



Séminaire scientifique « Contribution des CRB aux Sciences et Recherches Participatives » 30 septembre 2026

Biais d'observation dans les données participatives.

A. Benard,¹

⁽¹⁾ Biostatistique et processus spatiaux, F84914 Avignon, France

Impliquer des citoyens et citoyennes dans la production de données scientifiques répond à un double objectif : démultiplier les données disponibles et créer un lien entre recherche et société. Ces projets prennent des formes très variées, de l'étude de communautés microbiennes à partir d'échantillons envoyés volontairement par des particuliers à la surveillance de maladies via des signalements citoyens de vecteurs, en passant par l'identification d'espèces à partir de photographies prises par le grand public et suivi de la phénologie ou de la répartition d'espèces via des observations naturalistes... Or l'inférence statistique repose sur une hypothèse centrale : celle d'un échantillon représentatif, tiré indépendamment du phénomène que l'on cherche à mesurer. C'est cette hypothèse qui permet de généraliser d'une observation à une conclusion.

Faire reposer la collecte de données sur l'initiative individuelle de contributeurs non coordonnés interroge directement cette hypothèse, et ce de plusieurs manières : erreurs d'identification ou de mesure, hétérogénéité des protocoles d'un contributeur à l'autre, répartition spatiale non uniforme des observateurs, ou encore prise de décision en amont de ce qui mérite d'être signalé. Un échantillon peut ainsi devenir, silencieusement, un tirage non représentatif de la population d'étude. Dans cette présentation, nous nous concentrerons sur le cas le plus extrême : celui de l'absence totale de protocole d'échantillonnage, propre aux données d'observation entièrement opportunistes (Pl@ntNet, Visionature...). Je présenterai le cadre mathématique permettant de catégoriser les sources de biais, quelques exemples concrets de quantification de ces biais, ainsi que l'impact sur les conclusions d'une étude lorsqu'ils ne sont pas pris en compte. Je donnerai également un panorama des méthodes statistiques disponibles pour tenter d'y répondre, dont certaines sont actuellement à l'étude par l'unité Biosp.