

LES PROCÉDURES SAS D'ANALYSE DE DONNÉES D'ENQUÊTE

Olivier Sautory ¹

¹ *Département des méthodes statistiques, INSEE*

Dans les procédures standard d'analyse statistique proposées par SAS (**MEANS**, **FREQ**, **REG**, **LOGISTIC**), l'inférence statistique est fondée sur l'hypothèse que les données sont issues d'une population infinie par sondage aléatoire simple. Ces procédures peuvent donc ne pas être adaptées à l'analyse de données recueillies par des enquêtes par sondage probabiliste.

Les procédures **SURVEYMEANS**, **SURVEYFREQ**, **SURVEYREG**, **SURVEYLOGISTIC**, qui seront présentées dans l'atelier, permettent de mettre en oeuvre les méthodes classiques d'analyse en tenant compte du plan d'échantillonnage mis en oeuvre : sondage stratifié, à plusieurs degrés, à probabilités inégales, avec des méthodes d'estimation de variance fondées sur des développements de Taylor.

Plan de l'atelier

1. La procédure **SURVEYMEANS** permet de calculer des estimateurs de totaux, de moyennes, de proportions, de ratios, ainsi que les estimateurs de variance associés ; elle permet aussi de réaliser des estimations par domaine.
2. La procédure **SURVEYFREQ** permet d'estimer des effectifs de modalités de variables catégorielles, et des proportions, ainsi que les estimateurs de variance associés, dans des tableaux simples ou à plusieurs entrées ; elle réalise également des tests d'ajustement à une loi et des tests d'indépendance entre variables.
3. La procédure **SURVEYREG** permet de mettre en oeuvre des méthodes de régression linéaire (avec la possibilité d'introduire des variables explicatives catégorielles en utilisant la même syntaxe que dans **GLM**) : estimation des coefficients de la régression, de leur matrice de variance-covariance, test d'hypothèse sur les effets, sur les paramètres, etc.
4. La procédure **SURVEYLOGISTIC** permet de mettre en oeuvre des méthodes de régression logistique (avec la possibilité d'introduire des variables explicatives catégorielles), avec plusieurs fonctions de lien possibles : estimation des coefficients de la régression par la méthode du maximum de vraisemblance, de leurs variances, calcul des odds- ratios, etc.

Des exemples d'utilisation de ces procédures seront présentés en atelier.

Bibliographie

- [1] Binder, D. A. (1983). On the Variances of Asymptotically Normal Estimators from Complex Surveys, *International Statistical Review*, 51, pp. 279-292.
- [2] Fuller, W. A. (1975). Regression Analysis for Sample Survey, *Sankhya*, 37, Series C, pp. 117-132.
- [3] Roberts, G., and Rao, J. N. K., and Kumar, S. (1987). Logistic Regression Analysis of Sample Survey Data, *Biometrika*, 74, pp. 1-12.
- [4] Skinner, C. J., and Holt, D., and Smith, T. M. F. (1989). *Analysis of Complex Surveys*, New York, Wiley.
- [5] Woodruff, R. S. (1971). A Simple Method for Approximating the Variance of a Complicated Estimate, *Journal of the American Statistical Association*, 66, pp. 411-414.