

# Le renouvellement partiel d'échantillons d'entreprises

Arnaud Fizzala

**Résumé** — Lorsque les enquêtes auprès des entreprises sont répétées dans le temps et que l'on s'intéresse à des évolutions, l'échantillon n'est généralement pas renouvelé intégralement à chaque édition de l'enquête, mais seulement en partie - la moitié, un tiers, un quart... - selon un arbitrage à réaliser entre la charge de réponse que l'on pense tolérable pour les entreprises et l'importance accordée à l'évaluation des évolutions<sup>1</sup> dans les objectifs de l'enquête. On parle alors d'échantillon rotatif au sens où la partie de l'échantillon à renouveler change ("tourne") à chaque nouvelle édition de l'enquête. Deux techniques sont utilisées à l'Insee pour gérer le tirage de ces échantillons rotatifs d'entreprises<sup>2</sup> : la technique dite du numéro Hexal est appliquée en "régime courant", c'est-à-dire lorsqu'il y a peu de changements (de champ, de taille d'échantillon...) entre deux renouvellements, tandis que la technique dite de coordination positive est plutôt utilisée en cas de modification importante du champ de l'enquête ou du plan de sondage.

## I. EN SITUATION COURANTE : LA TECHNIQUE DU NUMÉRO HEXAL

La technique la plus couramment utilisée à l'Insee pour gérer les rotations d'échantillon d'entreprises se base sur un numéro permanent - dit numéro hexal - attribué aléatoirement à chaque unité de la base de sondage correspondant à la première édition de l'enquête. On présente dans la suite le cas le plus courant qui correspond aux renouvellements par moitié, mais la méthode se généralise sans difficultés aux renouvellements par tiers, par quart, par cinquième et par sixième. Pour réaliser le renouvellement par moitié d'un échantillon, la base de sondage de l'année initiale est divisée en deux parties équivalentes via l'attribution aléatoire, à chaque unité, d'un numéro "hexal" permanent compris entre 1 et 60<sup>3</sup>. Les unités possédant un numéro hexal impair forment la première partie de la base de sondage, et les unités possédant un numéro hexal pair la seconde<sup>4</sup>. L'échantillon de l'année initiale se trouve de fait lui aussi partitionné<sup>5</sup> en deux sous-échantillons de tailles équivalentes.

Par suite, pour la gestion courante de l'échantillon une année N donnée, le renouvellement par moitié des unités de l'échantillon consiste alors :

1. Plus les échantillons se recouvrent et plus l'estimation d'une évolution sera précise.

2. Plusieurs enquêtes auprès des ménages se basent aussi sur des échantillons rotatifs, par exemple l'enquête Loyers et Charges renouvelée au cinquième chaque trimestre et l'enquête Statistiques sur les Ressources et Conditions de Vie (SRCV) renouvelée au neuvième chaque année. La méthodologie utilisée pour le tirage de ces échantillons rotatifs n'est pas la même que celle présentée dans cette note qui n'est actuellement appliquée que pour les enquêtes auprès des entreprises.

3. Ce choix s'explique par le fait que 60 est divisible par tous les entiers compris entre deux et six : un numéro hexal compris entre 1 et 60 permet donc de partitionner la base de sondage en deux, trois, quatre, cinq ou six parties et donc de gérer des renouvellements par moitié, par tiers, par quart, par cinquième ou par sixième.

4. On peut aussi choisir de former la première partie avec les unités dont le numéro hexal est compris entre 1 et 30 et la seconde avec les unités dont le numéro hexal est compris entre 31 et 60.

5. En fonction de la parité des numéros hexal des unités le composant.

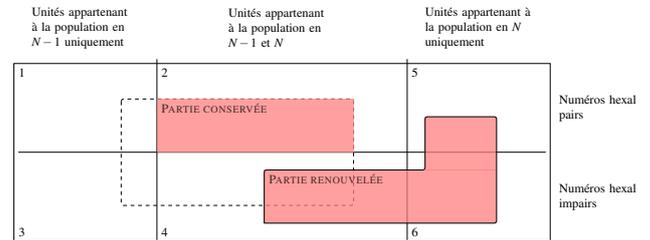


Fig. 1. Renouvellement par moitié concernant les numéros hexal impairs

- ▶ d'une part à conserver l'une des deux parties de la base de sondage (dite partie conservée) de l'année N-1, privée des unités qui ne sont plus dans le champ de l'enquête en N, ainsi que la partie correspondante de l'échantillon (dite partie conservée de l'échantillon) ;
- ▶ et d'autre part à tirer un nouvel échantillon dans une base de tirage constituée des unités de l'autre partie de la base de sondage (dite partie renouvelée) de l'année N-1, également privée des unités qui ne sont plus dans le champ de l'enquête en N, auxquelles sont ajoutées les unités nouvelles, c'est-à-dire entrées dans le champ de l'enquête en N. On attribue aléatoirement à ces unités nouvelles un numéro hexal qui peut être pair ou impair et qui sera conservé lors des prochains tirages.

Dans la figure 1, l'échantillon de l'année N est constitué de l'ensemble des parties roses. Le tirage effectivement réalisé l'année N n'a concerné que les unités de numéro hexal impair présentes à la fois dans la base de sondage de l'année N-1 et de l'année N - partie 4 -, ainsi que les unités nouvelles, c'est-à-dire absentes de la base de sondage en N-1, que leur numéro hexal soit pair ou impair - parties 5 et 6. Ces unités effectivement tirées en N correspondent donc sur la figure 1 au "L couché rose" et sont extrapolables à l'univers 4 + 5 + 6. Les unités de la partie conservée de l'échantillon en N sont extrapolables<sup>6</sup> à l'univers 2.

Dans une démarche de réduction<sup>7</sup> de la charge statistique imposée aux entreprises, cette méthode est généralement utilisée en coordonnant négativement [2] la partie de l'échantillon renouvelée l'année N avec<sup>8</sup> l'échantillon des années précédentes. Dans le cas des renouvellements par

6. Notons que le poids de sondage d'une unité de la partie conservée est issu du tirage réalisé l'année N-1, et doit donc lui aussi être conservé, même si les caractéristiques de l'unité ont évolué et l'auraient placée dans une autre strate si le tirage avait été effectué en N.

7. Ou dans le cas présent d'une meilleure répartition de la charge statistique entre les entreprises "statistiquement équivalentes".

8. Dans le cadre de la coordination négative générale des échantillons d'entreprises, la partie renouvelée est aussi coordonnée négativement avec les échantillons d'autres enquêtes récentes auprès des entreprises, mais en cherchant d'abord à minimiser le recouvrement avec l'échantillon de l'édition précédente de l'enquête. Techniquement, cela est réalisé en appliquant un "paramètre de charge" moins important aux autres enquêtes lors de la mise en œuvre de la coordination négative.

moitié, on cherche ainsi à interroger les unités deux ans, mais pas plus.

## II. EN SITUATION EXCEPTIONNELLE : LA COORDINATION POSITIVE

Notons qu'en cas de modifications substantielles du plan de sondage d'une enquête ayant un impact important sur l'échantillon, la méthode du numéro hexal décrite précédemment ne convient pas au renouvellement partiel de l'échantillon du fait de l'impossibilité d'en modifier la partie conservée. Par exemple, dans le cas extrême où la taille de l'échantillon devrait être divisée par deux une année N donnée, on devrait se limiter à une partie renouvelée de l'échantillon de quelques unités seulement (voire aucune s'il n'y a pas eu d'unités entrées ou sorties du champ de l'enquête), créant ainsi un déséquilibre les années suivantes entre les tailles des deux parties de l'échantillon. Lorsque ces cas problématiques se présentent, on utilise plutôt, pour renouveler partiellement l'échantillon, une technique plus "souple" dite de coordination positive. Cette technique utilise les mêmes outils que la technique de coordination négative, mais avec un objectif opposé, c'est-à-dire en visant un recouvrement maximal entre les échantillons.

Le principal inconvénient de la technique de coordination positive est la difficulté de bien maîtriser le recouvrement entre les échantillons et la durée de présence des unités dans le panel. C'est pourquoi, lorsqu'il n'y a pas de modification de l'enquête ayant un impact important sur l'échantillon, on lui préfère généralement la méthode du numéro hexal. Cependant, avec la mise en œuvre de tirages d'échantillons au niveau entreprise dans les années à venir, cette règle risque de changer. En effet, les contours des entreprises évoluant d'une année sur l'autre, plusieurs questions<sup>9</sup> seront à résoudre pour continuer à utiliser la méthode du numéro hexal plutôt conçue pour être mise en œuvre sur des bases de sondage stables dans le temps.

## III. EXEMPLES D'UTILISATION À L'INSEE

Tous les ans, la division Sondages de l'INSEE tire par exemple l'échantillon d'établissements de l'enquête annuelle sur les consommations d'énergie dans l'industrie (EACEI) et l'échantillon d'unités légales du volet entreprise de l'enquête sur les technologies d'information et de communication (TIC – entreprises) qui sont tous les deux renouvelés par moitié chaque année en appliquant la technique du numéro hexal. La division Sondages tire également l'échantillon de très petites<sup>10</sup> entreprises de l'enquête sur l'activité et conditions d'emploi de la main d'oeuvre (ACEMO - TPE) qui est lui renouvelé par quart chaque année en appliquant la technique du numéro hexal.

La méthode de coordination positive d'échantillons a par exemple été utilisée en 2008, lorsque les enquêtes annuelles entreprises (EAE), dont l'échantillon était renouvelé par moitié chaque année selon la technique du numéro hexal, ont été remplacées par l'enquête sectorielle annuelle (ESA). Le champ de l'enquête avait été revu à cette occasion et la taille

9. Comment gérer la modification du contour d'une entreprise appartenant à la partie conservée de la base de sondage par exemple. En particulier si elle absorbe une autre unité de la partie conservée.

10. De 1 à 9 salariés.

de la partie non exhaustive de l'échantillon avait été divisée par deux. La méthode du numéro hexal ne semblait pas envisageable avec de telles modifications et c'est donc la méthode de coordination positive qui a été utilisée<sup>11</sup>. Dès 2009, le plan de sondage s'étant stabilisé, un renouvellement par moitié de l'échantillon de l'ESA s'appuyant sur la technique du numéro hexal a été remis en place et l'échantillon a été géré de cette façon jusqu'en 2015. En 2016, une coordination positive a été à nouveau mise en place afin de gérer le passage au concept d'entreprise pour le tirage.

## REFERENCES

- [1] Demoly E., Fizzala A., Gros E. (2014). Méthodes et pratiques des enquêtes entreprises à l'Insee. *Journal de la Société Française de Statistique*, 155, 4, 134 à 159.
- [2] Fiche méthodologique "La coordination des échantillons des enquêtes auprès des entreprises".
- [3] Cotton F., Hesse C. (1992) Tirages coordonnés d'échantillons. *Document de travail E9206*, Insee.



*Département des méthodes statistiques  
Version n° 1, diffusée le 24 janvier 2018*

11. Il s'agissait à l'époque de l'ancienne méthode de coordination d'échantillons[3], la méthode utilisée actuellement n'étant pas encore opérationnelle.