

MÉMOIRE DE PREMIÈRE ANNÉE DE MASTER

Baromètre Covid-19



Alban EXPOSITO
Antoine LUCIANO
Yanis BAKHTAOUI

Responsables de mémoire :
Brigitte DORMONT
Paul DOURGNON
Responsable de majeur :
Julien STOEHR

2020-2021

Table des matières

Remerciements	2
0 Introduction	3
1 Données	5
1.1 Datacovid	5
1.2 Description des données	5
1.3 Représentativité des données	7
1.3.1 Sexe	7
1.3.2 Âge	7
1.3.3 Régions	8
1.3.4 Emploi	8
1.3.5 Diplôme	9
1.3.6 Evolution de la Covid-19 en France	9
1.4 Traitement des données	10
2 Les gestes barrières	11
2.1 Description des gestes	11
2.1.1 Introduction gestes barrières	11
2.1.2 La temporalité dans le respect des gestes barrières	12
2.1.3 Corrélation entre les gestes barrières	13
2.2 Regroupement des gestes	14
2.2.1 Préparation du regroupement	14
2.2.2 Analyse de la qualité du regroupement	18
2.3 Regroupement numérique des gestes dans les 3 groupes et traitement des données	21
3 Les travailleurs hors de leur domicile	23
3.1 Qui sont-ils ?	23
3.1.1 La variable HorsDomicile	23
3.1.2 Régression logistique	24
3.1.3 Profil des travailleurs à l'extérieur	26
3.1.4 Le respect des gestes barrières parmi ceux travaillant à l'extérieur	30
3.2 Exposition à l'infection à la Covid-19	31
3.2.1 La variable Infecte	31
3.2.2 Quels sont les facteurs augmentant les risques d'infection ?	32
3.3 La santé mentale en période de pandémie	33
3.3.1 Le score PHQ4	33
3.3.2 Régression logistique polytomique ordonnée	34
3.3.3 Profil des personnes souffrant de problèmes de santé mentale durant la crise	36

4	Conclusion	38
	Bibliographie	39
A	Questionnaire de la vague 1	40

Tous les résultats de ce mémoire sont reproductibles. Les fichiers R utilisés sont disponibles ici.

Remerciements

Nous tenions à remercier nos deux encadrants, Brigitte Dormont et Paul Dourgnon, pour leur accompagnement tout au long de ce projet. Avoir les deux points de vue économétrique et statistique fut très intéressant et formateur, et nous a permis d'appréhender ce sujet sous différents angles.

Nous souhaitions également remercier Pierre Cardaliaguet pour nous avoir permis de réaliser ce mémoire malgré le contexte défavorable, et Julien Stoehr pour son soutien tout au long de ce projet.

Enfin, merci à Sophie Morin, research manager chez Ipsos, et Frédéric Lemoine, responsable Data de DataCovid, pour leurs réponses à nos interrogations respectivement sur la création du panel et sur l'API de Datacovid.

0 Introduction

Ce mémoire nous a été proposé par Mme Brigitte Dormont en collaboration avec Mr Paul Dourgnon. Le sujet consiste à étudier l'impact de la Covid-19 en France sur les différentes classes sociales. Pour cela, nous avions à disposition une base de données tirée du site DataCovid. Nous allons plus spécifiquement nous intéresser à la différence d'exposition au virus entre les personnes de classes sociales différentes. Une première approche naïve de la question serait de considérer que les « pauvres » respecteraient moins les gestes barrières et les interdictions mises en place pour stopper le virus du fait d'un accès à l'information moindre, et qu'ils seraient donc plus exposés. Une seconde intuition consisterait à penser qu'ils sont plus exposés au virus puisqu'ils auraient moins accès aux masques et au gel qui les protégeraient. Ils seraient également obligés de travailler en dehors de chez eux, à l'inverse des plus riches qui auraient plus de possibilités de télétravailler.

Cette problématique s'inscrit dans le thème des inégalités sociales de santé. Selon un document coordonné par Thierry Lang, épidémiologiste, et co-écrit par des médecins et des épidémiologistes, les inégalités sociales de santé se traduisent par une différence d'espérance de vie de 7 ans entre ouvriers et cadre supérieurs. Pour citer le document : « les déterminants de santé ne concernent pas seulement les soins individuels et notamment l'accès aux soins. De nombreuses études ont mis en évidence d'autres déterminants socio-économiques : l'éducation, l'accès à l'emploi, les conditions de travail, l'âge de la retraite, la politique du logement, les relations sociales, les politiques redistributives à travers la fiscalité et les aides financières directes, mais aussi des comportements individuels de santé largement liés aux facteurs sociaux ».

Notre objectif est donc de trouver les déterminants de santé expliquant l'exposition à la Covid-19, à l'aide d'une approche mathématique et analytique. Pour cela, nous avons accès à une base de données en open-data sur le site DataCovid. Cette base de données a été construite grâce à un questionnaire effectué sur un panel IPSOS sur 8 vagues.

Nos axes de recherches principaux seront les suivants : nous allons d'abord étudier les gestes barrières et le travail à l'extérieur, pour ensuite analyser le lien entre le fait de travailler à l'extérieur et l'exposition à la Covid-19. Nous compléterons ensuite cette analyse par le lien entre exposition et classe sociale, pour enfin déterminer les facteurs de dépression durant cette épidémie.

Pour faciliter l'interprétation des résultats, il est important de relier les différentes vagues de l'étude à la temporalité. C'est dans ce but que nous avons créé une frise chronologique avec les grandes dates de la Covid-19 en France (Figure 1). Pour rappel, le premier cas de Covid-19 en France est apparu le 25 janvier et le premier décès est survenu le 27 février. L'annonce du premier confinement par le président de la République a été effectuée le 16 mars. Celle-ci visait à faire peur de façon à faire comprendre aux français les dangers de cette pandémie, et a pu agir sur la santé mentale de la population française. Notre première vague commence le 7 avril, en plein confinement. La vague 6 débute le lendemain du déconfinement, le 12 mai. En résumé, les vagues 1 à 5 ont lieu lors du confinement et les vagues 6, 7 et 8 ont lieu durant le déconfinement.

La COVID-19 en FRANCE

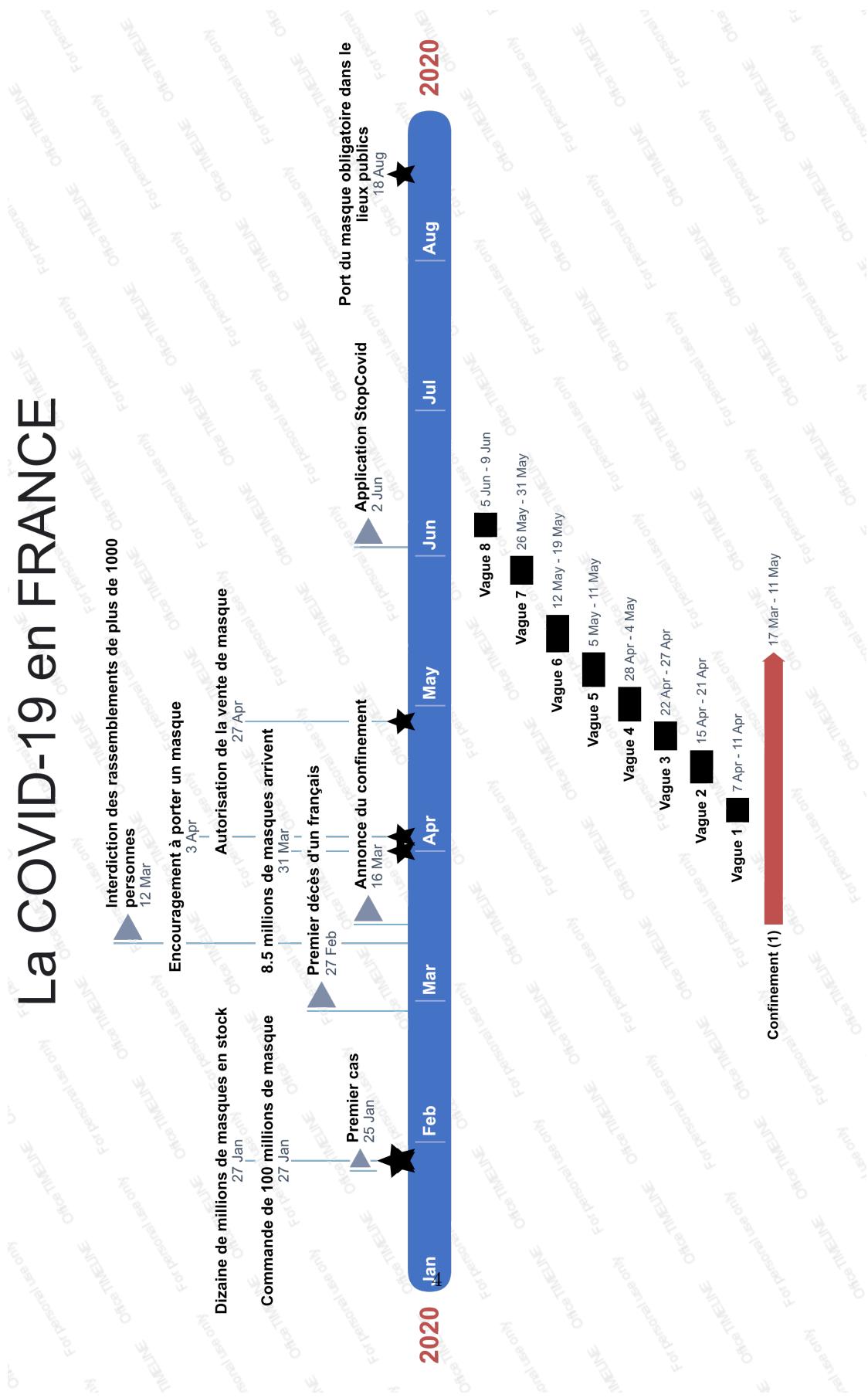


FIGURE 1 – Frise chronologique de la Covid-19 en France

1 Données

Dans cette partie, nous allons détailler le travail préliminaire à notre étude, c'est-à-dire tout ce que nous avons mis en œuvre afin d'utiliser les données le plus efficacement. Dans un premier temps, nous allons présenter l'initiative Datacovid. Dans un second temps, nous allons décrire les données et le questionnaire auquel ont répondu les panélistes. Enfin, nous allons nous intéresser à la représentativité des données. Puis, pour conclure, nous décrirons la manière dont nous avons traité les données.

1.1 Datacovid

Datacovid est une initiative visant "à collecter et mettre à disposition de tous des données construites, utiles et complémentaires des sources existantes, pour la gestion à court et long terme de l'épidémie et de ses multiples impacts". C'est une initiative d'une équipe de scientifiques voulant stimuler la collaboration entre spécialistes afin d'obtenir les meilleurs résultats sur l'épidémie dans le but d'en comprendre ses causes et conséquences. Plus précisément, Datacovid est un site internet répertoriant différentes données en libre-service de panélistes sondés grâce à l'organisme IPSOS, partenaire de cette initiative. Ces données ont été obtenues grâce à un questionnaire auquel les panélistes ont répondu sur huit périodes, qu'on appellera vagues. Sur chaque vague, 5000 personnes sont interrogées et choisies selon la méthode IPSOS (la 2ème vague comporte 5001 observations). Au total, nous avons donc 40001 observations.

Qu'est-ce que l'organisme IPSOS ?

IPSOS est une entreprise de sondages française et une société internationale de marketing d'opinion. Dans le but de comprendre au mieux les données, nous avons contacté IPSOS pour qu'ils nous expliquent la façon dont est sélectionné leur panel. L'objectif d'IPSOS est de créer un panel le plus représentatif possible de la population que l'on veut étudier. C'est dans ce but que le panel IPSOS respecte les critères suivants :

- Ce sont des individus volontaires et qualifiés (bons profils) pour un usage à long terme.
- C'est plus de 4.2 millions de panélistes dans 61 pays du monde.
- En France, on observe 208 000 panélistes.
- Les répondants sont réels, disponibles (on n'intérroge pas une même personne dans un court délai), engagés et uniques.
- Ils sont recrutés en effectuant un questionnaire pour déterminer leur profil.

Les données sont accessibles sous forme de classeur Excel mais aussi d'API. Elles sont publiées avec le questionnaire de la vague en question et une analyse des résultats émise sous forme de diaporama.

Il y a également un blog dans lequel des articles au sujet du Covid-19 sont publiés.

1.2 Description des données

Les données sont récoltées à travers un questionnaire en ligne. On peut donc déjà avancer que les personnes les moins aisées de notre panel ont tout de même accès à internet et à un ordinateur, nous n'avons donc pas accès aux classes sociales les plus basses. Les données sont longitudinales, en effet la même population est observée sur plusieurs vagues.

Le questionnaire est constitué de deux parties : le questionnaire cœur, commun à chaque vague, et un module qui diffère selon les vagues. Le questionnaire dure en moyenne une dizaine de minutes. Lorsqu'un individu commence le questionnaire, il va dans la plupart des cas au bout du questionnaire, ainsi les données que nous avons ne comportent que peu de valeurs manquantes. Toutefois, plus nous avançons dans les vagues, plus le nombre de participants n'allant pas au

bout du questionnaire augmente. Lors de la dernière vague, 18% des panélistes ne sont pas allés au bout du questionnaire. Ce taux pourrait s'expliquer par une augmentation du nombre de questions au fur et à mesure des vagues. Cependant, le questionnaire cœur est au début du questionnaire, et ce sont ces réponses qui nous intéressent. Nous avons donc les réponses de la majorité des individus.

Le questionnaire cœur nous renseigne sur l'attitude générale de l'individu face à la COVID, sur le fait que l'individu ait contracté ou été en contact avec la COVID, s'il s'est fait tester, s'il respecte les différents gestes barrières et à quelle fréquence, s'il est dans la capacité de les respecter, et beaucoup d'autres informations. Il nous renseigne également sur sa catégorie socio-professionnelle et sa santé mentale. Ceci est un bref résumé du questionnaire que l'on peut trouver sur le site de DataCovid (Voir l'Annexe A).

Les modules, ou éclairages ponctuels, quant à eux, traitent des sujets suivants :

- Vague 1 : Prise de température sur la disposition du panel à utiliser une application mobile ou de communiquer des informations personnelles afin d'endiguer l'épidémie.
- Vague 2 : Questions générales sur le respect des gestes barrières et sur les attentes liées au futur.
- Vague 3 : Pas de section d'éclairages ponctuels.
- Vague 4 : Impact du confinement sur la vie de l'individu, évolution de la mise en place de des mesures sanitaires au travail et compréhension des mesures et des test mis en place.
- Vague 5 : Impact du confinement sur la vie de l'individu, confiance en les institutions (gouvernement, santé, scientifiques, média, ...).
- Vague 6 : Réaction de l'individu par rapport au déconfinement et à la gestion de la crise par les pouvoirs publics français.
- Vague 7 : Confiance en la science, disposition à utiliser une application mobile.
- Vague 8 : Analyse économique des revenus et de la disposition de l'individu à épargner.

Le seul module que nous utiliserons dans notre étude est celui de la vague 8, qui nous donne des informations sur le revenu. Nous n'utiliserons pas le reste car ce qui nous intéresse est d'analyser l'évolution des comportements sur les 8 vagues. De plus, analyser des données sur un panel de 5000 personnes et sur une courte durée est moins pertinent.

Voici les variables qui vont principalement nous intéresser. Ce sont toutes des variables qualitatives :

- l'âge, réparti en tranches dans la variable **AGE2**.
- le sexe dans la variable **SEXE**.
- la catégorie socio-professionnelle répartie en différentes catégories : Agriculteurs, Indépendants, Cadres supérieurs, Professions intermédiaires, Employés, Ouvriers, Retraités et Inactifs (Variable **PCSI**).
- le niveau de diplôme divisé en modalités : les personnes sans diplôme, celles ayant un certificat de formation générale (CFG), un certificat d'études primaires, ou le Diplôme national du brevet, puis celles qui détiennent un certificat d'aptitude professionnelle (CAP) ou Brevet d'études professionnelles (BEP), enfin les personnes ayant un baccalauréat ou un diplôme équivalent, un bac + ou 2, un bac +3 ou plus. (Variable **C10**)
- les gestes barrières : au nombre de 9, ils comportent tous les mêmes modalités de fréquence de respect (Variables **B2_0** à **B2_8**).

1.3 Représentativité des données

Pour garantir l'efficacité de notre étude future, nous avons commencé par étudier la représentativité des données par rapport à la population française. Nous nous concentrerons sur la représentativité des variables suivantes : le sexe, l'âge, l'emploi et l'éducation. Si nos données ne représentent pas bien la population, nous ne pourrons tirer aucune conclusion pertinente.

1.3.1 Sexe

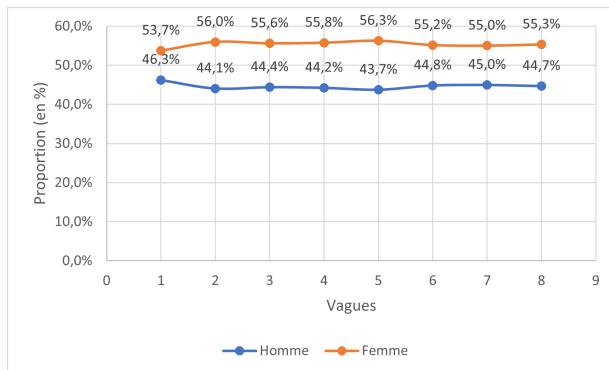


FIGURE 2 – Comparaison de la proportion d'hommes et de femmes dans le panel sur les huit vagues

La proportion d'hommes et de femmes dans la population française est de 52% de femmes pour 48% d'hommes. Dans le panel, elle est en moyenne sur les 8 vagues de 44.6% d'hommes pour 55.35% de femmes. On observe une légère sur-représentation de femmes dans le panel par rapport à la population française, que l'on peut considérer comme négligeable.

1.3.2 Âge

Nous avons comparé la répartition de la moyenne d'âge sur les huit vagues du panel Ipsos avec la répartition de l'âge de la population française (source INSEE). Nous ignorons les moins de 18 ans dans notre étude et donc dans notre comparaison.

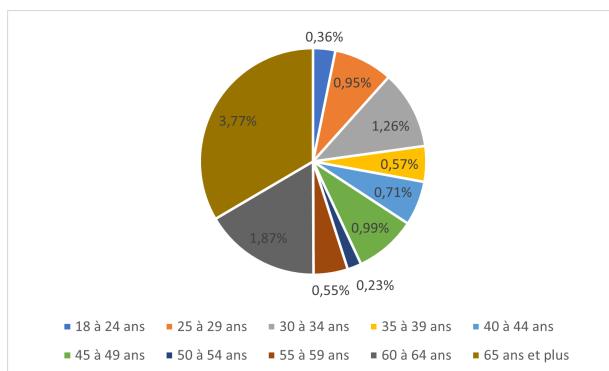


FIGURE 3 – Différence des proportions des tranches d'âges entre la population française et le panel, en moyenne sur les huit vagues

On observe une différence de l'ordre de 4% au niveau des 65 ans et plus, qui est la différence la plus importante. Cette différence est faible et peut être expliquée par le grand nombre de tranches d'âges sélectionnées. Elle peut donc être négligée. Le panel étudié vérifie globalement bien les proportions d'âge de la population française.

1.3.3 Régions

Nous avons aussi étudié la répartition du panel dans les différentes régions de France. On a observé la différence entre le panel et la répartition de la population française en fonction des régions.

Représentativité par région des sondés

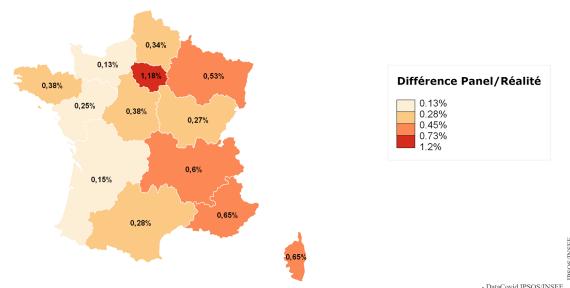


FIGURE 4 – Ecart de la population française avec la population du panel par régions

La plus grosse différence est de l'ordre de 1.2% en Ile-de-France, ce qui est relativement négligeable, puis proche de 0.7% dans le sud-est de la France. Le panel est donc représentatif selon ce critère.

1.3.4 Emploi

Il est très difficile de trouver des chiffres selon les mêmes critères sur l'emploi en France. Cependant, d'après l'INSEE, la part de salariés à temps partiel (hors apprentis) serait de l'ordre de 18%, la proportion d'étudiants dans la population française serait d'environ de 14% et 21% seraient des retraités. Nous pouvons donc comparer ces chiffres avec la distribution de l'emploi du panel.

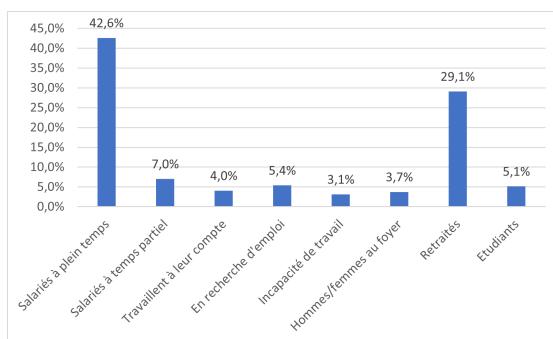


FIGURE 5 – Histogramme de la proportion de l'Emploi dans le panel

On peut donc en conclure que le panel n'est pas représentatif de l'emploi en France. En effet, il y a une sous-représentation des étudiants ainsi que des salariés à temps partiels et une sur-représentation des retraités.

1.3.5 Diplôme

Selon les chiffres de l'INSEE, en France, en 2020, 18.5% de la population sont sans diplôme ou avec un brevet des collèges, 23.8% ont un BEP, CAP ou équivalent, 17.8% ont le baccalauréat, 14.8% ont un bac+2, et 24.8% ont un diplôme supérieur à un bac+2.

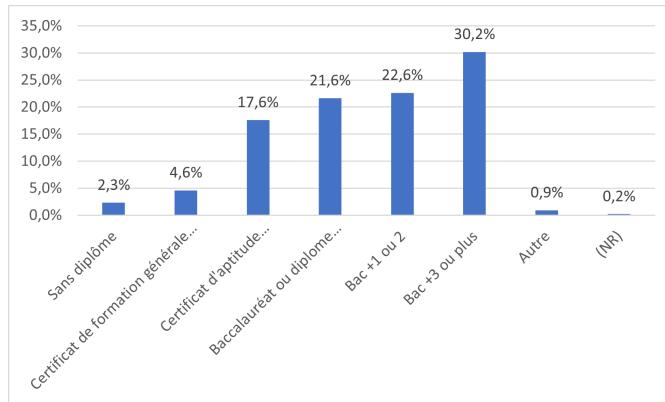


FIGURE 6 – Histogramme de la proportion de l'Education dans le panel

En comparant avec la base de données, on observe une sur-représentativité non-négligeable des diplômés d'au moins bac+3, compensée par une sous-représentativité des personnes sans diplôme ou avec seulement le brevet.

1.3.6 Evolution de la Covid-19 en France

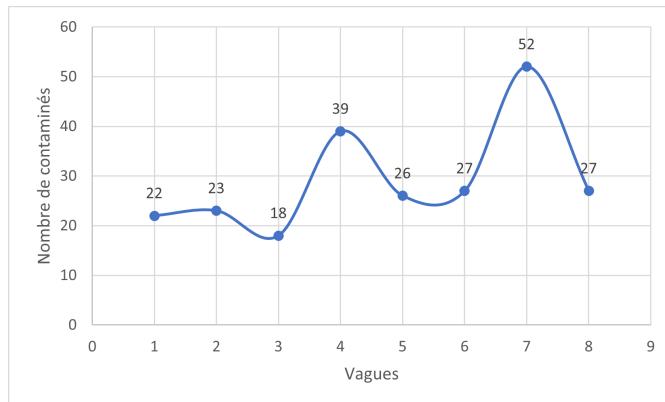


FIGURE 7 – Courbe de l'évolution des contaminés dans le panel

Nous voulons vérifier que le nombre de contaminés du panel augmente au fur et à mesure des vagues. En observant la tendance de la courbe , on conclut que c'est le cas, malgré une baisse lors de la vague 8. Mais, globalement, la pente est cohérente.

1.4 Traitement des données

Afin de faciliter l'accès aux données, DataCovid a mis en place une API. Celle-ci permet d'accéder directement aux données depuis R, à l'aide d'un code simple. Nous compptions l'utiliser, mais des contraintes se révélèrent au fur et à mesure de son utilisation. Déjà, nous nous sommes rendu compte qu'il manquait des données (des variables n'étaient pas présentes), ainsi que l'entièreté de la vague 8. Nous avons donc contacté le responsable de l'API, mais entre chaque échange de mail pour corriger les soucis que nous relevions, nous découvrions d'autres problèmes. De plus, l'API ne permettait de charger que 10 variables maximum. Nous devions alors bricoler lorsque nous avions besoin de plus de 10 variables. Nous avons donc décidé de ne pas utiliser l'API, mais les tableaux Excel fournis sur DataCovid.

Ainsi, afin d'importer nos données sur R, nous avons converti les 8 fichiers correspondant aux 8 vagues, du format TSV au format CSV. Nous avons ensuite créé une fonction *CSVtoR* qui prend en argument une liste de noms de variables et retourne la dataframme composée des 40001 observations sur les variables passées en argument. Enfin, afin de différencier les vagues entre elles , nous avons crée une variable Vague qui prend une valeur entière de 1 à 8 selon la vague à laquelle l'individu a répondu au questionnaire.

L'inconvéniant de cette méthode d'importation est que l'utilisateur doit télécharger les 8 fichiers CSV et placer le fichier R et le dossier comportant les fichiers au même endroit.

2 Les gestes barrières

2.1 Description des gestes

2.1.1 Introduction gestes barrières

Le 6 décembre 2019, en complément de la campagne saisonnière de communication sur la vaccination anti-grippale, l'Assurance Maladie sensibilise pour la première fois aux « gestes barrières », alors au nombre de quatre. « Tousser ou éternuer dans son coude », « Utiliser un mouchoir à usage unique », « Se laver les mains régulièrement » et « Porter un masque jetable quand vous êtes malade ». 4 mois après, en mars 2020, et alors que la France entre en confinement, ces gestes barrières font partie du vocabulaire de tous les jours. À ces 4 déjà utilisés pour la lutte contre la grippe hivernale, se rajoutent « Respecter une distance d'1m », « Eviter les rassemblements », « Rester chez soi », « Utiliser du gel hydroalcoolique » et « Eviter les embrassades ». Progressivement, le port du masque devient recommandé aux personnes saines, et en juin 2020 il est rendu obligatoire dans tous les espaces fermés, puis dans certaines villes. Bien que recommandés, ces gestes ne furent pas respectés par tous les français.

Dans le questionnaire DataCovid, il fut demandé à quelle fréquence ces gestes étaient respectés. Nous allons utiliser ces 9 variables gestes barrières.

- Questions B2_0 à B2_8 : À quel point arrivez-vous à mettre en œuvre le geste suivant ?**
- | | |
|---------------------|---|
| — 1 : Tout le temps | 5 : Jamais même si je pourrais le faire |
| — 2 : Souvent | — 6 : Jamais car je n'ai pas la possibilité de le faire |
| — 3 : Parfois | — 7 : Non réponse |
| — 4 : Rarement | |

	Tout le temps	Souvent	Parfois	Rarement	Jamais même si je pourrais le faire	Jamais car je n'ai pas la possibilité de le faire	Non réponse
lavageMain	65.53086	27.784305	4.589885	1.259969	0.3924902	0.1574961	0.2849929
gelHydroAlcoolique	27.31432	33.634159	17.639559	9.027274	3.7999050	8.3172921	0.2674933
tousserCoude	55.96360	24.889378	9.547261	5.467363	1.8999525	1.9624509	0.2699933
mouchoireUnique	65.66336	17.699558	7.019825	4.652384	2.9824254	1.6524587	0.3299918
distance1M	62.77093	26.614335	5.107372	2.244944	1.2074698	1.7649559	0.2899928
eviteSerrerMainEmbrasser	84.94288	7.804805	2.212445	1.397465	2.0074498	1.3749656	0.2599935
eviteRegroupements	76.94808	14.139647	3.254919	1.477463	1.1824704	2.7449314	0.2524937
resterChezSoi	41.85645	42.886428	7.919802	2.812430	0.7474813	3.3874153	0.3899903
porterMasque	18.48704	28.154296	17.927052	11.642209	6.2948426	17.2170696	0.2774931

FIGURE 8 – Fréquence de respect des gestes barrières au cours des 8 vagues confondues

On observe que certains gestes sont respectés dans l'ensemble, tels que « Eviter de s'embrasser » (85% le font tout le temps), mais que d'autres le sont moins, comme « Gel hydroalcoolique » et « Porter le Masque » (seulement 27.3% et 18.5%, respectivement, le font systématiquement). Ces deux gestes nécessitent du matériel, et sont d'ailleurs ceux pour lesquels le plus de gens ont répondu qu'ils n'avaient pas la possibilité de les appliquer (8.3% et 17.2%).

2.1.2 La temporalité dans le respect des gestes barrières

La première chose que l'on peut remarquer est que, pour les masques et le gel hydroalcoolique, beaucoup de sondés ont considéré ne pas les respecter car n'ayant pas la possibilité de le faire. Etant donné que la France a subi une pénurie de masque au début de l'épidémie, ce n'est pas étonnant. Durant les premières vagues de sondages de DataCovid, beaucoup de personnes n'avaient pas accès à cette ressource.

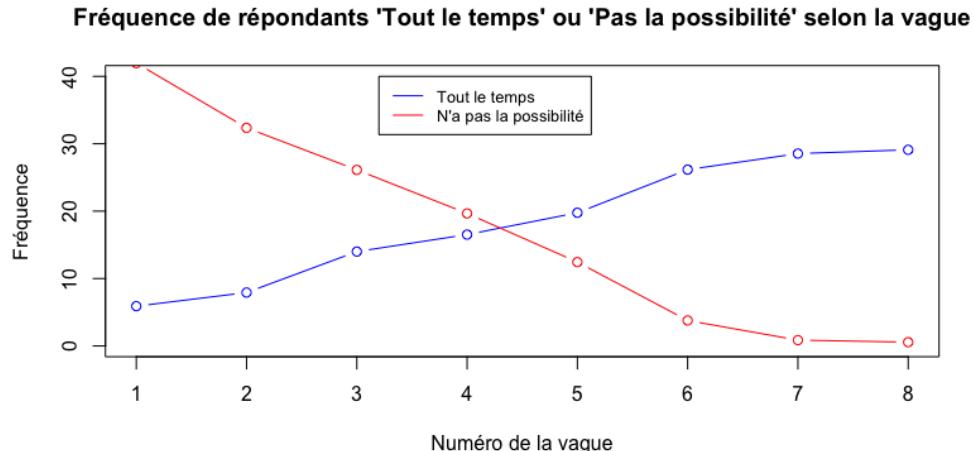


FIGURE 9 – Évolution du port du masque au fil du temps

On observe qu'effectivement plus le temps passe et plus les sondés ont trouvé un moyen d'accès à cette ressource. Cela se traduit par une décroissance du nombre de personnes répondant « Je n'en porte jamais car je n'en ai pas la possibilité » et une croissance du nombre de répondants « Tout le temps ».

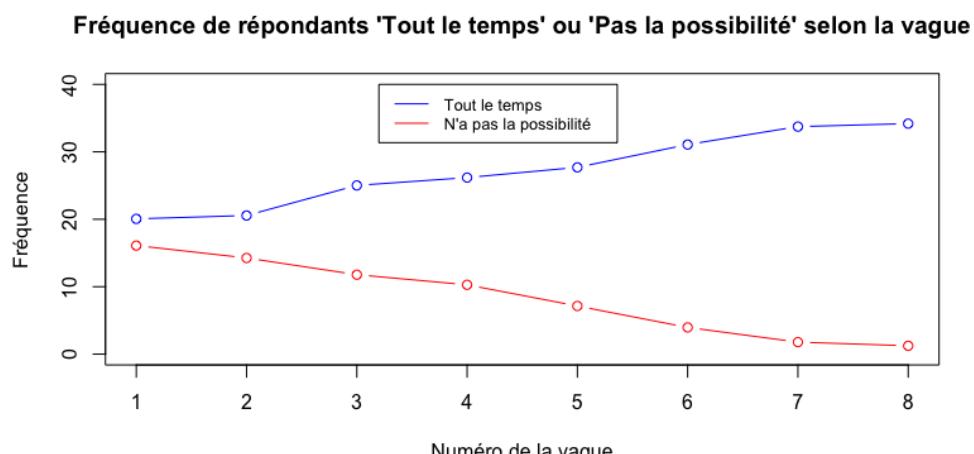


FIGURE 10 – Évolution de l'utilisation de gel hydroalcoolique au fil du temps

On observe la même tendance que pour le port du masque, bien que celle-ci soit moins marquée. Le respect des gestes barrières « Porter un masque » et « Utiliser du gel hydroalcoolique » paraît donc intimement lié à la temporalité. Néanmoins, l'accès aux masques était encore plus compliqué, avec 41.96% de personnes déclarant ne pas avoir accès à des masques, contre 16.10% pour le gel hydroalcoolique. Nous allons vérifier notre intuition à l'aide d'un corrplot. On s'attend à avoir une forte corrélation entre la variable **Vague** et les variables **porterMasque** et **gelHydroalcoolique**.

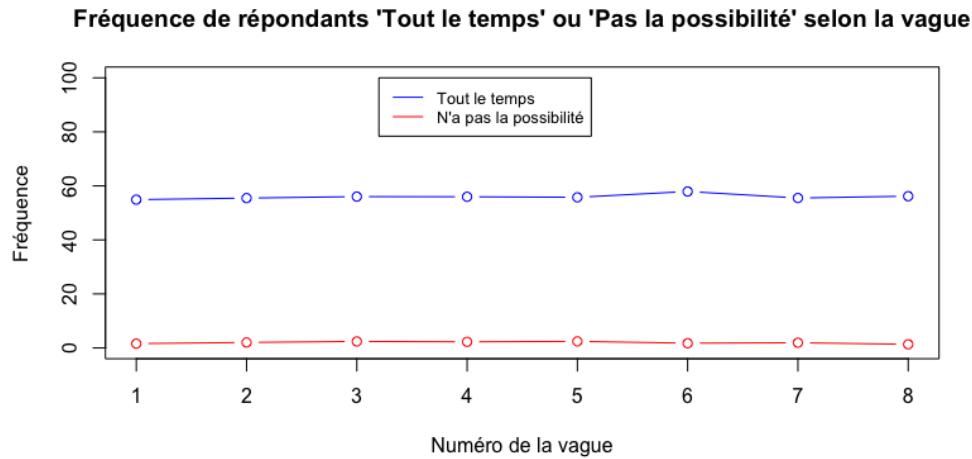


FIGURE 11 – Évolution du geste tousserCoude au fil du temps

A titre de comparaison, nous remarquons que pour un autre geste qui ne nécessite pas de matériel spécifique, et qui a pu être respecté dès le début (**tousserCoude** ici), la temporalité ne modifie pratiquement pas le comportement.

2.1.3 Corrélation entre les gestes barrières

Remarque : Nos variables gestes barrières sont des variables qualitatives. Elles comportent chacune 7 modalités. On ne devrait donc pas pouvoir calculer de corrélation entre elles. Néanmoins, les modalités de réponse sont "croissantes" en intensité de respect : Jamais, Rarement, Souvent, Tout le temps. Dans le cadre de cette analyse descriptive, nous allons donc les considérer comme des variables quantitatives. De plus, le but de cette section est de classifier ceux qui respectent les gestes barrières de ceux qui ne les respectent pas. Nous allons donc retirer les Non Réponse, qui pourraient avoir un poids dans la classification.

Coefficient de corrélation linéaire : Afin de quantifier la relation linéaire entre deux variables quantitatives X et Y, on peut calculer le coefficient de corrélation linéaire.

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\frac{1}{n} \sum_{i,j=1}^n (y_i - \bar{y})(x_j - \bar{x})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} \sqrt{\frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2}}$$

Plus la corrélation entre deux variables est proche de 0, plus ces variables ont un comportement indépendant l'une de l'autre. À l'inverse, plus elle tend en valeur absolue vers 1, plus les variables

sont liées entre elles. Si la corrélation est négative, c'est que l'augmentation de l'une entraîne la diminution de l'autre, tandis qu'une corrélation positive implique que les deux variables augmentent ensemble et diminuent ensemble.

Un corrplot est un graphique qui met en avant les corrélations entre chaque variable. Plus le rond est gros, plus la corrélation est importante. De même, plus la couleur est foncée, plus elle est importante (bleu pour une corrélation positive, rouge pour une corrélation négative)

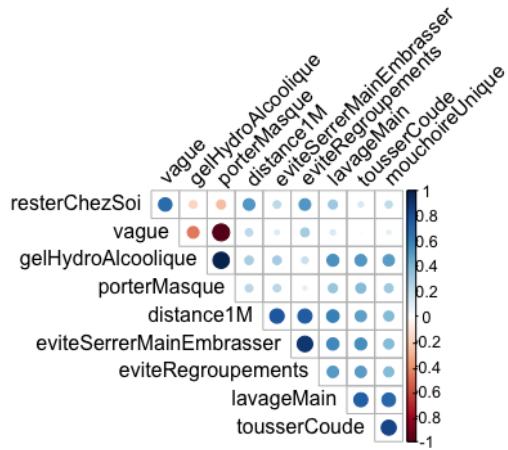


FIGURE 12 – Corrplot des variables gestes barrières et vague

On observe sur ce corrplot que, comme on le pensait, les deux variables les plus corrélées à la temporalité sont le respect du gel hydroalcoolique et du port du masque (La corrélation entre la vague et le port du masque est de -0.422, tandis que celle entre la vague et tousser dans son coude est de -0.003). De plus, vague est négativement corrélée à ces deux variables. C'est logique, car comme vu précédemment ces deux gestes furent de plus en plus respectés au fil du temps. Donc, quand la vague augmente, la valeur de **porterMasque** et **gelHydroAlcoolique** baisse (les modalités sont classées en ordre décroissant de respect).

Cette corrélation n'est tout de fois pas nulle avec les autres variables. Le respect général des gestes barrières a donc évolué au fil du temps.

2.2 Regroupement des gestes

2.2.1 Préparation du regroupement

Certaines variables sont corrélées entre elles et ont un comportement similaire. On a vu que l'utilisation de gel et le port du masque suivaient la même tendance. De même, on peut s'attendre à ce que les gestes « Je respecte la distance de 1M » et « J'évite les regroupements » se ressemblent. Nous allons créer des groupes de gestes barrières, en regroupant ensemble les variables qui ont le même comportement. Pour cela, nous allons utiliser une classification ascendante hiérarchique (abrégée CAH dans la suite). Le but de ce regroupement sera d'alléger le traitement des gestes barrières dans une future régression. Si on les laissait telles quelles, on aurait 9 variables avec chacune 7 modalités, soit $9 \times (7 - 1) = 54$ lignes dans la régression.

Pour la mise en place de cette CAH, nous nous sommes grandement aidés de la thèse de Matthias Studer¹, soutenue à l'Université de Genève. Ce travail a donné lieu au guide d'utilisation de la librairie WeightedCluster², que nous avons utilisée.

Cette CAH va nous permettre de créer des groupes de gestes. Pour cela, on commence par créer une matrice des effectifs de gestes barrière.

Cette matrice, de $\mathcal{M}_{9 \times 7}$, donne l'effectif de chaque modalité pour chaque variable. Ainsi, l'élément $M_{i,j}$ correspond au nombre de personnes ayant répondu la j^{eme} modalité à la question sur le i^{eme} geste barrière.

	Tout le temps	Souvent	Parfois	Rarement	Jamais même si je pourrais le faire	Jamais car je n'ai pas la possibilité de le faire
porterMasque	7284	11105	7080	4598	2483	6805
lavageMain	25890	10965	1803	491	149	57
tousserCoude	22111	9829	3763	2139	744	769
gelHydroAlcoolique	10768	13270	6973	3564	1498	3282
mouchoireUnique	25950	6988	2768	1830	1171	648
distance1M	24781	10517	2012	882	474	689
eviteSerrerMainEmbrasser	33567	3072	863	545	772	536
eviteRegroupements	30406	5560	1282	575	455	1077
resterChezSoi	16549	16952	3129	1105	291	1329

FIGURE 13 – Matrice des effectifs, avec les gestes barrières en ligne et les modalités en colonne

Nous avons centré réduit cette matrice.

	Tout le temps	Souvent	Parfois	Rarement	Jamais même si je pourrais le faire	Jamais car je n'ai pas la possibilité de le faire
porterMasque	0.250	1.569	0.180	-0.677	-1.407	0.085
lavageMain	1.867	0.426	-0.459	-0.586	-0.619	-0.628
tousserCoude	1.866	0.392	-0.336	-0.530	-0.698	-0.695
gelHydroAlcoolique	0.905	1.442	0.089	-0.644	-1.088	-0.704
mouchoireUnique	1.985	0.044	-0.388	-0.484	-0.552	-0.605
distance1M	1.875	0.407	-0.468	-0.584	-0.626	-0.604
eviteSME	2.036	-0.263	-0.429	-0.453	-0.436	-0.454
eviteRegroupements	2.014	-0.084	-0.446	-0.506	-0.516	-0.463
resterChezSoi	1.257	1.307	-0.431	-0.686	-0.789	-0.658

FIGURE 14 – Matrice des effectifs, centrée réduite

-
1. http://garcia-project.eu/wp-content/uploads/2014/10/Studer-2012_these.pdf
 2. <https://cran.r-project.org/web/packages/WeightedCluster/vignettes/WeightedCluster.pdf>

Afin de procéder à la CAH, nous avons besoin d'une matrice de distance. Un grand nombre de distances existent (Manhattan, Canberra, ...). Nous avons choisi la distance euclidienne car c'est celle que nous avons utilisée dans le cadre de notre cours sur l'Analyse de Données. Ici encore, il n'est pas préjudiciable de considérer les gestes barrières comme des variables quantitatives (étant donné que nous prenons la matrice des effectifs, nos variables deviennent quantitatives).

Distance euclidienne : Soit $M \in \mathcal{M}_{n,p}$. En notant x^j les lignes de la matrice, x_i les colonnes et x_i^j l'élément $M_{i,j}$, alors la distance euclidienne entre la variable i et la variable i' est :

$$d(x_i, x_j) = \sqrt{\sum_{j=1}^p (x_i^j - x_j^j)^2}$$

Exemple : Nous allons calculer "à la main" la distance entre les variables *porterMasque* et *gelHydroAlcoolique*.

On note x_{pM} la variable *porterMasque* et x_{gHA} la variable *gelHydroAlcoolique*.

On a

$$x_{pM} = \begin{pmatrix} 0.2501419 \\ 1.56877913 \\ 0.17974096 \\ -0.6768038 \\ -1.4066959 \\ 0.08483773 \end{pmatrix} \text{ et } x_{gHA} = \begin{pmatrix} 0.9045626 \\ 1.44229260 \\ 0.08894108 \\ -0.6437214 \\ -1.0877462 \\ -0.70432881 \end{pmatrix}$$

$$\begin{aligned} \text{Et donc } d(x_{pM}, x_{gHA})^2 &= (0.2501419 - 0.9045626)^2 + (1.56877913 - 1.44229260)^2 \\ &\quad + (0.17974096 - 0.08894108)^2 + (-0.6768038 + 0.6437214)^2 \\ &\quad + (-1.4066959 + 1.0877462)^2 + (0.08483773 + 0.70432881)^2 \\ &= 1.178117 \end{aligned}$$

$$\text{Donc } d(x_{pM}, x_{gHA}) = 1.085411.$$

On obtient une matrice symétrique, comprenant la distance entre chaque geste. La diagonale vaut 0, la distance entre deux mêmes variables étant nulle.

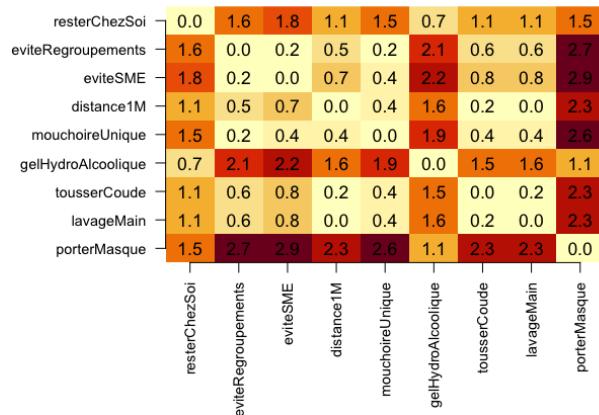


FIGURE 15 – Matrice des distances entre chaque geste

Une fois que nous avons la matrice de distances, l'algorithme fonctionne de la façon suivante :

Algorithme 1 : Algorithme de Clustering

```

Initialisation;
Nombre_de_groupes = 9;
Indice_groupe_1 = 1;
Indice_groupe_2 = 2;
min = distance(Indice_groupe_1, Indice_groupe_2);
Pour i = 1 à Nombre_de_groupes Faire
    Pour j = 1 à Nombre_de_groupes Faire
        Si i != j et distance(i,j) < min Alors
            min = distance(i,j);
            Indice_groupe_1 = i;
            Indice_groupe_2 = j;
        Fin Si
    Fin Pour
Fin Pour
Ajouter Groupe formé de i et j;
Supprimer Groupe i;
Supprimer Groupe j;
```

A la fin de cet algorithme, on a donc un unique groupe regroupant toutes les variables.

Le schéma agglomératif, c'est-à-dire la succession des regroupements effectués, représente la procédure de clustering sous la forme d'un arbre que l'on appelle dendrogramme. Une fois le schéma agglomératif construit, on sélectionne le nombre de groupes.

On peut représenter le regroupement ainsi créé.

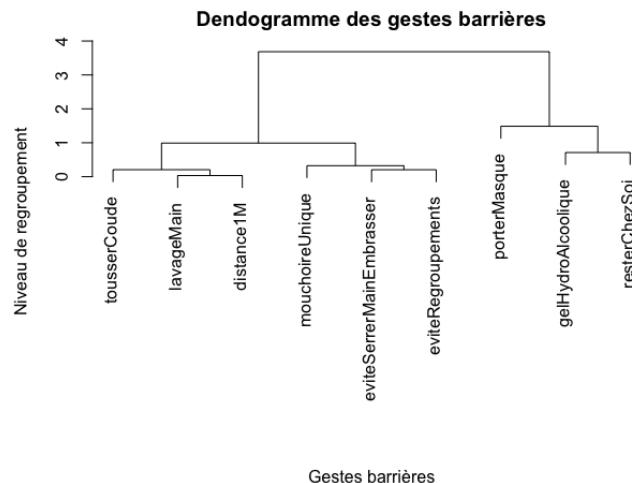


FIGURE 16 – Dendrogramme représentant la partition de nos gestes barrières

Ce dendrogramme se lit de bas en haut. Pour analyser le parcours de regroupement de l'algorithme, on part du bas et on remonte. Ainsi, à la première étape, l'algorithme a regroupé *lavageMain* et *distance1M*. On le savait déjà grâce à la matrice de distances : c'est pour ces deux variables que la distance était la plus faible.

La forme de notre dendrogramme corrobore l'intuition que nous avions eu à l'aide de l'analyse naïve de ces variables, avec les statistiques descriptives : le respect du port du masque et de l'utilisation de gel hydroalcoolique ne réagissent pas de la même façon que les autres gestes (sauf *resterChezSoi*). Leur respect est conditionné à l'accès à des biens matériels. A l'inverse, tousser dans son coude est accessible à tous. Notre échantillon ayant été sondé au tout début de l'épidémie, tout le monde n'avait pas accès à ces ressources. Le comportement face à ces deux gestes barrières était donc différent du respect des autres, plus accessibles. Leur ressemblance avec la variable *resterChezSoi* s'explique de la même façon : quand l'accès à ces ressources fut plus simple, nous avons été déconfinés.

2.2.2 Analyse de la qualité du regroupement

Maintenant que nous avons défini des possibilités de regroupement, il reste à déterminer en combien de groupes on sépare nos 9 gestes barrières : si l'on garde trop de groupes, cela ne va pas nous être utile pour notre étude, et si l'on en retire trop on risque de perdre trop d'information. Selon la documentation du package, 9 critères sont présents pour mesurer la qualité de la partition.

Les colonnes de ce tableau représentent :

- Nom : Nom du critère
- Abrv. : Abrévation utilisée dans le package
- Min/Max : Est-ce qu'une bonne partition maximise ou minimise ce critère ?
- Interprétation : Interprétation de la valeur

Nom	Abrv.	Étendue	Min/Max	Interprétation
Point Biserial Correlation	PBC	[−1; 1]	Max	Mesure de la capacité du clustering à reproduire les distances.
Hubert's Gamma	HG	[−1; 1]	Max	Mesure de la capacité du clustering à reproduire les distances (ordre de grandeur).
Hubert's Somers D	HGSD	[−1; 1]	Max	Mesure de la capacité du clustering à reproduire les distances (ordre de grandeur) avec prise en compte des égalités sur les distances.
Hubert's C	HC	[0; 1]	Min	Écart entre la partition obtenue et la meilleure partition qu'il serait théoriquement possible d'obtenir avec ce nombre de groupes et ces distances.
Average Silhouette Width	ASW	[−1; 1]	Max	Cohérence des assignations. Une cohérence élevée indique des distances intergroupes élevées et une forte homogénéité intragroupe.
Calinski-Harabasz index	CH	[0; +∞[Max	Pseudo F calculé à partir des distances.
Calinski-Harabasz index	CHsq	[0; +∞[Max	Idem que précédent, mais en utilisant les distances <i>au carré</i> .
Pseudo R^2	R2	[0; 1]	Max	Part de la dispersion expliquée par la solution de clustering (uniquement pour comparer des partitions avec nombre de groupes identiques).
Pseudo R^2	R2sq	[0; 1]	Max	Idem que précédent, mais en utilisant les distances <i>au carré</i> .

FIGURE 17 – Mesure de la qualité des regroupements

Comme conseillé dans la thèse de Monsieur Studer, nous allons utiliser 4 critères : HC, HG, PBC et ASW.

Point Biserial Correlation : Cette mesure prend en paramètre 2 matrices X et Y. La matrice X est notre matrice de distances. La matrice Y est une matrice dichotomique construite comme telle :

$$\forall (i, j) \in \llbracket 1, 9 \rrbracket, Y_{i,j} = \begin{cases} 0 & \text{si } i \text{ et } j \text{ sont dans le même groupe} \\ 1 & \text{sinon} \end{cases}$$

On divise la matrice X en deux groupes : le groupe des distances ayant "0" dans Y, et celui des distances ayant "1" dans Y. On note r_{pb} le coefficient.

$$r_{pb} = \frac{M_1 - M_0}{s_n} \sqrt{\frac{n_1 n_0}{n^2}} \text{ avec } s_n = \sqrt{\frac{1}{n} \sum_{i,j=1}^n (X_{i,j} - \bar{X})^2}$$

$$M_0 = \frac{1}{n_0} \sum_{i,j=1}^n X_{i,j} \mathbb{1}_{Y_{i,j}=0}$$

$$M_1 = \frac{1}{n_1} \sum_{i,j=1}^n X_{i,j} \mathbb{1}_{Y_{i,j}=1}$$

n_0 et n_1 le nombre de 0 (resp. de 1) dans Y

La mesure **Hubert's Gamma** fonctionne de la même façon, avec une statistique différente.

La mesure HC, ou **Hubert's C**, compare la partition avec la meilleure partition que l'on aurait pu obtenir avec cette matrice de distance. On cherche à minimiser ce critère.

Enfin, la mesure **Average Silhouette Width** est celle qui nous intéresse le plus. Elle se base sur la cohérence de l'assignation d'une observation à un groupe donné en mettant en parallèle la distance moyenne pondérée d'une observation aux autres membres de son groupe et la distance moyenne pondérée au groupe le plus proche. Une valeur est calculée pour chaque observation, mais on regarde plutôt la silhouette moyenne. Si celle-ci est faible, cela signifie que les groupes ne sont pas clairement séparés les uns des autres ou que l'homogénéité des groupes est faible.

Les deux mathématiciens à l'origine de cette mesure ont proposé un tableau d'évaluation de la mesure, que nous allons utiliser.

ASW	Interprétation proposée
0.71-1.00	Structure forte identifiée.
0.51-0.70	Structure raisonnable identifiée.
0.26-0.50	La structure est faible et pourrait être artificielle. Essayer d'autres algorithmes.
≤ 0.25	Aucune structure.

FIGURE 18 – Ordres de grandeur pour interpréter la mesure ASW

Nous allons donc chercher à maximiser cette mesure. En ne changeant rien et en gardant 9 groupes, elle vaudra 1. Elle aura ensuite tendance à baisser avec le nombre de groupes gardés. On cherchera à avoir une partition dont la mesure ASW est d'au moins 0.50.

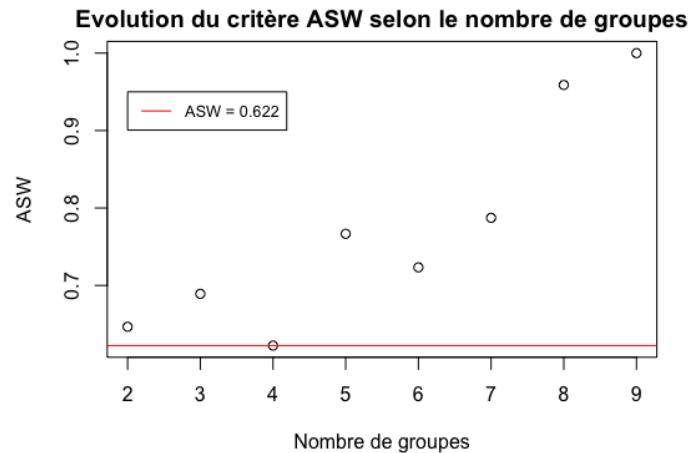


FIGURE 19 – Valeur de la mesure ASW selon le nombre de groupes retenus

On remarque que cette valeur n'est jamais inférieure à 0,5. Quelque soit le regroupement, il est donc de qualité. A partir de 6 groupes, le critère devient croissant, jusqu'à atteindre 1.

Nous allons maintenant afficher les 4 mesures pour déterminer combien de groupes nous allons garder dans notre regroupement.

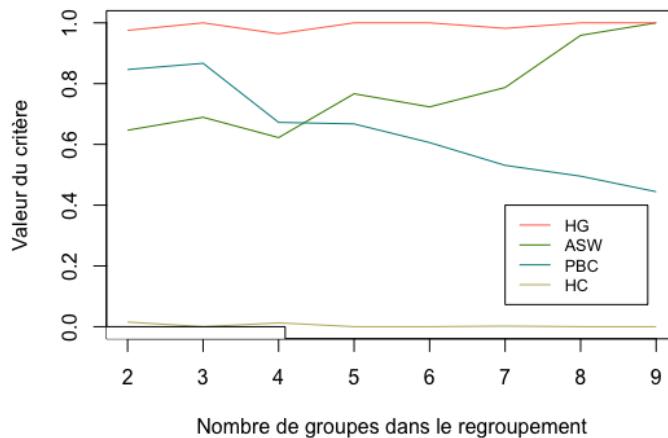


FIGURE 20 – Évolution des 4 coefficients de mesure de la qualité de la partition selon le nombre de groupes retenus

Pour rappel, nous souhaitons maximiser HG, ASW et PBC, et minimiser HC. On remarque que pour un regroupement en 3 groupes, nous avons un maximum global pour HG et PBC, et un maximum local pour ASW. Il s'agit également d'un minimum local pour HC. De plus, l'ASW pour le regroupement en 3 groupes est de 0.69, ce qui correspond à une structure raisonnable selon le tableau d'ordres de grandeur du coefficient ASW (et presque à une structure forte).

Nous allons donc garder ce regroupement en 3 groupes.

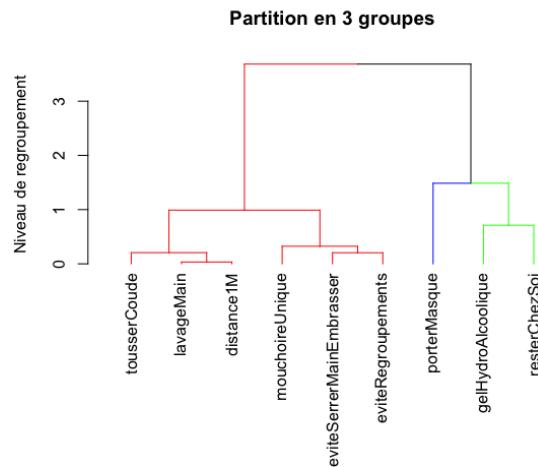


FIGURE 21 – Regroupement en 3 groupes des 9 gestes barrières

A présent que nous avons déterminé quelle partition nous allons utiliser, nous allons traiter numériquement ce regroupement.

2.3 Regroupement numérique des gestes dans les 3 groupes et traitement des données

Nous allons maintenant créer nos 3 variables de groupes, que nous utiliserons dans la section suivante à la place des 9 gestes barrières. Chacun de ces groupes sera une variable qualitative, avec un nombre de modalités proportionnel au nombre de gestes barrières regroupés dans ce groupe.

Pour cela, nous avons adopté la stratégie suivante :

1. Pour chaque geste barrière, on met les valeurs inférieures ou égales à 2 à 1, et les autres à 0
2. On additionne les valeurs des gestes présents dans ce groupe

Nous avons choisi de couper à 2 pour regrouper d'un côté les "Tout le temps" et "Souvent" et de l'autre côté le reste.

	Niveau 1	Niveau 2	Niveau 3	Niveau 4	Niveau 5	Niveau 6	Niveau 7
groupe1	2028	17115	20212	NA	NA	NA	NA
groupe2	20966	18389	NA	NA	NA	NA	NA
groupe3	282	321	547	1511	3797	8882	24015

FIGURE 22 – Tables du respect des 3 groupes de gestes barrières

Dans le premier groupe, nous avions 2 gestes barrières (*gelHydroAlcoolique* et *resterChezSoi*), nous avons donc 3 modalités :

- Niveau 1 si l'individu ne respecte ni l'utilisation de gel, ni le fait de rester chez soi.
- Niveau 2 si l'individu respecte l'un mais pas l'autre.
- Niveau 3 s'il respecte les 2.

Dans le second groupe, il n'y a que le port du masque. La variable finale correspond donc à la dichotomisation de la variable *porterMasque*.

Enfin, dans le dernier groupe sont regroupés tous les autres gestes. Il y en a 6, donc cette variable possède 7 modalités (Niveau 1 si l'individu ne respecte aucun des 6 gestes, Niveau 7 s'il les respecte tous).

3 Les travailleurs hors de leur domicile

3.1 Qui sont-ils ?

3.1.1 La variable HorsDomicile

Afin d'observer l'exposition des travailleurs hors de leur domicile, nous devons créer une variable dichotomique **HorsDomicile** permettant de les caractériser. Pour cela, on s'intéresse à la question suivante :

Question C6 : Quelle est votre situation professionnelle actuellement ? Cochez toutes les réponses qui correspondent à votre situation (C6_0 : Première réponse cochée, ..., C6_3 :4ème réponse cochée)

- 1 : Je travaille hors de mon domicile
- 2 : Je travaille à domicile (comme d'habitude)
- 3 : Je travaille à distance (télétravail)
- 4 : Je suis en arrêt maladie (exclusif avec C6_5)
- 5 : Je suis en chômage partiel
- 6 : Je suis en congé (exclusif avec C6_3)
- 7 : Non réponse
- 8 : Je travaille dans le secteur de la santé

Notre variable **HorsDomicile** est donc une variable binaire. Elle prend la valeur 1 lorsque la personne travaille à l'extérieur (modalité 1), et 0 si elle travaille à domicile (toute autre modalité).

Remarque : Cette question n'a été posée qu'aux personnes qui travaillent, nous décidons donc de retirer les inactifs et les retraités pour l'instant. Ils ne seront donc pas pris en compte dans les prochaines analyses et régressions.

	Travaille chez soi	Travaille à l'extérieur
Nombre	15456	7870
Fréquence	66.261 %	33.739 %

FIGURE 23 – Nombre et fréquence des personnes travaillant Hors Domicile dans notre échantillon

Nous avons donc, dans nos 40001 observations, à peu près 2/3 des personnes qui ont travaillé de leur domicile, et 1/3 qui travaillaient en dehors. Cette variable va donc pouvoir être exploitable. Elle ne l'aurait pas été si nous avions eu trop peu d'invidus travaillant dehors.

L'évolution de la proportion des travailleurs hors de leur domicile au fil du temps est très intéressante.

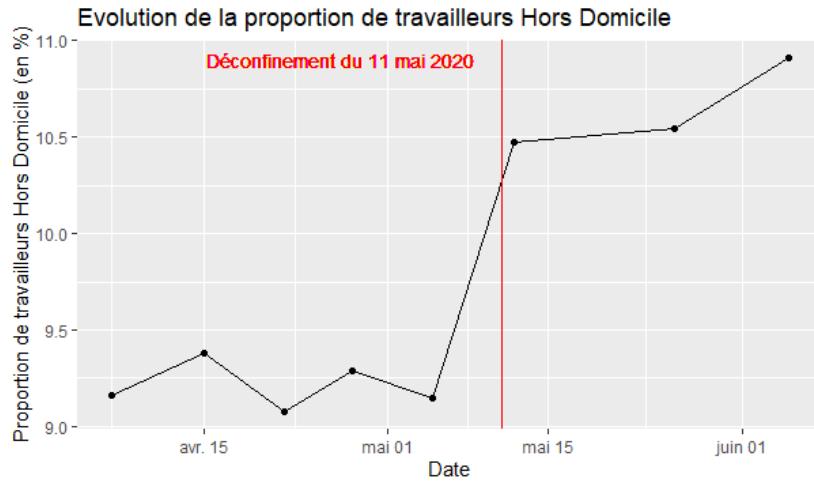


FIGURE 24 – Évolution de la proportion des travailleurs Hors Domicile en fonction du temps

En effet, le graphique ci-dessus montre clairement l'impact qu'a eu la fin du confinement sur le travail hors domicile. Comme on pouvait s'y attendre, la réouverture de nombreux magasins ainsi que la reprise partielle de la vie économique de notre pays le 11 mai 2020 a fait bondir la proportion d'individus travaillant hors de leur domicile. Toutefois, même si l'augmentation est significative, cette proportion reste relativement basse, ne dépassant pas les 11%, ce qui est bien inférieure à une période normale.

Remarque : A partir de cette même question, nous créons également une variable *ProfSante* que nous utiliserons plus tard. Cette variable prend la valeur 1 si l'individu travaille dans le secteur de la santé et 0 sinon.

De nombreux facteurs peuvent expliquer le fait de travailler dehors durant le confinement ou le début du déconfinement : le type de métier, l'âge, etc. Nous allons effectuer une régression logistique (qui est un modèle linéaire généralisé particulier) afin de déterminer les facteurs explicatifs du fait de travailler dehors ou non.

3.1.2 Régression logistique

Modèle linéaire généralisé : Le modèle linéaire généralisé est la donnée d'une loi de probabilité pour les Y_i et d'une fonction g appelée fonction de lien telle que

$$g(\mathbb{E}[Y_i]) = x_i \beta$$

Dans notre cas, Y correspond au vecteur de \mathbb{R}^{40001} contenant nos 40001 observations de la variable HorsDomicile. On a donc $y_i \in \{0, 1\} \forall i \in [1, 40001]$.

Nous allons naturellement utiliser la loi de Bernoulli pour modéliser Y_i :

$$(Y_i | X_i = x_i) \sim \mathcal{B}(p_i) \text{ avec } \mathbb{E}[Y_i | X = x] = p_i \in [0, 1]$$

Soit $g : [0, 1] \rightarrow \mathbb{R}$, bijective. On pose :

$$g(\mathbb{E}[Y_i | X_i = x_i]) = x_i \beta \Leftrightarrow \mathbb{E}[Y_i | X_i = x_i] = g^{-1}(x_i)$$

X correspond à la matrice contenant les variables explicatives.

Nous avons p variables, et on note $(p_i)_{1 \leq i \leq n}$ le nombre de modalités de chaque variable.

On note $m := \sum_{i=1}^p p_i$

Notre matrice X sera une matrice de $\mathcal{M}_{40001 \times m}$.

Elle est construite de la façon suivante :

- Chaque ligne de X correspond à une observation.
- Chaque colonne correspond à une modalité (les p_1 premières colonnes correspondant aux modalités de la première variable, etc.).
- $X_{i,j}$ vaut 1 si l'individu i a répondu la modalité j à la variable en question, et 0 sinon.

On ajoute une colonne de 1 au début de la matrice pour introduire un intercept.

Au final, X sera donc une matrice de $\mathcal{M}_{40001 \times (m+1)}$. Enfin, $\beta = (\beta_0, \dots, \beta_m)^t \in \mathbb{R}^{(m+1)}$ est le vecteur des coefficients que nous allons déterminer à l'aide de la régression.

Fonction de lien : Toute bijection de $[0, 1]$ dans \mathbb{R} peut être choisie comme fonction de lien. Néanmoins, nous choisissons souvent ce que l'on appelle la **fonction de lien canonique**.

Famille exponentielle naturelle : Soit f_Y la densité de la variable Y . f_Y appartient à la famille exponentielle naturelle si elle s'écrit sous la forme :

$$f_Y(y) = \exp\left(\frac{1}{\gamma(\phi)}(y\theta - b(\theta)) + c(y, \phi)\right)$$

La loi de Bernoulli (à densité par rapport à la mesure de comptage de $\{0, 1\}$) fait partie de la famille exponentielle naturelle. En effet,

$$\begin{aligned} f_Y(y) &= p^y(1-p)^{1-y} \\ &= \exp(y\ln(p) + (1-y)\ln(1-p)) \\ &= \exp\left(y\ln\left(\frac{p}{1-p}\right) + \ln(1-p)\right) \end{aligned}$$

On reconnaît la forme de la famille exponentielle avec

$$\begin{aligned} \gamma : x \rightarrow 1, \quad \theta &= \ln\left(\frac{p}{1-p}\right), \quad b(\theta) = -\ln(1-p) \text{ et } c(y, \phi) = 0 \\ \Leftrightarrow p &= \frac{e^\theta}{1+e^\theta}, \quad b(\theta) = -\ln\left(1 - \frac{e^\theta}{1+e^\theta}\right) = \ln\left(1 + e^\theta\right) \end{aligned}$$

Fonction de lien canonique : Il s'agit de la fonction de lien qui transforme l'espérance $\mathbb{E}[Y_i]$ en θ :

$$g = (b')^{-1}$$

Dans notre cas, nous avons $b'(\theta) = \frac{e^\theta}{1+e^\theta}$ et donc $g(x) = \ln(\frac{x}{1-x})$. Le nom de ce lien est logit. D'où le nom de régression logistique.

Régression : La régression va, à l'aide des données, estimer le paramètre β qui minimise la différence entre $g(\mathbb{E}[Y_i])$ et $x_i^t \beta$.

En particulier, pour chaque β_k , nous allons effectuer le test

$$\mathbf{H}_0 : \beta_k = 0 \text{ vs } \mathbf{H}_1 : \beta_k \neq 0$$

pour déterminer si la modalité en question est significative ou non dans notre modèle. Nous nous placerons à un niveau de 95%, et rejeterons donc les p-valeur inférieures à 0,05.

Statistique de test : Nous prenons la statistique

$$T(Y_n) = \frac{\hat{\beta}_{k,n}}{\sqrt{\hat{s}_k^2}} \text{ où } \hat{s}_k^2 = (I_{n,\hat{\phi}_n}(\hat{\beta}_n)^{-1})_{k,k}$$

avec $I_{n,\hat{\phi}_n}(\beta^*) = \frac{1}{\gamma(\phi^*)} X'W(\beta^*)X$ où $W(\beta)$ est une matrice diagonale n x n telle que

$$W_{ii}(\beta) = b''(x_i^t \beta) \quad \forall i = 1, \dots, n$$

$I_{n,\hat{\phi}_n}(\beta^*)$ s'appelle l'information de Fisher.

Cette statistique est de loi asymptotique normale centrée réduite sous \mathbf{H}_0 . Par conséquent, on rejette \mathbf{H}_0 si $|T| > q_{1-\frac{\alpha}{2}}$ où $q_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1 - \frac{\alpha}{2}$ d'une loi gaussienne centrée réduite.

Odds ratio : Afin d'analyser les résultats de notre régression, nous allons utiliser les odds ratios. Pour les β_k pour lesquels on rejette \mathbf{H}_0 , nous notons OR_k leurs odds ratio respectifs. Nous avons $OR_k = \frac{p_k}{1-p_k} = \exp(\beta_k)$.

L'intuition des odds ratio est la suivante : pour une modalité, on calcule le rapport des chances entre cette modalité et la modalité que l'on a prise en référence.

Prenons l'exemple de notre régression, avec comme variable d'intérêt HorsDomicile. Pour la variable PCSI, prenons comme modalité de référence Cadres Sups. Si l'odds ratio associé à la modalité Agriculteurs est de 1.5, c'est que d'après notre modèle un agriculteur aura 50% de chances en plus de travailler en extérieur qu'un cadre supérieur. Au contraire, si l'odds ratio est de 0.7, il aura 30% de chances en moins de travailler dehors.

3.1.3 Profil des travailleurs à l'extérieur

A présent nous allons déterminer qui sont les individus les plus susceptibles de travailler dehors, et surtout à quelle classe sociale ils appartiennent.

Pour cela nous allons nous intéresser aux variables **Sexe**, **Age**, **PCSI** et **Diplome**, qui découlent des questions suivantes :

- SEXE** : Genre du participant
- 1 : Homme
 - 2 : Femme

On crée donc une variable dichotomique **Sexe** prenant deux modalités "Homme" et "Femme".

- AGE** : Âge du participant
- | | |
|-------------------|-----------------------|
| — 1 : 18 à 24 ans | 6 : 45 à 49 ans |
| — 2 : 25 à 29 ans | — 7 : 50 à 54 ans |
| — 3 : 30 à 34 ans | — 8 : 55 à 59 ans |
| — 4 : 35 à 39 ans | — 9 : 60 à 64 ans |
| — 5 : 40 à 44 ans | — 10 : 65 ans et plus |

Ici, contrairement à nos attentes, l'âge du participant n'est pas une variable quantitative mais une variable qualitative avec 10 modalités. Afin de faciliter la lecture de nos travaux, nous avons décidé de réduire le nombre de modalités de la variable ***Age*** à 5 en regroupant les seuils 2 à 2 de la façon suivante :

- 18 à 29 ans
- 25 à 29 ans
- 30 à 39 ans
- 40 à 49 ans
- 50 à 59 ans
- 60 ans et plus

PCSI : Catégorie Socio-Professionnelle du participant

- 1 : Agriculteurs exploitants
- 2 : Professions Indépendantes
- 3 : Cadres Supérieurs
- 4 : Professions intermédiaires
- 5 : Employés
- 6 : Ouvriers
- 7 : Retraités
- 8 : Inactifs

Nous gardons cette variable telle quelle sous le nom ***PCSI*** (Profession ou Catégorie Socio-professionnelle de l'Individu).

Question C10 : Quel est le diplôme le plus élevé que vous ayez obtenu ?

- 1 : Sans diplôme
- 2 : Certificat de formation générale (CFG), Certificat d'études primaires, Diplôme national du brevet (BEPC ou Brevet des Collèges)
- 3 : Certificat d'aptitude professionnelle (CAP), Brevet d'études professionnelles (BEP)
- 4 : Baccalauréat ou diplôme équivalent
- 5 : Bac +1 ou 2
- 6 : Bac +3 ou plus
- 7 : Autre, précisez
- 8 : Non Réponse

Toujours dans l'objectif de faciliter la compréhension, nous avons décidé de réduire le nombre de modalités à 4 en regroupant certaines d'entre elles de la manière suivante dans la variable ***Diplome*** (et en retirant les Non Réponse et les Autre) :

- Sans Diplôme
- - de BAC
- BAC à BAC+2
- BAC+3 ou +

Afin d'analyser le profil des personnes travaillant à l'extérieur, nous allons effectuer une régression logistique sur la variable ***HorsDomicile***, avec comme variables explicatives ***PSCI***, ***Sexe***, ***Age***, ***Diplome***, et ***ProfSante***. Nous avons retiré les non réponses de cette régression.

Interprétation : Afin de travailler avec un niveau $\alpha = 5\%$, nous rejetterons H_0 si la p-value est supérieure à 0.05.

Nous avons retiré de la régression les inactifs et les retraités, qui ne travaillent pas et donc à fortiori ne travaillent pas à l'extérieur. La valeur de référence pour ***PSCI*** est Cadres sups. Ce choix a été fait car c'est la catégorie socio-professionnelle qui correspond au rang social le plus élevé. L'analyse des odds ratios est sans équivoque : tous les odds ratios sont supérieurs à 1 : les cadres supérieurs sont donc ceux ayant le moins de probabilité de travailler en dehors du domicile. Un agriculteur a en effet presque 3x plus de chances de travailler à l'extérieur qu'un cadre, un ouvrier 2,5x. Cela paraît logique : les cadres supérieurs sont plus susceptibles de pouvoir télétravailler. Au contraire, les ouvriers ou les agriculteurs font un métier qui nécessite de travailler hors de son domicile.

Concernant le sexe, il semble que les hommes soient 42% plus susceptibles de travailler en dehors de leur domicile. Nous avons retiré de cette analyse les hommes/femmes au foyer, ainsi que les inactifs et retraités. Une hypothèse sur cette différence est que, les écoles ayant été fermées durant le confinement, les parents ont du s'organiser pour garder les enfants. Dans notre société, il y a plus de chances que la femme ait rempli ce rôle.

Nous remarquons également que les personnes âgées ont plus de chances de travailler en dehors de leur domicile (sauf pour les 60 ans et plus). Cela peut s'expliquer par le fait que les jeunes générations ont plus de facilités avec les technologies et ont pu mieux s'organiser pour télétravailler. Nous pouvons également noter que la tranche 30-39 n'est pas significative : les moins de 40 ans ont réagi à peu près de la même façon.

Concernant le diplôme, le fait d'avoir un diplôme de niveau BAC+3 est très peu significatif. Ca l'est un peu plus pour les individus ayant un diplôme inférieur à BAC+2. Nous supposons que c'est parce que parmi les métiers visés par des formations type CAP et BEP, beaucoup sont des métiers manuels qui nécessitent de travailler hors de son domicile. Néanmoins, au vu des p-valeurs, cette variable n'est pas très intéressante.

Enfin, les professionnels de la santé sont presque 2x plus exposés que ceux exerçant d'autres professions. On s'en doutait, étant donné qu'ils étaient en première ligne face au virus.

Ainsi, les travailleurs à l'extérieur sont principalement des professionnels de la santé, ou bien ont un profil d'homme âgé, non cadre supérieur.

Characteristic	OR [†]	95% CI [†]	p-value
PCSI			
Cadres sups	—	—	
Agriculteurs	2.68	1.98, 3.62	<0.001
Employés	1.73	1.56, 1.90	<0.001
Indépendants	1.24	1.07, 1.44	0.003
Ouvriers	2.23	1.98, 2.50	<0.001
Prof. intermédiaires	1.99	1.81, 2.18	<0.001
Sexe			
Femme	—	—	
Homme	1.42	1.34, 1.51	<0.001
Age			
18-29	—	—	
30-39	1.08	0.99, 1.18	0.10
40-49	1.25	1.14, 1.37	<0.001
50-59	1.16	1.05, 1.27	0.003
60+	0.65	0.56, 0.74	<0.001
Diplome			
Sans Diplome	—	—	
- de BAC	1.30	1.06, 1.62	0.014
BAC à BAC+2	1.30	1.06, 1.60	0.014
BAC+3 ou +	1.06	0.85, 1.31	0.6
ProfSante			
Non	—	—	
Oui	1.93	1.65, 2.26	<0.001

[†] OR = Odds Ratio, CI = Confidence Interval

Afin de mieux différencier les classes sociales, nous aurions aimé intégrer le revenu dans cette régression. La question portant sur le revenu ne fait malheureusement pas partie du questionnaire cœur. Celle-ci n'apparaît que dans le module complémentaire de la vague 8. Nous ne disposons donc que de 5000 observations.

Question Q45 : Quel est le revenu total mensuel de votre ménage ? (C'est-à-dire l'ensemble de vos revenus, transferts, pensions et allocations, nets d'impôts et de cotisations sociales)

- 1 : Moins de 500 €/mois
- 2 : De 500 à 800€/mois
- 3 : De 800 à 1000€/mois
- 4 : De 1000 à 1500 €/mois
- 5 : De 1500 à 2000 €/mois
- 6 : De 2000 à 3000 €/mois
- 7 : Plus de 3000 €/mois
- 8 : Ne souhaite pas répondre

De plus, nous trouvons cette question maladroitement posée : nous aurions préféré l'avoir en qualitatif, et les tranches nous paraissent assez aléatoires. En effet, 3 tranches sur 7 couvrent un revenu inférieur au SMIC. Nous supposons que les personnes ayant un revenu mensuel inférieur au SMIC sont inactives ou bien retraitées. Elles ne rentreraient donc pas dans notre étude sur le travail à l'extérieur. La dernière tranche recouvre tous ceux gagnant plus de 3000€. On ne peut donc pas faire de distinctions entre les personnes aisées.

Comme la question couvre le ménage total, cela revient à une tranche plus haute pour un salaire de 1500€ pour chaque conjoint (dans le cas des ménages composés de plus d'une personne, ce qui représente la majorité de nos observations). En 2017, d'après l'INSEE, le revenu médian en France était de 1789€ nets mensuels, soit plus de 3000€ pour un ménage. L'analyse de la pertinence de ces tranches pourrait faire l'étude d'une étude complète, mais dans le cadre de notre mémoire cette question n'est pas très informative. Nous n'utiliserons donc pas cette variable.

3.1.4 Le respect des gestes barrières parmi ceux travaillant à l'extérieur

Maintenant que nous avons déterminé le profil de ceux qui travaillent hors de leur domicile, nous pouvons nous demander si ces derniers respectent plus les gestes barrières. Nous allons effectuer une nouvelle régression logistique sur ***HorsDomicile***. Cette fois-ci nous utiliserons comme variables explicatives les 3 catégories de gestes que nous avons créées dans le chapitre 2.

Interprétation : Le premier constat que nous pouvons faire est que toutes les p-values sont très faibles. Tous les groupes de gestes sont donc significatifs dans l'explication du travail hors domicile.

Le groupe 1 regroupe les gestes ***resterChezSoi*** et ***gel-Hydroalcoolique***. Nous rappelons que le score de ce groupe correspond au nombre de gestes que l'on respecte parmi ces deux. Il est évident que les personnes respectant ce groupe ne travaillent pas en dehors de leur domicile. En effet, la variable ***resterChezSoi*** est dedans, et quelqu'un qui reste chez soi est par définition quelqu'un qui ne travaille pas à l'extérieur. Ainsi, quelqu'un qui met du gel hydroalcoolique et qui reste chez lui a 85% de chances de plus de ne pas travailler hors de son domicile que quelqu'un qui ne respecte aucun de ces deux gestes.

Le groupe 2, qui correspond seulement au port du masque, est très influant également. Quelqu'un qui porte le masque a en effet 2,2x plus de chances de travailler dehors que quelqu'un qui ne le porte pas. C'est logique, car ceux travaillant hors de leur domicile doivent porter le masque sur leur lieu de travail. Au contraire, quelqu'un qui travaille de chez soi n'a pas d'utilité à porter le masque. Comme on a vu dans le chapitre 2, les masques étaient peu disponibles à la vente aux particuliers à cette période. Cela explique que ceux qui restaient chez eux n'en portaient pas.

Enfin, le groupe 3 regroupe tous les autres gestes (tels que tousser dans son coude, éviter les rassemblements, etc.). Pour ce groupe, nous remarquons que plus quelqu'un respecte de gestes barrières, plus il est probable qu'il travaille à l'extérieur. Les odds ratios sont tous supérieurs à 1, et tous croissants (sauf pour la modalité 6). Cela peut s'expliquer par le fait que travailler dehors mène à une plus grande méfiance vis à vis du virus. On se sent plus menacé, et on peut également craindre de ramener le virus dans son foyer.

Characteristic	OR [†]	95% CI [†]	p-value
groupe1			
0	—	—	
1	0.23	0.20, 0.26	<0.001
2	0.15	0.13, 0.17	<0.001
groupe2			
0	—	—	
1	2.19	2.07, 2.33	<0.001
groupe3			
0	—	—	
1	1.59	1.04, 2.44	0.031
2	2.32	1.57, 3.45	<0.001
3	2.27	1.60, 3.24	<0.001
4	2.61	1.87, 3.66	<0.001
5	2.81	2.02, 3.92	<0.001
6	2.55	1.84, 3.55	<0.001

[†] OR = Odds Ratio, CI = Confidence Interval

3.2 Exposition à l'infection à la Covid-19

3.2.1 La variable Infecte

Afin de déterminer si le fait de travailler en extérieur expose d'avantage à la Covid-19, et de déterminer les profils les plus exposés, nous cherchons maintenant à créer une variable dichotomique **Infecte**. **Infecte** prendra la valeur 1 si l'individu a déjà été infecté par la Covid-19 et 0 sinon. Pour la créer, nous nous intéressons à deux questions de notre questionnaire qui sont les suivantes :

Question A1 : Avez-vous ou pensez-vous avoir été infecté par le coronavirus ?

- 1 : Oui
- 2 : Non

Question A2 : Vous avez ou pensez avoir été infecté par le coronavirus car...

- 1 : J'ai été diagnostiqué positif par un test de dépistage
- 2 : J'ai ou j'ai eu les symptômes caractéristiques de la maladie et j'ai contacté un médecin
- 3 : J'ai ou j'ai eu les symptômes caractéristiques de la maladie mais je n'ai pas contacté de médecin
- 4 : Je n'ai pas eu de symptômes caractéristiques de la maladie mais j'ai été en contact proche avec au moins une personne diagnostiquée positive par un test de dépistage
- 5 : Je n'ai pas eu de symptômes caractéristiques de la maladie mais j'ai été en contact proche avec au moins une personne qui pense avoir été infectée par le coronavirus
- 6 : Je ne connais pas assez bien les symptômes pour répondre

La question A2 n'ayant été posée qu'aux individus ayant répondu favorablement à la question A1, nous avons décidé d'attribuer la valeur 0 à tous ceux qui ont répondu Non à la première. Nous avons alors du choisir judicieusement les modalités de la question A2 que nous allions attribuer à notre variable **Infecte**. L'accès aux tests s'étant démocratisé en France seulement à la mi-mai 2020, nous avons fait le choix de considérer comme ayant été infectés les individus qui ont répondu la première ou la deuxième modalité à la question A2. En effet, il nous paraissait bien trop laxiste de prendre en compte des "auto-diagnostic" dans une période durant laquelle nous faisions tous parfois preuve de paranoïa.

Nous obtenons l'effectif suivant pour la variable **Infecte** :

	N'a pas contracté le virus	A contracté le virus
Nombre	35173	1538
Fréquence	95.811 %	4.189 %

FIGURE 25 – Nombre et fréquence des infectés dans notre échantillon

3.2.2 Quels sont les facteurs augmentant les risques d'infection ?

La variable binaire ***Infecte*** créée, nous pouvons maintenant essayer de l'expliquer à l'aide d'une régression logistique afin de déterminer la population la plus touchée et les facteurs augmentant la contamination. Nous prendrons comme variables explicatives les mêmes que pour la régression sur ***HorsDomicile***, avec en plus cette dernière.

Interprétation : Nous observons tout d'abord que toutes les modalités de notre variable ***Age*** sont significatives, ce qui nous informe que les tranches d'âge ne sont pas touchées de la même manière par la Covid-19. Nous pouvons également observer une décroissance des odds ratios pour cette variable. Ainsi, plus l'individu est âgé moins il a de chances d'avoir été infecté. Les plus touchés étant sans surprise la tranche des 18-29 ans (une personne de 60 ans et plus a plus de 2x moins de chances d'être infecté que quelqu'un entre 18 et 29 ans). Cela peut s'expliquer par la crainte plus élevée des personnes âgées vis-à-vis du virus et de ses formes graves.

Parmi les différentes catégories socio-professionnelles, ce sont les agriculteurs qui sont les plus infectés. Nous n'avons pas réellement d'interprétation pour ce phénomène. Le faible nombre d'agriculteurs dans la base de données rend l'interprétation de cette donnée compliquée. Sinon, ce sont les cadres supérieurs qui sont les plus touchés par le virus. Cela rejoint le fait que les cadres travaillent plus à l'extérieur. Les ouvriers, les employés et les étudiants sont quant à eux plus épargnés.

Les professionnels de santé sont bien plus touchés, en moyenne 56% plus que les autres. Cela s'explique facilement par leur contact journalier avec des malades.

Sur le sexe, il n'y a pas de grande différence d'infection. La p-value est supérieure à 0.05.

Enfin, l'information majeure de cette régression se trouve au niveau de la variable ***HorsDomicile***. Cette dernière est bien significative avec une p-valeur inférieure à 0.05, ce qui montre que le fait de travailler hors de son domicile a bien une influence sur l'infection à la Covid-19. De plus, le modèle nous retourne un odds ratio valant 1.21, ce qui est clairement supérieur à 1. Nous pouvons donc conclure que les individus travaillant hors de leur domicile pendant la crise sanitaire ont **21%** de plus de probabilité d'avoir été infecté par la Covid-19. Cela confirme bien notre intuition de départ qui était que ces travailleurs étaient plus exposés que les autres.

Le diplôme, quant à lui, n'est pas significatif.

Nous avons donc sensiblement les mêmes conclusions que pour le fait de travailler à l'extérieur. L'infection à la Covid-19 a principalement frappé les professionnels de la santé, ce qui est logique. Ensuite, ce sont les cadres supérieurs qui travaillaient hors de leur domicile qui ont le plus contracté le virus, et les jeunes adultes.

Characteristic	OR ¹	95% CI ¹	p-value
PCSI			
Cadres sups	—	—	
Agriculteurs	1.48	0.84, 2.44	0.15
Employés	0.70	0.59, 0.85	<0.001
Etudiants	0.42	0.31, 0.56	<0.001
H/F au Foyer	0.49	0.34, 0.68	<0.001
Indépendants	1.01	0.77, 1.32	>0.9
Ouvriers	0.57	0.44, 0.73	<0.001
Prof. intermédiaires	0.75	0.63, 0.90	0.002
Retraités	0.45	0.33, 0.62	<0.001
Sans Emploi	0.75	0.54, 1.03	0.088
HorsDomicile			
0	—	—	
1	1.21	1.06, 1.37	0.004
Sexe			
Femme	—	—	
Homme	0.88	0.79, 0.99	0.028
Age			
18-29	—	—	
30-39	0.86	0.73, 1.01	0.070
40-49	0.69	0.58, 0.82	<0.001
50-59	0.46	0.38, 0.55	<0.001
60+	0.40	0.30, 0.53	<0.001
Diplome			
Sans Diplome	—	—	
- de BAC	0.76	0.54, 1.08	0.11
BAC à BAC+2	0.75	0.54, 1.05	0.084
BAC+3 ou +	0.68	0.49, 0.97	0.027
ProfSante			
Non	—	—	
Oui	1.56	1.15, 2.08	0.003

¹ OR = Odds Ratio, CI = Confidence Interval

3.3 La santé mentale en période de pandémie

La crise sanitaire sans précédent à laquelle nous faisons face depuis le début de l'année 2020 affecte la population de différentes manières. En effet, l'infection à la Covid-19 est malheureusement loin d'être la seule conséquence sanitaire dûe à la pandémie. Nous nous intéresserons ici aux conséquences psychologiques et plus précisément à la dépression et à l'anxiété qui touchent les français depuis le début de la pandémie.

Le 12 mars 2021, la Direction de la recherche, des études, de l'évaluation et des statistiques (Drees) a publié une étude sur l'état de santé des Français qui comprend un questionnaire portant sur la santé mentale des enquêtés pendant le confinement du printemps 2020. Le constat est sans appel : "En mai 2020, 13,5% des personnes âgées de 15 ans ou plus vivant en France déclarent des symptômes évocateurs d'un état dépressif ", contre 10,9% en 2019.

Dans notre questionnaire DataCovid, l'état mental du panel est abordé à la question C5 par ces quatre sous-questions :

Question C5 : Au cours des deux dernières semaines, à quelle fréquence avez-vous été dérangé(e) par les problèmes suivants ?

- C5_0 : Vous sentir triste, déprimé(e) ou désespéré(e)
- C5_1 : Éprouver un sentiment de nervosité, d'anxiété ou de tension
- C5_2 : Être incapable d'arrêter de vous inquiéter ou de contrôler vos inquiétudes
- C5_3 : Avoir peu d'intérêt ou de plaisir à faire des choses

Valeurs possibles :

- 1 : Presque tous les jours
- 2 : Plus de la moitié des jours
- 3 : Plusieurs jours
- 4 : Jamais
- 5 : Non Réponse

Ces questions correspondent en réalité à un test de santé mentale : ***The Patient Health Questionnaire-4 (PHQ-4).***

3.3.1 Le score PHQ4

The Patient Health Questionnaire-4 (PHQ-4) est un questionnaire très bref constitué comme son nom l'indique de quatre questions. Il a été développé et validé par Kroenke, Spitzer, Williams et Löwe en 2009 afin de prévenir les premiers signes d'anxiété et de dépression qui sont deux des maladies psychiatriques les plus répandues dans la population. L'avantage de ce test est sa brièveté, ce qui le rend très rapide à compléter, évitant ainsi toute perte de concentration chez les patients anxieux ou déprimés.

Le PHQ-4 est un questionnaire à quatre réponses sur une échelle de type Likert. Son objectif est de permettre une mesure ultra brève et précise des principaux symptômes et signes de dépression et d'anxiété. Il combine la mesure à deux éléments (PHQ-2) - constituée des critères de base de la dépression - ainsi qu'une mesure à deux éléments de l'anxiété (GAD - 2), qui se sont tous deux révélés indépendamment être de bons outils de dépistage rapide.

A partir de ce test nous pouvons construire le score PHQ-4 en suivant le barème suivant :

Au cours des deux dernières semaines, à quelle fréquence avez-vous été dérangé(e) par les problèmes suivants ?	Presque tous les jours	Plus de la moitié des jours	Plusieurs jours	Jamais
Vous sentir triste, déprimé(e) ou désespéré(e)	3	2	1	0
Éprouver un sentiment de nervosité, d'anxiété ou de tension	3	2	1	0
Être incapable d'arrêter de vous inquiéter ou de contrôler vos inquiétudes	3	2	1	0
Avoir peu d'intérêt ou de plaisir à faire des choses	3	2	1	0

Nous obtenons un score PHQ-4 de 0 à 12.

Remarque : Toutefois, un score PHQ-4 élevé n'est pas un diagnostic, mais plutôt un indicateur pouvant mener à une enquête plus approfondie afin d'établir la présence ou l'absence d'un trouble clinique justifiant un traitement. Des recherches antérieures ont établi qu'un score de 3 ou plus sur la sous-échelle de la dépression représente un point de coupure raisonnable pour identifier les cas potentiels de dépression. Un score de 3 ou plus est donc signe de dépression et l'état du patient doit ainsi être évalué de manière plus approfondie par un autre questionnaire PHQ-9 ou par un suivi médical. De même, un score de 3 ou plus sur la sous-échelle Anxiété représente un point de coupure raisonnable. Un score de 3 ou plus est positif et doit être évalué par GAD-7 ou un examen en santé mentale doit être effectué.

Notre premier réflexe a été de traiter ce score comme une variable quantitative et d'effectuer une régression linéaire dessus, ce qui s'avère maladroit pour ce type de score. En effet, passer de 0 à 1 n'est pas pareil que de passer de 11 à 12. Nous avons finalement préféré traiter cette variable de façon qualitative, en nous appuyant sur les seuils de détresse psychologique associés à ce test :

- Non : 0-2
- Bénigne : 3-5
- Modérée : 6-8
- Sévère : 9-12

Nous avons donc une variable avec 4 modalités ordonnées que nous voulons expliquer afin de comprendre les facteurs favorisant la détresse psychologique durant la pandémie.

Nous allons donc effectuer une régression logistique polytomique ordonnée.

3.3.2 Régression logistique polytomique ordonnée

Le modèle polytomique ordonné peut être présenté comme une généralisation du modèle logistique dichotomique présenté précédemment. Cette fois Y ne prend plus 2 modalités mais $K > 2$ modalités ordonnées. Par exemple ici nos 4 niveaux de détresse psychologique que l'on numérottera de 1 à 4.

On introduit alors $K - 1$ seuils déterministes notés $\alpha^1, \dots, \alpha^{K-1}$ tels que :

$$(Y|X = x) = \begin{cases} 1 & \text{si } Y^* < \alpha^1 \\ k & \text{si } \alpha^{k-1} \leq Y^* < \alpha^k \quad \text{pour } k = 2, \dots, K - 1 \text{ avec } Y^* = x\hat{\beta} \\ K & \text{si } Y^* \geq \alpha^{K-1} \end{cases}$$

On reprend les notations vues précédemment pour le modèle linéaire généralisé. Nous sommes ici aussi dans une régression logistique, nous avons donc :

$$g(p_\beta^k) = g(\mathbb{P}_\beta^j(Y \leq k|X = x)) = \alpha^k - x\beta \text{ avec } g(x) = \log\left(\frac{x}{1-x}\right) \quad \forall k \in \llbracket 1, K-1 \rrbracket$$

Nous prenons ici $p_\beta^j = \mathbb{P}_\beta^j(Y \leq j|X = x)$ car nous sommes dans un modèle logistique polytomique **cumulatif**. Ce modèle nous permet de comparer une catégorie avec toutes les catégories qui lui sont inférieures.

Nous avons alors $K-1$ équations LOGITS de la forme :

$$\log\left(\frac{\mathbb{P}_\beta^k(Y \leq k|X = x)}{\mathbb{P}_\beta^k(Y > k|X = x)}\right) = \alpha^k - \beta_1^k X_1 - \dots - \beta_m^k X_m, \quad \forall k \in \llbracket 1, K-1 \rrbracket \quad (1)$$

Notre modèle nous retournera alors $K-1$ estimations pour chaque coefficient.

Cependant, évaluer les influences niveau par niveau est très délicat, voir périlleux, il nous faut introduire une contrainte supplémentaire pour rendre les résultats de la régression exploitables : la **contrainte sur les pentes**.

Contrainte sur les pentes : Cette contrainte également appelée hypothèse de proportionnalité des odds ratios, consiste à supposer que les coefficients des co-variables sont les mêmes quel que soit le niveau étudié. C'est à dire que chaque variable explicative agit de la même manière sur la variable expliquée, quel que soit le niveau k .

Nous obtenons alors un unique modèle simplifié

$$\log\left(\frac{\mathbb{P}_\beta^k(Y \leq k|X = x)}{\mathbb{P}_\beta^k(Y > k|X = x)}\right) = \alpha^k - \beta_1 X_1 - \dots - \beta_m X_m, \quad \forall k \in \llbracket 1, K-1 \rrbracket \quad (2)$$

Ainsi, dans ce modèle, quelle que soit la modalité k considérée, une variable explicative donnée a la même influence sur $\mathbb{P}_\beta^k(Y \leq k|X = x)$. Les β sont les mêmes pour les différentes modalités k , sauf pour les α^k qui correspondent aux différents intercepts de chaque modalité. Ce modèle nécessite alors l'estimation de $m + K - 1$ coefficients (m pentes et K - 1 constantes).

Pour tester la pertinence de ce dernier modèle, nous devons effectuer le **test d'égalité des pentes**.

Test d'égalité des pentes : On se pose la question de vérifier si la modélisation en terme de seuils aléatoires est "raisonnable" vis-à-vis de nos données. Un moyen de vérifier si cette hypothèse est raisonnable consiste à tester l'hypothèse d'égalité des pentes. Cette approche consiste tout simplement à comparer le modèle en question (le modèle (2) avec égalité des pentes) à un modèle où on lève l'égalité des pentes (le modèle (1)). Il s'agit de considérer les hypothèses

$$\mathbf{H}_0 : \forall j \in \llbracket 1, m \rrbracket, \beta_j^1 = \dots = \beta_j^{K-1} \text{ vs } \mathbf{H}_1 : \exists k, k' \in \llbracket 1, K-1 \rrbracket, \exists j \in \llbracket 1, m \rrbracket \text{ tels que } \beta_j^k \neq \beta_j^{k'}$$

Sous \mathbf{H}_0 , on retrouve le modèle de l'équation (2). Ce test revient simplement à faire un test d'égalité, nous pouvons alors choisir la façon dont on souhaite l'effectuer. Nous choisissons ici d'effectuer un test du rapport de vraisemblance.

On choisit donc la statistique de test :

$$S = -2\log\left(\frac{\mathcal{L}_n(\hat{\gamma}_0)}{\mathcal{L}_n(\hat{\gamma})}\right)$$

avec $\hat{\gamma}_0 = (\hat{\alpha}^1, \dots, \hat{\alpha}^{K-1}, \hat{\beta}_1, \dots, \hat{\beta}_m)$ correspondant aux coefficients du modèle (2) qui correspond également au modèle (1) sous \mathbf{H}_0 et $\hat{\gamma} = (\hat{\alpha}^1, \dots, \hat{\alpha}^{K-1}, \hat{\beta}_1^1, \dots, \hat{\beta}_m^1, \dots, \hat{\beta}_1^{K-1}, \dots, \hat{\beta}_m^{K-1})$ correspondant aux coefficients du modèle (1).

Sous \mathbf{H}_0 , on a : $S \sim \chi^2_{p(K-2)}$. Ainsi, p-valeur = $1 - F_{\chi^2_{p(K-2)}}(S)$

Remarque : Le nombre de degrés de liberté s'obtient en faisant la différence entre la dimension du modèle (1) ($K - 1 + p(K - 1)$) et celle du modèle sous \mathbf{H}_0 , le modèle (2) ($K - 1 + p$). En effet $K - 1 + p(K - 1) - (K - 1 + p) = p(K - 2)$.

Le modèle passera avec succès le test d'égalité des pentes si la p-valeur > 5%.

3.3.3 Profil des personnes souffrant de problèmes de santé mentale durant la crise

Notre objectif est ici d'expliquer notre variable ***PHQ_seuil***, qui prend des valeurs de 1 à 4 correspondant aux niveaux de détresse psychologique suivants vus précédemment :

- 1 : Non — 3 : Modérée
- 2 : Bénigne — 4 : Sévère

Comme pour l'exposition à l'infection, nous cherchons à expliquer cette variable grâce aux variables ***PSCI***, ***HorsDomicile***, ***Sexe***, ***Age***, ***Diplome***, ***Infecte*** et ***ProfSante***, mais dans un modèle logistique polytomique ordonné.

Remarque : Nous avons précédemment présenté le modèle polytomique ordonné cumulatif "classique". Mais dans notre cas, nous cherchons à quantifier le passage à un niveau de détresse psychologique supérieur. Nous utilisons donc $p_\beta^j = \mathbb{P}_\beta^j(Y > j | X = x)$.

On récupère alors les coefficients de ces deux modèles afin d'effectuer le test d'égalité des pentes comme expliqué ci-dessus. Nous obtenons une p-valeur de l'ordre de 10^{-10} , ce qui est bien inférieur à 5%. Ainsi, on rejette \mathbf{H}_0 .

Nous concluons que l'influence de nos variables explicatives n'est pas le même pour chaque passage aux différents niveaux de détresse psychologique. Nous pouvons par exemple observer les différentes valeurs du coefficient de la variable ***HorsDomicile***.

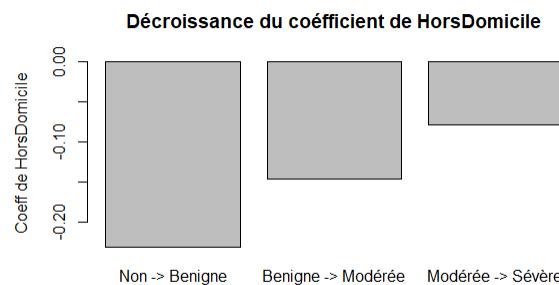


FIGURE 26 – Décroissance du coefficient de HorsDomicile

Tout d'abord, pour les trois régressions correspondant aux trois passages au niveau de détresse psychologique supérieur, le coefficient est négatif. Le fait de travailler en extérieur a donc

un effet positif sur la santé mentale. Cependant, nous pouvons voir que ce coefficient se rapproche grandement de 0 pour la troisième régression. Ainsi, plus le niveau de détresse psychologique est préoccupant, moins le fait de travailler dehors joue un rôle important.

Toujours dans le but de déterminer le profil des personnes touchées par ces problèmes psychologiques, nous allons à la place effectuer une régression logistique. Nous prendrons comme variable à expliquer **PHQ4**, qui a été dichotomisée et vaut 1 si l'individu souffre de dépression sévère ou modérée, et 0 sinon. En variables explicatives, nous allons utiliser **PSCI**, **HorsDomicile**, **Sexe**, **Age**, **Diplome**, **Infecte** et **ProfSante**.

Interprétation : En terme de p-valeurs, toutes les variables sont significatives, sauf la variable **ProfSante** et les modalités "H/F au foyer", "Etudiants" et "Prof. intermédiaires" de la variable **PSCI**. Pour **ProfSante**, c'est assez étrange, nous pensions justement que les personnes travaillant en milieu hospitalier étaient mentalement très touchées par la crise. Quant à la non-significativité de la modalité "Etudiants", elle couvre peut-être la même information que la tranche d'âge "18-29".

Le premier résultat intéressant concerne la variable **HorsDomicile**. On observe un odds ratio de 0.85 pour les personnes travaillant à l'extérieur. Ainsi, quelqu'un travaillant dehors a presque 20% de chances en moins d'être gravement déprimé que quelqu'un qui travaille de chez lui. Cela peut s'expliquer par la durée du confinement, qui pour les personnes travaillant à domicile pouvait être synonyme d'isolement total. Les personnes travaillant dehors ont eu plus de chances de croiser du monde et de garder un contact social.

La différence sur la dépression est frappante entre les hommes et les femmes. Une femme a 1,4x plus de chances d'être fortement déprimée qu'un homme. Il faut néanmoins être modéré sur ce résultat. Les hommes ont tendance à moins consulter de psychologues que les femmes, et sont moins enclins à reconnaître des difficultés psychologiques. Il est donc probable qu'un grand nombre d'hommes répondants au sondage aient minimisé leur anxiété.

Tout comme pour les risques d'infection, les odds ratio de la variable **Age** sont décroissants. Plus la personne est âgée, et moins elle a eu de chances de développer une dépression sévère durant le début de la pandémie. Nous pensons que les personnes âgées souffriraient mentalement de l'isolement, et de la peur du virus, mais ceux-ci ont près de 50% moins de chances de souffrir d'une maladie mentale que quelqu'un de moins de 30 ans.

Parmi les différentes catégories socio-professionnelles, ce sont les cadres supérieurs et les professions intermédiaires qui ont le moins souffert de dépression. Les agriculteurs, eux, en ont fortement souffert. Néanmoins, la dépression dans le secteur agricole est un phénomène qui existe depuis longtemps, et cela n'a donc pas forcément

Characteristic	OR [†]	95% CI [†]	p-value
PSCI			
Cadres sups	—	—	
Agriculteurs	1.95	1.35, 2.76	<0.001
Employés	1.20	1.08, 1.34	0.001
Etudiants	1.10	0.94, 1.29	0.2
H/F au Foyer	0.97	0.82, 1.15	0.7
Indépendants	1.25	1.06, 1.48	0.008
Ouvriers	1.21	1.06, 1.39	0.006
Prof. intermédiaires	0.93	0.84, 1.05	0.2
Retraités	0.83	0.71, 0.97	0.016
Sans Emploi	1.56	1.32, 1.84	<0.001
HorsDomicile			
0	—	—	
1	0.85	0.78, 0.91	<0.001
Sexe			
Femme	—	—	
Homme	0.59	0.55, 0.62	<0.001
Age			
18-29	—	—	
30-39	0.78	0.71, 0.86	<0.001
40-49	0.68	0.61, 0.75	<0.001
50-59	0.55	0.49, 0.61	<0.001
60+	0.53	0.46, 0.61	<0.001
Diplome			
Sans Diplome	—	—	
- de BAC	0.78	0.65, 0.92	0.004
BAC à BAC+2	0.68	0.57, 0.80	<0.001
BAC+3 ou +	0.64	0.54, 0.76	<0.001
Infecte			
N'a pas contracté le virus	—	—	
A contracté le virus	2.47	2.21, 2.76	<0.001
ProfSante			
Non	—	—	
Oui	0.87	0.70, 1.08	0.2

[†] OR = Odds Ratio, CI = Confidence Interval

de rapport avec la pandémie. Par contre, les sans emplois en souffrent beaucoup (à peu près 60% de plus que les cadres). Leur situation professionnelle et la situation économique de l'état durant cette période expliquent sûrement ce phénomène.

Pour le diplôme, qui n'est pas extrêmement significatif, ce sont les personnes sans diplôme qui souffrent le plus de dépression, et il semblerait que plus le diplôme d'une personne est de haut niveau, moins elle a eu de probabilité de développer une dépression sévère durant la pandémie.

Enfin, la variable la plus frappante est **Infecte**. Avec un odds ratio de 2.47 pour la modalité "A contracté le virus", elle nous apprend que quelqu'un ayant déjà contracté le virus a 2,5x plus de chances d'avoir été fortement déprimé que quelqu'un n'ayant pas eu la Covid-19. C'est très élevé, et cela montre à quel point quelqu'un n'ayant pas été au contact du virus a tendance à minimiser ce dernier.

Les personnes ayant le plus de chances d'avoir été fortement déprimées durant le début de la pandémie sont donc les femmes, assez jeunes, qui soit ne travaillaient pas, soit travaillaient à domicile, et qui ont suivi un parcours scolaire moins poussé.

4 Conclusion

Nous trouvons les résultats exposés lors de ce mémoire très intéressants. En effet, ces derniers ne correspondent pas exactement à ce que nous avions supposé au début de notre analyse. Nous pensions alors naïvement que les personnes de classe sociale plus modeste, contraints de travailler hors de leur domicile, seraient plus exposées au virus. Finalement, il s'avère que ce sont les cadres supérieurs qui travaillaient le plus dehors, et qui furent les plus contaminés durant la période étudiée.

Concernant les troubles psychologiques, au contraire, nos conclusions vont dans un sens plus prévisible. En effet, ce sont les personnes de rang social moins élevé et les plus jeunes qui ont le plus souffert de dépression et d'anxiété.

Il y a donc bien des disparités dans l'exposition à la Covid-19 (exposition au virus, et à la peur de ce dernier), et l'obligation de travail à l'extérieur, selon les caractéristiques démographiques des individus.

La base de données DataCovid a été pour nous un sujet d'étude très intéressant, nous permettant de travailler avec un très grand nombre d'observations issues d'un panel représentant fidèlement notre population. Elle nous a permis de traiter le sujet proposé par Mme Dormont et d'en tirer les résultats exposés ci-dessus. Toutefois, nous aurions aimé poursuivre notre étude sur d'autres axes abordant d'avantage les inégalités sociales, mais nos données ne nous ont pas permis d'aller dans ce sens. En effet, nous trouvons certaines formulations de questions maladroites et nous nous sommes également parfois interrogé sur le sérieux de certaines réponses au questionnaire.

Bibliographie

- Haut Conseil de la Santé, adsp n°73 : *Les inégalités sociales de santé*, 2010
- Richard BLUNDELLE, Monica COSTA DIAS, Robert JOYCE, Xiaowei XU : *CO-VID_19 and inequalities*, 2020
- Caitlin S. BROWN, Martin RAVALLION : *Inequality and the coronavirus : socioeconomic covariates of behavioral responses and viral outcomes across US counties*, 2020
- A. COLIN CAMERON, Pravin K. TRIVEDI : *Microeconomics, Methods and Applications*, 2005
- M. STUDER, *Étude des inégalités de genre en début de carrière académique à l'aide de méthodes innovatrices d'analyse de données séquentielles*, 2012
- RSTUDIO : *WeightedCluster Library Manual*
- DREES : *Confinement du printemps 2020 : une hausse des syndromes dépressifs, surtout chez les 15-24 ans*, 2020
- Laurent ROUVIÈRE : *Régression logistique avec R*
- Université de Lyon : *Régression logistique polytomique ordinaire*
- K. KROENKE : *The Patient Health Questionnaire-4 (PHQ-4)*

Annexe A

Questionnaire de la vague 1

datacovid.org - Baromètre COVID-19

Description des données de la Vague 1

Méthode

La première vague du baromètre COVID19 a été administrée sur internet par l'institut de sondage IPSOS auprès d'un échantillon de 5000 personnes représentatif de la population française métropolitaine. L'échantillon a été constitué à par tir du panel IPSOS propriétaire (voir la description du panel ci-dessous) selon la méthode des quotas (sexe, âge, région, catégorie d'agglomération, Situation professionnelle, CSP et situation familiale). Le questionnaire administré est disponible dans l'archive.

Description du panel

L'échantillon est constitué de 5000 personnes, constituant un échantillon représentatif de la population française agée de 18 ans et plus (voir tableau 1 pour un descriptif détaillé). Ces personnes ont été interrogées du 7 au 14 Avril 2020 par internet.

Le panel IPSOS est constitués d'individus :

- Volontaires pour participer à l'étude et uniquement invités aux études Ipsos ;
- Recrutés et gérés pour un usage à long terme ;
- Qualifiés afin de cibler les bons profils selon les besoins des études.

	Population (Eurostat 2016)	Panel IPSOS (Juillet 2019)
ÂGE		
16-24	15%	22%
25-34	17%	20%
35-44	18%	22%
45-54	19%	19%
55-74	32%	17%
GENRE		
Homme	49%	38%
Femme	51%	62%
REGION		
Île de France	19%	15%
Bassin Parisien	17%	19%
Nord	6%	8%
Est	8%	8%
Ouest	14%	16%
Sud-Ouest	11%	11%
Centre est	12%	11%
Aire méditerranéenne	13%	12%

Tableau 1 Description du panel IPSOS

Baromètre COVID 19

Un projet de science citoyenne

24.04.2020

LICENCE



Le « baromètre COVID 19 » est un ensemble de données ouvertes, disponibles sous la licence libre « [Licence ouverte 2.0 / Open licence 2.0](#) » accordée par la [etalab](#).

Vous êtes libres de

- Communiquer, le reproduire, le copier ;
- L'adapter, le modifier, l'extraire et le transformer, notamment pour créer des « Informations dérivées » ;
- Le diffuser, le redistribuer, le publier et le transmettre, de l'exploiter à titre commercial, par exemple en le combinant avec d'autres informations, ou en l'incluant dans votre propre produit ou application.

Sous réserve de :

- Mentionner la paternité de l'« Information » : sa source, « Baromètre COVID 19, <https://datacovid.org> », et la date de la dernière mise à jour du baromètre réutilisée.

Le [texte légal](#) complet détaille vos droits et responsabilités.

Comment créditer le « baromètre COVID 19 » ?

Nous demandons que votre crédit comporte la mention « © les contributeurs du baromètre COVID 19 ».

Vous devez également préciser clairement que les données sont disponibles sous la licence « LICENCE OUVERTE 2.0/OPEN LICENCE 2.0 ». Vous pouvez mentionner ceci avec un lien hypertexte vers [cette page de mentions légales](#). Alternativement, et obligatoirement si vous distribuez le « baromètre COVID 19 » sous forme de données brutes, vous pouvez directement nommer et fournir un lien vers la licence. Sur les supports où les liens hypertexte sont impossibles (par exemple sur un support papier), nous vous suggérons de rediriger vos lecteurs vers le site datacovid.org (éventuellement en développant « Baromètre COVID 19 » en son adresse complète datacovid.org), vers <https://www.etalab.gouv.fr/licence-ouverte-open-licence>.

Conditions d'utilisation des jeux de données issus du Baromètre COVID-19

Les jeux de données publiés par datacovid.org sur son site datacovid.org/data, datacovid.org/api, et sur data.gouv.fr, sont issus du [Baromètre Covid-19](#) opéré par [IPSOS](#) en partenariat avec datacovid.org, association à but non lucratif régie par la loi française de 1901.

Ces jeux de données :

- Sont régis par le droit français et par les [conditions d'utilisation](#) du site datacovid.org,
- Sont publiés sur Internet dans un objectif non lucratif, scientifique et citoyen pour combler le déficit d'information sur les dispositifs de gestion sociétale des épidémies,
- Ont été préalablement expurgés de toute donnée permettant d'identifier une personne ayant répondu au [Baromètre Covid-19](#),
- Sont ouverts, c'est-à-dire qu'ils peuvent être consultés, utilisés et partagés par tous, en particulier à des fins de recherche, notamment scientifique, historique et statistique,
- Ne doivent pas donner lieu à un usage commercial, sauf à ce que les résultats dérivés de ces jeux de données, directement ou indirectement, puissent être ouverts eux aussi, au



Baromètre COVID 19

Un projet de science citoyenne

24.04.2020

sens défini ci-dessus et portés à la connaissance de datacovid.org pour en assurer l'ouverture,

- Ne doivent pas être rapprochés avec d'autres jeux de données ou d'autres ressources dans des conditions qui permettraient à un tiers, par corrélation, par inférence ou par quelque moyen que ce soit, d'identifier une personne ayant répondu au [Baromètre Covid-19](#).

En conséquence,

- Toute utilisation de ces jeux de données et de tout ou partie de leurs éléments constitutifs qui ne respecterait pas chacune des conditions énumérées ci-dessus est interdite, en particulier :
- Tout usage commercial non ouvert au sens défini ci-dessus, est interdit et serait passible de poursuites et de sanctions civiles, administratives et pénales en application de la réglementation française en vigueur,
- toute corrélation ou toute inférence entre les éléments constitutifs de ces jeux de données et d'autres sources de données, qui permettrait à un tiers d'identifier une personne ayant répondu au [Baromètre Covid-19](#), engagerait la responsabilité d'un tel tiers à l'égard de datacovid.org et de toute personne concernée et serait passible de poursuites et de sanctions civiles, administratives et pénales en application de la réglementation française en vigueur.

En téléchargeant ou en rendant accessibles à un tiers des données issues du [Baromètre Covid-19](#), je m'engage à respecter les objectifs ci-dessus et les [conditions d'utilisation des jeux de données](#) publiés par datacovid.org. Si j'ai des questions ou des doutes, j'interroge contact@datacovid.org tout en restant responsable de mes actes ou de ceux de mes préposés et prestataires.

NB : Les données de trafic relatives à l'accès au site datacovid.org sont traitées par datacovid.org et ses prestataires afin d'en mesurer la fréquentation et d'assurer la disponibilité, l'intégrité et la sécurité du site et de ses contenus, dans des conditions et selon des durées de conservation conformes à la réglementation française. Toute personne physique justifiant de son identité peut écrire à contact@datacovid.org pour exercer [les droits qui lui sont garantis](#) par la réglementation [française](#) et [européenne](#) en vigueur relative à la protection des données personnelles et de la vie privée, notamment les droits d'accès, d'opposition ou de suppression des données personnelles la concernant traitées par datacovid.org.

Disponibilité des données

Les résultats du questionnaire de la vague 1 du Baromètre COVID-19 de dataCovid.org sont disponibles sous deux formes : 1) Dans un fichier au format TSV (« Tabulation Separated Values ») téléchargeable sur datacovid.org/data; 2) Via une api « REST ». Les deux formes sont décrites dans ce document, sous les angles suivants :

- 1- Nom des champs
- 2- Partie du questionnaire à laquelle il se réfère
- 3- Signification des codes réponses

Fichier TSV

Le fichier TSV, disponible sur datacovid.org/data, est constitué de 81 colonnes, correspondant aux réponses aux différentes questions du questionnaire. La signification de chacun des codes réponses de chacune des questions est également fourni dans le fichier Dictionnaire_codes.tsv.

- 1: SERIAL – Identifiant de la réponse/du participant. Valeurs possibles:
 - o (de 1 à 5000)
- 2: JOUR – Jour de la réponse au questionnaire. Valeurs possibles:
 - o 1: Mardi 7 Avril
 - o 2: Mercredi 8 Avril
 - o 3: Jeudi 9 Avril
 - o 4: Vendredi 10 Avril



Baromètre COVID 19

Un projet de science citoyenne

24.04.2020

- 5 : Samedi 11 Avril
 - 6 : Dimanche 12 Avril
 - 7 : Lundi 13 Avril
 - 8 : Mardi 14 Avril
- 3 : DEVICE - Type d'appareil utilisé pour répondre au questionnaire. Valeurs possibles :
 - 1: Ordinateur
 - 2: Smartphone
 - 3: Tablette
 - 4: SmartTV
 - 5: Aucun
- 4 : SEXE – Genre du participant. Valeurs possibles :
 - 1: Homme
 - 2: Femme
- 5 : AGE – Age du participant. Valeurs possibles :
 - 1: 18 à 24 ans
 - 2: 25 à 29 ans
 - 3: 30 à 34 ans
 - 4: 35 à 39 ans
 - 5: 40 à 44 ans
 - 6: 45 à 49 ans
 - 7: 50 à 54 ans
 - 8: 55 à 59 ans
 - 9: 60 à 64 ans
 - 10 : 65 ans et plus
- 6 : DEP – Département. Valeurs possibles :
 - 1-96 : Numéro des départements (à l'exception de la corse, voir annexe 1 pour le détail)
- 7 : REG12 – Région. Valeurs possibles :
 - 1: Ile-de-France
 - 2: Centre-Val de Loire
 - 3: Bourgogne-Franche-Comté
 - 4: Normandie
 - 5: Hauts-de-France
 - 6: Grand-Est
 - 7: Pays de la Loire
 - 8: Bretagne
 - 9: Nouvelle Aquitaine
 - 10 : Occitanie
 - 11 : Auvergne-Rhône-Alpes
 - 12 : PACA + Corse
- 8 : AGGLO5 – Catégorie d'agglomération. Valeurs possibles :
 - 1: Rural
 - 2: 2.000 à 19.999 hab.
 - 3: 20.000 à 99.999 hab.
 - 4: 100.000 hab. et plus
 - 5: Agglomération de Paris
- 9 : EMP - Situation professionnelle actuelle. Valeurs possibles :
 - 1: Salarié(e) à plein temps
 - 2: Salarié(e) à temps partiel
 - 3: Travaille à mon compte
 - 4: Ne travaille pas actuellement, mais en recherche d'emploi
 - 5: Ne travaille pas actuellement et ne recherche pas d'emploi / en incapacité de travail
 - 6: Homme/femme au foyer
 - 7: Retraité(e)
 - 8: Etudiant(e)/élève
- 10 : PCSI – Catégorie Socio-Professionnelle du participant. Valeurs possibles :
 - 1: Agriculteurs exploitants
 - 2: Professions Indépendantes
 - 3: Cadres Supérieurs
 - 4: Professions intermédiaires
 - 5: Employés
 - 6: Ouvriers
 - 7: Retraités
 - 8: Inactifs
- 11 : MAR – Situation familiale. Valeurs possibles :
 - 1: Célibataire
 - 2: En union libre



Baromètre COVID 19

Un projet de science citoyenne

24.04.2020

- 3 : Marié(e)
 - 4 : Séparé(e)
 - 5 : Divorcé(e)
 - 6 : Veuf/ve
 - 7 : Pacsé(e)
- 12-14 : Question A0 : Parmi les sujets suivants, quelles sont les trois qui vous semblent les plus préoccupants aujourd'hui, pour vous personnellement ? (Colonne A0_0 : Préoccupation 1, A0_1 : Préoccupation 2, A0_2 : Préoccupation 3). Valeurs possibles:
 - 1 : L'épidémie de coronavirus
 - 2 : Le pouvoir d'achat
 - 3 : Le changement climatique
 - 4 : Les inégalités sociales
 - 5 : Le chômage
 - 6 : Les déficits publics et de la dette de l'État
 - 7 : L'éducation
 - 8 : Le système de santé
 - 9 : L'insécurité
 - 10 : Le terrorisme
- 15 : Question A1 : Avez-vous ou pensez-vous avoir été infecté par le coronavirus ? Valeurs possibles :
 - 1: Oui
 - 2: Non
- 16 : Question A2 : Pourquoi avez-vous ou pensez-vous avoir été infecté par le coronavirus ? Valeurs possibles :
 - Vide : Non à la question précédente
 - 1: J'ai été diagnostiquée positif par un test de dépistage
 - 2: J'ai ou j'ai eu les symptômes caractéristiques de la maladie et j'ai contacté un médecin
 - 3: J'ai ou j'ai eu les symptômes caractéristiques de la maladie mais je n'ai pas contacté de médecin
 - 4: Je n'ai pas eu de symptômes caractéristiques de la maladie mais j'ai été en contact proche avec au moins une personne diagnostiquée positive par un test de dépistage
 - 5: Je n'ai pas eu de symptômes caractéristiques de la maladie mais j'ai été en contact proche avec au moins une personne qui pense avoir été infectée par le coronavirus
 - 6: Je ne connais pas assez bien les symptômes pour répondre.
- 17-18 : Question A3 : Depuis le début de l'épidémie de coronavirus, à votre connaissance, avez-vous été en contact avec au moins une personne qui a été diagnostiquée comme malade du coronavirus avec un test de dépistage ? Plusieurs réponses possibles (A3_0 : Première réponse, A3_1 : Deuxième réponse). Valeurs possibles :
 - 1: Oui, j'ai la certitude d'avoir été en contact avec au moins une personne au sein de mon foyer
 - 2: Oui, j'ai la certitude d'avoir été en contact avec au moins une personne en dehors de mon foyer
 - 3: Non, je n'ai pas la certitude d'avoir été en contact avec au moins une personne [EXCLUSIVE]
- 19 : Question A4 : Au cours de ces dernières 48 heures, pensez-vous avoir eu de la fièvre (frissons, sueurs) ? Valeurs possibles :
 - 1: Oui
 - 2: Non
- 20 : Question A5 : Lorsque vous avez eu de la fièvre au cours de ces dernières 48 heures, quelle a été votre température la plus élevée ? Soyez le plus précis possible et donner un chiffre après la virgule.
 - Vide : Non à la question précédente
 - 0,0 : Je ne sais pas, je n'ai pas pris ma température
 - Autre : Température (1 décimale)
- 21-25 : Question A6 : Durant ces derniers jours, avez-vous rencontré les problèmes suivants ?
 - Colonnes:
 - A6_0 : Une toux ou une augmentation de votre toux habituelle
 - A6_1 : Une forte diminution de votre goût ou de votre odorat
 - A6_2 : Un mal de gorge ou des douleurs musculaires ou des courbatures inhabituelles
 - A6_3 : Une fatigue inhabituelle qui vous oblige à vous reposer plus de la moitié de la journée
 - A6_4 : Des maux de tête inhabituels
 - Valeurs possibles:
 - 1: Oui
 - 2: Non
 - 3: Non réponse
- 26-30 : Question A7 : Durant ces dernières 24 heures avez-vous rencontré les problèmes suivants ?
 - Colonnes:
 - A7_0 : Une impossibilité de vous alimenter ou de boire
 - A7_1 : Un manque de souffle inhabituel lorsque vous parlez ou faites un petit effort



Baromètre COVID 19

Un projet de science citoyenne

24.04.2020

- A7_2 : Des diarrhées (au moins 3 selles molles)
- A7_3 : Des pertes d'équilibre inhabituelles
- A7_4 : Une conjonctivite à un niveau inhabituel (Une conjonctivite est une inflammation qui se manifeste notamment par la rougeur d'un ou des deux yeux, associée à des démangeaisons, des picotements, un écoulement clair ou purulent).
- Valeurs possibles :
 - 1: Oui
 - 2: Non
 - 3: Non réponse
- 31 : Question B1 : Sur une échelle de 0 à 10, comment évaluez-vous la gravité de l'épidémie de coronavirus aujourd'hui ? Valeurs possibles :
 - 0 (Pas grave du tout) à 10 (Extrêmement grave)
- 32-40 : Question B2 : À quel point arrivez-vous à mettre en œuvre les gestes suivants ?
 - Colonnes :
 - B1_0 : Se laver les mains plusieurs fois par jour à l'eau et au savon
 - B1_1 : Utiliser du gel hydroalcoolique
 - B1_2 : Tousser ou éternuer dans son coude ou un mouchoir
 - B1_3 : Utiliser des mouchoirs à usage unique et les jeter
 - B1_4 : Maintenir une distance d'au moins 1 mètre avec les personnes hors de mon foyer
 - B1_5 : Ne pas serrer la main ni embrasser
 - B1_6 : Éviter les regroupements (transports en commun, fêtes, etc.)
 - B1_7 : Rester chez soi
 - B1_8 : Porter un masque
 - Valeurs possibles :
 - 1: Tout le temps
 - 2: Souvent
 - 3: Parfois
 - 4: Rarement
 - 5: Jamais même si je pourrais le faire
 - 6: Jamais car je n'ai pas la possibilité de le faire
 - 7: Non Réponse
- 41 : Question B3 : Au cours des dernières 24 heures, combien de temps êtes-vous sorti de votre foyer (hors jardin) ? Prenez en compte toutes vos sorties. Vous pouvez répondre en heures et en minutes pour être le plus précis possible. Valeurs possibles :
 - Valeur en minutes
- 42-46 : Question B4 : Cochez toutes raisons pour lesquelles vous êtes sorti de votre foyer (hors jardin) au cours des dernières 24 heures (B4_0 : Première réponse cochée, ..., B4_4 : 5^{ème} réponse cochée) ? Valeurs possibles :
 - 1: Faire des achats de première nécessité
 - 2: Pratiquer un sport ou me promener (seul, avec des membres de mon foyer, ou un animal de compagnie)
 - 3: Me déplacer pour une consultation médicale, un motif familial impérieux, ou une convocation judiciaire ou administrative (par ex. tâche d'intérêt général).
 - 4: Me rendre sur mon lieu de travail
 - 5: Autre
 - Vide : Pas d'autre réponse cochée
- 47 : Question B5 : Au cours des dernières 24 heures, avec combien de personnes avez-vous été en contact proche (à moins de 1 mètre) en dehors des personnes de votre foyer ? Essayer de prendre en compte toutes les personnes que vous avez rencontrées (dans les magasins, dans la rue, etc.). Valeurs possibles :
 - Nombre entier
- 48 : Questions C1C2 : Indice de Masse corporelle. Valeurs possibles :
 - 1: Insuffisance pondérale (inférieur à 18.5)
 - 2: Poids normal (entre 18.5 et 25)
 - 3: Surpoids (entre 25 et 30)
 - 4: Obésité (supérieur à 30)
 - 5: Non réponse
- 49 : Question C3 : Êtes-vous enceinte ?
 - 1: Oui
 - 2: Non
 - 3: Non réponse
 - Vide : Non applicable
- 50-57 : Question C4 : Parmi les problèmes de santé suivants, cochez tous ceux dont vous souffrez.
 - Colonnes :
 - C4_0 : Diabète
 - C4_1: Cancer (ces trois dernières années)
 - C4_2 : Maladie respiratoire



Baromètre COVID 19

Un projet de science citoyenne

24.04.2020

- C4_3 : Insuffisance rénale chronique dialysée
- C4_4 : Maladie chronique du foie
- C4_5 : Hypertension artérielle mal équilibrée, maladie cardiaque ou vasculaire nécessitant un traitement à visée cardiologique
- C4_6 : Maladie connue pour diminuer vos défenses immunitaires
- C4_7 : Maladie nécessitant un traitement immunosupresseur (C'est un traitement qui diminue vos défenses contre les infections)
- Valeurs possibles:
 - 1: Oui
 - 2: Non
 - 3: Ne sais pas
- 58-61 : Question C5 : Au cours des deux dernières semaines, à quelle fréquence avez-vous été dérangé(e) par les problèmes suivants ?
 - Colonnes:
 - C5_0 : Vous sentir triste, déprimé(e) ou désespéré(e)
 - C5_1 : Éprouver un sentiment de nervosité, d'anxiété ou de tension
 - C5_2 : Être incapable d'arrêter de vous inquiéter ou de contrôler vos inquiétudes
 - C5_3 : Avoir peu d'intérêt ou de plaisir à faire des choses
 - Valeurs possibles:
 - 1: Presque tous les jours
 - 2: Plus de la moitié des jours
 - 3: Plusieurs jours
 - 4: Jamais
 - 5: Non Réponse
- 62-65 : Question C6 : En cette période, quelle est votre situation professionnelle ? Cochez toutes les réponses qui correspondent à votre situation (C6_0 : Première réponse cochée, ..., C6_3 : 4^{ème} réponse cochée). Valeurs possibles:
 - C6_0 : Je travaille hors de mon domicile
 - C6_1 : Je travaille à domicile (comme d'habitude)
 - C6_2 : Je travaille à distance (télétravail)
 - C6_3 : Je suis en arrêt maladie (exclusif avec C6_5)
 - C6_4 : Je suis en chômage partiel
 - C6_5 : Je suis en congé (exclusif avec C6_3)
- 66-68 : Question C7 : Veuillez indiquer le code postal et la commune dans laquelle vous résidez en ce moment, pendant cette période de confinement ?
 - Colonnes:
 - C7 : Résidence habituelle ? (1: Oui, 2: Non)
 - C7DEP : Département (1-96, Vide, voir annexe 1)
 - C7REG12 : Région (1-13, voir champs 7, REG12)
- 69 : Question C8 : Combien de pièces possède le logement dans lequel vous vivez cette période de confinement ? (SANS COMPTER la salle de bain et toilettes)
 - Entier
 - Vide : Ne sais pas
- 70-72 : Question C9 : Combien de personnes vivent actuellement dans ce logement (y compris vous-même) ?
 - Colonnes:
 - C9_QA : Nombre total de personnes
 - C9_QB : Nombre de personnes de moins de 18 ans
 - C9_QC : Nombre de personnes de plus de 70 ans
 - Valeurs possibles:
 - Entier
- 73 : Question C10 : Quel est le diplôme le plus élevé que vous ayez obtenu ?
 - 1: Sans diplôme
 - 2: Certificat de formation générale (CFG), Certificat d'études primaires, Diplôme national du brevet (BEPC ou Brevet des Collèges)
 - 3: Certificat d'aptitude professionnelle (CAP), Brevet d'études professionnelles (BEP)
 - 4: Baccalauréat ou diplôme équivalent
 - 5: Bac +1 ou 2
 - 6: Bac +3 ou plus
 - 7: Autre, précisez
 - 8: Non réponse
- 74 : ALEA : Aspect aléatoire de l'ordre des questions. Valeurs possibles:
 - 1: D1, D2, D3 en premier ;
 - 2: D4, D5, D6 en premier.
- 75 : Question D1 : Certaines applications mobiles peuvent aider à ralentir la circulation de virus. Les utilisateurs se protègeraient et protègeraient les autres en fournissant de façon anonyme des



Baromètre COVID 19

Un projet de science citoyenne

24.04.2020

informations sur leurs déplacements et leurs contacts, ce qui permettrait de prévenir et éventuellement de tester et traiter les personnes potentiellement exposées. L'application n'accèderait pas à vos données personnelles ni à vos coordonnées GPS, et les informations collectées seraient anonymes et seraient effacées dans un délai de quelques semaines.

Pour mieux connaître et prévenir les risques dans le cadre de la sortie du confinement, accepteriez-vous d'utiliser une application installée sur votre téléphone portable ? Valeurs possibles :

- 1: Oui, certainement
 - 2: Oui, probablement
 - 3: Non, probablement pas
 - 4: Non, certainement pas
 - 5: Je ne sais pas
- 76 : Question D2 : Une telle installation se ferait dans le respect Règlement général sur la protection des données européen sur le traitement des données personnelles (RGPD). L'application n'accèderait pas à vos données personnelles ni à vos coordonnées GPS, et les données collectées seraient anonymes et effacées dans un délai de quelques semaines. Vous auriez la possibilité de désinstaller l'application si vous ne désirez pas l'utiliser.

Pensez-vous que ce serait une bonne chose que le gouvernement demande aux opérateurs mobiles d'installer une telle application automatiquement sur votre téléphone portable ?

Valeurs possibles

- 1: Tout à fait favorable
 - 2: Plutôt favorable
 - 3: Plutôt défavorable
 - 4: Tout à fait opposé
 - 5: Je ne sais pas
- 77 : Question D3 : Personnellement, désinstalleriez-vous une telle application ? Valeurs possibles :
 - 1: Oui
 - 2: Non
 - 3: Je ne sais pas
- 78-80 : Question D4 : Accepteriez-vous de communiquer les éléments suivants à des chercheurs qui vous le demanderaient pour s'en servir uniquement pour lutter contre le coronavirus et qui s'engageraient à ne jamais les divulguer ou les revendre ?
 - Colonnes :
 - D4_0 : Des informations issues de votre dossier médical (maladies passées et présentes, traitements suivis, facteurs de risque, etc.)
 - D4_1 : Des résultats de prélèvements pas ou peu invasifs que vous avez déjà réalisés (salive, urine, selles, autotest sanguin, prélèvement oral ou nasal, etc.)
 - D4_2 : Des résultats de prélèvements plus invasifs que vous avez déjà réalisés (prise de sang, biopsie, etc.)
 - Valeurs possibles :
 - 1: Oui
 - 2: Non
 - 3: Je ne sais pas suffisamment ce que cela signifie ou implique
- 81 : Question D5 : La connaissance précise du nombre total de personnes immunisées dans la population est essentielle dans la lutte contre l'épidémie, ce que permettrait de nouveaux tests rapides.

Si on vous le proposait, accepteriez-vous de recevoir gratuitement un test à pratiquer à la maison qui vous permettrait de savoir en quelques minutes si vous avez contracté le coronavirus et de nous communiquer le résultat de façon anonyme ? Valeurs possibles :

- 1: Oui
 - 2: Non
 - 3: Je ne sais pas suffisamment ce que cela signifie ou implique
- 82 : Question D6 : Selon l'Organisation Mondiale de la Santé (OMS), la mise en place de recherches à partir de la population générale est particulièrement importante pour mieux comprendre et lutter contre le coronavirus.

Même dans le cas où vous êtes en bonne santé, accepteriez-vous de participer à un groupe de personnes qui est suivi dans la durée (panel) supportant ces recherches pour mieux comprendre et lutter contre le coronavirus ? Valeurs possibles :

- 1: Oui
 - 2: Non
 - 3: Je ne sais pas suffisamment ce que cela signifie ou implique



API « REST »

Les données peuvent également être interrogées en ligne à travers l'API REST accessible depuis l'URL <http://datacovid.org/api>. Les réponses de l'API sont au format JSON, de la forme suivante :

```
{  
    "statusCode": 200,  
    "license": "Licence Ouverte (https://www.etalab.gouv.fr/licence-ouverte-open-licence)",  
    "disclaimer": "...",  
    "data": {  
        ...  
    }  
}
```

L'api dispose de 2 principaux points d'entrées : 1) Interrogation, par vague, de l'ensemble des réponses, et 2) Interrogation des réponses individuelles.

Token d'accès à l'API

La création d'un token d'accès est un prérequis à l'interrogation des données. Pour créer un token d'accès, il suffit de se rendre sur <http://datacovid.org/data> et remplir le formulaire de demande de token.

Un mail contenant le nouveau token est ensuite envoyé.

Interrogation de l'ensemble des réponses

Le point d'entrée pour interroger l'ensemble des réponses par vague est : <http://datacovid.org/api/answers/<V>/<C>?token=<token>>

Avec :

- V= le numéro de la vague
- C= les champs que l'on veut interroger (voir ci-dessous pour la description de l'ensemble des champs), sous la forme champs1;champs2;champs3;etc.
- token= la clé permettant l'interrogation de l'API

Note : Il n'est possible d'interroger que 10 champs par requête. Les champs à partir du 11^{ème} ne seront pas pris en compte.

Exemple :

<http://datacovid.org/api/answers/1/age;diplome?token=<token>>

Retourne l'âge et le diplôme des 5000 participants de la vague 1:

```
{  
    "statusCode": 200,  
    "license": "Licence Ouverte (https://www.etalab.gouv.fr/licence-ouverte-open-licence)",  
    "data": {  
        "age": [  
            "7",  
            "1",  
            "7",  
            "2",  
            "4",  
            "1",  
            "1",  
            "1",  
            "1",  
            "1"  
        ]  
    }  
}
```



```
"1",
...
"diplome": [
    "4",
    "6",
    "4",
    "6",
    "4",
    ...
]
}
```

Interrogation des réponses individuelles

Il est possible d'interroger les réponses individuellement, en utilisant le point d'entrée suivant :

<http://datacovid.org/api/answers/<V>/id/<I>?token=<token>>

Avec :

- V : Le numéro de la vague ;
- I : L'identifiant de la réponse (de 1 à 5000 par exemple).
- token: la clé permettant d'interroger l'API

Exemple :

<http://datacovid.org/api/answers/1/id/1?token=sdklsjk>

Retourne la première réponse de la vague 1, avec l'ensemble des champs :

```
{
    "statusCode": 200,
    "license": "Licence Ouverte (https://www.etalab.gouv.fr/licence-ouverte-open-licence)",
    "data": {
        "vague": 1,
        "serial": "1",
        "jour": "1",
        "device": "1",
        "sexe": "1",
        "age": "7",
        "departement": "24",
        ...
    }
}
```

Description des champs

Chaque réponse possède les champs suivants (entre parenthèse, le numéro du champs correspondant dans le fichier TSV pour lequel les valeurs possibles sont décrites):

- 1) "vague": Le numéro de la vague dont provient la réponse
- 2) "serial": L'identifiant de la réponse au sein de la vague (1 à 5000 typiquement)
- 3) "jour": Le jour de la réponse (2),
- 4) "device": Type d'appareil utilisé (3),
- 5) "sexe": Genre du participant (4),
- 6) "age": Catégorie d'âge du participant (5),
- 7) "departement": Département (6),



Baromètre COVID 19

Un projet de science citoyenne

24.04.2020

- 8) "region": Région (7),
- 9) "agglo": Catégorie d'agglomération (8),
- 10) "emploi": Situation professionnelle (9),
- 11) "pcsi": Catégorie socio-professionnelle (10),
- 12) "famille": Situation familiale (11),
- 13) "preoccupation1": Question A0, Préoccupation 1 (12),
- 14) "preoccupation2": Question A0, Préoccupation 2 (13),
- 15) "preoccupation3": Question A0, Préoccupation 3 (14),
- 16) "penseInfecte": Question A1, le participant pense avoir été infecté (15),
- 17) "penseInfecteCause": Question A2, pourquoi le participant pense avoir été infecté (16),
- 18) "contactInfecteeDansFoyer": Question A3, en contact avec des personne infectée DANS le foyer (17). 0 :non ou 1 : oui
- 19) "contactInfecteeEnDehorsFoyer": Question A3, en contact avec des personne infectée EN DEHORS du foyer (18). 0 : non, 1 : oui,
- 20) "fievre": Question A4, Fièvre au cours des 48 dernières heures (19),
- 21) "maxFievre": Question A5, température max si fièvre dans les dernières 48h (20),
- 22) "toux": Question Question A6_0, Toux ou augmentation de la toux (21),
- 23) "anosmie": Question A6_1, Diminution du goût ou de l'odorat (22),
- 24) "malGorge": Question A6_2, Mal de gorge, douleurs musculaires, courbatures (23),
- 25) "fatigue": Question A6_3, Fatigue inhabituelle (24),
- 26) "mauxTete": Question A6_4, Maux de tête inhabituels (25),
- 27) "pasAlimentation": Question A7_0, Impossibilité de s'alimenter ou de boire (26),
- 28) "manqueSouffle": Question A7_1, Manque de souffle inhabituel (27),
- 29) "diarrhee": Question A7_2, Diarrhées (28),
- 30) "perteEquilibre": Question A7_3, Pertes d'équilibre inhabituelles (29),
- 31) "conjonctivite": Question A7_4, Conjonctivite à un niveau inhabituel (30),
- 32) "graviteEpidemie": Question B1, Évaluation de la gravité de l'épidémie (31),
- 33) "lavageMain": Question B2_0, Gestes barrières : Lavage des mains (32),
- 34) "gelHydroAlcoolique": Question B2_1, Gestes barrières : Utilisation de gel hydroalcoolique (33),
- 35) "tousserCoude": Question B2_2, Gestes barrières : Tousser dans son coude ou un mouchoir (34),
- 36) "mouchoireUnique": Question B2_3, Gestes barrières : Lavage des mains (35),
- 37) "distance1M": Question B2_4, Gestes barrières : Lavage des mains (36),
- 38) "eviteSerrerMainEmbrasser": Question B2_5, Gestes barrières : Ne pas serrer la main ni embrasser (37),
- 39) "eviteRegroupements": Question B2_6, Gestes barrières : Eviter les regroupements (38),
- 40) "resterChezSoi": Question B2_7, Gestes barrières : Rester chez soi (39),
- 41) "porterMasque": Question B2_8, Gestes barrières : Porter un masque (40),
- 42) "sortiesEn24h": Question B3, Temps de sortie dans les 24h (41),
- 43) "sortiesEn24hAchats": Question B4, Sortie en B3 pour des achats (42-46). 0 : Non, 1 : Oui,
- 44) "sortiesEn24hSport": Question B4, Sortie en B3 pour ballade ou une activité sportive (42-46). 0 : Non, 1 : Oui,
- 45) "sortiesEn24hMedical": Question B4, Sortie en B3 pour consultation médicale ou motif familial ou convocation judiciaire (42-46). 0 : Non, 1 : Oui,
- 46) "sortiesEn24hTravail": Question B4, Sortie en B3 pour se rendre au travail (42-46). 0 : Non, 1 : Oui,
- 47) "sortiesEn24hAutre": Question B4, Sortie en B3 pour une autre raison (42-46). 0 : Non, 1 : Oui,
- 48) "sortiesEn24hNA": Question B4, Non réponse (42-46). 0 : Non, 1 : Oui,
- 49) "nbContact24h": Question B5, Nombre de contacts proches en dehors des personnes du foyer (47)
- 50) "imc": Questions C1 et C2, Indice de masse corporel (48)
- 51) "enceinte": Question C3, la participante est-elle enceinte ? (49)
- 52) "diabete": Question C4, Problèmes de santé, diabète (50),
- 53) "cancer": Question C4, Problèmes de santé, cancer (51),
- 54) "maladieRespiratoire": Question C4, Problèmes de santé, maladie respiratoire (52),
- 55) "insuffisanceRenale": Question C4, Problèmes de santé, insuffisance rénale chronique dialysée (53),
- 56) "maladieChroniqueFoie": Question C4, Problèmes de santé, maladie chronique du foie (54),
- 57) "hypertensionArterielle": Question C4, Problèmes de santé, hypertension, maladie cardiaque ou vasculaire nécessitant un traitement (55),
- 58) "maladieImmunitaire": Question C4, Problèmes de santé, maladie diminuant les défenses immunitaires (56),
- 59) "traitementImmunosupresseur": Question C4, Problèmes de santé, maladie nécessitant un traitement immunosupresseur (57),
- 60) "frequenceTriste": Question C5, fréquence d'état de tristesse, de dépression dans les 2 dernières semaines (58),
- 61) "frequenceNervosite": Question C5, fréquence d'un sentiment de nervosité, anxiété, tension dans les 2 dernières semaines (59),
- 62) "frequenceInquietude": Question C5, fréquence d'état d'incapacité à arrêter de s'inquiéter dans les 2 dernières semaines (60),
- 63) "frequenceDesinteret": Question C5, fréquence de perte d'intérêt ou de plaisir à faire les choses (61),



Baromètre COVID 19

Un projet de science citoyenne

24.04.2020

-
- 64) "travailHorsDomicile": Question C6, Situation professionnelle actuelle, travail hors domicile (62-65), 0 : non, 1: oui,
 - 65) "travailDomicile": Question C6, Situation professionnelle actuelle, travail habituel à domicile (62-65), 0 : non, 1,oui,
 - 66) "travailADistance": Question C6, Situation professionnelle actuelle, travail à distance, télétravail (62-65), 0 : non, 1 : oui,
 - 67) "travailArretMaladie": Question C6, Situation professionnelle actuelle, arrêt maladie (62-65), 0 : non, 1 : oui,
 - 68) "travailChomagePartiel": Question C6, Situation professionnelle actuelle, chômage partiel (62-65), 0 : non, 1 : oui,
 - 69) "travailConge": Question C6, Situation professionnelle, congés (62-65), 0 : non, 1 : oui,
 - 70) "travailNA": Question C6, Situation professionnelle, Pas de réponse (62-65), 0 : non, 1 : oui,
 - 71) "confinementResidenceHabituelle": Question C7, résidez-vous actuellement dans votre résidence habituelle (66),
 - 72) "confinementDepartement": Question C7, Département de la résidence de confinement (67),
 - 73) "confinementRegion": Question C7, Région de la résidence de confinement (68),
 - 74) "confinementPieces": Question C8, Nombre de pièces dans la résidence de confinement (69),
 - 75) "confinementNbPersonnes": Question C9, Nombre de personnes total dans la résidence de confinement (70),
 - 76) "confinementNbPersonnesMineures": Question C9, Nombre de personnes >18 ans dans la résidence de confinement (71),
 - 77) "confinementNbPersonnesPlus70": Question C9, Nombre de personnes >70 ans dans la résidence de confinement (72),
 - 78) "diplome": Question C10, Diplôme le plus élevé obtenu (73),
 - 79) "alea": ALEA, aspect aléatoire de l'ordre des questions (74)
 - 80) "okAppTelephone": Question D1, D'accord pour utiliser une application mobile (75)
 - 81) "okAppTelephoneGouvernement": Question D2, D'accord pour que le gouvernement demande aux opérateurs mobiles d'installer une application automatiquement (76),
 - 82) "appTelephoneDesinstall": Question D3, Désinstallerait l'application (77)
 - 83) "okDossierMedical": Question D4, Communiquer des informations médicales à des chercheurs sur COVID19 (78)
 - 84) "okResultatsPrelevements": Question D4, Communiquer des résultats de tests peu invasifs à des chercheurs sur COVID19 (79)
 - 85) "okResultatsPrelevementsInvasifs": Question D4, Communiquer des résultats de tests plus invasifs à des chercheurs sur COVID19 (80)
 - 86) "okTestMaison": Question D5, D'accord pour recevoir un test [sérologique] gratuit à la maison et communiquer les résultats (81),
 - 87) "okparticipationPanel": Question D6, D'accord pour participer à un panel supportant ces recherche (82).



Baromètre COVID 19

Un projet de science citoyenne

24.04.2020

Annexe 1 : Liste des numéros des départements

- | | |
|----------------------------|---------------------------|
| 1: Ain | 51: Marne |
| 2: Aisne | 52: Haute-Marne |
| 3: Allier | 53: Mayenne |
| 4: Alpes-de-Haute-Provence | 54: Meurthe-et-Moselle |
| 5: Hautes-Alpes | 55: Meuse |
| 6: Alpes-Maritimes | 56: Morbihan |
| 7: Ardèche | 57: Moselle |
| 8: Ardennes | 58: Nièvre |
| 9: Ariège | 59: Nord |
| 10: Aube | 60: Oise |
| 11: Aude | 61: Orne |
| 12: Aveyron | 62: Pas-de-Calais |
| 13: Bouches-du-Rhône | 63: Puy-de-Dôme |
| 14: Calvados | 64: Pyrénées-Atlantiques |
| 15: Cantal | 65: Hautes-Pyrénées |
| 16: Charente | 66: Pyrénées-Orientales |
| 17: Charente-Maritime | 67: Bas-Rhin |
| 18: Cher | 68: Haut-Rhin |
| 19: Corrèze | 69: Rhône |
| 20: Corse-du-Sud | 70: Haute-Saône |
| 21: Côte-d'Or | 71: Saône-et-Loire |
| 22: Côtes-d'Armor | 72: Sarthe |
| 23: Creuse | 73: Savoie |
| 24: Dordogne | 74: Haute-Savoie |
| 25: Doubs | 75: Paris |
| 26: Drôme | 76: Seine-Maritime |
| 27: Eure | 77: Seine-et-Marne |
| 28: Eure-et-Loir | 78: Yvelines |
| 29: Finistère | 79: Deux-Sèvres |
| 30: Gard | 80: Somme |
| 31: Haute-Garonne | 81: Tarn |
| 32: Gers | 82: Tarn-et-Garonne |
| 33: Gironde | 83: Var |
| 34: Hérault | 84: Vaucluse |
| 35: Ille-et-Vilaine | 85: Vendée |
| 36: Indre | 86: Vienne |
| 37: Indre-et-Loire | 87: Haute-Vienne |
| 38: Isère | 88: Vosges |
| 39: Jura | 89: Yonne |
| 40: Landes | 90: Territoire de Belfort |
| 41: Loir-et-Cher | 91: Essonne |
| 42: Loire | 92: Hauts-de-Seine |
| 43: Haute-Loire | 93: Seine-Saint-Denis |
| 44: Loire-Atlantique | 94: Val-de-Marne |
| 45: Loiret | 95: Val-d'Oise |
| 46: Lot | 96: Haute-Corse |
| 47: Lot-et-Garonne | |
| 48: Lozère | |
| 49: Maine-et-Loire | |
| 50: Manche | |

