

II / Face aux sources : corpus et échantillons

La remise en cause du quantitatif depuis les années 1980 est largement passée par des questions sur l'usage des sources. Plutôt que d'utiliser des statistiques du passé ou de rechercher une mise en série systématique, beaucoup préfèrent dorénavant plonger dans des sources plus « micro », plus nominatives, et ne pas en sacrifier la richesse en les catégorisant (*cf.* chapitre III). De plus, au vertige des courbes a souvent succédé celui des bases de données exhaustives comme « façon de retrouver le paradis perdu de la totalité » [Lepetit, 1989]. Comme le dit André Zysberg [1987], on a pu entendre des universitaires « affirmer, non sans gourmandise : *et le corpus ne sera jamais achevé...* » Ce vertige est entretenu, depuis les années 2000, par l'objectif de la mise en ligne : si ces bases de données sont des outils pour la recherche, ne faut-il pas qu'elles puissent répondre à toutes les questions possibles ? En réalité, une telle base de données n'existe pas, ne serait-ce que parce qu'elle repose, implicitement ou explicitement, sur la définition d'un corpus.

C'est la première étape, souvent difficile, à laquelle tout chercheur est confronté : comment constituer des sources en corpus de données quantifiables ? Quels sont les écueils à éviter pour produire des chiffres à la fois interprétables et qui ne relèvent pas de la surinterprétation ? Pour cela, il est nécessaire de connaître quelques grands principes. Si aucune source n'impose ni n'exclut *a priori* la quantification, il faut arriver à ne comparer que ce qui est comparable, se garder d'épiloguer sur des différences faibles entre corpus exigus, mais aussi et sans doute surtout savoir échantillonner à bon escient : la quantification n'a de sens

que si les résultats obtenus sont à la mesure des efforts fournis, en particulier lors de la saisie.

Le choix du corpus

La diversité des sources quantifiables

Les traitements quantitatifs peuvent s'appliquer à des sources très variées : statistiques ou nominatives, mais aussi textuelles ou iconographiques. Recensements de la population, listes de membres d'institutions, dossiers de personnel, actes d'état civil, inventaires après décès, relevés stratigraphiques de sites archéologiques, correspondances, discours, pièces de céramique ou encore œuvres d'art : la liste n'est pas exhaustive des sources qui peuvent faire l'objet de traitements quantifiés. Séverine Lepape [2007] interprète ainsi la genèse et les transformations d'un thème iconographique classique du Moyen Âge, l'arbre de Jessé, grâce à une analyse quantifiée de plusieurs centaines d'objets, enluminures, gravures, verrières, sculptures, tapisseries, peintures murales, etc. Sur la base de la composition de trésors monétaires découverts en Europe et au Moyen-Orient, il est possible de proposer des hypothèses sur les liens économiques et politiques entre des contrées très éloignées, dans l'Antiquité et au Moyen Âge [Depeyrot, 2002]. Quant à Hans Joachim Voth [2000], c'est en étudiant les dépositions dans des procès criminels de quelques milliers de témoins, qui indiquent incidemment leurs activités à différents moments de la journée, qu'il revisite l'histoire du temps de travail dans l'Angleterre de la révolution industrielle...

Le point commun entre les sources quantifiables est donc à chercher du côté non pas de leur type ou de leur forme, mais de leur adéquation aux contraintes de la quantification, en fonction des problématiques de recherche. Il est en effet nécessaire de se poser, dès le début, des questions de définition, de comparabilité et de représentativité : a-t-on l'information pour tout le monde ? Les sources dont on dispose peuvent-elles raisonnablement être mises en série ?

Multiplier les corpus

Les historiens novices en matière de quantification sont parfois prisonniers d'une recherche *du* corpus ou de *l'échantillon* unique et idéal. Or rien n'interdit, au contraire, d'étudier plusieurs groupes (d'individus, de villages, de livres, de tableaux...) de différentes tailles, inclus ou non les uns dans les autres. Les groupes les plus nombreux peuvent alors faire l'objet de recherches moins poussées, mais qui fournissent des données de cadrage ou comparatives. Les plus petits corpus servent quant à eux à des études détaillées, croisant un grand nombre de variables et permettant d'élaborer, voire de tester, des hypothèses plus fines. Si les grandes réussites de la « prosopographie de masse » restent réservées à des enquêtes au long cours et au budget énorme [Rutman et Rutman, 1984], rien n'interdit, même en calibrant une recherche en fonction de la durée d'un master ou d'une thèse, de fournir des résultats marquants.

Entre biographie et quantification

Faire de la prosopographie, ou de la « biographie collective », « faire la sociologie d'un groupe » : telle est sans doute la pratique la plus répandue parmi les historiens qui utilisent des bases de données et/ou des comptages.

La prosopographie vient de l'histoire de l'Antiquité [Charle, 2001]. Grâce à l'étude de listes de personnes partageant un attribut (sénateurs romains, professeurs d'université...), il s'agit de définir ce qu'un groupe a en commun. La démarche s'impose presque naturellement quand on étudie une institution, dont on commence par lister les membres ; mais elle comporte des risques, en particulier celui de produire une accumulation de fiches qui ne sont pas de vraies biographies, sans pour autant pouvoir être réellement comparées de façon quantifiée. Celui ou celle qui se lance dans une prosopographie doit donc, d'emblée, définir des principes de collecte des données, en fonction de ses objectifs : cherche-t-on à constituer de véritables biographies ou une base de données permettant la quantification ? Vise-t-on une recherche exhaustive ou pas, et sur quelles informations ?

Surtout, que veut-on montrer ? S'il s'agit de situer un groupe par rapport à son environnement, il n'est pas très judicieux de n'étudier que ce groupe. La prosopographie en tant que telle est

plutôt propre à cerner les différences internes, dans une visée typologique qui peut être liée, notamment, à un traitement en termes de générations. Ainsi, Marie-Bénédicte Daviet-Vincent [2006] étudie un échantillon de hauts fonctionnaires allemands de l'Empire et de la République de Weimar. Cela lui permet à la fois de discuter des phénomènes de concurrence entre élites nobles et bourgeoises, et de souligner la relative continuité du personnel administratif après la Première Guerre mondiale, malgré le changement de régime politique.

Le « questionnaire » et l'historien

La constitution d'un corpus se doit également de respecter la spécificité des sources. L'historien ne peut se transformer en sociologue faisant passer à ses « enquêtés » un questionnaire rétrospectif à questions fermées. La visée quantitative ne doit pas conduire à négliger ce qui différencie les sources historiques des questionnaires en face-à-face : il faut respecter l'hétérogénéité des sources et conserver systématiquement le lien avec elles (qui permet vérification ou recodage). Lorsqu'elles se déploient dans le temps, cet aspect ne doit pas être masqué par la saisie, par exemple si l'on ne prévoit qu'une case « profession » ou « adresse », alors que, en général, les sources fournissent plusieurs indications différentes. Non seulement les individus changent, mais les réponses dépendent de la situation : auto-déclaration ou interprétation par un fonctionnaire, enjeux concrets (fiscaux par exemple) ou symboliques (présentation de soi dans le *Who's Who*)...

Prendre en compte ces éléments complique certes le travail de saisie et de codage des données, mais permet aussi de ne pas retomber dans les travers liés à la constitution de séries homogènes. Les « sources nominatives », qu'il s'agisse d'actes d'état civil ou notariés, ou de dictionnaires biographiques anciens, ne sont pas plus transparentes que les séries statistiques du XIX^e siècle : elles ne peuvent être traitées qu'au prix d'une réflexion sur leur construction.

Microhistoire et sources nominatives

Depuis la naissance de techniques informatiques permettant de relier plusieurs informations à un même individu, les

« sources nominatives » ont été exploitées de façon à la fois qualitative et quantitative. Les plus utilisées sont présentes sur la plupart des terrains et à la plupart des époques : ce sont les actes de naissance, mariage, décès, vente, crédit... conservés par les notaires, les paroisses ou les mairies et dont l'étude a largement renouvelé tant la démographie historique que l'histoire économique et sociale.

Les microhistoriens italiens ont fait un grand usage de ces sources et des possibilités de changement d'échelle qu'elles offrent. Il est vrai qu'ils bénéficient de la richesse des archives de ce type en Italie [Chauvard, 2004]. Sur d'autres terrains, il peut être utile de se poser quelques questions en termes de rendement des dépouillements avant de se lancer. Une étude exhaustive ou même représentative est en général impossible : mieux vaut alors l'assumer, en adoptant une démarche expérimentale fondée sur un petit nombre de cas [Dolan, 2004].

Il faut également se souvenir que, bien qu'elles soient faciles à mettre en série, puisque structurées dès le départ en rubriques, ces sources n'ont pas été constituées pour répondre aux questions de l'histoire sociale. Un travers classique de la prosopographie est ainsi de vouloir remplir à toute force certaines « cases », comme « origine sociale » ou « fortune ». Cela peut amener à surinterpréter par exemple une fortune au décès, connue par les sources fiscales, mais qui correspond selon les âges à un maximum ou à ce qui reste après des donations ; ou encore une profession du père à la naissance de l'individu, alors que ce père peut avoir connu ensuite une importante mobilité sociale.

Seuls l'inventivité et les allers et retours entre sources disponibles et questions de recherche permettent de construire une histoire sociale, politique ou culturelle quantifiée à partir de données sur des individus. Ainsi, au lieu de traquer ces derniers un à un dans des sources particulièrement lourdes et ne fournissant que des renseignements très généraux, mieux vaut réfléchir en termes de rendement, de comparabilité des résultats et de spécificité des questionnements. Pour l'étude d'une institution, par exemple, on peut privilégier les questions de cumuls de mandats en croisant la liste d'individus étudiés avec les listes de membres d'autres institutions : la technique, relativement rapide, permet en outre de prendre en compte plusieurs dates. Constituer des indicateurs d'assiduité ou de travail dans l'institution peut également être au moins aussi important, pour

Le témoin, cet inconnu

Un risque classique de surinterprétation des actes d'état civil ou notariés concerne les témoins toujours mentionnés sur ces actes. Leurs liens avec les personnes concernées par l'acte sont parfois indiqués, en général brièvement, qu'il s'agisse d'un lien familial ou du statut d'« ami ». Il faut toutefois prendre garde à ne pas mettre en série ce type d'indication sans autre forme de procès, comme si elle renvoyait à un phénomène social homogène, et plus encore en faire un indicateur direct des « relations sociales » ou des « structures de l'amitié ».

Le phénomène de témoignage informe, certes, sur les frontières de la « famille » et son ouverture sur l'extérieur [Hontebeyrie, 2005], ou même sur les relations ou les frontières entre groupes sociaux. Mais son utilisation comme indice nécessite un traitement contrôlé, qui croise autant que

possible les sources et prenne en compte toutes les informations que l'on peut obtenir sur les circonstances de rédaction des actes [Garden, 1998]. Il faut ainsi s'interroger sur les possibilités concrètes de présence de témoins sur place (un cas classique est celui du commerçant voisin de la mairie qui témoigne pour ceux qui n'ont pas de proches présents...), le sens de mots comme « ami » ou « cousin » dans la société considérée, etc.

De plus, si les actes d'état civil ou notariés sont la seule trace qui nous reste de bien des individus, notamment de leur vie familiale, ils ne se réfèrent qu'à des moments ponctuels et à des événements précis. On a ainsi reproché au concept de « stratégie familiale », très présent en histoire de la famille, de donner des échanges au sein des familles une vision trop purement économique [Rosental, 2000]. Cette tendance est en partie liée à un biais de sources, centrées sur les ventes, contrats de mariage ou héritages.

élaborer une typologie des membres, que de rechercher des actes de mariage.

Échantillons et significativité

Si certaines méthodes formalisées, on va le voir, sont particulièrement adaptées au traitement d'informations denses sur de petits échantillons, si l'exhaustivité constitue en général une tentation plutôt qu'un choix raisonnable, cela ne veut pas dire pour autant que tout chiffre, calculé à partir de tout corpus, soit interprétable, encore moins généralisable. Malgré les spécificités des sources historiques, la connaissance de quelques principes élémentaires de statistique s'impose pour éviter soit de se lancer dans des dépouillements surdimensionnés, soit de surinterpréter une différence de quelques points de pourcentage.

Combien de cas prendre en compte ?

Il arrive, en prosopographie par exemple, que l'on s'intéresse à une population, comme celle des membres ou des dirigeants d'une institution sur quelques décennies, qui compte seulement quelques dizaines de personnes, définies par une liste claire. Il peut alors être raisonnable de chercher des renseignements sur tous, quitte, pour certaines sources plus lourdes, à se centrer sur un sous-corpus ou, surtout, à étudier plus en profondeur quelques cas, situés par rapport à l'ensemble.

La situation est différente quand on s'intéresse *a priori* à un groupe très vaste — les salariés d'une grande entreprise, les immigrés pendant les trente glorieuses, les étudiants dans l'Europe moderne... — pour lequel, souvent, on ne dispose pas d'un comptage d'ensemble, encore moins d'une liste.

La statistique nous fournit dans ce cas quelques garde-fous. Tout d'abord, la qualité des conclusions chiffrées obtenues à partir d'un échantillon aléatoire, c'est-à-dire la marge d'erreur qu'il faut appliquer aux pourcentages calculés sur la population observée pour les généraliser à la population d'ensemble visée, dépend de l'effectif de l'échantillon et pas du taux de sondage. Du point de vue de la significativité des résultats, étudier 1 000 des 60 000 galériens d'A. Zysberg, 1 000 des 15 000 sans-culottes de M. Vovelle ou bien 1 000 des 60 millions de Français revient strictement au même.

L'effectif total de 1 000 correspond en effet à une marge d'erreur usuellement considérée comme acceptable. À cette échelle, si on trouve un pourcentage de 20 % pour telle ou telle caractéristique dans l'échantillon, on peut considérer qu'il y a des chances raisonnables pour que la part de cette caractéristique dans la population entière soit comprise entre 19 et 21 %, de très bonnes chances pour qu'elle soit entre 18 et 22 % et une quasi-certitude qu'elle soit entre 15 et 25 % [Martin, 2005]. Ces marges d'erreur sont donc loin d'être nulles. Encore faut-il souligner que, si on s'intéresse à des comportements plus rares (avec un pourcentage observé de 5 %, par exemple), les marges d'erreur sont plus importantes.

Par ailleurs, si l'échantillon comporte seulement 100 cas, on n'a déjà plus du tout de certitude de se situer entre 15 et 25 % pour un pourcentage observé de 20 %. Ces marges, qui doivent être considérées comme indicatives, donnent la mesure des

risques de surinterprétation. Encore partent-elles du principe d'un échantillonnage aléatoire, rarement respecté dans les pratiques historiennes.

Échantillons et comparaisons

En fait, tout dépend de ce que l'on veut faire des données. En effet, il est rare, en histoire, de rechercher des sources pour 500 ou 1 000 personnes et de se contenter, ensuite, de caractériser ce groupe par deux ou trois pourcentages d'ensemble.

Or, si l'on veut pouvoir opposer des sous-groupes en son sein, il faut que chacun de ces sous-groupes soit raisonnablement représenté. Typiquement, si l'on s'intéresse au comportement des femmes et que celles-ci sont très peu présentes dans la population d'origine, il est indispensable de constituer un échantillon séparé de femmes, afin qu'elles soient en nombre suffisant (quelques centaines par exemple). On parle alors d'*échantillons stratifiés*, souvent employés aussi lorsque les données renvoient à plusieurs lieux différents. Par exemple, pour comparer les carrières des curés de trois doyennés du Brabant aux xv^e et xvi^e siècles, il faut constituer trois échantillons aléatoires, avec des taux de tirage différents, sans quoi les chiffres pour le plus petit doyenné ne seraient pas interprétables [Bijsterveld et Mandemakers, 1994].

En plus de ces comparaisons internes à l'échantillon étudié, il est souvent utile de le confronter à une population plus générale. Ainsi, il est délicat de tenir un discours sur les entrepreneurs étrangers si l'on ne dispose d'aucune donnée sur les entrepreneurs français de la même époque. Lorsque des données chiffrées directement comparables ne sont pas fournies par d'autres sources, il est utile d'étudier un *échantillon de contrôle* dans la population non ciblée à l'origine (hommes *versus* femmes, Français *versus* étrangers...) ou la population générale, quitte à ne saisir que quelques données pour comparaison.

Comment choisir un échantillon ?

Les effectifs donnés plus haut valent pour des échantillons aléatoires au sens statistique, c'est-à-dire issus d'un tirage au sort parmi la population totale. Cela suppose que l'on dispose d'une liste de cette dernière, que l'on puisse accéder aussi facilement

aux données sur tous les individus — et, en toute rigueur, que chaque individu puisse être tiré plusieurs fois, mais ce point est peu gênant en pratique. Ces conditions ne sont pas toujours réunies.

L'organisation matérielle des archives, en particulier, peut donner envie d'utiliser d'autres logiques de choix, par exemple de dépouiller un dossier sur cinq d'une série, ou de choisir tous les individus dont le nom commence par la même lettre. Ce sont parfois les seules possibilités ; il faut toutefois prendre garde aux biais d'échantillonnage, qui peuvent être graves. Une lettre de l'alphabet peut sur- ou sous-représenter une origine nationale ou régionale ; des dossiers régulièrement répartis, s'ils sont issus d'une liste par ordre chronologique, privilégieront certaines périodes. Les échantillons fondés sur des *pas* (saisir une personne toutes les 300 entrées, par exemple) ne prennent pas en compte les rythmes d'inscription, qui obéissent bien souvent à des logiques propres.

C'est pourquoi l'aléatoire pur, quand il est possible, est toujours moins risqué ; tout tableur peut fournir à la demande une série de nombres aléatoires. Un tirage aléatoire n'empêche pas de cibler une sous-population d'étude dans une source plus générale : on sélectionne alors le premier étranger, la première femme, etc. que l'on trouve *après* le dossier ou le nom correspondant au nombre aléatoire tiré. Si tout choix aléatoire est impossible, il faut au moins veiller à ce que les biais de sélection de l'échantillon ne soient pas liés au thème étudié.

L'aspect longitudinal des données historiques pose d'autres problèmes qui ne doivent pas être éludés. Sélectionner un échantillon à partir de certaines sources privilégie mécaniquement, même avec un tirage aléatoire, les individus les plus stables ou dotés de la plus grande longévité. Il en va ainsi des dossiers de retraite. Pour résoudre ce problème, il faut si possible croiser les sources, sinon constituer des échantillons stratifiés, avec une surreprésentation volontaire des individus moins stables.

Enfin, la structure de la source et le temps de dépouillement disponible peuvent imposer des coupes dans le temps, visant à une comparaison entre échantillons issus de certaines années, voire de certains mois ou jours, tandis que les autres dates sont laissées de côté. La prudence s'impose alors dans la généralisation des résultats à ces autres dates, puisque les évolutions ne sont pas forcément linéaires.

Présenter des résultats en pourcentages

Avant tout, il est utile de respecter quelques bonnes pratiques en matière de pourcentages, qui améliorent la lisibilité des résultats et évitent d'en exagérer la précision.

1. Tout tableau en pourcentages doit indiquer l'effectif global de la population étudiée (souvent désigné comme « N »). S'il est inférieur à 100, donc si 1 % représente un individu ou moins, il n'est pas raisonnable de donner des pourcentages avec décimales, même si elles sont souvent fournies par l'ordinateur. De même, pour moins de 1 000 cas, une seule décimale suffit.

2. Les pourcentages, proportions issues de divisions, ne peuvent pas faire l'objet de simples additions ou soustractions. De ce fait, l'expression « points » a été créée pour rendre compte d'écart entre pourcentages. On dit donc qu'une valeur qui passe de 10 à 20 % a augmenté de 100 % ou de 10 *points*.

3. On utilise en général deux types de pourcentages. Ceux qui sont issus de *tris à plat* représentent la *distribution* de la population étudiée entre les *modalités* (les différentes valeurs) d'une *variable*. Exemple :

Activités des membres de la chambre de commerce de Paris (1803-1870)

Banquiers	Négociants	Autres	Total (N = 157)
31 %	31 %	38 %	100 %

Les tris à plat donnent une première idée des données mais, en général, on s'intéresse plus aux *tris croisés*, qui présentent en même temps la distribution de deux variables, donc donnent une première idée des liens, ou de l'absence de liens, entre elles. Fournir ces tris croisés en pourcentages plutôt qu'en nombres absolus permet au lecteur de se rendre compte des particularités de certaines parties de la population par rapport à l'ensemble, pour peu que l'on utilise judicieusement les *pourcentages en ligne ou en colonne*. Pour lire de tels tableaux, il faut d'abord regarder la ligne ou la colonne qui donne le tri à plat pour toute la population, puis y comparer le *profil* des autres lignes ou colonnes. Exemple :

Quand les chiffres ont-ils un sens ?

Une fois un échantillon raisonnable choisi et les données traitées, il est bon d'indiquer au lecteur quel degré de confiance il peut accorder aux différents chiffres qui lui sont présentés. La pratique est pourtant encore rare en histoire, où l'on commente souvent à l'infini des différences de quelques points de pourcentage — d'autant plus qu'elles ont été obtenues non sans peine, au prix de semaines voire de mois de travail de saisie. Or des outils existent, qui sont suffisamment utiles et accessibles pour justifier un petit effort d'apprentissage.

Activités et périodes d'entrée des membres à la chambre de commerce de Paris (1803-1870)

	Banquiers	Négociants	Autres	Total
1803-1813	43 %	34 %	23 %	100 %
1814-1831	54 %	26 %	21 %	100 %
1832-1848	16 %	39 %	45 %	100 %
1849-1869	13 %	29 %	58 %	100 %
Total (N = 157)	31 %	31 %	38 %	100 %

	Banquiers	Négociants	Autres	Total (N = 157)
1803-1813	31 %	24 %	14 %	22 %
1814-1831	47 %	22 %	15 %	27 %
1832-1848	10 %	24 %	24 %	20 %
1849-1869	12 %	29 %	47 %	31 %
Total	100 %	100 %	100 %	100 %

C'est le repérage des cases « 100 % » qui nous permet de dire que le premier tableau est en *pourcentages en ligne*, le second en *pourcentages en colonne*. Le premier compare les profils d'activité selon la date d'entrée, le second les profils de date d'entrée selon l'activité. Certains chiffres paraissent remarquables. Par exemple, 54 % des membres entrés en 1814-1831 sont des banquiers, contre 31 % de l'ensemble des membres et 13 % seulement de ceux entrés en 1849-1869. Et 47 % des « autres » sont entrés en 1849-1869, contre 31 % de l'ensemble des membres et 12 % seulement des banquiers. Cependant, l'effectif total peut nous faire douter de l'intérêt de ces résultats. Après tout, les banquiers entrés en 1849-1869 ne sont que six. Que peut-on en conclure ? À partir de quel écart en pourcentage peut-on se permettre un commentaire, qui évoquerait ici des transformations importantes du recrutement de l'institution [Lemerrier, 2003] ?

Il existe des règles pour interpréter les pourcentages en fonction des effectifs. Ce sont nécessairement des conventions, qui fixent des seuils de significativité en fonction des marges d'erreur considérées comme admissibles. En tant que telles, elles sont néanmoins fort utiles. Ainsi, *grosso modo*, on peut considérer que si l'on compare deux sous-groupes de 100 personnes chacun, il faut une différence de 12 à 14 points entre les pourcentages caractérisant chacun de ces groupes pour que l'on puisse se permettre de considérer ces pourcentages comme réellement différents avec certitude. Pour deux groupes de 1 000 personnes chacun, 4 points de différence

L'analyse de la variance

Une des techniques de test de la significativité d'écart entre résultats a été parfois utilisée par les historiens, surtout dans les années 1980 : l'analyse de la variance, ou ANOVA [Cellier et Cocard, 2001].

Son objectif est de déterminer si les valeurs moyennes d'une variable sont significativement différentes pour plusieurs sous-populations que l'on compare. Contrairement aux tableaux croisés évoqués jusqu'ici, qui traitent de variables quantitatives ou qualitatives réparties en quelques *classes* ou *modalités*, l'ANOVA s'applique à des variables *quantitatives*, par exemple un montant de fortune ou d'impôt.

Si on observe le montant moyen de l'impôt payé par les contribuables de quelques villages, on peut noter des différences de quelques livres : 9 livres dans un village, 13 ailleurs, 20 dans un troisième cas, etc. Peut-on en conclure que ces villages sont très différents de ce point de vue, et donc

essayer de comprendre ces différences, ou bien doit-on se méfier de ces chiffres ? L'écart pourrait en effet simplement être dû au hasard, du fait d'un trop faible nombre d'observations dans chaque village.

Pour répondre à cette question, on décompose les différences de niveau de la variable que l'on observe : une partie de ces différences relève des écarts entre groupes, ici les villages (elles rendent compte de la différence entre les moyennes de ces groupes) ; une autre partie s'explique par d'autres différences, à l'intérieur de chaque groupe.

Comme pour le chi-2, le résultat essentiel est une probabilité, celle que les différences entre groupes soient dues seulement au hasard des variations entre individus. Si cette probabilité est très faible, on peut conclure que les différences entre groupes sont significatives, par exemple que les habitants de plusieurs villages sont inégalement imposés, même si dans chaque cas les inégalités internes sont fortes.

suffisent ; 7 points pour des groupes de 300 et 500 personnes, etc. [Martin, 2005].

Le chi-2

Plus précisément, et en particulier quand on travaille sur des groupes plus restreints, il est possible, directement en ligne ou par quelques manipulations assez simples dans un tableur [Cellier et Cocard, 2001 ; Saly-Giocanti, 2005], de calculer des indicateurs de liaison entre deux variables, en particulier le chi-2. Celui-ci permet de dire avec une certitude raisonnable si les différences observées dans un tableau croisé peuvent être dues au simple hasard, du fait de faibles effectifs, ou si elles renvoient à une *attraction* ou une *opposition* réelle entre les modalités de variables considérées [Cibois,

2004]. Il restera ensuite, bien sûr, à interpréter ces liens entre variables, qui ne renvoient pas forcément à une causalité.

Le chi-2 — aussi noté « khi-deux », « chi carré », χ^2 — est une mesure des écarts entre la situation observée et une situation imaginaire. Cette situation, appelée en général *théorique*, est celle où il n'y aurait aucun lien entre les variables considérées : elles seraient totalement indépendantes, il n'y aurait pas de modalités associées de façon préférentielle. Dans notre exemple, il y aurait *indépendance* parfaite entre activité et période d'entrée si, pour toutes les périodes, on retrouvait à la chambre de commerce 31 % de banquiers, 31 % de négociants et 38 % d'autres activités. Le calcul de chi-2 compare les effectifs *théoriques*, ceux qui correspondent à cette situation d'indépendance imaginaire, aux effectifs *observés*, pour juger de la plausibilité de l'indépendance.

Le chiffre important renvoyé par le logiciel est alors une *probabilité*, souvent notée *p* ou *prob*. C'est la probabilité que les chiffres que nous observons correspondent à la situation d'*indépendance*, avec simplement une petite différence pouvant être due au hasard. La question est : est-il vraisemblable que, compte tenu de ce que nous observons, les deux variables soient indépendantes ? La probabilité représente ce degré de vraisemblance.

De façon conventionnelle, on considère en général que cette probabilité doit être inférieure à 5 % pour que l'on en conclue que les variables sont liées. C'est-à-dire qu'il y a moins de 5 % de risques pour que la liaison que nous avons l'impression de voir soit simplement due aux trop faibles effectifs observés. Parfois, des seuils de 1 % ou 10 % sont aussi utilisés.

Il faut toutefois souligner que la technique du chi-2 trouve ses limites pour les effectifs vraiment faibles : typiquement, quand certaines cases du tableau croisé renvoient à un effectif théorique, en situation d'indépendance, de moins de cinq cas, il faut se méfier des résultats [Martin, 2005]. Selon les effectifs et les variables considérés, d'autres tests que le chi-2 peuvent s'imposer [Georgin et Gouet, 2005].

Chi-2 et significativité des résultats

La question de la significativité statistique d'un tri croisé peut avoir un véritable sens historique, donnant au chi-2 un statut argumentatif fort. Il peut ainsi aider à comprendre la fabrication du recensement des juifs lensois par les autorités préfectorales à la fin de 1940, notamment le rôle de l'autodéclaration [Mariot et Zalc, 2007].

On peut en particulier se demander si certaines caractéristiques sociales des individus jouent un rôle dans le fait que les chefs de ménage identifiés, à un autre moment de la guerre, comme « juifs » par l'administration se déclarent ou non comme tels en décembre 1940. Pour le savoir, on construit des tableaux qui croisent les caractéristiques des chefs de ménage et le fait de se déclarer ou non ; pour statuer sur l'ampleur des différences entre pourcentages dans ces tableaux, eu égard aux effectifs concernés (parfois seulement quelques dizaines de cas), il est utile d'y ajouter un test de chi-2.

Ainsi, on peut observer un lien entre déclaration et composition du foyer : d'après le test, les *écarts à l'indépendance* sont significatifs au seuil de 1 % (ce qu'on indique souvent par trois étoiles dans les publications). En revanche, l'âge n'a pas d'effet significatif (mention « NS »). Notons que les effectifs totaux diffèrent légèrement, car l'information manque sur certains âges.

Composition du foyer***

	Célibataire	Couple sans enfants	Foyer avec enfants	Total
Non déclarants identifiés	51 %	8 %	41 %	100 % (N = 37)
Déclarants	16 %	14 %	70 %	100 % (N = 150)
Total	23 %	13 %	64 %	100 % (N = 187)

Âge du chef de ménage en 1942 (NS)

	16-30 ans	31-45 ans	46 ans et plus	Total
Non-déclarants identifiés	18 %	53 %	29 %	100 % (N = 34)
Déclarants	8 %	53 %	39 %	100 % (N = 148)
Total	10 %	53 %	37 %	100 % (N = 182)

L'absence de significativité statistique peut elle-même constituer un résultat important.

Déportation des juifs lensois (NS)

	Non-déportés	Déportés	Total
Non-déclarants identifiés	43 %	57 %	100 % (N = 37)
Déclarants	46 %	54 %	100 % (N = 150)
Total	45 %	55 %	100 % (N = 187)

Ici, un troisième test de chi-2 conclut à une probable indépendance entre le fait de se déclarer ou de ne pas le faire et le fait d'être ou non déporté par la suite : ne pas se déclarer ne protège pas de la déportation — qui, en revanche, selon un quatrième tableau non reproduit ici, frappe plus durement les étrangers que les Français.

Ce résultat permet d'écarter résolument tout jugement de valeur quant au « choix » d'autodéclaration des chefs de famille.