

Enquête 2010 auprès de la Génération 2007 Bilan des redressements

Pascale Rouaud

Département Entrées et évolutions dans la vie active (Deeva), Céreq.
rouaud@cereq.fr

Céreq
10, place de la Joliette BP21321
13 567 Marseille Cedex 02

Ce document est présenté sur le site du Céreq afin de favoriser la diffusion et la discussion de résultats de travaux d'études et de recherches. Il propose un état d'avancement provisoire d'une réflexion pouvant déboucher sur une publication. Les hypothèses et points de vue qu'il expose, de même que sa présentation et son titre, n'engagent pas le Céreq et sont de la responsabilité des auteurs.

août 2012

SYNTHÈSE

Dans le cadre de l'enquête Génération 2007, 25 000 jeunes ont été interrogés en 2010 sur leurs premières années de vie active. Retenus comme appartenant au champ de l'enquête Céreq, ils sont représentatifs des 739 000 jeunes sortis pour la première fois de formation initiale en 2006-2007, tous niveaux confondus.

Ce document méthodologique présente de manière détaillée le mode d'élaboration de l'échantillon, ainsi que les procédures de pondération et de redressement qui permettent d'obtenir cette représentativité.

L'enquête 2010 auprès de la Génération 2007, contrairement aux enquêtes précédentes, présente la particularité d'avoir un échantillon essentiellement associé à des demandes d'**extensions** ciblées sur des formations ou des territoires particuliers. Ainsi, le nombre de répondants visé initialement pour l'enquête nationale hors extensions était de 3 000, auquel s'ajoutaient 27 000 répondants attendus supplémentaires répartis entre les différentes extensions.

Le poids des extensions dans l'échantillon a rendu plus complexe la procédure de redressement. Dans certains cas, le champ des individus répondants était plus large que celui défini par le Céreq, nécessitant des redressements *ad hoc*. Certains modules de questionnement ont concerné uniquement un sous-échantillon des répondants, leur exploitation nécessitant des redressements spécifiques. Ce mode d'échantillonnage complexe a impliqué de satisfaire à plusieurs impératifs. Il s'agissait d'une part, pour les répondants appartenant au champ de l'enquête du Céreq, d'établir une pondération nationale qui intègre de manière cohérente les données de chaque extension. D'autre part, pour chaque extension réalisée sur un champ différent de celui du Céreq, il fallait mettre en œuvre un mode de redressement spécifique. Enfin, il a fallu s'assurer de la cohérence globale des jeux de pondération en comparant notamment la valeur des indicateurs clés obtenus à champ équivalent.

De la constitution de la base de sondage, aux coefficients de redressement, en passant par les taux de couverture, de nombreuses dimensions statistiques sont abordées et ce pour l'ensemble des pondérations développées.

Pour en savoir plus sur les enquêtes Génération :

<http://www.cereq.fr/index.php/themes/Acces-aux-donnees-Themes/Enquetes-d-insertion-Generation>

SOMMAIRE

1.	LE DISPOSITIF GÉNÉRATION	5
2.	PRÉSENTATION DE L'ENQUÊTE 2010 AUPRÈS DE LA GÉNÉRATION 2007.....	7
2.1.	Le champ de l'enquête	7
2.2.	Les extensions	8
2.3.	La constitution de la base de sondage.....	10
2.4.	Le plan de sondage.....	11
2.5.	Les coefficients intervenant dans la pondération	14
3.	ESTIMATION DE LA COUVERTURE	15
3.1.	Les établissements bénéficiant d'une couverture exhaustive	15
3.2.	Les établissements dont la couverture a été estimée à partir d'une référence	16
3.2.1.	Le champ des Centres de Formation par Apprentissage (CFA) (type=P).....	18
3.2.2.	Le champ des formations du domaine de la Santé (type=E) et du Social (type=D).....	21
3.2.3.	Le champ des Instituts Universitaires de Formation des Maîtres (IUFM) (type=H).....	23
3.2.4.	Le champ des écoles de commerce (type=B) et ingénieurs (type=A).....	25
3.2.5.	Le champ des formations agricoles (type=Q)	25
3.3.	Les établissements dont la couverture a été estimée sans information auxiliaire	28
3.4.	Récapitulatif	29
4.	PONDÉRATION DES REpondants DANS LE CHAMP DU CEReq.....	31
4.1.	La reconstitution des poids de tirage.....	31
4.1.1.	La méthodologie	31
4.1.2.	Les poids de tirage de l'échantillon principal	32
4.1.3.	L'utilisation de l'échantillon de réserve.....	34
4.1.4.	Les poids de tirage de l'échantillon utilisé (PDS).....	34
4.2.	La modélisation de l'envoi en plateau d'enquête	40
4.3.	La modélisation de la probabilité de réponse après envoi en enquête	43
4.4.	La suppression des individus hors du champ de l'enquête et calage aux marges	46
4.5.	Le traitement des poids élevés.....	48
4.6.	La pondération arrondie	49
4.7.	Récapitulatif	50
5.	AUTRES PONDÉRATIONS	51
5.1.	L'extension des 5 000	51
5.1.1.	La reconstitution des poids de tirage.....	51
5.1.2.	Le calcul de la pondération des 5 000	52
5.1.3.	La pondération des stages	53
5.2.	L'extension Pôle Emploi	54
5.3.	L'extension Drees.....	54

5.3.1.	Le contexte.....	54
5.3.2.	Les résultats de l'enquête.....	56
5.3.3.	Les redressements	56
5.4.	L'extension Sport	58
5.4.1.	Le contexte.....	58
5.4.2.	Les résultats de l'enquête.....	59
5.4.3.	Les redressements	59
5.5.	L'extension DGAFP	61
5.5.1.	Le contexte.....	61
5.5.2.	Les résultats de l'enquête.....	64
5.5.3.	Les redressements	65
ANNEXES		67
	La modélisation de l'envoi en enquête.....	67
	La modélisation de la probabilité de réponse suite à l'envoi en enquête.....	70
INDEX DES SIGLES.....		87

1. LE DISPOSITIF GÉNÉRATION

Le Centre d'études et de recherches sur les qualifications (Céreq) est un établissement public placé sous la tutelle des ministères en charge de l'Éducation nationale et du Travail. Il a pour missions de développer des études et des recherches, de collecter et d'exploiter des données originales dans le domaine de la relation formation-emploi, et de formuler des avis et propositions destinés à éclairer les choix en matière de politiques de formation.

A la fin des années quatre-vingt-dix, le Céreq a mis en place un dispositif d'enquêtes original qui permet d'étudier l'accès à l'emploi des jeunes à l'issue de leur formation initiale. Tous les trois ans, une nouvelle enquête est réalisée auprès de jeunes qui ont en commun d'être sortis du système éducatif la même année quel que soit le niveau ou le domaine de formation atteint, d'où la notion de Génération.

L'enquête permet de reconstituer les parcours des jeunes au cours de leurs trois premières années de vie active et d'analyser ces parcours au regard notamment du parcours scolaire et des diplômes obtenus. Certaines cohortes sont interrogées plusieurs fois pour suivre les débuts de carrière.

En s'appuyant sur un calendrier qui décrit mois par mois la situation des jeunes et sur des informations plus précises concernant le premier emploi et l'emploi occupé au bout de trois années passées sur le marché du travail, ce dispositif permet non seulement d'analyser les trajectoires d'entrée dans la vie active mais aussi de distinguer, d'une Génération à l'autre, les aspects structurels et conjoncturels de l'insertion.

L'enquête 2010 auprès de la Génération 2007 a été collectée au printemps 2010 auprès d'environ 29 000 jeunes sortis du système éducatif en France métropolitaine, au cours ou à la fin de l'année scolaire 2006-2007. Ils ont été interrogés sur leurs parcours scolaires et leurs premiers pas dans la vie active. Cette enquête de la Statistique publique, au sens de la loi n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques, s'inscrit dans le dispositif Génération mis en place par le Céreq à la fin des années quatre-vingt-dix.

L'intérêt et les atouts du dispositif reposent sur cinq dimensions majeures. Tout d'abord, un cadre d'analyse homogène et cohérent : le questionnement, la méthodologie et l'analyse sont les mêmes pour tous, quels que soient leur niveau ou spécialité de formation. Les informations recueillies sont riches et diversifiées. L'interrogation trois ans après la sortie du système éducatif permet un recul temporel suffisant. La description de la situation de l'enquêté mois par mois autorise un suivi longitudinal. Enfin, par construction, la conjoncture est la même pour tous les jeunes interrogés.

Figure 1

Enchaînement des enquêtes Génération

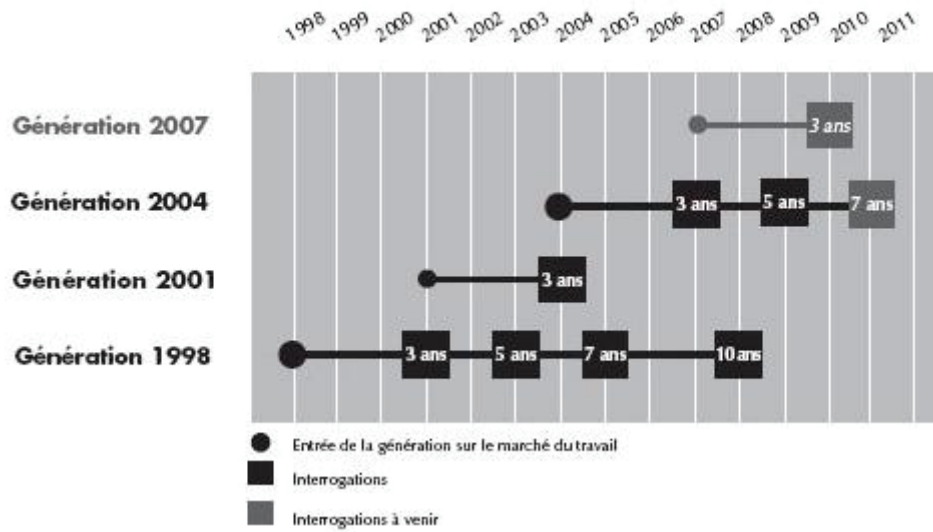


Tableau 1

Nombre de répondants au questionnaire à la première interrogation

	Année de la première interrogation	Nombre de répondants	Dont liés aux extensions d'échantillon
Génération 1992	1997	26 000	
Génération 1998	2001	55 000	19 000
Génération 2001	2004	25 000	15 000
Génération 2004	2007	65 000	37 000
Génération 2007	2010	29 000	26 000

Le cœur de l'enquête, pour la première interrogation comme pour les autres quand il y a lieu, est constitué par le calendrier professionnel qui permet de suivre mois par mois la situation des jeunes à l'issue de leur formation initiale, et de décrire les séquences successives d'emploi et de non emploi.

2. PRÉSENTATION DE L'ENQUÊTE 2010 AUPRÈS DE LA GÉNÉRATION 2007

Les enquêtes Génération permettent aussi de répondre à des demandes d'extensions d'échantillon ciblées qui conduisent à grossir la taille de l'échantillon. Pour des raisons budgétaires, l'échantillon national de base de l'enquête 2010 est très réduit. Contrairement aux enquêtes précédentes, l'essentiel de l'échantillon de l'enquête est associé à des demandes d'extensions ciblées sur des formations ou des territoires particuliers. Si les extensions larges et transversales, comme l'extension relative aux sortants de l'enseignement supérieur, contribuent à améliorer l'échantillon national de base, il n'en est pas de même des extensions plus spécifiques. Autrement dit, l'échantillon de l'enquête 2010 n'a pas la même représentativité qu'un échantillon de même taille qui aurait été tiré de façon aléatoire dans l'ensemble des formations et sur l'ensemble du territoire. De ce fait, les analyses trop détaillées ne sont pas possibles (par exemple les différenciations de trajectoire selon le pays de naissance des parents).

L'objectif principal de l'enquête 2010 auprès de la Génération 2007 était donc de produire et de mettre à jour des indicateurs standards d'insertion à trois ans (notamment le taux d'emploi, le taux de chômage, le salaire médian, la part des emplois à durée indéterminée) selon le plus haut diplôme obtenu, agrégé en dix grandes catégories :

- non diplômés ;
- diplômés d'un CAP ou BEP industriel ;
- diplômés d'un CAP ou BEP tertiaire ;
- diplômés d'un baccalauréat professionnel ou technologique industriel ;
- diplômés d'un baccalauréat professionnel tertiaire ;
- diplômés d'un baccalauréat général ou technologique tertiaire ;
- diplômés d'un bac+2 industriel ;
- diplômés d'un bac+2 santé ou tertiaire ;
- diplômés d'un second cycle (L3 et M1) ;
- diplômés d'un troisième cycle, d'une école de commerce ou d'ingénieur ou d'un doctorat.

L'enquête permet cependant d'affiner les résultats pour les diplômés de l'enseignement supérieur du fait d'une extension financée par la Direction de l'enseignement supérieur et de l'insertion professionnelle (DGESIP) sur ce champ.

2.1. Le champ de l'enquête

L'enquête 2010 auprès de la Génération 2007 concerne les « primo sortants » de formation initiale en 2006-2007 (année scolaire). Les post initiaux (sortants de formation qui avaient déjà interrompu leurs études pendant au moins un an avant l'année scolaire considérée) sont hors champ. Tous les niveaux et domaines de formations sont concernés. De façon plus précise, les critères d'éligibilité pour être dans le champ retenu, nommé ensuite « champ Céreq », sont les suivants :

- avoir été inscrit dans un établissement de formation en France métropolitaine durant l'année scolaire 2006-2007 ;

- avoir quitté le système éducatif entre octobre 2006 et octobre 2007 ;
- ne pas avoir interrompu ses études durant une année ou plus avant l'année scolaire 2006-2007 (sauf pour raison de santé) ;
- ne pas avoir repris ses études pendant l'année qui a suivi l'entrée sur le marché du travail ;
- avoir 35 ans ou moins en 2007 ;
- être localisé en France métropolitaine au moment de l'enquête (ce qui exclut donc notamment les personnes poursuivant des études à l'étranger ou y allant pour un premier emploi).

Ces conditions sont cumulatives.

Quelques points particuliers concernent l'application de ces critères :

Bien qu'il s'agisse de contrats de travail au même titre que les contrats de professionnalisation ou que les contrats de qualification, les contrats d'apprentissage sont considérés comme relevant de la formation initiale. Une personne repérée comme sortant de formation en 2006-2007 qui poursuit par un contrat d'apprentissage en 2007-2008 est donc considérée en poursuite d'études, donc hors champ. Ce n'est pas le cas s'il s'agit d'un contrat de professionnalisation ou de qualification.

Une personne sortie d'un établissement de formation en 2006-2007 qui poursuit des cours par correspondance ou des cours du soir en 2007-2008 est considérée comme en poursuite d'études, donc hors champ, si elle n'a pas d'emploi en parallèle. Si elle a un emploi en parallèle, sa situation d'emploi prime et elle est dans le champ de l'enquête.

Une personne sortie d'un établissement de formation en 2006-2007 mais poursuivant ses études en 2007-2008 comme élève fonctionnaire n'est pas considérée comme étant en poursuite d'études ; elle est considérée en emploi. Les personnes en deuxième année d'IUFM sont donc, en particulier, considérées comme étant en emploi.

2.2. Les extensions

Les acteurs intervenant dans le domaine de la formation sont multiples, avec leurs questionnements propres mais aussi avec un questionnement commun sur le devenir de leur public cible. A la demande de certains d'entre eux, ministères, conseils régionaux ou observatoires de branches, des extensions sont régulièrement adossées aux enquêtes Génération. Elles permettent de réaliser des analyses ciblées sur des publics, des filières ou des territoires particuliers dans un contexte plus large, par rapprochement avec des indicateurs de référence nationaux.

Plusieurs types d'extensions existent, parfois combinées :

- l'extension d'échantillon, pour disposer d'un nombre de répondants suffisant pour permettre des analyses statistiquement pertinentes sur les catégories de sortants ciblées ;
- l'extension de champ, pour interroger également sur les catégories de sortants ciblées les personnes non retenues dans le champ Céreq (par exemple en levant la restriction aux seuls primo sortants) ;
- l'extension de questionnement, pour poser quelques questions complémentaires sur des catégories de sortants ciblées.

Pour l'enquête 2010 auprès de la Génération 2007, le nombre de répondants visé initialement pour l'enquête nationale hors extensions était de 3 000, auquel s'ajoutaient 27 000 répondants attendus supplémentaires répartis entre les différentes extensions (voir Tableau 3), avec des phénomènes de saturation sur certaines strates.

¹ Sauf pour les formations du Sport et de l'Animation, de la Santé et du Social, ainsi que les doctorants, où les jeunes sortis entre novembre 2007 et décembre 2007 sont également enquêtés.

Tableau 2

Bilan des demandes d'extensions sur Génération 2007

Institution ou organisme demandeur d'extension	Cible de sortants	Objectif de questionnaires supplémentaires	Extension de champ	Extension de questionnement
Direction de l'Evaluation de la Prospective et de la Performance (DEPP)	Sortants des expérimentations de baccalauréats professionnels en trois ans et sortants de trois filières professionnelles	3 100	Non	Non
DGES-IP	Sortants de l'enseignement supérieur (y c docteurs hors santé)	4 500	Sorties jusqu'en décembre 2007 pour les docteurs	Module Thèse
Direction de la Recherche, des Etudes, de l'Evaluation et de la Statistique du Ministère de la Santé et des Affaires Sociales (Drees)	Formations sanitaires et sociales	5 300 (3 700 santé et 1 600 social)	Inclusion des post-initiaux, sorties jusqu'en décembre 2007, pas de limite d'âge	Module Post-initiaux
Pôle Emploi	Sortants de Zones Urbaines Sensibles (ZUS)	750	Non	Non
Pôle Emploi	Non Diplômés	500	Non	Non
Pôle Emploi	Echantillon des "5 000" (tous niveaux)	..	Non	Module Intermédiaires marché du travail
CGDD	Formations de l'environnement	3 000	Non	Non
Ministère de la Jeunesse, des Sports et de la Vie Associative	Diplômés des formations du Sport	2 000	Inclusion des post-initiaux, sorties jusqu'en décembre 2007, pas de limite d'âge	Modules Post-initiaux et Multi-activité
Région Nord-Pas-de-Calais	Tous les sortants de Nord-Pas-de-Calais (pour obtenir une représentativité au niveau régional en 10 niveaux de diplôme*)	2 000	Non	Non
Franche-Comté	Tous les sortants de Franche-Comté (pour obtenir une représentativité au niveau régional en 10 niveaux de diplôme*)	2 000	Non	Non
Région Poitou-Charentes	Tous les sortants de Poitou-Charentes (pour obtenir une représentativité au niveau régional en 10 niveaux de diplôme*)	2 000	Non	Non
Région Bretagne	Tous les sortants de Bretagne (pour obtenir une représentativité au niveau régional en 10 niveaux de diplôme*)	2 000	Non	Non
Région Aquitaine	Tous les sortants d'Aquitaine (pour obtenir une représentativité au niveau régional en 10 niveaux de diplôme*)	2 000	Non	Non
DGAFP	Sortants des écoles de la fonction publique	1 200	Tous les sortants des écoles collectées	Non
DGAFP	Echantillon des "5 000" (tous niveaux)	..	Non	Module Stage

*10 niveaux de diplômes : 1- jeunes sortis sans aucun diplôme 2- diplômés d'un Cap ou d'un Bep industriel 3- diplômés d'un Cap ou d'un Bep tertiaire 4- diplômés d'un Baccalauréat professionnel ou technologique industriel 5- diplômés d'un Baccalauréat professionnel tertiaire 6- diplômés d'un Baccalauréat général ou technologique tertiaire 7- diplômés d'un Bac+2 industriel 8- diplômés d'un Bac+2 santé ou tertiaire 9- diplômés d'un Second Cycle (L3 et M1) 10- diplômés d'un Troisième Cycle ou Ecoles (M2 et D)

2.3. La constitution de la base de sondage

Il n'existe pas de base d'élèves centralisée couvrant l'ensemble des formations initiales dispensées en France métropolitaine. Il faut donc la constituer spécifiquement pour l'enquête, à partir de différentes sources, dans le cadre d'une collecte ad hoc. La base ainsi constituée présente un défaut de couverture dû aux établissements qui n'ont pas fourni de bases d'élèves lors de la constitution de la base de sondage.

Par ailleurs, les données récupérées sont souvent des listes de sortants d'un établissement : il s'agit de personnes inscrites en 2006-2007 dans l'établissement et non réinscrites en 2007-2008 dans le même établissement. Elles peuvent néanmoins s'être réinscrites ensuite ailleurs et sont alors hors du champ de l'enquête. La base comporte aussi des doublons : les jeunes inscrits dans plusieurs établissements ou dans plusieurs filières. Il n'est possible de traiter que partiellement ces deux aspects lors de la constitution de la base de sondage.

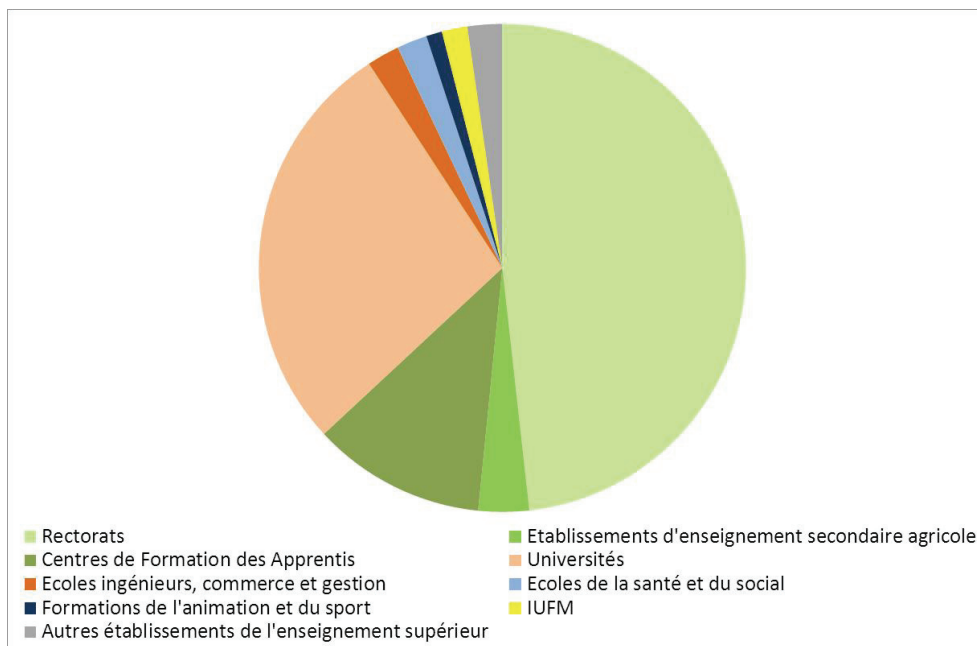
Pour chaque individu scolarisé pendant l'année scolaire 2006-2007 (présupposé sortant), les informations suivantes sont collectées : nom et prénom, adresse et numéros de téléphone, date de naissance, sexe, diplôme et spécialité de formation, obtention ou non du diplôme préparé.

Les rectorats disposent des listes d'inscrits dans les établissements de leur ressort : collèges, lycées et BTS dépendant du ministère en charge de l'éducation nationale (y compris les établissements privés sous contrat), grâce au système informatique Scolarité. Le Céreq a réalisé lui-même la collecte des informations souhaitées directement auprès des rectorats. La couverture des établissements est ainsi exhaustive sur le champ rectoral. Pour les autres types d'établissements, une collecte spécifique a été organisée. Les 3 700 établissements dispensant de la formation initiale répertoriés par la Banque Centrale des Etablissements (BCE) ont été sollicités par le Céreq pour transmettre leurs listes d'inscrits. Parmi eux, 65 % ont renvoyé les informations demandées par le Céreq.

Les fichiers reçus ont ensuite été compilés et corrigés : correction des numéros de téléphone, harmonisation des adresses, codification des diplômes et spécialités, suppression des individus âgés de plus de 35 ans, repérage et suppression des doublons. A l'issue de cette étape, la base de sondage contenait environ 1 250 000 individus.

Figure 2

Origine des sortants dans la base de sondage



2.4. Le plan de sondage

Tout comme pour les précédentes enquêtes Génération, l'enquête 2010 auprès de la Génération 2007 s'appuie sur un échantillonnage aléatoire. Par opposition à la méthode des quotas, si le plan de sondage est adapté et la collecte bien réalisée, le tirage d'un échantillon aléatoire permet d'obtenir des estimateurs d'Horvitz-Thompson des totaux de variables d'intérêt qui coïncident avec les vrais totaux de ces mêmes variables.

Le plan de sondage de l'enquête est un tirage aléatoire à un seul degré avec probabilités inégales. La probabilité de tirage individuelle est calculée de façon itérative : un taux de sondage initial permet d'abord de simuler un échantillon répondant aux seuls besoins du Céreq (réalisation de 3 000 questionnaires interprétables sur 10 niveaux de plus haut diplôme). Des surpondérations sont ensuite estimées de façon à répondre aux besoins des différentes extensions d'échantillon (pour atteindre un objectif global de 30 000 questionnaires). Une fois ces calculs réalisés, les poids de sondage finaux calculés permettent de procéder au tirage en une seule étape. Une variable agrégée de strates de formation (Strate07) a été construite afin de faciliter le tirage.

La définition du plan de sondage a nécessité de résoudre trois défis :

- dimensionner l'échantillon au plus juste afin de ne réaliser ni trop, ni trop peu de questionnaires, en faisant des hypothèses sur les taux de succès qui tiennent compte à la fois de la probabilité de joindre la personne recherchée, de la probabilité qu'elle accepte de répondre et de la probabilité qu'elle soit bien dans le champ de l'enquête ;
- tenir compte d'une clé de passage entre les niveaux de sortie, connus dans la base de sondage, et le plus haut diplôme attendu, inconnu avant l'enquête mais qui constitue la catégorisation cible pour la publication des résultats de base ;
- répartir l'échantillon entre les nombreuses extensions, point particulièrement crucial dans l'enquête 2010 puisque ces extensions constituent l'essentiel de l'échantillon.

Pour faire face à un éventuel taux de réponse insuffisant, un échantillon de « réserve », injectable en cours d'enquête en cas de besoin, a également été constitué.

En pratique, l'échantillon et la réserve ont été calibrés en faisant des hypothèses de taux de succès sur la base des résultats de l'enquête 2007 auprès de la Génération 2004. Il a été tenu compte de la forte amélioration de la qualité de la base de sondage en ce qui concerne la disponibilité de numéros de téléphone : 75% des individus présents dans la base de sondage de Génération 2007 ont au moins un numéro de téléphone contre 27% pour la base de sondage de la Génération 2004. Les hypothèses de taux de succès retenues ont été calculées sur la base des taux observés dans l'enquête 2007 en croisant la strate de sortie (Strate07) avec une variable indiquant la présence ou non d'un numéro de téléphone dans la base de sondage.

Pour les individus issus des fichiers des rectorats, aucune référence n'existait puisque les numéros de téléphone n'étaient pas transmis auparavant. Le taux de succès anticipé a donc été estimé sur la base du taux observé dans l'enquête 2007 pour les jeunes pour lesquels un numéro de téléphone avait été retrouvé sur la base du nom et de l'adresse du jeune dans la phase d'enrichissement de l'échantillon en numéros de téléphone (recherches dites « phase A »).

Pour passer du plus haut diplôme au niveau de sortie, une clef de passage a été estimée là encore à partir des résultats de l'enquête 2007 auprès de la Génération 2004.

Pour le sous-échantillon des « 5000 », concerné notamment par le module « stage », la difficulté du tirage consistait à tirer un échantillon « équilibré » sur le plus haut diplôme dans un échantillon global lui-même très déséquilibré par rapport à cette variable du fait des nombreuses extensions d'échantillon ciblées. Un coefficient de tirage a donc été calculé pour moduler le taux de tirage au sein de l'échantillon global en fonction de la strate de sortie (Strate07).

Le tirage a été réalisé sous le logiciel SAS avec l'aide de la macro procédure CUBE2 développée par l'INSEE. Ce type de tirage autorise les probabilités de tirage individuelles.

Le tirage est stratifié par région pour les 5 régions d'extension : Aquitaine, Franche-Comté, Poitou-Charentes, Bretagne et Nord-Pas de Calais. Un tirage indépendant est également effectué pour l'Île-de-France car cette région présente un taux de réponse et une structure en termes de diplômes/spécialités particulière. Les régions restantes sont regroupées dans 3 classes par proximité de taux de chômage et de taux de réponse.

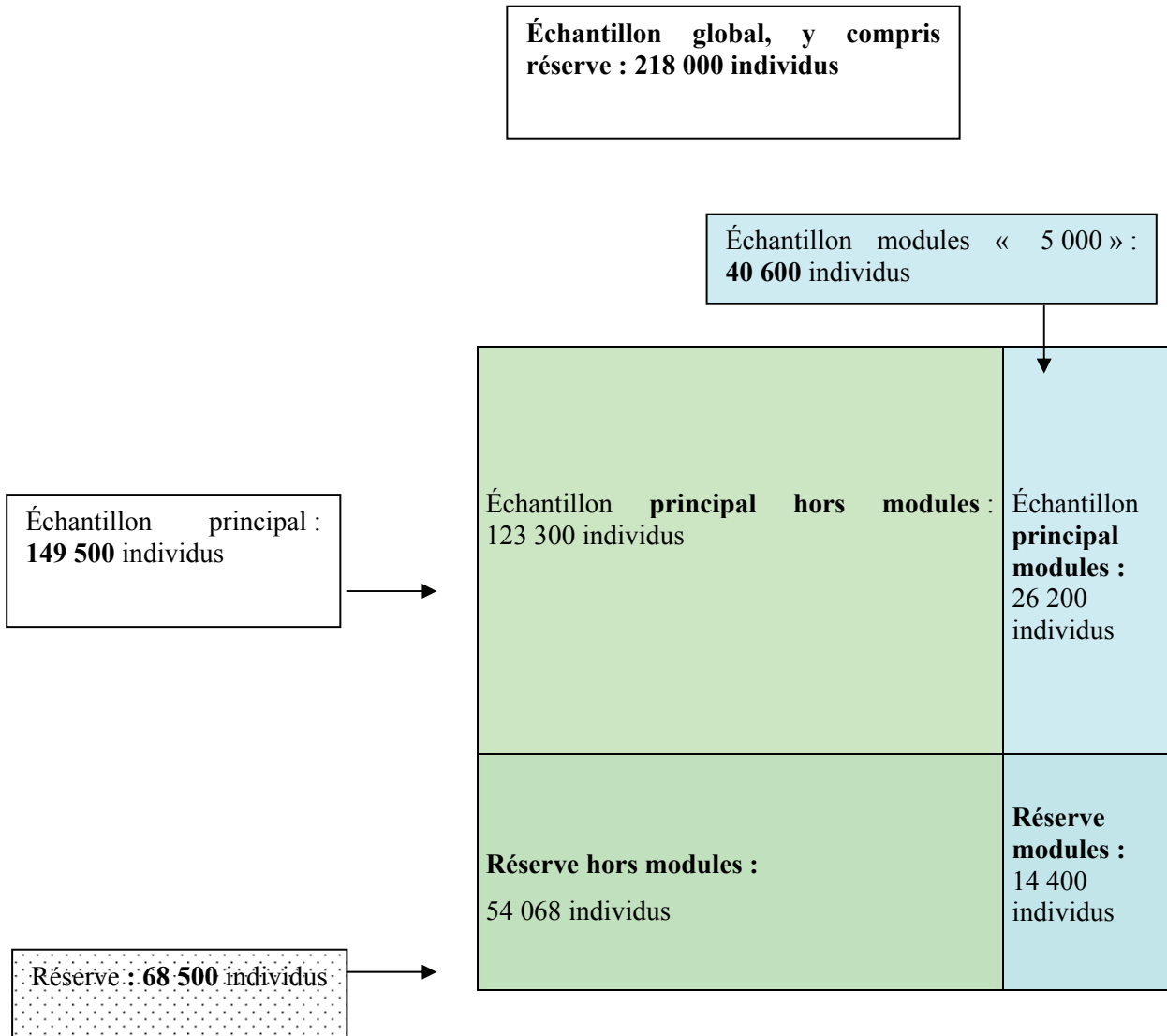
On a commencé par effectuer le tirage de l'échantillon y compris les réserves. Cet échantillon contenait 218 000 individus. Il incluait un sous-échantillon emboîté destiné à la production des 5 000 questionnaires du sous-échantillon des « 5 000 », pour les modules « Stages » et « Intermédiaires du marché du travail ». Un échantillon de 40 600 individus (réserve comprise) a été tiré à l'intérieur de l'échantillon global.

Un dernier tirage, au sein de l'échantillon des 218 000 individus, a permis de distinguer l'échantillon principal de l'échantillon de réserve.

² Pour plus d'informations, consulter : http://www.insee.fr/fr/methodes/outils/cube/documentation_cube_web.pdf

Figure 3

Constitution de l'échantillon de Génération 2007



2.5. Les coefficients intervenant dans la pondération

Les pondérations finales sont obtenues en tenant compte des paramètres suivants³ :

- couverture de la base de sondage ;
- probabilité individuelle d'appartenance à l'échantillon utilisé ;
- probabilité d'envoi en enquête (conditionné à la présence d'au moins un numéro de téléphone à l'issue des recherches de coordonnées téléphoniques) ;
- probabilité de réponse sachant que l'individu a été envoyé en enquête ;
- coefficient de calage sur marge.

La correction du taux de couverture vise à palier les défauts d'exhaustivité de la base de sondage, qui résultent de la non réponse de certains établissements à la phase de collecte des fichiers d'inscrits. Le taux de couverture est calculé dans la mesure du possible par rapport aux effectifs attendus et non à partir du seul taux de réponse des établissements. Pour cela, diverses sources externes ont été mobilisées (cf. infra). Les coefficients de couverture individuels sont calculés une seule fois et utilisés pour l'ensemble des pondérations.

La probabilité d'appartenance à l'échantillon est égale, pour les individus de l'échantillon utilisé, à leur probabilité d'appartenir à cet échantillon. Elle se calcule à partir de la probabilité individuelle d'appartenance à l'échantillon principal, à l'échantillon réserve comprise et à l'information sur l'utilisation de l'échantillon de réserve. Deux probabilités distinctes ont été calculées : une pour l'échantillon global et une pour le sous-échantillon des « 5 000 ».

La probabilité d'envoi en enquête a été calculée pour l'échantillon global. Les coefficients obtenus ont été utilisés pour la pondération du sous-échantillon des « 5 000 ». La probabilité de réponse pour les individus envoyés en enquête a été calculée de manière indépendante pour la pondération globale et celle du sous-échantillon des « 5 000 ».

Un coefficient de calage sur marge différent a été calculé pour chaque pondération.

³ Ce paragraphe ne concerne pas la pondération de l'extension DGAFP portant sur les élèves de la fonction publique. Cette pondération spécifique fera l'objet d'explications infra.

3. ESTIMATION DE LA COUVERTURE

L'estimation de la couverture de la base de sondage, qui renvoie au degré d'exhaustivité de la collecte des fichiers d'établissement, a été effectuée en s'appuyant sur diverses sources administratives. Il s'agit d'une étape préalable indispensable pour calculer le poids des individus répondants.

Sauf mention contraire, les taux de couverture présentés portent sur les effectifs couverts et non sur le nombre d'établissements couverts.

Le coefficient 1 ($0 < \text{coeffl} \leq 1$) de redressement de la couverture attribué à chaque individu présent dans la base de sondage rend compte de la probabilité de l'individu d'appartenir à la base de sondage, en fonction de ses caractéristiques (niveau et spécialité de sortie, région et type d'établissement de formation).

Le poids de redressement de la couverture attribué à chaque individu est $p1 = 1 / \text{coeffl}$.

Les coefficients de redressement de la couverture ont été calculés à partir des informations présentées ci-dessous.

3.1. Les établissements bénéficiant d'une couverture exhaustive

Lorsque les établissements ont transmis l'intégralité de leur fichiers de sortants, le coefficient de redressement de la couverture est égal à 1 ($\text{coeffl}=1$; $p1 = 1$).

Tous les rectorats ($\text{type}=\text{N}$) ont transmis leurs bases d'élèves donc sur ce champ, on a considéré que la couverture de la base de sondage était exhaustive.

En l'absence de référence complémentaire fiable, on a considéré que sur le champ des docteurs, la base de sondage était exhaustive. Le coefficient de redressement des Conventions Industrielles de Formation par la Recherche (CIFRE) ($\text{type}=\text{X}$) et des allocataires thèse ($\text{type}=\text{Y}$) est donc fixé à 1.

Le champ des formations de la Jeunesse et du Sport ($\text{type}=\text{F}$) a été défini à partir d'une liste de formations transmise par le ministère en charge des sports (à partir d'informations issues des directions régionales de la jeunesse et des sports). Toutes les formations figurant sur la liste transmise par le ministère demandeur de l'extension ayant été collectées, le champ des diplômés des formations du ministère des Sports a été considéré comme exhaustif.

Tableau 3

Nombre d'individus présents dans la base de sondage sur les champs exhaustifs

Type d'établissement	Libellé	Effectifs
F	Directions régionales jeunesse et sports	13 238
N	Rectorats	604 844
X	Thèses CIFRE	964
Y	Allocataires thèses	4 802
TOTAL		623 848

3.2. Les établissements dont la couverture a été estimée à partir d'une référence

Le champ des Universités (type=C)

Chacune des 79 universités dispensant de la formation initiale en France métropolitaine a transmis au Céreq ses listes d'inscrits pour constituer la base de sondage.

Les effectifs de la base de sondage considérés comme sortants après le repérage des doublons sur le champ université ont été comparés à ceux fournis par le système d'information de suivi des étudiants (Sise) de la Direction de l'évaluation, de la prospective et de la performance (Depp).

La base de sondage contient 328 130 individus issus du champ des Universités.

Au niveau national, la couverture du champ universitaire est plutôt bonne, avec un taux de couverture de 90 %. La couverture varie selon les niveaux et les disciplines de formation. Par exemple, la couverture est de 73% pour les écoles d'ingénieurs universitaires contre 94% pour les 2nde année licence – IUT.

Tableau 4

Estimation de la couverture par niveau

Niveau	Effectif collecté par le Céreq	Effectif présent dans SISE	Taux de couverture estimé
1 ^{ère} année licence - IUT	6 987	7 908	88 %
1 ^{ère} année licence universitaire	97 930	105 889	92 %
2 nd e année licence - IUT	18 850	19 997	94 %
2 nd e année licence universitaire	28 352	31 347	90 %
3 ^{ème} année licence professionnelle	23 271	26 386	88 %
3 ^{ème} année licence universitaire	53 702	59 571	90 %
1 ^{ère} année de Master	45 547	51 424	89 %
2 nd e année de Master	66 164	72 398	91 %
Formations diverses du supérieur	9 532	11 874	80 %
Doctorats	11 621	14 004	83 %
Ecoles d'ingénieurs universitaires	5 369	7 357	73 %
Total	367 325	408 155	90 %

Tableau 5

Estimation de la couverture par discipline

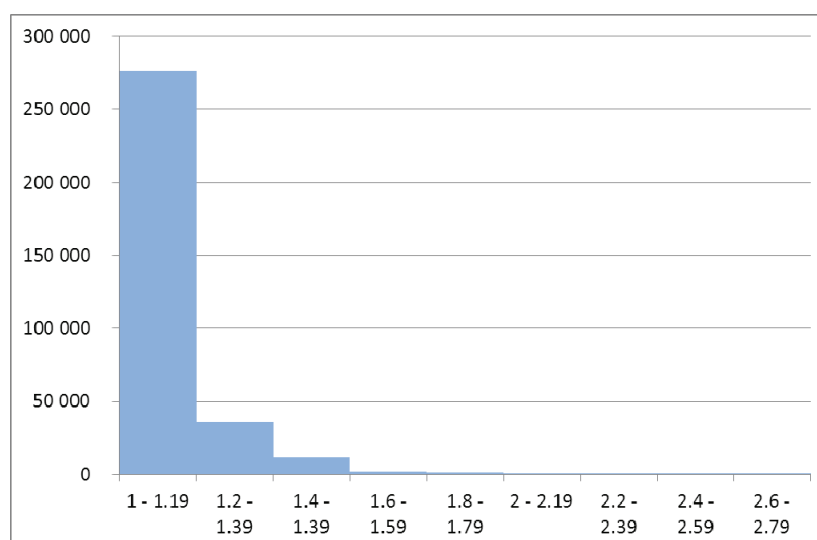
Discipline	Effectif collecté par le Céreq	Effectif présent dans SISE	Taux de couverture estimé
Secteur secondaire	23 908	26 316	91 %
Secteur tertiaire	25 200	27 975	90 %
Ecoles d'ingénieurs universitaires	5 369	7 357	73 %
Formations de la santé	9 532	11 874	80 %
Droit, sciences politiques	41 547	45 980	90 %
Economie, gestion	51 597	57 822	89 %
Lettres et arts	62 687	67 839	92 %
Sciences humaines et sociales	65 660	71 160	92 %
Médecine	14 876	16 786	89 %
Sciences fondamentales appliquées	66 949	75 046	89 %
Total	367 325	408 155	90 %

La couverture a été calculée sur le croisement du niveau, de la discipline et de la région. Pour les écoles d'ingénieurs universitaires, la couverture a été calculée au niveau global. Dans les régions Champagne Ardennes, Poitou Charente et Languedoc Roussillon, les disciplines ont été regroupées pour le calcul de la couverture. Dans les autres cas, la couverture a été estimée pour chaque région sur un croisement de niveau et de discipline.

Le coefficient p1 de redressement de la couverture est finalement compris entre 1 et 2,7 pour les individus issus du champ universitaire (moyenne=1,11, écart-type=0,15).

Figure 4

Distribution du coefficient de redressement de la couverture sur le champ des Universités



3.2.1. Le champ des Centres de Formation par Apprentissage (CFA) (type=P)

Un total de 988 centres de formation des apprentis a été sollicité par le Céreq. Parmi eux, 777 (79 %) ont renvoyé les fichiers demandés.

La base de sondage contient 144 206 individus issus du champ des CFA.

La couverture a été estimée à partir de l'enquête annuelle Système d'Information sur la Formation des Apprentis (Sifa) pilotée par la Depp du ministère de l'Éducation nationale sur les apprentis formés en CFA⁴. Pour les précédentes enquêtes, la couverture était calculée à partir de l'enquête 51 qui portait sur l'ensemble des apprentis mais cette enquête a été supprimée.

Afin de limiter le nombre d'individus hors champ dans la base de sondage, le Céreq a collecté auprès des CFA la liste des inscrits durant l'année scolaire 2007-2008, afin de les coupler avec les inscrits de 2006-2007. D'autre part, le Céreq s'est procuré auprès de la Depp les effectifs d'inscrits en CFA durant l'année 2007-2008. La couverture a été calculée en rapportant les effectifs bruts d'inscrits en 2007-2008 collectés par le Céreq⁵ à ceux issus de Sifa. Le coefficient de couverture moyen est de 67 % pour le champ des CFA.

⁴ L'enquête fournit, sur différents échelons géographiques, des statistiques sur les effectifs par sexe, par âge, par niveau et spécialité de formation, par secteur d'activité de l'employeur (code APE). Les données sont collectées de façon agrégée.

⁵ Avant dédoublement et apurements de la base de sondage.

Tableau 6

Estimation de la couverture des CFA par Région

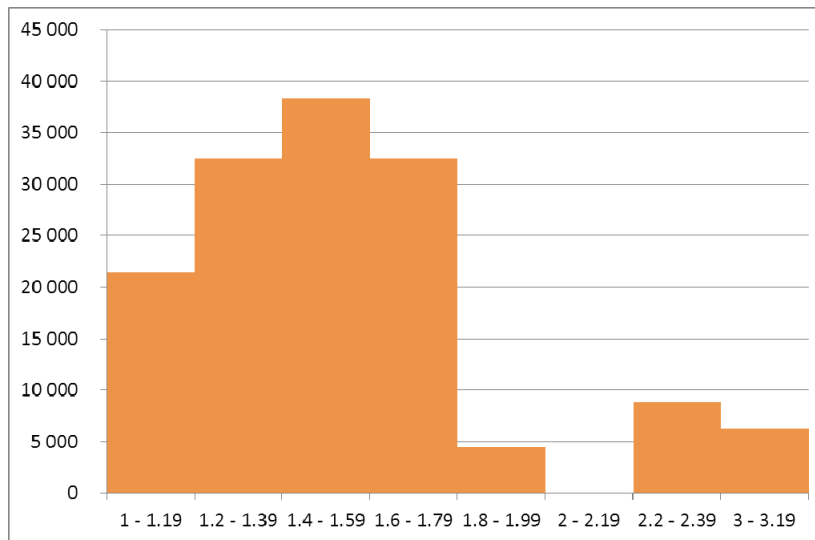
Région de l'établissement de formation	Effectif d'inscrits 2007-2008		Taux couverture estimé
	Collectés par le Céreq	Effectif présent dans SIFA	
ILE DE FRANCE	41 456	70 945	58 %
CHAMPAGNE-ARDENNES	7 019	8 361	84 %
PICARDIE	10 252	12 189	84 %
HAUTE-NORMANDIE	4 707	14 568	32 %
CENTRE	13 780	18 102	76 %
BASSE-NORMANDIE	7 592	10 565	72 %
BOURGOGNE	6 689	12 523	53 %
NORD-PAS DE CALAIS	13 110	18 774	70 %
LORRAINE	13 718	16 458	83 %
ALSACE	12 833	14 035	91 %
FRANCHE-COMTE	7 441	9 543	78 %
PAYS DE LA LOIRE	18 961	32 120	59 %
BRETAGNE	8 055	17 783	45 %
POITOU-CHARENTES	10 976	13 294	83 %
AQUITAINE	13 312	17 688	75 %
MIDI-PYRENEES	13 473	16 544	81 %
LIMOUSIN	3 244	3 789	86 %
RHONE-ALPES	30 412	42 936	71 %
AUVERGNE	5 901	8 341	71 %
LANGUEDOC-ROUSSILLON	6 089	13 779	44 %
PACA ET CORSE	24 120	35 472	68 %
TOTAL	273 140	407 809	67 %

L'estimation de la couverture s'est effectuée au niveau régional. Aucun individu en CFA n'est retrouvé en Corse dans la base de sondage, la Corse a donc été regroupée avec la région PACA pour le calcul du taux de couverture.

Le coefficient de redressement de la couverture est finalement compris entre 1,09 et 3,09 pour les individus issus du champ CFA (moyenne=1,56, écart-type=0,43).

Figure 5

Distribution du coefficient de redressement de la couverture sur le champ des CFA



3.2.2. Le champ des formations du domaine de la Santé (type=E) et du Social (type=D)

Le champ a été défini à partir d'une liste de formations transmise par la Direction de la recherche, des études, de l'évaluation et des statistiques (Drees). Au total, on comptabilisait 25 378 diplômés des écoles de la santé et du social en 2007 dans la base de sondage. Selon le fichier national des établissements sanitaires et sociaux (Finess)⁶, ils étaient 63 883, soit une couverture moyenne de 40 % pour la base de sondage en termes d'effectifs (contre 43 % pour Génération 2004). Cette faible couverture résulte essentiellement de la non-réponse totale ou partielle des établissements concernés.

Tableau 7

Estimation de la couverture des diplômés de la santé et du social par région

Région de l'établissement de formation	Effectif collecté Céreq	Effectif présent dans le fichier Finess	Taux de couverture estimé
ILE DE FRANCE	3 742	11 177	33 %
CHAMPAGNE-ARDENNES	639	1 326	48 %
PICARDIE	1 176	1 780	66 %
HAUTE-NORMANDIE	1 077	1 727	62 %
CENTRE	1 413	2 812	50 %
BASSE-NORMANDIE	1 109	1 497	74 %
BOURGOGNE	870	1 616	54 %
NORD-PAS DE CALAIS	2 362	4 432	53 %
LORRAINE	871	2 758	32 %
ALSACE	985	2 164	46 %
FRANCHE-COMTE	373	1 220	31 %
PAYS DE LA LOIRE	1 403	3 980	35 %
BRETAGNE	1 343	3 014	45 %
POITOU-CHARENTES	509	1 420	36 %
AQUITAINE	1 251	3 277	38 %
MIDI-PYRENEES	367	2 421	15 %
LIMOUSIN	705	1 214	58 %
RHONE-ALPES	1 774	6 289	28 %
AUVERGNE	378	1 508	25 %
LANGUEDOC-ROUSSILLON	1 185	2 706	44 %
PACA	1 772	5 358	33 %
CORSE	74	187	40 %
TOTAL	25 378	63 883	40 %

⁶ Le fichier Finess est un fichier géré par le ministère en charge de la santé et des affaires sociales pour répondre d'abord à des besoins de gestion, puisqu'il permet l'immatriculation par les Agences régionales de la santé (ex DDASS et DRASS) des établissements et entités juridiques porteurs d'une autorisation ou d'un agrément dans les trois domaines : le sanitaire et médico-social, le social, l'enseignement des personnels en santé et en social.

Tableau 8

Estimation de la couverture des diplômés de la santé et du social par diplôme

Diplôme	Effectif collecté Céreq	Effectif FINESS	Taux de couverture estimé
Formations de la santé, type=E			
Infirmier	10533	21813	48 %
Masseur-kinésithérapeute	758	1603	47 %
Pédicure-podologue	165	378	44 %
Manipulateur d'électro-radiologie	98	552	18 %
Ergothérapeute	61	331	18 %
Psychomotricien	89	239	37 %
Sage-femme	132	842	16 %
Puéricultrice	95	988	10 %
Aide-soignant	6547	22847	29 %
Auxiliaire de puériculture	637	3118	20 %
Formations du social, type=D			
Conseiller en économie sociale et familiale	577	958	60 %
Assistant de service social	1158	2381	49 %
Educateur de jeunes enfants	672	1202	56 %
Educateur spécialisé	2762	4167	66 %
Moniteur éducateur	1094	2464	44 %
Total	25378	63883	40 %

Pour le diplôme d'infirmier, la couverture est calculée au niveau régional.

Les aides-soignants et les auxiliaires de puériculture (diplômes de la santé de niveau V) sont regroupés pour le calcul de la couverture au niveau régional.

Pour le diplôme de masseur-kinésithérapeute, la couverture est calculée au niveau national.

Les autres diplômes de la santé de niveau bac+2 (pédicure podologue, manipulateur d'électro-radiologie, ergothérapeute, psychomotricien, sage-femme, puéricultrice) présentent des effectifs faibles et sont donc regroupés pour le calcul d'une couverture nationale.

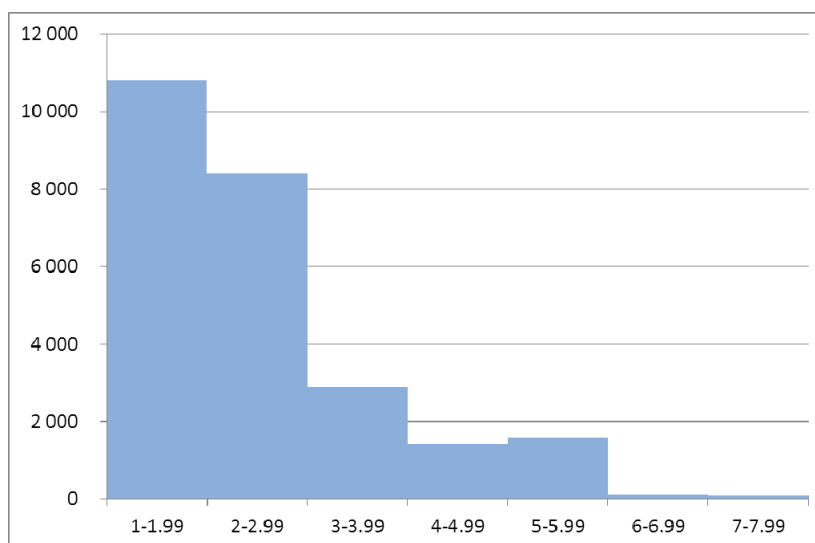
La couverture est calculée au niveau national pour chaque diplôme du social (conseiller en économie sociale et familiale, assistant de service social, éducateur de jeunes enfants, éducateur spécialisé, moniteur éducateur).

La couverture est ramenée à 1 pour les infirmiers de Haute-Normandie qui présentent un taux de couverture de 1,1.

Le coefficient de redressement de la couverture est finalement compris entre 1 et 7,69 pour les individus issus du champ des formations de la santé et du social (moyenne=2,49, écart-type=1,17).

Figure 6

Distribution du coefficient de redressement de la couverture sur le champ des formations de la santé et du social



3.2.3. Le champ des Instituts Universitaires de Formation des Maîtres (IUFM) (type=H)

Un total de 26 IUFM ont été sollicités par le Céreq. Parmi eux, 19 (73 %) ont renvoyé les informations demandées. La base de sondage contient finalement 18 403 individus sur le champ des IUFM.

En parallèle, La Direction générale pour l'enseignement supérieur et l'insertion professionnelle (DGESIP) a transmis au Céreq un fichier national des inscrits en IUFM (équivalent de Sise) en 2006-2007 et 2007-2008.

Le champ défini pour la comparaison est constitué des inscrits en première année en 2006-2007 :

- absents d'IUFM en 2007-2008,
- ou inscrits en deuxième année d'IUFM en 2007-2008.

On a comparé sur ce champ les effectifs issus des fichiers bruts collectés par le Céreq et ceux du fichier national transmis par la DGESIP.

En termes d'effectifs, on dénombrait 23 070 sortants d'IUFM dans la base brute collectée par le Céreq. Selon les informations transmises par la DEPP/DGESIP, le taux de couverture atteint environ 64 % avec des différences assez fortes selon les régions.

Tableau 9

Estimation de la couverture des IUFM par région

Région de l'établissement de formation	Effectif collecté Céreq	Effectif issu de la source Dgesip	Taux de couverture estimé
ILE DE FRANCE	5 807	7 218	80 %
CHAMPAGNE-ARDENNES	695	713	97 %
PICARDIE	932	932	100 %
HAUTE-NORMANDIE	0	896	0 %
CENTRE	1 224	1 227	100 %
BASSE-NORMANDIE	0	767	0 %
BOURGOGNE	753	983	77 %
NORD-PAS DE CALAIS	6	3 033	0 %
LORRAINE	1063	1 309	81 %
ALSACE	958	960	100 %
FRANCHE-COMTE	0	675	0 %
PAYS DE LA LOIRE	1 441	1 751	82 %
BRETAGNE	1 495	1 536	97 %
POITOU-CHARENTES	767	810	95 %
AQUITAINE	1 239	1 501	83 %
MIDI-PYRENEES	1463	1 466	100 %
LIMOUSIN	323	332	97 %
RHONE-ALPES	1 869	3 618	52 %
AUVERGNE	3	697	0 %
LANGUEDOC-ROUSSILLON	1 324	1 340	99 %
PACA	1 536	2 804	55 %
CORSE	172	178	97 %
TOTAL	23 070	34 746	66 %

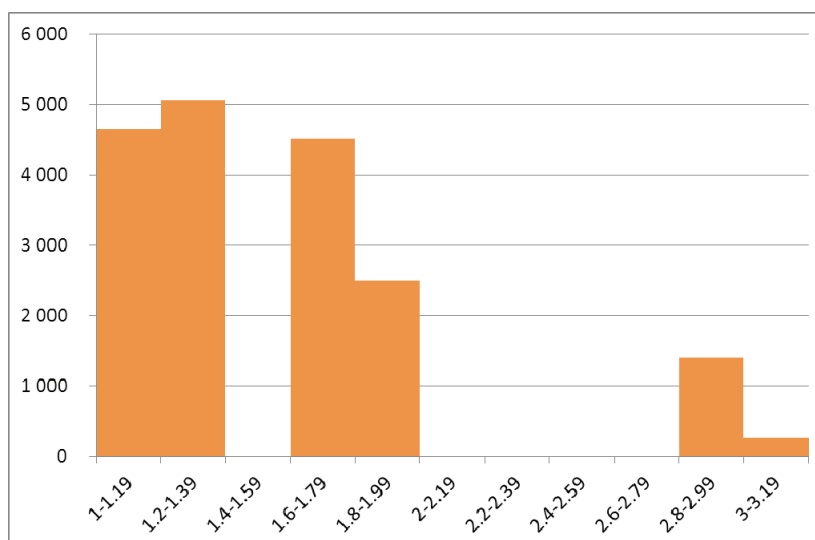
Le coefficient de couverture a été calculé au niveau régional, en effectuant cependant les regroupements de régions suivants :

- Auvergne et Limousin ;
- Nord Pas de Calais, Champagne Ardennes et Picardie ;
- Haute-Normandie, Basse-Normandie, Pays de Loire et Bretagne ;
- Franche Comté, Bourgogne et Lorraine.

Le coefficient de redressement de la couverture est finalement compris entre 1 et 3,16 pour les individus issus du champ des IUFM (moyenne=1,52, écart-type=0,54).

Figure 7

Distribution du coefficient de redressement de la couverture sur le champ des IUFM



3.2.4. Le champ des écoles de commerce (type=B) et ingénieurs (type=A)

Pour Génération 2007, le Céreq a sollicité 280 écoles de commerce et 201 écoles d'ingénieurs. Un total de 125 écoles de commerce (45 %) et de 126 écoles d'ingénieurs (63 %) a répondu favorablement à cette demande. Pour chaque établissement, le Céreq demandait les informations pour les diplômés sortants. La base de sondage contient 26 947 individus sur ce champ.

La référence pour le calcul du taux de couverture est issue de la publication "Repères et références statistiques 2011" (RERS 2011 - p. 261,262) de la Depp. Les chiffres portent sur le nombre de diplômés d'écoles de commerce et ingénieurs en France métropolitaine + DOM.

Nous avons contacté la Depp pour avoir pour avoir la distinction entre France métropolitaine et DOM. En 2007, un total de 24 369 individus ont été diplômés d'écoles de commerce en France métropolitaine (28 diplômés dans les DOM) et 27 488 ont été diplômés d'écoles d'ingénieurs en France métropolitaine

Le taux de couverture estimé varie donc selon le type d'école dans les proportions suivantes :

- écoles de commerce : 13 540 sortants collectés sur 24 369 estimés soit une couverture de 55,6 % ;
- écoles d'ingénieurs : 13 407 sortants collectés sur 27 488 estimés soit une couverture de 48,8 %.

Le coefficient de redressement de la couverture est finalement égal à 1,8 pour les individus du champ des écoles de commerce et 2,1 pour les individus du champ des écoles d'ingénieurs.

3.2.5. Le champ des formations agricoles (type=Q)

Sur 866 établissements d'enseignement secondaire agricole sollicités (type=Q), un total de 670 (77 %) a répondu à la collecte. La base de sondage contenait finalement 42 066 individus issus du champ des formations agricoles.

Pour le calcul de la couverture, la référence utilisée est interne au Céreq. Il s'agit de la base des élèves inscrits à la rentrée 2006 dans les établissements sous tutelle du Ministère de l'agriculture (destinée à

construire la base Reflet). Nous nommerons cette base RefletCer. On y trouve essentiellement des élèves préparant des diplômes du Ministère de l'agriculture mais aussi des élèves préparant des diplômes du Ministère de l'Éducation nationale.

Parmi les 866 établissements sollicités par le Céreq, 776 figurent dans Reflet. C'est sur cette base commune que la couverture est calculée.

Les informations présentes dans les deux bases ne sont pas à champ comparable puisque :

- dans Reflet, on collecte les élèves inscrits à la rentrée 2006
- pour la base de sondage, on collecte les sortants de formation à l'issue de l'année scolaire 2006/2007.

On part du principe que lorsque les établissements ont envoyé leurs fichiers, ces derniers étaient complets. Sur les 563 établissements ayant renvoyé leurs informations au Céreq et figurant dans la base Reflet, on a calculé le rapport entre les effectifs collectés auprès des établissements et ceux présents dans RefletCer. On constate qu'en moyenne, les effectifs de sortants représentent 78 % des effectifs d'inscrits. On a fait l'hypothèse que ce taux était stable et on a estimé que parmi les individus présents dans RefletCer, 78 % devraient être présents dans les bases transmises au Céreq par les établissements de formation.

Tableau 10

Estimation de la couverture des établissements de l'enseignement agricole par région (établissements sollicités par le Céreq et présents dans RefletCer)

REGION	Effectif collecté	Effectif RefletCer	Effectif estimé	Taux de couverture
ALSACE	405	905	713	57 %
AQUITAINE	2 478	5 684	4 476	55 %
AUVERGNE	1 419	2 073	1 633	87 %
BOURGOGNE	1 903	3 177	2 502	76 %
BRETAGNE	4 481	8 110	6 387	70 %
CENTRE	1 549	2 757	2 171	71 %
CHAMPAGNE-ARDENNE	1 127	2 434	1 917	59 %
CORSE	0	65	51	0 %
FRANCHE-COMTE	1 823	3 023	2 381	77 %
ILE DE France	1 194	1 868	1 471	81 %
LANGUEDOC-ROUSSILLON	2 758	3 444	2 712	102 %
LIMOUSIN	906	1 783	1 404	65 %
LORRAINE	838	1 466	1 155	73 %
MIDI-PYRENEES	2 713	4 615	3 634	75 %
NORD-PAS DE CALAIS	1 452	3 665	2 886	50 %
NORMANDIE BASSE	2 280	3 625	2 855	80 %
NORMANDIE HAUTE	983	1 594	1 255	78 %
PAYS DE LA LOIRE	5 854	8 236	6 486	90 %
PICARDIE	1 450	2 375	1 870	78 %
POITOU-CHARENTES	2 590	3 815	3 004	86 %
PROVENCE ALPES COTE D'AZUR	2 009	3 640	2 867	70 %
RHONE-ALPES	5 283	8 772	6 908	76 %
TOTAL	32 133	51 921	40 889	79 %

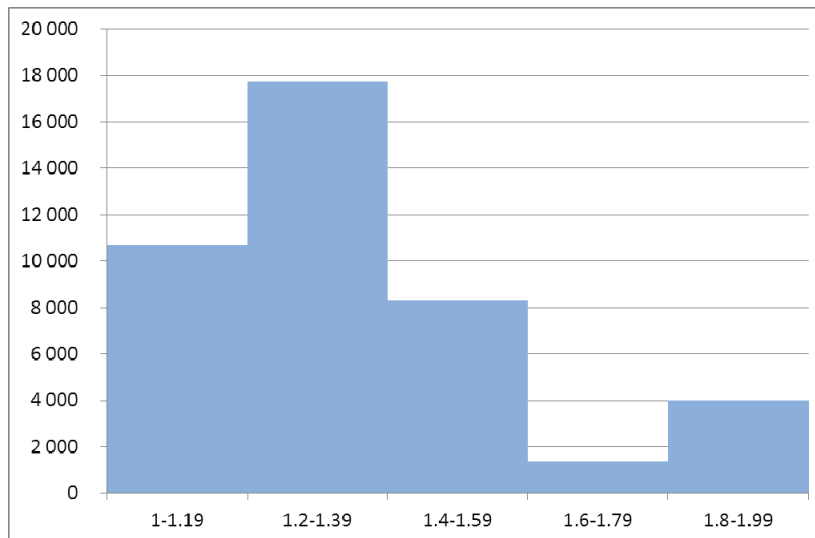
Lecture : En Alsace, sur le champ des établissements présents dans Reflet et sollicités par le Céreq, 405 individus présumés sortants en 2006-2007 ont été collectés par le Céreq. Dans Reflet, on comptait 905 individus inscrits à la rentrée en septembre 2006, soit 713 ($=905 \times 0,78$) présumés sortants. Le taux de couverture en Alsace est donc estimé à 57% ($=405/713$).

Les taux de couverture présentés dans le tableau précédent sont appliqués à l'ensemble des individus issus d'établissements de l'enseignement agricole, selon leur Région. Pour le calcul de la couverture, la Corse et Paca sont regroupées. Le taux de couverture de Languedoc Roussillon est ramené à 1.

Le coefficient de redressement de la couverture est finalement compris entre 1 et 1,99 pour les individus issus du champ des formations agricoles (moyenne=1,35 ; écart-type=0,23).

Figure 8

Distribution du coefficient de redressement de la couverture sur le champ des formations agricoles



3.3. Les établissements dont la couverture a été estimée sans information auxiliaire

Pour l'ensemble des établissements suivants, nous n'avons trouvé aucune référence statistique fiable pour calculer la couverture. Au total, 25 236 individus présents dans la base de sondage sont concernés. On a donc appliqué tout simplement comme taux de couverture le taux de réponse global à la collecte.

Tableau 11

Taux de couverture établissement des autres établissements

Type d'établissement	Libellé	Nombre d'établissements répondants	Nombre d'établissements sollicités	Taux de couverture estimé
J	Ecoles dans le secteur services	47	132	36 %
K	Ecoles dans le secteur industriel	12	22	55 %
L	Ecoles de formation agricole	14	15	93 %
M	Centres ou facultés privés	6	12	50 %
R	Centres de formation pédagogiques privés	18	32	56 %
S	Ecoles d'administration publique	5	6	83 %
T	Ecoles Normales Supérieures	4	5	80 %
U	Ecoles de la DGA	1	3	33 %
V	Instituts d'études politiques	8	11	73 %
W	Ecoles d'architecture et artistiques	158	332	48 %
Z	Nouvelles écoles collectées pour la DGAFP	11	27	41 %
TOTAL		284	597	48%

Vu le faible nombre d'établissements concernés, la couverture a été calculée tous types d'établissements confondus. Ainsi, le coefficient de redressement de la couverture est égal à 2,1 pour tout individu appartenant au champ des types d'établissements précédents.

3.4. Récapitulatif

Au global, le taux de couverture de la base de sondage est estimé à 84 %. L'estimation de la taille de la base de sondage exhaustive est de 1 465 000 individus.

Tableau 12

Coefficients de redressement moyens par type d'établissement

Type d'établissement	Nombre d'individus dans la base de sondage	Coefficient de redressement moyen	Estimation exhaustive
A - Ecoles ingénieurs	13 407	2,05	27 484
B - Ecoles de commerce	13 540	1,8	24 372
C - Universités	328 130	1,11	364 224
D - Ecoles professions sociales	6 188	1,81	11 200
E - Ecoles professions de la santé	19 096	2,72	51 941
F - DRJS	13 238	1	13 238
H - IUFM	18 403	1,52	27 973
J - Ecoles secteur service	2 996	2,1	6 292
K - Ecoles secteur industriel	1 251	2,1	2 627
L - Ecoles formations agricoles	1 023	2,1	2 148
M - Facultés privées	2 595	2,1	5 450
N - Lycées et collèges MEN	604 844	1	604 844
P - Centre de Formation des Apprentis	144 206	1,56	224 961
Q - Lycées agricoles	42 066	1,35	56 789
R - Centres privés d'enseignement	947	2,1	1 989
S - Ecoles administrations publiques	609	2,1	1 279
T - Ecoles Normales Supérieures	816	2,1	1 714
U - Ecoles de la DGA	144	2,1	302
V - IEP	2 255	2,1	4 736
W - Ecoles ministère de la culture	9 918	2,1	20 828
X - CIFRE	964	1	964
Y - ALLOCATAIRES THESES	4 802	1	4 802
Z - DGAFP	2 682	2,1	5 632
TOTAL	1 234 120	1,2	1 465 789

Tableau 13

Coefficients de redressement moyens en fonction de la strate de formation

Strate de formation (variable strate07)	Nombre d'individus dans la base de sondage	Coefficient de redressement moyen	Estimation exhaustive
Collège	150 621	1	150 621
CAP/BEP (ttes années) scolaire industriel	56 755	1,05	59 593
CAP/BEP (ttes années) scolaire tertiaire	59 189	1,05	62 148
CAP/BEP (ttes années) apprentissage industriel	58 452	1,55	90 601
CAP/BEP (ttes années) apprentissage tertiaire	26 549	1,53	40 620
Lycée (2nde à BAC) apprentissage	34 264	1,55	53 109
Lycée (2nde à BAC) pro/techno industriel	72 662	1,05	76 295
Lycée (2nde à BAC) pro/techno tertiaire	116 147	1,01	117 308
Lycée (2nde à BAC) généraux	100 160	1	100 160
BAC+1 à BAC+5 apprentissage	27 818	1,6	44 509
BAC+1 BTS-DUT	23 120	1	23 120
BAC+1 Université	82 586	1,08	89 193
BAC+2 BTS	65 676	1,04	68 303
BAC+2 DUT	23 596	1,08	25 484
BAC+2 Université	27 176	1,1	29 894
DRJS	13 238	1	13 238
SOCIAL	6 971	1,69	11 781
SANTE	19 115	2,72	51 993
IUFM	18 403	1,52	27 973
Licences Professionnelles hors licences IUP	23 214	1,13	26 232
Licences Université	43 898	1,12	49 166
M1-MST-MSG-Maitrise IUP	40 240	1,14	45 874
Ecoles diverses tous niveaux	24 941	2,1	52 376
Ecoles de commerce tous niveaux	12 638	1,8	22 748
Ecoles d'ingénieurs	13 680	2,03	27 770
DESS + master pro	45 389	1,1	49 928
DEA + master recherche + divers degrés 6 hors docteurs	32 915	1,16	38 181
Doctorat	14 707	1,13	16 619

4. PONDÉRATION DES REpondANTS DANS LE CHAMP DU CEReq

4.1. La reconstitution des poids de tirage

4.1.1. La méthodologie

Pour Génération 2007, l'échantillon a été constitué de manière à satisfaire aux exigences du Céreq et des différentes extensions. On a affecté une probabilité d'inclusion dans l'échantillon (probabilité de tirage) à chaque individu présent dans la base de sondage, en fonction de ses caractéristiques et du nombre de questionnaires visé. La constitution de l'échantillon repose sur des hypothèses relatives aux taux de réponse des individus. Au regard de l'incertitude existant sur ces taux de réponse, le Céreq a constitué un échantillon de réserve. Ce dernier devait être utilisé en cours de plateau d'enquête uniquement si les taux de réponse constatés se révélaient inférieurs aux taux de réponse anticipés sur une cible donnée.

Le tirage de l'échantillon pour Génération 2007 s'est déroulé en deux étapes :

- tirage de l'échantillon incluant la réserve (ECHRES) ;
- tirage de l'échantillon principal (ECHPRINC) à l'intérieur de l'échantillon incluant la réserve.

A l'issue de ces deux étapes, chaque individu de l'échantillon incluant la réserve présentait un poids individuel de tirage (PDSECHRES), inverse de sa probabilité d'appartenance à l'échantillon.

Les individus de l'échantillon principal présentaient également un poids individuel de tirage spécifique à leur appartenance à l'échantillon principal (PDSPRINC). Pour les individus de l'échantillon principal, $1 \leq \text{PDSECHRES} \leq \text{PDSPRINC}$.

Lorsque sur un champ particulier, la réserve n'était pas sollicitée en cours d'enquête, l'échantillon utilisé correspondait à l'échantillon principal. Dans ce cas, le poids de chaque individu de l'échantillon utilisé était égal à PDSPRINC.

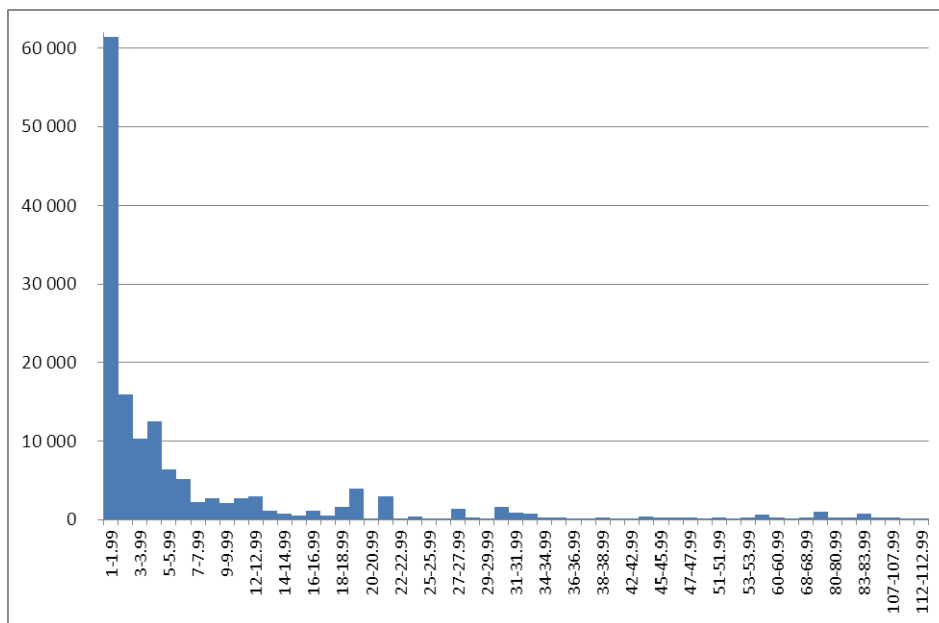
Étant donné que l'utilisation de la réserve a été partielle pour Génération 2007, l'échantillon utilisé correspondait à l'échantillon principal et à une partie de la réserve. Les poids de tirage « finaux » des individus de l'échantillon utilisé appartenant à des strates où la réserve a été sollicitée ont donc été recalculés.

4.1.2. Les poids de tirage de l'échantillon principal

L'échantillon principal comprenait 149 490 individus. Leur poids de tirage était compris entre 1 (individu sélectionné de manière certaine) et 112 (individu ayant le minimum de chances d'appartenir à l'échantillon principal). En moyenne, PDSPRINC= 8,33 (ET=14,9).

Figure 9

Distribution des poids de tirage de l'échantillon principal



En raison des nombreuses extensions d'échantillon et du faible nombre de questionnaires financés par le Céreq, les poids de tirage de l'échantillon principal présentent une dispersion très forte.

Les poids de tirage les plus élevés se trouvent sur les strates de l'enseignement secondaire (collèges, CAP/BEP, lycées) pour lesquelles il n'y avait pas d'extension spécifique.

Tableau 14

Poids de tirage dans l'échantillon principal selon la strate de formation

Strate de formation (strate07)	Nombre d'individus dans l'échantillon principal	PDSPRIN moyen	Ecart-type de la strate	Somme des PDSPRIN
Collège	8 584	18,01	23,35	154 598
CAP/BEP (ttes années) scolaire industriel	4 436	12,97	18,7	57 535
CAP/BEP (ttes années) scolaire tertiaire	5 197	11,83	17,88	61 481
CAP/BEP (ttes années) apprentissage industriel	4 989	12,48	18,74	62 263
CAP/BEP (ttes années) apprentissage tertiaire	2 187	11,88	16,56	25 982
Lycée (2nde à BAC) apprentissage	3 328	10,72	19,02	35 676
Lycée (2nde à BAC) pro/techno industriel	6 194	11,99	21,4	74 266
Lycée (2nde à BAC) pro/techno tertiaire	9 008	12,81	21,46	115 392
Lycée (2nde à BAC) généraux	6 386	15,18	22,44	96 939
BAC+1 à BAC+5 apprentissage	3 020	9,42	16,88	28 448
BAC+1 BTS-DUT	2 053	11,1	10,05	22 788
BAC+1 Université	8 361	9,67	7,2	80 851
BAC+2 BTS	5 532	11,82	18,87	65 388
BAC+2 DUT	3 253	7,65	5,47	24 885
BAC+2 Université	2 720	10,15	7,13	27 608
DRJS	12 593	1,06	1,6	13 349
SOCIAL	6 971	1	0	6 971
SANTE	16 018	1,19	0,15	19 061
IUFM	2 219	8,29	4,73	18 396
Licences Professionnelles hors licences IUP	1 827	12,97	16,19	23 696
Licences Université	4 899	9,01	8,18	44 140
M1-MST-MSG-Maitrise IUP	4 848	8,4	7,39	40 723
Ecoles diverses tous niveaux	4 874	5,13	18,42	25 004
Ecoles de commerce tous niveaux	1 649	7,75	6,24	12 780
Ecoles d'ingénieurs	2 345	5,8	4,73	13 601
DESS + master pro	5 306	8,7	10,04	46 162
DEA + master recherche + divers degrés 6 hors docteurs	4 022	8,25	8,47	33 182
Doctorat	6 671	2,22	0,6	14 810
Total	149 490	8,33		1 245 975

4.1.3. L'utilisation de l'échantillon de réserve

Un total de 42 478 individus, soit 62 % de la réserve, a finalement été utilisé. Des injections successives de réserve ont été réalisées afin d'atteindre les objectifs en nombre de questionnaires sur les champs suivants :

- Aquitaine, Bretagne, Franche Comté, Nord Pas de Calais, Poitou-Charentes (extensions régionales) ;
- expérimentations de baccalauréat (extension DEPP) ;
- formations environnementales (extension CGDD) ;
- docteurs (extension DGESIP) ;
- sortants du secondaire ;
- échantillon des 5 000.

Lorsque sur un champ donné, l'utilisation de la réserve a été totale, le poids final d'un individu échantillonné (qu'il appartienne à l'échantillon principal ou à la réserve) est égal à son PDSECHRES.

Lorsque sur un champ donné, l'utilisation de la réserve est partielle, les poids de tirage finaux sont égaux à :

$$- PDSECHRES * \frac{|réserve\ dans\ le\ champ|}{|réserve\ utilisée\ dans\ le\ champ|}$$

si l'individu appartient à la réserve utilisée

$$- \frac{1}{\frac{1}{PDS\ PRIN} + \left(\frac{1}{PDSECHRES} - \frac{1}{PDS\ PRIN} \right) \frac{|réserve\ dans\ le\ champ|}{|réserve\ utilisée\ dans\ le\ champ|}}$$

si l'individu appartient à l'échantillon principal.

4.1.4. Les poids de tirage de l'échantillon utilisé (PDS)

L'échantillon utilisé comprenait 191 968 individus. Leur poids de tirage était compris entre 1 (individu sélectionné de manière certaine) et 112 (individu ayant le minimum de chances d'appartenir à l'échantillon principal). En moyenne, PDS= 6,44 (ET=11,4).

Figure 10

Distribution des poids de tirage de l'échantillon utilisé

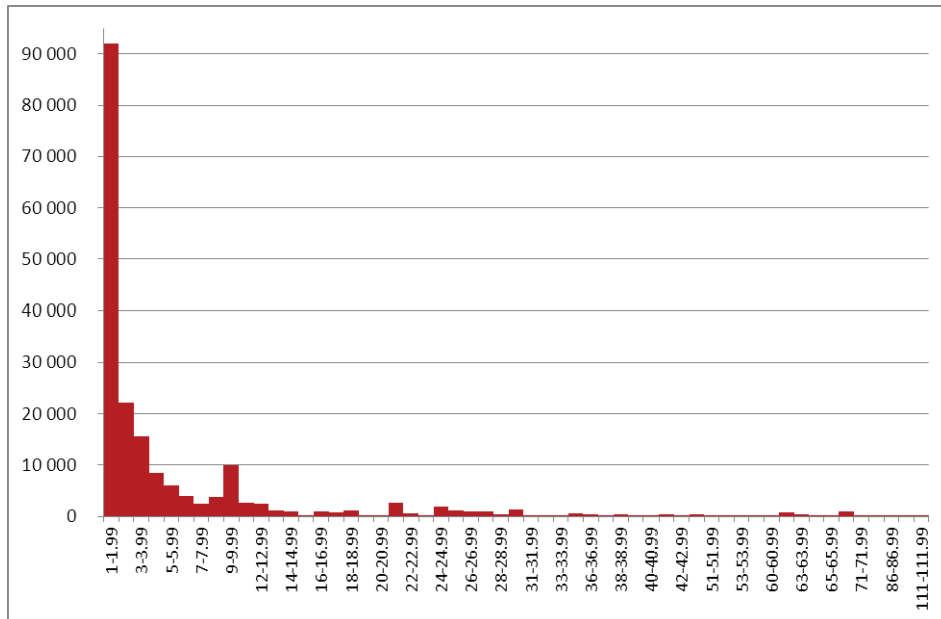


Tableau 15

Poids de tirage dans l'échantillon utilisé selon le type d'établissement

Type d'établissement	Nombre d'individus dans l'échantillon utilisé	PDS moyen	ET sur le type d'établissement	Somme des PDS
A - Ecoles ingénieurs	2 324	5,75	4,75	13 363
B - Ecoles de commerce	2 200	6,17	5,96	13 574
C - Universités	56 847	5,82	6,38	330 850
D - Ecoles professions sociales	6 188	1	0	6 188
E - Ecoles professions de la santé	15 999	1,19	0,15	19 039
F - DRJS	12 593	1,06	1,6	13 349
H - IUFM	2 219	8,29	4,73	18 396
J - Ecoles secteur service	226	13,7	30,5	3 096
K - Ecoles secteur industriel	26	58,29	51,87	1 516
L - Ecoles formations agricoles	446	2,36	10,72	1 053
M - Facultés privées	223	10,2	23,37	2 275
N - Lycées et collèges MEN	53 913	11,22	16,71	604 904
P - Centre de Formation des Apprentis	23 802	6,08	10,5	144 716
Q - Lycées agricoles	8 427	4,99	9,9	42 051
R - Centres privés d'enseignement	49	22,48	37,67	1 102
S - Ecoles administrations publiques	609	1	0	609
T - Ecoles Normales Supérieures	17	50,53	42,48	859
U - Ecoles de la DGA	27	5,55	0,74	150
V - IEP	70	31,49	41,3	2 204
W - Ecoles ministère de la culture	515	18,75	35,53	9 656
X - CIFRE	479	1,41	0,3	675
Y - ALLOCATAIRES THESES	2 089	1,66	0,3	3 468
Z - DGAFP	2 680	1	0	2 680
Total	191 968	6,44	11,42	1 235 770

Tableau 16

Poids de tirage dans l'échantillon utilisé selon la strate de formation

Strate de formation (Strate07)	Nombre d'individus dans l'échantillon utilisé	PDS moyen	Ecart-type de la strate	Somme des PDS
Collège	10 732	14,37	18,69	154 219
CAP/BEP (ttes années) scolaire industriel	6 051	9,34	14,23	56 516
CAP/BEP (ttes années) scolaire tertiaire	6 898	8,88	13,14	61 254
CAP/BEP (ttes années) apprentissage industriel	9 206	6,32	9,57	58 182
CAP/BEP (ttes années) apprentissage tertiaire	4 246	6,2	8,43	26 325
Lycée (2nde à BAC) apprentissage	5 952	5,8	9,91	34 522
Lycée (2nde à BAC) pro/techno industriel	8 272	8,66	15,84	71 636
Lycée (2nde à BAC) pro/techno tertiaire	11 262	10,3	17,03	115 999
Lycée (2nde à BAC) généraux	7 873	12,31	18,02	96 917
BAC+1 à BAC+5 apprentissage	4 954	5,7	13,92	28 238
BAC+1 BTS-DUT	2 529	9	8,11	22 761
BAC+1 Université	16 515	4,93	3,57	81 419
BAC+2 BTS	7 508	8,66	14,36	65 019
BAC+2 DUT	5 863	3,95	2,79	23 159
BAC+2 Université	5 162	5,36	3,46	27 668
DRJS	12 593	1,06	1,6	13 349
SOCIAL	6 971	1	0	6 971
SANTE	16 018	1,19	0,15	19 061
IUFM	2 219	8,29	4,73	18 396
Licences Professionnelles hors licences IUP	1 973	12,1	14,46	23 873
Licences Université	4 899	9,01	8,18	44 140
M1-MST-MSG-Maitrise IUP	5 785	7,04	5,95	40 726
Ecoles diverses tous niveaux	4 874	5,13	18,42	25 004
Ecoles de commerce tous niveaux	2 064	6,19	5,06	12 776
Ecoles d'ingénieurs	2 377	5,72	4,75	13 596
DESS + master pro	6 040	7,61	8,43	45 964
DEA + master recherche + divers degrés 6 hors docteurs	4 161	7,98	8,46	33 205
Doctorat	8 971	1,63	0,31	14 623
TOTAL	191 968	6,44		1 235 517

On souhaite savoir si l'échantillonnage modifie la structure de répartition des poids selon les strates de formation. Pour chaque strate de formation, on commence par comparer le nombre d'individus présents dans la base de sondage et la somme des poids de tirage des individus de l'échantillon utilisé. En théorie, ces grandeurs sont égales. On constate que pour chaque strate de formation, l'écart relatif est inférieur à 3 % en valeur absolue. On compare ensuite l'effectif pondéré redressé de la couverture (le poids attribué à chaque individu de l'échantillon utilisé est égal à $COEFF1 * PDS$) à l'effectif estimé pour une base de sondage

exhaustive. A nouveau, on constate que pour chaque strate de formation, l'écart relatif est inférieur à 3 % en valeur absolue.

Tableau 17

Déformations potentielles induites par l'échantillonnage selon la strate de formation

Strate de formation (strate07)	Somme des PDS	Nombre d'individus présents dans la base de sondage	Ecart relatif	Effectif pondéré redressé couverture	Nombre d'individus estimés pour une base de sondage exhaustive	Ecart relatif
Collège	154 219	150 621	2 %	154 219	150 621	2 %
CAP/BEP (ttes années) scolaire industriel	56 516	56 755	0 %	59 421	59 593	0 %
CAP/BEP (ttes années) scolaire tertiaire	61 254	59 189	3 %	64 220	62 148	3 %
CAP/BEP (ttes années) apprentissage industriel	58 182	58 452	0 %	89 851	90 601	-1 %
CAP/BEP (ttes années) apprentissage tertiaire	26 325	26 549	-1 %	39 997	40 620	-2 %
Lycée (2nde à BAC) apprentissage	34 522	34 264	1 %	54 044	53 109	2 %
Lycée (2nde à BAC) pro/techno industriel	71 636	72 662	-1 %	75 523	76 295	-1 %
Lycée (2nde à BAC) pro/techno tertiaire	115 999	116 147	0 %	117 350	117 308	0 %
Lycée (2nde à BAC) généraux	96 917	100 160	-3 %	97 153	100 160	-3 %
BAC+1 à BAC+5 apprentissage	28 238	27 818	2 %	44 883	44 509	1 %
BAC+1 BTS-DUT	22 761	23 120	-2 %	22 761	23 120	-2 %
BAC+1 Université	81 419	82 586	-1 %	88 190	89 193	-1 %
BAC+2 BTS	65 019	65 676	-1 %	67 647	68 303	-1 %
BAC+2 DUT	23 159	23 596	-2 %	24 976	25 484	-2 %
BAC+2 Université	27 668	27 176	2 %	30 507	29 894	2 %
DRJS	13 349	13 238	1 %	13 349	13 238	1 %
SOCIAL	6 971	6 971	0 %	11 781	11 781	0 %
SANTE	19 061	19 115	0 %	52 059	51 993	0 %
IUFM	18 396	18 403	0 %	28 137	27 973	1 %
Licences Professionnelles hors licences IUP	23 873	23 214	3 %	26 991	26 232	3 %
Licences Université	44 140	43 898	1 %	49 382	49 166	0 %
M1-MST-MSG-Maitrise IUP	40 726	40 240	1 %	46 222	45 874	1 %
Ecoles diverses tous niveaux	25 004	24 941	0 %	52 542	52 376	0 %
Ecoles de commerce tous niveaux	12 776	12 638	1 %	22 972	22 748	1 %
Ecoles d'ingénieurs	13 596	13 680	-1 %	27 692	27 770	0 %
DESS + master pro	45 964	45 389	1 %	50 615	49 928	1 %
DEA + master recherche + divers degrés 6 hors docteurs	33 205	32 915	1 %	38 281	38 181	0 %
Doctorat	14 623	14 707	-1 %	16 955	16 619	2 %
TOTAL	1 235 517	1 234 120	0 %	1 467 721	1 464 836	0 %

4.2. La modélisation de l'envoi en plateau d'enquête

L'échantillon utilisé était constitué de 191 968 individus. Parmi eux, 166 444 ont été envoyés en enquête. Pour les 25 524 autres (appelés les cas 4), aucune coordonnée téléphonique n'était disponible dans la base de sondage ni à l'issue des différentes recherches de coordonnées mises en œuvre.

La modélisation de la probabilité d'envoi en enquête s'est faite grâce à une régression logistique sur l'échantillon. On a modélisé la probabilité d'être envoyé en enquête en fonction des variables suivantes :

- présence d'un nom de voie sur l'adresse de l'individu en 2007 ;
- spécialité de formation (disciplinaire, industriel, tertiaire) ;
- présence d'un code postal sur l'adresse de l'individu en 2007 ;
- région de l'établissement de formation en 2007 ;
- information sur l'adresse de l'individu en formation (Zus, non Zus, inconnu) ;
- genre ;
- strate07.

Les individus dont l'adresse en 2007 ne disposait pas de code postal, disposaient d'un nom de voie erroné, n'avaient pas pu être géolocalisés ou habitaient en Zus étaient moins souvent envoyés en enquête que les autres. Les jeunes issus des rectorats étaient moins souvent envoyés en enquête. Les femmes étaient moins souvent envoyées en enquête que les hommes. Les individus issus de formations disciplinaires ou industrielles étaient plus souvent envoyés en enquête que ceux issus des filières tertiaires. Le programme et les résultats de la modélisation logistique sont consultables en annexe.

Le poids d'envoi en enquête retenu est égal à l'inverse de la probabilité d'envoi en enquête estimée par la procédure logistique. Il est compris entre 1 et 10,1 pour les individus effectivement envoyés en enquête. En moyenne, $PDSENVENQ = 1,16$ ($ET = 0,25$).

Figure 11.

Distribution des poids liés à l'envoi en enquête des individus envoyés en plateau d'enquête

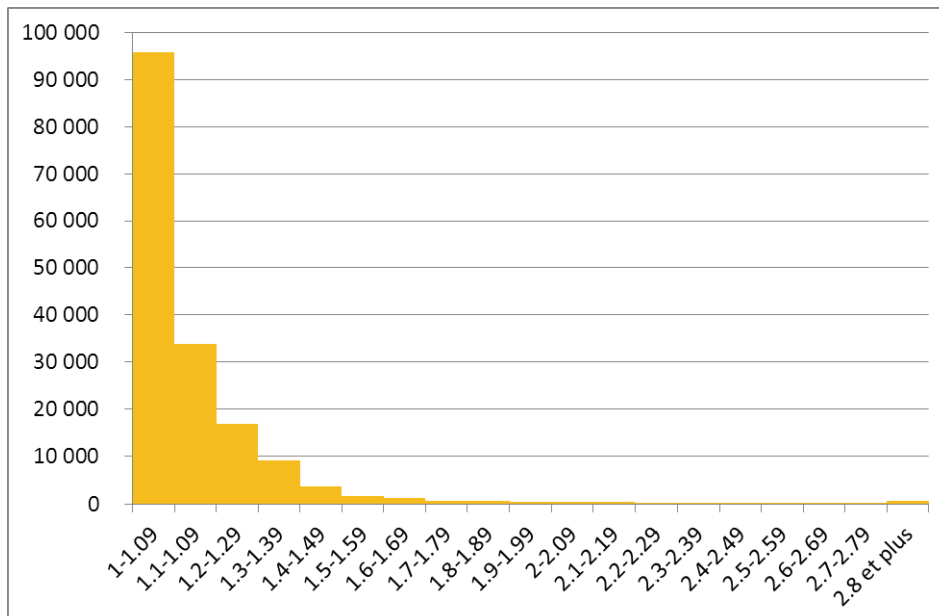


Tableau 18

Poids d'envoi en enquête dans l'échantillon envoyé en enquête selon la strate de formation

Strate de formation (strate07)	Nombre d'individus envoyés en enquête	poids moyen d'envoi en enquête	Ecart-type dans la strate	Somme des poids d'envoi en enquête
Collège	7 616	1,43	0,37	10 891
CAP/BEP (ttes années) scolaire industriel	5 048	1,2	0,09	6 058
CAP/BEP (ttes années) scolaire tertiaire	5 255	1,33	0,34	6 989
CAP/BEP (ttes années) apprentissage industriel	8 755	1,05	0,06	9 193
CAP/BEP (ttes années) apprentissage tertiaire	4 013	1,06	0,06	4 254
Lycée (2nde à BAC) apprentissage	5 704	1,05	0,06	5 989
Lycée (2nde à BAC) pro/techno industriel	7 208	1,15	0,15	8 289
Lycée (2nde à BAC) pro/techno tertiaire	8 776	1,29	0,26	11 321
Lycée (2nde à BAC) généraux	5 763	1,37	0,23	7 895
BAC+1 à BAC+5 apprentissage	4 726	1,05	0,07	4 962
BAC+1 BTS-DUT	1 943	1,3	0,23	2 526
BAC+1 Université	15 588	1,05	0,06	16 367
BAC+2 BTS	6 555	1,15	0,13	7 538
BAC+2 DUT	5 548	1,05	0,06	5 825
BAC+2 Université	4 350	1,16	0,25	5 046
DRJS	10 305	1,28	0,4	13 190
SOCIAL	6 559	1,06	0,06	6 953
SANTE	14 855	1,08	0,12	16 043
IUFM	2 205	1,01	0	2 227
Licences Professionnelles hors licences IUP	1 892	1,04	0,03	1 968
Licences Université	4 192	1,14	0,19	4 779
M1-MST-MSG-Maitrise IUP	5 098	1,12	0,18	5 710
Ecoles diverses tous niveaux	3 741	1,33	0,6	4 976
Ecoles de commerce tous niveaux	1 599	1,33	0,6	2 127
Ecoles d'ingénieurs	1 947	1,23	0,4	2 395
DESS + master pro	5 557	1,07	0,1	5 946
DEA + master recherche + divers degrés 6 hors docteurs	3 726	1,11	0,16	4 136
Doctorat	7 920	1,12	0,19	8 870
TOTAL	166 444	1,15		192 463

Tableau 19

Déformations potentielles induites par le coefficient d'envoi en enquête selon la strate de formation

Strate de formation	Somme des poids d'envoi en enquête	Nombre d'individus échantillonnés	Ecart relatif
Collège	10 891	10 732	1 %
CAP/BEP (ttes années) scolaire industriel	6 058	6 051	0 %
CAP/BEP (ttes années) scolaire tertiaire	6 989	6 898	1 %
CAP/BEP (ttes années) apprentissage industriel	9 193	9 206	0 %
CAP/BEP (ttes années) apprentissage tertiaire	4 254	4 246	0 %
Lycée (2nde à BAC) apprentissage	5 989	5 952	1 %
Lycée (2nde à BAC) pro/techno industriel	8 289	8 272	0 %
Lycée (2nde à BAC) pro/techno tertiaire	11 321	11 262	1 %
Lycée (2nde à BAC) généraux	7 895	7 873	0 %
BAC+1 à BAC+5 apprentissage	4 962	4 954	0 %
BAC+1 BTS-DUT	2 526	2 529	0 %
BAC+1 Université	16 367	16 515	-1 %
BAC+2 BTS	7 538	7 508	0 %
BAC+2 DUT	5 825	5 863	-1 %
BAC+2 Université	5 046	5 162	-2 %
DRJS	13 190	12 593	5 %
SOCIAL	6 953	6 971	0 %
SANTE	16 043	16 018	0 %
IUFM	2 227	2 219	0 %
Licences Professionnelles hors licences IUP	1 968	1 973	0 %
Licences Université	4 779	4 899	-2 %
M1-MST-MSG-Maitrise IUP	5 710	5 785	-1 %
Ecoles diverses tous niveaux	4 976	4 874	2 %
Ecoles de commerce tous niveaux	2 127	2 064	3 %
Ecoles d'ingénieurs	2 395	2 377	1 %
DESS + master pro	5 946	6 040	-2 %
DEA + master recherche + divers degrés 6 hors docteurs	4 136	4 161	-1 %
Doctorat	8 870	8 971	-1 %
TOTAL	192 463	191 968	0 %

4.3. La modélisation de la probabilité de réponse après envoi en enquête

Parmi les 166 444 individus envoyés en enquête, 48 436 sont répondants (29,1 %). On considère qu'un individu est répondant lorsqu'on a réussi à joindre l'individu recherché, qu'il a accepté de répondre à l'enquête. Ainsi, les individus qui se sont révélés hors du champ de l'enquête à l'issue du questionnaire filtre sont considérés comme des répondants.

On a étudié l'impact sur la probabilité de répondre du genre, de l'origine et du type des coordonnées téléphoniques disponibles, du lieu d'habitat en 2007 (Zus, non Zus), du passage par la phase de préqualification, du fait d'avoir réalisé sa formation par apprentissage, de la strate de formation, de la région de formation ainsi que de la spécialité de formation.

La modélisation de la probabilité de répondre s'est faite grâce à des régressions logistiques mises en œuvre de manière indépendante sur chaque région ayant donné lieu à une extension (Aquitaine, Bretagne, Franche-Comté, Nord-Pas-de-Calais, Poitou-Charente), sur l'Ile de France et sur les autres régions regroupées.

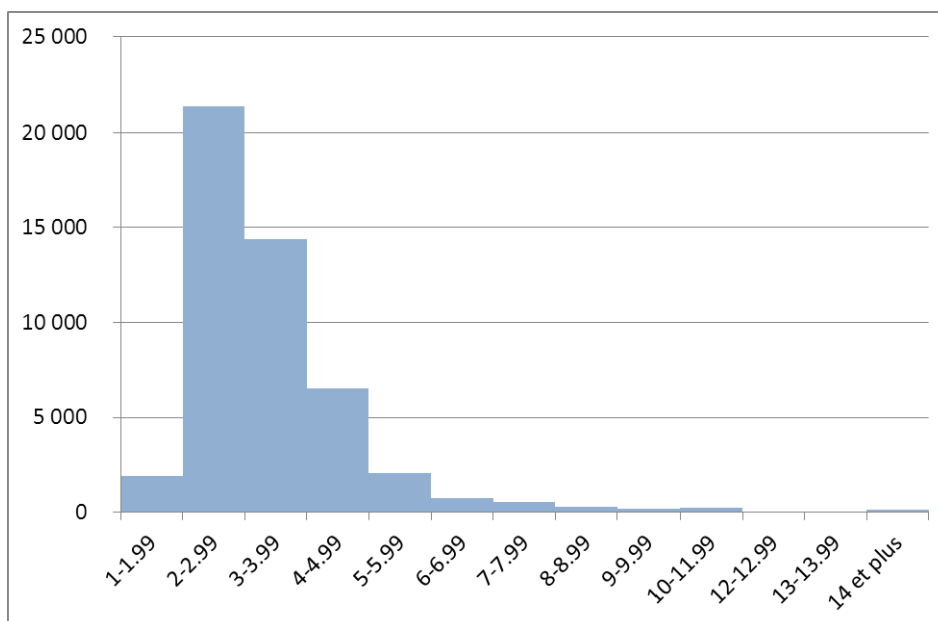
De manière générale, les individus domiciliés en Zus en 2007, ceux passés par la phase de préqualification, ceux issus d'une formation du tertiaire présentent des taux de réponse plus faibles. Les individus pour lesquels on dispose d'un numéro de téléphone portable répondent plus fréquemment que les autres. Selon les strates de formation, les niveaux de réponse sont également très contrastés.

Le coefficient de redressement corrigeant la non réponse est calculé pour tous les répondants. Il est égal à l'inverse de la probabilité de réponse estimée par la procédure logistique. Il est compris entre 1,34 et 37,9 (moyenne : 3,42 ; ET = 1,51).

Les programmes et les résultats détaillés sont consultables en annexe.

Figure 12

Distribution des poids liés à la probabilité de réponse des individus répondants



Les individus qui ont le moins souvent répondu sont les Docteurs ainsi que les sortants de CAP/BEP en apprentissage. Les individus sortants de licence et d'IUFM sont ceux qui ont le plus souvent répondu.

Tableau 20

Poids de réponse des répondants selon la strate de formation

Strate de formation (Strate07)	Nombre d'individus répondants	poids moyen de réponse	Ecart-type dans la strate	Somme des poids des répondants
Collège	2 324	3,21	1,28	7 460
CAP/BEP (ttes années) scolaire industriel	1 408	3,49	1,4	4 914
CAP/BEP (ttes années) scolaire tertiaire	1 329	3,77	1,58	5 010
CAP/BEP (ttes années) apprentissage industriel	2 149	4,16	1,43	8 940
CAP/BEP (ttes années) apprentissage tertiaire	802	5,02	1,69	4 026
Lycée (2nde à BAC) apprentissage	1 589	3,72	1,29	5 911
Lycée (2nde à BAC) pro/techno industriel	2 219	3,35	1,42	7 434
Lycée (2nde à BAC) pro/techno tertiaire	2 231	3,91	1,87	8 723
Lycée (2nde à BAC) généraux	1 854	3,03	0,77	5 618
BAC+1 à BAC+5 apprentissage	1 361	3,57	1,26	4 859
BAC+1 BTS-DUT	568	3,31	1,53	1 880
BAC+1 Université	4 756	3,3	1	15 695
BAC+2 BTS	2 142	3,09	1,2	6 619
BAC+2 DUT	1 726	3,26	1,07	5 627
BAC+2 Université	1 290	3,32	1,13	4 283
DRJS	2 614	3,86	2,55	10 090
SOCIAL	2 052	3,24	1,38	6 648
SANTE	4 394	3,37	1,26	14 808
IUFM	1 039	2,13	0,37	2 213
Licences Professionnelles hors licences IUP	816	2,3	0,63	1 877
Licences Université	1 533	2,66	0,81	4 078
M1-MST-MSG-Maitrise IUP	1 494	3,38	1,24	5 050
Ecoles diverses tous niveaux	1 214	2,88	1,16	3 496
Ecoles de commerce tous niveaux	431	3,6	1,42	1 552
Ecoles d'ingénieurs	685	2,81	1,03	1 925
DESS + master pro	1 927	2,86	0,78	5 511
DEA + master recherche + divers degrés 6 hors docteurs	1 119	3,26	1,08	3 648
Doctorat	1 370	5,63	2,6	7 713
TOTAL	48 436	3,42	1,51	165 606

Tableau 21

**Déformations potentielles induites par le coefficient de redressement de la non réponse
selon la strate de formation**

Strate de formation	Somme des poids d'envoi en enquête	Nombre d'individus envoyés en enquête	Ecart relatif
Collège	7 460	7 616	-2 %
CAP/BEP (ttes années) scolaire industriel	4 914	5 048	-3 %
CAP/BEP (ttes années) scolaire tertiaire	5 010	5 255	-5 %
CAP/BEP (ttes années) apprentissage industriel	8 940	8 755	2 %
CAP/BEP (ttes années) apprentissage tertiaire	4 026	4 013	0 %
Lycée (2nde à BAC) apprentissage	5 911	5 704	4 %
Lycée (2nde à BAC) pro/techno industriel	7 434	7 208	3 %
Lycée (2nde à BAC) pro/techno tertiaire	8 723	8 776	-1 %
Lycée (2nde à BAC) généraux	5 618	5 763	-3 %
BAC+1 à BAC+5 apprentissage	4 859	4 726	3 %
BAC+1 BTS-DUT	1 880	1 943	-3 %
BAC+1 Université	15 695	15 588	1 %
BAC+2 BTS	6 619	6 555	1 %
BAC+2 DUT	5 627	5 548	1 %
BAC+2 Université	4 283	4 350	-2 %
DRJS	10 090	10 305	-2 %
SOCIAL	6 648	6 559	1 %
SANTE	14 808	14 855	0 %
IUFM	2 213	2 205	0 %
Licences Professionnelles hors licences IUP	1 877	1 892	-1 %
Licences Université	4 078	4 192	-3 %
M1-MST-MSG-Maitrise IUP	5 050	5 098	-1 %
Ecoles diverses tous niveaux	3 496	3 741	-7 %
Ecoles de commerce tous niveaux	1 552	1 599	-3 %
Ecoles d'ingénieurs	1 925	1 947	-1 %
DESS + master pro	5 511	5 557	-1 %
DEA + master recherche + divers degrés 6 hors docteurs	3 648	3 726	-2 %
Doctorat	7 713	7 920	-3 %
TOTAL	165 606	166 444	-1 %

4.4. La suppression des individus hors du champ de l'enquête et calage aux marges

Pour la pondération nationale, les individus qui se sont révélés hors du champ de l'enquête à l'issue du questionnaire filtre (en particulier les individus qui ont poursuivi leurs études en 2007-2008) ont été supprimés. Un total de 23 857 individus (49 %) s'est ainsi révélé hors du champ de l'enquête Céreq à l'issue du questionnaire filtre.

Les taux de hors champ sont très élevés chez les sortants de collège et de lycée général (poursuites d'études non identifiées au niveau de la base de sondage), ainsi que pour les formations du Sport et de l'Animation (nombreux sortants post initiaux).

Tableau 22

Appartenance au champ de l'enquête

strate07	Dans le champ	Hors du champ	Total	Taux de hors champ
Collège	257	2 067	2 324	89 %
CAP/BEP (ttes années) scolaire industriel	620	788	1 408	56 %
CAP/BEP (ttes années) scolaire tertiaire	701	628	1 329	47 %
CAP/BEP (ttes années) apprentissage industriel	1 227	922	2 149	43 %
CAP/BEP (ttes années) apprentissage tertiaire	461	341	802	43 %
Lycée (2nde à BAC) apprentissage	1 014	575	1 589	36 %
Lycée (2nde à BAC) pro/techno industriel	1 088	1 131	2 219	51 %
Lycée (2nde à BAC) pro/techno tertiaire	1 289	942	2 231	42 %
Lycée (2nde à BAC) généraux	316	1 538	1 854	83 %
BAC+1 à BAC+5 apprentissage	826	535	1 361	39 %
BAC+1 BTS-DUT	255	313	568	55 %
BAC+1 Université	2 089	2 667	4 756	56 %
BAC+2 BTS	1 463	679	2 142	32 %
BAC+2 DUT	974	752	1 726	44 %
BAC+2 Université	549	741	1 290	57 %
DRJS	772	1 842	2 614	70 %
SOCIAL	991	1 061	2 052	52 %
SANTE	2 184	2 210	4 394	50 %
IUFM	751	288	1 039	28 %
Licences Professionnelles hors licences IUP	684	132	816	16 %
Licences Université	820	713	1 533	47 %
M1-MST-MSG-Maitrise IUP	867	627	1 494	42 %
Ecoles diverses tous niveaux	254	960	1 214	79 %
Ecoles de commerce tous niveaux	252	179	431	42 %
Ecoles d'ingénieurs	421	264	685	39 %
DESS + master pro	1 596	331	1 927	17 %
DEA + master recherche + divers degrés 6 hors docteurs	710	409	1 119	37 %
Doctorat	1 148	222	1 370	16 %
Total	24 579	23 857	48 436	49 %

Ainsi, on estime avant calage à 796 000 le nombre de sortants du système éducatif au cours ou à l'issue de l'année scolaire 2006-2007, dont 11 % de jeunes sortis sans aucun diplôme, 19 % de jeunes ayant un CAP ou un BEP comme plus haut diplôme, 27 % ayant un baccalauréat comme plus haut diplôme et 43 %, un diplôme de l'enseignement supérieur. Le calage est effectué sur les marges des niveaux de diplôme (en 6 postes) de l'Enquête Emploi réalisée par l'INSEE. Cette enquête sert de référence dans les publications du ministère de l'Education nationale

Tableau 23

Les marges de calage

	Répartition utilisée pour le calage		Répartition déduite de la pondération avant calage de l'enquête 2010
	de l'enquête 2010 auprès de la Génération 2007	de l'enquête 2007 auprès de la Génération 2004	
Sans qualification (niveau VI et Vbis)	8	6,6	4,3
Sans diplôme mais qualifié (non diplômés de niveau V et IV)	10	10,1	6,6
CAP – BEP et équivalent (diplômés de niveau V)	17	17,3	18,9
BAC et équivalent (diplômés de niveau IV)	23	24	27,0
Supérieur court (diplômés de niveau III)	16	18,6	14,5
Supérieur long (diplômés de niveau II et I)	26	23,4	28,7
Ensemble	100	100	100
Effectifs totaux	739 000	737 000	796 000
Source	Insee, Enquêtes Emploi 2006, 2007, 2008, (DEPP, L'Etat de l'Ecole, nov. 2010)	Insee Enquêtes Emploi 2003, 2004, 2005, (DEPP, L'Etat de l'Ecole, nov. 2010)	Cereq, enquête 2010 auprès de la génération 2007, pondération avant le calage sur marge

Les résultats avant calage sont légèrement différents de ceux observés dans l'enquête Emploi. En particulier, le nombre de sortants est surestimé, notamment au niveau du baccalauréat et de l'enseignement supérieur long. A contrario, le nombre de jeunes sortis sans diplôme est sous-estimé. C'est pourquoi, un calage sur marges a été effectué de façon à retrouver des effectifs totaux de sortants et une répartition par plus haut diplôme conformes aux publications de la DEPP. Il a été effectué à l'aide de la macro-procédure CALMAR, de l'INSEE. Le tableau 25 présente les marges utilisées, estimées en compilant les enquêtes trimestrielles de trois années successives, 2006 à 2008.

4.5. Le traitement des poids élevés

A l'issue du calage sur marge, le poids final associé à certains individus est très élevé, ce qui peut conduire à leur donner une influence très importante dans certaines analyses, par niveau du plus haut diplôme par exemple. Une telle situation n'est pas satisfaisante parce qu'elle peut conduire à des résultats peu robustes. Les poids les plus élevés se concentrent essentiellement sur les sortants non diplômés, puis dans une moindre mesure sur les sortants de CAP ou BEP, de BTS ou DUT. Les domaines et régions de l'échantillon concernés par des extensions sont épargnés par cette difficulté.

Un total de 24 579 individus dans le champ du Céreq ont répondu à l'enquête Génération 2007. Ils sont représentatifs des 739 000 jeunes sortis du système éducatif en 2007. En moyenne, le poids d'un individu ayant répondu à l'enquête après calage est donc de 30.

Le nombre de questionnaires financé par le Céreq était de 3 000, les autres étaient destinés à répondre aux besoins des partenaires à l'origine d'une extension (régionale ou thématique). Les individus répondants n'appartenant à aucune extension présentent initialement des taux de tirage très faibles et, de façon mécanique, des poids après calage très élevés. Ce phénomène est particulièrement vrai pour les moins qualifiés, qui présentent à la fois des taux de tirage et des taux de réponse à l'enquête particulièrement faibles.

Afin d'améliorer la robustesse des résultats et de limiter la variance, nous avons décidé d'écarter les poids supérieurs à 400. Un total de 147 individus est concerné par l'écarterage. Le basculement de poids a été fait à la faveur d'appariements avec des individus aux caractéristiques initiales proches mais présentant des poids plus faibles. Plusieurs étapes ont été nécessaires pour effectuer le basculement des poids. Durant chacune de ces étapes, lorsque l'appariement est possible (i.e. lorsque le basculement de poids ne conduit pas pour le(s) jumeau(x) à un poids supérieur à 400), on effectue un transfert de poids. Les individus dont le poids a été affecté lors d'une étape antérieure ne sont plus concernés par les étapes suivantes :

- Étape 1 : pour chaque poids élevé, on cherche des jumeaux au sein du même établissement de formation et de la même strate de formation agrégée (strate07). A l'issue de cette étape, 31 individus sont écartés et 57 jumeaux voient leur poids augmenter.
- Étape 2 : pour chaque poids élevé restant, on cherche des jumeaux au sein de la même région de formation et de la même strate fine de formation (strateg). A l'issue de cette étape, 51 individus sont écartés et 349 jumeaux voient leur poids augmenter.
- Étape 3 : pour chaque poids élevé restant, on cherche des jumeaux au sein de la même strate fine de formation (strateg). A l'issue de cette étape, 45 individus sont écartés et 4 140 jumeaux voient leur poids augmenter.
- Étape 4 : il reste 20 individus non écartés. On bascule leur supplément de poids sur les 19 878 individus non encore affectés par les redressements et dont le poids est inférieur à 390.

L'écarterage ne modifie pas la somme des poids car chaque surplus de poids d'un individu à poids élevé a été basculé sur un autre individu. Par contre, le basculement de poids se fait au regard d'informations présentes dans la base de sondage (région et strate de formation) et non des réponses données durant l'enquête (plus haut diplôme obtenu). Les marges de calage sont ainsi très légèrement modifiées par l'écarterage des poids élevés.

Pour retrouver les marges de calage, on effectue un nouveau calage à partir des poids corrigés par l'écarterage. Les poids définitifs après recalage sont inférieurs à 480.

4.6. La pondération arrondie

La multiplication des différents coefficients ayant servi pour obtenir la pondération conduit à des poids non arrondis. Selon les logiciels statistiques et les procédures utilisés, la gestion de la partie non entière du poids est différente.

Ainsi pour le logiciel SAS, la PROC FREQ utilisée avec l'option WEIGHT donne des résultats exacts tandis que la PROC TABULATE avec l'option FREQ ne tient compte que de la partie entière.

La livraison d'une pondération arrondie présente deux avantages :

- elle permet d'obtenir des statistiques indépendantes des logiciels ou procédures mis en œuvre
- elle donne des effectifs ronds sur les statistiques descriptives.

Les arrondis ont été calculés de la manière suivante.

Première étape :

Les individus ont été regroupés par strates de poids. Une strate est constituée de l'ensemble des individus qui possèdent la même partie entière de poids et qui sont situés dans la même modalité de calage. Pour chaque individu de la strate, le poids final (p_{arr}) sera égal soit à la partie entière de son poids initial (p_{narr}), soit à la partie entière de son poids initial+1.

A l'intérieur de chaque strate, on effectue la somme des poids des individus et on retranche la partie entière du poids des individus multipliée par le nombre d'individus présents dans la strate. Le reste (sa partie entière plus exactement) obtenu correspond au poids qui doit être réaffecté aux individus de la strate.

On appelle p_{1arr} le poids de chaque individu à la fin de cette première étape.

Supposons que la strate soit composée de n individus et que la partie entière du reste soit égal à r ($r < n$). Les individus au sein de la strate sont classés par ordre de poids croissant. On affecte la partie entière de leur poids initial aux $n-r$ individus de la strate dont le poids initial était le plus faible ($p_{1arr} = \lfloor p_{narr} \rfloor$) et la partie entière de leur poids initial augmentée de un aux r individus de la strate dont le poids initial était le plus élevé ($p_{1arr} = \lfloor p_{narr} \rfloor + 1$).

Seconde étape :

Pour chaque marge de calage, la première étape peut faire varier très légèrement la somme des poids. On va donc corriger les marges.

On calcule pour chaque individu d'une marge donnée la quantité $\delta = p_{1arr} - p_{arr}$ ($-1 < \delta < 1$).

On calcule ensuite la différence sur la somme des poids de la marge avant et après arrondi et on ré-équilibre.

Par exemple, si la marge de calage a perdu p individus en pondéré, on trie par δ croissant et on ajoute un au poids des p premiers individus ($p_{arr} = p_{arr} + 1$).

Si la marge de calage a gagné q individus en pondéré, on trie par δ décroissant et on retranche un au poids des q premiers individus ($p_{arr} = p_{arr} - 1$).

Au final, les marges de calage sont strictement respectées et la différence de poids entre le poids non arrondi et le poids arrondi est strictement inférieure à un en valeur absolue.

4.7. Récapitulatif

Le tableau 24 rend compte des différents coefficients utilisés afin d'obtenir le poids définitif pour chaque individu répondant dans le champ du Céreq. Les coefficients les plus élevés s'observent au niveau du tirage de l'échantillon (les individus n'appartenant pas à une extension présentent des poids de tirage élevés) et de la réponse suite à envoi en enquête.

Tableau 24

Distribution des différentes pondérations successives et de leurs composantes

Coefficient	Min	5 %	10 %	25 %	Médiane	75 %	90 %	95 %	Max
Poids initial	1	1	1	1,21	2,18	6,91	18,83	29,06	111,7
Correction du taux de couverture	1	1	1	1	1,08	1,41	2,05	2,23	7,69
Correction du taux d'envoi en enquête	1	1,02	1,03	1,04	1,06	1,13	1,25	1,32	6,7
Correction de la non réponse	1,34	2,02	2,17	2,47	3,05	3,87	4,9	5,91	37,91
Correction du calage sur marge	0,79	0,79	0,79	0,79	0,84	1,02	1,02	1,4	1,74
1er poids final	1,57	2,68	3,26	5,24	9,96	24,93	68,57	127,86	2093,92
Poids corrigé des poids élevé avant recalage	1,6	2,75	3,34	5,41	10,35	25,69	71,11	135,06	400
Poids corrigé calé et arrondi	1	3	3	5	10	25	70	134	476

Champ : 24 579 répondants dans le champ de l'enquête du Céreq

5. AUTRES PONDÉRATIONS

Le champ de l'enquête pour les extensions régionales est celui du Céreq. Les extensions régionales ont été traitées lors de la pondération nationale et ne donnent pas lieu à une pondération spécifique. Dans la pondération nationale, la non réponse a été modélisée au niveau régional pour chaque région d'extension. Les extensions régionales ont fait l'objet de contrôles spécifiques.

5.1. L'extension des 5 000

5.1.1. La reconstitution des poids de tirage

La reconstitution des poids de tirage s'est effectuée de la même manière que pour l'échantillon des 30 000, en fonction des utilisations partielles ou totales de la réserve. Un total de 36 856 individus constitue l'échantillon des 5 000. Le tableau 25 rend compte des écarts observés entre la somme des poids observés et attendus, eu égard aux effectifs présents dans la base de sondage.

Hormis pour les individus issus d'une formation de la santé qui possèdent des poids très élevés par rapport à l'attendu, les sommes des poids par strate de formation sont cohérentes avec les effectifs initialement présents dans la base de sondage.

Tableau 25
Poids de tirage

Strate de formation (Strate07)	Nombre d'individus dans la base de sondage	Nombre d'individus échantillonnés	Somme des poids	Ecart entre l'observé et l'attendu
Collège	150 621	4 088	154 437	3 %
CAP/BEP (ttes années) scolaire industriel	56 755	1 488	56 190	-1 %
CAP/BEP (ttes années) scolaire tertiaire	59 189	1 695	61 661	4 %
CAP/BEP (ttes années) apprentissage industriel	58 452	2 296	58 180	0 %
CAP/BEP (ttes années) apprentissage tertiaire	26 549	1 150	26 280	-1 %
Lycée (2nde à BAC) apprentissage	34 264	1 101	34 448	1 %
Lycée (2nde à BAC) pro/techno industriel	72 662	1 389	72 608	0 %
Lycée (2nde à BAC) pro/techno tertiaire	116 147	2 327	115 306	-1 %
Lycée (2nde à BAC) généraux	100 160	2 588	96 866	-3 %
BAC+1 à BAC+5 apprentissage	27 818	723	23 026	-17 %
BAC+1 BTS-DUT	23 120	358	22 032	-5 %
BAC+1 Université	82 586	3 440	81 566	-1 %
BAC+2 BTS	65 676	1 491	65 249	-1 %
BAC+2 DUT	23 596	970	22 795	-3 %
BAC+2 Université	27 176	1 168	27 976	3 %
DRJS	13 238	514	12 343	-7 %
SOCIAL	6 971	463	7 647	10 %
SANTE	19 115	1 433	29 113	52 %
IUFM	18 403	548	18 263	-1 %
Licences Professionnelles hors licences IUP	23 214	284	22 500	-3 %
Licences Université	43 898	893	45 755	4 %
M1-MST-MSG-Maitrise IUP	40 240	1 949	40 347	0 %
Ecoles diverses tous niveaux	24 941	550	20 789	-17 %
Ecoles de commerce tous niveaux	12 638	841	12 758	1 %
Ecoles d'ingénieurs	13 680	378	13 289	-3 %
DESS + master pro	45 389	1 281	45 039	-1 %
DEA + master recherche + divers degrés 6 hors docteurs	32 915	887	34 034	3 %
Doctorat	14 707	563	15 048	2 %
Total	1 234 120	36 856	1 235 546	0 %

5.1.2.. Le calcul de la pondération des 5 000

Pour l'échantillon des 5 000, les coefficients individuels utilisés pour le redressement de la couverture et l'envoi en enquête sont ceux utilisés pour la pondération des 30 000. Un total de 30 680 individus ont effectivement été envoyés en plateau d'enquête. L'échantillon des 5 000 possédant des spécificités par rapport à celui des 30 000, le coefficient de redressement de la non réponse après envoi en enquête a fait l'objet d'une modélisation spécifique. Les variables retenues dans cette modélisation sont le fait d'habiter en ZUS, la strate et la spécialité de formation, la région de l'établissement de formation, le genre, le type de coordonnées téléphoniques disponibles et le fait d'être passé par le plateau de pré-qualification.

Un total de 9 397 individus a répondu au questionnaire filtre. Après suppression des individus hors champ, il reste 4 695 individus dans le champ de l'enquête. Parmi eux, 179 présentent un poids élevé avant calage (supérieur au poids moyen de leur strate augmenté de deux écart-types). Pour ces individus, le poids avant calage est borné (au poids moyen de leur strate augmenté de deux écart-types). Les marges de calage

appliquées sont les mêmes que pour l'échantillon des 30 000. A l'issue du calage, une procédure similaire à celle effectuée sur l'échantillon des 30 000 est mise en œuvre pour arrondir les poids.

Tableau 26

**Distribution des différentes pondérations successives et de leurs composantes
pour l'échantillon des 5 000**

Coefficient	Min	5%	10%	25%	Médiane	75%	90%	95%	Max
Poids initial	1,96	20,9	22,1	24,9	30,1	44,5	70,6	74,5	89,6
Correction du taux de couverture	1	1	1	1	1,1	1,4	1,8	2,1	7,7
Correction du taux d'envoi en enquête	1	1	1	1	1,1	1,2	1,3	1,4	6,7
Correction de la non réponse	1,4	1,9	2,1	2,4	3	3,7	4,5	5,2	14,3
Gestion des poids élevés et correction du calage sur marge	0,3	0,8	0,8	0,9	0,9	1	1,1	1,5	1,8
Poids corrigé calé et arrondi	7	54	63	84	123	190	295	382	1005

Champ : 4 695 répondants dans le champ de l'enquête du Céreq

5.1.3. La pondération des stages

Le module stages est destiné à obtenir des informations sur les stages d'une durée supérieure ou égale à un mois effectués par les individus pendant leurs études. Sont concernés par ce module les primo-sortants de l'échantillon des 5 000, qui n'étaient pas en apprentissage ou élèves fonctionnaires durant leur dernière année d'études.

La formulation du module est différente selon que les individus sont issus ou non d'une formation sanitaire ou sociale. En effet, dans les formations sanitaires et sociales, le nombre de stages durant les études est souvent très important, les individus réalisant plusieurs stages pendant la même année scolaire. Il apparaissait d'une part fastidieux de lister l'ensemble des stages (temps de passation très important et effet mémoire), d'autre part les stages sont souvent de même nature.

Pour les individus sortants des formations sanitaires et sociales, on demande le nombre de stages effectués par an, la durée totale de ces stages, le nom de l'entreprise où a été effectué le plus long stage, la durée de ce stage et le secteur de cette entreprise (public ou privé).

Pour les autres individus, on demande la liste des stages, du plus récent au plus ancien, dans la limite de 10 stages.

Pour chaque individu, on décrit ensuite entre zéro et deux stages. On commence par identifier le stage le plus long. Si ce stage a été effectué dans une entreprise appartenant au secteur public, il est décrit dans tous les cas. S'il a été effectué dans une entreprise du secteur privé, il est décrit, de manière aléatoire une fois sur deux.

Si l'individu a effectué deux stages ou plus, on décrit au hasard un stage hors stage le plus long.

Par exemple, si l'individu a effectué deux stages, son stage le moins long sera obligatoirement décrit. Imaginons un individu qui a effectué 5 stages. Son stage le plus long sera obligatoirement décrit s'il a été effectué dans le secteur public et aura une chance sur deux d'être décrit s'il a été effectué dans le secteur privé. Un de ses 4 autres stages pris au hasard sera également décrit.

Seuls les stages décrits sont pondérés. Le poids d'un stage décrit est obtenu en divisant le poids de l'individu par la probabilité de décrire ce stage.

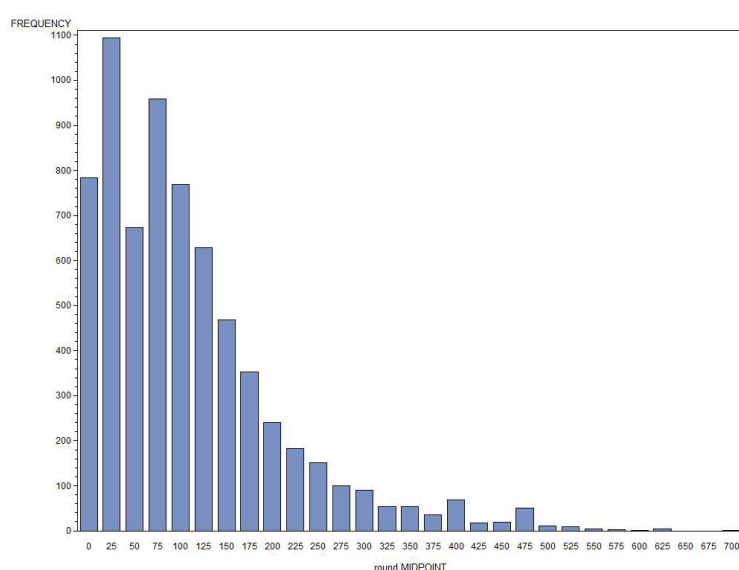
5.2. L'extension Pôle Emploi

Le champ de cette extension est le suivant : habitant en ZUS en 2007, non diplômé (plus haut diplôme) ou appartenant à l'échantillon des 5 000. Un total de 6 825 individus est concerné.

Les individus de ZUS et les non diplômés conservent dans un premier temps leur poids issu de l'échantillon global et les autres leur poids issu de la pondération des 5 000. Un calage est ensuite effectué afin que la somme des poids soit égale à 739 000. Les poids sont arrondis afin de simplifier l'analyse.

Figure 13

Distribution des poids de réponse des individus répondants à l'extension Pôle Emploi



5.3. L'extension Drees

5.3.1. Le contexte

L'extension commanditée par la Direction de la recherche, des études, de l'évaluation et des statistiques (Drees) concerne les formations suivantes :

Dans le secteur de la santé : Infirmier, Sage-femme, Masseur-Kinésithérapeute, Aide-soignant, Puéricultrice, Manipulateur en électroradiologie médicale, Ergothérapeute, Psychomotricien, Pédicure-podologue, Auxiliaire de puériculture

Dans le secteur social : Conseiller en économie sociale familiale, Assistant de service social, Éducateur spécialisé, Éducateur de jeunes enfants, Moniteur-Éducateur

Pour chacune de ces formations, la Drees souhaite interroger tous les individus ayant terminé leurs études en 2007. Sont ainsi concernés les primo-sortants qui appartiennent au champ Céreq mais également les individus post-initiaux (i.e. ceux qui avant l'année 2007 avaient interrompu leurs études pendant une année ou plus).

Les répondants primo-sortants sont intégrés dans les tables du Céreq et pondérés de la même manière que les autres répondants appartenant au champ de l'enquête Céreq. Ainsi, la Drees peut effectuer pour les individus primo-sortants de son champ des comparaisons par rapport aux autres sortants de la Génération 2007.

Etant donné que la Drees souhaite interroger les post-initiaux, une pondération spécifique sur l'ensemble du champ de la Drees comprenant les primo-sortants et les post-initiaux, adaptée à l'analyse des sortants de formations sanitaires et sociales en 2007 est également nécessaire. La suite du paragraphe concerne opérations effectuées pour le redressement de l'ensemble des individus de la Drees ('primo-sortants' et 'post-initiaux').

Pour calculer la couverture, les effectifs présents dans la base de sondage ont été rapportés aux effectifs issus du Fichier national des établissements sanitaires et sociaux (Finess).

Tableau 27

Couverture sur le champ de la Drees

Diplôme	Secteur	Effectifs présents dans la base de sondage	Effectifs présents dans FINESS	Taux de couverture
Infirmier	Santé	10533	21813	48 %
Masseur-kinésithérapeute	Santé	758	1603	47 %
Pédicure-Podologue	Santé	165	378	44 %
Manipulateur d'électro-radiologie	Santé	98	552	18 %
Ergothérapeute	Santé	61	331	18 %
Psychomotricien	Santé	89	239	37 %
Sage-femme	Santé	132	842	16 %
Puéricultrice	Santé	95	988	10 %
Aide-soignant	Santé	6547	22847	29 %
Auxiliaire de puéricultrice	Santé	637	3118	20 %
Conseiller en économie sociale et familiale	Santé	577	958	60 %
Assistant de service social	Social	1115	2381	47 %
Educateur de jeunes enfants	Social	672	1202	56 %
Educateur spécialisé	Social	2711	4167	65 %
Moniteur éducateur	Social	1094	2464	44 %
Total		25284	63883	40 %

5.3.2. Les résultats de l'enquête

Un total de 22 187 individus issus de formations sanitaires et sociales a été échantillonné dans la base de sondage dont 6 262 (28 %) ont accepté de répondre à l'enquête. Parmi eux, 860 se sont révélés hors du champ de la Drees à l'issue du questionnaire filtre. Un total de 5 402 questionnaires a finalement été collecté à l'issue de l'enquête.

Tableau 28

Taux de réponse selon le diplôme

Diplôme	Echantillonnés	Nombre de répondants		Taux de réponse
		dans le champ Drees	hors champ Drees	
Infirmier	8849	2052	330	27 %
Masseur-kinésithérapeute	624	151	15	27 %
Pédicure-Podologue	130	34	2	28 %
Manipulateur d'électro-radiologie	75	13	2	20 %
Ergothérapeute	48	18	0	38 %
Psychomotricien	68	19	3	32 %
Sage-femme	109	29	1	28 %
Puéricultrice	80	24	2	33 %
Aide-soignant	5480	1335	255	29 %
Auxiliaire de puéricultrice	555	101	8	20 %
Conseiller en économie sociale et familiale	577	130	25	27 %
Assistant de service social	1115	336	48	34 %
Educateur de jeunes enfants	672	210	19	34 %
Educateur spécialisé	2711	681	109	29 %
Moniteur éducateur	1094	269	41	28 %
Total	22187	5402	860	28 %

5.3.3. Les redressements

Pour réaliser la pondération sur le champ de la Drees, nous utilisons l'ensemble des coefficients issus de la pondération non calée réalisée par le Céreq sur l'ensemble des répondants, en incluant les individus hors du champ Céreq de l'enquête.

Nous restreignons ensuite le champ aux formations sanitaires et sociales sélectionnées par la Drees et réalisons le calage à partir des marges issues de Finess.

Sur les diplômes où il y a plus de 100 répondants (y compris hors champ), la marge appliquée est l'effectif fourni sur le diplôme par le fichier Finess. Lorsque ce n'est pas le cas, on opère un regroupement. Ainsi, dans les formations de la santé, les diplômes de Pédicure-Podologue, Manipulateur d'électroradiologie, Ergothérapeute, Psychomotricien, Sage-femme et Puéricultrice ont été regroupés.

Le calage sur marges se rapporte aux effectifs bruts et non aux pourcentages (différence sensible pour les marges faibles). Il s'effectue à partir de l'ensemble des individus répondants, y.c. les hors-champs.

A l'issue de la suppression des individus hors champ, le nombre de répondants dans le champ de la Drees s'élève à 5 402 et la somme des poids à 54 555. Autrement dit, sur les 63 883 individus diplômés d'après la source FINESS, on considère que 54 555 font partie du champ de l'enquête défini pour la Drees.

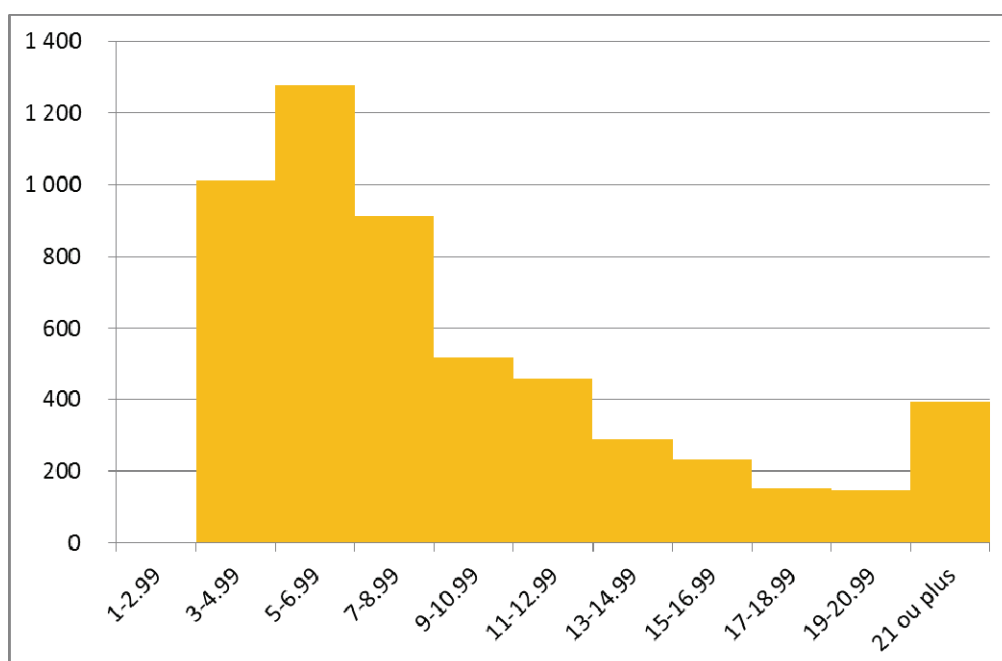
Tableau 29

Marges de calage, poids moyens des individus et estimation du nombre de sortants dans le champ de la DREES

Diplôme	Marge	Nombre de répondants dans le champ	Poids moyen	Estimation nb d'individus dans le champ
Infirmier	21 813	2 052	9,06	18 601
Masseur-kinésithérapeute	1 603	151	9,23	1 394
Pédicure-Podologue, Manipulateur d'électro-radiologie, Ergothérapeute, Psychomotricien, Sage-femme, Puéricultrice	3 330	137	22,33	3 059
Aide-soignants	22 847	1 335	14,25	19 030
Auxiliaire de puéricultrice	3 118	101	28,28	2 856
Conseiller en économie sociale et familiale	958	130	6,18	804
Assistant de service social	2 381	336	6,13	2 061
Educateur de jeunes enfants	1 202	210	5,28	1 109
Educateur spécialisé	4 167	681	5,18	3 530
Moniteur éducateur	2 464	269	7,84	2 110
Total	63 883	5 402	10,10	54 554

Figure 14.

Distribution des poids de réponse des individus répondants dans le champ de la Drees



5.4. L'extension Sport

5.4.1. Le contexte

L'extension commanditée par le Ministère de la Santé et des Sports concerne les formations du sport et de l'animation suivantes :

- Le Brevet d'État d'Éducateur Sportif du 1er degré (BEES)
- Le Brevet Professionnel de la Jeunesse, de l'Éducation Populaire et du Sport (BP JEPS) spécialités animation
- Le Brevet Professionnel de la Jeunesse, de l'Éducation Populaire et du Sport (BP JEPS) spécialités sportives
- Le Brevet d'Aptitude Professionnelle d'Assistant Animateur Technicien (BAPAAT)
- Le Brevet d'État d'Animateur Technicien de l'Éducation Populaire (BEATEP)
- Le Diplôme d'État Fonction Animation (DEFA).

Pour chacune de ces formations, le ministère de la Santé et des Sports souhaite interroger tous les individus ayant terminé leurs études en 2007. Sont ainsi concernés les primo-sortants qui appartiennent au champ Céreq mais également les individus post-initiaux (i.e. ceux qui avant l'année 2007 avaient déjà interrompu leurs études pendant une année ou plus).

Les répondants primo-sortants sont intégrés dans les tables du Céreq et pondérés de la même manière que les autres répondants appartenant au champ de l'enquête Céreq. Ainsi, le Ministère de la Santé et des Sports peut effectuer pour les individus primo-sortants de son champ des comparaisons par rapport aux autres sortants de la Génération 2007.

Une pondération spécifique sur l'ensemble du champ du ministère de la Santé et des Sports comprenant les primo-sortants et les post-initiaux, adaptée à l'analyse des sortants de formations du sport et de l'animation en 2007 est également nécessaire.

Les fichiers de diplômés des formations du sport et de l'animation sont traditionnellement collectés auprès des Directions Régionales de la Jeunesse, du Sport et de la Cohésion Sociale (DRJSCS). Pour Génération 2004, on ne disposait pas d'autre source d'information sur les effectifs de diplômés dans ces formations. Le Ministère de la Santé et des Sports ayant récemment créé une base nationale contenant les effectifs de diplômés par diplôme et région, le Céreq s'est procuré cette base afin d'estimer la couverture de la base de sondage pour Génération 2007 grâce à cette nouvelle source.

Pour Génération 2007, il est à noter que toutes les DRJSCS ont transmis leur fichier de diplômés au Céreq.

La comparaison entre les informations issues de la base de sondage et de la base nationale donne un volume global équivalent (environ 13 000 diplômés), avec des différences sur deux régions : Ile de France où le fichier du Céreq semble plus exhaustif que celui transmis par le Ministère, Aquitaine où le fichier du Ministère semble plus exhaustif.

Pour l'Ile de France, le Ministère estime que sa base n'est pas complète et que toutes les informations ne lui ont pas été transmises donc il semble que les informations issues de la base de sondage du Céreq soient plus justes.

Pour l'Aquitaine, les informations transmises au Céreq ne concernent que les BEES alors que dans la base ministère, d'autres diplômes sont représentés dans la Région. La base du Céreq est certainement moins fiable à ce niveau.

Étant donné que la base nationale n'est pas encore stabilisée et qu'elle ne permet pas de distinguer les diplômés au niveau le plus fin (spécialité sportive ou d'animation), nous avons décidé de considérer que la

base de sondage était exhaustive sur les formations du sport et de l'animation et de ne pas procéder à un redressement de la couverture.

Tableau 30

Comparaison des effectifs présents dans la base de sondage et dans la base nationale du ministère de la Santé et des Sports

Diplôme	Effectifs présents dans la base de sondage du Céreq	Effectifs présents dans la base nationale
BEES premier degré	6 716	5 574
Brevet professionnel (sport et animation)	3 887	5 066
Autre diplôme	1 947	2 263
Total	12 550	12 903

5.4.2. Les résultats de l'enquête

Le ministère de la Santé et des Sports souhaitait obtenir le maximum de questionnaires sur le champ des diplômés des formations du sport et de l'animation. Tous les individus présents dans la base de sondage, soit 12 550, ont donc été échantillonnés.

Parmi eux, 2 609 (21 %) ont accepté de répondre à l'enquête. 604 se sont révélés hors du champ du ministère de la Santé et des Sports à l'issue du questionnaire filtre. Un total de 2 005 questionnaires a finalement été collecté à l'issue de l'enquête.

Tableau 31

Taux de réponse selon le diplôme

Diplôme	Echantillonnés	Nombre de répondants		Taux de réponse
		dans le champ du sport	hors champ sport	
BEES premier degré	6 716	956	340	19 %
Brevet professionnel spécialité animation	1 891	377	65	23 %
Brevet professionnel spécialité sportive	1 996	354	96	23 %
Autre diplôme du sport et de l'animation	1 947	318	103	22 %
Total	12 550	2 005	604	21 %

5.4.3. Les redressements

Pour réaliser la pondération sur le champ du ministère de la Santé et des Sports, nous partons des coefficients issus de la pondération non calée réalisée par le Céreq sur l'ensemble des répondants, incluant les individus hors du champ Céreq de l'enquête.

Nous restreignons ensuite le champ des formations du sport et de l'animation étudiées par le Ministère de la Santé et des Sports et réalisons le calage à partir des marges issues de la base de sondage du Céreq.

Le calage sur marges se rapporte aux effectifs bruts et non aux pourcentages (différence sensible pour les marges faibles). Il s'effectue à partir de l'ensemble des individus répondants, dans le champ de l'extension du Sport.

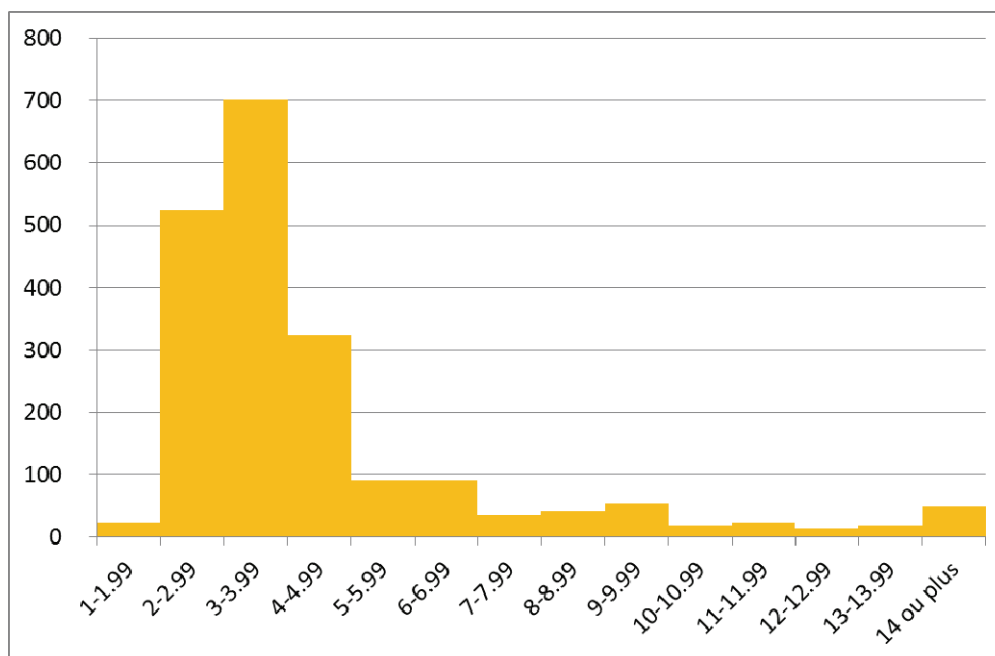
Tableau 32

Marges de calage, poids moyens des individus et estimation du nombre de sortants dans le champ du sport

Diplôme	Marge	Nombre de répondants dans le champ	Poids moyen	Estimation nb d'individus dans le champ
BEES premier degré	6 716	956	4,99	4 772
Brevet professionnel spécialité animation	1 891	377	4,08	1 538
Brevet professionnel spécialité sportive	1 996	354	4,19	1 482
Autre diplôme du sport et de l'animation	1 947	318	4,57	1 453
Total	12 550	2 005	4,61	9 245

Figure 15

Distribution des poids de réponse des individus répondants dans le champ du sport



5.5. L'extension DGAFP

5.5.1. Le contexte

La Direction générale de l'administration et de la fonction publique (DGAFP) a souhaité réaliser, de manière expérimentale, une extension d'échantillon sur les sortants des écoles de la fonction publique. A partir de la liste des établissements concernés⁷ transmise au Céreq par la DGAFP, le Céreq a sollicité 41 établissements.

Un total de 22 établissements ont finalement envoyé les informations concernant leurs élèves. La DGAFP souhaitant obtenir le maximum de questionnaires, tous les individus de ces établissements ont été échantillonnés.

Un nombre important d'écoles (19 sur 41), dont de nombreuses écoles de police ou des écoles de très haut niveau (Mines, ENSAE, télécom) n'ont pas donné une suite favorable à la collecte. On ne dispose d'aucune information sur les effectifs de sortants de ces écoles.

⁷ La liste transmise par la DGAFP concernait 42 établissements mais suite à un problème technique, l'école nationale des télécommunications n'a pas été sollicitée par le Céreq.

Tableau 33

Liste des écoles de l'extension DGAFP ayant transmis leurs informations au Céreq

Matricule de l'établissement	Nom de l'établissement	Etablissement présent dans la base de sondage de Génération 2004	Type d'école	Effectifs inscrits dans les fichiers transmis
069F	Centres de formation de la police nationale de Chassieu	NON		195
031F	Centres de formation de la police nationale Toulouse	NON	FONCTIONNAIRE	90
0672678X	ENA	OUI	FONCTIONNAIRE	90
0310153Y	Ecole nationale des services vétérinaires de Toulouse	OUI	MIXTE	116
0690193K	Ecole nationale des services vétérinaires de Lyon	OUI	MIXTE	79
0332239E	Ecole nationale de la magistrature	NON	FONCTIONNAIRE	277
092F	Ecole nationale de la statistique et de l'analyse de l'information	NON	MIXTE	111
0211660R	Ecole nationale des greffes	NON	FONCTIONNAIRE	268
0630107Z	Ecole nationale des impôts	NON	MIXTE	1019
0772517T	Ecole nationale des Ponts et Chaussées	NON	MIXTE	239
0772496V	Ecole nationale des sciences géographiques	NON	MIXTE	108
075F	Ecole nationale des techniciens de l'équipement	NON	FONCTIONNAIRE	108
0692566P	Ecole nationale des travaux publics de l'Etat	NON	MIXTE	200
0441916U	Ecole nationale des services vétérinaires de Nantes	OUI	MIXTE	97
0940608A	Ecole nationale des services vétérinaires d'Alfort	OUI	MIXTE	139
0350095N	Ecole des hautes études en santé publique	OUI	FONCTIONNAIRE	344
0670189S	Ecole nationale du génie de l'eau et de l'environnement de Strasbourg.	OUI	MIXTE	128
7200090K	IRA Bastia	OUI	FONCTIONNAIRE	131
0441618V	IRA Nantes.	OUI	FONCTIONNAIRE	132
0572647T	IRA Metz	OUI	FONCTIONNAIRE	137
0691861Y	IRA Lyon	OUI	FONCTIONNAIRE	121
069F	Institut national du travail de l'emploi et de la formation professionnelle (INTEFP)	NON	FONCTIONNAIRE	65
EFFECTIF TOTAL COLLECTE				4 194

Tableau 34

Liste des écoles de l'extension DGAFP n'ayant pas transmis leurs informations au Céreq

Matricule de l'établissement	Nom de l'établissement	Etablissement dans la base de sondage de l'enquête Génération 2004
0311256X	Ecole nationale de l'aviation civile	NON
031F	Ecole nationale d'application de la police nationale	NON
0420094Z	Ecole nationale supérieure des mines de Saint-Etienne	NON
0470953F	Ecole nationale d'administration pénitentiaire	NON
0540128W	Ecole nationale supérieure des mines de Nancy	OUI
0593648V	IRA Lille	OUI
0596695G	Ecole nationale des douanes	NON
059F	Ecole nationale de la protection judiciaire de la jeunesse	NON
0690189F	Ecole nationale supérieure de police	NON
0753493P	Ecole nationale supérieure des mines de Paris	NON
0753503A	Ecole nationale du génie rural, des eaux et forêts	OUI
0755026F	Institut national du patrimoine	NON
0755222U	Ecole de Chaillot	OUI
075F	Centre de formation de l'INSEE à Libourne	NON
0771939P	Ecole nationale du Trésor public	NON
0772316Z	Ecole nationale supérieure des officiers de police	NON
0921682D	Ecole nationale de la statistique et de l'administration économique	NON
093F	Ecole nationale de la concurrence, de la consommation et de la répression des fraudes	NON
098F	Centres de formation de la police nationale Calédonie	NON
	Ecole nationale supérieure des télécommunications	NON

5.5.2. Les résultats de l'enquête

Parmi les 4 194 individus de la DGAFP échantillonnés, 1 007 (24 %) ont accepté de répondre à l'enquête. Parmi eux, 164 se sont révélés hors champ à l'issue du questionnaire filtre. Un total de 843 questionnaires a finalement été collecté à l'issue de l'enquête.

Tableau 35

Taux de réponse pour les 21 établissements envoyés en enquête

Matricule de l'établissement	Nom de l'établissement	Nombre d'individus dans l'échantillon	Nombre de répondants		Taux de réponse
			dans le champ DGAFP	hors champ	
0211660R	Ecole nationale des greffes	268	114	6	45%
0310153Y	Ecole nationale des services vétérinaires de Toulouse	116	27	37	55%
031F	Centres de formation de la police nationale Toulouse	90	4	0	4%
0332239E	Ecole nationale de la magistrature	277	4	0	1%
0350095N	Ecole des hautes études en santé publique	344	87	7	27%
0441618V	IRA Nantes.	132	54	0	41%
0441916U	Ecole nationale des services vétérinaires de Nantes	97	39	25	66%
0572647T	IRA Metz	137	38	2	29%
0630107Z	Ecole nationale des impôts	1 019	48	5	5%
0670189S	Ecole nationale du génie de l'eau et de l'environnement de Strasbourg.	128	43	0	34%
0672678X	ENA	90	52	0	58%
0690193K	Ecole nationale des services vétérinaires de Lyon	79	19	25	56%
0691861Y	IRA Lyon	121	42	0	35%
0692566P	Ecole nationale des travaux publics de l'Etat	200	54	3	29%
069F	Centres de formation de la police nationale de Chassieu	195	32	1	17%
069F	Institut national du travail de l'emploi et de la formation professionnelle (INTEFP)	65	3	0	5%
075F	Ecole nationale des techniciens de l'équipement	108	14	16	28%
0772496V	Ecole nationale des sciences géographiques	108	25	2	25%
0772517T	Ecole nationale des Ponts et Chaussées	239	71	17	37%
092F	Ecole nationale de la statistique et de l'analyse de l'information	111	38	5	39%
0940608A	Ecole nationale des services vétérinaires d'Alfort	139	12	11	17%
7200090K	IRA Bastia	131	23	2	19%
TOTAL		4 194	843	164	24%

5.5.3. Les redressements

En l'absence d'information sur les volumes d'inscrits dans les 23 écoles n'ayant pas transmis d'information au Céreq, le calcul du taux de couverture sur l'ensemble des écoles de la fonction publique n'est pas possible. Par ailleurs, rien ne permet d'affirmer que les sortants des écoles non répondantes présentent des caractéristiques similaires à ceux des écoles répondantes.

Dans ces conditions, il n'est pas envisageable de procéder à un redressement général pour les écoles de la fonction publique. On restreint donc le champ aux écoles de la fonction publique ayant transmis des informations au Céreq, soit 22 écoles.

On réalise pour l'extension DGAFP une pondération ad hoc, complètement indépendante de la pondération des 30 000. Elle ne tient pas compte des poids individuels avant calage et s'appuie sur un simple calage aux marges des répondants.

Le calage aux marges au niveau établissement conduirait à une répartition des poids très déséquilibrée car les taux de réponse varient fortement d'un établissement à l'autre. On constate en particulier un très faible taux de réponse des individus issus de certains établissements (centre de formation de la police nationale de Toulouse, école nationale de la magistrature,...).

Les marges de calage obtenues grâce à un regroupement d'établissements ont permis de garantir un volume de répondants suffisant pour chaque marge. Les regroupements d'établissements tiennent compte des taux de réponse, de la proportion de hors champ, du type d'école (fonctionnaire ou mixte, domaine,...), des caractéristiques d'insertion des répondants dans le champ (niveau de rémunération, catégorie socioprofessionnelle).

Les regroupements suivants ont été retenus :

- Groupe 1 : écoles de vétérinaires (mixtes) : 0310153Y, 0441916U, 0940608A, 0690193K
- Groupe 2 : IRA (établissements fonctionnaires sauf une mixte avec 3 répondants non fonctionnaires) : 0441618V, 0572647T, 7200090K, 0691861Y
- Groupe 3 : Centres de formation de police (un établissement fonctionnaire, un mixte sans répondant non fonctionnaire) : 031F04, 069F02
- Groupe 4 : Ponts et chaussée, travaux publics et école de l'eau de Strasbourg, sciences géographiques (4 établissements mixtes) : 0772517T, 0692566P, 0670189S, 0772496V
- Groupe 5 : école des techniciens de l'électricité : 075F01
- Groupe 6 : ENA, magistrature, EHESP, travail et emploi (1 établissement mixte sans répondant non fonctionnaire, 4 établissements fonctionnaires) : 0672678X, 0332239E, 0350095N, 069F11
- Groupe 7 : ENSAI : 092F07
- Groupe 8 : Greffes : 0211660R
- Groupe 9 : Impôts : 0630107Z

Tableau 36

Taux de réponse par groupe d'établissements envoyés en enquête

Groupe	Nombre d'individus dans l'échantillon	Nombre de répondants		Taux de réponse
		dans le champ DGAFP	hors champ DGAFP	
1	431	97	98	45 %
2	521	157	4	31 %
3	285	36	1	13 %
4	675	193	22	32 %
5	108	14	16	28 %
6	776	146	7	20 %
7	111	38	5	39 %
8	268	114	6	45 %
9	1 019	48	5	5 %
Total	4 194	843	164	24 %

Le calage aux marges s'effectue par rapport aux effectifs présents dans la base de sondage, à partir des individus répondants, y.c. les hors-champs.

Tableau 37

Marges de calage, poids des individus et estimation du nombre de sortants dans le champ

Groupe	Marge de calage	Nombre de répondants dans le champ	Poids d'un répondant dans le champ	Estimation du nombre d'individus dans le champ
1	431	97	2,2	215
2	521	157	3,2	507
3	285	36	7,7	277
4	675	193	3,1	606
5	108	14	3,6	51
6	776	146	5,1	740
7	111	38	2,5	96
8	268	114	2,2	255
9	1 019	48	19,2	923
Total	4 194	843		3 671

Les répondants dans le champ ont un poids moyen de 4,4 (min=2,2 – max=19,2, écart-type=3,9).

Pour les 22 écoles ayant transmis des informations au Céreq, le nombre de sortants dans le champ de l'enquête est estimé à 3 671 individus. Les résultats de l'enquête sont exploitables au niveau global mais pas au niveau établissement.

ANNEXES

La modélisation de l'envoi en enquête

Le programme SAS mis en œuvre est le suivant :

```
proc logistic data=pond.var_coef4 descending outest=mod4;  
class no_adress specialite no_cp region type_zus SEXE strate07;  
output out=envenq pred=invcoef4;  
model envoye=no_adress specialite no_cp region type_zus SEXE strate07/rsquare;  
run;
```

Les sorties associées sont présentées ci-dessous.

Statistiques d'ajustement du modèle		
Critère	Coordonnée à l'origine uniquement	Coordonnée à l'origine et covariables
AIC	150495.09	128740.64
SC	150505.26	129309.88
-2 Log L	150493.09	128628.64

R-Square	0.1076	Max-rescaled R-Square	0.1981
----------	--------	-----------------------	--------

Test de l'hypothèse nulle globale : BETA=0			
Test	Khi 2	DF	Pr > Khi 2
Likelihood Ratio	21864.4541	55	<.0001
Score	25291.2700	55	<.0001
Wald	18499.7046	55	<.0001

Analyse des effets Type 3			
Effet	DF	Khi 2 de Wald	Pr > Khi 2
no_adress	1	75.7918	<.0001
specialite	2	253.8551	<.0001
no_cp	1	3082.5783	<.0001

Analyse des effets Type 3			
Effet	DF	Khi 2 de Wald	Pr > Khi 2
region	21	605.3380	<.0001
type_zus	2	1445.0545	<.0001
SEXE	1	79.2165	<.0001
strate07	27	9657.3931	<.0001

Estimations des rapports de cotes			
Effet	Point Estimate	95% Limites de confiance de Wald	
no_adress 0 vs 1	1.538	1.396	1.694
specialite 1 vs 3	1.587	1.447	1.741
specialite 2 vs 3	1.924	1.767	2.095
no_cp 0 vs 1	6.244	5.853	6.662
region 11 vs 94	0.823	0.585	1.158
region 21 vs 94	0.964	0.669	1.390
region 22 vs 94	0.687	0.482	0.979
region 23 vs 94	0.926	0.649	1.320
region 24 vs 94	0.805	0.566	1.145
region 25 vs 94	1.042	0.726	1.496
region 26 vs 94	0.998	0.695	1.433
region 31 vs 94	0.773	0.549	1.089
region 41 vs 94	1.259	0.880	1.801
region 42 vs 94	0.661	0.465	0.939
region 43 vs 94	1.243	0.881	1.754
region 52 vs 94	0.901	0.633	1.282
region 53 vs 94	0.907	0.644	1.278
region 54 vs 94	1.125	0.798	1.587
region 72 vs 94	0.853	0.606	1.202
region 73 vs 94	1.263	0.887	1.799
region 74 vs 94	0.927	0.635	1.352
region 82 vs 94	0.903	0.640	1.273
region 83 vs 94	0.568	0.399	0.808
region 91 vs 94	0.783	0.553	1.109
region 93 vs 94	0.711	0.504	1.003
type_zus 1 vs 5	2.156	2.067	2.250
type_zus 2 vs 5	1.355	1.275	1.439
SEXE 1 vs 2	1.151	1.116	1.188
strate07 01 vs 28	0.162	0.148	0.177
strate07 02 vs 28	0.267	0.231	0.309

Estimations des rapports de cotes			
Effet	Point Estimate	95% Limites de confiance de Wald	
strate07 03 vs 28	0.359	0.315	0.409
strate07 04 vs 28	1.146	0.977	1.343
strate07 05 vs 28	1.911	1.604	2.277
strate07 06 vs 28	1.936	1.630	2.299
strate07 07 vs 28	0.373	0.323	0.429
strate07 08 vs 28	0.387	0.342	0.439
strate07 09 vs 28	0.180	0.165	0.198
strate07 10 vs 28	1.862	1.561	2.220
strate07 11 vs 28	0.319	0.275	0.371
strate07 12 vs 28	1.465	1.326	1.619
strate07 13 vs 28	0.575	0.503	0.657
strate07 14 vs 28	1.372	1.195	1.574
strate07 15 vs 28	0.640	0.573	0.714
strate07 16 vs 28	0.514	0.471	0.561
strate07 17 vs 28	1.915	1.641	2.235
strate07 18 vs 28	1.790	1.568	2.044
strate07 19 vs 28	13.035	7.673	22.145
strate07 20 vs 28	1.896	1.491	2.411
strate07 21 vs 28	0.641	0.572	0.718
strate07 22 vs 28	0.788	0.704	0.881
strate07 23 vs 28	0.923	0.817	1.042
strate07 24 vs 28	0.616	0.524	0.723
strate07 25 vs 28	0.541	0.460	0.635
strate07 26 vs 28	1.156	1.022	1.306
strate07 27 vs 28	0.855	0.752	0.972

Association des probabilités prédites et des réponses observées			
Percent Concordant	76.7	Somers' D	0.540
Percent Discordant	22.7	Gamma	0.543
Percent Tied	0.6	Tau-a	0.124
Pairs	4248316656	c	0.770

La modélisation de la probabilité de réponse suite à l'envoi en enquête

Région Aquitaine :

```
proc logistic data=mod descending outest=mod5aqui;
where region='72';
title "Modélisation non réponse Aquitaine";
output out=aqui pred=invcoef5;
class zus prequal tel_rec specialite portable strate07 prequal ;
model repondant = zus prequal tel_rec specialite portable strate07 prequal
/rsquare;run;
```

Statistiques d'ajustement du modèle		
Critère	Coordonnée à l'origine uniquement	Coordonnée à l'origine et covariables
AIC	20522.658	19779.059
SC	20530.398	20057.683
-2 Log L	20520.658	19707.059

R-Square	0.0468	Max-rescaled R-Square	0.0667
-----------------	--------	------------------------------	--------

Test de l'hypothèse nulle globale : BETA=0			
Test	Khi 2	DF	Pr > Khi 2
Likelihood Ratio	813.5992	35	<.0001
Score	785.3582	35	<.0001
Wald	741.5641	35	<.0001

Analyse des effets Type 3			
Effet	DF	Khi 2 de Wald	Pr > Khi 2
ZUS	2	49.2636	<.0001
prequal	1	89.5960	<.0001
tel_rec	2	232.0412	<.0001
specialite	2	5.6091	0.0605
PORTABLE	1	50.7185	<.0001
strate07	27	236.1461	<.0001

Estimations des rapports de cotes			
Effet	Point Estimate	95% Limites de confiance de Wald	
ZUS 1 vs 3	0.916	0.741	1.132
ZUS 2 vs 3	1.401	1.248	1.573
prequal 1 vs 2	0.505	0.438	0.582
tel_rec Tel BS seule vs rech élargie	1.098	1.002	1.204
tel_rec rech adresse vs rech élargie	1.817	1.676	1.971
specialite 1 vs 3	1.251	0.983	1.592
specialite 2 vs 3	1.176	0.991	1.395
PORTABLE 0 vs 1	0.730	0.669	0.796
strate07 01 vs 28	2.558	1.940	3.374
strate07 02 vs 28	1.575	1.071	2.317
strate07 03 vs 28	1.609	1.108	2.337
strate07 04 vs 28	1.036	0.718	1.495
strate07 05 vs 28	0.849	0.582	1.237
strate07 06 vs 28	1.157	0.803	1.666
strate07 07 vs 28	1.714	1.167	2.518
strate07 08 vs 28	1.648	1.158	2.347
strate07 09 vs 28	2.892	2.138	3.911
strate07 10 vs 28	1.196	0.850	1.683
strate07 11 vs 28	1.631	1.053	2.526
strate07 12 vs 28	1.192	0.942	1.507
strate07 13 vs 28	1.973	1.383	2.814
strate07 14 vs 28	1.345	1.028	1.761
strate07 15 vs 28	1.062	0.779	1.450
strate07 16 vs 28	0.909	0.625	1.322
strate07 17 vs 28	1.794	1.217	2.645
strate07 18 vs 28	1.780	1.255	2.525
strate07 19 vs 28	2.750	1.982	3.814
strate07 20 vs 28	2.006	1.435	2.802
strate07 21 vs 28	1.660	1.214	2.271
strate07 22 vs 28	1.323	0.984	1.779
strate07 23 vs 28	1.480	1.033	2.119
strate07 24 vs 28	1.483	0.972	2.263
strate07 25 vs 28	2.429	1.522	3.876
strate07 26 vs 28	1.329	0.997	1.772
strate07 27 vs 28	1.516	1.132	2.029

Association des probabilités prédites et des réponses observées			
Percent Concordant	62.8	Somers' D	0.267
Percent Discordant	36.1	Gamma	0.270
Percent Tied	1.1	Tau-a	0.111
Pairs	59641660	c	0.634

Région Bretagne :

```
proc logistic data=mod descending outest=mod5bret;
where region='53';
title "Modélisation non réponse Bretagne";
output out=bret pred=invcoef5;
class zus cfa prequal tel_rec specialite portable special strate07 prequal ;
model repondant = zus cfa prequal tel_rec specialite portable special strate07
prequal /rsquare;run;
```

Statistiques d'ajustement du modèle		
Critère	Coordonnée à l'origine uniquement	Coordonnée à l'origine et covariables
AIC	22094.993	21206.367
SC	22102.767	21501.796
-2 Log L	22092.993	21130.367

R-Square	0.0533	Max-rescaled R-Square	0.0745
----------	--------	-----------------------	--------

Test de l'hypothèse nulle globale : BETA=0			
Test	Khi 2	DF	Pr > Khi 2
Likelihood Ratio	962.6261	37	<.0001
Score	914.3695	37	<.0001
Wald	860.9486	37	<.0001

Analyse des effets Type 3			
Effet	DF	Khi 2 de Wald	Pr > Khi 2
ZUS	2	32.4149	<.0001
CFA	1	2.8294	0.0926
prequal	1	64.8112	<.0001
tel_rec	2	286.7425	<.0001
specialite	2	24.5439	<.0001
PORTABLE	1	59.5598	<.0001
SPECIAL	1	6.7421	0.0094
strate07	27	176.2983	<.0001

Estimations des rapports de cotes			
Effet	Point Estimate	95% Limites de confiance de Wald	
ZUS 1 vs 3	0.755	0.594	0.959
ZUS 2 vs 3	1.240	1.111	1.384
CFA 1 vs 2	0.492	0.215	1.124
prequal 1 vs 2	0.565	0.491	0.649
tel_rec Tel BS seule vs rech élargie	1.138	1.025	1.264
tel_rec rech adresse vs rech élargie	1.916	1.766	2.080
specialite 1 vs 3	1.894	1.471	2.440
specialite 2 vs 3	1.123	0.961	1.313
PORTABLE 0 vs 1	0.704	0.644	0.769
SPECIAL 0 vs 1	1.134	1.031	1.247
strate07 01 vs 28	3.165	2.323	4.311
strate07 02 vs 28	2.636	1.769	3.930
strate07 03 vs 28	2.504	1.678	3.738
strate07 04 vs 28	3.572	1.436	8.884
strate07 05 vs 28	2.505	0.986	6.366
strate07 06 vs 28	3.484	1.394	8.708
strate07 07 vs 28	3.072	2.069	4.561
strate07 08 vs 28	2.713	1.850	3.977
strate07 09 vs 28	3.124	2.246	4.346
strate07 10 vs 28	3.981	1.603	9.888
strate07 11 vs 28	3.784	2.447	5.851
strate07 12 vs 28	1.390	1.053	1.834
strate07 13 vs 28	2.738	1.876	3.995
strate07 14 vs 28	1.670	1.233	2.262
strate07 15 vs 28	1.603	1.164	2.209

Estimations des rapports de cotes			
Effet	Point Estimate	95% Limites de confiance de Wald	
strate07 16 vs 28	1.114	0.795	1.559
strate07 17 vs 28	2.745	1.828	4.122
strate07 18 vs 28	2.611	1.767	3.858
strate07 19 vs 28	2.002	1.400	2.861
strate07 20 vs 28	1.968	1.306	2.963
strate07 21 vs 28	1.965	1.430	2.700
strate07 22 vs 28	1.716	1.235	2.384
strate07 23 vs 28	2.642	1.760	3.967
strate07 24 vs 28	2.195	1.332	3.618
strate07 25 vs 28	4.913	3.277	7.366
strate07 26 vs 28	1.938	1.411	2.662
strate07 27 vs 28	1.807	1.268	2.577

Association des probabilités prédites et des réponses observées			
Percent Concordant	63.4	Somers' D	0.279
Percent Discordant	35.5	Gamma	0.282
Percent Tied	1.1	Tau-a	0.122
Pairs	67467798	c	0.640

Région Franche-Comté :

```
proc logistic data=mod descending outest=mod5fc;
where region='43';
title "Modélisation non réponse Franche Comté";
output out=fc pred=invcoef5;
class zus sexe prequal tel_rec special strate07 prequal ;
model repondant = zus sexe prequal tel_rec special strate07 prequal
/rsquare;run;
```

Statistiques d'ajustement du modèle		
Critère	Coordonnée à l'origine	
	uniquement	et covariables
AIC	17022.311	16483.069
SC	17029.827	16746.137
-2 Log L	17020.311	16413.069

R-Square	0.0437	Max-rescaled R-Square	0.0612
----------	--------	-----------------------	--------

Test de l'hypothèse nulle globale : BETA=0			
Test	Khi 2	DF	Pr > Khi 2
Likelihood Ratio	607.2416	34	<.0001
Score	589.9815	34	<.0001
Wald	565.0800	34	<.0001

Analyse des effets Type 3			
Effet	DF	Khi 2 de Wald	Pr > Khi 2
ZUS	2	23.7124	<.0001
SEXE	1	8.4660	0.0036
prequal	1	32.6077	<.0001
tel_rec	2	198.1868	<.0001
SPECIAL	1	5.6301	0.0177
strate07	27	229.6536	<.0001

Estimations des rapports de cotes			
Effet	Point Estimate	95% Limites de confiance de Wald	
ZUS 1 vs 3	0.844	0.669	1.066
ZUS 2 vs 3	1.242	1.053	1.465
SEXE 1 vs 2	1.128	1.040	1.223
prequal 1 vs 2	0.640	0.549	0.746
tel_rec Tel BS seule vs rech élargie	1.078	0.965	1.205
tel_rec rech adresse vs rech élargie	1.800	1.642	1.973
SPECIAL 0 vs 1	1.142	1.023	1.274
strate07 01 vs 28	2.259	1.455	3.509
strate07 02 vs 28	1.640	1.047	2.571
strate07 03 vs 28	1.471	0.949	2.279
strate07 04 vs 28	1.195	0.771	1.853
strate07 05 vs 28	0.788	0.502	1.238
strate07 06 vs 28	1.786	1.151	2.770
strate07 07 vs 28	2.119	1.355	3.314
strate07 08 vs 28	1.649	1.071	2.540
strate07 09 vs 28	2.973	1.890	4.677
strate07 10 vs 28	1.283	0.827	1.990
strate07 11 vs 28	1.866	1.126	3.095
strate07 12 vs 28	1.800	1.175	2.758
strate07 13 vs 28	1.824	1.182	2.815
strate07 14 vs 28	1.819	1.153	2.871
strate07 15 vs 28	1.836	1.137	2.967
strate07 16 vs 28	1.395	0.842	2.312
strate07 17 vs 28	1.356	0.794	2.317
strate07 18 vs 28	1.275	0.776	2.094
strate07 19 vs 28	4.286	2.597	7.074
strate07 20 vs 28	2.453	1.208	4.983
strate07 21 vs 28	2.433	1.530	3.867
strate07 22 vs 28	1.712	1.087	2.697
strate07 23 vs 28	2.130	1.209	3.754
strate07 24 vs 28	2.365	1.266	4.420
strate07 25 vs 28	0.972	0.567	1.666
strate07 26 vs 28	2.084	1.301	3.339
strate07 27 vs 28	1.158	0.713	1.883

Association des probabilités prédites et des réponses observées			
Percent Concordant	62.3	Somers' D	0.255
Percent Discordant	36.7	Gamma	0.258
Percent Tied	1.0	Tau-a	0.111
Pairs	40107605	c	0.628

Région Ile de France :

```
proc logistic data=mod descending outest=mod5idf;
where region ='11';
title "Modélisation non réponse IdF";
output out=idf pred=invcoef5;
class zus cfa sexe prequal tel_rec specialite portable strate07 prequal;
model repondant = zus cfa sexe prequal tel_rec specialite portable strate07
prequal/rsquare;run;
```

Statistiques d'ajustement du modèle		
Critère	Coordonnée à l'origine uniquement	Coordonnée à l'origine et covariables
AIC	22585.297	21493.101
SC	22593.234	21794.699
-2 Log L	22583.297	21417.101

R-Square	0.0548	Max-rescaled R-Square	0.0825
----------	--------	-----------------------	--------

Test de l'hypothèse nulle globale : BETA=0			
Test	Khi 2	DF	Pr > Khi 2
Likelihood Ratio	1166.1957	37	<.0001
Score	1063.6022	37	<.0001
Wald	955.7752	37	<.0001

Analyse des effets Type 3			
Effet	DF	Khi 2 de Wald	Pr > Khi 2
ZUS	2	13.7069	0.0011
CFA	1	3.9670	0.0464
SEXE	1	2.8947	0.0889

Analyse des effets Type 3			
Effet	DF	Khi 2 de Wald	Pr > Khi 2
prequal	1	197.8553	<.0001
tel_rec	2	172.7202	<.0001
specialite	2	4.6503	0.0978
PORTABLE	1	83.3843	<.0001
strate07	27	500.5130	<.0001

Estimations des rapports de cotes				
Effet		Point Estimate	95% Limites de confiance de Wald	
ZUS	1 vs 3	0.888	0.778	1.013
ZUS	2 vs 3	1.070	0.967	1.183
CFA	1 vs 2	1.306	1.004	1.699
SEXE	1 vs 2	1.065	0.991	1.144
prequal	1 vs 2	0.273	0.228	0.327
tel_rec	Tel BS seule vs rech élargie	1.120	1.033	1.213
tel_rec	rech adresse vs rech élargie	1.737	1.595	1.891
specialite	1 vs 3	1.245	0.967	1.603
specialite	2 vs 3	1.208	0.966	1.511
PORTABLE	0 vs 1	0.687	0.633	0.744
strate07	01 vs 28	4.678	3.509	6.236
strate07	02 vs 28	2.095	1.344	3.265
strate07	03 vs 28	3.845	2.632	5.617
strate07	04 vs 28	1.733	1.070	2.810
strate07	05 vs 28	1.516	0.912	2.521
strate07	06 vs 28	2.563	1.642	3.999
strate07	07 vs 28	3.756	2.619	5.385
strate07	08 vs 28	3.315	2.386	4.605
strate07	09 vs 28	11.015	8.278	14.656
strate07	10 vs 28	3.386	2.148	5.335
strate07	11 vs 28	3.250	2.137	4.941
strate07	12 vs 28	2.624	2.251	3.058
strate07	13 vs 28	3.069	2.129	4.425
strate07	14 vs 28	2.663	2.140	3.314
strate07	15 vs 28	2.463	2.020	3.003
strate07	16 vs 28	2.095	1.703	2.576
strate07	17 vs 28	3.319	2.433	4.527
strate07	18 vs 28	2.701	2.000	3.649
strate07	19 vs 28	5.630	4.382	7.233

Estimations des rapports de cotes			
Effet	Point Estimate	95% Limites de confiance de Wald	
strate07 20 vs 28	3.824	2.693	5.430
strate07 21 vs 28	2.688	2.156	3.350
strate07 22 vs 28	2.055	1.704	2.478
strate07 23 vs 28	3.761	3.016	4.689
strate07 24 vs 28	3.369	2.262	5.018
strate07 25 vs 28	2.691	1.813	3.994
strate07 26 vs 28	2.504	2.075	3.022
strate07 27 vs 28	2.310	1.902	2.806

Association des probabilités prédites et des réponses observées			
Percent Concordant	64.5	Somers' D	0.301
Percent Discordant	34.4	Gamma	0.304
Percent Tied	1.1	Tau-a	0.108
Pairs	77023822	c	0.650

Région Nord-Pas-de-Calais :

```
proc logistic data=mod descending outest=mod5npdc;
where region='31';
title "Modélisation non réponse NPDC";
output out=npdc pred=invcoef5;
class zus prequal tel_rec specialite portable strate07 prequal ;
model repondant = zus prequal tel_rec specialite portable strate07 prequal
/rsquare;run;
```

Statistiques d'ajustement du modèle		
Critère	Coordonnée à l'origine uniquement	Coordonnée à l'origine et covariables
AIC	22363.400	21295.497
SC	22371.256	21570.462
-2 Log L	22361.400	21225.497

R-Square	0.0578	Max-rescaled R-Square	0.0837
----------	--------	-----------------------	--------

Test de l'hypothèse nulle globale : BETA=0			
Test	Khi 2	DF	Pr > Khi 2
Likelihood Ratio	1135.9033	34	<.0001
Score	1069.4270	34	<.0001
Wald	986.8414	34	<.0001

Analyse des effets Type 3			
Effet	DF	Khi 2 de Wald	Pr > Khi 2
ZUS	2	23.8084	<.0001
prequal	1	155.8267	<.0001
tel_rec	2	423.1573	<.0001
specialite	2	5.0897	0.0785
PORTABLE	1	33.4329	<.0001
strate07	26	184.5409	<.0001

Estimations des rapports de cotes				
Effet		Point Estimate	95% Limites de confiance de Wald	
ZUS	1 vs 3	1.042	0.895	1.212
ZUS	2 vs 3	1.257	1.124	1.406
prequal	1 vs 2	0.393	0.340	0.455
tel_rec	Tel BS seule vs rech élargie	1.390	1.273	1.518
tel_rec	rech adresse vs rech élargie	2.291	2.116	2.481
specialite	1 vs 3	1.364	1.024	1.816
specialite	2 vs 3	1.115	0.938	1.325
PORTABLE	0 vs 1	0.790	0.729	0.856
strate07	01 vs 28	3.281	2.423	4.442
strate07	02 vs 28	2.045	1.334	3.135
strate07	03 vs 28	2.028	1.332	3.088
strate07	04 vs 28	1.554	1.031	2.342
strate07	05 vs 28	1.644	1.095	2.468
strate07	06 vs 28	1.871	1.251	2.798
strate07	07 vs 28	2.470	1.625	3.754
strate07	08 vs 28	2.017	1.348	3.019
strate07	09 vs 28	4.428	3.219	6.093
strate07	10 vs 28	1.867	1.253	2.781
strate07	11 vs 28	2.989	1.883	4.746
strate07	12 vs 28	1.783	1.373	2.316
strate07	13 vs 28	2.778	1.866	4.134

Estimations des rapports de cotes			
Effet	Point Estimate	95% Limites de confiance de Wald	
strate07 14 vs 28	1.462	1.085	1.971
strate07 15 vs 28	1.785	1.314	2.425
strate07 16 vs 28	2.208	1.633	2.987
strate07 17 vs 28	2.664	1.775	4.000
strate07 18 vs 28	2.366	1.597	3.505
strate07 20 vs 28	2.650	1.776	3.956
strate07 21 vs 28	2.298	1.684	3.135
strate07 22 vs 28	1.675	1.215	2.308
strate07 23 vs 28	1.735	1.153	2.610
strate07 24 vs 28	1.435	0.882	2.334
strate07 25 vs 28	1.951	1.203	3.163
strate07 26 vs 28	1.922	1.436	2.574
strate07 27 vs 28	2.377	1.707	3.310

Association des probabilités prédites et des réponses observées			
Percent Concordant	64.4	Somers' D	0.299
Percent Discordant	34.4	Gamma	0.303
Percent Tied	1.2	Tau-a	0.119
Pairs	72202014	c	0.650

Région Poitou-Charentes :

```
proc logistic data=mod descending outest=mod5pc;
where region='54';
title "Modélisation non réponse Poitou Charente";
output out=pc pred=invcoef5;
class zus prequal tel_rec specialite portable strate07 prequal ;
model repondant = zus prequal tel_rec specialite portable strate07 prequal
/rsquare;run;
```

Statistiques d'ajustement du modèle		
Critère	Coordonnée à l'origine uniquement	Coordonnée à l'origine et covariables
AIC	17683.525	17104.559
SC	17691.073	17376.265
-2 Log L	17681.525	17032.559

R-Square	0.0453	Max-rescaled R-Square	0.0631
-----------------	--------	------------------------------	--------

Test de l'hypothèse nulle globale : BETA=0			
Test	Khi 2	DF	Pr > Khi 2
Likelihood Ratio	648.9662	35	<.0001
Score	637.7557	35	<.0001
Wald	609.5629	35	<.0001

Analyse des effets Type 3			
Effet	DF	Khi 2 de Wald	Pr > Khi 2
ZUS	2	14.7827	0.0006
prequal	1	31.3016	<.0001
tel_rec	2	263.8440	<.0001
specialite	2	23.7492	<.0001

PORTABLE	1	18.3590	<.0001
strate07	27	119.7838	<.0001

Estimations des rapports de cotes			
Effet	Point Estimate	95% Limites de confiance de Wald	
ZUS 1 vs 3	0.926	0.724	1.185
ZUS 2 vs 3	1.226	1.075	1.397
prequal 1 vs 2	0.631	0.537	0.742
tel_rec Tel BS seule vs rech élargie	0.946	0.855	1.047
tel_rec rech adresse vs rech élargie	1.836	1.685	2.001
specialite 1 vs 3	1.987	1.463	2.700
specialite 2 vs 3	1.319	1.108	1.570
PORTABLE 0 vs 1	0.804	0.727	0.888
strate07 01 vs 28	1.994	1.395	2.849
strate07 02 vs 28	1.429	0.885	2.306
strate07 03 vs 28	1.484	0.932	2.364
strate07 04 vs 28	1.103	0.699	1.741
strate07 05 vs 28	1.145	0.715	1.836
strate07 06 vs 28	1.299	0.827	2.041
strate07 07 vs 28	1.383	0.857	2.233
strate07 08 vs 28	1.451	0.927	2.273
strate07 09 vs 28	1.586	1.091	2.306
strate07 10 vs 28	1.450	0.915	2.296
strate07 11 vs 28	1.921	1.112	3.320
strate07 12 vs 28	1.138	0.825	1.569
strate07 13 vs 28	2.079	1.322	3.267
strate07 14 vs 28	0.860	0.608	1.217
strate07 15 vs 28	1.194	0.811	1.756
strate07 16 vs 28	1.407	0.980	2.020
strate07 17 vs 28	2.261	1.343	3.808
strate07 18 vs 28	1.590	0.966	2.616
strate07 19 vs 28	1.728	1.139	2.622
strate07 20 vs 28	1.486	0.890	2.480
strate07 21 vs 28	1.630	1.153	2.305
strate07 22 vs 28	1.141	0.778	1.674
strate07 23 vs 28	1.814	0.977	3.368
strate07 24 vs 28	1.667	0.571	4.870
strate07 25 vs 28	1.746	0.994	3.064
strate07 26 vs 28	1.046	0.713	1.535
strate07 27 vs 28	0.885	0.579	1.352

Association des probabilités prédites et des réponses observées			
Percent Concordant	62.0	Somers' D	0.258
Percent Discordant	36.2	Gamma	0.263
Percent Tied	1.9	Tau-a	0.113
Pairs	43097535	c	0.629

Autres régions :

```
proc logistic data=mod descending outest=mod5aut;
where region not in ('11', '31', '43', '53', '54', '72');
title "Modélisation non réponse autres régions";
output out=aut pred=invcoef5;
class region zus prequal tel_rec specialite portable strate07 prequal ;
model repondant = region zus prequal tel_rec specialite portable strate07
prequal /rsquare;run;
```

Statistiques d'ajustement du modèle		
Critère	Coordonnée à l'origine uniquement	Coordonnée à l'origine et covariables
AIC	77925.011	74508.059
SC	77934.086	74970.895
-2 Log L	77923.011	74406.059

R-Square	0.0530	Max-rescaled R-Square	0.0756
----------	--------	-----------------------	--------

Test de l'hypothèse nulle globale : BETA=0			
Test	Khi 2	DF	Pr > Khi 2
Likelihood Ratio	3516.9520	50	<.0001
Score	3314.5135	50	<.0001
Wald	3088.2660	50	<.0001

Analyse des effets Type 3			
Effet	DF	Khi 2 de Wald	Pr > Khi 2
region	15	70.8071	<.0001
ZUS	2	194.5132	<.0001
prequal	1	519.6060	<.0001
tel_rec	2	778.8899	<.0001

Analyse des effets Type 3			
Effet	DF	Khi 2 de Wald	Pr > Khi 2
specialite	2	12.9179	0.0016
PORTABLE	1	337.4992	<.0001
strate07	27	864.8410	<.0001

Estimations des rapports de cotes			
Effet	Point Estimate	95% Limites de confiance de Wald	
region 21 vs 94	0.934	0.750	1.164
region 22 vs 94	0.823	0.664	1.019
region 23 vs 94	0.891	0.721	1.103
region 24 vs 94	0.892	0.724	1.100
region 25 vs 94	0.976	0.789	1.209
region 26 vs 94	0.811	0.654	1.005
region 41 vs 94	0.897	0.727	1.107
region 42 vs 94	0.935	0.757	1.155
region 52 vs 94	1.000	0.813	1.230
region 73 vs 94	0.959	0.778	1.181
region 74 vs 94	1.001	0.800	1.253
region 82 vs 94	0.899	0.734	1.101
region 83 vs 94	0.930	0.747	1.157
region 91 vs 94	0.815	0.663	1.003
region 93 vs 94	0.786	0.640	0.964
ZUS 1 vs 3	0.932	0.853	1.017
ZUS 2 vs 3	1.356	1.283	1.434
prequal 1 vs 2	0.421	0.391	0.454
tel_rec Tel BS seule vs rech élargie	1.163	1.110	1.218
tel_rec rech adresse vs rech élargie	1.801	1.725	1.880
specialite 1 vs 3	1.289	1.106	1.502
specialite 2 vs 3	1.142	1.024	1.274
PORTABLE 0 vs 1	0.689	0.662	0.717
strate07 01 vs 28	4.245	3.671	4.908
strate07 02 vs 28	2.306	1.862	2.855
strate07 03 vs 28	2.963	2.395	3.667
strate07 04 vs 28	1.731	1.417	2.115
strate07 05 vs 28	1.786	1.422	2.242
strate07 06 vs 28	1.977	1.616	2.418
strate07 07 vs 28	3.232	2.637	3.960

Estimations des rapports de cotes			
Effet	Point Estimate	95% Limites de confiance de Wald	
strate07 08 vs 28	2.461	2.000	3.030
strate07 09 vs 28	5.229	4.391	6.227
strate07 10 vs 28	2.004	1.614	2.487
strate07 11 vs 28	2.509	1.930	3.262
strate07 12 vs 28	1.808	1.629	2.006
strate07 13 vs 28	2.430	1.986	2.973
strate07 14 vs 28	1.748	1.534	1.993
strate07 15 vs 28	1.766	1.535	2.032
strate07 16 vs 28	1.984	1.791	2.198
strate07 17 vs 28	2.423	2.003	2.933
strate07 18 vs 28	2.223	1.853	2.665
strate07 19 vs 28	3.666	3.125	4.299
strate07 20 vs 28	3.025	2.597	3.523
strate07 21 vs 28	2.354	2.048	2.706
strate07 22 vs 28	1.708	1.501	1.944
strate07 23 vs 28	3.325	2.808	3.938
strate07 24 vs 28	1.838	1.448	2.333
strate07 25 vs 28	2.855	2.280	3.576
strate07 26 vs 28	2.235	1.981	2.520
strate07 27 vs 28	1.798	1.549	2.086

Association des probabilités prédites et des réponses observées			
Percent Concordant	63.8	Somers' D	0.282
Percent Discordant	35.6	Gamma	0.284
Percent Tied	0.6	Tau-a	0.117
Pairs	860738144	c	0.641

INDEX DES SIGLES

BCE	Banque centrale des établissements (répertoire d'établissements de formation géré par l'Education nationale)
BEP	Brevet d'études professionnelles
BTS	Brevet de technicien supérieur
CAP	Certificat d'aptitude professionnelle
Céreq	Centre d'études et de recherches sur les qualifications
CFA	Centre de formation d'apprentis
Cifre	Convention industrielle de formation par la recherche
Cnil	Commission nationale informatique et liberté
Cnis	Conseil national de l'information statistique
CNU	Conseil national des universités
Dares	Direction de l'animation de la recherche, des études et des statistiques (service statistique ministériel du Travail)
Depp	Direction de l'évaluation, de la prospective et de la performance (service statistique ministériel de l'Education nationale)
DGAFFP	Direction générale de l'administration et de la fonction publique
Dgesip	Direction générale pour l'enseignement supérieur et l'insertion professionnelle
Drees	Direction de la recherche, des études, de l'évaluation et des statistiques (service statistique ministériel de la Santé et des Affaires sociales)
DRJSCS	Direction régionale de la jeunesse, des sports et de la cohésion sociale
Finess	Fichier national des établissements sanitaires et sociaux (géré par la DREES)
INE	Identifiant national d'étudiant
Insee	Institut national de la statistique et des études économiques
IUFM	Institut universitaire de formation des maîtres
IUT	Institut universitaire de technologie
MEN	Ministère de l'Education nationale
Nes	Nomenclature économique de synthèse (gérée par l'INSEE)
NSF	Nomenclature des spécialités de formation (gérée par l'Education nationale)
PCS	Profession et catégories socioprofessionnelle (nomenclature gérée par l'INSEE)
Reflét	Regards sur les flux de l'enseignement technique et professionnel (base de données statistiques gérée par le Céreq)
RERS	Repères et références statistiques
Sicore	Système informatique de codage des réponses aux enquêtes (INSEE)
Sifa	Système d'information sur la formation des apprentis
Sise	Système d'information et de suivi des étudiants

ISSN 1776-3177
Marseille, 2012.