

傾向スコア解析



北海道大学 医学統計学
横田 勲

1

今回の内容

2

- ▶ 交絡と傾向スコア
- ▶ 重み付け解析と標的集団
- ▶ マッチング

2

バックドア基準 back-door criterion

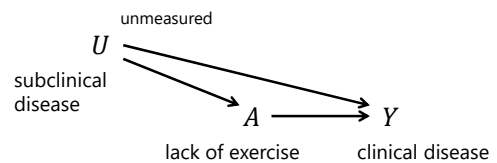
3

- ▶ AはYの非子孫
- ▶ 次の2条件を満たす頂点集合SはA-Yについてバックドア基準を満たす
 - ▶ AからSの任意の要素へ有向道がない
 - ▶ Aから出る矢線をすべて除いたグラフにおいて、SがAとYを有向分離する
- ▶ S, A, Yが観察されていれば、AからYへの因果効果は識別可能

3

交絡の例①

4

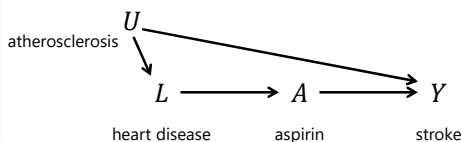


- ▶ 逆因果 reverse causation
- ▶ $A \leftarrow U \rightarrow Y$ のバックドアパスが存在
 - ▶ Uをもし観察できれば、条件付けることでバックドアパスをブロック

4

交絡の例②

5



- ▶ 適応による交絡 confounding by indication
- ▶ UかLを条件付ければ $A \leftarrow L \leftarrow U \rightarrow Y$ をブロック
 - ▶ LはUを介してYに影響するため、経験的同定基準でも交絡因子として定義される

5

平均因果効果 Average Causal Effects

6

- ▶ $E[Y^{a=1}] - E[Y^{a=0}]$
 - ▶ 集団全員が曝露を受けた場合と集団全員が曝露を受けなかった場合の差
- ▶ Null hypothesis of no average causal effect
 - ▶ $E[Y^{a=1}] = E[Y^{a=0}]$
 - ▶ Sharp causal null hypothesisに加え、Helpedな人とHurtな人が同数いる場合も成立

6

周辺構造モデル

7

- ▶ Marginal Structural Model
 - ▶ 潜在アウトカムの周辺平均に関するモデル
- ▶ $E[Y^a] = \Pr[Y^a = 1] = \beta_0 + \beta_1 a$
 - ▶ β_1 が平均因果リスク差
- ▶ logit $\Pr[Y^a = 1] = \beta_0 + \beta_1 a$
 - ▶ β_1 が平均因果オッズ比

7

擬似集団 pseudo-population

8

- ▶ 因果リスクを知る上で必要な、全員が曝露／非曝露である場合の仮想集団
- ▶ 実際に曝露を受けた人は9/12
 - ▶ 曝露を受けた割合の逆数をかけてみよう
 - ▶ $\frac{1}{\frac{9}{12}} = \frac{12}{9}$ 倍
- ▶ 曝露を受けなかった人は3/12
 - ▶ 逆数をかけよう； $\frac{1}{\frac{3}{12}} = 4$ 倍

8

擬似集団にてモデル化

9

- ▶ $E[Y^a] = E_{ps}[Y|a]$
 - ▶ 擬似集団において実際に曝露 a を受けた人のアウトカムの期待値
- ▶ 周辺構造モデルにおける因果パラメータを擬似集団の関連指標として推定
 - ▶ $E[Y^a] = \beta_0 + \beta_1 a$
 - ▶ $E_{ps}[Y|a] = \theta_0 + \theta_1 a$

9

逆確率重み付け法

10

- ▶ Inverse probability weighting (IPW) 法
 - ▶ Horvitz-Thompson推定量(1952, *J Am Stat Assoc*)
$$\hat{E} \left[\frac{I(A=1)Y}{\Pr(A=1|L)} \right]$$
- ▶ 生成した擬似集団での関連効果指標は
 - ▶ 擬似集団での因果効果指標
 - ▶ 元の集団での因果効果指標 に同じ
 - ▶ $Y^a \perp\!\!\!\perp A|L$ for all a ゆえ

10

Hajek推定量

11

- ▶ Modified Horvitz-Thompson推定量
$$\frac{\hat{E} \left[\frac{I(A=1)Y}{\Pr(A=1|L)} \right]}{\hat{E} \left[\frac{I(A=1)}{\Pr(A=1|L)} \right]}$$
 - ▶ 二項アウトカム Y においてHajek推定量は0~1
 - ▶ Positivityが成立しない場合を自動的に除外

11

(非)曝露確率の逆数での重み

12

- ▶ IPW法による操作
 - ▶ 実際に曝露を受けた集団から全員が曝露を受けた場合の擬似集団
 - ▶ 実際に曝露を受けなかった集団から全員が曝露を受けなかった場合の擬似集団
- ▶ サンプルサイズが2倍に
 - ▶ $\frac{1}{\Pr(A|L)}$ を $\frac{0.5}{\Pr(A|L)}$ とすれば1:1ランダム化相当

12

Stabilized weight 安定化重み

13

- ▶ 集団全体の曝露確率に寄せる

$$\frac{\Pr(A)}{\Pr(A|L)}$$

- ▶ 飽和モデルでない場合に効率上昇
 - ▶ 曝露が連続量、時間依存性である場合

13

安定化重み stabilized weight

14

- ▶ $\frac{1}{\Pr(A|L)}$ ではなく $\frac{\Pr(A)}{\Pr(A|L)}$ を用いた擬似集団

- ▶ Hajek 推定量

$$\frac{\hat{E} \left[\frac{I(A=1)Y}{\Pr(A=1|L)} \right]}{\hat{E} \left[\frac{I(A=1)}{\Pr(A=1|L)} \right]}$$

- ▶ 二項アウトカムYにおいてHajek推定量は0~1
- ▶ Positivityが成立しない場合を自動的に除外

14

傾向スコア $\pi(L)$

15

Rosenbaum, Rubin. Biometrika. 1983.

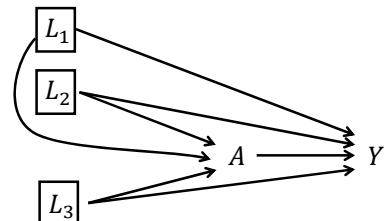
- ▶ 対象者それぞれの曝露（治療）の受けやすさ（確率）
- ▶ 1:1割付のランダム化研究
 - ▶ $\pi(L)$ は全員0.5で共通
- ▶ 観察研究
 - ▶ $\pi(L)$ を対象者ごとに予測
 - ▶ $Y^a \perp\!\!\!\perp A|L \Rightarrow Y^a \perp\!\!\!\perp A|\pi(L)$
 - ▶ 交絡因子をすべて特定している
 - ▶ 交絡因子から傾向スコアを正しく求められている
 - ▶ $\pi(L) = 0, 1$ となる対象者がいない

15

交絡因子はたいてい複数

16

- ▶ ただし、すべて観察できている



すべてブロックしよう

16

交絡因子の候補が10個

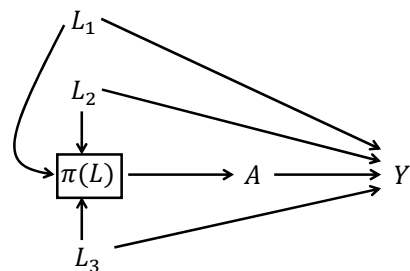
17

- ▶ 層別解析、マッチング
 - ▶ すべて2値変数だったとしても1024層
 - ▶ マッチングでも10個の変数が完全に一致する相手はまれ
 - ▶ 連続変数だったら絶望的
- ▶ 回帰モデルの利用
 - ▶ 10パラメータで済む単純なモデル
 - ▶ 交絡因子間の関係は無視
 - ▶ 曝露効果の大きさは交絡因子によらない

17

傾向スコアを因果DAGで表現

18



18

解析の流れ

19

- ▶ 対象者ごとの傾向スコアを推定
- ▶ 傾向スコアで調整した解析
 - ▶ 回帰分析で共変量に
 - ▶ 限定、層別解析
 - ▶ マッチング
 - ▶ 重み付け解析
- ▶ 交絡因子のバランスがとれたかを確認
- ▶ 平均因果効果の推定
+ 感度解析

19

傾向スコアの作り方

20

- ▶ 対象者ごとの曝露確率の予測値を得たい
 - ▶ 交絡因子のバランスがとれるような予測値
- ▶ どんな統計手法を使ってもよい
 - ▶ ロジスティック回帰
 - ▶ 層別解析
 - ▶ プロビット回帰やノンパラメトリック回帰
 - ▶ 分類木
 - ▶ 機械学習の手法
 - ▶ . . .

20

NHEFSデータの一部

21

National Health and Nutrition Examination Survey Data I Epidemiologic Follow-up Study

- ▶ 25歳から74歳の喫煙者1,566名の11年追跡
- ▶ 曝露 A
 - ▶ ベースライン時に喫煙習慣をやめた($A = 1$)
 - ▶ 喫煙習慣はやめなかった($A = 0$)
- ▶ アウトカム Y : 体重変化量(kg)
- ▶ 交絡 L
 - ▶ 年齢、性別、人種、学歴、ベースライン体重、喫煙本数、喫煙期間、運動習慣、活動性など

21

傾向スコアモデル

22

- ▶ 対象者ごとに喫煙習慣をやめる確率を予測するモデルを作成
 - ▶ 喫煙習慣をアウトカムとして交絡を説明変数に加えたロジスティック回帰
- ▶ 交絡因子 (ベースライン時)
 - ▶ 性別、人種、年齢(2次項まで)、教育歴、喫煙本数(2次項まで)、喫煙期間(2次項まで)、運動習慣、活動性、体重(2次項まで)

22

ベースライン情報①

23

平均, 頻度(割合)	喫煙グループ $A = 0$ ($N = 1,163$)	禁煙グループ $A = 1$ ($N = 403$)
性別 (男性)	542 (47%)	220 (55%)
(女性)	621 (53%)	183 (45%)
人種 (白人)	993 (85%)	367 (91%)
(その他)	170 (15%)	36 (9%)
年齢	42.8	46.2
教育歴 (大卒以上)	115 (10%)	62 (15%)
(大学中退)	92 (8%)	29 (7%)
(高卒)	480 (41%)	157 (39%)
(高校中退)	266 (23%)	74 (18%)
(中卒以下)	210 (18%)	81 (20%)

23

ベースライン情報②

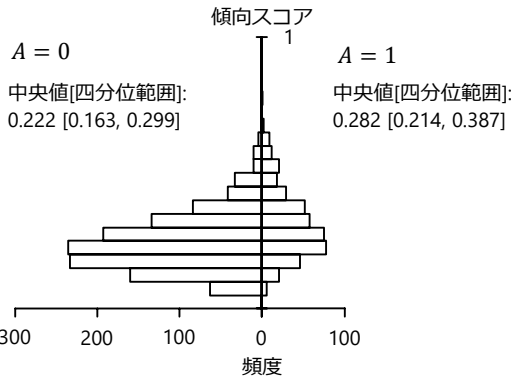
24

平均, 頻度(割合)	喫煙グループ $A = 0$ ($N = 1,163$)	禁煙グループ $A = 1$ ($N = 403$)
喫煙本数/日	21.2	18.6
喫煙年数	24.0	26.0
運動 (とてもする)	237 (20%)	63 (16%)
(そこそこ)	485 (42%)	176 (44%)
(ほぼしない)	441 (38%)	164 (41%)
活動 (とても活動的)	532 (46%)	170 (42%)
(そこそこ)	527 (45%)	188 (47%)
(非活動的)	104 (9%)	45 (11%)
体重	70.3	72.4

24

傾向スコア（曝露確率）の分布

25



25

標的集団

26

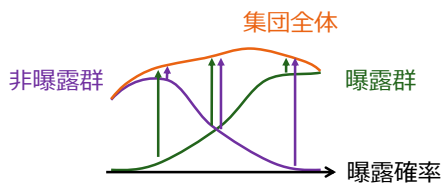
- ▶ 議論を行う集団
 - ▶ 特定の疾患を診断され、全員に同様の治療を実施できるような場合
 - ▶ 集団全体が標的集団
 - ▶ 診療報酬データのような一般的なデータ
 - ▶ 当該疾患に罹り、特定の治療を受けるような集団

26

IPW解析での擬似集団①

27

- ▶ 重みの作り方で、いろいろな集団に
 - ▶ Average treatment effect
 - ▶ 曝露群では曝露確率の逆数
 - ▶ 非曝露群では非曝露確率の逆数

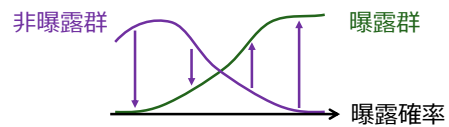


27

IPW解析での擬似集団②

28

- ▶ ATT; Average treatment effect as treated (exposed)
 - ▶ 曝露群はそのまま
 - ▶ 非曝露群では曝露確率のオッズ
- ▶ 曝露を受けるような集団での議論

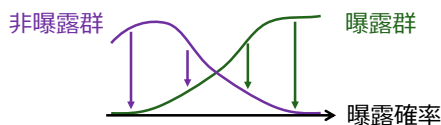


28

IPW解析での擬似集団③

29

- ▶ Matching weight
- ▶ マッチできるような集団での議論
 - ▶ 曝露確率の低い群にあわせる



29

IPW解析での擬似集団④

30

- ▶ Overlap weight
 - ▶ 曝露群では非曝露を受ける確率で
 - ▶ 非曝露群では曝露を受ける確率で重み
- ▶ 傾向スコアの重なりが小さいときに安定した推定を行いやすい
- ▶ 推定対象が平均因果効果ではない

30

練習① 傾向スコア解析

31

- ▶ NHEFSデータで傾向スコア解析をIPW法で行い、ATE、ATTを求めよう
 - ▶ 傾向スコアモデルはロジスティックモデル
 - ▶ 体重変化量の平均因果差
- ▶ データは以下より入手
 - ▶ <https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/>

31

重み付け解析の結果

32

- ▶ ATEを推定
 - ▶ 3.38 (95%CI: 2.42, 4.35)
- ▶ ATTを推定
 - ▶ 3.36 (95%CI: 2.40, 4.33)

32

重み付け解析の利点・弱点

33

- ▶ 利点
 - ▶ 狙った集団での平均因果効果を自在に推定
- ▶ 弱点
 - ▶ 傾向スコアが0や1に近い場合、大きな重みとなりすぎる
 - ▶ 結果が不安定に
 - ▶ 傾向スコアのトリム(Crump RK, Biometrika. 2009)がおすすめ
 - ▶ Stürmer T, et al. Am J Epidemiol. 2010, 2021
 - ▶ Conover MM, et al. Stat Med. 2021

33

解析手段

34

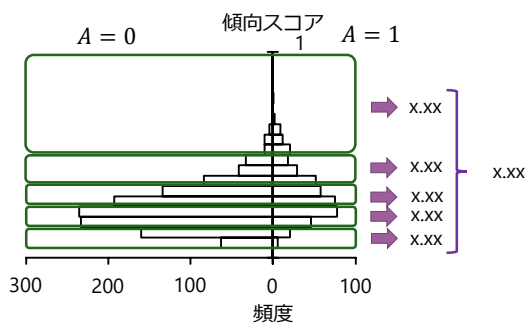
- ▶ 回帰分析の説明変数に加える
 - ▶ 限定、層別解析
 - ▶ マッチング
 - ▶ 重み付け解析
- 治療を受ける確率は40%
-
- 治療を受けた $\times \frac{1}{0.4} = 5$
- 治療を受けなかった $\times \frac{1}{0.6} = 5$

34

層別解析

35

- ▶ 傾向スコアでサブグループ化



35

層別解析の前提

36

- ▶ 残差交絡がない
 - ▶ 層内では交絡が起こっていない
 - ▶ 層別で十分に交絡を調整できる
- ▶ 異なる傾向スコアをもつ対象者が同じ層にいるのに残差交絡がない？
 - ▶ 離散的な傾向スコアならいいかも
- ▶ 重み付け解析に比べて利点がほぼない

36

マッチング

37

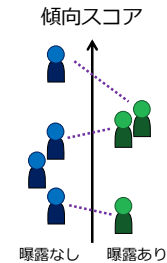
- ▶ 同じ傾向スコアを持つ対象者同士を組に
- ▶ マッチングの種類
 - ▶ マッチングアルゴリズム
 - ▶ 最近傍マッチング、最適マッチング
 - ▶ キャリパーの設定
 - ▶ 復元か非復元か
 - ▶ マッチする人数比

37

最近傍マッチング

38

- ▶ Greedyマッチングのひとつ
- ▶ 傾向スコアが最も近い人とマッチ

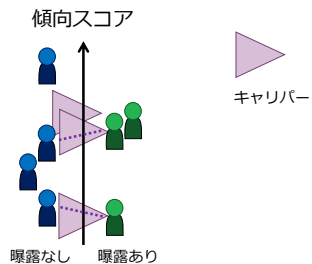


38

キャリパー付き最近傍マッチング

39

- ▶ 許容されるマッチ候補との傾向スコアの差
- ▶ キャリパー内に候補複数いたらランダムに選ぶことも

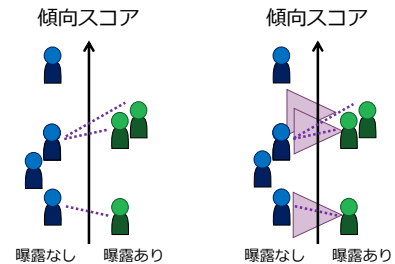


39

復元マッチング

40

- ▶ マッチされる相手として繰り返し登場可
- ▶ 非復元：重婚を認めない

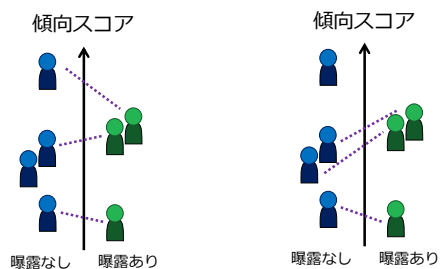


40

初期値による影響

41

- ▶ マッチさせる順番
 - ▶ 小さいほうから
 - ▶ 大きいほうから



41

マッチする人数比

42

- ▶ 1:1 マッチ
 - ▶ 群間で人数がそろろう
- ▶ 1:k マッチ
- ▶ 変動比マッチ
 - ▶ 対象者によってマッチする人数が異なる
 - ▶ 曝露グループの人数がとても少ない場合

42

最適マッチング

43

- ▶ 傾向スコアの差の合計が最小になるようあらゆるペアを探索

43

マッチングの利点・弱点

44

- ▶ 利点
 - ▶ ペアになるよう人を選ぶだけでわかりやすい
 - ▶ あたかもランダム化したような気分になれる
- ▶ 弱点
 - ▶ 自分で作った傾向スコアが同じになるよう人を選んだので別にランダム化ではない
 - ▶ 選ばれなかった人の情報を捨てる効率ロス
 - ▶ マッチングの手順を考慮した解析が必要？

44

練習② マッチング

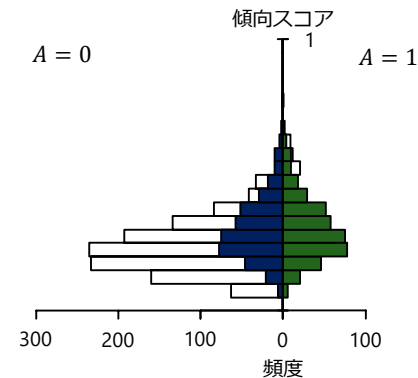
45

- ▶ 以下の傾向スコア
 - ▶ 曝露グループ
0.10, 0.20, 0.40, 0.50, 0.80
 - ▶ 非曝露グループ
0.03, 0.15, 0.30, 0.35, 0.60
- ▶ マッチング方法
 - ▶ 最近傍1:1
 - ▶ (小さいほうから・大きいほうから)
 - ▶ キャリパーなし、キャリパー：±0.05
 - ▶ 最適マッチング

45

マッチされた人・されない人

46



46

1:1最近傍マッチングでの背景①

47

平均, 頻度(割合)	喫煙グループ A = 0 (N=396)	禁煙グループ A = 1 (N=396)
性別 (男性)	227 (57%)	216 (55%)
(女性)	169 (43%)	180 (45%)
人種 (白人)	364 (92%)	360 (91%)
(その他)	32 (8%)	36 (9%)
年齢	46.2	45.9
教育歴 (大卒以上)	57 (14%)	62 (16%)
(大学中退)	31 (8%)	27 (7%)
(高卒)	152 (38%)	154 (39%)
(高校中退)	78 (20%)	74 (19%)
(中卒以下)	78 (20%)	79 (20%)

47

1:1最近傍マッチングでの背景②

48

平均, 頻度(割合)	喫煙グループ A = 0 (N=396)	禁煙グループ A = 1 (N=396)
喫煙本数/日	19.3	18.7
喫煙年数	26.7	26.1
運動 (とともする)	70 (18%)	62 (16%)
(そこそこ)	176 (44%)	174 (44%)
(ほぼしない)	150 (38%)	160 (40%)
活動 (とても活動的)	163 (41%)	167 (42%)
(そこそこ)	184 (46%)	186 (47%)
(非活動的)	49 (12%)	43 (11%)
体重	71.8	72.2

48

標準化差で評価

49

- ▶ 曝露グループ間での背景因子のバランス
 - ▶ 検定では確認しない
 - ▶ 標準化差が±0.1、±0.25以内に入れればよい

$$\frac{\bar{X}_{a=1} - \bar{X}_{a=0}}{s.d.}$$

- ▶ ただし、 $s.d. = \sqrt{\frac{\text{Var}(X_{a=1}) + \text{Var}(X_{a=0})}{2}}$

49

1:1最近傍マッチングでの標準化差

50

	マッチング前	マッチング後
性別	-0.16	0.06
人種	-0.18	0.03
年齢	0.28	0.02
喫煙本数/日	-0.22	-0.05
喫煙年数	0.16	-0.05
体重	0.13	0.03

- ▶ 交絡因子のバランスがうまくとれたよう

50

練習③ 標準化差

51

- ▶ スライド#47のマッチング前後での性別の標準化差を計算してみよう

51

体重変化量Yの解析

52

- ▶ 通常のt検定
 - ▶ マッチング前 : 2.54 (95%CI: 1.66, 3.43)
 - ▶ マッチング後 : 3.39 (95%CI: 2.29, 4.50)

52

まとめ

53

- ▶ 傾向スコア
 - ▶ 平均因果効果を推定する強力な手段
 - ▶ 交絡因子がたくさんある場合に有用
- ▶ マッチング
- ▶ 重み付け解析
 - ▶ それぞれの利点・弱点

53