



Traitement des données manquantes à l'aide de SAS

Lundi 21 et mardi 22 mars 2011, de 9h à 16h

Lieu de la formation

Laboratoire du CIQSS, 3535 Queen-Mary, bureau 420, Montréal.

Objectifs

- Maîtriser les principes théoriques du traitement des données manquantes.
- Comprendre la distinction entre les schémas (*pattern*) (l'attrition par exemple) et les mécanismes de données manquantes (mieux connus comme : MAR, MCAR et MNAR).
- Appliquer une approche des analyses en termes de sensibilité des résultats.

Responsable

Stéphane Paquin, candidat au doctorat en sociologie, Université de Montréal, Groupe de recherche sur l'inadaptation psychosociale chez l'enfant, Hôpital Ste-Justine.

Description générale

Les données manquantes constituent un problème récurrent lors de l'analyse de données. De nombreuses études ont démontré que l'analyse des cas complets peut entraîner un biais dans l'estimation d'un paramètre lorsque les données manquantes ne le sont pas de manière aléatoire.

Prérequis

Une connaissance de base du logiciel SAS. Connaissance des concepts de moyenne et variance ainsi que de la régression.

Plan de cours

Jour 1 : AM (9h – 12h)

- Exemples et définition d'une donnée manquante
- Traitement classique et leurs implications (exemples à partir de simulations)
- Description des données manquantes à l'aide d'un indicateur binaire (0 =observée; 1 =manquante)
- Distinction entre le schéma et le mécanisme

Jour 1 : PM (13h – 16h)

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

Jour 2 : AM (9h – 12h)

- Analyse par mélange de schémas
 - Imputation multiple
- Modèle de sélection
 - Pondération
- Principe de l'analyse de sensibilité

Jour 2 : PM (13h – 16h)

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

Références bibliographiques

(les deux premières références en gras sont des lectures d'introduction)

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.

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

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, 8(1), 128-141.

Fielding, S., Maclennan, G., Cook, J. a, & Ramsay, C. R. (2008). A review of RCTs in four medical journals to assess the use of imputation to overcome missing data in quality of life outcomes. *Trials*, 9, 51. doi : 10.1186/1745-6215-9-51.

Gadbury, G. L., Coffey, C. S., & Allison, D. B. (2003). Modern statistical methods for handling missing repeated measurements in obesity trial data: beyond LOCF. *Obesity reviews*, 4, 175-84. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12916818>.

Graham, J. W., & Donaldson, S. I. (1993). Evaluating Interventions With Differential Attrition : The Importance of Nonresponse Mechanisms and Use of Follow-Up Data. *Journal of Applied Psychology*, 78(1), 119-128.

Graham, J. W., Hofer, S. M., & Piccinin, A. M. (1994). Analysis with missing data in drug prevention research. *NIDA research monograph*, 142(October 1995), 13-63. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9243532>.

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 : the official journal of the Society for Prevention Research*, 8(3), 206-13. doi : 10.1007/s11121-007-0070-9.

Harel, O., & Zhou, X.-hua. (2007). Multiple imputation : Review of theory , implementation and software. *Aging*, (January), 3057-3077. doi : 10.1002/sim.

Hogan, J. W., Roy, J., & Korkontzelou, C. (2004). Handling drop-out in longitudinal studies. *Statistics in medicine*, 23(9), 1455-97. doi : 10.1002/sim.1728.

Lee, K. J., & Carlin, John B. (2010). Multiple imputation for missing data : fully conditional specification versus multivariate normal imputation. *American journal of epidemiology*, 171(5), 624-32. doi : 10.1093/aje/kwp425.

Shaffer, M. L., & Chinchilli, V. M. (2007). Including multiple imputation in a sensitivity analysis for clinical trials with treatment failures. *Contemporary clinical trials*, 28(2), 130-7. doi : 10.1016/j.cct.2006.06.006.

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.

Stuart, E. a, Azur, M., Frangakis, C., & Leaf, P. (2009). Multiple imputation with large data sets : a case study of the Children's Mental Health Initiative. *American journal of epidemiology*, 169(9), 1133-9. doi : 10.1093/aje/kwp026.