

医学統計勉強会

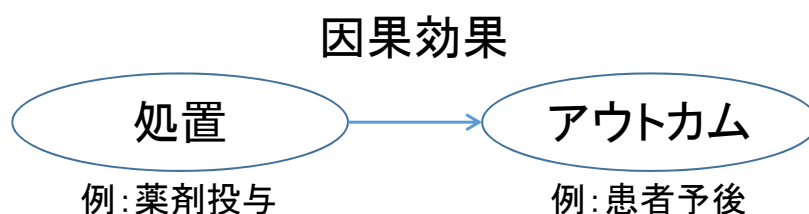
東北大学病院循環器内科・東北大学臨床研究推進センター 共催

東北大学大学院医学系研究科EBM開発学寄附講座

宮田 敏

傾向スコア

新しい処置 = 医学的介入の因果効果の検証



「処置」と「アウトカム」は、様々な交絡因子（予後因子）の影響を受ける。

無作為化比較試験 (Randomized Controlled Trial, RCT) では、介入のランダムな割り付けで予後因子を均等化する。

無作為化が困難な場合 ⇒ 傾向スコア

無作為化比較試験 (Randomized Controlled Trial, RCT)

- 新しい処置を受ける群 + 対照群 (control)
- 処置群と対照群の割り付けは、無作為割り付け。(randomization)
- ヒトを対象とする場合、情報開示と文書による同意。(informed consent)

無作為割り付けにより、**既知**および**未知**の交絡因子に関して、処置群と対照群が確率的な誤差を除いて**同等**。

観察研究では、無作為割り付けが可能とは限らない。

- 臨床では状態の悪い患者さんに、優先して投薬。
- 割り付けの基準はときに不明確，不可解。
- 喫煙など有害な習慣。

交絡因子による影響の調整

- **層別，マッチング**：共変量が「似た」サブグループに分割し、サブグループごとに解析
⇒ 共変量が多いと層の数が多くなり、各層ごとの標本数が少なくなる。
- **多変量解析**：回帰分析，ロジスティック回帰，Cox比例ハザードモデル，等。
⇒ 共変量とアウトカムの関係性を、適切にモデル化する必要がある。**変数選択，線形モデル**の仮定等。
- **傾向スコア (Propensity Score, PS)**：「処置」の有無に関わる全交絡因子の情報を集約する。

例：lindnerデーターPCI施行後のAbciximab投与の効果

lifepres： PCI施行6ヶ月経過後の平均余命。死亡=0，生存=11.6（死亡は、996例中26例）
 cardbill： PCI施行後6ヶ月間の心臓病関連医療費（ドル）
 abcix： Abciximab投与有り=1，投与なし=0
 stent： ステント
 height： 身長
 female： 性別（女性）
 diabetic： 糖尿病
 acutemi： 7日以内のAMI
 ejecfrac： 左室駆出率

	lifepres	cardbill	abcix	stent	height	female	diabetic	acutemi	ejecfrac
1	0	14301	1	0	163	1	1	0	56
2	11.6	3563	1	0	168	0	0	0	56
3	11.6	4694	1	0	188	0	0	0	50

傾向スコア解析においては、まず共変量（背景因子）の介入群と非介入群の間での比較を行う。

例：lindnerデーターPCI施行後のAbciximab投与の効果

Name	abcix=0 : (n=298)	abcix=1: (n=698)	p-value	standardized-difference
cardbill	14614.22 ± 14514	16126.68 ± 9383.83	0.098	0.124
height	171.45 ± 10.59	171.44 ± 10.69	0.996	0
ejecfrac	52.29 ± 10.3	50.4 ± 10.42	0.009	-0.182
stent	174 (58.4%)	492 (70.5%)	0	0.255
female	115 (38.6%)	231 (33.1%)	0.11	-0.115
diabetic	80 (26.8%)	143 (20.5%)	0.031	-0.15
acutemi	18 (6%)	125 (17.9%)	0	0.372

心臓病関連医療費、身長、性別を除く共変量でAbciximab投与群と非投与群に有意差がある。
 検定：Welch's t-test, Fisher's exact test

介入群、非介入群の比較には検定の他に以下の **Standardized difference** の検討も行う。

$$d = \frac{(\bar{x}_{treatment} - \bar{x}_{control})}{\sqrt{\frac{s_{treatment}^2 + s_{control}^2}{2}}},$$

連続変数の場合

$$d = \frac{(\hat{p}_{treatment} - \hat{p}_{control})}{\sqrt{\frac{\hat{p}_{treatment}(1 - \hat{p}_{treatment}) + \hat{p}_{control}(1 - \hat{p}_{control})}{2}}}$$

離散変数の場合

|Standardized difference| < 0.1 の場合、差は軽微なものともみなされる。lindnerデータの場合、差が軽微なのは身長のみ。

傾向スコア

【定義】 **傾向スコア (Propensity Score, PS)**

処置の有無に關与する複数の共変量を用いた、処置群、対照群への**割り付け確率を予測するスコア**。

$$PS = P(\text{treatment group} \mid x_i = X)$$

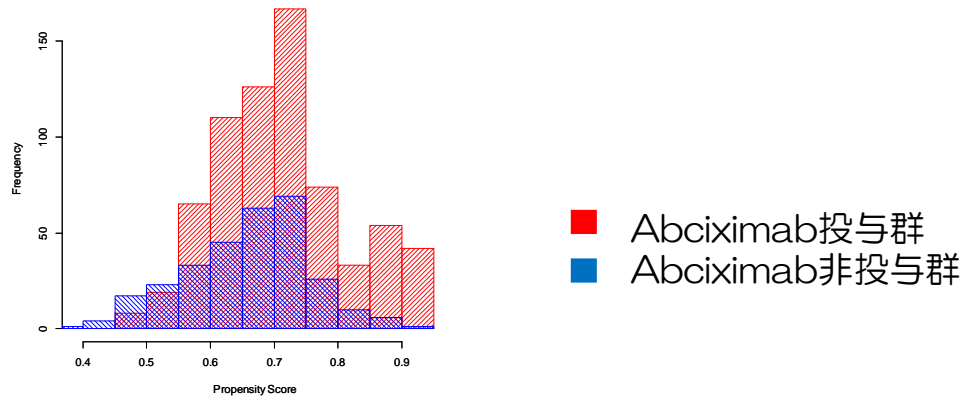
通常は、PSは**ロジスティック回帰モデル**によって推定される。処置の有無に關与する共変量は全てPS予測に用いられ、PS予測に用いられた共変量の情報は、全てPSに**吸収**されたと解釈する（それ以降の解析では、PS推定に用いた共変量は使用しない。）

例：lindnerデータのPropensity Score

ロジスティック回帰モデルによるPS

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1(\text{height})_i + \beta_2(\text{stent})_i + \beta_3(\text{female})_i + \beta_4(\text{diabetic})_i + \beta_5(\text{acutemi})_i + \beta_6(\text{ejectfrac})_i$$

p_i : Abciximab投与群への割り付け確率 = 傾向スコア



傾向スコアの特徴：

- 群の割り付けに関する多くの交絡因子を、一次元の情報に集約し調整する。（交絡因子の数イベントに比べて多いとき、次元の縮小に有用）
- 同じ傾向スコアを持つ標本は、処置群と対照群の多くの共変量が同時にバランスされている。

傾向スコアマッチング後の共変量の比較

Name	abcix=0 (n=275)	abcix=1 (n=275)	p-value
cardbill	14841.53 ± 902.45	16009.2 ± 536.21	0.267
height	171.5 ± 0.64	171.37 ± 0.63	0.884
ejecfrac	51.83 ± 0.62	51.41 ± 0.61	0.631
stent	174 (63.3%)	185 (67.3%)	0.37
female	99 (36%)	109 (39.6%)	0.429
diabetic	62 (22.5%)	63 (22.9%)	1
acutemi	18 (6.5%)	15 (5.5%)	0.72

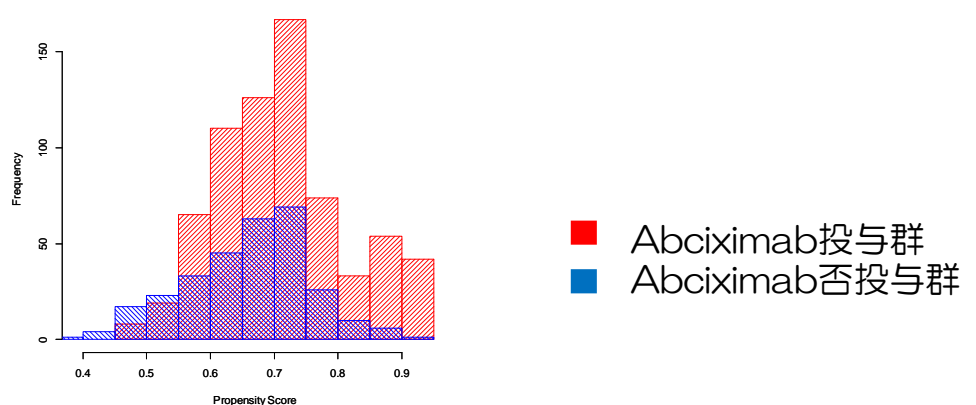
傾向スコアのマッチング (Propensity Score Matching)

傾向スコアが同じサンプルは、交絡因子の分布が同じになる傾向がある。 傾向スコアは実数値であるので、適当な有効桁数に丸める (caliper)。

処置群と対照群で、同じ傾向スコアを持つサンプルをマッチさせていく。もし、同じ傾向スコアを持つサンプルが複数あれば、順番はランダムにマッチングさせる。処置群と対照群で同じ傾向スコアを持つ相手がいない場合は、データから除かれる。

例：lindnerデータのPropensity Score

傾向スコアの分布



処置群、対照群のPS分布の重なり部分のサンプルがマッチングされる。重なっていない部分のサンプルは、マッチングデータからは削除される。

傾向スコアモデルのチェック

- 傾向スコアを推定するのに用いたロジスティック回帰モデルの、パフォーマンスを評価する
 - ⇒ Hosmer-Lemeshowの適合度検定
 - ⇒ ROC曲線
- 同じ傾向スコアを持つサンプルは、交絡因子の分布が同じになる傾向がある。

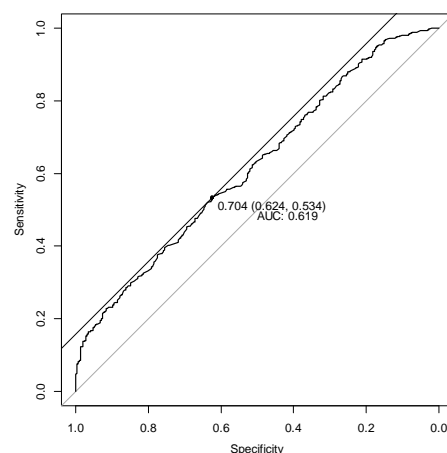
傾向スコアマッチング後の共変量の比較

Name	abcix=0 : (n=275)	abcix=1: (n=275)	p-value	standardized-difference
cardbill	14841.53 ± 14965.5	15778.05 ± 9073.46	0.375	0.076
height	171.5 ± 10.63	172.26 ± 10.95	0.409	0.07
ejecfrac	51.83 ± 10.35	52.13 ± 9.65	0.724	0.03
stent	174 (63.3%)	180 (65.5%)	0.656	0.046
female	99 (36%)	90 (32.7%)	0.473	-0.069
diabetic	62 (22.5%)	63 (22.9%)	1	0.009
acutemi	18 (6.5%)	16 (5.8%)	0.86	-0.03

ロジスティック回帰モデルによって傾向スコア (PS) を推定したら、その推定のパフォーマンスを評価する。

- Hosmer-Lemeshow検定:
P-value = 0.060 (ぎりぎり有意ではない。モデルは適合している)

- ROC解析とAUC
AUC = 0.619 あまりよくない



傾向スコアモデルのチェック（続き）

- マッチングされた後のサンプル数を確認する。処置群と対照群で傾向スコアの分布の乖離があまりに大きいと、マッチされるサンプルが少なくなり、解析に適さない。

⇒ 傾向スコアの乖離が大きい場合は、マッチングを行わず他の方法で傾向スコアを用いることを検討すべき。（傾向スコアによる層別化、重み付け、共変量としての利用、他）

- 本日の解析では、マッチング後のデータの比較に Welch's t-test と Fisher's exact test を用いた。しかし、マッチングされたペアを対応のあるデータと考えると、paired t-test, McNemar test を用いる場合もある。

傾向スコアを用いた解析法

- Matching: 傾向スコアマッチングを行い、その後は、処置群、対照群の別による **log-rank 検定**（生存時間解析）など、**単変量解析**を行う。
- 層別：処置群、対照群をプールした上で、傾向スコアの値で層別する（通常5層に分ける）。その上で、層別に解析を行う。（例：stratified log-rank test）
- 共変量としての傾向スコア：マッチングは行わず、全てのサンプルを用い、PSを共変量とした多変量モデルを当てはめる。

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \gamma G_i + \beta_1(\text{PS})_i, G_i = \begin{cases} 1: \text{treatment group} \\ 0: \text{control group} \end{cases}$$

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \gamma G_i + \beta_1(\text{PS})_i + \beta_2(\text{other})_i + \dots$$

例：lindnerデータのPropensity Score解析例

1. 傾向スコアを用いない、通常のロジスティック回帰

```
fit <- glm(lifepres ~ abcix, family=binomial, data=Data)
```

	Estimate	Std.Error	z value	Pr(> z)	
(Intercept)	2.9374	0.265	11.086	< 2.000E-16	***
abcix	1.197	0.4032	2.969	0.00299	**

2. 傾向スコアマッチング

```
fit1 <- glm(lifepres ~ abcix, family=binomial,  
data=Data.PS.matching)
```

	Estimate	Std.Error	z value	Pr(> z)	
(Intercept)	2.8526	0.2655	10.743	< 2.000E-16	***
abcix	1.6546	0.6384	2.592	0.00955	**

2. 傾向スコアマッチング

```
fit1 <- glm(lifepres ~ abcix, family=binomial,  
data=Data.PS.matching)
```

	Estimate	Std.Error	z value	Pr(> z)	
(Intercept)	2.8526	0.2655	10.743	< 2.000E-16	***
abcix	1.6546	0.6384	2.592	0.00955	**

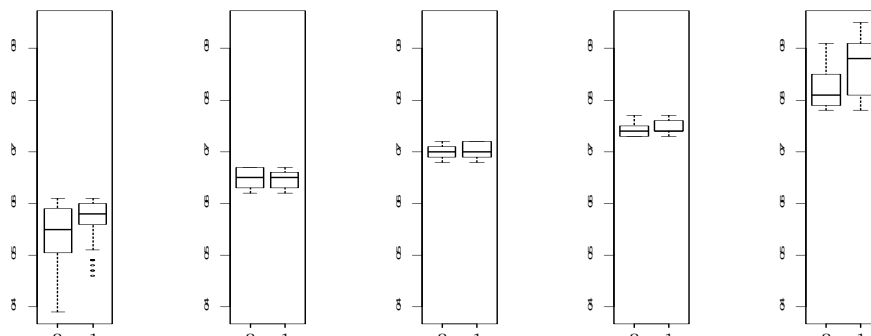
傾向スコアの推定に用いる共変量は、介入の有無にかかわる変数。傾向スコアの推定には用いられないが、アウトカムに影響を与える可能性のある変数を含める場合もある。

```
fit1.2 <- glm(lifepres ~ cardbill + abcix, family=binomial,  
data=Data.PS.matching)
```

	Estimate	Std.Error	z value	Pr(> z)	
(Intercept)	3.096	3.29E-01	9.401	< 2.000E-16	***
cardbill	0.000	1.01E-05	-1.429	0.15301	
abcix	1.649	6.39E-01	2.58	0.00989	**

2. 傾向スコアによる層別

データ全体を傾向スコアの値によって層別する（5つ程度）各層ごとに、単変量解析を行う。



PS		Estimate	Std.Error	z value	Pr(> z)
0% ~ 20%	abcix	0.7196	0.924	0.779	0.436
20% ~ 40%	abcix	0.4036	0.9257	0.436	0.663
40% ~ 60%	abcix	18.6216	2461.816	0.008	0.994
60% ~ 80%	abcix	1.9891	1.1657	1.706	0.0879
80% ~ 100%	abcix	1.5939	0.7034	2.266	0.02344 *

2015/11/19

東北大学 医学統計勉強会

19

3. 傾向スコアを共変量として用いる解析

```
fit3 <- glm(lifepres ~ abcix + PS, family=binomial, data=Data)
```

	Estimate	Std.Error	z value	Pr(> z)	
(Intercept)	5.6837	1.45	3.92	8.86E-05	***
abcix	1.42	0.4231	3.356	0.000789	***
PS	-4.0268	2.0431	-1.971	0.048728	*

```
fit3.2 <- glm(lifepres ~ cardbill + abcix + PS, family=binomial, data=Data)
```

	Estimate	Std.Error	z value	Pr(> z)	
(Intercept)	6.184	1.50E+00	4.123	3.75E-05	***
cardbill	0.000	9.33E-06	-2.384	0.017107	*
abcix	1.420	4.27E-01	3.329	0.000873	***
PS	-4.184	2.08E+00	-2.012	0.044271	*

2015/11/19

東北大学 医学統計勉強会

20

4. 傾向スコアの層を共変量として用いる解析

```
fit4 <- glm(lifepres ~ abcix + PS.strata, family=binomial,  
data=Data)
```

	Estimate	Std.Error	z value	Pr(> z)	
(Intercept)	3.10344	0.46811	6.63	3.36E-11	***
abcix	1.4152	0.42179	3.355	0.000793	***
PS.strataII	-0.17637	0.64583	-0.273	0.784778	
PS.strataIII	0.32065	0.74211	0.432	0.665686	
PS.strataIV	-0.01347	0.68398	-0.02	0.984289	
PS.strataV	-1.12426	0.59119	-1.902	0.057212	.

以下のモデルも可能

```
fit4.2 <- glm(lifepres ~ cardbill + abcix + PS.strata,  
family=binomial, data=Data)
```

5. IPTW (Inverse probability of treatment weighting)

傾向スコアの逆数を重み付けとして、各種多変量モデルを推定する。

$$W_i = \frac{Z_i}{PS_i} + \frac{1-Z_i}{1-PS_i}, Z_i = \begin{cases} 1: \text{treatment group} \\ 0: \text{control group} \end{cases}$$
$$= \begin{cases} \frac{1}{PS_i} : \text{treatment group (介入群では、介入する確率の逆数)} \\ \frac{1}{1-PS_i} : \text{control group (対照群では、介入しない確率の逆数)} \end{cases}$$

介入の有無の判断が確定的なサンプルは重みを軽く、介入するかしないか判断が微妙な場合については重みを重くして重点的に学習する。

5. **IPTW** (Inverse probability of treatment weighting) (続き)
 線形回帰モデル、ロジスティック回帰モデル、Cox比例ハザードモデルなどには、重み付き回帰モデルが存在するので、その重みに傾向スコアの逆数を用いる。

```
fit5 <- glm(lifepres ~ abcix, weight=W, family=binomial, data=Data)
```

	Estimate	Std.Error	z value	Pr(> z)	
(Intercept)	2.6946	0.1309	20.592	< 2.000E-16	***
abcix	1.4575	0.2881	5.059	4.21E-07	***

```
fit5.2 <- glm(lifepres ~ cardbill + abcix, weight=W, family=binomial, data=Data)
```

	Estimate	Std.Error	z value	Pr(> z)	
(Intercept)	3.086	1.69E-01	18.217	< 2.000E-16	***
cardbill	0.000	5.59E-06	-4.092	4.28E-05	***
abcix	1.464	2.89E-01	5.06	4.19E-07	***

傾向スコアを用いた解析法には、未だ最終的な答えはない。上記いずれかの方法を、少なくとも複数用いて、同じような結果が出ることを確かめるべき。少なくとも、**傾向スコアマッチング**と**IPTW**の二つは行い、同様の結果が得られることを示す。

傾向スコアを推定するための共変量に何を選ぶか、傾向スコアを用いた解析にどの手法を選ぶかで、結果が変わってくることもある。

Take Home Message

1. 因果効果と交絡因子
2. 傾向スコア
3. 傾向スコアマッチング
4. 傾向スコアモデルのチェック
5. 傾向スコアを用いた解析法
 - i. 傾向スコアマッチング：対応づけられたデータに対する単変量解析
 - ii. 層別
 - iii. 共変量としての傾向スコア
 - iv. 重み付け：Inverse-probability score-based weighted methods (IPTW)

参考文献

Rosenbaum & Rubin, “The central role of the propensity score in observational studies for causal effects” *Biometrika* (1983) 70 (1): 41-55.

星野 崇宏「調査観察データの統計科学—因果推論・選択バイアス・データ融合」
岩波書店 (2009/7/29) ISBN-10: 4000069721