

格差/因果/比較分析のためのデータ分析 (ver 0.2.1)

Balancing weights を軸とした手法整理と機械学習の活用

川田恵介

2025-02-24

Table of contents

Preface	5
第 1 章 バランス後の比較	7
1.1 不動産市場の年次比較	7
1.2 バランス後の平均差	10
1.2.1 Overlap の仮定	11
1.3 応用事例	12
1.3.1 既存店ベースの比較	12
1.3.2 合計特殊出生率の時点/地域間比較	12
1.3.3 物価指数	13
1.3.4 因果効果	13
1.3.5 格差分析	14
1.4 応用上の課題	14
第 2 章 Weight	17
2.1 データ上の平均値	17
2.2 ターゲット上の平均値	18
2.2.1 バランス後の平均との類似性	19
2.3 加重平均値	19
第 3 章 Balancing Weight	21
3.1 Balancing Weight の定義	21
3.2 応用上の課題	22
第 4 章 OLS による特徴のバランス	23
4.1 OLS による平均値のバランス	23
4.1.1 例	24
4.2 OLS による分散や共分散のバランス	25
4.3 母集団の推定方法	26
4.4 実践への示唆	26
4.5 実践上の課題	27

4.5.1	定式化	27
4.5.2	ターゲット割合	28
4.5.3	負のウェイト	28
4.6	R による実践例	29
4.6.1	準備	30
4.6.2	OLS によるバランス	30
第 5 章	直接的な Balancing Weight の算出	35
5.1	算出方法	35
5.2	R による実践例	36
5.2.1	準備	36
5.2.2	Entropy Weight によるバランス	38
5.3	発展	41
第 6 章	機械学習の活用: 残差回帰	43
6.1	OLS の一般化	43
6.2	残差回帰	44
6.2.1	OLS の再解釈	44
6.2.2	教師付き学習の応用	45
6.3	母集団の推定方法	46
6.4	応用上の課題	46
6.5	R による実践例	47
6.5.1	準備	47
6.5.2	残差回帰	48
第 7 章	機械学習の活用: AIPW	51
7.1	Inverse probability weight	51
7.2	Augmented Inverse Probability Weight	52
7.3	母集団の推定方法	53
7.4	残差回帰との比較	53
7.5	R による実践例	53
7.5.1	準備	54
7.5.2	ATE	54
	Reference	57

Preface

定量的な比較分析の方法を、R での実装とともに紹介します。

比較分析は、社会における「グループ間の違い」を明らかにすることを目標とします。本ノートでは、データ上のある変数 D 間での、別の変数 Y の平均値の差を推定する方法を紹介します。例えば性別 (D) 間での賃金 (Y) の平均格差を推定します。

比較分析は、社会/市場を理解するための方法として、中核的な位置を占めています。通常、これらの研究関心となる D については、 Y 以外の違いがあることが一般的です。研究課題に応じて、これらの違いの一部を、データ分析上の処理として「解消 (バランス)」することが必要です。

例えば、性別 (D) 間で見られる職業経験の違い X をバランスさせることで、職業経験以外の要因によって生じる男女間の賃金 (Y) 格差が推定できます (Vafa, Athey, and Blei 2024)。あるいは、ある職業訓練プログラム (D) が就業確率や就業後の賃金 (Y) に与える因果的効果を推定を試みます。このような研究課題では、異なる職業訓練プログラムへの参加者間で、背景属性をバランスさせた上での、就業状態や賃金の比較が求められます (Behaghel, Crépon, and Gurgand 2014; Kallus 2023)。

本ノートの中心的なコンセプトは、Balancing Weight です。Balancing Weight は、以下のアプローチを統合的整理できる極めて有益な概念です。

- 重回帰 (OLS) や Penalized Regression による調整 (Chattopadhyay and Zubizarreta 2023; Bruns-Smith et al. 2023)
- 傾向スコア (Propensity Score) の活用 (Imai and Ratkovic 2014)
- 機械学習などを用いた Debiased Machine Learning (Chernozhukov et al. 2018; Chernozhukov, Escanciano, et al. 2022)
- Entropy Weight (Hainmueller 2012) や Stable Weight (Zubizarreta 2015)、Energy Weight (Huling and Mak 2024) による調整
- 機械学習などを用いた Auto Debiased Machine Learning/Augmented Balancing Weights (Chernozhukov, Newey, and Singh 2022a, 2022b; Chernozhukov, Newey,

et al. 2022; Bruns-Smith et al. 2023)

簡易な入門としては [Chattopadhyay and Zubizarreta \(2024\)](#)、詳細な入門としては [Chattopadhyay, Hase, and Zubizarreta \(2020\)](#)、[Ben-Michael et al. \(2021\)](#) などを参考にしてください。

本ノートの構成は以下のとおりです。

- Chapter 1: 本ノートの Estimand である「バランス後の比較」を定義します。
- Chapter 2: 本ノートの中核概念である Balancing Weight を紹介する準備として、荷重 (Weight) を定義します。
- Chapter 3: Balancing Weight を定義します。また Balancing Weight の直感的な算出方法が、利用できない状況が多いことを指摘します。
- Chapter 4: より幅広い状況で算出できる近似的な Balancing Weight を紹介します。また標準的な OLS 推定が、暗黙のうちに近似的な Balancing Weight を算出した、バランス後の比較と解釈できることを示します。
- Chapter 5: 近似的な Balancing Weight を、明示的な最適化問題として算出する方法を紹介します。代表的な方法の一つである Entropy weight ([Hainmueller 2012](#)) は、その計算効率や分析の透明性の高さから、幅広く用いられています。
- Chapter 6: 機械学習を用いた、よりデータ主導のアプローチを紹介します。残差回帰に機械学習を補助的に用いることで、近似的な Balancing Weight を算出した、バランス後の比較を行うことができます。OLS とは異なり、事例数が十分大きければ、「母集団上で、完全な Balance を達成した後の比較分析」、の優れた推定値と見做せることを紹介します。
- Chapter 7: 機械学習を用いた Balanced comparison の代替的な方法である、Augmented inverse probability weighting への活用を紹介します。残差回帰よりも Estimand の解釈が容易な一方で、 X の分断が激しいケースでは推定結果が不安定になりやすいという問題点があります。

第 1 章

バランス後の比較

典型的なバランス後の比較 (Balanced Comparison) 分析では、グループ D の間で、 X についての差を解消した後に、 Y についての平均差を推定します^{*1}。このような比較は、格差分析や因果効果の肝となります。

まず実例から紹介します。

1.1 不動産市場の年次比較

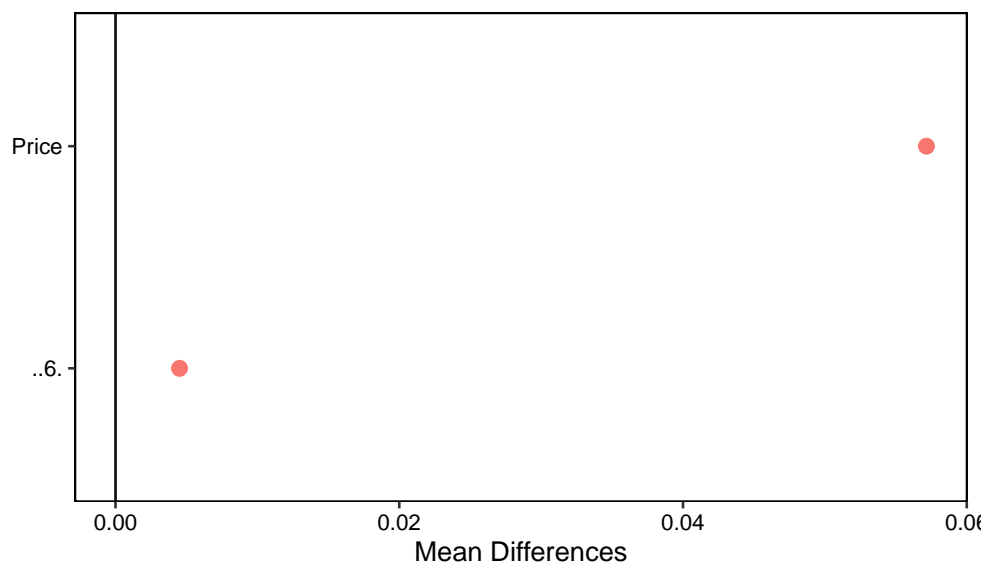
東京 23 区の中古マンション市場において、2022 年と 2021 年の取引価格と立地 (中心 6 区 (港区、中央区、文京区、千代田区、渋谷区、新宿区)、かそれ以外か) について、平均的な差を図示します。

^{*1} Gelman, Hullman, and Kennedy (2024)

Example

平均価格 (100 万円)	中心 6 区	取引年	事例割合
37.748	0	2021	0.784
60.474	1	2021	0.216
39.150	0	2022	0.779
64.814	1	2022	0.221

Covariate Balance



2022 年の平均取引価格は、2021 年に比べて上昇しており、不動産市場における価格上昇が続いているように見えます。しかしながら、同時に中心 6 区の物件割合も増加しています。一般に中心 6 区に立地する物件の方が、高い取引価格が予想されます。このため物件の立地の変化によって、取引価格の上昇が”底上げ”されている可能性があります。

もし中心 6 区の物件割合が不変であった場合、平均取引価格にどのような差が残るでしょうか？ このような問いに対して、バランス後の比較分析は回答できます。

以下では実際のデータを用いて、取引年と立地 (中心 6 区であれば 1、それ以外であれば 0) ごとに、平均取引価格を示しています。また各取引年における取引事例の立地割合も算出しています。

繰り返し期待値の法則を用いると 2022 年と 2021 年の平均取引価格の差を算出できます。

！ 繰り返し期待値の法則

- 変数 Y, D と他の変数 X について、 Y の平均値 は以下のように書き換えられる。

$$\begin{aligned}
 & (D = d) \text{ における } Y \text{ の平均値} \\
 &= (D = d, X = x_1) \text{ を満たす事例の } Y \text{ の平均値} \\
 & \quad \times (D = d, X = x_1) \text{ を満たす事例の割合} \\
 & \quad + \dots \\
 & \quad + (D = d, X = x_L) \text{ を満たす事例の } Y \text{ の平均値} \\
 & \quad \times (D = d, X = x_L) \text{ を満たす事例の割合}
 \end{aligned}$$

- ただし X がとりえる値は、 x_1, \dots, x_L とする。

Example に適用すると、2022 年の平均取引価格は以下のように計算できます。

$$\begin{aligned}
 & \text{2022年の平均価格} \\
 &= \underbrace{64.814}_{\text{中心6区における平均価格}} \times \underbrace{0.221}_{\text{中心6区に立地する事例の割合}} \\
 &+ \underbrace{39.150}_{\text{中心6区以外における平均価格}} \times \underbrace{0.779}_{\text{中心6区以外に立地する事例の割合}} \\
 &= 44.810
 \end{aligned}$$

2021 年の平均取引価格も同様に計算できます。

$$\text{2021年の平均価格} = 60.474 \times 0.216 + 37.748 \times 0.784 = 42.658$$

2022 年と 2021 の平均差を計算すると、2022 年の平均取引価格の方が、2.153 ほど高くなっていることが確認できます。

繰り返し期待値の法則自体は、シンプルな計算ルールですが、「 X のバランス」の意義を明確にします。繰り返し期待値の法則から、平均価格の違いは、同じ立地内での平均取引価格の違いと立地の割合の違いによって生じることがわかります。2021 年と比べると、

平均価格	中心6区	取引年	事例割合	ターゲットとなる割合
37.748	0	2021	0.784	0.750
60.474	1	2021	0.216	0.250
39.150	0	2022	0.779	0.750
64.814	1	2022	0.221	0.250

2022年に取引された物件の中で、中心6区の物件割合が21.6%から22.1%に上昇しています。一般に中心6区の方が、物件の価格が高い傾向にあります。このため、中心6区の物件割合の違いが、平均取引価格上昇の一因となっている可能性があります。

バランス後の比較分析では、「取引されている物件に占める中心6区の割合が、変化しなかった場合」の平均取引価格の変化を推定します。より一般的には、「 X についての差を解消した状態で」 D 間で Y を比較します。

1.2 バランス後の平均差

本ノートにおいて、バランス後の比較分析の一つである「バランス後の平均差」を推定します。

！ バランス後の平均値

$(D = d)$ における Y の平均値

$= (D = d, X = x_1)$ を満たす事例の Y の平均値
 $\times (X = x_1)$ についてのターゲットとなる割合

+..

$+(D = d, X = x_L)$ を満たす事例の Y の平均値
 $\times (X = x_L)$ についてのターゲットとなる割合

- ターゲットとなる割合は、 D の値にかかわらず同じ値として、研究者が定める。

バランス後の平均値の計算例として、以下ではターゲットを取引年にかかわらず、「中心6区が0.25、その他が0.75」と設定し、Exampleに適用します。

ターゲットのもとでは、取引年にかかわらず、中心6区とそれ以外に立地する事例の割合

は一定となります。このため平均価格の変化を生み出す要因から、事例割合の変化を排除できます。

事例割合をターゲットに差し替えて、繰り返し期待値の法則を適用すると、バランス後の平均価格は以下のように計算できます。

2022年のバランス後の平均価格

$$= 64.814 \times \underbrace{0.25}_{\text{中心6区についてのターゲット}} + 39.150 \times \underbrace{0.75}_{\text{中心6区以外についてのターゲット}} = 45.566$$

$$\text{2021年のバランス後の平均価格} = 60.474 \times 0.25 + 37.748 \times 0.75 = 43.4295$$

バランス後の平均差は、2.1365 であり、バランス前 (2.153) に比べて縮小しました。

1.2.1 Overlap の仮定

バランス後の平均差が計算できる前提条件は、Overlap の仮定 (Positivity の仮定とも呼ばれます) が成り立っていることです。

! Important 1: Overlap の仮定

- ターゲットとなる割合が 0 よりも大きい X の組み合わせについて、 $D = 1$ の事例も $D = 0$ の事例も、**両方存在する**：

$$1 > \Pr[D = d|X] > 0$$

ただし $\Pr[D = d|X]$ は $(X = x)$ 内での $D = d$ の割合 ($D = 1$ の割合)

Overlap が成り立っていない場合、全ての D について平均値が計算できず、厳密なバランス後の比較は**根本的に**不可能です。

例えば、教育経験 ($= X$) をバランスさせた男女間 ($= D$) での賃金格差を推定したいとします。もしデータにおける男女間での教育経験の分断が極めて大きい場合、大学卒以上の女性は存在しない可能性があります。この場合、 $X = \text{大学卒}$ の女性割合は 0 であり、大学卒内で男女間比較は不可能です。このため大学卒について、ターゲットとなる割合を”0”としない限り、バランス後の比較は不可能です。

1.3 応用事例

バランス後の比較分析の応用例は、大量に存在します。以下では実務においてよく利用されている指標を紹介します。

1.3.1 既存店ベースの比較

バランス後の比較は、企業の経営戦略を考える上でも用いられます。

小売や飲食/宿泊業などでは、しばしば既存店に絞った上での、売上比較がなされます。例えば、あるコンビニチェーンで、店舗あたりの平均売り上げが1000万円増大したとします。同時に去年から今年にかけて、新規出店も大きく増加したとします。新規店の方が売上が高くなる傾向がある場合、新規店割合の違いが、平均売上の上昇をもたらした可能性があります。

既存店割合をバランスさせるシンプルな方法として、既存店だけに絞った平均売上を比較がよく行われます。このような分析では、既存店比率は全ての年について100%となり、完全なバランスが達成されます。

既存店ベースの比較におけるターゲット割合は、新規店については1、新規店以外については0となります。

1.3.2 合計特殊出生率の時点/地域間比較

バランス後の比較分析は、社会/政策分析においても幅広く利用されています。代表例は、合計特殊出生率の国家間/時代間比較です。

出生数の動向を把握する上で、新生児数を年次や国家間比較は、有益だとみなされてきました。合計特殊出生率は、成人の年齢構造の違いをバランスさせるために利用されている指標です。単純な出生率（一年間に生まれた子供の数/女性の数）は、成人の年齢構造の影響を強く受ける可能性があります。社会において高齢者の比率が高まれば、出生率は低下することが予想されるからです。対して合計特殊出生率は、「仮に年齢構造が同じであった場合」の出生率を、以下の方法で推定しています

$$\frac{15\text{歳の女性が産んだ子供の数}}{15\text{歳の女性の数}} + \dots + \frac{49\text{歳の女性が産んだ子供の数}}{49\text{歳の女性の数}}$$

合計特殊出生率の計算においては、15歳から49歳の女性の年齢（= X）層が同じ割合になることを目指しています。すなわちターゲット割合は、15歳から49歳までについて1/34、それ以外の層については“0”となります。

1.3.3 物価指数

より複雑な比較が要求される指標の代表例は、物価指数です。物価指数は、“同じような消費生活”を送るために必要な支出額を、比較するために算出されます ([wiki](#))。この際に問題となるのは、実際に消費する商品の組み合わせや量は、年毎に変化することです。すなわち D = 時点、 Y = 名目価格、 X = “品目” とし、消費する”品目”をバランスさせる必要があります。

Bajari et al. (2023) は、機械学習の手法と商品の画像やテキストデータも用いることで、“質”も含めてバランスする試みを行なっています。このような分析によって、“ステレス値上げ”や“商品の質的改善”にも対応できる指標が作成できる可能性があります。

1.3.4 因果効果

バランス後の比較は、因果効果を推定する際にも、中心的な役割を果たします。因果効果を記述する枠組みは、多数提案されています (Potential outcome/Structural Equation Model (DAG)/ 経済モデルなど)。これらは共通して、[思考実験](#)、(特に[反実仮想実験](#)) の重要性を指摘しています。

ここでは反実仮想を明らかにできる仮想的な実験 (Target trial) を想像します。典型的には、「無限大の被験者を用意し、被験者間の相互作用がない環境で、ランダムに選ばれた一部の被験者に介入を行う」実験です。このような実験結果をもし得ることができれば、介入を行った/行わなかった被験者を比較することで、介入の因果効果を明らかにできます。介入は完全ランダムに決まっているため、介入を行った/行わなかった集団間で、背景属性について大きな差が生じないことを担保できます^{*2}。

現実には、このような大規模な実験を行うことは不可能です。このため現実のデータから、理想的な実験結果を可能な限り再現することを目指します。

Egami et al. (2024) は、TV ゲームが精神的健康状態に与える因果効果について、実証的な証拠を提供しています。同論文では、コロナ下で生じたゲーム機への過大需要と、それに対応するために実施された「ゲーム機購入権くじ」を利用しています。同時期においては、ゲーム機購入希望者の中で、くじに当選した人だけゲームを購入することができました。このため理想的な実験に近い状況が生じています。

しかしながら、理想的な実験とは異なり、背景属性に幾つかの違いが生じる可能性があります。例えば地域によってくじの当選確率が異なるのであれば、ゲーム機購入者と非購入者の間で、居住地に偏りが生じてしまいます。

^{*2} 完全ランダムかつ無限大の被験者が存在すれば、格差は完全に消失します。現実の被験者数のもとでは、背景属性について偏りは生じますが、信頼区間などの統計的な手法でその影響を評価できます。

以上に対応するために、 $D =$ くじに当選、 $Y =$ 健康状態、 $X =$ 居住地等とし、当選/非当選者間で背景属性をバランスさせた推定を行なっています。

1.3.5 格差分析

格差分析においても、バランス後の比較は重要です。格差分析の出発点は、通常格差/差別/不平等の大きさの推定です。このためには、研究対象とする格差を明確に定義する必要があります。

因果効果と同様に、格差の定義についても、多様な議論があります (ヒュームの法則; Jackson and VanderWeele (2018); Rose (2023); Mossé et al. (2025))。代表的なものとして、格差や差別を因果的に定義するのは難しく、あくまで規範的に定義する立場があります。すなわち「本来であれば存在すべきではない差」を格差として定義します。

例えば、両親の教育経験間での、子供の所得格差に関心があるとします。両親の属性に基づく差は、機会の不平等の代表的な例であり、多くの研究が行われてきました。

実証上の論点は、両親の教育経験が異なれば、子供の年齢の分布も異なる可能性への対処です。日本のような教育年数が急激に伸びた国においては、教育年数の長い両親の子供は、年齢が若い可能性があります。また年功的賃金体系が残存していることを踏まえれば、年齢が若ければ賃金が低い傾向があります。このため単純比較を行い、「教育年数の長い両親の子供の方が、短い両親の子供よりも所得が低い」、という結果を得たとしても、この差は年齢分布の違いを反映している可能性があります。

もし研究関心が、「同じ世代なのに両親の教育経験によって生じる格差」にあるのであれば、子供の生まれ年を X としてバランスする必要があります。

またバランスは格差の要因分析にも応用できます。このような分析は、分解分析 (Decomposition analysis) と呼ばれます (Opacic, Wei, and Zhou 2023)。Vafa, Athey, and Blei (2024) では、男女間賃金格差の要因分析として、 $X =$ 職歴、 $D =$ 性別、 $Y =$ 賃金、とするバランス後の比較を行なっています。職歴のバランス後にも、賃金格差が残存しているのであれば、採用時や個別賃金交渉における男女間での異なる取り扱いが男女間格差の要因であると解釈できます。

1.4 応用上の課題

バランス後の比較は、シンプルな枠組みです。上記の例の通り、大規模なデータを用いて、少ない X をバランスさせるのであれば、単純な計算でバランスできます。

しかしながら、多くの応用では複数の X を同時にバランスさせることが求められます。このような応用では、同じ属性を持つ事例数が少なくなり、実際の計算は困難になります。

多くの応用で活用できる、より実践的な推定方法が求められます。このような推定手法は、Balancing Weight (Chapter 3) を実質的に推定していると解釈できます。Chapter 2 では Balancing Weight を理解するための準備として、一般的な Weight を紹介します。

第 2 章

Weight

Balancing Weights を導入する準備として、より一般的な概念である Weights を紹介します。Weights は、データ上の事例の分布を、統計的な処理によって、変化させるために用いられます。

以下、サンプリングの偏りへの対応を例とします。前章のバランス後の比較分析と議論の多くが類似している点に注意してください。

2.1 データ上の平均値

今、ある「不動産研究所」が調査員を千代田区、文京区、板橋区に派遣し、中古マンションの取引事例を収集したとします。各調査員は、全く同じ数の事例を収集します。

以下では、立地 (District) ごとに、平均取引価格とデータ全体に対する事例の割合をまとめています。全ての区について、同数の事例が収集されていることに注意してください。

繰り返し期待値の法則を用いると、この情報のみからデータ全体の平均取引価格は計算できます。

- 平均取引価格 は

Example		
事例の割合	平均価格	District
0.333	66.5	千代田区
0.333	47.3	文京区
0.333	29.5	板橋区

$$\begin{aligned}
&= \underbrace{66.5}_{\text{千代田区の事例の平均取引価格}} \times \underbrace{0.333}_{\text{千代田区の事例の割合}} \\
&+ \underbrace{47.3}_{\text{文京区の実例の平均取引価格}} \times \underbrace{0.333}_{\text{文京区の実例の割合}} \\
&+ \underbrace{29.5}_{\text{板橋区の実例の平均取引価格}} \times \underbrace{0.333}_{\text{板橋区の実例の割合}} \\
&= 47.8
\end{aligned}$$

繰り返し期待値の法則から、データ上の平均値は、 X についてのサブグループ内での平均値とサブグループの割合の掛け算の総和となります。このため、もしサブグループの割合が研究関心から乖離している場合、サブグループ内での平均値が妥当な値であったとしても、ミスリードな平均値が計算されてしまいます。

2.2 ターゲット上の平均値

今研究関心は、「もし**実際の取引履歴をすべて収集したデータ**を用いて計算された平均取引価格」、であるとしします。このような平均値を計算したい仮想的な割合を、バランス後の比較分析と同様に、**ターゲット割合**と呼びます。

今、ターゲット割合は、千代田区が 0.161、文京区が 0.33、板橋区が 0.509 であることが判明しているとしします。もしデータ上の各区の取引割合を、ターゲットの取引割合と一致させた場合、平均値はどのように変化するでしょうか？

Example における調整された平均取引価格は、以下のように算出されます。

$$\begin{aligned}
&= \underbrace{66.5}_{\text{千代田区の事例の平均取引価格}} \times \underbrace{0.161}_{\text{千代田区の事例の割合}} \\
&+ \underbrace{47.3}_{\text{文京区の実例の平均取引価格}} \times \underbrace{0.33}_{\text{文京区の実例の割合}} \\
&+ \underbrace{29.5}_{\text{板橋区の実例の平均取引価格}} \times \underbrace{0.509}_{\text{板橋区の実例の割合}} \\
&= 41.3
\end{aligned}$$

データ上の平均値は 47.8 であったので、過大であったことがわかります。これは、平均取引価格が高い傾向にある千代田区の物件割合が、現実の取引割合 (0.161) よりも、データ上の割合 (0.333) が過大であることに起因します。

2.2.1 バランス後の平均との類似性

以上の議論は、バランス後の平均値と本質的には同じものです。ターゲットとデータ上の分布が乖離しているため、平均値と調整された平均値は乖離しています。唯一の違いは、バランス後の平均値を定義する際には、ターゲットを D の値に依存しないように設定する必要があり、Overlap の仮定 Important 1 に注意する必要がある点のみです。

このため以下の加重平均を用いた調整された平均値の計算方法は、バランス後の平均値を求める際にも利用できます。

2.3 加重平均値

ターゲット割合へ調整を行うための有力な方法は、加重平均 (Weighted mean) を計算することです。一般に加重平均は以下のように定義されます。

! 加重平均値

- 調整された平均値

$$= \left(\underset{\text{Weight}}{\omega} \times Y \right) \text{の平均値}$$

- Weight に対して、 ω の平均値 = 1 を制約とする。

Weight は各事例の Y の値が、最終的な平均値に反映される度合いをコントロールします。例えば、もし $\omega = 0$ であれば、その事例は平均値の計算に一切反映されません。

Weight は、データとターゲットにおける X の分布を揃えるように設定されます。すなわち

$$\omega \times \text{データ上の } X \text{ の割合} = \text{ターゲットとなる } X \text{ の割合}$$

を達成するように ω を算出します。両辺を事例割合で割ると、

$$\omega = \frac{\text{ターゲットとなる割合}}{\text{データ上の割合}}$$

となります。すなわちターゲットよりも過大に収集されているグループは小さめに、ターゲットよりも過小なグループは大きめに反映させます。

平均価格	District	ターゲットとなる割合	事例の割合	Weight
66.5	千代田区	0.161	0.333	0.483
47.3	文京区	0.330	0.333	0.991
29.5	板橋区	0.509	0.333	1.529

Example に適用すると、以下となります。

加重平均値は以下のように算出できます。

$$\left[\underbrace{\underbrace{\underbrace{44}_{\text{取引価格}} \times \underbrace{0.483}_{\text{Weight}}}_{\text{千代田区的事例}} + 70.0 \times 0.483 + \dots}_{\text{千代田区的事例}} + \underbrace{75.0 \times 0.991 + \dots}_{\text{文京区的事例}} + \underbrace{59.0 \times 1.529 + \dots}_{\text{板橋区的事例}} \right] \text{の平均} = \underbrace{41.3}_{\text{加重平均}}$$

第 3 章

Balancing Weight

D 間での X の分布をバランスを達成する実用的な手法は、数多く提案されています (Chattopadhyay, Hase, and Zubizarreta 2020; Bruns-Smith et al. 2023)。このような手法を整理し、活用していくためには、**Balancing weight** という概念を導入することが有益です。

Balancing weight は、前章で導入した Weight の一種であり、 D 間での X の分布の乖離を調整するために用いられます。

3.1 Balancing Weight の定義

! Balancing Weight

- Balancing weight $\omega(x, d)$ は、 D 間での X の分布の乖離を調整するために導入され、以下のように定義する。

$$\begin{aligned} & (D = 1) \text{ における事例割合} \times \omega(x, 1) \\ &= (D = 0) \text{ における事例割合} \times \omega(x, 0) \\ &= \text{ターゲットとなる割合} \end{aligned}$$

- 定義式を変形すると

$$\omega(x, d) = \frac{\text{ターゲットとなる割合}}{(D = d) \text{ グループにおける割合}}$$

- ターゲットに比べて過大な事例割合が過大なグループに対しては小さい、過小なグループに対しては大きな Weight を付与する。

ターゲットは、原理的には研究者が指定する必要があります。代表的なものは、データ全

平均価格	中心6区	取引年	事例割合	ターゲット割合	Weight
37.748	0	2021	0.784	0.782	0.997
60.474	1	2021	0.216	0.218	1.010
39.150	0	2022	0.779	0.782	1.003
64.814	1	2022	0.221	0.218	0.989

体における X の分布です^{*1}。

先のデータに適用すると、以下のような Balancing Weight が計算されます。

2022 年と 2021 年を結合したデータ全体のうち、中心 6 区に立地する物件割合は 21.8%、それ以外が 78.2% であったので、ターゲットとして設定しています。

Balancing weights を用いると、バランス後の平均値は以下のように計算できます

2021年のバランス後の平均値 = 2021年の $(\omega(X_i, 2021) \times Y_i)$ の平均値

2022年のバランス後の平均値 = 2022年の $(\omega(X_i, 2022) \times Y_i)$ の平均値

3.2 応用上の課題

X の組み合わせの種類に比べて、十分な事例数が存在するのであれば、Balancing weight は、データ上での X の割合を用いて計算できます^{*2}。

しかしながら多くの応用研究では、バランスの対象となる X の数が多く、 $D = 1$ または $D = 0$ のどちらかの事例しか存在しない組み合わせが対象に発生します。このような組み合わせについては、Balancing weights を計算することができず、Balanced comparison が不可能となります。この問題を解決するために、次節以降で紹介する、OLS (Chapter 4) や機械学習 (Chapter 6) などを用いた「近似的なバランス法」の活用が有用です。

^{*1} 因果推論の文脈では、平均効果 (Average Treatment Effect) と呼ばれています。

^{*2} この方法は Exact Matching や Stratified Estimation (Wager 2024) として知られる方法による推定結果と完全に一致します。例えば Exact Matching は、[MatchIt package](#) (Stuart et al. 2011) などを利用して実装できます。

第4章

OLS による特徴のバランス

X の組み合わせが多く、Balancing weights を計算することが困難な場合、近似的なバランスが有力な方法となります。

本節では代表的な統計手法である OLS（重回帰）が、近似的な Balancing weights を暗黙のうちに計算する手法であることを紹介します。OLS は分布の特徴を、研究者が定める定式化に応じて、要求する Balance の精度を変更できることが利点です。ただし、 D 間での X の分布の分断が激しい場合、予期せぬ挙動を示しやすいことに注意が必要です。

4.1 OLS による平均値のバランス

近年の研究により、線型モデルの OLS 推定は、近似的な Balance を達成することが確認されています (Chattopadhyay and Zubizarreta 2023)。本ノートでは、 $D = \{0, 1\}$ を前提とし、その議論の骨子を紹介します。

! OLS の性質 [Chattopadhyay 2023 implied]

- $Y \sim D + X_1 + \dots + X_L$ を OLS で推定し算出される D の係数値は、以下の方法で計算される値と完全に一致する

1. すべての X_l について、以下を満たす $\omega(x, d)$ を探す。

$$(D = 1) \text{ における } (\omega(x, 1) \times X_l) \text{ の平均値}$$

$$= (D = 0) \text{ における } (\omega(x, 0) \times X_l) \text{ の平均値}$$

2. 1. を満たす $\omega(x, d)$ の中で、最小の分散を持つ $\omega(x, d)$ を Balancing Weights とする

3. β_D を以下のように計算する。

$$\beta_D = (D = 1) \text{ における } (\omega(x, 1) \times Y) \text{ の平均値}$$

$$-(D = 0) \text{ における } (\omega(x, 0) \times Y) \text{ の平均値}$$

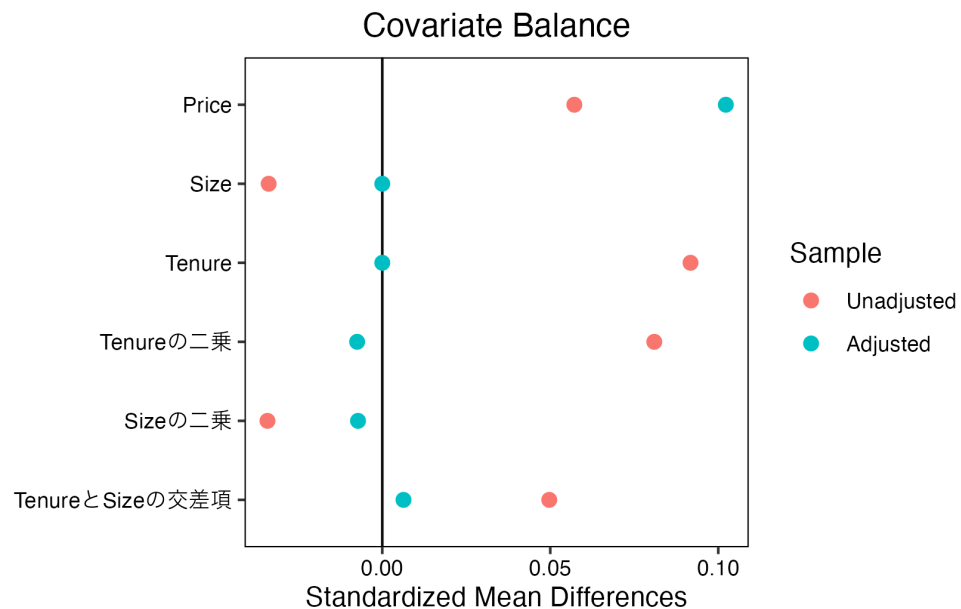
重回帰による推定は、 D 間で X の平均値をバランスさせた上で、平均値を比較しています。また「母集団における OLS の結果」の推論に悪影響を与える、Weight の分散も可能な限り削減しています。

4.1.1 例

部屋の広さ (Size) と 築年数 (Tenure) をバランスさせた後に、2022($D = 1$)/2021($D = 0$) 年の平均取引価格差を推定します。 $Price \sim D + Size + Tenure$ を OLS で推定します。

```
estimatr::lm_robust(
  Price ~ D + Size + Tenure,
  Data)
```

この推定によって得られる D の係数値は、以下のようなバランスを達成する Balancing Weights を用いた平均差と一致します。



赤点 (Unadjusted) は、バランス前の単純平均差を表します。価格が大きく上昇していますが、取引物件の部屋の広さは狭くなり、築年数は古くなっています。青点 (Adjusted)

は、OLS による暗黙のバランス後の差を示しています。Size や Tenure の平均値は完全にバランスしており、結果平均取引価格差も上昇しています。Tenure2 や Size2 は、築年数や部屋の広さの二乗項 (分散)、Tenure_Size は交差項 (共分散) を示しており、これらについてはバランスしていません。

4.2 OLS による分散や共分散のバランス

$Price \sim D + Size + Tenure$ を推定しても、Size や Tenure の平均値のみしかバランスできません。一見、これは OLS の致命的な弱点のように見えますが、簡単な修正によって解決できます

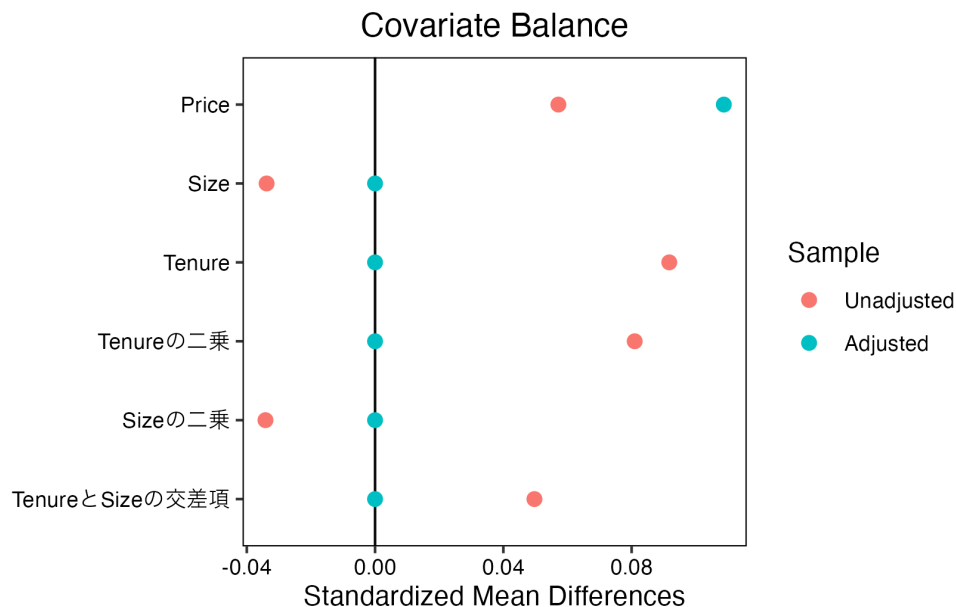
分散や共分散もバランスさせるためには、二乗項や交差項もモデルに導入したモデル

$$Price \sim D + Size + Tenure + Size^2 + Tenure^2 + Tenure \times Size$$

を OLS 推定します。

```
estimatr::lm_robust(
  Price ~ D + Size + Tenure +
    I(Size^2) + I(Tenure^2) +
    (Size + Tenure)**2,
  Data)
```

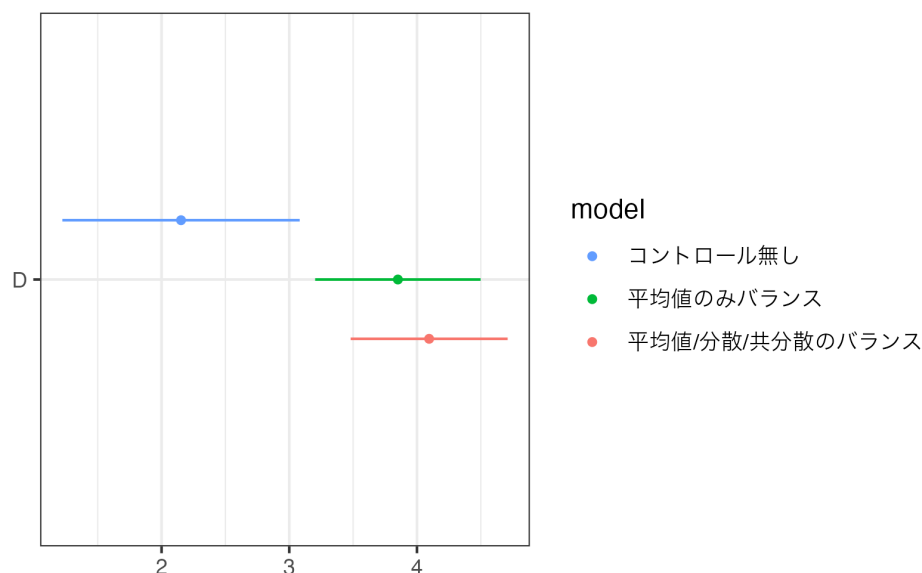
これによって、Size の二乗の”平均値”などもバランスさせることができます。これは各変数の分散や共分散をバランスを意味します。結果、以下の図の通り、分散や共分散も Balance します。



4.3 母集団の推定方法

推定するパラメタ (β_0, \dots, β_L) に比べて事例数が十分に大きければ、データ上の OLS の結果は、「母集団における OLS の結果」の優れた推定値です^{*1}。特に近似的に計算される信頼区間を用いれば、母集団における OLS の結果を、定量的に議論できます。

以下では一切の X をバランスさせないケース、Size と Tenure の平均値のみをバランスさせるケース、平均に加えて分散と共分散もバランスさせたケースの推定結果を、95% 信頼区間とともに比較しています。



コントロールをしないケースに比べて、Size や Tenure をバランスさせると、2022 年と 2021 年の平均差が大きくなりました。一方で本データについては、平均値に加えて分散や共分散をバランスさせたとしても、あまり大きな変化は生じませんでした。

また信頼区間に焦点を当てるとバランス後の平均差の方は、前に比べて、より「狭く」なっています。これは、母集団における OLS の結果とデータ上の OLS の結果が、大きく乖離する可能性が減少している (と推測できる) ことを反映しています。

4.4 実践への示唆

実践における課題は、 X について適切な複雑さを持つモデルを設定することです。しかしながら適切な複雑性は、母集団の特徴や事例数などに依存しており、一般的な解決は困難で

^{*1} 詳細に関心がある読者は、Chernozhukov et al. (2024) の 1 章などを参照ください。

す。ただし過去の実践の問題点、および実践上の示唆を得ることはできます。

多くの研究で X の二乗項や交差項を導入しない推定が行われてきました。しかしながらこのような推定は、平均値のみのバランスにとどまり、不十分な可能性が高いと考えられます。例えば、NBER Summer Institute 2018 における、[Esther Duflo のチュートリアル](#) では、連続変数については二乗項、および全変数について交差項を導入した推定を行っています。

複雑な推定は、母集団における推定結果の推論を困難にします。この問題は、事例数が少ない小規模データを用いた推定において深刻です。しかしながら現代的な分析環境のもとでは、1000 事例を超えるデータを用いた推定が一般的になっています。このため、バランスさせたい X の数が少ない場合、その二乗項や交差項を加えたとしても、悪影響は小さいと考えられます。少なくとも小規模事例を用いた推定よりも、より複雑なモデルを推定すべきであると考えられます。

バランスさせたい変数 X の数が多い場合、二乗項や交差項を導入するとモデルが爆発的に複雑化し、OLS では推定できなくなります。このような場合は、Chapter 6 で議論する通り、機械学習を応用が有力です。

4.5 実践上の課題

OLS により暗黙のうちに計算される Weight は、平均値をバランスします。しかしながら、Balancing weights に求められる他の性質は必ずしも満たされません。以下、OLS の抱える問題点を列挙します。

以降の章で紹介する手法の利点は、この OLS が抱える問題点を改善することにあります。このため発展的な手法を理解し、活用するためにも、OLS の問題点をしっかり認識する必要があります。

4.5.1 定式化

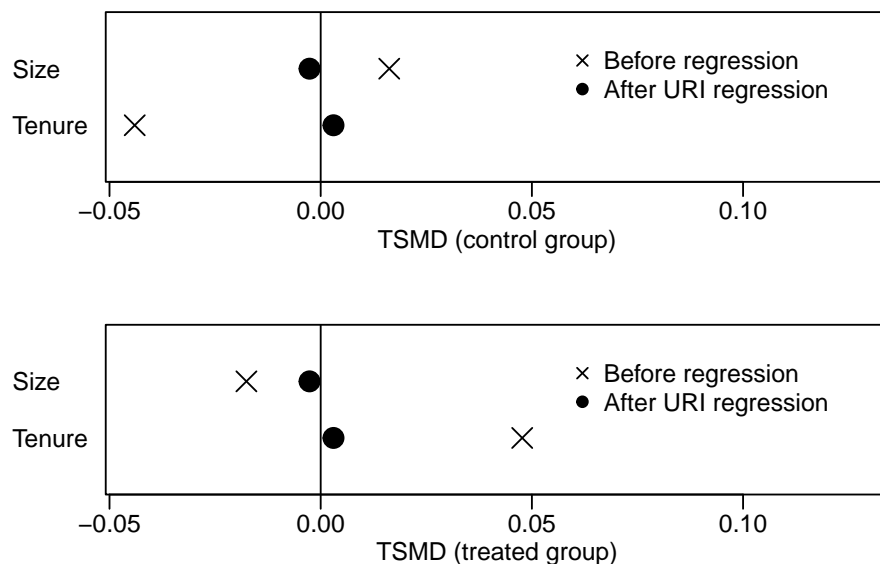
OLS においては、分布の特徴をどこまでバランスさせるのかが問題となります。事例数が十分あれば、3 乗項などの多くの特徴をバランスさせることが可能です。しかしながら事例数が少ない場合、大量のモーメントをバランスさせると、推定誤差が大きくなってしまいます。このため事例数に応じて、バランスの対象とする特徴の数を選択する必要があります。しかしながら実際の応用において、このような選択を適切に行うことは困難です。

この問題に対して、Chapter 6 では機械学習を用いた改善方法を紹介します。

4.5.2 ターゲット割合

バランス後の X の平均値がどのような水準になるのか、一般に不透明です。結果を解釈するためには、 X の平均値は明確な水準、例えばデータ全体での平均値と一致させることが望まれます。しかしながら、OLS はそのような水準との一致を保証しません。

OLS によるバランス後の X の平均値について、[lmw package](#) により診断できます。



黒丸は OLS によるバランス後、ばつ印はバランス前の平均値を示しています。Control group は、 $D = 0$ (2021 年)、Treatment group は、 $D = 1$ (2022 年) の値です。0 線は、サンプル平均を示しています。

同図からバランス前は、2022 年については Size がサンプル平均よりも小さく、Tenure は長くなっています。黒丸を見ると OLS によるバランス後は、2022 年と 2021 年の間で平均差がなくなることが確認できます。ただし 0 線からは乖離しており、サンプル平均とは一致していないことが確認できます。

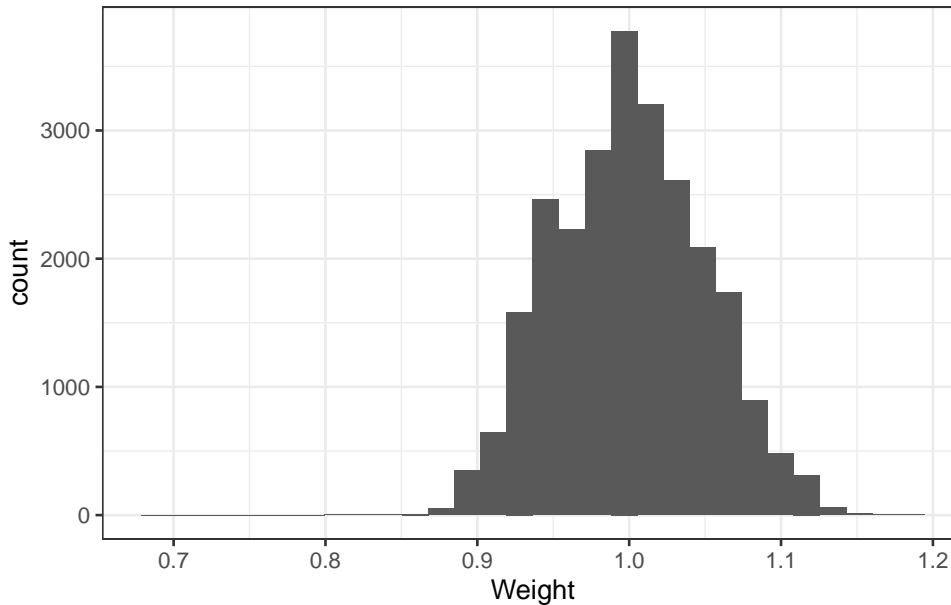
この問題の解決としては、サンプル平均にバランスさせることを明示的に要求した Balancing Weight の算出 (Chapter 5) が有力です。

4.5.3 負のウェイト

Balancing weights は、正の値を取ることが望まれます。しかしながら OLS が生成する Weight は、負の値を取る可能性があり、ミスリードな推定結果をもたらす可能性があります。

ます。

lmw package は、OLS が生成する weights の値を計算します。例えば hist 関数により、ヒストグラムとして可視化できます。



本応用例では、負の weights は発生していないことが確認できました。

負の weights が発生しない方法としては、明示的な Balancing Weight の算出 (Chapter 5) や機械学習を活用した柔軟な推定 (Chapter 6) が有力です。

4.6 R による実践例

- D と X の交差項を含めたモデルの OLS 推定、およびその性質の診断は、以下のパッケージを用いて実装できます。

- readr (tidyverse に同梱): データの読み込み

- lmw: OLS が計算する balance weights を計算

- * [Repository](#)

- estimatr: OLS を Robust standard error とともに計算

- * [Repository](#)

- dotwhisker: 信頼区間の可視化

- * [Repository](#)

4.6.1 準備

データを取得します。 D として、取引年が 2022 か 2021 かで、1/0 となる変数を定義します。シンプルな比較分析について信頼区間は、データ分割は不要です。

```
Data = readr::read_csv("Public.csv") # データ読み込み

Data = dplyr::mutate(
  Data,
  D = dplyr::if_else(
    TradeYear == 2022, 1, 0
  ) # 2022 年に取引されれば 1、2021 年に取引されていれば 0
)
```

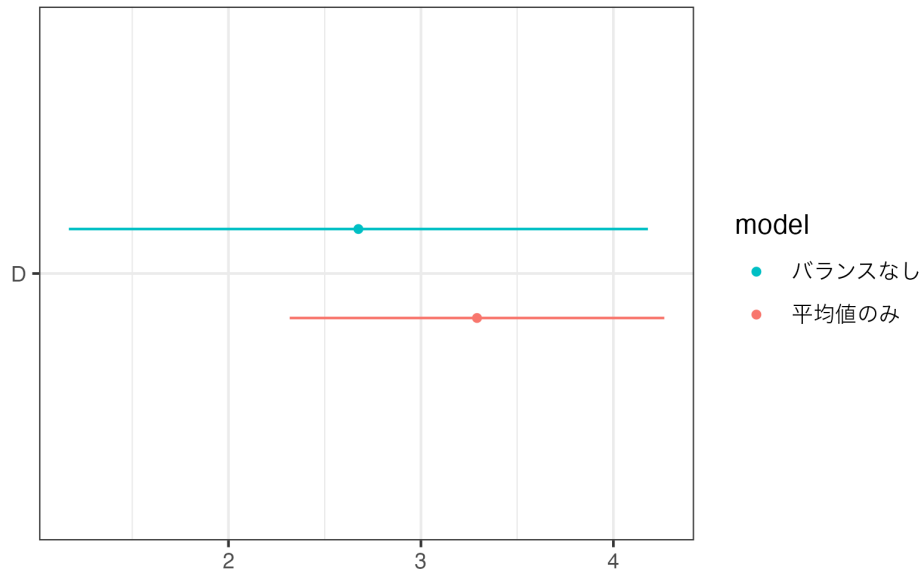
4.6.2 OLS によるバランス

D 間で Size, Tenure, StationDistance の平均値をバランスさせ、Price の平均値を比較します。また比較のために、一才のバランスを行わない比較結果も併記します。

```
Model_NoBalance = estimatr::lm_robust(
  Price ~ D,
  Data) # OLS 推定

Model = estimatr::lm_robust(
  Price ~ D + Size + Tenure + StationDistance,
  Data) # OLS 推定

dotwhisker::dwplot(
  list(バランスなし = Model_NoBalance, 平均値のみ = Model),
  vars_order = "D") + # 信頼区間の可視化
  ggplot2::theme_bw() # 背景を白地化
```



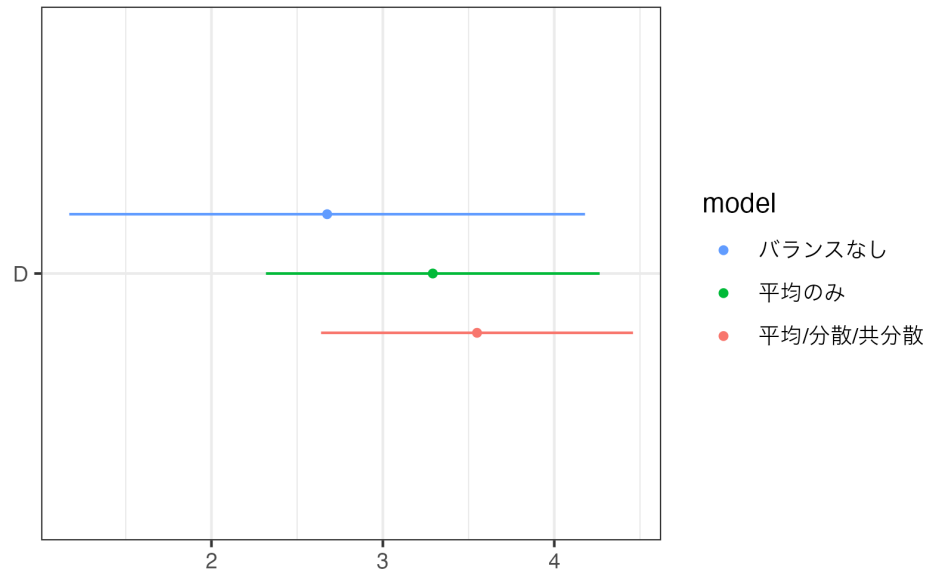
D の係数値は 3.29 であり、バランス後も中心 6 区の物件の方が平均取引価格が高いことがわかります。また信頼区間を考慮することで、母集団においても、中心 6 区の物件の方が平均取引価格が高い傾向があることがわかります。

次に各変数の分散と共分散もバランスさせます

```
ModelLong = estimatr::lm_robust(
  Price ~ D +
    (Size + Tenure + StationDistance)**2 + # 交差項の作成
    I(Size^2) + I(Tenure^2) + I(StationDistance^2), # 分散
  Data)
```

バランスをしない単純比較も含めて、推定結果を比較すると以下ようになります。

```
dotwhisker::dwplot(
  list(
    バランスなし = Model_NoBalance,
    平均のみ = Model,
    `平均/分散/共分散` = ModelLong
  ),
  vars_order = "D"
) +
ggplot2::theme_bw()
```



バランスすることで、推定値が大きくなり、信頼区間が縮小する (推定精度が改善する) ことが確認できます。

4.6.2.1 Balanced Weight

lmw パッケージの lmw 関数を用いれば、OLS が算出している Balance weights を計算できます。

```
Match = lmw::lmw(
  ~ D + I(Size^2) + I(Tenure^2) + I(StationDistance^2) +
    (Size + Tenure + StationDistance)**2, # 平均、分散、共分散をバランス
  Data
) # Weight の算出

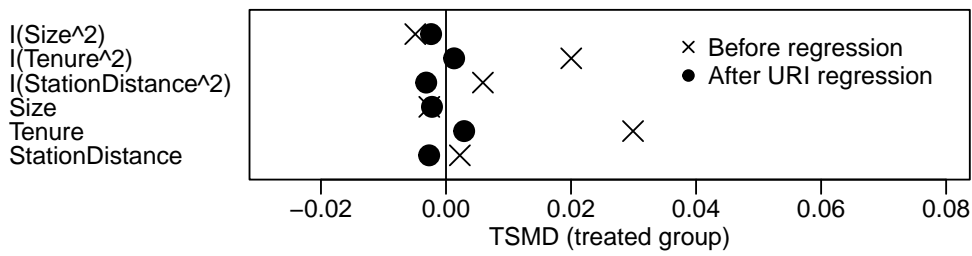
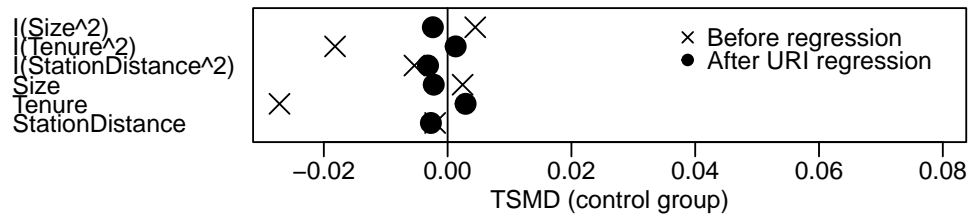
summary(Match$weights) # Weight の記述統計量を図示
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.7005	0.9685	1.0014	1.0000	1.0335	1.3118

負の Weight が発生していないことが確認できます。

データ全体での平均値との乖離も、以下のとおり確認できます。


```
plot(summary(Match), abs = FALSE)
```



第 5 章

直接的な Balancing Weight の算出

Hainmueller (2012) や Zubizarreta (2015) では、Balancing Weight を明示的な”最小化問題”の解として算出します。このようなアプローチは、満たすべき条件 (負の荷重が生じない/サンプル平均にバランスするなど) を課した上で、Balancing Weight が算出できます。このため OLS による暗黙の Balancing に比べて、より透明性の高い分析が可能です。

5.1 算出方法

Balancing Weight について、以下のような制約を課します。

1. 全ての X_l について、その加重平均をデータ全体の平均値に一致させる:

$$\begin{aligned} & (D=1) \text{ における } (\omega(x, 1) \times X_l) \text{ の平均値} \\ &= (D=0) \text{ における } (\omega(x, 0) \times X_l) \text{ の平均値} \\ &= X_l \text{ のデータ全体での平均値} \end{aligned}$$

2. 全ての Weight は非負の値をとる:

$$\omega(x, d) \geq 0$$

上記の制約を満たす $\omega(d, x)$ のなかで、最もばらつきが小さいものを Balancing Weight とします。ばらつきの測定方法は、いくつかの提案があります。

- Hainmueller (2012) : $\omega(x, d)$ の entropy divergence $\omega(x, d) \log(\omega(x, d)/q)$
 - q は base weight と呼ばれ、例えば $q = 1/\text{事例数}$ が用いられます。
- Zubizarreta (2015) : $\omega(x, d)$ の分散

特に entropy divergence を用いる Hainmueller (2012) の方法は、実際の計算も早く実用的です。以下では、実際の実装方法を紹介します。

5.2 R による実践例

- D と X の交差項を含めたモデルの OLS 推定、およびその性質の診断は、以下のパッケージを用いて実装できます。
 - readr (tidyverse に同梱): データの読み込み
 - WeightIt: entroy weight の計算
 - * [Repository](#)
 - marginaleffects: WeightIt パッケージが計算する Weight を用いた推定
 - * [Repository](#)

5.2.1 準備

データを取得します。 D として、立地が中心 6 区か否かで、1/0 となる変数を定義します。

```
Data = readr::read_csv("Public.csv") # データ読み込み

Data = dplyr::mutate(
  Data,
  D = dplyr::if_else(
    LargeDistrict == "中心 6 区", 1, 0
  ) # 中心 6 区に立地していれば 1、それ以外は 0
)
```

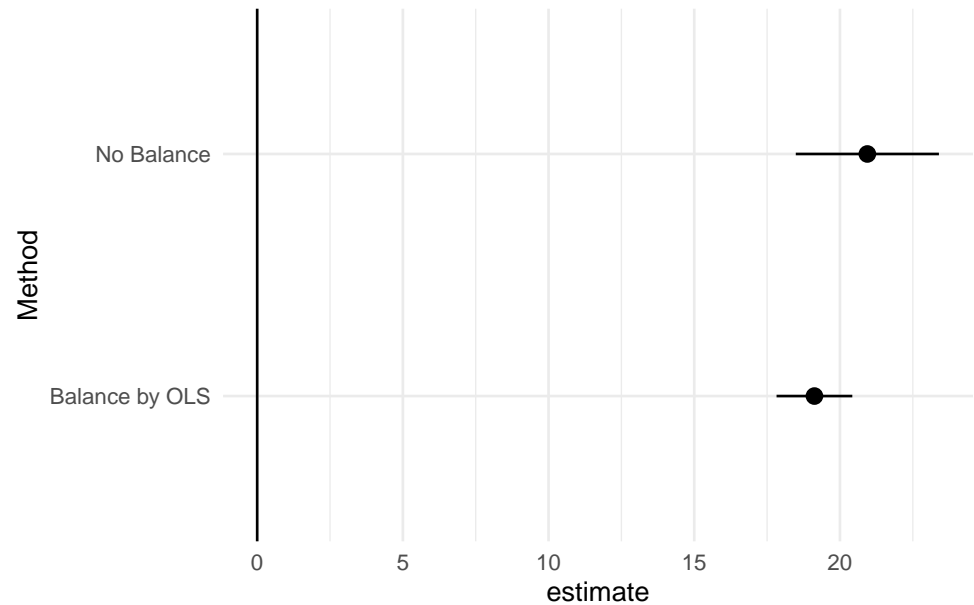
ベンチマークとして、一切のバランスを行わない単純差、および OLS によるバランスを行った結果を示します。

```
Model_No = estimatr::lm_robust(Price ~ D, Data)

Model_OLS = estimatr::lm_robust(
  Price ~ D + (Size + Tenure + StationDistance)**2 +
  I(Size^2) + I(Tenure^2) + I(StationDistance^2),
  Data)

Model_No |>
  generics::tidy() |>
  dplyr::mutate(
```

```
  Method = "No Balance"
) |>
dplyr::bind_rows(
  Model_OLS |>
  generics::tidy() |>
  dplyr::mutate(
    Method = "Balance by OLS"
  )
) |>
dplyr::filter(
  term == "D"
) |>
ggplot2::ggplot(
  ggplot2::aes(
    y = Method,
    x = estimate,
    xmin = conf.low,
    xmax = conf.high
  )
) +
ggplot2::geom_pointrange() +
ggplot2::theme_minimal() +
ggplot2::geom_vline(
  xintercept = 0
)
```



5.2.2 Entropy Weight によるバランス

D 間で Size, Tenure, StationDistance の平均値をバランスさせ、Price の平均値を比較します。まず WeightIt パッケージを用いて、Entropy Weight を計算します。また比較のために lmw パッケージを用いて、OLS Weight も計算します。

```
Entropy = WeightIt::weightit(
  D ~ (Size + Tenure + StationDistance)**2 +
    I(Size^2) + I(Tenure^2) + I(StationDistance^2), # 平均、分散、共分散をバランス
  data = Data,
  method = "ebal", # Entropy Weight を計算
  estimand = "ATE")

OLS = lmw::lmw(
  ~ D + (Size + Tenure + StationDistance)**2 +
    I(Size^2) + I(Tenure^2) + I(StationDistance^2),
  Data
)
```

summary 関数を用いて、Weight の分布を確認します。

```
summary(Entropy$weights)
```

```
Min. 1st Qu. Median Mean 3rd Qu. Max.
```

```
0.2533 0.8622 0.9706 1.0000 1.0877 16.5677
```

```
summary(OLS$weights)
```

```
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.2343  0.7615  1.0007  1.0000  1.2484  2.7393
```

OLS を用いると負の Balancing weight が発生していることが確認できます。対して Entropy weight では、そのような weight は生じません。

Entropy Weight を用いたバランス後の平均差は、marginaleffects パッケージの avg_comparison 関数を用いて、以下のように計算できます。

```
WeightIt::lm_weightit(
  Price ~ D*(I(Size^2) + I(Tenure^2) + I(StationDistance^2) +
    (Size + Tenure + StationDistance)**2),
  data = Data,
  weightit = Entropy
) |>
marginaleffects::avg_comparisons(variables = "D")
```

```
Estimate Std. Error    z Pr(>|z|)      S 2.5 % 97.5 %
      19.9      0.625 31.8  <0.001 736.2  18.7   21.1
```

```
Term: D
```

```
Type: probs
```

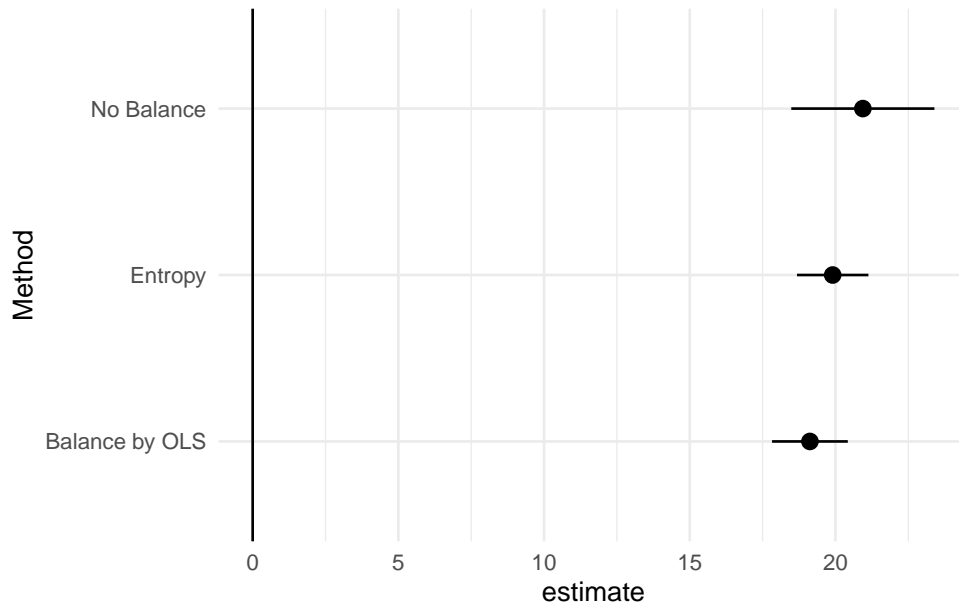
```
Comparison: 1 - 0
```

引き続き 20 前後の平均価格差が算出されました。

最後にバランスなし、OLS によるバランス、Entropy Weight によるバランス、の結果を図示します。

```
Model_Ent = WeightIt::lm_weightit(
  Price ~ D*(I(Size^2) + I(Tenure^2) + I(StationDistance^2) +
    (Size + Tenure + StationDistance)**2),
  data = Data,
  weightit = Entropy
) |>
marginaleffects::avg_comparisons(variables = "D")
```

```
Model_No |>
  generics::tidy() |>
  dplyr::mutate(
    Method = "No Balance"
  ) |>
  dplyr::bind_rows(
    Model_OLS |>
    generics::tidy() |>
    dplyr::mutate(
      Method = "Balance by OLS"
    )
  ) |>
  dplyr::filter(
    term == "D"
  ) |>
  dplyr::bind_rows(
    tibble::tibble(
      Method = "Entropy",
      estimate = Model_Ent$estimate,
      conf.low = Model_Ent$conf.low,
      conf.high = Model_Ent$conf.high
    )
  ) |>
  ggplot2::ggplot(
    ggplot2::aes(
      y = Method,
      x = estimate,
      xmin = conf.low,
      xmax = conf.high
    )
  ) +
  ggplot2::geom_pointrange() +
  ggplot2::theme_minimal() +
  ggplot2::geom_vline(
    xintercept = 0
  )
```

5.3 発展

Balancing Weight を直接算出する方法は、近年改めて注目を集めており、さまざまな提案がなされています。例えば Auto debiased machine learning (Chernozhukov, Newey, and Singh 2022a, 2022b; Chernozhukov, Newey, et al. 2022) と呼ばれる手法も、分布をバランスする Weight を機械学習などを用いて直接算出する手法であるとの解釈も提案されています (Bruns-Smith et al. 2023)。Energy Weight (Huling and Mak 2024) も、特徴ではなく分布そのものをバランスする Weight の推定を行なっています。

直接計算する手法は、AIPW などの間接的な方法 (Chapter 7) に比べて、推定結果が安定しやすいことが利点として挙げられます。

第 6 章

機械学習の活用：残差回帰

X の数が多い場合、OLS を用いた分布の特徴のバランスは、難しくなります。これはどのような特徴をバランスさせるのか、研究者の裁量の余地が大きくなりすぎるためです。このような研究者による推定式の定式化に起因する推定結果のブレは、Model Uncertainty (Varian 2014) と呼ばれ、分析の信頼性を脅かす大きな要因となります。

このような問題への対応として、機械学習を有効活用した、よりデータ主導のアプローチが注目されて来ました (Varian 2014; Urminsky, Hansen, and Chernozhukov 2019)。近年、機械学習の中でも教師付き学習の手法を、バランス後の比較に応用する方法について理解が急速に進みました。その中で教師付き学習 (より一般にはノンパラメトリック/データ適合的な推定) が持つ弱点である「収束の遅さ」*1 を補完する方法が開発されています。このような方法を用いることで、信頼区間の近似計算など、母集団への含意がより明確な推定が可能となります。

本章は、OLS の直接的な拡張と解釈できる **残差回帰** への応用を紹介します*2。

6.1 OLS の一般化

OLS を用いて、 X の分布を完全にバランスさせるには、以下のようなモデルを推定する必要があります。

$$Y \sim D + \underbrace{f(X_1, \dots, X_L)}_{X \text{ についての極めて複雑な式}}$$

*1 「収束」の具体的な定義に関心がある方は、[wikipedia の記事](#)などを参照ください。

*2 機械学習は、より幅広い推定方法に応用できます。一般的な入門としては、[Chernozhukov et al. \(2024\)](#) を推奨します。他の入門資料としても、[Wager \(2024\)](#)、[Schuler and Laan \(2024\)](#)、[Ichimura and Newey \(2022\)](#)、[Fisher and Kennedy \(2021\)](#)、[Hines et al. \(2022\)](#)、[Kennedy \(2024\)](#)、[Renson et al. \(2025\)](#) などの優れた教材が利用できます。簡潔な入門としては、[Díaz \(2020\)](#) がお勧めです。

$f(X_1, \dots, X_L)$ は、一般に無限個のパラメータを持つ式となります。

例えば 1 つの連続変数, X_1 , のみのバランスが目標であるとします。この場合でも、推定式は以下のように無限個のパラメータを持つ式として定式化できます。

$$f(X_1) = \underbrace{\beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_1^3 \dots}_{\text{無限和}}$$

このような無限個のパラメータを持つ式を、OLS で推定することは不可能です。あるいはパラメータの数が有限個であったとしても、データの事例数を超えると推定が不可能になります*3。

6.2 残差回帰

残差回帰は、 $Y \sim D + f(X_1, \dots, X_L)$ が以下のように書き換えることができることに基づいています。

$$\underbrace{Y - (X\text{内での})Y\text{の平均値}}_{Y\text{の残差}} \sim \underbrace{D - (X\text{内での})D\text{の平均値}}_{D\text{の残差}}$$

すなわち Y の残差と D の残差を単回帰すれば、バランス後の平均値の比較を達成できます。

残る課題は、 Y と D の平均値をデータから推定することです。このような推定は、 X が多くある場合、困難な課題でした。以下、OLS 推定を再解釈することで、どのような問題が生じるのか示します。

6.2.1 OLS の再解釈

FWL 定理 ([wiki](#)) は、OLS も残差回帰として解釈できることを示しています。

OLS によって推定された D の係数値は、以下の手順で計算された値と厳密に一致します。

! FWL(Frisch–Waugh–Lovell) 定理

0. $Y/D/X$ を設定

1. 全データを用いて、 Y , D の平均値の推定値 $f_Y(X), f_D(X)$ を $Y \sim \beta_0 + \beta_1 X_1 + \dots + \beta_L X_L$ および $D \sim \beta_0 + \beta_1 X_1 + \dots + \beta_L X_L$ を OLS 推定することで算出

*3 パラメータと事例数、推定精度の詳細な関係性については、[Chernozhukov et al. \(2024\)](#) の 1 章とその引用文献を参照ください。

2. 残差 $Y - f_Y(X)$ と $D - f_D(X)$ を計算
3. 残差同士を回帰する: $Y - f_Y(X) \sim D - f_D(X)$

X が多い場合、Step 1. が大きな問題を抱えます。一般にパラメタ (β_0, \dots, β_L) の数が、事例数に近い値になると、平均値の推定値は Y, D の実際の値に限りなく近づいていきます。このような現象は、過剰適合、あるいは過学習と呼ばれています。

機械学習の応用では、Step 1. を Y と D を X から予測する問題と捉えます。このような予測問題においては、OLS も含めたさまざまな推定手法を用いることができます。

6.2.2 教師付き学習の応用

ある一つの連続/カテゴリー変数 Y や D と (大量の) 変数 X の関係性の推定は、教師付き学習の中心的な研究課題です。中でも、(X 内での) Y や D の平均値の推定については、膨大な研究成果が蓄積されています^{*4}。

一般に教師付き学習は、ある結果変数の予測を目的としています。予測の精度を平均二乗誤差で測定する場合、理論上最善の予測値は、(X 内での) 母平均であることが容易に証明できます。このため、(X 内での) 母平均を近似できるモデルの推定が、教師付き学習の実質的な目標となります。この性質を利用することで、既存の機械学習の方法で推定された予測モデルを、平均値を近似するモデルとして利用することができます。

推定に用いるアルゴリズムとしては、複数の予測モデルを組み合わせる Stacking 法が推奨されます (Ahrens et al. 2024; Naimi, Mishler, and Kennedy 2023)。これは多くの社会分析において、高い予測性能を期待できるアルゴリズムを事前に選択することが困難であるためです。

予測モデルを推定する際には、Section 6.3 で議論する優れた推定上の性質を達成するために、交差推定を活用することが、一般に推奨されます (Naimi, Mishler, and Kennedy 2023)^{*5}。

! 交差推定 (Cross Validation)

0. データを細かく分割 (第 1, ..., 10 サブグループなど)
1. 第 1 サブグループ以外で推定して、第 1 サブグループの予測値を算出
2. 第 2...サブグループについて、繰り返し、全事例に対して予測値を算出

交差推定を活用した、残差回帰の具体的な推定手順は以下となります。

^{*4} 詳細は、James et al. (2021) などを参照ください

^{*5} 交差推定を活用すべきかについては、Chen, Syrgkanis, and Austern (2022) などで発展的な議論が行われています。

! 機械学習を活用した残差回帰

0. $Y/D/X$ および予測に使用するアルゴリズムを設定
1. 交差推定を用いて、 Y と D を予測するモデル $f_Y(X)$ と $f_D(X)$ を推定する
2. 残差 $Y - f_Y(X)$ と $D - f_D(X)$ を計算する
 - 予測誤差とも解釈できます。
3. 残差同士を回帰する: $Y - f_Y(X) \sim D - f_D(X)$

6.3 母集団の推定方法

交差推定を併用した残差回帰の利点は、「母集団上で X を完全にバランスさせた後に計算した平均差」について、推論が可能なことです。特に交差推定と Stacking などの柔軟な方法で予測モデルを推定すれば、緩やかな仮定のみで信頼区間を近似計算が可能です (Chernozhukov et al. 2018)。

6.4 応用上の課題

機械学習を活用した残差回帰は、OLS の持つ幾つかの問題点を改善します。

十分な事例数があれば、「母集団上で X を完全にバランスさせた後に計算した平均差」を近似するため、負の荷重問題は生じません。

定式化そのものではなく、推定方法を選ぶため、Researcher degrees of freedom を適切に軽減できる可能性があります (Urminsky, Hansen, and Chernozhukov 2019; Ludwig, Mullainathan, and Spiess 2019)。特に複数の推定方法を組み合わせる Stacking 法とシード値の適切な管理^{*6}によっては、研究者への依存度を低下させられます。

問題点としては、ターゲットの解釈が容易ではないことが知られています。母集団上での残差回帰は、 X を完全にバランスさせますが、そのターゲットは、以下となります。

$$\frac{(X内での)D = 1の割合 \times (X内での)D = 0の割合}{[(X内での)D = 1の割合 \times (X内での)D = 0の割合]の平均値}$$

このようなターゲットは、Overlap Weight とも呼ばれています。

Overlap Weight をターゲットとすると、 D のばらつきが大きいグループの平均値を、最終的な比較結果に強く反映します。 D のばらつきが無いグループ ($D = 0$ または $D = 1$ しか存在しない) の加重は 0 であり、最終的な比較結果に一才の影響を与えません。このため Overlap の仮定 (Important 1) を「必ず」満たすことができます。

^{*6} Naimi, Yu, and Bodnar (2024), Schader et al. (2024); Zivich (2024)

このような優れた性質を持つ一方で、このターゲットをどのように解釈すればいいのか、一般に不透明です。特に X 間での D の偏りが大きい場合、Overlap Weight のばらつきも大きくなり、解釈が一層難しくなります。Zhou and Opacic (2022) など、Overlap Weight の解釈を提供する研究は行われていますが、確立された解釈は現状ありません。

6.5 R による実践例

- 以下のパッケージを使用
 - readr (tidyverse に同梱): データの読み込み
 - ddml: 残差回帰の実施
 - * [manual](#)

6.5.1 準備

ddml パッケージの関数を用いる場合、事前に Y と D はベクトルとして、 X は行列として定義する必要があります。

```
set.seed(111) # シード値を固定

Data = readr::read_csv("Public.csv")

Data = dplyr::mutate(
  Data,
  D = dplyr::if_else(
    TradeYear == 2022, 1, 0
  ) # 2022 年に取引されれば 1、2021 年に取引されていれば 0
)

Y = Data$Price # Y の定義
D = Data$D # D の定義
X = Data |>
  select(
    District,
    Size,
    Tenure,
    StationDistance
  ) |>
```

```
data.matrix() # X の定義
```

6.5.2 残差回帰

Stacking 法を活用した残差回帰は、ddml パッケージの `ddm_plm` 関数を用いて実装できます。

```
Model = ddml::ddml_plm(
  y = Y, # Y の指定
  D = D, # D の指定
  X = X, # X の指定
  learners = list(
    list(fun = ddml::ols), # OLS を使用
    list(fun = ddml::mdl_ranger) # RandomForest を使用
  ),
  shortstack = TRUE, # 簡略化した Stacking 法を利用
  silent = TRUE # Message を非表示
)

Model |> summary()
```

PLM estimation results:

```
, , nnls
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.374	0.157	-2.38	1.73e-02
D_r	3.730	0.328	11.38	5.01e-30

`summary` 関数は、推定値 (Estimate)、標準誤差 (Std. Error)、t 値 (t value)、p 値 ($\Pr(>|t|)$) を報告しています。バランス後の平均差は、`D_r` です。

95% 信頼区間は以下のように計算できます。

```
Sum = summary(Model)

Est = Sum[2,1,1] # 推定値
SD = Sum[2,2,1] # 標準誤差

c(Est - 1.96*SD, Est + 1.96*SD) # 信頼区間の下限、上限
```

```
[1] 3.088172 4.372682
```


第 7 章

機械学習の活用: AIPW

D 間での X の分布を近似的にバランスする手法としては、Inverse Probability Weight (IWP) も有力です。本手法は、(X 内での) D の分布を推定し、その逆数をウェイトとして使用する方法です。この D の分布は、傾向スコア (Propensity Score) と呼ばれています。特に Y の推定されたモデルと組みわせる Augmented Inverse Probability Weight は、機械学習などを用いた柔軟な推定と信頼区間の近似計算などの統計的推論と両立できます。

7.1 Inverse probability weight

傾向スコアを活用した代表的な推定手法は、Inverse probability weight です。この手法は、ベイズルールによる Balancing weights の書き換えに基づいています。ただしこの手法は、傾向スコアの推定精度に強く依存します。また一般に推定精度が悪い傾向があります^{*1}。このため本ノートでは、Section 7.2 で紹介する Argumented inverse probability weight (AIPW) の活用を推奨します。

Balancing Weight $\omega(d, x)$ は以下のように定義されました。

$$\omega(d, x) = \frac{\text{ターゲット割合}}{D = d \text{ における } X = x \text{ の事例割合}}$$

ベイズルールを適用すると $\omega(d, x)$ は以下のように書き換えられます。

! Balancing Weight の書き換え

$$\omega(d, x) = \frac{1}{\underbrace{X = x \text{ における } D = d \text{ の割合}}_{\text{傾向スコアの逆数}}} \times (D = d \text{ の割合})$$

^{*1} 詳細は、Wager (2024) の第 2 章などを参照ください。

$$\times \frac{\text{ターゲット割合}}{X=x \text{の割合}}$$

ターゲット割合を $X = x$ の割合 とするのであれば、以下のように単純化できる。

$$\omega(d, x) = \text{傾向スコアの逆数} \times D = d \text{の割合}$$

$D = d$ の割合は、データ内での $D = d$ の割合などを用いて、容易に推定できます。よって傾向スコアが推定できれば、Balancing Weight の計算が可能です。推定された傾向スコアを用いた Balanced comparison は Inverse probability weight と呼ばれ、以下の手順としてまとめられます。

! Inverse probability weight

1. 傾向スコアを何らかの手法で推定する
2. Balancing weights $\omega(d, x)$ を計算する
3. Balancing weights で荷重した平均差を計算する

$D = 1$ における $(\omega(x, 1) \times Y)$ の平均値

$-D = 0$ における $(\omega(x, 0) \times Y)$ の平均値

この手法は傾向スコアの推定には、 D を X で回帰する標準的な方法を用いることができます。例えば Logit や Probit などのパラメトリックな手法や Random Forest や Boosting、Deep Learning などの機械学習の手法も活用可能です。

しかしながらどちらの手法を用いたとしても、推定上の問題があります。IPW による推定結果は、傾向スコアの推定精度に強く依存します。このため Logit や Probit を用いた場合、推定モデルの定式化を正しく行う必要がありますが、多くの応用で困難です。対して機械学習の手法を用いた場合、推定結果の収束の遅さが問題となり、信頼区間の近似計算などが難しくなります。

この問題に対して、 Y の平均値の推定値 $f_Y(D, X)$ を併用する Augmented Inverse Probability Weight が有力です。

7.2 Augmented Inverse Probability Weight

Augmented Inverse Probability Weight の手順は以下のようにまとめられます。

! Augmented Inverse Probability Weight

1. 交差推定を用いて、 Y, D を予測するモデル $f_Y(D, X)$ と $f_D(X)$ を推定する
2. AIPW の推定値を以下のように補正し、AIPW の推定値を計算する

$$+f_Y(1, X_i) \frac{D - f_D(X)}{f_D(X)} \text{の平均値}$$

$$-f_Y(0, X_i) \frac{(1 - D) - (1 - f_D(X))}{1 - f_D(X)} \text{の平均値}$$

7.3 母集団の推定方法

母集団上での AIPW とデータ上での AIPW は、残差回帰とよく似た性質を持ちます。交差推定と Stacking などの柔軟な方法を組み合わせて予測モデルを推定すれば、緩やかな仮定のみで、「母集団上で計算した Augmented Inverse Probability Weight」の結果について、信頼区間の近似計算が可能です。AIPW は、データ全体での X の分布をターゲット割合としたバランス後の比較です。このため「母集団上での X の分布をターゲットとした、バランス後の比較」について、信頼区間計算ができます。

7.4 残差回帰との比較

残差回帰と AIPW は、よく似た性質を持ち、同じような動機に基づいています^{*2}。

大きな違いは、ターゲット割合です。残差回帰におけるターゲット割合は、Overlap Weight であり (Chapter 6)、その解釈は困難でした。対して AIPW では、 X の分布であり、解釈が容易です。すなわち「もし X の分布がデータ全体のものと一致していた場合、平均差はどのようなものになるか？」という疑問に答えることができます。このような解釈が容易さが、AIPW の大きな魅力です。

対して AIPW の弱点は、傾向スコアの値が 0 や 1 に近い事例がある場合、推定結果が不安定になりやすい点です。これは傾向スコアの逆数が推定式に含まれているために生じます。弊害の具体例としては、計算された信頼区間の拡大や信頼区間の近似計算そのものの信頼性低下、などです。

7.5 R による実践例

引き続き ddml パッケージを使用し、AIPW 推定は実装できます。

^{*2} どちらも Efficient influence function を用いて、推定式を導出できます (Ichimura and Newey 2022; Fisher and Kennedy 2021; Hines et al. 2022; Renson et al. 2025)。

- 以下のパッケージを使用
 - readr (tidyverse に同梱): データの読み込み
 - ddml: AIPW の実施
 - * [manual](#)

7.5.1 準備

Y と D はベクトル、 X は行列として定義します。

```
set.seed(111) # シード値を固定

Data = readr::read_csv("Public.csv")

Data = dplyr::mutate(
  Data,
  D = dplyr::if_else(
    TradeYear == 2022, 1, 0
  ) # 2022 年に取引されれば 1、2021 年に取引されていれば 0
)

Y = Data$Price
D = Data$D
X = Data |>
  dplyr::select(
    District,
    Size,
    Tenure,
    StationDistance
  )
```

7.5.2 ATE

Stacking を用いた AIPW 推定は、`ddml_ate` 関数を用いて、以下のように実装できます。

```
Model = ddml::ddml_ate(
  y = Y,
  D = D,
  X = data.matrix(X),
```

```
learners = list(  
  list(fun = ddml::ols),  
  list(fun = ddml::mdl_ranger)  
)  
shortstack = TRUE, # 簡略化した Stacking 法を利用  
silent = TRUE # Message を非表示  
)  
  
Model |> summary()
```

ATE estimation results:

	Estimate	Std. Error	t value	Pr(> t)
nnls	3.68	0.339	10.8	2.52e-27

summary 関数は、推定値 (Estimate)、標準誤差 (Std. Error)、t 値 (t value)、p 値 (Pr(>|t|)) を報告しています。バランス後の平均差は、nnls です。

95% 信頼区間は以下のように計算できます。

```
Sum = summary(Model)  
  
Est = Sum[1,1] # 推定値  
SD = Sum[1,2] # 標準誤差  
  
c(Est - 1.96*SD, Est + 1.96*SD) # 信頼区間の下限、上限
```

```
[1] 3.010346 4.340967
```


Reference

- Ahrens, Achim, Christian B Hansen, Mark E Schaffer, and Thomas Wiemann. 2024. “Model Averaging and Double Machine Learning.” *arXiv Preprint arXiv:2401.01645*.
- Bajari, Patrick, Zhihao Cen, Victor Chernozhukov, Manoj Manukonda, Suhas Vijaykumar, Jin Wang, Ramon Huerta, et al. 2023. “Hedonic Prices and Quality Adjusted Price Indices Powered by AI.” *arXiv Preprint arXiv:2305.00044*.
- Behaghel, Luc, Bruno Crépon, and Marc Gurgand. 2014. “Private and Public Provision of Counseling to Job Seekers: Evidence from a Large Controlled Experiment.” *American Economic Journal: Applied Economics* 6 (4): 142–74.
- Ben-Michael, Eli, Avi Feller, David A Hirshberg, and José R Zubizarreta. 2021. “The Balancing Act in Causal Inference.” *arXiv Preprint arXiv:2110.14831*.
- Bruns-Smith, David, Oliver Dukes, Avi Feller, and Elizabeth L Ogburn. 2023. “Augmented Balancing Weights as Linear Regression.” *arXiv Preprint arXiv:2304.14545*.
- Chattopadhyay, Ambarish, Christopher H Hase, and José R Zubizarreta. 2020. “Balancing Vs Modeling Approaches to Weighting in Practice.” *Statistics in Medicine* 39 (24): 3227–54.
- Chattopadhyay, Ambarish, and José R Zubizarreta. 2023. “On the Implied Weights of Linear Regression for Causal Inference.” *Biometrika* 110 (3): 615–29.
- Chattopadhyay, Ambarish, and José R. Zubizarreta. 2024. *Harvard Data Science Review* 6 (1).
- Chen, Qizhao, Vasilis Syrgkanis, and Morgane Austern. 2022. “Debiased Machine Learning Without Sample-Splitting for Stable Estimators.” *Advances in Neural Information Processing Systems* 35: 3096–3109.
- Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. 2018. “Double/Debiased Machine Learning for Treatment and Structural Parameters.” *The Econometrics Journal* 21 (1): C1–68.
- Chernozhukov, Victor, Juan Carlos Escanciano, Hidehiko Ichimura, Whitney K Newey, and James M Robins. 2022. “Locally Robust Semiparametric Estimation.” *Econo-*

- metrica* 90 (4): 1501–35.
- Chernozhukov, Victor, Christian Hansen, Nathan Kallus, Martin Spindler, and Vasilis Syrgkanis. 2024. “Applied Causal Inference Powered by ML and AI.” *arXiv Preprint arXiv:2403.02467*.
- Chernozhukov, Victor, Whitney K Newey, and Rahul Singh. 2022a. “Automatic Debiased Machine Learning of Causal and Structural Effects.” *Econometrica* 90 (3): 967–1027.
- . 2022b. “Debiased Machine Learning of Global and Local Parameters Using Regularized Riesz Representers.” *The Econometrics Journal* 25 (3): 576–601.
- Chernozhukov, Victor, Whitney Newey, Victor M Quintas-Martinez, and Vasilis Syrgkanis. 2022. “Riesznet and Forestriesz: Automatic Debiased Machine Learning with Neural Nets and Random Forests.” In *International Conference on Machine Learning*, 3901–14. PMLR.
- Díaz, Iván. 2020. “Machine Learning in the Estimation of Causal Effects: Targeted Minimum Loss-Based Estimation and Double/Debiased Machine Learning.” *Biostatistics* 21 (2): 353–58.
- Egami, Hiroyuki, Md Shafiur Rahman, Tsuyoshi Yamamoto, Chihiro Egami, and Takahisa Wakabayashi. 2024. “Causal Effect of Video Gaming on Mental Well-Being in Japan 2020–2022.” *Nature Human Behaviour* 8 (10): 1943–56.
- Fisher, Aaron, and Edward H Kennedy. 2021. “Visually Communicating and Teaching Intuition for Influence Functions.” *The American Statistician* 75 (2): 162–72.
- Gelman, Andrew, Jessica Hullman, and Lauren Kennedy. 2024. “Causal Quartets: Different Ways to Attain the Same Average Treatment Effect.” *The American Statistician* 78 (3): 267–72.
- Hainmueller, Jens. 2012. “Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies.” *Political Analysis* 20 (1): 25–46.
- Hines, Oliver, Oliver Dukes, Karla Diaz-Ordaz, and Stijn Vansteelandt. 2022. “Demystifying Statistical Learning Based on Efficient Influence Functions.” *The American Statistician* 76 (3): 292–304.
- Huling, Jared D, and Simon Mak. 2024. “Energy Balancing of Covariate Distributions.” *Journal of Causal Inference* 12 (1): 20220029.
- Ichimura, Hidehiko, and Whitney K Newey. 2022. “The Influence Function of Semiparametric Estimators.” *Quantitative Economics* 13 (1): 29–61.
- Imai, Kosuke, and Marc Ratkovic. 2014. “Covariate Balancing Propensity Score.” *Journal of the Royal Statistical Society Series B: Statistical Methodology* 76 (1): 243–63.
- Jackson, John W, and Tyler J VanderWeele. 2018. “Decomposition Analysis to Identify Intervention Targets for Reducing Disparities.” *Epidemiology (Cambridge,*

- Mass.*) 29 (6): 825.
- James, Gareth, Daniela Witten, Trevor Hastie, Robert Tibshirani, et al. 2021. *An Introduction to Statistical Learning*. Vol. 112. Springer.
- Kallus, Nathan. 2023. “Treatment Effect Risk: Bounds and Inference.” *Management Science* 69 (8): 4579–90.
- Kennedy, Edward H. 2024. “Semiparametric Doubly Robust Targeted Double Machine Learning: A Review.” *Handbook of Statistical Methods for Precision Medicine*, 207–36.
- Ludwig, Jens, Sendhil Mullainathan, and Jann Spiess. 2019. “Augmenting Pre-Analysis Plans with Machine Learning.” In *Aea Papers and Proceedings*, 109:71–76. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.
- Mossé, Milan, Kara Schechtman, Frederick Eberhardt, and Thomas Icard. 2025. “Modeling Discrimination with Causal Abstraction.” *arXiv Preprint arXiv:2501.08429*.
- Naimi, Ashley I, Alan E Mishler, and Edward H Kennedy. 2023. “Challenges in Obtaining Valid Causal Effect Estimates with Machine Learning Algorithms.” *American Journal of Epidemiology* 192 (9): 1536–44.
- Naimi, Ashley I, Ya-Hui Yu, and Lisa M Bodnar. 2024. “Pseudo-Random Number Generator Influences on Average Treatment Effect Estimates Obtained with Machine Learning.” *Epidemiology* 35 (6): 779–86.
- Opacic, Aleksei, Lai Wei, and Xiang Zhou. 2023. “Disparity Analysis: A Tale of Two Approaches.”
- Renson, Audrey, Lina Montoya, Dana E Goin, Iván Díaz, and Rachael K Ross. 2025. “Pulling Back the Curtain: The Road from Statistical Estimand to Machine-Learning Based Estimator for Epidemiologists (No Wizard Required).” *arXiv Preprint arXiv:2502.05363*.
- Rose, Evan K. 2023. “A Constructivist Perspective on Empirical Discrimination Research.” *Journal of Economic Literature* 61 (3): 906–23.
- Schader, Lindsey, Weishan Song, Russell Kempker, and David Benkeser. 2024. “Don’t Let Your Analysis Go to Seed: On the Impact of Random Seed on Machine Learning-Based Causal Inference.” *Epidemiology* 35 (6): 764–78.
- Schuler, Alejandro, and Mark van der Laan. 2024. “Introduction to Modern Causal Inference.” preparation.
- Stuart, Elizabeth A, Gary King, Kosuke Imai, and Daniel Ho. 2011. “MatchIt: Non-parametric Preprocessing for Parametric Causal Inference.” *Journal of Statistical Software*.
- Urminsky, Oleg, Christian Hansen, and Victor Chernozhukov. 2019. “The Double-Lasso Method for Principled Variable Selection.”
- Vafa, Keyon, Susan Athey, and David M Blei. 2024. “Estimating Wage Disparities Using Foundation Models.” *arXiv Preprint arXiv:2409.09894*.

- Varian, Hal R. 2014. “Big Data: New Tricks for Econometrics.” *Journal of Economic Perspectives* 28 (2): 3–28.
- Wager, Stefan. 2024. “Causal Inference: A Statistical Learning Approach.” preparation.
- Zhou, Xiang, and Aleksei Opacic. 2022. “Marginal Interventional Effects.” *arXiv Preprint arXiv:2206.10717*.
- Zivich, Paul N. 2024. “Commentary: The Seedy Side of Causal Effect Estimation with Machine Learning.” *Epidemiology* 35 (6): 787–90.
- Zubizarreta, José R. 2015. “Stable Weights That Balance Covariates for Estimation with Incomplete Outcome Data.” *Journal of the American Statistical Association* 110 (511): 910–22.