



ATELIER DU CIQSS

Initiation à l'analyse des données d'enquêtes à l'aide de SAS : des tableaux de fréquences à la régression logistique

Lieu : CIQSS, 3535 chemin Queen-Mary, Suite 420, Montréal
Dates : 6 et 7 février 2007
Heures : 8h30 à 12h00 et 13h30 à 16h00

DESCRIPTION GÉNÉRALE

Le séminaire d'une durée de deux jours présente l'analyse des données d'enquêtes à l'aide du progiciel SAS, principalement par le biais de l'analyse du tableau de fréquences (tableau de contingence), de l'analyse de corrélation ainsi que de la régression multiple (linéaire et logistique).

La formation cherchera à limiter les explications de nature mathématique. Les considérations mathématiques limitées à l'essentiel, la formation se veut pratique, axée principalement sur la compréhension et l'interprétation des données d'enquête produites notamment à l'aide du progiciel SAS. Les procédures SAS suivantes seront mises à contribution : FREQ, UNIVARIATE, CORR, REG, LOGISTIC, GPLOT. L'atelier est articulé à l'exploitation des données de *l'Enquête sur la santé des collectivités canadiennes (ESCC)* réalisée par Statistique Canada. Son recours permet aux participants de se familiariser avec la manipulation, la préparation et l'analyse statistique de données catégorielles. Il s'agit non seulement de produire les résultats statistiques à l'aide de SAS mais il est aussi question d'interpréter correctement les épreuves informatisées. Tout au long de la formation, il sera donc question de mettre en relief :

- Les conditions d'utilisation des procédures;
- Les précautions qui s'imposent lors de leur utilisation;
- La lecture et l'interprétation des épreuves informatisées;
- La programmation informatique (syntaxe) nécessaire pour produire les épreuves informatisées.

Les présentations formelles se donnent en français mais les questions peuvent être posées en français ou en anglais.

PROFIL DES PARTICIPANTS

Cette formation s'adresse plus particulièrement aux chercheurs d'organismes ou de ministères qui souhaitent rafraîchir, actualiser ou compléter leurs connaissances en matière d'analyse quantitative de données d'enquêtes. Elle s'adresse également aux étudiants gradués qui souhaitent mettre en pratiques les notions théoriques acquises dans le cadre des cours de statistiques en vue de la réalisation d'un mémoire ou d'une thèse basé sur l'exploitation de données d'enquêtes. **Une connaissance de base du logiciel SAS est exigée.**

PLAN DE LA FORMATION

PREMIÈRE JOURNÉE – AVANT-MIDI - 8H30 À 10H00

1. ANALYSE DES TABLEAUX DE FRÉQUENCES

Le tableau de fréquences (tableau de contingence) est sans contredit parmi les approches analytiques les plus souvent utilisées en recherche sociale. Or son exploitation efficace renvoie à des règles précises de confection et de lecture. Accompagnant presque invariablement les tableaux de contingence, le chi-carré constitue vraisemblablement la mesure statistique la plus utilisée en recherche sociale mais trop souvent, à tort. En effet, son exploitation efficace renvoie notamment à la nature et à la taille de l'échantillon et au niveau de mesure des variables. Un regard particulier sera posé sur les notions de signification statistique et de rapport de cotes (*odds ratio*).

1.1 UN BREF RAPPEL DE QUELQUES NOTIONS

- Niveaux de mesure d'une variable (types de variables)
- Les opérations mathématiques permises selon le niveau de mesure

1.2 LE TABLEAU DE CONTINGENCE

- Règles de confection du tableau de contingence
- Règles de calcul et de lecture des pourcentages

1.3 L'ANALYSE DU TABLEAU DE CONTINGENCE

- Test d'indépendance du chi-carré et ses conditions d'utilisation
- Rapport de cotes (*Odds ratio*)
- Risque relatif (*Relative risk*)
- Regroupement des catégories

PREMIÈRE JOURNÉE – AVANT-MIDI - 10H30 À 12H00

2. L'ANALYSE DE CORRÉLATION

L'analyse de corrélation n'appartient pas à proprement parler au champ de l'analyse des données catégorielles; elle renvoie pour l'essentielle à l'analyse des données métriques. Toutefois, l'analyse de corrélation constitue un chemin logique fréquemment emprunté qui se situe au carrefour de l'analyse de l'indépendance et de la régression. D'ailleurs, l'analyse de corrélation demeure inextricable de l'analyse de régression. En effet, la recherche ne s'interroge pas seulement sur le type de la relation entre les phénomènes mais aussi sur la force, l'intensité de cette relation. *L'analyse de régression* vise à répondre à la première de ces préoccupations alors que *l'analyse de corrélation* répond à la seconde. Une attention particulière sera accordée à la notion de variance.

2.1 LECTURE DU TAUX DE CORRÉLATION

- Direction de la relation
- Intensité mathématique de la relation
- Intensité historique de la relation
- Corrélation tétracorique et polycorique

2.2 CORRÉLATION ET VARIANCE

- Le coefficient de détermination

PREMIÈRE JOURNÉE – APRÈS-MIDI - 13H30 À 16H00

3. LA RÉGRESSION LINÉAIRE (SIMPLE ET MULTIPLE)

La régression linéaire est l'une des techniques statistiques les plus utiles et l'une de celles qu'on emploie de plus en plus couramment dans le cas d'une variable dépendante continue. De plus, parce qu'on peut l'étendre au-delà des données bivariées en l'appliquant à une situation multivariée, la régression se révèle un outil très utile de la recherche sociale. L'analyse de régression multiple permet de construire une équation explicative d'un phénomène donné. On identifie alors les variables indépendantes les plus significatives, ce qui permet de «prédire» les comportements non mesurés directement.

3.1 NORMALITÉ ET RÉGRESSION

- Vérification des postulats de normalité, de colinéarité et d'homoscédasticité
- Que faire puisqu'elle n'est pas normale?

3.2 RÉGRESSION LINÉAIRE SIMPLE ET RÉGRESSION MULTIPLE

- Choix des variables et production du modèle
- Interprétation des coefficients des paramètres du modèle
- Mesures d'ajustement du modèle
- Pondération et énoncé *weight*
- Plan de sondage complexe (l'effet de plan) : effet sur les estimés et leur variance

3.3 CE QUI DISTINGUE LA RÉGRESSION LOGISTIQUE DE LA RÉGRESSION LINÉAIRE

- À quel moment recourir à la régression logistique?
- Pourquoi recourir à la régression logistique?

DEUXIÈME JOURNÉE – AVANT-MIDI – 8H30 À 12H00

4. LA RÉGRESSION LOGISTIQUE BINOMIALE

La régression logistique est un outil qui permet de mettre en relation des variables explicatives à une variable réponse dichotomique, c'est-à-dire qui ne peut prendre qu'un nombre limité de valeurs, le cas classique étant celui d'une variable réponse (dépendante) binaire. Cette situation est fréquente dans divers champs d'application, particulièrement dans les sciences sociales.

4.1 LE CONTEXTE D'UTILISATION DE LA RÉGRESSION LOGISTIQUE

- Exploration? Prédiction? Explication?
- Comment construire un modèle de régression logistique?

4.2 LE CAS D'UNE VARIABLE RÉPONSE (DÉPENDANTE) BINAIRE

- Confection du modèle
- Les variables explicatives qualitatives
- Les variables explicatives continues
- Quelques éléments de paramétrisation (énoncé *class (param=ref ref='level')*; énoncé *contrast*)
- L'ajustement du modèle
- L'intérêt des coefficients *b* (ou *beta*)
- La valeur discriminante du modèle

DEUXIÈME JOURNÉE – APRÈS-MIDI – 13H30 À 14H00

L'après-midi de la deuxième journée de formation sera consacré aux travaux pratiques à l'aide de SAS. Des activités d'apprentissage spécifiques seront offertes aux participants qui pourront alors confronter leur compréhension des techniques statistiques présentées durant le séminaire-atelier. Le formateur recevra les questions et apportera les précisions qui s'imposent.

FORMATEUR

Sociologue de formation, Gilles Valiquette est actif en recherche depuis plus de 30 ans dans divers domaines tels le marketing, l'évaluation de programmes, la formation et la planification urbaine. Il a notamment enseigné les techniques d'analyse des données pendant plus de 15 ans au Collège Mérici dans le cadre du programme Techniques de recherche, d'enquête et de sondage. À titre de consultant, M. Valiquette a œuvré auprès de divers organisme et ministères dont le MSSQ, le MEQ, le MAPAQ et le MRCIQ.

MODALITÉS D'INSCRIPTION

Les frais de participation sont de 40 \$ pour les étudiants et stagiaires postdoctoraux, de 80 \$ pour les professeurs et chercheurs des universités membres du CIQSS, de 150 \$ pour les professeurs et chercheurs des autres universités et de 250 \$ pour toute autre catégorie. Les frais devront être payés sous forme de mandat ou de chèque libellé au nom du « CIQSS – Université de Montréal ».

Les candidats seront sélectionnés sur la base de la pertinence de la formation par rapport à leur parcours d'étude, de recherche ou d'enseignement. Les candidats retenus recevront une confirmation par courriel. Tous les dossiers doivent être soumis par notre site: www.ciqss.umontreal.ca. La capacité d'accueil est limitée à 15 places.

Amélie Gagnon

Centre interuniversitaire québécois de statistiques sociales
Université de Montréal
C.P. 6128, Succursale Centre-ville
Montréal, Québec H3C 3J7

Téléphone : (514) 343-6111 poste 4942
Télécopieur : (514) 343-2328
Courriel : a.gagnon@umontreal.ca