交通調査データにおける無回答バイアスの取り扱い方法†
How to Correct Non-response Biases in Transport Survey Data

藤原章正††，杉恵頼寧†††，原田慎也††††
By Akimasa FUJIWARA, Yoriyasu SUGIE and Shinya Harada

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を見出すために，交通データに存在する無回答バイアスの問題についてレビューし，その対応方法について検討する．ここで無回答バイアスとは回答が欠損することよって観測される交通行動や行動と要因との因果関係が歪むことをいう．

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

2. 交通調査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

---

† キーワード：調査論，EMアルゴリズム

†† 正員，工博，広島大学大学院国際協力研究科
（東広島市鏡山 1-5-1Phone&Fax: 0824-24-6921）
††† 正員，工博，広島大学大学院国際協力研究科
（東広島市鏡山 1-5-1Phone&Fax: 0824-24-6919）
†††† 学生員，広島大学大学院国際協力研究科
（東広島市鏡山 1-5-1Phone&Fax: 0824-24-6921）

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）無回答の個人情報を全て削除する方法

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許容誤差として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 法を適用しこれらの誤差を修正すること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回答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[3),4)]

(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．

例えば，交通需要モデルの無回答バイアスを回避するためにEMアルゴリズムを適用する場合の基本的な手順を以下の通りである．

1) 前節で述べた単純な imputation 法によって欠損値の初期推定値を計算し，観測値と推定値で仮の完全データを作る．
2) 最尤法によってモデルパラメータを推定し，それを仮の完全データに適用する．
3) モデルのパラメータなどの期待値を用いて欠損値を再推定する．
4) 欠損値の再推定値を用いてモデルパラメータを再推定する．
5) step 3 と 4 を収束するまで繰り返す．

4. アイテム無回答の修正

(1) 分析方法

9 アイテム 1000 ユニットの仮想データを作成する．9 アイテムのうち 1 つは鉄道と自動車の選択結

果を表す2値データであり，残り8アイテムは交通機関選択を説明する要因（費用，乗車時間，アクセス時間，待ち時間）である．アイテム間には各々相関が存在するよう乱数を発生させてデータを作成する．

このデータを完全データと考え，交通機関選択を除く8つのアイテムの中から一部を欠損させた場合を不完全データとする．欠損の仕方を変化させて，多様な無回答パターンの下でEMアルゴリズム法の適用効果を測定する．

(2) 平均値のバイアスの修正結果

図1に1アイテム（鉄道の費用）のみ欠損させた場合において，修正前のバイアスを含む平均値と，EMアルゴリズム法により修正した後の平均値を，欠損率を 20%～80%の範囲で4段階に変えながら比較した結果を示す．欠損率が高くなればなるほど欠損データの平均値は下がるが，修正後のデータの平均値は完全データの平均値に近いことが分かる．特に欠損率が 60%までであれば，EMアルゴリズム法によりアイテムの平均値をほぼ正確に修正できることが分かる．

次に複数のアイテムが同時に欠損した場合について検討する．図2は1～3個のアイテムが同時に20%ずつ欠損した場合の欠損アイテム数とEMアルゴリズム法による平均値の改善率との関係を示したものである．ここで改善率 κ は以下の式で算出した．

$$\kappa = \left[1 - \left\{\left(\theta - \hat{\theta}\right)/\left(\theta - \tilde{\theta}\right)\right\}\right] \times 100 \qquad (1)$$

ここで

θ ：完全データの平均値（真値）

$\tilde{\theta}$ ：修正前データの平均値

$\hat{\theta}$ ：修正後データの平均値

なお，2アイテム（3アイテム）欠損の場合の改善率は，当該アイテムと他の各アイテムとの5通り（10 通り）の組合せにおいて得られた改善率の平均値である．

1アイテム欠損の場合に比べて3アイテム欠損の場合は，改善率が約 72～84%まで低下する．図2の中でアクセス時間（鉄道）と待ち時間（鉄道）

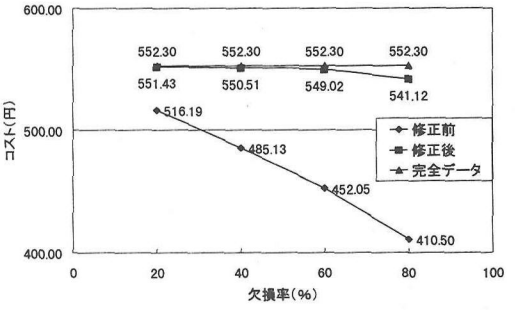

図1　費用(鉄道)の平均値と欠損率との関係

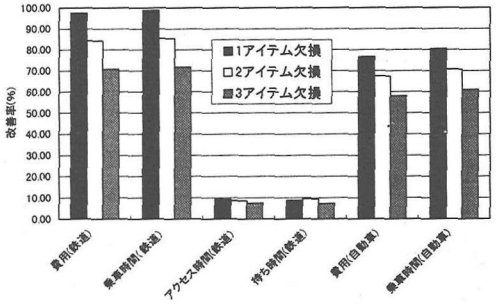

図2　欠損アイテムの数と平均値の改善率との関係

表3　EMアルゴリズムによる推定パラメータの修正結果

	完全データ	20%欠損	EM 修正値	改善率(%)
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

の改善率が低いのは，他のアイテム間に比べてアクセス時間（鉄道）と待ち時間（鉄道）との相関が小さく設定されているためと考えられる．通常の交通データではアイテム間にある程度の相関があるので，EMアルゴリズム法によりアイテム無回答に伴うバイアスが修正されることが期待できる．

(3) モデルパラメータのバイアスの修正結果

交通需要予測において無回答バイアスの重大な問題は予測モデルの推定パラメータにバイアスが生じることである．そこで，交通機関選択の需要予測で頻繁に用いられる非集計ロジットモデルを事例として，不完全データとEMアルゴリズム法による修正後のデータに適用し，推定パラメータの比較を行ってアイテム無回答がモデルパラメータに及ぼす影響

について分析する．なお，分析に用いるモデルは2項選択ロジット型であり，4つの共通変数からなる線形効用関数をもつとする．

表3は費用（鉄道）のみが 20%欠損した場合の各変数（アイテム）のパラメータ推定値とその改善率を示したものである．欠損した場合，待ち時間を除くパラメータ推定値は完全データの時に求められる真値よりも絶対値が小さく過小評価となる．しかしEMアルゴリズム法の適用によりこのようなバイアスは 90%以上改善されており，アイテム無回答に対するこの修正法の有効性が認められた．

5. ユニット無回答の修正

ユニット無回答の問題は他のデータソースを活用することによって対処することが考えられる．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，これらの個人情報を基にアイテム無回答と同様の方法でユニット無回答バイアスを修正することが理論上可能である．そこで，アイテム無回答の分析で用いた仮想データと同様の変数に性別，公共交通機関までの距離，自動車運転免許の保有状態という3つの属性変数を加えた，新たな仮想データを作り分析を進めることとする．

図4は，ユニットの欠損率と平均値の改善率との関係を，各アイテム別にグラフで表したものである．改善率は 12～47%程度であり図2のアイテム無回答の場合に比べてEMアルゴリズム法による改善効果は低い．その中でアクセス時間(鉄道)に関してはやや改善効果が認められる．これは，バイアス修正のために採用した外部情報の中に「公共交通機関までの距離」というアクセス時間と相関の高いアイテムが含まれていたためであろう．換言するとユニット無回答のバイアスの修正は使用する外部情報に大きく依存することになる．

本分析で想定したユニット無回答とアイテム無回答の大きな違いは，モデルの目的変量である交通機関選択データが欠損するか否かである．すなわちユニット無回答のように目的変量が欠損する場合においては，アイテム無回答のような修正効果は期待できないことが予想される．

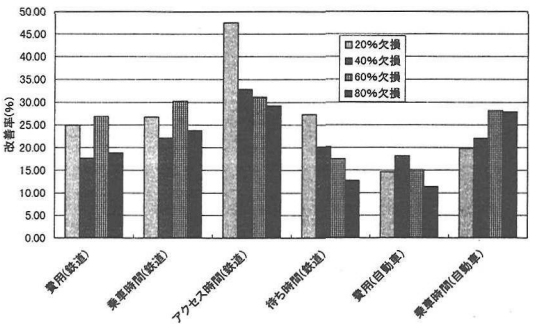

図4　ユニット欠損率と平均値の改善率との関係

6. おわりに

本研究では仮想的な交通データのもとで，アイテム無回答とユニット無回答に伴うバイアスの修正方法としてEMアルゴリズムを取り上げ，改善効果について検討した．1つの重要な結果として，アイテム無回答によるバイアスはEMアルゴリズムの適用によって修正可能であることが確認された．

一方，ユニット無回答に伴うバイアスの修正に関しては，あまり修正効果が認められなかった．特に交通需要モデルを前提とした調査データを扱う場合には，目的変量が欠損するため，アイテム無回答と同様の扱いでは不十分な場合があると考えられる．またユニット無回答バイアスを扱う際には，外部情報が重要となるが，例えば昨今普及が著しい GIS を活用することにより，交通サービス水準に関する客観値を外部情報として利用するなど，情報の収集法についても今後の研究課題としたい．

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ソントリップ調査報告書－3 現況集計編, p.6, 1989.
3) J.Polak, X.L.Han : Iterative Imputation Based Methods for Unit and Item Non-Response in Travel Diary Surveys, 8th Meeting of the IATBR, Austin, pp.21-25 1997.
4) A.Dempster, N.Laird and D.Rubin : Maximum Likelihood from Incomplete Data via the EM Algorithm, Journal of the Royal Statistical Society B, No.39, pp.1-38, 1977.