Lundi 14 mars 2005 : 17h15-17h40 *(en parallèle avec la session 4)*

Régression sur données d'enquêtes par sondage : deux nouvelles procédures dans le logiciel SAS.

Josiane LE GUENNEC
CEPE

Régression linéaire, analyse de la variance et régression logistique se pratiquent, avec SAS, au moyen des procédures REG, GLM, LOGISTIC. Appropriées à la recherche de relations linéaires dans une population, ces procédures ont l'inconvénient de ne pas estimer la précision de façon correcte lorsque les données analysées proviennent d'un échantillon aléatoire. La possibilité de pondérer les observations, notamment par leur poids de sondage, permet bien d'obtenir le bon estimateur d'un coefficient de régression, mais le calcul de variance ne tient aucun compte du plan de sondage, et estime donc imparfaitement le caractère significatif ou non d'une variable du modèle.

Les nouvelles versions du logiciel ont comblé cette lacune. La procédure SURVEYREG, introduite dès la version 8, réalise la régression linéaire, l'analyse de la variance et de la covariance sur des données d'échantillon. La procédure SURVEYLOGISTIC, apparue dans la version 9, effectue la régression logistique sur échantillon aléatoire. Dans ces deux nouvelles procédures, la variance est estimée en tenant compte du plan de sondage, ce qui permet une appréciation plus précise de l'apport d'une variable exogène dans un modèle linéaire.

L'exposé pourra s'articuler en deux parties :

- présentation des méthodes de calcul mises en œuvre par le logiciel
- mise en évidence des différences que l'on peut attendre de l'usage de SURVEYREG ou de SURVEYLOGISTIC, par rapport aux procédures REG, GLM ou LOGISTIC appliquées au même échantillon, à partir des données d'application utilisées pour tester ces nouvelles procédures.