

# 一般用マイクロデータの 試作データ(詳細版)作成手法

～平成21年全国消費実態調査  
に基づく擬似データ～

統計センター 伊原 一  
総務省統計局 平澤 鋼一郎

平成28年12月10日  
研究会発表資料

信頼に応じて作る統計表

National  
Statistics  
Center

## 発表内容

### 1 . 一般用ミクロデータ

#### < 試作データ(詳細版) >

- 2 . 試作データ(詳細版)の仕様
- 3 . 特化係数による作成手法
- 4 . 按分計算結果の乱数化

#### < 分析結果 >

- 5 . 構成比グラフ
- 6 . 特化係数グラフ
- 7 . 410品目分類の相関
- 8 . 10大費目間の相関
- 9 . 変数記号と品目分類の対応表 ( 抜粋 )

### 10 . まとめ

# 1 . 一般用マイクロデータ

既存結果表とのバランスを考慮した統計量に基づく集計表を作成、公表したのち、その統計量に基づき乱数を発生させて作成した、マイクロデータ形式のデータ

- ・平成21年全国消費実態調査に基づく度数、平均及び標準偏差について集計したものを公表（統計局）
- ・年間収入、消費支出、10大費目の12項目
- ・対数正規乱数による擬似データ作成
- ・平成28年3月提供開始

一般用マイクロデータについて（総務省統計局URL）  
<<http://www.stat.go.jp/useful/ippanmicro.htm>>

一般用マイクロデータの利用（統計センターURL）  
<<http://www.nstac.go.jp/services/ippan-microdata.html>>


# 1 . 一般用ミクロデータ

## 一般用ミクロデータの改善

相関係数反映版（改訂版）を公表予定  
（年内見込）

## 一般用ミクロデータの改善

特化係数を反映した按分計算による410品目  
分類の詳細版の作成手法について研究

 本日は、改善 について  
試作データ(詳細版)の作成方法等を紹介

## < 試作データ(詳細版) >

### 2 . 試作データ(詳細版)の仕様

作成した試作データ(詳細版)の仕様は下記のとおり。

レコード数	約46,000
ファイル形式	C S V
ファイル容量	約80MB
基データ	平成21年全国消費実態調査(集計表)
世帯属性	全世帯
属性項目	3大都市圏か否か、世帯人員、有業人員、住宅所有、 年齢5歳階級、65歳以上か否か
数量項目	復元ウェイト、年間収入、消費支出、10大費目、 410品目分類
統計量	平均と標準偏差は近似的に一致 (対数正規乱数による誤差を含む)
支出傾向	世帯属性別特化係数により近似的に一致
相関係数	あり(擬似的相関)

### 3 . 特化係数による作成手法

世帯属性別の支出品目の構成比について特化係数に着目し、10大費目から小品目の按分計算を行う際の按分比率に世帯属性別の特化係数の最大値及び最小値を反映することにより支出の特徴が反映されるようにデータを作成

#### < 特化係数 >

地域分析を行う際によく用いられる計算方法

特化係数が1.0より大きければ構成比が全体平均よりも高く、逆に1.0より小さければ構成比が全体平均よりも低いことを示す。

### 3 . 特化係数による作成手法

#### 構成比計算に用いたデータ（平成21年全国消費実態調査）

世帯属性	報告書結果表（品目編）
3大都市圏か否か	表1 地域，品目別1世帯当たり1か月間の支出
世帯人員	表4 世帯主の職業・世帯人員・世帯類型・主な年間収入の種類，品目別1世帯当たり1か月間の支出
年齢階級（勤労者世帯） + 無職世帯	表3 年間収入五分位階級・世帯主の年齢階級，購入形態，品目別1世帯当たり1か月間の支出
	表4 世帯主の職業・世帯人員・世帯類型・主な年間収入の種類，品目別1世帯当たり1か月間の支出
年間収入階級	表2 年間収入階級，品目別1世帯当たり1か月間の支出
世帯属性	報告書結果表（家計収支編）
住居所有 （住宅品目のみ）	表23 住居の所有関係別1世帯当たり1か月間の収入と支出

### 3 . 特化係数による作成手法

#### 按分比率の計算方法

(1) 世帯属性別に各品目構成比の特化係数を計算

住居以外の品目

3 大都市圏か否か(2 区分)

世帯人員(2 区分)

年齢 ( 勤労者世帯 9 区分 )

+ 無職世帯(1 区分)

年間収入(19区分)

住居

上記に 住宅所有(2 区分)を追加



### 3 . 特化係数による作成手法

#### 按分比率の計算方法

##### (2) 複合特化係数を計算

- ・ 世帯属性別特化係数が全て1.0以上の場合  
最大値 × 1.0 ( 基準値1.0を含めて最大値判定 )
- ・ 世帯属性別特化係数が全て1.0未満の場合  
最小値 × 1.0 ( 基準値1.0を含めて最小値判定 )
- ・ 最大値が 1 以上、最小値が1.0未満の場合  
最大値 × 最小値

##### (3) 構成比の合計補正

全体構成比に複合特化係数を乗じて世帯属性別の費目構成比を求めた上で合計が100%になるように補正

### 3 . 特化係数による作成手法

計算例 ( 参考 )

( 1 ) パン ( E004 ) の特化係数 [ 最大値 ]

1.00000 基準値

1.06542 3大都市圏

( + 店舗が多くパンを買いやすい )

1.05623 3人以上世帯

( + パン食の子供がいる )

1.05614 30 ~ 39歳就業

( + パン食が多い若い世代 )

 1.09671 ~ 300万円

( + 主食比率が高い )

複合特化係数  $1.00000 \times 1.09671 = 1.09671$

### 3 . 特化係数による作成手法

計算例 ( 参考 )

( 2 ) パン ( E004 ) の特化係数 [ 最小値 ]

1.00000 基準値

0.95361 3大都市圏外

( - 店舗が少なくパンを買いにくい )

 0.89119 2人世帯

( - パン食の子供がいない )

0.94971 65歳以上就業

( - 高齢でパン食が少ない )

0.93962 ~ 1000万円

( - 主食比率が低い )

複合特化係数  $1.00000 \times 0.89119 = 0.89119$

### 3 . 特化係数による作成手法

計算例 ( 参考 )

( 3 ) 米 ( E003 ) の特化係数 [ 最小値 × 最大値 ]

1.00000 基準値

➡ 0.84528 3大都市圏  
( - 店舗が多くパンを買いやすい )

0.94142 3人以上世帯  
( - パン食の子供がいる )

➡ 1.65527 65歳以上就業  
( + 高齢でパン食が少ない )

1.48744 ~ 300万円  
( + 主食比率が高い )

複合特化係数  $1.65527 \times 0.84528 = 1.39917$

## 4 . 按分計算結果の乱数化

特化係数を反映した按分比率410品目分類により  
簡易データから按分値を作成し、乱数化

按分値の乱数化

(1) 按分値を対数変換

(2) 世帯属性別に按分値の平均と標準偏差を求め、期待値がこの平均と標準偏差に一致する対数正規乱数を作成

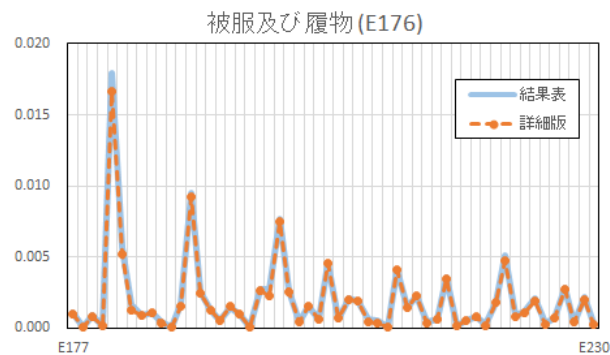
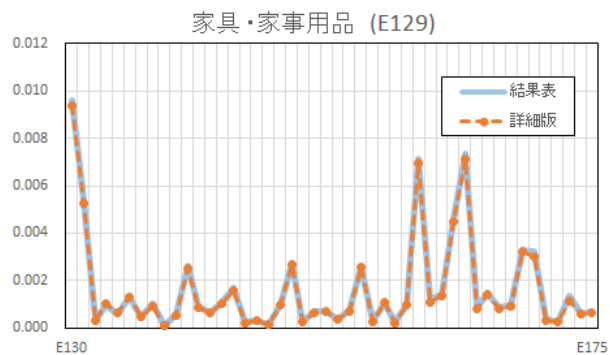
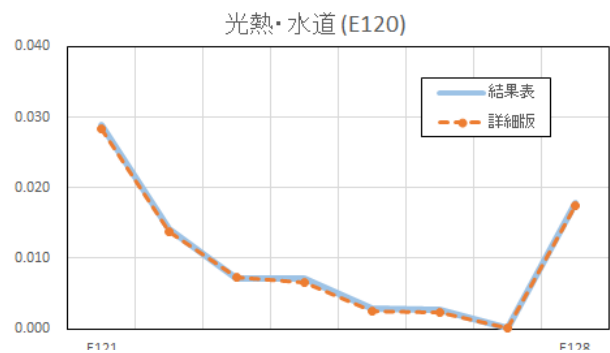
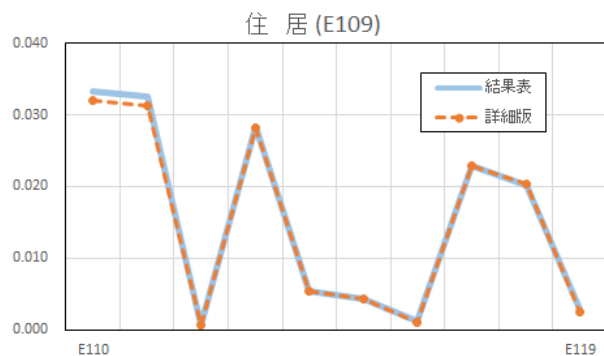
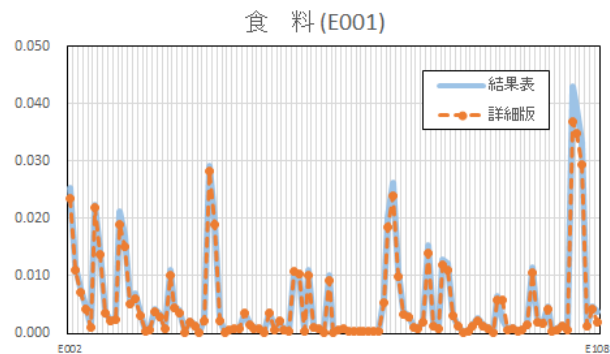
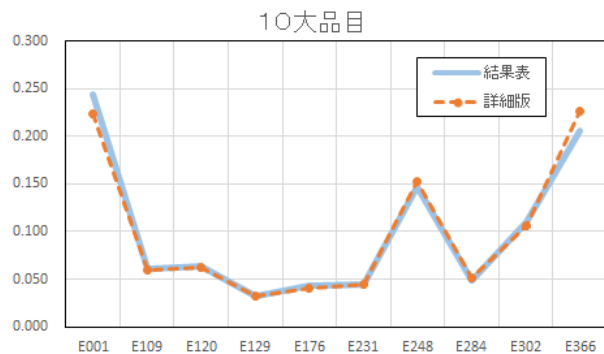
(3) 逆対数変換により乱数化按分値を作成

< 分析結果 >

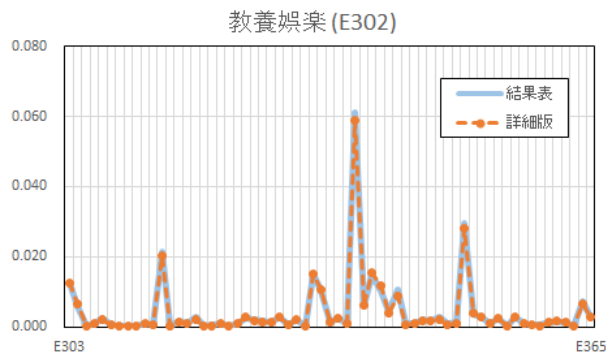
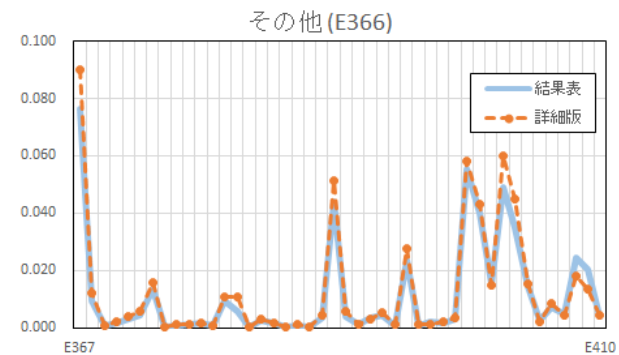
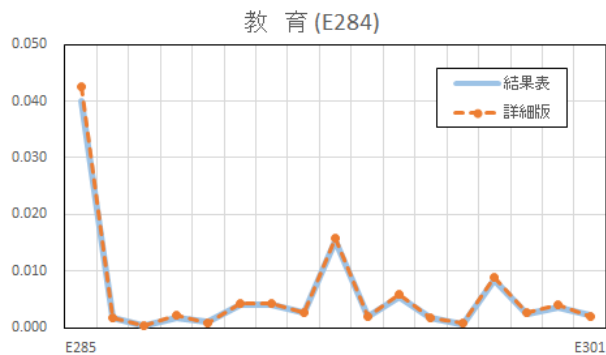
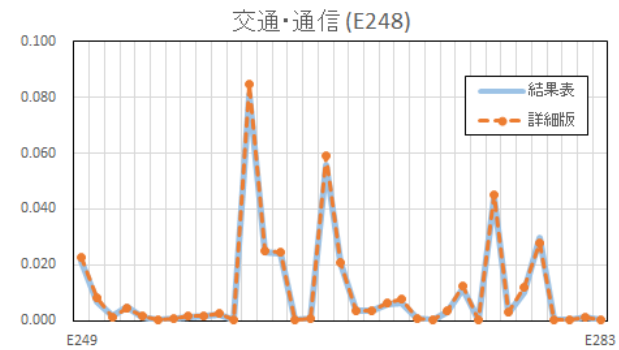
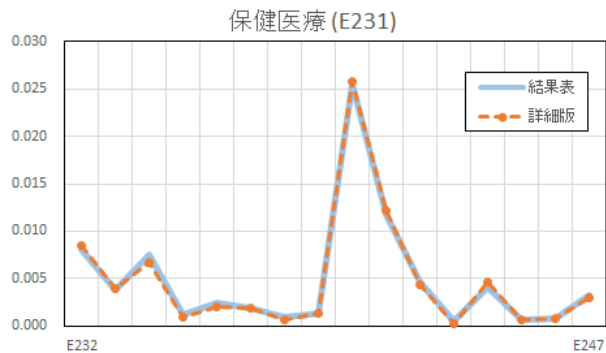
## 5 . 構成比グラフ

410品目分類の按分計算を行った構成比の結果グラフから、世帯属性の合計レベルでは410品目分類の平均支出額の構成比は報告書結果表と概ね同様の値になることが確認できる。

# 5 . 構成比グラフ



# 5 . 構成比グラフ



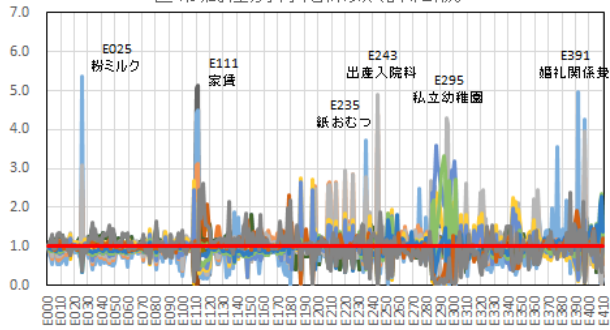


## 6 . 特化係数グラフ

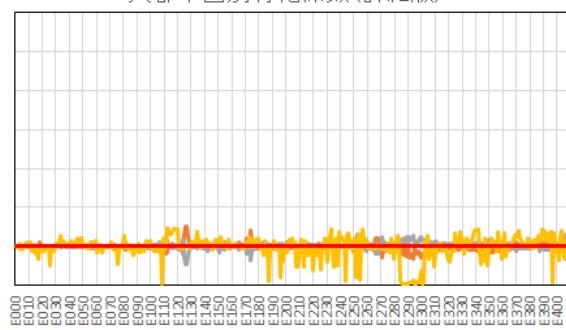
410品目分類の詳細データ(試作版)について特化係数の集計を行った結果グラフから、世帯属性別の410品目分類の特化係数について報告書結果表の傾向が反映されていることが確認できる。

# 6 . 特化係数グラフ

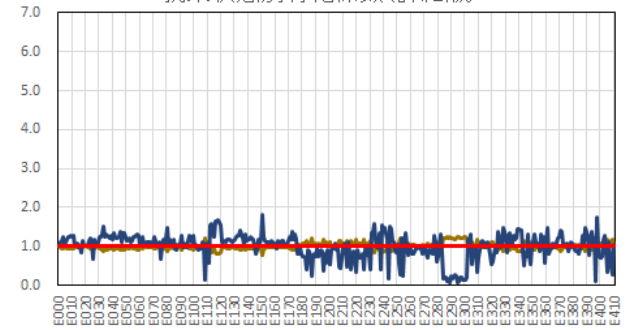
世帯属性別特化係数(詳細版)



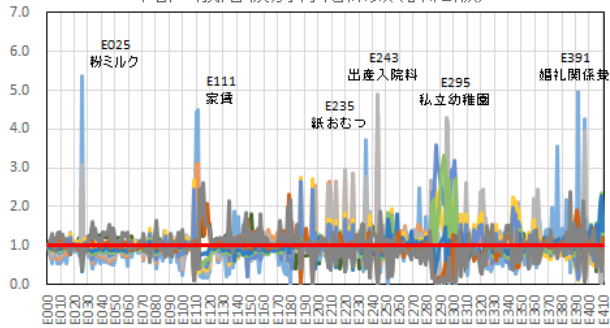
3大都市圏別特化係数(詳細版)



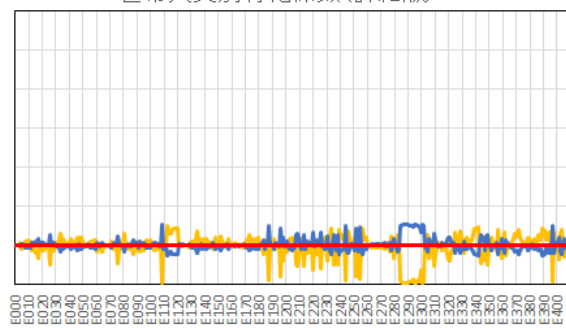
就業状態別特化係数(詳細版)



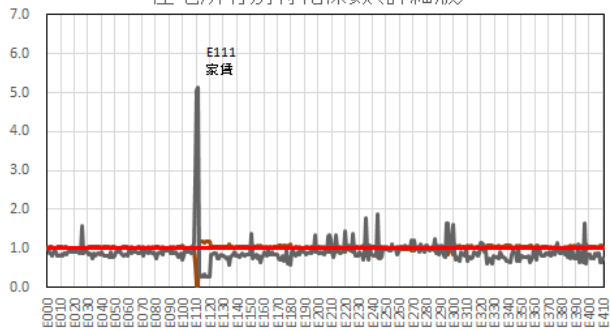
年齢5歳階級別特化係数(詳細版)



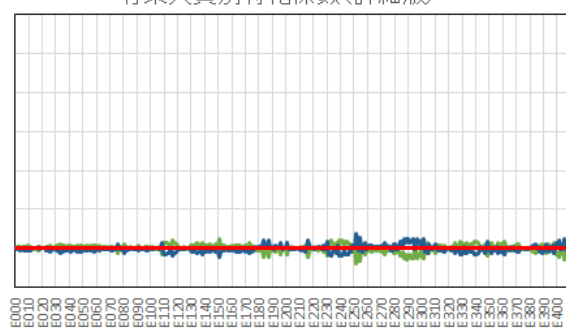
世帯人員別特化係数(詳細版)



住宅所有別特化係数(詳細版)



有業人員別特化係数(詳細版)



## 7 . 410品目分類の相関

特化係数を反映させた按分計算結果の乱数化により、世帯属性の合計レベルでは410品目分類間で相関が擬似的に再現されることが観察される。

注) 各品目とその品目を含む小計との相関については自己相関になるため、相関係数の計算から除外

## 7 . 410品目分類の相関

### 7 . 1 粉ミルク（E025）との相関係数

特化係数が最も高い粉ミルクに着目

#### 正の相関

保育所費用、出産入院料、子供靴などの育児  
関連の費目

#### 負の相関

授乳に影響するウイスキー、焼酎、清酒などの  
酒類、緑茶やつくだ煮、漬物など高齢者世  
帯に多いと見られる品目

## 7 . 410品目分類の相関

### 7 . 1 粉ミルク（E025）との相関係数

#### 粉ミルクと正の相関

相関係数	変数名	品目コード	品目名
+0.48	E396	954	保育所費用
+0.38	E243	723	出産入院料
+0.35	E225	676	子供靴
+0.34	E108	39X	学校給食
+0.34	E219	657	子供用靴下
+0.33	E208	596	子供用シャツ
+0.33	E213	620	子供用下着類
+0.29	E319	836	テレビゲーム機
+0.29	E294	778	国公立幼稚園
+0.28	E321	837	他のがん具

## 7 . 410品目分類の相関

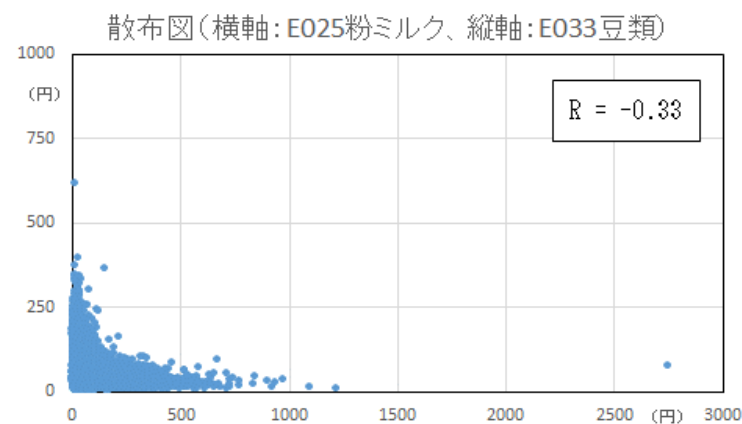
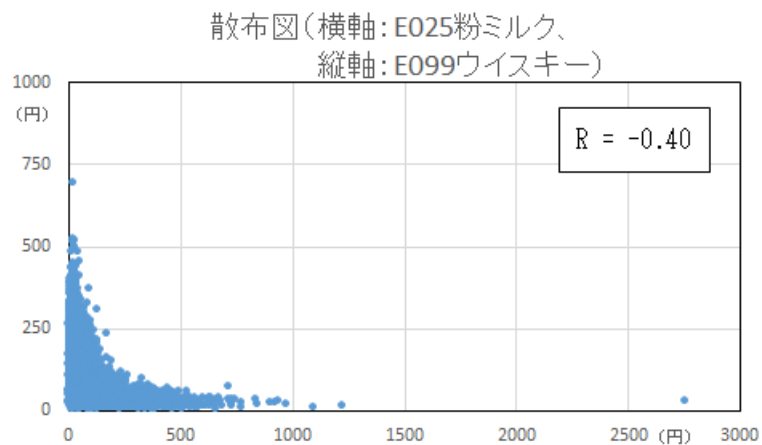
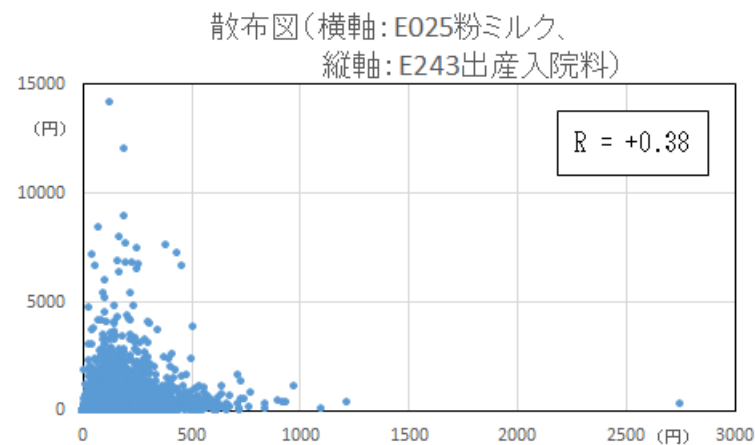
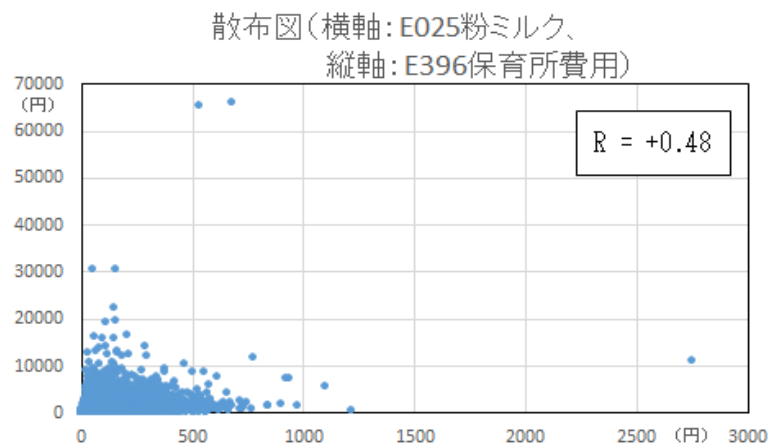
### 7 . 1 粉ミルク（E025）との相関係数

#### 粉ミルクと負の相関

相関係数	変数名	品目コード	品目名
-0.40	E099	3X5	ウイスキー
-0.33	E033	273	豆類
-0.33	E080	380	緑茶
-0.30	E045	296	野菜・海藻のつくだ煮
-0.30	E048	300	生鮮果物
-0.27	E097	3X3	焼ちゅう
-0.27	E044	294	野菜の漬物
-0.26	E096	3X1	清酒
-0.26	E059	327	酢
-0.26	E008	170	生鮮魚介

## 7 . 410品目分類の相関

### 7 . 1 粉ミルク (E025) との相関係数



## 7 . 410品目分類の相関

### 7 . 2 粉ミルク以外の相関係数

#### 正の相関

果実・野菜ジュースと他の飲料のその他、  
冷凍調理食品と学校給食 等

#### 負の相関

学校給食や保育所費用などの育児関連の品目  
と、豆類、緑茶、つくだ煮などの品目の組み  
合わせ



## 7 . 410品目分類の相関

### 7 . 2 粉ミルク以外の相関係数

#### 正の相関（粉ミルク以外）

相関係数	変数名	品目名	
+0.67	E089E094	385果実・野菜ジュース	389他の飲料のその他
+0.52	E075E108	370冷凍調理食品	39X学校給食
+0.51	E075E089	370冷凍調理食品	385果実・野菜ジュース
+0.50	E017E075	22X合いびき肉	370冷凍調理食品
+0.49	E089E090	385果実・野菜ジュース	387炭酸飲料
+0.49	E017E108	22X合いびき肉	39X学校給食
+0.48	E017E089	22X合いびき肉	385果実・野菜ジュース
+0.48	E087E108	386ココア・ココア飲料	39X学校給食
+0.47	E087E089	386ココア・ココア飲料	385果実・野菜ジュース
+0.46	E089E092	385果実・野菜ジュース	38A乳飲料

## 7 . 410品目分類の相関

### 7 . 2 粉ミルク以外の相関係数

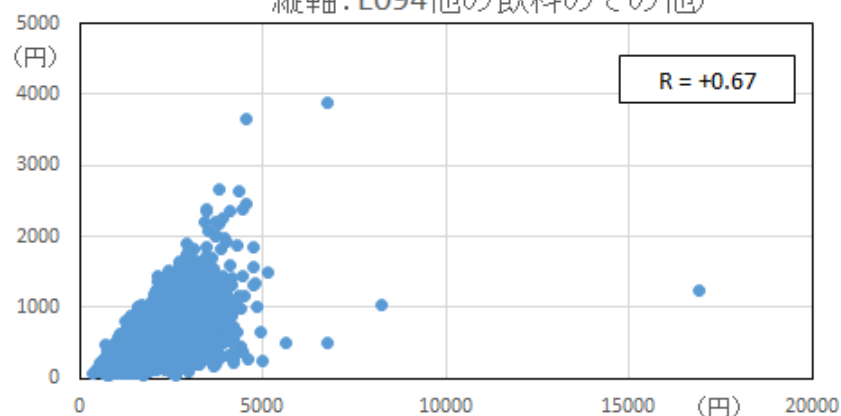
#### 負の相関（粉ミルク以外）

相関係数	変数名	品目名	
-0.36	E033E108	273豆類	39X学校給食
-0.35	E080E108	380緑茶	39X学校給食
-0.28	E099E396	3X5ウイスキー	954保育所費用
-0.27	E045E108	296野菜・海藻のつくだ煮	39X学校給食
-0.27	E033E396	273豆類	954保育所費用
-0.27	E080E396	380緑茶	954保育所費用
-0.25	E048E108	300生鮮果物	39X学校給食
-0.25	E033E219	273豆類	657子供用靴下
-0.24	E033E309	273豆類	811ビデオカメラ
-0.24	E033E225	273豆類	676子供靴

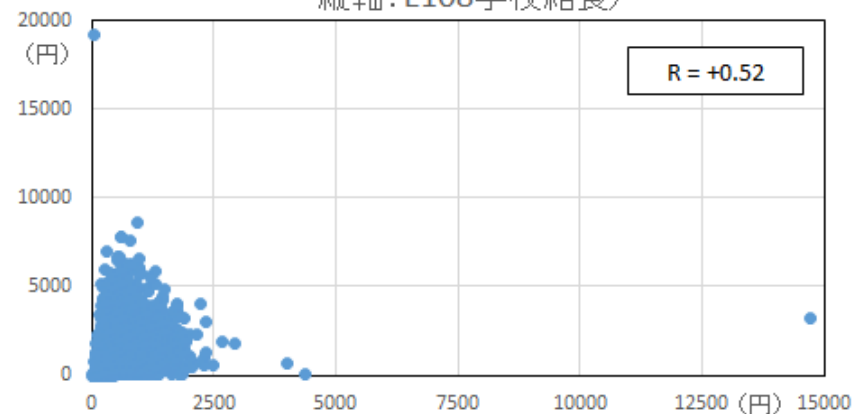
# 7 . 410品目分類の相関

## 7 . 2 粉ミルク以外の相関係数

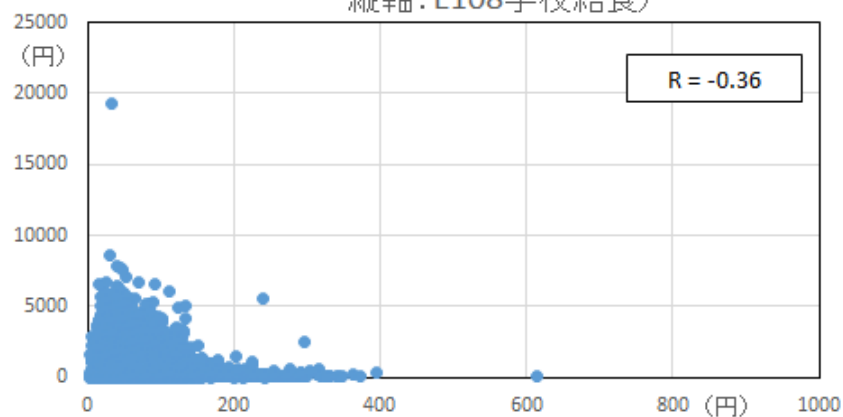
散布図(横軸: E089果実・野菜ジュース、  
縦軸: E094他の飲料のその他)



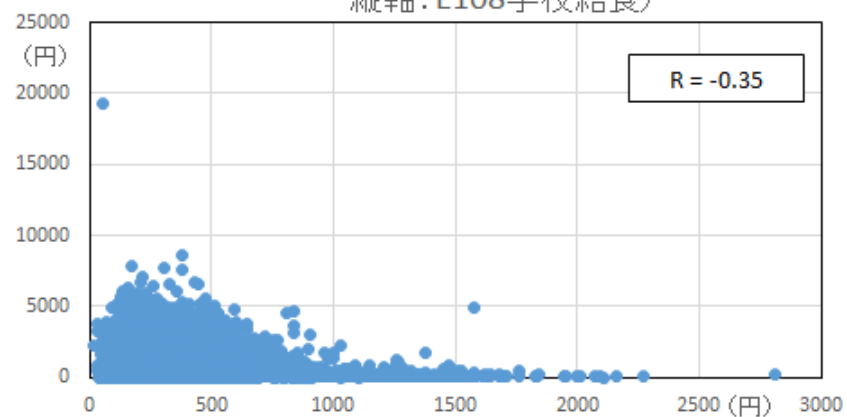
散布図(横軸: E075冷凍調理食品、  
縦軸: E108学校給食)



散布図(横軸: E033豆類、  
縦軸: E108学校給食)



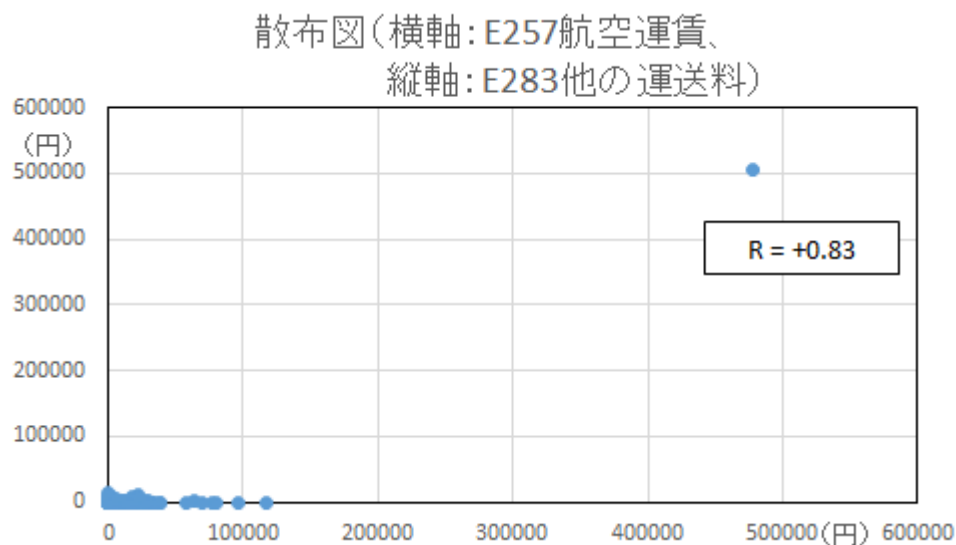
散布図(横軸: E080緑茶、  
縦軸: E108学校給食)



## 7 . 410品目分類の相関

### 7 . 3 外れ値の相関係数

本稿のデータは対数正規乱数により作成しているため、乱数の組み合わせによって相関係数に外れ値による特異値が発生する場合がある。



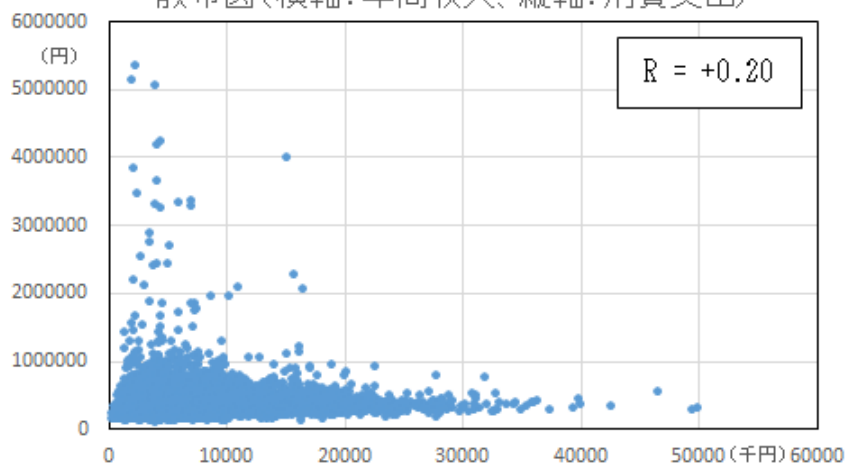
## 8 . 10大費目間の相関

10大費目間の相関については、相関係数は高いとはいえず、相関性の再現は不十分。相関係数と散布図（衣食住）は下記のとおり。

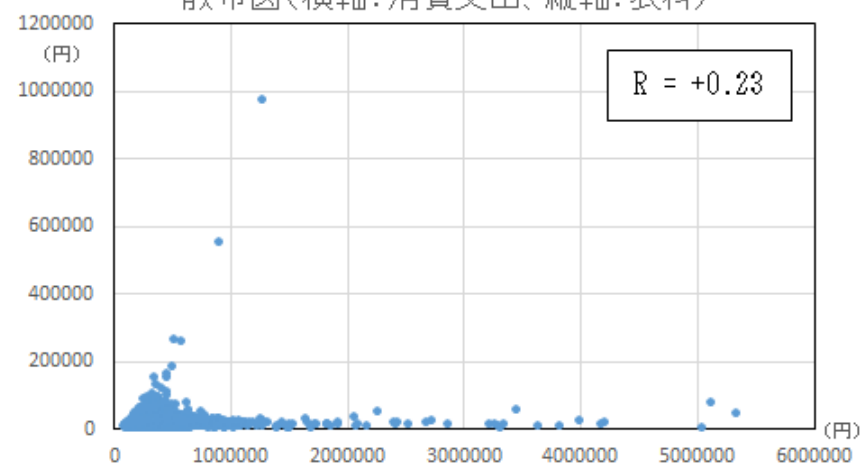
	Y_Income	L_Expenditure	Food	Housing	LFW	Furniture
Y_Income	1.00	0.20	0.32	-0.06	0.23	0.05
L_Expenditure	0.20	1.00	0.34	0.35	0.21	0.10
Food	0.32	0.34	1.00	-0.09	0.39	0.08
Housing	-0.06	0.35	-0.09	1.00	-0.10	-0.02
LFW	0.23	0.21	0.39	-0.10	1.00	0.07
Furniture	0.05	0.10	0.08	-0.02	0.07	1.00
Clothes	0.15	0.23	0.20	-0.01	0.09	0.03
Health	-0.02	0.09	0.03	-0.01	0.01	0.02
Transport	0.06	0.64	0.06	0.00	0.05	0.01
Education	0.15	0.41	0.23	-0.01	0.14	0.00
Recreation	0.09	0.19	0.22	-0.06	0.07	0.03
OL_Expenditure	0.15	0.53	0.16	-0.03	0.12	0.04

## 8 . 10大費目間の相関

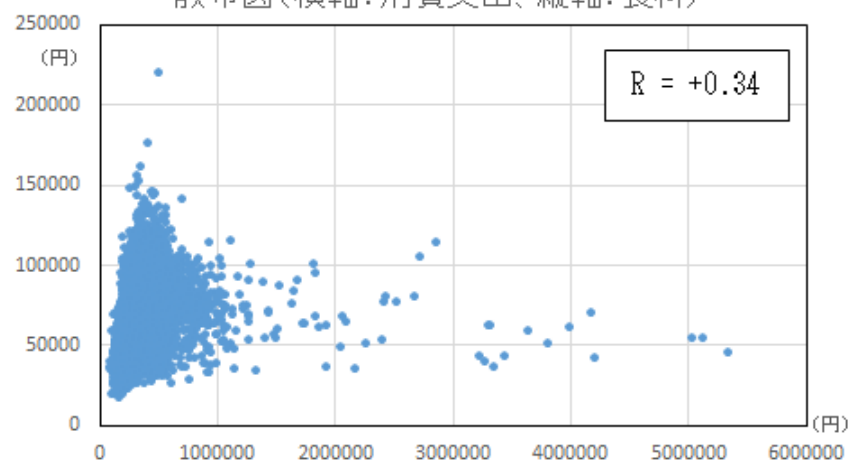
散布図(横軸:年間収入、縦軸:消費支出)



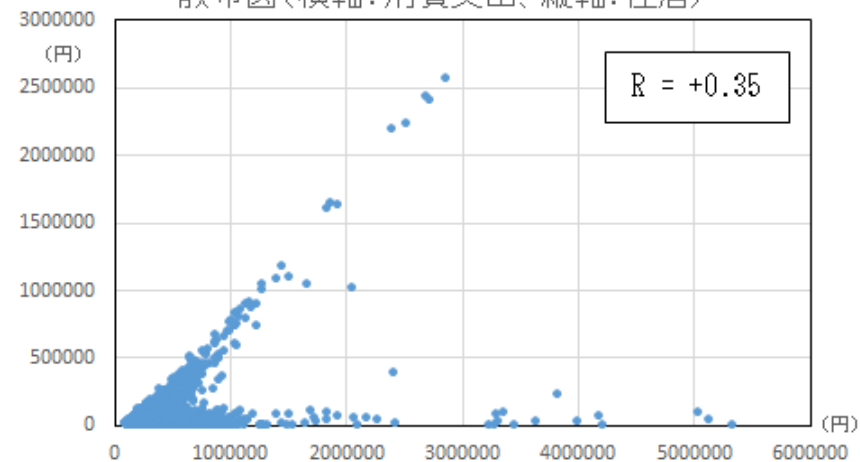
散布図(横軸:消費支出、縦軸:衣料)



散布図(横軸:消費支出、縦軸:食料)



散布図(横軸:消費支出、縦軸:住居)



## 9 . 変数記号と品目分類の対応表（抜粋）

E000	102 ~ 981	消費支出
E001	102 ~ 398・39X・39Y	食料
E109	400 ~ 429	住居
E120	430 ~ 440	光熱・水道
E129	451 ~ 543・54X	家具・家事用品
E176	550 ~ 694	被服及び履物
E231	700 ~ 729	保健医療
E248	730 ~ 769	交通・通信
E284	770 ~ 792	教育
E302	801 ~ 889・88A・88B・88X・88Y	教養娯楽
E366	890 ~ 981	その他の消費支出
E001	102 ~ 398・39X・39Y	食料
E002	102 ~ 140	穀類
E003	102	米
E004	120	パン
E005	130	めん類
E006	140	他の穀類
E007	170 ~ 213	魚介類
E008	170	生鮮魚介
E009	195	塩干魚介
E010	203	魚肉練製品

## 10 . まとめ

### 一般用ミクロデータの試作データ(詳細版)

- ・ 世帯属性別の平均と標準偏差
- ・ 410品目分類の品目構成比
- ・ 410品目分類について品目間の特化係数
- ・ 410品目分類間の相関係数

 上記について、擬似的な再現性を確認

### 課題

10大費目間の相関については再現性が不十分なため、さらに改善が必要。



## 10 . まとめ

### 従来の擬似データ

- 高次元集計表を集計して作成
- 乱数化によって秘匿性は確保
  
- 秘匿上の問題から高次元集計表は公表困難
- 未公表の集計表を基に作成したデータ
- 利用上の制約を受ける

➡ 公表されている統計表のみから擬似データを作成する手法の開発が必要

## 10 . まとめ

### 特化係数を反映させた乱数化按分値

- ・ 本研究において開発された新しい手法
- ・ 今後の擬似データ作成を進める上で有用な手法

本稿における作成手法は、公表結果のみを用いて特化係数を反映させた乱数化按分値により擬似データを作成することでこの問題を解消しており、統計教育等の目的に使用できるオープンな擬似データを作成する上で新たなルートを開拓できたという意味で、意義のあるものであるといえる。