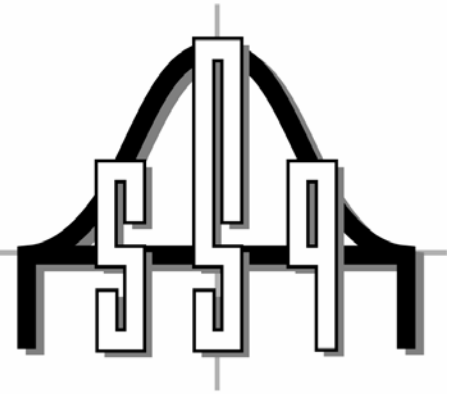


Convergence

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



Volume XI • Numéro 3

Décembre 2006

Chers membres,

Sans doute l'avez-vous déjà noté : *Convergence* a un nouveau rédacteur en chef ! Après avoir consacré beaucoup de temps et d'efforts à la relance du bulletin d'information de l'association depuis janvier 2005, Mireille Guay souhaitait en effet prendre un peu de recul et relever de nouveaux défis au cours de son second mandat à la direction des communications de l'association.

L'ASSQ est extrêmement redevable à Mireille d'avoir accepté au pied levé la lourde tâche de remettre *Convergence* sur les rails. Notre admiration et notre reconnaissance lui sont acquises pour le brio avec lequel elle a mené cette opération. Sa succession a été confiée en août dernier à Jean-François Quessy, pour un mandat de deux ans. Ce numéro est le premier qu'il réalise.

Originaire de Shawinigan, Jean-François Quessy a fait ses études de premier cycle en mathématiques à l'Université du Québec à Trois-Rivières (UQTR). Le professeur Bruno Rémillard (de HEC Montréal) et moi avons ensuite eu le plaisir de codiriger ses travaux de recherche de 2^e et de 3^e cycle à l'Université Laval. Pédagogue hors pair et chercheur dynamique ayant déjà à son actif plus d'une demi-douzaine de publications scientifiques, Jean-François est professeur de statistique à l'UQTR. Il a reçu cette année le prix Pierre-Robillard récompensant la meilleure thèse ès statistique soutenue dans une université canadienne. L'ASSQ est privilégiée de pouvoir compter sur ses services.

Une autre nomination d'importance est survenue à la fin de l'été : celle de Jean-François Plante au titre de webmestre de l'association. Une relève est ainsi assurée à Mike Sirois, qui avait exprimé le légitime désir de consacrer plus de temps à sa famille après quatre ans de bons et loyaux services. Mike était apprécié pour le soin méticuleux et constant qu'il apportait à l'entretien et à la mise à jour de la vitrine de l'association sur la toile. L'ASSQ lui devait beaucoup et le remplacer n'allait certes pas être chose aisée. Le Conseil d'administration croit néanmoins avoir trouvé en Jean-François Plante la perle rare. Ce dernier assume ses nouvelles fonctions depuis le 1^{er} septembre.

Né à Lévis, Jean-François Plante est titulaire d'un baccalauréat spécialisé en mathématiques et d'une maîtrise en statistique réalisés à l'Université Laval. Il est actuellement en rédaction de thèse à l'Université de la Colombie-Britannique, où il travaille sous la direction du professeur James V. Zidek. Coauteur de trois articles scientifiques avant même d'avoir entrepris ses études doctorales et déjà consacré « professeur étoile » à Laval pour de remarquables prestations de cours, gageons que Jean-François est déjà dans la mire de plusieurs universités québé-

coises. L'ASSQ peut néanmoins s'enorgueillir de l'avoir recruté la première, si ce n'est qu'à titre bénévole !

Nous nous réjouissons de l'arrivée des deux « Jean-François, » qui permettra de soutenir le dynamisme de notre association. Vous trouverez dans les pages suivantes moult autres preuves de sa vitalité retrouvée et, notamment deux comptes rendus des populaires « Jeudis de l'ASSQ » tenus cet automne à Québec (le 12 octobre) et à Montréal (le 2 novembre). Novices et récidivistes des « 5 à 7 » sont invités à retenir les lieux et les dates des prochaines rencontres : Trois-Rivières (le 22 février) et Québec (le 29 mars).

Mais surtout, réservez d'ores et déjà le vendredi 8 juin 2007 pour une réédition du « rendez-vous de Saint-Antoine, » selon la même formule éprouvée : deux ou trois exposés abordables présentés en plénière, assemblée générale de rigueur (mais pas rigoriste pour autant), petit coquetel et souper gastronomique dans le cadre champêtre du Lac Delage, près de la capitale nationale ! Tous les détails vous seront communiqués dans le prochain numéro de *Convergence*.

Toujours soucieux de répondre à vos attentes, le Conseil d'administration de l'ASSQ vous remercie de votre confiance et de votre appui. Recevez nos meilleurs vœux à l'occasion du Nouvel An.

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
Les Jeudis de l'ASSQ	5
Le tournoi de golf	8
Chronique historique	9
Au-dessus de la moyenne	10
Statistique et sport	11
« Pi populaire » et « Pi quoi encore »	12
Chronique SAS	16
Statistique et vie	3, 5, 7, 13, 14, 18
Conférences à venir	19
Nouvelles publications	20
Suivre son cours	21

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 expédier au rédacteur en chef leur document électronique prêt pour l'impression avant la date de tombée du prochain numéro. Les membres institutionnels bénéficient d'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 : les avis de cours et de séminaires en dernière page sont gratuits) :

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, 2006

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,

Lorsque notre président, Christian Genest, m'a contacté pour me proposer de remplacer Mireille Guay à titre de rédacteur en chef de *Convergence*, j'ai accepté sur-le-champ. D'abord, l'ASSQ est une sympathique organisation, et voilà, me dis-je, une belle opportunité de m'impliquer concrètement dans le milieu de la statistique au Québec. Compte tenu de l'excellent travail accompli par Mireille, j'accepte ce défi avec beaucoup d'humilité et espère simplement maintenir ses normes de qualité.

Au moment de la passation du flambeau, Mireille m'a mentionné que le style employé dans ce « Mot du rédacteur en chef » était libre. À la limite, dit-elle, tu pourrais l'écrire sous la forme d'un poème ! Soyez sans crainte, Dame poésie n'a pas frappé à ma porte depuis belle lurette. Je m'en tiendrai donc à décrire sobrement les contributions de nos précieux collaborateurs. Soit dit en passant, la qualité des textes et le dévouement des auteurs pour bâtir un numéro de *Convergence* m'ont impressionné.

Peut-être inspiré des conseils de Mireille à mon égard, Claude Ouimet s'est commis dans une œuvre poétique qui décrit le dernier tournoi de golf de l'ASSQ tenu le 9 septembre dernier, avec photos en prime. Toujours dans la sphère créatrice, un texte de David Bellhouse vous est présenté pour commémorer le 250^e anniversaire de naissance de Mozart. Vous verrez que stochastique et artistique vont parfois de pair.

Dans un texte de son cru, Pierre Lavallée nous montre que pour performer « au-dessus de la moyenne » jour après jour, on est condamné à toujours se surpasser, puisque cette moyenne va elle-même augmenter au fil du temps. Pas question donc, de s'asseoir sur ses lauriers ! Faisant suite à son texte sur les A's d'Oakland paru dans le numéro du mois d'août

dernier, Martin Rioux nous explique cette fois comment les Packers de Green Bay, de la Ligue Nationale de Football, utilisent les outils statistiques pour améliorer leurs performances sur le terrain. Parions que d'autres équipes professionnelles tenteront aussi de développer leurs propres méthodes.

Pierre Lavallée nous revient avec deux textes sur le fascinant nombre π . Dans le premier, on verra que virtuellement n'importe quelle suite de chiffres se retrouve dans le développement décimal de π . Je peux ainsi affirmer sans risque que la suite des huit chiffres de votre date de naissance (JJMMAAAA) y apparaît quelque part ! Dans son deuxième texte, Pierre décrit un moyen mnémotechnique pour retenir les cinquante-quatre premières décimales de π , et par le fait même impressionner vos amis ! Devinez quoi ? Ce truc est basé sur un poème !

Un membre de l'ASSQ, Nathalie Madore, a gentiment attiré mon attention sur un texte qui critique un sondage paru peu de temps après l'effondrement du viaduc de la Concorde à Laval. L'article est reproduit ici en entier. Finalement, dans sa « Chronique SAS », Sylvain Tremblay nous entretient de la nouvelle fonctionnalité *PROC GLIMMIX*, qui permet l'analyse de données via les *modèles linéaires généralisés mixtes*.

Pour terminer, je transmets à tous les membres de l'ASSQ mes meilleurs vœux de bonheur et de sérénité pour l'année 2007. Puisse ce temps de repos avec vos proches vous apporter l'énergie et la motivation nécessaires pour relever les défis de la prochaine année.

[Jean-François Quessy](#), Ph. D.

Statistique et vie :

Aimez-vous les uns les autres...

Lasse des agressions dont sont victimes près de 15 % des prélats, l'Église d'Angleterre a décidé d'initier les membres du clergé aux sports de défense. La pratique des arts martiaux devrait leur permettre de neutraliser les mendiants violents, les notables furieux de se voir refuser une date de baptême ou de mariage et les voleurs de troncs. En cas d'agression, les hommes d'église ne tendront plus l'autre joue mais un poing ferme sur le nez d'en face...

[Tiré de l'*Almanach Vermot* 2005.]

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 CA

[Claude Ouimet](#), secrétaire

Quelques réunions ont été tenues depuis le retour des vacances d'été. À la lecture de l'article qui suit, vous verrez que les décisions prises au CA ne sont pas de nature à ébranler les fondements de la statistique. Nous vous en faisons part tout de même :

- Le post-mortem du Colloque tenu le 9 juin ayant été somme toute très positif, les démarches pour organiser une seconde édition en juin 2007 sont à leurs premiers balbutiements.
- Concernant les « 5 à 7 » organisés à Montréal et Québec, la formule connaît un retentissant succès. Elle a réussi à attirer plus d'une trentaine de participants à Québec et une vingtaine à Montréal. En Outaouais, les volontaires pour s'impliquer dans l'organisation ne se ruent pas à notre porte. Pour en avoir discuté avec quelques membres de l'ASSQ travaillant à Statistique Canada, la multitude de séminaires offerts à l'interne rend peut-être moins nécessaire la tenue d'un « 5 à 7 » dans cette région. Ceci expliquerait l'absence de volontaires. Pour l'instant, cette région ne sera donc pas desservie tant que la demande ne sera pas plus manifeste.
- Petite pause remerciements au travers des annonces... ça fait toujours plaisir d'en recevoir quand les implications sont bénévoles :
 - à Mireille Guay et Jean-François Quessy, pour le travail effectué et à venir à la tête de *Convergence* ;
 - à Mike Sirois et Jean-François Plante, pour les efforts passés et futurs comme webmestre ;
 - à Éric Gagnon de l'ISQ, pour avoir accepté la présidence d'élection pour les postes à combler au CA.
- L'ASSQ s'est départi récemment de plusieurs articles, sous forme de dons dans certains cas ou en échange de sommes symboliques dans d'autres cas. L'ordinateur et l'imprimante utilisés jadis par Isabelle Marchand pour la mise en page du *Convergence* ainsi que le micro-ordinateur portable utilisé par notre ancien webmestre Mike Sirois ne font plus partie des actifs de l'association.
- Si vous n'aviez pas renouvelé votre adhésion pour deux ans, vous avez probablement reçu des nouvelles de notre registraire Michel Fluet, la campagne d'adhésion ayant été devancée de quelques mois cette année. Afin d'accroître le nombre de membres, la sollicitation devrait s'effectuer également auprès des universités Laval, de Montréal et de Sherbrooke via un envoi destiné à l'ensemble des étudiants dans le domaine des math/stat.
- Enfin, le CA s'est penché sur le problème survenu dans la dernière année concernant les messages acheminés dans la boîte de nouvelles pour lesquels aucun suivi n'était donné, puisque les messages n'étaient plus transmis automatiquement à qui de droit au CA.
- Si vous avez des commentaires ou autres concernant les sujets traités ou encore d'autres thèmes que vous aimeriez voir abordés, parlez-en : votre CA est à l'écoute de vos besoins. Ou encore, manifestez-vous via le groupe de courriels.

Les Jeudis de l'ASSQ...

Le 12 octobre...

Dans l'enceinte de l'Université Laval, le 12 octobre dernier, se sont rassemblés une trentaine de membres de l'ASSQ pour écouter attentivement ce que M. Paul Bergeron, conférencier invité pour l'occasion, avait à dire sur ses expériences passées au Ministère des Transports. Ayant cumulé au-delà d'une vingtaine d'années d'expérience au sein de ce ministère, M. Bergeron a su captiver l'attention des jeunes et des moins jeunes concernant ses expériences vécues dans le cadre des nombreuses fonctions qu'il a occupées. Soulignons la diversité de provenance de l'auditoire : une dizaine d'étudiants, une douzaine d'employés de la fonction publique, cinq représentants de l'Université Laval et cinq représentants du secteur privé. L'oraison de M. Bergeron fut appréciée de tous.

Il va de soi de souligner la participation conjointe de Michel Fluet de SOM et Diane Leroux du Ministère des Transports pour l'organisation de cette activité, avec la complicité de Thierry Duchesne de l'Université Laval.

Les nombreux participants ont terminé en beauté cette rencontre en bavardant autour de succulentes petites bouchées préparées pour l'occasion, avec, il va sans dire, des rafraîchissements ainsi que quelques vins soigneusement choisis pour accompagner le tout.

Claude Ouimet
claudiouimet@mtq.gouv.qc.ca

Statistique et vie :

Il n'y a pas de sottises taxes

Les dealers taxés, c'est au Québec. Le hasch et la coke vont doper les finances du fisc québécois.

Revenu Québec réclame le paiement de la taxe de vente du Québec (TVQ) à cinq gros trafiquants - des Hells Angels ayant écoulé pour 40 millions de dollars canadiens de drogue (plus de 24 millions d'euros) avant leur arrestation.

La cotisation réclamée correspond à 7,5 % du montant des ventes, plus les pénalités de retard. L'avis de cotisation le plus important (2,1 millions de dollars canadiens) a été adressé à Gerald Matticks, membre influent du "gang de l'Ouest". "Tous les biens vendus dans le cadre d'une activité commerciale sont sujets à taxation", rappelle le Ministre du Revenu, Lawrence S. Bergman. Seuls sont exemptés de TVQ l'épicerie et des articles vendus par les sociétés dont le chiffre d'affaires n'excède pas 30000 dollars canadiens par an". Les petits dealers peuvent dormir sur leurs deux oreilles.

[Tiré de l'Almanach Vermot 2005.]

Les Jeudis de l'ASSQ...

Le 2 novembre...

Pour une deuxième année consécutive, l'UQAM-SCAD a été l'hôte d'un « Jeudi de l'ASSQ » qui s'est déroulé le 2 novembre dernier à Montréal.

Dix-neuf personnes, dont quatre étudiants, ont participé à ce « 5 à 7 ». Plusieurs personnes en ont profité pour rejoindre les rangs de l'association.

Le conférencier, M. Victor Tremblay, président de *Statplus*, experts-conseils en statistique, avait choisi de nous entretenir de la « *Problématique de la non-réponse dans les sondages* », laquelle affecte toute étude statistique basée sur la collecte de données. De nombreux graphiques et des simulations actives ont permis aux participants de suivre l'impact de la non-réponse sur la précision des résultats.

Après l'exposé, une dégustation de vins et fromages a été offerte aux participants. À cette occasion, de nombreux contacts ont été établis entre le conférencier, les professionnels, les professeurs et les étudiants.

Nous adressons nos remerciements les plus sincères au SCAD, le *Service de Consultation en Analyse de Données*, pour avoir mis gracieusement à notre disposition la salle de conférence et l'équipement nécessaire.

Nous vous attendons encore plus nombreux aux prochains « Jeudis de l'ASSQ » qui se tiendront à Trois-Rivières le 22 février 2007 et à Québec le 29 mars 2007. À noter à votre agenda !

Le conférencier en action



L'assistance suit avec intérêt la gestion de la non-réponse

Joseph Nader
joseph-n@mtl.feric.ca

Les prochains « Jeudis de l'ASSQ »

Deux rencontres des désormais traditionnels « Jeudis de l'ASSQ » auront lieu pendant la saison hivernale. Comme à l'habitude, un conférencier viendra nous entretenir d'un sujet intéressant en statistique. La présentation sera suivie d'échanges cordiaux entre les participants, le tout agrémenté, bien sûr, d'une dégustation de canapés et de vins.

Quand : 22 février 2007
Lieu : Université du Québec à Trois-Rivières, Pavillon Ringuet

Quand : 29 mars 2007
Lieu : Université Laval, Pavillon Alexandre-Vachon

Les noms des conférenciers ainsi que les titres des présentations seront annoncés au début de l'année 2007. Chose certaine, les exposés sauront intéresser le statisticien en vous !

Soyez-y nombreux !

Statistique et vie :

Air Gauloise

On pensait que le nouveau confort des avions (relatif en classe touriste) ferait s'adoucir les moeurs des passagers. Il n'en est rien. Air France faisait part de 500 incidents en 1998. Cette année, le nombre sera en augmentation de 400 % ! L'incivilité, les libations des fêtards, le stress de certains passagers, les retards dans les horaires, tout cela entraîne des dérapages préoccupants.

La privation de tabac est aussi à l'origine de nombreuses altercations entre des passagers et le personnel navigant. Un espace fumeur sera peut-être envisagé, dit-on, si la situation devenait plus préoccupante.

Tournoi de golf annuel du 9 septembre

Samedi le 9 septembre dernier
Malgré un début de tournoi sous une ondée
Et un ciel ennuagé le reste de la journée
19 golfeurs bien comptés

Se sont tout de même rassemblés
Pour cet événement comptant une douzaine d'habités
Avec comme intention de la petite blanche la frapper
Pour tenter dans le trou de l'y envoyer

Mission accomplie pour ces 19 braves
Malgré une Dame Nature et ses quelques entraves
De se faire mouiller un peu n'est pas bien grave
Se dirent à l'unisson nos 19 braves



Pierre Lavallée, Steve Méthot, Miho Shibata et Serge Thiffault

À nouveau la région des Bois-Francs nous a bien reçus
Au golf Drummond, à St-Majorique, tous se sont rendus
Nul désistement, malgré le mauvais temps prévu
Un gros merci à tous les participants d'y être venus

Pour boire, bien manger et quoi d'autre ?
Au resto « À la bonne vôtre »
Tous étaient attendus
Chacun en ressortant rassasié et bien repu

Je vous convie à l'an prochain, pour la neuvième édition
Le deuxième samedi de septembre comme de raison
Sur un terrain à déterminer, autant que la région
Espérant que de nouveaux adeptes s'y ajouteront



Claude Roy, Nathalie Ebnoether, Louise Bisson et Gilles Therrien



Suzie Bond, Daniel Hurtubise, Élyse Archambault et Sylvain Masse



Marco DeThomasis, Alfonso Troisi, Alain Nadeau et Julien Beauséjour

Les piles de l'appareil photo ayant rendu l'âme, le dernier trio n'a malheureusement pas pu être photographié. Merci à Gilles Grenon et Anne-Marie Lemire qui ont complété mon groupe de trois pour cette occasion.

[Claude Ouimet](#), secrétaire

Anniversaire historique: Wolfgang Amadeus Mozart

David Bellhouse, University of Western Ontario

Le 27 janvier 2006, le monde a célébré le 250^e anniversaire de naissance de Wolfgang Amadeus Mozart. Les célébrations ont aussi été associées à une série de deux ans de récitals et autres événements musicaux en son honneur. Mozart était un enfant prodige au violon et au piano. Il a composé son premier menuet à l'âge de cinq ans et sa première symphonie à neuf ans. À sa mort en 1791, le nombre de ses compositions s'élevait à plus de six cents.



Qu'est-ce que Mozart a à voir avec les probabilités et la statistique ? La réponse est de nature historique. Au XVIII^e siècle, nombreux étaient ceux qui étaient fascinés par les percées réalisées autour des mathématiques du hasard. Grâce à elle, le hasard n'était plus perçu comme un chaos complet ou un caprice des dieux ; il était dorénavant sujet à des lois mathématiques connues. La puissance de la déesse Fortuna avait été battue par quelques mathématiciens mortels !

On sait maintenant que le hasard a une structure et c'est là que Mozart entre en jeu. Est-il possible de composer une œuvre musicale qui soit en quelque sorte imprévisible, ou sujette au hasard, mais qui, en même temps, ait une structure musicale agréable ? C'est ce que Mozart a fait en 1787 dans son *Musikalisches Würfelspiel* en do (K.516f) pour piano. La pièce, qui compte 32 mesures, est un menuet de 16 mesures suivi d'un trio de 16 mesures. Chaque mesure est générée par un jet de dés. Deux dés sont lancés 16 fois, un jet pour chaque mesure, afin de déterminer le menuet, et un dé est lancé 16 fois pour déterminer chaque mesure du trio. Pour le menuet, Mozart a écrit 176 mesures au total. Un tableau de 11 x 16 a été produit ; les lignes correspondant à la somme des faces des dés, et les colonnes correspondant aux numéros des mesures du menuet.

Les valeurs du tableau sont les nombres 1 à 176 correspondant aux 176 mesures que Mozart a écrites. En parcourant les colonnes du tableau, les jets répétés des dés indiquent quelle ligne consulter dans le tableau pour obtenir la mesure appropriée à choisir. Un autre tableau est donné pour le trio, ce dernier de taille 6 x 16, et le même principe peut être appliqué pour déterminer ce morceau. Ceci est illustré dans le tableau ci-contre (montrant seulement les huit premières mesures) d'une autre œuvre musicale par dés attribuée à Mozart et publiée en 1798. Un examen attentif de la partition complète montre qu'il n'y a que deux différentes mesures d'ouverture données pour le menuet, et deux différentes mesures de clôture. Un peu d'arithmétique montre qu'au total, cette méthode conduit à $11^{14} \times 2^2 = 1,518,999,334,332,964$ menuets possibles.

ZAHLENTAFEL.
TABLE de CHIFFRES.

Erster Theil.
Premiere Partie.

	A	B	C	D	E	F	G	H
2	96	22	141	41	105	122	11	30
3	32	6	128	63	146	46	134	81
4	69	95	158	13	153	55	110	24
5	40	17	113	85	161	2	159	100
6	148	74	163	45	80	97	36	107
7	104	157	27	167	154	68	118	91
8	152	60	171	53	99	133	21	127
9	119	54	114	50	140	86	169	94
10	98	142	42	156	75	129	62	123
11	5	87	165	61	125	47	147	33
12	54	130	10	103	28	97	106	5

Aujourd'hui, vous pouvez composer votre propre menuet ou trio en allant sur le site <http://sunsite.univie.ac.at/Mozart/dice/>

[Article aussi paru dans *ISI Newsletter*, volume 30, no 1, 2006.]

Au-dessus de la moyenne

Pierre Lavallée, Statistique Canada

Il y a quelque temps, je discutais avec mon patron du rendement des employés. Il a alors mentionné tout bonnement : « Ils sont tous au-dessus de la moyenne. » Il s'est alors arrêté quelques secondes et a ajouté : « en autant que cette phrase veuille dire quelque chose. » En effet, quoiqu'elle ait un sens vantant le mérite des employés, cette phrase n'a pas de sens du point de vue statistique. Il est, en fait, mathématiquement impossible que tous les employés soient au-dessus de la moyenne. Soit x_i , la valeur de l'employé i , et soit $\bar{X} = \sum_{i=1}^N x_i / N$, la moyenne des N employés. Si $x_i > \bar{X}$ pour tout i , on en déduit que $\sum_{i=1}^N x_i > \sum_{i=1}^N \bar{X} = N\bar{X} = \sum_{i=1}^N x_i$ et donc $\sum_{i=1}^N x_i > \sum_{i=1}^N x_i$, ce qui est impossible. Il faut voir dans la phrase de mon patron que la moyenne dont on parle est en fait une moyenne historique basée sur des données d'une période antérieure. Cette période antérieure peut avoir trois définitions possibles :

- (1) elle peut comprendre tous les anciens employés, c'est-à-dire ceux qui ont maintenant quitté l'entreprise ;
- (2) elle peut comprendre tous les anciens employés, ainsi que tous les employés actuels qui ne font pas objet de l'évaluation de ce rendement ;
- (3) elle peut comprendre tous les employés actuels, excluant ceux qui font l'objet de l'évaluation de rendement.

Dans le langage courant, on s'entend généralement à dire que la définition (2) est probablement celle à laquelle on fait le plus souvent référence.

Si $\bar{X}_a = \sum_{i=1}^{N_a} x_i / N_a$ représente la moyenne des employés de la période antérieure, on peut alors avoir sans problème $x_i > \bar{X}_a$ pour $i = 1, \dots, N$. Maintenant, au fur et à mesure que le temps passe, les employés de la période courante s'ajoutent à la période antérieure et on calcule alors la nouvelle moyenne antérieure \bar{X}_{a+c} sur les deux ensembles d'employés, c'est-à-dire

$$\begin{aligned}\bar{X}_{a+c} &= \frac{\sum_{i=1}^{N_a} x_i + \sum_{i=1}^N x_i}{N_a + N} \\ &= \frac{N_a \bar{X}_a + N\bar{X}}{N_{a+c}}\end{aligned}$$

où $N_{a+c} = N_a + N$. Si, comme le mentionne mon patron, $x_i > \bar{X}_a$ pour les N employés, alors

$$\begin{aligned}\bar{X}_{a+c} &= \frac{\sum_{i=1}^{N_a} x_i + \sum_{i=1}^N x_i}{N_{a+c}} \\ &> \frac{\sum_{i=1}^{N_a} x_i + \sum_{i=1}^N \bar{X}_a}{N_{a+c}} \\ &= \frac{N_a \bar{X}_a + N\bar{X}_a}{N_{a+c}} = \bar{X}_a\end{aligned}$$

et donc $\bar{X}_{a+c} > \bar{X}_a$. On voit alors que la moyenne des employés va croître avec le temps, ce qui implique que les employés ont un rendement supérieur aux précédents, ou que les précédents étaient moins bons que les nouveaux. Après vingt-et-un ans à l'emploi de Statistique Canada, dois-je y voir une incitation à la retraite ? Ça mérite réflexion, mais histoire de ne pas trop déprimer, continuons la discussion...

Il est intéressant de voir que le raisonnement ci-dessus s'applique à tous les jours dans l'annonce des températures saisonnières. En effet, en disant « il fait aujourd'hui deux degrés au-dessus des normales saisonnières », ces mêmes degrés vont éventuellement faire partie du calcul de la moyenne des années ultérieures, et ainsi de suite. Si les températures sont toujours au-dessus des normales saisonnières, on assiste alors à une croissance de la moyenne au fil du temps, ce qui crée alors une escalade qui peut être dangereuse pour l'environnement. Mais ne vous inquiétez pas, mon voisin Georges m'a dit l'autre jour que tout est sous contrôle et que si la moyenne augmente, il saura la faire redescendre. Parlait-il de l'environnement, ou de son propre rendement ?

La statistique dans le sport :

Le département de R&D des Packers de Green Bay

Martin Rioux, Promaintech Novaxa

L'équipe des Packers de Green Bay, de la Ligue Nationale de Football américain (NFL), bien connue pour sa tradition d'excellence, est à l'avant-garde de la prise de décision objective à l'aide de l'analyse statistique. Le leader de cet effort chez les Packers est Mike Eayrs, premier et seul directeur du département de recherche et développement (R&D) depuis sa création par les Packers le 9 mars 2001.

Depuis six ans, Mike Eayrs est responsable de gérer et de développer des bases de données en lien avec un match de football. À l'aide de ces données, il est en mesure d'appuyer les entraîneurs et les dirigeants de l'équipe dans leurs diverses prises de décision en temps réel, sur les meilleures tactiques de jeux à employer en cours de match (patron de jeu offensif versus formation défensive et vice-versa, tendances tactiques particulières de l'adversaire, etc.).

L'un des aspects les plus importants du travail de Eayrs est l'utilisation de la vidéo numérique pour ramasser des données sur la performance des joueurs en fonction de diverses situations de jeu. La force de la base de données est la possibilité de trier et d'analyser conjointement de l'information quantitative et qualitative. Il y est possible d'obtenir des réponses de type « Qui ? » « Quoi ? » « Quand ? ». Ensuite, grâce à la vidéo, il est possible de visionner les séquences correspondantes afin de répondre aux questions « Pourquoi ? » et « Comment ? ». Cela permet d'analyser rapidement la performance ou l'échec de toute séquence de jeu ou de chacun des joueurs en présence. En documentant les performances de jeu de chaque joueur, quantifiées en termes de nombre de passes, courses, verges gagnées ou perdues, et en analysant ensuite la technique et le succès de chaque jeu complété, les Packers parviennent à obtenir la vision de ce qu'un joueur exécute bien et des aspects du jeu qu'il doit améliorer. Le personnel d'entraîneurs peut ensuite entrevoir rapidement les problèmes, initier des actions correctives, et même améliorer davantage les bonnes performances des joueurs étoiles. Pour accélérer l'acquisition de données utiles, mêmes les séquences de jeu lors des entraînements de l'équipe sont souvent filmées et mises à contribution pour alimenter la base de données.

En complément aux données qui sont la propriété des Packers, la base de données est alimentée par les statistiques officielles de chaque match. Le rapport de chaque match provient du bureau de la ligue qui diffuse sous une forme standardisée, disponible à toutes les équipes, les alignements, les statistiques individuelles et d'équipes (passes complétées, plaqués, interceptions, etc.), en plus des verges gagnées de chacune des séquences de jeu du match. À ces informations, les Packers vont même ajouter la séquence de jeu de l'entraîneur adverse en réaction au jeu précédent, ceci afin de déceler des tendances particulières dans la stratégie de l'adversaire.

Les jours précédents un match des Packers, l'équipe R&D est responsable de déterminer le profil de l'opposant basé sur ses performances face à des situations et des équipes similaires et d'en faire rapport aux entraîneurs. Les images vidéos et les diverses statistiques peuvent être jumelées, ce qui permet aux entraîneurs de faire des requêtes vidéos triées par situation — si long 3^e jeu, si premier jeu, si jeu de course, si passe voilée, etc. — et par joueurs impliqués. Il est aussi possible par ces requêtes de faire une compilation de l'efficacité des différents jeux truqués connus à ce jour ou de tenter de reconnaître les jeux à venir de l'adversaire parce que, par exemple, « le numéro 65 lève son pied gauche avant tout jeu de passe... ».

Durant la partie, chaque choix de jeu de l'adversaire est recensé pour déterminer si le tout correspond au profil pré-établi. Sinon les entraîneurs sont avisés aussitôt. Lorsqu'une combinaison offensive-défensive a une forte tendance probabiliste de gains ou de pertes de terrain, les entraîneurs sont également avisés en temps réel de cette information (ce qui explique certains temps d'arrêt demandés tout juste avant d'amorcer un jeu qui, de toute évidence, ne sera pas réussi). Le logiciel de prises de décision en temps réel développé par les Packers pour appuyer les choix de jeu considère plus de 70 variables !

Au fil des années, l'équipe R&D a aussi consacré plusieurs heures de recherche pour déterminer des statistiques valides en mesure de prédire les probabilités de gagner un match. Certaines de ces mesures (tenues secrètes par les Packers) sont jusqu'ici fortement corrélées avec la victoire ! Cela permet aux Packers d'estimer la contribution nécessaire de chaque joueur pour espérer augmenter les probabilités que l'équipe gagne ses matchs. Ces analyses permettent de conseiller la direction dans le repêchage et la sélection du personnel des joueurs et dans l'amélioration de la performance individuelle de chacun des joueurs.

Les équipes de la NFL suivantes seraient les plus avancées technologiquement sur le plan de l'analyse statistique combinée à la séquence vidéo : Packers de Green Bay, Patriots de la Nouvelle-Angleterre, Chiefs de Kansas City, Broncos de Denver, Jaguars de Jacksonville et Bears de Chicago. Il est intéressant de constater que la plupart de ces équipes sont des habituées des récentes séries éliminatoires et qu'en date du 11 décembre 2006, leur fiche combinée pour la saison actuelle est de 46 victoires, 31 défaites. Est-ce qu'une fois de plus, la statistique serait utile pour gagner des championnats ?!!!

RÉFÉRENCES

- Conférence de John Jones, président des Packers de Green Bay, tenue à Milwaukee le 3 mai 2006 dans le cadre du « 60^e World Conference on Quality and Improvement », organisée par l'American Society for Quality
- <http://www.footballoutsiders.com/ramblings.php?p=226&cat=8>
- <http://www.eweek.com/article2/0,1895,1596885,00.asp>

π populaire

Le nombre π est fascinant. C'est un des nombres les plus populaires en mathématiques, et ce pour plusieurs raisons. D'abord, le seul fait de comprendre ce qu'il est, ce qu'il représente et comment il peut être utilisé a intrigué les mathématiciens depuis la nuit des temps. Son histoire au cours des quatre derniers millénaires a amusé le monde entier et a mené à des découvertes importantes, en plus de se révéler un défi persistant. Travailler sans relâche à obtenir une valeur de plus en plus exacte de π en déterminant combien de décimales les ordinateurs peuvent générer et à quelle vitesse ils peuvent le faire, est devenu un défi d'aujourd'hui pour les ordinateurs et les informaticiens. Les mathématiciens, pour leur part, continuent à chercher des algorithmes plus élégants (et efficaces) pour accomplir ces tâches. Bien que nous connaissions déjà des billions (10^{12}) de décimales de π , qui sait jusqu'où on pourra pousser la capacité des ordinateurs ? Cette quête de précision semble être le test technologique ultime !

Il y a une attraction touristique unique au monde par laquelle l'enthousiasme pour π a été démontré. En 1937, dans le hall 31 du Palais de la Découverte, — aujourd'hui le Musée des sciences de Paris, sur l'avenue Franklin D. Roosevelt —, la valeur de π a été affichée avec de grands chiffres de bois disposés en spirale au plafond d'une coupole. Cette réalisation, qui se voulait un hommage à ce nombre célèbre, comportait toutefois une erreur : elle s'appuyait sur l'approximation générée en 1874 par William Shanks, laquelle comportait une erreur à

la 528^e décimale ! Cette erreur a été détectée en 1946 et corrigée au plafond du musée en 1949.

Certains ont une fixation sur la fréquence des chiffres composant les décimales de π . Est-ce que les entiers 0 à 9 apparaissent avec la même fréquence dans l'ensemble du développement décimal de π ? Pour le déterminer, on peut construire un tableau de contingence qui donne le nombre d'apparitions de chaque nombre, dans un intervalle donné. Pour les premières 100 décimales de π , les fréquences sont-elles égales ? Sinon, sont-elles à peu près égales ? Espérer des fréquences égales dans les cent premières décimales de π ne serait pas très raisonnable. Lorsqu'on regarde les résultats, on s'aperçoit que la distribution est en fait loin d'être uniforme.

Il existe des tests statistiques pour déterminer si les petits écarts observés par rapport à l'hypothèse d'équidistribution sont attribuables au hasard. Le cas échéant, on dit que la distribution n'est pas significativement différente d'une loi uniforme. Il se trouve que tel est le cas pour la distribution des chiffres composant les décimales de π . La distribution des décimales de π donnée ci-dessous a été compilée par Kanada pour les premières 10^n décimales de π (on ne s'attarde, en fait, qu'aux décimales de π). Se référer à « Sample Digits for Decimal Digits of Pi », 18 janvier 2003, www.supercomputing.org/pi-decimal_current.html). Elle montre qu'il n'y a pas d'écart statistiquement significatif par rapport à la distribution uniforme. Kanada donne la distribution des $1,24 \times 10^{12}$ premières décimales, qui constitue le record pour la valeur de n obtenue à la fin de 2002.

The number of times the digits appear within the first 10^n places of π											
Digits	1 to 10^2	1 to 10^3	1 to 10^4	1 to 10^5	1 to 10^6	1 to 10^7	1 to 10^8	1 to 10^9	1 to 10^{10}	1 to 10^{11}	1 to 10^{12}
0	8	93	968	9999	99959	999440	999922	99993942	999967995	10000104750	99999485134
1	8	116	1026	10137	99758	999333	10002,475	99997334	1000037790	9999937631	99999945664
2	12	103	1021	9908	100026	1000306	10001092	100002410	1000017271	10000026432	100000480057
3	11	102	974	10025	100229	999964	9998442	99986911	999976483	9999912396	99999787805
4	10	93	1012	9971	100230	1001093	10003863	100011958	999937688	10000032702	100000357857
5	8	97	1046	10026	100359	1000466	9993478	99998885	1000007928	9999963661	99999671008
6	9	94	1021	10029	99548	999337	9999417	100010387	999985731	9999824088	99999807503
7	8	95	970	10025	99800	1000207	9999610	99996061	1000041330	10000084530	99999818723
8	12	101	948	9978	99985	999814	10002180	100001839	999991772	10000157175	100000791469
9	14	106	1014	9902	100106	1000040	9999521	100000273	1000036012	9999956635	99999854780

Nombre d'apparitions de chacun des entiers de 0 à 9 dans les 10^n premières décimales de π .

Kanada nous a aussi divertis avec l'obtention de son record de précision pour la valeur de π . Par exemple, il a cerné les répétitions de chiffres —12 pour être exact— à certaines positions des $1,24 \times 10^{12}$ premières décimales. Voici la liste de ces répétitions et la position de la décimale où elles commencent :

777777777777: à partir de la 368 299 898 266^e décimale de π
999999999999: à partir de la 897 831 316 556^e décimale de π
111111111111: à partir de la 1 041 032 609 98^e décimale de π
888888888888: à partir de la 1 141 385 905 180^e décimale de π
666666666666: à partir de la 1 221 587 715 177^e décimale de π

On trouve aussi des suites naturelles de chiffres (avec des zéros à chaque bout de la séquence) à différents endroits dans les premières $1,24 \times 10^{12}$ décimales de π . Les voici, avec l'endroit où elles commencent :

01234567890 : à partir de la 53 217 681 704^e décimale de π
01234567890 : à partir de la 148 425 641 592^e décimale de π
01234567890 : à partir de la 461 766 198 041^e décimale de π
01234567890 : à partir de la 542 229 022 495^e décimale de π
01234567890 : à partir de la 674 836 914 243^e décimale de π
01234567890 : à partir de la 731 903 047 549^e décimale de π
01234567890 : à partir de la 751 931 754 993^e décimale de π
01234567890 : à partir de la 884 326 441 338^e décimale de π
01234567890 : à partir de la 1 073 216 766 668^e décimale de π

Elles peuvent aussi être trouvées en ordre inverse :

09876543210 : à partir de la 42 321 758 803^e décimale de π .

Ceci ne constitue que quelques-uns des aspects divertissants des décimales de π . En effet, puisque celles-ci s'étendent à l'infini (même si nous n'en connaissons que les $1,24 \times 10^{12}$ premières !), on devrait trouver n'importe quelle suite de chiffres dans le développement décimal complet. Par exemple, la date de fondation des États-Unis (7-4-1776), c'est-à-dire 741776, apparaît à la 21 134^e décimale de π . Vous pouvez vous amuser à trouver n'importe quelle autre suite de chiffres. La manière la plus simple est de chercher sur Internet un site Web qui fait cela pour vous¹.

[Traduction libre par Pierre Lavallée d'un extrait de Posamentier, A.S., Lehmann, I. (2004), *Pi: A Biography of the World's Most Mysterious Number*, publié chez Prometheus Books.]

Statistique et vie :

Souffrez de ma présence

Vous avez mal ? Alors pensez à vous soigner.

Sept millions de Français souffrent de douleurs chroniques. Bien sûr, on fait des efforts depuis des années et on se soigne mieux que nos ancêtres, mais la prise en charge reste insuffisante et les médicaments efficaces trop peu utilisés.

On se demande pourquoi, dans beaucoup de cas, on persiste à souffrir...

[Tiré de l'*Almanach Vermot* 2005.]

¹ NdT : Certains diront « π après? », mais bon...

π quoi encore !

[Pierre Lavallée](#), Statistique Canada

Tout le monde connaît le nombre π . En statistique, il apparaît d'abord et avant tout dans l'expression de la loi normale. On le retrouve souvent en sondage pour dénoter les poids. Cependant, dans ce cas particulier, il ne s'agit pas du nombre π mais de la lettre grecque.

Quand il s'agit de donner la valeur de π , la plupart des gens mentionnent 3,1416. Ceux qui, comme moi, ont eu dans leur jeunesse une calculatrice scientifique *Texas Instruments* n'ont pas trop de problème à répondre 3,141592654. Maintenant, comment peut-on faire pour retenir plus de décimales de π ? Pour en retenir 126, apprenez par cœur ce petit poème :

*Que j'aime à faire apprendre un nombre utile aux sages !
Glorieux Archimède, artiste, ingénieur,
Toi de qui Syracuse aime encore la gloire,
Soit ton nom conservé par de savants grimoires !
Jadis, mystérieux, un problème bloquait
Tout l'admirable procédé, l'œuvre grandiose
Que Pythagore découvrit aux anciens Grecs.
O quadrature ! Vieux tourment du philosophe
Insoluble rondeur, trop longtemps vous avez
Défié Pythagore et ses imitateurs.
Comment intégrer l'espace plan circulaire ?
Former un triangle auquel il équivaudra ?
Nouvelle invention : Archimède inscrira
Dedans un hexagone ; appréciera son aire
Fonction du rayon. Pas trop ne s'y tiendra :
Dédoublera chaque élément antérieur ;
Toujours de l'orbe calculée approchera ;
Définira limite ; enfin, l'arc, le limiteur
De cet inquiétant cercle, ennemi trop rebelle
Professeur, enseignez son problème avec zèle*

Pourquoi ce poème en particulier ? Je vous laisse deviner. Et ma raison ne va pas de mal en π ...

Statistique et vie :

Dormir naturellement

Si l'on en croit la dernière étude faite sur le sujet, les Français et les Françaises restent fidèles aux traditions. À savoir: 49 % des personnes interrogées dorment avec une chemise de nuit ou un pyjama. Cette habitude se vérifie surtout parmi les personnes de plus de cinquante ans. Pour les 18-35 ans, c'est le tee-shirt qui a la cote des Français et des Françaises. Quant à la plus « naturelle » des tenues pour dormir, celle d'Adam ou d'Ève, ils sont 25 % (surtout les hommes) à la préférer. Il faut remarquer que 1 % des personnes interrogées n'ont rien eu à répondre. Des insomniaques, sans doute...

[Tiré de l'*Almanach Vermot* 2005.]

Quand opportunisme

devient irresponsabilité

Lundi, le 2 octobre passé, moins de quarante-huit heures après l'effondrement du viaduc de la Concorde à Laval, le *Journal de Montréal* publiait un sondage, dirigé par Léger Marketing, sur la perception de la population face à l'événement. On y lisait notamment que « les Québécois ont été tellement ébranlés par l'effondrement du viaduc à Laval qu'ils ont perdu toute confiance dans les infrastructures et affirment avoir peur désormais de passer sous ce type de construction ».

La conclusion aurait pu difficilement être différente en sondant les gens si rapidement après un événement spectaculaire dont on ne connaissait (et ne connaît toujours pas) réellement les causes. En fait, tout dans ce sondage menait à une conclusion alarmiste. Pour vendre de la copie, on capitalisait et profitait de l'émotion spontanée des gens pour poser des questions douteuses, pour ne pas dire biaisées. Voici l'analyse des quatre premières questions du sondage.

Question 1- Est-ce que vous pensez que cet accident au viaduc de la Concorde aurait pu être évité ?

Par définition, un accident est un événement imprévisible, mais qui rétrospectivement peut toujours être évité. De toute façon, comment les

gens pourraient-ils répondre sans connaître les causes de l'effondrement ?

Question 2 - Après la découverte des premiers débris tombés sous le viaduc, les autorités auraient-elles dû interdire l'accès à ce viaduc ?

Formulé de cette façon, l'énoncé est chargé et biaisé et ne peut qu'engendrer une réponse positive - 92% en fait. De plus, la rapidité avec laquelle les événements se sont produits ne permet pas réellement de parler « des autorités » mais plutôt d'un simple employé du ministère. Déjà dans la question, on sent un autre reproche, le déplacement de la responsabilité est entamé.

Question 3 - Est-ce que le gouvernement investit suffisamment d'argent ou non dans l'entretien des infrastructures routières au Québec ?

Et voilà ! Nous y sommes ! Encore la faute du gouvernement qui n'investit pas suffisamment. Le problème, c'est que la question ne tient pas compte des arbitrages que le gestionnaire des fonds publics doit faire et qu'après une catastrophe, les priorités sont bousculées. On aurait obtenu des résultats similai-

res le lendemain de la diffusion de n'importe quel problème de santé publique.

Question 4 - Êtes-vous favorable ou défavorable à l'imposition de péages sur les autoroutes afin de financer les infrastructures routières au Québec ?

Jamais le péage n'aura été si populaire. Poser la question à ce moment décrédibilise le résultat et les décideurs ne pourront jamais utiliser cette information en toute confiance.

En conclusion, ce qui importe dans ce dossier, c'est de réaliser que les sondeurs, qu'ils le veuillent ou non, sont investis d'une certaine responsabilité civique ou sociale. Les journalistes et les citoyens leur font généralement confiance et ils sont tenus à un minimum de rigueur et de retenue dans l'élaboration de nos instruments de mesure. En cautionnant une telle consultation publique, le sondeur ne fait qu'accroître l'anxiété collective tout en apportant peu au débat. Pour aider à vendre de la copie, le sondeur a agi de façon incompétente, mais surtout irresponsable.

[Tiré de la chronique « Point de vue » du 16 octobre 2006 sur le site de Ipsos (<http://www.ipsos-na.com>)]

Chronique SAS

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

Proc GLIMMIX - Modèles Linéaires Généralisés Mixtes

La statistique est un domaine en constante évolution. Conscient de cette réalité, SAS tient son module SAS/STAT à jour en ajoutant de nouvelles procédures et en améliorant les plus anciennes. Une des dernières additions aux procédures de SAS/STAT est Proc GLIMMIX.

Proc GLIMMIX est la suite logique de Proc MIXED pour les modèles mixtes. En effet, cette nouvelle procédure permet de faire de l'estimation et de l'inférence pour les modèles linéaires généralisés mixtes (GLMM). Elle sera livrée avec la version 9.2 de SAS mais elle est actuellement téléchargeable à partir du site support.sas.com de SAS. Elle est en version production (et non expérimentale) et est compatible avec la version 9.1 de SAS sous Windows.

Le but de cet article est de donner un bref aperçu des modèles linéaires généralisés mixtes, d'illustrer la syntaxe de base de Proc GLIMMIX et d'aider les statisticiens à identifier correctement quelle procédure de SAS/STAT ils doivent utiliser dans le contexte des modèles linéaires.

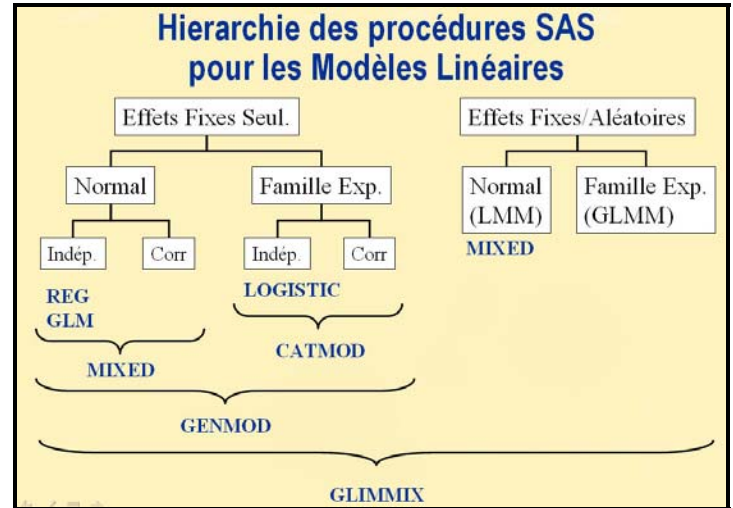
Modèles Linéaires Généralisés Mixtes (GLMM)

Ces modèles sont utilisés pour modéliser des données corrélées dont la variance n'est pas constante lorsque la variable dépendante n'est pas nécessairement normalement distribuée. Tout comme pour les modèles linéaires mixtes, les modèles GLMM supposent la normalité des effets aléatoires. Conditionnellement à ces effets aléatoires, les données peuvent suivre n'importe quelle loi issue de la famille exponentielle.

Cette dernière comprend plusieurs des distributions élémentaires discrètes et continues. Les lois de Bernoulli, binomiale, Poisson et binomiale négative (ou de Pascal) en font partie. Les distributions gaussienne, bêta, gamma et khi-deux y appartiennent aussi. En l'absence d'effets aléatoires, la procédure GLIMMIX effectue le même travail que la procédure GENMOD : les modèles linéaires généralisés.

Dans le but de faciliter la tâche de sélection de la bonne procédure de SAS/STAT pour l'ajustement d'un modèle linéaire, nous avons inclus le tableau suivant (voir Encadré 1). En se basant sur le type d'effet, la distribution de la variable dépendante et l'indépendance des observations, ce dernier simplifie le choix de l'outil pertinent.

Encadré 1



La procédure GLIMMIX généralise les procédures MIXED et GENMOD de deux façons importantes. Premièrement, la variable dépendante peut avoir une distribution non normale. La procédure MIXED suppose la normalité pour la variable dépendante. Deuxièmement, la procédure GLIMMIX incorpore les effets aléatoires dans le modèle et permet donc de faire de l'inférence (conditionnelle) sur les sujets et (marginale) sur la population. La procédure GENMOD ne permet que l'inférence marginale.

Exemple d'une application

Illustrons l'utilité de cette classe de modèles au moyen d'un exemple provenant du marketing : l'évaluation de l'efficacité de diverses campagnes de promotion. Des données ont été recueillies chez des agents d'assurance utilisant des programmes promotionnels différents auprès de leurs détenteurs de polices (*holders*). Le but de ces campagnes est de susciter l'achat de nouvelles polices d'assurance. L'analyse vise à évaluer quel programme est le plus efficace. Le résultat (achat ou non d'une nouvelle police) a été enregistré pour chaque détenteur.

Il est raisonnable de supposer que la variable **agent** représente un échantillon aléatoire d'une plus grande population d'agents d'assurance (il s'agit donc d'un effet aléatoire) et que la variable **new** (nombre de nouvelles polices ajoutées (soit le nombre de succès) suit une loi binomiale, étant donné l'effet aléatoire **agent**).

Cela constitue un modèle mixte comprenant l'effet fixe **promotions** et l'effet aléatoire **agent**. Étant donné que la variable dépendante **new** suit une distribution binomiale, l'hypothèse de normalité des modèles linéaires mixtes n'est pas respectée. Il faut donc utiliser un modèle linéaire généralisé mixte pour ces données.

La syntaxe de GLIMMIX pour cet exemple serait la suivante (Encadré 2).

Encadré 2

```
proc glimmix
  data=sasuser.policy;
  class promotions agent;
  model new/holders = promotions
    / dist=binomial
    link=logit
    solution ;
  random agent;
run;
```

Notes :

L'instruction **Model** sert à spécifier la variable dépendante et les effets fixes.

La syntaxe 'succès/essais' (**new/holders**) est propre aux modèles avec des données binomiales.

L'instruction **Random** sert à spécifier les effets aléatoires.

L'option 'solution' a pour but de produire une estimation des paramètres des effets fixes.

Dans la sortie générée (Encadré 3), il est important de consulter la section 'Model Information' afin de vérifier que SAS a utilisé la bonne fonction de lien ('link function') et la bonne distribution (binomiale dans notre cas).

Encadré 3

The SAS System		
The GLIMMIX Procedure		
Model Information		
Data Set	SASUSER.POLICY	
Response Variable (Events)	new	
Response Variable (Trials)	holders	
Response Distribution	Binomial	
Link Function	Logit	
Variance Function	Default	
Variance Matrix	Not blocked	
Estimation Technique	Residual PL	
Degrees of Freedom Method	Containment	
Class Level Information		
Class	Levels	Values
promotions	3	A B C
agent	10	1 2 3 4 5 6 7 8 9 10
Number of Observations Read	30	
Number of Observations Used	30	
Number of Events	242	
Number of Trials	622	
Dimensions		
G-side Cov. Parameters	1	
Columns in X	4	
Columns in Z	10	
Subjects (Blocks in U)	1	
Max Obs per Subject	30	

La table 'Solutions for Fixed Effects' (Encadré 4) nous permet de répondre à la question initiale, à savoir si les campagnes de promotion ont un effet différent sur la probabilité d'acheter une nouvelle police. L'estimé pour la promotion A représente la différence entre les promotions A et C sur l'échelle logit. L'estimé pour la promotion B représente la différence entre les promotions B et C sur l'échelle logit. Les seuils significatifs indiquent que les promotions A et B sont différentes de la promotion C en terme de probabilité d'achat d'une nouvelle police. La table 'Type III Tests of Fixed Effects' nous indique que l'effet fixe **promotions** est un facteur significatif.

Encadré 4

Solutions for Fixed Effects						
Effect	promotions	Estimate	Standard Error	DF	t Value	Pr > t
Intercept		-1.4402	0.2265	9	-6.36	0.0001
promotions	A	1.3291	0.2281	18	5.83	<.0001
promotions	B	1.4247	0.2258	18	6.31	<.0001
promotions	C	0
Type III Tests of Fixed Effects						
Effect	Num DF	Den DF	F Value	Pr > F		
promotions	2	18	23.03	<.0001		

Pour conclure, Proc GLIMMIX est une heureuse addition au module SAS/STAT. Cette procédure permet enfin de traiter des modèles linéaires généralisés mixtes avec SAS et donne ainsi l'occasion aux statisticiens d'étendre leur champ d'analyse à de nouvelles applications.

Références :

A Comparison of the Mixed Procedure and the Glimmix Procedure - SUGI 31
 Jennifer Ferrell, University of Louisville, Louisville, KY
<http://www2.sas.com/proceedings/sugi31/189-31.pdf>

Introducing the GLIMMIX Procedure for Generalized Linear Mixed Models - SUGI 30
 Oliver Schabenberger, SAS Institute Inc., Cary, NC
<http://www2.sas.com/proceedings/sugi30/196-30.pdf>

Statistique et vie :

La carte de vœux a un coup de vieux !

La carte de vœux n'est pas en voie de disparition mais elle fait de moins en moins recette. Chaque année les ventes diminuent. De grands organismes comme l'Unicef, qui écoule près de 12 millions de cartes de vœux pendant les Fêtes, ont vu leurs ventes baisser de 300 000 unités par an ces deux dernières années.

Les Français n'écriraient plus à leur famille, leurs amis ou leurs clients à l'occasion du Nouvel An ? Bien au contraire, nous a-t-on dit. Envoyer des vœux n'a jamais été aussi branché mais on ne les envoie plus autant sur papier. Les jeunes se sont mis aux SMS et aux courriels. L'an passé, les abonnés au téléphone cellulaire ont envoyé près de 100 millions de SMS, soit le double de l'année précédente.

Les courriels sont très prisés car on peut trouver sur Internet des idées de vœux originaux et surtout adaptables par chacun. Photos

truquées, dessins drôles, textes arrangés, tout est bon pour envoyer des vœux personnalisés qui sauvent du temps sur l'achat d'une carte, la rédaction d'un texte bien senti et l'envoi par la poste.

Les entreprises qui se veulent à la pointe du modernisme font parfois « plancher » leur personnel pour trouver des idées de vœux originaux en symbiose avec l'entreprise.

La Poste démentait l'année dernière une chute d'activité. « Le volume de cartes de vœux acheminées est stable depuis des années ». Il n'empêche que la Poste a lancé un service: les c@rtes. Il permet de choisir ou de créer sa carte de vœux sur le site <http://www.laposte.net/> et de l'envoyer ensuite par la poste.

[Tiré de l'*Almanach Vermot 2005*.]

**Vous cherchez
de la formation
ou de la consultation**

LES SERVICES CONSEILS
HARDY

- ☞ 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!

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

4715, des Replats, Bureau 260
Québec (QC) G2J 1B8

440, boul. René-Lévesque Ouest
Bureau 1250, Montréal (QC) H2Z 1V7

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
7èmes Journées Francophones "Extraction et Gestion des Connaissances" - EGC 2007	Namur (Belgique)	Institut d'Informatique FUNDP - Les Facultés Universitaires Notre-Dame de la Paix (Note : Deux prix scientifiques seront décernés)	23 au 26 janvier 2007
SADA'07: Colloque international de Statistique Appliquée pour le Développement en Afrique	Cotonou (Bénin)	Université de Pau et des Pays de l'Adour , Université de Abomey-Calavi	26 février au 2 mars 2007
Conférence sur la technologie de l'information (TI) 2007	Ottawa (Canada)	Statistique Canada	4 et 5 avril 2007
SAS GLOBAL FORUM 2007	Orlando, FL (États-Unis)	SAS Global Forum ans SAS Institute	16 au 19 avril 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	St John's (Terre-Neuve)	Société statistique du Canada	9 au 13 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 (États-Unis)	American Statistical Association (ASA)	29 juillet au 2 août 2007
56e Congrès de L'IIS	Lisbonne (Portugal)	Institut international de statistique (IIS)	22 au 29 août 2007 [Note : en anglais]
			2008
36e Congrès annuel de la SSC – 2008 Note : auquel sera jumelé le Congrès 2008 de l'ASSQ	Ottawa (Canada)	Société statistique du Canada ASSQ (Note : l'ASSQ prévoit jumeler son congrès annuel à celui de la SSC ; pour info, communiquer avec les membres du CA)	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 est 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 diverses
Statistique et probabilités pour l'ingénieur Aide-mémoire	Renée Veyseyre	Dunod	Octobre 2006	Cet aide-mémoire rassemble toutes les définitions, lois et formules du calcul des probabilités et de la statistique utiles à l'ingénieur en activité aussi bien qu'à l'étudiant en formation. (476 pages)
Dictionary of Statistics	Graham Upton Ian Cook	Oxford University Press	Août 2006	Un dictionnaire de 490 pages avec plus de 2000 entrées sur tous les aspects de la statistique incluant la programmation, les mathématiques, la recherche opérationnelle et les probabilités. [en anglais]
Assimiler et utiliser les statistiques Résumés de cours et 180 exercices corrigés	Luc Pibouleau	Ellipses	Août 2006	L'ouvrage présente et illustre les méthodes classiques utilisées en calcul des probabilités et statistique. Il est constitué de neuf chapitres consacrés au calcul de probabilités, à l'estimation, à la régression et aux tests statistiques. Chaque chapitre commence par un résumé de cours et se poursuit par vingt exercices corrigés dont le niveau de difficulté croissant est indiqué. (282 pages)
Le dilemme du prisonnier	Nicolas Eber	La Découverte	Juin 2006	L'objet de ce livre est de présenter les avancées les plus récentes des recherches sur le dilemme du prisonnier [suggéré par une expérience menée en 1950 par les mathématiciens Melvin Dresher et Merrill Flood] et ses généralisations (jeu du bien public, jeu de la ressource commune, jeux de confiance). (122 pages)
Méthodes d'enquêtes et sondages - Pratiques européenne et nord-américaine	Pierre Lavallée Louis-Paul Rivest	Dunod	Mai 2006	Ce recueil fait le point sur la pratique des enquêtes et des sondages en Amérique du Nord et en Europe. Il s'adresse aux élèves des écoles d'ingénieurs, aux étudiants de Master, aux professionnels des sondages et à tout utilisateur de données statistiques soucieux d'acquérir une vision large des pratiques d'enquêtes. (432 p.)
Le modèle linéaire par l'exemple - Régressions, analyse de la variance et plans d'expériences illustrés avec R, SAS, et Splus	Jean-Marc Azaïs Jean-Marc Bardet	Dunod	Avril 2006	Cet ouvrage présente des variations sur le thème du modèle linéaire (régression linéaire, analyse de la variance, modèles mixtes et plans d'expériences). Les programmes permettant leurs traitements avec les logiciels R, SAS et Splus sont présentés, tout comme les résultats numériques et graphiques. (326 p.)

REMARQUE : Les « Notes diverses » sont extraites des notices descriptives préparées par les éditeurs.

Pour ceux que cela intéresse, vous trouverez sur le site de l'[AMQ](#) dans la rubrique « Lu pour vous » des présentations d'un choix de livres de mathématique 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. Notez que l'information fournie dans la grille des cours est sujette à changement. Le lecteur est invité à entrer en communication avec la ou le responsable des cours pour corroborer et compléter l'information présentée (frais d'inscription, dates des cours, etc.).

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.gc.ca schardy@schardy.gc.ca
Institut SAS Montréal, Québec, Ottawa	Grand public, entreprises, débutant ou expert	Forage de données (Data Mining), langage macro, programmation, rédaction de rapports, statistique.	Nathaly Renaud (514) 395-8922, poste 4071 www.sas.com/formation sastrain@can.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 Note : 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 la ou le responsable pour plus de détails.

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