製表技術参考資料 29

ISSN　1348－7345
ISSN－L 1348－7329

質的変数に関する擬似ミクロデータの作成方法について

－平成 14 年就業構造基本調査データを用いた研究－

NSTAC
Working Paper No.29

平成 27 年 3 月

独立行政法人　統計センター

製表技術参考資料は、独立行政法人 統計センターの職員がその業務に関連して行った製表技術に関する研究の結果を紹介するためのものである。

ただし、本資料に示された見解は、執筆者の個人的見解である。

目　次

質的変数に関する擬似ミクロデータの作成方法について
－平成 14 年就業構造基本調査データを用いた研究－

滝澤 有美*、堀川 顕子*

要　　旨

平成 21 年の新統計法の全面施行に伴い、公的統計の二次的利用拡大に向けた匿名データの提供及びオーダーメード集計が開始された。平成 26 年 5 月 1 日時点において、匿名データは 7 調査の提供を、オーダーメード集計は 26 調査の集計サービスを行っている†。

新統計法に基づく二次的利用データの提供開始後、研究者や教育関係者等から、二次的利用促進のための統計教育・訓練用の擬似的なミクロデータの必要性が指摘され、また統計委員会の審議の場においても同様の指摘を受けた。これらの指摘に応じる形で、（独）統計センターでは擬似的なミクロデータ（以下「擬似ミクロデータ」と言う。）の作成方法に関する研究を開始した。

先行研究の秋山ほか(2012)では、平成 16 年全国消費実態調査を用いた高次元クロス集計表を作成し、集計表のセルごとに量的変数の統計量（平均、分散及び共分散）に基づく多変量正規乱数を発生させ、これを用いて擬似ミクロデータを作成した。この擬似ミクロデータは、平成 23 年 8 月より統計センターホームページで試行提供中である。

しかしながら、試行提供の対象が量的変数のみであったため、利用者から質的変数への対象拡大の要望があり、新たに質的変数（産業分類、職業分類等）に関する作成方法の研究を行うこととなった。そこで、平成 24 年度から、就業構造基本調査データを用いた擬似ミクロデータ作成のための研究を行った。

はじめに、公表結果表での使用頻度等を考慮し使用変数を選定した。次に、使用変数全てを用いた高次元クロス集計表を作成した。さらに、 (1) 集計表セル単位の量的変数の統計量に基づく多変量正規乱数（以下「多変量正規乱数」と略す。）と(2)集計表セル度数の出現確率に基づく一様乱数（以下「一様乱数」と略す。）を生成し、それぞれから擬似ミクロデータを作成した。

作成した擬似ミクロデータと実データとを比較検証した結果、一様乱数は多変量正規乱数よりも実データ分布への近似が良好であり、集計用乗率を考慮した分布へも適用することが可能であることがわかった。

この結果、先行研究の量的変数の擬似ミクロデータ作成に続き、質的変数の擬似ミクロデータの作成研究でも成果が得られ、両変数に対応できたことから、他調査への応用も可能となった。今後は、擬似ミクロデータのもとになる集計表の作成方法の改善が課題であると考えている。

* 統計センター統計情報・技術部統計技術研究課

† 匿名データ提供及びオーダーメード集計を実施している調査の種類は、総務省政策統括官（統計基準担当）（2014）「委託による統計の作成等及び匿名データの作成・提供に関する年度計画一覧（平成 26 年度）」を参照されたい。

1　はじめに　－研究の背景及び目的－

新統計法（平成十九年法律第五十三号）は、統計に関する基本法として、旧統計法（昭和二十二年法律第十八号）の全面改正及び統計報告調整法（昭和二十七年法律第百四十八号）の廃止とともに、平成 19 年 5 月 16 日に成立し、平成 21 年 4 月 1 日に全面施行された。

新統計法でのミクロデータの二次的利用には、①旧統計法より継続している第三十三条による調査票情報の利用、②新たな仕組みである第三十四条による委託による統計の作成等（オーダーメード集計）及び③第三十五条・三十六条による匿名データの作成・提供がある。

しかしながら、新統計法の全面施行の準備段階から、有識者より、二次的利用を促進するためにはミクロデータを統計教育・訓練用で利用できることが必要との議論が出ていた[1]。

二次的利用データの匿名データも教育目的での利用が可能であるが、実際には統計法令上の制約により、多くの学生を対象にした大学での統計演習等の利用には困難が生じる場合がある。

そこで、上記の制約を受けずに使用できるデータとして、調査票情報を集計して調査票情報との関連を断ち切り、その集計表からミクロデータの形式を持つ擬似的なデータ(以下「擬似ミクロデータ」)の作成に関する研究を、平成21年度から（独）統計センターで開始した。

本稿では、最初に全国消費実態調査の量的変数を用いた先行研究（秋山ほか(2012)）の概要を述べる。次に就業構造基本調査の質的変数を用いた擬似ミクロデータの作成方法及びその検証結果を示し、最後に今後の課題を提示する。

[1] 統計委員会（2009）及び統計委員会匿名データ部会（2009）を参照されたい。

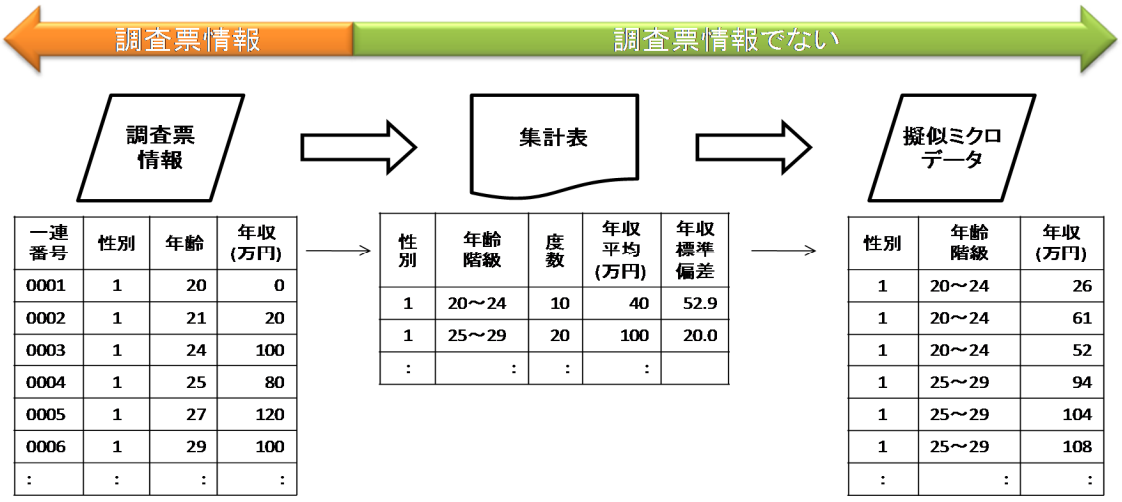
2　先行研究

2.1 量的変数の擬似ミクロデータの概要

（独）統計センターでは、欧米諸国で調査研究が進められている「ミクロアグリゲーション(micro-aggregation)」に着目するとともに、ミクロアグリゲーションの適用可能性に関する研究を行ってきた(伊藤(2008)，伊藤ほか(2008，2009))。こうしたミクロアグリゲーションの方法論を援用することによって、高次元にクロス集計した集計表から実データに近い分布と特性を持つ擬似的なミクロデータの作成が可能になることから(伊藤(2008，19 頁))、この研究成果を応用する形で、平成 21 年度から擬似ミクロデータの作成に着手した。

擬似ミクロデータ作成の流れは、図2-1のとおり、調査票情報から集計表を作成し、この集計表から実データに近似したミクロデータを擬似的に生成することである。

図 2-1　擬似ミクロデータ作成の基本イメージ



一連番号	性別	年齢	年収(万円)
0001	1	20	0
0002	1	21	20
0003	1	24	100
0004	1	25	80
0005	1	27	120
0006	1	29	100
:	:	:	:

性別	年齢階級	度数	年収平均(万円)	年収標準偏差
1	20～24	10	40	52.9
1	25～29	20	100	20.0
:	:	:	:	

性別	年齢階級	年収(万円)
1	20～24	26
1	20～24	61
1	20～24	52
1	25～29	94
1	25～29	104
1	25～29	108
:	:	:

秋山ほか(2012)では、平成16年全国消費実態調査データの量的変数（収支項目）の統計量（平均、分散及び共分散）に基づく多変量正規乱数を発生させ、擬似的な値を生成するという方法を用いた。

作成方法の処理手順は以下のとおりである。

(1) 変数の選択

使用する変数を、使用頻度等を考慮して表 2-1 のとおり選定した。

表 2-1　全国消費実態調査の擬似ミクロデータの変数[2]

質的変数　14変数		量的変数　184変数	
世帯事項	世帯区分 世帯人員区分 就業人員区分 住居の構造 住居の建て方 住宅の所有関係 入居時期･入居年	年間収入	支出(総額)
		収入(総額)	実支出
		実収入 実収入以外の収入 繰入金	消費支出
			食料
			穀類
			米 …
			住居
			…
			光熱･水道
			…
			家具･家事用品
			…
			被服及び履物
			…
			交通･通信
			…
			保健医療
			…
			教育
			…
			教養娯楽
			…
			その他の消費支出
			…
			非消費支出
			実支出以外の支出
			繰越金
世帯員事項･世帯主	性別 年齢5歳階級 就業･非就業の別 企業区分 企業規模 産業符号 職業符号		

[2] 変数の詳細は、秋山ほか(2012，29-32 頁)を参照されたい。

(2) 集計表作成

質的変数（14 変数）を用いたクロス集計表を作成し、集計表のセルの度数が２以下の場合[3]、質的変数が類似である組み合わせにグループ化し、度数を３以上にした。グループ内で同一でない符号は「V」または「VV」に置き換えた。どのグループにも統合できないセルは削除した。

(3) 統計量の算出

量的変数のうち収支項目大分類に相当する 21 変数について、数値を対数変換した[4]後、平均及び分散・共分散を算出した。

(4) 多変量正規乱数の発生

平均、分散及び共分散に基づく多変量正規乱数を発生させて擬似的な値を生成した。

(5) 一部の量的変数を 0 に置き換え

多変量正規乱数では 0 は生成されないが、実データには量的変数が 0 のケースもあるため（例：世帯人員 2 名で教育費が 0)、実データの 0 の発現パターンに基づき、一部の値を 0 に置き換えた。

(6) 収支バランスの調整

表 2-1 の支出（総額）以下に含まれる中分類（例：穀類…）及び小分類（例：米…）について、上位大分類（例：食料）の内訳の構成比を算出した上で、収支バランスを調整した。

(7) 集計用乗率の付与

集計表のセルごとに平均した集計用乗率を、各レコードに付与した。

[3] 全国消費実態調査の公表結果表では度数２以下を秘匿しているため、これに準拠した。
[4] 収支金額は、実数のままでは低い値に集中し、必ずしも正規分布に従わないため、対数正規分布に従うことを仮定した。

実データと擬似ミクロデータとを比較した結果は以下のとおりである。

(1) 基本統計量及びヒストグラム（表 2-2 及び図 2-2）

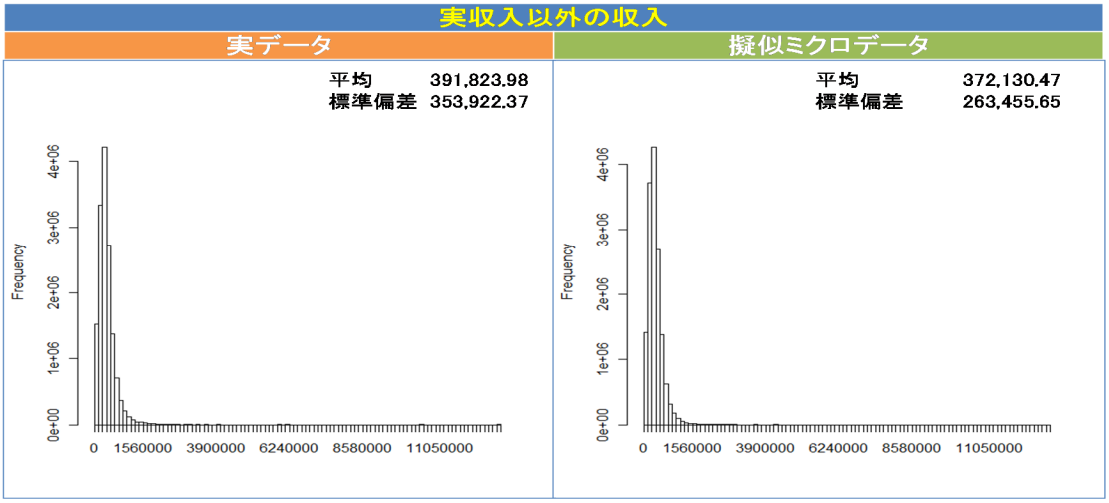
基本統計量の標準偏差では「実収入以外の収入」の差（表 2-2 の注を参照）が最も大きかったが、「実収入以外の収入」のヒストグラムでは両者の分布形状は近似している。

表 2-2　実データと擬似ミクロデータの比較－基本統計量[5]

	平均				標準偏差			
	実データ	擬似ミクロデータ	差		実データ	擬似ミクロデータ	差
年間収入	740.18	729.81	-0.01	年間収入	358.18	337.69	-0.06
収入総額	971,789.24	946,779.03	-0.03	収入総額	541,290.74	473,480.73	-0.13
実収入	502,133.73	497,655.92	-0.01	実収入	280,695.92	261,558.27	-0.07
実収入以外の収入	391,823.98	372,130.47	-0.05	実収入以外の収入	353,922.37	263,445.65	-0.26
繰入金	77,831.53	76,992.65	-0.01	繰入金	87,036.21	98,947.04	0.14
支出総額	971,789.24	946,779.03	-0.03	支出総額	541,290.74	473,480.73	-0.13
実支出	415,809.39	403,746.63	-0.03	実支出	224,419.69	219,290.60	-0.02
消費支出	339,199.37	328,139.70	-0.03	消費支出	194,501.15	192,447.21	-0.01
食料	73,738.54	72,883.42	-0.01	食料	30,149.02	28,064.49	-0.07
住居	19,387.99	17,687.21	-0.09	住居	52,962.36	60,587.32	0.14
光熱・水道	19,395.36	19,237.81	-0.01	光熱・水道	8,009.23	7,690.12	-0.04
家具・家事用品	9,783.81	9,204.04	-0.06	家具・家事用品	15,977.65	14,933.13	-0.07
被服及び履物	14,649.44	14,137.63	-0.03	被服及び履物	18,837.04	19,823.09	0.05
保健医療	11,936.01	11,366.36	-0.05	保健医療	19,763.39	19,284.07	-0.02
交通・通信	50,740.68	47,960.92	-0.05	交通・通信	85,021.69	84,654.38	0.00
教育	22,332.15	22,269.65	0.00	教育	51,989.72	64,157.45	0.23
教養娯楽	32,472.95	31,389.49	-0.03	教養娯楽	32,161.60	32,723.04	0.02
その他の消費支出	84,762.44	82,003.18	-0.03	その他の消費支出	95,898.83	102,040.97	0.06
非消費支出	76,610.02	75,606.93	-0.01	非消費支出	56,199.75	66,378.49	0.18
実支出以外の支出	475,947.80	464,318.09	-0.02	実支出以外の支出	394,805.29	334,227.09	-0.15
繰越金	80,032.04	78,714.31	-0.02	繰越金	96,421.45	118,055.82	0.22

注　集計表作成段階で、度数２以下のセルを削除しているため、本表の実データの数値は公表結果表とは一致しない。また、本表の「差」の計算式は「（擬似ミクロデーター実データ）/実データ」である。

図 2-2　実データと擬似ミクロデータの比較－ヒストグラム[6]



[5] 秋山ほか（2012，20 頁，表 4-1）

[6] 秋山ほか（2012，21 頁，図 4-1）

(2) 相関係数及び散布図（表 2-3、表 2-4 及び図 2-3）

図 2-3 は、実データと擬似ミクロデータの相関係数の差が最も大きい「年間収入」×「非消費支出」の散布図である。

多変量正規乱数発生の際に実データより大きく外れた値が出現したため、相関係数に 0.20 の差が生じているが、散布図の分布形状は両者とも右上がりで近似している。

表 2-3　実データの相関係数[7]

	年間収入	収入総額	実収入	実収入以外の収入	繰入金	支出総額	実支出	消費支出	食料	住居	光熱・水道	家具・家事用品	被服及び履物	保健医療	交通・通信	教育	教養娯楽	その他の消費支出	非消費支出	実支出以外の支出	繰越金
年間収入	1.00																				
収入総額	0.60	1.00																			
実収入	0.66	0.78	1.00																		
実収入以外の収入	0.35	0.85	0.36	1.00																	
繰入金	0.19	0.26	0.14	0.04	1.00																
支出総額	0.60	1.00	0.78	0.85	0.26	1.00															
実支出	0.60	0.73	0.56	0.63	0.17	0.73	1.00														
消費支出	0.49	0.66	0.45	0.61	0.16	0.66	0.97	1.00													
食料	0.47	0.42	0.37	0.31	0.17	0.42	0.52	0.50	1.00												
住居	-0.02	0.11	0.00	0.16	0.01	0.11	0.24	0.28	-0.03	1.00											
光熱・水道	0.32	0.24	0.22	0.16	0.11	0.24	0.28	0.27	0.44	-0.07	1.00										
家具・家事用品	0.15	0.25	0.12	0.26	0.09	0.25	0.26	0.27	0.17	0.07	0.10	1.00									
被服及び履物	0.30	0.30	0.24	0.24	0.10	0.30	0.39	0.38	0.29	0.02	0.12	0.16	1.00								
保健医療	0.11	0.16	0.10	0.15	0.07	0.16	0.24	0.25	0.15	0.01	0.07	0.08	0.09	1.00							
交通・通信	0.14	0.33	0.15	0.37	0.04	0.33	0.54	0.57	0.12	0.01	0.05	0.05	0.10	0.06	1.00						
教育	0.18	0.23	0.15	0.23	0.03	0.23	0.37	0.39	0.24	-0.03	0.19	0.02	0.09	0.04	0.07	1.00					
教養娯楽	0.32	0.35	0.27	0.30	0.12	0.35	0.44	0.42	0.32	0.02	0.10	0.15	0.26	0.10	0.10	0.09	1.00				
その他の消費支出	0.39	0.46	0.38	0.37	0.12	0.46	0.66	0.66	0.21	0.01	0.13	0.12	0.19	0.11	0.12	0.04	0.16	1.00			
非消費支出	0.70	0.63	0.70	0.38	0.12	0.63	0.62	0.43	0.35	-0.02	0.19	0.12	0.26	0.08	0.17	0.14	0.29	0.34	1.00		
実支出以外の支出	0.44	0.90	0.72	0.79	0.04	0.90	0.40	0.32	0.25	0.01	0.14	0.18	0.17	0.08	0.14	0.11	0.22	0.23	0.49	1.00	
繰越金	0.16	0.24	0.13	0.06	0.86	0.24	0.13	0.12	0.13	0.02	0.10	0.07	0.07	0.05	0.02	0.02	0.08	0.10	0.10	0.01	1.00

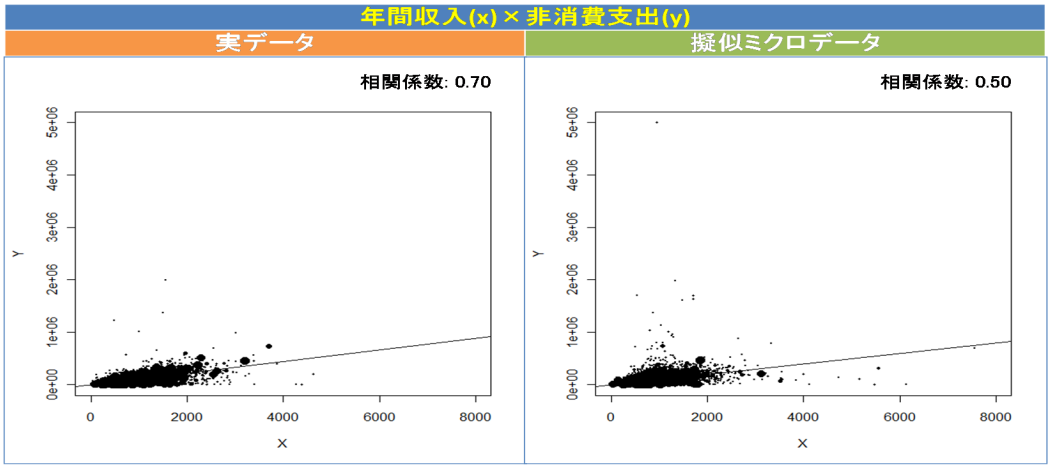
表 2-4　擬似ミクロデータの相関係数[8]

	年間収入	収入総額	実収入	実収入以外の収入	繰入金	支出総額	実支出	消費支出	食料	住居	光熱・水道	家具・家事用品	被服及び履物	保健医療	交通・通信	教育	教養娯楽	その他の消費支出	非消費支出	実支出以外の支出	繰越金
年間収入	1.00																				
収入総額	0.58	1.00																			
実収入	0.63	0.85	1.00																		
実収入以外の収入	0.38	0.83	0.48	1.00																	
繰入金	0.12	0.32	0.15	0.05	1.00																
支出総額	0.58	1.00	0.85	0.83	0.32	1.00															
実支出	0.52	0.71	0.59	0.64	0.14	0.71	1.00														
消費支出	0.42	0.63	0.49	0.60	0.14	0.63	0.96	1.00													
食料	0.46	0.40	0.36	0.32	0.13	0.40	0.45	0.43	1.00												
住居	-0.05	0.08	0.04	0.09	0.03	0.08	0.24	0.28	-0.06	1.00											
光熱・水道	0.32	0.25	0.23	0.18	0.09	0.25	0.26	0.25	0.44	-0.07	1.00										
家具・家事用品	0.12	0.15	0.11	0.14	0.04	0.15	0.19	0.19	0.15	0.00	0.10	1.00									
被服及び履物	0.21	0.23	0.19	0.20	0.06	0.23	0.29	0.28	0.20	0.01	0.08	0.12	1.00								
保健医療	0.07	0.13	0.09	0.13	0.04	0.13	0.19	0.20	0.11	0.00	0.06	0.05	0.05	1.00							
交通・通信	0.12	0.30	0.17	0.35	0.04	0.30	0.50	0.54	0.10	-0.01	0.05	0.03	0.06	0.04	1.00						
教育	0.14	0.24	0.18	0.24	0.02	0.24	0.38	0.41	0.18	-0.02	0.16	0.01	0.04	0.02	0.04	1.00					
教養娯楽	0.26	0.30	0.24	0.28	0.06	0.30	0.35	0.34	0.26	-0.01	0.06	0.12	0.18	0.07	0.07	0.05	1.00				
その他の消費支出	0.33	0.44	0.38	0.37	0.11	0.44	0.63	0.65	0.17	-0.02	0.11	0.07	0.11	0.06	0.09	0.04	0.10	1.00			
非消費支出	0.50	0.50	0.52	0.35	0.07	0.50	0.53	0.26	0.24	-0.04	0.14	0.07	0.14	0.05	0.09	0.07	0.18	0.21	1.00		
実支出以外の支出	0.45	0.85	0.77	0.74	0.07	0.85	0.32	0.25	0.25	-0.05	0.15	0.08	0.13	0.05	0.09	0.09	0.18	0.18	0.35	1.00	
繰越金	0.10	0.28	0.14	0.05	0.82	0.28	0.07	0.07	0.09	0.00	0.08	0.03	0.03	0.02	0.01	0.00	0.02	0.06	0.04	0.00	1.00

[7] 秋山ほか（2012，22 頁，表 4-2）

[8] 秋山ほか（2012，22 頁，表 4-3）

図 2-3　実データと擬似ミクロデータの比較－散布図[9]



[9] 秋山ほか（2012，23 頁，図 4-2）

2.2 量的変数の擬似ミクロデータの提供

平成 16 年全国消費実態調査の擬似ミクロデータは、平成 23 年８月より統計センターホームページ経由で一般への試行提供を開始した。平成 24 年度から、利用者の要望に応えレコード数及び変数を少なくした簡易データを追加し、現在は表 2-5 の仕様で試行提供中である。

表 2-5　試行提供中の擬似ミクロデータ

<table>
<tr><td colspan="6">平成16年全国消費実態調査（平成23年度より試行提供開始）
http://www.nstac.go.jp/services/giji-microdata.html</td></tr>
<tr><td rowspan="2"></td><td rowspan="2">レコード数</td><td rowspan="2">収録項目</td><td colspan="3">収録項目の内訳</td></tr>
<tr><td>世帯属性等</td><td>支出項目</td><td>収入項目</td></tr>
<tr><td>大規模データ
（CSV形式）</td><td>約3万2千

二人以上の
勤労者世帯</td><td>197項目</td><td rowspan="2">14項目

世帯主の年齢、住居の種類など</td><td>149項目

用途分類</td><td>34項目

年間収入等</td></tr>
<tr><td>簡易データ
（CSV形式
Excel形式）</td><td>約8千

世帯人員が4名で
有業人員が
1～2人の世帯のみ</td><td>25項目</td><td>11項目

消費支出及び十大費目のみ</td><td>なし</td></tr>
</table>

3　質的変数の擬似ミクロデータに関する作成方法

先行研究で作成した擬似ミクロデータの試行提供後、多くの利用者から、質的変数への対象拡大の要望が寄せられた。しかしながら、質的変数は量的変数と異なり、平均や分散に基づく乱数発生が困難であったため、最初に平成16年全国消費実態調査の質的変数で加工方法を試行した後（別添１参照）、平成14年就業構造基本調査のデータを用いて、質的変数の擬似ミクロデータの作成方法の検証を行った。

なお、本章の検証に用いたプログラム言語は、Microsoft Excel の Visual Basic for Applications 及び R である。

3.1 多変量正規乱数及び一様乱数の検証

最初に、第２章で述べた多変量正規乱数及び集計表セル度数の出現確率に基づく一様乱数（以後「一様乱数」と略す。）による作成方法を検証することとした。

3.1.1　使用したデータ及び変数

使用したデータは、平成14年就業構造基本調査[10]（968,628レコード）の「世帯主有業者」[11]281,489レコードである。変数は、公表結果表での使用頻度、調査項目の利用状況等を考慮し、表3-1のとおり選定した[12]。

表 3-1　使用した変数

集計表使用変数（15変数）		乱数発生変数（2変数）
基本変数（8変数）	加工対象変数（7変数）	個人所得 継続就業期間
性別 配偶者の有無 続き柄 教育（教育区分） 教育（学校区分） 年齢（5歳階級） 1年前との就業異動 前職の有無	従業上の地位（5区分） 雇用形態 産業（大分類） 職業（大分類） 従業者規模 年間就業日数 週間就業時間	

基本変数：　集計表使用変数のうち、公表結果表における頻度の高い変数。
集計表の度数２以下の際に符号の置き換えを行わない。
加工対象変数：　集計表使用変数のうち、基本変数以外の変数。
集計表の度数２以下の際に符号の置き換え対象となる。
乱数発生変数：　集計表には使用せず、集計表作成後に多変量正規乱数及び一様乱数を発生させる変数。

[10] 平成19年データの使用も検討したが、当該年次は匿名データが未作成である。
[11] 全国消費実態調査における擬似ミクロデータの作成手法の適用可能性を検証するため、対象レコードも全国消費実態調査のデータ形式（世帯単位）に合わせ、世帯主有業者のみとした。
[12] 各変数における符号については、別添２を参照されたい。

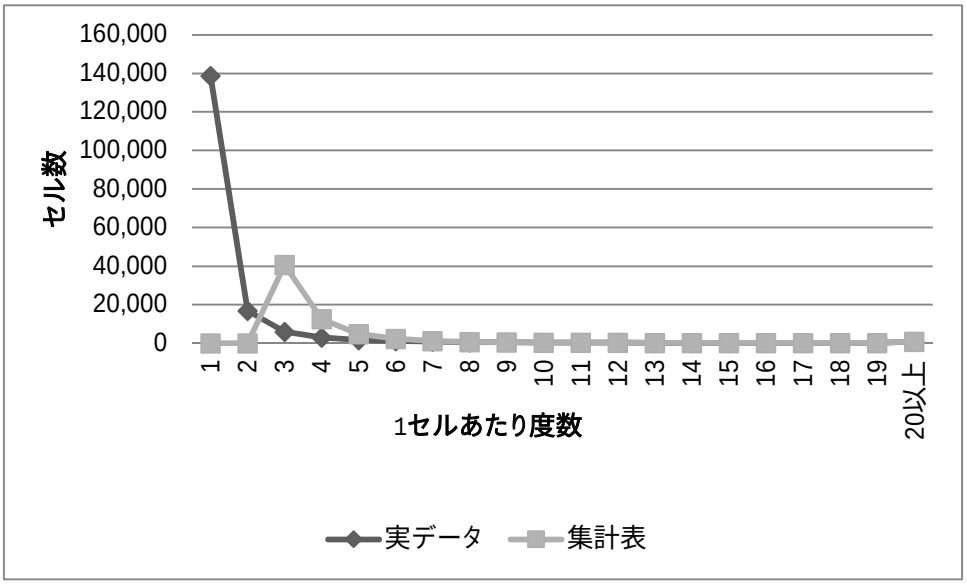
3.1.2 集計表作成

集計表使用変数（15変数）のクロス集計表を作成し、集計表のセルの度数が２以下の場合、質的変数が類似である組み合わせにグループ化し、度数を３以上にした。グループ内で同一でない符号は「V」または「VV」に置き換えた。どのグループにも統合できないレコードは削除した。

不詳置き換えの優先順位は、各変数の秘匿性を考慮し、週間就業時間、年間就業日数、従業者規模、職業（大分類)、産業（大分類)、雇用形態、従業上の地位（５区分）の順番とした。

集計の結果、実データで171,130であった集計セル数は64,976に減少した。また、１セルあたり度数は、実データの約61.2%を占めていた度数２以下のセルが全て度数３～14となった（図3-1参照)。どのグループにも統合できず削除されたレコードは、281,489レコード中2,139レコード（約0.8%）であった。

図 3-1 集計表作成によるセル数及び１セルあたり度数の変化



3.1.3 乱数発生方法

「個人所得」及び「継続就業期間」[13]について、次の２つの方法で乱数を発生させた。

(1) 多変量正規乱数

量的変数の平均、分散及び共分散に基づく多変量正規乱数を発生させ擬似的な値を生成する方法で、処理手順は以下のとおりである[14]。

① 個人所得の区分を、量的変数（中央値）に変換し、対数変換を行った[15]。
② 継続就業期間の区分を、量的変数（月値）に変換し、対数変換を行った[16]。
③ ①及び②の数値に対し、集計表のセルごとに平均、分散及び共分散を算出し、これらに基づく多変量正規乱数を発生させた。
④ 乱数は対数の状態であるため、実数に戻した後、区分へ対応させた。

(2) 一様乱数

集計表セル度数の出現確率に基づく一様乱数を発生させる方法で、処理手順は以下のとおりである。なお、作成イメージについては別添3-1を参照されたい。

① 実データの集計表から、各セルの度数の出現確率を算出した。
② ①の出現確率を累積確率に変換した。
③ 0～1 の一様乱数を発生させて②の累積確率と照合し、該当する変数組み合わせの擬似レコードを生成した。この処理を実データのレコード分繰り返した。
④ ③の擬似レコードを変数組み合わせ順にソートし、擬似ミクロデータを完成させた。

---

[13] 各変数における符号については、別添２を参照されたい。
[14] 両変数とも、符号「不詳」については量的変数への変換を行っていない。
[15] 個人所得は、実数のままでは低い値に集中し、必ずしも正規分布に従わないため、対数正規分布に従うことを仮定した。
[16] 継続就業期間は、実数のままでは高い値に集中し、必ずしも正規分布に従わないため、対数正規分布に従うことを仮定した。

3.1.4 検証結果

3.1.3 で作成した擬似ミクロデータの「個人所得」及び「継続就業期間」について、実データとの比較検証を行った[17]。

(1) クロス集計表

自営業主と雇用者[18]、前職ありと前職なし[19]の区分を追加し、クロス集計表を作成したところ、図 3-2 及び図 3-3 の棒グラフ及び折れ線グラフのとおり、一様乱数の方が多変量正規乱数に比べて実データにより近似した。

＜グラフの凡例＞

図 3-2 個人所得 自営業主×前職あり（上）自営業主×前職なし（下）

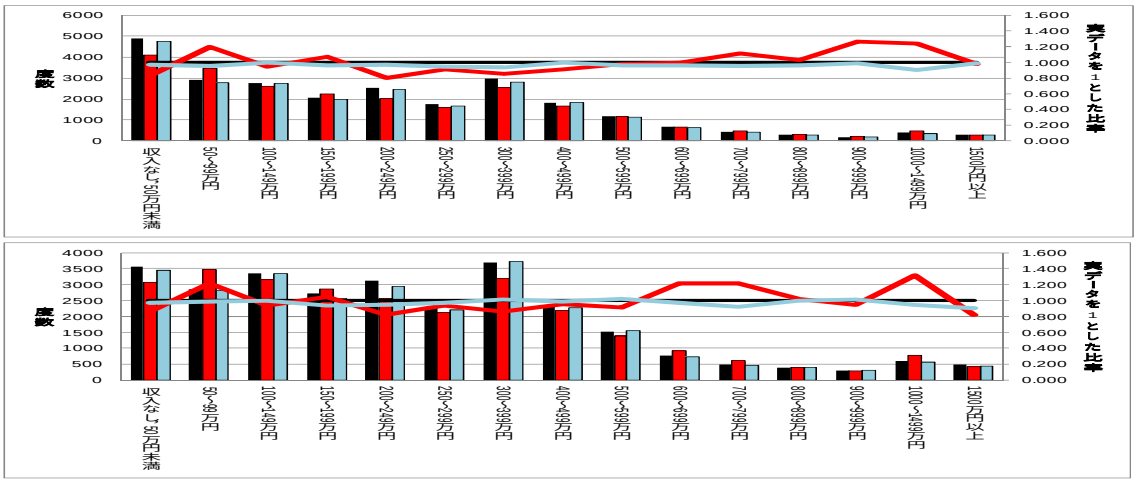
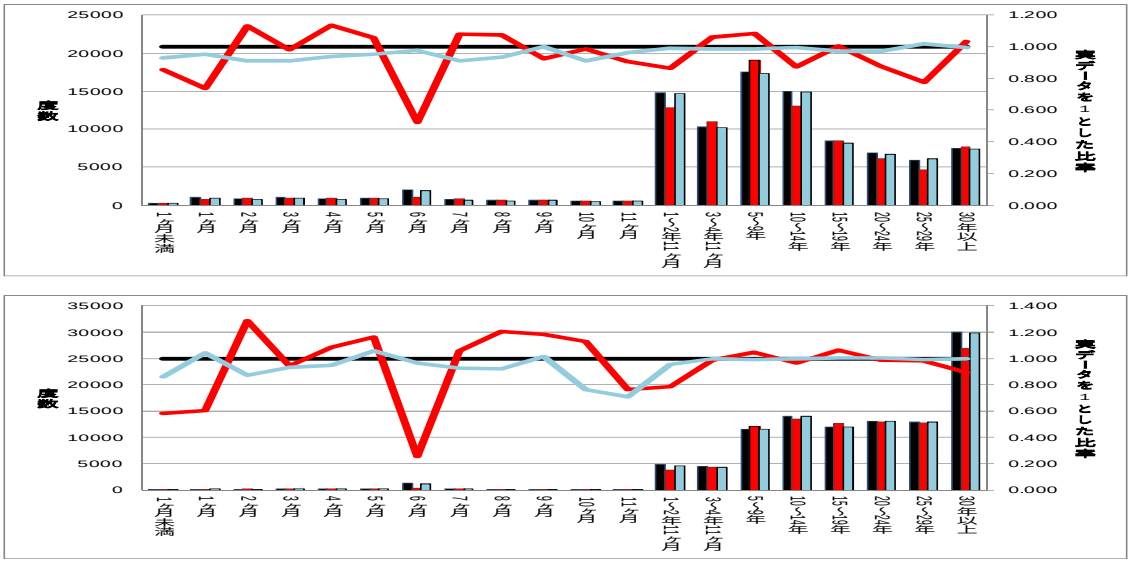


図 3-3 継続就業期間 雇用者×前職あり（上）雇用者×前職なし（下）



[17] 個人所得は「不詳」、継続就業期間は「不詳」及び「1 年未満（月不詳）」を除いた結果である。
[18] 従業上の地位における「1：内職者以外の自営業主」「4：民間の役員・常雇」をそれぞれ指す。
[19] 前職の有無における「2：あり」と「1：なし」をそれぞれ指す。

(2) カイ二乗検定

実データ及び擬似ミクロデータの個人所得及び継続就業期間のクロス集計表について、帰無仮説を「両データの分布に差がない」と設定し、独立性のカイ二乗検定[20]を行った結果、表 3-2 のとおり、実データ×多変量正規乱数の p 値はすべて 0.05 未満で有意差が認められたが、実データ×一様乱数の p 値は、「雇用者×前職あり」「自営業主×前職あり」を除き 0.05 以上で有意差が認められなかった。

表 3-2　独立性のカイ二乗検定

検定対象変数		実データ×多変量正規乱数			実データ×一様乱数		
		カイ二乗統計量	自由度	p値	カイ二乗統計量	自由度	p値
個人所得	自営業主×前職あり	219.0766	14	< 2.2e-16	4.8401	14	0.9879
	自営業主×前職なし	250.2059	14	< 2.2e-16	11.3582	14	0.6577
	雇用者×前職あり	222.7423	14	< 2.2e-16	32.9551	14	0.0029
	雇用者×前職なし	397.8419	14	< 2.2e-16	19.3856	14	0.1507
継続就業期間	自営業主×前職あり	240.5263	19	< 2.2e-16	23.0714	19	0.2342
	自営業主×前職なし	167.3398	19	< 2.2e-16	37.7853	19	0.0063
	雇用者×前職あり	830.5278	19	< 2.2e-16	18.6413	19	0.4801
	雇用者×前職なし	838.8219	19	< 2.2e-16	14.8047	19	0.7349

以上の検証結果から、一様乱数の方が多変量正規乱数よりも実データの分布をよりよく再現すると判断し、以後の検証は一様乱数に絞って進めることとした。

[20] クロス集計表の各セルで、擬似ミクロデータの度数をx^i、実データの度数をe^iとすると、独立性のカイ二乗統計量は$\sum_{i=1}^{n}\left(x^i - e^i\right)^2 / e^i$となる。本稿の検定結果によれば、サンプルサイズ（n）が大きいため検出力が強くなっている可能性があるものの、p 値が 0.05 未満であることから、有意差がある（帰無仮説を棄却）と判断することができる。

3.2 集計用乗率を考慮した一様乱数の検証

作成方法の評価を正確に行うためには公表結果表との比較が必要であるため、使用するデータを「有業者」に拡大するとともに、乱数発生の際に集計用乗率を考慮し、公表結果表との比較検証を行った。

3.2.1 使用したデータ及び変数

平成 14 年就業基本構造調査（968,628 レコード）のうち「有業者」567,699 レコードを用い、変数を表 3-3 のとおり設定した[21]。

表 3-3　使用した変数

集計用使用変数（15変数）	
基本変数（8変数）	加工対象変数（7変数）
性別 配偶者の有無 続き柄 教育（教育区分） 教育（学校区分） 年齢（5歳階級） 1年前との就業異動 前職の有無	従業上の地位（5区分） 雇用形態 産業（大分類） 職業（大分類） 従業者規模 年間就業日数 週間就業時間

基本変数：　集計表使用変数のうち、公表結果表における頻度の高い変数。
集計表の度数２以下の際に符号の置き換えを行わない。

加工対象変数：　集計表使用変数のうち、基本変数以外の変数。
集計表の度数２以下の際に符号の置き換え対象となる。

[21] 各変数における符号については、別添２を参照されたい。

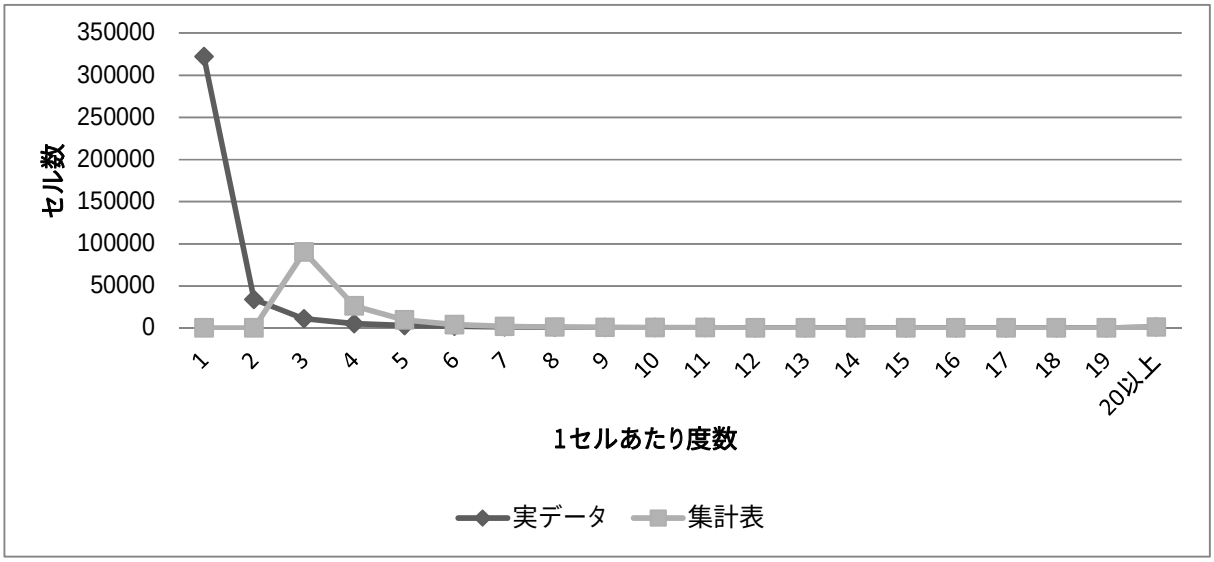
3.2.2 集計表作成

集計表使用変数（15変数）を用いて、集計表のセルの度数が２以下の場合、質的変数が類似である組み合わせにグループ化し、度数を３以上にした。グループ内で同一でない符号は「V」または「VV」に置き換えた。どのグループにも統合できないレコードは削除した。

不詳置き換えの優先順位は、各変数の秘匿性を考慮し、週間就業時間、年間就業日数、従業者規模、職業（大分類）、産業（大分類）、雇用形態、従業上の地位（５区分）の順番とした。

集計の結果、実データで383,837であった集計セル数は138,120に減少した。１セルあたり度数は、図3-4のとおり、実データの68.7%を占めていた度数２以下のセルが全て度数３～13となった（図3-4参照）。どのグループにも統合できず削除されたレコードは、567,699レコード中12,032レコード（約2.1%）であった。

図 3-4　集計表作成によるセル数及び１セルあたり度数の変化



3.2.3 乱数発生方法

3.1.3の方法を応用し、集計用乗率を考慮した一様乱数を発生させた。処理手順は表3-4の左列のとおりである。作成イメージについては別添3-2を参照されたい。

なお、3.1.3の検証では変数を固定せずに一様乱数発生を行ったが、本章の検証では、公表結果表の分布を再現する目的から、基本変数を固定して一様乱数発生を行った。詳細は別添4を参照されたい。

表 3-4　一様乱数による乱数発生方法

集計表乗率を考慮した一様乱数	（参考）度数に基づく一様乱数 (3.1.3(2)記載)
①　実データに基づいたクロス集計表から、各セルの集計表乗率合計の出現確率を算出した。	①　実データに基づいたクロス集計表から、各セルの度数の出現確率を算出した
②　①の出現確率を累積確率に変換した。	②　①の出現確率を累積確率に変換した。
③　0～1 の一様乱数を発生させて②の累積確率と照合し、該当する変数組み合わせの擬似レコードを生成した。この処理を実データの乗率合計分繰り返した。	③　0～1 の一様乱数を発生させて②の累積確率と照合し、該当する変数組み合わせの擬似レコードを生成した。この処理を実データのレコード分繰り返した。
④　③のレコードを再集計した。(乱数発生後のため、①の集計表とは集計表乗率合計が異なる)	―
⑤　④の集計表の各セルを複数レコードに分割した。その際、データセット全体のレコード数が実データのレコード数と一致し、かつ各レコードの集計表乗率がセルの集計表乗率合計と一致するように分割した。	―
⑥　擬似レコードを変数組み合わせ順にソートし、擬似ミクロデータを完成させた。	④　擬似レコードを変数組み合わせ順にソートし、擬似ミクロデータを完成させた。

3.2.4 検証結果

以下の変数について、公表結果表、集計表及び乱数発生結果の比較検証を行った[22]。

① 産業（公表結果表第 18 表(報告書第 9 表)と比較）

② 週間就業時間（うち雇用者・男・200 日未満就業者・うち規則的就業）
（公表結果表第 40 表(報告書第 18 表)と比較）

乱数発生結果をより正確に評価するため、乱数発生回数 3 回、20 回、100 回の平均度数も参考値として算出し、併せて検証を行った。

産業の度数は、表 3-5 のとおり、公表結果表と集計表には差が認められたが、一様乱数については乱数発生回数を増やすほど集計表に近似した。

次に、表 3-5 の公表結果表と集計表、集計表と一様乱数について、帰無仮説を「両データの分布に差がない」と設定し、独立性のカイ二乗検定を行ったところ、表 3-6 のとおり、前者は p 値が 0.05 未満で有意差が認められたが、後者は p 値が 0.7 を上回り、有意差が認められなかった。

表 3-5 産業別有業者数

産業	公表結果表	集計表	一様乱数発生			
			1回	(参考)		
				3回平均	20回平均	100回平均
農業	2,703,700	2,340,100	2,339,700	2,338,000	2,339,700	2,340,000
林業	58,500	12,800	12,800	12,800	12,800	12,800
漁業	265,500	158,000	158,000	158,100	158,100	158,100
鉱業	40,100	7,300	7,500	7,300	7,300	7,300
建設業	6,086,100	4,903,700	4,902,600	4,904,100	4,904,400	4,904,000
製造業	12,202,000	10,564,100	10,568,500	10,563,500	10,564,700	10,564,100
電気･ガス･熱供給･水道業	376,800	247,300	247,100	247,100	247,000	247,200
情報通信業	1,766,100	1,062,900	1,062,200	1,062,200	1,062,900	1,062,900
運輸業	3,327,300	2,380,900	2,382,000	2,382,200	2,381,300	2,380,900
卸売･小売業	11,699,200	9,775,400	9,774,000	9,776,200	9,774,600	9,775,000
金融･保険業	1,781,300	1,251,600	1,251,900	1,251,900	1,251,900	1,251,800
不動産業	916,200	377,400	377,000	377,200	377,500	377,500
飲食店，宿泊業	3,632,000	2,544,200	2,545,400	2,546,400	2,544,700	2,544,400
医療，福祉	4,891,700	3,867,900	3,868,300	3,868,100	3,867,500	3,867,800
教育，学習支援業	2,826,400	2,067,800	2,069,800	2,068,800	2,068,100	2,068,100
複合サービス事業	769,200	477,200	476,500	476,700	477,000	477,200
サービス業（他に分類されないもの）	8,460,200	6,321,300	6,322,300	6,320,200	6,321,100	6,321,100
公務（他に分類されないもの）	2,174,000	1,923,400	1,922,200	1,924,100	1,923,500	1,923,300
（集計表 － 一様乱数）の絶対値の和			**16,700**	**12,400**	**5,400**	**2,200**

表 3-6 独立性のカイ二乗検定

検定対象データ			公表結果表			集計表		
			カイ二乗統計量	自由度	p値	カイ二乗統計量	自由度	p値
集計表			613715.2	17	<2.2e-16			
一様乱数発生	1回		613818.1	17	<2.2e-16	13.5261	17	0.7003
	(参考)	3回平均	613396.4	17	<2.2e-16	6.9057	17	0.9847
		20回平均	613475.6	17	<2.2e-16	1.1264	17	1
		100回平均	613412.6	17	<2.2e-16	0.2737	17	1

[22] 公表結果表と集計表については、集計表作成時の不詳置き換え処理の影響で数値の差が生じている。詳細は別添５を参照されたい。

週間就業時間（うち雇用者・男・200 日未満就業者・うち規則的就業）の度数は、表 3-7 のとおり、公表結果表と集計表に差が認められたが、一様乱数については乱数発生回数を増やすほど集計表に近似した。

次に、表 3-7 の公表結果表と集計表、集計表と一様乱数について、帰無仮説を「両データの分布に差がない」と設定し、独立性のカイ二乗検定を行ったところ、表 3-8 のとおり、前者は p 値が 0.05 未満で有意差が認められたが、後者は、乱数発生回数 3 回平均以上で p 値が 0.8 を上回り、有意差が認められなかった。

表 3-7　週間就業時間「うち雇用者・男・200 日未満就業者・うち規則的就業」別有業者数

週間就業時間	公表結果表	集計表	一様乱数発生			
			1回	(参考)		
				3回平均	20回平均	100回平均
15時間未満	300,700	38,400	38,400	38,400	38,400	38,400
15～19時間	224,800	21,900	21,900	21,900	21,900	21,900
20～21時間	147,700	5,600	5,700	5,600	5,600	5,600
22～29時間	286,000	28,000	28,000	28,000	28,000	28,000
30～34時間	264,800	20,200	20,400	20,200	20,200	20,200
35～42時間	771,600	186,900	186,400	186,900	186,800	186,900
43～45時間	258,800	17,900	18,300	18,200	18,000	17,900
46～48時間	252,000	25,000	24,700	24,900	25,000	24,900
49～59時間	206,900	11,500	11,500	11,500	11,500	11,500
60時間以上	139,200	7,100	6,900	7,100	7,100	7,100
(集計表 － 一様乱数) の絶対値の和			**1,700**	**400**	**200**	**100**

表 3-8　　独立性のカイ二乗検定

検定対象データ			公表結果表			集計表		
			カイ二乗統計量	自由度	p値	カイ二乗統計量	自由度	p値
集計表			121622.3	9	<2.2e-16			
一様乱数発生	1回		120536.2	9	<2.2e-16	23.0467	9	0.006092
	(参考)	3回平均	121328.7	9	<2.2e-16	5.3147	9	0.8061
		20回平均	121350.3	9	<2.2e-16	0.6122	9	0.9999
		100回平均	121700.0	9	<2.2e-16	0.3725	9	1

以上の検証において、公表結果表と集計表の分布には差が見られたが、集計表と一様乱数の分布の差は小さく乱数発生回数の増加に従い近似度が増すことがわかった。

従って、一様乱数は、集計用乗率を考慮した分布再現にも有効であると言える。

4　おわりに　－結論及び課題－

質的変数の擬似ミクロデータ作成に関する検証の結果、集計表セル度数の出現確率に基づく一様乱数は、実データ分布に概ね近似しており、集計用乗率を考慮した分布へも適用可能であることが明らかになった。

先行研究の量的変数の擬似ミクロデータ作成に関する研究成果に続き、質的変数の擬似ミクロデータ作成の研究において成果が得られたことは、他の調査への応用も可能となることから、擬似ミクロデータの作成方法について進展があったと考える。

今後の課題は集計表の作成方法の改善である。本研究では、度数２以下のセルに対する不詳置き換えにより分布再現性が低下する傾向が見られた。このため、区分統合、リコーディング及び乗率による度数調整といった不詳置き換え以外の方法を用いた擬似ミクロデータの作成可能性の検討を、さらに進める必要があると考える。

別添１　質的変数の加工方法

本文第３章における検証の前に試行した、平成 16 年全国消費実態調査を用いた質的変数の加工方法は、以下のとおりである。

① 集計表セル度数の出現確率に基づく一様乱数を発生

集計表の各セルの度数の、全度数における割合（出現確率）を求め、その確率に基づく一様乱数を発生させて度数を加工する方法である。

② 集計表の度数に加法ノイズを付加

集計表の度数に加法ノイズを付加して、その合計がプラスマイナスゼロになるよう調整する方法である。

③ 集計表の度数２以下を度数３へ丸め

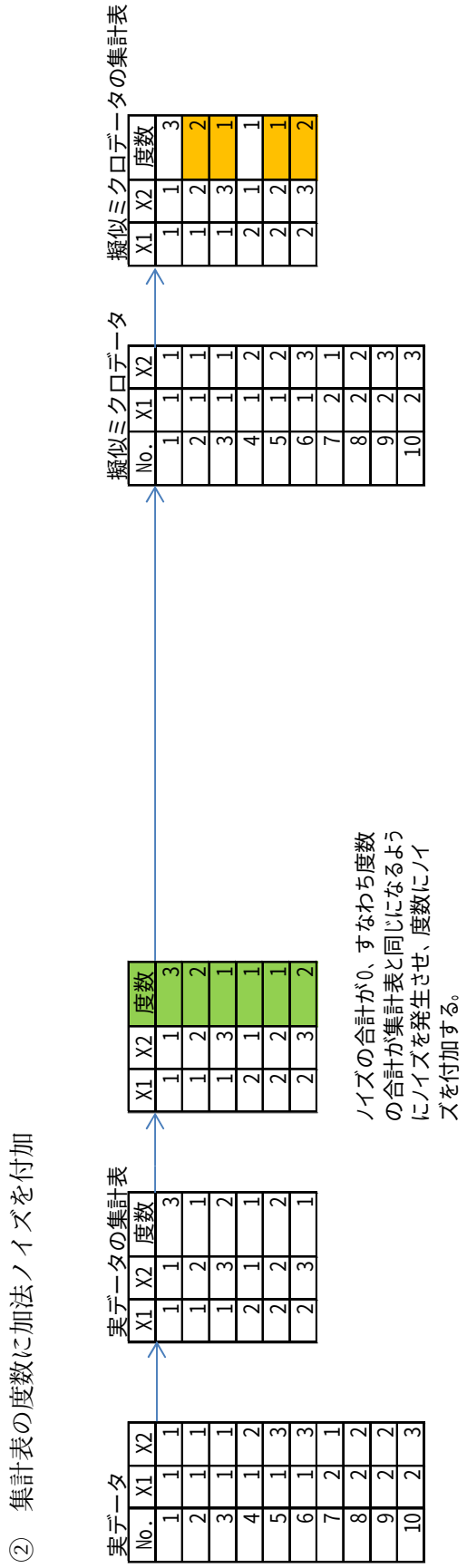
集計表の度数２以下のレコードを度数３に増加させた後に集計表乗率を変更し、度数分布を調整する方法である。

④ 集計表の周辺分布の再現

集計表を２つに分割し、そのうち１つの集計表にレコード数分の乱数を付与して乱数順にソートしてから、もう一方の集計表と合体する方法である。

上記の各方法を、手順の容易さ及び生成結果の安定性等で検討した結果、①の「集計表の出現確率に基づく乱数を発生させる」方法を、就業構造基本調査を用いた質的変数の擬似ミクロデータの作成研究に用いることとした。

②～④の加工イメージは以下のとおりである。


② 集計表の度数に加法ノイズを付加
実データ
No. X1 X2
1 1 1
2 1 1
3 1 1
4 1 2
5 1 3
6 1 3
7 2 1
8 2 2
9 2 2
10 2 3
実データの集計表
X1 X2 度数
1 1 3
1 2 1
1 3 2
2 1 1
2 2 2
2 3 1
X1 X2 度数
1 1 3
1 2 2
1 3 1
2 1 1
2 2 1
2 3 2
擬似ミクロデータ
No. X1 X2
1 1 1
2 1 1
3 1 1
4 1 2
5 1 2
6 1 3
7 2 1
8 2 2
9 2 3
10 2 3
擬似ミクロデータの集計表
X1 X2 度数
1 1 3
1 2 2
1 3 1
2 1 1
2 2 1
2 3 2
ノイズの合計が0、すなわち度数の合計が集計表と同じになるようにノイズを発生させ、度数にノイズを付加する。

③　集計表の度数２以下を度数３に丸め

実データ

No.	X1	X2
1	1	1
2	1	1
3	1	1
4	1	2
5	1	3
6	1	3
7	2	1
8	2	2
9	2	2
10	2	3

実データの集計表

X1	X2	度数
1	1	3
1	2	1
1	3	2
2	1	1
2	2	2
2	3	1

X1	X2	度数
1	1	3
1	2	3
1	3	3
2	1	3
2	2	3
2	3	3

度数1、2について、
度数3にする。

擬似ミクロデータ

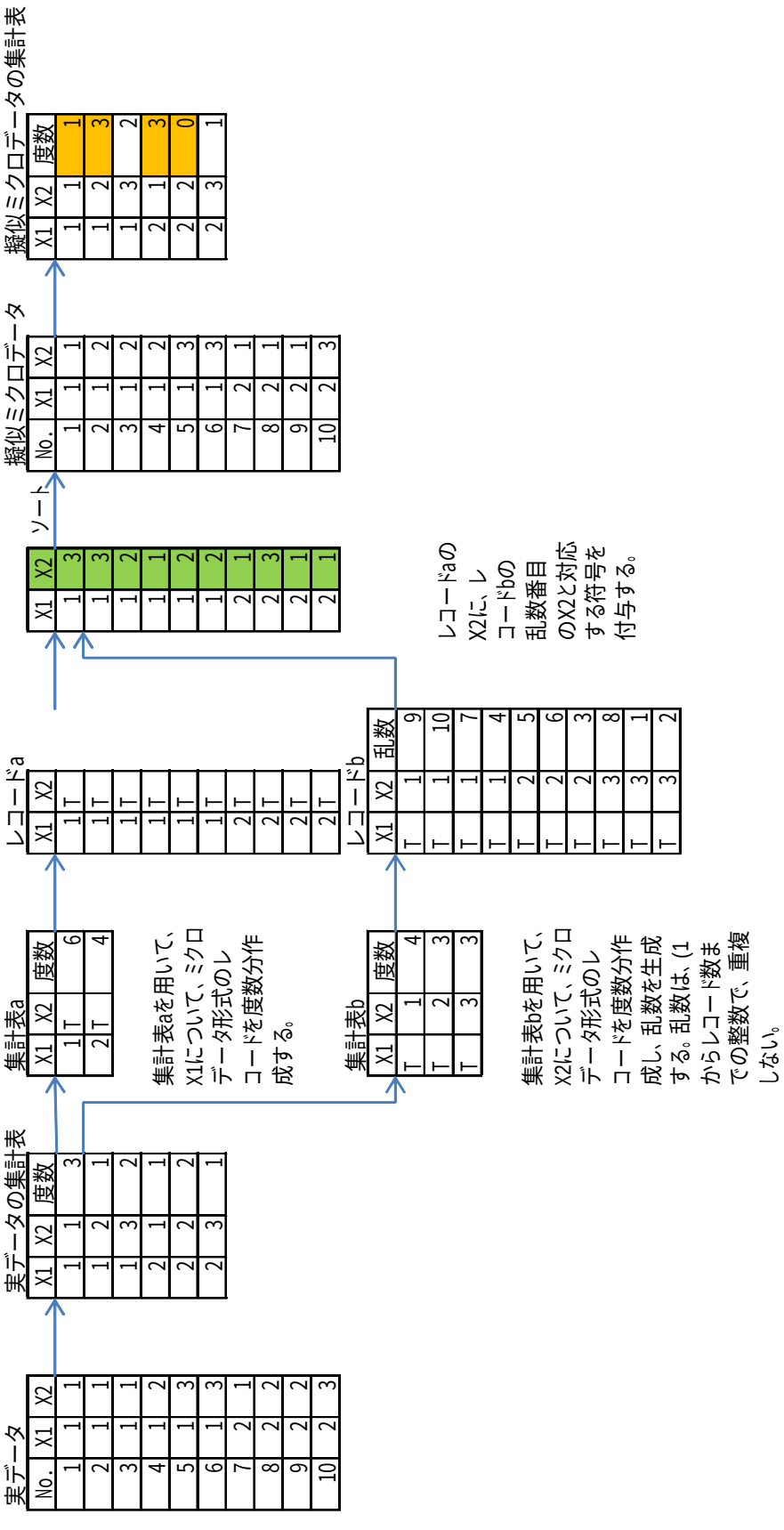
No.	X1	X2
1	1	1
2	1	1
3	1	1
4	1	2
5	1	2
6	1	2
7	1	3
8	1	3
9	1	3
10	2	1
11	2	1
12	2	1
13	2	2
14	2	2
15	2	2
16	2	3
17	2	3
18	2	3

擬似ミクロデータの集計表

X1	X2	度数
1	1	3
1	2	3
1	3	3
2	1	3
2	2	3
2	3	3

この後に、集計表乗
率で適宜調整すれば、
実データに近似する。

④ 集計表の周辺分布の再現



別添２　擬似ミクロデータの符号表

①　基本変数

項目名	対象	符号	符号内容
性別		1	男
		2	女
配偶者の有無		1	配偶者あり
		2	配偶者なし
		V	不詳
続き柄		01	世帯主
		02	世帯主の配偶者
		03	子及び子の配偶者
		04	世帯主の父母及び世帯主の配偶者の父母
		05	その他
		VV	不詳
(教育) 教育区分		1	在学中
		2	卒業
		3	在学したことがない
		V	不詳
(教育) 学校区分	教育区分が1，2	1	小学･中学
		2	高校･旧制中
		3	短大･高専
		4	大学･大学院
		V	不詳
(年齢) 5歳階級		01	15～19歳
		02	20～24歳
		03	25～29歳
		04	30～34歳
		05	35～39歳
		06	40～44歳
		07	45～49歳
		08	50～54歳
		09	55～59歳
		10	60～64歳
		11	65～69歳
		12	70～74歳
		13	75歳以上
1年前との就業異動		1	継続就業者：1年前にも現在と同じ勤め先 (企業) で就業していた者
		2	転職者：1年前の勤め先 (企業) と現在の勤め先とが異なっている者
		3	新規就業者：1年前には仕事をしていなかったが，この1年間に仕事についた者
		4	離職者：1年前には仕事をしていたが，その仕事をやめて現在仕事をしていない者
		5	継続非就業者：1年前も現在も仕事をしていない者
		V	不詳
(有業者) 前職の有無	有業者	1	ある
		2	ない
		V	不詳

② 加工対象変数

項目名	対象	符号	符号内容
(有業者・本業・従業上の地位) 5区分	有業者	1	内職者以外の自営業主
		2	内職者
		3	家族従業者
		4	民間の役員・常雇
		5	臨時雇・日雇
		V	不詳
(有業者・本業) 雇用形態	有業者で民間の役員を除く雇用者(常雇・臨時雇・日雇)	1	正規の職員・従業員
		2	パート
		3	アルバイト
		4	労働者派遣事業所の派遣社員
		5	契約社員・嘱託
		6	その他
		V	不詳
(有業者・本業・産業) 大分類	有業者	01	農業
		02	林業
		03	漁業
		04	鉱業
		05	建設業
		06	製造業
		07	電気・ガス・熱供給・水道業
		08	情報通信業
		09	運輸業
		10	卸売・小売業
		11	金融・保険業
		12	不動産業
		13	飲食店・宿泊業
		14	医療・福祉
		15	教育・学習支援業
		16	複合サービス業
		17	サービス業(他に分類されないもの)
		18	公務(他に分類されないもの)
		VV	分類不能の産業
(有業者・本業・職業) 大分類	有業者	01	専門的・技術的職業従事者
		02	管理的職業従事者
		03	事務従事者
		04	販売従事者
		05	サービス職業従事者
		06	保安職業従事者
		07	農林漁業作業者
		08	運輸・通信従事者
		09	技能工，採掘・製造・建設作業及び労務従事者
		VV	分類不能の職業
(有業者・本業) 従業者規模	有業者	01	1～4人
		02	5～9人
		03	10～19人
		04	20～29人
		05	30～49人
		06	50～99人
		07	100～299人
		08	300～499人
		09	500～999人
		10	1000人以上
		11	官公庁
		VV	不詳
(有業者・本業) 年間就業日数	有業者	1	50日未満
		2	50～99日
		3	100～149日
		4	150～199日
		5	200～249日
		6	250日以上
		V	不詳
(有業者・本業) 週間就業時間	有業者で年間就業日数が200日未満で就業の規則性がだいたい規則的及び年間就業日数が200日以上	01	15時間未満
		02	15～19時間
		03	20～21時間
		04	22～29時間
		05	30～34時間
		06	35～42時間
		07	43～45時間
		08	46～48時間
		09	49～59時間
		10	60時間以上
		VV	不詳

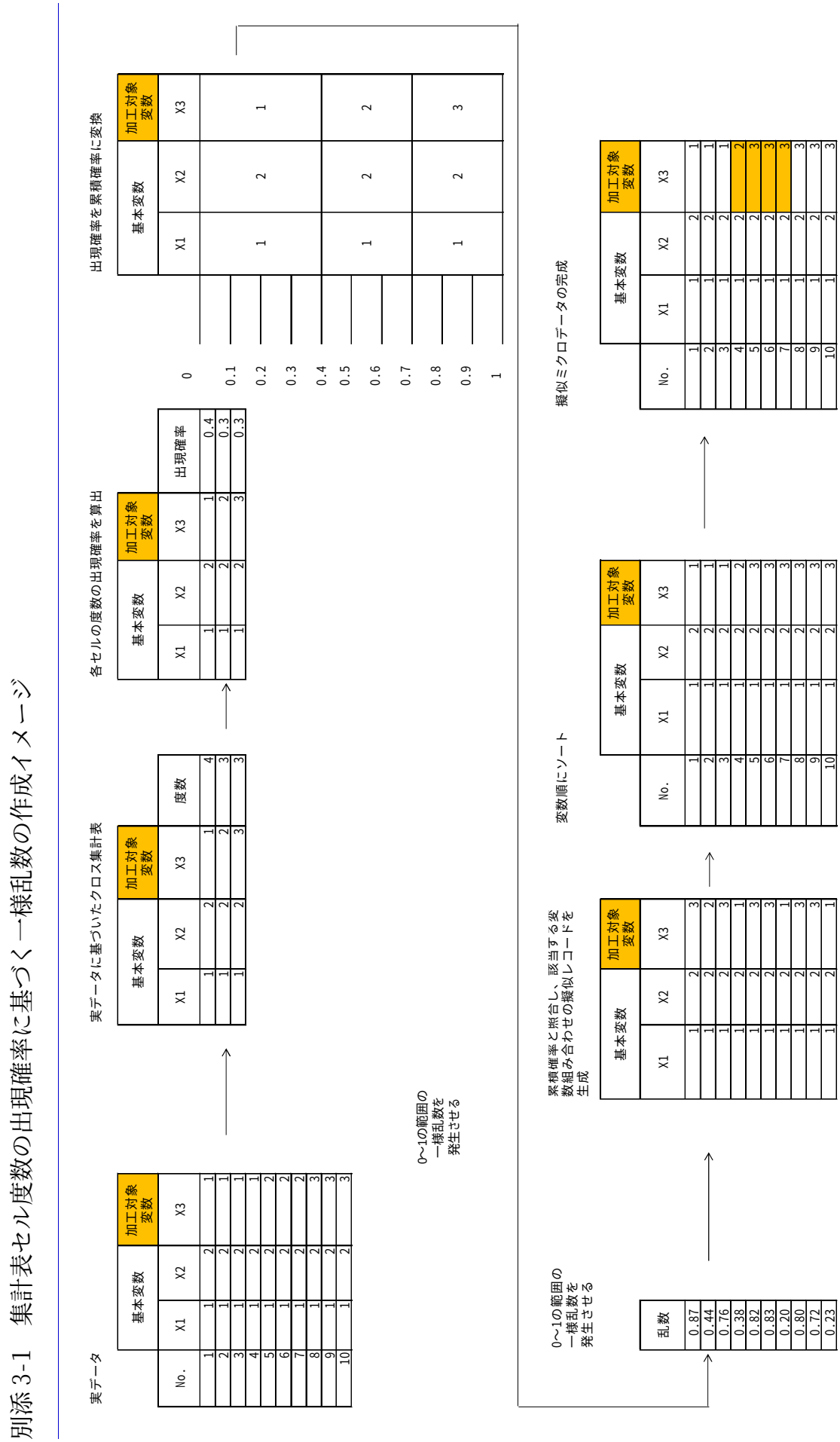
③　乱数発生変数

項目名	符号	符号内容	項目名	符号	符号内容
個人所得	01	収入なし，５０万円未満	継続就業期間	01	１ヶ月未満
	02	５０～９９万円		02	１ヶ月
	03	１００～１４９万円		03	２ヶ月
	04	１５０～１９９万円		04	３ヶ月
	05	２００～２４９万円		05	４ヶ月
	06	２５０～２９９万円		06	５ヶ月
	07	３００～３９９万円		07	６ヶ月
	08	４００～４９９万円		08	７ヶ月
	09	５００～５９９万円		09	８ヶ月
	10	６００～６９９万円		10	９ヶ月
	11	７００～７９９万円		11	１０ヶ月
	12	８００～８９９万円		12	１１ヶ月
	13	９００～９９９万円		13	１～２年１１ヶ月
	14	１０００～１４９９万円		14	３～４年１１ヶ月
	15	１５００万円以上		15	５～９年
				16	１０～１４年
				17	１５～１９年
				18	２０～２４年
				19	２５～２９年
				20	３０年以上
				21	１年未満（月不詳）

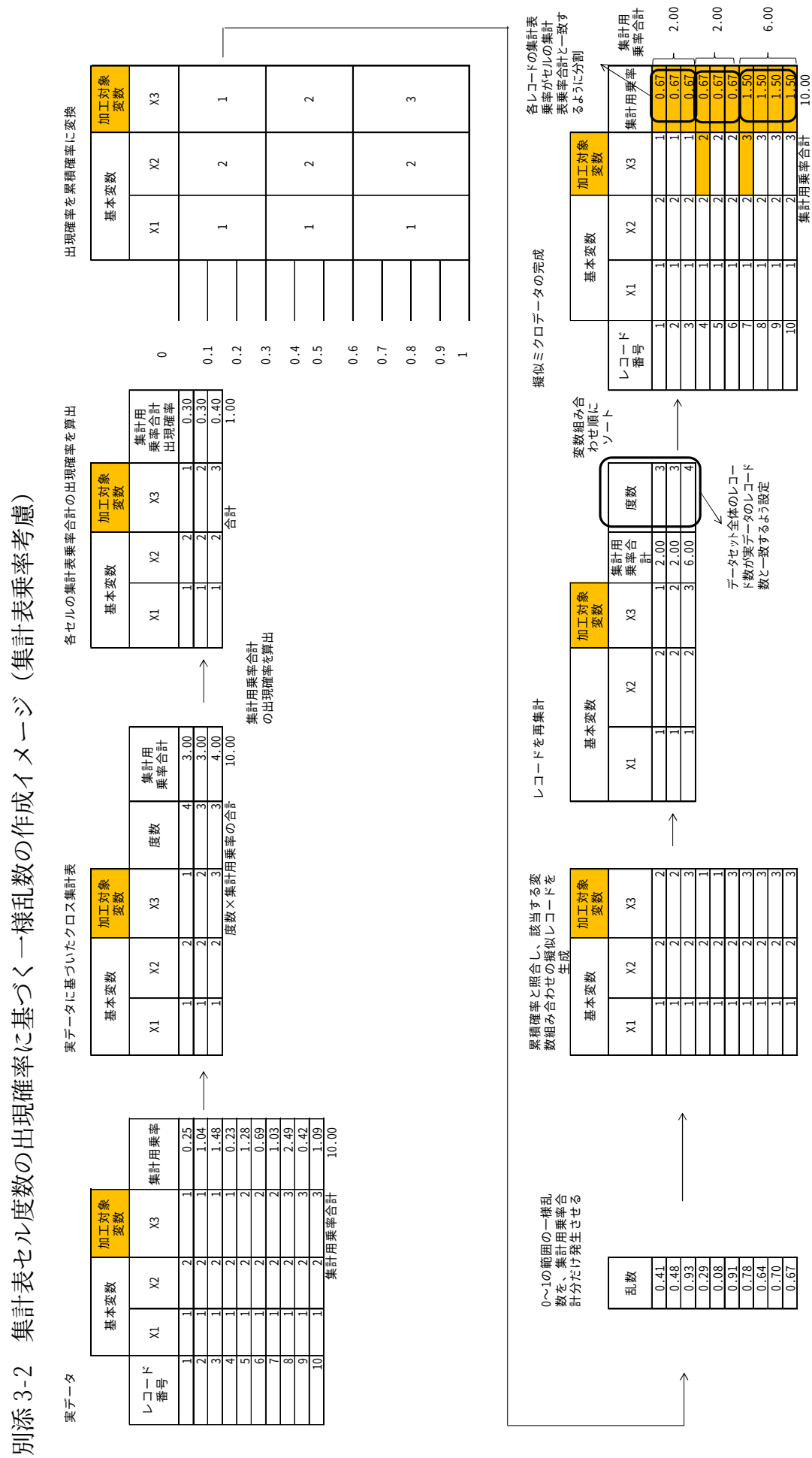
④　集計用乗率

項目名	対象	符号	符号内容
集計用乗率		000000.00 0000～	7桁目は小数点。以下6桁は小数点以下を表す。

別添 3-1　集計表セル度数の出現確率に基づく一様乱数の作成イメージ



別添 3-2　集計表セル度数の出現確率に基づく一様乱数の作成イメージ（集計表乗率考慮）



別添４　一様乱数発生における変数固定の有無の検証

集計表の一部の変数を固定し、その範囲で本文 3.1.3 の手順による一様乱数を発生させれば、実データの特徴をより反映できると仮定し、

① 高次元クロス集計表（15 変数）の組み合わせを固定し、残りの変数について乱数を発生（以下「15 変数固定」）

② 基本変数（8 変数）の組み合わせを固定し、残りの変数について乱数を発生（以下「8 変数固定」）

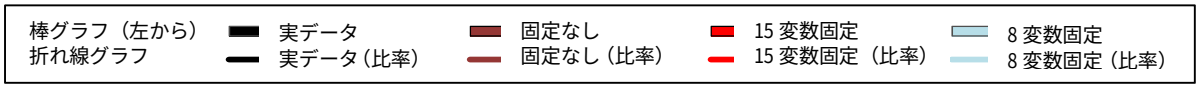
以上の２パターンの追加検証を行った。

この追加検証において使用したデータ及び変数は本文 3.1.1 と、集計表は本文 3.1.2 と同一のものであり、固定変数と乱数発生変数の設定は以下のとおりである。

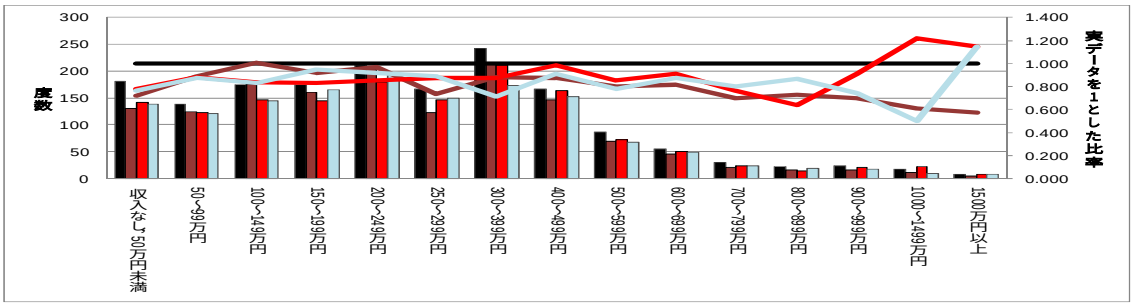
変数名		（参考）固定なし		15変数固定		8変数固定	
		固定	乱数発生	固定	乱数発生	固定	乱数発生
基本変数	性別		○	○		○	
	配偶者の有無		○	○		○	
	続き柄		○	○		○	
	教育（教育区分）		○	○		○	
	教育（学校区分）		○	○		○	
	年齢（5歳階級）		○	○		○	
	1年前との就業異動		○	○		○	
	前職の有無		○	○		○	
加工対象変数	従業上の地位（5区分）		○	○			○
	雇用形態		○	○			○
	産業（大分類）		○	○			○
	職業（大分類）		○	○			○
	従業者規模		○	○			○
	年間就業日数		○	○			○
	週間就業時間		○	○			○
個人所得			○		○		○
継続就業期間			○		○		○

作成した擬似ミクロデータを実データと比較検証した結果、以下のとおり、クロス統計表の度数分布及び独立性のカイ二乗検定のいずれの結果からも、変数の固定の有無による明確な影響は認められなかった。

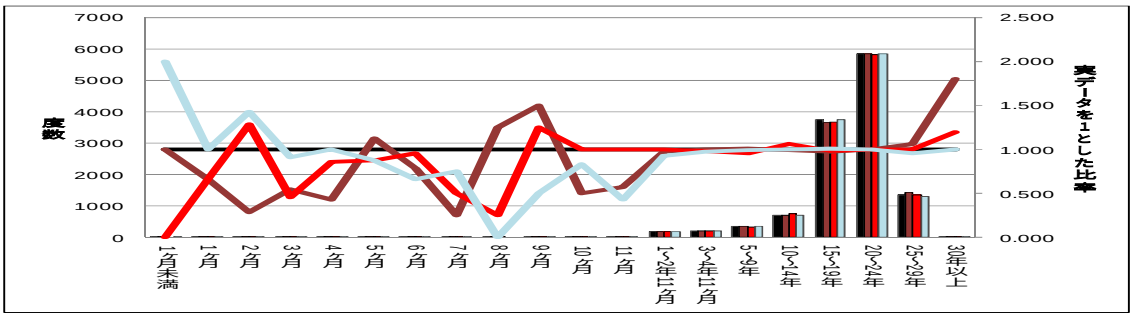
＜グラフの凡例＞



個人所得　製造業×前職あり×自営業主



継続就業期間　40～44 歳×前職なし×雇用者



独立性のカイ二乗検定

検定対象変数		実データ×固定なし			実データ×15変数固定			実データ×8変数固定		
		カイ二乗統計量	自由度	p値	カイ二乗統計量	自由度	p値	カイ二乗統計量	自由度	p値
個人所得	自営業主×前職あり	4.8401	14	0.9879	8.7625	14	0.8460	6.7895	14	0.9425
	自営業主×前職なし	11.3582	14	0.6577	8.7022	14	0.8496	13.6727	14	0.4744
	雇用者×前職あり	32.9551	14	0.0029	18.4491	14	0.1871	14.2894	14	0.4284
	雇用者×前職なし	19.3856	14	0.1507	18.8247	14	0.1718	14.1053	14	0.4419
継続就業期間	自営業主×前職あり	23.0714	19	0.2342	25.8269	19	0.1351	36.8588	19	0.0083
	自営業主×前職なし	37.7853	19	0.0063	44.3986	19	0.0008	44.0053	19	0.0009
	雇用者×前職あり	18.6413	19	0.4801	18.8610	19	0.4658	20.5444	19	0.3625
	雇用者×前職なし	14.8047	19	0.7349	14.3196	19	0.7647	15.3661	19	0.6991

別添５　集計表における加工処理変数の度数分布の変化

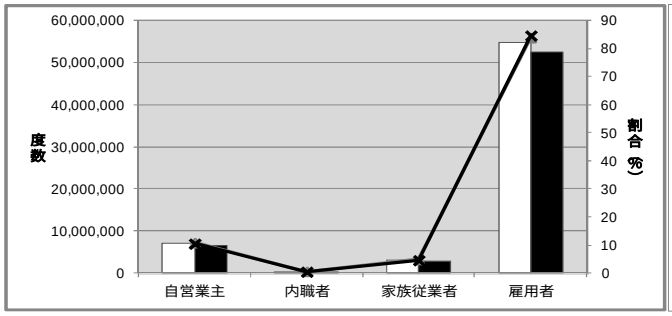
集計表作成時に不詳置き換え処理を行う加工処理変数は、集計表作成後に不詳以外の度数が減少し、特に不詳置き換え処理の優先順位が高い「週間就業時間」で大幅な減少が見られた（以下の棒グラフを参照）。

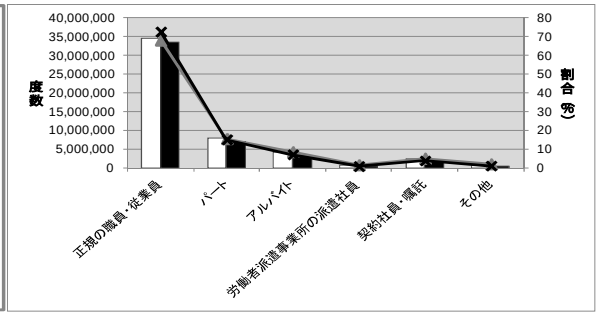
しかし、不詳以外の区分の度数合計を１とする割合（%）については、以下の折れ線グラフのとおり、いずれの区分も公表結果表の分布と近似した。

＜グラフの凡例＞

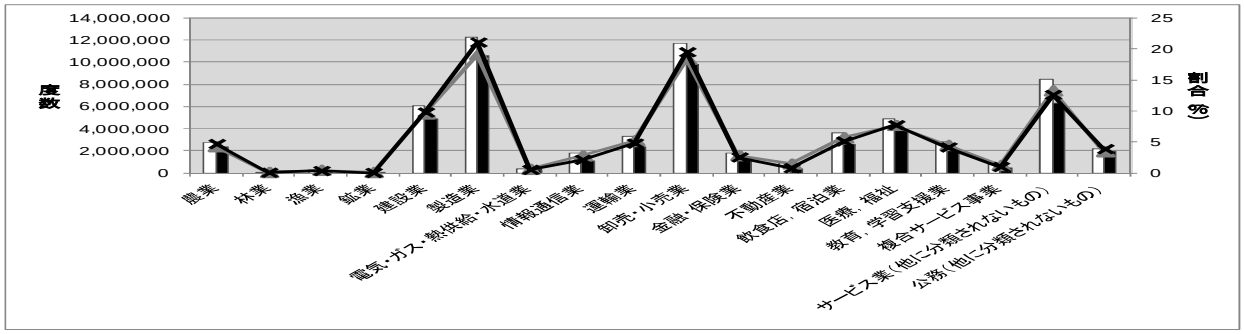
棒グラフ：　　白：公表結果表、黒：集計表 折れ線グラフ：黒：公表結果表、灰色：集計表

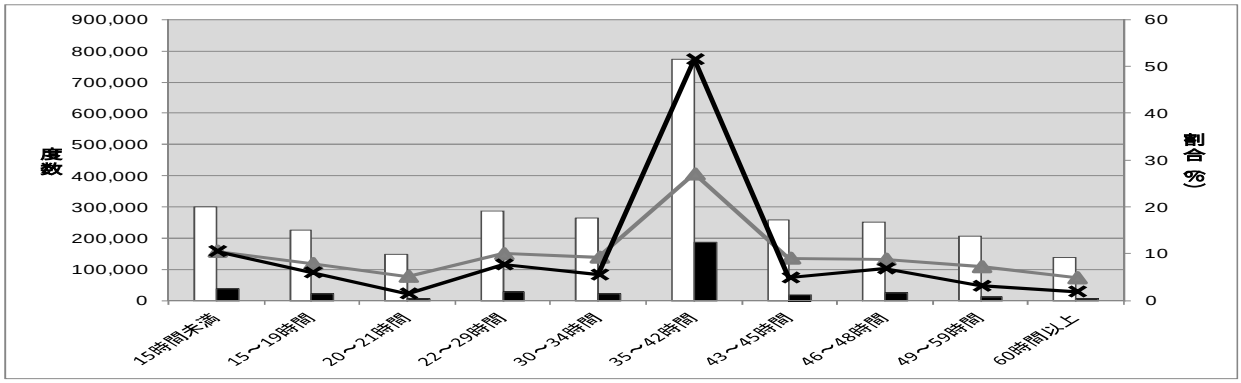
従業上の地位別（左）及び雇用形態別（右）有業者数





産業別（上）及び週間就業時間別（下）有業者数





参考文献

伊藤伸介（2008）「ミクロアグリゲーションに関する研究動向」『製表技術参考資料』No.10
伊藤伸介・磯部祥子・秋山裕美(2008)「匿名化技法としてのミクロアグリゲーションの有効性に関する研究－全国消費実態調査を例に－」『製表技術参考資料』No.10
伊藤伸介・磯部祥子・秋山裕美(2009)「秘匿性の評価方法に関する実証研究 －全国消費実態調査のミクロアグリゲートデータを用いて－」 『製表技術参考資料』No.11
槙田直木(2012)「擬似ミクロデータの試行提供」平成 24 年度公的統計のミクロデータの利用に関する研究集会・報告資料
秋山裕美・山口幸三・伊藤伸介・星野なおみ・後藤武彦(2012) 「教育用擬似ミクロデータの開発とその利用～平成 16 年全国消費実態調査を例として～」 『製表技術参考資料』No.16
P.スペクター著　石田基広・石田和枝訳(2012)　「R データ自由自在」丸善出版
金明哲編　藤井良宜著(2010)「R で学ぶデータサイエンス 1　カテゴリカルデータ解析」共立出版
青木繁伸(2009)　「R による統計解析」オーム社
舟尾暢男(2009)　「The R Tips 第 2 版　－データ解析環境 R の基本技・グラフィックス活用集－」オーム社
統計委員会（2009）「第 20 回統計委員会議事録」
統計委員会匿名データ部会（2009）「第 1 回匿名データ部会議事概要」
総務省政策統括官（統計基準担当）(2011)「平成 22 年度　統計法施行状況報告」
総務省政策統括官（統計基準担当）(2014)「公的統計の整備に関する基本的な計画」
総務省政策統括官（統計基準担当）(2014)「委託による統計の作成等及び匿名データの作成・提供に関する年度計画一覧（平成 26 年度）」
総務省統計局（2006）「平成 16 年全国消費実態調査報告　第 1 巻　家計収支編」
総務省統計局（2004）「平成 14 年就業構造基本調査報告　全国編」

製 表 技 術 参 考 資 料 29
平成 27 年 3 月発行

編集・発行　　独立行政法人　統計センター

〒162－8668
東京都新宿区若松町 19－1
電　話　代表　03 (5273) 1200

統計情報・技術部統計技術研究課　　TEL : 03－5273－1368
E-mail : research@nstac.go.jp