

# Erreur écologique

L'erreur écologique, [efn\_note]terme sans relation avec l'écologie comme discipline scientifique[/efn\_note] désigne le processus par lequel on énonce des conclusions fausses sur des comportements individuels à partir de données observées à un échelon plus agrégé. Cette expression « erreur écologique » (ecological fallacy) fut inventée en 1958 par Hanan C. Selvin dans un article consacré à Emile Durkheim .

L'erreur de raisonnement consiste dans l'interprétation de résultats statistiques obtenus à partir de données agrégées en exprimant des conclusions à l'échelon individuel qui laissent penser que les relations entre des variables observées au niveau des groupes s'appliquent aussi à chaque individu (Robinson, 1950). La difficulté fondamentale avec de telles inférences est que beaucoup d'interactions possibles au niveau individuel peuvent générer des observations variables au niveau agrégé. L'erreur en question est qualifiée d'« écologique », parce que l'unité d'analyse n'est pas une personne individuelle mais porte sur un agrégat. L'erreur écologique participe ainsi de la confusion fréquente entre les notions de corrélation et de causalité. Par exemple, à partir des résultats électoraux d'une circonscription et de sa composition démographique, on peut être tenté de tirer des conclusions sur le comportement électoral des différents groupes sociaux qui la compose. L'erreur écologique est l'inverse de l'erreur atomiste qui, elle, provient de l'interprétation de données individuelles étendue au niveau de l'ensemble du groupe, en ne tenant pas compte du milieu dans lequel évoluent les individus.

William S. Robinson fut un des premiers à attirer l'attention sur le problème de l'erreur écologique. Dans un texte fondateur de 1950, il montrait que les corrélations entre deux caractéristiques mesurées de façon binaire sur des individus ou par des proportions sur des agrégats ne donnaient pas des résultats identiques. Il avait calculé la relation entre le taux d'analphabétisme et la proportion de la population née hors des États-Unis pour chacun des 48 États (+ district of Columbia) à partir du recensement de 1930. Il montra que ces deux variables étaient associées à une corrélation négative de -0,53 (ecological correlation) en d'autres termes, plus la proportion d'immigrés dans un Etat est élevée, plus l'analphabétisme moyen est faible. Cependant, lorsqu'on faisait ces calculs au niveau des individus (individual correlation), et non plus sur des pourcentages, le coefficient de corrélation passait à +0,12 - les immigrants étaient en moyenne plus analphabètes que les citoyens autochtones. Cette inversion de valeur du coefficient de corrélation au niveau des populations des Etats était due au fait que les immigrants avaient tendance à s'installer dans des États les plus prospères où la population autochtone était plus instruite. Il montrait ainsi que les corrélations entre deux caractéristiques mesurées de façon binaire sur des individus ou, au contraire selon des pourcentages sur des zones géographiques ne donnaient pas des résultats équivalents. Il attirait ainsi l'attention contre toute déduction abusive au niveau des individus.

En géographie, la prise en compte de la dimension spatiale du phénomène de corrélation est à la fois une source considérable d'enrichissement de l'analyse, mais aussi un danger pouvant conduire à des erreurs d'interprétation du fait que l'on travaille sur des unités spatiales constituées d'agrégats d'individus. Avec ces valeurs agrégées il faut alors éviter d'affecter à l'ensemble des individus d'une zone les caractéristiques du comportement moyen observé sur la zone d'étude. Au niveau agrégé on enregistre des moyennes qui masquent des variations intra, par exemple, des pourcentages régionaux masquent les variations départementales, qui elles-mêmes masquent des relations portant sur des sous-ensembles différents (ex : la co-existence du chômage et de l'attractivité migratoire dans les régions du sud de la France) (Pumain, Saint Julien, 1997). Dans un article paru en 1984, S. Openshaw fut un des premiers à attirer l'attention sur les risques d'erreurs d'interprétation des données du recensement. Au Royaume-Uni, les données de recensement étaient alors disponibles que sous une forme agrégée et pour des zones géographiques arbitraires. Les unités spatiales utilisées pour présenter les données du recensement (districts, secteurs de recensement, quartiers, unités de gouvernement local) ne présentaient aucun caractère de significativité géographique. Déplorant qu'il n'y ait aucun moyen d'être en mesure de déterminer si une donnée spatialisée particulière allait produire des résultats proches des valeurs individuelles. S.Openshaw suggérait des pistes. Notamment d'examiner plus étroitement les effets de l'agrégation des données sur la précision des estimations des paramètres.

Dans les recherches contemporaines, la question de l'erreur écologique soulève plusieurs enjeux théoriques importants dans les sciences sociales en général. La première porte sur le développement des analyses multiniveaux, et sur le rôle des initiatives individuelles. Un des problèmes porte sur le passage du niveau macro au niveau micro, et plus généralement sur les découpages de l'ensemble observé en sous-groupes pertinents. L'expression « paradoxe de Simpson », [efn\_note] décrit par E.Simpson en 1951[/efn\_note] est utilisée pour décrire un biais d'échantillonnage dans lequel une tendance observée sur plusieurs groupes peut s'inverser lorsque les découpages sont réalisés différemment, ce qui conduit à des

conclusions opposées. C'est ce que fait fréquemment Emmanuel Todd en associant et dissociant plusieurs catégories sociales et électorales à différents niveaux d'échelle pour démontrer que les systèmes familiaux sont localement invariants dans la longue durée. Il s'appuie sur une approche réductrice du structuralisme partant de l'idée que les faits sociaux ont une existence propre indépendamment des individus qui les vivent, qu'ils sont dotés d'une grande stabilité dans le temps et qu'ils fonctionnent à toutes les échelles. Pour appuyer ses démonstrations, le même auteur produit des [cartes choroplèthes](#) suggérant des corrélations entre deux phénomènes sans justification dans les discrétisations des légendes. Au-delà de ce type de biais méthodologique, de réels problèmes scientifiques demeurent posés par les inférences en géographie. Citons notamment le biais statistique induit par l'utilisation d'unités surfaciques. Souvent mise en avant, cette question est connue en géographie sous le nom de MAUP (Modifiable Areal Unit Problem, ou problème d'unité spatiale modifiable) (Josselin, 2015). Ce problème repose sur le fait que la manière d'agréger les données sous la forme d'unités spatiales a un impact significatif sur le résultat. Pour autant, si ce biais ne doit pas être ignoré, sa prise en compte présente un intérêt heuristique. La variation des résultats en fonction du découpage spatial adopté reflète le caractère multiscalaire du phénomène étudié, et représente en tant que tel un apport de connaissance sur ce phénomène (L. Sanders).

B.E.

voir aussi: [statistique spatiale](#), [variable quantitative](#), [analyse spatiale](#)

Documents joints

[Réflexions sur la causalité en sciences sociales](#)

## Bibliographie

Bibliographie :

- Courgeau. D. Réflexions sur la causalité en sciences sociales. In: Recherches et Prévisions, n°60, 2000. Chômage et famille.
- Durkheim E., Le suicide, PUF, Paris, 1930 (première édition 1896).
- Josselin.D, Genre-Grandpierre.C, Aryal.J, 2015, Impact du changement d'échelle sur l'étude des causes des feux de forêts du sud-est de la France, SAGEO, 2015
- Openshaw, S, 1984, Ecological fallacies and the analysis of areal census data; Environment and Planning A, 1984, volume 16
- Pumain.D, Saint-Julien.T , 1997, Analyse spatiale : les localisations, Paris, Armand Colin,

-Robinson W. (1950). Ecological correlations and the behaviour of individuals. American Sociological Review, vol. 15, <https://academic.oup.com/ije/article/38/2/337/658252>

-Selvin.H.C, 1958, Durkheim's Suicide and Problems of Empirical Research, The American Journal of Sociology, Vol. 63, No. 6, <https://www.journals.uchicago.edu/doi/10.1086/222356>

HYPERGEEO