

バランス後の母平均差の推定 (ver 0.0)

川田恵介

2024-11-04

Table of contents

Preface	5
第 1 章 Balanced Comparison	7
1.1 実例	7
1.1.1 不動産市場の年次比較	7
1.1.2 合計特殊出生率	8
1.1.3 既存店ベースの比較	8
1.2 推定対象	9
1.2.1 バランス後の平均値	9
第 2 章 Balancing Weight	11
2.1 仮定: Overlap	11
2.2 推定方法	12
2.3 R による実践例	13
2.3.1 準備	13
2.3.2 Balanced Weight	13
第 3 章 特徴のバランス	17
3.1 OLS	17
3.1.1 例	18
3.1.2 Post selection	19
3.2 OLS の問題点と解決策	20
3.2.1 解釈の難しさ	20
3.2.2 負の荷重	21
3.2.3 解決策	22
3.3 R による実践例	22
3.3.1 準備	23
3.3.2 Balanced comparison by OLS	23
3.4 Reference	25
第 4 章 Inverse Probability Weight	27

4.1	Inverse probability weighting	27
4.2	傾向スコアの推定方法	28
4.3	問題点	28
4.4	他の方法	28
4.5	Reference	28
第 5 章	Augmentation	29
5.1	動機	29
5.2	Augmented Inverse Probability Weighting	29
5.2.1	交差推定	30
5.3	R による実践例	31
5.3.1	準備	31
5.3.2	Balanced comparson by OLS	32
5.4	Reference	32
	Reference	33

Preface

比較分析 (グループ間での違いを比較する) の方法を紹介します。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。例えば、ある職業訓練プログラムが就業確率や就業後賃金に与える因果的効果を推定するためには、職業訓練プログラムへの参加者と非参加者の間で見られる就業状態や賃金の比較分析が求められます。「男女」間賃金格差を推定するためには、「男性」と「女性」の間で賃金を比較を行います。

あるグループ間で、変数の平均値 (“条件つき” 平均) を推定する方法を紹介します。R の実例では、中古マンションの取引データを用いて、物件の属性 (部屋の広さ、駅からの距離など) ごとに、平均取引価格 を推定します。

平均の推定値は、さまざまな実務で活用されています。中でも「の値を予測」という課題においては、代表的な予測値となっています。

伝統的には、OLS が母平均の推定方法として用いられてきました。近年では、OLS は母平均の仮想的な線型モデル (“補助線”) を推定する手法として、解釈できることが強調されています (Angrist and Pischke 2009; Aronow and Miller 2019)。モデルの定式に誤りがあったとしても、推定結果は常に” 解釈” できることがその理由です。

また近年では、機械学習の手法も積極的に活用されています。本ノートでは、OLS が” 研究者が設定した平均値のシンプルなモデル” を推定する方法として優れていることを強調します。もしより複雑なモデルを推定したい場合は、OLS はその有効性を失います。このような” 複雑なモデル” を推定する優れた方法として、LASSO を紹介します 1。

最後に平均値の特徴を捉えることに成功した予測モデルは、予測においても有効であることを論じます。

- (Imbens 2015; Chattopadhyay and Zubizarreta 2023)
- (Ben-Michael et al. 2021; Bruns-Smith et al. 2023; Chattopadhyay, Hase, and Zubizarreta 2020)
- (Van Der Laan and Rubin 2006; Chernozhukov et al. 2018, 2022)

第 1 章

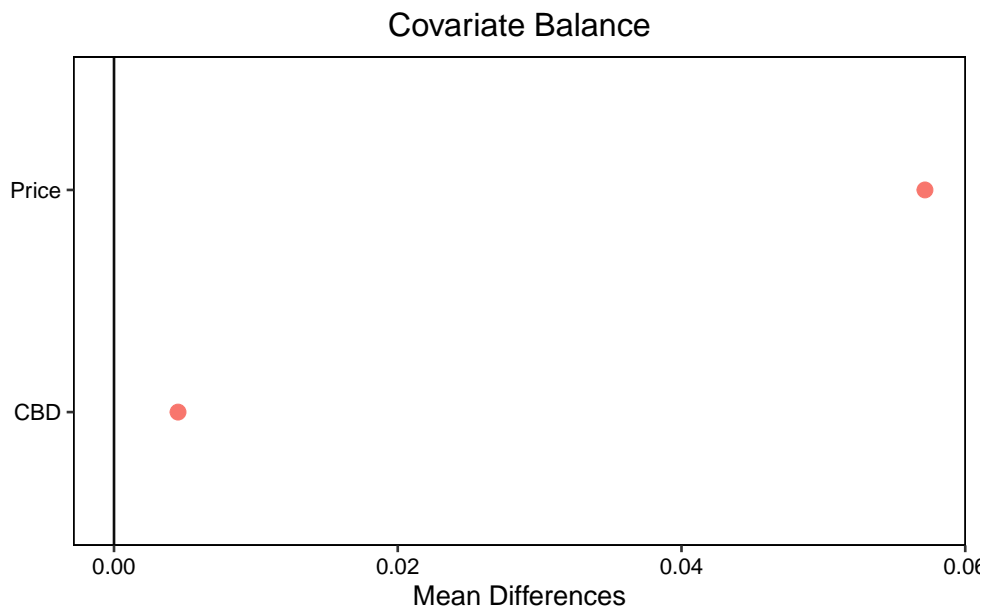
Balanced Comparison

バランス後の比較では、 D 間での X についての格差が解消された場合の、 Y についての差の推定を目指します。このような比較は、因果効果や格差の推定の肝となります。詳細な入門としては、[Chattopadhyay and Zubizarreta \(2024\)](#) などを参考にしてください。

1.1 実例

1.1.1 不動産市場の年次比較

例えば、2022 年と 2021 年の東京 23 区の中古マンション市場における、取引価格と立地 (中心 6 区かそれ以外か) について平均的な差は以下の通りです。



2022 年においては、平均取引価格が 2021 年に比べて上昇していますが、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 区の物件の方が取引価格が高い傾向が予想されるので、その分取引価格の上昇が”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。

1.1.2 合計特殊出生率

合計特殊出生率の国家間/時代間比較は、バランス後の比較の代表例です。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。比較的高齢の成人の比率が高まれば、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

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

シンプルな枠組みであり、大規模なデータが活用可能な状況では、有効だと考えられます。一方で、年齢以外の属性（教育歴、居住地等々）もバランスさせる場合、同じ属性を持つ事例数が少なくなり、適用が難しくなります。

1.1.3 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。合計特殊出生率と同様に、新規店比率のみをバランスさせるのであれば、非常に実践的な方法です。しかしながら他の属性、例えば客層の変化などもバランスさせたい場合は、事例数が不足する可能性が高くなります。

1.2 推定対象

1.2.1 バランス後の平均値

以上の推定対象は、一般に以下のように定義できます。

グループ d における Y の平均値は、一般に以下のように書き換えることができます。

$$\begin{aligned} & (d) \text{ における } Y \text{ の平均値} \\ &= \left\{ (x \ \& \ d) \text{ における } Y \text{ の平均値} \times (d) \text{ における } (x) \text{ の割合} \right\} \\ & \text{の } x \text{ についての総和} \end{aligned}$$

D 間での平均差を生み出す要因は、以下に分解できます。

- D の間での Y の平均値の違い
- D の間での X の分布の違い

バランス後の比較における推定対象は、 X の分布の違いを排除したバランス後の平均値の差として定義します。 d のバランス後の平均値は、

$$\begin{aligned} & (d) \text{ における } Y \text{ のバランス後平均値} \\ &= \left\{ (x \ \& \ d) \text{ における } Y \text{ の平均値} \times (x) \text{ へのウェイト} \right\} \\ & \text{の } x \text{ についての総和} \end{aligned}$$

バランス後の平均値の差は、 X 内での平均差の (x) へのウェイト を用いた集計値として書き換えることができます:

$$\begin{aligned} & \left\{ \left[(x \ \& \ D = 1) \text{ における } Y \text{ の平均値} - (x \ \& \ D = 0) \text{ における } Y \text{ の平均値} \right] \right. \\ & \left. \times (x) \text{ へのウェイト} \right\} \text{ の } x \text{ についての総和} \end{aligned}$$

$(X = x)$ へのウェイト は、原理的には研究者が設定できます。以下、最も代表的なウェイトである、母集団全体での $X = x$ の割合をウェイトとして、バランス後の平均値の推定を目指します。因果推論においては、母集団全体での割合を用いた集計値を平均効果と呼んでおります。

平均価格	CBD	TradeYear	事例割合
37.7	0	2021	0.784
60.5	1	2021	0.216
39.2	0	2022	0.779
64.8	1	2022	0.221

1.2.1.1 実例

例えば、立地と取引年ごとの平均取引価格は以下です。

2022/2021 年の平均取引価格差は、データ全体では 2.2 ですが、 $CBD = 1$ では $64.8 - 60.5 = 4.3$ 、 $CBD = 0$ では $39.2 - 37.7 = 1.4$ となります。よって立地をバランスさせた後の平均差は 4.3 と 1.4 の”平均値”となります。例えばデータ上での立地の割合、 $CBD=1$ について 0.22、 $CBD=0$ について 0.78、をウェイトとして用いるのであれば 2 がバランス後の平均差となります。

第2章

Balancing Weight

バランス後の平均値は、**Balancing weight** を用いた加重平均としても計算できます。バランス後の平均値を以下のように書き換えられることで、Balancing weight $\omega(x)$ は定義できます。

バランス後の平均値

$$(d) \text{ における } Y \text{ のバランス後平均値} = \left\{ (x \ \& \ d) \text{ における } Y \text{ の平均値} \right. \\ \left. \times \underbrace{\frac{x \text{ へのウェイト}}{d \text{ における } x \text{ の割合}}}_{\equiv \omega(d, x)} \times (d) \text{ における } (x) \text{ の割合} \right\} \\ \text{の } x \text{ についての総和}$$

バランス後の平均値は、 $D_i = d$ を満たす事例についての加重平均としても算出できます。

$$\frac{D = 1 \text{ の事例についての } \omega(d) \times Y \text{ の総和}}{D = 1 \text{ を満たす事例数}}$$

ただし N は全体の事例数、 $I(D_i = d)$ は事例 i の D の値が d であれば 1、それ以外であれば 0 となる変数です。よって $D_i = d$ を満たす事例について、Balancing weights を掛けたものの総和を計算し、 $D_i = d$ を満たす事例数で割ることで計算できます。

例えば、 $(X = x)$ へのウェイトとして $\text{CBD}=1$ が 0.22、 $\text{CBD}=0$ が 0.78 を用いるのであれば、Balancing Weight は以下のように算出できます。

2.1 仮定: Overlap

母集団全体での $X = x$ の割合をウェイトとした、バランス後の平均値を推定対象とするためには、**母集団**に対して以下を推定する必要があります。

平均価格	CBD	TradeYear	事例割合	Balancing Weight
37.7	0	2021	0.41	1.90
60.5	1	2021	0.11	2.00
39.2	0	2022	0.37	2.11
64.8	1	2022	0.11	2.00

! Overlap の仮定

- 全ての X の組み合わせについて、 $D = 1$ の事例も $D = 0$ の事例も、母集団上には両方存在する:

$$1 > E[D|X] > 0$$

ただし $E[D|X]$ は D の母平均 ($D = 1$ の割合)

Positivity が成り立っていない場合、 $D = 1$ または $D = 0$ しか存在しない X の組み合わせが存在することになります。結果、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。ここで関心となる母集団は、男女間での教育経験の分断が極めて大きく、大学卒以上の女性は存在しないとします。この場合、大学卒の女性割合は 0 であり、どのような Balancing weight $\omega(\text{大学卒})$ を用いたとしても、男性/女性の大学卒比率を揃えることは不可能です。言い換えるならば、大学卒の女性が存在しないため、大学卒内で男女間の賃金を比較できないため、バランス後の比較は不可能です。

@(有限標本では OK?)

((では?))、Positivity の不成立についての対処を論じます)

2.2 推定方法

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。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。

Exact matching や Stratified Estimation は、非常に直感的な推定方法ですが、 X の組み合わせが増えると、実行不可能です。例えば X に、両親の年収や資産などの連続変数が含まれている場合は、 X の組み合わせが非常に大きくなり、Balancing weights を計算することは事実上不可能となります。

この問題を解決するために、次節以降で紹介する、OLS や傾向スコアの逆数 (Inverse

probability weights) の活用が有用です。

2.3 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
 - matchit: Exact matching を含む多様な Matching を実装
 - * [manual](#)

2.3.1 準備

データを取得します。 D として、中心 6 区かそれ以外で、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 = mutate(  
  Data,  
  D = if_else(  
    LargeDistrict == "中心 6 区", 1, 0  
  )  
)
```

2.3.2 Balanced Weight

MatchIt パッケージ内の matchit 関数を用いて、Balanced weights を計算します。例えば立地別の平均取引価格とその信頼区間は、以下で計算できます。

```
Match = MatchIt::matchit(  
  D ~ Size + Tenure + StationDistance, # D ~ X を指定  
  Data, # 用いるデータの指定  
  method = "exact", # Balanced weight を計算するために、exact matching を実行  
  target = "ATE" # サンプル全体の X の分布をターゲット  
)
```

```
DataWeight = MatchIt::match.data(  
  Match,  
  drop.unmatched = FALSE # Balance weight が計算できない事例も含む
```

```
) # Balance weight を含んだデータを生成
```

```
Match # Balance weight の特徴を表示
```

```
A matchit object
```

```
- method: Exact matching
- number of obs.: 6378 (original), 1702 (matched)
- target estimand: ATT
- covariates: Size, Tenure, StationDistance
```

number of obs. において、元々の事例数 (6378) と balanced weight を計算できた事例数 (1702) を表示しています。事例が大きく減少しており、balanced weight を計算できない事例が多かったことを示しています。この理由は、Size, Tenure, StationDistance が完全に一致する事例が、 $D = 1$ または $D = 0$ のどちらかしか存在しない場合が多いためです。

Balanced weight が算出できた事例について、バランス後の平均差を計算すると以下となります。

```
lm(Price ~ D,
    DataWeight,
    weights = weights # Balancing weights を使用
)
```

Call:

```
lm(formula = Price ~ D, data = DataWeight, weights = weights)
```

Coefficients:

(Intercept)	D
36.05	12.55

単純比較の結果は以下であり、大きく異なることが確認できます。

```
lm(Price ~ D,
    DataWeight)
```

Call:

```
lm(formula = Price ~ D, data = DataWeight)
```

Coefficients:

(Intercept)	D
-------------	---

38.04

20.94

ただし今回のように、多くの事例が分析から除外されてしまった場合、単純比較とバランス後の平均差が乖離する理由は不明瞭です。 X をバランスさせることで平均差が変化した可能性があります、分析事例の限定も値を変化させます。このため次節以降の方法を用いて、極力分析事例を除外しない方法を用いることが望ましいです。

第 3 章

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、 X の分布を近似的にバランスさせることが有力です。このような方法は大きく、balancing approach と modelling approach に大別できます (Chattopadhyay, Hase, and Zubizarreta 2020)。本節では balancing approach の代表例である、分布を特徴づける代表的な値である平均値や分散などをバランスさせるアプローチを紹介します。

3.1 OLS

近年の研究により、線型モデルの OLS 推定は、Moment Balance を達成することが確認されています (Imbens 2015; Chattopadhyay and Zubizarreta 2023)。Chattopadhyay and Zubizarreta (2023) は、以下を証明しました。

! OLS の性質

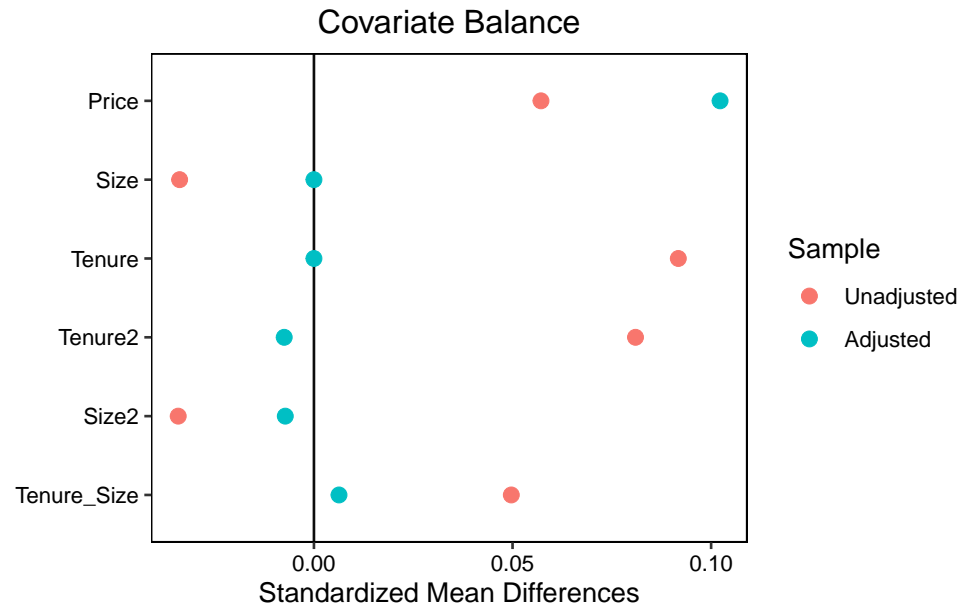
- $D = \{0, 1\}$

?@eq-balance は、 X の平均値を $D = \{0, 1\}$ 間で均質化していることを意味しています。?@eq-normal は、ウェイトとして総和を 1 に基準化しています。すなわち Step 1 において、平均値をバランスさせるウェイトの中から、最も分散が小さいものを選んでることを意味しています。Step 2 において、このウェイトを用いた平均差を計算しています。

ウェイトの分散は、最終的な推定結果の推定誤差に影響を与えます。一般に、ウェイトの分散が小さいと、推定誤差が削減される傾向があります。OLS は、平均値のバランスを達成するウェイトから分散が最小となるものを選ぶため、推定誤差が小さくなる傾向を持ちます。

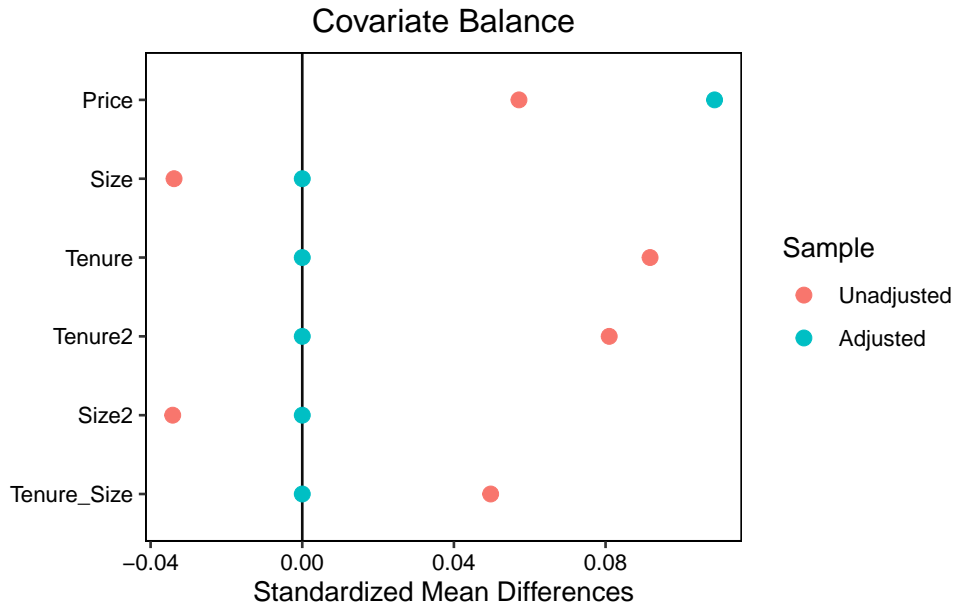
3.1.1 例

部屋の広さ (Size) と 築年数 (Tenure) をバランスさせた後に、2022/2021 年の平均取引価格差を推定します。 $Price \sim D + Size + Tenure$ を OLS 回帰すると、以下のようなバランスが達成されます。



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。結果、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 OLS を行ったとしてもバランスしていません。

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 $Price \sim D + Size + Tenure + Size2 + Tenure2 + TenureSize$ を OLS 推定します。結果、以下の図の通り、分散や共分散も Balance します。



3.1.2 Post selection

OLS においては、分布の特徴をどこまでバランスさせるのが問題となります。事例数が十分あれば、3 乗項などの高次項もバランスさせることは可能です。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。

このような問題に対して、OLS 推定を行う前に重要な変数のみを選択することが必要となります。変数選択を行う方法としては、Chernozhukov, Hansen, and Spindler (2015) が Post Double Selection というデータ主導の手法を提案しており、幅広く応用されています。Angrist and Frandsen (2022) は、変数選択について、より入門的な紹介を行っています。

Post Double Selection では、 X の中から、 Y または D について予測モデルを推定した際に使用される変数のみを使用し、OLS 推定を行います。予測モデルは、LASSO により推定されます。ここで“または”であることに注意してください。例えば Y の予測モデルからは排除された変数であったとしても、 D の予測モデルに利用されているのであれば、OLS 推定に採用されます。このような“慎重な”変数選択によって、信頼区間の計算可能性などの統計的性質を保証しています。

当該手法は [hdm package](#) を用いて実装できます。

```
Y = Data$Price # Outcome
D = Data$D # Treatment
X = select(
```

```

Data,
Size,
Size2,
Tenure,
Tenure2,
Tenure_Size) # Control

Fit = hdm::rlassoEffect(
  x = as.matrix(X),
  d = D,
  y = Y
) # Fit post double selection

summary(Fit) # Show results

```

```

[1] "Estimates and significance testing of the effect of target variables"
      Estimate. Std. Error t value Pr(>|t|)
d1      3.8506      0.3304   11.65  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

選択された変数は、以下のように表示できます。部屋の広さと築年数の平均値のみが、選択されたことが確認できます。

```
Fit$selection.index # Show selected X
```

Size	Size2	Tenure	Tenure2	Tenure_Size
TRUE	FALSE	TRUE	FALSE	FALSE

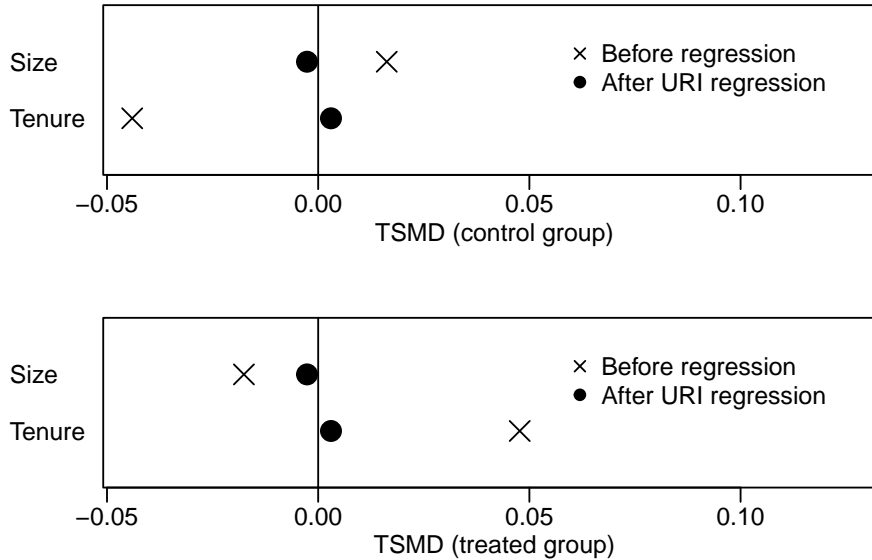
3.2 OLS の問題点と解決策

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?@eq-balance。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。

3.2.1 解釈の難しさ

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しいです。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。



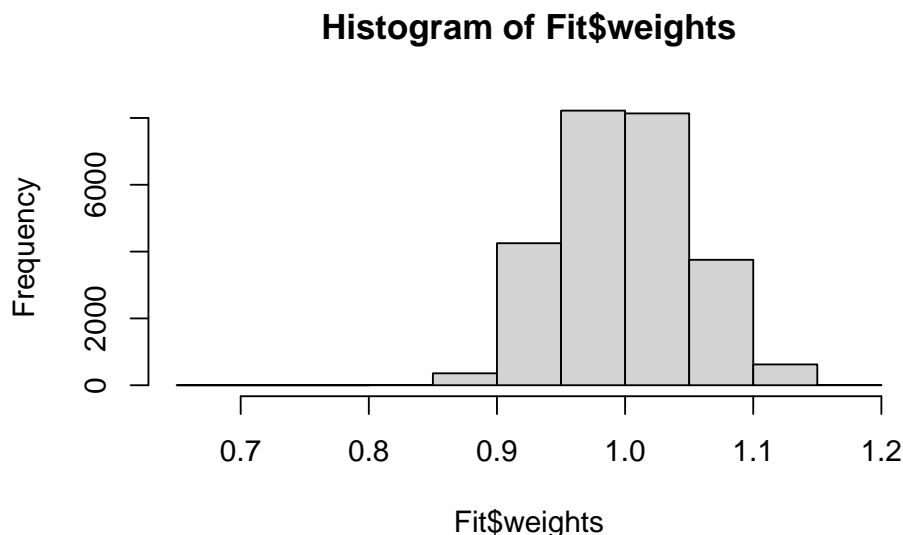
黒丸は OLS によるバランス後、ばつ印はバランス前の平均値を示しています。Control group は、 $D = 0$ (2021 年)、Treatment group は、 $D = 1$ (2022 年) の値です。0 線は、サンプル平均を示しています。

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 は長いことが確認できます。黒丸を見ると、OLS によるバランス後はどちらも 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。

3.2.2 負の荷重

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、ミスリーディングな推定結果をもたらす可能性があります。

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。

3.2.3 解決策

D と X の交差項を含めた以下のモデルを推定した、 $\beta_D, \dots, \beta_{DL}$ の平均値は、サンプル全体の X の平均値とバランス後の平均値を一致させる Weight を活用した平均の差と一致します (Chattopadhyay and Zubizarreta 2023)。

$$Y \sim D \times (\beta_D + \beta_{D1}X_1 + \dots + \beta_{DL}X_L) + \beta_0 + \beta_1X_1 + \dots + \beta_LX_L.$$

ただし負の weight は以前として生じる可能性があります。

負の Weight を発生させない方法としては、Entropy weights (Hainmueller 2012) や Stable weights (Zubizarreta 2015) が有力です。これらの手法では、サンプル平均との一致や正の値を取ることを条件として課した上で、weight を計算します。このため OLS が持つ問題点の多くを克服しており、より信頼できるバランス後の比較分析が可能です。これらの手法は [WeightIt package](#) (Greifer 2024) で容易に実装できます。

3.3 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](#)

– `marginalEffects`: 交差項を含めたモデルでの推定

* [Manual](#)

3.3.1 準備

データを取得します。 D として、中心6区かそれ以外で、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(
    LargeDistrict == "中心6区", 1, 0
  ) # 中心6区であれば1、それ以外であれば0
)
```

3.3.2 Balanced comparison by OLS

交差項を含む OLS により Balance させた推定結果は以下で導出できます。`Size`, `Tenure`, `StationDistance` の平均値、分散、共分散を、中心6区とそれ以外で一致させています

```
OLS = lm(
  Price ~
    D*(I(Size^2) + I(Tenure^2) + I(StationDistance^2) +
      (Size + Tenure + StationDistance)**2) + # DとXの交差項
    I(Size^2) + I(Tenure^2) + I(StationDistance^2) +
      (Size + Tenure + StationDistance)**2, # X,
  Data
)

marginalEffects::avg_comparisons(
  OLS, # 計算の元となるモデル
  variables = "D", # Dについて平均差を推定
  vcov = "HC4") # ロバストな信頼区間
```

Estimate	Std. Error	z	Pr(> z)	S	2.5 %	97.5 %
19.4	0.705	27.6	<0.001	554.2	18.1	20.8

Term: D

Type: response

Comparison: mean(1) - mean(0)

Columns: term, contrast, estimate, std.error, statistic, p.value, s.value, conf.l

バランスさせない平均差 (以下) と比べると、平均差が少し減少しています。

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

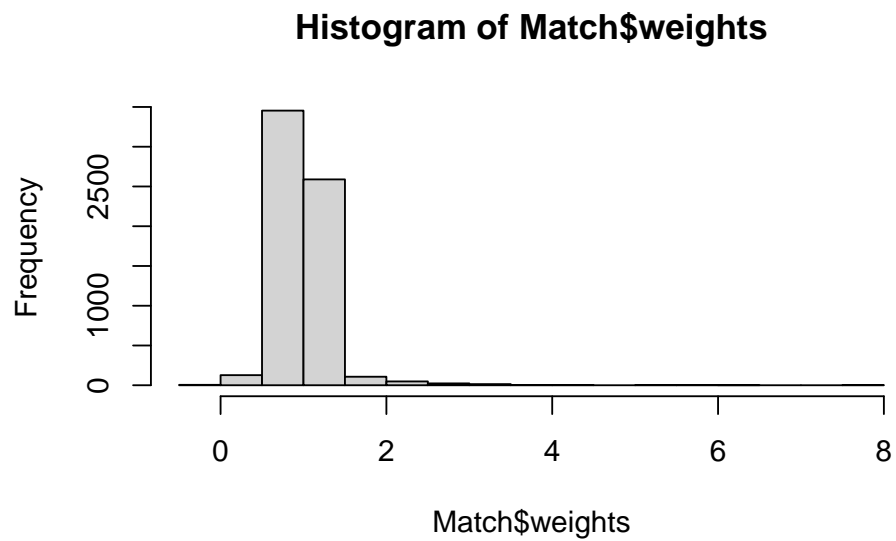
	Estimate	Std. Error	t value	Pr(> t)	CI Lower	CI Upper	DF
(Intercept)	38.03972	0.3182179	119.5399	0.000000e+00	37.41591	38.66354	6376
D	20.94057	1.2529064	16.7136	2.084367e-61	18.48446	23.39669	6376

3.3.2.1 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,
  method = "MRI" # D と X の交差項を導入
) # Weight の算出

hist(Match$weights) # Weight のヒストグラムを算出
```

少数ですが負の Weight が発生しています。

3.4 Reference

第 4 章

Inverse Probability Weight

Balancing weights を計算する別のアプローチとして、傾向スコア (Propensity score; Rosenbaum and Rubin (1983)) の活用が挙げられます。

! 傾向スコアの定義

- X 内で $D = d$ である事例の母集団における割合 ($p_d(X)$)

4.1 Inverse probability weighting

($X = x$) へのウェイトとして、 X の母分布をターゲットとするのであれば、Balancing weight $\omega(X)$ と傾向スコア $p_d(X)$ の間に以下の関係性が成り立ちます。

$$\omega(X) = \frac{D = d \text{ の割合}}{p_d(X)}$$

よってバランス後の平均値は以下のように書き換えられます。

$$\begin{aligned} & \frac{D = d \text{ について } (\omega(X) \times Y) \text{ の総和}}{D = d \text{ の事例数}} \\ &= \frac{D = d \text{ について } [(D = d \text{ の割合} / p_d(X)) \times Y] \text{ の総和}}{D = d \text{ の割合} \times \text{総事例数}} \\ &= \frac{D = d \text{ について } [(1/p_d(X)) \times Y] \text{ の総和}}{\text{総事例数}} \end{aligned}$$

よって、傾向スコアが推定できれば、その逆数 ($1/p_d(X)$) を掛ることでバランス後の平均値が推定できます。

4.2 傾向スコアの推定方法

傾向スコアの推定方法については、予測を目的とする要約 (**sec-Prediction**) の手法を応用することができます。例えば以下のモデルを、シンプルに OLS を用いて推定することも可能です。

$$D \sim \beta_0 + \beta_1 X_1 + ..$$

ただ多くの応用で、Logit や Probit など、予測値が $[0, 1]$ の範囲に収まることが保証される方法を用いることが多くなっています。

近年では、より柔軟な推定手法 (LASSO や RandomForest, Boosting などの機械学習の手法) を用いることが増えています。既存の機械学習の手法が容易に応用できる点は、傾向スコアを用いることの大きな利点となります。

4.3 問題点

Inverse probability weighting を用いたバランス後の平均値推定は、傾向スコアの推定精度に決定的に依存します。シンプルなモデルを Logit など推定した場合、モデル定式化の誤りが深刻となりえます。機械学習等のより柔軟な推定手法を用いれば、モデル定式化の誤りは減少します。しかしながら、依然として推定精度は不十分な場合が多いことが知られています (Chernozhukov et al. 2018)^{*1}。

以上の問題を克服するために、次節では Y の予測モデルも併用する推定方法 (Augmented inverse probability weighting) を紹介します。

4.4 他の方法

ここまで Balancing weights を推定する方法として、OLS や傾向スコアを用いた方法を紹介しました。他の有力な方法として、Imai and Ratkovic (2014) は、 X のバランスと $p_d(X)$ を極力両立するように推定する方法を提案しています。Iacus, King, and Porro (2012), Huling and Mak (2024) では、 X の分布の距離 (Energy distance) を最小化するように weight を推定します。

4.5 Reference

^{*1} この問題は収束の遅さとして知られています

第 5 章

Augmentation

5.1 動機

バランス後の平均差の推定結果は、balancing weights の推定精度に強く依存します。例えば傾向スコアの逆数 (`?@sec-WhatIfPF`) を用いた推定では、傾向スコア (D の母平均) の高い精度での推定が必要ですが、実践は困難です。 D のモデルを、Logit などの伝統的な方法で推定する場合、研究者が事前に設定するモデルの定式化に推定結果が強く依存します。LASSO などより柔軟な推定方法を用いると、モデルへの依存度は低下させられますが、データへの依存度は上昇してしまいます。このためより推定結果が不安定になりやすく、例えば信頼区間の計算 (`?@sec-SimpleComparison`) などが困難になります。

Balancing weights の推定結果への依存度を下げる方法として、 Y のモデルを用いた補強 (Augment) が提案されています (Ben-Michael et al. 2021)。このような手法は、特に機械学習を用いた推定との相性が良く、近年改めて注目されています (Chernozhukov et al. 2018)。以下では補強を行なった推定値として、Augmented Inverse Probability Weight (Robins and Rotnitzky 1995) を紹介します。

5.2 Augmented Inverse Probability Weighting

! Augmented Inverse Probability Weighting の定義

- バランス後の平均値は、以下の通り推定できます ただし $f(D = d, X)$ は Y の予測値, $p_d(X)$ は傾向スコアの推定値です。

Augmented inverse probability weighting はいくつかの利点を持ちます。実践上の大きな利点は、 Y や D の予測モデル ($f(D, X), p_d(X)$) 推定に機械学習を利用できることです。

- Y や D の予測値 $f(D, X)$ や $p_d(X)$ は、母平均の**一致推定量**である必要があります

す。一致推定量とは、無限大の事例数で推定できた場合、予測値と母平均が一致する推定値のことです。この要求は、機械学習を活用することで比較的緩やかな仮定のもとで保証されます。

- 機械学習の弱点である、モデルがデータにより依存してしまふ、という問題についても、その弊害を軽減できます。限られた事例数のもとで、 $f(D, X)$ や $p_d(X)$ の推定誤差が十分に削減できなかったとしても、それらを用いて推定されるバランス後の平均値は、より高い精度で推定でき、信頼区間も計算できます。これは Y と D の予測モデルの内、**どちらか一方の推定精度が不十分であったとしても**、バランス後の平均値の推定結果が大きな影響を受けないように設計されているためです。

$f(D, X), p_d(X)$ を OLS や Logit、LASSO など推定する場合、Augmented Inverse Probability Weight を用いた推定値は、以下のような性質を持ちます。

! Augmented Inverse Probability Weighting の性質

- 以下の仮定のもとで、Augmented Inverse Probability Weight を用いた推定値について、信頼区間が計算できる

仮定 1. Overlap が成り立つ

仮定 2. $f(D, X), p_d(X)$ の推定精度は、ある程度高い

仮定 3. $f(D, X), p_d(X)$ は、交差推定で推定されている

仮定 1 (Positivity) は、バランス後の比較の根本的な仮定であり、推定方法に拘らず必要です。

仮定 2 が要求する推定精度を保証することは、厳密には困難です。このため実用においては、さまざまな推定方法 (OLS, LASSO など) の予測精度をデータ分割 (Cross Prediction) などで検証し、最善の方法を用いる必要があります。

仮定 3 が要求する交差推定は、以下の手順で実行できます。

5.2.1 交差推定

ある事例 i について、予測値を算出する予測モデルは、当該事例 i を用いずに推定される必要があります。このような推定を効率的に行う方法として、**交差推定 (Cross fitting)** が頻繁に用いられます。

! 交差推定

1. 事例をランダムにいくつかのグループ (第 1, ..., K グループ)
2. 第 1 グループに対する予測値を算出する。その際に用いる予測モデルは、第

- 1 グループ以外的事例を用いて推定する。
3. 第2グループについて、手順2を繰り返し、全事例に対して予測値を算出する。
4. 第3グループ以降についても、順次、手順2を繰り返す

5.3 R による実践例

Ahrens et al. (2024) の手法を実装します。

- 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: Augmented Inverse Probability Weighting の実装

5.3.1 準備

データを取得します。 D として、中心6区かそれ以外で、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(
    LargeDistrict == "中心6区", 1, 0
  ) # 中心6区であれば1、それ以外であれば0
) # Dの定義

Y = Data$Price
D = Data$D
X = model.matrix(
  ~ 0 + poly(Size, 2) + poly(Tenure, 2) + poly(StationDistance, 2),
  Data
)
X = scale(X)
```

5.3.2 Balanced comparison by OLS

```
ATE = ddml::ddml_ate(
  y = Y,
  D = D,
  X = X,
  learners = list(
    list(fun = ddml::mdl_glmnet)
  ),
  shortstack = TRUE,
  sample_folds = 2,
  silent = TRUE
)

ATE$oos_pred$ED_X |> summary()
```

```
      nnls
Min.   :-0.01363
1st Qu.: 0.15098
Median : 0.22598
Mean    : 0.22209
3rd Qu.: 0.29186
Max.    : 0.52154
```

```
ATE |> summary()
```

ATE estimation results:

	Estimate	Std. Error	t value	Pr(> t)
nnls	20.4	0.721	28.3	1.08e-175

5.4 Reference

Reference

- Ahrens, Achim, Christian B Hansen, Mark E Schaffer, and Thomas Wiemann. 2024. “Model Averaging and Double Machine Learning.” *arXiv Preprint arXiv:2401.01645*.
- Angrist, Joshua D, and Brigham Frandsen. 2022. “Machine Labor.” *Journal of Labor Economics* 40 (S1): S97–140.
- 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. “Causation, Comparison, and Regression.” *Harvard Data Science Review* 6 (1).
- 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. <https://doi.org/10.1111/ectj.12097>.
- Chernozhukov, Victor, Juan Carlos Escanciano, Hidehiko Ichimura, Whitney K Newey, and James M Robins. 2022. “Locally Robust Semiparametric Estimation.” *Econometrica* 90 (4): 1501–35.
- Chernozhukov, Victor, Christian Hansen, and Martin Spindler. 2015. “Valid Post-Selection and Post-Regularization Inference: An Elementary, General Approach.” *Annu. Rev. Econ.* 7 (1): 649–88.
- Greifer, Noah. 2024. *WeightIt: Weighting for Covariate Balance in Observational Studies*. <https://ngreifer.github.io/WeightIt/>.
- 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.
- Huling, Jared D, and Simon Mak. 2024. “Energy Balancing of Covariate Distributions.” *Journal of Causal Inference* 12 (1): 20220029.
- Iacus, Stefano M, Gary King, and Giuseppe Porro. 2012. “Causal Inference Without Balance Checking: Coarsened Exact Matching.” *Political Analysis* 20 (1): 1–24.
- Imai, Kosuke, and Marc Ratkovic. 2014. “Covariate Balancing Propensity Score.” *Journal of the Royal Statistical Society Series B: Statistical Methodology* 76 (1): 243–63.
- Imbens, Guido W. 2015. “Matching Methods in Practice: Three Examples.” *Journal of Human Resources* 50 (2): 373–419.
- Robins, James M., and Andrea Rotnitzky. 1995. “Semiparametric Efficiency in Multivariate Regression Models with Missing Data.” *Journal of the American Statistical Association* 90 (429): 122129. <https://doi.org/10.1080/01621459.1995.10476494>.
- Rosenbaum, Paul R, and Donald B Rubin. 1983. “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” *Biometrika* 70 (1): 41–55.
- Stuart, Elizabeth A, Gary King, Kosuke Imai, and Daniel Ho. 2011. “MatchIt: Non-parametric Preprocessing for Parametric Causal Inference.” *Journal of Statistical Software*.
- Van Der Laan, Mark J, and Daniel Rubin. 2006. “Targeted Maximum Likelihood Learning.” *The International Journal of Biostatistics* 2 (1).
- Wager, Stefan. 2024. “Causal Inference: A Statistical Learning Approach.” preparation.
- 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.