



Chapitre 1

PRÉSENTATION DE L'ENQUÊTE ET DES MÉTHODES STATISTIQUES

1.1	PRÉSENTATION DE L'ENQUÊTE	10
1.2	DESCRIPTION DE L'ÉCHANTILLON RÉGIONAL	12
1.3	STATISTIQUES DESCRIPTIVES UNIVARIÉES ET BIVARIÉES	14
1.4	ANALYSES MULTIVARIÉES	15
1.5	COMPARAISON DES RÉSULTATS RÉGION/FRANCE	18
	Bibliographie	18

L'ESSENTIEL

L'enquête Baromètre santé environnement dans les Pays de La Loire

Le Baromètre santé environnement des Pays de la Loire a été réalisé par l'Observatoire régional de la santé (ORS), grâce au soutien financier de l'État (Direction régionale des affaires sanitaires et sociales) et du Conseil régional des Pays de la Loire.

Il s'agit de la déclinaison régionale d'une enquête conduite au niveau national par l'Institut de prévention et d'éducation pour la santé (Inpes), selon un protocole identique.

Les opinions de la population ont été recueillies au moyen d'un questionnaire téléphonique (d'une durée de 40 minutes environ), auprès d'un échantillon représentatif de la population. Le recueil des données s'est déroulé entre janvier et mai 2007, dans la région comme au plan national.

Au total, dans la région, 2 294 individus âgés de 18 à 75 ans, ont été interrogés. 380 Ligériens l'ont été dans le cadre de l'enquête réalisée au plan national par l'Inpes (dont l'effectif total est de 6 007 individus), et 1 914 dans le cadre de sa déclinaison régionale mise en place par l'ORS (suréchantillon régional). Les résultats présentés dans ce rapport concernent les 2 294 Ligériens interrogés.

Les taux de refus (personnes contactées mais qui n'ont pas souhaité participer à l'enquête ou qui n'ont pas répondu au questionnaire dans son intégralité), sont respectivement de 37,5 % dans la région et de 42 % au niveau national.

Les données ont été redressées sur la base de la structure par sexe, croisée par âge, taille d'agglomération et département de la population au recensement de 1999. Le redressement garantit la représentativité de l'échantillon au niveau régional, pour les variables sélectionnées.

Les méthodes statistiques utilisées

Des analyses bivariées ont été réalisées pour examiner l'influence de différentes variables sociodémographiques (sexe, âge, catégorie socioprofessionnelle...) sur les réponses des Ligériens aux différentes questions de l'enquête. Des analyses multivariées (régressions logistiques) ont également été mises en œuvre afin de déterminer l'effet d'une variable, indépendamment des effets de structure liés aux autres variables introduites dans l'analyse.

Les enquêtes nationale et régionale ayant été conduites selon un protocole identique, des comparaisons région/France ont également pu être réalisées. Des modèles de régression logistiques ont été utilisés afin de comparer les réponses des Ligériens et celles des Français, en s'affranchissant d'éventuelles différences de caractéristiques sociodémographiques entre ces deux populations. Les comparaisons ont ainsi pu être menées à structure par âge, sexe, catégorie socioprofessionnelle, diplômes, revenus, taille d'agglomération identique entre la région et la France.

1.1 Présentation de l'enquête

Les enquêtes Baromètres santé

L'Institut national de prévention et d'éducation pour la santé (Inpes) conduit depuis 1992 des enquêtes « Baromètres Santé » dont l'objectif est le suivi régulier (tous les 3 à 5 ans) d'un ensemble d'indicateurs décrivant les opinions, les attitudes et les comportements ainsi que les connaissances de la population à l'égard de thèmes de santé. Ces enquêtes permettent de disposer d'éléments de

connaissances utiles notamment à l'orientation et à l'évaluation des programmes de prévention et d'éducation à la santé, et plus généralement, des politiques dans ce domaine. Elles sont menées en partenariat avec les institutions et les équipes de recherche qui interviennent dans le champ de la santé publique. Depuis 1999, certains Observatoires régionaux de la santé (ORS) déclinent ces enquêtes en région.

Pour la première fois en 2007, dans le cadre du Plan national santé environnement 2004-2008, l'Inpes a réalisé un Baromètre santé environnement consacré aux connaissances de la population sur les relations entre l'environnement et la santé. Cette enquête, qui doit être reconduite tous les cinq ans, a également été menée dans sept régions de France métropolitaine, dont la région Pays de la Loire.

Le Baromètre santé environnement Pays de la Loire 2007

Le Baromètre santé environnement des Pays de la Loire a été réalisé grâce au soutien financier de l'Etat (Direction régionale des affaires sanitaires et sociales) et du Conseil régional des Pays de la Loire.

Les enquêtes régionale et nationale ont été menées selon un protocole identique, permettant des comparaisons entre la région et la France.

Le questionnaire¹ téléphonique a été conçu au niveau national par un groupe d'experts pluridisciplinaire. Les entretiens téléphoniques, d'une durée moyenne de 40 minutes ont été réalisés par l'institut de sondage Atoo. En Pays de la Loire, l'enquête s'est déroulée du 24 janvier au 15 mai 2007, parallèlement à l'enquête nationale, menée entre le 22 janvier et le 21 mai 2007.

L'effectif de l'échantillon global de la région des Pays de la Loire s'élève à 2 294 individus, âgés de 18 à 75 ans. 380 Ligériens ont été interrogés dans le cadre de l'enquête réalisée au plan national par l'Inpes (dont l'effectif total est de 6 007 individus), et 1 914 dans le cadre de sa déclinaison régionale mise en place par l'ORS (suréchantillon régional).

La méthode d'échantillonnage

Les échantillons (suréchantillon régional de 1 914 individus, échantillon national de 6 007 individus) ont été obtenus par une méthode de sondage aléatoire à deux degrés.

La première étape de cette méthode a consisté à tirer au sort des ménages en composant au hasard des numéros de téléphone inscrits sur les listes blanche, orange et rouge, après stratification sur la taille de l'agglomération et le département (les numéros sur liste rouge ont été obtenus en incrémentant de un les derniers chiffres des numéros de l'annuaire téléphonique).

La deuxième étape du sondage avait pour objectif de sélectionner de manière aléatoire une seule personne âgée de 18 à 75 ans par ménage. Cette

sélection a été réalisée sur la base de la date d'anniversaire la plus proche à venir. Seule cette personne a été invitée à participer à l'enquête. Les ménages ne comportant aucun individu âgé de 18 à 75 ans n'ont pas été retenus.

Les personnes qui ne disposent pas de ligne fixe, ne sont pas prises en compte par cette méthode de sondage. C'est en particulier le cas des personnes détentrices uniquement d'un téléphone portable, qui représentent à l'heure actuelle 15 % de l'ensemble de la population et se distinguent par certaines spécificités socio-démographiques (population jeune, milieux moins favorisés) [1]. Leur exclusion, justifiée par des contraintes budgétaires, peut entraîner un biais de représentativité de l'enquête. Mais elle permet en contrepartie de disposer d'un échantillon de plus grande taille, plus robuste, et de réaliser des analyses détaillées.

Information des personnes enquêtées et taux de refus

Avant tout appel, une lettre de présentation de l'étude a été adressée aux ménages sélectionnés, à l'exception des ménages sur liste rouge pour lesquels son envoi a été proposé lors du premier contact téléphonique. Pour les ménages sélectionnés dans le cadre du suréchantillon régional, la lettre présentait l'enquête réalisée en Pays de la Loire et était adressée par l'ORS. Les ménages ligériens sélectionnés dans le cadre de l'enquête nationale ont reçu le courrier de l'enquête nationale transmis par l'Inpes.

Les taux de refus sont respectivement de 44 % pour les enquêtés sur liste rouge et de 36 % pour les enquêtés sur liste blanche ou orange, soit un taux de refus global de 37,5 % (42 % au niveau national) [1] (cf. annexe 1.1, p. 232).

Pondération et redressement

Les données ont été pondérées afin de tenir compte de la probabilité de sélection d'un individu au sein du ménage (un individu d'un grand ménage a moins de chance d'être tiré au sort qu'un individu d'un petit ménage) et redressées sur la base de la structure par sexe, croisée par âge, taille d'agglomération et département de la population au recensement de 1999 (cf. annexe 1.2, p. 233). Le redressement permet notamment de compenser en partie la différence de structure liée à l'exclusion des personnes ne possédant pas de ligne fixe [1].

1. Le questionnaire de l'enquête est disponible sur le site internet www.santepaysdelaloire.com.



1.2 Description de l'échantillon régional

Représentativité de l'échantillon du Baromètre santé environnement

En raison du redressement, l'échantillon présente la même structure par sexe, croisée par âge, taille d'agglomération et département que celle décrite par le recensement de 1999. Cette structure de la population régionale a évolué au cours de la dernière décennie mais les données des enquêtes annuelles de recensement disponibles au moment de l'exploitation des résultats de l'enquête² ne sont pas suffisamment détaillées pour l'actualiser.

Concernant les répartitions de la population selon les caractéristiques socioprofessionnelles et les niveaux de diplôme, des différences sont observées entre l'échantillon du Baromètre santé environnement et les données du recensement.

- Dans l'échantillon du Baromètre santé environnement, les catégories sociales les moins favorisées sont sous-représentées (56 % des actifs de l'échantillon sont agriculteurs exploitants, employés, ouvriers ou sans activité professionnelle vs 62 % selon les données 2004-2006 du recensement), en raison d'une proportion moindre d'ouvriers (25 % vs 30 %).
- De même, bien que les classes d'âge ne soient pas les mêmes (18-75 ans dans le Baromètre santé environnement et 14 ans et plus dans les enquêtes annuelles de recensement), les personnes les moins diplômées apparaissent sous-représentées dans l'échantillon du Baromètre santé environnement (50 % des personnes enquêtées sont titulaires d'un diplôme de niveau inférieur au bac vs 66 % selon les données du recensement 2004-2007). Les personnes les plus diplômées apparaissent au contraire

surreprésentées (18 % de personnes titulaires d'un diplôme de niveau bac+3 et plus ou de grandes écoles vs 8 %).

Spécificités régionales

Néanmoins, malgré ces différences, les spécificités des Pays de la Loire par rapport à la France décrites par les données des recensements de population réalisés par l'Insee (1990, 1999, enquêtes annuelles depuis 2004) se retrouvent dans les échantillons régional et national du Baromètre santé environnement.

- La région se distingue ainsi par une moindre proportion de cadres (12 % des habitants de la région selon les résultats du Baromètre santé environnement vs 15 % au niveau national, $p < 0,01$), et par une surreprésentation des ouvriers (22 % vs 17 %, $p < 0,001$) et des agriculteurs (4 % vs 3 %, $p < 0,001$).
- En cohérence avec ces caractéristiques socio-professionnelles, la part des personnes disposant de revenus supérieurs à 1 500 € par unité de consommation (UC) est inférieure à celle observée au niveau national (32 % dans la région vs 36 % en France, $p < 0,001$).
- Concernant le niveau de diplôme, la part plus importante de personnes titulaires de diplômes de niveau inférieur au bac dans la région (49 % vs 45 %, $p < 0,001$) et la moindre proportion de diplômés de niveau bac+3 et plus (18 % vs 22 %, $p < 0,001$) se retrouvent également dans l'échantillon.
- Enfin, la région se caractérise par une plus forte proportion de personnes vivant en zone rurale (28 % vs 17 % en France, $p < 0,001$) (tableau 1).

Tableau 1 : description des échantillons des enquêtes Baromètres santé environnement réalisées en Pays de la Loire et en France métropolitaine selon certaines caractéristiques sociodémographiques

	Pays de la Loire (échantillon global) (n=2 294)	Pays de la Loire (suréchantillon régional) (n=1 914)	France métropolitaine (n=6 007)
	%	%	%
Sexe			NS
Homme	49,4	49,4	49,0
Femme	50,6	50,6	51,0
Age			NS
18 à 25 ans	15,6	15,6	14,6
26 à 34 ans	17,6	17,6	18,3
35 à 44 ans	20,3	20,3	20,6

2. 2004-2006 ou 2004-2007

	Pays de la Loire (échantillon global) (n=2 294)	Pays de la Loire (suréchantillon régional) (n=1 914)	France métropolitaine (n=6 007)
	%	%	%
45 à 54 ans	19,4	19,4	19,7
55 à 64 ans	12,9	12,9	13,1
65 à 75 ans	14,2	14,2	13,7
Professions et catégories socioprofessionnelles (PCS)¹			***
Agriculteurs exploitants (PCS-)	4,3	4,3	2,8
Artisans, commerçants et chefs d'ent. (PCS+)	4,9	5,1	6,3
Cadres et prof. intell. sup. (PCS+)	12,6	12,5	15,1
Professions intermédiaires (PCS+)	21,6	22,2	21,8
Employés (PCS-)	25,0	24,7	25,9
Ouvriers (PCS-)	22,0	22,2	16,7
Autres pers. sans activité prof. (PCS-) ²	9,5	9,1	11,4
PCS en deux classes³			**
PCS (-)	54,7	54,2	50,6
PCS (+)	45,3	45,8	49,4
Niveau d'éducation			***
Aucun diplôme ou inf. au Bac	49,6	49,5	45,0
Bac	21,4	21,2	20,8
Bac+2	11,2	11,4	12,7
Bac+3 et sup. et grandes écoles	17,8	17,9	21,5
Revenus mensuels⁴			***
< 900 € par unité de consommation	29,3	29,1	25,8
900-1 500 € par unité de consommation	39,6	39,1	37,7
> 1 500 € par unité de consommation	31,1	31,8	36,5
Zonage⁵			***
Rural	28,2	27,5	17,0
Urbain	71,8	72,5	83,0
Taille agglomération			***
Commune rurale	34,0	34,0	24,2
2 000-19 999 habitants	23,4	23,4	16,7
20 000-99 999 habitants	7,6	7,6	13,4
100 000 habitants et plus	35,0	35,0	45,7

Source : Baromètre santé environnement Pays de la Loire 2007. Inpes, ORS Pays de la Loire

Résultats des tests de comparaison entre le suréchantillon Pays de la Loire et l'échantillon France :

NS : non significatif ; * différence significative au seuil de 5 %, ** différence significative au seuil de 1 %, *** différence significative au seuil de 0,1 %.

Note de lecture : l'échantillon global Pays de la Loire est constitué des 1 914 individus interrogés dans le cadre de l'enquête régionale et 380 Ligériens interrogés dans le cadre de l'enquête nationale.

1. Les retraités et les chômeurs ayant déjà travaillé sont classés en fonction de leur dernière profession.
2. Personnes sans profession : chômeurs n'ayant jamais travaillé, élèves, étudiants, femmes au foyer, inactifs
3. PCS (-) : agriculteurs exploitants, employés, ouvriers, autres personnes sans profession à l'exception des élèves et des étudiants
PCS (+) : cadres et professions intellectuelles supérieures, professions intermédiaires, artisans, commerçants et chefs d'entreprise, élèves et étudiants
4. Le revenu mensuel total du foyer (y compris allocations, rentes, bénéficiaires...) est rapporté au nombre de personnes le constituant selon la conversion suivante : le premier adulte du foyer correspond à une unité de consommation (UC), les autres individus de 14 ans ou plus à 0,5 UC et les enfants de moins de 14 ans à 0,3 UC.
5. Le Zonage en aires urbaines (ZAU), défini par l'Insee, décline le territoire en quatre catégories. La première représente l'espace à dominante rurale qui comprend à la fois des petites unités urbaines et des communes rurales. Les trois autres constituent l'espace à dominante urbaine : ce sont les pôles urbains, les couronnes périurbaines et les communes multipolarisées.

1.3 Statistiques descriptives univariées et bivariées

Les résultats régionaux sont présentés sous la forme de proportions, calculées après agrégation des données du suréchantillon régional (1 914 individus) et des données concernant les Ligériens enquêtés dans le cadre de l'enquête nationale (380 individus). Les estimations de ces proportions prennent en compte le plan de sondage, mais pour des raisons de simplification, les intervalles de confiance associés ne sont pas présentés dans cette étude.

Analyses bivariées (tris croisés)

Ces proportions ont été comparées entre sous-groupes de population selon certaines caractéristiques (sociodémographiques ou autres comme vivre à proximité d'une source de pollution...).

Ces croisements ont été systématiquement réalisés pour une sélection de variables socio-démographiques (tableau 2).

Des tests statistiques de significativité (test du khi-deux d'indépendance) ont été utilisés pour déterminer si la différence de résultats observée entre deux ou plusieurs groupes de population (par exemple, les différences de proportions de

personnes se déclarant satisfaites de l'information sur les risques de santé liés à l'environnement selon leur catégorie socioprofessionnelle) peut être considérée comme statistiquement significative ou bien due aux fluctuations d'échantillonnage.

Le résultat du test est donné par la probabilité « p » de se tromper en rejetant l'hypothèse d'absence de lien entre les deux variables (satisfaction à l'égard de l'information et catégorie socioprofessionnelle). Plus cette probabilité est faible et plus l'hypothèse d'absence de lien (hypothèse d'indépendance) est peu vraisemblable. Trois degrés de significativité sont habituellement utilisés en statistiques : $p < 0,05$, $p < 0,01$ et $p < 0,001$. Pour un risque d'erreur de 5 % ($p < 0,05$), l'hypothèse d'indépendance peut être rejetée : la différence entre les proportions observées est significative et la probabilité d'erreur associée à cette décision est inférieure à 5 % (il y a moins de 5 chances sur 100 pour que les deux caractères soient indépendants).

Les estimations de proportions et les analyses bivariées ont été réalisées avec le module « Echantillons complexes » du logiciel SPSS. Ce module permet de prendre en compte le plan de sondage utilisé lors de l'enquête pour les estimations statistiques.

Tableau 2 : variables sociodémographiques de croisement sélectionnées

Variable	Modalités
Sexe	Homme, Femme
Age	18 à 25 ans, 26 à 34 ans, 35 à 44 ans, 45 à 54 ans, 55 à 64 ans, 65 à 75 ans
Professions et catégories socioprofessionnelles (PCS) en deux classes ¹	PCS (-) : agriculteurs exploitants, employés, ouvriers, autres personnes sans profession à l'exception des élèves et des étudiants ; PCS (+) : cadres et professions intellectuelles supérieures, professions intermédiaires, artisans, commerçants et chefs d'entreprise, élèves et étudiants
Niveau de diplôme ²	Aucun diplôme ou inférieur au Bac, Bac, Bac+2, Bac+3 et sup. et grandes écoles
Revenus mensuels	Inférieurs à 900 € par unité de consommation, 900-1 500 € par unité de consommation, supérieurs à 1 500 € par unité de consommation
Taille agglomération	Commune rurale, 2 000-19 999 habitants, 20 000-99 999 habitants, 100 000 habitants et plus

Source : Baromètre santé environnement Pays de la Loire 2007. Inpes, ORS Pays de la Loire

1. Les effectifs régionaux ne permettent pas une analyse par catégorie socioprofessionnelle détaillée.

2. De même, pour des raisons d'effectifs, les catégories de diplôme « Bac+3 ou 4 » et « Bac+5 et grandes écoles » ont été regroupées.

1.4 Analyses multivariées

Régression logistique : effet des variables sociodémographiques, « toutes choses égales par ailleurs »

Lorsqu'une différence significative est mise en évidence par l'analyse bivariée, par exemple, la différence de proportions de personnes se déclarant satisfaites de l'information selon leur catégorie socioprofessionnelle, elle peut ne pas être directement liée à la variable étudiée (catégorie socioprofessionnelle) mais à une ou plusieurs autre(s) variable(s) (âge, niveau de diplôme...). Ainsi, à âge, diplôme (...) identiques, il n'y a pas de différence de satisfaction selon la catégorie socioprofessionnelle.

Des régressions logistiques ont été utilisées afin de contrôler l'existence d'éventuels effets de structure liés à une ou plusieurs caractéristiques de la population. Les variables systématiquement introduites dans les régressions logistiques sont les mêmes que celles sélectionnées pour les tris croisés (tableau 2). La sélection systématique des mêmes variables ne conduit pas nécessairement à un modèle de régression satisfaisant. En outre, certaines variables du modèle ne sont pas indépendantes. Cependant, les modèles de régressions logistiques ne sont pas utilisés ici à des fins prédictives, mais uniquement pour quantifier la force du lien entre la variable à expliquer et les variables explicatives.

L'existence d'une association entre l'évènement étudié (par exemple, la satisfaction à l'égard de l'information) et certaines caractéristiques de la population (âge, sexe...) est évaluée à l'aide de l'indicateur « Odds Ratio » (Rapport de cotes). Cet indicateur permet d'évaluer l'existence d'une augmentation de la probabilité d'occurrence de l'évènement étudié (satisfaction à l'égard de l'information) associé à une caractéristique de la population (être âgé de 65 à 75 ans) par rapport à une caractéristique de référence (être âgé de 18 à 25 ans).

L'Odds Ratio est très utilisé en épidémiologie, notamment pour évaluer les risques de développer une maladie dans certains groupes de population. Il peut être assimilé à une estimation de l'augmentation du risque relatif (augmentation du risque de contracter la maladie dans un groupe de population par rapport à un autre), mais uniquement si la grandeur observée est rare (inférieure à 1 %). Cette approximation est valable dans le cas de maladies peu fréquentes, mais ne

peut pas être utilisée lorsqu'il s'agit d'opinions ou de comportements plus fréquents, tels que ceux étudiés par le Baromètre santé environnement.

De manière générale, seules les associations entre deux variables significatives en régression logistique au seuil de 5 %, c'est-à-dire après ajustement sur les autres variables introduites dans le modèle de régression, sont commentées dans cette étude. Pour des raisons de simplification, ces associations sont commentées et illustrées à partir des proportions et des résultats de l'analyse bivariée (test du khi-deux). Dans les cas rares où l'effet d'une variable est mis en évidence par la régression logistique, mais n'est pas retrouvé en analyse bivariée, la relation statistique est présentée dans l'étude à partir des résultats de la régression logistique.

Exemple : satisfaction à l'égard de l'information reçue, en général, sur les risques de santé liés à son environnement

Les résultats des analyses bivariées mettent en évidence des différences statistiquement significatives de la proportion de personnes « satisfaites à l'égard de l'information » selon les caractéristiques sociodémographiques - sexe, âge, catégorie socioprofessionnelle et niveau de diplôme (tests du khi-deux).

Mais le modèle de régression logistique montre que, après contrôle des effets de structure liés aux caractéristiques de la population, seuls les effets du sexe et de l'âge sont confirmés. Ainsi, par exemple, à structure par catégorie socioprofessionnelle, diplômes, revenus et taille d'agglomération identiques, les personnes âgées de 65 à 75 ans ont une probabilité supérieure (OR=2,6) de se déclarer satisfaites de l'information par rapport aux personnes âgées de 18 à 25 ans (modalité de référence) (tableau 3).

Dans l'étude, seuls les effets du sexe et de l'âge seront donc commentés, et présentés de la manière suivante :

- Les hommes se déclarent plus souvent satisfaits que les femmes (53 % vs 47 %, $p < 0,01$).
- De manière générale, la satisfaction à l'égard de l'information augmente avec l'âge, de 43 % des personnes âgées de 18 à 25 ans à 64 % des 65-75 ans ($p < 0,001$).

Les régressions logistiques ont été réalisées avec le module « Echantillons complexes » du logiciel SPSS.

Tableau 3 : exemple de tableau de résultats des analyses statistiques

Satisfaction à l'égard de l'information sur les risques liés à son environnement
selon certaines caractéristiques sociodémographiques : analyses univariées et multivariées

	Se déclarer plutôt satisfait(e) de l'information reçue sur les risque liés à son environnement	
	(n=2239)	
	%	OR ajusté
Total	49,9 [47,6-52,2]	
Sexe	**	***
Homme	53,2	Référence
Femme	46,7	0,7 ***
Age	***	***
18 à 25 ans	42,7	Référence
26 à 34 ans	43,0	1,2
35 à 44 ans	44,8	1,2
45 à 54 ans	51,0	1,5 *
55 à 64 ans	59,1	2,3 ***
65 à 75 ans	64,4	2,6 ***
Professions et catégories socioprofessionnelles (PCS)	*	NS
PCS (-)	52,7	-
PCS (+)	46,6	-
Niveau d'éducation	**	NS
Aucun diplôme ou inf. au Bac	54,2	-
Bac	48,6	-
Bac+2	39,7	-
Bac+3 et sup. et grandes écoles	47,2	-
Revenus mensuels	NS	NS
< 900 €/UC	52,5	-
900-1 500 €/UC	49,6	-
> 1 500 €/UC	46,5	-
Taille agglomération	NS	NS
Commune rurale	51,6	
2 000-19 999 habitants	49,4	
20 000-99 999 habitants	51,9	
100 000 habitants et plus	48,3	

Source : Baromètre santé environnement Pays de la Loire 2007. Inpes, ORS Pays de la Loire

NS : non significatif ; * : p<0,05 ; ** : p<0,01 ; *** : p<0,001

Note de lecture :

La première colonne présente les proportions et les résultats des tests du khi-deux.

La deuxième colonne présente les résultats de la régression logistique. Lorsqu'une variable a un effet significatif, le nombre d'étoiles indique la significativité de la variable dans le modèle. Les Odds Ratio associés à chaque modalité et leur significativité sont alors mentionnés.

Analyse factorielle et typologie des individus

Pour certains chapitres du Baromètre santé environnement, une Analyse des correspondances multiples (ACM) suivie d'une Classification ascendante hiérarchique (CAH) ont été réalisées afin de dégager des profils d'individus homogènes au regard de leurs représentations des problèmes environnementaux et des risques sanitaires associés [2, 3].

L'ACM est une méthode de réduction du nombre de variables pour permettre des représentations géométriques des variables et des individus. La réduction du nombre de variables ne se fait pas par une simple sélection de certaines d'entre elles mais par la construction de nouvelles variables synthétiques (les facteurs). Par exemple, si on considère les 2 300 individus de l'enquête et leurs réponses à dix questions, géométriquement, ce tableau correspond à un nuage de 2 300 points situés dans un espace à dix dimensions. Un tel espace n'étant pas visualisable, l'ACM permet de déterminer un sous-espace de dimension réduite, « compréhensible » visuellement, sur lequel est projeté le nuage de points des individus et les variables qualitatives. Cette méthode constitue donc une extension de la statistique descriptive au cas multidimensionnel et autorise une analyse simultanée de plusieurs variables qualitatives [4].

Les méthodes de classification ou de typologie ont pour but de regrouper les individus en un nombre restreint de classes homogènes. La CAH est une des méthodes de classification qui a pour objectif d'obtenir, à partir des facteurs issus de l'ACM, des classes d'individus les plus cohérentes possibles : soit en constituant les groupes les plus homogènes ou, à l'inverse, les groupes qui se distinguent le plus les uns des autres. La mesure communément utilisée pour juger de la qualité

d'une CAH est le quotient de l'inertie interclasse sur l'inertie totale du nuage des individus : plus il est élevé et plus les individus regroupés forment des groupes homogènes nettement différenciés les uns des autres [4].

Ces analyses ont été réalisées à l'aide du logiciel SPAD. Les variables actives retenues pour l'ACM sont les différentes questions de santé-environnement, les variables sociodémographiques³ étant projetées en variables supplémentaires sur les axes factoriels (elles ne participent pas à la construction de ces axes). Les modalités des variables représentant moins de 5 % des réponses des enquêtés ont été exclues de l'analyse.

Le nombre d'axes factoriels de l'ACM retenus pour la classification a été déterminé de manière à représenter au moins 80 % de l'inertie totale.

Le logiciel SPAD utilise une procédure itérative de calcul (appelée méthode de « consolidation ») afin de déterminer les partitions de qualité optimale pour le critère d'homogénéité des classes. Plusieurs partitions sont proposées par la procédure automatique du logiciel SPAD. Les typologies présentées dans ce document sont celles qui résumaient au mieux les résultats détaillés du chapitre.

Seuls les résultats de la typologie sont présentés. Les modalités des variables caractéristiques de chaque groupe d'individus présentées sont les modalités les plus discriminantes. Ainsi, si la proportion d'individus ayant répondu à une modalité dans le groupe identifié par la CAH est significativement supérieure à la proportion d'individus ayant répondu à cette même modalité dans l'ensemble de l'échantillon régional, cette modalité permet de distinguer le groupe [5].

3. Sexe, âge, catégorie socioprofessionnelle en deux classes, niveau de diplôme, niveau de revenu, taille d'agglomération



1.5 Comparaison des résultats région/France

Afin d'assurer l'indépendance de l'échantillon des Pays de la Loire et de l'échantillon national, ainsi que la représentativité de l'échantillon national, les 380 Ligériens interrogés dans le cadre de l'enquête nationale n'ont pas été intégrés à l'échantillon régional pour ces comparaisons. Ce dernier compte ainsi 1 914 individus, et l'échantillon France 6 007.

Les proportions présentées dans les tris à plat et les tris croisés régionaux (calculées sur 2 294 individus) peuvent ainsi différer légèrement de celles utilisées pour la comparaison région/France (calculées respectivement sur 1 914 et 6 007 individus). Pour des raisons de simplification, seules les proportions estimées à partir des 2 294 individus sont présentées, y compris dans les comparaisons région/France.

Les structures par âge et sexe dans la région et en France étant peu différentes, les résultats n'ont pas fait l'objet d'une standardisation préalable sur ces deux variables pour les comparaisons région/France.

La région se distinguant par d'autres caractéristiques (structures par catégorie socioprofessionnelle ou niveau de diplôme différentes dans la région et au niveau national par exemple), un modèle de régression logistique a été utilisé

afin d'intégrer ces spécificités et d'effectuer ainsi une comparaison des résultats entre les Pays de la Loire et la France, « toutes choses égales par ailleurs ». Dans le modèle de régression, les mêmes variables socio-économiques que pour les analyses des résultats régionaux ont été insérées : âge, sexe, catégorie socioprofessionnelle en deux classes, niveau de diplôme, niveau de revenu et taille d'agglomération (tableau 2). La régression logistique permet de comparer les résultats entre la région et la France à structure par âge, sexe, catégorie socioprofessionnelle en deux classes, diplômes, revenus et taille d'agglomération identique.

Pour certains chapitres (pollution de l'air extérieur, bruit), compte tenu de la situation très particulière de la région Ile-de-France à l'égard de ces problèmes environnementaux, des comparaisons ont été réalisées entre les Pays de la Loire (2 294 individus), la région Ile-de-France (743 individus) et les autres régions de France métropolitaine (4 883 individus). La représentativité des sous-échantillons « Ile-de-France » et « autres régions » n'étant pas garantie, ces comparaisons n'ont qu'une valeur indicative.

Les régressions logistiques ont été réalisées avec le module « Echantillons complexes » du logiciel SPSS.

Bibliographie

- [1] Méthodologie d'enquête. F. Beck, C. Léon, R. Guignard. Dans *Baromètre santé environnement 2007*. Sous la direction de C. Ménard, D. Girard, C. Léon, F. Beck. Ed. Inpes, 2008, pp. 43-55 (Coll. Baromètres Santé)
- [2] Analyses factorielles multiples. X. Bry. Ed. Economica, 1996, 112 p.
- [3] Analyses factorielles simples. X. Bry. Ed. Economica, 1995, 112 p.
- [4] Probabilités, analyses des données statistiques. G. Saporta. Ed. Technip, 1990, 493 p.
- [5] SPAD profiling. Manuel d'aide à l'interprétation. T. Le Nouvel. SPAD, 2006, 129 p.