

2022年3月27日
第24回春の合宿セミナー（日本行動計量学会）
（統計的因果推論入門）

講義6 傾向スコアの導入 と

傾向スコアマッチング

長崎大学 情報データ科学部 准教授

高橋 将宜

博士（理工学）

m-takahashi@nagasaki-u.ac.jp

概要

- バランシングスコア
 - 傾向スコア
 - 傾向スコアのモデル化
 - マッチング
 - 復元マッチングと非復元マッチング
 - マッチング方法
 - 傾向スコアマッチングの実践
 - 傾向スコアをマッチングに使うべきでない？
- 教科書
Ch.10
- 教科書
Ch.11



balancing score

バランシングスコア (balancing score)

- 観測される共変量 X の関数 $b(X)$
 - $b(X)$ が与えられたときの X の条件付き分布が、処置群と統制群において同じとなる関数
 - 傾向スコアはバランシングスコア的一种

Rコード : data10aの読み込み (教科書p.137)

- ❑ `data10a <- read.csv(file.choose())`
- ❑ `attach(data10a)`
- ❑ `data10a`

```
> data10a
  y0t ylt tl  y xl
1   63  74  0 63 66
2   55  70  0 55 70
3   59  69  0 59 70
4   71  77  1 77 73
5   73  78  0 73 73
6   66  77  0 66 74
7   69  79  0 69 75
8   70  81  1 81 77
9   62  77  1 77 77
10  72  81  0 72 78
11  65  79  0 65 78
12  68  81  1 81 78
13  69  73  1 73 79
14  79  93  1 93 83
15  75  85  1 85 84
16  80  91  1 91 87
17  75  89  0 75 88
18  82  85  1 85 91
19  91  97  1 97 92
20  74  81  1 81 92
```

data10aについて

- 20人の被験者
- 潜在的結果変数： y_{0t} と y_{1t}
 - y_{0t} と y_{1t} は、実際の場面では観測されないが、解説のために掲載している情報
- 処置： t_1
- 観測結果： y
- 共変量： x_1

```
> data10a
      y0t  y1t  t1   y  x1
1      63   74   0  63  66
2      55   70   0  55  70
3      59   69   0  59  70
4      71   77   1  77  73
5      73   78   0  73  73
6      66   77   0  66  74
7      69   79   0  69  75
8      70   81   1  81  77
9      62   77   1  77  77
10     72   81   0  72  78
11     65   79   0  65  78
12     68   81   1  81  78
13     69   73   1  73  79
14     79   93   1  93  83
15     75   85   1  85  84
16     80   91   1  91  87
17     75   89   0  75  88
18     82   85   1  85  91
19     91   97   1  97  92
20     74   81   1  81  92
```

処置効果

- $\text{mean}(y_{1t}) - \text{mean}(y_{0t})$
- $\text{mean}(y_{1t}[t_1 == 1]) - \text{mean}(y_{0t}[t_1 == 1])$
- $\text{mean}(y[t_1 == 1]) - \text{mean}(y[t_1 == 0])$

> `mean(y1t) - mean(y0t)` 平均処置効果（ATE）の真値

[1] 9.95

> `mean(y1t[t1==1]) - mean(y0t[t1==1])` 処置群の平均処置効果（ATT）の真値

[1] 9.090909

> `mean(y[t1==1]) - mean(y[t1==0])` ナイーブな推定値

[1] 17.39394

交絡因子

- このデータは、**共変量x1**の値に応じて**昇順**で並んでいる
- データをよく見てみると、**共変量x1**の値が大きくなるにつれて、**t1=1**が増えている
- したがって、交絡が発生しており、**共変量x1**が**交絡因子**である

```
> data10a
```

| | y0t | y1t | t1 | y | x1 |
|----|-----|-----|----|----|----|
| 1 | 63 | 74 | 0 | 63 | 66 |
| 2 | 55 | 70 | 0 | 55 | 70 |
| 3 | 59 | 69 | 0 | 59 | 70 |
| 4 | 71 | 77 | 1 | 77 | 73 |
| 5 | 73 | 78 | 0 | 73 | 73 |
| 6 | 66 | 77 | 0 | 66 | 74 |
| 7 | 69 | 79 | 0 | 69 | 75 |
| 8 | 70 | 81 | 1 | 81 | 77 |
| 9 | 62 | 77 | 1 | 77 | 77 |
| 10 | 72 | 81 | 0 | 72 | 78 |
| 11 | 65 | 79 | 0 | 65 | 78 |
| 12 | 68 | 81 | 1 | 81 | 78 |
| 13 | 69 | 73 | 1 | 73 | 79 |
| 14 | 79 | 93 | 1 | 93 | 83 |
| 15 | 75 | 85 | 1 | 85 | 84 |
| 16 | 80 | 91 | 1 | 91 | 87 |
| 17 | 75 | 89 | 0 | 75 | 88 |
| 18 | 82 | 85 | 1 | 85 | 91 |
| 19 | 91 | 97 | 1 | 97 | 92 |
| 20 | 74 | 81 | 1 | 81 | 92 |

処置群と統制群の分布

- `summary(x1[t1==0])`
- `summary(x1[t1==1])`

```
> summary(x1[t1==0])
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 66.00  70.00   74.00   74.67  78.00   88.00

> summary(x1[t1==1])
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 73.0   77.5   83.0   83.0   89.0   92.0
```

処置群 ($t1==1$) と統制群 ($t1==0$)
におけるx1の分布は異なっている

最も自明なバランシングスコア

- バランシングスコア $b(X)$ を使って, 共変量 x_1 の分布が処置群と統制群で同じになるようにする
- 最も自明なバランシングスコア
 - 共変量 X そのもの

共変量x1の値

- 「2と3」
- 「4と5」
- 「8と9」
- 「10と11と12」
- 「19と20」
 - 以上が同一の値となっている

```
> data10a
```

| | y0t | y1t | t1 | y | x1 |
|----|-----|-----|----|----|----|
| 1 | 63 | 74 | 0 | 63 | 66 |
| 2 | 55 | 70 | 0 | 55 | 70 |
| 3 | 59 | 69 | 0 | 59 | 70 |
| 4 | 71 | 77 | 1 | 77 | 73 |
| 5 | 73 | 78 | 0 | 73 | 73 |
| 6 | 66 | 77 | 0 | 66 | 74 |
| 7 | 69 | 79 | 0 | 69 | 75 |
| 8 | 70 | 81 | 1 | 81 | 77 |
| 9 | 62 | 77 | 1 | 77 | 77 |
| 10 | 72 | 81 | 0 | 72 | 78 |
| 11 | 65 | 79 | 0 | 65 | 78 |
| 12 | 68 | 81 | 1 | 81 | 78 |
| 13 | 69 | 73 | 1 | 73 | 79 |
| 14 | 79 | 93 | 1 | 93 | 83 |
| 15 | 75 | 85 | 1 | 85 | 84 |
| 16 | 80 | 91 | 1 | 91 | 87 |
| 17 | 75 | 89 | 0 | 75 | 88 |
| 18 | 82 | 85 | 1 | 85 | 91 |
| 19 | 91 | 97 | 1 | 97 | 92 |
| 20 | 74 | 81 | 1 | 81 | 92 |

処置群と統制群に分かれている組

- 「4と5」
- 「11と12」
- 「4と5」を事実上の同一人物と見なす
- 「11と12」を事実上の同一人物と見なす

```
> data10a
      y0t  y1t  t1   y  x1
1      63   74   0  63  66
2      55   70   0  55  70
3      59   69   0  59  70
4      71   77   1  77  73
5      73   78   0  73  73
6      66   77   0  66  74
7      69   79   0  69  75
8      70   81   1  81  77
9      62   77   1  77  77
10     72   81   0  72  78
11     65   79   0  65  78
12     68   81   1  81  78
13     69   73   1  73  79
14     79   93   1  93  83
15     75   85   1  85  84
16     80   91   1  91  87
17     75   89   0  75  88
18     82   85   1  85  91
19     91   97   1  97  92
20     74   81   1  81  92
```

処置群と統制群に分かれている組

- 処置群（処置=1）
 - 被験者4と12
 - x1の平均値
 - $(73+78)/2 = 75.5$
- 統制群（処置=0）
 - 被験者5と11
 - x1の平均値
 - $(73+78)/2 = 75.5$
- 標準偏差も3.54で同一

処置群と統制群の間で共変量の分布が同じであることを、**共変量のバランスが取れている**という

```
> data10a
      y0t  y1t  t1  y  x1
1      63   74   0 63  66
2      55   70   0 55  70
3      59   69   0 59  70
4      71   77   1 77  73
5      73   78   0 73  73
6      66   77   0 66  74
7      69   79   0 69  75
8      70   81   1 81  77
9      62   77   1 77  77
10     72   81   0 72  78
11     65   79   0 65  78
12     68   81   1 81  78
13     69   73   1 73  79
14     79   93   1 93  83
15     75   85   1 85  84
16     80   91   1 91  87
17     75   89   0 75  88
18     82   85   1 85  91
19     91   97   1 97  92
20     74   81   1 81  92
```

処置群と統制群に分かれている組

- 処置群における「結果」の平均

- $(77+81)/2 = 79$

- 統制群における「結果」の平均

- $(73+65)/2 = 69$

- 処置効果

- $79-69 = 10$ と推定

- ATEの真値

- 9.95

- ATTの真値

- 9.09

```
> data10a
      y0t  y1t  t1  y  x1
1      63   74   0 63  66
2      55   70   0 55  70
3      59   69   0 59  70
4      71   77   1 77  73
5      73   78   0 73  73
6      66   77   0 66  74
7      69   79   0 69  75
8      70   81   1 81  77
9      62   77   1 77  77
10     72   81   0 72  78
11     65   79   0 65  78
12     68   81   1 81  78
13     69   73   1 73  79
14     79   93   1 93  83
15     75   85   1 85  84
16     80   91   1 91  87
17     75   89   0 75  88
18     82   85   1 85  91
19     91   97   1 97  92
20     74   81   1 81  92
```

最も細かなバランシングスコア

□ 共変量 X

- 最も細かなバランシングスコア（finest balancing score）という
- 共変量 X そのものを比べているので、最も細かなレベルで共変量の違いを見ているのである

傾向スコア

Rコード : data10bの読み込み (教科書p.139)

- ❑ data10b <- read.csv(file.choose())
- ❑ attach(data10b)
- ❑ data10b

```
> data10b
  y0t ylt tl  y x1  x2  x3  ps1
1   63  74  0 63 66  76  75 0.12
2   55  70  0 55 70  75  55 0.05
3   59  69  0 59 70  60  73 0.19
4   71  77  1 77 73  76  82 0.43
5   73  78  0 73 73  79  78 0.35
6   66  77  0 66 74  72  79 0.41
7   69  79  0 69 75  78  73 0.34
8   70  81  1 81 77  75  83 0.62
9   62  77  1 77 77  66  77 0.50
10  72  81  0 72 78  76  74 0.48
11  65  79  0 65 78  79  64 0.28
12  68  81  1 81 78  83  79 0.57
13  69  73  1 73 79  74  60 0.26
14  79  93  1 93 83  89  88 0.86
15  75  85  1 85 84  81  90 0.90
16  80  91  1 91 87  82  86 0.91
17  75  89  0 75 88  78  82 0.90
18  82  85  1 85 91  84  85 0.95
19  91  97  1 97 92 102 101 0.99
20  74  81  1 81 92  77  73 0.90
```

ps1は傾向スコアだが、
これについては後述する。

共変量Xは3変量

処置群と統制群の分布

□ `summary(x1[t1==0])`

```
> summary(x1[t1==0])
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
66.00   70.00   74.00   74.67   78.00   88.00
```

□ `summary(x1[t1==1])`

```
> summary(x1[t1==1])
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
73.0    77.5    83.0    83.0    89.0    92.0
```

□ `summary(x2[t1==0])`

```
> summary(x2[t1==0])
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
60.00   75.00   76.00   74.78   78.00   79.00
```

□ `summary(x2[t1==1])`

```
> summary(x2[t1==1])
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
66.00   75.50   81.00   80.82   83.50  102.00
```

□ `summary(x3[t1==0])`

```
> summary(x3[t1==0])
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
55.00   73.00   74.00   72.56   78.00   82.00
```

□ `summary(x3[t1==1])`

```
> summary(x3[t1==1])
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
60.00   78.00   83.00   82.18   87.00  101.00
```

- 処置群と統制群の間に系統的な差がある
- このまま結果変数yを処置群と統制群で比較することは適切な因果推論ではない

傾向スコアの定義

$$e(X) = Pr(T_i = 1|X)$$

- 共変量 X が与えられたとき，処置に割付けられる確率
- 傾向スコア $e(X)$ は，共変量 X を条件としたときに， $T_i = 1$ となる確率

最も粗いバランシングスコア

- 傾向スコアは最も粗いバランシングスコア
 - 英語では, coarsest balancing scoreという
- 最も細かなバランシングスコア
 - 多変量の X を縮約せずにそのまま使ったので, 共変量 X は最も細かなバランシングスコアだった
- 最も粗いバランシングスコア
 - 多変量の X を単変量の $e(X)$ に縮約するため, 傾向スコアは最も粗い
 - 最も単純といってもよい
 - その結果, どうすれば共変量のバランシングが達成できるか, 一目瞭然であり, 統計的因果推論において重宝されている

data08bを傾向スコアで並び替え

| 被験者 | 処置 | 結果 | x1 | x2 | x3 | ps1 |
|-----|----|----|----|-----|-----|------|
| 2 | 0 | 55 | 70 | 75 | 55 | 0.05 |
| 1 | 0 | 63 | 66 | 76 | 75 | 0.12 |
| 3 | 0 | 59 | 70 | 60 | 73 | 0.19 |
| 13 | 1 | 73 | 79 | 74 | 60 | 0.26 |
| 11 | 0 | 65 | 78 | 79 | 64 | 0.28 |
| 7 | 0 | 69 | 75 | 78 | 73 | 0.34 |
| 5 | 0 | 73 | 73 | 79 | 78 | 0.35 |
| 6 | 0 | 66 | 74 | 72 | 79 | 0.41 |
| 4 | 1 | 77 | 73 | 76 | 82 | 0.43 |
| 10 | 0 | 72 | 78 | 76 | 74 | 0.48 |
| 9 | 1 | 77 | 77 | 66 | 77 | 0.50 |
| 12 | 1 | 81 | 78 | 83 | 79 | 0.57 |
| 8 | 1 | 81 | 77 | 75 | 83 | 0.62 |
| 14 | 1 | 93 | 83 | 89 | 88 | 0.86 |
| 15 | 1 | 85 | 84 | 81 | 90 | 0.90 |
| 17 | 0 | 75 | 88 | 78 | 82 | 0.90 |
| 20 | 1 | 81 | 92 | 77 | 73 | 0.90 |
| 16 | 1 | 91 | 87 | 82 | 86 | 0.91 |
| 18 | 1 | 85 | 91 | 84 | 85 | 0.95 |
| 19 | 1 | 97 | 92 | 102 | 101 | 0.99 |

- 傾向スコアをどのようにして計算するかは後述する.
- data08bのps1（傾向スコア）の値が似ている個体について、処置群と統制群に分けてみよう.
- このデータは、傾向スコアの値に応じて昇順で並べ替えてある.
- どれぐらいの差までを似ていると判断するかについては後述するが、ここでは直感的にだいたい同じぐらいの数字と考えよう.

傾向スコアでペアにする

| 被験者 | 処置 | 結果 | x1 | x2 | x3 | ps1 |
|-----|----|----|----|-----|-----|------|
| 2 | 0 | 55 | 70 | 75 | 55 | 0.05 |
| 1 | 0 | 63 | 66 | 76 | 75 | 0.12 |
| 3 | 0 | 59 | 70 | 60 | 73 | 0.19 |
| 13 | 1 | 73 | 79 | 74 | 60 | 0.26 |
| 11 | 0 | 65 | 78 | 79 | 64 | 0.28 |
| 7 | 0 | 69 | 75 | 78 | 73 | 0.34 |
| 5 | 0 | 73 | 73 | 79 | 78 | 0.35 |
| 6 | 0 | 66 | 74 | 72 | 79 | 0.41 |
| 4 | 1 | 77 | 73 | 76 | 82 | 0.43 |
| 10 | 0 | 72 | 78 | 76 | 74 | 0.48 |
| 9 | 1 | 77 | 77 | 66 | 77 | 0.50 |
| 12 | 1 | 81 | 78 | 83 | 79 | 0.57 |
| 8 | 1 | 81 | 77 | 75 | 83 | 0.62 |
| 14 | 1 | 93 | 83 | 89 | 88 | 0.86 |
| 15 | 1 | 85 | 84 | 81 | 90 | 0.90 |
| 17 | 0 | 75 | 88 | 78 | 82 | 0.90 |
| 20 | 1 | 81 | 92 | 77 | 73 | 0.90 |
| 16 | 1 | 91 | 87 | 82 | 86 | 0.91 |
| 18 | 1 | 85 | 91 | 84 | 85 | 0.95 |
| 19 | 1 | 97 | 92 | 102 | 101 | 0.99 |

- 赤字のps1が似ているペアと判断したとしよう。
- ps1が0.34と0.35は似ている値だが、被験者7と被験者5の処置は同じ0なので、これら2つは比較対象となるペアとして選ぶことができない。
- 処置が異なる集団において、傾向スコアの近い値をペアにする。
- 傾向スコアに関して、処置群と統制群の両方に観測値があることを**共通部分** (common support) という。

処置群と統制群に整理

| 被験者 | 処置 | 結果 | x1 | x2 | x3 | ps1 |
|-----|----|----|----|----|----|------|
| 11 | 0 | 65 | 78 | 79 | 64 | 0.28 |
| 6 | 0 | 66 | 74 | 72 | 79 | 0.41 |
| 10 | 0 | 72 | 78 | 76 | 74 | 0.48 |
| 17 | 0 | 75 | 88 | 78 | 82 | 0.90 |
| 13 | 1 | 73 | 79 | 74 | 60 | 0.26 |
| 4 | 1 | 77 | 73 | 76 | 82 | 0.43 |
| 9 | 1 | 77 | 77 | 66 | 77 | 0.50 |
| 15 | 1 | 85 | 84 | 81 | 90 | 0.90 |

統制群（処置＝0）

被験者11・6・10・17

処置群（処置＝1）

被験者13・4・9・15

共変量のバランシング

| 被験者 | 処置 | 結果 | x1 | x2 | x3 | ps1 |
|-----|----|----|----|----|----|------|
| 11 | 0 | 65 | 78 | 79 | 64 | 0.28 |
| 6 | 0 | 66 | 74 | 72 | 79 | 0.41 |
| 10 | 0 | 72 | 78 | 76 | 74 | 0.48 |
| 17 | 0 | 75 | 88 | 78 | 82 | 0.90 |
| 13 | 1 | 73 | 79 | 74 | 60 | 0.26 |
| 4 | 1 | 77 | 73 | 76 | 82 | 0.43 |
| 9 | 1 | 77 | 77 | 66 | 77 | 0.50 |
| 15 | 1 | 85 | 84 | 81 | 90 | 0.90 |

統制群（処置=0）

被験者11・6・10・17

処置群（処置=1）

被験者13・4・9・15

x1, x2, x3の分布が処置群と統制群で同じになったはずである。
→2つの集団を事実上の同一集団と考える。

2群の結果の比較が可能になる

| 被験者 | 処置 | 結果 | x1 | x2 | x3 | ps1 |
|-----|----|----|----|----|----|------|
| 11 | 0 | 65 | 78 | 79 | 64 | 0.28 |
| 6 | 0 | 66 | 74 | 72 | 79 | 0.41 |
| 10 | 0 | 72 | 78 | 76 | 74 | 0.48 |
| 17 | 0 | 75 | 88 | 78 | 82 | 0.90 |
| 13 | 1 | 73 | 79 | 74 | 60 | 0.26 |
| 4 | 1 | 77 | 73 | 76 | 82 | 0.43 |
| 9 | 1 | 77 | 77 | 66 | 77 | 0.50 |
| 15 | 1 | 85 | 84 | 81 | 90 | 0.90 |

統制群（処置＝0）

被験者11・6・10・17

処置群（処置＝1）

被験者13・4・9・15

x1, x2, x3の分布が処置群と統制群で同じになったはずである。
→2つの集団を事実上の同一集団と考える。

Rコード : data10cの読み込み (教科書p.141)

- ❑ data10c <- read.csv(file.choose())
- ❑ attach(data10c)
- ❑ data10c

```
> data10c
  y0t ylt t1  y x1 x2 x3  ps1
1  65  79  0 65 78 79 64 0.28
2  66  77  0 66 74 72 79 0.41
3  72  81  0 72 78 76 74 0.48
4  75  89  0 75 88 78 82 0.90
5  69  73  1 73 79 74 60 0.26
6  71  77  1 77 73 76 82 0.43
7  62  77  1 77 77 66 77 0.50
8  75  85  1 85 84 81 90 0.90
```

このデータは、被験者11・6・10・17・13・4・9・15の8人だけに絞ったものである。

処置群と統制群の分布

| | |
|-----------------------------------|---|
| □ <code>summary(x1[t1==0])</code> | > <code>summary(x1[t1==0])</code> Min. 1st Qu. Median Mean 3rd Qu. Max. 74.0 77.0 78.0 79.5 80.5 88.0 |
| □ <code>summary(x1[t1==1])</code> | > <code>summary(x1[t1==1])</code> Min. 1st Qu. Median Mean 3rd Qu. Max. 73.00 76.00 78.00 78.25 80.25 84.00 |
| □ <code>summary(x2[t1==0])</code> | > <code>summary(x2[t1==0])</code> Min. 1st Qu. Median Mean 3rd Qu. Max. 72.00 75.00 77.00 76.25 78.25 79.00 |
| □ <code>summary(x2[t1==1])</code> | > <code>summary(x2[t1==1])</code> Min. 1st Qu. Median Mean 3rd Qu. Max. 66.00 72.00 75.00 74.25 77.25 81.00 |
| □ <code>summary(x3[t1==0])</code> | > <code>summary(x3[t1==0])</code> Min. 1st Qu. Median Mean 3rd Qu. Max. 64.00 71.50 76.50 74.75 79.75 82.00 |
| □ <code>summary(x3[t1==1])</code> | > <code>summary(x3[t1==1])</code> Min. 1st Qu. Median Mean 3rd Qu. Max. 60.00 72.75 79.50 77.25 84.00 90.00 |

- 傾向スコアを使ったことで処置群と統制群における共変量x1, x2, x3の分布は、ほぼ同じになっている。
- バランシングしていることが確認できる。これが、傾向スコアはバランシングスコアという意味である。

処置効果の推定

□ $\text{mean}(y[t1==1]) - \text{mean}(y[t1==0])$

```
> mean(y[t1==1]) - mean(y[t1==0])  
[1] 8.5
```

- $t1$ が1のときの y の平均値と $t1$ が0のときの y の平均値の差を計算する.
- 処置群における平均処置効果 (ATT) の真値 : 9.09
- ATTの真値にかなり近くなっている.

傾向スコア定理1： バランシング

$$X \perp T | e(X)$$

- 処置の割付け T と観測された共変量 X は，傾向スコア $e(X)$ が与えられたとき，条件付き独立である
 - 傾向スコア $e(X)$ が同じ値であれば，処置群と統制群における多変量の共変量 X の分布は同じであることを意味している

傾向スコア定理2：条件付き独立性

$$\{Y(1), Y(0)\} \perp T | e(X)$$

- 傾向スコア $e(X)$ が与えられれば，潜在的結果変数 $\{Y(1), Y(0)\}$ と割付け変数 T は条件付き独立である
 - 傾向スコアが同じ値となる個体であれば，処置への割付けは無作為と見なせるということである

前提条件

- 定理1と定理2をあわせて傾向スコア定理という
- 傾向スコア定理が成り立つための前提条件
 - 強い意味での無視可能な割付けのための条件
 - 条件付き独立性（無交絡性）
$$\{Y(1), Y(0)\} \perp T | X$$
 - 条件付き正值性
$$0 < Pr(T_i = 1 | X) < 1$$
 - オーバーラップ条件ともいう

前提条件と傾向スコア定理の違い

□ $\{Y(1), Y(0)\} \perp T|X$

- 処置の割付けに影響を与えるのは観測された共変量のみということ
- 観察研究では、この条件を満たすために、できるだけ多くの変数をモデルに取り込む必要がある
- **すべての共変量 X が与えられたときという条件**

□ $\{Y(1), Y(0)\} \perp T|e(X)$

- 傾向スコア $e(X)$ が与えられれば、潜在的結果変数 $\{Y(1), Y(0)\}$ と割付け変数 T は条件付き独立である
- **単変量の傾向スコア**が同じ値となる個体であれば、処置への割付けは無作為と見なせる
- これが傾向スコアの利点の1つ

傾向スコアのモデル化

ロジスティック回帰モデル

- 傾向スコアはロジスティック回帰モデルでモデル化できる
 - ロジスティック回帰モデルの全容を解説するには講義時間が足りないが、傾向スコアを理解するために必要な部分、すなわち確率のモデル化について、以下で解説する

初期設定

$$T_i^* = \beta_0 + \beta_1 X_i + \varepsilon_i$$

□ T_i^* : 連続変数

- T_i^* は理論的に存在すると想定される変数だが、実際には観測されない変数とする.
- 上式のとおり, 観測される共変量 X_i と関連しているものとする.

□ X_i : 共変量

□ ε_i : 誤差項

- $E[\varepsilon|X] = 0$, $var[\varepsilon|X] = \pi^2/3$ とする.
- π は円周率である. この誤差項の分散の設定は, モデルの識別のために必要なものである.

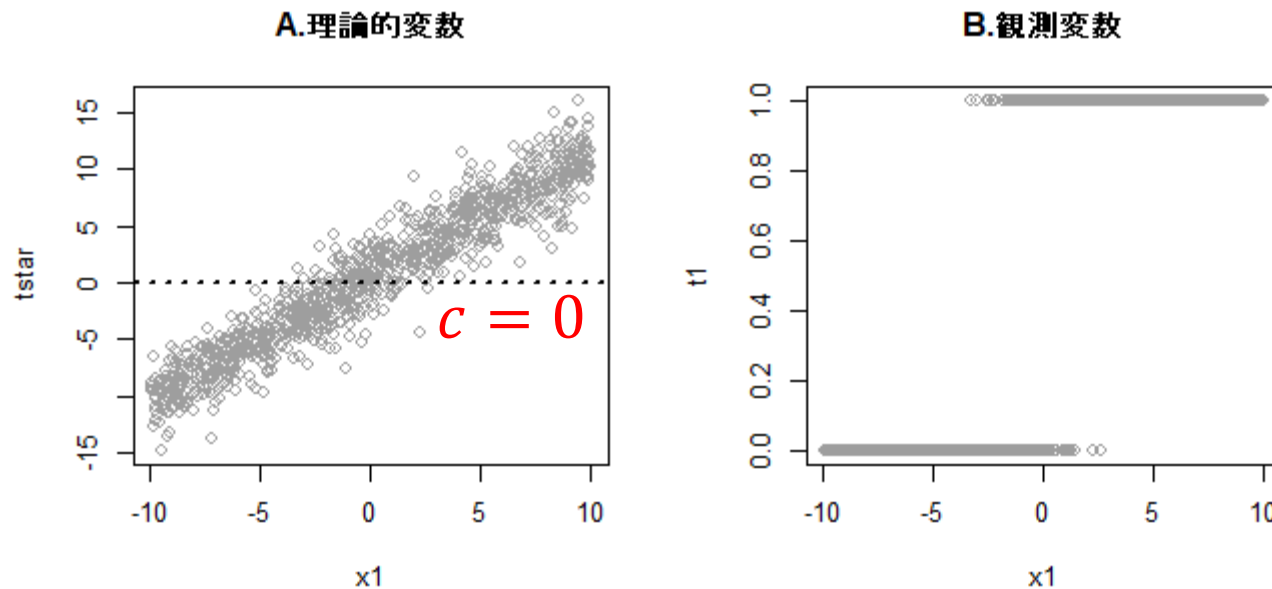
リンク関数

$$T_i = \begin{cases} 1 & \text{if } T_i^* > c \\ 0 & \text{if } T_i^* \leq c \end{cases}$$

- T_i^* は，上式を通じて，実際に観測される二値変数 T_i とリンクされているものとする.

- 具体的例
 - 理論的な連続変数 T_i^* ：入試の点数
 - 実際に観測される二値の変数 T_i ：入試の合否（0＝不合格，1＝合格）

ここまでの設定を図示（1）

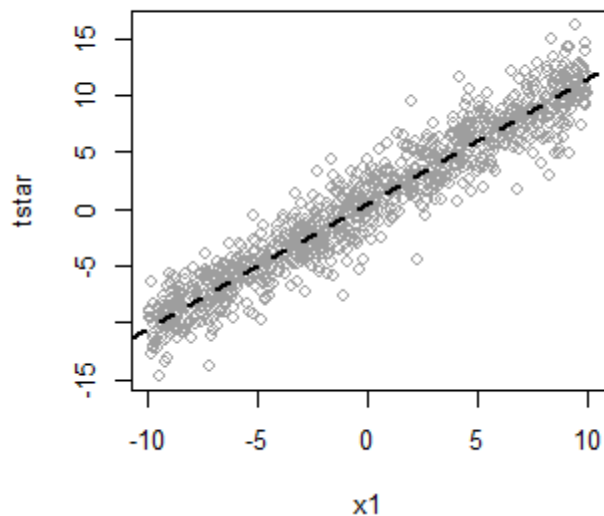


共変量 x_1 と理論的変数
 t_{star} との散布図

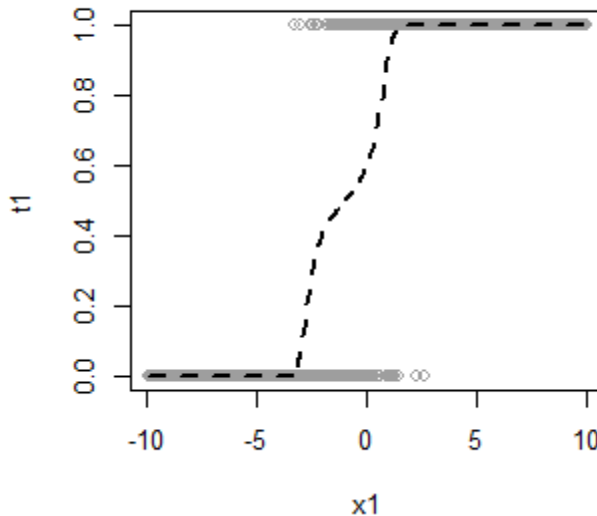
- t_{star} の値が点線よりも下にある
とき、 t_1 は0として観測される。
- t_{star} の値が点線よりも上にある
とき、 t_1 は1として観測される。

ここまでの設定を図示（2）

C. 理論的変数と回帰直線



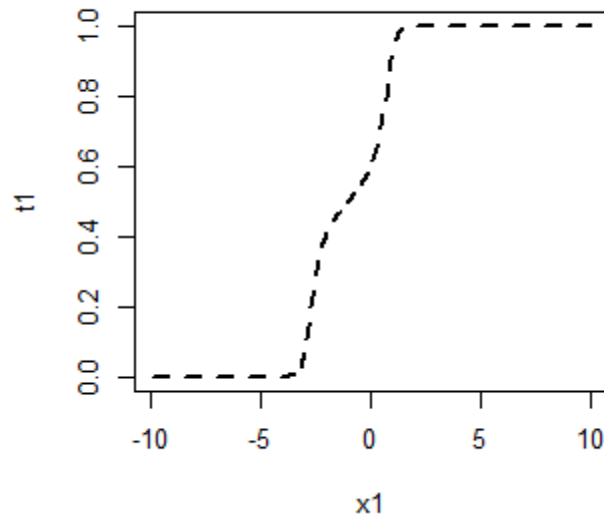
D. 観測変数とLOWESS



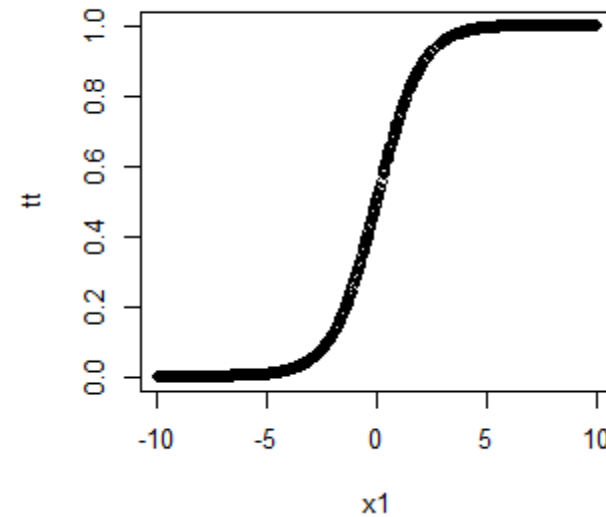
lowessとは、locally weighted scatterplot smoothingの略で、局所的な平滑化を行うノンパラメトリックな回帰モデルである。

ここまでの設定を図示 (3)

D. 観測変数とLOWESS



E. ロジスティック曲線



ロジスティック回帰モデル

$$Pr(T_i = 1|X_i) = \frac{\exp(\beta_0 + \beta_1 X_i)}{1 + \exp(\beta_0 + \beta_1 X_i)}$$

- この式の左辺 $Pr(T_i = 1|X_i)$ こそが傾向スコア $e(X)$ である.
 - $\beta_0 + \beta_1 X_i$ の指数をとったものに関するロジスティック分布の累積分布関数（CDF: cumulative distribution function）である.

R関数glm

- `model1 <- glm(t1 ~ x1 + x2 + x3 + x4, family=binomial(link="logit"))`
- `p1 <- model1$fitted.values`

RパッケージMatchIt

- ❑ `library(MatchIt)`
- ❑ `m.out <- matchit(t1 ~ x1 + x2 + x3 + x4, data=dataXX)`
- ❑ `ps2 <- m.out$model$fitted.values`

パッケージとは、Rにインストールして使用できる関数群のことです。

傾向スコアモデリングの注意点

□ 実験研究

- 観測されない共変量のバランシングまで保証する.

□ 傾向スコアを用いたモデリング

- 処置の割付けに影響を与えるものは観測される共変量のみと仮定している.
- 観測されない共変量のバランシングまでは保証しない.
- 少なくとも観測された共変量についてはバランシングが保証されていると理解できる.
- 共変量の選び方：傾向スコアモデリングにおいて必要となる共変量の選び方は、第7回の講義の議論に準じるので、改めて第7回の講義資料を確認して欲しい

距離：使用できる選択肢

| | distanceの引数の意味 |
|---------------------|--|
| glm | 一般化線形モデル (generalized liner model) |
| gam | 一般化加法モデル (generalized additive model) |
| rpart | 分類木 (classification tree) |
| randomforest | ランダムフォレスト (random forest) |
| nnet | ニューラルネットワーク (neural network) |
| cbps | 共変量バランシング傾向スコア (covariate balancing propensity score) |
| bart | ベイズ的加法回帰木 (BART: Bayesian additive regression trees) |
| mahalanobis | マハラノビスの距離 (Mahalanobis distance) |

- RパッケージMatchItのmatchit関数の引数として、distance=の右辺に距離を設定する.
 - 個体*i*と個体*j*の距離は、傾向スコアの差の絶対値 $|e_i - e_j|$ によって測定.

距離：デフォルトの選択肢

| | distanceの引数の意味 |
|---------------------|--|
| glm | 一般化線形モデル (generalized liner model) |
| gam | 一般化加法モデル (generalized additive model) |
| rpart | 分類木 (classification tree) |
| randomforest | ランダムフォレスト (random forest) |
| nnet | ニューラルネットワーク (neural network) |
| cbps | 共変量バランシング傾向スコア (covariate balancing propensity score) |
| bart | ベイズ的加法回帰木 (BART: Bayesian additive regression trees) |
| mahalanobis | マハラノビスの距離 (Mahalanobis distance) |

- 引数glmからbartまでは、各々の手法からモデル化した傾向スコアを距離として用いる.
- デフォルトは、**glm**で、**ロジスティック回帰によってモデル化した傾向スコア**である.

マッチング

マッチングとは？

- 実際には異なる個体同士について、**観測される共変量が同じ個体をペア**にして、事実上の同じ個体として扱うこと

具体例

| 被験者 | 処置 | 結果 | 年齢 | 性別 |
|-----------|----------|-----------|-----------|----------|
| 1 | 0 | 63 | 20 | 男 |
| 2 | 0 | 55 | 25 | 女 |
| 3 | 0 | 59 | 38 | 男 |
| 4 | 1 | 77 | 46 | 女 |
| 5 | 1 | 73 | 32 | 男 |
| 6 | 0 | 66 | 66 | 女 |
| 7 | 0 | 69 | 18 | 男 |
| 8 | 1 | 81 | 26 | 女 |
| 9 | 1 | 77 | 20 | 男 |
| 10 | 1 | 81 | 54 | 女 |

- 被験者10人のデータ
- 処置：0または1
- 結果：55から81までの数値
- 共変量：年齢と性別（共変量 X は2次元）
- 被験者1：処置を受けていない20歳男性
- 被験者9：処置を受けている20歳男性

年齢と性別でマッチングした後のデータ

| 被験者 | 処置 | 結果 | 年齢 | 性別 |
|-----|----|----|----|----|
| 1 | 0 | 63 | 20 | 男 |
| 9 | 1 | 77 | 20 | 男 |

- この2人をマッチングした理由
- 年齢と性別は同じ
- この2人の「結果」の違い ($77 - 63 = 14$) は、
年齢と性別が原因ではないと考えられるから

無交絡性の仮定

高橋（2022, 第3章）も参照されたい

$$\{Y(1), Y(0)\} \perp T | X$$

□ 因果推論の妥当性

- 「未観測の交絡因子がないという仮定」の妥当性に依存
- 条件付き独立性（無交絡性）が満たされている必要がある

□ 未観測の交絡因子がないという仮定

- この仮定が正しければ、この2人の「結果」の違い（ $77 - 63 = 14$ ）は、「処置」の違い（0と1）が原因と主張できる。

多数の共変量

- 年齢と性別以外の共変量
 - 身長, 体重, 年収, 貯蓄, 学歴, 職歴, 居住地など, さまざまな情報が交絡因子として存在している可能性がある.
- 今回の二変量では十分な情報を持っていないかもしれない.
 - 共変量 X は多変量である必要がある.
- 多変量の共変量をもとにしてマッチングすることは困難

傾向スコアの活用

- 多変量の共変量 X を条件としたときに、処置に割付けられる確率として**単変量の傾向スコアに情報を縮約**することができる。
- 傾向スコア定理
 - 傾向スコアが似ている者同士をマッチングすれば、処置群と対照群における多変量の共変量 X の分布は同じになると考えることができる。
 - **傾向スコアが似ている個体同士をマッチング**して分析を行う。

復元マッチングと非復元マッチング

2種類の抽出方法

□ 初級的な確率統計の授業

- 箱からボールを取り出す方法について議論することがよくある.
- 『統計学入門』（1991, p.110）も参照されたい.

□ 復元抽出（sampling with replacement）

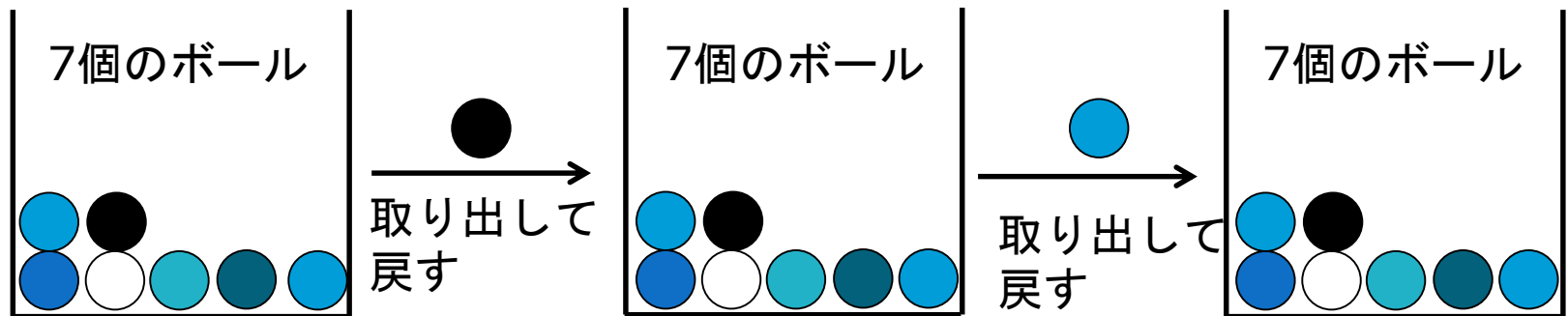
- 一度取り出したボールを箱に戻す抽出の仕方

□ 非復元抽出（sampling without replacement）

- 一度取り出したボールを箱に戻さない抽出の仕方

復元抽出

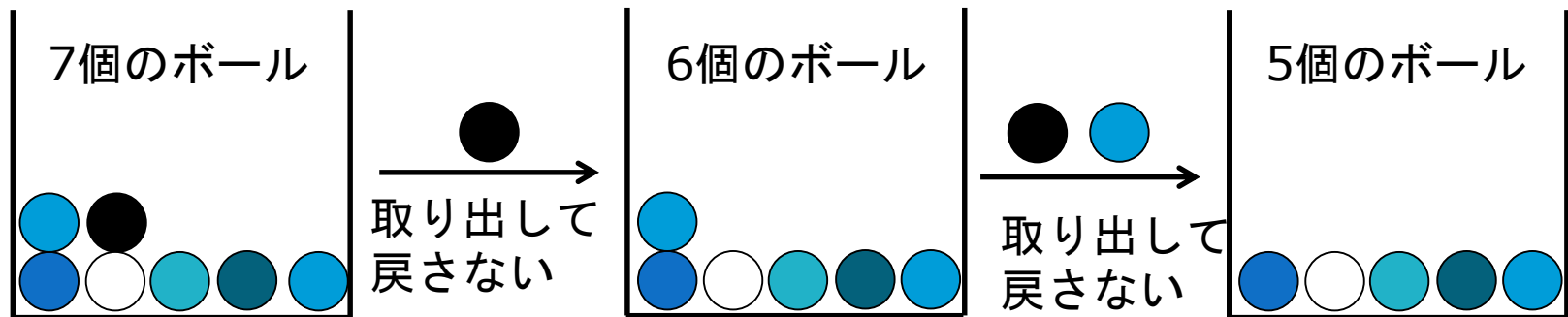
- ある対象を選んだら、その対象を標本空間に**戻す**ことで、毎回、選ぶうる対象の数は**同じ**となる抽出方法



- n 個の対象から m 回の復元抽出によって選ぶなら、対象を選ぶ方法は n^m 通り

非復元抽出

- ある対象を選んだら、その対象を標本空間に**戻さない**ことで、毎回、選ぶ対象の数は**減っていく**抽出方法



- n 個の対象から m 回の非復元抽出によって選ぶなら、対象を選ぶ方法は:

$$n \times (n - 1) \times (n - 2) \times \cdots \times (n - m + 1)$$

具体例：男女のマッチング

| ID | 性別 | 年齢 |
|----|----|----|
| 1 | 女 | 28 |
| 2 | 女 | 28 |
| 3 | 男 | 28 |
| 4 | 男 | 60 |
| 5 | 男 | 65 |

- 5人の男女を年齢だけでマッチングして，年齢の近い男女をペアにする.
- 年齢が30歳以上離れた場合，マッチングの候補としては不適切と考えることにしよう.

復元抽出によるマッチング

| ID | 性別 | 年齢 |
|----|----|----|
| 1 | 女 | 28 |
| 2 | 女 | 28 |
| 3 | 男 | 28 |
| 4 | 男 | 60 |
| 5 | 男 | 65 |

- ID1は28歳の女性であり，ID3，ID4，ID5がマッチング候補の男性である.
 - 男性3人の中でID3の年齢が28歳なので，ID1とマッチングしてペアとしよう.
- 次に，ID2は28歳の女性であり，復元の場合には一度選ばれたID3も再び候補者に含めるので，ID3，ID4，ID5がマッチング候補の男性である.
 - したがって，男性3人の中でID3の年齢が28歳なので，ID2とマッチングしてペアとする.
- よって，2組のペアが出来上がる.

非復元抽出によるマッチング

| ID | 性別 | 年齢 |
|----|----|----|
| 1 | 女 | 28 |
| 2 | 女 | 28 |
| 3 | 男 | 28 |
| 4 | 男 | 60 |
| 5 | 男 | 65 |

- 先ほどと同様に，まずはID1の28歳の女性に対してID3の28歳の男性をマッチングしてペアとする.
- 次に，ID2の28歳の女性に対してマッチング相手を探す
が，非復元の場合には一度選ばれたID3は候補者に含めない
ので，ID4とID5がマッチング候補の男性である.
 - 次点の候補としてはID4になるが，年齢が30歳以上離れた場合，
マッチングの候補としては不適切と考えているから，適切な
マッチング相手ではない.
- よって，1組のペアが出来上がる.

復元抽出の長所と短所（1）

- 現実のお見合い用マッチングアプリ
 - 復元によるマッチングでは同じ人が2人以上とペアになってしまい、大変に困ったことになる。

- データ解析
 - データに含まれる情報が無駄になりにくいという長所があり、解析結果の偏りが小さくなる傾向がある。

復元抽出の長所と短所（2）

□ 復元抽出の長所

- 復元では2組のペアが出来上がった
- 非復元では1組のペアしか出来上がっていない
- 非復元ではデータの情報を十分に活用できていない可能性がある。

□ 復元抽出の短所

- 復元の場合、ID3の男性が2つのペアに登場するわけであるから、**マッチングされたデータは独立でない**
- マッチングの重みを考慮するなど、推定の仕方が複雑になるという短所もある

マッチング方法

使用できる選択肢

| | methodの引数 |
|-----------------|---------------------------------------|
| nearest | 最近隣法マッチング (nearest neighbor matching) |
| optimal | 最適ペアマッチング (optimal pair matching) |
| genetic | 遺伝的マッチング (genetic matching) |
| exact | 厳密マッチング (exact matching) |
| cem | 単純化厳密マッチング (coarsened exact matching) |
| subclass | 層化 (subclassification) |
| full | 最適フルマッチング (optimal full matching) |

- RパッケージMatchItのmatchit関数の引数として、method=の右辺にマッチング方法を設定する。

最近隣法マッチング (nearest neighbor matching)

| | methodの引数 |
|-----------------|---------------------------------------|
| nearest | 最近隣法マッチング (nearest neighbor matching) |
| optimal | 最適ペアマッチング (optimal pair matching) |
| genetic | 遺伝的マッチング (genetic matching) |
| exact | 厳密マッチング (exact matching) |
| cem | 単純化厳密マッチング (coarsened exact matching) |
| subclass | 層化 (subclassification) |
| full | 最適フルマッチング (optimal full matching) |

- 処置群における各々の個体に対して，傾向スコアの値が近い個体を1つずつ順番にマッチングする．
 - 貪欲マッチング (greedy matching) ともいわれる
 - 個別の距離を最小化することを目指しており，必ずしもデータ全体での距離は最小化しない

最適ペアマッチング (optimal pair matching)

| | methodの引数 |
|----------------|--|
| nearest | 最近隣法マッチング (nearest neighbor matching) |
| optimal | 最適ペアマッチング (optimal pair matching) |
| genetic | 遺伝的マッチング (genetic matching) |
| exact | 厳密マッチング (exact matching) |
| cem | 単純化厳密マッチング (coarsened exact matching) |
| subclass | 層化 (subclassification) |
| full | 最適フルマッチング (optimal full matching) |

- マッチングされたすべてのペアに関して，平均絶対距離 (average absolute distance) が最も小さくなるようにマッチングを行うもの

他の選択肢

| | methodの引数 |
|-----------------|---------------------------------------|
| nearest | 最近隣法マッチング (nearest neighbor matching) |
| optimal | 最適ペアマッチング (optimal pair matching) |
| genetic | 遺伝的マッチング (genetic matching) |
| exact | 厳密マッチング (exact matching) |
| cem | 単純化厳密マッチング (coarsened exact matching) |
| subclass | 層化 (subclassification) |
| full | 最適フルマッチング (optimal full matching) |

□ 高橋（2022, pp.161-162）も参照されたい

最近隣法（貪欲マッチング）と最適ペアマッチングの違い

| 距離行列の例 | | | | |
|--------|-----|----------|-----|------|
| | | 統制群（対照群） | | |
| | | 個体4 | 個体5 | 個体6 |
| 処置群 | 個体1 | 0.4 | 0.1 | 10.0 |
| | 個体2 | 100.0 | 0.2 | 4.0 |
| | 個体3 | 5.0 | 3.0 | 0.3 |

- 処置群の個体1・個体2・個体3と統制群の個体4・個体5・個体6をマッチングする.
- 簡単のため, **非復元のマッチング**を考えよう.
- 数字は個体間の距離を表しているものとする.
 - 数字が小さいほど, 個体間の距離が近く, 似ている個体と見なされる.

最近隣法（1）

| 距離行列の例 | | | | |
|--------|-----|----------|-----|------|
| | | 統制群（対照群） | | |
| | | 個体4 | 個体5 | 個体6 |
| 処置群 | 個体1 | 0.4 | 0.1 | 10.0 |
| | 個体2 | 100.0 | 0.2 | 4.0 |
| | 個体3 | 5.0 | 3.0 | 0.3 |

- 最近隣法（貪欲マッチング）は，早い者勝ちの方法
- 最も小さな距離は，**個体1と個体5の距離0.1**である．
- したがって，最初に**個体1と個体5がペア**となる．

最近隣法（2）

| 距離行列の例 | | | | |
|--------|-----|----------|-----|------|
| | | 統制群（対照群） | | |
| | | 個体4 | 個体5 | 個体6 |
| 処置群 | 個体1 | 0.4 | 0.1 | 10.0 |
| | 個体2 | 100.0 | 0.2 | 4.0 |
| | 個体3 | 5.0 | 3.0 | 0.3 |

- 次に小さな距離は**個体2**と**個体5**の距離**0.2**であるが、個体5はすでにペアとなっており、非復元のマッチングを考えているので、この数字は無視する。

最近隣法 (3)

| 距離行列の例 | | | | |
|--------|-----|-----------|-----|------|
| | | 統制群 (対照群) | | |
| | | 個体4 | 個体5 | 個体6 |
| 処置群 | 個体1 | 0.4 | 0.1 | 10.0 |
| | 個体2 | 100.0 | 0.2 | 4.0 |
| | 個体3 | 5.0 | 3.0 | 0.3 |

- 次に小さな距離は個体3と個体6の距離0.3であるから、個体3と個体6をペアとする。

最近隣法（4）

| 距離行列の例 | | | | |
|--------|-----|----------|-----|------|
| | | 統制群（対照群） | | |
| | | 個体4 | 個体5 | 個体6 |
| 処置群 | 個体1 | 0.4 | 0.1 | 10.0 |
| | 個体2 | 100.0 | 0.2 | 4.0 |
| | 個体3 | 5.0 | 3.0 | 0.3 |

- 残ったペアは、**個体2と個体4**であるが、この2つの個体の距離は**100.0**もある。

最近隣法（5）

| 距離行列の例 | | | | |
|--------|-----|----------|-----|------|
| | | 統制群（対照群） | | |
| | | 個体4 | 個体5 | 個体6 |
| 処置群 | 個体1 | 0.4 | 0.1 | 10.0 |
| | 個体2 | 100.0 | 0.2 | 4.0 |
| | 個体3 | 5.0 | 3.0 | 0.3 |

□ 出来上がったペア

- 個体1と個体5
- 個体3と個体6
- 個体2と個体4

□ 距離の総和

- $0.1 + 0.3 + 100.0 = 100.4$

□ 平均絶対距離

- $100.4 / 3 = 33.5$

最適ペアマッチング

| 距離行列の例 | | | | |
|--------|-----|----------|-----|------|
| | | 統制群（対照群） | | |
| | | 個体4 | 個体5 | 個体6 |
| 処置群 | 個体1 | 0.4 | 0.1 | 10.0 |
| | 個体2 | 100.0 | 0.2 | 4.0 |
| | 個体3 | 5.0 | 3.0 | 0.3 |

- 平均絶対距離が最も小さくなるようにマッチングを行う方法.
 - 赤い数字で表されている対角線の距離の総和は $0.4 + 0.2 + 0.3 = 0.9$ で、平均絶対距離は $0.9/3 = 0.3$ となって、これが最も小さい.
 - 個体1と個体4、個体2と個体5、個体3と個体6が最適なペアと判定される.

結論：最近隣法と最適ペアマッチングの違い（1）

| 距離行列の例 | | | | |
|--------|-----|----------|-----|------|
| | | 統制群（対照群） | | |
| | | 個体4 | 個体5 | 個体6 |
| 処置群 | 個体1 | 0.4 | 0.1 | 10.0 |
| | 個体2 | 100.0 | 0.2 | 4.0 |
| | 個体3 | 5.0 | 3.0 | 0.3 |

- 最近隣法（貪欲マッチング）では，個別の距離を最小化することを目指している.
- しかし，よくマッチングされた個体を使いたいなら，最適マッチングがよいと指摘されている.

結論：最近隣法と最適ペアマッチングの違い（2）

| 距離行列の例 | | | | |
|--------|-----|----------|-----|------|
| | | 統制群（対照群） | | |
| | | 個体4 | 個体5 | 個体6 |
| 処置群 | 個体1 | 0.4 | 0.1 | 10.0 |
| | 個体2 | 100.0 | 0.2 | 4.0 |
| | 個体3 | 5.0 | 3.0 | 0.3 |

- 最近隣法（貪欲マッチング）は，早い者勝ちの方法なので，結果として，個体をよくマッチングする方法として適切でないおそれがある．
- この例では，最初に選んだ個体1と個体5のペアを固定してしまっても再検討しないため，結局，個体2と個体4はよくマッチングされていない個体のペアとなってしまったのである．

共変量の標準化平均差の絶対値

- 傾向スコアのモデリングをする際には、処置群と統制群における各々の共変量の標準化平均差の絶対値が小さくなる手法がよいとされている.
- 具体的には0.1未満になるような手法が推奨されている.

傾向スコアマッチングの実践

2種類の推定対象 (estimand)

□ 平均処置効果 (ATE)

- $\tau_{ATE} = E[Y_i(1) - Y_i(0)]$
- $\tau_{ATE} = E[Y_i(1)] - E[Y_i(0)]$

□ 処置群の平均処置効果 (ATT)

- $\tau_{ATT} = E[Y_i(1) - Y_i(0) | T_i = 1]$
- $\tau_{ATT} = E[Y_i(1) | T_i = 1] - E[Y_i(0) | T_i = 1]$

傾向スコアマッチングにおける推定対象

□ ATEではなく, **ATT**

□ 理由 :

- 処置群における個体に対して, 対照群からマッチングする候補を選んでくるため, マッチング後のデータは処置群の個体を中心として構成されているから
- 傾向スコアによってATEを推定したい場合
 - 傾向スコアによる層化解析 (subclassification)
 - 傾向スコアによる重み付け法 (weighting)
 - 詳しくは高橋 (2022, 第12章) を参照

data11の読み込み（教科書p.156）

- ❑ data11 <- read.csv(file.choose())
- ❑ attach(data11)
- ❑ summary(data11)

```
> summary(data11)
```

| y0t | y1t | y3 | t1 |
|----------------|----------------|----------------|----------------|
| Min. : -30.30 | Min. : -31.43 | Min. : -30.30 | Min. : 0.000 |
| 1st Qu.: 16.23 | 1st Qu.: 18.91 | 1st Qu.: 16.88 | 1st Qu.: 0.000 |
| Median : 30.79 | Median : 34.70 | Median : 30.56 | Median : 0.000 |
| Mean : 31.29 | Mean : 35.05 | Mean : 32.42 | Mean : 0.389 |
| 3rd Qu.: 47.54 | 3rd Qu.: 51.85 | 3rd Qu.: 49.91 | 3rd Qu.: 1.000 |
| Max. : 101.65 | Max. : 109.19 | Max. : 109.19 | Max. : 1.000 |

| x1 | x2 | x3 |
|-----------------|-----------------|-----------------|
| Min. : -1.8499 | Min. : -1.8679 | Min. : -2.3885 |
| 1st Qu.: 0.3613 | 1st Qu.: 0.2532 | 1st Qu.: 0.3073 |
| Median : 0.9937 | Median : 0.9921 | Median : 0.9687 |
| Mean : 0.9816 | Mean : 0.9727 | Mean : 1.0053 |
| 3rd Qu.: 1.5948 | 3rd Qu.: 1.6457 | 3rd Qu.: 1.7140 |
| Max. : 3.8706 | Max. : 4.7807 | Max. : 4.0386 |

| x4 | x5 | x6 |
|-----------------|-----------------|-----------------|
| Min. : -1.8857 | Min. : -2.4033 | Min. : -1.9252 |
| 1st Qu.: 0.3407 | 1st Qu.: 0.3220 | 1st Qu.: 0.2879 |
| Median : 1.0123 | Median : 0.9727 | Median : 1.0183 |
| Mean : 1.0254 | Mean : 0.9843 | Mean : 1.0138 |
| 3rd Qu.: 1.7158 | 3rd Qu.: 1.6205 | 3rd Qu.: 1.7328 |
| Max. : 4.9507 | Max. : 4.6297 | Max. : 3.5217 |

ATTの真値

□ $\text{mean}(y1t[t1==1]) - \text{mean}(y0t[t1==1])$

```
> mean(y1t[t1==1]) - mean(y0t[t1==1])  
[1] 2.888651
```

処置群の平均処置効果 (ATT)

$$\tau_{ATT} = E[Y_i(1) - Y_i(0) | T_i = 1]$$

$$\tau_{ATT} = E[Y_i(1) | T_i = 1] - E[Y_i(0) | T_i = 1]$$

傾向スコアマッチング

- ❑ `library(MatchIt)`
- ❑ `m.out1 <- matchit(t1 ~ x1 + x2 + x3 + x4 + x5 + x6, data = data11, replace = TRUE, distance = "glm", method = "nearest")`
- ❑ `m.data1 <- match.data(m.out1)`

解析モデル

- `model1 <- lm(y3 ~ t1 + x1 + x2 + x3 + x4 + x5 + x6, data = m.data1, weights = weights)`
- `summary(model1)`

Coefficients:

| | Estimate | Std. Error |
|-------------|----------|------------|
| (Intercept) | 1.8485 | 0.5049 |
| t1 | 2.9174 | 0.4026 |
| x1 | 1.7283 | 0.2465 |
| x2 | 1.0956 | 0.2480 |
| x3 | -2.1161 | 0.2437 |
| x4 | 7.8469 | 0.2553 |
| x5 | 9.9916 | 0.2479 |
| x6 | 11.2766 | 0.2413 |

標準誤差

- ❑ library(lmtest)
- ❑ library(sandwich)
- ❑ coeftest(model1, vcov. = vcovCL, cluster = ~weights)

| | Estimate | Std. Error |
|-------------|----------|------------|
| (Intercept) | 1.84855 | 1.08109 |
| t1 | 2.91741 | 0.67454 |
| x1 | 1.72827 | 0.33283 |
| x2 | 1.09558 | 2.06205 |
| x3 | -2.11607 | 3.03974 |
| x4 | 7.84693 | 0.33587 |
| x5 | 9.99163 | 0.25037 |
| x6 | 11.27663 | 0.42177 |

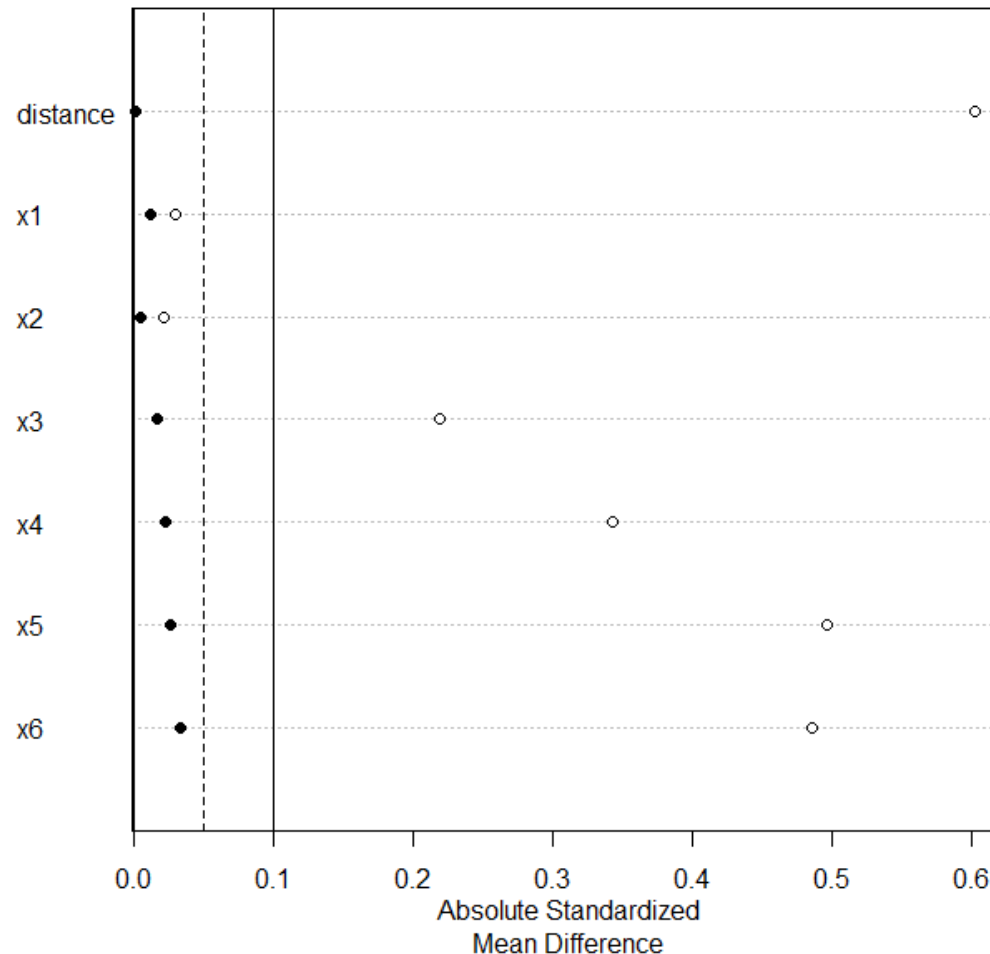
信頼区間

- `coefci(model1, level = 0.95, vcov. = vcovCL, cluster = ~weights)`

| | 2.5 % | 97.5 % |
|-------------|------------|-----------|
| (Intercept) | -0.2745683 | 3.971664 |
| t1 | 1.5927145 | 4.242102 |
| x1 | 1.0746396 | 2.381904 |
| x2 | -2.9539977 | 5.145151 |
| x3 | -8.0857046 | 3.853565 |
| x4 | 7.1873315 | 8.506527 |
| x5 | 9.4999391 | 10.483317 |
| x6 | 10.4483347 | 12.104929 |

ラブプロットによるバランシングの評価

□ `plot(summary(m.out1), position=NULL)`



傾向スコアをマッチングに使うべきでない？

King and Nielsen (2019)

- 「傾向スコアをマッチングに使うべきでない理由」というタイトルの論文
 - マハラノビス距離マッチング (Mahalanobis distance matching) や単純化厳密マッチング (CEM) と比較して、傾向スコアマッチングでは、用いたモデルに依存して結果が変わってしまう可能性が高く、実際には共変量がバランシングしないことがある。

傾向スコアマッチングのパラドックス

- データが元からバランシングしているのであれば、傾向スコアマッチングは悪影響を及ぼす.
- データが元からバランシングしていないのであれば、傾向スコアマッチングがバランシングを改善したとしても、そのようなデータはそもそも因果推論に適していない.

問題のポイント（1）

- King and Nielsen（2019）が問題視しているのは、傾向スコアマッチングである.
- 傾向スコアそのものではない点である.
- 次回の授業で扱う傾向スコアによる層化解析法および重み付け法は批判の対象にはなっていない.

問題のポイント（2）

- 傾向スコアマッチングで考慮すべきこと
 - 距離の求め方
 - マッチング方法
 - 復元と非復元の選択
 - さまざまなことを考慮しなければいけなかった.
- King and Nielsen（2019）の批判している傾向スコアマッチング
 - 最近隣法マッチング（貪欲マッチング）による非復元の1対1マッチングのこと
 - 必ずしもすべての傾向スコアマッチングを批判しているわけではない.

この批判の適用範囲は極めて限定的

- King and Nielsen (2019) の主張それ自体は正しいものの、その適用範囲は極めて限定的というのが、大勢の見方である
 - Nguyen et al., 2017; Jann, 2017; Guo, Fraser, and Qi, 2020; Cunningham, 2021
- King and Nielsen (2019) の主張は、最近隣法マッチング（貪欲マッチング）による非復元の1対1の傾向スコアマッチングのような単純な手法を機械的に適用してはいけないという教訓と理解できる。