

# 生態学的データ利用における誤謬の問題

——ロビンソンの生態学的誤謬問題を中心として——

森 幸 雄

- |                        |                 |
|------------------------|-----------------|
| 1. はじめに                | 4. 生態学的誤謬をめぐる議論 |
| 2. 個人相関と生態学的相関         | 5. 生態学的データ利用の問題 |
| 3. ロビンソンによる生態学的誤謬問題の提起 |                 |

## 1. はじめに

社会学的な実証研究において、地域的に集計された統計データを研究の重要な手がかりとして利用するものは少なくない。デュルケームの『自殺論』や、非行問題についてのショー (Shaw, C. R.) の『都市地域における少年非行 (Juvenile Delinquency in Urban Areas)』(1942) などでは、統計データの利用は研究上不可欠であった。また、研究対象として特定の地域を選択したり、また特定の地域での研究結果を一般化したりする際にそうした統計データを用いることも少なくない。

しかしながら、地域的に集計された統計データを利用する研究において、データを恣意的に利用したり、データの扱いに慣れていなかったといった研究者個人の問題のためからではなく、研究方法自体のもつ問題のために事実を誤認してしまうことがあると指摘したのが、ロビンソン (Robinson, W. S.) である。ただし、これは統計データを利用した研究一般に対する疑義をしめしたのではない。彼は地域的に集計された統計データを利用する際に生じるこのような誤りを「生態学的誤謬 (ecological fallacy)」と呼んだが、これは、社会学ばかりではなく、統計的データを用いることの多い研究分野、とりわけ、投票行為の研究をおこなっていた政治学でも多くの議論を巻き起こした。

本稿では、ロビンソンがとりあげた「生態学的誤謬」の問題、およびこれをめぐる議論を中心にして、生態学的データを利用した研究の問題点をみていきたい。

## 2. 個人相関と生態学的相関

まず、本稿での用語や概念の説明を行ないたい。本稿でとりあげるデータは地域的な区分で集計されたデータであり、集計単位はアグレゲーション (aggregation) となる。アグレゲーションは社会学や統計学における集団 (group) とは異なる。社会的に集団とよべるような成員間の共同性を欠いているのは無論のこと、統計学的用法での集団とよべるような共通する属性をも欠いているためである。つまり、アグレゲーションはなんらかの基準により区分された地域ごとに集計されたデータ群であり、その地域は、近隣や統計区、あるいは市町村、県、地方、国、さらに大きな単位までとることができる。

なお、こうしたアグレゲーションと人間生態学でのコミュニティの概念との関係についてふれておきたい。どちらの概念も地域的に区画されたものであるが、自然地域として形成された相対的に同質的な地域とされるコミュニティとアグレゲーションとは一致するとは限らない。近隣や統計区のようにかなり狭い地域を単位とするアグレゲーションのときはこれをコミュニティとみなすことができるが、かなり広い範囲を単位とするアグレゲーションのときはこれをコミュニティとすることはできないからである。

本稿でとりあげた議論ではアグレゲーションを形成する統計学的単位 (unit) は個人である。そうした個人の属性についてのデータを「個人データ」とよんでいきたい。これに対して、黒人の比率や文盲の比率といった、アグレゲーションを単位とし、比率であらわされるような属性についてのデータを「生態学的データ」とよんでいきたい。そして、二つの変数間の関係を示す際に相関関係を利用し、個人データを用いた場合が「個人相関」、生態学的データを用いた場合が「生態学的相関」となるのである。

この関係を表1のような $2 \times 2$ のクロス集計表を用いて確認してみると、次のようになる。個人データは  $(AB)$   $(A\beta)$   $(\alpha B)$   $(\alpha\beta)$  といった同時分布の

表1 模式化した $2 \times 2$ のクロス表

	B	$\beta$	合計
A	$(AB)$	$(A\beta)$	$(A)$
$\alpha$	$(\alpha B)$	$(\alpha\beta)$	$(\alpha)$
合計	$(B)$	$(\beta)$	N

数字を利用し、生態学的データは(A) ( $\alpha$ ) (B) ( $\beta$ ) といった周辺分布の数字を利用したものである。この同時分布間の相関が個人相関となり、周辺分布の構成比間の相関が生態学的相関となるのである。

### 3. ロビンソンによる問題提起

ロビンソンは、1950年の論文「生態学的相関と個人の行動 (Ecological Correlation and the Behavior of Individuals)」<sup>(1)</sup>において、生態学的誤謬を例をあげて示し、誤謬の起きる理由を数学的に説明している。ここでは、この論文を中心に、ロビンソンがとりあげた生態学的誤謬についてやや詳しくみていきたい。

#### (1) 生態学的誤謬の生じる実例

ロビンソンは1930年センサスの結果などをもとにして、生態学的相関を個人相関の代用として利用する際に生じる誤謬を、実際に両者の数値を計算したうえで示している。

彼がまず示したのは、生態学的相関を用いたために相関が誇張される例である。10歳以上人口を対象にして、黒人と文盲との生態学的相関を、合衆国を9区分した九地区分類の散布図で見たものが図1である。この場合のピアソン積率相関係数を計算すると、0.946となり、かなり高い正の相関がみられること

図1. 人種と文盲の散布図 (九地区分類)  
1930年合衆国センサスデータ

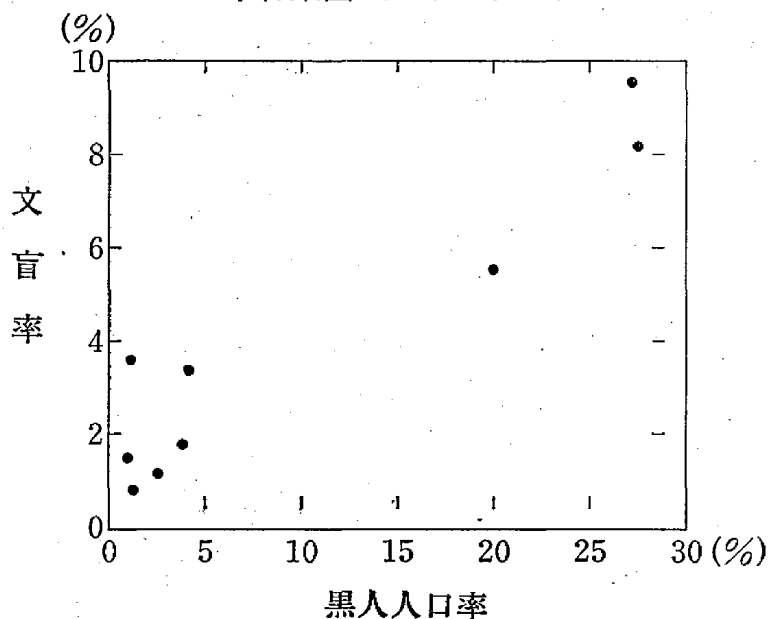


表2 人種と文盲との個人相関  
1930年合衆国センサスデータ

	黒人	白人	合計
文盲	1,512	2,406	3,918
識字者	7,780	85,574	93,354
合計	9,292	87,980	97,272

(10歳以上 単位千人)

を示す値となる。

しかしながら、個人の属性である人種と文盲の相関を、地域別の百分比という間接的な方法ではなく、人種と文盲の同時分布として直接みていったものが表2である。このときのピアソン四分点相関係数を求めると0.203となる。この値は生態学的相関を用いた場合の0.946に比べかなり低く、黒人と文盲の相関関係はかなり弱いことをしめす。地域をより細分化させて州単位の生態学的相関で相関係数を求めても、その値は0.773であり、個人相関との差は大きい。なお、ピアソン四分点相関係数は、 $2 \times 2$ のクロス表を用いた場合にピアソン積率相関係数と数学的に同様のものである。

生態学的相関と個人相関による値がこれほど違うのは、生態学的相関も個人相関も共に一定の地域内での個人データを集計して求められるが、それぞれが依存するデータが異なっているためである。つまり、個人相関は同時分布に依存しているのに対して、生態学的相関は周辺分布に依存している。周辺分布は同時分布により決定されるが、同時分布は周辺分布により決定されるものではない。このため、同じ周辺分布をとるような同時分布の組み合わせは多数存在することになり、生態学的相関を用いて、個人相関を推定することはできないとしている。

このような理由から、生態学的相関と個人相関とが全く逆の相関関係を示す場合すらあるとして、出生地と文盲との関係の例をあげている。外国出生者の比率と文盲の比率との関係を九地区分類による生態学的相関でみると、その相関係数は0.118となる。これは、相関の程度はあまり大きくないが、外国出生者が多い地域は文盲の率が高く、外国出生者には文盲が多いことをしめしていると予測される。これに対して、外国出生者と文盲との個人相関をとってみると、四分点相関係数は $-0.526$ となり、生態学的相関とは逆に、外国出生者には文盲が少ないとの傾向を示している。

(2) 誤謬の起きる数学的理由

生態学的相関と個人相関の値が異なる理由をロビンソンは、つぎのようにして数学的に示している。

まず、以下のように仮定する。

- (a) 全数Nの集団があり、それはXとYの2つの変数属性によって、特徴づけられている。これらの属性には年齢や収入のように数値であらわされる変数もあるし、性や識字のように二分法的属性によってあらわされる変数もある。
- (b) 全数Nの全集団は、センサス地区や町村、郡、州、地方などの地理的状况に応じて、m個の下位集団に入れることができる。これらの下位集団をmの値の第三の変数Aとして定義する。

このように仮定したうえで、全個人相関 (total individual correlation)  $r$ 、生態学的相関 (ecological correlation)  $r_e$ 、地域内個人相関 (within-areas individual correlation)  $r_w$  の関係を、個人の属性 XY と地域をしめす変数 A を用いて、次のよう定式化している。ここで、全個人相関 ( $r$ ) は集団の全成員NについてのXとYとの単純ピアソン相関であり、もしXとYとが二分法的な属性のときは、全個人相関は四分点相関となる。生態学的相関 ( $r_e$ ) は、下位集団に属するXとYとの百分比のm個の組の重みつき相関となる。地域内個人相関 ( $r_w$ ) はXとYとのm個の地域間での個人相関の重みつき平均であり、それぞれの地域内相関にそれぞれの属する集団の大きさによる重みづけを行なったものである。さらにこの関係式では2つの相関比  $\eta_{XA}$  と  $\eta_{YA}$  が用いられる。これは地域ごとのXとYとのそれぞれの値の示す程度をあらわすものである。たとえば、Xを文盲率とすると  $\eta_{XA}$  の値が大きいときは、地域による文盲率のばらつきが大きいことを示す。

こうして定式化された関係は次のようになっている。

$$r_e = k_1 r - k_2 r_w \dots\dots\dots(1)$$

$$\text{ただし } k_1 = \frac{1}{\eta_{XA}\eta_{YA}}$$

$$k_2 = \frac{\sqrt{1-\eta_{XA}^2} \sqrt{1-\eta_{YA}^2}}{\eta_{YA}\eta_{XA}}$$

ここで、個人相関と生態学的相関とが等しくなる場合、つまり  $r_e = r$  となる場合を仮定してみると、その時には次の関係が成り立つことになる。

$$r_w = k_3 r \dots\dots\dots(2)$$

$$\text{ただし } k_3 = \frac{1 - \eta_{XA}\eta_{YA}}{\sqrt{1 - \eta_{XA}^2} \sqrt{1 - \eta_{YA}^2}}$$

しかしながら、 $k_3$ の値に注目すると、

$$\text{分子}^2 - \text{分母}^2 = (\eta_{XA} - \eta_{YA})^2 \geq 0$$

となり、 $k_3 \geq 1$  となっている。さらに、下位集団がいかに同質的であっても、XとYとの相関は下位集団が集団全体よりも大きくなることはない。このため  $r_e = r$  との仮定は誤りとなる。このため、(2)式の右辺と左辺との関係は、

$$r_w < k_3 r$$

となり、これにより、生態学的相関は全個人相関よりも大きくなることが導かれる。

また同じ(1)式から、生態学的相関の大きさは下位地域の数に依存しており、地域数が減少する、つまり単位となる地域が大きくなるほど生態学的相関の値が大きくなることがわかるとしている。小さな地域が結合されて大きな地域にまとめられると、(1)式の右辺の値が変わってくるが、この場合、次のような変化が起きる。

- (a) 下位地域の異質性が増加するため、地域内個人相関の平均的規模を増大させる。これは生態学的相関の値を低下させる
- (b) 下位地域でのXとYと同質性が減少するため、 $\eta_{XA}$ と $\eta_{YA}$ の値は減少する。これは生態学的相関の値を増加させる。

しかしながら、この二つの変化が全体の数値にあたえる効果は同じではない。(a)と(b)の変化の効果を比較すると(b)の効果のほうが大きくなり、このため、地域単位が大きくなるほど生態学的相関の値が大きくなることがわかるのである。

### (3) 結 論

こうした実例と数学的説明によると生態学的相関と個人相関との数値が等しくなることは理論的にはありえるが、通常データを利用する場合にはありえない、とロビンソンは述べている。実際にデータを利用する際に、「生態学的相関と個人相関との関係で確実なことは、生態学的相関はそれに対応する個人相関とは等しくならないということである」と結論づけている。さらに、「この結論は、近年の重要な研究の多くに対して、その有効性に疑問をなげかけるものである。将来の研究で意味のない相関〔生態学的相関〕の計算をさけることができ、個人の属性間の意味のある相関〔個人相関〕が用いられるとしたら、本論の目的は果たされるであろう」として論文を締めくくっている。

(4) ロビンソンの関係式の補足

ロビンソンが示した全個人相関・生態学的相関・地域内個人相関の三つについての関係式の数学的説明は、ロビンソン自らの論文では必ずしも十分ではない。そこで、この関係式が成り立つ理由をアルカー (Alker JR., H. R.) の説明により補足したい<sup>(2)</sup>。

アルカーはロビンソンの関係式の再構成を行なうに際して、時系列的要素を加えているが、ここではロビンソンがとりあげたように地域的に分属している個人についてのデータとしてみていきたい。

ロビンソンの関係式は、基本的には $N$ 個の統計学的単位における変数 $X$ 、変数 $Y$ についての共分散の定義式から導き出される。ここで統計学的単位となるのは個人であり、それぞれの個人は全部で $R$ となる地域 $r$ にそれぞれ属する。

ここで、まず次のような等式が成立を確認したい。すなわち

$$(X_{ir} - X_{..}) = (X_{ir} - X_{.r}) + (X_{.r} - X_{..})$$

- ただし  $X_{ir}$  : 個人それぞれの $X$ の値
- $X_{..}$  : 全体についての $X$ の平均値
- $X_{.r}$  : 地域ごとの $X$ の平均値

これを $X$ と $Y$ との共分散の定義式にあてはめると、

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N (X_{ir} - X_{..})(Y_{ir} - Y_{..}) &= \frac{1}{N} \sum_{i=1}^N (X_{ir} - X_{.r})(Y_{ir} - Y_{.r}) \\ &+ \frac{1}{N} \sum_{i=1}^N (X_{.r} - X_{..})(Y_{.r} - Y_{..}) + \frac{1}{N} \sum_{i=1}^N (X_{ir} - X_{.r})(Y_{..} - Y_{..}) \\ &+ \frac{1}{N} \sum_{i=1}^N (X_{.r} - X_{..})(Y_{ir} - Y_{.r}) \end{aligned}$$

となるが、後ろの2項は相殺しあうので、結局のところ、

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N (X_{ir} - X_{..})(Y_{ir} - Y_{..}) \\ = \frac{1}{N} \sum_{i=1}^N (X_{ir} - X_{.r})(Y_{ir} - Y_{.r}) + \frac{1}{N} \sum_{i=1}^N (X_{.r} - X_{..})(Y_{.r} - Y_{..}) \dots \dots (3) \end{aligned}$$

となる。(3)を $C_{XY}$  ( $X$ と $Y$ についての共分散),  $WC_{XY}$  ( $X$ と $Y$ についての地域内共分散),  $EC_{XY}$  ( $X$ と $Y$ についての地域間共分散, あるいは生態学的共分散) として書き改めると、

$$C_{XY} = WC_{XY} + EC_{XY} \dots \dots \dots (4)$$

この(4)式を相関係数を用いた関係式に改めるため、両辺を $\sqrt{C_{XX}C_{YY}}$  ( $C_{XX}$ ,  $C_{YY}$  はそれぞれ $X$ と $Y$ についての分散) で除し、右辺を地域内相関, 生態学的

相関で表わそうとすると、(4)式は次のようになる。

$$\begin{aligned}
 R_{XY} &= \frac{WC_{XY} + EC_{XY}}{\sqrt{C_{XX}C_{YY}}} \\
 &= \frac{WC_{XY}}{\sqrt{WC_{XX}WC_{YY}}} \cdot \sqrt{\frac{WC_{XX}WC_{YY}}{C_{XX}C_{YY}}} \\
 &\quad + \frac{EC_{XY}}{\sqrt{EC_{XX}EC_{YY}}} \cdot \sqrt{\frac{EC_{XX}EC_{YY}}{C_{XX}C_{YY}}} \dots\dots\dots(5)
 \end{aligned}$$

ただし  $R_{XY}$  :  $X$  と  $Y$  についての全個人相関

$WC_{XX}WC_{YY}$  :  $X$  (あるいは  $Y$ ) についての地域内分散

$EC_{XX}EC_{YY}$  :  $X$  (あるいは  $Y$ ) についての生態学的分散

さらに、 $WC_{XX} = C_{XX} - EC_{XX}$ 、 $WC_{YY} = C_{YY} - EC_{YY}$  との分散の関係式を用い、また相関比を用いると、(5)式は次のように改めることができる。

$$R_{XY} = WR_{XY} \sqrt{1 - \eta_{XR}^2} \sqrt{1 - \eta_{YR}^2} + ER_{XY} \eta_{XR} \eta_{YR} \dots\dots\dots(6)$$

ただし、 $WR_{XY}$  :  $X$  と  $Y$  についての地域内相関

$ER_{XY}$  :  $X$  と  $Y$  についての生態学的相関

この(6)式は、まさにロビンソンが提示した全個人相関・生態学的相関・地域内個人相関の関係式と同じものである。

### (5) ロビンソンにみる問題提示の特徴

こうして提示された生態学的誤謬問題をめぐっては多くの議論がなされており、本稿ではその一部をとりあげるにすぎない。生態学的誤謬問題がこれほど多くの関心をあつめたのは、何といたっても生態学的相関を利用した研究が無意味であるとの主張があたえた問題の深刻さがあげられよう。しかしながら、多くの議論をよんだことには問題提示の仕方との関連も深い。

まず、取り上げたデータが個人についての二分法的な外在的属性のデータであり、しかも合衆国全体といった広範囲の地域についての全数調査であるため、データをめぐるやっかいな議論のいくつかを避けることができ、問題の論点を絞りこむことができたことがあげられよう。それらはたとえば次のようなことである。外在的に分類可能な属性を取り上げたことにより、意識や態度についてのデータを利用する際のように多くの議論が集まる分類枠そのものの問題を避けることができる。また、二分法的属性をとりあつかうことにより多くの分析や検定の方法を比較的容易に利用できるため、集計の次元を異にする生態学的相関と個人相関との比較が可能となっている。さらに、全数調査によるデータによっているため、サンプリング調査の際の母集団の再現性の議論を避けることができる。



また、相関させた属性そのものに多くの人々の関心を集める要素があったこともあげられよう。人種問題や移民問題への関心が高い時代に、あえて黒人や移民と文盲との相関関係を取りあげ、それぞれの相関関係が生態学的相関により誇張されたり、逆転されたりするとの研究結果は、さまざまな立場からの関心をよんだものとおもわれる。

#### 4. 生態学的誤謬をめぐる議論

ロビンソンの生態学的誤謬の指摘に対する反応として、2つの方向のものがある。一つはロビンソンの指摘をうけて、生態学的データにより個人データの推定をおこないうるような、生態学的相関に代わる分析手法を提唱したものである。別の一つは、生態学的相関の研究は個人相関の代替として意義をもつだけでなく、生態学的相関をふくめた生態学的データそのものが社会学的研究において意義をもつとして、生態学的データを用いた研究を行なおうとするものである。次にこうした議論のいくつかをみていきたい。

##### (1) 個人データ推測のための生態学的データの分析法

生態学的データを利用して個人データを推定しようとする手法を代表するものとして、ダンカンとデイビスによる限界法的手法とグッドマンによる回帰分析を取りあげたい。

##### (a) ダンカンとデイビスによる限界法的手法

ダンカンとデイビス (Duncan, O. D. & Davis, B.) は、生態学的相関の研究それ自体の意義を認めたくて、「たとえ個人相関に関心がある研究者においても、生態学的相関が最良の方法ではないにしても、地域的に分類されたデータは役に立つであろう」として、生態学的相関により個人相関の研究の可能性を認めている。その方法として、生態学的相関によって知ることができる地域ごとの周辺度数から、同時分布の範囲を知り、個人相関の範囲を推定するものであり、「地域的に分類されたデータから個人相関をかなりの近似値まで確定できる」としている<sup>(3)</sup>。

彼らは周辺度数を用いて個人相関を推定する実例を、1940年と1945年のセンサス結果を用いてしめしている。まずとりあげたのは1940年のシカゴ市のセンサスによる「人種と女性労働者の家事労働 (domestic service) 従事率」の相関である。市域全体での個人相関は四分点相関係数で0.289となっていた。これに対して、市内にある935カ所の統計区ごとの四分点相関係数は-0.008～

表3 女性の職業と人種との個人相関  
1940年シカゴ第625センサス地区

	家事以外の職業	家事労働	全就業者
白人			95
非白人			782
合計	638	239	

a. 非白人の家事労働者が最大の場合

	家事以外の職業	家事労働	全就業者
白人	95	0	95
非白人	543	239	782
合計	638	239	877

b. 非白人の家事労働者が最小の場合

	家事以外の職業	家事労働	全就業者
白人	0	95	95
非白人	638	144	782
合計	638	239	877

0.898のかなり広い範囲の値を示している。しかしながら、統計区ごとに同時分布のとりうる最大値と最小値とを推定することにより、その範囲をかなり狭めることができる。

その推定は次のようにして行なう。第625センサス地区のデータを例にすると、人種と女性就業者の職業の周辺分布は表3に示すようになっていた。これを用いて、非白人の家事労働者が最大の場合と非白人の家事労働者が最小の場合とを想定するのである。非白人の家事労働者が最大の場合の同時分布は表3-aのようになり、非白人の家事労働者が最小の場合には表3-bのようになる。これをそれぞれの統計区ごとに最大値・最小値の推定をおこない、そこからシカゴ市全体の同時分布の最大値・最小値を求めたうえで、推定されたシカゴ市全体の四分点相関係数の範囲は0.126~0.355となり、統計区ごとの四分点相関係数の範囲-0.008~0.898に比べて、範囲はかなり狭まっている。

同様に、「人種と自己所有住宅率」をみると、統計区(935)ごとの四分点相関係数は-0.521~0.168の値を示すのに対し、先ほどと同様の手法で計算し推定した四分点相関係数の範囲は0.058~0.138となり、範囲はかなり狭まる。実際の市全体の四分点相関係数は0.116であった。

限界法的手法は、このようにして周辺分布のデータから同時分布のとりうる範囲を推定する方法であり、確率論的な手続きを要しないという利点がある。

しかしながら、求めた範囲内で特定の値をしめす確率などを知ることはできないし、また求めた範囲の中央値付近に実際の値が現れるわけでもない。さらに推定値の範囲が広い場合にはその利用に困る場合もある。

### (b) グッドマンなどによる回帰分析的手法

ロビンソン問題の解決法として、回帰分析を用いて生態学的データから個人データを推定しようとする研究が進み、その手法も精緻化している。

グッドマンは、ロビンソン論文へのコメントを加えた短い論文<sup>(4)</sup>のなかで、生態学的データを利用した研究が、回帰分析の手法を用いることで個人の行動の推定に役立つ例として次の2つをあげている。

一つは、2変数間に直線的回帰関係があるかどうかを知りたい場合で、ロビンソンの黒人率と文盲率の関係でいうと、高い黒人人口率の地域が高い文盲率を示すかどうかを知りたい場合である。こうした関係は、生態学的変数間の回帰の分析によって知ることができる。もう一つは、生態学的変数間の回帰の研究を「個人の行動」を推測するものとして利用しようとするものであり、同じ例でいうと、黒人であることと文盲であることとの関係を推測しようとするものである。この場合、特定の条件のもとでは回帰分析による推測が可能となるとしている。

のちに、ロビンソンおよびダンカンとデイビスの論文をふまえて、自らの回帰分析の手法を詳しく示した論文<sup>(5)</sup>のなかで、グッドマンはこうした考えを補足している。彼はまず、先の論文で用いた「個人の行動」とは個人の属性についての変数を示すとしたうえで、回帰分析により個人の属性を推測しうる場合の条件として、個人データが未知であり、個人の集団についての二分法的属性のクロス分析であることをあげている。こうした限定を加えたうえで、ロビンソンの黒人率と文盲率の関係についての個人データの推測を回帰分析により行なっている。合衆国九地区分類による生態学的データに対する回帰分析から得られた個人データを利用すると、全個人相関の平均値は0.38と推定される。この値は、ロビンソンが行なった生態学的相関の値が0.95となり、またダンカンとデイビスの限界法を利用して算出した推定値が-0.07~0.60となるのに比べると、実際の全個人相関の値である0.20にかなり近いものとなっているとしている。

こうした生態学的相関に代わる分析手法が示されても、生態学的データを用いる研究において、それらは必ずしも広く利用されているわけではない。その理由として、シベリー (Shively) はそれぞれの手法に次のように難点があるためとしている。「ダンカンとデイビスの限界法は、実際にその範囲がかなり

ひろくなってしまう、可能な値の範囲ではなく、ひとつの値をもとめたがるためである」とし、生態学的回帰が用いられないのは、その推定の前提となる手続きが繁雑であるうえ、「〔構成比の〕推定値がマイナスや100パーセント以上の値をとってしまうことがあり、研究者をためらわせるためである」としている。「これに対して、生態学的相関による生態学的推定があやまりであっても、生態学的相関の値は-1.0から1.0と使いやすいため」であるとしている<sup>(6)</sup>。

しかしながら、生態学的相関に代わる分析手法が広く用いられない理由は、そうした手法が利用しにくいという技術的な理由からだけではない。むしろ生態学的相関それ自体を利用した研究に意義をみいだす研究者が少なくないためであろうと思われる。

## (2) 生態学的データ自体を利用した研究への肯定的評価

個人データの推定のために生態学的データを利用するのではなく、生態学的相関を含めた生態学的データそのものにも研究上の意義があるとする見解は、ロビンソンが生態学的誤謬を指摘した時点でもかなり強いものであった。そうした立場を代表するのがメンツェル (Menzel, H.) である。

メンツェルはロビンソン論文掲載直後に、短いコメントを載せている<sup>(7)</sup>。ここでは、ロビンソン論文に対して「一般的な誤謬に対する有効で重要な警告」との評価をのべたあと、次のような反論を加えている。生態学的相関を用いた研究では、生態学的相関は個人の属性間の相関が利用出来ないときにのみ用いられるものであり、地域それ自体の属性間の相関に関心を持つためではないとのロビンソンのコメントに対して、生態学的相関が個人相関を反映しなくても、生態学的相関の研究そのものに研究上の価値があるとした。

生態学的相関により「地域それ自体の固有な属性」を知ることができる。たとえば、まず、問題とする二つの属性の原因として共通する他の変数を知ることができる場合をあげている。そうした例のひとつとして、メンツェルは社会解体の指標となるようなものの地域相関をあげている。逮捕者と離婚との相関は高い値を示すが、この傾向は逮捕された人間が離婚しやすいとの一般的印象を与えがちである。しかし、この傾向はむしろ、「地域それ自体に固有な属性」、文化的葛藤、社会解体やそれに類するものが、その地域に潜在的に存在することの論拠として主張されるものである、としている。さらにロビンソンの取り上げた黒人と文盲との生態学的相関にみられる傾向は、多数の黒人の流入、学校制度の不備などの歴史的状況や経済発展の状態などが共通の原因となっているためであるとみることができるとしている。また、一つの変数が他の変数の原因となっている場合にも生態学的相関を利用した研究は有効であると

している。その例として、「ユダヤ人人口率と反ユダヤ主義」をあげ、ユダヤ人が反ユダヤ主義を示すのではないにもかかわらず、両者の相関が高いこと自体が研究上の価値をもつとしている。同様の例として「人口あたりの外科医数と乳児死亡率」の生態学的相関などもあるとしている。

こうした点から、個人相関を知ることは重要なことではあるが、だからといって生態学的相関を知ることの重要性を損なうものではないと主張している。

## 5. 生態学的データの利用の問題

### (1) 生態学的データ利用による誤謬

生態学的誤謬を生態学的データ、あるいは集団データを用いた場合に生じる誤謬とすると、生態学的誤謬が起りうるのはロビンソンが指摘した場合に限られるわけではない。こうした生態学的データを用いた場合に起りうる誤謬について、アルカーは次のような整理を行なっている<sup>(8)</sup>。

それによると、個人と集団や、統計区と市町村といったような分析する単位を異にしたデータについて、共分散や相関関係を利用した研究では、従来、証明されない多くの推論が用いられているとし、生態学的推論における誤謬の可能性を三つの水準から指摘できるとしている。

第一の場合は、アグレゲーションについてより広い地域のデータからより狭い地域のデータを推論する、つまり、アグレゲーションの高い水準から低い水準を推論する際の誤まりである。これにはロビンソンの提示したような生態学的データから個人データを推定するような誤まりも含まれる。

第二の場合は、分析のレベルを不正確に拡大して推論する際の誤まりである。これは特定の個人や、小さな集団あるいはアグレゲーションについての行動から、より大きな集団あるいはアグレゲーションの集会的関係を一般化しようとする際に起きるものであり、第1の場合と逆方向におこなる推論の際の誤まりである。アルカーはその例として、労働者が社会的にラジカルであるからといって、労働者が高い割合を占める経済的先進国においてラジカルな世論が形成されるといえないことをあげている。

第三の場合は、下位集団あるいはアグレゲーションのあるものから、他の集団あるいはアグレゲーションを推測しようとする際の誤まりである。これはサンプリングの技術と大きく関連する問題である。

さらに別な観点からすると、「誤謬性」そのものの意味が、データの社会的位置づけや研究上の位置づけにより違ってくる点を忘れてはなるまい。データとなった属性の持っている社会的意味の相違により、誤謬性のもつ重要度は異

なつてこよう。さらにデータの研究上の位置づけによつても誤謬性の持つ意味が異なる。データによつて示したいものが、強調すべきことであるのか、統計的關係であるのか、あるいは因果關係であるのかによつて、誤謬の許容される範囲は劇的に異なつてこよう。

## (2) 生態学的データ利用の研究の意義

ロビンソンは個人の属性間の研究において意味のある相関は個人相関のみであるとし、生態学的相関を意味のない相関であるとして、生態学的相関を用いた属性間の研究に意義を見出さなかつた。しかし、個人の属性間の研究において、生態学的相関を含めた生態学的データが役立たないわけではない。

まずあげられるのは、個人的データの代用として生態学的データを使わざるをえないという点である。ロビンソンが批判した研究の多くもそうであつたように、個人データを利用できない場合には生態学的データを使わざるをえない。特に、現在ではプライバシー保護の観点からロビンソンの時代以上に個人データに入手は困難となつてきており、生態学的データの消極的利用ではあるが、個人データの代用としての利用の重要性はますます高くなろう。

生態学的データからの個人データの推測には、コンピュータ利用法の発展の寄与するものが大である。生態学的データを利用して個人の属性間の關係を知ろうとする分析手法はかなり面倒な計算を要するため、従来はその利用は容易ではなかつたが、現在ではコンピュータの発達、特にパッケージ・プログラムの利用が進むことにより比較的容易にそうした手法が利用できるよつたになっている。さらにコンピュータを利用して、個人データによる個人の属性についても回帰分析やクラスター分析、因子分析など多変量解析が盛んにおこなわれるよつたになっている。もっとも、こうしたコンピュータの利用には問題点もあり、多変量解析による分析では、分析結果の検定方法がないことや、データの変化に敏感すぎるため、データの僅かな変化によつて結果に大きな影響が出ることなどなどがある。また、コンピュータ利用によつて、分析過程がブラック・ボックス化してしまい、分析の手順を追いきれず、分析手法にふりまわされてしなう恐れがある。

さて、より積極的な意味で、生態学的データ、とくに生態学的相関を利用することの意義もある。それは、個人の属性に対して集団があたえる影響を知ることであり、それには生態学的データや生態学的相関を利用しなければならないからである。こうした分析に有効であるのは、集団の分析で用いられるエラボレーション (elabolation) の手法である。この手法を實際に用いて、集団や地域性の影響を生態学的データからとりだそうとするものには、ブラウなどの

構造効果や、サイモンとブレイバックの因果推論などの分析法があり、法則定立的な研究には有効な手法である。

こうした点からすると、個人属性間の関係の分析法についても考え直す必要がある。ロビンソンが個人属性間の関係を反映するとみた全個人相関には、地域や集団による影響が考えられる。このため、集団の影響を除かなければ、実際の個人属性間の関係を知ることはできない。実際に、こうした観点からの推論を試みた研究がある。それはアルカーによるもので<sup>(9)</sup>、生態学的相関による値が0.95であったものが、個人相関により0.20まで下がったにしても、この相関係数は黒人と文盲との関係を示すものとしては大きすぎるとし、第三変数として産業化あるいは都市化を考えたらうえて、サイモンとブレイバックの因果推論の手法を用いて黒人と文盲の関係を考察している。そこで、より確かな相関として地域内個人相関で考えてみると、黒人と文盲との相関は0.05以下となり、黒人と文盲との関係はより弱くなってくるのである。メンツェルが指摘した、黒人と文盲との相関においては、黒人の流入や学校制度の不備など両者に共通の原因の影響を考えるべきだとした考えは、こうした手法によって確認が可能となろう。

このようにみると生態学的データの利用についての議論は、ある意味では出発点にもどったとの観もある。しかしながら、ロビンソンが指適した問題をめぐる議論のなかで、統計的データのさまざまな分析手法が提示され、また従来からのものも含めて、それぞれの分析法の適合範囲も明らかになってきた。こうした過程で、社会学の古典的研究での分析法の再評価もおこなわれるようになった。しかしながら、他方では統計的データを利用した研究に不信感を持つ研究者も少なくない。統計的データを利用する研究において、より説得力のあるデータの利用を行なうため、データの吟味や分析法の選択が今後も重要な課題でありつづけよう。

〔注〕

- (1) Robinson, W. S. "Ecological Correlations and the Behavior of Individuals" *A. S. R.* 15 (1950) pp. 351-357
- (2) Alker JR., R. A. "A Typology of Ecological Fallacies" in M. Dogan & S. Rokkan eds. *Quantitative Ecological Analysis in the Social Sciences* (1969)
- (3) Duncan, O. D. & Davis, B. "An Alternative to Ecological Correlation" *A. S. R.* 18 (1953) pp. 665-666
- (4) Goodman, L. A. "Ecological Regressions and Behavior of Individuals" *A. S. R.* 18 (1953) pp. 663-664

- (5) Goodman, L. A. "Some Alternatives to Ecological Correlation" *A. J. S.* 64 (1959) pp. 610-625
- (6) Shively, W. P. "'Ecological' Inference: The Use of Aggregate Data to Study Individuals" *The American Political Science Review* 63 (1969) pp. 1183-1196
- (7) Menzel, H. "Comment on Robinson's 'Ecological Correlations and the Behavior of Individuals'" *A. S. R.* 15 (1950) p. 674
- (8) Alker JR., *op. cit.*
- (9) *ibid.*