

1. はじめに

交通容量の拡大を目指した時代から交通需要を管理する時代へと変遷する中で、交通主体の意思決定をより直接的に扱う交通需要予測手法として非集計アプローチが急速に普及してきた。それに伴って交通行動調査手法も大規模でマクロな調査から小規模でミクロな調査へとニーズが広がってきた。対象とする交通政策の時間スケール（長期、中期、短期）、空間スケール（ネットワーク全体、幹線網、地区）、対象者（受益者、被害者、非利用者）に応じて、適切な交通調査手法を選別し実施することが今問われている。またある目的で収集した調査データを他の目的でも活用できるようデータの公開、共有と標準化、一般化に対する要望も強い。

一方、調査に対する費用対効果の視点も重視されてきた。サンプルサイズを増し母集団に対する網羅性を高めることはデータの精度を高める反面、膨大な調査費用を要する。一般論として精度と費用のバランス点をとるような調査を設計することが解決策とされるが、交通調査に対するニーズが多様化するにつれて調査方法の工夫だけに解を見つけることは困難である。

そこで、調査手法の洗練化については別の論文に委ねることとし、本稿では交通調査から得られたデータの効率的な利用方法に焦点を絞り、交通データに存在する無回答問題についてレビューし、その対応方法について考察する。

2. 交通データにおける無回答問題

パーソントリップ（PT）調査などのアンケート調査では無回答の問題は不可避である。無回答は次の2つに分類される：

a) アイテム無回答

個人としては多くの質問項目に答えているものの、一部の質問項目について明らかに誤っていたり、答えなかったものがある場合

b) ユニット無回答

個人が白票で返却したり、回答を拒否したりして

表1 KONTIVにおけるアイテム無回答

調査項目	欠損率(%)	調査項目	欠損率(%)
<個人属性>		<交通属性>	
性別	3.00	目的地	15.0
年齢	4.50	目的	10.0
結婚	3.50	手段	54.7
学歴	9.40	所要時間	20.3
職業	9.50		
運転免許	9.50		

表2 広島都市圏 PT 調査におけるユニット無回答

市・町	人口	回収全数	有効標本数	欠損率(%)
広島市	1,042,308	83,342	72,822	12.62
呉市	225,040	18,211	16,053	11.85
大竹市	32,576	2,543	2,283	10.22
廿日市市	55,080	4,328	3,718	14.09
府中町	49,440	3,920	3,465	11.61
海田町	30,510	2,447	2,167	11.44
熊野町	26,089	2,109	1,911	9.39
坂町	13,401	1,034	951	8.03
大野町	24,405	1,938	1,751	9.65
合計	1,498,849	119,872	105,121	12.31

当該個人の回答に関する情報が一切入手できない場合

前者の例としてドイツの大規模交通実態調査である KONTIV の報告結果を表1に示す¹⁾。個人属性に関する項目について欠損率は10%未満と低いものの、交通属性に関する項目ではデータの欠損率が高い。特に交通手段に関しては半数を超える回答者が無回答になっている。1日の行動を逐一思い出すことの困難さや代替の交通サービスの情報欠如などが主な原因とされている。

後者の例として広島都市圏で1987年に実施されたPT調査におけるユニット無回答の報告結果を表2に示す²⁾。全体として約12%のユニット無回答の存在が報告されており、最大で4.6%の地域間格差がみられる。ユニット無回答が特定の社会階層や地域に偏って発生する可能性があると言える。

以上の無回答データに対して、これまでの対応方法としては次のようなものが代表的であった。

a) 無回答の個人情報全てを削除する方法

b) 回答データに相応の重み付けをする方法

a) は調査効率を無視した対応であり、通常バイ

アスも大きい。b) の対応方法としては、無回答や無効票の内容は有効回収票と同質であると仮定して、適切な拡大率により処理するものである。この方法は無回答がランダムに発生する場合それほど問題は生じないが、調査への関心度や抵抗など非観測要因に起因して無回答が生じる場合、回答と無回答の間に偏りすなわちバイアスが生じ、結果が歪められる危険性がある。

また、都市の幹線交通をとらえることを主目的とする大規模調査では無回答の問題は無視し得ることもあるが、交通行動分析を前提とした小規模調査では重大な問題になることがある。例えば、単身世帯、共稼ぎ世帯などで有効回答率が著しく低い場合、自動車保有や休日の買物など、世帯タイプと密接に関連する交通行動を過小に見積もってしまうことが多い。このようなデータに基づいて需要予測をした場合は、交通計画全体が誤ったものになる。

3. 無回答バイアスの修正方法

(1) Imputation 法

無回答を防止するためには、まず調査段階において、サンプリング方法、調査の設計、調査の道具、調査の管理などに細心の注意を払うことが先決である。しかし一度ランダムな誤差や系統的な誤差が生じた場合は、Imputation 法が実用的であると考えられる。

Imputation 法とはアイテム無回答を以下の方法で補完し、疑似完全データとして扱うものである。

- a) hot-deck imputation：無回答をサンプル中の完全データでそのまま置換する方法
- b) 平均値 imputation：無回答を回答データから求められる平均値で代用する方法
- c) 回帰 imputation：無回答アイテムを回答されている他のアイテム値で回帰推計する方法

(2) EM アルゴリズム法

(1) の Imputation 法では個々のアイテム無回答を補完した後に完全データと同様の手順で分析を進めるが、この方法は個々のデータを推定するのではなく、推定母数に含まれる無回答バイアスを修正し、正しい十分統計量を推定するものである。特徴は他の方法に比べて汎用性が高い点にある。

EM アルゴリズムは E (Expectation) ステップ

表3 EM アルゴリズムによる無回答バイアスの修正結果

	完全データ	20%欠損	EM 修正値	修正率(%)
費用	-0.059	-0.001	-0.058	99.0
乗車時間	-0.554	-0.081	-0.530	94.9
アクセス時間	-1.319	-0.260	-1.228	91.4
待ち時間	-0.187	-0.255	-0.225	45.0

と M (Maximization) ステップから成る。まず M ステップで、無回答データは存在しないものとして母数 θ を通常最尤法により推定する。次に E ステップで、回答データを用いて欠損値の条件付き期待値を計算する。この 2 ステップを反復して無回答バイアスを修正する。

一例として表 3 に無回答を含む交通機関選択データに EM アルゴリズム法を適用した場合の修正結果を示す。4 説明変数をもつ 2 項選択ロジットモデルの推定パラメータの真値 (無回答のない完全データで推定した値)、バイアス値 (1 つのアイテムが 20% 欠損した場合の無回答バイアスを含む値)、EM 修正値とその修正率を示している。EM アルゴリズム法によりモデルパラメータが大幅に修正されることが明らかである。

(3) 他のデータソースの活用

以上の方法はアイテム無回答の対応方法であるが、他のデータソースを活用することによってユニット無回答の問題に対応することもできる。例えば、交通行動調査の被験者の住所、年齢、自動車保有台数などがサンプリングの段階で国勢調査データ等から別途得られている場合には、これらの個人情報に基づいて上記の方法でユニット無回答バイアスを補正することが理論上可能である。

また普及が著しい GIS を活用することにより、移動の OD と時刻さえ回答されていれば、利用可能なすべての経路の所用時間や乗換回数などの客観値を算定することが可能である。この方法により一定水準の精度で無回答が補完できることが実証されれば、交通属性に関する質問項目の削減により有効回答率の向上、調査の効率化を図ることができる。

参考文献

- 1) A.Richardson, E.Ampt, A.Meyburg : Survey Methods for Transport Planning, Eucalyptus Press, p.314,1995
- 2) 広島都市圏交通計画協議会：昭和 63 年度広島都市圏パースントリップ調査報告書－3 現況集計編, p.6, 1989.