

Module 3. Modèles linéaires généralisés : principes et applications dans le logiciel R

- Date: 21 au 25 Mai 2018
- Heure : 9h-12h
- Lieu : Salle de conférence du LABEF
- Formateurs: Prof. Romain Glèlè Kakaï, Wilfried Bonou (PhD Student), Bruno Lokonon (PhD-student)

En matière de modélisation, le modèle linéaire général (*ex.*, régression linéaire, ANOVA, ANCOVA, etc.) s'est longtemps imposé dans de nombreuses situations courantes. Mais, ce modèle repose sur une hypothèse forte: **le terme d'erreur suit une loi normale et de même variance**. Mais nombreux sont également ces phénomènes (des sciences biologiques, physiques, sociales, etc.) où cette hypothèse de départ est généralement loin (voire très loin) d'être vérifiée. Le décompte d'événements rares en est un bon exemple. En effet, pour cet exemple, considérer la variable à expliquer (et son erreur) comme suivant une loi normale suppose une distribution symétrique autour de la moyenne, alors que ce n'est très probablement pas le cas, puisque la majorité des données collectées pourront, par exemple, avoir plus fréquemment pour valeurs zéro ou un. De plus, au niveau du modèle linéaire général, l'on part du postulat que les variables prédictives ont un effet linéaire sur la variable mesurée (effet qui se traduit par les coefficients de régression), or certains de ces effets ne sont peut-être pas linéaires en réalité. Par exemple, le nombre d'œufs pondus par une poule ne change peut-être pas linéairement avec son âge. Afin de permettre l'analyse de telles données, plusieurs solutions s'offrent à nous. La plus répandue consiste à trouver une transformation mathématique de la variable à expliquer pour la rendre normale. Ces transformations ne sont pas toutes efficaces, et leur effet normalisant est parfois difficile à quantifier. Par ailleurs, il reste évidemment préférable d'utiliser les données d'origine plutôt que leurs valeurs transformées ne serait-ce que pour rendre l'interprétation des résultats plus aisée. C'est donc pour ces multiples raisons que les modèles linéaires généralisés ont été développés par Nelder et Wedderburn (1972) pour pallier les insuffisances du modèle linéaire général. Ces modèles ont connu des développements récents dont nombreux sont déjà disponibles ou programmables sous le logiciel libre et gratuit R pour une analyse efficace des données.

Sous la direction du **Prof. Romain GLELE KAKAI** (Professeur Titulaire de biométrie et foresterie, Directeur du LABEF), ce module des formations modulaires du LABEF vous introduira dans l'univers des modèles linéaires généralisés. Au cours du module, vous:

- apprendrez à identifier les facteurs en présence ainsi que la distribution de la variable réponse ;
- serez introduit aux principes des modèles linéaires généralisés ;
- écrirez des scripts R relatifs aux modèles linéaires généralisés de la famille de Poisson et Binomiale;
- interprétez et rapporterez les résultats d'une analyse de données utilisant les modèles linéaires généralisés.