$ et $\chi^2_{4, dl, \alpha=5\%} = 9.49$ donc on rejette H_0 avec une **p-valeur de** $2.19 \cdot 10^{-6}$. Autrement dit il existe un lien fort entre l'activité professionnelle et la santé perçue des sujets de la cohorte EPP3.

Lorsque l'on fait le test avec la variable activproCAH plutôt que activpro (en retirant la classe NSP puisqu'elle est évidemment vide) on obtient également une p-valeur significative ($Q_{calc} = 21.51$, $Q_{th} = 7.81$ et $p = 8.24 \cdot 10^{-6}$).

3.3 Variables d'ajustement et facteurs de confusion

Rappelons qu'un facteur de confusion est une variable à la fois liée à l'exposition et à la variable d'intérêt. Notre variable d'exposition est l'activité professionnelle des sujets et la variable d'intérêt leur santé perçue.

Le facteur de confusion le plus évident est donc l'âge du sujet puisque celui-ci est déterminant dans l'exercice d'une activité professionnelle mais joue également un rôle dans la perception qu'ont les sujets de leur santé.

Les autres variables d'ajustement sont choisies en se basant sur la littérature : sexe, alcool, tabac, score de dépression, statut marital, indice de masse corporel, diabète, activité sportive et niveau d'éducation.

Rappelons que l'indice de masse corporel est le rapport entre le poids et la taille carrée des individus.

Le score de dépression est quant à lui la somme de 13 variables binaires :

1. En ce moment ma vie me semble vide.
2. J'ai du mal à me débarrasser des mauvaises pensées qui me passent par la tête.
3. Je suis sans énergie.
4. Je me sens bloqué(e) ou empêché(e) devant la moindre chose à faire.
5. Je suis déçu(e) et dégouté(e) de moi-même.

6. Je suis obligé(e) de me forcer pour faire quoi que ce soit.
7. J'ai du mal à faire les choses que j'avais l'habitude de faire.
8. En ce moment je suis triste.
9. J'ai l'esprit moins clair que d'habitude.
10. J'aime moins qu'avant faire les choses qui me plaisent ou m'intéressent.
11. Ma mémoire me semble moins bonne que d'habitude.
12. Je suis sans espoir pour l'avenir.
13. En ce moment je me sens moins heureux(se) que la plupart des gens.

On considère qu'un sujet est déprimé si son score de dépression est supérieur ou égal à 7 ou s'il déclare prendre des antidépresseurs.

J'ai également choisi d'ajouter quelques autres variables qui m'ont semblé pertinentes : maladie indice de précarité (EPICE) de vie et score de stress (PSS4) qui sont deux indicateurs très utilisés en épidémiologie et que de nombreuses études ont jugés consistants [13] [12].

Le score EPICE est un score Évaluant de la Précarité et des Inégalités de santé dans les Centres d'Examens de santé) prenant en compte diverses informations :

1. Rencontrez-vous parfois un travailleur social (assistante sociale, éducateur) ?
coeff = 10,06
2. Bénéficiez-vous d'une assurance maladie complémentaire (mutuelle) coeff = -11,83
3. Vivez-vous en couple ? coeff = -8,28
4. Etes-vous propriétaire de votre logement (ou accédant à la propriété) ?
coeff = -8,28
5. Y-a-t-il des périodes dans le mois où vous rencontrez de réelles difficultés financières à faire face à vos besoins (alimentation, loyer, EDF...) ?
coeff = 14,80
6. Vous est-il arrivé de faire du sport au cours des 12 derniers mois ?
coeff = -6,51
7. Etes-vous allé au spectacle (cinéma, théâtre...) au cours des 12 derniers mois ?
coeff = -7,10
8. Etes-vous parti en vacances au cours des 12 derniers mois ?
coeff = -7,10
9. Au cours des 6 derniers mois, avez-vous eu des contacts avec des membres de votre famille autres que vos parents ou vos enfants.
coeff = -9,47
10. En cas de difficultés (financières, familiales, de santé...) y-a-t-il dans votre entourage des personnes sur qui vous puissiez compter pour vous héberger quelques jours en cas de besoin ?
coeff = -9,47
11. En cas de difficultés (financières, familiales, de santé...), y-a-t-il dans votre entourage des personnes sur qui vous puissiez compter pour vous apporter une aide matérielle (y compris un prêt) ?
coeff = -7,10

Lorsque la réponse à une des questions est oui on ajoute le coefficient correspondant au score constant qui vaut 75,14.

On considère un seuil de 30,17 au delà duquel le sujet est considéré comme vulnérable tout âge confondus. Ce seuil n'a évidemment de sens que si les 13 questions sont renseignées.

Le score de stress PSS 4 (Perceived Stress Scale) quant à lui est un score à quatre items :

- Vous a-t-il semblé difficile de contrôler les choses importantes de votre vie ?
- Vous êtes-vous senti(e) confiant(e) en vos capacités à prendre en main vos problèmes personnels ?
- Avez-vous senti que les choses allaient comme vous le vouliez ?
- Avez-vous trouvé que les difficultés s'accumulaient à tel point que vous ne pouviez les contrôler ?

pour les questions 1 et 4 les coefficients de chaque réponse sont :

0. Jamais
1. Presque jamais
2. Parfois
3. Assez souvent
4. Souvent

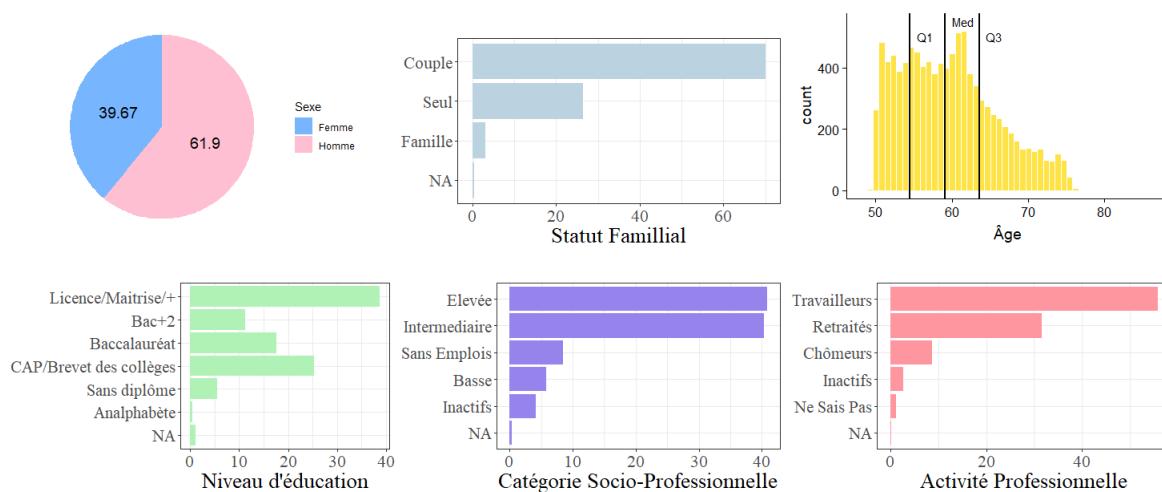
Pour les question 2 et 3 :

0. Souvent
1. Assez souvent
2. Parfois
3. Presque jamais
4. Jamais

Le score est obtenu en sommant les items et est donc compris entre 0 et 16 (0 pour bon niveau de stress et 16 pour très mauvais).

Voici une description visuelle de la cohorte EPP3 en fonction des variables préalablement citées :

1. Variables Socio-Administratives



On observe ainsi que la cohorte EPP3 est masculine à plus de 60 %. De plus pres de 70 % des individus déclarent vivre en couple tandis que 30 % déclarent vivre seuls et moins de 10 % en famille.

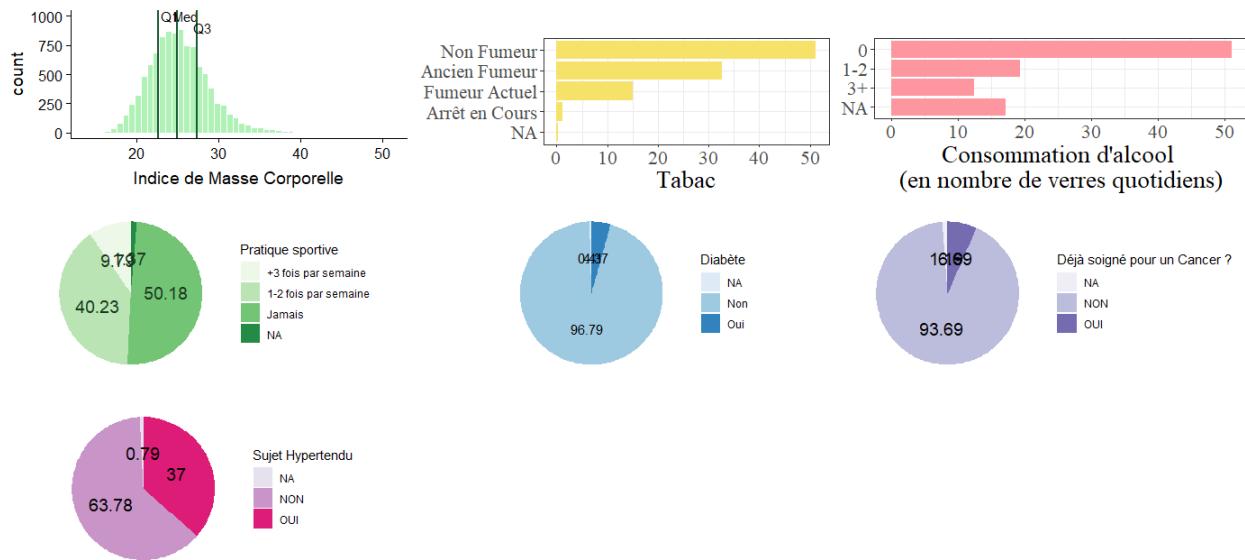
Sans surprise l'histogramme des âges nous indique que les sujets sont globalement âgés de 50 à 75 ans (c'était un des critères d'inclusion). 50 % de la population a moins de 59 ans et 75 % moins de 63. Il n'est donc pas étonnant d'observer un peu plus de 25 % de Retraités contre un peu moins de 60 % de Travailleurs Actifs.

Il est à noté que la distribution de l'âge ne semble pas suivre une loi normale. Cette hypothèse est confirmée après un test de Shapiro effectué sur des échantillons aléatoires d'individus de

5000 sujets.

Enfin la cohorte EPP3 se situe majoritairement dans les catégories socio-professionnelles élevée et intermédiaires. Ils semblent de plus avoir un bon niveau d'éducation puisqu'à peine moins de 6 % se déclarent sans diplômes.

2. Variables Style de vie

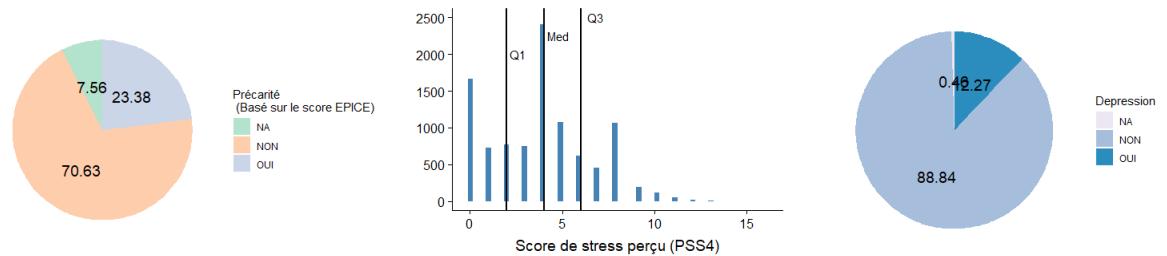


La cohorte EPP3 est dans l'ensemble en bonne santé puisque globalement non diabétique (> 96 %) et non cancéreuse (> 96%).

De plus une large majorité des sujets ne fume pas ou plus tandis que 70 % déclarent boire au plus 2 verres d'alcool par jours (tout alcool confondus).

Cependant une part non négligeable des sujets est hypertendue mais très peu d'individus déclarent être atteint d'une maladie cardiovasculaire.

3. Variables Santé Psychique



Enfin, on observe que très peu d'individus sont dépressifs mais que plus de 20 % sont en situation de précarité.

3.4 Données Manquantes

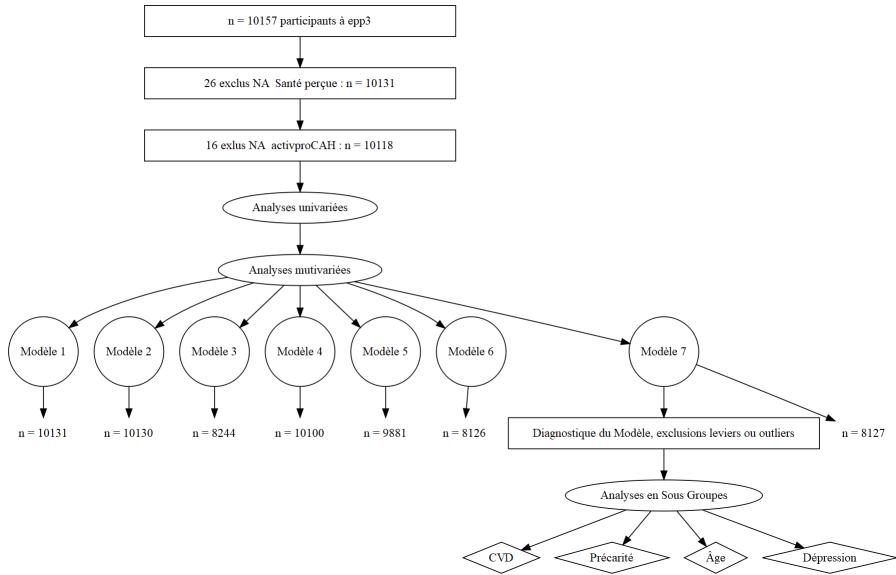
Ci dessous, le tableau du nombre de données manquantes pour chacune des variables de l'analyse.

Nb Na	% NA	Variables
1744	17.17	Alcool (Nb de verres)
756	7.44	Score Epice de Précarité
137	1.35	Sport
134	1.32	Maladies Cardiovasculaires
119	1.17	Cancer
108	1.06	Niveau d'Education
86	0.85	Score de Stresse Perçu (PSS4)
79	0.78	Hypertension Arterielle
46	0.45	Dépression
42	0.41	Situation Familliale
41	0.40	Diabète
30	0.30	Catégorie Socio-Professionnelle
29	0.29	Activité Professionnelle (CAH)
26	0.26	Santé Perçue Binaire
20	0.20	Tabac
16	0.16	Activité Professionnelle (Codage Manuel)
1	0.01	Âge
0	0.00	Sexe
0	0.00	Indice de Masse Corporelle(Classes)

Le nombre de données manquantes par variable de dépasse jamais les 20 %, j'ai donc décider d'exclure les individus ayant au moins une des variables ci-dessus non renseignée.

4 Statistiques

4.1 Flowchart



4.2 Analyses univariées

Afin d'avoir une première idée des variables significativement liées à la santé perçue on effectue une régression logistique entre la santé perçue binarisée des sujets (on coupe à la médiane : si la santé perçue est > 7 on la considère comme bonne) et chacune des variables explicatives.

La régression logistique :

On cherche à modéliser la probabilité qu'une variable prenne une certaine valeur en fonction de variables explicatives.

La spécificité de la régression logistique est le caractère dichotomique de la variables à expliquer. Cela permet de modéliser l'appartenance à une classe de la variable à expliquer par une loi de Bernoulli.

Soit Y la variables à expliquer et X_1, X_2, \dots, X_p les p variables explicatives.

On note Y_i l'observation de la variable Y pour l'individu i , $i \in \llbracket 0, n \rrbracket$ (où n est le nombre total d'individus) et $X_i = (X_{i1}, X_{i2}, \dots, X_{ip})^t$ le vecteur des variables explicatives pour le sujet.

On note $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)^t$, le vecteur de dimension $p + 1$ des paramètres du modèle.

Le modèle logistique ne fait qu'une seule hypothèse à priori :

$$Y_i \text{ iid } \sim \mathfrak{B}(\pi(X_i))$$

Où :

$$\begin{aligned} \pi(X_i) &= \mathbb{E}(Y_i|X_i) \\ &= \mathbb{P}(Y_i = 1|X_i) \\ &= \frac{e^{(\beta_0 + \sum_{k=1}^p \beta_k X_{ik})}}{1 + e^{(\beta_0 + \sum_{k=1}^p \beta_k X_{ik})}} \\ &= \text{sigmoïde}(e^{(\beta_0 + \sum_{k=1}^p \beta_k X_{ik})}) \end{aligned}$$

Ainsi on peut écrire :

$$\begin{aligned} \text{Logit}(\pi(X_i)) &= \frac{\log(\pi(X_i))}{1 - \pi(X_i)} \\ &= \beta_0 + \sum_{k=1}^p \beta_k X_{ik} \end{aligned}$$

$\forall k \in \llbracket 1, p \rrbracket$, e^{β_k} est l'odd ratio de la variable k .

L'estimation du vecteur de régression se fait par la méthode du maximum de vraisemblance :

$$\mathfrak{L}(\beta) = \prod_{i=1}^n \left[\left(\frac{e^{(\beta_0 + \sum_{k=1}^p \beta_k X_{ik})}}{1 + e^{(\beta_0 + \sum_{k=1}^p \beta_k X_{ik})}} \right)^{Y_i} \left(\frac{1}{1 + e^{(\beta_0 + \sum_{k=1}^p \beta_k X_{ik})}} \right)^{(1-Y_i)} \right]$$

Ainsi :

$$\log(\mathfrak{L}(\beta)) = \sum_{k=1}^n \left[Y_i \left(\beta_0 + \sum_{k=1}^p \beta_k X_{ik} \right) - \log \left(1 + e^{\beta_0 + \sum_{k=1}^p \beta_k X_{ik}} \right) \right]$$

Et donc $\forall k \in \llbracket 0, p \rrbracket$:

$$\frac{\partial \log(\mathfrak{L}(\beta))}{\partial \beta_k} = \sum_{i=1}^n X_{ik} \left(Y_i - \frac{e^{X_i^t \beta}}{1 + e^{X_i^t \beta}} \right)$$

Avec $X_{i0} = 1$ et $X_i^t \beta = \sum_{k=1}^p \beta_k X_{ik}$

L'équation $\frac{\partial \log(\mathfrak{L}(\beta))}{\partial \beta} = 0$, qui permet d'obtenir $\hat{\beta}$, est résolue grâce à des algorithmes itératifs.

On a de plus $\forall k \in \llbracket 0, p \rrbracket$:

$$\hat{\beta}_k \underset{n \rightarrow \infty}{\sim} \mathcal{N} \left(\beta_k, \hat{\mathbb{V}} \left(\hat{\beta}_j \right) \right)$$

Ainsi au niveau de confiance $-\alpha$ on a :

$$IC(\beta_k) = \left[\hat{\beta}_k \pm |z_{\frac{\alpha}{2}}| \sqrt{\hat{\mathbb{V}}(\hat{\beta}_k)} \right]$$

Où $z_{\frac{\alpha}{2}}$ est t.q $\mathbb{P}(Z < z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$ pour $Z \sim \mathcal{N}(0, 1)$

On a alors les intervalles de confiance des odd ration pour chacune des variables :

$$\begin{aligned} IC(e^{\beta_k}) &= IC(OR_k) \\ &= e^{\left[\hat{\beta}_k \pm |z_{\frac{\alpha}{2}}| \sqrt{\hat{\mathbb{V}}(\hat{\beta}_k)} \right]} \\ &= \left[e^{\hat{\beta}_k - |z_{\frac{\alpha}{2}}| \sqrt{\hat{\mathbb{V}}(\hat{\beta}_k)}} ; e^{\hat{\beta}_k + |z_{\frac{\alpha}{2}}| \sqrt{\hat{\mathbb{V}}(\hat{\beta}_k)}} \right] \end{aligned}$$

Ici la variable Y est la santé perçue binarisée, les autres variables sont les $p = 16$ variables explicatives.

J'ai choisi de catégoriser 3 des variables continues (IMC, âge et score épice de précarité afin de rendre l'interprétation des résultats plus évidente).

Ainsi l'âge a été découpé en quartiles.

Pour l'indice de masse corporelle, après discussion avec des médecins j'ai fixé à 25 le seuil de normalité et à 30 le seuil de sur poids. Au delà de 30 on parle d'obésité.

Enfin, le score de précarité a un seul officiel, comme expliqué plus haut les sujets ayant un score supérieur à 30.17 sont considéré en situation de précarité.

TABLE 1: Analyses univariées

	0 N=4390	1 N=5741	OR	p.ratio
Activité Professionnelle :				
Travailleurs	2446 (55.7%)	3213 (56.0%)	Ref.	Ref.
Retraités	1315 (30.0%)	1892 (33.0%)	1.10 [1.00 ; 1.20]	0.042
Chômeurs	424 (9.66%)	458 (7.98%)	0.82 [0.71 ; 0.95]	0.007
Inactifs	142 (3.23%)	128 (2.23%)	0.69 [0.54 ; 0.88]	0.003
Ne Sais Pas	63 (1.44%)	50 (0.87%)	0.60 [0.41 ; 0.88]	0.008
Activité Professionnelle (CAH) :				
Travailleurs	2364 (53.9%)	3089 (53.9%)	Ref.	Ref.
Chômeurs	403 (9.19%)	438 (7.64%)	0.83 [0.72 ; 0.96]	0.013
Inactifs	210 (4.79%)	206 (3.59%)	0.75 [0.61 ; 0.92]	0.005
Retraités	1408 (32.1%)	2000 (34.9%)	1.09 [1.00 ; 1.19]	0.059
Sexe :				
Hommes	2424 (55.2%)	3750 (65.3%)	Ref.	Ref.
Femmes	1966 (44.8%)	1991 (34.7%)	0.65 [0.60 ; 0.71]	0.000
Situation Familiale :				
Couple	2929 (66.8%)	4180 (73.1%)	Ref.	Ref.
Famille	134 (3.06%)	188 (3.29%)	0.98 [0.78 ; 1.23]	0.881
Seul	1320 (30.1%)	1354 (23.7%)	0.72 [0.66 ; 0.79]	<0.001
Âge :				
≤ 54	1134 (25.8%)	1395 (24.3%)	Ref.	Ref.
]54,59]	1140 (26.0%)	1393 (24.3%)	0.99 [0.89 ; 1.11]	0.906
]59,63]	1051 (23.9%)	1482 (25.8%)	1.15 [1.03 ; 1.28]	0.016
	1065 (24.3%)	1470 (25.6%)	1.12 [1.00 ; 1.25]	0.042
Éducation :				
CAP/Brevet des collèges	1212 (27.9%)	1354 (23.8%)	Ref.	Ref.
Analphabète	26 (0.60%)	24 (0.42%)	0.83 [0.47 ; 1.45]	0.508
Bac+2	485 (11.2%)	662 (11.6%)	1.22 [1.06 ; 1.41]	0.005
Baccalauréat	812 (18.7%)	977 (17.2%)	1.08 [0.95 ; 1.22]	0.230
Licence/Maitrise/+	1500 (34.5%)	2435 (42.8%)	1.45 [1.31 ; 1.61]	<0.001
Sans diplôme	313 (7.20%)	239 (4.20%)	0.68 [0.57 ; 0.82]	<0.001
Catégorie Socio-Professionnelle :				
Intermédiaire	1908 (43.5%)	2186 (38.1%)	Ref.	Ref.
Basse	300 (6.84%)	294 (5.13%)	0.86 [0.72 ; 1.02]	0.076
Élevée	1541 (35.2%)	2605 (45.4%)	1.48 [1.35 ; 1.61]	0.000
Inactifs	219 (5.00%)	207 (3.61%)	0.83 [0.68 ; 1.01]	0.059
Sans Emplois	416 (9.49%)	441 (7.69%)	0.93 [0.80 ; 1.07]	0.302
Indice de Masse Corporelle :				
Normal	2080 (47.4%)	3105 (54.1%)	Ref.	Ref.
Obésité	551 (12.6%)	427 (7.44%)	0.52 [0.45 ; 0.60]	0.000
Sur-poids	1759 (40.1%)	2209 (38.5%)	0.84 [0.77 ; 0.91]	<0.001

continued on next page

TABLE 1 – *continued from previous page*

	0 N=4390	1 N=5741	OR	p.ratio
Tabac :				
Non Fumeurs	2269 (51.7%)	2919 (50.9%)	Ref.	Ref.
Ancien Fumeurs	1327 (30.2%)	1979 (34.5%)	1.16 [1.06 ;1.27]	0.001
Arrête en Cours	61 (1.39%)	49 (0.85%)	0.62 [0.43 ;0.91]	0.015
Fumeurs	733 (16.7%)	790 (13.8%)	0.84 [0.75 ;0.94]	0.002
Diabète :				
Non	4132 (94.3%)	5537 (96.7%)	Ref.	Ref.
Oui	250 (5.71%)	187 (3.27%)	0.56 [0.46 ;0.68]	<0.001
Activité Sportive :				
Jamais	2531 (58.4%)	2480 (43.7%)	Ref.	Ref.
1-2 fois par semaine	1524 (35.2%)	2496 (44.0%)	1.67 [1.54 ;1.82]	0.000
≥ 3 fois par semaine	277 (6.39%)	702 (12.4%)	2.59 [2.23 ;3.01]	0.000
Hypertension Artérielle :				
Non	2564 (58.9%)	3801 (66.7%)	Ref.	Ref.
Oui	1792 (41.1%)	1898 (33.3%)	0.71 [0.66 ;0.78]	<0.001
Consommation d'alcool :				
0	2262 (63.7%)	2922 (60.1%)	Ref.	Ref.
1-2	759 (21.4%)	1205 (24.8%)	1.23 [1.11 ;1.37]	<0.001
3+	528 (14.9%)	734 (15.1%)	1.08 [0.95 ;1.22]	0.248
Cancer :				
Non	3997 (91.9%)	5364 (94.5%)	Ref.	Ref.
Oui	354 (8.14%)	315 (5.55%)	0.66 [0.57 ;0.78]	<0.001
Score EPICE de précarité :				
Situation non précaire	2827 (69.5%)	4230 (79.5%)	Ref.	Ref.
Situation précaire	1240 (30.5%)	1094 (20.5%)	0.59 [0.54 ;0.65]	0.000
Dépression :				
Non	3533 (80.7%)	5344 (93.4%)	Ref.	Ref.
Oui	847 (19.3%)	377 (6.59%)	0.29 [0.26 ;0.33]	0.000
Score de stresse perçu (PSS4)	.	.	0.82 [0.81 ; 0.84]	< 2.2e-16

On remarque un lien significatif entre l'activité professionnelle et la santé perçue et ce, que l'on considère la classification manuelle ou la classification ascendante hiérarchique de la variable activité professionnelle. Ainsi, les retraités ont une santé perçue de 10 % supérieure aux travailleurs tandis que la SRH des chômeurs est 40 % moins bonne et celle des inactifs 30 \$.

Le groupe des individus non classés manuellement (NSP) s'apparente aux inactifs en terme de baisse de santé perçue. Ce résultat est cohérent avec l'observation faite plus haut : la CAH classe majoritairement ces individus en inactifs.

Les femmes ont une santé perçue de 35 % inférieure à celle des hommes. On peut peu-être expliquer

cela par le fait que les grossesses rendent la santé des femmes plus fragile. De plus certaines études montrent que les femmes sont plus enclines à la dépression que les hommes [11]. Or la santé perçue des sujets dépressifs est de 70 % inférieure à celle des sujets non dépressifs.

Un résultat intéressant est le lien entre la santé perçue et l'entourage du sujet. Les individus déclarant vivre seuls ont une santé perçue significativement inférieure à celle des sujets en couple. Il aurait été éclairant de pouvoir tester cette liaison plus en profondeur, par exemple en demandant aux individus de noter leur sentiment de bonheur au quotidien et d'ensuite coupler ces résultats avec le fait de vivre en couple ou non.

Sans surprise on observe que les sujets ayant une catégorie scio-professionnelle basse se perçoivent en moins bonne santé que ceux de la classe intermédiaire (de même que les inactifs comme nous l'avons déjà vu grâce à la variable d'activité professionnelle), tandis que les sujets d'une CSP élevée voient leur santé perçue augmentée de 50 %. On peut facilement imaginer que plus on descend dans les catégories socio-professionnelles, plus l'accès au soin dans son sens le plus large (ne serait-ce que l'accès à l'information de santé et à la prévention) est difficile.

Dans le même ordre d'idée les sujets de la cohorte se perçoivent en meilleur santé lorsqu'ils ont eu un parcours éducatif plus long. Ainsi les détenteurs d'un BAC+2 ont une santé perçue de 22 % supérieure à ceux ayant un CAP ou le brevet des collèges. Cette augmentation passe à 45 % pour les détenteurs d'une licence ou d'un autre diplôme supérieur. Toutefois les sujets ne sachant ni lire ni écrire voient leur santé perçue baisser de plus de 20 % par rapport aux diplômés d'un CAP ou d'un Brevet des collèges.

Parmi les variables décrivant la santé réelle des sujets de l'étude, observons que les fumeurs ont une santé perçue de plus de 15 % inférieure à celle des non fumeurs. Il est intéressant de remarquer que les anciens fumeurs s'estiment en meilleur santé que ceux qui sont non fumeurs (augmentation de 20 %). Cela s'explique peut-être par le fait que ces individus ont vu un réel effet de l'arrêt du tabac sur leur corps et qu'ils s'estiment, par conséquent, en très bonne santé à présent qu'ils ne fument plus.

Ce qui paraît étonnant par contre c'est que les sujets déclarant être en période de sevrage ont une baisse de santé perçue plus importante que les fumeurs par rapport aux sujets non fumeurs. La difficulté psychologique de l'arrêt de la cigarette ainsi que la prise de poids qui le suit souvent sont des explications probables.

L'un des résultats les plus étonnantes de cette analyse est celui concernant la consommation quotidienne d'alcool (en nombre de verres).

Par rapport aux abstiens, on trouve chez les sujets buvant entre 1 et 2 verres d'alcool par jour une augmentation significative de plus de 20 % de la santé perçue. On pourrait expliquer ce phénomène par le plaisir de boire un verre entre amis ou avec son conjoint. S'il est impossible de vérifier que les individus déclarant boire 1 à 2 verres par jour boivent seuls ou avec des amis ou de la famille, on peut cependant remarquer que la grande majorité de ces individus vivent en couple et qu'ils sont pour la plupart non dépressifs (ils ont donc probablement moins tendance à s'isoler et ont donc plus probablement une vie sociale développée).

couple	1472	0	1767
famille	61	1	196
seul	426	NA	3

4.3 Analyses multivariées

4.3.1 Sélection des variables

On prend en compte dans le modèle l'ensemble des variables décrites dans l'analyse univariée puisque cette dernière montre qu'elles sont toutes significatives.

Nous avions précédemment établi que l'âge était un facteur de confusion évident pour le lien entre la santé perçue et l'activité professionnelle : la comparaison des deux modèles suivants le montre assez clairement :

TABLE 2: Mise en évidence de l'âge comme facteur de confusion

	Modèle 1			Modèle 2		
	OR	IC	p	OR	IC	p
Intercept	1.52	[1.43; 1.62]	<2.2e-16	1.44	[1.33; 1.57]	<2.2e-16
Activité Professionnelle :						
Travailleurs	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Retraités	1.15	[1.05; 1.26]	0.002	1.06	[0.94; 1.20]	0.36
Chômeurs	0.82	[0.71; 0.95]	0.006	0.81	[0.70; 0.94]	0.004
Inactifs	0.88	[0.68; 1.13]	0.30	0.84	[0.65; 1.08]	0.18
Ne Sais Pas	0.68	[0.47; 0.99]	0.048	0.66	[0.45; 0.97]	0.03
Sexe :						
Hommes	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Femmes	0.65	[0.6; 0.71]	<2.2 e-16	0.65	[0.6; 0.71]	<2.2 e-16
Âge :						
<54.5				Ref.	Ref.	Ref.
[54.5; 59.1]				1.04	[0.93; 1.16]	0.53
[59.1; 63.7]				1.18	[1.04; 1.34]	0.008
[63.7; 86.2]				1.13	[0.97; 1.31]	0.11

Lorsque l'on n'ajuste pas le modèle sur l'âge, on montre que les retraités ont une meilleures santé perçue que les travailleurs : augmentation de 15 % avec une p-valeur de 0.002.

Dès lors que l'on ajuste sur l'âge cette information est masquée par l'augmentation de 18 % de la santé perçue des sujets dont l'âge est compris entre 59 et 64 ans : tranche d'âge de près d'un tiers des retraités (97 % ont plus de 59 ans).

Les variables considérées par l'étude peuvent être divisées en 3 catégories : celles qui traitent du style de vie des sujets (indice de masse corporelle, statut tabagique, diabète, pratique sportive, hypertension, consommation d'alcool), les variables "socio-administratives" (sexe, statut familiale, âge, niveau d'éducation, catégorie socio-professionnel, activité professionnelle) et celles traitant de la santé psychique des sujets (dépression, stresse).

J'ai fait une sélection de variables dans chacun des trois blocs en me basant sur le critère AIC⁵ (méthode stepwise) en incluant dans chaque modèle l'activité professionnelle qui est notre variable d'exposition, l'âge et le sexe. Chacun des trois modèles élimine l'âge et conserve toutes les autres variables.

On inclut ces dernières dans le modèle :

5. $AIC = -2 \log(\mathcal{L}(\hat{\beta})) + 2p$

TABLE 3: Modèle de sélection de variables

	OR	IC	p
Intercepte :	2.41	[2.00 ; 2.89]	< 2.2e-16
Activité professionnelle :			
Travailleurs	Ref.	Ref.	Ref.
Retraités	1.21	[1.08 ; 1.36]	0.001
Chômeurs	0.95	[0.37 ; 2.43]	0.91
Inactifs	1.03	[0.57 ; 1.86]	0.92
NSP	0.74	[0.39 ; 1.40]	0.36
Sexe			
Hommes	Ref.	Ref.	Ref.
Femmes	0.81	[0.72 ; 0.91]	0.0003
Catégorie socio-professionnelle			
Intermédiaire	Ref.	Ref.	Ref.
Elevée	1.17	[1.04 ; 1.33]	0.01
Basse	1.02	[0.80 ; 1.31]	0.87
Inactifs	1.07	[0.64 ; 1.80]	0.81
Sans Emplois	1.18	[0.45 ; 3.01]	0.73
Éducation			
CAP/Brevet des collèges	Ref.	Ref.	Ref.
Licence/Maitrise/+	1.12	[0.99 ; 1.29]	0.097
Bac+2	1.05	[0.88 ; 1.25]	0.60
Baccalauréat	0.97	[0.84 ; 1.14]	0.74
Sans diplôme	0.74	[0.57 ; 0.95]	0.02
Analphabète	1.60	[0.60 ; 4.37]	0.35
Situation Familiale			
Couple	Ref.	Ref.	Ref.
Famille	1.08	[0.81 ; 1.46]	0.60
Seul	0.93	[0.83 ; 1.05]	0.23
Activité Sportive			
Jamais	Ref.	Ref.	Ref.
1-2 fois par semaine	1.45	[1.31 ; 1.61]	1.65e-13
≥3 fois par semaine	2.37	[1.97 ; 2.86]	< 2.2e-16
Indice de Masse Corporelle			
Normal	Ref.	Ref.	Ref.
Sur-poids	0.80	[0.72 ; 0.89]	5.59e-05
Obésité	0.61	[0.51 ; 0.72]	1.67e-09
Hypertension Artérielle			
Non	Ref.	Ref.	Ref.
Oui	0.76	[0.69 ; 0.84]	2.81e-07
Statut Tabagique			
Non fumeurs	Ref.	Ref.	Ref.
Ancien Fumeurs	1.10	[0.98 ; 1.22]	0.11

continued on next page

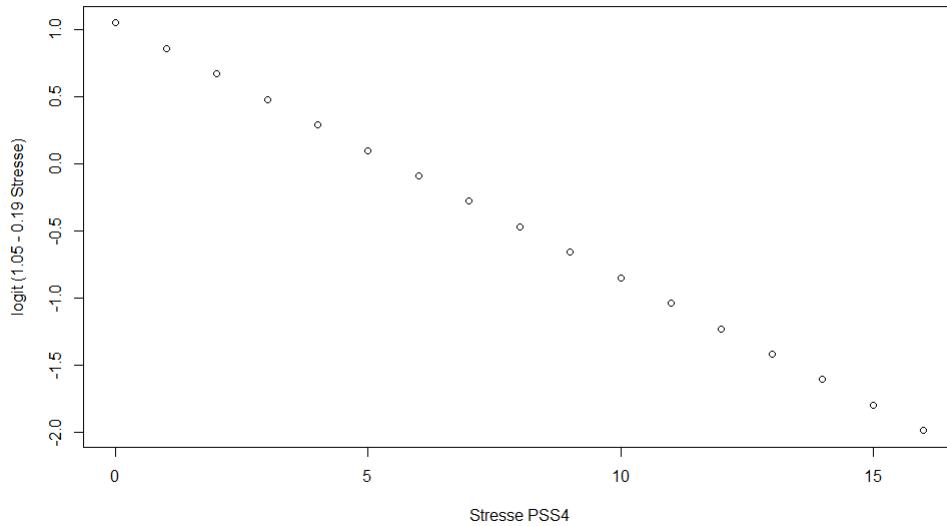
TABLE 3 – *continued from previous page*

	OR	IC	p
Arrêt en cours	0.50	[0.31 ; 0.79]	0.004
Fumeurs actuels	0.94	[0.81 ; 1.08]	0.39
Dépression			
Non	Ref.	Ref.	Ref.
Oui	0.47	[0.40 ; 0.56]	< 2.2e-16
Stresse	0.87	[0.85 ; 0.88]	< 2.2e-16

4.3.2 Diagnostique du modèle

Lien linéaire des variables quantitatives

Vérification de la linéarité du lien entre le logit et l'âge



Le Logit semble baisser lorsque le score de stresse augmente ce qui est logique puisque plus le PSS4 est élevé plus le sujet est considéré comme stressé.

Adéquation du modèle : test des résidus de Pearson

$\forall i \in \llbracket 1, n \rrbracket :$

$$e_i^o = \frac{Y_i - \hat{\pi}(X_i)}{\sqrt{\hat{\pi}(X_i)(1 - \hat{\pi}(X_i))}}$$

(H_0) : Le modèle est pertinent : $\pi(X_i) = \text{sigmoïde}(\beta_0 + \sum_{k=1}^p \beta_k X_{ik})$

(H_1) : Le modèle n'est pas pertinent : $\pi(X_i) \neq \text{sigmoïde}(\beta_0 + \sum_{k=1}^p \beta_k X_{ik})$

Statistique de test :

$$S^2 = \sum_{i=1}^n \left(\overset{\circ}{e_i} \right)^2 \underset{H_0}{\sim} \chi^2(n - (p + 1))$$

Application numérique :

On a $S_{obs}^2 = 7495.388$, $n = 7501$, $p = 11$

On obtient donc $(n - (p + 1)) = 7489$ et $pvaleur = 0.43$.

Le modèle est donc adéquat aux données.

Observations atypiques

On utilise la distance de Cook et les résidus standardisés de Pearson pour détecter les observations atypiques.

$\forall i \in [1, n]$ les résidus standardisés de Pearson sont donnés par :

$$\overset{*}{e_i} = \frac{\overset{\circ}{e_i}}{\sqrt{1 - \left[W^{\frac{1}{2}} X (X^t W X)^{-1} X^t W^{\frac{1}{2}} \right]_{i,i}}}$$

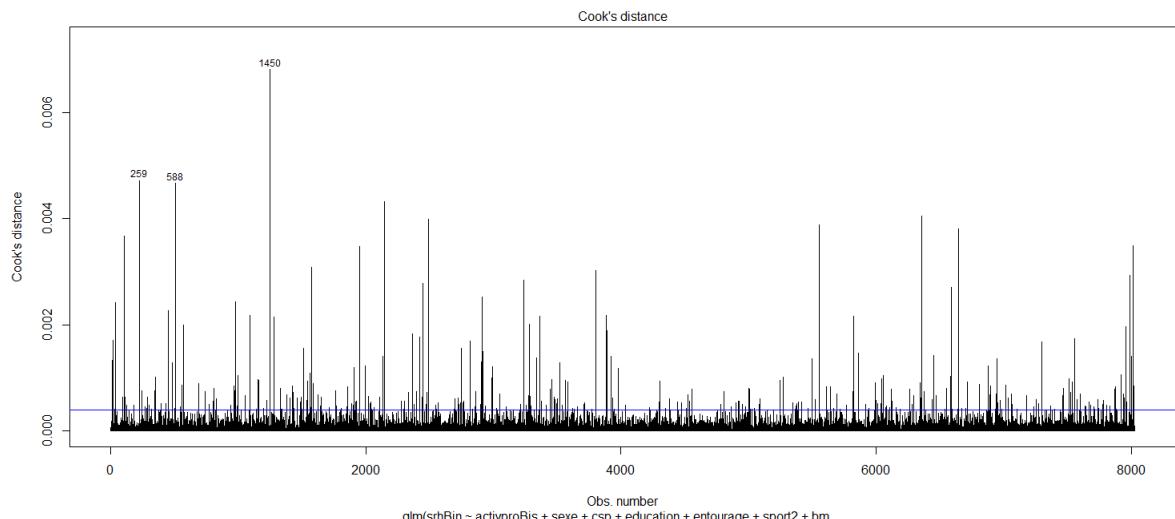
Où $W = diag(\pi(X_1)(1 - \pi(X_1)), \dots, \pi(X_n)(1 - \pi(X_n)))$

et les distances de Cook par :

$$d_i = \frac{\left[W^{\frac{1}{2}} X (X^t W X)^{-1} X^t W^{\frac{1}{2}} \right]_{i,i}}{(p + 1) \left(1 - \left[W^{\frac{1}{2}} X (X^t W X)^{-1} X^t W^{\frac{1}{2}} \right]_{i,i} \right)}$$

On choisit $\frac{4}{n-(p+1)} = \frac{4}{8016} = 0.0004$ comme seuil pour la distance de Cook.

On élimine les individus ayant une distance de Cook supérieure à ce seuil et des résidus inférieurs à $|2|$ (57 individus).



Les résultats de l'analyse multivariée ainsi effectuée sont similaires à ceux du modèle explicité plus haut.

4.4 Analyse en sous-groupes

Dans cette partie, nous appliquons le modèle précédent dans deux sous populations différentes.

Dans la population composée des sujets ayant déclaré avoir un cancer ou une maladie cardiovasculaire, on observe que la variable d'hypertension devient significative ($OR = 0.76$) ce qui est logique puisque l'hypertension artérielle est un facteur de risque connu des maladies cardiovasculaires. De plus la variable dépression a un OR beaucoup plus bas : population générale 0.47 vs sous population 0.28. Résultat qui ne surprend pas puisqu'il est aisément d'imaginer que les individus souffrant d'une pathologie lourde comme le cancer et une CVD ont une moindre santé psychique.

On remarque également que dans cette sous population , aucune des catégories de la variable activité professionnelle n'a un lien significatif sur la santé perçue.

Enfin, les individus dont la catégorie socio-professionnelle est "inactifs" présentent un OR à 3.89 ($p = 0.06$) alors que dans la population globale cette catégorie n'était pas significative.

Lorsque l'on effectue la même démarche dans la sous population des individus en situation de précarité on remarque plusieurs choses.

Tout d'abord comme dans la sous population précédente, l'activité professionnelle n'est plus significativement liée à la santé perçue.

En deuxième lieu nous remarquons que la baisse de santé perçue des femmes est accentuée : population générale 0.81 vs sous population 0.76.

Enfin, on observe également qu'avoir une activité sportive plus de 3 fois par semaine améliore plus la santé perçue des individus en situation de précarité ($OR = 2.68$) que les individus de la population globale ($OR = 2.37$).

4.5 Analyse de sensibilité

Comme nous l'avons vu dans les parties précédentes j'ai construit deux variables traitant de l'activité professionnelle des sujets : la première manuellement (au cas par cas) et la deuxième avec

un algorithme(Classification Ascendante Hiérarchique).

Un test du χ^2 avait montré que les deux classifications n'étaient pas assez similaires pour être considérées comme significativement identiques : que donne le modèle finalement choisi lorsque la variable d'exposition est la CAH de l'activité professionnelle ? En effectuant la même régression logistique on retrouve des résultats similaires.

5 Perspectives : Ce que j'ai fait, découvert, appris

Comme expliqué dans la première partie, j'ai effectué mon stage dans l'unité INSERM 970 qui est spécialisée dans les maladies cardiovasculaires.

J'y ai passé cinq mois : du 01/04/2019 au 31/08/2019, durant lesquels j'ai beaucoup appris dans des domaines très variés.

Tout d'abord j'ai découvert le monde de la recherche, qui m'a beaucoup plu. C'était un monde entièrement neuf pour moi et la longue durée de mon stage a été un atout non négligeable dans son succès : en effet j'ai eu le temps de m'adapter à l'environnement, aux us et coutumes, au jargon, à l'équipe.

Cela m'a également donné le temps de me documenter sur l'épidémiologie : domaine qui m'était inconnu et qui, malgré sa ressemblance avec les statistiques a ses propres règles, son propre vocabulaire. En m'intégrant dans l'équipe, j'ai découvert le travail de groupe nécessaire à la rédaction de chaque article scientifique et l'émulsion que cela occasionne. J'ai pu apprendre les grandes étapes de l'élaboration d'une étude épidémiologique, le cadre légal, le recrutement des individus, leur suivi ainsi que la saisie et le nettoyage des données.

J'ai également participé au recueil de certaines données en appelant les hôpitaux pour récupérer des comptes redus hospitaliers de sujets de l'étude qui étaient utiles aux médecins investigateurs pour confirmer une pathologie déclarée par lesdits sujets dans les questionnaires de suivis. J'ai ainsi pu me confronter à l'épineux problème de la confidentialité des données de santé.

J'ai également eu l'occasion de participer au reviewing d'un article (pas officiellement juste pour la pédagogie). J'ai ainsi pu apprendre toutes les étapes de la rédaction d'un article : le travail de recherche, la mise en forme des idées puis la recherche d'une revue susceptible de publier l'article, l'étape de relecture par des pairs et la prise en compte de leurs remarques...

Au-delà du monde de la recherche, j'ai dû apprendre à m'intégrer dans un groupe où il y avait une grande pluralité des parcours (médecins, statisticiens, attachés de recherches cliniques, ingénieurs...). Cela supposait de se rendre intelligible par tous, et parfois de verser dans une forme de vulgarisation de certains concepts statistiques ce qui m'a beaucoup enrichie⁶.

J'ai également beaucoup profité du savoir des médecins qui m'entouraient notamment sur le fonctionnement de la santé en France ainsi que sur certaines notions médicales élémentaires. Ainsi j'ai pu percevoir la différence entre une "jolie" analyse statistique et une analyse qui avait un sens/intérêt clinique.

Au niveau des statistiques, plutôt que de me lancer dans de techniques "avancées", mon directeur de stage et moi avons décidé de nous concentrer sur les bases comme la visualisation de données et la régression logistique. J'ai pris l'initiative de tenter des méthodes exploratoires comme l'ACM ou la CAH.

Régulièrement l'équipe se réunissait pour des réunions de méthodologie durant lesquelles j'ai pu me familiariser avec d'autres notions dont j'espérais étudier le versant mathématiques en M2 (modèles de suivi, mété-analyse, données répétées et GEE...).

Ces réunions avaient pour support un article dont on faisait une lecture critique en tentant de comprendre d'une part les intérêts clinique de l'étude menée et d'autre part les méthodes statistiques mises en œuvre.

Profitant du temps que j'avais, je me suis familiarisée avec de nouveaux outils qui se sont révélés très utiles : github pour le versionnage de mon code et de ce rapport, pubmed pour lire ce qui s'était déjà fait sur mon sujet et zotero pour créer ma bibliographie.

Conclusion

Pour conclure, ce stage était une très bonne expérience, très enrichissante qui m'a donné envie de poursuivre dans la recherche en bio-statistiques.

6. Voir les chaînes Youtube StatQuest et science4all (notamment la série de vidéo sur la formule de Bayes) qui ne se contentent pas d'expliquer vaguement quelques concepts très connus mais qui vont au fond des choses tout en les rendant accessibles au "grand public"

6 Annexes

6.1 Annexe 1 : Questionnaire IPC

QUESTIONNAIRE
SOCIO-ADMINISTRATIF – 2011
Annexe au questionnaire médical



Collez votre étiquette à
l'intérieur de ce cadre.

1. Comment avez-vous eu connaissance du centre ?

Invitation du Centre	<i>Adm 1</i>	<input type="checkbox"/>	Assistante sociale (ou travailleur social)	<i>Adm 6</i>	<input type="checkbox"/>
Famille, voisin, ami	<i>Adm 2</i>	<input type="checkbox"/>	Médecin traitant	<i>Adm 7</i>	<input type="checkbox"/>
Foyer, centre d'insertion, mission locale, autre	<i>Adm 3</i>	<input type="checkbox"/>	Caisse d'assurance maladie	<i>Adm 8</i>	<input type="checkbox"/>
Mutuelle, caisse de retraite	<i>Adm 4</i>	<input type="checkbox"/>	Presse	<i>Adm 9</i>	<input type="checkbox"/>
Médecin travail, employeur, comité entreprise	<i>Adm 5</i>	<input type="checkbox"/>			

2. Si vous n'êtes pas en CDI, êtes-vous : (plusieurs réponses possibles) ? *Adm 11 et Adm 11a*

Etudiant(e)	<i>1</i>	<input type="checkbox"/>	En contrat emploi-solidarité, intérim, CDD	<i>6</i>	<input type="checkbox"/>
Jeune en cours de formation	<i>2</i>	<input type="checkbox"/>	Pré-retraité(e)	<i>7</i>	<input type="checkbox"/>
A la recherche d'un emploi	<i>3</i>	<input type="checkbox"/>	Retraité(e)	<i>8</i>	<input type="checkbox"/>
Chômeur depuis moins de six mois	<i>4</i>	<input type="checkbox"/>	Personne au foyer	<i>9</i>	<input type="checkbox"/>
Chômeur depuis plus de six mois	<i>5</i>	<input type="checkbox"/>	En formation professionnelle	<i>10</i>	<input type="checkbox"/>

3. Depuis quand n'exercez-vous plus d'activité professionnelle ? *Adm 11*

Moins d'un an	<i>1</i>	<input type="checkbox"/>	3 ans ou plus	<i>4</i>	<input type="checkbox"/>
1 an	<i>2</i>	<input type="checkbox"/>	Vous n'avez jamais travaillé	<i>5</i>	<input type="checkbox"/>
2 ans	<i>3</i>	<input type="checkbox"/>	<i>travail</i> : <i>0</i>		

4. Si vous travaillez quelle est votre profession ? *Adm 10*

					<input type="checkbox"/>	<input type="checkbox"/>
--	--	--	--	--	--------------------------	--------------------------

5. Rencontrez-vous parfois un travailleur social (assistante sociale, éducateur) ? Oui Non *Adm 13*

6. Quel est votre niveau d'études ? *Adm 14*

Ne sait pas lire ou écrire le français	<i>1</i>	<input type="checkbox"/>	BAC, BTS	<i>4</i>	<input type="checkbox"/>
Sans diplôme	<i>2</i>	<input type="checkbox"/>	BAC + 2	<i>5</i>	<input type="checkbox"/>
CAP, BEPC, Brevet des collèges	<i>3</i>	<input type="checkbox"/>	Science, maîtrise et +	<i>6</i>	<input type="checkbox"/>

7. Bénéficiez-vous : d'une mutuelle *Adm 15* **d'une CMU** *Adm 15a* **CMU complémentaire** *Adm 15b* **de 100% ALD** *Adm 15c*

8. Avez-vous une difficulté d'accès aux soins ? *Adm 16* Oui Non

9. Avez-vous un médecin traitant ? *Adm 17* Oui Non

10. Estimez-vous que votre situation professionnelle est précaire ? *Adm 18* Oui Non

11. Vivez-vous : Qm 2 seul(e) en couple en famille

12. Combien avez-vous d'enfants ? *Qm 3*

13. Etes-vous propriétaire de votre logement (ou accédant à la propriété) ? *Adm 21* Oui Non

14. Y-a-t-il des périodes dans le mois où vous rencontrez de réelles difficultés financières à faire face à vos besoins (alimentation, loyer, EDF,) ? *Adm 22* Oui Non

15. Etes-vous allé(e) au spectacle (cinéma, théâtre, ...) au cours des 12 derniers mois ? *Adm 23* Oui Non

16. Etes-vous parti(e) en vacances au cours des 12 derniers mois ? *Adm 24* Oui Non

17. Au cours des 6 derniers mois, avez-vous eu des contacts avec des membres de votre famille (autres que vos parents ou vos enfants) ? *Adm 25* Oui Non

18. En cas de difficultés (financières, familiales, de santé) y-a-t-il des personnes sur lesquelles vous pouvez compter pour ?

- vous héberger quelques jours en cas de besoin *Adm 26* Oui Non
- vous apporter une aide matérielle (y compris un prêt) *Adm 26a* Oui Non

Ne pas remplir cette partie

Code Accueil :



Signature :

16. Utilisez-vous des margarines enrichies en stérol (proactiv® et autres) ? Margarine oui non

17. Consommez-vous régulièrement des boissons autres : boisson Aucune 1 A base de cola 2 A base de taurine (dites énergisantes) 3 Autre soda 4

17a. Si oui, quelle quantité : (une seule réponse) qt-boisson

1 à 4 verres par jour au maximum 1 Plus de 4 verres par jour 2

TABAC

18. Vous êtes : non fumeur (1) fumeur actuel (2) ancien fumeur (3) en cours d'arrêt - 1 an (4) Qm 16

19. Si fumeur actuel (Réponse 2):

19a. Combien de cigarettes ou de cigarellas par JOUR ? Qm 17a

--	--

19b. Depuis combien d'années fumez-vous ? Qm 17b

--	--

19c. Inhaliez-vous (ou avalez-vous) la fumée ? Qm 17c

oui non

19d. Fumez-vous la pipe ? Qm 17d

oui non

19e. Fumez-vous le narguilé ou la chicha ? Chicha

oui non

19f. Fumez-vous le cigare ou autre ? Qm 17e

oui non

20. Si ancien fumeur (Réponse 3):

20a. Combien de cigarettes ou de cigarellas fumiez-vous par JOUR ? Qm 18a

--	--

20b. Pendant combien d'années avez-vous fumé ? Qm 18c

--	--

20c. Depuis combien d'années avez-vous cessé de fumer ? Qm 18d

--	--

ANTÉCÉDENTS FAMILIAUX

Ces questions concernent vos père, mère, frères et sœurs

Parmi vos père, mère, frères et sœurs, quelqu'un a-t-il eu :

21. une hypertension artérielle ? Qm 19

oui non

22. un infarctus du myocarde ? Qm 20

oui non

23. une mort subite (mort brutale non accidentelle et non volontaire) avant l'âge de 60 ans ? Qm 22

oui non

24. un taux de cholestérol trop élevé ? Qm 23

oui non

25. un diabète ? Qm 24

oui non

26. une hémochromatose (trop de fer dans le sang) ? Qm 25

oui non

27. un cancer de la prostate ? AFK prost

oui non

28. un cancer de l'intestin ou du rectum ? Qm 27

oui non

29. un cancer du sein ? Qm 28

oui non

30. un cancer de l'utérus ? Qm 29

oui non

31. une maladie d'Alzheimer ? AF-Alzheimer

oui non

32. une maladie allergique (asthme, eczéma ...) ? Qm 31

oui non

33. un glaucome, une tension oculaire élevée ? Qm 32

oui non

ANTÉCÉDENTS PERSONNELS

Avez-vous eu :

34. une tension artérielle égale ou supérieure à 140 mm Hg à plusieurs reprises ? ap-ha

oui non

35. un infarctus du myocarde ? Qm 34

oui non

36. une angine de poitrine (maladie coronaire confirmée par votre médecin) ? Qm 35

oui non

37. un pontage coronaire, une angioplastie ou un stent coronaire ? AP-pontage

oui non

38. un autre problème cardiaque (fibrillation auriculaire, fibrillation atriale, flutter) ? Qm 37

oui non

39. une maladie des artères ? Qm 38

oui non

40. une phlébite ? Qm 39

oui non

41. une embolie pulmonaire ? Qm 40

oui non

42. un accident vasculaire cérébral ? **Qm 41**
43. des crises d'épilepsie ou des convulsions ? **Qm 42**
44. de l'asthme ? **Qm 43**
45. une maladie allergique (rhume des foins, eczéma, ...) ? **Qm 44**
46. une tuberculose ? (*plusieurs réponses possibles*) du poumon (1) du rein (2) autre localisation (3) **Qm 48**
47. la rubéole **Qm 49**
48. une affection ophthalmologique (maladie des yeux) ? **Qm 51**
49. une affection O.R.L. (maladie du nez, de la gorge, des oreilles) ? **Qm 52**
50. une affection de la glande thyroïde (goitre, nodule) ? **Qm 53**
51. un ulcère de l'estomac ou du duodénum ? **Qm 54**
52. une hernie hiatale ou reflux gastro-oesophagien? **am 55**
53. une cirrhose? **AP_cirr**
54. une chirurgie digestive de l'obésité : anneau gastrique, sleeve (ou ballon), bypass ? **chir_obes**
55. un traitement des rides par Botox, collagène ? **Botox**
56. une hépatite (avec ou sans jaunisse) ? **Qm 58**
- Qm 58a** *si oui*, 56a. (*plusieurs réponses possibles*) hépatite B 1 hépatite C 2 autre 3
57. des polypes de l'intestin ou du rectum ? **Qm 59**
58. une coloscopie au cours des cinq dernières années ? **Qm 60**
59. des crises de coliques néphrétiques ou calculs du rein ? **Qm 61**
60. une autre maladie rénale ? **Qm 62**
61. des infections urinaires (cystites à répétition) ? **Qm 63**
62. une dépression diagnostiquée par un médecin ? **Qm 64**
63. Avez-vous déjà fait une tentative de suicide ? **TS**
64. une maladie sexuellement transmissible ? **Qm 65**
- HST1** *Si oui*, 64a. blennorragie, mycose, chlamydiae (1) mycoplasme, trichomonas, gonococcie (2)
(plusieurs réponses possibles) syphilis (3) herpès (4) HPV (5) autre (6)
65. Test de dépistage du SIDA ? **Qm 66**
- Qm 66a** *Si oui*, 65a. - Etais-ce dans les six derniers mois ?
66. Avez-vous été soigné(e) pour une tumeur précancéreuse ou un cancer ? **Qm 67**
- Si oui*, 66a. - Qu'elle était sa localisation (*plusieurs réponses possibles*)
- Tumeur_1**
- | | | |
|---|---------------------------------------|---|
| <input type="checkbox"/> pulmonaire (1) | <input type="checkbox"/> O.R.L.(5) | <input type="checkbox"/> de la peau (9) |
| <input type="checkbox"/> digestive 2) | <input type="checkbox"/> prostate (6) | <input type="checkbox"/> utérus col (J) |
| <input type="checkbox"/> utérus corps (3) | <input type="checkbox"/> sein (7) | <input type="checkbox"/> ovaire (K) |
| <input type="checkbox"/> rein (4) | <input type="checkbox"/> vessie (8) | <input type="checkbox"/> autre (L) |
67. Avez-vous été opéré(e) : (*plusieurs réponses possibles*) : **Qm 68**
- | | | |
|--|--|---|
| <input type="checkbox"/> de l'estomac ou du duodénum (1) | <input type="checkbox"/> de la vésicule biliaire (2) | <input type="checkbox"/> de l'intestin (3) |
| <input type="checkbox"/> du poumon ou de la plèvre (4) | <input type="checkbox"/> du cœur (5) | <input type="checkbox"/> des reins (6) |
| <input type="checkbox"/> de la vessie (7) | <input type="checkbox"/> des artères (sauf varices)(8) | <input type="checkbox"/> autre (appendice, etc...)(9) |
- Opération_2**
68. Avez-vous bénéficié d'une transfusion de sang avant 1991 ? **trsf_sang** oui non
69. Vous a-t-on trouvé un taux trop élevé dans le sang de : cholestérol (1) triglycérides(2) sucre (3) **Qm 71**
(plusieurs réponses possibles)
70. Etes-vous diabétique ? **diabète** oui non

VACCINATIONS

71. Avez-vous eu il y a moins de 10 ans, une vaccination ou un rappel **Qm 72**
- (plusieurs réponses possibles)*
- | | | | | |
|---|--|--|---|--------------------------------------|
| <input type="checkbox"/> tétanos (1) | <input type="checkbox"/> polio (2) | <input type="checkbox"/> rubéole (3) | <input type="checkbox"/> hépatite B (4) | <input type="checkbox"/> BCG (5) |
| <input type="checkbox"/> grippe (automne dernier) (6) | <input type="checkbox"/> oreillons (7) | <input type="checkbox"/> pneumocoque (8) | <input type="checkbox"/> coqueluche (9) | <input type="checkbox"/> autre(s)(M) |
| <input type="checkbox"/> contre le papillomavirus (HPV) (J) | | <input type="checkbox"/> méningite (K) | <input type="checkbox"/> hépatite A (L) | |
- Vaccine** *←*
- Vaccin 3** *" "*

ÉTAT DE SANTÉ ACTUEL													
72. Indiquez par une note comprise entre 0 et 10, votre état de santé tel que vous le ressentez													
Qm 73	mauvais →	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6	<input type="checkbox"/> 7	<input type="checkbox"/> 8	<input type="checkbox"/> 9	<input type="checkbox"/> 10	← excellent
73. Avez-vous des troubles de l'équilibre ou des vertiges?	Qm 75	<input type="checkbox"/> oui	<input type="checkbox"/> non										
74. Avez-vous fait plus d'une chute dans l'année écoulée?	chute	<input type="checkbox"/> oui	<input type="checkbox"/> non										
75. Avez-vous habituellement des troubles du sommeil?	Qm 77	<input type="checkbox"/> oui	<input type="checkbox"/> non										
76. En moyenne combien d'heures dormez-vous par nuit?	heur_som	<input type="checkbox"/>	<input type="checkbox"/>										
77. Souhaitez-vous que soient abordées des questions sur votre vie sexuelle?	Qm 78	<input type="checkbox"/> oui	<input type="checkbox"/> non										
78. Avez-vous des troubles de mémoire?	tbl_memoire	<input type="checkbox"/> oui	<input type="checkbox"/> non										
79. Avez-vous déjà ressenti une douleur, une gêne pénible, une sensation de poids en pleine poitrine?	Qm 82	<input type="checkbox"/> oui	<input type="checkbox"/> non										
Si oui,													
79a. cette douleur siégeait-elle au milieu de la poitrine?	Qm 82a	<input type="checkbox"/> oui	<input type="checkbox"/> non										
79b. gagnait-elle la mâchoire, ou les 2 épaules, ou les 2 bras?	Qm 82b	<input type="checkbox"/> oui	<input type="checkbox"/> non										
79c. s'est-elle accompagnée d'une sensation de malaise, de sueurs?	Qm 82c	<input type="checkbox"/> oui	<input type="checkbox"/> non										
79d. cette sensation est-elle survenue pendant la marche?	Qm 82d	<input type="checkbox"/> oui	<input type="checkbox"/> non										
Si oui, 79da. avez-vous dû ralentir fortement la marche, ou arrêter de marcher?	Qm 82da	<input type="checkbox"/> oui	<input type="checkbox"/> non										
79e. cette douleur a-t-elle duré moins de 5 minutes?	Qm 82e	<input type="checkbox"/> oui	<input type="checkbox"/> non										
79f. avez-vous eu cette douleur plus d'une fois?	Qm 82f	<input type="checkbox"/> oui	<input type="checkbox"/> non										
80. Vous est-il arrivé de ressentir que votre cœur s'emballait?	Qm 83	<input type="checkbox"/> oui	<input type="checkbox"/> non										
81. Vous est-il arrivé de ressentir que votre cœur avait des ratés?	Qm 84	<input type="checkbox"/> oui	<input type="checkbox"/> non										
82. Etes-vous essoufflé(e) quand vous marchez à côté d'une personne de votre âge?	Qm 85	<input type="checkbox"/> oui	<input type="checkbox"/> non										
83. Avez-vous les jambes lourdes?	Qm 86	<input type="checkbox"/> oui	<input type="checkbox"/> non										
84. Lorsque vous marchez, vous arrive-t-il fréquemment de ressentir une douleur au(x) mollet(s)?	Qm 87	<input type="checkbox"/> oui	<input type="checkbox"/> non										
Si oui, 84a. cette douleur vous oblige-t-elle à ralentir ou arrêter votre marche?	Qm 87a	<input type="checkbox"/> oui	<input type="checkbox"/> non										
84b. disparaît-elle en moins de cinq minutes?	Qm 87b	<input type="checkbox"/> oui	<input type="checkbox"/> non										
85. Avez-vous remarqué depuis moins de six mois que votre appétit a nettement diminué?	Qm 88	<input type="checkbox"/> oui	<input type="checkbox"/> non										
86. Avez-vous observé une perte de poids involontaire de plus de 5 kg (sans faire de régime), au cours des derniers mois?	Qm 89	<input type="checkbox"/> oui	<input type="checkbox"/> non										
87. Avez-vous noté sur la peau, des boutons ou des croûtes qui ne guérissent pas, un changement d'aspect d'un grain de beauté?	Qm 90	<input type="checkbox"/> oui	<input type="checkbox"/> non										
88. Avez-vous des ganglions (glandes, grosses) au niveau de l'aisselle, de l'aine ou du cou?	Qm 91	<input type="checkbox"/> oui	<input type="checkbox"/> non										
89. Toussez-vous ou crachez-vous presque tous les jours, trois mois de suite chaque année depuis au moins deux ans?	Qm 92	<input type="checkbox"/> oui	<input type="checkbox"/> non										
90. Vous est-il arrivé récemment de cracher du sang en toussant?	Qm 93	<input type="checkbox"/> oui	<input type="checkbox"/> non										
91. Etes-vous enrôlé(e) en permanence depuis quelques semaines?	Qm 94	<input type="checkbox"/> oui	<input type="checkbox"/> non										
92. Avez-vous des difficultés récentes pour avaler, particulièrement le pain ou la viande?	Qm 95	<input type="checkbox"/> oui	<input type="checkbox"/> non										
93. Avez-vous des douleurs, des crampes ou des brûlures d'estomac?	Qm 96	<input type="checkbox"/> oui	<input type="checkbox"/> non										
94. Avez-vous une ou plusieurs hernie(s) de la paroi abdominale (ombilicale, inguinale droite ou gauche)?	Qm 97	<input type="checkbox"/> oui	<input type="checkbox"/> non										
95. Avez-vous fréquemment des douleurs abdominales?	Qm 98	<input type="checkbox"/> oui	<input type="checkbox"/> non										
96. Etes-vous constipé(e) chronique?	Qm 99	<input type="checkbox"/> oui	<input type="checkbox"/> non										
97. Avez-vous souvent la diarrhée?	Qm 100	<input type="checkbox"/> oui	<input type="checkbox"/> non										
98. Vos selles se sont-elles récemment modifiées (diarrhée, constipation, alternance des deux)?	Qm 101	<input type="checkbox"/> oui	<input type="checkbox"/> non										
99. Avez-vous du sang dans les selles ou saignez-vous par l'anus?	Qm 102	<input type="checkbox"/> oui	<input type="checkbox"/> non										
Si oui, 99a - hémorroïdes connues	Qm 102a	<input type="checkbox"/> oui	<input type="checkbox"/> non										
100. Avez-vous fait un test de recherche de sang dans les selles, type Hémocult?	hemocult	<input type="checkbox"/> oui	<input type="checkbox"/> non										
101. Urinez-vous du sang?	Qm 103	<input type="checkbox"/> oui	<input type="checkbox"/> non										
102. Avez-vous une gêne pour uriner et/ou vous levez-vous plusieurs fois la nuit pour uriner?	Qm 104	<input type="checkbox"/> oui	<input type="checkbox"/> non										

103. Avez-vous des douleurs rhumatismales ? **Qm 105**

oui non

Si oui, 103a. (plusieurs réponses possibles) :

- | | |
|--|---------------------------------------|
| <input type="checkbox"/> mains (1) | <input type="checkbox"/> poignets (2) |
| <input type="checkbox"/> épaules (4) | <input type="checkbox"/> pieds (5) |
| <input checked="" type="checkbox"/> genoux (7) | <input type="checkbox"/> hanches (8) |
| <input type="checkbox"/> dos (10) | <input type="checkbox"/> reins (11) |

- | |
|--|
| <input type="checkbox"/> coudes (3) |
| <input type="checkbox"/> chevilles (6) |
| <input type="checkbox"/> cou (9) |

Qm 105a

Qm 105b

104. Avez-vous eu une ou plusieurs crises de sciatique ? **Qm 106**

oui non

105. Avez-vous des maux de tête ? **Qm 107**

oui non

Qm 107a Si oui, 105a.

migraine 1

arthrose cervicale 2

autre 3

106. Ces deux dernières années avez-vous consulté un dentiste ? (une seule réponse possible) **Qm 108**

- jamais 1 1 fois par an 2 depuis moins de 6 mois 3 davantage 4

107. Ces deux dernières années avez-vous consulté un médecin (généraliste ou spécialiste) ? (une seule réponse possible) **Qm 109**

- jamais 1 1 fois par an 2 2 fois par an 3 davantage 4

108. Avez-vous eu depuis moins de 1 an : électrocardiogramme mesure de la tension artérielle **exa_1an**

(plusieurs réponses possibles)

108a - Utilisez-vous un appareil pour prendre vous-même votre tension (automesure) ? **auto-HTA**

oui non

109. Avez-vous donné du sang depuis moins de 3 mois ? **don_sang**

oui non

110. Etes-vous suivi(e) ou traité(e) pour une ou plusieurs maladies ? **Qm 110**

oui non

Si oui, 110a. -- laquelle (ou lesquelles) ? 

111. Utilisez-vous régulièrement une BPAP / CPAP (masque de ventilation nocturne) contre l'apnée du sommeil ? oui non **BPAP**

MEDICAMENTS PRIS ACTUELLEMENT REGULIEREMENT

112. Aucun **Qm 113**

oui non

113. Aspirine **Qm 114**

oui non

114. Cortisone **Qm 115**

oui non

115. Anti-inflammatoire (autre qu'aspirine et cortisone) **Qm 116**

oui non

116. Laxatif **Qm 117**

oui non

117. Diurétique : (plusieurs réponses possibles) **Qm 118** pour faire baisser la tension 1 pour maigrir 2

pour uriner 3

118. Médicament contre l'hypertension artérielle (autre que diurétique) **Qm 119**

oui non

119. Médicament contre l'insuffisance cardiaque (autre que diurétique) **tt_inscard**

oui non

120. Médicament contre l'asthme **tt_asthme**

oui non

121. Médicament contre la bronchite chronique **tt_bronc**

oui non

122. Anticoagulant (autre qu'aspirine) **Qm 120**

oui non

123. Médicament pour faire baisser le cholestérol **Qm 121**

oui non

124. Médicament pour faire baisser l'acide urique **Qm 122**

oui non

125. Médicament contre le diabète **Qm 123**

oui non

126. Médicament pour maigrir (autre que diurétique) **Qm 124**

oui non

127. Médicament pour dormir **Qm 125**

oui non

128. Médicament contre l'anxiété, la dépression **Qm 126**

oui non

129. Médicament contre l'épilepsie **Qm 127**

oui non

130. Autres médicaments **Qm 128**

oui non

Sexe :

Homme

Femme

QUESTIONNAIRE POUR LES HOMMES

131. Avez-vous des troubles de l'érection ? **tbl_ejec**

oui non

132. Avez-vous noté une augmentation récente du volume d'un testicule ? **Qm 130**

oui non

133. Avez-vous été opéré (plusieurs réponses possibles) **Qm 131** des testicules (1)

de la prostate (2)

134. Prenez-vous des médicaments pour la prostate ? **Qm 132**

oui non

135. Prenez-vous des médicaments contre les troubles de l'érection ? **tt_ejec**

oui non

QUESTIONNAIRE POUR LES FEMMES					
136. Etes-vous enceinte ? <i>Qm 133</i>	<input type="checkbox"/> oui	<input type="checkbox"/> non			
137. Combien avez-vous eu d'enfant(s) ? <i>Qm 134</i>	<input type="checkbox"/> <input type="checkbox"/>				
138. Avez-vous eu un (ou des) enfant(s) d'un poids de naissance égal ou supérieur à 4 kilos ? <i>Qm 135</i>	<input type="checkbox"/> oui	<input type="checkbox"/> non			
139. Au cours d'une grossesse, avez-vous eu de l'hypertension artérielle ? <i>Qm 136</i>	<input type="checkbox"/> oui	<input type="checkbox"/> non			
Si oui 139a. - Avez-vous eu de l'albumine dans les urines ? <i>Qm 136a</i>	<input type="checkbox"/> oui	<input type="checkbox"/> non			
140. Avez-vous eu une (ou plusieurs) fausse(s) couche(s) ou avortement(s) spontané(s) ? <i>Qm 137</i>	<input type="checkbox"/> oui	<input type="checkbox"/> non			
141. Avez-vous eu un (ou des) avortement(s) provoqué(s) ou interruption(s) volontaire(s) de grossesse (I.V.G.) ? <i>Qm 138</i>	<input type="checkbox"/> oui	<input type="checkbox"/> non			
142. Avez-vous été opérée : (<i>plusieurs réponses possibles</i>) <i>Qm 139</i>	<input type="checkbox"/> du sein (en dehors d'une intervention esthétique) (1)	<input type="checkbox"/> de l'utérus (2)	<input type="checkbox"/> des ovaires (3)		
142a. Si la réponse est 2 (<i>une seule réponse possible</i>) <i>Qm 139a-2</i>	<input type="checkbox"/> hysterectomie totale 1	<input type="checkbox"/> hysterectomie partielle 2	<input type="checkbox"/> autre 3		
143. Utilisez-vous une méthode contraceptive ? <i>Qm 140</i>	<input type="checkbox"/> oui	<input type="checkbox"/> non			
Si oui, 143a : <i>ctcep</i>	<input type="checkbox"/> pilule orale 1	<input type="checkbox"/> stérilet 2	<input type="checkbox"/> préservatif 3	<input type="checkbox"/> hormones transdermiques 4	
	<input type="checkbox"/> hormones transvaginales 5	<input type="checkbox"/> ligature des trompes 6	<input type="checkbox"/> autre 7	<i>ctcep 3</i>	
144. Si vous ne prenez pas la pilule, l'avez-vous prise antérieurement pendant plus d'1 an ? <i>Qm 141</i>	<input type="checkbox"/> oui	<input type="checkbox"/> non			
Si oui 144a. L'avez-vous interrompue pour raison médicale ? <i>Qm 141a</i>	<input type="checkbox"/> oui	<input type="checkbox"/> non			
145. Avez-vous remarqué dans l'un de vos seins (ou les deux) une grosseur même de petit volume ? <i>Qm 142</i>	<input type="checkbox"/> oui	<input type="checkbox"/> non			
146. Avez-vous remarqué un écoulement du mamelon ? <i>Qm 143</i>	<input type="checkbox"/> oui	<input type="checkbox"/> non			
147. Avez-vous eu une mammographie au cours des deux dernières années ? <i>Qm 144</i>	<input type="checkbox"/> oui	<input type="checkbox"/> non			
148. Vos dernières règles datent de (<i>une seule réponse</i>) <i>Qm 145</i>	<input type="checkbox"/> moins de 2 mois (1)	<input type="checkbox"/> 2 à 6 mois (2)	<input type="checkbox"/> 6 mois à 1 an (3)	<input type="checkbox"/> 1 à 5 ans (4)	<input type="checkbox"/> plus de 5 ans (5)
Si vous avez répondu 1 ou 2 à la question 148 :					
149. Avez-vous vos règles à intervalles réguliers ? <i>Qm 146</i>	<input type="checkbox"/> oui	<input type="checkbox"/> non			
150. Vos règles sont-elles abondantes et/ou accompagnées de caillots ? <i>Qm 148</i>	<input type="checkbox"/> oui	<input type="checkbox"/> non			
151. Avez-vous remarqué des pertes de sang (rouges, rosées, ou marrons) en dehors de vos règles, <i>Qm 149</i> ou après un rapport sexuel ?	<input type="checkbox"/> oui	<input type="checkbox"/> non			
Si vous avez répondu 3, 4 ou 5 à la question 148 :					
152. Etes-vous ménopausée ? <i>Qm 147</i>	<input type="checkbox"/> Oui 1	<input type="checkbox"/> Non 2	<input type="checkbox"/> Ne sait pas 3	<i>menopause</i>	
Si oui, (<i>une seule réponse possible</i>)					
152a. depuis combien de temps: 1 <input type="checkbox"/> moins d'un an 2 <input type="checkbox"/> de un à deux ans 3 <input type="checkbox"/> de deux ans à 5 ans 4 <input type="checkbox"/> plus	<i>tps_meno</i>				
152b. Vous arrive-t-il depuis l'arrêt de vos règles, d'avoir des saignements, même minimes ? <i>Qm 150a</i>	<input type="checkbox"/> oui	<input type="checkbox"/> non			
152c. Avez-vous des bouffées de chaleur ? <i>Qm 150b</i>	<input type="checkbox"/> oui	<input type="checkbox"/> non			
152d. Avez-vous ou avez-vous eu un traitement pour la ménopause ? <i>Qm 150c</i>	<input type="checkbox"/> oui	<input type="checkbox"/> non			
152e. Si oui, lequel ? (<i>une seule réponse</i>) <i>Qm 150d</i>	<input type="checkbox"/> oral 1	<input type="checkbox"/> transdermique 2	<input type="checkbox"/> transvaginal 3	<i>tt_menop</i>	
153f. depuis combien d'années ? <i>tps_ttmen</i>	<input type="checkbox"/> <input type="checkbox"/>				
152g. Si arrêté, combien d'années l'avez-vous pris en tout ? <i>Qm 150e</i>	<input type="checkbox"/> <input type="checkbox"/>		<i>stop_ttmen</i>		
153. Avez-vous un suivi gynécologique régulier (une fois par an) ? <i>Qm 151</i>	<input type="checkbox"/> oui	<input type="checkbox"/> non			
154. En ce qui concerne le frottis de dépistage du cancer du col de l'utérus, cochez parmi les propositions suivantes celle qui correspond à votre cas (une seule réponse) :	<i>frottis</i>				
1 <input type="checkbox"/> je n'ai jamais eu de frottis 2 <input type="checkbox"/> j'ai eu un frottis dans les 12 derniers mois 3 <input type="checkbox"/> j'ai eu un frottis il y a moins de 3 ans	<input type="checkbox"/> oui	<input type="checkbox"/> non			
155. Etes-vous gênée par des pertes blanches abondantes ? <i>Qm 153</i>	<input type="checkbox"/> oui	<input type="checkbox"/> non			
156. Avez-vous des envies impérieuses d'uriner ? <i>Qm 154</i>	<input type="checkbox"/> oui	<input type="checkbox"/> non			
157. Perdez-vous parfois vos urines lorsque vous toussez, courez, riez, etc ... ? <i>Qm 155</i>	<input type="checkbox"/> oui	<input type="checkbox"/> non			

Lieu de naissance de votre père (Pays ou département) ? Qm 260

Lieu de naissance de votre mère (Pays ou département) ? Qm 261

Votre lieu de naissance (Pays ou département) ? Qm 262

ÉQUILIBRE ET STRESS

Répondre par **Oui ou Non** (en cochant la case correspondante).
Les propositions suivantes concernent votre état actuel.

1. En ce moment ma vie me semble vide.
2. J'ai du mal à me débarrasser des mauvaises pensées qui me passent par la tête.
3. Je suis sans énergie.
4. Je me sens bloqué(e) ou empêché(e) devant la moindre chose à faire.
5. Je suis déçu(e) et dégoûté(e) de moi-même.
6. Je suis obligé(e) de me forcer pour faire quoi que ce soit.
7. J'ai du mal à faire les choses que j'avais l'habitude de faire.
8. En ce moment je suis triste.
9. J'ai l'esprit moins clair que d'habitude.
10. J'aime moins qu'avant faire les choses qui me plaisent et m'intéressent.
11. Ma mémoire me semble moins bonne que d'habitude.
12. Je suis sans espoir pour l'avenir.
13. En ce moment je me sens moins heureux(se) que la plupart des gens.

Qm 201

Qm 213

- | | |
|------------------------------|------------------------------|
| <input type="checkbox"/> oui | <input type="checkbox"/> non |
| <input type="checkbox"/> oui | <input type="checkbox"/> non |
| <input type="checkbox"/> oui | <input type="checkbox"/> non |
| <input type="checkbox"/> oui | <input type="checkbox"/> non |
| <input type="checkbox"/> oui | <input type="checkbox"/> non |
| <input type="checkbox"/> oui | <input type="checkbox"/> non |
| <input type="checkbox"/> oui | <input type="checkbox"/> non |
| <input type="checkbox"/> oui | <input type="checkbox"/> non |
| <input type="checkbox"/> oui | <input type="checkbox"/> non |
| <input type="checkbox"/> oui | <input type="checkbox"/> non |
| <input type="checkbox"/> oui | <input type="checkbox"/> non |
| <input type="checkbox"/> oui | <input type="checkbox"/> non |
| <input type="checkbox"/> oui | <input type="checkbox"/> non |
| <input type="checkbox"/> oui | <input type="checkbox"/> non |
| <input type="checkbox"/> oui | <input type="checkbox"/> non |

Au cours du dernier mois, combien de fois :

Cochez la case correspondant à votre choix (une seule réponse)

14. Vous a-t-il semblé difficile de contrôler les choses importantes de votre vie? Qm 214
- Jamais 1 Presque jamais 2 Parfois 3 Assez souvent 4 Souvent
15. Vous êtes-vous senti(e) confiant(e) en vos capacités à prendre en main vos problèmes personnels ? Qm 215
- Jamais 1 Presque jamais 2 Parfois 3 Assez souvent 4 Souvent
16. Avez-vous senti que les choses allaient comme vous le vouliez ? Qm 216
- Jamais 1 Presque jamais 2 Parfois 3 Assez souvent 4 Souvent
17. Avez-vous trouvé que les difficultés s'accumulaient à tel point que vous ne pouviez les contrôler ? Qm 217
- Jamais 1 Presque jamais 2 Parfois 3 Assez souvent 4 Souvent

NUTRITION

1. Votre Poids à la naissance? Qm 259

kg	gr	

2. Votre poids à 18 ans ? Qm 251

	kg
--	----

3. Votre poids maximum depuis l'âge de 18 ans ? Qm 252

	kg
--	----

4. Votre poids minimum depuis l'âge de 18 ans ? Qm 253

	kg
--	----

Au cours des 3 derniers mois, votre poids :

5. est-il stable ? Qm 254
6. en augmentation de 2 kilos ou plus ? Qm 255
7. en diminution de 2 kilos ou plus ? Qm 256
8. Contrôlez-vous régulièrement votre alimentation ? Qm 257
9. Faites-vous un régime pour ne pas prendre du poids ou en perdre (Atkins, Dukan, etc.) ? Type-reg
10. S'il vous arrive de faire des régimes, combien de fois depuis l'âge de 18 ans avez-vous perdu et repris 5 kilos à la suite de ces régimes ? Qm 258

- | | |
|------------------------------|------------------------------|
| <input type="checkbox"/> oui | <input type="checkbox"/> non |
| <input type="checkbox"/> oui | <input type="checkbox"/> non |
| <input type="checkbox"/> oui | <input type="checkbox"/> non |
| <input type="checkbox"/> oui | <input type="checkbox"/> non |
| <input type="checkbox"/> oui | <input type="checkbox"/> non |
| <input type="checkbox"/> oui | <input type="checkbox"/> non |

--	--



Merci d'avoir pris le temps de répondre à ce questionnaire

Ne pas remplir cette partie

Code Accueil :

--	--

P 8

MEDICAL



6.2 Annexe 2 : Codes travail IPC

Code	Profession	Code	Profession	Code	Profession
01	Acteur	30	Docker	59	Marin
02	Agriculteur	31	Dynamiteur	60	Mécanicien
03	Aide-soignante	32	Ebéniste	61	Mineur
04	Aiguilleur de train	33	Eboueur	62	Mouleur
05	Animateur	34	Echantillonneur	63	Pêcheur
06	Assembleur	35	Electricien	64	Peintre en bâtiment
07	Assistant dentiste	36	Eleveur	65	Pisciculteur
08	Athlète	37	Employé de cuisine	66	Plâtrier
09	Bagagiste	38	Employé de La Poste	67	Plombier
10	Barman	39	Employé de maintenance	68	Portier et/ou concierge
11	Boucher	40	Employé forestier	69	Poseur de papier peint
12	Boulanger	41	Enseignant en maternelle	70	Professeur d'éducation physique
13	Bûcheron	42	Facteur	71	Réparateur d'ascenseur
14	Capitaine de navire de pêche	43	Femme de chambre	72	Réparateur équipement agricole
15	Carrossier	44	Femme de ménage	73	Réparateur d'équipement lourd
16	Carrelleur	45	Forgeron	74	Réparateur de machine industrielle
17	Charpentier	46	Garde forestier	75	Réparateur téléphonique
18	Chasseur	47	Gardien d'animaux	76	Repassuseuse
19	Chauffagiste	48	Géomètre	77	Représentant
20	Cordonnier	49	Guide	78	Serveur
21	Courrier	50	Horticulteur	79	Tapissier
22	Couturière	51	Infirmier	80	Terrassier
23	Couvreur	52	Ingénieur civil	81	Tôlier
24	Cuisinier	53	Ingénieur des forêts	82	Tourneur
25	Cultivateur	54	Inspecteur de construction	83	Tourneur/fraiseur
26	Danseur	55	Installateur téléphonique	84	Trieur de produit agricole
27	Déménageur	56	Jardinier	85	Vitrer
28	Dératiseur	57	Machiniste	99	Autres activités professionnelles
29	Désinsectisateur	58	Maçon		

Si vous ne trouvez pas votre profession dans la liste, cochez 99 - Autre profession

Code	Activité sportive	Code	Activité sportive	Code	Activité sportive
01	Aérobic/Cardio-training/Vélo d'appartement/Rameur	22	Delta-plane/Parapente	43	Randonnée montagne
02	Apnée	23	Football	44	Roller
03	Aquagym	24	Frîsbee	45	Rugby
04	Arbitrage	25	Golf	46	Skateboard
05	Athlétisme multiple	26	Gymnastique/Stretching	47	Ski-alpin/Surf
06	Aviron	27	Hand-ball	48	Ski de fond/Raquettes
07	Badminton	28	Hockey/Glace ou gazon	49	Ski nautique
08	Base-ball	29	Jeu de boule/Pétanque	50	Spéléologie
09	Basket-ball	30	Karaté/Judo/Kobudo/Rav-maga/Jujitsu/Taekwondo/Boxe thaï/Kendoë	51	Sport nautique
10	Bowling	31	Marche	52	Squash
11	Boxe/Lutte	32	Moto/Karting	53	Surf de mer
12	Canoë/Kayak	33	Musculation	54	Tai chi chuan/Aïkido
13	Cerf volant à traction	34	Natation	55	Tennis
14	Chasse	35	Paint-Ball	56	Tir à l'arc
15	Course/Jogging/Running	36	Patinage	57	Trampoline
16	Danse	37	Parachutisme	58	Vélo/VTT
17	Double dutch	38	Pêche	59	Planche à voile
18	Entraîneur	39	Pelote basque/Jeu de Paume	60	Voile
19	Equitation	40	Ping-Pong	61	Vol à voile
20	Escalade	41	Plongée marine	62	Volley
21	Escrime	42	Polo	63	Water-polo

**PCS 2003 - Niveau 3 - Liste des catégories socioprofessionnelles détaillées**[Niveau 1](#) | [Niveau 2](#) | [Niveau 3](#) | [Niveau 4](#) [Télécharger au format Excel \(5 Ko\)](#) [Télécharger au format dBase \(4 Ko\)](#)

- [11 Agriculteurs sur petite exploitation](#)
- [12 Agriculteurs sur moyenne exploitation](#)
- [13 Agriculteurs sur grande exploitation](#)
- [21 Artisans](#)
- [22 Commerçants et assimilés](#)
- [23 Chefs d'entreprise de 10 salariés ou plus](#)
- [31 Professions libérales](#)
- [33 Cadres de la fonction publique](#)
- [34 Professeurs, professions scientifiques](#)
- [35 Professions de l'information, des arts et des spectacles](#)
- [37 Cadres administratifs et commerciaux d'entreprise](#)
- [38 Ingénieurs et cadres techniques d'entreprise](#)
- [42 Professeurs des écoles, instituteurs et assimilés](#)
- [43 Professions intermédiaires de la santé et du travail social](#)
- [44 Clergé, religieux](#)
- [45 Professions intermédiaires administratives de la fonction publique](#)
- [46 Professions intermédiaires administratives et commerciales des entreprises](#)
- [47 Techniciens](#)
- [48 Contremaîtres, agents de maîtrise](#)
- [52 Employés civils et agents de service de la fonction publique](#)
- [53 Policiers et militaires](#)
- [54 Employés administratifs d'entreprise](#)
- [55 Employés de commerce](#)
- [56 Personnels des services directs aux particuliers](#)
- [62 Ouvriers qualifiés de type industriel](#)
- [63 Ouvriers qualifiés de type artisanal](#)
- [64 Chauffeurs](#)
- [65 Ouvriers qualifiés de la manutention, du magasinage et du transport](#)
- [67 Ouvriers non qualifiés de type industriel](#)
- [68 Ouvriers non qualifiés de type artisanal](#)
- [69 Ouvriers agricoles](#)
- [71 Anciens agriculteurs exploitants](#)
- [72 Anciens artisans, commerçants, chefs d'entreprise](#)
- [74 Anciens cadres](#)
- [75 Anciennes professions intermédiaires](#)
- [77 Anciens employés](#)
- [78 Anciens ouvriers](#)
- [81 Chômeurs n'ayant jamais travaillé](#)
- [83 Militaires du contingent](#)
- [84 Elèves, étudiants](#)
- [85 Personnes diverses sans activité professionnelle de moins de 60 ans \(sauf retraités\)](#)
- [86 Personnes diverses sans activité professionnelle de 60 ans et plus \(sauf retraités\)](#)



Institut national de la statistique et des études économiques

PCS 2003 - Niveau 2 - Liste des catégories socioprofessionnelles
Niveau 1 | Niveau 2 | Niveau 3 | Niveau 4 [Télécharger au format Excel \(4 Ko\)](#) [Télécharger au format dBase \(3 Ko\)](#)

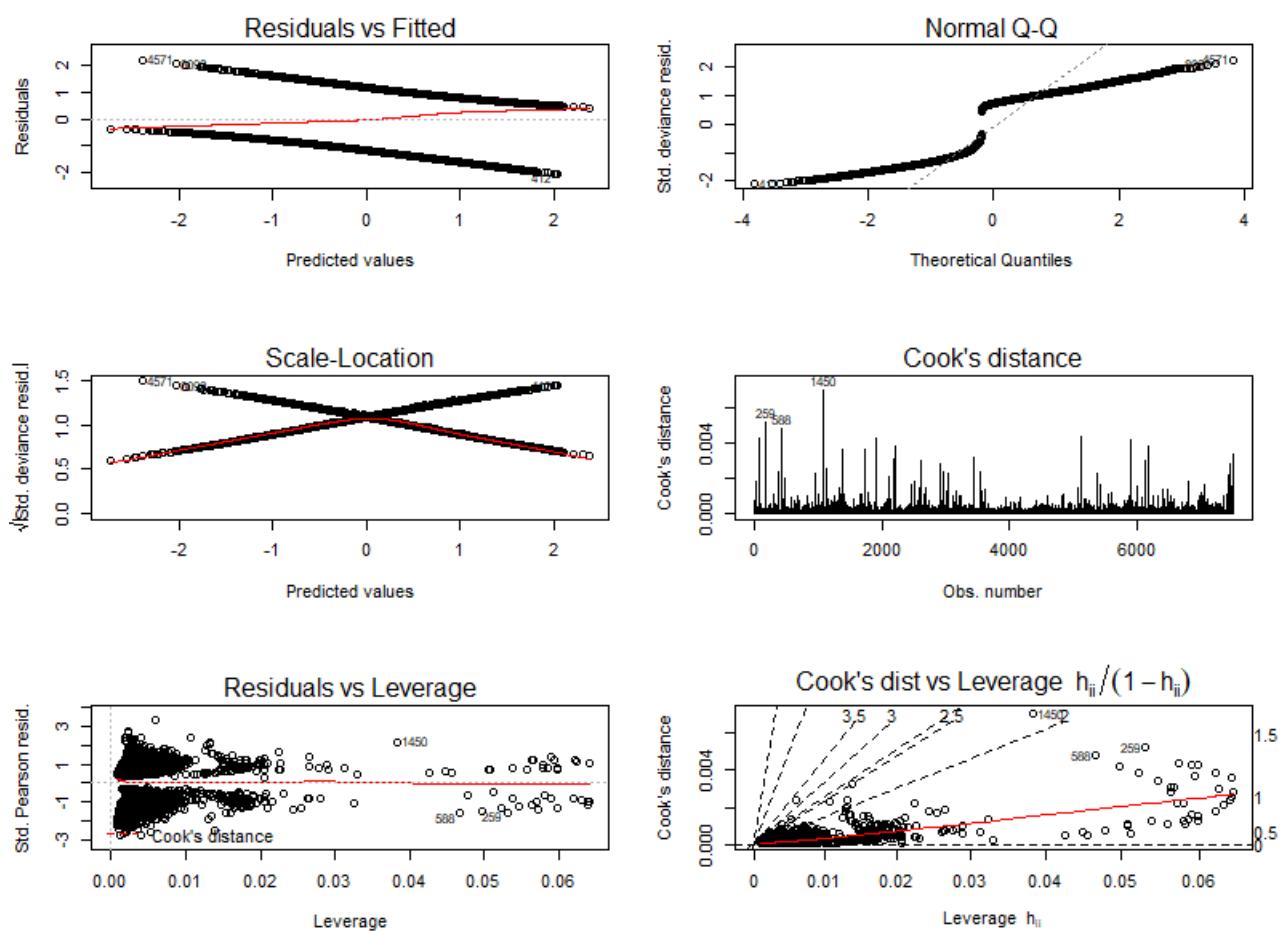
- [10 Agriculteurs exploitants](#)
- [21 Artisans](#)
- [22 Commerçants et assimilés](#)
- [23 Chefs d'entreprise de 10 salariés ou plus](#)
- [31 Professions libérales et assimilés](#)
- [32 Cadres de la fonction publique, professions intellectuelles et artistiques](#)
- [36 Cadres d'entreprise](#)
- [41 Professions intermédiaires de l'enseignement, de la santé, de la fonction publique et assimilés](#)
- [46 Professions intermédiaires administratives et commerciales des entreprises](#)
- [47 Techniciens](#)
- [48 Contremaîtres, agents de maîtrise](#)
- [51 Employés de la fonction publique](#)
- [54 Employés administratifs d'entreprise](#)
- [55 Employés de commerce](#)
- [56 Personnels des services directs aux particuliers](#)
- [61 Ouvriers qualifiés](#)
- [66 Ouvriers non qualifiés](#)
- [69 Ouvriers agricoles](#)
- [71 Anciens agriculteurs exploitants](#)
- [72 Anciens artisans, commerçants, chefs d'entreprise](#)
- [73 Anciens cadres et professions intermédiaires](#)
- [76 Anciens employés et ouvriers](#)
- [81 Chômeurs n'ayant jamais travaillé](#)
- [82 Inactifs divers \(autres que retraités\)](#)

		Assures	Ayants-droits
10	Exploitants agricoles	80	81-82-83-84
21	Artisans	.10-80	81-82-83-84
22	Commerçants et assimilés	.10-80	81-82-83-84
23	Chef d'entreprise de 10 Salariés ou +	80	81-82-83-84
31	Professions libérales	.10-80	81-82-83-84
32	Cadre de la fonction publique Professions intellectuelles et artistiques	80	81-82-83-84
36	Cadres d'entreprise	80	81-82-83-84
41	Professions intermédiaires de l'enseignement , de la santé, de la fonction publique et assimilés	.00	.00
46	Professions intermédiaires administratives et commerciales des entreprises	.00	.00
47	Techniciens		
48	Contremaires, agents de maîtrise		
51	Employés de la fonction publique		
54	Employés administratifs d'entreprise		
55	Employés de commerce		
56	Personnels de service directs aux particuliers		
61	Ouvriers qualifiés		
66	Ouvriers non qualifiés		
69	Ouvriers agricoles		
71	Anciens agriculteurs et exploitants		
72	Anciens artisans, commerçants, chefs d'entreprise	20	21-22-23-24
74	Anciens cadres	20	21-22-23-24
75	Anciennes professions intermédiaires	20	21-22-23-24
77	Anciens employés	20	21-22-23-24
78	Anciens ouvriers	20	21-22-23-24
81	Chômeurs n'ayant jamais travaillé	10	.01
83	Militaires du contingent		
84	Elèves, étudiants, apprentis, formation professionnelle	.01-10	.01
85	Personnes au foyer sans activité professionnelle	.01	.01
86	Personnes diverses sans activité professionnelle	.01-10-20	.01-21-22-23-24
91	Agriculteurs exploitants chômeurs	85	86-87-88-89
92	Artisans, commerçants, Chefs d'entreprise chô- meurs	30	31-32-33-34
93	Cadres et professions intellectuelles chômeurs	30	31-32-33-34
94	Professions intermédiaires chômeurs	30	31-32-33-34
95	Employés chômeurs	30	31-32-33-34
96	Ouvriers chômeurs	30	31-32-33-34

6.3 Annexe 3 : Classification des codes IPC par activité professionnelle

CODES	
Retraités	71 72 73 74 75 76 77 78
Chômeurs	91 92 93 94 95 96
Inactifs	81 82 84 85 86
Ne sais pas	88
Travailleurs	Le reste

6.4 Annexe 4 : Graphes des résidus du modèle de sélection de variables



6.5 Annexe 5 : Fiche d'appréciation de l'employeur



ANNEXE 1 APPRECIATION DU STAGIAIRE PAR L'EMPLOYEUR

NOM PRENOM : EDPAÑA JEAN-PHILIPPE FONCTION : DÉCEPTEUR RECH. INVEST. de l'employeur
 ACTIVITÉ de l'Entreprise : IMEKA CODE APE : EPST
 ADRESSE : 5C Rue Letellier 91015 Paris du stagiaire
 NOM PRENOM : GODET DELIA MAIL : Odelia-g@katomil.fr du stagiaire
 TEL : 06 27 66 22 75 DIPLOME : P1 NENT
 NIVEAU : L1 L2 L3 M1 M2 DUREE DU STAGE : 12 semaine(s) 5 mois

COMPORTEMENT / PERSONNALITE

ASSIDUITE PONCTUALITE	Toujours ponctuel pas d'absence <input checked="" type="checkbox"/>	Retards et absences exceptionnels <input type="checkbox"/>	Quelques retards ou absences <input type="checkbox"/>	Retards ou absences réguliers <input type="checkbox"/>
AUTONOMIE INITIATIVE	Esprit d'initiative, dynamique et judicieux <input checked="" type="checkbox"/>	Autonome, prend souvent des initiatives <input type="checkbox"/>	Sollicite une assistance fréquemment <input type="checkbox"/>	Aucune indépendance <input type="checkbox"/>
MOTIVATION	Très motivé, vif intérêt pour le projet <input checked="" type="checkbox"/>	Curiosité éveillée <input type="checkbox"/>	Intérêt limité <input type="checkbox"/>	Non motivé <input type="checkbox"/>
CONSCIENCE PROFESSIONNELLE	Excellent Conscience professionnelle <input checked="" type="checkbox"/>	Bonne Conscience professionnelle <input type="checkbox"/>	Moyenne Conscience professionnelle <input type="checkbox"/>	Conscience professionnelle insuffisante <input type="checkbox"/>

PARTAGE DU SAVOIR ET DU SAVOIR-FAIRE

ECOUTE SOCIABILITE	Attentif aux autres, a l'esprit d'équipe <input checked="" type="checkbox"/>	Peut se rendre disponible pour les autres <input type="checkbox"/>	Peu sociable <input type="checkbox"/>	Réservé, en retrait <input type="checkbox"/>
OUVERTURE D'ESPRIT ADAPTABILITE	Très ouvert, étudie toutes les idées <input checked="" type="checkbox"/>	Ouvert, accepte les conseils <input type="checkbox"/>	Reste sur son idée même si elle est fausse <input type="checkbox"/>	Systématiquement contrariant <input type="checkbox"/>
PARTICIPATION INTEGRATION	Offre spontanément son aide, présente son travail <input checked="" type="checkbox"/>	Participation irrégulière <input type="checkbox"/>	Présente ses résultats suite à des demandes <input type="checkbox"/>	Ne collabore pas <input type="checkbox"/>
METHODOLOGIE ORGANISATION	Très bonne organisation, bonne coordination <input checked="" type="checkbox"/>	Bien organisé <input type="checkbox"/>	Organisation moyenne <input type="checkbox"/>	Pas organisé <input type="checkbox"/>

COMPETENCES/CAPACITES

COMPREHENSION	Excellent compréhension <input checked="" type="checkbox"/>	Bonne compréhension <input type="checkbox"/>	Compréhension limitée <input type="checkbox"/>	A des difficultés de compréhension <input type="checkbox"/>
RYTHME DE TRAVAIL	Rapide et efficace <input checked="" type="checkbox"/>	Rapide avec des erreurs <input type="checkbox"/>	Rapidité moyenne <input type="checkbox"/>	Lenteur <input type="checkbox"/>
COMPETENCES TECHNIQUES	Au delà des compétences attendues <input checked="" type="checkbox"/>	En adéquation avec la mission <input type="checkbox"/>	Juste convenable, fait des efforts <input type="checkbox"/>	Niveau insuffisant <input type="checkbox"/>
CONNAISSANCES THEORIQUES	Au delà de nos attentes <input checked="" type="checkbox"/>	Cohérentes par rapport à la mission <input type="checkbox"/>	Juste convenable, fait des efforts <input type="checkbox"/>	Niveau insuffisant <input type="checkbox"/>

Estimez-vous que l'étudiant ait mis en œuvre, pendant ce stage, les compétences requises pour intégrer un poste de travail similaire?

OUI

NON

Recommanderiez-vous cet étudiant à une entreprise ? OUI

NON

Appréciation d'ensemble/commentaires :

Excellent étudiant, vif, très motivé. Un réel plaisir de travailler avec lui - Guendj!

Bibliographie

- [1] Code de la sécurité sociale - Article L351-8.
- [2] Javier Alvarez-Galvez, Maria Luisa Rodero-Cosano, Emma Motrico, Jose A. Salinas-Perez, Carlos Garcia-Alonso, and Luis Salvador-Carulla. The impact of socio-economic status on self-rated health : study of 29 countries using European social surveys (2002-2008). *International Journal of Environmental Research and Public Health*, 10(3) :747–761, February 2013.
- [3] Philippe Bizouarn. L'éco-épidémiologie - Vers une épidémiologie de la complexité. *médecine/sciences*, 32(5) :500–505, May 2016.
- [4] Thomas F. Crossley and Steven Kennedy. The reliability of self-assessed health status. *Journal of Health Economics*, 21(4) :643–658, July 2002.
- [5] Jean-Philippe Empana, Kathy Bean, Catherine Guibout, Frédérique Thomas, Annie Bingham, Bruno Pannier, Pierre Boutouyrie, Xavier Jouven, and PPS3 Study Group. Paris Prospective Study III : a study of novel heart rate parameters, baroreflex sensitivity and risk of sudden death. *European Journal of Epidemiology*, 26(11) :887–892, November 2011.
- [6] Enrique Fatas, Juan A. Lacomba, and Francisco Lagos. An Experimental Test on Retirement Decisions. *Economic Inquiry*, 45(3) :602–614, 2007.
- [7] Mathilde Frérot, Annick Lefebvre, Simon Aho, Patrick Callier, Karine Astruc, and Ludwig Serge Aho Glélé. What is epidemiology ? Changing definitions of epidemiology 1978-2017. *PLoS ONE*, 13(12), December 2018.
- [8] Carol Mansyur, Benjamin C. Amick, Ronald B. Harrist, and Luisa Franzini. Social capital, income inequality, and self-rated health in 45 countries. *Social Science & Medicine* (1982), 66(1) :43–56, January 2008.
- [9] José C. Millán-Calenti, Alba Sánchez, Trinidad Lorenzo, and Ana Maseda. Depressive symptoms and other factors associated with poor self-rated health in the elderly : gender differences. *Geriatrics & Gerontology International*, 12(2) :198–206, April 2012.
- [10] J. M. Mossey and E. Shapiro. Self-rated health : a predictor of mortality among the elderly. *American Journal of Public Health*, 72(8) :800–808, August 1982.
- [11] Susan Nolen-Hoeksema. Gender differences in depression. *Current directions in psychological science*, 10(5) :173–176, 2001.
- [12] Catherine Sass, R. Guéguen, J.-J. Moulin, L. Abric, V. Dauphinot, C. Dupré, J.-P. Giordanna, F. Girard, C. Guenot, Émilie Labbé, Emilio La Rosa, P. Magnier, Emmanuelle Martin, B. Royer, M. Rubirola, and Laurent Gerbaud. Comparaison du score individuel de précarité des Centres d'exams de santé, EPICES, à la définition socio-administrative de la précarité. *Sante Publique*, Vol. 18(4) :513–522, 2006.
- [13] Sheryl L Warttig, Mark J Forshaw, Jane South, and Alan K White. New, normative, English-sample data for the Short Form Perceived Stress Scale (PSS-4). *Journal of Health Psychology*, 18(12) :1617–1628, December 2013.

- [14] Hugo Westerlund, Mika Kivimäki, Archana Singh-Manoux, Maria Melchior, Jane E. Ferrie, Jaana Pentti, Markus Jokela, Constanze Leineweber, Marcel Goldberg, Marie Zins, and Jussi Vahtera. Self-rated health before and after retirement in France (GAZEL) : a cohort study. *Lancet (London, England)*, 374(9705) :1889–1896, December 2009.
- [15] Shunquan Wu, Rui Wang, Yanfang Zhao, Xiuqiang Ma, Meijing Wu, Xiaoyan Yan, and Jia He. The relationship between self-rated health and objective health status : a population-based study. *BMC public health*, 13 :320, April 2013.