

Meyssa BEDDAR  
Félita DONOR

# DÉPRESSION ET TABAGISME

2020-2021

*Dossier présenté dans le cadre de l'UE Statistiques – Medas 1*

# TABLE DES MATIÈRES

## Contenu

Introduction	1
Données	2
Méthode	5
Statistiques descriptives	6
Etude des liaisons	18
Effet du tabagisme sur la dépression	20
Effet de la dépression sur le tabagisme	28
Conclusion	36
Bibliographie	37
Annexes	39

# EFFET DE LA DÉPRESSION SUR LE TABAGISME

## Introduction

Selon Salt et Osborne (2020), le taux de tabagisme est 50% plus élevé parmi les personnes souffrant de problèmes de santé mentale que dans la population générale et d'autant plus lorsqu'elles vivent sous le seuil de pauvreté.

Effectivement, les problèmes de santé mentale augmenteraient le risque de tabagisme, qui, à son tour, augmenterait encore davantage les risques de problème de santé mentale tels que la dépression, l'anxiété, les symptômes liés aux traumatismes ou encore les risques suicidaires (Lee, 2021), principalement chez les jeunes adultes (Mojtabai et Crum, 2013).

Il existerait toutefois des différences selon le sexe, les relations entre le tabagisme et la santé mentale (trouble dépressif majeur, anxiété, trouble panique et déficiences graves) étant plus prononcées parmi les femmes que chez les hommes (Richardson et al, 2012).

Aussi, cette population présenterait plus de difficulté à maintenir l'abstinence, bien que le désir d'arrêter ne soit pas moindre. Ainsi, selon Lee (2021), les personnes ayant des problèmes de santé mentale constituerait une sous population à haut risque, et ce, tout groupe d'âge confondu.

Patten et al. (2001) avaient par ailleurs montré que les fumeurs éprouvaient des niveaux plus élevés de fatigue et d'insomnie que les non-fumeurs.

Cesser de fumer permettrait alors aux fumeurs de diminuer l'anxiété, le stress et la dépression et ainsi d'améliorer l'humeur positive ainsi que la qualité de vie au même titre que les antidépresseurs (Salt et Osborne, 2020).

Toutefois, malgré les liens précédemment explicités, les études concernant la direction de la relation entre le tabagisme et la dépression produisent des résultats contradictoires. (Richardson, 2012).

L'objectif de la présente étude sera donc d'évaluer le sens de cette relation à partir des données du BRFSS (Behavioral Risk Factor Surveillance System) qui est le premier système américain « d'enquêtes téléphoniques liées à la santé qui collectent des données sur les résidents américains concernant leurs comportements à risque liés à la santé, leurs problèmes de santé chroniques et l'utilisation des services de prévention ». Effectivement, cette enquête fournit des informations concernant le tabagisme des répondants ainsi que sur leur santé mentale avec notamment une question sur la dépression.

Ainsi, dans un premier temps, nous définirons les données utilisées et la méthodologie employée pour cette étude. Nous présenterons ensuite les principaux résultats des statistiques descriptives ainsi que les liaisons entre les différentes variables, pour enfin évaluer le lien causal entre tabagisme et dépression à l'aide de régression logistique après avoir effectué deux méthodes utilisant les scores de propension sur l'échantillon.

# EFFET DE LA DÉPRESSION SUR LE TABAGISME

## Données

### SOURCES ET ÉCHANTILLON

#### Sources

Le BRFSS (Behavioral Risk Factor Surveillance System) est un système de sondage développé par le CDC (Centers for Disease Control and Prevention). Le CDC est un ensemble de centres formant l'agence fédérale américaine de santé publique. Elle fait partie du département de la Santé et des Services Sociaux, et à ce titre, produit des informations afin d'aider à la prise de décision en matière de santé publique. Le BRFSS a été initié en 1984 et est toujours en cours.

Les données ainsi que les codebooks sont accessibles à ce lien : [https://www.cdc.gov/brfss/annual\\_data/annual\\_data.htm](https://www.cdc.gov/brfss/annual_data/annual_data.htm).

Un autre jeu de donnée sur le prix du tabac a été utilisé dans les séries temporelles mais n'est que descriptif et n'a pas été pris en compte dans les modèles.

#### Mode de recueil des données

Les sondages sont conduits selon une technique de randomisation, le RDD (Random Digit Dialing), par téléphone fixe ainsi que par téléphone portable depuis 2009.

Le questionnaire est composé d'une partie commune et standardisée et de questions optionnelle. Chaque Etat peut également ajouter des questions plus spécifiques, par exemple sur l'obésité, l'exercice ou la sécurité en voiture

#### Définition de l'échantillon

Les répondants sont des citoyens américains de plus de 18 ans. Chaque ligne correspond donc à un individu.

Etant donné la taille de l'échantillon et la technique de randomisation, les données sont généralisables à l'ensemble de la population américaine.

#### Taille

Au total, le nombre d'adultes interrogés chaque année approche les 400 000.

Leur nombre varie d'Etat en Etat en fonction des fonds disponibles pour conduire les interviews.

# EFFET DE LA DÉPRESSION SUR LE TABAGISME

## VARIABLES

### Définition des variables expliquées et explicatives

L'objectif du BRFSS est de recueillir des informations sur les six premières causes de mortalité prématurée et de morbidité : le tabac, l'alcool, l'activité physique, l'alimentation, l'hypertension et le port de la ceinture. Le questionnaire est découpé en une partie principale, des modules optionnels et une partie variables calculées, puis chaque partie découpée en sections.

Nous avons choisi de sélectionner une trentaine de variables parmi les 300 de tout le questionnaire.

Elles ont été regroupées en 3 grands thèmes : socio-démographie, tabac et santé mentale. Ceux-ci sont représentés par 3 couleurs différentes dans la Google Sheet du Dictionnaire des Données.

Ce dernier détaille le nom de la variable utilisé chaque année à partir de 2013 (celui change dans de rares cas) ainsi que le rang de valeurs possibles et leurs descriptions.

Nos variables à expliquer sont le tabac (en jaune) et les variables explicatives sont les facteurs socio-démographiques (bleu) et la santé mentale (vert).

## DÉFIS RENCONTRÉS

### Import

Il y a un fichier disponible par année de 1984 à 2020. Le nombre et parfois le nom des variables varie d'année en année. Pour cela, il fallait tester chaque variable pour savoir si elle apparaissait ou avait disparu.

Une fois les colonnes identifiées, il a été possible de faire des jointures, en plusieurs fois pour éviter les erreurs dues à une mémoire insuffisante.

### Nettoyage

Deux nettoyages ont été faits, un avec facteurs, c'est-à-dire en gardant les valeurs d'origine, et un sans facteurs, en remplaçant l'ordinalité par une description. Par exemple, 1 pour "Très Satisfait", 2 pour "Satisfait", 3 "Insatisfait", etc.

Des règles concernant les NA ont été appliquées : elles remplacent les réponses du type "Ne sais pas" ou "N'a pas répondu".

# EFFET DE LA DÉPRESSION SUR LE TABAGISME

## Nettoyage des Time Series

Des problèmes ont été rencontrés avec les différentes librairies de séries temporelles. Certaines fonctions ne fonctionnent qu'avec un format ts ou bien xts.

De plus, certaines variables comportent des valeurs manquantes pour des années entières. Dans ce cas, une interpolation a été faite mais celle-ci ne représente pas la variabilité et est donc linéaire.

## Mise en forme

Les noms des variables et du contenu tel que les dates ont été uniformisés afin de constituer un seul tableau final regroupant toutes les variables de 1984 à aujourd'hui.

## Mise en forme Shiny

L'utilisation de flexdashboard avait pour but de visualiser en même temps plusieurs graphiques et ainsi aider à la compréhension des différentes étapes. Cela avait été réalisé dans l'optique que la projection soit sur un écran suffisamment grand tel celui d'un amphithéâtre.

## Développement

Etant donné le volume de données final (8 millions de lignes, 30 variables), le chargement, le nettoyage et les analyses ont été faits dans des modules séparés. Chaque module produit un csv ou un RData qui est ensuite réutilisé par le module suivant.

Cela a évité les erreurs et facilité le fait de pouvoir exécuter et vérifier l'output d'une seule partie du code.

Globalement, la taille des fichiers et le nombre de variables a rendu cette étape assez longue bien qu'elle ait pu être compensée par la propreté des jeux de données originaux.

## Méthode

Afin de mieux appréhender les éventuelles relations existant entre les variables et ainsi avoir une première idée des covariables à contrôler pour notre étude, nous avons d'abord effectué des statistiques descriptives ainsi qu'une analyse factorielle.

Ensuite, à partir de ces premières informations et de l'étude des corrélations, liaisons et dépendances entre variables, nous avons pu déterminer les facteurs que nous souhaitons contrôler.

Nous avons par la suite étudié l'effet du tabagisme (VI) sur la dépression (VD) à l'aide d'une régression logistique après avoir effectué deux méthodes s'appuyant sur les scores de propension (une correspondance exacte puis une pondération). L'objectif était, à partir de ces deux méthodes, d'une part de limiter l'influence des covariables, et, d'autre part, de s'assurer de la robustesse de notre modèle par la comparaison des résultats obtenus.

Enfin, de façon similaire, nous avons étudié l'effet inverse soit l'impact de la dépression sur le tabagisme.

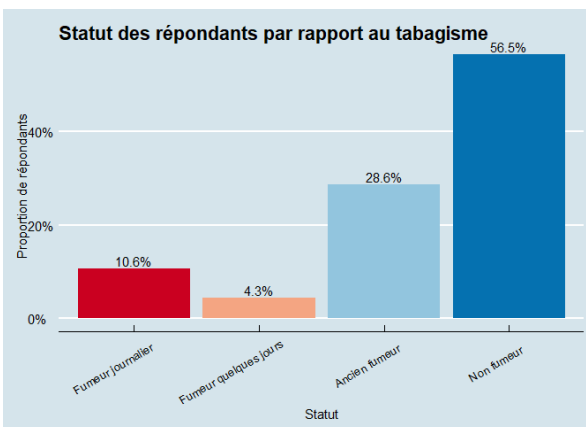
# EFFET DE LA DÉPRESSION SUR LE TABAGISME

## Statistiques descriptives

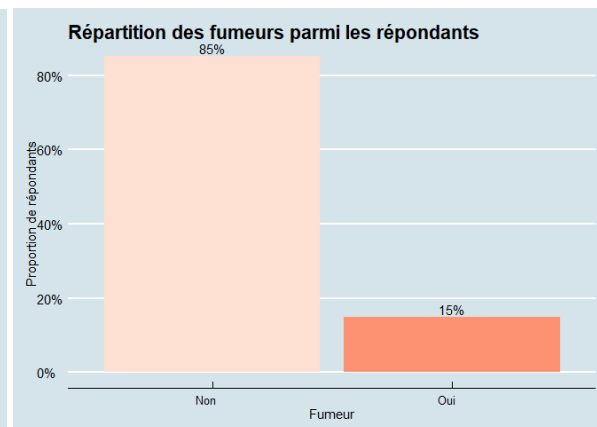
L'objectif des statistiques descriptives est non seulement de connaître la répartition de l'échantillon au regard du tabagisme et de la santé mentale, mais également de voir les éventuels liens existant entre tabagisme, santé mentale et socio-démographie.

### TABAGISME

Graphique 1 :



Graphique 2 :



On remarque qu'environ 56% des répondants n'ont jamais fumé, 29% sont d'anciens fumeurs et 15% fument toujours à l'heure actuelle (10.6% de manière quotidienne et 4.3% de manière moins récurrente). Ainsi, approximativement, 85% des répondants sont non fumeurs et 15% sont fumeurs.

### SANTÉ MENTALE

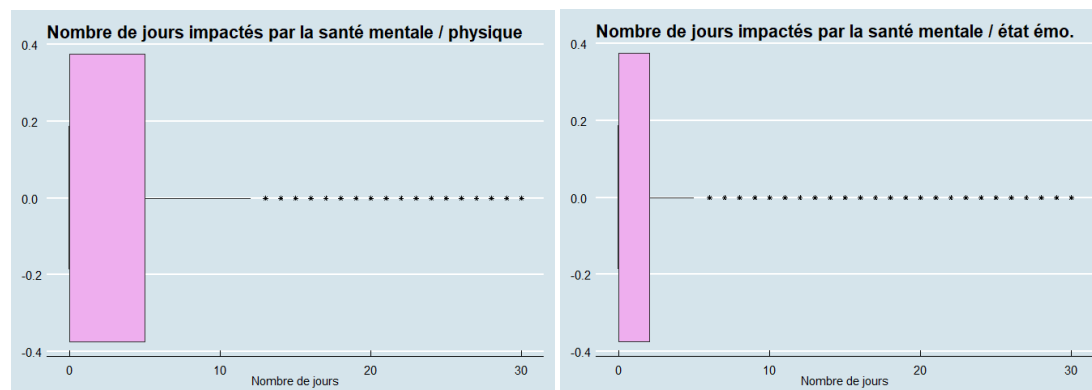
1. Mauvaise santé mentale et/ou physique et mauvaise santé mentale

Graphique 3 :

Graphique 4 :



# EFFET DE LA DÉPRESSION SUR LE TABAGISME

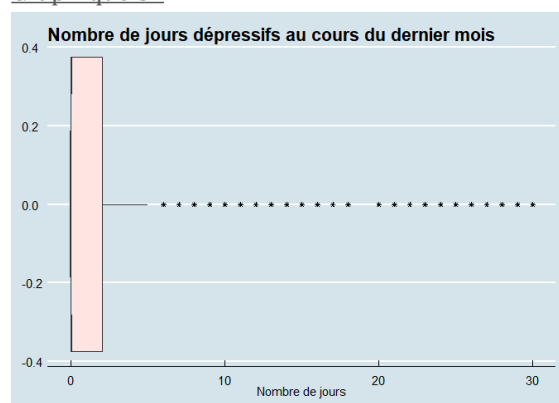


Une grande majorité (75%) des répondant indique ne pas être impactée par leur santé mentale ou physique plus de 5 jours par mois. Un répondant sur 4 indique toutefois être impacté entre 5 et 12 jours par mois par sa santé mentale ou physique, ce qui est non négligeable.

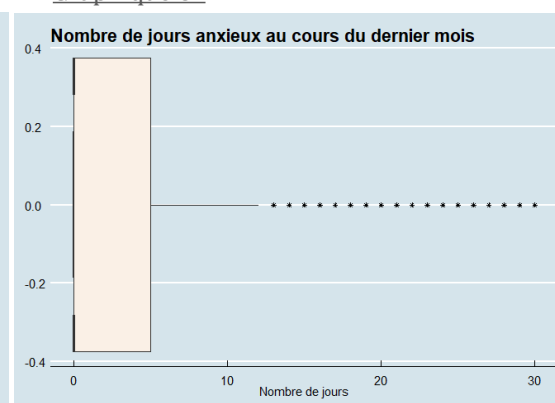
Au vu du second graphique, il semblerait que la santé mentale pèse presque autant que la santé physique pour les répondants. Effectivement, ici, 75% des répondants indique ne pas être impacté négativement plus de 2 ou 3 jours par mois par leur santé mentale ou état émotionnel, contre 25% qui le seraient entre 2/3 à 5 jours par mois.

## 2. Humeurs anxieuse et dépressive

Graphique 5 :



Graphique 6 :



Au vu du graphique 5, on peut se demander si ce ne sont pas les humeurs dépressives dont il est question dans la question relative à la mauvaise santé mentale étant donné les résultats similaires entre ces deux questions.

Etonnament, on observe un nombre de jours impactés par l'anxiété plus important que le nombre de jours impactés par la santé mentale ou l'état émotionnel. Effectivement, 75% des répondants indique souffrir

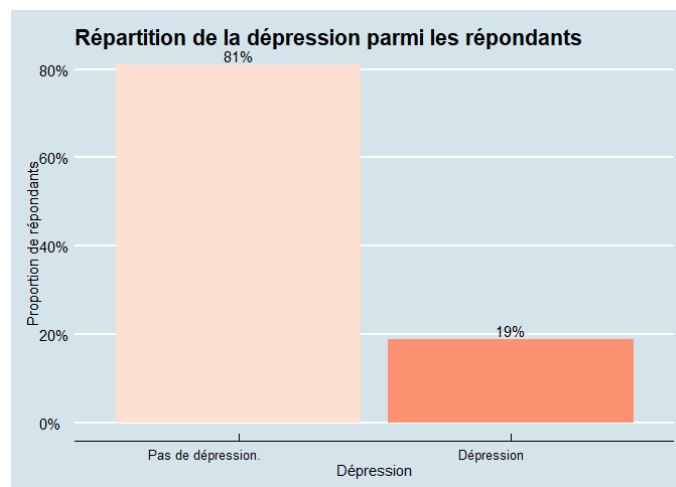
# EFFET DE LA DÉPRESSION SUR LE TABAGISME

d'anxiété jusqu'à 5 jours par mois et les 25% restants entre 5 et 12 jours par mois.

Ce sont par ailleurs les mêmes résultats que ceux de la question concernant l'impact de la santé physique et/ou mentale.

## 3. Dépression

Graphique 7 :

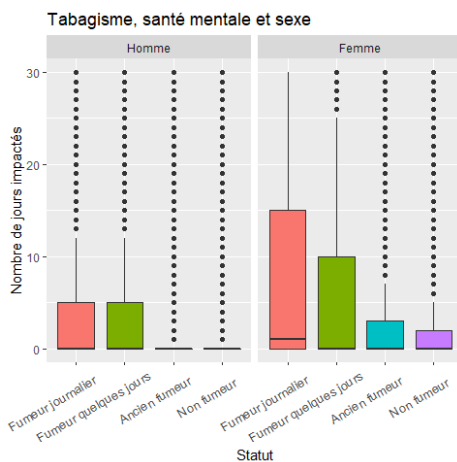


On observe que près de 1 répondant sur 5 indique avoir eu un diagnostic de dépression.

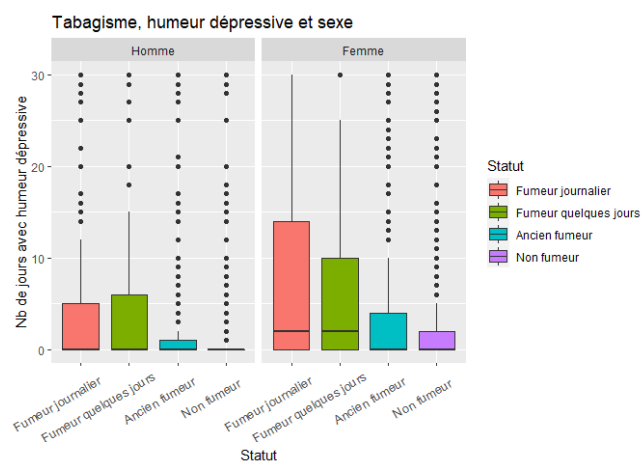
## SANTÉ MENTALE, TABAGISME ET SOCIO-DÉMOGRAPHIE

### 1. Tabagisme, santé mentale et sexe

Graphique 8 :



Graphique 9 :

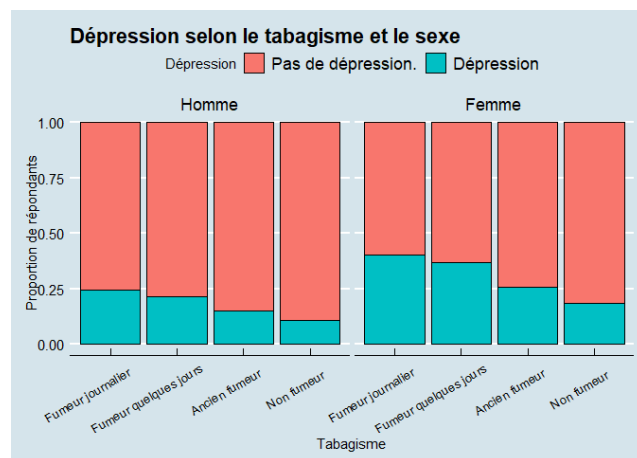


# EFFET DE LA DÉPRESSION SUR LE TABAGISME

On observe de nettes différences en terme de santé mentale selon que les répondants soient fumeurs ou non fumeurs. Aussi, les femmes déclarent plus de jours impactés négativement par leur santé mentale que les hommes.

Si l'humeur dépressive semble varier selon le tabagisme, elle diffère également selon le sexe, les femmes déclarant davantage de jours dépressifs que les hommes.

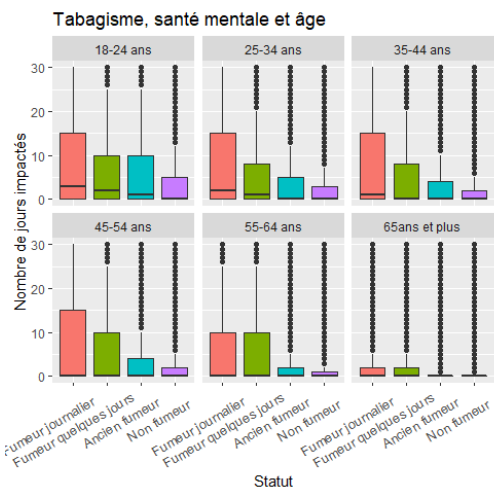
Graphique 10 :



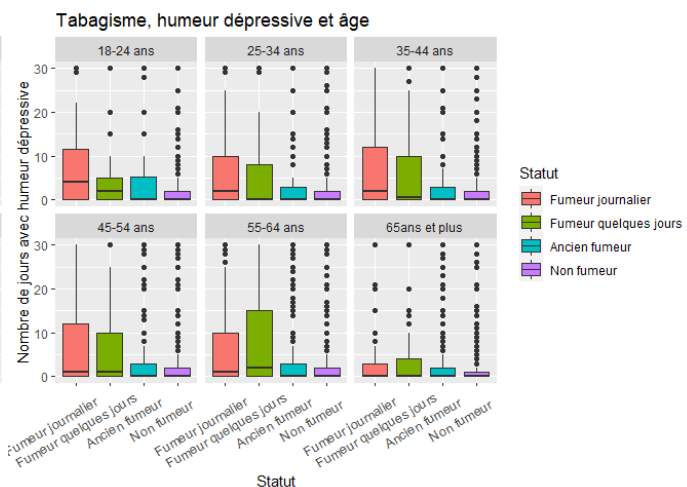
La dépression semble davantage présente auprès des femmes fumeuses. Toutefois, quel que soit le sexe, on observe une variation de la prévalence de la dépression selon le statut par rapport au tabac.

## 2. Tabagisme, santé mentale et âge

Graphique 11 :



Graphique 12 :

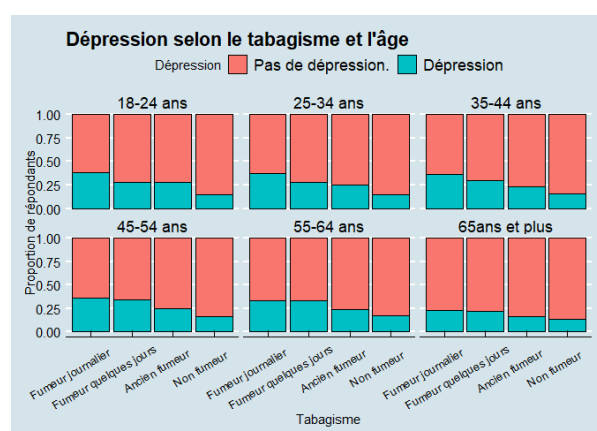


# EFFET DE LA DÉPRESSION SUR LE TABAGISME

On remarque notamment une augmentation de jours impactés par la santé mentale avec la jeunesse. Effectivement, plus les répondants sont jeunes, plus ils déclarent un nombre de jours par mois impacté par leur santé mentale élevé. Aussi, on remarque une distinction fumeur / non fumeur quelque soient les âges.

Une nette distinction est également visible selon le fait d'être fumeur ou non fumeur concernant l'humeur dépressive. On remarque cependant que les répondants de 65 ans et plus déclarent moins de jours dépressifs que les répondants de moins de 65 ans.

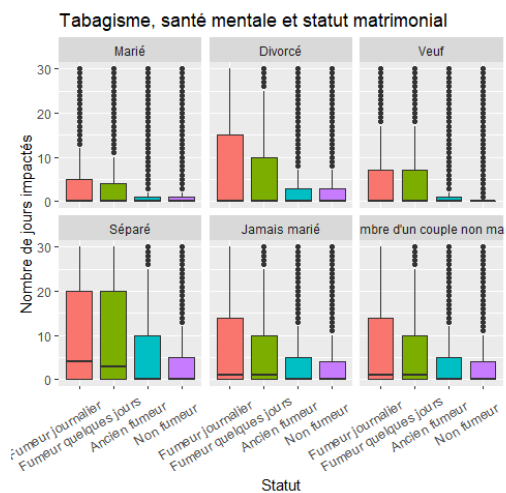
Graphique 13 :



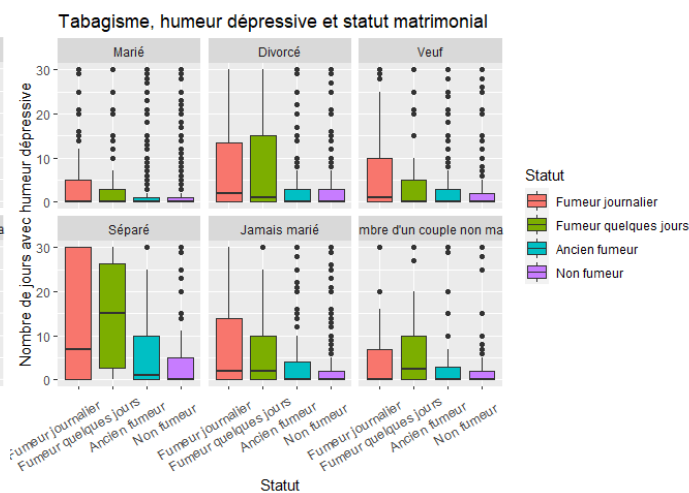
Il semblerait que la prévalence de la dépression soit inférieure auprès des répondants de 65 ans et plus, ce qui coïncide avec le graphique précédent. Aussi, quelque soit l'âge du répondant, on observe une distinction selon le fait d'être fumeur ou non fumeur.

## 3. Tabagisme, santé mentale et statut matrimonial

Graphique 14 :



Graphique 15 :

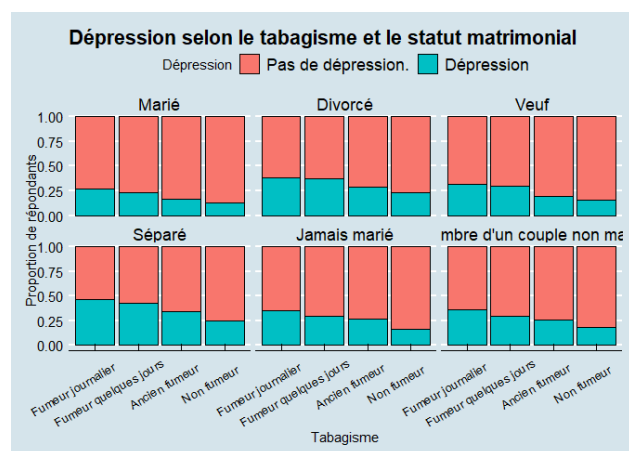


# EFFET DE LA DÉPRESSION SUR LE TABAGISME

On remarque surtout que les fumeurs séparés semblent plus impactés par leur santé mentale. Aussi, le fait d'être marié pourrait être un facteur protecteur étant donné le nombre de jours impacté plus faible.

L'humeur dépressive semble davantage présente chez les répondants divorcés, séparés et jamais mariés. Aussi, quels que soient le statut matrimonial, on observe une différence d'humeur dépressive selon le fait d'être fumeur ou non fumeur, bien qu'elle soit plus marquée auprès des séparés et divorcés.

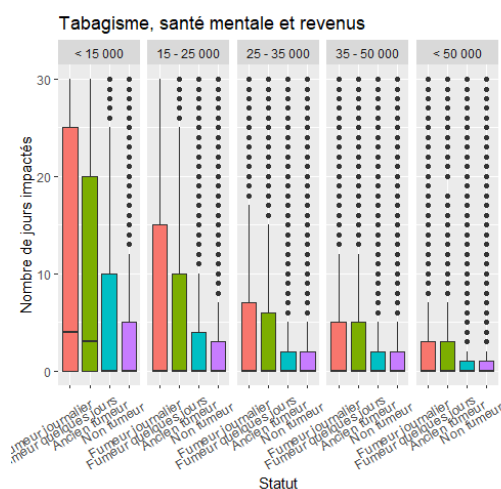
Graphique 16 :



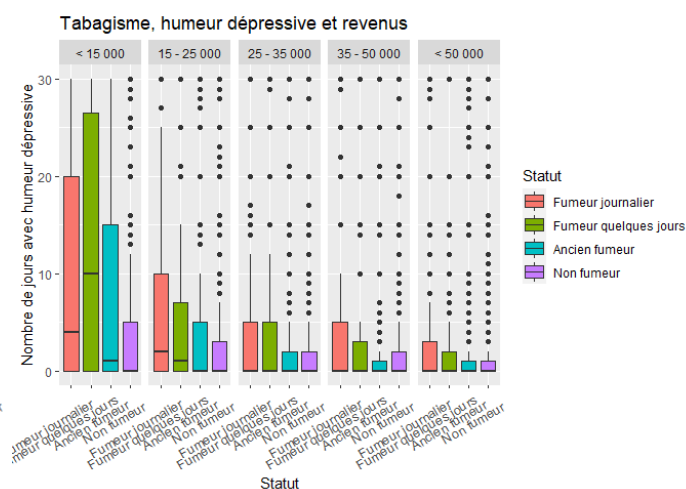
La prévalence de la dépression semble moins importante auprès des répondants mariés.

## 4. Tabagisme, santé mentale et niveau de revenus

Graphique 17 :



Graphique 18 :

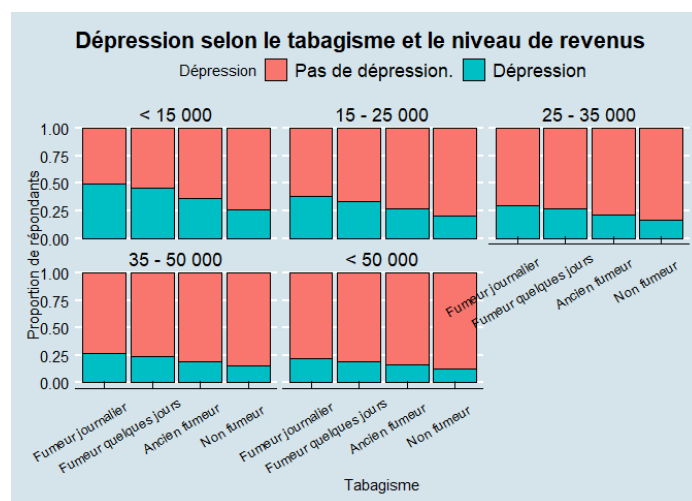


# EFFET DE LA DÉPRESSION SUR LE TABAGISME

Les revenus semblent également impacter de manière importante la santé mentale des répondants. Effectivement, plus les revenus sont bas, plus les répondants déclarent un nombre de jours impactés par la santé mentale important. Les distinctions fumeurs/non fumeurs sont fortement visibles sur ce premier graphique.

L'humeur dépressive semble également diminuer avec les revenus. Les différences sont également visibles ici selon le fait d'être fumeur ou non fumeur.

Graphique 19 :



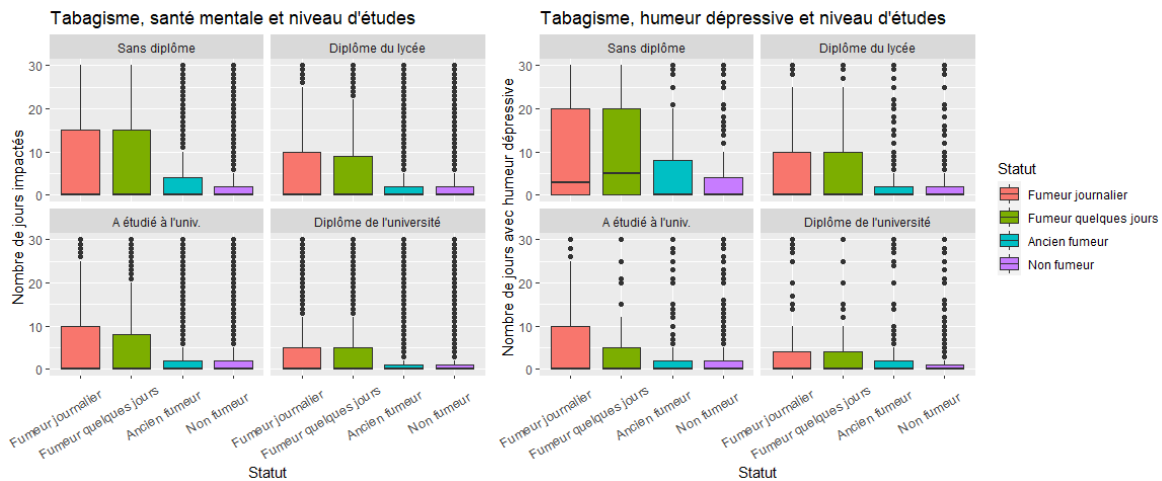
Il semblerait également, au vu du graphique, que la prévalence de la dépression diminue avec les revenus.

## 5. Tabagisme, santé mentale et niveau d'études

Graphique 20 :

Graphique 21 :

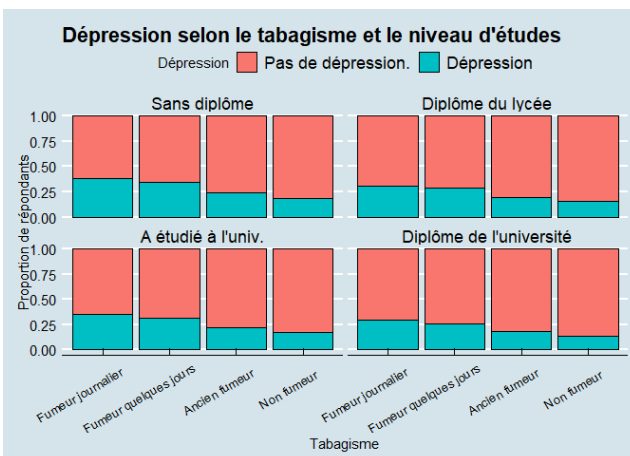
# EFFET DE LA DÉPRESSION SUR LE TABAGISME



Il semblerait que les fumeurs sans diplôme soient les plus impactés négativement par leur santé mentale d'après le premier graphique.

De même, les différences d'humeur sont davantage visibles auprès des répondants fumeurs sans diplôme ou ayant un diplôme du lycée.

Graphique 22 :



La dépression semble davantage présente auprès des répondants sans diplôme. Contrairement aux résultats concernant l'humeur dépressive, on ne distingue pas de différence entre les diplômés du lycée et les répondants ayant étudié à l'université.

# EFFET DE LA DÉPRESSION SUR LE TABAGISME

## ANALYSE DES CORRESPONDANCES MULTIPLES

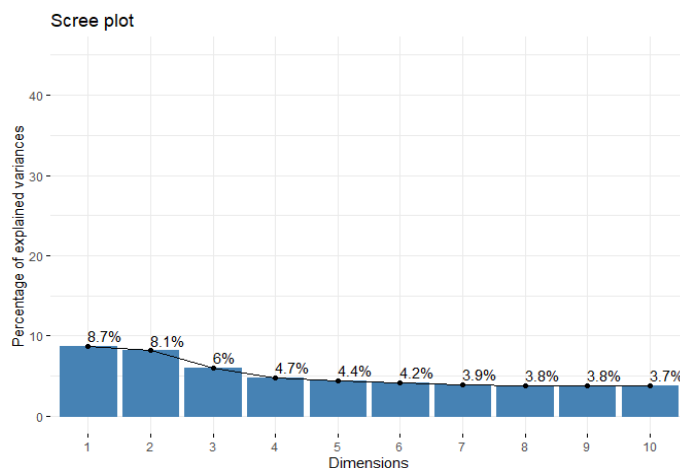
L'analyse des correspondances multiples, résumant l'information d'un grand nombre de variables, a été réalisée afin de faciliter l'interprétation des corrélations existants entre ces dernières et en particulier déterminer les modalités corrélées entre-elles.

Les variables sélectionnées pour créer les dimensions sont les suivantes :

- SEX (sexe),
- X\_AGE\_G\_FACTOR (âge),
- MARITAL (statut matrimonial),
- X\_EDUCAG\_FACTOR (niveau d'études),
- EMPLOY1\_FACTOR (CSP),
- RENTHOM1\_FACTOR (statut par rapport au logement),
- X\_INCOMG\_FACTOR (niveau de revenus)

Les variables relatives à la dépression et au tabagisme ont été ajoutées de manière supplémentaire.

Graphique 23 :



Ainsi, les axes 1 et 2 de l'ACM expliquent à eux deux près de 17% de la variance.

Pour

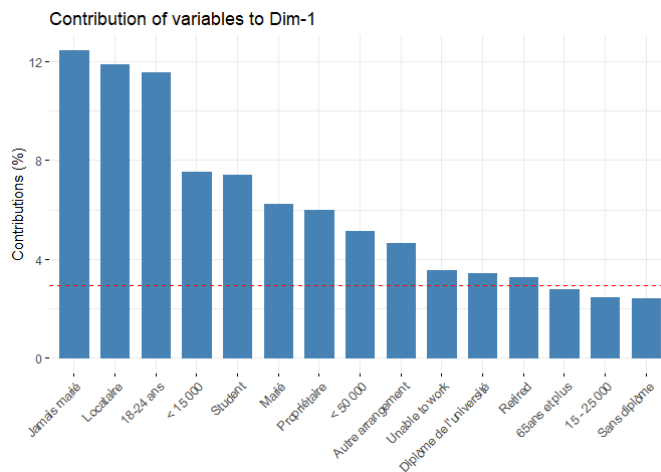
comprendre la signification de ces deux axes, il importe d'identifier quelles sont les variables/ modalités qui ont le plus



# EFFET DE LA DÉPRESSION SUR LE TABAGISME

contribué à leur création.

Graphique 24 :

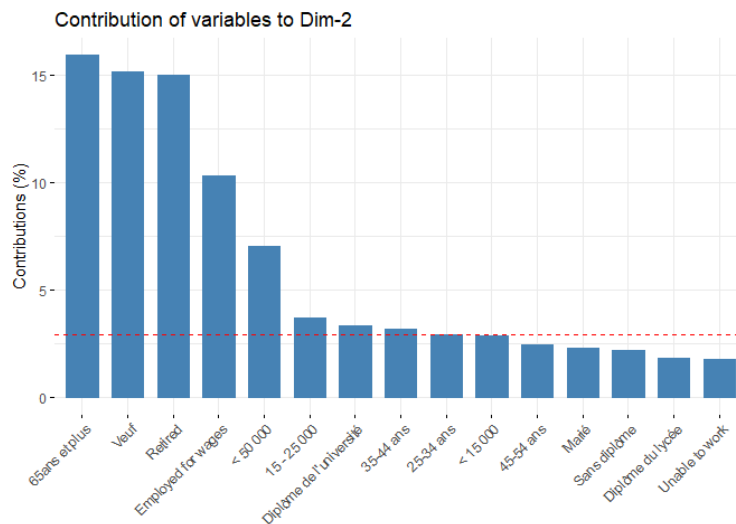


Les modalités :

- Jamais marié,
- Locataire,
- 18-24 ans,
- Revenus < 15000\$ annuels,
- Etudiant

sont celles ayant le plus contribué à l'axe 1.

Graphique 25 :



Les modalités :

- 65ans et plus,
- Veuf,
- Retraité,
- Employé,
- Revenus supérieurs à 50 000\$ annuels

sont celles ayant le plus contribué à l'axe 2.

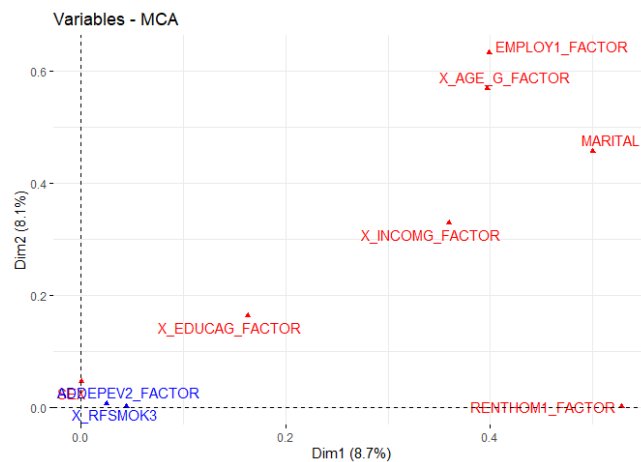
Graphique 26 :

# EFFET DE LA DÉPRESSION SUR LE TABAGISME

On remarque que les variables relatives à l'âge, les CSP, le statut matrimonial et le statut par rapport au logement sont les plus corrélées à l'axe 1.

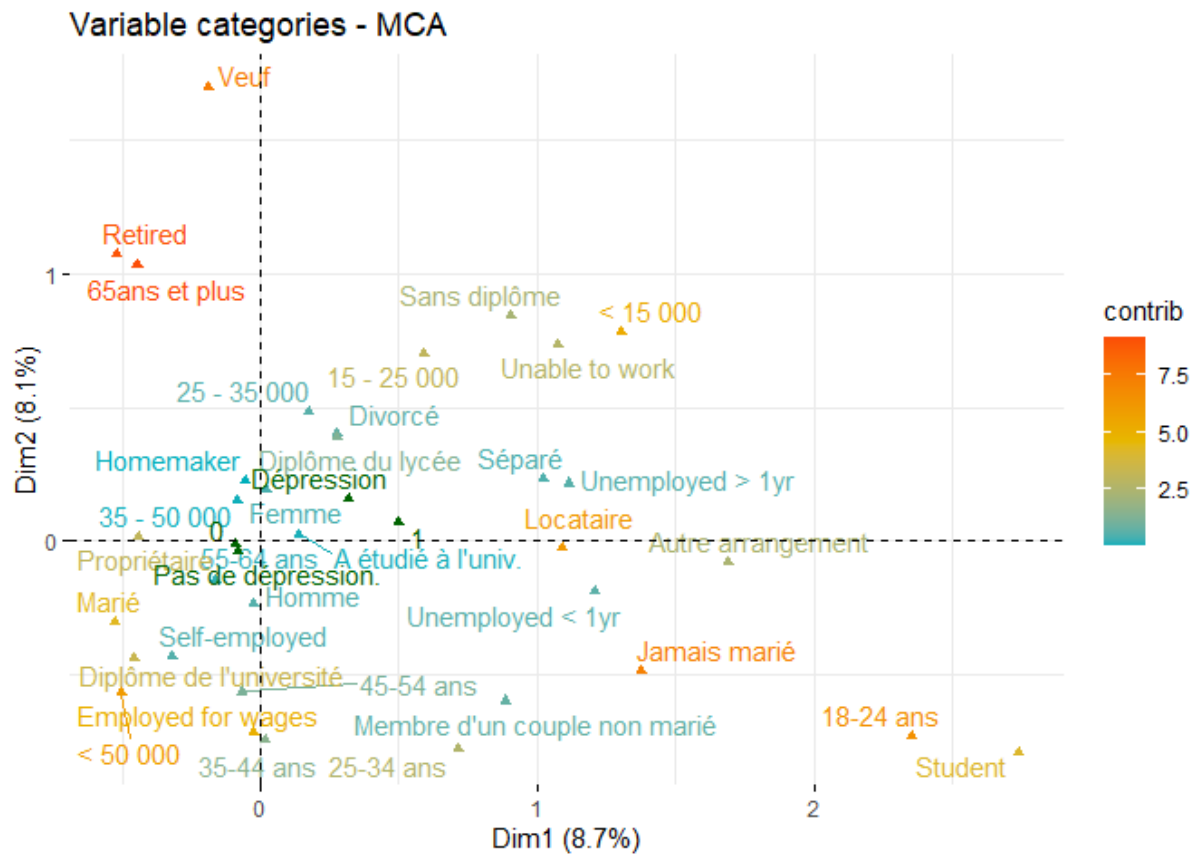
Par rapport à l'axe 2, ce sont le statut matrimonial, l'âge et les CSP qui sont les plus corrélées.

Le sexe, le niveau d'éducation, la dépression ainsi que le tabagisme sont peu corrélés aux



Graphique 27 : Représentation des modalités et de leur contribution aux axes 1 et 2

# EFFET DE LA DÉPRESSION SUR LE TABAGISME



Ce graphique nous amène à nous interroger sur les liens entre les modalités :

- Dépression et Femme, Divorcé, Diplôme du lycée, Fumeur
- Pas de dépression et Homme, 55-64ans, Non fumeur, revenus de 35 à 50 000\$

# EFFET DE LA DÉPRESSION SUR LE TABAGISME

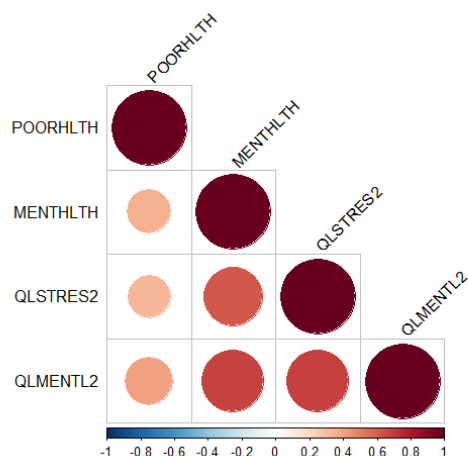
## Etude des liaisons

### VARIABLES QUANTITATIVES : CORRÉLATIONS ET ASSOCIATIONS

Après avoir visualisé la distribution des variables quantitatives à l'aide d'histogrammes et comparé graphiquement les distributions avec une distribution normale, nous avons pu constater qu'aucune de nos variables quantitative ne semble suivre une loi normale.

Nous avons donc effectué des tests de corrélation de Spearman afin de mesurer les corrélations entre variables, et avons en particulier remarqué que les variables « QLMENTL2 » (concernant l'humeur dépressive) et « QLSTRES2 » (concernant l'humeur anxieuse) mais aussi « MENTHLTH » (relative à l'impact de la santé mentale sur la vie quotidienne) et « QLMENTL2 » étaient fortement corrélés, avec respectivement des rho à 0.63 et 0.67.

Graphique 28 : Matrice de corrélation



Nous avons également réalisé la matrice de corrélation suivante afin de visualiser les variables liées entre-elles.

On peut notamment relever que l'humeur dépressive est corrélée fortement à l'humeur anxieuse ainsi qu'à une mauvaise santé mentale.

### VARIABLES QUALITATIVES : INDÉPENDANCE

Nous avons également effectué des tests d'indépendance du  $\chi^2$  afin de voir les éventuelles liaisons existant entre nos variables.

Nous avons en particulier pu remarquer que:

# EFFET DE LA DÉPRESSION SUR LE TABAGISME

- L'âge, le sexe, le statut matrimonial, le niveau d'éducation ainsi que le niveau de revenus ne sont pas indépendants du tabagisme,
- L'impact de la santé mentale sur la vie quotidienne, l'humeur dépressive, l'humeur anxieuse et la dépression ne sont pas indépendants du tabagisme.

Effectivement, au vu des p-values largement inférieures à 0.05, les différences sont suffisamment significatives de sorte qu'on ne pourrait pas les expliquer par le hasard. (Annexe 1)

## Effet du tabagisme sur la dépression

Etant données les visualisations ainsi que les non indépendances relevées plus tôt entre plusieurs variables sociodémographiques avec le tabagisme, nous avons décidé, afin de limiter les biais et de nous rapprocher des essais randomisés contrôlés, de nous appuyer sur des méthodes d'appariement pour étudier l'effet du tabagisme sur la dépression.

Plus précisément, nous effectuerons deux appariements différents : un appariement exact puis un appariement par score de propension. L'effet sera déterminé, suite aux appariements, à l'aide d'une régression logistique.

### MÉTHODE 1 : RÉGRESSION LOGISTIQUE SUITE À LA CORRESPONDANCE EXACTE

#### 1. Appariement exact

L'appariement exact est une forme d'appariement de strates qui implique la création de sous-classes basées sur des combinaisons uniques de valeurs de covariables et l'affectation de chaque unité dans leur sous-classe correspondante de sorte que seules les unités avec des valeurs de covariables identiques soient placées dans la même sous-classe.

Toutes les unités appartenant à des sous-classes dépourvues d'unités traitées ou contrôlées seront supprimées. L'appariement exact est la méthode d'appariement la plus puissante en ce sens qu'aucune hypothèse de forme fonctionnelle n'est requise sur le modèle de traitement ou de résultat pour que la méthode élimine les facteurs de confusion dus aux covariables mesurées. Les distributions de covariables sont parfaitement équilibrées.

Pour effectuer une correspondance exacte sur toutes les covariables, nous nous sommes basées sur le package MatchIT et notamment la fonction « matchit » en définissant l'argument correspondant à la méthode sur "exact".

Nous avons décidé, au vu des résultats précédents, d'effectuer l'appariement sur les facteurs suivants :

- sexe,
- âge,
- statut matrimonial,
- niveau d'éducation,
- niveau de revenus.

# EFFET DE LA DÉPRESSION SUR LE TABAGISME

L'appariement exact avec la prise en compte des poids a été effectif étant donné que les variables sont parfaitement équilibrées comme le montre le tableau ci-contre :

Tableau 1 : Equilibre des variables et de leurs modalités

Stratifié par X_RFSMOK3				
		Non fumeur	Fumeur	p
n		2174258	391069	
SEX = Femme (%)		1217273 (56.0)	203747 (52.1)	1.000
Âge	18-24ans	0.05 (0.26)	0.05 (0.21)	1.000
	25-34ans	0.14 (0.59)	0.14 (0.35)	1.000
	35-44ans	0.16 (0.59)	0.16 (0.36)	1.000
	45_54ans	0.21 (0.68)	0.21 (0.41)	1.000
	55_64ans	0.25 (0.66)	0.25 (0.43)	1.000
	> 64 ans	0.19 (0.34)	0.19 (0.39)	1.000
Statut matrimonial	Marié	0.37 (0.54)	0.37 (0.48)	1.000
	Divorcé	0.23 (0.77)	0.23 (0.42)	1.000
	Veuf	0.10 (0.38)	0.10 (0.30)	1.000
	Séparé	0.04 (0.37)	0.04 (0.20)	1.000
	Jamais marié	0.20 (0.66)	0.20 (0.40)	1.000
	Couple non mari	0.05 (0.37)	0.05 (0.22)	1.000
Niveau d'études	Sans diplôme	0.13 (0.63)	0.13 (0.33)	1.000
	Diplôme lycée	0.38 (0.85)	0.38 (0.48)	1.000
	Suivi université	0.32 (0.68)	0.32 (0.46)	1.000
	Diplôme univ.	0.18 (0.30)	0.18 (0.38)	1.000
Niveau de revenus	< 15000	0.19 (0.76)	0.19 (0.39)	1.000
	15000 à 25000	0.24 (0.72)	0.24 (0.43)	1.000
	25000 à 35000	0.13 (0.47)	0.13 (0.33)	1.000
	35000 à 50000	0.14 (0.46)	0.14 (0.35)	1.000
	> 50000	0.29 (0.46)	0.29 (0.46)	1.000

## 2. Régression logistique

# EFFET DE LA DÉPRESSION SUR LE TABAGISME

Nous avons ensuite effectué la régression logistique en fournissant les poids de stratification afin de prédire la dépression à partir du tabagisme sur l'échantillon apparié.

La fonction « summary » (voir Annexe 2 pour les résultats) se basant sur l'hypothèse d'homoscédasticité (c'est à dire qu'elle suppose que la variance des erreurs est constante), nous avons donc estimé l'erreur standard robuste avec la fonction « coeftest » en paramétrant « vcov. » sur « vcovHC ».

« vcovHC » est une fonction d'estimation d'une matrice de variance-covariance robuste de paramètres pour un modèle de panel à effets fixes ou à effets aléatoires selon la méthode White que l'on peut utiliser en tant que paramètre dans la fonction « coeftest ».

Les résultats de l'estimation de l'erreur standard robuste régulière de la régression dont les poids de stratification ont été fournis sont les suivants :

Tableau 2 : Estimation de l'erreur standard robuste

*z test of coefficients:*

	Estimate	Std. Error	Z value	Pr (> z )
Intercept	-1.2787372	0.0024553	-520.80	< 2.2e-16 ***
X_RFSMOK3	0.5160420	0.0042211	122.25	< 2.2e-16 ***

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

On observe donc que le tabagisme a un effet positif significatif sur la dépression. (Coefficient estimé à 0.52 et p. value < 2e-16). On rejette donc l'hypothèse H0 selon laquelle la probabilité d'être dépressif chez les fumeurs est similaire à celle des non-fumeurs.

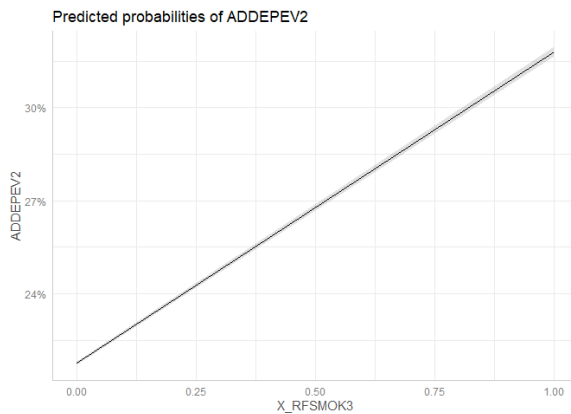
On obtient alors l'équation suivante :

$$\log \frac{\text{prob}(ADDEPEV2 = 1)}{1 - \text{prob}(ADDEPEV2 = 1)} = -1.278737 + 0.516042 * X_{RFSMOK3}$$

Graphique 29 : Probabilités prédites de la dépression en fonction du tabagisme



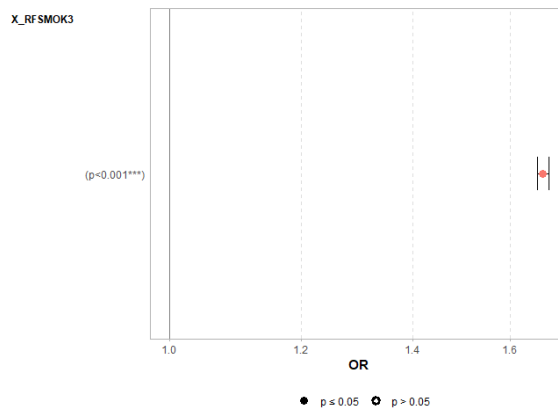
# EFFET DE LA DÉPRESSION SUR LE TABAGISME



Les fumeurs ont environ 33% de chance de souffrir de dépression, tandis que les non-fumeurs 22% d'après le modèle de régression logistique.

On peut également évaluer que les fumeurs ont approximativement 1.67 fois plus de chance de souffrir de dépression que les non-fumeurs.

Graphique 30 : Odds-ratio



## MÉTHODE 2 : RÉGRESSION LOGISTIQUE SUITE À LA PONDÉRATION AVEC WEIGHTIT

Le package « WeightIt » contient plusieurs fonctions pour estimer et évaluer les poids d'équilibrage. Ainsi, WeightIt fait pour pondérer ce que MatchIt a été fait pour la correspondance.

La fonction « Weightit » génère des pondérations à partir d'une formule donnée et d'une entrée de données selon des méthodes et d'autres paramètres que l'on spécifie.

Nous avons utilisé la méthode « ps » pour les poids des scores de propension et ATT comme estimateur puisque nous nous intéressons à l'effet moyen du traitement sur les traités.

# EFFET DE LA DÉPRESSION SUR LE TABAGISME

Tableau 3 : Balance des variables avant la pondération - Bilan des différences moyennes

On remarque qu'il existe avant la pondération plusieurs variables déséquilibrées (10 modalités), c'est à dire que les différences de proportion sont supérieures à 0,05.

	Count
Balanced, <0.05	11
Not Balanced, >0.05	10

Tout d'abord, nous avons utilisé des pondérations de probabilité inverses à partir des scores de propension générés par la régression logistique. Nous avons donc fourni à `weightit()` la formule du modèle, l'ensemble de données, l'estimation (ATT) et la méthode d'estimation ("ps").

L'évaluation des poids comporte deux éléments: évaluer l'équilibre des covariables produit par les poids et évaluer si les poids permettront une précision suffisante dans l'estimation de l'effet éventuel.

Tableau 4 : Balance des variables après la pondération - Bilan des différences moyennes

	Count
Balanced, <0.05	22
Not Balanced, >0.05	0

Tableau 5 : Balance des variables après la pondération – Mesures

	Type	Diff.Adj	M.Threshold
prop.score	Distance	-0.0055	Balanced, <0.05

# EFFET DE LA DÉPRESSION SUR LE TABAGISME

Âge	18-24ans	Binary	0.0002	Balanced, <0.05
	25-34ans	Binary	0.0005	Balanced, <0.05
	35-44ans	Binary	-0.0008	Balanced, <0.05
	45_54ans	Binary	0.0012	Balanced, <0.05
	55_64ans	Binary	0.0003	Balanced, <0.05
	> 64 ans	Binary	-0.0014	Balanced, <0.05
Statut matrimonial	Marié	Binary	0.0020	Balanced, <0.05
	Divorcé	Binary	-0.0011	Balanced, <0.05
	Veuf	Binary	0.0011	Balanced, <0.05
	Séparé	Binary	-0.0011	Balanced, <0.05
	Jamais marié	Binary	0.0010	Balanced, <0.05
	Couple non mari	Binary	-0.0020	Balanced, <0.05
Niveau d'études	Sans diplôme	Binary	-0.0015	Balanced, <0.05
	Diplôme lycée	Binary	0.0006	Balanced, <0.05
	Suivi université	Binary	0.0004	Balanced, <0.05
	Diplôme univ.	Binary	0.0005	Balanced, <0.05
Niveau de revenus	< 15000	Binary	-0.0005	Balanced, <0.05
	15000 à 25000	Binary	-0.0011	Balanced, <0.05
	25000 à 35000	Binary	-0.0004	Balanced, <0.05
	35000 à 50000	Binary	-0.0001	Balanced, <0.05
	> 50000	Binary	0.0021	Balanced, <0.05
		V.Ratio.Adj		
prop.score		0.9638		

Tableau 6 : Balance des variables après la pondération - Tailles d'échantillon

	Control	Treated
Unadjusted	2174258	391069
Adjusted	1164882	391069

# EFFET DE LA DÉPRESSION SUR LE TABAGISME

Nous remarquons que suite à la pondération, toutes les variables sont équilibrées.

Tableau 7 : Variation des poids

	Min	Max
<b>Treated</b>	<b>1</b>	<b>1</b>
<b>Control</b>	<b>0.0239</b>	<b>1.0937</b>

Tableau 8 : Statistiques des poids

	Coef of Var	MAD	Entropy	Zeros
<b>Treated</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>Control</b>	<b>0.931</b>	<b>0.661</b>	<b>0.343</b>	<b>0</b>

Des pondérations à faible variabilité étant souhaitables car elles améliorent la précision de l'estimateur, la variabilité des poids semble donc correcte. Aussi, avec un ESS de 1156794, on peut estimer que la taille de l'échantillon est très satisfaisante.

Nous avons donc ensuite évalué l'erreur type de l'estimation de l'effet, dont les résultats sont les suivants :

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	<b>-1.279518</b>	<b>0.002398</b>	<b>-533.6</b>	<b>&lt;2e-16 ***</b>
X_RFSMOK3	<b>0.516847</b>	<b>0.004188</b>	<b>123.4</b>	<b>&lt;2e-16 ***</b>

Tableau 9 : Evaluation de l'erreur type

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

(Dispersion parameter for binomial family taken to be 1)

Number of Fisher Scoring iterations: 4

# EFFET DE LA DÉPRESSION SUR LE TABAGISME

L'équation de la régression est donc la suivante :

$$\log \frac{\text{prob}(ADDEPEV2 = 1)}{1 - \text{prob}(ADDEPEV2 = 1)} = -1.279518 + 0.516847 * X_{RFSMOK3}$$

Tableau 10 : Exponentiel des coefficients

<b>Intercept</b>	<b>0.2781714</b>
<b>X_RFSMOK3</b>	<b>1.6767324</b>

Tableau 11 : Intervalles de confiance

	<b>2.5 %</b>	<b>97.5%</b>
<b>Intercept</b>	0.276867	0.2794819
<b>X_RFSMOK3</b>	<b>1.663025</b>	<b>1.6905526</b>

On obtient donc les mêmes résultats qu'avec la première méthode d'appariement exact. On peut ainsi conclure que le tabagisme aurait un effet positif sur la dépression. En particulier, les fumeurs auraient 1.7 fois plus de chance de souffrir de dépression que les non-fumeurs. Le tabagisme serait donc un facteur de risque de la dépression.

# EFFET DE LA DÉPRESSION SUR LE TABAGISME

## Effet de la dépression sur le tabagisme

### MÉTHODE 1 : RÉGRESSION LOGISTIQUE SUITE À LA CORRESPONDANCE EXACTE

Pour l'étude de l'effet de la dépression sur le tabagisme, nous avons décidé de garder les mêmes covariables à pondérer que celles de la causalité inverse.

Les 2 groupes de répondants dépressifs/non dépressifs n'étant pas équilibrés au regard des variables sociodémographiques, nous avons dû les appairer de la même manière que dans l'étude précédente.

Tableau 12 : Equilibre des variables et de leurs modalités

Stratifié par ADDEPEV2				
		Absence de dépression	Présence de dépression	p
n		2061467	503769	
Âge	18-24ans	0.05 (0.23)	0.05 (0.21)	1.000
	25-34ans	0.11 (0.38)	0.11 (0.31)	1.000
	35-44ans	0.13 (0.45)	0.13 (0.34)	1.000
	45_54ans	0.19 (0.59)	0.19 (0.39)	1.000
	55_64ans	0.26 (0.65)	0.26 (0.44)	1.000
	> 64 ans	0.27 (0.42)	0.27 (0.44)	1.000
Statut matrimonial	Marié	0.43 (0.49)	0.43 (0.49)	1.000
	Divorcé	0.21 (0.69)	0.21 (0.41)	1.000
	Veuf	0.12 (0.38)	0.12 (0.32)	1.000
	Séparé	0.04 (0.31)	0.04 (0.19)	1.000
	Jamais marié	0.17 (0.48)	0.17 (0.38)	1.000
	Couple non mari	0.04 (0.25)	0.04 (0.19)	1.000
Niveau d'études	Sans diplôme	0.09 (0.42)	0.09 (0.28)	1.000
	Diplôme lycée	0.27 (0.59)	0.27 (0.45)	1.000
	Suivi université	0.31 (0.68)	0.31 (0.46)	1.000
	Diplôme univ.	0.33 (0.52)	0.33 (0.47)	1.000

# EFFET DE LA DÉPRESSION SUR LE TABAGISME

Niveau de revenus	< 15000	0.19 (0.74)	0.19 (0.39)	1.000
	15000 à 25000	0.22 (0.58)	0.22 (0.41)	1.000
	25000 à 35000	0.11 (0.36)	0.11 (0.32)	1.000
	35000 à 50000	0.13 (0.36)	0.13 (0.34)	1.000
	> 50000	0.35 (0.41)	0.35 (0.41)	1.000

Suite à l'appariement et la prise en compte des poids, l'ensemble des modalités sont maintenant équilibrées.

Nous avons ensuite réalisé la régression en fournissant les poids de stratification puis évalué l'erreur standard robuste, dont voici les résultats :

Tableau 13 : Evaluation de l'erreur standard robuste

*z test of coefficients:*

	Estimate	Std. Error	Z value	Pr (> z )
Intercept	-1.6163585	0.0026045	-620.61	< 2.2e-16 ***
ADDEPEV2	0.5011592	0.0041784	119.94	< 2.2e-16 ***

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

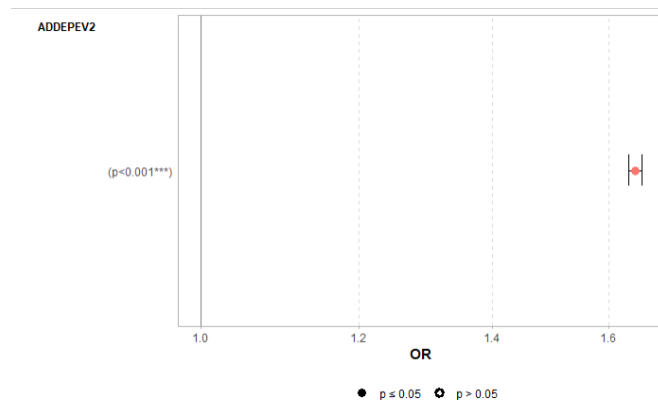
On observe donc que le tabagisme a un effet positif significatif sur la dépression. (coefficient évalué à 0.50 et p. value <2e-16).

On obtient alors l'équation de régression suivante :

$$\log \frac{\text{prob}(X_{RFSMOK3} = 1)}{1 - \text{prob}(X_{RFSMOK3} = 1)} = -1.6163585 + 0.5011592 * ADDEPEV2$$

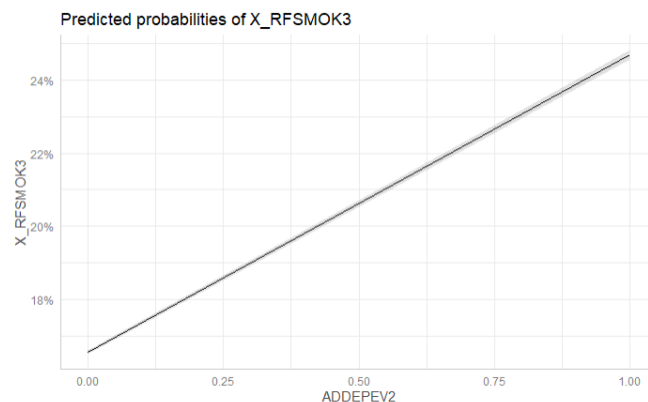
# EFFET DE LA DÉPRESSION SUR LE TABAGISME

Graphique 31 : Odds-ratio



On peut donc supposer que les dépressifs ont 1.65 fois plus de chance d'être fumeur que les non dépressifs.

Graphique 32 : Probabilités prédites du tabagisme en fonction de la dépression



Les dépressifs ont environ 23% de chance d'être fumeur, tandis que les non dépressifs 17% d'après le modèle de régression logistique.

## MÉTHODE 2 : RÉGRESSION LOGISTIQUE SUITE À LA PONDÉRATION AVEC WEIGHT

Après avoir vérifié l'équilibre des variables et la variabilité des poids, nous avons pu évaluer l'effet de la dépression sur le tabagisme à l'aide de la régression logistique dont les résultats sont les suivants :



# EFFET DE LA DÉPRESSION SUR LE TABAGISME

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	<b>-1.616869</b>	<b>0.002366</b>	<b>-683.4</b>	<b>&lt;2e-16 ***</b>
ADDEPEV2	<b>0.501696</b>	<b>0.004034</b>	<b>124.4</b>	<b>&lt;2e-16 ***</b>

Tableau 14 : Evaluation de l'erreur type

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

(Dispersion parameter for binomial family taken to be 1)

Number of Fisher Scoring iterations: 5

On obtient donc l'équation suivante :

$$\log \frac{\text{prob}(X_{RFSMOK3} = 1)}{1 - \text{prob}(X_{RFSMOK3} = 1)} = -1.616869 + 0.501696 * ADDEPEV2$$

Le calcul de l'exponentiel des coefficients ainsi que des intervalles de confiance sont les suivants :

Tableau 15 : Exponentiel des coefficients

<b>Intercept</b>	<b>0.1985194</b>
<b>ADDEPEV2</b>	<b>1.6515203</b>

Tableau 16 : Intervalles de confiance

	<b>2.5 %</b>	<b>97.5%</b>
<b>Intercept</b>	0.1976009	0.1994421
<b>ADDEPEV2</b>	<b>1.6385141</b>	<b>1.6646298</b>

On obtient donc les mêmes résultats qu'avec la première méthode matchIT.

# EFFET DE LA DÉPRESSION SUR LE TABAGISME

## MÉTHODE 3 : RÉGRESSION DYNAMIQUE DE SÉRIES TEMPORELLES

L'objectif d'utiliser des séries temporelles était d'exploiter la dimension historique des données disponibles, de 1984 à aujourd'hui.

En particulier, le modèle de régression dynamique de séries temporelles permet de prendre en compte la dynamique temporelle des séries en plus d'inclure plusieurs variables explicatives .

En effet, les modèles d'ARIMA classique et exponential smoothing permettent d'utiliser les observations passées d'une série mais ne permettent pas l'inclusion d'autres variables, dites exogènes. Les modèles de régression simples eux, permettent d'inclure d'autres variables mais ne prennent pas en compte la dynamique subtile des séries temporelles tel que le fait ARIMA.

### Equations

- Equation d'une série temporelle :

$$y_t = S_t + T_t + R_t,$$

L'équation de base d'une série temporelle y a un moment t représente les différentes composantes :

S : la saisonnalité, la répétition d'un certain schéma sur une période régulière

T : la tendance et/ou le cycle

la tendance indique les changements de direction, augmentation et diminution

le cycle est comme la saisonnalité, une fluctuation, mais sans période fixe et en général plus longue (au moins 2 ans)

R : le résidu, c'est-à-dire la série une fois que les deux premières composantes ont été retirées

- Equation modèle ARIMA :

$$y_t = \beta_0 + \beta_1 x_{1,t} + \cdots + \beta_k x_{k,t} + \varepsilon_t$$

# EFFET DE LA DÉPRESSION SUR LE TABAGISME

- Equation modèle ARIMAX (ARIMA avec variable exogène) :

$$y_t = \beta_0 + \beta_1 x_{1,t} + \dots + \beta_k x_{k,t} + \eta_t$$

- y représente la variable expliquée
- x les variables explicatives
- beta 0 représente l'intercept
- beta 1...k représente la pente, c'est-à-dire la variation moyenne prédite de y pour une unité de x en plus

La différence entre les deux modèles se situe au niveau des erreurs. Dans un modèle ARIMA classique, l'erreur epsilon est du bruit blanc, tandis que dans un ARIMAX, l'erreur  $\eta_t$  suit un modèle ARIMA et contient donc de l'autocorrélation ainsi que sa propre erreur epsilon.

## Stationnarisation des variables

Pour appliquer ce modèle, toutes les variables doivent être stationnaires si elles ne le sont déjà.

En reprenant l'équation des time series, le but est d'enlever S et T (excepté du cycle), afin que les propriétés de la série ne dépendent pas du moment où on l'observe. La série d'une variable stationnaire ressemble à du bruit blanc.

Afin de vérifier cette condition, plusieurs opérations sont mises en place. Dans l'application Shiny, il est possible d'appliquer cette vérification de manière plus automatique en sélectionnant la variable désirée.

- Vérification de la stabilité de la variance et de la moyenne au début, milieu et fin de chaque série
- Unit Root Test : permet de déterminer si une différenciation est nécessaire avec :
  - o  $H_0$  : la série est stationnaire
  - o p-value < 0.05 : une différenciation est requise
- Test de saisonnalité
- Lag plot : permet de détecter la saisonnalité

# EFFET DE LA DÉPRESSION SUR LE TABAGISME

- o Si les différentes couleurs forment une diagonale, indique qu'il y a un ordre chronologique fort
- ACF plot : indique la stationnarité
  - o Si 95% des lags sont contenus entre les deux lignes bleues, cela représente du bruit blanc et la série est donc stationnaire

Le résultat des tests de stationnarité a indiqué que la plupart des variables requièrent une différenciation pour stabiliser la moyenne et aucune n'était saisonnière.

## Mise en place du modèle

La fonction `auto.arima` a été utilisée pour trouver le modèle ARIMA. Cette fonction effectue directement la différenciation lorsqu'elle est nécessaire en utilisant un Unit Root test.

La variable expliquée choisie est SMOKE100 (individu ayant fumé au moins 100 cigarettes dans leur vie) étant donné un nombre de réponses plus important.

Le modèle a été appliqué trois fois pour la variable SMOKE100 en changeant les variables explicatives. Dans un premier temps, toutes les variables explicatives ont été prises en compte : sociodémographiques, la santé mentale et l'utilisation de substituts.

Dans un second temps, les variables explicatives ont été limitées à la santé mentale.

Dernièrement, une seule variable exogène explicative fut sélectionnée en fonction des résultats de l'autre modèle : ADDEPEV2 (a été dit dépressif).

Suite aux résultats, la variable expliquée a été changée en X\_RFSMOK3 (fumeur) avec pour variable explicative ADDEPEV2.

## Estimation

Avec des variables stationnaires, l'estimation du modèle peut se faire en considérant uniquement les résidus des erreurs ARIMA. Celles-ci doivent ressembler à du bruit blanc pour que le modèle soit valide.

Dans les quatre cas, le meilleur modèle est un ARIMA(5, 1, 0) :

- auto-regressivité 5
- 1 différenciation

# EFFET DE LA DÉPRESSION SUR LE TABAGISME

- moyenne mouvante d'ordre 0

## Résultats

La validation du modèle passe par la vérification des résidus des erreurs ARIMA qui doivent avoir une distribution normale.

Nos résultats, même s'ils semblent de premier abord ressembler à du bruit blanc, n'en sont pas. En effet il y a de grandes variations, et l'ACF montre des lags qui sont autocorrélés

Le Ljung-Box Test permet de rejeter l'hypothèse nulle que les résidus sont indépendamment distribués, où la p-value doit être supérieure à 0.05.

L'analyse des séries temporelles pour expliquer si un individu a fumé plus de 100 cigarettes dans sa vie ou bien s'il est fumeur ne sont pas concluants. Il n'est pas possible de prédire de manière sûre avec les données sociodémographiques, d'utilisation de substituts ni de santé mentale.

## Conclusion

Ainsi, nos résultats vont dans le sens de la revue de littérature étant donné que nous avons observé non seulement un effet du tabagisme sur la prévalence de la dépression, mais également un effet de la dépression sur le tabagisme. Il semblerait donc qu'il existe une relation bidirectionnelle entre le tabagisme et la dépression.

Toutefois, nous devons souligner le fait que la dépression ici n'a pas été mesurée sur les participants à l'aide d'un questionnaire standardisé, leur réponse étant déclarative. Il se pourrait ainsi que certaines personnes aient indiqué ne pas avoir reçu de diagnostic de dépression sans pour autant ne pas souffrir de dépression. De la même façon, nous n'avons pas d'indication sur la date du diagnostic qui pourrait le rendre obsolète, bien que ce syndrome persiste dans la durée par sa chronicité et les rechutes fréquentes.

Il pourrait également être intéressant de se concentrer sur les différentes modalités du tabagisme, en prenant en compte la régularité de la consommation de tabac ainsi que le fait d'être un ancien fumeur. Effectivement, il semblerait, au vu des statistiques descriptives, qu'il existe des différences au regard de la santé mentale selon le statut des répondants par rapport au tabagisme.

Aussi, des études futures pourraient évaluer les relations entre tabagisme et dépression en prenant en compte la gravité de la maladie et si la personne diagnostiquée est en rémission partielle ou totale.

## Bibliographie

Grover, K. W., Goodwin, R. D., & Zvolensky, M. J. (2012). Does current versus former smoking play a role in the relationship between anxiety and mood disorders and nicotine dependence? *Addictive Behaviors*, 37(5), 682–685. <https://doi-org.proxybib-pp.cnam.fr/10.1016/j.addbeh.2012.01.014>

Lee, B. (2021). Longitudinal relationship between tobacco product use and mental health in adolescence and adulthood (2020-86256-179; Numéros 4-B) [ProQuest Information & Learning]. <http://proxybib-pp.cnam.fr/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2020-86256-179&site=ehost-live&scope=site>

Mojtabai, R., & Crum, R. M. (2013). Cigarette smoking and onset of mood and anxiety disorders. *American Journal of Public Health*, 103(9), 1656–1665. <https://doi.org/10.2105/AJPH.2012.300911>

Patten, C. A., Gillin, J. C., Golshan, S., Wolter, T. D., & Kelsoe, J. (2001). Relationship of Mood Disturbance to Cigarette Smoking Status Among 252 Patients With a Current Mood Disorder. *The Journal of Clinical Psychiatry*, 62(5), 0–0.

Richardson, A., He, J.-P., Curry, L., & Merikangas, K. (2012). Cigarette smoking and mood disorders in U S adolescents: Sex-specific associations with symptoms, diagnoses, impairment and health services use. *Journal of Psychosomatic Research*, 72(4), 269–275. <https://doi-org.proxybib-pp.cnam.fr/10.1016/j.jpsychores.2012.01.013>

Salt, V., & Osborne, C. (2020). Mental health, smoking and poverty : Benefits of supporting smokers to quit. *BJPsych Bulletin*, 44(5), 213–218. <https://doi.org/10.1192/bjb.2020.88>

Stone, M. D., Audrain-McGovern, J., & Leventhal, A. M. (2017). Association of Anhedonia With Adolescent Smoking Susceptibility and Initiation. *Nicotine & Tobacco Research*, 19(6), 738–742. <https://doi.org/10.1093/ntr/ntw177>

# EFFET DE LA DÉPRESSION SUR LE TABAGISME



# EFFET DE LA DÉPRESSION SUR LE TABAGISME

## Annexes

### ANNEXE 1 : TESTS D'INDÉPENDANCE

- Age et tabagisme

```
##  
## Pearson's Chi-squared test  
##  
## data: brfss$X_AGE_G_FACTOR and brfss$SMOKDAY2_FACTOR  
## X-squared = 134364, df = 10, p-value < 2.2e-16
```

- Sexe et tabagisme

```
##  
## Pearson's Chi-squared test  
##  
## data: brfss$SEX and brfss$SMOKDAY2_FACTOR  
## X-squared = 682.91, df = 2, p-value < 2.2e-16
```

- Statut matrimonial et tabagisme

```
##  
## Pearson's Chi-squared test  
##  
## data: brfss$MARITAL and brfss$SMOKDAY2_FACTOR  
## X-squared = 81809, df = 10, p-value < 2.2e-16
```

- Niveau d'éducation et tabagisme

```
##  
## Pearson's Chi-squared test  
##  
## data: brfss$X_EDUCAG_FACTOR and brfss$SMOKDAY2_FACTOR  
## X-squared = 52200, df = 6, p-value < 2.2e-16
```

- Niveau de revenus et tabagisme

# EFFET DE LA DÉPRESSION SUR LE TABAGISME

```
##  
## Pearson's Chi-squared test  
##  
## data:  tab  
## X-squared = 52858, df = 8, p-value < 2.2e-16
```

- Humeur dépressive et tabagisme

```
##  
## Pearson's Chi-squared test  
##  
## data:  tab  
## X-squared = 71.777, df = 4, p-value = 9.565e-15
```

- Dépression et tabagisme

```
##  
## Pearson's Chi-squared test  
##  
## data:  tab  
## X-squared = 21789, df = 2, p-value < 2.2e-16
```

# EFFET DE LA DÉPRESSION SUR LE TABAGISME

## ANNEXE 2 : RÉSULTATS DE LA RÉGRESSION LOGISTIQUE AVEC LA FONCTION SUMMARY – EFFET DU TABAGISME SUR LA DÉPRESSION

```
##
## Call:
## glm(formula = ADDEPEV2 ~ X_RFSMOK3, family = quasibinomial(link = "logit"),
##      data = m.final, weights = m.final$weights)
##
## Deviance Residuals:
##      Min        1Q      Median        3Q        Max
## -2.8623  -0.8264  -0.5180  -0.3417   7.1308
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.278737   0.001643  -778.2  <2e-16 ***
## X_RFSMOK3    0.516042   0.003807   135.6  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 1.000039)
##
##      Null deviance: 2785552  on 2565239  degrees of freedom
## Residual deviance: 2767962  on 2565238  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
```

## ANNEXE 3 : RÉSULTATS DE LA RÉGRESSION LOGISTIQUE AVEC LA FONCTION SUMMARY – EFFET DE LA DÉPRESSION SUR LE TABAGISME

# EFFET DE LA DÉPRESSION SUR LE TABAGISME

```
Call: glm(formula = X_RFSMOK3 ~ ADDEPEV2, family = quasibinomial(link = "logit"),
  data = m.final, weights = m.final$weights)
```

```
Coefficients:
(Intercept)      ADDEPEV2
   -1.6164         0.5012
```

```
Degrees of Freedom: 2565235 Total (i.e. Null);  2565234 Residual
Null Deviance:      2431000
Residual Deviance: 2414000      AIC: NA
```

```
Call:
glm(formula = X_RFSMOK3 ~ ADDEPEV2, family = quasibinomial(link = "logit"),
  data = m.final, weights = m.final$weights)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1091  -0.7102  -0.5496  -0.3838   6.6433
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.616359   0.001873  -862.9  <2e-16 ***
ADDEPEV2     0.501159   0.003766   133.1  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for quasibinomial family taken to be 1.000004)
```

```
Null deviance: 2431284 on 2565235 degrees of freedom
Residual deviance: 2414377 on 2565234 degrees of freedom
AIC: NA
```

```
Number of Fisher Scoring iterations: 4
```