
LA COMMUNICATION AUTOUR DE L'INCERTITUDE DES STATISTIQUES PUBLIQUES

Ronan LE SAOUT

Crest-Ensay

ronan.le-saout@ensai.fr

Mots-clés : incertitude, qualité, statistiques publiques

Résumé

En tant que statisticien-économiste, il est souvent surprenant de constater que la précision des statistiques ne soit pas publiée par les instituts de statistique publique. Les raisons invoquées lors de discussions informelles sont multiples, le manque de moyens, les difficultés théoriques, l'absence de demande sociale, ou la difficulté de la pédagogie... Elles illustrent une absence de connaissance mais aussi une trop forte focalisation sur la précision au détriment d'une vision plus large de l'incertitude. Deux notions d'aléas statistiques interagissent également, un aléa lié à l'observation d'un échantillon (si l'ensemble de la population est observée, il n'y a plus d'incertitude), un aléa associé à un modèle statistique (les observations demeurent aléatoires même sur l'ensemble de la population). L'objet de cet article est dans un premier temps d'illustrer, à partir des statistiques diffusées par l'INSEE, la typologie de Manski (2015) définissant l'incertitude selon trois composantes, permanente, transitoire et conceptuelle. Cette notion d'incertitude ne se limite ainsi pas à la précision et se rapproche de la notion de qualité. Ce caractère pluriel fait que les choix en termes de communication sont nécessairement multiples. Ils vont dépendre des produits statistiques (issues d'enquête, de sources administratives ou de sources multiples), de l'approche retenue (descriptive ou d'étude économique), et du public visé. Une évolution historique vers une plus grande transparence est mise en avant. Dans un second temps, les conséquences de l'émergence des données massives seront discutées. Trois évolutions seront mises en avant : 1) la réappropriation de la notion de qualité des données par les chercheurs en sciences sociales, qui de simples utilisateurs (par exemple de bases de données d'enquêtes) deviennent également collecteurs des données (à travers les sources de données issues d'internet) ; 2) l'appréhension de la notion d'incertitude lorsque la population est observée exhaustivement; 3) le rôle de la data visualisation dans la communication auprès du grand public.