
Régression logistique et fraction attribuable moyenne

Alioscha Massein^{*1,2}

¹Maison des Sciences de l'Homme Lyon Saint-Etienne (MSH Lyon St-Etienne) – École Normale Supérieure - Lyon, Université Lumière - Lyon 2, Sciences Po Lyon - Institut d'études politiques de Lyon, Université Jean Moulin - Lyon 3, Université Claude Bernard Lyon 1, Université Jean Monnet - Saint-Etienne, Centre National de la Recherche Scientifique – France

²TGIR PROGEDO – École des Hautes Études en Sciences Sociales, Centre National de la Recherche Scientifique – France

Résumé

De façon classique, le sociologue se pose la question d'expliquer un fait social à partir de différents facteurs. La méthode la plus utilisée, la **régression logistique**, permet d'identifier, toutes choses égales par ailleurs dans le modèle, **l'effet propre à chacun de ces facteurs**. Cette analyse permet d'obtenir une hiérarchie des facteurs quant à leur capacité à expliquer le phénomène étudié (Selz and Deauvieau, 2011). **Cependant**, outre, ce premier résultat, une autre question fondamentale se pose : si l'on a bien une idée de la hiérarchie des effets, les effectifs concernés restent inconnus. En effet, lorsque l'on travaille par exemple sur les inégalités, l'un des objectifs est aussi de s'interroger sur l'effet démographique, c'est-à-dire sur le volume engendré pour chaque facteur toutes choses égales par ailleurs dans le modèle.

Pour répondre à cette question, les épidémiologistes utilisent une méthode spécifique : **la fraction attribuable moyenne (ou average attributable fraction)**. Par exemple, en santé publique, il s'agit de choisir pour une campagne de prévention, un ou plusieurs leviers qui permettent de réduire significativement la prévalence d'une maladie (Ancelle, 2017). Ce levier ne sera pas forcément l'effet le plus important relevé par la régression logistique.

Une **innovation** serait d'utiliser ce type de méthode en sciences humaines et sociales. Cet outil statistique permet de répartir la contribution de chacun des facteurs de "risque" sur l'ensemble des individus touchés par le phénomène (Eide, 1995; Ferguson et al., 2018). La multiplicité des facteurs n'empêche pas le calcul d'une proportion unique pour chacun d'entre eux : en effet, on peut calculer précisément l'impact d'un facteur parmi tous les autres sur un fait social au sein d'une population. C'est en ce sens une nouveauté pour l'analyse d'un phénomène : cette manière de produire une analyse donne une lecture originale par rapport à la régression logistique, permettant de donner pour chaque facteur explicatif la proportion d'individus engendrée.

L'usage de cette méthode est présenté au travers d'un exemple tiré d'un jeu de données de l'enquête nationale sur les pratiques physiques et sportives des françaises et des français 2020 à l'aide du logiciel R. Plus précisément, les packages `averisk` et `graphPAF` (Ferguson, 2017; O'Connell and Ferguson, 2022) sont utilisés pour calculer la **fraction attribuable moyenne** de chaque facteur et son intervalle de confiance.

Finalement, un tableau complet de **présentation** d'une analyse, toute chose égale par ailleurs dans le modèle, est proposé

*Intervenant

Mots-Clés: PROGEDO, méthodologie, statistique, quantitatif, régression, PAF, AAF, average attributable fraction