

## Analyse et traitement des données manquantes

**Lieu de la formation :** CIQSS, 3535 chemin Queen-Mary, bureau 420, Montréal

**Dates :** 18 et 19 juin 2012

*Cette session de formation bénéficie de l'appui financier des universités membres du CIQSS<sup>1</sup>.*

### Objectif

Ce cours avancé en statistique sociale a pour but de permettre aux participants d'évaluer l'impact des données incomplètes sur les conclusions d'une analyse. Les méthodes classiques de traitement des données manquantes (utilisation des cas complets, imputation par la moyenne) ne permettent généralement pas de corriger les biais causés par les données manquantes. Les méthodes récentes présentent l'avantage d'évaluer la sensibilité aux données manquantes d'un résultat.

### Contenu

À la suite de ce cours, les participants seront capables :

- 1) De maîtriser les implications des traitements classiques de données manquantes.
- 2) De maîtriser la distinction entre les schémas (*pattern*) et les mécanismes (MAR, MCAR et MNAR) de données manquantes.
- 3) De produire des analyses permettant de poser différentes hypothèses sur le mécanisme des données manquantes.
- 4) De faire une analyse par mélange de schémas, par modèle de sélection ou par imputation multiple.

### Formateur

Stéphane Paquin, Doctorant, Département de sociologie de l'Université de Montréal.

### Déroulement

Ce cours utilise un matériel pédagogique développé dans le but de faciliter l'auto apprentissage. Des fichiers de données, syntaxes et exercices seront mis en ligne avant la tenue du cours. Le matériel pédagogique de cours (capsules vidéos, syntaxe, fichiers des données d'exercices) restera disponible aux participants après la formation.

Le cours se donnera en français de **9h00 à 17h00**, mais on pourra répondre aux questions formulées en anglais. Les séances offertes dans les locaux du CIQSS alterneront présentation théorique et réalisation d'exercices pratiques.

### Plan de cours

Jour 1 AM

- Implication du traitement classique des données manquantes
- Description des données manquantes à l'aide d'un indicateur factice (binaire)
- Distinction entre le schéma et le mécanisme des données manquantes
- L'analyse de sensibilité

---

<sup>1</sup> Université de Montréal, INRS-UCS, Université McGill, Université Concordia, Université Laval, Université du Québec, Université de Sherbrooke.

#### Jour 1 PM

- Exemples appliqués
  - Exemple de description des données manquantes
  - Exemple d'hypothèses sur le mécanisme
  - Rappel du fonctionnement de SAS (si nécessaire)
- Séance de laboratoire

#### Jour 2 AM

- L'analyse par mélange de schémas
- Un modèle de sélection
- L'imputation multiple

#### Jour 2 PM

- Exemples appliqués
  - Exemple d'une analyse par mélange de schémas
  - Exemple d'une analyse par modèle de sélection
  - Exemple d'une analyse par imputation multiple
- Séance de laboratoire

### **Références bibliographiques (les lectures d'introduction sont en gras)**

Carpenter, James R, Kenward, Michael G, & White, Ian R. (2007). Sensitivity analysis after multiple imputation under missing at random: a weighting approach. *Statistical methods in medical research*, 16(3), 259-75. doi: 10.1177/0962280206075303.

Carpenter, James R., & Kenward, Michael G. (2006). A comparison of multiple imputation and inverse probability weighting for analyses with missing data. *J. Roy. Statist. Soc. Ser. A*, 169, 571-584.

Enders, C. K. (2001). TEACHER ' S CORNER A Primer on Maximum Likelihood Algorithms Available for Use With Missing Data, *Structural Equation Modeling*, 8(1), 128-141.

Enders, C. K. (2011). Missing Not at Random Models for Latent Growth Curve Analyses. *Psychological Methods*, 16, 1-16. doi: 10.1037/a0022640

**Graham, J. W. (2009). Missing data analysis: making it work in the real world. *Annual review of psychology*, 60, 549-76. doi: 10.1146/annurev.psych.58.110405.085530.**

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention science*, 8(3), 206-13. doi: 10.1007/s11121-007-0070-9.

Hedeker, D., & Gibbons, R. D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods*, 2, 64-78. doi: 10.1037//1082-989x.2.1.64

Molenberghs, G., Verbeke, G., & Kenward, M. (2009). Sensitivity analysis for incomplete data. In G. Fitzmaurice, M. Davidian, G. Verbeke & G. Molenberghs (Eds.), *Longitudinal Data analysis* (pp. 501-551). Boca Raton: Chapman and Hall/CRC Press.

**Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147-177.**

Sterne, J. a C., White, I. R, Carlin, J. B, Spratt, M., Royston, P., Kenward, M. G, et al. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *Bmj*, 338(jun29 1), b2393-b2393. doi: 10.1136/bmj.b2393.

Yang, X., & Shoptaw, S. (2005). Assessing missing data assumptions in longitudinal studies: an example using a smoking cessation trial. *Drug and Alcohol Dependence*, 77, 213-225. doi: 10.1016/j.drugalcdep.2004.08.018

### **Profil des participants**

Ce cours s'adresse aux étudiants gradués, aux chercheurs du milieu universitaire (professeurs, agents de recherche, chercheurs postdoctoraux) et gouvernemental. Ce cours nécessite une connaissance préalable de la statistique descriptive et inférentielle et des analyses multivariées (régression linéaire et logistique) ainsi que du logiciel SAS.

On acceptera un maximum de quinze participants et ceux-ci seront sélectionnés en fonction de la pertinence du cours pour leur programme d'étude ou leurs activités professionnelles.

### **Inscription**

La période d'inscription s'étendra du **19 mars au 15 avril 2012**. L'annonce des résultats de la sélection des participants sera faite au cours de la semaine du **16 avril 2012**. Pour toute autre information et pour vous inscrire, vous êtes invités à consulter notre site Internet : [www.ciqss.umontreal.ca](http://www.ciqss.umontreal.ca).

### **Information**

Luc St-Pierre

[l.st-pierre@umontreal.ca](mailto:l.st-pierre@umontreal.ca)