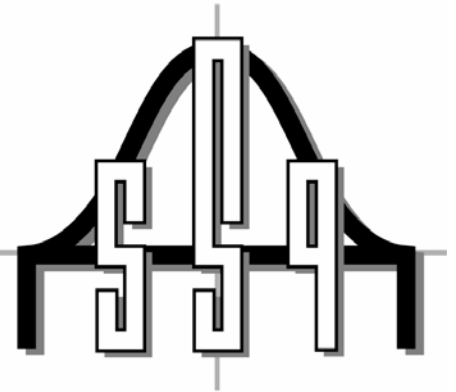


Convergence

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



Volume XII • Numéro 1

Avril 2007

Chers membres,

Nombreux sont ceux d'entre vous qui vous en êtes voulu d'avoir loupé le Colloque 10^e anniversaire de l'ASSQ tenu à Saint-Antoine-de-Tilly en juin dernier. L'après-midi avait été consacré à trois conférences sur des thèmes d'actualité, livrées avec verve et brio par des spécialistes soucieux de captiver leur auditoire. Cette prestation avait été suivie de l'assemblée générale annuelle et s'était ensuite prolongée par un coquetel et un repas cinq services agrémenté de vins capiteux. Le tout s'était déroulé dans un cadre bucolique et une atmosphère conviviale.

Comme vous avez déjà pu le lire dans le *Convergence* de décembre, vous aurez la chance de vous reprendre le 8 juin, au Manoir du Lac Delage. Blotti au pied des Laurentides, à une trentaine de kilomètres au nord de Québec, ce lieu de villégiature réputé conjugue nature et plaisirs de la table. La formule de l'an dernier ayant plu, elle sera reprise intégralement, quoiqu'à un tarif légèrement majoré : 50 \$ pour les membres en règle et 25 \$ pour les membres étudiants. Un hébergement à prix modique a également été prévu pour les participants provenant de l'extérieur de la région.

Le programme scientifique qui a été concocté pour l'occasion est à la fois séduisant et varié. Notre conférencier de marque sera Keith Worsley, professeur à l'Université McGill et récipiendaire de la médaille d'or de la Société statistique du Canada en 2004. Il nous entretiendra de ses importants travaux concernant le traitement statistique des images de tomodensitométrie, et en français s'il vous plaît !

Dans un deuxième temps, Chantal Girard nous présentera les conclusions d'une vaste étude sur les perspectives démographiques du Québec et de ses régions réalisée en collaboration avec deux de ses collègues de l'Institut de la statistique du Québec, Esther Letourneau et Normand Thibault.

Enfin, sur une note plus ludique, notre populaire chroniqueur Martin Rioux nous fera partager sa passion com-

municative pour tout ce qui touche les applications de la statistique dans le « merveilleux monde du sport. »

L'assemblée générale annuelle du 8 juin prochain sera aussi l'occasion de faire le bilan d'une pleine mesure d'activités.

Plusieurs « 5 à 7 » ont été organisés depuis septembre (deux à Québec, un à Montréal et un à Trois-Rivières). Grâce aux efforts concertés des deux Jean-François, le magazine et la vitrine électronique de l'ASSQ projettent en outre une image positive de l'association sur les scènes nationale et internationale. Que tous les artisans de ces succès trouvent ici l'expression renouvelée de notre gratitude.

Et puisque j'en suis à l'article des remerciements, permettez-moi de souligner combien Mireille Guay et moi sommes sensibles au témoignage de confiance que vous nous avez rendu en renouvelant notre mandat, ne fut-ce qu'implicitement, lors des dernières élections. Nous saurons nous en montrer dignes.

Au plaisir de vous retrouver tous lors du colloque du 8 juin !

Le président de l'ASSQ, [Christian Genest](#), Ph. D.

Dans ce numéro :

Le mot du rédacteur en chef	3
Les échos du C.A.	4
Un mot sur la campagne de renouvellement 2007	5
Des statistiques au menu sur le site Web de l'ASSQ	6
La 20 ^e édition de la Journée de la statistique	8
Les « Jeudis de l'ASSQ »	9
Le panel : avantages et limites d'application	11
Compte rendu de lecture	13
Le paradoxe de Simpson	15
Chronique historique	17
Statistiquement vrai ?	19
Chronique SAS	20
Conférences à venir	22
Nouvelles publications	23
Suivre son cours	24

CONVERGENCE

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

Rédacteur en chef

Jean-François Quessy, Université du Québec à Trois-Rivières (Jean-Francois.Quessy@uqtr.ca)

Rédacteurs adjoints

Steve Méthot, Agriculture et agroalimentaire Canada
Myrto Mondor, C. H. affilié universitaire de Québec
Nathalie Gaudreault, SOM inc.

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 envoyés, sous forme de fichiers Microsoft Word, à l'adresse électronique du rédacteur en chef (voir ci-haut). Les dates de tombée sont les 1^{er} mars, 1^{er} juillet et 27 octobre pour les numéros d'avril, d'août et de décembre, respectivement. 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 par courriel au rédacteur en chef leur document électronique prêt pour l'impression avant la date de tombée du prochain numéro. Les membres institutionnels ont une gratuité de publicité allant jusqu'à une page par numéro. Les tarifs pour la parution dans un numéro de *Convergence* sont les suivants (note : l'annonce de cours et de séminaires en dernière page est gratuite) :

Tarifs

Carte d'affaires	15 \$
1/4 page	40 \$
1/2 page	80 \$
Page entière	150 \$

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 81094, succursale Loretteville
Québec (Québec) Canada G2B 3W6
Courriel : association_assq@yahoo.ca
Page Internet : <http://www.association-assq.qc.ca>
Dépôt légal - Bibliothèque nationale du Québec, 2007

MEMBRES INSTITUTIONNELS :



Département de mathématiques et de statistique



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 types d'adhésion 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) 10 \$ (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 : Christian Genest (*Université Laval*)

Vice-président : Joseph Nader (*FERIC*)

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

Trésorière : Marie-Ève Tremblay (*Institut de la statistique du Québec*)

Registraire : Michel Fluet (*SOM Inc.*)

Directrice des communications : Mireille Guay (*Santé Canada*)

Mot du rédacteur en chef

Chers membres,

Une fois de plus, les collaborateurs de Convergence vous ont préparé une édition printanière des plus intéressantes. Outre les chroniques régulières auxquelles vous êtes habitués depuis quelque temps déjà, vous constaterez que quelques membres de l'ASSQ se sont commis dans des articles qui viennent pimenter la parution actuelle. Est-il besoin de rappeler que ces initiatives sont hautement appréciées par celui, en l'occurrence moi, qui ne sert que de « relais » ? N'hésitez pas à vous laisser inspirer par ces initiatives !

Une de ces collaborations spéciales provient de l'ancien webmestre de la vitrine Internet de l'ASSQ, Mike Sirois. Celui-ci nous présente quelques statistiques de fréquentation du site web de l'association et explique comment accéder à une multitude d'informations relatives aux visites du site. Je soupçonne que le (la) statisticien(ne) en vous sera très tenté(e) par l'exploration de toutes ces fonctionnalités dès la fin de la lecture de l'article !

Toujours dans la catégorie des collaborations inhabituelles, Brahim Baccouche nous entretient des pour et des contre de l'utilisation de panels en recherche marketing. Une attention particulière est portée à l'évolution de l'appréciation de produits dans le temps. Quant à elle, Marie-Hélène Toupin, une étudiante de l'Université du Québec à Trois-Rivières, nous livre ses commentaires et

ses impressions suite à la lecture d'un livre sur les erreurs courantes en statistique. Il est toujours bon de se faire rappeler les pièges potentiels lors de la mise en œuvre de méthodes statistiques. Dans la même lignée, je vous présente une traduction d'un texte concernant le paradoxe de Simpson.

En plus d'avoir donné son nom à une distribution statistique bien connue, Vilfredo Federico Samaso Pareto est surtout un scientifique au parcours inhabituel. Dans sa chronique historique, Pierre Lavallée nous décrit les grandes lignes de la vie de celui qui a découvert la « courbe de la loi pour la richesse et le revenu ». Continuant dans ce filon historique, vous avez sûrement entendu parler de ce documentaire de James Cameron dans lequel seraient établies des preuves irréfutables de la découverte du tombeau de Jésus. À la lecture de l'article « Statistiquement vrai ? Jésus aurait été marié et père », porté à l'attention de Convergence par Pierre Lavallée, vous constaterez le rôle central de la statistique pour arriver à cette étonnante conclusion.

Enfin, la chronique SAS de Sylvain Tremblay porte sur quelques nouvelles fonctionnalités de la populaire procédure PROC LOGISTIC. Comme à l'habitude, des exemples clairs viennent illustrer ses descriptions.

Bonne lecture !

[Jean-François Quessy](#), Ph. D.

Statistique et vie :

Statistique gérontologie

Lorsqu'on n'a plus vingt ans et que l'on se rend au supermarché faire ses courses, on s'attend à quoi ? Aux étiquettes illisibles, car écrites trop petites, aux produits trop hauts ou trop bas dans les rayons et aux bousculades. Chez *Adeg Aktiv Markt 50+*, tout cela est révolu. Ce magasin, situé dans la banlieue de Salzbourg et à Vienne, a tout prévu pour les seniors. Les prix sont écrits en grand, la lumière n'est plus violente et des loupes sont à la disposition des clients aux rayons de crèmerie et de produits de beauté. Plus besoin, donc, de se casser les reins ou de se hisser sur la pointe des pieds, car les produits sont à la portée de la main et le sol est antidérapant. Des coins repos sont mêmes aménagés et les chariots à roulettes sont prévus pour faire sièges !

Les responsables de ces magasins prévoient un succès de la formule, car un quart de la population autrichienne et un tiers de la population allemande devraient avoir plus de 60 ans en 2015.

[Tiré de l'*Almanach Vermot* 2006.]

À propos de l'ASSQ

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

Les échos du C.A.

Claude Ouimet, secrétaire

Pas (ou peu) de nouvelles... bonnes nouvelles ! Non pas que nous n'ayons rien à vous annoncer dans cette chronique, mais l'air d'aller que s'est donné le Conseil d'administration (CA) depuis quelque temps fait en sorte que les réunions sont un peu plus espacées. La chimie qui est maintenant installée entre ses membres permet au CA d'être très efficace. Il ne faudrait pas pour autant que vous croyiez que les personnes actuellement en poste seront là éternellement... Et surtout, que leur connivence ne vous empêche pas de songer à devenir vous-même membre du CA, de peur d'avoir de la difficulté à vous y intégrer ! N'empêche que la machine semble mieux huilée que jamais et que depuis quelque temps, les choses tombent en place rondement au plan logistique.

- L'ASSQ sera partenaire du congrès que la Société statistique du Canada et la Société française de statistique tiendront conjointement à Ottawa en mai 2008. L'ASSQ parrainera l'événement mais n'y contribuera pas financièrement. Pour cette raison, aucun rabais d'inscription ne sera consenti à ses membres. Une salle sera néanmoins offerte gratuitement à l'ASSQ pour qu'elle puisse tenir son assemblée générale annuelle et la rencontre sera annoncée dans le programme du congrès. En échange, *Convergence* fera la promotion de cette conférence d'envergure internationale auprès de ses lecteurs.
- L'organisation du Colloque ASSQ prévu le 8 juin prochain va bon train. L'affiche annonçant l'événement circule depuis peu. Nous espérons que le « Mot du président » de Christian vous aura mis l'eau à la bouche et que vous serez des nôtres. Les dernières réunions ont permis de figurer les détails concernant le déroulement de la journée et les frais de participation au colloque.
- Petite pause remerciements au travers des annonces... ça fait toujours plaisir d'en recevoir quand on s'engage à titre de bénévole dans une association :
- à Jean-François Quessy et son collègue Mhamed Mesfioui pour l'organisation du « Jeudi de l'ASSQ » tenu à Trois-Rivières, le 22 février dernier. Cette initiative a connu un franc succès, compte tenu du bassin réduit de statisticiens en Mauricie.
- à Thierry Duchesne et à Aurélie Labbe, qui ont pris en charge l'accueil et la logistique des salles dans le cadre des deux « Jeudi de l'ASSQ » tenus à l'Université Laval cette année.
- La campagne de renouvellement des adhésions a été un succès, quoique plusieurs anciens membres ne semblent plus intéressés à cotiser. Il faut mentionner que le rabais dont les adhérents à l'ASSQ bénéficiaient lors du Colloque francophone sur les sondages, tenu à Québec en 2005, a contribué à hausser artificiellement le nombre de cotisants cette année-là.
- L'ASSQ a accepté avec plaisir de parrainer le colloque « Une mesure pour l'avenir » qui aura lieu à Québec le 18 octobre prochain. Cette activité, organisée par le Réseau d'échange sur la mesure de la satisfaction de la clientèle du Centre d'expertise des grands organismes, traitera de l'utilité des sondages dans la fonction publique québécoise. De plus amples détails sur l'événement vous parviendront sous peu.
- Vous aurez peut-être remarqué que quelques annonces concernant la Société statistique de Montréal ont récemment été diffusées sur le site web de l'ASSQ. Cette initiative s'inscrit dans le cadre d'un programme informel de coopération entre les deux associations. Nous espérons que vous tirerez profit de ces informations.
- Si vous avez des commentaires à formuler concernant les sujets traités ou des suggestions de thèmes à aborder, n'hésitez pas ! Votre CA est à l'écoute de vos besoins ; écrivez-nous ou manifestez-vous via le groupe de courriel.

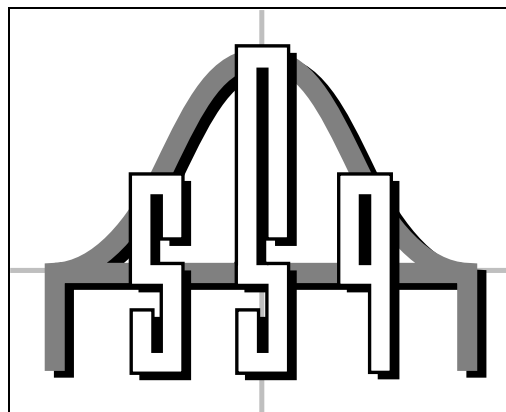
Un mot sur la campagne de renouvellement 2007

Lors de la campagne de renouvellement et d'abonnements pour 2006, l'ASSQ offrait la possibilité d'une adhésion pour deux ans au prix de 75 \$. Cent huit membres se sont alors prévalus de cette occasion. Ceci constituait une base solide de membres pour l'année 2007.

Pour la campagne de renouvellement 2007, quatre-vingt-douze membres réguliers et huit membres institutionnels ont été sollicités. En date du 1^{er} mars 2007, l'ASSQ comptait cent quarante-quatre membres réguliers et six membres institutionnels.

Enfin, au moment où vous lirez ces lignes, plusieurs auront reçu un dernier rappel par courrier pour les encourager à renouveler leur adhésion pour 2007. J'espère que la plupart en auront profité pour rejoindre les rangs de l'ASSQ.

[Michel Fluet](#), Registraire



Des statistiques au menu sur le site Web de l'ASSQ

Les statistiques sont le gagne-pain des membres de l'Association des statisticiennes et des statisticiens du Québec (ASSQ). À l'image de notre profession, le site Web de l'ASSQ regorge de données de toutes sortes. La section « Membres » fournit à chaque année le nombre de membres de l'association, telle une série chronologique. Les numéros archivés de *Convergence* contiennent quant à eux les tableaux de fréquences des rapports annuels de la trésorière. Pour ceux et celles dont l'appétit statistique n'en serait pas rassasié pour autant, il existe également une section consacrée exclusivement aux statistiques de fréquentation du site Web de l'ASSQ. Voici ce qui est au menu à cet endroit.

Des ustensiles pour le chef cuisinier

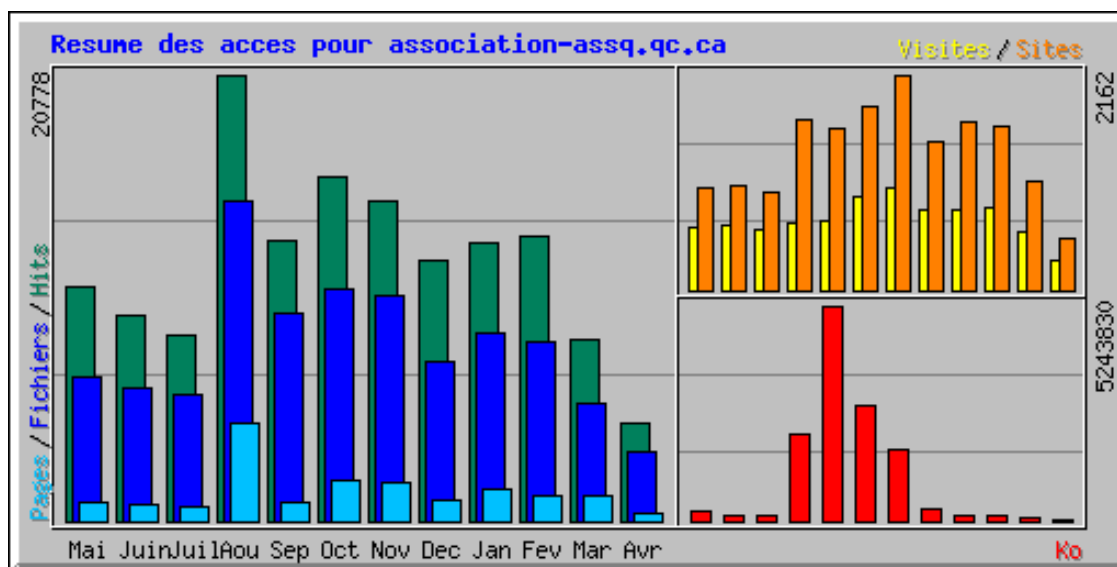
Pour concocter les statistiques du site, le gratuiciel *The Webalizer*, disponible à l'adresse <http://www.mrunix.net/webalizer/>, est utilisé. En plus d'être gratuit, cet utilitaire offre de nombreux avantages, notamment la possibilité d'être configuré pour obtenir des statistiques personnalisées. Fait à noter, il traite correctement la langue française, avec ses accents particuliers.

Un réfrigérateur bien garni

Pour ouvrir la porte menant aux statistiques de fréquentation, tel un cuisinier qui cherche ses ingrédients, vous devez vous rendre à l'adresse <http://www.association-assq.qc.ca/stats>. À noter qu'il n'y a pas de lien direct vers cette page à partir de l'accueil du site. De cet endroit, on accède à une quantité gargantuesque de statistiques de fréquentation du site. Un tableau sommaire pour les douze derniers mois est offert comme entrée.

La définition des éléments du tableau se retrouve à http://www.mrunix.net/webalizer/webalizer_help.html. La statistique la plus importante est probablement le nombre total de visites, c'est-à-dire le nombre d'accès au site par des personnes différentes.

À noter d'emblée que l'information pour le mois d'avril indique une basse fréquentation, puisque le mois n'était pas encore terminé au moment de la publication du texte.



La suite...

De façon générale, on peut dire que l'été est la période la plus creuse. L'automne et l'hiver présentent les fréquences les plus élevées, ce qui correspond à l'activité intense des séminaires, des conférences, des offres d'emploi ainsi que du tournoi de golf annuel de l'ASSQ ! C'est aussi la période de fréquentation scolaire des étudiants.

Fait non négligeable, de nombreux sites Web associés à des écoles et des établissements divers à travers le monde possèdent un lien vers le site de l'ASSQ. Donc, on peut penser que des étudiants de tous endroits viennent y grignoter régulièrement ! On possède d'ailleurs des statistiques sur la provenance des navigateurs.

Ceux qui sont membres de l'ASSQ depuis longtemps connaissent la popularité du site à l'extérieur du pays via l'ancien forum de discussion. Régulièrement, des individus de l'extérieur du Canada posaient des questions sur la théorie statistique et les possibilités d'emploi dans le domaine.

Pour le webmestre et le conseil d'administration, connaître quelles sont les pages les plus populaires s'avère important. Par exemple, si on s'apercevait qu'une page n'est jamais ou presque jamais visitée, le conseil pourrait décider de revoir son contenu et sa pertinence. Comme la page des nouvelles apparaît par défaut, c'est sans surprise qu'elle est la plus consultée. Toutefois, lorsqu'une nouvelle édition de *Convergence* est rendue disponible sur le site, le lien devient rapidement très fréquenté.

Un congélateur de réserves

Un autre avantage du gratuiciel qui produit les statistiques de fréquentation du site est qu'il permet l'archivage des données de manière automatique. Le webmestre peut alors consulter, via un accès informatique à distance, l'historique des données de fréquentation. Cela permet de tracer des bilans, de juger de la popularité du site selon les mois, les années et d'identifier les sections moins populaires.

La facture

Le logiciel est gratuit et le service rendu est fort utile. Ces informations recueillies sont en quelque sorte des commentaires d'appréciation des clients suite à un repas au restaurant. Ces traces, tout à fait confidentielles, permettent de réajuster le tir au besoin et de s'interroger à savoir s'il y a place à amélioration. Vous aussi, les membres, avez votre mot à dire. N'hésitez donc pas à contacter votre nouveau webmestre Jean-François Plante à jeff@stat.ubc.ca si vous avez des questions, des commentaires ou des suggestions concernant le site Web de l'association. À votre tour de découvrir les statistiques du site.

Bon appétit!

Mike Sirois, Statistique Canada

La 20^e édition de la Journée de la statistique



Le 2 mars dernier a eu lieu à Québec la 20^e édition annuelle de la Journée de la statistique. Cette rencontre vise à faire connaître la profession et à préparer les futurs diplômés en statistique au monde du travail ; elle est organisée par un groupe d'étudiants de premier cycle en statistique de l'Université Laval, le CASUL (Comité pour l'Avancement de la Statistique de l'Université Laval), placé cette année sous la présidence de Mme Chantal Caouette.

Réunis à l'Hôtel Classique, une quarantaine de participants ont eu l'occasion d'entendre trois conférenciers. Éric Shink a décrit son cheminement professionnel et présenté quelques applications de la statistique chez Anapharm. Après avoir dépeint la journée type d'une méthodologiste à l'emploi de Statistique Canada, Isabelle Michaud a brossé un tableau des opérations nécessaires au traitement des données de recensement. Enfin, François Bellavance, professeur à HEC Montréal, a fait état de divers travaux de recherche portant sur les risques d'accidents automobiles liés à l'emploi du téléphone portable au volant. Un « 5 à 7 » au restaurant « La pointe des Amériques » a clôturé la journée.

Mieux vaut se taire, parfois ...

Un politicien français, à n'en pas douter rempli de bonnes intentions, s'est un jour indigné du fait que « 50% des Français ont un quotient intellectuel (QI) inférieur à 100 ». Il s'est ensuite lancé dans une série de propositions visant à améliorer le système d'éducation français pour ultimement, souhaitait-il, améliorer le QI de ses concitoyens. Ce que notre valeureux politicien ignorait sans doute, c'est que la distribution des scores QI est artificiellement ramenée à une moyenne de 100 et un écart-type de 15 ! Grosso modo, le score arbitraire « 100 » est placé à la médiane des QI estimée sur un échantillon de la population française. Malheureusement pour notre pauvre homme, même advenant une augmentation significative du QI moyen des Français, la médiane sera toujours placée à 100 et l'affirmation « 50% des français ont un QI inférieur à 100 » sera toujours d'actualité !

[Adaptation libre d'un extrait du livre « Inventer l'homme », de Albert Jacquard]

Les jeudis de l'ASSQ...

Le 22 février...

Le 22 février dernier s'est tenu le premier « Jeudi de l'ASSQ » de l'histoire de l'UQTR ! L'auditoire composé d'une quinzaine de participants, en majorité des étudiants de premier cycle en mathématiques, a écouté attentivement le conférencier Ali Assani, professeur à l'UQTR. Géographe de formation, M. Assani a entretenu les participants du rôle central que joue la statistique dans la compréhension de nombreux phénomènes naturels. Plusieurs illustrations ont été fournies, dont la plus remarquable concernait la modélisation du nombre annuel d'ouragans en Amérique du Nord. Une discussion animée a suivi l'exposé et le tout a culminé en échanges cordiaux autour d'un bon vin et de quelques victuailles. De l'avis général, l'expérience méritera d'être répétée l'an prochain !

[Jean-François Quessy](#)

Le mystérieux 142857

À première vue, ce nombre n'a rien de particulier. Certes, mais multiplions-le et regardons ce qui se passe :

$$\begin{aligned}142857 \times 1 &= 142857 \\142857 \times 2 &= 285714 \\142857 \times 3 &= 428571 \\142857 \times 4 &= 571428 \\142857 \times 5 &= 714285 \\142857 \times 6 &= 857142\end{aligned}$$

Ce sont toujours les mêmes chiffres qui apparaissent, changeant simplement de place en avançant comme un ruban. Si on continue, voyez ce qui arrive :

$$142857 \times 7 = 999999 !$$

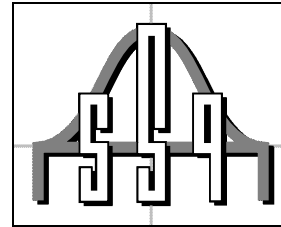
Or, en additionnant $142 + 857$, on obtient 999, et $14 + 28 + 57 = 99$!

Vous en voulez encore ? Le carré de 142857 donne 20408122449, et ce dernier nombre est formé de 20408 et 122449, dont la somme est ... 142857 !

[Tiré de l'*Almanach Vermot* 2006]

COLLOQUE ASSQ ET AGA

VENDREDI 8 JUIN 2007 – 13H



L'ASSQ vous convie à son colloque annuel, qui aura lieu le vendredi 8 juin 2007 au Manoir du Lac Delage (www.lacdelage.com), site de villégiature situé à quelque 30 kilomètres au nord de Québec. Cette activité sera suivie de l'assemblée générale annuelle des membres de l'association, d'un coquetel et d'un repas gastronomique.

Programme

13h	Ouverture du colloque
13h30	L'imagerie numérique et son analyse statistique à des fins cliniques : quelques applications Keith Worsley (Université McGill)
14h30	Si la tendance se maintient... Perspectives démographiques, Québec et régions, 2001-51 Chantal Girard (Institut de la statistique du Québec)
15h30	Pause santé
16h	L'analyse statistique appliquée à la LNH Martin Rioux (Promaintech Novaxa)
17h	AGA 2007
18h	Coquetel
19h	Souper

Lieu
Manoir du Lac Delage



40, avenue du Lac, Stoneham

www.lacdelage.com/hotel-quebec/nousContacter.cfm

Tarifs avant le 25 mai⁽¹⁾⁽²⁾

	Étudiant	Autre
Membre de l'ASSQ :	25 \$	50 \$
Non-membre de l'ASSQ ⁽³⁾	35 \$	100 \$

⁽¹⁾ Après le 25 mai, il faut ajouter 25 \$ au tarif

⁽²⁾ Taxes comprises

⁽³⁾ Incluant l'adhésion à l'ASSQ pour l'année 2007

Il vous est possible de réserver une chambre au tarif de 125 \$ + taxes en occupation double, petit déjeuner exclus. Un bloc de 10 chambres nous est réservé jusqu'au 9 mai. Lors de la réservation, mentionnez votre participation au colloque ASSQ du 8 juin pour bénéficier du tarif préférentiel. Les droits d'inscription au colloque précisés ci-dessus comprennent les conférences de l'après-midi, une pause santé, un coquetel et un souper gastronomique. Il est possible d'inviter les conjoints à se joindre au souper au prix coûtant de 60 \$.

Pour réservation : Marie-Ève Tremblay

Par courriel : marie-eve.tremblay@stat.gouv.qc.ca

Par téléphone : (418) 691-2410 poste 3050

Par télécopieur : (418) 643-4129

Pour confirmer votre réservation, vous devez faire parvenir un chèque libellé à l'ordre de l'ASSQ au soin de Mme Marie-Ève Tremblay à l'adresse suivante :

**Direction de la méthodologie, de la démographie
et des enquêtes spéciales**

Institut de la statistique du Québec

200 chemin Sainte-Foy, 3^e étage

Québec, Québec G1R 5T4

Réservation

Prénom et nom :

Employeur :

Téléphone : ()

Courriel électronique :

Membre étudiant : 25 \$

Non-membre étudiant : 35 \$

Membre – autre : 50 \$

Non-membre – autre : 100 \$

Conjoint : 60 \$

**Seuls les chèques seront acceptés et doivent nous parvenir
avant le 30 mai pour confirmer votre participation au tarif réduit.
Le nombre de places est limité à 80.**

Pour réservation : ASSOCIATION DES STATISTIENNES ET DES STATISTIENS DU QUÉBEC, a/s Marie-Ève Tremblay

Téléphone : (418) 691-2410 poste 3050 - **Télécopieur :** (418) 643-4129

Courriel : marie-eve.tremblay@stat.gouv.qc.ca

Le panel :

avantages et limites d'application

Brahim Baccouche, Impact Recherche - Groupe Cossette Communication

En recherche marketing, nous sommes souvent amenés à réaliser des enquêtes pré et post-campagne publicitaire, à mesurer l'évolution de comportements et de perceptions ou encore à analyser l'évolution de la fidélité à une marque dans le temps. À cet effet, une alternative se présente au chercheur : (1) l'*étude longitudinale* où il constitue un panel et maintient le même échantillon à travers le temps et (2) l'*étude transversale* où il utilise des échantillons indépendants tirés de la même population à chaque collecte de données. Le choix de l'une ou de l'autre de ces méthodes dépend principalement du budget alloué à l'enquête, de l'objectif de cette dernière, de la nature du phénomène étudié et de l'intervalle séparant les différentes collectes de données. Dans cet article, je cherche à présenter les deux méthodes de recherche tout en mettant en exergue les avantages et les inconvénients de chacune en vue de fournir des recommandations aux professionnels de la recherche quant aux conditions appropriées pour la considération de chacune des méthodes.

Dans le domaine de la recherche marketing, nous entendons très souvent parler de « panel ». Dans la majorité des cas, ce terme est utilisé pour désigner un ensemble d'individus sollicités pour répondre à des sondages, généralement via le Web, portant sur des sujets divers. Dans le présent article, l'attention porte sur le concept de panel en tant que méthode permettant de s'adresser « à un échantillon d'individus à travers le temps pour fournir de multiples observations de chaque individu dans l'échantillon »¹. Quoiqu'il soit difficile dans la pratique de maintenir un panel à l'identique dans le temps, puisque les chercheurs sont souvent amenés à remplacer certains répondants, le panel permet d'avoir

des observations plus riches de l'évolution du phénomène observé par rapport à une enquête qui étudierait l'évolution du phénomène entre deux ou plusieurs périodes sur des échantillons indépendants. Un exemple de cette dernière approche serait de demander à un échantillon de consommateurs représentatifs de la population d'exprimer leur fidélité à une marque donnée à l'instant t et d'interroger à l'instant $t+1$ un deuxième échantillon également représentatif à propos du même phénomène. Une telle méthode permet certes d'avoir la variation globale nette de la fidélité à la marque, mais ne permet en aucun cas de connaître la fréquence et la direction de cette variation. Pour reprendre l'exemple de l'étude de la fidélité à la marque, une étude longitudinale auprès d'un panel de consommateurs permettrait non seulement d'avoir l'évolution de la fidélité à la marque étudiée, mais d'identifier également les consommateurs qui changeraient de marque et de déceler ainsi les marques concurrentes les plus menaçantes.

L'avantage que présente le panel en termes de richesse et de précision de l'information collectée dépend de la corrélation entre les observations du phénomène étudié ainsi que de l'intervalle séparant les périodes de l'enquête. Dans le cas où la corrélation est négative et l'intervalle est court, l'utilisation de deux échantillons indépendants serait la méthode la plus appropriée. Dans ce cas, la variance du phénomène étudié d'une période à une autre est faible. À titre d'exemple, l'étude des habitudes d'achat de biens de consommation courante comme l'épicerie hebdomadaire se prêterait mieux à une étude auprès d'échantillons indépendants. A contrario, le choix du panel comme méthode de recherche serait plus judicieux pour l'étude des comportements d'achat de biens de consommation durable ou encore l'étude de l'évolution d'une valeur sociale comme l'entraide dans une société donnée. Cela m'amène à parler

¹ Cheng Hsiao, 1986, *Analysis of Panel*, Cambridge University Press.

d'un autre facteur déterminant dans le choix de l'une ou l'autre de ces méthodes : la nature du phénomène à étudier et sa stabilité dans le temps. Les changements au niveau des aspects socio-économiques sont plus fréquents que ceux relatifs aux phénomènes socio-psychologiques. Ainsi, la perception du don au sein d'une société devrait être plus stable que la perception d'une marque pour laquelle on accroît le budget publicitaire.

Outre l'avantage que présente le panel en termes de richesse et de précision de l'information, cette approche est également avantageuse en ce qui a trait au coût de recrutement des panélistes. Cependant, ces avantages s'accompagnent de certaines limites. La principale réside dans « la mortalité du panel ». Ce phénomène consiste en la perte de panélistes souvent due à une baisse de leur motivation à participer à l'étude ou à un changement de leurs coordonnées durant l'enquête.

On reproche également au panel le contexte plutôt artificiel dans lequel les individus interviewés sont placés. Les personnes interrogées risqueraient de manquer de spontanéité dans leurs réponses puisqu'elles sont conscientes du phénomène étudié. Ainsi, lorsque l'objet de l'étude porte sur un type de consommation courante pour lequel les individus ont développé un réflexe quasi mécanique, l'utilisation du panel avec des enquêtes répétées risque de pousser les personnes interviewées à mener des réflexions sur leur consommation et de créer par conséquent une situation artificielle porteuse de biais. A contrario, dans le cas de l'étude de phénomènes ayant fait l'objet de réflexions ou de concertations au préalable comme le tabagisme ou l'alcool au volant, la probabilité d'apparition d'une situation arti-

ficielle serait plus faible. Néanmoins, le biais généré par une éventuelle apparition d'une situation artificielle peut être détecté par la constitution de groupes-contrôles. Le groupe-contrôle réside dans la constitution simultanée d'un panel principal qui sera sollicité plusieurs fois et d'un groupe test qui servira à réaliser deux enquêtes séparées par le même intervalle de temps et réalisées à la même période que le panel principal. La comparaison des données recueillies permettra de déceler un éventuel biais dû à la création d'une situation artificielle et de prendre les mesures nécessaires pour corriger la situation.

En résumé, et comme le montre l'analyse des avantages et des inconvénients de chacune des méthodes, le choix du panel ou des échantillons indépendants pour analyser l'évolution d'un phénomène donné est tributaire de l'analyse préalable de la nature du phénomène étudié, de la fréquence de variation du phénomène et de la corrélation entre les observations du phénomène étudié d'une période à une autre. Cependant, dans la pratique, la contrainte budgétaire demeure un facteur prépondérant dans le choix de la méthode à considérer.

Bibliographie :

Cheng Hsiao, 1986, *Analysis of Panel*, Cambridge University Press.

François Chazel, La méthode du panel et ses possibilités d'application à la structure politique française, 1966, *Revue française de sociologie*, VII.

M. Neveu, 2005, *Techniques d'enquête et méthodes de sondages*.

Erreurs courantes en statistique

Voici une collection de quelques commentaires suite à ma lecture récente du livre « Common Errors in Statistics », de Phillip I. Good et James W. Hardin.

Les statisticiens sont souvent appelés à donner leur avis d'experts dans des applications touchant divers domaines tels la médecine, les sciences sociales et la finance, ou lors d'enquêtes et sondages. Leurs opinions exprimées sont souvent critiquées lors de la prise de décisions importantes. Il y a donc peu de place pour l'erreur dans cette profession. Le livre *Common Errors in Statistics* de Phillip I. Good et James W. Hardin dresse une liste d'erreurs souvent commises lors de la pratique statistique. L'objectif des auteurs est, d'une certaine façon, d'éveiller les consciences des praticiens de la statistique pour leur éviter autant que possible d'éventuels faux pas.

L'ouvrage est composé de trois parties : 1) les erreurs courantes dans les applications de base ; 2) les tests d'hypothèses et l'estimation de paramètres ; et 3) la construction de modèles probabilistes. Le livre couvre plusieurs étapes d'une étude statistique, en passant par la formulation d'hypothèses, la collecte de données, la vulgarisation des résultats et leur interprétation. Le texte est généralement facile à comprendre pour quiconque possède des connaissances de base en statistique. Pour rendre la lecture encore plus accessible, les auteurs utilisent des résumés pour faire ressortir les items jugés les plus importants. Parmi les items sur lesquels les auteurs se sont penchés, on retrouve

1. Les hypothèses : le pourquoi de votre recherche.
2. La collecte de données.
3. L'estimation.
4. La vulgarisation des résultats.

1. Les hypothèses : le pourquoi de votre recherche

Les auteurs rappellent, à juste titre, qu'une hypothèse statistique doit être à la fois quantifiable et testable. Ainsi, l'hypothèse doit faire intervenir des quantités mesurables ou référer à des éléments qui peuvent être affectés à des catégories mutuellement exclusives. Dans ce dernier cas, on parle de variables qualitatives ou catégoriques.

Avant même de débiter une expérience, les auteurs suggèrent de clairement

- définir les objectifs et établir un plan de recherche ;
- formuler les hypothèses et toutes les alternatives associées et dresser la liste des résultats, conclusions et actions possibles.

2. La collecte des données

Lors de la prise de données, les auteurs rappellent qu'il faut

- considérer les valeurs exactes lorsque cela est possible, même si on prévoit grouper par catégories pour les analyses ;
- s'assurer, lors de sondages, que les questions ne révèlent pas le but de l'étude ;
- toujours commencer une étude avec plus d'observations que le nombre minimum requis afin d'obtenir la puissance désirée s'il y a des non-réponses ;
- s'assurer que l'échantillon représente bien la population, et non un sous-groupe non représentatif de celle-ci.

Selon eux, la collecte des observations devrait faire en sorte que les observations soient indépendantes et identiquement distribuées. Comme exception évidente à cette affirmation, cependant, on pense d'emblée à l'analyse de séries chronologiques.

Pour s'assurer de l'indépendance des observations lors d'une étude où les questionnaires sont envoyés par la poste, il ne faut surtout pas envoyer plus d'un formulaire par résidence pour éviter que les personnes interrogées ne communiquent entre elles au sujet de l'étude. Pour que les observations soient identiquement distribuées, il faut par ailleurs éviter de changer d'instrument de mesure au cours d'une expérience et ne pas modifier en cours d'étude des questions mal comprises dans un questionnaire.

3. L'estimation

Parmi les mises en garde des auteurs concernant les méthodes d'estimations de paramètres, mentionnons qu'un bon estimateur doit être

- impartial, convergent, efficace, robuste et engendrer un minimum de perte. Ces notions sont expliquées ci-dessous.
 - Impartialité : l'estimateur ne dépend pas de l'unité de mesure. Par exemple, on devrait pouvoir prendre un ensemble de mesures en mètres, estimer les paramètres en centimètres, reconvertir en mètres et obtenir la même valeur que si on avait estimé en mètres au départ.
 - Convergence : la valeur estimée doit s'approcher de la vraie valeur lorsque la taille de l'échantillon augmente.
 - Efficacité : l'estimateur le plus efficace est celui dont le nombre d'observations requises est le plus petit pour atteindre une précision visée.
 - Robustesse : se dit d'un estimateur qui est peu affecté par d'éventuelles valeurs aberrantes. Par exemple, lorsqu'on mesure une tendance cen-

trale, enlever les valeurs maximales et minimales de l'échantillon rend généralement l'estimateur plus robuste.

- Minimum de pertes : Si θ^* est un estimateur de θ , la fonction de perte $|\theta^* - \theta|$ ou $(\theta^* - \theta)^2$ doit être minimale.

- sans biais (au moins asymptotiquement) et posséder une variance minimale.

4. La vulgarisation des résultats

À l'étape de l'écriture d'un rapport statistique, quelques conseils sont donnés, dont

- fournir les détails sur la puissance et les calculs de tailles d'échantillons requises ;
- expliquer, si possible, la raison et / ou la provenance des données manquantes ;
- utiliser des mesures de dispersion (variance, étendue) pertinentes à l'étude ;
- préférer des intervalles de confiance plutôt que des seuils observés (*p-values*);
- identifier les sources potentielles de biais.



Parfois, on utilise la statistique comme un ivrogne utilise un poteau de lumière: comme un support et pas pour éclairer...

[Tiré du site <http://alea.ine.pt/frances/html/trivial/html/humor4.html>]

Le paradoxe de Simpson : quand les grands jeux de données mentent

Traduction libre d'un article tiré du site Internet « *Amazing applications of probability and statistics* », par Jean-François Quesy

Une règle générale en statistique consiste à considérer que plus un jeu de données est volumineux, plus les conclusions qu'on peut en tirer sont fiables. Le paradoxe de Simpson, cependant, remet sérieusement en question cette croyance. En effet, on peut montrer que la combinaison de petits jeux d'observations pour en obtenir un plus grand mène quelquefois à des conclusions moins fiables. Ceci contredit évidemment l'intuition. Parfois, on peut même arriver à des conclusions complètement opposées à celles émises sur les petits jeux de données. Une illustration de cette apparente incohérence est donnée dans la suite.

Le tableau 1 montre les résultats moyens obtenus en physique par des étudiants en génie. Ce cours est considéré comme très rigoureux et difficile. Parmi ces étudiants, la plupart ont suivi préalablement un cours de physique au niveau secondaire, ce qui semble mieux les préparer au cours universitaire par rapport à ceux qui n'ont pas de formation préalable, comme en fait foi la différence de 10 points entre les deux groupes. Le tableau 2 présente le même genre de statistiques, cette fois pour des étudiants en art. Encore une fois, les étudiants possédant une formation préalable en physique possèdent un avantage de 10 points sur les autres. Jusqu'ici, pas de problème ! Les résultats semblent indiquer une nette supériorité pour les étudiants ayant suivi un cours de physique au secondaire.

Les choses se compliquent si on combine les deux jeux de données. Les résultats se retrouvent aux tableaux 3 et 4. À noter que la moyenne des résultats combinés en physique s'obtient en additionnant le total des scores pour les étudiants en génie et en art, puis en divisant par le nombre total d'étudiants. Comme vous pouvez le constater, les étudiants qui n'ont pas suivi une formation préalable en physique au secondaire ont des résultats supérieurs ! Mais quelle est la source de cette conclusion en totale contradiction avec celles tirées des tableaux 1 et 2 ?

D'abord, il est clair que la combinaison des jeux de données dresse un portrait erroné de la situation, et

ce pour deux raisons majeures. La première est que les résultats pour les deux groupes principaux (étudiants en génie et étudiants en art) sont influencés par une variable sous-jacente, à savoir le niveau de difficulté du cours. En fait, les étudiants en génie ont suivi un cours beaucoup plus difficile que les étudiants en art, ces derniers n'ayant pas besoin d'aller autant dans les détails de la compréhension. Par conséquent, les notes en génie ont été beaucoup plus basses. La deuxième provient de l'hétérogénéité des tailles des échantillons. En effet, bien que les tailles des groupes soient les mêmes, en l'occurrence 55, les tailles des sous-groupes sont différentes, ce qui a causé un déséquilibre lors de la concaténation.

En conclusion, il y avait quatre groupes dans cette étude, à savoir

- Étudiants en génie avec cours préalable en physique
- Étudiants en génie sans cours préalable en physique
- Étudiants en art avec cours préalable en physique
- Étudiants en art sans cours préalable en physique

Si ces quatre groupes avaient été de même taille, les résultats auraient effectivement concédé un avantage de 10 points aux étudiants ayant suivi une formation préalable en physique au secondaire, peu importe qu'ils soient en génie ou en art. Le paradoxe de Simpson apparaît quand on combine des échantillons de tailles inégales et quand une variable sous-jacente n'est pas prise en compte. L'exemple présenté en fournit une belle illustration.

Ce problème est contré par une expérience correctement planifiée, où toutes les variables susceptibles d'influencer les résultats sont identifiées et contrôlées. Ceci peut vouloir dire de simplement les éliminer, les garder constantes ou les inclure comme faisant partie intégrante de l'étude.

	Cours au secondaire	Pas de cours	Amélioration
n	50	5	---
Moyenne	80	70	10

Tableau 1. Scores moyens en physique des étudiants en génie

	Cours au secondaire	Pas de cours	Amélioration
n	5	50	---
Moyenne	95	85	10

Tableau 2. Scores moyens en physique des étudiants en art

	n	Moyenne	Total
Génie	50	80	4000
Art	5	95	475
Total	55		4475
Moyenne	---	81.4	---

Tableau 3. Scores moyens en physique pour les étudiants ayant suivi un cours de physique au secondaire

	n	Moyenne	Total
Génie	5	70	350
Art	50	85	4250
Total			4600
Moyenne		83.6	

Tableau 4. Scores moyens en physique pour les étudiants n'ayant pas suivi un cours de physique au secondaire

Vous cherchez de la formation ou de la consultation

- ☞ sur les logiciels SAS, SPSS, MINITAB ou Statistica,
- ☞ offerte par des professionnels,
- ☞ en séminaires publics ou dans votre organisation,
- ☞ adaptée à vos besoins,
- ☞ neutre et appuyée par 30 ans d'expérience?

CONTACTEZ-NOUS!

LES SERVICES CONSEILS
HARDY

Tél.: (514) 866-0871 (418) 626-1666
www.schardy.qc.ca

4715, des Replats, Bur. 260
Québec (Québec) G2J 1B8

440, boul. René-Lévesque O, Bur. 1250
Montréal (Québec) H2Z 1V7



Vilfredo Federico Samaso Pareto

15 juillet 1848 – 19 août 1923

Vilfredo Pareto est né à Paris où son père, le marquis Raffaele Pareto (1812-1882), un ingénieur d'une vieille famille génoise, était en exil pour des raisons politiques. Sa mère, Marie Métenier (1816-1889) était Française, et ainsi le français était la langue maternelle de Pareto. Ce ne fut que lors du retour de la famille dans le Piémont qu'il a appris l'italien. Il a eu deux sœurs qui n'ont pas eu d'enfant.

Sa première épouse, qu'il a mariée en 1889, était une aristocrate russe, Alexandrina (Dina) Bakounine, qui l'a laissé en 1900. Il a ensuite vécu pendant 20 ans avec une Française, Jeanne Regis, avec laquelle il s'est remarié deux ans avant sa mort. Comme il n'a pas eu d'enfant, il lui a laissé une fortune assez considérable, ayant auparavant fait don de sa bibliothèque à l'Université de Lausanne en 1908. Il a été enterré à Céligny, près de Genève, où il habitait depuis 1900.

Sa vie intellectuelle et professionnelle se divise assez clairement en trois phases.

1. La période de sa naissance à 1870 a été la phase de formation. Elle fut principalement scientifique et technique, même si Pareto a aussi appris le grec et le latin par lui-même et qu'il est devenu un connaisseur sagace de ces langues. Il a d'abord fréquenté l'Institut technique Leardi à Casale Monferrato (Piémont) de 1859 à 1861, et ensuite l'Institut technique royal de Turin (1861-1864). Ses études se poursuivirent à l'Université de Turin, où il a obtenu un diplôme en sciences mathématiques et physiques en 1867, ainsi qu'à l'école d'ingénierie appliquée de cette université. Il a obtenu son diplôme de génie après une thèse sur la *Théorie de l'élasticité des corps solides et l'intégration des équations différentielles qui définissent l'équilibre*.

2. De 1870 à 1873, Pareto a travaillé comme ingénieur sur les chemins de fer italiens, et à partir de 1873 pour une firme de matériel de chemin de fer à San Giovanni Valardo, en Toscane. Il a géré cette firme de 1880 à 1890 et il est devenu consultant de 1890 à 1893. Ce fut durant cette période, lorsqu'il a fait face aux problèmes économiques et sociaux de la gestion de la firme, qu'il s'est intéressé avec entrain à l'économie et la sociologie.

En 1874, il a pris part à la création de la Société Adam Smith. Durant cette période, il a écrit plusieurs articles pour *L'economista* de 1876, le *Journal des économistes* de 1887, et pour le *Monde économique* et le *Giornale degli Economisti*. Il a aussi présenté plusieurs communications à des congrès.

Il a été résolument en faveur de l'économie libérale, mais il a aussi été un pacifiste et un anticolonialiste. Il s'est présenté à l'élection de la Chambre des députés, mais sans succès.

Il a beaucoup voyagé à l'étranger, résidant occasionnellement en France, en Allemagne, en Autriche, en Angleterre, en Belgique et en Suisse. En 1891, il a rencontré Léon Walras à qui il a succédé en 1893 à la chaire d'économie politique à l'Université de Lausanne.

3. À partir de 1893, il s'est voué à l'enseignement et à la recherche. En tant qu'économiste, il a été un disciple de Antoine-Augustin Cournot et Léon Walras en économie mathématique. Sa principale innovation a été une définition de l'équilibre (l'optimum de Pareto) qu'on utilise toujours aujourd'hui. Mais Pareto n'a considéré la théorie qu'en tant que test pour la réalité. Ce fut l'une des raisons pour son intérêt envers la statistique et son travail dans ce domaine.

En tant que sociologue, il a été convaincu que les faits économiques étaient inséparables de la société qui les engendrait. Donc, de 1897 à sa retraite, il a aussi enseigné la sociologie.

Ce fut durant cette période que ses trois principaux livres ont été publiés : *Cours d'économie politique*; *Manuale di economica politica con una introduzione alla scienza sociale* (aussi publié en français sous le titre *Manuel d'économie politique*); *Trattato di sociologia* (aussi publié en français sous le titre *Traité de sociologie*).

En statistique, Pareto a travaillé dans trois domaines :

(a) Les méthodes d'interpolation et d'ajustement sur lesquelles il a écrit plusieurs articles, en particulier ceux dans le *Journal de la Société statistique de Paris*, et dans le *Zeitschrift für schweizerische Statistik*, et aussi dans les *Actes du IVème Congrès international de psychologie*.

(b) Les études d'actuariat sur les systèmes mutuels d'assurance et le calcul des pensions. Il a travaillé là-dessus en 1905 en réponse à une demande des employés des chemins de fer fédéraux suisses.

Les principaux travaux qu'il a publiés dans les domaines (a) et (b) ont été réédités en 1889 où son article sur les « Fonctions génératrices d'Abel » en mathématiques pures, publié dans le *Journal für die reine und angewandte Mathematik* de Kronecker, est aussi à découvrir.

(c) Mais ce fut surtout sa découverte en 1895 de la « courbe de loi pour la richesse et le revenu » qui a conféré à Pareto sa reconnaissance dans le monde statistique. Il a montré empiriquement que la loi observée des revenus, comme celle de la richesse, s'ajustait bien à une loi théorique du type

$$P(X > x) = \frac{K}{(x+c)^\alpha}, \quad x \geq x_0$$

où K et c sont respectivement des paramètres d'échelle et de position, tandis que l'exposant α , pour la série étudiée par Pareto, prend des valeurs entre 1 et 2.

Plus tard, cette sorte de loi s'est retrouvée dans différents domaines, non seulement dans les sciences sociales, mais aussi dans les sciences naturelles et physiques. Plusieurs modèles probabilistes mènent à ce type de loi : Pareto lui-même a tenté de développer des modèles probabilistes dès 1896. Trente ans plus tard, Maurice Fréchet a perfectionné les modèles de Pareto.

La motivation fondamentale qui conduit aux lois de Pareto a été donnée en 1935 par Paul Lévy. Ce dernier a montré que des lois stables, autres que la loi normale de Laplace-Gauss, montrent un comportement asymptotique de type (1) avec $0 < \alpha < 2$.

De plus, quand les conditions de convergence pour la loi normale dans le théorème central limite ne s'appliquent pas, notamment lorsqu'on joue avec des termes de sommes indépendants et identiquement distribués où les grandes valeurs ont des probabilités non négligeables (lois leptocurtiques), alors un résultat central limite s'applique encore, avec la loi stable avec l'exposant α figurant comme loi limite, au lieu de la normale.

C'est donc à Pareto que tout le crédit revient d'avoir brisé le monopole presque total de la loi gaussienne (normale) et ses domaines d'attraction en statistique. Malheureusement, ça ne lui a valu que le manque de compréhension et l'hostilité de ses pairs, en particulier Francis Ysidro Edgeworth.

[Traduction libre, par Pierre Lavallée, de Heyde, C.C., Seneta, E. (2001), *Statisticians of the Centuries*, Springer, New-York, pp. 232-235.]

Statistiquement vrai ?

Jésus aurait été marié et père

L'hypothèse n'est pas neuve, mais elle demeure explosive : de son vivant, Jésus-Christ aurait eu une épouse et un fils, aux côtés desquels il aurait été enterré. C'est nul autre que le réalisateur de *Titanic*, James Cameron, qui défend cette hérésie, qui se retrouve au coeur du documentaire *Le tombeau perdu de Jésus*, dont il est producteur. Le documentaire a été réalisé par Simcha Jacobovici, Canadien d'origine israélienne. « En tant que documentariste, je ne dois pas avoir peur de chercher la vérité, » a déclaré Cameron lors d'une conférence de presse à la *New York Public Library*, où ont été dévoilées les boîtes taillées dans le calcaire où auraient été recueillis les ossements de Jésus et de sa femme, Marie-Madeleine.

Engagé depuis 10 ans dans des projets cinématographiques touchant à la science et à l'exploration, Cameron a évoqué des « preuves tangibles, archéologiques et scientifiques » de l'existence d'un Jésus-Christ marié et père de famille. Mais il n'a pas fallu chercher loin pour entendre un autre son de cloche. « Je suis sceptique, c'est dans ma nature », a déclaré l'archéologue britannique Shimon Gibson, qui a participé en 1980 à l'exhumation de la tombe du quartier de Talpiot, à Jérusalem, où se trouvaient dix boîtes funéraires appartenant à une famille juive du premier siècle de notre ère. « Tout ce que je peux dire, c'est que l'hypothèse mérite d'être étudiée davantage », a ajouté Gibson, lors du point de presse new-yorkais.

Selon l'Autorité des antiquités d'Israël, les ossuaires de la tombe de Talpiot portent divers noms, dont ceux de « Jésus, fils de Joseph », « Judas, fils de Jésus », « Maria », « Joseph » et « Mariamne » (soit Marie-Madeleine en grec). En 1996, l'année où ces inscriptions ont été rendues publiques, la chaîne de télévision britannique BBC a présenté un court documentaire affirmant que la tombe de Talpiot était probablement celle de Jésus-Christ. Archéologues et religieux avaient sévèrement critiqué la chaîne britannique pour avoir diffusé de telles conjectures.

Simcha Jacobovici, réalisateur du *Tombeau perdu de Jésus*, revient néanmoins à la charge, apportant ce qu'il qualifie de nouvelles « preuves ». Il a notamment fait appel à un statisticien pour abattre l'argument de la fréquence des noms de Jésus, Judas, Maria et Joseph à l'époque. En tenant compte de ces noms, Andrey Feuerwerker, professeur de statistique à l'Université de Toronto, a conclu que la probabilité est de 600 contre 1 que le tombeau de Talpiot est bien celui de Jésus-Christ. Simcha Jacobovici a également fait appel à un expert de l'ADN qui a démontré que « Jésus, fils de Joseph » n'avait pas de lien sanguin avec « Mariamne » (Marie-Madeleine). C'est donc dire qu'ils étaient probablement mari et femme.

Mais ce Judas, qui serait né de leur union, pourquoi son nom n'est-il pas mentionné dans les Évangiles, comme l'est celui de Marie-Madeleine, la femme la plus importante du Nouveau Testament après Marie ? Les artisans du *Tombeau perdu de Jésus* ont une réponse à cette question, mais ils n'ont pas de preuve tangible. Ils soutiennent que Judas est ce « disciple bien-aimé » auquel font référence les Évangiles. « Si les évangélistes avaient nommé le fils de Jésus, ils auraient signé son arrêt de mort, » a déclaré l'archéologue et auteur Charles Pellegrino, qui a cosigné avec Simcha Jacobovici un livre portant le même titre que le documentaire. « Le disciple bien-aimé est là lors de la dernière Cène, et il se retrouve avec Marie-Madeleine au pied de la croix. »

Deux autres universitaires ont également participé à la conférence de presse, dont James Tabor, spécialiste en archéologie biblique, qui a tenté de réconcilier l'hypothèse de Cameron et cie et le dogme chrétien de la résurrection. L'ossuaire de « Jésus, fils de Joseph » ne prouve-t-il pas qu'il ne soit pas revenu à la vie après sa mort ? « Il y aura un débat théologique sur cette question, comme il y en a un sur la question de l'évolution », a dit Tabor qui enseigne à l'Université de Caroline du Nord. « Les fondamentalistes refuseront d'aller au-delà d'une interprétation littérale des Évangiles. Les autres pourront citer Paul, qui parle d'une résurrection spirituelle. »

Simcha Jacobovici s'est défendu d'avoir orchestré un coup publicitaire pour attirer l'attention sur son documentaire. « Ce n'est pas un coup publicitaire, je vous présente aujourd'hui des documents authentiques, » a-t-il déclaré, en faisant référence aux ossuaires de « Jésus, fils de Joseph » et de « Mariamne ».

[Tiré d'un article paru dans le journal *Le Droit*, le 27 février 2007.]

Chronique SAS

[Sylvain Tremblay](#), Groupe de formation – Institut SAS (Canada) inc.

Proc LOGISTIC lave plus blanc !

Nouveau et amélioré ! Il n'y a rien de tel qu'un bon slogan marketing pour attirer l'attention. En effet, avec la version 9 de SAS, la procédure LOGISTIC s'est enrichie de nouvelles instructions et options qui facilitent la vie aux utilisateurs pour la validation du modèle de régression logistique et son application pour le calcul de prévisions.

Notamment, l'option OUTMODEL = permet de sauvegarder dans un fichier de données SAS les informations relatives à un modèle que l'on vient d'estimer. Par la suite, on peut utiliser la procédure LOGISTIC à nouveau dans le but de calculer les prévisions (scores) pour un nouveau jeu de données en utilisant la nouvelle instruction SCORE et l'option INMODEL =. On peut ainsi calculer les probabilités prédites sans avoir à estimer de nouveau le modèle.

La première étape consiste à utiliser la procédure LOGISTIC pour estimer le modèle (voir l'encadré 1) :

Encadré 1

```
PROC LOGISTIC
  DATA=sasuser.b_sales_inc
    OUTMODEL=work.fitted_model
  ;
  CLASS gender (PARAM=REF REF='Male')
    inclevel (PARAM=REF REF='3')
  ;
  MODEL purchase(event='1')=gender
    inclevel
    gender*inclevel
  / LINK=logit
    OUTROC=work.roc
  ;
RUN;
QUIT;
```

L'option OUTMODEL = de l'instruction PROC LOGISTIC sert à spécifier un fichier de données SAS dans lequel seront écrites les informations (valeurs estimées des paramètres, etc.) du modèle résultant (estimé). Ce fichier de données est très compact et ne contiendra que le minimum d'informations requises pour calculer les probabilités prédites pour un nouveau jeu de données.

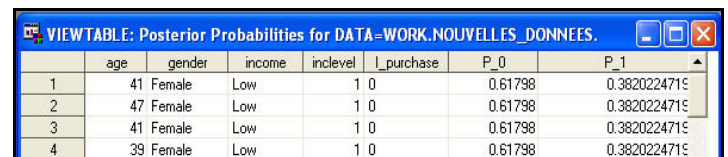
Ensuite, pour calculer les valeurs prédites de la probabilité de l'événement pour un nouveau jeu de données à partir du modèle que l'on vient d'estimer, on utilise à nouveau PROC LOGISTIC, mais cette fois avec l'option INMODEL =. Voir à ce sujet l'encadré 2.

Encadré 2

```
PROC LOGISTIC
  INMODEL=work.fitted_model
  ;
  SCORE DATA=work.nouvelles_donnees
    OUT =work.nouvelles_donnees_scored
  ;
RUN;
QUIT;
```

L'option INMODEL = sert à spécifier le fichier de données SAS contenant les informations nécessaires au calcul des probabilités prédites (modèle estimé), celui-ci ayant été créé à l'étape précédente. Ensuite, on utilise l'instruction SCORE avec l'option DATA = pour spécifier les nouvelles données pour lesquelles on désire des prévisions. Ce jeu de données n'a qu'à contenir les variables indépendantes requises par le modèle. De plus, on peut sauvegarder les prévisions dans un fichier de données SAS grâce à l'option OUT = de l'instruction SCORE. Dans l'exemple suivant, le fichier de données 'nouvelles_donnees_scored' va contenir de nouvelles variables, dont une portant le nom de P_1 (voir l'encadré 3). Celle dernière contient la probabilité prédite.

Encadré 3



	age	gender	income	inclevel	l_purchase	P_0	P_1
1	41	Female	Low	1	0	0.61798	0.3820224715
2	47	Female	Low	1	0	0.61798	0.3820224715
3	41	Female	Low	1	0	0.61798	0.3820224715
4	39	Female	Low	1	0	0.61798	0.3820224715

Pour la validation du modèle, il est maintenant simple d'obtenir un graphique de la courbe ROC ('Receiver Operating Characteristic Curve'), illustrant la relation entre la sensibilité et 1-spécificité de notre modèle. Pour ce faire, il suffit d'utiliser l'option OUTROC = de l'instruction MODEL lors de l'estimation du modèle (voir l'encadré 1) pour sauvegarder dans un fichier de données SAS les statistiques d'adéquation du modèle associées au ROC. Dans notre exemple, nous obtenons les variables suivantes dans le fichier de données 'work.roc' (voir l'encadré 4) :

Encadré 4

#	Variable	Type	Len	Label
7	_1MSPEC_	Num	8	1 - Specificity
5	_FALNEG_	Num	8	No. of Events Predicted as Nonevents
4	_FALPOS_	Num	8	No. of Nonevents Predicted as Events
3	_NEG_	Num	8	No. of Correctly Predicted Nonevents
2	_POS_	Num	8	No. of Correctly Predicted Events
1	_PROB_	Num	8	Probability Level
6	_SENSIT_	Num	8	Sensitivity

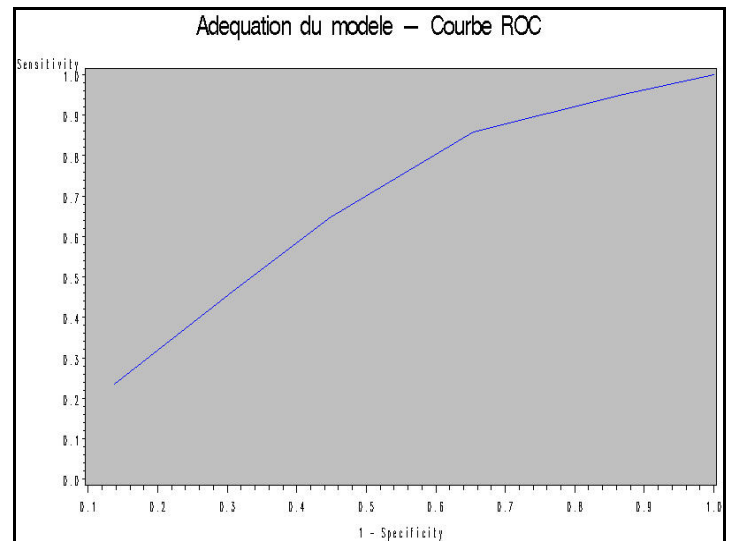
On peut ensuite utiliser SAS/GRAPH et ce fichier de données pour générer la courbe ROC (voir l'encadré 5) :

Encadré 5

```

SYMBOL1 i=join v=none c=blue;
PROC GPGLOT
  DATA = work.roc
  ;
  TITLE 'Adequation du modele - Courbe ROC'
  ;
  PLOT _sensit_*_1mspec_ = 1
      / vaxis=0 to 1 by .1 cframe=ligr
  ;
RUN;
QUIT;

```



En plus de la nouvelle instruction SCORE sous SAS version 9, la procédure LOGISTIC permet maintenant aussi d'effectuer des régressions logistiques conditionnelles, très utilisées dans les études de type 'cas-témoins' (c'est-à-dire 'case-control'). Pour ce faire, il suffit d'utiliser la nouvelle instruction STRATA qui permet de nommer les variables qui définissent les strates. À noter que la procédure PHREG permet également de réaliser les régressions logistiques conditionnelles, mais la procédure LOGISTIC possède l'avantage de pouvoir utiliser les instructions CLASS et CONTRAST.

Ne vous affolez pas si dans l'éditeur de SAS, les instructions SCORE et STRATA sont en rouge : elles sont valides. Il s'agit simplement d'une petite anomalie de l'éditeur, due au fait que ces nouvelles procédures n'ont pas encore été référencées. En attendant qu'elles le soient dans les prochaines versions, il est possible de déclarer ces nouvelles instructions de la façon suivante : Outils -> Options -> Éditeur amélioré -> General -> Mot-clés définis par l'utilisateur -> Ajouter -saisir "score" et "strata". Elles seront alors reconnues par l'éditeur et n'apparaîtront plus en rouge.

Donc pour conclure, la procédure LOGISTIC gagne encore en convivialité et en fonctionnalités. D'autres améliorations ont été apportées à cette procédure avec la version 9 de SAS, notamment l'ajout de nouveaux paramètres pour les variables de classe. Elles se retrouvent dans la section 'What's New in SAS/STAT® 9 and 9.1' de la documentation en ligne que je vous invite à consulter : <http://support.sas.com/onlinedoc/913/docMainpage.jsp>

Conférences à venir

Voici un choix d'activités futures, annoncées par différentes organisations statistiques. Si vous avez des conférences à annoncer, n'hésitez pas à nous contacter !

Quoi ?	Où ?	Organisateurs	Quand ?
			2007
8^{ème} Séminaire européen de statistique 2007	La Manga del Mar Minor (Espagne)	Universidad Politécnica de Cartagena (Note : certaines subventions sont disponibles)	6 au 12 mai 2007
Conférence socioéconomique 2007	Ottawa (Canada)	Statistique Canada	28 et 29 mai 2007
35e Congrès annuel de la SSC – 2007	Saint-Jean (Terre-Neuve)	Société statistique du Canada	9 au 13 juin 2007
39èmes Journées de statistique	Angers (France)	Société française de statistique	11 au 15 juin 2007
Conférence internationale en probabilité et statistique - Journées ESAIM P&S	Toulouse (France)	Université Paul-Sabatier - Toulouse III	14 et 15 juin 2007
3e Conférence internationale sur les enquêtes auprès des établissements (ICES-III)	Montréal (Québec)	American Statistical Association (ASA) , Société statistique du Canada , International Association of Survey Statisticians	18 au 21 juin 2007 [Note : en anglais] [Événement septennal]
JSM 2007 (Joint Statistical Meeting)	Salt Lake City (Utah, États-Unis)	American Statistical Association (ASA)	29 juillet au 2 août 2007
56e Congrès de l'Institut international de statistique	Lisbonne (Portugal)	Institut international de statistique (IIS)	22 au 29 août 2007 [Note : en anglais]
4èmes Rencontres Internationales Analyse Statistique Implicative	Castellón de la Plana (Espagne)	Universidad Jaume I	18 au 21 octobre 2007
Colloque francophone sur les sondages	Marseille (France)	Société française de statistique	5 au 7 novembre 2007
			2008
10èmes Journées européennes agro-industrie et méthodes statistiques	Louvain-la-Neuve (Belgique)	Société française de statistique	23 au 25 janvier 2008
36e Congrès annuel de la SSC – 2008 auquel sera jumelé le Congrès 2008 de l'ASSQ	Ottawa (Canada)	Société statistique du Canada , Société française de statistique et ASSQ	25 au 29 mai 2008

Pour ceux que cela intéresse, le site de l'ISI (Institut international de statistique, basé à La Haie, aux Pays-Bas), <http://isi.cbs.nl/calendar.htm>, dresse une liste assez exhaustive des événements internationaux importants à venir. L'horaire ci-dessus peut être sujet à changement. Consultez le site web des organisateurs pour de plus amples informations.

Nouvelles publications

Voici un choix de monographies, de livres de cours et autres ouvrages reliés à la statistique récemment publiés, principalement en français, Si d'autres ouvrages vous semblent intéressants, n'hésitez pas à nous contacter ; nous nous ferons un plaisir de les ajouter à notre répertoire !

Titre	Auteur	Éditeur	Date de parution	Notes ⁽¹⁾
Statistique descriptive - Appliquée à la gestion et à l'économie	Jérôme Hubler	Bréal	Mars 2007	Après une présentation des principaux concepts de la statistique descriptive, l'ouvrage insiste sur quelques méthodes et instruments souvent utilisés en économie et en gestion : l'analyse des relations entre deux variables, l'ajustement d'une droite de régression et l'étude des corrélations. Il met enfin l'accent sur la construction et l'interprétation des indices statistiques. (220 pages)
Statistiques et probabilités - En 30 fiches	Daniel Fredon	Dunod	Janvier 2007	Présentation de la statistique et des probabilités en trente fiches pédagogiques. Chaque fiche comporte : <ul style="list-style-type: none"> ▪ les idées centrales à connaître ; ▪ la méthode à mettre en œuvre ; ▪ des exercices corrigés.
Régression - Théorie et applications	Pierre-André Cornillon & Eric Matzner-Lober	Springer	Décembre 2006	Cet ouvrage expose de manière détaillée, exemples à l'appui, l'une des méthodes statistiques les plus courantes : la régression. Les premiers chapitres sont consacrés à la régression linéaire simple et multiple. Les auteurs expliquent les fondements de la méthode, tant au niveau des choix opérés que des hypothèses et de leur utilité. Ensuite, sont développés les outils permettant de vérifier les hypothèses de base de la régression. Une présentation simple des modèles d'analyse de la variance et de la covariance est faite. Les derniers chapitres sont consacrés au choix de modèles ainsi qu'à certaines généralisations de la régression. (302 pages)
The Cambridge Dictionary of Statistics	Brian S. Everitt	Cambridge University Press	Novembre 2006	Un dictionnaire de 442 pages avec plus de 3600 termes en statistique théorique et appliquée. Inclut aussi une courte biographie de plus de 100 statisticiens. [en anglais]
Simulation à événements discrets	Gérard Fleury Philippe Lacomme Alain Tanguy	Eyrolles	Novembre 2006	Ce livre introduit les techniques de simulation à événements discrets par des exemples de complexité croissante, en évitant autant que possible les exposés théoriques a priori. Avec CD-Rom d'accompagnement. (444 pages)
Analyse statistique des données spatiales	Jean-Jacques Droesbecke, Michel Lejeune & Gilbert Saporta	Technip	Novembre 2006	L'analyse statistique des données spatiales et spatiotemporelles est un champ de recherches intense, tant sur le plan théorique que sur le plan des applications. Cet ouvrage fait le point sur les développements les plus récents dans ce domaine. (470 pages)

⁽¹⁾ Les "Notes" sont extraites des notices descriptives préparées par les éditeurs.

Pour les intéressés, vous trouverez sur le site de l'[AMQ](#) dans la rubrique « Lu pour vous » des présentations d'un choix de livres de mathématiques et de statistique. Le bulletin de [mars 2006](#) est spécialement consacré au thème de la statistique.

Suivre son cours ...

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

Cette grille présente les cours offerts par différents organismes. La liste des cours universitaires est disponible sur demande auprès de l'ASSQ. L'information fournie dans cette grille est sujette à changements. Le lecteur est invité à communiquer avec les responsables des cours pour corroborer et compléter l'information présentée (frais d'inscription, dates).

Organisme, Lieu	Clientèle (basée sur la matière du cours)	Types de cours	Contacts
Creascience, Montréal	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 27 www.creascience.com info@creascience.com
Services conseils Hardy, Québec, Montréal	Débutant, intermédiaire et avancé	Différentes sessions de formation portant sur les logiciels SAS, SPSS, MINITAB, Statistica, Crystal Reports.	Monique Trempe (418) 626-1666 www.schardy.qc.ca schardy@schardy.qc.ca
Institut SAS Montréal, Québec, Ottawa	Grand public, entreprises, débutant ou expert	Data mining, langage macro, programmation, rédaction de rapport, statistique.	Chantal Lessard (514) 395-4071 http://support.sas.com/training/canada/ Chantal.Lessard@sas.com
École de technologie supérieure, Montréal	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
Statistique Canada, Ottawa	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.	Céline Charette (613) 951-1044 www.statcan.ca charcel@statcan.ca <u>Note</u> : Certains cours de Statistique Canada peuvent se dispenser à l'extérieur.

SÉMINAIRES

Des séminaires ont lieu régulièrement aux endroits suivants. N'hésitez pas à contacter les responsables pour plus de détails.

UQAM Pascale Rousseau Tél. (514) 987-3000, #3224 Fax (514) 987-8935 www.uqam.ca rousseau.pascale@uqam.ca	Université de Montréal Martin Bilodeau Tél. (514) 343-2410 Fax (514) 343-5700 www.umontreal.ca bilodeau@dms.umontreal.ca	Université de Sherbrooke Bernard Colin Tél. (819) 821-8000, #2012 Fax (819) 821-8200 www.usherbrooke.ca bernard.colin@dmi.usherb.ca	Université Laval Thierry Duchesne Tél. (418) 656-5077 Fax (418) 656-2817 www.ulaval.ca duchesne@mat.ulaval.ca
---	---	--	--