

L'ANALYSE DE SONDAGE PAR ORDINATEUR: EXPÉRIMENTATION D'UNE ACTIVITÉ DE LABORATOIRE EN SOCIOLOGIE

par JACQUES FOURNIER

professeur de sociologie
Cégep de Sherbrooke

Le projet de recherche présenté à la DGEC*, l'année dernière, comportait deux grands objectifs : une étude de faisabilité pour déterminer parmi les bases de données numériques pertinentes à l'enseignement de la sociologie celles qu'il est techniquement et financièrement possible d'utiliser dans le cadre d'une activité de laboratoire de sociologie et une expérimentation pédagogique pour réaliser concrètement un laboratoire informatisé de sociologie, évaluer l'apprentissage des élèves et connaître les réactions que suscite cette méthode d'enseignement.

Dans les pages qui suivent, nous fournissons d'abord des précisions sur le sens de quelques réalités mises en cause dans notre recherche. Nous décrivons ensuite les caractéristiques principales des bases de données en liaison directe et les avantages que présente l'achat d'un fichier de données. Pour terminer, nous présentons les objectifs et les modalités de réalisation de l'expérimentation pédagogique que nous avons menée dans le cadre d'un cours d'initiation à l'analyse sociologique.

* Cet exposé résume le rapport d'une recherche effectuée grâce à une subvention de la DGEC dans le cadre du programme PROSIP. Ce rapport publié au mois d'août 1985 au Collège de Sherbrooke s'intitule : L'analyse de sondage par ordinateur : expérimentation d'une activité de laboratoire en sociologie.

1. Principales caractéristiques des bases de données recherchées

Avant d'établir les caractéristiques des bases de données utiles à l'enseignement, nous aimerions préciser le sens de quelques mots.

Banque de données (*databank*) : Un ensemble d'informations directement exploitables et recouvrant un domaine particulier de connaissances.

Base de données (*database*) : Un rassemblement des données nécessaires au traitement d'une application sous tous ses aspects. La distinction entre banque de données et base de données n'est pas très nette. On peut considérer que le fichier constitue la base de données et que l'organisation qui permet de traiter le fichier constitue la banque de données. Par exemple « CANSIM » est la base de données de Statistique Canada, la banque de données est la compagnie qui permet d'accéder à CANSIM.

Fichier (*file*) : Ensemble d'informations de même nature ou concernant un même sujet.

Progiciel (*package*) : Produit et logiciel, ensemble de programmes cohérent et indépendant, facilement utilisable pour réaliser des traitements informatiques standard. Exemple : SPSS, STATPAC...

Les bases de données en liaison directe (*online*)

On distingue habituellement deux grands types de bases de données : les bases de données de référence (*reference data-bases*) et les bases de données informatives (*sources data-bases*)*.

Les bases de données de référence sont les plus connues, elles peuvent contenir des fichiers de référence bibliographiques ou des fichiers d'un autre type de référence comme un annuaire des professionnels, des industries, etc.

Les bases de données informatives peuvent contenir des fichiers de données numériques, ce qui nous intéresse plus particulièrement, comme les données originales d'un sondage ou les statistiques produites par une manipulation des données originales. Les bases de données informatives peuvent aussi contenir des fichiers de données à la fois textuels et numériques et des fichiers uniquement textuels, c'est-à-dire un fichier plein texte.

La liaison directe à une base de données est très développée dans le cas des bases de données bibliographiques et dans le cas des bases de données numériques reliées au monde des affaires et du commerce. La liaison directe est plus restreinte dans le cas des données numériques à caractère sociologique.

La liaison directe offre des avantages considérables comme la mise à jour régulière des informations, la facilité du traitement. Par contre, le coût d'utilisation — payer pour chacune des demandes — et la rareté des informations à caractère sociologique nous incitent à suggérer d'acheter les fichiers de données numériques sous forme de bandes magnétiques ou de disquettes.

L'achat d'un fichier de données

L'achat d'un fichier de données permet d'accéder à une plus grande variété d'informations. De plus, l'utilisateur peut effectuer plusieurs traitements sans coût supplémentaire. Il doit cependant être capable d'utiliser des logiciels de traitements comme SPSS ou STATPAC.

Les données numériques en sociologie peuvent servir à décrire l'écologie sociale. On utilise des taux, des indicateurs ou des caractéristiques démogra-

* *Directory of online databases*, volume 4, no 3, Printemps 1983, CUADRA Associates Inc., California. Cet ouvrage contient des informations au sujet de plus de 1600 bases de données à travers le monde.

phiques pour décrire une population. L'ensemble des faits sociaux ainsi disponibles est très varié. Malheureusement, les données du recensement ne sont pas encore disponibles sous formes de disquettes prêtes à être utilisées. On peut cependant espérer que la situation change comme c'est le cas aux États-Unis. (Certaines données du recensement américain sont maintenant disponibles sous forme de disquettes de micro-ordinateur.)

Les sondages d'opinion scientifique* constituent une deuxième forme de données sociales pertinentes à l'enseignement. Les sondages de qualité sont très variés et mesurent des phénomènes importants pour un sociologue.

Conclusion

L'enseignant qui est à la recherche de données numériques pour décrire les faits sociaux a d'abord tendance à vouloir utiliser les banques de données commerciales en liaison directe. Les informations sont récentes et assez faciles à obtenir. Malheureusement ces banques de données contiennent peu d'informations à caractère sociologique et sont encore trop coûteuses pour être utilisées par les élèves.

L'achat d'un fichier de sondage ou de recensement est la meilleure façon d'accéder aux bases de données numériques typiques de la sociologie. Il permet de constituer une activité de laboratoire de sociologie où les élèves pourront effectuer des commandes individuelles de traitement.

2. Expérimentation pédagogique

Le but de l'expérimentation pédagogique était d'évaluer les possibilités d'utilisation des résultats d'un sondage scientifique pour constituer un exercice de laboratoire dans le cadre d'un cours d'initiation à l'analyse sociologique.

L'analyse des résultats d'un sondage scientifique permet d'initier les élèves à l'analyse par variables et aux modèles d'explications caractéristiques de la sociologie empirique. L'exercice de laboratoire informatique permet d'utiliser une pédagogie active (faire des hypothèses et les vérifier) pour analyser une question sociale.

L'expérimentation visait à déterminer si les élèves possèdent les habiletés nécessaires pour effectuer l'analyse des résultats d'une enquête. Il s'agissait

* Le Centre de Sondage de l'Université de Montréal vend les fichiers des enquêtes qu'il effectue régulièrement auprès de la population du Québec.

de connaître les difficultés qu'ils rencontrent, mesurer l'apprentissage qu'ils réalisent et observer l'intérêt et la réaction des élèves devant cette activité.

Cadre de l'expérimentation

Moment : 4 rencontres de trois heures, mars 1985.

Groupe : 2 groupes d'étudiants en sciences humaines sans mathématiques dans le cadre d'un cours d'initiation à la sociologie.

Modalités : travaux d'équipe, commandes de tableaux différents pour chacune des équipes.

Sondage : Omnibus (*i.e.* un sondage comprenant des questions sur divers sujets) — automne 1983, Centre de Sondage de l'Université de Montréal. Sondage représentatif de l'ensemble de la population du Québec, 2053 répondants.

Étapes de l'expérimentation

Choix d'une variable dépendante. L'exercice a été conçu comme une analyse de l'effet de variables explicatives sur une variable dépendante. Le sondage retenu contient quatre questions concernant l'enseignement au Québec. Ces questions simples et générales suscitent habituellement l'intérêt des élèves*.

Q. 18 Dans votre région, considérez-vous comme très bon, plutôt bon, plutôt mauvais ou très mauvais l'enseignement donné : a) à l'école primaire ? b) à l'école secondaire ? c) au cégep ? d) à l'université ?

Q. 19 Êtes-vous plutôt satisfait ou plutôt insatisfait du travail fait auprès des élèves par les : a) enseignants de l'école primaire ? b) enseignants de l'école secondaire ? c) non-enseignants de l'école primaire ? d) non-enseignants de l'école secondaire ? e) directeurs d'école au primaire ? f) directeurs d'école au secondaire ?

Q. 20 Dans l'ensemble, trouvez-vous qu'on est trop exigeant, assez exigeant ou pas assez exigeant dans les examens passés par les élèves à la fin de : a) l'école primaire ? b) l'école secondaire ? c) du cégep et du collège ? d) de l'université ?

Q. 21 Au cours des dix dernières années, croyez-vous que l'éducation s'est améliorée, est restée pareille ou s'est détériorée : a) dans les écoles primaires ? b) dans les écoles secondaires ? c) dans les cégeps et collèges ? d) dans les universités ?

* Le cahier d'exercices remis aux élèves contient l'ensemble de ces questions et les fréquences correspondantes.

Choix des caractéristiques sociales des répondants. L'analyse par variables consiste à mesurer l'effet d'une variable indépendante (dans notre cas les caractéristiques sociales des individus) sur la variable dépendante. Les sondages omnibus contiennent des informations concernant le sexe, l'âge, le niveau de scolarité, l'occupation, le revenu, etc. des répondants.

Formulation d'une hypothèse et commande d'un tableau croisé. Après avoir choisi la variable dépendante à étudier et avoir identifié les variables indépendantes, l'élève devait rédiger une hypothèse qui précise la réponse attendue. Voici à titre d'exemple une hypothèse qui précise la relation attendue entre la question 20C et le niveau de scolarité : « Les répondants qui ont une scolarité plus élevée auront tendance à trouver que les professeurs ne sont pas assez exigeants dans les examens passés par les élèves à la fin du cégep. »

EXIGENCES DANS LES EXAMENS PASSÉS PAR LES ÉLÈVES AU CÉGEP SELON LA SCOLARITÉ DES RÉPONDANTS

Q. 20C	Nombre d'années de scolarité	
	11 ans et moins	12 ans et plus
trop exigeant	47 (10,5)	33 (4,5)
assez exigeant	301 (67,3)	448 (61,1)
pas assez exigeant	99 (22,2)	252 (34,4)
TOTAL	447	733

Significatif à .0000

Observations manquantes : 868

Source : Centre de Sondage de l'Université de Montréal, Omnibus-universitaire, Automne 1983.

Analyse des résultats. L'analyse des résultats amène les élèves à confirmer ou à rejeter leur première hypothèse. Une deuxième, une troisième commandes à l'ordinateur permettent de vérifier des hypothèses reformulées.

Déroulement et résultats de l'expérimentation

L'analyse des résultats d'une enquête sociale suppose la maîtrise de trois types d'habiletés :

Formulation d'une hypothèse. Les élèves possèdent les habiletés nécessaires pour formuler des hypothèses suffisamment précises pour qu'elles puissent être vérifiées à l'aide d'un tableau croisé. Peu habitués à une logique typique des sciences expérimentales, les élèves ont cependant besoin de quelques heures pour apprendre à formuler une réponse attendue.

Description d'une relation entre deux variables. Il est étonnant de constater que les élèves de sciences humaines n'éprouvent pas de difficultés à comprendre les tableaux croisés. D'une façon générale, ils décrivent correctement les relations entre deux variables. Certains insistent sur des points marginaux mais la majorité des élèves décrivent bien la tendance de la relation.

Explication d'une relation entre deux variables. Les élèves possèdent les habiletés requises pour construire un modèle explicatif simple. Ces explications sont faibles dans les cas plus complexes.

Un élève qui désirait vérifier si la croyance religieuse influence l'attitude au sujet de la peine de mort* écrit : « Notre hypothèse est confirmée en partie car le tableau dévoile, peu importe la religion, que le sujet de la peine de mort demeure personnel. Mais le tableau nous dévoile une surprise : ceux qui n'ont pas de religion semblent moins favorables à la peine de mort que ceux qui en ont une. Inexplicable. »

L'enseignement consiste à fournir aux élèves les informations nécessaires pour expliquer ces relations à première vue inexplicables.

Conclusion de l'expérimentation

L'expérimentation pédagogique permet de dégager les points forts et les limites de l'analyse de sondage par ordinateur.

Points forts. L'analyse d'une enquête sociale permet de :

- susciter l'intérêt : permet de voir que les opinions varient selon les différents milieux sociaux ;
- susciter la curiosité : permet de développer un questionnement sociologique ;
- favoriser l'intégration d'une connaissance d'ordre général : formuler des hypothèses et développer un modèle d'analyse simple pour expliquer les relations observées ;
- développer la rigueur dans la description et l'analyse sociale.

Limites. L'expérimentation pédagogique montre que les élèves ont de la difficulté à intégrer en même temps la problématique, le questionnaire et l'analyse par variables. Les élèves ont de la difficulté à élaborer un modèle d'analyse complexe et à effectuer une synthèse. Les progiciels d'analyse statistique, suffisam-

ment « performants » pour traiter un grand nombre de cas, ne sont pas interactifs. Le contact avec l'ordinateur est beaucoup moins intéressant que dans le cas d'un jeu éducatif.

3. Recommandations

L'exercice de laboratoire avait été conçu comme un moyen d'analyser l'effet des variables indépendantes sur une variable dépendante. Il s'agit alors de découvrir l'ensemble des variables explicatives.

Pour aider les élèves à développer des modèles explicatifs, il faudrait effectuer l'analyse de l'effet de quelques variables sociologiquement plus importantes sur plusieurs variables dépendantes et même dans plusieurs sondages. En mettant l'accent sur les variables indépendantes on pourrait mieux développer les modèles d'explications qui font appel aux caractéristiques sociales les plus importantes.

Il est nécessaire de fournir aux élèves un guide d'analyse qui présente la problématique de l'enquête et les principales formes d'explication du phénomène étudié.

Conclusion

Cette recherche a permis de constater que les grandes banques de données en liaison directe (*online*) ne répondent pas adéquatement aux besoins des professeurs de sociologie. Par contre, les fichiers de sondages ou de recensements disponibles sous forme de bandes magnétiques ou disquettes de micro-ordinateurs constituent des bases de données numériques très utiles dans l'enseignement.

Pour assurer une utilisation efficace des résultats d'une enquête sociale dans un laboratoire de sociologie, il faudrait que les données soient faciles d'utilisation. Par exemple, une disquette d'ordinateur prête à être étudiée à l'aide d'un progiciel d'analyse statistique. Un cahier d'exercices devrait présenter la problématique de l'enquête, résumer les principales théories explicatives et suggérer la démarche à suivre pour effectuer l'activité de laboratoire.

Ces ensembles de données sociales pourraient être développées prioritairement dans le cas des thèmes d'étude les plus « classiques » en sociologie. Il faudrait aussi choisir des domaines d'étude où les données sont mises à jour régulièrement, comme c'est le cas en sociologie électorale, en sociologie de l'éducation, etc.

* Il s'agit dans ce cas d'un autre sondage.