

Enquête communautaire sur l'innovation

Community Innovation Survey (CIS)

Méthodologie de l'enquête CIS2010

L'enquête Innovation, également appelée « Community Innovation Survey » (CIS) cherche à comprendre les mécanismes de l'innovation au sein des entreprises, ses objectifs et ses déterminants.

Présentation et historique de l'enquête CIS

Le dispositif d'enquête sur l'innovation, ou Community Innovation Survey (CIS), répond à une **demande politique européenne** de statistiques sur l'innovation exprimée au sommet de Lisbonne en 1997. Ce dernier s'appuie sur le manuel d'Oslo, élaboré sous l'autorité de l'OCDE et qui définit la méthodologie des enquêtes et les notions de l'innovation.

La première édition de l'enquête, CIS1, est menée en 1993 et s'intéresse uniquement aux **innovations de produits** (biens ou prestations de services) **ou de procédés**. Le champ de l'enquête est celui des entreprises industrielles.

Le champ de l'enquête et son contenu évoluent au cours des années et l'enquête devient plus régulière. De nouveaux sujets apparaissent dans CIS2 : existence de projets d'innovation n'ayant pas abouti, recours à des financements publics, obstacles à l'innovation etc. En 2001, CIS3 étend le champ de l'enquête aux entreprises du commerce et des services. En 2005, CIS4 couvre tous les secteurs d'activité avec un élargissement de la notion d'innovation en examinant également les **innovations d'organisation** et de **marketing**. Il s'agit d'un tournant pour l'enquête car c'est à partir de CIS4 que le questionnaire de l'enquête Innovation va prendre la forme qu'il a actuellement.

A partir de 2009, différents modules additionnels apparaissent dans CIS : un module environnemental pour CIS2008, un module concernant les compétences et la créativité pour CIS2010 etc.

Depuis CIS2008 (CIS6), l'enquête est gérée par l'Insee. Jusqu'alors, les enquêtes étaient réalisées par le Service des études et des statistiques industrielles (SESSI) du ministère en charge de l'industrie qui devient en 2009 le Pôle Statistiques Industrielles de l'Insee.

Les **principaux thèmes** abordés dans l'enquête sont les suivants :

- **fréquence de l'innovation par type d'innovation** (produits, procédés, organisation, marketing) : nombre d'entreprises ayant réalisé de telles innovations au cours de la période 2008 à 2010 ;
- degré de l'innovation : **part en chiffre d'affaires (2010) des innovations** de produits ou services introduites entre 2008 et 2010 suivant le degré de nouveauté : produits ou services nouveaux pour le marché, produits ou services nouveaux pour l'entreprise ;
- **dépenses liées aux activités d'innovation** de produits et de procédés suivant une nomenclature simple : dépenses de R&D (internes/externes), dépenses d'équipement liées au développement des innovations (hors R&D) ; dépenses pour acquérir des connaissances externes
- autres aspects du processus d'innovation concernant les innovations de produits et de procédés (questions qualitatives uniquement) : existence de **financements publics, sources d'information** et formes de **coopération** pour le développement des innovations, **objectifs des innovations** réalisées ;
- **facteurs freinant les activités d'innovation** en produits ou en procédés ;
- **créativité et compétences** acquises en interne ou externe (développement de logiciels, conception de sites Web etc.)

Champ de l'enquête

Période de référence : CIS2010 porte sur la période 2008-2010.

Année de référence : 2010

Certaines questions ne portent que sur l'année 2010 et non pas sur la période 2008 à 2010, telle que la question 5.2 sur les dépenses d'innovation

Champ : pour CIS2010, les entreprises interrogées sont :

- actives, marchandes et exploitantes au 31/12/2010
- de 10 salariés ou plus (en effectif ETP)
- implantées sur le territoire français (métropole et Dom)
- de secteurs d'activité correspondant, dans la nomenclature NAF rev2, aux divisions 05 à 81, sauf 45, 47, 55, 56 et 75.

Les divisions en NAF rev2 faisant partie du champ de l'enquête sont :

- industrie extractive (divisions 05 à 09) ;
- industrie manufacturière (divisions 10 à 33) ;
- production et distribution d'électricité, de gaz, de vapeur et d'air conditionné (division 35) ;
- production et distribution d'eau, assainissement, gestion des déchets et dépollution (divisions 36 à 39) ;
- la construction (divisions 41 à 43) ;
- commerce de gros, à l'exception des automobiles et des motocycles (division 46) ;
- transports et entreposage (divisions 49 à 53) ;
- édition (division 58) ;
- production de films cinématographiques, de vidéo et de programme de télévision ; enregistrement sonore et édition musicale / programme et diffusion (divisions 59 et 60) ;
- télécommunications (division 61) ;
- programmation, conseil et autres activités informatiques (division 62) ;
- services d'information (division 63) ;
- activités financières et d'assurance (division 64 à 66) ;
- activités immobilières (division 68) ;

- activités juridiques et comptables (division 69) ;
- activités des sièges sociaux ; conseil de gestion (division 70) ;
- activités d'architectures et d'ingénierie ; activités de contrôle et analyses (division 71) ;
- recherche-développement scientifique (division 72) ;
- publicité et études de marché (division 73) ;
- autres activités spécialisées, scientifiques et techniques (division 74) ;
- activités de location et location-bail (division 77) ;
- activités liées à l'emploi (division 78) ;
- activités des agences de voyages, voyagistes, services de réservation et activités connexes (division 79) ;
- enquêtes et sécurités (division 80) ;
- services relatifs aux bâtiments et aménagement paysager (division 81).

L'enquête Innovation 2010 s'inscrit dans le cadre du **règlement européen n°1450/2004** qui définit les indicateurs statistiques à produire sur le champ commun pour tous les participants européens à l'enquête et établit les délais à respecter.

Le champ obligatoire, imposé par Eurostat, recouvre les divisions : 05-09, 10-33, 35, 36-39, 46, 49-53, 58, 61, 62-63, 64-66, 71. Les autres divisions couvertes par l'Insee dans CIS2010 sont optionnelles.

La réalisation de l'enquête

L'enquête comporte une **partie nationale** et une **partie régionale**.

La **partie nationale**, qui concerne près des trois-quarts des entreprises, permet d'obtenir des informations l'intégralité du champ. Pour la partie nationale, les entreprises retenues dans l'échantillon selon leur activité et tranche d'effectifs. Les **extensions régionales** viennent compléter l'information par région sur certains secteurs, en tirant un nombre suffisant d'entreprises pour chaque région. Elles permettent en effet d'augmenter la précision des statistiques par région, en augmentant la marge de certitude des résultats.

Période de collecte : la collecte a eu lieu du 14 juin au 31 décembre 2011 pour la partie nationale, de septembre à fin décembre pour la partie régionale.

Mode de collecte : l'enquête est effectuée par internet, avec possibilité de répondre par papier sur demande. Les entreprises reçoivent un courrier avec la lettre de présentation de l'enquête contenant le lien internet ainsi que l'identifiant et le mot de passe permettant de répondre en ligne.

Extensions régionales : les extensions régionales viennent compléter l'information par région sur certains secteurs. Pour certaines régions, l'extension porte exhaustivement sur un ou plusieurs secteurs, c'est-à-dire que toutes les entreprises de la région dans ce ou ces secteurs sont retenues. Le champ des extensions régionales est celui des entreprises « régionales », c'est-à-dire ayant 80 % de leurs postes de travail situés dans une seule région. De plus parmi ces entreprises, seules celles ayant moins de 250 salariés sont prises en considération (les entreprises de 250 salariés ou plus étant prises en compte exhaustivement dans la partie nationale).

Les extensions régionales concernent dix régions. Sept régions font l'objet d'extensions exhaustives sur l'activité considérée : La Guadeloupe, la Martinique, la Réunion, la Haute-Normandie, la Franche-Comté, le Poitou-Charentes et le Languedoc-Roussillon. Les trois dernières, qui sont l'Île-de-France, la Bretagne et le Rhône-Alpes, font l'objet d'extensions non exhaustives.

Pour plus d'informations : se reporter à l'[Annexe 1](#).

Plan de sondage

Unités interrogées : l'unité statistique interrogée est l'entreprise au sens d'unité légale. Dans les cas des unités profilées (Renault, PSA etc.), seule la tête de groupe est interrogée.

Base de sondage : la base est issue du répertoire SIRENE et contient 116 714 entreprises qui entrent dans le champ.

Taille de l'échantillon : l'échantillon comporte 29 370 entreprises.

- 21 318 au titre du tirage national
- 8 052 au titre d'extensions régionales.

Type de sondage : sondage stratifié.

Échantillon national

Pour le tirage national, la base de sondage est stratifiée selon **trois critères** :

- un **critère d'activité** (niveau division de la NAF, sauf pour l'industrie agro-alimentaire où la stratification sectorielle est plus fine)
- un **critère de taille** (4 tranches de salariés en ETP, A : 10 à 19 salariés, B : 20 à 49 salariés, C : 50 à 249 salariés, D : 250 salariés et plus)
- un **critère de localisation** (regroupements de régions principales d'implantation)

Les **allocations** ont été calculées en **trois temps** :

- 1) Dans un premier temps l'allocation est optimisée à un niveau de stratification agrégé (A38 x tranche d'effectif)¹.
- 2) Les taux de sondage obtenus sont ensuite appliqués dans chaque sous-strate (au niveau « division » de la NAF x tranche d'effectif) en imposant de tirer au moins 10 unités par sous-strate.
- 3) Enfin les taux de sondage obtenus dans le deuxième temps sont appliqués à chaque strate de tirage (activité x tranche d'effectif x regroupements de régions).

Remarques :

une stratification sectorielle plus fine est faite dans le secteur des industries agroalimentaires où le plan de sondage est optimisé au niveau A38 * tranche d'effectif.

Exhaustivité : toutes les entreprises ayant plus de 250 salariés sont interrogées dans l'enquête. Elles représentent 3,7% des entreprises de l'échantillon.

Échantillon des extensions régionales

Comme nous l'avons dit plus haut, les extensions régionales ne s'intéressent qu'aux entreprises de moins de 250 salariés ayant au moins 80 % de leurs effectifs dans une seule région.

Les **extensions régionales** sont exhaustives dans les secteurs concernés sauf pour les régions Bretagne, Île-de-France, Rhône-Alpes. Pour ces trois régions, le plan de sondage est en étroite relation avec celui défini au niveau national : une allocation optimale a ainsi été calculée au niveau (A38 x tranche d'effectif), les taux de sondage ainsi obtenus ayant été appliqués au niveau des strates de tirage (Activité x tranche d'effectif).

Pour les autres régions, toutes les entreprises du secteur d'activité retenu sont sélectionnées.

¹ Le calcul des allocations par strate d'optimisation a été effectué en combinant deux allocations :

- une allocation de Neyman sur le taux d'entreprises innovantes (en produit ou procédé)
- une allocation proportionnelle au nombre d'entreprises innovantes

Traitement de la non-réponse et calage

Le taux de réponse à l'enquête est de 81,1 % pour CIS2010. Afin d'avoir les meilleurs résultats possibles, il est nécessaire d'effectuer des redressements de la non-réponse et un calage.

Les principales étapes de traitement de la non-réponse sont :

- 1) les **travaux préliminaires** aux redressements
- 2) la **redressement de la non-réponse**
 - a. partielle
 - b. totale
- 3) le **calage**

Les travaux préliminaires aux redressements comprennent aussi un contrôle de la cohérence des réponses fournis qui peuvent amener à les modifier : il s'agit de la phase d'apurement.

Entreprises « substituables » / « non-substituables »

Lors du traitement de la non-réponse, on distingue les entreprises « **substituables** » des entreprises « **non-substituables** ».

- Entreprises « non-substituables »

Les entreprises dites « **non-substituables** » sont les entreprises ayant un comportement singulier par rapport aux variables d'intérêt de l'enquête. Elles sont ici au nombre de 141 (dont une hors-champ) et ont les caractéristiques suivantes :

- un effectif au lancement au moins égal à 500
- un chiffre d'affaires au lancement au moins égal à 4 000 000 k€, ou un montant total de dépenses d'innovation - variable RTOT - au moins égal à 50 000 k€, ou encore un montant déclaré dans l'enquête R&D en dépenses réalisées en interne ou en externe, supérieur à 50 000 k€

Ces entreprises ont un protocole de contrôle particulier (détaillé dans la suite).

- Entreprises « substituables »

Les entreprises ne vérifiant pas ces conditions sont dites « **substituables** »

Remarque : la condition sur l'effectif fait notamment qu'aucune entreprise échantillonnée au titre d'une extension régionale n'est non-substituable, toutes les entreprises de 250 salariés ou plus étant tirées dans l'échantillon national.

1 - Les travaux préliminaires aux redressements

Répartition « répondant / non-répondant / hors-champ »

Les entreprises échantillonnées sont réparties en trois groupes :

- « H » pour les unités repérées comme étant hors-champ ;
- « R » pour les unités répondantes dans le champ ;
- « N » pour les unités non-répondantes dans le champ.

Phase de typage

Chaque observation d'entreprise dispose, pour chaque variable, d'un indicateur (une variable dite de « **statut** »). Ces statuts permettent ainsi de repérer les éléments modifiés lors du redressement de la non-réponse partielle. Cet indicateur prend les modalités suivantes :

- « B » s'il s'agit d'une donnée brute/authentique (c'est-à-dire fournie par l'unité interrogée) ou si la donnée est issue de la phase d'apurement ;
- « M » s'il s'agit d'une valeur manquante : cet item se réfère à l'absence de donnée alors qu'on devrait avoir une réponse ;
- « N » s'il s'agit d'une entreprise non concernée (questions filtrées).

Ces variables permettent notamment de calculer les taux de non-réponse partielle (égaux au rapport du nombre de statuts « M » sur le nombre de statuts « B » ou « M ») pour chaque variable d'une part et pour chaque entreprise d'autre part. Lorsque, pour une entreprise, ce taux est supérieur à 75%, l'entreprise est requalifiée en non-répondante totale.

2 - Le redressement de la non-réponse

a. Redressement de la non-réponse partielle

- Entreprises « substituables »

Sources externes

Des sources externes sont utilisées pour certains redressements, telle l'enquête R&D (menée par le Ministère de l'Enseignement Supérieur et de la Recherche) sur les questions relatives aux dépenses de recherche et développement. La base de donnée de l'année précédente est également utilisée pour valider ou non certaines réponses.

Méthodes

Les méthodes utilisées (décrites dans la suite) sont des méthodes de redressement **par donneur** ou **d'imputation aléatoire**. Chacune de ces méthodes utilise les réponses des entreprises répondantes pour imputer une entreprise non-répondante.

Dans le cas où l'entreprise à imputer serait une entreprise d'une région à extension, le donneur est choisi au sein du fichier national : il peut donc en particulier être d'une région d'implantation principale distincte de celle du receveur.

Variables quantitatives

Pour le tableau 2.3, on **choisira un donneur** pour chacune des trois variables (TURNMAR, TURNIN, TURNUNG), après une recherche de liens. On débutera par le redressement de la variable ayant les valeurs les plus importantes parmi les répondantes. On s'assurera après chacun de ces trois redressements que la somme reste inférieure ou égale à 100. Lorsque la somme dépassera 100, on procédera à la troncature de la dernière valeur redressée de façon à retrouver un total de 100.

Variables qualitatives

On commence par la recherche de liens pour imputer une variable qualitative. On obtient ainsi des variables auxiliaires. Cette recherche de liens sera menée selon l'importance que l'on créditera à la variable qualitative considérée. On définira ainsi des **variables « prioritaires »** pour lesquelles on s'efforcera de rechercher un lien **optimal**.

Pour ces **variables prioritaires**, des liens seront cherchés avec d'autres variables potentielles telles que taille, activité, région principale d'implantation ou d'autres variables du questionnaire.

La recherche de liens sera moins poussée pour les **autres variables** : nous redresserons soit en fonction du couple de variable taille-activité ; soit comme la variable prioritaire dominante du même bloc.

Une fois cette recherche de liens effectuée, nous utiliserons la méthode **d'imputation aléatoire**. Cette méthode consiste à exploiter la distribution observée, pour la variable à imputer, sur les répondantes. Ainsi, si 50% des entreprises des entreprises répondantes de 10 à 49 salariés innovent en matière d'organisation (leur statut, pour cette variable, étant égal à « B »), une entreprise non-répondante à cette question (statut = « M ») et ayant de 10 à 49 salariés aura une chance sur deux de se voir imputer « oui » (on a supposé ici que l'innovation en matière d'organisation était liée à la variable TEFF désignant la taille de l'entreprise, regroupée par tranches).

Les premières variables qui seront redressées selon cette méthode seront les **questions-filtre** (questions 2.1, 3.1, 4.1, 5.1A1, 5.3a3, 6.2, 9.1, 10.1). Une fois ces imputations effectuées, on procédera à la correction éventuelle des statuts des questions filtrées correspondantes (en effet, l'imputation de la question filtre peut modifier le statut de l'entreprise, considérée initialement comme non concernée par la question filtrée). S'agissant des questions 9.1 et 10.1, lorsque les questions

filtrées seront déjà, au moins partiellement, renseignées par une réponse positive ('Elevé', 'Moyen' ou 'Faible'), on s'assurera qu'au moins un 'Oui' aura été imputé à la question-filtre.

Le tableau 6.3 sera l'objet d'un traitement particulier : en effet, on n'utilisera pas ici la méthode d'imputation aléatoire mais la **méthode du donneur** (un donneur **pour le tableau** entier). Celui-ci sera recherché parmi les entreprises de « même type » que l'entreprise à imputer. Le « type » de l'entreprise sera défini au moyen d'une variable auxiliaire liée à la modalité la plus cochée dans le tableau, parmi l'ensemble des questionnaires. Les entreprises n'appartenant à aucun groupe ou réseau d'enseigne (GP = RE = « Non ») se verront toutefois attribuer des 0 (cases non cochées) sur toute la première ligne. Une conséquence de ce traitement particulier est, notamment, qu'une variable statut unique sera constituée pour le tableau dans son ensemble : il n'y aura pas une variable de statut pour chacune des variables qui le composent.

Contrôle du redressement

A l'issue du redressement de la non-réponse partielle, le programme de contrôle de l'apurement est exécuté de nouveau pour s'assurer que le redressement n'aura pas introduit d'incohérence. Dans le cas contraire, on procèdera à des corrections « manuelles ».

- Entreprises non substituables

Dans un premier temps, ces entreprises ont été redressées via des données externes, constituées par les données de CIS 2008 (après expertise de la maîtrise d'ouvrage) et les résultats de l'enquête R&D. A noter que les variables redressées de cette façon ont un statut égal à « M ».

Lorsque cette méthode n'a pas permis l'imputation, on a utilisé les mêmes méthodes que celles appliquées aux entreprises substituables.

b. Redressement de la non-réponse totale

- Entreprises « substituables »

Repondération pour non réponse

La méthode générique employée pour corriger la non-réponse totale, pour les entreprises substituables, est la **repondération**. Il s'agit de considérer le fait de répondre ou non comme une nouvelle phase d'échantillonnage où, pour chaque entreprise, la probabilité d'être sélectionnée est égale à sa probabilité de réponse. Le coefficient de « repondération » à utiliser est l'inverse de cette probabilité de réponses.

La méthode de repondération est la suivante :

- les entreprises non-répondantes totales substituables se voient attribuer un poids de 0 en fin de traitement
- les entreprises répondantes substituables voient leur poids se dilater suite à la repondération.

La méthode employée pour la repondération est celle des groupes de réponses homogènes où on partage la population en groupes d'unités supposée avoir le même comportement de réponse.

- Entreprises « non substituables »

Les entreprises « **non-substituables** » ont été **exclues** des traitements de la non-réponse totale, notamment car il peut-être gênant de mettre un poids supérieur à 1 à de grandes entreprises. Le traitement relatif à ces entreprises est décrit dans la partie « redressement de la non-réponse partielle ».

Groupes de réponses homogènes

Les **groupes de réponses homogènes** sont constitués à l'aide de caractéristiques disponibles à la fois sur les répondants et les non-répondants, et qui se révèlent avoir une influence significative sur la probabilité de réponse. Ces groupes se définissent par rapport à ces caractéristiques.

La méthode consiste à bâtir un modèle explicatif du comportement de réponse et à retenir les variables qui paraissent significatives. On ne prend pas en compte tous les croisements possibles entre les variables retenues, mais on rassemble les groupes « adjacents » (c'est-à-dire des groupes aux caractéristiques proches et pour lesquels la différence de taux de réponse n'est pas significative). Les variables à retenir doivent être connues sur les répondants et sur les autres, ce qui exclut le recours aux questions de l'enquête. Les variables envisagées sont : la tranche de taille de l'entreprise ; le secteur d'activité ; la localisation du siège ou la région de principale implantation ; le comportement de réponse à l'enquête précédente ; des variables économiques et financières sur les entreprises : chiffre d'affaires, taux d'investissement, taux d'exportation par exemple ; des variables relatives à la composition de la main d'œuvre.

Une fois les groupes de réponses homogènes constitués, on effectue une repondération pour corriger la non-réponse totale. A cet effet, on estimera la probabilité de réponse à l'intérieur de ces groupes, comme le rapport (pondéré par les poids de lancement) entre le nombre de répondants et l'effectif total du groupe.

3 - Le calage

Les méthodes de **calage** d'un échantillon consistent à **changer les poids** (poids consécutifs à la correction de la non-réponse totale dans le cas présent) pour que les estimations de totaux de variables quantitatives soient égales aux vrais totaux connus par ailleurs sur la population. Ceci permet aussi de traiter le cas de variable qualitative, pour lesquelles on se calera sur les effectifs de chaque modalité. On parlera aussi de « totaux ». Les poids initiaux conduisant à des estimations sans biais, la méthode de calage consiste à modifier le moins possible ces poids pour retrouver les totaux connus sur des variables auxiliaires. De la sorte, on améliore les estimations sur toutes les questions liées à ces variables auxiliaires.

Rappelons que les variables considérées doivent constituer une information « extérieure à l'enquête », disponible sur les unités enquêtées et dont le total est connu sur l'ensemble de la population. L'idée générale est ici d'opérer un seul calage sur l'ensemble du fichier. De fait les entreprises des régions à extension ont le même poids entre le fichier national et le fichier régional de la région considérée. Cela supposera d'introduire, au niveau des strates de calage, chaque région à extension et de regrouper l'ensemble des régions sans extension

A l'issue de la phase de calage, chaque unité est dotée d'un **nouveau poids** (poids_cal), qui sera son **poids définitif pour l'exploitation des données**. Comme pour les enquêtes passées, les unités non substituables sont exclues de la phase de calage, afin de conserver leur poids de 1.

Une fois l'ensemble du redressement terminé, **des calculs de contribution** sont effectués pour chaque variable quantitative. Cela permet notamment de voir s'il n'est pas nécessaire de considérer de manière particulière quelques unités fortement influentes dans les résultats. Le redressement ne prend fin que lorsque ces unités influentes ont été traitées.

ANNEXE 1 : le champ des extensions

Région (8052)	Champ des extensions	Taille des extensions
Guadeloupe (342)	Industries extractives ; Industrie manufacturière Eau - Assainissement - Gestion des déchets Construction Services technologiques ; Services intellectuels Services administratifs et de soutien	3 ; 104 9 99 13 ; 38 76
Martinique (347)	Industries extractives ; Industrie manufacturière Eau - Assainissement - Gestion des déchets Construction Services technologiques ; Services intellectuels Services administratifs et de soutien	5 ; 107 9 105 9 ; 53 59
La Réunion (449)	Industries extractives ; Industrie manufacturière Eau - Assainissement - Gestion des déchets Services technologiques ; Services intellectuels Services administratifs et de soutien	5 ; 229 14 22 ; 82 97
Île-de-France (1000)	Industrie manufacturière	1000
Haute-Normandie (670)	Industrie manufacturière	670
Franche-Comté (847)	Industrie manufacturière Services technologiques ; Services intellectuels	714 22 ; 111
Bretagne (1000)	Industrie manufacturière Services technologiques ; Services intellectuels réduits	752 106 ; 142
Poitou-Charentes (938)	Industrie manufacturière Services technologiques ; Services intellectuels	731 37 ; 170
Rhône-Alpes (1500)	Industrie manufacturière Eau - Assainissement - Gestion des déchets Services technologiques ; Services intellectuels	390 86 363 ; 661
Languedoc-Roussillon (959)	Industrie manufacturière Services technologiques ; Services intellectuels	560 90 ; 309

Champ des extensions	Contour en Naf rev2 (NAF 88)
Industries extractives	05 à 09, en pratique uniquement 08
Industrie manufacturière	10 à 33
Eau - Assainissement - Gestion des déchets	36 à 39
Construction	41 à 43
Services technologiques	58 à 63 et 72
Services intellectuels	69 - 70 - 71 -73 champ réduit : 71 et 73
Services administratifs et de soutien	77 à 82