

What If 分析

労働経済学 2

川田恵介

Table of contents

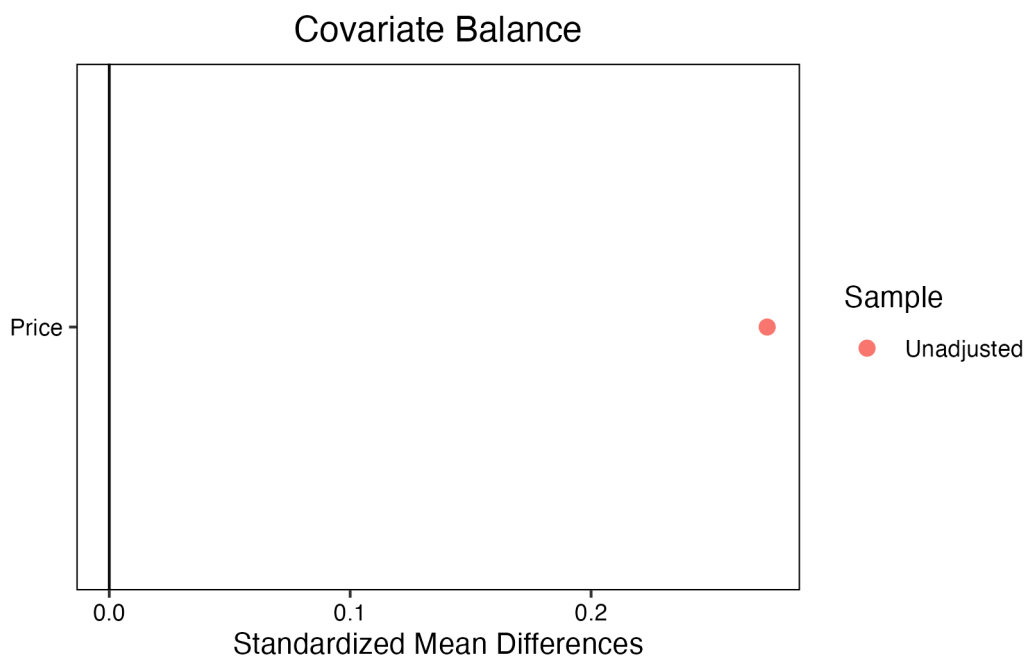
1	分析例	2
1.1	例: 2023/2021 年の単純比較	2
1.2	例: 可能な説明	2
1.3	例: 理論予測	3
1.4	例: 他の属性比較	3
1.5	例: 他の属性の変化	3
1.6	例: Size を調整	4
1.7	実例. 既存店ベースの比較	4
1.8	年齢構造のバランス: 経済成長	4
1.9	年齢構造のバランス: 出生率	5
1.10	まとめ	5
1.11	限界	5
2	単純な例: データ上でのバランス	6
2.1	目標	6
2.2	準備: 繰り返し期待値の法則	6
2.3	正式な定義: 平均値の分解	6
2.4	例	6
2.5	含意	7
2.6	バランス後の平均	7
2.7	Target となる割合	7
2.8	例	7
2.9	例: Balanced Mean	8
2.10	仮定: Overlap	8
2.11	まとめ	8
3	OLS による実装	8

3.1	例: ダミー化	9
3.2	例: OLS	9
3.3	例: 実例	10
	Reference	10

1 分析例

- 差が生じている”理由”の(部分的)説明に挑戦
- 広義には What if 分析 (もし~ならば、どうなるか?) の一部
 - 近年、機械学習の活用が注目されている

1.1 例: 2023/2021 年の単純比較



- 2023 年に比べて、2023 年の平均取引価格は上昇している
 - なぜ上昇したのか?

1.2 例: 可能な説明

- 経済学理論は”差”についての、豊富な潜在的説明を提供する
- 平均取引価格が上昇したとしても、“市場が好調”とは言えない

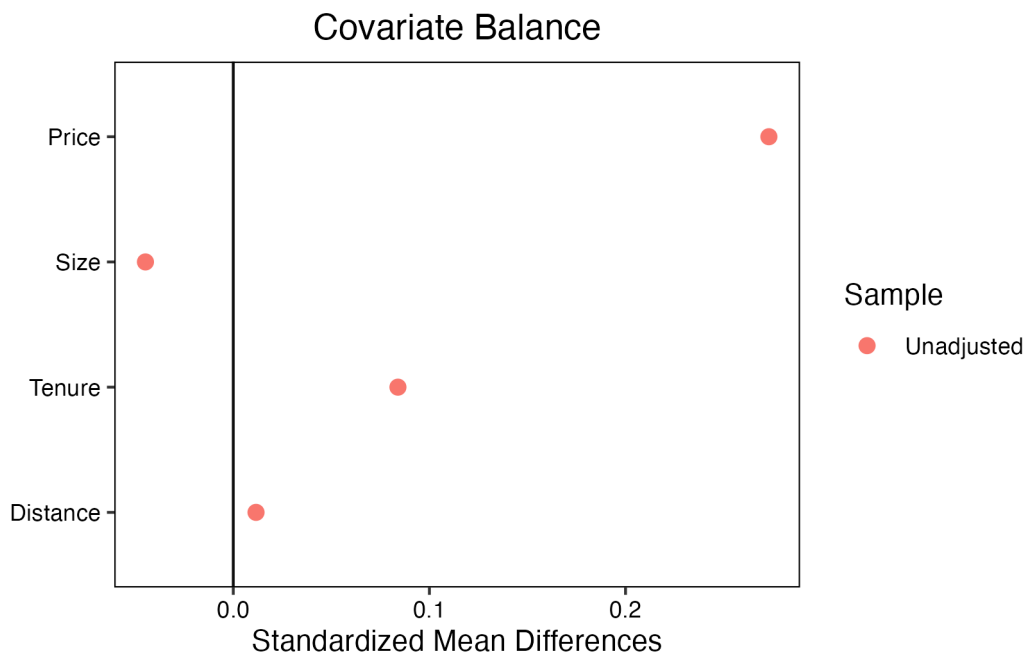
- 不動産への需要上昇だけでなく、供給減少によっても取引価格は上昇する
- 平均価格も取引数量も上昇するのであれば需要要因、取引数量が減少するのであれば供給要因

* 2023 年の取引数量は 3038, 2021 年は 2539

1.3 例: 理論予測

- 需要減少によって、“条件の悪い物件”の取引が大きく減少したかもしれない
 - 限られた条件のいい物件のみ取引されているのであれば、平均取引価格は上昇する

1.4 例: 他の属性比較

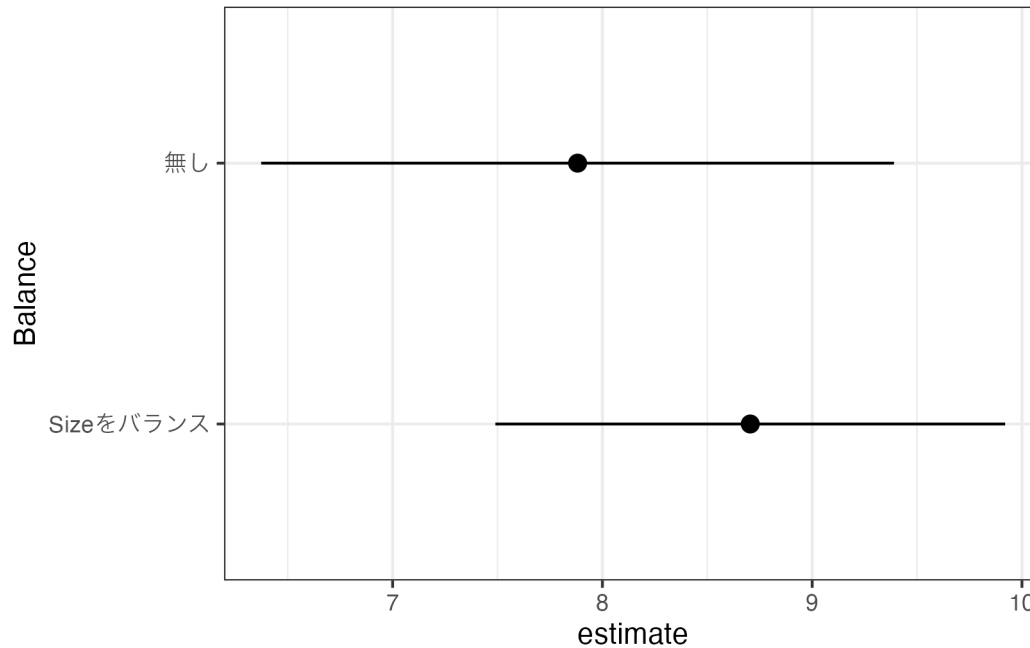


1.5 例: 他の属性の変化

- 実際に取り引された物件の情報のみを集計
 - 価格のみならず、取引される物件の特徴も変化
- 価格の以外の属性 (部屋の広さ、築年数、駅からの距離) も変化している
- より狭く、築年が古い物件が取引される傾向
 - 平均取引価格の変化の”一部は”、取引物件の性質の違いから生じているのではないか?

1.6 例: Size を調整

- Size の変化を調整し、バランスさせて比較する



1.7 実例. 既存店ベースの比較

- あるコンビニチェーンで、店舗あたりの平均売り上げが 1000 万円増大した

$$E[Y|D=1] - E[Y|D=0] = 1000\text{万円}$$

- 去年から今年にかけて、新規出店が大きく増加した
 - 売り上げが大きくなる傾向のある新規店の割合が大きく、結果、平均売り上げが増大したのではないかな?
- 既存店割合をバランスさせるために、既存店に絞って、比較する
 - セブン&アイ

1.8 年齢構造のバランス: 経済成長

- 一人当たり GDP(付加価値) の成長率は異なる
 - 国によって、(付加価値を主に生み出すと想定される) 生産年齢人口 (15-65 歳) の割合は異なる

- GDP 全体/人口と GDP 全体/生産年齢の成長率は異なる (Fernández-Villaverde, Ventura, and Yao 2023)
 - 後者では、日本の成長率が先進国で最も高い
 - * “労働者が生み出す付加価値” の成長率は、日本が最も高い (?)
 - ・ 議論が続く主張だが、少なくともバランスの重要性を示唆

1.9 年齢構造のバランス: 出生率

- 合計特殊出生率
 - 出生率の国比較や時系列比較に理由される
 - * 年齢構造も異なる
- 合計特殊出生率

$$= \frac{15\text{歳の女性が一年間で生んだ子供の数}}{15\text{歳の女性人口}} + \dots + \frac{49\text{歳の女性が一年間で生んだ子供の数}}{49\text{歳の女性人口}}$$

1.10 まとめ

- 既存店ベース: 今年と去年で、既存店の割合が異なる
 - 既存店比率を 100% に揃える
- 経済成長: 国によって、生産年齢人口が異なる
 - 生産年齢人口割合を 100% に揃える (+ 批判の残る仮定)
- 出生率: 国によって、年齢構造が異なる
 - “すべての年齢が同じ割合” に揃える

1.11 限界

- 実務でよく使われてきた手法は、バランスさせたい X の種類が少ない (既存店か否か、年齢)
 - 事例数が十分あれば、バランスできる
- X の組み合わせが大量にある場合 (部屋の広さ、駅からの距離、築年数、立地)、バランスさせることが難しくなる
 - 近年では、機械学習の活用が注目を集める ([EconML](#))

2 単純な例: データ上でのバランス

2.1 目標

- 取引年 ($= D$) 間で、Size ($= X$) を Balance させた後の、価格 ($= Y$) の平均差を算出
 - “よく似た物件” の平均取引価格の変化
 - * Balanced Comparison (バランス後の比較) と呼ばれる
- Point: Balance とは何か?

2.2 準備: 繰り返し期待値の法則

- 繰り返し期待値の法則

$$\begin{aligned} & \text{武蔵大学の平均身長} \\ &= (\text{武蔵大学の})\text{経済学部生の平均身長} \\ & \quad \times \text{経済学部生の割合} \\ & \quad + \text{社会学部生の平均身長} \\ & \quad \times \text{社会学部生の割合} + \dots \end{aligned}$$

2.3 正式な定義: 平均値の分解

- 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

$$\begin{aligned} & d\text{年における平均取引価格} \\ &= (30\text{平米} \ \& \ d\text{年})\text{における平均取引価格値} \\ & \quad \times d\text{年における30平米の物件割合} \\ & \quad + (35\text{平米} \ \& \ d\text{年})\text{における平均取引価格値} \\ & \quad \times d\text{年における35平米の物件割合} + \dots \end{aligned}$$

2.4 例

Y の平均値	D	Size	N	Size の割合
62.7	0	75	301	0.796
71.7	1	75	414	0.838
84.0	0	90	77	0.204
101.9	1	90	80	0.162

- 2023 年 ($D = 1$) の平均取引価格 $= 0.838 \times 71.7 + 0.162 \times 101.9$

2.5 含意

- D 間での格差を生み出す要因は 2 種類に分解できる
 - Y の平均値 の違い (X 内での違い)
 - X の割合 の違い (X の違い)
- バランス後の比較: 後者を排除

2.6 バランス後の平均

- “ X の格差” を排除した平均値

$$\begin{aligned}
 &= (30\text{平米} \ \& \ d\text{年})\text{における平均取引価格値} \\
 &\quad \times \underbrace{30\text{平米の物件割合の調整目標}}_{Target} \\
 &+ (35\text{平米} \ \& \ d\text{年})\text{における平均取引価格値} \\
 &\quad \times \underbrace{35\text{平米の物件割合の調整目標}}_{Target} + ..
 \end{aligned}$$

2.7 Target となる割合

- Target となる割合は D 間で共通
 - 研究者が設定する必要がある
- 代表例として
 - データ全体での Size の割合
 - 一定比率 (合計特殊出生率と同じ)

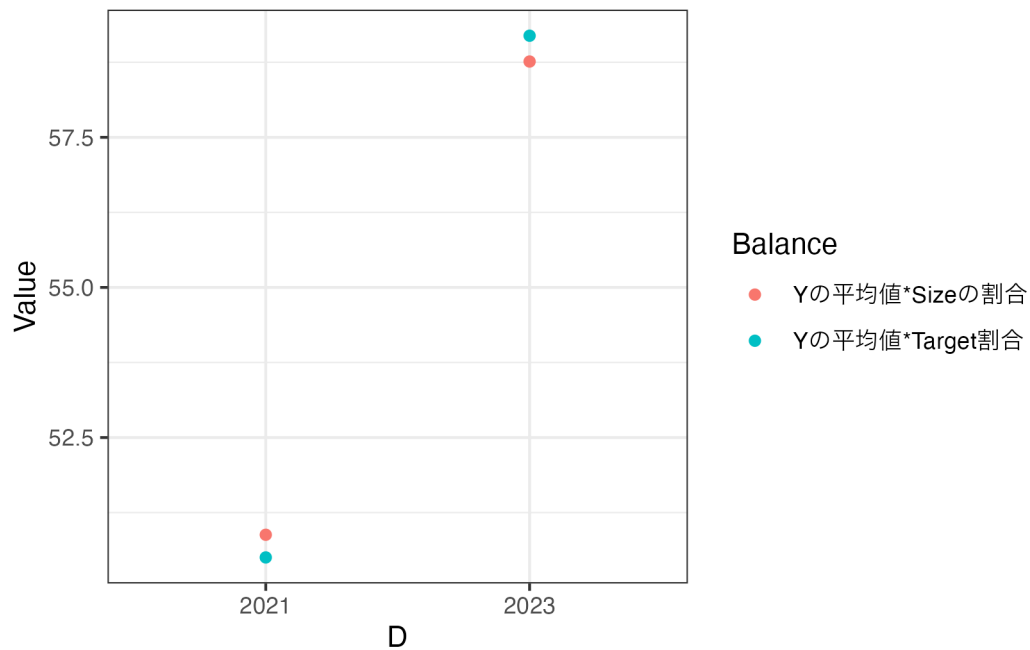
2.8 例

- データ全体での $Size = 75$ の割合 $(301 + 414)/(301 + 414 + 77 + 80) = 0.82$
- $Size = 90$ の割合 $(77 + 80)/(301 + 414 + 77 + 80) = 0.18$

Y の平均値	D	Size	N	Target	Size の割合
62.7	0	75	301	0.82	0.796
71.7	1	75	414	0.82	0.838

84.0	0	90	77	0.18	0.204
101.9	1	90	80	0.18	0.162

2.9 例: Balanced Mean



2.10 仮定: Overlap

- バランス後の比較が可能な前提
- “すべての X の組み合わせについて、 $D = 1$ となる事例も $D = 0$ となる事例も存在する”

2.11 まとめ

- Overlap の仮定が成り立てば、バランス後の比較を実行できる
- 多くの実践で、 X の組み合わせに対して、事例数が不十分であり、近似的なバランスしかできない

3 OLS による実装

- Balanced Comparison は、OLS でも実装できる
- X のすべての組み合わせについて、ダミー変数か (0/1 変換)

$$Y \sim D + \underbrace{factor(X)}_{\text{すべての } X \text{ を } 0/1 \text{ 変数に変形する}}$$

3.1 例: ダミー化

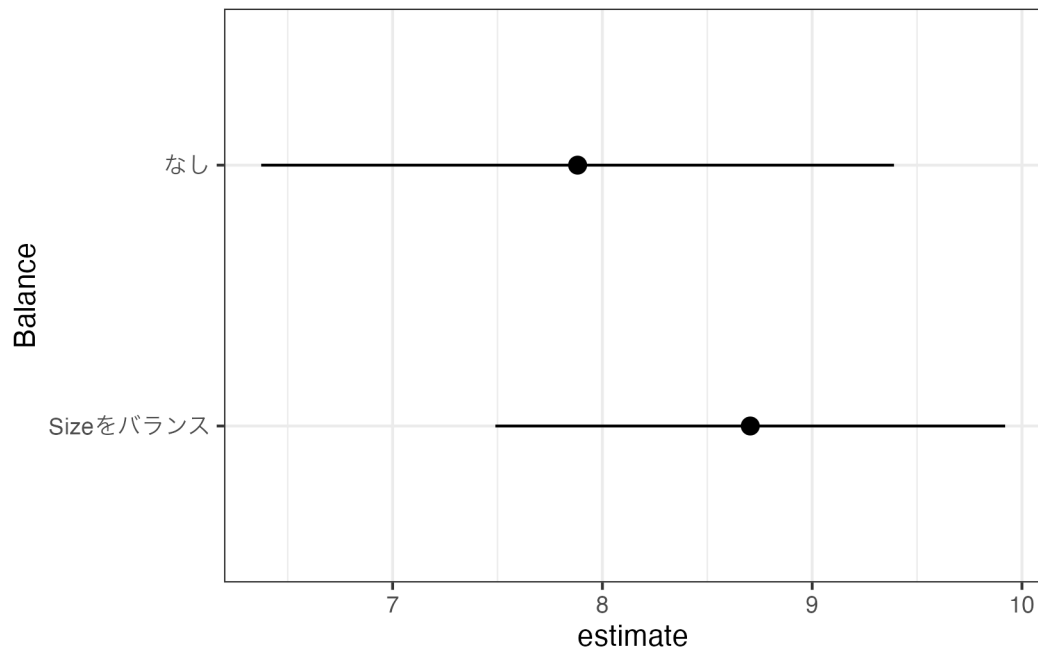
Size	Size_X40	Size_X50	Size_X60
40	1	0	0
50	0	1	0
60	0	0	1
60	0	0	1
40	1	0	0
60	0	0	1

3.2 例: OLS

平均価格	D	Size	データ上の割合	Target
35.70	0	40	0.252	0.253
39.06	1	40	0.253	0.252
37.82	0	50	0.220	0.209
47.85	1	50	0.196	0.209
52.18	0	60	0.528	0.539
59.13	1	60	0.551	0.538

- Target は”自動的に”選ばれる (後で議論)

3.3 例: 実例



- 「母集団上でバランス後の比較結果」を 95% の確率で含む信頼区間も計算できる

Reference

Fernández-Villaverde, Jesús, Gustavo Ventura, and Wen Yao. 2023. "The Wealth of Working Nations." National Bureau of Economic Research.