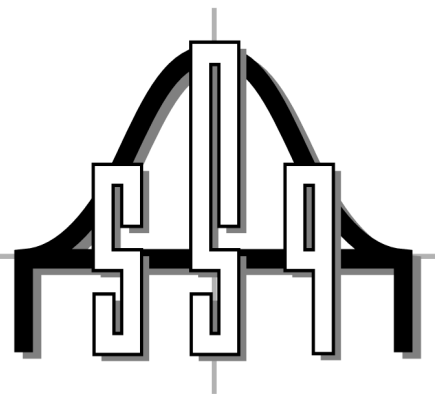


# Convergence

Le journal de l'Association des statisticiennes et statisticiens du Québec



Volume VIII • Numéro 2

Août 2003

Dans le précédent numéro de Convergence, je faisais allusion aux préparatifs entourant l'organisation d'un événement qui était prévu pour mai 2003, un projet que l'exécutif de l'association avait mis de l'avant pour tenir la rencontre annuelle des membres et par le fait même, notre assemblée générale annuelle. Malheureusement, faute d'inscriptions en nombre suffisant, ce projet n'a pu se concrétiser comme nous le souhaitions.

À l'occasion de l'assemblée générale annuelle, qui s'est finalement tenue à Québec en juin dernier, nous en avons profité pour discuter avec les membres présents de ce qui pourrait être fait dans le futur pour augmenter les chances de succès d'un tel événement. L'exécutif a évidemment pris bonne note des remarques et suggestions qui ont été formulées par les membres. Pour les deux prochaines années, l'association vise à s'adjoindre à d'autres organisations pour tenir notre rencontre annuelle. Ainsi, en 2004, la Société statistique du Canada tiendra son Congrès annuel à Montréal et nous sommes actuellement en pourparlers pour nous greffer à ce congrès. Pour 2005, le Colloque francophone sur les sondages, réunissant entre autres des membres de la Société française de statistique, aura lieu à Québec, et l'ASSQ y est déjà représentée au sein du comité organisateur.

Nous croyons qu'en agissant de la sorte, nous mettrons toutes les chances de notre côté pour obtenir une participation maximale des membres à notre rencontre annuelle. Évidemment, ceci n'exclut d'aucune façon que l'association organise certaines journées de formation à l'échelle

régionale ou provinciale, si la demande pour de tels événements le justifie, cela va de soi !

Je termine en vous invitant à vous manifester sur le groupe de discussion (envoyez un courriel à l'adresse : [association\\_assq@groupeyahoo.ca](mailto:association_assq@groupeyahoo.ca)). Pourquoi ne pas profiter de la période estivale pour commenter la position décrite plus haut, ou encore suggérer des activités que vous souhaiteriez voir mises en branle par l'association pour les prochains mois ?

Bonne fin d'été à tous et à toutes !

Sylvain Végiard, président ■

## Dans ce numéro :

<b>Mot de la rédactrice</b>	3
<b>À propos de l'ASSQ</b> (Claude Ouimet, Mike Sirois)	4
<b>Histoire du mot imputation</b> (Éric Rancourt)	6
<b>Livre « Exercices corrigés de méthodes de sondage » de Ardilly et Tillé</b> (Pierre Lavallée)	7
<b>La chronique SAS</b> (Caty Banchette)	8
<b>Les fleurs et le pot</b> (Pierre Lavallée)	10
<b>Nettoyer une liste de noms : l'algorithme du « soundex »</b> (Thierry Petitjean-Roget)	11
<b>Conférences à venir</b>	15
<b>Suivre son cours</b>	16

# CONVERGENCE

*Convergence*, le journal de l'Association des statisticiennes et statisticiens du Québec (ASSQ), est publié trois fois par année en avril, août et décembre. Il est distribué gratuitement aux membres de l'ASSQ.

## Rédactrice en chef

Myrto Mondor, Centre de recherche, hôpital du St-Sacrement

## Rédacteurs adjoints

Caty Blanchette, Centre de recherche, hôpital du St-Sacrement

Sylvie Gauthier, Statistique Canada

Isabelle Marchand, Statistique Canada

Steve Méthot, Agriculture et agroalimentaire Canada

## AVIS AUX AUTEURS

La rédaction de *Convergence* invite les statisticiens et toutes les personnes intéressées par la statistique et ses applications à lui faire parvenir leurs articles, questions, commentaires, soumissions et résolutions de problèmes. Les textes doivent être soumis sous forme de fichiers de traitement de texte Microsoft Word. Les auteurs doivent faire parvenir, à l'adresse de l'ASSQ, leur fichier sur disquette 3,5 pouces (ou un fichier convertible et lisible sur courrier électronique) ainsi qu'une copie papier de leur texte avant la date de tombée du prochain numéro. La rédaction ne s'engage pas à publier tous les textes reçus et se réserve le droit de n'en publier que des extraits sur approbation de l'auteur.

## AVIS AUX ANNONCEURS / EMPLOYEURS

Les entreprises ou les personnes qui désirent faire paraître de la publicité ou des offres d'emploi dans *Convergence* doivent faire parvenir, à l'adresse de l'ASSQ, leur document prêt pour l'impression avant la date de tombée du prochain numéro. Les tarifs pour la parution dans un numéro de *Convergence* sont les suivants :

Tarifs	page intérieure	endos(publicité seul.)
Carte d'affaires	15 \$	20 \$
1/4 page	40 \$	50 \$
1/2 page	80 \$	100 \$
page entière	150 \$	200 \$

Note liminaire: la forme masculine est employée dans le but d'alléger le texte et désigne les deux sexes, à moins d'une mention contraire de l'auteur.

La rédaction de *Convergence* laisse aux auteurs l'entière responsabilité de leurs opinions. La reproduction des articles est autorisée, sous réserve de mention de la source.

## Toute correspondance doit être adressée à :

### *Convergence*

Association des statisticiennes et statisticiens du Québec

Boîte postale 94

Loretteville (Québec), G2B 3W6, Canada

Fax/répondeur : (418) 871-1945

Adresse élect. : [association\\_assq@yahoo.ca](mailto:association_assq@yahoo.ca)

Page internet : <http://www.association-assq.qc.ca>

Dépôt légal - Bibliothèque nationale du Québec, 2003

## MEMBRES INSTITUTIONNELS :

Québec 

Institut de  
la statistique  
du Québec



Statistique  
Canada

UNIVERSITÉ  
LAVAL

Département de mathématiques et de statistique



LES SERVICES CONSEILS

**HARDY**  
CONSULTING SERVICES

NOVAXA  
Conseil



SOCIÉTÉ DES CASINOS  
DU QUÉBEC INC.

## Mission

L'ASSQ a pour mission de regrouper les statisticiennes et les statisticiens de tous les domaines en vue de promouvoir la statistique et d'en favoriser la bonne utilisation.

## Membres

L'ASSQ offre deux catégories de membres aux personnes intéressées par ses activités :

Membre statisticien : Toute personne possédant au moins un baccalauréat en statistique ou l'équivalent (baccalauréat avec au moins 24 crédits de cours reconnus en statistique ou probabilité). Les personnes ne répondant pas à cette condition peuvent accéder à la catégorie de membre statisticien si leur expérience professionnelle est jugée équivalente aux connaissances acquises lors de la formation académique.

Membre affilié : Toute personne qui souhaite faire partie de l'ASSQ.

Frais d'adhésion annuels pour chacune des deux catégories :  
50 \$ (régulier) 20 \$ (étudiant)

Les organismes peuvent devenir membres institutionnels de l'ASSQ au coût de 300 \$ par année et ainsi bénéficier de plusieurs privilèges dont l'adhésion gratuite comme membres statisticiens ou affiliés pour trois de leurs employés.

## Conseil d'administration

Président : Sylvain Végiard (*Institut de la statistique du Québec*)

Vice-président : Michel Guillet (*Creascience*)

Secrétaire : Claude Ouimet (*Ministère des transports du Québec*)

Trésorière : Isabelle Auger (*Ministère des ressources naturelles du Québec*)

Registraire : Christine Gamelin (*Bell Canada*)

Directeur des communications : Mike Sirois (*Statistique Canada*)

# Mot de la rédactrice

Bonjour,

Après un été où les activités sont habituellement ralenties, voici votre numéro de Convergence du mois d'août qui, espérons-le, vous aidera à repasser vos neurones du mode vacances au mode travail.

Une chose est sûre, Pierre Lavallée, un collaborateur fréquent et de longue date de Convergence, était bien éveillé lorsqu'il a étudié le livre *Exercices corrigés de méthodes de sondage* d'Ardilly et Tillé. Il nous en fait un compte-rendu détaillé, nous permettant par la même occasion de nous rafraîchir la mémoire et d'avoir un aperçu des derniers développements dans le domaine des sondages. Dans ce numéro, vous retrouverez aussi la chronique historique qui cette fois nous relate l'origine du mot "imputation". La chronique informatique nous explique différents algorithmes qui permettent de regrouper des mots similaires phonétiquement. Cette technique peut s'avérer utile lorsque l'on doit fusionner deux banques de données sur la base de variables telles que nom et prénom plutôt que sur une bonne vieille variable de numéro d'identification unique.

La chronique "Les fleurs et le pot" nous rapporte encore une fois le fruit de la vigilance d'un de nos membres. J'aime bien cette chronique puisqu'elle fait le lien entre l'univers parfois fermé où nous produisons des résultats statistiques et le vaste monde où ces résultats sont utilisés. Les échos du CA et de l'assemblée annuelle font quant à eux le lien entre votre association et vous tandis que Claude Ouimet nous encou-

rage à créer des liens entre les membres en participant au traditionnel tournoi de golf de l'ASSQ.

Finalement, je suis bien contente d'annoncer le retour de la chronique SAS. Jean Hardy qui s'en est fort bien occupé pendant très longtemps manquait de temps pour continuer sa participation. Caty Blanchette, une de mes collègues de travail, a donc proposé de prendre la relève. Je la remercie, d'autant plus qu'elle a également joint l'équipe de rédaction pour nous donner un coup de main. Si son exemple vous inspire, ne vous gênez pas pour en faire autant, toute nouvelle chronique est la bienvenue.

J'en profite par la même occasion pour remercier les autres membres de l'équipe de rédaction. Isabelle Marchand s'occupe de la mise en page tandis que Steve Méthot et Sylvie Gauthier sont nos réviseurs officiels. Ce sont des collaborateurs de longue date : Steve et Sylvie font partie de l'équipe depuis 1997 et 1998, respectivement, et Isabelle s'est jointe au groupe en 2000. Ils ont pallié mon manque d'expérience et m'ont remise dans la bonne voie quelques fois lors du dernier numéro !

Profitez-bien de ce numéro et, toujours, dans l'espoir de vous lire dans le prochain ! ■

Myrto Mondor  
Rédactrice en chef, stat. ASSQ  
myrto.mondor@uresp.ulaval.ca

## Du plaisir garanti....en bonne compagnie Claude Ouimet

Quoi de mieux que de prendre une marche de santé sur des allées bien découpées aux odeurs de pelouse fraîchement coupée et de fouler, pourquoi pas, des cratères de sable doré sous un ciel ensoleillé. Et tant qu'à y être, se laisser aller à frapper une ridicule balle blanche pour l'amener dans un petit trou inaccessible. Venez découvrir cette intrigante activité découverte par les Écossais il y a de cela de nombreuses années.

L'invitation est donc lancée de participer à une activité que plusieurs hésitent à tort à découvrir...eh oui...le golf ! En lisant ceci, les date et lieu de cette activité vous auront été dévoilés via le groupe de nouvelles. Allez, laissez-vous aller à l'essayer. Au plaisir de vous y retrouver. ■

# À propos de l'ASSQ

Cette rubrique est préparée par les membres du Conseil d'administration et a pour but de fournir de l'information continue aux membres de l'ASSQ.

## Les échos du CA Claude Ouimet, Secrétaire

Deux réunions du CA ont eu lieu dans les derniers mois, soit en mars et en juin, la dernière s'étant tenue la journée même de l'Assemblée générale annuelle.

Le sujet qui a le plus retenu l'attention était bien sûr l'organisation du 1<sup>er</sup> Colloque ASSQ. Les dates et lieu de la tenue de l'événement, la recherche de conférenciers, la publicité de l'événement, le matériel promotionnel, la procédure d'inscription sont les principaux points qui ont fait l'objet des préoccupations les plus constantes. Lors de la dernière réunion, un examen de l'annulation de l'événement a été fait. Les principales remarques à cet effet se retrouvent dans l'article des échos de l'assemblée annuelle (voir chronique suivante). En plus, disons que l'inscription de moins d'une vingtaine de participants a justifié à elle seule l'annulation du colloque par respect pour les conférenciers qui n'étaient pas très chauds à l'idée de présenter leur atelier devant seulement quelques personnes.

À la première des deux réunions, le renouvellement du membership a fait l'objet de plusieurs discussions. Doit-on changer quelque chose à la procédure actuelle? Est-ce qu'un marketing agressif, telle la gratuité pour les nouveaux membres, qu'ils soient étudiants ou non, serait une bonne façon d'augmenter sensiblement le nombre de membres, lequel stagne depuis plusieurs années? Un regard sur l'ajout d'institutions comme membres institutionnels est également porté vers l'Université de Montréal, l'Université de Sherbrooke et l'UQAM.

Afin de faciliter la logistique entourant l'organisation de cours et/ou de conférences, l'ASSQ est fière de s'adjoindre la collaboration de trois membres, soit Nathalie Hamel pour Ottawa, Diane Leroux pour Québec et Christine Gamelin pour Montréal.

Quant au document de présentation que certains membres de l'ASSQ ont peaufiné à l'intention des cégeps, la version actuelle peut servir de canevas à quiconque veut préparer sa propre présentation sur le métier de statisticien. Cette version est ou sera sous peu disponible sur le site de l'Association.

Enfin, une discussion sur l'accessibilité des diverses sections du site internet de l'ASSQ par un membre ou un non-membre fait l'objet d'une réflexion au sein du CA. ■

## Échos de l'assemblée annuelle Claude Ouimet, Secrétaire

Viva le quantum...viva le quorum. Suite à l'assemblée de l'an dernier, qui s'était avérée non officielle par manque de participants, saluons la participation relativement massive de 22 membres cette année, le 19 juin dernier, dans les locaux de l'Université Laval à Québec. Présidée de main de maître par Pierre Lavallée, l'assemblée s'est déroulée dans une atmosphère très conviviale, le vin et le léger buffet aidant particulièrement.

Le tout a débuté officiellement à 17h45 par une présentation du rapport d'activités. La présence conjointe sur la tribune du président sortant Pierre Lavallée ainsi que du président actuel Sylvain Végiard a donné lieu à une certaine interaction, laquelle a permis d'alléger quelque peu ce genre de discours qui, avouons-le, n'est pas des plus faciles à rendre.

Isabelle Auger, entre deux boires et/ou changements de couches, nous a présenté brièvement le bilan financier de l'année 2002. Chapeau à Isabelle qui a su présenter un exercice comptable assez complet sans que l'Association ait recours à un vérificateur externe, ce que l'ensemble des membres présents à l'Assemblée a approuvé à l'unanimité.

Après l'approbation du maintien de la grille tarifaire actuelle pour une autre année, une discussion entre les membres s'ensuit. Le point le plus largement discuté a été l'achoppement du premier Colloque annuel ASSQ. Ce qui en ressort principalement, c'est la difficulté d'attirer une masse critique suffisante pour des ateliers ou des conférences tenus en parallèle. Quelques membres ont souligné que le moment choisi pour l'événement n'était pas tout à fait adéquat et que l'annonce un peu tardive du colloque, alors que la planification des congrès dans leur milieu de travail était déjà faite, avait causé un problème de planification budgétaire. La tenue de cet événement sur une seule

journée, avec une première conférence en matinée et une seconde en après-midi, entrecoupée de l'Assemblée annuelle sur l'heure du midi pourrait connaître plus de succès selon l'avis de plusieurs.

Enfin, le faible achalandage des étudiants dans la concentration statistique à l'Université Laval et la fermeture des programmes de statistique aux Collège Mérici et au Collège de Rosemont laissent perplexes les membres présents. Il faudrait peut-être que l'ASSQ soit plus présente dans les cours d'orientation de choix de carrière au secondaire ainsi que dans les cégeps.

La clôture de l'Assemblée générale annuelle 2003 a été prononcée vers 19h30.

Encore une fois, un merci à tous les membres présents. ■

## **L'accréditation professionnelle pour les statisticiens**

**Mike Sirois,**  
**directeur des communications**

### **Réunion spéciale**

Une réunion ayant pour thème principal l'accréditation professionnelle pour les statisticiens s'est déroulée à Ottawa le 16 mai dernier. Monsieur Neville Davies de la Royal Statistical Society était invité pour l'occasion. Des membres de divers organismes étaient présents. Cet article rapporte les grandes lignes de la réunion. Veuillez consulter les sites Web des organismes en question pour obtenir plus de détails sur l'accréditation professionnelle, la formation et le perfectionnement et tout autre sujet qui a été abordé au cours de cette réunion spéciale :

Royal Statistical Society :

<http://www.rss.org.uk/>

Société statistique du Canada :

<http://www.ssc.ca>

Société statistique d'Ottawa :

<http://macnash.admin.uottawa.ca/~sso/>

Association des statisticiennes et statisticiens du Québec :

<http://www.association-assq.qc.ca>

Statistique Canada :

<http://www.statcan.ca>

### **Accréditation professionnelle**

Certaines professions exigent une accréditation de ses travailleurs. On n'a qu'à penser aux actuaire qui passent une suite d'examens pour prouver leur niveau de compétence ou encore aux infirmières québécoises qui doivent réussir un examen du gouvernement pour pouvoir exercer leur profession. Les diplômes de formation obtenus sont souvent utilisés comme un des critères menant à l'accréditation. Le concept d'accréditation est vaste et peut prendre différentes formes. En statistique, il n'y a rien de formel en ce sens pour l'instant. Toute personne avec le moindre de formation ou d'expérience en statistique peut clamer avoir les connaissances et les compétences

pour faire une analyse de données, un sondage, etc. Cette réunion a offert une occasion d'échanger sur les processus d'accréditation en place ou en voie de l'être pour la profession de statisticien ainsi que les moyens d'améliorer les connaissances et les compétences des statisticiens.

### **Royal Statistical Society (RSS)**

Monsieur Neville Davies a entamé la réunion avec une présentation portant sur l'accréditation des statisticiens qui recherchent un statut professionnel reconnu internationalement. Le processus d'accréditation de la RSS attestant des qualifications d'une personne est fort complexe. Il tient compte de l'expérience (temps plein, temps partiel), de la formation reçue (durée, institution d'enseignement, etc.) et de plusieurs autres facteurs. Il a aussi parlé de continuité pour maintenir une accréditation acquise (formation et perfectionnement, mentorat, etc.). De plus, il a parlé de correspondance entre les diplômes obtenus à l'étranger et ceux obtenus ici ainsi que leur évaluation pour des fins d'accréditation. On peut imaginer qu'un processus d'accréditation universelle exige de faire des correspondances entre les différents systèmes d'éducation et organismes de travail.

### **Société statistique du Canada (SSC)**

Plusieurs membres de la SSC ont participé à cette rencontre, dont l'organisateur M. Don Royce. Les activités de la SSC sont nombreuses et plusieurs comités spécifiques ont été créés pour mieux les coordonner. On retrouve donc, entre autres, un comité sur le perfectionnement professionnel, un comité d'éducation en statistique, une section sur les méthodes d'enquêtes et un comité pour la réalisation de l'accréditation professionnelle. Des représentants siégeant à différents comités ont fait part des réalisations de la SSC. Les comités doivent collaborer ensemble assez régulièrement puisque les activités sont souvent interreliées. Le sujet d'intérêt, l'accréditation des statisticiens professionnels, a soulevé de nombreuses interrogations sur la manière dont elle devrait être faite. Le Comité de la SSC sur l'accréditation est à mettre au point actuellement un processus d'accréditation des statisticiens qui pourrait bientôt voir le jour.

### **Société statistique d'Ottawa (SSO)**

Monsieur Bill Ross a fait part des différentes activités organisées par la SSO. D'autres membres de la SSO étaient aussi présents pour cet échange. La SSO est une association régionale de la SSC.

### **Association des statisticiennes et statisticiens du Québec (ASSQ)**

Monsieur Pierre Lavallée, président sortant de l'ASSQ, a présenté le cas de l'ASSQ. On peut dire que l'ASSQ a été innovatrice en ce qui a trait à l'accréditation. En effet, le système de classement des nouveaux membres (étudiant, membre affilié, membre statisticien, membre institutionnel) tel que décrit sur le formulaire d'inscription de l'ASSQ représente une certaine forme d'accréditation en soi. La compétence des membres est établie en prenant en

**Suite à la p.9**

# Histoire du mot imputation

Éric Rancourt, Statistique Canada

De nos jours, le mot imputation est monnaie courante dans le jargon des enquêtes et recensements, mais d'où vient-il au juste?

Pour connaître la réponse, il faut d'abord se tourner vers la langue latine. Imputation et imputer viennent du verbe *imputare* qui possède plusieurs significations. La première est celle de porter au compte. Cette signification a été utilisée la première par Columelle, agronome sous les empereurs romains Tibère et Claude. C'est ce sens qui sera plus tard grandement répandu dans la théorie économique du XIX<sup>e</sup> siècle. Ensuite, le mot a été utilisé par Tacite et Pline le jeune pour signifier attribuer quelque chose à quelqu'un. Enfin, Martial, auteur espagnol d'épigrammes sous Titus et Domitien, se sert du mot imputation pour signifier donner ou assigner.

À la fin du moyen-âge, plusieurs auteurs utilisent déjà le mot imputation à divers usages. Entre autres, on note Rutebeuf au XIII<sup>e</sup> siècle, de même que Calvin et, plus tard, Shakespeare qui fait dire à Falstaff dans *Henri IV* à propos de Shallow que ses serviteurs lui ressemblent tellement qu'il oserait leur imputer ses qualités.

Un autre sens a été utilisé en théologie pour signifier l'attribution des mérites aux hommes. Par exemple, l'Épître aux Romains dans la Bible utilise six fois le mot imputation dans ce sens (en traduction car le terme original était *λογίζομαι*).

Par la suite, plusieurs des utilisations mentionnées ci-dessus sont faites du mot imputation jusqu'au XIX<sup>e</sup> siècle, où semble se dessiner une tendance à l'utiliser surtout en économie, finance et comptabilité pour le fait d'affecter une somme à un compte. Ce n'est qu'au milieu du XX<sup>e</sup> siècle que le mot imputation prend son sens actuel dans le domaine des sondages, après que le concept d'imputation ait été signifié par des termes tels que assignations, corrections ou estimations.

La première utilisation du mot imputation dans son sens actuel est vraisemblablement due à Hansen, Hurwitz et Madow (1953) qui utilisent ce terme la première fois (page 546) pour rappeler les corrections faites dans l'enquête américaine *Survey of retail shares* de 1948. Ensuite, il y a peu d'utilisations du mot imputation dans les deux décennies suivantes, mais à partir

des années 70, il y a prolifération de son utilisation à la suite de l'automatisation et de la formalisation des processus de vérification.

Enfin, les années 90 ont vraiment été la période où l'imputation a été propulsée à l'avant-plan de la statistique.

À partir de ce moment, le nombre d'articles sur l'imputation de même que les méthodes et approches se sont multipliées. On donne maintenant toutes sortes de qualificatifs à l'imputation : manuelle, automatique, simple, multiple, massive ou en chaîne.

Lorsqu'on parle d'imputation, on fait donc utilisation d'un terme riche en étymologie qui a été judicieusement choisi et qui représente bien le concept d'attribution d'une valeur à une unité.

### Références :

- Dictionnaire illustré Latin-Français, Félix Gaffiot, Hachette, 1934.
- Hansen, M.H., Hurwitz, W.N. et Madow, W.G. (1953). *Sample Survey Methods and Theory*, Vol. I et II. New York : John Wiley and Sons.
- The complete works of William Shakespeare, Vol. 1. Nelson Doubleday Inc., 1853.
- The Oxford English Dictionary. (1989). Second Edition. Clarendon Press, Oxford.
- Trésor de la langue française. Dictionnaire de la langue du XIX<sup>e</sup> et du XX<sup>e</sup> siècle. (1971). Éditions du Centre National de la recherche scientifique, Paris.

Note : Cet article a paru dans Le Bulletin d'imputation, Vol. 1, No. 1, printemps 2001, publié par la Direction de méthodologie de Statistique Canada. Le Bulletin d'imputation paraît deux fois par année, au printemps et à l'automne. Si vous êtes intéressé à recevoir le Bulletin d'imputation, vous pouvez faire ajouter votre nom à la liste de distribution en contactant Éric Rancourt (eric.rancourt@statcan.ca). ■

## Chronique livre

# « Exercices corrigés de méthodes de sondage » de Ardilly et Tillé

Pierre Lavallée, Statistique Canada

En matière de sondage, on trouve des ouvrages de cours sans trop de difficulté, mais les livres d'exercices corrigés sont très rares. À ma connaissance, le seul autre ouvrage de ce genre est le livre de Dussaix et Grosbras, *Exercices de sondages*. Notons que la plupart des livres de cours possèdent une section qui propose des exercices à la fin de chaque chapitre. Si des exercices sont déjà présents dans ces ouvrages, alors pourquoi écrire le livre *Exercices corrigés de méthodes de sondage* et en quoi ce livre peut-il être utile ? La réponse à ces questions est multiple.

Premièrement, ce livre tire son utilité de l'importance du sujet qu'il traite soit les méthodes de sondage. Quoiqu'elle possède des fondements mathématiques, la méthodologie des sondages est avant tout une discipline axée sur la pratique. En effet, la raison ultime du sondage est la collecte et le traitement de l'information afin d'augmenter notre savoir sur un ou des sujets donnés. La base mathématique derrière les méthodes de sondage n'est présente que pour calculer les différentes statistiques requises et pour optimiser les processus d'obtention de l'information. Pour un étudiant, l'aspect pratique de la matière fait souvent défaut puisque ce dernier ne possède généralement pas encore d'expérience dans les sondages.

Le livre de Ardilly et Tillé comble en partie le manque d'expérience pratique des étudiants. En effet, les différents exercices touchent beaucoup de situations rencontrées dans la « vraie vie ». On retrouve, bien entendu, le cas des sondages électoraux (exercice 2.9). Il y a aussi le cas délicat du traitement de la non-réponse quand celle-ci est reliée au calcul du poids total des passagers d'un avion (exercice 9.1). D'autres exercices sont plus farfelus, comme le redressement de l'estimation de la taille moyenne du pied droit à partir de la taille moyenne de l'individu (exercice 6.5). Un autre exercice digne de mention est la stratification des 100 éléphants d'un éleveur (exercice 4.3) où l'on peut y voir un clin d'œil au célèbre exemple de Basu sur le sondage à probabilités inégales et l'admissibilité des estimateurs. Quelle que soit la situation présentée dans les exercices, l'étudiant peut y trouver un intérêt parce qu'il peut enfin voir « à quoi ça sert... ».

se trouve du côté des professeurs. Ces derniers peuvent voir dans le livre de Ardilly et Tillé une source de problèmes et d'exemples à donner aux étudiants dans des travaux pratiques. Cependant, puisque le livre présente les solutions à la suite de l'énoncé des exercices, il va de soit que les exercices ne peuvent être directement utilisés par les professeurs à titre d'examen si les étudiants disposent du livre. Les professeurs peuvent tout de même s'inspirer de ces exercices pour en développer d'autres pour les étudiants. D'ailleurs, toutes les solutions sont largement commentées. Ardilly et Tillé mettent en évidence les éventuels pièges que recèlent les problèmes et tentent de porter, lorsque c'est nécessaire, un regard critique sur les méthodes, avec des éléments de comparaison pour que le lecteur se fasse une bonne idée de ce qui est souhaitable et de ce qui ne l'est pas. Chaque chapitre est précédé d'un bref rappel de cours qui reprend les points « à savoir » et qui fournit les notations importantes.

Pour les professeurs qui enseignent à un auditoire d'adultes, ce livre peut s'avérer l'outil rêvé. Les études démontrent que, contrairement aux enfants et aux adolescents, les adultes ont besoin de se rattacher à des exemples pratiques pour bien apprendre. Le livre de Ardilly et Tillé répond très bien à cette façon de présenter les méthodes de sondage. Il utilise en fait deux types d'approche : on y trouve alternativement des exercices assez théoriques et des exercices tout à fait pratiques laissant une large part aux applications numériques. Parfois, il s'agit d'applications assez directes de formules de cours – afin que ces dernières soient bien comprises – et parfois, ce sont des développements plus subtils, qui nécessitent une compréhension plus profonde des mécanismes reliés à la théorie des sondages.

Une troisième raison pour trouver ce livre particulièrement utile est la diversité des sujets couverts. Bien entendu, on ne peut ici parler d'exhaustivité, étant donné la vaste étendue de sujets auxquels s'applique la méthodologie des sondages. Cependant, le livre de Ardilly et Tillé touche à des sujets rarement ou encore jamais vus dans les livres de cours conventionnels. Notons

**Suite à la p.13**

La deuxième raison pour laquelle ce livre s'avère utile



# La chronique SAS

**Caty Blanchette, Unité de recherche en santé des populations**  
**cblanche@uresp.ulaval.ca**

Il me fait plaisir de prendre le relais de Jean Hardy et Jacques Pagé pour poursuivre la chronique SAS. Par le passé, j'ai toujours lu avec assiduité cette chronique qui m'a permis d'apprendre une multitude de petits trucs très pertinents. C'est maintenant à mon tour de vous communiquer mes « petits secrets » que j'ai accumulés au fil des années en tant qu'utilisatrice de SAS.

Pour cette première chronique, je vais vous dresser un portrait de quelques sections intéressantes à consulter sur le site Web de SAS afin de tirer profit de toute la documentation disponible.

Comme vous le savez probablement, l'adresse principale du site est <http://www.sas.com/> et on y trouve les détails sur la compagnie. En tant qu'utilisateur, il est plus intéressant de visiter le Centre de support aux clients dont l'adresse est <http://support.sas.com/>. Les sections suivantes sont alors disponibles :

- Technical Support
- Documentation
- Code Samples
- Communities
- Software Downloads
- Training

Il est possible d'accéder à de nombreux exemples de programmes à l'adresse [http://support.sas.com/techsup/sample/sample\\_library.html](http://support.sas.com/techsup/sample/sample_library.html). En particulier, en cliquant sur le lien [Data Analysis Samples \(SAS/STAT\)](#), il est possible de télécharger des programmes d'analyse facilement (et gratuitement...). L'an dernier par exemple, je cherchais une façon de calculer l'aire sous une courbe. En cherchant par mots-clés, j'ai trouvé un programme dans la banque ci-dessus nommé area.sas (<http://ftp.sas.com/techsup/download/stat/area.sas>) qui utilise PROC EXPAND. J'ai adapté ce programme selon mes besoins et mon problème fut résolu en quelques minutes!

Il est aussi possible de consulter les FAQ (questions fréquentes). Par exemple, vous trouverez les FAQ de la procédure GENMOD à l'adresse suivante : [http://support.sas.com/techsup/faq/stat\\_proc/genproc.html](http://support.sas.com/techsup/faq/stat_proc/genproc.html)

Pour ceux qui ne l'ont pas déjà d'installé sur leur poste, la documentation SAS Online pour la version 8 (et même pour la version 9 qui sera sur le marché en 2004) est accessible à l'adresse suivante : <http://support.sas.com/documentation/onlinedoc/index.html>, il suffit de s'enregistrer gratuitement pour obtenir un code d'utilisateur et un mot de passe. Cette documentation remplace tous les livres de référence désuets et s'avère être un outil de consultation très performant. L'essayer, c'est l'adopter!

Finalement, pour tout savoir sur les nombreuses publications de SAS, il faut se rendre à l'adresse <http://support.sas.com/publishing/index.html>. On a alors accès en ligne à tout l'inventaire des livres et il est possible de faire des recherches par mots-clés facilement. En plus d'avoir une description détaillée, on peut visualiser la table des matières et un des chapitres pour certains de ces livres. De plus, il peut être très intéressant de consulter toute la programmation des exemples ainsi que les bases de données utilisées, cela est possible pour les livres de la liste suivante : <http://support.sas.com/documentation/onlinedoc/code.samples.html>.

La catégorie des livres « Books by Users » est particulièrement intéressante car ces ouvrages traitent d'applications concrètes. J'aime bien consulter « A Step-by-Step Approach to Using SAS for Factor Analysis and Structural Equation Modeling » et « SAS System for Mixed Models », par exemple. Si vous cherchez un petit livre d'introduction à SAS (pour vous, un collègue moins expérimenté ou un étudiant), je vous recommande le livre suivant : « The Little SAS Book: A Primer, Revised Second Edition ». Vous pouvez le découvrir à l'adresse <http://www.sas.com/apps/pubscat/bookdetails.jsp?catid=1&pc=58788>. C'est le livre avec lequel j'aurais aimé apprendre SAS il y a 10 ans, ça m'aurait peut-être semblé moins compliqué... Il y a même un chapitre consacré au « débogage » de nos programmes qui explique comment interpréter les notes du style « Lost Card » qui apparaissent parfois dans la fenêtre LOG.

Je vous souhaite donc de faire de belles découvertes, ce ne sont pas les opportunités qui manquent. ■



## Suite de la p.5

considération certains critères, dont la formation. Il a aussi été question de la vidéothèque de l'ASSQ, du groupe de nouvelles, des présentations que l'Association organise sur le thème de la statistique à Ottawa, Montréal et Québec, de même que de la possibilité de la tenue d'un colloque de l'Association. Ces derniers aspects concernent l'échange entre statisticiens, la formation et le perfectionnement nécessaires pour faire progresser la statistique et pour assurer l'excellence de ces professionnels de la statistique.

### Statistique Canada

Pour conclure la réunion, M. Hew Gough de Statistique Canada a fait une présentation sur la formation et le perfectionnement des statisticiens. Il a présenté un éventail des différents cours et ateliers offerts à Statistique Canada, leurs publics cibles et leurs objectifs respectifs. Il a aussi donné un aperçu de certains cours offerts à Statistique Canada et à l'extérieur (ailleurs au Canada et à l'étranger) ainsi que du fardeau de préparation et de mise à jour de

ces cours.

### Conclusion

Le débat sur les processus d'accréditation pour les statisticiens, que cette dernière soit obligatoire ou non, normalisée à l'intérieur d'une association, d'une province, d'un pays ou dans le monde, ne fait que commencer. Nous sommes loin d'un système aussi sophistiqué que celui des avocats ou des ingénieurs. Le titre de statisticien n'est pas encore explicitement défini et le jour où cette profession sera officiellement reconnue et encadrée est encore loin. Nous sommes tous concernés par ce mouvement et il est important d'y participer. D'autres réunions de ce genre auront lieu très prochainement et vous en serez informés.

Pour obtenir plus d'information au sujet de la réunion qui s'est déroulée le 16 mai dernier et les présentations qui y ont été faites ou pour me faire part de vos commentaires concernant les différents processus d'accréditation ou de vos idées pour les prochaines discussions, n'hésitez pas à communiquer avec moi : [Mike.Sirois@statcan.ca](mailto:Mike.Sirois@statcan.ca) ■

## Statistiques et vie

### Drôles ?

Un sondage tout ce qu'il y a de sérieux nous en apprend de belles... À la question « Êtes-vous drôles ? », 71 % des personnes interrogées ont dit oui, pour 29 % qui ont répondu non.

Vous ne vous doutiez pas que vous étiez entourés de tant de gens drôles... ou du moins qui se croient drôles. Parmi les 71 % de gens drôles qui vous côtoient, il doit y avoir la concierge, la contractuelle, votre chef de bureau ou même le patron et vous ne le saviez pas ! Mais maintenant qu'on vous l'a dit, vous verrez certainement le monde environnant d'un autre œil. Puisque les trois quarts des gens sont drôles ! ■


[Tiré de l'*Almanach Vermot* 2003.]



# Un tableau magnifique et un pauvre graphique

Pierre Lavallée, Statistique Canada

Dans son édition du 6 mai 2003, le journal Le Droit présentait deux articles à la page 23 portant sur l'économie : un premier article sur les compressions faites par Air Transat et un autre article sur la hausse du dollar canadien par rapport au dollar américain.



*Transat AT (dont le siège social se situe à Montréal) a mis à pied lundi 18 % de sa main-d'oeuvre.*

**Mises à pied par secteur:**

- Air Transat: 400
- Vacances Air Transat et Nolitour: 100

**Mises à pied par région:**

- Montréal: 50 %
- Toronto: 40 %
- Vancouver: 10 %

**Main-d'oeuvre restante par secteur:**

- Air Transat: 1700
- Voyageurs: 600

**Résultats financiers:**

- Exercice 2002: Bénéfice: 7,8 M\$
- 1<sup>er</sup> trimestre 2003: Perte: (7 M\$)

**Cours de l'action**  
(à la Bourse de Toronto clôture de lundi): 3,5 \$

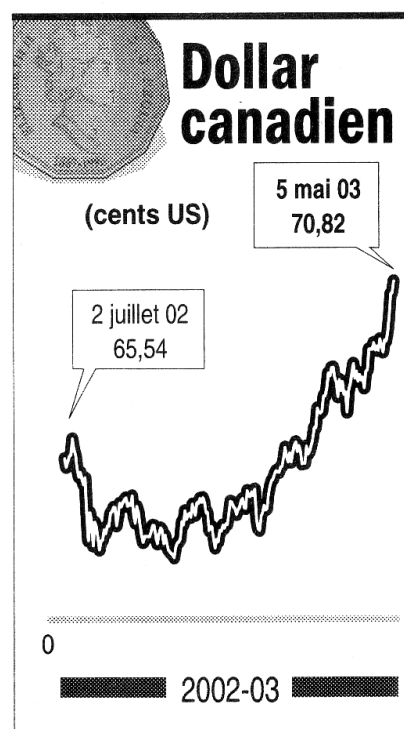
Source: Transat A.T. inc.

PC

L'article sur Air Transat était accompagné d'un tableau récapitulatif résumant les compressions d'Air Transat. Toute l'information a été résumée dans un simple tableau qui parle de lui-même. En regardant le tableau reproduit ici, on n'a même plus besoin de lire l'article. Après la mention que les mises à pied représentent 18% de la main-d'oeuvre d'Air Transat (ce qui met en contexte l'importance du problème), on y retrouve les mises à pied par secteur, les pourcentages par région, la main-d'oeuvre restante, les résultats financiers avec les bénéfices de 2002 et les pertes du 1<sup>er</sup> trimestre de 2003 (ce qui explique les compressions) et le cours de l'action (pour les investisseurs soucieux). Toute cette information

a été présentée de façon claire et concise, ce qui mérite certainement des fleurs.

L'article sur le dollar canadien comprenait le graphique reproduit ici. Ce graphique prétend montrer les fluctuations et la hausse du dollar canadien, ce qui est effectivement le cas. Cependant, en voulant réduire ce graphique à sa plus simple expression, on a alors perdu de l'information essentielle. Par exemple, le graphique ne possède pas d'axe vertical. Si on se fie à la barre horizontale où la valeur zéro apparaît sans explication, on peut alors penser que le point de départ à 65,54 cents US n'est certainement pas à l'échelle. En fait, on a même l'impression à la fin du graphique que la valeur du dollar canadien aurait, à toute fin pratique, doublée (trop beau pour être vrai...). Notez que l'axe horizontal est aussi pratiquement inutile puisqu'il ne donne pas d'information exacte sur les dates du début et de la fin de la période représentée, les graphistes ayant opté d'inscrire ces dates dans les bulles à l'intérieur du graphique. Ce graphique ne mérite finalement rien de plus que le pot. ■



— PC

## Nettoyer une liste de noms : l'algorithme du « Soundex »

Thierry Petitjean-Roget, IRSST

### Problématique

Le site Internet de l'Institut de recherche Robert-Sauvé en santé et en sécurité du travail ([www.irsst.qc.ca](http://www.irsst.qc.ca)) fait l'objet d'une refonte majeure. Un des buts recherchés est de valoriser le chercheur en associant ses publications à sa page du bottin, ou de façon générique, d'associer un individu à l'ensemble de ses publications à partir d'un nom d'auteur.

Dans un premier temps, la liste des noms d'auteurs a été triée par ordre alphabétique pour n'en retenir que les noms distincts. Mais, dans une liste relativement importante, il est difficile de repérer les fautes de frappe, telles que « Bélair », « Belair » ou « Beliar »; ces noms ne sont pas nécessairement consécutifs dans une liste ordonnée. Dans un deuxième temps, afin de repérer ces cas, l'algorithme du Soundex, qui traduit un mot en code « phonétique », a été utilisé.

### Méthode

L'algorithme du Soundex a été développé en 1918 par Margaret K. Odell et Robert C. Russel du bureau américain des archives. Les informaticiens en connaissent surtout la description faite par Donald Knuth (1973). Relativement simple, l'algorithme se déroule en 8 étapes :

1. Supprimer les « espaces initiaux »;
2. Mettre le mot en majuscule, en remplaçant les lettres accentuées par leur équivalent non accentué;
3. Garder la première lettre;
4. Supprimer les lettres A,E,I,O,U,Y,H et W;
5. Remplacer les lettres restantes par le chiffre associé dans la table ci-contre, basée sur un regroupement de la prononciation des lettres;
6. Supprimer les paires consécutives de chiffres répétés;
7. Si le code obtenu contient moins de quatre éléments, compléter à droite par des zéros;
8. Si le code obtenu contient plus de quatre éléments, n'en conserver que les quatre premiers.

Code	Phonèmes	Anglais	Français
1	Bilabiales	B,F,P,V	B,P
2	Labiodentales	C,G,J,K,Q,S, X,Z	C,K,Q
3	Dentales	D,T	D,T
4	Alvéolaires	L	L
5	Vélaires	M,N	M,N
6	Laryngales	R	R
7			G,J
8			X,Z,S
9			F,V

À titre d'exemple :

1. « Lavallée Pierre » → « LavalléePierre »
2. « LavalléePierre » → « LAVALLEEPIERRE »
3. « LAVALLEEPIERRE » → « L »
4. « L » → « LVLLPRR »
5. « LVLLPRR » → « L944166 »
6. « L944166 » → « L9416 »
7. « L9416 » → « L9416 »
8. « L9416 » → « L941 »

On obtient un code équivalent avec « Laval Paul » ou bien « Léveillée Patrice » !

Le Soundex est un algorithme d'association de mots par leur consonance, et non de tri traditionnel. On l'utilise beaucoup pour des recherches généalogiques, là où les noms sont souvent d'origine phonétique, mais aussi dans des moteurs de recherche. Utilisé dans de nombreuses bases de données, cet algorithme est basé sur la prononciation anglaise; la table de concordance a donc été « révisée » pour la langue française. Il a fallu aussi convertir le « ç » en s, et le « œ » en « e ».

L'hôpital général de Toronto l'a utilisé il y a quelques années lors de l'informatisation de ses dossiers, parmi d'autres outils informatiques; près de 20% de ces dossiers étaient mal classés, dont une bonne partie due à des erreurs phonétiques. Dans le cadre de la restructuration du site Web de l'IRSST, c'est ce type d'erreur que l'on veut corriger. Dans notre base de données d'auteurs, on retrouve sous le même code « B460 » : Bellair, Bélair, Belair, Boilar et Beaulier; après vérification, il s'agit de la même personne dans les trois premiers cas.

### **Autres méthodes**

L'algorithme Soundex a subi plusieurs améliorations au fil du temps. Soundex 2 repose sur l'algorithme de Gus Baird (Georgia Tech) tel que décrit dans le livre « SQL avancé » de Joe Celko paru en 1995 chez Thomson International Publishing. Il utilise l'ensemble des lettres sous une forme non accentuée et remplace des groupes de lettres et préfixes par leur consonance. Une adaptation française a été faite par Frédéric Brouard.

Dans sa page Web, Monsieur Brouard propose un troisième algorithme, le « Phonex<sup>®</sup> ». L'algorithme Phonex est optimisé pour le langage français et sait reconnaître différents types de sons comme les sons 'on', 'ai', 'ein', etc. On obtient un résultat sous la forme d'un nombre réel en double précision. À titre d'exemple, Brouard écrit : « ... Soundex2 récupère moins d'occurrences que Soundex. Mais dans le cas de FAURE, la différence est très importante. Soundex récupère FERREY, FERREY, FREY et FUERI, tandis qu'il oublie FORT ! En revanche Soundex2 se montre plus tolérant et récupère FORT et PHAURE. Quand à Phonex, il récupère très peu de noms : FARRE, FAURE, FORT, FOURR, PHAURE, VARD et VAURE. »

### **Conclusion**

Ayant testé la version anglaise originale du Soundex, la version française et le Soundex 2, nous avons choisi la version française du Soundex pour répondre à nos besoins de repérage des noms à consonance similaire. La version anglaise originale ne permet pas de différencier des noms tels que « Archambeault » et « Arsenault ». Le Soundex 2, dont le coût d'exploitation en temps de calcul est assez élevé, semble peu approprié dans le cas de petites bases de données d'auteurs, dont les patronymes sont d'origine linguistique diversifiée. De plus, en associant le prénom au code Soundex du nom, le nombre de corrections à effectuer dans notre liste de noms a été réduit.

J'ai sans doute utilisé la machinerie lourde là où ce n'était pas nécessaire, mais cela m'a permis de partager avec vous mes redécouvertes. ■

### Références :

Donald Knuth, "The Art Of Computer Programming, vol. 3: Sorting And Searching", Addison-Wesley (1973), pp. 391-392.

Toute cette information se retrouve sur le site Web de Frédéric Brouard : Le SQL de A à Z - L'art des SOUNDEX [http://sqlpro.developpez.com/Soundex/SQL\\_AZ\\_soundex.html](http://sqlpro.developpez.com/Soundex/SQL_AZ_soundex.html)

On peut aussi consulter :

Année Universitaire 2000/2001 Examen Partiel : Valeur d'Accueil et de Reconversion en Informatique <http://lwh.free.fr/pages/vari/VARI2001A.pdf>

Soundex Calculators <http://www.myatt.demon.co.uk/soundex.htm>

## Suite de la p. 7

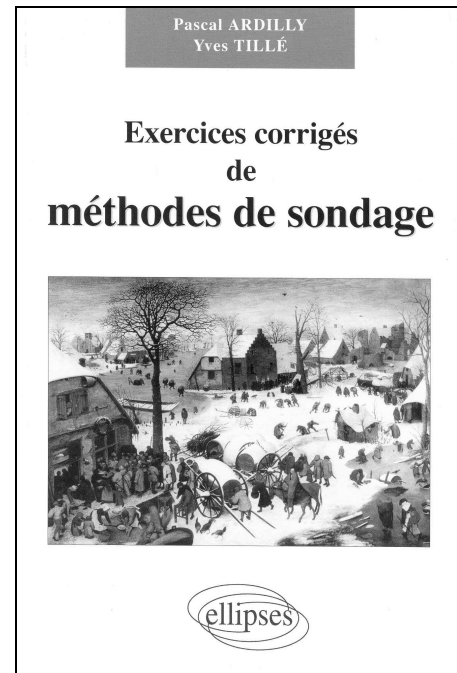
que bien des sujets, comme la capture-recapture (exercice 2.15) et la méthode des quotas (exercice 3.27), sont directement tirés de la pratique, tandis que d'autres sujets, comme la Rao-Blackwellisation (exercice 3.22), relèvent d'un cadre beaucoup plus théorique.

Bien qu'il touche à plusieurs sujets, l'ensemble du livre est structuré de manière assez classique. On présente en premier lieu les méthodes d'échantillonnage (aléatoire simple, stratifié, à probabilités inégales, à plusieurs degrés, équilibré, utilisant la méthode des quotas, ...) qui constituent la première étape naturelle du processus de sondage. On passe ensuite aux redressements suivis d'un chapitre sur l'estimation de la variance et d'un chapitre consacré à la correction de la non-réponse. Tout au long du livre de Ardilly et Tillé, la « vedette » est la variance d'échantillonnage, véritable fil rouge de l'ouvrage, dont on trouve trace dans la grande majorité des exercices, au même titre que les considérations sur le biais des estimateurs.

Après le premier chapitre d'introduction, le deuxième chapitre traite du sondage aléatoire simple. Dans ce chapitre, parmi les sujets rarement vus dans les ouvrages de cours, mentionnons d'abord l'enquête répétée (exercice 2.8) qui constitue la base de la pratique des sondages dans les instituts nationaux de statistique. En effet, la plupart des enquêtes que l'on retrouve dans ces instituts sont répétées annuellement, trimestriellement ou mensuellement de manière à ce qu'on puisse suivre des évolutions au cours du temps. La mesure de la tendance revêt ainsi une importance plus grande que la mesure du total ou de la moyenne ponctuelle. Le cas de l'enquête répétée est aussi étudié au sixième chapitre (Redressement sur une variable) en rapport avec la régression (exercice 6.11).

L'estimation dans un domaine (exercice 2.12) et sa variance (exercice 2.13) constitue un autre sujet pratiquement ignoré des ouvrages de cours conventionnels. Cependant, cette méthode est primordiale et presque toujours utilisée en pratique pour la production des estimations. Son utilisation vient du fait que les bases de sondage disponibles pour la sélection des échantillons sont rarement ou sinon jamais à jour. Ainsi, il peut y avoir une différence entre l'information fournie par la base de sondage et celle collectée sur le terrain. Par exemple, une population a été stratifiée par région géographique, mais certaines unités ont déménagé entre la création de la base de sondage et l'enquête elle-même. Ainsi, les estimations pour les régions actuelles constituent des domaines d'intérêt qui diffèrent, bien entendu, des strates de tirage.

La capture-recapture (exercice 2.15) et la recapture avec remise (exercice 2.17) constituent des méthodes utilisées en biologie (pour l'estimation de populations de poissons, par exemple) ou en démographie (pour évaluer la sous-couverture ou la sur-couverture des recensements). Ce sont des méthodes de sondage très utiles en pratique, mais que l'on ne retrouve généralement pas dans les livres de sondage. À ma connaissance, le seul livre de sondage qui présente la méthode de capture-recapture est le livre *Sampling* de Thompson.



Parmi les sujets rarement vus dans les livres, le troisième chapitre (Sondage à probabilités inégales) présente le plan équilibré (exercice 3.20). Cette méthode consiste à tirer un échantillon de sorte que l'estimation pour une variable donnée corresponde à la valeur connue de la population. L'utilisation du plan équilibré est en pleine expansion. Longtemps dénigrée par manque d'algorithme satisfaisant pour assurer une certaine randomisation des échantillons, cette méthode a récemment été utilisée pour le découpage du territoire français dans le cadre du recensement continu. Précisons que le recensement continu vise à couvrir annuellement un cinquième de la France de sorte que le territoire complet soit recensé au bout de cinq ans. Le livre de Ardilly et Tillé présente aussi un exercice sur des équilibrages successifs (exercice 3.28).

La méthode des quotas (exercice 3.27) est certainement la méthode de sondage la plus utilisée par les instituts de sondage privés, en particulier pour les sondages d'opinion. Cette méthode consiste à tirer des individus au hasard jusqu'à ce que des quotas,

par groupe d'âge-sexe par exemple, soient satisfaits. Les quotas sont la plupart du temps proportionnels au nombre de personnes dans la population entière. La méthode des quotas s'avère très utile quand on ne dispose pas de liste ou de base de sondage adéquate pour cerner la population, ce qui est souvent le cas pour les instituts de sondage privés. Notons que l'absence de base de sondage est aussi étudiée à l'exercice 3.29.

Au quatrième chapitre (Stratification), l'exercice sur l'optimalité nationale et régionale (exercice 4.16) relève d'une méthode de répartition de l'échantillon rarement vue dans les ouvrages de cours. Cette méthode se nomme répartition avec puissance (ou *power allocation*). La répartition avec puissance permet, avec l'utilisation d'un paramètre entre 0 et 1, d'obtenir une répartition de l'échantillon qui soit un compromis entre la répartition de Neyman – qui minimise la variance de l'estimation nationale au détriment de celle des estimations régionales – et une répartition visant à produire des estimations régionales avec des précisions relatives égales, au détriment de l'estimation nationale. Cette méthode de répartition est donc très utile pour les instituts nationaux de statistique qui doivent produire des estimations à la fois pour le pays entier et ses différentes régions ou provinces.

Les cinquième et sixième chapitres (Sondage à plusieurs degrés et Redressements sur une variable) présentent, somme toute, des sujets que l'on retrouve généralement dans les livres de sondage conventionnels.

Le septième chapitre est entièrement consacré à un sujet rarement traité dans les livres de cours : les redressements sur plusieurs variables ou, autrement dit, le calage sur marges. Notons que le calage sur marges n'est en fait maîtrisé que depuis une dizaine d'années. Comme son nom le suggère, cette technique vise à ajuster les poids d'estimation de sorte que les estimations produites correspondent à des totaux connus. L'application la plus souvent rencontrée de cette technique dans les instituts nationaux de statistique est certainement le redressement des enquêtes sur la population active (ou enquêtes force de travail). Pour effectuer ces enquêtes, on a le plus souvent recours à des bases de sondage aréolaires. Comme on sélectionne des aires géographiques, on ne contrôle pas l'échantillon au niveau de l'âge et du sexe des personnes enquêtées. Les estimations des effectifs de personnes par groupe d'âge-sexe peuvent donc différer des valeurs réelles de la population. Par ailleurs, grâce aux recensements de la population, ces effectifs sont connus et permettent donc d'effectuer un calage sur marges afin que les estimations d'effectifs par groupe d'âge-sexe correspondent à ces totaux.

Le huitième chapitre (Estimation de variance) traite en son ensemble d'un sujet rarement abordé : l'estimation de variance en utilisant la linéarisation. Notons que la linéarisation permet d'obtenir des formules de variance pour des estimateurs relativement complexes comme la moyenne géométrique (exercice 8.3), le coefficient de détermination  $R^2$  (exercice 8.7) que l'on rencontre en régression linéaire, ainsi que le coefficient d'asymétrie (exercice 8.8).

Le neuvième et dernier chapitre du livre aborde le traitement de la non-réponse. De par son aspect inévitable, la non-réponse en sondage est fréquemment rencontrée dans les livres de cours.

Le niveau technique initial nécessaire pour profiter de ce livre n'est pas très élevé. De manière peut-être paradoxale, la plupart des composantes des plans de sondage ne font pas appel à des mathématiques très élaborées. Quant à l'aspect probabiliste et statistique, il s'accommode d'un niveau de premier cycle. Pour la plupart des exercices, une connaissance de base en calcul de probabilité et en statistique descriptive suffit.

Le livre *Exercices corrigés de méthodes de sondage* de Ardilly et Tillé est un outil extrêmement utile aux étudiants, professeurs, praticiens et théoriciens désireux d'en savoir davantage sur la méthodologie des sondages par l'entremise d'exercices sur des sujets très divers et bien choisis. Ce livre constitue aussi un excellent complément aux livres respectifs des deux auteurs : *Les techniques de sondages* de Ardilly et *Théorie des sondages* de Tillé. ■

#### Références :

Ardilly, P., Tillé, Y. (2003) *Exercices corrigés de méthodes de sondage*. Ellipses.

Dussaix, A.-M., Grosbras, J.-M. (1992) *Exercices de sondages*. Economica.

#### Exemple de D. Basu :

"An essay on the logical foundations of survey sampling, Part One," du livre *Foundations of Statistical Inference*, (1971) édité par Godambe, V.P. et Sprott, D.A. (pp. 203-242), Toronto.

(pour un résumé, voir [www-unix.oit.umass.edu/~cluster/ed/outline/c00ed72.PDF](http://www-unix.oit.umass.edu/~cluster/ed/outline/c00ed72.PDF))

Thompson, Steven K. (2002) *Sampling*. Wiley, 2<sup>e</sup> édition.

Ardilly, P. (1994) *Les techniques de sondages*. Technip.

Tillé, Y. (2001) *Théorie des sondages. Échantillonnage et estimation en populations finies*. Dunod.

# Conférences à venir

Si vous avez des conférences à annoncer, n'hésitez pas à nous contacter!

Quoi?	Où?	Organisateur / site Web	Quand?
Congrès de l'ACFAS 2004	Montréal	ACFAS / <a href="http://www.acfas.ca">www.acfas.ca</a>	2004
Congrès annuel de la SSC 2004	Montréal	SSC / <a href="http://www.ssc.ca">www.ssc.ca</a>	30 mai au 2 juin 2004
Joint Statistical Meetings 2004	Toronto	ASA / <a href="http://www.amstat.org">www.amstat.org</a>	8 au 12 août 2004
Colloque francophone sur les sondages 2005	Québec	SFdS / <a href="http://www.sfds.asso.fr">www.sfds.asso.fr</a>	26 au 27 mai 2005
Congrès de la SFdS 2005	Sherbrooke	SFdS / <a href="http://www.sfds.asso.fr">www.sfds.asso.fr</a>	30 mai au 3 juin 2005
Congrès de l'ACFAS 2005	Chicoutimi	ACFAS / <a href="http://www.acfas.ca">www.acfas.ca</a>	2005
Congrès annuel de la SSC 2005	Saskatoon	SSC / <a href="http://www.ssc.ca">www.ssc.ca</a>	12 au 15 juin 2005

L'horaire ci-dessus peut être sujet à changement. Consultez le site web de chaque association pour de plus amples informations.

ACFAS : Association francophone pour le savoir (anciennement connu sous le nom de Association Canadienne-Française pour l'Avancement des Sciences).

ASA : American Statistical Association

SSC : Société Statistique du Canada

SFdS : Société Française de Statistique

## De quoi se faire des cheveux

Le laboratoire de biotechnologie britannique Phytopharm ne fera pas fortune avec sa dernière invention : la lotion pour faire repousser les cheveux. Au cours d'un « banc d'essai » public, sa nouvelle crème miracle a montré que les hommes qui avaient, soi-disant, constaté un mieux dans la repousse ou un coup d'arrêt dans la chute de leurs cheveux, étaient ceux à qui on avait proposé... un placebo, c'est-à-dire une crème où il n'y avait aucun produit quel qu'il soit.

À la suite de ces résultats, l'action Phytopharm a chuté de 15 % à la bourse dans les heures qui ont suivi. Et là, ce sont les patrons de l'entreprise qui se sont fait tirer les cheveux... ■

[Tiré de l'*Almanach Vermot* 2003.]

## Statistiques et vie

### La tête de l'emploi

« Être chauve diminue considérablement les chances de trouver un emploi... »

En Allemagne, c'est à cette conclusion qu'est parvenue une étude menée par un cabinet d'experts. Six hommes ont présenté chacun deux C.V. L'un accompagné d'une photo où il apparaissait avec des cheveux, l'autre avec une photo retouchée où le candidat apparaissait chauve. 588 demandes d'emploi ont été soumises à 98 responsables de recrutement. Le résultat est probant : 41 % des candidats avec cheveux ont été retenus pour un entretien, contre seulement 32 % des mêmes candidats, mais qui étaient chauves.

Les chevelus sont considérés comme dynamiques (49 %) contre seulement 29 % des chauves. 36 % des chauves ont droit au qualificatif de « physique agréable », mais les chevelus sont 49 % à avoir ce « physique agréable ». Heureusement, les responsables des ressources humaines qui ont répondu à l'étude ont affirmé, ensuite, qu'ils ne prenaient pas en compte l'apparence physique des candidats. Néanmoins, quand on est chauve, on n'a pas forcément la tête de l'emploi.

Et une situation peut parfois ne tenir qu'à un cheveu... ■

[Tiré de l'*Almanach Vermot* 2003.]



# Suivre son cours ...

*La vie suit son cours, mais qu'en est-il du statisticien qui sommeille en vous?*

## COURS

Cette grille présente les cours offerts par différents organismes. La liste des cours universitaires est disponible sur demande à l'ASSQ. Notez que l'information fournie dans la grille des cours est sujette à changement. Le lecteur est invité à entrer en communication avec le ou la responsable des cours pour corroborer et compléter l'information présentée (frais d'inscription, dates des cours). Le lecteur pourra par le fait même faire ses propres démarches pour suivre le ou les cours de son choix.

Organisme, Lieu	Clientèle (basée sur la matière du cours)	Types de cours	Contacts
<b>Creascience, Montréal</b>	Débutant, intermédiaire, chercheurs, techniciens et autres professionnels de la R&D	Planification d'expériences, Analyse en composantes principales, Contrôle statistique de la qualité.	Natalie Rodrigue Montréal (514) 840-9220 poste 227 creascience.com info@creascience.com
<b>École de Technologie Supérieure, Montréal</b>	Débutant	Planification d'expérience dans le domaine industriel, Contrôle statistique de la qualité.	Suzanne LeBel (514) 396-8830 (Service de perfectionnement) www.perf.etsmtl.ca perf@etsmtl.ca
<b>Institut SAS</b>	Grand public, entreprises, débutant ou expert	Data mining, langage macro, programmation, rédaction de rapport, statistique.	Institut SAS (514) 395-8922 poste 4071 www.sas.com/formation sastrain@can.sas.com
<b>Services conseils Hardy, Québec, Montréal</b>	Débutant, intermédiaire et avancé	Différentes sessions de formation portant sur les logiciels SAS, SPSS, MINITAB, Statistica, Crystal Reports.	Monique Trempe (418) 626-1666 http://www.schardy.qc.ca schardy@schardy.qc.ca
<b>Statistique Canada, Ottawa</b>	Intermédiaire et avancé	Différents cours portant sur les méthodes d'analyse de données (séries chronologiques, contrôle statistique de la qualité, analyse de données de survie, ...), méthodologie d'enquête.	Hew Gough (613) 951-3067 Céline Charette (613) 951-1044 www.statcan.ca infostats@statcan.ca  <u>Note</u> : Certains cours de Statistique Canada peuvent se donner à l'extérieur.

## SÉMINAIRES

Des séminaires ont lieu de façon régulière aux endroits suivants. N'hésitez pas à contacter le ou la responsable pour plus de détails.

<b>UQAM</b> Pascale Rousseau Tél. (514) 987-3000, #3224 Fax (514) 987-8935 rousseau.pascale@uqam.ca	<b>Université de Montréal</b> Martin Bilodeau Tél. (514) 343-2410 Fax (514) 343-5700 bilodeau@dms.umontreal.ca	<b>Université de Sherbrooke</b> Bernard Colin Tél. (819) 821-8000, #2012 Fax (819) 821-8200 bernard.colin@dmi.usherb.ca	<b>Université Laval</b> Christian Genest Tél. (418) 656-5280 Fax (418) 656-2817 genest@mat.ulaval.ca
---	--	---	--