



Le mot de la présidente

Par Nathalie Madore

Depuis quelques années déjà, comme vous le savez, nous dirigeons certains de nos efforts à sensibiliser les jeunes aux joies de la statistique et nous tentons de promouvoir notre profession, particulièrement auprès des étudiants. Je suis fière de vous annoncer une nouvelle initiative que nous avons prise en ce sens ce printemps. Lors de la finale provinciale des Expo-Sciences qui s'est tenue à Terrebonne du 10 au 13 avril dernier, une première bourse de l'Association des statisticiennes et statisticiens du Québec a été offerte au projet qui a démontré la meilleure utilisation de la statistique. Les critères qui étaient évalués pour l'attribution de cette bourse étaient :

- Le choix de la méthode (illustration graphique, méthode d'analyse, type de statistique utilisée, etc.) en fonction des objectifs poursuivis;
- Le bien-fondé de l'utilisation de la statistique (que celle-ci ait vraiment un apport, qu'elle soit une valeur ajoutée dans le projet);
- La qualité de l'utilisation de la statistique et des conclusions tirées à partir de son utilisation (bien conclure la bonne chose).

C'est notre secrétaire, Lise Charette, qui s'est rendue sur place le 13 avril pour remettre la bourse de 500 \$ au gagnant. J'espère que ce sera le début d'un long partenariat avec cette organisation qui fait la promotion de la science auprès des jeunes, un objectif qui est tout-à-fait cohérent avec celui de notre association. Je vous invite à consulter le site www.exposciences.qc.ca pour connaître le projet gagnant puisqu'au moment d'écrire ces lignes, la remise des prix n'a pas encore eu lieu.

Vous avez pu constater une autre nouveauté ce printemps : l'adhésion en ligne à votre association préférée. Depuis le début de mon mandat à titre de présidente, je harcelais mes collègues du CA pour trouver une façon de procéder en ligne au renouvellement des adhésions, considérant que notre «vieux» formulaire avait bien fait son temps et n'était plus tellement à la page. C'est notre nouvelle registraire, Véronique Tremblay, qui a eu l'idée de génie cette année d'utiliser le logiciel de sondage en ligne de SOM pour concevoir notre nouveau formulaire de renouvellement. N'hésitez pas à lui faire part de vos commentaires, positifs ou négatifs, sur cette initiative. Au besoin, nous pourrions réajuster certains détails l'an prochain. J'en profite pour remercier chaleureusement l'équipe de SOM qui a fait un excellent travail dans ce dossier.

Sur une note moins joyeuse, j'ai appris il y a quelques semaines le départ de Sylvain Tremblay à titre de collaborateur au Convergence. Je voudrais donc le remercier très sincèrement de sa précieuse contribution

pendant de nombreuses années. Un petit mot plus formel est publié dans cette édition pour souligner son apport à notre journal.

Enfin, je m'en voudrais de ne pas terminer en vous invitant à réserver votre après-midi du 6 juin prochain pour vous joindre à nous lors de notre colloque annuel qui s'annonce vraiment intéressant encore cette année. Le programme reste le même : conférences en après-midi suivies de l'assemblée générale annuelle, d'un cocktail et d'un souper. Cette fois-ci, nous vous accueillerons au Château Laurier, situé en plein cœur de la ville de Québec, sur la Grande-Allée et adjacent à la Place George-V. J'espère avoir le plaisir de vous y rencontrer en grand nombre.

Nathalie Madore, présidente

Les échos du CA

Par Lise Charette

Depuis janvier, votre CA a quelque quatre réunions derrière la cravate et a prévu un souper annuel, notamment avec les membres sortants du CA. Ce sera pour nous l'occasion de faire nos au revoir à Andréanne Moisan-Roy et à Éric Lacroix qui quittent le CA et de souhaiter la bienvenue à Yona Bernardo qui se joint au CA à titre de trésorière!

Le Colloque annuel et l'AGA ont retenu notre attention en ce début d'année. Le choix des sujets qui vous intéressent fait évidemment partie de nos discussions et, cette année, nous vous ferons grâce de modifications aux statuts et règlements, venant de notre part en tout cas!

J'ai eu la chance de représenter l'ASSQ pour la remise du tout nouveau Prix de l'Association des statisticiennes et statisticiens du Québec à la finale provinciale d'Expo-Sciences Hydro-Québec 2014. C'est le jeune Nicolas Roulier du Collège Stanislas de Montréal qui en est le récipiendaire. Je n'ai malheureusement pas pu visiter les stands de ces jeunes talentueux, mais leur fougue et leur passion étaient contagieuses à la cérémonie de remise des prix. Les équipes étaient présentées par région, avec un cri ou une chanson de ralliement qui ont ramené pas mal de souvenirs de ma lointaine jeunesse. Je le dis haut et fort, nous pouvons être fiers de nos jeunes, ils sont beaux, ils sont bons, ils sont capables.

Nous ne pouvons présumer pour l'instant de la portée qu'aura cette tribune pour la science qu'est la statistique. Votre association peut à tout le moins être googlée dans l'univers de la [presse écrite](#)... Mon sentiment est que ce prix est une très belle occasion de rejoindre la communauté scientifique de demain. À suivre...

Lise Charette, secrétaire

Le mot du rédacteur en chef

Par Denis Talbot

Chronique SAS

C'est avec beaucoup de regret que j'ai appris que Sylvain Tremblay ne serait plus disponible dans l'avenir pour rédiger son habituelle Chronique SAS. Notre présidente a d'ailleurs écrit un message au nom de notre association pour le remercier pour sa contribution de longue date. Personnellement, je n'ai jamais connu le Convergence sans la chronique de Sylvain. Ses chroniques étaient toujours bien rédigées, pédagogiques et ô combien instructives. Je tiens à le remercier personnellement pour m'avoir tant appris sur le logiciel SAS.

Son départ m'amène par ailleurs à chercher un nouveau collaborateur régulier au Convergence. Je ne dirais pas que je recherche quelqu'un pour remplacer Sylvain, parce que je ne crois pas qu'il est remplaçable, mais plutôt quelqu'un qui voudrait écrire de façon régulière pour le Convergence une chronique visant des objectifs similaires

à la Chronique SAS. Je suis ouvert aux idées en ce qui a trait au type exact de chronique; il pourrait s'agir d'une nouvelle Chronique SAS, mais également d'une Chronique logiciels par exemple. Évidemment, le nouveau rédacteur n'aurait pas à essayer de répliquer l'écriture de Sylvain et pourrait donner à sa chronique une nouvelle mouture. Si vous êtes intéressé par « l'emploi », vous pouvez communiquer avec moi.

Chronique revue de livre

Tant qu'à chercher un nouveau collaborateur, aussi bien en chercher un deuxième! Tel que je l'avais déjà mentionné, notre association offre la possibilité de rembourser le coût d'achat d'un livre lié à la statistique en échange de l'écriture d'une chronique dans le Convergence sur ledit livre. Si vous êtes intéressé, communiquez avec moi pour obtenir des détails sur la façon de procéder.

Denis Talbot, rédacteur en chef

Merci Sylvain !

Par Nathalie Madore

C'est dans le numéro d'avril 2006 que Sylvain Tremblay a fait son entrée dans la famille des rédacteurs du Convergence. Plus d'une vingtaine de chroniques plus tard, Sylvain nous quitte avec, nous l'espérons, la satisfaction du devoir accompli. Pendant ces huit années de collaboration, il a contribué à parfaire les connaissances de nos membres à l'égard d'un des principaux outils de travail du statisticien : le logiciel SAS. Grâce à lui, GLIMMIX, DATASETS, PLM, GLMSELECT et autres ne sont plus pour nous de simples combinaisons aléatoires de lettres mais bien des procédures réelles qui facilitent l'exploitation et l'analyse des données.

Tous les membres du CA se joignent aux rédacteurs en chef actuel et passés pour te remercier, Sylvain, de ces heures consacrées à tes chroniques et te souhaiter la meilleure des chances dans tes projets futurs.

Nathalie Madore, présidente

L'ASSQ décerne un prix!

Par Lise Charette

Nicolas Roulier est tout jeune, mais il a déjà démontré comment digérer le « Big Data ». Après une médaille d'or dans la catégorie junior et une bourse de 175\$ du ministère des Transports du Québec à l'Expo-Sciences Hydro-Québec régionale, cet élève du Collège Stanislas de Montréal en 2e année du secondaire a remporté le prix de l'Association des statisticiennes et statisticiens du Québec, d'un montant de 500\$, pour sa présentation portant sur l'essence même du profit des compagnies pétrolières selon les fluctuations saisonnières et les périodes de grands déplacements.

Mise en contexte, objectifs, méthodologie, résultats, interprétation, analyses statistiques et graphiques à l'appui, tout était sur les panneaux du stand de ce jeune chercheur (ou statisticien) en herbe. La relève est là, *don't worry, be happy!* Bravo Nicolas!

L'ASSQ tient à féliciter tous les jeunes qui ont participé aux différentes étapes de l'Expo-Sciences Hydro-Québec 2014 ainsi que les enseignants et parents qui les ont supportés.

Lise Charette, secrétaire

Portrait de membre

Par Préambule Communication

Pierre Lavallée

Détenteur d'un baccalauréat en mathématiques combiné à une mineure en informatique, d'une maîtrise en statistique et d'un doctorat en théorie des sondages, monsieur Pierre Lavallée pratique le métier de statisticien depuis 1985. Dans ce portrait, il nous fait non seulement part de son cheminement de carrière, mais il nous explique aussi de quelle façon il a contribué au milieu de la statistique.

Depuis près de 30 ans, M. Lavallée exerce la profession de méthodologiste d'enquête chez Statistique Canada. Au cours de sa carrière, il a travaillé sur divers types d'enquêtes, par exemple des enquêtes économiques, agricoles, sociales, financières, d'indices de prix, par panels ou avec des données administratives. Petit à petit, M. Lavallée a gravi les échelons, passant de MA03, soit méthodologiste junior, à MA07 en devenant directeur adjoint. Après cinq ans à œuvrer dans le domaine des enquêtes sociales, il se concentre maintenant principalement sur les enquêtes auprès des entreprises : registre des entreprises, financement des industries, etc. M. Lavallée fait également beaucoup de consultations méthodologiques et participe activement au sein de divers comités techniques.

L'application de la théorie

M. Lavallée croit que la capacité d'application de la théorie est quelque chose que chaque statisticien devrait être capable de faire : « La statistique n'existe pas par elle-même, elle est un outil pour calculer des choses ou en prédire d'autres. Lorsque l'on possède seulement les chiffres, nous n'avons rien. Si un statisticien n'est pas capable d'appliquer la théorie, il ne pourra pas être apte à produire quelque chose seulement qu'avec ces chiffres ». L'homme se dit d'ailleurs passionné par la théorie des sondages et l'aspect mathématique qui y est lié.

Un lien particulier avec l'Europe

M. Lavallée a souvent eu à collaborer avec des statisticiens qui se trouvent en France et ailleurs en Europe, en ayant par exemple travaillé avec l'École Nationale de la Statistique et de l'Analyse de l'Information (ENSAI) et l'Institut National de la Statistique et des Études Économiques (INSEE). En plus de s'occuper des rapports de pays pour l'Association Internationale des Statisticiens d'Enquête, il donne des cours sur des enquêtes par panels un peu partout en Europe. Mais ce lien avec l'Europe ne s'arrête pas là. M. Lavallée a fait sa thèse de doctorat à l'Université Libre de Bruxelles où il a développé le sondage indirect : « Si je suis quand même assez connu un peu partout à travers le monde, c'est surtout grâce au sondage indirect. Tout ce qui a été fait par la suite, soit après ma thèse, découle de cela ». (N.D.L.R. : plusieurs publications sur le sondage indirect existent, dont [celle-ci](#), à la page 165, qui est accessible gratuitement sur le site Web de Statistique Canada)

La surabondance de données : un phénomène en expansion

Quand on lui demande comment il voit l'avenir dans le domaine de la statistique, Pierre Lavallée répond que l'on passe peu à peu du problème de l'obtention de petits échantillons de données à une surabondance de données. Le problème est qu'il faut non seulement savoir quoi faire avec cette quantité importante de données, mais qu'il faut également conserver un esprit de rigueur dans le traitement statistique de ces données. Si les futurs statisticiens ne font pas attention, cela risque de devenir un problème.

Formule de Thabit pour les nombres aimables

Par Pierre Lavallée

[Lien vers le document](#)

Réfléchir grâce à la statistique

Par Pierre Lavallée

L'autre jour, je lisais l'excellent roman de Paul Cleave, « Nécrologie » (Éditions France Loisirs), lorsque je suis tombé sur ce passage :

«Quentin James a été rattrapé en moins d'une heure. Son 4 x 4 avec les barres antibuffles à l'avant, qui n'avait jamais roulé ailleurs que sur des routes depuis quatre ans qu'il le possédait, a été confisqué. Il a été inculpé pour homicide involontaire et conduite imprudente, mais il aurait dû être inculpé pour meurtre. Je n'ai jamais compris ça. Il avait choisi de rouler saoul. Il avait choisi de le faire chaque foutu jour de sa vie. Ce qui signifie qu'il n'était pas question de destin ni de malchance, mais d'un choix conscient. Et aussi de statistique. C'était mathématique. Ça devait arriver. Envoyez un ivrogne sur les routes chaque jour et il finira par tuer quelqu'un. C'est forcé, de la même manière que si vous lancez encore et encore une pièce en l'air elle finira obligatoirement par atterrir du côté pile.»

Plusieurs personnes croient que la statistique, et surtout les statistiques, ne sont que de la production de chiffres destinés à garnir les tablettes de nos dirigeants. C'est vrai que quelquefois on se pose la question, à savoir à quoi peut bien servir de produire telle statistique ou telle autre, et pourquoi il est important qu'elle soit précise. Ajoutons à cela le calcul de la probabilité de certains événements. Quelle que soit la probabilité calculée, on finit toujours par s'étonner, que l'événement ait eu lieu ou non.

C'est lorsqu'on voit des exemples simples de l'usage des statistiques et des probabilités, ou des anecdotes reliées à ces dernières que l'on se met à réfléchir sur ce que devrait être leur première utilisation : la prise de décision éclairée. Pour ce faire, Pieter Hofstra, logicien au département de mathématiques et de statistique de la Faculté des sciences de l'Université d'Ottawa, a créé le cours « Probabilités et jeux de hasard : Poker 101 ». Il s'agit d'une introduction à la théorie des probabilités. La théorie et les fondements mathématiques des jeux de hasard sont démontrés par leur application au jeu de poker.

« La raison pour laquelle j'ai créé le cours, c'est parce qu'à la base, j'aime le poker et les mathématiques. Je pense que les deux sont reliés. C'est important de bien comprendre les mathématiques pour bien jouer. Avec l'engouement autour du poker, je trouvais que c'était une belle opportunité d'intéresser un public plus large aux mathématiques », a expliqué M. Hofstra au journal *Le Droit* du 4 mars 2014.

Ce cours se veut une façon d'encourager la réflexion selon des principes mathématiques lors de situations vécues quotidiennement. M. Hofstra a donné en exemple des gens qui doivent négocier un salaire ou toute autre situation dans le domaine des affaires. Dans son cours, les étudiants analysent des mains de poker, voient quelle stratégie un joueur aurait dû privilégier et d'autres facettes des jeux de hasard.

Avec le contenu de ce cours, peut-être que Quentin James n'aurait pas conduit en état d'ébriété sachant que ça allait lui arriver un jour; il a fait ce choix sans connaître les probabilités.

Est-ce que ça vous fait réfléchir? Moi si.

Pierre Lavallée, statistique Canada

Encore (et toujours?) la valeur-p!

Par Denis Talbot

Vous vous souvenez de [cet article](#), paru dans *Le Soleil* et *La Presse*, qui essayait de vulgariser les conclusions de [l'article](#) scientifique de Johnson paru dans *Proceedings of the National Academy of Sciences of the United States of America* (PNAS), lequel suggérait de fortement réviser à la baisse le seuil critique utilisé en statistique? Je me souviens que mon *Facebook* a été inondé de partages de cet article. Bon, inondé est une exagération... mais disons que l'article a soulevé l'intérêt de mes amis dans le domaine de la recherche en général et de mes amis statisticiens en particulier. J'aimerais entamer une réflexion et si possible une discussion autour de la valeur-p.

Qu'est-ce que c'est exactement?

J'imagine que les membres de notre association n'ont pas vraiment besoin qu'on leur rappelle ce qu'est une valeur-p, mais juste au cas... Une valeur-p, ce n'est pas la probabilité que l'hypothèse nulle soit vraie ou la probabilité que l'hypothèse alternative soit fausse. En fait, ces deux quantités n'ont aucun sens d'un point de vue fréquentiste puisque les paramètres, sur lesquels portent les hypothèses statistiques, sont considérés comme des quantités fixes, bien qu'inconnues. Ce qui est considéré comme aléatoire en statistique fréquentiste, ce sont les données. La valeur-p est généralement définie comme la probabilité, si l'hypothèse nulle est vraie, d'observer un résultat aussi ou plus extrême que celui observé dans les données. Si la valeur-p est très faible, le résultat qu'on a observé est fortement improbable sous l'hypothèse nulle. La logique veut donc que si on obtient très peu de support en faveur de l'hypothèse nulle, on la décrète comme étant fausse et on accepte l'hypothèse alternative. Dans le cas contraire, on est dans une situation un peu floue. Ce n'est pas clair que si la valeur-p est élevée, c'est parce que l'hypothèse nulle obtient beaucoup de support de la part des données. Une valeur-p élevée peut aussi survenir parce que les données sont tout simplement peu informatives.

Le problème soulevé dans PNAS et la solution proposée

Dans l'article scientifique, que j'ai mis en référence, on soulève un problème bien réel et préoccupant : la non-reproductibilité des résultats d'études scientifiques. Évidemment, ce problème n'a pas une cause unique, mais bien des causes multiples qui passent de la planification des expériences scientifiques à leur analyse statistique. L'article publié dans PNAS semble attribuer la majeure partie du taux de non-reproductibilité à l'utilisation d'un seuil critique trop libéral. Il est proposé de réduire les seuils critiques usuels de 0.05 et de 0.01 à 0.005 et à 0.001 respectivement.

Afin de comprendre l'argumentaire de Johnson, il convient d'effectuer une brève introduction à (ou un bref rappel de) la statistique bayésienne. En statistique bayésienne, on considère grosso-modo les paramètres comme des variables aléatoires et les données comme des quantités fixes. On peut donc fixer une loi *a priori* aux paramètres, c'est-à-dire qu'on attribue à ces paramètres une distribution statistique basée sur les connaissances dont on dispose avant de conduire l'étude. À partir de ces lois *a priori*, on peut déterminer la probabilité *a priori* que l'hypothèse alternative soit vraie et la probabilité *a priori* que l'hypothèse nulle soit vraie ainsi que le rapport de ces deux quantités. En combinant les connaissances apportées par les données aux connaissances *a priori*, on peut déterminer le rapport *a posteriori* de la probabilité que l'hypothèse alternative soit vraie sur la probabilité que l'hypothèse nulle soit vraie.

Les conclusions d'un test d'hypothèses en statistique bayésienne se basent généralement sur le facteur de Bayes qui correspond au facteur d'augmentation du rapport *a posteriori* sur le rapport *a priori* des probabilités que l'hypothèse alternative soit vraie et que l'hypothèse nulle soit vraie. Le facteur de Bayes correspond donc d'une certaine manière au niveau de preuve contenu dans les données en faveur de l'hypothèse alternative. Bien qu'il n'y ait pas de consensus exact quant aux « valeurs critiques » pour le facteur de Bayes, des valeurs supérieures à 50 correspondent à des niveaux de preuves importants en faveur de l'hypothèse alternative et des valeurs supérieures à 150 correspondent à des niveaux de preuve très importants. (Notons qu'on pourrait argumenter que

les lois *a priori* sont attribuées de façon un peu arbitraire. Il est ainsi commun de chercher à diminuer leur impact sur les conclusions. Dans ce cas, des lois *a priori* dites non-informatives peuvent être utilisées.)

L'argumentaire utilisé dans l'article du PNAS est basé sur une correspondance, pour certains tests statistiques simples, entre la valeur-p d'un test fréquentiste et le facteur de Bayes d'un test bayésien. Il se trouve que dans ces cas, un facteur de Bayes de 50 correspond à une valeur-p d'environ 0.005 et un facteur de Bayes de 150 correspond à une valeur-p d'environ 0.001.

Les problèmes avec la solution proposée dans l'article de PNAS

De l'aveu même de l'auteur de l'article paru dans PNAS, la solution implique des coûts supplémentaires pour effectuer des études. En effet, pour obtenir un même niveau de puissance, beaucoup plus de données devraient être récoltées. Il faut également mentionner que, pour des raisons logistiques ou éthiques, il est parfois impossible de récolter davantage de données.

Mon opinion sur le problème

À partir d'ici, j'énonce mon opinion personnelle sur le sujet qui n'implique ni l'ASSQ de façon générale, ni son conseil d'administration en particulier, ni les réviseurs du *Convergence*. Je crois que le problème énoncé dans PNAS est bien réel et mérite une attention particulière. Je suis convaincu que les statisticiens peuvent jouer un rôle afin de lutter contre ce problème. Cependant, je n'adhère pas à la solution proposée dans PNAS pour diverses raisons. D'une part, cette solution n'est pas toujours réaliste pour les raisons énoncées dans la section précédente. D'autre part, cette solution me semble trop rigide et ne tient pas suffisamment compte de la diversité des réalités des différents domaines d'application de la statistique. Voici donc mes quatre premières recommandations :

1. Rapporter les valeurs-p exactes;

2. Rapporter les intervalles de confiance;

3. Favoriser une interprétation des résultats basée sur les intervalles de confiance;

4. Combiner l'interprétation statistique des résultats et leur interprétation pratique.

Ces premières recommandations sont de plus en plus courantes dans la littérature scientifique. L'idée est de donner autant d'information que possible au lecteur pour lui permettre de bien interpréter les résultats. Les intervalles de confiance sont souvent considérés comme étant plus informatifs qu'une simple valeur-p parce qu'ils donnent un intervalle de « valeurs plausibles » pour les paramètres. Il faut également noter que la valeur estimée d'un paramètre peut être « statistiquement différente de 0 », sans être importante d'un point de vue pratique. Afin de ne pas donner de fausses impressions aux lecteurs, l'interprétation statistique des résultats devrait être combinée à une interprétation pratique.

Les quatre prochaines recommandations visent à reconnaître que les études n'ont pas toutes le même contexte et à donner une interprétation plus appropriée à la valeur-p selon le contexte :

5. Reconnaître la nature exploratoire ou confirmatoire des études;

6. Favoriser les tests choisis *a priori* pour les études confirmatoires;

7. Favoriser la correction des valeurs-p pour la multiplicité des tests dans les études confirmatoires;

8. Rapporter avec précision les différentes étapes effectuées dans l'analyse des données .

Dans une étude exploratoire, on cherche généralement à identifier des associations potentielles ou à identifier des facteurs de risque potentiels. On peut se permettre dans ce contexte d'explorer les données, d'effectuer de la sélection de variables ou de modèles basée sur les données ou d'effectuer plusieurs tests statistiques sans nécessairement corriger les seuils des tests. Par contre, il est important d'être conscient qu'en « partant à la

pêche », on augmente la probabilité d'obtenir de faux positifs, c'est-à-dire des tests erronément « statistiquement significatifs ». On doit également être honnête dans la façon de rapporter les résultats pour que les lecteurs sachent que les conclusions obtenues sont préliminaires et auront besoin d'être confirmées ultérieurement.

Dans une étude confirmatoire, on cherche à confirmer des hypothèses précises, élaborées avant même que les données soient récoltées. Autrement dit, on ne choisit pas les hypothèses à vérifier en fonction des données observées. Afin d'obtenir les conclusions les plus solides possible, il convient de choisir les tests à effectuer et les modèles à utiliser sans se baser sur les données observées autant que possible. Si certains choix sont basés sur les données, les lecteurs doivent en être informés. Aussi, afin d'éviter une augmentation du niveau global de faux positifs, il peut convenir dans certaines circonstances d'effectuer une correction des valeurs-p pour la multiplicité des tests (par exemple, une correction de type *Bonferroni*); sans correction, si 20 tests statistiques sont effectués au niveau 5% alors que l'hypothèse nulle est vraie, en moyenne 1 test aura une valeur-p inférieure à 0.05 simplement par le fruit du hasard!

Votre opinion

C'est maintenant à votre tour de vous exprimer! Partagez votre opinion sur la recommandation de l'article de *PNAS* ainsi que sur les miennes, que vous soyez en accord ou non. Vous avez vos propres recommandations? Partagez-les également!

Denis Talbot, rédacteur en chef