



Le mot de la présidente

Par Nathalie Madore

Depuis mon élection à titre de présidente de l'ASSQ, je reviens constamment sur le même sujet : l'implication et la participation des membres à la vie de notre association. Encore en juin dernier, lors de l'assemblée générale, j'ai martelé ce message qui est selon moi un incontournable à la longévité de l'ASSQ.

Mais qui prône l'implication doit aussi faire preuve de reconnaissance à l'égard de ceux qui entendent son message. J'aimerais donc aujourd'hui profiter de la tribune qui m'est offerte pour souligner certaines de ces contributions qui seront peut-être pour vous source d'inspiration.

Le 8 septembre s'est tenu le 14e tournoi de golf de l'ASSQ. Quatorze éditions toutes organisées par un seul et même homme : Claude Ouimet. Chaque année, pas besoin de le lui rappeler ou de quêter son aide. Claude prend les devants, trouve le terrain, le restaurant, envoie les invitations, recueille les inscriptions et les paiements, s'assure du bon déroulement. Et en 14 ans, jamais de plaintes ! Merci Claude pour cette activité mais aussi pour toutes les autres fois où tu as contribué à l'ASSQ en étant membre du CA, rédacteur pour le Convergence, livreur de vin pour les 5 à 7, et quoi d'autre encore.

Les dernières élections, et les précédentes aussi, se sont déroulées sous la direction de Marie-Ève Tremblay qui a généreusement accepté d'agir à titre de présidente d'élections. Oui, je sais, il y a bien longtemps que l'ASSQ n'a pas vécu de « vraies » élections puisque les représentants au CA sont généralement élus par acclamation. Mais ça ne veut pas dire qu'il n'y a rien à faire ! Il faut mettre en place le processus, établir le calendrier, solliciter les candidatures, annoncer les résultats, être prêt, au cas où... Marie-Ève aussi a déjà eu l'occasion de s'impliquer autrement, en particulier à titre de membre du CA. Merci de contribuer à la vivacité de notre association.

Saviez-vous qu'à chaque édition du Convergence, en plus de Denis Talbot le rédacteur en chef, deux personnes assument les fonctions de rédacteurs adjoints et relisent chacun des textes pour en assurer la meilleure qualité ? Depuis 2005, Myrto Mondor et Steve Méthot consacrent de nombreuses heures à réviser et corriger les textes qui leur sont soumis. Mais leur contribution date de bien plus longtemps car Steve agit comme rédacteur adjoint presque depuis les débuts du Convergence soit depuis 1997. Quant à Myrto, elle a occupé le poste de rédactrice en chef en 2003. Merci pour ce travail qui, bien que réalisé dans l'ombre, contribue à l'amélioration de notre principal véhicule de communication.

Est-ce que j'ai fait le tour ? Bien sûr que non. Je n'ai pas parlé des nombreuses personnes qui se sont succédées au conseil d'administration, des précieux collaborateurs qui rédigent fidèlement des articles pour le Convergence (dont Pierre Lavallée et Sylvain Tremblay), du personnel de SOM qui nous offre un support administratif inestimable, de ceux qui ont consacré leur temps à l'organisation des activités sociales et professionnelles de

l'ASSQ par leur support logistique, ou par la tenue d'une conférence, ou de toute autre façon! Hé oui, si notre association peut aujourd'hui se targuer de fêter ses 17 ans, c'est grâce à vous tous.

Vous avez le goût de souligner la contribution d'un collègue dont je n'ai pas parlé ? Pourquoi ne pas envoyer un petit texte pour la prochaine édition. Vous n'avez pas encore eu l'occasion de vous impliquer vous-même activement ? N'ayez crainte, il n'est jamais trop tard. Votre CA vous prépare toutes sortes d'occasions de vous manifester. Ne vous gênez pas, on compte sur vous.

Nathalie Madore, présidente

Les échos du C.A.

Par Lise Charette

J'amorce ici ma 2e chronique avec un peu d'appréhension, faisant déjà la girouette en changeant mon style nature (cf. C.A. de la chronique précédente) pour adopter un style givré (CA de la même chronique). Ce changement de cap n'est pas dû à Isaac qui fait la pluie et le mauvais temps ces jours-ci mais plutôt à un état second de ma personne qui se consacre à des vacances grâce auxquelles j'ai pu vérifier scientifiquement qu'un corps au repos tend à garder son état de repos, loi qui est directement reliée à la loi du moindre effort...

Si vous lisez entre les lignes, vous comprendrez que vos humbles serviteurs du CA étaient au repos cet été, après un blitz pré-colloque. En effet, en bonne gestionnaire – et malgré sa promotion et ses vacances printanières ainsi que celles de notre trésorière, Andréanne – Nathalie nous avait mis à l'horaire des rencontres judicieusement planifiées serrées pour que nos objectifs pour le colloque soient rencontrés pour votre plus grand plaisir, nous l'espérons.

Préparer un colloque implique plusieurs étapes et engagements. L'opération s'est avérée particulièrement intéressante du fait de la participation de plusieurs personnes, à commencer par les conférenciers qui ont répondu positivement. Le lieu que nous avons eu le privilège de réserver, malgré nos contraintes de dates et le fait que nous étions tardifs, disons, était spectaculaire. J'ouvre ici sur ce dernier adjectif car un constat ici s'impose : en cette année olympique – Nathalie en sait quelque chose – votre CA s'est entraîné au sprint plutôt qu'au demi-fond!

Qui dit colloque dit assemblée générale annuelle et la préparation de cette activité a également retenu une partie de nos efforts, étant donné les modifications proposées aux règlements à l'admissibilité des membres et aux statuts. Nathalie a mené ces projets de modifications d'une main de maître et a gardé le cap malgré le délai serré pour que ces propositions se rendent jusqu'aux membres, j'en suis un témoin privilégié.

J'ose ici aborder à nouveau le sujet de l'efficacité des membres du CA lors de nos rencontres. En avril, j'avais mentionné celle de Jean-François à faire plusieurs choses en même temps, lorsqu'il était à la maison. Sachez que Jean-François est également multitâche en réunion, de même qu'Éric et Mhamed, il faut le dire, pour nous dénicher l'information en temps réel sur le Web. Plusieurs sujets abordés peuvent ainsi être réglés sur le champ, ce qui évite les longues chaînes de courriels par la suite. Bravo les gars!

Parmi les sujets qui retiendront notre attention pour les prochains mois, on retrouve notamment l'évolution du site Web, les Jeudis de l'ASSQ à renouveler, la préparation d'une présentation sur la carrière en statistique et l'Année internationale de la statistique en 2013. Ou peut-être d'autres sujets, selon les commentaires et les suggestions des membres? Faites-nous part de vos attentes et de vos besoins! [Du pain et des jeux](#), disaient les Romains? Nous avons du pain sur la planche et nous vous invitons à mettre la main à la pâte!

Lise Charette, secrétaire

Les échos de l'AGA 2012

Par Lise Charette

L'Assemblée générale annuelle 2012 de l'Association a été tenue tout de suite après le Colloque à la salle Grande Course de l'Aquarium du Québec. Trente-sept membres y ont participé, le quorum étant ainsi atteint. Les autres personnes invitées, principalement les conférenciers, ont pu profiter du lieu magnifique qu'est le site de l'Aquarium en attendant le souper sans que la pluie forte tombée plus tôt durant l'après-midi ne joue les trouble-fête.

Faits saillants :

Le compte-rendu de l'assemblée précédente a été lu (merci pour les corrections suggérées).

L'Association est en bonne santé financière (voir bilan 2011), l'encaisse étant de quelque 22 000 \$, ce qui amène le CA à inviter plus que jamais les membres à suggérer des activités, à en organiser ou à proposer des collaborations avec des événements existants. Un formulaire est disponible sur le site de l'Association.

La gratuité du membership est dorénavant offerte aux étudiants d'un programme reconnu par l'ASSQ pour toute la durée de leurs études. Le montant de la contribution des autres catégories de membres demeure inchangé. Le mode de communication privilégié pour rejoindre les étudiants est de passer par le représentant des étudiants et les professeurs des différentes universités.

Le CA est en bonne santé, dix rencontres ont eu lieu en 2011 et déjà cinq réunions avaient eu lieu en 2012 au moment de l'AGA (le chiffre de quatre réunions avait alors été mentionné à l'AGA mais c'était plutôt cinq, foi de la personne qui fait les comptes-rendus!).

Un comité sur l'évolution du site Web et du Convergence a produit de bons résultats, le Convergence étant dorénavant en mode électronique. Le comité est toujours actif et poursuit son mandat en ce qui concerne l'évolution du site Web.

La création d'une bourse de l'ASSQ est abordée et les membres sont invités à réfléchir à la forme que cela pourrait prendre et à faire des suggestions à ce sujet.

Le poste de représentant des étudiants, formalisé dans les statuts, devient vacant et les étudiants ont jusqu'au 31 août de chaque année pour présenter leur candidature.

Bien que nécessaires au maintien d'une certaine rigueur, la proposition de modifications aux statuts et règlements est une opération ardue. Cette partie de la rencontre a malheureusement empiété sur le temps de discussion. Les membres sont donc invités à émettre leurs commentaires et leurs suggestions via les différents médias à leur disposition. Exprimez-vous, l'ASSQ, c'est vous !

À noter qu'un compte-rendu plus complet est disponible. Les commentaires et les suggestions de tous sont les bienvenus.

Lise Charette, Secrétaire

La minute financière

Par Andréanne Moisan-Roy

Depuis plusieurs années, le bilan financier de l'ASSQ affiche des surplus. En 2011, il n'y a toujours pas d'exception à la règle; l'association continue de bien se porter. Quoique le bilan annuel de 2011 ait été présenté

lors de l'assemblée générale annuelle des membres tenue en juin dernier, en voici un résumé pour ceux qui auraient manqué l'événement.

Le montant perçu pour les cotisations et pour les inscriptions à toutes les activités organisées par l'ASSQ en 2011 totalise 8 960 \$. En incluant les dons provenant des membres et les intérêts de placement, les revenus totaux se chiffrent à 9 166,02 \$. Plus précisément, les cotisations comptent pour plus de 70 % de ce montant avec 6 440 \$. Les inscriptions au colloque, au tournoi de golf et aux « Jeudis de l'ASSQ » ont quant à elles généré respectivement 2 185 \$, 210 \$ et 125 \$.

Pour ce qui est des dépenses, 7 860,85 \$ ont été utilisés au cours de l'année. Les frais reliés au colloque s'élèvent à 4 458,12 \$ et ils correspondent à la plus grande dépense de l'association. Des sommes de 605,09 \$ et de 495,05 \$ ont été déployées respectivement pour les « Jeudis de l'ASSQ » et pour le tournoi de golf. Un montant de 506,23 \$ a été utilisé pour le conseil d'administration. Les frais de télécommunications, qui incluent entre autres l'hébergement du site Web, ont engendré des dépenses de 641,22 \$. Les autres dépenses, soient les frais bancaires et les commandites, se chiffrent à 1 155,14 \$.

Ainsi, l'année financière 2011 se termine avec un surplus de 1 305,17 \$. L'encaisse est donc passée de 21 207,15 \$ au 1er janvier 2011 à 22 512,32 \$ au 31 décembre 2011. Les détails de ce bilan sont présentés dans le tableau suivant :

Encaisse au 1^{er} janvier 2011 21 207,15 \$

← Source	Dépenses (\$)	Revenus (\$)	Solde (\$)
Cotisations		6 440,00	6 440,00
Colloque	4 458,12	2 185,00	(2 273,12)
Tournoi de golf	495,05	210,00	(285,05)
Jeudis de l'ASSQ	605,09	125,00	(480,09)
C.A. – Réunions	506,23		(506,23)
Télécommunications	641,22		(641,22)
Autres	1 155,14	206,02	(949,12)
Totaux	7 860,85	9 166,02	1 305,17

Encaisse au 31 décembre 2011 22 512,32 \$

Andréanne Moisan-Roy, trésorière

À venir dans le *Convergence*

Par Denis Talbot

Au cours de la dernière année, le *Convergence* a pris un virage Web. Ce « changement de format » permet un *Convergence* un peu plus dynamique, avec des hyperliens qui permettent facilement d'accéder à d'autres sites Web ou à d'autres articles du *Convergence*. On constate également quelques changements au niveau du contenu, entre autres, des hyperliens intéressants suggérés par les membres sont ajoutés au *Convergence*. Il y a encore plusieurs changements à venir, je vous en présente quelques-uns dans cet article.

Chronique « Revue de livre »

Une chronique revue de livre devrait bientôt faire son apparition dans le *Convergence*. L'idée serait que l'ASSQ offre aux membres la possibilité d'obtenir gratuitement un livre touchant à la statistique. En échange, le membre recevant le livre devrait en effectuer une revue qui servirait de chronique dans le *Convergence*.

Chronique « Bon coups d'un membre »

L'ASSQ souhaite inclure dans le *Convergence* une chronique « Bon coup d'un membre ». Cette chronique permettrait de faire connaître aux lecteurs du *Convergence* les réussites en statistique ou dans d'autres domaines d'un membre de l'ASSQ. Cette chronique pourrait voir le jour dès la prochaine édition du *Convergence*! Il n'en tient qu'à vous de me transmettre un texte à mon adresse courriel denis.talbot@mat.ulaval.ca.

Chronique « Présentation d'un membre »

La chronique « Présentation d'un membre » est un peu similaire à la chronique « Bon coup d'un membre » puisqu'il s'agit dans les deux cas d'une chronique sur un membre de l'ASSQ. Par contre, la chronique « Articles sur les membres » ferait un portrait plus en profondeur d'un membre de l'ASSQ. La chronique prendrait la forme d'une entrevue. Cette chronique serait probablement réalisée par un étudiant en journalisme et pourrait voir le jour dans le courant de la prochaine année.

Si vous avez d'autres idées de chroniques qui pourraient être incluses dans le *Convergence*, n'hésitez pas à m'en faire part par courriel. Surtout, si vous souhaitez écrire des articles pour le *Convergence*, n'hésitez pas à me les soumettre.

Denis Talbot, rédacteur en chef

La chronique SAS

Par Sylvain Tremblay

Imputation multiple

Que faire avec les données manquantes?

La plupart des procédures statistiques de SAS procèdent à l'analyse des cas complets, c'est-à-dire qu'elles excluent les observations qui ont une valeur manquante pour une au moins une des variables utilisées dans l'analyse. Cette approche simple se comprend d'un point de vue calcul informatique, mais résulte en une perte d'information qui peut entraîner une perte de puissance. De plus, elle dernière peut engendrer un biais et les inférences produites pourraient ne pas s'appliquer à la population. Afin d'éviter cela, une stratégie est d'utiliser l'imputation simple. Par exemple, chaque valeur manquante est remplacée par la moyenne de la variable pour les cas complets. Cette approche peut changer la forme des distributions et fausser les relations entre les variables en plus de diminuer artificiellement la variance.

Afin de produire des inférences valides (les estimateurs sont convergents, les valeurs de p sont correctes, etc.), il faut tenir compte du mécanisme de non-réponse. Une technique d'imputation qui est préférable à l'imputation

simple est l'imputation multiple. Elle consiste à remplacer chaque valeur manquante avec un ensemble de valeurs plausibles qui représentent l'incertitude quant à la juste valeur à imputer. Les multiples ensembles de données imputés sont ensuite analysés à l'aide des procédures standard et les résultats produits sont par la suite combinés et analysés afin de produire des inférences valides.

L'imputation multiple avec SAS s'effectue avec les procédures MI (« Multiple Imputation ») et MIANALYZE du module SAS/STAT. Cet article vise à expliquer les bases de ces procédures.

L'imputation multiple dans le cadre d'une analyse de régression

Afin de présenter les procédures MI et MIANALYZE, nous allons utiliser un exemple dans lequel on désire faire une analyse de régression linéaire avec des données continues pour lesquelles il y a des valeurs manquantes.

Le code suivant (voir encadré 1) génère le fichier de données Fitness1 (n=31 observations) que nous allons utiliser dans notre exemple. Il contient trois variables : Oxygen, RunTime et RunPulse. Certaines valeurs ont été changées pour des valeurs manquantes afin que le fichier de données résultant contienne un patron de non-réponse arbitraire.

Encadré 1

```
data Fitness1;
  input Oxygen RunTime RunPulse @@;
  datalines;
44.609 11.37 178 45.313 10.07 185
54.297 8.65 156 59.571 . .
49.874 9.22 . 44.811 11.63 176
. 11.95 176 . 10.85 .
39.442 13.08 174 60.055 8.63 170
50.541 . . 37.388 14.03 186
44.754 11.12 176 47.273 . .
51.855 10.33 166 49.156 8.95 180
40.836 10.95 168 46.672 10.00 .
46.774 10.25 . 50.388 10.08 168
39.407 12.63 174 46.080 11.17 156
45.441 9.63 164 . 8.92 .
45.118 11.08 . 39.203 12.88 168
45.790 10.47 186 50.545 9.93 148
48.673 9.40 186 47.920 11.50 170
47.467 10.50 170
;
run;
```

L'inférence par imputation multiple comporte trois phases distinctes :

1. Les données manquantes sont remplacées m fois par des données imputées pour générer m ensembles de données complets;Chacun des m ensembles de données complets est analysé à l'aide des procédures standards;
2. Les résultats de ces analyses sont combinés pour l'inférence.

Étape 1

Avec SAS, cette première étape est réalisée avec la procédure MI. Elle crée de multiples ensembles de données imputées pour des données incomplètes multivariées. Elle utilise des méthodes qui incorporent une variabilité appropriée à travers les m imputations. Le choix de la méthode d'imputation repose sur les patrons de non-réponses dans les données et le type de la variable imputée.

Pour les données qui ont un motif de non-réponse monotone (fréquent dans les études longitudinales), l'imputation d'une variable continue peut se faire avec les méthodes suivantes :

- Régression,
- « predictive mean matching »,
- score de propension.

Pour les données qui ont un patron de non-réponse arbitraire, on peut utiliser les méthodes suivantes :

- Monte-Carlo par chaînes de Markov (MCMC),
- Spécification complètement conditionnelle (« Fully Conditional Specification »)

Le code SAS suivant (voir encadré 2) effectue les imputations six fois ($m=6$) par la méthode de Monte-Carlo par chaînes de Markov pour le jeu de données Fitness1.

Encadré 2

```
proc mi data=Fitness1 seed=1234 nimpute=6 out=outmi;
  mcmc chain=multiple displayinit initial=em(itprint);
  var Oxygen RunTime RunPulse;
run;
```

Ce code va produire deux types de sortie :

1. un fichier de données nommé « outmi » de $n=186$ observations ($m \times n = 6 \times 31$). Ce fichier contient 6 copies du fichier initial où les données manquantes ont toutes été remplacées par des valeurs imputées
2. une sortie imprimée

Dans la sortie imprimée, on retrouve entre autres, un sommaire des patrons de non-réponses, où les X représentent une valeur non manquante et les points une valeur manquante (voir encadré 3) ainsi que la table des paramètres estimés suite à l'imputation multiple (voir encadré 4).

Encadré 3

Group	Oxygen	Run Time	Run Pulse	Freq	Percent	-----Group Means-----		
						Oxygen	RunTime	RunPulse
1	X	X	X	21	67.74	46.353810	10.809524	171.666667
2	X	X	.	4	12.90	47.109500	10.137500	.
3	X	.	.	3	9.68	52.461667	.	.
4	.	X	X	1	3.23	.	11.950000	176.000000
5	.	X	.	2	6.45	.	9.885000	.

Encadré 4

Parameter Estimates							
Variable	Mean	Std Error	95% Confidence Limits		DF	Minimum	Maximum
Oxygen	47.083034	0.967941	45.0997	49.0664	27.815	46.947704	47.190294
RunTime	10.525174	0.271169	9.9696	11.0807	27.897	10.484734	10.560146
RunPulse	171.788248	1.911910	167.8235	175.7530	22.031	170.832457	172.403380

L'algorithme EM (« Expectation-Maximization ») est utilisé pour trouver les estimés par le maximum de vraisemblance pour les données manquantes. On peut également utiliser l'algorithme EM pour trouver un mode postérieur, les estimations des paramètres qui maximisent la densité postérieure des données observées. Le mode postérieur offre une bonne valeur de départ pour la méthode MCMC.

Étape 2

Les m=6 ensembles de données complets sont analysés un à la suite de l'autre (en raison de l'instruction `BY _imputation_` dans le code de l'encadré 5) à l'aide de la procédure REG pour notre exemple de régression linéaire. Le code suivant va générer le fichier de données « outreg » qui va contenir les coefficients de régression et les matrices de covariance correspondantes calculées à partir des ensembles de données imputées.

Encadré 5

```
proc reg data=outmi outest=outreg covout noprint;
  model Oxygen = RunTime RunPulse;
  by _Imputation_;
run;
```

Le fichier de données « outreg » ainsi produit contient les informations suivantes : (voir encadré 6)

Encadré 6

	Imputation	_MODEL_	_TYPE_	_NAME_	_DEPVAR_	_RMSE_	Intercept	RunTime	RunPulse
1	1	MODEL1	PARMS		Oxygen	2.5662891	91.2516479	-2.936	-0.079
2	1	MODEL1	COV	Intercept	Oxygen	2.5662891	76.6020744	-0.163	-0.436
3	1	MODEL1	COV	RunTime	Oxygen	2.5662891	-0.1633269	0.1	-0.005
4	1	MODEL1	COV	RunPulse	Oxygen	2.5662891	-0.4362483	-0.005	0.0029
5	2	MODEL1	PARMS		Oxygen	2.5349188	90.5333856	-3.084	-0.064
6	2	MODEL1	COV	Intercept	Oxygen	2.5349188	77.5017603	-0.451	-0.423
7	2	MODEL1	COV	RunTime	Oxygen	2.5349188	-0.4507362	0.1055	-0.004
8	2	MODEL1	COV	RunPulse	Oxygen	2.5349188	-0.422561	-0.004	0.0027
9	3	MODEL1	PARMS		Oxygen	2.5717514	89.9683537	-3.053	-0.061
10	3	MODEL1	COV	Intercept	Oxygen	2.5717514	70.5244728	-0.68	-0.366
11	3	MODEL1	COV	RunTime	Oxygen	2.5717514	-0.6797772	0.1098	-0.003
12	3	MODEL1	COV	RunPulse	Oxygen	2.5717514	-0.3662019	-0.003	0.0023
13	4	MODEL1	PARMS		Oxygen	2.4741949	85.9442417	-3.014	-0.041
14	4	MODEL1	COV	Intercept	Oxygen	2.4741949	54.0455434	-0.375	-0.29
15	4	MODEL1	COV	RunTime	Oxygen	2.4741949	-0.3745242	0.096	-0.004
16	4	MODEL1	COV	RunPulse	Oxygen	2.4741949	-0.2897646	-0.004	0.0019
17	5	MODEL1	PARMS		Oxygen	2.6434842	86.7830236	-2.708	-0.066
18	5	MODEL1	COV	Intercept	Oxygen	2.6434842	63.4677317	-0.449	-0.339
19	5	MODEL1	COV	RunTime	Oxygen	2.6434842	-0.4491762	0.0984	-0.003
20	5	MODEL1	COV	RunPulse	Oxygen	2.6434842	-0.3394113	-0.003	0.0022
21	6	MODEL1	PARMS		Oxygen	2.7739438	96.997889	-3.293	-0.089
22	6	MODEL1	COV	Intercept	Oxygen	2.7739438	86.7060823	-0.304	-0.487
23	6	MODEL1	COV	RunTime	Oxygen	2.7739438	-0.3041574	0.1344	-0.007
24	6	MODEL1	COV	RunPulse	Oxygen	2.7739438	-0.4873522	-0.007	0.0033

Étape 3

Les estimations obtenues à l'étape précédente sont ensuite combinées pour produire des inférences statistiques valides pour notre modèle de régression. Pour ce faire, nous utilisons la procédure MIANALYZE (voir encadré 7).

Encadré 7

```
proc mianalyze data=outreg edf=28;
  modeleffects Intercept RunTime RunPulse;
run;
```

L'option edf sert à spécifier le nombre de degrés de liberté. Dans un modèle de régression avec 3 paramètres et 31 observations, le nombre de degrés de liberté est 28. Dans les résultats produits, on retrouve la table des paramètres estimés avec chacun leur erreur-type et leur intervalle de confiance (voir encadré 8)

Encadré 8

Parameter Estimates					
Parameter	Estimate	Std Error	95% Confidence Limits		DF
Intercept	90.246424	9.460502	70.36132	110.1315	17.883
RunTime	-3.014608	0.387784	-3.84460	-2.1846	14.316
RunPulse	-0.066596	0.053276	-0.17704	0.0438	22.167

À la fin des trois étapes du processus d'imputation multiple, nous avons donc

- évité de perdre de l'information précieuse en raison des valeurs manquantes et
- généré des inférences valides.

Conclusion

L'imputation multiple ne cherche pas à estimer chaque valeur manquante par l'entremise de valeurs simulées, mais plutôt à représenter un échantillon aléatoire de ce que pourrait être les vraies valeurs. Ce processus aboutit à des inférences statistiques valides qui reflètent bien l'incertitude due aux des valeurs manquantes. C'est une technique d'imputation préférable à l'imputation simple.

La procédure MI propose de nombreuses méthodes d'imputation qui tiennent compte des patrons de non-réponse dans les données et du type des variables imputées. Après avoir analysé les multiples fichiers de données imputées avec les procédures standards de SAS/STAT, la procédure MIANALYZE vous permet de facilement combiner les résultats de vos analyses afin de produire des inférences valides. Évidemment, nous n'avons pas abordé tous les aspects théoriques reliés à l'utilisation de cette méthode (patrons de non-réponse pour lesquels l'imputation multiple est valide, choix des variables à inclure dans le modèle d'imputation, choix de m, etc) mais ce survol donne un aperçu de la façon de la mettre en pratique avec SAS.

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

Références

[Yuan, Y. C., Multiple Imputation for Missing Data: Concepts and New Development, SAS Institute Inc.](#)

[Documentation des procédures MI et MIANALYZE](#)

Six livres de polygraphie

Par Pierre Lavallée

Johannes Trithemius (1462-1516), Abu ibn Ishaq Al-Sabbah Al-Kindi (801-873)

Aujourd'hui, la théorie mathématique est un aspect essentiel de la cryptographie. Cependant, à une époque reculée, c'était les cryptogrammes par substitution, dans lesquelles les lettres d'un message étaient remplacées par d'autres lettres, qui étaient souvent utilisés. Par exemple, CHAT devient DIBU si nous remplaçons chaque lettre du mot CHAT par la lettre suivante de l'alphabet. Bien sûr, des cryptogrammes aussi élémentaires deviennent très faciles à décoder après la découverte de l'analyse des fréquences par l'érudite arabe Al-Kindi au IXe siècle, par exemple. Cette méthode analyse les lettres qui apparaissent le plus fréquemment dans une langue donnée, comme ETAION SHRDLU en anglais, et utilise ces informations pour résoudre les codes de substitution. Il est également possible d'utiliser des statistiques plus complexes, comme le nombre de paires de lettres. Par exemple, en français, la lettre Q est presque toujours associée à la lettre U.

Le premier ouvrage imprimé sur la cryptographie, *Polygraphiae Libri Sex* (Six livres de polygraphie), fut écrit par l'abbé allemand Johannes Trithemius et publié après sa mort, en 1518. Il contient des centaines de colonnes de mots latins, disposés sous forme de deux colonnes par page. Chaque mot correspond à une lettre de l'alphabet. La première page se présente ainsi :

a : Deus	a : clemens
b : Creator	b : clementissimus
c : Conditor	c : pius

Pour coder un message, on utilise un mot pour représenter une lettre. Trithemius a construit les tables de telle sorte que le passage codé ressemble à une véritable prière. Par exemple, si les deux premières lettres d'un message étaient CA, la prière commençait par les deux mots Conditor clemens (Créateur miséricordieux). Les autres livres présentent des méthodes cryptographiques plus sophistiquées, ainsi que des tables, destinées à masquer les informations de façon plus créative.



Trithemius

Un autre écrit célèbre de Trithemius, *Steganographia* (Stéganographie), écrit en 1499 et publié en 1606, fut mis à l'index par l'église catholique, car considéré comme un ouvrage sur la magie noire, alors qu'il ne s'agissait en réalité que d'un autre livre sur l'art de crypter les messages.

[Tiré de Pickover, C.A. (2010), *Le beau Livre des Maths, De Pythagore à la 57^e dimension*, Dunod, Paris.]

Let's Predict the Next "Voice of ..." Finalist

Lien suggéré par Simon Olivier Fournier

[Lien vers la page](#)

Why Maximum Likelihood is Better Than Multiple Imputation

Lien suggéré par Simon Olivier Fournier

[Lien vers la page](#)

The first SAS PROCs

Lien suggéré par Simon Olivier Fournier

[Lien vers la page](#)