

# **(3) 全国家計構造調査「年平均推定値」 の推定方法について**

**家計調査を利用した  
全国消費実態調査の時点間補完  
に向けて—t分布の利用—**

**慶應義塾大学 経済学部・大学院経済学研究科  
(兼)総務省統計研究研修所 星野崇宏**

**慶應義塾大学大学院経済学研究科修士課程 慶野有輝**

# 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

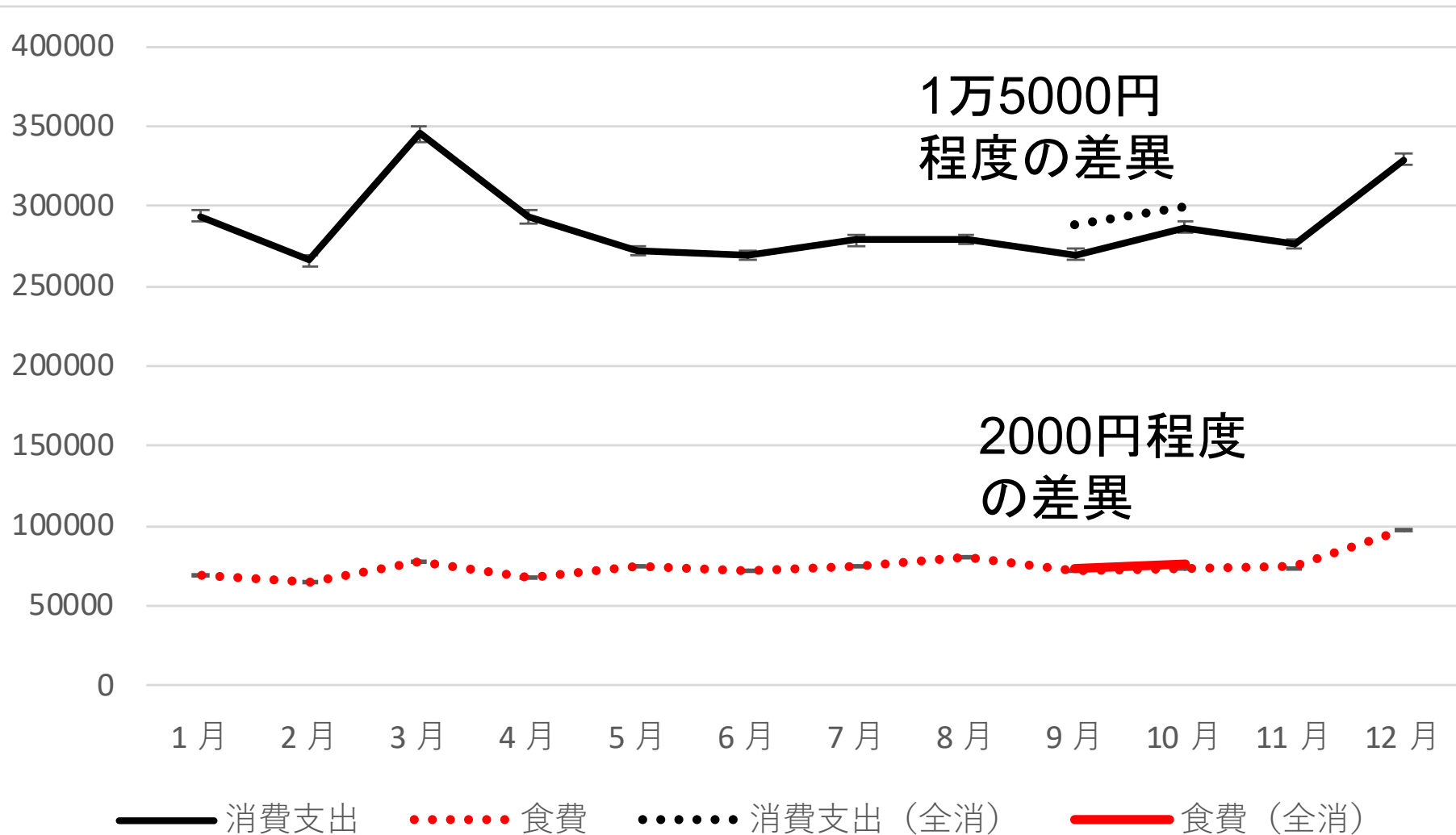
通年での推測をしたい

○個票データでの補完 ×時系列データとしての補完

⇒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)

# 家計調査と全国消費実態調査の調査の違い

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

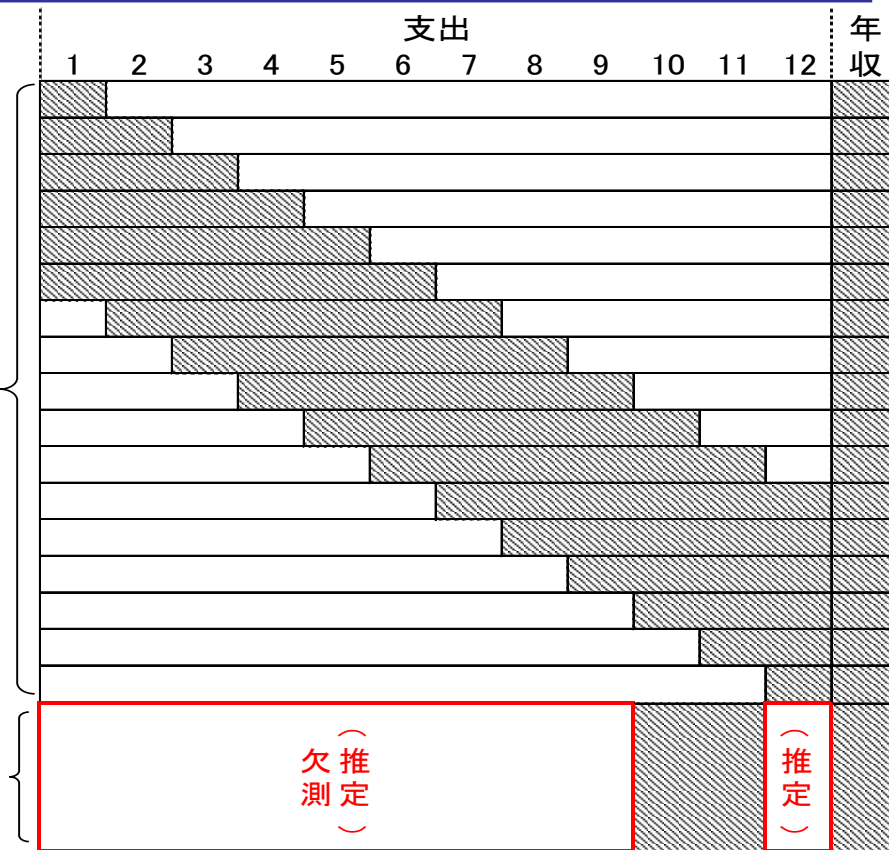


調査モードの違い

# 提案手法

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) 両者は共に代表性が  
あるとして月次の平均は共通

家計調査

全国消費  
実態調査

上図) 2人以上世帯の場合  
「全国消費実態調査・年平均値推定ロードマップ案」より引用

\* 仮定Bは無くても識別可能

## 【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

# 3月の発表内容

---

【目的】 月次の変動が大きい消費について両調査を融合させて通年での推測／年平均の計算

【方法論】 個票データで欠測が存在していることを仮定した  
欠測データ解析

具体的にはEMアルゴリズムを用いた平均と相関構造の推定

\* 時系列にして季節性調整とは異なる個票データ解析

【結果：「全消」と「家計」の違いの考慮】

①質問項目の違い②調査時期の違い③調査対象の違い

⇒これらを考慮して

「全消の対象者が1年通じて調査に答えた場合の年平均」算出

---

# いくつかの分析

---

(1) 2調査の異質性を考慮しない場合

(2) 2調査の異質性を考慮した場合

(3) 年収を共変量として利用する場合 (2調査の異質性なし)

(4) 年収を共変量として利用する場合 (2調査の異質性考慮)

\* 2調査の異質性

⇒ 「調査モード」と「2調査の標本の標本誤差」両者が存在

---

# 年平均の値について

(A)全消での9・10月の平均（月）      (B)家計調査での9・10月の平均

(C)家計調査での年平均

(D)上記を融合した結果から“全国消費実態調査”の調査対象者が年間で答えたと仮定した場合の推計値      ⇒方法(1)と方法(2)の融合

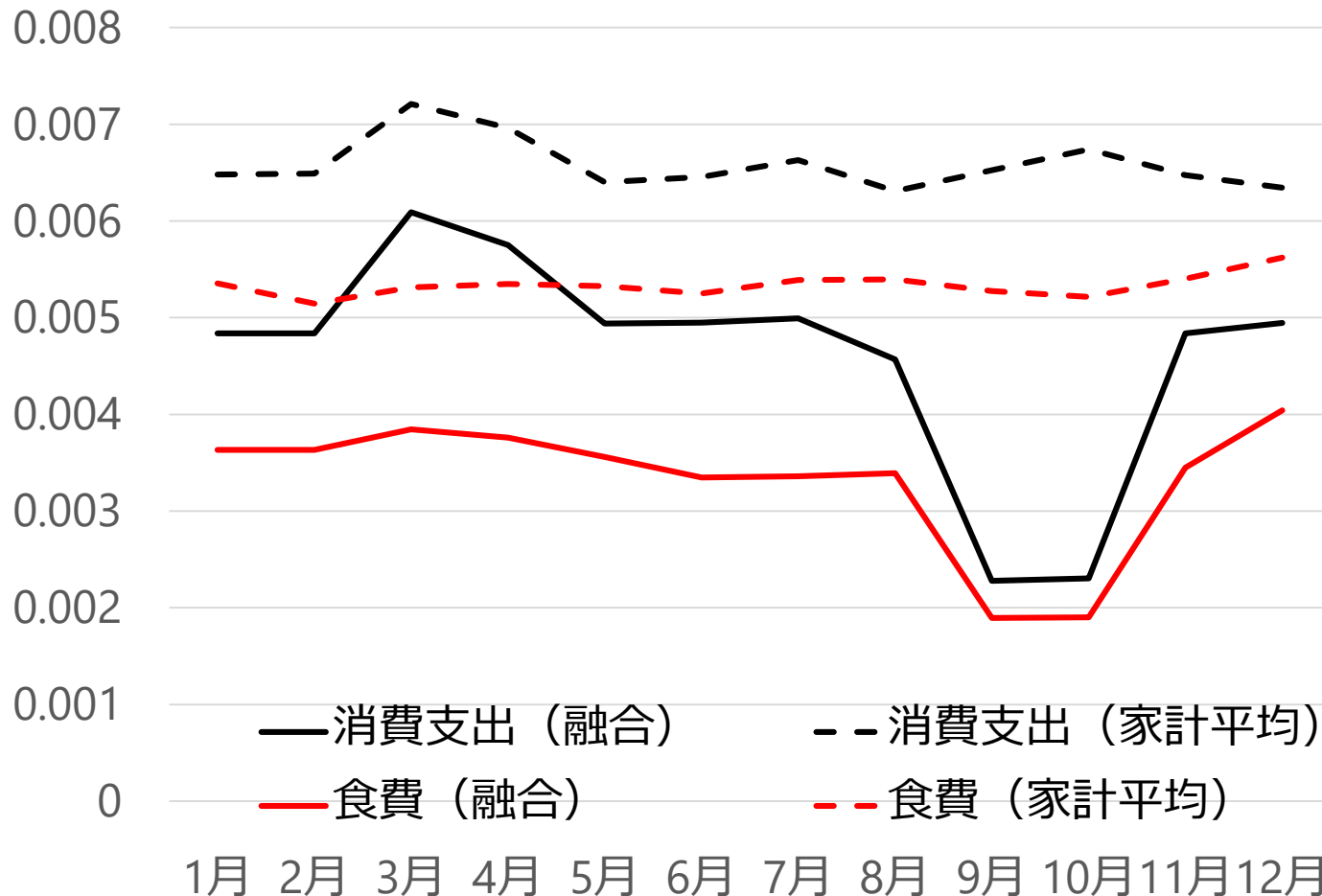
(E)上記を融合：収入を共変量とし調査モードが両者が異なることを仮定した場合      ⇒方法(4)の場合

方法	(A)	(B)	(C)	(D)	(E)
消費支出	294090	278534	288536	302886	296716
食費	74803	72900	75168	76979	76196



# 解析結果（標準誤差：両者共通平均共分散）

## 標準誤差について



全国消費実態調査が加わる9・10月はもちろん、他の月も改善。

# 前回のまとめ

---

## 【まとめ】

- ・ 全国消費実態調査は9・10月の実施だが消費の季節変動は大

⇒単にこれを6倍して年間の消費額や食費とすると生活保護世帯への支給額の算定など様々な政策的決定を誤る危険性

- ・ 今回は集計時系列上での季節調整ではなくミクロレベルでの調整

⇒標準偏差が減少し融合した家計調査にもメリット

## 【今後実施すべき項目】

- ・ 多変量t分布の実装

\* 外れ値が多いため対数にしても正規分布では不適切？

- ・ 単身世帯でのプログラム作成と2人以上との統合

- ・ 代入の可能性の検討(多重代入でないとは分散過小評価)
-

# 多変量 t 分布の仮定

- 家計調査のデータと全国消費実態調査のデータ(9・10月)を融合したことの効果としての平均値と標準誤差の変化

(家計調査単独)

(全消と家計調査との融合)

12か月の各平均、その標準誤差

- ここでのモデルは家計調査も全国消費実態調査も同じ多変量 t 分布に従うと仮定(自由度 $\nu$ : $\nu \rightarrow \infty$ で多変量正規分布)

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_{12} \end{bmatrix} \sim t_{12} \left( \mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_{12} \end{bmatrix}, \Phi = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1,12} \\ \vdots & \ddots & \vdots \\ \sigma_{1,12} & \cdots & \sigma_{12}^2 \end{bmatrix}, \nu \right)$$

# 多変量 t 分布とEMアルゴリズム

---

Liu and Rubin(1995)よりこの多変量 t 分布に従うYは

$$Y \sim N_{12}(\mu, \Phi/\tau)$$

とし $\tau$ がガンマ分布

$$\tau \sim \text{Gamma}(v/2, v/2)$$

に従うと考えるてもよい。ここで $\tau$ を観測できない潜在変数と考えてEMアルゴリズムを実行すれば

①  $\tau$ が分かっているとした場合にYの中の観測値部分（全消なら9・10月）が得られているときの欠測値部分（同9・10以外）の期待値を計算

② この期待値を代入して $\mu$ と $\Phi$ を推定

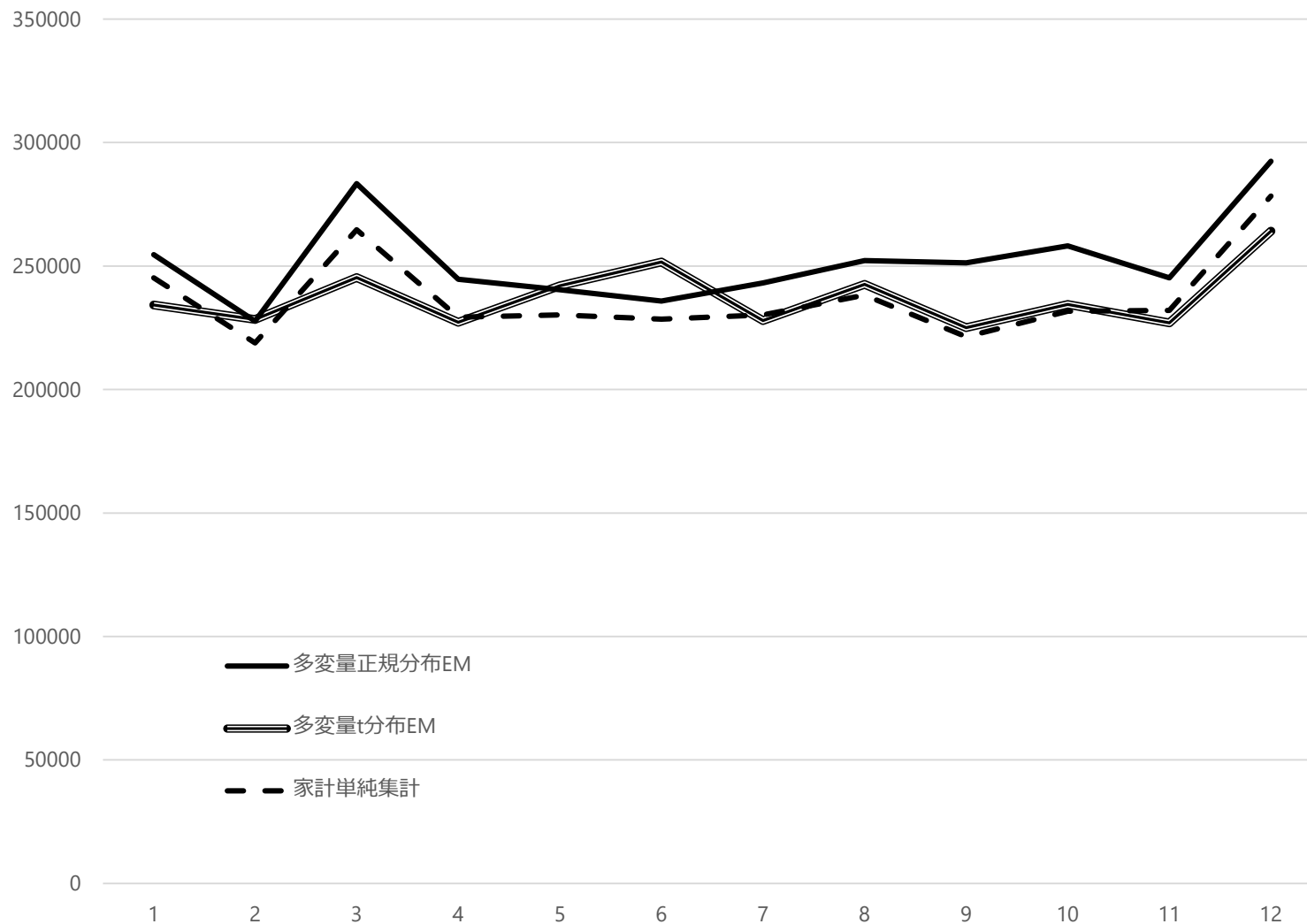
③  $\tau$ の期待値を代入して $v$ を推定

のEMアルゴリズムが利用可能

---

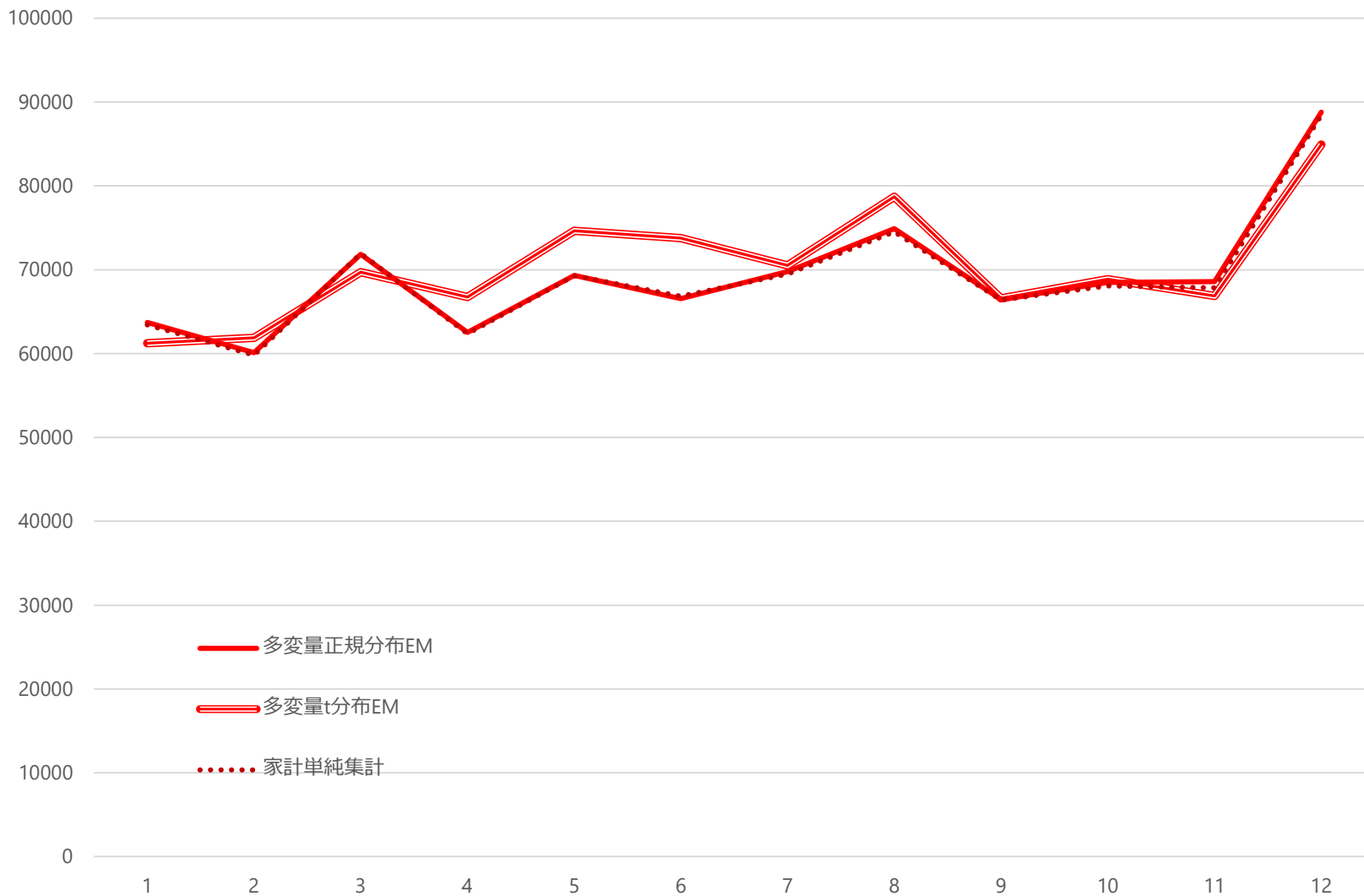
# 解析結果（消費支出中央値）

横軸：月，縦軸：消費支出



# 解析結果（食費中央値）

横軸：月，縦軸：食費



# 年中央値の値について

---

(A)全消での9・10月の中央値

(B)家計調査での9・10月の中央値

(C)家計調査での年中央値

(D)消費支出に多変量正規分布を仮定した場合のEMアルゴリズムを仮定した場合の年中央値推定

(E)消費支出に多変量t分布を仮定した場合のEMアルゴリズムを仮定した場合の年中央値推定

	(A)	(B)	(C)	(D)	(E)
消費支出	242435	230439	237381.3	238948.8	237468.1
食費	68776	66863	69042.83	70377.3	70377.3

---

# ヒストグラムと対数正規、対数 t の当てはまり

実際の消費支出と推定結果を用いた確率密度関数の比較

ヒストグラム：消費支出

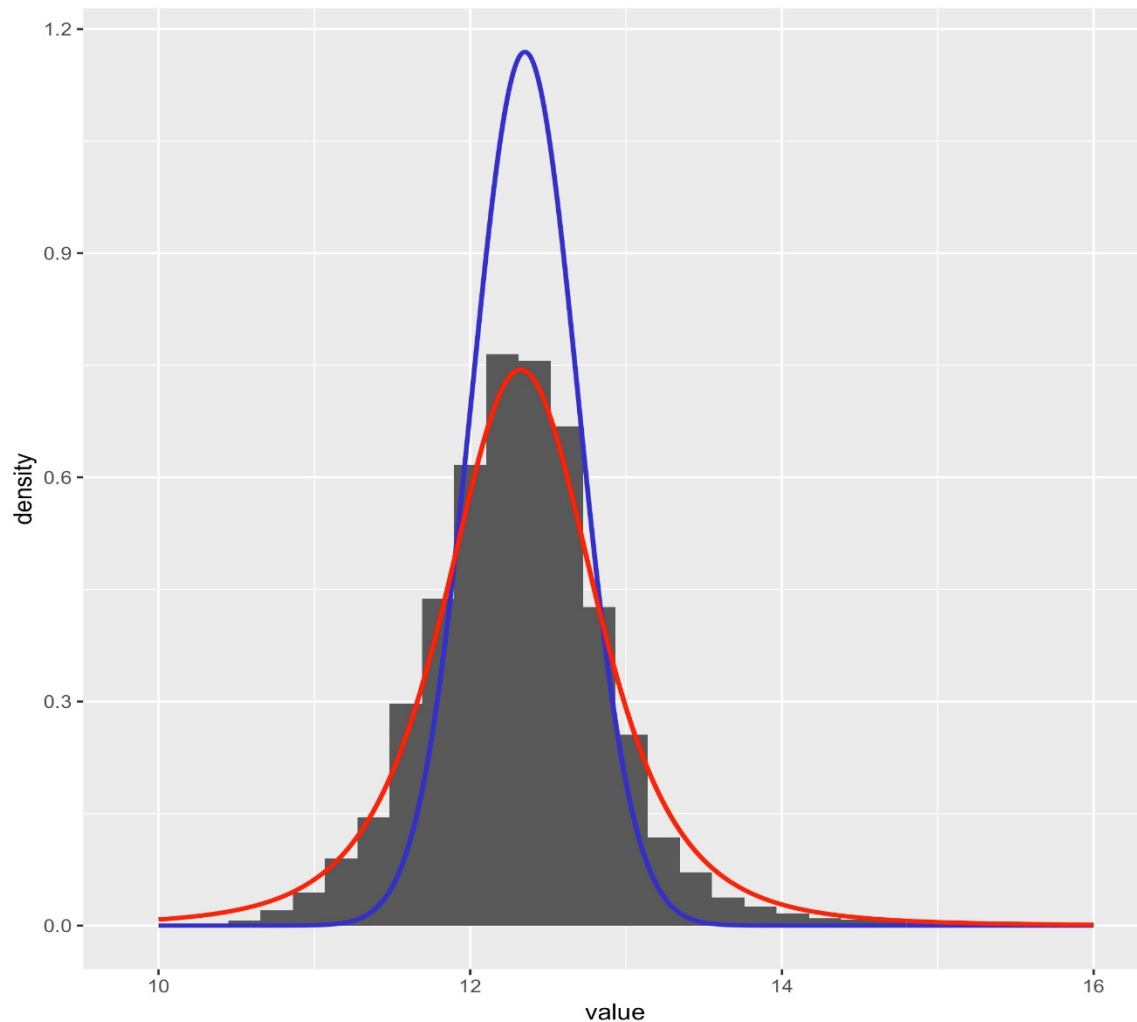
青線：

多変量正規分布を仮定した  
もの

赤線：

多変量t分布を仮定した  
もの

※図はそれぞれ2014年9月  
のものを使用





# 参考文献

---

- Kim, JK. and Shao, J. (2014) *Statistical Methods for Handling Incomplete Data*. Chapman & Hall.
- Little, R.A.J and Rubin, D.B. (2002) *Statistical Analysis with Missing Data*, 2<sup>nd</sup> ed. Wiley.
- Liu, C. and Rubin, D.B. (1995) "ML Estimation Of The t-Distribution Using EM and Its Extensions, ECM and ECME", *Statistica Sinica*, 5, 19-39.
- Ridder, G. & Moffitt, R. (2007) The Econometrics of Data Combination, in J. J. Heckman & E. E. Leamer, eds, 'Handbook of Econometrics', Vol. 6, Elsevier, chapter 75, pp.5469-5547.
- Schafer, J.L. (1997) *Analysis of Incomplete Multivariate Data*, Chapman & Hall.
- 星野崇宏(2009)「調査観察データの統計科学」岩波書店
- 高井啓二・星野崇宏・野間久史 (2016)「欠測データの統計科学」岩波書店
-