

GISを活用した交通日誌の欠損データの補填方法* Imputation Method of Missing Data of Travel Diary by Using GIS*

藤原章正**・杉恵頼寧**・山下大輔***

By Akimasa FUJIWARA**, Yoriyasu SUGIE ** and Daisuke YAMASHITA ***

1. はじめに

より短期でミクロな交通計画の需要予測を行うための分析技法として、非集計モデルの研究が1970年代後半より始められ、今日まで様々な手法によりモデルの改良が行われてきた¹⁾。一方、モデルを構築する際に使用するデータに関しては、パーソントリップ調査で代表されるような交通日誌調査から得た主観値データを用いることが多く、その精度や効率性に関しては1980年代後半に議論がなされた²⁾。しかし最近になって、モデル技法の精緻化に伴って必要とされるデータの量と質が高くなり、さらに道路地図のデジタル化や交通量実測データの蓄積が進んだ結果、交通日誌データの有効な活用法について再検討が必要となってきた。

本研究では近年普及が著しいGISを用いたネットワーク法³⁾により得られる客観値データを用いて交通日誌調査の欠損データを補填する方法を開発することを目的とする。具体的には、アンケートやインタビューで得た主観値データの中の欠損部分を、客観値データすなわち交通センサスなどの実測データをGISによる計算データで補填する方法を提案する(図1)。またこの方法が非集計モデルの推定結果に及ぼす影響について検討する。

これまで行われてきた欠損データの補填方法は、交通ネットワーク法⁴⁾とImputation法⁵⁾に大別される。前者はネットワークから求められる客観値をもって欠損データを補填する方法であり、本研究はこの方法の改善策としてGISを活用するものである。後者は完全な主観値データ(回答データ)により欠損データを補填する方法であり、例えば回答データの平均値で補填する平均値Imputation法などが一般的である。

本研究の流れは以下の通りである。

- 1) 交通ネットワークを作成し、個人別に交通サービス変数の客観値データを作成する。その際、既存統

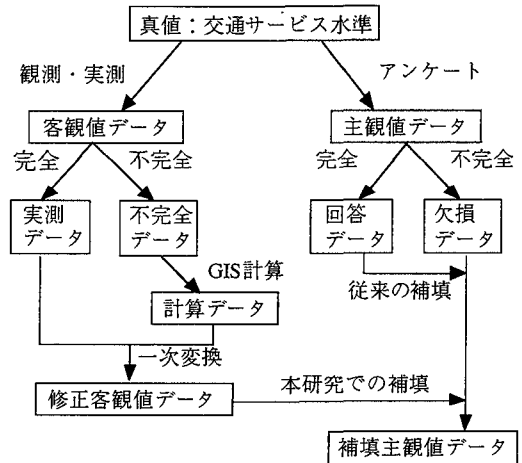


図1 データの定義と欠損値の補填方法

計データから直接得られない不完全データについてはGISを用いて計算データを算出する。

- 2) ネットワークから得られた客観値(実測データと計算データ)と主観値の回答データとの関係式を導き、客観値を主観値に変換する。変換した修正客観値により主観値の欠損データを補填する。
- 3) 客観値及び変換した主観値のそれぞれで補填した場合のデータでモデルを構築し、尤度比、的中率等のモデルの適合度について、従来の補填方法との比較を行う。

2. 使用データの分析

交通日誌データを用いて非集計行動モデルを推定する際、従来、欠損データや論理的に誤った値のデータは無効データとして分析から除外されたり、修正したりすることが多かった。ここでは、1994年に行われた広島市新交通システムの利用実態調査から得た交通日誌データを事例として、このような欠損データや誤データがどれくらい含まれているかについて確認する。同調査は家庭訪問留置回収により実施された自己記入式アンケートである。

* キーワード：交通日誌データ、調査論、GIS

** 正会員，工博，広島大学大学院国際協力研究科

(東広島市鏡山1-5-1，TEL & FAX: 0824-24-6921)

*** 正会員，工修，復建調査設計(株)

(広島市東区光町2-10-11，TEL: 082-506-1853)

(1)交通日誌の欠損率

表1から、個人属性である年齢については完全に回答されているが、交通サービス水準に関しては約40%以上が欠損している。欠損の理由としては、普段利用しなくてわからない、記入もれなどが考えられる。サンプル全体を通して、完全回答率は21.1%に過ぎず、明らかに調査費用のロスは大い。

(2) 誤記入や完全情報の欠如による誤差

アンケートに回答する場合において、普段から利用する主要交通機関のサービス水準については比較的正確な回答がなされると考えられるが、普段は利用しない代替交通機関に関しては、個人の知覚の誤り、誤記入、完全情報の欠如による誤差などがより多く含まれていると考えられる。

そこで、主要交通機関と代替交通機関のサービス水準の回答データの平均値と分散の違いを新交通システムのアクセス時間を例に分析した。表2に示すように、第二高取団地からのアクセス時間以外は、主要交通機関と代替交通機関の間で分散に大きな差が見られた。

この結果の原因の一つは、主要交通機関に比べて代替交通機関の有効サンプル数が少ない、すなわち欠損データが多いことが考えられる。そこで本研究では、GISで計算した客観値（計算データ）を利用して代替交通機関のサービス水準の欠損データを補填することとする。

3. 回答値と客観値の比較

(1)回答値と客観値の関係

アンケート調査から得られる主観値は、回答者個人の知覚値や態度の指標であり、GISから計算される客観値と比較すると異なった特性を持っており、主観値の方が効用に対する個人間の知覚の違いが反映されている。そこで、本研究では客観値すなわち実測データやGISによる計算データを簡単な変換式によって主観値に変換する作業を主要交通機関と代替交通機関に分けて行った。例として自動車の所要時間の関係を（図2, 3）に示した。

決定係数から判断して、主観値と客観値間の関係は単純な変換式によって十分に表すことができることが確認された。

(2)修正客観値と主観値の比較

前節で求めた変換式を用いて各交通サービス水準ごとに修正客観値を求め、主観値と比較し、変換の精度について調べる（表3, 4）。

表1 交通サービス水準の欠損率

項目 個人	新交通 乗車時間	新交通 AC時間	新交通 AC手段	駐車料金	自動車 乗車時間	年齢
1		5		10000		43
2	25		1	0		25
3	30	7			35	38
4		4	2	30000		33
.	.	.	.	.	.	.
.	.	.	.	.	.	.
6 1 1		9	1	15000	20	55
欠損率	40%	42%	39%	52%	48%	0%
完全回答率	21.1%					

表2 主要交通機関と代替交通機関の回答データ

団地	サンプル数	平均値		分散	
		主要	代替	主要	代替
第一高取	主要：45 代替：39	6.95	7.46	4.86	11.57
第二高取	主要：23 代替：19	10.00	11.32	12.55	12.12
長楽寺	主要：25 代替：17	7.20	9.35	7.08	48.20
相田	主要：11 代替：11	10.18	10.36	13.96	8.65
毘沙門	主要：41 代替：28	14.76	15.54	8.69	19.15

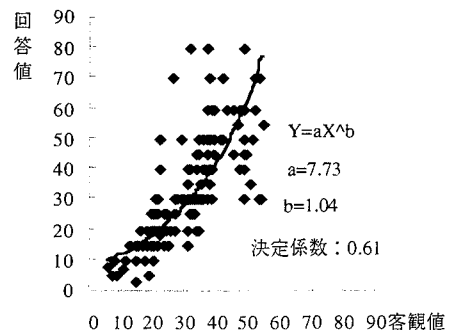


図2 自動車の所要時間の回答値と客観値の比較（自動車利用者）

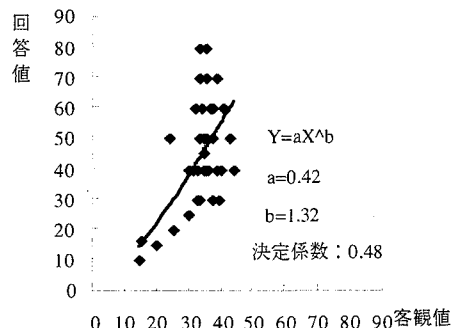


図3 自動車の所要時間の回答値と客観値の比較（自動車非利用者）

車の所要時間、新交通の乗車時間、アクセス時間とも主要交通機関に関しては、修正客観値と主観値の間に分散、平均値とも差が見られないという結果になった。

代替交通機関に関しても、平均値については有意な差は認められないことから、ある程度正確に修正できることがわかる。しかし、普段利用しない交通機関のため回答値の分散が大きくなり、修正客観値と主観値の分散には有意な差が出るという結果となった。

4. 欠損データの補填

(1) 補填方法と有効サンプル数

アンケート調査などから得た主観値データに含まれる欠損データを従来の方法によって補填した場合と、GISを用いて補填した場合の有効サンプル数の違いを図4に示す。

主観値データ284サンプルのうちアンケートで完全に回答された回答データは126サンプル（44.3%）に過ぎなかったが、回答データから欠損データを推定する従来の補填方法によると200サンプル（70.4%）まで増加する。またGISを活用した補填を行うとさらに16サンプル増加する。補填によるサンプルの増加率は交通機関により異なり、欠損の大きい交通手段ほど補填率が高い。

(2) モデル推定結果の比較

これらのサンプルを用いて非集計交通機関選択モデルを推定し、結果を比較する。推定したモデルは、新交通と自動車の2項選択ロジットモデルであり、説明変数として、個人属性（性別、年齢、職業）、交通サービス水準（自動車の所要時間、新交通アクセス時間、新交通乗車時間）を用いた。

表5から、回答データと補填主観値データとの間で性別と職業のパラメータに大きな差がある。これは、性別では全データのうち女性が1人、職業では無職者が2人と主観値データから不完全データを取り除いたことにより、回答データの母集団の代表性が欠如したことが原因である。この結果より、不完全データの補填を行いサンプルの偏りの修正が必要であることがわかる。

このモデルの説明変数から性別と職業を除いたモデル結果を表6に示す。従来の平均値imputation法による補填主観値データ(従来)のモデルと、GISを用いた補填主観値データ(提案)のモデルのパラメータ推定値には大きな違いが見られる。また、客観値をそのまま補填に用いた場合（客観値データ）と、主観値の欠損デ

表3 乗車時間の平均値と分散

		主観値	修正客観値	差
平均値	主要：自動車	34.7	33.1	1.6
	代替：自動車	46.1	44.2	1.9
	主要：新交通	22.6	22.5	0.1
	代替：新交通	16.1	15.8	0.3
分散	主要：自動車	317.9	255.5	62.4
	代替：自動車	278.4	114.4	164.0
	主要：新交通	31.3	24.6	6.7
	代替：新交通	58.7	48.4	10.3

表4 アクセス時間の平均値と分散

		主観値	修正客観値	差
平均値	主要：新交通	8.0	8.1	0.1
	代替：新交通	8.9	8.3	0.6
分散	主要：新交通	9.5	10.0	0.5
	代替：新交通	26.6	10.0	16.6

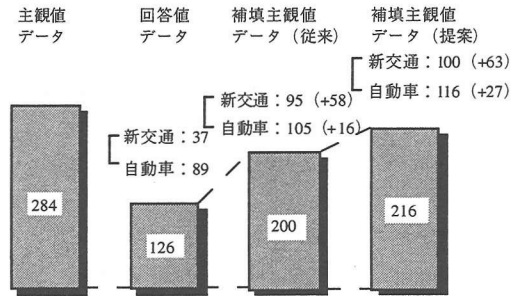


図4 補填方法と有効サンプル数

ータに修正客観値を用いて補填した場合（補填主観値データ(提案)）においても、推定パラメータに差が認められる。補填方法によりモデル推定結果に違いが現れることがわかる。

次に、補填主観値データを用いたモデルがどの程度の精度のものであるか調べるため、回答データ126サンプルのうち半分のデータを補填主観値データに置き換えてモデルに適用した。回答データを用いたモデルパラメータを真値とし、補填主観値データを用いたモデルパラメータの偏りを調べるためである。各モデルの推定結果を表7に示す。

表7から補填主観値データを用いたモデルの尤度比、的中率共に回答データを用いたモデルと同程度の値を示した。またパラメータベクトルの差の尤度比検定を行ったところ、 $\chi^2 = -1.0$ （自由度5）となり有意差は認められなかった。このことから、GISによる補填主

観値データを用いたモデルの精度は、完全に主観値データが収集できた場合のモデルと同程度であり、十分高いと言える。

5. 結論

本研究は、交通日誌調査データの効率的な活用を目指して、アンケート調査から得られる主観値と、道路交通センサス等から得られる客観値を活用したデータ補填の方法を提案した。特に主観値データに含まれる欠損データをGISを用いて算出した計算データにより補填する、補填主観値データの作成法の有効性を実証的に示した。この方法の有効性の評価指標としては、有効サンプル数の増加量、補正したデータの精度、非集計モデルの推定精度を採用した。

以下、本研究で得られた結果をまとめる。

- 1) ネットワークから得られた客観値を一次変換により主観値に近づけることにより、主要交通機関に関しては、主観値と有意差のないデータを作成することができた。
- 2) 従来の補填方法と比較して、本研究で提案したGISによる補填方法を採用すると有効サンプルの増加数が大きくより効率的であることが明らかとなった。
- 3) 無回答データをGISによる方法で補填した場合のモデル推定結果は、欠損の全く無い主観値データを用いたモデル推定結果とほぼ等しくなることが確認された。

参考文献

- 1) 屋井鉄雄，森川高行：交通需要モデル研究のダイナミズムー10年の軌跡，土木計画学研究講演集，No.14(2)，pp1-8，1991
- 2) 原田昇：交通行動調査のバイアスに関する研究のレビュー，交通工学，No.5 pp73-80，1989
- 3) Choi,K. and W. Jang：Transit Network Development and Transit Demand Modeling within Integrated GIS Using Spational Analysis and Dynamic Segmentation, Paper Presented at '98 WCTR in Antwerp, 1998.
- 4) 杉恵頼寧・薬師寺聡：公共輸送機関の経路探索が非集計交通機関選択モデルに及ぼす影響，土木計画学研究・論文集，No.6，pp.233-240，1988
- 5) 藤原章正・杉恵頼寧・原田慎也：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，土木計画学研究・講演集，No.21(2)，pp.61-64，1998

表5 補填方法の違いとモデル推定結果(1)

説明変数	回答データ	補填主観値データ (従来)	補填主観値データ (提案)
性別	-24.1 (0.001)	-5.15 (3.18)	-1.38 (1.68)
年齢	0.033 (1.11)	0.054 (2.68)	0.015 (1.03)
職業	20.4 (0.009)	1.21 (1.36)	0.42 (0.62)
アクセス (分)	-0.038 (0.86)	-0.067 (0.20)	-0.04 (1.13)
NTS乗車時間 (分)	0.046 (1.59)	0.062 (2.68)	0.097 (4.17)
自動車所要時間(分)	-0.015 (1.21)	-0.0024 (0.25)	-0.016 (1.65)
初期尤度	87.34	138.63	149.72
最終尤度	68.75	125.75	133.77
尤度比	0.20	0.07	0.09
的中率 (%)	70.6	66.5	67.1
サンプル数	126	200	216

(括弧内：t 値)

表6 補填方法の違いとモデル推定結果(2)

説明変数	補填主観値データ (従来)	補填主観値データ (提案)	客観値データ
年齢	0.042 (2.26)	0.037 (1.96)	0.037 (1.99)
アクセス (分)	-0.066 (0.20)	-0.023 (0.73)	-0.045 (1.41)
NTS乗車時間 (分)	0.063 (2.74)	0.099 (4.26)	0.061 (2.69)
自動車所要時間(分)	-0.067 (0.70)	-0.016 (1.54)	-0.018 (1.69)
定数項	-3.43 (3.23)	-3.13 (2.93)	-3.57 (3.33)
初期尤度	138.63	149.72	149.72
最終尤度	128.65	132.42	133.42
尤度比	0.06	0.11	0.10
的中率 (%)	68.0	68.5	66.7
サンプル数	200	216	216

(括弧内：t 値)

表7 補填主観値データを用いたモデルの精度

説明変数	回答データ	補填主観値データ (提案)
年齢	0.023 (0.82)	0.024 (0.82)
アクセス (分)	-0.040 (0.84)	-0.019 (0.52)
NTS乗車時間 (分)	0.050 (1.74)	0.070 (2.27)
自動車所要時間(分)	-0.015 (1.20)	-0.0048 (0.34)
定数項	-3.31 (2.14)	-3.76 (2.41)
初期尤度	87.34	87.34
最終尤度	70.71	71.21
尤度比	0.18	0.17
的中率 (%)	70.6	68.3
サンプル数	126	126

(括弧内：t 値)