

全国家計構造調査

「年平均推定値」の推定方法について

—10大費目への適用にむけて

共変量選択による修正と標準誤差等の考慮—

2021年6月25日

慶應義塾大学 経済学部／理化学研究所AIPセンター

星野崇宏

慶應義塾大学大学院経済学研究科

／理化学研究所AIPセンター

清水祐弥・二荒麟・松永将志

問題意識

全国消費実態調査（全国家計構造調査）のメリット／デメリット

サンプルサイズが大きい→詳細な区分での分析

2か月(前回3か月) 5年に一度

家計調査のメリット／デメリット

通年で結果が得られるローテーションパネル

比較的サンプルサイズが小さい

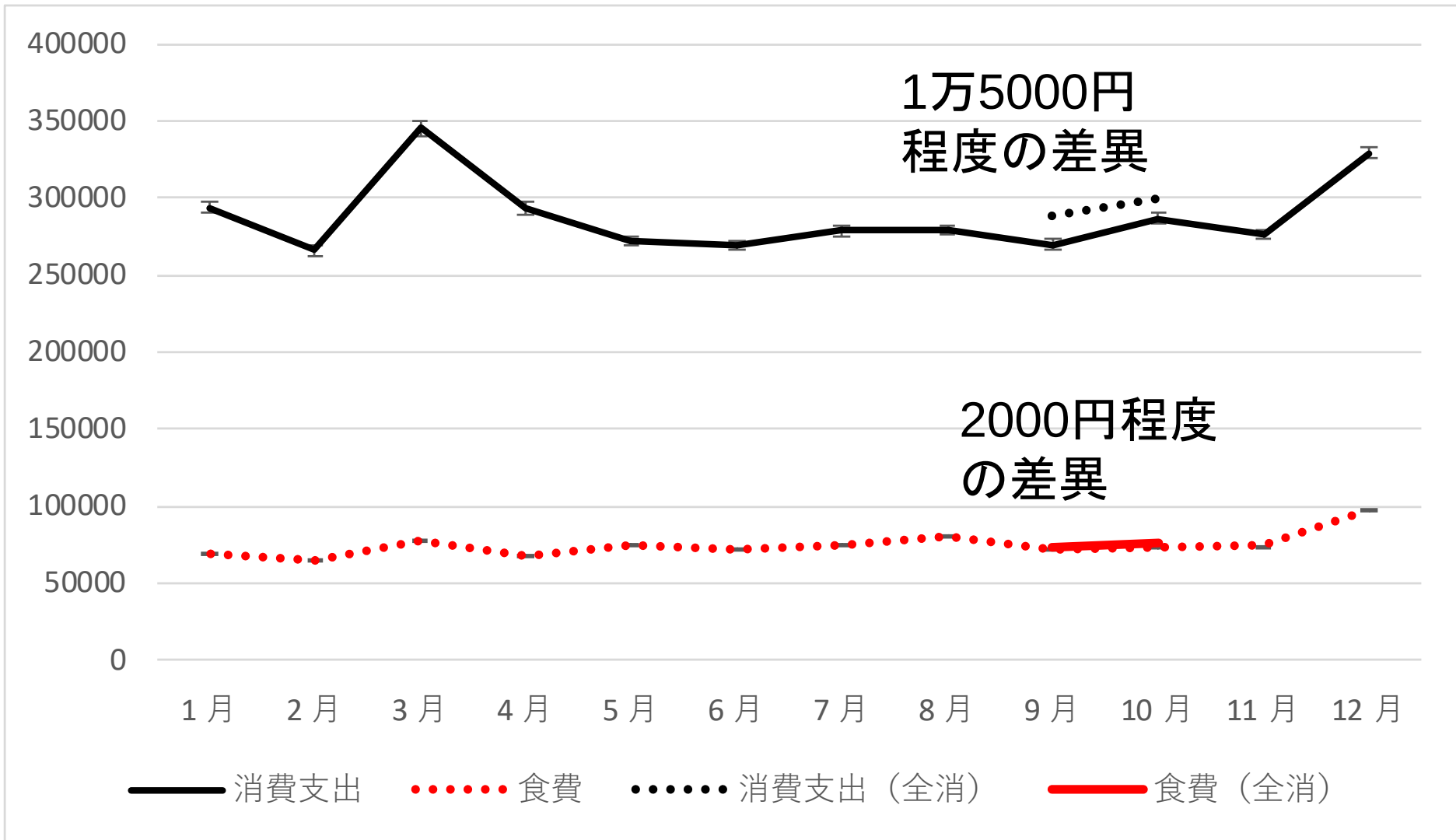
食費など月次の変動が大きいものについて両者を融合させて

通年での推測をしたい

○個票データでの補完 ×時系列データでの調整は不可能

⇒両調査の調査モードや標本の違いを反映し精度の高い推定

例：月次での支出の変動



* 消費や食費で家計調査が低いのは過小記入バイアスや儉約化か？
 (過小記入バイアス Deaton&Irish, 1984; 牧, 2007; 儉約化・調査疲れ Stephens&Unayama, 2011)

家計調査と家計構造(旧全消)の違い

“選択バイアス” = 回答集団の違い

“調査・データ取得モードの違い” = 取り方の違い(今回は質問)

⇒両者の違いが混ざっているので分離して議論したい

* 但し今回の両者の違いは非標本誤差ではなく標本誤差
 回答集団の違い

家計調査回答者

全国消費実態調査回答者

家計調査の 調査票	家計調査の結果	欠測
	調査モードの違い	
全消の調査票	欠測	全消の結果
補助変数・ 共変量	回答集団間の違いが生じる属性 (性年代・職種・収入等)	

本研究の目的

全国家計構造調査の年平均の推定

このために家計調査を補助情報として利用する

【家計調査と全消(以前の全国家計構造調査)の乖離の原因】

同じ月の比較で全体として全消の方が大

- ①標本の違い 収入 家計 < 全消 $\Rightarrow$ 全消の消費大
- ②季節性 全消の対象月は↓の傾向
- ③調査継続による効果 家計調査は6か月継続のため↓の傾向
- ④調査項目対象(財・サービスの対象)

*①が対象集団の違い②は時期の違い ③・④調査モードの違い

今回は②と③を調整

*厳密には回帰によって①も一定程度考慮する

分析手法(費目別)

「0」が含まれる割合を基準として分析手法を分類する。

「0」が含まれない/わずかな費目は重回帰分析、「0」が多く含まれる費目はTobit-Type2を利用

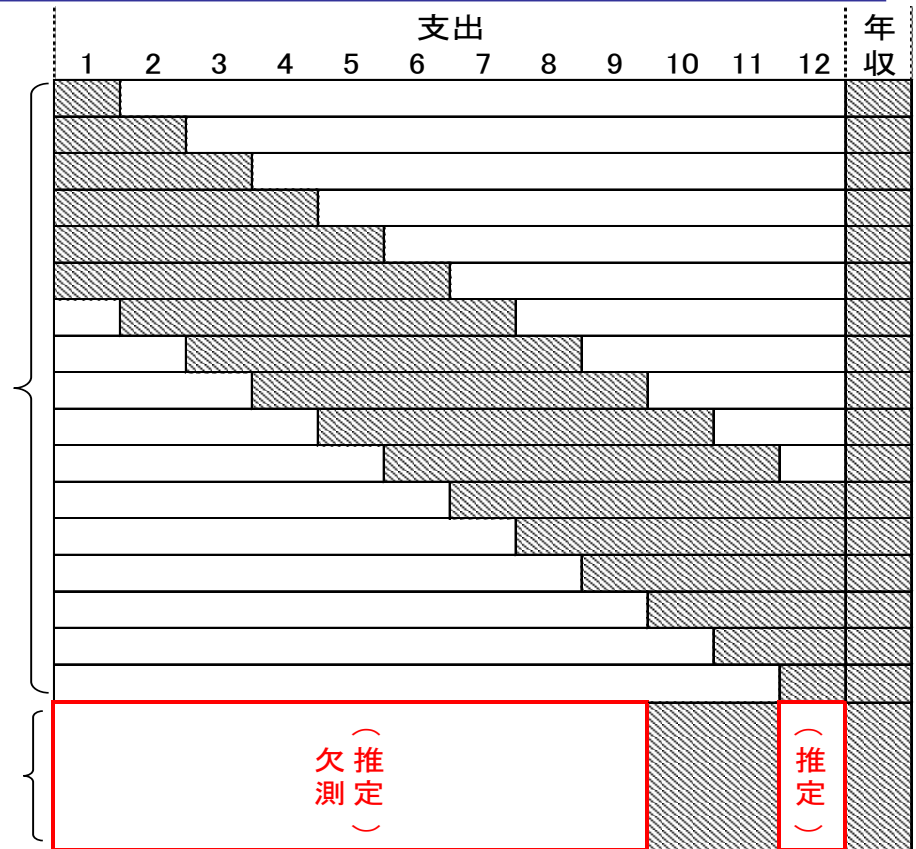
但し重回帰は「0でないグループに適用」(混合分布)

	0でない数	0でない割合(%)	分析手法
食料	210623	100	重回帰分析
住居	79039	37.52629105	Tobit-Type2
光熱. 水道	205795	97.70775271	重回帰分析
家具. 家事用品	205247	97.4475722	重回帰分析
被服及び履物	172415	81.85953101	Tobit-Type2
保健医療	190743	90.56133471	Tobit-Type2
交通. 通信	206743	98.15784601	重回帰分析
教育	48264	22.91487634	Tobit-Type2
教養娯楽	207357	98.44936213	重回帰分析
その他の消費支出	206414	98.00164275	重回帰分析

提案手法の仮定

ローテーションパネルである
家計調査と全国消費実態調査の
関係を右図のように整理し、
欠測データの構造であること
を考慮し両者を併合させた分析
仮定A) 欠測が「ランダムな欠測」
仮定B) 両者の回答集団の差異は
利用した共変量にのみ依存

家計調査

全国消費
実態調査

上図) 2人以上世帯の場合

【利点】

- ・ 家計調査単体より精度が向上
- ・ 全消の諸変数と相関算出可能

式としては

家計*i*の当年*t*月のデータ（またはそれ+1の対数）を

$$y_{it} = \sum_{m=1}^{12} \alpha_m D_{it} + \sum_{k=2}^6 \beta_k E_{itk} + \varepsilon_{it}$$

$$\text{但し } D_{it} = \begin{cases} 1 & t = m \\ 0 & \text{otherwise} \end{cases}, \quad E_{itk} = \begin{cases} 1 & \text{家計} i \text{ が } t \text{ 月に } k \text{ か月目} \\ 0 & \text{otherwise} \end{cases}$$

と表現する。

（前提条件）

家計調査と全国家計構造調査の違いは「季節性」「調査に加わって何か月目の調査か？」に依存

属性情報による回帰とTobit-type2 model

さらに家計に関する属性情報(例えば子供の数や世帯収入) x を用いて

$$y_{it} = \gamma^T x_i + \sum_{m=1}^{12} \alpha_m D_{it} + \sum_{k=2}^6 \beta_k E_{itk} + \varepsilon_{it}$$

のように表現

⇒回帰による回答者の違いの考慮

加えて

$$y_{it}^* = \lambda^T x_i + \phi + \epsilon_{it}$$

を背後に考え

教育費などゼロが多い項目のことも考慮して

y ではなく v が観測されるとする。

$$v_{it} = \begin{cases} y_{it} & \text{if } y_{it}^* \geq 0 \\ 0 & \text{if } y_{it}^* < 0 \end{cases}$$

多変量分析の困難さ

12か月分の結果を統合するには
12変数のTobit-type2モデルを
構築する(正確には24変数)必要
があり数値計算的に困難

方法1) マルコフ連鎖モンテカルロ
法によるベイズ推定
方法2) モンテカルロEM
アルゴリズム

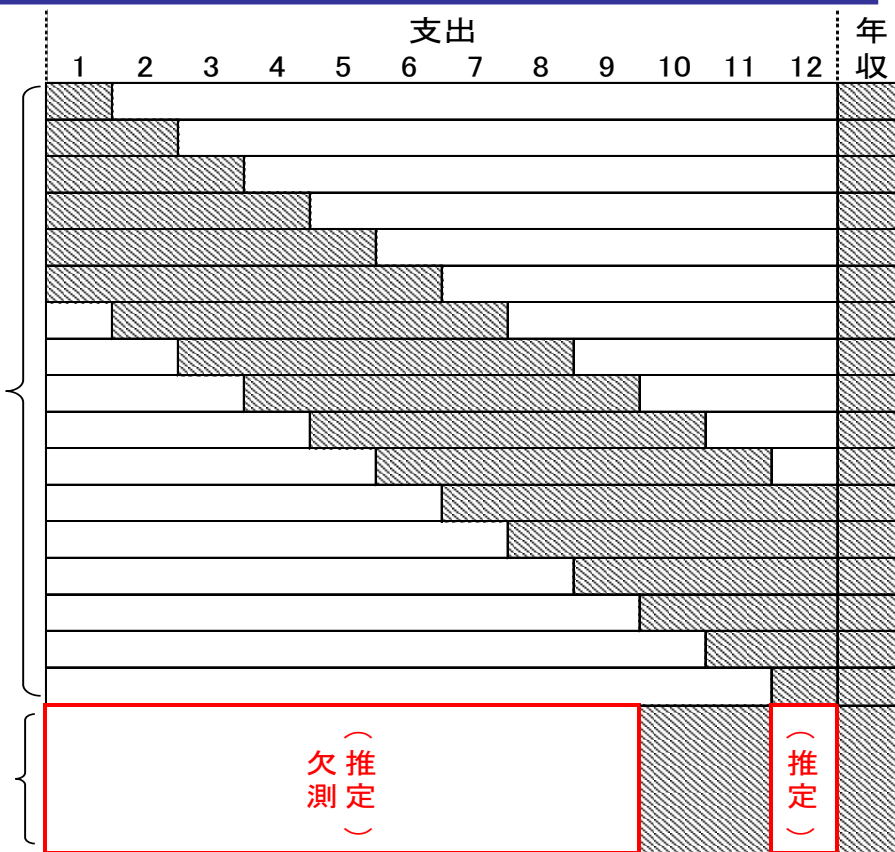
どれも乱数発生の問題、今回は

方法3) 誤差相関を無視した方法 を利用する

⇒平均値構造の母数推定については一致性を持たせる

家計調査

全国消費
実態調査



上図) 2人以上世帯の場合

誤差相関を無視した分析

本来はTobit type 2 モデルを考えているので毎月 2 変数

12か月で24変数の多変量Tobit type2モデルの推定

⇒膨大な数値積分のため困難

そこで本研究では「すべての月の値」をスタックして利用

具体的には上のデータを下に変換

問題点

$$y_{it}^* = \lambda_t^T x_i + \phi + \epsilon_{it}$$

$$y_{it} = \gamma^T x_i + \sum_{m=1}^{12} \alpha_m D_{it} + \sum_{k=2}^6 \beta_k E_{itk} + \epsilon_{it}$$

ID	開始	5月	6月	7月
1	5月	A	B	C
2	1月	D	E	欠測
3	6月	欠測	F	G



の誤差相関($\epsilon_{it}, \epsilon_{it}$)は同じiでは相関

⇒一般に線形のモデルでは系列相関

があっても期待値構造の推定には

一貫性(不偏性も)

ID'	ID	月	何か月目	値
1	1	5	1	A
2	1	6	2	B
3	1	7	3	C
4	2	5	5	D
5	2	6	6	E
6	3	6	1	F
7	3	7	2	G

誤差相関を無視した分析

今回は多変量Tobit type 2 モデルにおいて系列相関を無視してよいかどうか？

これについては今回は詳細は省くが平均構造の母数の一致性を示すことが可能(漸近論の概略は資料参照)

【今回の簡便法の利点】

- ・ 計算が簡単（積分計算のための乱数発生が不要）
- ・ 一致性がある

【今回の簡便性の欠点】

- ・ 効率性は多変量の同時推定に劣る

分析で出した指標

「全国消費実態調査（添字 zi ）を1年やった場合」の結果

① 1年間「家計調査での平均的な“調査継続による効果”」の元での年平均

$$\mu_{avg} = E \left[\gamma^T x_{zi} + \sum_{m=1}^{12} \alpha_m D_{zit} + \frac{1}{5} \sum_{k=2}^6 \beta_k + \sigma \lambda_{mills, zi} \mid y_{zi} > 0 \right] P(y_{zi} > 0)$$

の推定、具体的には

$$\widehat{\mu_{avg}} = \frac{1}{\#(y_{zi} > 0)} \sum_{i: y_{zi} > 0} \left[\hat{\gamma}^T x_{zi} + \hat{\alpha}_m D_{zit} + \frac{1}{5} \sum_{k=2}^6 \hat{\beta}_k + \hat{\sigma} \hat{\lambda}_{mills, zi} \right] \frac{\#(y_{zi} > 0)}{N_Z}$$

②なるべく調査継続による効果が一番少ない(最も記入が多い)場合の下での年平均

$$\mu_{min} = E \left[\gamma^T x_{zi} + \sum_{m=1}^{12} \alpha_m D_{zit} + \max(\beta_k) + \sigma \lambda_{mills} \mid y_{zi} > 0 \right] P(y_{zi} > 0)$$

の推定、具体的には

$$\widehat{\mu_{min}} = \frac{1}{\#(y_{zi} > 0)} \sum_{i: y_{zi} > 0} \left[\hat{\gamma}^T x_{zi} + \hat{\alpha}_m D_{zit} + \max(\hat{\beta}_k) + \hat{\sigma} \hat{\lambda}_{mills, zi} \right] \frac{\#(y_{zi} > 0)}{N_Z}$$

を計算する。但し $\hat{\gamma}, \hat{\alpha}_m, \hat{\beta}_k, \hat{\sigma}$ と逆ミルズ比 λ_{mills} のパラメータは家計調査から推定する。

分析で出した指標

「全国消費実態調査を1年やった場合」の結果

③ 1, 2 か月目の調査継続による効果を1年間継続var.1(9,10月は全消利用)

$$\frac{N_{Z9}\bar{y}_{Z9}}{N} + \frac{N_{Z10}\bar{y}_{Z10}}{N} + \sum_{m \neq 9,10}^{12} \frac{N_{Am}}{N} \frac{1}{\#(y_{Zi} > 0)} \sum_{i:y_{Zi} > 0} \left[\hat{\alpha}_m + \frac{\hat{\beta}_1 + \hat{\beta}_2}{2} + \hat{\gamma}^T x_{Zi} + \hat{\sigma} \hat{\lambda}_{mills,Zi} \right] \frac{\#(y_{Zi} > 0)}{N_Z}$$

* N_{Zm} , N_{Am} は全国消費実態調査と家計調査のm月の人数,

* N_Z は全国消費実態調査の人数,

* N_A は家計調査の10,11月以外の月の人数,

* $N = N_Z + N_A$,

* $\bar{y}_{Zm}$ は全国消費実態調査のm月平均. ここで,

$$\frac{1}{\#(y_{Zi} > 0)} \sum_{i:y_{Zi} > 0} \left[\hat{\alpha}_m + \frac{\hat{\beta}_1 + \hat{\beta}_2}{2} + \hat{\gamma}^T x_{Zi} + \hat{\sigma} \hat{\lambda}_{mills,Zi} \right] \frac{\#(y_{Zi} > 0)}{N_Z}$$

は $E \left[\alpha_m + \frac{\beta_1 + \beta_2}{2} + \gamma^T x_{Zi} + \sigma \lambda_{mills} | y_{Zi} > 0 \right] P(y_{Zi} > 0) + 0 * P(y_{Zi} = 0)$ の推定値

分析で出す指標

「全国消費実態調査を1年やった場合」の結果、

④ 1, 2 か月目の調査継続による効果を1年間継続var. 2 (9, 10月は全消利用)

* **共変量調整を利用せず**に平均値をもとに推定する場合

= 1 ~ 8、11・12月は「1, 2か月目の調査継続による効果にした家計調査」

$$\frac{N_{Z9}\bar{y}_{Z9}}{N} + \frac{N_{Z10}\bar{y}_{Z10}}{N} + \sum_{m \neq 9,10}^{12} \frac{N_{Am}}{N} \left[\bar{y}_{Am} + \left(\frac{\hat{\beta}_1 + \hat{\beta}_2}{2} - \frac{1}{5} \sum_{k=2}^6 \hat{\beta}_k \right) \frac{\#(y_{Aim} > 0)}{N_{Am}} \right]$$

* N_{Zm}, N_{Am} は全国消費実態調査と家計調査のm月の人数,

* N_Z は全国消費実態調査の人数

* N_A は家計調査の10,11月以外の月の人数,

* $N = N_Z + N_A$,

* $\bar{y}_{Zm}$ は全国消費実態調査のm月平均

再解析の結果

共変量の選択と標準誤差の算出

利用する共変量(tobit)

- 1段階目(probit)に用いた共変量

* 尚, 共変量選択に当たり, 可能な組み合わせを試行した上で, BIC最小の共変量による結果を選択している.

	住居	被服及び履物	保健医療	教育
年間収入	○	○	○	○
世帯人員	○	○	○	○
世帯主の性別	○	○	○	
世帯主の年齢	○	○	○	
住居の延面積	○			
就学者人数		○		○
産業(併合: パターンA)	○			
職業(併合: パターンA)	○			
産業(併合: パターンB)				
職業(併合: パターンB)				
住居の所有関係(併合)	○			
住居の構造(併合)	○			

※2段階目(OLS)の共変量、重回帰の共変量は共に「年間収入」「世帯人員」であるため表は割愛

利用する共変量(重回帰分析)

- 用いた共変量

* 尚，共変量選択に当たり，可能な組み合わせを試行した上で，BIC最小の共変量による結果を選択している。

	食料	光熱・水道	家具・家事用品	交通・通信	教養娯楽	その他
年間収入	○	○	○	○	○	○
世帯人員	○	○	○	○	○	○
世帯主の性別	○	○	○	○	○	○
世帯主の年齢	○	○	○	○	○	○
住居の延面積						
就学者人数						
産業(併合: パターンA)				○	○	○
職業(併合: パターンA)				○	○	○
産業(併合: パターンB)	○	○				
職業(併合: パターンB)	○	○				
住居の所有関係(併合)						
住居の構造(併合)						

年平均の推定・標準誤差(Tobit-type2 model)

【家計調査+全消9月10月】

(A)年推定 (調査継続効果平均)① (B)年推定 (調査継続効果最小)②

(C)調査継続効果1,2か月目での年推定(回帰利用)③

(D)調査継続効果1,2か月目での年推定(平均利用)④

費目	統計量	(A)	(B)	(C)	(D)
住居	平均	17458	18954	16934	17080
	標準誤差	802.9023	1173.205	338.9962	321.2874
被服及び履物	平均	11176	11679	11719	11562
	標準誤差	173.9253	209.659	104.9876	77.4758
保健医療	平均	12339	12751	12359	12566
	標準誤差	216.2998	268.1611	95.18456	94.23587
教育	平均	12833	13484	12928	11751
	標準誤差	493.6315	611.2683	212.1888	187.2068

* 標準誤差は、(家計IDの) ブートストラップによる推定

年平均の推定・標準誤差(Tobit-type2 model)

【家計調査+全消9月10月】

(E)年平均 (家計)

(F) 9,10 月平均 (家計)

(G) 9,10 月平均 (全消)

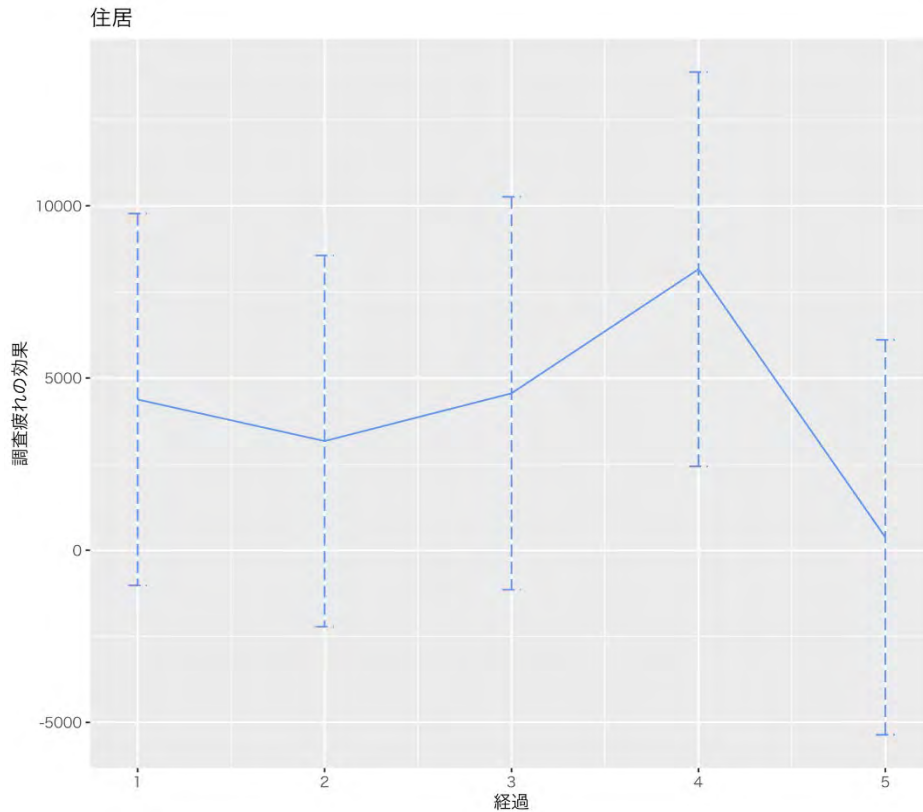
(H)年平均 (合計 = 家計と全消混合)

(I)年平均 (家計 + 全消の効果 = $E + (G - F)$)

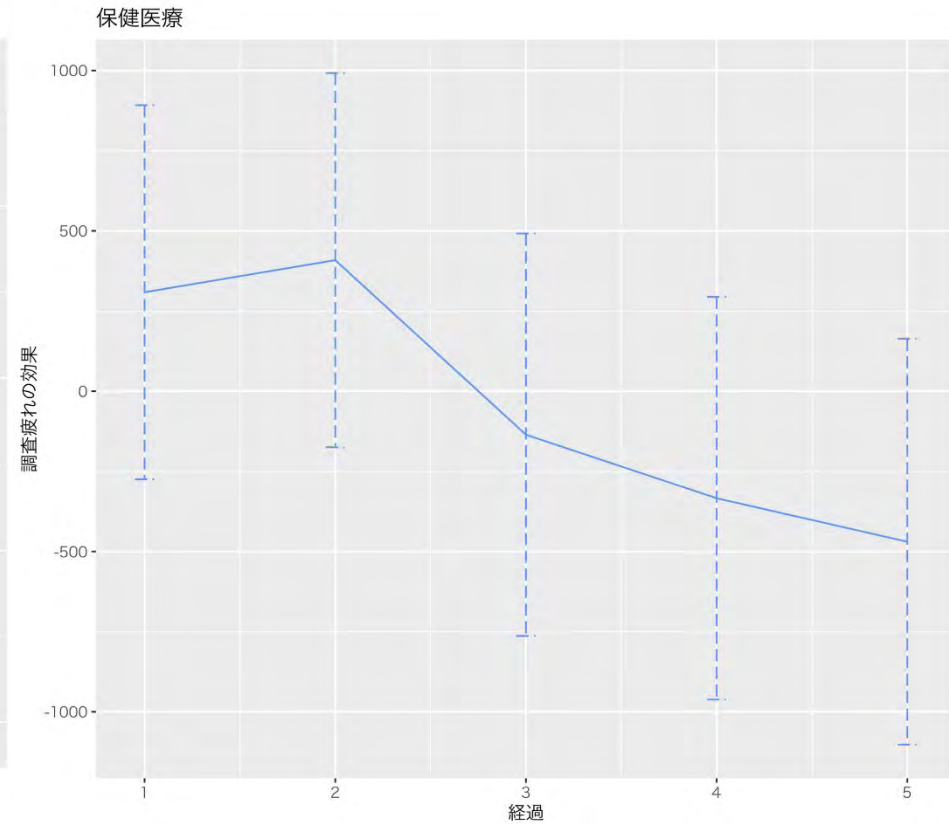
費目	統計量	(E)	(F)	(G)	(H)	(I)
住居	平均	17947	17393	16391	16910	16945
	標準誤差	325.02	803.77	282.35	188.55	911.81
被服及び履物	平均	11874	10412	10780	11688	12243
	標準誤差	77.303	162.88	63.69	44.58	191.21
保健医療	平均	12118	12093	12631	12309	12656
	標準誤差	83.87	207.08	81.06	51.11	237.67
教育	平均	9793	11376	13670	10905	12088
	標準誤差	174.84	449.59	208.54	113.42	525.54

調査継続による効果(Tobit-type2 model)

【住居】



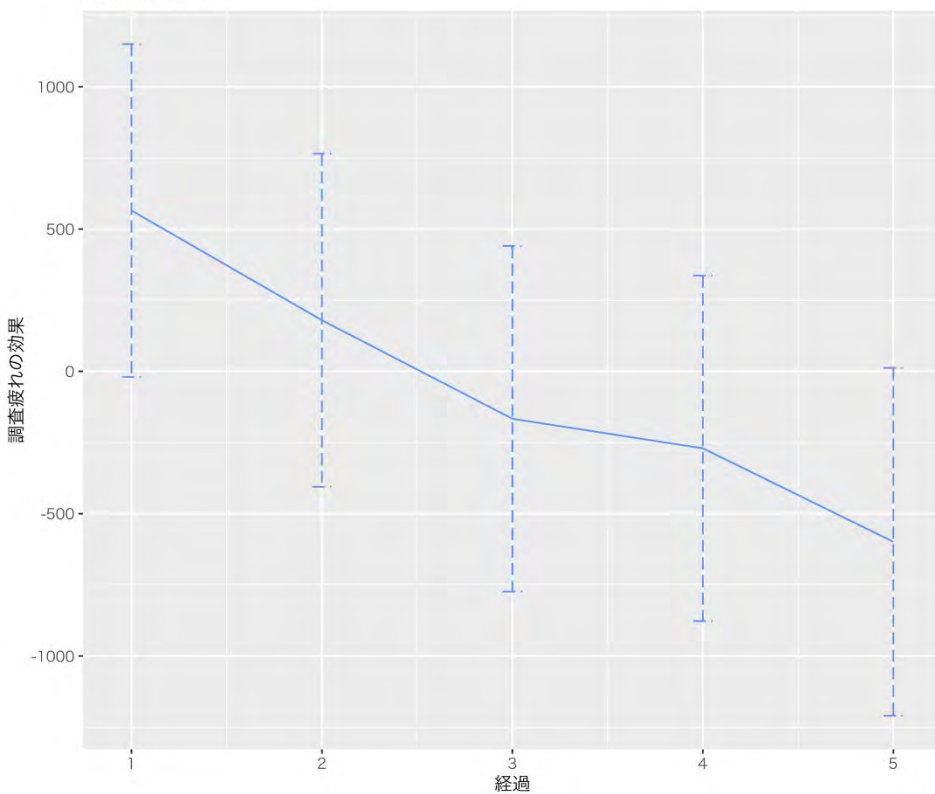
【保健医療】



調査継続による効果(Tobit-type2 model)

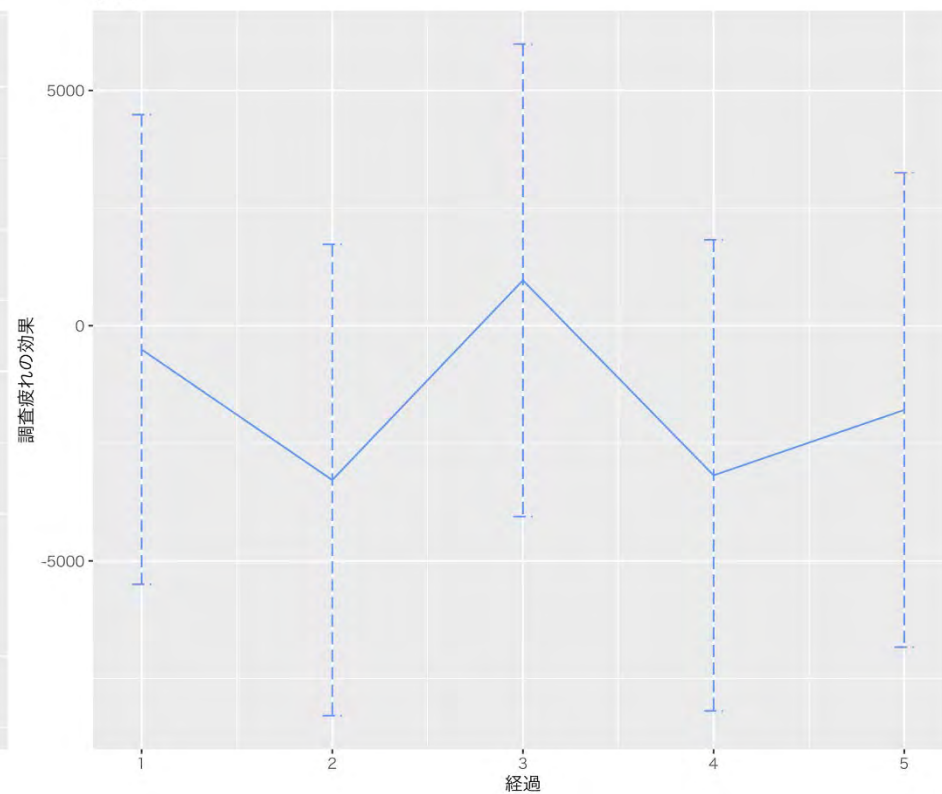
【被服及び履物】

被服及び履物



【教育】

教育



年平均の推定・標準誤差(重回帰分析)

【家計調査+全消9月10月】

(A)年推定 (調査継続効果平均)① (B)年推定 (調査継続効果最小)②

(C)調査継続効果1,2か月目での年推定(回帰利用)③

(D)調査継続効果1,2か月目での年推定(平均利用)④

費目	統計量	(A)	(B)	(C)	(D)
食料	平均	72220	72654	70538	73195
	標準誤差	381.3707	454.7355	183.7283	157.7977
光熱.水道	平均	20124	20306	24112	21841
	標準誤差	92.65474	113.2095	75.26221	46.48548
家具.家事用品	平均	9170	9378	10240	10353
	標準誤差	175.8226	215.3511	92.97096	81.58676
交通.通信	平均	43559	45245	45780	44060
	標準誤差	1203.247	1422.709	753.2545	432.9158
教養娯楽	平均	28265	29267	29484	29546
	標準誤差	381.0545	455.0518	230.8463	157.7977
その他の消費支出	平均	51465	51952	57746	53790
	標準誤差	973.9815	1087.824	499.6399	354.1751

* 標準誤差は, (家計IDの) ブートストラップによる推定

年平均の推定・標準誤差(重回帰分析)

【家計調査+全消9月10月】

(E)年平均 (家計)

(F) 9,10 月平均 (家計)

(G) 9,10 月平均 (全消)

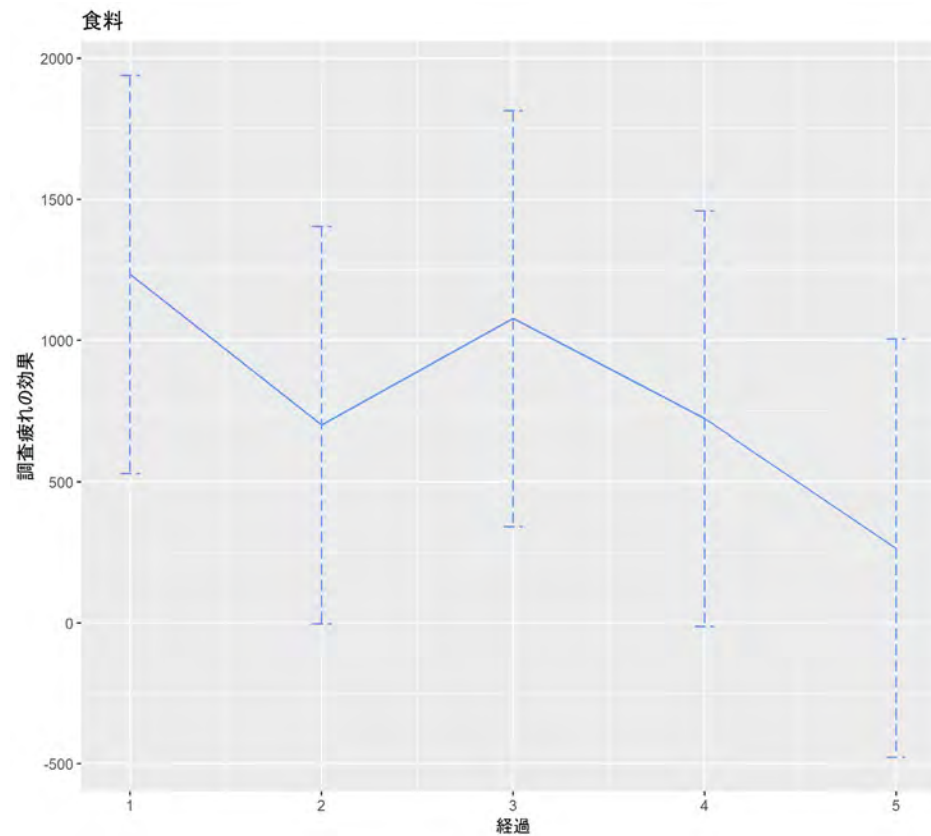
(H)年平均 (合計 = 家計と全消混合)

(I)年平均 (家計 + 全消の効果 = $E + (G - F)$)

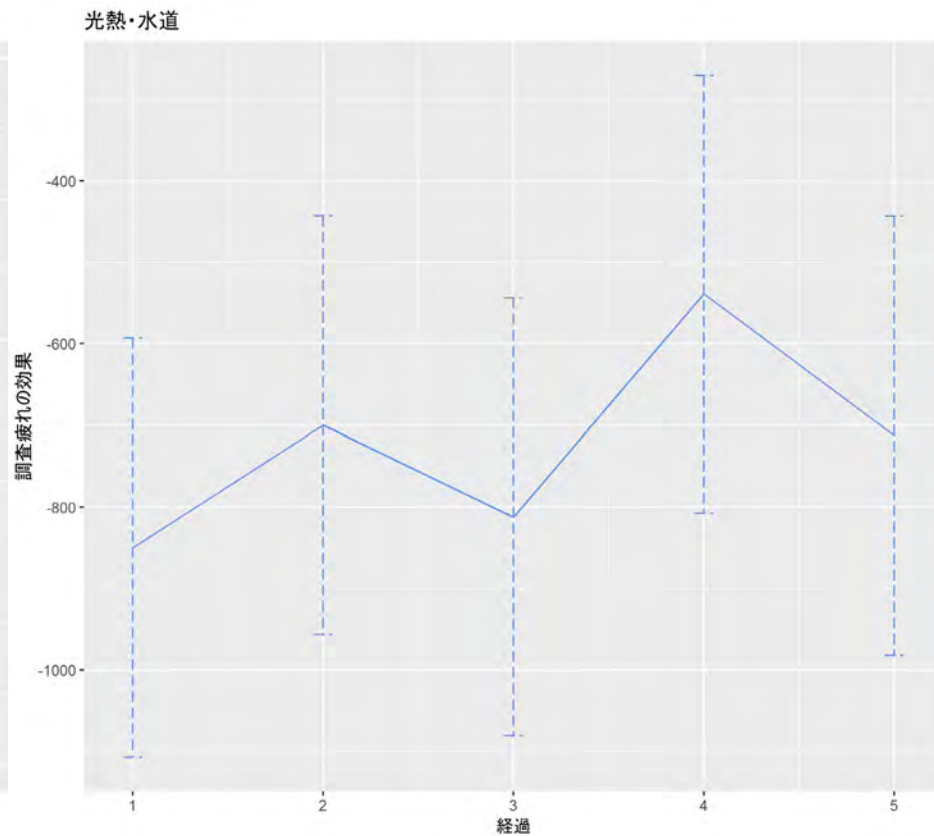
費目	統計量	(E)	(F)	(G)	(H)	(I)
食料	平均	72471	70245	73284	72410	75510
	標準誤差	124.32	266.12	112.99	73.87	314.71
光熱.水道	平均	22808	19352	20525	21489	23981
	標準誤差	45.13	89.63	34.33	24.75	106.06
家具.家事用品	平均	10307	8823	10019	10198	11503
	標準誤差	84.66	159.59	76.41	50.30	196.15
交通.通信	平均	40013	40932	46763	42674	45845
	標準誤差	446.32	1128.89	415.45	255.80	1283.04
教養娯楽	平均	28791	27478	29281	28474	30594
	標準誤差	157.75	346.03	138.80	89.74	404.83
その他の消費支出	平均	51618	49542	55025	52307	57101
	標準誤差	333.39	918.00	339.47	203.55	1033.98

調査継続による効果(重回帰分析)

【食料】



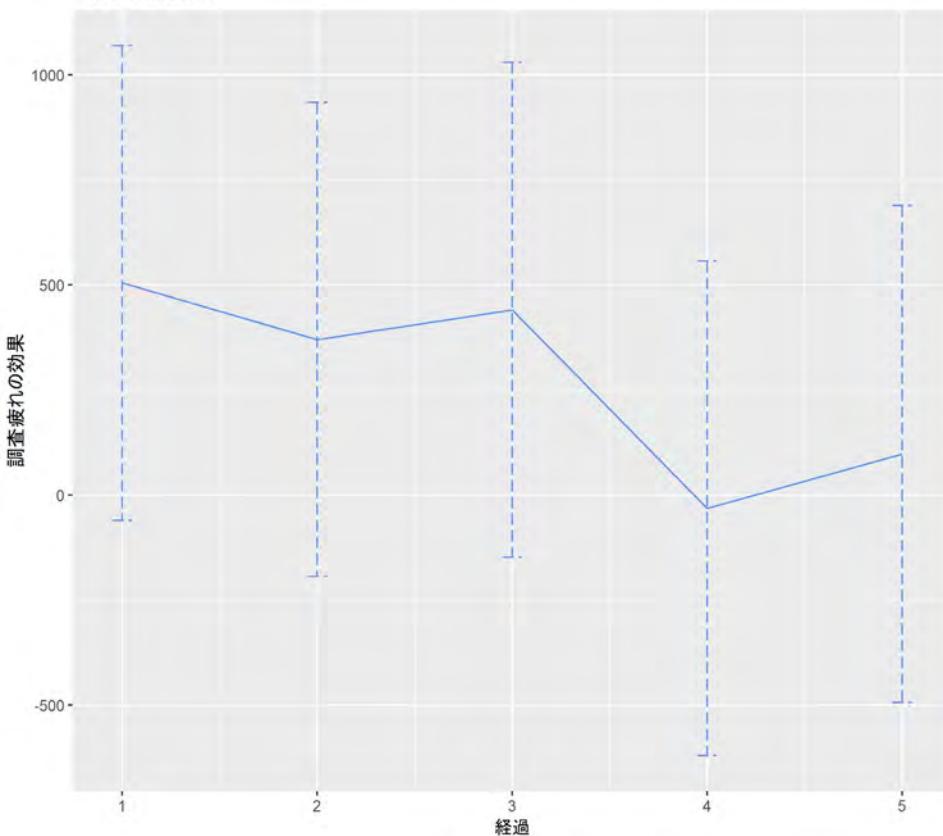
【光熱.水道】



調査継続による効果(重回帰分析)

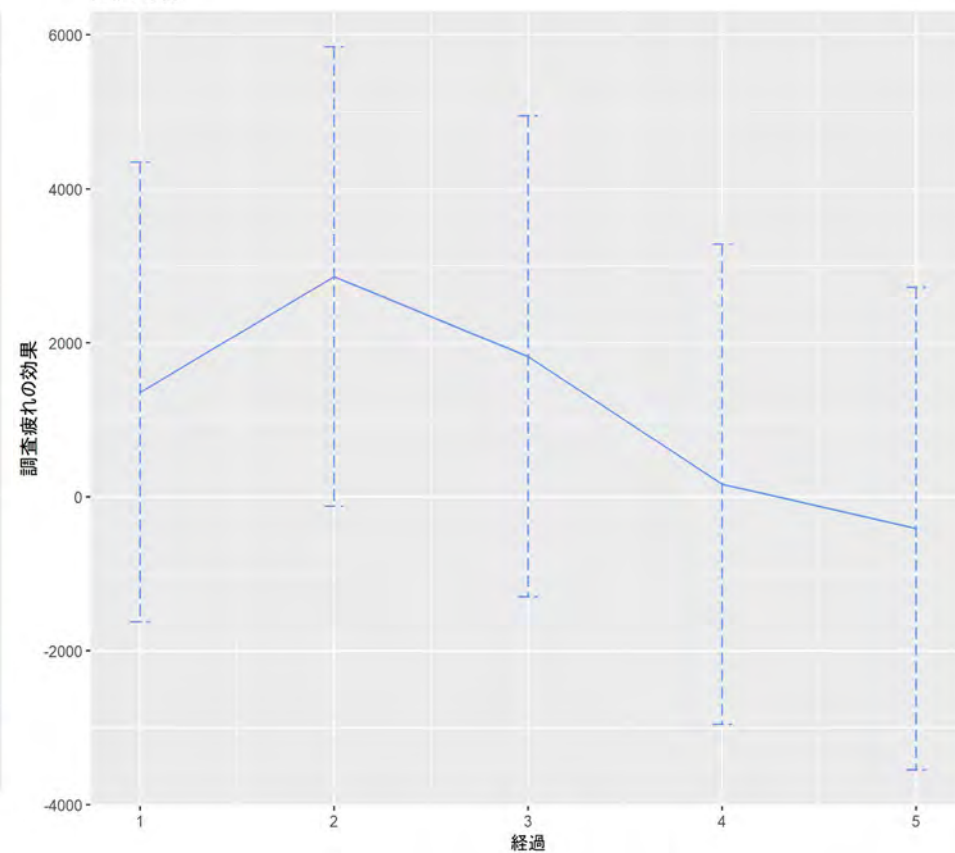
【家具・家事用品】

家具・家事用品



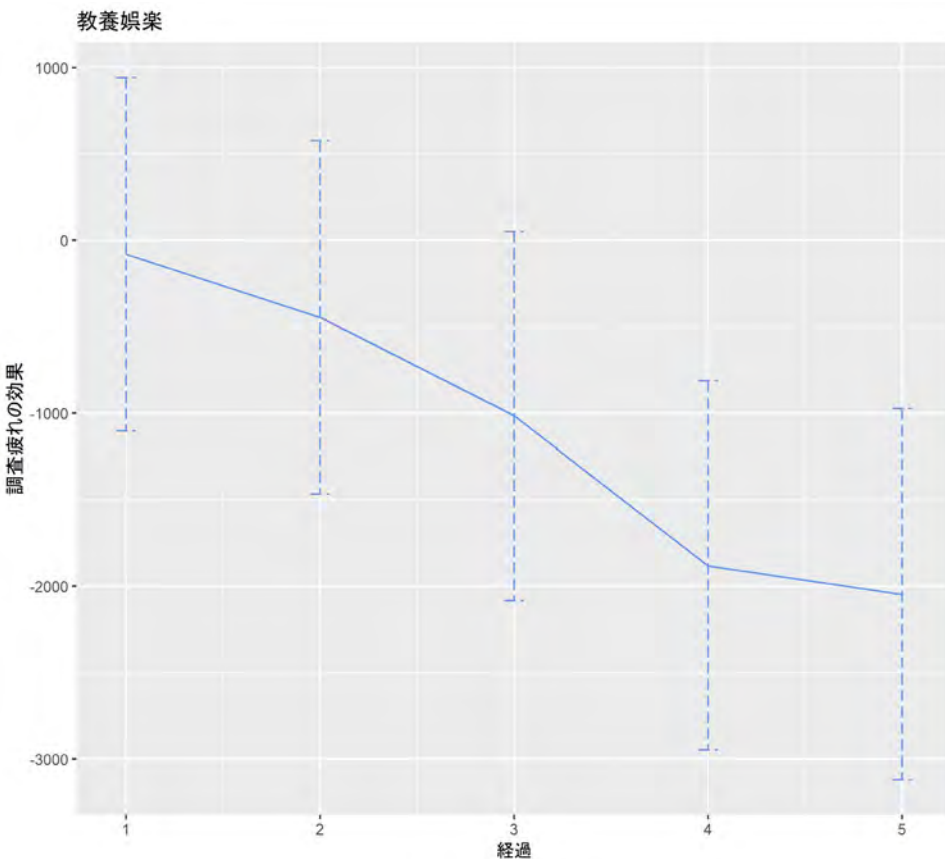
【交通・通信】

交通・通信

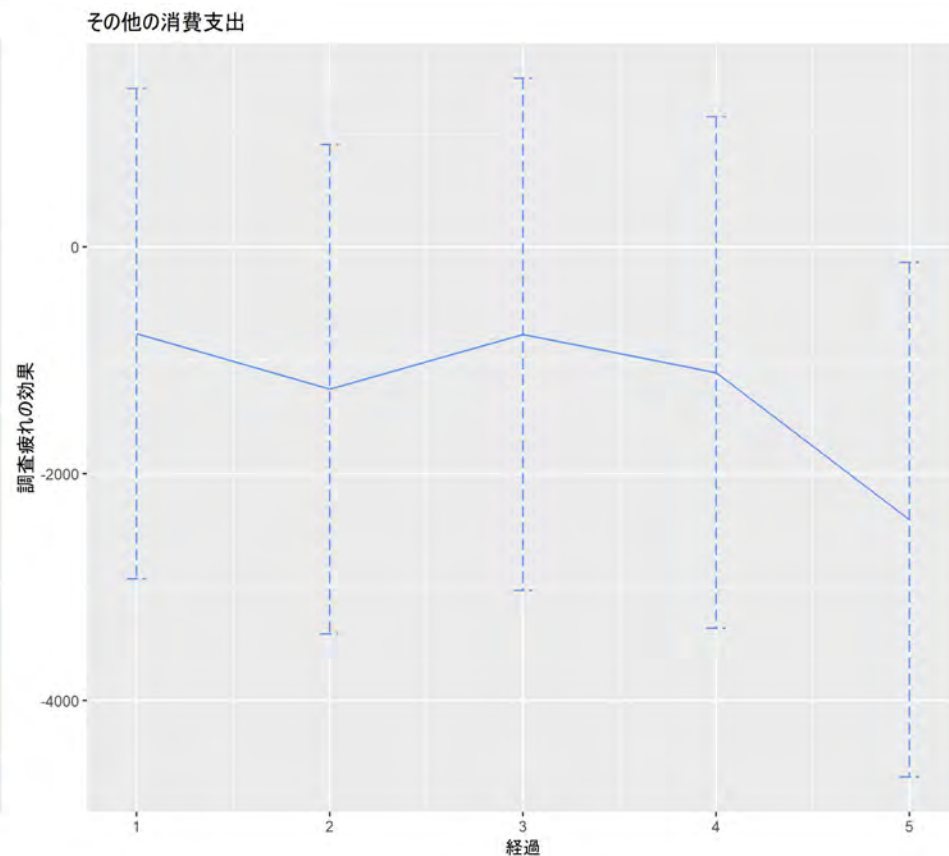


調査継続による効果(重回帰分析)

【教養娯楽】



【その他の消費支出】

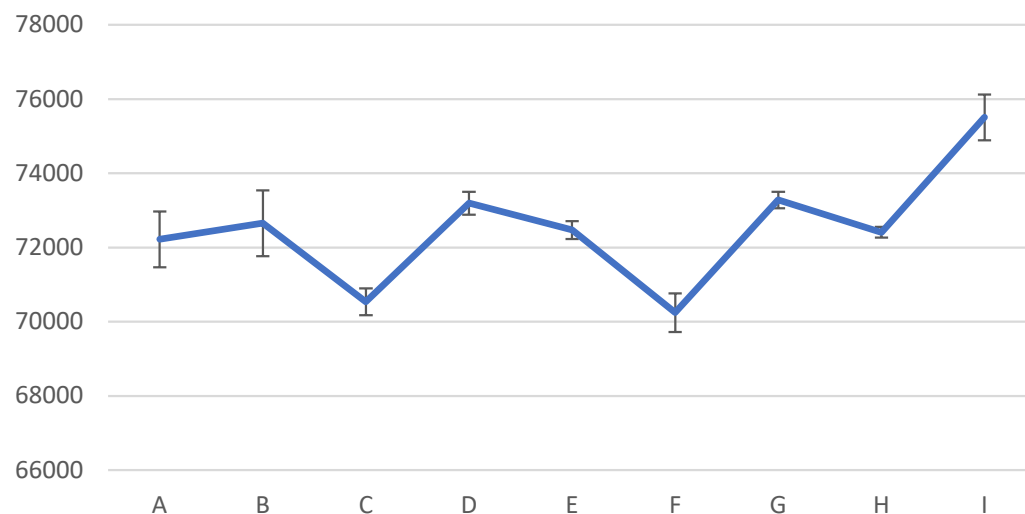


図による比較

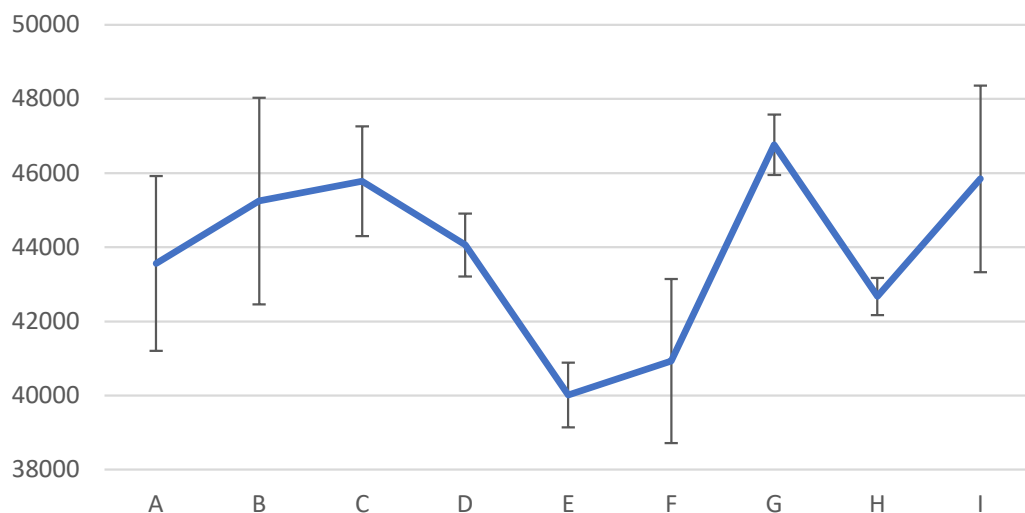
バーは95%信頼区間

- (A)年推定 (調査継続効果平均)①
- (B)年推定 (調査継続効果最小)②
- (C)調査継続効果1,2か月目での年推定
(回帰利用)③
- (D)調査継続効果1,2か月目での年推定
(平均利用)④
- (E)年平均 (家計)
- (F) 9,10 月平均 (家計)
- (G) 9,10 月平均 (全消)
- (H)年平均 (合計 = 家計と全消混合)
- (I)年平均 (家計 + 全消の効果 = $E + (G - F)$)

食料 平均



交通.通信 平均



図による比較

バーは95%信頼区間

(A)年推定 (調査継続効果平均)①

(B)年推定 (調査継続効果最小)②

(C)調査継続効果1,2か月目での年推定

(回帰利用)③

(D)調査継続効果1,2か月目での年推定

(平均利用)④

(E)年平均 (家計)

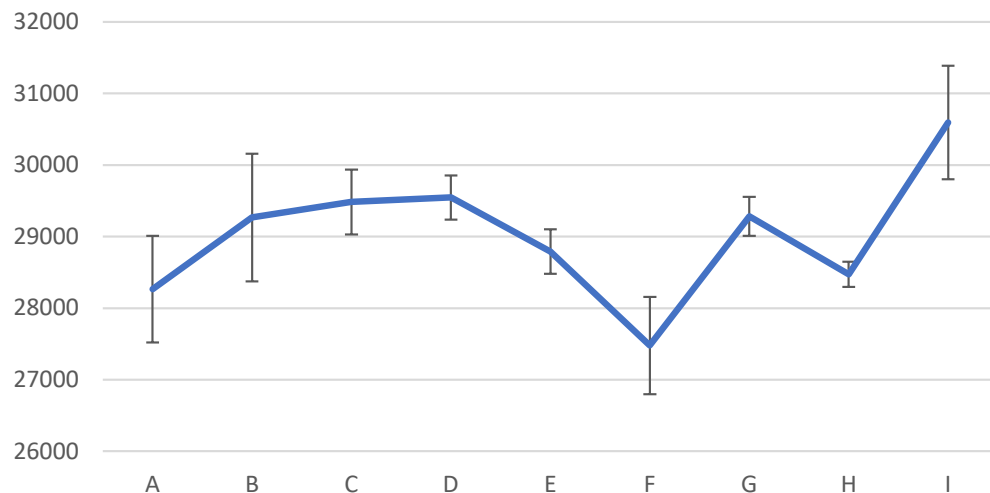
(F) 9,10 月平均 (家計)

(G) 9,10 月平均 (全消)

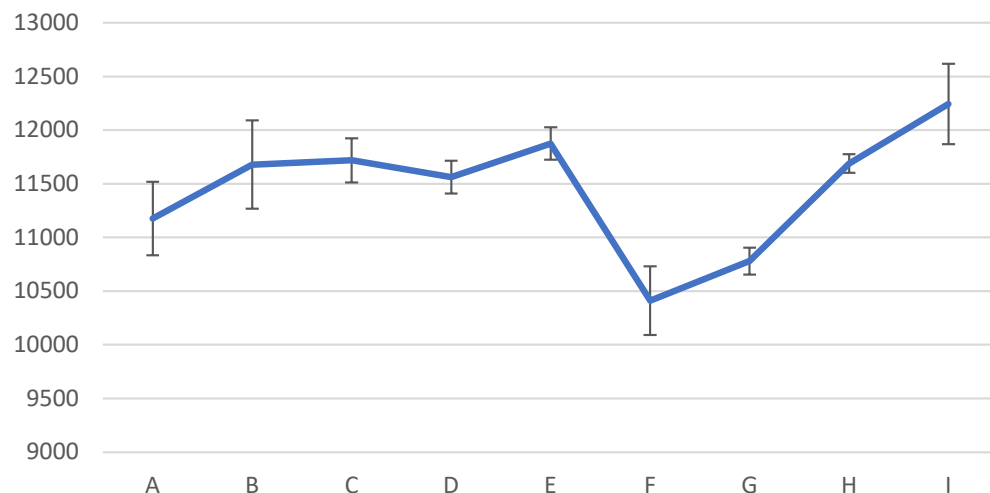
(H)年平均 (合計 = 家計と全消混合)

(I)年平均 (家計 + 全消の効果 = $E + (G - F)$)

教養娯楽 平均



被服及び履物 平均



調査継続による効果に関する検定

10大費目	推定方法	検定方法	F値或いはWald統計量	P値
食料	重回帰分析	F検定	1295.00	0
住居	Tobit-Type2	Wald検定	42.04	5.79E-08
光熱、水道	重回帰分析	F検定	958.30	0
家具、家事用品	重回帰分析	F検定	105.22	0
被服及び履物	Tobit-Type2	Wald検定	23.84	0.00023
保険医療	Tobit-Type2	Wald検定	17.63	0.00345
交通、通信	重回帰分析	F検定	68.11	0
教育	Tobit-Type2	Wald検定	10.45	0.06339
教養娯楽	重回帰分析	F検定	270.70	0
その他の消費支出	重回帰分析	F検定	292.60	0

推定された調査継続効果パラメーターと調査継続効果パラメーター（5つのダミー変数）を「0」とした場合が有意に異なるかどうかの検定：

– $F\text{値} = ((SSE(R) - SSE(F)) / K) / (SSE(F) / (n - K))$

* SSE: 残差平方和, K: 説明変数の数, n: サンプルサイズ

R: 「調査継続効果パラメーター=0」のモデル

F: 推定された調査継続効果パラメーターによるモデル

– $\text{Wald統計量} = (\theta(F) - \theta(R))' V(\theta(F))^{-1} (\theta(F) - \theta(R))$

* $\theta(R)$, $\theta(F)$: それぞれのモデルにおける推定値

2調査の差と調査継続による効果の大きさ

10大費目	推定方法	9.10月での平均の 2調査の乖離(絶対値)	共変量の群間 差による説明 分	調査継続効果 の最小値最大 値の差	家計調査 の年平均	0でない割 合(%)	継続効果バ イアス(%)
食料	重回帰分析	3039	2009.34	970.32	72471	100.000	1.339
住居	Tobit-Type2	1002	169.46	2889	17947	37.526	6.041
光熱、水道	重回帰分析	1173	568.01	307.53	22808	97.708	1.317
家具、家事用品	重回帰分析	1196	315.09	524.599	10307	97.448	4.960
被服及び履物	Tobit-Type2	368	25.83	938	11874	81.860	6.467
保険医療	Tobit-Type2	538	33.50	798	12118	90.561	5.964
交通、通信	重回帰分析	5831	1204.99	3241.94	40013	98.158	7.953
教育	Tobit-Type2	2294	178.37	1096	9793	22.915	2.565
教養娯楽	重回帰分析	1803	1075.00	1945.45	28791	98.449	6.652
その他の消費支出	重回帰分析	5483	1976.38	1613.14	51618	98.002	3.063

食費などは乖離が大きいがこれは「2調査の収入と世帯人員の差」で説明できる
一方それ以外では調査継続効果による説明の可能性

注1) 共変量の群間差による説明分＝ゼロでない割合/100

$\times \sum (\text{共変量}i \text{の係数} \times (\text{全消の共変量}i \text{の平均} - \text{家計の共変量}i \text{の平均}))$

ただし、重回帰分析の共変量*i*は年間収入,世帯人員,世帯主の性別,世帯主の年齢,
Tobit-Type2の共変量*i*は年間収入,世帯人員.

注2) 調査継続効果によるバイアス＝ゼロでない割合/100

$\times (\text{調査継続効果の最大値と最小値の差} / \text{家計調査の年平均})$

まとめと今後の課題

【まとめ】全国家計構造調査の年次集計に家計調査を利用する場合、過去の家計調査と全消の乖離の原因を理解する必要

“同じ月の比較で全体として全消の方が大”

①標本の違い 収入 家計 < 全消 ⇒ 全消の消費大

②季節性 全消の対象月は↓

③調査継続による効果 家計調査は6か月継続のため↓

本研究では①～③を考慮する方法を提案し合理的な値を推計

前回との違いは「共変量選択」「標準誤差」「検定」の実施

【課題】

(1)まだ一部費目(光熱費)のみ説明不能⇒2019年調査への適用？

(2)各費目の地域ごとの集計値の算出

(3)最終的に何を除去するのか？何を目標にするのか？について

の議論が必要⇒何を考慮した“年間支出”か？ 次ページ問2

議論：課題3について（前回同様）

最終的に何を除去するのか？何を目標にするのか？

家計調査と全消(以前の全国家計構造調査)の乖離の原因である

①標本の違い ②季節性 ③調査継続による効果

④調査項目対象(財・サービスの対象)

のうち②と④が解決されるのは必須として

問1) ①標本の違いをさらにコントロールすべきか？

今回は回帰により一定の考慮 ここさらに共変量調整？

問2) 2調査は期間が異なるので③を考慮する必要がある

但し何を目的とするか？

「指導の結果としての“調査で記入してもらえる最大限”」か？

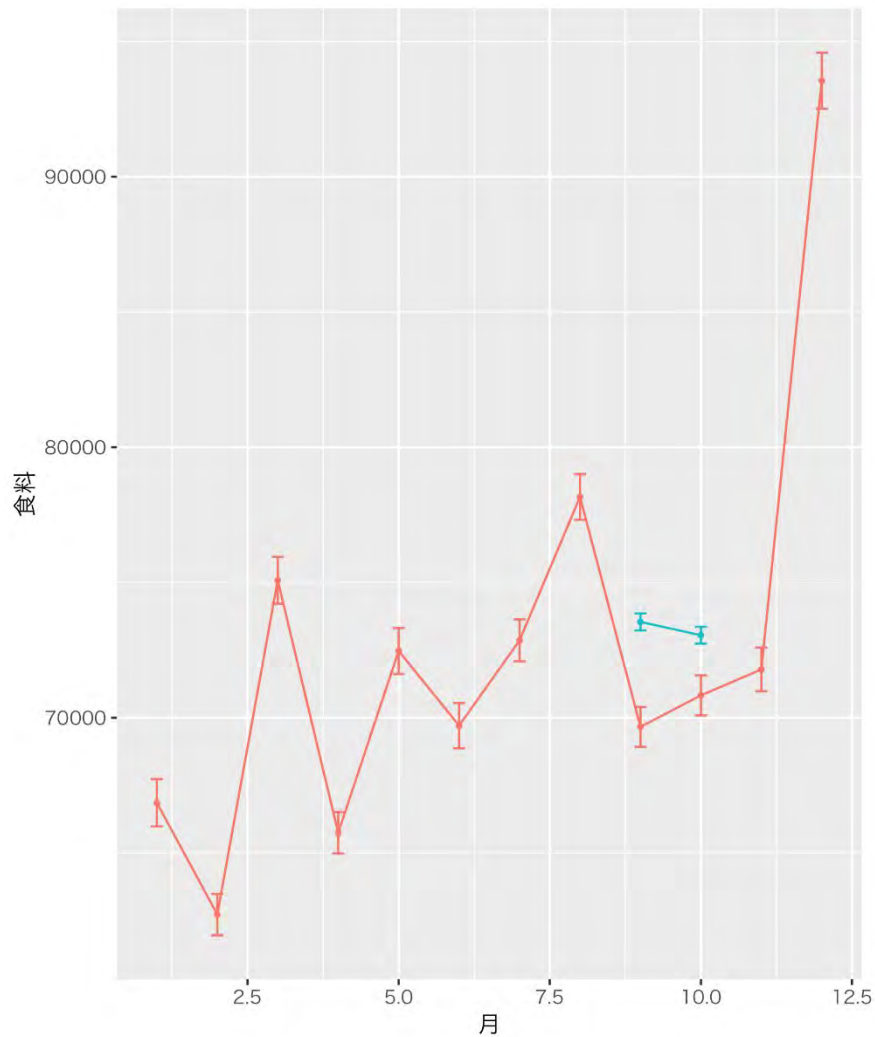
「各月の家計調査が“全消”と同じ1・2か月の“調査継続による効果”の場合を求めるのか？」

* 家計調査では2か月目に「指導」が行われることの効果も存在

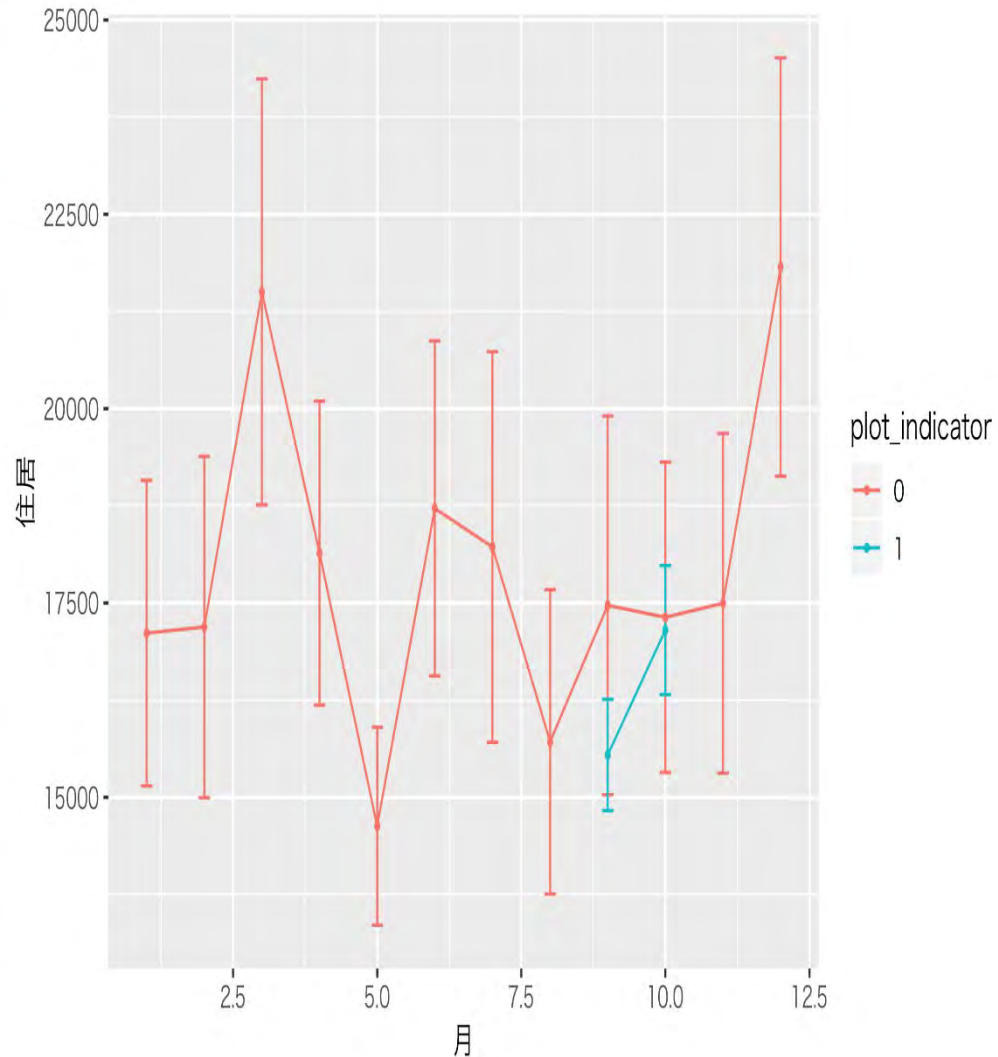
資料編

月平均とその標準誤差(費目別)

【食料】

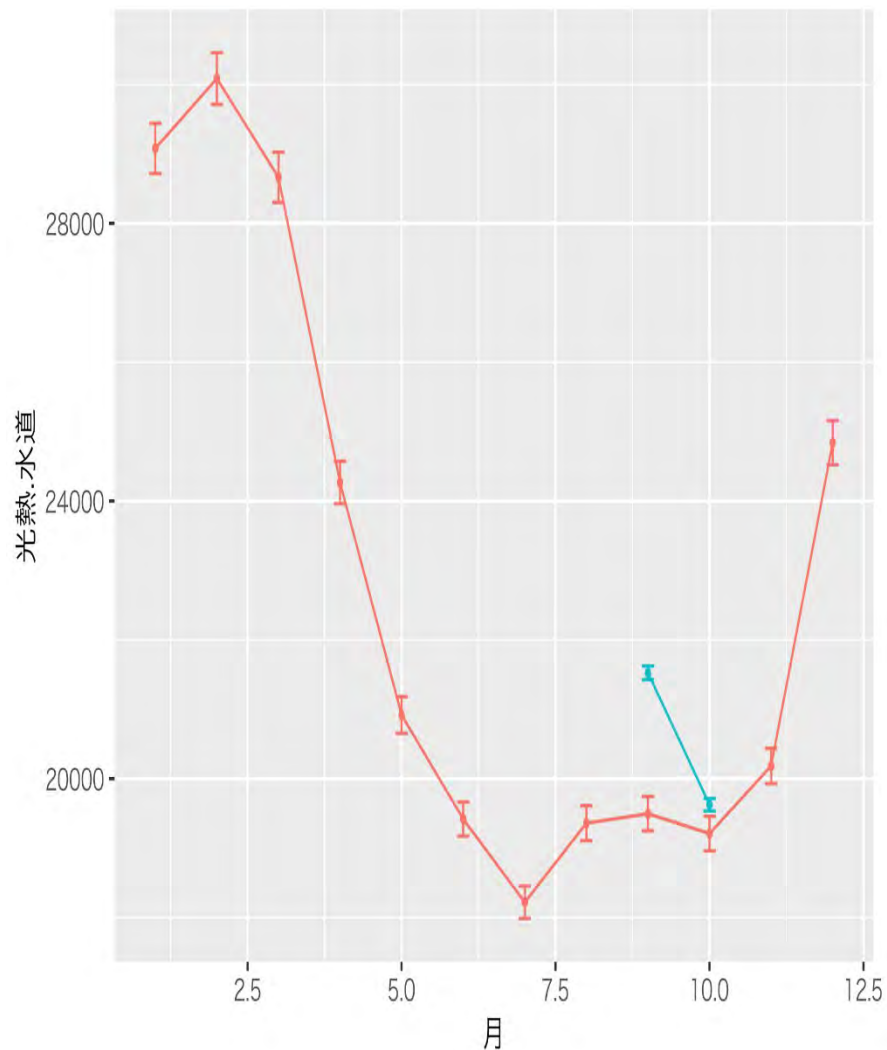


【住居】

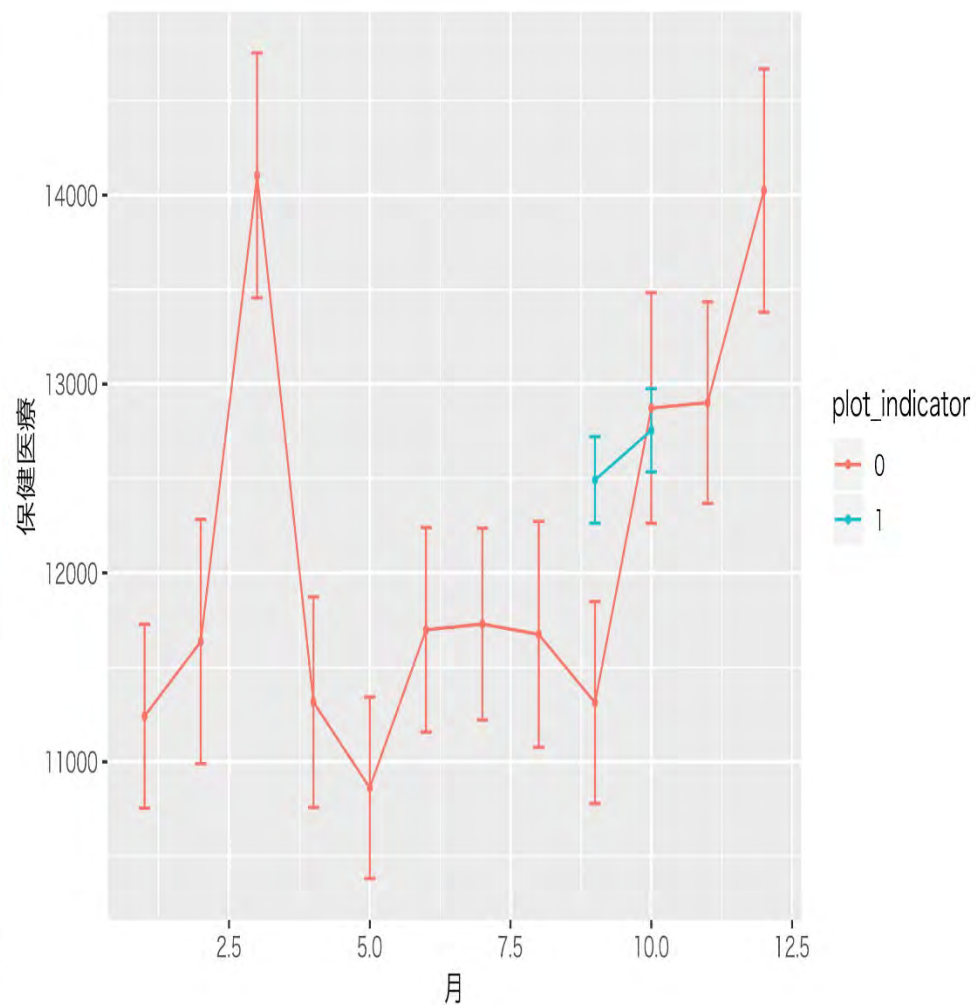


月平均とその標準誤差(費目別)

【光熱.水道】



【保健医療】

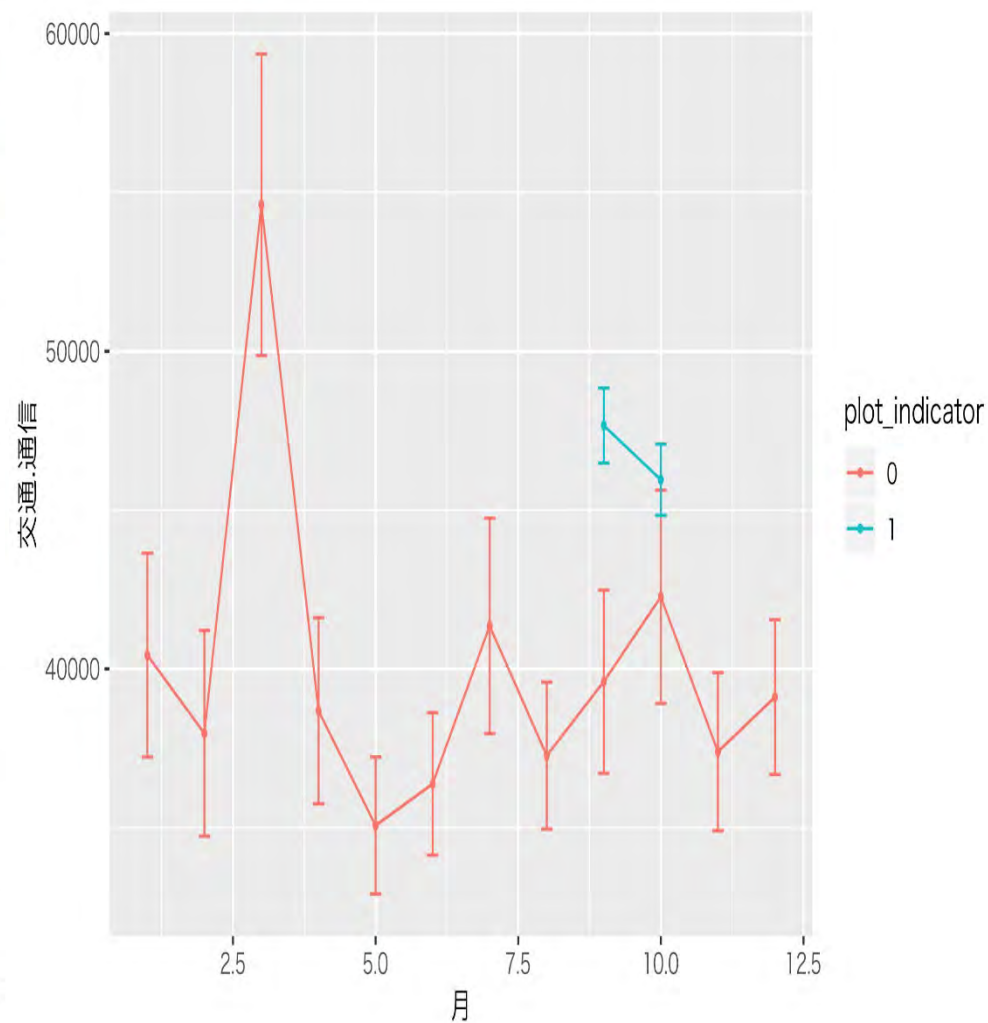


月平均とその標準誤差(費目別)

【被服及び履物】

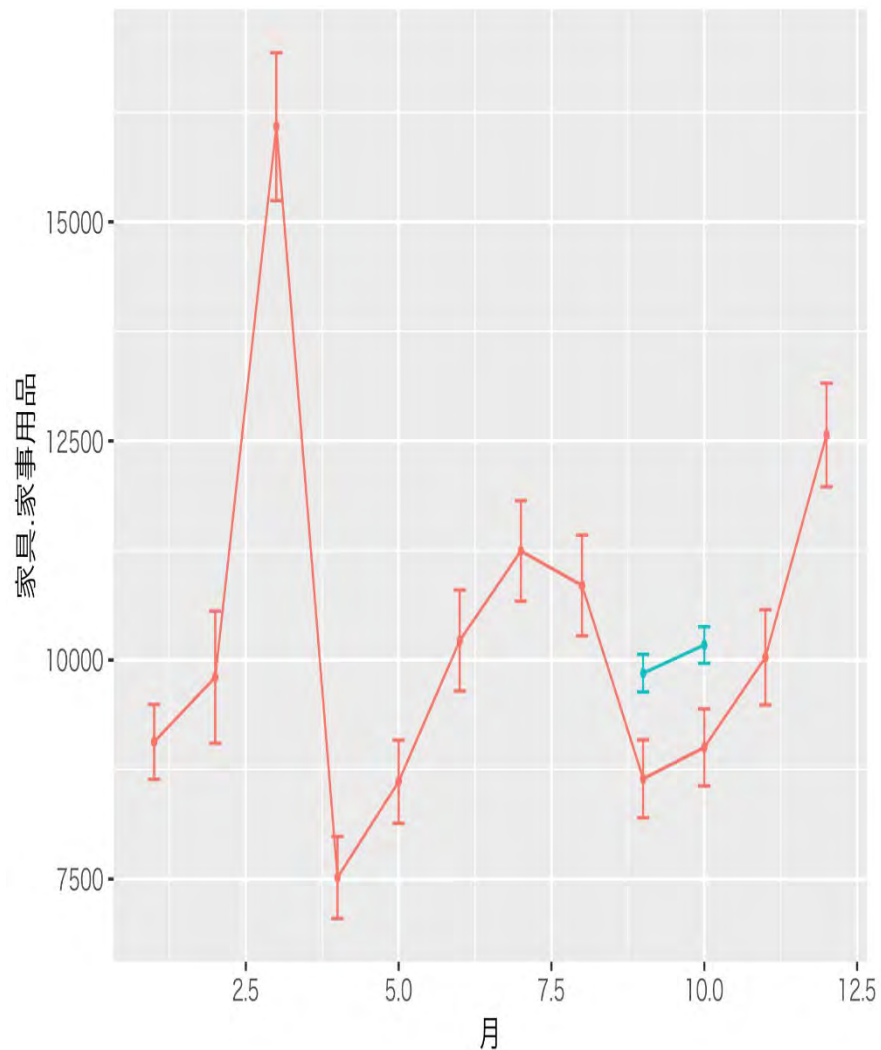


【交通・通信】

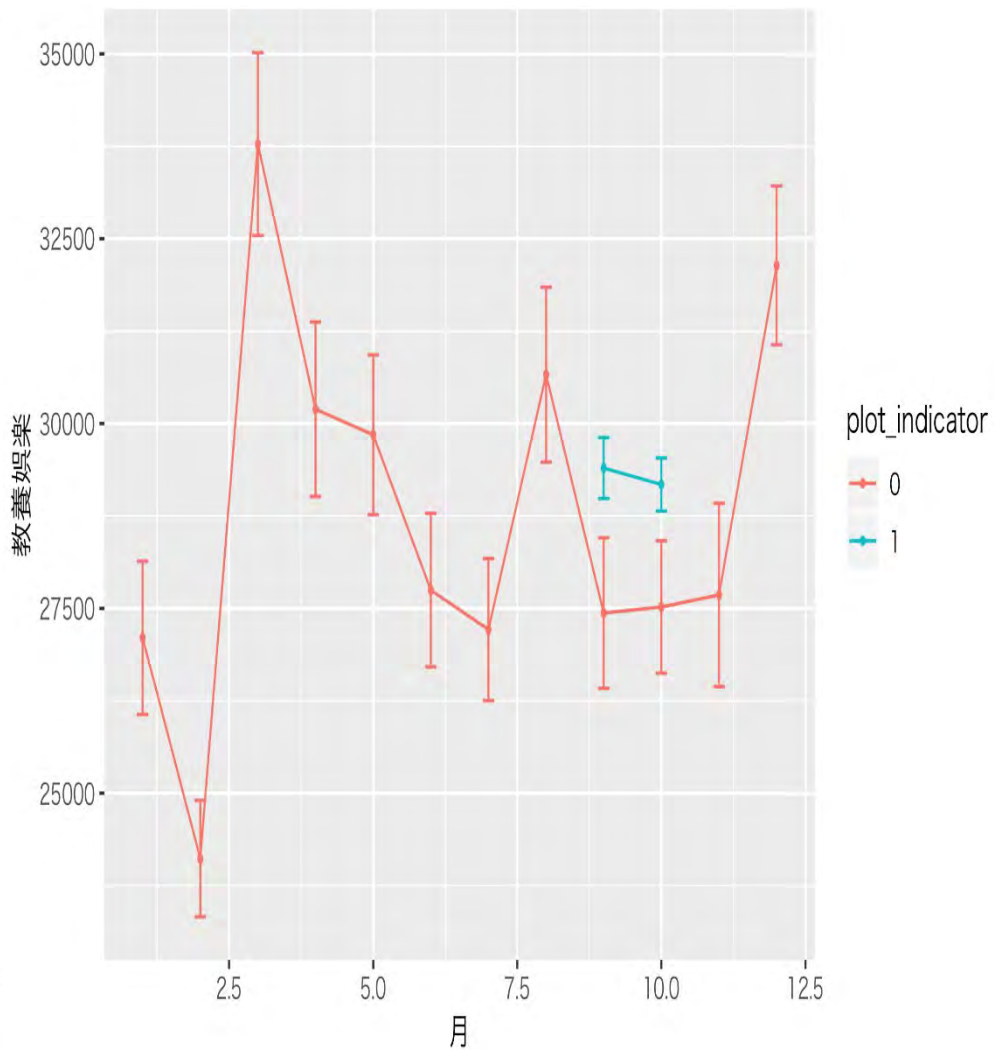


月平均とその標準誤差(費目別)

【家具.家事用品】

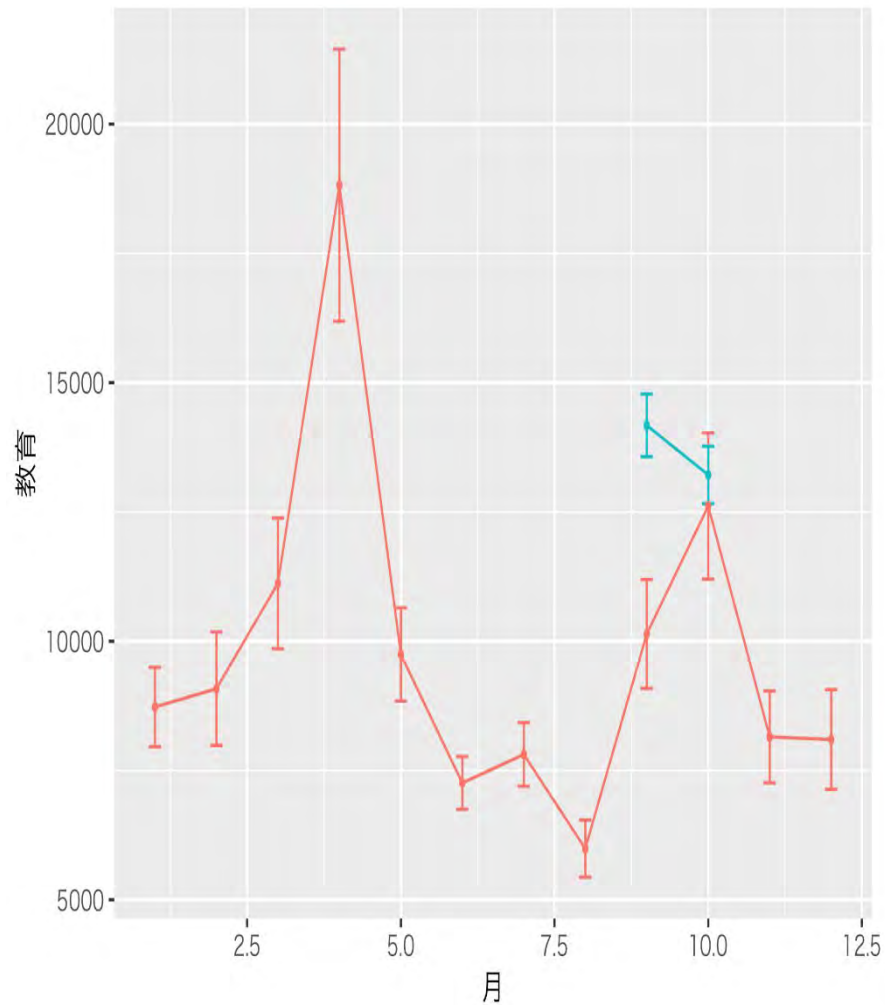


【教養娯楽】

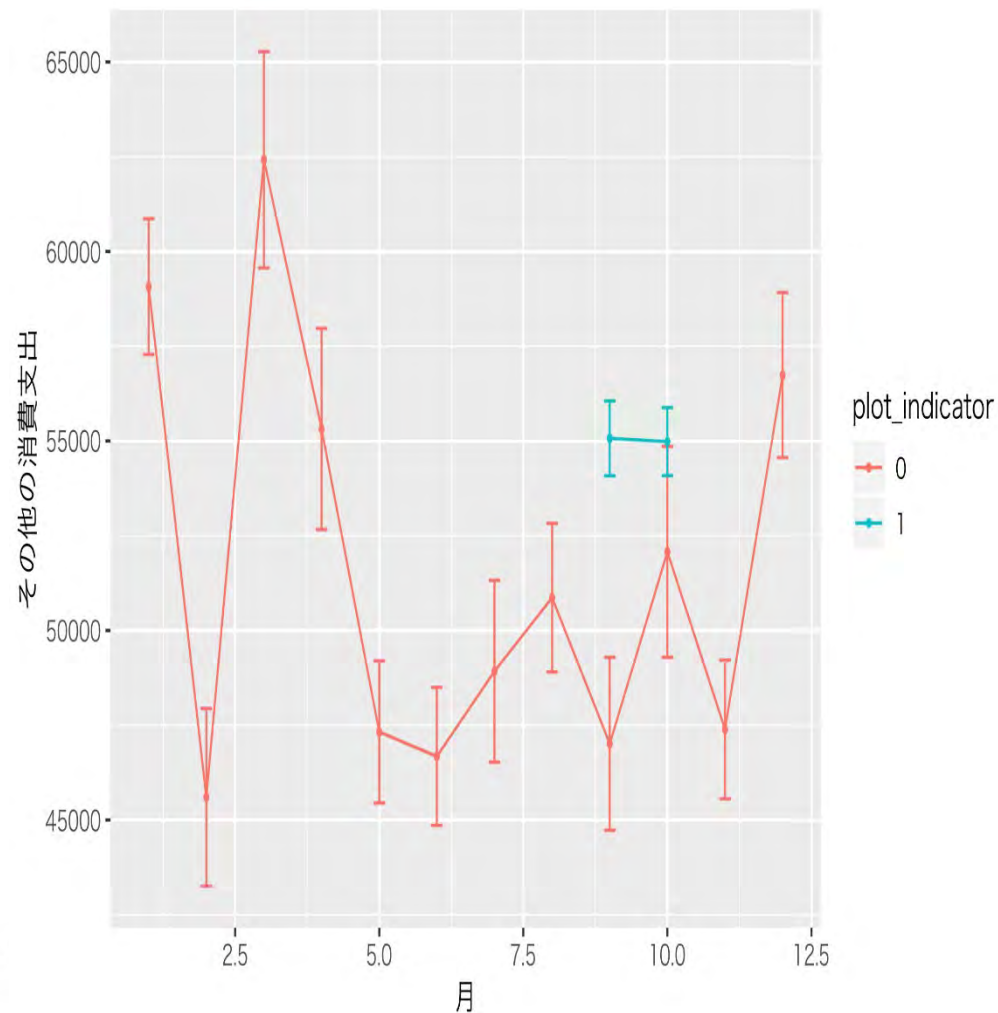


月平均とその標準誤差(費目別)

【教育】

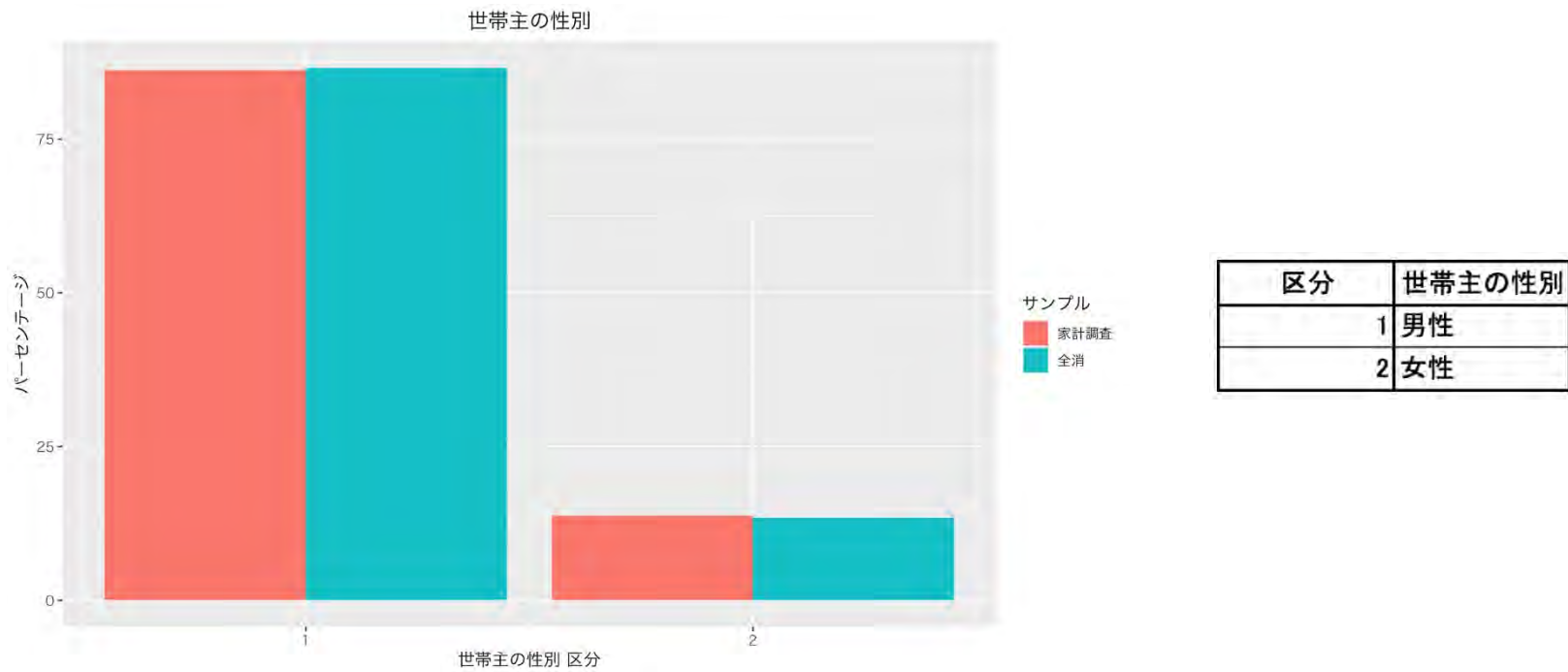


【その他の消費支出】

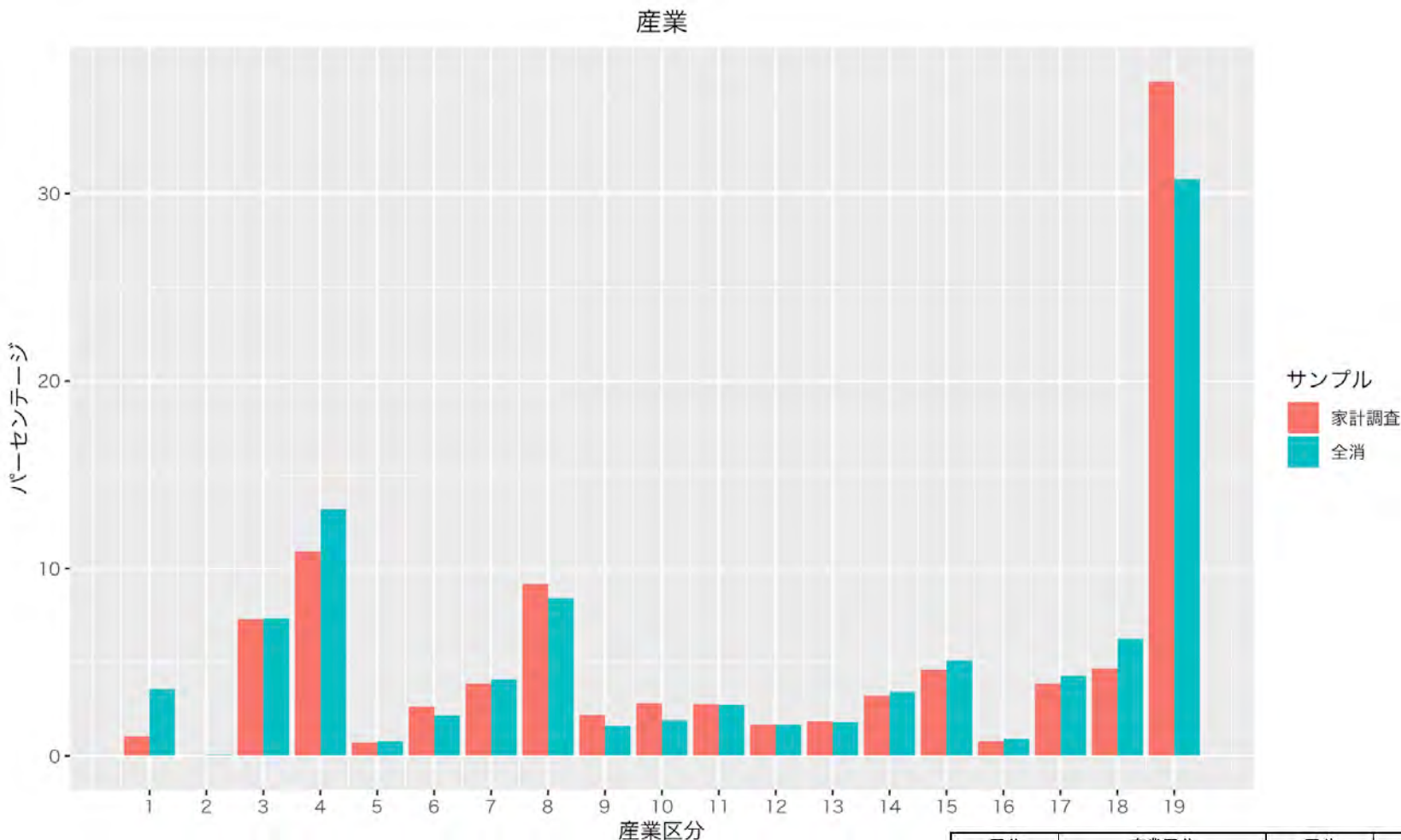


家計調査と全国消費実態調査の違い

	全消		家計調査	
	平均	標準偏差	平均	標準偏差
年間収入	600.7839	415.3084	570.4629	376.4146
世帯人員	2.9222	1.2119	2.8456	1.2003
就業人員	1.366	1.0204	1.2208	0.9681
世帯主の年齢	58.0043	14.6944	58.6914	15.3966
住居の延面積	125.7086	90.1157	115.5035	61.5522
就学者人数	0.5619	0.9159	0.5101	0.8761

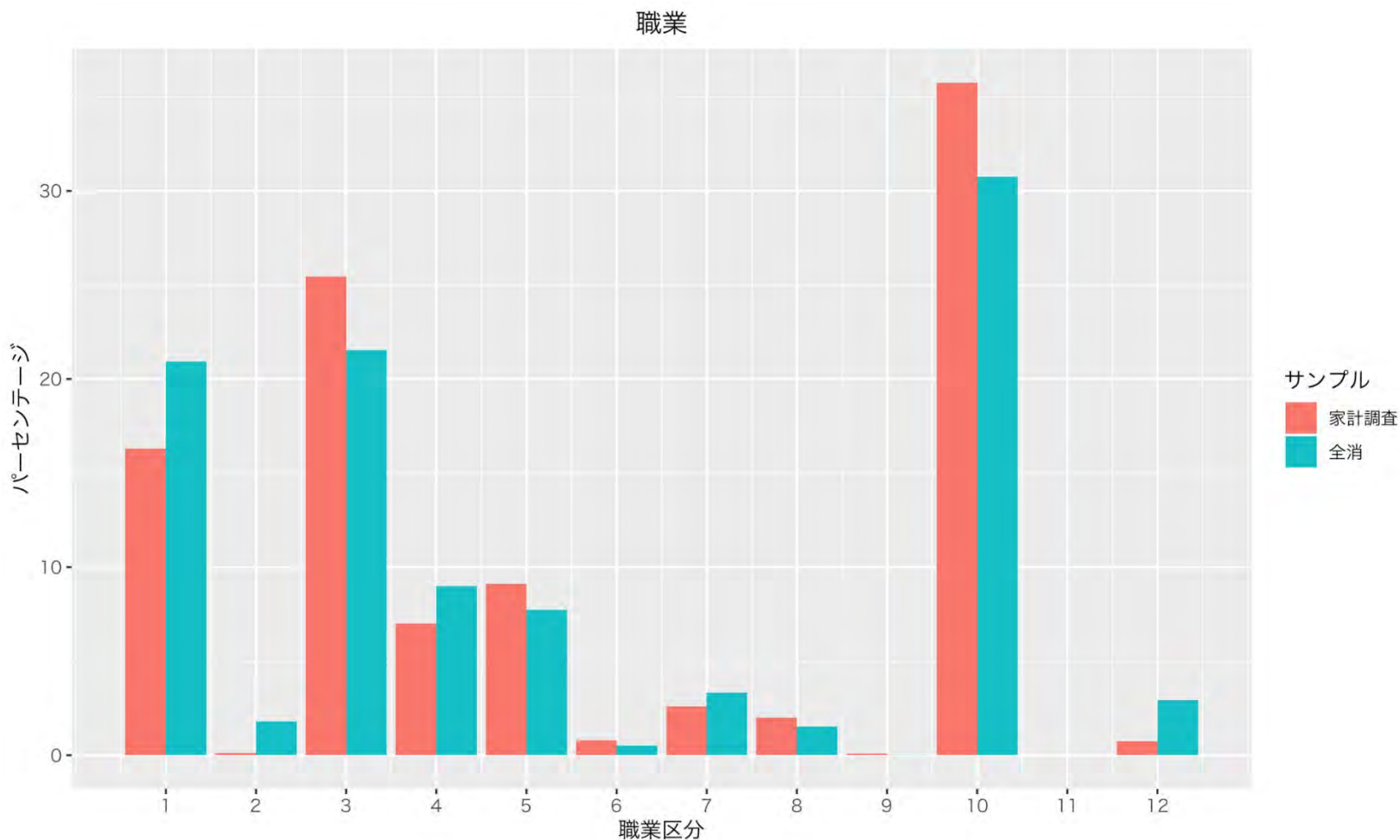


家計調査と全国消費実態調査の違い



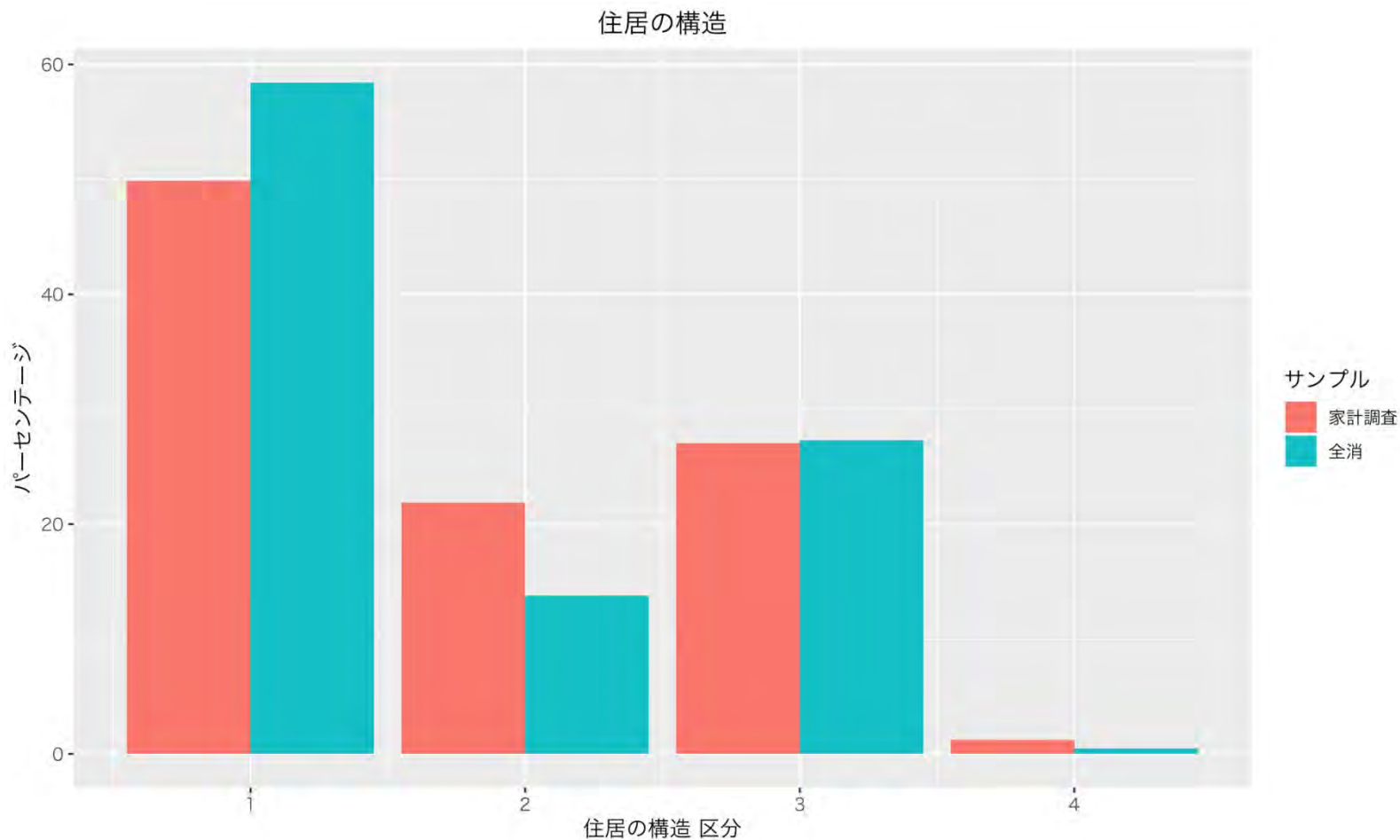
区分	産業区分	区分	産業区分
1	農業・林業・漁業	11	学術研究・専門・技術サービス業
2	鉱業・採石業・砂利採取業	12	宿泊業・飲食サービス業
3	建設業	13	生活関連サービス業・娯楽業
4	製造業	14	教育・学習支援業
5	電気・ガス・熱供給・水道業	15	医療・福祉
6	情報通信業	16	複合サービス事業
7	運輸業・郵便業	17	サービス業(他に分類されないもの)
8	卸売業・小売業	18	公務(他に分類されないもの)
9	金融業・保険業	19	その他・無職
10	不動産業・物品賃貸業		

家計調査と全国消費実態調査の違い



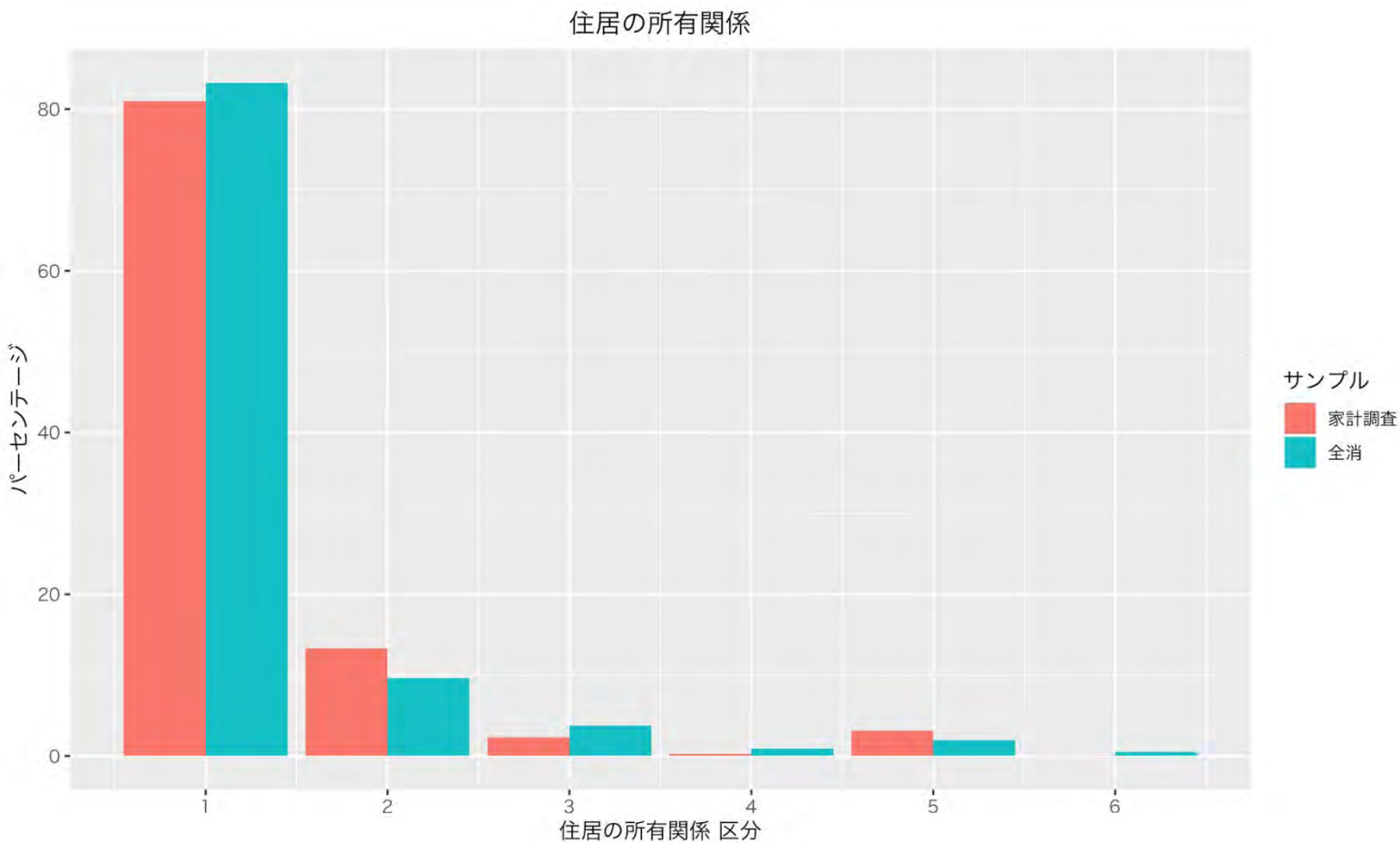
区分	職業区分	区分	職業区分
1	常用労務者	7	法人経営者
2	臨時及び日々雇労務作業	8	自由業者
3	民間職員	9	その他
4	官公職員	10	無職
5	商人及び職人	11	家族従業者
6	個人経営者	12	農林漁業従事者

家計調査と全国消費実態調査の違い



区分	住居の構造
1	木造
2	防災木造
3	鉄骨・鉄筋コンクリート造
4	その他(ブロック造, レンガ造など)

家計調査と全国消費実態調査の違い



区分	住居の所有関係
1	持ち家
2	民営の賃貸住宅
3	都道府県・市区町村営賃貸住宅
4	都市再生機構・公社等の賃貸住宅
5	社宅・公務員宿舎(借上げの社宅及び寮・寄宿舎を含む)
6	借間

背景

○ミクロ統計（家計調査など）とマクロ統計（国民経済計算）の乖離

- 牧（2007）：ミクロ統計/マクロ統計＝約80%（世帯類型、用途分類などの調整後）
 - ミクロ統計/マクロ統計の比率：とりわけ「医療・保険」分野で低い
 - ミクロ統計の問題：誤記入や標本の代表性の問題など
 - マクロ統計の問題：持ち家の帰属家賃の推定方法やコモディティー・フロー法の耐久消費財/非耐久消費財の家計と企業の配分方法の問題など
- 特に、誤記入仮説はミクロ統計とマクロ統計の乖離をサポートする形になっている

⇒しかし、この研究ではその理由やどうやって補正すべきかは明らかではない

背景

○乖離の理由

- 家計調査は、全国消費実態調査や家計消費状況調査と比較しても、特に消費支出額が過少になる（宇南山, 2015）
- 理由の一つは、「調査継続による効果」によるもの
 - 家計調査や全国消費実態調査は、日記型と呼ばれる調査手法（他に、想起型がある）
 - 日記型の問題点：調査負担の重さから、調査慣れが生じる（Browning et al., 2014）
 - 調査慣れ→ 1. 報告行動の変化（調査疲れ;宇南山, 2015）、2. 消費行動の変化（調査参加効果による自己モニタリング; Crossley et al., 2017）
 - 他にも、イレギュラーな消費が抜けやすい問題（Purchase Infrequency）がある（Browning et al., 2014）
- 特に、家計調査は調査期間が長く、調査継続による効果の影響が考えられる

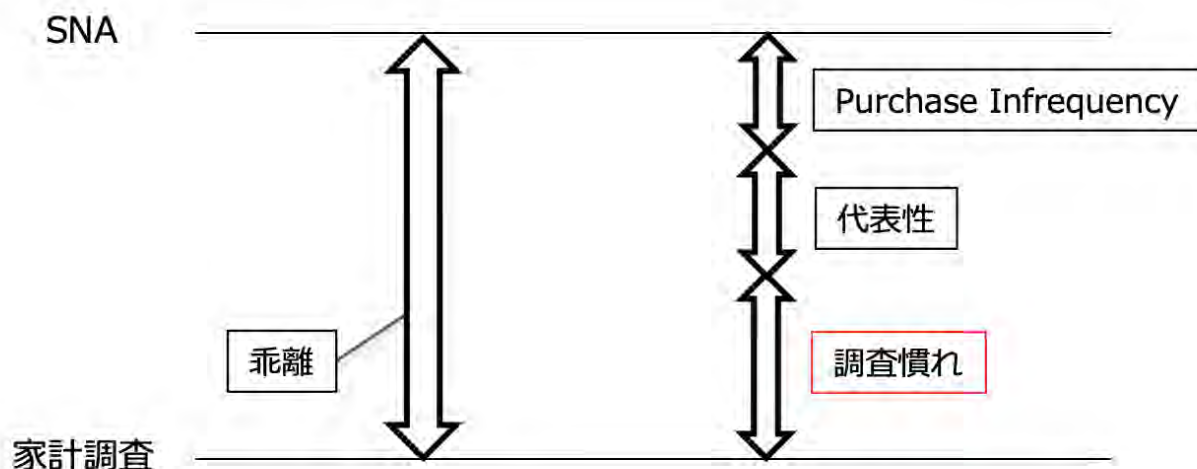
背景

○乖離の補正

今回の報告は、家計調査と全国消費実態調査の乖離に着目する中で調査継続による効果を考慮する

結果としてミクロ統計とマクロ統計の差の乖離を埋め合わせる議論にもつながるのではと考えられる

* 厳密には全消と家計の調査項目の違いがあるが、全国家計構造調査では調査項目自体の乖離は解消されるはず



再度まとめ

【本研究の目的】

全国家計構造調査の年平均の推定

このために家計調査を補助情報として利用する

【家計調査と全消(以前の全国家計構造調査)の乖離の原因】

同じ月の比較で全体として全消の方が大

- ①標本の違い 収入 家計 < 全消 $\Rightarrow$ 全消の消費大
- ②季節性 全消の対象月は↓
- ③調査継続による効果 家計調査は6か月継続のため↓
- ④調査項目対象(財・サービスの対象)

今回は②と③を調整

* 厳密には回帰によって①も一定程度考慮する

これまでの方法の問題点

家計調査の情報を利用した全国家計構造調査での年平均の推計

⇒ E M アルゴリズムの適用

* これまでは総支出と食費に限定（ゼロの家計いない）

① 一方、10大費目については教育・保険医療など“ゼロ”の存在する家計は十分多いはず

⇒ これまでの $\log(\text{数値})$ での扱いではうまくいかない

⇒ $\log(\text{数値})$ のTobitタイプのモデリングが必要では？

但し、EM適用なら24次元で積分計算も膨大で難しい

② 家計調査と家計構造調査の違いは“対象”“季節性”“継続効果”

全国家計構造調査は2か月vs家計調査は6か月

⇒ 何か月目の調査か、の項を導入

③ Breakdown(地域別性年代)しても推定できる母数の推定

共変量の併合（産業：パターンA）

併合前	併合後
1 農業，林業，漁業	第一次産業及び第二次産業
2 鉱業，採石業，砂利採取業	第一次産業及び第二次産業
3 建設業	第一次産業及び第二次産業
4 製造業	第一次産業及び第二次産業
5 電気・ガス・熱供給・水道業	第三次産業
6 情報通信業	第三次産業
7 運輸業，郵便業	第三次産業
8 卸売業，小売業	第三次産業
9 金融業，保険業	第三次産業
10 不動産業，物品賃貸業	第三次産業
11 学術研究，専門・技術サービス業	第三次産業
12 宿泊業，飲食サービス業	第三次産業
13 生活関連サービス業，娯楽業	第三次産業
14 教育，学習支援業	第三次産業
15 医療，福祉	第三次産業
16 複合サービス事業	第三次産業
17 サービス業(他に分類されないもの)	第三次産業
18 公務（他に分類されないもの）	第三次産業
19 その他	その他

共変量の併合（職業：パターンA）

併合前		併合後
1	常用労務者	勤労者世帯
2	臨時及び日々雇労務作業者	勤労者世帯
3	民間職員	勤労者世帯
4	官公職員	勤労者世帯
5	商人及び職人	非勤労者世帯
6	個人経営者	非勤労者世帯
7	法人経営者	非勤労者世帯
8	自由業者	非勤労者世帯
9	その他	非勤労者世帯
10	無職	非勤労者世帯
11	家族従業者	非勤労者世帯
12	農林漁業従事者	非勤労者世帯

* 併合基準は家計調査の職業分類表に基づく。

共変量の併合（産業：パターンB）

併合前	併合後
1 農業，林業，漁業	その他産業
2 鉱業，採石業，砂利採取業	その他産業
3 建設業	平均賃金グループ2
4 製造業	平均賃金グループ3
5 電気・ガス・熱供給・水道業	その他産業
6 情報通信業	平均賃金グループ1
7 運輸業，郵便業	平均賃金グループ4
8 卸売業，小売業	平均賃金グループ2
9 金融業，保険業	平均賃金グループ1
10 不動産業，物品賃貸業	その他産業
11 学術研究，専門・技術サービス業	平均賃金グループ1
12 宿泊業，飲食サービス業	平均賃金グループ4
13 生活関連サービス業，娯楽業	平均賃金グループ4
14 教育，学習支援業	平均賃金グループ1
15 医療，福祉	平均賃金グループ3
16 複合サービス事業	平均賃金グループ4
17 サービス業(他に分類されないもの)	平均賃金グループ4
18 公務（他に分類されないもの）	公務
19 その他	その他

* 平均賃金グループに関しては、賃金構造統計調査の平均賃金に基づく。

共変量の併合（職業：パターンB）

併合前		併合後
1	常用労務者	勤労者世帯
2	臨時及び日々雇労務作業者	勤労者世帯
3	民間職員	勤労者世帯
4	官公職員	勤労者世帯
5	商人及び職人	個人営業世帯
6	個人経営者	個人営業世帯
7	法人経営者	その他
8	自由業者	その他
9	その他	その他
10	無職	その他
11	家族従業者	その他
12	農林漁業従事者	個人営業世帯

* 併合基準は家計調査の職業分類表に基づく。

共変量の併合（住居の所有関係・住居の構造）

住居の所有関係

併合前	併合後
1 持ち家	持ち家
2 民営の賃貸住宅	借家
3 都道府県・市区町村営賃貸住宅	借家
4 都市再生機構・公社等の賃貸住宅	借家
5 社宅・公務員宿舎	借家
6 借間	借家

住居の構造

併合前	併合後
1 木造	木造
2 防災木造	木造
3 鉄骨・鉄筋コンクリート造	その他
4 その他（ブロック造・レンガ造等）	その他

参考：漸近論

- $N \rightarrow \infty$ のときの推定量の性質を述べる.
- 簡単化のため, ローターショナルパネルの欠測は無視する.
- 家計 i に対する確率変数のベクトルを z_i で表す.
- 真のデータ生成過程がタイプII Tobitであり, その誤差項が z_i と独立かつ, 次の多変量正規分布にi.i.d.に従うと仮定する.

$$\begin{pmatrix} \epsilon_i \\ \varepsilon_i \end{pmatrix} \equiv \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{i12} \\ \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{i12} \end{pmatrix} \sim N(O, \Sigma) = N\left(O, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right),$$

参考：漸近論

- ただし, Σ は次を満たすように定義された行列である.

$$\mathbb{E}[\epsilon_{t1}\epsilon_{s1}] = \begin{cases} 1 & \text{if } t = s \\ \sigma_{\text{cross}}^{(1)} & \text{if } t \neq s \end{cases},$$

$$\mathbb{E}[\varepsilon_{it}\varepsilon_{is}] = \begin{cases} \sigma^{(2)2} & \text{if } t = s \\ \sigma_{\text{cross}}^{(2)} & \text{if } t \neq s \end{cases},$$

$$\mathbb{E}[\epsilon_{it}\varepsilon_{is}] = \begin{cases} \sigma^{(12)} & \text{if } t = s \\ \sigma_{\text{cross}}^{(12)} & \text{if } t \neq s \end{cases}$$

- 誤差の月間の相関を捉えるために, 次ページの12変量プロビットモデルと非線形モデルを考える.

参考：漸近論

- **1st step**. 12変量プロビットモデル (パラメータ π_1)

$$P(w_i | \pi_1) = \int_{A_{i12}} \dots \int_{A_{i1}} \phi_{12}(u | 0, \Sigma_{11}) du$$

ただし, $w_{it} = \mathbf{1}\{y_{it}^* > 0\}$, $w_i = (w_{i1}, w_{i2}, \dots, w_{i12})^T$, $\phi_{12}(u | 0, \Sigma_{11})$

は平均0, 分散 Σ_{11} の 12 変量正規分布の確率密度関数,

A_{it} は区間

$$A_{it} = \begin{cases} (-\infty, \lambda^T x_i + \phi) & \text{if } w_{it} = 1 \\ [\lambda^T x_i + \phi, \infty) & \text{if } w_{it} = 0 \end{cases}$$

- **2nd step**. 非線形モデル (パラメータ π_1, π_2)

$$y_i = H(Z_i, \pi_1, \pi_2) + \sigma^{(12)} h(x_i, \pi_1) + U_i$$

ただし,

$$H(Z_i, \pi_1, \pi_2) = \begin{pmatrix} \phi(\lambda^T x_i + \phi) \left(\gamma^T x_i + \sum_{m=1}^{12} \alpha_m D_{i1} + \sum_{k=2}^6 \beta_k E_{i1k} \right) \\ \phi(\lambda^T x_i + \phi) \left(\gamma^T x_i + \sum_{m=1}^{12} \alpha_m D_{i2} + \sum_{k=2}^6 \beta_k E_{i2k} \right) \\ \vdots \\ \phi(\lambda^T x_i + \phi) \left(\gamma^T x_i + \sum_{m=1}^{12} \alpha_m D_{i12} + \sum_{k=2}^6 \beta_k E_{i12k} \right) \end{pmatrix}, h(x_i, \pi_1) = \begin{pmatrix} \phi(\lambda^T x_i + \phi) \\ \phi(\lambda^T x_i + \phi) \\ \vdots \\ \phi(\lambda^T x_i + \phi) \end{pmatrix}.$$

参考：漸近論

- 1st step目の推定方程式は，対数尤度を偏微分して

$$m_1(Z_i, \pi_1) = \frac{\partial}{\partial \pi_1} \log P(w_i | \pi_1)$$

- 2nd step目の推定方程式は，非線形最小二乗法の一階の条件より

$$m_2(Z_i, \pi_1, \pi_2) = \frac{\partial}{\partial (\pi_1^T, \pi_2^T)^T} \left\{ y_i - H(Z_i, \pi_1, \pi_2) - \sigma^{(12)} h(x_i, \pi_1) \right\}^T \left\{ y_i - H(Z_i, \pi_1, \pi_2) - \sigma^{(12)} h(x_i, \pi_1) \right\}$$

- $\pi = (\pi_1^T, \pi_2^T)^T$ とおく．
- スタックした推定方程式

$$m(Z_i, \pi) = \begin{bmatrix} m_1(Z_i, \pi_1) \\ m_2(Z_i, \pi_1, \pi_2) \end{bmatrix}$$

をモーメント条件として扱う．

- 正則条件のもとで一致性と漸近正規性を得る．

参考：漸近論

- 特に, 2nd step目のモデルの推定量は次の漸近正規性を持つ.

$$\sqrt{N}(\hat{\pi}_2 - \pi_{20}) \xrightarrow{d} N(0, V)$$

ただし,

$$V = M_{22}^{-1} \mathbb{E} \left[\{m_2(Z_i, \pi_{10}, \pi_{20}) + M_{21} \psi(Z_i)\} \{m_2(Z_i, \pi_{10}, \pi_{20}) + M_{21} \psi(Z_i)\}^T \right] (M_{22}^{-1})^T$$

かつ,

$$M_{22} = \mathbb{E} \left[\frac{\partial}{\partial \pi_2^T} m_2(Z_i, \pi_{10}, \pi_{20}) \right], M_{21} = \mathbb{E} \left[\frac{\partial}{\partial \pi_1^T} m_2(Z_i, \pi_{10}, \pi_{20}) \right],$$

$$M_{11} = \mathbb{E} \left[\frac{\partial}{\partial \pi_1^T} m_1(Z_i, \pi_{10}) \right], \psi(Z_i) = -M_{11}^{-1} m_1(Z_i, \pi_{10})$$

- 証明は例えばNewey and McFadden (1994) を参照.

参考：BICによるモデル選択における共変量候補

10大費目	推定方法	パターン1	パターン2	パターン3	パターン4
食料	重回帰分析	1,2	1,2,7,8	1,2,7,8,13a,14a	1,2,7,8,13b,14b
住居	Tobit-Type2	1,2,11	1,2,11,15,16	1,2,7,8,11,13a,14a,15,16,17,18	1,2,7,8,11,13b,14b,15,16,17,18
光熱、水道	重回帰分析	1,2	1,2,7,8	1,2,7,8,13a,14a	1,2,7,8,13b,14b
家具、家事用品	重回帰分析	1,2	1,2,7,8	1,2,13a,14a	1,2,7,8,13a,14a
被服及び履物	Tobit-Type2	1,2,7,8	1,2,7,8,12	1,2,7,8,12,13a,14a	1,2,7,8,12,13b,14b
保険医療	Tobit-Type2	1,2,7,8	1,2,7,8,13a,14a	1,2,7,8,12,13a,14a	1,2,7,8,12,13b,14b
交通、通信	重回帰分析	1,2	1,2,7,8	1,2,13a,14a	1,2,7,8,13a,14a
教育	Tobit-Type2	1,2,12	1,2,7,8,12	1,2,7,8,12,13a,14a	1,2,7,8,12,13b,14b
教養娯楽	重回帰分析	1,2	1,2,7,8	1,2,13a,14a	1,2,7,8,13a,14a
その他の消費支出	重回帰分析	1,2	1,2,7,8	1,2,13a,14a	1,2,7,8,13a,14a

共変量候補 (上図の赤文字の組合せが、4種類の組合せの中で、BIC最小の共変量の組合せ):

1. 年間収入, 2. 世帯人員, 3. 就業人員, 4. 企業規模, 5. 産業, 6. 職業, 7. 世帯主の性別, 8. 世帯主の年齢, 9. 住居の構造, 10. 住居の所有関係, 11. 住居の延面積, 12. 就学者人数, 13a. 産業併合A, 14a. 職業併合A, 13b. 産業併合B, 14b. 職業併合B, 15. 住居の所有関係併合, 16. 住居の構造併合, 17. 居住室数, 18. 居住室の畳数

*産業分類	13a	13b
1. 農業・林業・漁業	第一次・第二次産業	その他産業
2. 鉱業・採石業・砂利採取業	第一次・第二次産業	その他産業
3. 建設業	第一次・第二次産業	民間・平均賃金2
4. 製造業	第一次・第二次産業	民間・平均賃金3
5. 電気・ガス・熱供給・水道業	第三次産業	その他産業
6. 情報通信業	第三次産業	民間・平均賃金1
7. 運輸業・郵便業	第三次産業	民間・平均賃金4
8. 卸売業・小売業	第三次産業	民間・平均賃金2
9. 金融業・保険業	第三次産業	民間・平均賃金1
10. 不動産業・物品賃貸業	第三次産業	その他産業
11. 学術研究・専門・技術サービス業	第三次産業	民間・平均賃金1
12. 宿泊業・飲食サービス業	第三次産業	民間・平均賃金4
13. 生活関連サービス業・娯楽業	第三次産業	民間・平均賃金4
14. 教育・学習支援業	第三次産業	民間・平均賃金1
15. 医療・福祉	第三次産業	民間・平均賃金3
16. 複合サービス事業	第三次産業	民間・平均賃金4
17. サービス業(他に分類されないもの)	第三次産業	民間・平均賃金4
18. 公務(他に分類されないもの)	第三次産業	公務
19. その他	その他	その他

*職業分類	14a	14b
1. 常用労働者	勤労者世帯	勤労者世帯
2. 臨時及び日々雇労働作業者	勤労者世帯	勤労者世帯
3. 民間職員	勤労者世帯	勤労者世帯
4. 官公職員	勤労者世帯	勤労者世帯
5. 商人及び職人	勤労者以外の世帯	個人営業世帯
6. 個人経営者	勤労者以外の世帯	個人営業世帯
7. 法人経営者	勤労者以外の世帯	その他
8. 自由業者	勤労者以外の世帯	その他
9. その他	勤労者以外の世帯	その他
10. 無職	勤労者以外の世帯	その他
11. 家族従業者	勤労者以外の世帯	その他
12. 農林漁業従事者	勤労者以外の世帯	個人営業世帯