

丸め誤差発生確率モデルによるトリップ出発時刻の回答誤差の推定

熊本大学 学生会員 ○佐藤 嘉洋

熊本大学 正会員 円山 琢也

1. 研究の背景と目的

一般的な社会調査において、回答には少なからず誤差が含まれていることが知られている。交通調査の分野における非標本誤差の一つとして「丸め誤差」が挙げられ、トリップ出発時刻、到着時刻にこの誤差が見られることが多いが、PT 調査などの大規模交通調査において時刻の丸め誤差に関する実態の把握はほとんど行われていない。しかし、丸められて回答された時刻を基に旅行時間が算出され、手段選択モデルの構築や時間帯別交通需要量分析等で利用されることを考慮すると、回答時刻の信頼性に関する精査は重要な課題と言える。そこで本研究では、丸められて回答されたデータに含まれる回答誤差をトリップ目的別・手段別に推定し、定量的な評価を行うことを目的とする。

2. 使用データと丸め誤差発生確率モデルの構築

本研究では、2012 年に熊本都市圏で実施されたパーソントリップ調査のデータを使用し、トリップの出発時刻に着目してトリップ数0およびトリップ出発時刻が不明のデータを除いた 240,006 トリップを本研究の分析対象とする。モデル構築にあたり、オランダで行われた郵送調査をもとに丸め誤差発生確率のモデリングを行った Rietveld¹⁾をもとに改良を行った。本研究では、図-1 の分布から丸められる値を 5,10,15,30,60 分単位とした。回答者の実際の出発時刻が m 分のとき、5 分単位での丸め誤差の発生確率を $p_{m,5}$ とすると、パラメータ a_5, b_5 を用いて丸め誤差発生確率を以下のように定義する。

$$p_{m,5} = a_5 + b_5 \cdot d_{m,5} \quad d_{m,5} = 1,2 \tag{1}$$

$d_{m,5}$ は m から最も近い 5 分単位の時刻までの差を表す。10 ～60 分単位も同様に定義するが、先行研究 ²⁾の結果より $b_{10}, \dots, b_{60} = 0$ とする。回答者が真に m 分に出発した確率を g_m とおくと、例えば m 分にもっとも近い 5 分単位の時刻に丸められる確率は、丸め誤差発生確率を用いて $g_m \cdot p_{m,5}$ と表すことができる。回答者が m 分と回答する確率を q_m とおくと、 m が丸められうる値でないときは式(2)、 m が丸められうる値のとき、例えば $m = 15$ のときは、式(3)のように表される。

$$q_m = g_m \cdot [1 - p_{m,5} - p_{m,10} - p_{m,15} - p_{m,30} - p_{m,60}] \tag{2}$$

$$q_{15} = [g_8 \cdot p_{8,15} + \dots + g_{22} \cdot p_{22,15}] + [g_{13} \cdot p_{13,5} + \dots + g_{17} \cdot p_{17,5}] + g_{15} \cdot [1 - p_{15,10} - p_{15,30} - p_{15,60}] \tag{3}$$

ここで、トリップの発生は一様に分布しているとの仮定をおき、 $g_m = 1/60$ とする。式(2),(3)より同時生起確率を定義し、対数尤度関数を用いて最尤法によりパラメータ推定を行った。推定結果を表-1 に示す。5 分単位の値 (a_5) が最も大きい一方、15 分単位の値 (a_{15}) が他と比べて小さい。これは 15 分単位での丸め誤差発生確

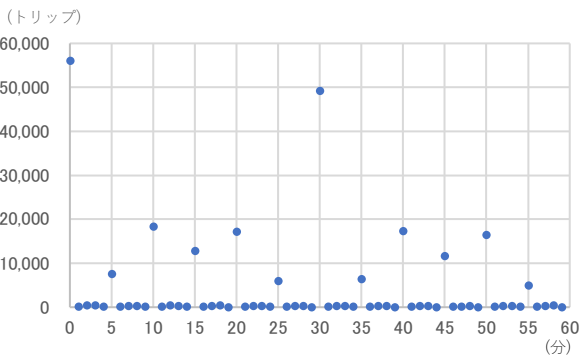


図-1 トリップ出発時刻の分布（熊本 PT 調査）

表-1 丸め誤差発生確率モデル推定結果

係数	最尤推定値	標準誤差	係数	最尤推定値	標準誤差
a_5	0.375	0.0027	b_5	-0.0511	0.0011
a_{10}	0.278	0.0023	b_{10}	—	—
a_{15}	0.101	0.0018	b_{15}	—	—
a_{30}	0.216	0.0022	b_{30}	—	—
a_{60}	0.028	0.0013	b_{60}	—	—
対数尤度(LL)			-606280		
初期対数尤度(L0)			-982667		

キーワード 丸め誤差，パーソントリップ調査，出発時刻，トリップ手段別，トリップ目的別

連絡先 〒860-8555 熊本市中央区黒髪 2-39-1 くまもと水循環・減災研究教育センター TEL 096-342-3601

率が低いことを意味しており、Rietveld の結果とは大きく違う傾向を示している点である。

3. 実際の出発時刻の分布とトリップ目的別・手段別回答誤差の推定

回答者が m 分と回答したとき、実際の出発時刻が n 分である確率を $P(n|m)$ とすると、以下のように表される。

$$P(n|m) = \frac{P(m|n)P(n)}{P(m)}$$

(4)

$P(n)$ は実際の出発した時刻が n 分である確率であり、2.での仮定より $P(n) = 1/60$ である。 $P(m|n)$ は丸め誤差発生確率、 $P(m)$ は回答者が m 分と回答する確率なので、すでに定式化されている。つまり、モデル推定結果を用いれば、 $P(n|m)$ は計算が可能であり、実際の出発時刻の分布を推定することができる。2.で示したトリップ出発時刻の分布より、30分単位での回答が多いため、本研究では $m = 30$ として分析を行う。推定の結果、実際の出発時刻が30分である確率をもっとも高いが、1割にも満たないことが図-2の分布から確認できる。この分布から、丸め誤差が発生しうる場合の回答誤差を定量的に評価することが可能となる。

トリップ目的別および代表交通手段別でそれぞれパラメータ推定を行い、実際の出発時刻の累積確率を表-2、表-3に示す。目的別では、業務目的のトリップで丸め誤差が発生していない確率および5分、10分以内の区間に実際の出発時刻が含まれる確率ともに最も低い。5分以内の区間では、他の目的と比べておよそ5%程度低く推定された。業務目的では他目的と比べ00分、30分と回答されたトリップの割合が多く、これが分布の裾が広い背景となっていることが考えられる。手段別では、公共交通の区分において丸められない確率および5分以内、10分以内の区間に実際の出発時刻が含まれると推定される確率が他目的よりやや高い。公共交通では00分、30分単位での回答が他の手段より少なく、普段利用する公共交通機関の出発時刻が回答の目安となっていることが示唆される。また、タクシーにおいて分布の裾が広く推定されている。推定された一部のパラメータが統計的に有意ではないので妥当性について留意が必要であるが、公共交通または自家用車とは異なる傾向を示していることは興味深い。

4. おわりに

本研究では、熊本PT調査でのトリップ出発時刻の回答誤差の分布に関する推定を行った。具体的には、30分と回答されたトリップに着目し、丸められず回答したと推定されるのは1割に満たないことを明らかにした。また、目的別では業務目的で回答誤差が大きく、代表交通手段別では、公共交通を利用したトリップが他の手段に比べて回答誤差がやや小さい傾向にあることを確認した。今後の展望としては、例えば性別・職業等の基礎属性や紙/Webの回答方式別でも推定を行い、丸め誤差発生に関する俯瞰的な評価を行いたい。

参考文献

1) Rietveld P : Rounding of Arrival and Departure Times in Travel Surveys: An Interpretation in Terms of Scheduled Activities, *Journal of Transportation Statistics*, Vol.5, Issue 1, pp. 71–82, 2001.

2) 佐藤嘉洋, 円山琢也 : 交通調査の出発時刻の丸め誤差に関するモデル分析, 第 59 回土木計画学研究発表会, 投稿中, 2019.

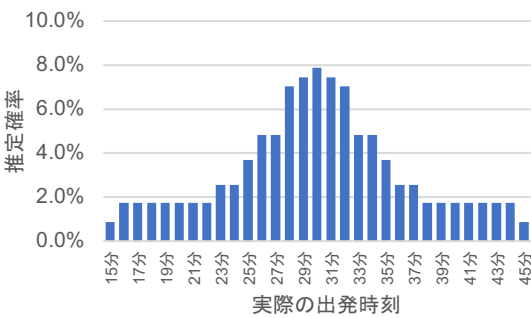


図-2 実際の出発時刻の分布（回答：30分）

表-2 目的別 実際の出発時刻の累積確率

誤差 実際の出発時刻		なし 30分	5分以内 25分～35分	10分以内 20分～40分
推定 確率	通勤	8.0%	63.5%	85.7%
	通学	9.3%	72.5%	88.8%
	帰宅	8.3%	64.5%	84.7%
	私用	7.6%	61.8%	83.3%
	業務	6.4%	58.0%	81.5%

表-3 代表交通手段別 実際の出発時刻の累積確率

誤差 実際の出発時刻		なし 30分	5分以内 25分～35分	10分以内 20分～40分
推定 確率	徒歩等	8.0%	64.2%	84.6%
	自家用車	7.7%	63.2%	84.1%
	貨物車	7.9%	63.3%	84.1%
	公共交通	8.7%	65.4%	85.2%
	タクシー	6.6%	57.8%	81.1%

* 厳密には、実際の出発時刻が30.0000分である確率は限りなく0に近いが、PT調査は1分単位での回答であるため、ここでは30分と記載する。