

用平均值断定个人：生态谬误（Ecological Fallacy）

文章网址

<https://www.gklibrarykor.com/zh-hans/article/10651/>

PDF版式预览 - 仅含节选

假设某家地方报纸刊登了一篇报道，称“收入越高的地区，某个政党的支持率就越高”。许多人读完后会自然而然地想：“原来有钱人喜欢那个政党。”然而，如果真的逐一调查每个人，有时却会得到截然相反的结果。把群体的数字原封不动地套到个人身上时产生的这种错觉，就是“生态谬误（Ecological Fallacy）”。简单来说，生态谬误是指把以地区、国家、学校等群体为单位收集的数据所得出的结论，推断为同样适用于属于该群体的每个人的错误。这里的“生态学”并非指生物学，而是一个统计术语，指的是围绕个人的群体层面的数据。本文将依次探讨生态谬误这一名称的由来、群体数据与个人数据出现分歧的原理、现实中经常遇到的实例，以及避开这一陷阱的方法。

群体的数字为何无法代替个人 “生态谬误”这一名称的由来 生态谬误是在一位社会学家重新计算美国人口普查数据之后才广为人知的。（1）罗宾逊1950年的论文美国社会学家威廉·罗宾逊（William S. Robinson）利用1930年人口普查数据，逐州比较了外国出生人口的比例与文盲率。以州为单位来看，移民越多的州，文盲率反而显得越低。然而以个人为单位重新计算后，却得出了方向相反的结果：移民比本土出生者成为文盲的可能性略高。州层面的相关系数为负0.53，而个人层面的相关系数为正0.12。2011年有一项重新审视指出，罗宾逊在州层面计算中使用的部分数据有误，正确值应为负0