

Master 1 Econométrie appliquée

Cours d'économétrie linéaire avancé

**Analyse du temps d'écran chez les Français de plus de 18 ans
(sur un échantillons de 98 personnes en 2023)**

Jasmine Dupau, Isabel Palacio, Valorys Trillaud

Résumé :

L'objectif de notre étude est d'identifier les facteurs influençant la consommation numérique hebdomadaire chez les adultes Français, à partir de 18 ans, à l'échelle nationale. Nous avons fait appel à la méthode des moindres carrés (MCO) ou communément appelée : la régression linéaire, qui nous a permis d'obtenir un modèle fiable validant toutes les conditions requises pour son application. Pour ce faire, nous avons opté pour l'utilisation d'un jeu de données se rapprochant le plus possible de la réalité. Par conséquent, nous avons directement interrogé la population à l'aide d'un questionnaire se basant sur la modélisation faite au préalable à partir de la littérature. Nous avons obtenu un échantillon de 98 individus non représentatif dû à un ratio femmes/hommes élevé.

Le modèle nous a donné d'une part des résultats attendus. En effet, les personnes ayant une vie sociale importante ou ne travaillant pas toute la journée sur un écran ont une consommation numérique hebdomadaire moindre par rapport aux casaniers et ceux exerçant un métier pour lequel l'écran est le seul outil de travail. Puis, d'autre part, des résultats inattendus. En effet, une personne possédant un nombre d'appareils moindre a une consommation plus importante que celle ayant plus d'appareils ou encore les personnes n'ayant pas la fibre passent plus de temps sur les écrans que ceux ayant la fibre.

Abstract :

The aim of our study is to identify the factors influencing weekly digital consumption among French adults aged 18 and over, on a national scale. We used the method of the ordinary least squares (OLS) or, more commonly, linear regression, which allowed us to obtain a reliable model validating all the conditions required for its application. To do this, we opted for a dataset as close to reality as possible. As a result, we surveyed the population directly, using a questionnaire based on the modeling previously carried out in the literature. We obtained a non-representative sample of 98 individuals, due to a high female/male ratio.

On the one hand, the model gave us the expected results. Indeed, people with an important social life or who don't work all day on a screen have a lower weekly digital consumption than those who live at home or have a job for which the screen is the only work tool. Then, there were the unexpected results. People with fewer equipments consume more than

those with more numeric equipments, and people without fiber spend more time on screens than those with fiber.

Sommaire :

1. Introduction.....	3
2. Analyse des variables.....	4
3. Analyse statistique du jeu de données.....	22
4. Analyse économétrique.....	39
5. Conclusion et discussion.....	49
6. Annexes.....	51
7. Bibliographie.....	53
8. Table des matières.....	58

1. Introduction

Nous pouvons dire aujourd'hui que le numérique occupe une grande place dans notre quotidien grâce à Internet. En effet, le progrès numérique a bouleversé les habitudes des individus avec l'apparition de nouvelles infrastructures numériques ou encore de technologies, inconcevables au début du XXI^{ème} siècle. Par exemple, il était inimaginable au siècle dernier d'utiliser une montre pour connaître la météo ou suivre son rythme cardiaque car cet objet était initialement conçu pour lire l'heure. Par conséquent, la société a dû s'adapter à ces grandes innovations poussées par une dématérialisation rapide et une émergence continue de nouvelles technologies.

Par ailleurs, la crise sanitaire de 2020 a joué un rôle important dans l'adaptation du modèle économique des pays. En effet, les confinements ont d'une part incité les professionnels à se réinventer en proposant de nouveaux services comme la livraison à domicile pour les restaurants, la téléconsultation pour le médical ou encore le télétravail à l'aide de plateformes nettement utilisées aujourd'hui telles que Teams ou encore Zoom. D'autre part, cette pandémie a également impacté la manière de consommer dû à la fermeture de magasins non essentiels, des établissements scolaires et d'infrastructures culturelles tels les cinémas. Par conséquent, les jeunes ont particulièrement été sujets à une dépendance accrue des écrans avec les cours en distanciel et de la vidéo à la demande. Ce constat est appuyé par une étude dont l'un des coordinateurs de l'Unaf affirme que « *les écrans ont trouvé une place au sein de la famille et c'est difficile de revenir en arrière* » (Ouest-France, 2022).

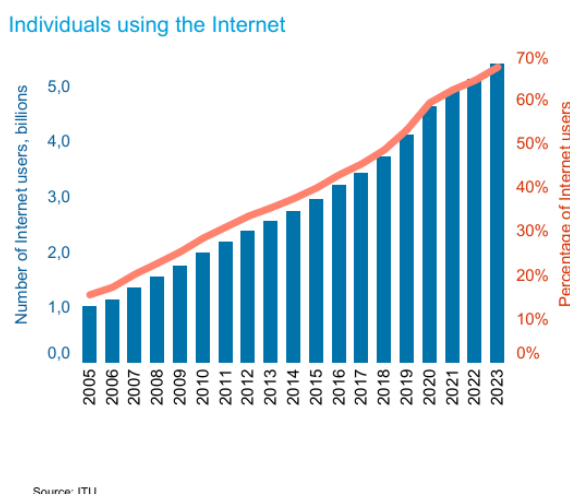


Figure 1 : évolution du nombre mondial d'internautes

En termes de chiffres clés, nous avons relevé plus de 5 milliards d'internautes représentant une part de 67% de la population mondiale en 2023 contre seulement 1 milliard en 2005 (voir figure ci-contre). Selon Statista et Internet lives stats, cette croissance exponentielle est aussi valable en France avec 60 millions d'utilisateurs soit 85% des français en 2023 contre seulement

8 millions soit 14% de la population française en 2000.

Ainsi, nous trouvions intéressant d'étudier ce phénomène à travers une application économétrique de la consommation hebdomadaire numérique des adultes français. Cette étude économétrique va nous permettre d'identifier les facteurs influençant positivement ou négativement le temps hebdomadaire des français passé sur les écrans.

Pour mener à bien notre étude, nous allons tout d'abord présenter les différentes variables trouvées à l'aide de notre intuition et de la littérature qui nous ont semblé nécessaires pour expliquer le temps hebdomadaire numérique. Dans un second temps, nous développerons le contexte de notre analyse. Ensuite, nous allons nous familiariser avec notre base de données à travers le nettoyage et les différentes statistiques effectuées. Enfin, nous réaliserons des estimations économétriques qui permettront de mettre en évidence les facteurs impactant ou non de manière positive ou négative la consommation numérique hebdomadaire des adultes français.

2. Analyse des variables.

Dans cette partie, nous allons étudier la pertinence des variables choisies pour expliquer le temps d'écrans. Dans un premier temps, nous expliquerons la variable à expliquer puis la raison de notre choix des différentes variables explicatives. Notre modèle peut être résumé par le schéma ci-dessous.

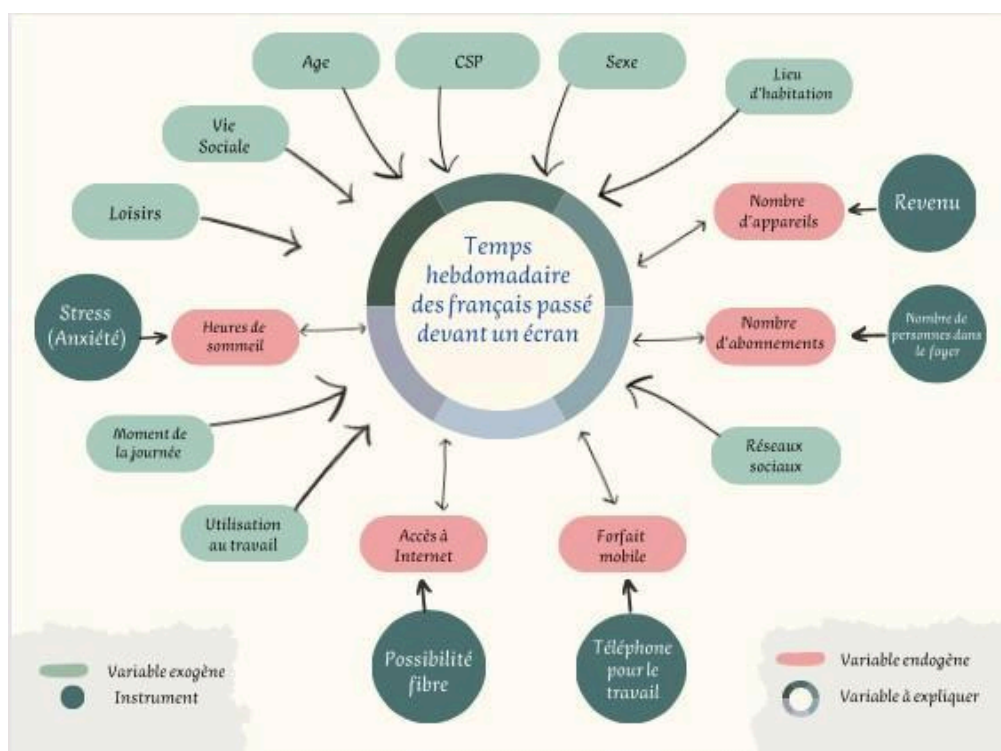


Figure 2 : schéma du modèle

2.1 Variable à expliquer.

Notre étude a pour but de comprendre l'évolution de la consommation numérique hebdomadaire chez les adultes venant de toute la France. Nous trouvons intéressant de mesurer de manière globale la place du numérique dans la vie des français puisque chacun est concerné par cette thématique comme nous vivons aujourd'hui dans un monde connecté. De plus, ce sujet est ouvert et vaste avec une variété d'études possibles selon la manière dont il est abordé. Dans notre cas, nous trouvons intéressant d'inclure les dimensions socio-économique et médicale.

La consommation numérique hebdomadaire correspond au temps moyen passé sur les écrans par semaine prenant en compte l'activité professionnelle. En termes de chiffres, le temps hebdomadaire moyen des Français est de 32 heures par semaine en 2022¹.

Par le mot « écrans », nous faisons référence à tous les équipements numériques utilisés par les individus pour s'informer, communiquer ou encore visionner du contenu tels que le téléphone portable, l'ordinateur portable, la tablette, la télévision, etc.

Nous allons par conséquent faire appel à plusieurs variables pour identifier les facteurs incitant les Français à faire usage de ces écrans.

2.2 Variables explicatives.

2.2.1) Réseaux sociaux.

Initialement conçus pour faciliter la communication et le partage entre amis et familles, les réseaux sociaux ont rapidement évolué pour devenir des outils puissants pour les entreprises, les médias et même la politique. Ils offrent aux utilisateurs la possibilité de publier des textes, des images, des vidéos, et d'interagir à travers des commentaires, des likes, et des partages.

¹ Leroy T, « En moyenne, les français regardent un écran 32 heures par semaine », 2023, BFMTV : https://www.bfmtv.com/tech/actualites/en-moyenne-les-francais-regardent-un-ecran-32-heures-par-semaine_AD-202301300358.html (consulté le 14/01)

Le paysage des réseaux sociaux est varié, allant de plateformes généralistes comme Facebook et X, à des plateformes plus spécialisées comme LinkedIn pour les professionnels, ou Instagram axé sur le partage de photos et de vidéos.

Chaque plateforme possède ses propres caractéristiques et est utilisée à des fins différentes, mais toutes partagent le même objectif fondamental de connecter les gens entre eux.

Toutefois, il est crucial de comprendre que l'omniprésence des réseaux sociaux ne vise pas uniquement à faciliter le partage; elle sert essentiellement à asseoir la pérennité de leur modèle économique, qui est principalement fondé sur la publicité. Que cela soit justifié ou non, ce modèle encourage une exploitation intensive de nos données personnelles et un profilage de plus en plus précis. Plus nous sommes connus par le réseau social, plus il peut augmenter la valeur de ses espaces publicitaires.

Ainsi, pour nous inciter à passer plus de temps sur la plateforme, les réseaux sociaux mettent en avant des contenus qui favorisent l'engagement des utilisateurs et à rester connectés.

Comme expliqué par les auteurs Abiteboul et Cattan dans leur article sur les réseaux sociaux et la régulation² « Notre attention est devenue un énorme marché ».

En France, pendant l'année 2022, 6 français sur 10 passent en moyenne 52 minutes par jour sur des réseaux sociaux et messageries, soit 7 minutes de plus qu'il y a un an.

L'usage des réseaux sociaux varie considérablement selon l'âge. Par exemple, les jeunes de 15 à 24 ans y consacrent en moyenne 2 heures et 19 minutes par jour. Cela équivaut à 61% du temps total qu'ils passent sur Internet chaque jour, une proportion en hausse par rapport à 55% en 2021 et 50% en 2020. En comparaison, l'internaute moyen dédie 'seulement' 37% de son temps de navigation sur Internet aux réseaux sociaux.³

² Abiteboul S. et J. Cattan, « Nos réseaux sociaux, notre régulation », RED, vol. 1, no. 1, 2020, pp. 36-44, consulté le 28/12 <https://www.cairn.info/revue-red-2020-1-page-36.htm#s1n1>

³ L'Année Internet 2022, consulté le 30/12/2023
<https://www.mediametrie.fr/fr/annee-internet-2022>

D'après la littérature, une corrélation positive entre notre variable illustrative 'la consommation hebdomadaire des écrans' et la variable explicative 'réseaux sociaux' est probable.

Cependant, un aspect particulièrement intéressant réside dans l'influence de cette dernière sur la consommation globale des médias. En effet, les réseaux sociaux ont, d'une certaine manière, modifié les habitudes de consommation d'information. Ils ont transformé la manière dont les contenus sont consommés, au point de menacer, par exemple, l'existence de la télévision publique⁴. Cette tendance incite les utilisateurs à se détourner progressivement des médias traditionnels au profit des réseaux sociaux pour s'informer et se divertir. Cette évolution a une conséquence remarquable : bien qu'elle contribue à augmenter le temps global passé devant les écrans, elle entraîne simultanément une réduction du temps consacré à la télévision.

2.2.2) L'âge.

Notre exposition aux écrans débute dès le plus jeune âge, généralement sous la supervision d'adultes. Cependant, dans le cadre de notre étude, nous avons délibérément choisi de nous concentrer sur la population de plus de 18 ans. Cette décision a été prise afin de cibler spécifiquement les comportements et les habitudes d'utilisation des écrans chez les adultes, sans les influences possibles de la surveillance ou du contrôle parental.

L'hypothèse qui nous amène à intégrer l'âge dans nos variables explicatives est basée sur l'approche des explications générationnelles.

Cette hypothèse repose sur l'idée que les expériences uniques vécues par chaque génération pendant ses années de formation (adolescence et études) ont une influence déterminante sur les comportements et habitudes ultérieurs, notamment en ce qui concerne la consommation des écrans.

De la génération des baby-boomers (nés entre 1946 et 1964) à la génération Alpha (nés après 2010), en passant par les générations X (1965-1980), Y (1981-1996) et Z

⁴ Sonnac N, "Quelle place pour l'audiovisuel public face aux réseaux sociaux ?", consulté le 30/12/2023 <https://www.vie-publique.fr/parole-dexpert/290432-quelle-place-pour-laudiovisuel-public-face-aux-reseaux-sociaux>

(1997-2009), chaque groupe présente des comportements distincts vis-à-vis de la technologie.

Les circonstances et contextes dans lesquels chaque génération a grandi jouent un rôle majeur dans la formation de leurs habitudes plus tard dans la vie, notamment en ce qui concerne l'utilisation des technologies numériques. Par exemple, bien que les usages numériques soient devenus prédominants chez les jeunes, avec une utilisation intensive des vidéos en ligne, des réseaux sociaux et des jeux vidéo, cette tendance n'a pas éliminé leur intérêt pour les activités en dehors du monde numérique. Les jeunes, notamment ceux âgés de 15 à 24 ans, continuent de fréquenter activement les lieux culturels tels que les cinémas, les théâtres, les bibliothèques et les sites patrimoniaux⁵.

En apparence, il pourrait sembler que les jeunes générations soient les plus grandes consommatrices d'écrans, alimentant ainsi les critiques des adultes et les discussions sur un 'fossé générationnel'⁶. Cependant, cette perspective simpliste ne rend pas justice à la complexité de la situation. L'utilisation des écrans ne se limite pas aux smartphones, mais englobe également la télévision et les ordinateurs. De plus, des études récentes ont montré que l'adoption des nouvelles technologies par les seniors, qui a augmenté pendant le confinement, persiste. Il est à noter que les adultes plus âgés peuvent également passer jusqu'à 10 heures par jour sur divers appareils numériques⁷.

Ainsi, nous considérons que l'âge est un indicateur essentiel pour comprendre les différences dans l'utilisation et l'interaction des écrans. Il joue un rôle significatif dans la détermination des habitudes de consommation des écrans, mettant en évidence la complexité et la diversité des usages à travers les différentes générations.

Pour notre variable explicative 'âge', nous avons posé la question suivante aux participants : 'À quelle catégorie d'âge appartenez-vous ?' Ensuite, nous avons classé les réponses en utilisant les tranches d'âge définies dans notre cours d'économétrie linéaire:

⁵ P. Lombardo, L. Wolff, "Cinquante ans de pratiques culturelles en France", consulté le 30/12/2023 <https://www.cairn.info/revue-culture-etudes-2020-2-page-1.htm>

⁶ A. Cournoyer, "Internet et les écrans: le fossé des générations", consulté le 31/12/2023 <https://pausetonecran.com/internet-et-les-ecrans-le-fosse-des-generations/>

⁷ J. Jargon, "Les seniors aussi peuvent être accros aux écrans", consulté le 02/01/2024 <https://www.lopinion.fr/economie/les-seniors-aussi-peuvent-etre-accros-aux-ecrans>

18- 24

Nés entre 1999 et 2005. Ils appartiennent principalement à la Génération Z (nés approximativement entre 1997 et 2012)

25-49

Nés entre 1974 et 1998. Cette tranche d'âge couvre la fin de la Génération X (nés approximativement entre 1965 et 1980) et la majorité de la Génération Y ou Millennials (nés approximativement entre 1981 et 1996)

50 -69

Nés entre 1954 et 1973. Cette tranche d'âge inclut la fin de la Génération du Baby-Boom (nés approximativement entre 1946 et 1964) et le début de la Génération X.

70 et +

Nés en 1953 ou avant.

Ils appartiennent principalement à la Génération du Baby-Boom ou même à la génération précédente, souvent appelée la Génération Silencieuse (nés avant 1946).

2.2.3) Le forfait mobile et son instrument le téléphone portable professionnel.

Dans notre étude, nous avons interrogé les participants sur leurs dépenses mensuelles pour leur forfait mobile, avec des options de réponse allant de 0€ à plus de 35€. Cette question nous aide à comprendre l'intérêt des utilisateurs pour les fonctionnalités avancées du téléphone, qui se manifeste souvent par des forfaits plus coûteux offrant plus de données internet et d'autres avantages.

Nous avons inclus la variable des dépenses mensuelles de forfait mobile dans notre étude car, selon notre hypothèse, un montant plus élevé consacré au forfait mobile pourrait indiquer une préoccupation accrue pour le bon fonctionnement et les fonctionnalités du téléphone portable. Cela peut inclure l'accès à un appareil plus récent, une plus grande quantité de données internet, une utilisation à l'étranger, entre autres.

En d'autres termes, nous supposons qu'une dépense plus importante dans le forfait mobile est potentiellement liée à une utilisation plus intensive et plus grande des fonctionnalités du téléphone.

Nous avons choisi de nous concentrer sur le montant du forfait mobile car le smartphone est devenu un outil central dans la gestion des petits temps quotidiens, tels que les trajets en transports en commun ou l'attente dans les files. De plus, son rôle s'étend à des tâches de plus en plus complexes, comme effectuer des virements bancaires ou gérer des aspects liés à notre santé.

En effet, les forfaits mobiles les plus coûteux facilitent l'accès à Internet en tout lieu, à tout moment, et pour effectuer des tâches plus complexes ou personnelles, grâce à l'accès à des données mobiles, assurant à l'utilisateur de ne pas être restreint dans son usage par le réseau wifi ou des données insuffisantes.

En 2022, 95% des personnes de 12 ans ou plus résidant en France métropolitaine disposent d'un téléphone mobile. De plus, le smartphone s'impose comme la norme : 87% des Français de 12 ans et plus ont désormais un smartphone⁸.

D'après notre intuition, les utilisateurs disposant de forfaits mobiles supérieurs ont tendance à utiliser davantage leur smartphone pour accéder à du contenu en ligne, comme des vidéos en streaming ou diverses applications, surtout lorsqu'ils sont en dehors de chez eux puisqu'ils ont accès à davantage de données mobiles. Cette facilité d'accès à internet se traduit par une augmentation de leur temps passé devant les écrans.

Selon l'Arcep (Autorité de régulation des communications électroniques, des postes et de la distribution de la presse), au 30 juin 2023, le prix moyen d'un forfait mobile en France était de 15€, ce qui le positionne comme le deuxième prix moyen le plus avantageux de l'Union européenne.⁹ Cette tarification compétitive reflète non seulement l'évolution de la consommation, mais aussi l'importance croissante des smartphones dans la vie quotidienne des Français. Cela suggère une accessibilité accrue à la technologie mobile,

⁸ BAROMÈTRE DU NUMÉRIQUE - édition 2022 – RAPPORT (30 janvier 2023), page 29, consulté le 31/12
https://www.arcep.fr/uploads/tx_gspublication/rapport-barometre-numerique-edition-2022-Rapport.pdf

⁹ Gaudiaut T, "Le coût de l'Internet mobile dans le monde", consulté le 01/01/2024
<https://fr.statista.com/infographie/29106/comparaison-cout-internet-mobile-dans-le-monde-prix-moyen-1-go-donnees-forfaits-mobiles/>

répondant ainsi aux besoins variés des consommateurs pour la connectivité et les services numériques.

En intégrant la variable 'forfait mobile' dans notre modèle, nous considérons la possibilité d'une simultanéité entre cette variable explicative et notre variable illustrative 'La consommation hebdomadaire des écrans'. En d'autres termes, le montant alloué au forfait mobile peut influencer le temps passé devant l'écran. De plus, une utilisation accrue des écrans peut amener à accorder plus d'importance au budget dédié à cette utilisation. Cette interaction bidirectionnelle peut créer un problème d'endogénéité, impactant la précision et la validité de nos résultats.

Pour approfondir notre compréhension de l'impact des téléphones portables dans le cadre professionnel, notre étude a inclus une question spécifique adressée aux participants : "Dans le cadre de votre activité professionnelle ou de vos études, utilisez-vous votre téléphone portable ?" Cette question vise à déterminer l'importance et la fréquence de l'utilisation du téléphone portable dans un contexte professionnel ou académique, ce qui peut fournir des insights précieux sur le rôle des smartphones dans la productivité, la communication et la gestion du temps.

Dans notre modèle, l'usage du téléphone professionnel est employé comme variable instrumentale, sous l'hypothèse d'une corrélation avec le choix du forfait mobile personnel, tout en étant indépendant des erreurs de l'équation de consommation d'écrans. Cette approche est justifiée car l'utilisation fréquente d'un téléphone pour le travail peut orienter vers des forfaits mobiles personnels plus étendus, répondant ainsi aux besoins professionnels de communication. Cette spécificité rend le téléphone professionnel un instrument efficace pour isoler l'impact de la variable endogène "prix du forfait mobile" sur la consommation d'écrans.

2.2.4) Le nombre d'appareils numériques et son instrument le revenu.

Dans l'analyse de la consommation d'écrans, nous reconnaissons que le terme "écran" évoque une gamme étendue d'appareils et d'expériences. Autrefois réservés à la diffusion de contenus médiatiques traditionnels, les écrans se retrouvent aujourd'hui au

cœur de nos interactions avec des technologies variées, de la navigation Internet à la gestion de la domotique.

Cette généralisation, qui inclut désormais des interfaces interactives avec diverses machines (imprimantes, voitures, réfrigérateurs, etc), illustre parfaitement l'extension et la complexification de notre environnement numérique.

Dans notre étude, nous nous concentrons sur les appareils qui facilitent la transmission, l'enregistrement, la création, le partage ou l'échange d'informations. Cela inclut non seulement la télévision, qui a longtemps été le média prédominant, mais aussi un large éventail d'appareils comme les ordinateurs, les tablettes et les smartphones. Ces derniers sont souvent classés dans la catégorie des Technologies de l'Information et de la Communication (TIC)¹⁰, formant un écosystème technologique étendu et intégré.

Notre variable "nombre d'appareils numériques" cherche à capturer cette multiplicité des points d'accès numériques TIC qui, format de poche ou écran géant, redéfinissent notre consommation hebdomadaire des écrans.

Le nombre d'appareils auxquels un individu a accès peut influencer directement la quantité de temps passé devant un écran. Plus il y a d'appareils disponibles, plus les occasions de consommer du contenu numérique sont nombreuses.

Selon le baromètre numérique de 2022, les activités numériques sont devenues plus fréquentes durant les temps libres, avec une augmentation significative dans divers domaines : navigation internet (70%, +40 points), visionnage de vidéos (50%, +37 points), écoute de musique (63%, +12 points), et jeux vidéo (47%, +23 points). Cette tendance croissante vers le numérique semble éclipser en partie les activités non-numériques¹¹.

Parallèlement, le nombre d'appareils numériques dans les foyers continue de s'accroître, comme l'illustre la figure ci-dessous:

¹⁰ Technologies de l'information et de la communication, consulté le 02/01/2024

<https://uis.unesco.org/fr/glossary-term/technologies-de-linformation-et-de-la-communication-tic>

¹¹ BAROMÈTRE DU NUMÉRIQUE - édition 2022, page 9 - consulté le 31/12/2023
https://www.arcep.fr/uploads/tx_gspublication/rapport-barometre-numerique-edition-2022-Rapport.pdf

Evolution du taux d'équipement des répondants (%)

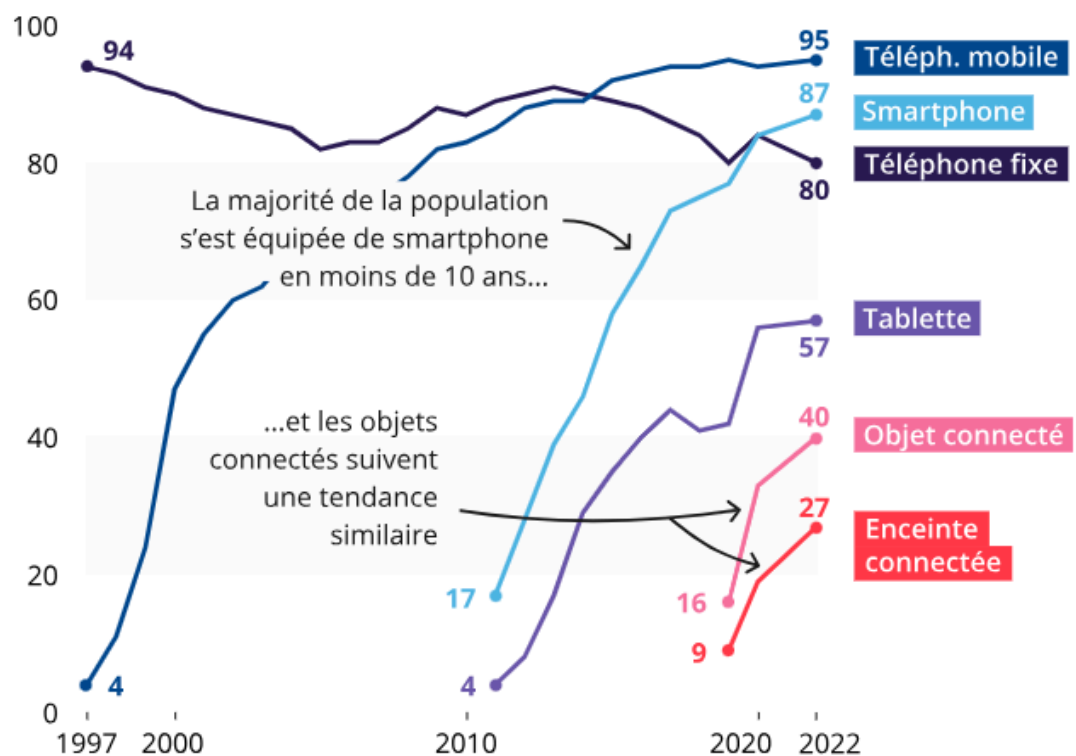


Figure 3 : évolution du taux d'équipement, source l'arcep, le baromètre du numérique janvier 2023

Nous pourrions être confrontées à une causalité inverse ou à une simultanéité concernant l'usage de multiples appareils. L'accès à un grand nombre d'appareils peut augmenter le temps passé devant les écrans. Inversement, une consommation accrue des écrans pourrait conduire à l'acquisition de plus d'appareils. Cette relation réciproque présente un potentiel problème d'endogénéité.

Notre variable "Revenu", issue du Revenu Fiscal de Référence (RFR) des foyers, est utilisée dans notre modèle comme un instrument pour corriger l'endogénéité de la variable "Nombre d'appareils numériques".

Les données de RFR, allant de moins de 10 000€ à plus de 80 000€, sont extraites des informations fournies par les participants. Nous supposons qu'un revenu plus élevé permet un accès à un plus grand nombre d'appareils, mais n'influence pas à priori le nombre

d'heures consacré aux écrans. Cette approche nous permet de contrôler l'impact potentiellement biaisé de la variable endogène.

2.2.5) Vie sociale.

Une variable que nous avons considérée comme importante est le temps passé avec sa famille ou ses amis en dehors d'un intermédiaire comme un téléphone ou tout autre écran. Nous avons résumé cette variable par la vie sociale. La vie sociale est un élément important de notre société. En effet, elle « influe sur le niveau de bonheur et augmente l'espérance de vie, elle renforce l'immunité, réduit la dépression et les troubles de l'anxiété »¹². Nous avons choisi cette variable car nous sommes parties du principe que plus on passe de temps avec les autres, moins nous passons de temps sur les écrans, car nous sommes occupés à faire d'autres activités. Cependant, différentes études ont montré que la hausse du temps consacré aux écrans favorise le risque de se replier sur soi-même entraînant ainsi moins d'interactions sociales. En effet, « l'utilisation excessive des écrans peut entraîner un isolement social »¹³. Cette variable doit donc être considérée comme endogène c'est-à-dire que le temps passé avec son entourage impacte le temps passé sur les écrans et inversement. Il faut donc une variable instrument pour avoir un modèle prenant en compte cette endogénéité. Afin de pallier ce problème, l'utilisation d'un instrument est importante. Cependant, nous n'avons pas anticipé de questions dans notre questionnaire. Pour résoudre ce manquement, nous avons pensé que nous pourrions utiliser le stress comme instrument. En effet, «une revue a révélé que les personnes ayant des cercles sociaux restreints avaient des niveaux plus élevés de cortisol, l'hormone du stress »¹⁴. Il faudra pour cela garder une seule des variables endogènes que l'on peut expliquer par le stress dans le modèle final. Le stress ayant été mesuré dans notre questionnaire pour anticiper l'endogénéité de la variable sommeil qui sera abordée dans la suite.

¹²<https://ensemble.aesio.fr/aesio-mag/conseils-vie-sociale-epanouie#:~:text=Une%20vie%20sociale%20est%20faite,autres%20et%20savoir%20les%20%C3%A9couter>. (consulté le 22/01/24)

¹³<https://fondation-ramsayssante.com/questions/addiction-aux-ecrans/addict-aux-ecrans-queelles-sont-les-consequences-sociales#:~:text=L'utilisation%20excessive%20des%20%C3%A9crans,enfants%20les%20rendraient%20moins%20sociaux>. (consulté le 22/01/24)

¹⁴ <https://www.livi.fr/en-bonne-sante/isolement-social/> (consulté le 27/07/2024)

Pour la variable vie sociale, nous avons demandé au participant, « En moyenne, en dehors de votre activité professionnelle, de vos études et des réseaux sociaux, combien de temps consacrez-vous à votre vie sociale (famille, amis, autres) par semaine (en heures) ? ».

2.2.6) Loisir

Nous avons voulu connaître le temps attribué au loisir en dehors des écrans car depuis le confinement « Les Français passent désormais plus de 60% de leur temps devant un écran »¹⁵. Le temps passé à pratiquer une activité hors des écrans peut alors réduire le temps d'écrans. Une personne très investie dans ses loisirs pourrait avoir tendance à moins utiliser les écrans. En effet, pour quelqu'un consacrant par exemple quatre heures par semaine à des activités sportives représentent autant de périodes où la personne n'est pas devant un écran. Nous pensons donc que le temps consacré au loisir peut diminuer le temps passé sur les écrans. Les loisirs peuvent être de nature très diverse, la lecture, les activités sportives, la cuisine ou toute autre activités pratiquées.

Pour notre variable loisir nous avons demandé aux participants : « Combien d'heures passez-vous par semaine à pratiquer une activité en dehors des écrans ? (Sports, lecture, cuisine) ».

2.2.7) Nombre d'abonnements et son instrument le nombre de personnes dans le foyer.

Les abonnements ont une place de plus en plus importante depuis plusieurs années. En effet, les abonnements numériques font partie intégrante de la façon de consommer les écrans aujourd'hui. Prendre en compte le nombre d'abonnements paraît alors cohérent pour expliquer le temps d'écran. Nous avons défini les abonnements comme tous les abonnements payant que ce soit pour la presse, les plateformes de streaming vidéo ou les chaînes de télévision. Nous avons pris en compte toutes les styles d'abonnement pour prendre en compte les divers centres d'intérêts possibles du film aux informations en

¹⁵https://www.bfmtv.com/sante/les-francais-passent-plus-de-60-de-leur-temps-libre-devant-un-ecran-selon-un-e-etude_AN-202204070485.html#:~:text=RMC%20D%C3%A9couverte-,Les%20Fran%C3%A7ais%20passent%20plus%20de%2060%25%20de%20leur%20temps%20libre,un%20%C3%A9cran%2C%20selon%20une%20%C3%A9tude&text=De%20la%20t%C3%A9l%C3%A9vision%20au%20smartphone,tendance%20acc%C3%A9l%C3%A9r%C3%A9e%20par%20les%20confinements (consulté le 22/01/24)

passant par la musique. Ce choix repose sur l'hypothèse que la diversité des abonnements, comme l'association d'un abonnement sportif payant à plusieurs plateformes de streaming, pourrait encourager une utilisation prolongée des écrans. En effet, cette variété d'options offre en permanence du contenu nouveau à découvrir, ce qui pourrait inciter à passer plus de temps devant un écran.

Toutefois, nous supposons que cette variable est en partie endogène, c'est-à-dire qu'elle peut être autant une cause qu'une conséquence du temps passé devant les écrans. En d'autres termes, plus de temps passé sur les écrans pourrait motiver à rechercher de nouveaux contenus, conduisant ainsi à l'acquisition de nouveaux abonnements. Pour atténuer cette endogénéité, nous avons choisi d'utiliser le nombre de personnes vivant dans le foyer comme une variable instrumentale pour le nombre d'abonnements.

Nous partons du principe que la taille du foyer pourrait influencer le nombre d'abonnements souscrits. Par exemple, un foyer composé de plusieurs personnes pourrait nécessiter une plus grande diversité d'abonnements pour satisfaire les préférences et les intérêts variés de ses membres. Cette approche instrumentale nous permet d'examiner plus précisément l'impact du nombre d'abonnements sur le temps passé devant les écrans en prenant en compte le contexte familial.

Nous avons alors posé deux questions pour la variable endogène et son instrument. Les deux questions sont respectivement : « Combien d'abonnements payants avez-vous? (netflix, ebooks kindle, presse, etc) » et « Avec combien de personnes vivez-vous dans votre logement principal ? (Par exemple si vous êtes en couple répondez 1) ».

2.2.8) Genre

Nous avons cherché à expliquer le temps d'écran par le genre des individus. Tout d'abord, les différences entre les genres se font dès le plus jeune âge, l'appropriation des outils numériques ne bénéficiant pas de la même approbation selon le genre. En effet, « dès l'enfance, la prise en main des outils numériques n'est pas valorisée de la même manière chez les filles et les garçons »¹⁶. Ces disparités dans l'accès et l'utilisation des appareils

¹⁶ https://www.arcep.fr/uploads/tx_gspublication/rapport-barometre-numerique-edition-2022-Rapport.pdf (consulté le 27/01/2024)

électroniques dès le plus jeune âge peuvent exercer une influence profonde sur les habitudes des hommes et des femmes quant à l'utilisation des écrans à l'âge adulte. En effet, selon une précédente étude, les hommes passeraient plus de temps sur les écrans que les femmes, « Les femmes sont moins consommatrices d'écrans que les hommes. 34% des hommes passent au moins cinq heures par jour devant les écrans, alors que cette proportion n'est que de 26% chez les femmes ». Ces constats montrent l'importance du genre dans le temps que l'on peut passer sur un écran, cela nous permet de penser que cette variable est importante à prendre en compte dans notre étude.

Pour notre variable lieu d'habitation nous avons demandé au participants : « A quel genre vous identifiez-vous ? », les réponses proposées étaient : « Homme », « Femme », « Autre » et « Ne souhaite pas répondre ».

2.2.9) Lieu d'habitation.

La considération du lieu de résidence s'avère être une composante cruciale dans notre analyse, étant donné son impact sur l'accès aux ressources technologiques. Nous avons pris en compte cette variable car les personnes vivant dans un milieu rural ont un accès à un ordinateur plus faible que les personnes vivant dans une agglomération de plus de 2 000 (« 78% des personnes résidant dans des communes rurales, 86% des personnes résidant dans des agglomérations comptant entre 2 000 et 20 000 habitants »¹⁷). Prendre en compte le lieu d'habitation peut alors être cohérent, les personnes vivant en ville auront alors un accès plus facile à un ordinateur ou à internet, ils pourront alors passer facilement plus de temps devant des écrans.

Cette variable de localisation géographique revêt ainsi une importance majeure, car elle peut être un indicateur crucial de l'opportunité d'accès aux outils numériques. Les habitants des zones urbaines, par exemple, bénéficient généralement d'une connexion Internet plus rapide et d'une accessibilité accrue aux ordinateurs. En effet, « Les habitants de zones rurales sont moins satisfaits que les autres (- 6 points par rapport à la moyenne pour les données mobiles à l'intérieur du logement et - 8 points pour les données mobiles à l'extérieur du logement) ». Par conséquent, les habitants des zones urbaines auront plus de

¹⁷ https://www.arcep.fr/uploads/tx_gspublication/rapport-barometre-numerique-edition-2022-Rapport.pdf (consulté le 27/01/2024)

facilité à se retrouver immergés dans l'utilisation des écrans pour diverses activités telles que l'éducation, le divertissement ou le travail.

Pour notre variable lieu d'habitation nous avons demandé au participants : « Où habitez-vous ? », les réponses proposées étaient : « Ville », « Périurbaine/Banlieue » et « Campagne ».

2.2.10) Catégorie socio-professionnelle.

Nous avons pris en considération la catégorie socioprofessionnelle (CSP) car il existe une corrélation entre la CSP des parents et la durée d'utilisation des écrans chez leurs enfants. Les enfants issus de milieux sociaux plus défavorisés ont tendance à passer davantage de temps devant les écrans, « Presque deux heures par jour de consommation séparent les tweens les plus favorisés (plus de 100 000 dollars de revenus par an pour le foyer) des plus pauvres (moins de 35 000 dollars par an) »¹⁸ (Tweens : jeunes de 8 à 12). Ainsi, notre objectif était d'évaluer si ces disparités observées entre les milieux sociaux, mesurées dans notre étude grâce aux catégories socioprofessionnelles, continuent d'influencer le temps d'écran une fois que les individus atteignent l'âge adulte.

Il est plausible de considérer que ces différences persistent, mais il est également pertinent d'envisager une autre perspective. Par exemple, les professions exigeant un travail de bureau ou des responsabilités de cadres supérieurs pourraient être associées à une utilisation plus intensive des écrans en raison de la nature même de ces emplois, où l'ordinateur est un outil de travail constant. En revanche, les professions manuelles qui ne nécessitent pas une interaction soutenue avec un ordinateur pourraient être associées à une utilisation moins fréquente des écrans.

Ainsi, l'influence de la catégorie socioprofessionnelle sur la durée d'écran des individus est un aspect complexe que nous avons cherché à explorer, tenant compte à la fois des disparités socio-économiques antérieures et des exigences professionnelles qui pourraient impacter différemment l'utilisation des écrans chez les adultes.

¹⁸<https://www.blogdumoderateur.com/les-enfants-pauvres-passeraient-deux-heures-de-plus-par-jour-devant-un-ecran-que-les-enfants-riches/#:~:text=Un%20temps%20d'%C3%A9cran%20quotidien%20consid%C3%A9rable&text=S'il%20s'agit%20d,tweens%2C%20h27%20pour%20les%20teens> (consulté le 26/01/2024)

Pour notre variable catégorie socioprofessionnelle nous avons demandé aux participants : « Quelle est votre catégorie socioprofessionnelle ? », les réponses proposées étaient : «Agriculteurs exploitants », « Artisan/commerçant/Chef d'entreprise », « Cadres et professions intellectuelles supérieures » , « Professions intermédiaires », « Employés », « Ouvriers », « Retraités » et « Autres personnes sans activité professionnelle (par exemple les chômeurs ou les étudiants) ».

2.2.11) Sommeil et son instrument le stress.

Le sommeil est vital pour notre organisme, occupant plus d'un tiers du temps de notre vie. Une bonne qualité de sommeil est essentielle pour être en bonne santé car celui-ci agit directement sur le cerveau : « Il est déterminant pour la croissance, la maturation cérébrale, le développement et la préservation de nos capacités cognitives »¹⁹.

Cependant, la qualité du sommeil est mise à mal aujourd'hui. En effet, l'article²⁰ stipule que cette dégradation s'explique par une réduction du temps de sommeil, favorisée par l'habitude des français à se coucher de plus en plus tard en semaine. Cette dernière est apparue avec la perpétuelle évolution du mode de vie amplifiée par les nouvelles technologies comme l'intelligence artificielle. Par conséquent, les individus vivent dans un monde où le sommeil est vu comme une « perte de temps » selon la rédactrice. Ce constat est appuyé par la recommandation des spécialistes d'avoir un « couvre feu digital »²¹ pour améliorer la qualité du sommeil face aux chiffres alarmants : 44% des français se connectent dans leur lit avant d'éteindre pour dormir²².

Ainsi, le sommeil et la consommation numérique ont une relation négative. En effet, plus un individu trouvera le sommeil rapidement, moins il sera tenté d'utiliser un appareil se

¹⁹ Institut national du sommeil et de la vigilance, « SOMMEIL, un carnet pour mieux comprendre » :https://sante.gouv.fr/IMG/pdf/Sommeil_un_carnet_pour_mieux_comprendre.pdf (consulté le 14/01)

²⁰ Henrich E., « La baisse inquiétante du temps de sommeil des Français : comment y faire face ? », 2019 : <https://youmatter.world/fr/temps-sommeil-francais-baisse-conseils/> (consulté le 14/01)

²¹ « Un "couvre-feu digital" d'une heure avant de se coucher permettrait de mieux dormir », 2019, madame le figaro : <https://madame.lefigaro.fr/bien-etre/bien-dormir-linstitut-national-du-sommeil-recommande-un-couvre-feu-pour-les-ecrans-sondage-190319-164350> (consulté le 14/01)

²² Enquête publiée par l'institut national du sommeil en partenariat avec le groupe mgen sur un échantillon d'un échantillon représentatif de 1014 personnes de 18 à 75 ans : https://institut-sommeil-vigilance.org/wp-content/uploads/2019/10/DP_JS2019.pdf (consulté le 14/01)

trouvant à proximité. De plus, nous suspectons une endogénéité car la majorité des articles ou études trouvées sur internet affirment que « les écrans perturbent le sommeil »²³.

Pour pallier ce problème, nous avons considéré le stress comme une variable instrumentale du sommeil. En effet, un sommeil perturbé peut aussi être dû à des troubles d'anxiété qui favorisent un endormissement plus lent voire même des réveils nocturnes²⁴.

2.2.12) Moment de la journée.

La consommation numérique fluctue d'une personne à l'autre suivant le quotidien de chacun et a tendance à être plus importante le soir. En effet, une grande majorité de français consacre leur soirée à utiliser les écrans jusque dans leur lit. Cela s'explique par le fait qu'on a majoritairement plus de temps libre le soir. Ce phénomène est encouragé par exemple par le déploiement de multiples plateformes de streaming ou encore de réseaux sociaux qui permettent à l'individu de s'éloigner des tracas du quotidien. En effet, un individu fatigué de sa journée, va naturellement se tourner vers les activités avec écrans car elles ne nécessitent pas d'efforts au dépend des activités hors écrans comme la lecture, qui, quant à elles nécessitent de la concentration. Ce processus est appelé « paresse intellectuelle »²⁵.

Ainsi, le moment de la journée a un effet positif sur la consommation numérique.

2.2.13) Utilisation d'écran du travail.

La mesure de la consommation numérique prend aussi en compte l'utilisation des écrans dans le cadre professionnel. En effet, l'omniprésence du numérique a poussé les entreprises à se digitaliser et a fait émerger une nouvelle catégorie d'emplois sur le marché du travail introduisant plusieurs secteurs d'activités, appelés « métiers du numérique ». Nous pouvons y trouver les data analyst, data scientists, les développeurs de logiciels, les

²³ « Sommeil et écrans : les dessous du problème », Alptis : <https://www.alptis.org/complementaire-sante/sommeil/ecran/> (consulté le 14/01)

²⁴ « Le lien étroit entre le stress et les troubles du sommeil », 2022, Le figaro: <https://www.lefigaro.fr/services/le-lien-etroit-entre-le-stress-et-les-troubles-du-sommeil-20220228> (consulté le 14/01)

²⁵ « Les Français passent la majorité de leur temps libre devant les écrans, selon une étude », 2022, Ouest-France: <https://www.ouest-france.fr/societe/les-francais-passent-la-majorite-de-leur-temps-libre-devant-les-ecrans-selon-une-etude-7715659> (consulté le 14/01)

influenceurs ou encore les youtubers. Tous ces métiers ont comme essence l'usage des écrans. Ce constat est appuyé par l'étude réalisée en 2021²⁶ estimant que les Français consacraient en moyenne 56 heures de leur temps par semaine sur Internet dont 20 heures au travail. De plus, la crise sanitaire a aussi contribué à la hausse de la consommation numérique par le biais du télétravail, unique solution pour faire tourner l'activité des entreprises lorsque cela était possible. Aujourd'hui, les entreprises proposent différentes formes de la « journée type » du salarié tel que le mode hybride comptabilisant 53% des salariés²⁷.

Ainsi, l'utilisation d'écrans dans le cadre professionnel a un effet positif sur la consommation numérique.

2.2.14) Accès à internet et son instrument la fibre.

L'usage des écrans nécessite l'accès à Internet. En effet, une personne n'ayant pas de connexion internet chez soi est obligée de se tourner vers les activités hors écrans comme la lecture, le jardinage, les sorties ou encore le ménage. Cela s'explique par le fait qu'Internet a joué le rôle de précurseur dans l'essor du numérique dès les années 50. Ce réseau mondial de communication a incité les ingénieurs et chercheurs à mettre en place de nombreuses innovations telles que les Web 2 et 3 incluant les plateformes²⁸. Autrement dit, si nous assimilons un équipement numérique à une voiture, l'internet serait le carburant. Ainsi, s'il n'y pas d'internet, il n'y a pas d'intérêt pour les personnes souhaitant pratiquer des activités nécessitant une connexion comme lire la presse, regarder des vidéos en ligne ou encore communiquer via les réseaux sociaux, de faire appel aux écrans.

Ainsi, l'accès à internet a un effet positif sur la consommation numérique. Cependant, nous supposons la présence d'endogénéité car une personne ayant une forte consommation numérique implique logiquement une forte consommation de la connexion internet. Pour pallier ce problème, nous définissons l'accès à la fibre comme étant une

²⁶ « Combien de temps de nos vies passons-nous en ligne », 2021, NordVPN : <https://nordvpn.com/fr/blog/temps-passe-en-ligne/> (consulté le 14/01)

²⁷ « Bilan de l'adoption du télétravail en France ? », 2023, SFR business : <https://www.sfrbusiness.fr/room/communications-unifiees/teletravail-bilan-et-perspectives.html> (consulté le 14/01)

²⁸ Cours d'économie numérique

variable instrumentale de la connexion internet. En effet, le débit de la connexion dépend des technologies utilisées comme la fibre qui permet d'avoir une connexion plus rapide comparée à l'ADSL. Par conséquent, une personne ayant un très bon débit de connexion va être incitée à faire usage des écrans.

3. Analyse statistique du jeu de données.

Tous les tableaux et graphiques introduits dans la section sont issus du logiciel R.

3.1 Présentation de la base de données.

Le jeu de données est composé de 98 individus dont 72 femmes et 26 hommes venant de toute la France (voir annexe 1). Il est issu d'un questionnaire élaboré par nos soins avec la plateforme Google forms. Ce dernier a été diffusé sur nos réseaux sociaux puis partagé sur d'autres supports par les répondants.

Avant d'exposer les résultats statistiques, nous trouvons intéressant de détailler le processus effectué pour préparer au mieux la base de données avant l'exploitation de celle-ci.

3.1.1) Représentativité de notre échantillon

Afin de faire l'analyse de la consommation des écrans dans la population française de plus de 18 ans, nous avons diffusé un questionnaire. Nous l'avons diffusé via nos proches et sur nos réseaux sociaux, ce qui amène à un premier biais car notre entourage est composé principalement de jeunes, d'étudiants, ou de personnes plus âgées habitant en campagne. Nous avons ainsi pu obtenir 98 réponses à notre questionnaire.

Nous allons maintenant tester la représentativité de notre échantillon via la distribution homme femme.

Figure 4 : Répartition hommes/femmes de notre échantillon

	Hommes	Femmes	Total
Échantillon observé	72	26	98
% de l'échantillon	73,5 %	26,5 %	100 %

Figure 5 : Répartition hommes/femmes en France

Genre	Hommes	Femmes	Total
Population en %	52,13 %	47,87%	100 %

Genre	Hommes	Femmes	Total
Effectifs observés	72	26	98
Effectifs théoriques	51,09 (1)	46,91	98

Calcul des effectifs théoriques :

$$(1) = 52,13\% * 98$$

Test du Khi-deux :

$$Khi2 \text{ théorique (pour un seuil de risque à 5\%)} = 3,84$$

3.1.2) Nettoyage et traitement des valeurs manquantes.

Le processus de nettoyage de la base de données est nécessaire après la mise en ligne d'un questionnaire. En effet, les répondants ont pu répondre aux questions avec des valeurs non souhaitées comme l'exemple de « ça dépend des jours » lorsque l'on voulait une valeur numérique ou encore des valeurs telles que « ne souhaite pas répondre ». Nous avons choisi de mettre ce type de réponses en NA (Not Available). Enfin, nous avons dû faire quelques ajustements tels que remplacer les données des cellules par la moyenne des valeurs lorsqu'on avait affaire à des encadrements de deux valeurs ou encore recalculer par nous-mêmes des valeurs lorsque les données étaient journalières et non hebdomadaires. Nous avons au total 41 valeurs manquantes.

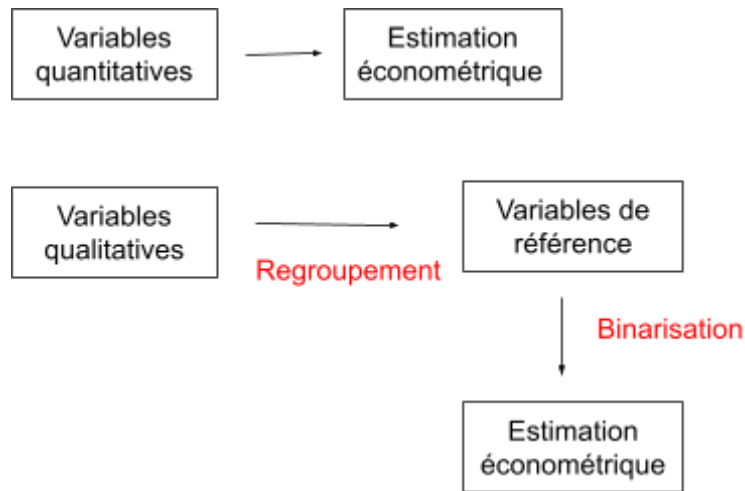
Pour éviter de perdre beaucoup d'information en les supprimant, nous avons fait appel à la méthode des plus proches voisins (KNN) consistant à affecter la moyenne des valeurs des k voisins aux valeurs manquantes²⁹. En d'autres termes, l'algorithme a cherché un individu (une ligne) ayant des caractéristiques les plus proches de celui qui comportait des valeurs manquantes.

3.1.3) Travail sur les types de données.

Notre étude consiste à expliquer la consommation numérique hebdomadaire à l'aide des variables explicatives détaillées précédemment par un modèle de régression linéaire multiple. Or, notre jeu de données comporte quatre variables quantitatives et seize variables qualitatives. Cependant, cette modélisation nécessite que toutes les variables soient quantitatives. Par conséquent, nous devons procéder à la binarisation des variables qualitatives pour les transformer en variables quantitatives. En d'autres termes, nous élaborons des catégories pour les mettre en binaire par la suite. Cependant, afin de respecter la condition selon laquelle le nombre d'observations doit être supérieur au nombre de variables explicatives, nous avons regroupé les catégories les moins représentées en groupes plus conséquents. Cette démarche a été poursuivie par la création de catégories binaires et le choix d'une modalité de référence, correspondant à la modalité regroupant le plus grand nombre d'observations : voir Annexe 1 et 2

²⁹ Cours de classification – semestre 1 : k voisins = k observations pour lesquelles les distances sont très petites.

Figure 6 : schéma transformation des variables



3.2 Analyse statistique.

3.2.1) Analyse des variables quantitatives et des instruments quantitatifs.

3.2.1.A) Analyse univariée

L'analyse univariée des variables qualitatives peut être assimilée à l'élaboration de la carte d'identité de notre échantillon. En effet, à l'aide d'un tableau de contingence accompagné de diagrammes en bâtons, nous pourrions connaître les principales caractéristiques de notre population d'étude.

Ayant une quinzaine de variables qualitatives, nous allons uniquement présenter les principales dans cette partie.

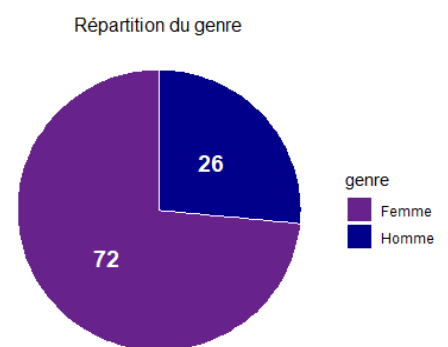
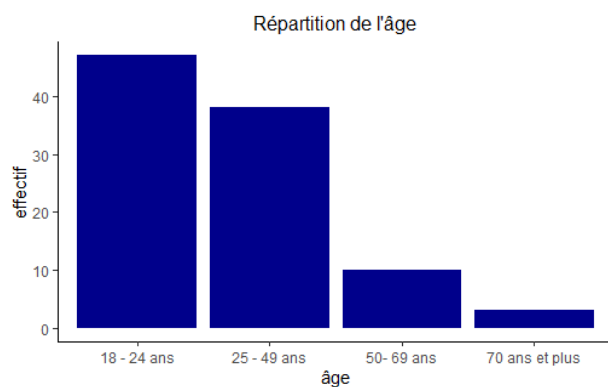
Pour commencer, nous décidons de réaliser un tableau de contingence entre les variables "genre" et "age".

Figure 7 : tableau de contingence sexe/âge

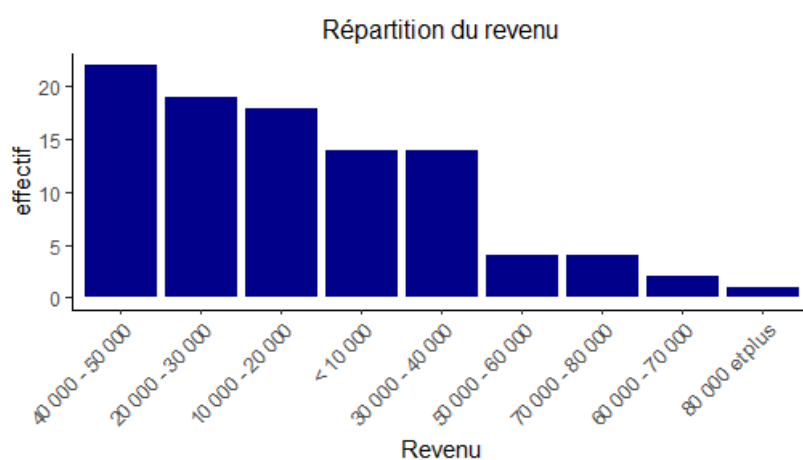
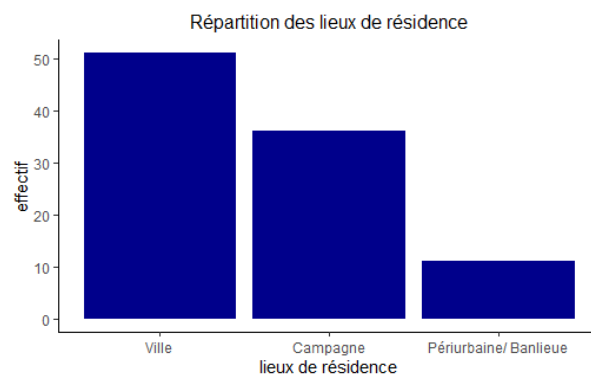
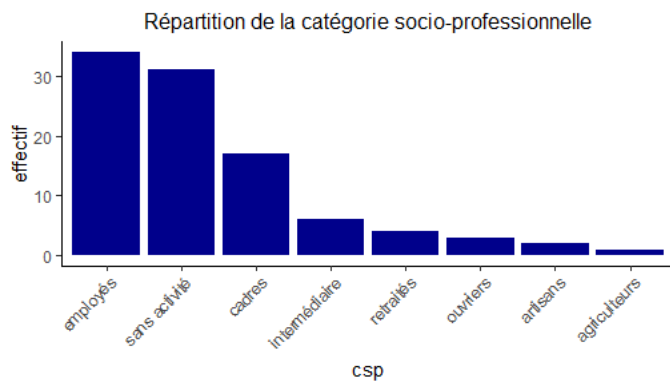
Age Genre	18 - 24 ans	25 - 49 ans	50- 69 ans	70 ans et plus
Femme	36	28	7	1

Homme	11	10	3	2
-------	----	----	---	---

Au regard de ce tableau, nous pouvons voir que la majorité de nos individus ont entre 18 ans et 49 ans avec une dominance du genre féminin. En effet, cette différence est bien représentée par le diagramme en bâtons et diagramme circulaire suivants. Par ailleurs, nous pouvons voir que les générations des années 50-60 constituent une minorité.

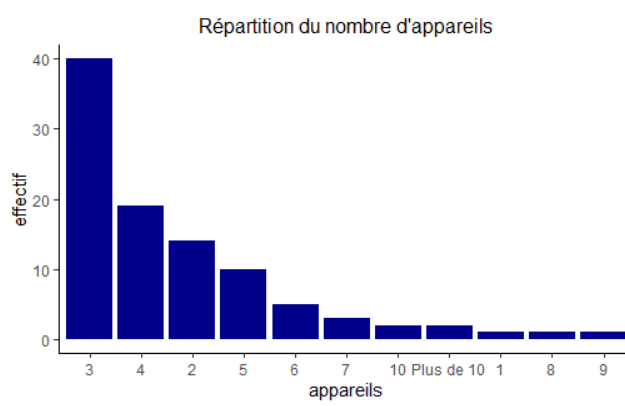
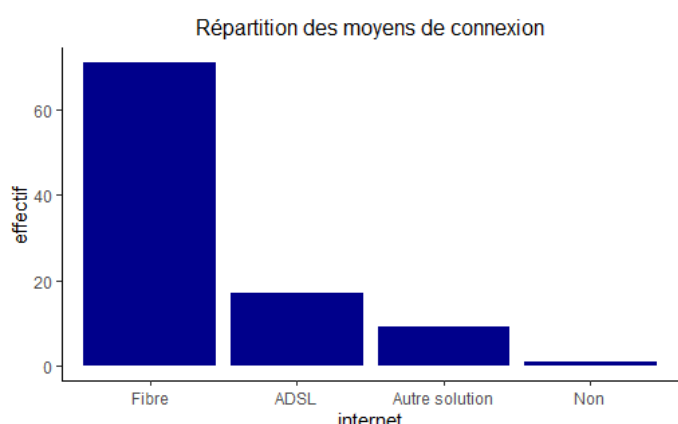


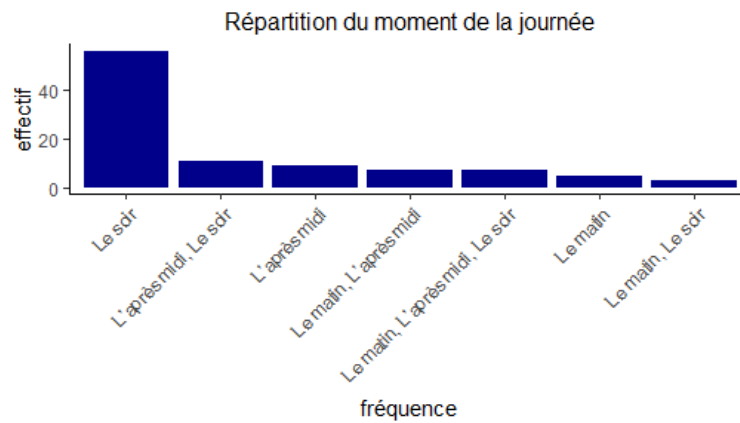
Ensuite, nous allons nous intéresser aux variables “revenu”, “csp” et “lieu_de_vie”. Ayant beaucoup de modalités, nous décidons de réaliser uniquement des diagrammes en barres.



A la vue de ces trois graphiques, nous pouvons d’abord observer une dominance des employés et des personnes n’exerçant pas d’activité professionnelle tels que les étudiants. Ensuite, la proportion d’individus vivant en ville ou en campagne est aux alentours des 90% contre seulement 10% pour la banlieue. Enfin, le revenu est plus étiré avec une majorité d’individus ayant un revenu inférieur à 50 000 euros.

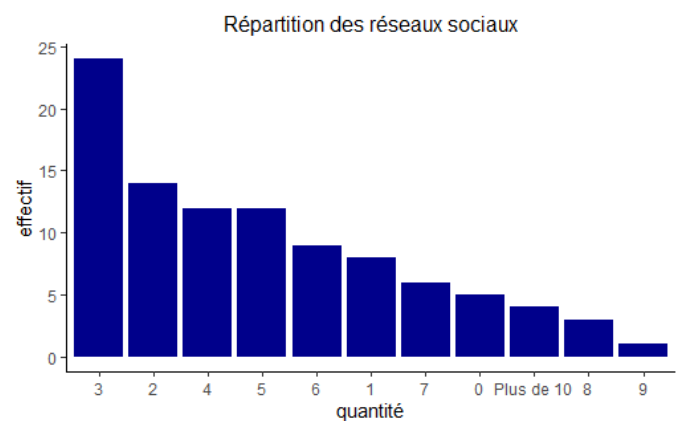
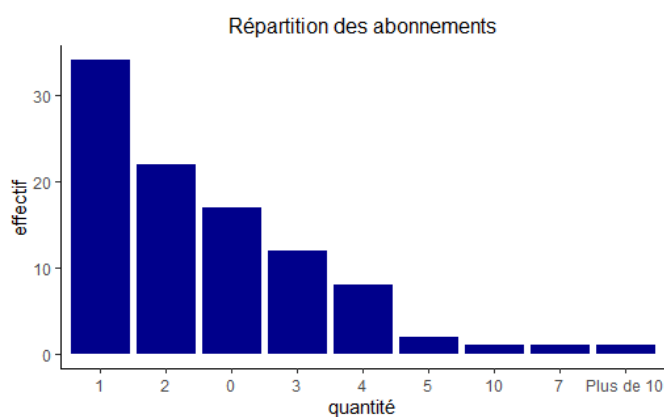
Nous allons poursuivre avec une vue globale de la consommation numérique en dehors de l’activité professionnelle avec l’analyse des variables suivantes : “internet”, “nb_appareils” et “ecran_jour”.



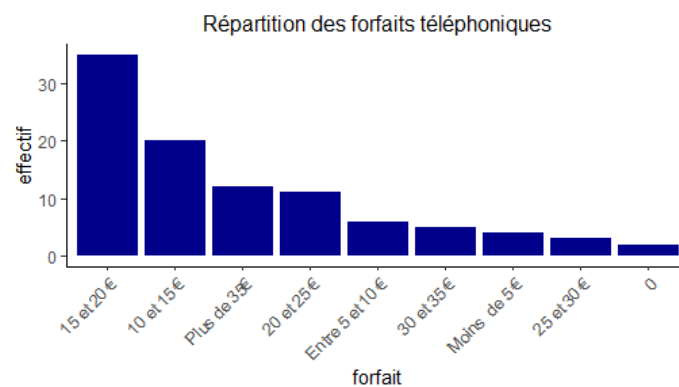


Grâce à ces trois graphiques, nous constatons que la fibre est le moyen largement utilisé par la majorité des individus pour obtenir de l'internet. La plupart des individus ont une consommation importante le soir et possèdent au moins 3 appareils.

Nous finissons par une rapide observation des différents outils utilisés incitant les individus à passer du temps sur les écrans.



A l'observation de ces trois graphiques, nous pouvons conclure que les individus ont tous au moins un abonnement complété par le paiement d'un forfait téléphonique compris entre 15 et 20 euros pour la majorité des cas. Enfin, nous observons également que les réseaux sociaux sont bien présents avec une utilisation de réseaux sociaux multiples pour une grande partie des individus.

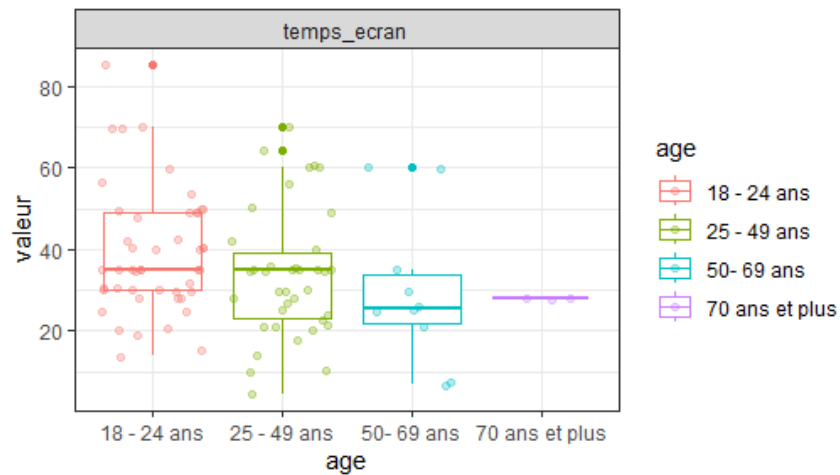


Nous allons poursuivre notre étude statistique par une analyse bivariée.

3.2.1.B) Analyse bivariée

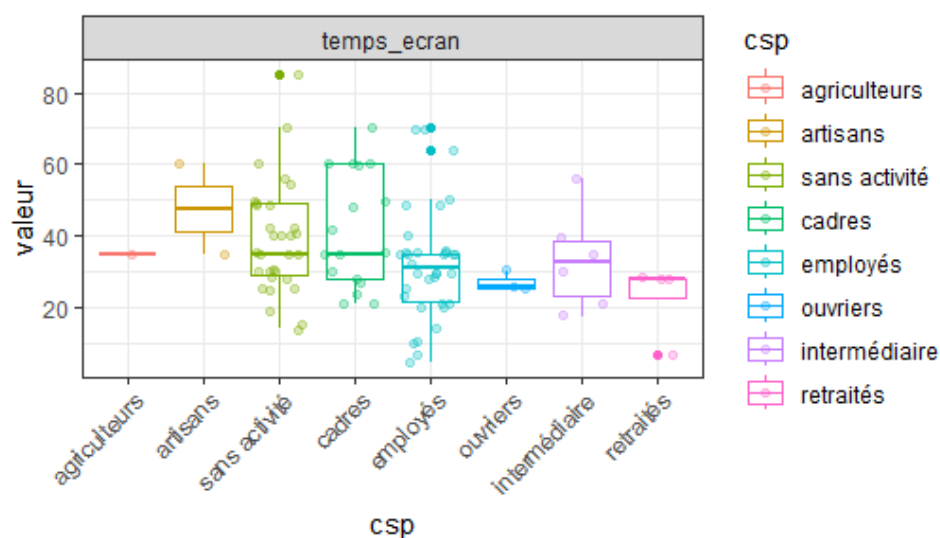
L'analyse bivariée quantitative - qualitative va nous permettre d'observer l'évolution de la consommation numérique (temps_ecran) en fonction des variables explicatives qualitatives. Pour cette partie, nous allons seulement réaliser l'analyse avec les variables "age" et "csp" car ce sont les plus générales.

Nous allons commencer par l'analyse entre la variable "age" et "temps_ecran". Nous trouvons que cette analyse a du sens car tout le monde a tendance à dire que les jeunes consacrent plus de temps aux activités avec écrans que les aînés.



Ce graphique donne raison à la société. En effet, la population passant plus d’heures sur les écrans est la plus jeune avec l’observation de la boîte centrale qui nous indique que 50% de ces derniers dénombrent environ entre 30 et 50 heures passées sur les écrans par semaine.

Nous terminons par l’analyse entre la variable “csp” et “temps_ecran”. Nous trouvons que cette analyse a du sens car les individus seraient plus sujets à utiliser des supports numériques exerçant un métier dans un bureau comme les cadres en comparaison des métiers requérant des compétences manuelles et s’exercent à l’extérieur tels que les agriculteurs ou encore les ouvriers.



Cet ensemble de boxplots confirme notre hypothèse. Les cadres représentent la catégorie la plus gourmande en termes d'heures passées sur les écrans, suivis par les artisans et les individus n'ayant pas d'activité professionnelle.

Maintenant que l'analyse descriptive des variables qualitatives est aboutie, nous allons poursuivre avec l'analyse descriptive des variables quantitatives.

3.2.2 Analyses des variables quantitatives et instruments quantitatifs.

3.2.2.A) Analyse univariée avec les valeurs atypiques

3.2.2.A) a) Mesures de dispersion

	nb_personnes	temps_ecran	loisir	vie_social
Min	0	4,5	0	0
Q1	1	26,25	4	5
Médiane	1	35	10	10
Moyenne	1,622	35,76	12,53	15,92

Q3	3	42	15	22,5
Max	5	85	100	80

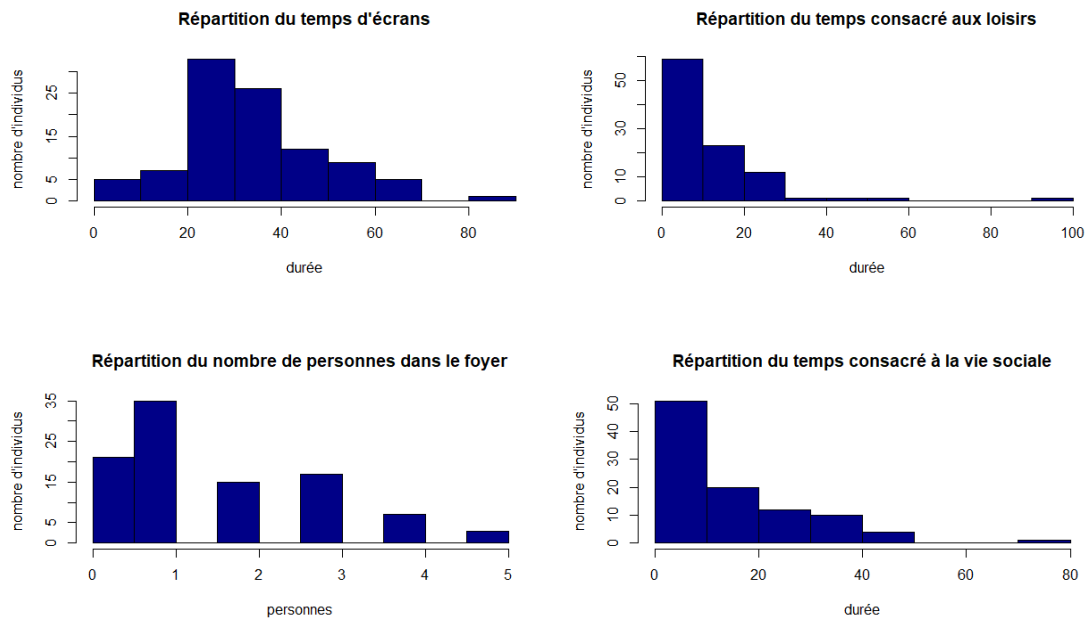
Nous pouvons remarquer que les paramètres de position, moyenne et médiane, sont assez éloignés l'une de l'autre pour la majorité des variables explicatives. Autrement dit, la distribution des valeurs est relativement asymétrique dans sa globalité.

A propos de la variable illustrative « temps écran », nous voyons que la médiane vaut 35. Cela signifie que 50% des individus ont un temps hebdomadaire consacré aux écrans inférieur ou égal à 35 heures et 50% compte plus de 35 heures, ce qui est éloigné de la moyenne. L'écart entre la moyenne et la médiane s'explique par le fait qu'une minorité d'individus font face à une consommation numérique considérable, assez pour tirer la moyenne vers le haut. Cela suppose la présence de valeurs atypiques. Par conséquent, nous pouvons supposer que la distribution du temps d'écrans est modérément disparate.

Nous pouvons compléter notre étude avec l'observation des quartiles de notre variable illustrative. La boîte centrale délimitée par le Q1 et le Q3 nous indique que 50% des individus comptent entre 26 et 42 heures passées sur les écrans. Le troisième quartile signifie que seulement 25% des pays dénombrent plus de 42 heures. Enfin, la moyenne valant environ 36 heures est donc inférieure au troisième quartile.

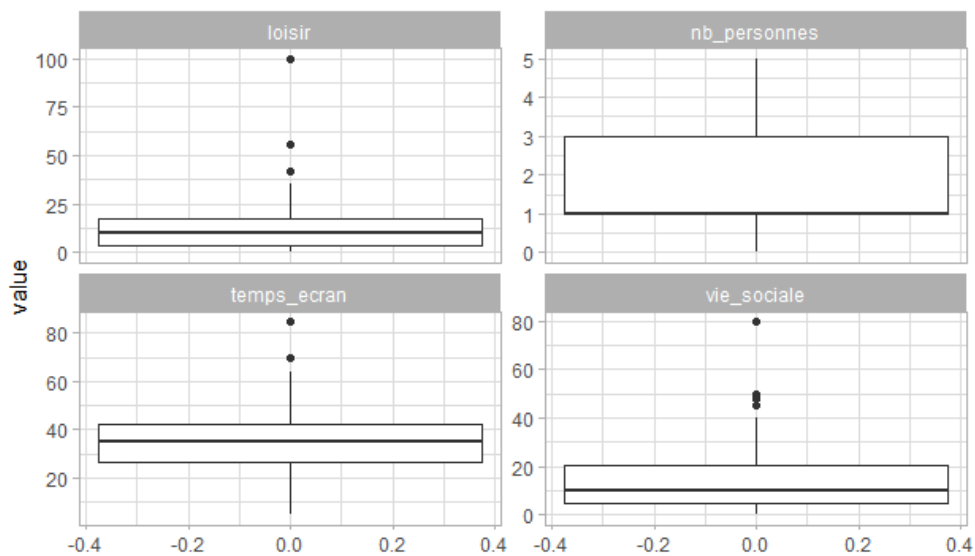
Nous allons poursuivre par la représentation graphique de ces variables qui nous permet d'avoir un meilleur visuel de ces mesures de dispersion.

3.2.2.A) b) Histogrammes



Nous pouvons observer que la distribution de la variable "temps_ecran" ressemble à une distribution de la loi normale mais est asymétrique, ce qui appuie ce qui a été dit précédemment. De plus, nous remarquons que les distributions pour les variables "vie_sociale" et "loisir" sont asymétriques ce qui nous incite à suspecter aussi la présence de valeurs atypiques pour ces variables. Cependant, nous pouvons voir que la distribution pour la variable «nb_personnes » est relativement symétrique avec une discontinuité causée par son caractère discret et non continu.

3.2.2.A) c) Boxplots et atypicité



La visualisation des boxplots confirme la présence de valeurs atypiques pour les variables "loisir", "temps_ecran" et "vie_sociale".

Nous allons utiliser le test de Rosner pour identifier les observations liées aux valeurs extrêmes afin de les supprimer par la suite.

Variables	Nombres d'individus potentiellement atypiques	Test	Individu et sa valeur réellement atypiques
Loisir	3	Rosner	61 : 100 heures 34 : 56 heures
Temps_ecran	2	Rosner	Aucun
Vie_sociale	4	Rosner	9 : 80 heures

D'après le test de Rosner, nous avons 2 valeurs atypiques pour la variable "loisir" et 1 valeur atypique pour la variable "vie_sociale". Les observations 80, 61 et 34 sont donc des valeurs atypiques. Par conséquent, nous les supprimons et notre base de données comporte dorénavant 95 individus.

Nous allons maintenant réaliser l'analyse sans les valeurs atypiques avec les mêmes outils.

3.2.2.B) Analyse univariée sans les valeurs atypiques

3.2.2.B) a) Mesures de dispersion

	nb_personnes	temps_ecran	loisir	vie_social
Min	0	4,5	0	0
Q1	1	26,5	4	5

Médiane	1	35	10	10
Moyenne	1,663	35,88	10,96	15,26
Q3	3	45	15	22
Max	5	85	42	50

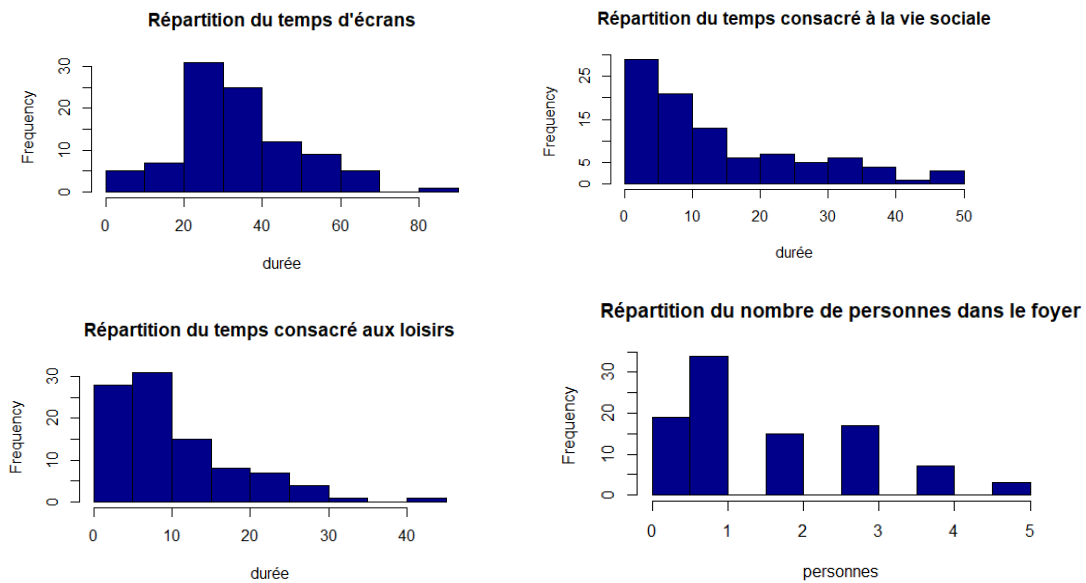
Nous pouvons remarquer que les paramètres de position, moyenne et médiane, sont maintenant assez proches l'une de l'autre pour la majorité des variables actives comme par exemple « loisir » et « vie_social » par rapport à précédemment. Autrement dit, la distribution des valeurs est relativement symétrique dans sa globalité.

A propos de la variable illustrative « temps_ecran », nous voyons que la médiane vaut toujours 35 heures. Cependant, nous pouvons remarquer que l'écart avec la moyenne a augmenté. Cette hausse s'explique par le fait que la quantité d'individus faisant face à une consommation numérique considérable a augmenté, assez pour tirer la moyenne vers le haut. Par conséquent, nous pouvons supposer que la distribution du temps d'écrans est encore disparate.

Nous pouvons compléter notre étude avec l'observation des quartiles de notre variable illustrative. La boîte centrale délimitée par le Q1 et le Q3 nous indique que 50% des individus comptent désormais entre 26,5 et 45 heures passées sur les écrans. Le troisième quartile signifie que seulement 25% des individus dénombrent plus de 45 heures contre 42 heures dans le cas précédent. Enfin, la moyenne se rapproche plus des 36 heures par rapport à précédemment due à une hausse du troisième quartile passant de 42 à 45 heures.

Nous allons poursuivre par la représentation graphique de ces variables qui nous permet d'obtenir un meilleur visuel de ces mesures de dispersion.

3.2.2.B) b) Histogrammes



Nous pouvons observer que la distribution de la variable "temps_ecran" ressemble à une distribution de la loi normale mais son asymétrie a augmenté, ce qui appuie ce qui a été interprété précédemment. De plus, nous remarquons que les distributions pour les variables "vie_sociale" et « loisir » sont désormais continues avec moins de ruptures. Cela s'explique par le fait que l'écart entre la moyenne et la médiane a bien diminué pour ces dernières, surtout pour la variable « loisir ». Par conséquent, leur distribution est moins asymétrique.

Nous aurions pu de nouveau supprimer les valeurs atypiques jusqu'à avoir une belle symétrie des distributions mais cela aurait considérablement diminué notre échantillon d'étude. Ainsi, nous n'avons eu recours à la suppression qu'une seule fois.

3.2.2.C) Analyse bivariée sans les valeurs atypiques

L'analyse bivariée permet d'étudier la relation entre deux variables. La partie va être traitée en deux temps.

Tout d'abord, nous allons étudier la corrélation entre uniquement nos variables explicatives afin de savoir si nous pouvons toutes les conserver pour l'analyse économétrique. Puis, nous analyserons la corrélation de chaque variable explicative avec la variable illustrative.

3.2.2.C) a) Corrélation entre les variables explicatives

La corrélation ou la relation entre deux variables est mesurée par le coefficient de corrélation linéaire qui définit l'intensité et le sens d'une relation linéaire entre deux variables quantitatives. Les valeurs sont toujours comprises entre -1 et 1. Lorsque le coefficient est proche de 1, cela indique que la relation entre deux variables est positive et forte sans pour autant dire qu'il y a de la causalité. A contrario, une valeur proche de -1 indique une relation négative forte mais ne signifie pas pour autant qu'il n'y a pas de relation entre les deux variables car il peut exister une relation non linéaire entre ces dernières.³⁰

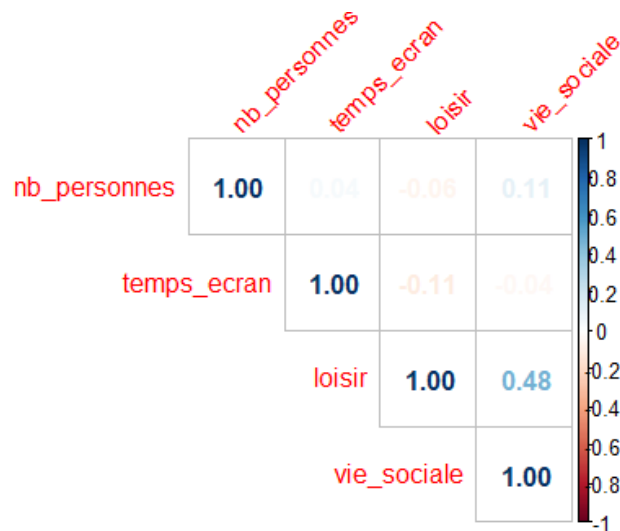
Lorsque nous parlons de coefficient de corrélation, nous sous-entendons la présence de deux coefficients, celui de Pearson et celui de Spearman. Ces derniers se différencient par la normalité des variables. En effet, le coefficient de Pearson nécessite la normalité de toutes les variables. Par conséquent, nous avons vérifié pour une des variables la présence de normalité à l'aide du test de Kolmogorov- Smirnov, adapté aux échantillons comptabilisant plus de 50 observations. D'après le test, la variable ne suit pas une loi normale (voir figure ci-dessous). Nous allons par conséquent utiliser le coefficient de Spearman.

Variable	Test de Kolmogorv-Smirnov
nb_personne	p - value = 1.062e-05 => rejet de la normalité

La datavisualisation est généralement préférée aux tableaux pour l'analyse des corrélations car cela permet d'avoir une compréhension plus rapide avec le jeu des couleurs et parlante pour une personne qui n'a par exemple aucune connaissance sur cette notion. Ainsi, nous trouvons que graphique de la matrice des corrélations, appelée aussi corrélogramme, est le plus approprié pour avoir une interprétation claire et précise.

³⁰ Cours de biostatistiques - semestre 1

La matrice des corrélations (figure ci-dessous) nous donne un aperçu plus visuel sur les différentes relations qu'ont les variables quantitatives entre elles. En effet, elle indique toutes les relations directes et indirectes existantes entre elles avec leur intensité.



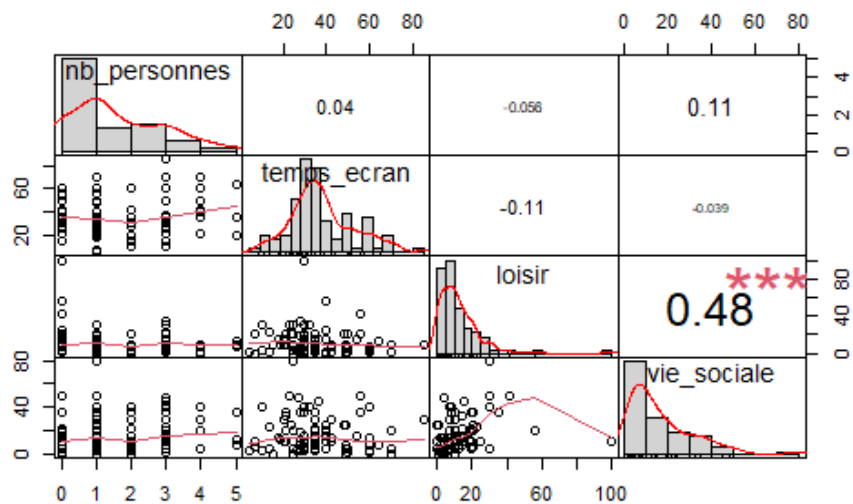
Nous pouvons observer qu'il n'y a pas de corrélation entre toutes les variables explicatives "nb_personnes", "loisir" et "vie_sociale" car le coefficient est toujours inférieur à 0,6 en valeur absolue. Par conséquent, nous allons toutes les conserver dans notre modèle.

De plus, nous remarquons aussi que notre variable illustrative "temps_ecran" n'est corrélée avec aucune autre, ce qui nous amène à étudier la corrélation entre celle dernière et chacune des variables.

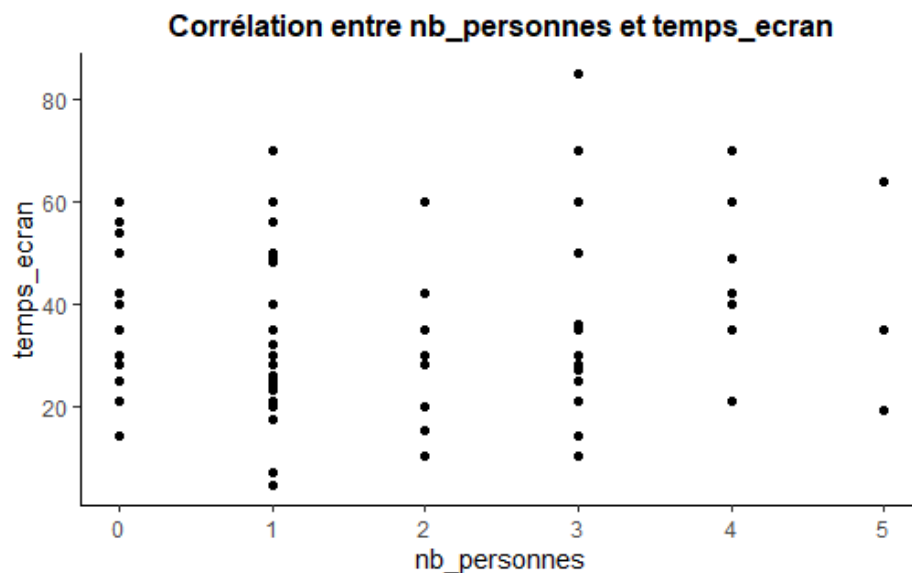
3.2.2.C) b) Corrélation entre les variables illustratives et les variables explicatives.

Le recours aux nuages de points est un bon outil pour observer le sens et l'intensité du lien entre "temps_ecran" et les variables explicatives.

Par conséquent, nous allons présenter le corrélogramme vu précédemment complété des histogrammes et des nuages de points de chaque variable quantitative.

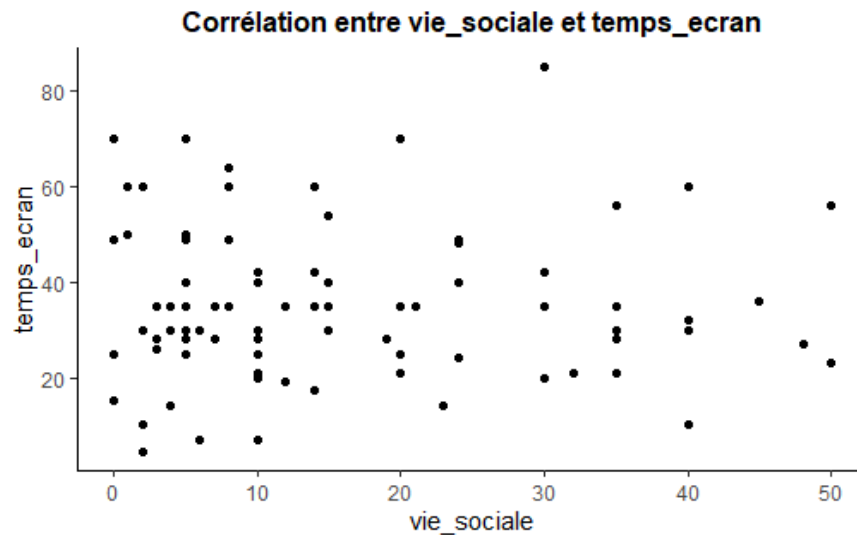


Nous pouvons en effet voir avec les nuages de points que la tendance est approximativement rectiligne entre la variable “temps_ecran” et chacune des variables dû à un coefficient faible. Cependant, nous n’avons pas d’informations précises qui expliquent ce phénomène. Par conséquent, nous allons analyser les quatre nuages de points un par un.

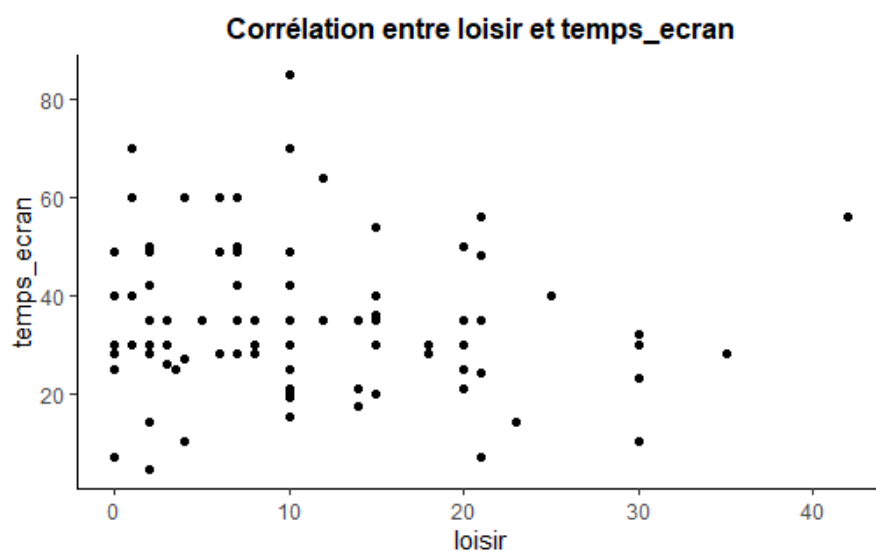


Le graphique nous informe que le nombre de personnes présentes dans le foyer n’influence pas vraiment le temps passé sur les écrans. En effet, une personne vivant seule passe autant de temps qu’une personne vivant avec 2 personnes ou encore qu’une personne

vivant avec 5 personnes. Par conséquent, nous pouvons conclure que la corrélation est presque inexistante. Cela est confirmé par un coefficient de corrélation qui prend la valeur de 0,04 sur le corrélogramme.



Ce nuage de points nous indique que le temps consacré à la vie sociale n'a pas d'incidence sur le temps passé sur les écrans. En effet, un individu qui consacre peu de temps aux sorties passe autant d'heures sur les écrans que celui qui consacre par exemple en moyenne 50 heures par semaine pour les sorties. Par conséquent, nous pouvons conclure que la corrélation est presque inexistante. Cela est confirmé par un coefficient de corrélation valant -0,039 sur le corrélogramme.



Enfin, ce dernier graphique nous informe que le temps consacré aux loisirs a une influence faible sur le temps passé sur les écrans. En effet, une personne qui ne s'octroie aucun moment de la semaine consacré aux loisirs est plus incitée à utiliser les écrans qu'une personne réservant par exemple 30 heures pour les loisirs. Cependant, nous pouvons voir une contradiction pour les individus s'octroyant 10 heures pour les loisirs. En effet, leur temps passé devant un écran est plus élevé qu'une personne qui ne réserve aucune heure pour les loisirs. Par conséquent, nous pouvons conclure que la corrélation est négative et faible. Cela est confirmé par un coefficient de corrélation de -0,11 sur le corrélogramme.

Maintenant que l'étude de notre base de données est accomplie, nous allons désormais procéder l'analyse économétrique.

4. Analyse économétrique.

Après avoir effectué l'analyse descriptive et prétraité nos données, nous entamons une première approche économétrique. Nous allons chercher le modèle le plus pertinent pour ensuite faire une analyse en partant de l'hypothèse que nos variables explicatives sont exogènes. Nous traiterons enfin l'endogénéité de notre modèle.

Nous avons constitué deux bases de données distinctes. La première base se caractérise par des variables explicatives qui n'intègrent pas les instruments. Quant à la deuxième base, elle inclut ces instruments. Dans les deux cas, nous excluons les modalités de référence.

4.1 Modèle initial

Pour évaluer la pertinence de l'utilisation de la méthode des moindres carrés ordinaires (MCO), nous procéderons à la vérification des hypothèses nécessaires. Cette étape sera réalisée au moyen de tests statistiques, permettant ainsi de confirmer ou de réfuter l'applicabilité de cette méthode à notre analyse économétrique.

Nous utiliserons la première base de données, excluant les instruments et les modalités de référence, mais intégrant toutes les autres variables.

4.1.1 Sélection des variables explicatives significatives - Modèle linéaire

Notre modèle initial comportant 26 variables, sa pertinence peut être limitée en raison de l'inclusion d'un grand nombre de variables. Par conséquent, nous allons nous tourner vers la méthode "step" pour identifier la combinaison la plus optimale de variables, affinant ainsi notre modèle afin d'assurer une analyse plus significative. Cela permet d'avoir un modèle avec un nombre de variables plus faibles pour simplifier le modèle tout en gardant une bonne qualité de l'ajustement du modèle. Nous allons alors procéder aux trois méthodes de step, la "forward" soit ascendante qui ajoute progressivement les variables les plus significatives. Puis la "backward", c'est-à-dire la méthode descendante qui supprime progressivement les variables les moins significatives. Et enfin, la méthode bidirectionnelle qui combine les deux méthodes précédentes.

Sélection des variables explicatives significatives - Modele lineaire

Méthode	Variables retenues	AIC
Ascendante	ecran_travail_peu, homme, cat_reseau_6_sup6, ecran_travail_bcp, internet_autre, vie_sociale Nombre de variables : 6	513.47
Descendante	vie_sociale, cat_reseau_6_sup6, tel_sup_25, appareils_1_2, abo_0, homme, pas_ville, cat_cadres, ecran_travail_bcp, ecran_travail_peu, internet_autre Nombre de variables : 11	511.42
Double	ecran_travail_peu, homme, cat_reseau_6_sup6, ecran_travail_bcp, internet_autre, vie_sociale Nombre de variables : 6	513.47

Les méthodes ascendante et double produisent des résultats identiques, avec un critère AIC de 513.47. Tandis que la méthode descendante aboutit à un résultat affichant un AIC de 511,42. Nous avons opté pour le modèle présentant un critère d'Akaike de 511,42, en effet un AIC plus bas indique un meilleur compromis entre l'ajustement du modèle et sa complexité. De plus, dans notre cas cette méthode permet de ne pas supprimer un nombre de variables trop important. Le modèle que nous allons analyser sera donc le suivant :

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{6i} + \beta_3 X_{5i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} + \beta_7 X_{7i} + \beta_8 X_{8i} + \beta_9 X_{9i} + \beta_{10} X_{10i} + \beta_{11} X_{11i} + \varepsilon_i$$

(X_{1i} = vie_sociale, X_{2i} = cat_reseau_6_sup6, X_{3i} = tel_sup_25, X_{4i} =appareils_1_2, X_{5i} = abo_0, X_{6i} =homme, X_{7i} =pas_ville , X_{8i} =cat_cadres, X_{9i} =ecran_travail_bcp, X_{10i} =ecran_travail_peu, X_{11i} =internet_autre)

4.1.2 Hypothèse: Les résidus du modèle estimé suivent une loi normale

Hypothèse nulle (H₀): Les erreurs suivent une distribution normale au seuil de risque de 5%.

Étant donné que notre jeu de données comprend plus de 50 observations, le test de normalité des résidus Kolmogorov-Smirnov est le plus adapté pour évaluer cette hypothèse. Les résultats obtenus sont les suivants : D =0.073643 et une p-value de 0.654.

Une p-value supérieure au seuil généralement accepté de 0.05 indique qu'il n'existe pas de preuves suffisantes pour rejeter l'hypothèse nulle. Cela suggère qu'il n'y a pas de preuves significatives indiquant que les résidus ne suivent pas une distribution normale.

Avec une p-value de 0.654, nous acceptons l'hypothèse nulle de normalité des résidus au seuil de 5%, les résidus du modèle sont bien distribués selon une loi normale.

4.1.3 Hypothèse d'homoscédasticité des erreurs

Notre hypothèse nulle (H₀) dans le cadre du test de Breusch-Pagan est que les résidus sont homoscédastiques, signifiant que les variances des résidus restent constantes et ne sont pas influencées par les valeurs prédites.

L'hypothèse nulle ne sera pas rejetée si la valeur observée lors du test est inférieure à la valeur théorique correspondante à un seuil de risque donné, étant donné que cette valeur théorique suit une distribution du Khi-deux à k degrés de liberté, ou si la p -value du test est supérieure au seuil de risque prédéfini de 0,05. Avec une p -value de 0.4792, nous conservons l'hypothèse nulle d'homoscédasticité des erreurs, indiquant l'absence de preuves significatives contre cette homoscédasticité.

4.1.4 Hypothèse : absence de multicolinéarité entre les variables explicatives

Pour examiner la multicolinéarité parmi les variables explicatives de notre modèle, nous avons recours au critère du facteur d'inflation de la variance (VIF). Un VIF excédant 10 indiquerait une colinéarité prononcée au sein du modèle.

Nos observations indiquent l'absence de problèmes significatifs de multicolinéarité parmi les variables explicatives considérées. Ces valeurs indiquent que chaque variable apporte des informations uniques au modèle, et qu'il n'y a pas de redondance substantielle.

(voir annexe N°)

4.1.5 Forme fonctionnelle du modèle

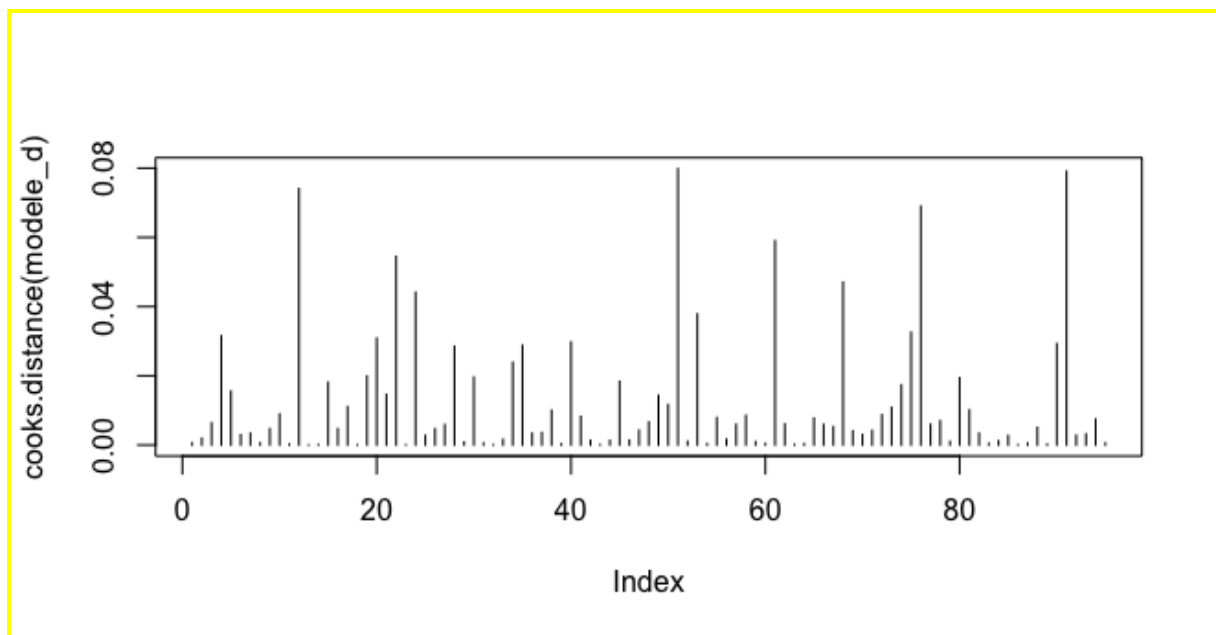
Le test de Ramsey (RESET), utilisé pour évaluer la forme fonctionnelle du modèle, repose sur l'hypothèse nulle (H_0) selon laquelle la forme fonctionnelle spécifiée pour le modèle est adéquate. Cela signifie que si la p -value est supérieure au seuil de risque de 0,05, le modèle est jugé correctement spécifié en termes de sa forme fonctionnelle.

Dans notre cas, étant donné que la p -value du test RESET est de 0.4907, l'hypothèse nulle de ce test n'est pas rejetée. Par conséquent, nous acceptons la forme fonctionnelle linéaire spécifiée pour notre modèle, en considérant le seuil de risque de 5%. Ainsi, la forme fonctionnelle spécifiée pour le modèle est considérée comme adéquate.

4.1.6 Distance de Cook

Nous pouvons observer dans le graphique ci-dessous, que la valeur maximale est légèrement inférieure à 0.08, ce qui signifie qu'aucun individu ne possède une distance de Cook excédant 1. En conséquence, nous déduisons qu'il n'existe aucune observation exerçant une influence notable sur notre modèle estimé.

Graphique N°



4.2 Interprétation des résultats de la méthode des moindres carré ordinaire (MCO)

Maintenant que les conditions nécessaires pour l'utilisation des MCO ont été vérifiées et que dans cette partie nous considérons toutes nos variables exogènes, nous allons pouvoir interpréter les résultats. Le cas de l'endogénéité sera traité dans la partie suivante.

Les résultats de la MCO sont résumés dans le tableau suivant :

Variabes	Estimate	Pr(> t)
Vie_sociale	-0.2519	0.04043*

Tel_sup_25		-5.3294		0.16477	
Appareils_1_2		11.8095		0.01484**	
Abo_0		-6.8988		0.09552*	
Internet_autre		7.2202		0.03442**	
Cat_reseau_6_sup6		12.3794		0.00263***	
Homme		6.4234		0.08300*	
Pas_ville		6.3371		0.05136*	
Cat_cadres		6.4451		0.18471	
Ecran_travail_bcp		-10.5870		0.00758***	
Ecran_travail_peu		-12.7396		0.00178***	
Multiple R-squared	0.3231	Adjusted R-squared	0.2334	p-value	0.000352

Significativité code : « * * * » 0.01, « ** » 0.05 et « * » 0.1

Tout d'abord le test de Fisher pose comme hypothèse nulle qu'aucune variable n'est statistiquement différente de zéro, pour accepter cette hypothèse la p-value du test de Fisher doit être supérieur à 0,05. Dans notre cas la p-value du test est de 0,000352, cela signifie qu'au moins une variable est statistiquement significative au seuil de 5%. Nous rejetons alors l'hypothèse nulle car au moins une variable a un coefficient statistiquement différent de zéro.

Nous pouvons alors interpréter les coefficients des variables, à la condition que leur test de Student conclu à une significativité du coefficient. Pour cela la p-value du test de Student doit être inférieur à 0,01 pour un seuil de risque de 1%, ou 0,05 pour un seuil de risque de 5%.

Notre modèle à trois variables significatif au seuil de 1%. La première est « cat_reseau_6_sup » qui correspond aux personnes étant inscrites sur 6 ou plus réseaux sociaux. Cela signifie que les personnes de cette catégorie ont en moyenne un temps d'écran

supérieur de 12,38 heures par semaine à la catégorie de référence qui sont les personnes étant sur 3 à 5 réseaux sociaux, ce qui est en accord avec la littérature.

Les deux autres sont « *ecran_travail_bcp* » et « *ecran_travail_peu* » correspondant aux personnes qui utilisent un écran au travail plus de la moitié de la journée (mais pas toutes la journée) et ceux qui utilisent un écran moins de la moitié du temps. La catégorie de référence pour ces deux variables sont les personnes passant tout leur temps au travail sur un écran. Les personnes travaillant plus de la moitié du temps sur un écran passent en moyenne 10,59 heures de moins par semaine sur les écrans que les personnes de la catégorie de référence. Tandis que les personnes travaillant moins de la moitié du temps sur un écran passent en moyenne 12,74 heures de moins par semaine sur les écrans que les personnes qui utilisent un écran toute la journée au travail, ces deux résultats sont en accord avec la littérature.

Le modèle contient aussi trois variables avec des coefficients significativement différents de 0 au seuil de risque de 5%. La première variable est la « *vie_sociale* » qui mesure le temps passé avec sa famille ou ses amis en dehors des écrans. Les personnes dont la vie sociale augmente d'une heure ont un temps d'écran qui diminue de 0,2519 heures, ce qui est cohérent selon la littérature.

La seconde variable est « *appareils_1_2* » qui correspond aux personnes ayant 1 ou 2 appareils numériques. Les personnes avec 1 ou 2 appareils ont un temps d'écran en moyenne supérieur de 11,81 heures par rapport aux personnes avec 3 ou 4 appareils qui est la catégorie de référence. Ce résultat est en contradiction avec la littérature. Nous pouvons l'expliquer par le fait que les personnes passent le plus de temps sur leur smartphone en comparaison aux autres écrans³¹. Cela peut expliquer que les personnes avec peu d'écran passent plus de temps que les autres sur les écrans par semaine.

Enfin, la troisième variable significative au seuil de 5% est « *internet_autre* », elle correspond aux personnes ayant accès à internet chez eux par l'ADSL ou tout autre possibilité différente de la fibre. Les personnes ayant internet par un autre moyen que la fibre passent en moyenne 7,22 heures par semaine en plus sur les écrans que les personnes

³¹<https://fr.statista.com/statistiques/1218692/temps-passe-sur-internet-par-les-jeunes-francais-par-type-de-support/> (consulté le 26/01/2024)

ayant internet chez eux via la fibre. Cette conclusion n'est pas celle que l'on attendait au vu de la littérature.

Nous avons aussi trois variables significatives au seuil de 10%, « abo_0 », « homme » et « pas_ville » correspondant chacune aux personnes sans abonnements payant, aux hommes et aux personnes ne vivant pas en ville. Les personnes sans abonnement payant passent en moyenne 6,90 heures de moins par semaine sur les écrans que les personnes avec 1 abonnement payant qui est la variable de référence, ce résultat est cohérent avec nos attentes.

La seconde variable "homme" a un coefficient de 6,42 signifiant que les hommes passent 6,42 heures par semaine de plus en moyenne sur les écrans que la catégorie de référence, les femmes. Ce résultat vient conforter les informations récoltées lors de notre recherche dans la littérature.

La troisième variable, "pas_ville", nous indique que les personnes ne vivant pas en ville passent en moyenne 6,34 heures de plus sur les écrans que la catégorie de référence qui sont les personnes habitant en ville. Ce résultat n'est pas celui attendu, nous pensions trouver une relation inverse, cela peut s'expliquer par le fait que même s'ils sont moins bien équipés les répondants de notre étude ont tous un appareil numérique car notre questionnaire a été fait via une plateforme numérique. L'explication peut donc venir d'autres différences qui peuvent être liées au lieu d'habitation comme le nombre d'activités plus important proposée en ville, éloignant les personnes des écrans.

Nous n'interpréterons pas les autres variables car elles ne sont pas significatives au seuil de risque de 10%. Les variables non significatives sont « tel_sup_25 » et « cat_cadres » qui correspondent respectivement aux personnes avec un forfait téléphone supérieure à 25 euros et aux personnes occupant un poste de cadres et de professions intellectuelles supérieures.

Enfin, nous allons interpréter le R^2 du modèle qui est de 32,31%. Nous pouvons dire que notre modèle explique 32,31% de la variance du temps d'écran par semaine ce qui peu satisfaisant. Cependant, pour pouvoir utiliser les résultats de la MCO, toutes les variables

doivent être considérées comme endogènes ce qui n'est pas le cas dans notre modèle nous allons donc traiter de cette endogénéité dans la partie suivante.

4.3 Prise en compte de l'endogénéité

L'exogénéité des variables du modèle est nécessaire pour pouvoir utiliser la méthode des MCO. Cependant, dans la première partie nous avons vu que certaines variables peuvent être endogènes. Les variables concerné dans notre modèle sont, « vie_sociale », « tel_sup_25 », « appareils_1_2 », « abo_0 » et « internet_autre ». L'endogénéité de toutes les variables sauf vie_sociale avait été prévue par notre questionnaire. Cependant, la variable « sommeil » ne faisant pas partie du modèle final nous décidons de prendre l'instrument « stress » initialement prévu pour cette variable comme instrument de la « vie sociale » (les modalités seront « stress_bcp » et « stress_peu »). La modalité « tel_sup_25 » qui signifie que le forfait téléphone à un prix supérieur à 25 euros a comme instrument la variable utilisation du portable au travail (« tel_pro_non »). La modalité « appareils_1_2 » a comme instrument le revenu fiscal après impôt (« cat_inf_20000 » et « cat_20_40000 »). La modalité « abo_0 » a comme instrument le nombre de personnes vivant dans le foyer (« nb_personnes »). Enfin, la modalité « internet_autre » a comme instrument la possibilité d'être raccordé à la fibre ou non (« fibre_non »). Nous avons alors 5 variables endogènes et 5 instruments (soit 7 modalités car certains instruments sont qualitatifs).

Variables	Estimate	Pr(> t)
Vie_sociale	-0.3865	0.541446
Tel_sup_25	-17.0728	0.435751
Appareils_1_2	2.7568	0.914262
Abo_0	-33.7151	0.217196
Internet_autre	25.8429	0.025057*
Cat_reseau_6_sup6	11.3062	0.278015

Homme		3.5460		0.567533	
Pas_ville		5.5460		0.512034	
Cat_cadres		9.8941		0.320198	
Ecran_travail_bcp		-14.8904		0.104563	
Ecran_travail_peu		-7.9819		0.346820	
Multiple R-squared	-0.4848	Adjusted R-squared	-0.6816	p-value	0.1367

	df1	df2	statistic	p-value
Weak instruments (appareils_1_2)	7	81	1.303	0.25989
Weak instruments (abo_0)	7	81	0.811	0.580768
Weak instruments (internet_autre)	7	81	4.839	0.000136***
Weak instruments (vie_sociale)	7	81	1.495	0.180661
Weak instruments (tel_sup_25)	7	81	1.034	0.413900
Wu-Hausman	5	78	1.814	0.119685
Sargan	2	NA	2.208	0.331536

Significativité code : « *** » 0.001, « * * » 0.01 et « * » 0.05

Pour connaître la pertinence de nos instruments, il faut regarder le résultat au test de Weak instrument. Dans notre modèle, seul le test de Weak pour la variable « internet_autre » a une p-value inférieur à 0.05. Cela signifie que les instruments choisis sont pertinents

pour la variable « internet_autre », les instruments sont corrélés avec cette variable. Mais les instruments ne sont pas pertinents pour toutes les autres variables que nous avons considérées endogènes, en effet les p-value sont supérieures à 0.05. Les instruments n'ont pas d'impact sur les variables « tel_sup_25 », « vie_sociale », « abo_0 » et « appareils_1_2 ».

Nous devons maintenant vérifier le résultat au test de Sargan car nous avons plus d'un instrument par variable « endogène ». Ce test permet de vérifier si les instruments ne sont pas corrélés avec les résidus. La p-value du test étant supérieure à 0.05 la non corrélation des instruments avec les erreurs du modèle est acceptée.

Ensuite nous vérifions le test de Wu-Hausman qui a pour hypothèse nulle, que les variables testées sont supposées exogènes. La p-value de notre modèle est supérieure à 0.05, cela signifie que l'hypothèse nulle est acceptée. Nos variables peuvent donc être considérées comme exogènes et l'interprétation des résultats obtenus par la méthode de MCO peut être gardée.

5. Conclusion et discussion

5.1 Conclusion

Le but de notre étude est d'analyser les facteurs pouvant expliquer le temps d'écrans des Français. Pour cela, nous avons créé un questionnaire construit à partir de la littérature afin de trouver les variables les plus pertinentes. Nous avons eu 98 résultats, cependant les résultats de notre questionnaire ne sont pas représentatifs du ratio femmes/hommes élevés. Nous ne pourrions donc pas utiliser nos résultats à une échelle plus grande comme la France.

Cependant, nous pouvons analyser les résultats de notre échantillon. Dans un premier temps, nous avons choisi d'utiliser la fonction step dans le but d'avoir un modèle plus simple, gardant les variables les plus significatives. Notre modèle final contient 11 variables : vie_sociale, cat_reseau_6_sup6, tel_sup_25, appareils_1_2, abo_0, homme, pas_ville, cat_cadres, ecran_travail_bcp, ecran_travail_peu, internet_autre. Une fois, ce modèle obtenu, le but est d'analyser les liens de causalité entre le temps d'écrans et les variables par la méthode des moindres carrés ordinaires. Pour cela, nous avons d'abord dû

vérifier les conditions d'utilisations des MCO. Notre modèle ne contient pas de problème, les résidus suivent une loi normale et sont homoscédastique, les variables ne sont pas multicolinéarité et la forme fonctionnelle linéaire est acceptée. De plus, l'exogénéité des variables à été acceptée, l'utilisation des MCO est donc possible.

L'analyse avec les MCO nous permet d'identifier 5 variables significatives au seuil de risque de 5%, "vie_sociale", "ecran_travail_bcp", "ecran_travail_peu", "internet_autre" et "appareils_1_2". Ainsi que 3 variables significatives au seuil des risques de 10%, "abo_0", "homme" et "pas_ville". Cependant, pour les variables "internet_autre", "appareils_1_2" et "pas_ville" l'orientation des résultats est inattendue, allant à l'encontre de la tendance que nous espérons observer. Avoir une vie sociale importante diminue le temps d'écran par semaine, de plus, les personnes travaillant plus de la moitié du temps sur un écran ou les personnes travaillant moins de la moitié du temps sur un écran passent en moyenne moins de temps par semaine sur les écrans que les personnes qui utilisent un écran toute la journée au travail. De la même manière avoir un seul abonnement payant fait que le temps passé sur les écrans est inférieur aux personnes ayant 1 abonnement payant. A l'inverse, avoir 1 ou 2 appareils numériques augmente le temps d'écran par rapport aux personnes avec 3 ou 4 appareils chez eux. De plus, les personnes ayant accès à internet par une autre méthode que la fibre passe plus de temps sur les écrans que les personnes avec la fibre. Les hommes passent en moyenne plus de temps sur les écrans par semaine que les femmes. Enfin, ne pas habiter en ville engendre un temps d'écran supérieur par semaine que les personnes vivant en ville.

5.2 Discussion

6. Annexes

Annexe 1 : les variables qualitatives

Variable	Modalités	Catégories	Codage
Réseaux sociaux	0 : 5 1 : 8 2 : 14 3 : 24 4 : 12 5 : 12 6 : 9 7 : 6 8 : 3 9 : 1 Plus de 10 : 4	0 à 2 : 27 3 à 5 : 48 6 et plus : 23	cat_reseau_0_2 cat_reseau_3_5 cat_reseau_6_sup6
Age	18-24 : 47 25-49 : 38 50-69 : 10 Plus de 70 : 3	18-24 : 47 25-49 : 38 50 et plus : 13	cat_18_24 cat_25_49 cat_plus_50
Forfait téléphone	0 : 2 Moins de 5 : 4 5-10 : 6 10-15 : 20 15-20 : 35 20-25 : 11 25-30 : 3 30-35 : 5 Plus de 35 : 12	Inférieure à 15 : 32 15-25 : 46 Plus de 25 : 20	tel_inf_15 tel_15_25 tel_sup_25
Nombre d'appareils	1 : 1 2 : 14 3 : 40 4 : 19 5 : 10 6 : 5 7 : 3 8 : 1 9 : 1 10 : 2 Plus de 10 : 2	1-2 : 15 3-4 : 59 4 et plus : 24	appareils_1_2 appareils_3_4 appareils_plus_4
Nombre d'abonnement	0 : 17 1 : 34 2 : 22 3 : 12 4 : 8 5 : 2 6 : 0 7 : 1 8 : 0 9 : 0 10 : 1 Plus de 10 : 1	0 : 17 1 : 34 2 : 22 3 et plus : 25	abo_0 abo_1 abo_2 abo_3_sup

Variable	Modalités	Catégories	Codage
Lieu de vie	Campagne : 36 Périurbaine/Banlieue : 11 Ville : 51	Ville : 51 Pas ville : 47	pas_ville ville
Genre	Homme : 26 Femme : 72	Homme : 26 Femme : 72	homme femme
CSP	Employés : 34 Autres personnes sans activité professionnelle (par exemple les chômeurs ou les étudiants : 31 Cadres et professions intellectuelles supérieures : 17 Professions intermédiaires : 6 Retraités : 4 Ouvriers : 3 Artisan/commerçant/Chef d'entreprise : 2 Agriculteurs exploitants : 1	Employés : 34 Sans activité pro : 31 Cadres : 17 Autres : 16	cat_Employes cat_sans_activite cat_cadres cat_autres
Ecran travail	Toute la journée : 38 Fréquemment : 26 Parfois : 18 Rarement : 12 Jamais : 4	Toute la journée : 38 Fréquemment : 26 Peu fréquemment : 34	ecran_travail_plein ecran_travail_bcp ecran_travail_peu
Internet	Non : 1 Fibre : 71 ADSL : 17 Autre : 9	Non : 1 Oui_Fibre : 71 Oui_Autre : 25	pas_internet internet_autre internet_fibre
Sommeil	3-5H : 9 5-7H : 32 7-9H : 54 Plus de 9H : 3	Moins de 7 heures : 41 Plus de 7 heures : 57	nuit_inf_7 nuit_sup_7
Ecran moment de la journée	"Le matin, L'après midi, Le soir" : 7 « L'après midi, Le soir" : 11 "Le matin, L'après midi" : 7 "Le matin, Le soir" : 3 "L'après midi" : 9 "Le matin" : 5 "Le soir" : 56	3 fois par jours : 7 2 fois par jours : 21 1 fois par jours : 70	jour_3 jour_2 jour_1

Annexe 2 : les variables instruments qualitatives

Variable	Modalités	Catégories	Codage
Revenu	<10 000 : 11 10 000 20 000 : 17 20 000 30 000 : 22 30 000 40 000 :12 40 000 50 000 :26 50 000 60 000 :3 60 000 70 000 :2 70 000 80 000 :4 80 000 et plus :1	<20 000 : 28 20 000 40 000 : 34 Plus de 40 000 : 36	cat_inf_20000 cat_20_40000 cat_sup_40000
Fibre	Oui : 62 Non : 6	Oui : 62 Non : 6	fibre_oui fibre_non
Téléphone professionnelle	Oui : 65 Non : 33	Oui : 65 Non : 33	tel_pro_non tel_pro_oui
Stress	Toujours : 16 Fréquemment : 26 Parfois : 39 Rarement : 15 Jamais : 2	Toujours : 16 Fréquemment : 26 Parfois, rarement, jamais : 56	stress_toujours stress_bcp stress_peu

7. Bibliographie

Cours de biostatistiques - semestre 1

Cours de classification - semestre 1

Cours d'économie numérique - semestre 1

Abiteboul S. et J. Cattan, « Nos réseaux sociaux, notre régulation », RED, vol. 1, no. 1, 2020, pp. 36-44, (consulté le 28/12) <https://www.cairn.info/revue-red-2020-1-page-36.htm#s1n1>

Aésion mutuelle, “Des conseils pour une vie sociale épanouie”, (consulté le 22/01/24)
<https://ensemble.aesio.fr/aesio-mag/conseils-vie-sociale-epanouie#:~:text=Une%20vie%20sociale%20est%20faite,autres%20et%20savoir%20les%20%C3%A9coute>

Alptis, « Sommeil et écrans : les dessous du problème » :
<https://www.alptis.org/complementaire-sante/sommeil/ecran/> (consulté le 14/01)

BAROMÈTRE DU NUMÉRIQUE - édition 2022 – RAPPORT (30 janvier 2023), page 9 et 29,
(consulté le 31/12)
https://www.arcep.fr/uploads/tx_gspublication/rapport-barometre-numerique-edition-2022-Rapport.pdf

BDM, “Les enfants pauvres passeraient deux heures de plus par jour devant un écran que les enfants riches”,
<https://www.blogdumoderateur.com/les-enfants-pauvres-passeraient-deux-heures-de-plus-par-jour-devant-un-ecran-que-les-enfants-riches/#:~:text=Un%20temps%20d'%C3%A9cran%20quotidien%20consid%C3%A9rable&text=S'il%20s'agit%20d,tweens%2C%20h27%20pour%20les%20teens> (consulté le 26/01/2024)

BFMTV, “De la télévision au smartphone, en passant par les jeux vidéos, la population française consacre la plus grande partie de son temps libre aux écrans. Une tendance accélérée par les confinements.” :
https://www.bfmtv.com/sante/les-francais-passent-plus-de-60-de-leur-temps-libre-devant-un-ecran-selon-une-etude_AN-202204070485.html#:~:text=RMC%20D%C3%A9couverte-,Les%20Fran%C3%A7ais%20passent%20plus%20de%2060%25%20de%20leur%20temps%20libre,un%20%C3%A9cran%2C%20selon%20une%20%C3%A9tude&text=De%20la%20t%C3%A9l%C3%A9vision%20au%20smartphone,tendance%20acc%C3%A9l%C3%A9r%C3%A9e%20par%20les%20confinements (consulté le 22/01/24)

A.Cournoyer, “Internet et les écrans: le fossé des générations”, (consulté le 31/12/2023)

<https://pausetonecran.com/internet-et-les-ecrans-le-fosse-des-generations/>

Gaudiaut T, "Le coût de l'Internet mobile dans le monde", (consulté le 01/01/2024)

<https://fr.statista.com/infographie/29106/comparaison-cout-internet-mobile-dans-le-monde-prix-moyen-1-go-donnees-forfaits-mobiles/>

Henrich E., « La baisse inquiétante du temps de sommeil des Français : comment y faire face ? », 2019 : <https://youmatter.world/fr/temps-sommeil-francais-baisse-conseils/> (consulté le 14/01)

Fondation Ramsay Santé, "Addict aux écrans, quelles sont les conséquences sociales ?", (consulté le 22 /01/2024)

<https://fondation-ramsaysante.com/questions/addiction-aux-ecrans/addict-aux-ecrans-quelles-sont-les-consequences-sociales#:~:text=L'utilisation%20excessive%20des%20%C3%A9crans,enfants%20les%20rendraient%20moins%20sociaux>

Institut national du sommeil et de la vigilance, « SOMMEIL, un carnet pour mieux comprendre »

:https://sante.gouv.fr/IMG/pdf/Sommeil_un_carnet_pour_mieux_comprendre.pdf (consulté le 14/01)

Enquête publiée par l'institut national du sommeil en partenariat avec le groupe mgen sur un échantillon d'un échantillon représentatif de 1014 personnes de 18 à 75 ans :

https://institut-sommeil-vigilance.org/wp-content/uploads/2019/10/DP_JS2019.pdf (consulté le 14/01)

J. Jargon, "Les seniors aussi peuvent être accros aux écrans", (consulté le 02/01/2024)

<https://www.lopinion.fr/economie/les-seniors-aussi-peuvent-etre-accros-aux-ecrans>

L'Année Internet 2022, (consulté le 30/12/2023)

<https://www.mediametrie.fr/fr/lannee-internet-2022>

Le figaro, « Le lien étroit entre le stress et les troubles du sommeil », 2022 :
<https://www.lefigaro.fr/services/le-lien-etroit-entre-le-stress-et-les-troubles-du-sommeil-20220228> (consulté le 14/01)

Leroy T, « En moyenne, les français regardent un écran 32 heures par semaine », 2023, BFMTV:
https://www.bfmtv.com/tech/actualites/en-moyenne-les-francais-regardent-un-ecran-32-heures-par-semaine_AD-202301300358.html (consulté le 14/01/2024)

Livi, “L’impact de l’isolement social sur votre santé mentale”, (consulté le 27/01/2024)
<https://www.livi.fr/en-bonne-sante/isolement-social/>

Madame le figaro, « Un "couvre-feu digital" d'une heure avant de se coucher permettrait de mieux dormir », 2019:
<https://madame.lefigaro.fr/bien-etre/bien-dormir-linstitut-national-du-sommeil-recommande-un-couvre-feu-pour-les-ecrans-sondage-190319-164350> (consulté le 14/01

P. Lombardo, L. Wolff, “Cinquante ans de pratiques culturelles en France”,(consulté le 30/12/2023) <https://www.cairn.info/revue-culture-etudes-2020-2-page-1.htm>

Ouest-France,« Les Français passent la majorité de leur temps libre devant les écrans, selon une étude », 2022 (consulté le 14/01):<https://www.ouest-france.fr/societe/les-francais-passent-la-majorite-de-leur-temps-libre-devant-les-ecrans-selon-une-etude-7715659>

NordVPN, « Combien de temps de nos vies passons-nous en ligne », 2021 :
<https://nordvpn.com/fr/blog/temps-passe-en-ligne/> (consulté le 14/01)

Statista, “En général, combien de temps passez-vous par jour devant les écrans suivants, dans le cadre de vos loisirs ?”, (consulté le 27/01/2024)

<https://fr.statista.com/statistiques/1218692/temps-passe-sur-internet-par-les-jeunes-francais-par-type-de-support/>

SFR business, « Bilan de l'adoption du télétravail en France ? », 2023 :

<https://www.sfrbusiness.fr/room/communications-unifiees/teletravail-bilan-et-perspectives.html> (consulté le 14/01)

Sonnac N, "Quelle place pour l'audiovisuel public face aux réseaux sociaux ?", (consulté le 30/12/2023)

<https://www.vie-publique.fr/parole-dexpert/290432-quelle-place-pour-laudiovisuel-public-face-aux-reseaux-sociaux>

Technologies de l'information et de la communication, (consulté le 02/01/2024)

<https://uis.unesco.org/fr/glossary-term/technologies-de-linformation-et-de-la-communication-tic>

8. Table des matières

1. Introduction.....	3
2. Analyse des variables.....	4
2.1 Variable à expliquer.....	5
2.2 Variables explicatives.....	5
2.2.1) Réseaux sociaux.....	5
2.2.2) L'âge.....	7
2.2.3) Le forfait mobile et son instrument le téléphone portable professionnel.....	9
2.2.4) Le nombre d'appareils numériques et son instrument le revenu.....	11
2.2.5) Vie sociale.....	14
2.2.6) Loisir.....	15
2.2.7) Nombre d'abonnements et son instrument le nombre de personnes dans le foyer.....	15
2.2.8) Genre.....	16
2.2.9) Lieu d'habitation.....	17
2.2.10) Catégorie socio-professionnelle.....	18
2.2.11) Sommeil et son instrument le stress.....	19
2.2.12) Moment de la journée.....	20
2.2.13) Utilisation d'écran du travail.....	20
2.2.14) Accès à internet et son instrument la fibre.....	21
3. Analyse statistique du jeu de données.....	21
3.1 Présentation de la base de données.....	22
3.1.1) Nettoyage et traitement des valeurs manquantes.....	22
3.1.2) Travail sur les types de données.....	22
3.2 Analyse statistique.....	23
3.2.1) Analyse des variables quantitatives et des instruments quantitatifs.....	23
3.2.1.A) Analyse univariée avec les valeurs atypiques.....	23
3.2.1.A.a) Mesures de dispersion.....	23
3.2.1.A.b) Histogrammes.....	24
3.2.1.A.c) Boxplots et atypicité.....	25
3.2.1.B) Analyse bivariée sans les valeurs atypiques.....	26
3.2.1.B.a) Corrélation entre les variables explicatives.....	26
3.2.1.B.b) Corrélation entre la variable illustrative et les variables explicatives..	28
4. Analyse économétrique.....	31
4.1 Choix du modèle.....	31
4.2 Test statistique pour le modèle choisi.....	31
4.2.1 Normalité.....	31
4.2.2 Forme fonctionnelle.....	31

4.2.3 Multicolinéarité.....	31
4.2.4 Homoscédasticité des erreurs.....	31
4.2.5 Distance de Cook.....	31
4.2.6 Interprétation des résultats de la méthode des moindres carré ordinaire (MCO).	31
4.3 Prise en compte de l'endogénéité.....	34
5. Conclusion et discussion.....	37
5.1 Conclusion.....	37
5.2 Discussion.....	37
6. Annexes.....	37
7. Bibliographie.....	39
8. Table des matières.....	40