

交通調査における出発時刻丸め回答の 選択モデル分析：個人属性と調査手法の影響

佐藤 嘉洋¹・円山 琢也²

¹学生会員 熊本大学大学院自然科学教育部工学専攻（〒860-8555 熊本市中央区黒髪2-39-1）

E-mail:yo-sato@kumamoto-u.ac.jp

²正会員 熊本大学准教授 くまもと水循環・減災研究教育センター（〒860-8555 熊本市中央区黒髪2-39-1）

E-mail: takumaru@kumamoto-u.ac.jp

パーソントリップ調査などの交通調査では、トリップの出発・到着時刻を繰り上げまたは繰り下げて回答する「丸め誤差」によって、5,10,30分単位での回答が多いことが知られている。本研究では、個人やトリップの属性および調査手法に着目して丸め回答の選択行動を分析する。熊本都市圏パーソントリップ調査を利用した分析の結果、個人属性では男性、年齢が高齢または20代、就業者のときに、トリップ属性では業務トリップのときに30分単位で回答される傾向にあることを明らかにした。また、トリップ出発時刻の丸め回答の選択行動を順序プロビットモデルで分析し、到着時刻から受ける影響が認められること、調査手法によって丸め回答に影響を与える変数が異なることなどを示した。

Key Words: rounding error, travel survey, departure time

1. 研究の背景と目的

社会調査において、得られた回答には少なからず誤差が存在しており、その誤差が結果に与える影響が問題となっている。誤差は大きく標本誤差と非標本誤差の2つに分けられるが、パーソントリップ調査(以下、PT調査)をはじめとする交通調査においては、回答されたトリップ出発、到着時刻^{1,2}に非標本誤差が見られることが多く、実際の時刻をある値に繰り上げまたは繰り下げ、切りの良い数字として回答する「丸め誤差」と呼ばれる非標本誤差が起きやすいことが知られている^{3,4}。これら回答時刻の丸め誤差は、その後の分析に与える影響が懸念される。例えば、回答時刻から旅行時間が算出される場合は交通手段選択モデルの推定結果や、旅行時間が別の回答として取得されていた場合でも時間帯別交通需要量への影響である⁵。この影響を鑑みれば、回答時刻の信頼性に関する精査は重要な課題と言える。

本研究の実施にあたり、先行研究では、熊本都市圏PT調査において回答された出発時刻の分布を示し⁶、欧米の調査と回答の傾向が異なることなどの知見を得た^{7,8}。ただし、先行研究では主にトリップ出発時刻のみに着目した集計的な分析に留まっており、回答者の属性やトリップの特徴、調査手法が回答時刻の分布に与える影響を俯瞰的には示せていない。そこで本研究では、

- ・回答者の個人属性、トリップの特徴および調査手法に着目し、トリップ出発時刻の分布を比較すること
- ・トリップ出発時刻の回答傾向に関するモデルを推定し、属性および調査手法とトリップ出発時刻の回答傾向との関係を明らかにすること

以上を目的とする。分析方法として、既存研究で利用したRieveld⁹によるモデルの分析結果を援用した順序プロビットモデルを利用する。

丸め誤差は回答時刻のみならず、移動距離¹⁰、年収¹¹、年齢¹²、ミクロ経済学分野でのWTP (Willingness to Pay)¹³、HIV感染リスクを考慮した性交渉人数¹⁴など多くの調査または調査項目で見られ、heaping modelと呼ばれる丸め誤差を補正する手法も提案されている。ただし、旅行時間ではなく回答されたトリップの時刻に着目した研究例は殆ど行われておらず、調査手法の違いの影響の分析事例も筆者の知りうる限り存在しない。この点に本研究の新規性・独自性がある。

本稿の構成は以下のとおりである。2章で使用データを説明し、分析で使用する丸め回答の区分を定義する。続いて3章では、回答者の属性およびトリップ属性別に丸め回答の分析を行う。さらに4章では、丸め誤差発生確率モデル⁸を用いて出発時刻が到着時刻から受ける影響を定式化し、これを考慮した出発時刻の回答傾向に関する順序プロビットモデルを構築、推定する。最後に、

5章で本研究のまとめと今後の展望について整理する。

2. 調査概要と使用データ

(1) 使用するデータと回答区分の設定

本研究では、2012年熊本都市圏PT調査(表-1)のデータを使用する。調査は紙面調査とWeb調査を併用して行われた。記録されている約27万トリップのうち、分析に使用する変数に欠損および不明が存在しない227,690トリップを本研究の分析対象とする。そのうち、紙面調査での回答が205,283トリップ、Web調査での回答が22,407トリップとなった。

図-1に調査手法別のトリップ出発時刻の分布を示した。全体では、調査手法間での大きな偏りは見られない。分布が示すとおり、トリップ出発時刻は5分ごとの回答が多く見られ、回答の多くが丸められて回答されている可能性が示されている。分析対象とするトリップ数は多少異なるが、先行研究⁷⁾にて同調査のトリップ出発時刻の丸め誤差に関するモデル分析を行っている。分析の結果、出発時刻は主に5分、10分、30分単位で丸められて回答する傾向にあること、欧米では15分単位での丸め誤差が発生しやすい¹⁶⁾が、それに対して熊本PTでは15分単位での丸め誤差発生確率が低いことなどが示されている。よって本研究では、先行研究の結果にならない表-2のように4つの回答区分を設定する。区分が下にいくほど、大きく丸められていることを示している。しかし、これらの回答の中には実際にその時刻に出発した回答も含まれているが、出発時刻の真値は把握できないため、本稿ではこれら丸め回答を「丸められて回答されたものを含む回答」と定義する。以降は表-2の4つの区分を「丸めなし」「5分単位」「10分単位」「30分単位」と表記する。

表-3には4区分のトリップ出発時刻の回答構成比を示した。丸めなしおよび5分単位でWeb調査の回答構成比がわずかに低く、10分または30分単位で紙面調査の回答構成比が高い。紙面調査の方が大きく丸められて回答され、Web調査では丸めなしまたは5分単位の小さい丸め回答が多い傾向にある。ただし、この回答傾向は調査票やWeb入力フォームのデザインなど調査設計に大きく影響されるため、この結果は熊本PT調査の事例であり、一般的な傾向とは限らないことに注意されたい。

表-1 2012年熊本都市圏PT調査概要¹⁵⁾

調査期間	2012年10月～11月
調査方法	郵送配布・郵送回収での紙面調査 およびWeb調査(併用)
調査対象	5市6町1村
調査回答者	97,109人
回答総トリップ数	271,143トリップ (0トリップは1トリップ分として集計) うち紙面調査: 245,945トリップ, Web調査: 25,198トリップ
本研究におけるサンプルサイズ	分析に使用する変数すべてに欠損・不明がない 227,690トリップ うち紙面調査: 205,283トリップ, Web調査: 22,407トリップ

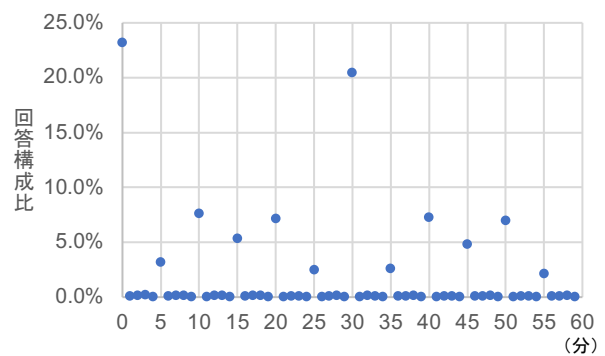


図-1(a) トリップ出発時刻の分布(紙面調査)

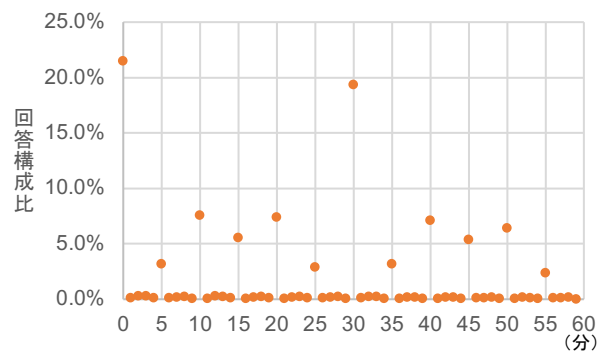


図-1(b) トリップ出発時刻の分布(Web調査)

表-2 トリップ出発時刻の回答区分

区分	対象出発時刻(分)
丸めなし回答 (丸め誤差の可能性がない回答)	下記3つに該当しない場合
5分単位での丸め回答 (5分単位で丸められた可能性がある回答)	5分/15分/25分/35分/45分/55分
10分単位での丸め回答 (10分単位で丸められた可能性がある回答)	10分/20分/40分/50分
30分単位での丸め回答 (30分単位で丸められた可能性がある回答)	00分/30分

(2) 調査区分/基礎属性別の基礎分析

ここでは、前節で設定した回答区分を用いて、その回答構成比の分布を分析していく。

a) 回答者の属性

表-4は回答者の属性別に回答区分の分布を整理した。男女別では、女性に比べて男性が紙面調査、Web調査のどちらも30分単位での構成比が高く、それ以外の構成比は低い。男性の方が大きく丸めて回答する傾向にあることが示唆されている。

年代別では、どちらの調査方式でも年代が高くなるほど大きく丸められた回答に偏る傾向がある。ただし、20代はその周辺の年代と比べて30分単位での回答に偏っており、大きく丸められて回答する傾向が比較的高い年代といえる。調査手法別では、若年層では全体の傾向と大きな違いは見られない一方、70代以上の高齢者ではWeb調査での30分単位での回答が紙面調査のそれを上回っており、他年代と異なる傾向を示す点は興味深い。当該区分のサンプルサイズが小さいことには留意が必要であるが、Web調査で回答する高齢者の割合はさほど多くないと考えられるため、同居家族が代理回答¹⁶⁾している可能性が考えられる。例えば午前中に買い物へ行った程度の情報など、高齢者の正確な行動を把握できていない場合は、回答しやすい30分単位で入力を行った可能性がある。

就業別では、全体と傾向はほぼ変わらないが、就業者の30分単位での回答で非就業者に比べて紙面調査とWeb調査の差が大きい。後述するトリップ目的にも関係すると思われるが、就業者が紙調査で回答を行う場合、30分単位で大きく丸めて回答する可能性が高くなることが考えられる。

b) トリップの属性

表-5では回答されたトリップの属性別に回答の傾向を整理している。まずトリップ目的別では、どちらの調査手法も業務区分でもっとも30分単位の回答構成比が高く、丸めなしと5分単位を合わせた回答が5区分中もっとも低い。手法別では、通学目的において30分単位での構成比の差が他目的と逆転しており、差がほとんどないことが特徴である。加えて、他目的に比べ通学目的は10分単位と30分単位の構成比に大きな差がないことは興味深い。通学範囲など地域特性に大きく左右されると思われるが、大都市に比べて熊本都市圏は比較的通学範囲が狭く、交通手段も徒歩または自転車であることが多い。例えば自動車で移動することの多い通勤や業務トリップと比較して短時間で移動している場合、10分単位での回答の割合が高くなったものと考えられる。

次に代表交通手段別では、他の手段と比べ公共交通で30分単位での回答構成比がもっとも多く、丸めなしでの回答の構成比が突出して高い。ただし5分単位、10分単位での回答構成比は他の手段と大きく変わらないことか

表-3 調査手法別 トリップ出発時刻の回答構成比

(単位：パーセント, Nはトリップ数)

区分	丸めなし	5分単位	10分単位	30分単位
紙面(N=94,310)	62	20.8	29.2	43.8
Web(N=10,985)	7.8	22.7	28.6	40.9
構成比の差(紙面-Web)	-1.6	-1.9	0.6	2.8

表-4 個人属性別 回答構成比

(単位：パーセント, Nはトリップ数)

区分	丸めなし	5分単位	10分単位	30分単位	
性別					
男性	紙面(N=94,310)	5.5	19.4	28.5	46.5
	Web(N=10,985)	6.9	21.1	28.1	43.9
女性	紙面(N=110,973)	6.8	22.1	29.8	41.4
	Web(N=11,422)	8.6	24.3	29.0	38.1
年代別					
10代以下	紙面(N=26,562)	4.8	23.8	32.1	39.3
	Web(N=4,634)	6.4	24.8	30.5	38.3
20代	紙面(N=15,886)	5.2	18.6	26.9	49.4
	Web(N=2,091)	7.2	19.9	29.2	43.8
30代	紙面(N=27,754)	5.3	21.7	28.8	44.2
	Web(N=4,560)	8.3	23.2	27.3	41.2
40代	紙面(N=29,012)	5.8	21.7	29.3	43.2
	Web(N=5,152)	7.2	22.9	27.6	42.3
50代	紙面(N=33,662)	6.0	20.9	29.3	43.8
	Web(N=3,290)	8.3	21.6	29.3	40.8
60代	紙面(N=40,888)	7.1	20.1	29.0	43.8
	Web(N=1,877)	11.2	22.7	28.9	37.1
70代	紙面(N=23,125)	8.1	19.3	28.5	44.1
	Web(N=545)	9.4	16.9	25.9	47.9
80代以上	紙面(N=8,394)	8.4	17.8	27.6	46.2
	Web(N=258)	4.3	22.1	24.4	49.2
就業状態					
就業	紙面(N=113,124)	5.5	20.4	29.0	45.2
	Web(N=13,305)	7.6	22.5	28.5	41.5
非就業	紙面(N=92,159)	7.1	21.4	29.5	42.0
	Web(N=9,102)	8.0	23.1	28.8	40.2

ら、公共交通を使用するトリップは5分または10分の比較的小さく丸めて回答される可能性はあるものの、30分単位で大きく丸めて回答されることは少ない傾向にあることがわかる。先行研究⁸⁾でも同様の分析を行っているが、公共交通を使用する場合は、例えば毎朝7時15分のバスに乗るので5分前の7時10分に家を出発するなど、利用する交通機関の時刻表が回答の目安となり、自身の行動をある程度正確に記憶していることが理由として考えられる。調査手法別では全体の傾向と大きく変わりはないが、貨物車を利用したトリップにおいて、紙面調査とWeb調査の30分単位での回答構成比の差が他と比べて大きい。

次に、「1日に行う行動の回数が多いほど、記憶が曖昧になり丸めて回答する傾向が強くなる」との仮説を立て、1日の行動順、つまりトリップ番号に着目して回答の傾向を分析する。表-5に結果を示しているが、1日の行動順が後になるほど、つまりトリップ番号が大きくなるほど小さく丸めて回答する傾向となり、仮説とは逆の結果となったことは興味深い。カイ二乗検定(独立性の検定)を行うと、どちらの調査手法も有意差が認められ($P<0.001$)、この結果からは仮説は支持されないとと言える。このような結果となった要因の一つとして、回答するトリップ数が多い回答者は自身の行動を正確に答えようとしている、つまり回答者の性格(几帳面さ)が現れている可能性も考えられる。調査手法間では、全体の傾向と特筆すべき差は見られなかった。

最後に、トリップの出発時間帯に着目して分析を行う。回答されたトリップは丸められて回答している可能性が考えられるため参考程度にはなるが、ピーク時間帯を朝：7時～9時、夕方：17時～19時と設定し、その前後のオフピーク時間を時間が早い方からそれぞれ早朝/昼/夜間オフピーク時間帯とし、5つに分けて回答傾向を整理した。表-5より、30分単位での回答構成比に着目すると朝ピーク～夕方ピークの時間帯は時間が遅くなるにしたがって回答構成比が上昇していることがわかる。加えて、丸めなし～5分単位での時間帯ごとの回答傾向に大きな差はないが、10分単位での回答では朝ピーク～昼オフピークに比べて夕方ピーク時間帯の構成比が低い傾向にある。朝ピークは通勤/通学、夕方ピークは帰宅時間帯に相当するため、目的別の通勤、帰宅の関係と同様の傾向を示していると考えられる。

また、6時台までの早朝オフピーク時間帯では、30分単位での回答構成比は5時間帯中もっとも少ないが、丸めなし回答の構成比は他時間帯と特筆すべき差はなく、10分単位の回答構成比が高くなっていることが特徴である。交通量が少ないため、移動時間も比較的短時間で済むことが多く、丸められて回答されたとしても比較的小さく丸められることが考えられる。

表-5 トリップの属性別 回答構成比

(単位：パーセント, Nはトリップ数)

区分		丸めなし	5分単位	10分単位	30分単位
目的					
通勤	紙面(N=35,767)	4.6	21.8	31.0	42.6
	Web(N=4,718)	6.7	24.8	30.7	37.8
通学	紙面(N=11,822)	4.9	25.3	34.6	35.1
	Web(N=2,203)	6.4	26.4	31.8	35.5
帰宅	紙面(N=76,019)	7.3	21.6	28.9	42.2
	Web(N=8,216)	8.2	22.3	28.0	41.5
私用	紙面(N=69,205)	6.3	19.5	28.2	46.0
	Web(N=5,890)	8.5	21.2	27.4	43.0
業務	紙面(N=12,470)	4.7	16.9	26.2	52.2
	Web(N=1,380)	7.8	18.8	24.9	48.5
代表交通手段					
徒歩 自転車 バイク	紙面(N=63,709)	5.8	21.8	29.5	42.9
	Web(N=7,254)	7.2	23.4	29.4	39.9
自動車	紙面(N=113,993)	5.9	20.8	29.4	43.8
	Web(N=12,767)	7.5	22.6	28.2	41.8
貨物車	紙面(N=12,855)	5.5	17.1	26.3	51.1
	Web(N=907)	8.9	19.5	28.9	42.7
公共交通	紙面(N=13,327)	11.3	20.6	28.8	39.3
	Web(N=1,414)	12.2	22.3	27.9	37.6
タクシー	紙面(N=1,399)	4.6	16.7	28.4	50.3
	Web(N=65)	7.7	23.1	26.2	43.1
トリップ番号					
トリップ 1	紙面(N=71,147)	4.2	19.8	29.4	46.6
	Web(N=8,402)	6.0	22.9	29.0	42.1
トリップ 2	紙面(N=64,481)	5.8	20.2	28.6	45.4
	Web(N=7,059)	7.0	21.5	27.3	44.2
トリップ 3	紙面(N=30,361)	7.1	21.2	29.6	42.1
	Web(N=2,857)	8.7	22.4	29.5	39.4
トリップ 4	紙面(N=19,526)	8.5	23.3	29.5	38.7
	Web(N=1,934)	11.3	23.5	29.7	35.4
トリップ 5以降	紙面(N=19,768)	11.0	23.7	29.6	35.6
	Web(N=2,155)	12.8	25.8	28.8	32.7
出発時間帯					
早朝 オフピーク	紙面(N=8,644)	5.1	22.6	35.3	37.0
	Web(N=905)	6.9	25.9	35.9	31.4
朝 ピーク	紙面(N=44,473)	4.6	22.2	30.1	43.1
	Web(N=6,153)	6.5	25.1	29.3	39.1
昼 オフピーク	紙面(N=95,982)	7.1	20.5	29.3	43.1
	Web(N=8,169)	8.3	21.4	28.1	42.2
夕方 ピーク	紙面(N=36,936)	5.6	19.4	26.6	48.4
	Web(N=4,392)	8.1	20.5	26.5	44.9
夜間 オフピーク	紙面(N=19,248)	6.9	21.6	28.9	42.7
	Web(N=2,788)	8.6	23.9	29.3	38.2

4. 順序プロビットモデルを用いた出発時刻回答傾向モデルの推定

(1) モデルの概要

本章では、3.で分析を行った結果を用いて、順序プロビットモデルを構築および推定してトリップ出発時刻の回答傾向を把握する。

順序プロビットモデル(順序ロジットモデル)は、被説明変数のとりうる選択肢が3つ以上あり、その選択肢に順序がある場合に用いられるモデルである。本稿では、2.で示した回答区分を被説明変数としてモデルを構築する。ただし、丸められず回答されたトリップは回答数が少ないため、被説明変数は「丸められないまたは5分単位で回答される($y = 1$)」、「10分単位で回答される($y = 2$)」、「30分単位で回答される($y = 3$)」の3肢とする。回答者*i*がトリップ*j*の出発時刻を回答する際に選ぶ選択肢を y_{ij} とすると、 y_{ij} が大きくなるほどより大きく丸めて回答する傾向にあると言える。

このとき、潜在変数を y_i^* を次式のように表す。

$$y_{ij}^* = \beta x_{ij} + \varepsilon_{ij} \quad (1)$$

式(1)のうち x_{ij} は説明変数で構成されるベクトルであり、 β は本モデルで推定するパラメータである。 ε_{ij} は正規分布にしたがう確率項である。このとき、観測される選択肢 y_{ij} は潜在変数を用いて次のように表される。

$$y_i = \begin{cases} 3 & \text{if } y_{ij}^* > \theta_2 \\ 2 & \text{if } \theta_1 < y_{ij}^* \leq \theta_2 \\ 1 & \text{if } y_{ij}^* \leq \theta_1 \end{cases} \quad (2)$$

ここで、 θ_1 および θ_2 はそれぞれの選択肢間の閾値である。3.での分析結果を踏まえ、説明変数を表-6のように設定した。

(2) 到着時刻から受ける影響の考慮

a) 到着時刻から受ける影響の定式化

3.までは、出発時刻のみに着目した分析を行ったが、トリップ到着時刻の影響は無視できないといえる。例を挙げれば、10時00分に出発し、10時05分に到着したトリップがあったとすると、出発時刻の00分だけで考えると丸めて回答していない~60分単位で丸めて回答した可能性のいずれも考えられるが、到着時刻を考慮する、つまり旅行時間が5分であることを考えると、大きく丸めて回答した可能性は少なく、丸めて回答したとしても小さな単位で可能性の方が大きいと考えるのがより自然であろう。ここで、「トリップの回答を行うとき、旅行時間はある程度覚えている」との仮定を置き、出発時刻の回答の際に到着時刻(または旅行時間)から受ける影響について、以下の2つの仮説を設定する。

表-6 モデル推定に用いた変数と基本統計量

変数名	説明	紙面	Web
		平均 (SD)	平均 (SD)
回答者の属性			
女性	女性1, 男性0のダミー変数	0.541 (0.498)	0.510 (0.500)
就業者	就業者: 1, 非就業者/学生/未就学児: 0のダミー変数	0.551 (0.497)	0.594 (0.491)
年齢	対数変換した回答者の5歳階級年齢	3.662 (0.675)	3.397 (0.709)
20代	回答者が20代: 0, それ以外: 0のダミー変数	0.213 (0.409)	0.297 (0.457)
トリップの属性			
通学	トリップ目的が通学: 0, それ以外: 0のダミー変数	0.058 (0.233)	0.098 (0.298)
業務	トリップ目的が業務: 1, それ以外: 0のダミー変数	0.061 (0.239)	0.062 (0.240)
公共交通	代表交通手段が公共交通: 1, それ以外: 0のダミー変数	0.065 (0.246)	0.063 (0.243)
貨物車	代表交通手段が貨物車: 1, それ以外: 0のダミー変数	0.063 (0.242)	0.040 (0.197)
トリップ番号	回答日における行動順	2.372 (1.548)	2.361 (1.782)
朝ピーク	出発時刻が朝ピーク(7:00~9:00): 1, それ以外: 0のダミー変数	0.217 (0.412)	0.275 (0.446)
夕方ピーク	出発時刻が夕方ピーク(17:00~19:00): 1, それ以外: 0のダミー変数	0.180 (0.384)	0.196 (0.397)

- (i) 旅行時間が長いほど、出発時刻の記憶はより曖昧になり、丸めた値で回答されやすい
- (ii) 到着時刻が丸められている可能性が高い時刻で答えられているほど、出発時刻も丸めた値で回答されやすい

これら2つの仮説に基づいてモデルを推定するため、式(1)に以下のような説明変数の導入を検討する。回答者*i*がトリップ*j*の出発時刻 $t_{ijstart}$ を回答するとき、到着時刻 t_{ijend} から受ける効用 V_{ij} を以下のように定義する。

$$z_{ij} = y_{ij} \frac{w_{jend,m}}{|t_{ijend} - t_{ijstart}|} \quad (3)$$

$w_{jend,m}$ は回答された到着時刻 t_{ijend} の「 m 分」の確からしさを表す指標とする。確からしさとは、本稿では「実際に m 分に出発し、 m 分と回答したと推定される確率」と定義する。

式(3)を式(1)に導入し、以下とする。式(2)の y_{ij}^* も \tilde{y}_{ij}^* に読み替える。 y_{ij} が大きいほどより大きい値で丸めて回答することを意味するため、 z_{ij} は負の値を取ることが望ましいと考えられる。

$$\tilde{y}_{ij}^* = \beta x_{ij} + z_{ij} + \varepsilon_{ij} \quad (4)$$

$w_{jend,m}$ を推定するにあたって、先行研究で行った丸め誤差発生確率モデルを用いる。本モデルはRietveld⁹⁾が開発し、筆者らが熊本PTでの丸め誤差発生確率をより良好に推定できるために改良を行ったものである。本稿

の範囲を外れるため詳細は文献^{7,9)}を参照して頂きたいが、以下簡潔にモデルの構造を紹介する。

b) 丸め誤差発生確率モデル

丸め誤差の発生は5分、10分、15分、30分、60分単位のいずれかで起こるとして、回答者の実際の到着時刻が 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。

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

ここで、 $d_{m,5}$ は m から最も近い5分単位の時刻までの差を表す。他の丸められる値については一定の確率であるとした。つまり、次の式(6)~式(9)で表される。

$$p_{m,10} = a_{10} \quad (6)$$

$$p_{m,15} = a_{15} \quad (7)$$

$$p_{m,30} = a_{30} \quad (8)$$

$$p_{m,60} = a_{60} \quad (9)$$

回答者が真に m 分に到着した確率を g_m をとおくと、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。ここで、回答者が m 分と回答する確率を q_m とおくと、 m が丸められる値でないときは式(10)、 m が丸められる値のとき、例えば $m=15$ のときは、式(11)のように記述できる。

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

$$q_{15} = [g_8 \cdot p_{8,15} + \dots + g_{22} \cdot p_{22,15}] + [g_{13} \cdot p_{13,5} + \dots + g_{17} \cdot p_{17,5}] + g_{15} \cdot [1 - p_{15,10} - p_{15,30} - p_{15,60}] \quad (11)$$

式(10)の第2項はどの値にも丸められない確率を示している。式(11)の第1項および第2項は周辺の時刻で実際に到着し、かつ15分丸められて回答された確率の和を表している。回答者が実際に m 分に到着した確率 g_m は、「トリップの到着時刻の分は一様に分布している」という仮定を置く。つまり、 $g_m = 1/60$ である。

このとき、式(10)および式(11)を用いて、 N_m を到着時刻が m 分と回答されたトリップの総数とすると、同時生起確率 L は以下のように表される。

$$L = \prod_{m=0}^{59} q_m^{N_m} \quad (12)$$

q_m およびその中に含まれる $p_{m,5}, \dots, p_{m,60}$ は確率変数であるため、以下の条件を付加した制約条件つき尤度最大化問題としてパラメータを推定する。

$$0 \leq q_m, p_{m,5}, p_{m,10}, p_{m,15}, p_{m,30}, p_{m,60} \leq 1 \quad (13)$$

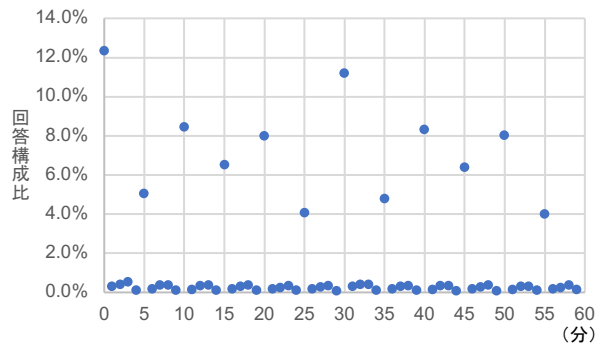


図-2(a) トリップ到着時刻の分布 (紙面調査)

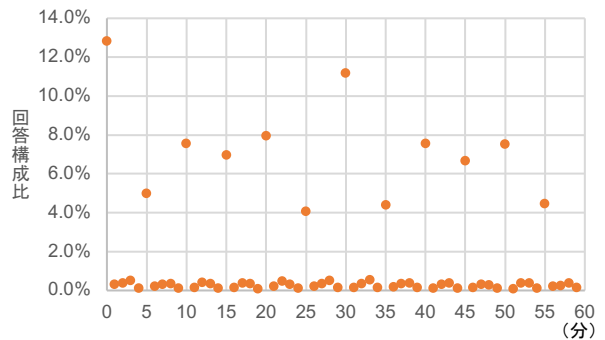


図-2(b) トリップ到着時刻の分布 (Web 調査)

表-7 到着時刻の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	0.687	0.0035	b_5	-0.120	0.0018
a_{10}	0.220	0.0022			
a_{15}	0.080	0.0017			
a_{30}	0.021	0.0018			
a_{60}	0.012	0.0010			
サンプルサイズ	227,690				
対数尤度 (LL)	-679,543				
初期対数尤度 (L0)	-932,241				

c) 使用データと丸め誤差発生確率モデル推定結果

先行研究⁷⁾では出発時刻の「分」を用いてパラメータを推定したが、到着時刻の確からしさである $w_{jend,m}$ を求めるため、本稿では到着時刻の「分」を用いて同じモデルによるパラメータ推定を行う。図-2に、熊本PT調査におけるトリップ到着時刻の分布を示す。2.で紹介したトリップ出発時刻の分布と比較して、00分および30分単位での回答構成比が大きく減少している。例えば通勤/通学トリップは、始業の5分前にいつも到着するようにしているなど、目安となる時間が存在するため、出発時刻に比べて到着時刻は大きな値で丸められにくいことが考えられる。手法間では、10分単位および15分単位での構成比に若干の差があるが、分布に大きな違いはない。

詳細は後述するが、紙面調査とWeb調査に分けてモデルを推定するため、式(3)において紙面調査とWeb調査それぞれの $w_{jend,m}$ を用いるとパラメータ γ_{ij} がこの影響を受け、モデル推定結果の比較が困難となるため、 $w_{jend,m}$ の推定にあたっては調査方法の別を問わず、すべての到着時刻のデータを用いてパラメータ推定を行った。

丸め誤差発生確率モデルの推定結果を表-7に示す。出発時刻の推定⁸⁾で高かった30分単位での丸め誤差発生確率 a_{30} は到着時刻の推定では低くなり、5分単位および10分単位での丸め誤差発生が支配的となる結果を示した。

d) 到着時刻の「分」の確からしさの推定

前項の結果をもとに、 $w_{jend,m}$ を推定していく。まず回答者が 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、ベイズの定理を用いて以下のように変形できる。

$$P(n|m) = \frac{P(m|n)P(n)}{P(m)} \quad (13)$$

右辺のうち、 $P(n)$ は実際の出発した時刻が n 分である確率であり、これはモデルでの家庭に基づき $P(n) = 1/60$ である。 $P(m|n)$ は実際の出発時刻が n 分であるとき、回答者が m 分と回答する確率なので、これは丸め誤差発生確率に他ならない。 $P(m)$ は回答者が m 分と回答する確率なので、すでに式(7)、式(8)で定式化されている。つまり、これまで定式化してきた確率と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。図-3は、例として $m = 30$ のときの実際の出発時刻の分布を示したものである。この図から分かるように、30分に出発したと回答されたトリップのうち、実際に30分に出発したトリップは14.7%程度しかないと考えられる。この推定された確率を本稿では確からしさ、つまり $w_{jend,m}$ と定義する。

図-3と同様に、他の分についても実際の到着時刻の分布を推定し、表-8に $w_{jend,m}$ の一覧を整理する。図-2の分布で説明すると、回答が集中しているところほど $w_{jend,m}$ は小さくなる。また、回答された m が丸められる値でないときは、丸めて回答されたとはみなされないため、5分単位以外の時刻は $w_{jend,m} = 1$ とする。

(3) 出発時刻丸め回答選択モデル推定結果

モデル推定結果を表-9に示す。本稿では(2)で述べた、到着時刻からの影響を考慮したモデル(a)と、考慮しないモデル(b)も比較のため推定を行った。モデル(a)とモデル(b)の比較では、紙面調査、Web調査ともにモデル(a)のAICの方が小さく、到着時刻から受ける影響を考慮することでより良好な結果が得られることが示されている。

順序プロビットモデルを用いた推定のため、パラメー

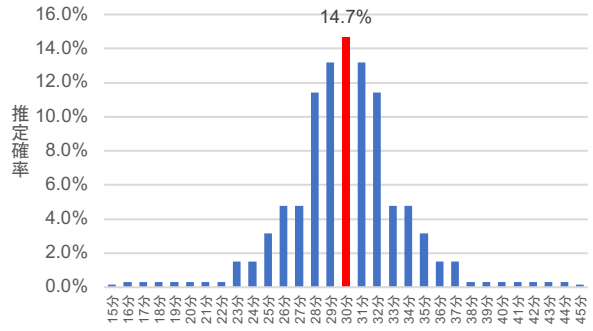


図-3 実際の到着時刻時刻の分布 (回答：hh時30分)

表-8 到着時刻の「分」の確からしさの推定

丸め誤差	回答された到着時刻の分(m)	$w_{jend,m}$
なし	下記以外	1.000
5分単位	5 / 25 / 35 / 55分	0.248
5/10分単位	10 / 20 / 40 / 50分	0.181
5/15分単位	15 / 45分	0.192
5/10/15/30分単位	30分	0.147
5/10/15/30/60分単位	00分	0.134

タは説明変数が被説明変数、つまり回答傾向へ与える影響を示している。正の値はより大きい値で丸められる傾向を、負の値はより小さい値で丸められる傾向を表していると言える。いずれの推定結果もパラメータの符号条件は3.での分析結果と一致しており、回答傾向を良好に示している。以下、より良好な結果が得られたモデル(a)に絞って考察する。

a) 同一調査手法内でのパラメータの比較

回答者の属性では、年齢を除く3つのダミー変数の中で性別がもっとも絶対値の大きい(負の)値を示しており、男性に比べて女性の方が丸めて回答する傾向が低いことを示している。トリップ属性別では、目的：業務トリップがもっとも絶対値の大きな値を示した。正の値を示していることから、3.でも述べたように、業務トリップはより大きな値で丸められる傾向にあると言える。到着時刻から受ける影響 γ_{ij} は負の値を示しており、(2)(a)で設定した仮説を概ね支持していると考えられる。

b) 調査手法間内でのパラメータの比較

紙面調査とWeb調査の推定結果を比較すると、紙面調査では就業状態のパラメータが有意とはならなかったが、Web調査では有意な結果が得られている。ただしどちらも絶対値は大きくなく、影響は限定的と言える。トリップの属性では、Web調査で代表交通手段：貨物車でパラメータが有意とはならなかった。表-5で示しているように対象のトリップ数が少ないことには留意が必要だが、今回の熊本PT調査のWeb調査回答者で貨物車を代表交通手段とするトリップが丸め回答行動に与える影響は検出されていない。

表-9 出発時刻丸め回答選択の順序プロビットモデル推定結果

説明変数	(a) 到着時刻から受ける影響を考慮				(b) 到着時刻から受ける影響を考慮しない				
	紙面調査		Web 調査		紙面調査		Web 調査		
	推定値	t 値	推定値	t 値	推定値	t 値	推定値	t 値	
回答者の個人属性									
女性	-0.108	-20.73 ***	-0.137	-8.725 ***	-0.112	-21.71 ***	-0.143	-9.12 ***	
就業者	-0.010	-1.72	-0.071	-3.542 ***	-0.011	-1.86	-0.063	-3.15 ***	
年齢 (対数変換)	0.047	12.17 ***	0.053	3.833 ***	0.046	9.75 ***	0.050	3.63 ***	
20 代	0.083	10.01 ***	0.054	3.005 ***	0.086	12.62 ***	0.056	3.17 ***	
トリップの属性									
目的：通学	-0.139	-10.03 ***	-0.084	-2.444 **	-0.133	-9.59 ***	-0.073	-2.15 **	
目的：業務	0.185	15.99 ***	0.186	5.422 ***	0.190	16.45 ***	0.191	5.56 ***	
代表交通手段：公共交通	-0.164	-15.95 ***	-0.154	-4.899 ***	-0.147	-14.29 ***	-0.132	-4.21 ***	
代表交通手段：貨物車	0.132	11.83 ***	0.025	0.621	0.132	11.89 ***	0.036	0.89	
トリップ番号	-0.091	-49.66 ***	-0.075	-15.071 ***	-0.094	-51.10 ***	-0.079	-15.97 ***	
出発時間帯：朝ピーク	-0.082	-11.07 ***	-0.115	-5.319 ***	-0.083	-11.17 ***	-0.117	-5.43 ***	
出発時間帯：夕方ピーク	0.184	25.80 ***	0.146	6.997 ***	0.186	26.16 ***	0.148	7.10 ***	
到着時刻から受ける影響									
γ_{ij}	-0.565	-24.67 ***	-0.811	-10.174 ***	-	-	-	-	
閾値									
θ_1 (1 2)	-0.720	-38.179 ***	-0.648	-12.910 ***	-0.707	-37.522 ***	-0.631	-12.595 ***	
θ_2 (2 3)	0.061	3.262 ***	0.104	2.068 **	0.073	3.870 ***	0.118	2.358 **	
サンプルサイズ	205,283		22,407		205,283		22,407		
AIC	436,197		48,142		436,801		48,244		
Log-Likelihood	-218,084		-24,057		-218,388		-24,109		

*** 1%水準有意, ** 5%水準有意

選択肢 1：丸めなしまたは 5 分単位の回答 選択肢 2：10 分単位の回答 選択肢 3：30 分単位の回答

朝ピーク時間帯：7 時～9 時，夕方ピーク時間帯：17 時～19 時

5. おわりに

本研究では、熊本都市圏PT調査のデータを用いて、個人およびトリップの属性、調査手法に着目してトリップ出発時刻の丸め回答の選択行動を分析した。結果を以下に示す。

- 1) 回答者の個人属性では、男性、年齢が高齢または20代、就業者のとき、30分単位の大きく丸めたと考えられる時刻での回答が多い傾向にある。
- 2) トリップの属性別では、目的が業務である、朝ピーク時間帯に比べて夕方ピーク時間帯のとき、より大きく丸めたと考えられる時刻で回答される傾向にある一方、代表交通手段が公共交通、目的が通学のとき小さく丸めたと考えられる時刻で回答される傾向にある。
- 3) トリップ出発時刻の丸め回答選択モデルを推定し、到着時刻から受ける影響を加味することで良好な推定結果が得られる知見を得た。

- 4) 推定の結果、個人属性では性別、トリップ属性別では目的：業務であることがそれぞれ回答傾向に与える影響が大きいことを示した。また、調査手法別の比較では、紙面調査における個人属性：就業者、Web調査における代表交通手段：貨物車など、回答に影響を与えない変数が調査手法によって異なることを示した。

今後の課題として、3)で示したように出発時刻と到着時刻の回答傾向は相互に作用していると考えられるため、4.(2)で紹介した丸め誤差発生確率モデルの改良が挙げられる。実際の出発時刻の分布の推定を目的とするとき、例えば Yamamoto *et al.*¹⁰⁾はトリップ移動距離の丸め誤差補正を試みているため、本手法を参考にしながら改良を行っていきたい。

謝辞：本研究は、JSPS 科研費 JP18H01561 の支援を受けた成果の一部です。

参考文献

- 1) Bricka S.G., Sen S., Paleti R., Bhat C.R. : An analysis of the factors influencing differences in survey-reported and GPS-recorded trips, *Transportation Research Part C*, Vol.21, pp.67-88, 2012.
- 2) Montini, L., S. Prost, J. Schrammel, N. Rieser-Schüssler, Axhausen K. W. : Comparison of Travel Diaries Generated from Smartphone Data and Dedicated GPS Devices, *Transportation Research Procedia*, Vol. 11, pp. 227-241, 2015.
- 3) Levinson D. M., Kumar A : A Multi-Modal Trip Distribution Model: Structure and Application, *Transportation Research Record*, Vol.1466, pp.124-131, 1994.
- 4) Murakami E., Wagner D. P. : Can Using Global Positioning System (GPS) Improve Trip Reporting?, *Transportation Research Part C*, Volume 7, pp.149-165, 1999.
- 5) Varotto S. F., Glerum A., Stathopoulos A., Bierlaire M., Longo G. : Mitigating the Impact of Errors in Travel Time Reporting on Mode Choice Modelling, *Journal of Transport Geography*, Volume 62, pp. 236-246, 2017.06.
- 6) 上原一輝, 佐藤嘉洋, 円山琢也 : 交通調査の出発・到着時刻の丸め誤差と代理回答バイアス, 第 59 回土木計画学研究発表会講演集, 2019.06.
- 7) 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講演集, 2019.06.
- 8) 佐藤嘉洋, 円山琢也 : 交通調査のトリップ出発時刻の丸め誤差: 都市間比較, 第 60 回土木計画学研究発表会講演集, 2019.11.
- 9) 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.
- 10) Yamamoto, T., Madre, J.L., de Lapparent, M., Collet, R.: A random heaping model of annual vehicle kilometres travelled considering heterogeneous approximation in reporting, *Transportation*, pp.1-19, 2018.
- 11) Drechsler, J., Kiesl, H. : Beat the heap: An imputation strategy for valid inferences from rounded income data, *Journal of Survey Statistics and Methodology*, Vol.4, Issue 1, pp.22-42, 2016.
- 12) Bar, H.Y., Lillard, D.R. : Accounting for heaping in retrospectively reported event data - a mixture-model approach. *Statistics in Medicine*, Vol. 31, pp.3347-3365, 2012.
- 13) Page, I.B., Lichtenberg, E., Saavoss, M. : Estimating Recreation Demand When Survey Responses are Rounded. In: 2015 AAEA & WAEA Joint Annual Meeting, July 26-28, San Francisco, California. Agricultural and Applied Economics Association, 2015.
- 14) Crawford, F.W., Weiss, R.E., Suchard, M.A. : Sex, lies and self-reported counts: Bayesian mixture models for heaping in longitudinal count data via birth-death processes. *The Annals of Applied Statistics*, Vol.9, pp.572-596, 2015.
- 15) 熊本都市圏総合交通計画協議会 : 熊本都市圏都市交通マスタープラン, 2016.
- 16) Kitamura, R: Time-of-day characteristics of travel: an analysis of 1990 NPTS data. In: Special Reports on Trip and Vehicle Attributes, 1990 NPTS Report Series, Publication No. FHWA-PL-95-033, US Department of Transportation, Federal Highway Administration, Washington, DC, February, pp 4-1-4-56, 1995.
- 17) Maruyama, T., Hosotani, K., and Kawano, T.: Inferring proxy response in household travel surveys with unknown completer using a group-based choice model, *Transportation*, in press, 2019.

(2020.3.8 受付)

MODELING THE ROUNDING CHOICE OF REPORTED DEPARTURE TIMES IN TRAVEL SURVEYS: THE DEMOGRAPHIC AND SURVEY MODE EFFECTS

Yoshihiro SATO and Takuya MARUYAMA