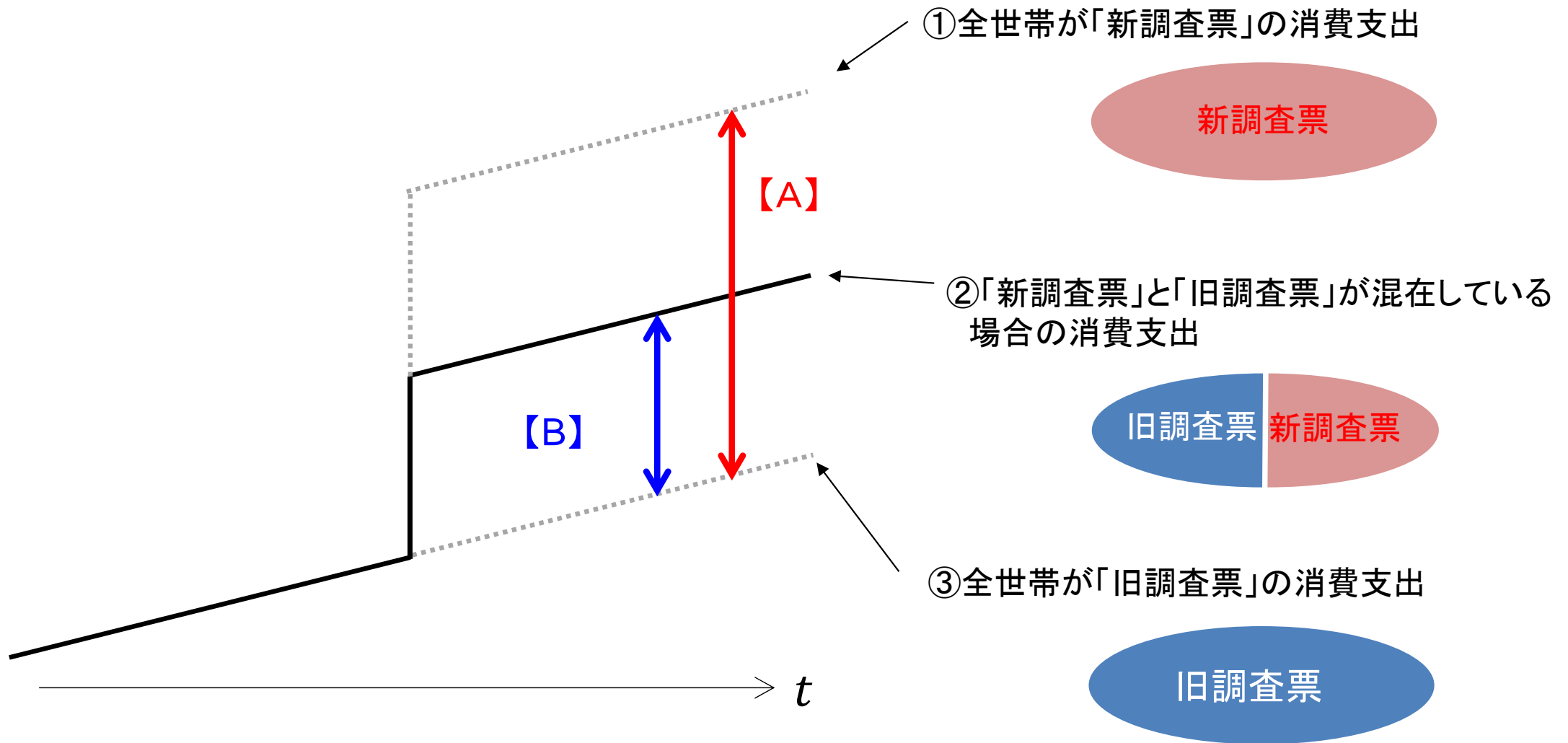


# 家計調査の調査方法変更による断層の 推定方法について

平成30年1月25日  
総務省統計局

# 家計調査の調査方法変更による断層のイメージ



【A】: 全てが新調査票とした場合と全てが旧調査票とした場合の断層

【B】: 一部が新調査票に切り替わった場合の断層

# 断層の推定方法（案）

## (1) 重回帰モデルに基づく方法

$$\log(y_i) = \beta_0 + \beta_1 x_{i1} + \beta_1 x_{i2} \cdots + \beta_p x_{ip} + \gamma D_i + \varepsilon_i$$

( $y_i$  : 消費支出)

各種属性 (log(年収)、年齢、性別、職業、地域等) をコントロール

$D_i := 0$  (旧調査票)  
 $= 1$  (新調査票)

$\Rightarrow \exp(\hat{\gamma})$  が新・旧の断層 (倍率) の推定値。

(※全てが新調査票の場合と全てが旧調査票の場合との差になることに注意)

## (2) 差分の差 (difference in differences) 推定に基づく方法

	切替え:あり	切替え:無し
2017年9月 (T = 0)	旧【B】 (D = 0)	旧【D】 (D = 0)
2017年10月 (T = 1)	新【A】 (D = 1)	旧【C】 (D = 0)

A - B = 「新調査票の効果」 + 「9月から10月の変化」

C - D = 「9月から10月の変化」

(A - B) - (C - D) = 新調査票の効果。

$$\log(y_i) = \beta_0 + \beta_1 x_{i1} + \beta_1 x_{i2} \cdots + \beta_p x_{ip} + \gamma D_i + \lambda T_i + \delta D_i T_i + \varepsilon_i$$

各種属性 (log(年収)、年齢、性別、職業、地域等) をコントロール

$\Rightarrow \exp(\hat{\delta})$  が新調査票の効果の推定値。

## 断層の推定方法（案）（続き）

### (3) 傾向スコア・IPW（Inverse Probability Weighting）推定に基づく方法

- ・ 新調査票（旧調査票）に割り当てられた場合に1(0)となる割当て変数( $Z_i$ )を導入。
- ・ 属性変数を用いて新調査票への割当て確率  $e_i$ （傾向スコア）を推定する  
ロジスティック回帰モデルを構築。

$$\log(e_i/(1 - e_i)) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$$

$$\left[ \begin{array}{l} e_i : \text{新調査票への割当て確率} \quad Z_i := 0 \text{ (旧調査票)} = 1 \text{ (新調査票)} \\ x_{i1}, x_{i2}, \dots, x_{ip} : \text{コントロール変数 (log(年収)、性別、年齢、職業、地域 等)} \end{array} \right]$$

- ・ 傾向スコアの推定値 ( $\hat{e}_i$ ) に基づくIPW推定により、仮に全てが「旧調査票」であった場合の母数団全体の平均値  $\hat{E}^{IPW}(y_0)$  を推定。

$$\hat{E}^{IPW}(y_0) = \sum_{i=1}^N \frac{(1 - z_i)w_i y_i}{(1 - e_i)} / \sum_{i=1}^N \frac{(1 - z_i)w_i}{1 - e_i}$$

# 断層の推定：試算の方法

## (1) 分析に使用したデータ

### (1) 家計調査：

- ・平成29年9月分結果（8,424世帯：新調査票対象は4,175世帯）
- ・平成29年10月分結果（8,413世帯：新調査票対象は4,183世帯）

### (2) 家計消費状況調査：

- ・平成29年9月分結果（22,020世帯）
- ・平成29年10月分結果（22,205世帯）

### ※使用した調査事項：

家計消費状況調査の家計調査補完品目及び対応する家計調査の支出項目

## (2) 試算用データの作成

- ・家計調査における新調査票対象世帯について「類似」した家計消費状況調査の世帯を統計的にマッチング（「類似」度は共通の調査事項に基づく「Gower距離」で計測）。
- ・家計消費状況調査の特定50品目への消費支出額と、対応する家計調査の消費支出額との差を、家計調査の世帯の消費支出に加算。
- ・マッチング前後の差を「真」の断層とみなして分析。

# 断層の推定結果

## 今回作成した分析用データにおける1世帯当たりの消費支出

- (1) 全てが旧調査票の場合 : 238,030円
- (2) 一部が新調査票の場合 : 257,090円  
⇒「真」の断層  $((2) \div (1))$  は、1.08007。

## 断層の推定結果

推定手法	断層の推定値	
(1) 重回帰モデルに基づく方法	<u>1.10223</u>	← 5,275円 $((1) \times 0.02216)$ の差
(2) 差分の差推定に基づく方法	<u>1.07491</u>	← 1,228円 $((1) \times 0.00516)$ の差
(3) 傾向スコアに基づく方法	<u>1.08247</u>	← 571円 $((1) \times 0.00240)$ の差