Decomposition Analysis

労働経済学 2

川田恵介

Table of contents

1　近似的なバランス

- バランス後の比較は、Nonparametric Decomposition Analysis おいても有益
  - 格差の「要因」について知見を得られる
  - 伝統的な Kitagawa-Oaxaca-Blinder 分解に比べて、定式化への依存度を下げることができ、より頑強な分析が可能
  - 詳細は、Opacic, Wei, and Zhou (2023) を参照

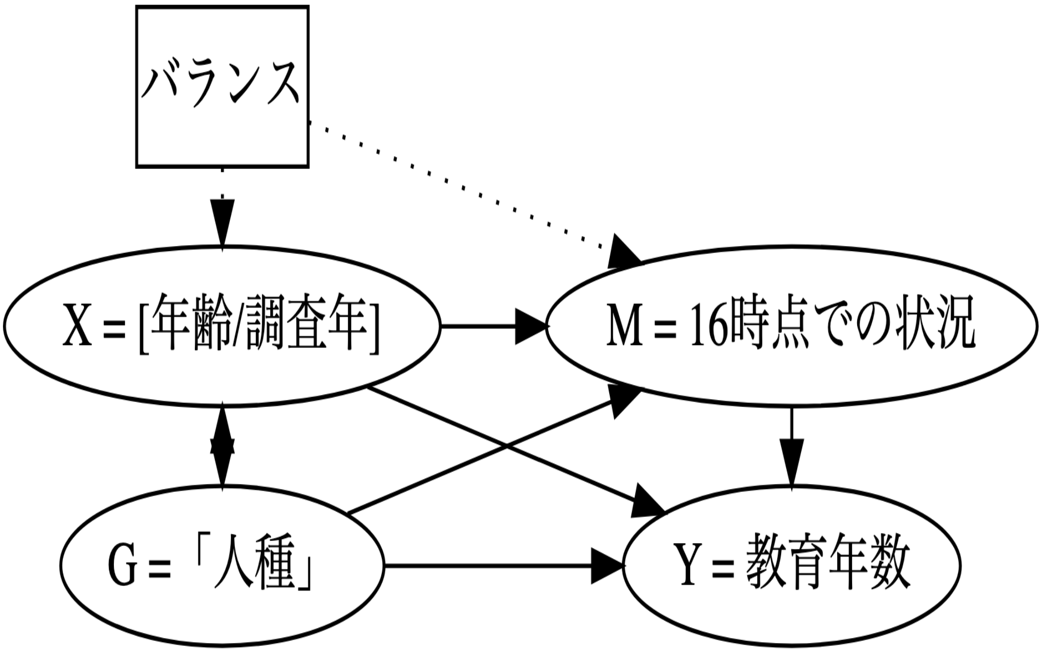
1.1　分解分析

- 差が生じる要因について議論
- 例: 人種間教育格差について、16 才時点での経済/社会状況がどの程度関係しているのか？
  - 記述的な分解分析:「会計的に」16 歳時点での経済/社会状況をバランスさせたら、格差はどのように変化するのか？
  - 因果的分解: 16 歳時点での経済/社会状況をバランスさせるような介入を行うと、格差はどのように変化するのか？

1.2　格差の分解

- 格差を定義するためのバランス (例: 年齢と調査年のバランス) のみを行った平均差と、格差の要因となるうる変数 (16 歳時点での状況) も含めてバランスさせた平均差、を比較する

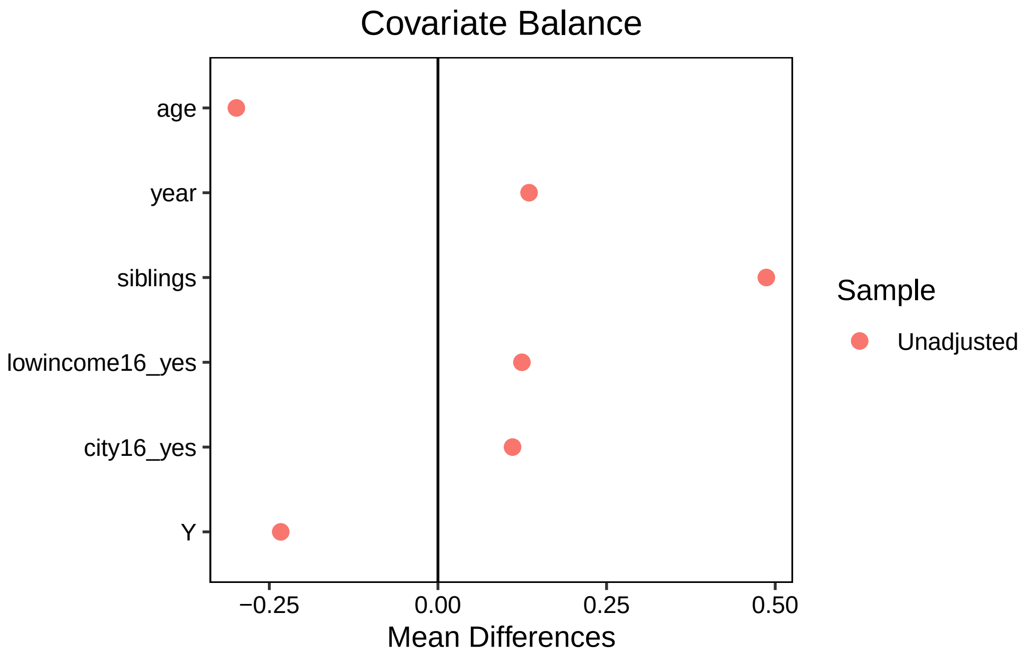
1.3 イメージ



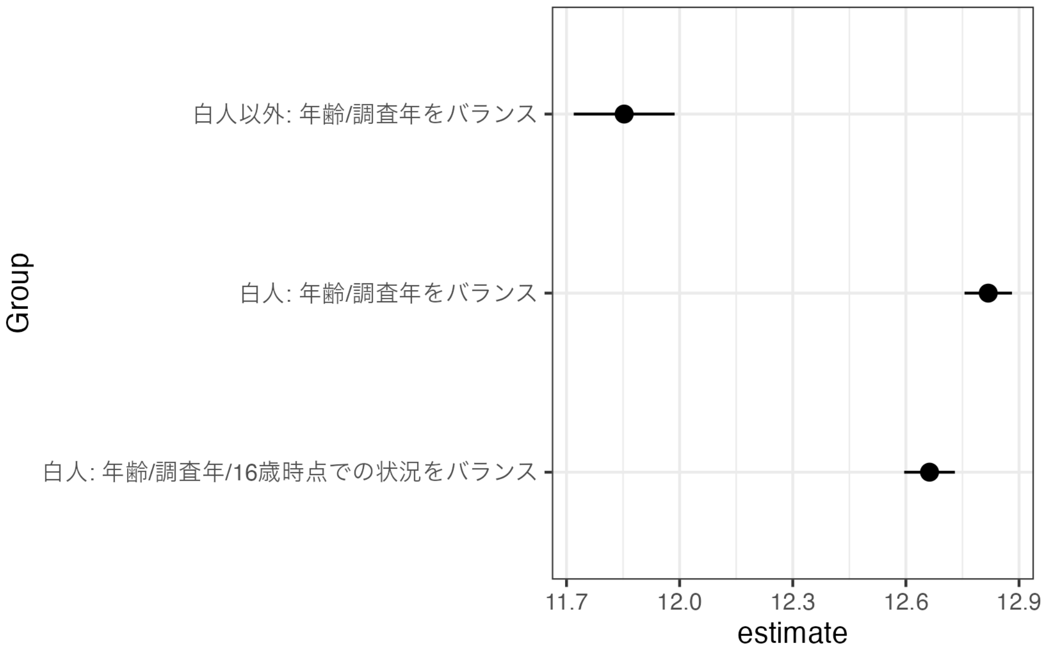
1.4 実例: 教育格差

- 「人種」間教育格差を推定
  - 同じ生まれ年での平均差を格差と定義
- 「人種」が教育格差をもたらす要因として、子供の時期の経済/社会状態の格差がありえる
  - もし生まれ年 + 子供の時期の経済/社会状態の格差をバランスさせた場合、格差はどのように変化するか?
    - 経済/社会状態の格差の「貢献」

1.5　実例: 教育格差（「白人」以外-　「白人」）



1.6　実例: 教育格差



2 Sequential Decomposition

- 先の事例を詳細に説明
- Duncun/Generalized-Kitagawa-Oaxaca-Blinder Decomposition (Opacic, Wei, and Zhou 2023)

2.1 記号

- d, x 内での、条件付き平均値 $E[Y|d,x]$
- d 内での、$X = x$ の条件付き割合: $f(x|d)$

2.2 復習: 繰り返し期待値の法則

- 一般にデータ上の Y の平均値は以下のように書き換えられる

$$\underbrace{d\text{における}Y\text{の平均値}}_{E[Y|d]} = \left\{ \underbrace{(x\ \&\ d)\text{における}Y\text{の平均値}}_{E[Y|d,x]} \right.$$

$$\left. \times \underbrace{d\text{における}x\text{の割合}}_{f(x|d)} \right\} \text{の}x\text{についての総和}$$

2.3 Counterfactual quantity

- X の分布が $D = 1$ と同じ場合の、Y の平均値

$$C_X[Y|D=0] = \left\{ \underbrace{(x\ \&\ 0)\text{における}Y\text{の平均値}}_{E[Y|0,x]} \right.$$

$$\left. \times \underbrace{1\text{における}x\text{の割合}}_{f(x|1)} \right\} \text{の}x\text{についての総和}$$

2.4 平均格差

-

$$E[Y|D=1] - C_X[Y|D=0]$$

$$= \left\{ [\underbrace{E[Y|D=1,X] - E[Y|D=0,X]}_{X\text{内格差}}] \right.$$

$$\left. \times \underbrace{1\text{における}x\text{の割合}}_{f(x|1)} \right\} \text{の}x\text{についての総和}$$

2.5 実装

```r
library(WeightIt)
library(marginaleffects)

Weight = weightit(
  D ~ age + year,
  Data,
  method = "ebal",
  estimand = "ATT" # D=1 の分布に揃える
  )

Fit = lm_weightit(
  Y ~ D + age + year,
  Data,
  weightit = Weight)

avg_predictions(Fit,variables = "D")
```

```
 D Estimate Std. Error   z Pr(>|z|)   S 2.5 % 97.5 %
 0     12.8     0.0321 399   <0.001 Inf  12.8   12.9
 1     11.9     0.0682 174   <0.001 Inf  11.7   12.0

Type:  probs
Columns: D, estimate, std.error, statistic, p.value, s.value, conf.low, conf.high
```

2.6 実装: バランスの追加

- X, M の分布が $D=1$ と同じ場合の、Y の平均値

$$C_{XM}[Y|D=0]=\Biggl\{\underbrace{(x\ \&\ m\ \&\ 0)\text{における}Y\text{の平均値}}_{E[Y|0,x,m]}$$

$$\times\underbrace{1\text{における}x,m\text{の割合}}_{f(x,m|1)}\Biggr\}\text{の}x\text{についての総和}$$

2.7 実装: 16 時点での状況もバランス

```
Weight = weightit(
  D ~ age + year + siblings + lowincome16 + city16, # 兄弟数と 16 際時点での家計所得/居住地を加える
  Data,
  method = "ebal",
  estimand = "ATT" # D=1 の分布に揃える
  )

Fit = lm_weightit(
  Y ~ D + age + year + siblings + lowincome16 + city16,
  Data,
  weightit = Weight)

avg_predictions(Fit,variables = "D")
```

```
 D Estimate Std. Error   z Pr(>|z|)   S 2.5 % 97.5 %
 0     12.7     0.0343 370   <0.001 Inf  12.6   12.7
 1     12.1     0.0689 175   <0.001 Inf  11.9   12.2

Type:  probs
Columns: D, estimate, std.error, statistic, p.value, s.value, conf.low, conf.high
```

2.8 Sequential decomposition

- $$E[Y|D=1]-C_X[Y|D=0]$$
$$=\underbrace{E[Y|D=1]-C_{MX}[Y|D=0]}_{Unexplained\ by\ M\ and\ X}$$
$$+\underbrace{C_{MX}[Y|D=0]-C_X[Y|D=0]}_{Explained\ by\ M}$$

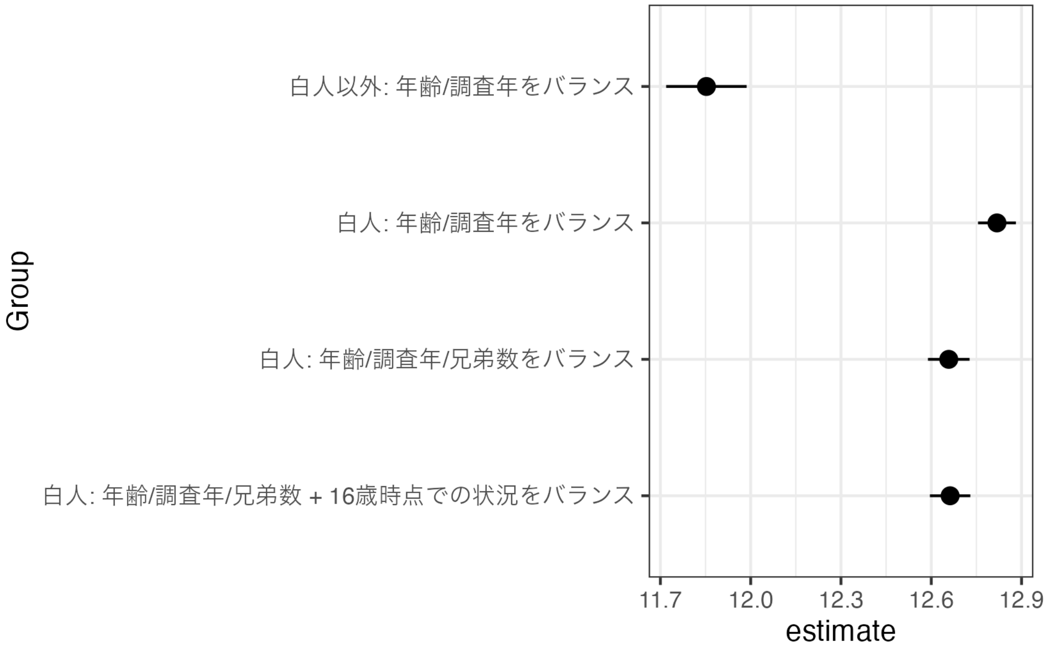
2.9 Sequential decomposition: 発展

- M に加えて、N もバランスさせた Counterfactual は、$C_{NMX}[Y|D=0]$
- $$E[Y|D=1]-C_X[Y|D=0]$$
$$=\underbrace{E[Y|D=1]-C_{NMX}[Y|D=0]}_{Unexplained\ by\ N\ M\ and\ X}$$

$$+\underbrace{C_{NMX}[Y|D=0]-C_{MX}[Y|D=0]}_{Explained\ by\ N}$$

$$+\underbrace{C_{MX}[Y|D=0]-C_{X}[Y|D=0]}_{Explained\ by\ M}$$

2.10 実例: 教育格差



2.11 順番問題

- Balance させる順番によって、分解結果は異なる
  - M = 兄弟数, N = 16歳時点での状況 と M = 16歳時点での状況, N = 兄弟数 で兄弟数の貢献は異なる値として推定される

2.12 因果的解釈

- $C_{MX}[Y|D=0]-C_{X}[Y|D=0]$ を「M の格差をなくす介入」による格差の縮小と見做せるか?
  - Causal Decomposition Analysis (Opacic, Wei, and Zhou 2023)
- 重要な仮定: M への介入によって、$E[Y|d,X,M]$ は変化しない
  - 通常の因果推論と同じような仮定が必要
    - No interference

* d, X 内で M はランダムに決まっている

2.13 補論: KBO 分解

- Kitagawa-Blinder-Oaxaca 分解
- $E[Y|D=d,X] = \beta_{0,d} + \beta_{1,d}X_1 + .. + \beta_{L,d}X_L$ を仮定
  - $\beta_{0,d}, .., \beta_{L,d}$ を OLS で推定

2.14 補論: KBO 分解

- $$\begin{aligned} & E[Y|D=1] - E[Y|D=0] \\ = & \underbrace{\beta_{1,0} - \beta_{0,0} + (\beta_{1,1} - \beta_{1,0})E[X_1|D=0]}_{Unexplained} + .. \\ & + \underbrace{\beta_{1,1}(E[X_1|D=1] - E[X_1|D=0]) + ..}_{Explained} \end{aligned}$$
- 線型モデルの仮定に強く依存

2.15 Referene

Opacic, Aleksei, Lai Wei, and Xiang Zhou. 2023. "Disparity Analysis: A Tale of Two Approaches."