

Utilisation d'une méthode de scoring pour des estimations locales à partir d'une enquête nationale : exemple avec l'enquête TIC-ménages

Documents de travail

N° 2023-15



Résumé

Dans les enquêtes ménages nationales, la taille de l'échantillon régional s'avère souvent trop faible pour produire directement des estimations infrarégionales, ou même régionales, fiables. Les méthodes d'estimation sur petits domaines peuvent toutefois être mobilisées. Parmi elles, la méthode de scoring consiste à utiliser l'échantillon national pour estimer, via un modèle économétrique, une relation entre un phénomène étudié par l'enquête et les caractéristiques sociodémographiques de ses répondants (sexe, âge, situation professionnelle...). Ce modèle permet ainsi d'estimer la probabilité qu'un individu soit concerné. Ces probabilités peuvent ensuite être projetées sur les habitants d'une région grâce à une source externe, comme le Recensement de la Population.

Cette méthode n'est toutefois pertinente que si deux hypothèses sont satisfaites. D'une part, les données individuelles, disponibles à la fois dans l'enquête et dans la source externe, doivent permettre une bonne prédiction du phénomène étudié. D'autre part, le phénomène étudié doit se produire de manière similaire dans la région d'intérêt et dans l'ensemble de l'échantillon national.

Ce document de travail a pour objectif de proposer un cas concret d'application de cette méthode à partir de l'enquête TIC-ménages 2019. Le phénomène en question est l'illectronisme. La méthode est testée et appliquée sur le taux d'illectronisme au sein des 12 régions de France métropolitaine continentale.

Abstract

In national surveys, the size of the regional sample is often too small to directly produce reliable sub-regional estimates. However, small domain estimation methods can fix this problem. Among them, the scoring method gathers all the individuals in the national sample to estimate, thanks to an econometric model, a relationship between a given phenomenon and a set of socio-demographic explanatory variables (sex, age, occupational situation, etc.). From this estimation, individual probabilities to be faced with this phenomena are calculated. Then, these probabilities are projected on the population of a region, thanks to an external source, such as the census. It is only possible if two important hypotheses are respected. On the one hand, the available individual data in the survey and external source must correctly grasp the analysed phenomenon causes. On the other hand, the analysed phenomena must be relatively similar between the region of interest and the whole national sample.

The purpose of this working paper is to give a concrete example by using the 2019 ICT-Household Survey. In particular, it was tested and applied on the illectronism rate in 12 regions in France.

TABLE DES MATIERES

Introduction	p. 3
1. Estimation sur petits domaines : la méthode de scoring	p. 3
1.1. Principes généraux	p. 3
1.2. Application de la méthode de scoring à l'enquête TIC-ménages 2019	p. 4
1.2.1. Présentation générale	p. 4
1.2.2. Choix de l'échantillon de répondants	p. 5
2. Conformité des données régionales avec l'échantillon de répondants	p. 5
3. Construction et utilisation d'un score pour estimer des taux d'illectronisme localisés	p. 8
3.1. Choix du modèle d'apprentissage	p. 8
3.2. Estimation des taux d'illectronisme localisés	p. 10
4. Effets régionaux et perfectionnement du scoring	p. 10
4.1. Existe-t-il un effet du contexte local ?	p. 10
4.2. Existe-t-il des spécificités régionales et comment en tenir compte ?	p. 11
4.3. Faut-il pondérer les modèles logistiques ?	p. 13
4.4. Pourquoi recalculer les pondérations sur marges régionales et France de Province ?	p. 14
5. Comparaison des taux d'illectronisme observés, calés et scorés pour toutes les régions de France de Province	p. 14
6. Résultats détaillés en Bourgogne-Franche-Comté	p. 16
6.1. L'illectronisme au niveau infrarégional	p. 16
6.2. Estimation d'autres indicateurs	p. 17
6.3. Indicateurs scorés ventilés selon les principales caractéristiques individuelles	p. 18
7. Analyse territoriale au niveau France métropolitaine	p. 19
Conclusion	p. 23
Bibliographie	p. 24
Annexes	p. 25
Annexe 1. Distribution initiale des variables auxiliaires selon la source (avant calage sur marges)	p. 25
Annexe 2. Recherche du meilleur modèle prédictif de l'illectronisme	p. 26
Annexe 3. Pertinence de la méthode de scoring : exemple avec l'emploi, une donnée connue dans le Recensement de la Population	p. 29

Introduction

Dans les enquêtes nationales, la taille de l'échantillon s'avère souvent trop faible pour produire directement des estimations infra-régionales ou même régionales fiables des phénomènes observés par l'enquête. Les extensions d'enquêtes étant coûteuses quand elles mobilisent des enquêteurs, une solution alternative est d'utiliser des méthodes d'estimation sur petits domaines.

Ce document de travail présente l'application d'une de ces méthodes pour mesurer l'illectronisme en région grâce à l'enquête TIC-ménages¹, laquelle ne fournit que des informations représentatives au niveau national. Une personne est considérée comme en situation d'illectronisme lorsqu'elle est dépourvue de toute compétence numérique de base (envoyer des courriers électroniques, consulter ses comptes en ligne, utiliser des logiciels...) ou qu'elle ne s'est pas servie d'Internet au cours des douze derniers mois (par incapacité ou impossibilité matérielle). En France, en 2019, 17 % des personnes de 15 ans ou plus sont en situation d'illectronisme.

La crise sanitaire liée à la Covid-19 a accru les usages numériques avec le recours massif au télétravail, à la télémédecine, aux classes virtuelles ou encore aux loisirs numériques. Dans ce contexte, la lutte contre l'illectronisme est devenue une priorité dans les politiques publiques récentes. Dans le cadre du Plan France Relance, initié en 2020, 10 millions d'euros sont par exemple mobilisés pour mieux outiller et former les aidants numériques. Une information finement territorialisée est donc nécessaire pour organiser et coordonner les actions de lutte contre l'illectronisme.

L'objectif de ce document est de présenter une méthode permettant d'obtenir des estimations locales à partir de l'enquête nationale TIC-ménages, initialement non-prévue pour cela. Cette méthode sera essentiellement appliquée sur l'illectronisme mais aussi sur la maîtrise faible ou élevée des outils numériques, l'usage/non-usage d'Internet dans l'année, l'accès/non-accès à Internet et les types d'usage d'Internet (envoyer/recevoir des mails, s'informer, contacter l'administration...).

1. Estimation sur petits domaines : la méthode de scoring

1.1. Principes généraux

Certaines enquêtes nationales font l'objet d'une ou plusieurs extensions régionales pour obtenir des estimations fiables des phénomènes observés par l'enquête au niveau régional². Toutefois, la pression sur

1. L'objectif de cette enquête est de collecter des informations décrivant l'équipement et les usages des ménages dans le domaine des TIC (informatique, Internet fixe et mobile). Pour l'équipement, l'enquête mesure l'équipement en ordinateur et l'accès à Internet. Pour les usages, elle mesure la fréquence, les lieux et les types d'usages (administration en ligne et commerce électronique) mais aussi les compétences individuelles en informatique et en Internet. Elle est annuelle depuis 2007 et multimode (Internet, papier, téléphone).

2. Produire des statistiques régionales à partir d'enquêtes nationales ne va pas de soi. En effet, les plans de sondage n'assurent la représentativité des échantillons qu'au niveau national ; de plus, la taille de l'échantillon total est en général trop faible pour obtenir des estimations locales suffisamment stables avec les méthodes classiques de type Horvitz-Thompson. En particulier, ces estimateurs classiques, dits « directs » au sens où ils ne s'appuient que sur les observations propres au sous-groupe considéré, ont une variance inversement proportionnelle à la taille du sous-échantillon considéré. Ainsi, si la taille de la population d'intérêt représente 10 % de l'échantillon total, alors la variance de l'estimateur associé est multipliée par 10, ce qui rend les estimations peu fiables. Cet écueil se présente lorsqu'il est question de passer d'une estimation nationale (sur plusieurs milliers d'individus interrogés) à son équivalent régional (quelques centaines d'individus pour certaines régions).

les ressources budgétaires conduit à privilégier des alternatives à ces extensions d'enquête, très coûteuses lorsqu'elles recourent à des enquêteurs. Ces alternatives consistent à utiliser des méthodes spécifiques : les méthodes d'estimation sur petits domaines³ (Ardilly, 2006). À l'Insee, elles ont par exemple été mises en œuvre pour estimer des taux de pauvreté régionaux à partir des Statistiques sur les Ressources et les Conditions de Vie (Sautory 2018). Elles ont également été utilisées pour estimer la prévalence des handicaps dans certains départements (Couet, 2022), des taux de chômage localisés (Ardilly 2012), ou encore des indicateurs régionaux de déplacements et d'usage des moyens de transport (Le Guennec, 2012).

Parmi ces méthodes, celle d'estimation indirecte avec modélisation implicite, plus communément appelée « scoring » (Le Guennec, 2012), consiste à mobiliser l'ensemble des individus de l'échantillon national pour estimer une relation entre un phénomène donné et un ensemble de critères sociodémographiques économétriquement explicatifs de ce dernier (sexe, âge, situation professionnelle...). Ces critères doivent être présents dans une source externe susceptible de donner de l'information au niveau régional et infrarégional, comme l'enquête du Recensement de la Population. C'est donc une approche qui permet, grâce à un modèle économétrique, d'estimer une probabilité théorique pour chaque individu d'être confronté à un phénomène donné et donc de déduire la population d'individus qui l'est dans la région. Cette technique permet potentiellement de réduire la sensibilité des estimations à la composition de l'échantillon, car les estimations sont calculées sur un plus grand nombre d'individus répondants. Par rapport à une extension d'enquête, cette méthode a l'avantage d'avoir un coût faible mais au prix de deux hypothèses importantes. D'une part les données individuelles disponibles doivent permettre de prédire correctement le phénomène étudié. D'autre part, on doit supposer qu'au-delà des critères sociodémographiques retenus pour construire le modèle, les phénomènes étudiés sont relativement similaires entre la région étudiée et l'ensemble de la France métropolitaine. Ce type d'estimation n'est donc envisageable que si ces deux hypothèses sont vérifiées.

1.2. Application de la méthode de scoring à l'enquête TIC-ménages 2019

1.2.1. Présentation générale

La démarche méthodologique consiste à :

- 1/ Vérifier que l'échantillon de répondants à l'enquête TIC-ménages 2019 reste structurellement proche de la population régionale cible afin de réduire les risques d'erreur par ailleurs,
- 2/ Construire un score via une régression logistique (ou « modèle d'apprentissage ») en veillant à tenir compte d'éventuels effets régionaux à l'aide d'une indicatrice région, éventuellement croisée,
- 3/ Projeter les probabilités obtenues par scoring sur les habitants de 15 ans ou plus des territoires cibles via les données du Recensement de la Population.

3. Les petits domaines désignent des sous-groupes de population pour lesquels la taille de l'échantillon est trop petite pour que les estimations directes soient suffisamment fiables pour être publiées

Cette méthode sera principalement testée sur la région Bourgogne-Franche-Comté mais pourra être appliquée dans toutes les autres régions de France de Province, les situations de l'Île-de-France et des DOM étant particulières (cf. infra).

1.2.2. Choix de l'échantillon de répondants

Pour une région de France de Province, comme Bourgogne-Franche-Comté, l'échantillon le plus pertinent pour réaliser le scoring est celui des répondants à TIC-ménages 2019 de France de Province. Plusieurs raisons motivent ce choix. En retirant les habitants d'Île-de-France et des DOM, l'échantillon est plus homogène du point de vue de la structure de la population et des caractéristiques contextuelles tout en étant de taille suffisante. Les échantillons de répondants à TIC-ménages au niveau régional sont en effet de faibles tailles : pour certaines modalités de variables, même en agrégeant les données avec celles d'un autre millésime ou d'une autre région ayant des caractéristiques proches de la région cible, les effectifs restent faibles⁴. Par ailleurs, empiler les millésimes pourrait conduire à une sous-évaluation de l'équipement et l'usage d'Internet, ceux-ci évoluant rapidement dans le temps en fonction de l'offre notamment. Au demeurant, la principale variable d'intérêt, l'illectronisme, n'est pas disponible pour chaque millésime de l'enquête TIC-ménages avant 2021.

En revanche, pour réaliser des estimations infrarégionales en Île-de-France, un modèle de scoring propre à cette région est envisageable à partir de l'échantillon de répondants franciliens, celui-ci étant plus conséquent (N > 2000). A l'inverse, la méthode de scoring ne peut être envisagée pour les DOM : en raison des faibles effectifs de répondants pour chacun d'entre eux, elle ne peut y être appliquée pour obtenir des données infradépartementales. La seule option serait d'utiliser les données brutes de l'enquête TIC de 2017 qui a bénéficié d'une extension d'enquête permettant de fournir des résultats à ce niveau géographique.

2. Conformité des données régionales avec l'échantillon de répondants

Une étape préalable au scoring consiste à vérifier que les données entre l'échantillon de répondants retenu (en l'occurrence ceux de France de Province) et la population cible (individus de 15 ans ou plus vivant en logement ordinaire dans la région cible) varient peu sur les informations auxiliaires, c'est-à-dire disponibles pour l'ensemble de la population visée via le Recensement de la Population. L'absence de distorsion trop importante étaye la confiance que l'on peut avoir dans les estimations localisées. Dans l'Insee Première n°1780 (Legleye et Rolland, 2019), les variables communes aux deux sources ayant le plus d'influence sur le risque d'illectronisme sont identifiées. Nous les mobilisons pour comparer les deux populations. Ces variables sont présentées dans le tableau 1.

4. Un essai a été effectué sur les répondants des régions Bourgogne-Franche-Comté (N=622) et Centre-Val-de-Loire (N=577), connues pour avoir des structures de population proches.

Tableau 1 : Présentation des variables auxiliaires d'intérêt

Variables auxiliaires	Type de variables
Tranche d'âge	Variables individuelles
Sexe	
Niveau de diplôme le plus élevé	
Situation professionnelle	
Catégorie socio-professionnelle (proxy du niveau de vie, non disponible dans le RP)	
Situation familiale	
Catégorie de l'Aire d'Attraction des Villes (2020)	Variables contextuelles
Part communale de locaux éligibles au Très Haut Débit fixe	
Quintile de niveau de vie communal	

La part des locaux éligibles au Très Haut Débit (THD) fixe est une variable construite à partir des données de l'Autorité de Régulation des Communications Electroniques, des Postes et de la distribution de la Presse (ARCEP) au 3ème trimestre 2019. Celles-ci ont toutefois plusieurs limites :

- Celles concernant les communes rurales ne sont pas toutes disponibles,
- Elles donnent une information sur la couverture théorique mais pas sur la couverture réelle. En effet, ces données reflètent la disponibilité d'accès à un service, à l'extérieur des bâtiments, tel que les opérateurs l'affichent sur leurs cartes de couverture. Or, les débits annoncés peuvent être parfois plus élevés que ceux observés dans la réalité puisque le niveau de couverture peut varier selon un certain nombre d'éléments, tels que la typologie du terrain (montagnes, vallées, forêts...), la sensibilité du terminal ou encore le fait d'être à l'intérieur d'un bâtiment,
- Enfin, elles excluent les réseaux locaux de couverture numérique.

La variable « quintile de niveau de vie communal » a quant à elle été construite au niveau France métropolitaine à partir du Fichier Localisé Social et Fiscal et appariée dans le Recensement de la Population grâce à la commune de résidence. Le dernier quintile n'apparaît pas dans les modèles portant sur France de Province, les répondants à l'enquête TIC-ménage qui résident dans les communes les plus aisées de l'Hexagone étant tous en Île-de-France.

Si la structure de l'échantillon de répondants à une enquête nationale est très proche de la structure d'une population régionale sur les variables auxiliaires retenues, et si celles-ci prédisent efficacement une variable d'intérêt (ici, l'illectronisme), alors cet échantillon produira des estimations fiables de cette variable d'intérêt au niveau régional.

En pratique, l'évaluation du degré de similarité des deux sources sur les variables auxiliaires se fait par un calcul de différences standardisées et un R-indicateur global (Schouten et al., 2009 ; Schouten et al., 2012). Les différences standardisées (ou *d* de Cohen) permettent d'évaluer l'équilibrage (identité de distribution) de deux échantillons (ici, répondants à TIC-ménages et RP régional) pour une variable auxiliaire donnée. Plus les différences sont faibles, plus les écarts sont négligeables, même s'ils peuvent être statistiquement significatifs⁵. Pour fournir une mesure synthétique de l'équilibrage sur les variables auxiliaires catégorielles,

5. Pour une présentation détaillée de la distribution des observations par source (TIC et RP) et par variables, voir annexe 1.

la moyenne des valeurs absolues des distances d de toutes les modalités est calculée. Il s'agit de les comparer ensuite aux seuils de référence afin d'évaluer l'importance de l'écart (tableau 2). Les différences standardisées sont préférables à un test pour deux raisons. D'une part, elles quantifient l'écart et non la probabilité qu'il soit dû au hasard : il n'y a donc pas de risque d'erreur d'effectuer une généralisation abusive du résultat de l'échantillon pour la population. D'autre part, elles ne sont pas sensibles à la taille de l'échantillon contrairement aux tests dont la puissance croît avec celle-ci.

Tableau 2 : Règles d'évaluation basées sur les différences standardisées

Seuils de référence du D de Cohen	
Evaluation	Seuil (en valeur absolue)
Très faible écart	0,01
Faible écart	0,2
Écart moyen	0,5
Écart important	0,8
Écart très important	1,2
Écart immense	2

Les résultats du tableau 3 indiquent que les moyennes des valeurs absolues des différences standardisées sont faibles. Ainsi, même s'il existe des écarts statistiquement significatifs, ceux-ci sont suffisamment faibles pour être ignorés. En conclusion, l'échantillon de répondants de France de Province paraît suffisamment bien équilibré sur ces variables auxiliaires pour être structurellement proche de la population de Bourgogne-Franche-Comté.

Tableau 3 : Résultats des tests d'équilibrage par variables auxiliaires

Variables auxiliaires	Nombre de modalités	Valeur des moyennes absolues des d de toutes les modalités	Probchi	Significativité	Indicatrice>seuil
Sexe	2	0,015	0,000	< 0,01	0
Tranche d'âge	6	0,031	0,000	< 0,01	0
Niveau de diplôme le plus élevé	4	0,055	0,000	< 0,01	0
Catégorie socioprofessionnelle	7	0,045	0,000	< 0,01	0
Situation professionnelle	5	0,039	0,000	< 0,01	0
Situation familiale	5	0,081	0,000	< 0,01	0
Catégorie d'AAV	5	0,104	0,000	< 0,01	0
Part de locaux éligibles en Très Haut Débit (THD) fixe dans la commune de résidence (en tranches)	5	0,119	0,000	< 0,01	0
Quintile de niveau de vie communal	4	0,198	0,000	< 0,01	0

Cette conclusion est corroborée par le R-indicateur (ou indicateur de représentativité). En effet, celui-ci, compris entre 0 et 1, permet de mesurer, sans utiliser les variables d'intérêt ou des covariables, le degré de similarité d'un échantillon par rapport à la population cible, 1 signifiant qu'il y a une représentativité parfaite. Concernant la Bourgogne-Franche-Comté, le R-indicateur est de 0,93 ce qui confirme que l'échantillon retenu est équilibré et donc représentatif de la population cible sur les variables auxiliaires envisagées. Le recours à l'échantillon de Province en lieu et place de l'échantillon de la région d'intérêt est donc licite en première analyse. Pour rappel, cela permet de pallier la faible taille de ce dernier, qui risquerait de fragiliser les analyses s'il était utilisé comme unique ressource d'apprentissage du modèle de scoring.

Les tests d'équilibrage réalisés sur les autres régions de France de Province mènent globalement aux mêmes conclusions (tableau 4).

Tableau 4 : Synthèse des résultats des tests d'équilibrage sur toutes les régions de France de Province

Régions	Variables individuelles						Variables contextuelles			Conclusion échantillon*	R-indicateur
	Sexe	Tranche d'âge	Diplôme	CSP	Situation professionnelle	Situation familiale	Catégorie de l'AAV	Part de locaux éligibles au THD fixe (en tranches)	Quintile de niveau de vie communal		
Bourgogne-Franche-Comté	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Ajusté	0,93
Centre-Val-de-Loire	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Ajusté	0,93
Normandie	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Ajusté	0,94
Hauts de France	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Ajusté	0,95
Grand Est	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Ajusté	0,96
Pays de la Loire	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Ajusté	0,94
Bretagne	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Ajusté	0,94
Nouvelle Aquitaine	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Ajusté	0,97
Languedoc-Roussillon	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Ajusté	0,96
Auvergne-Rhône-Alpes	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Ajusté	0,97
PACA	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Déséquilibrée (d=0,22)	Équilibrée	Équilibrée	Ajusté hors AAV	0,96
Corse	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Équilibrée	Déséquilibrée (d=0,25)	Équilibrée	Équilibrée	Ajusté hors AAV	0,79

Note : Une variable est équilibrée si la taille de son effet (mesuré par le d de Cohen) est inférieure au seuil de 0,2 ; en dessous de ce seuil, les écarts moyens de distributions entre l'échantillon (répondants à TIC-ménages de 15 ans ou plus de France de Province) et la population régionale sont faibles et peuvent donc être ignorés.

* Echantillon ajusté sur les variables auxiliaires et donc représentatif de la population régionale sur ces variables (sauf AAV pour 2 régions)

D'après les R-indicateurs, l'échantillon de répondants à TIC-ménages de France de Province est globalement très bien ajusté à la population de chaque région, sauf pour la Corse.

3. Construction et utilisation d'un score pour estimer des taux d'illectronisme localisés.

3.1. Choix du modèle d'apprentissage

Suite au choix de l'échantillon par rapport aux populations cibles, il s'agit de trouver un lien solide entre les variables d'intérêt (équipement, usage, compétences, illectronisme) et les variables auxiliaires via un modèle économétrique, le « modèle d'apprentissage ». Plus précisément, celui-ci doit permettre de mesurer l'association de chaque variable auxiliaire (dite aussi explicative) à une variable d'intérêt en maintenant toutes les autres à un niveau constant. C'est le principe du « toutes choses égales par ailleurs ». Ici, le modèle est logistique, l'habitant i pouvant être en situation d'illectronisme ($Y_i=1$) ou non ($Y_i=0$)

L'annexe 2 détaille les principaux modèles testés et la sélection de celui qui semble avoir les meilleures qualités prédictives. Le modèle retenu intègre comme variables explicatives toutes les variables auxiliaires, à savoir : l'âge, le diplôme, la situation professionnelle, la catégorie socioprofessionnelle, la situation familiale, le sexe, la catégorie de l'aire d'attraction des villes de résidence, le quintile de niveau de vie communal et la part communale de locaux éligible au THD fixe (en tranches). Le modèle permet d'estimer la probabilité qu'un habitant de la région considérée soit en situation d'illectronisme, à partir de ses caractéristiques

décrites par les variables auxiliaires. Ces probabilités sont calculées y compris pour les habitants des territoires dans lesquels il n'y a pas de répondants à l'enquête TIC-ménages.

Dans ce modèle, les variables ont les effets attendus (tableau 5). Ainsi, toutes choses égales par ailleurs, plus une personne est âgée, plus elle risque d'être en situation d'illectronisme. L'âge, le diplôme et la catégorie socioprofessionnelle sont les principaux déterminants de l'illectronisme. Par ailleurs, les estimations des variables contextuelles ne sont pas significatives, aux seuils usuels, lorsque les variables individuelles sont prises en compte. Cela fait sens dans la mesure où l'équipement et l'usage relèvent davantage d'un choix et de capacités individuelles que de l'offre d'équipement et donc du lieu de vie.

Tableau 5 : Estimations du modèle logistique prédictif de l'illectronisme, avant prise en compte des effets régionaux

Modèle de base			
	Paramètres	Estimation	Pr > khi-2
	Intercept	-3,154	0,000
Age	De 15 à 29 ans	-1,076	0,000
	De 30 à 44 ans	-0,994	0,000
	De 45 à 59 ans	Réf	Réf
	De 60 à 69 ans	0,484	0,000
	De 70 à 79 ans	1,438	0,000
	80 ans ou plus	2,439	0,000
Niveau de diplôme le plus élevé	Brevet des collèges, BEPC	1,182	0,000
	CAP-BEP	0,552	0,000
	Baccalauréat	Réf	Réf
	Diplômes du supérieur	-0,355	0,013
Situation Professionnelle	Actifs occupés	Réf	Réf
	Etudiants	-0,890	0,063
	Actifs inoccupés	-0,050	0,791
	Retraités	0,313	0,018
	Autres inactifs	0,474	0,003
Catégorie socio-professionnelle	Agriculteurs	1,180	0,000
	Artisans et chefs d'entreprises < 10 sal.	0,362	0,008
	Cadres et professions intellect. sup.	-0,608	0,000
	Professions intermédiaires	Réf	Réf
	Employés	0,443	0,000
	Ouvriers	1,179	0,000
	Autres inactifs	1,142	0,000
Situation familiale	Personnes seules	Réf	Réf
	Couples sans enfant(s)	-0,666	0,000
	Couples avec enfant(s)	-0,999	0,000
	Familles monoparentales	-1,422	0,000
	Autres	-0,493	0,000
Quintile de niveau de vie communal	1er quintile (communes modestes)	0,259	0,004
	2e quintile	0,225	0,011
	3e quintile	0,035	0,680
	4e quintile (communes aisées)	Réf	Réf
Sexe	Femmes	-0,060	0,382
	Hommes	Réf	Réf
Catégorie de l'aire d'attraction des villes	Commune-centre	0,129	0,104
	Commune d'un pôle	0,048	0,641
	Commune de la couronne	Réf	Réf
	Commune hors attraction des pôles	0,109	0,300
Part de locaux éligibles au THD fixe dans la commune de résidence	Moins de 20 %	0,042	0,652
	De 20 % à 39 %	0,149	0,119
	De 40 % à 59 %	0,058	0,560
	De 60 % à 79 %	Réf	Réf
	80 % ou plus	-0,029	0,773

Source : enquête TIC-ménages 2019,
Champ : individus âgés de 15 ans ou plus vivant en logement ordinaire en France de Province

Afin d'améliorer la prédiction de la situation d'illectronisme en région, une indicatrice régionale est ensuite ajoutée aux autres variables explicatives de ce modèle⁶. Cette indicatrice régionale (par exemple, habiter en Bourgogne-Franche-Comté ou non) pourra être croisée avec certaines des variables auxiliaires pour tenir compte des éventuels effets régionaux (cf. supra). Il y a donc autant de modèles que de régions.

3.2. Estimations des taux localisés

Les taux d'illectronisme localisés sont estimés à partir du modèle retenu pour chaque région. Plus précisément, les principales étapes consistent à :

- Calculer \tilde{Y}_i à partir des coefficients estimés du modèle et des caractéristiques de l'individu i ;
- Appliquer la formule $\frac{\exp(\tilde{Y}_i)}{1+\exp(\tilde{Y}_i)}$ pour calculer la probabilité que l'individu i soit en situation d'illectronisme ;
- Multiplier cette probabilité par le poids de l'individu dans le Recensement de la Population pour obtenir des effectifs.
- Agréger ces effectifs à la maille territoriale retenue et calculer des taux.

4. Effets régionaux et perfectionnement du scoring

4.1. Existe-t-il un effet du contexte local ?

L'effet des variables observées du contexte local sur le fait d'être en situation d'illectronisme apparaît faible par rapport à l'effet des caractéristiques individuelles, notamment l'âge, qui est le plus fort déterminant de l'illectronisme. En particulier, la variable relative au taux de couverture en très haut débit fixe n'est pas significative aux seuils usuels : avec les obligations légales de couverture numérique du territoire, les « zones blanches » deviennent plus marginales. On peut cependant noter que le quintile de niveau de vie communal a un effet significatif, les habitants des communes les plus modestes étant plus concernés par l'illectronisme. De fait, les effets des caractéristiques individuelles étant prépondérants, le modèle de scoring pour estimer localement l'illectronisme semble particulièrement pertinent pour calculer des indicateurs locaux.

Cela n'est toutefois vrai que s'il n'existe pas de politique(s) locale(s) de lutte contre l'illectronisme, qui ne peu(ven)t être prise(s) en compte dans l'analyse. Si tel est le cas alors le modèle pourrait surestimer les taux d'illectronisme par rapport à ce qu'ils sont réellement. Néanmoins, compte tenu de la mise en place « récente » de ce type de politique, l'utilisation du millésime 2019 de l'enquête TIC-ménages devrait minimiser le risque d'un écart significatif.

6. Pour l'Île-de-France, le même modèle a été réalisé mais sur les répondants à l'enquête TIC-ménages résidant dans cette région uniquement, l'échantillon de répondants étant plus conséquent. L'ajout d'une indicatrice région n'a donc pas lieu d'être.

À titre de comparaison, dans l'annexe 3, la méthode est appliquée pour retrouver la part des personnes se déclarant en emploi dans le Recensement de la Population, pour chaque EPCI et chaque département de la région Bourgogne-Franche-Comté. Quoique l'emploi dépende certainement davantage du contexte local que l'illectronisme, les écarts entre parts observées et parts estimées par la méthode de scoring d'individus se déclarant en emploi restent assez minces. Cela contribue à étayer la confiance que l'on peut avoir dans les taux d'illectronisme estimés localement.

4.2. Existe-t-il des spécificités régionales et comment en tenir compte ?

Raisonnant sur l'échantillon de répondants de France de Province, il faut s'assurer qu'il n'y ait pas de spécificités régionales. En d'autres termes, il faut s'assurer que certaines variables (socio-démographiques et/ou contextuelles) n'ont pas d'effet particulier sur la variable d'intérêt (l'illectronisme) dans la région cible et qui ne s'observerait pas au niveau France de Province. Pour vérifier cela, l'effet et la significativité de la variable d'interaction entre une indicatrice région (1=Bourgogne-Franche-Comté ; 0=Reste de France de Province) et chaque variable auxiliaire du modèle retenu doivent être testés.

Les résultats des tests joints des 9 modèles logistiques ainsi réalisés⁷, laissent apparaître que deux variables croisées avec l'indicatrice région Bourgogne-Franche-Comté sont significatives au seuil de 5 % : il s'agit des variables « situation professionnelle » et « sexe ». Cela traduit que ces variables ont un effet légèrement différent en tant que variables explicatives de l'illectronisme en Bourgogne-Franche-Comté et dans le reste de France de Province. Par exemple, toutes choses égales par ailleurs, le fait d'être une femme en Bourgogne-Franche-Comté diminue significativement la probabilité d'être en situation d'illectronisme par rapport au fait d'être un homme alors que cet effet s'avère non significatif au niveau France de Province. Il est également possible d'interpréter ainsi : le fait de vivre en Bourgogne-Franche-Comté augmente la probabilité d'être en situation d'illectronisme par rapport au fait de vivre ailleurs en France de Province sauf pour les femmes pour lesquelles cette probabilité diminue.

Pour neutraliser cet effet régional différencié sur ces deux variables, l'indicatrice région seule et croisée avec les variables « sexe » et « situation professionnelle » sont intégrées dans le modèle logistique portant sur les répondants TIC de France de Province pour estimer l'illectronisme en Bourgogne-Franche-Comté. C'est ce modèle final qui servira alors pour le scoring propre à cette région.

7. Neuf modèles logistiques pour tester l'interaction entre l'indicatrice région et les 9 variables auxiliaires. Ils ne sont pas présentés ici.

Tableau 6 : Résultats du modèle logistique visant à expliquer l'illectronisme avec prise en compte des effets régionaux

Modèle final (BFC)			
	Paramètres	Estimation	Pr > khi-2
	Intercept	-3,219	0,000
Age	De 15 à 29 ans	-1,085	0,000
	De 30 à 44 ans	-1,012	0,000
	De 45 à 59 ans	Réf	Réf
	De 60 à 69 ans	0,479	0,000
	De 70 à 79 ans	1,434	0,000
	80 ans ou plus	2,437	0,000
Niveau de diplôme le plus élevé	Brevet des collèges, BEPC	1,179	0,000
	CAP-BEP	0,542	0,000
	Baccalauréat	Réf	Réf
	Diplômes du supérieur	-0,356	0,013
Situation Professionnelle	Actifs occupés	Réf	Réf
	Etudiants	-0,780	0,106
	Actifs inoccupés	-0,063	0,749
	Retraités	0,369	0,006
	Autres inactifs	0,518	0,001
Catégorie socio-professionnelle	Agriculteurs	1,167	0,000
	Artisans et chefs d'entreprises < 10 sal.	0,367	0,007
	Cadres et professions intellect. sup.	-0,603	0,000
	Professions intermédiaires	Réf	Réf
	Employés	0,446	0,000
	Ouvriers	1,180	0,000
	Autres inactifs	1,151	0,000
Situation familiale	Personnes seules	Réf	Réf
	Couples sans enfant(s)	-0,662	0,000
	Couples avec enfant(s)	-0,992	0,000
	Familles monoparentales	-1,416	0,000
	Autres	-0,495	0,000
Quintile de niveau de vie communal	1 ^{er} quintile (communes modestes)	0,267	0,003
	2 ^e quintile	0,228	0,010
	3 ^e quintile	0,037	0,665
	4 ^e quintile (communes aisées)	Réf	Réf
Sexe	Femmes	-0,017	0,805
	Hommes	Réf	Réf
Catégorie de l'aire d'attraction des villes	Commune-centre	0,129	0,103
	Autre commune d'un pôle	0,047	0,651
	Commune de la couronne	Réf	Réf
	Commune hors attraction des pôles	0,108	0,305
Part de locaux éligibles au THD fixe dans la commune de résidence	Moins de 20 %	0,039	0,674
	De 20 % à 39 %	0,150	0,117
	De 40 % à 59 %	0,052	0,597
	De 60 % à 79 %	Réf	Réf
	80 % ou plus	-0,037	0,719
		Indicatrice régionale (BFC)	0,932
	Etudiants*BFC	-10,702	0,964
	Actifs inoccupés*BFC	0,317	0,648
	Retraités*BFC	-0,734	0,010
	Autres inactifs*BFC	-0,702	0,187
	Femmes*BFC	-0,689	0,006

Source : enquête TIC-ménages 2019,
Champ : individus âgés de 15 ans ou plus vivant en logement ordinaire en France de Province

Les indicateurs de qualité et d'ajustement du modèle final sont présentés ci-après (tableau 7).

Tableau 7 : Résultats des tests de robustesse et d'ajustement entre modèle retenu avec et sans effets régionaux

		Effets régionaux	
		Non contrôlés	Contrôlés
		Modèle de base	Modèle final
	AIC	7145,2	7138,4
	AUC	0,893	0,894
	R ²	32,7%	32,8%
	% concordants	89,3%	89,4%
Test Hosmer – Lemeshow	khi-2	12,4	12,3
	Pr > Khi-2	0,14	0,14

Le modèle final prédit correctement 89,4 % des observations : il est donc de très bonne qualité prédictive et bien ajusté, même s'il n'améliore pas particulièrement le modèle de base qui ne prenait pas en compte les spécificités régionales. Cela démontre que celles-ci ont un impact relativement marginal. De plus, le test d'Hosmer-Lemeshow indique que l'on ne peut pas conclure à un écart important entre les valeurs observées et les valeurs prédites, que ce soit par le modèle de base ou le modèle final. Les deux modèles semblent donc bien calibrés. Le modèle final est retenu, le pourcentage de valeurs concordantes étant légèrement plus élevé.

Somme toute, pour améliorer la précision du score, il est raisonnable d'utiliser l'échantillon de France de Province avec des indicatrices de résidence pour tenir compte des principales spécificités régionales tout en profitant de la robustesse conférée par la plus grande taille de l'échantillon pan-régional. Cette démarche s'étant avérée concluante pour la Bourgogne-Franche-Comté, elle peut être réitérée pour toutes les régions de France de Province.

4.3. Faut-il pondérer les modèles logistiques ?

Comme le scoring est établi au niveau France de Province, les pondérations de l'enquête ont a priori moins de sens puisque celles-ci ont été calculées au niveau France métropolitaine. Par ailleurs, dans le cas où les variables qui impactent le fait d'être répondant ou non à l'enquête sont incluses dans les variables explicatives des modèles logistiques, l'estimateur pondéré est convergent⁸ mais non efficace⁹, tandis que l'estimateur non pondéré est convergent et efficace (Davezies & D'Haultfoeuille, 2009). Les modèles logistiques ne devraient donc pas, a priori, être pondérés.

Toutefois, les estimateurs pondérés sont aussi considérés comme étant plus robustes que les estimateurs non pondérés (Davezies & D'Haultfoeuille, 2009). En outre, la pondération dans l'enquête TIC corrige deux biais de sélection liés à l'illectronisme, au moins via l'usage d'Internet. D'une part, un biais de sélection de niveau ménage, d'autre part un biais de niveau individu. Dans les deux cas, ce sont les ménages où les individus sont les plus à l'aise avec les outils numériques qui répondent davantage à l'enquête TIC, en lieu et place des autres. Ces individus sont généralement les plus jeunes. Ne pas tenir compte des pondérations

8. C'est-à-dire que la probabilité de s'éloigner de la valeur à estimer de plus de ε tend vers 0 quand la taille de l'échantillon augmente. Autrement dit, un estimateur est convergent lorsqu'il converge vers sa vraie valeur.

9. Un estimateur efficace est caractérisé par une faible variance, indiquant une faible déviance entre la valeur estimée et la valeur "réelle".

dans le scoring conduirait donc potentiellement à sous-estimer les taux d'illectronisme, même si les variables de redressement de TIC sont incluses dans le modèle. En effet, la pondération de l'enquête TIC corrige ces effets de sélection liés à l'âge car elle est constituée bloc d'âge par bloc d'âge, celui-ci étant un déterminant du simple fait de répondre à l'enquête. La pondération de l'enquête TIC n'est donc pas une moyenne. De fait, la considération de Hautefoeuille et Davezies (2009) n'est pas valide ici puisqu'elle vaut dans un cadre sans effet de sélection. C'est pourquoi, il semble préférable de pondérer les modèles de scoring, après avoir recalé les poids sur les marges de France de Province¹⁰.

4.4. Pourquoi recaler les pondérations sur marges régionales et France de Province ?

Le taux d'illectronisme observé à l'échelle d'une région dans l'enquête TIC-ménages est obtenu à partir des poids calculés au niveau France métropolitaine. Il est donc possible qu'il ne soit pas représentatif de la population de la région cible. Il s'agit donc de modifier les poids des répondants de manière à ce que l'échantillon ainsi repondéré soit représentatif de certaines variables auxiliaires dont on connaît les totaux sur la population (Sautory, 1993). Cette méthode, dite de calage sur marges, doit permettre d'améliorer la précision des estimateurs et ainsi obtenir les meilleurs résultats possibles au niveau régional¹¹.

Le calage sur marges ne permet toutefois pas d'obtenir des résultats au niveau infrarégional, les effectifs devenant alors trop faibles. C'est pourquoi le recours à la méthode de scoring est nécessaire pour obtenir des estimations du taux d'illectronisme à un niveau territorial plus fin que la région.

La comparaison des taux d'illectronisme régionaux selon la méthode retenue (calage ou scoring) doit cependant permettre de s'assurer de la pertinence de l'approche « modèle ». En effet, si les hypothèses de la méthode de scoring sont respectées, alors les écarts doivent être minimes. Cela doit alors rassurer quant à la fiabilité des taux d'illectronisme estimés au niveau infrarégional.

5. Comparaison des taux d'illectronisme observés, calés et scorés pour toutes les régions de France de Province

Après avoir estimé les taux d'illectronisme pour toutes les régions de France de Province par la méthode de scoring, il apparaît effectivement que les taux théoriques pondérés sont proches des taux régionaux calés sur les marges régionales (tableau 8). Ces résultats confirment la robustesse de la méthode choisie en tenant compte des pondérations recalées sur les marges de France de Province dans les modélisations. Seuls les résultats pour la Corse ne sont pas concluants avec des écarts significatifs : il est donc préférable de ne pas réaliser d'estimations locales des taux d'illectronisme dans cette région.

10. Le modèle logistique étant réalisé sur France de Province et non sur la région par manque d'observations.

11. Les variables ayant servi au calage sur marges régionales de l'échantillon sont : la tranche d'âge, le niveau de diplôme, la situation professionnelle, la situation familiale, la catégorie socio-professionnelle, la catégorie de l'AAV de résidence et le sexe. Seules les variables quintile de niveau de vie communal et part communale de locaux éligibles au THD fixe n'ont pas été retenues.

Tableau 8 : Comparaison des taux d'illectronisme observés (non calés et recalés) et théoriques

Taux d'illectronisme	Observés			Théoriques	Ecart
	Non calés (1)	Calés sur marges régionales (2)	Taille de l'échantillon régional dans TIC	Scorés – modèles pondérés (poids recalés sur les marges de France de Province...*) (3)	
Bourgogne Franche-Comté	18,5	18,2	622	18,8	-0,6
Centre-Val-de-Loire	19,8	19,7	577	19,1	0,6
Normandie	19,7	19,4	699	19,3	0,1
Hauts-de-France	18,2	18,3	1 171	18,0	0,3
Grand-Est	18,0	17,2	1 202	16,8	0,4
Pays de la Loire	13,1	14,9	798	14,7	0,2
Bretagne	13,8	15,6	752	15,2	0,4
Nouvelle-Aquitaine	16,3	17,0	1 299	16,9	0,1
Occitanie	17,9	18,2	1 204	17,8	0,4
Auvergne-Rhône-Alpes	15,1	15,7	1 581	15,4	0,3
PACA	16,5	16,0	1 016	15,7	0,3
Corse	22,9	ns	56	ns	ns
France de Province	16,9	17,1	10 977	16,9	0,2

Champ : individus de 15 ans ou plus vivant en logement ordinaire
Sources : TIC-ménages 2019 et Recensement de la Population 2018

* ...le modèle de scoring portant sur les répondants de France de Province. Les poids ont été normés.
Variables de calage sur marges de France de Province : âge, sexe, niveau de diplôme, situation professionnelle, catégorie socioprofessionnelle, situation familiale, catégorie de l'AAV de résidence

Pour chaque région (hors Bourgogne-Franche-Comté, Normandie, Grand-Est et Nouvelle-Aquitaine) :
Taux scorés d'illectronisme (Y) = [âge, sexe, niveau de diplôme, situation professionnelle, catégorie socioprofessionnelle, situation familiale, catégorie de l'AAV de résidence, niveau de vie communal, part communale de locaux éligibles au THD fixe + indicatrice région]

Pour les régions Bourgogne-Franche-Comté, Normandie, Grand-Est et Nouvelle-Aquitaine : ajout des effets croisés régionaux au modèle précédent :
Bourgogne-Franche-Comté : $Y = [...] + I_BFC * \text{sexe} + I_BFC * \text{situation professionnelle}$
Normandie : $Y = [...] + I_NOR * \text{quintile de niveau de vie communal}$
Grand-Est : $Y = [...] + I_GE * \text{quintile de niveau de vie communal}$
Nouvelle-Aquitaine : $Y = [...] + I_NA * \text{tranche d'âge}$

Les résultats pour la Corse étant moins concluants, ceux-ci ne sont pas présentés

Pour la France de Province : mêmes variables explicatives utilisées, hors indicatrices régions (ou somme des effectifs estimés par région)

Au niveau France de Province, le taux d'illectronisme scoré (16,9 %) est égal au taux observé dans l'enquête (16,9 %) ; il est également très proche du taux régional recalé sur les marges de France de Province (17,1 %). Cette tendance vaut également pour la plupart des régions. Ainsi, en Bourgogne-Franche-Comté, les taux d'illectronisme observés non-calés et calés sur les marges régionales atteignent respectivement 18,5 % et 18,2 % dans la région. Le taux scoré est quant à lui de 18,8 %. Les écarts sont donc assez faibles. Dans les autres régions, les écarts sont plus minces encore. La méthode de scoring semble donc prédire des taux d'illectronisme relativement proches des données observées, preuve de sa pertinence.

6. Résultats détaillés en Bourgogne-Franche-Comté

6.1. L'illectronisme au niveau infrarégional

En Bourgogne-Franche-Comté, 18,8 % des habitants seraient exposés au risque d'illectronisme, soit presque un point de pourcentage de plus qu'au niveau France entière. La Nièvre, la Saône-et-Loire, l'Yonne et la Haute-Saône ont les taux d'illectronisme estimés les plus élevés de la région, en lien avec leur population vieillissante et leur faible densité de population (tableau 9).

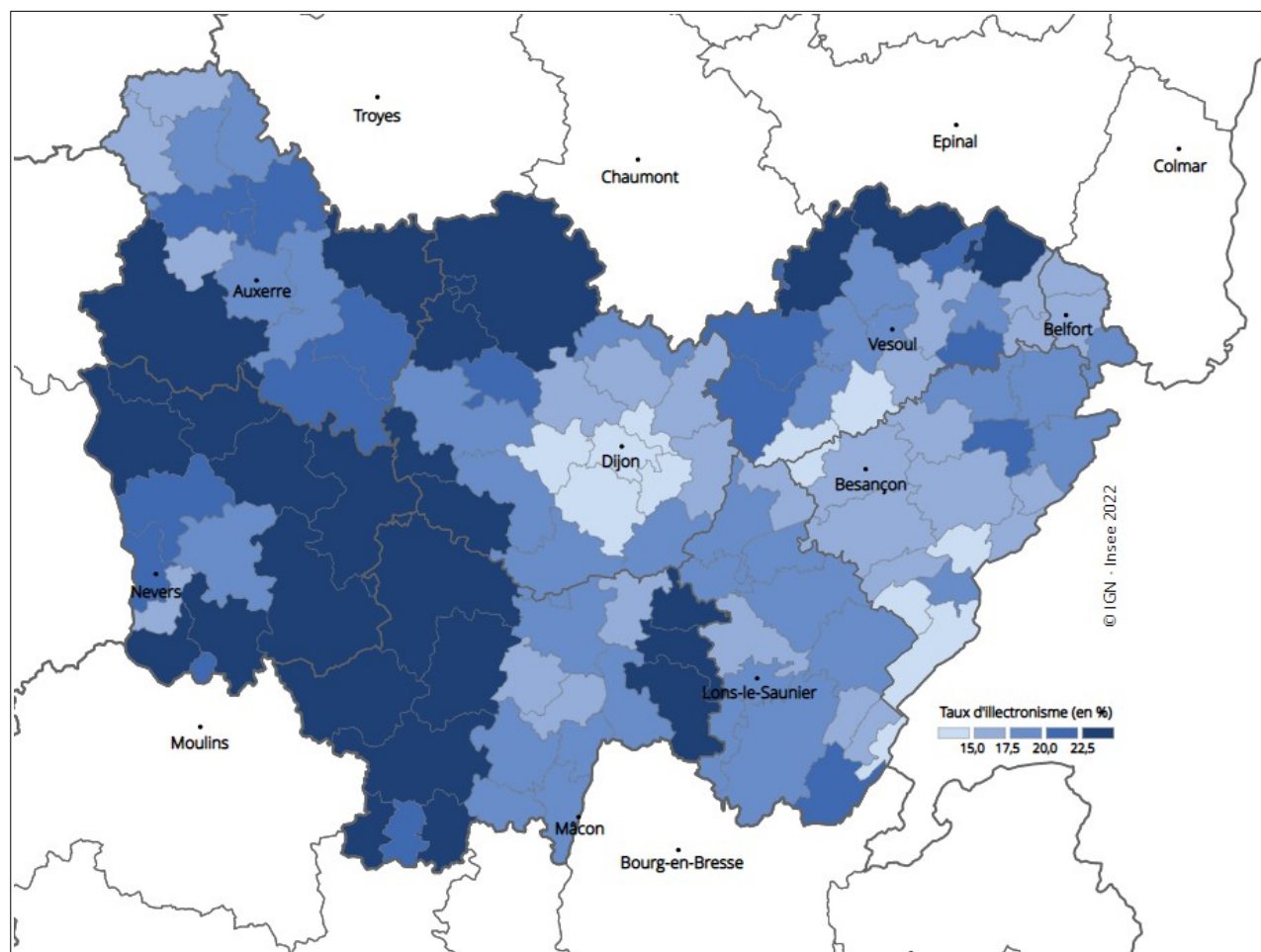
Tableau 9 : Nombre et pourcentage d'habitants de Bourgogne-Franche-Comté exposés au risque d'illectronisme, par département

	En situation d'illectronisme		Population de 15 ans ou plus
	Effectifs	%	
Côte d'Or	72 100	16,2 %	445 600
Doubs	74 200	16,8 %	441 700
Jura	40 000	18,6 %	215 400
Nièvre	40 600	23,1 %	175 800
Haute-Saône	38 300	19,6 %	195 000
Saône-et-Loire	98 200	21,2 %	462 500
Yonne	55 300	19,8 %	279 200
Territoire de Belfort	18 750	16,1 %	116 400
Bourgogne-Franche-Comté	437 500	18,8 %	2 331 600

Champ : individus de 15 ans ou plus vivant en logement ordinaire en Bourgogne-Franche-Comté
Source : Recensement de la population 2018 ; Enquête TIC-ménages 2019

L'intérêt de la méthode de scoring est de pouvoir calculer ces taux à des niveaux géographiques plus fins encore. Cependant, une des principales limites de ces estimations est qu'elles ne prennent pas en compte d'éventuelles mesures locales de lutte contre l'illectronisme (cf. partie 4.1.). La carte 1, ci-après, fournit par exemple les taux d'illectronisme estimés à la maille EPCI. La variation du taux d'illectronisme par EPCI est cohérente avec les caractéristiques de la région. Ce sont dans les territoires où la part des personnes les plus âgées est la plus élevée que l'exposition au risque d'illectronisme est le plus important, l'âge étant le plus fort déterminant de l'illectronisme.

Carte 1 : Taux d'illectronisme estimés par EPCI



Champ : individus de 15 ans ou plus vivant en logement ordinaire en Bourgogne-Franche-Comté
Source : Recensement de la population 2018 ; Enquête TIC-ménages 2019

6.2. Estimation d'autres indicateurs

La méthode de scoring a été généralisée à d'autres indicateurs d'intérêt au niveau de la région, comme la part d'individus n'ayant pas accès à Internet ou encore la part d'individus ayant un usage quasi-quotidien d'Internet (tableau 10).

Tableau 10 : Taux observés, taux calés sur les marges régionales et taux scorés¹² des autres indicateurs - Bourgogne-Franche-Comté

		Observé	Calé	Scoré	Ecart calé-scoré
Compétences numériques	Illectronisme	18,5	18,2	18,8	-0,6
	A au moins une incapacité	34,2	34,3	32,3	2,0
	Maîtrise élevée	24,8	24,0	24,7	-0,7
Équipement Internet	Pas d'accès à Internet à domicile	13,1	14,2	14,3	-0,1
Fréquence d'utilisation d'Internet	Non usage d'Internet dans l'année	16,4	16,4	16,8	-0,4
	Usage quasi-quotidien d'Internet	66,7	66,1	66,8	-0,7
Usage d'Internet au cours des 12 derniers mois pour...	... envoyer/recevoir des mails	88,4	88,4	87,9	0,5
	... s'informer	80,9	79,1	77,4	1,7
	... participer à la vie civique ou politique	9,9	9,3	9,7	-0,4
	... chercher un emploi ou postuler à une offre d'emploi	16,8	16,7	16,3	0,4
	... mettre des documents en ligne (photos, vidéos, textes...)	20,2	19,5	19,3	0,2
	... écouter de la musique en ligne	49,3	50,0	50,7	-0,7
	... vendre des biens et/ou services en ligne	25,3	25,2	26,1	-0,9
	... accéder aux comptes bancaires	67,4	66,2	65,7	0,5
	... sauvegarder des documents	27,2	26,4	26,3	0,1
	... contacter un service public ou une administration	80,3	78,9	77,3	1,6
	... faire des achats en ligne	75,1	73,3	72,3	1,0

Champ : individus de 15 ans ou plus vivant en logement ordinaire en Bourgogne-Franche-Comté.

Source : Recensement de la population 2018 ; Enquête TIC-ménages 2019

Lecture (taux observés) :

a) En Bourgogne-Franche-Comté, 25 % des habitants de 15 ans ont une maîtrise élevée des outils numériques.

b) En Bourgogne-Franche-Comté, 88 % des usagers d'Internet au cours des 12 derniers mois ont effectué des achats en ligne.

Globalement, les taux scorés ne s'éloignent pas trop des taux calés, sauf pour trois variables (maîtrise faible des outils numériques, usage d'Internet au cours des 12 derniers mois pour s'informer ou contacter un service public), qu'il est donc plus prudent de ne pas commenter.

6.3. Indicateurs scorés ventilés selon les principales caractéristiques individuelles

Le tableau 11 présente les différents indicateurs scorés en fonction des caractéristiques socio-démographiques des individus de 15 ans ou plus de Bourgogne-Franche-Comté.

Les résultats sont tout à fait cohérents avec ce qui est attendu. En particulier, le taux d'illectronisme estimé augmente avec l'âge et baisse à mesure que le niveau de diplôme s'élève. Par effet « miroir », la part d'individus de 15 ans ou plus susceptibles d'avoir de fortes compétences numériques est la plus élevée chez les moins de 30 ans et la plus faible chez les personnes âgées de 70 ans ou plus. De fortes disparités existent aussi selon la situation familiale : les couples avec enfants ont des taux d'illectronisme estimés très faibles (moins de 6%) et des taux d'équipement et d'usage quasi-quotidien d'Internet estimés beaucoup plus élevés que ceux des autres types de famille, ce qui témoigne de l'influence des jeunes dans l'acquisition des compétences de leurs parents. Par ailleurs, en lien avec leur âge, il est estimé que 100 % des étudiants vivant en Bourgogne-Franche-Comté ont un usage d'Internet quasi-quotidien, ce qui est également cohérent avec le fait que tous semblent avoir Internet chez eux. Enfin, les femmes semblent être davantage éloignées des outils numériques et de leur usage que les hommes : d'après les estimations, elles

12. Pour rappel, les taux scorés ci-dessus sont issus de modèles logistiques pondérés des poids recalés sur les marges de France de Province, comme pour l'illectronisme.

sont en proportion plus nombreuses à ne pas avoir Internet chez elles ou à ne pas avoir utilisé cette technologie au cours des 12 derniers mois, expliquant en partie qu'elles soient aussi plus souvent exposées au risque d'illectronisme. Toutefois, cet écart avec les hommes est davantage lié à la forte surreprésentation des femmes après 75 ans dans la population de Bourgogne-Franche-Comté qu'au fait d'être une femme.

Tableau 11 : Taux scorés relatifs aux compétences numériques, à l'équipement Internet et à sa fréquence d'utilisation selon les caractéristiques socio-démographiques des individus vivant en Bourgogne-Franche-Comté

		Part d'individus de 15 ans ou plus ayant...				
		...aucun usage d'Internet ou aucune compétence numérique (illectronisme)	... des compétences numériques fortes	...aucun équipement Internet à domicile	...eu aucun usage d'Internet au cours des 12 derniers mois	... un usage quasi-quotidien d'Internet
Age	15 à 29 ans	3 %	53 %	2 %	3 %	92 %
	30 à 44 ans	4 %	33 %	5 %	4 %	85 %
	45 à 59 ans	11 %	21 %	10 %	9 %	71 %
	60 à 69 ans	21 %	12 %	14 %	18 %	57 %
	70 à 79 ans	45 %	5 %	24 %	39 %	36 %
	80 ans et plus	74 %	1 %	67 %	70 %	12 %
Diplôme	Sans diplôme ou brevet	38 %	14 %	30 %	35 %	46 %
	BEP-CAP	17 %	13 %	11 %	15 %	65 %
	Baccalauréat	7 %	33 %	5 %	6 %	80 %
	Diplôme du supérieur	4 %	45 %	4 %	4 %	85 %
Sexe	Femmes	23 %	25 %	19 %	21 %	64 %
	Hommes	14 %	25 %	9 %	13 %	70 %
Situation professionnelle	En emploi	7 %	33 %	5 %	6 %	81 %
	Étudiants	0 %	73 %	0 %	0 %	100 %
	Chômeurs	11 %	19 %	10 %	11 %	77 %
	Retraités	42 %	7 %	30 %	38 %	39 %
	Autres inactifs	22 %	12 %	28 %	19 %	50 %
Situation familiale	Personnes seules	34 %	16 %	32 %	31 %	49 %
	Couples sans enfant(s)	24 %	17 %	15 %	20 %	58 %
	Couples avec enfant(s)	5 %	32 %	3 %	5 %	84 %
	Familles monoparentales	9 %	21 %	11 %	8 %	75 %
	Autres	13 %	41 %	10 %	12 %	79 %
Sans ventilation		19 %	25 %	14 %	17 %	67 %

Champ : individus de 15 ans ou plus vivant en logement ordinaire en Bourgogne-Franche-Comté

Source : Recensement de la population 2018 ; Enquête TIC-ménages 2019

7. Analyse territoriale au niveau France métropolitaine

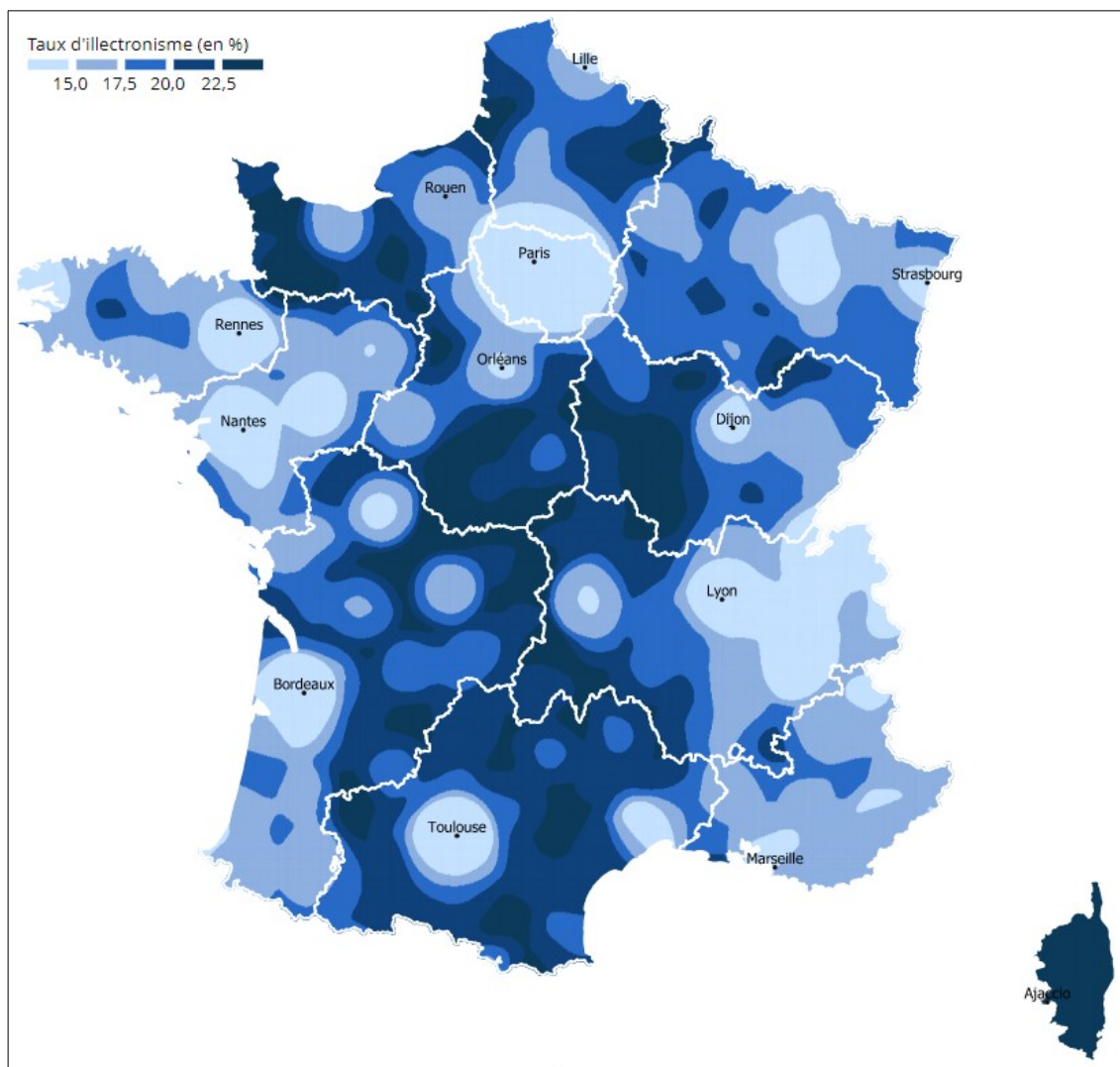
La même démarche que pour la Bourgogne-Franche-Comté a été réalisée pour chaque région de France métropolitaine, ce qui permet d'avoir des estimations infrarégionales de l'illectronisme et des usages d'Internet pour toute la France Métropolitaine, sauf pour la Corse¹³. Cela permet d'en dresser une analyse territoriale fine même si ces variables sont finalement très dépendants de la structure de la population, leurs principaux déterminants étant quasi-exclusivement liés aux caractéristiques individuelles de la population (cf. supra).

Ainsi, les taux d'illectronisme estimés sont particulièrement forts dans la « diagonale des faibles densités », allant des Ardennes au Sud-Ouest (hormis la métropole de Toulouse) en passant par le Massif central. À l'inverse, les taux estimés sont plus faibles dans les grandes aires d'attraction des villes (carte 2). Les

13. Les résultats de la Corse (cf parties 2 et 5) ne sont pas assez robustes : ils ne peuvent donc être présentés.

territoires français sous influence de grands pôles situés à l'étranger semblent également moins touchés par l'illectronisme. En effet, ils sont globalement plus favorisés socialement, notamment ceux situés à proximité des pôles du Luxembourg ou de Genève-Annemasse, compte tenu des salaires plus élevés des travailleurs frontaliers

Carte 2 : carte lissée des taux d'illectronisme estimés au niveau intercommunal pour chaque région



Lecture : le taux d'illectronisme estimé est inférieur à 15 % dans la majorité des plus grandes aires d'attraction des villes comme Paris, Lyon ou Bordeaux.

Champ : France métropolitaine, personnes de 15 ans ou plus vivant en logement ordinaire.

Source : Insee, enquête TIC-ménages 2019, recensement de la population 2018.

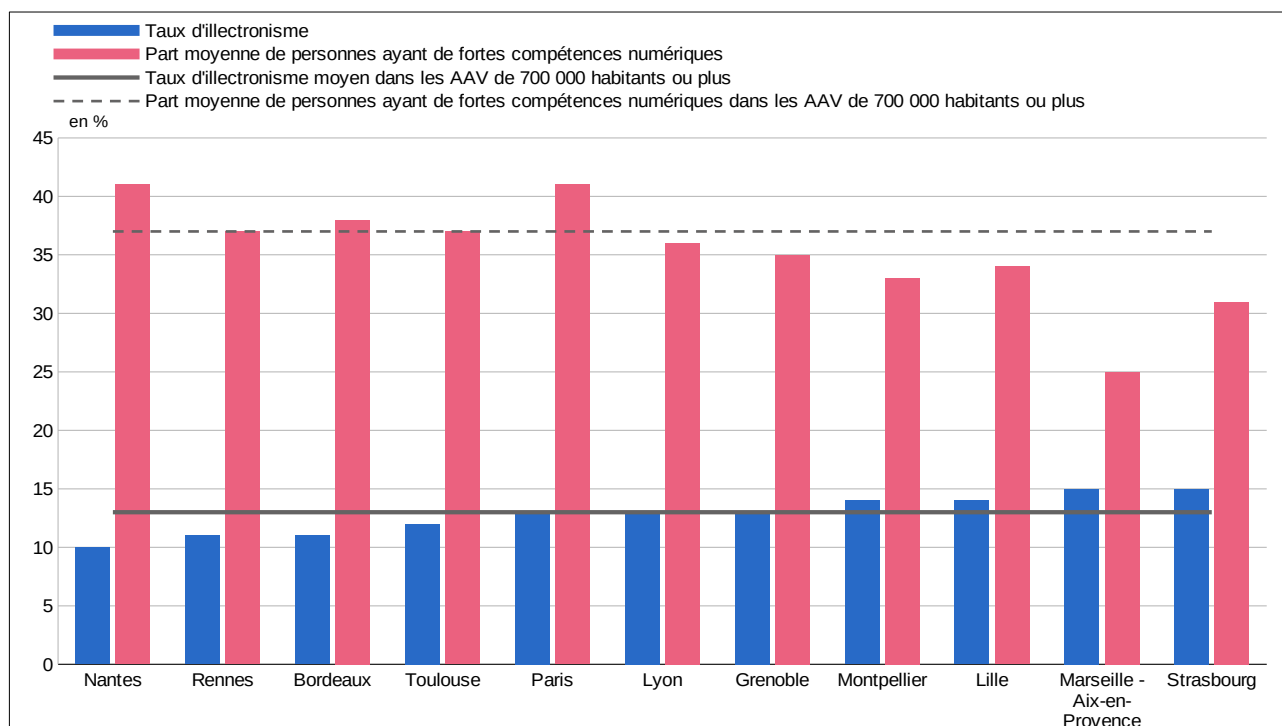
Parmi les aires d'attraction des villes de plus de 700 000 habitants, Strasbourg, Marseille-Aix-en-Provence, Lille et, dans une moindre mesure, Montpellier se démarqueraient avec des taux d'illectronisme estimés relativement plus élevés (graphique 1). Sur la carte 2, cette moins bonne situation par rapport aux autres grandes aires d'attraction des villes, se matérialise par un périmètre beaucoup plus restreint autour de la ville centre concerné par un taux d'illectronisme estimé inférieur à 15 %. La part d'habitants susceptibles d'avoir des compétences numériques fortes y est aussi plus faible, tout comme celle utilisant Internet tous les jours ou presque. Les écarts de niveaux de vie expliquent en partie ces différences. Les niveaux de vie annuels médians sont inférieurs à 22 000 euros à Marseille-Aix-en-Provence, Lille et Montpellier, alors qu'ils

dépassent 23 000 euros dans toutes les autres aires de même taille. Dans les premières, le taux de pauvreté oscille ainsi entre 15 % et 18 % quand il avoisine plutôt les 10 % dans celles de Nantes ou Rennes, où seul un habitant sur dix est exposé au risque d'illectronisme.

Par ailleurs, la population des aires d'attraction de Montpellier et de Marseille-Aix-en-Provence est plus âgée : un habitant sur cinq a au moins 65 ans, soit davantage que dans le reste du groupe. De même, Lille et Marseille-Aix-en-Provence ont une population moins qualifiée, avec respectivement 23 % et 21 % des habitants de 15 ans ou plus peu ou pas diplômés, contre 18 % en moyenne dans les autres aires d'attraction de même taille. Ces facteurs contribuent à expliquer que les populations de ces grandes aires d'attraction soient moins enclines à utiliser les outils numériques.

Le taux d'illectronisme estimé plutôt élevé dans l'aire d'attraction de Strasbourg est quant à lui le résultat de fortes disparités entre son pôle et sa couronne. En effet, il atteint 13 % dans le pôle, contre 17 % dans la couronne côté français. C'est l'écart le plus important parmi les aires d'attraction de plus de 700 000 habitants. Cette différence s'explique à nouveau par les caractéristiques des populations respectives. Par exemple, 23 % des habitants de la couronne strasbourgeoise sont ouvriers ou l'étaient avant de partir en retraite, contre 18 % en moyenne dans les couronnes des villes de province de même taille.

Graphique 1 : Taux d'illectronisme estimés et part estimée des personnes ayant de fortes compétences numériques, dans les plus grandes aires d'attraction des villes

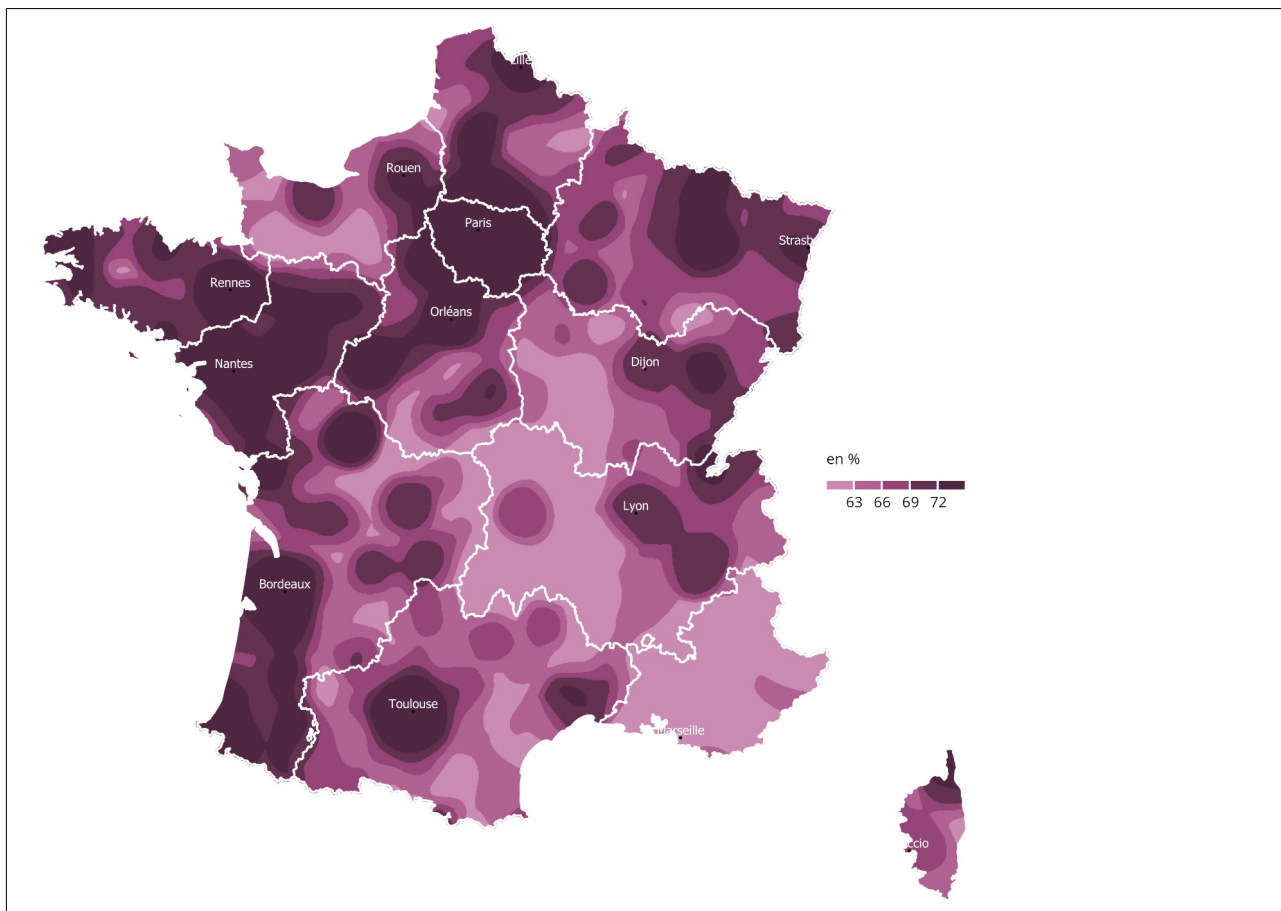


Lecture : il est estimé que 10 personnes sur 100 sont exposés au risque d'illectronisme dans l'aire d'attraction de la ville de Nantes. Champ : personnes de 15 ans ou plus résidant dans une aire d'attraction des villes de 700 000 habitants ou plus dont le pôle est situé en France.

Source : Insee, enquête TIC-ménages 2019, recensement de la population 2018.

Enfin, en 2019, 71 % des habitants de France métropolitaine utilisent Internet tous les jours ou presque. La part de ces usagers semble la plus élevée sur toute la façade atlantique, dans les Pays de la Loire, mais aussi en Île-de-France et sa périphérie (carte 3). Elle semble également supérieure à la moyenne dans les grandes aires d'attraction des villes, comme Bordeaux, Nantes, Rennes ou Toulouse. Pour autant, elle paraît aussi élevée dans des aires plus petites, mais attractives pour les étudiants, comme Angers, Besançon, Metz, Nancy, Poitiers, Tours, Amiens, Rouen ou encore Orléans. Situées à moins de 150 kilomètres de Paris, les trois dernières sont également attractives pour les actifs qui travaillent dans la capitale.

Carte 3 : carte lissée de la part estimée d'usagers quasi-quotidiens d'Internet au niveau intercommunal pour chaque région



Lecture : la part estimée d'usagers quasi quotidiens d'Internet dépasse 72 % dans toute l'Île-de-France.

Champ : France métropolitaine, personnes de 15 ans ou plus vivant en logement ordinaire.

Source : Insee, enquête TIC-ménages 2019, recensement de la population 2018.

D'autres résultats régionaux et infrarégionaux sont présentés dans Branche-Seigeot (2023).

Conclusion

En Bourgogne-Franche-Comté, sur 17 variables où la méthode de scoring a été appliquée, les résultats sont conformes aux estimations calées pour 14 d'entre elles, et plus discutables pour 3 autres (au moins une incompétence, utiliser Internet pour s'informer ou contacter un service public ou une administration). Cette cohérence est un gage de robustesse dans l'estimation infra-régionale de phénomènes habituellement étudiés au niveau national. Cette démarche pourrait répondre à des demandes importantes de la part des acteurs publics régionaux sur les compétences numériques et les usages d'Internet. Elle pourrait aussi, au besoin, être appliquée à d'autres mailles géographiques infrarégionales, ou à d'autres phénomènes à condition qu'ils dépendent essentiellement de caractéristiques individuelles.

Bien entendu, cette méthode de scoring n'est pas exempte de limites. La principale tient à la disponibilité et aux choix des variables mises en jeu dans les modèles. Si le contexte influence peu l'illectronisme, il ne faut pas pour autant écarter la recherche de meilleures variables contextuelles (ici : couverture numérique, enclavement, caractère montagneux...). Le cas de l'emploi présenté en annexe 3 semble d'ailleurs assez révélateur à cet égard : une variable de distance à la frontière par exemple, dont l'intérêt est connu à l'aide des analyses d'emploi de l'Insee, permettrait sans doute d'expliquer les écarts observés. Cette méthode ne tient pas non plus compte des éventuels effets portés par des politiques publiques locales. Enfin, si des données sur le revenu existaient dans la source externe (ici, le Recensement de la Population), des variables relatives au niveau de vie des ménages pourraient également être mobilisées, ce qui affinerait les estimations au-delà de la simple catégorie-socioprofessionnelle ou du quintile de niveau de vie de la commune de résidence.

Cette méthode d'infrarégionalisation permet toutefois de comparer assez justement les régions et les départements entre eux (ou autre maille géographique plus fine encore) du point de vue des compétences, de l'équipement et de la fréquence d'utilisation d'Internet.

Bibliographie

- P. Ardilly, « Panorama des principales méthodes d'estimation sur les petits domaines », Document de travail Insee, M0602, 2006. <https://www.insee.fr/fr/statistiques/fichier/1380679/m0602.pdf>
- P. Ardilly, « Estimation sur petits domaines à l'Insee: premier bilan et pistes de progrès » Actes Des 7e Colloque Francophone Sur Les Sondages. 2012.
http://sondages2012.ensai.fr/wp-content/uploads/2011/01/Article_Rennes_2012_ARDILLY.pdf
- A. Branche-Seigeot, "Davantage d'illectronisme dans les communes éloignées des villes et les petits pôles urbains", Insee Analyses n°85, 2023.
- C. Couet, « Estimations locales dans le cadre de l'enquête HID, Document de travail Insee, F0207, 2002.
<https://www.bnsp.insee.fr/ark:/12148/bc6p06zqvpq.pdf>
- L. Davezies & X. D'Haultfoeuille, « Faut-il pondérer ?... Ou l'éternelle question de l'économètre confronté à des données d'enquête », Documents de travail N° G2009/06, Insee, juillet 2009.
<https://www.insee.fr/fr/statistiques/1380863>
- A. Fievet, S. Hilary & A. Moineau, « 800 000 habitants en situation d'illectronisme », Insee Analyses Hauts-de-France n°118, décembre 2020. <https://www.insee.fr/fr/statistiques/4986976>
- S. Legleye & A. Rolland, « Une personne sur six n'utilise pas Internet, plus d'un usager sur trois manque de compétences numériques de base », Insee Première n° 1780, octobre 2019
<https://www.insee.fr/fr/statistiques/4241397>
- J. Le Guennec, « Estimations régionales dans l'enquête nationale Transport et Déplacements 2007-2008 », Document de travail Insee, H2012/03, 2012.
<https://www.insee.fr/fr/statistiques/fichier/1381022/H2012-03.pdf>
- A. Rolland, « L'usage des technologies de l'information et de la communication par les ménages entre 2009 et 2018 », Insee Résultats, mars 2019.
<https://www.insee.fr/fr/statistiques/4463497?sommaire=4466268>
- O. Sautory, « Redressement d'un échantillon par calage sur marges », Document de travail de la DSDS n°F9310, Insee, 1993. <https://www.insee.fr/fr/statistiques/fichier/2838097/5-les-methodes-de-calage.pdf>.
- Schouten, B., Cobben, F. et Bethlehem, J. (2009). Indicateurs de la représentativité de la réponse aux enquêtes. Techniques d'enquête, 35(1), 107-121.
- Schouten, B., Bethlehem, J., Beullens, K., Kleven, Ø., Loosveldt, G., Luiten, A., Rutar, K., Shlomo, N. et Skinner, C. (2012). Evaluating, comparing, monitoring, and improving representativeness of survey response through R-indicators and partial R-indicators. International Statistical Review, 80, 382 -39.

ANNEXES :

Annexe 1. Distribution initiale des variables auxiliaires selon la source (avant calage sur marges)

	TIC-MENAGES		RP
	FRANCE DE PROVINCE	Bourgogne-Franche-Comté	Bourgogne-Franche-Comté
	10 977	622	580 949
	%	%	%
De 15 à 29 ans	18,5%	16,8%	19,5%
De 30 à 44 ans	24,2%	23,2%	21,2%
De 45 à 59 ans	23,6%	22,7%	24,3%
De 60 à 69 ans	16,4%	18,5%	16,1%
De 70 à 79 ans	10,3%	11,1%	10,4%
80 ans ou plus	7,0%	7,6%	8,6%
Hommes	47,4%	47,0%	48,2%
Femmes	52,6%	53,0%	51,8%
Brevet des collèges, BEPC	31,1%	33,1%	31,9%
CAP-BEP	23,4%	22,3%	27,3%
Baccalauréat	18,8%	20,0%	17,4%
Diplômes du supérieur	26,8%	24,6%	23,3%
Actifs occupés	49,2%	48,0%	48,0%
Etudiants	8,6%	8,3%	6,9%
Actifs inoccupés	5,3%	6,9%	6,8%
Retraités	31,1%	33,3%	31,7%
Autres inactifs	5,8%	3,6%	6,6%
Agriculteurs	2,3%	3,4%	3,1%
Artisans et chefs d'entreprises de moins de 10 salariés	5,7%	5,1%	5,7%
Cadres et professions intellectuelles supérieures	12,4%	11,6%	8,4%
Professions intermédiaires	19,7%	17,0%	19,1%
Employés	25,5%	26,9%	25,1%
Ouvriers	21,3%	23,4%	24,2%
Autres inactifs	13,0%	12,5%	14,5%
Personnes seules	20,3%	19,1%	20,9%
Couples sans enfant(s)	28,9%	32,1%	32,0%
Couple avec enfant(s)	22,3%	20,9%	26,1%
Familles monoparentales	3,9%	3,6%	4,7%
Autres	24,5%	24,3%	16,3%
Commune-centre	29,2%	31,3%	31,7%
Autre commune d'un pôle	13,5%	5,3%	6,4%
Commune de la couronne	48,6%	54,0%	51,3%
Commune hors attraction des pôles	8,7%	9,4%	10,6%

Champ : individus de 15 ans ou plus vivant en logement ordinaire en France de Province

Source : enquête TIC-ménages, Recensement de la population 2018 ; données pondérées.

Note : ici, la catégorie socio-professionnelle est celle des actifs, occupés ou non, mais aussi celle du dernier poste occupé par les retraités avant leur sortie du marché du travail.

Annexe 2. Recherche du meilleur modèle prédictif de l'illectronisme

Afin de sélectionner les variables catégorielles et continues liées à l'illectronisme, plusieurs modèles logistiques ont été testés selon la technique du stepwise, avec ou sans répliquions bootstrap¹⁴, et avec ou sans interactions entre les variables auxiliaires. Les résultats indiquent que :

- En bivarié, chaque variable auxiliaire est significativement corrélée à l'illectronisme.
- En multivarié, l'âge, le diplôme, la catégorie socio-professionnelle et la situation familiale sont systématiquement liés à la variable « illectronisme ». C'est également presque systématiquement le cas pour les variables « situation professionnelle » et « quintile de niveau de vie communal ». En revanche, le sexe, la part communale de locaux éligibles au THD fixe ou encore la catégorie de l'aire d'attraction des villes sont quant à elles des variables qui ressortent beaucoup moins souvent.

Suite à ces indications, trois modèles logistiques ont été retenus et testés pour le scoring à partir des données non-pondérées :

- **Modèle 1 (modèle de base)** : un modèle sans interaction avec toutes les variables auxiliaires, individuelles et contextuelles.

- **Modèle 2** : le meilleur modèle sans interaction identifié. Il s'agit du modèle 1 (modèle de base) duquel ont été retirés la catégorie de l'AAV, la part de locaux éligibles au THD fixe et le sexe (ces trois variables ne sont d'ailleurs pas ressorties significatives dans le modèle de base)

- **Modèle 3** : le meilleur modèle avec interactions identifié. Celui-ci correspond au modèle 2 duquel a été retiré la variable « quintile de niveau de vie communal » et auquel a été ajouté une variable croisée « Age*CS ».

Au regard des tests de robustesse et d'ajustement, les trois modèles sont quasi-équivalents :

Résultats des tests de robustesse et d'ajustement des modèles

		Modèle 1 (modèle de base)	Modèle 2	Modèle 3
AIC		7145,2	7135,9	7142,9
AUC		0,893	0,893	0,894
R ²		32,7%	32,7%	32,9%
% concordants		89,3%	89,2%	89,3%
Test Hosmer – Lemershow	khi-2	12,41	13,33	3,61
	Pr > Khi-2	0,14	0,10	0,89

14. Permet de créer des nouveaux échantillons par tirage avec remise, à partir de l'échantillon initial.

En particulier, le pouvoir prédictif et la qualité de l'ajustement des modèles sans interaction (modèle 1 et 2) sont similaires et très bons. Concernant le modèle 3, la variable croisée « Age*CS » n'améliore pas ou peu le modèle initial. Enfin, au regard des tests d'Hosmer-Lemeshow, le risque d'un écart important entre les valeurs prédites et les valeurs observées semble davantage écarté dans les modèles 1 et 3 que dans le modèle 2. De fait, le modèle de base retenu est le modèle 1 (présenté dans la partie 3.2). Le modèle 3, avec interaction, ne l'est pas pour éviter tout risque de surajustement du modèle, le nombre de modalités croisées étant important.

Modèle 2			
	Paramètres	Estimation	Pr > khi-2
	Intercept	-3,096	0,000
Age	De 15 à 29 ans	-1,073	0,000
	De 30 à 44 ans	-0,992	0,000
	De 45 à 59 ans	Réf	Réf
	De 60 à 69 ans	0,475	0,000
	De 70 à 79 ans	1,430	0,000
	80 ans ou plus	2,430	0,000
Niveau de diplôme le plus élevé	Brevet des collèges, BEPC	1,178	0,000
	CAP-BEP	0,550	0,000
	Baccalauréat	Réf	Réf
	Diplômes du supérieur	-0,362	0,012
Situation Professionnelle	Actifs occupés	Réf	Réf
	Etudiants	-0,879	0,066
	Actifs inoccupés	-0,042	0,825
	Retraités	0,318	0,016
	Autres inactifs	0,472	0,003
Catégorie socio-professionnelle	Agriculteurs	1,173	0,000
	Artisans et chefs d'entreprises < 10 sal.	0,366	0,007
	Cadres et professions intellect. sup.	-0,599	0,000
	Professions intermédiaires	Réf	Réf
	Employés	0,417	0,000
	Ouvriers	1,182	0,000
Situation familiale	Autres inactifs	1,127	0,000
	Personnes seules	Réf	Réf
	Couples sans enfant(s)	-0,661	0,000
	Couples avec enfant(s)	-1,007	0,000
	Familles monoparentales	-1,436	0,000
Quintile de niveau de vie communal	Autres	-0,496	0,000
	1er quintile (communes modestes)	0,299	0,000
	2e quintile	0,267	0,002
	3e quintile	0,064	0,439
	4e quintile (communes aisées)	Réf	Réf

Source : enquête TIC-ménages 2019,

Champ : individus âgés de 15 ans ou plus vivant dans un logement ordinaire en France de Province

Modèle 3			
	Paramètres	Estimation	Pr > khi-2
	Intercept	-2,985	0,000
Age	De 15 à 29 ans	0,616	0,168
	De 30 à 44 ans	-1,117	0,042
	De 45 à 59 ans	Réf	Réf
	De 60 à 69 ans	0,252	0,389
	De 70 à 79 ans	1,359	0,000
	80 ans ou plus	2,513	0,000
Niveau de diplôme le plus élevé	Brevet des collèges, BEPC	1,186	0,000
	CAP-BEP	0,559	0,000
	Baccalauréat	Réf	Réf
	Diplômes du supérieur	-0,408	0,005
Situation Professionnelle	Actifs occupés	Réf	Réf
	Etudiants	-0,807	0,173
	Actifs inoccupés	0,067	0,723
	Retraités	0,403	0,006
	Autres inactifs	0,603	0,001
Catégorie socio-professionnelle	Agriculteurs	1,822	0,000
	Artisans et chefs d'entreprises < 10 sal.	0,820	0,016
	Cadres et professions intellect. sup.	-0,169	0,663
	Professions intermédiaires	Réf	Réf
	Employés	0,412	0,107
	Ouvriers	1,178	0,000
	Autres inactifs	0,706	0,038
Situation familiale	Personnes seules	Réf	Réf
	Couples sans enfant(s)	-0,668	0,000
	Couples avec enfant(s)	-1,062	0,000
	Familles monoparentales	-1,478	0,000
	Autres	-0,510	0,000
Variable croisée Age x catégorie-socioprofessionnelle	AGEG1*Csc1	-10,922	0,954
	AGEG1*Csc2	0,017	0,989
	AGEG1*Csc3	-8,588	0,849
	AGEG1*Csc5	-2,605	0,019
	AGEG1*Csc6	-3,392	0,002
	AGEG1*Csc7	-1,323	0,063
	AGEG2*Csc1	2,185	0,012
	AGEG2*Csc2	0,636	0,425
	AGEG2*Csc3	0,005	0,996
	AGEG2*Csc5	-0,003	0,997
	AGEG2*Csc6	-0,138	0,822
	AGEG2*Csc7	0,570	0,387
	AGEG3*Csc1	-0,981	0,090
	AGEG3*Csc2	-0,149	0,739
	AGEG3*Csc3	-0,410	0,431
	AGEG3*Csc5	0,217	0,497
	AGEG3*Csc6	0,197	0,516
	AGEG3*Csc7	0,980	0,018
	AGEG5*Csc1	-0,521	0,349
	AGEG5*Csc2	-0,637	0,121
	AGEG5*Csc3	-0,374	0,410
	AGEG5*Csc5	-0,018	0,953
	AGEG5*Csc6	0,134	0,654
	AGEG5*Csc7	0,503	0,206
	AGEG6*Csc1	-0,925	0,093
	AGEG6*Csc2	-0,826	0,054
	AGEG6*Csc3	-0,670	0,159
	AGEG6*Csc5	-0,050	0,882
	AGEG6*Csc6	-0,038	0,913
	AGEG6*Csc7	0,146	0,717

Source : enquête TIC-ménages 2019,

Champ : individus âgés de 15 ans ou plus vivant dans un logement ordinaire en France de Province

Annexe 3. Pertinence de la méthode de scoring : exemple avec l'emploi, une donnée connue dans le Recensement de la Population

Afin de tester la robustesse de la méthode de scoring pour obtenir des résultats au niveau infra-régional, elle a été appliquée sur la variable « être en emploi versus ne pas l'être ». En effet, « être en emploi ou non » est une variable « explicite » dans les données individuelles du Recensement de la Population, contrairement à la variable « être en situation d'illectronisme ou non » : il sera alors possible de comparer les chiffres obtenus avec la méthode de scoring (sur les répondants de France de Province à TIC-ménages) avec les chiffres « réels » du RP régional (ici, de Bourgogne-Franche-Comté).

La démarche mise en œuvre est la même que pour l'illectronisme :

- déterminer les variables les plus souvent identifiées comme liées à l'emploi, sans et avec interactions, sur l'échantillon de répondants à France de Province de TIC-ménages 2019.
- réaliser un modèle logistique sur ce même échantillon en y intégrant les variables les plus pertinentes : âge, diplôme, sexe, situation familiale, la catégorie de l'aire d'attraction de la commune de résidence¹⁵.
- regarder s'il y avait des spécificités régionales à l'aide d'une indicatrice région croisée aux variables socio-démographiques. Il s'avère qu'aucune n'était significative au seuil de 5 %. Aucune indicatrice région croisée n'a donc été ajoutée au modèle.
- calculer, dans le RP 2018 régional cette fois, la probabilité pour chaque individu d'être en emploi à partir des coefficients issus du modèle logistique réalisé sur l'échantillon de répondants de France de Province à TIC-ménages
- réaliser une comparaison des données du modèle (données théoriques) avec les données « réelles », par département et par EPCI.

Les résultats de cette comparaison sont mis en évidence sous forme de graphique pour les départements de la région Bourgogne-Franche-Comté et sous forme de carte pour les intercommunalités (cf. supra).

Au niveau départemental, s'il existe parfois des écarts un peu plus importants entre les données théoriques et réelles, la structure reste globalement la même entre les départements, ce qui importe in fine. En effet, même si l'estimation n'est pas parfaite (et elle ne peut pas l'être), elle conserve cependant les écarts structurels entre les territoires, ce qui ne modifie pas les messages en infra-régional. La figure ci-dessous donne les parts de personnes se déclarant en emploi d'après les données théoriques issues de la méthode de scoring et les données réelles. Au niveau régional, l'écart est de 1,3 points de pourcentage, ce qui peut être considéré comme acceptable.

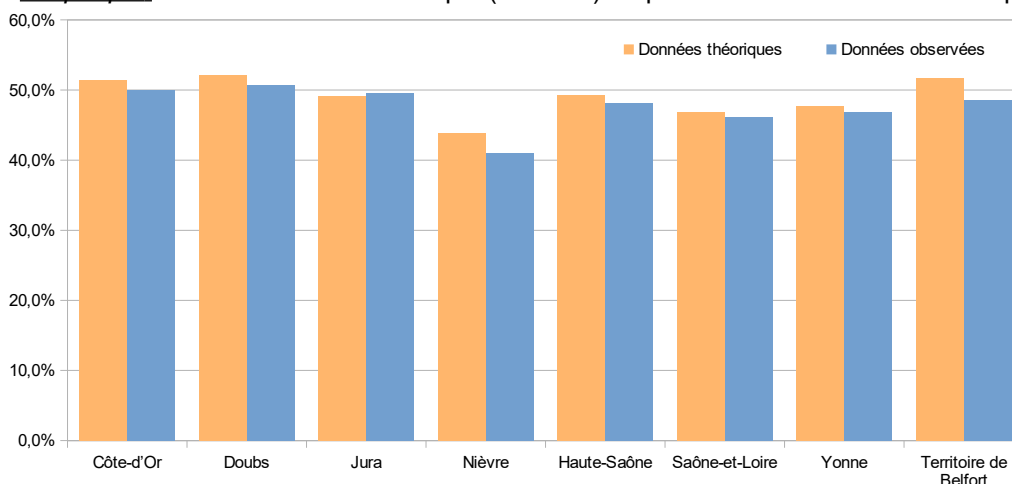
Globalement, par département mais aussi par EPCI, les écarts sont assez minces, ce qui montre que la méthode de scoring donne des résultats plutôt satisfaisants par rapport aux données réelles (carte).

Seul le Doubs a plusieurs EPCI où les écarts sont un peu plus marqués (entre 3 et 7 points de pourcentage de différence). Cependant, ce sont principalement des EPCI de la bande frontalière avec la Suisse qui sont

15. De nombreuses variables croisées sont ressorties comme significatives mais le modèle s'avérait être mal ajusté lorsqu'elles étaient introduites dans le modèle logistique (test de Hosmer et Lemeshow significatif). De même, comme il est préférable ne pas ajouter trop d'interactions pour éviter tout surapprentissage du modèle, ces variables croisées n'ont pas été intégrées compte tenu des nombreuses modalités ainsi générées.

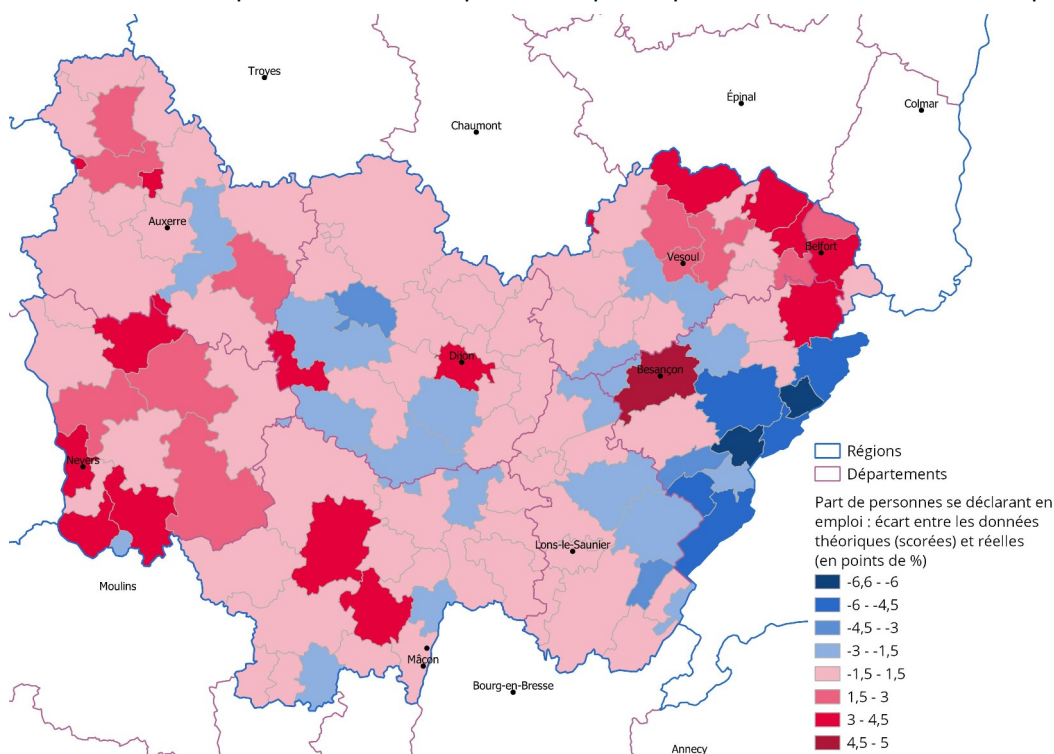
donc soumis à des effets particuliers liés à l'emploi non pris en compte par le modèle. Ainsi, celui-ci sous-estime la situation, la Suisse représentant un vivier d'emplois pour ceux qui résident non loin. A l'inverse, dans certains EPCI, la part de personnes se déclarant en emploi peut être surestimée. Cela peut correspondre à une baisse de l'offre d'emploi dans le secteur privé compte tenu des aléas conjoncturels et/ou dans le secteur public sous l'effet de la réduction des effectifs publics. Le modèle de scoring négligeant ces effets d'offre, les écarts entre la part théorique et la part réelle de personnes se déclarant en emploi peuvent alors être importants. Il serait néanmoins tout à fait possible d'intégrer ces effets d'éloignement à la frontière dans le modèle et les données.

Graphique : Part observée et théorique (=scorée) de personnes se déclarant en emploi



Champ : individus de 15 ans ou plus en Bourgogne-Franche-Comté
 Source : Recensement de la population 2018 ; Enquête TIC-ménages 2019

Carte : Ecart entre la part observée et la part théorique de personnes se déclarant en emploi



Champ : individus de 15 ans ou plus vivant en Bourgogne-Franche-Comté
 Source : Recensement de la population 2018 ; Enquête TIC-ménages 2019

Série des Documents de Travail « Action régionale »

H2011/01

Panorama des villes moyennes
Jean-Michel Floch et Bernard Morel

H2012/01

Les déterminants régionaux de l'innovation
Benoît Buisson, Lionel Doisneau, Claire Kubrak, Michelle Mongo et Corinne Autant-Bernard

H2012/02

Estimations locales du handicap dans l'enquête Handicap-Santé 2008
Josiane Le Guennec

H2012/03

Estimations régionales dans l'enquête nationale Transport et Déplacements 2007-2008
Josiane Le Guennec

H2012/04

Détection des disparités socio-économiques L'apport de la statistique spatiale
Jean-Michel Floch

H2013/01

Concentration et spécialisation des activités économiques : des outils pour analyser les tissus productifs locaux
Claire Kubrak

H2014/01

La ségrégation spatiale dans les grandes unités urbaines de France métropolitaine : une approche par les revenus
Gaëlle Dabet Jean-Michel Floch

H2014/02

Rapport du groupe de travail DREES/INSEE/DRJSCS sur les indicateurs locaux de suivi du plan pluriannuel de lutte contre la pauvreté et pour l'inclusion sociale (PPLPIS)

H2016/01

Élaboration des Produits Intérieurs Bruts régionaux en base 2010 : Principes et méthodologie
Benoît Hurpeau

H2018/01

Structurel, résiduel, géographique : Principe et mise en oeuvre des approches comptable et économétrique
Claire Kubrak

H2018/02

Les dynamiques de la qualité de vie dans les territoires
Robert Reynard Pascal Vialette

H2018/03

Guide de sémiologie graphique

H2018/04

Guide de sémiologie cartographique

H2019/01

Les entreprises régionales Concepts, usages, mesure statistique
Lionel Doisneau

H2021/01

Des projections nationales aux projections locales
Laure Crusson Jérôme Fabre

H2021/02

Décomposition de l'évolution de l'emploi par catégorie d'entreprise

Hervé Bacheré Benoît Mirouse Zoé Brassier

2022-18

La grille communale de densité à 7 niveaux
Simon Beck, Marie-Pierre De Bellefon, Jocelyn Forest, Mathilde Gerardin, David Levy,

2023-01

Cas d'utilisation des indices de Theil pour mesurer les disparités spatiales selon les revenus à l'infracommunal
Mathilde Gerardin Julien Pramil

2023-11

La rétopolation en 2010 du zonage en aires d'attraction des villes de 2020
Simon Beck Olivier Pégaz-Blanc Adem Khamallah

2023-15

Utilisation d'une méthode de scoring pour des estimations locales à partir d'une enquête nationale : exemple avec l'enquête TIC-ménages
Aline Branche-Seigeot