

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

“Data! data! data!” he cried impatiently. “I can't make bricks without clay.”

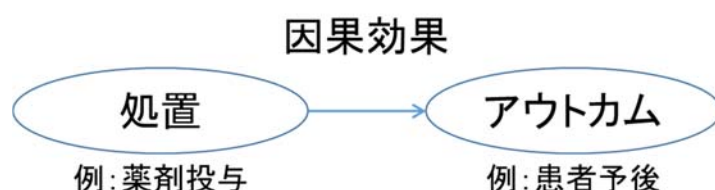
From The Adventure of the Copper Beeches, The Adventure of Sherlock Holmes.

「データ！データ！データ！」ホームズはいらいらして叫んだ。「粘土が無ければレンガは作れない」

第7回 傾向スコア

1. 因果効果と交絡因子

医学において新規薬剤が開発されたり、新たな手術技法が考案されたりしたとき、これらの新しい**処置** (=医学的介入) の**因果効果**を検証するための研究が行われます。



通常「処置」と「アウトカム」は、様々な**交絡因子** (= **confounding factor**, **予後因子**、**背景因子**、**共変量**) の影響を受けるため、処置を行った対象だけを検討 (**single arm**) してもアウトカムに対する効果は分かりません。因果効果の検証のための方法として、**無作為化比較試験 (Randomized Controlled Trial, RCT)** と呼ばれる方法があります。(RCTについては、次回詳しく取り上げます。) RCTにおいては、

- 新しい**処置を受ける群** (**treatment group**) と共に、標準的な処置 (あるいはプラセボ) を受ける**対照群** (**control group**) を置く。(control)
- 処置群と対照群の割り付けは、**無作為割り付け**による。(randomization)
- ヒトを対象とする場合、試験について十分な情報開示を行った上で、被験者から文書による同意を得る。(informed consent)

RCT では、処置群と対照群の無作為割り付けを行うことにより、処置群と対照群における処置の有無以外の要因は、全て偶然の誤差によるものと解釈出来ます。このことで交絡因子の影響を除き、真の因果効果のみを検証することが出来るようになります。

しかし、臨床における観察研究では、群の割り付けを無作為化することが困難で、処置群と対照群の間で交絡因子の分布に大きな違いが出ることがあります。例えば、ある薬剤の投薬の効果を検証する際、医師は適応がある患者、最も治

療を必要としている患者に優先して投与するかもしれません。また、喫煙のような有害な習慣の場合、無作為に割り付けて喫煙を強要するわけにも生きません。このように処置がランダムに割り付けられない場合、処置の因果効果は様々な交絡因子の影響を受けるため、交絡の影響をコントロールする研究デザインが必要になります。近年交絡をコントロールする手法として注目されている手法の一つが**傾向スコア (Propensity Score, PS)** になります。傾向スコアは「処置」の有無に関わる全交絡因子の情報を集約する手法で、層別や多変量解析など従来用いられた手法に代わるものとして盛んに用いられています。

例：lindner データ

オハイオ州 lindner センターで経皮的冠動脈形成術 (Percutaneous Coronary Interventions (PCIs)) を施行した 996 名の患者のうち、PCI のみの患者 298 名と PCI+ Abciximab (アブシキマブ) 投与の患者 698 名に対して、PCI 施行 6 ヶ月経過後の平均余命を比較したデータ (1997 年)。

lifepres : PCI 施行 6 ヶ月経過後の平均余命. 死亡=0, 生存=11.4

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

Name	abcix=0 (n=298)	abcix=1 (n=698)	p-value
cardbill	14614.22 ± 840.77	16126.68 ± 355.18	0.098
height	171.45 ± 0.61	171.44 ± 0.4	0.996
ejecfrac	52.29 ± 0.6	50.4 ± 0.39	0.009
stent	174 (58.4%)	492 (70.5%)	0
female	115 (38.6%)	231 (33.1%)	0.11
diabetic	80 (26.8%)	143 (20.5%)	0.031
acutemi	18 (6%)	125 (17.9%)	0

Lindner データからは、Abciximab 投与群と非投与群では、身長を除く全ての
変数に有意な差があることが分かります。

2. 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。

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

通常傾向スコアは**ロジスティック回帰モデル**によって推定されます。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**されたと解釈されます。従って、それ以降の解析では、
PS 予測に用いた共変量は使用されません。

傾向スコアの特徴：

- 群の割り付けに關する多くの交絡因子を、次元の情報に集約し調整する。
(交絡因子の数が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。

例：Lindner データの傾向スコア

ロジスティック回帰モデルによる傾向スコアの推定：

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

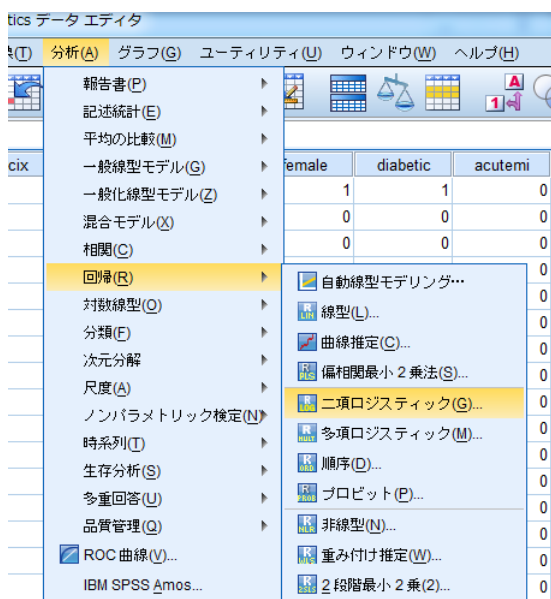
ただし、 p_i : Abciximab 投与群への割り付け確率 = **傾向スコア**。上の式の右
辺で、傾向スコアの推定には、height から ejectfrac までの共変量が用いられて
いますが、cardbill は用いられていないことに注意してください。これは、
cardbill が治療群選択後にかかった医療費を表していて、Abciximab 投与群への
割り付け確率=傾向スコアには關係がないからです。

このように、傾向スコアの推定では群の割り付けに關わる共変量は全て用い

られるのが原則であり、論文には**どの共変量が傾向スコアの推定に用いられたか明記する**必要があります。

例：SPSS による傾向スコアの推定（ロジスティック回帰による）

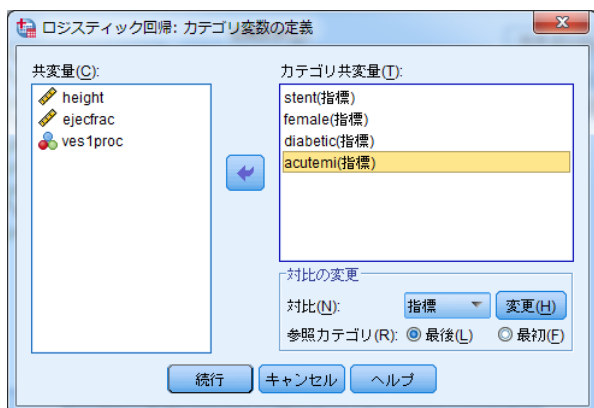
1. “lindner.sav” を読み込む。
2. 「分析」→「回帰」→「二項ロジスティック」



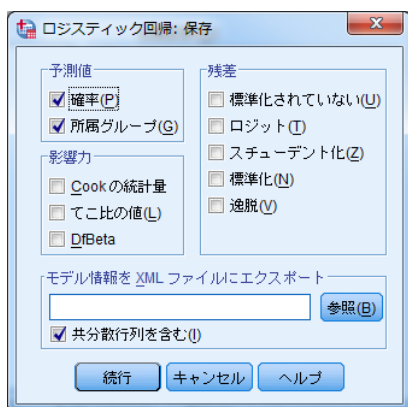
3. 従属変数：abcix、共変量：stent ~ ejectfrac



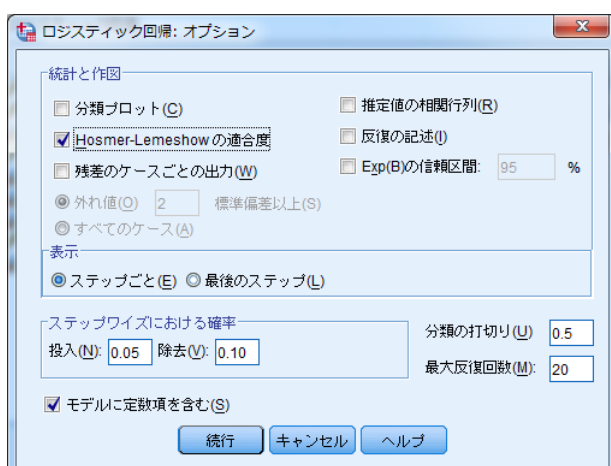
4. 「カテゴリ」ボタン：カテゴリ変数として、stent, female, diabetic, acutemi を選択 → 「続行」



5. 「保存」ボタン：予測値の「確率」、「所属グループ」を選択 → 「続行」



6. 「オプション」ボタン：「Hosmer-Lemeshow の適合度」にチェック → 「続行」



7. 「ロジスティック回帰」ウィンドで、「OK」

3. 傾向スコアマッチング (Propensity Score Matching)

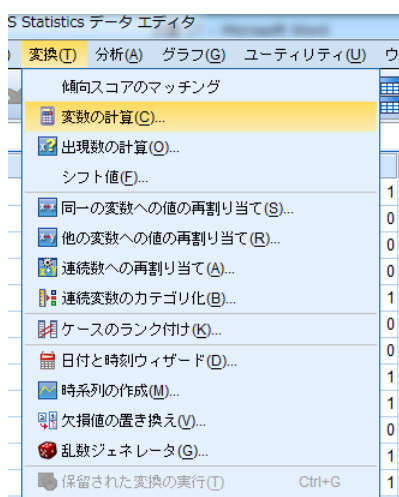
傾向スコアを用いた解析では、傾向スコアが同じサンプルは、交絡因子の分布が同じになる傾向があるとされています。(数学的な証明は省略します)

そこで、傾向スコアを推定した後、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ングさせることを考えます。傾向スコアは実数値ですので、まずこれを適当な桁数に丸めます。その後で、以下の方法でマッチング (対応付け) を行います。

- 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 (対応付け) する。
- もし、同じ傾向スコアを持つペアが複数存在するときは、順番はランダムにマッチングさせる。
- 同じ傾向スコアを持つサンプルが他方の群に存在しないときは、そのサンプルはマッチデータから削除する。

例：SPSS を用いた傾向スコアマッチング (lindner データ)

1. 傾向スコアの推定まで行う。
2. 「変換」 → 「変数の計算」



3. ロジスティック回帰で推定した群の割り付け確率=傾向スコアを、適当な桁数に丸める。新たに、丸めた後の変数名を“PS”とし、「数式」は以下の通り

$\text{RND}(\text{PRE}_1, 0.01)$



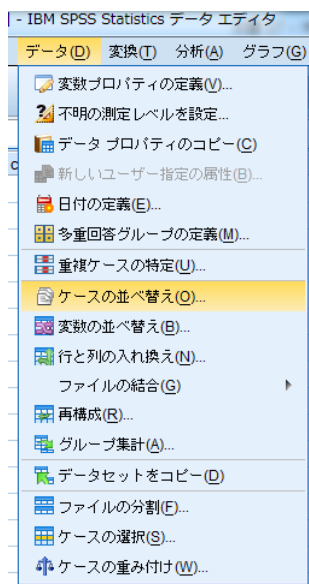
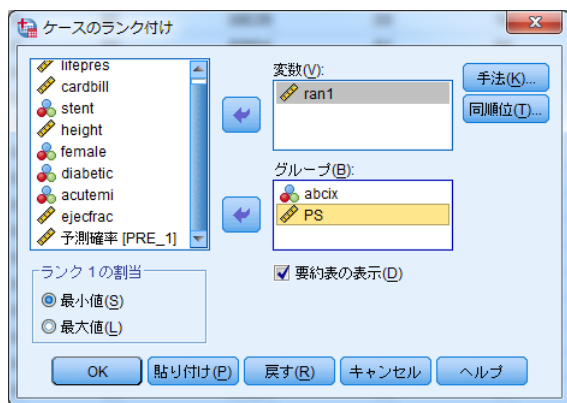
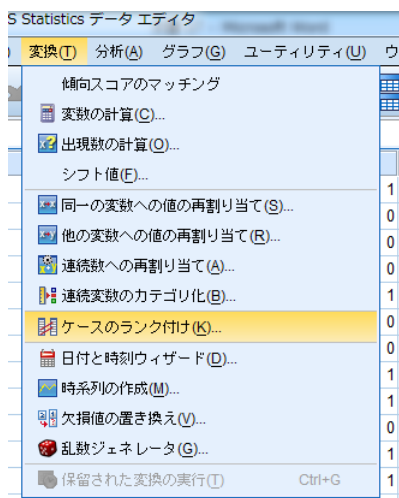
ここでは小数点以下第2位まで丸めたが、丸める程度は状況による。もし小数点以下の桁数が一桁のように少なければ、マッチするサンプルは増えますがマッチングの程度は荒くなります。逆に、小数点以下第3位、4位と高次まで求めると、マッチするサンプル数が少なくなってしまいます。

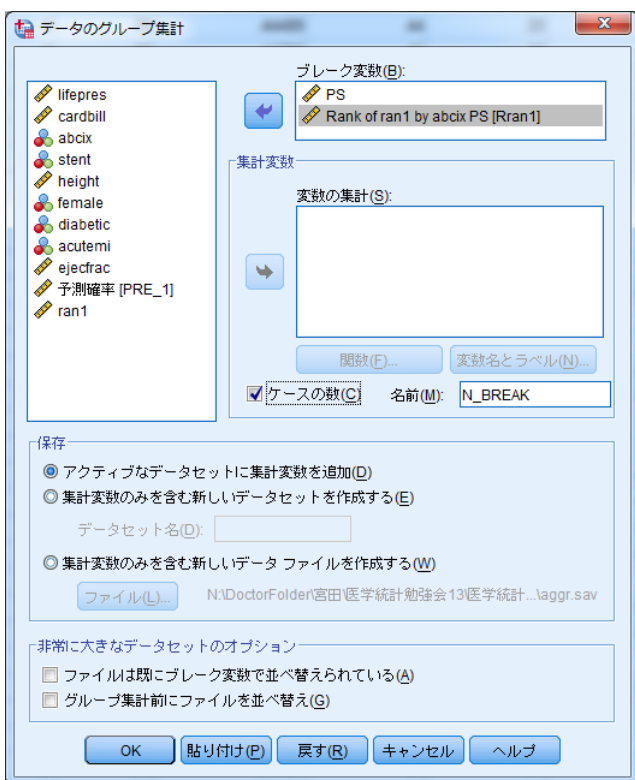
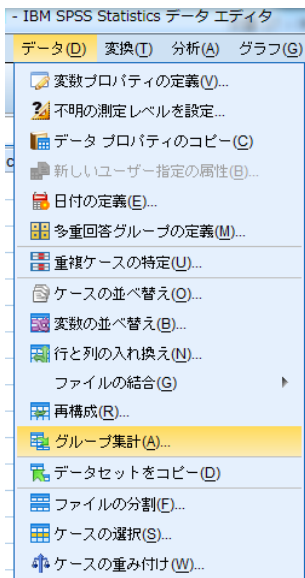
4. 傾向スコアが同じペアが複数あった場合、ランダムにマッチングを行うため、乱数を生成しておきます。

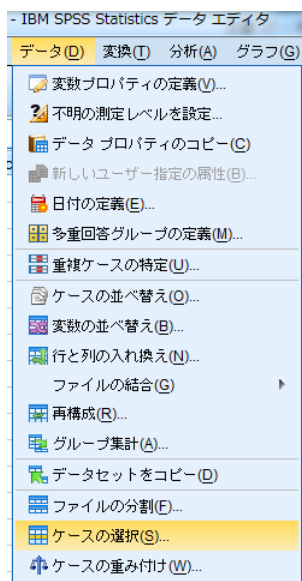
「変換」 → 「変数の計算」 → 変数名 : ran1 数式 : uniform(1)



5. 以下、傾向スコアの順にサンプルをソート（並べ直し）し、処置群と対照群で同じ傾向スコアを持つペアを抽出していきます。







4. 傾向スコアモデルのチェック

傾向スコアを推定した後、傾向スコアの粋手がうまくいっているかどうか検証する必要があります。

- 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
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

lindner データの場合、傾向スコアの推定に用いた共変量だけでなく、PS の推定に用いられなかった cardbill も結果的に差がなくなっています。

- 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、解析に適しません。

5. 傾向スコアを用いた解析法

以上述べてきたような方法で傾向スコアを推定した場合、推定した傾向スコアを用いた解析はいくつかの場合に分けられます。

5.1 傾向スコアマッチングを用いた解析

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ます。その場合、マッチする相手がいなかったサンプルは対応付けされたデータからは除かれます。従って、処置群と対照群で傾向スコアの分布が大きく異なるときは、マッチされたデータの数は大幅に少なくなり、元データの限られた一部分しか見ていないこととなります。そうではなく、処置群と対照群の多くのサンプルがマッチされた場合は、共変量の情報は全て傾向スコアに吸収されていますので、**群の違いのみに注目した単変量解析**に帰着します。回帰分析であれば単回帰、ロジスティック回帰モデルであれば、対応のあるデータに対する**条件付きロジスティック回帰**、生存時間解析であれば群の違いに注目した単変量の**log-rank検定**を行うこととなります。

5.2 マッチングを行わずに、傾向スコアを用いる解析

5.2.1. 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 Cox proportional hazard model**）

5.2.2. 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、群の割り付けを表すダミー変数 G と傾向スコアの値 PS を共変量とした多変量モデルを当てはめる。

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

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

上の式は、ロジスティック回帰を例に説明していますが、他のモデルでも同様です。第一式では、傾向スコアのみを共変量として用いています。一方第二式では傾向スコアの他にモデルに残っている共変量をモデルに入れています。ただし第二式のようなモデルでは、傾向スコアの推定に使った共変量をモデルに入れてはいけません。そのような共変量の情報は、全て傾向スコアに吸収されていると考えるからです。Lindner データの例でいえば、**cardbill** : PCI 施行後6ヶ月間の心臓病関連医療費（ドル）のように傾向スコアの推定に用いられなかった共変量が、付け加えられる可能性があります。

5.2.3. 重み付け : Inverse-probability score-based weighted methods (IPTW)

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。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いずれの解析モデルにも、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ということ。

傾向スコアを用いた解析法には、未だ最終的な答えはありません。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だと思われまます。

以上

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